Ann. Inst. Statist. Math.
Vol. 52, No. 1, 160–172 (2000)

CONSISTENCY AND ASYMPTOTIC NORMALITY OF ESTIMATORS IN A PROPORTIONAL HAZARDS MODEL WITH INTERVAL CENSORING AND LEFT TRUNCATION

SOMNATH DATTA[1]*, GLEN A. SATTEN[2] AND JOHN M. WILLIAMSON[2]**

[1] *Department of Statistics, University of Georgia, Athens, GA 30602, U.S.A.*
[2] *Division of HIV/AIDS Prevention: Surveillance and Epidemiology, MS E-48, National Center for HIV/AIDS, STD and TB Prevention, Centers for Disease Control and Prevention, Atlanta, GA 30333, U.S.A.*

(Received August 18, 1997; revised November 12, 1998)

Abstract. Satten *et al.* (1998, *J. Amer. Statist. Assoc.*, **93**, 318–327) proposed an approach to the proportional hazards model for interval censored data in which parameter estimates are obtained by solving estimating equations which are the score equations for the full data proportional hazards model, averaged over all rankings of imputed failure times consistent with the observed censoring intervals. In this paper, we extend this approach to incorporate data that are left-truncated and right censored (dynamic cohort data). Consistency and asymptotic normality of the estimators obtained in this way are established.

Key words and phrases: Cox model, current status data, interval censoring, left truncation, survival analysis.

1. Introduction

The proportional hazards or Cox model (Cox (1972)) is used for assessing the effects of covariates on survival time. The standard approach of estimating the model parameters via the partial likelihood is applicable when a distinct failure or censoring time is observed for each individual or experimental unit. However, in many instances, one may only know a time interval in which the failure occurred, in which case we say the data are interval censored. More specifically, for each individual, instead of a failure time, we observe a censoring interval $[\ell_i, u_i)$ which is known to contain the actual failure time. Special cases of interval censoring include grouped data (Kalbfleisch and Prentice (1973)), where individuals are each seen at the same times (e.g., every six months or longer), and current status data (Grummer-Strawn (1993)), where individuals are seen only once after study enrollment.

When fitting the proportional hazards model to interval censored or grouped data, several approaches are currently available. Finkelstein (1986) considered a parametric method, in which the baseline distribution is fit simultaneously along with regression coefficients, by maximizing the full likelihood of the observed data; Diamond *et al.* (1986) proposed a similar model for current status data. Many authors have considered use of the marginal likelihood of all possible rankings for grouped data (Peto (1972), Kalbfleisch and Prentice (1972, 1973), Sinha *et al.* (1994), DeLong *et al.* (1994)). Although actual

* Research supported in part by NSA grant MDA 904-96-1-0049.
** Research supported in part by NIH grant T32-AI07442.

failure times are not observed, a rank-based approach to inference on regression parameters in the proportional hazards model that does not require knowledge of the baseline distribution is available (Satten (1996)).

Very recently, Satten *et al.* (1998) proposed a semi-parametric approach where a parametric form of the baseline distribution is assumed. Regression parameter estimates are then obtained by solving estimating equations which are the partial likelihood score equations for the full data proportional hazards model, averaged over all rankings of imputed failure times consistent with the observed censoring intervals. While a parametric baseline hazard is fit to provide a model for imputing the missing failure times, the method is semi-parametric in that the full data estimating equations for the regression parameter do not include information on this baseline hazard. The cost of this flexibility is that the resulting estimating equations for the regression parameters must be solved using Monte-Carlo methods, specifically, stochastic approximations. An advantage to this approach is that it can be adapted to handle data that are both left-truncated (Andersen *et al.* (1993)) and interval-censored; this generalization is considered here for the first time. Such data arises in dynamic cohort studies (Weinstock *et al.* (1998)), where the baseline hazard is a function of calendar time, rather than time on study. In these studies, rolling admission of individuals who have not yet experienced a failure event results in left-truncated data, while subsequent periodic observation causes failure-time data to be interval censored. In the next section, we discuss this approach in more details.

The main emphases of this paper are: to provide arguments for the consistency and asymptotic normality of the estimators proposed by Satten *et al.* (1998), which did not contain any proofs; to extend the Satten *et al.* approach to left-truncated as well as interval censored data; and to assess the performance of these estimators by a simulation study. Finally, the details of the proofs are themselves interesting as they are extendible to other estimating equations that are averages of full data estimating equations.

The asymptotic results and their proofs are presented in Sections 2 and 3. Section 4 contains a small simulation study to assess the effect of mis-specification of the baseline hazard function on the estimated regression parameters. The paper ends with a brief discussion in Section 5.

2. The estimators and their asymptotic properties

Following Satten *et al.* (1998), we consider an estimating equation which is the expected value of the partial likelihood score function for the full-data proportional hazards model with respect to the joint distribution of the rank order of the imputed failure times for all the left-truncated and interval-censored observations. Specifically, let $\boldsymbol{t} = (t_1, t_2, \ldots, t_n)$ be the vector of (possibly right-censored) failure times; let $\boldsymbol{\delta}$ be a vector of right-censoring indicators with components $\delta_i = 0$ if the i-th individual is right censored and $\delta_i = 1$ otherwise. Let τ_i be the left truncation time for the i-th individual; due to left truncation, data from the i-th individual is only observed because her failure time exceeds τ_i. If data are not subject to left truncation, then the distribution of τ_i is taken to be degenerate at 0. Let ℓ_i and u_i be the left and right end-points, respectively, of the censoring interval. Without loss of generality, we take $\ell_i \geq \tau_i$. Let $\boldsymbol{\ell}$, $\boldsymbol{u}$ and $\boldsymbol{\tau}$ denote vectors with i-th component ℓ_i, u_i and τ_i respectively. Let $\boldsymbol{x}_i$ be a vector of covariates for the i-th observation, and let $\boldsymbol{x}$ denote the matrix whose i-th row is $\boldsymbol{x}_i$. Let $\boldsymbol{d}_i$ denote the vector $(\boldsymbol{x}_i, \ell_i, u_i, \tau_i, \delta_i)$ and let $\boldsymbol{d}$ denote the matrix whose i-th row is $\boldsymbol{d}_i$. We assume that the observed data comprise iid samples of $\boldsymbol{d}_i$. Let $F(\boldsymbol{t} \mid \boldsymbol{d}; \boldsymbol{\beta}, \boldsymbol{\theta})$ be a parametric family of conditional distributions of failure times $\boldsymbol{t}$ given the observed censoring intervals, truncation times, covariates, and right-censoring indicator $\boldsymbol{\delta}$, and assume this family of distributions contains the true distribution of $\boldsymbol{t}$ and satisfies the proportional hazards assumption when $\boldsymbol{t}$ is unrestricted. For this family of distributions, the conditional expectation of $S_\beta(\boldsymbol{t} \mid \boldsymbol{\delta}, \boldsymbol{\tau}, \boldsymbol{x}; \boldsymbol{\beta})$, the partial likelihood score function for the full data (left-truncated) proportional hazards model, given the observed censoring intervals, is

$$\mathbb{S}_\beta(\boldsymbol{\beta},\boldsymbol{\theta}) = E_{F(\cdot|\boldsymbol{d};\boldsymbol{\theta})}[S_\beta(\boldsymbol{t} \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta})] = \int S_\beta(\boldsymbol{t} \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta}) dF(\boldsymbol{t} \mid \boldsymbol{d}; \boldsymbol{\beta}, \boldsymbol{\theta}); \tag{2.1}$$

the proposal is to estimate $\boldsymbol{\beta}$ by the solution to

$$\mathbb{S}_\beta(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}(\boldsymbol{\beta})) = \boldsymbol{0}, \tag{2.2}$$

where $\hat{\boldsymbol{\theta}}(\boldsymbol{\beta})$ is an estimate of the parameters $\boldsymbol{\theta}$, as a function of $\boldsymbol{\beta}$, required to specify F.

In equation (2.1) above, the partial likelihood score function for the full data (left-truncated) proportional hazards model $S_\beta(\boldsymbol{t} \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta})$ can be written as follows (see e.g. Andersen *et al.* (1993)). Let

$$\mathcal{R}_i = \{j : t_j \geq t_i \text{ and } \tau_j < t_i\};$$

note that if $\tau_i = 0 \forall i$ (no left truncation) then the risk sets $\mathcal{R}_i$ reduce to the usual risk sets for the proportional hazards model. As when left truncation is absent, the partial likelihood score function can be written as

$$S_\beta(\boldsymbol{t} \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta}) = \sum_{i=1}^n \delta_i \left\{ x_i - \frac{\sum_{j\in\mathcal{R}_i} x_j e^{\beta^T x_j}}{\sum_{j\in\mathcal{R}_i} e^{\beta^T x_j}} \right\}.$$

The distribution F has the product form

$$F(\boldsymbol{t} \mid \boldsymbol{d}; \boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{i=1}^n F(t_i \mid \ell_i, u_i, \boldsymbol{x}_i; \boldsymbol{\beta}, \boldsymbol{\theta})^{\delta_i} I(t_i \geq \ell_i)^{1-\delta_i}, \tag{2.3}$$

where $I(C) = 1$ if C is true and 0 otherwise. It is assumed that the censoring is independent of failure times and the truncation mechanism so that

$$F(t_i \mid \ell_i, u_i, \boldsymbol{x}_i; \boldsymbol{\beta}, \boldsymbol{\theta}) = \frac{F(t_i \mid \boldsymbol{x}_i; \boldsymbol{\beta}, \boldsymbol{\theta}) - F(\ell_i \mid \boldsymbol{x}_i; \boldsymbol{\beta}, \boldsymbol{\theta})}{F(u_i \mid \boldsymbol{x}_i; \boldsymbol{\beta}, \boldsymbol{\theta}) - F(\ell_i \mid \boldsymbol{x}_i; \boldsymbol{\beta}, \boldsymbol{\theta})} I(\ell_i \leq t_i < u_i).$$

As $F(t_i \mid \boldsymbol{x}_i; \boldsymbol{\beta}, \boldsymbol{\theta})$ is in the proportional hazards family, we have

$$F(t_i \mid \boldsymbol{x}_i; \boldsymbol{\beta}, \boldsymbol{\theta}) = 1 - [1 - F_o(t_i \mid \boldsymbol{\theta})]^{e^{\boldsymbol{\beta}\cdot\boldsymbol{x}_i}},$$

where $F_o(t \mid \boldsymbol{\theta})$ is the cumulative distribution function of failure-times in the baseline group; we will refer to F_o as the baseline distribution. We will denote by $\lambda_o(t)$ the hazard function for the distribution F_o.

The log-likelihood of the observed data (ℓ_i, u_i, δ_i), $1 \leq i \leq n$, given $\boldsymbol{x}$ and $\boldsymbol{\tau}$, is

$$\mathcal{L}^o = \sum_{i=1}^n \delta_i \ln \left[\frac{F(u_i \mid \boldsymbol{x}_i; \boldsymbol{\beta}, \boldsymbol{\theta}) - F(\ell_i \mid \boldsymbol{x}_i; \boldsymbol{\beta}, \boldsymbol{\theta})}{1 - F(\tau_i \mid \boldsymbol{x}_i; \boldsymbol{\beta}, \boldsymbol{\theta})} \right] + (1 - \delta_i) \ln \left[\frac{1 - F(\ell_i \mid \boldsymbol{x}_i; \boldsymbol{\beta}, \boldsymbol{\theta})}{1 - F(\tau_i \mid \boldsymbol{x}_i; \boldsymbol{\beta}, \boldsymbol{\theta})} \right]. \tag{2.4}$$

For a given $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ can be estimated by maximizing $\mathcal{L}^o$ with respect to $\boldsymbol{\theta}$. Estimating $\boldsymbol{\beta}$ using (2.2) and $\hat{\boldsymbol{\theta}}(\boldsymbol{\beta})$ which maximizes (2.4) with respect to $\boldsymbol{\theta}$ is equivalent to solving the estimating equations

$$\mathbb{S}_\beta(\boldsymbol{\beta},\boldsymbol{\theta}) = \mathbf{0} \tag{2.5}$$

$$\mathbb{U}^o_\theta(\boldsymbol{\beta},\boldsymbol{\theta}) \equiv \sum_{i=1}^{n} \mathcal{U}^o_\theta(\boldsymbol{d}_i;\boldsymbol{\beta},\boldsymbol{\theta}) = \mathbf{0} \tag{2.6}$$

simultaneously for $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$, where

$$\mathcal{U}^o_\theta(\boldsymbol{d}_i;\boldsymbol{\beta},\boldsymbol{\theta}) = \delta_i \frac{\partial \ln \left[\dfrac{F(u_i \mid \boldsymbol{x}_i;\boldsymbol{\beta},\boldsymbol{\theta}) - F(\ell_i \mid \boldsymbol{x}_i;\boldsymbol{\beta},\boldsymbol{\theta})}{1 - F(\tau_i \mid \boldsymbol{x}_i;\boldsymbol{\beta},\boldsymbol{\theta})} \right]}{\partial \boldsymbol{\theta}} + (1-\delta_i) \frac{\partial \ln \left[\dfrac{1 - F(\ell_i \mid \boldsymbol{x}_i;\boldsymbol{\beta},\boldsymbol{\theta})}{1 - F(\tau_i \mid \boldsymbol{x}_i;\boldsymbol{\beta},\boldsymbol{\theta})} \right]}{\partial \boldsymbol{\theta}}. \tag{2.7}$$

For future use, we define the corresponding partial derivatives with respect to $\boldsymbol{\beta}$ as $\mathcal{U}^o_\beta(\boldsymbol{d}_i;\boldsymbol{\beta},\boldsymbol{\theta})$, and define the maximum likelihood score function for $\boldsymbol{\beta}$ obtained from (2.4) as

$$\mathbb{U}^o_\beta(\boldsymbol{\beta},\boldsymbol{\theta}) = \frac{\partial \mathcal{L}^o}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \mathcal{U}^o_\beta(\boldsymbol{d}_i;\boldsymbol{\beta},\boldsymbol{\theta}).$$

We also define the score functions which would be used in maximum likelihood estimation from the 'complete' data problem, i.e. those that would result if we had observed the actual failure times for the (left-truncated) observations which were interval censored. Specifically, define

$$\mathcal{U}^c_\beta(t_i,\delta_i,\boldsymbol{x}_i,\tau_i;\boldsymbol{\beta},\boldsymbol{\theta}) = \delta_i \frac{\partial \ln \dfrac{f(t_i \mid \boldsymbol{x}_i,\boldsymbol{\beta},\boldsymbol{\theta})}{1 - F(\tau_i \mid \boldsymbol{x}_i;\boldsymbol{\beta},\boldsymbol{\theta})}}{\partial \boldsymbol{\beta}} + (1-\delta_i) \frac{\partial \ln \left[\dfrac{1 - F(t_i \mid \boldsymbol{x}_i,\boldsymbol{\beta},\boldsymbol{\theta})}{1 - F(\tau_i \mid \boldsymbol{x}_i;\boldsymbol{\beta},\boldsymbol{\theta})} \right]}{\partial \boldsymbol{\beta}}$$

$$\mathcal{U}^c_\theta(t_i,\delta_i,\boldsymbol{x}_i,\tau_i;\boldsymbol{\beta},\boldsymbol{\theta}) = \delta_i \frac{\partial \ln \dfrac{f(t_i \mid \boldsymbol{x}_i,\boldsymbol{\beta},\boldsymbol{\theta})}{1 - F(\tau_i \mid \boldsymbol{x}_i;\boldsymbol{\beta},\boldsymbol{\theta})}}{\partial \boldsymbol{\theta}} + (1-\delta_i) \frac{\partial \ln \left[\dfrac{1 - F(t_i \mid \boldsymbol{x}_i,\boldsymbol{\beta},\boldsymbol{\theta})}{1 - F(\tau_i \mid \boldsymbol{x}_i;\boldsymbol{\beta},\boldsymbol{\theta})} \right]}{\partial \boldsymbol{\theta}}$$

$$\mathbb{U}^c_\beta(\boldsymbol{t} \mid \boldsymbol{\delta},\boldsymbol{x},\boldsymbol{\tau};\boldsymbol{\beta},\boldsymbol{\theta}) \equiv \sum_{i=1}^{n} \mathcal{U}^c_\beta(t_i,\delta_i,\boldsymbol{x}_i,\tau_i;\boldsymbol{\beta},\boldsymbol{\theta}),$$

and

$$\mathbb{U}^c_\theta(\boldsymbol{t} \mid \boldsymbol{\delta},\boldsymbol{x},\boldsymbol{\tau};\boldsymbol{\beta},\boldsymbol{\theta}) \equiv \sum_{i=1}^{n} \mathcal{U}^c_\theta(t_i,\delta_i,\boldsymbol{x}_i,\tau_i;\boldsymbol{\beta},\boldsymbol{\theta}).$$

Note the relation between the 'observed' and 'complete' data score functions gives

$$\mathbb{U}^o_\beta(\boldsymbol{\beta},\boldsymbol{\theta}) = \int \mathbb{U}^c_\beta(\boldsymbol{t} \mid \boldsymbol{\delta},\boldsymbol{x},\boldsymbol{\tau};\boldsymbol{\beta},\boldsymbol{\theta}) dF(\boldsymbol{t} \mid \boldsymbol{d};\boldsymbol{\beta},\boldsymbol{\theta}).$$

Let $\boldsymbol{\beta}_o$ and $\boldsymbol{\theta}_o$ denote the true values of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. First, note that by equation (2.1), assuming $F_o(t \mid \boldsymbol{\theta}_o)$ and the proportional hazards assumption are correct, the unconditional expected value of $\mathbb{S}_\beta(\boldsymbol{\beta}_o, \boldsymbol{\theta}_o)$ is equal to the unconditional expected value of $\boldsymbol{S}_\beta(\boldsymbol{t} \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta})$, which in turn is zero (Andersen *et al.* (1993)). Hence, the estimating equation (2.5) is unbiased. By the usual properties of maximum likelihood estimation, the estimating equation (2.6) is also unbiased. If $\mathbb{S}_\beta(\boldsymbol{\beta}, \boldsymbol{\theta})$ were a sum of iid terms for each observation, we could use standard results for estimating equations to conclude asymptotic consistency and normality of $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$. This not being the case, a deeper analysis is required to establish these asymptotic results. Also note that the usual martingale theory does not seem to apply to interval censored data, as conditioning failure times to lie in a finite interval involves both past and future information in the hazard function.

Following similar arguments as in Lin and Wei (1989), one can show that $n^{-1/2}\boldsymbol{S}_\beta(\boldsymbol{t} \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta})$ is expressible as a sum of iid terms up to $o_p(1)$; specifically,

$$n^{-1/2}\boldsymbol{S}_\beta(\boldsymbol{t} \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta}_o) = n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\phi}(t_i, \boldsymbol{x}_i, \tau_i, \delta_i; \boldsymbol{\beta}_o) + o_p(1), \tag{2.8}$$

where

$$\begin{aligned}\boldsymbol{\phi}(t_i, \boldsymbol{x}_i, \tau_i, \delta_i; \boldsymbol{\beta}) = {} & \delta_i \left(\boldsymbol{x}_i - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t_i)}{s^{(0)}(\boldsymbol{\beta}, t_i)}\right) \\ & - e^{\boldsymbol{\beta}' \cdot \boldsymbol{x}_i}\int_0^\infty Y_i(t)\left(\boldsymbol{x}_i - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)}\right)\frac{dH(t)}{s^{(0)}(\boldsymbol{\beta}, t_i)}\end{aligned} \tag{2.9}$$

and where

$$\begin{aligned} Y_i(t) &= I[t_i \geq t, \tau_i < t], \\ H(t) &= P[t_1 \leq t, \delta_1 = 1, \tau_1 < t], \\ \boldsymbol{S}^{(r)}(\boldsymbol{\beta}, t) &= n^{-1}\sum_{i=1}^{n} I[t_i \geq t, \tau_i < t] e^{\boldsymbol{\beta}^T \cdot \boldsymbol{x}_i}\boldsymbol{x}_i^{\otimes r}, \qquad r = 0, 1, 2 \end{aligned} \tag{2.10}$$

and

$$\boldsymbol{s}^{(r)}(\boldsymbol{\beta}, t) = E[\boldsymbol{S}^{(r)}(\boldsymbol{\beta}, t)]. \tag{2.11}$$

In the above, for a vector $\boldsymbol{a}$, $\boldsymbol{a}^{\otimes r}$ stands for 1, $\boldsymbol{a}$ and $\boldsymbol{a}\boldsymbol{a}^T$, respectively for $r = 0$, 1 and 2. We show that integration of (2.8) allows us to write

$$n^{-1/2}\mathbb{S}_\beta(\boldsymbol{\beta}_o, \boldsymbol{\theta}_o) = n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\psi}(\boldsymbol{d}_i; \boldsymbol{\beta}_o, \boldsymbol{\theta}_o) + o_p(1), \tag{2.12}$$

where

$$\boldsymbol{\psi}(\boldsymbol{d}_i; \boldsymbol{\beta}_o, \boldsymbol{\theta}_o) = \begin{cases} \displaystyle\int_{\ell_i}^{u_i} \boldsymbol{\phi}(t_i, \boldsymbol{x}_i, \tau_i, \delta_i; \boldsymbol{\beta}_o) dF(t_i \mid \ell_i, u_i, \boldsymbol{x}_i; \boldsymbol{\beta}_o, \boldsymbol{\theta}_o) & \text{if} \quad \delta_i = 1 \\ \boldsymbol{\phi}(\ell_i, \boldsymbol{x}_i, \tau_i, \delta_i; \boldsymbol{\beta}_o) & \text{if} \quad \delta_i = 0. \end{cases} \tag{2.13}$$

As mentioned earlier, we assume that $\boldsymbol{d}_i = (\boldsymbol{x}_i, \ell_i, u_i, \tau_i, \delta_i)$ are iid, and that $\boldsymbol{x}$ is bounded. Let $\|\boldsymbol{a}\|$ denote the Euclidean norm of $\boldsymbol{a}$ if $\boldsymbol{a}$ is a vector and the operator norm if $\boldsymbol{a}$ is a matrix. For notational convenience, let $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_o$ denote $(\boldsymbol{\beta}, \boldsymbol{\theta})$ and $(\boldsymbol{\beta}_o, \boldsymbol{\theta}_o)$, respectively. We will assume, without specifying it on a case by case basis, that certain quantities are differentiable in the parameters $\boldsymbol{\gamma}$, at least in a neighborhood of the true value γ_o, and that the order of certain differentiation and integration can be interchanged. For $\epsilon > 0$ and $\boldsymbol{b} \in \mathcal{R}^p$ for some p, let $N_\epsilon(\boldsymbol{b}) = \{\boldsymbol{b}' \mid \|\boldsymbol{b}' - \boldsymbol{b}\| < \epsilon\}$ denote the ϵ-neighborhood of $\boldsymbol{b}$. Throughout, $\vee$ will denote the suprema of functions of $\boldsymbol{\gamma}$ over $N_{K/\sqrt{n}}(\boldsymbol{\gamma}_o)$, where K is a constant. If the function does not involve $\boldsymbol{\theta}$, then $\vee$ denotes the supremum over $N_{K/\sqrt{n}}(\boldsymbol{\beta}_o)$.

We say that an iid sequence indexed by $\boldsymbol{\gamma}$, $\{y_i(\boldsymbol{\gamma}), i \geq 1\}$, satisfies the U-WLLN condition at $\boldsymbol{\gamma}_o$ if $E(y_1(\boldsymbol{\gamma}))$ is continuous at $\boldsymbol{\gamma}_o$, and if for some $\delta > 0$, $E(y_1^*) < \infty$, where $y_1^* = \sup_{\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_o\|<\delta} \|y_1(\boldsymbol{\gamma})\|$, and if

$$\limsup_{\rho \downarrow 0} E[V_1(\boldsymbol{\gamma})] = 0 \tag{2.14}$$

for each $\boldsymbol{\gamma} \in N_\delta(\boldsymbol{\gamma}_o)$, where $V_1(\boldsymbol{\gamma}) = \sup\{\|y_1(\boldsymbol{\gamma}') - y_1(\boldsymbol{\gamma})\| : \boldsymbol{\gamma}' \in N_\rho(\boldsymbol{\gamma})\}$. If this condition holds, then $n^{-1}\sum y_i(\boldsymbol{\gamma}) \xrightarrow{P} E[y_1(\boldsymbol{\gamma})]$ uniformly in $\boldsymbol{\gamma} \in N_{\delta/2}(\boldsymbol{\gamma}_o)$; see e.g. Datta (1988). Note that (2.14) can be replaced with more familiar sufficient conditions such as

$$\|y_1(\boldsymbol{\gamma}') - y_1(\boldsymbol{\gamma})\| \leq d(\|\boldsymbol{\gamma}' - \boldsymbol{\gamma}\|)W_1 \quad \forall \boldsymbol{\gamma}, \boldsymbol{\gamma}' \in N_\delta(\boldsymbol{\gamma}_o),$$

where $d(x) \to 0$ as $x \to 0$, and where $E[W_1] < \infty$.

Let $\mathcal{U}^o(\boldsymbol{d}_i;\boldsymbol{\beta},\boldsymbol{\theta}) = (\mathcal{U}^o_\beta(\boldsymbol{d}_i;\boldsymbol{\beta},\boldsymbol{\theta})^T, \mathcal{U}^o_\theta(\boldsymbol{d}_i;\boldsymbol{\beta},\boldsymbol{\theta})^T)^T$ and $\mathcal{U}^c(t_i,\delta_i,\boldsymbol{x}_i,\tau_i;\boldsymbol{\beta},\boldsymbol{\theta}) = (\mathcal{U}^c_\beta(t_i,\delta_i,\boldsymbol{x}_i,\tau_i;\boldsymbol{\beta},\boldsymbol{\theta})^T, \mathcal{U}^c_\theta(t_i,\delta_i,\boldsymbol{x}_i,\tau_i;\boldsymbol{\beta},\boldsymbol{\theta})^T)^T$. Assuming that the following regularity conditions hold:

C1: $E\|\mathcal{U}^o(\boldsymbol{d}_1;\boldsymbol{\beta}_o,\boldsymbol{\theta}_o)\|^2 < \infty$; $\|\mathcal{U}^c(t_i,\delta_i,\boldsymbol{x}_i,\tau_i;\boldsymbol{\beta}_o,\boldsymbol{\theta}_o)\|$ and $w(\boldsymbol{d}_1,t_1,\delta_1) = \sup_{\boldsymbol{\gamma}\in N_\epsilon(\boldsymbol{\gamma}_o)}\|\frac{\partial}{\partial\boldsymbol{\gamma}}\mathcal{U}^c(t_1;\boldsymbol{d}_1;\boldsymbol{\beta},\boldsymbol{\theta})\|$, for some $\epsilon > 0$, have finite moment generating functions in some neighborhood of 0,

C2: The sequences $\left\{\frac{\partial\boldsymbol{\phi}}{\partial\boldsymbol{\gamma}}\right\}$, $\left\{\frac{\partial\mathcal{U}^o}{\partial\boldsymbol{\gamma}}\right\}$, $\{\boldsymbol{\phi}\mathcal{U}^{oT}\}$ and $\{\boldsymbol{\phi}\mathcal{U}^{cT}\}$ each satisfy the U-WLLN condition at $\boldsymbol{\gamma}_o$,

we obtain the following theorem establishing the consistency of the solution of the proposed estimating equations.

THEOREM 2.1. *Under* C1 *and* C2, *given any* $\epsilon > 0$, *there exist a* $K < \infty$ *and an integer* n_o *such that*

$$\Pr\{(2.5)\ \textit{and}\ (2.6)\ \textit{have a solution}\ \hat{\gamma}\ \textit{with}\ n^{1/2}\|\hat{\gamma} - \boldsymbol{\gamma}_o\| \leq K\} \geq 1 - \epsilon \ \ \forall n \geq n_o.$$

If we assume further that

C3: The matrix $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}$ is non-singular, where

$$\boldsymbol{A}_{11} = \boldsymbol{Q}(\boldsymbol{\beta}_o) - E\{\boldsymbol{\phi}(t_1,\boldsymbol{x}_1,\tau_1,\delta_1;\boldsymbol{\beta}_o) \cdot [\mathcal{U}^c_\beta(t_1,\delta_1,\boldsymbol{x}_1,\tau_1;\boldsymbol{\beta}_o,\boldsymbol{\theta}_o) - \mathcal{U}^o_\beta(\boldsymbol{d}_1;\boldsymbol{\beta}_o,\boldsymbol{\theta}_o)]^T\}, \tag{2.15}$$

$$\boldsymbol{A}_{12} = E\{\boldsymbol{\phi}(t_1,\boldsymbol{x}_1,\tau_1,\delta_1,\boldsymbol{\beta}_0) \cdot [\mathcal{U}^c_\theta(t_1,\delta_1,\boldsymbol{x}_1,\tau_1;\boldsymbol{\beta}_o,\boldsymbol{\theta}_o) - \mathcal{U}^o_\theta(\boldsymbol{d}_1;\boldsymbol{\beta}_o,\boldsymbol{\theta}_o)]^T\}, \tag{2.16}$$

$$\boldsymbol{A}_{21} = E\left\{\left.\frac{\partial\mathcal{U}^o_\theta(\boldsymbol{d}_1;\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\beta}}\right|_{\substack{\boldsymbol{\beta}=\boldsymbol{\beta}_0,\\ \boldsymbol{\theta}=\boldsymbol{\theta}_o}}\right\}, \tag{2.17}$$

$$\boldsymbol{A}_{22} = E\left\{\left.\frac{\partial\mathcal{U}^o_\theta(\boldsymbol{d}_1;\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\substack{\boldsymbol{\beta}=\boldsymbol{\beta}_0,\\ \boldsymbol{\theta}=\boldsymbol{\theta}_o}}\right\} \tag{2.18}$$

and

$$Q(\boldsymbol{\beta}) = \int \left\{ \frac{s^{(2)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)} - \frac{s^{(1)}(\boldsymbol{\beta}, t)^{\otimes 2}}{s^{(0)}(\boldsymbol{\beta}, t)^2} \right\} s^{(0)}(\boldsymbol{\beta}, t)\lambda_o(t)dt,$$

then we obtain the following theorem establishing the asymptotic distribution of the solutions of the proposed estimating equations.

THEOREM 2.2. *Let* $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$ *solve* (2.5) *and* (2.6) *and* $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_o) = O_p(1)$. *Then under* C1–C3,

$$n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_o) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{V})$$

where the variance-covariance matrix $\boldsymbol{V}$ *has the sandwich form*

$$\boldsymbol{V} = \boldsymbol{A}^{-1} \cdot \boldsymbol{\Psi} \cdot \boldsymbol{A}^{-1}, \tag{2.19}$$

with $\boldsymbol{\Psi}$ *is given by*

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} \\ \boldsymbol{\Psi}_{12}^T & \boldsymbol{\Psi}_{22} \end{bmatrix},$$

where

$$\boldsymbol{\Psi}_{11} = E[\boldsymbol{\psi}(\boldsymbol{d}_1; \boldsymbol{\beta}_o, \boldsymbol{\theta}_o)\boldsymbol{\psi}(\boldsymbol{d}_1; \boldsymbol{\beta}_o, \boldsymbol{\theta}_o)^T], \tag{2.20}$$

$$\boldsymbol{\Psi}_{12} = E[\boldsymbol{\psi}(\boldsymbol{d}_1; \boldsymbol{\beta}_o, \boldsymbol{\theta}_o)\boldsymbol{\mathcal{U}}_\theta^o(\boldsymbol{d}_1; \boldsymbol{\beta}_o, \boldsymbol{\theta}_o)^T], \tag{2.21}$$

and

$$\boldsymbol{\Psi}_{22} = E[\boldsymbol{\mathcal{U}}_\theta^o(\boldsymbol{d}_1; \boldsymbol{\beta}_o, \boldsymbol{\theta}_o)\boldsymbol{\mathcal{U}}_\theta^o(\boldsymbol{d}_1; \boldsymbol{\beta}_o, \boldsymbol{\theta}_o)^T]. \tag{2.22}$$

3. Proofs

The following two lemmas will be used to prove the theorems stated in Section 2. The first lemma shows that the estimating equations with interval censored data inherit an asymptotic iid representation from the full data estimating equations.

LEMMA 3.1. *As* $n \to \infty$,

$$n^{-1/2}\mathbb{S}_\beta(\boldsymbol{\gamma}) = n^{-1/2}\tilde{\mathbb{S}}_\beta(\boldsymbol{\gamma}) + o_p(1) \tag{3.1}$$

uniformly in $\boldsymbol{\gamma} \in N_{K/\sqrt{n}}(\boldsymbol{\gamma}_o)$, *for each* $K < \infty$, *where*

$$\tilde{\mathbb{S}}_\beta(\boldsymbol{\gamma}) \equiv \sum_{i=1}^n \boldsymbol{\psi}(\boldsymbol{d}_i; \boldsymbol{\gamma}).$$

PROOF. Consider $\boldsymbol{R}_n(\boldsymbol{\beta}) = n^{-1/2}\{S_\beta(\boldsymbol{t} \mid \boldsymbol{\delta}, \boldsymbol{x}; \boldsymbol{\beta}) - \tilde{S}_\beta(\boldsymbol{t} \mid \boldsymbol{\delta}, \boldsymbol{x}; \boldsymbol{\beta})\}$, where

$$\tilde{S}_\beta(\boldsymbol{t} \mid \boldsymbol{\delta}, \boldsymbol{x}; \boldsymbol{\beta}) \equiv \sum_{i=1}^n \boldsymbol{\phi}(t_i, \boldsymbol{x}_i, \tau_i, \delta_i; \boldsymbol{\beta}).$$

Then

$$n^{-1/2}\{\mathbb{S}_\beta(\gamma) - \tilde{\mathbb{S}}_\beta(\gamma)\} = \int \boldsymbol{R}_n(\boldsymbol{\beta}) dF(\boldsymbol{t} \mid \boldsymbol{d}; \boldsymbol{\beta}, \boldsymbol{\theta})$$

$$= \int \boldsymbol{R}_n(\boldsymbol{\beta}) \exp\{n^{-1/2}[\mathbb{U}^c(\boldsymbol{t} \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\gamma}^*) - \mathbb{U}^o(\boldsymbol{\gamma}^*)]\} \cdot dF(\boldsymbol{t} \mid \boldsymbol{d}; \boldsymbol{\beta}_o, \boldsymbol{\theta}_o),$$

where $\boldsymbol{\gamma}^*$ lies on the line segment joining $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_o$. Therefore,

$$\bigvee \|n^{-1/2}\{\mathbb{S}_\beta(\boldsymbol{\gamma}) - \tilde{\mathbb{S}}_\beta(\boldsymbol{\gamma})\}\| \le A_n B_n, \tag{3.2}$$

where

$$A_n = \exp\left\{K \cdot \bigvee \|n^{-1/2}\mathbb{U}^o(\boldsymbol{\gamma})\|\right\}$$

and

$$B_n = E\left[(\vee\|\boldsymbol{R}_n(\boldsymbol{\beta})\|)\exp\left\{K \cdot \bigvee \|n^{-1/2}\mathbb{U}^c(\boldsymbol{\gamma})\|\right\} \mid \boldsymbol{d}; \boldsymbol{\gamma}_o\right].$$

Note that by a Taylor expansion,

$$\bigvee \|n^{-1/2}\mathbb{U}^o(\boldsymbol{\gamma})\| \le \|n^{-1/2}\mathbb{U}^o(\boldsymbol{\gamma}_o)\| + K \cdot \bigvee \left\|n^{-1}\frac{\partial}{\partial \boldsymbol{\gamma}}\mathbb{U}^o(\boldsymbol{\gamma})\right\|.$$

The first term above is $O_p(1)$ since the central limit theorem ensures that $n^{-1/2}\mathbb{U}^o(\boldsymbol{\gamma}_o)$ converges to a zero-mean normal random vector. The second term converges by the U-WLLN condition to $K \cdot \|E\frac{\partial}{\partial \boldsymbol{\gamma}}\mathcal{U}^o(\boldsymbol{d}_1; \boldsymbol{\gamma})\,|_{\boldsymbol{\gamma}_o}\|$; hence, $A_n = O_p(1)$.

To estimate the magnitude of B_n, consider its expected value:

$$E[B_n] \le (E[R_n^{*3/2}])^{2/3}\left(E\left[\exp\left\{3K \cdot \bigvee \|n^{-1/2}\mathbb{U}^c(\boldsymbol{\gamma})\|\right\}\right]\right)^{1/3},$$

where $R_n^* = \vee\|\boldsymbol{R}_n(\boldsymbol{\beta})\|$. By a similar Taylor expansion as used above,

$$\bigvee \|n^{-1/2}\mathbb{U}^c(\boldsymbol{\gamma})\| \le \|n^{-1/2}\mathbb{U}^c(\boldsymbol{\gamma}_o)\| + K \cdot n^{-1}\sum_{i=1}^n w(\boldsymbol{x}_i, t_i, \delta_i), \tag{3.3}$$

where $w(\boldsymbol{x}_i, t_i, \delta_i)$ is defined in C1. By elementary calculations with iid summands, the moment generating function of the rhs of (3.3) is bounded in n (and, in fact, is convergent). Note also that

$$R_n^* \le \|\boldsymbol{R}_n(\boldsymbol{\beta}_o)\| + K \cdot \bigvee\|n^{-1}\mathcal{I}_\beta(\boldsymbol{t} \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta}) - \boldsymbol{Q}(\boldsymbol{\beta}_o)\| + K \cdot \bigvee\left\|-n^{-1}\sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\beta}}\boldsymbol{\phi}(t_i, \boldsymbol{x}_i, \tau_i, \delta_i, \boldsymbol{\beta}) - \boldsymbol{Q}(\boldsymbol{\beta}_o)\right\|, \tag{3.4}$$

where $\mathcal{I}_\beta(\boldsymbol{t} \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} S_\beta(\boldsymbol{t} \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta})$ is the partial likelihood information matrix for the full data (left-truncated) proportional hazards model. By similar arguments as in Theorem 2.1 of Lin and Wei (1989) we get that $\|\boldsymbol{R}_n(\boldsymbol{\beta}_o)\| = o_p(1)$. By an adaptation of

Theorem 3.2 (and Theorem 4.2) of Andersen and Gill (1982) to the case of left truncated data, the second term of the right hand side of (3.4) is $o_p(1)$. Because the U-WLLN condition can be shown to hold for $\partial\boldsymbol{\phi}/\partial\boldsymbol{\beta}$, the third term of the right hand side of (3.4) is $o_p(1)$ as well. Therefore by (3.4), $R_n^* = o_p(1)$. To prove that $E[R_n^{*3/2}] \to 0$, it is now enough to show that the second moment of R_n^* is bounded. It follows from the properties of $\mathbb{S}_\beta(\boldsymbol{\gamma}_o)$ and $\bar{\mathbb{S}}_\beta(\boldsymbol{\gamma}_o)$ that

$$\limsup_{n\to\infty} E\|\boldsymbol{R}_n(\boldsymbol{\beta}_o)\|^2 < \infty. \tag{3.5}$$

Since the space of the explanatory variables $\boldsymbol{x}$ is bounded, it follows that

$$\bigvee \|n^{-1}\boldsymbol{\mathcal{I}}_\beta(t \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta})\| \le \bigvee \int \left\| \frac{S^{(2)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)} - \frac{S^{(1)}(\boldsymbol{\beta}, t)^{\otimes 2}}{S^{(0)}(\boldsymbol{\beta}, t)^2} \right\| d\bar{N}(t) \le M,$$

where $M < \infty$ is a constant independent of n, and where

$$\bar{N}(t) = n^{-1}\sum_{i=1}^{n} I(t_i \le t, \delta_i = 1, t_i > \tau_i).$$

In addition, it can be verified that $\bigvee \|n^{-1}\sum_{i=1}^n \frac{\partial}{\partial\boldsymbol{\beta}}\boldsymbol{\phi}(t_i, x_i, \tau_i, \delta_i, \boldsymbol{\beta}_o)\|$ has finite second moment. Therefore, it follows from (3.4) that $\text{limsup}_n E[R_n^{*2}] < \infty$, which in turn proves that $E[R_n^{*3/2}] \to 0$. Therefore, $E[B_n] \to 0$ and hence $A_nB_n = o_p(1)$, proving the lemma by (3.2). □

Let

$$\boldsymbol{A}_n(\boldsymbol{\beta}, \boldsymbol{\theta}) \equiv \begin{bmatrix} \boldsymbol{A}_{11}^n(\boldsymbol{\beta}, \boldsymbol{\theta}) & \boldsymbol{A}_{12}^n(\boldsymbol{\beta}, \boldsymbol{\theta}) \\ \boldsymbol{A}_{21}^n(\boldsymbol{\beta}, \boldsymbol{\theta}) & \boldsymbol{A}_{22}^n(\boldsymbol{\beta}, \boldsymbol{\theta}) \end{bmatrix} = -\frac{1}{n} \begin{bmatrix} \dfrac{\partial \mathbb{S}_\beta(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial\boldsymbol{\beta}} & \dfrac{\partial \mathbb{S}_\beta(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \\ \dfrac{\partial \mathbb{U}_\theta^o(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial\boldsymbol{\beta}} & \dfrac{\partial \mathbb{U}_\theta^o(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \end{bmatrix}.$$

$\boldsymbol{A}_{21}^n$ and $\boldsymbol{A}_{22}^n$ are easily obtained by differentiation of rhs(2.7) with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, respectively, while $\boldsymbol{A}_{11}^n$ and $\boldsymbol{A}_{12}^n$ are given by

$$\begin{aligned} \boldsymbol{A}_{11}^n(\boldsymbol{\beta}, \boldsymbol{\theta}) = \frac{1}{n}\int [&\boldsymbol{\mathcal{I}}_\beta(t \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta}) \\ &- S_\beta(t \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta})[\mathbb{U}_\beta^c(t \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta}, \boldsymbol{\theta}) - \mathbb{U}_\beta^o(\boldsymbol{\beta}, \boldsymbol{\theta})]^T]dF(t \mid \boldsymbol{d}; \boldsymbol{\beta}, \boldsymbol{\theta}) \end{aligned}$$

and

$$\boldsymbol{A}_{12}^n(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{1}{n}\int S_\beta(t \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta})[\mathbb{U}_\theta^c(t \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta}, \boldsymbol{\theta}) - \mathbb{U}_\theta^o(\boldsymbol{\beta}, \boldsymbol{\theta})]^T dF(t \mid \boldsymbol{d}; \boldsymbol{\beta}, \boldsymbol{\theta}).$$

The next lemma establishes the local uniform convergence of the Hessian matrix.

LEMMA 3.2. *For any $K < \infty$,*

$$\bigvee_{N_{Kn^{-1/2}}(\boldsymbol{\gamma}_0)} \|\boldsymbol{A}_{ij}^n(\boldsymbol{\gamma}) - \boldsymbol{A}_{ij}\| = o_p(1).$$

PROOF. Convergence of $\boldsymbol{A}_{21}^n(\boldsymbol{\gamma})$ and $\boldsymbol{A}_{22}^n(\boldsymbol{\gamma})$ follows immediately by the U-WLLN condition. As shown before, $\bigvee\|n^{-1}\boldsymbol{\mathcal{I}}_\beta(t \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta}) - \boldsymbol{Q}(\boldsymbol{\beta}_o)\| = o_p(1)$ and is bounded;

therefore the convergence in probability holds in L_1 as well. By a similar calculation as in (3.2),

$$\Delta_n \equiv \bigvee \left\| n^{-1} \int \mathcal{I}_\beta(t \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta}) dF(t \mid \boldsymbol{d}; \boldsymbol{\gamma}) - \boldsymbol{Q}(\boldsymbol{\beta}_o) \right\| \leq A_n E[C_n \mid \boldsymbol{d}, \boldsymbol{\gamma}_o], \tag{3.6}$$

where A_n is as in (3.2) and

$$C_n = \vee \| n^{-1} \mathcal{I}_\beta(\boldsymbol{t} \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta}) - \boldsymbol{Q}(\boldsymbol{\beta}_o) \| \exp\left\{ K \cdot \bigvee \| n^{-1/2} \mathbb{U}^c(\boldsymbol{\gamma}) \| \right\} = o_p(1).$$

Note also for large n,

$$0 \leq C_n \leq M \exp\left\{ K \| n^{-1/2} \mathbb{U}^c(\gamma_o) \| + K^2 n^{-1} \sum_{i=1}^{n} w(\boldsymbol{x}_i, t_i, \delta_i) \right\}, \tag{3.7}$$

for some constant M independent of n. Using elementary calculations for the iid summands, the expected value of the right hand side of (3.7) can be shown to converge to that of its in probability limit. Therefore, by the extended dominated convergence theorem, $E[C_n] \to 0$ and hence

$$\Delta_n \xrightarrow{p} 0 \tag{3.8}$$

by (3.6). Using similar arguments as above we can show that

$$\bigvee n^{-1} \int \boldsymbol{R}_n(\boldsymbol{\beta}) [\mathbb{U}^c_\beta(\boldsymbol{t} \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta}, \boldsymbol{\theta}) - \mathbb{U}^o_\beta(\boldsymbol{\beta}, \boldsymbol{\theta})]^T dF(\boldsymbol{t} \mid \boldsymbol{d}; \boldsymbol{\gamma}) \xrightarrow{p} 0. \tag{3.9}$$

Note also that

$$\begin{aligned} n^{-1} \int & \tilde{\boldsymbol{S}}_\beta(\boldsymbol{t} \mid \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{\tau}; \boldsymbol{\beta}) [\mathbb{U}^c_\beta(\boldsymbol{t} \mid \boldsymbol{\delta}, \boldsymbol{x}; \boldsymbol{\beta}, \boldsymbol{\theta}) - \mathbb{U}^o_\beta(\boldsymbol{\beta}, \boldsymbol{\theta})]^T dF(\boldsymbol{t} \mid \boldsymbol{d}; \boldsymbol{\gamma}) \\ &= n^{-1} \int \sum_{i=1}^{n} \boldsymbol{\phi}(t_i, \boldsymbol{x}_i, \tau_i, \delta_i, \boldsymbol{\beta}) [\boldsymbol{\mathcal{U}}^c_\beta(t_i, \delta_i, \boldsymbol{x}_i, \tau_i; \boldsymbol{\beta}, \boldsymbol{\theta}) - \boldsymbol{\mathcal{U}}^o_\beta(\boldsymbol{d}_i; \boldsymbol{\beta}, \boldsymbol{\theta})]^T \\ &\qquad \cdot dF(\boldsymbol{t} \mid \boldsymbol{d}; \boldsymbol{\gamma}) \\ &\xrightarrow{p} E\{\boldsymbol{\phi}(t_1, \boldsymbol{x}_1, \tau_1, \delta_1, \boldsymbol{\beta}_o) [\boldsymbol{\mathcal{U}}^c_\beta(t_1, \delta_1, \boldsymbol{x}_1, \tau_1; \boldsymbol{\beta}_o, \boldsymbol{\theta}_o) - \boldsymbol{\mathcal{U}}^o_\beta(\boldsymbol{d}_1; \boldsymbol{\beta}_o, \boldsymbol{\theta}_o)]^T\}, \end{aligned} \tag{3.10}$$

uniformly in $\boldsymbol{\gamma} \in N_{K/\sqrt{n}}(\boldsymbol{\gamma}_o)$. Combining (3.8)–(3.10) we obtain the uniform convergence of $\boldsymbol{A}^n_{11}(\boldsymbol{\gamma})$. Convergence of $\boldsymbol{A}^n_{12}(\boldsymbol{\gamma})$ can be proved in by an argument which is similar to that leading to (3.9) and (3.10). □

The two theorems of Section 2 can now be proved.

PROOF OF THEOREM 2.1. Let $\boldsymbol{T}(\boldsymbol{\gamma}) = [\mathbb{S}_\beta(\boldsymbol{\gamma}), \mathbb{U}^o_\beta(\boldsymbol{\gamma})]^T$. We may assume without loss of generality that $\boldsymbol{A}$ is positive definite by replacing $\boldsymbol{T}(\boldsymbol{\gamma})$ by its product with an orthogonal matrix if necessary. By the usual Taylor expansion,

$$n^{-1/2} \boldsymbol{T}(\boldsymbol{\gamma}) = n^{-1/2} \boldsymbol{T}(\boldsymbol{\gamma}_o) - \boldsymbol{A}_n(\boldsymbol{\gamma}^*) n^{1/2} (\boldsymbol{\gamma} - \boldsymbol{\gamma}_o), \tag{3.11}$$

where $\boldsymbol{\gamma}^*$ is on the line segment joining $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_o$. By Lemma 3.1 and the central limit theorem, $n^{-1/2} \boldsymbol{T}(\boldsymbol{\gamma}_o) = O_p(1)$. Let ϵ be arbitrary. Then there exists $M < \infty$ such that for all large enough n, on a set of probability at least $1-\epsilon/2$, $\|n^{-1/2}\boldsymbol{T}(\boldsymbol{\gamma}_o)\| \le M$. Let $\Delta > 0$ be the minimum eigenvalue of $\boldsymbol{A}$. Choose K large enough that $K^2\Delta/2 - KM > 0$. Then using (3.11) and Lemma 3.2, we find that for all large enough n, on a set of probability at least $1-\epsilon$, $(\boldsymbol{\gamma}-\boldsymbol{\gamma}_o)^T\boldsymbol{T}(\boldsymbol{\gamma}) \ge n^{-1}(K^2\Delta/2 - KM) > 0$, for $\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_o\| = Kn^{-1/2}$. Hence, by a version of Brouwer's fixed point theorem (Lemma 2 of Aitchison and Silvey (1958)), on the same set of probability at least $1-\epsilon$, $\boldsymbol{T}(\boldsymbol{\gamma}) = 0$ has a solution in $N_{K/\sqrt{n}}(\boldsymbol{\gamma}_o)$. □

Proof of Theorem 2.2. Since $\boldsymbol{T}(\hat{\boldsymbol{\gamma}}) = 0$, by Taylor expansion we have

$$A_n(\boldsymbol{\gamma}^*)n^{1/2}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_o) = n^{-1/2}\boldsymbol{T}(\boldsymbol{\gamma}_o) \tag{3.12}$$

where $\boldsymbol{\gamma}^*$ lies on the line segment joining $\boldsymbol{\gamma}$ and $\hat{\boldsymbol{\gamma}}$. Since $\|\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_o\| = O_p(n^{-1/2})$, by Lemma 3.2, $\boldsymbol{A}_n(\boldsymbol{\gamma}^*) \xrightarrow{p} \boldsymbol{A}$. Also by the central limit theorem, $n^{-1/2}\boldsymbol{T}(\boldsymbol{\gamma}_o) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Psi})$. Hence, the theorem follows from (3.12). □

4. A simulation study

The results of Sections 2 and 3 establish the consistency and asymptotic normality of the solution to the score equations (2.5) and (2.6). However, the integrals in equations (2.1), (2.15)–(2.18) will generally be intractable. The Monte-Carlo procedure proposed by Satten *et al.* (1998) must be used to solve the score equations and provide variance estimates for the point estimates. This Monte-Carlo procedure is easily extended to the case of left-truncated data considered here; the only change required is that the partial likelihood score equation for $\boldsymbol{\beta}$ and the maximum likelihood score equation for $\boldsymbol{\theta}$ be replaced by the versions presented here which account for left truncation.

In this section, we present the results of a simulation study to assess the performance of our method when the assumed baseline distribution is mis-specified. For comparison, we include the performance of the fully parametric method, i.e. maximization of the likelihood (2.4) with respect to both $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. Results for the semi-parametric method were obtained using the Monte Carlo methodology described above.

We generated 100 data sets each containing 500 failure times, a binary covariate, and a left-truncation time. In each data set, 250 observations were generated with $x = 0$ and $x = 1$, respectively. The data were generated using $\beta = \ln(2) \approx 0.693$ corresponding to a hazard ratio for covariate values $x = 1$ to $x = 0$ of $\exp(\beta) = 2.0$. The baseline distribution used to generate these data was a log-logistic distribution with shape parameter 4 and location parameter 0.01. For each observation, we also generated 18 renewal times from a renewal process begun at time 0; the increments followed a lognormal distribution with mean 15 and standard error 109.8. We used the fourth renewal time as the left-truncation time. Observations for which the failure time was earlier than the left-truncation time were rejected, and a new pair of failure times and left-truncation times were generated until 250 observations were generated for each covariate value. On average, 93.66 observations in each data set were discarded due to left truncation.

Each simulated data set was then subject to three types of censoring: "light", "heavy", and "current-status". In light censoring, the censoring interval for the i-th ordered failure time overlaps only with the $(i-1)$-th and $(i+1)$-th ordered failure times. Heavy censoring was achieved by using the renewal times described above. If the failure time was before the 18th renewal time, then the first renewal times before and after the failure time was used as the censoring interval; otherwise, the observation was

Table 1. Simulation results for log logistic baseline analyzed with Weibull distribution[a].

Method	β^*	S.E.(β^*)	S.E.($\hat{\beta} - \beta^*$)
		No Censoring	
Exact[b]	0.703 (0.475, 0.919)	0.0945 (0.0892, 0.103)	
		Light Censoring	
Parametric[c]	0.840 (0.642, 1.092)	0.0975 (0.0926, 0.106)	
Semiparametric[d]	0.703 (0.476, 0.919)	0.0943 (0.0889, 0.103)	5.1×10^{-6} (3.2×10^{-6}, 13.9×10^{-6})
		Heavy Censoring	
Parametric	0.820 (0.531, 1.152)	0.131 (0.120, 0.151)	
Semiparametric	0.711 (0.357, 0.993)	0.128 (0.0980, 0.159)	5.8×10^{-4} (3.8×10^{-4}, 7.6×10^{-4})
		Current Status	
Parametric	0.743 (0.380, 1.166)	0.152 (0.133, 0.167)	
Semiparametric	0.709 (0.367, 1.064)	0.161 (0.113, 0.231)	6.4×10^{-4} (4.1×10^{-4}, 10.6×10^{-4})

a. Values shown are means of 100 simulations, and the ranges of the 100 simulations are shown in parentheses. *b.* Exact refers to fitting the left-truncated proportional hazards model to the actual failure times with no censoring. *c.* Parametric refers to a fully parametric model for analyzing left-truncated interval-censored data. *d.* Semiparametric refers to the proposed estimating equation approach for left-truncated interval-censored data using the right-censoring form of the proportional hazards model.

considered to be right censored at the 18th renewal time. For the heavy censored data, approximately 26.8% of the observations were right censored. The average number of intervals each interval overlapped with (a measure of the extent of censoring used by Satten *et al.* (1998)) was 71%. Current status data were created by generating a single independent log-logistic random variate s_i for observation in each data set and choosing as the censoring interval which of the intervals $[\tau_i, \tau_i + s_i)$ or $[\tau_i + s_i, \infty)$ contained the true failure time. Approximately 50.3% of the observations fell into intervals of the form $[\tau_i + s_i, \infty)$.

All analyses using the proposed semiparametric method and the fully parametric method assumed that the baseline distribution was a Weibull distribution. For each censoring type, we estimated the log-hazard ratio β and the standard error of β using the proposed semiparametric method and the fully parametric method. The Monte-Carlo procedure for parameter estimation with our proposed semiparametric method used 400 steps and a block size of 50 (see Satten *et al.* (1998) for details); the standard error of the Monte Carlo approximant to the true parameter estimate was also estimated. For comparison, we also estimated β and its standard error using the exact failure times using the standard proportional hazards model.

Table 1 summarizes the results of our simulations. β^* denotes the approximate solution using the Monte-Carlo procedure. For the light censoring, the proposed semiparametric approach gives point estimates and standard errors that are very close to the values obtained using exact failure times. For the heavily censored and current-status data sets, the mean values of estimated β are also close to the values obtained using the exact failure times, and in both cases the difference between these means is not significant. The mean values of estimated β obtained using the parametric method were significantly different from ln(2) for all three censoring types, while those from the semi-parametric method were not. In addition, the estimated mean square error was smaller for the semi-parametric than for the parametric method for light censored data (0.009 vs. 0.031) and heavy censored data (0.017 vs. 0.033) and nearly equal for current status data (0.0265 vs. 0.0255).

5. Discussion

A natural question regarding our method is, why use the estimating equation (2.5) when a parametric likelihood has already been assumed and is used for estimating nuisance parameters $\boldsymbol{\theta}$. As shown in Section 4, the answer is that the ranking carried out in (2.5) results in an estimating equation that is more robust to mis-specification of the baseline distribution than the parametric maximum likelihood estimator of $\boldsymbol{\beta}$. Similar results were demonstrated in Satten *et al.* (1998), but without left truncation. In practice, we recommend that a flexible spline model be used to model the baseline distribution, so that there is additional confidence that important features of the baseline distribution which might effect estimation of $\boldsymbol{\beta}$ are not missed.

References

Aitchison, J. and Silvey, S. D. (1958). Maximum likelihood estimation of parameters subject to restraints, *Ann. Math. Statist.*, **29**, 813–828.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study, *Ann. Statist.*, **10**, 1100–1120.

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer, New York.

Cox, D. R. (1972). Regression models and life tables (with discussion), *J. Roy. Statist. Soc. Ser. B*, **26**, 187–220.

Datta, S. (1988). Asymptotically optimal Bayes compound and empirical Bayes estimators in exponential families with compact parameter spaces, Ph. D. Dissertation, Department of Statistics and Probability, Michigan State University.

DeLong, D. M., Guirguis, G. H. and So, Y. C. (1994). Efficient computation of subset selection probabilities with application to Cox regression, *Biometrika*, **81**, 607–611.

Diamond, I. D., McDonald, J. W. and Shah, I. H. (1986). Proportional hazards models for current status data: Application to the study of differentials in age at weaning in Pakistan, *Demography*, **23**, 607–620.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data, *Biometrics*, **42**, 845–854.

Grummer-Strawn, L. M. (1993). Regression analysis of current-status data: An application to breast-feeding, *J. Amer. Statist. Assoc.*, **88**, 758–765.

Kalbfleisch, J. D. and Prentice, R. L. (1972). Discussion of paper by D. R. Cox, *J. Roy. Statist. Soc. Ser. B*, **26**, 215–216.

Kalbfleisch, J. D. and Prentice, R. L. (1973). Marginal likelihood based on Cox's regression and life model, *Biometrika*, **60**, 267–278.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model, *J. Amer. Statist. Assoc.*, **89**, 1074–1078.

Peto, R. (1972). Discussion of paper by D. R. Cox, *J. Roy. Statist. Soc. Ser. B*, **26**, 205–207.

Satten, G. A. (1996). Rank-based inference in the proportional hazards model for interval censored data, *Biometrika*, **83**, 355–370.

Satten, G. A., Datta, S. and Williamson, J. M. (1998). Inference based on imputed failure times for the proportional hazards model with interval censored data, *J. Amer. Statist. Assoc.*, **93**, 318–327.

Sinha, D., Tanner, M. A. and Hall, W. J. (1994). Maximization of the marginal likelihood of grouped survival data, *Biometrika*, **81**, 53–60.

Weinstock, H., Sweeney, S., Satten, G. A. and Gwinn, M. (1998). HIV seroincidence and risk factors among patients attending sexually transmitted disease clinics in the United States, 1991 to 1996, *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, **19**, 506–512.