Web-based Supplementary Materials

for “Efficient Inferences for Linear Transformation Models with Doubly Censored Data”

by Sangbum Choi and Xuelin Huang

A Asymptotic Properties

A.1 Asymptotic results and conditions

In this supplementary material, we provide the asymptotic results of the proposed estimator. For ease of presentation, we restate asymptotic results along with required regularity condition.

(a) The function $A_0(t) \in \mathcal{A}$ and $\alpha_0 = A_0'$ exists, where $\mathcal{A}$ is the collection of strictly increasing and continuously differentiable functions satisfying $0 < \inf_{t\in[0,\tau]} \alpha_0(t) \le \sup_{t\in[0,\tau]} \alpha_0(t) < \infty$. The parameter θ_0 lies in the interior of a compact set Θ in R^d.

(b) The covariate $Z \in R^d$ has a bounded support and for any $b \neq b_0$, the probability $P(b^T Z \neq b_0^T Z) > 0$.

(c) The underlying distribution of failure time T is continuous, and τ denotes the time at the end of the study that satisfies $P\{Y(\tau) = 1|Z\} > 0$ and $P(X \ge \tau, \delta = 0|Z) > 0$ with probability one. Also, there exists $0 < \sigma < \tau$ such that $P(X \ge \sigma, \delta = 1|Z) = 1$ and $0 < \eta_0 < A_0(\sigma) < A_0(\tau) < \infty$ for some η_0.

(d) For any positive constant a_0, $\limsup_{x\to\infty}\{G(a_0 x)\}^{-1} \log\{x \sup_{y\le x} g(y)\} = 0$.

Under these conditions, we claim asymptotic consistency and normality for the proposed estimator $(\hat\theta, \hat A)$ as follows.

Theorem A.1. *Suppose that assumptions (a)–(d) hold. Then,* $\|\hat\theta - \theta_0\| \to 0$ *and* $\sup_{t\in[0,\tau]} |\hat A(t) - A_0(t)| \to 0$ *almost surely.*

Theorem A.2. *Suppose that assumptions (a)–(d) hold and that* $n_1/n \to \alpha_1$ *as* $n \to \infty$, *with* $0 < \alpha_1 \le 1$. *Then*

$$n^{1/2}(\hat\theta - \theta_0) \;\to\; N(0, \Sigma^{-1}) \qquad \text{and} \qquad n^{1/2}(\hat A - A_0) \;\to\; \mathcal{G}$$

in distribution, where Σ *is the information matrix with respect to* θ *and* $\mathcal{G}$ *is a Gaussian process. In addition,* $\hat\theta$ *is semiparametrically efficient.*

A.1.1 Proof of Theorem 1

The proofs of the asymptotic properties of the nonparametric maximum likelihood estimator essentially follow the steps outlined by Zeng and Lin (2007). Let $l^\infty[0,\tau]$ denote the space of functions with bounded total variation in $[0,\tau]$ under the supremum norm $\|\cdot\|_{l^\infty[0,\tau]}$, and $\|\mu\|_{BV[0,\tau]}$ be the total variation of $\mu(t)$ in $[0,\tau]$. Also, define $\mathcal{H}=\{\mu(t):\|\mu\|_{BV[0,\tau]}\le 1\}$. Then we can regard $\hat{A}(\cdot)$ as a bounded linear function in $l^\infty(\mathcal{H})$,and $\{\hat{\theta}-\theta_0,\hat{A}(\cdot)-A_0(\cdot)\}$ as a random element in the metric space $R^d\times l^\infty(\mathcal{H})$.

First, observe that $l(\theta,A)\le l_1(\theta,A)$, which is bounded by

$$O(1)+\sum_{i=1}^{n_1}\big[\log\{dA(\tilde{T}_i)\sup_{t\le A(\tilde{T}_i)e^M} g(t)\}-G\{e^{-M}A(\tilde{T}_i)\}\big], \tag{1}$$

where $O(1)$ is some positive constant and $M>0$ is a constant satisfying

$$\exp(-M)\le \inf_{t,\theta,Z}[\exp\{\theta^TZ_i(t)\}]\le \sup_{t,\theta,Z}[\exp\{\theta^TZ_i(t)\}]\le \exp(M),$$

such that M exists by conditions (a) and (b). By condition (d), expression (1) would diverge to $-\infty$ if $dA(\tilde{T}_i)$ is infinite for some $\tilde{T}_i$. Thus, the jump sizes of A must be finite.

Let $\tilde{A}(\cdot)=\hat{A}(\cdot)/\hat{A}(\tau)$. Note that

$$\begin{aligned}&0\le n^{-1}\{l(\hat{\theta},\hat{A})-l(\hat{\theta},\tilde{A})\}\\&\quad\le O(1)+n^{-1}\Bigg[\sum_{i=1}^{n_1}\log\{\hat{A}(\tau)\sup_{t\le\hat{A}(\tau)e^M}g(t)\}-\sum_{i=n_1+1}^{n}(1-\delta_i)I(\tilde{T}_i=\tau)G\{e^{-M}\hat{A}(\tau)\}\Bigg].\end{aligned}$$

By conditions (c) and (d), the right-hand side of the above expression would be negative if $\hat{A}(\tau)$ diverges to $-\infty$, which is a contradiction of the definition of $(\hat{\theta},\hat{A})$. Thus, $\hat{A}$ must be bounded almost surely over $[0,\tau]$. By Helly's selection theorem, there is thus a convergent subsequence such that $\hat{A}_n\to A_*$ and $\hat{\theta}_n\to\theta_*$.

By taking a derivative of $l(\theta,A)$ with respect to the jump size dA, we obtain

$$\hat{A}(t)=n^{-1}\int_0^t\frac{\sum_{i=1}^n dN_i(s)}{|\omega_n(s;\hat{\theta}_n,\hat{A}_n)|}, \tag{2}$$

where $N_i(s)=I(\tilde{T}_i\le s,\varepsilon_i=1)$ is the counting process that counts the number of events observed by time t on the ith subject, with $\varepsilon_i=1$ or 0 denoting whether the subjectÕs data belongs to the exact data group or not. Here,

$$\omega_n(s;\hat\theta_n,\hat A_n) = n^{-1}\sum_{i=1}^{n_1} Y_i(s)\exp(\hat\theta_n^T Z_i)\{g_i(t_i;\hat\theta_n,\hat A_n) - \psi_i(t_i;\hat\theta_n,\hat A_n)\}$$
$$+ n^{-1}\sum_{i=n_1+1}^{n} Y_i(s)\exp(\hat\theta_n^T Z_i)\{(1-\delta_i) - \delta_i\Lambda_i(t_i;\hat\theta_n,\hat A_n)\}g_i(t_i;\hat\theta_n,\hat A_n).$$

By the Glivenko-Cantelli theorem, $\omega_n(t;\hat\theta_n,\hat A_n)$ converges uniformly to a continuously differentiable function $\omega_*(t;\theta_*,A_*)$. Note that for $t \le \tau_0$,

$$|\omega_n(t;\hat\theta_n,\hat A_n)| \le (e^M/n)\sum_{i=1}^{n} Y_i(t)\sup_{t\in[0,\tau_0]}\Big|g_i(t_i;\hat\theta_n,\hat A_n) + |\psi_i(t_i;\hat\theta_n,\hat A_n)|\Big| < \infty;$$

thus there exists $\eta > 0$ such that $\hat A(\tau_0) > \eta$ for all sufficiently large n. Also, by condition (c), $X_i \ge \tau_0$ for censored observations, and thus $\omega_n(t;\hat\theta_n,\hat A_n) \ge \omega_n(\tau_0;\hat\theta_n,\hat A_n)$ for $t \le \tau_0$. For $t \ge \tau_0$, $\omega_n(t;\hat\theta_n,\hat A_n) \ge [e^{-M}g\{e^{-M}\hat A(\tau_0)\}]\{n^{-1}\sum_{i=n_1+1}^{n}(1-\delta_i)Y_i(t)\}$. Thus, it can be argued that $\min_{t\in[0,\tau]}|\omega_*(t;\theta_*,A_*)| = \min_{t\in[\tau_0,\tau]}|\omega_*(t;\theta_*,A_*)|$ must be bounded above from zero.

Define

$$\check A(t) = n^{-1}\int_0^t \frac{\sum_{i=1}^n dN_i(s)}{|\omega_n(s;\theta_0,A_0)|}. \tag{3}$$

Then, $\check A$ converges to A_0 uniformly by the Glivenko-Cantelli theorem. It follows from (2), (3) and the strict positivity of $|\omega_n|$ that $\hat A(t)$ is absolutely continuous with respect to $\check A(t)$ and that $d\hat A/d\check A$ converges uniformly to some bounded function. Note that $n^{-1}\{l(\hat\theta,\hat A) - l(\theta_0,\check A)\} \ge 0$. This implies that the Kullback-Leibler distance between the density that is indexed by (θ_*,A_*) and the true density would be negative. Therefore, almost surely

$$\begin{aligned}
&\int_0^\tau [\log\{dA_*(t)\} + {\theta_*}^T Z_i(t) + \log\{g_i(t;\theta_*,A_*)\} - G_i(t;\theta_*,A_*)]dN_i(t)\\
&\qquad + (1-\varepsilon_i)[\delta_i\log[1-\exp\{-G_i(X_i;\theta_*,A_*)\}] - (1-\delta_i)G_i(X_i;\theta_*,A_*)]\\
&= \int_0^\tau [\log\{dA_0(t)\} + {\theta_0}^T Z_i(t) + \log\{g_i(t;\theta_0,A_0)\} - G_i(t;\theta_0,A_0)]dN_i(t)\\
&\qquad + (1-\varepsilon_i)[\delta_i\log[1-\exp\{-G_i(X_i;\theta_0,A_0)\}] - (1-\delta_i)G_i(X_i;\theta_0,A_0)].
\end{aligned}$$

By condition (c), the above equality holds in the special case where $Y_i(t) = 1$ for $t \in [0,\tau]$. For observations with $\varepsilon_i = \delta_i = 0$, consider the equality under $N_i(\tau) = 0$ and $X_i \ge \tau$ and the equality under $N_i(\tau-) = 0$ and $N_i(\tau) = 1$. The difference between these two equalities and the identifiability condition (b) imply that $\theta_* = \theta_0$ and $A_* = A_0$. We have thus established the consistency property. By the continuity of $A_0(\cdot)$, we have the uniform convergence of $\hat A(t)$ to $A(t)$ in $t \in [0,\tau]$.

A.1.2 Proof of Theorem 2

As the model involves an infinite-dimensional parameter, we need to derive semiparametric analogues of the parametric score equation and the information matrix. The overall idea is to utilize these operators to establish the asymptotic normality of $n^{1/2}(\hat{\vartheta}-\vartheta_0)$ by checking the conditions of Theorem 3.3.1 of van der Vaart and Weller (1996).

Likelihood equations for ϑ can be obtained by inserting a smooth family ϑ_s that approaches ϑ as $s \to 0$. In particular, given a real vector $h_1 \in R^d$, a bounded and measurable function $h_2 \in \mathcal{H}$ and every sufficiently small number $|s|$, we can consider a parametric submodel $s \mapsto \vartheta_s \equiv \vartheta + s(h_1, \int h_2 dA)$. Taking a derivative with respect to s of this submodel leads to the likelihood equation $\mathcal{S}_n(\hat{\vartheta})(h) = 0$ for any $h = (h_1, h_2)$, where the score operator $\mathcal{S}_n(\vartheta)(h)$ is defined by

$$\mathcal{S}_n(\vartheta)(h) = \left.\frac{(\partial/\partial s) l(\vartheta_s)}{n}\right|_{s=0} = h_1^T \mathcal{S}_{1_n}(\vartheta) + \mathcal{S}_{2_n}(\vartheta)(h_2),$$

where

$$\begin{aligned}
\mathcal{S}_{1_n}(\vartheta) &= \frac{n_1}{n} \frac{\sum_{i=1}^{n_1} \int_0^\tau [Z_i(t) - \{g_i(t;\theta,A) - \psi_i(t;\theta,A)\} \int_0^t Y_i(s) Z_i(s) e^{\theta^T Z_i(s)} dA(s)] dN_i(t)}{n_1} \\
&\quad + \frac{n_2}{n} \frac{\sum_{i=n_1+1}^{n} [\{\delta_i \Lambda_i(X_i;\theta,A) - (1-\delta_i)\} g_i(X_i;\theta,A) \int_0^{X_i} Y_i(s) Z_i(s) e^{\theta^T Z_i(s)} dA(s)]}{n_2} \\
&\equiv (n_1/n) P_{n_1} \dot{l}_{1\theta}(\theta,A) + (n_2/n) P_{n_2} \dot{l}_{2\theta}(\theta,A),
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{S}_{2_n}(\vartheta)(h_2) &= \frac{n_1}{n} \frac{\sum_{i=1}^{n_1} \int_0^\tau [h_2(t) - \{g_i(t;\theta,A) - \psi_i(t;\theta,A)\} \int_0^t Y_i(s) h_2(s) e^{\theta^T Z_i(s)} dA(s)] dN_i(t)}{n_1} \\
&\quad + \frac{n_2}{n} \frac{\sum_{i=n_1+1}^{n} [\{\delta_i \Lambda_i(X_i;\theta,A) - (1-\delta_i)\} g_i(X_i;\theta,A) \int_0^{X_i} Y_i(s) h_2(s) e^{\theta^T Z_i(s)} dA(s)]}{n_2} \\
&\equiv (n_1/n) P_{n_1} \dot{l}_{1A}(\theta,A)(h_2) + (n_2/n) P_{n_2} \dot{l}_{2A}(\theta,A)(h_2).
\end{aligned}$$

Here, $P_n = (P_{n_1}, P_{n_2})$ denotes the empirical joint measure of the exact observations and the case-1 interval-censored observations. Let $P_0 = (P_{10}, P_{20})$ be the corresponding true distribution of an exact observation and a censored observation.

Note that $\mathcal{S}_n(\vartheta) = \{\mathcal{S}_{1_n}(\vartheta), \mathcal{S}_{2_n}(\vartheta)\}$ is the element of $R^d \times l^\infty(\mathcal{H})$. The asymptotic version of

$\mathcal{S}_n$, denoted by $\mathcal{S}(\vartheta) = \{\mathcal{S}_1(\vartheta), \mathcal{S}_2(\vartheta)\}$ of $R^d \times l^\infty(\mathcal{H})$, is obtained by replacing P_n with P_0 in the definition of $\mathcal{S}_n$ and given by

$$\mathcal{S}_1(\vartheta) = \alpha_1 P_{10}\dot{l}_{1\theta} + \alpha_2 P_{20}\dot{l}_{2\theta}$$

$$\begin{aligned}\mathcal{S}_2(\vartheta) &= \alpha_1 P_{10}\dot{l}_{1A}(h_2) + \alpha_2 P_{20}\dot{l}_{2A}(h_2)\\ &= \alpha_1 \int \dot{l}_{1A}(h_2)dP_{10} + \alpha_2 \int \dot{l}_{2A}(h_2)dP_{20}.\end{aligned}$$

Clearly, $\mathcal{S}_n(\hat{\vartheta}) = 0$ and $\mathcal{S}(\vartheta_0) = 0$. It can be shown that the elements in $\mathcal{S}(\vartheta)$ are P_0-Donsker classes. Also, by conditions (a), (b) and the Donsker theorem, $n^{1/2}\{(\mathcal{S}_n - \mathcal{S})(\hat{\vartheta}) - (\mathcal{S}_n - \mathcal{S})(\vartheta_0)\} = o_p(1)$ in the metric space $R^d \times l^\infty(\mathcal{H})$. Furthermore, $n^{1/2}\{\mathcal{S}_n(\vartheta_0) - \mathcal{S}(\vartheta_0)\} = n^{1/2}\mathcal{S}_n(\vartheta_0)$ weakly converges to a tight Gaussian process. The Fréchet-differentiability of $\mathcal{S}$ at $\vartheta = \vartheta_0$ can be directly checked under the assumed conditions.

We now verify that the inverse of the derivative $\dot{\mathcal{S}}$ exists and is continuous at ϑ_0. Let

$$\ddot{l}_{i\theta\theta}(\vartheta) = \frac{\partial}{\partial\theta}\dot{l}_{i\theta}(\theta, A) \qquad \text{and} \qquad \ddot{l}_{i\theta A}(\vartheta)(h_2) = \frac{\partial}{\partial s}\dot{l}_{i\theta}(\theta, A_s)\bigg|_{s=0},$$

$$\ddot{l}_{iA\theta}(\vartheta)(h_2) = \frac{\partial}{\partial\theta}\dot{l}_{iA}(\theta, A)(h_2) \qquad \text{and} \qquad \ddot{l}_{iAA}(\vartheta)(h_2, h_2') = \frac{\partial}{\partial s}\dot{l}_{iA}(\theta, A_{2s})(h_2)\bigg|_{s=0},$$

where $i = 1, 2$, h_2, $h_2' \in \mathcal{H}$ and $A_{2s} = \int(1 + sh_2')dA$. Note that the derivative of $\mathcal{S}$ at ϑ_0, denoted by $\dot{\mathcal{S}}_{\vartheta_0}$, is given by the map

$$(\theta - \theta_0, A - A_0) \mapsto \begin{pmatrix}\dot{\mathcal{S}}_{11} & \dot{\mathcal{S}}_{12}\\ \dot{\mathcal{S}}_{21} & \dot{\mathcal{S}}_{22}\end{pmatrix}\begin{pmatrix}\theta - \theta_0\\ A - A_0\end{pmatrix}(h) = \begin{pmatrix}h_1^T\dot{\mathcal{S}}_{11}(\theta - \theta_0) + h_1^T\dot{\mathcal{S}}_{12}(A - A_0)\\ \dot{\mathcal{S}}_{21}(\theta - \theta_0)(h_2) + \dot{\mathcal{S}}_{22}(A - A_0)(h_2)\end{pmatrix}$$

where

$$\begin{aligned}\dot{\mathcal{S}}_{11}(\theta - \theta_0) &= (\alpha_1 P_{10}\ddot{l}_{1\theta\theta} + \alpha_2 P_{20}\ddot{l}_{2\theta\theta})(\theta - \theta_0) = B_0(\theta - \theta_0),\\ \dot{\mathcal{S}}_{12}(A - A_0) &= \alpha_1 \int \ddot{l}_{1\theta A}(A - A_0)dP_{10} + \alpha_2 \int \ddot{l}_{2\theta A}(A - A_0)dP_{20} = \mathcal{C}_0^*(A - A_0),\\ \dot{\mathcal{S}}_{21}(\theta - \theta_0)(h_2) &= (\alpha_1 P_{10}\ddot{l}_{1A\theta} + \alpha_2 P_{20}\ddot{l}_{2A\theta})(\theta - \theta_0)(h_2) = \mathcal{C}_0(\theta - \theta_0)(\int h_2 dA_0),\\ \dot{\mathcal{S}}_{22}(A - A_0)(h_2) &= \alpha_1 \int \ddot{l}_{1AA}(A - A_0)(h_2)dP_{10} + \alpha_2 \int \ddot{l}_{2AA}(A - A_0)(h_2)dP_{20}\\ &= \int \{-a_0(t)I + \mathcal{D}_0\}(h_2)d(A - A_0),\end{aligned}$$

where $a_0(t) > 0$, I is the identity operator, $\mathcal{C}_0$ and $\mathcal{D}_0$ are both linear operators, $\mathcal{C}_0^*$ is the dual operator of $\mathcal{C}_0$, and B_0 is the Fisher information matrix for θ in the situation when A_0 is known.

Then the information operator $\Omega = (\Omega_1, \Omega_2)$ satisfies

$$\dot{\mathcal{S}}_{\vartheta_0}(\hat{\theta} - \theta_0, \hat{A} - A_0)(h) = \Omega_1(h)^T(\hat{\theta} - \theta_0) + \int \Omega_2(h) d(\hat{A} - A_0),$$

where Ω_1 is a linear map from $R^d \times \mathcal{H}$ to R^d and Ω_2 is a linear map from $R^d \times \mathcal{H}$ to $l^\infty[0, \tau]$.

Essentially, the invertibility of $\dot{\mathcal{S}}_{\vartheta_0}$ can be obtained by checking that the information operator Ω is one-to-one and Òonto.Ó The "onto" property is easy to verify and has been omitted here. For the "one-to-one" property, fix an $h = (h_1, h_2) \in R^d \times \mathcal{H}$, for which $\Omega(\vartheta_0)(h) = 0$ and we wish to show $h = 0$. Define the one-dimensional submodel $s \mapsto \vartheta_{0s} = \vartheta_0 + s(h_1, \int h_2 dA)$. Note that $\Omega(\vartheta_0)(h) = 0$ implies that $P_0\{\partial^2 l(\vartheta_{0s})/(\partial s)^2|_{s=0}\} = -P_0\{\mathcal{S}(\vartheta_0)(h)\}^2 = 0$, which implies $\mathcal{S}(\vartheta_0)(h) = 0$. It can be seen that it holds almost surely as long as

$$\begin{aligned} h_1^T Z(t) + h_2(t) = &-[\psi(t; \theta_0, A_0) - \delta g(t; \theta_0, A_0)\{1 + \Lambda(t; \theta_0, A_0)\}] \\ &\times \int_0^t Y(s)\{h_1^T Z(s) + h_2(t)\} \exp\{\theta_0^T Z(s)\} dA(s) \end{aligned}$$

for all $t \in [0, \tau]$ with respect to the conditional distribution of (T, X) given Z, where we let $\delta = 1$ for left censoring, otherwise, 0. This equation is a homogeneous integral equation for the function $h_1^T Z(t) + h_2(t)$, which leads to $h_1^T Z(t) + h_2(t) = 0$. Then identifiability condition (b) implies that $h = (h_1, h_2) = 0$, which ensures the invertibility of $\dot{\mathcal{S}}$.

It now follows from Theorem 3.3.1 of van der Vaart and Wellner (1996) that, in the metric space $R^d \times l^\infty(\mathcal{H})$, $\sqrt{n}(\hat{\theta} - \theta_0, \hat{A} - A_0)$ weakly converges to some Gaussian process. Furthermore,

$$\begin{aligned} n^{1/2}\dot{\mathcal{S}}_{\vartheta_0}(\hat{\theta} - \theta_0, \hat{A} - A_0)(h) &= n^{1/2}\Omega_1(h)^T(\hat{\theta} - \theta_0) + n^{1/2}\int \Omega_2(h) d(\hat{A} - A_0) \\ &= n^{1/2}\{\mathcal{S}_n(\vartheta_0)(h) - \mathcal{S}(\vartheta_0)(h)\} + o_p(1). \end{aligned}$$

Since the invertibility of $\dot{\mathcal{S}}_{\vartheta_0}$ implies the invertibility of the map Ω, we have that

$$n^{1/2}h_1^T(\hat{\theta} - \theta_0) + n^{1/2}\int h_2(u) d\{\hat{A}(u) - A_0(u)\} = n^{1/2}\{\mathcal{S}_n(\vartheta_0)(\Omega^{-1}(h)) - \mathcal{S}(\vartheta_0)(\Omega^{-1}(h))\} + o_p(1),$$

uniformly in h as $n \to \infty$. Setting $h_2 = 0$ leads to $n^{1/2}h_1^T(\hat{\theta} - \theta_0) = n^{1/2}\{\mathcal{S}_n(\vartheta_0)(\Omega^{-1}(h_1, 0)) - \mathcal{S}(\vartheta_0)(\Omega^{-1}(h_1, 0))\} + o_p(1)$, and by the central limit theorem, $n^{1/2}h_1^T(\hat{\theta} - \theta_0)$ is asymptotically normal with mean zero and variance $P_0\{\mathcal{S}(\vartheta_0)(\Omega^{-1}(h_1, 0))\}^2 = \{\Omega_1^{-1}(h_1, 0)\}^T h_1 = h_1^T \Sigma h_1$, where $\Sigma = \{\Omega_1^{-1}(e_1, 0), ..., \Omega_1^{-1}(e_d, 0)\}$, with e_k denoting the d-dimensional vector with the kth component equal to 1 and elsewhere 0. Also, letting $h_1 = 0$ and $h_2 = I(u \leq t)$ in the above equation gives $n^{1/2}\{\hat{A}(t) - A_0(t)\} = n^{1/2}\{\mathcal{S}_n(\vartheta_0)(\Omega^{-1}(0, I(u \leq t))) - \mathcal{S}(\vartheta_0)(\Omega^{-1}(h_1, 0, I(u \leq t)))\} + o_p(1)$, which has an asymptotically normal distribution with mean zero and variance $\int_0^t \Omega_2^{-1}(0, I(u \leq t))dA_0(u)$. Moreover, the above argument shows that $\hat{\theta}$ is an asymptotically linear estimator for θ_0 and that its influence function lies on the space spanned by the score function, implying that $\hat{\theta}$ is semiparametrically efficient. The consistency of the variance estimators can be justified along the lines of Zeng and Lin (2007), by showing that the linear operator constructed from the negative Hessian matrix of the log-likelihood function approximates the information operator. The detailed derivation is omitted here.

References

Zeng, D. and Lin, D. Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of Royal Statistical Society, Series B* **69**, 507–564.

van der Vaart, A. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* New York: Springer.