CONFIDENCE INTERVALS FOR HIGH-DIMENSIONAL COX MODELS

Yi Yu[1], Jelena Bradic[2] and Richard J. Samworth[3]

[1]*Department of Statistics, University of Warwick*

[2]*Department of Mathematics, University of California at San Diego*

[3]*Statistical Laboratory, University of Cambridge*

Supplementary Material

This document consists of technical lemmas, detailed proofs of Propositions 1, 2 and 3 in the main document, and further numerical results.

S1 Technical lemmas

We first recall the following basic facts about the operator norms of a matrix.

Lemma S1. *For any matrix* $\boldsymbol{A} = (A_{ij}) \in \mathbb{R}^{m\times m}$*, we have*

$$\|\boldsymbol{A}\|_{\mathrm{op},1} = \max_{j=1,\ldots,m} \sum_{i=1}^{m} |A_{ij}| \quad \text{and} \quad \|\boldsymbol{A}\|_{\mathrm{op},\infty} = \max_{i=1,\ldots,m} \sum_{j=1}^{m} |A_{ij}|$$

Proof. For any $\boldsymbol{v} = (v_1, \ldots, v_m)^\top \in \mathbb{R}^m \setminus \{\mathbf{0}\}$, we have

$$\frac{\|\boldsymbol{A}\boldsymbol{v}\|_1}{\|\boldsymbol{v}\|_1} = \frac{\sum_{i=1}^m \bigl|\sum_{j=1}^m A_{ij} v_j\bigr|}{\sum_{j=1}^m |v_j|} \leq \frac{\sum_{j=1}^m |v_j| \sum_{i=1}^m |A_{ij}|}{\sum_{j=1}^m |v_j|} \leq \max_{j=1,\ldots,m} \sum_{i=1}^m |A_{ij}|.$$

On the other hand, suppose in the first instance that $\boldsymbol{A} \neq 0$. Let $j^* \in \mathrm{argmax}_{j=1,\ldots,m} \sum_{i=1}^m |A_{ij}|$, and let $\boldsymbol{v} = (v_1, \ldots, v_m)^\top \in \mathbb{R}^m$ be given by $v_i = \mathrm{sgn}(A_{ij^*}) \mathbb{1}_{\{i=j^*\}}$ for $i = 1, \ldots, m$. Then

$$\max_{j=1,\ldots,m} \sum_{i=1}^m |A_{ij}| = \sum_{i=1}^m |A_{ij^*}| = \frac{\|\boldsymbol{A}\boldsymbol{v}\|_1}{\|\boldsymbol{v}\|_1}.$$

This equality also holds for any $\boldsymbol{v} \neq 0$ when $\boldsymbol{A} = 0$, so the result for the case of $\|\cdot\|_{\mathrm{op},1}$ follows. The argument for the case of $\|\cdot\|_{\mathrm{op},\infty}$ is similar and is omitted. □

The lemma below is key in deriving both the asymptotic normality of the leading term and the asymptotic negligibility of the residual terms. It provides conditions under which we can control the deviations of the process $\{\bar{\boldsymbol{Z}}(t, \boldsymbol{\beta}) : t \in [0, t_*)\}$ from $\{\boldsymbol{\mu}(t, \boldsymbol{\beta}) : t \in [0, t_*)\}$, where $t_* \in [0, t_+)$ is an (initially arbitrary) truncation. It will be convenient to define some notation. Let

$$L_{\boldsymbol{\mu}}(\|\boldsymbol{\beta}\|_1) := \frac{e^{2\|\boldsymbol{\beta}\|_1 K_Z}(L + K_Z\|f_T\|_\infty + 2K_Z\|\boldsymbol{\beta}\|_1 L) + K_Z e^{2\|\boldsymbol{\beta}\|_1 K_Z}(\|f_T\|_\infty + 2\|\boldsymbol{\beta}\|_1 L)}{\bar{F}_T(t_*)}, \tag{S1.1}$$

$$L_{\bar{\boldsymbol{Z}}}(\|\boldsymbol{\beta}\|_1) := L + 4K_Z\|\boldsymbol{\beta}\|_1 L. \tag{S1.2}$$

Lemma S2. *Assume **(A1)** and **(A2)(a)** and let $t_* \in \mathcal{T}$. For $\boldsymbol{\beta} \in \mathbb{R}^p$, let $\epsilon_n = \epsilon_n(\|\boldsymbol{\beta}\|_1) := \frac{3K_Z e^{2\|\boldsymbol{\beta}\|_1 K_Z}}{n\bar{F}_T(t_*) - n^{1/2}(\log n)\bar{F}_T^{1/2}(t_*)}$ and fix $\epsilon > \epsilon_n$. Set*

$$h_0 = h_0(n,\epsilon) := \min\left\{\frac{1}{2(\|\boldsymbol{\beta}\|_1 + 1)L}, \frac{\epsilon - \epsilon_n}{3L_{\bar{\boldsymbol{Z}}}(\|\boldsymbol{\beta}\|_1)}, \frac{\epsilon}{3L_{\boldsymbol{\mu}}(\|\boldsymbol{\beta}\|_1)}, \frac{\bar{F}_T(t_*)}{2\|f_T\|_\infty}\right\}.$$

Then, writing $M_0 = M_0(n,\epsilon) := t_*/h_0 + n^3 + 1$, we have

$$\begin{aligned}\mathbb{P}\left(\sup_{t\in[0,t_*)} \|\bar{\boldsymbol{Z}}(t,\boldsymbol{\beta}) - \boldsymbol{\mu}(t,\boldsymbol{\beta})\|_\infty > \epsilon\right) &\le \frac{1}{2n} + e^{-(\log^2 n)/2} \\ &+ (2p+2)M_0 \exp\left(-\frac{n\epsilon^2 \bar{F}_T(t_*)}{1152K_Z e^{4\|\boldsymbol{\beta}\|_1 K_Z}(K_Z + \frac{\epsilon}{36})}\right) \\ &+ 2M_0 \exp\left(-\frac{3n\bar{F}_T(t_*)}{28e^{4\|\boldsymbol{\beta}\|_1 K_Z)}}\right).\end{aligned}$$

Remark: Consider an asymptotic regime in which **(A3)(a)** holds and $\bar{F}(t_*) = O\big(n^{-(1/2-\delta)}\big)$ for some $\delta \in (0, 1/2)$. If $\int_0^{t_+} t^\alpha f_T(t)\,dt < \infty$ for some $\alpha > 0$ (i.e. **(A2)(b)** holds), then for such t_*,

$$n^{-(1/2-\delta)} = \int_{F_T^{-1}(1-n^{-(1/2-\delta)})}^{t_+} f_T \le \frac{1}{\big\{F_T^{-1}(1 - n^{-(1/2-\delta)})\big\}^\alpha} \int_0^{t_+} t^\alpha f_T(t)\,dt;$$

in other words, $F_T^{-1}(1-n^{-(1/2-\delta)}) = O(n^{(1/2-\delta)/\alpha})$. It therefore follows from

Lemma S2 that if **(A1)**, **(A2)** and **(A3)(a)** hold, then

$$\sup_{t\in[0,t_*)} \|\bar{\boldsymbol{Z}}(t,\boldsymbol{\beta}) - \boldsymbol{\mu}(t,\boldsymbol{\beta})\|_\infty \xrightarrow{p} 0.$$

Proof. As a first step, we prove that the process $\{\boldsymbol{\mu}(t,\boldsymbol{\beta}) : t \in [0,t_*)\}$ inherits a Lipschitz property from $\{\boldsymbol{Z}(t,\boldsymbol{\beta}) : t \in [0,t_*)\}$. In fact, writing $\tilde{w}(t,\boldsymbol{\beta}) := Y(t)e^{\boldsymbol{\beta}^\top \boldsymbol{Z}(t)}$, for $t, t+h \in [0,t_*)$ with $h \in (0,h_0]$ (so in particular $(\|\boldsymbol{\beta}\|_1 + 1)Lh \leq 1/2$),

$$\begin{aligned}
&\bigl|\mathbb{E}\bigl\{\tilde{w}(t+h,\boldsymbol{\beta})\bigr\} - \mathbb{E}\bigl\{\tilde{w}(t,\boldsymbol{\beta})\bigr\}\bigr| \\
&\leq \bigl|\mathbb{E}\bigl\{\tilde{w}(t+h,\boldsymbol{\beta})\bigr\} - \mathbb{E}\bigl\{Y(t)e^{\boldsymbol{\beta}^\top \boldsymbol{Z}(t+h)}\bigr\}\bigr| \\
&\qquad\qquad + \bigl|\mathbb{E}\bigl\{Y(t)e^{\boldsymbol{\beta}^\top \boldsymbol{Z}(t+h)}\bigr\} - \mathbb{E}\bigl\{\tilde{w}(t,\boldsymbol{\beta})\bigr\}\bigr| \\
&\leq e^{\|\boldsymbol{\beta}\|_1 K_Z}\|f_T\|_\infty h + e^{\|\boldsymbol{\beta}\|_1 K_Z}(e^{\|\boldsymbol{\beta}\|_1 Lh} - 1) \leq e^{\|\boldsymbol{\beta}\|_1 K_Z}(\|f_T\|_\infty + 2\|\boldsymbol{\beta}\|_1 L)h.
\end{aligned}$$

Similarly, again for $t, t+h \in [0,t_*)$ and $h \in (0,h_0]$,

$$\begin{aligned}
&\bigl\|\mathbb{E}\bigl\{\boldsymbol{Z}(t+h)\tilde{w}(t+h,\boldsymbol{\beta})\bigr\} - \mathbb{E}\bigl\{\boldsymbol{Z}(t)\tilde{w}(t,\boldsymbol{\beta})\bigr\}\bigr\|_\infty \\
&\qquad\qquad \leq e^{\|\boldsymbol{\beta}\|_1 K_Z} Lh + K_Z e^{\|\boldsymbol{\beta}\|_1 K_Z}\|f_T\|_\infty h + K_Z e^{\|\boldsymbol{\beta}\|_1 K_Z}(e^{\|\boldsymbol{\beta}\|_1 Lh} - 1) \\
&\qquad\qquad \leq e^{\|\boldsymbol{\beta}\|_1 K_Z}(L + K_Z\|f_T\|_\infty + 2K_Z\|\boldsymbol{\beta}\|_1 L)h.
\end{aligned}$$

It follows that provided $h \in (0,h_0]$, so that $\bar{F}_T(t+h) \geq \bar{F}_T(t)/2$ for $t, t+h \in$

$[0, t_*)$, we have

$$\begin{aligned}
&\|\boldsymbol{\mu}(t+h, \boldsymbol{\beta}) - \boldsymbol{\mu}(t, \boldsymbol{\beta})\|_\infty \\
&= \left\| \frac{\mathbb{E}\big\{\boldsymbol{Z}(t+h)\tilde{w}(t+h, \boldsymbol{\beta})\big\}\mathbb{E}\tilde{w}(t, \boldsymbol{\beta}) - \mathbb{E}\big\{\boldsymbol{Z}(t)\tilde{w}(t, \boldsymbol{\beta})\big\}\mathbb{E}\tilde{w}(t+h, \boldsymbol{\beta})}{\mathbb{E}\tilde{w}(t+h, \boldsymbol{\beta})\mathbb{E}\tilde{w}(t, \boldsymbol{\beta})} \right\|_\infty \\
&\le L_{\boldsymbol{\mu}}(\|\boldsymbol{\beta}\|_1)h, \qquad \text{(S1.3)}
\end{aligned}$$

where $L_{\boldsymbol{\mu}}(\|\boldsymbol{\beta}\|_1)$ was defined in (S1.1). We now aim to prove a similar property for the process $\{\bar{\boldsymbol{Z}}(t, \boldsymbol{\beta}) : t \in [0, t_*)\}$ (though this process may have jumps). Let $M := n^3$, and let $s_j := F_T^{-1}(j/M)$ for $j = 0, 1, \ldots, M-1$. Let $E_j := \sum_{i=1}^n \mathbb{1}_{\{T_i \in [s_j, s_{j+1})\}}$, and let $\Omega_0 := \cap_{j=1}^M \{E_j \le 1\}$. Then

$$\mathbb{P}(\Omega_0^c) \le M\left\{1 - \left(1 - \frac{1}{M}\right)^n - \frac{n}{M}\left(1 - \frac{1}{M}\right)^{n-1}\right\} \le \frac{1}{2n}.$$

Now, fix $t \in [0, t_*)$ and $h \in (0, h_0]$ such that $t, t+h \in [s_j, s_{j+1})$ for some j, and let $R_t := \{i : Y_i(t) = 1\}$ denote the risk set at time t. If $\sum_{i \in R_t} \mathbb{1}_{\{T_i \in [t_i, t_i+h)\}} = 0$, i.e. there are no observed events in $[t, t+h)$, then

$$\begin{aligned}
&\sum_{i \in R_t} |w_i(t+h, \boldsymbol{\beta}) - w_i(t, \boldsymbol{\beta})| \\
&= \sum_{i \in R_t} \left| \frac{e^{\boldsymbol{\beta}^\top \boldsymbol{Z}_i(t+h)} \sum_{j \in R_t} e^{\boldsymbol{\beta}^\top \boldsymbol{Z}_j(t)} - e^{\boldsymbol{\beta}^\top \boldsymbol{Z}_i(t)} \sum_{j \in R_t} e^{\boldsymbol{\beta}^\top \boldsymbol{Z}_j(t+h)}}{\big(\sum_{j \in R_t} e^{\boldsymbol{\beta}^\top \boldsymbol{Z}_j(t+h)}\big)\big(\sum_{j \in R_t} e^{\boldsymbol{\beta}^\top \boldsymbol{Z}_j(t)}\big)} \right| \\
&\le 2(e^{\|\boldsymbol{\beta}\|_1 Lh} - 1).
\end{aligned}$$

On the other hand, if there is one observed event (corresponding to the individual i^*) in $[t, t+h)$, then

$$\sum_{i \in R_t} |w_i(t+h, \boldsymbol{\beta}) - w_i(t, \boldsymbol{\beta})| = \sum_{i \in R_t \setminus \{i^*\}} |w_i(t+h, \boldsymbol{\beta}) - w_i(t, \boldsymbol{\beta})| + w_{i^*}(t, \boldsymbol{\beta})$$
$$\leq 2(e^{\|\boldsymbol{\beta}\|_1 Lh} - 1) + \frac{e^{2\|\boldsymbol{\beta}\|_1 K_Z}}{|R_t|}.$$

It follows that on the event Ω_0, if $t \in [0, t_*)$ and $h \in (0, h_0]$ are such that $t, t+h \in [s_j, s_{j+1})$ for some j, then

$$\|\bar{\boldsymbol{Z}}(t+h, \boldsymbol{\beta}) - \bar{\boldsymbol{Z}}(t, \boldsymbol{\beta})\|_\infty$$
$$= \left\| \sum_{i \in R_t} \boldsymbol{Z}_i(t+h) w_i(t+h, \boldsymbol{\beta}) - \sum_{i \in R_t} \boldsymbol{Z}_i(t) w_i(t, \boldsymbol{\beta}) \right\|_\infty$$
$$\leq L_{\bar{\boldsymbol{Z}}}(\|\boldsymbol{\beta}\|_1) h + \frac{K_Z e^{2\|\boldsymbol{\beta}\|_1 K_Z}}{|R_t|},$$

where $L_{\bar{\boldsymbol{Z}}}(\|\boldsymbol{\beta}\|_1)$ was defined in (S1.2).

Now let $\Omega_1 := \bigl\{|R_{t_*}| \geq n\bar{F}_T(t_*) - n^{1/2}(\log n)\bar{F}_T^{1/2}(t_*)\bigr\}$, so that by a standard Binomial tail bound (e.g. Shorack and Wellner, 1986, p. 440),

$$\mathbb{P}(\Omega_1^c) \leq e^{-(\log^2 n)/2}.$$

Fix $\epsilon > \epsilon_n$, and partition $[0, t_*)$ into at most $\lceil t_*/h_0 \rceil + M \leq M_0$ intervals $\{[r_j, r_{j+1}) : j = 0, \ldots, M_0 - 1\}$ such that for each j, there exists k for which

$[r_j, r_{j+1}) \subseteq [s_k, s_{k+1})$, and such that $|r_{j+1} - r_j| \le h_0$. Then

$$\mathbb{P}\left(\sup_{h\in(0,h_0]} \max_{j=0,1,\ldots,M_0-1} \|\bar{\boldsymbol{Z}}(r_j + h, \boldsymbol{\beta}) - \bar{\boldsymbol{Z}}(r_j, \boldsymbol{\beta})\|_\infty > \frac{\epsilon}{3}\right) \le \mathbb{P}(\Omega_0^c) + \mathbb{P}(\Omega_1^c) \le \frac{1}{2n} + e^{-(\log^2 n)/2}. \tag{S1.4}$$

Finally, we seek to control the difference between $\bar{\boldsymbol{Z}}(\cdot, \boldsymbol{\beta})$ and $\boldsymbol{\mu}(\cdot, \boldsymbol{\beta})$ at $r_0, \ldots, r_{M_0}$. To this end, note that for any $t \in [0, t_*)$,

$$\begin{aligned}\big\|\bar{\boldsymbol{Z}}(t, \boldsymbol{\beta}) - \boldsymbol{\mu}(t, \boldsymbol{\beta})\big\|_\infty &= \left\|\frac{n^{-1}\sum_{i=1}^n \boldsymbol{Z}_i(t)\tilde{w}_i(t, \boldsymbol{\beta})}{n^{-1}\sum_{j=1}^n \tilde{w}_j(t, \boldsymbol{\beta})} - \frac{\mathbb{E}\big\{\boldsymbol{Z}(t)\tilde{w}(t, \boldsymbol{\beta})\big\}}{\mathbb{E}\big\{\tilde{w}(t, \boldsymbol{\beta})\big\}}\right\|_\infty \\ \le \frac{1}{n^{-1}\sum_{j=1}^n \tilde{w}_j(t, \boldsymbol{\beta})}&\left\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{Z}_i(t)\tilde{w}_i(t, \boldsymbol{\beta}) - \mathbb{E}\big\{\boldsymbol{Z}(t)\tilde{w}(t, \boldsymbol{\beta})\big\}\right\|_\infty \\ &+ K_Z\mathbb{E}\big\{\tilde{w}(t, \boldsymbol{\beta})\big\}\left|\frac{1}{n^{-1}\sum_{j=1}^n \tilde{w}_j(t, \boldsymbol{\beta})} - \frac{1}{\mathbb{E}\big\{\tilde{w}(t, \boldsymbol{\beta})\big\}}\right|.\end{aligned} \tag{S1.5}$$

Let

$$\Omega_2 := \left\{\frac{1}{n}\sum_{j=1}^n \tilde{w}_j(t, \boldsymbol{\beta}) \ge \frac{1}{2}\mathbb{E}\big\{\tilde{w}(t, \boldsymbol{\beta})\big\}\right\},$$

so that by Bernstein's inequality,

$$\begin{aligned}\mathbb{P}(\Omega_2^c) &\le \exp\left(-\frac{n\mathbb{E}^2\tilde{w}(t, \boldsymbol{\beta})}{8\big\{\mathbb{E}\tilde{w}^2(t, \boldsymbol{\beta}) + e^{\|\boldsymbol{\beta}\|_1 K_Z}\mathbb{E}\tilde{w}(t, \boldsymbol{\beta})/6\big\}}\right) \\ &\le \exp\left(-\frac{3n\bar{F}_T(t)}{28\exp(4\|\boldsymbol{\beta}\|_1 K_Z)}\right).\end{aligned} \tag{S1.6}$$

Then, by Bernstein's inequality again,

$$\begin{aligned}\mathbb{P}\biggl(\biggl\{\frac{1}{n^{-1}\sum_{j=1}^n \tilde{w}_j(t,\boldsymbol{\beta})}\biggl\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{Z}_i(t)\tilde{w}_i(t,\boldsymbol{\beta}) - \mathbb{E}\bigl\{\boldsymbol{Z}(t)\tilde{w}(t,\boldsymbol{\beta})\bigr\}\biggr\|_\infty > \frac{\epsilon}{2}\biggr\} \\ \bigcap \Omega_2\biggr) \\ \leq \mathbb{P}\biggl(\biggl\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{Z}_i(t)\tilde{w}_i(t,\boldsymbol{\beta}) - \mathbb{E}\bigl\{\boldsymbol{Z}(t)\tilde{w}(t,\boldsymbol{\beta})\bigr\}\biggr\|_\infty > \frac{\epsilon\mathbb{E}\bigl\{\tilde{w}(t,\boldsymbol{\beta})\bigr\}}{4}\biggr) \\ \leq 2p\exp\biggl(-\frac{n\epsilon^2\bar{F}_T(t)}{32K_Z\exp\bigl(4\|\boldsymbol{\beta}\|_1K_Z\bigr)(K_Z+\epsilon/12)}\biggr). \qquad \text{(S1.7)}\end{aligned}$$

Now the mean value theorem, for $x, y > 0$ with $x \geq y/2$,

$$\biggl|\frac{1}{x} - \frac{1}{y}\biggr| \leq \sup_{x_*=(1-t)x+ty:t\in[0,1]} \frac{|x-y|}{x_*^2} \leq \frac{4|x-y|}{y^2}.$$

Hence, by another application of Bernstein's inequality,

$$\begin{aligned}\mathbb{P}\biggl(\biggl\{K_Z\mathbb{E}\bigl\{\tilde{w}(t,\boldsymbol{\beta})\bigr\}\biggl|\frac{1}{n^{-1}\sum_{i=1}^n \tilde{w}_i(t,\boldsymbol{\beta})} - \frac{1}{\mathbb{E}\bigl\{\tilde{w}(t,\boldsymbol{\beta})\bigr\}}\biggr| > \frac{\epsilon}{2}\biggr\} \bigcap \Omega_2\biggr) \\ \leq \mathbb{P}\biggl(\biggl|\frac{1}{n}\sum_{i=1}^n \tilde{w}_i(t,\boldsymbol{\beta}) - \mathbb{E}\bigl\{\tilde{w}(t,\boldsymbol{\beta})\bigr\}\biggr| > \frac{\epsilon\mathbb{E}\bigl\{\tilde{w}(t,\boldsymbol{\beta})\bigr\}}{8K_Z}\biggr) \\ \leq 2\exp\biggl(-\frac{n\epsilon^2\bar{F}_T(t)}{128K_Z\exp(4\|\boldsymbol{\beta}\|_1K_Z)(K_Z+\epsilon/24)}\biggr). \\ \text{(S1.8)}\end{aligned}$$

It follows from (S1.5), (S1.6), (S1.7) and (S1.8) that for any $\epsilon > 0$ and

$t \in [0, t_*)$,

$$\begin{aligned}\mathbb{P}\big(\big\|\bar{\boldsymbol{Z}}(t,\boldsymbol{\beta}) - \boldsymbol{\mu}(t,\boldsymbol{\beta})\big\|_\infty > \epsilon\big) \\ \leq (2p+2)\exp\biggl(-\frac{n\epsilon^2\bar{F}_T(t)}{128K_Z\exp\bigl(4\|\boldsymbol{\beta}\|_1K_Z\bigr)(K_Z+\epsilon/12)}\biggr) \\ + \exp\biggl(-\frac{3n\bar{F}_T(t)}{28\exp(4\|\boldsymbol{\beta}\|_1K_Z)}\biggr).\end{aligned} \tag{S1.9}$$

From (S1.3), (S1.4) and (S1.9), together with the fact that $L_{\boldsymbol{\mu}}(\|\boldsymbol{\beta}\|_1)h_0 \leq \epsilon/3$, we conclude that for $\epsilon > \epsilon_n$,

$$\begin{aligned}&\mathbb{P}\biggl(\sup_{t\in[0,t_*)}\|\bar{\boldsymbol{Z}}(t,\boldsymbol{\beta}) - \boldsymbol{\mu}(t,\boldsymbol{\beta})\|_\infty > \epsilon\biggr) \\ &\leq \mathbb{P}\biggl(\sup_{h\in(0,h_0]}\max_{j=0,1,\ldots,M_0-1}\|\bar{\boldsymbol{Z}}(r_j+h,\boldsymbol{\beta}) - \bar{\boldsymbol{Z}}(r_j,\boldsymbol{\beta})\|_\infty > \frac{\epsilon}{3}\biggr) \\ &\qquad + \mathbb{P}\biggl(\max_{j=0,1,\ldots,M_0-1}\|\bar{\boldsymbol{Z}}(r_j,\boldsymbol{\beta}) - \boldsymbol{\mu}(r_j,\boldsymbol{\beta})\|_\infty > \frac{\epsilon}{3}\biggr) \\ &\leq \frac{1}{2n} + (2p+2)M_0\exp\biggl(-\frac{n\epsilon^2\bar{F}_T(t_*)}{1152K_Z\exp\bigl(4\|\boldsymbol{\beta}\|_1K_Z\bigr)(K_Z+\epsilon/36)}\biggr) \\ &\qquad + e^{-(\log^2 n)/2} + M_0\exp\biggl(-\frac{3n\bar{F}_T(t_*)}{28\exp(4\|\boldsymbol{\beta}\|_1K_Z)}\biggr),\end{aligned}$$

as required. □

The lemma below is used several times in controlling the residual terms in (2.7).

Lemma S3. *Assume* ***(A1)****,* ***(A2)(a)****,* ***(A3)(b)*** *and* ***(A4)(a)****. For* $\widehat{\boldsymbol{\beta}}$

in (2.2)*, let* $\lambda \asymp n^{-1/2}\log^{1/2}(np)$*. Then, for every* $\eta \in (0, 1/3)$*,*

$$\|\widehat{\mathcal{V}}(\widehat{\boldsymbol{\beta}})\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}\|_\infty = O_p\biggl\{\max\biggl(\|\boldsymbol{\Sigma}^{-1}\|_{\mathrm{op},1}\frac{d_o\log^{1/2}(np)}{n^{1/2}}, \; \|\boldsymbol{\Sigma}^{-1}\|_{\mathrm{op},1}n^{-(1/3-\eta)}\biggr)\biggr\}.$$

Proof. Writing $W_1 := n^{-1}\sum_{i=1}^n \int_{\mathcal{T}}\{\boldsymbol{Z}_i(s) - \boldsymbol{\mu}(s, \boldsymbol{\beta}^o)\}^{\otimes 2}\, dN_i(s) - \boldsymbol{\Sigma}$, we have

$$\begin{aligned}
&\|\widehat{\mathcal{V}}(\widehat{\boldsymbol{\beta}}) - \boldsymbol{\Sigma}\|_\infty \\
&\leq \biggl\|\frac{1}{n}\sum_{i=1}^n\int_{\mathcal{T}}\bigl\{(\boldsymbol{Z}_i(s) - \bar{\boldsymbol{Z}}(s, \widehat{\boldsymbol{\beta}}))^{\otimes 2} - (\boldsymbol{Z}_i(s) - \boldsymbol{\mu}(s, \boldsymbol{\beta}^o))^{\otimes 2}\bigr\}\, dN_i(s)\biggr\|_\infty \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad + \|W_1\|_\infty \\
&\leq \biggl\|\frac{1}{n}\sum_{i=1}^n\int_{\mathcal{T}}\bigl(\boldsymbol{Z}_i(s) - \bar{\boldsymbol{Z}}(s, \widehat{\boldsymbol{\beta}})\bigr)\bigl(\bar{\boldsymbol{Z}}(s, \widehat{\boldsymbol{\beta}}) - \boldsymbol{\mu}(s, \boldsymbol{\beta}^o)\bigr)^\top dN_i(s)\biggr\|_\infty \\
&\quad + \biggl\|\frac{1}{n}\sum_{i=1}^n\int_{\mathcal{T}}\bigl(\boldsymbol{Z}_i(s) - \boldsymbol{\mu}(s, \boldsymbol{\beta}^o)\bigr)\bigl(\bar{\boldsymbol{Z}}(s, \widehat{\boldsymbol{\beta}}) - \boldsymbol{\mu}(s, \boldsymbol{\beta}^o)\bigr)^\top dN_i(s)\biggr\|_\infty + \|W_1\|_\infty \\
&\leq \frac{4K_Z}{n}\sum_{i=1}^n\int_{\mathcal{T}}\|\bar{\boldsymbol{Z}}(s, \widehat{\boldsymbol{\beta}}) - \boldsymbol{\mu}(s, \boldsymbol{\beta}^o)\|_\infty\, dN_i(s) + \|W_1\|_\infty \\
&\leq \frac{4K_Z}{n}\sum_{i=1}^n\int_{\mathcal{T}}\|\bar{\boldsymbol{Z}}(s, \widehat{\boldsymbol{\beta}}) - \bar{\boldsymbol{Z}}(s, \boldsymbol{\beta}^o)\|_\infty\, dN_i(s) \\
&\qquad\qquad + \frac{4K_Z}{n}\sum_{i=1}^n\int_{\mathcal{T}}\|\bar{\boldsymbol{Z}}(s, \boldsymbol{\beta}^o) - \boldsymbol{\mu}(s, \boldsymbol{\beta}^o)\|_\infty\, dN_i(s) + \|W_1\|_\infty.
\end{aligned} \tag{S1.10}$$

Now, for any $s \in \mathcal{T}$,

$$\begin{aligned}
&\|\bar{\boldsymbol{Z}}(s,\widehat{\boldsymbol{\beta}}) - \bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o)\|_\infty \\
&= \left\| \frac{\sum_{i=1}^n \boldsymbol{Z}_i(s) Y_i(s) e^{\widehat{\boldsymbol{\beta}}^\top \boldsymbol{Z}_i(s)}}{\sum_{i=1}^n Y_i(s) e^{\widehat{\boldsymbol{\beta}}^\top \boldsymbol{Z}_i(s)}} - \frac{\sum_{i=1}^n \boldsymbol{Z}_i(s) Y_i(s) e^{\boldsymbol{\beta}^{o\top} \boldsymbol{Z}_i(s)}}{\sum_{i=1}^n Y_i(s) e^{\boldsymbol{\beta}^{o\top} \boldsymbol{Z}_i(s)}} \right\|_\infty \\
&\le 2K_Z (e^{K_Z \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^o\|_1} - 1). \qquad \text{(S1.11)}
\end{aligned}$$

It therefore follows from Lemma 1(ii) that

$$\frac{4K_Z}{n} \sum_{i=1}^n \int_{\mathcal{T}} \|\bar{\boldsymbol{Z}}(s,\widehat{\boldsymbol{\beta}}) - \bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o)\|_\infty \, dN_i(s) = O_p\biggl(\frac{d_o \log^{1/2}(np)}{n^{1/2}} \biggr). \qquad \text{(S1.12)}$$

Next, fix an arbitrary $\eta \in (0, 1/3)$ and let $t_* := F_T^{-1}(1 - n^{-(1/3-\eta)})$. Recalling that $R_t = \{i : Y_i(t) = 1\}$, set $\Omega_* := \bigl\{ n\bar{F}(t_*)/2 \le |R_{t_*}| \le 3n\bar{F}(t_*)/2 \bigr\}$. Then by Hoeffding's inequality,

$$\mathbb{P}(\Omega_*^c) \le 2e^{-n^{1/3+2\eta}/2}. \qquad \text{(S1.13)}$$

Moreover, on Ω_*,

$$\begin{aligned}
&\frac{4K_Z}{n} \sum_{i=1}^n \int_{\mathcal{T}} \|\bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o) - \boldsymbol{\mu}(s,\boldsymbol{\beta}^o)\|_\infty \, dN_i(s) \\
&\le 4K_Z \sup_{s \in [0,t_*)} \|\bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o) - \boldsymbol{\mu}(s,\boldsymbol{\beta}^o)\|_\infty + 12K_Z^2 n^{-(1/3-\eta)} \\
&= O_p(n^{-(1/3-\eta)}), \qquad \text{(S1.14)}
\end{aligned}$$

where the final bound follows from Lemma S2. Finally, by **(A1)(a)** and Hoeffding's inequality, we have that for every $x > 0$,

$$\mathbb{P}\bigl(\|W_1\|_\infty > x\bigr) \le p(p+1)e^{-nx^2/(32K_Z^4)},$$

so that

$$\|W_1\|_\infty = O_p\biggl(\frac{\log^{1/2}(np)}{n^{1/2}}\biggr). \tag{S1.15}$$

We conclude from (S1.10), (S1.12), (S1.13), (S1.14) and (S1.15) that for every $\eta \in (0, 1/3)$,

$$\begin{aligned}\|\widehat{\mathcal{V}}(\widehat{\boldsymbol{\beta}})\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}\|_\infty &\leq \|\boldsymbol{\Sigma}^{-1}\|_{\mathrm{op},1}\|\widehat{\mathcal{V}}(\widehat{\boldsymbol{\beta}}) - \boldsymbol{\Sigma}\|_\infty \\ &= O_p\biggl\{\max\biggl(\|\boldsymbol{\Sigma}^{-1}\|_{\mathrm{op},1}\frac{d_o \log^{1/2}(np)}{n^{1/2}}, \|\boldsymbol{\Sigma}^{-1}\|_{\mathrm{op},1} n^{-(1/3-\eta)}\biggr)\biggr\},\end{aligned}$$

as required. □

The following result is a consequence of Cai et al. (2016, Lemma 7.1).

Lemma S4. *Let $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_p)^\top = (\Theta_{ij}) \in \mathbb{R}^{p\times p}$ be symmetric and let $\check{\boldsymbol{\Theta}} = (\check{\boldsymbol{\Theta}}_1, \ldots, \check{\boldsymbol{\Theta}}_p)^\top$ be an estimator of $\boldsymbol{\Theta}$. On the event*

$$\bigl\{\|\check{\boldsymbol{\Theta}}_j\|_1 \leq \|\boldsymbol{\Theta}_j\|_1,\, j = 1, \ldots, p\bigr\},$$

we have

$$\|\check{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_{\mathrm{op},\infty} \leq 12\|\check{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_\infty \max_{j=1,\ldots,p}\sum_{i=1}^p \mathbb{1}_{\{\Theta_{ij}\neq 0\}}.$$

S2 Proofs of Propositions

Proof of Proposition 1. Writing M_i for the mean-zero martingale in the Doob–Meyer decomposition of N_i (cf. (2.3)), we have

$$\begin{aligned} n^{1/2}\boldsymbol{c}^\top \boldsymbol{\Sigma}^{-1}\dot{\ell}(\boldsymbol{\beta}^o) &= \frac{1}{n^{1/2}}\sum_{i=1}^n \int_{\mathcal{T}} \boldsymbol{c}^\top \boldsymbol{\Sigma}^{-1}\big\{\boldsymbol{Z}_i(s) - \bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o)\big\}\,dN_i(s) \\ &= \frac{1}{n^{1/2}}\sum_{i=1}^n \int_{\mathcal{T}} \boldsymbol{c}^\top \boldsymbol{\Sigma}^{-1}\big\{\boldsymbol{Z}_i(s) - \bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o)\big\}\,dM_i(s) \\ &= \frac{1}{n^{1/2}}\sum_{i=1}^n \int_{\mathcal{T}} \boldsymbol{c}^\top \boldsymbol{\Sigma}^{-1}\big\{\boldsymbol{Z}_i(s) - \boldsymbol{\mu}(s,\boldsymbol{\beta}^o)\big\}\,dM_i(s) \\ &\quad - \frac{1}{n^{1/2}}\sum_{i=1}^n \int_{\mathcal{T}} \boldsymbol{c}^\top \boldsymbol{\Sigma}^{-1}\big\{\bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o) - \boldsymbol{\mu}(s,\boldsymbol{\beta}^o)\big\}\,dM_i(s) \\ &=: \frac{1}{n^{1/2}}\sum_{i=1}^n U_{ni} - \frac{1}{n^{1/2}}\sum_{i=1}^n V_{ni}, \end{aligned}$$

say. Now, for each $n \in \mathbb{N}$, we have that $U_{n1},\ldots,U_{nn}$ are independent and identically distributed, with $\mathbb{E}(U_{n1}) = 0$ and $\mathrm{Var}(U_{n1}) = \boldsymbol{c}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{c}$. Moreover, for every $\epsilon > 0$,

$$\begin{aligned} &\frac{1}{n\boldsymbol{c}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{c}}\sum_{i=1}^n \mathbb{E}\big(U_{ni}^2 \mathbb{1}_{\{|U_{ni}|>\epsilon n^{1/2}(\boldsymbol{c}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{c})^{1/2}\}}\big) \\ &\qquad = \frac{1}{\boldsymbol{c}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{c}}\mathbb{E}\big(U_{n1}^2 \mathbb{1}_{\{|U_{n1}|>\epsilon n^{1/2}(\boldsymbol{c}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{c})^{1/2}\}}\big) \to 0 \end{aligned}$$

as $n \to \infty$. It follows by the Lindeberg–Feller central limit theorem (e.g. Gut, 2005, Theorem 7.2.1) that

$$\frac{1}{n^{1/2}}\sum_{i=1}^n U_{ni} \xrightarrow{d} \mathcal{N}(0,\nu^2).$$

Next, we observe that $V_{n1}, \dots, V_{nn}$ are exchangeable, with $\mathbb{E}(V_{n1}) = 0$. Moreover, by, e.g., Andersen et al. (1993, pp. 74–75),

$$\begin{aligned} &\mathrm{Var}\biggl(\frac{1}{n^{1/2}}\sum_{i=1}^{n} V_{ni}\biggr) = \mathrm{Var}(V_{n1}) \\ &= \boldsymbol{c}^\top \boldsymbol{\Sigma}^{-1}\mathbb{E}\biggl(\int_0^{t_+} \bigl\{\bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o) - \boldsymbol{\mu}(s,\boldsymbol{\beta}^o)\bigr\}^{\otimes 2}\tilde{w}_1(s,\boldsymbol{\beta}^o)\lambda_0(s)\,ds\biggr)\boldsymbol{\Sigma}^{-1}\boldsymbol{c}. \end{aligned} \quad \text{(S2.1)}$$

Now write $t_* := F_T^{-1}(1 - n^{-1/2})$ and $\boldsymbol{\mathcal{Z}}_1 := \{\boldsymbol{Z}_1(t) : t \in \mathcal{T}\}$ and let $S_1 := -\log \bar{F}_{\tilde{T}_1|\boldsymbol{\mathcal{Z}}_1}(T_1) \leq -\log \bar{F}_{T_1|\boldsymbol{\mathcal{Z}}_1}(T_1) =: Q_1$, say, where $Q_1|\boldsymbol{\mathcal{Z}}_1 \sim \mathrm{Exp}(1)$. Then

$$\begin{aligned} &\mathbb{E}\biggl(\int_0^{t_+} \bigl\|\bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o) - \boldsymbol{\mu}(s,\boldsymbol{\beta}^o)\bigr\|_\infty^2 \tilde{w}_1(s,\boldsymbol{\beta}^o)\lambda_0(s)\,ds\biggr) \\ &\quad\leq \mathbb{E}\biggl\{\sup_{s\in[0,t_*)}\bigl\|\bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o) - \boldsymbol{\mu}(s,\boldsymbol{\beta}^o)\bigr\|_\infty^2 \mathbb{E}(S_1|\boldsymbol{\mathcal{Z}}_1)\biggr\} \\ &\qquad\qquad + 4K_Z^2\mathbb{E}\biggl(\int_{t_*}^{T_1} e^{\boldsymbol{\beta}^{o\top}\boldsymbol{Z}_1(s)}\lambda_0(s)\,ds\,\mathbb{1}_{\{T_1\geq t_*\}}\biggr) \\ &\quad\leq 2^{1/2}\biggl[\mathbb{E}\biggl\{\sup_{s\in[0,t_*)}\bigl\|\bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o) - \boldsymbol{\mu}(s,\boldsymbol{\beta}^o)\bigr\|_\infty^4\biggr\}\biggr]^{1/2} \\ &\qquad\qquad + 4K_Z^2\mathbb{E}\bigl\{\mathbb{E}\bigl(S_1\mathbb{1}_{\{T_1\geq t_*\}} \bigm| \boldsymbol{\mathcal{Z}}_1\bigr)\bigr\}. \end{aligned} \quad \text{(S2.2)}$$

First note that

$$\begin{aligned} \mathbb{E}\bigl\{\mathbb{E}\bigl(S_1\mathbb{1}_{\{T_1\geq t_*\}} \bigm| \boldsymbol{\mathcal{Z}}_1\bigr)\bigr\} &\leq \mathbb{E}\bigl\{\mathbb{E}\bigl(Q_1\mathbb{1}_{\{Q_1\geq -\log \bar{F}_{T_1|\boldsymbol{\mathcal{Z}}_1}(t_*)\}} \bigm| \boldsymbol{\mathcal{Z}}_1\bigr)\bigr\} \\ &= \mathbb{E}\bigl[\bigl\{1 - \log \bar{F}_{T_1|\boldsymbol{\mathcal{Z}}_1}(t_*)\bigr\}\bar{F}_{T_1|\boldsymbol{\mathcal{Z}}_1}(t_*)\bigr] \\ &\leq \mathbb{E}\bar{F}_{T_1|\boldsymbol{\mathcal{Z}}_1}(t_*) - \mathbb{E}\bar{F}_{T_1|\boldsymbol{\mathcal{Z}}_1}(t_*)\log \mathbb{E}\bar{F}_{T_1|\boldsymbol{\mathcal{Z}}_1}(t_*) \\ &= \frac{1}{n^{1/2}} + \frac{\log n}{2n^{1/2}}, \end{aligned} \quad \text{(S2.3)}$$

where the second inequality follows by Jensen's inequality. Now let $C = C(\|\boldsymbol{\beta}^o\|_1) := 1152^{1/2} K_Z \exp(2\|\boldsymbol{\beta}^o\|_1 K_Z)$ and choose $\epsilon_{n,*} > 0$ such that

$$8(p+1)n^{\max\{3,1+1/(2\alpha)\}} \exp\biggl(-\frac{n^{1/2}\epsilon_{n,*}^2}{2C^2}\biggr) = 1,$$

where $\alpha > 0$ is taken from **(A2)(b)**. Thus, by **(A3)(a)**, we have $\epsilon_{n,*} = o(n^{-1/4+\delta})$ for every $\delta > 0$ and $n^{1/4}\epsilon_{n,*} \to \infty$. Observe that

$$\frac{1}{n^{1/2}} = \int_{t_*}^{\infty} dF_T(t) \leq \frac{1}{t_*^{\alpha}} \int_0^{\infty} t^{\alpha}\, dF_T(t),$$

so by **(A2)(b)**, $t_* = O(n^{1/(2\alpha)})$. Noting the definition of $h_0 = h_0(n, \epsilon)$ in Lemma S2 in the Appendix, we choose $n_0 \in \mathbb{N}$ large enough that the following conditions hold for $n \geq n_0$:

1. $\epsilon_{n,*} \leq 2K_Z$

2. $n\bar{F}_T(t_*) - n^{1/2}(\log n)\bar{F}_T^{1/2}(t_*) = n^{1/2} - n^{1/4}\log n \geq n^{1/2}/2$

3. $\epsilon_{n,*} - \frac{6K_Z e^{\|\boldsymbol{\beta}^o\|_1 K_Z}}{n^{1/2}} \geq \epsilon_{n,*}/2$

4. $1 + t_*/h_0(n, \epsilon_{n,*}) \leq n^{1+1/(2\alpha)}$.

It follows that $M_0^* := M_0(n, \epsilon_{n,*})$, defined in Lemma S2, satisfies $M_0^* \leq 2n^{\max\{3,1+1/(2\alpha)\}}$ for $n \geq n_0$. Write

$$g(n) := \frac{1}{2n} + e^{-(\log^2 n)/2} + 4n^{\max\{3,1+1/(2\alpha)\}} \exp\biggl\{-\frac{3n^{1/2}}{28\exp(4\|\boldsymbol{\beta}^o\|_1 K_Z)}\biggr\}.$$

Then, by Lemma S2, for $n \geq n_0$,

$$\begin{aligned}
&\mathbb{E}\biggl\{\sup_{s\in[0,t_*)}\bigl\|\bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o)-\boldsymbol{\mu}(s,\boldsymbol{\beta}^o)\bigr\|_\infty^4\biggr\}\\
&=\int_0^{16K_Z^4}\mathbb{P}\biggl(\sup_{s\in[0,t_*)}\bigl\|\bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o)-\boldsymbol{\mu}(s,\boldsymbol{\beta}^o)\bigr\|_\infty^4>\delta\biggr)\,d\delta\\
&\leq \epsilon_{n,*}^4+4\int_{\epsilon_{n,*}}^{2K_Z}\epsilon^3\mathbb{P}\biggl(\sup_{s\in[0,t_*)}\bigl\|\bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o)-\boldsymbol{\mu}(s,\boldsymbol{\beta}^o)\bigr\|_\infty>\epsilon\biggr)\,d\epsilon\\
&\leq \epsilon_{n,*}^4+8(p+1)M_0^*\int_{\epsilon_{n,*}}^{\infty}\epsilon^3\exp\biggl(-\frac{n^{1/2}\epsilon^2}{2C^2}\biggr)\,d\epsilon+16K_Z^4g(n)\\
&=\epsilon_{n,*}^4+\frac{8(p+1)M_0^*C^4}{n}\int_{n^{1/4}\epsilon_{n,*}/C}^{\infty}t^3e^{-t^2/2}\,dt+16K_Z^4g(n)\\
&\leq \epsilon_{n,*}^4+\frac{2C^4\bigl\{\log\bigl(16(p+1)n^{\max\{3,1+1/(2\alpha)\}}\bigr)+1\bigr\}}{n}+16K_Z^4g(n)=o(n^{-(1-\delta)}),
\end{aligned} \tag{S2.4}$$

for every $\delta > 0$. From (S2.1), (S2.2), (S2.3) and (S2.4) and **(A4)(c)**, we deduce that

$$\begin{aligned}
&\mathrm{Var}\biggl(\frac{1}{n^{1/2}}\sum_{i=1}^n V_{ni}\biggr)\\
&\leq\|\boldsymbol{\Sigma}^{-1}\|_{\mathrm{op},1}^2\mathbb{E}\biggl(\int_0^{t_+}\bigl\|\bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o)-\boldsymbol{\mu}(s,\boldsymbol{\beta}^o)\bigr\|_\infty^2\tilde{w}_1(s,\boldsymbol{\beta}^o)\lambda_0(s)\,ds\biggr)\to 0,
\end{aligned}$$

as required. □

Proof of Proposition 2. Define the event $\mathcal{A} := \{\|\boldsymbol{\Sigma}^{-1}\widehat{\mathcal{V}}(\widehat{\boldsymbol{\beta}}) - \boldsymbol{I}\|_\infty \leq \lambda_n\} = \{\|\widehat{\mathcal{V}}(\widehat{\boldsymbol{\beta}})\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}\|_\infty \leq \lambda_n\}$. Then by construction of $\widehat{\boldsymbol{\Theta}} = (\widehat{\boldsymbol{\Theta}}_1, \ldots, \widehat{\boldsymbol{\Theta}}_p)^\top$ as an estimator of $\boldsymbol{\Sigma}^{-1} = \bigl((\boldsymbol{\Sigma}^{-1})_1, \ldots, (\boldsymbol{\Sigma}^{-1})_p\bigr)^\top$, on the event $\mathcal{A}$, we have

$$\|\widehat{\boldsymbol{\Theta}}_j\|_1 \leq \|(\boldsymbol{\Sigma}^{-1})_j\|_1, \quad j = 1, \ldots, p,$$

so in particular, $\|\widehat{\mathbf{\Theta}}\|_{\text{op},\infty} \le \|\mathbf{\Sigma}^{-1}\|_{\text{op},\infty}$, and

$$\big\|\widehat{\mathbf{\Theta}}\widehat{\mathcal{V}}(\widehat{\boldsymbol{\beta}}) - \boldsymbol{I}\big\|_\infty \le \lambda_n.$$

Hence, using the fact that $\mathbf{\Sigma}$ and $\widehat{\mathcal{V}}(\widehat{\boldsymbol{\beta}})$ are symmetric, on the event $\mathcal{A}$,

$$\begin{aligned}&\|\widehat{\mathbf{\Theta}} - \mathbf{\Sigma}^{-1}\|_\infty = \big\|\widehat{\mathbf{\Theta}}\big(\boldsymbol{I} - \widehat{\mathcal{V}}(\widehat{\boldsymbol{\beta}})\mathbf{\Sigma}^{-1}\big) + \big(\widehat{\mathbf{\Theta}}\widehat{\mathcal{V}}(\widehat{\boldsymbol{\beta}}) - \boldsymbol{I}\big)\mathbf{\Sigma}^{-1}\big\|_\infty \\ &\le \|\widehat{\mathbf{\Theta}}\|_{\text{op},\infty}\|\widehat{\mathcal{V}}(\widehat{\boldsymbol{\beta}})\mathbf{\Sigma}^{-1} - \boldsymbol{I}\|_\infty + \|\mathbf{\Sigma}^{-1}\|_{\text{op},1}\|\widehat{\mathbf{\Theta}}\widehat{\mathcal{V}}(\widehat{\boldsymbol{\beta}}) - \boldsymbol{I}\|_\infty \le 2\lambda_n\|\mathbf{\Sigma}^{-1}\|_{\text{op},1}.\end{aligned} \tag{S2.5}$$

Hence, from Lemma S4, on the event $\mathcal{A}$,

$$\begin{aligned}|\boldsymbol{c}^\top(\widehat{\mathbf{\Theta}} - \mathbf{\Sigma}^{-1})\dot{\ell}(\boldsymbol{\beta}^o)| &\le \|\widehat{\mathbf{\Theta}} - \mathbf{\Sigma}^{-1}\|_{\text{op},\infty}\|\dot{\ell}(\boldsymbol{\beta}^o)\|_\infty \\ &\le 24\lambda_n\|\mathbf{\Sigma}^{-1}\|_{\text{op},1}\|\dot{\ell}(\boldsymbol{\beta}^o)\|_\infty \max_{j=1,\dots,p} r_j\end{aligned}$$

The desired conclusion therefore follows from Lemmas 1(i) and S3, together with **(A4)(c)**. □

Proof of Proposition 3. Note that

$$\begin{aligned}&\|\widehat{\mathbf{\Theta}}\boldsymbol{M}(\widetilde{\boldsymbol{\beta}}) + \boldsymbol{I}\|_\infty \\ &\le \|\widehat{\mathbf{\Theta}} - \mathbf{\Sigma}^{-1}\|_{\text{op},\infty}\|\boldsymbol{M}(\widetilde{\boldsymbol{\beta}}) + \mathbf{\Sigma}\|_\infty \\ &\qquad + \|\mathbf{\Sigma}^{-1}\|_{\text{op},1}\|\boldsymbol{M}(\widetilde{\boldsymbol{\beta}}) + \mathbf{\Sigma}\|_\infty + \|\widehat{\mathbf{\Theta}} - \mathbf{\Sigma}^{-1}\|_{\text{op},\infty}\|\mathbf{\Sigma}\|_\infty.\end{aligned} \tag{S2.6}$$

For $j = 1, \dots, p$, let $\ddot{\ell}_j(\boldsymbol{\beta}^o)$ denote the jth column of $\ddot{\ell}(\boldsymbol{\beta}^o)$. Then

$$\begin{aligned}
&\|\ddot{\ell}_j(\widetilde{\boldsymbol{\beta}}_j) - \ddot{\ell}_j(\boldsymbol{\beta}^o)\|_\infty \\
&\quad \leq \biggl\|\sum_{i=1}^n \int_{\mathcal{T}} \Bigl[\bigl\{\boldsymbol{Z}_i(s) - \bar{\boldsymbol{Z}}(s, \widetilde{\boldsymbol{\beta}}_j)\bigr\}^{\otimes 2} w_i(s, \widetilde{\boldsymbol{\beta}}_j) \\
&\qquad\qquad - \bigl\{\boldsymbol{Z}_i(s) - \bar{\boldsymbol{Z}}(s, \boldsymbol{\beta}^o)\bigr\}^{\otimes 2} w_i(s, \boldsymbol{\beta}^o)\Bigr]\, d\bar{N}(s)\biggr\|_\infty \\
&\quad \leq \sup_{s \in \mathcal{T}} \biggl\|\sum_{i=1}^n \Bigl[\bigl\{\boldsymbol{Z}_i(s) - \bar{\boldsymbol{Z}}(s, \widetilde{\boldsymbol{\beta}}_j)\bigr\}^{\otimes 2} w_i(s, \widetilde{\boldsymbol{\beta}}_j) \\
&\qquad\qquad - \bigl\{\boldsymbol{Z}_i(s) - \bar{\boldsymbol{Z}}(s, \boldsymbol{\beta}^o)\bigr\}^{\otimes 2} w_i(s, \boldsymbol{\beta}^o)\Bigr]\biggr\|_\infty \\
&\quad \leq 4K_Z \sup_{s \in \mathcal{T}} \bigl\|\bar{\boldsymbol{Z}}(s, \widetilde{\boldsymbol{\beta}}_j) - \bar{\boldsymbol{Z}}(s, \boldsymbol{\beta}^o)\bigr\|_\infty + 4K_Z^2 \sup_{s \in \mathcal{T}} \sum_{i=1}^n |w_i(s, \widetilde{\boldsymbol{\beta}}_j) - w_i(s, \boldsymbol{\beta}^o)|.
\end{aligned} \tag{S2.7}$$

But, for any $s \in \mathcal{T}$,

$$\begin{aligned}
&\bigl\|\bar{\boldsymbol{Z}}(s, \widetilde{\boldsymbol{\beta}}_j) - \bar{\boldsymbol{Z}}(s, \boldsymbol{\beta}^o)\bigr\|_\infty \\
&\qquad = \left\|\frac{\sum_{i=1}^n \boldsymbol{Z}_i(s) Y_i(s) e^{\widetilde{\boldsymbol{\beta}}_j^\top \boldsymbol{Z}_i(s)}}{\sum_{i=1}^n Y_i(s) e^{\widetilde{\boldsymbol{\beta}}_j^\top \boldsymbol{Z}_i(s)}} - \frac{\sum_{i=1}^n \boldsymbol{Z}_i(s) Y_i(s) e^{\boldsymbol{\beta}^{o\top} \boldsymbol{Z}_i(s)}}{\sum_{i=1}^n Y_i(s) e^{\boldsymbol{\beta}^{o\top} \boldsymbol{Z}_i(s)}}\right\|_\infty \\
&\qquad \leq 2K_Z \sum_{i=1}^n \frac{Y_i(s) |e^{\widetilde{\boldsymbol{\beta}}_j^\top \boldsymbol{Z}_i(s)} - e^{\boldsymbol{\beta}^{o\top} \boldsymbol{Z}_i(s)}|}{\sum_{\ell=1}^n Y_\ell(s) e^{\widetilde{\boldsymbol{\beta}}_j^\top \boldsymbol{Z}_\ell(s)}} \\
&\qquad \leq 2K_Z (e^{K_Z \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^o\|_1} - 1).
\end{aligned} \tag{S2.8}$$

Similarly, for any $s \in \mathcal{T}$,

$$\begin{aligned}
\sum_{i=1}^n |w_i(s, \widetilde{\boldsymbol{\beta}}_j) - w_i(s, \boldsymbol{\beta}^o)| &= \sum_{i=1}^n \left|\frac{Y_i(s) e^{\widetilde{\boldsymbol{\beta}}_j^\top \boldsymbol{Z}_i(s)}}{\sum_{\ell=1}^n Y_\ell(s) e^{\widetilde{\boldsymbol{\beta}}_j^\top \boldsymbol{Z}_\ell(s)}} - \frac{Y_i(s) e^{\boldsymbol{\beta}^{o\top} \boldsymbol{Z}_i(s)}}{\sum_{\ell=1}^n Y_\ell(s) e^{\boldsymbol{\beta}^{o\top} \boldsymbol{Z}_\ell(s)}}\right| \\
\leq 2 \sum_{i=1}^n \frac{Y_i(s) |e^{\widetilde{\boldsymbol{\beta}}_j^\top \boldsymbol{Z}_i(s)} - e^{\boldsymbol{\beta}^{o\top} \boldsymbol{Z}_i(s)}|}{\sum_{\ell=1}^n Y_\ell(s) e^{\widetilde{\boldsymbol{\beta}}_j^\top \boldsymbol{Z}_\ell(s)}} &\leq 2\bigl(e^{K_Z \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^o\|_1} - 1\bigr).
\end{aligned} \tag{S2.9}$$

It follows from (S2.7), (S2.8) and (S2.9) that

$$\|\boldsymbol{M}(\widetilde{\boldsymbol{\beta}}) - \ddot{\ell}(\boldsymbol{\beta}^o)\|_\infty = \max_{j=1,\dots,p} \|\ddot{\ell}_j(\widetilde{\boldsymbol{\beta}}_j) - \ddot{\ell}_j(\boldsymbol{\beta}^o)\|_\infty \le 16K_Z^2(e^{K_Z\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^o\|_1} - 1). \tag{S2.10}$$

Moreover,

$$\begin{aligned}
&\|\ddot{\ell}(\boldsymbol{\beta}^o) + \boldsymbol{\Sigma}\|_\infty \\
&\qquad \le \left\| \frac{1}{n}\sum_{i=1}^n \int_{\mathcal{T}} \sum_{j=1}^n \big\{\boldsymbol{Z}_j(s) - \bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o)\big\}^{\otimes 2} w_j(s,\boldsymbol{\beta}^o)\, dM_i(s) \right\|_\infty \\
&\qquad\qquad + \left\| \frac{1}{n}\sum_{i=1}^n \int_0^{t_+} \big\{\boldsymbol{Z}_i(s) - \bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o)\big\}^{\otimes 2} \tilde{w}_i(s,\boldsymbol{\beta}^o)\lambda_0(s)\, ds - \boldsymbol{\Sigma} \right\|_\infty .
\end{aligned} \tag{S2.11}$$

The first term is the entrywise maximum absolute norm of a random $p \times p$ matrix. Fixing $j, k \in \{1, \dots, p\}$, it is convenient to write its (j,k)th entry as $n^{-1}\sum_{i=1}^n \int_{\mathcal{T}} a(s)\, dM_i(s)$, with

$$a(s) = a_{j,k}(s) := \sum_{i=1}^n \big\{Z_{ij}(s) - \bar{Z}_j(s,\boldsymbol{\beta}^o)\big\}\big\{Z_{ik}(s) - \bar{Z}_k(s,\boldsymbol{\beta}^o)\big\} w_i(s,\boldsymbol{\beta}^o),$$

where $Z_{ij}(s)$ and $\bar{Z}_j(s,\boldsymbol{\beta}^o)$ are the jth components of $\boldsymbol{Z}_i(s)$ and $\bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o)$ respectively. For $t \in \mathcal{T}$, we also define the right-continuous martingale

$$W_t := \frac{1}{n}\sum_{i=1}^n \int_{[0,t]} a(s)\, dM_i(s),$$

and claim that $(W_t : t \in \mathcal{T})$ is uniformly integrable. To see this, note that

$$\begin{aligned}\sup_{t\in\mathcal{T}} \mathbb{E}(W_t^2) &= \frac{1}{n}\sup_{t\in\mathcal{T}} \mathbb{E}\int_0^t a(s)^2 Y(s) e^{\boldsymbol{\beta}^{o\top}\boldsymbol{Z}(s)}\lambda_0(s)\,ds \\ &\le \frac{16K_Z^4}{n}\sup_{t\in\mathcal{T}} \mathbb{E}\big[\mathbb{E}\{-\log \bar{F}_{\tilde{T}|\boldsymbol{\mathcal{Z}}}(T\wedge t)|\boldsymbol{\mathcal{Z}}\}\big] \\ &\le \frac{16K_Z^4}{n}\mathbb{E}\big[\mathbb{E}\big\{-\log \bar{F}_{\tilde{T}|\boldsymbol{\mathcal{Z}}}(\tilde{T})|\boldsymbol{\mathcal{Z}}\big\}\big] = \frac{16K_Z^4}{n},\end{aligned}$$

which establishes the desired uniform integrability. Thus, by, e.g., Karatzas and Shreve (1991, p.18) there exists a random variable W_{t_+} such that $\mathbb{E}(|W_{t_+}|) < \infty$ and $W_t \overset{\text{a.s.}}{\to} W_{t_+}$ as $t \to t_+$. Now let $t_0 := 0$, let $t_j := \inf\big\{t \in \mathcal{T} : \sum_{i=1}^n \mathbb{1}_{\{T_i \le t\}} = j\big\}$ be the jth observed or censored event for $j = 1, \ldots, n$, and let $t_{n+1} := t_+$. Then $\{t_j\}$ is a sequence of increasing stopping times. For $j = 0, \ldots, n+1$, define $X_j := W_{t_j}$, as well as the σ-algebra $\mathcal{F}_j = \mathcal{F}_{t_j}$ consisting of those events A for which $A \cap \{t_j \le t\} \in \mathcal{F}_t$ for every $t \in \mathcal{T}$. Then, writing $d_j := X_j - X_{j-1}$, we have by the optional sampling theorem (e.g. Karatzas and Shreve, 1991, Theorem 1.3.22) that $\{d_j : j = 1, \ldots, n+1\}$ is a martingale difference sequence with respect to the filtration $\{\mathcal{F}_j : j = 0, 1, \ldots, n+1\}$.

We seek to control $\mathbb{E}(|d_j|^k|\mathcal{F}_{j-1})$ for $k \in \mathbb{N}$ with $k \ge 2$. Writing $s_j := \min_{\ell:\ell\in R_{t_{j-1}}} \tilde{T}_\ell$ for $j = 1, \ldots, n$ and $s_{n+1} := t_+$, note that $d_{n+1} = 0$ and for

$j = 1, \dots, n$,

$$
\begin{aligned}
|d_j| &= \frac{1}{n}\left|\sum_{i=1}^{n} \int_{(t_{j-1},t_j]} a(s)\, dM_i(s)\right| \\
&= \frac{1}{n}\left|\sum_{i=1}^{n} \int_{(t_{j-1},t_j]} a(s)\, dN_i(s) - \sum_{i=1}^{n} \int_{(t_{j-1},t_j]} a(s)\, d\Lambda_i(s, \boldsymbol{\beta}^o)\right| \\
&\le \frac{4K_Z^2}{n}\left(1 + \sum_{i=1}^{n} \int_{(t_{j-1},t_j]} d\Lambda_i(s, \boldsymbol{\beta}^o)\right) \\
&\le \frac{4K_Z^2}{n}\left(1 + \sum_{i=1}^{n} \int_{(t_{j-1},s_j]} d\Lambda_i(s, \boldsymbol{\beta}^o)\right),
\end{aligned}
$$

where the final inequality follows because for every $j = 1, \dots, n$, if t_j is the time of a censored event, then $s_j > t_j$; if t_j is the time of an observed event, then $s_j = t_j$. Now let $i^* \in \{1, \dots, n\}$ denote the smallest index in $R_{t_{j-1}}$, so that i^* is $\mathcal{F}_{t_{j-1}}$-measurable. Then

$$
\begin{aligned}
\sum_{i=1}^{n} \int_{(t_{j-1},s_j]} d\Lambda_i(s, \boldsymbol{\beta}^o) &= \sum_{i \in R_{t_{j-1}}} \int_{(t_{j-1},s_j]} Y_i(s) e^{\boldsymbol{\beta}^{o\top} \boldsymbol{Z}_i(s)} \lambda_0(s)\, ds \\
&\le e^{2\|\boldsymbol{\beta}^o\|_1 K_Z} (n - j + 1) \int_{(t_{j-1},s_j]} e^{\boldsymbol{\beta}^{o\top} \boldsymbol{Z}_{i^*}(s)} \lambda_0(s)\, ds.
\end{aligned}
$$

But, writing $\mathcal{Z}^{(n)} := \{\boldsymbol{Z}_i(t) : i = 1, \ldots, n, t \in \mathcal{T}\}$, for any $x > 0$,

$$\begin{aligned}
&\mathbb{P}\biggl(\int_{(t_{j-1},s_j]} e^{\boldsymbol{\beta}^{o\top}\boldsymbol{Z}_{i^*}(s)}\lambda_0(s)\,ds > x \biggm| \mathcal{F}_{j-1}, \mathcal{Z}^{(n)}\biggr) \\
&= \mathbb{P}\biggl(-\log \bar{F}_{\tilde{T}_{i^*}|\mathcal{F}_{j-1},\mathcal{Z}^{(n)}}(s_j) > x \biggm| \mathcal{F}_{j-1}, \mathcal{Z}^{(n)}\biggr) \\
&= \mathbb{P}\biggl(\min_{\ell:\ell\in R_{t_{j-1}}} \tilde{T}_\ell > \bar{F}^{-1}_{\tilde{T}_{i^*}|\mathcal{F}_{j-1},\mathcal{Z}^{(n)}}(e^{-x}) \biggm| \mathcal{F}_{j-1}, \mathcal{Z}^{(n)}\biggr) \\
&= e^{-x} \prod_{\ell\in R_{t_{j-1}}\setminus\{i^*\}} \exp\biggl\{-\int_{t_{j-1}}^{\bar{F}^{-1}_{\tilde{T}_{i^*}|\mathcal{F}_{j-1},\mathcal{Z}^{(n)}}(e^{-x})} e^{\boldsymbol{\beta}^{o\top}\boldsymbol{Z}_\ell(s)}\lambda_0(s)\,ds\biggr\} \\
&\leq e^{-x} \prod_{\ell\in R_{t_{j-1}}\setminus\{i^*\}} \exp\biggl\{-e^{-2\|\boldsymbol{\beta}^o\|_1 K_Z}\int_{t_{j-1}}^{\bar{F}^{-1}_{\tilde{T}_{i^*}|\mathcal{F}_{j-1},\mathcal{Z}^{(n)}}(e^{-x})} e^{\boldsymbol{\beta}^{o\top}\boldsymbol{Z}_{i^*}(s)}\lambda_0(s)\,ds\biggr\} \\
&\leq \exp\Bigl\{-(n-j+1)e^{-2\|\boldsymbol{\beta}^o\|_1 K_Z} x\Bigr\}.
\end{aligned}$$

We deduce that

$$\begin{aligned}
&\sum_{i=1}^n \int_{(t_{j-1},s_j]} d\Lambda_i(s,\boldsymbol{\beta}^o) \biggm| \mathcal{F}_{j-1} \\
&\qquad\qquad \leq_{\mathrm{st}} e^{2\|\boldsymbol{\beta}^o\|_1 K_Z}(n-j+1)\mathrm{Exp}\bigl((n-j+1)e^{-2\|\boldsymbol{\beta}^o\|_1 K_Z}\bigr),
\end{aligned}$$

where $\leq_{\mathrm{st}}$ denotes the usual stochastic ordering. In particular,

$$\begin{aligned}
\mathbb{E}(|d_j|^k|\mathcal{F}_{j-1}) &\leq \Bigl(\frac{4K_Z^2}{n}\Bigr)^k \sum_{l=0}^k \binom{k}{l} \mathbb{E}\biggl\{\Bigl(\sum_{i=1}^n \int_{(t_{j-1},s_j]} d\Lambda_i(s,\boldsymbol{\beta}^o)\Bigr)^l \biggm| \mathcal{F}_{j-1}\biggr\} \\
&\leq \Bigl(\frac{4K_Z^2}{n}\Bigr)^k e^{4k\|\boldsymbol{\beta}^o\|_1 K_Z} \sum_{l=0}^k \binom{k}{l} l! = \Bigl(\frac{4K_Z^2}{n}\Bigr)^k e^{4k\|\boldsymbol{\beta}^o\|_1 K_Z} k! e.
\end{aligned}$$

Hence, for all $k \in \mathbb{N}$, we have

$$\mathbb{E}(|d_j|^{2k}|\mathcal{F}_{j-1}) \leq \Bigl(\frac{4e^{4\|\boldsymbol{\beta}^o\|_1 K_Z + 1}K_Z^2}{n}\Bigr)^{2k}(2k)!. \tag{S2.12}$$

From (S2.12), and writing $\nu := 8e^{4\|\boldsymbol{\beta}^o\|_1 K_Z + 1} K_Z^2$, we can apply Boucheron et al. (2013, Theorem 2.3) and the fact that $d_1, \ldots, d_{n+1}$ have zero mean to deduce that each $d_j | \mathcal{F}_{j-1}$ is a sub-gamma random variable with parameters ν^2/n^2 and ν/n. Now let $\mathcal{G} := \sigma(d_1, \ldots, d_{n+1})$. It follows from Section 2 of de la Peña (1999) that for the sequence (d_j), there exists a tangent sequence (e_j) satisfying

$$d_j | \mathcal{F}_{j-1} \overset{d}{=} e_j | \mathcal{F}_{j-1} \overset{d}{=} e_j | \mathcal{G}$$

and such that $e_1, \ldots, e_{n+1}$ are conditionally independent given $\mathcal{G}$. Thus, for $x > 0$,

$$\begin{aligned}
\mathbb{P}\biggl(\sum_{j=1}^{n+1} d_j \geq x\biggr) &\leq \inf_{s>0} e^{-sx} \mathbb{E}\biggl\{\exp\biggl(s \sum_{j=1}^{n} d_j\biggr)\biggr\} \\
&\leq \inf_{s>0} e^{-sx} \biggl\{\mathbb{E}\exp\biggl(2s \sum_{j=1}^{n} e_j\biggr)\biggr\}^{1/2} \\
&= \inf_{s>0} e^{-sx} \biggl[\mathbb{E}\biggl\{\mathbb{E}\exp\biggl(2s \sum_{j=1}^{n} e_j \biggm| \mathcal{G}\biggr)\biggr\}\biggr]^{1/2} \\
&\leq \inf_{0<s<n/\nu} \exp\biggl(-sx + \frac{\nu^2 s^2}{n - 2\nu s}\biggr) \leq \exp\biggl\{-\frac{nx^2}{4(\nu x + \nu^2)}\biggr\},
\end{aligned}$$

where the second inequality follows from Corollary 3.1 in de la Peña (1999), the third inequality follows from the conditional independence of the sequence (e_j) and the sub-gamma tail behaviour, and the last inequality holds by taking

$$s = \frac{nx}{2(\nu x + \nu^2)} < \frac{n}{\nu}.$$

Therefore, for $x > 0$,

$$\mathbb{P}\bigg\{\bigg\|\frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{T}}\sum_{j=1}^{n}\big\{\boldsymbol{Z}_j(s) - \bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o)\big\}^{\otimes 2} \quad w_j(s,\boldsymbol{\beta}^o)\,dM_i(s)\bigg\|_\infty \geq x\bigg\}$$
$$\leq 2p^2\exp\bigg\{-\frac{nx^2}{4(\nu x + \nu^2)}\bigg\}.$$

We deduce that

$$\bigg\|\frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{T}}\sum_{j=1}^{n}\big\{\boldsymbol{Z}_j(s) - \bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o)\big\}^{\otimes 2} w_j(s,\boldsymbol{\beta}^o)\,dM_i(s)\bigg\|_\infty = O_p\bigg(\frac{\log^{1/2}(np)}{n^{1/2}}\bigg). \tag{S2.13}$$

For the second term in (S2.11), observe that

$$\begin{aligned}\bigg\|\frac{1}{n}\sum_{i=1}^{n}\int_0^{t_+}\big\{\boldsymbol{Z}_i(s) - \bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o)\big\}^{\otimes 2}\tilde{w}_i(s,\boldsymbol{\beta}^o)\lambda_0(s)\,ds - \boldsymbol{\Sigma}\bigg\|_\infty \\ \leq \bigg\|\frac{1}{n}\sum_{i=1}^{n}\int_0^{t_+}\big\{\boldsymbol{Z}_i(s) - \boldsymbol{\mu}(s,\boldsymbol{\beta}^o)\big\}^{\otimes 2}\tilde{w}_i(s,\boldsymbol{\beta}^o)\lambda_0(s)\,ds - \boldsymbol{\Sigma}\bigg\|_\infty \\ + \frac{1}{n}\sum_{i=1}^{n}\int_0^{t_+}\big\|\bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o) - \boldsymbol{\mu}(s,\boldsymbol{\beta}^o)\big\|_\infty^2\tilde{w}_i(s,\boldsymbol{\beta}^o)\lambda_0(s)\,ds.\end{aligned} \tag{S2.14}$$

For the first term in (S2.14), we note that it is the maximum absolute value of a random vector, each of whose components is a sample average of independent and identically distributed random variables that are bounded in absolute value by $8K_Z^2$ and have expectation zero. Thus

$$\bigg\|\frac{1}{n}\sum_{i=1}^{n}\int_0^{t_+}\big\{\boldsymbol{Z}_i(s) - \boldsymbol{\mu}(s,\boldsymbol{\beta}^o)\big\}^{\otimes 2}\tilde{w}_i(s,\boldsymbol{\beta}^o)\lambda_0(s)\,ds - \boldsymbol{\Sigma}\bigg\|_\infty = O_p\bigg(\frac{\log^{1/2}(np)}{n^{1/2}}\bigg). \tag{S2.15}$$

For the second term in (S2.14), by (S2.2), (S2.3), (S2.4) and Markov's inequality, we have that for any $\delta > 0$,

$$\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{t_+}\left\|\bar{\boldsymbol{Z}}(s,\boldsymbol{\beta}^o)-\boldsymbol{\mu}(s,\boldsymbol{\beta}^o)\right\|_\infty^2 \tilde{w}_i(s,\boldsymbol{\beta}^o)\lambda_0(s)\,ds = o_p(n^{-(1/2-\delta)}). \tag{S2.16}$$

We deduce from (S2.11), (S2.13), (S2.14), (S2.15) and (S2.16) that for every $\delta > 0$,

$$\|\ddot{\ell}(\boldsymbol{\beta}^o)+\boldsymbol{\Sigma}\|_\infty = o_p(n^{-(1/2-\delta)}). \tag{S2.17}$$

Combining (S2.6) with (S2.5) from the proof of Proposition 2, as well as (S2.10), (S2.17), Lemma 1(ii), **(A4)(b)** and **(A4)(c)**, we have

$$\left|\boldsymbol{c}^\top\big(\widehat{\boldsymbol{\Theta}}\boldsymbol{M}(\widetilde{\boldsymbol{\beta}})+\boldsymbol{I}\big)(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^o)\right| \le \|\widehat{\boldsymbol{\Theta}}\boldsymbol{M}(\widetilde{\boldsymbol{\beta}})+\boldsymbol{I}\|_\infty\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^o\|_1 = o_p(n^{-1/2}),$$

as required. □

S3 Further numerical results

We now include a high-dimensional simulation setting, with $n = 200$ and $p = 500$. The covariates are multivariate Gaussian random vectors with identity covariance matrix, the sparsity parameter $d_o = 3$, the non-zero coefficients are all 1, and the censoring rate is approximately 25%.

We use our method for screening and then apply the standard maximum partial likelihood estimator (MPLE) to obtain confidence intervals for the variables that survive the screening. The corresponding results are collected in Table 1. The debiased estimator and MPLE columns are the results based on our method and an additional refining step, respectively. The results are in the form of mean(standard deviation) over 100 repetitions. The columns EC, Width and p-values are for empirical coverage, interval width and p-values, respectively. The false positive (FP) and false negative (FN) are the corresponding results after the screening.

	Debiased estimators			MPLE				
	EC	Width	p-values	EC	Width	p-values	FP	FN
S	.436(.050)	.355(.006)	0(0)	.945(.033)	.378(0.004)	0(0)	3.131(.104)	0(0)
N	.977(.015)	.267(.003)	.612(.026)	.977(.015)	.010(.006)	.596(.026)		

Table 1: The high-dimensional setting. S and N rows are results for signal and noise variables respectively.

There are different ways to use the outputs of our method to conduct the first step screening. To produce Table 1, we sorted the p-values in the increasing order, namely $p_{(1)} \leq \ldots \leq p_{(p)}$, and included the variables corresponding to the m smallest p-values, with m chosen to satisfy

$$\sum_{j=1}^{m} p_{(j)} \leq 0.05, \quad \text{and} \quad \sum_{j=1}^{m+1} p_{(j)} > 0.05.$$

We then pass on these variables to the second stage, where the MPLE is deployed for refinement.

It can be seen from the false positive (FP) and false negative (FN) columns that the screening can identify the important variables effectively. Comparing the empirical coverage (EC) columns, we see that this two-stage method can significantly improve the empirical coverage of the confidence intervals for the signal variables, since the original debiasing does not completely remove the bias for the signal variables. For further discussion of the bias, we refer the readers to Section 4.1.4.

In the main text, we have assumed a boundedness condition on the covariates, but have generated the covariates from Gaussian distributions in Section 4. Here we include a simulation with uniformly distributed covariates, where our boundedness condition is satisfied. In this setting, we let $(n, p, d_p) = (1000, 10, 1)$, non-zero coefficients be 1, and covariates be generated as independent $\mathrm{Unif}(-2, 2)$ random variables. The rest of the settings and the methods are the same as those in Section 4.1.4. The results are collected in Table 2, which shows the same patterns as what we have seen for the Gaussian cases.

β_j^o	$\hat{\beta}_j - \beta_j^o$	$\hat{b}_j - \beta_j^o$	EC	Wid	p-values	$\hat{b}_j - \beta_j^o$	EC	Width	p-values
		$\widehat{\Theta}$, λ_{CV}				$\widehat{\Theta}$, $0.1\lambda_{\text{CV}}$			
1	-.047(.002)	-.002(.002)	.922(.013)	.143(.000)	.000(.000)	.000(.002)	.927(.013)	.145(.000)	.000(.000)
0	.000(.001)	.000(.001)	.941(.011)	.107(.000)	.519(.013)	.000(.001)	.946(.011)	.108(.000)	.512(.014)
0	.001(.001)	.000(.001)	.939(.012)	.107(.000)	.549(.014)	.000(.001)	.941(.011)	.108(.000)	.542(.014)
0	.000(.001)	.001(.001)	.962(.009)	.107(.000)	.536(.014)	.002(.001)	.958(.010)	.108(.000)	.532(.014)
0	.000(.001)	.000(.001)	.955(.010)	.107(.000)	.532(.014)	.000(.001)	.955(.010)	.108(.000)	.529(.014)
0	.000(.001)	.000(.001)	.946(.011)	.107(.000)	.515(.014)	.000(.001)	.946(.011)	.108(.000)	.510(.014)
0	-.001(.001)	.000(.001)	.948(.011)	.107(.000)	.534(.014)	-.001(.001)	.946(.011)	.108(.000)	.527(.014)
0	-.001(.001)	-.002(.001)	.955(.010)	.107(.000)	.529(.014)	-.002(.001)	.950(.011)	.108(.000)	.526(.014)
0	.000(.001)	-.001(.001)	.934(.012)	.107(.000)	.517(.014)	-.001(.001)	.932(.012)	.108(.000)	.512(.014)
0	.000(.001)	.000(.001)	.950(.011)	.107(.000)	.525(.015)	.000(.001)	.946(.011)	.108(.000)	.521(.015)
		$\widetilde{\Theta}$				$\widehat{\Theta}$, FLARE			
1	-.047(.002)	-.002(.002)	.946(.011)	.158(.000)	.000(.000)	-.007(.002)	.922(.013)	.138(.000)	.000(.000)
0	.000(.001)	.000(.001)	.965(.009)	.118(.000)	.548(.013)	.000(.001)	.939(.012)	.105(.000)	.528(.014)
0	.001(.001)	.000(.001)	.953(.010)	.118(.000)	.577(.014)	.000(.001)	.939(.012)	.105(.000)	.559(.014)
0	.000(.001)	.001(.001)	.972(.008)	.118(.000)	.566(.014)	.002(.001)	.956(.010)	.105(.000)	.545(.014)
0	.000(.001)	.000(.001)	.972(.008)	.118(.000)	.565(.013)	.000(.001)	.958(.010)	.105(.000)	.547(.014)
0	.000(.001)	.000(.001)	.962(.009)	.118(.000)	.546(.014)	.000(.001)	.946(.011)	.105(.000)	.526(.014)
0	-.001(.001)	.000(.001)	.972(.008)	.118(.000)	.560(.014)	-.001(.001)	.941(.012)	.104(.000)	.543(.014)
0	-.001(.001)	-.002(.001)	.974(.008)	.118(.000)	.560(.014)	-.002(.001)	.956(.010)	.105(.000)	.540(.014)
0	.000(.001)	-.001(.001)	.955(.010)	.118(.000)	.546(.014)	-.001(.001)	.932(.013)	.104(.000)	.527(.015)
0	.000(.001)	.000(.001)	.969(.008)	.118(.000)	.556(.014)	.000(.001)	.944(.011)	.105(.000)	.540(.015)
		Merge				MPLE			
1	-.047(.002)	-.004(.002)	.921(.013)	.142(.000)	.000(.000)	.009(.002)	.942(.012)	.159(.000)	.000(.000)
0	.000(.001)	.000(.001)	.943(.011)	.106(.000)	.527(.014)	.000(.002)	.945(.011)	.119(.000)	.491(.014)
0	.001(.001)	.000(.001)	.938(.012)	.106(.000)	.554(.014)	.000(.002)	.942(.012)	.119(.000)	.515(.014)
0	.000(.001)	.002(.001)	.963(.009)	.106(.000)	.544(.014)	.002(.001)	.958(.010)	.120(.000)	.512(.014)
0	.000(.001)	.000(.001)	.958(.010)	.106(.000)	.545(.014)	.000(.001)	.960(.010)	.119(.000)	.510(.014)
0	.000(.001)	.000(.001)	.948(.011)	.106(.000)	.524(.014)	.000(.002)	.950(.011)	.119(.000)	.489(.014)
0	-.001(.001)	-.001(.001)	.953(.010)	.106(.000)	.538(.014)	-.001(.002)	.945(.011)	.119(.000)	.506(.015)
0	-.001(.001)	-.002(.001)	.958(.010)	.106(.000)	.541(.014)	-.002(.001)	.955(.010)	.119(.000)	.509(.014)
0	.000(.001)	-.001(.001)	.931(.013)	.106(.000)	.522(.015)	-.002(.002)	.940(.012)	.120(.000)	.489(.015)
0	.000(.001)	.000(.001)	.943(.011)	.106(.000)	.536(.015)	.000(.002)	.950(.011)	.119(.000)	.502(.015)

Table 2: Uniformly distributed covariate simulation setting.

Bibliography

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

Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

Breslow, N. E. (1972). Contribution to discussion of paper by D. R. Cox. *J. Roy. Statist. Soc., Ser. B*, **34**, 216–217.

Cai, T. T., Liu, W. and Zhou, H. H. (2016). Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *Ann. Statist.*, **44**, 455–88.

de la Peña, V. (1999). A general class of exponential inequalities for martingales and ratios. *Ann. Probab.*, **27**, 537–564.

Gut, A. (2005) *Probability: A Graduate Course*. Springer, New York.

Karatzas, I., and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus (2nd edition)*. Springer-Verlag, New York.

Shorack, G. R. and Wellner, J. A. (1986) *Empirical Processes with Applications to Statistics*. Wiley, New York.