

A General Approach to Maximum Likelihood Estimation of Semiparametric Models with Massive Sample Sizes

Jianqiao Wang¹, Donglin Zeng² and Dan-Yu Lin¹

¹*University of North Carolina at Chapel Hill*, ²*University of Michigan*

Supplementary Material

The Supplementary Material contains technical details of the proposed methods and additional simulation studies.

S1 Proof of Theorem 1

We first establish the convergence rate of $\dot{\tilde{\eta}}_{\theta}$.

Lemma 1. *Under Conditions 1-4, $\|\dot{\tilde{\eta}}_{\theta} + h_0\|_{\mathcal{F}} = O_P(\|\theta - \theta_0\|) + O_P(m^{-\beta})$ as $\theta \rightarrow_P \theta_0$.*

Proof. Since $\tilde{\eta}_{\theta}$ maximizes $\mathbb{P}_m l(\theta, \eta)$ over $\mathcal{F}_m$, it must satisfy $\mathbb{P}_m \dot{l}_{\eta}(\theta, \tilde{\eta}_{\theta})[h] = 0$ for every $h \in \dot{\mathcal{F}}_m$, where $\dot{\mathcal{F}}_m$ is the tangent space of $\mathcal{F}_m$. By differentiating this equation with respect to θ at θ_0 , we obtain

$$\mathbb{P}_m \ddot{l}_{\theta\eta}(\theta_0, \tilde{\eta}_{\theta_0})[h] + \mathbb{P}_m \ddot{l}_{\eta\eta}(\theta_0, \tilde{\eta}_{\theta_0})[\dot{\tilde{\eta}}_{\theta_0}, h] = 0.$$

Condition 3 implies that $\tilde{\eta}_{\theta_0} \rightarrow_P \eta_0$ and $\dot{\tilde{\eta}}_{\theta_0} \rightarrow_P \tilde{h}_0$. These facts, along with the Glivenko-Cantelli properties given in Condition 4, entail that $\mathbb{P}_m \ddot{l}_{\theta\eta}(\theta_0, \tilde{\eta}_{\theta_0})[h] \rightarrow_P E \ddot{l}_{\theta\eta}(\theta_0, \eta_0)[h]$ and $\mathbb{P}_m \ddot{l}_{\eta\eta}(\theta_0, \tilde{\eta}_{\theta_0})[\dot{\tilde{\eta}}_{\theta_0}, h] \rightarrow_P E \ddot{l}_{\eta\eta}(\theta_0, \eta_0)[\tilde{h}_0, h]$. By taking the limit of the left side of the above equation, we have

$$E \ddot{l}_{\theta\eta}(\theta_0, \eta_0)[h] + E \ddot{l}_{\eta\eta}(\theta_0, \eta_0)[\tilde{h}_0, h] = 0 \tag{S1.1}$$

for any $h \in \dot{\mathcal{F}}_m$. For any $h' \in \mathcal{F}$, there exists a sequence $\{h_m\} \subset \dot{\mathcal{F}}_m$ that converges to h' .

Thus, (S1.1) holds for any $h \in \dot{\mathcal{F}}$. The uniqueness of the least favorable direction provided in Condition 2 yields $\tilde{h}_0 = -h_0$. It then follows from Condition 3 that as $\boldsymbol{\theta} \rightarrow_P \boldsymbol{\theta}_0$, $\dot{\tilde{\eta}}_{\boldsymbol{\theta}}$ converges in probability to $-h_0$. Thus, Lemma 1 holds. $\square$

We are now ready to prove Theorem 1.

Proof. For simplicity, we assume that $\boldsymbol{\theta}$ is one-dimensional. Let $\hat{\boldsymbol{\theta}}_f$ and $\hat{\boldsymbol{\theta}}_{f,\delta_n}$ denote the final one-step estimators using $D_{\boldsymbol{\theta}}l(\mathbf{O}; \tilde{\boldsymbol{\theta}}_m, \tilde{\eta}_{\tilde{\boldsymbol{\theta}}_m})$ and $D_{\boldsymbol{\theta},\delta_n}l(\mathbf{O}; \tilde{\boldsymbol{\theta}}_m, \tilde{\eta}_{\tilde{\boldsymbol{\theta}}_m})$, respectively. By Condition 6, $\hat{\boldsymbol{\theta}}_{f,\delta_n} = \hat{\boldsymbol{\theta}}_f + o_P(n^{-1/2})$. Thus, it suffices to show that $n^{1/2}(\hat{\boldsymbol{\theta}}_f - \boldsymbol{\theta}_0) \rightarrow \mathbf{N}(0, \mathbf{I}_0^{-1})$.

Clearly, $\mathbb{P}_m \left\{ \dot{l}_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}_m, \tilde{\eta}_{\tilde{\boldsymbol{\theta}}_m}) + \dot{l}_{\eta}(\tilde{\boldsymbol{\theta}}_m, \tilde{\eta}_{\tilde{\boldsymbol{\theta}}_m})[\dot{\tilde{\eta}}_{\tilde{\boldsymbol{\theta}}_m}] \right\} = 0$. Thus,

$$\hat{\boldsymbol{\theta}}_f = \tilde{\boldsymbol{\theta}}_m + \left(\tilde{\mathbf{I}}_m/m \right)^{-1} \mathbb{P}_n \left\{ \dot{l}_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}_m, \tilde{\eta}_{\tilde{\boldsymbol{\theta}}_m}) + \dot{l}_{\eta}(\tilde{\boldsymbol{\theta}}_m, \tilde{\eta}_{\tilde{\boldsymbol{\theta}}_m})[\dot{\tilde{\eta}}_{\tilde{\boldsymbol{\theta}}_m}] \right\}.$$

Let $\mathbb{G}_n \equiv n^{1/2}(\mathbb{P}_n - P_{\boldsymbol{\theta}_0, \eta_0})$ denote the empirical process with $\mathbf{O}_1, \dots, \mathbf{O}_n$. Note that

$$\begin{aligned} & \mathbb{P}_n \left\{ \dot{l}_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}_m, \tilde{\eta}_{\tilde{\boldsymbol{\theta}}_m}) + \dot{l}_{\eta}(\tilde{\boldsymbol{\theta}}_m, \tilde{\eta}_{\tilde{\boldsymbol{\theta}}_m})[\dot{\tilde{\eta}}_{\tilde{\boldsymbol{\theta}}_m}] \right\} \\ &= n^{-1/2} \mathbb{G}_n \left\{ \dot{l}_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}_m, \tilde{\eta}_{\tilde{\boldsymbol{\theta}}_m}) + \dot{l}_{\eta}(\tilde{\boldsymbol{\theta}}_m, \tilde{\eta}_{\tilde{\boldsymbol{\theta}}_m})[\dot{\tilde{\eta}}_{\tilde{\boldsymbol{\theta}}_m}] \right\} \\ &+ E \left\{ \dot{l}_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}_m, \tilde{\eta}_{\tilde{\boldsymbol{\theta}}_m}) + \dot{l}_{\eta}(\tilde{\boldsymbol{\theta}}_m, \tilde{\eta}_{\tilde{\boldsymbol{\theta}}_m})[\dot{\tilde{\eta}}_{\tilde{\boldsymbol{\theta}}_m}] \right\}. \end{aligned}$$

Let $\mathbf{l}_0(\mathbf{O}; \boldsymbol{\theta}, \eta_0)$ denote the efficient score function for $\boldsymbol{\theta}_0$ with $\mathbf{l}_0(\mathbf{O}; \boldsymbol{\theta}, \eta_0) = \dot{l}_{\boldsymbol{\theta}}(\mathbf{O}; \boldsymbol{\theta}, \eta_0) - \dot{l}_{\eta}(\mathbf{O}; \boldsymbol{\theta}, \eta_0)[h_0]$. By Lemma 1, $\dot{\tilde{\eta}}_{\tilde{\boldsymbol{\theta}}_m}$ converges in probability to $-h_0$. In addition, $(\tilde{\boldsymbol{\theta}}_m, \tilde{\eta}_{\tilde{\boldsymbol{\theta}}_m}) \rightarrow_P (\boldsymbol{\theta}_0, \eta_0)$ by Condition 3, and $\dot{l}_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}_m, \tilde{\eta}_{\tilde{\boldsymbol{\theta}}_m}) + \dot{l}_{\eta}(\tilde{\boldsymbol{\theta}}_m, \tilde{\eta}_{\tilde{\boldsymbol{\theta}}_m})[\dot{\tilde{\eta}}_{\tilde{\boldsymbol{\theta}}_m}]$ is dominated by a square-integrable function under Condition 4. Thus, the dominated convergence theorem implies that

$$E \left\{ \dot{l}_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}_m, \tilde{\eta}_{\tilde{\boldsymbol{\theta}}_m}) + \dot{l}_{\eta}(\tilde{\boldsymbol{\theta}}_m, \tilde{\eta}_{\tilde{\boldsymbol{\theta}}_m})[\dot{\tilde{\eta}}_{\tilde{\boldsymbol{\theta}}_m}] - \mathbf{l}_0 \right\}^2 \rightarrow 0.$$

The Donsker property given in Condition 4 then yields

$$\mathbb{G}_n \left\{ \dot{l}_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}_m, \tilde{\eta}_{\tilde{\boldsymbol{\theta}}_m}) + \dot{l}_{\eta}(\tilde{\boldsymbol{\theta}}_m, \tilde{\eta}_{\tilde{\boldsymbol{\theta}}_m})[\dot{\tilde{\eta}}_{\tilde{\boldsymbol{\theta}}_m}] \right\} = \mathbb{G}_n \mathbf{l}_0 + o_P(1).$$

Therefore,

$$\begin{aligned} m^{-1} \tilde{\mathbf{I}}_m n^{1/2} (\hat{\boldsymbol{\theta}}_f - \boldsymbol{\theta}_0) &= m^{-1} \tilde{\mathbf{I}}_m n^{1/2} (\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0) + \mathbb{G}_n \mathbf{l}_0 \\ &\quad + n^{1/2} E \left\{ \dot{l}_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}_m, \tilde{\boldsymbol{\eta}}_{\tilde{\boldsymbol{\theta}}_m}) + \dot{l}_{\boldsymbol{\eta}}(\tilde{\boldsymbol{\theta}}_m, \tilde{\boldsymbol{\eta}}_{\tilde{\boldsymbol{\theta}}_m}) [\dot{\tilde{\boldsymbol{\eta}}}_{\tilde{\boldsymbol{\theta}}_m}] \right\} + o_P(1). \end{aligned}$$

By the Taylor series expansions of $E \left\{ \dot{l}_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}_m, \tilde{\boldsymbol{\eta}}_{\tilde{\boldsymbol{\theta}}_m}) + \dot{l}_{\boldsymbol{\eta}}(\tilde{\boldsymbol{\theta}}_m, \tilde{\boldsymbol{\eta}}_{\tilde{\boldsymbol{\theta}}_m}) [\dot{\tilde{\boldsymbol{\eta}}}_{\tilde{\boldsymbol{\theta}}_m}] \right\}$,

$$\begin{aligned} &m^{-1} \tilde{\mathbf{I}}_m n^{1/2} (\hat{\boldsymbol{\theta}}_f - \boldsymbol{\theta}_0) \\ &= \mathbb{G}_n \mathbf{l}_0 + m^{-1} \tilde{\mathbf{I}}_m n^{1/2} (\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0) + n^{1/2} E \left\{ \dot{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}_0}) + \dot{l}_{\boldsymbol{\eta}}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}_0}) [\dot{\tilde{\boldsymbol{\eta}}}_{\boldsymbol{\theta}_0}] \right\} \\ &\quad + n^{1/2} E \left\{ \ddot{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}_0}) + 2\ddot{l}_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}_0}) [\dot{\tilde{\boldsymbol{\eta}}}_{\boldsymbol{\theta}_0}] + \ddot{l}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}_0}) [\dot{\tilde{\boldsymbol{\eta}}}_{\boldsymbol{\theta}_0}, \dot{\tilde{\boldsymbol{\eta}}}_{\boldsymbol{\theta}_0}] \right. \\ &\quad \left. + \dot{l}_{\boldsymbol{\eta}}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}_0}) [\ddot{\tilde{\boldsymbol{\eta}}}_{\boldsymbol{\theta}_0}] \right\} (\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0) + n^{1/2} O_P(\|\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0\|^2) + o_P(1) \\ &= \mathbb{G}_n \mathbf{l}_0 + n^{1/2} O_P(\|\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0\|^2) + o_P(1) + n^{1/2} E \left\{ \dot{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}_0}) + \dot{l}_{\boldsymbol{\eta}}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}_0}) [\dot{\tilde{\boldsymbol{\eta}}}_{\boldsymbol{\theta}_0}] \right\} \\ &\quad + n^{1/2} \left[m^{-1} \tilde{\mathbf{I}}_m + E \left\{ \ddot{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}_0}) + 2\ddot{l}_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}_0}) [\dot{\tilde{\boldsymbol{\eta}}}_{\boldsymbol{\theta}_0}] \right. \right. \\ &\quad \left. \left. + \ddot{l}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}_0}) [\dot{\tilde{\boldsymbol{\eta}}}_{\boldsymbol{\theta}_0}, \dot{\tilde{\boldsymbol{\eta}}}_{\boldsymbol{\theta}_0}] + \dot{l}_{\boldsymbol{\eta}}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}_0}) [\ddot{\tilde{\boldsymbol{\eta}}}_{\boldsymbol{\theta}_0}] \right\} \right] (\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0). \end{aligned}$$

Since $\tilde{\boldsymbol{\theta}}_m$ is the maximum profile likelihood estimator, $m^{1/2}(\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0) \rightarrow_d \mathbf{N}(0, \mathbf{I}_0^{-1})$. Under Conditions 3 and 5, $\tilde{\mathbf{I}}_m/m \rightarrow_P \mathbf{I}_0$ and $n/m^2 \rightarrow 0$. Therefore, the conclusion of Theorem 1 will follow from Slutsky's theorem if the following results hold:

$$E \left\{ \dot{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}_0}) + \dot{l}_{\boldsymbol{\eta}}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}_0}) [\dot{\tilde{\boldsymbol{\eta}}}_{\boldsymbol{\theta}_0}] \right\} = o_P(n^{-1/2}), \quad (\text{S1.2})$$

and

$$\begin{aligned} &m^{-1} \tilde{\mathbf{I}}_m + E \left\{ \ddot{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}_0}) + 2\ddot{l}_{\boldsymbol{\theta}\boldsymbol{\eta}}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}_0}) [\dot{\tilde{\boldsymbol{\eta}}}_{\boldsymbol{\theta}_0}] + \ddot{l}_{\boldsymbol{\eta}\boldsymbol{\eta}}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}_0}) [\dot{\tilde{\boldsymbol{\eta}}}_{\boldsymbol{\theta}_0}, \dot{\tilde{\boldsymbol{\eta}}}_{\boldsymbol{\theta}_0}] + \dot{l}_{\boldsymbol{\eta}}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}_0}) [\ddot{\tilde{\boldsymbol{\eta}}}_{\boldsymbol{\theta}_0}] \right\} \\ &= o_P(m^{1/2} n^{-1/2}). \end{aligned}$$

To verify (S1.2), we apply the Taylor series expansions and Bartlett identities to obtain

$$E \left\{ \dot{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}_0}) + \dot{l}_{\boldsymbol{\eta}}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_{\boldsymbol{\theta}_0}) [\dot{\tilde{\boldsymbol{\eta}}}_{\boldsymbol{\theta}_0}] \right\}$$

$$\begin{aligned}
&= E \left\{ \dot{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \eta_0) + \dot{l}_{\eta}(\boldsymbol{\theta}_0, \eta_0)[\dot{\tilde{\eta}}_{\boldsymbol{\theta}_0}] + \ddot{l}_{\boldsymbol{\theta}\eta}(\boldsymbol{\theta}_0, \eta_0)[\tilde{\eta}_{\boldsymbol{\theta}_0} - \eta_0] + \ddot{l}_{\eta\eta}(\boldsymbol{\theta}_0, \eta_0)[\dot{\tilde{\eta}}_{\boldsymbol{\theta}_0}, \tilde{\eta}_{\boldsymbol{\theta}_0} - \eta_0] \right\} \\
&\quad + O_P(\|\tilde{\eta}_{\boldsymbol{\theta}_0} - \eta_0\|_{\mathcal{F}}^2). \\
&= E \left\{ \ddot{l}_{\boldsymbol{\theta}\eta}(\boldsymbol{\theta}_0, \eta_0)[\tilde{\eta}_{\boldsymbol{\theta}_0} - \eta_0] + \ddot{l}_{\eta\eta}(\boldsymbol{\theta}_0, \eta_0)[\dot{\tilde{\eta}}_{\boldsymbol{\theta}_0}, \tilde{\eta}_{\boldsymbol{\theta}_0} - \eta_0] \right\} + O_P(\|\tilde{\eta}_{\boldsymbol{\theta}_0} - \eta_0\|_{\mathcal{F}}^2).
\end{aligned}$$

Under Condition 2, the least favorable direction h_0 solves the equation

$$E\ddot{l}_{\boldsymbol{\theta}\eta}(\boldsymbol{\theta}_0, \eta_0)[h] - E\ddot{l}_{\eta\eta}(\boldsymbol{\theta}_0, \eta_0)[h_0, h] = 0 \quad (\text{S1.3})$$

for every $h \in \mathcal{F}$. Thus,

$$E \left\{ \dot{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \tilde{\eta}_{\boldsymbol{\theta}_0}) + \dot{l}_{\eta}(\boldsymbol{\theta}_0, \tilde{\eta}_{\boldsymbol{\theta}_0})[\dot{\tilde{\eta}}_{\boldsymbol{\theta}_0}] \right\} = E\ddot{l}_{\eta\eta}(\boldsymbol{\theta}_0, \eta_0)[\dot{\tilde{\eta}}_{\boldsymbol{\theta}_0} + h_0, \tilde{\eta}_{\boldsymbol{\theta}_0} - \eta_0] + O_P(\|\tilde{\eta}_{\boldsymbol{\theta}_0} - \eta_0\|_{\mathcal{F}}^2).$$

Under Condition 3, $\|\tilde{\eta}_{\boldsymbol{\theta}_0} - \eta_0\|_{\mathcal{F}} = O_P(m^{-\alpha})$. Condition 1 implies that $E\ddot{l}_{\eta\eta}(\boldsymbol{\theta}_0, \eta_0)[\cdot, \cdot]$ is a bounded bilinear operator. These facts, together with Lemma 1, entail that

$$E \left\{ \dot{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \tilde{\eta}_{\boldsymbol{\theta}_0}) + \dot{l}_{\eta}(\boldsymbol{\theta}_0, \tilde{\eta}_{\boldsymbol{\theta}_0})[\dot{\tilde{\eta}}_{\boldsymbol{\theta}_0}] \right\} = O_P(m^{-2\alpha} + m^{-\alpha-\beta}). \quad (\text{S1.4})$$

Equation (S1.2) then follows from the fact that $n^{1/2}/m^{2\alpha} \rightarrow 0$ and $n^{1/2}/m^{\alpha+\beta}$ under Condition 5.

Next, we verify (S1.3). By the Taylor series expansion and Bartlett identities,

$$\begin{aligned}
&E \left\{ \ddot{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \tilde{\eta}_{\boldsymbol{\theta}_0}) + 2\ddot{l}_{\boldsymbol{\theta}\eta}(\boldsymbol{\theta}_0, \tilde{\eta}_{\boldsymbol{\theta}_0})[\dot{\tilde{\eta}}_{\boldsymbol{\theta}_0}] + \ddot{l}_{\eta\eta}(\boldsymbol{\theta}_0, \tilde{\eta}_{\boldsymbol{\theta}_0})[\dot{\tilde{\eta}}_{\boldsymbol{\theta}_0}, \dot{\tilde{\eta}}_{\boldsymbol{\theta}_0}] + \dot{l}_{\eta}(\boldsymbol{\theta}_0, \tilde{\eta}_{\boldsymbol{\theta}_0})[\ddot{\tilde{\eta}}_{\boldsymbol{\theta}_0}] \right\} \\
&= E \left\{ \ddot{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \eta_0) + 2\ddot{l}_{\boldsymbol{\theta}\eta}(\boldsymbol{\theta}_0, \eta_0)[\dot{\tilde{\eta}}_{\boldsymbol{\theta}_0}] + \ddot{l}_{\eta\eta}(\boldsymbol{\theta}_0, \eta_0)[\dot{\tilde{\eta}}_{\boldsymbol{\theta}_0}, \dot{\tilde{\eta}}_{\boldsymbol{\theta}_0}] + \dot{l}_{\eta}(\boldsymbol{\theta}_0, \eta_0)[\ddot{\tilde{\eta}}_{\boldsymbol{\theta}_0}] \right\} \\
&\quad + O_P(\|\tilde{\eta}_{\boldsymbol{\theta}_0} - \eta_0\|_{\mathcal{F}}) \\
&= E \left\{ \ddot{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \eta_0) + 2\ddot{l}_{\boldsymbol{\theta}\eta}(\boldsymbol{\theta}_0, \eta_0)[\dot{\tilde{\eta}}_{\boldsymbol{\theta}_0}] + \ddot{l}_{\eta\eta}(\boldsymbol{\theta}_0, \eta_0)[\dot{\tilde{\eta}}_{\boldsymbol{\theta}_0}, \dot{\tilde{\eta}}_{\boldsymbol{\theta}_0}] \right\} + O_P(\|\tilde{\eta}_{\boldsymbol{\theta}_0} - \eta_0\|_{\mathcal{F}}).
\end{aligned}$$

The efficient information matrix $\mathbf{I}_0$ takes the form

$$\mathbf{I}_0 = E\mathbf{l}_0\mathbf{l}_0^T = -E \left\{ \ddot{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \eta_0) - 2\ddot{l}_{\boldsymbol{\theta}\eta}(\boldsymbol{\theta}_0, \eta_0)[h_0] + \ddot{l}_{\eta\eta}(\boldsymbol{\theta}_0, \eta_0)[h_0, h_0] \right\}.$$

Thus,

$$\begin{aligned}
& E \left\{ \ddot{l}_{\theta\theta}(\boldsymbol{\theta}_0, \tilde{\eta}_{\theta_0}) + 2\ddot{l}_{\theta\eta}(\boldsymbol{\theta}_0, \tilde{\eta}_{\theta_0})[\dot{\tilde{\eta}}_{\theta_0}] + \ddot{l}_{\eta\eta}(\boldsymbol{\theta}_0, \tilde{\eta}_{\theta_0})[\dot{\tilde{\eta}}_{\theta_0}, \dot{\tilde{\eta}}_{\theta_0}] + \dot{l}_{\eta}(\boldsymbol{\theta}_0, \tilde{\eta}_{\theta_0})[\ddot{\tilde{\eta}}_{\theta_0}] \right\} \\
&= -\mathbf{I}_0 + E \left\{ 2\ddot{l}_{\theta\eta}(\boldsymbol{\theta}_0, \eta_0)[\dot{\tilde{\eta}}_{\theta_0} + h_0] + \ddot{l}_{\eta\eta}(\boldsymbol{\theta}_0, \eta_0)[\dot{\tilde{\eta}}_{\theta_0} + h_0, \dot{\tilde{\eta}}_{\theta_0} - h_0] \right\} \\
&\quad + O_P(\|\tilde{\eta}_{\theta_0} - \eta_0\|_{\mathcal{F}}) \\
&= -\mathbf{I}_0 + E\ddot{l}_{\eta\eta}(\boldsymbol{\theta}_0, \eta_0)[\dot{\tilde{\eta}}_{\theta_0} + h_0, \dot{\tilde{\eta}}_{\theta_0} + h_0] + O_P(\|\tilde{\eta}_{\theta_0} - \eta_0\|_{\mathcal{F}}).
\end{aligned}$$

The second equality comes from (S1.3). By the arguments in proving (S1.4), we can show that

$$\begin{aligned}
& E \left\{ \ddot{l}_{\theta\theta}(\boldsymbol{\theta}_0, \tilde{\eta}_{\theta_0}) + 2\ddot{l}_{\theta\eta}(\boldsymbol{\theta}_0, \tilde{\eta}_{\theta_0})[\dot{\tilde{\eta}}_{\theta_0}] + \ddot{l}_{\eta\eta}(\boldsymbol{\theta}_0, \tilde{\eta}_{\theta_0})[\dot{\tilde{\eta}}_{\theta_0}, \dot{\tilde{\eta}}_{\theta_0}] + \dot{l}_{\eta}(\boldsymbol{\theta}_0, \tilde{\eta}_{\theta_0})[\ddot{\tilde{\eta}}_{\theta_0}] \right\} \\
&= -\mathbf{I}_0 + O_P(m^{-\alpha} + m^{-2\beta}).
\end{aligned}$$

Under Condition 3, $\tilde{\mathbf{I}}_m/m = \mathbf{I}_0 + O_P(m^{-\gamma})$, implying that

$$\begin{aligned}
& m^{-1}\tilde{\mathbf{I}}_m + E \left\{ \ddot{l}_{\theta\theta}(\boldsymbol{\theta}_0, \tilde{\eta}_{\theta_0}) + 2\ddot{l}_{\theta\eta}(\boldsymbol{\theta}_0, \tilde{\eta}_{\theta_0})[\dot{\tilde{\eta}}_{\theta_0}] + \ddot{l}_{\eta\eta}(\boldsymbol{\theta}_0, \tilde{\eta}_{\theta_0})[\dot{\tilde{\eta}}_{\theta_0}, \dot{\tilde{\eta}}_{\theta_0}] + \dot{l}_{\eta}(\boldsymbol{\theta}_0, \tilde{\eta}_{\theta_0})[\ddot{\tilde{\eta}}_{\theta_0}] \right\} \\
&= O_P(m^{-\alpha} + m^{-2\beta} + m^{-\gamma}).
\end{aligned}$$

Equation (S1.3) then follows from the fact that $n^{1/2}/m^{1/2+\alpha} \rightarrow 0$, $n^{1/2}/m^{1/2+2\beta} \rightarrow 0$, and $n^{1/2}/m^{1/2+\gamma} \rightarrow 0$ under Condition 5. $\square$

S2 Technical details on the verification of regularity conditions

S2.1 Partially linear model

Following Chen (2007), we assume the following conditions:

Condition S1. $\boldsymbol{\theta}_0$ belongs to the interior of a compact subset $B \subset \mathbb{R}^p$.

Condition S2. σ_0^2 is positive and bounded.

Condition S3. g_0 and $E(\mathbf{X} \mid Z)$ belong to $\mathcal{G}$.

Condition S4. Both $E(\mathbf{X}\mathbf{X}^\top)$ and $E[\{\mathbf{X} - E(\mathbf{X} \mid Z)\}\{\mathbf{X} - E(\mathbf{X} \mid Z)\}^\top]$ are positive definite.

Under Conditions S1–S4 and with the choice of $k_m = O\{m^{1/(2q+1)}\}$,

$$m^{1/2}(\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0) \rightarrow_d \mathbf{N}(0, \mathbf{I}_0^{-1}) \quad (\text{S2.5})$$

and

$$\|\tilde{g}_\theta - g_0\|_2 = O_P\{m^{-q/(2q+1)}\} + O_P(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|). \quad (\text{S2.6})$$

as $\boldsymbol{\theta} \rightarrow_P \boldsymbol{\theta}_0$, where $\mathbf{I}_0 = \sigma_0^{-2}[\{\mathbf{X} - E(\mathbf{X} \mid Z)\}\{\mathbf{X} - E(\mathbf{X} \mid Z)\}^\top]$ is the efficient information matrix for $\boldsymbol{\theta}$ (Chen, 2007).

We verify Conditions 1–4 under Conditions S1–S4. Let $\psi = \sigma^{-2}$, $\psi_0 = \sigma_0^{-2}$ and $\tilde{\psi}_\theta = \tilde{\sigma}_\theta^{-2}$.

We write the log-likelihood function for a single observation $\mathbf{O}$ as

$$l(\mathbf{O}; \boldsymbol{\theta}, g, \psi) = -\frac{1}{2}\psi\{Y - \boldsymbol{\theta}^\top \mathbf{X} - g(Z)\}^2 + \frac{1}{2}\log \psi.$$

Condition 1 then follows from the definition of differentiability, with $\dot{l}_\theta(\mathbf{O}; \boldsymbol{\theta}, g, \psi) = \psi\{Y - \boldsymbol{\theta}^\top \mathbf{X} - g(Z)\}x$, $\dot{l}_\psi(\mathbf{O}; \boldsymbol{\theta}, g, \psi) = \{Y - \boldsymbol{\theta}^\top \mathbf{X} - g(Z)\}^2/2 + 1/(2\psi)$, and $\dot{l}_g(\mathbf{O}; \boldsymbol{\theta}, g, \psi)[h] = \psi\{Y - \boldsymbol{\theta}^\top \mathbf{X} - g(Z)\}h$, for any $h \in \mathcal{G}$.

Let (h_0, t_0) denote the least favorable direction, which minimizes $E\|\dot{l}_\theta(\boldsymbol{\theta}_0, g_0, \psi_0) - \dot{l}_g(\boldsymbol{\theta}_0, g_0, \psi_0)[h] - \dot{l}_\psi(\boldsymbol{\theta}_0, g_0, \psi_0)t\|^2$ by definition. Condition 2 then follows from simple algebra showing that $h_0 = E(\mathbf{X} \mid Z)$ and $t_0 = 0$.

We now verify Condition 4. By the chain rule,

$$\begin{aligned} & \dot{l}_\theta(\mathbf{O}; \tilde{\boldsymbol{\theta}}_m, \tilde{g}_{\tilde{\boldsymbol{\theta}}_m}, \tilde{\psi}_{\tilde{\boldsymbol{\theta}}_m}) + \dot{l}_g(\mathbf{O}; \tilde{\boldsymbol{\theta}}_m, \tilde{g}_{\tilde{\boldsymbol{\theta}}_m}, \tilde{\psi}_{\tilde{\boldsymbol{\theta}}_m})[\dot{\tilde{g}}_{\tilde{\boldsymbol{\theta}}_m}] + \dot{l}_\psi(\mathbf{O}; \tilde{\boldsymbol{\theta}}_m, \tilde{g}_{\tilde{\boldsymbol{\theta}}_m}, \tilde{\psi}_{\tilde{\boldsymbol{\theta}}_m})[\dot{\tilde{\psi}}_{\tilde{\boldsymbol{\theta}}_m}] \\ &= \dot{\tilde{\psi}}_{\tilde{\boldsymbol{\theta}}_m}\{Y - \tilde{\boldsymbol{\theta}}_m^\top \mathbf{X} - \tilde{g}_{\tilde{\boldsymbol{\theta}}_m}(Z)\}\{\mathbf{X} + \dot{\tilde{g}}_{\tilde{\boldsymbol{\theta}}_m}(Z)\}. \end{aligned}$$

We first show that $\tilde{g}_{\tilde{\theta}_m}(Z) - g_0(Z)$ belongs to a Donsker class. We consider the function class $\mathcal{G}_{m,\delta} = \{g(Z) - g_0(Z) : \|g - g_0\|_2 \leq \delta, g \in \mathcal{G}_m\}$, where $\delta = O(k_m^{-q'})$ for some $q' \in (1/2, q)$, and $\mathcal{G}_m$ is a function class that contains polynomial splines of order r with k_m knots. Let $N_{[]}(\epsilon, \mathcal{G}_{m,\delta}, \|\cdot\|_2)$ denote the bracketing number of $\mathcal{G}_{m,\delta}$. Clearly, $\log N_{[]}(\epsilon, \mathcal{G}_{m,\delta}, \|\cdot\|_2) \lesssim (k_m + r) \log(1/\epsilon)$, where $a \lesssim b$ means that a is bounded by a constant times b (Chen, 2007).

Elementary calculus then yields

$$\begin{aligned} & \int_0^{2\delta} \sqrt{\log N_{[]}(\epsilon, \mathcal{G}_{m,\delta}, \|\cdot\|_2)} d\epsilon \\ & \lesssim (k_m + r)^{1/2} \epsilon \sqrt{\log(1/\epsilon)} - (k_m + r)^{1/2} \int_0^{\sqrt{\log(1/\epsilon)}} e^{-t^2} dt \Big|_0^{2\delta} \\ & \lesssim (k_m + r)^{1/2} 2\delta \sqrt{\log(1/2\delta)} - (k_m + r)^{1/2} \int_{\sqrt{\log(1/2\delta)}}^{+\infty} e^{-t^2} dt. \end{aligned}$$

By the concentration inequality, the right side of the above inequality converges to zero as $m \rightarrow +\infty$ with $\delta = O(k_m^{-q'})$. Theorem 2.5.6 in van der Vaart and Wellner (1996) then implies that $\mathcal{G}_{m,\delta}$ is Donsker and thus $\tilde{g}_{\tilde{\theta}_m}(Z) - g_0(Z)$ belongs to a Donsker class. By Theorem 2.7.1 and Corollary 2.7.2 in van der Vaart and Wellner (1996), $\mathcal{G}$ is a Donsker class. Therefore, $\tilde{g}_{\tilde{\theta}_m}(Z)$ belongs to a Donsker class by Theorem 2.10.6 in van der Vaart and Wellner (1996). By similar arguments, $\dot{\tilde{g}}_{\tilde{\theta}_m}(Z)$ also belongs to a Donsker class. Thus, $\tilde{\psi}_{\tilde{\theta}_m}\{Y - \tilde{\theta}_m^T \mathbf{X} - \tilde{g}_{\tilde{\theta}_m}(Z)\}\{\mathbf{X} + \dot{\tilde{g}}_{\tilde{\theta}_m}(Z)\}$ belongs to a Donsker class. Likewise, for any fixed $h \in \mathcal{G}_m$, the functions $\ddot{l}_{gg}(\mathbf{O}; \boldsymbol{\theta}_0, \tilde{g}_{\boldsymbol{\theta}_0}, \tilde{\psi}_{\boldsymbol{\theta}_0})[\tilde{g}_{\boldsymbol{\theta}_0}, h]$, $\ddot{l}_{\psi\psi}(\mathbf{O}; \boldsymbol{\theta}_0, \tilde{g}_{\boldsymbol{\theta}_0}, \tilde{\psi}_{\boldsymbol{\theta}_0})$, $\ddot{l}_{g\psi}(\mathbf{O}; \boldsymbol{\theta}_0, \tilde{g}_{\boldsymbol{\theta}_0}, \tilde{\psi}_{\boldsymbol{\theta}_0})[h]$, $\ddot{l}_{\theta g}(\mathbf{O}; \boldsymbol{\theta}_0, \tilde{g}_{\boldsymbol{\theta}_0}, \tilde{\psi}_{\boldsymbol{\theta}_0})[h]$, and $\ddot{l}_{\theta\psi}(\mathbf{O}; \boldsymbol{\theta}_0, \tilde{g}_{\boldsymbol{\theta}_0}, \tilde{\psi}_{\boldsymbol{\theta}_0})$ belong to some Donsker classes. Hence, Condition 4 holds.

Finally, we verify Condition 3. Note that

$$\tilde{\psi}_{\boldsymbol{\theta}} = \left[m^{-1} \sum_{i=1}^m \{Y_i - \boldsymbol{\theta}^T \mathbf{X}_i - \tilde{g}_{\boldsymbol{\theta}}(Z_i)\}^2 \right]^{-1}.$$

By the arguments used to verify Condition 4, we can show that $\{Y - \boldsymbol{\theta}^T \mathbf{X} - \tilde{g}_{\boldsymbol{\theta}}(Z)\}^2$ belongs

to a Donsker class in the neighbourhood of $\boldsymbol{\theta}_0$. Thus,

$$\tilde{\psi}_{\boldsymbol{\theta}} = O_P(m^{-1/2}) + E\{Y - \boldsymbol{\theta}^T \mathbf{X} - \tilde{g}_{\boldsymbol{\theta}}(Z)\}^2.$$

Equation (S2.6) then entails that $|\tilde{\psi}_{\boldsymbol{\theta}} - \psi_0| = O_P\{m^{-q/(2q+1)}\} + O_P(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|)$ as $\boldsymbol{\theta} \rightarrow_P \boldsymbol{\theta}_0$.

By the definition of differentiability, $\tilde{g}_{\boldsymbol{\theta}}$ and $\tilde{\psi}_{\boldsymbol{\theta}}$ are thrice differentiable in the neighborhood of $\boldsymbol{\theta}_0$, and the derivatives are bounded. Specifically,

$$\dot{\tilde{g}}_{\boldsymbol{\theta}}(Z) = -\mathbf{p}(Z)^T \left\{ \sum_{i \in \mathcal{I}_1} \mathbf{p}(Z_i) \mathbf{p}(Z_i)^T \right\}^{-1} \sum_{i \in \mathcal{I}_1} \mathbf{p}(Z_i) \mathbf{X}_i^T$$

and

$$\dot{\tilde{\psi}}_{\boldsymbol{\theta}} = 2\tilde{\psi}_{\boldsymbol{\theta}}^2 m^{-1} \sum_{i=1}^m \{Y_i - \boldsymbol{\theta}^T \mathbf{X}_i - \tilde{g}_{\boldsymbol{\theta}}(Z_i)\} \{\mathbf{X}_i + \dot{\tilde{g}}_{\boldsymbol{\theta}}(Z_i)\}.$$

Since $\tilde{g}_{\boldsymbol{\theta}}(Z)$ is the sieve least-squares estimator of $-E(\mathbf{X} \mid Z)$ with the polynomial splines, Condition S3 implies that

$$\|\dot{\tilde{g}}_{\boldsymbol{\theta}} + E(\mathbf{X} \mid Z)\|_{\mathcal{G}} = O_P\{m^{-q/(2q+1)}\} + O_P(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|) \quad (\text{S2.7})$$

as $\boldsymbol{\theta} \rightarrow_P \boldsymbol{\theta}_0$. Using the arguments for showing the convergence rate of $\tilde{\psi}_{\boldsymbol{\theta}}$, we can show that $\dot{\tilde{\psi}}_{\boldsymbol{\theta}} = O_P\{m^{-q/(2q+1)}\} + O_P(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|)$.

We now derive the convergence rate of $m^{-1}\tilde{\mathbf{I}}_m$. Note that $m^{-1}\tilde{\mathbf{I}}_m = \tilde{\psi}_{\tilde{\boldsymbol{\theta}}_m} m^{-1} \sum_{i=1}^m \{\mathbf{X}_i + \dot{\tilde{g}}_{\tilde{\boldsymbol{\theta}}_m}(Z_i)\} \{\mathbf{X}_i + \dot{\tilde{g}}_{\tilde{\boldsymbol{\theta}}_m}(Z_i)\}^T$. Since $\mathbf{X} + \dot{\tilde{g}}_{\tilde{\boldsymbol{\theta}}_m}(Z)$ belongs to a Donsker class, $\{\mathbf{X} + \dot{\tilde{g}}_{\tilde{\boldsymbol{\theta}}_m}(Z)\} \{\mathbf{X} + \dot{\tilde{g}}_{\tilde{\boldsymbol{\theta}}_m}(Z)\}^T$ belongs to a Donsker class. Thus,

$$m^{-1}\tilde{\mathbf{I}}_m = O_P(m^{-1/2}) + \tilde{\psi}_{\tilde{\boldsymbol{\theta}}_m} E\{\mathbf{X} + \dot{\tilde{g}}_{\tilde{\boldsymbol{\theta}}_m}(Z)\} \{\mathbf{X} + \dot{\tilde{g}}_{\tilde{\boldsymbol{\theta}}_m}(Z)\}^T.$$

Equations (S2.5) and (S2.7) then imply that $\|m^{-1}\tilde{\mathbf{I}}_m - \mathbf{I}_0\|_2 = O_P\{m^{-q/(2q+1)}\}$. Therefore, Condition 3 holds.

S2.2 Cox proportional hazards model with interval-censored data

Following Zeng et al. (2016), we assume the following conditions:

Condition S5. The true parameter value $\boldsymbol{\theta}_0$ lies in the interior of a known compact set $\mathcal{B}$ in $\mathbb{R}^p$, and the true value of $\Lambda(\cdot)$, denoted by $\Lambda_0(\cdot)$, is continuously differentiable with positive derivatives over $[0, \tau]$, which is the union of the supports of $(U_1, \dots, U_K)$.

Condition S6. The vector $\mathbf{Z}(t)$ has total bounded variation over $[0, \tau]$. In addition, if there exist a deterministic function $h_1(t)$ and a constant vector h_2 such that $h_1(t) + h_2^\top \mathbf{Z}(t) = 0$ for all $t \in [0, \tau]$ with probability 1, then $h_1(t) = 0$ for $t \in [0, \tau]$ and $h_2 = 0$.

Condition S7. The number of monitoring times, K , is positive, with $E(K) < \infty$. The conditional probability $P(U_K = \tau \mid K, \mathbf{Z})$ is greater than some positive constant c_1 . In addition, $P(U_{j+1} - U_j \geq \eta \mid K, \mathbf{Z}) = 1$ ($j = 1, \dots, K$) for some positive constant c_2 . Finally, the densities of (U_j, U_{j+1}) conditional on $\mathbf{Z}$ and K , denoted by $g_j(u, v \mid \mathbf{Z}, K)$ ($j = 1, \dots, K$), have continuous second-order partial derivatives with respect to u and v when $v - u > c_2$ and are continuously differentiable with respect to $\mathbf{Z}$.

Under Conditions S5–S7, $m^{1/2}(\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0) \rightarrow_d \mathbf{N}(0, \mathbf{I}_0^{-1})$ and $\|\tilde{\Lambda}_{\boldsymbol{\theta}} - \Lambda_0\|_2 = O_P(m^{-1/3}) + O_P(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|)$ as $\boldsymbol{\theta} \rightarrow_P \boldsymbol{\theta}_0$ (Zeng et al., 2016).

We verify Condition 3 under Conditions S5–S7. Let h_0 denote the least favorable direction. Because $\dot{\tilde{\Lambda}}_{\boldsymbol{\theta}}$ converges to $-h_0$ in probability as $\boldsymbol{\theta} \rightarrow_P \boldsymbol{\theta}_0$ (Xu et al., 2022), Condition 3 is satisfied if the following results hold:

$$\|\dot{\tilde{\Lambda}}_{\tilde{\boldsymbol{\theta}}_m} + h_0\|_2 = O_P(m^{-1/3}), \quad (\text{S2.8})$$

and

$$\|m^{-1}\tilde{\mathbf{I}}_m - \mathbf{I}_0\|_2 = O_P(m^{-1/3}). \quad (\text{S2.9})$$

To prove (S2.8), we introduce some notation:

$$Q(u, v; \mathbf{Z}, \boldsymbol{\theta}, \Lambda) = \exp \left[- \int_0^u \exp\{\boldsymbol{\theta}^\top \mathbf{Z}(t)\} d\Lambda(t) \right] - \exp \left[- \int_0^v \exp\{\boldsymbol{\theta}^\top \mathbf{Z}(t)\} d\Lambda(t) \right],$$

$$B(t, u, v; \mathbf{Z}, \boldsymbol{\theta}, \Lambda) = \frac{\exp\{\boldsymbol{\theta}^\top \mathbf{Z}(t)\}}{Q(u, v; \mathbf{Z}, \boldsymbol{\theta}, \Lambda)} \left(- \exp \left[- \int_0^u \exp\{\boldsymbol{\theta}^\top \mathbf{Z}(s)\} d\Lambda(s) \right] I(u \geq t) \right. \\ \left. + \exp \left[- \int_0^v \exp\{\boldsymbol{\theta}^\top \mathbf{Z}(s)\} d\Lambda(s) \right] I(v \geq t) \right),$$

and

$$g_0(t; \boldsymbol{\theta}, \Lambda) = \sum_{k=0}^K \Delta_k B(t, U_k, U_{k+1}; \mathbf{Z}, \boldsymbol{\theta}, \Lambda),$$

$$g_1(t; \boldsymbol{\theta}, \Lambda) = \sum_{k=0}^K \Delta_k \mathbf{Z}(t) B(t, U_k, U_{k+1}; \mathbf{Z}, \boldsymbol{\theta}, \Lambda).$$

In addition, define

$$R(t, u, v; \tau, \mathbf{Z}, \boldsymbol{\theta}, \Lambda) = \frac{\exp\{\boldsymbol{\theta}^\top \mathbf{Z}(t)\}}{Q(u, v; \mathbf{Z}, \boldsymbol{\theta}, \Lambda)} \{I(u \geq t) - I(v \geq t)\} \\ \times \exp \left[- \frac{1}{2} \int_0^\tau \exp\{\boldsymbol{\theta}^\top \mathbf{Z}(r)\} \{I(u \geq r) + I(v \geq r)\} d\Lambda(r) \right],$$

and

$$W_0(t, s; \boldsymbol{\theta}, \Lambda) = \sum_{k=0}^K \Delta_k R(t, U_k, U_{k+1}; \tau, \mathbf{Z}, \boldsymbol{\theta}, \Lambda) R(s, U_k, U_{k+1}; \tau, \mathbf{Z}, \boldsymbol{\theta}, \Lambda),$$

$$W_1(t, s; \boldsymbol{\theta}, \Lambda) = \sum_{k=0}^K \Delta_k \mathbf{Z}(t) R(t, U_k, U_{k+1}; \tau, \mathbf{Z}, \boldsymbol{\theta}, \Lambda) R(s, U_k, U_{k+1}; \tau, \mathbf{Z}, \boldsymbol{\theta}, \Lambda).$$

Finally, let

$$g_m(t) = \mathbb{P}_m \{g_1(t; \tilde{\boldsymbol{\theta}}_m, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m})\} - \int_0^\tau \mathbb{P}_m \{W_1(s, t; \tilde{\boldsymbol{\theta}}_m, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m})\} d\tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m}(s),$$

$$g(t) = E \{g_1(t; \boldsymbol{\theta}_0, \Lambda_0)\} - \int_0^\tau E \{W_1(s, t; \boldsymbol{\theta}_0, \Lambda_0)\} d\Lambda_0(s),$$

and

$$\Gamma_m(h) = \int_0^\tau \mathbb{P}_m \{W_0(s, t; \tilde{\boldsymbol{\theta}}_m, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m})\} h(s) d\tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m}(s) - \mathbb{P}_m \{g_0(t; \tilde{\boldsymbol{\theta}}_m, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m})\} h(t),$$

$$\Gamma(h) = \int_0^\tau E\{W_0(s, t; \boldsymbol{\theta}_0, \Lambda_0)\}h(s)d\Lambda_0(s) - E\{g_0(t; \boldsymbol{\theta}_0, \Lambda_0)\}h(t).$$

Clearly, $\dot{\tilde{\Lambda}}_{\boldsymbol{\theta}}$ solves the equation $g_m = \Gamma_m(\dot{\tilde{\Lambda}}_{\boldsymbol{\theta}})$, and h_0 solves the equation $g = \Gamma(-h_0)$.

Note that

$$\begin{aligned} & \Gamma_m(\dot{\tilde{\Lambda}}_{\tilde{\boldsymbol{\theta}}_m}) - \Gamma(-h_0) \\ &= (\mathbb{P}_m - E) \left\{ \int_0^\tau W_0(s, t; \tilde{\boldsymbol{\theta}}_m, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m}) \dot{\tilde{\Lambda}}_{\tilde{\boldsymbol{\theta}}_m}(s) d\tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m}(s) - g_0(t; \tilde{\boldsymbol{\theta}}_m, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m}) \dot{\tilde{\Lambda}}_{\tilde{\boldsymbol{\theta}}_m}(t) \right\} \\ &+ \int_0^\tau E \left\{ W_0(s, t; \tilde{\boldsymbol{\theta}}_m, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m}) - W_0(s, t; \boldsymbol{\theta}_0, \Lambda_0) \right\} \dot{\tilde{\Lambda}}_{\tilde{\boldsymbol{\theta}}_m}(s) d\tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m}(s) \\ &- E \left\{ g_0(t; \tilde{\boldsymbol{\theta}}_m, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m}) - g_0(t; \boldsymbol{\theta}_0, \Lambda_0) \right\} \dot{\tilde{\Lambda}}_{\tilde{\boldsymbol{\theta}}_m}(t) \\ &+ \int_0^\tau E \left\{ W_0(s, t; \boldsymbol{\theta}_0, \Lambda_0) \right\} \dot{\tilde{\Lambda}}_{\tilde{\boldsymbol{\theta}}_m}(s) \{d\tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m}(s) - d\Lambda_0(s)\} \\ &+ \Gamma(\dot{\tilde{\Lambda}}_{\tilde{\boldsymbol{\theta}}_m} + h_0). \end{aligned} \tag{S2.10}$$

Note also that $W_d(s, t; \tilde{\boldsymbol{\theta}}_m, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m})$ and $g_d(t; \tilde{\boldsymbol{\theta}}_m, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m})$ ($d = 0, 1$) belong to some Donsker class and that $\dot{\tilde{\Lambda}}_{\tilde{\boldsymbol{\theta}}_m}(\cdot)$ has bounded total variation over $[0, \tau]$. Thus, the first term on the right side of (S2.10) is $O_P(m^{-1/2})$. The second, third and forth terms are all $O_P(m^{-1/3})$ as $m^{1/2}(\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0) \rightarrow_d \mathbf{N}(0, \mathbf{I}_0^{-1})$, $\|\tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m} - \Lambda_0\|_2 = O_P(m^{-1/3}) + O_P(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|)$ and $\dot{\tilde{\Lambda}}_{\tilde{\boldsymbol{\theta}}_m}(t) \rightarrow_P -h_0(t)$.

Therefore,

$$\Gamma_m(\dot{\tilde{\Lambda}}_{\tilde{\boldsymbol{\theta}}_m}) - \Gamma(-h_0) = O_P(m^{-1/3}) + \Gamma(\dot{\tilde{\Lambda}}_{\tilde{\boldsymbol{\theta}}_m} + h_0).$$

By similar arguments, $g_m = g + O_P(m^{-1/3})$. Thus, (S2.8) follows from the fact that Γ is bounded and invertible.

To verify (S2.9), we assume, for simplicity, that $\boldsymbol{\theta}$ is one-dimensional, and we consider

$$\tilde{\mathbf{I}}_m = m \left\{ \mathbb{P}_m \left(\left[h_m^{-1} \left\{ l(\tilde{\boldsymbol{\theta}}_m + h_m, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m + h_m}) - l(\tilde{\boldsymbol{\theta}}_m, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m}) \right\} \right]^2 \right) \right\}.$$

Since the third-order derivative of $l(\mathbf{O}; \boldsymbol{\theta}, \tilde{\Lambda}_{\boldsymbol{\theta}})$ with respect to $\boldsymbol{\theta}$ at $\tilde{\boldsymbol{\theta}}_m$ belongs to some

Glivenko-Cantelli class, the Taylor series expansion yields

$$m^{-1}\tilde{\mathbf{I}}_m = \mathbb{P}_m \left\{ \dot{l}_\theta(\tilde{\boldsymbol{\theta}}_m, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m}) + \dot{l}_\Lambda(\tilde{\boldsymbol{\theta}}_m, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m})[\dot{\tilde{\Lambda}}_{\tilde{\boldsymbol{\theta}}_m}] \right\}^2 + O_P(h_m).$$

Since $\dot{l}_\theta(\tilde{\boldsymbol{\theta}}_m, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m}) + \dot{l}_\Lambda(\tilde{\boldsymbol{\theta}}_m, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m})[\dot{\tilde{\Lambda}}_{\tilde{\boldsymbol{\theta}}_m}]$ belongs to some Donsker class (Xu et al., 2022), $\left\{ \dot{l}_\theta(\tilde{\boldsymbol{\theta}}_m, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m}) + \dot{l}_\Lambda(\tilde{\boldsymbol{\theta}}_m, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m})[\dot{\tilde{\Lambda}}_{\tilde{\boldsymbol{\theta}}_m}] \right\}^2$ belongs to some Donsker class. Therefore, with $h_m = O_P(m^{-1/2})$,

$$m^{-1}\tilde{\mathbf{I}}_m = E \left\{ \dot{l}_\theta(\tilde{\boldsymbol{\theta}}_m, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m}) + \dot{l}_\Lambda(\tilde{\boldsymbol{\theta}}_m, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m})[\dot{\tilde{\Lambda}}_{\tilde{\boldsymbol{\theta}}_m}] \right\}^2 + O_P(m^{-1/2}).$$

By the Taylor series expansion,

$$m^{-1}\tilde{\mathbf{I}}_m = E \left\{ \dot{l}_\theta(\boldsymbol{\theta}_0, \tilde{\Lambda}_{\boldsymbol{\theta}_0}) + \dot{l}_\Lambda(\boldsymbol{\theta}_0, \tilde{\Lambda}_{\boldsymbol{\theta}_0})[\dot{\tilde{\Lambda}}_{\boldsymbol{\theta}_0}] \right\}^2 + O_P(\|\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0\|) + O_P(m^{-1/2}).$$

Thus, (S2.9) follows from the facts that $m^{1/2}(\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0) = \mathbf{N}(0, \mathbf{I}_0^{-1})$, $\|\tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m} - \Lambda_0\|_2 = O_P(m^{-1/3})$, $\|\dot{\tilde{\Lambda}}_{\tilde{\boldsymbol{\theta}}_m} + h_0\|_2 = O_P(m^{-1/3})$, and $\mathbf{I}_0 = E \left\{ \dot{l}_\theta(\boldsymbol{\theta}_0, \Lambda_0) - \dot{l}_\Lambda(\boldsymbol{\theta}_0, \Lambda_0)[h_0] \right\}^2$

S2.3 Proportional hazards or intensity model with random effects for dependent failure times

Following Zeng and Lin (2007), we assume the following conditions:

Condition S8. The true parameter value $(\boldsymbol{\theta}_0^T, \boldsymbol{\gamma}_0^T)^T$ belongs to the interior of a compact set Θ , and $\Lambda'_{0k}(t) > 0$ for all $t \in [0, \tau]$, $k = 1, \dots, K$.

Condition S9. With probability 1, $\mathbf{Z}_{ikj}(\cdot)$ and $\tilde{\mathbf{Z}}_{ikj}(\cdot)$ are left continuous in $[0, \tau]$ with uniformly bounded left derivatives.

Condition S10. With probability 1, $P(C_{ikj} \geq \tau \mid \mathbf{Z}_{ikj}) > \delta_0 > 0$ for some constant δ_0 .

Condition S11. With probability 1, n_{ik} is bounded by some integer n_0 . In addition, $E\{N_{ik\cdot}(\tau)\} \leq \infty$.

Condition S12. For any constant $a_1 > 0$,

$$\sup_{\gamma} \left\{ E \left(\int_{\mathbf{b}} \exp [a_1 \{N_{ik}^*(\tau) + 1\} |\mathbf{b}|] f(\mathbf{b}; \gamma) d\mathbf{b} \right) \right\} < \infty, \quad (\text{S2.11})$$

and there is a constant $a_2 > 0$ such that, for any γ ,

$$\left| \frac{f'_{\gamma}(\mathbf{b}; \gamma)}{f(\mathbf{b}; \gamma)} \right| + \left| \frac{f''_{\gamma}(\mathbf{b}; \gamma)}{f(\mathbf{b}; \gamma)} \right| + \left| \frac{f_{\gamma}^{(3)}(\mathbf{b}; \gamma)}{f(\mathbf{b}; \gamma)} \right| \leq O(1) \exp \{a_2(1 + |\mathbf{b}|)\}. \quad (\text{S2.12})$$

Condition S13. If there are $\mathbf{c}(t)$ and $\mathbf{v}$ such that

$$\mathbf{c}(t) + \mathbf{v}^T \mathbf{Z}_{ikj}(t) = 0 \quad (\text{S2.13})$$

with probability 1 for $k = 1, \dots, K$ and $j = 1, \dots, n_{ik}$, then $\mathbf{c}(t) = 0$ and $\mathbf{v} = 0$. In addition, there is some $t \in [0, \tau]$ such that $\{\tilde{\mathbf{Z}}_{ikj}(t), k = 1, \dots, K, j = 1, \dots, n_{ik}\}$ spans the whole space of $\mathbf{b}$.

Condition S14. $f(\mathbf{b}; \gamma) = f(\mathbf{b}; \gamma_0)$ if and only if $\gamma = \gamma_0$ if $\mathbf{v}^T f'(\mathbf{b}; \gamma_0) = 0$, then $\mathbf{v} = 0$.

Let $\boldsymbol{\beta} = (\boldsymbol{\theta}^T, \boldsymbol{\gamma}^T)^T$ and $\eta = (\Lambda_1, \dots, \Lambda_K)$. Define $\mathcal{V} = \{v \in R^d, |v| \leq 1\}$ and $\mathcal{Q} = \{h(t) : \|h(t)\|_{V[0, \tau]} \leq 1\}$, where $\|\cdot\|_{V[0, \tau]}$ denotes the total variation on $[0, \tau]$. Under Conditions S8–S14, $m^{1/2}(\tilde{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_0, \tilde{\eta}_{\tilde{\boldsymbol{\beta}}_m} - \eta_0)$ converges weakly to a zero-mean Gaussian process in $l^\infty(\mathcal{V} \times \mathcal{Q}^K)$ and $\|\tilde{\eta}_{\tilde{\boldsymbol{\beta}}} - \eta_0\|_{V[0, \tau]} = O_P(m^{-1/2}) + O_P(\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|)$ as $\boldsymbol{\beta} \rightarrow_P \boldsymbol{\beta}_0$ (Zeng and Lin, 2010).

We verify Condition 3 under Conditions S8–S14. Let $\int h_0 d\eta_0$ denote the least favorable direction. Condition 3 holds if the following results are satisfied

$$\|\dot{\tilde{\eta}}_{\tilde{\boldsymbol{\beta}}_m} + \int h_0 d\eta_0\|_{V[0, \tau]} = O_P(m^{-1/2}) \quad (\text{S2.14})$$

and

$$\|m^{-1} \tilde{\mathbf{I}}_m - \mathbf{I}_0\| = O_P(m^{-1/2}). \quad (\text{S2.15})$$

To prove (S2.14), we introduce some notation:

$$q_{ikl}(t) = \int_0^t R_{ikl}(s) \exp\{\boldsymbol{\theta}^\top \mathbf{Z}_{ikl}(s) + \mathbf{b}^\top \tilde{\mathbf{Z}}_{ikl}(s)\} d\Lambda_k(s),$$

$$\Omega_{ikl}(\mathbf{b}; \boldsymbol{\beta}, \Lambda_k) = \prod_{t \leq \tau} \left\{ R_{ikl}(t) e^{\boldsymbol{\theta}^\top \mathbf{Z}_{ikl}(t) + \mathbf{b}^\top \tilde{\mathbf{Z}}_{ikl}(t)} \right\}^{dN_{ikl}^*(t)} \exp\{-q_{ikl}(\tau)\},$$

and

$$\Psi(\mathbf{O}_i; \boldsymbol{\beta}, \eta) = \int_{\mathbf{b}} \prod_{k=1}^K \prod_{l=1}^{n_{ik}} \Omega_{ikl}(\mathbf{b}; \boldsymbol{\beta}, \Lambda_k) f(\mathbf{b}; \boldsymbol{\gamma}) d\mathbf{b},$$

$$\psi_k(t; \mathbf{O}_i, \boldsymbol{\beta}, \eta) = \frac{- \int_{\mathbf{b}} \prod_{j=1}^K \prod_{l=1}^{n_{ij}} \Omega_{ijl}(\mathbf{b}; \boldsymbol{\beta}, \Lambda_j) f(\mathbf{b}; \boldsymbol{\gamma}) \sum_{l=1}^{n_{ik}} R_{ikl}(t) e^{\boldsymbol{\theta}^\top \mathbf{Z}_{ikl}(t) + \mathbf{b}^\top \tilde{\mathbf{Z}}_{ikl}(t)} d\mathbf{b}}{\Psi(\mathbf{O}_i; \boldsymbol{\beta}, \eta)}.$$

In addition, let $BV[0, \tau]$ denote the space of functions with bounded total variations in $[0, \tau]$,

and define

$$B_k(\mathcal{W}; \mathbf{O}_i, \boldsymbol{\beta}, \eta) = h_k(t) \psi_k(t; \mathbf{O}_i, \boldsymbol{\beta}, \eta) + \sum_{j=1}^K \int_0^\tau \psi_{kj}(s, t; \mathbf{O}_i, \boldsymbol{\beta}, \eta) h_j(s) d\Lambda_j(s),$$

where $\mathcal{W} = (h_1, \dots, h_K) \in \{BV[0, \tau]\}^K$ and $\psi_{kj}(s, t; \mathbf{O}_i, \boldsymbol{\beta}, \eta)$ is the derivative of ψ_k with respect to Λ_j along the direction $\Lambda_j - \Lambda_{0j}$. Let $\psi_{k\boldsymbol{\beta}}(t; \mathbf{O}_i, \boldsymbol{\beta}, \eta)$ denote the derivative of ψ_k with respect to $\boldsymbol{\beta}$, $B = (B_1, \dots, B_K)$, and $\psi_{\cdot\boldsymbol{\beta}} = (\psi_{1\boldsymbol{\beta}}, \dots, \psi_{K\boldsymbol{\beta}})$.

By definition, $\tilde{\eta}_{\boldsymbol{\beta}}$ solves

$$\mathbb{P}_m \dot{l}_{\Lambda_k}(\boldsymbol{\beta}, \tilde{\eta}_{\boldsymbol{\beta}})[h] = 0, k = 1, \dots, K \quad (\text{S2.16})$$

for any $h \in BV[0, \tau]$, and therefore $\tilde{\Lambda}_{k,\boldsymbol{\beta}}(t)$ takes the form

$$\tilde{\Lambda}_{k,\boldsymbol{\beta}}(t) = \int_0^t \sum_{i=1}^n \sum_{l=1}^{n_{ik}} \frac{1}{\sum_{u=1}^n \psi_k(s; \mathbf{O}_u, \boldsymbol{\beta}, \tilde{\eta}_{\boldsymbol{\beta}})} R_{ikl}(s) dN_{ikl}^*(s).$$

By differentiating this equation with respect to $\boldsymbol{\beta}$, we can express $\tilde{\dot{\Lambda}}_{k,\boldsymbol{\beta}}(t)$ in the form $\int_0^t \tilde{h}_k(s) d\tilde{\Lambda}_{k,\boldsymbol{\beta}}(s)$ for some function $\tilde{h}_k(s)$. Define $\tilde{h} = (\tilde{h}_1, \dots, \tilde{h}_K)$.

By the definition of least favorable direction, h_0 solves

$$EB(-h_0; \boldsymbol{\beta}_0, \eta_0) + E\psi_{\cdot\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \eta_0) = 0, \quad (\text{S2.17})$$

and by differentiating (S2.16) with respect to β , $\tilde{h}$ solves

$$\mathbb{P}_m B(\tilde{h}; \tilde{\beta}_m, \tilde{\eta}_{\tilde{\beta}_m}) + \mathbb{P}_m \psi_{\cdot\beta}(\tilde{\beta}_m, \tilde{\eta}_{\tilde{\beta}_m}) = 0. \quad (\text{S2.18})$$

We can show that $\psi_k(t; \mathbf{O}, \tilde{\beta}_m, \tilde{\eta}_{\tilde{\beta}_m})$, $\psi_{kj}(s, t; \mathbf{O}, \tilde{\beta}_m, \tilde{\eta}_{\tilde{\beta}_m})$ and $\psi_{k\beta}(t; \mathbf{O}, \tilde{\beta}_m, \tilde{\eta}_{\tilde{\beta}_m})$ belong to some Donsker classes. Thus, $\mathbb{P}_m B(h; \tilde{\beta}_m, \tilde{\eta}_{\tilde{\beta}_m}) \rightarrow EB(h; \beta_0, \eta_0)$ for any $h \in \{BV[0, \tau]\}^K$ and $\mathbb{P}_m \psi_{\cdot\beta}(\tilde{\beta}_m, \tilde{\eta}_{\tilde{\beta}_m}) \rightarrow E\psi_{\cdot\beta}(\beta_0, \eta_0)$. Because $EB(\cdot; \beta_0, \eta_0)$ is bounded and invertible (Zeng and Lin, 2010), $\mathbb{P}_m B(\cdot, \tilde{\beta}_m, \tilde{\eta}_{\tilde{\beta}_m})$ is bounded and invertible when m is large enough, and $\left\{ \mathbb{P}_m B(\cdot, \tilde{\beta}_m, \tilde{\eta}_{\tilde{\beta}_m}) \right\}^{-1}(h) \rightarrow \{EB(\cdot; \beta_0, \eta_0)\}^{-1}(h)$.

In light of (S2.17) and (S2.18), and together with the uniform boundedness principle,

$$\begin{aligned} & \tilde{h} \\ &= \left\{ \mathbb{P}_m B(\cdot, \tilde{\beta}_m, \tilde{\eta}_{\tilde{\beta}_m}) \right\}^{-1} \left\{ \mathbb{P}_m \psi_{\cdot\beta}(\tilde{\beta}_m, \tilde{\eta}_{\tilde{\beta}_m}) \right\} \\ &= \left[\left\{ \mathbb{P}_m B(\cdot, \tilde{\beta}_m, \tilde{\eta}_{\tilde{\beta}_m}) \right\}^{-1} - \{EB(\cdot; \beta_0, \eta_0)\}^{-1} \right] \left\{ \mathbb{P}_m \psi_{\cdot\beta}(\tilde{\beta}_m, \tilde{\eta}_{\tilde{\beta}_m}) \right\} \\ &\quad + \{EB(\cdot; \beta_0, \eta_0)\}^{-1} \left\{ \mathbb{P}_m \psi_{\cdot\beta}(\tilde{\beta}_m, \tilde{\eta}_{\tilde{\beta}_m}) \right\} \\ &\rightarrow \{EB(\cdot; \beta_0, \eta_0)\}^{-1} \{E\psi_{\cdot\beta}(\beta_0, \eta_0)\} \\ &= -h_0. \end{aligned}$$

Note that

$$\begin{aligned} & \mathbb{P}_m B(\tilde{h}; \tilde{\beta}_m, \tilde{\eta}_{\tilde{\beta}_m}) - EB(-h; \beta_0, \eta_0) \\ &= (\mathbb{P}_m - E)B(\tilde{h}; \tilde{\beta}_m, \tilde{\eta}_{\tilde{\beta}_m}) \\ &\quad + E \left\{ \psi_{\cdot}(\tilde{\beta}_m, \tilde{\eta}_{\tilde{\beta}_m}) - \psi_{\cdot}(\beta_0, \eta_0) \right\} \tilde{h} \\ &\quad + \sum_{k=1}^K \int_0^\tau E \left\{ \psi_{\cdot k}(\tilde{\beta}_m, \tilde{\eta}_{\tilde{\beta}_m}) - \psi_{\cdot k}(\beta_0, \eta_0) \right\} \tilde{h} d\tilde{\Lambda}_{k, \tilde{\beta}_m} \\ &\quad + \sum_{k=1}^K \int_0^\tau E\psi_{\cdot k}(\beta_0, \eta_0) \tilde{h} \left\{ d\tilde{\Lambda}_{k, \tilde{\beta}_m} - d\Lambda_{0k} \right\} \end{aligned}$$

$$+ EB(\tilde{h} + h_0; \beta_0, \eta_0).$$

By similar arguments to those used in Section S2.2, $\mathbb{P}_m B(\tilde{h}; \tilde{\beta}_m, \tilde{\eta}_{\tilde{\beta}_m}) - EB(-h; \beta_0, \eta_0) = O_P(m^{-1/2}) + EB(\tilde{h} + h_0; \beta_0, \eta_0)$. Thus, (S2.14) follows from the fact that $EB(\cdot; \beta_0, \eta_0)$ is bounded and invertible and that $\mathbb{P}_m \psi_{\beta}(\tilde{\beta}_m, \tilde{\eta}_{\tilde{\beta}_m}) - E\psi_{\beta}(\beta_0, \eta_0) = O_P(m^{-1/2})$. Finally, (S2.15) can be proved using identical arguments to those detailed in Section S2.2.

S3 Techincal details of time complexity calculations

S3.1 Partially linear model

We first analyze the time complexity of performing sieve maximum likelihood estimation with the entire data, which is determined by the time required to minimize the sum of squared residuals $\sum_{i=1}^n \{Y_i - \theta^T \mathbf{X}_i - \omega^T \mathbf{p}_n(Z_i)\}^2$, where $\mathbf{p}_n(Z)$ is the basis of the space of polynomial splines of order r with k_n knots, and $k_n = O\{n^{1/(2q+1)}\}$. Clearly, the time complexity is $O\{n(p + r + k_n)^2\} = O\{n^{1+2/(1+2q)}\}$.

Next, we determine the time complexity of the proposed method. The time required to compute $\tilde{\theta}_m$ is $O\{m^{1+2/(1+2q)}\}$ by the above result. For the second block, we need to evaluate $\sum_{i \in \mathcal{J}_2} (\tilde{\sigma}_{\tilde{\theta}_m}^2)^{-1} \left\{ Y_i - \tilde{\theta}_m^T \mathbf{X}_i - \tilde{g}_{\tilde{\theta}_m}(Z_i) \right\} \left\{ \mathbf{X}_i + \tilde{g}_{\tilde{\theta}_m}(Z_i) \right\}$, which requires $O\{(n - m) \times m^{1/(1+2q)}\}$ calculations. Combining the calculations in the first and second blocks, we conclude that the total time complexity of the proposed method is $O\{n^{r_0+2r_0/(1+2q)} + n^{1+r_0/(1+2q)}\}$ provided that $m = O(n^{r_0})$.

S3.2 Cox proportional hazards model with interval-censored data

We first describe the EM algorithm of Zeng et al. (2016) on the entire data. Let $t_1 < \dots < t_{n_0}$ denote the unique values of $L_i > 0$ and $R_i < \infty$ ($i = 1, \dots, n$), and let $\lambda_1, \dots, \lambda_{n_0}$ denote the values of $\lambda(t)$ at $t_1, \dots, t_{n_0}$. In addition, write $\mathbf{Z}_{ik} = \mathbf{Z}_i(t_k)$ ($i = 1, \dots, n; k = 1, \dots, n_0$). Finally, define $R_i^* = R_i I(R_i < \infty) + L_i I(R_i = \infty)$, where $I(\cdot)$ is the indicator function.

In the E-step, for $i = 1, \dots, n$ and $k = 1, \dots, n_0$, if $L_i < R_i$, we set

$$w_{ik} = \begin{cases} \frac{\lambda_k e^{\boldsymbol{\theta}^T \mathbf{Z}_{ik}}}{1 - \exp\left(-\sum_{L_i < t_l \leq R_i} \lambda_l e^{\boldsymbol{\theta}^T \mathbf{Z}_{il}}\right)} & \text{if } L_i < t_k \leq R_i < \infty \\ 0 & \text{otherwise.} \end{cases}$$

If $L_i = R_i$, we set $w_{ik} = I(L_i = t_k)$. In the M-step, we set

$$\boldsymbol{\theta}_{\text{new}} = \boldsymbol{\theta} + \Sigma^{-1} \sum_{i=1}^n \sum_{k=1}^{n_0} I(t_k \leq R_i^*) w_{ik} \left\{ \mathbf{Z}_{ik} - \frac{\sum_{j=1}^n I(t_k \leq R_j^*) e^{\boldsymbol{\theta}^T \mathbf{Z}_{jk}} \mathbf{Z}_{jk}}{\sum_{j=1}^n I(t_k \leq R_j^*) e^{\boldsymbol{\theta}^T \mathbf{Z}_{jk}}} \right\},$$

where

$$\Sigma = \sum_{i=1}^n \sum_{k=1}^{n_0} I(t_k \leq R_i^*) w_{ik} \left[\frac{\sum_{j=1}^n I(t_k \leq R_j^*) e^{\boldsymbol{\theta}^T \mathbf{Z}_{jk}} \mathbf{Z}_{jk} \mathbf{Z}_{jk}^T}{\sum_{j=1}^n I(t_k \leq R_j^*) e^{\boldsymbol{\theta}^T \mathbf{Z}_{jk}}} - \frac{\left\{ \sum_{j=1}^n I(t_k \leq R_j^*) e^{\boldsymbol{\theta}^T \mathbf{Z}_{jk}} \mathbf{Z}_{jk} \right\} \left\{ \sum_{j=1}^n I(t_k \leq R_j^*) e^{\boldsymbol{\theta}^T \mathbf{Z}_{jk}} \mathbf{Z}_{jk} \right\}^T}{\left\{ \sum_{j=1}^n I(t_k \leq R_j^*) e^{\boldsymbol{\theta}^T \mathbf{Z}_{jk}} \right\}^2} \right].$$

Then, we calculate

$$\lambda_k^{\text{new}} = \frac{\sum_{i=1}^n I(t_k \leq R_i^*) w_{ik}}{\sum_{i=1}^n I(t_k \leq R_i^*) e^{\boldsymbol{\theta}_{\text{new}}^T \mathbf{Z}_{ik}}}, \quad k = 1, \dots, n_0.$$

The time required for a single iteration in the EM algorithm is determined by the evaluation of Σ and w_{ik} ($i = 1, \dots, n; k = 1, \dots, n_0$), which is linear in $n_0 n$. Since $n_0 = O(n)$ and the number of iterations is N_{iter} , the time complexity for the standard method is $O(N_{\text{iter}} n^2)$.

Next, we determine the time complexity of the proposed method. The time required to

compute $\tilde{\boldsymbol{\theta}}_m$ is $O(N_{iter}m^2)$ by the above result. For the second block, we need to evaluate

$$\sum_{i \in \mathcal{J}_2} (2\delta_n)^{-1} \left\{ l(\mathbf{O}_i; \tilde{\boldsymbol{\theta}}_m + \delta_n \mathbf{e}_j, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m + \delta_n \mathbf{e}_j}) - l(\mathbf{O}_i; \tilde{\boldsymbol{\theta}}_m - \delta_n \mathbf{e}_j, \tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m - \delta_n \mathbf{e}_j}) \right\}_{j=1, \dots, p}. \quad (\text{S3.19})$$

Note that $\tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m + \delta_n \mathbf{e}_j}$ and $\tilde{\Lambda}_{\tilde{\boldsymbol{\theta}}_m - \delta_n \mathbf{e}_j}$ ($j = 1, \dots, p$) are step functions with m_0 jumps, which are obtained from the first block analysis. Here, m_0 is the number of distinct monitoring times for the first block. Thus, the time complexity for (S3.19) is $O\{\max(n - m, m_0)\} = O(n - m)$ after the monitoring times in the second block are sorted. Combining the calculations in the first and second blocks, we conclude that the total time complexity of the proposed method is $O\{N_{iter}n^{2r_0} + n - m + (n - m) \log(n - m)\} = O(N_{iter}n^{2r_0} + n \log n)$ given $m = O(n^{r_0})$.

S4 Additional simulation results

S4.1 Proportional intensity model for recurrent events

We generated two types of failures with cumulative hazard functions $\Lambda_k(t) \exp(\theta_1 Z_{1i} + \theta_2 Z_{2i} + b_i)$ ($k = 1, 2; i = 1, \dots, n$), where Z_{1i} and Z_{2i} are independent Bernoulli and uniform $[0, 1]$ variables, b_i is standard normal, $\theta_1 = -\theta_2 = 1$, $\Lambda_1(t) = 0.3t$, and $\Lambda_2(t) = 0.15t^2$. We generated censoring times from uniform $[0, 3]$ and set $\tau = 4$, resulting in approximately 64% and 70% censoring for $k = 1$ and $k = 2$, respectively. We set $n = 10^6$. The computational setup for all methods was identical to that used in the simulation study for the random effects proportional hazards model with clustered events.

Table S1 presents the results based on 1000 replicates. The standard method took approximately 22 minutes, and the average total CPU time for the divide-and-conquer method with $M = 5, 10$ and 20 was 30, 38, and 44 minutes, respectively. By contrast, the average total CPU time for the proposed method with $k = 0.05, 0.1$ and 0.2 was only 3, 5, and 8

Table S1: Summary statistics for the simulation study for multiple failure types[†]

Method		$\theta_1 = 1$				$\theta_2 = -1$			
		Bias	SE	SEE	CP	Bias	SE	SEE	CP
		$\times 10^{-3}$	$\times 10^{-3}$	$\times 10^{-3}$	$\times 100\%$	$\times 10^{-3}$	$\times 10^{-3}$	$\times 10^{-3}$	$\times 100\%$
Standard		0.11	3.61	3.60	95	-0.39	6.04	5.93	95
Proposed	$k = 0.05 \quad \delta_n = 0.0005$	-0.01	3.61	3.61	95	-0.24	6.04	6.00	95
	$\delta_n = 0.005$	0.01	3.61	3.61	95	-0.30	6.04	5.96	95
	$k = 0.1 \quad \delta_n = 0.0005$	0.05	3.61	3.61	95	-0.32	6.04	5.98	95
	$\delta_n = 0.005$	0.06	3.61	3.60	95	-0.35	6.04	5.95	95
	$k = 0.2 \quad \delta_n = 0.0005$	0.08	3.61	3.60	95	-0.35	6.04	5.96	95
	$\delta_n = 0.005$	0.09	3.61	3.60	95	-0.37	6.04	5.94	95
Divide-and-conquer	$M = 20$	0.08	3.61	3.60	95	-0.37	6.04	5.93	95
	$M = 10$	0.10	3.61	3.60	95	-0.38	6.04	5.93	95
	$M = 5$	0.10	3.61	3.60	95	-0.39	6.04	5.93	95

[†]Bias, bias of the parameter estimator; SE, standard error of the parameter estimator; SEE, average of the standard error estimator; CP, coverage percentage of the 95% confidence interval.

minutes, respectively.

S4.2 Proportional hazards model for multiple types of events

We generated recurrent event times with cumulative intensity function $\Lambda_0(t) \exp(\theta_1 Z_{i1} + \theta_2 Z_{i2} + b_i)(i = 1, \dots, n)$, where Z_{1i} and Z_{2i} are independent Bernoulli and uniform $[0, 1]$ variables, b_i is standard normal, $\theta_1 = -\theta_2 = 0.5$, and $\Lambda_0(t) = 0.1 \log(1 + t)$. We sampled censoring times from uniform $[1, 3]$ and set $\tau = 2$. The censoring rate was approximately 85%. We set $n = 10^6$ and used the identical computational setup described in the simulation study for the proportional hazards model with clustered events.

Table S2: Summary statistics for the simulation study for recurrent events[†]

Method		$\theta_1 = 0.5$				$\theta_2 = -0.5$			
		Bias	SE	SEE	CP	Bias	SE	SEE	CP
		$\times 10^{-3}$	$\times 10^{-3}$	$\times 10^{-3}$	$\times 100\%$	$\times 10^{-3}$	$\times 10^{-3}$	$\times 10^{-3}$	$\times 100\%$
Standard		-0.15	6.04	6.00	95	0.08	2.74	2.75	96
Proposed	$k = 0.05 \quad \delta_n = 0.0005$	-0.21	6.04	6.03	95	0.12	2.74	2.80	94
	$\delta_n = 0.005$	-0.18	6.04	6.00	95	0.08	2.74	2.78	95
	$k = 0.1 \quad \delta_n = 0.0005$	-0.18	6.05	6.03	95	0.10	2.74	2.77	95
	$\delta_n = 0.005$	-0.17	6.05	6.00	95	0.09	2.74	2.75	95
	$k = 0.2 \quad \delta_n = 0.0005$	-0.16	6.05	6.01	95	0.08	2.74	2.75	95
	$\delta_n = 0.005$	-0.16	6.05	6.00	95	0.07	2.74	2.75	95
Divide-and-conquer	$M = 20$	-0.15	6.04	6.00	95	0.07	2.74	2.75	96
	$M = 10$	-0.15	6.05	6.00	95	0.07	2.74	2.75	96
	$M = 5$	-0.15	6.04	6.00	95	0.07	2.74	2.75	96

[†]Bias, bias of the parameter estimator; SE, standard error of the parameter estimator; SEE, average of the standard error estimator; CP, coverage percentage of the 95% confidence interval.

Table S2 summarizes the results based on 1000 replicates. The standard method took approximately 27 minutes, and the average total CPU time for the divide-and-conquer method with $M = 5, 10$ and 20 was 25, 49, and 65 minutes, respectively. By contrast, the average total CPU time for the proposed method with $k = 0.05, 0.1$ and 0.2 was only 3.5, 5.7, and 6.2 minutes, respectively.

Bibliography

- Chen, X. (2007). Large sample sieve estimation of semi-nonparametric model. In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume VI, pp. 5549–5632. New York: Elsevier Science B.V.
- van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York: Springer.
- Xu, Y., D. Zeng, and D. Y. Lin (2022, 11). Marginal proportional hazards models for multivariate interval-censored data. *Biometrika* 110(3), 815–830.
- Zeng, D. and D. Lin (2010). A general asymptotic theory for maximum likelihood estimation in semiparametric regression models with censored data. *Statistica Sinica* 20(2), 871.
- Zeng, D. and D. Y. Lin (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 69(4), 507–564.
- Zeng, D., L. Mao, and D. Y. Lin (2016). Maximum likelihood estimation for semiparametric transformation models with interval-censored data. *Biometrika* 103, 253–271.

Table S3: Estimation results for hypertension in the UK Biobank study

Covariates	Standard method		Proposed method			Divide-and-conquer method		
			$k = 0.2$	$k = 0.1$	$k = 0.05$	$M = 5$	$M = 10$	$M = 20$
Sex	0.3945 (3.979)	0.3952 (3.985)	0.3962 (3.992)	0.3947 (4.008)	0.3946 (3.997)	0.3947 (3.996)	0.3949 (4.010)	
Race	0.0980 (8.730)	0.1009 (8.862)	0.1043 (8.779)	0.0875 (8.422)	0.0983 (8.768)	0.0979 (8.719)	0.0984 (8.663)	
Birthplace	-0.0355 (5.811)	-0.0348 (5.800)	-0.0331 (5.826)	-0.0382 (5.860)	-0.0355 (5.842)	-0.0355 (5.847)	-0.0355 (5.834)	
PC1	0.0012 (0.034)	0.0012 (0.028)	0.0013 (0.021)	0.0012 (0.013)	0.0012 (0.026)	0.0012 (0.020)	0.0012 (0.014)	
PC2	0.0003 (0.085)	0.0003 (0.088)	0.0004 (0.090)	0.0003 (0.087)	0.0003 (0.089)	0.0003 (0.091)	0.0003 (0.092)	
PC3	0.0005 (0.146)	0.0005 (0.143)	0.0005 (0.143)	0.0004 (0.135)	0.0005 (0.145)	0.0005 (0.143)	0.0005 (0.140)	
PC4	0.0024 (0.217)	0.0024 (0.220)	0.0023 (0.222)	0.0023 (0.224)	0.0024 (0.219)	0.0024 (0.221)	0.0024 (0.223)	
PC5	0.0007 (0.294)	0.0007 (0.293)	0.0008 (0.294)	0.0006 (0.290)	0.0007 (0.294)	0.0007 (0.293)	0.0007 (0.291)	
PC6	-0.0022 (0.489)	-0.0021 (0.525)	-0.0023 (0.445)	-0.0019 (0.378)	-0.0022 (0.478)	-0.0022 (0.469)	-0.0023 (0.465)	
PC7	-0.0014 (0.437)	-0.0014 (0.445)	-0.0014 (0.446)	-0.0013 (0.438)	-0.0014 (0.438)	-0.0014 (0.439)	-0.0014 (0.441)	
PC8	0.0022 (0.471)	0.0021 (0.486)	0.0022 (0.471)	0.0023 (0.452)	0.0022 (0.472)	0.0023 (0.472)	0.0023 (0.476)	
PC9	0.0015 (0.457)	0.0016 (0.461)	0.0015 (0.458)	0.0016 (0.469)	0.0015 (0.459)	0.0015 (0.461)	0.0015 (0.463)	
PC10	0.0024 (0.514)	0.0024 (0.512)	0.0024 (0.524)	0.0027 (0.523)	0.0024 (0.515)	0.0024 (0.516)	0.0024 (0.518)	
PRS	0.3158 (2.159)	0.3155 (2.165)	0.3156 (2.176)	0.3156 (2.160)	0.3158 (2.159)	0.3160 (2.162)	0.3163 (2.166)	

Standard error estimates ($\times 10^{-3}$) are shown in parentheses.