Bootstrap likelihood ratio confidence bands for survival functions under random censorship and its semiparametric extension

Sundarraman Subramanian
Center for Applied Mathematics and Statistics
Department of Mathematical Sciences
New Jersey Institute of Technology

Abstract

Simultaneous confidence bands (SCBs) for survival functions, from randomly right censored data, can be computed by inverting likelihood ratio functions based on appropriate thresholds. Sometimes, however, the requisite asymptotic distributions are intractable, or thresholds based on Brownian bridge approximations are not easy to obtain when SCBs over only sub-regions are possible or desired. We obtain the thresholds by bootstrapping (i) a nonparametric likelihood ratio function via censored data bootstrap and (ii) a semiparametric adjusted likelihood ratio function via a two-stage bootstrap that utilizes a model for the second stage. These two scenarios are grounded respectively in standard random censorship and its semiparametric extension introduced by Dikta. The two bootstraps, which are different in the way resampling is done, are shown to have asymptotic validity. The respective SCBs are neighborhoods of the well-known Kaplan–Meier estimator and the more recently developed Dikta's semiparametric counterpart. As evidenced by a validation study, both types of SCBs provide approximately correct coverage. The model-based SCBs, however, are tighter than the nonparametric ones. Two sensitivity studies reveal that the model-based method performs well when standard binary regression models are fitted, indicating its robustness to misspecification as well as its practical applicability. An illustration is given using real data.

KEY WORDS: Binary response, Cauchit, Empirical coverage, Gaussian process, Lagrange multiplier, Maximum likelihood estimator.

1 Introduction

In this paper, we develop bootstrap simultaneous confidence bands (SCBs) for survival functions from randomly censored data using the likelihood ratio (LR) approach. Unlike pointwise confidence intervals (PCIs), SCBs are global, allow simultaneous conclusions at multitude time points, present correct estimate of treatment difference over a region, in turn promoting correct decision making. We first develop the SCBs under the random censorship model (RCM), regarded as the de facto framework in which the event and censoring time random variables are independent, their distributions are completely unspecified and, in particular, censoring is noninformative. We then develop alternate SCBs from a semiparametric extension, called SRCMs henceforth, a framework that incorporates informative censoring into RCM through a model for the conditional non-censoring probability and one, which, in particular, is more flexible than a fully parametric specification of the censoring distribution. To emphasize intent, we indicate that a major thrust of the paper is to develop asymptotically valid thresholds for computing nonparametric as well as semiparametric SCBs via the LR.

Let T denote the failure time and let $S_0(t)$ denote the survival function of T. Most SCBs for S_0 under the RCM are Wald-type, based on the weak limit of the normalized Kaplan–Meier (KM) process (Hall and Wellner, 1980; Nair, 1984). In a landmark paper, Thomas and Grunkemeier (1975) introduced the nonparametric LR

E-mail address: sundars@njit.edu.

and inverted it to obtain PCIs for survival probabilities, which gave superior small-sample performance over the normal approximation based methods. PCIs, however, only guarantee correct coverage for each isolated point separately but not for a multitude of points jointly. Hollander, McKeague, and Yang (1997), HMY henceforth, used the nonparametric LR and developed SCBs as "neighborhoods" around S_n, the KM estimator of S_0.

To address the issues precisely, for an independent censoring time C we write $X = \min(T, C)$, $\delta = I(T \leq C)$, and assume that τ_H is such that $H(\tau_H) < 1$, where H is the distribution function for X. Write $S_-(t) = S(t-)$. The observed data are $\{(X_i, \delta_i), 1 \leq i \leq n\}$, where $T_1, \ldots, T_n$ are iid failure times having the survival function $S_0(t)$, and $C_1, \ldots, C_n$ are iid censoring times independent of the T_i's. Let Γ denote the set of survival functions supported by the uncensored lifetimes and $0 < p < 1$. For each fixed t, Thomas and Grunkemeier (1975) obtained PCIs for $S_0(t)$ using the nonparametric LR statistic

$$R_{\mathrm{TG}}(p,t) \quad = \quad \frac{\sup\{\mathrm{Lik}(S) : S(t) = p, S \in \Gamma\}}{\mathrm{Lik}(S_n)}, \tag{1.1}$$

where

$$\mathrm{Lik}(S) \quad = \quad \prod_{i=1}^{n} \left[S_-(X_i) - S(X_i)\right]^{\delta_i} \left[S(X_i)\right]^{1-\delta_i}. \tag{1.2}$$

An asymptotic $100(1-\alpha)\%$ PCI for $S_0(t)$, $0 < \alpha < 1$, is obtained by inverting $-2\log R_{\mathrm{TG}}(p,t)$ using the threshold $\chi^2_{1,\alpha}$, the upper-α quantile of the chi-squared distribution with one degree of freedom. That is, for each fixed t, the collection of points $\{p : -2\log R_{\mathrm{TG}}(p,t) \leq \chi^2_{1,\alpha}\}$ is a $100(1-\alpha)\%$ PCI for $S_0(t)$.

The approach of HMY was to extend Thomas and Grunkemeier's "pointwise" framework to all of $[0, \tau_H]$, the objective being to find an envelope that encloses all survival functions with support over the uncensored time points, and which includes S_0 with $100(1-\alpha)\%$ confidence. HMY's $100(1-\alpha)\%$ SCB for S_0 is given by

$$\mathcal{B}_{\mathrm{NP}} \quad = \quad \{S(t) : \ -2\log R_{\mathrm{TG}}(S(t), t) \leq \tilde{\rho}_n(t), \qquad t \in [0, \tau_H]\}, \tag{1.3}$$

where $\tilde{\rho}_n(t)$ is a threshold determined by $\sigma_n^2(t)$ [cf. Eq. (2.3)] and the percentiles of an appropriate supremum of the absolute value of a Brownian bridge process.

Owen (1988, 1990) gave the first theoretical treatment of the nonparametric LR method. For censored data, the fundamental work of Li (1995) gave theoretical justification of the nonparametric LR. Li, Hollander, McKeague, and Yang (1996) derived nonparametric LR-based SCBs for individual quantile functions. Einmahl and McKeague (1999) generalized Li et al.'s (1997) approach to the k-sample case and developed simultaneous confidence tubes for multiple quantile plots. In this paper, we develop bootstrap LR SCBs for $S_0(t)$ over any $[\epsilon, t_2] \in (0, \tau_H]$ under the framework of RCM as well as for a semiparametric extension. The nonparametric method and its proposed semiparametric counterpart are each based on a separate bootstrap.

For the nonparametric scenario, write $\mathcal{L}_n(S(t), t) \equiv -2\log R_{\mathrm{TG}}(S(t), t)$, which is a function of t, and has the representation given by Eq. (2.1), see HMY. Using S_n, the KM estimator, our first proposal is to bootstrap the distribution of

$$\mathcal{L}_n\left((S_0(t), t\right), \qquad t \in [\epsilon, t_2] \subset (0, \tau_H].$$

We determine ρ_n, the threshold, using our bootstrap approximation of the distribution of the supremum of $\mathcal{L}_n\left(S_0(t), t\right)$ over $[t_1, t_2]$ combined with a weight function based on $\sigma_n^2(t)$. We obtain the nonparametric SCBs

$$\mathcal{B}_{\mathrm{NP}} \quad = \quad \{S(t) : \ \mathcal{L}_n(S(t), t) \leq \rho_n(t), \qquad t \in [\epsilon, t_2] \subset (0, \tau_H]\}, \tag{1.4}$$

see subsection 2.2 for a detailed development of the procedure. For this purpose, we employ the censored data bootstrap (Efron, 1981; Akritas, 1986). The asymptotic justification for the bootstrap approximation of the nonparametric LR can be derived using the techniques that we provide in considerable detail for the more challenging semiparametric scenario. This approach is detailed in subsections 2.1 and 2.2.

Turning to the second scenario, SRCMs present an alternate framework for obtaining SCBs for S_0. By exploiting $\hat{S}$, a semiparametric estimator of S_0 (Dikta, 1998), which is also semiparametric efficient (Dikta, 2014), Wald-type SCBs for S_0 have been developed (Subramanian and Zhang, 2013). However, the resulting SCBs were not based on LR. A compelling and most fundamental argument for the relevance of SRCMs arises from the observation that, under independence of the event and censoring times, $m(x) = P(\delta = 1|X = x)$ is the ratio of the event time hazard to the total hazard, the latter being the sum of event time and censoring hazards (Dikta, 1998; Yuan, 2005). In turn, the censoring hazard is linked to the event time hazard through the multiplicative factor $\exp(-logit(m))$, which is a smooth function of the conditional odds of non-censoring given X. The RCM is indifferent to this relationship, utilizing only the censoring indicator δ in the analysis (noninformative censoring). Alternately, in lieu of δ, a nonparametric estimator such as a kernel or nearest-neighbor estimator can be employed for m, retaining noninformative censoring. However, the latter approach is unappealing due to (i) its reliance on optimal bandwidths which are cumbersome to compute; (ii) possible complications in the LR analysis; and (iii) additional complications arising due to conditional estimation when some censoring indicators are missing. We employ SRCMs, which exploit the link by utilizing a model for $m(x)$. The relationship suggests using logistic regression, which is a standard tool in the analysis of binary response data. Tsiatis, Davidian and McNeney (2002), among others, indicate that careful modeling would provide a good approximation to a conditional probability function, here $m(x)$. However, any appropriate model could be used. McCullagh and Nelder (1989) discuss link functions such as logistic, probit, complementary log-log, and log-log. The probit and Cauchy links, the latter also known as the cauchit, are nested within the Gosset family of links (Koenker and Yoon, 2009). Dikta (1998) discussed the generalized proportional hazards model, which arises when the event and censoring times are each Weibull distributed. Morgan and Smith (1992) used an empirical example in which the cauchit appeared to perform better than the probit link. Our two simulations as well as the real example illustration employ the cauchit link, producing SCBs with approximately correct empirical coverage probabilities and smaller average enclosed areas and widths.

Accordingly, specify a model $m(x, \boldsymbol{\theta})$ for $m(x)$, where $\boldsymbol{\theta} \in I\!R^k$ with true value $\boldsymbol{\theta}_0$. Then, when the model is correctly specified, $m(x, \boldsymbol{\theta}_0) \equiv m(x)$. Dikta (1998, 2014) estimated $\boldsymbol{\theta}$ using $\hat{\boldsymbol{\theta}}$, its maximum likelihood estimator (MLE) and proposed $\hat{S}(t)$, his semiparametric efficient survival function estimator. Replacing δ_i with $m(X_i, \hat{\boldsymbol{\theta}})$ in representation (2.1) produces $\hat{\mathcal{L}}(S(t), t)$ given by Eq. (2.14), see Subramanian (2012), who showed that $\hat{\mathcal{L}}(S(t), t) := -2 \log R(S(t), t)$ for a certain *ad hoc* LR statistic R given by Eq. (2.13) (section 2.4).

Using $\hat{S}$, our second proposal is to bootstrap the distribution of

$$\hat{\mathcal{L}}(S_0(t), t), \qquad t \in [\epsilon, t_2] \subset (0, \tau_H].$$

We then develop our proposed model-based LR SCBs for $S_0(t)$ given by

$$\mathcal{B}_{\mathrm{SP}} = \{S : \hat{\mathcal{L}}(S(t), t) \leq \hat{\rho}(t), \qquad t \in [\epsilon, t_2] \subset (0, \tau_H]\}, \tag{1.5}$$

where $\hat{\rho}(t)$ is an appropriate data-based threshold, calibrated from a bootstrap approximation of the distribution of the supremum of $\hat{\mathcal{L}}(S_0(t), t)$ over $[\epsilon, t_2]$ combined with a weight function based on a variance estimate. Two choices for the variance estimate, namely $\hat{\sigma}^2(t)$ obtained from Eq. (2.10) and $\tilde{\sigma}^2(t)$ given by Eq. (2.16), yield two types of semiparametric SCBs, which we indicate as "SRCM I" and "SRCM II" respectively, see section 3. We employ the two-stage bootstrap introduced by Subramanian and Zhang (2013), which combines classical bootstrap with model-based regeneration of censoring indicators (Dikta et al., 2006). Note that SRCMs require a resampling mechanism that exploits information available through a model for $m(x)$; Efron's (1981) censored data bootstrap is not efficient for SRCMs. The model-based LR SCBs for S_0 are developed in subsections 2.4 and 2.6. Included are a derivation of an asymptotic representation for the semiparametric adjusted LR and proof of asymptotic justification of its (two-stage) bootstrap approximation. These require some new theoretical results and also exploit some results derived by Subramanian and Zhang (2013).

HMY developed their nonparametric SCBs over $[0, \tau_H]$. Note, however, that the validity of the LR-based SCBs rely on the *uniform* convergence rate of a Lagrange multiplier. The pointwise rate derived by Li (1995) can be extended to apply uniformly over compact regions that are proper subsets of $[0, \tau_H]$; in particular, over $[\epsilon, \tau_H]$ for ϵ exceeding, but arbitrarily close to, zero, provided the sample size is adequately large. The 'full' extension to $[0, \tau_H]$, however, requires that $\Psi_0(t)$, the cumulative hazard function of T, needs to be bounded away from 0 over $[0, \tau_H]$, which conflicts with the requirement that $\Psi_0(0) = 0$; see lemma 1 given in the Appendix. We therefore focus on $[\epsilon, \tau_H]$, where $\epsilon > 0$ can be chosen as close to 0 as desired for adequately large sample sizes. For practical data sets, anyway, the SCBs can be constructed only over the region supported by the range of the observed data. Alternate construction of HMY-type SCBs over $[\epsilon, \tau_H]$

would require reliance on special-purpose tables, specific to the type of SCB desired (Hall–Wellner type, equal-precision type), that would need to be generated (Chung, 1986, 1987). This increases the complexity when implementing SCBs over any arbitrary interval that excludes the lower end point of zero.

The article is organized as follows. We develop the bootstrap LR SCBs in section 2. We present our numerical studies (a validation study, two sensitivity studies, and a real example) in section 3. We present some extensions and further discussion in section 4. Technical complements are detailed in the Appendix.

2 Proposed SCBs

To develop the nonparametric and semiparametric bootstrap based SCBs for $S_0(t)$, we define $\tilde{N}(t) = \sum_{i=1}^{n} I(X_i \le t, \delta_i = 1)$, $N(t) = \sum_{i=1}^{n} I(X_i \le t)$, and $Y(t) = \sum_{i=1}^{n} I(X_i \ge t)$. Following HMY, it would be convenient to designate the jump, say, of N, by ΔN.

2.1 Nonparametric LR function

Let $\tilde{D}(t) = \max_{s \le t}(\Delta\tilde{N}(s) - Y(s))$. Recall from HMY that $\mathcal{L}_n(S(t), t) = -2 \log R_{\mathrm{TG}}(S(t), t)$, where

$$\begin{aligned} \mathcal{L}_n\left(S(t), t\right) \quad = \quad & -2\left[\sum_{s \le t}\left(Y(s) - \Delta\tilde{N}(s)\right)\log\left(1 + \frac{\lambda_n(t)}{Y(s) - \Delta\tilde{N}(s)}\right)\right. \\ & \left. - \sum_{s \le t} Y(s)\log\left(1 + \frac{\lambda_n(t)}{Y(s)}\right)\right]. \end{aligned} \tag{2.1}$$

Here $\lambda_n(t) = \lambda_n(S(t), t)$ is the Lagrange multiplier that solves uniquely the equation

$$\prod_{s \le t}\left(1 - \frac{\Delta\tilde{N}(s)}{Y(s) + \lambda(t)}\right) \quad = \quad S(t), \tag{2.2}$$

on $(\tilde{D}(t), \infty)$. This choice for $\tilde{D}(t)$ guarantees that $0 \le \Delta\tilde{N}(s)/(Y(s) + \lambda(t)) < 1$ for all $s \le t$. The Lagrange multiplier that solves Eq. (2.2) with $S = S_0$ will be denoted by $\lambda_{n,0}(t) \equiv \lambda_{n,0}(S_0(t), t)$.

Unless specified otherwise explicitly, we will use $\|h\|_{t_1}^{t_2} \equiv \sup_{t \in [t_1, t_2]} |h(t)|$ to denote the sup-norm of h over $[t_1, t_2]$. However, we will write $\|h\|$ for the sup-norm of h over $[0, \tau_H]$. The LR approach sets the right hand side (RHS) of Eq. (2.1) equal to a threshold value and solves for $\lambda_n(t)$. The threshold is typically taken as the 95th percentile of the distribution of $\|\mathcal{L}_n\left(S_0(t), t\right)\|_{t_1}^{t_2}$. There are two roots for $\lambda_n(t)$, one negative and one positive (Thomas and Grunkemeier, 1975; Li, 1995; HMY), which, when plugged into Eq. (2.2), produces the lower and upper confidence limits for $S_0(t)$; for more details, see the end of section 2.2. Define

$$\sigma_n^2(t) \quad = \quad n\sum_{s \le t}\frac{\Delta\tilde{N}(s)}{Y(s)(Y(s) - \Delta\tilde{N}(s))} \qquad [\, = O_{\mathrm{p}}(1)\,]. \tag{2.3}$$

From HMY, when S_0 is continuous, an asymptotic representation for $\mathcal{L}_n\left(S_0(t), t\right)$ is given by

$$\mathcal{L}_n\left(S_0(t), t\right) \quad = \quad \frac{1}{\sigma_n^2(t)}\left\{n^{1/2}\left(\log S_n(t) - \log S_0(t)\right)\right\}^2 + o_{\mathrm{p}}(1), \tag{2.4}$$

uniformly for $t \in [\epsilon, t_2] \subset (0, \tau_H]$. Motivated by Eq. (2.4), HMY applied a Brownian bridge approximation to the distribution of the supremum over $[0, \tau_H]$ of a scaled signed-root LR to obtain their threshold. Its validity over $[0, \tau_H]$ is not conclusive, however, see Eq. (2.4), since the stated uniformity result holds only over $[\epsilon, \tau_H]$.

2.2 Nonparametric SCBs

Following the standard censored data bootstrap, $\{(X_i^*, \delta_i^*), 1 \le i \le n\}$ are obtained by resampling the observed data $\{(X_i, \delta_i), 1 \le i \le n\}$. For bootstrap analysis, all quantities in subsection 2.1 are replaced with the corresponding bootstrap ones. Specifically, $\tilde{N}^*$ and Y^* are the bootstrap versions of $\tilde{N}$ and Y respectively and $\tilde{D}^*(t) = \max_{s\le t}(\Delta\tilde{N}^*(s) - Y^*(s))$. Also, $\lambda_n^*(t) = \lambda_n^*(S(t), t)$ will denote the bootstrap Lagrange multiplier, which is the unique solution over $(\tilde{D}^*(t), \infty)$ of the bootstrap version of Eq. (2.2), obtained by replacing $\Delta\tilde{N}$ and Y with their bootstrap counterparts. When $S = S_n$, the corresponding Lagrange multiplier will be denoted by $\lambda_{n,0}^*(t) \equiv \lambda_{n,0}^*(S_n(t), t)$. The bootstrap nonparametric LR function, denoted by $\mathcal{L}_n^*(S(t), t)$, is likewise obtained from Eq. (2.1). This choice for $\tilde{D}^*(t)$ ensures that $0 \le \Delta\tilde{N}^*(s)/(Y^*(s) + \lambda(t)) < 1$ for all $s \le t$. We write $\mathbb{P}_n$ and $\mathbb{E}_n$ for the probability measure and expectation associated with the bootstrap. Then, for almost all sample sequences $\{(X_i, \delta_i), 1 \le i \le n\}$, uniformly for $t \in [t_1, t_2]$, it can be shown that

$$\mathcal{L}_n^*(S_n(t), t) \quad = \quad \frac{1}{\sigma_n^2(t)}\left\{n^{1/2}\left(\log S_n^*(t) - \log S_n(t)\right)\right\}^2 + o_{\mathbb{P}_n}(1). \tag{2.5}$$

The proof of Eq. (2.5) can be obtained from the detailed proofs (see the proofs for lemma 1, lemma 2, and lemma 3 in the Appendix) that we give for the more difficult semiparametric scenario, and, hence, omitted.

Recall that, for almost all sample sequences $\{(X_i, \delta_i), 1 \le i \le n\}$, the process $n^{1/2}(S_n^* - S_n)$ converges weakly in $D[0, \tau_H]$ to the same limit as $n^{1/2}(S_n - S)$ (Akritas, 1986). Let $w(t)$ denote a weight function and let $w_n(t)$ denote a consistent estimate of $w(t)$. By the continuous mapping theorem and Eqs. (2.4) and (2.5) it follows that, for almost all sample sequences $\{(X_i, \delta_i), 1 \le i \le n\}$, both $\|w_n(\cdot)\,\mathcal{L}_n^*(S_n(\cdot), \cdot)\|_{t_1}^{t_2}$ and $\|w_n(\cdot)\,\mathcal{L}_n((S_0(\cdot), \cdot)\|_{t_1}^{t_2}$ have the same limit distribution. This, in turn, permits the calibration of percentiles of the latter using the former. Specifically, for each $i = 1, \ldots, n$, when $X_i \in [t_1, t_2]$ and $\delta_i = 1$,

1. Compute $\lambda_{n,0}^*(X_i)$, the bootstrap Lagrange multiplier, by solving the version of Eq. (2.2), obtained by replacing $\Delta\tilde{N}, Y$ and S in Eq. (2.2) with $\Delta\tilde{N}^*, Y^*$ and S_n respectively.
2. Compute $\mathcal{L}_n^*(S_n(X_i), X_i)$ and obtain $B = \max_{\{X_i : X_i \in [t_1, t_2], \delta_i = 1\}} w_n(X_i)\mathcal{L}_n^*(S_n(X_i), X_i)$.
3. Compute B in (2) M times and obtain q_α, the $100M(1-\alpha)$ ordered $B_1, \ldots, B_M$.

Let $\rho_n(t) = q_\alpha / w_n(t)$. The nonparametric LR-based SCBs for S_0 are given by

$$\mathcal{B}_{\mathrm{NP}} \quad = \quad \{S(t) : \; \mathcal{L}_n(S(t), t) \le \rho_n(t), \qquad t \in [t_1, t_2] \subset (0, \tau_H]\}. \tag{2.6}$$

In simulations reported in section 3, we used $w_n(t) = \sigma_n(t)/(1 + \sigma_n^2(t))$.

Remark 1 The jumps of $\tilde{N}^*$ constitute a subset of that of $\tilde{N}$. Therefore the functions in steps (1) and (2) above require computation only at the event times (that is, uncensored times).

Note that $\rho_n(t)$ has jump discontinuities at the distinct uncensored times $X_i \in [t_1, t_2]$. For each fixed t, $\mathcal{L}_n(S(t), t) \equiv \mathcal{L}_n(\lambda, t)$, regarded as a function of λ, is strictly decreasing over $(\tilde{D}(t), 0]$, is zero at 0, and is increasing over $[0, \infty)$; furthermore, $\mathcal{L}_n(\lambda, t) \to \pm\infty$ as $\lambda \to \pm\infty$ (Thomas and Grunkemeier, 1975; Li, 1995; HMY). Hence, for each *uncensored* X_i, exactly two solutions, say $\lambda_{L,i} < 0 < \lambda_{U,i}$, can be obtained by solving the equation $\mathcal{L}_n(\lambda, X_i) = \rho_n(X_i)$. The piecewise constant nonparametric confidence limits over $[X_i, \cdot) \subset [t_1, t_2]$, where the '$\cdot$' indicates the next uncensored time, are

$$\left[\prod_{s \le X_i}\left(1 - \frac{\Delta\tilde{N}(s)}{Y(s) + \lambda_{L,i}}\right) \quad , \quad \prod_{s \le X_i}\left(1 - \frac{\Delta\tilde{N}(s)}{Y(s) + \lambda_{U,i}}\right)\right]. \tag{2.7}$$

Remark 2 For the case of $w_n(t) = 1$, the resulting nonparametric "linear" SCBs provide monotone upper limit. Note for each fixed t that the search for $\lambda(t) \equiv \lambda$, solving $\mathcal{L}_n(S(t), t) \equiv \mathcal{L}_n(\lambda, t) = q_\alpha$, is over $(\tilde{D}(t), \infty)$, see Eq. (2.2). Writing $\mathcal{L}_n'(\lambda, t)$ for the partial derivative of $\mathcal{L}_n(S(t), t) \equiv \mathcal{L}_n(\lambda, t)$ with respect to λ, this implies that the two denominator terms of the integrand in

$$\mathcal{L}_n'(\lambda, t) \quad = \quad \lambda \int_0^t \frac{d\tilde{N}(s)}{(Y(s) - \Delta\tilde{N}(s) + \lambda)(Y(s) + \lambda)} \tag{2.8}$$

are positive; hence the sign of $\mathcal{L}'_n(\lambda, t)$ is the same as that for λ. By Eq. (2.8), therefore, $|\mathcal{L}'_n(\lambda, t_1)| \leq |\mathcal{L}'_n(\lambda, t_2)|$ for $t_1 \leq t_2$. Together with the argument preceding Eq. (2.7), it follows that the two curves are concave upward, with the curve $\mathcal{L}_n(\lambda, t_2)$ nested within $\mathcal{L}_n(\lambda, t_1)$. See the attached graphs in Figure 1, from our simulation study, of $\mathcal{L}_n(\lambda, t)$ as function of λ for 5 values of the event time t.

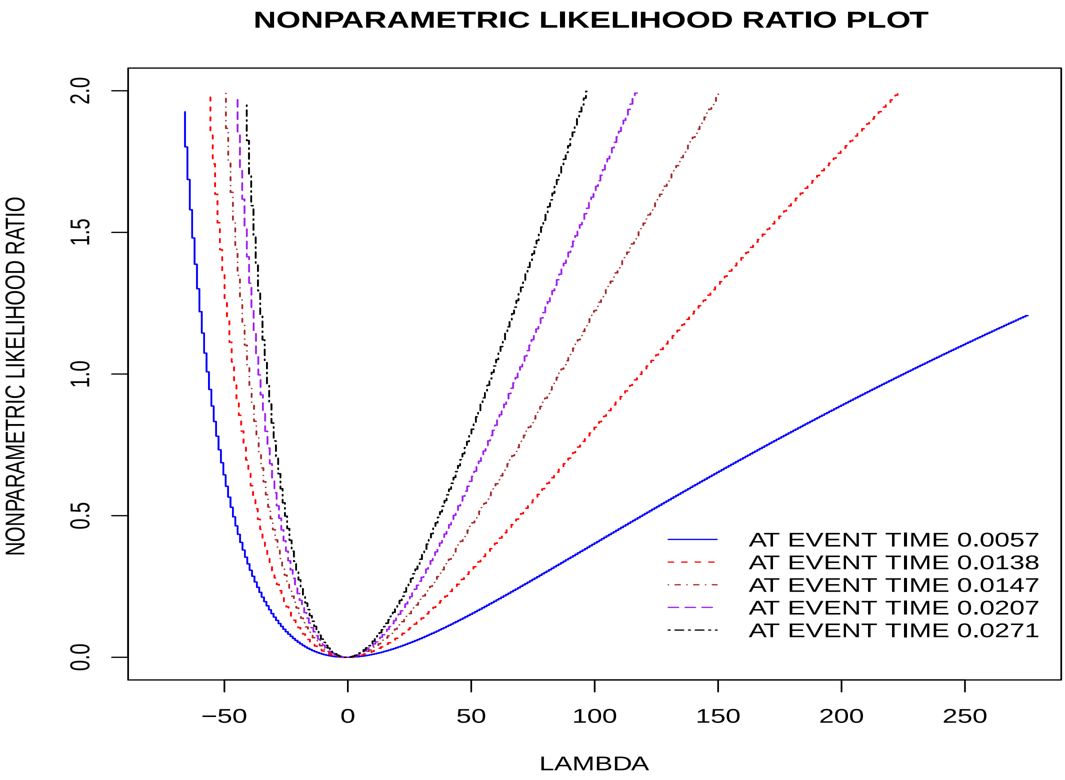

Figure 1: Sample nonparametric likelihood ratio function plot at five event times.

Therefore, (i) $\lambda_{L,t_1} \leq \lambda_{L,t_2} < 0$ and (ii) $0 < \lambda_{U,t_2} \leq \lambda_{U,t_1}$. However, (ii) holds if and only if $0 < 1 - a/(b + \lambda_{U,t_2}) \leq 1 - a/(b + \lambda_{U,t_1}) \leq 1$, where a and b are nonnegative. Note that each term in the product of the second expression inside the square brackets of Eq. (2.7) obeys this monotonicity relation for $t_1 < t_2$. Hence monotone upper limit follows readily. For arbitrary weight functions, however, non-monotone upper limits can appear at the tails.

Remark 3 The case for the monotonicity of the lower limit appears to be less transparent. Our numerical studies indicate that lack of monotone lower limits may be a problem for both the RCM- and SRCM-based SCBs near 0. The apparent lack of monotone lower limits may be due to inadequate data near 0 causing the bootstrap to give a less than reliable threshold (used to compute the two Lagrange multiplier values). Recall also that the rate of convergence of $\hat{\lambda}_0$, the Lagrange multiplier estimate, is uniform over $[\epsilon, \tau_H]$, see lemma 1. As we move closer to the lower end of zero, we need more data to accurately estimate the Lagrange multiplier.

Remark 4 The lack of monotonicity discussed in *Remark 2* and *Remark 3* can be readily addressed, however. A tighter monotone lower limit with identical coverage can be constructed by assigning the maximum value, among the subset exhibiting lack of monotonicity, as the lower limit at all the time points in that subset. The lower limit is flat over that subset but coverage is maintained due to the monotonicity of $S_0(t)$. Likewise for the upper limit with the obvious modifications.

2.3 The semiparametric survival function estimator

We write $\hat{\Psi}(t)$ for the semiparametric cumulative hazard estimator of $\Psi_0(t)$. Recall that $N(t) = \sum_{i=1}^n I(X_i \leq t)$. Dikta's (1998) semiparametric estimator of $S_0(t)$ is given by

$$\hat{S}(t) \;:=\; \prod_{s\leq t}\left(1-\hat{\Psi}(ds)\right) \;=\; \prod_{s\leq t}\left(1-\frac{m(s,\hat{\boldsymbol{\theta}})\Delta N(s)}{Y(s)}\right), \tag{2.9}$$

where $\Delta N(s)/Y(s)$ is defined as 0 when $Y(s)=0$. Dikta (1998) derived a functional central limit theorem for $\hat{\mathbb{W}} = n^{1/2}(\hat{S}-S_0)$ and Dikta (2014) proved semiparametric efficiency. The variance function of the limiting Gaussian process $\mathbb{W}$ at t is $S_0^2(t)\sigma^2(t)$, where

$$\sigma^2(t) \;=\; \int_0^t \frac{m^2(s,\boldsymbol{\theta}_0)}{(1-H(s))^2}dH(s) + \int_0^t\int_0^t \frac{\alpha(u,v)}{(1-H(u))\,(1-H(v))}dH(v)dH(u), \tag{2.10}$$

$\alpha(u,v) = (\mathrm{Grad}(m(u,\boldsymbol{\theta}_0)))^T\, I_0^{-1}\mathrm{Grad}(m(v,\boldsymbol{\theta}_0))$, $\mathrm{Grad}(m(t,\boldsymbol{\theta})) = [D_1(t,\boldsymbol{\theta}),\ldots,D_k(t,\boldsymbol{\theta})]^T$, $D_r(m(t,\boldsymbol{\theta}))$ is the partial derivative of $m(t,\boldsymbol{\theta})$ with respect to $\theta_r, r=1,\ldots,k$, and

$$I_0 \;=\; E\left[\frac{\mathrm{Grad}(m(X,\boldsymbol{\theta}_0))\mathrm{Grad}^T(m(X,\boldsymbol{\theta}_0))}{m(X,\boldsymbol{\theta}_0)(1-m(X,\boldsymbol{\theta}_0))}\right]. \tag{2.11}$$

We denote by $\hat{\sigma}^2(t)$ the variance estimate obtained by substituting unknown quantities in Eq. (2.10) with standard (empirical) estimates.

2.4 Semiparametric LR function

Let $\mathcal{D}$ denote the space of survival functions on $[0,\infty)$ supported by $X_1,\ldots,X_n$ (*all* observed times, uncensored as well as censored). Note that $\hat{S}\in\mathcal{D}$. For any $S\in\mathcal{D}$, define

$$L(S) \;=\; \prod_{i=1}^n [S_-(X_i)-S(X_i)]^{m(X_i,\hat{\boldsymbol{\theta}})}\,[S(X_i)]^{1-m(X_i,\hat{\boldsymbol{\theta}})}, \tag{2.12}$$

which is analogous to Eq. (1.2), but with $m(X_i,\hat{\boldsymbol{\theta}})$ now playing the role of δ_i. Subramanian (2012) showed that $\hat{S}(t)$ maximizes $L(S)$ over $\mathcal{D}$. In the spirit of Eq. (1.1), he proposed the semiparametric adjusted LR statistic

$$R(p,t) \;=\; \frac{\sup\{L(K): K(t)=p, K\in\mathcal{D}\}}{L(\hat{S})}. \tag{2.13}$$

Recall that $\hat{\mathcal{L}}(S(t),t) := -2\log R(S(t),t)$. For *fixed* t, an asymptotic $100(1-\alpha)\%$ PCI for $S_0(t)$ can be obtained by inverting using a threshold depending on $\chi^2_{1,\alpha}$ and some variance estimates. Writing $D(t) = \max_{s\leq t}(m(s,\hat{\boldsymbol{\theta}})\Delta N(s) - Y(s))$, we recall (cf. Subramanian, 2012) that

$$\begin{aligned}\hat{\mathcal{L}}(S(t),t) \;=\; -2&\left[\sum_{s\leq t}\left(Y(s)-m(s,\hat{\boldsymbol{\theta}})\Delta N(s)\right)\log\left(1+\frac{\hat{\lambda}(t)}{Y(s)-m(s,\hat{\boldsymbol{\theta}})\Delta N(s)}\right)\right.\\ &\left.-\sum_{s\leq t}Y(s)\log\left(1+\frac{\hat{\lambda}(t)}{Y(s)}\right)\right],\end{aligned} \tag{2.14}$$

where $\hat{\lambda}(t) = \hat{\lambda}(S(t),t)$, the Lagrange multiplier, is the unique solution of

$$\prod_{s\leq t}\left(1-\frac{m(s,\hat{\boldsymbol{\theta}})\Delta N(s)}{Y(s)+\lambda(t)}\right) \;=\; S(t), \tag{2.15}$$

on $(D(t), \infty)$. Note that $0 \leq m(s, \hat{\boldsymbol{\theta}})\Delta N(s)/(Y(s) + \lambda(t)) < 1$ for all $s \leq t$. The Lagrange multiplier that solves Eq. (2.15) with $S = S_0$ will be denoted by $\hat{\lambda}_0(t) \equiv \hat{\lambda}(S_0(t), t)$. Define

$$\tilde{\sigma}^2(t) \quad = \quad n \sum_{s \leq t} \frac{m(s, \hat{\boldsymbol{\theta}})\Delta N(s)}{Y(s)(Y(s) - m(s, \hat{\boldsymbol{\theta}})\Delta N(s))} \qquad [\, = O_{\mathrm{p}}(1)\,]. \tag{2.16}$$

We then have the following semiparametric analog of Eq. (2.4). The proof is given in the Appendix.

Theorem 1 *Uniformly for $t \in [t_1, t_2] \subset (0, \tau_H]$,*

$$\hat{\mathcal{L}}\left((S_0(t), t\right) \quad = \quad \frac{1}{\tilde{\sigma}^2(t)} \left\{ n^{1/2} \left(\log \hat{S}(t) - \log S_0(t) \right) \right\}^2 + o_{\mathrm{p}}(1). \tag{2.17}$$

To establish asymptotic justification, an analogous result will be given for the bootstrap semiparametric LR.

2.5 The two-stage bootstrap

The two-stage bootstrap proposed by Subramanian and Zhang (2013) is given as follows:
(1) Generate $X_i^*, i = 1, \ldots, n$, from $\hat{H}(t)$, where $\hat{H}(t)$ is the empirical estimator of $H(t)$.
(2) Generate $\delta_i^*, i = 1, \ldots, n$, from a Bernoulli distribution having success probability $m(X_i^*, \hat{\boldsymbol{\theta}})$.

We retain the notation $\mathbb{P}_n$ and $\mathbb{E}_n$ for the probability measure and expectation associated with the two-stage bootstrap. For proving the asymptotic justification in subsection 2.6, we prove some new results and recall some existing ones. We will assume the regularity conditions given in Appendix A of Subramanian and Zhang (2013), most of which were essentially introduced by Dikta (1998). For completeness, these conditions are included here. The normalized log likelihood function is given by $l_n(\boldsymbol{\theta}) = \sum_{i=1}^n w(X_i, \delta_i, \boldsymbol{\theta})/n$, where

$$w(X_i, \delta_i, \boldsymbol{\theta}) \quad = \quad \delta_i \log(m(X_i, \boldsymbol{\theta})) + (1 - \delta_i) \log(1 - m(X_i, \boldsymbol{\theta})), \qquad i = 1, \ldots, n. \tag{2.18}$$

Write $\mathrm{Grad}(m(t, \boldsymbol{\theta})) = [D_1(t, \boldsymbol{\theta}), \ldots, D_k(t, \boldsymbol{\theta})]^T$, where $D_r(m(t, \boldsymbol{\theta}))$ is the partial derivative of $m(t, \boldsymbol{\theta})$ with respect to $\theta_r, r = 1, \ldots, k$. Also denote the second order partial derivatives by $D_{r,s}(\cdot)$. From Dikta (1998), the MLE $\hat{\boldsymbol{\theta}} \in \boldsymbol{\Theta} \subset \mathbb{R}^k$ is a measurable solution of $\mathrm{Grad}(l_n(\boldsymbol{\theta})) = 0$ satisfying $\hat{\boldsymbol{\theta}} \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0$. Write $w_1(x, \boldsymbol{\theta}) = \log(m(x, \theta))$ and $w_2(x, \boldsymbol{\theta}) = \log(1 - m(x, \boldsymbol{\theta}))$. The regularity conditions are:

$\mathbb{A}_1$ For $1 \leq r, s \leq k$, and $i = 1, 2$, the quantities $D_{r,s} w_i(x, \boldsymbol{\theta})$ exist at each $\boldsymbol{\theta} \in \boldsymbol{\Theta}, x \in \mathbb{R}$, 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 $E(M^2(X)) < \infty$, such that $\sum_{i=1}^2 D_{r,s}(w_i(x, \boldsymbol{\theta})) + \sum_{i=1}^2 D_r(w_i(x, \boldsymbol{\theta})) \leq M(x)$ for all $\theta \in V(\boldsymbol{\theta}_0)$, $x \geq 0$, and $1 \leq r, s \leq k$.
$\mathbb{A}_2$ The matrix I_0 defined in section 2.3 of the paper is positive definite.
$\mathbb{A}_3$ The function $m(t, \boldsymbol{\theta})$ has continuous second order partial derivatives with respect to $\boldsymbol{\theta}$ & t.
$\mathbb{A}_4$ For each $\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0) \subset \boldsymbol{\Theta}$, $\left\| \int_0^\tau |d\left(\mathrm{Grad}\left(m(x, \boldsymbol{\theta})\right)\right)| \right\| \leq M < \infty$
$\mathbb{A}_5$ For $1 \leq r \leq k$ and $x \in [0, \tau]$ the function $D_r(m(x, \boldsymbol{\theta}_0))$ is Lipschitz. This means that for an appropriate constant c, and any $x, y \in [0, \tau_H]$, with $H(\tau_H) < 1$, the following holds:

$$|D_r(x, \boldsymbol{\theta}) - D_r(y, \boldsymbol{\theta})| \leq c|x - y|.$$

Conditions $\mathbb{A}_1, \mathbb{A}_2, \mathbb{A}_3$, and $\mathbb{A}_5$ were given by Dikta (1998) to prove the asymptotic normality of $\hat{\boldsymbol{\theta}}$ and to derive a functional central limit theorem for the SRCM-based survival function estimator of $S_0(t)$. Condition $\mathbb{A}_4$ was given by Subramanian and Zhang (2013) for deriving the asymptotic validity of the two-stage bootstrap for the SRCM-based survival function estimator of $S_0(t)$, see their theorem 2 as well as proposition 3 of the present paper. Condition $\mathbb{A}_5$ is also needed for bootstrapping the SRCM-based survival function estimator of $S_0(t)$. Note that condition $\mathbb{C}_1$ of Subramanian and Zhang (2013) is implied by theorem 1 proved below. In addition, we will need condition **A** below (where, $\| \cdot \|$ is the Euclidean norm) for proving Eq. (2.20) and proposition 2 :

Condition A: For a neighborhood $V_{\boldsymbol{\theta}_0} \subset \boldsymbol{\Theta}$ of $\boldsymbol{\theta}_0$, $\sup_{(x, \boldsymbol{\theta}) \in [0, \tau_H] \times V_{\boldsymbol{\theta}_0}} \|\mathrm{Grad}(m(x, \boldsymbol{\theta})\| < \infty$.

Let $\hat{\boldsymbol{\theta}}^*$ denote the bootstrap MLE of $\boldsymbol{\theta}$. Proof of theorem 2 below is given in the Appendix.

Theorem 2 *For almost all sample sequences* $\{(X_i, \delta_i), 1 \leq i \leq n\}$, $\hat{\boldsymbol{\theta}}^* = \hat{\boldsymbol{\theta}} + o(1)$, $I\!P_n$ *a.s.*

Let $\mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote a k-variate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. For each $i = 1, \ldots, n$, write $w(X_i, \delta_i, \boldsymbol{\theta}) = \delta_i \log(m(X_i, \boldsymbol{\theta})) + (1 - \delta_i) \log(1 - m(X_i, \boldsymbol{\theta}))$. Subramanian and Zhang (2013) proved proposition 1 stated below. The implication is that, for almost all sample sequences $\{(X_i, \delta_i), 1 \leq i \leq n\}$, we have $(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}) = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_{I\!P_n}(n^{-1/2})$.

Proposition 1 *Suppose that* $m(x, \boldsymbol{\theta})$ *is correctly specified. For almost all sample sequences* $\{(X_i, \delta_i), 1 \leq i \leq n\}$, *then,* $n^{1/2}\left(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}\right)$ *is asymptotically* $\mathcal{N}_k(\boldsymbol{0}, I_0^{-1})$. *In particular,*

$$n^{1/2}\left(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}\right) = n^{-1/2} \sum_{i=1}^{n} I_0^{-1} \mathrm{Grad}(w(X_i^*, \delta_i^*, \hat{\boldsymbol{\theta}})) + o_{I\!P_n}(1). \tag{2.19}$$

A Taylor's expansion of $m(x, \hat{\boldsymbol{\theta}}^*)$ about $\hat{\boldsymbol{\theta}}$, together with condition **A**, yields

$$\|m(\cdot, \hat{\boldsymbol{\theta}}^*) - m(\cdot, \hat{\boldsymbol{\theta}})\| = o_{I\!P_n}(1). \tag{2.20}$$

Note that $\hat{H}^*(t) = \sum_{j=1}^{n} I(X_j^* \leq t)/n$ satisfies the Glivenko–Cantelli theorem: For all sample sequences $\{X_i, 1 \leq i \leq n\}$, $\|\hat{H}^* - \hat{H}\| = o(1)$ $I\!P_n$ a.s. Furthermore, theorem 1, Eq. (2.20), the strong law of large numbers, and lemma 3.6 of Dikta (1998) imply the following proposition:

Proposition 2 *Under condition* ***A****, the bootstrap versions of* $\hat{\Psi}(t)$ *and* $\hat{S}(t)$*, denoted by* $\hat{\Psi}^*(t)$ *and* $\hat{S}^*(t)$ *respectively, satisfy, for almost all sample sequences* $\{(X_i, \delta_i), 1 \leq i \leq n\}$,

$$\|\hat{\Psi}^* - \hat{\Psi}\| = o_{I\!P_n}(1), \qquad \|\hat{S}^* - \hat{S}\| = o_{I\!P_n}(1).$$

Write $\hat{\mathbb{Z}}^* = n^{1/2}(\hat{\Psi}^* - \hat{\Psi})$ and $\hat{\mathbb{W}}^* = n^{1/2}(\hat{S}^* - \hat{S})$, the bootstrap versions of $\hat{\mathbb{Z}} = n^{1/2}(\hat{\Psi} - \Psi_0)$ and $\hat{\mathbb{W}}$ respectively. Subramanian and Zhang (2013) proved proposition 3 below.

Proposition 3 *For almost all sample sequences* $\{(X_i, \delta_i), 1 \leq i \leq n\}$, $\hat{\mathbb{Z}}^*$ $(\hat{\mathbb{W}}^*)$ *has the same weak limit as* $\hat{\mathbb{Z}}$ $(\hat{\mathbb{W}})$. *Furthermore, for almost all sample sequences* $\{(X_i, \delta_i), 1 \leq i \leq n\}$, $\|\hat{\mathbb{Z}}^*\|$ $(\|\hat{\mathbb{W}}^*\|)$ *has the same limit distribution as that of* $\|\hat{\mathbb{Z}}\|$ $(\|\hat{\mathbb{W}}\|)$.

2.6 Semiparametric SCBs

The bootstrap semiparametric LR function $\hat{\mathcal{L}}^*(S(t), t) = -2 \log \mathcal{L}^*(S(t), t)$ is given by

$$\begin{aligned} \hat{\mathcal{L}}^*(S(t), t) = & \left[\sum_{s \leq t} \left(Y^*(s) - m(s, \hat{\boldsymbol{\theta}}^*) \Delta N^*(s) \right) \log \left(1 + \frac{\hat{\lambda}^*(t)}{Y^*(s) - m(s, \hat{\boldsymbol{\theta}}^*) \Delta N^*(s)} \right) \right. \\ & \left. - \sum_{s \leq t} Y^*(s) \log \left(1 + \frac{\hat{\lambda}^*(t)}{Y^*(s)} \right) \right], \end{aligned} \tag{2.21}$$

and $\hat{\lambda}^*(t) = \hat{\lambda}^*(S(t), t)$, the bootstrap Lagrange multiplier, is the unique solution of

$$\prod_{s \leq t} \left(1 - \frac{m(s, \hat{\boldsymbol{\theta}}^*) \Delta N^*(s)}{Y^*(s) + \lambda(t)} \right) = S(t), \tag{2.22}$$

on the interval $(D^*(t), \infty)$, where $D^*(t) = \max_{s \leq t}(m(s, \hat{\boldsymbol{\theta}}^*) \Delta N^*(s) - Y^*(s))$. This choice for $D^*(t)$ ensures that $0 \leq m(s, \hat{\boldsymbol{\theta}}^*) \Delta N^*(s)/(Y^*(s) + \lambda(t)) < 1$ for all $s \leq t$. The $\lambda(t)$ solving Eq. (2.22) with $S = \hat{S}$ will be denoted by $\hat{\lambda}_0^*(t) \equiv \hat{\lambda}^*(\hat{S}(t), t)$. Using theorem 1, it can be shown that $\|(\tilde{\sigma}^*(t))^2 - \tilde{\sigma}^2(t)\| = o_{I\!P_n}(1)$, where

$$(\tilde{\sigma}^*(t))^2 = n \sum_{s \leq t} \frac{m(s, \hat{\boldsymbol{\theta}}^*) \Delta N^*(s)}{Y^*(s)(Y^*(s) - m(s, \hat{\boldsymbol{\theta}}^*) \Delta N^*(s))} \tag{2.23}$$

defines the bootstrap version of $\tilde{\sigma}^2(t)$. We then have the following two-stage bootstrap analog of theorem 1. The proof is given in the Appendix.

Theorem 3 *For almost all sample sequences* $\{(X_i,\delta_i), 1\le i\le n\}$, *uniformly for* $t\in[t_1,t_2]\subset(0,\tau_H]$,

$$\hat{\mathcal{L}}^*\left(\hat{S}(t),t\right) \quad = \quad \frac{1}{\tilde{\sigma}^2(t)}\left\{n^{1/2}\left(\log\hat{S}^*(t)-\log\hat{S}(t)\right)\right\}^2+o_{\mathbb{P}_n}(1). \tag{2.24}$$

Eqs. (2.17) and (2.24), the continuous mapping theorem, and proposition 3 establish, for almost all sample sequences $\{(X_i,\delta_i), 1\le i\le n\}$, that $\hat{\mathcal{L}}^*(\hat{S}(t),t)$ has the same limit distribution as $\hat{\mathcal{L}}(S_0(t),t)$. The steps described in subsection 2.2, with appropriate modifications for the semiparametric scenario, can now be followed to obtain q_α, needed for computing the semiparametric SCBs. Note in particular that $\hat{\lambda}_0^*(t)$ and, in turn, $\hat{\mathcal{L}}^*(\hat{S}(t),t)$ are computed at *each* $X_i, i=1,\ldots,n$. The semiparametric LR SCBs for S_0 are given by

$$\mathcal{B}_{\mathrm{SP}} \quad = \quad \{S(t):\ \hat{\mathcal{L}}(S(t),t)\le\hat{\rho}(t), \qquad t\in[t_1,t_2]\subset(0,\tau_H]\}, \tag{2.25}$$

where $\hat{\rho}(t)=q_\alpha/\hat{w}(t)$, with $\hat{w}(t)$ taken as $\hat{\sigma}(t)/(1+\hat{\sigma}^2(t))$ or $\tilde{\sigma}(t)/(1+\tilde{\sigma}^2(t))$, giving the semiparametric SCBs "SRCM I" and "SRCM II" respectively, which are piecewise constant confidence limits over $[X_i,X_{i+1})\subset[t_1,t_2], i=1,\ldots,n$, akin to Eq. (2.7).

Remark 5 For each fixed t, note that the search for $\lambda(t)\equiv\lambda$ is over $(D(t),\infty)$, see Eq. (2.15). Writing $\hat{\mathcal{L}}'(\lambda,t)$ for the partial derivative of $\hat{\mathcal{L}}(S(t),t)\equiv\hat{\mathcal{L}}(\lambda,t)$ with respect to λ, this implies that the two denominator terms of the integrand in

$$\hat{\mathcal{L}}'(\lambda,t)=\lambda\int_0^t\frac{m(s,\hat{\boldsymbol{\theta}})}{(Y(s)-m(s,\hat{\boldsymbol{\theta}})\Delta N(s)+\lambda)(Y(s)+\lambda)}dN(s)$$

are positive; hence the sign of $\hat{\mathcal{L}}'(\lambda,t)$ is the same as that for λ. Therefore, $\hat{\mathcal{L}}(\lambda,t)$ is concave upward – strictly decreasing over $(D(t),0]$, zero at 0, and strictly increasing over $[0,\infty)$. Because of this there are two solutions for $\hat{\mathcal{L}}(\lambda,t)=\hat{\rho}(t)$, one negative and one positive. As in the nonparametric case, the graphs of $\hat{\mathcal{L}}(\lambda,t)$ as function of λ are inversely nested. While the linear SCBs provide monotone upper limits, the case of lower limits is less transparent. Lack of monotone limits can be addressed, however, via the adjustment already discussed in *Remark 4* for the nonparametric case.

3 Numerical studies

Through a simulation study, we first showcase the *validity* of the new SCBs for practical (moderate) sample sizes. We demonstrate that under correct parametric specification our methods indeed produce SCBs having correct empirical coverage probabilities (ECPs). More specifically, we report the proportion of 2,000 SCBs, computed from $k=2{,}000$ simulated data sets, that include $S_0(t)$ for all $t\in[t_1,t_2]$. Further comparisons between *correctly* performing SCBs then are necessary for determining the *best* performing method. For this purpose comparisons are based on plots of percentage relative reduction (PRR henceforth) in the *estimated average enclosed areas* (EAEAs) and the *estimated average widths* (EAWs), of the semiparametric SCBs over the nonparametric ones, as a function of the censoring rate (CR). Specifically, on the interval $[x_{m_1},x_{m_2}]\supset[t_1,t_2]$, where x_{m_1} and x_{m_2} are the largest and smallest distinct uncensored times smaller and larger than t_1 and t_2 respectively,

$$\mathrm{EAEA}=\frac{1}{k}\sum_{i=1}^{k}\left\{\sum_{j=m_1}^{m_2}l_j\,\Delta_{x_j}\right\}; \qquad \mathrm{EAW}=\frac{1}{k}\sum_{i=1}^{k}\left\{\sum_{j=m_1}^{m_2}l_j\Delta_{\hat{S}_j}\right\},$$

where l_j denotes the width of the band computed at point x_j, $\Delta_{x_j}=x_{j+1}-x_j$, and $\Delta_{\hat{S}_j}$ is the jump size of $\hat{S}$, the estimator of S, at x_j. The sums inside the braces change with i but our notation, to keep things simple, does not reflect it. Note that for the nonparametric case the inner sum is based on widths and jump sizes computed at each distinct uncensored point belonging to $[x_{m_1},x_{m_2}]$, with S_n replacing $\hat{S}$. After the validation study, we present the results of two sensitivity studies to investigate the operating characteristics of our proposed SCBs, where a misspecified (cauchit) model is always fitted for m. Finally, we provide a real example illustration.

The SCBs are computed over the entire length spanned by the data, with the caveat that we can only compute the SCBs where the data permit the computation of the Lagrange multiplier values, for determining the threshold as well as the lower and upper limits. Specifically, where possible, we compute ordered pairs (a_k, b_k), $k = 1, \ldots, n$, such that $P(a_1 \le S_0(X_1) \le b_1, \ldots, a_n \le S_0(X_n) \le b_n) \approx 1 - \alpha$. The range covered (almost) all the observed times for SRCMs and observed event (uncensored) times for RCM. The excepted regions occur near the left and right tails, typically the first and last uncensored data points in the case of RCM, where dearth of data leads to the breakdown of Brent's method (Press et al., 1992) of root finding. For the RCM, note that, because of the shape of the likelihood ratio function, see Figure 1, Brent's method is ideal but requires that the two roots (one each for the lower and upper limits) be bracketed within their respective cut-offs, specifically, $[\tilde{D}(t), 0)$ and $(0, \infty)$. At the problem areas, the function (likelihood ratio minus the threshold) is supposed to give values with opposite signs at the extremities $\tilde{D}(t)$ and 0, but typically fails to do so. This obduracy may be due to paucity of data near the misbehaving points. Likewise for SRCMs.

3.1 Validation study (simulation)

The failure time was Weibull, with $F(x) = 1 - S(x) = 1 - \exp(-(\theta x)^2)$. The censoring was exponential with mean 1. Then, $m(x, \theta) := P(\delta = 1 | X = x) = 2\theta^2 x/(1 + 2\theta^2 x)$. The CR, expressed as a function of the parameter θ, is given by

$$C(\theta) = \frac{\sqrt{\pi}}{\theta} exp\left(\frac{1}{4\theta^2}\right)\left\{1 - \Phi\left(\frac{1}{\theta\sqrt{2}}\right)\right\},$$

where $\Phi(x)$ denotes the standard normal distribution function. Simulations were run for 12 different censoring percentages between 19% ($\boldsymbol{\theta} = 4.0$) and 44% ($\boldsymbol{\theta} = 1.4$). For each simulation run, involving one value of θ, the ECPs of the proposed SCBs as well as the PRR in EAEA and EAW of the semiparametric (SRCM) over the nonparametric (RCM) SCBs were calculated. For each simulation run, the SCBs were based on sample size 100 and were calculated repeatedly 2,000 times; the critical value for computing each SCB was based on bootstrap sample size 100 and 1,500 bootstrap replications. We used $w_n = \sigma_n/(1 + \sigma_n^2)$ for the RCM-based SCBs, and $\hat{w} = \hat{\sigma}/(1 + \hat{\sigma}^2)$ and $\tilde{\sigma}/(1 + \tilde{\sigma}^2)$ for the two types of SRCM-based SCBs, referred as "SRCM I" and "SRCM II" respectively. The plots presented in Figure 2 indicate that both the RCM- and SRCM-based SCBs provide approximately correct ECPs, close to the nominal 0.95. The plots presented in Figure 3 and Figure 4 indicate that the PRR in EAEA and EAW of the SRCM-based SCBs over the RCM-based SCB amounted to between 4% (for 19% CR) and 10%-11% (for 44% CR).

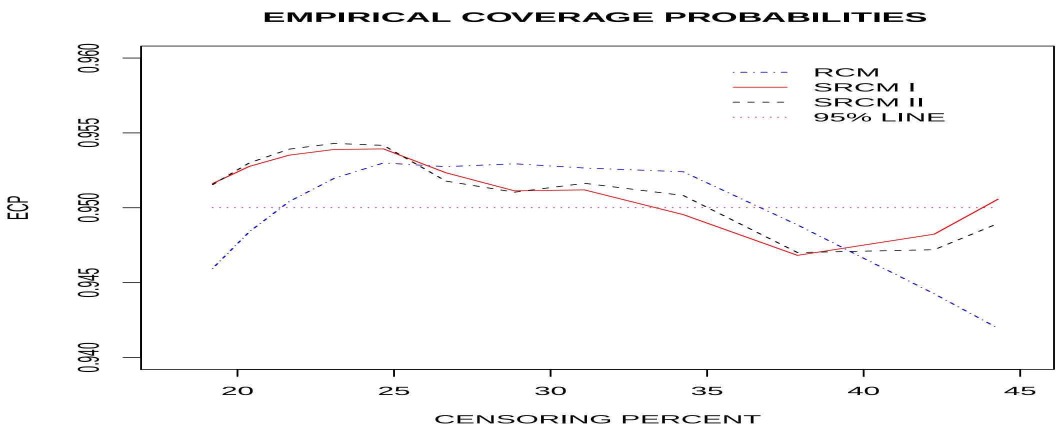

Figure 2: Validation study. Empirical coverage probabilities (ECPs) of the 95% SCBs for $S_0(t)$ over the region supported by the data are presented.

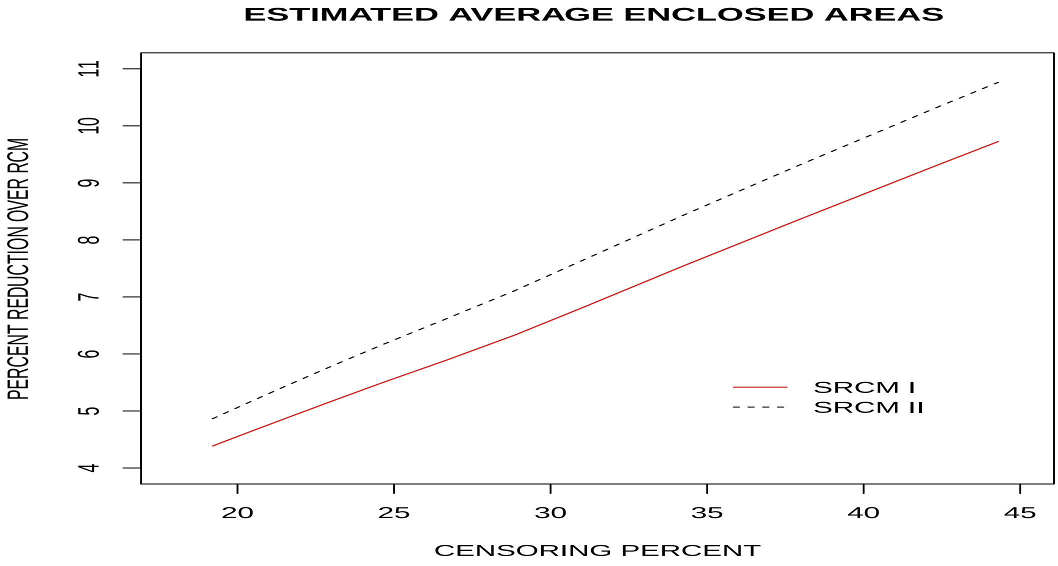

Figure 3: Validation study. The percent relative reduction in estimated average enclosed areas (EAEAs) of the two semiparametric SCBs over the (RCM) nonparametric SCB are presented.

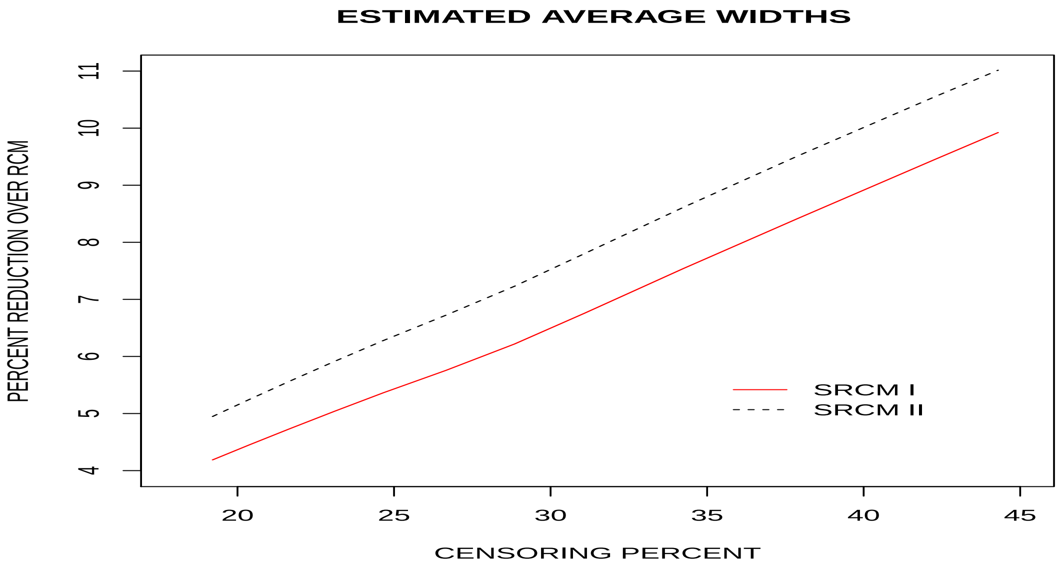

Figure 4: Validation study. The percent relative reduction in estimated average widths (EAWs) of the two semiparametric SCBs over the (RCM) nonparametric SCB are presented.

3.2 Sensitivity studies (simulations)

For the first study, we introduced misspecification of $m(x)$ by always fitting the Cauchy model

$$m^{\text{cauchit}}(x, \boldsymbol{\gamma}) = \frac{1}{\pi}\left\{0.5 + \arctan(\gamma_0 + \gamma_1 x)\right\}. \tag{3.26}$$

to the binary response data which, however, were generated using the model $m(x, \theta) = 2\theta^2 x/(1 + 2\theta^2 x)$, see subsection 3.1. The ECPs and the PRR in EAEA and EAW values, for each of twelve values of θ between 4 and 1.4, were calculated over the entire range of the observed data. The sample size was 100 and the number of replications was 2,000. The bootstrap sample size was 100 and the number of bootstrap replications was 1,500. Note that this study affects only the SRCMs-based SCBs; the RCM-based SCBs remain the same as in the validation study. The plots presented in Figure 5, indicate that the SRCM-based SCBs provided ECPs that were not significantly downgraded, but, in fact, were close to the nominal 95% level. Furthermore, the plots presented in Figure 6 and Figure 7 indicate that the PRR in EAEA and EAW of the SRCM-based SCBS over the RCM-based ones amounted to between 2% and 6%. The plots suggest increasing gains at higher CRs.

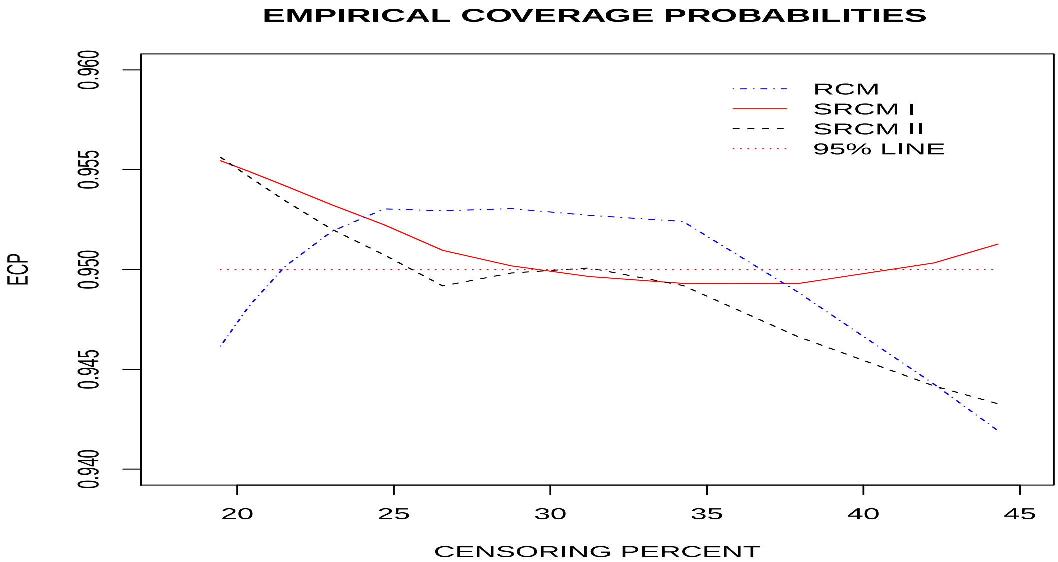

Figure 5: First misspecification study using the cauchit link. Empirical coverage probabilities (ECPs) of the 95% SCBs for $S_0(t)$ over the region supported by the data are presented.

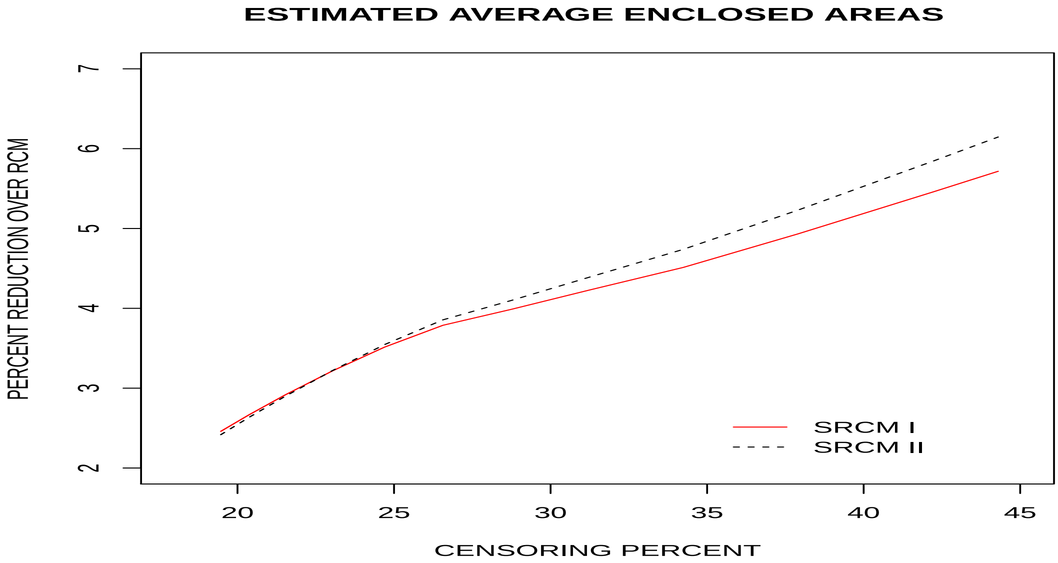

Figure 6: First misspecification study using the cauchit link. The percent relative reduction in estimated average enclosed areas (EAEAs) of the two semiparametric SCBs over the (RCM) nonparametric SCB are presented.

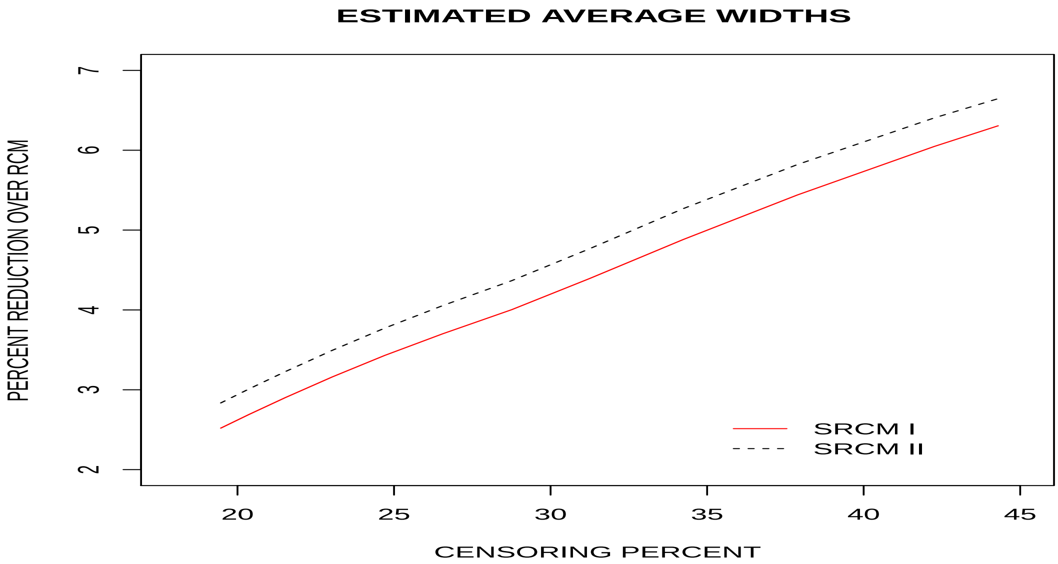

Figure 7: First misspecification study using the cauchit link. The percent relative reduction in estimated average widths (EAWs) of the two semiparametric SCBs over the (RCM) nonparametric SCB are presented.

For the second misspecification study, an alternate data-generating mechanism was followed, requiring the specification of the distribution of X and the model m. Accordingly, the minimum was taken as uniform over $(0, 1)$ and the true model for m was taken as $m(t, \boldsymbol{\theta}) = 1 - \exp(-\exp(\theta_1 + \theta_2 t))$, where θ_2 was fixed at -5.92

and θ_1 was varied over a grid of values between 4.4 and 2.85 to give various censoring percentages between 20 and 44. The true survival function is then given by

$$S_0(t) = (1-t)\exp\left(\int_0^t \frac{\exp(-\exp(\theta_1+\theta_2 y))}{1-y}dy\right).$$

For each of 13 simulation runs, the sample size was 100, number of replications 2000, and the bootstrap sample size and replications were 100 and 1500 respectively. We fitted the cauchit link for m, given by Eq. (3.26), which indicates a misspecified model. The ECPs and PRR in EAEA and EAW values of the SRCM-based SCBs over the RCM-based ones are now presented in Figure 8, Figure 9 and Figure 10. The ECPs of the SCBs computed via SRCMs are largely above 94.5%, and approach the nominal 95% for higher CRs. Furthermore, when compared with RCM-based SCBs, they provided a percent reduction in enclosed areas amounting to between 1.5% and 4.25% and a reduction in average widths amounting to between 5% and 10%.

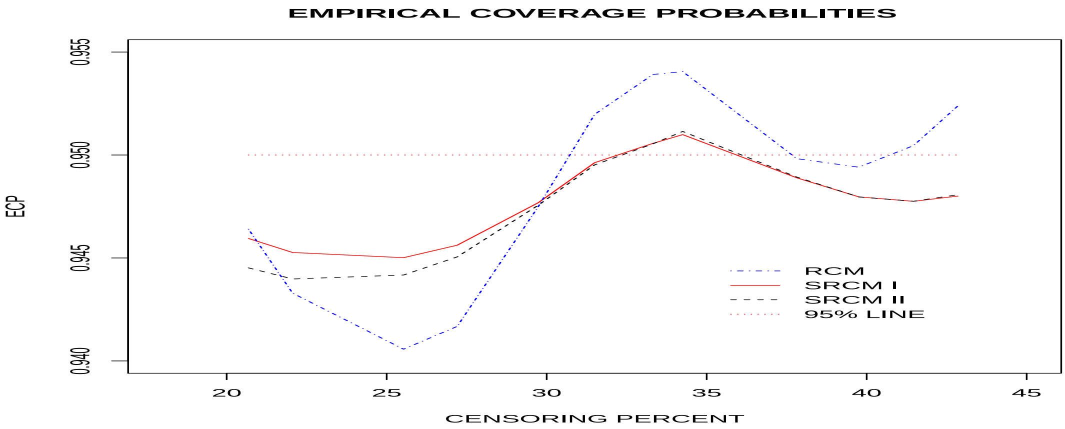

Figure 8: Second misspecification study using the cauchit link. Empirical coverage probabilities (ECPs) of the 95% SCBs for $S_0(t)$ over the region supported by the data are presented.

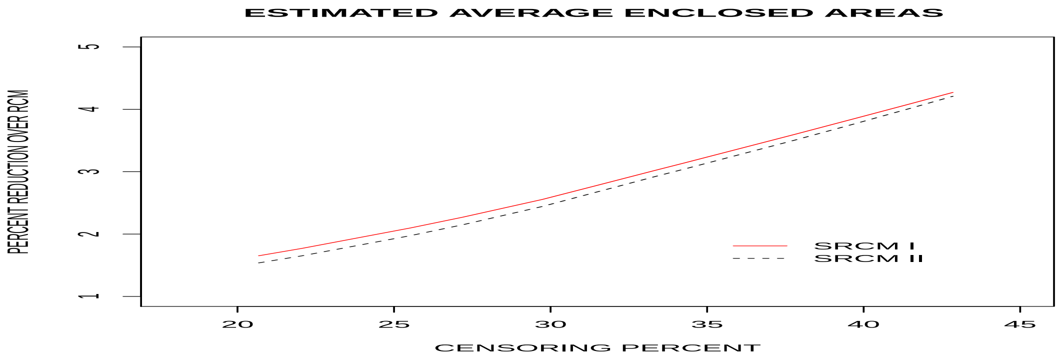

Figure 9: Second misspecification study using the cauchit link. The percent relative reduction in estimated average enclosed areas (EAEAs) of the two semiparametric SCBs over the (RCM) nonparametric SCB are presented.

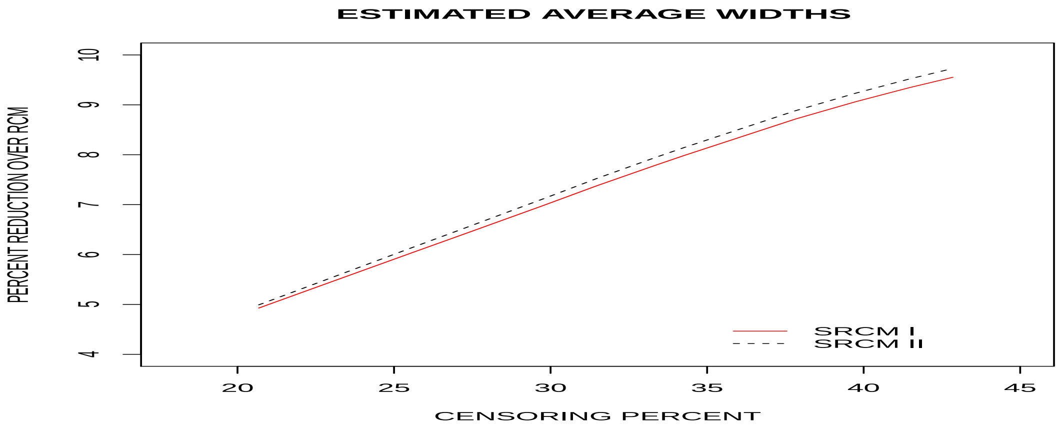

Figure 10: Second misspecification study using the cauchit link. The percent relative reduction in estimated average widths (EAWs) of the two semiparametric SCBs over the (RCM) nonparametric SCB are presented.

3.3 JASA time-to-first-review data analysis (real example)

HMY analyzed data giving the time-to-first-review of 432 papers submitted to the Theory and Methods Section of JASA between January 1, 1994 and December 13, 1994. Here X is the number of days between a manuscript's submission and its first review or the cut-off date, and $\delta = 1$ or 0 according as whether a paper received its first review by the cut-off date or not. There are 275 uncensored times and 157 censored times, giving 36% CR. We employed a model-based resampling procedure (Dikta, Kvesic, and Schmidt, 2006) to check the adequacy of three models for fitting the binary response data. They are the Cauchy model, see Eq. (3.26), and the complementary log-log, logistic, and probit models given by

$$m^{\text{Complementary log-log}}(x,\boldsymbol{\gamma}) = 1 - \exp\left(-\exp(\gamma_0 + \gamma_1 x)\right), \tag{3.27}$$

$$m^{\text{Logistic}}(x,\boldsymbol{\gamma}) = \frac{e^{\gamma_0+\gamma_1 x}}{1+e^{\gamma_0+\gamma_1 x}}, \tag{3.28}$$

$$m^{\text{Probit}}(x,\boldsymbol{\gamma}) = \Phi(\gamma_0 + \gamma_1 x), \tag{3.29}$$

where $\Phi(x)$ denotes the standard normal cumulative distribution function. We computed the Kolmogorov–Smirnov (KS) and Cramér-von Mises (CvM) test statistics from the observed data and then computed their (model-based) bootstrap counterparts 5,000 times. The KS and CvM statistics are based on Stute's (1997) marked empirical process adapted to binary regression (Dikta et al., 2006). The proportion of the 2,000 bootstrap values that exceeded the test statistics gives the p-values of the tests. There is some indication that all the four models may be adequate, see Table 1 below, where the p-values are presented.

Table 1: Observed significance level of KS and CvM tests for checking model adequacy. Test statistics computed from observed data are placed within parenthesis.

Test	*Cauchy*	*Complementary log-log*	*Logistic*	*Probit*
KS	0.1280 (0.3815)	0.1843 (0.3775)	0.1450 (0.3791)	0.1500 (0.3787)
CvM	0.1630 (0.0251)	0.2080 (0.0249)	0.1803 (0.0250)	0.1760 (0.0250)

The RCM- and SRCM-based SCBs are presented in Figure 11, along with the log-transformed equal precision (EP) and Hall–Wellner (HW) bands. Where required, the SCBs were adjusted near the tails to satisfy monotonicity, as discussed in Remark 4. The EP bands perform best near the tails, while the performance of the HW bands is considerably worse in those regions. The two SRCM-based SCBs were developed using a

Cauchy fitted m and performed better than the other competing SCBs over a greater region, specifically, the region represented between the 25th and 80th percentiles of $1 - \hat{S}(t)$, the SRCM-based estimator of $1 - S_0(t)$.

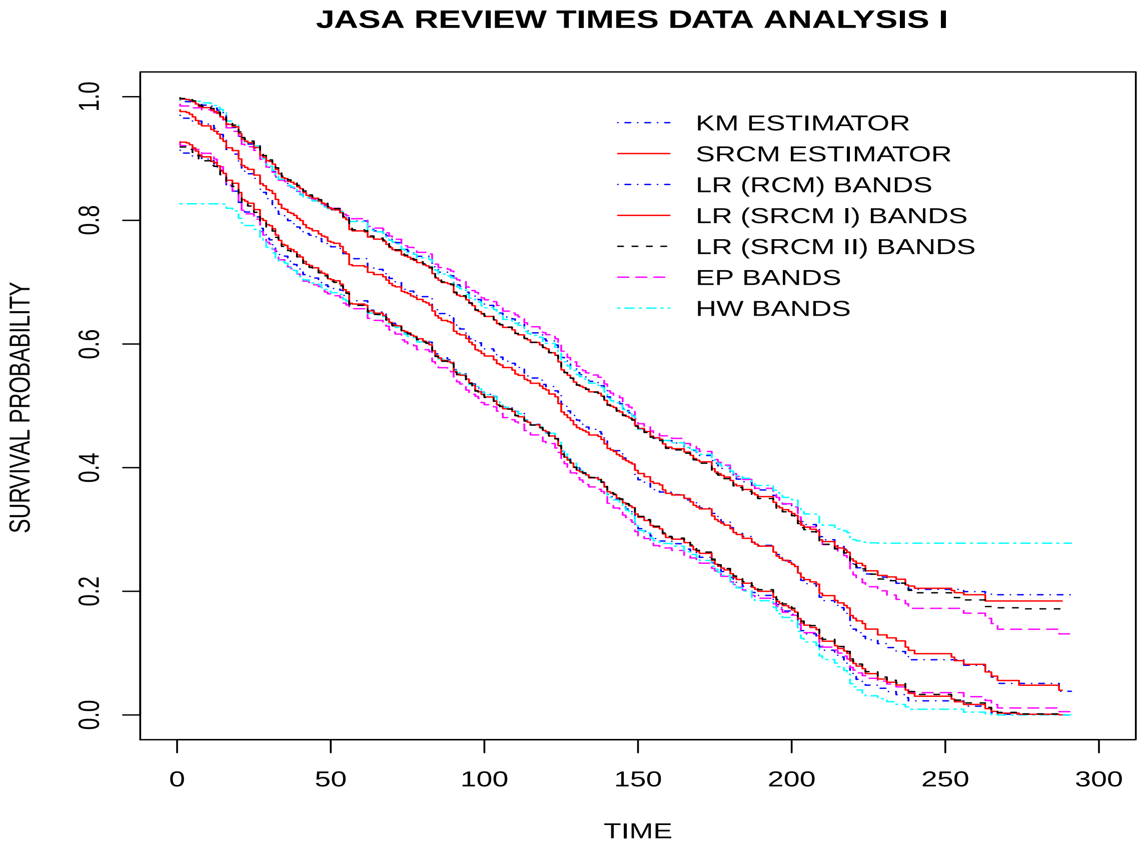

Figure 11: Confidence bands for $S_0(t)$. The cauchit link used for the semiparametric bands.

Since the HW bands performed relatively worse than the other competing SCBs, in Table 2 we present the PRR in EAEA and EAW of the proposed SCBs over Nair's EP bands. The SRCM I and SRCM II SCBS provided a PRR exceeding 8% over the EP bands.

Table 2: Percent relative reduction in EAEA and EAW of SRCM I, II, and RCM bands over EP bands.

Measure	*RCM*	*SRCM I*	*SRCM II*
EAEA	−1.70%	9.69%	8.33%
EAW	3.33%	12.38%	12.01%

We also developed all the SCBs separately over the region represented between approximately the 25th and 80th percentiles of $1 - \hat{S}(t)$. The plots, presented in Figure 12, indicates that the SRCM I and SRCM II bands performed better overall. From Table 3, they provided PRR in EAEA and EAW values exceeding 10% over the EP bands.

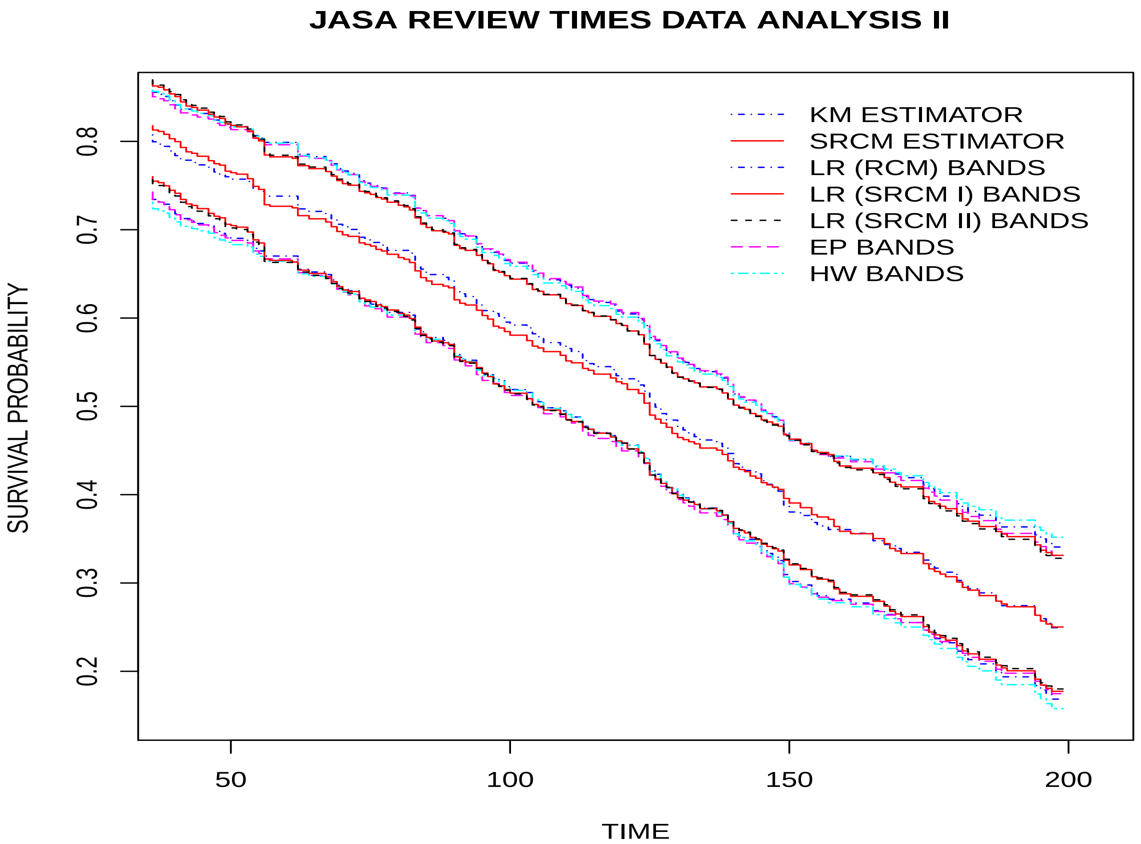

Figure 12: Restricted range SCBs for $S_0(t)$ over the region between the 80th and 25th percentiles of the SRCM survival function estimator. The cauchit link used for the semiparametric bands.

Table 3: Percent relative reduction in EAEA and EAW of SRCM I, II, RCM, and HW bands over EP bands.

Measure	*RCM*	*SRCM I*	*SRCM II*	*HW*
EAEA	1.00%	10.99%	11.14%	−1.43%
EAW	1.39%	10.23%	10.50%	−0.63%

4 Concluding discussion and extensions

The bootstrap procedure, implemented in this paper separately for the nonparametric and semiparametric scenarios, facilitates performing comparisons between the two approaches on an "equal footing". It eliminates the need to adjust for two different methods, which, otherwise, may be necessary to ensure a proper comparison study. The model-based LR method offers a new research direction that has remained unexplored so far and, in this paper, has been shown to provide an appealing alternative for constructing SCBs for survival functions from one-sample randomly censored data. The notion that improved SCBs would arise as "neighborhoods" of a semiparametric-efficient survival function estimator is an important insight recognized and pursued in this paper. Toward this end, the proposed semiparametric method aims to exploit the power of available good-fitting models and good model-fitting procedures for binary response data to produce more informative semiparametric SCBs. To examine the validity of the semiparametric method, we carried out a simulation study incorporating correct parametric specification, which successfully yielded SCBs having the desired coverage. To showcase its applicability for practical data analysis, we carried out two sensitivity studies in each of which we fitted the cauchit model to the binary response data that were, in reality, generated from a different model, and found that the proposed SCBs still provided adequate coverage, close to the nominal level. The semiparametric SCBs were found to be always tighter than the proposed nonparametric SCBs, in terms of producing smaller average enclosed areas and widths, with considerable reduction evidenced at higher CRs. This should not be surprising, since, when the censoring is rather heavy, the KM estimator, having support only at uncensored points, may exhibit flat regions due to which it can contribute toward less informative nonparametric SCBs. Since the semiparametric survival function estimator has support at all observed time points (censored and uncensored), SRCMs have good potential to avoid such a limitation when there is heavy censoring. Our implementation of both types of LR-based SCBs provides a user with the option and flexibility to choose the one that would best fit the user's requirements.

It is important to note that the semiparametric approach, although it introduces informative censoring, is different from direct (parametric) modelling of the censoring distribution. Indeed, since the distribution of X is unspecified, a *semiparametric* model is indicated for the censoring distribution, so there is perhaps less danger of flirting with misspecification than when a (parametric) model is used for the true censoring distribution. Interestingly, contrary to popular perception, informative censoring can manifest even when the event time distribution and the censoring time distribution do not have shared parameters. Specifically, when the failure and censoring times are each Weibull distributed, $m(x)$ is the generalized proportional hazards, see Example 2.9 of Dikta (1998).

Evidently, the success of the model-based approach is predicated on the user being able to supply a good-fitting model for the conditional non-censoring probability given the observed time. For the analysis of binary response data, however, there are a number of models to choose from, such as logistic, complementary log-log, generalized proportional hazards, probit, cauchit, among others. As indicated by Koenker and Yoon (2009), who provide some comments on the cauchit model, both the probit and cauchit are nested within the family of Gosset links. The search for the right model can be narrowed down by discarding ill-fitting models using model-checking methods. As evidenced by other simulation studies as well, the cauchit link often yields good estimates (Subramanian 2012; Mondal and Subramanian, 2014).

When missing censoring indicators (MCIs) an RCM analysis is problematic. The nonparametric LR is inapplicable since the right side of Eq. (2.1) and the left side of Eq. (2.2) are not computable. Nonparametric adjustments are perhaps cumbersome and hence might be unsatisfactory, with the need to supply conditional function estimates and attendant optimal bandwidths. SRCMs pose no such limitations, since model parameters can be estimated consistently from the complete cases (Subramanian, 2004), allowing as before the computation of the right side of Eq. (2.14) and the left side of Eq. (2.15).

Both the nonparametric and the semiparametric approaches can produce SCBs with non-monotone limits near the tails. However, as noted in a remark in section 2.2, tighter monotone limits can be provided by a simple modification. This was incorporated in both the simulations and data analysis.

Matthews (2013) introduced a censored-data analog of the Owen–Jager–Wellner method (Owen, 1995; Jager and Wellner, 2005) and proposed nonparametric SCBs for $S_0(t)$. In contrast to inversion of the scaled log LR, Matthews (2013) inverted a modified Berk–Jones statistic (Berk and Jones, 1979), representing a "binomial measure of discrepancy" between $S_n(t)$ and $S_0(t)$. Unlike the unbiased empirical distribution function associated with the original Berk–Jones statistic, however, the KM estimator is biased for finite sample sizes, hence the effect of the bias on the SCBs needs to be investigated. Furthermore, his sampling scheme requires n fully observed event times and would be inapplicable when some of them are missing at random. The random sample size implied by his data generating mechanism appears to be different from the standard RCM setup. Large-sample SCBs based on his modified Berk–Jones statistic provide a worthwhile direction for future research.

"Restricted-range" SCBs seem important enough for Matthews (2013) to indicate that "In some study settings, investigators may only wish to obtain simultaneous confidence bands for a restricted interval on the time scale.". Aside from the fact that SCBs plotted over a restricted range provide magnified viewing to help determine visually the best performing SCBs, one can envision scenarios where restricted-range SCBs would in fact be desirable. Such restricted-range SCBs may perhaps provide a more reliable conclusion for a hypothesis in question than “full-range” SCBs – simply because the latter were based on larger thresholds and, therefore, were too wide and lacked the “punch” to provide any meaningful conclusion over and above what was known already. For example, when comparing allogenic bone marrow transplant vs conventional chemotherapy, where difference is one key measure, positive difference between survival rates typically occurs after some years (Zhang and Klein, 2001). SCBs for the entire range may provide no new vital information near the left tail other than reinforcing the “status quo” that the difference of survival rates is negative. Furthermore, with such full-range SCBs, masking can be a possibility. Specifically, SCBs constructed over the entire range may lack the power to zoom-in on the time period beyond which the transplant actually becomes beneficial. In fact, the unrestricted SCBs may provide a conservative estimate and indicate zero or negative difference over a larger range, which could quite conceivably affect decision making regarding the type of treatment to be given to a patient. As seen from Figure 12, our approach can also provide SCBs over any desired data range quite easily. The user will only need to specify the window over which the SCBs are desired and the computation of the bootstrap thresholds for inversion will be automatic over the specified window.

The argument for the proposed semiparametric LR inference should enjoy stronger sustenance when implemented for the two-sample and, more generally, k-sample setting. For the two-sample case, the difference of survival functions, among others, is a key parameter. For example, the efficacy of a treatment (e.g., low dose) over control (standard dose) can be gauged using the difference of the survival functions. The simple and oft-used approach of displaying areas around point estimates of the difference using PCIs can lead to poor judgement regarding treatment efficacy. Indeed, since PCIs do not capture the global variability that SCBs do, an artifact of this limitation is that they typically indicate presence of treatment difference over a broader time range, potentially allowing an investigator to conclude incorrectly that the treatment is better (or inferior) over a larger time range than may really be the case. McKeague and Zhao (2005) developed LR-based SCBs for the difference, combining nonparametric LR with a plug-in KM estimator. They employed the Gaussian multiplier bootstrap to obtain the critical values required for SCB construction. The semiparametric approach for this problem would entail combining our proposed adjusted LR with the SRCM estimator of $S_0(t)$ as a plug-in. It is expected that the semiparametric SCBs for the difference, using thresholds computed by applying (1) the Gaussian multiplier bootstrap and (2) the two-stage bootstrap, should perform better. The model-based SCBs would more accurately depict the range of positive or negative treatment difference, facilitating improved decision making regarding type of treatment. Research on this front is currently in progress.

The proposed methods can be implemented for comparing two or more distributions via the corresponding quantile functions as well, see Einmahl and McKeague (1999), who developed Q-Q plot methods for two-sample comparisons and, more generally, simultaneous confidence tubes (SCTs) for multiple quantile plots. Q-Q plot methods and SCTs for multiple quantile plots using SRCMs would be a topic for future research.

We also envision the proposed methods as providing promise to reinforce and even modify, in borderline cases, past conclusions in cancer studies. A case in point is the analysis of AIDS clinical trial data by Parzen, Wei, and Ying (1997), who concluded on the basis of their Wald-type SCBs that a reduced dose of AZT was at least as good as the standard one with respect to a patient's survival. The zero line of no difference is barely included within their SCB when the observation times are between 575 and 650 days, however, see figure 2 of Parzen et al. (1997). With the proposed semiparametric approach yielding tighter SCBs, a second, and perhaps conclusive, analysis using the proposed methods would be certainly useful, since, based on Parzen et al.'s (1997) study, the low dose has “become the standard AZT mono-therapy for treating AIDS patients”.

Finally, the recent monograph of Claeskens and Hjort (2008) appears relevant for our problem. Indeed, uncertainties involved in parametric specifications engender inference following model selection. More specifically, model averaging methods can provide better assessment of confidence intervals, and, by extension, SCBs. This important topic is not pursued here, but would be a worthwhile direction for future research.

Appendix

A.1 Proof of theorem 1

We will need two lemmas. The first lemma establishes a strong uniform convergence rate for the Lagrange multiplier $\hat{\lambda}_0(t)$ and the second lemma derives an asymptotic representation for $\hat{\lambda}_0(t)$. Lemma 1 needs to be proved in advance of any application of Taylor series expansions needed in lemma 2 and Eq. (2.17). Recall that $\hat{\mathbb{Z}} = n^{1/2}(\hat{\Psi} - \Psi_0)$ and that $\hat{\mathbb{W}} = n^{1/2}(\hat{S} - S_0)$. Note that $\|\hat{\mathbb{Z}}\| = o((\log n)^{1/2})$ a.s. and $\|\hat{\mathbb{W}}\| = o((\log n)^{1/2})$ a.s., see, for example, Appendix A of Subramanian and Bean (2008). We also recall that $\|h\|_{t_1}^{t_2} \equiv \sup_{t\in[t_1,t_2]} |h(t)|$ denotes the sup-norm of any function h over $[t_1, t_2]$, and that $\|h\|$ denotes the sup-norm of h over $[0, \tau_H]$.

Lemma 1 *Suppose that S_0 is continuous and, for $[t_1, t_2] \subset (0, \tau_H]$, $\Psi_0(t_1) > \delta$ for some $\delta > 0$. When $m(x, \boldsymbol{\theta})$ is correctly specified, then, almost surely,*

$$\|\hat{\lambda}_0\|_{t_1}^{t_2} := \sup_{t\in[t_1,t_2]} |\hat{\lambda}_0(t)| = o\left((n \log n)^{1/2}\right).$$

Proof When $\hat{\lambda}_0(t) < 0$ for $t \in [t_1, t_2]$, we follow the proof of lemma 2.2 of Li (1995) to obtain the inequality $-\log S_0(t) \equiv \Psi_0(t) \geq \hat{\Psi}(t)$ and [cf. Eq. (2.12) of Li (1995)]

$$|\hat{\lambda}_0(t)| \quad \leq \quad \frac{n\left\{-\log S_0(t) - \hat{\Psi}(t)\right\}}{-\log S_0(t)} \quad = \quad \frac{n\left\{\Psi_0(t) - \hat{\Psi}(t)\right\}}{\Psi_0(t)} \quad \leq \quad \frac{n^{1/2}\|\hat{\mathbb{Z}}\|_{t_1}^{t_2}}{\Psi_0(t_1)}.$$

Hence the lemma is proved when $\hat{\lambda}_0(t) < 0$ for $t \in [t_1, t_2]$.

When $\hat{\lambda}_0(t) > 0$, follow the steps leading to Eq. (2.13) on page 101 of Li (1995) to obtain

$$\hat{\Psi}(t) + \log \hat{S}(t) - \log S_0(t) \quad \leq \quad \hat{\Psi}(t)\left(\frac{n}{n + |\hat{\lambda}_0(t)|}\right), \tag{A.1}$$

and in turn conclude that $\log S_0(t) \geq \log \hat{S}(t)$. Furthermore, the strong uniform consistency of $\hat{\Psi}(t)$ and $\hat{S}(t)$ over $[0, \tau_H]$ implies that, given $\epsilon < \Psi_0(t_1)$, we can find n so large that

$$\left\|\left(\hat{\Psi}(t) - \Psi_0(t)\right) + \left(\log \hat{S}(t) - \log S_0(t)\right)\right\| < \epsilon.$$

We then have, uniformly for $t \in [t_1, t_2] \subset (0, \tau_H]$,

$$\begin{aligned}\hat{\Psi}(t) + \log \hat{S}(t) - \log S_0(t) \quad &= \quad \Psi_0(t) + \left(\hat{\Psi}(t) - \Psi_0(t)\right) + \left(\log \hat{S}(t) - \log S_0(t)\right) \\ &> \quad \Psi_0(t_1) - \epsilon \ > \ 0,\end{aligned}$$

allowing us to conclude from Eq. (A.1) that for large enough n

$$|\hat{\lambda}_0(t)| \quad \leq \quad \frac{n\left\{\log S_0(t) - \log \hat{S}(t)\right\}}{\hat{\Psi}(t) + \log \hat{S}(t) - \log S_0(t)} \quad \leq \quad \frac{n\|\log \hat{S} - \log S_0\|_{t_1}^{t_2}}{\Psi_0(t_1) - \epsilon}.$$

Now apply Taylor's expansion of $\log \hat{S}(t)$ about $S_0(t)$ to complete the proof. □

Lemma 2 *Under the conditions of lemma 1, uniformly for $t \in [t_1, t_2] \subset (0, \tau_H]$,*

$$\hat{\lambda}_0(t) \quad = \quad -\frac{n}{\tilde{\sigma}^2(t)}\left(\log \hat{S}(t) - \log S_0(t)\right) + o_{\mathrm{p}}(\log n). \tag{A.2}$$

Proof Recall from Eq. (2.15) that $\hat{\lambda}_0(t) \equiv \hat{\lambda}(S_0(t), t)$ satisfies $f(\hat{\lambda}_0(t)) = \log S_0(t)$, where

$$f(\lambda) := \sum_{s \le t} \log\left(1 - \frac{m(s, \hat{\boldsymbol{\theta}})\Delta N(s)}{Y(s) + \lambda}\right).$$

Then $f(0) = \log \hat{S}(t)$ and

$$\begin{aligned} f'(\lambda) &= \sum_{s \le t} \frac{m(s, \hat{\boldsymbol{\theta}})\Delta N(s)}{(Y(s) + \lambda)(Y(s) + \lambda - m(s, \hat{\boldsymbol{\theta}})\Delta N(s))}, \\ f''(\lambda) &= -\sum_{s \le t} \frac{m(s, \hat{\boldsymbol{\theta}})\Delta N(s)\left\{2(Y(s) + \lambda) - m(s, \hat{\boldsymbol{\theta}})\Delta N(s)\right\}}{\left(Y(s) + \lambda - m(s, \hat{\boldsymbol{\theta}})\Delta N(s)\right)^2 (Y(s) + \lambda)^2}, \end{aligned} \tag{A.3}$$

so that $f'(0) = \tilde{\sigma}^2(t)/n$, see Eq. (2.16). Taylor's expansion of $f(\hat{\lambda}_0(t))$ about 0 yields

$$\log(S_0(t)) \equiv f(\hat{\lambda}_0(t)) = f(0) + \hat{\lambda}_0(t) f'(0) + \hat{\lambda}_0^2(t) f''(\eta(t))/2,$$

where $|\eta(t)| \le |\hat{\lambda}_0(t)|$, from which we arrive at the equation

$$\hat{\lambda}_0(t) = -\frac{n}{\tilde{\sigma}^2(t)}\left\{\left(\log \hat{S}(t) - \log S_0(t)\right) + \frac{1}{2}\hat{\lambda}_0^2(t) f''(\eta(t))\right\}.$$

Eq. (A.2) follows from lemma 1 provided that $f''(\eta(t)) = O_{\mathrm{p}}\left(n^{-2}\right)$, uniformly for $t \in [t_1, t_2]$. To obtain an upper bound for $f''(\eta(t))$, we show, for all $s \le t \le \tau_H$, that

$$Y(s) + \eta(t) - m(s, \hat{\boldsymbol{\theta}})\Delta N(s) > 0, \tag{A.4}$$

and, hence, $Y(s) + \eta(t) > 0$, which is critical for the denominator quantities of $f''(\eta(t))$, see Eq. (A.3). First note from Eq. (2.15) and the definition of $\hat{\lambda}_0(t)$, that, for all $s \le t \le \tau_H$,

$$Y(s) + \hat{\lambda}_0(t) - m(s, \hat{\boldsymbol{\theta}})\Delta N(s) > 0. \tag{A.5}$$

When $\hat{\lambda}_0(t) < 0$, then $0 > \eta(t) > \hat{\lambda}_0(t)$, and, hence, inequality (A.4) follows from (A.5). When $\hat{\lambda}_0(t) > 0$, then $0 < \eta(t) < \hat{\lambda}_0(t)$ and hence (A.4) again follows from (A.5) and the inequality

$$Y(s) - m(s, \hat{\boldsymbol{\theta}})\Delta N(s) \ge Y(s) - \Delta N(s) > 0, \qquad \text{for all } s \le t \le \tau_H. \tag{A.6}$$

We next obtain uniform lower bounds for the denominator quantities in Eq. (A.3). Since $1 - H(\tau_H) > 0$, choose $0 < \epsilon < 1 - H(\tau_H)$ such that, for large enough n,

$$\frac{1}{n}\left[Y(\tau_H) - \Delta N(\tau_H)\right] > \epsilon$$

and, hence, $Y(\tau_H)/n > \epsilon$ for large enough n. By lemma 1, for n large enough, $-\epsilon < \eta(t)/n < \epsilon$ for all $t \in [t_1, t_2]$. Note that the following hold for all $s \in [0, \tau_H]$:

$$Y(t) - \Delta N(t) \ge Y(s) - \Delta N(s), \qquad s > t. \tag{A.7}$$

Then, for large n,

$$Y(s) - m(s, \hat{\boldsymbol{\theta}})\Delta N(s) + \eta(t)) > (Y(s) - \Delta N(s)) - n\epsilon \ge (Y(\tau_H) - \Delta N(\tau_H)) - n\epsilon > 0.$$

Since $\sum_{s \le t} m(s, \hat{\boldsymbol{\theta}})\Delta N(s) \le \sum_{s \le t} \Delta N(s) \le n$, it follows from lemma 1 that

$$|f''(\eta(t))| \le \frac{n(2(n + \sup_{t \in [t_1, t_2]} |\eta(t)|) + n)}{\{(Y(\tau_H) - \Delta N(\tau_H)) - n\epsilon\}^2 (Y(\tau_H) - n\epsilon)^2} = O_{\mathrm{p}}\left(\frac{1}{n^2}\right). \quad \square$$

Proof of Eq. (2.17) Apply $\log(1+x) = x - x^2/2 + x^3/3 - x^4/4 + O(x^5)$ as $x \to 0$ to get

$$\begin{aligned}
-2\log\mathcal{L}\left(S_0(t),t\right) &= \hat{\lambda}_0^2(t)\sum_{s\le t}\left\{\frac{1}{Y(s)-m(s,\hat{\boldsymbol{\theta}})\Delta N(s)} - \frac{1}{Y(s)}\right\} \\
&\quad - \frac{2}{3}\hat{\lambda}_0^3(t)\sum_{s\le t}\left\{\left(\frac{1}{Y(s)-m(s,\hat{\boldsymbol{\theta}})\Delta N(s)}\right)^2 - \left(\frac{1}{Y(s)}\right)^2\right\} \\
&\quad + \frac{1}{2}\hat{\lambda}_0^4(t)\sum_{s\le t}\left\{\left(\frac{1}{Y(s)-m(s,\hat{\boldsymbol{\theta}})\Delta N(s)}\right)^3 - \left(\frac{1}{Y(s)}\right)^3\right\} + O_{\mathrm{p}}(n^{-1/2}) \\
&\equiv I_1(t) + I_2(t) + I_3(t) + o_{\mathrm{p}}(1).
\end{aligned}$$

By lemma 2 and using Eq. (2.16), uniformly for $t \in [t_1, t_2]$,

$$I_1(t) = \hat{\lambda}_0^2(t)\frac{\tilde{\sigma}^2(t)}{n} = \frac{\tilde{\sigma}^2(t)}{(\tilde{\sigma}^2(t))^2}\left\{n^{1/2}\left(\log\hat{S}(t) - \log S_0(t)\right) + o_{\mathrm{p}}\left(\frac{\log n}{n^{1/2}}\right)\right\}^2.$$

Applying Eqs. (A.7) and (A.6), and then lemma 1, it follows that $\|I_2\|_{t_1}^{t_2}$ is bounded above by

$$\begin{aligned}
&\frac{2}{3}\left(\|\hat{\lambda}_0\|_{t_1}^{t_2}\right)^3\left(\frac{1}{Y(t_2)-\Delta N(t_2)} + \frac{1}{Y(t_2)}\right)\sum_{t\le t_2}\left(\frac{1}{Y(t)-m(t,\hat{\boldsymbol{\theta}})\Delta N(t)} - \frac{1}{Y(t)}\right) \\
=\; & o_{\mathrm{p}}\left((n\log n)^{3/2}\right)O_{\mathrm{p}}\left(n^{-1}\right)\tilde{\sigma}^2(t_2)/n.
\end{aligned}$$

Since $\tilde{\sigma}^2(t_2) = O_{\mathrm{p}}(1)$, it follows that $\|I_2\|_{t_1}^{t_2} = o_{\mathrm{p}}(1)$. Likewise, $\|I_3\|_{t_1}^{t_2}$ is bounded above by

$$\begin{aligned}
&o_{\mathrm{p}}\left((n\log n)^2\right)\left[\frac{1}{(Y(t_2)-\Delta N(t_2))^2} + \frac{1}{Y(t_2)(Y(t_2)-\Delta N(t_2))} + \frac{1}{Y^2(t_2)}\right]\frac{\tilde{\sigma}^2(t_2)}{n} \\
=\; & o_{\mathrm{p}}\left((n\log n)^2\right)O_{\mathrm{p}}\left(n^{-2}\right)O_{\mathrm{p}}\left(n^{-1}\right) = o_{\mathrm{p}}(1).
\end{aligned}$$

This completes the proof of theorem 1. □

A.2 Proof of theorem 2

For $\eta > 0$, let $V_\eta(\boldsymbol{\theta})$ denote a η-neighborhood of $\boldsymbol{\theta} \in \bar{\boldsymbol{\Theta}}$, the closure of $\boldsymbol{\Theta}$. Let $w_1(x,\boldsymbol{\theta}) = \log m(x,\boldsymbol{\theta})$ and $w_2(x,\boldsymbol{\theta}) = \log(1-m(x,\boldsymbol{\theta}))$. Recall that the normalized log-likelihood is given by

$$l_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n w(X_i,\delta_i,\boldsymbol{\theta}) \equiv \frac{1}{n}\sum_{i=1}^n\left[\delta_i w_1(X_i,\boldsymbol{\theta}) + (1-\delta_i)w_2(X_i,\boldsymbol{\theta})\right].$$

By the strong law of large numbers (SLLN), we have $l_n(\boldsymbol{\theta}) \xrightarrow{\text{a.s.}} L_H(\boldsymbol{\theta}_0,\boldsymbol{\theta})$, where

$$L_H(\boldsymbol{\theta}_0,\boldsymbol{\theta}) = \int_0^\infty\left[m(s,\boldsymbol{\theta}_0)w_1(s,\boldsymbol{\theta}) + (1-m(s,\boldsymbol{\theta}_0))w_2(s,\boldsymbol{\theta})\right]dH(s).$$

The function $L_H(\boldsymbol{\theta}_0,\cdot)$ has a unique maximum at $\boldsymbol{\theta}_0$, see lemma 2.6 of Stute (1992), and $l_n(\boldsymbol{\theta}_0) \xrightarrow{\text{a.s.}} L_H(\boldsymbol{\theta}_0,\boldsymbol{\theta}_0)$. Recall that the bootstrap log-likelihood function is given by

$$l_n^*(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n w(X_i^*,\delta_i^*,\boldsymbol{\theta}) \equiv \frac{1}{n}\sum_{i=1}^n\left[\delta_i^* w_1(X_i^*,\boldsymbol{\theta}) + (1-\delta_i^*)w_2(X_i^*,\boldsymbol{\theta})\right].$$

We show for large n, for almost all sample sequences $\{(X_i, \delta_i), 1 \leq i \leq n\}$, that $l_n^*(\hat{\boldsymbol{\theta}})$ is arbitrarily close to $L_H(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0), I\!P_n$ a.s. Applying iterated expectation by conditioning on X^*,

$$
\begin{aligned}
&I\!E_n\left(\delta^* w_1(X^*, \hat{\boldsymbol{\theta}}) + (1-\delta^*) w_2(X^*, \hat{\boldsymbol{\theta}})\right) \\
&\qquad = \frac{1}{n}\sum_{i=1}^{n}\left[m(X_i, \hat{\boldsymbol{\theta}}) w_1(X_i, \hat{\boldsymbol{\theta}}) + (1 - m(X_i, \hat{\boldsymbol{\theta}})) w_2(X_i, \hat{\boldsymbol{\theta}})\right]. \qquad \text{(A.8)}
\end{aligned}
$$

Since $\hat{\boldsymbol{\theta}} \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0$, we can apply a contiguity argument as in the proof of lemma A.1 of Dikta et al. (2006) and the SLLN to conclude that the right side of Eq. (A.8) is arbitrarily close to $L_H(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0)$ a.s.; that is, given $\epsilon > 0$, we can find $n_0 \equiv n_0(\epsilon)$ such that, for all $n \geq n_0(\epsilon)$,

$$
\left|\frac{1}{n}\sum_{i=1}^{n}\left[m(X_i, \hat{\boldsymbol{\theta}}) w_1(X_i, \hat{\boldsymbol{\theta}}) + (1 - m(X_i, \hat{\boldsymbol{\theta}})) w_2(X_i, \hat{\boldsymbol{\theta}})\right] - L_H(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0)\right| \leq \epsilon. \qquad \text{(A.9)}
$$

By invoking the SLLN again, we have, for almost all sample sequences $\{(X_i, \delta_i), 1 \leq i \leq n\}$, that $l_n^*(\hat{\boldsymbol{\theta}})$ is arbitrarily close to the left member of Eq. (A.8), $I\!P_n$ a.s. That is, we can find $n_1 \equiv n_1(\epsilon)$, which, if needed, can be chosen greater than n_0, satisfying Eq. (A.9) and

$$
\left|l_n^*(\hat{\boldsymbol{\theta}}) - I\!E_n\left(\delta^* w_1(X^*, \hat{\boldsymbol{\theta}}) + (1-\delta^*) w_2(X^*, \hat{\boldsymbol{\theta}})\right)\right| \leq \epsilon. \qquad \text{(A.10)}
$$

Hence, for almost all sample sequences $\{(X_i, \delta_i), 1 \leq i \leq n\}$, $l_n^*(\hat{\boldsymbol{\theta}})$ is arbitrarily close to $L_H(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0), I\!P_n$ a.s. Since $\hat{\boldsymbol{\theta}}^*$ maximizes $l_n^*(\boldsymbol{\theta})$, the above arguments indicate that it suffices to establish, for almost all sample sequences $\{(X_i, \delta_i), 1 \leq i \leq n\}$, and for any $\eta > 0$, that

$$
I\!P_n\left(\sup_{\boldsymbol{\theta} \in \bar{\boldsymbol{\Theta}} \backslash V_\eta(\hat{\boldsymbol{\theta}})}\left[l_n^*(\boldsymbol{\theta}) - l_n^*(\hat{\boldsymbol{\theta}})\right] < 0\right) = 1. \qquad \text{(A.11)}
$$

To prove Eq. (A.11), we adapt to our setting Stute's (1992) method of proof. Define

$$
\bar{w}_1(z, \boldsymbol{\theta}) = w_1(z, \boldsymbol{\theta}) - w_1(z, \hat{\boldsymbol{\theta}}) \qquad , \qquad \bar{w}_2(z, \boldsymbol{\theta}) = w_2(z, \boldsymbol{\theta}) - w_2(z, \hat{\boldsymbol{\theta}}).
$$

For each $\tilde{\boldsymbol{\theta}} \in \bar{\boldsymbol{\Theta}} \backslash V_\eta(\hat{\boldsymbol{\theta}})$, we have that, as $\eta \downarrow 0$,

$$
\sup_{\boldsymbol{\theta} \in V_\eta(\tilde{\boldsymbol{\theta}})}\left[\delta^* \bar{w}_1(X^*, \boldsymbol{\theta}) + (1-\delta^*) \bar{w}_2(X^*, \boldsymbol{\theta})\right] \downarrow \left[\delta^* \bar{w}_1(X^*, \tilde{\boldsymbol{\theta}}) + (1-\delta^*) \bar{w}_2(X^*, \tilde{\boldsymbol{\theta}})\right], \qquad I\!P_n \text{ a.s.}
$$

By the monotone convergence theorem, as $\eta \downarrow 0$,

$$
I\!E_n\left(\sup_{\boldsymbol{\theta} \in V_\eta(\tilde{\boldsymbol{\theta}})}\left[\delta^* \bar{w}_1(X^*, \boldsymbol{\theta}) + (1-\delta^*) \bar{w}_2(X^*, \boldsymbol{\theta})\right]\right) \quad \rightarrow \quad I\!E_n\left(\delta^* \bar{w}_1(X^*, \tilde{\boldsymbol{\theta}}) + (1-\delta^*) \bar{w}_2(X^*, \tilde{\boldsymbol{\theta}})\right),
$$

with the term on the right side above being arbitrarily close to $L_H(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}) - L_H(\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}) < 0$, for large enough n, see also Eqs. (A.8) and (A.9). That is, by choosing η sufficiently small and n sufficiently large, $I\!E_n\left(\sup_{\boldsymbol{\theta} \in V_\eta(\tilde{\boldsymbol{\theta}})}\left[\delta^* \bar{w}_1(X^*, \boldsymbol{\theta}) + (1-\delta^*) \bar{w}_2(X^*, \boldsymbol{\theta})\right]\right)$ is strictly negative. With such a chosen η, large enough n, and $\tilde{\boldsymbol{\theta}} \in \bar{\boldsymbol{\Theta}} \backslash V_\eta(\hat{\boldsymbol{\theta}})$, we now focus our calculations on each $\boldsymbol{\theta} \in V_\eta(\tilde{\boldsymbol{\theta}})$. By the SLLN, for almost all sample sequences $\{(X_i, \delta_i), 1 \leq i \leq n\}$, the quantity

$$
l_n^*(\boldsymbol{\theta}) - l_n^*(\hat{\boldsymbol{\theta}}) \quad = \quad \frac{1}{n}\sum_{i=1}^{n}\left[w(X_i^*, \delta_i^*, \boldsymbol{\theta}) - w(X_i^*, \delta_i^*, \hat{\boldsymbol{\theta}})\right]
$$

can be made as close to $\mathbb{E}_n\left(\left[w(X^*,\delta^*,\boldsymbol{\theta}) - w(X^*,\delta^*,\hat{\boldsymbol{\theta}})\right]\right)$ as possible. However,

$$\begin{aligned}\mathbb{E}_n\left(\left[w(X^*,\delta^*,\boldsymbol{\theta}) - w(X^*,\delta^*,\hat{\boldsymbol{\theta}})\right]\right) &\leq \mathbb{E}_n\left(\sup_{\boldsymbol{\theta}\in V_\eta(\tilde{\boldsymbol{\theta}})}\left[w(X^*,\delta^*,\boldsymbol{\theta}) - w(X^*,\delta^*,\hat{\boldsymbol{\theta}})\right]\right)\\ &= \mathbb{E}_n\left(\sup_{\boldsymbol{\theta}\in V_\eta(\tilde{\boldsymbol{\theta}})}\left[\delta^*\bar{w}_1(X^*,\boldsymbol{\theta}) + (1-\delta^*)\bar{w}_2(X^*,\boldsymbol{\theta})\right]\right).\end{aligned}$$

Thus, for each $\boldsymbol{\theta} \in V_\eta(\tilde{\boldsymbol{\theta}})$, we have shown that for large n the quantity $l_n^*(\boldsymbol{\theta}) - l_n^*(\hat{\boldsymbol{\theta}})$ is bounded above by a strictly negative value and, therefore, so is its supremum taken over $\theta \in V_\eta(\tilde{\boldsymbol{\theta}})$. Hence, for each $\tilde{\boldsymbol{\theta}} \in \bar{\boldsymbol{\Theta}}\backslash V_\eta(\hat{\boldsymbol{\theta}})$, conclude for all sample sequences $\{(X_i,\delta_i), 1 \leq i \leq n\}$ that

$$\mathbb{P}_n\left(\sup_{\boldsymbol{\theta}\in V_\eta(\tilde{\boldsymbol{\theta}})}\left[l_n^*(\boldsymbol{\theta}) - l_n^*(\hat{\boldsymbol{\theta}})\right] < 0\right) = 1.$$

By compactness, $\bar{\boldsymbol{\Theta}}\backslash V_\eta(\hat{\boldsymbol{\theta}})$ can be covered by finitely many $V_\eta(\tilde{\boldsymbol{\theta}})$ proving Eq. (A.11). □

A.3 Proof of theorem 3

Under conditions of lemma 1, we first derive an asymptotic representation for $\hat{\lambda}_0^*(t)$.

Lemma 3 *For almost all sample sequences $\{(X_i,\delta_i), 1 \leq i \leq n\}$, uniformly for $t \in [t_1, t_2]$,*

$$\hat{\lambda}_0^*(t) = -\frac{n}{(\tilde{\sigma}(t))^2}\left\{\left(\log\hat{S}^*(t) - \log\hat{S}(t)\right) + O_{\mathbb{P}_n}\left(\frac{1}{n^2}\right)\right\}. \tag{A.12}$$

Proof From section 2.6, $\hat{\lambda}_0^*(t) \equiv \hat{\lambda}^*(\hat{S}(t),t)$ satisfies $g(\lambda(t)) = \log\hat{S}(t)$, where

$$g(\lambda) := \sum_{s\leq t}\log\left(1 - \frac{m^*(s,\hat{\boldsymbol{\theta}}^*)\Delta N^*(s)}{Y^*(s)+\lambda}\right). \tag{A.13}$$

Then $g(0) = \log\hat{S}^*(t)$ and $g'(0) = (\tilde{\sigma}^*(t))^2/n$, where

$$g'(\lambda) = \sum_{s\leq t}\frac{m(s,\hat{\boldsymbol{\theta}}^*)\Delta N^*(s)}{(Y^*(s)+\lambda)(Y^*(s)+\lambda - m(s,\hat{\boldsymbol{\theta}}^*)\Delta N^*(s))}.$$

Note, by definition, that $\hat{\lambda}_0^*(t) > m(s,\theta^*)\Delta N^*(s) - Y^*(s)$ for all $s \leq t \in [t_1,t_2]$. Then, we have the following inequality, that we will need later.

$$Y^*(s) + \hat{\lambda}_0^*(t) - m(s,\theta^*)\Delta N^*(s) > 0, \qquad \text{for all } s \leq t \in [t_1,t_2]. \tag{A.14}$$

First we show, for almost all sample sequences $\{(X_i,\delta_i), 1 \leq i \leq n\}$, that $|\hat{\lambda}_0^*(t)| = O_{\mathbb{P}_n}(n^{1/2})$, uniformly for $t \in [t_1,t_2]$. We need to consider the two cases $\hat{\lambda}_0^*(t) < 0$ and $\hat{\lambda}_0^*(t) > 0$.

When $\hat{\lambda}_0^*(t) < 0$, use Eq. (A.13) and follow the proof of lemma 2.2 of Li (1995) to obtain $-\log\hat{S}(t) \geq \hat{\Psi}^*(t)$ and

$$|\hat{\lambda}_0^*(t)| \leq \frac{n\left\{-\log\hat{S}(t) - \hat{\Psi}^*(t)\right\}}{-\log\hat{S}(t)} \leq \frac{n\left\{-\log\hat{S}(t) - \hat{\Psi}(t)\right\} + n\left\{\hat{\Psi}(t) - \hat{\Psi}^*(t)\right\}}{-\log\hat{S}(t_1)}. \tag{A.15}$$

Then follow the proof of lemma 7.1 of Breslow and Crowley (1974) to obtain

$$-\log\hat{S}(t) - \hat{\Psi}(t) \leq \frac{\hat{H}(t)}{n(1-\hat{H}(t))} \leq \frac{\hat{H}(\tau_H)}{n(1-\hat{H}(\tau_H))}, \tag{A.16}$$

with a similar inequality holding for the bootstrap data. Plugging (A.16) into (A.15) yields

$$|\hat{\lambda}_0^*(t)| \quad \leq \quad \frac{\hat{H}(\tau_H)/(1-\hat{H}(\tau_H)) + n^{1/2}\|\mathbb{Z}^*\|_{t_1}^{t_2}}{-\log \hat{S}(t_1)}.$$

Now apply proposition 3 to conclude, for almost all sample sequences $\{(X_i,\delta_i), 1 \leq i \leq n\}$, that $|\hat{\lambda}_0^*(t)| = O_{\mathbb{P}_n}(n^{1/2})$, uniformly for $t \in [t_1, t_2]$.

When $\hat{\lambda}_0^*(t) > 0$, follow the steps leading to inequality (2.13) of Li (1995) to obtain

$$\hat{\Psi}^*(t)\left(\frac{n}{n+|\hat{\lambda}_0^*(t)|}\right) \geq \hat{\Psi}^*(t) + \log \hat{S}^*(t) - \log \hat{S}(t). \tag{A.17}$$

However, applying the bootstrap counterpart of (A.16) to the right side of (A.17) yields

$$\hat{\Psi}^*(t) + \log \hat{S}^*(t) - \log \hat{S}(t) \quad \geq \quad -\frac{\hat{H}^*(t_2)}{n(1-\hat{H}^*(t_2))} - \log \hat{S}(t_1) \;\equiv\; C. \tag{A.18}$$

Since $\hat{H}^*(t_2)/(1-\hat{H}^*(t_2)) = O(1), \mathbb{P}_n$ a.s. , we shall consider an adequately large n that makes C in (A.18) strictly positive. Hence, for almost all sample sequences $\{(X_i,\delta_i), 1 \leq i \leq n\}$,

$$\hat{\Psi}^*(t) + \log \hat{S}^*(t) - \log \hat{S}(t) \quad > \quad 0, \qquad \mathbb{P}_n \text{ a.s.} \tag{A.19}$$

Finally, note from inequality (A.17), for almost all sample sequences $\{(X_i,\delta_i), 1 \leq i \leq n\}$, that $\log \hat{S}(t) \geq \log \hat{S}^*(t), \mathbb{P}_n$ a.s. Hence, inequality (A.17) can be manipulated to obtain an upper bound for $|\hat{\lambda}_0^*(t)|$, which, together with the continuous mapping theorem, yields, for almost all sample sequences $\{(X_i,\delta_i), 1 \leq i \leq n\}$, the desired rate, uniformly for $t \in [t_1, t_2]$:

$$|\hat{\lambda}_0^*(t)| \quad \leq \quad \frac{n\left\{\log \hat{S}(t) - \log \hat{S}^*(t)\right\}}{\hat{\Psi}^*(t) + \log \hat{S}^*(t) - \log \hat{S}(t)} \quad \leq \quad \frac{n}{C}\left\|\log \hat{S} - \log \hat{S}^*\right\|_{t_1}^{t_2} \;=\; O_{\mathbb{P}_n}(n^{1/2}).$$

Recall from Eq. (2.22) that $\hat{\lambda}_0^*(t) \equiv \hat{\lambda}_0^*(\hat{S}(t), t)$ satisfies $g(\hat{\lambda}_0^*(t) = \log \hat{S}(t)$, where

$$g(\lambda^*) \;:=\; \sum_{s \leq t} \log\left(1 - \frac{m(s,\hat{\boldsymbol{\theta}}^*)\Delta N^*(s)}{Y^*(s)+\lambda^*}\right).$$

Let $|\eta^*(t)| \leq |\hat{\lambda}_0^*(t)|$. As in the proof of lemma 2, Taylor's expansion of $g(\hat{\lambda}_0^*(t))$ about 0 yields

$$\log \hat{S}(t) \;\equiv\; g(\hat{\lambda}_0^*(t)) = g(0) + \hat{\lambda}_0^*(t)g'(0) + (\hat{\lambda}_0^*(t))^2 g''(\eta^*(t))/2.$$

Note that $g'(0) = (\tilde{\sigma}^*(t))^2/n$ and that $g''(\lambda^*)$ is the bootstrap equivalent of Eq. (A.3), where the data-based quantities are replaced with their bootstrap counterparts. It follows that

$$\hat{\lambda}_0^*(t) \quad = \quad -\frac{n}{(\tilde{\sigma}^*(t))^2}\left\{\left(\log \hat{S}^*(t) - \log \hat{S}(t)\right) + \frac{1}{2}(\hat{\lambda}_0^*(t))^2 g''(\eta^*(t))\right\}.$$

To prove the negligibility of $g''(\eta^*(t))$, note as in the proof of lemma 2 that

$$Y^*(s) + \eta^*(t) - m(s,\hat{\boldsymbol{\theta}}^*)\Delta N^*(s) > 0, \qquad \text{for all } s \leq t \leq \tau_H.$$

Hence, also, $Y^*(s) + \eta^*(t) > 0$, for all $s \leq t \leq \tau_H$. We then have

$$\begin{aligned} |g''(\eta^*(t))| \quad &= \quad \sum_{s \leq t}\left\{\frac{1}{(Y^*(s)+\eta^*(t) - m(s,\hat{\boldsymbol{\theta}}^*)\Delta N^*(s))^2} - \frac{1}{(Y^*(s)+\eta^*(t))^2}\right\} \\ &\leq \quad \frac{2}{Y^*(t)+\eta^*(t)}\sum_{s \leq t}\left\{\frac{1}{Y^*(s)+\eta^*(t) - m(s,\hat{\boldsymbol{\theta}}^*)\Delta N^*(s)} - \frac{1}{Y^*(s)+\eta^*(t)}\right\}. \end{aligned}$$

Let $(\breve{\sigma}^*(t))^2/n$ denote the sum on the right side above. Note, for almost all sample sequences $\{(X_i, \delta_i), 1 \leq i \leq n\}$, that $\|(\breve{\sigma}^*(t))^2 - \tilde{\sigma}^2(t)\| = o_{\mathbb{P}_n}(1)$. Hence, uniformly for $t \in [t_1, t_2]$,

$$|g''(\eta^*(t))| \;\leq\; \frac{2}{Y^*(t) + \eta^*(t)} \frac{(\breve{\sigma}^*(t_2))^2}{n} \;=\; O_{\mathbb{P}_n}\left(n^{-1}\right) O_{\mathbb{P}_n}\left(n^{-1}\right) \;=\; O_{\mathbb{P}_n}\left(n^{-2}\right). \qquad \square$$

Remark By the functional delta method, proposition 3, Eqs. (A.2) and (A.12), we have $\hat{\lambda}_0^*(t) = \hat{\lambda}_0(t) + o_{\mathbb{P}_n}(1)$.

Proof of Eq. (2.24) As in the proof of Eq. (2.17), we have, for almost all sample sequences $\{(X_i, \delta_i), 1 \leq i \leq n\}$, that $-2\log \mathcal{L}^*\left(\hat{S}(t), t\right) = \sum_{k=1}^3 I_k^*(t) + o_{\mathbb{P}_n}(1)$, where the $I_k^*(t)$ are the bootstrap counterparts of the $I_k(t)$ defined in the proof of Eq. (2.17). Then, by lemma 3,

$$\begin{aligned} I_1^*(t) &= (\hat{\lambda}_0^*(t))^2 \sum_{s \leq t} \left\{ \frac{1}{Y^*(s) - m(s, \hat{\boldsymbol{\theta}}^*)\Delta N^*(s)} - \frac{1}{Y^*(s)} \right\} \equiv (\hat{\lambda}_0^*(t))^2 \frac{(\breve{\sigma}^*(t))^2}{n} \\ &= \frac{(\breve{\sigma}^*(t))^2}{((\tilde{\sigma}^*(t))^2)^2} \left\{ n^{1/2}\left(\log \hat{S}^*(t) - \log \hat{S}(t)\right) + O_{\mathbb{P}_n}\left(n^{-3/2}\right)\right\}^2, \text{ uniformly for } t \in [t_1, t_2]. \end{aligned}$$

As in the proof of Eq. (2.17), applying lemma 3, we have

$$\begin{aligned} \|I_2^*\|_{t_1}^{t_2} &\leq \left(\|\hat{\lambda}_0^*\|_{t_1}^{t_2}\right)^3 \left(\frac{1}{Y^*(t_2) - \Delta N^*(t_2)} + \frac{1}{Y^*(t_2)}\right) \frac{(\tilde{\sigma}^*(t_2))^2}{n} \\ &= O_{\mathbb{P}_n}\left(n^{3/2}\right) O_{\mathbb{P}_n}\left(n^{-1}\right) O_{\mathbb{P}_n}\left(n^{-1}\right) \;=\; o_{\mathbb{P}_n}(1). \end{aligned}$$

Likewise, $\|I_3^*\|_{t_1}^{t_2}$ is bounded above by

$$\begin{aligned} & O_{\mathbb{P}_n}(n^2)\left[\frac{1}{(Y^*(t_2) - \Delta N^*(t_2))^2} + \frac{1}{Y^*(t_2)(Y^*(t_2) - \Delta N^*(t_2))} + \frac{1}{(Y^*(t_2))^2}\right] \frac{(\tilde{\sigma}^*(t_2))^2}{n} \\ =\; & O_{\mathbb{P}_n}(n^2)\, O_{\mathbb{P}_n}\left(n^{-2}\right) O_{\mathbb{P}_n}\left(n^{-1}\right) \;=\; o_{\mathbb{P}_n}(1). \end{aligned}$$

The proof of Theorem 3 is completed. $\square$

References

M.G. Akritas, Bootstrapping the Kaplan–Meier estimator, J. Amer. Statist. Assoc. 81 (1986) 1032–1038.

R. Berk, D. Jones, Goodness-of-fit test statistics that dominate the Kolmogorov statistics, Z. Wahrsch. Verw. Geb. 47 (1979) 47–59.

N.E. Breslow, J.J. Crowley, A large sample study of the life table and prodct limit estimates under random censorship, Ann. Statist. 2 (1973) 437–453.

C.F. Chung, Formulas for probabilities associated with Wiener and Brownian bridge processes, Technical report 79, Laboratory for Research in Statistics and Probability, Carleton University, Ottawa, Canada, 1986.

C.F. Chung, Wiener pack: a subroutine package for computing probabilities associated with Wiener and Brownian bridge processes, Paper 87-12 Geological Survey of Canada, 1987.

G. Claeskens, N.L. Hjort, Model Selection and Model Averaging, Cambridge University Press, 2008.

G. Dikta, On semiparametric random censorship models, J. Statist. Plann. Inference 66 (1998) 253–279.

G. Dikta, Asymptotically efficient estimation under semi-parametric random censorship models, J. Multivariate Anal. 124 (2014) 10–24.

G. Dikta, M. Kvesic, C. Schmidt, Bootstrap approximations in model checks for binary data, J. Amer. Statist. Assoc. 101 (2006) 521–530.

B. Efron, Censored data and the bootstrap, J. Amer. Statist. Assoc. 76 (1981) 312–319.

J.H.J Einmahl, I.W. McKeague, Confidence tubes for multiple quantile plots via empirical likelihood, Ann. Statist. 27 (1999) 1348–1367.

W. Hall, J.A. Wellner, Confidence bands for a survival curve from censored data, Biometrika 67 (1980) 133–143.

M. Hollander, I.W. McKeague, J. Yang, Likelihood ratio-based confidence bands for survival functions, J. Amer. Statist. Assoc. 92 (1997) 215-226.

L. Jager, J.A. Wellner, A new goodness of fit test: the reversed Berk–Jones statistic, Technical Report 443, Department of Statistics, University of Washington, 2005.

R. Koenker, J. Yoon, Parametric links for binary choice models: A Fisherian-Bayesian colloquy, J. Econometrics 152 (2009) 120–130.

G. Li, On nonparametric likelihood ratio estimation of survival probabilities for censored data, Statist. Probab. Lett. 25 (1995) 95–104.

G. Li, M. Hollander, I.W. McKeague, J. Yang, Nonparametric likelihood ratio confidence bands for quantile functions from incomplete survival data, Ann. Statist. 23 (1996) 628-640.

D. Matthews, Exact Nonparametric Confidence Bands for the Survivor Function, I. J. Biostatist. 9 (2013) 185–204.

P. McCullagh, J.A. Nelder, Generalized linear models (Second edition), Chapman & Hall, New York, 1989.

I.W. McKeague, Y. Zhao, Comparing distribution functions via empirical likelihood, Int. J. Biostatist. 1 (2005) Article 5.

S. Mondal, S. Subramanian, Model assisted Cox regression, J. Multivariate Anal. 123 (2014) 281–303.

B.J.T. Morgan, D.M. Smith, A note on Wadley's problem with overdispersion, Appl. Statist. 41 (1992) 349–354.

V. Nair, Confidence bands for survival functions with censored data: a comparative study, Technometrics 14 (1984) 265–275.

A. Owen, Empirical likelihood ratio confidence intervals for a single functional, Biometrika 75 (1988) 265–275.

A. Owen, Empirical likelihood ratio confidence regions, Ann. Stat. 18 (1990) 90–120.

A. Owen, Nonparametric coonfidence bands for a distribution function, J. Amer. Statist. Assoc. 90 (1995) 516–521.

M.I. Parzen, L.J. Wei, Z. Ying, Simultaneous confidence intervals for the difference of two survival functions, Scand. J. Statist. 24 (1997) 309–314.

W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical recipes in Fortran 77: The art of scientific computing (Second edition), Cambridge University Press, 1992.

W. Stute, Nonparametric model checks for regression, Ann. Statist. 25 (1997) 613–641.

S. Subramanian, The Missing censoring-indicator model of random censorship, in: N. Balakrishnan, C.R. Rao. (Eds.), Advances in Survival Analysis, Handbook of Statistics, 23 (2004) 123–141.

S. Subramanian, D. Bean, Hazard function estimation from homogeneous right censored data with missing censoring indicators, Statist. Methodol. 5 (2008) 515–527.

S. Subramanian, Model-based likelihood ratio confidence intervals for survival functions, Statist. Probab. Lett. 82 (2012) 626–635.

S. Subramanian, P. Zhang, Model-based confidence bands for survival functions, J. Statist. Plann. Inf. 143 (2013) 1355–1374.

A.A. Tsiatis, M. Davidian, B. McNeney, Multiple imputation methods for testing treatment differences in survival distributions with missing cause of failure, Biometrika 89 (2002) 238–244.

D.R. Thomas, G.L. Grunkemeier, Confidence interval estimation of survival probabilities for censored data, J. Amer. Statist. Assoc. 70 (1975) 865-871.

M. Yuan, Semiparametric censorship model with covariates, Test 14 (2005) 1–26.

M.J. Zhang, J. Klein Confidence bands for the difference of two survival curves under proportional hazards model, Lifetime Data Anal. 7 (2001) 243–254.