Supplementary Material for "Sieve estimation of a class of partially linear transformation models with interval-censored competing risks data"

Xuewen Lu[1], [1] Yan Wang [1], Dipankar Bandyopadhyay[2], and Giorgos Bakoyannis[3]

[1] *University of Calgary, Calgary, AB, T2N 1N4, Canada*

[2] *Virginia Commonwealth University, VA 23298-0032, USA*

[3] *Indiana University, Indianapolis, IN, USA*

The Supplementary Material consists of two sections. While Section SM1 contains detailed derivation and proofs of the theoretical results presented in Section 3 (main paper); Section SM2 presents Scenarios 2 and 3, the tables and figures summarizing all the simulation results in Section 4 (main paper).

SM1. Derivations and Proofs

Proof of Theorem 1: Consistency

Techniques from empirical processes (Shorack and Wellner, 2009) will be applied to derive the consistency of the B-spline and Bernstein polynomial sieve MLEs. Define $Pf = \int_{\mathcal{X}} f(x)dP(x)$ and $\mathbb{P}_n f = n^{-1}\sum_{i=1}^{n} f(X_i)$, the empirical process indexed by the function f evaluated at $X_i = (V_i, U_i, \delta_{i1}, \ldots, \delta_{ik}, \delta_{i1}^1, \ldots, \delta_{ik}^1, \delta_i, Z_i)$, the sample point of individual i, with $\mathcal{X}$ denoting the sample space. Also, let K and C denote general constants which may differ according to the settings. The following regularity conditions are assumed:

C1. $E(ZZ^\top)$ is non-singular and Z is bounded, i.e., there exists a $z_0 > 0$ such that $P(\|Z\| \le z_0) = 1$.

C2. $\beta_j \in \mathcal{B}_j$, where $\mathcal{B}_j$ is a compact subset of $\mathbb{R}^d$ for every $j = 1, \ldots, k$, where d is a fixed integer and $1 \le d \ll n$.

C3. There exists an $\eta > 0$ such that $P(U - V \ge \eta) = 1$ and the union of the supports of V and U are contained in $[a, b]$, where $0 < a < b < \infty$ and $0 < \min_{j\in\{1,\ldots,k\}} F_j(a) < \sum_{j=1}^{k} F_j(b) < 1$.

C4. Functions $\phi_{0,j} \in \Phi$, $j = 1, \ldots, k$, where Φ is a class of functions with bounded p-th derivative in $[a, b]$ for $p \ge 1$ and the first derivative of $\phi_{0,j}$ is strictly positive and continuous on $[a, b]$. Functions $\psi_{0,je} \in \Psi$, $e = 1, \ldots, q$, where Ψ is a class of functions with bounded r-th derivative in $[a_w^e, b_w^e]$ for $r \ge 1$ and the first derivative of $\psi_{0,je}$ is continuous, q is a fixed integer and $1 \le q \ll n$.

C5. The conditional density of (V, U) given (Z, W) has bounded partial derivatives with respect to (v, u) and the bounds of these derivatives do not depend on (v, u, z, w).

C6. For some $\kappa \in (0, 1)$, $a^\top \mathrm{Var}(Z|V, W)a \ge \kappa a^\top E(ZZ^\top|V, W)a$ a.s. and $a^\top \mathrm{Var}(Z|U, W)a \ge \kappa a^\top E(ZZ^\top|U, W)a$ a.s. for all $a \in \mathbb{R}^d$.

C7. The number of internal knots $N_j = O(n^\nu)$ of the B-splines and the degree of the Bernstein polynomials $m_w = O(n^\nu)$ for all $j = 1, \ldots, k$, where ν satisfies $1/[2(1 + \sigma)] < \nu < 1/(2\sigma)$, $\sigma = \min(p, r/2)$.

[1]Corresponding author. Email: lux@math.ucalgary.ca. Phone: (1-403)220-6620. Fax: (1-403)282-5150

C8. For the σ defined in (C7), $\sigma > 1$ and $\nu > 1/(4\sigma)$.

C9. The semiparametric information matrix $I(\beta_0)$ is positive definite.

Let $\mathcal{F}_1 = \{l(\theta; X) : \theta \in \Theta_n\}$ denote the class of log-likelihood functions indexed by the parameter space $\Theta_n = \prod_{j=1}^{k}(\mathcal{B}_j \otimes \mathcal{M}_{n,j} \otimes \mathcal{W}_{n,j})$. The Euclidean parameters are $\beta_j \in \mathcal{B}_j$ and the functional parameters are $\phi_j \in \mathcal{M}_{n,j}$ and $\psi_j \in \mathcal{W}_{n,j}$, respectively, for $j = 1, \ldots, k$. The functional parameter space $\mathcal{M}_{n,j} \equiv \mathcal{M}_n(\gamma_j, m_{\phi,n,j})$ is the space of monotone B-spline functions defined on the observation time interval $[a, b]$, and $\mathcal{W}_{n,j} \equiv \mathcal{W}_n(\alpha_j, m_{\psi,n,j})$ is the space of Bernstein polynomial functions without constraints defined on the interval $[a_w, b_w]$. Convergence will be proved in the L_2-metric

$$d(\theta_1, \theta_2) = \left(\sum_{j=1}^{k} \|\beta_j^{(1)} - \beta_j^{(2)}\|^2 + \sum_{j=1}^{k} \|\phi_j^{(1)} - \phi_j^{(2)}\|_{\Phi}^2 + \sum_{j=1}^{k} \|\psi_j^{(1)} - \psi_j^{(2)}\|_{\Psi}^2 \right)^{\frac{1}{2}},$$

for $\theta_1 = (\beta^{(1)\top}, \phi^{(1)\top}, \psi^{(1)\top})^\top$ and $\theta_2 = (\beta^{(2)\top}, \phi^{(2)\top}, \psi^{(2)\top})^\top$, where

$$\|\phi_j^{(1)} - \phi_j^{(2)}\|_{\Phi}^2 = E\left[\phi_j^{(1)}(V) - \phi_j^{(2)}(V)\right]^2 + E\left[\phi_j^{(1)}(U) - \phi_j^{(2)}(U)\right]^2, \quad j = 1, \ldots, k$$

and

$$\|\psi_j^{(1)} - \psi_j^{(2)}\|_{\Psi}^2 = \sum_{e=1}^{q} \|\psi_{je}^{(1)} - \psi_{je}^{(2)}\|_{\Psi}^2 = \sum_{e=1}^{q} E\left[\psi_{je}^{(1)}(W_e) - \psi_{je}^{(2)}(W_e)\right]^2, \quad j = 1, \ldots, k,$$

and $\|\cdot\|$ denotes the Euclidean norm. Let $\mathbb{M}(\theta) = Pl(\theta; X)$ and $\mathbb{M}_n(\theta) = \mathbb{P}_n l(\theta; X)$, which leads to $\mathbb{M}_n(\theta) - \mathbb{M}(\theta) = (\mathbb{P}_n - P)l(\theta; X)$ for each $\theta \in \Theta_n$. In order to prove that $d(\hat{\theta}_n, \theta_0) \overset{a.s.}{\to} 0$, we need to verify the following conditions:

(1) $\sup_{\theta \in \Theta_n} |\mathbb{M}_n(\theta) - \mathbb{M}(\theta)| \overset{a.s.}{\to} 0$.

(2) $\sup_{\theta \,:\, d(\theta, \theta_0) \geq \epsilon} \mathbb{M}(\theta) < \mathbb{M}(\theta_0)$.

(3) The sequence of estimators $\hat{\theta}_n$ satisfy $\mathbb{M}_n(\hat{\theta}_n) \geq \mathbb{M}_n(\theta_0) - o_{a.s.}(1)$, where $o_{a.s.}(1)$ represents a random term tending to zero almost surely.

Proof of (1): We first define the covering number of the class $\mathcal{F}_1$. For any $\varepsilon > 0$, define the covering number $N(\varepsilon, \mathcal{F}_1, L_1(P_n))$ as the smallest value of κ for which there exists $\{\theta_{(1)}, \ldots, \theta_{(\kappa)}\}$ such that

$$\min_{j \in \{1, \ldots, \kappa\}} |\mathbb{M}_n(\theta) - \mathbb{M}(\theta_{(j)})| < \varepsilon$$

for all $\theta \in \Theta_n$ and $\theta_{(j)} = (\beta^{(j)}, \phi^{(j)}, \psi^{(j)}) \in \Theta_n$.

Assume that Conditions C1-C5 hold. Then, the covering number $N(\epsilon, \mathcal{F}_n, L_1(P_n))$ of the set of functions $\mathcal{F}_n = \{\ell(\theta, X) : \theta \in \Theta_n\}$ satisfies

$$N(\epsilon, \mathcal{F}_n, L_1(P_n)) \leq K M_n^{k(m+q(m_w+1))} \varepsilon^{-[d+k(m+q(m_w+1))]},$$

where K denotes an arbitrary constant, that varies according to the inequality, $M_n = O(n^a)$ with $0 < a < 1/2$, which controls the size of the space Θ_n, m is the number of the B-spline basis functions, m_w is the degree of Bernstein polynomials, kp is the dimension of a regression parameter β, and $k(1+q)$ is the number of nonparametric functions. We now proceed to calculate the covering number $N(\epsilon, \mathcal{F}_n, L_1(P_n))$. For any $\theta^1 = (\beta^1, \phi^1, \psi^1)$, $\theta^2 = (\beta^2, \phi^2, \psi^2) \in \Theta_n$, and using the Mean Value Theorem, we have

$$|\ell(\theta^1; X) - \ell(\theta^2; X)| \leq K(\|\beta_1 - \beta_2\| + \|\phi^1 - \phi^2\|_\infty + \|\psi^1 - \psi^2\|_\infty),$$

where $\|\phi^1 - \phi^2\|_\infty = \sum_{j=1}^k \|\phi_j^1 - \phi_j^2\|_\infty$, and $\|\psi^1 - \psi^2\|_\infty = \sum_{j=1}^k \sum_{e=1}^q \|\psi_{je}^1 - \psi_{je}^2\|_\infty$.

Denote $\gamma_j^i = (\gamma_{j1}^i, \ldots, \gamma_{jm}^i)^\top$, $i = 1, 2$, $j = 1, \ldots, k$, the B-spline coefficients corresponding to ϕ_j^i, and $\alpha_{je}^i = (\alpha_{(je)1}^i, \ldots, \alpha_{(je)m_w}^i)^\top$, $j = 1, \ldots, k$, $e = 1, \ldots, q$, the Bernstein coefficients corresponding to ψ_{je}^i. Then, we have

$$\begin{aligned}
\|\phi^1 - \phi^2\|_\infty &= \sum_{j=1}^k \sup_t \|\phi_j^1 - \phi_j^2\| \\
&= \sum_{j=1}^k \sup_t \|\sum_{s=1}^m \gamma_{js}^1 B_s(t, m, a, b) - \sum_{s=1}^m \gamma_{js}^2 B_s(t, m, a, b)\| \\
&\leq \sum_{j=1}^k \max_{1 \leq s \leq m} |\gamma_{js}^1 - \gamma_{js}^2| \sum_{s=1}^m B_s(t, m, a, b) \\
&= \sum_{j=1}^k \max_{1 \leq s \leq m} |\gamma_{js}^1 - \gamma_{js}^2| \\
&= \sum_{j=1}^k \|\gamma_j^1 - \gamma_j^2\|,
\end{aligned}$$

since $\sum_{s=1}^m B_s(t, m, a, b) = 1$, $B_s(t, m, a, b) \geq 0$, where $\|\gamma_j^1 - \gamma_j^2\| = \max_{1 \leq s \leq m} |\gamma_{js}^1 - \gamma_{js}^2|$. Similarly, we have

$$\begin{aligned}
\|\psi^1 - \psi^2\|_\infty &= \sum_{j=1}^k \sum_{e=1}^q \|\psi_{je}^1 - \psi_{je}^2\|_\infty \\
&= \sum_{j=1}^k \sum_{e=1}^q \sup_{w_e} |\sum_{s=0}^{m_w} \alpha_{(je)s}^1 \{B_s^e(w_e, m_w, a_w^e, b_w^e) - B_s^e(a_w^e, m_w, a_w^e, b_w^e)\} \\
&\quad - \sum_{s=0}^{m_w} \alpha_{(je)s}^2 \{B_s^e(w_e, m_w, a_w^e, b_w^e) - B_s^e(a_w^e, m_w, a_w^e, b_w^e)\}| \\
&\leq \sum_{j=1}^k \sum_{e=1}^q [\sup_{w_e} |\sum_{s=0}^{m_w} \alpha_{(je)s}^1 B_s^e(w_e, m_w, a_w^e, b_w^e) - \sum_{s=0}^{m_w} \alpha_{(je)s}^2 B_s^e(w_e, m_w, a_w^e, b_w^e)| \\
&\quad + \sup_{w_e} |\sum_{s=0}^{m_w} \alpha_{(je)s}^1 B_s^e(a_w^e, m_w, a_w^e, b_w^e) - \sum_{s=0}^{m_w} \alpha_{(je)s}^2 B_s^e(a_w^e, m_w, a_w^e, b_w^e)|] \\
&\leq 2 \sum_{j=1}^k \sum_{e=1}^q \sup_{w_e} |\sum_{s=0}^{m_w} \alpha_{(je)s}^1 B_s^e(w_e, m_w, a_w^e, b_w^e) - \sum_{s=0}^{m_w} \alpha_{(je)s}^2 B_s^e(w_e, m_w, a_w^e, b_w^e)| \\
&\leq 2 \sum_{j=1}^k \sum_{e=1}^q \max_{0 \leq s \leq m_w} |\alpha_{(je)s}^1 - \alpha_{(je)s}^2| \sum_{s=0}^{m_w} B_s^e(w_e, m_w, a_w^e, b_w^e) \\
&= 2 \sum_{j=1}^k \sum_{e=1}^q \|\alpha_{(je)}^1 - \alpha_{(je)}^2\|,
\end{aligned}$$

since $\sum_{s=0}^{m_w} B_s^e(w_e, m_w, a_w^e, b_w^e) = 1$, $B_s^e(w_e, m_w, a_w^e, b_w^e) \geq 0$, where $\|\alpha_{(je)}^1 - \alpha_{(je)}^2\| = \max_{0 \leq s \leq m_w} |\alpha_{(je)s}^1 - \alpha_{(je)s}^2|$.

Thus, from the above results, we obtain

$$|\ell(\theta^1; X) - \ell(\theta^2; X)| \quad \leq \quad K\|\beta_1 - \beta_2\| + K \sum_{j=1}^k \|\gamma_j^1 - \gamma_j^2\| + K \sum_{j=1}^k \sum_{e=1}^q \|\alpha_{(je)}^1 - \alpha_{(je)}^2\|.$$

Let $C_{kq} = 1 + k + kq$. Then, from Lemma 2.5 of van de Geer (2000), we show that $\{\beta \in \mathbb{R}^d, \|\beta\| \leq M\}$ is covered by $\{5M/(\varepsilon/(C_{kq}K))\}^d$ balls with radius $\varepsilon/(C_{kq}K)$, $\{\gamma_j \in \mathbb{R}^m, \max_{1 \leq s \leq m} |\gamma_{js}| \leq M_n\}$ is covered by $\{5M_n/(\varepsilon/(C_{kq}K))\}^m$ balls with radius $\varepsilon/(C_{kq}K)$, $j = 1, \ldots, k$, and $\{\alpha_{(je)} \in \mathbb{R}^{m_w+1}, \max_{0 \leq s \leq m_w} |\alpha_{(je)s}| \leq M_n\}$ is covered by $\{5M_n/(\varepsilon/(C_{kq}K))\}^{m_w+1}$

balls with radius $\varepsilon/(C_{kq}K)$, $j=1,\ldots,k$ and $e=1,\ldots,q$. Therefore, the $L_1(P_n)$ covering number of $\mathcal{F}_n$ is bounded by

$$\begin{aligned} N(\epsilon,\mathcal{F}_n,L_1(P_n)) &\le \left(\frac{5C_{kq}KM}{\varepsilon}\right)^d \prod_{j=1}^k \left[\left(\frac{5C_{kq}KM_n}{\varepsilon}\right)^m\right] \prod_{j=1}^k \prod_{e=1}^q \left[\left(\frac{5C_{kq}KM_n}{\varepsilon}\right)^{m_w+1}\right] \\ &\le KM_n^{k(m+q(m_w+1))}\varepsilon^{-[d+k(m+q(m_w+1))]}. \end{aligned}$$

Now, we are ready to show the condition (1) holds. Given $|\ell(\theta,X)|$ is bounded under Conditions C1-C5, we assume $\sup_{\theta\in\Theta}|\ell(\theta,X)| \le M_a$, where $M_a>0$ is a constant. Then, $P\{\ell(\theta,X)\} \le P\{\sup_{\theta\in\Theta}|\ell(\theta,X)|\}^2 \le M_a^2$. Let $\alpha_n = n^{-(1/2-\zeta)}(\log n)^{1/2}$ and $\epsilon_n=\epsilon\alpha_n$ with $\nu/2<\zeta<1/2$ and $\epsilon>0$. Then, $\{\epsilon_n\}$ is a non-increasing sequence of positive numbers, with $\epsilon_n\to 0$ when $n\to\infty$. Then, for a sufficiently large n and any $\theta\in\Theta_n$, we have

$$\mathrm{Var}\{P_n\ell(\theta,\mathcal{X})\}/(4\epsilon_n)^2 \le \frac{(1/n)P\ell^2(\theta,X)}{16\epsilon^2\alpha_n^2} \le \frac{1}{16\epsilon^2 n\alpha_n^2 M_a} = \frac{1}{16\epsilon^2 n^{2\zeta}\log n M_a} \ll \frac{1}{2}.$$

Let P_n^o denote the signed measure that places mass $\pm n^{-1}$ at each of the observations $\{X_1,\ldots,X_n\}$, with the random $\pm$ signs being decided independently of the X_i's. Then by Pollard (1984) (refer to page 31), and $\mathrm{Var}\{P_n\ell(\theta,\mathcal{X})\}/(4\epsilon_n)^2 \le 1/2$, we obtain the following symmetrization inequality:

$$P\left\{\sup_{\theta\in\Theta_n}|P_n\ell(\theta,X)-P\ell(\theta,X)|>8\epsilon_n\right\} \le 4P\left\{\sup_{\theta\in\Theta_n}|P_n^o\ell(\theta,X)|>2\epsilon_n\right\}.$$

Let $\mathcal{X}=\{X_1,\ldots,X_n\}$ and $J_n=N(\epsilon_n/2,\mathcal{L}_n,L_1(P_n))$. Given $\mathcal{X}$, choose $\theta_j\in\Theta_n$, $j=1,\ldots,J_n$, such that for all $\theta\in\Theta_n$, we have

$$\min_{j\in\{1,\ldots,J_n\}} P_n|\ell(\theta,X)-\ell(\theta^{(j)},X)|<\epsilon_n/2.$$

Then, for each $\theta\in\Theta_n$, there exists a $j'\in\{1,\ldots,J_n\}$ such that $P_n|\ell(\theta,X)-\ell(\theta^{(j')},X)|<\epsilon_n/2$, hence, $|P_n^o\{\ell(\theta,X)-\ell(\theta^{(j')},X)\}| = |n^{-1}\sum_{i=1}^n \pm\{\ell(\theta,X_i)-\ell(\theta^{(j')},X_i)\}| \le n^{-1}\sum_{i=1}^n|\ell(\theta,X_i)-\ell(\theta^{(j')},X_i)| = P_n|\ell(\theta,X)-\ell(\theta^{(j')},X)|<\epsilon_n/2$ and $|P_n^o\{\ell(\theta^{(j')},X)| = |P_n^o\ell(\theta,X)-P_n^o\{\ell(\theta,X)-\ell(\theta^{(j')},X)\}| > P_n^o|\ell(\theta,X)|-\epsilon_n/2$. Then, we obtain

$$\max_{j\in\{1,\ldots,J_n\}}|P_n^o\{\ell(\theta^{(j)},X)| \ge \sup_{\theta\in\Theta_n}P_n^o|\ell(\theta,X)|-\epsilon_n/2.$$

Therefore, we have

$$\begin{aligned} \Pr(\sup_{\theta\in\Theta_n}P_n^o|\ell(\theta,X)|>2\epsilon|\mathcal{X}) &\le \Pr(\max_{j\in\{1,\ldots,J_n\}}|P_n^o\{\ell(\theta^{(j)},X)|>3\epsilon_n/2|\mathcal{X}) \\ &\le \Pr(\max_{j\in\{1,\ldots,J_n\}}|P_n^o\{\ell(\theta^{(j)},X)|>\epsilon_n|\mathcal{X}) \\ &\le N(\epsilon_n/2,\mathcal{L}_n,L_1(P_n))\max_{j\in\{1,\ldots,J_n\}}\Pr(|P_n^o\{\ell(\theta^{(j)},X)|>\epsilon_n|\mathcal{X}). \end{aligned}$$

From Hoeffding's inequality (Pollard, 1984, Appendix B), for each $\theta^{(j)}$, we have

$$\begin{aligned} \Pr(|P_n^o\ell(\theta^{(j)},X)|>\epsilon_n|\mathcal{X}) &= \Pr(|\sum_{i=1}^n \pm\ell(\theta^{(j)},X_i)|>n\epsilon_n|\mathcal{X}) \\ &\le 2\exp[-2(n\epsilon_n)^2/\sum_{i=1}^n(2\ell(\theta^{(j)},X_i))^2] \\ &\le 2\exp\{-n\epsilon_n^2/(2M_a)\}. \end{aligned}$$

By the three inequalities above, we obtain

$$\begin{aligned} \Pr(\sup_{\theta\in\Theta_n}P_n^o|\ell(\theta,X)|>2\epsilon_n|\mathcal{X}) &\le 2N(\epsilon_n/2,\mathcal{L}_n,L_1(P_n))\exp\{-n\epsilon_n^2/(2M_a)\} \\ &\le 2KM_n^{k(m+q(m_w+1))}\varepsilon^{-[d+k(m+q(m_w+1))]}\exp\{-n\epsilon_n^2/(2M_a)\}. \end{aligned}$$

Using this result and the symmetrization inequality derived above, we obtain

$$\begin{aligned}
&\Pr\left\{\sup_{\theta\in\Theta_n}|P_n\ell(\theta,X)-P\ell(\theta,X)dP|>8\epsilon_n\right\}\\
&\leq\ 4P\left\{\sup_{\theta\in\Theta_n}|P_n^o l(\theta,Z,W)|>2\epsilon_n\right\}\\
&\leq\ 8KM_n^{k(m+q(m_w+1))}\varepsilon^{-[d+k(m+q(m_w+1))]}\exp\{-n\epsilon_n^2/(2M_a)\}\\
&\leq\ 8K\exp\Big[ak\{m+q(m_w+1)\}-\{d+k(m+q(m_w+1))\}\log\epsilon-n\epsilon_n^2/(2M_a)\Big]\\
&=\ 8K\exp\Big[-\{\epsilon^2/(2M_a)+o(1)\}n^{2\zeta}\log n\Big]\\
&\leq\ 8K\exp\Big(-K_1n^{2\zeta}\log n\Big),
\end{aligned}$$

where $0<K_1<\epsilon^2/(2M_a)$. Thus $\sum_{n=1}^{\infty}P\left\{\sup_{\theta\in\Theta_n}|P_n\ell(\theta,X)-P\ell(\theta,X)|>8\epsilon_n\right\}<\infty$. By the Borel-Cantelli lemma, we have $\sup_{\theta\in\Theta_n}|P_n\ell(\theta,X)-P\ell(\theta,X)|\to 0$ almost surely, which completes the proof of (1). □

Next, we are going to show that condition (2), i.e., $\sup_{\theta\,:\,d(\theta,\theta_0)\geq\epsilon}\mathbb{M}(\theta)<\mathbb{M}(\theta_0)$, holds. By definition we have:

$$\begin{aligned}
\mathbb{M}(\theta_0)-\mathbb{M}(\theta)\ =\ & E\Bigg\{\sum_{j=1}^{k}F_j^0(V,Z,W)\log\frac{F_j^0(V,Z,W)}{F_j(V,Z,W)}\\
&+\sum_{j=1}^{k}\left[F_j^0(U,Z,W)-F_j^0(V,Z,W)\right]\log\frac{F_j^0(U,Z,W)-F_j^0(V,Z,W)}{F_j(U,Z,W)-F_j(V,Z,W)}\\
&+\left[1-\sum_{j=1}^{k}F_j^0(U,Z,W)\right]\log\frac{1-\sum_{j=1}^{k}F_j^0(U,Z,W)}{1-\sum_{j=1}^{k}F_j(U,Z,W)}\Bigg\}\\
=\ & E\Bigg\{\sum_{j=1}^{k}F_j(V,Z,W)m\left[\frac{F_j^0(V,Z,W)}{F_j(V,Z,W)}\right]\\
&+\sum_{j=1}^{k}\left[F_j(U,Z,W)-F_j(V,Z,W)\right]m\left[\frac{F_j^0(U,Z,W)-F_j^0(V,Z,W)}{F_j(U,Z,W)-F_j(V,Z,W)}\right]\\
&+\left[1-\sum_{j=1}^{k}F_j(U,Z,W)\right]m\left[\frac{1-\sum_{j=1}^{k}F_j^0(U,Z,W)}{1-\sum_{j=1}^{k}F_j(U,Z,W)}\right]\Bigg\},
\end{aligned}$$

where, as in Zhang et al. (2010), $m(x)=x\log x-x+1\geq(1/4)(x-1)^2$ for $x\in[0,5]$. Hence, for any θ in a sufficiently small neighbourhood of θ_0, we have

$$
\begin{aligned}
&\mathbb{M}(\theta_0) - \mathbb{M}(\theta) \\
&\geq \frac{1}{4}E\Bigg\{\sum_{j=1}^{k} F_j(V,Z,W)\left[\frac{F_j^0(V,Z,W)}{F_j(V,Z,W)} - 1\right]^2 \\
&\quad + \sum_{j=1}^{k}[F_j(U,Z,W) - F_j(V,Z,W)]\left[\frac{F_j^0(U,Z,W) - F_j^0(V,Z,W)}{F_j(U,Z,W) - F_j(V,Z,W)} - 1\right]^2 \\
&\quad + \left[1 - \sum_{j=1}^{k} F_j(U,Z,W)\right]\left[\frac{1 - \sum_{j=1}^{k} F_j^0(U,Z,W)}{1 - \sum_{j=1}^{k} F_j(U,Z,W)} - 1\right]^2\Bigg\} \\
&= \frac{1}{4}E\Bigg\{\sum_{j=1}^{k}\frac{1}{F_j(V,Z,W)}\left[F_j^0(V,Z,W) - F_j(V,Z,W)\right]^2 \\
&\quad + \sum_{j=1}^{k}\frac{1}{F_j(U,Z,W) - F_j(V,Z,W)}\left[F_j^0(U,Z,W) - F_j^0(V,Z,W) - F_j(U,Z,W) + F_j(V,Z,W)\right]^2 \\
&\quad + \frac{1}{F_j(U,Z,W)}\left[\sum_{j=1}^{k} F_j^0(U,Z,W) - \sum_{j=1}^{k} F_j(U,Z,W)\right]^2\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
&\geq \frac{1}{4}E\Bigg\{\sum_{j=1}^{k}\left[F_j^0(V,Z,W) - F_j(V,Z,W)\right]^2 \\
&\quad + \sum_{j=1}^{k}\left[F_j^0(U,Z,W) - F_j^0(V,Z,W) - F_j(U,Z,W) + F_j(V,Z,W)\right]^2 \\
&\quad + \left[\sum_{j=1}^{k} F_j^0(U,Z,W) - \sum_{j=1}^{k} F_j(U,Z,W)\right]^2\Bigg\} \\
&\geq \frac{1}{4}E\left(\sum_{j=1}^{k}\left[F_j^0(V,Z,W) - F_j(V,Z,W)\right]^2 + \left\{\sum_{j=1}^{k}\left[F_j^0(U,Z,W) - F_j(U,Z,W)\right]\right\}^2\right) \\
&\geq \frac{1}{4k}E\left(\left\{\sum_{j=1}^{k}\left[F_j^0(V,Z,W) - F_j(V,Z,W)\right]\right\}^2 + \left\{\sum_{j=1}^{k}\left[F_j^0(U,Z,W) - F_j(U,Z,W)\right]\right\}^2\right) \\
&= CE\Bigg(\left\{\sum_{j=1}^{k}\left[(\beta_{0,j} - \beta_j)^\top Z + (\phi_{0,j} - \phi_j)(V) + \sum_{e=1}^{q}(\psi_{0,je} - \psi_j)(W_e)\right]\right\}^2 \\
&\quad + \left\{\sum_{j=1}^{k}\left[(\beta_{0,j} - \beta_j)^\top Z + (\phi_{0,j} - \phi_j)(U) + \sum_{e=1}^{q}(\psi_{0,je} - \psi_j)(W_e)\right]\right\}^2\Bigg),
\end{aligned}
$$

where the last equality follows from a Taylor expansion around the value of the linear predictor $\eta_j(X)$. Using the same arguments as in Wellner et al. (2007), pp. 2126-2127, along with regularity conditions C1-C6, it follows that

$$
\mathbb{M}(\theta_0) - \mathbb{M}(\theta) \geq C\left(\sum_{j=1}^{k}\|\beta_j - \beta_{j,0}\|^2 + \sum_{j=1}^{k}\|\phi_j - \phi_{j,0}\|_\Phi^2 + \sum_{j=1}^{k}\|\psi_j - \psi_{j,0}\|_\Psi^2\right) = Cd^2(\theta_0,\theta).
$$

Consequently

$$
\begin{aligned}
\sup_{\theta\,:\,d(\theta,\theta_0)\geq\epsilon}\mathbb{M}(\theta) &\leq \sup_{\theta\,:\,d(\theta,\theta_0)\geq\epsilon}\left[\mathbb{M}(\theta_0) - Cd^2(\theta_0,\theta)\right] \\
&= \mathbb{M}(\theta_0) - \inf_{\theta\,:\,d(\theta,\theta_0)\geq\epsilon} Cd^2(\theta_0,\theta) \\
&= \mathbb{M}(\theta_0) - C\epsilon^2 \\
&< \mathbb{M}(\theta_0).
\end{aligned}
$$

Thus, we prove condition (2).

Finally, we prove condition (3), i.e., that the sequence of estimators $\hat{\theta}_n$ satisfy $\mathbb{M}_n(\hat{\theta}_n) \geq \mathbb{M}_n(\theta_0) - o_{a.s.}(1)$. Letting $\theta_{0,n} = (\beta_0, \phi_{0,n})$ it follows that

$$\begin{aligned} &\mathbb{M}_n(\hat{\theta}_n) - \mathbb{M}_n(\theta_0) \\ &= \mathbb{M}_n(\hat{\theta}_n) - \mathbb{M}_n(\theta_{0,n}) + \mathbb{M}_n(\theta_{0,n}) - \mathbb{M}_n(\theta_0) \\ &\geq \mathbb{M}_n(\theta_{0,n}) - \mathbb{M}_n(\theta_0) \\ &= \mathbb{M}_n(\theta_{0,n}) - \mathbb{M}_n(\theta_0) + \mathbb{M}(\theta_{0,n}) - \mathbb{M}(\theta_{0,n}) + \mathbb{M}(\theta_0) - \mathbb{M}(\theta_0) \\ &= \{\mathbb{M}_n(\theta_{0,n}) - \mathbb{M}(\theta_{0,n})\} + \{\mathbb{M}(\theta_0) - \mathbb{M}_n(\theta_0)\} + \{\mathbb{M}(\theta_{0,n}) - \mathbb{M}(\theta_0)\}. \end{aligned} \tag{SM11}$$

Now, from condition (1) proved earlier, we obtain $\{\mathbb{M}_n(\theta_{0,n}) - \mathbb{M}(\theta_{0,n})\} = o_{a.s.}(1)$. By the SLLN, we obtain $\{\mathbb{M}(\theta_0) - \mathbb{M}_n(\theta_0)\} = o_{a.s.}(1)$. Next, for the remaining term in (SM11), we can again use the boundedness of the loglikelihood function for any sample point $X \in \mathcal{X}$. Furthermore, with $\|\phi_{0,n,j} - \phi_{0,j}\|_\Phi = O(n^{-p\nu})$ and $\|\psi_{0,n,je} - \psi_{0,je}\|_\Psi = O(n^{-r\nu/2})$ along with the dominated convergence theorem, it follows that

$$\mathbb{M}(\theta_{0,n}) - \mathbb{M}(\theta_0) > -o(1)$$

as $n \to \infty$. Thus

$$\mathbb{M}_n(\hat{\theta}_n) - \mathbb{M}_n(\theta_0) \geq o_{a.s.}(1) - o(1) = -o_{a.s.}(1)$$

and this proves the final consistency condition. Consequently

$$\left(\sum_{j=1}^{k} \|\hat{\beta}_{j,n} - \beta_{j,0}\|^2 + \sum_{j=1}^{k} \|\hat{\phi}_{j,n} - \phi_{j,0}\|_\Phi^2 + \sum_{j=1}^{k} \|\hat{\psi}_{j,n} - \psi_{j,0}\|_\Psi^2 \right)^{\frac{1}{2}} \overset{a.s.}{\to} 0. \qquad \blacksquare$$

Proof of Theorem 2: Rate of Convergence

Derivation of the rate of convergence will be based on Theorem 3.2.5 of van der Vaart and Wellner (1996). We showed in the proof of consistency that

$$\mathbb{M}(\theta_0) - \mathbb{M}(\theta) \geq Cd^2(\theta_0, \theta)$$

and that

$$\mathbb{M}_n(\hat{\theta}_n) - \mathbb{M}_n(\theta_0) \geq I_{1,n} + I_{2,n},$$

where $I_{1,n} = (\mathbb{P}_n - P)[l(\beta_0, \phi_{0,n}, \psi_{0,n}; X) - l(\beta_0, \phi_0, \psi_0; X)]$ and $I_{2,n} = P[l(\beta_0, \phi_{0,n}, \psi_{0,n}; X) - l(\beta_0, \phi_0, \psi_0; X)]$. Applying a Taylor expansion leads to

$$\begin{aligned} I_{1,n} &= (\mathbb{P}_n - P)\left[\dot{l}_{2,\phi}(\beta_0, \tilde{\phi}, \tilde{\psi}; X)(\phi_{0,n} - \phi_0) + \dot{l}_{2,\psi}(\beta_0, \tilde{\phi}, \tilde{\psi}; X)(\psi_{0,n} - \psi_0)\right] \\ &= n^{-\sigma\nu+\epsilon}(\mathbb{P}_n - P)\left[\dot{l}_{2,\phi}(\beta_0, \tilde{\phi}, \tilde{\psi}; X)\frac{\phi_{0,n} - \phi_0}{n^{-\sigma\nu+\epsilon}} + \dot{l}_{2,\psi}(\beta_0, \tilde{\phi}, \tilde{\psi}; X)\frac{\psi_{0,n} - \psi_0}{n^{-\sigma\nu+\epsilon}}\right] \end{aligned}$$

for any $0 < \epsilon < 1/2 - \sigma\nu$. The uniform boundedness of $\dot{l}_{2,\phi}(\beta_0, \tilde{\phi}, \tilde{\psi}; X)$ due to the Conditions C1-C4 and the fact that $\|\phi_{0,n} - \phi_0\|_\infty = O(n^{-p\nu}) = O(n^{-\sigma\nu})$ (Lu, 2007) and $\|\psi_{0,n} - \psi_0\|_\infty = O(n^{-r\nu/2}) = O(n^{-\sigma\nu})$ (Lorentz, 1986) lead to

$$P\left[\dot{l}_{2,\phi}(\beta_0, \tilde{\phi}, \tilde{\psi}; X)\frac{\phi_{0,n} - \phi_0}{n^{-\sigma\nu+\epsilon}}\right]^2 \to 0, \quad P\left[\dot{l}_{2,\psi}(\beta_0, \tilde{\phi}, \tilde{\psi}; X)\frac{\psi_{0,n} - \psi_0}{n^{-\sigma\nu+\epsilon}}\right]^2 \to 0.$$

Next, we consider the class of functions $\mathcal{F}_2 = \{l(\beta_0, \phi, \psi; X) - l(\beta_0, \phi_0, \psi_0; X) : \phi_j \in \mathcal{M}_{j,n}, \|\phi_j - \phi_{0,j}\|_\Phi \leq Cn^{-p\nu}, \psi_{je} \in \mathcal{W}_{je,n}, \|\psi_{je} - \psi_{0,je}\|_\Psi \leq Cn^{-r\nu/2}$ for any $j = 1, \ldots, k,\ e = 1, \ldots, q\}$, and construct a set of ϵ-brackets with $L_2(P)$-norm bounded by

$(1/\epsilon)^{\sum_{j=1}^{k} C_j m + \sum_{j=1}^{k}\sum_{e=1}^{q} C_{je}(m_w+1)} = (1/\epsilon)^{mK_1+(m_w+1)K_2}$. Consequently, the corresponding bracketing integral

$$\begin{aligned} J_{[]}(1,\mathcal{F}_2,L_2(P)) &\equiv \int_0^1 \sqrt{\log N_{[]}(\epsilon,\mathcal{F}_2,L_2(P))}d\epsilon \\ &\le \sqrt{mK_1+(m_w+1)K_2}\int_0^1 \sqrt{\log(1/\epsilon)}d\epsilon \\ &= \sqrt{mK_1+(m_w+1)K_2}\int_0^\infty u^{1/2}e^{-u}du \\ &< \infty, \end{aligned}$$

because the integral above is bounded by 1 (Kosorok, 2008). Hence, using Theorem 19.5 of van der Vaart (2000), $\mathcal{F}_2$ is P-Donsker. Since $\mathcal{F}_2$ is P-Donsker and by Corollary 2.3.12 of van der Vaart and Wellner (1996), we have that

$$(\mathbb{P}_n - P)\left[\dot{l}_2(\beta_0,\tilde{\phi},\tilde{\psi};X)\frac{\phi_{0,n}-\phi_0}{n^{-\sigma\nu+\epsilon}} + \dot{l}_2(\beta_0,\tilde{\phi},\tilde{\psi};X)\frac{\psi_{0,n}-\psi_0}{n^{-\sigma\nu+\epsilon}}\right] = o_p(n^{-1/2}).$$

Consequently,

$$I_{1,n} = o_p(n^{-\sigma\nu+\epsilon}n^{-1/2}) = o_p(n^{-2\sigma\nu}).$$

Since the function $m(x) = x\log x - x + 1$ is bounded by $\le (x-1)^2$ in a neighbourhood of $x = 1$, it follows that

$$\mathbb{M}(\theta_0) - \mathbb{M}(\theta_{0,n}) \le C\{\|\phi_0 - \phi_{0,n}\|_\Phi^2 + \|\psi_0 - \psi_{0,n}\|_\Psi^2\} = O(n^{-2\sigma\nu})$$

and thus

$$I_{2,n} = \mathbb{M}(\theta_{0,n}) - \mathbb{M}(\theta_0) \ge -O(n^{-2\sigma\nu}).$$

Consequently,

$$\mathbb{M}_n(\hat{\theta}_n) - \mathbb{M}_n(\theta_0) \ge -O_p(n^{-2\sigma\nu}) = -O_p(n^{-2\min(\sigma\nu,(1-\nu)/2)}).$$

Defining the class of functions $\mathcal{F}_3(\eta) = \{\ell(\theta;X) - \ell(\theta_0;X) : \phi_j \in \mathcal{M}_n^j, \psi_{je} \in \mathcal{W}_n^{je}$, for any $j = 1,\dots,k$, $e = 1,\dots,q$, and $d(\theta,\theta_0) \le \eta\}$, and using similar arguments as in the proofs of consistency, we have that

$$\log N_{[]}(\epsilon,\mathcal{F}_3(\eta),L_2(P)) \le C\{kd + kq(m_w+1) + km\}\log(\eta/\epsilon)$$

and thus

$$\begin{aligned} J_{[]}(\eta,\mathcal{F}_3(\eta),L_2(P)) &= \int_0^\eta \sqrt{1+\log N_{[]}(\epsilon,\mathcal{F}_3(\eta),L_2(P))}d\epsilon \\ &\le CQ_n\int_0^\eta \sqrt{\log(\eta/\epsilon)}d\epsilon = CQ_n\{\Gamma(3/2)\}\eta, \end{aligned}$$

where $Q_n = (kd + kq(m_w+1) + km)^{1/2}$. Now, using the uniform boundedness of $l(\theta;X)$ as a result of Conditions C1-C4, and Theorem 3.4.1 of van der Vaart and Wellner (1996), the function $\phi_n(\eta)$ of Theorem 3.2.5 (van der Vaart and Wellner, 1996) is

$$\phi_n(\eta) = Q_n\eta + \frac{Q_n^2}{n^{1/2}}.$$

Also, noticing that $Q_n = O(n^\nu)$, we have that

$$n^{2\sigma\nu}\phi_n(1/n^{\sigma\nu}) = n^{\sigma\nu}n^{\nu/2} + \frac{n^{2\sigma\nu}n^\nu}{n^{1/2}} = n^{1/2}\left[n^{\sigma\nu-(1-\nu)/2} + n^{2\sigma\nu-(1-\nu)}\right].$$

Consequently, if $\sigma\nu \le (1-\nu)/2$, then $n^{2\sigma\nu}\phi_n(1/n^{\sigma\nu}) \le n^{1/2}$. Based on this and in accordance with Zhang et al. (2010), we conclude that if we choose $r_n = n^{\min(\sigma\nu,(1-\nu)/2)}$, then $r_n^2\phi_n(1/r_n) \le n^{1/2}$ and $\mathbb{M}_n(\hat{\theta}_n) - \mathbb{M}_n(\theta_0) \ge -O_p(r_n^{-2})$. Therefore

$$d(\hat{\theta}_n,\theta_0) = O_p(r_n^{-1}). \qquad \blacksquare$$

Proof of Theorem 3: Asymptotic Normality

The presence of more than one nonparametric component in the marginal model for each competing risk event precludes us from using Theorem 3 of Zhang et al. (2010) to prove the asymptotic normality of the estimator for Eucidean parameter $\hat{\beta}_n$. Instead, we use the method in Chen et al. (2006), or Zhou et al. (2017).

Let Ω be the linear span of $\Theta - \theta_0$, and $\delta_n = n^{-\min\{(1-\nu)/2, \sigma\nu\}}$ denote the rate of convergence obtained in Theorem 2, where $\sigma = \min(p, r/2)$. For any $\theta \in \{\theta \in \Theta : d(\theta, \theta_0) = O(\delta_n)\}$, define the first order directional derivative of $\ell(\theta, X)$ at the direction $\omega \in \Omega$, given as

$$\dot{\ell}(\theta, X)[\omega] = \left.\frac{d\ell(\theta + s\omega, X)}{ds}\right|_{s=0}$$

and the second order directional derivative at the directions $\omega, \tilde{\omega} \in \Omega$ as

$$\ddot{\ell}(\theta, X)[\omega, \tilde{\omega}] = \left.\left.\frac{d^2\ell(\theta + s\omega + \tilde{s}\tilde{v}, X)}{d\tilde{s}ds}\right|_{s=0}\right|_{\tilde{s}=0} = \left.\frac{d\dot{\ell}(\theta + \tilde{s}\tilde{\omega}, X)[\tilde{\omega}]}{d\tilde{s}}\right|_{\tilde{s}=0}.$$

Define the Fisher inner product on the space Ω as $\langle \omega, \tilde{\omega} \rangle = P\left\{\dot{\ell}\left(\theta_0, X\right)[\omega]\dot{\ell}\left(\theta_0, X\right)[\tilde{\omega}]\right\}$, and the Fisher norm for $\omega \in \Omega$ as $\|\omega\|^2 = \langle \omega, \omega \rangle$. For a vector $b = (b_1^\top, \dots, b_k^\top)^\top$ of dimension $k \times d$ with $\|b\| \leq 1$, we define a smooth functional of θ as $\eta(\theta) = b_1^\top \beta_1 + \cdots + b_k^\top \beta_k$. Then, for any $\omega^\top = (\omega_\beta^\top, \tilde{\phi}_1, \dots, \tilde{\phi}_k, \tilde{\psi}_{11}, \dots, \tilde{\psi}_{1q}, \dots, \tilde{\psi}_{k1}, \dots, \tilde{\psi}_{kq})$, where $\omega_\beta^\top = (\omega_{11}, \dots, \omega_{k1}, \dots, \omega_{k1}, \dots, \omega_{kd})$, we have

$$\dot{\eta}\left(\theta_0\right)[\omega] = \left.\frac{d\eta\left(\theta_0 + s\omega\right)}{ds}\right|_{s=0} = b'\omega_\beta.$$

Let $\overline{\Omega}$ be the closure of the linear span Ω under the Fisher norm. Then, under the Fisher norm $(\overline{\omega}, \|\cdot\|)$ is a Hilbert space. By the Riesz representation theorem (Hartig, 1983), there exists $\omega^* \in \overline{\Omega}$, such that $\dot{\eta}\left(\theta_0\right)[\omega] = \langle \omega, \omega^* \rangle$ for all $\omega \in \overline{\Omega}$ and $\|\omega^*\|^2 = \|\dot{\eta}\left(\theta_0\right)\|^2$, where $\omega^{*\top} = (\omega_\beta^*, \tilde{\phi}_1^*, \dots, \tilde{\phi}_k^*, \tilde{\psi}_{11}^*, \dots, \tilde{\psi}_{1q}^*, \dots, \tilde{\psi}_{k1}^*, \dots, \tilde{\psi}_{kq}^*)$ is derived in the sequel. We observe that $b^\top(\hat{\beta}_n - \beta_0) = \eta(\hat{\theta}_n) - \eta\left(\theta_0\right) = \dot{\eta}(\theta_0)[\hat{\theta}_n - \theta_0] = \langle \hat{\theta}_n - \theta_0, \omega^* \rangle$. Therefore, it follows from the Cramér-Wold device that to prove Theorem 3, it suffices to show that

$$\sqrt{n}\langle \hat{\theta}_n - \theta_0, \omega^* \rangle \to_d N(0, b^\top I^{-1}(\beta_0)b). \tag{SM12}$$

In fact, this holds, since we can show that $\sqrt{n}\langle \hat{\theta}_n - \theta_0, \omega^* \rangle \to_d N(0, \|\omega^*\|^2)$ and $\|\omega^*\|^2 = b^\top I^{-1}(\beta_0)b$. In the following, we prove these two results.

To prove the first result, we first note that by Condition C4, the result of Lu (2007) and Theorem 1.6.2 of Lorentz (1986), there exists $\Pi_n\omega^* \in \Theta_n - \theta_0$, such that $\|\Pi_n\omega^* - \omega^*\| = O(n^{-\nu\sigma})$. Moreover, under the assumptions $\sigma > 1$ and $\nu > 1/(4\sigma)$ in Condition C8, we obtain $\delta_n\|\Pi_n\omega^* - \omega^*\| = o(n^{-1/2})$. Define $r[\theta - \theta_0, X] = \ell(\theta, X) - \ell(\theta_0, X) - \dot{\ell}(\theta_0, X)[\theta - \theta_0]$ and let $\varepsilon_n = o(n^{-1/2})$ be any positive sequence. Then, by the definition of $\hat{\theta}_n$ and noticing that $P\dot{\ell}(\theta_0, X)[\Pi_n\omega^*] = 0$, we have

$$
\begin{aligned}
0 &\le P_n\left\{\ell\left(\hat{\theta}_n, X\right)-\ell\left(\hat{\theta}_n \pm \varepsilon_n \Pi_n \omega^*, X\right)\right\} \\
&=\left(P_n-P\right)\left\{\ell\left(\hat{\theta}_n, X\right)-\ell\left(\hat{\theta}_n \pm \varepsilon_n \Pi_n \omega^*, X\right)\right\}+P\left\{\ell\left(\hat{\theta}_n, X\right)-\ell\left(\hat{\theta}_n \pm \varepsilon_n \Pi_n \omega^*, X\right)\right\} \\
&=\mp \varepsilon_n P_n \dot{\ell}\left(\theta_0, X\right)\left[\Pi_n \omega^*\right]+\left(P_n-P\right)\left\{r\left(\hat{\theta}_n-\theta_0, X\right)-r\left(\hat{\theta}_n \pm \varepsilon_n \Pi_n \omega^*-\theta_0, X\right)\right\} \\
&\quad+P\left\{r\left(\hat{\theta}_n-\theta_0, X\right)-r\left(\hat{\theta}_n \pm \varepsilon_n \Pi_n \omega^*-\theta_0, X\right)\right\} \\
&=\mp \varepsilon_n P_n \dot{\ell}\left(\theta_0, X\right)\left[\omega^*\right] \mp \varepsilon_n P_n \dot{l}\left(\theta_0, X\right)\left[\Pi_n \omega^*-\omega^*\right] \\
&\quad+\left(P_n-P\right)\left\{r\left(\hat{\theta}_n-\theta_0, X\right)-r\left(\hat{\theta}_n \pm \varepsilon_n \Pi_n \omega^*-\theta_0, X\right)\right\} \\
&\quad+P\left\{r\left(\hat{\theta}_n-\theta_0, X\right)-r\left(\hat{\theta}_n \pm \varepsilon_n \Pi_n \omega^*-\theta_0, X\right)\right\} \\
&=\mp \varepsilon_n P_n \dot{\ell}\left(\theta_0, X\right)\left[\omega^*\right] \mp I_1+I_2+I_3 .
\end{aligned}
$$

We can show that

$$
I_1=\varepsilon_n P_n \dot{\ell}\left(\theta_0, X\right)\left[\Pi_n \omega^*-\omega^*\right]=\varepsilon_n \times o_p\left(n^{-1/2}\right), \tag{SM13}
$$

$$
I_2=\left(P_n-P\right)\left\{r\left(\hat{\theta}_n-\theta_0, X\right)-r\left(\hat{\theta}_n \pm \varepsilon_n \Pi_n \omega^*-\theta_0, X\right)\right\}=\varepsilon_n \times o_p\left(n^{-1/2}\right), \tag{SM14}
$$

and

$$
\begin{aligned}
I_3 &=P\left\{r\left(\hat{\theta}_n-\theta_0, X\right)-r\left(\hat{\theta}_n \pm \varepsilon_n \Pi_n \omega^*-\theta_0, X\right)\right\} \\
&=\pm \varepsilon_n\left\langle\hat{\theta}_n-\theta_0, \omega^*\right\rangle+\varepsilon_n \times o_p\left(n^{-1/2}\right) .
\end{aligned} \tag{SM15}
$$

Using (SM13), (SM14) and (SM15) together with the fact that $P\dot{l}\left(\theta_0, X\right)\left[\omega^*\right]=0$, we obtain

$$
\begin{aligned}
0 &\le P_n\left\{\ell\left(\hat{\theta}_n, X\right)-\ell\left(\hat{\theta}_n \pm \varepsilon_n \Pi_n \omega^*, X\right)\right\} \\
&=\mp \varepsilon_n\left(P_n-P\right)\left\{\dot{\ell}\left(\theta_0, X\right)\left[\omega^*\right]\right\} \pm \varepsilon_n\left\langle\hat{\theta}_n-\theta_0, \omega^*\right\rangle+\varepsilon_n \times o_p\left(n^{-1/2}\right) .
\end{aligned}
$$

Hence, we have

$$
\sqrt{n}\left\langle\hat{\theta}_n-\theta_0, \omega^*\right\rangle=\sqrt{n}\left(P_n-P\right)\left\{\dot{\ell}\left(\theta_0, X\right)\left[\omega^*\right]\right\}+o_p(1).
$$

Note that $\|\omega^*\|^2=\left\|\dot{\ell}\left(\theta_0, X\right)\left[\omega^*\right]\right\|^2$. Now, by virtue of the Central Limit Theorem, we have

$$
\sqrt{n}\left\langle\hat{\theta}_n-\theta_0, \omega^*\right\rangle=\sqrt{n}\left(P_n-P\right)\left\{\dot{\ell}\left(\theta_0, X\right)\left[\omega^*\right]\right\}+o_p(1) \rightarrow_d N\left(0,\|\omega^*\|^2\right).
$$

Now, we will prove that equations (SM13), (SM14) and (SM15) hold. First, one can easily show using Conditions A1 and A3, Chebyshev's inequality, and $\|\Pi_n \omega^*-\omega^*\|=o(1)$ that $I_1=o_p\left(n^{-1/2}\right)$. To establish I_2, we have (using the Mean Value Theorem)

$$
\begin{aligned}
I_2 &=\left(P_n-P\right)\left\{\ell\left(\hat{\theta}_n, X\right)-\ell\left(\hat{\theta}_n \pm \varepsilon_n \Pi_n \omega^*, X\right) \pm \varepsilon_n \ell\left(\theta_0, X\right)\left[\Pi_n \omega^*\right]\right\} \\
&=\mp \varepsilon_n\left(P_n-P\right)\left\{\left(\dot{\ell}(\tilde{\theta}, Z)-\dot{\ell}\left(\theta_0, X\right)\right)\left[\Pi_n \omega^*\right]\right\},
\end{aligned}
$$

where $\tilde{\theta} \in \Theta_n$ lies between $\hat{\theta}_n$ and $\hat{\theta}_n \pm \varepsilon_n \Pi_n \omega^*$. Define

$$
\mathcal{F}_{1n}=\left\{\dot{\ell}(\theta, Z)\left[\Pi_n \omega^*\right]-\dot{\ell}\left(\theta_0, X\right)\left[\Pi_n \omega^*\right]: \theta \in \Theta_n\right\}.
$$

For any $\theta^j=\left(\beta^j, \phi^j, \psi^j\right) \in \Theta_n, j=1,2$, we define their distance $\tilde{d}\left(\theta^1, \theta^2\right)=\left\|\beta^1-\beta^2\right\|+\left\|\phi^1-\phi^2\right\|_{\infty}+\left\|\psi^1-\psi^2\right\|_{\infty}$. Then, by Conditions A1 and A3 with $r=2$, it follows that for any $a_1, a_2 \in \mathcal{F}_{1n}$, $|a_1-a_2|=\left|l\left(\theta^1, X\right)\left[\Pi_n \omega^*\right]-l\left(\theta^2, X\right)\left[\Pi_n \omega^*\right]\right| \le K\tilde{d}\left(\theta^1, \theta^2\right)$. By Theorem 2.7.11 of van der Vaart and Wellner (1996), we have

$$
N_{[]}\left(2K\epsilon, \mathcal{F}_{1n}, L_2(P)\right) \le N\left(\epsilon, \Theta_n, \tilde{d}\right).
$$

Also, using Lemma 2.5 of van de Geer (2000), and Lemma 2 of Chen and Shen (1998), we have

$$N\left(\epsilon, \Theta_n, \tilde{d}\right) \leq K\left(\frac{20M}{\epsilon}\right)^{d}\left(\frac{20K}{\epsilon}\right)^{k\times m}\left(\frac{20K}{\epsilon}\right)^{k\times q\times(m_w+1)}.$$

The definitions of $N\left(\epsilon, \Theta_n, \tilde{d}\right)$ and $N_{[]}\left(2K\epsilon, \mathcal{F}_{1n}, L_2(P)\right)$ can be found in Section 2.1 of van der Vaart and Wellner (1996). If $\epsilon \in [\max(20M, 20K), \infty)$, then $N\left(\epsilon, \Theta_n, \tilde{d}\right) = 1$. Then, via simple calculations, we obtain

$$\begin{aligned}&\int_0^\infty \sqrt{\log N_{\mathbb{D}}\left(2K\epsilon, \mathcal{F}_1, L_2(P)\right)}d\epsilon \\ &\leq \int_0^{20M} d\log\tfrac{20M}{\epsilon}d\epsilon + \int_0^{20K}(k\times m)\log\tfrac{20K}{\epsilon}d\epsilon + \int_0^{20K}\{k\times q\times(m_w+1)\}\log\tfrac{20K}{\epsilon}d\epsilon < \infty.\end{aligned}$$

Now, from Theorem 2.8.4 of van der Vaart and Wellner (1996), we know that $\mathcal{F}_{1n}$ is Donsker class. Hence, following Corollary 2.3.12 of van der Vaart and Wellner (1996), we have $I_2 = \varepsilon_n \times\, o_p\left(n^{-1/2}\right)$.

To establish I_3, note that

$$\begin{aligned}P\left\{r\left(\theta-\theta_0, X\right)\right\} =& P\left\{\ell(\theta, Z) - \ell\left(\theta_0, X\right) - \dot{\ell}\left(\theta_0, X\right)\left[\theta-\theta_0\right]\right\} \\ =& 2^{-1}P\left\{\ddot{\ell}(\tilde{\theta}, Z)\left[\theta-\theta_0, \theta-\theta_0\right] - \ddot{\ell}\left(\theta_0, X\right)\left[\theta-\theta_0, \theta-\theta_0\right]\right\} \\ &+ 2^{-1}P\left\{\ddot{\ell}\left(\theta_0, X\right)\left[\theta-\theta_0, \theta-\theta_0\right]\right\} \\ =& 2^{-1}P\left\{\dot{\ell}\left(\theta_0, X\right)\left[\theta-\theta_0, \theta-\theta_0\right]\right\} + \varepsilon_n \times o_p\left(n^{-1/2}\right),\end{aligned}$$

where $\tilde{\theta} \in \Theta_n$ lies between θ_0 and θ, with the last equality following from a Taylor expansion and Conditions C1-C4 and C8. Also, with $\|\omega\|^2 = -P\left(\ddot{\ell}\left(\theta_0, X\right)[\omega, \omega]\right)$ and $\sigma > 1$, we have

$$\begin{aligned}I_3 &= -2^{-1}\left\{\left\|\hat{\theta}_n - \theta_0\right\|^2 - \left\|\hat{\theta}_n \pm \varepsilon_n \Pi_n \omega^* - \theta_0\right\|^2\right\} + \varepsilon_n \times o_p\left(n^{-1/2}\right) \\ &= \pm\varepsilon_n\left\langle\hat{\theta}_n - \theta_0, \Pi_n\omega^*\right\rangle + 2^{-1}\left\|\varepsilon_n\Pi_n\omega^*\right\|^2 + \varepsilon_n \times o_p\left(n^{-1/2}\right) \\ &= \pm\varepsilon_n\left\langle\hat{\theta}_n - \theta_0, \omega^*\right\rangle + 2^{-1}\left\|\varepsilon_n\Pi_n\omega^*\right\|^2 + \varepsilon_n \times o_p\left(n^{-1/2}\right) \\ &= \pm\varepsilon_n\left\langle\hat{\theta}_n - \theta_0, \omega^*\right\rangle + \varepsilon_n \times o_p\left(n^{-1/2}\right),\end{aligned}$$

where, the last equality holds because of $\delta_n\left\|\Pi_n\omega^* - \omega^*\right\| = o_p\left(n^{-1/2}\right)$, the use of the Cauchy-Schwartz inequality, $\left\|\Pi_n\omega^*\right\|^2 \to \left\|\omega^*\right\|^2$ and $\varepsilon_n = o\left(n^{-1/2}\right)$. Thus, we have established (SM13), (SM14) and (SM15). Hence, the former claim that $\sqrt{n}\left\langle\hat{\theta}_n - \theta_0, \omega^*\right\rangle \to_d N\left(0, \left\|\omega^*\right\|^2\right)$ holds.

Next, we prove $\left\|\omega^*\right\|^2 = b'\Sigma b$. To do so, we start from calculating the efficient score S_β and the information matrix $I(\beta_0)$ in Theorem 3. Generically, the likelihood for an individual observation is

$$\begin{aligned}l(\beta, \phi; X) &= \sum_{j=1}^{k}\delta_j^1\log\left[F_j(V; Z, \alpha_j)\right] + \sum_{j=1}^{k}\delta_j\log\left[F_j(U; Z, \alpha_j) - F_j(V; Z, \alpha_j)\right] \\ &+ (1-\delta)\log\left[1 - \sum_{j=1}^{k}F_j(U; Z, \alpha_j)\right],\end{aligned}$$

where

$$F_j(t; Z, \alpha_j) = \begin{cases} 1 - \left\{1 + \alpha_j\exp\left[\phi_j(t) + \beta_j^\top Z\right]\right\}^{-\frac{1}{\alpha_j}} & \text{if } 0 < \alpha_j < \infty \\ \\ 1 - \exp\{-\exp\left[\phi_j(t) + \beta_j^\top Z\right]\} & \text{if } \alpha_j = 0.\end{cases}$$

We will carry out the calculations for the proportional subdistribution hazards model (i.e., $\alpha_j = 0$). The proof for any other combination of link functions with $0 \leq \alpha_j < \infty$, follows from similar arguments and calculations. Let $\gamma_j(Z, W; \beta_j, \psi_j) =$

$e^{\beta_j^\top Z+\sum_{e=1}^q \psi_{je}(W_e)}$, where $\beta_j = (\beta_{j1}, \ldots, \beta_{jd})^\top$ and $\psi_j = (\psi_{j1}, \ldots, \psi_{jq})^\top$. Then, the likelihood for an individual observation under $\alpha_j = 0$ for all $j = 1, \ldots, k$ is:

$$
\begin{aligned}
\ell(\beta,\phi,\psi;X) &= \sum_{j=1}^k \delta_j \log\left[e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)} - e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}\right] \\
&+ \sum_{j=1}^k \delta_j^1 \log\left[1 - e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)}\right] \\
&+ (1-\delta)\log\left\{1 - \sum_{j=1}^k \left[1 - e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}\right]\right\},
\end{aligned}
$$

where $\Lambda_j = \exp(\phi_j)$ is the cumulative sub-distribution hazard for the j-th cause of failure (Fine and Gray, 1999). The score function for β_j $(j = 1, \ldots, k)$ is

$$
\begin{aligned}
\dot{\ell}_{\beta_j} &= \delta_j Z\gamma_j(Z,W;\beta_j,\psi_j)\frac{e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)}[-\Lambda_j(V)] - e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}[-\Lambda_j(U)]}{e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)} - e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}} \\
&+\delta_j^1 Z\gamma_j(Z,W;\beta_j,\psi_j)\frac{e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)}\Lambda_j(V)}{1 - e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)}} \\
&+(1-\delta)Z\gamma_j(Z,W;\beta_j,\psi_j)\frac{e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}[-\Lambda_j(U)]}{1 - \sum_{j=1}^k\left[1 - e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}\right]}
\end{aligned}
$$

$$
\begin{aligned}
&= -\delta_j Z\gamma_j(Z,W;\beta_j,\psi_j)\frac{\Lambda_j(V)e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)} - \Lambda_j(U)e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}}{e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)} - e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}} \\
&+\delta_j^1 Z\gamma_j(Z,W;\beta_j,\psi_j)\frac{\Lambda_j(V)e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)}}{1 - e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)}} \\
&-(1-\delta)Z\gamma_j(Z,W;\beta_j,\psi_j)\frac{\Lambda_j(U)e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}}{1 - \sum_{j=1}^k\left[1 - e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}\right]}.
\end{aligned}
$$

Let $h_j = \frac{\partial\Lambda_{j,s}}{\partial s}\Big|_{s=0}$. Then, the score operator for Λ_j $(j = 1, \ldots, k)$ is

$$
\begin{aligned}
\dot{\ell}_s^j[h_j] &= -\delta_j \gamma_j(Z,W;\beta_j,\psi_j)\frac{h_j(V)e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)} - h_j(U)e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}}{e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)} - e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}} \\
&+\delta_j^1 \gamma_j(Z,W;\beta_j,\psi_j)\frac{h_j(V)e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)}}{1 - e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)}} \\
&-(1-\delta)\gamma_j(Z,W;\beta_j,\psi_j)\frac{h_j(U)e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}}{1 - \sum_{j=1}^k\left[1 - e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}\right]}.
\end{aligned}
$$

Similarly, let $h_{je} = \frac{\partial\psi_{je,s}}{\partial s}\Big|_{s=0}$. The corresponding score operator for ψ_{je} $(j = 1, \ldots, k, e = 1, \ldots, q)$ is

$$
\begin{aligned}
\dot{\ell}_s^{je}[h_{je}] &= -\delta_j h_{je}(W_e)\gamma_j(Z,W;\beta_j,\psi_j)\frac{\Lambda_j(V)e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)} - \Lambda_j(U)e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}}{e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)} - e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}} \\
&+\delta_j^1 h_{je}(W_e)\gamma_j(Z,W;\beta_j,\psi_j)\frac{\Lambda_j(V)e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)}}{1 - e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)}} \\
&-(1-\delta)h_{je}(W_e)\gamma_j(Z,W;\beta_j,\psi_j)\frac{\Lambda_j(U)e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}}{1 - \sum_{j=1}^k\left[1 - e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}\right]}.
\end{aligned}
$$

Let

$$\begin{aligned}\alpha_1^j &= \frac{e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)}}{1-e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)}},\ \alpha_2^j = \frac{e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)}}{e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)}-e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}} = \alpha_3^j+1,\\ \alpha_3^j &= \frac{e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}}{e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)}-e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}},\ \alpha_4^j = \frac{e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}}{1-\sum_{j=1}^k\left[1-e^{-\Lambda_j(U)\gamma_j(Z,W;\beta_j,\psi_j)}\right]},\\ \frac{\alpha_1^j}{1+\alpha_1^j} &= e^{-\Lambda_j(V)\gamma_j(Z,W;\beta_j,\psi_j)}.\end{aligned}$$

Then,

$$\begin{aligned}\dot\ell_{\beta_j} &= Z\gamma_j(Z,W;\beta_j,\psi_j)\left[-\delta_j\alpha_3^j\Lambda_j(U)+\delta_j\alpha_2^j\Lambda_j(V)+\delta_j^1\alpha_1^j\Lambda_j(V)-(1-\delta)\Lambda_j(U)\alpha_4^j\right]\\ &= Z\gamma_j(Z,W;\beta_j,\psi_j)\left\{\delta_j^1\Lambda_j(V)\alpha_1^j-\delta_j\left[\Lambda_j(V)\alpha_3^j-\Lambda_j(U)\alpha_2^j\right]-(1-\delta)\Lambda_j(U)\alpha_4^j\right\},\end{aligned}$$

$$\begin{aligned}\dot\ell_s^{je}[h_{je}] &= h_{je}(W_e)\gamma_j(Z,W;\beta_j,\psi_j)\left[-\delta_j\alpha_3^j\Lambda_j(U)+\delta_j\alpha_2^j\Lambda_j(V)+\delta_j^1\alpha_1^j\Lambda_j(V)-(1-\delta)\Lambda_j(U)\alpha_4^j\right]\\ &= h_{je}(W_e)\gamma_j(Z,W;\beta_j,\psi_j)\left\{\delta_j^1\Lambda_j(V)\alpha_1^j-\delta_j\left[\Lambda_j(V)\alpha_3^j-\Lambda_j(U)\alpha_2^j\right]-(1-\delta)\Lambda_j(U)\alpha_4^j\right\},\end{aligned}$$

and

$$\dot\ell_s^j[h_j] = \gamma_j(Z,W;\beta_j,\psi_j)\left\{\delta_j^1h_j(V)\alpha_1^j-\delta_j\left[h_j(V)\alpha_3^j-h_j(U)\alpha_2^j\right]-(1-\delta)h_j(U)\alpha_4^j\right\}.$$

For each component of β_{ju}, $j=1,\ldots,k$, $u=1,\ldots,d$, we denote by $\tilde\phi_{ju}^{*\top}=(h_{1,ju}^*,\ldots,h_{k,ju}^*)$, the value of $\tilde\phi_{ju}^\top=(h_{1,ju},\ldots,h_{k,ju})$, and $\tilde\psi_{ju}^{*\top}=(h_{11,ju}^*,\ldots,h_{1q,ju}^*,\ldots,h_{k1,ju}^*,\ldots,h_{kq,ju}^*)$, the value of $\tilde\psi_{ju}^\top=(h_{11,ju},\ldots,h_{1q,ju},\ldots,h_{k1,ju},\ldots,h_{kq,ju})$ minimizing

$$E\{\dot\ell_{\beta_j}\cdot e_u-\dot\ell_s^1[h_{1,ju}]-\cdots-\dot\ell_s^k[h_{k,ju}]-\dot\ell_s^{11}[h_{11,ju}]-\cdots-\dot\ell_s^{1q}[h_{1q,ju}]-\cdots-\dot\ell_s^{k1}[h_{k1,ju}]-\cdots-\dot\ell_s^{kq}[h_{kq,ju}]\},$$

where, e_u is a d-dimensional vector with 1 in the element u and zeros elsewhere. The (ju)-th element of S_β has the form $\dot\ell_{\beta_j}\cdot e_u-\dot\ell_s^1[h_{1,ju}]-\cdots-\dot\ell_s^k[h_{k,ju}]-\dot\ell_s^{11}[h_{11,ju}]-\cdots-\dot\ell_s^{1q}[h_{1q,ju}]-\cdots-\dot\ell_s^{k1}[h_{k1,ju}]-\cdots-\dot\ell_s^{kq}[h_{kq,ju}]$. We define $I(\beta_0)$ as $E(S_\beta S_\beta^\top)$. Then, by Condition C9, the semiparametric information matrix $I(\beta_0)$ is positive definite, with $\Sigma=[I(\beta_0)]^{-1}$. Let $\omega^*=(\omega_\beta^{*\top},\tilde\phi_1^*,\ldots,\tilde\phi_k^*,\tilde\psi_{11}^*,\ldots,\tilde\psi_{1q}^*,\ldots,\tilde\psi_{k1}^*,\ldots,\tilde\psi_{kq}^*)^\top$, where $\omega_\beta^*=\Sigma b$, $\tilde\phi_i^*=-(h_{i,ju}^*,j=1,\ldots,k,u=1,\ldots,d)^\top\cdot\omega_\beta^*$ for $i=1,\ldots,k$, $\tilde\psi_{ie}^*=-(h_{ie,ju}^*,j=1,\ldots,k,u=1,\ldots,d)^\top\cdot\omega_\beta^*$ for $i=1,\ldots,k$, $e=1,\ldots,q$, and $x\cdot y$ representing the inner product of two column vectors x and y. Let $\omega_\phi=(\tilde\phi_1,\ldots,\tilde\phi_k)^\top$ and $\omega_\psi=(\tilde\psi_{11},\ldots,\tilde\psi_{1q},\ldots,\tilde\psi_{k1},\ldots,\tilde\psi_{kq})^\top$. Then, following similar calculations in Chen et al. (2006) (Section 3.2), we obtain

$$\begin{aligned}\|\omega^*\|^2=\|\dot\eta(\theta_0)\|^2 &\equiv \sup_{\omega\in\overline\Omega:\|\omega\|>0}\frac{|\dot\eta(\theta_0)[\omega]|^2}{\|\omega\|^2}\\ &=\sup_{\omega_\beta}\frac{|\dot\eta(\theta_0)[\omega_\beta]|^2}{\inf_{\omega_\phi,\omega_\psi}\|\omega\|^2}\\ &=\sup_{\omega_\beta}\frac{b^\top\omega_\beta\omega_\beta^\top b}{\omega_\beta^\top E(S_\beta S_\beta^\top)\omega_\beta}\\ &=\sup_{\omega_\beta}\frac{b^\top\omega_\beta\omega_\beta^\top b}{\omega_\beta^\top I(\beta_0)\omega_\beta}\\ &=b^\top\Sigma b<\infty.\end{aligned}$$

This completes the proof of Theorem 3. ■

SM2. Simulation Studies: Scenarios 2 and 3, tables & figures for Scenarios 1-3

Scenario 2: Here, the design is similar to that in Scenario 1, except that $Z_2 = W + \varepsilon$, and $\varepsilon \sim N(0,1)$, and the true non-linear regression functions are $\psi_1(W) = 3\sin(W)$ and $\psi_2(W) = -3\sin(W)$ for the two causes of failure. Under this simulation setup, W is a common cause for both Z_2, and the survival times T_1 and T_2. Thus, it behaves as a confounder, requiring adjustment for consistent estimation of β_1 and β_2. We fit two competing risks transformation models, with the nonlinear predictor $\eta_1 = \beta_1 Z_1 + \beta_2 Z_2 + \psi_1(W)$ (true model), and the linear predictor $\eta_2 = \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 W$ (misspecified linear model), respectively, and examine the relative biases of the estimates of β_1 and β_2 under the two models. When the true value of a parameter β is β_0, the relative bias is defined as Rel.bias(%) $= [(\hat{\beta} - \beta_0)/\beta_0]100\%$. The results are summarized in Table S2. The biases under the misspecified linear model are larger than those under the true nonlinear model, with the differences increasing under larger sample sizes.

Scenario 3: As suggested by an anonymous referee, the setting of this scenario is similar to Scenario 1, except that the true non-linear regression functions corresponding to the first and second causes of failure are now $\psi_1(W) = W\sin(1.5W)$, and $\psi_2(W) = -W\sin(1.5W)$, respectively, which are more complex functions compared to Scenario 1.

Due to the complexity of these non-linear regression functions, more basis functions are needed for implementing both BP, or B-splines approaches. Here, we choose 7 basis functions in both approaches. To mitigate the enhanced computing time resulting from the use of more basis functions, we set the number of simulation runs at 100 (instead of 200 as in Scenario 1). For the sample size $n = 100$ and $n = 500$, we report the simulation results in Table S3, and Figures S9 – S16 (see Supplementary Materials). Under smaller sample sizes (such as $n = 100$), the "B-spline+B-spline" approach tends to yield larger bias and larger coverage probability in parameter estimation than the"B-spline+Bernstein polynomials" approach, while the former shows less bias in curve estimation. When sample size is large, e.g., $n = 500$, both approaches give similar and satisfactory results in parameter estimation, although the coverage probability is somewhat under or over the nominal level due to biased estimation of the standard errors. However, "B-spline+Bernstein polynomials" approach shows better performance in curve estimation.

In summary, the overall performance of Bernstein polynomials is superior to that of B-splines. On the other hand, for more complex non-linear regression functions in the partially linear transformation models, more basis functions and large sample sizes are required in either approaches to obtain satisfactory results.

Table S1: Simulation Results from Scenario 1. The values presented are the summary statistics for the parameter vector $\beta = (\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22})^\top = (0.5, -0.3, -0.5, 0.3)^\top$. The summary values correspond to AEST: Average estimates of β, MCSD: Monte Carlo standard deviation of estimates, ASE: Average standard error of estimates, and ECP: Empirical coverage probability. B-spline+Bernstein polynomial: The baseline cumulative incidence functionals $\phi_j(t)$ were approximated by B-spline functions and the non-linear regression functions $\psi_j(w)$ were approximated by BP functions. B-spline+B-spline: Both $\phi_j(t)$ and $\psi_j(w)$ were approximated by B-spline functions.

		B-spline+Bernstein polynomial				B-spline+B-spline			
n	Statistics	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
100	AEST	0.525	-0.335	-0.536	0.333	0.527	-0.335	-0.537	0.333
	MCSD	0.468	0.215	0.466	0.209	0.468	0.214	0.473	0.211
	ASE	0.446	0.214	0.441	0.213	0.484	0.233	0.481	0.232
	ECP	0.950	0.940	0.945	0.955	0.970	0.970	0.950	0.975
500	AEST	0.521	-0.305	-0.509	0.307	0.521	-0.305	-0.510	0.306
	MCSD	0.174	0.081	0.174	0.082	0.174	0.080	0.174	0.082
	ASE	0.177	0.085	0.175	0.084	0.181	0.087	0.179	0.086
	ECP	0.960	0.945	0.955	0.950	0.960	0.970	0.965	0.955

Table S2: Simulation results for Scenario 2. The values presented are the relative biases (%) of the estimates of $\beta = (\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22})^\top = (0.5, -0.3, -0.5, 0.3)^\top$, under the correct and misspecified models.

n	Model	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
100	True model	9.19	9.94	11.9	8.81
	Misspecified model	-16.2	-18.8	-18.8	-18.2
500	True model	3.99	0.85	3.11	1.17
	Misspecified model	-19.6	-22.6	-21.0	-20.9

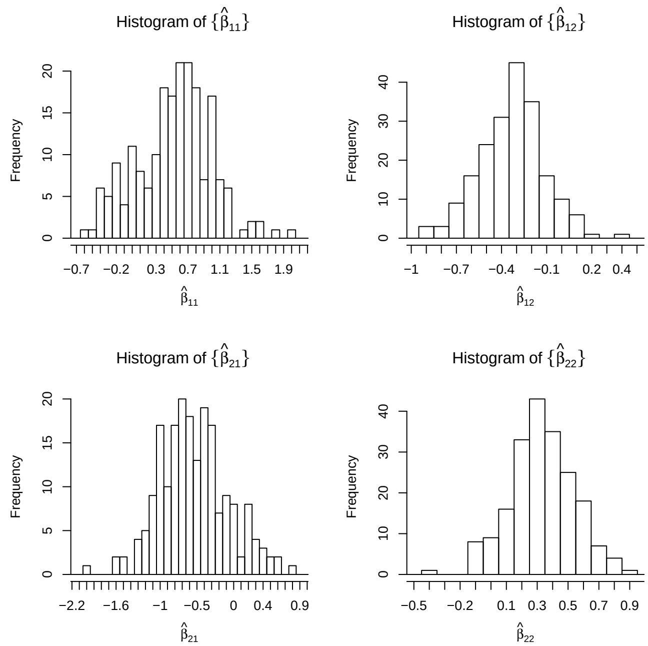

Figure S1: Scenario 1 (B-spline+Bernstein polynomial): Histograms of the estimates of β, based on 200 simulation replications, and sample size $n = 100$.

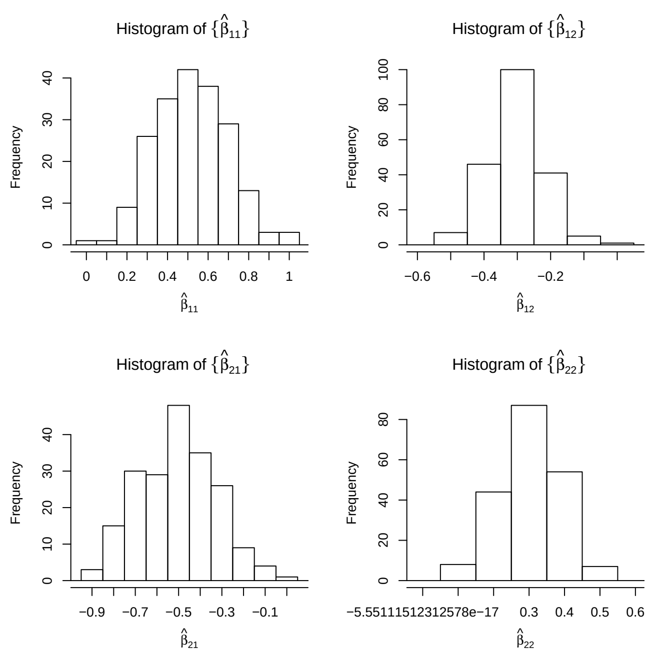

Figure S2: Scenario 1 (B-spline+Bernstein polynomial): Histograms of the estimates of β, based on 200 simulation replications, and sample size $n = 500$.

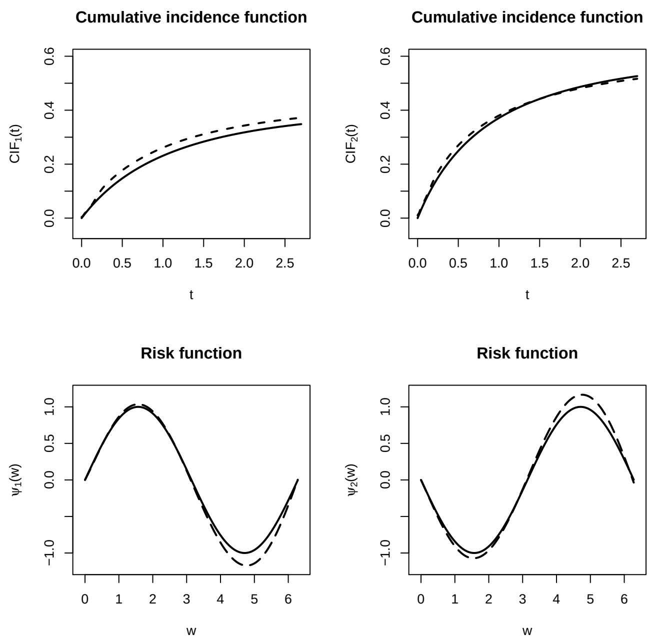

Figure S3: Scenario 1 (B-spline+Bernstein polynomial): Plots of the baseline CIFs $\phi_j(t)$ and the nonlinear risk functions $\psi_j(w)$, based on 200 simulation replications, and sample size $n = 100$. The true and average estimated functions are represented by solid, and dashed lines, respectively.

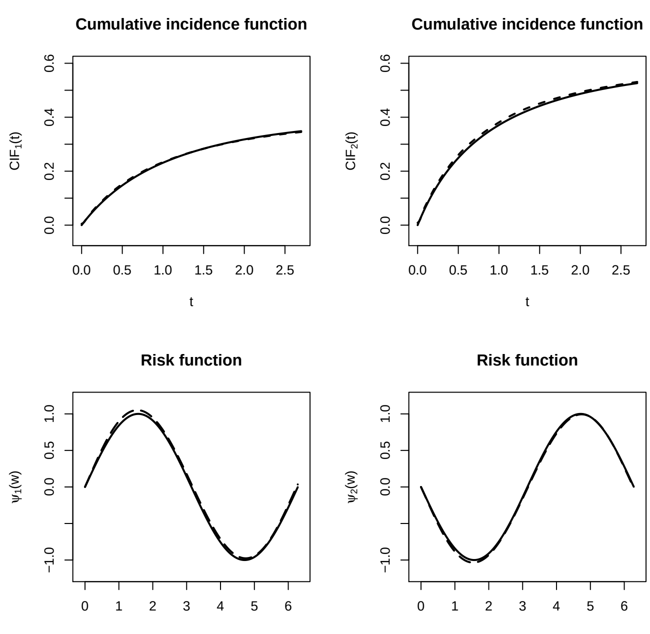

Figure S4: Scenario 1 (B-spline+Bernstein polynomial): Plots of the baseline CIFs $\phi_j(t)$ and the nonlinear risk functions $\psi_j(w)$, based on 200 simulation replications, and sample size $n = 500$. The true and average estimated functions are represented by solid, and dashed lines, respectively.

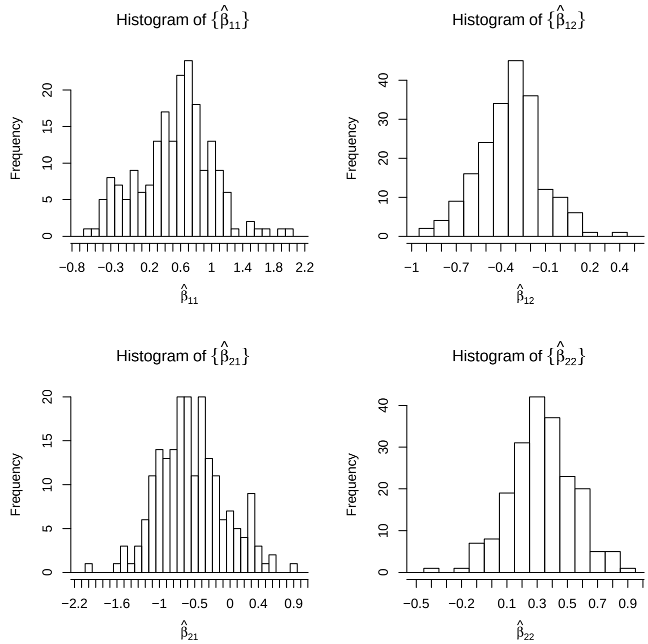

Figure S5: Scenario 1 (B-spline+B-spline): Histograms of the estimates of β, based on 200 simulation replications, and sample size $n = 100$.

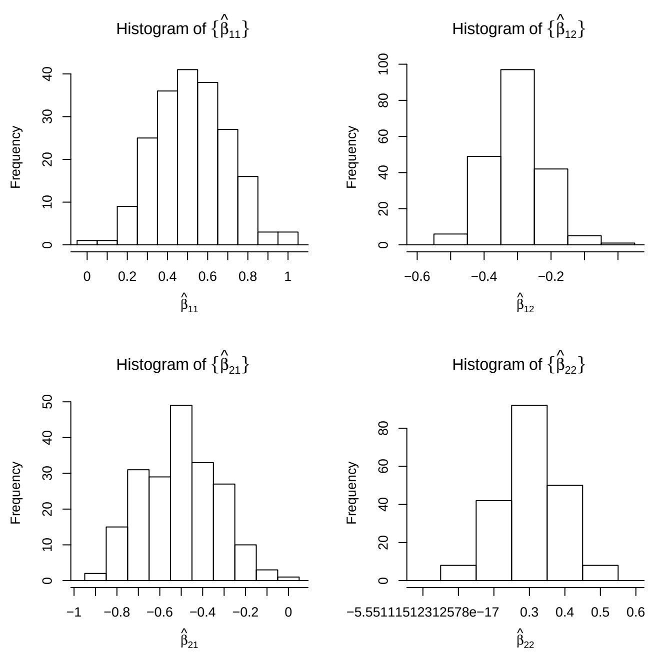

Figure S6: Scenario 1 (B-spline+B-spline): Histograms of the estimates of β, based on 200 simulation replications, and sample size $n = 500$.

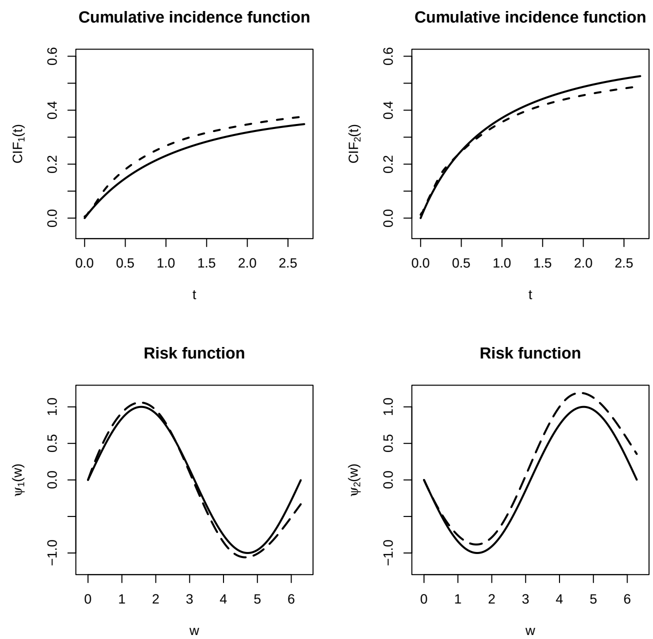

Figure S7: Scenario 1 (B-spline+B-spline): Plots of the baseline CIFs $\phi_j(t)$ and the nonlinear risk functions $\psi_j(w)$, based on 200 simulation replications, and sample size $n = 100$. The true and average estimated functions are represented by solid, and dashed lines, respectively.

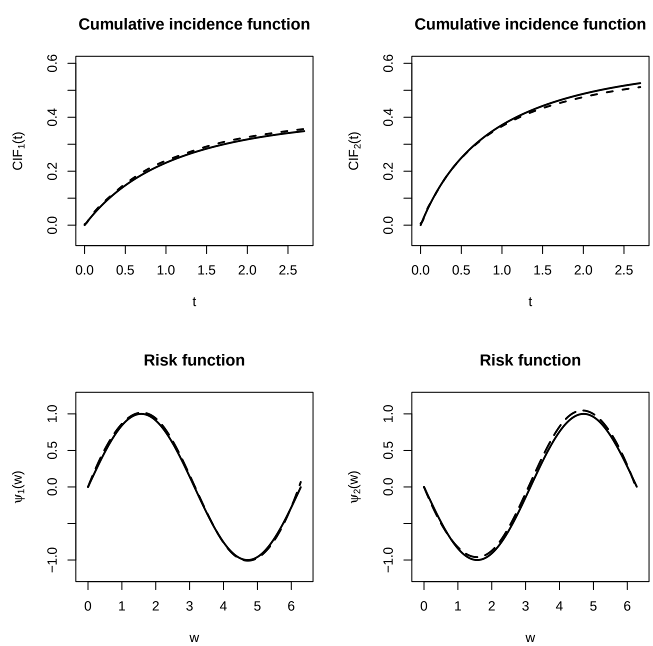

Figure S8: Scenario 1 (B-spline+B-spline): Plots of the baseline CIFs $\phi_j(t)$ and the nonlinear risk functions $\psi_j(w)$, based on 200 simulation replications, and sample size $n = 500$. The true and average estimated functions are represented by solid, and dashed lines, respectively.

Table S3: Simulation Results from Scenario 3. The values presented are the summary statistics for the parameter vector $\beta = (\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22})^\top = (0.5, -0.3, -0.5, 0.3)^\top$. The summary values correspond to AEST: Average estimates of β, MCSD: Monte Carlo standard deviation of estimates, ASE: Average standard error of estimates, and ECP: Empirical coverage probability. B-spline+Bernstein polynomial: The baseline cumulative incidence functionals $\phi_j(t)$ were approximated by B-spline functions and the non-linear regression functions $\psi_j(w)$ were approximated by BP functions. B-spline+B-spline: Both $\phi_j(t)$ and $\psi_j(w)$ were approximated by B-spline functions.

		B-spline+Bernstein polynomial				B-spline+B-spline			
n	Statistics	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
100	AEST	0.556	-0.299	-0.597	0.301	0.567	-0.323	-0.632	0.323
	MCSD	0.539	0.251	0.545	0.251	0.562	0.257	0.565	0.256
	ASE	0.434	0.214	0.449	0.215	0.570	0.272	0.586	0.273
	ECP	0.900	0.900	0.900	0.910	0.960	0.970	0.960	0.970
500	AEST	0.494	-0.289	-0.505	0.292	0.499	-0.294	-0.516	0.298
	MCSD	0.181	0.098	0.180	0.097	0.178	0.101	0.184	0.101
	ASE	0.171	0.084	0.176	0.084	0.202	0.095	0.209	0.095
	ECP	0.930	0.930	0.950	0.910	0.980	0.920	0.970	0.910

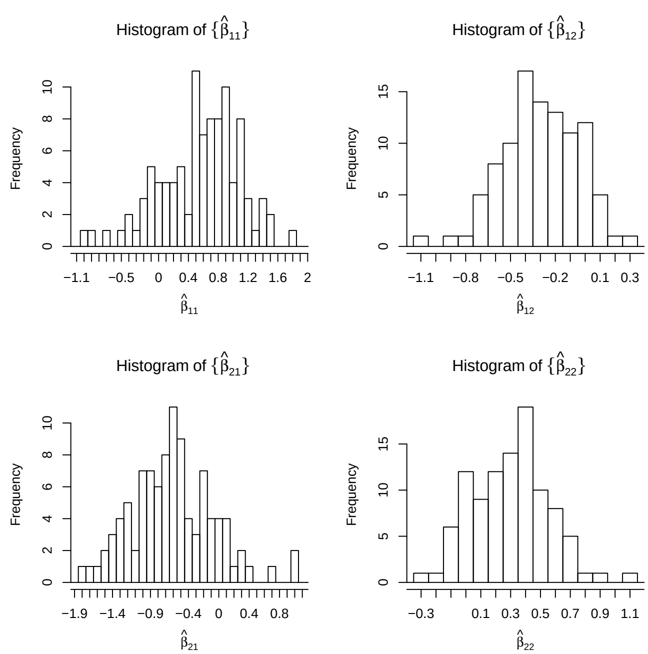

Figure S9: Scenario 3 (B-spline+Bernstein polynomial): Histograms of the estimates of β, based on 100 simulation replications, and sample size $n = 100$.

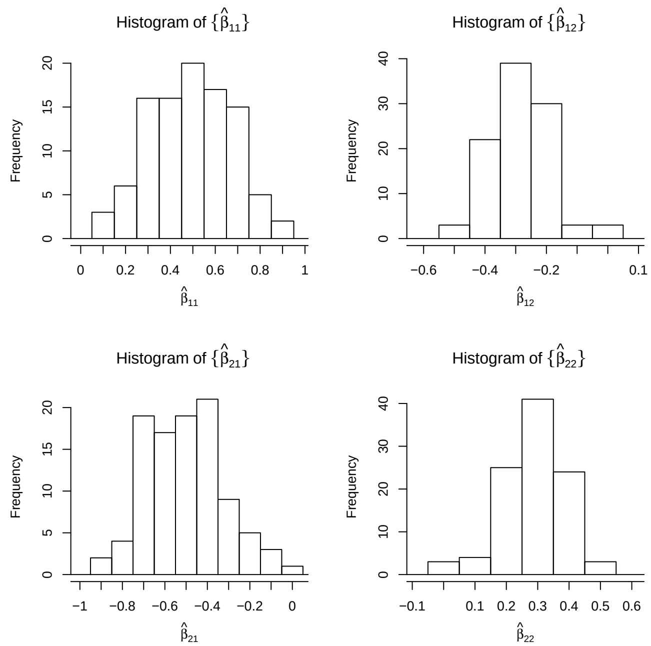

Figure S10: Scenario 3 (B-spline+Bernstein polynomial): Histograms of the estimates of β, based on 100 simulation replications, and sample size $n = 500$.

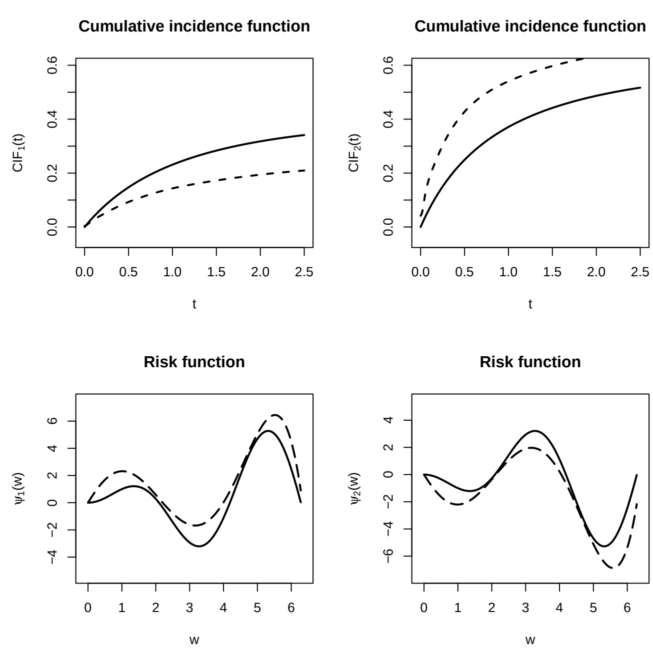

Figure S11: Scenario 3 (B-spline+Bernstein polynomial): Plots of the baseline CIFs $\phi_j(t)$ and the nonlinear risk functions $\psi_j(w)$, based on 100 simulation replications, and sample size $n = 100$. The true and average estimated functions are represented by solid, and dashed lines, respectively.

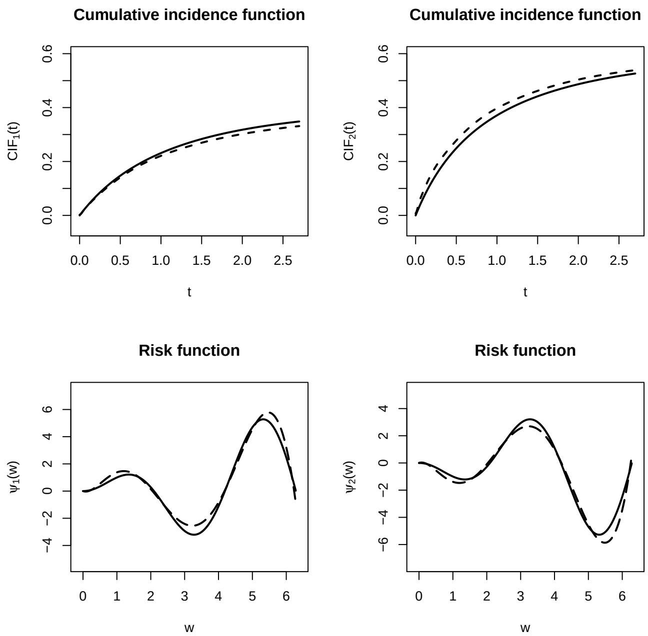

Figure S12: Scenario 3 (B-spline+Bernstein polynomial): Plots of the baseline CIFs $\phi_j(t)$ and the nonlinear risk functions $\psi_j(w)$, based on 100 simulation replications, and sample size $n = 500$. The true and average estimated functions are represented by solid, and dashed lines, respectively.

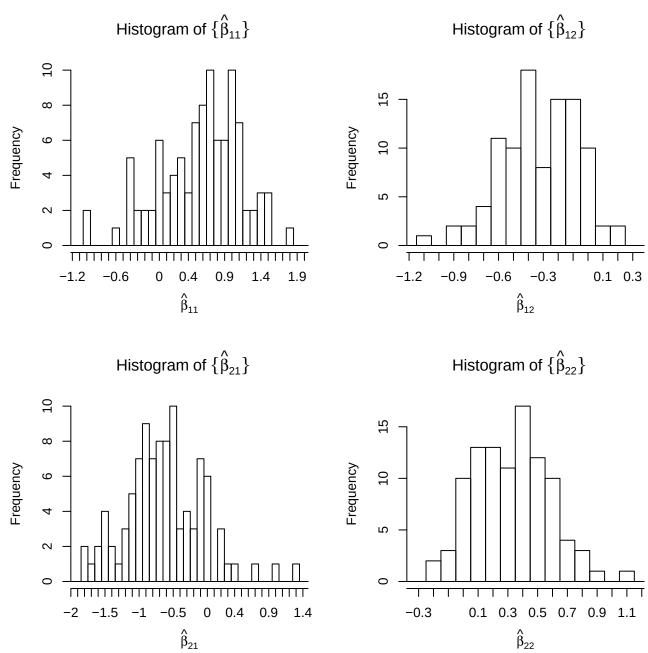

Figure S13: Scenario 3 (B-spline+B-spline): Histograms of the estimates of β, based on 100 simulation replications, and sample size $n = 100$.

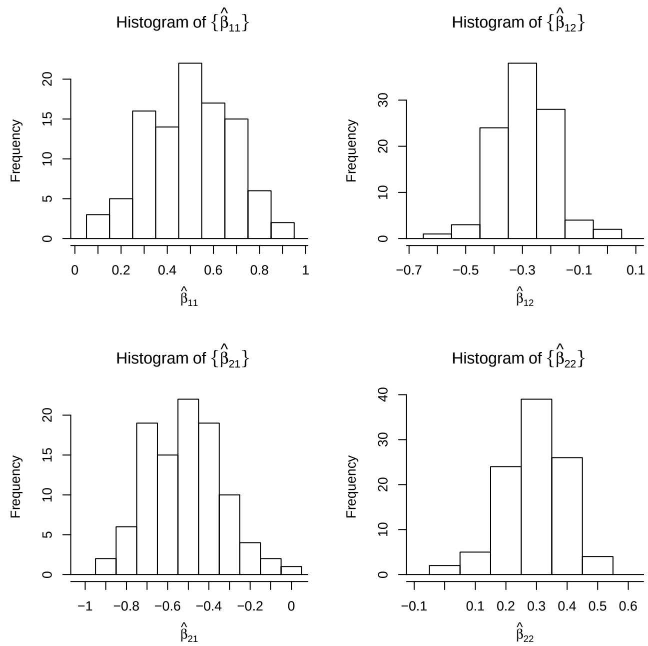

Figure S14: Scenario 3 (B-spline+B-spline): Histograms of the estimates of β, based on 100 simulation replications, and sample size $n = 500$.

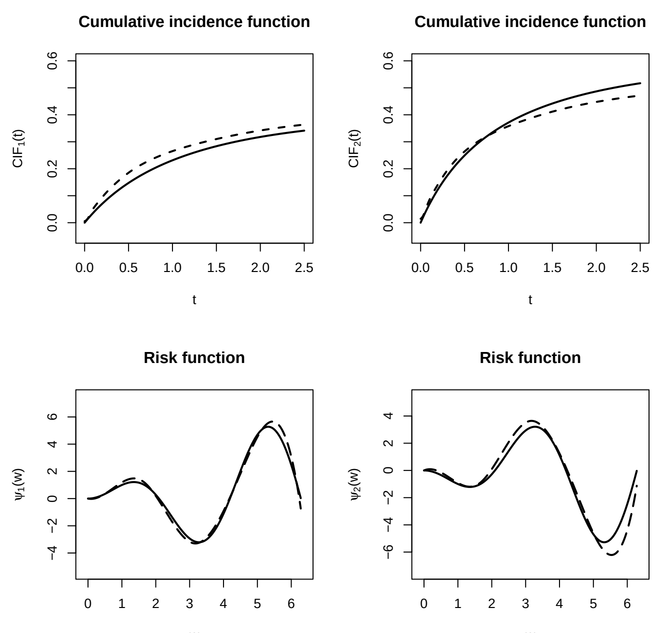

Figure S15: Scenario 3 (B-spline+B-spline): Plots of the baseline CIFs $\phi_j(t)$ and the nonlinear risk functions $\psi_j(w)$, based on 100 simulation replications, and sample size $n = 100$. The true and average estimated functions are represented by solid, and dashed lines, respectively.

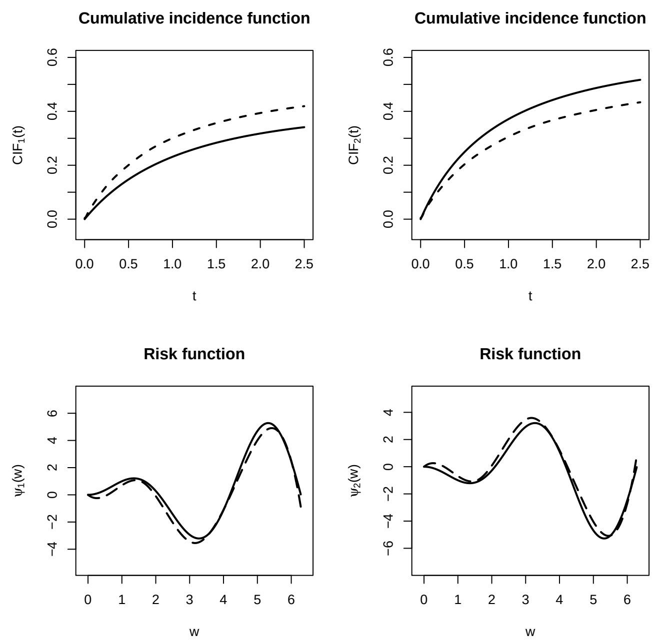

Figure S16: Scenario 3 (B-spline+B-spline): Plots of the baseline CIFs $\phi_j(t)$ and the nonlinear risk functions $\psi_j(w)$, based on 100 simulation replications, and sample size $n = 500$. The true and average estimated functions are represented by solid, and dashed lines, respectively.

Bibliography

Chen, X., Fan, Y., and Tsyrennikov, V. (2006). Efficient estimation of semiparametric multivariate copula models. *Journal of the American Statistical Association*, 101(475):1228–1240.

Chen, X. and Shen, X. (1998). Sieve extremum estimates for weakly dependent data. *Econometrica*, 66:289–314.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446):496–509.

Hartig, D. G. (1983). The Riesz representation theorem revisited. *The American Mathematical Monthly*, 90(4):277–280.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer.

Lorentz, G. G. (1986). *Bernstein Polynomials.* New York: Chelsea Publishing Co.

Lu, M. (2007). Monotone spline estimations for panel count data. *Department of Biostatistics, University of Iowa.*

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer-Verlag.

Shorack, G. R. and Wellner, J. A. (2009). *Empirical Processes with Applications to Statistics.* SIAM.

van de Geer, S. (2000). *Empirical Processes in M-estimation*, volume 6. 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 and empirical processes with applications to statistics. *Journal of the Royal Statistical Society-Series A Statistics in Society*, 160(3):596–608.

Wellner, J. A., Zhang, Y., et al. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *Annals of Statistics*, 35(5):2106–2142.

Zhang, Y., Hua, L., and Huang, J. (2010). A spline-based semiparametric maximum likelihood estimation method for the cox model with interval-censored data. *Scandinavian Journal of Statistics*, 37(2):338–354.

Zhou, Q., Hu, T., and Sun, J. (2017). A sieve semiparametric maximum likelihood approach for regression analysis of bivariate interval-censored failure time data. *Journal of the American Statistical Association*, 112(518):664–672.