A Supplementary Materials

This supplementary document contains the detailed proofs for the asymptotic properties of the proposed PLAC estimator. The proofs are provided under the following regularity conditions, which should hold in most practical scenarios, such as in the RRI-CKD study demonstrated in Section 4.

(C1) The true regression coefficients vector $\boldsymbol{\beta}_0$ lies in the interior of a compact set $B \subset \mathbb{R}^p$. The true cumulative baseline hazard function $\Lambda_0(t)$ is continuously differentiable and strictly increasing on $[0, \tau]$, and satisfies $\Lambda_0(0) = 0$.

(C2) The covariates vector $\mathbf{Z}$ is bounded almost surely. If there exist a deterministic function $b_0(t)$ and a vector $b \in \mathbb{R}^p$, such that $b_0(t) + b^T\mathbf{Z} = 0$ with probability one, then $b_0(t) = 0$ and $b = 0$.

(C3) With probability one, there exists a constant $\delta_1 > 0$ such that $\Pr(A^* < T^* \leqslant A^* + C \,|\, \mathbf{Z},\, A^*,\, C) > \delta_1$, $\Pr(A + C \geqslant \tau \,|\, \mathbf{Z}) > \delta_1$, and that $\Pr(T \geqslant \tau \,|\, \mathbf{Z}) > \delta_1$.

(C4) With probability one, there exists a constant $\delta_2 > 0$ such that $\Pr(A^* \geqslant T^* | \mathbf{Z}) > \delta_2$.

(C5) Let $b \in \mathbb{R}^p$ and h be a function with bounded total variation on $[0, \tau]$, then the information operator corresponding to the conditional likelihood evaluated at the true parameters $(\boldsymbol{\beta}_0,\, \Lambda_0)$,

$$J_0^C(b,\, h) = \left(\lim_{n \to \infty} \frac{\partial U^C(\boldsymbol{\beta}, \Lambda)}{\partial(\boldsymbol{\beta}, \Lambda)} \bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0,\, \Lambda=\Lambda_0} \right) (b,\, h)$$

is invertible.

Conditions (C1)-(C3) and (C5) are standard assumptions for the Cox model under left-trucation, which are necessary to prove the identifiability of the parameters as well as the existence and uniqueness of the PLAC estimator. The continuity of $\Lambda_0(t)$ facilitates the uniform convergence proof of $\hat{\Lambda}(t)$, and the strictly monotonicity suggests that events can happen at any time during the follow-up. The boundness assumption in (C2) is important for the uniform convergence proofs for the function classes involved, and the second assumption ensures the covariates are not degenerate and that the parameters are identifiable. The first and second assumptions of (C3) imply that for any covariate pattern, subjects with the events happen between 0 and τ have a positive chance to be observed, i.e., not all of them are censored or truncated; whereas the third assumption implies that some subjects could be still at risk by the end of the study. Putting altogether, (C3) ensures the denominator of $\mathcal{L}_C$ is bounded away from zero. Condition (C4) is necessary for $\mathcal{L}_P$ to be non-degenerate so that we can attain efficiency gains using the PLAC estimator beyond the conditional approach estimator. Condition (C5), which is used to show the root of the composite score equations is unique, is adapted from the classic weak convergence proof for the Cox model (see Andersen et al., 1993, Condition VII2.1(e)).

We use Ω to denote the set of all possible observations. For convenience, we adopt the de Finetti's linear functional notations (Pollard, 2002), where $\mathbb{P}_n$ denotes the empirical measure of the observations $\mathcal{O}_i$, $i = 1, \ldots, n$, P_0 denotes the true probability measure on Ω, and $\mathbb{U}_{n,2}$ is the empirical measure of pairs $(\mathcal{O}_i, \mathcal{O}_j)$ such that $1 \leqslant i < j \leqslant n$.

A.1 Identifiability of $(\boldsymbol{\beta}_0, \Lambda_0)$

Lemma 1. *Under Conditions (C1)-(C3), both $\boldsymbol{\beta}_0$ and Λ_0 are identifiable. Specifically, if there exist parameters $(\boldsymbol{\beta}, \Lambda)$ such that Λ is absolutely continuous with respect to Λ_0, $\ell_n^C(\boldsymbol{\beta}, \Lambda) = \ell_n^C(\boldsymbol{\beta}_0, \Lambda_0)$ and that $\ell_n^P(\boldsymbol{\beta}, \Lambda) = \ell_n^P(\boldsymbol{\beta}_0, \Lambda_0)$ with probability one under P_0, then we have $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ and $\Lambda = \Lambda_0$, where ℓ_n^C and ℓ_n^P are the conditional and pairwise log-likelihood functions, respectively.*

Proof. Denote the density and distribution functions of $\mathbf{Z}$ as $f_{\mathbf{Z}}$ and $F_{\mathbf{Z}}$, respectively. First, suppose we have $\ell_n^C(\boldsymbol{\beta}, \Lambda) = \ell_n^C(\boldsymbol{\beta}_0, \Lambda_0)$, i.e.,

$$\begin{aligned} & \int_0^\tau \left(\log \lambda(s) + \mathbf{Z}^T \boldsymbol{\beta}\right) dN(s) - \int_0^\tau Y(s)\lambda(s) e^{\mathbf{Z}^T \boldsymbol{\beta}} ds \\ = & \int_0^\tau \left(\log \lambda_0(s) + \mathbf{Z}^T \boldsymbol{\beta}_0\right) dN(s) - \int_0^\tau Y(s)\lambda_0(s) e^{\mathbf{Z}^T \boldsymbol{\beta}_0} ds \end{aligned}$$

holds almost everywhere under P_0. By Conditions (C1), (C3) and the fact that the support of A^* includes zero, outside a set with zero probability, for any $0 \leqslant a < u \leqslant \tau$ and any $\mathbf{z}$ in the bounded support of $f_{\mathbf{Z}}$, the equality holds for the case with $N(u-) = 0$ and $N(u) = 1$. Taking anti-log transformation on both sides and rearranging the equation, we have

$$\frac{\lambda(u)}{\lambda_0(u)} = e^{\mathbf{z}^T(\boldsymbol{\beta}_0 - \boldsymbol{\beta})} \frac{\exp\left\{(\Lambda(u) - \Lambda(a)) e^{\mathbf{z}^T \boldsymbol{\beta}}\right\}}{\exp\left\{(\Lambda_0(u) - \Lambda(a)) e^{\mathbf{z}^T \boldsymbol{\beta}_0}\right\}}.$$

Let $a \to 0$, we get

$$\frac{\lambda(u)}{\lambda_0(u)} = e^{\mathbf{z}^T(\boldsymbol{\beta}_0 - \boldsymbol{\beta})} \frac{\exp\{\Lambda(u) e^{\mathbf{z}^T \boldsymbol{\beta}}\}}{\exp\{\Lambda_0(u) e^{\mathbf{z}^T \boldsymbol{\beta}_0}\}}. \tag{A.1}$$

Moreover, there exists a sequence $\{u_k\}_{k \geqslant 1}$ converging to 0 from above such that (A.1) holds for almost every $\mathbf{z} \in D_k = \{\mathbf{z} : \Pr(C \geqslant u_k | \mathbf{Z} = \mathbf{z}) > 0\}$. Note that $D_k \uparrow D = \{\mathbf{z} : \Pr(C \geqslant 0 \,|\, \mathbf{Z} = \mathbf{z}) > 0\}$, with $\Pr(D) = 1$ under $F_{\mathbf{Z}}$. As $k \to \infty$, the limit for the right-hand side evaluated at u_k, by the continuity of Λ_0 and absolute continuity of Λ with respect to Λ_0, is $e^{\mathbf{z}^T(\boldsymbol{\beta}_0 - \boldsymbol{\beta})}$. The left-hand side must also converge, but the limit is independent of $\mathbf{z}$. Hence, the variable $\mathbf{z}^T(\boldsymbol{\beta}_0 - \boldsymbol{\beta})$ is degenerate, which by (C2) implies $\boldsymbol{\beta} = \boldsymbol{\beta}_0$.

Substituting $\boldsymbol{\beta}$ with $\boldsymbol{\beta}_0$ in (A.1), on a non-empty set of $\mathbf{z}$ such that $A + C \geqslant \tau$, we have the equality

$$-\lambda(u) e^{\mathbf{z}^T \boldsymbol{\beta}_0} \exp\{-\Lambda(u) e^{\mathbf{z}^T \boldsymbol{\beta}_0}\} = -\lambda_0(u) e^{\mathbf{z}^T \boldsymbol{\beta}_0} \exp\{-\Lambda_0(u) e^{\mathbf{z}^T \boldsymbol{\beta}_0}\}$$

holds for almost every $u \in [0, \tau]$. Integrating both sides from 0 to t yields

$$\int_0^t d\exp\{-\Lambda(u)e^{\mathbf{z}^T\boldsymbol{\beta}_0}\} = \int_0^t d\exp\{-\Lambda_0(u)e^{\mathbf{z}^T\boldsymbol{\beta}_0}\}$$

for all $t \leqslant \tau$. Therefore, we have $\Lambda(t) = \Lambda_0(t)$ for all $t \in [0, \tau]$.

For the pairwise likelihood, outside a set with zero probability, by Condition (C1), for almost every pair of $(\mathbf{z}_1, \mathbf{z}_2)$ in the support of f_Z and every pair of (A_1, A_2) such that $\Lambda_0(A_1) - \Lambda_0(A_2) > 0$, we have

$$\frac{\Lambda(A_1) - \Lambda(A_2)}{\Lambda_0(A_1) - \Lambda_0(A_2)} = \frac{e^{\mathbf{z}_1^T\boldsymbol{\beta}_0} - e^{\mathbf{z}_2^T\boldsymbol{\beta}_0}}{e^{\mathbf{z}_1^T\boldsymbol{\beta}} - e^{\mathbf{z}_2^T\boldsymbol{\beta}}}. \tag{A.2}$$

Thus (A.2) implies that the ratios on both sides are the same constant c. By Condition (C1), the left-hand side then gives $\Lambda(t) = c\Lambda_0(t)$ for t in the support of A. On the other hand, the right-hand side is degenerate if it equals c when $(\mathbf{z}_1, \mathbf{z}_2)$ vary, this again implies $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ thus $c = 1$. Note that we need the support of A includes the follow-up period to identify Λ_0 solely from the pairwise likelihood.

It is worth noting that $\Lambda_0(t)$ is not identifiable for $0 < t < w_1$ (Wang et al., 1993). However, since the support of A^* includes zero, by (C3), w_1 is usually close to zero; thus, the identifiability issue is less likely to occur. □

A.2 Consistency of $(\hat{\boldsymbol{\beta}}, \hat{\Lambda})$

The PLAC estimator falls in the category of Z-estimators. To follow the consistency proof of the general Z-estimators, a complication brought by the pairwise structure is to show the uniform convergence of the involved bivariate function classes. We tackle this difficulty through bounding the bracketing numbers (entropies) of these function classes using the U-processes theory (see De la Peña and Giné, 1999, Chapter 5). For $k = 0, 1, 2$, the function classes $\{(\mathbf{z}_1, \mathbf{z}_2) \mapsto \mathbf{z}_1^{\otimes k} e^{\mathbf{z}_1^T\boldsymbol{\beta}} - \mathbf{z}_2^{\otimes k} e^{\mathbf{z}_2^T\boldsymbol{\beta}} : \mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^p;\ \boldsymbol{\beta} \in B\}$ are Euclidean (Nolan and

Pollard, 1987); thus, their bracketing numbers in $L_1(P^2)$ are finite, where $P^2 \equiv P \otimes P$, and P is any probability measure. Bounds for classes only consisting of indicator functions can be shown using the VC theory (see De la Peña and Giné, 1999, Section 5.2). Denoting the class of cumulative baseline hazard functions satisfying (C1) as $\mathcal{H}_\Lambda$, then

Lemma 2. *The bivariate function class* $\mathcal{H}_\Lambda^D = \{(s,t) \mapsto \Lambda(s) - \Lambda(t) : s, t \in [0,\tau];\ \Lambda \in \mathcal{H}_\Lambda\}$ *has finite bracketing numbers in* $L_1(P^2)$ *for all* $\varepsilon > 0$.

Proof. To avoid technicality, we assume all bivariate function classes involved in this and the following proofs are measurable (see De la Peña and Giné, 1999, Section 3.5). Theorem 2.7.5 of van der Vaart and Wellner (1996) indicates that for a fixed $\varepsilon > 0$, there exists a constant K_1 such that the bracketing entropy

$$\log N_{[]}(\varepsilon, \mathcal{H}_\Lambda, L_1(P)) < \frac{K_1}{\varepsilon} < \infty$$

for any probability measure P. For a given $\Lambda \in \mathcal{H}_\Lambda$, suppose an ε-bracket containing it in $L_1(P)$ is (Λ_l, Λ_u); thus, we have $\Lambda_l(t) < \Lambda(t) < \Lambda_u(t)$, $\forall t \in [0,\tau]$ and that

$$\int |\Lambda_u(s) - \Lambda_l(s)|\, dP < \varepsilon.$$

Then for the corresponding bivariate function in $\mathcal{H}_\Lambda^D$, we have

$$\Lambda_l(s) - \Lambda_u(t) < \Lambda(s) - \Lambda(t) < \Lambda_u(s) - \Lambda_l(t), \quad \forall s, t \in [0,\tau]\,.$$

By triangle inequality,

$$\begin{aligned} & \iint |\Lambda_u(s) - \Lambda_l(t) - \Lambda_l(s) + \Lambda_u(t)|\, dP^2 \\ \leq\ & \int\int |\Lambda_u(s) - \Lambda_l(s)|\, dP dP + \int\int |\Lambda_u(t) - \Lambda_l(s)|\, dP dP \\ =\ & \int |\Lambda_u(s) - \Lambda_l(s)|\, dP + \int |\Lambda_u(t) - \Lambda_l(s)|\, dP < 2\varepsilon. \end{aligned}$$

Therefore, $(\Lambda_l(s) - \Lambda_u(t),\, \Lambda_u(s) - \Lambda_l(t))$ is a 2ε-bracket for $\Lambda(s) - \Lambda(t)$ in $L_1(P^2)$, thus there is a constant $K_2 > 0$ such that the bracketing entropy

$$\log N_{[]}(\varepsilon, \mathcal{H}_\Lambda^D, L_1(P^2)) < \frac{K_2}{\varepsilon} < \infty.$$

Since ε is arbitrary, the class $\mathcal{H}_\Lambda^D$ has finite bracketing numbers in $L_1(P^2)$. □

Remark 1. By Corollary 5.2.5 of De la Peña and Giné (1999), the finite bracketing numbers imply the corresponding function classes satisfy the uniform law of large numbers of U-processes. The uniform law of large numbers for $U^P(\boldsymbol{\beta}, \Lambda)$ and its derivatives then follow, because they are Lipschitz functions of the component functions with finite bracketing numbers (van der Vaart and Wellner, 1996).

Proof of Theorem 1. We first re-write the modified composite log-likelihood (3) and the composite score functions using the linear functional notations. Let $N_i(s) = \Delta_i I(X_i \leqslant s)$ be the observed event counting process for subject i, then (3) can be written as

$$\begin{aligned}\ell_n^c(\boldsymbol{\beta}, \Lambda) &= \mathbb{P}_n \int_0^\tau \left\{ (\log \Lambda\{s\} + \mathbf{Z}^T\boldsymbol{\beta}) dN(s) - Y(s) e^{\mathbf{Z}^T\boldsymbol{\beta}} d\Lambda(s) \right\} \\ &\quad - \mathbb{U}_{n,2} \log(1 + R(\boldsymbol{\beta}, \Lambda)).\end{aligned}$$

Differentiating it with respect to $\boldsymbol{\beta}$ yields the composite score function for $\boldsymbol{\beta}$:

$$\begin{aligned}U_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \Lambda) &= \mathbb{P}_n \int_0^\tau \mathbf{Z} \left\{ dN(s) - Y(s) e^{\mathbf{Z}^T\boldsymbol{\beta}} d\Lambda(s) \right\} \\ &\quad - \mathbb{U}_{n,2} \left\{ \frac{R(\boldsymbol{\beta}, \Lambda)}{1 + R(\boldsymbol{\beta}, \Lambda)} \int_0^\tau Q^{(1)}(s; \boldsymbol{\beta}) d\Lambda(s) \right\}.\end{aligned}$$

For $0 \leqslant t \leqslant \tau$ and $h(\cdot) = I(\cdot \leqslant t)$, define a perturbation of Λ by $d\Lambda_\varepsilon = (1 + \varepsilon h) d\Lambda$. The derivative of $\ell_n^c(\boldsymbol{\beta}, \Lambda_\varepsilon)$ with respect to ε evaluated at $\varepsilon = 0$ yields the composite score function for Λ in the direction of h:

$$\begin{aligned} U_\Lambda(\boldsymbol{\beta},\Lambda)(h) &= \mathbb{P}_n \int_0^\tau h(s)\left\{dN(s) - Y(s)e^{\mathbf{Z}^T\boldsymbol{\beta}} d\Lambda(s)\right\} \\ &\quad -\mathbb{U}_{n,2}\left\{\frac{R(\boldsymbol{\beta},\Lambda)}{1+R(\boldsymbol{\beta},\Lambda)}\int_0^\tau Q^{(0)}(s;\boldsymbol{\beta})h(s)d\Lambda(s)\right\}. \end{aligned}$$

As in Section 2.3, we can write the composite score function

$$U(\boldsymbol{\beta},\Lambda) = \begin{pmatrix} U_\beta(\boldsymbol{\beta},\Lambda) \\ U_\Lambda(\boldsymbol{\beta},\Lambda)(h) \end{pmatrix}$$

as the summation of $U^C(\boldsymbol{\beta},\Lambda)$ and $U^P(\boldsymbol{\beta},\Lambda)$; the former is the conditional approach score function and has expectation zero. We can also show that $\mathrm{E}_0\{U^P(\boldsymbol{\beta}_0,\Lambda_0)\} = 0$, since the summand of U^P satisfies $\mathrm{E}_0\{U^P_{ij}(\boldsymbol{\beta}_0,\Lambda_0)\} = 0$, $1 \leqslant i < j \leqslant n$. To see this, note that the pair $(A_i,\ A_j)$ has a binary distribution after conditioning on $(\mathbf{Z}_i,\mathbf{Z}_j)$ and the order statistics of $(A_i,\ A_j)$; thus, by double expectation, we have

$$\begin{aligned} \mathrm{E}_0\{U^P_{ij}(\boldsymbol{\beta},\Lambda)\} &= \mathrm{E}_0\left\{\frac{1}{1+R^{-1}_{ij}(\boldsymbol{\beta},\Lambda)}\cdot\frac{1}{1+R_{ij}(\boldsymbol{\beta}_0,\Lambda_0)}\begin{pmatrix}\int Q^{(1)}(s;\boldsymbol{\beta})d\Lambda(s) \\ \int h(s)Q^{(0)}(s;\boldsymbol{\beta})d\Lambda(s)\end{pmatrix}\right. \\ &\quad \left. -\frac{1}{1+R_{ij}(\boldsymbol{\beta},\Lambda)}\cdot\frac{1}{1+R^{-1}_{ij}(\boldsymbol{\beta}_0,\Lambda_0)}\begin{pmatrix}\int Q^{(1)}(s;\boldsymbol{\beta})d\Lambda(s) \\ \int h(s)Q^{(0)}(s;\boldsymbol{\beta})d\Lambda(s)\end{pmatrix}\right\} \end{aligned} \quad \text{(A.3)}$$

The two terms in the bracket cancel if and only if $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ and $\Lambda = \Lambda_0$ by Lemma 1.

Since $\log \mathcal{L}^P_n$ is always negative, by the similar arguments as in Zeng and Lin (2006), we can show that the PLAC estimator has finite jump sizes, and that $\hat{\Lambda}(\tau)$ is bounded a.s. when $n \to \infty$. Because $\ell^c_n(\boldsymbol{\beta},\Lambda)$ is maximized at the PLAC estimator $(\hat{\boldsymbol{\beta}},\hat{\Lambda})$ over the whole model, it is certainly maximized along the parametric sub-model $(\hat{\boldsymbol{\beta}},\Lambda_\varepsilon)$ at $\varepsilon = 0$. Thus by the regularity conditions, the PLAC estimator is the solution to the composite score equations $U_\beta(\boldsymbol{\beta},\Lambda) = 0$ and $U_\Lambda(\boldsymbol{\beta},\Lambda)(h) = 0$. Interchanging the summations and integrals in the second equation and rearranging the resulting terms, we have

$$\mathbb{P}_n \int_0^\tau h(s) dN(s) = \int_0^\tau h(s) \left\{ \mathbb{P}_n Y(s) e^{\mathbf{Z}^T \hat{\boldsymbol{\beta}}} + \mathbb{U}_{n,2} \frac{R(\hat{\boldsymbol{\beta}}, \hat{\Lambda})}{1 + R(\hat{\boldsymbol{\beta}}, \hat{\Lambda})} Q^{(0)}(s; \hat{\boldsymbol{\beta}}) \right\} d\hat{\Lambda}(s). \quad \text{(A.4)}$$

Let

$$M_n(s; \hat{\boldsymbol{\beta}}, \hat{\Lambda}) = \mathbb{P}_n Y(s) e^{\mathbf{Z}^T \hat{\boldsymbol{\beta}}} + \mathbb{U}_{n,2} \frac{R(\hat{\boldsymbol{\beta}}, \hat{\Lambda})}{1 + R(\hat{\boldsymbol{\beta}}, \hat{\Lambda})} Q^{(0)}(s; \hat{\boldsymbol{\beta}})$$

denote the random function in the brackets. Replacing $h(s)$ with $h(s)/M_n(s; \hat{\boldsymbol{\beta}}, \hat{\Lambda})$ on both sides of (A.4) yields the self-consistency solution of Λ:

$$\hat{\Lambda}(t) = \mathbb{P}_n \int_0^t \frac{dN(s)}{M_n(s; \hat{\boldsymbol{\beta}}, \hat{\Lambda})}.$$

The rest of the proof follows closely to Murphy et al. (1997), yet the technical details are different due to the pairwise pseudo-likelihood. Inspired by the form of $\hat{\Lambda}$, we define another random step function

$$\tilde{\Lambda}(t) = \mathbb{P}_n \int_0^t \frac{dN(s)}{M_n(s; \boldsymbol{\beta}_0, \Lambda_0)}.$$

Let $M_0(s; \boldsymbol{\beta}_0, \Lambda_0) = P_0\{Y(s) e^{\mathbf{Z}^T \beta_0}\}$. Since $\mathrm{E}_0\{U(\boldsymbol{\beta}_0, \Lambda_0)\} = 0$ and $\mathrm{E}_0\{U^P(\boldsymbol{\beta}_0, \Lambda_0)\} = 0$, the same algebra as we used to get $\hat{\Lambda}$ yields

$$\Lambda_0(t) = P_0 \int_0^t \frac{dN(s)}{M_0(s; \boldsymbol{\beta}_0, \Lambda_0)}.$$

Under the regularity conditions (C2)-(C3), by Lemma 2, and the double expectation argument as we used in (A.3), $s \mapsto M_n(s; \boldsymbol{\beta}_0, \Lambda_0)$ is uniformly bounded away from zero and infinity, and is of uniformly bounded variation when n is sufficiently large. Therefore, by the Glivenko-Cantelli theorem and Remark 1, we have

$$\|M_n(s; \boldsymbol{\beta}_0, \Lambda_0) - M_0(s; \boldsymbol{\beta}_0, \Lambda_0)\|_{L_\infty[0,\tau]} \stackrel{a.s.}{\to} 0$$

and

$$\left\|\mathbb{P}_n \int_0^t \frac{dN(s)}{M_n(s;\boldsymbol{\beta}_0,\Lambda_0)} - P_0 \int_0^t \frac{dN(s)}{M_n(s;\boldsymbol{\beta}_0,\Lambda_0)}\right\|_{L_\infty[0,\tau]} \xrightarrow{a.s.} 0,$$

where $\|\cdot\|_{L_\infty[0,\tau]}$ is the supreme norm over $[0,\tau]$. These results combined with the dominated convergence theorem yield

$$\left\|\tilde{\Lambda}(t) - \Lambda_0(t)\right\|_{L_\infty[0,\tau]} \xrightarrow{a.s.} 0.$$

By the definition of the PLAC estimator, the log-composite-likelihood evaluated at $(\hat{\beta},\hat{\Lambda})$ is greater than that evaluated at $(\beta_0,\tilde{\Lambda})$:

$$\begin{aligned} & \mathbb{P}_n \int_0^\tau \left\{\log \frac{\hat{\Lambda}}{\tilde{\Lambda}}\{s\} + \mathbf{Z}^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right\} dN(s) \\ -\mathbb{P}_n \left\{e^{\mathbf{Z}^T\hat{\beta}} \int_0^\tau Y(s)d\hat{\Lambda}(s) - e^{\mathbf{Z}^T\beta_0}\int_0^\tau Y(s)d\tilde{\Lambda}(s)\right\} - \mathbb{U}_{n,2}\log\frac{1+R(\hat{\boldsymbol{\beta}},\hat{\Lambda})}{1+R(\boldsymbol{\beta}_0,\tilde{\Lambda})} & \geqslant 0. \end{aligned}$$

By assumption, $\boldsymbol{\beta}$ is in a compact set, and that $\hat{\Lambda}(t) \leqslant \hat{\Lambda}(\tau)$ is bounded for $t \in [0,\tau]$ with probability one. Thus, by the Bolzano–Weierstrass theorem and the Helly's selection lemma, for every subsequence of $(\hat{\boldsymbol{\beta}},\hat{\Lambda})$, we can find a further subsequence (still denoted as $(\hat{\boldsymbol{\beta}},\hat{\Lambda})$) along which $\hat{\boldsymbol{\beta}} \to \boldsymbol{\beta}^*$ for some $\boldsymbol{\beta}^*$ and $\hat{\Lambda}(t) \to \Lambda^*(t)$, $\forall t \in [0,\tau]$ for some monotone function Λ^* almost surely.

Note that $\hat{\Lambda}(t)$ is absolutely continuous with respect to $\tilde{\Lambda}(t)$. Let $\eta(t) = \lim_{n\to\infty} d\hat{\Lambda}/d\tilde{\Lambda}$ be a bounded measurable function, then $\Lambda^*(t) = \int_0^t \eta(s)d\Lambda_0(s)$ (Zeng and Lin, 2006). By (C1), $\Lambda^*(t)$ is absolutely continuous with respect to the Lebesgue measure and we denote its derivative as $\lambda^*(t)$. Thus we have the ratio $d\hat{\Lambda}/d\tilde{\Lambda}$ converges to $\eta(t) = \lambda^*(t)/\lambda_0(t)$. Again, by the Glivenko-Cantelli theorem, Lemma 2, Remark 1 and the dominant convergence theorem, the difference of the log-composite-likelihoods converges to

$$\begin{aligned} & P_0 \int_0^\tau \left\{\log \frac{\lambda^*}{\lambda_0}(s) + \mathbf{Z}^T(\boldsymbol{\beta}^* - \boldsymbol{\beta}_0)\right\} dN(s) \\ -P_0 \left\{e^{\mathbf{Z}^T\beta^*} \int_0^\tau Y(s)d\Lambda^*(s) - e^{\mathbf{Z}^T\beta_0}\int_0^\tau Y(s)d\Lambda_0(s)\right\} - P_0\log\frac{1+R(\boldsymbol{\beta}^*,\Lambda^*)}{1+R(\boldsymbol{\beta}_0,\Lambda_0)} & \geqslant 0. \end{aligned}$$

The left-hand side is the composite Kullback-Leibler divergence (Varin and Vidoni, 2005) of the density indexed by $(\boldsymbol{\beta}^*, \Lambda^*)$ from the true density, which by Lemma 1 should be strictly negative unless $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0$ and $\Lambda^* = \Lambda_0$. Since every subsequence of $(\hat{\boldsymbol{\beta}}, \hat{\Lambda})$ has a further subsequence converging to $(\boldsymbol{\beta}_0, \Lambda_0)$, we have the convergence of the entire sequence to the same limit. Finally, the uniform convergence of $\hat{\Lambda}(t)$ to $\Lambda_0(t)$ over $[0, \tau]$ follows from the continuity of Λ_0. □

A.3 Asymptotic Normality of $(\hat{\boldsymbol{\beta}}, \hat{\Lambda})$

We first establish a lemma on the $\sqrt{n}$-uniform convergence rate and asymptotic normality of the log-generalized odds ratio. This is achieved by the projection of the U-process.

Lemma 3. *Under Conditions (C1)-(C4), the class of the log-generalized odds ratios*

$$\mathcal{R} = \left\{(\mathcal{O}_i, \mathcal{O}_j) \mapsto r_{ij}(\boldsymbol{\beta}, \Lambda) : \mathcal{O}_i, \mathcal{O}_j \in \Omega, \boldsymbol{\beta} \in B, \Lambda \in \mathcal{H}_\Lambda\right\},$$

where $r_{ij}(\boldsymbol{\beta}, \Lambda) = (e^{\mathbf{Z}_i^T \beta} - e^{\mathbf{Z}_j^T \beta})(\Lambda(A_i) - \Lambda(A_j))$, *satisfies the uniform central limit theorem for U-processes:*

$$\sqrt{n}(\mathbb{U}_{n,2} r(\boldsymbol{\beta}, \Lambda) - P_0^2 r(\boldsymbol{\beta}, \Lambda)) \rightsquigarrow \mathbb{G}_r,$$

where $\mathbb{G}_r$ *is a tight mean-zero Gaussian process.*

Proof. To establish the weak convergence, we first show that

$$\left\|\mathbb{U}_{n,2} r(\boldsymbol{\beta}, \Lambda) - P_0^2 r(\boldsymbol{\beta}, \Lambda) - \hat{\mathbb{U}}_{n,2} r(\boldsymbol{\beta}, \Lambda)\right\|_{\beta,\Lambda} = o_p(n^{-1/2}),$$

where

$$\hat{\mathbb{U}}_{n,2} r(\boldsymbol{\beta}, \Lambda) = \sum_{i=1}^n \mathrm{E}\left(\mathbb{U}_{n,2} r(\boldsymbol{\beta}, \Lambda) - P_0^2 r(\boldsymbol{\beta}, \Lambda) | \mathcal{O}_i\right)$$

is the Hájek projection of $\mathbb{U}_{n,2} r(\boldsymbol{\beta}, \Lambda) - P_0^2 r(\boldsymbol{\beta}, \Lambda)$ (van der Vaart, 2000), and $\|\cdot\|_{\beta,\Lambda}$ is the supreme norm over the parameter space.

It can be verified that $P_0^2 r(\boldsymbol{\beta}, \Lambda) = 2\text{Cov}(e^{\mathbf{Z}^T\boldsymbol{\beta}}, \Lambda(A))$. Moreover, since the pair $\mathcal{O}_i$ and $\mathcal{O}_j$ are i.i.d.,

$$\begin{aligned} \text{E}\,(\, r_{ij}(\boldsymbol{\beta}, \Lambda) \,|\, \mathcal{O}_i) &= \text{E}\left\{ (e^{\mathbf{Z}_i^T\boldsymbol{\beta}} - e^{\mathbf{Z}_j^T\boldsymbol{\beta}})(\Lambda(A_i) - \Lambda(A_j)) \,\middle|\, A_i, \mathbf{Z}_i \right\} \\ &= e^{\mathbf{Z}_i^T\boldsymbol{\beta}}\Lambda(A_i) - \Lambda(A_i)\text{E}e^{\mathbf{Z}_i^T\boldsymbol{\beta}} - e^{\mathbf{Z}_i^T\boldsymbol{\beta}}\text{E}\Lambda(A_i) + \text{E}(e^{\mathbf{Z}_i^T\boldsymbol{\beta}}\Lambda(A_i)). \end{aligned}$$

Thus we have

$$\begin{aligned} \hat{\mathbb{U}}_{n,2} r(\boldsymbol{\beta}, \Lambda) &= \sum_{i=1}^{n} \text{E}\left\{ \binom{n}{2}^{-1} \sum_{j<k} r_{jk}(\boldsymbol{\beta}, \Lambda) - P_0^2 r(\boldsymbol{\beta}, \Lambda) \,\middle|\, \mathcal{O}_i \right\} \\ &= \frac{2}{n}\sum_{i=1}^{n}\left\{ e^{\mathbf{Z}_i^T\boldsymbol{\beta}}\Lambda(A_i) - \Lambda(A_i)\text{E}e^{\mathbf{Z}_i^T\boldsymbol{\beta}} - e^{\mathbf{Z}_i^T\boldsymbol{\beta}}\text{E}\Lambda(A_i) + \text{E}(e^{\mathbf{Z}_i^T\boldsymbol{\beta}}\Lambda(A_i)) \right\} \\ &\quad - 4\text{Cov}(e^{\mathbf{Z}^T\boldsymbol{\beta}}, \Lambda(A)). \end{aligned}$$

Direct calculation gives

$$\tilde{\mathbb{U}}_{n,2} \equiv \mathbb{U}_{n,2} r(\boldsymbol{\beta}, \Lambda) - P^2 r(\boldsymbol{\beta}, \Lambda) - \hat{\mathbb{U}}_{n,2} r(\boldsymbol{\beta}, \Lambda) = \frac{1}{\binom{n}{2}} \sum_{i<j} \tilde{\mathbb{U}}_{n,2}^{(i,j)}.$$

The summand of $\tilde{\mathbb{U}}_{n,2}$ is given by

$$\begin{aligned} \tilde{\mathbb{U}}_{n,2}^{(i,j)} &= e^{\mathbf{Z}_i^T\boldsymbol{\beta}}\Lambda(A_i) - e^{\mathbf{Z}_j^T\boldsymbol{\beta}}\Lambda(A_i) - e^{\mathbf{Z}_i^T\boldsymbol{\beta}}\Lambda(A_j) + e^{\mathbf{Z}_j^T\boldsymbol{\beta}}\Lambda(A_j) - 2\text{Cov}(e^{\mathbf{Z}^T\boldsymbol{\beta}}, \Lambda(A)) \\ &\quad - \left\{ e^{\mathbf{Z}_i^T\boldsymbol{\beta}}\Lambda(A_i) - \Lambda(A_i)\text{E}e^{\mathbf{Z}_i^T\boldsymbol{\beta}} - e^{\mathbf{Z}_i^T\boldsymbol{\beta}}\text{E}\Lambda(A_i) + \text{E}(e^{\mathbf{Z}_i^T\boldsymbol{\beta}}\Lambda(A_i)) \right\} \\ &\quad - \left\{ e^{\mathbf{Z}_j^T\boldsymbol{\beta}}\Lambda(A_j) - \Lambda(A_j)\text{E}e^{\mathbf{Z}_j^T\boldsymbol{\beta}} - e^{\mathbf{Z}_j^T\boldsymbol{\beta}}\text{E}\Lambda(A_j) + \text{E}(e^{\mathbf{Z}_j^T\boldsymbol{\beta}}\Lambda(A_j)) \right\} \\ &\quad + 4\text{Cov}(e^{\mathbf{Z}^T\boldsymbol{\beta}}, \Lambda(A)) \\ &= -(e^{\mathbf{Z}_i^T\boldsymbol{\beta}} - \text{E}e^{\mathbf{Z}_i^T\boldsymbol{\beta}})\,(\Lambda(A_j) - \text{E}\Lambda(A_j)) - (e^{\mathbf{Z}_j^T\boldsymbol{\beta}} - \text{E}e^{\mathbf{Z}_j^T\boldsymbol{\beta}})\,(\Lambda(A_i) - \text{E}\Lambda(A_i))\,, \end{aligned}$$

where the second equality holds by the definition of the covariance and the i.i.d. property of the observations. Therefore, we have

$$\begin{aligned}\tilde{\mathbb{U}}_{n,2} &= -\frac{1}{\binom{n}{2}}\sum_{i=1}^{n}\sum_{j=1}^{n}(e^{\mathbf{Z}_i^T\boldsymbol{\beta}} - \mathrm{E}e^{\mathbf{Z}_i^T\boldsymbol{\beta}})\left(\Lambda(A_j) - \mathrm{E}\Lambda(A_j)\right)\\ &\asymp -2\cdot\frac{1}{n}\sum_{i=1}^{n}(e^{\mathbf{Z}_i^T\boldsymbol{\beta}} - \mathrm{E}e^{\mathbf{Z}_i^T\boldsymbol{\beta}})\cdot\frac{1}{n}\sum_{j=1}^{n}\left(\Lambda(A_j) - \mathrm{E}\Lambda(A_j)\right),\end{aligned}$$

where $\asymp$ means asymptotically equivalent. Since both summations in the last line are empirical processes of Donsker classes, we have

$$\left\|\tilde{\mathbb{U}}_{n,2}\right\|_{\boldsymbol{\beta},\Lambda} \lesssim \left\|n^{-1/2}\mathbb{G}_n e^{\mathbf{Z}^T\boldsymbol{\beta}}\right\|_{\boldsymbol{\beta}}\cdot\left\|n^{-1/2}\mathbb{G}_n\Lambda\right\|_{\Lambda} = O_p(n^{-1/2})O_p(n^{-1/2}) = o_p(n^{-1/2}),$$

where $\lesssim$ means the inequality holds up to a multiplicative constant and $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P_0)$.

Therefore, $\mathbb{U}_{n,2}r(\boldsymbol{\beta},\Lambda) - P_0^2 r(\boldsymbol{\beta},\Lambda)$ is equivalent to its projection $\hat{\mathbb{U}}_{n,2}r(\boldsymbol{\beta},\Lambda)$ up to a term of $o_p(n^{-1/2})$. The weak convergence of $\hat{\mathbb{U}}_{n,2}r(\boldsymbol{\beta},\Lambda)$ can be established using the empirical process theory. Combining these two facts leads to the weak convergence of $\mathbb{U}_{n,2}r(\boldsymbol{\beta},\Lambda)$. □

Proof of Theorem 2. Let θ denote the parameters $(\boldsymbol{\beta},\Lambda)$. We proceed by checking the four conditions in Theorem 3.3.1 of van der Vaart and Wellner (1996). Note that $\sqrt{n}U(\theta_0)$ can be decomposed into $\sqrt{n}U^C(\theta_0) + \sqrt{n}U^P(\theta_0)$. Following the martingale theory, the first term converges weakly to a mean-zero Gaussian process $\mathbb{G}_{U^C}$, and the linear functional

$$\sqrt{n}\left\{b_1^T U_{\boldsymbol{\beta}}^C(\theta_0) + U_{\Lambda}^C(\theta_0)(h)\right\}$$

converges weakly to a mean-zero normal random variable with the variance that can be consistently estimated by $b^T\hat{V}^C b$, where b is defined as in Section 2.3. For the second term, by Lemma 3, the preservation theorem of Lipschitz functions and Theorem 5.3.1 of (De la Peña and Giné, 1999), it also converges weakly to a mean-zero Gaussian process $\mathbb{G}_{U^P}$, and the linear functional

$$\sqrt{n}\left\{b_1^T U_{\boldsymbol{\beta}}^P(\theta_0) + U_\Lambda^P(\theta_0)(h)\right\}$$

converges weakly to a mean-zero normal random variable with the variance that can be consistently estimated by $b^T \hat{V}^P b$. Note also that given $\{(A_i, \mathbf{Z}_i)\}_{i=1}^n$, $U^C(\theta_0)$ is a martingale, whereas $U^P(\theta_0)$ is a function of A_i and $\mathbf{Z}_i$ only, thus by the double expectation

$$\begin{aligned}
\mathrm{E}_0\{U^C(\theta_0) \cdot U^P(\theta_0)\} &= \mathrm{E}_0\left\{\mathrm{E}_0\left(U^C(\theta_0)\Big| (A_i, Z_i), i = 1, \ldots, n\right) \cdot U^P(\theta_0)\right\} \\
&= \mathrm{E}_0\{0 \cdot U^P(\theta_0)\} = 0,
\end{aligned}$$

where $\cdot$ denotes the inner product of the underlying space. This indicates that the $U^C(\theta_0)$ and $U^P(\theta_0)$ are asymptotically independent (van der Vaart and Wellner, 1996, Example 1.4.6) at $\theta_0$that $\sqrt{n}U(\theta_0)$ converges weakly to a mean-zero Gaussian process $\mathbb{G}_U$. In addition, $\sqrt{n}\left\{b_1^T U_{\boldsymbol{\beta}}(\theta_0) + U_\Lambda(\theta_0)(h)\right\}$ converges weakly to a mean-zero normal random variable with asymptotic variance that can be consistently estimated by $b^T(\hat{V}^C + \hat{V}^P)\, b$. Therefore, the two stochastic conditions are satisfied by the consistency of $\hat{\theta}$, Lemma 3 and Lemma 3.3.5 of van der Vaart and Wellner (1996). The fourth condition holds since $\hat{\theta}$ is a zero of $U(\theta)$ and that $u(\theta_0) \equiv \mathrm{E}_0 U(\theta_0) = 0$ by the the arguments in the consistency proof.

To complete the proof, we only need to verify that the Fréchet-derivative of u at θ_0 exists and is continuous invertible. The Fréchet-differentiability can be check directly. For the continuous invertibility, note that the derivative $J \equiv \partial u(\theta)/\partial\theta|_{\theta=\theta_0}$ can be decomposed into J^C and J^P. By (C5) and the classic Cox model results, the first part is continuously invertible. Thus, it suffices to show J^P is a compact operator and that J is one-to-one by the Fredholm theory.

Following Example 3.3.10 of van der Vaart and Wellner (1996), we find the derivate J^P has the form

$$\begin{pmatrix} \boldsymbol{\beta} - \boldsymbol{\beta}_0 \\ \Lambda - \Lambda_0 \end{pmatrix} \mapsto \begin{pmatrix} J^P_{\boldsymbol{\beta}\boldsymbol{\beta}} & J^P_{\boldsymbol{\beta}\Lambda} \\ J^P_{\Lambda\boldsymbol{\beta}} & J^P_{\Lambda\Lambda} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} - \boldsymbol{\beta}_0 \\ \Lambda - \Lambda_0 \end{pmatrix},$$

where

$$
\begin{aligned}
J^P_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}-\boldsymbol{\beta}_0) &= -P_0\left\{\frac{R_0(\int Q_0^{(1)}d\Lambda_0)(\int Q_0^{(1)}d\Lambda_0)^T}{(1+R_0)^2}+\frac{R_0(\int Q_0^{(2)}d\Lambda_0)}{1+R_0}\right\}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\\
J^P_{\boldsymbol{\beta}\Lambda}(\Lambda-\Lambda_0) &= -P_0\left\{\frac{R_0(\int Q_0^{(1)}d\Lambda_0)\int Q_0^{(0)}d(\Lambda-\Lambda_0)}{(1+R_0)^2}\right\}\\
J^P_{\Lambda\boldsymbol{\beta}}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)h &= -P_0\left\{\frac{R_0(\int Q_0^{(1)}d\Lambda_0)^T\int Q_0^{(0)}hd\Lambda_0}{(1+R_0)^2}\right\}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\\
J^P_{\Lambda\Lambda}(\Lambda-\Lambda_0)h &= -P_0\left\{\frac{R_0\int Q_0^{(0)}hd\Lambda_0\cdot\int Q_0^{(0)}hd(\Lambda-\Lambda_0)}{(1+R_0)^2}+\frac{R_0\int Q_0^{(0)}hd(\Lambda-\Lambda_0)}{1+R_0}\right\},
\end{aligned}
$$

where the functions with subscript zero are evaluated at the true parameter θ_0. Note that for $J^P_{\boldsymbol{\beta}\boldsymbol{\beta}}$ and $J^P_{\Lambda\Lambda}$, the second terms in the brackets have expectation zero, by the similar double expectation arguments as in (A.3). Since bounded linear operators with finite dimensional ranges are compact, we only need to show the compactness of $J^P_{\Lambda\boldsymbol{\beta}}$ and $J^P_{\Lambda\Lambda}$. That is to say, for a sequence of functions h_n in the unit ball, $J^P_{\Lambda\boldsymbol{\beta}}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)h_n$ and $J^P_{\Lambda\Lambda}(\Lambda-\Lambda_0)h_n$ have convergent subsequences. In fact, by (C1)-(C2) and the bounded variation properties of the functions involved, the convergent subsequences can be selected using the Helly's lemma; thus, the operator J^P is compact.

We now show J is one-to-one. For $(b,h)\in\mathbb{R}^p\times BV[0,\tau]$, we need to show $J(b,h)=0$ implies $b=0$ and $h(t)=0$. Similar to the arguments in Zeng and Lin (2006), some algebra gives

$$
\begin{aligned}
J(b,h) &= P_0\left\{\left(b^T\int_0^\tau \mathbf{Z}(dN-Ye^{\mathbf{Z}^T\beta_0}d\Lambda_0)+\int_0^\tau hdN-\int_0^\tau Ye^{\mathbf{Z}^T\beta_0}hd\Lambda_0\right)^2\right.\\
&\quad\left.+\frac{1}{R_0}\left\{\frac{R_0}{1+R_0}b^T\int_0^\tau Q_0^{(1)}d\Lambda_0+\frac{R_0}{1+R_0}\int_0^\tau Q_0^{(0)}hd\Lambda_0\right)^2\right\}.
\end{aligned}
$$

Comparing the expressions of J^C and J^P with V^C and V^P, we note that although the second Bartlett equality for the pairwise likelihood does not hold (Varin et al., 2011), the non-negativity of quadratic functions and R_0 indicate that, with probability one, the conditional score along the path $(\boldsymbol{\beta}_0 + b, \Lambda_0 + \varepsilon \int h d\Lambda_0)$

$$b^T \int_0^\tau \mathbf{Z}\{dN(s) - Y(s)e^{\mathbf{Z}^T\beta_0} d\Lambda_0(s)\} + \int_0^\tau h(s)dN(s) - \int_0^\tau Y(s)e^{\mathbf{Z}^T\beta_0} h(s)d\Lambda_0(s) = 0$$

By (C1) and (C3), considering the case of $N(\tau) = 0$ and $A + C \geqslant \tau$ and the case of $N(t) = I(t \geqslant t_0)$ for some $t_0 \in [0, \tau]$ and $A + C \geqslant \tau$, we obtain two equalities. Taking the difference, we have

$$\int_0^\tau \left(b^T\mathbf{Z} + h(s)\right) e^{\mathbf{Z}^T\beta_0} d\Lambda_0(s) + b^T\mathbf{Z} + h(t_0) = 0.$$

The only solution to the above equations is trivial, thus

$$b^T\mathbf{Z} + h(t) = 0, \quad \forall t \in [0, \tau].$$

It follows from the identifiability condition (C2) that $b = 0$ and $h(t) = 0$.

With all four conditions verified, by Theorem 3.3.1 of van der Vaart and Wellner (1996), we have

$$n^{1/2}(\hat{\theta} - \theta_0) \rightsquigarrow -J^{-1}\mathbb{G}_U,$$

where $\mathbb{G}_U$ is a mean-zero Gaussian process. Since linear maps preserve the Gaussian property, $\sqrt{n}(\hat{\theta} - \theta_0)$ also converge weakly to a mean-zero Gaussian process. In addition, the linear functional (8) converges weakly to a mean-zero Gaussian random variable with the variance estimator given by (9). The matrices $\hat{J}^C$ and $\hat{J}^P$ are given by

$$\begin{aligned}
\hat{J}^C &= -\frac{1}{n}\sum_{i=1}^n \partial U_i^C(\boldsymbol{\beta}, \boldsymbol{\lambda})/\partial(\boldsymbol{\beta}^T, \boldsymbol{\lambda}^T)\Big|_{\beta=\hat{\beta}, \boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}}, \\
\hat{J}^P &= \frac{-1}{n(n-1)}\sum\sum_{i \neq j} \partial U_{ij}^P(\boldsymbol{\beta}, \boldsymbol{\lambda})/\partial(\boldsymbol{\beta}^T, \boldsymbol{\lambda}^T)\Big|_{\beta=\hat{\beta}, \boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}}.
\end{aligned}$$

The summand of the above matrices $\partial U_i^C(\boldsymbol{\beta}, \boldsymbol{\lambda})/\partial(\boldsymbol{\beta}^T, \boldsymbol{\lambda}^T)$ and $\partial U_{ij}^P(\boldsymbol{\beta}, \boldsymbol{\lambda})/\partial(\boldsymbol{\beta}^T, \boldsymbol{\lambda}^T)$ take the forms

$$-\begin{pmatrix} \mathbf{Z}_i^{\otimes 2} e^{\mathbf{Z}_i^T\boldsymbol{\beta}} \sum_{k=1}^m \lambda_k Y_i(w_k) & \mathbf{Z}_i e^{\mathbf{Z}_i^T\boldsymbol{\beta}} Y_i(w_1) & \cdots & \mathbf{Z}_i e^{\mathbf{Z}_i^T\boldsymbol{\beta}} Y_i(w_m) \\ \mathbf{Z}_i^T e^{\mathbf{Z}_i^T\boldsymbol{\beta}} Y_i(w_1) & I(X_i = w_1)\Delta_i/\lambda_1^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z}_i^T e^{\mathbf{Z}_i^T\boldsymbol{\beta}} Y_i(w_m) & 0 & \cdots & I(X_i = w_m)\Delta_i/\lambda_m^2 \end{pmatrix}$$

and

$$-R_{ij}\begin{pmatrix} \frac{\left(\Lambda(Q_{ij}^{(1)})\right)^{\otimes 2}}{(1+R_{ij})^2} + \frac{\Lambda(Q_{ij}^{(2)})}{(1+R_{ij})} & \frac{Q_{ij}^{(0)}(w_1)\Lambda(Q_{ij}^{(1)})}{(1+R_{ij})^2} + \frac{Q_{ij}^{(1)}(w_1)}{(1+R_{ij})} & \cdots & \frac{Q_{ij}^{(0)}(w_m)\Lambda(Q_{ij}^{(1)})}{(1+R_{ij})^2} + \frac{Q_{ij}^{(1)}(w_m)}{(1+R_{ij})} \\ \left\{\frac{Q_{ij}^{(0)}(w_1)\Lambda(Q_{ij}^{(1)})}{(1+R_{ij})^2} + \frac{Q_{ij}^{(1)}(w_1)}{(1+R_{ij})}\right\}^T & \frac{\left(Q_{ij}^{(0)}(w_1)\right)^2}{(1+R_{ij})^2} & \cdots & \frac{Q_{ij}^{(0)}(w_1)Q_{ij}^{(0)}(w_m)}{(1+R_{ij})^2} \\ \vdots & \vdots & \ddots & \vdots \\ \left\{\frac{Q_{ij}^{(0)}(w_1)\Lambda(Q_{ij}^{(1)})}{(1+R_{ij})^2} + \frac{Q_{ij}^{(1)}(w_1)}{(1+R_{ij})}\right\}^T & \frac{Q_{ij}^{(0)}(w_1)Q_{ij}^{(0)}(w_m)}{(1+R_{ij})^2} & \cdots & \frac{\left(Q_{ij}^{(0)}(w_m)\right)^2}{(1+R_{ij})^2} \end{pmatrix},$$

respectively, where

$$\Lambda(Q_{ij}^{(l)}) = \sum_{k=1}^m \lambda_k Q_{ij}^{(l)}(w_k).$$

The consistency of variance estimator (9) follows from the Glivenkon-Cantelli theorem and Remark 1. □

References

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical models based on counting processes*. Springer Science & Business Media.

De la Peña, V. and Giné, E. (1999). *Decoupling: from dependence to independence.* Springer.

Murphy, S., Rossini, A., and van der Vaart, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association*, 92(439):968–976.

Nolan, D. and Pollard, D. (1987). U-processes: rates of convergence. *The Annals of Statistics*, pages 780–799.

Pollard, D. (2002). *A user's guide to measure theoretic probability*, volume 8. Cambridge University Press.

van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge University Press.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence*. Springer.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42.

Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, 92(3):519–528.

Wang, M.-C., Brookmeyer, R., and Jewell, N. P. (1993). Statistical models for prevalent cohort data. *Biometrics*, 49(1):1–11.

Zeng, D. and Lin, D. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*, 93(3):627–640.