Spring 5-31-2011

Confidence bands for survival functions under semiparametric random censorship models

Peixin Zhang
New Jersey Institute of Technology

ABSTRACT

CONFIDENCE BANDS FOR SURVIVAL FUNCTIONS UNDER SEMIPARAMETRIC RANDOM CENSORSHIP MODELS

by
Peixin Zhang

In medical reports point estimates and pointwise confidence intervals of parameters are usually displayed. When the parameter is a survival function, however, the approach of joining the upper end points of individual interval estimates obtained at several points and likewise for the lower end points would not produce bands that include the entire survival curve with a given confidence. Simultaneous confidence bands, which allow confidence statements to be valid for the entire survival curve, would be more meaningful.

This dissertation focuses on a novel method of developing one-sample confidence bands for survival functions from right censored data. The approach is model-based, relying on a parametric model for the conditional expectation of the censoring indicator given the observed minimum, and derives its chief strength from easy access to a good-fitting model among a plethora of choices currently available for binary response data. The substantive methodological contribution is in exploiting an available semiparametric estimator of the survival function for the one-sample case to produce improved simultaneous confidence bands. Since the relevant limiting distribution cannot be transformed to a Brownian Bridge unlike for the normalized Kaplan–Meier process, a two-stage bootstrap approach that combines the classical bootstrap with the more recent model-based regeneration of censoring indicators is proposed and a justification of its asymptotic validity is also provided. Several different confidence bands are studied using the proposed approach. Numerical studies, including robustness of the proposed bands to misspecification, are carried out to check efficacy. The method is illustrated using two lung cancer data sets.

CONFIDENCE BANDS FOR SURVIVAL FUNCTIONS UNDER SEMIPARAMETRIC RANDOM CENSORSHIP MODELS

by
Peixin Zhang

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology and
Rutgers, The State University of New Jersey – Newark
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Mathematical Sciences

Department of Mathematical Sciences, NJIT
Department of Mathematics and Computer Science, Rutgers-Newark

May 2011

APPROVAL PAGE

CONFIDENCE BANDS FOR SURVIVAL FUNCTIONS UNDER SEMIPARAMETRIC RANDOM CENSORSHIP MODELS

Peixin Zhang

Sundarraman Subramanian, PhD, Dissertation Advisor — Date
Associate Professor, Department of Mathematical Sciences, NJIT

Manish Bhattacharjee, PhD, Committee Member — Date
Professor, Department of Mathematical Sciences, NJIT

Sunil Dhar, PhD, Committee Member — Date
Associate Professor, Department of Mathematical Sciences, NJIT

Wenge Guo, PhD, Committee Member — Date
Assistant Professor, Department of Mathematical Sciences, NJIT

Kaifeng Lu, PhD, Committee Member — Date
Associate Director, Biostatistics, Forest Laboratories Inc.

BIOGRAPHICAL SKETCH

Author: Peixin Zhang

Degree: Doctor of Philosophy

Date: May 2011

Undergraduate and Graduate Education:

- Doctor of Philosophy in Mathematical Sciences,
  New Jersey Institute of Technology, Newark, NJ, 2011
- Bachelor of Sciences in Mathematics,
  Beijing University of Aeronautics and Astronautics, Beijing, 2007

Major: Mathematical Sciences

Presentations and Publications:

Peixin Zhang, Sundarraman Subramanian, "A Generalized Inverse Censoring Weighted Survival Function Estimator," *the Sixth Annual Conference on Frontiers in Applied and Computational Mathematics (FACM '09)*, Department of Mathematical Sciences, NJIT, June 1-2, 2009.

Peixin Zhang, Sundarraman Subramanian, "A Generalized Inverse Censoring Weighted Survival Function Estimator with Applications," *the Fifth Annual Graduate Research Day*, NJIT, Newark, November 11, 2009.

Peixin Zhang, Sundarraman Subramanian, "Confidence Bands for Survival Functions under Semiparametric Random Censorship Models," *the Seventh Annual Conference on Frontiers in Applied and Computational Mathematics (FACM '10)*, Department of Mathematical Sciences, NJIT, May 21-23, 2010.

Peixin Zhang, "Confidence Bands for Survival Functions under Semiparametric Random Censorship Models," *Applied Mathematics Seminar*, Department of Mathematical Sciences, NJIT, July 13, 2010.

Peixin Zhang, Sundarraman Subramanian, "Model-based Confidence Bands for Survival Functions," *the Sixth Annual Graduate Research Day*, NJIT, Newark, November 4, 2010.

Sundarraman Subramanian, Peixin Zhang, “The Inverse Censoring Weighted Approach for Estimation of Survival Functions from Left and Right Censored Data,” *Recent Advances in Biostatistics: False Discovery, Survival Analysis and Other Topics, Series in Biostatistics, Vol 4, 2011.*

Only those who have the patience to do simple things perfectly ever acquire the skill to do difficult things easily.

—Friedrich Schiller

ACKNOWLEDGMENT

First of all, I would like to thank my dissertation advisor, Professor Sundarraman Subramanian for his tremendous guidance with this dissertation. Ever since I joined the PhD program of the Department of Mathematical Sciences in NJIT, he has been a steadfast source of information, ideas, support, and energy. I am deeply grateful for the patient, encouraging and resourceful guidance he has been giving me throughout the last few years, and I sincerely wish for him all the best in all his future endeavors.

I also would like to extend my gratitude to the other members in my committee for their kindly encouragement and support to me throughout all these years while I was in the graduate school as well as throughout the process of working on my dissertation: Professor Manish Bhattacharjee, Professor Sunil Dhar, Professor Wenge Guo, and Dr. Kaifeng Lu. Especially, I would like to thank Professor Manish Bhattacharjee for recommending me to the department for admission into the PhD program with assistantship. I would like to thank Professor Daljit Ahluwalia, Chair of the Department of Mathematical Sciences, for his support.

I also want to thank all the other faculty, staff members and graduate students in the Department of Mathematical Sciences for all the help they have been giving me in the last four years. It has been a great honor to me to get to know all of them.

Last, but not the least, I would like to give the most sincere gratitude to my beloved wife. While all has been encouraging and enlightening, I simply would not have accomplished even a little bit without the love, support and understanding from my wife, Feiyan Chen.

TABLE OF CONTENTS

Chapter	Page
1 INTRODUCTION	1
1.1 General Overview	1
1.2 Semiparametric Random Censorship Models	2
1.3 Two-stage Resampling	3
1.4 Proposed Confidence Bands	4
2 REVIEW OF EXISTING CONFIDENCE BANDS	6
2.1 Hall–Wellner Band	6
2.2 Nair's Equal Precision Band	7
2.3 Akritas Band	8
2.4 Other Bands	8
3 NEW SIMULTANEOUS CONFIDENCE BANDS	10
3.1 Review of Semiparametric Random Censorship Models	10
3.2 Resampling Procedure and Large Sample Justification	12
3.3 Proposed Simultaneous Confidence Bands	16
4 NUMERICAL STUDIES	20
4.1 Simulation Studies	20
4.2 Misspecification Studies	23
4.3 Real Example Illustrations	26
5 CONCLUSION	35
APPENDIX A PROOFS	36
A.1 Proof of Theorem 1	36
A.2 Proof of Theorem 2	40
A.3 Proof of Theorem 3	45
REFERENCES	47

LIST OF TABLES

Table **Page**

4.1 First Lung Cancer Study Example: Percent Reduction in Width and Enclosed Area of Proposed over Competing Untransformed Bands Computed over the Interval [0.2274, 3.6027] 29

4.2 First Lung Cancer Study Example: Percent Reduction in Width and Enclosed Area of Proposed over Competing Transformed Bands Computed over the Interval $[0.2274, 3.6027]$ 29

4.3 First Lung Cancer Study Example: Percent Reduction in Width and Enclosed Area of Proposed over Competing Transformed Bands Computed over the Interval $[0.6077, 3.6027]$ 32

4.4 Second Lung Cancer Study Example: Percent Reduction in Width and Enclosed Area of Proposed over Competing Untransformed Bands Computed over the Interval $[0.991, 22.539]$ 33

4.5 Second Lung Cancer Study Example: Percent Reduction in Width and Enclosed Area of Proposed over Competing Transformed Bands Computed over the Interval $[0.991, 22.539]$ 33

4.6 Second Lung Cancer Study Example: Percent Reduction in Width and Enclosed Area of Proposed over Competing Transformed Bands Computed over the Interval $[2.99, 22.539]$ 33

LIST OF FIGURES

Figure **Page**

4.1 Empirical coverage probabilities of several confidence bands for different censoring rates. 20

4.2 Estimated average widths of several confidence bands for different censoring rates. 21

4.3 Estimated average enclosed areas of several confidence bands for different censoring rates. 22

4.4 Empirical coverage probabilities of proposed and other currently existing confidence bands for the first misspecification study. 24

4.5 Estimated average widths of proposed and other currently existing confidence bands for the first misspecification study. 25

4.6 Estimated average enclosed areas of proposed and other currently existing confidence bands for the first misspecification study. 26

4.7 Empirical coverage probabilities of proposed and other currently existing confidence bands for the second misspecification study. 27

4.8 Estimated average widths of proposed and other currently existing confidence bands for the second misspecification study. 28

4.9 Estimated average enclosed areas of proposed and other currently existing confidence bands for the second misspecification study. 30

4.10 Transformed simultaneous confidence bands of the survival function in a lung cancer study. 31

4.11 Untransformed simultaneous confidence bands of the survival function in a second lung cancer study. 34

CHAPTER 1

INTRODUCTION

1.1 General Overview

This research seeks to investigate a novel approach of constructing *simultaneous confidence bands* (SCBs) for survival functions from right censored data. In medical reports it is typical to display the estimate of a survival curve along with *pointwise* confidence intervals, which are two curves one of which connects the upper endpoints and the other connects the lower end points of interval estimates obtained at several points. It is not possible to make confidence statements for the entire survival curve with such intervals, however. SCBs do not have this problem and hence may be more meaningful to report. Recent advances offer the prospect of producing SCBs with improved coverage and which are potentially more informative than existing ones based on the Kaplan–Meier (KM) [32] estimator and its large sample properties. We implement such a new procedure for the one-sample setting.

SCBs, which were first applied to linear models [47, 57], are random regions within which an entire curve to be estimated lies with a pre-specified probability. Subsequently, SCBs have been obtained for generalized linear models [54, 43], isotonic or convex functions [16], regression curves [33, 26, 53, 58, 9], distribution functions [5, 25], density functions [45], receiver operating characteristic curves [27], isotonic dose-response curves [34], load curves [4], hazard rates [8], and the additive regression model [59], among others.

In the analysis of censored-time-to-event data, special attention has been devoted to obtaining SCBs for the *cumulative hazard* and *survival functions*, some of them based on empirical likelihood as well [23, 28, 18, 42, 1, 12, 6, 30, 36, 24, 29, 35, 40]; or for the difference/ratio of two survival functions [44, 61, 39]. In the one-sample setting of the random censorship model the data are n independent and identically distributed (i.i.d) copies of (Z, δ), where $Z = \min(T, C)$, T is the failure time of interest, C is an independent censoring variable, and $\delta = I(T \leq C)$ is the censoring indicator. Significantly, however, all existing methods for this setting employ the KM estimator to develop SCBs for the survival function $S(t) = P(T > t)$. Some recent advances indicate that improved SCBs can be obtained through an alternative approach, which has not yet been investigated and implemented.

1.2 Semiparametric Random Censorship Models

The alternative approach that we implement recognizes that one or more good-fitting models for $m(t)$, the conditional expectation of δ given $Z = t$, are available from the literature on binary response data and may be utilized for improved SCB construction. Indeed, choices such as logistic, probit, complementary log-log, and generalized proportional hazards, among others, may be investigated for zeroing-in on an apt model for $m(t)$; see, for example, [11, 10]. *Semiparametric random censorship models* (SRCMs) exploit this facility to replace each observed δ with its estimated conditional expectation, so that the censoring indicator δ figures in subsequent analysis only through its surrogate, namely the estimated $m(t)$. Thus, SRCMs derive their rationale from their ability to gainfully utilize parametric ideas within the (nonparametric) random censorship environment. Indeed, when the model for $m(t)$ is correctly chosen, asymptotically, the resulting SRCM-based estimator of $S(t)$ is more efficient than the KM estimator [13], and so we expect this efficiency to reflect in improved SRCM-based SCBs for $S(t)$. Unlike for the standard random censorship model, the SRCM approach is flexible enough to include missing censoring indicators as well with no additional effort, which may be an added plus [48].

1.3 Two-stage Resampling

The main issue in constructing one-sample SCBs for $S(t)$ lies in the specification of critical values. It is well known that the scaled KM process converges weakly to a time-transformed Brownian Bridge [28, 1, 19], using which the percentiles of its supremum can be obtained from tables calculated for this purpose. This is not possible with the SRCM approach, however, because the limiting distribution of the normalized cumulative hazard estimator does not have independent increments. A similar problem also arises when constructing subject-specific SCBs for $S(t)$, see [36], who developed simulated-process SCBs for $S(t)$ in the Cox regression framework. Their rationale was to utilize a representation for the limit of the normalized cumulative hazard process to produce an approximation whose distribution they generated using simulation. We employ a different strategy in that we propose and implement a novel two-stage resampling scheme that is specifically tailored to SRCMs and which we show produces asymptotically correct critical values for the *supremum* statistic leading to improved SCBs for $S(t)$.

In many instances the classical bootstrap allows calibration of percentiles of intractable distributions. Akritas (1986) [1] and Horváth and Yandell (1987) [31] showed that the approach [18] of obtaining bootstrap replicates by drawing at random and with replacement from $\{(Z_i, \delta_i), i = 1, \ldots, n\}$ yields asymptotically correct SCBs for $S(t)$. Sun, Sun, and Diao (2001) [55] used the same approach to derive SCBs for quantile functions. This approach, however, would be unsatisfactory for obtaining SRCM-based SCBs, which calls for a resampling mechanism that takes into account information available in the form of an assumed model for $m(t)$. We propose a two-stage resampling plan where we combine the classical bootstrap with model-based regeneration of censoring indicators. Regeneration of binary responses has been employed in multiple imputations estimation and model checking [38, 56, 14, 49, 50, 15].

1.4 Proposed Confidence Bands

The range-respecting SCBs that we propose are based on the asymptotic validity of our bootstrap, which means proving that for almost all samples the suprema over a certain time interval $[0, \tau]$ of normalized bootstrap processes, from which the desired critical values will be calibrated, have the same limit distribution as the basic ones they are intended to approximate. The method of proof involves first deriving functional central limit theorems for the bootstrap versions of certain basic estimators and later invoking Gill and Johansen's (1990) [22] functional delta method to prove via a series of compactly differentiable mappings the desired asymptotic validity. For different choices of a weight function we are then able to obtain the approximating critical values for constructing the SCBs for $S(t)$ on any desired interval $[t_1, t_2] \subset [0, \tau]$, see Section 3.3.

Our simulation studies focus on two cases one of which is when the model for $m(t)$ is specified correctly and the second pertains to performance in the face of misspecification. For the first case we perform comparisons with competing confidence bands proposed in the literature in terms of measures such as the empirical coverage probability (ECP), the estimated average enclosed area (EAEA), and the estimated average width (EAW). The ECP is the proportion of SCBs, computed from several generated data sets each of a certain sample size, that include $S(t)$ for all $t \in [t_1, t_2]$. The area enclosed by an SCB is computed as the sum of products of the widths of the SCB at each point of jump of the estimator of $S(t)$ and the *distance between points of jump*. The EAEA is the average of many such quantities. The average width of an SCB is a weighted average of the widths of the SCB at each jump point weighted by the *jump size of the estimator*. The EAW of the SCB is the average of many such quantities. The numerical study for the case of no misspecification is intended (i) to validate the proposed SCBs in the sense of providing verification that their ECPs match the nominal level $1 - \alpha$ and (ii) to showcase the superiority of the proposed

SCBs in the sense that they produce smaller EAEAs and EAWs in comparison with competing SCBs. On the other hand, the numerical robustness study that we report is to check how the ECP, EAEA, and EAW of the proposed SCBs deviate, with increased misspecification, from their values when there is no misspecification. As is typical of such studies, we estimate the parameters from models which are different from the ones from which the original data are generated.

CHAPTER 2

REVIEW OF EXISTING CONFIDENCE BANDS

It is of interest to compare the proposed SCBs with some of the existing bands based on the KM estimator. Here we provide a brief review of the various approaches that have been proposed in the literature.

2.1 Hall–Wellner Band

Hall and Wellner (1980) [28] derived their large-sample SCB based on the asymptotic property of the KM process

$$Z^*(t) = n^{1/2}(\hat{S}_{KM}(t) - S(t)), \qquad 0 \le t \le \tau.$$

They showed that for continuous *cumulative distribution functions* (cdf) $F(t)$ and $G(t)$, where $F(t) = 1 - S(t)$ and $G(t)$ is the cdf for the censoring time C, the limiting process $\{Z^*(t)\}_{0\le t\le\tau}$ is related to a Brownian bridge process, B^0, "by a rescale of the state space and a monotone transformation of the time axis". More specifically, asymptotically,

$$Z^*(t) = \frac{S(t)}{1-K(t)} B^0(K(t)), \qquad 0 \le t \le \tau,$$

in distribution, where

$$K(t) = C(t)\{1 + C(t)\}^{-1}, \qquad C(t) = \int_0^t (1 - F(s))^{-2} (1 - G(s))^{-1}\, dF(s).$$

Defining $\hat{K}(t) = \hat{C}(t)/(1+\hat{C}(t))$, where $\hat{C}(t)$ is a certain consistent estimator of $C(t)$, Hall and Wellner (1980) [28] proved that, with a properly estimated critical value, λ, as $n \to \infty$,

$$P\left(n^{1/2}\left|\hat{S}_{KM}(t) - S(t)\right| \frac{1-\hat{K}(t)}{\hat{S}_{KM}(t)} \leq \lambda, 0 \leq t \leq \tau\right)$$
$$\to \quad P\left(|Z^*(t)| \frac{1-K(t)}{S(t)} \leq \lambda, 0 \leq t \leq \tau\right).$$

It is useful to note that the Hall–Wellner band reduces to the Kolmogorov–Smirnov band, one of the most well-known bands used with complete data.

2.2 Nair's Equal Precision Band

Based on the normal approximation of the KM estimator and *Greenwood's variance formula*, one can derive the pointwise confidence interval for $S(t)$ on a fixed t by,

$$\hat{S}_{KM}(t) \pm z_{\alpha/2} n^{-1/2} \hat{S}_{KM}(t) \hat{C}(t)^{1/2}, \tag{2.1}$$

where $z_{\alpha/2}$ is the $100(1-\alpha/2)\%$ percentile of the standard normal distribution. Nair (1984) [42] showed that a large-sample SCB can be obtained by replacing $z_{\alpha/2}$ in (2.1) by an appropriately larger critical value. Specifically, under the random censorship model, he proved that for fixed $0 < a < b < 1$,

$$P\left(n^{1/2}|\hat{S}_{KM}(t) - S(t)| \leq e_\alpha \hat{S}_{KM}(t)\hat{C}(t)^{1/2}, \forall t : a \leq \hat{K}(t) \leq b\right) \quad \to \quad 1-\alpha,$$

as $n \to \infty$ with $e_\alpha = e_\alpha(a, b)$ satisfying

$$P\left(\sup_{a \leq x \leq b} \frac{|B^0(x)|}{(x(1-x))^{1/2}} \leq e_\alpha\right) \quad = \quad 1-\alpha.$$

Note that, Nair's SCB is defined simultaneously on $t \in [0, \tau]$ for which $a < \hat{K}(t) < b$, and the band width is proportional to the estimated standard deviation at each t. In this sense, the band has equal precision at all the valid t points. It is also of interest to note that in the absence of censoring, the equal precision band reduces to

$$\hat{S}_{KM}(t) \pm e_\alpha n^{-1/2} \left(\hat{S}_{KM}(t)\left(1 - \hat{S}_{KM}(t)\right)\right)^{1/2},$$

which is another well-known confidence band in the uncensored case.

2.3 Akritas Band

Akritas (1986) [1] suggested that for the random censorship model, bootstrapping may be carried out in two different ways [18, 46]. He showed, however, only Efron's (1981) [18] approach could be applied to produce SCBs for $S(t)$. Akritas (1986) [1] derived the weak convergence of the bootstrapped KM process, specifically that

$$n^{1/2}\left(\hat{S}^*_{KM}(t) - \hat{S}_{KM}(t)\right)\frac{1-\hat{K}(t)}{\hat{S}_{KM}(t)} \to B^0(K(t)),$$

conditionally almost surely on $[0, \tau]$, where $\hat{S}^*_{KM}$ denotes the KM estimator based on the bootstrapped data. As a result, this allows us to choose the critical values from the bootstrap distribution to construct SCBs. Akritas (1986) [1] demonstrated in his simulation studies that the SCBs using Efron's (1981) [18] procedure provide asymptotically correct convergence for samples even with small sample size and for discrete data as well. Independently, using different methodologies, Lo and Singh (1986) [37], and Horváth and Yandell (1987) [31] also proved that Efron's bootstrapping approach is correct to estimate the asymptotic distribution of the KM process. Lo and Singh (1986) [37] established a representation of the KM process as an i.i.d mean of a set of bounded random variables and gave a corresponding bootstrap version of the representation. Horváth and Yandell (1987) [31] approximated the bootstrapped KM process with a Wiener process. In the last two studies, rates of convergence of their approximations and similar results for the bootstrapped KM quantile processes were also provided.

2.4 Other Bands

Besides the three most widely used SCBs mentioned above, some other kinds of SCBs have also been introduced in the literature.

Efron (1967) [17] noted that asymptotically, the KM process can be transformed to a standard Brownian motion, or Wiener process, W. Specifically, he showed that for $s \in [0, C(\tau)]$,

$$n^{1/2}\left(\hat{S}_{KM}\left(C^{-1}(s)\right) - S\left(C^{-1}(s)\right)\right)/S\left(C^{-1}(s)\right) \quad \rightarrow \quad W(s),$$

weakly, as $n \rightarrow \infty$, where $C^{-1}(t)$ is the inverse of the increasing function, $C(t)$ for $t \in [0, \tau]$. Based on this version of transformation, instead of the transformation to Brownian bridge process used by Hall and Wellner (1980) [28], Gillespie and Fisher (1979) [23] developed their SCB by their proved fact that for fixed $c_1 < 0$, $c_2 > 0$ and specially chosen d_1, d_2,

$$\begin{aligned} & P\left(n^{1/2}\frac{\hat{S}_{KM}(t)}{n^{1/2} + c_2 + d_2\hat{C}(t)} \leq S(t) \leq n^{1/2}\frac{\hat{S}_{KM}(t)}{n^{1/2} + c_1 + d_1\hat{C}(t)}, 0 \leq t \leq \tau\right) \\ \rightarrow \quad & P\left(c_1 + d_1 s \leq W(s) \leq c_2 + d_2 s, 0 \leq s \leq C(\tau)\right) \\ = \quad & 1 - P\left(c_1, d_1, c_2, d_2, C(\tau)\right), \end{aligned}$$

as $n \rightarrow \infty$, where $P\left(c_1, d_1, c_2, d_2, C(\tau)\right)$ denotes the probability that $W(s)$ hits one of the nonintersecting straight lines, $c_1 + d_1 s$ and $c_2 + d_2 s$, [3]. Nair (1980) [41] adopted the same idea to generate a general version of large-sample SCB which releases the condition of linear boundaries. Both bands fail to reduce to the Kolmogorov–Smirnov band when there is no censoring. As we can see from previous study, Hall and Wellner (1980) [28] developed their SCB using the distribution of $\sup_{0 \leq t \leq \tau} |B^0(K(t))|$; a parallel approach can be seen in Gill (1980) [21] where he constructs SCB for $S(t)$ based on the known distribution of $\sup_{0 \leq s \leq C(\tau)} |W(s)|$.

CHAPTER 3

NEW SIMULTANEOUS CONFIDENCE BANDS

We denote the distribution function of Z by $H(t)$, its empirical estimator by $\hat{H}(t)$, and assume that $H(\tau) > 0$ where $\tau > 0$. We denote the cumulative hazard function of T by $\Lambda(t)$.

3.1 Review of Semiparametric Random Censorship Models

Dikta (1998) [13] first derived a functional central limit theorem for the SRCM-based estimator $\hat{S}(t)$. Here, however, we will employ the following modular procedure for our derivations. The basic building block is the subdistribution $Q(t) = P(Z \leq t, \delta = 1)$, which is given by

$$Q(t) = \int_0^t m(s)dH(s). \tag{3.1}$$

Following Dikta (1998) [13], we specify a parametric model $m(t, \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset I\!R^k$ is the model parameter, which we estimate via maximum likelihood. We denote the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$ and the model-based estimator of $m(t)$ by $m(t, \hat{\boldsymbol{\theta}})$. We write $D_r(m(t, \boldsymbol{\theta}))$ for the partial derivative of $m(t, \boldsymbol{\theta})$ with respect to θ_r, $r = 1, 2, \ldots, k$, denoting it by $D_r(m(t, \boldsymbol{\theta}^*))$ when it is evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}^*$. We also write $\text{Grad}(m(t, \boldsymbol{\theta})) = [D_1(t, \boldsymbol{\theta}), \ldots, D_k(t, \boldsymbol{\theta})]^T$ and $C_{\boldsymbol{\theta}}(t) = \text{Grad}(m(t, \boldsymbol{\theta}))\,(\text{Grad}(m(t, \boldsymbol{\theta})))^T$. When $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, we denote the matrix $C_{\boldsymbol{\theta}_0}(t)$ by $C_0(t)$. Let $I(\boldsymbol{\theta}_0) \equiv I_0 = E[C_0(Z)/(m(Z, \boldsymbol{\theta}_0)(1 - m(Z, \boldsymbol{\theta}_0)))]$ and let $\alpha(u, v) = (\text{Grad}(m(u, \boldsymbol{\theta}_0)))^T\, I_0^{-1} \text{Grad}(m(v, \boldsymbol{\theta}_0))$. We denote the second order partial derivatives by $D_{r,s}(\cdot)$. Note that the (r, s) element of the information matrix I_0 is given by

$$i_{r,s} \;=\; E\left(\frac{D_r(m(Z, \boldsymbol{\theta}_0))D_s(m(Z, \boldsymbol{\theta}_0))}{m(Z, \boldsymbol{\theta}_0)(1 - m(Z, \boldsymbol{\theta}_0))}\right). \tag{3.2}$$

The SRCM-based estimator of $Q(t)$, denoted by $\hat{Q}(t)$, is obtained by replacing $m(s)$ and $H(s)$ on the right hand side of Equation (3.1) with the estimates $m(s, \hat{\boldsymbol{\theta}})$ and $\hat{H}(s)$ respectively. We can show that $\hat{\mathbb{Z}}(t) = n^{1/2}\left(\hat{Q}(t) - Q(t)\right)$ tends weakly to a centered Gaussian process $\mathbb{Z}$ on $[0, \tau]$, where the covariance structure of $\mathbb{Z}$, $\mathrm{Cov}\,(\mathbb{Z}(t_1), \mathbb{Z}(t_2))$ is given for $0 \leq t_1 \leq t_2 \leq \tau$ by

$$\int_0^{t_1} m^2(x, \boldsymbol{\theta}_0) dH(x) - Q(t_1)Q(t_2) + \int_0^{t_2}\int_0^{t_1} \alpha(u, v) dH(u) dH(v). \tag{3.3}$$

See, for example Subramanian and Bandyopadhyay (2010) [51], where the influence function for a related process and the expression for its asymptotic variance are both given.

Writing $\hat{H}_-(t)$ for $\hat{H}(t-)$, the SRCM-based estimator of $\Lambda(t)$, denoted by $\hat{\Lambda}(t)$, is obtained via the following sequence of mappings, see Gill and Johansen (1990) [22] or Subramanian (2009) [49]:

$$(\hat{Q}, 1 - \hat{H}_-) \rightarrow \left(\hat{Q}, \frac{1}{1 - \hat{H}_-}\right) \rightarrow \int_{[0,\cdot]} \frac{1}{1 - \hat{H}_-} d\hat{Q} \;\equiv\; \hat{\Lambda}. \tag{3.4}$$

The SRCM-based estimator of $S(t)$, denoted by $\hat{S}(t)$, follows via the product integral mapping:

$$\underset{[0,\cdot]}{\pi}\left(1 - \frac{1}{1 - \hat{H}_-} d\hat{Q}\right) \;\equiv\; \underset{[0,\cdot]}{\pi}\left(1 - d\hat{\Lambda}\right) \;\equiv\; \hat{S}. \tag{3.5}$$

When the model for $m(t)$ is correctly specified, one may use the weak convergence of the basic bivariate process $n^{1/2}(\hat{Q}(t) - Q(t), \hat{H}_-(t) - H_-(t)) \equiv (\hat{\mathbb{Z}}(t), \hat{\mathbb{H}}(t))$ and appeal to Gill and Johansen's (1990) [22] functional version of the delta method to obtain, successively, the weak convergence of each of the intermediate processes in the above sequence, finally culminating in the weak convergence of $\hat{\mathbb{W}}(t) = n^{1/2}(\hat{S}(t) - S(t))$ in $D[0, \tau]$; see, for example, Subramanian (2009) [49] for an application of this approach. In particular, $\hat{\mathbb{W}}$ converges weakly to a centered Gaussian process with variance at t, denoted by $V(t)$ (see Equation (1.7) also), given by

$$S^2(t)\left[\int_0^t \frac{m(s)}{1-H(s)}d\Lambda(s)+\int_0^t\int_0^t \frac{\alpha(u,v)}{(1-H(u))(1-H(v))}dH(u)dH(v)\right]. \qquad (3.6)$$

When $m(t)$ is correctly specified, $V(t)$ is no greater than the asymptotic variance of the Nelson–Aalen estimator of $\Lambda(t)$ [13]. In this article we demonstrate that this efficiency of the SRCM-based approach leads to improved SCBs for $S(t)$. Given the plethora of choices available for fitting binary response data, identifying a suitable model for $m(t)$ should not be difficult; furthermore, model checking methods for testing the adequacy of a chosen model for binary response data are also readily available [14]. Therefore, it is clear that the SRCM approach would not involve significant investment of additional effort. As was noted in the introduction section, however, the limiting process does not have independent increments, the latter property being crucial for approximating a scaled version of the normalized survival function process with a Brownian Bridge process from which desired critical values could be calibrated. To compute critical values, we now introduce our two-stage resampling procedure and state our main results.

3.2 Resampling Procedure and Large Sample Justification

We obtain the bootstrap data $(Z_1^*, \delta_1^*), \dots, (Z_n^*, \delta_n^*)$ in the following way:

(1) Generate $Z_i^*, i = 1, \dots, n$ from $\hat{H}(t)$.

(2) For each $i = 1, \dots, n$, generate the censoring indicator δ_i^* from a Bernoulli distribution having success probability $m(Z_i^*, \hat{\boldsymbol{\theta}})$.

We write $\hat{\boldsymbol{\theta}}^*$ for the bootstrap MLE of $\boldsymbol{\theta}$ and we let $\hat{H}^*(t)$ and $\hat{Q}^*(t)$ denote the bootstrap versions of $\hat{H}(t)$ and $\hat{Q}(t)$ respectively. In turn, Equation (3.4) and Equation (3.5) then determine the bootstrap versions of the SRCM-based estimators of $\Lambda(t)$ and $S(t)$, which we will denote by $\hat{\Lambda}^*(t)$ and $\hat{S}^*(t)$ respectively. Our new SCBs will be based on large-sample justification of the proposed resampling, which involves deriving a new bootstrap version of the functional central limit theorem for the normalized SRCM-based cumulative hazard and survival function processes. We prove our result in the form of several technical modules.

We employ some of the notations and regularity conditions introduced by Dikta et al. (2006) [14]. We let $I\!P_n, I\!E_n, \mathbb{V}_n$, and Cov_n denote respectively the probability measure, expectation, variance, and covariance associated with our bootstrap sample. Write $w_1(x, \boldsymbol{\theta}) = \ln(m(x, \boldsymbol{\theta}))$, $w_2(x, \boldsymbol{\theta}) = \ln(1-m(x, \boldsymbol{\theta}))$, and $w(\delta, Z, \boldsymbol{\theta}) = \delta w_1(Z, \boldsymbol{\theta})+(1-\delta)w_2(Z, \boldsymbol{\theta})$. Then, the normalized log likelihood function based on the original data takes the form

$$l_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\left\{\delta_i w_1(Z_i, \boldsymbol{\theta}) + (1-\delta_i)w_2(Z_i, \boldsymbol{\theta})\right\},$$

Noting that $l_n^*(\boldsymbol{\theta})$ is the bootstrap version of $l_n(\boldsymbol{\theta})$, we now state some standard regularity conditions; see [13, 14].

$\mathbb{C}_1$ There exists a measurable solution $\hat{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$ of $\mathrm{Grad}(l_n(\boldsymbol{\theta})) = 0$ satisfying $\hat{\boldsymbol{\theta}} \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0$.

$\mathbb{C}_2$ For almost all sample sequences, $(Z_1, \delta_1), \ldots, (Z_n, \delta_n)$, there exists a measurable solution $\hat{\boldsymbol{\theta}}^* \in \boldsymbol{\Theta}$ of $\mathrm{Grad}(l_n^*(\boldsymbol{\theta})) = 0$ such that $\hat{\boldsymbol{\theta}}^* \xrightarrow{I\!P_n} \boldsymbol{\theta}_0$.

$\mathbb{A}_1$ For $1 \le r, s \le k$, and $i = 1, 2$, the quantities $D_{r,s}(w_i(x, \boldsymbol{\theta}))$ exist at each $\boldsymbol{\theta} \in \boldsymbol{\Theta}, x \in I\!R$, and $D_r(w_i(\cdot, \boldsymbol{\theta}))$ and $D_{r,s}(w_i(\cdot, \boldsymbol{\theta}))$ are measurable for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. There exists a neighborhood $V(\boldsymbol{\theta}_0) \subset \boldsymbol{\Theta}$ of $\boldsymbol{\theta}_0$ and a measurable function M, with $E(M^2(Z)) < \infty$, such that $\sum_{i=1}^{2} |D_{r,s}(w_i(x, \boldsymbol{\theta}))| + \sum_{i=1}^{2} |D_r(w_i(x, \boldsymbol{\theta}))| \le M(x)$ for all $\theta \in V(\boldsymbol{\theta}_0)$, $x \ge 0$, and $1 \le r, s \le k$.

$\mathbb{A}_2$ The matrix I_0 with elements given by Equation (3.2) is positive definite.

Let $\mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote a k-variate normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. We state our first result, which describes a central limit theorem for the bootstrap version of the MLE $\hat{\boldsymbol{\theta}}$.

Theorem 1 *Suppose that* $\boldsymbol{\Theta} \subset I\!R^k$ *is connected and open, conditions* $\mathbb{C}_1, \mathbb{C}_2, \mathbb{A}_1$, *and* $\mathbb{A}_2$ *hold, and* H *is continuous. Under the proposed two-stage resampling, with* *probability 1,*

$$n^{1/2}\left(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}\right) = n^{-1/2}\sum_{i=1}^{n} I_0^{-1}\mathrm{Grad}(w(\delta_i^*, Z_i^*, \hat{\boldsymbol{\theta}})) + o_{\mathbb{P}_n}(1). \tag{3.7}$$

Furthermore, $n^{1/2}\left(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}\right)$ *is asymptotically distributed as* $\mathcal{N}_k(\boldsymbol{0}, I_0^{-1})$ *with probability 1.*

The proof of Theorem 1, given in the Appendix, uses a standard Taylor's expansion of $S_n^*(\hat{\boldsymbol{\theta}}^*) \equiv \mathrm{Grad}(l_n^*(\hat{\boldsymbol{\theta}}^*))$ about $\hat{\boldsymbol{\theta}}$ to obtain an asymptotic representation for $\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}$ as the average of $\mathrm{Grad}(w_i(\delta_i^*, Z_i^*, \hat{\boldsymbol{\theta}})), i = 1, \ldots, n$, multiplied by $D_{r,s}(l_n^*(\boldsymbol{\theta}))$ evaluated at an intermediate value $\tilde{\boldsymbol{\theta}}$ joining the line segment connecting $\hat{\boldsymbol{\theta}}^*$ and $\hat{\boldsymbol{\theta}}$. Repeated use of a continuity argument and verification of Lindeberg's condition then completes the proof.

We next derive a functional central limit theorem for $\hat{\mathbb{Z}}^*(t) = n^{1/2}(\hat{Q}^*(t) - \hat{Q}(t))$ (Theorem 2 below). Recalling that $\alpha(u, v) = (\mathrm{Grad}(m(u, \boldsymbol{\theta}_0)))^T I_0^{-1}\mathrm{Grad}(m(v, \boldsymbol{\theta}_0))$, we introduce $\hat{\alpha}(u, v) = \left(\mathrm{Grad}(m(u, \hat{\boldsymbol{\theta}}))\right)^T I_0^{-1}\mathrm{Grad}(m(v, \hat{\boldsymbol{\theta}}))$. It is straightforward to show that the following processes which figure in the asymptotic representation for $\hat{\mathbb{Z}}^*(t)$ proved in Theorem 2 below are centered, that is, they have bootstrap expectation zero:

$$A^*(t) := A^*(Z^*, t) = m(Z^*, \hat{\boldsymbol{\theta}})1_{[Z^* \le t]} - \int_0^t m(x, \hat{\boldsymbol{\theta}})d\hat{H}(x), \tag{3.8}$$

$$B^*(t) := B^*(Z^*, \delta^*, t) = \frac{\delta^* - m(Z^*, \hat{\boldsymbol{\theta}})}{m(Z^*, \hat{\boldsymbol{\theta}})\left(1 - m(Z^*, \hat{\boldsymbol{\theta}})\right)}\int_0^t \hat{\alpha}(x, Z^*)\, d\hat{H}(x). \tag{3.9}$$

To prove Theorem 2, we will need some additional regularity conditions which we state now.

$\mathbb{A}_3$ The function $m(t, \boldsymbol{\theta})$ has continuous partial derivatives of second order with respect to $\boldsymbol{\theta}$ and t. Furthermore, for each $\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0) \subset \boldsymbol{\Theta}$, a neighborhood of $\boldsymbol{\theta}_0$,

$$\left\| \int_0^\tau |d\left(\mathrm{Grad}\left(m(x, \boldsymbol{\theta})\right)\right)| \right\| \le M < \infty.$$

$\mathbb{A}_4$ For $1 \leq r \leq k$, $[\partial m(\cdot, \boldsymbol{\theta})/\partial \theta_r]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is Lipschitz on $[0, \tau]$. This means that for an appropriate constant c and any $x, y \in [0, \tau]$, with $H(\tau) < 1$, the following holds:

$$\left| \frac{\partial m(x, \boldsymbol{\theta}_0)}{\partial \theta_r} - \frac{\partial m(y, \boldsymbol{\theta}_0)}{\partial \theta_r} \right| \;\; \leq \;\; c\,|x - y|\,.$$

Remark 1 The first part of condition $\mathbb{A}_3$ is standard [14]. The second part of $\mathbb{A}_3$ is needed in the proof of Theorem 2 to show that the remainder term is $o_{I\!P_n}(1)$ uniformly for $t \in [0, \tau]$; see the remainder term $I_3^*(t)$ occurring at the beginning of the proof of Theorem 2. It simply requires that the total variation of $D_r(m(x, \boldsymbol{\theta}))$ is bounded over $[0, \tau]$ for each $r = 1, \ldots, k$.

Remark 2 Condition $\mathbb{A}_4$ is also standard [14]. It is needed to prove tightness of a centered process $\beta_n^*(x)$ defined in the proof of Theorem 3.

Theorem 2 *Suppose that $\boldsymbol{\Theta} \subset I\!R^k$ is connected and open, conditions $\mathbb{C}_1, \mathbb{C}_2, \mathbb{A}_1 - \mathbb{A}_4$ hold, and H is continuous. Under the proposed two-stage resampling, with probability 1,*

$$\hat{\mathbb{Z}}^*(t) \;\; = \;\; n^{-1/2} \sum_{i=1}^{n} \left(A^*\left(Z_i^*, t\right) + B^*\left(Z_i^*, \delta_i^*, t\right)\right) + o_{I\!P_n}(1). \tag{3.10}$$

In particular, with probability 1, $\hat{\mathbb{Z}}^*$ has the same limit distribution as $\hat{\mathbb{Z}}$.

The proof of Theorem 2 is given in the Appendix. The derived asymptotic representation for $\hat{\mathbb{Z}}^*(t)$ given by Equation (3.10) is used for verifying finite dimensional convergence and tightness, and that the limiting covariance structure matches the expression given by Equation (3.3).

Our final result pertains to a functional central limit theorem for the normalized bootstrap SRCM-based survival function process, denoted by $\hat{\mathbb{W}}^*(t) = n^{1/2}(\hat{S}^*(t) - \hat{S}(t))$, which is the bootstrap version of $\hat{\mathbb{W}}(t)$. First note that, with probability 1, the bivariate process $(\hat{\mathbb{Z}}^*, \hat{\mathbb{H}}^*)$ converges weakly to $(\mathbb{Z}^*, \mathbb{H}^*)$, where $(\mathbb{Z}^*, \mathbb{H}^*)$ is a bivariate Gaussian process with the same covariance structure as the bivariate gaussian process

$(\mathbb{Z}, \mathbb{H})$, the latter being the limit distribution of the bivariate process $(\hat{\mathbb{Z}}, \hat{\mathbb{H}})$. A sequence of mappings operated on the basic bivariate process $(\hat{\mathbb{Z}}^*, \hat{\mathbb{H}}^*)$ produces the process $\hat{\mathbb{W}}^*$. The mappings are compactly differentiable, which by the functional delta method (see Theorem II.8.1 of Andersen et al. (1993) [2]) allows us to deduce the weak convergence of $\hat{\mathbb{W}}^*$.

Theorem 3 *Under the conditions of Theorem 2, with probability 1, $\hat{\mathbb{W}}^*(t)$ has the same limit distribution as $\hat{\mathbb{W}}(t)$.*

The proof of Theorem 3 is given in the Appendix. As discussed in section 3.1, the proof employs Gill and Johansen's (1990) [22] functional delta method.

Remark 3 Since the supremum of the absolute value of a process on a closed interval, $[0, \tau]$ in this case, is a continuous mapping, we may deduce from Theorem 3 and the continuous mapping theorem (see Billingsley (1968) [7]) that with probability 1 both

$$n^{1/2} \sup_{0 \le t \le \tau} \left| \hat{S}(t) - S(t) \right|, \qquad n^{1/2} \sup_{0 \le t \le \tau} \left| \hat{S}^*(t) - \hat{S}(t) \right|$$

have the same limit distribution, permitting us to calibrate the critical values of the first from those of the second.

3.3 Proposed Simultaneous Confidence Bands

In the following, q_α refers to a generic upper α quantile of the distribution of the bootstrap processes. We first introduce untransformed SCBs and then develop further refinements.

Let t_1, t_2 be such that $[t_1, t_2] \subset [0, \tau]$. Since, by the results of the preceding subsection, the processes $\hat{\mathbb{W}}(t) = n^{1/2}(\hat{S}(t) - S(t))$ and $\hat{\mathbb{W}}^*(t) = n^{1/2}(\hat{S}^*(t) - \hat{S}(t))$ are asymptotically equivalent, their quantiles are approximately equal. This yields q_α satisfying the equation

$$P(\sup_{t_1 \le t \le t_2} |\hat{\mathbb{W}}^*(t)| \le q_\alpha) \;=\; 1 - \alpha. \tag{3.11}$$

A $100(1-\alpha)\%$ fixed-width SCB for $S(t)$, referred as "Proposed I" henceforth, is given by

$$\left[\hat{S}(t) - n^{-1/2} q_\alpha, \hat{S}(t) + n^{-1/2} q_\alpha\right].$$

Alternatively, let $\hat{V}(t)$ denote a consistent estimate of $V(t)$, obtained by replacing $S(t)$, $\Lambda(t)$, $H(t)$, and $\boldsymbol{\theta}$ in Equation (3.6) with their estimates. The processes $W_2(t) = \hat{\mathbb{W}}(t)/(S(t)V(t)^{1/2})$ and $W_2^*(t) = \hat{\mathbb{W}}^*(t)/(\hat{S}(t)\hat{V}(t)^{1/2})$ are asymptotically equivalent, which yields q_α satisfying

$$P(\sup_{t_1 \le t \le t_2} |W_2^*(t)| \le q_\alpha) \quad = \quad 1 - \alpha. \tag{3.12}$$

A $100(1-\alpha)\%$ variable-width SCB for $S(t)$, referred as "Proposed II" henceforth, is given by

$$\left[\hat{S}(t) - n^{-1/2}\hat{S}(t)\hat{V}(t)^{1/2} q_\alpha, \hat{S}(t) + n^{-1/2}\hat{S}(t)\hat{V}(t)^{1/2} q_\alpha\right].$$

These SCBs can, however, yield values outside the interval $(0,1)$ and hence require truncation.

Let $g(t) \ge 0$ denote a bounded weight function on $[t_1, t_2]$. We now obtain range-respecting SCBs over $[t_1, t_2] \subset [0, \tau]$. Introducing the transformed process

$$\Psi(t) \quad = \quad n^{1/2} g(t) \left\{ \log\left(-\log\left(\hat{S}(t)\right)\right) - \log\left(-\log\left(S(t)\right)\right) \right\},$$

we note by the functional delta method that $\Psi(t)$ is asymptotically equivalent to the process $U(t) = g(t)\hat{\mathbb{W}}(t)/(S(t)(\log S(t)))$, which, by our results of the preceding subsection, can be approximated by the process $U^*(t) = \hat{g}(t)\hat{\mathbb{W}}^*(t)/(\hat{S}(t)(\log \hat{S}(t)))$, based on the bootstrap data, where $\hat{g}(t)$ is a consistent estimate of $g(t)$. Our choices for g are: $g(t) = S(t)\log(S(t))$ and $g(t) = V(t)^{-1/2}\log S(t)$, where $V(t)$ is given by Equation (3.6). For the first choice $U^*(t)$ reduces to $\hat{\mathbb{W}}^*(t)$ and Equation (3.11) gives q_α. For this case a $100(1-\alpha)\%$ SCB for $S(t)$, referred henceforth as "Proposed III", is given by

$$\left[\hat{S}(t)^{\exp\left(-n^{-1/2}q_\alpha/\left(\hat{S}(t)\log\hat{S}(t)\right)\right)}, \hat{S}(t)^{\exp\left(n^{-1/2}q_\alpha/\left(\hat{S}(t)\log\hat{S}(t)\right)\right)}\right].$$

For the second choice of g, the process $U^*(t)$ reduces to $W_2^*(t)$ and Equation (3.12) gives q_α. A $100(1-\alpha)\%$ SCB for $S(t)$, referred henceforth as "Proposed IV", is then given by

$$\left[\hat{S}(t)^{\exp\left(-n^{-1/2}q_\alpha\hat{V}^{1/2}(t)/\log\hat{S}(t)\right)}, \hat{S}(t)^{\exp\left(n^{-1/2}q_\alpha\hat{V}^{1/2}(t)/\log\hat{S}(t)\right)}\right].$$

Note that, Proposed I band was constructed based on the asymptotic distribution of the process, $\hat{\mathbb{W}}(t) = n^{1/2}(\hat{S}(t) - S(t))$ under the semiparametric random censorship models. This is similar to the construction of Hall–Wellner band [28] and Akritas band [1], both of which were constructed based on the asymptotic distribution of the KM process, $n^{1/2}(\hat{S}_{KM}(t) - S(t))$ under the standard random censorship model. In fact, Akritas (1986) [1] constructed his (nonparametric) bootstrap confidence bands following the Hall–Wellner approach. From this perspective, Proposed I band can be actually seen as a semiparametric version of Hall–Wellner band.

Proposed II band was constructed based on the asymptotic distribution of the normalized process, $W_2(t) = n^{1/2}(\hat{S}(t)-S(t))/(S(t)V(t)^{1/2})$ under the semiparametric random censorship models. This is similar to the construction of Nair's equal precision band [42] developed based on the asymptotic distribution of the normalized KM process, $n^{1/2}(\hat{S}_{KM}(t) - S(t))/(S(t)C(t)^{1/2})$ under the standard random censorship model. So, Proposed II band can be actually seen as a semiparametric version of Nair's equal precision band.

For Proposed III band in which $g(t) = S(t)\log(S(t))$, based on the delta method, the process $\Psi(t)$ is asymptotically equivalent to $\hat{\mathbb{W}}(t)$ from which critical values are calibrated for constructing SCBs. So, Proposed III band is actually a transformed version of Proposed I band.

Similarly, for Proposed IV band in which $g(t) = V(t)^{-1/2} \log S(t)$, the process $\Psi(t)$ is asymptotically equivalent to $W_2(t)$ from which we calibrated critical values to construct SCBs. So, Proposed IV band is actually a transformed version of Proposed II band.

CHAPTER 4

NUMERICAL STUDIES

For our simulation and misspecification studies estimates were calculated repeatedly over 1000 data sets (replications) each with sample size 100. For each simulated data set of size 100, critical values were estimated based on 500 bootstrap resamples.

4.1 Simulation Studies

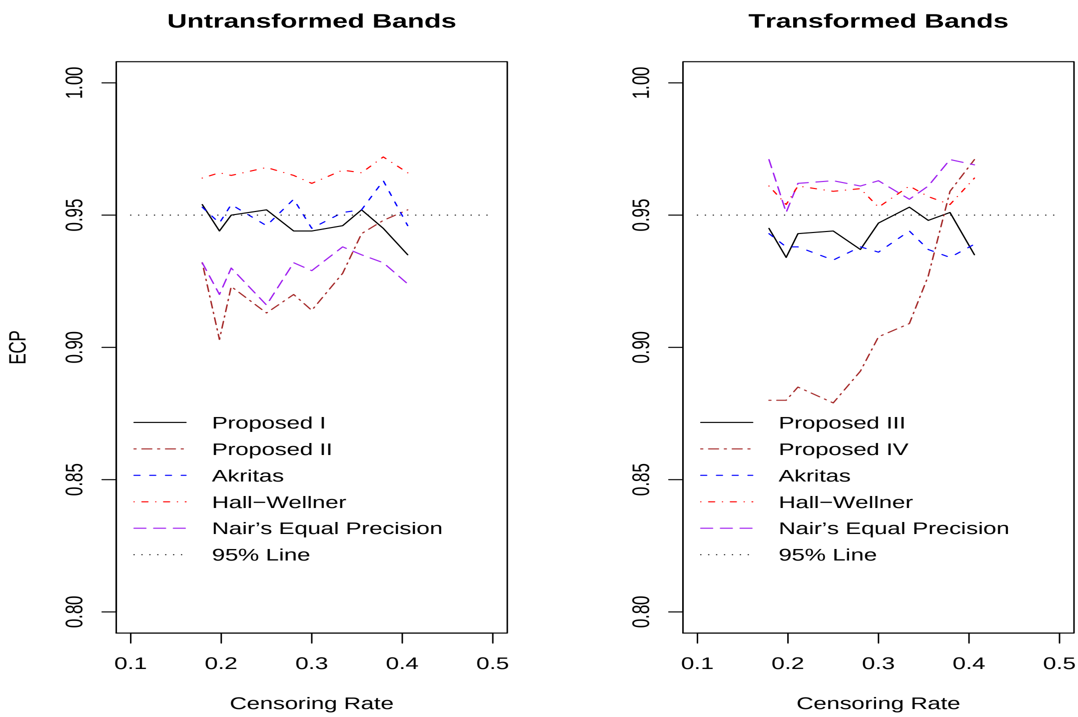

Figure 4.1 Empirical coverage probabilities of several confidence bands for different censoring rates.

The failure and censoring times are generated from two independent Weibull distributions respectively, with $F(t) = 1 - \exp\left(-(t/\beta_1)^{\alpha_1}\right)$, $t \geq 0$ and $G(t) = 1 - \exp\left(-(t/\beta_2)^{\alpha_2}\right)$, $t \geq 0$. Introducing new parameterizations $\theta_1 = (\alpha_1\beta_1^{-\alpha_1})/(\alpha_2\beta_2^{-\alpha_2})$

and $\theta_2 = \alpha_2 - \alpha_1$, and letting $\boldsymbol{\theta}' = (\theta_1, \theta_2)$, the true model for the conditional probability $P(\delta = 1|Z = t)$, called the generalized proportional hazards model (GPHM), is given by $m(t, \boldsymbol{\theta}) = \theta_1/(\theta_1 + t^{\theta_2})$, see Dikta (1998) [13]. Taking $(\alpha_1, \beta_1, \beta_2) = (2, 3, 4.5)$ and varying α_2 in a fine grid of values between 1.1 and 5.5, the censoring rates varied between 18% ($\alpha_2 = 5.5$) and 40% ($\alpha_2 = 1.1$). The ECPs of the proposed and several existing 95% confidence bands are presented in Figure 4.1. The Proposed I, Proposed III, transformed and untransformed Akritas, and transformed Hall–Wellner bands perform well achieving the nominal 95% level. The other bands do not perform as well.

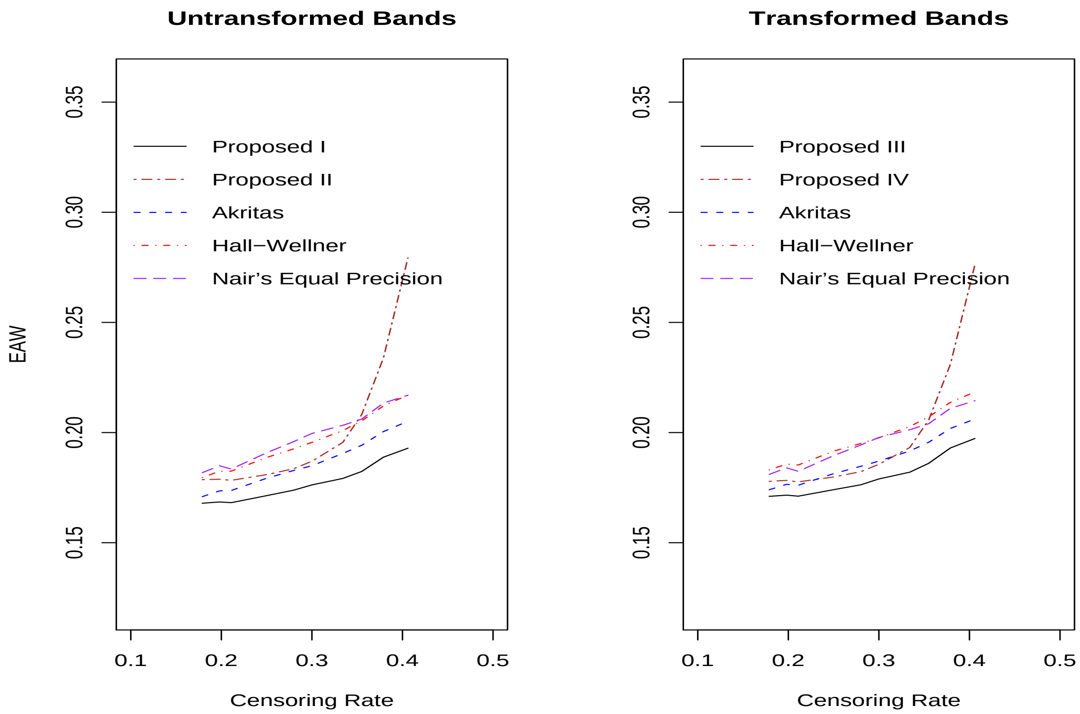

Figure 4.2 Estimated average widths of several confidence bands for different censoring rates.

In Figure 4.2, we present the EAWs of the proposed and several existing SCBs at 95% confidence level. Proposed I and Proposed III perform the best. In particular, for the untransformed scenario, depending on the censoring rate, the Proposed I SCB

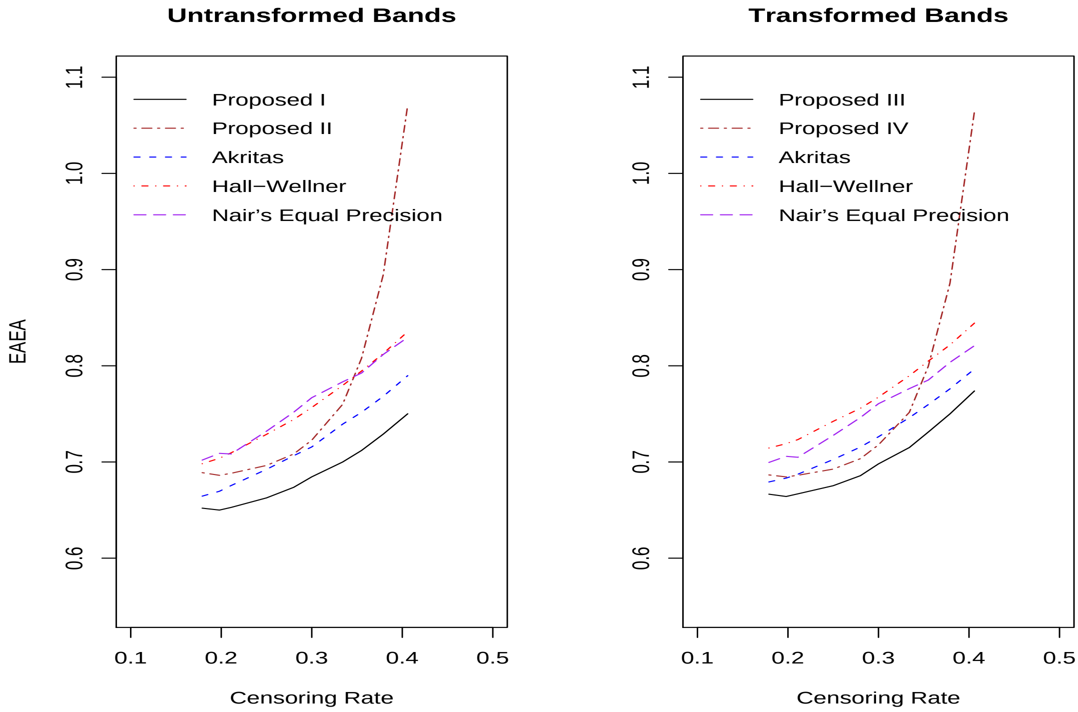

Figure 4.3 Estimated average enclosed areas of several confidence bands for different censoring rates.

offers a reduction of between 1.74% and 6.12% in EAW over its nearest competitor, which is the SCB of Akritas. Likewise for the transformed scenario, depending on the censoring rate, the Proposed III SCB offers a reduction between 1.72% and 5.02% in EAW over its nearest competitor, which is also the SCB of Akritas. The Proposed II and Proposed IV SCBs do not perform well.

Finally, in Figure 4.3, we present the EAEAs of the proposed and several existing 95% SCBs. Proposed I and Proposed III perform the best. In particular, for the untransformed scenario, depending on the censoring rate, the Proposed I SCB offers a reduction of between 1.88% and 5.32% in EAEA over its nearest competitor, which is the SCB of Akritas. Likewise for the transformed scenario, depending on the censoring rate, the Proposed III SCB offers a reduction between 1.87% and 4.16% in EAEA over its nearest competitor, which is also the SCB of Akritas. The Proposed II and

Proposed IV SCBs do not perform well. Therefore, it seems clear that the studentized processes, may not provide improved untransformed or transformed SCBs.

4.2 Misspecification Studies

For our first misspecification study, the minimum Z was uniform on $(0, 1)$ and we generated $m(x)$ according to the following complementary log-log model $m(x, \boldsymbol{\alpha}) = 1 - \exp(-\exp(\alpha_1 + \alpha_2 x))$, where $\boldsymbol{\alpha}^T = (\alpha_1, \alpha_2)$, the parameter α_2 was fixed at -5.92, and the parameter α_1 was assigned several values between 3 and 6, giving censoring rates of between 40% ($\alpha_1 = 3$) and 3% ($\alpha_1 = 6$). Misspecification of $m(t)$ was introduced by fitting the model $m(x, \boldsymbol{\alpha}) = 1 - \exp(-\exp(4 + \alpha_2 x))$ to the generated data. It may be noted that the misspecification of $m(x, \boldsymbol{\alpha})$ increases when α_1 departs from 4. Figure 4.4 gives the ECPs of Proposed I, Proposed III, and other competing 95% SCBs for different values of α_1. The Proposed II and Proposed IV bands were not investigated in view of their poor performance when there was no misspecification. The proposed SCBs provide adequate coverage, comparable to the Hall–Wellner and Akritas bands.

In Figure 4.5, we present the EAWs of Proposed I, Proposed III, and other competing 95% SCBs under misspecification, computed over the entire interval $[3, 6]$. The proposed SCBs outperform their nearest competitor, the Akritas bands. In particular, for both the transformed and untransformed scenarios, our proposed SCBs offer a reduction of up to 1.4% over the SCBs of Akritas. Since the proposed and Akritas bands have comparable coverage over $[3, 6]$, lower EAWs for the proposed bands may be seen as evidence of their marginal superiority over the Akritas bands, for the range of α_1 values that we have considered.

In Figure 4.6, we present the EAEAs of the proposed and the competing 95% SCBs. Proposed I and Proposed III perform the best. The proposed SCBs outperform their nearest competitor, the Akritas bands. Specifically, for both the untransformed

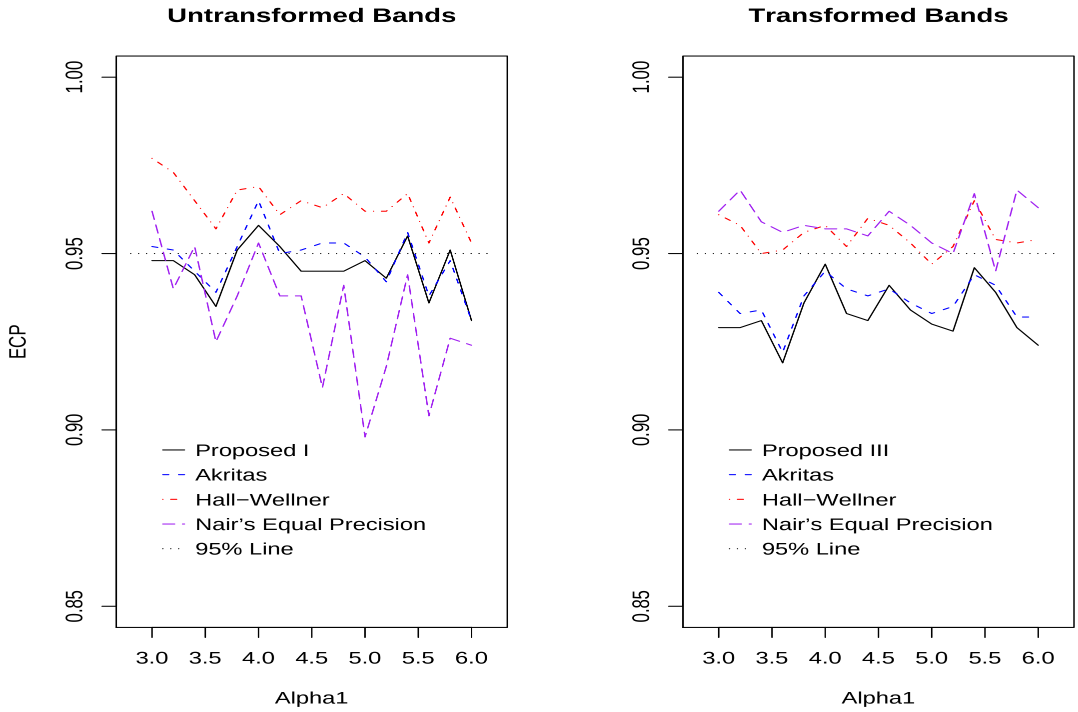

Figure 4.4 Empirical coverage probabilities of proposed and other currently existing confidence bands for the first misspecification study.

and transformed scenarios, depending on the censoring rate, our proposed SCBs offer a reduction of up to 2% over the SCBs of Akritas. Since the proposed and Akritas bands have comparable coverage, lower EAEAs for the proposed bands may be seen as evidence of their superiority over the Akritas bands, for the range of α_1 values that we have considered.

For the second misspecification study, to generate the data we employed the same model as for our simulation study but took $\alpha_1 = 0.8, \beta_1 = 2/3$ and $\beta_2 = 10$. We assigned values for α_2 in a fine grid between 0.3 and 1.3, which gave censoring rates from 32% ($\alpha_2 = 0.3$) to 4% ($\alpha_2 = 1.3$). Although the true model was $m(t, \boldsymbol{\theta}) = \theta_1/(\theta_1 + t^{\theta_2})$, we fit a constant model $m(t) = k$ to the generated data and computed our proposed SCBs using the estimated misspecified model. It may be noted that when $\alpha_2 = 0.8$ there is no misspecification, since $\theta_2 = 0$ and m reduces to a constant.

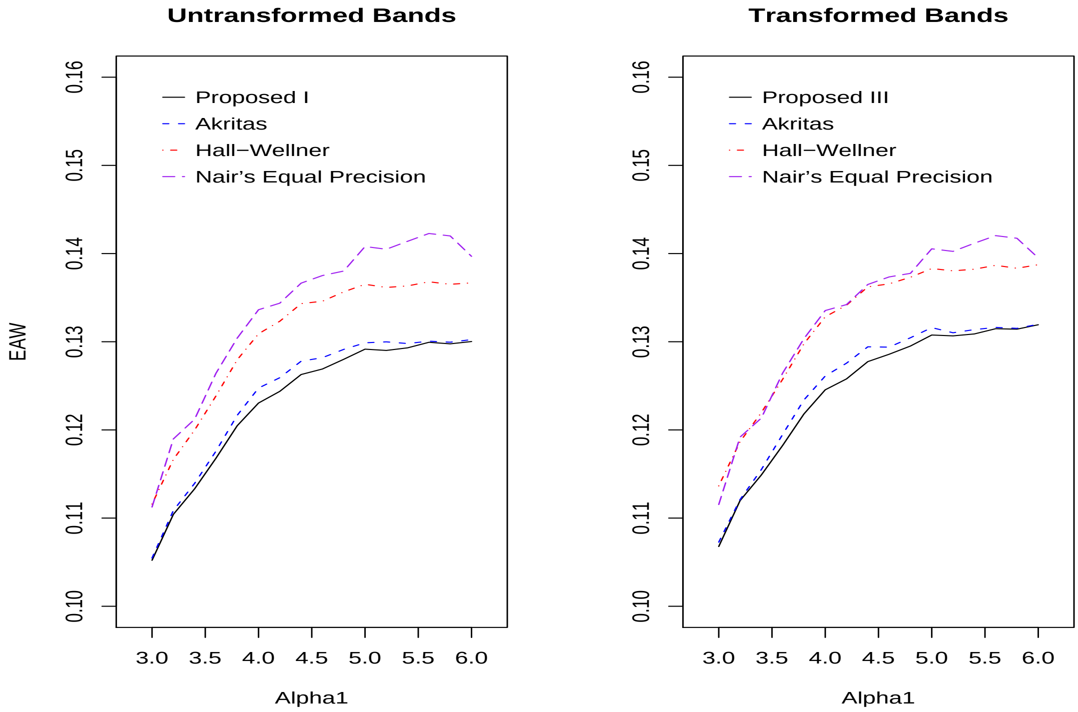

Figure 4.5 Estimated average widths of proposed and other currently existing confidence bands for the first misspecification study.

Figure 4.7 gives the ECPs of Proposed I, Proposed III, and other competing 95% SCBs for different values of α_2. The Proposed II and Proposed IV bands were not investigated in view of their poor performance when there was no misspecification. For values of $\alpha_2 \in [0.55, 1.3]$, the proposed SCBs provide adequate coverage, comparable to the Hall–Wellner and Akritas bands. The proposed bands provide poor coverage for values of α_2 lower than 0.55.

In Figure 4.8, we present the EAWs of Proposed I, Proposed III, and other competing 95% SCBs under misspecification. The proposed bands offer comparable coverage only for $\alpha_2 \in [0.55, 1.3]$. So we focus on this region for following comparisons. The proposed SCBs outperform their nearest competitor, the Akritas bands. In particular, for both the untransformed and transformed scenarios, depending on the censoring rate, our proposed SCBs offer a reduction of up to 10.6% over the SCBs of

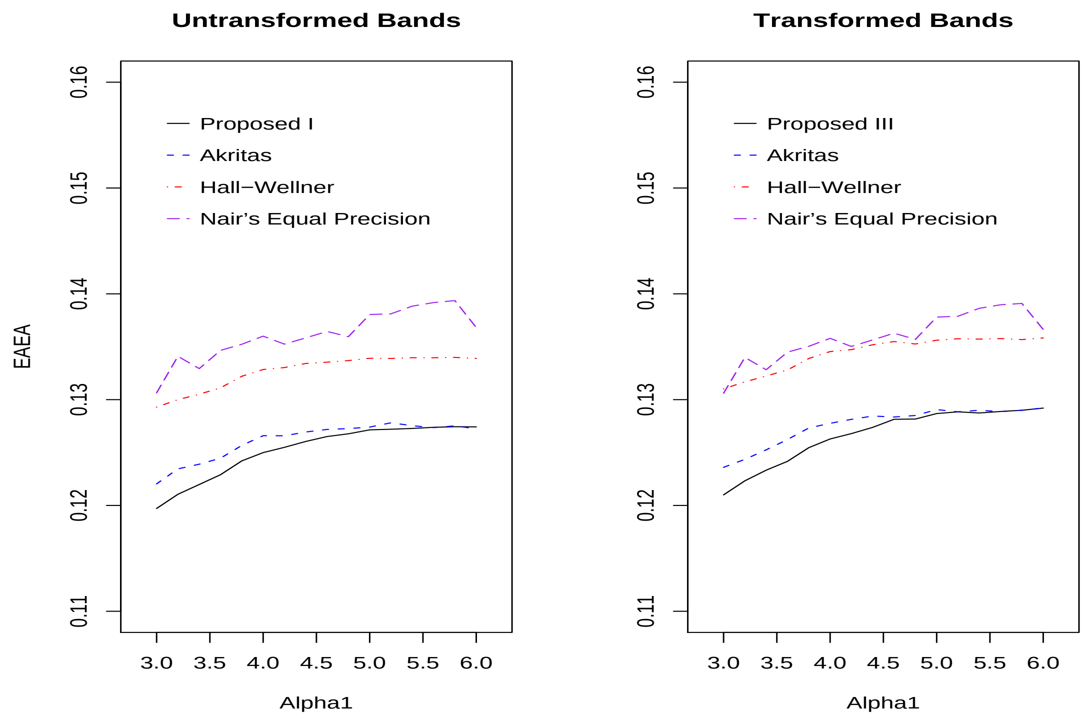

Figure 4.6 Estimated average enclosed areas of proposed and other currently existing confidence bands for the first misspecification study.

Akritas. Lower EAWs for the proposed bands may be seen as evidence of their clear superiority over the Akritas bands.

In Figure 4.9, we present the EAEAs of the proposed and the competing 95% SCBs. The proposed SCBs outperform their nearest competitor, the Akritas bands. Specifically, for both the untransformed and transformed scenarios, depending on the censoring rate, our proposed SCBs offer a reduction of up to 10% over the SCBs of Akritas. Lower EAEAs for the proposed bands may be seen as evidence of their clear superiority over the Akritas bands.

4.3 Real Example Illustrations

We illustrate the new SCBs through two examples. To facilitate comparisons with alternative SCBs, the Hall–Wellner, Nair, and Akritas bands are also plotted. The

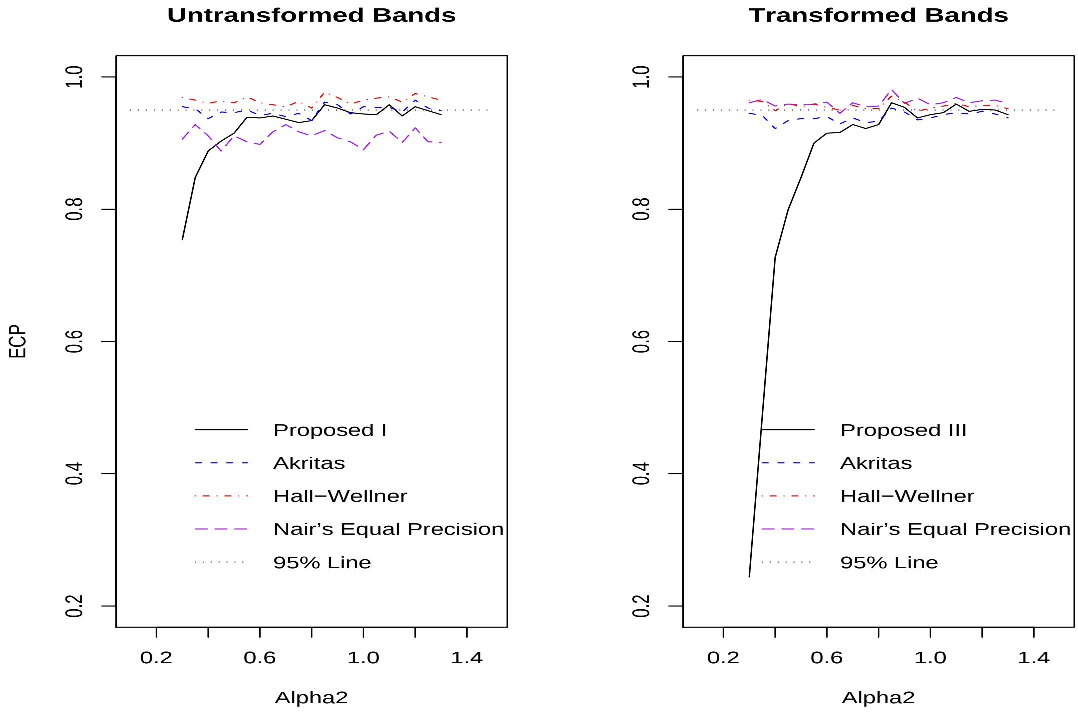

Figure 4.7 Empirical coverage probabilities of proposed and other currently existing confidence bands for the second misspecification study.

objectives of our illustrations are two fold. First, to showcase our methodology as being able to produce bands that are as informative as existing bands based on the defacto choice, which is the KM estimator. Our second objective is to convince the practitioner that obtaining good parametric fits, required by our methodology, is not a cumbersome activity and that our proposed methodology would indeed offer a viable and alternative option to existing KM-based bands. These objectives, we feel, can be achieved using publicly available real data sets. Accordingly, the first data that we utilize for our illustrations are from a lung cancer study reported by Ying, Jung, and Wei (1995) [60]; the second data are from a lung cancer study reported by Gatsonis, Hsieh and Korwar (1985) [20].

A lung cancer study Patients with small cell lung cancer were assigned randomly to two treatments. The response is the base 10 log failure time, and age

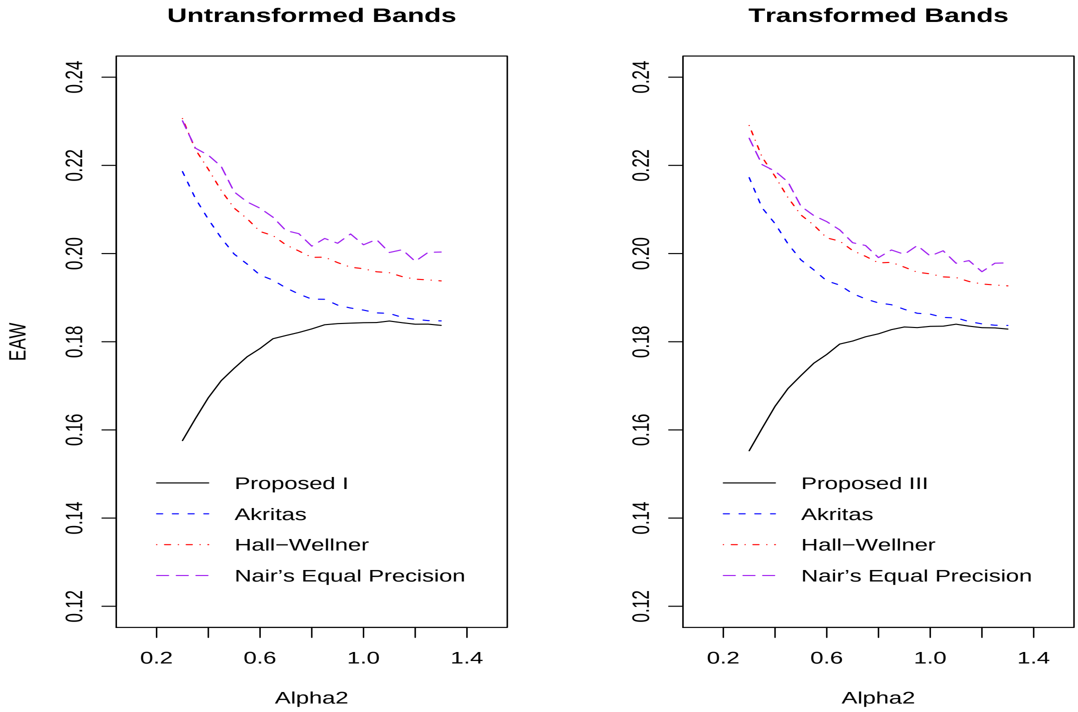

Figure 4.8 Estimated average widths of proposed and other currently existing confidence bands for the second misspecification study.

and treatment indicator were the two covariates. The estimation of the censoring distribution was of particular interest for inverse censoring weighted median regression. Since censoring was administrative, it is free of the covariate, and the two choices are the KM and SRCM-based estimators of its distribution. Subramanian and Dikta (2009) [52] fit the model $m(x, \boldsymbol{\theta}) = (10^x/365)^{\theta_2}/(\theta_1 + (10^x/365)^{\theta_2})$ to the data $(Z_i, 1-\delta_i), i = 1, \ldots, n$. Here $\boldsymbol{\theta} = (\theta_1, \theta_2)'$, with $10^x/365$ being just the original failure time expressed in years. The parameter estimates were reported as $\hat{\theta}_1 = 610$ and $\hat{\theta}_2 = 6.01$ and the SRCM-based survival function estimator showed good agreement with KM estimator suggesting that misspecification may not be an issue. They also performed a formal goodness of fit test, which indicated that the above model was appropriate.

We calculated the SCBs over the interval $[t_1, t_2]$, where t_1 was slightly larger than 0.2273, the smallest uncensored observation, and t_2 was slightly smaller than

Table 4.1 First Lung Cancer Study Example: Percent Reduction in Width and Enclosed Area of Proposed over Competing Untransformed Bands Computed over the Interval [0.2274, 3.6027]

	Width			*Enclosed Area*		
Proposed	*Akritas*	*Hall–Wellner*	*Nair*	*Akritas*	*Hall–Wellner*	*Nair*
I	3.34%	7.77%	7.65%	3.4%	7.77%	7.65%
II	−2.45%	2.25%	2.13%	2.52%	7.07%	4.58%

Table 4.2 First Lung Cancer Study Example: Percent Reduction in Width and Enclosed Area of Proposed over Competing Transformed Bands Computed over the Interval $[0.2274, 3.6027]$

	Width			*Enclosed Area*		
Proposed	*Akritas*	*Hall–Wellner*	*Nair*	*Akritas*	*Hall–Wellner*	*Nair*
III	3.14%	7.42%	−8.71%	2.59%	6.2%	−18.47%
IV	15.23%	18.97%	4.86%	21.63%	24.53%	4.68%

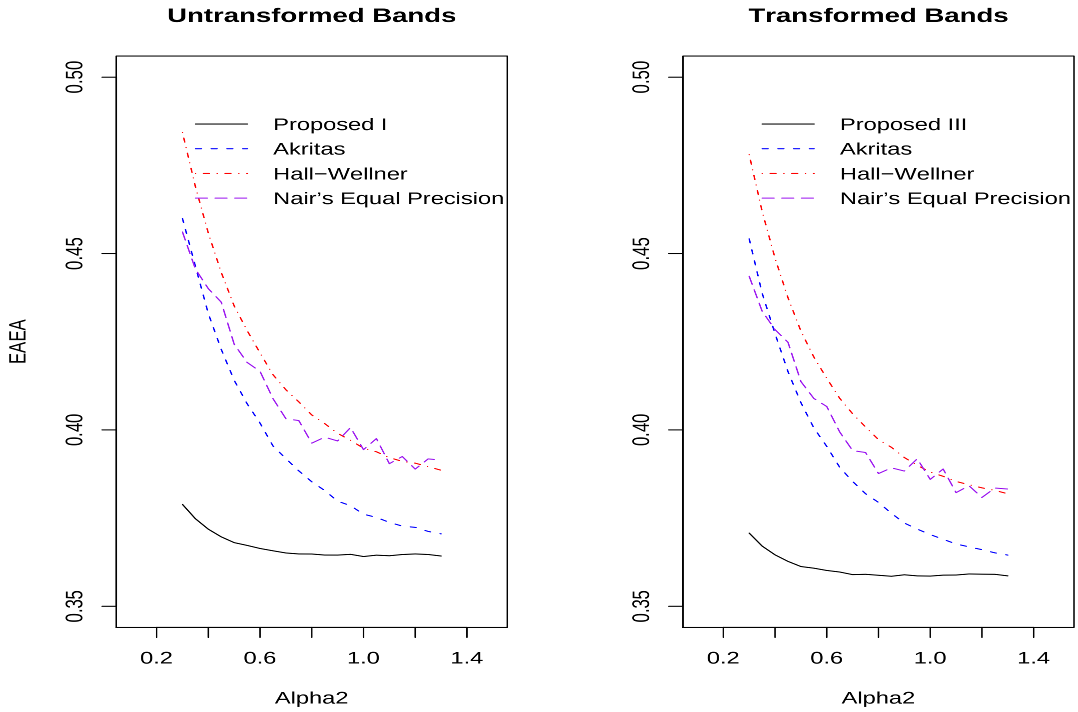

Figure 4.9 Estimated average enclosed areas of proposed and other currently existing confidence bands for the second misspecification study.

3.6028, the largest uncensored observation. The rationale for such truncation was to allow fair comparisons with the Hall–Wellner and Nair bands, which cannot be calculated outside this bound. Tables 4.1 and 4.2 give the percent reduction in empirical width and enclosed area of the proposed over competing bands. Table 4.1 represents untransformed bands and Table 4.2 is for transformed bands. The proposed I and IV bands provide more informative bands than each of the Akritas, Hall–Wellner, and Nair SCBs in terms of the empirical width and enclosed area.

In Figure 4.10 above, we present the KM and SRCM-based estimators of the censoring distribution, together with Proposed III, Proposed IV, and the transformed Akritas, Hall–Wellner, and Nair bands. The untransformed SCBs, consisting of Proposed I, Proposed II, and the other three competing bands, followed nearly the

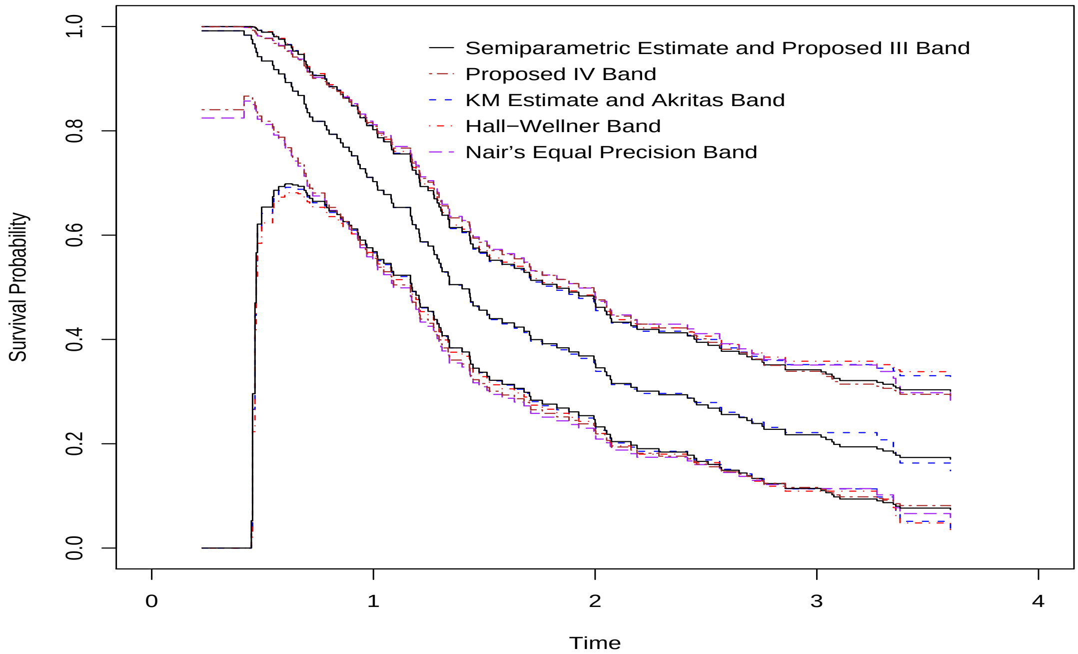

Figure 4.10 Transformed simultaneous confidence bands of the survival function in a lung cancer study.

same pattern as the displayed one except that the irregularity near the lower end of the "Time" axis was absent.

In Table 4.2, we reported the widths and enclosed areas over the interval $[0.2274, 3.6027]$. The poor performance of the Proposed III, transformed Akritas and Hall–Wellner bands in comparison with Proposed IV and Nair bands may be seen to be an artifact of the irregularity near the lower end, see Figure 4.10. We now report figures calculated over the interval $[0.6077, 3.6027]$ in Table 4.3. The Proposed III bands perform best over this interval.

Second lung cancer study Memorial Sloan-Kettering Institute conducted a study of the effects of cisplatin based chemotherapy on lung cancer patients. The data reported by Gatsonis et al. (1985) [20] pertain to survival or censoring times of

Table 4.3 First Lung Cancer Study Example: Percent Reduction in Width and Enclosed Area of Proposed over Competing Transformed Bands Computed over the Interval $[0.6077, 3.6027]$

	Width			*Enclosed Area*		
Proposed	*Akritas*	*Hall–Wellner*	*Nair*	*Akritas*	*Hall–Wellner*	*Nair*
III	5.28%	9.77%	11.51%	1.83%	6.52%	9.12%
IV	−0.47%	4.29%	6.14%	−3.46%	1.49%	4.23%

97, Stage III, non-small cell lung cancer patients who had received no chemotherapy. Approximately one quarter of the patients were still alive at the end of the study and the observations on these patients were treated as censored. We performed formal goodness of fit tests of each of three models for $m(x)$, via the model-based resampling test of Dikta et al. (2006) [14]. The models were the GPHM given by $m(t, \boldsymbol{\theta}) = \theta_1/(\theta_1+t^{\theta_2})$, see also the simulation studies section, and the logistic and probit models. The p-values for the Kolmogorov–Smirnov test, based on 500 bootstrap resamples, were 0.967, 0.778 and 0.810, respectively. The p-values for the Cramer-vón Mises test were 0.914, 0.641 and 0.679, respectively. Since all the three models were not rejected, we used the GPHM for obtaining our parametric fit for $m(x)$.

We constructed the SCBs over the interval $[0.991, 22.539]$. As for the first example we present in Tables 4.4 and 4.5 the percent reduction in empirical width and enclosed area of the proposed over competing bands. Table 4.4 represents untransformed bands and Table 4.5 is for transformed bands. The proposed IV bands perform the best.

In Figure 4.11 below we plot the KM and SRCM-based estimators of the survival function, together with the proposed and other SCBs. Here we report only the untransformed bands as the patterns were similar for transformed bands.

Table 4.4 Second Lung Cancer Study Example: Percent Reduction in Width and Enclosed Area of Proposed over Competing Untransformed Bands Computed over the Interval $[0.991, 22.539]$

	Width			*Enclosed Area*		
Proposed	*Akritas*	*Hall–Wellner*	*Nair*	*Akritas*	*Hall–Wellner*	*Nair*
I	−2.82%	1.89%	3.35%	1.53%	6.39%	10.1%
II	−2.66%	2.05%	3.50%	−1.76%	3.26%	7.1%

Table 4.5 Second Lung Cancer Study Example: Percent Reduction in Width and Enclosed Area of Proposed over Competing Transformed Bands Computed over the Interval $[0.991, 22.539]$

	Width			*Enclosed Area*		
Proposed	*Akritas*	*Hall–Wellner*	*Nair*	*Akritas*	*Hall–Wellner*	*Nair*
III	−1.4%	2.97%	−13.52%	2.94%	7.09%	−4.61%
IV	14%	17.71%	3.73%	13.64%	17.33%	6.92%

Table 4.6 Second Lung Cancer Study Example: Percent Reduction in Width and Enclosed Area of Proposed over Competing Transformed Bands Computed over the Interval $[2.99, 22.539]$

	Width			*Enclosed Area*		
Proposed	*Akritas*	*Hall–Wellner*	*Nair*	*Akritas*	*Hall–Wellner*	*Nair*
III	−1.5%	3.25%	6.04%	3.90%	8.46%	11.92%
IV	−4.52%	0.37%	3.24%	−1.69%	3.13%	6.80%

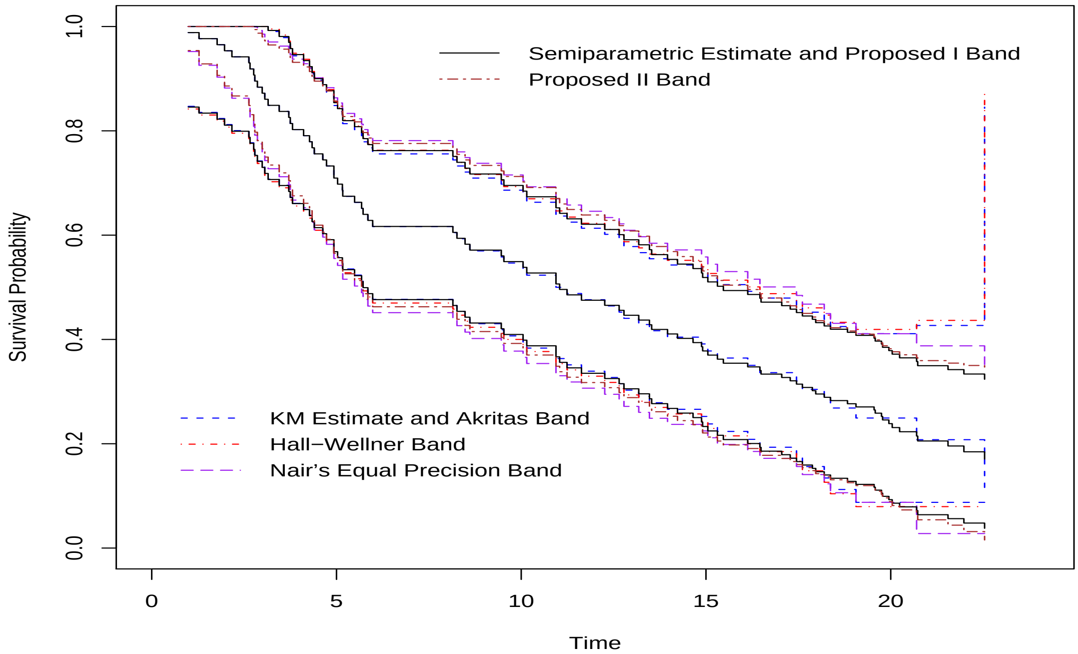

Figure 4.11 Untransformed simultaneous confidence bands of the survival function in a second lung cancer study.

The irregularity noted in the first example was found here as well. In Table 4.6, we report the figures computed over the interval $[2.99, 22.539]$ from which we infer that the Akritas and Proposed III bands perform equally well.

CHAPTER 5

CONCLUSION

The model-based approach of constructing SCBs for survival curves proposed in this paper would be a viable alternative to the existing and established paradigm based on the KM estimator for the one-sample case. In fact, the approach of replacing the censoring indicator with a model-based estimate of its conditional expectation given the covariates applies equally well to the one-sample as well as subject-specific settings and is the first of its kind to be proposed for improved SCB construction for survival curves. A novel extension that addresses SCBs for subject-specific survival would require non-trivial analysis and, for this reason, was not pursued in this article. The proposed approach has sound merit due to the availability of good-fitting models and good model-fitting procedures for binary response data and would be all the more attractive because it is expected to produce more informative SCBs for the survival curve with very little additional effort.

The idea underlying the proposed approach is that parametric specifications, when employed judiciously, lead to more efficient estimation and inference. A novel methodology developed in this project is the bootstrap of the SRCM-based survival function estimator, where the two-stage resampling combines classical bootstrap with model-based regeneration of the censoring indicators to yield a bootstrap that would produce asymptotically correct critical values needed for constructing the proposed SCBs for the one-sample survival curve. This hinges on a new functional central limit theorems for the normalized cumulative hazard and survival function processes in the context of semiparametric random censorship.

APPENDIX A

PROOFS

A.1 Proof of Theorem 1

The normalized bootstrap log likelihood function is given by

$$l_n^*(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\{\delta_i^* w_1(Z_i^*, \boldsymbol{\theta}) + (1-\delta_i^*) w_2(Z_i^*, \boldsymbol{\theta})\}. \tag{1.1}$$

Let $\tilde{\boldsymbol{\theta}}^*$ denote a point inside the line segment joining $\hat{\boldsymbol{\theta}}^*$ and $\hat{\boldsymbol{\theta}}$. Write $A_n^*(\tilde{\boldsymbol{\theta}}^*) = \left(a_{r,s}^n(\tilde{\boldsymbol{\theta}}^*)\right)_{1\le r,s\le k}$, where $a_{r,s}^n(\tilde{\boldsymbol{\theta}}^*) = [\partial^2 l_n^*(\boldsymbol{\theta})/\partial\theta_r\partial\theta_s]_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}^*} = \partial \boldsymbol{S}_n^*(\boldsymbol{\theta})/\partial\boldsymbol{\theta}|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}^*}$ for $1 \le r, s \le k$. A Taylor expansion of $\boldsymbol{S}_n^*(\hat{\boldsymbol{\theta}}^*) = \mathrm{Grad}(l_n^*(\hat{\boldsymbol{\theta}}^*)) \equiv 0$ about $\hat{\boldsymbol{\theta}}$ leads to $\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}} = -\left(A_n^*(\tilde{\boldsymbol{\theta}}^*)\right)^{-1}\cdot \boldsymbol{S}_n^*(\hat{\boldsymbol{\theta}})$. The j-th element of the vector is given by

$$\frac{1}{n}\sum_{i=1}^{n}\left(\delta_i^*\left[\frac{\partial w_1(Z_i^*,\boldsymbol{\theta})}{\partial\theta_j}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} + (1-\delta_i^*)\left[\frac{\partial w_2(Z_i^*,\boldsymbol{\theta})}{\partial\theta_j}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial w(\delta_i^*, Z_i^*,\boldsymbol{\theta})}{\partial\theta_j}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

from which we have

$$\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}} = -\left(A_n^*(\tilde{\boldsymbol{\theta}}^*)\right)^{-1}\cdot\frac{1}{n}\sum_{i=1}^{n}\mathrm{Grad}\left(w(\delta_i^*, Z_i^*, \hat{\boldsymbol{\theta}})\right). \tag{1.2}$$

It remains to prove that, with probability 1, $A_n^*(\tilde{\boldsymbol{\theta}}^*) = -I_0 + o_{\mathbb{P}_n}(1)$ and then derive the asymptotic normality of the average displayed above. Using conditions $\mathbb{C}_1$ and $\mathbb{C}_2$, by Markov's inequality, for any $\varepsilon > 0$ and $1 \le r, s \le k$,

$$\mathbb{P}_n\left(|a_{r,s}^n(\tilde{\boldsymbol{\theta}}^*) - a_{r,s}^n(\boldsymbol{\theta}_0)| > \varepsilon\right) \le \frac{1}{\varepsilon}\mathbb{E}_n\left(\sup_{\boldsymbol{\theta}\in V_\gamma(\boldsymbol{\theta}_0)}\left|a_{r,s}^n(\boldsymbol{\theta}) - a_{r,s}^n(\boldsymbol{\theta}_0)\right|\right) + o_{\mathbb{P}_n}(1),$$

where $V_\gamma(\boldsymbol{\theta}_0)$ is γ-neighborhood of $\boldsymbol{\theta}_0$. The expectation on the right is bounded by

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} I\!E_n\Bigg(\sup_{\boldsymbol{\theta}\in V_\gamma(\boldsymbol{\theta}_0)} \left| \frac{\partial^2 w_1(Z_i^*,\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s} - \left[\frac{\partial^2 w_1(Z_i^*,\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right| \\
&\qquad\qquad + \sup_{\boldsymbol{\theta}\in V_\gamma(\boldsymbol{\theta}_0)} \left| \frac{\partial^2 w_2(Z_i^*,\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s} - \left[\frac{\partial^2 w_2(Z_i^*,\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right| \Bigg) \\
= \; &\frac{1}{n}\sum_{j=1}^{n} \Bigg(\sup_{\boldsymbol{\theta}\in V_\gamma(\boldsymbol{\theta}_0)} \left| \frac{\partial^2 w_1(Z_j,\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s} - \left[\frac{\partial^2 w_1(Z_j,\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right| \\
&\qquad\qquad + \sup_{\boldsymbol{\theta}\in V_\gamma(\boldsymbol{\theta}_0)} \left| \frac{\partial^2 w_2(Z_j,\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s} - \left[\frac{\partial^2 w_2(Z_j,\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right| \Bigg)
\end{aligned}$$

which, by the strong law of large numbers (SLLN), tends with probability 1 to

$$\begin{aligned}
E\Bigg\{ \sup_{\boldsymbol{\theta}\in V_\gamma(\boldsymbol{\theta}_0)} \Bigg(&\left| \frac{\partial^2 w_1(Z,\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s} - \left[\frac{\partial^2 w_1(Z,\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right| \\
&\qquad + \left| \frac{\partial^2 w_2(Z,\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s} - \left[\frac{\partial^2 w_2(Z,\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right| \Bigg) \Bigg\}.
\end{aligned}$$

When $\gamma \to 0$, the above expectation goes to 0 by Lebesgue's theorem. Since γ is arbitrary, we conclude that with probability 1, $A_n^*(\tilde{\boldsymbol{\theta}}^*) = A_n^*(\boldsymbol{\theta}_0) + o_{I\!P_n}(1)$. Furthermore, for $1 \le r, s \le k$, we have that $I\!E_n\left(a_{r,s}^*(\boldsymbol{\theta}_0)\right) = I\!E_n\left([\partial^2 l_n^*(\boldsymbol{\theta})/\partial\theta_r\partial\theta_s]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)$, so that

$$\begin{aligned}
&I\!E_n\left(a_{r,s}^*(\boldsymbol{\theta}_0)\right) \\
= \; &\frac{1}{n}\sum_{i=1}^{n} I\!E_n\left(\delta_i^* \left[\frac{\partial^2 w_1(Z_i^*,\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + (1-\delta_i^*)\left[\frac{\partial^2 w_2(Z_i^*,\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) \\
= \; &I\!E_n\left(\delta^* \left[\frac{\partial^2 w_1(Z^*,\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + (1-\delta^*)\left[\frac{\partial^2 w_2(Z^*,\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) \\
= \; &I\!E_n\Bigg(-\left(\frac{\delta^*}{m^2(Z^*,\boldsymbol{\theta}_0)} + \frac{1-\delta^*}{(1-m(Z^*,\boldsymbol{\theta}_0))^2}\right)\left[\frac{\partial m(Z^*,\boldsymbol{\theta})}{\partial\theta_r}\frac{\partial m(Z^*,\boldsymbol{\theta})}{\partial\theta_s}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&\qquad\qquad + \left(\frac{\delta^*}{m(Z^*,\boldsymbol{\theta}_0)} - \frac{1-\delta^*}{1-m(Z^*,\boldsymbol{\theta}_0)}\right)\left[\frac{\partial^2 m(Z^*,\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \Bigg).
\end{aligned}$$

The second term has conditional expectation given Z^* equal to zero. Using iterated expectation with conditioning on Z^*, it follows by the SLLN that, with probability

1,

$$\begin{aligned}\mathbb{E}_n\left(a^*_{r,s}(\boldsymbol{\theta}_0)\right) &= -\frac{1}{n}\sum_{i=1}^{n}\frac{[(\partial m(Z_i,\boldsymbol{\theta})/\partial\theta_r)(\partial m(Z_i,\boldsymbol{\theta})/\partial\theta_s)]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}}{m(Z_i,\boldsymbol{\theta}_0)(1-m(Z_i,\boldsymbol{\theta}_0))}\\ &\to -E\left(\frac{[(\partial m(Z,\boldsymbol{\theta})/\partial\theta_r)(\partial m(Z,\boldsymbol{\theta})/\partial\theta_s)]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}}{m(Z,\boldsymbol{\theta}_0)(1-m(Z,\boldsymbol{\theta}_0))}\right) \equiv -i_{r,s}.\end{aligned}$$

Therefore, Chebyshev's inequality and $\mathbb{A}_1$ yield that, $A^*_n(\tilde{\boldsymbol{\theta}}^*) = -I_0 + o_{\mathbb{P}_n}(1)$.

To prove the asymptotic normality of the average displayed in Equation (1.2), note that for $i = 1, 2, \ldots, n$,

$$\left[\frac{\partial w(\delta^*_i, Z^*_i, \boldsymbol{\theta})}{\partial\theta_r}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \left(\frac{\delta^*_i}{m(Z^*_i,\hat{\boldsymbol{\theta}})} - \frac{1-\delta^*_i}{1-m(Z^*_i,\hat{\boldsymbol{\theta}})}\right)\left[\frac{\partial m(Z^*_i,\boldsymbol{\theta})}{\partial\theta_r}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

which, using iterated expectation with conditioning on Z^*_i, has mean 0 with probability 1, and hence $\mathrm{Grad}\left(w(\delta^*_i, Z^*_i, \hat{\boldsymbol{\theta}})\right)$ is centered for each $i = 1, 2, \ldots, n$. Furthermore, for fixed $\tilde{\boldsymbol{a}} = (a_1, a_2, \ldots, a_k)' \in \mathbb{R}^k$, we can show that with probability 1,

$$\mathbb{V}_n\left(n^{-1/2}\sum_{i=1}^{n}\tilde{\boldsymbol{a}}'\mathrm{Grad}\left(w(\delta^*_i, Z^*_i, \hat{\boldsymbol{\theta}})\right)\right) \to \tilde{\boldsymbol{a}}' I_0 \tilde{\boldsymbol{a}}. \tag{1.3}$$

Indeed, the term on the left hand side of Equation (1.3) can be expressed as

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\mathbb{V}_n\left(\tilde{\boldsymbol{a}}'\mathrm{Grad}\left(w(\delta^*_i, Z^*_i, \hat{\boldsymbol{\theta}})\right)\right) &= \mathbb{V}_n\left(\tilde{\boldsymbol{a}}'\mathrm{Grad}\left(w(\delta^*, Z^*, \hat{\boldsymbol{\theta}})\right)\right)\\ &= \mathbb{E}_n\left(\left(\tilde{\boldsymbol{a}}'\mathrm{Grad}\left(w(\delta^*, Z^*, \hat{\boldsymbol{\theta}})\right)\right)^2\right),\end{aligned}$$

which equals

$$\sum_{1\le r,s\le k} a_r a_s \mathbb{E}_n\left(\left(\frac{\delta^*}{m^2(Z^*,\hat{\boldsymbol{\theta}})} + \frac{1-\delta^*}{1-m^2(Z^*,\hat{\boldsymbol{\theta}})}\right)\left[\frac{\partial m(Z^*,\boldsymbol{\theta})}{\partial\theta_r}\frac{\partial m(Z^*,\boldsymbol{\theta})}{\partial\theta_s}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right).$$

Using iterated expectation with conditioning on Z^*, the expectation $\mathbb{E}_n$ above becomes

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\frac{[(\partial m(Z_i,\boldsymbol{\theta})/\partial\theta_r)(\partial m(Z_i,\boldsymbol{\theta})/\partial\theta_s)]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}}{m(Z_i,\hat{\boldsymbol{\theta}})(1-m(Z_i,\hat{\boldsymbol{\theta}}))}\\ &\qquad\to E\left(\frac{[(\partial m(Z,\boldsymbol{\theta})/\partial\theta_r)(\partial m(Z,\boldsymbol{\theta})/\partial\theta_s)]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}}{m(Z,\boldsymbol{\theta}_0)(1-m(Z,\boldsymbol{\theta}_0))}\right) = i_{r,s},\end{aligned}$$

with probability 1 as $n \to \infty$, by condition $\mathbb{C}_1$ and the SLLN, proving Equation (1.3). It remains to verify Lindeberg's condition. That is, we need to prove that for every $\varepsilon > 0$, with probability 1,

$$\begin{aligned} & L_n(\varepsilon) \\ = \; & \frac{1}{n}\sum_{i=1}^{n} I\!E_n\left(\frac{\left(\tilde{\boldsymbol{a}}'\mathrm{Grad}(m(Z_i^*,\hat{\boldsymbol{\theta}}))\right)^2\left(\delta_i^* - m(Z_i^*,\hat{\boldsymbol{\theta}})\right)^2}{m^2(Z_i^*,\hat{\boldsymbol{\theta}})\left(1-m(Z_i^*,\hat{\boldsymbol{\theta}})\right)^2} 1_{\left[\left|\frac{\tilde{\boldsymbol{a}}'\mathrm{Grad}(m(Z_i^*,\hat{\boldsymbol{\theta}}))}{m(Z_i^*,\hat{\boldsymbol{\theta}})\left(1-m(Z_i^*,\hat{\boldsymbol{\theta}})\right)}\right|>n^{1/2}\varepsilon\right]}\right) \\ = \; & I\!E_n\left(\frac{\left(\tilde{\boldsymbol{a}}'\mathrm{Grad}(m(Z^*,\hat{\boldsymbol{\theta}}))\right)^2\left(\delta^* - m(Z^*,\hat{\boldsymbol{\theta}})\right)^2}{m^2(Z^*,\hat{\boldsymbol{\theta}})\left(1-m(Z^*,\hat{\boldsymbol{\theta}})\right)^2} 1_{\left[\left|\frac{\tilde{\boldsymbol{a}}'\mathrm{Grad}(m(Z^*,\hat{\boldsymbol{\theta}}))}{m(Z^*,\hat{\boldsymbol{\theta}})(1-m(Z^*,\hat{\boldsymbol{\theta}}))}\right|>n^{1/2}\varepsilon\right]}\right) \;\to\; 0. \end{aligned}$$

Using iterated expectation with conditioning on Z^*, we have that,

$$\begin{aligned} & L_n(\varepsilon) \\ = \; & I\!E_n\left(\frac{\left(\tilde{\boldsymbol{a}}'\mathrm{Grad}(m(Z^*,\hat{\boldsymbol{\theta}}))\right)^2\left(\delta^* - m(Z^*,\hat{\boldsymbol{\theta}})\right)^2}{m^2(Z^*,\hat{\boldsymbol{\theta}})\left(1-m(Z^*,\hat{\boldsymbol{\theta}})\right)^2} 1_{\left[\left|\frac{\tilde{\boldsymbol{a}}'\mathrm{Grad}(m(Z^*,\hat{\boldsymbol{\theta}}))}{m(Z^*,\hat{\boldsymbol{\theta}})(1-m(Z^*,\hat{\boldsymbol{\theta}}))}\right|>n^{1/2}\varepsilon\right]}\right) \\ = \; & I\!E_n\left(\frac{\left(\tilde{\boldsymbol{a}}'\mathrm{Grad}(m(Z^*,\hat{\boldsymbol{\theta}}))\right)^2}{m(Z^*,\hat{\boldsymbol{\theta}})} 1_{\left[\left|\tilde{\boldsymbol{a}}'\mathrm{Grad}(w_1(Z^*,\hat{\boldsymbol{\theta}}))\right|>n^{1/2}\varepsilon\right]}\right. \\ & \qquad \left. + \frac{\left(\tilde{\boldsymbol{a}}'\mathrm{Grad}(m(Z^*,\hat{\boldsymbol{\theta}}))\right)^2}{1-m(Z^*,\hat{\boldsymbol{\theta}})} 1_{\left[\left|\tilde{\boldsymbol{a}}'\mathrm{Grad}(w_2(Z^*,\hat{\boldsymbol{\theta}}))\right|>n^{1/2}\varepsilon\right]}\right) \\ = \; & \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\left(\tilde{\boldsymbol{a}}'\mathrm{Grad}(m(Z_i,\hat{\boldsymbol{\theta}}))\right)^2}{m(Z_i,\hat{\boldsymbol{\theta}})} 1_{\left[\left|\tilde{\boldsymbol{a}}'\mathrm{Grad}(w_1(Z_i,\hat{\boldsymbol{\theta}}))\right|>n^{1/2}\varepsilon\right]}\right. \\ & \qquad \left. + \frac{\left(\tilde{\boldsymbol{a}}'\mathrm{Grad}(m(Z_i,\hat{\boldsymbol{\theta}}))\right)^2}{1-m(Z_i,\hat{\boldsymbol{\theta}})} 1_{\left[\left|\tilde{\boldsymbol{a}}'\mathrm{Grad}(w_2(Z_i,\hat{\boldsymbol{\theta}}))\right|>n^{1/2}\varepsilon\right]}\right) \\ = \; & \frac{1}{n}\sum_{i=1}^{n}\sum_{1\le r,s\le k} a_r a_s \frac{\left[(\partial m(Z_i,\boldsymbol{\theta})/\partial\theta_r)(\partial m(Z_i,\boldsymbol{\theta})/\partial\theta_s)\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}}{m(Z_i,\hat{\boldsymbol{\theta}})\left(1-m(Z_i,\hat{\boldsymbol{\theta}})\right)} \\ & \times\left(\left(1-m(Z_i,\hat{\boldsymbol{\theta}})\right)1_{\left[\left|\tilde{\boldsymbol{a}}'\mathrm{Grad}(w_1(Z_i,\hat{\boldsymbol{\theta}}))\right|>n^{1/2}\varepsilon\right]} + m(Z_i,\hat{\boldsymbol{\theta}})1_{\left[\left|\tilde{\boldsymbol{a}}'\mathrm{Grad}(w_2(Z_i,\hat{\boldsymbol{\theta}}))\right|>n^{1/2}\varepsilon\right]}\right). \end{aligned}$$

The remainder of the proof is analogous to Dikta et al. (2006) [14]. By condition $\mathbb{C}_1$, $\hat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0$ with probability 1. We apply condition $\mathbb{A}_1$ and the Cauchy–Schwarz inequality to bound

$$L_n(\varepsilon) \leq \frac{2}{n}\sum_{i=1}^{n}\sum_{1\leq r,s\leq k} |a_r||a_s|M^2(Z_i)1_{[\|\tilde{\mathbf{a}}\|\sqrt{k}M(Z_i)>n^{1/2}\varepsilon]}.$$

Now, fixing $\lambda > 0$ and using SLLN, we have that for properly selected $c < \infty$,

$$\limsup_{n\to\infty} L_n(\varepsilon) \leq cE\left(M^2(Z)1_{[M(Z)>\lambda]}\right),$$

with probability 1. By condition $\mathbb{A}_1$, however, the right hand side of the inequality above tends to 0 as $\lambda \to \infty$. Therefore, Lindeberg's condition is finally verified. It now follows that asymptotically, $n^{1/2}\left(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}\right)$ and $n^{1/2}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)$ have the same distribution. $\square$

A.2 Proof of Theorem 2

We can write $\hat{\mathbb{Z}}^*(t) := n^{1/2}(\hat{Q}^*(t) - \hat{Q}(t)) = I_1^*(t) + I_2^*(t) + I_3^*(t)$, where

$$\begin{aligned}
I_1^*(t) &= n^{1/2}\int_0^t m(x,\hat{\boldsymbol{\theta}})d\left(\hat{H}^*(x) - \hat{H}(x)\right) \equiv n^{-1/2}\sum_{i=1}^{n} A_i^*(t),\\
I_2^*(t) &= n^{1/2}\int_0^t \left(m(x,\hat{\boldsymbol{\theta}}^*) - m(x,\hat{\boldsymbol{\theta}})\right)d\hat{H}(x),\\
I_3^*(t) &= n^{1/2}\int_0^t \left(m(x,\hat{\boldsymbol{\theta}}^*) - m(x,\hat{\boldsymbol{\theta}})\right)d\left(\hat{H}^*(x) - \hat{H}(x)\right).
\end{aligned}$$

We will denote by $\tilde{\boldsymbol{\theta}}^*$ a point on the line segment joining $\hat{\boldsymbol{\theta}}^*$ and $\hat{\boldsymbol{\theta}}$, with the understanding that it will change with each application of Taylor's or the mean value theorems. We have

$$\begin{aligned}
I_2^*(t) &= n^{1/2}\int_0^t \left(\mathrm{Grad}\left(m(x,\hat{\boldsymbol{\theta}})\right)\right)^T\left(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}\right)d\hat{H}(x)\\
&\quad + n^{1/2}\frac{1}{2}\sum_{1\leq r,s\leq k}\int_0^t \left[\frac{\partial^2 m(x,\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s}\right]_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}^*}\left(\hat{\theta}_r^* - \hat{\theta}_r\right)\left(\hat{\theta}_s^* - \hat{\theta}_s\right)d\hat{H}(x).
\end{aligned}$$

We apply Theorem 1 to deduce that the first term of $I_2^*(t)$ above can be expressed as the sum of the following centered term given the original data plus a remainder term $o_{\mathbb{P}_n}(1)$:

$$n^{-1/2}\sum_{i=1}^{n}\frac{\delta_i^* - m(Z_i^*,\hat{\boldsymbol{\theta}})}{m(Z_i^*,\hat{\boldsymbol{\theta}})\left(1-m(Z_i^*,\hat{\boldsymbol{\theta}})\right)}\int_0^t \hat{\alpha}\left(x,Z_i^*\right)d\hat{H}(x) \;\equiv\; n^{-1/2}\sum_{i=1}^{n}B_i^*(t),$$

uniformly for $t \in [0,\tau]$. By conditions $\mathbb{C}_1 - \mathbb{A}_1$ and Theorem 1 the supremeum over $[0,\tau]$ of the second term of $I_2^*(t)$ is $O\left(n^{1/2}\|\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}\|^2 \mathbb{E}_n\left(M(Z^*)\right)\right) = o_{\mathbb{P}_n}(1)$, uniformly for $t \in [0,\tau]$.

We next show that $I_3^*(t) = o_{\mathbb{P}_n}(1)$ uniformly for $t \in [0,\tau]$. The mean value theorem yields

$$I_3^*(t) \;=\; n^{1/2}(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}})^T\int_0^t \mathrm{Grad}\left(m(x,\tilde{\boldsymbol{\theta}}^*)\right) d\left(\hat{H}^*(x) - \hat{H}(x)\right).$$

After integration by parts it follows from Theorem 1 and the Glivenko–Cantelli lemma that, uniformly for $t \in [0,\tau]$,

$$I_3^*(t) \;=\; n^{1/2}(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}})\int_0^t \left(\hat{H}^*(x) - \hat{H}(x)\right) d\left(\mathrm{Grad}\left(m(x,\tilde{\boldsymbol{\theta}}^*)\right)\right) + o_{\mathbb{P}_n}(1).$$

By condition $\mathbb{A}_3$, the leading term is

$$O\left(n^{1/2}\|\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}\|\left\|\int_0^\tau \left|d\left(\mathrm{Grad}\left(m(x,\tilde{\boldsymbol{\theta}}^*)\right)\right)\right|\right\| \|\hat{H}^* - \hat{H}\|\right) = o_{\mathbb{P}_n}(1),$$

uniformly for $t \in [0,\tau]$, by the Glivenko–Cantelli lemma and Theorem 1. Therefore, for $t \in [0,\tau]$,

$$\hat{\mathbb{Z}}^*(t) = n^{-1/2}\sum_{i=1}^{n}\left[A_i^*(t) + B_i^*(t)\right] + o_{\mathbb{P}_n}(1) \;=\; \gamma_n^*(t) + \int_0^t \beta_n^*(x)d\hat{H}(x) + o_{\mathbb{P}_n}(1),$$

where $\gamma_n^*(t) = n^{-1/2}\sum_{i=1}^{n} A_i^*(t)$ and

$$\beta_n^*(x) \;=\; n^{-1/2}\sum_{i=1}^{n}\frac{\delta_i^* - m(Z_i^*,\hat{\boldsymbol{\theta}})}{m(Z_i^*,\hat{\boldsymbol{\theta}})\left(1-m(Z_i^*,\hat{\boldsymbol{\theta}})\right)}\hat{\alpha}\left(x,Z_i^*\right).$$

The weak convergence of $\hat{\mathbb{Z}}^*(t)$ in $D[0,\tau]$ would follow from the continuous mapping theorem and the weak convergence in $(D[0,\tau])^2$ of (γ_n^*, β_n^*). Finite dimensional convergence to a multivariate normal distribution, with probability 1, can be shown by verifying Lindeberg's condition as in the proof of Theorem 1. Tightness of γ_n^* and β_n^* under the probability measure $I\!P_n$ follows as in lemma 3.13 of Dikta (1998) [13]. For example, to show the tightness of β_n^*, let $\bar{i}_{r,s}$ denote the (r,s) element of I_0^{-1} and write $\beta_n^*(x) = \sum_{1\le r,s\le k} \beta_n^{*r,s}(x)$, where

$$\begin{aligned}\beta_n^{*r,s}(x) \quad = \quad & \left(n^{-1/2} \sum_{i=1}^{n} \frac{\delta_i^* - m(Z_i^*, \hat{\boldsymbol{\theta}})}{m(Z_i^*, \hat{\boldsymbol{\theta}}) \left(1 - m(Z_i^*, \hat{\boldsymbol{\theta}})\right)} \left[\frac{\partial m(Z_i^*, \boldsymbol{\theta})}{\partial \theta_s} \right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right) \\ & \cdot \bar{i}_{r,s} \left[\frac{\partial m(x, \boldsymbol{\theta})}{\partial \theta_r} \right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} .\end{aligned}$$

By condition $\mathbb{A}_3$, the process $\beta_n^{*r,s} \in C[0,\tau]$. Since for any $x_1, x_2 \in [0,\tau]$ and for any $\epsilon > 0$,

$$\begin{aligned} & I\!P_n \left(|\beta_n^{*r,s}(x_2) - \beta_n^{*r,s}(x_1)| \ge \epsilon \right) \\ \le \quad & \frac{1}{\epsilon^2} I\!E_n \left((\beta_n^{*r,s}(x_2) - \beta_n^{*r,s}(x_1))^2 \right) \\ = \quad & \bar{i}_{r,s}^2 \left(\left[\frac{\partial m(x_2, \boldsymbol{\theta})}{\partial \theta_r} \right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} - \left[\frac{\partial m(x_1, \boldsymbol{\theta})}{\partial \theta_r} \right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right)^2 \\ & \times E_n \left(\left(\frac{\delta^* - m(Z^*, \hat{\boldsymbol{\theta}})}{m(Z^*, \hat{\boldsymbol{\theta}}) \left(1 - m(Z^*, \hat{\boldsymbol{\theta}})\right)} \left[\frac{\partial m(Z^*, \boldsymbol{\theta})}{\partial \theta_s} \right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right)^2 \right) \\ \le \quad & c\, |x_2 - x_1|^2 , \end{aligned}$$

for a specially chosen positive constant c, the tightness of $\beta_n^{*r,s}$ follows from condition $\mathbb{C}_1$ and conditions $\mathbb{A}_3$ and $\mathbb{A}_4$ according to Theorem 12.3 in Billingsley (1968) [7]. So, β_n^* is tight in $C[0,\tau]$, which implies tightness in $D[0,\tau]$. Tightness of γ_n^* and β_n^* implies that (γ_n^*, β_n^*) is tight. Therefore we conclude that $\hat{\mathbb{Z}}^*$ converges weakly to a centered Gaussian process $\mathbb{Z}^*$ on $[0,\tau]$.

We now calculate the covariance structure for the limiting process $\mathbb{Z}^*$. Since, for $0 \le t_1 \le t_2 \le \tau$, $\mathrm{Cov}\left(\mathbb{Z}^*(t_1), \mathbb{Z}^*(t_2)\right) = \mathrm{Var}\left(\mathbb{Z}^*(t_1)\right) + \mathrm{Cov}\left(\mathbb{Z}^*(t_1), \mathbb{Z}^*(t_2) - \mathbb{Z}^*(t_1)\right)$, we need to calculate the two components of $\mathrm{Cov}\left(\mathbb{Z}^*(t_1), \mathbb{Z}^*(t_2)\right)$. We first compute $\mathrm{Var}(\mathbb{Z}^*(t))$. We have

$$\begin{aligned}
\mathbb{V}_n\left(A^*(t)\right) &= I\!\!E_n\left(m^2(Z^*, \boldsymbol{\theta}_0)1_{[Z^*\le t]} - 2\hat{Q}(t)m(Z^*, \boldsymbol{\theta}_0)1_{[Z^*\le t]} + \left(\hat{Q}(t)\right)^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} m^2(Z_i, \boldsymbol{\theta}_0)1_{[Z_i\le t]} - 2\hat{Q}(t)\frac{1}{n}\sum_{i=1}^{n} m(Z_i, \boldsymbol{\theta}_0)1_{[Z_i\le t]} + \left(\hat{Q}(t)\right)^2,
\end{aligned}$$

which converges with probability 1 to $\int_0^t m^2(x, \boldsymbol{\theta}_0)dH(x) - Q^2(t)$ as $n \to \infty$. We also have

$$\mathbb{V}_n\left(B^*(t)\right) = I\!\!E_n\left(\frac{\left(\delta^* - m(Z^*, \hat{\boldsymbol{\theta}})\right)^2}{m^2(Z^*, \hat{\boldsymbol{\theta}})\left(1 - m(Z^*, \hat{\boldsymbol{\theta}})\right)^2}\left(\int_0^t \hat{\alpha}\left(x, Z^*\right)d\hat{H}(x)\right)^2\right).$$

To obtain $\mathbb{V}_n\left(B^*(t)\right)$, recall $\hat{\alpha}\left(u, v\right) = \left(\mathrm{Grad}(m(u, \hat{\boldsymbol{\theta}}))\right)^T I_0^{-1}\mathrm{Grad}(m(v, \hat{\boldsymbol{\theta}}))$ and note that

$$\begin{aligned}
\left(\int_0^t \hat{\alpha}\left(x, Z^*\right)d\hat{H}(x)\right)^2 &= \int_0^t\int_0^t \hat{\alpha}(u, Z^*)\hat{\alpha}(Z^*, v)d\hat{H}(u)d\hat{H}(v) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \hat{\alpha}(Z_i, Z^*)\hat{\alpha}(Z^*, Z_j)I(Z_i \le t)I(Z_j \le t).
\end{aligned}$$

Using iterated expectation with conditioning on Z^*, it follows that $\mathbb{V}_n\left(B^*(t)\right)$ equals

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\mathrm{Grad}\left(m(Z_i, \hat{\boldsymbol{\theta}})\right)\right)^T I_0^{-1}I_n(\hat{\boldsymbol{\theta}})I_0^{-1}\mathrm{Grad}\left(m(Z_j, \hat{\boldsymbol{\theta}})\right)I(Z_i \le t)I(Z_j \le t),$$

where $I_n(\hat{\boldsymbol{\theta}})$ defined below is a consistent estimator of I_0 [That is, $I_n(\hat{\boldsymbol{\theta}}) \xrightarrow{P} I_0,$ as $n \to \infty$.]:

$$I_n(\hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{k=1}^{n}\frac{\mathrm{Grad}\left(m(Z_k, \hat{\boldsymbol{\theta}})\right)\left(\mathrm{Grad}\left(m(Z_k, \hat{\boldsymbol{\theta}})\right)\right)^T}{m(Z_k, \hat{\boldsymbol{\theta}})\left(1 - m(Z_k, \hat{\boldsymbol{\theta}})\right)}.$$

Therefore, we deduce that, with probability 1,

$$\mathbb{V}_n\left(B^*(t)\right) \to \int_0^t \int_0^t \alpha(u,v)dH(u)dH(v).$$

Finally, using iterated expectation with conditioning on Z^*, it is straightforward to show that $\mathrm{Cov}_n(A^*(t), B^*(t)) = 0$. Therefore we have that on $[0,\tau]$

$$\mathrm{Var}\left(\mathbb{Z}^*(t)\right) = \int_0^t m^2(x,\boldsymbol{\theta}_0)dH(x) - Q^2(t) + \int_0^t \int_0^t \alpha(u,v)dH(u)dH(v). \quad (1.4)$$

We now calculate $\mathrm{Cov}\left(\mathbb{Z}^*(t_1), \mathbb{Z}^*(t_2) - \mathbb{Z}^*(t_1)\right)$. It is straightforward to show that

$$\begin{aligned}\mathrm{Cov}_n\left(A^*(t_1), A^*(t_2) - A^*(t_1)\right) &= I\!E_n\left(A^*(t_1)\left(A^*(t_2) - A^*(t_1)\right)\right) \\ &= \left(\hat{Q}(t_1)\right)^2 - \hat{Q}(t_1)\hat{Q}(t_2),\end{aligned}$$

which converges to $Q^2(t_1) - Q(t_1)Q(t_2)$ with probability 1. Furthermore,

$$\begin{aligned}\mathrm{Cov}_n\left(B^*(t_1), B^*(t_2) - B^*(t_1)\right) &= I\!E_n\left(B^*(t_1)(B^*(t_2) - B^*(t_1))\right) \\ &\to \int_{t_1}^{t_2} \int_0^{t_1} \alpha(u,v)dH(u)dH(v),\end{aligned}$$

with probability 1. Using iterated expectation with conditioning on Z^*, we also have

$$\mathrm{Cov}_n\left(A^*(t_1), B^*(t_2) - B^*(t_1)\right) = \mathrm{Cov}_n\left(B^*(t_1), A^*(t_2) - A^*(t_1)\right) = 0,$$

which, along with the derivations above, lead to $\mathrm{Cov}\left(\mathbb{Z}^*(t_1), \mathbb{Z}^*(t_2) - \mathbb{Z}^*(t_1)\right)$ given by

$$Q^2(t_1) - Q(t_1)Q(t_2) + \int_{t_1}^{t_2} \int_0^{t_1} \alpha(u,v)dH(u)dH(v). \quad (1.5)$$

From Equation (1.4) and Equation (1.5), the limiting covariance structure of $\hat{\mathbb{Z}}^*(t)$ is exactly equal to that of $\mathbb{Z}(t)$ given by Equation (3.3). We conclude that both have the same asymptotic distribution and hence are asymptotically equivalent on $[0, \tau]$. $\square$

A.3 Proof of Theorem 3

The mapping $\phi : (x, y) \to (x, u) = (x, 1/(1-y))$ is compactly differentiable with derivative evaluated at (h, k) given by $d\phi\,(x, y) \cdot (h, k) = (h, k/(1-y)^2) = (h, j)$. Apply the functional delta method to deduce that with probability 1 the two bivariate processes

$$n^{1/2}\left[\left(\hat{Q}^*, \frac{1}{1-\hat{H}^*_-}\right) - \left(\hat{Q}, \frac{1}{1-\hat{H}_-}\right)\right], n^{1/2}\left[\left(\hat{Q}^*, \frac{\hat{H}^*}{(1-\hat{H}_-)^2}\right) - \left(\hat{Q}, \frac{\hat{H}}{(1-\hat{H}_-)^2}\right)\right]$$

have the same limit distribution. Next, the mapping $\psi : (x, u) \to v = \int_{[0,\cdot]} u dx$ is also compactly differentiable with derivative evaluated at (h, j) given by $d\psi(x, u)\cdot(h, j) = \int_{[0,\cdot]} j dx + \int_{[0,\cdot]} u dh = l$. Apply the functional delta method again to conclude that with probability 1 the following two processes have the same limit distribution:

$$n^{1/2}(\hat{\Lambda}^*(\cdot) - \hat{\Lambda}(\cdot)), \qquad \int_{[0,\cdot]} \frac{\hat{\mathbb{H}}^*(s)}{(1-\hat{H}_-(s))^2} d\,\hat{Q}(s) + \int_{[0,\cdot]} \frac{1}{1-\hat{H}_-(s)} d\,\hat{\mathbb{Z}}^*(s).$$

Finally the mapping $\xi : v \to z = \prod_{[0,\cdot]} (1 - dv)$, is compactly differentiable with derivative $d\xi(v)\cdot l = z \int_{[0,\cdot]} (z_-/z) dl$, so, by the functional delta method, with probability 1, the two processes

$$\hat{\mathbb{W}}^*(\cdot), \qquad -\hat{S}(\cdot) \int_{[0,\cdot]} \frac{1}{1-\Delta\hat{\Lambda}(s)} \left(\frac{\hat{\mathbb{H}}^*(s)}{(1-\hat{H}_-(s))^2} d\,\hat{Q}(s) + \frac{1}{1-\hat{H}_-(s)} d\,\hat{\mathbb{Z}}^*(s) \right)$$

have the same limit distribution. Basic calculations show that, with probability 1, $\hat{\mathbb{W}}^*(\cdot)$ and

$$-\hat{S}(\cdot) \int_{[0,\cdot]} \frac{d\,\hat{\mathbb{Z}}^*(s) + \hat{\mathbb{H}}^*(s) d\hat{\Lambda}(s)}{(1-\Delta\hat{\Lambda}(s))(1-\hat{H}_-(s))} \equiv -\hat{S}(\cdot) \int_{[0,\cdot]} \frac{d\,\hat{\mathbb{K}}^*(s)}{(1-\Delta\hat{\Lambda}(s))(1-\hat{H}_-(s))}$$

are asymptotically equivalent, where

$$\hat{\mathbb{K}}^*(t) = \hat{\mathbb{Z}}^*(t) + \int_0^t \hat{\mathbb{H}}^*(s) d\hat{\Lambda}(s).$$

Note that Theorem 2 gives the limiting covariance structure of $\hat{\mathbb{Z}}^*$, see Equation (3.3) for the final expression. From Theorem 2 we also have $\hat{\mathbb{Z}}^*(t) = A^*(t) + B^*(t) + o_{\mathbb{P}_n}(1)$. Writing

$$C^*(t) = \int_0^t (I(Z^* < s) - \hat{H}_-(s)) d\hat{\Lambda}(s),$$

the following expressions can be verified for $0 \le t_1 \le t_2 \le \tau$: As $n \to \infty$,

$$\begin{aligned}
\mathrm{Cov}_n\left(C^*(t_1), C^*(t_2)\right) &\to \int_0^{t_1} \Lambda(s) dQ(s) + \int_0^{t_1} (Q(t_2) - Q(s)) d\Lambda(s) - Q(t_1)Q(t_2), \\
\mathrm{Cov}_n\left(A^*(t_1), C^*(t_2)\right) &\to -\int_0^{t_1} \Lambda(s) dQ(s) + Q(t_1)Q(t_2), \\
\mathrm{Cov}_n\left(C^*(t_1), A^*(t_2)\right) &\to -\int_0^{t_1} (Q(t_2) - Q(s)) d\Lambda(s) + Q(t_1)Q(t_2).
\end{aligned}$$

From the above expressions it follows that as $n \to \infty$, $\mathrm{Cov}_n\left(\hat{\mathbb{K}}^*(t_1), \hat{\mathbb{K}}^*(t_2)\right)$ converges to

$$\int_0^{t_1} m^2(x, \theta_0) dH(x) + \int_0^{t_2} \int_0^{t_1} \alpha(u, v) dH(u) dH(v), \tag{1.6}$$

Thus, with probability 1, the process $\hat{\mathbb{W}}^*(\cdot)$ has the limiting covariance structure $V(t_1, t_2)$ given by

$$\int_0^{t_1} \frac{m^2(s, \theta_0)}{(1 - H(s))^2} dH(s) + \int_0^{t_1} \int_0^{t_2} \frac{\alpha(u, v)}{(1 - H(u))(1 - H(v))} dH(v) dH(u), \tag{1.7}$$

which is exactly the limiting covariance of $\hat{\mathbb{W}}(\cdot)$, see, example Dikta (1998) [13]. We conclude that with probability 1 both $\hat{\mathbb{W}}^*$ and $\hat{\mathbb{W}}$ have the same limit distribution completing the proof. $\square$

REFERENCES

[1] M. G. Akritas. Bootstrapping the Kaplan–Meier estimator. *Journal of the American Statistical Association*, 81:1032–1038, 1986.

[2] P. K. Andersen, O. Borgan, R. D. Gill, and N. Keiding. *Statistical Models Based on Counting Processes.* Springer-Verlag, New York., 1993.

[3] T. W. Anderson. A modification of the sequential probability ratio test to reduce the sample size. *Annals of Mathematical Statistics*, 31:165–197, 1960.

[4] J. M. Azais, S. Bercu, J. C. Fort, A. Lagnoux, and P. Lé. Simultaneous confidence bands in curve prediction applied to load curves. *Applied Statistics*, 59:Part 5, 2010.

[5] P. J. Bickel and M. Rosenblatt. On some global measures of the deviations of density function estimates. *The Annals of Statistics*, 1:1071–1095, 1973.

[6] O. Bie, Ø. Borgan, and K. Liestø l. Confidence intervals and confidence bands for the cumulative hazard rate function and their small sample properties. *Scandinavian Journal of Statistics*, 14:221–233, 1987.

[7] P. Billingsley. *Convergence of probability measures.* New York, Wiley., 1968.

[8] M. Cheng, P. Hall, and D. Tu. Confidence bands for hazard rate under random censorship. *Biometrika*, 93:357–366, 2006.

[9] G. Claeskens and I. van Keilegom. Bootstrap confidence bands for regression curves and their derivatives. *The Annals of Statistics*, 31:1852–1884, 2003.

[10] D. Collett. *Modelling Binary Data.* CRS Press, Boca Raton, FL., 2nd edition, 2002.

[11] D. R. Cox and E. J. Snell. *Analysis of Binary Data.* Chapman and Hall, London., 2nd edition, 1989.

[12] S. Csörgő and L. Horváth. Confidence bands from censored samples. *Canadian Journal of Statistics*, 14:131–144, 1986.

[13] G. Dikta. On semiparametric random censorship models. *Journal of Statistical Planning and Inference*, 66:253–279, 1998.

[14] G. Dikta, M. Kvesic, and C. Schmidt. Bootstrap approximations in model checks for binary data. *Journal of the American Statistical Association*, 101:521–530, 2006.

[15] G. Dikta, S. Subramanian, and T. Winkler. Bootstrap based model checks with missing binary response data. *Under Review*, 2010.

[16] L. Dümbgen. Optimal confidence bands for shape-restricted curves. *Bernoulli*, 9:423–449, 2003.

[17] B. Efron. The Two Sample Problem with Censored Data. *Fifth Berkeley symposium on mathematical statistics and probability*, 4:831–853, 1967.

[18] B. Efron. Censored data and the bootstrap. *Journal of the American Statistical Association*, 76:312–319, 1981.

[19] T. Fleming and D. Harrington. *Counting Processes and Survival Analysis.* Wiley, New Jersey., 2005.

[20] C. Gatsonis, H. Hsieh, and R. Korwar. Simple nonparametric tests for a known standard survival based on censored data. *Communications in Statistics – Theory and Methods*, 14:2137–2162, 1985.

[21] R. D. Gill. Censoring and Stochastic Integrals. *Mathematical Centre Tracts, Mathematisch Centrum, Amsterdam*, 124, 1980.

[22] R. D. Gill and S. Johansen. A survey of product-integration with a view toward application in survival analysis. *The Annals of Statistics*, 18:1501–1555, 1990.

[23] M. J. Gillespie and L. Fisher. Confidence band for the Kaplan–Meier survival curve estimate. *The Annals of Statistics*, 7:920–924, 1979.

[24] S. Gulati and W. J. Padgett. Families of smooth confidence bands for the survival function under the general random censorship model. *Lifetime Data Analysis*, 2:349–362, 1996.

[25] P. Hall. On bootstrap confidence intervals in nonparametric regression. *The Annals of Statistics*, 20:695–711, 1992.

[26] P. Hall and D. M. Titterington. On confidence bands in nonparametric density estimation and regression. *Journal of Multivariate Analysis*, 27:228–254, 1988.

[27] P. G. Hall, R. J. Hyndman, and Y. Fan. Nonparametric confidence intervals for receiver operating characteristic curves. *Biometrika*, 91:743–750, 2004.

[28] W. L. Hall and J. A. Wellner. Confidence bands for a survival curve from censored data. *Biometrika*, 67:133–143, 1980.

[29] M. Hollander, I. W. McKeague, and J. Yang. Likelihood ratio-based confidence bands for survival functions. *Journal of the American Statistical Association*, 92:215–226, 1997.

[30] M. Hollander and A. Pena. Families of confidence bands for the survival function under the general random censorship model and the KoziolGreen model. *Canadian Journal of Statistics*, 17:59–74, 1989.

[31] L. Horváth and B. S. Yandell. Convergence rates for the bootstrapped product-limit process. *The Annals of Statistics*, 15:1155–1173, 1987.

[32] E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481, 1958.

[33] G. Knafl, J. Sacks, and D. Ylvisaker. Confidence bands for regression functions. *Journal of the American Statistical Association*, 80:683–691, 1985.

[34] E. L. Korn. Confidence bands for isotonic dose-response curves. *Applied Statistics*, 31:59–63, 1982.

[35] G. Li and I. van Keilegom. Likelihood ratio confidence bands in non-parametric regression with censored data. *Scandinavian Journal of Statistics*, 29:547–562, 2002.

[36] D. Y. Lin, T. R. Fleming, and L. J. Wei. Confidence bands for survival curves under the proportional hazards model. *Biometrika*, 81:73–81, 1994.

[37] S.-H. Lo and K. Singh. The product-limit estimator and the bootstrap: some asymptotic representations. *Probability Theory and Related Field*, 71:455–465, 1986.

[38] K. Lu and A. A. Tsiatis. Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics*, 57:1191–1197, 2001.

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

[40] I. W. McKeague and Y. Zhao. Width-scaled confidence bands for survival functions. *Statistics and Probability Letters*, 76:327–339, 2006.

[41] V. N. Nair. Plots and tests for goodness of fit with randomly censored data. *Biometrika*, 68:99–103, 1980.

[42] V. N. Nair. Confidence Bands for survival functions with censored data: a comparative study. *Technometrics*, 26:265–275, 1984.

[43] J. L. Ojeda, J. A. Cristóbal, and J. T. Alcalá. Nonparametric confidence bands construction for GLM models with length biased data. *Journal of Nonparametric Statistics*, 16:421–441, 2004.

[44] M. I. Parzen, L. J. Wei, and Z. Ying. Simultaneous confidence intervals for the difference of two survival functions. *Scandinavian Journal of Statistics*, 24:309–314, 1997.

[45] D. Picard and K. Tribouley. Adaptive confidence interval for pointwise curve estimation. *The Annals of Statistics*, 28:298–335, 2000.

[46] N. Reid. Estimating the median survival time. *Biometrika*, 68:601–608, 1981.

[47] H. Scheffé. *The Analysis of Variance.* New York, Wiley, 1959.

[48] S. Subramanian. The Missing censoring-indicator model of random censorship. *Handbook of Statistics*, 23:123–141, 2004a.

[49] S. Subramanian. The multiple imputations based Kaplan-Meier estimator. *Statistics and Probability Letters*, 79:1906–1914, 2009.

[50] S. Subramanian. Multiple imputations and the missing censoring indicator model. *Journal of Multivariate Analysis*, 102:105–117, 2011.

[51] S. Subramanian and D. Bandyopadhyay. Doubly robust semiparametric estimation for the missing censoring indicator model. *Statistics and Probability Letters*, 80:621–630, 2010.

[52] S. Subramanian and G. Dikta. Inverse censoring weighted median regression. *Statistical Methodology*, 6:594–603, 2009.

[53] J. Sun and C. Loader. Simultaneous confidence bands for linear regression and smoothing. *The Annals of Statistics*, 22:1328–1345, 1994.

[54] J. Sun, C. Loader, and W. P. McCormick. Confidence bands in generalized linear models. *The Annals of Statistics*, 28:429–460, 2000.

[55] Y. Sun, S. Sun, and Y. Diao. Smooth quantile processes from right censored data and construction of simultaneous confidence bands. *Communications in Statistics – Theory and Methods*, 30:707–727, 2001.

[56] A. A. Tsiatis, M. Davidian, and B. McNeney. Multiple imputation methods for testing treatment differences in survival distributions with missing cause of failure. *Biometrika*, 89:238–244, 2002.

[57] J. Tukey. *The Problem of Multiple Comparisons.* Unpublished Manuscript, Princeton Univ., 1953.

[58] Y. Xia. Bias-corrected confidence bands in nonparametric regression. *Journal of the Royal Statistical Society Series B*, 60:797–811, 1998.

[59] L. Yang. Confidence band for additive regression model. *Journal of Data Science*, 6:207–217, 2008.

[60] Z. Ying, S. Jung, and L. J. Wei. Survival analysis with median regression models. *Journal of the American Statistical Association*, 90:178–184, 1995.

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