2015

# Empirical Likelihood Confidence Band

Shihong Zhu
*University of Kentucky*, shihong2015@gmail.com

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Shihong Zhu, Student

Dr. Mai Zhou, Major Professor

Dr. Constance Wood, Director of Graduate Studies

Empirical Likelihood Confidence Band

---

DISSERTATION

---

A dissertation submitted in partial
fulfillment of the requirements for the
degree of Doctor of Philosophy in the
College of Arts and Sciences at the
University of Kentucky

By
Shihong Zhu
Lexington, Kentucky

Director: Dr. Mai Zhou, Professor of Statistics
Lexington, Kentucky 2015

ABSTRACT OF DISSERTATION

Empirical Likelihood Confidence Band

The confidence band represents an important measure of uncertainty associated with a functional estimator and empirical likelihood method has been proved to be a viable approach to constructing confidence bands in many cases. Using the empirical likelihood ratio principle, this dissertation developed simultaneous confidence bands for many functions of fundamental importance in survival analysis, including the survival function, the difference and ratio of survival functions, the hazards ratio function, and other parameters involving residual lifetimes. Covariate adjustment was incorporated under the proportional hazards assumption. The proposed method can be very useful when, for example, an individualized survival function is desired or a non-proportional hazards ratio is present in treatment effect assessment. Its large sample property was theoretically verified and the competitive finite sample performance was demonstrated by extensive numerical simulation studies.

KEYWORDS: Confidence Band, Empirical Likelihood, Likelihood Ratio Test, Scaled Chi-Square, Survival Analysis

Author's signature: Shihong Zhu

Date: June 30, 2015

Empirical Likelihood Confidence Band

By
Shihong Zhu

Director of Dissertation: Mai Zhou

Director of Graduate Studies: Constance Wood

Date: June 30, 2015

# ACKNOWLEDGMENTS

First of all, I am deeply grateful to my advisor Dr. Mai Zhou for his persistent encouragement and guidance. Being his student is simply a bless.

Secondly, I want to express my sincere gratitude to my dissertation committee members, Dr. Bonner, Dr. Fardo, Dr. Griffith, and Dr. Zheng, for their devotion to improving my dissertation.

Finally, I am very thankful to the faculty and staff in the Department of Statistics for their generous financial support and constant effort to maintain the department as an ideal place for my graduate study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

# Chapter 1 Introduction

## 1.1 Proportional Hazards Model

Cox's proportional hazards (PH) model (Cox, 1972) is unarguably the most popular regression tool in time-to-event analysis. It is very easy to fit and allows the regression coefficient to be readily interpreted as the log-hazard-ratio. This dissertation addresses a few important questions arising from practical applications of the PH model.

It is assumed that there exists a non-negative event time $X$, denoting the time until the occurrence of some event or failure, together with a vector of covariates $\boldsymbol{Z}$ of dimension $p$. Research interest lies in uncovering the quantitative relationship between $X$ and $\boldsymbol{Z}$. In practice, one can not fully observe $X$. Instead, there is an underlying non-negative censoring time $C$ such that one can only observe a censored time $T = \min(X, C)$, and a censoring indicator $\delta = I\{X \leq C\}$, valued at $1$ if death is observed and $0$ otherwise. This type of censoring is called right censoring and represents the most common censoring type. Other types of censoring include, but not limited to, left censoring, interval censoring, and a mix of multiple censoring types. This dissertation is focused on right-censored data only.

The hazard function plays a fundamental role in survival analysis. It describes the instantaneous rate of failure per time unit. Mathematically, the hazard function of the random variable $X$ is defined as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\Pr(X < t + \Delta t | X \geq t)}{\Delta t},$$

assuming the existence of the limit. Integrating $\lambda(\cdot)$ over time will lead to the cumulative hazard function $\Lambda(t) = \int_0^t \lambda(s)\mathrm{d}s$. For discrete distributions where $\lambda(t)$ is not well-defined, the cumulative hazard function can be defined by the following connection to the cumula-

tive distribution function $\Lambda(t) = \int_0^t 1/\{1 - F(s-)\}\mathrm{d}F(s)$.

The semi-parametric PH model relates the hazard function of $X$ to the covariates vector $\boldsymbol{Z}$ by assuming

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t)\exp(\boldsymbol{\beta}_0^\top \boldsymbol{Z}),$$

where $\boldsymbol{\beta}_0$ is the unknown $p$-dimensional vector of regression coefficients and $\lambda_0(t)$ is the unspecified baseline hazard function that can be theoretically interpreted as the hazard function of subjects with covariates vector $\boldsymbol{Z} = \mathbf{0}$. Under the assumption that $C$ be independent of $X$ given $\boldsymbol{Z}$ (non-informative censoring), the PH model can be fitted via the partial likelihood methodology (Cox, 1975).

One defining characteristic of the PH model is its proportional hazards assumption — the ratio of hazards between two subjects with different covariates is a constant in time. Valid use of the Cox model requires careful examination of the appropriateness of this assumption. A popular visual examination diagnostic tool is the deviance residual plot. Caution should be used when applying this tool since other model mis-specification, including outliers and non-linear relationship, could also lead to a rejection of the null hypothesis. When evidence against this assumption is deemed sufficient, many alternatives in the literature can be investigated. The standard approach is adding an interaction term of time and the covariate that does not support proportional hazards. Typically, the functional form of the interaction is parametrically specified, at the risk of model misspecification. While this also applies to the categorical covariate case, for categorical covariate, especially when the level of covariates is not too large, the more frequently used alternative is the *stratified Cox* model, which uses that particular categorical variable as a stratification variable and allows each stratum to have its own baseline. Specifically, if a categorical variable has $m$ levels

distinguished by the group indicator $g$, the stratified PH model postulates

$$\lambda(t|\boldsymbol{Z}, g = i) = \lambda_{i0}(t) \exp(\boldsymbol{\beta}_0^\top \boldsymbol{Z}), \quad i = 1, 2, \cdots, m. \tag{1.1}$$

The stratified PH model allows non-proportional hazards across the $m$ groups by allowing the baseline hazards to be group dependent. Meanwhile it maintains reasonable model parsimony by postulating a shared $\boldsymbol{\beta}$ across the groups. This model enables one to study the effect of covariates other than the stratification variable, without explicitly investigating the effect of the stratification variable, since the baseline hazards $\{\lambda_{0i}(\cdot)\}_{i=1}^n$ are profiled out as nuisance parameters in the model fitting stage.

## 1.2 Empirical Likelihood

### EL Confidence Interval

Empirical Likelihood (EL) is a non-parametric inference tool based on likelihood ratio principle to conduct hypothesis test and construct confidence intervals. It allows us to use likelihood method without having to assume a known parametric family of distributions. Its origination is credited to the seminal work by Owen (see Owen 1988; Owen 1990), who firstly applied it to obtain a non-parametric confidence interval for the population mean, and rigorously justified such use. For completeness, here we briefly introduce the framework of Owen's work. Suppose we are interested in the mean of the distribution $F$, denoted by $\mu_0$, and what we have is an identically and independently observed sample $(x_1, x_2, \cdots, x_n)$ following $F$. The EL method starts with a non-parametric likelihood ratio test of $H_0 : \mu_0 = \mu$ versus $H_a : \mu_0 \neq \mu$, based on the following likelihood ratio statistic:

$$\Re(\mu) = \frac{\sup_{p_1,\cdots,p_n} \left\{ \prod_{i=1}^n p_i | \sum_{i=1}^n p_i = 1, p_i \geq 0, \sum_{i=1}^n p_i x_i = \mu \right\}}{\sup_{p_1,\cdots,p_n} \left\{ \prod_{i=1}^n p_i | \sum_{i=1}^n p_i = 1, p_i \geq 0 \right\}}.$$

The product $\prod_{i=1}^n p_i$ is the likelihood of the observed data when the distribution is multinominal with weight $p_i$ assigned to $x_i$, therefore the denominator functions as the full

model maximum likelihood while the numerator functions as the reduced model maximum likelihood. The constraint $\sum_{i=1}^{n} p_i x_i = \mu$ is added to reflect the null hypothesis $\mu_0 = \int_{-\infty}^{\infty} t \, \mathrm{d}F(t) = \mu$. Assuming that $F$ allow a finite variance, Owen showed that $-2 \log \Re(\mu_0)$ converges in distribution to a chi-square distribution with one degree of freedom, a result parallel to the parametric Wilks's theorem (Wilks, 1938). By inverting this likelihood ratio, Owen obtained the so-called EL confidence interval for $\mu_0$, which is the collection of $\mu$ for which we fail to reject the null hypothesis.

Owen's work sparked a huge interest among the researchers. Influential works following and extending his work, among others, include Qin and Lawless (1994) who established the general methodology of EL inference with estimating equations, and DiCiccio et al. (1991) who showed that EL confidence intervals are Bartlett-correctable. See Owen (2001) for a general review on the EL method.

While early development of the EL method is focused on complete data, collective effort soon generalizes this method to incorporate censored data. Let $(T_i, \delta_i)$, $i = 1, 2, \cdots, n$, be the observed right-censored sample. Li (1995) studied the EL confidence interval for the survival probability with right censored data. The author applied the empirical likelihood formulation based on discrete distribution functions defined by

$$\mathrm{EL}_1(F) = \prod_{i=1}^{n} \Big\{ \mathrm{d}F(T_i) \Big\}^{\delta_i} \Big\{ 1 - F(T_i) \Big\}^{1-\delta_i},$$

where $\mathrm{d}F(t) = F(t+) - F(t-)$. Maximizing this likelihood over all distribution functions leads to the well-known Kaplan-Meier estimator. In addition to this formulation, Murphy (1995) also studied the so-called Possion extension formulation based on discrete cumulative hazard functions:

$$\mathrm{EL}_2(\Lambda) = \prod_{i=1}^{n} \Big\{ \mathrm{d}\Lambda(T_i) \Big\}^{\delta_i} \exp \Big\{ - \Lambda(T_i) \Big\},$$

where $\mathrm{d}\Lambda(t) = \Lambda(t+) - \Lambda(t-)$. Maximizing the above likelihood over all cumulative hazard functions leads to the Nelson-Aalen hazard estimator. Murphy (1995) showed that both formulations lead to meaningful confidence intervals for survival probabilities. Pan and Zhou (2002) took the second formulation and showed that this version is able to gracefully handle general and multiple constraints involving linear functionals of the cumulative hazard function. Multiple general constraints involving linear functionals of the cumulative distribution functions under either formulation are also possible but more challenging, both theoretically and computationally. Alternatively, Wang and Jing (2001) took the estimating equation approach to study linear functionals of the distribution function, but the limiting distribution of the resulting likelihood rato is no-longer pivotal. See Li et al. (2005) for a review of recent developments of EL method in survival analysis.

Numerous studies have endorsed the EL method as a powerful way to establish confidence intervals, by demonstrating the possession of some or all of the following features by EL confidence intervals:

- It is range preserving and transformation respecting due to its likelihood ratio based nature.
- adapting to skewed sampling distribution and producing confidence intervals that have data-driven shapes.
- not requiring a variance estimator since the likelihood ratio is a pivotal statistics. This self-studentizing ability is rather appealing when a stable variance estimator is hard to get.
- incorporating auxiliary information seamlessly by simply adding more constraints to the reduced model likelihood.

### EL Confidence Band

For functional parameters, such as a survival function over a given time interval, a simultaneous *confidence band* may be more relevant in some cases, since we may want to monitor the global uncertainty in an estimate of the survival function in that particular time interval while holding the global coverage rate controlled. Compared to the vast EL confidence intervals literature, fewer research on EL confidence bands has been done. One fact is that a functional parameter is targeted in less cases than a scalar (or vector) parameter. The other fact that may have also contributed to this phenomenon is that the transition from a confidence interval to a confidence band is generally technically non-trivial. An asymptotic confidence band necessitates uniform convergence results and therefore a finer asymptotic analysis. Moreover the limiting process *may* have a complex structure such that merely computing the critical value of the process becomes challenging.

Li et al. (1996) constructed the EL confidence band for the quantile function of a single population. Hollander and McKeague (1997) derived the EL confidence band associated with the Kaplan-Meier estimator. Using kernel smoothing and an weighted empirical likelihood, Li and Van Keilegom (2002) obtained the EL confidence band for the conditional survival function. Wang and Qin (2010) studied the EL confidence band for the distribution function with missing response. EL confidence bands have also been established for the ratio of two survival functions (McKeague and Zhao, 2002), the difference of two survival functions (Shen and He, 2006), and the ratio of two distribution functions (McKeague and Zhao, 2006).

## 1.3 Overview of the Dissertation

The work in this dissertation consists of three parts that are inter-related but each has its own challenge.

In Chapter 2, based on the hazard type EL formulation in Murphy (1995) and Pan and Zhou (2002), we construct the EL confidence interval and confidence band for the individualized survival function $S(t|\boldsymbol{Z})$ under the PH model. The hazard type likelihood formulation is a more natural choice than the distribution type formulation, since the PH model itself is formulated in terms of hazard functions.

The traditional normal approximation based interval suffers from poor finite sample coverage performance, and as a partial remedy, lots of transformations are proposed by many people. The possible transformations include, but not limited to, log, log-minus-log, arcsine-root, and logit transformation. Most Statistical softwares only provide pointwise confidence interval for the predicted survival curve but offer multiple transformations for users to choose from. For example, both SAS (lifetest procedure) and R (survival package) provide all transformations mentioned above but the former defaults log-minus-log transformation while the latter defaults the log transformation. The lack of an "optimal" transformation makes it unclear, for a particular application, which one should be used. Therefore it is of interest to derive a transformation respecting confidence interval that persistently produces reliable coverage accuracy. Moreover, the proposed method can be readily extended to construct confidence intervals for other parameters including residual survival functions and quantiles of the lifetime.

Also included in Chapter 2 are the large sample properties of the proposed method, some numeric studies, and a brief discussion on related computational issues.

In Chapter 3, we extend the study in Chapter 2 to make EL based inference on the difference of two individualized survival functions. More specifically we consider inference on $S(t|\boldsymbol{Z}_0, g = 1) - S(t|\boldsymbol{Z}_0, g = 2)$ under the stratified PH model. What motivates this study is the frequently encountered non-proportional hazards in observational studies be-

tween two treatments that we want to compare. The difference allows us to examine when the two treatments differ at the patient level, and which patient subgroup is likely to benefit the most. This is particularly important when one treatment has a higher early survival but lower long term survival. Zhang and Klein (2001) applied this measure to compare the survival prospects of patients treated with the conventional chemotherapy against patients treated by an allogeneic bone marrow transplants. Parzen et al. (1997) also studied this measure assuming no covariates. Both papers used the normal approximation method that may produce inferior confidence intervals and bands. In their paper, Parzen et al. (1997) explicitly expressed such concern!

Technically, the extension of EL inference from one survival function to the difference of two survival functions is difficult, basically because the constraints in the reduced model likelihood would be otherwise non-linear in the hazard increments. For EL inference, directly working with non-linear constraints is not impossible (Wood et al., 1996) but almost non-existent in the literature due to the accompanying theoretical and computational difficulty. To overcome the difficulty we adopt a linearization technique that has been seen in the literature. The linearization technique introduces an intermediate variable over which we need an additional maximization procedure. This makes the theoretical proof challenging. In order to alleviate the added complexity of numerical computation, we have used a plug-in EL formulation, in the cost of losing a pivotal statistic. We derive the large sample properties of the proposed likelihood ratio and evaluate the finite sample performance of the associated confidence bands and intervals by simulations. Evidently, the simulation studies favor our EL intervals and bands over the normal approximation alternative.

In Chapter 4, we apply the idea in Chapter 3 to study the ratio of cumulative hazards under model (1.1) with two strata. Specifically, we consider the EL inference on $\Lambda_{10}(t)/\Lambda_{20}(t)$. Wei and Schaubel (2008) applied this as a measure of cumulative treatment effect when the

treatment groups do not support proportional hazards, and outlined a few of its advantages as compared to other treatment effect measures handling non-proportional hazards.

Dong and Matthews (2012) considered the EL inference on this measure. However, their methodology is flawed and the resulting confidence interval is under-covering. Moreover, their simultaneous confidence band requires a Bootstrap calibration and is therefore rather numerically burdensome. Using essentially the same formulation as they did, we prove a stronger result that enables us to efficiently construct both of the confidence intervals and bands. The success of the proposed method is evidenced by simulations.

Finally, possible directions for future research are discussed in Chapter 5.

# Chapter 2 Individualized Survival Function under Cox Model

## 2.1 Introduction

In medical studies, Cox's model is mostly employed as an explanatory tool to assess the association between the risk factors and the survival time, since the regression coefficient can be readily interpreted as the log-hazard ratio. Estimation and inference on the regression coefficient can be accessed through the partial likelihood methodology (Cox, 1975). More often than not, it is also of interest to investigate the survival prospect of a certain subject and then the Cox model comes in as a predictive tool. For example, the Cox model is an important mathematical tool to predict survival in the patient with end-stage primary biliary cirrhosis (PBC) who has not undergone transplantation (Dickson et al., 1989). Such a model based objective survival prediction would be valuable for improving selection of patients for and the timing of the liver transplantation.

After fitting the Cox model, an estimator of the survival function is obtainable through the partial likelihood estimator of $\boldsymbol{\beta}_0$ and the Breslow estimator for the baseline cumulative hazard. Suppose the observations consist of $n$ independent replicates of $(T, \delta, \boldsymbol{Z})$, denoted by $(T_i, \delta_i, \boldsymbol{Z}_i), i = 1, 2, \cdots, n$. Let $N_i(t) = I[T_i < t, \delta_i = 1]$ be the counting process of $T_i$ and define

$$\boldsymbol{S}^{(k)}(\boldsymbol{\beta}, t) = n^{-1} \sum_{i=1}^{n} I[T_i \geq t] \boldsymbol{Z}_i^{\otimes k} \exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i), \quad k = 0, 1, 2.$$

Here $\boldsymbol{Z}^{\otimes 0} = 1$, $\boldsymbol{Z}^{\otimes 1} = \boldsymbol{Z}$, and $\boldsymbol{Z}^{\otimes 2} = \boldsymbol{Z}\boldsymbol{Z}^\top$. The conditional survival function $S_0(t|\boldsymbol{Z}_0) = \Pr(X \geq t|\boldsymbol{Z}_0)$ could then be consistently estimated by

$$\widehat{S}_0(t|\boldsymbol{Z}_0) = \exp\big\{ - \widehat{\Lambda}_0(t) \exp(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{Z}_0)\big\},$$

where

$$\widehat{\Lambda}_0(t) = n^{-1} \sum_{i=1}^{n} \int_0^t \mathrm{d}N_i(s)/S^{(0)}(\widehat{\boldsymbol{\beta}}, s)$$

is the Breslow hazard estimator and $\widehat{\boldsymbol{\beta}}$, the partial likelihood estimator, solves the following score equation

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\infty} \Big\{ \boldsymbol{Z}_i - \boldsymbol{S}^{(1)}(\boldsymbol{\beta}, t)/S^{(0)}(\boldsymbol{\beta}, t) \Big\} \, \mathrm{d}N_i(t) = \mathbf{0}.$$

When the scientific interest is on a specific time spot, a confidence interval accompanying the estimated survival is desirable to calibrate the uncertainty of the estimation. When the interest is on the survival function over the entire time horizon, a simultaneous confidence band is of more relevance. Using different techniques, both Tsiatis (1981) and Lin et al. (1994) obtained the pointwise and simultaneous confidence bands by showing that $\sqrt{n}\big\{ \widehat{\boldsymbol{S}}_0(t|\boldsymbol{Z}_0) - \boldsymbol{S}_0(t|\boldsymbol{Z}_0) \big\}$ converges weakly to a mean zero Gaussian process with a rather complex covariance function. Since the survival probability is confined in $[0, 1]$, when the sample size is small, the approximation by a Gaussian distribution might not be ideal (Nair, 1984). To achieve better finite sample approximation, many transformations were suggested, including the log, log-log, logit, and arcsine-root transformations (Kalbfleisch and Prentice, 1980), but selecting the optimal transformation may be challenging since the optimal transformation, if exists, may vary from case to case and is generally unknown in practice. In addition, when dealing with such parameters as the difference of two survival functions or the ratio of two survival functions, it is unclear whether there is any simple transformation to improve the finite sample normal approximation. Motivated by these facts, we contribute in this manuscript to the toolbox an EL based method to construct the confidence band that is transformation respecting.

The remainder of this chapter is organized as follows: Section 2.2 investigates how the empirical likelihood method could be applied to address the current problem. Also presented are the asymptotic property of the likelihood ratio test statistic. Section 2.3 provides

a re-sampling scheme to construct the simultaneous confidence band. Section 2.4 devotes to a simulation study to compare the newly proposed method with its existing competitors in terms of the finite sample coverage probability and average length. An application to the Mayo Clinic PBC dataset is also included, followed by a brief discussion in Section 2.5. Finally we prove the theorem in Section 2.6.

## 2.2 Empirical Likelihood Confidence Band

Empirical likelihood and Cox's partial likelihood methodology has close relationship, as illustrated by Johansen (1983) that the profile empirical likelihood of the regression coefficients is equivalent to the partial likelihood. This fact makes inference within the Cox model employing the empirical likelihood method very natural.

There are several papers adopting the empirical likelihood method to obtain confidence regions of the regression coefficients. In particular, Qin and Jing (2001) derived an empirical likelihood confidence region of the regression coefficients assuming a known baseline hazard. To avoid the assumption of a known baseline hazard, Zhao and Jinnah (2012) applied a plug-in type empirical likelihood where the baseline hazard is semi-parametrically estimated. Ren and Zhou (2011) defined the full empirical likelihood function through the baseline survival function and also obtained such a confidence region. Zhou (2006) showed that the empirical likelihood could be employed to obtain an enhanced estimator of the regression coefficients when there is a certain form of side information on the baseline hazard. However, all of these works were focused on the regression coefficients and thus could not be directly used to draw inference on the infinitely dimensional baseline hazard or the associated survival function.

**Empirical Likelihood Ratio**

We consider the EL inference on the survival function $S_0(t|\boldsymbol{Z}_0)$ for $t$ in a prespecified interval. For notation simplicity and ease of presentation, we may assume $\boldsymbol{Z}_0 = 0$ and study the baseline survival $S_0(t) = \exp\{-\Lambda_0(t)\}$; otherwise, we shift the original covariates to $\tilde{\boldsymbol{Z}}_i = \boldsymbol{Z}_i - \boldsymbol{Z}_0$, and proceed as if $\tilde{\boldsymbol{Z}}_i$ were the observed covariates. The consequence of this shift is that the new baseline survival will refer to the survival of subjects with original vector of covariates $\boldsymbol{Z}_0$. This enables us to get rid of $\boldsymbol{Z}_0$ and focus on the baseline survival. Furthermore, due to the transformation respecting property of the EL confidence interval, we could firstly obtain the EL confidence interval of $\Lambda_0(t)$, and then transform the obtained interval back to one for $S_0(t)$.

In terms of the hazard function, the likelihood of the observed data is given by

$$\mathrm{EL}(\boldsymbol{\beta}, \Lambda) = \prod_{i=1}^{n} \Big\{ \mathrm{d}\Lambda(T_i) \exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i) \Big\}^{\delta_i} \exp\Big\{ -\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i) \Lambda(T_i) \Big\},$$

where $\mathrm{d}\Lambda(t) = \Lambda(t) - \Lambda(t-)$ denotes the increment of $\Lambda(t)$ at time $t$. The empirical likelihood methodology seeks to maximize the above likelihood function in the family of discrete hazards $\Lambda(t)$ that only has nonnegative increments on the observed time spots. To that end, let $\boldsymbol{w} = (w_1, \cdots, w_n)$ be the vector of nonnegative increments of $\Lambda(t)$ and write the likelihood function as

$$\mathrm{EL}(\boldsymbol{\beta}, \boldsymbol{w}) = \prod_{i=1}^{n} \Big\{ w_i \exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \Big\}^{\delta_i} \exp\Big\{ -\exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \sum_{j=1}^{n} I[T_j \geq T_i] w_j \Big\}. \quad (2.1)$$

It's a well known result (Johansen, 1983) that the above likelihood attains its maximum at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ and $w_i^0 = \mathrm{d}\widehat{\Lambda}_0(T_i)$. In addition to this unconstrained maximum likelihood, the EL

ratio also requires a constrained likelihood that takes into consideration the hypothesis we want to test. Since a discrete hazard function is simply the cumulative summation of its increments, we define the constrained maximum empirical likelihood at $\Lambda(t)$ as

$$\text{EL}_c\{\Lambda(t)\} \;=\; \sup_{\boldsymbol{\beta},\boldsymbol{w}} \text{EL}(\boldsymbol{\beta},\boldsymbol{w}) \quad \text{subject to} \quad \sum_{i=1}^{n} I[T_i \le t] w_i = \Lambda(t). \tag{2.2}$$

The empirical likelihood ratio statistic at $\Lambda(t)$ is then defined as

$$\Re\{\Lambda(t), t\} = \text{EL}_c\{\Lambda(t)\}/\text{EL}(\widehat{\boldsymbol{\beta}}, \boldsymbol{w}^0). \tag{2.3}$$

In order to solve the constrained optimization problem (2.2), we resort to the Lagrangian Multiplier method and write the target function as

$$G \;=\; \log \text{EL}(\boldsymbol{\beta},\boldsymbol{w}) - n\lambda \left\{ \sum_{i=1}^{n} g_i(t) w_i - \Lambda(t) \right\},$$

where $g_i(t) = I[T_i \le t]$ and $\lambda$ is the Lagrange Multiplier. Differentiate $G$ with respect to $w_i$, set the resulting derivative to zero, we then obtain

$$w_i \;=\; \frac{1}{n} \frac{\delta_i}{\boldsymbol{S}_0(\boldsymbol{\beta}, T_i) + \lambda g_i(t)}, \quad 1 \le i \le n. \tag{2.4}$$

Equating $\partial G/\partial \boldsymbol{\beta}$ and $\partial G/\partial \lambda$ with zero and making use of (2.4), we reduce the constrained optimization problem to the following set of equations of $\boldsymbol{\beta}$ and $\lambda$

$$\begin{cases} \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} \dfrac{\delta_i g_i(t)}{S^{(0)}(\boldsymbol{\beta}, T_i) + \lambda g_i(t)} - \Lambda(t) = 0, & (2.5) \\[2ex] \displaystyle\sum_{i=1}^{n} \delta_i \left\{ \boldsymbol{Z}_i - \dfrac{\boldsymbol{S}^{(1)}(\boldsymbol{\beta}, T_i)}{S^{(0)}(\boldsymbol{\beta}, T_i) + \lambda g_i(t)} \right\} = 0. & (2.6) \end{cases}$$

Note that (2.5) simply says the constraint should be satisfied while (2.6) is a variant of the partial likelihood score equation in the presence of constraint. Numerical methods such as Newton's method is required to solve these two equations for $\boldsymbol{\beta}$ and $\lambda$. Generally, the solution will depend on $t$ and $\Lambda(t)$, but for notation simplicity, we simply denote the solution by $(\boldsymbol{\beta}^*, \lambda_n)$. Once we obtain $(\boldsymbol{\beta}^*, \lambda_n)$, we could calculate the constrained maximum

likelihood by plugging the expression for $w_i$ into (2.1), which leads to the following log-likelihood ratio statistic:

$$\log \Re\{\Lambda(t), t\} \;=\; \sum_{i=1}^{n} \delta_i \left\{ (\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}^\top)\boldsymbol{Z}_i - \log \frac{S^{(0)}(\boldsymbol{\beta}^*, T_i) + \lambda_n g_i(t)}{S^{(0)}(\widehat{\boldsymbol{\beta}}, T_i)} \right\} + n\lambda_n \Lambda(t).$$

**Asymptotic Properties**

Let $\boldsymbol{s}^{(k)}(\boldsymbol{\beta}, t)$ be the expected value of $\boldsymbol{S}^{(k)}(\boldsymbol{\beta}, t)$, $k = 0, 1, 2$. We define

$$\begin{aligned}
\sigma^2(t) &= \int_0^t \left\{ \boldsymbol{s}^{(0)}(\boldsymbol{\beta}_0, u) \right\}^{-1} \mathrm{d}\Lambda_0(u), \\
\boldsymbol{h}(t) &= \int_0^t \boldsymbol{s}^{(1)}(\boldsymbol{\beta}_0, u) \big/ \boldsymbol{s}^{(0)}(\boldsymbol{\beta}_0, u)\, \mathrm{d}\Lambda_0(u), \\
\boldsymbol{\Sigma} &= \int_0^\infty \left\{ \frac{\boldsymbol{s}^{(2)}(\boldsymbol{\beta}_0, u)}{s^{(0)}(\boldsymbol{\beta}_0, u)} - \frac{\boldsymbol{s}^{(1)2}(\boldsymbol{\beta}_0, u)}{s^{(0)2}(\boldsymbol{\beta}_0, u)} \right\} \mathrm{d}\Lambda_0(u), \\
v(t, s) &= \sigma^2(t \wedge s) + \boldsymbol{h}(t)^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{h}(s). \qquad (2.7)
\end{aligned}$$

Notice that $\boldsymbol{\Sigma}$ is the information matrix and $\sigma^2$ is the variance of the Breslow cumulative hazard estimator when $\boldsymbol{\beta}_0$ is known a priori.

To derive the asymptotic distribution of the EL ratio $\Re\{\Lambda_0(t), t\}$, we impose the following regularity conditions:

(C1) The triplet $\{(T_i, \delta_i, \boldsymbol{Z}_i)\}_{i=1}^n$ are independent and identically distributed and $\boldsymbol{Z}_1$ has bounded support.

(C2) The information matrix $\boldsymbol{\Sigma}$ at $\boldsymbol{\beta}_0$ is positive definite.

(C3) $\Pr(\delta_i = 1) > 0$. In words, the probability of observing an event is positive.

These conditions are commonly seen in the literature studying large sample behavior of the PH model. It might be possible to relax condition (C1) following Andersen and Gill

(1982), but the identical distribution and bounded covariate assumptions significantly simplify our subsequent technical proof. Condition (C2) is essential for asymptotic normality of the partial likelihood estimator. Condition (C3) rules out the possibility that all observations are censored when the sample size is large enough. Finally, these four conditions, combined together, are stronger than the assumptions made by Andersen and Gill (1982) and thus ensure the asymptotic properties of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Lambda}_0(t)$ given in that paper. We now summarize the main result in the following theorem and defer the proof to Section 2.6.

**Theorem 2.1.** *Let $\tau_0 < \tau$ be positive numbers such that $0 < \Lambda_0(\tau_0) < \Lambda_0(\tau) < \infty$. Under conditions (C1)-(C3), $-2\log\Re\{\Lambda_0(t), t\}$ converges weakly to $U^2(t)/v(t,t)$ in $D[\tau_0, \tau]$, where $U(t)$ is a mean-zero Gaussian process with covariance function $v(t,s)$ defined in (2.7).*

**Corollary 2.1.1.** *For a given $t_0$ such that $0 < \Lambda_0(t_0) < \infty$, under the conditions (C1)-(C3), $-2\log\Re\{\Lambda_0(t_0), t_0\} \xrightarrow{D} \chi^2(1)$, as the sample size goes to infinity. Therefore, an asymptotic level $\alpha$ confidence interval for $\Lambda_0(t_0)$ is given by $\{\theta | -2\log\Re(\theta, t_0) \leq \chi^2_\alpha(1)\}$, where $\chi^2_\alpha(1)$ denotes the upper $\alpha$ percentile of $\chi^2(1)$.*

Due to the transformation respecting property, the EL confidence interval for the survival function $S_0(t_0)$ is given by

$$\left\{\theta \middle| -2\log\Re\{-\log\theta, t_0\} \leq \chi^2_\alpha(1)\right\},$$

or equivalently, the one for $\Lambda_0(t_0)$ transformed by the map: $x \to \exp(-x)$. Unlike the normal approximation based confidence interval, this EL ratio based confidence interval is transformation respecting. In addition, it does not require estimating the variance of $\widehat{S}_0(t_0)$, always contained in $[0, 1]$, almost surely asymmetric about $\widehat{S}_0(t_0)$.

**Corollary 2.1.2.** *Let* $C = C_\alpha(\tau_0, \tau)$ *be the upper* $\alpha$ *percentile of* $\sup_{\tau_0 \le t \le \tau} |U(t)|/\sqrt{v(t,t)}$. *The following gives a level* $\alpha$ *EL simultaneous confidence band for* $S_0(t)$ *on* $[\tau_0, \tau]$*:*

$$B_\alpha(\tau_0, \tau) = \Big\{ \big(S(t), t\big) \Big| \sup_{\tau_0 \le t \le \tau} -2 \log \Re\big\{ -\log S(t), t \big\} \le C^2, t \in [\tau_0, \tau] \Big\}.$$

## 2.3 Monte-Carlo Simulation for Confidence Bands

In order to put into Corollary 2.1.2 into application, the very first thing one needs to do is to calculate the critical value $C$. It is unclear if it can be obtained analytically or by a simple Monte-Carlo simulation, since the limiting process $U(t)$ dose not have independent increments, as indicated by the fact that $v(t,s) \neq v(t,t)$ when $0 < t < s$.

To overcome this difficulty, we observe that $U(t)$ could be decomposed into two independent components each of which can be easily simulated. Specifically, let $\boldsymbol{G}$ be a multivariate normal random variable with mean zero and covariance matrix $\boldsymbol{\Sigma}^{-1}$, and $V(t)$ be a Brownian motion with variance $\sigma^2(t)$ and independent of $\boldsymbol{G}$. Then one can easily show that $W(t) = \boldsymbol{h}(t)^\top \boldsymbol{G} + V(t)$ has the same distribution with $U(t)$. Due to its independent increment property, $V(t)$ can be easily simulated by sequentially generating its independent increments, so is $W(t)$.

Since we do not know the parameters $\sigma^2(t)$, $\boldsymbol{h}(t)$, $\boldsymbol{\Sigma}$, and $v(t,t)$, we need to use their estimates obtained by replacing in their definitions the unknown quantities with the corresponding empirical counterparts. For example, we could estimate $\sigma^2(t)$ by $\widehat{\sigma}^2(t) = \int_0^t 1/S^{(0)}(\widehat{\boldsymbol{\beta}}, u)\, \mathrm{d}\widehat{\Lambda}_0(u)$ and define $\widehat{\boldsymbol{h}}(t)$, $\widehat{\boldsymbol{\Sigma}}$, and $\widehat{v}(t,t)$ similarly. According to Andersen and Gill (1982), $\widehat{\sigma}^2(t)$, $\widehat{\boldsymbol{h}}(t)$, and $\widehat{v}(t,t)$ are uniformly consistent estimators of $\sigma^2(t)$, $\boldsymbol{h}(t)$, and $v(t,t)$, respectively. We observe that $\widehat{v}(t,t)$ and $\widehat{\sigma}^2(t)$ are piecewise constant and only jump at the event times, therefore we only need to sample $W(t)$ at those distinct event times. This amounts to simulating the sample path of $\tilde{W}(t)$ by treating the observed data

as fixed, where

$$\tilde{W}(t) = \widehat{\boldsymbol{h}}(t)^\top \boldsymbol{G} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i g_i(t) G_i}{S^{(0)}(\widehat{\boldsymbol{\beta}}, T_i)}, \tag{2.8}$$

and $G_1, \cdots, G_n$ are i.i.d. standard normal random variables independent of $\boldsymbol{G}$ which is a p-dimensional multivariate normal with covariance matrix $\widehat{\boldsymbol{\Sigma}}^{-1}$. So we could obtain, say $N$, sample paths of $\tilde{W}(t)$, denoted by $\{\tilde{W}_i(t)\}_{i=1}^N$, calculate $C_i = \max_{1 \le i \le n} |\tilde{W}(T_i)| / \sqrt{v(T_i, T_i)}$, and approximate $C$ by the $\alpha$ percentile of $\{C_i\}_{i=1}^N$.

**Algorithm 1** Estimating C for Survival Functions

**for** each $i \in \{1, 2, \cdots, N\}$ **do**

Generate $\boldsymbol{G}_i \leftarrow \text{MVN}(\boldsymbol{0}, \widehat{\boldsymbol{\Sigma}}^{-1})$

**for** $j \in \{1, 2, \cdots, n\}$ **do**

Generate $G_{ij} \leftarrow \text{N}(0, 1)$

**end for**

Set $\tilde{W}_i(t) \leftarrow \widehat{\boldsymbol{h}}(t)^\top \boldsymbol{G}_i + (\sqrt{n})^{-1} \sum_{i=1}^n \delta_i g_i(t) G_{ij} / S^{(0)}(\widehat{\boldsymbol{\beta}}, T_i)$

Set $C_i \leftarrow \sup_{\tau_0 \le t \le \tau} |\tilde{W}(t)| / \sqrt{\widehat{v}(t,t)}$

**end for**

**return** $C \leftarrow$ upper $\alpha$ percentile of $\{C_1, C_2, \cdots, C_N\}$

Note that the supremum in the algorithm can be only attained on some of the observed $T_i$s since both $\tilde{W}$ and $\widehat{v}$ are piece-wise constant functions. In practice, it is important to montior how the estimated $C$ changes with $N$. $N$ needs to be large enough to stabilize the estimate. It needs to be even larger if higher confidence level is desired, because the more extreme percentile is always harder to estimate.

This simulation strategy turns out to be quite similar to the one proposed by Lin et al. (1994)

with the only difference in the way we treat the multivariate normal random variable $\boldsymbol{G}$. Instead of directly sampling $\boldsymbol{G}$ from the normal distribution with covariance matrix $\widehat{\boldsymbol{\Sigma}}^{-1}$, Lin et al. (1994) approximates it by another summation of $n$ independent random variables involving the partial likelihood score function. Specifically, they replace $\boldsymbol{G}$ in (2.8) by

$$n^{-1/2}\sum_{i=1}^{n}\widehat{\boldsymbol{\Sigma}}^{-1}\delta_i\Big\{\boldsymbol{Z}_i-\boldsymbol{S}^{(1)}(\widehat{\boldsymbol{\beta}},T_i)/S^{(0)}(\widehat{\boldsymbol{\beta}},T_i)\Big\}G_i$$

based on a modification to the martingale representation of the Breslow cumulative hazard estimator.

Compared to the approach in Lin, our derivation of the simulation method is more intuitively justifiable and features a simpler form. Regarding the sampling of $\boldsymbol{G}$, we only need to deal with a $p-$vector, but they have to deal with a $n-$vector, this slight difference may be magnified when $n$ is large and a very large $N$ is needed (for example when we want $99\%$ confidence bands). Despite all these differences, simulation studies show that both methods are pretty fast and produce results consistent with each other.

## 2.4 Numerical Study and Real Data Example

### Coverage of Confidence Intervals

We compare by simulation studies the proposed EL confidence interval with its normal approximation based counterparts in terms of the empirical coverage accuracy and average length. We present the results of two numerical experiments.

In the *first* experiment, the underlying survival time $X$ follows a Weibull distribution with shape parameter $3$ and scale parameter $\exp(0.1Z)$, $Z$ being evenly spaced between $[-1,1]$. The censoring time $C$ follows an Exponential distribution with rate parameter $\alpha$ that will

be adjusted to achieve several prespecified levels of censoring. The parameter of interest is assumed to be $S_0(0.9|Z = 0.1)$ whose true value is approximating $0.9$, representing an early life stage. The *second* design differs the first one only in that the survival time follows an Exponential distribution with rate $\exp(0.5Z)$ and the parameter of interest is $S_0(1.5|Z_0 = 0.8)$ whose true value is about $0.1$, indicating a late life stage.

Under each scenario, $5000$ simulation replicates were obtained and the observed coverage probabilities together with the average lengths are summarized in Table 2.1 and Table 2.2, based on which we make a few observations:

- As expected, the average length of all methods decreases when the sample size grows or when the censoring rate decreases. Meanwhile, the coverage probability gets closer to the nominal $95\%$ level.

- In both cases, the plain confidence interval has noticeably lower than nominal coverage probabilities, even when the sample size is up to $100$ and the censoring rate as low as $10\%$, which justifies the need for an appropriate transformation. While the log transformation improves the coverage probabilities in the second case, it does not help much in the first case. It is also found that the Arcsine-root transformation does improve the coverage probability but not as impressively as the Log-log and Logit transformations. Although the Arcsine-root transformation yields a shorter average length, but it is more of a reflection of its smaller coverage probabilities.

- We proceed to compare the EL method with the Log-log and Logit transformations. We can see that all three methods yield comparable coverage probabilities. However, in the first case, EL clearly has the shortest average length; in the second case, Log-log has the shortest length, but its advantage over the EL method is negligible compared with the advantage of EL over the Logit transformation. This fact indicates

Table 2.1: Case One —- Coverage percentages and average lengths of CIs for survival probabilities*.

| $\alpha$ | C | $n$ | Plain | Log | Log-Log | Logit | Arcsine | EL |
|---|---|---|---|---|---|---|---|---|
| 0.12 | 10% | 20 | 88.04(0.22) | 87.72(0.22) | 87.88(0.29) | 88.48(0.28) | 88.92(0.23) | 89.48(0.24) |
| | | 30 | 86.06(0.20) | 86.24(0.20) | 93.58(0.24) | 94.06(0.23) | 93.26(0.21) | 95.48(0.21) |
| | | 50 | 92.02(0.17) | 92.08(0.17) | 95.92(0.19) | 96.30(0.18) | 92.58(0.17) | 95.04(0.17) |
| | | 100 | 92.26(0.12) | 92.34(0.12) | 95.70(0.13) | 95.72(0.13) | 94.08(0.12) | 94.60(0.12) |
| 0.4 | 30% | 20 | 86.42(0.23) | 86.34(0.22) | 85.42(0.31) | 86.08(0.29) | 87.00(0.24) | 87.18(0.25) |
| | | 30 | 85.20(0.21) | 84.56(0.22) | 93.30(0.26) | 93.82(0.25) | 93.86(0.22) | 94.40(0.22) |
| | | 50 | 89.88(0.18) | 90.06(0.18) | 96.26(0.20) | 96.64(0.19) | 93.64(0.18) | 95.16(0.18) |
| | | 100 | 92.74(0.13) | 92.46(0.13) | 96.12(0.14) | 96.06(0.13) | 94.42(0.13) | 95.20(0.13) |
| 0.8 | 50% | 20 | 81.62(0.23) | 81.66(0.22) | 81.22(0.33) | 81.80(0.30) | 82.44(0.25) | 84.00(0.27) |
| | | 30 | 86.46(0.22) | 85.56(0.21) | 91.16(0.28) | 91.54(0.26) | 91.68(0.23) | 91.90(0.23) |
| | | 50 | 89.50(0.19) | 88.72(0.19) | 96.16(0.21) | 96.44(0.21) | 92.28(0.19) | 93.50(0.19) |
| | | 100 | 91.94(0.14) | 92.04(0.14) | 95.76(0.15) | 95.78(0.15) | 93.78(0.14) | 94.74(0.14) |

* based on $5000$ replicates; $C$ is the approximate censoring rate; $n$ is the sample size.

Table 2.2: Case Two —- Coverage percentages and average lengths of CIs for survival probabilities*.

| $\alpha$ | C | $n$ | Plain | Log | Log-Log | Logit | Arcsine | EL |
|---|---|---|---|---|---|---|---|---|
| 0.11 | 10% | 20 | 84.56(0.33) | 92.08(0.77) | 96.58(0.38) | 95.00(0.53) | 89.82(0.37) | 94.76(0.40) |
| | | 30 | 87.68(0.28) | 92.80(0.51) | 95.60(0.32) | 94.84(0.40) | 91.90(0.31) | 94.62(0.33) |
| | | 50 | 89.48(0.24) | 94.82(0.33) | 95.50(0.25) | 96.08(0.29) | 93.08(0.25) | 94.20(0.26) |
| | | 100 | 92.64(0.18) | 94.50(0.21) | 95.30(0.18) | 95.20(0.20) | 94.14(0.18) | 94.98(0.18) |
| 0.25 | 20% | 20 | 84.38(0.35) | 91.90(0.83) | 96.06(0.40) | 94.78(0.58) | 89.30(0.39) | 94.78(0.42) |
| | | 30 | 86.40(0.30) | 93.06(0.59) | 95.56(0.34) | 95.42(0.44) | 90.84(0.33) | 94.36(0.35) |
| | | 50 | 89.76(0.25) | 94.04(0.36) | 95.42(0.27) | 95.66(0.31) | 92.94(0.26) | 94.78(0.27) |
| | | 100 | 91.26(0.19) | 94.68(0.22) | 94.76(0.19) | 95.54(0.21) | 93.40(0.19) | 94.16(0.19) |
| 0.43 | 30% | 20 | 83.90(0.37) | 90.86(0.89) | 96.10(0.44) | 94.38(0.63) | 88.66(0.42) | 94.24(0.45) |
| | | 30 | 85.68(0.32) | 92.02(0.68) | 95.92(0.36) | 94.48(0.49) | 90.18(0.35) | 94.14(0.38) |
| | | 50 | 89.50(0.26) | 93.62(0.42) | 95.26(0.29) | 95.42(0.34) | 92.40(0.28) | 94.76(0.29) |
| | | 100 | 91.12(0.20) | 93.82(0.25) | 94.52(0.21) | 94.92(0.23) | 93.16(0.21) | 94.70(0.21) |

* based on 5000 replicates; $C$ is the approximate censoring rate; $n$ is the sample size.

a certain "robustness" property of the EL method in the sense that it always offers competitive performance when compared with the best performing transformation.

The conclusion is that certain transformation may drastically improve the performance of the plain confidence interval, but it seems that the optimal transformation may vary from case to case. For a particular application, when systematic studies on which transformation produces the best result is unavailable, the application of the EL method should be emphasized, due to its "robustness" explained in the third observation we made above.

**Primary Biliary Cirrhosis Data Analysis**

The dataset is from the Mayo Clinic trial in primary biliary cirrhosis of the liver conducted between 1974 and 1984. A total of 424 PBC patients, referred to Mayo Clinic during that ten-year interval, met eligibility criteria for the randomized placebo controlled trial of the drug D-penicillamine. The first 312 cases in the data set participated in the randomized trial and contain largely complete data. The additional 112 cases did not participate in the clinical trial, but consented to have basic measurements recorded and to be followed for survival. Six of those cases were lost to follow-up shortly after diagnosis, so the data here are on an additional 106 cases as well as the 312 randomized participants.

The PBC data were used by Dickson et al. (1989) to build a Cox model (Mayo PBC model) to predict survival for individual patients. The Mayo PBC model is extremely useful in medical management by aiding in the selection of patients for and timing of orthotopic liver transplantation. The model includes five covariates, log(bilirubin), log(protime), log(albumin), age and edema. The estimated model parameters and the corresponding standard errors in Table 2.3 are from Lin et al. (1994). Using this model, we demonstrate our pointwise and simultaneous EL confidence bands of the predicted survival function over the interval of the first and the last observed deaths.

Table 2.3: Parameter estimates and standard errors in the Mayo PBC model.

| Variable | Estimate | Error | Est/SE |
|---|---|---|---|
| Age | 0.0394 | 0.0077 | 5.1508 |
| log (Albumin) | -2.5328 | 0.6482 | -3.9074 |
| log (Bilirubin) | 0.8707 | 0.0826 | 10.5372 |
| Oedema | 0.8592 | 0.2711 | 3.1688 |
| log (Prothrombin) | 2.3797 | 0.7666 | 3.1043 |

Following Lin et al. (1994), we consider a hypothetical patient with 51 years of age, $3.4$ gm/dl serum albumin, $1.8$ mg/pl serum bilirubin, $10.74$ seconds of prothrombin time and no oedema. All simultaneous confidence bands are constructed using the same critical value $3.074$ based on $10000$ simulation replicates. Figure 2.1 displays the pointwise and simultaneous $95\%$ EL confidence bands, as well as the plain normal simultaneous confidence bands. We can see that the long term survival estimate for this hypothetical patient is subject to substantial uncertainty, as shown by the wide confidence bands at the right end. It is also found that the simultaneous EL band is very consistent with, but slightly narrower than, the plain normal band. Figure 2.2 contrasts the simultaneous EL band with the Log, Log-log transformed normal bands. While the EL band is also similar to the Log-log transformed band, it is quite different from the Log transformed band. These observations seem to support the role of the EL confidence band as a very attractive alternative to its normal approximation counterparts and give us confidence in its applications.

Figure 2.1: 95% EL pointwise and simultaneous confidence bands.

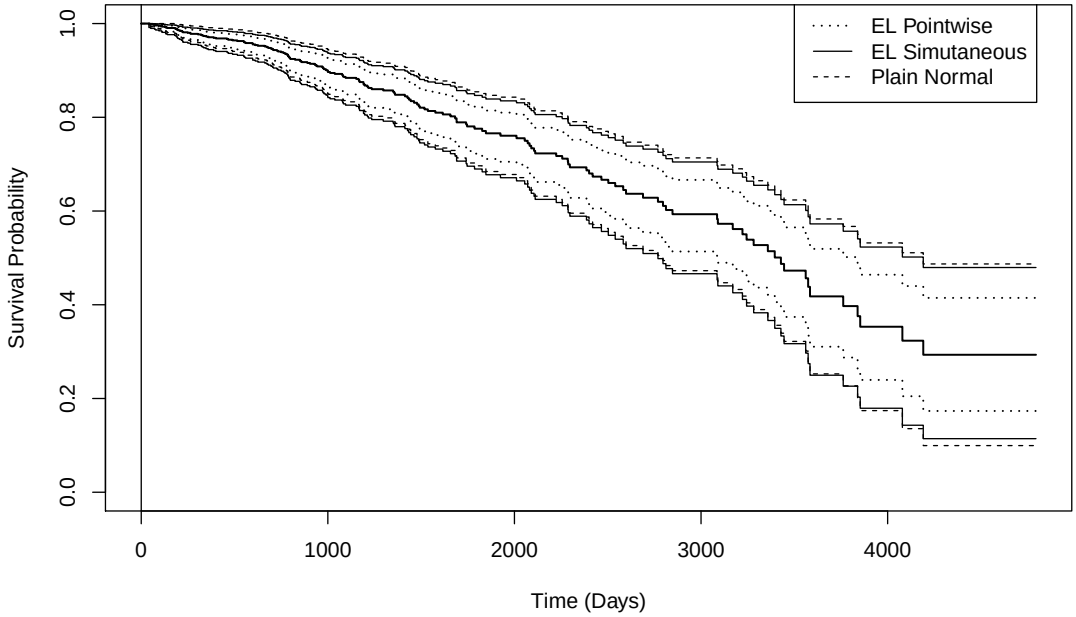
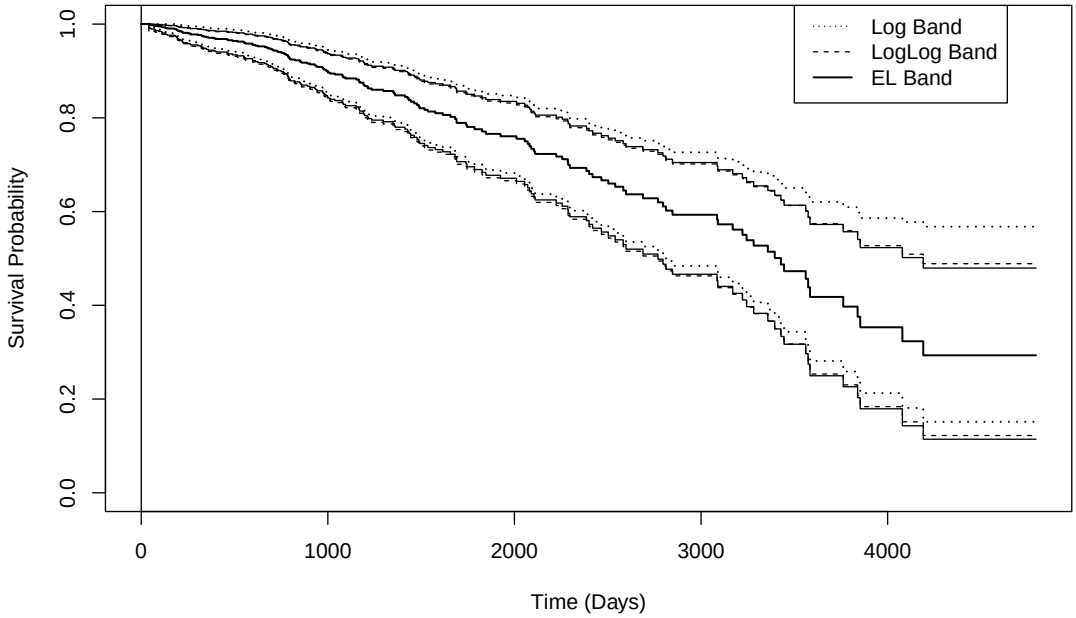



Figure 2.2: 95% EL, Log, and Log-Log transformed simultaneous confidence bands.




## 2.5 Summary

In this paper, by inverting the empirical likelihood ratio test on the hazard function, we derived both the pointwise and simultaneous confidence bands of the survival function for any

prespecified vector of covariates under the Proportional Hazards Model. The proposed EL band features several appealing properties, including transformation respecting and range respecting. Especially, the pointwise confidence interval does not require variance estimation. The performance of the pointwise confidence interval is compared with its normal approximation alternatives through a simulation study. It is found that the EL confidence interval has very competitive empirical coverage probabilities and average lengths. The simultaneous EL band is illustrated using the Mayo PBC dataset where the EL band is found to be very consistent with the normal approximation band with or without "log-log" transformation.

The proposed method is flexible enough to study other parameters of fundamental importance in survival analysis, including but not limited to

- residual survival function $S(t|\boldsymbol{Z}, X \geq \tau)$. It describes the conditional survival prospect given that the subject under consideration has survived up to some time instant, and therefore is more appropriate for existing patients.
- lifetime quantiles. It is the inverse of the cumulative distribution function and provides another perspective towards the survival prospect.

In many applications, the covariate may be longitudinally collected, resulting in time-dependent covariate. While the EL inference procedure in this chapter is able to handle time-dependent covariate with little effort, caution should be exercised regarding the interpretation of the predicted curve and its confidence limits (Fisher and Lin, 1999).

The proposed EL confidence band could also carry a weight function to adjust the relative width of the band at different time spots. Specifically, let $g_n(t) \geq 0$ be a weight function that converges in probability to a nonnegative bounded function $g(t)$ uniformly in $t \in [\tau_0, \tau]$, then a weighted EL confidence band will be given by

$$\Big\{\big(S(t),t\big)\Big| -2g_n(t)\log\Re\big[-\log S(t),t\big]\le C_\alpha^2(g)\Big\},$$

where $\Pr\big\{\sup_{\tau_0\le t\le\tau} g_n(t)\tilde{W}^2(t)/\widehat{v}(t,t)\le C_\alpha^2(g)\big\}=\alpha$, the probability being evaluated by simulation.

Clearly, the weighted EL band will be narrower (wider) than the unweighted band where $g_n(t)$ is larger (smaller). However, it is impossible to connect the weighted EL band with, for example, the well-known equi-precision normal approximation band described in Nair (1984), which is because the EL band does not have an explicit formula to allow examination of proportionality between the length of the pointwise interval and that of the simultaneous band.

A well-known weakness of the EL method is its computational intensity. The EL method we proposed in this paper is no exception. In order to calculate the confidence limits, we need to repeatedly calculate the likelihood ratio statistic for various hypothesized hazard values, which involves repeatedly solving (2.5) and (2.6) by an iterative method. Therefore, the EL method is substantially more computationally demanding than the normal approximation method. When the sample size is large, it could take a very long time to draw the entire simultaneous confidence band. The bottleneck is solving the set of equations (2.5) and (2.6). Although the Newton's method using $(\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}},\lambda=0)$ as the starting point succeeds in most cases, it does fail when the hypothetical value $\Lambda(t)$ is far away from the observed value $\widehat{\Lambda}(t)$. To overcome this problem, we could start by solving the following

equation using starting value ($\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}, \lambda = 0$):

$$\begin{cases} \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \dfrac{\delta_i g_i(t)}{S^{(0)}(\boldsymbol{\beta}, T_i) + \lambda g_i(t)} - \tilde{\Lambda}(t) = 0, \\ \\ \displaystyle\sum_{i=1}^{n} \delta_i \left\{ \boldsymbol{Z}_i - \dfrac{\boldsymbol{S}^{(1)}(\boldsymbol{\beta}, T_i)}{S^{(0)}(\boldsymbol{\beta}, T_i) + \lambda g_i(t)} \right\} = 0. \end{cases}$$

Here $\tilde{\Lambda}(t) = 0.5\{\widehat{\Lambda}(t) + \Lambda(t)\}$. Note that, compared to $\Lambda(t)$, $\tilde{\Lambda}(t)$ is closer to $\widehat{\Lambda}(t)$, which is why we make this adjustment. Once we obtain the solution to the above equations we could use this solution as the starting value for the original equations (2.5) and (2.6). We may need to repeat this procedure until convergence. This detour is pretty robust in our numerical study but could be time consuming.

## 2.6 Proof

One of the challenges is showing the uniform consistence of $\boldsymbol{\beta}^*$ as an estimator of $\boldsymbol{\beta}_0$, or $\lambda_n$ as an estimator of $0$. One of these two convergence results implies the other, so we start with $\boldsymbol{\beta}^*$. This is done by showing that $\boldsymbol{\beta}^*$ should be uniformly "close" to $\widehat{\boldsymbol{\beta}}$. Provided that $\widehat{\boldsymbol{\beta}}$ is consistent, we know $\boldsymbol{\beta}^*$ is also consistent. Once we obtain the consistence of both $\boldsymbol{\beta}^*$ and $\lambda_n$, we can then use Taylor expansion to the Lagrange multiplier equations, carefully controlling the remainder terms. This leads to asymptotic expressions for $\boldsymbol{\beta}^*$ and $\lambda_n$. Using a second Taylor expansion to the likelihood ratio statistic, one can then prove the theorem.

For notation simplicity, let $S_i(\boldsymbol{\beta}) = S^{(0)}(\boldsymbol{\beta}, T_i)$, $i = 1, 2, \cdots, n$. Define $\ell(\boldsymbol{\beta}, \lambda) = G(\boldsymbol{\beta}, \boldsymbol{w})$ with $w_i = n^{-1}\delta_i/\{S_i(\boldsymbol{\beta}) + \lambda g_i(t)\}$, where $G(\boldsymbol{\beta}, \lambda) = \log \mathrm{EL}(\boldsymbol{\beta}, \lambda)$. The unconstrained log-empirical likelihood $G(\widehat{\boldsymbol{\beta}}, \boldsymbol{w}^0)$ could be therefore denoted by $\ell(\widehat{\boldsymbol{\beta}}, 0)$ and the constrained log-empirical likelihood $\log \mathrm{EL}_c\{\Lambda_0(t)\}$ by $\ell(\boldsymbol{\beta}^*, \lambda_n)$. For every $t \in [\tau_0, \tau]$, define

$$r(t) = \ell(\widehat{\boldsymbol{\beta}}, 0) - \sup_{\boldsymbol{w}} \Big\{ G(\widehat{\boldsymbol{\beta}}, \boldsymbol{w}) \Big| \sum_{i=1}^{n} w_i g_i(t) = \Lambda_0(t) \Big\}.$$

$r(t)$ is simply the profile log-empirical likelihood ratio at $\Lambda_0(t)$ with $\boldsymbol{\beta}$ fixed at $\widehat{\boldsymbol{\beta}}$ in the constrained likelihood.

We will need the uniform boundedness of $r(t)$, namely, $\sup_{\tau_0 \le t \le \tau} r(t) = O_p(1)$. The proof of this result will be given at the end. Keeping this result in mind, we will prove the main theorem in several steps.

**Step 1**: The uniform consistency of $\boldsymbol{\beta}^*$.

***Proof:*** On the one hand, it follows from the definition of the constrained likelihood that

$$\begin{aligned}\ell(\boldsymbol{\beta}^*, \lambda_n) &= \sup_{\boldsymbol{\beta}} \Big\{ \sup_{\boldsymbol{w}} G(\boldsymbol{\beta}, \boldsymbol{w}) \Big| \sum_{i=1}^{n} g_i(t) w_i = \Lambda_0(t) \Big\}, \\ &\geq \sup_{\boldsymbol{w}} \Big\{ G(\widehat{\boldsymbol{\beta}}, \boldsymbol{w}) \Big| \sum_{i=1}^{n} g_i(t) w_i = \Lambda_0(t) \Big\},\end{aligned}$$

therefore $0 \leq \ell(\widehat{\boldsymbol{\beta}}, 0) - \ell(\boldsymbol{\beta}^*, \lambda_n) \leq r(t)$, hence $\sup_{\tau_0 \leq t \leq \tau} \big\{ \ell(\widehat{\boldsymbol{\beta}}, 0) - \ell(\boldsymbol{\beta}^*, \lambda_n) \big\} = O_p(1)$.

On the other hand, $\ell(\boldsymbol{\beta}^*, 0) = \sup_{\boldsymbol{w}} G(\boldsymbol{\beta}^*, \boldsymbol{w})$ while

$$\ell(\boldsymbol{\beta}^*, \lambda_n) = \sup_{\boldsymbol{w}} \Big\{ G(\boldsymbol{\beta}^*, \boldsymbol{w}) \Big| \sum_{i=1}^{n} g_i(t) w_i = \Lambda_0(t) \Big\}.$$

Because the constrained maximum will never exceed the unconstrained maximum, so $\ell(\boldsymbol{\beta}^*, \lambda_n) \leq \ell(\boldsymbol{\beta}^*, 0)$. By definition of $\widehat{\boldsymbol{\beta}}$, we have $\ell(\widehat{\boldsymbol{\beta}}, 0) \geq \ell(\boldsymbol{\beta}_0, 0)$, therefore

$$\ell(\boldsymbol{\beta}_0, 0) - \ell(\boldsymbol{\beta}^*, 0) \leq \ell(\widehat{\boldsymbol{\beta}}, 0) - \ell(\boldsymbol{\beta}^*, \lambda_n),$$

which then implies

$$\begin{aligned}\sup_{\tau_0 \leq t \leq \tau} \Big\{ \ell(\boldsymbol{\beta}_0, 0) - \ell(\boldsymbol{\beta}^*, 0) \Big\} &\leq \sup_{\tau_0 \leq t \leq \tau} \Big\{ \ell(\widehat{\boldsymbol{\beta}}, 0) - \ell(\boldsymbol{\beta}^*, \lambda_n) \Big\} \\ &= O_p(1). \qquad (2.9)\end{aligned}$$

Let $X_n(\boldsymbol{\beta}) = n^{-1} \big\{ \ell(\boldsymbol{\beta}_0, 0) - \ell(\boldsymbol{\beta}, 0) \big\}$. Then by the uniform boud we have just derived, we have $\sup_{\tau_0 \leq t \leq \tau} X_n(\boldsymbol{\beta}^*) \leq O_p(n^{-1})$. Andersen and Gill (1982) showed that $X_n(\boldsymbol{\beta})$ converges in probability uniformly in any compact set to a nonrandom convex function $X(\boldsymbol{\beta})$ which has a unique minimum 0 at $\boldsymbol{\beta}_0$. It is now a standard argument using the convexity of $X_n(\boldsymbol{\beta})$ to conclude that $\boldsymbol{\beta}^*$ converges to $\boldsymbol{\beta}_0$ in probability uniformly in $t$. □

**Step 2**: $\lambda_n$ converges in probability to zero uniformly in $t \in [\tau_0, \tau]$.

***Proof:*** For each $t \in [\tau_0, \tau]$, $\lambda_n$ satisfies $n^{-1} \sum_{i=1}^{n} \delta_i g_i(t) / \big( b_i + \lambda_n \big) = \Lambda_0(t)$ with $b_i = S_i(\boldsymbol{\beta}^*)$. Without loss of generality, we assume $T_i$ has been ordered in increasing order, thus

$b_1 \geq b_2 \geq \cdots \geq b_n > 0$. Let $\widehat{\Lambda}^*(t) = \sum_{i=1}^n n^{-1}\delta_i g_i(t)/b_i$, which is merely a variant of the Breslow baseline hazard estimator with $\widehat{\boldsymbol{\beta}}$ replaced by $\boldsymbol{\beta}^*$. A similar argument to those in Theorem 3.4 of Andersen and Gill (1982) shows that $\widehat{\Lambda}^*(t)$ is also a uniform consistent estimator of $\Lambda_0(t)$. If $\lambda_n < 0$, we have

$$\begin{aligned}\Lambda_0(t) &= \frac{1}{n}\sum_{i=1}^n \frac{\delta_i g_i(t)}{b_i}\frac{b_i}{b_i - |\lambda_n|} \geq \frac{b_1}{b_1 - |\lambda_n|}\frac{1}{n}\sum_{i=1}^n \frac{\delta_i g_i(t)}{b_i} \\ &= \frac{b_1}{b_1 - |\lambda_n|}\widehat{\Lambda}^*(t).\end{aligned}$$

Similarly, when $\lambda_n > 0$ we have

$$\begin{aligned}\Lambda_0(t) &= \frac{1}{n}\sum_{i=1}^n \frac{\delta_i g_i(t)}{b_i}\frac{b_i}{b_i + |\lambda_n|} \leq \frac{b_1}{b_1 + |\lambda_n|}\frac{1}{n}\sum_{i=1}^n \frac{\delta_i g_i(t)}{b_i} \\ &= \frac{b_1}{b_1 + |\lambda_n|}\widehat{\Lambda}^*(t).\end{aligned}$$

These two inequalities forces $|\lambda_n| \leq \Lambda_0^{-1}(t)\big|\Lambda_0(t) - \widehat{\Lambda}^*(t)\big|b_1$. Boundedness of $Z_i$ and uniform consistency of $\boldsymbol{\beta}^*$ imply that $b_1$ is uniformly bounded in probability. Therefore $\lambda_n$ converges in probability to zero uniformly in $t \in [\tau_0, \tau]$. □

**Step 3**: Asymptotic representation of $\lambda_n$ and $\boldsymbol{\beta}^*$ as shown in (2.16) and (2.17), respectively.

***Proof:*** We start with a Taylor expansion to equation (3.5) and (3.6). To that end, define

$$\begin{aligned}m_n(\boldsymbol{\beta}, \lambda) &= \frac{1}{n}\sum_{i=1}^n \frac{\delta_i g_i(t)}{S_i(\boldsymbol{\beta}) + \lambda g_i(t)} - \Lambda_0(t), \\ l_n(\boldsymbol{\beta}, \lambda) &= \frac{1}{n}\sum_{i=1}^n \delta_i \left\{ Z_i - \frac{S^{(1)}(\boldsymbol{\beta}, T_i)}{S_i(\boldsymbol{\beta}) + \lambda g_i(t)} \right\}.\end{aligned}$$

Since $\boldsymbol{\beta}^*$ and $\lambda_n$ converges to $\boldsymbol{\beta}_0$ and $0$, we may apply a Taylor expansion to $m_n(\boldsymbol{\beta}, \lambda)$ and obtain

$$0 = m_n(\boldsymbol{\beta}^*, \lambda_n) = m_n(\widehat{\boldsymbol{\beta}}, 0) + \left(\frac{\partial m_n}{\partial \boldsymbol{\beta}}, \frac{\partial m_n}{\partial \lambda}\right)^\top \Big|_{(\boldsymbol{\beta}_m, \lambda_m)} (\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}, \lambda_n),$$

where $(\boldsymbol{\beta}_m, \lambda_m)$ lies on the line segment connecting $(\widehat{\boldsymbol{\beta}}, 0)$ and $(\boldsymbol{\beta}^*, \lambda_n)$. Using the uniform consistency of $\boldsymbol{\beta}^*$ and $\lambda_n$, together with the boundedness on $s^{(k)}(\boldsymbol{\beta}, t)$ for $\boldsymbol{\beta}$ in any compact neigborhood of $\boldsymbol{\beta}_0$ and $t \in [\tau_0, \tau]$, we can show

$$\sup_{\tau_0 \le t \le \tau} \left| \frac{\partial m_n(\boldsymbol{\beta}_m, \lambda_m)}{\partial \boldsymbol{\beta}} + h(t) \right| \xrightarrow{p} 0 \tag{2.10}$$

and

$$\sup_{\tau_0 \le t \le \tau} \left| \frac{\partial m_n(\boldsymbol{\beta}_m, \lambda_m)}{\partial \lambda} + \sigma^2(t) \right| \xrightarrow{p} 0. \tag{2.11}$$

Since the proof of these two conclusions are similar, we only provide the details for the second conclusion. Note that

$$\frac{\partial m_n(\boldsymbol{\beta}_m, \lambda_m)}{\partial \lambda} \;=\; -\frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i g_i(t)}{S_i^2(\boldsymbol{\beta}_m)} \left\{ \frac{S_i(\boldsymbol{\beta}_m)}{S_i(\boldsymbol{\beta}_m) + \lambda_m} \right\}^2.$$

Since $g_i(t)$ is zero-one valued, we only need to see the property of $S_i(\boldsymbol{\beta}_m)/\{S_i(\boldsymbol{\beta}_m) + \lambda_m\}$ in the case of $g_i(t) = 1$. However, when $g_i(t) = 1$, $T_i \le t \le \tau$, therefore $S_i(\boldsymbol{\beta}_m) \ge S^{(0)}(\boldsymbol{\beta}_m, \tau)$, which converges in probability to $s^{(0)}(\boldsymbol{\beta}_0, \tau) > 0$ uniformly in $t \in [\tau_0, \tau]$ due to the boundedness of $Z_i$ and the uniform consistency of $\boldsymbol{\beta}_m$. Therefore, $S_i(\boldsymbol{\beta}_m)/\{S_i(\boldsymbol{\beta}_m) + \lambda_m\}$ converges uniformly to 1.

It remains to show

$$\sup_{\tau_0 \le t \le \tau} \left| \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i g_i(t)}{S_i^2(\boldsymbol{\beta}_m)} - \sigma^2(t) \right| \xrightarrow{p} 0.$$

The above convergence can be established by an application of the Lenglart Inequality and the procedures to show the consistency of the observed information matrix in Theorem 3.2 of Andersen and Gill (1982). The only place that needs attention is that $\boldsymbol{\beta}_m$ depends on $t$ while the proof in Theorem 3.2 of Andersen and Gill (1982) deals with a time independent intermediate variable "$\boldsymbol{\beta}^*$", but the key ingredient that

$$\sup_{\tau_0 \le t \le \tau} \left| S^{(k)}(\boldsymbol{\beta}_m, t) - s^{(k)}(\boldsymbol{\beta}_0, t) \right| \to 0$$

still holds for $k = 0, 1, 2$.

To conclude, we have

$$\boldsymbol{h}(t)^\top(\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}) + \sigma^2(t)\lambda_n = m_n(\widehat{\boldsymbol{\beta}}, 0) + o_p(||\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}|| + |\lambda_n|). \tag{2.12}$$

Using the same technique, we can also show that, for any sequence of random point $(\boldsymbol{\beta}_l, \lambda_l)$ lying on the line segment connecting $(\boldsymbol{\beta}_0, 0)$ and $(\boldsymbol{\beta}^*, \lambda_n)$,

$$\sup_{\tau_0 \le t \le \tau} \left| \frac{\partial l_n(\boldsymbol{\beta}_l, \lambda_l)}{\partial \boldsymbol{\beta}} + \boldsymbol{\Sigma} \right| \xrightarrow{p} 0 \tag{2.13}$$

and

$$\sup_{\tau_0 \le t \le \tau} \left| \frac{\partial l_n(\boldsymbol{\beta}_l, \lambda_l)}{\partial \lambda} - h(t) \right| \xrightarrow{p} 0. \tag{2.14}$$

Since the above properties are still true along each dimmension of $l_n(\boldsymbol{\beta}, \lambda)$, we could apply Taylor expansion to the vectored-valued function $l_n(\boldsymbol{\beta}, \lambda)$ and obtain ( using $l_n(\boldsymbol{\beta}^*, \lambda_n) = l_n(\widehat{\boldsymbol{\beta}}, 0) = 0$)

$$\boldsymbol{\Sigma}(\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}) - \boldsymbol{h}(t)\lambda_n = o_p(||\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}|| + |\lambda_n|). \tag{2.15}$$

Using martingale central limit theorem, Andersen and Gill (1982) showed that $\sqrt{n} m_n(\widehat{\boldsymbol{\beta}}, 0)$ converges weakly to a mean-zero Gaussian process in $D[\tau_0, \tau]$ with covariance

$$v(t, s) = \sigma^2(t \wedge s) + \boldsymbol{h}(t)^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{h}(s).$$

Therefore, $m_n(\widehat{\boldsymbol{\beta}}, 0)$ is uniformly of order $O_p(n^{-1/2})$, which, combined with (2.12) and (2.15), yields the following representation

$$\lambda_n = \frac{1}{v(t,t)} m_n(\widehat{\boldsymbol{\beta}}, 0) + o_p(n^{-1/2}), \tag{2.16}$$

$$\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}} = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{h}(t)}{v(t,t)} m_n(\widehat{\boldsymbol{\beta}}, 0) + o_p(n^{-1/2}). \tag{2.17}$$

□

**Step 4**: Asymptotic representation of the likelihood ratio test statistic, as shown in (2.19).

***Proof:*** The idea is to express the EL ratio into a quadratic form of $m_n(\boldsymbol{\beta}_0, 0)$ plus a negligible residual term. It's straightforward to see

$$\log \Re\{\Lambda_0(t), t\} = \sum_{i=1}^{n} \delta_i \boldsymbol{Z}_i^\top (\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}) + n\lambda_n \Lambda_0(t) - J_n,$$

where

$$J_n \;=\; \sum_{i=1}^{n} \delta_i \log \left\{ 1 + \frac{S_i(\boldsymbol{\beta}^*) - S_i(\widehat{\boldsymbol{\beta}}) + \lambda_n g_i(t)}{S_i(\widehat{\boldsymbol{\beta}})} \right\}.$$

Note that

$$\frac{S_i(\boldsymbol{\beta}^*) - S_i(\widehat{\boldsymbol{\beta}}) + \lambda_n g_i(t)}{S_i(\widehat{\boldsymbol{\beta}})} = \frac{S_i(\boldsymbol{\beta}^*) - S_i(\widehat{\boldsymbol{\beta}})}{S_i(\widehat{\boldsymbol{\beta}})} + \frac{g_i(t)}{S_i(\widehat{\boldsymbol{\beta}})}\lambda_n. \tag{2.18}$$

By the mean value theorem, there exists $\tilde{\boldsymbol{\beta}}_i$ on the line segement between $\boldsymbol{\beta}^*$ and $\widehat{\boldsymbol{\beta}}$ such that the first term is bounded by $||S_i^{(1)}(\tilde{\boldsymbol{\beta}}_i)||/S_i(\widehat{\boldsymbol{\beta}})||\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}||$. The boundedness of $\{Z_j\}, j = 1, 2, \cdots, n$, the consistency of $\boldsymbol{\beta}^*$ and $\widehat{\boldsymbol{\beta}}$ imply that $||S_i^{(1)}(\tilde{\boldsymbol{\beta}}_i)||/S_i(\boldsymbol{\beta}_0)$ is bounded by a constant $K$ that is independent of $i, n, t$, therefore the first term on the r.h.s. of (2.18) is of the order $O_p(n^{-1/2})$. The second term is also of the order $O_p(n^{-1/2})$, since when $g_i(t) > 0$, $S_i(\widehat{\boldsymbol{\beta}}) \geq S^{(0)}(\widehat{\boldsymbol{\beta}}, \tau)$ which converges in probability to $s^{(0)}(\boldsymbol{\beta}_0, \tau)$ uniformly in $t \in [\tau_0, \tau]$.

Therefore we may apply Taylor expansion to $J_n$ and obtain

$$\begin{aligned} J_n \;&=\; \sum_{i=1}^{n} \delta_i \left[ \frac{S_i(\boldsymbol{\beta}^*) - S_i(\widehat{\boldsymbol{\beta}})}{S_i(\widehat{\boldsymbol{\beta}})} + \frac{g_i(t)\lambda_n}{S_i(\widehat{\boldsymbol{\beta}})} - \frac{\left\{ S_i(\boldsymbol{\beta}^*) - S_i(\widehat{\boldsymbol{\beta}}) + \lambda_n g_i(t) \right\}^2}{2S_i^2(\widehat{\boldsymbol{\beta}})} \right] \\ &+\; O_p(n^{-1/2}). \end{aligned}$$

With a similar analysis to the residual term, we can easily show that

$$\begin{aligned}\frac{S_i(\boldsymbol{\beta}^*) - S_i(\widehat{\boldsymbol{\beta}})}{S_i(\widehat{\boldsymbol{\beta}})} &= \frac{(\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}})^\top \boldsymbol{S}^{(1)}(\widehat{\boldsymbol{\beta}}, T_i) + 2^{-1}(\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}})^\top \boldsymbol{S}^{(2)}(\widehat{\boldsymbol{\beta}}, T_i)(\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}})}{S_i(\widehat{\boldsymbol{\beta}})} \\ &+ r_{ni}.\end{aligned}$$

Here the residual term is of the order $O_p(n^{-3/2})$ uniformly in $i$ and $t$. This result, (2.10), (2.11), (2.13), (2.14), (2.17), and (2.16), simplify the EL ratio statistic into

$$-2\log \Re\{\Lambda_0(t), t\} = n\frac{m_n^2(\widehat{\boldsymbol{\beta}}, 0)}{v(t,t)} + o_p(1). \tag{2.19}$$

□

***Proof of the theorem:*** Given the representation in (2.19), the asymptotic distribution of $-2\log \Re\{\Lambda_0(t), t\}$ follows from that of $\sqrt{n}m_n(\widehat{\boldsymbol{\beta}}, 0)$, which is, as previousely mentioned, a zero-mean Gaussian process with covariance $v(t, s)$. □

**Step 5**: Boundedness of $\sup_{\tau_0 \le t \le \tau} r(t)$.

***Proof:*** We will show that $r(t)$ converges to the square of a mean-zero Gaussian process. In order to obtain $r(t)$, we apply the Lagrangian multiplier method to solve the constrained likelihood where $\boldsymbol{\beta}$ is fixed at $\widehat{\boldsymbol{\beta}}$. For notational simplicity, let $d_i = S_i(\widehat{\boldsymbol{\beta}})$. It can be shown that $r(t)$ is equal to

$$r(t) = \sum_{i=1}^{n} \delta_i \log\left\{1 + \frac{g_i(t)}{d_i}\lambda_n\right\} - n\lambda_n\Lambda_0(t),$$

where the lagrangian multiplier (still denoted by $\lambda_n$) satisfies the following equation:

$$\sum_{i=1}^{n} \frac{\delta_i g_i(t)}{d_i + \lambda_n} = n\Lambda_0(t).$$

It is easy to find that the above equation has a unique solution in a neighborhood of $0$ as long as $\sum_{i=1}^{n} N_i(t) > 0$. The argument in step 2 can be applied here to show

$$|\lambda_n| \le \left|m_n(\widehat{\boldsymbol{\beta}}, 0)\right| \Lambda_0^{-1}(t) d_1.$$

It follows from the boundedness on $d_1$ and the weak convergence of $\sqrt{n}m_n(\widehat{\boldsymbol{\beta}}, 0)$ that $\lambda_n = O_p(n^{-1/2})$ uniformly in $t \in [\tau_0, \tau]$.

Next we derive an asymptotic representation for $\lambda_n$. This is done by observing that

$$\begin{aligned} m_n(\widehat{\boldsymbol{\beta}}, 0) &= \lambda_n \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i g_i(t)}{d_i(d_i + \lambda_n)} \\ &= \lambda_n \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i g_i(t)}{d_i^2} \times \frac{d_i}{d_i + \lambda_n} \\ &= \lambda_n \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i g_i(t)}{d_i^2} - \lambda_n^2 \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i g_i(t)}{d_i^2} \frac{1}{d_i + \lambda_n}. \end{aligned}$$

Using the bound on $g_i(t)/d_i$, one can easily see that the second term on the r.h.s. of the above eqation is of the order $O_p(n^{-1})$.

We also recall from (2.10) and (2.11) that

$$\sup_{\tau_0 \le t \le \tau} \left| \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i g_i(t)}{d_i^2} - \sigma^2(t) \right| \xrightarrow{p} 0,$$

therefore

$$\lambda_n = m_n(\widehat{\boldsymbol{\beta}}, 0) \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i g_i(t)}{d_i^2} \right\}^{-1} + o_p(n^{-1/2}).$$

A Taylor expansion to $r(t)$ yields

$$\begin{aligned} 2r(t) &= 2 \sum_{i=1}^{n} \delta_i \left\{ \frac{g_i(t)}{d_i} \lambda_n - \frac{1}{2} \frac{g_i(t)}{d_i^2} \lambda_n^2 + \frac{1}{3} \frac{1}{(1+\xi_i)^3} \frac{g_i(t)}{d_i^3} \lambda_n^3 \right\} \\ &= 2n\lambda_n m_n(\widehat{\boldsymbol{\beta}}, 0) - \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i g_i(t)}{d_i^2} \lambda_n^2 \\ &+ \frac{2}{3} \sum_{i=1}^{n} \frac{1}{(1+\xi_i)^3} \frac{\delta_i g_i(t)}{d_i^3} \lambda_n^3, \end{aligned} \tag{2.20}$$

where $\xi_i$ is between $0$ and $\lambda_n g_i(t)/d_i$. We have shown that $g_i(t)/d_i$ is t-uniformly bounded in probability and therefore the last term in (2.20) is of order $O_p(n^{-1/2})$. The representation of $\lambda_n$ then leads to

$$2r(t) = nm_n^2(\widehat{\boldsymbol{\beta}}, 0) \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i g_i(t)}{d_i^2} \right\}^{-1} + O_p(n^{-1/2}).$$

The proof is completed by the the weak convergence of $\sqrt{n} m_n(\widehat{\boldsymbol{\beta}}, 0)$. We comment that this result shows $2r(t)$ and $-2 \log \Re\{\Lambda_0(t), t\}$ is asymptotically equal to each other up to a multiplicative function of $t$. □

# Chapter 3 Difference of Survival Functions with Covariate Adjustment

## 3.1 Introduction

In biomedical studies, it is of great importance to compare two treatments that give rise to censored time-to-event data. The conventional approach relies on the popular Cox Proportional Hazards Model to estimate a constant hazard ratio as the treatment effect measure. However, the proportional hazards assumption is often violated in practice and accordingly many alternative treatment effect measures have been proposed. Dabrowska et al. (1989) introduced a relative change function involving the survival functions for tow populations and constructed pointwise confidence intervals for this function. Schemper (1992) suggested the estimation of average hazard ratio of two populations through a weighted Cox model. Xu and O'Quigley (2000) estimated the average regression effect through weighted score equation, under a non-proportional hazards model with time dependent coefficients.

Particularly, Zhang and Klein (2001) extended the work of Parzen et al. (1997) and obtained a Wald-type confidence band for the difference of two individualized survival functions with covariates adjusted by a stratified Cox model. The major advantage of this measure is that it is easy to interpret. It directly compares the survival prospect at the patient level. Moreover, it does not make any parametric assumption on the hazards ratio between the two treatments. However, as pointed out by Bie et al. (1987), normal approximation based confidence intervals and bands for hazard or survival functions related parameters may have unsatisfactory small sample performance and may include values outside the natural range of the parameters. Motivated by the many desirable properties of the EL method, in particular, its transformation respecting property that renders transformations unnecessary, under the same framework of Zhang and Klein (2001), we provide in this chapter an EL based confidence band for the difference of two individualized survival

functions.

As an appealing nonparametric method, EL has been widely used in the literature to compare two censored samples. EL inference on the difference of mean and median survival times has been studied in Wang and Wang (2001); EL simultaneous confidence bands have been obtained for Q-Q plots (Einmahl and McKeague, 1999), ratio of survival functions (McKeague and Zhao, 2002), difference of survival functions (Shen and He, 2006), and ratio of instant hazard functions (Zhao and Zhao, 2011). However, these works all focus on nonparametric comparisons without covariate adjustment. In many applications, it is crucial to make adjustment for covariates to account for imbalanced baseline risk factors which are almost ubiquitous in observational studies.

Asymptotic properties of the proposed EL ratio statistic and a resampling scheme used to construct the EL confidence band are given in Section 3.2. A simulation study comparing the EL and normal approximation methods, together with a real data example, is presented in Section 3.3 followed by the technical proof in Section 3.4.

### 3.2 Plug-in Empirical Likelihood Confidence Band

Suppose that there are two strata and denote by $n_i$ the sample size of the $i$th stratum and $n = n_1 + n_2$ the total sample size. For $i = 1, 2; j = 1, 2, \cdots, n_i$, $X_{ij}$ and $C_{ij}$ are the underlying survival and censoring times, respectively. They are assumed to be independent conditional on the vector of covariates $\boldsymbol{Z}_{ij}$. In the presence of right censoring, we observe $T_{ij} = \min(X_{ij}, C_{ij})$ and the censoring indicator $\delta_{ij} = I[X_{ij} \leq C_{ij}]$. The stratified Cox model relates the survival time $X_{ij}$ to the vector of covariates $\boldsymbol{Z}_{ij}$ through its conditional cumulative hazard function by assuming

$$\Lambda_{ij}(t|\boldsymbol{Z}_{ij}) = \Lambda_i(t) \exp(\boldsymbol{Z}_{ij}^\top \boldsymbol{\beta}_0),$$

where $\boldsymbol{\beta}_0$ is the unknown vector of regression coefficients, and $\Lambda_i(t)$ is the unspecified cumulative baseline hazard function of stratum $i$. Note that the two strata share the same regression coefficients but each has its own baseline hazard. Let $\boldsymbol{Z}_0$ be a given vector of covariates and $S_i(t|\boldsymbol{Z}_0)$ be the survival function of a subject with covariate $\boldsymbol{Z}_0$ in the $i$th stratum ($i = 1, 2$). We are interested in the conditional survival difference function $\theta_0(t|\boldsymbol{Z}_0) = S_1(t|\boldsymbol{Z}_0) - S_2(t|\boldsymbol{Z}_0)$.

Similar to Chapter 2, we will focus on the case of $\boldsymbol{Z}_0 = \mathbf{0}$ such that $\theta_0(t|\boldsymbol{Z}_0)$ is merely the difference of the baseline survival function. When $\boldsymbol{Z}_0 \neq \mathbf{0}$, we may use the shifted vector $\boldsymbol{Z}_{ij} - \boldsymbol{Z}_0$ as if it were originally observed. Due to this shift, the difference of the new baseline survival functions will be equal to the difference of the original survival functions with covariate $\boldsymbol{Z}_0$. Therefore in what follows $\boldsymbol{Z}_0$ is assumed to be $\mathbf{0}$ and we simply write $\theta_0(t)$, instead of $\theta_0(t|\mathbf{0})$. Define

$$\boldsymbol{S}_i^{(k)}(\boldsymbol{\beta}, t) = n^{-1} \sum_{j=1}^{n_i} I[T_{ij} \geq t] \boldsymbol{Z}_{ij}^{\otimes k} \exp(\boldsymbol{Z}_{ij}^\top \boldsymbol{\beta}), \quad i = 1, 2; k = 0, 1, 2.$$

Clearly, $\theta_0(t)$ could be consistently estimated by

$$\widehat{\theta}_0(t) = \exp\left\{-\widehat{\Lambda}_2(t)\right\} - \exp\left\{-\widehat{\Lambda}_1(t)\right\}.$$

Here

$$\widehat{\Lambda}_i(t) = n^{-1} \sum_{j=1}^{n_i} \delta_{ij} Y_{ij}(t) / S_i^{(0)}(\widehat{\boldsymbol{\beta}}, t)$$

is the Breslow baseline hazard estimator. $\widehat{\boldsymbol{\beta}}$ the partial likelihood estimator of $\boldsymbol{\beta}_0$ and is the root of the following partial likelihood score function

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{2} \sum_{j=1}^{n_i} \int_0^\infty \left\{ \boldsymbol{Z}_{ij} - \boldsymbol{S}_i^{(1)}(\boldsymbol{\beta}, t) / S_i^{(0)}(\boldsymbol{\beta}, t) \right\} \mathrm{d} I[T_{ij} \leq t, \delta_{ij} = 1].$$

Using a simple martingale convergence result, Zhang and Klein (2001) showed that under certain regularity conditions $\sqrt{n}\{\widehat{\theta}_0(t) - \theta_0(t)\}$ converges weakly to a mean-zero Gaussian

process with covariance function

$$v(t,s) = \sum_{i=1}^{2} S_i(t)S_i(s)\int_0^{t\wedge s}\Big\{s_i^{(0)}(\boldsymbol{\beta}_0,u)\Big\}^{-1}\mathrm{d}\Lambda_i(u) + \Big\{S_1(t)\boldsymbol{h}_1(t)-S_2(t)\boldsymbol{h}_2(t)\Big\}^{\top}\boldsymbol{\Sigma}^{-1}\Big\{S_1(s)\boldsymbol{h}_1(s)-S_2(s)\boldsymbol{h}_2(s)\Big\}. \quad (3.1)$$

In the above we have used the following notations:

$$\boldsymbol{h}_i(t) = \int_0^t \boldsymbol{s}_i^{(1)}(\boldsymbol{\beta}_0,u)\big/s_i^{(0)}(\boldsymbol{\beta}_0,u)\mathrm{d}\Lambda_i(u), \quad i=1,2;$$

$$\boldsymbol{s}_i^{(k)}(\boldsymbol{\beta},u) = p_i\mathbf{E}\big\{n\boldsymbol{S}_i^{(k)}(\boldsymbol{\beta},u)/n_i\big\}, \quad i=1,2; k=0,1,2;$$

$$\boldsymbol{\Sigma} = \sum_{i=1}^{2}\int_0^{\infty}\left[\boldsymbol{s}_i^{(2)}(\boldsymbol{\beta}_0,t)-\Big\{\boldsymbol{s}_i^{(1)}(\boldsymbol{\beta}_0,t)\Big\}^{\otimes 2}\Big/s_i^{(0)}(\boldsymbol{\beta}_0,t)\right]\mathrm{d}\Lambda_i(t). \quad (3.2)$$

Here $p_i = \lim_{n\to\infty} n_i/n$, which is assumed to exist and be positive for $i=1,2$.

Based on this result, Zhang and Klein (2001) constructed a Wald-type confidence band for $\theta_0(t)$ using a resampling technique originally proposed by Lin et al. (1994). In the following, we give an EL based alternative confidence band for $\theta_0(t)$.

**Empirical Likelihood Ratio**

Let $P_1$ and $P_2$ be two cumulative hazard functions defined on $[0,\infty)$, the EL function of the observed data is given by

$$L(P_1,P_2,\boldsymbol{\beta}) = \prod_{i=1}^{2}\prod_{j=1}^{n_i}\Big\{p_{ij}\exp(\boldsymbol{Z}_{ij}^{\top}\boldsymbol{\beta})\Big\}^{\delta_{ij}}\exp\Big\{-\exp(\boldsymbol{Z}_{ij}^{\top}\boldsymbol{\beta})P_i(T_{ij})\Big\},$$

where $p_{ij} = P_i(T_{ij}) - P_i(T_{ij}-)$ denotes the increment of $P_i$ at time $T_{ij}$. For a fixed $t$, let $\theta\in[-1,1]$ be a hypothesized value of $\theta_0(t)$ and $\eta$ a hypothesized value of $S_1(t)$. Following the idea in Chapter 2, we may define the following likelihood ratio

$$\Re_1(\theta, t) = \frac{\sup_{P_1, P_2, \boldsymbol{\beta}} \left\{ L(P_1, P_2, \boldsymbol{\beta}) \middle| \exp(-P_1) - \exp(-P_2) = \theta \right\}}{\sup_{\boldsymbol{\beta}, P_1, P_2} L(P_1, P_2, \boldsymbol{\beta})}.$$

However, the constraint in the numerator is non-linear in $P_1$ and $P_2$, making it difficult, if not impossible, to profile out these two quantities. In order to avoid this difficulty, we need to adapt a linearization technique that basically breaks this single constraint into two constraints linked by an intermediate variable which will be profiled out eventually. Earlier use of this linearization technique can be found in Naik-Nimbalkar and Rajarshi (1997) and Shen and He (2006), among others. Comes with this linearization technique is the increased computation complexity, given that the computation in Chapter 2 is already challenging and we need another round of maximization to profile out the intermediate variable. To address this difficulty, we further apply a plug-in technique, which fixes $\boldsymbol{\beta}$ in the numerator by its partial likelihood estimator $\widehat{\boldsymbol{\beta}}$. By doing so, we do not need to maximize the numerator over the possibly multi-dimensional regression coefficients. The consequence is that, as will be seen later, the constrained optimization problem is reduced to solving two univariate equations. Similar plug-in strategy of using an estimated parameter has been widely used in the literature (Dong and Matthews, 2012; Hjort et al., 2009) when the true profile empirical likelihood is difficult to obtain.

Combining these two techniques, we define the following profile EL ratio at $\theta = \theta(t)$ and $\eta = \eta(t)$:

$$\Re(\theta, \eta, t) = \frac{\sup_{P_1, P_2} \left\{ L(P_1, P_2, \widehat{\boldsymbol{\beta}}) \middle| P_1(t) = -\log \eta, P_2(t) = -\log(\eta - \theta) \right\}}{\sup_{\boldsymbol{\beta}, P_1, P_2} L(P_1, P_2, \boldsymbol{\beta})}. \tag{3.3}$$

Note that when we are only interested in inference on $\theta_0(t)$ at a single instant $t$, we may simply use $\theta$ and $\eta$ for $\theta(t)$ and $\eta(t)$, respectively. But when we want simultaneous inference,

we need to consider likelihood ratio test at each instant, and therefore, it is more appropriate to recognize that both $\theta$ and $\eta$ are functions of $t$. Also note that the two constraints are now linear in $P_1$ and $P_2$. The restriction that $\eta \in \mathbf{\Theta} \equiv (\max(0,\theta), \min(1, 1+\theta))$ is imposed such that both $\eta$ and $\eta - \theta$ fall in $(0,1)$, the natural range of a survival probability.

A straightforward argument using Lagrange Multiplier method yields

$$\begin{aligned} \log \Re(\theta, \eta, t) &= -\sum_{i=1}^{2}\sum_{j=1}^{n_i} \delta_{ij} \log\left\{1 + \frac{I[T_{ij} \le t]}{S_i^{(0)}(\widehat{\boldsymbol{\beta}}, T_{ij})}\lambda_i\right\} \\ &- n\Big\{\lambda_1 \log\eta + \lambda_2 \log\big(\eta - \theta\big)\Big\}, \end{aligned} \tag{3.4}$$

where the Lagrange multipliers $\lambda_1$ and $\lambda_2$ are implicit functions of $\eta, \theta$, and $t$, described by the following equations:

$$\begin{cases} \dfrac{1}{n}\displaystyle\sum_{j=1}^{n_1} \frac{\delta_{1j} I[T_{1j} \le t]}{S_1^{(0)}(\widehat{\boldsymbol{\beta}}, T_{1j}) + \lambda_1} + \log\eta = 0, & (3.5) \\ \dfrac{1}{n}\displaystyle\sum_{j=1}^{n_2} \frac{\delta_{2j} I[T_{2j} \le t]}{S_2^{(0)}(\widehat{\boldsymbol{\beta}}, T_{2j}) + \lambda_2} + \log\big(\eta - \theta\big) = 0. & (3.6) \end{cases}$$

For any $\eta \in \mathbf{\Theta}$, as long as $\sum_j \delta_{1j} I[T_{1j} \le t] > 0$, using a monotonicity and continuity argument, we can show that there is a unique smooth solution $\lambda_1$ to (3.5) such that $S_1^{(0)}(\widehat{\boldsymbol{\beta}}, T_{1j}) + \lambda_1 > 0$ for all $j \in \{1, 2, \cdots, n_1\}$. Similar claim regarding equation (3.6) can be also made. Therefore, for given $(\theta, \eta)$, the log-likelihood ratio $\log \Re(\theta, \eta, t)$ in (3.4) can be easily calculated and. In order to find its maximum over $\eta$, we equate with $0$ its derivative with respect to $\eta$:

$$\begin{aligned} \frac{\partial \Re}{\partial \eta} &= -\sum_{i=1}^{2}\sum_{j=1}^{n_i} \delta_{ij} \frac{I[T_{ij} \le t]}{S_i^{(0)}(\widehat{\boldsymbol{\beta}}, T_{ij})} \Bigg/ \left\{1 + \frac{I[T_{ij} \le t]}{S_i^{(0)}(\widehat{\boldsymbol{\beta}}, T_{ij})}\right\} \frac{\partial \lambda_i}{\partial \eta} \\ &- n\left\{\frac{\lambda_1}{\eta} + \frac{\partial \lambda_1}{\partial \eta}\log\eta + \frac{\lambda_2}{\eta - \theta} + \log(\eta - \theta)\frac{\partial \lambda_2}{\partial \eta}\right\}. \end{aligned}$$

If we take derivative with respect to $\eta$ on both sides of (3.5) and (3.6), we can find

$$\sum_{j=1}^{n_1} \frac{\delta_{1j} I[T_j \le t]}{\left\{S_1^{(0)}(\widehat{\boldsymbol{\beta}}, T_{1j}) + \lambda_1\right\}^2} \frac{\partial \lambda_1}{\partial \eta} = \frac{n}{\eta} \quad \text{and} \quad \sum_{j=1}^{n_2} \frac{\delta_{2j} I[T_{2j} \le t]}{\left\{S_2^{(0)}(\widehat{\boldsymbol{\beta}}, T_{2j}) + \lambda_2\right\}^2} \frac{\partial \lambda_2}{\partial \eta} = \frac{n}{\eta - \theta}.$$

Using the above result, we simplify the equation $\partial \Re / \partial \eta = 0$ to the following equation.

$$\frac{\lambda_1}{\eta} + \frac{\lambda_2}{\eta - \theta} = 0. \tag{3.7}$$

One can claim via a continuity argument that under very general condition, (3.7) has at least one solution. Although we are unable to show that it has a unique solution, lots of numerical experiments we have done indicate that the solution is unique. See Figure 3.1 for a plot of $\partial \Re / \partial \eta$ as a function of $\eta$.

To study the asymptotic properties of the EL ratio, we impose the following regularity conditions that are standard in the literature on proportional hazards models:

C1) The triplet $\{T_{ij}, \delta_{ij}, \boldsymbol{Z}_{ij}\}_{j=1}^{n_i}$ are independent and identically distributed within each group and the covariate vector $\boldsymbol{Z}_{ij}$ has bounded support.

C2) The information matrix $\boldsymbol{\Sigma}$ defined in (3.2) is positive-definite and $\Pr(\delta_{i1} = 1) > 0$, $i = 1, 2$.

C3) As $n$ approaches infinity, $n_i/n \to p_i > 0$, $i = 1, 2$.

**Theorem 3.1.** *Assume conditions C1-C3 and let $\tau_0$ and $\tau$ be pre-specified numbers such that $0 < S_i(\tau) < S_i(\tau_0) < 1$. When $\theta(t) = \theta_0(t)$, (3.7) has a solution $\eta_n(t) = \eta_n\{\theta_0(t), t\}$ that converges uniformly in probability to $S_1(t)$, and*

$$-2\sigma^2(t) \log \Re\left\{\theta_0(t), \eta_n(t), t\right\} \xrightarrow{\mathscr{D}} U^2(t),$$

*where $U(t)$ is a mean-zero Gaussian process with covariance function $v(t,s)$ defined in (3.1) and $\sigma^2(t)$ is given by*

$$\sigma^2(t) = \sum_{i=1}^{2} S_i^2(t) \int_0^t \left\{ s_i^{(0)}(\boldsymbol{\beta}_0, u) \right\}^{-1} \mathrm{d}\Lambda_i(u). \tag{3.8}$$

It is worthwhile to note that $v(t,t)$ is the asymptotic variance of $\widehat{\theta}_0(t)$ while $\sigma^2(t)$ can be viewed as the variance of $\widehat{\theta}_0(t)$ when $\boldsymbol{\beta}_0$ is known. According to Zhang and Klein (2001), they can be consistently estimated by $\widehat{v}(t,t)$ and $\widehat{\sigma}^2(t)$ obtained by replacing the unknown $\boldsymbol{s}^{(k)}(\boldsymbol{\beta}_0, t)$ and $\Lambda_i$ in (3.1) (3.8) with their empirical counterparts $\boldsymbol{S}^{(k)}(\widehat{\boldsymbol{\beta}}, t)$ and $\widehat{\Lambda}_i$ respectively.

An immediate consequence of the above theorem is that for a fixed $t \in [\tau_0, \tau]$, as $n$ approaches infinity,

$$-2 \log \Re\big\{ \theta_0(t), \eta_n(t), t \big\} \xrightarrow{\mathscr{D}} \frac{v(t,t)}{\sigma^2(t)} \chi_1^2,$$

a scaled chi-square distribution with one degree of freedom. This result is different from what we obtained in Chapter 2 where the limiting distribution is a chi-squared distribution without scaling factor. Let $\chi_1^2(\alpha)$ be the upper $\alpha$ percentile of $\chi_1^2$, then the level $\alpha$ EL confidence interval for $\theta_0(t)$ is given by

$$\Big\{ \theta \Big| -2\widehat{\sigma}^2(t) \log \Re\big\{ \theta, \eta_n(\theta, t), t \big\} \le \widehat{v}(t,t) \chi_1^2(\alpha) \Big\}.$$

**Confidence Band**

Similar to what we have seen in Chapter 2, the theorem provides us the basis of constructing the simultaneous EL confidence band for $\theta_0(t)$. Let $C = C_\alpha(\tau_0, \tau_1)$ be the upper $\alpha$ percentile of $\sup_{\tau_0 \le t \le \tau} |U(t)| / \sqrt{v(t,t)}$, then a level $\alpha$ EL simultaneous confidence band for $\theta_0(t)$ on $[\tau_0, \tau]$ is given by

$$\Big\{\big(\theta(t),t\big)\Big|\sup_{\tau_0\le t\le\tau} -2\widehat{\sigma}^2(t)\log\Re\big[\theta(t),\eta_n\{\theta(t),t\},t\big]\le\widehat{v}(t,t)C^2, t\in[\tau_0,\tau]\Big\}.$$

Now we face the same task of estimating the critical value $C$. The resampling technique presented in Chapter 2 is still applicable after some modifications. Here we omit the mathematical details but present the procedure in Algorithm 2. Again, it is important to choose a sufficiently large $N$ to stabilize the estimate. In our simulation study, $N = 5000$ appears to work very well. This is consistent with Lin et al. (1994) and Zhang and Klein (2001) who reported in their respective studies that $N$ of about a few thousand was enough for a $95\%$ confidence band.

**Algorithm 2** Estimating C for Difference of Survival Functions

**for** each $k \in \{1, 2, \cdots, N\}$ **do**
  Generate $\boldsymbol{G}_k \leftarrow \text{MVN}(\mathbf{0}, \widehat{\boldsymbol{\Sigma}}^{-1})$

  **for** $i \in \{1, 2\}$ and $j \in \{1, 2, \cdots, n_i\}$ **do**
    Generate $G_{ij} \leftarrow \text{N}(0, 1)$
  **end for**

  Set $\tilde{W}_k(t) \leftarrow \Big\{\widehat{\boldsymbol{h}}_1(t)\widehat{S}_1(t) - \widehat{\boldsymbol{h}}_2(t)\widehat{S}_2(t)\Big\}^\top \boldsymbol{G}_k$
  $+ n^{-1/2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\widehat{S}_i(t)\delta_{ij}I[T_{ij}\le t]G_{ij}/S_i^{(0)}(\widehat{\boldsymbol{\beta}}, T_{ij})$

  Set $C_k \leftarrow \sup_{\tau_0\le t\le\tau}|\tilde{W}_k(t)|/\sqrt{\widehat{v}(t,t)}$

**end for**
**return** $C \leftarrow$ upper $\alpha$ percentile of $\{C_1, C_2, \cdots, C_N\}$

*Remark:* When there's no covariate, the problem reduces to non-parametric two sample comparison. In that case, the variance-covariance function of the limiting distribution $U(t)$ will reduce to the variance-covariance function found in (Shen and He, 2006). One consquence is that the pointwise limiting distribution will be a chi-square with one degree

of freedom, but not a scaled chi-square. The other consequence is that $U(t)$ will have independent increments, rendering estimation of the critical value much easier. However, the Possion extension EL formulation in our work is different from the one used by (Shen and He, 2006), therefore, the resulting confidence bands and intervals will be different. (Murphy, 1995) studied both types of formulation for the EL inference on survival probabilities and demonstrated that both of them lead to meaningful confidence intervals, and in many cases, the have very similar performance. We believe such similarities can be also expected in the case of survival differences.

## 3.3 Numerical Study and Real Data Example

### Simulation

We assume the underlying survival time $X_{ij}$ have a Weibull distribution with cumulative hazard function

$$\Lambda_{ij}(t|Z_{ij}) = \alpha_i t^{\gamma_i} \exp(Z_{ij}^\top \beta_0), \quad i = 1, 2; j = 1, 2, \cdots, n_i.$$

In this setup, the proportional hazards assumption holds within each stratum but not across the strata if $\gamma_1 \neq \gamma_2$. The parameter $\gamma_i$ controls the shape of the distribution of $X_{ij}$ and $\alpha_i$ adjusts its scale. The censoring times $C_{1j}$ and $C_{2j}$ are assumed to be exponentially distributed with mean $\theta_1$ and $\theta_2$, respectively. The univariate covariate $Z_{ij}$ is drawn from the uniform distribution on $[0, 1]$. In our simulation study, $\gamma_1 = 1.1, \gamma_2 = 1, \alpha_1 = 0.5, \alpha_2 = 1$, and $\beta_0 = 0.1$. Finally, $\theta_1$ and $\theta_2$ are adjusted to incorporate various censoring rates in both groups. We apply both the EL and normal approximation confidence bands for $\theta_0(t)$ on the interval $[\tau_0, \tau]$, where $\tau_0$ is the first observed death in either stratum 1 or 2, whichever comes later; $\tau$ is chosen such that the risk set in each group contains at least $10$ percent of the observations in the corresponding group. $C_\alpha(\tau_0, \tau)$ is estimated based on $5000$ replicates.

We also considered the normal approximation method with two transformations that simply mimic the *log-minus-log* and *arcsine-square-root* transformations in Bie et al. (1987). That is, we considered applying $f_1(x) = \log\left[-\log\left\{(x+1)/2\right\}\right]$ and $f_2(x) = \arcsin\sqrt{(x+1)/2}$ on $\widehat{\theta}_0(t)$. Observed coverage percentages of the confidence bands using $1000$ simulation replicates are summarized in Table 3.1. The coverage percentages of confidence intervals for $\theta_0(3)$ and $\theta_0(2)$ using $5000$ replicates are given in Table 3.2 and Table 3.4, respectively. The corresponding average lengths are given in Table 3.3 and Table 3.5. In these tables, $C$ denotes the overall censoring rate and $n$ is the sample size per group.

One can see that the EL intervals and bands have accurate coverage while the normal approximation intervals and bands are perceivably undercovering, especially when the sample size is less than $100$ per group. The *arcsine-square-root* transformation moderately improves the performance of confidence band when the sample size is small, but it does not seem to improve the performance of the confidence interval. The use of the *log-minus-log* transformation does not yield any improvement. The smaller average length achieved by the normal confidence interval is not surprising given the smaller coverage probability.

Additional plots not shown here indicate that, when the sample size is small, the sampling distribution of $\widehat{\theta}_0(t)$ is clearly skewed, which might explain why the normal confidence interval fails to achieve the nominal coverage level, but it is less of a problem for the EL confidence interval, because one well-known advantage of the EL method is its ability to adapt to skewed sampling distribution and generate confidence intervals that have a data-driven shape.

Table 3.1: Coverage of EL and normal confidence bands for survival difference functions.

| $\alpha$ | $\beta$ | $C$ | $n$ | Log-Log | Arcsine | Normal | Empirical |
|---|---|---|---|---|---|---|---|
| 16 | 9 | 10% | 25 | 91.4 | 92.7 | 91.5 | 95.8 |
| | | | 30 | 91.3 | 92.9 | 91.6 | 95.1 |
| | | | 50 | 92.2 | 93.3 | 93.5 | 95.8 |
| | | | 100 | 93.7 | 94.2 | 94.0 | 95.4 |
| 7.1 | 4 | 20% | 25 | 90.6 | 92.6 | 91.8 | 95.5 |
| | | | 30 | 91.9 | 92.7 | 92.2 | 95.1 |
| | | | 50 | 92.8 | 93.1 | 92.5 | 94.7 |
| | | | 100 | 93.2 | 93.7 | 93.5 | 95.6 |
| 4.3 | 2.2 | 30% | 25 | 90.6 | 92.4 | 91.1 | 95.1 |
| | | | 30 | 91.3 | 92.3 | 90.8 | 94.4 |
| | | | 50 | 92.5 | 92.9 | 92.3 | 94.8 |
| | | | 100 | 92.9 | 93.4 | 93.5 | 95.6 |

Table 3.2: Coverage percentages of EL and normal CIs for survival differences at $t = 3$.

| $\alpha$ | $\beta$ | $C$ | $n$ | Log-Log | Arcsine | Normal | Empirical |
|---|---|---|---|---|---|---|---|
| 16 | 9 | 10% | 25 | 91.26 | 91.14 | 91.14 | 95.28 |
| | | | 30 | 91.62 | 91.40 | 91.30 | 94.54 |
| | | | 50 | 93.44 | 93.44 | 93.08 | 95.10 |
| | | | 100 | 94.10 | 94.22 | 94.20 | 95.30 |
| 7.1 | 4 | 20% | 25 | 90.54 | 90.74 | 90.62 | 95.32 |
| | | | 30 | 91.20 | 91.76 | 91.84 | 95.26 |
| | | | 50 | 92.20 | 92.02 | 92.52 | 94.76 |
| | | | 100 | 93.76 | 93.82 | 93.96 | 95.24 |
| 4.3 | 2.2 | 30% | 25 | 90.06 | 91.02 | 91.10 | 94.68 |
| | | | 30 | 90.36 | 91.00 | 91.12 | 94.54 |
| | | | 50 | 92.30 | 92.92 | 92.36 | 94.50 |
| | | | 100 | 93.50 | 93.76 | 93.30 | 95.44 |

Table 3.3: Average lengths of EL and normal CIs for survival differences at $t = 3$.

| $\alpha$ | $\beta$ | $C$ | $n$ | Log-Log | Arcsine | Normal | Empirical |
|---|---|---|---|---|---|---|---|
| 16 | 9 | 10% | 25 | 0.448 | 0.447 | 0.427 | 0.448 |
| | | | 30 | 0.410 | 0.409 | 0.389 | 0.409 |
| | | | 50 | 0.319 | 0.319 | 0.302 | 0.313 |
| | | | 100 | 0.224 | 0.224 | 0.210 | 0.214 |
| 7.1 | 4 | 20% | 25 | 0.459 | 0.459 | 0.462 | 0.489 |
| | | | 30 | 0.421 | 0.420 | 0.426 | 0.447 |
| | | | 50 | 0.328 | 0.327 | 0.329 | 0.341 |
| | | | 100 | 0.232 | 0.232 | 0.233 | 0.238 |
| 4.3 | 2.2 | 30% | 25 | 0.514 | 0.513 | 0.514 | 0.537 |
| | | | 30 | 0.475 | 0.474 | 0.475 | 0.495 |
| | | | 50 | 0.371 | 0.371 | 0.372 | 0.384 |
| | | | 100 | 0.263 | 0.263 | 0.262 | 0.268 |

Table 3.4: Coverage of EL and normal CIs for survival differences at $t = 2$.

| $\alpha$ | $\beta$ | $C$ | $n$ | Log-Log | Arcsine | Normal | Empirical |
|---|---|---|---|---|---|---|---|
| 16 | 9 | 10% | 25 | 93.52 | 93.56 | 93.14 | 95.00 |
| | | | 30 | 93.90 | 94.04 | 93.68 | 94.84 |
| | | | 50 | 94.58 | 94.52 | 94.16 | 94.90 |
| | | | 100 | 94.76 | 94.90 | 94.80 | 95.10 |
| 7.1 | 4 | 20% | 25 | 93.28 | 93.56 | 93.28 | 94.74 |
| | | | 30 | 94.04 | 94.06 | 93.70 | 95.02 |
| | | | 50 | 94.34 | 94.38 | 94.22 | 95.08 |
| | | | 100 | 94.78 | 94.76 | 94.66 | 94.86 |
| 4.3 | 2.2 | 30% | 25 | 93.14 | 93.36 | 92.90 | 94.68 |
| | | | 30 | 93.38 | 93.56 | 93.22 | 94.94 |
| | | | 50 | 94.58 | 94.36 | 94.24 | 95.04 |
| | | | 100 | 94.44 | 94.60 | 94.56 | 95.02 |

Table 3.5: Average lengths EL and normal CIs for survival differences at $t = 2$.

| $\alpha$ | $\beta$ | $C$ | $n$ | Log-Log | Arcsine | Normal | Empirical |
|---|---|---|---|---|---|---|---|
| 16 | 9 | 10% | 25 | 0.483 | 0.481 | 0.487 | 0.490 |
| | | | 30 | 0.444 | 0.443 | 0.447 | 0.449 |
| | | | 50 | 0.346 | 0.346 | 0.348 | 0.349 |
| | | | 100 | 0.245 | 0.245 | 0.245 | 0.246 |
| 7.1 | 4 | 20% | 25 | 0.514 | 0.512 | 0.518 | 0.521 |
| | | | 30 | 0.472 | 0.471 | 0.475 | 0.477 |
| | | | 50 | 0.368 | 0.367 | 0.370 | 0.371 |
| | | | 100 | 0.262 | 0.262 | 0.263 | 0.263 |
| 4.3 | 2.2 | 30% | 25 | 0.557 | 0.554 | 0.562 | 0.564 |
| | | | 30 | 0.511 | 0.509 | 0.515 | 0.517 |
| | | | 50 | 0.405 | 0.404 | 0.407 | 0.407 |
| | | | 100 | 0.287 | 0.287 | 0.288 | 0.288 |

Out of many cases we examined, we present here an exemplary plot of $\partial\Re/\partial\eta$ over $\eta$, with varying values $t$ and $\theta = \theta_0(t)$. We used the first set of parameters in the above simulations to generate the data where sample size per group is $50$. It can be seen in these three examples the solution is unique.

**An Example**

The dataset we use to illustrate our proposal is available in Appendix 1 of Kalbfleisch and Prentice (1980). It is a subset of the data from a randomized trial comparing chemotherapy followed by radiation therapy versus radiation therapy alone in patients with carcinoma of the mouth or throat. The dataset comprises 195 patients and includes follow-up times in days, status at the end of follow-up (dead or censored), and six additional covariates: sex, age at diagnosis, general condition of the patient on a functional rating scale, tumor site, T stage (size of tumor), and N stage (extent of nodal metastases). See Zucker (1998) for more details on the dataset and the fitted Cox model stratified by the treatments.

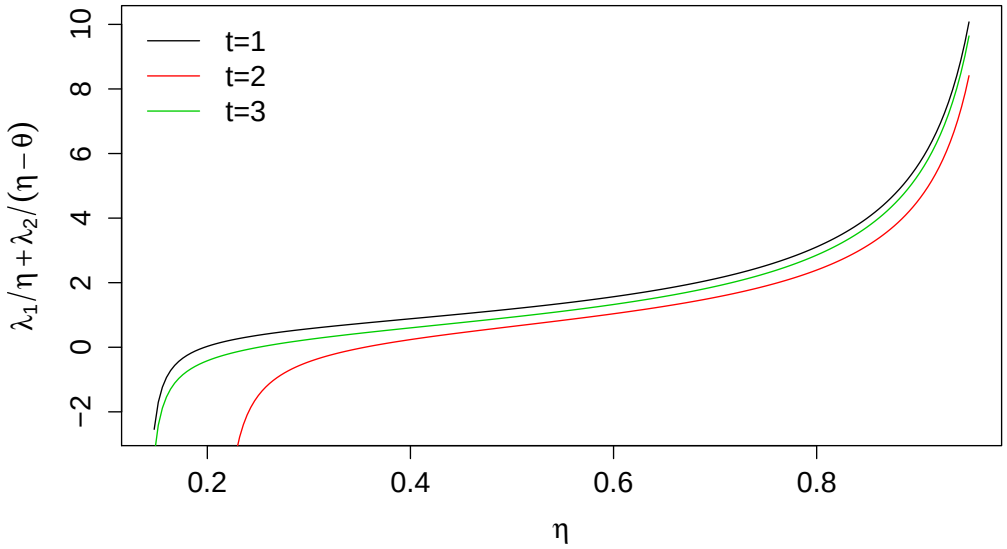



Figure 3.1: Plot of $\partial\Re/\partial\eta$.

For a male subject aged $60.4$ (average age across treatment groups) with a tumor site at faucial arch, having Grade category 1, Condition category 2, T staging category 4, and N stage category 0, we plot in Figure 3.2 the predicted survival difference and the associated simultaneous confidence band on the interval $(81, 1100)$. This interval is chosen because there's at least one death in each group beginning from day $81$ and the risk set on day $1100$ contains around $10\%$ of subjects in the corresponding group. The critical value $C_\alpha$ was estimated to be $8.53$ based on $10000$ simulations. For this subject, the estimated survival difference is subject to substantial variability as shown by the wide confidence band. The EL confidence band is pretty consistent with the normal confidence band, with only moderate distinction in the early and late ends of the time span.

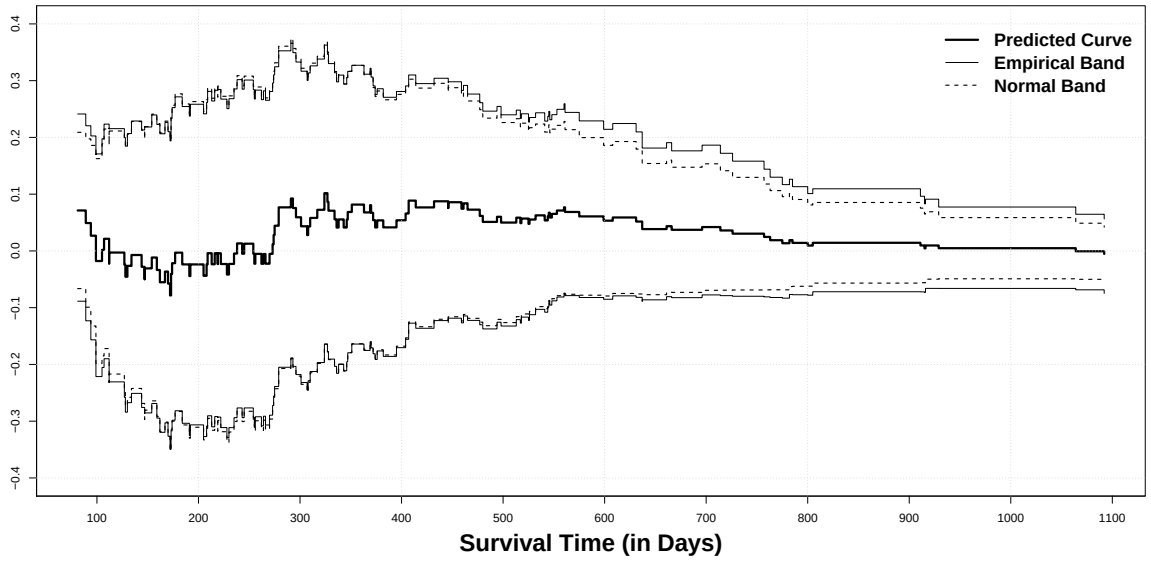



Figure 3.2: EL and normal approximation confidence bands for the individualized survival difference (Radiation vs. Chemotherapy + Radiation).

## 3.4 Proof

In what follows, "$O_p(\cdot)$" and "$O(\cdot)$" are to be understood uniformly in $t \in [\tau_0, \tau]$. We introduce the notations: $S_{ij} = S_i^{(0)}(\widehat{\boldsymbol{\beta}}, T_{ij})$, $\eta_0 = S_1(t)$, and $g_{ij} = I[T_{ij} \leq t]$. The major difficulty is showing the consistency of $\eta_n(t)$. However, it is intuitively clear that if $\eta$ is far away from $\eta_0$, the constraints are "unlikely" to be true, and therefore, the log-likelihood ratio is unlikely to be maximized. Therefore, in the following proof, we directly compare likelihood ratio at two values of $\eta$, one being close to $\eta_0$, one being away from $\eta_0$.

**Lemma 3.1.1.** *Let $M > 0$ be an arbitrarily fixed constant. If $\theta(t) = \theta_0(t)$ and $|\eta - \eta_0| \leq Mn^{-1/2}$, then the Lagrange Multiplier $\lambda_i$ defined in equations (3.5) and (3.6) are of the order $O_p(n^{-1/2})$.*

***Proof:*** Without loss of generality, we assume that $T_{ij}$ has been ordered such that $T_{i1} \leq T_{i2} \leq \cdots \leq T_{in_i}$. It is then clear that $S_{i1} \geq S_{i2} \geq \cdots \geq S_{in_i}$.

When $\lambda_1 > 0$, we have

$$\begin{aligned} -\log\eta &= \frac{1}{n}\sum_{j=1}^{n_1}\frac{\delta_{1j}g_{1j}}{S_{1j}+\lambda_1} = \frac{1}{n}\sum_{j=1}^{n_1}\frac{\delta_{1j}g_{1j}}{S_{1j}}\frac{S_{1j}}{S_{1j}+|\lambda_1|} \\ &\leq \frac{1}{n}\sum_{j=1}^{n_1}\frac{\delta_{1j}g_{1j}}{S_{1j}}\frac{S_{11}}{S_{11}+|\lambda_1|} \\ &= -\frac{S_{11}}{S_{11}+|\lambda_1|}\log\widehat{S}_1. \end{aligned}$$

Similarly, when $\lambda_1 < 0$, we have

$$\log\eta \leq S_{11}/(S_{11}-|\lambda_1|)\log\widehat{S}_1.$$

Therefore,

$$|\lambda_1| \leq S_{11}|\log\widehat{S}_1 - \log\eta|/|\log\eta|.$$

The uniform root-$n$ consistency of $\widehat{S}_1$, uniform boundedness of $S_{11}$ implied by the consistency of $\widehat{\boldsymbol{\beta}}$ and boundedness of $\boldsymbol{Z}_{ij}$, and the assumption that $|\eta_0 - \eta| \leq Mn^{-1/2}$ then imply $\lambda_1 = O_p(n^{-1/2})$.

Moreover, we can rewrite equation (3.5) into

$$\log\widehat{S}_1 - \log\eta = -\frac{\lambda_1}{n}\sum_{j=1}^{n_1}\frac{\delta_{1j}g_{1j}}{S_{1j}^2} + \frac{\lambda_1^2}{n}\sum_{j=1}^{n_1}\frac{g_{1j}}{S_{1j}^2(S_{1j}+\lambda_1)}.$$

Note that the first term on the r.h.s. is simply $-\lambda_1\widehat{\sigma}_1^2$. When $g_{1j} = I[T_{1j} \leq t] = 1$, we have $T_{1j} \leq t \leq \tau$, therefore $S_{1j} \geq S_i^{(0)}(\widehat{\boldsymbol{\beta}}, \tau)$ which converges in probability uniformly to $s_i^{(0)}(\boldsymbol{\beta}_0, \tau) > 0$ by a simple argument using SLLN. This shows that the second term is of

the order $O_p(n^{-1})$. To conclude,

$$\log\eta - \log\widehat{S}_1 = \lambda_1\widehat{\sigma}_1^2 + O_p(n^{-1}). \tag{3.9}$$

When $|\eta - \eta_0| \le Mn^{-1/2}$, there holds $|(\eta-\theta_0) - S_2| \le Mn^{-1/2}$, so similarly we have $\lambda_2 = O_p(n^{-1/2})$ and

$$\log(\eta-\theta_0) - \log\widehat{S}_2 = \lambda_2\widehat{\sigma}_2^2 + O_p(n^{-1}). \tag{3.10}$$

□

**Lemma 3.1.2.** *With probability approaching one, the likelihood ratio $\Re\big\{\theta_0(t),\eta,t\big\}$, as a function of $\eta$, allows a local maximizer $\eta_n\big\{\theta_0(t),t\big\}$ that converges uniformly to $\eta_0$.*

***Proof:*** For any $\eta$ such that $|\eta-\eta_0| \le Mn^{-1/2}$, using Taylor expansion to (3.4) and Lemma 3.1.1, we obtain

$$\begin{aligned}
-\log\Re\big\{\theta_0(t),\eta,t\big\} &= \sum_{i=1}^{2}\sum_{j=1}^{n_i}\delta_{ij}g_{ij}\left(\frac{\lambda_i}{S_{ij}} - \frac{\lambda_i^2}{2S_{ij}^2}\right) \\
&+ n\lambda_1\log\eta + n\lambda_2\log(\eta-\theta_0) + O_p(n^{-1/2}) \\
&= n\lambda_1\big(\log\eta - \log\widehat{S}_1\big) + n\lambda_2\big\{\log(\eta-\theta_0) - \log\widehat{S}_2\big\} \\
&- \frac{n}{2}\sum_{i=1}^{2}\lambda_i^2\widehat{\sigma}_i^2 + O_p(n^{-1/2}).
\end{aligned} \tag{3.11}$$

Due to (3.9), (3.10) and $\log(1+Mn^{-1/2}/S_1) = O(n^{-1/2})$, we have

$$\begin{aligned}
-\log\Re(\theta_0,\eta_0+Mn^{-1/2},t) &= \sum_{i=1}^{2}\frac{n\big\{\log(S_i+Mn^{-1/2}) - \log\widehat{S}_i\big\}^2}{2\widehat{\sigma}_i^2} + O_p(n^{-1/2}), \\
-\log\Re(\theta_0,\eta_0,t) &= \sum_{i=1}^{2}\frac{n\big(\log S_i - \log\widehat{S}_i\big)^2}{2\widehat{\sigma}_i^2} + O_p(n^{-1/2}).
\end{aligned}$$

Note that for $i=1,2$, there holds

$$\log(S_i+Mn^{-1/2}) - \log\widehat{S}_i = \Big(\log S_i - \log\widehat{S}_i\Big) + \log\Big(1+Mn^{-1/2}/S_i\Big).$$

The first term on the right hand side is of the order $O_p(n^{-1/2})$ while the second term is of the order $M \cdot O(n^{-1/2})$, hence by choosing a sufficiently large $M$, the probability that the second term eventually dominates the first term can be made as high as desired. When the second term dominates the first term, we have

$$\Re(\theta_0, \eta_0 + Mn^{-1/2}, t) < \Re(\theta_0, \eta_0, t).$$

Similar conclusion can be made when we replace $Mn^{-1/2}$ by $-Mn^{-1/2}$. This completes the proof. □

***Proof of the Theorem:*** Using Taylor expansion to the l.h.s. of (3.5) and (3.6) at $(\eta, \lambda_1, \lambda_2) = (\eta_0, 0, 0)$, one can find

$$\lambda_i = \frac{(-1)^{i-1} S_i}{S_1^2 \widehat{\sigma}_1^2 + S_2^2 \widehat{\sigma}_2^2} \left( S_1 \log \frac{S_1}{\widehat{S}_1} - S_2 \log \frac{S_2}{\widehat{S}_2} \right) + O_p(n^{-1}). \tag{3.12}$$

Using (3.9)-(3.11) we obtain

$$\begin{aligned} -2 \log \Re\{\theta_0, \eta_n(t), t\} &= 2n\lambda_1 \Big\{ \lambda_1 \widehat{\sigma}_1^2 + O_p\big(n^{-1}\big) \Big\} + 2n\lambda_2 \Big\{ \lambda_1 \widehat{\sigma}_1^2 + O_p\big(n^{-1}\big) \Big\} \\ &- \frac{n}{2} \sum_{i=1}^{2} \lambda_i^2 \widehat{\sigma}_i^2 + O_p\big(n^{-1/2}\big) \\ &= n\lambda_1^2 \widehat{\sigma}_1^2 + n\lambda_2^2 \widehat{\sigma}_2^2 + O_p(n^{-1/2}). \end{aligned}$$

Plugging the expression for $\lambda_i$ in (3.12) into the above equality, we have

$$-2 \log \Re\{\theta_0, \eta_n(t), t\} = \frac{n}{S_1^2 \widehat{\sigma}_1^2 + S_2^2 \widehat{\sigma}_2^2} \left( S_1 \log \frac{S_1}{\widehat{S}_1} - S_2 \log \frac{S_2}{\widehat{S}_2} \right)^2 + o_p(1).$$

It suffices to recall that $\sqrt{n}\big(S_1 \log S_1/\widehat{S}_1 - S_2 \log S_2/\widehat{S}_2\big) \to U(t)$ in distribution (Zhang and Klein, 2001). □

# Chapter 4 Hazard Ratio Function with Covariate Adjustment

## 4.1 Introduction

In the presence of non-proportional hazards, Wei and Schaubel (2008) considered the ratio of baseline cumulative hazard functions under stratified Cox model. By using the cumulative hazards, this measure describes the cumulative treatment effect, instead of an instantaneous one. The authors outlined a few advantages of this measure and derived both the pointwise and simultaneous confidence bands associated with the estimated ratio using normal approximations. Recently, Dong and Matthews (2012) (henceforth, DM) studied the EL inference of this hazard ratio and obtained an EL ratio test based pointwise confidence interval. However, their asymptotic result on the EL ratio statistic is generally invalid and the resulting confidence interval is asymptotically undercovering. This motivated us to investigate the correct EL inference. Using the idea in Chapter 3, we establish in this chapter a stronger result, which allows us to obtain the correct EL pointwise confidence intervals and construct simultaneous confidence bands without the computationally extensive Bootstrap calibration.

The remainder of this chapter is organized as follows. In Section 4.2, we provide a short introduction to the EL method and present the details on its applicability to our study. The theoretical justification of the proposed method is also given. In Section 4.3, we summarize the result of a numerical study to validate our method. A brief discussion about the proposed method is given in Section 4.4.

## 4.2 Plug-in Empirical Likelihood Confidence Band

Assuming the same data setup and notations as those in Chapter 3, we provide the details on how we could obtain the EL inference on $\theta_0(t) = \Lambda_{10}(t)/\Lambda_{20}(t)$.

### Empirical Likelihood Ratio

For $i \in \{1, 2\}$, let $P_i(t)$ be a hazard function defined on $[0, \infty)$ that has non-negative increment $p_{ij}$ at the observed survival time $T_{ij}$. The EL function takes the form of

$$L(P_1, P_2, \boldsymbol{\beta}) = \prod_{i=1}^{2} \prod_{j=1}^{n_i} \Big\{ p_{ij} \exp(\boldsymbol{Z}_{ij}^\top \boldsymbol{\beta}) \Big\}^{\delta_{ij}} \exp\Big\{ -\exp(\boldsymbol{Z}_{ij}^\top \boldsymbol{\beta}) \sum_{k=1}^{n_i} I[T_{ik} \le T_{ij}] p_{ik} \Big\}. \tag{4.1}$$

It follows from Johansen (1983) that the above likelihood is maximized at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$, and $P_i = \widehat{\Lambda}_i$. Moreover, the achieved maximum, denoted by $L_0$, is equal to the maximum partial likelihood up to a multiplicative constant that involves only the number of censored observations. In addition to this full model maximum likelihood, the EL ratio also requires a constrained likelihood to reflect the plausibility of our hypothesis. Let $\theta(t) > 0$ be a hypothetical value for $\theta_0(t)$. We formulate the following constrained maximum EL at $\theta(t)$ :

$$L_c\{\theta(t), t\} = \sup_{P_1, P_2} L(P_1, P_2, \widehat{\boldsymbol{\beta}}) \quad \text{subject to} \quad \sum_{j=1}^{n_1} p_{1j} g_{1j}(t) = \theta(t) \sum_{j=1}^{n_2} p_{2j} g_{2j}(t). \tag{4.2}$$

Here $g_{ij}(t) = I[T_{ij} \le t]$. The supremum is only taken over the baseline hazards while $\boldsymbol{\beta}$ is fixed at its partial likelihood estimator $\widehat{\boldsymbol{\beta}}$. Similar plug-in strategies of using an estimated quantity in the EL ratio have been discussed in Li and Wang (2003), and Hjort et al. (2009). As commented by DM, we could alternatively follow Zhou (2006) to update both $\boldsymbol{\beta}$ and the baseline hazards $P_1$ and $P_2$. However, the need to maximize over both $\boldsymbol{\beta}$ and the baseline hazards will substantially increase the computational burden. It is worthwhile to note

that the formulation of the constraint in (4.2) is mathematically equivalent to that in DM, although the latter takes a different form that could potentially complicates the subsequent numerical computation and theoretical development.

Using Lagrange multiplier method, we can show that the solution to (4.2) is given by

$$p_{ij} = \frac{\delta_{ij}}{nS_i^{(0)}(\widehat{\boldsymbol{\beta}}, T_{ij}) + n\lambda_n\{-\theta(t)\}^{i-1}g_{ij}(t)}. \tag{4.3}$$

Here the Lagrange multiplier $\lambda_n = \lambda_n\{\theta(t), t\}$ solves the following equation

$$\sum_{j=1}^{n_1} \frac{\delta_{1j}g_{1j}(t)}{S_1^{(0)}(\widehat{\boldsymbol{\beta}}, T_{1j}) + \lambda_n} - \theta(t)\sum_{j=1}^{n_2} \frac{\delta_{2j}g_{2j}(t)}{S_2^{(0)}(\widehat{\boldsymbol{\beta}}, T_{2j}) - \lambda_n\theta(t)} = 0. \tag{4.4}$$

Under very general conditions, the left hand side of (4.4) is monotonic in $\lambda_n$ and allows a unique solution in a neighborhood of $0$. This nice property greatly facilitates robust numerical search for $\lambda_n$. We could plug (4.3) and $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ into (4.1) to get the constrained maximum likelihood, which finally leads to the following likelihood ratio

$$\Re\{\theta(t), t\} \equiv \frac{L_c\{\theta(t), t\}}{L_0} = \prod_{\delta_{ij}=1}\left[1 + \frac{\{-\theta(t)\}^{i-1}g_{ij}(t)}{S_i^{(0)}(\widehat{\boldsymbol{\beta}}, T_{ij})}\lambda_n\right]^{-1}. \tag{4.5}$$

**Asymptotic Distribution**

In order to study the large sample behavior of the likelihood ratio statistic, we assume that as $n$ goes to infinity, $n_i/n$ converges to $\alpha_i \in (0, 1)$. We also require the following regularity conditions:

(C1) The triplet $(T_{ij}, \delta_{ij}, \boldsymbol{Z}_{ij})$ are independent and identically distributed within each stratum; $\boldsymbol{Z}_{ij}$ is bounded.

(C2) The information matrix $\boldsymbol{\Sigma}$ at $\boldsymbol{\beta}_0$ is positive definite, where

$$\boldsymbol{\Sigma} = \sum_{i=1}^{2} \int_0^{\infty} \big[\boldsymbol{s}_i^{(2)}(\boldsymbol{\beta}_0, t) - \big\{\boldsymbol{s}_i^{(1)}(\boldsymbol{\beta}_0, t)\big\}^{\otimes 2} / s_i^{(0)}(\boldsymbol{\beta}_0, t)\big] \mathrm{d}\Lambda_i(t),$$

$$\boldsymbol{s}_i^{(k)}(\boldsymbol{\beta}, t) = \alpha_i \mathbf{E}\big\{Y_{i1}(t) \boldsymbol{Z}_{i1}^{\otimes k} \exp(\boldsymbol{Z}_{i1}^{\top}\boldsymbol{\beta})\big\}.$$

(C3) $\Pr(\delta_{ij} = 1) > 0$ for $i = 1, 2$. In words, the probability of observing an event in each group is positive.

To state the main theorem, we introduce the following notations:

$$\boldsymbol{h}_i(t) = \int_0^t \boldsymbol{s}_i^{(1)}(\boldsymbol{\beta}_0, u) / s_i^{(0)}(\boldsymbol{\beta}_0, u) \mathrm{d}\Lambda_i(u), \quad i = 1, 2, \tag{4.6}$$

$$\sigma^2(t) = \int_0^t \mathrm{d}\Lambda_1(u) / s_1^{(0)}(\boldsymbol{\beta}_0, u) + \theta_0^2(t) \int_0^t \mathrm{d}\Lambda_2(u) / s_2^{(0)}(\boldsymbol{\beta}_0, u), \tag{4.7}$$

$$v(t, s) = \sigma^2(t \wedge s) + \big\{\boldsymbol{h}_1(s) - \theta_0(s)\boldsymbol{h}_2(s)\big\}^{\top} \boldsymbol{\Sigma}^{-1} \big\{\boldsymbol{h}_1(t) - \theta_0(t)\boldsymbol{h}_2(t)\big\}. \tag{4.8}$$

Under conditions (C1)-(C3), we could follow the proof of Theorem 3.2 in Andersen and Gill (1982) to show that the parameters defined in (4.6)–(4.8) could be consistently estimated by $\widehat{\boldsymbol{h}}_i(t)$, $\widehat{\sigma}^2(t)$, and $\widehat{v}(t, s)$, obtained by replacing in (4.6)–(4.8) the unknown $\boldsymbol{s}^{(k)}(\boldsymbol{\beta}_0, t)$, $\Lambda_i(t)$, and $\theta_0(t)$ with their corresponding empirical counterparts $\boldsymbol{S}^{(k)}(\widehat{\boldsymbol{\beta}}, t)$, $\widehat{\Lambda}_i(t)$ and $\widehat{\theta}_0(t)$.

**Theorem 4.1.** *Let* $0 < \tau_0 < \tau$ *be such that* $0 < \theta_0(\tau_0), \theta_0(\tau) < \infty$. *Under conditions (C1)-(C3),*

$$-2\log \Re\{\theta_0(t), t\} \to U^2(t)/\sigma^2(t) \text{ in } D[\tau_0, \tau],$$

*where* $U(t)$ *is a Gaussian process with mean zero and variance covariance function* $v(t, s)$ *given by (4.8).*

The proof is deferred to Section 4.5. For a particular $t \in [\tau_0, \tau]$, it follows from Theorem 4.1 that

$$-2\log \Re\big\{\theta_0(t), t\big\} \xrightarrow{D} v(t,t)/\sigma^2(t)\chi_1^2.$$

Thus, an asymptotic level $\alpha$ EL confidence interval for $\theta_0(t)$ is given by

$$\Big\{\theta | -2\log \Re(\theta, t) \le \widehat{v}(t,t)/\widehat{\sigma}^2(t)\chi_1^2(\alpha)\Big\}.$$

***Remark 1:*** DM stated that, for a fixed $t \in [\tau_0, \tau]$, the likelihood ratio $-2\log \Re\{\theta_0(t), t\}$ converges in distribution to $\chi_1^2$, which is incorrect in general. This is only valid when $\sigma^2(t) = v(t,t)$, or equivalently, $\boldsymbol{h}_1(t) = \theta_0(t)\boldsymbol{h}_2(t)$. In the general case, we will have $\sigma^2(t) < v(t,t)$, therefore treating the limiting distribution as a chi-square distribution will result in asymptotically undercovering confidence intervals. An example is given in Section 4.3 to illustrate this point.

***Remark 2:*** According to Theorem 4.1, an asymptotic level $\alpha$ simultaneous confidence band for $\theta_0(t)$ on the interval $[\tau_0, \tau]$ is given by

$$B(\alpha, \tau_0, \tau) = \Big\{\big(\eta(t), t\big)\Big| \sup_{\tau_0 \le t \le \tau} \Big[ -2\widehat{\sigma}^2(t)/\widehat{v}(t,t)\log \Re\{\eta(t), t\}\Big] \le C(\alpha), t \in [\tau_0, \tau]\Big\},$$

where $C(\alpha)$ satisfies $\Pr\left\{\sup_{t\in[\tau_0,\tau]} U^2(t)/v(t,t) \geq C(\alpha)\right\} = \alpha$. However, just like the situation in Chapter 3, it is difficult to estimate $C(\alpha)$ analytically or by directly simulating samples of $U(t)$. We adapt the resampling method in Chapter 3 to the current problem and outline the procedure in Algorithm 3. Clearly, this re-sampling technique is computation-

**Algorithm 3** Estimating C for Hazards Ratiol Functions

**for** each $k \in \{1, 2, \cdots, N\}$ **do**

Generate $\boldsymbol{G}_k \leftarrow \text{MVN}(\boldsymbol{0}, \widehat{\boldsymbol{\Sigma}}^{-1})$

**for** $i \in \{1, 2\}$ and $j \in \{1, 2, \cdots, n_i\}$ **do**

Generate $G_{ij} \leftarrow \text{N}(0, 1)$

**end for**

Set $\tilde{W}_k(t) \leftarrow \left\{\widehat{\boldsymbol{h}}_1(t) - \widehat{\theta}_0(t)\widehat{\boldsymbol{h}}_2(t)\right\}^\top \boldsymbol{G}_k$
$+ n^{-1/2} \sum_{i=1}^{2} \sum_{j=1}^{n_i} [\widehat{\theta}_0(t)]^{i=2} \delta_{ij} g_{ij}(t) G_{ij} / S_i^{(0)}(\widehat{\boldsymbol{\beta}}, T_{ij})$

Set $C_k \leftarrow \sup_{\tau_0 \leq t \leq \tau} |\tilde{W}_k(t)| / \sqrt{\widehat{v}(t,t)}$

**end for**

**return** $C \leftarrow$ upper $\alpha$ percentile of $\{C_1, C_2, \cdots, C_N\}$

ally less burdensome compared to the Bootstrap approach in DM. To see the computational burden of the Bootstrap approach, assume that $B$ Bootstrap samples are desired. We will need to solve the nonlinear equation (4.4) approximately $B$ times at each failure time falling in the interval $(\tau_0, \tau)$, which is rather time consuming even if there are only a few hundred observations.

### 4.3 Numerical Study

The purpose of the simulation study is to demonstrate that the unadjusted EL (UEL) confidence interval of DM is asymptotically undercovering while our adjusted EL interval (AEL) has an accurate coverage probability. For convenience of comparison, we use the same simulation design as DM did. Let the underlying lifetime $X_{ij}$ follow a Weibull distribution with hazard function

$$\lambda_{ij}(t|Z_{ij}) = \alpha_i \gamma_i t^{\gamma_i - 1} \exp(\beta_0 Z_{ij}), \quad i = 1, 2.$$

In this setup, the proportional hazards assumption holds within each stratum but not between strata if we choose $\gamma_1 \neq \gamma_2$. We can adjust $\alpha_i$ to control the level of censoring and check how the censoring rate affects the finite sample performance. Besides, we generate the underlying censoring time $C_{ij}$ from Uniform$(2.5, 5)$.

In DM, the univariate covariate $Z_{ij}$ follows a Binomial distribution with probability parameter $0.5$ for $i = 1, 2$. We modify this such that $Z_{1j}$ follows Binomial with parameter $0.25$ while $Z_{2j}$ follows Binomial with parameter $0.75$. We make this modification to reflect the covariate imbalance that is usually encountered in observational studies, and more importantly, to yield a larger value of $v(t,t)/\sigma^2(t)$ such that the discrepancy between the asymptotic distribution of the unadjusted likelihood ratio statistic and the $\chi^2_1$ distribution is easily discernible. See Figure 4.1 for a plot of the true value of $v(t,t)/\sigma^2(t)$ when the Binomial parameters are $0.25$ and $0.75$, where $\gamma_i$ and $\alpha_i$ are assigned the first set of values in Table 4.1. We find that when both Binomial parameters are $0.5$, the adjusting coefficient $v(t,t)/\sigma^2(t)$ is not exactly one, but close to one, which makes it hard to detect the undercoverage in simulation studies.

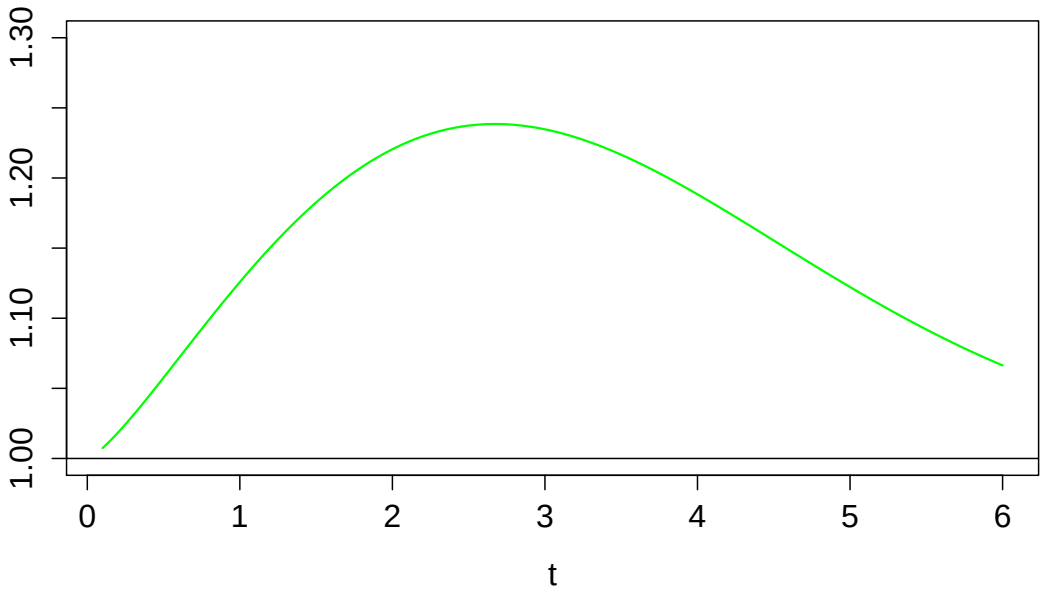



Figure 4.1: True value of $v(t,t)/\sigma^2(t)$ when there is no censoring.

Following their design, we consider the empirical likelihood confidence interval of $\theta_0(t_{0.75})$, where $t_{0.75}$ is the $75$th percentile of the combined survival times from the two groups. The observed coverage percentages of both the UEL and AEL methods using $50,000$ replicates are summarized in Table 4.1, where $C$ denotes the overall censoring rate and $n$ the sample size per group.

As expected, Table 4.1 shows that the UEL interval is considerably undercovering with the coverage probability clearly falling short of the nominal level even when the sample size per group reaches $400$. The AEL method, however, yields very accurate coverage probability even when the sample size per group drops to $30$ in the presence of $40\%$ censoring.

Table 4.1: Coverage percentages of the adjusted and un-adjusted EL CIs for hazards ratios.

| $\gamma_1$ | $\gamma_2$ | $\alpha_1$ | $\alpha_2$ | $t_{0.75}$ | $C$ | $n$ | $1-\alpha=0.90$ | | $1-\alpha=0.95$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | UEL | AEL | UEL | AEL |
| 1.4 | 1.2 | 0.4 | 0.35 | 2.273 | 0% | 30 | 84.498 | 89.822 | 90.788 | 94.504 |
| | | | | | | 50 | 85.118 | 89.884 | 91.358 | 94.744 |
| | | | | | | 100 | 85.898 | 90.200 | 92.040 | 94.986 |
| | | | | | | 400 | 86.058 | 89.984 | 92.096 | 94.936 |
| 1.4 | 1.2 | 0.4 | 0.35 | 2.273 | 10% | 30 | 84.060 | 89.662 | 90.416 | 94.316 |
| | | | | | | 50 | 84.638 | 89.638 | 91.016 | 94.452 |
| | | | | | | 100 | 85.400 | 89.952 | 91.516 | 94.802 |
| | | | | | | 400 | 85.836 | 90.094 | 91.986 | 95.022 |
| 1.0 | 1.5 | 0.2 | 0.10 | 3.503 | 40% | 30 | 83.590 | 89.434 | 90.038 | 94.220 |
| | | | | | | 50 | 84.572 | 89.856 | 90.946 | 94.704 |
| | | | | | | 100 | 85.368 | 90.012 | 91.702 | 94.886 |
| | | | | | | 400 | 85.612 | 90.048 | 91.794 | 94.966 |

## 4.4 Summary

We can also extend our method to the group-specific covariate adjustment model that involves fitting a separate Cox proportional hazards model in each of the treatment groups.

$$\Lambda_{ij}(t|\boldsymbol{Z}_{ij}) = \Lambda_i(t)\exp(\boldsymbol{Z}_{ij}^\top \boldsymbol{\beta}_i) \tag{4.9}$$

Clearly this model has greater flexibility than the stratified Cox model because it allows each group to have its own baseline hazard *and* regression coefficients. Under this model, DM studied the EL inference on the ratio of the two baseline cumulative hazards. In their simulation study, the reported coverage probabilities were substantially lower than the nominal levels, which was hypothesized to be the consequence of efficiency loss associated with a more complex model. As a result they considered employing Bootstrap to improve the coverage accuracy. However, we find that the EL ratio statistic should also be

adjusted by a multiplicative coefficient in a fashion similar to that in Section 4.2. Using the same simulation setup as DM did (DM, page 413), we summarize in Table 4.2 the observed coverage percentage of the adjusted EL method. Percentages for the UEL and Bootstrap methods were directly taken from Tables 2 and 3 in DM. One can see that the adjusted EL method drastically improves the coverage probabilities over the unadjusted EL and is better than the Bootstrap method in some cases.

Table 4.2: Coverage percentages of EL CIs for hazards ratios in the case of group-specific covariate adjustment.

| $\gamma_1$ | $\gamma_2$ | $\alpha_1$ | $\alpha_2$ | $c$ | $n$ | UEL | AEL | Bootstrap |
|---|---|---|---|---|---|---|---|---|
| 1.4 | 1.2 | 0.4 | 0.35 | 0% | 25 | 88.6 | 93.9 | 94.7 |
| | | | | | 50 | 89.1 | 94.6 | 94.5 |
| | | | | | 100 | 89.1 | 94.3 | 94.6 |
| | | | | | 250 | 89.4 | 95.2 | 94.6 |
| 1.4 | 1.2 | 0.4 | 0.35 | 10% | 25 | 87.6 | 93.7 | 94.6 |
| | | | | | 50 | 90.2 | 94.9 | 94.7 |
| | | | | | 100 | 89.6 | 94.4 | 94.6 |
| | | | | | 250 | 90.3 | 95.1 | 94.8 |
| 1 | 1.5 | 0.115 | 0.1 | 40% | 25 | 81.0 | 92.5 | 92.0 |
| | | | | | 50 | 82.2 | 93.4 | 92.2 |
| | | | | | 100 | 81.3 | 94.5 | 92.6 |
| | | | | | 250 | 82.4 | 94.7 | 92.7 |

One important feature of the EL confidence interval is that it typically does not require estimating the variance of the statistic, which is mostly appreciated in survival analysis where stable variance estimates could be very hard to obtain. However, the EL method in this paper does not have this feature, since we have to explicitly adjust the statistic by a quantity that involves the variance of the baseline hazards estimator. The need for the adjustment is due to the use of the plug-in type constraint which fixes $\boldsymbol{\beta}$ at its partial likelihood estimator. Additional analysis not shown here suggests that if we update in (4.2)

both the baseline hazards and $\boldsymbol{\beta}$, similar to what we did in Chapter 2, the resulting likelihood ratio statistic will have a limiting chi-square distribution with one degree of freedom. Apparently, that approach entails optimizing the constrained likelihood with respect to $\boldsymbol{\beta}$ and is therefore more computationally demanding. On the contrary, the plug-in type EL method used by DM is rather computationally friendly. The major computational effort is to calculate the likelihood ratio statistic, which only requires solving a univariate equation of the Lagrange multiplier. Monotonicity of the equation makes possible a fairly efficient numerical solution.

Without material modifications, the method we have proposed can be used to obtain an EL confidence band for the ratio of survival functions of two subjects in different strata with the same diagnostic covariates: $S_1(t|\boldsymbol{Z})/S_2(t|\boldsymbol{Z})$. This parameter compares the treatment on a relative scale, which might be more appropriate than the difference when both probabilities are small (McKeague and Zhao, 2006). Unlike the difference we studied in Chapter 3, the ratio allows a linear constraint and can be easily handled just like the ratio of cumulative hazards.

## 4.5 Proof

We start with three lemmas.

**Lemma 4.1.1.** *$\lambda_n$ uniquely exists with probability approaching 1 uniformly in $t \in [\tau_0, \tau]$.*

***proof:*** Note that when $t \in [\tau_0, \tau]$, the following term approaches 1 uniformly in $t$:

$$\Pr\left[\left\{\sum_{j=1}^{n_1} \delta_{1j} g_{1j}(t) > 0\right\} \cap \left\{\sum_{j=1}^{n_2} \delta_{2j} g_{2j}(t) > 0\right\}\right]$$

Therefore, we may assume $\sum_{j=1}^{n_1} \delta_{1j} g_{1j}(t) > 0$ and $\sum_{j=1}^{n_2} \delta_{2j} g_{2j}(t) > 0$. Without loss of generality we may assume that $T_{ij}$ has been rearranged in increasing order within each group. Let $m_i = \max\{j | N_{ij}(t) > 0\}$ and $b_{ij} = S_i^{(0)}(\widehat{\boldsymbol{\beta}}, T_{ij})$. We observe that the l.h.s. of equation (4.4) is montonically decreasing in $\lambda_n$; it approaches $+\infty$ and $-\infty$ as $\lambda_n$ approaches $-b_{1m_1}$ from the right and $b_{2m_2}/\theta_0(t)$ from the left, respectively. Therefore, equation (4.4) has a unique solution in $(-b_{1m_1}, b_{2m_2}/\theta_0(t))$. □

**Lemma 4.1.2.** *$\lambda_n = O_p(n^{-1/2})$ uniformly in $t \in [\tau_0, \tau]$.*

***proof:*** The order of $T_{ij}$ implies that $b_{i1} \geq b_{i2} \geq \cdots \geq b_{in_i}$, $i = 1, 2$, thus when $\lambda_n > 0$, we have

$$\sup_{1 \leq j \leq n_1} \frac{b_{1j}}{b_{1j} + \lambda_n} \leq \frac{b_{11}}{b_{11} + \lambda_n},$$

$$\sup_{j: N_{2j}(t) > 0} \frac{b_{2j}}{b_{2j} - \lambda_n \theta_0(t)} \geq \frac{b_{21}}{b_{21} - \lambda_n \theta_0(t)}.$$

It follows from equation (4.4) that

$$\begin{aligned}
\widehat{\Lambda}_1(t) \frac{b_{11}}{b_{11} + \lambda_n} &\geq \frac{1}{n} \sum_{j=1}^{n_1} \frac{\delta_{1j} g_{1j}(t)}{b_{1j}} \frac{b_{1j}}{b_{1j} + \lambda_n} \\
&= \theta_0(t) \frac{1}{n} \sum_{j=1}^{n_2} \frac{\delta_{2j} g_{2j}(t)}{b_{2j}} \frac{b_{2j}}{b_{2j} - \lambda_n \theta_0(t)} \\
&\geq \theta_0(t) \widehat{\Lambda}_2(t) \frac{b_{21}}{b_{21} - \lambda_n \theta_0(t)}.
\end{aligned}$$

Similarly, when $\lambda_n \leq 0$, we have

$$\widehat{\Lambda}_1(t)\frac{b_{11}}{b_{11}-|\lambda_n|} \leq \theta_0(t)\widehat{\Lambda}_2(t)\frac{b_{21}}{b_{21}+|\lambda_n|\theta_0(t)}.$$

Therefore, we always have

$$\begin{aligned} |\lambda_n| &\leq \frac{b_{11}b_{21}}{\theta_0(t)\{b_{11}\widehat{\theta}_0(t)+b_{21}\}}\left|\widehat{\theta}_0(t)-\theta_0(t)\right| \\ &\leq \frac{b_{11}}{\theta_0(t)}\left|\theta_0(t)-\widehat{\theta}_0(t)\right|, \end{aligned}$$

where $\widehat{\theta}_0(t) = \widehat{\Lambda}_1(t)/\widehat{\Lambda}_2(t)$. According to Wei and Schaubel (2008), $\widehat{\theta}_0(t) - \theta_0(t) = O_p(n^{-1/2})$ uniformly in $t \in [\tau_0, \tau]$; furthermore, consistency of $\widehat{\boldsymbol{\beta}}$, boundedness of $\boldsymbol{Z}_{ij}$ and a Central Limit Theorem (CLT) applied to $S_1^{(0)}(\boldsymbol{\beta}_0, T_{11})$ imply that $b_{11}$ is bounded in probability, which implies the lemma. □

**Lemma 4.1.3.** $\lambda_n = \left\{\widehat{\Lambda}_1(t) - \theta_0(t)\widehat{\Lambda}_2(t)\right\}/\widehat{\sigma}^2(t) + O_p(n^{-1})$ *uniformly in* $t \in [\tau_0, \tau]$.

***proof:*** Equation (4.4) implies

$$\frac{1}{n}\sum_{j=1}^{n_1}\frac{\delta_{1j}g_{1j}(t)}{b_{1j}}\left(1-\frac{\lambda_n}{b_{1j}+\lambda_n}\right) = \frac{\theta_0(t)}{n}\sum_{j=1}^{n_2}\frac{\delta_{2j}g_{2j}(t)}{b_{2j}}\left\{1+\frac{\lambda_n\theta_0(t)}{b_{2j}-\lambda_n\theta_0(t)}\right\},$$

therefore

$$\begin{aligned} &\widehat{\Lambda}_1(t) - \frac{\lambda_n}{n}\sum_{j=1}^{n_1}\frac{\delta_{1j}g_{1j}(t)}{b_{1j}^2}\left(1-\frac{\lambda_n}{b_{1j}+\lambda_n}\right) \\ &= \theta_0(t)\widehat{\Lambda}_2(t) + \frac{\lambda_n\theta_0^2(t)}{n}\sum_{j=1}^{n_2}\frac{\delta_{2j}g_{2j}(t)}{b_{2j}^2}\left\{1+\frac{\lambda_n\theta_0(t)}{b_{2j}-\lambda_n\theta_0(t)}\right\}. \end{aligned}$$

Recalling the definiton of $\widehat{\sigma}^2(t)$, we can rewrite the above equality into

$$\begin{aligned} \widehat{\Lambda}_1(t) - \theta_0(t)\widehat{\Lambda}_2(t) &= \lambda_n\widehat{\sigma}^2(t) - \lambda_n^2\frac{1}{n}\sum_{j=1}^{n_1}\frac{\delta_{1j}g_{1j}(t)}{b_{1j}^2}\frac{1}{b_{1j}+\lambda_n} \\ &+ \lambda_n^2\frac{\theta_0^3(t)}{n}\sum_{j=1}^{n_2}\frac{\delta_{2j}g_{2j}(t)}{b_{2j}^2}\frac{1}{b_{2j}-\lambda_n\theta_0(t)}. \end{aligned}$$

When $g_{ij}(t) = I[T_{ij} \leq t] > 0$, we know $T_{ij} \leq t \leq \tau$, therefore $g_{ij}(t) > 0$ implies $b_{ij} \geq S_i^{(0)}(\widehat{\beta}, \tau)$ which converges in probability to $s_i^{(0)}(\boldsymbol{\beta}_0, \tau) > 0$ by the consistence of $\widehat{\boldsymbol{\beta}}$, boundedness of $\boldsymbol{Z}_{ij}$ and the CLT. Combined with the uniform consistency of $\widehat{\sigma}^2(t)$, the uniform $O_p(n^{-1/2})$ bound of $\lambda_n$ implies that both of the last two terms on the r.h.s. of the above equation are of the order $O_p(n^{-1})$ uniformly in $t \in [\tau_0, \tau]$. This gives

$$\widehat{\Lambda}_1(t) - \theta_0(t)\widehat{\Lambda}_2(t) = \lambda_n \widehat{\sigma}^2(t) + O_p(n^{-1}).$$

Therefore,

$$\lambda_n = \big\{\widehat{\Lambda}_1(t) - \theta_0(t)\widehat{\Lambda}_2(t)\big\}/\widehat{\sigma}^2(t) + O_p(n^{-1}).$$

□

***Proof of the Theorem:*** We have demonstrated that $g_{ij}(t)/b_{ij}$ is bounded in probability uniformly in $i, j$ and $t \in [\tau_0, \tau]$, therefore, we may apply Taylor expansion to the right hand side of (4.5) (justified by Lemma 4.1.2) and use the expression for $\lambda_n$ in Lemma 4.1.3 to obtain

$$\begin{aligned}
-\log \Re\Big\{\theta_0(t), t\Big\} &= \sum_{i=1}^{2}\sum_{j=1}^{n_i} \delta_{ij}\left[\frac{\{-\theta_0(t)\}^{i-1} g_{ij}(t)}{b_{ij}}\lambda_n - \frac{1}{2}\frac{\{\theta_0(t)\}^{2i-2} g_{ij}(t)}{b_{ij}^2}\lambda_n^2\right] \\
&+ \sum_{i=1}^{2}\sum_{j=1}^{n_i} \delta_{ij}\frac{1}{3}\frac{\{-\theta_0(t)\}^{3i-3} g_{ij}(t)}{b_{ij}^3}\lambda_n^{*3} \\
&= \frac{\big\{\widehat{\Lambda}_1(t) - \theta_0(t)\widehat{\Lambda}_2(t)\big\}^2}{2\widehat{\sigma}^2(t)} + \sum_{i=1}^{2}\sum_{j=1}^{n_i} \delta_{ij}\frac{1}{3}\frac{\{-\theta_0(t)\}^{3i-3} g_{ij}(t)}{b_{ij}^3}\lambda_n^{*3},
\end{aligned}$$

where $\lambda_n^*$ is between 0 and $\lambda_n$ and is therefore $O_p(n^{-1/2})$, which then implies that the residual term is of the order $O_p(n^{-1/2})$. In summary,

$$-2\log \Re\Big\{\theta_0(t), t\Big\} = \Big\{\widehat{\Lambda}_1(t) - \theta_0(t)\widehat{\Lambda}_2(t)\Big\}^2/\widehat{\sigma}^2(t) + O_p(n^{-1/2}).$$

Since we have

$$\sqrt{n}\Big\{\widehat{\Lambda}_1(t) - \theta_0(t)\widehat{\Lambda}_2(t)\Big\} = \sqrt{n}\Big[\big\{\widehat{\Lambda}_1(t) - \Lambda_1(t)\big\} - \theta_0(t)\big\{\widehat{\Lambda}_2(t) - \Lambda_2(t)\big\}\Big],$$

the weak convergence of $\sqrt{n}\{\widehat{\Lambda}_1(t) - \theta_0(t)\widehat{\Lambda}_2(t)\}$ to $U(t)$ as described in the theorem is a simple application of the martingale central limit theorem (Andersen and Gill, 1982). $\square$

## Chapter 5 Future Work

Firstly, the difference of two individualized survival functions studied in Chapter 3 examins the treatment effect on the patient level, which is not appropriate when an aggregated difference is desired. One may consider the difference of two average survival functions defined as (Kim, 2001; Zucker, 1998)

$$\theta_0^*(t) = \int_{\boldsymbol{Z}} \big\{S_1(t|\boldsymbol{Z}) - S_2(t|\boldsymbol{Z})\big\} f(\boldsymbol{Z}) \mathrm{d}\boldsymbol{Z},$$

where $f(\cdot)$ is the density function of $\boldsymbol{Z}$. The example in Kim (2001) shows that this parameter may be comparable with $\theta_0(\cdot|\boldsymbol{Z}_0)$ in practice, where $\boldsymbol{Z}_0$ is the average vector of covariates. It is interesting to compare the EL confidence intervals and bands with the normal approximation competitors.

For the purpose of EL inference on $\theta_0^*(t)$, we can fix $\boldsymbol{\beta}$ at $\widehat{\boldsymbol{\beta}}$ and $f(\cdot)$ at its empirical estimator that puts equal weight on each observation, and formulate the following EL ratio at a hypothetical value $\theta^*$:

$$\Re(\theta^*, t) = \frac{\sup_{\eta, P_1, P_2} \big\{L(P_1, P_2, \widehat{\boldsymbol{\beta}}) \big| P_1(t) = -\log\eta, P_2(t) = -\log\phi\big\}}{\sup_{\boldsymbol{\beta}, P_1, P_2} L(P_1, P_2, \boldsymbol{\beta})},$$

where $\phi = \phi(\eta, \theta^*)$ is given by the equation

$$n^{-1} \sum_{i=1}^{2} \sum_{j=1}^{n_i} \big\{\eta^{\exp(\boldsymbol{Z}_{ij}^\top \widehat{\boldsymbol{\beta}})} - \phi^{\exp(\boldsymbol{Z}_{ij}^\top \widehat{\boldsymbol{\beta}})}\big\} = \theta^*.$$

Note that the left hand side of the above equation is monotonic in $\eta$ for a given $\phi$, therefore, the computational technique in Chapter 3 is still applicable here. The asymptotic property of the likelihood ratio may be more subtle than that in Chapter 3 due to the presence of $f(\boldsymbol{Z})$, but should be accessible, given that we know $\widehat{\boldsymbol{\beta}}$ and $\widehat{f}(\cdot)$ very well.

Secondly, the short-term and long-term hazards model proposed by Yang and Prentice

(2005) is an extension to the PH model. It incorporates short-term and long-term covariate effects by assuming the following hazard

$$\lambda(t|\boldsymbol{Z}) = \frac{\exp\left\{(\boldsymbol{\beta}+\boldsymbol{\gamma})^\top \boldsymbol{Z}\right\}}{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z})F(t) + \exp(\boldsymbol{\gamma}^\top \boldsymbol{Z})S(t)}\lambda(t),$$

where $\lambda(t)$ is the unspecified baseline hazard function, $S(t) = \exp\{-\int_0^t \lambda(s)\mathrm{d}s\}$ is the baseline survival function, and $F(t) = 1 - S(t)$ is the baseline cumulative distribution function. It can be seen that

$$\lim_{t\to 0}\frac{\lambda(t|\boldsymbol{Z}_1)}{\lambda(t|\boldsymbol{Z}_2)} = \exp\left\{\boldsymbol{\beta}^\top(\boldsymbol{Z}_1 - \boldsymbol{Z}_2)\right\}, \quad \lim_{t\to\infty}\frac{\lambda(t|\boldsymbol{Z}_1)}{\lambda(t|\boldsymbol{Z}_2)} = \exp\left\{\boldsymbol{\gamma}^\top(\boldsymbol{Z}_1 - \boldsymbol{Z}_2)\right\}.$$

Therefore, $\exp(\boldsymbol{\beta})$ and $\exp(\boldsymbol{\gamma})$ can be interpreted as the short-term and long-term hazards ratios, respectively. This features allows crossing hazards and crossing survivals that very happen in biomedical studies. Estimation and inference procedures for this model can be found in Diao et al. (2013).

With the increasing popularity of this model, it is of great interest to derive, under this model, confidence intervals and bands associated with the predicted survival function or hazard function for a given set of covariates. Unlike the PH model, there is no simple "partial likelihood" for this model, but it is still possible to profile out the baseline hazard $\lambda(t)$ using an iterative algorithm. This very possibility makes it very promising to derive EL confidence intervals and bands for the survival function using the methodology we have presented in Chapter 2.

# Appendix

## Computer Codes for Chapter 2

```
# the following code computes the empirical likelihood confidence  interval for S(t0)
# Time: survival time; delta: censoring indicator (1 for death)
# eps: step length. use a smaller value if fail to converge

library(nleqslv)
library(survival)
ci.surv <- function(Time, delta, Z, t0, eps = 0.1){

  # order the data by time
  ord <- order(Time); Time <- Time[ord]
  delta <- delta[ord]; Z <- Z[order,]
  fit <- coxph(Surv(time,delta)~Z); cox.beta <- fit$coef
  # unconstrained likelihood #
  old.eps <- eps; tol <- 1e-8; sum.dz <- apply(Z*delta,2,sum)
  tmp <- as.vector(exp(Z%*%cox.beta)); tmp <- rev(cumsum(rev(tmp)))
  w <- delta/tmp; lkhd0 <- sum(log(w[delta==1]+1e-13))+sum(cox.beta*sum.dz)-sum(delta)

  gvec <- as.numeric(Time<=t0); theta0 <- sum(gvec*w)
  up.theta1 <- theta0 + eps; up.theta0 <- theta0
  ind <- 1; beta <- cox.beta; lmda <- 0

  # upper boundary of the CI #
  while(ind){
   fit <- ratio.surv(Time,delta,Z, up.theta1,beta,lmda,cox.beta,lkhd0,t0)
    if(fit$lkhd.ratio<3.841459){
      if(fit$lkhd.ratio>3) eps = 0.5*old.eps
      beta <- fit$para[-1]; lmda <- fit$para[1]
      up.theta0 <- up.theta1; up.theta1 <- up.theta1 + eps
    }else{
      up.theta1 <- up.theta1 - 0.5*(up.theta1-up.theta0)
    }
    if((abs(up.theta1-up.theta0)<tol)||(abs(fit$lkhd.ratio-3.841459)<tol))
      ind <- 0
  }

  ind <- 1; beta <- cox.beta; lmda <- 0
  low.theta0 <- theta0; eps <- min(eps,0.1*theta0)
  low.theta1 <- theta0 - eps; old.eps <- eps

  # lower boundary of the CI #
  while(ind){
    # low.theta1=theta0; beta=cox.beta; lmda=0;
    fit <- ratio.surv(Time,delta,Z, low.theta1,beta,lmda,cox.beta,lkhd0,t0)
    if(fit$lkhd.ratio<3.841459){
      if(fit$lkhd.ratio > 3) eps = 0.5*old.eps
      beta <- fit$para[-1]; lmda <- fit$para[1]
      low.theta0 <- low.theta1; low.theta1 <- low.theta1 - eps
     }else{
      low.theta1 <- low.theta1 + (low.theta0-low.theta1)/2
    }
    if((abs(low.theta1-low.theta0)<tol)||(abs(fit$lkhd.ratio-3.841459)<tol))
      ind <- 0
  }
  return(exp(-c(low.theta0,up.theta0)))
}

# el ratio test needed
ratio.surv <- function(Time, delta, Z, theta, init.beta, init.lmda, cox.beta,
```

```
                      lkhd0, t0){

  # theta=theta0; init.beta=cox.beta; init.lmda=0
  gvec <- as.numeric(Time<=t0); n <- length(delta)
  n.cov <- dim(Z)[2]; sum.dz <- apply(Z*delta,2,sum)

  # joint equation of lambda and beta, and the jacobian
  nlefn <- function(para){
    beta <- para[-1]; lmda <- para[1]
    gam <- as.vector(exp(Z%*%beta)); gweight <- rev(cumsum(rev(gam)))
    z.gam <- Z*gam; gweight2 <- z.gam
    for(i in 1:n.cov)
      gweight2[,i] = rev(cumsum(rev(z.gam[,i])))

    Hw <- delta/(gweight+n^3*lmda*gvec); Hw2 <- Hw*gweight2
    func <- rep(0,n.cov+1); func[1] <- sum(Hw*gvec)-theta
    func[-1] = sum.dz-apply(Hw2,2,sum)
    return(func)
  }

  ### solve for beta and lmda using nleqslv
  cur.par <- nleqslv(x=c(init.lmda,init.beta),nlefn,
                     global="dbldog",control=list(xtol=1e-6,ftol=1e-8))$x
  beta <- cur.par[-1]; lmda <- cur.par[1]
  tmp <- as.vector(exp(Z%*%beta)); tmp <- rev(cumsum(rev(tmp)))
  w <- delta/(tmp+n^3*lmda*gvec)
  lkhd <- sum(log(w[delta==1]+1e-13))+sum(beta*sum.dz)-sum(w*tmp)

  return(list(lkhd.ratio=2*(lkhd0-lkhd),para=cur.par))
}
```

**Computer Codes for Chapter 3**

```
# codes for the empirical likelihood ratio test of
#    S1(t0) - S2(t0) = theta
# strata can be either 1 or 2; Z should be univariate


library(survival)
ratio.diff <- function(Time, delta, Z, strata, theta, t0){

  n1 <- sum(strata==1); n2 <- sum(strata==2);
  n  <- n1 + n2; tol <- 1e-9

  # order the data in strata and then time
  Mat  <- cbind(Time, Z, delta, strata)
  Mat  <- Mat[order(Mat[,4],Mat[,1]),]
  Time <- Mat[,1]; Z  <- Mat[,2]; delta  <- Mat[,3]

  gvec   <- as.numeric(Time <= t0); d.gvec <- delta*gvec
  events <- which(d.gvec==1)
  if(length(events)==0) stop("t0 too small")
  if(abs(theta)>=1) stop("theta out of range")

  # fit cox model
  fit   <- coxph(Surv(Time,delta)~Z+strata(strata))
  cox.beta <- fit$coef; names(cox.beta) <- NULL
  var.beta <- fit$var; names(var.beta) <- NULL

  # calculate adjusting coefficients
  gam <- exp(Z*cox.beta)
  S0  <- c(rev(cumsum(rev(gam[1:n1]))),
           rev(cumsum(rev(gam[(n1+1):n]))))/n
```

```
S1  <- c(rev(cumsum(rev(Z[1:n1]*gam[1:n1]))),
         rev(cumsum(rev((Z[(n1+1):n]*gam[(n1+1):n])))))/n

# death in each group
ind1 <- events[which(events <= n1)]
ind2 <- events[which(events > n1)]
j1 <- max(which((d.gvec)[1:n1]>0))
j2 <- n1 + max(which((d.gvec)[(n1+1):n]>0))

# function of eta to solve
f.eta1 <- function(eta){

  f.lmda1 <- function(x)
    sum(1/(S0[ind1]+x))/n + log(eta)

  f.lmda2 <- function(x)
    sum(1/(S0[ind2]+x))/n +
    log(eta-theta)

  # solve for lmda_i
  f10 <- f.lmda1(0)
  if(abs(f10)<tol) x1 = 0
  if(f10>tol){
    a <- 0; b <- 1; tmp = f10
    while(tmp > 0){
      b <-  2*b
      tmp <- f.lmda1(b)
    }
    x1 = uniroot(f.lmda1,c(a,b),tol=tol)$root
  }
  if(f10< -tol){
    a <- 0; b <- a
    while(f10 < 0){
      b <- 0.5*(b - S0[j1])
      f10 <- f.lmda1(b)
    }
    x1 = uniroot(f.lmda1,c(b,a),tol=tol)$root
  }

  f20 <- f.lmda2(0)
  if(abs(f20)<tol) x2 = 0
  if(f20>tol){
    a <- 0; b <- 1; tmp = f20
    while(tmp > 0){
      b <- 2*b
      tmp <- f.lmda2(b)
    }
    x2 = uniroot(f.lmda2, c(a,b),tol=tol)$root
  }
  if(f20< -tol){
    a <- 0; b <- a
    while(f20 < 0){
      b <- 0.5*(b - S0[j2])
      f20 <- f.lmda2(b)
    }
    x2 = uniroot(f.lmda2,c(b,a),tol=tol)$root
  }
  return(c(x1, x2, x1*theta-(x1+x2)*eta))
}

# solve for eta.n
f.eta <- function(eta) f.eta1(eta)[3]
a = max(0,theta)+1e-9 ; b=min(1,1+theta)-1e-9
eta.n = uniroot(f.eta, lower=a, upper=b, tol=tol)$root
lmda <- f.eta1(eta.n)[1:2]

# adjusting coefficient
```

```
  tmp1 = d.gvec/S0
  hat.S1 = exp(-sum(tmp1[1:n1])/n)
  hat.S2 = exp(-sum(tmp1[(n1+1):n])/n)
  hat.theta = hat.S1 - hat.S2

  tmp2 = tmp1*S1/S0
  h1 <- sum(tmp2[1:n1])/n
  h2 <- sum(tmp2[(n1+1):n])/n
  h1_h2 <- hat.S1*h1 - hat.S2*h2

  tmp3 = tmp1/S0
  v11 <- sum(tmp3[1:n1])/n
  v22 <- sum(tmp3[(n1+1):n])/n
  v2 <- (hat.S1^2*v11 + hat.S2^2*v22)
  epstt <- v2 + (h1_h2)^2*n*var.beta

  # -2log-likelihood ratio
  lrt <- 2*(sum(log(1+1/S0[ind1]*lmda[1]))+
              sum(log(1+1/S0[ind2]*lmda[2])))+
    (2*n)*(lmda[1]*log(eta.n)+log(eta.n-theta)*lmda[2])
  lrt <- v2/epstt*lrt
  return(list(lrt=lrt, p=pchisq(lrt, 1, lower.tail=FALSE)))
}
```

```
# calculate the critical value for confidence bands
# (t0, t1): the interval over which CB is constructed
# alpha: confidence level; m: number of replicates

c.diff = function(Time, delta, Z, strata, t0, t1, alpha=0.95, m=5000){

    # order the data by strata & time
    Mat  <- cbind(Time, Z, delta, strata)
    Mat  <- Mat[order(Mat[,4],Mat[,1]),]
    Time <- Mat[,1]; Z  <- Mat[,2]; delta  <- Mat[,3]

    fit   <- coxph(Surv(Time,delta)~Z+strata(strata))
    cox.beta <- fit$coef; var.beta <- fit$var

    # calculate adjusting coefficients
    n1 <- sum(strata==1); n2 <- sum(strata==2)
    n  <- n1 + n2
    gam <- exp(Z*cox.beta)
    S0  <- c(rev(cumsum(rev(gam[1:n1]))),
             rev(cumsum(rev(gam[(n1+1):n]))))/n
    S1  <- c(rev(cumsum(rev(Z[1:n1]*gam[1:n1]))),
             rev(cumsum(rev((Z[(n1+1):n]*gam[(n1+1):n])))))/n

    # death ids when time is ordered across groups
    ord <- order(Time); Time1 <- Time[ord]; delta1 <- delta[ord]
    S01 <- S0[ord]; S11 <- S1[ord]; strata1 <- strata[ord]
    ind1 = which((delta1==1)&id1); ind2 = which((delta1==1)&(1-id1))
    id1 <- as.numeric(strata1==1)

    # quantities involved in the M-C simulation
    a = max(ind1[1],ind2[1]); b = min(ind1[length(ind1)],ind2[length(ind2)])
    ind <- sort(c(ind1,ind2)); ind <- ind[which((ind >=a)&(ind <= b))]
    ind <- ind[which((Time1[ind] <= t1)&(Time1[ind] >= t0))]
    k <- length(ind)

    tmp  <- delta1/S01; tmp1 <- tmp/S01; tmp2 <- tmp1*S11
    hat.S1 = exp(-(cumsum(id1*tmp)[ind])/n)
    hat.S2 = exp(-(cumsum((1-id1)*tmp)[ind])/n)
    hat.theta = hat.S1 - hat.S2
    h1 <- (cumsum(id1*tmp2)[ind])/n
    h2 <- (cumsum((1-id1)*tmp2)[ind])/n
    h1_h2 <- hat.S1*h1 - hat.S2*h2
```

```
    v11 <- cumsum(id1*tmp1)[ind]; v22 <- cumsum((1-id1)*tmp1)[ind]
    v2 <- (hat.S1^2*v11+hat.S2^2*v22)/n
    epstt <- v2 + (h1_h2)^2*n*var.beta

    # calculate the critical value by simulation
    c_alpha = rep(0,m)
    for(j in 1:m){
      G  <- rnorm(1); G0 <- rnorm(n)*tmp
      W  <- (h1_h2)*G*sqrt(n*var.beta)+
        (hat.S1*cumsum(id1*G0)[ind]+
           hat.S2*cumsum((1-id1)*G0)[ind])/sqrt(n)
      c_alpha[j] <- max(abs(W)/sqrt(epstt))
    }
    return(quantile(c_alpha, alpha))
}
```

## Bibliography


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.

Bie, O., Borgan, Ø., and Liestøl, K. (1987). Confidence intervals and confidence bands for the cumulative hazard rate function and their small sample properties. *Scandinavian Journal of Statistics*, pages 221–233.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B*, 34:187–202.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.

Dabrowska, D. M., Doksum, K. A., and Song, J.-K. (1989). Graphical comparison of cumulative hazards for two populations. *Biometrika*, 76(4):763–773.

Diao, G., Zeng, D., and Yang, S. (2013). Efficient semiparametric estimation of short-term and long-term hazard ratios with right-censored data. *Biometrics*, 69(4):840–849.

DiCiccio, T., Hall, P., and Romano, J. (1991). Empirical likelihood is bartlett-correctable. *The Annals of Statistics*, 19(2):1053–1061.

Dickson, E. R., Grambsch, P. M., Fleming, T. R., Fisher, L. D., and Langworthy, A. (1989). Prognosis in primary biliary cirrhosis: model for decision making. *Hepatology*, 10(1):1–7.

Dong, B. and Matthews, D. E. (2012). Empirical likelihood for cumulative hazard ratio estimation with covariate adjustment. *Biometrics*, 68(2):408–418.

Einmahl, J. H. and McKeague, I. W. (1999). Confidence tubes for multiple quantile plots via empirical likelihood. *The Annals of Statistics*, 27(4):1348–1367.

Fisher, L. D. and Lin, D. Y. (1999). Time-dependent covariates in the cox proportional-hazards regression model. *Annual Review of Public Health*, 20(1):145–157.

Hjort, N. L., McKeague, I. W., and Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *The Annals of Statistics*, 37:1079–1111.

Hollander, M. and McKeague, I. W. (1997). Likelihood ratio-based confidence bands for survival functions. *Journal of the American Statistical Association*, 92(437):215–226.

Johansen, S. (1983). An extension of Cox's regression model. *International Statistical Review/Revue Internationale de Statistique*, 51:165–174.

Kalbfleisch, J. D. and Prentice, R. L. (1980). The statistical analysis of failure time data. 1980. *New York: John Wilev & Sons*.

Kim, J. (2001). Confidence intervals for the difference of median survival times using the stratified Cox proportional hazards model. *Biometrical Journal*, 43(6):781–790.

Li, G. (1995). On nonparametric likelihood ratio estimation of survival probabilities for censored data. *Statistics & Probability Letters*, 25(2):95–104.

Li, G., Hollander, M., McKeague, I. W., and Yang, J. (1996). Nonparametric likelihood ratio confidence bands for quantile functions from incomplete survival data. *The Annals of Statistics*, pages 628–640.

Li, G., Li, R., and Zhou, M. (2005). Empirical likelihood in survival analysis. In *Contemporary Multivariate Analysis and Design of Experiments*. Edited by J. Fan and G. Li, pp. 337-350. The World Scientific Publisher.

Li, G. and Van Keilegom, I. (2002). Likelihood ratio confidence bands in non–parametric regression with censored data. *Scandinavian Journal of Statistics*, 29(3):547–562.

Li, G. and Wang, Q.-H. (2003). Empirical likelihood regression analysis for right censored data. *Statistica Sinica*, 13(1):51–68.

Lin, D., Fleming, T., and Wei, L. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika*, 81(1):73–81.

McKeague, I. W. and Zhao, Y. (2002). Simultaneous confidence bands for ratios of survival functions via empirical likelihood. *Statistics & Probability Letters*, 60(4):405–415.

McKeague, I. W. and Zhao, Y. (2006). Comparing distribution functions via empirical likelihood. *The International Journal of Biostatistics*, 1(1):1–20.

Murphy, S. (1995). Likelihood ratio-based confidence intervals in survival analysis. *Journal of the American Statistical Association*, 90(432):1399–1405.

Naik-Nimbalkar, U. and Rajarshi, M. (1997). Empirical likelihood ratio test for equality of k medians in censored data. *Statistics & Probability Letters*, 34(3):267–273.

Nair, V. N. (1984). Confidence bands for survival functions with censored data: a comparative study. *Technometrics*, 26(3):265–275.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249.

Owen, A. B. (1990). Empirical likelihood confidence regions. *The Annals of Statistics*, 18(90):90–120.

Owen, A. B. (2001). *Empirical likelihood*. CRC press.

Pan, X.-R. and Zhou, M. (2002). Empirical likelihood ratio in terms of cumulative hazard function for censored data. *Journal of Multivariate Analysis*, 80(1):166–188.

Parzen, M., Wei, L., and Ying, Z. (1997). Simultaneous confidence intervals for the difference of two survival functions. *Scandinavian Journal of Statistics*, 24(3):309–314.

Qin, G. and Jing, B.-Y. (2001). Empirical likelihood for Cox regression model under random censorship. *Communications in Statistics-Simulation and Computation*, 30(1):79–90.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, pages 300–325.

Ren, J.-J. and Zhou, M. (2011). Full likelihood inferences in the Cox model: an empirical likelihood approach. *Annals of the Institute of Statistical Mathematics*, 63(5):1005–1018.

Schemper, M. (1992). Cox analysis of survival data with non-proportional hazard functions. *The Statistician*, pages 455–465.

Shen, J. and He, S. (2006). Empirical likelihood for the difference of two survival functions under right censorship. *Statistics & Probability Letters*, 76(2):169–181.

Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *The Annals of Statistics*, 9(1):93–108.

Wang, Q. and Qin, Y. (2010). Empirical likelihood confidence bands for distribution functions with missing responses. *Journal of Statistical Planning and Inference*, 140(9):2778–2789.

Wang, Q. and Wang, J.-L. (2001). Inference for the mean difference in the two-sample random censorship model. *Journal of Multivariate Analysis*, 79(2):295–315.

Wang, Q.-H. and Jing, B.-Y. (2001). Empirical likelihood for a class of functionals of survival distribution with censored data. *Annals of the Institute of Statistical Mathematics*, 53(3):517–527.

Wei, G. and Schaubel, D. E. (2008). Estimating cumulative treatment effects in the presence of nonproportional hazards. *Biometrics*, 64(3):724–732.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62.

Wood, A. T., Do, K.-A., and Broom, B. (1996). Sequential linearization of empirical likelihood constraints with application to U-statistics. *Journal of Computational and Graphical Statistics*, 5(4):365–385.

Xu, R. and O'Quigley, J. (2000). Estimating average regression effect under non-proportional hazards. *Biostatistics*, 1(4):423–439.

Yang, S. and Prentice, R. (2005). Semiparametric analysis of short-term and long-term hazard ratios with two-sample survival data. *Biometrika*, 92(1):1–17.

Zhang, M.-J. and Klein, J. P. (2001). Confidence bands for the difference of two survival curves under proportional hazards model. *Lifetime Data Analysis*, 7(3):243–254.

Zhao, Y. and Jinnah, A. (2012). Inference for Coxs regression models via adjusted empirical likelihood. *Computational Statistics*, 27(1):1–12.

Zhao, Y. and Zhao, M. (2011). Empirical likelihood for the contrast of two hazard functions with right censoring. *Statistics & Probability Letters*, 81(3):392–401.

Zhou, M. (2006). The Cox proportional hazards model with partially known baseline. In *Random Walk, Sequential Analysis and Related Topics*. Edited by A. C. Hsiung, Z. Ying and C. Zhang, pp. 215-232. The World Scientific Publisher.

Zucker, D. M. (1998). Restricted mean life with covariates: modification and extension of a useful survival analysis method. *Journal of the American Statistical Association*, 93(442):702–709.

**Vita**

# Shihong Zhu

## Education

**M.S. in Mathematics** 2009
Institute of Mathematics, Chinese Academy of Sciences Beijing

**B.S. in Mathematics** 2006
Nanjing University Nanjing

## Employment

Teaching Assistant & Research Assistant 2010-2015
Department of Statistics, University of Kentucky Lexington

## Publication

Shihong Zhu, Yifan Yang, Mai Zhou. A note on the empirical likelihood confidence band for hazards ratio with covariate adjustment. *Biometrics*, accepted.