Marginal screening for high-dimensional predictors of survival outcomes

Tzu-Jung Huang, Ian W. McKeague, Min Qian

Department of Biostatistics, Columbia University

Supplementary Materials

For readibility, below we recap all the assumptions which have been listed in the main text.

(A.1) The predictors U_j, $j = 1, \dots, p$, are bounded, and $|\text{Corr}(U_j, U_k)| < 1$ for all $j \neq k$.

(A.2) The error term ε in (7) has zero mean, finite variance, and is uncorrelated with $\boldsymbol{U}$.

(A.3) The censoring time C is independent of $(T, \boldsymbol{U})$ and bounded above by τ (the time to the end of the follow-up).

(A.4) The marginal survival function of the censoring, G_0, is continuous on $\mathcal{T}$, and there exists a positive constant c_g such that $G_0(\tau) > c_g > 0$. Also, the marginal survival function of T, F_0, is continuous on $\mathcal{T}$, and there exists a positive constant c_f such that $F_0(\tau) > c_f > 0$.

S1 First and second moments of $\tilde{\varepsilon}$

Given conditions (A.1)–(A.4), $\tilde{\varepsilon} = \tilde{Y} - \alpha_0 - \boldsymbol{U}^T\boldsymbol{\beta}_0$ has zero mean and finite variance (the square integrability of $\tilde{\varepsilon}$), and is uncorrelated with $\boldsymbol{U}$. The relevant proof goes as follows.

Proof. Because we have the equality $E[\tilde{Y}|\boldsymbol{U}] = E[T|\boldsymbol{U}]$, it ensures that

$$E[\tilde{\varepsilon}|\boldsymbol{U}] = E[\tilde{Y}|\boldsymbol{U}] - \alpha_0 - \boldsymbol{U}^T\boldsymbol{\beta}_0 = E[T|\boldsymbol{U}] - \alpha_0 - \boldsymbol{U}^T\boldsymbol{\beta}_0 = 0, \tag{S1.1}$$

which indicates the zero mean of $\tilde{\varepsilon}$ through taking expectation on both sides of (S1.1). Recall that (A.4) implies that $G_0(t)$ is bounded away from zero for all $t \in \mathcal{T} = (-\infty, \tau]$, where τ denotes the end of the follow-up. We show the boundedness of $E[\tilde{\varepsilon}^2|\boldsymbol{U}]$ as follows, which implies the finite variance of $\tilde{\varepsilon}$. It is straightforward to have that

$$\begin{aligned}
E[\tilde{\varepsilon}^2|\boldsymbol{U}] &= E[(\tilde{Y} - \alpha_0 - \boldsymbol{U}^T\boldsymbol{\beta}_0)^2|\boldsymbol{U}] = E[\tilde{Y}^2|\boldsymbol{U}] - (\alpha_0 + \boldsymbol{U}^T\boldsymbol{\beta}_0)^2 \\
&= E[T^2 G_0(T)^{-1}|\boldsymbol{U}] - (\alpha_0 + \boldsymbol{U}^T\boldsymbol{\beta}_0)^2 \\
&\leq G_0(\tau-)^{-1}E[\varepsilon^2|\boldsymbol{U}] + (G_0(\tau-)^{-1} - 1)(\alpha_0 + \boldsymbol{U}^T\boldsymbol{\beta}_0)^2.
\end{aligned} \tag{S1.2}$$

Taking expectation on both sides of (S1.2) yields that

$$E[\tilde{\varepsilon}^2] \leq G_0(\tau-)^{-1}E[\varepsilon^2] + (G_0(\tau-)^{-1} - 1)E[(\alpha_0 + \boldsymbol{U}^T\boldsymbol{\beta}_0)^2].$$

Because (A.1) and (A.2) give the finite second moment of ε and predictors U_j, $j = 1, \ldots, p$, and (A.4) yields non-zero $G_0(\tau)$, it is easy to see that $E[\tilde{\varepsilon}^2]$ is bounded above by a finite constant. Hence, we show that $\tilde{\varepsilon}$ has a finite variance. Note that (A.2) gives $\text{Cov}(U_j, \varepsilon) = 0$ for all j. Since we have that $\text{Cov}(U_j, \tilde{Y}) = \text{Cov}(U_j, T)$, it is easy to see that $\text{Cov}(U_j, \tilde{\varepsilon}) = \text{Cov}(U_j, \tilde{Y}) - \text{Cov}(U_j, \alpha_0 + \boldsymbol{U}^T\boldsymbol{\beta}_0) = \text{Cov}(U_j, T) - \text{Cov}(U_j, \alpha_0 + \boldsymbol{U}^T\boldsymbol{\beta}_0) = \text{Cov}(U_j, \varepsilon) = 0$, for all j. □

S2 Pollard's Functional Central Limit Theorem

We state Pollard's functional central limit theorem below for readers' convenience. Consider random processes developed from a triangular array $\{\tilde{f}_{ni}(t), i = 1, \ldots, m_n, t \in \mathcal{T}, n \in \mathbb{N}\}$, with the $\{\tilde{f}_{ni}\}$ independent within each row and $\mathcal{T}$ being the index set. In addition, we can define

$$\mathbb{W}_n(t) = \sum_{i \le m_n} (\tilde{f}_{ni}(t) - E\tilde{f}_{ni}(t))\,;\ \rho_n(s,t) = (\sum_{i \le m_n} E|\tilde{f}_{ni}(s) - \tilde{f}_{ni}(t)|^2)^{1/2}.$$

Let $UC(\mathcal{T}, \rho)$ denote the space of all bounded functions $\tilde{f} : \mathcal{T} \to \mathbb{R}$ which are uniformly ρ-continuous, that is, with any appropriately selected semimetric ρ,

$$\lim_{\tilde{\delta} \downarrow 0} \sup_{\rho(s,t) < \tilde{\delta}} |\tilde{f}(s) - \tilde{f}(t)| = 0.$$

Moreover, $\mathcal{T}$ is totally bounded by ρ (equivalently, $(\mathcal{T}, \rho)$ is totally bounded) if for every $\epsilon > 0$, there exists a finite collection $\mathcal{T}_m = \{t_1, \ldots, t_m\} \subset \mathcal{T}$ such that for all $t \in \mathcal{T}$, we have $\rho(s,t) \le \epsilon$ for some $s \in \mathcal{T}_m$. We would like to indicate that if the weak limit is a Gaussian process W, then the semimetric ρ can be selected as $\rho(s,t) = (E|W(s) - W(t)|^2)^{1/2}$.

Theorem (Pollard, (1990))**.** *Suppose the processes from the triangular array $\{\tilde{f}_{ni}(t)\}$ are independent within rows, and satisfies*

(A) the $\{\tilde{f}_{ni}\}$ are manageable, with envelopes $\{\tilde{F}_{ni}\}$ which are also independent within rows;

(B) $V(s,t) = \lim_{n\to\infty} E\mathbb{W}_n(s)\mathbb{W}_n(t)$ exists for every s, $t \in \mathcal{T}$;

(C) $\limsup_{n\to\infty} \sum_{i=1}^{m_n} E\tilde{F}_{ni}^2$ is finite;

(D) For each $\eta > 0$, $\lim_{n\to\infty} \sum_{i=1}^{m_n} E\tilde{F}_{ni}^2 1(\tilde{F}_{ni} > \eta) = 0$ (an analogy to the Lindeberg condition);

(E) For every s, $t \in \mathcal{T}$, $\rho(s,t) = \lim_{n\to\infty} \rho_n(s,t)$ exists. And for all deterministic sequences $\{s_n\}$ and $\{t_n\}$ in $\mathcal{T}$, $\rho_n(s_n, t_n) \to 0$ if $\rho(s_n, t_n) \to 0$.

Then, we have (i) $\mathcal{T}$ is totally bounded under the pseudometric (semimetric) ρ; (ii) the finite dimensional distributions of $\mathbb{W}_n$ have Gaussian limits, with zero means and covariances given by

V, which uniquely determine a Gaussian distribution concentrated on $UC(\mathcal{T},\rho)$; (iii) $\mathbb{W}_n$ converge weakly on ℓ^∞_τ to a tight mean zero Gaussian process W concentrated on $UC(\mathcal{T},\rho)$ with $V(s,t)$ as covariance.

S3 Proof for Theorem 1

Theorem 1 follows from a series of lemmas below. To keep notational simplicity, we suppress the superscript "(n)" under the local model in this proof unless otherwise stated. Define the sample space by $\mathcal{X}=\mathcal{T}\times\{0,1\}\times\mathbb{R}^p$ (with σ-algebra $\mathcal{A}$), where we observe a random sample $\{X_i,\delta_i,\boldsymbol{U}_i\}_{i=1}^n$. In addition, $(X,\delta,\boldsymbol{U})$ follows a distribution P belonging to the set of probability measure $\mathcal{P}$ on $(\mathcal{X},\mathcal{A})$, and its empirical distribution is denoted by $\mathbb{P}_n$. In the following, we outline how all the lemmas play their roles to prove Theorem 1 and thereafter list these lemmas along with their corresponding proofs.

In Lemma 3.1, we take advantage of Stute's Theorem 1.1 (Stute (1995)), and express $\sqrt{n}[\hat{G}_n(t)-G_0(t)]$ as an i.i.d. sum, for any fixed t. Let $\mathbb{G}_n=\sqrt{n}(\mathbb{P}_n-P)$, and $\{\sqrt{n}[\hat{G}_n(t)-G_0(t)],t\in\mathcal{T}\}$ can be approximated by an empirical process (with probability approaching to one) $\mathbb{L}_n\colon\mathcal{X}\mapsto\ell^\infty_\tau$, which is

$$\{\mathbb{G}_n[\phi_t(X)\gamma_0(X)(1-\delta)+\gamma_1(X,t)\delta-\gamma_2(X,t)-G_0(t)],t\in\mathcal{T}\}$$

with ϕ_t, γ_0, γ_1 as well as γ_2 stated in Lemma 3.1. Moreover, we define a function $\Psi_j:\mathbb{R}\times\ell^\infty_\tau\times\mathcal{P}\to\mathbb{R}$, where

$$\Psi_j(m,h,Q)=m+Q\left[\frac{(U_j-EU_j)\tilde{Y}h(X)}{G_0(X)}\right], \tag{S3.1}$$

and $\widetilde{\mathbb{M}}_n=\{\widetilde{\mathbb{M}}_{n,j},j=1,\ldots,p\}$ with

$$\widetilde{\mathbb{M}}_{n,j}=\mathbb{G}_n\left(\tilde{\varepsilon}_n+(\boldsymbol{U}-\mathbb{P}_n\boldsymbol{U})^T\boldsymbol{\beta}_0-(U_j-\mathbb{P}_nU_j)C_j^T\boldsymbol{\beta}_0/V_j\right)(U_j-\mathbb{P}_nU_j),\ j=1,\ldots,p. \tag{S3.2}$$

We further introduce Lemma 3.2 to indicate

$$\sqrt{n}(\hat{\theta}_n-\theta_n)S^2_{\hat{j}_n}=\Psi_{j_0}(\widetilde{\mathbb{M}}_{n,j_0},\mathbb{L}_n,\mathbb{P}_n).$$

To attain the limiting distribution of $\sqrt{n}\hat{\theta}_n$ when $\boldsymbol{\beta}_0\neq\mathbf{0}$, we need to derive the joint weak limit of $(\widetilde{\mathbb{M}}_n,\mathbb{L}_n)$. It suffices to show that the empirical process $\mathbb{W}_n=\{\mathbb{W}_n(t)=\mathbb{L}_n(t)+\sum_{j=1}^p a_j\widetilde{\mathbb{M}}_{n,j},t\in\mathcal{T}\}$ converges weakly to a mean-zero Gaussian process $\mathbb{W}$ with covariance function σ_W, where for

$(a_1, \ldots, a_p) \in \mathbb{R}^p$,

$$\begin{aligned}\mathbb{W}_n(t) &= \mathbb{L}_n(t) + \sum_{j=1}^{p} a_j \widetilde{\mathbb{M}}_{n,j} \\ &= \mathbb{G}_n\{\phi_t(X)\gamma_0(X)(1-\delta) + \gamma_1(X,t)\delta - \gamma_2(X,t) - G_0(t) \\ &\quad + \sum_{j=1}^{p} a_j(\tilde{\varepsilon}_n + (\boldsymbol{U} - \mathbb{P}_n\boldsymbol{U})^T\boldsymbol{\beta}_0 - (U_j - \mathbb{P}_nU_j)C_j^T\boldsymbol{\beta}_0/V_j)(U_j - \mathbb{P}_nU_j)\}\end{aligned} \tag{S3.3}$$

and

$$\sigma_W(s,t) = \sum_{j=1}^{p}\sum_{k=1}^{p} a_j a_k \sigma_M(j,k) + \sum_{j=1}^{p} a_j \sigma_{ML}(j,s) + \sum_{j=1}^{p} a_j \sigma_{ML}(j,t) + \sigma_L(s,t).$$

In Lemma 3.3, we obtain this desired result by checking some regularity conditions for Pollard's functional central limit theorem stated in Section S2. This result equivalently ensures the joint weak convergence of $(\widetilde{\mathbb{M}}_n, \mathbb{L}_n)$ to $(\boldsymbol{M}, \mathbb{L})$, where $(\boldsymbol{M}, \mathbb{L})$ is a mean-zero joint Gaussian process. Furthermore, multivariate central limit theorem implies that $\widetilde{\mathbb{M}}_n$ converges in distribution to a normal random vector $\boldsymbol{M}$, and the weak convergence of $\mathbb{L}_n$ to $\mathbb{L}$ can be developed by applying Pollard's functional central limit theorem again in a similar fashion as in Lemma 3.3.

In ensuing Lemma 3.4, we prove the continuity of Ψ_j on $\mathbb{R} \times \ell_\tau^\infty \times \mathcal{P}$ almost surely (a.s.) for all given j, and validate the application of continuous mapping theorem for empirical processes (Kosorok (2008), Chap. 7, Sec. 7.2.1). Based on Lemma 3.3 and 3.4, we develop the limiting distribution of $\sqrt{n}\hat{\theta}_n$ when $\boldsymbol{\beta}_0 \neq \mathbf{0}$ in Lemma 3.5. Moreover, we show the oracle property of $\hat{j}_n$ when $\boldsymbol{\beta}_0 \neq \mathbf{0}$ in Lemma 3.6. When $\boldsymbol{\beta}_0 = \mathbf{0}$, the joint limiting distribution of $\sqrt{n}\widehat{\boldsymbol{\theta}}$ and $n[S_Y^2\mathbf{1}_p - \widehat{\boldsymbol{R}}]$ is constructed in Lemma 3.7. To establish the limiting distribution when $\boldsymbol{\beta}_0 = \mathbf{0}$, we use similar arguments to those in MQ's work (McKeague and Qian (2015)) and state one of their crucial lemmas as Lemma 3.8.

Let $\boldsymbol{Z} = (Z_1, \ldots, Z_p)^T$ in which the j-th component $Z_j = M_j + \varphi_j(\mathbb{L})$, and $\boldsymbol{Z}$ can be regarded as a function from $\mathbb{R}^{2p}$ to $\mathbb{R}^p$. Let $f(\boldsymbol{z}, \boldsymbol{b})_j = (z_j + C_j^T\boldsymbol{b})^2/V_j$, for all j. Since $\boldsymbol{Z}$ is a random vector and $|\mathrm{Corr}(U_j, U_k)| < 1$ for $j \neq k$, it is indicated that $f(\boldsymbol{Z}, \boldsymbol{b}_0)_j \neq f(\boldsymbol{Z}, \boldsymbol{b}_0)_k$ for any $j \neq k$, a.s. Using Lemma 3.8, it further ensures that $J = \arg\max_{j=1,\ldots,p} f(\boldsymbol{Z}, \boldsymbol{b}_0)_j$ is uniquely determined a.s. For a p-variate real vector $\boldsymbol{t} = (t_1, \ldots, t_p)^T$, define

$$\boldsymbol{h}(\boldsymbol{t}) = (1(\arg\max_j t_j = 1), \ldots, 1(\arg\max_j t_j = p)).$$

We then show that $\sqrt{n}(\hat{\theta}_n - \theta_n)$ is a continuous function of $\sqrt{n}\widehat{\boldsymbol{\theta}}\boldsymbol{h}(n[S_Y^2\mathbf{1}_p - \widehat{\boldsymbol{R}}])$. Note that $\hat{j}_n$ is a unique maximizer to $n[S_Y^2\mathbf{1}_p - \widehat{\boldsymbol{R}}]$. Since both $\hat{j}_n$ and J are uniquely determined and $\boldsymbol{h}$ is continuous at $\boldsymbol{t}$ when $\arg\max_j t_j$ is unique, then in Lemma 3.9 we develop the desired limiting distribution of

$\sqrt{n}\hat{\theta}_n$, applying continuous mapping theorem on the joint distribution of $\sqrt{n}\widehat{\boldsymbol{\theta}}$ and $n[S_Y^2\mathbf{1}_p - \widehat{\boldsymbol{R}}]$ obtained from Lemma 3.7. According to all the lemmas and results thereof (stated below in order), we finally complete this proof and show the desired theorem.

Lemma 3.1. *Suppose that (A.4) holds. Uniformly over $t \le \tau$,*

$$\sqrt{n}[\hat{G}_n(t) - G_0(t)] = -\mathbb{L}_n(t) + o_p(1).$$

Proof. For all $x \in \mathbb{R}$ and $t \le \tau$, let

$$\begin{aligned}
\tilde{H}^0(x) &= P(X \le x, \delta = 0) = -\int_{-\infty}^{x} F_0(y)G_0(dy);\\
\tilde{H}^1(x) &= P(X \le x, \delta = 1) = -\int_{-\infty}^{x} G_0(y)F_0(dy);\\
\gamma_0(x) &= \exp\{\int_{-\infty}^{x} \frac{\tilde{H}^1(dy)}{H_0(y)}\};\\
\gamma_1(x,t) &= \frac{1}{H_0(x)}\int 1_{(x,\infty)}(w)\phi_t(w)\gamma_0(w)\tilde{H}^0(dw);\\
\gamma_2(x,t) &= \int\int \frac{1_{(-\infty,x)}(v)1_{(-\infty,w)}(v)\phi_t(w)\gamma_0(w)}{H_0(v)^2}\tilde{H}^1(dv)\tilde{H}^0(dw),
\end{aligned}$$

where $\phi_t(\cdot) \equiv 1_{(-\infty,t]}(\cdot)$. We will apply Theorem 1.1 of Stute (1995), for which we need to check two conditions below:

$$\begin{aligned}
&\int \phi_t^2(x)\gamma_0^2(x)\tilde{H}^0(dx) = \int [\phi_t(X)\gamma_0(X)(1-\delta)]^2 dP_n < \infty;\\
&-\int |\phi_t(x)|\Gamma^{1/2}(x)G_0(dx) < \infty, \text{ where } \Gamma(x) = \int_{-\infty}^{x} \frac{-F_0(dy)}{H_0(y)F_0(y)}.
\end{aligned}$$

Note that $\gamma_0(x) = F_0(x)^{-1}$ and the value of $1-\delta$ is either zero or one. The first condition then follows since

$$\int [\phi_t(X)\gamma_0(X)(1-\delta)]^2 dP_n \le \int_{-\infty}^{t} F_0(X)^{-2} dP_n < \frac{1}{F_0(\tau)^2} < \infty,$$

where we have $F_0(\tau)^{-2}$ finite in the above display because (A.4) indicates that $F_0(\tau) > 0$. Moreover for all $u \le \tau$, we have that

$$\Gamma(u) \le \frac{-1}{H_0(\tau)}\left[\int_{-\infty}^{u} \frac{F_0(dy)}{F_0(y)}\right] = \frac{1}{H_0(\tau)F_0(\tau)}\left(1 - F_0(u)\right),$$

which further implies that for all $t \leq \tau < \tau_H$,

$$-\int |\phi_t(x)|\Gamma^{1/2}(x)G_0(dx) = -\int_{-\infty}^{t} \Gamma^{1/2}(x)G_0(dx) \leq -\int_{-\infty}^{t} \Gamma^{1/2}(\tau)G_0(dx)$$
$$\leq \frac{(1-G_0(t))\sqrt{(1-F_0(\tau))}}{\sqrt{F_0(\tau)H_0(\tau)}} < \infty.$$

The second condition is satisfied, and the result then follows from Stute's theorem. □

Lemma 3.2. *Suppose that (A.1)–(A.4) hold; that $\boldsymbol{\beta}_0 \neq \mathbf{0}$, and that j_0 is unique when $\boldsymbol{\beta}_0 \neq \mathbf{0}$.*

$$\sqrt{n}(\hat{\theta}_n - \theta_n)S^2_{\hat{j}_n} = \Psi_{j_0}(\widetilde{\mathbb{M}}_{n,j_0}, \mathbb{L}_n, \mathbb{P}_n) + o_p(1),$$

where $\widetilde{\mathbb{M}}_{n,j}$, $\mathbb{L}_n$, and Ψ_j are as previously defined.

Proof. Because (A.3) implies the property that $\text{Cov}(U_j, T) = \text{Cov}(U_j, \tilde{Y})$ for all j, we can have that

$$\begin{aligned}\sqrt{n}(\hat{\theta}_n - \theta_n)S^2_{\hat{j}_n} &= \sqrt{n}\left(\frac{1}{S^2_{\hat{j}_n}}\mathbb{P}_n(U_{\hat{j}_n} - \mathbb{P}_nU_{\hat{j}_n})Y - \frac{\text{Cov}(U_{j_n}, T)}{V_{j_n}}\right)S^2_{\hat{j}_n} \\ &= \sqrt{n}\left(\mathbb{P}_n(U_{\hat{j}_n} - \mathbb{P}_nU_{\hat{j}_n})Y - \frac{\text{Cov}(U_{j_n}, \tilde{Y})}{V_{j_n}}S^2_{\hat{j}_n}\right).\end{aligned} \tag{S3.4}$$

Meanwhile, we can further observe that

$$\begin{aligned}\sqrt{n}\mathbb{P}_n(U_{\hat{j}_n} - \mathbb{P}_nU_{\hat{j}_n})Y &= \sqrt{n}\mathbb{P}_n(U_{\hat{j}_n} - \mathbb{P}_nU_{\hat{j}_n})\tilde{Y} + \sqrt{n}\mathbb{P}_n(U_{\hat{j}_n} - \mathbb{P}_nU_{\hat{j}_n})(Y - \tilde{Y}) \\ &= \sqrt{n}\mathbb{P}_n(U_{\hat{j}_n} - \mathbb{P}_nU_{\hat{j}_n})\tilde{Y} + \sqrt{n}\mathbb{P}_n(U_{\hat{j}_n} - \mathbb{P}_nU_{\hat{j}_n})\delta X\left[\frac{1}{\hat{G}_n(X)} - \frac{1}{G_0(X)}\right] \\ &= \sqrt{n}\mathbb{P}_n(U_{j_0} - \mathbb{P}_nU_{j_0})\tilde{Y} + \mathbb{P}_n\frac{(U_{j_0} - \mathbb{P}_nU_{j_0})\delta X\mathbb{L}_n(X)}{G_0(X)^2} + o_p(1),\end{aligned} \tag{S3.5}$$

where the second term is contributed by the effect of estimating G_0 by the Kaplan–Meier estimator $\hat{G}_n$. The last equality in (S3.5) can be ensured by $\hat{j}_n \overset{a.s.}{\to} j_0$ (shown in Lemma 3.6 implied by (A.1)); the first order Taylor expansion around G_0, and Lemma 3.1 (where (A.4) is required).

Recall that $\tilde{Y} = \alpha_0 + \boldsymbol{U}^T\boldsymbol{\beta}_n + \tilde{\varepsilon}_n$, and then (S3.4)–(S3.5) further imply that

$$\begin{aligned}\sqrt{n}(\hat{\theta}_n - \theta_n)S^2_{\hat{j}_n} &= \sqrt{n}\mathbb{P}_n\left(\tilde{\varepsilon}_n + (\boldsymbol{U} - \mathbb{P}_n\boldsymbol{U})^T\boldsymbol{\beta}_n - (U_{j_0} - \mathbb{P}_nU_{j_0})C^T_{j_0}\boldsymbol{\beta}_n/V_{j_0}\right)(U_{j_0} - \mathbb{P}_nU_{j_0}) \\ &\quad + \mathbb{P}_n\frac{(U_{j_0} - \mathbb{P}_nU_{j_0})\delta X\mathbb{L}_n(X)}{G_0(X)^2} + o_p(1).\end{aligned} \tag{S3.6}$$

Because (A.2) implies that $\tilde{\varepsilon}_n$ and $\boldsymbol{U}$ are uncorrelated, it ensures that for any j,

$$P\left(\tilde{\varepsilon}_n + (\boldsymbol{U} - \mathbb{P}_n \boldsymbol{U})^T \boldsymbol{\beta}_n - (U_j - \mathbb{P}_n U_j) C_j^T \boldsymbol{\beta}_n / V_j\right)(U_j - \mathbb{P}_n U_j) = 0.$$

Since $\tilde{Y} = \delta X / G_0(X)$, $\mathbb{P}_n U_{j_0} \stackrel{a.s.}{\to} P U_{j_0}$ along with (S3.6), we further have that

$$\begin{aligned}\sqrt{n}(\hat{\theta}_n - \theta_n) S^2_{\hat{j}_n} &= \mathbb{G}_n\left(\tilde{\varepsilon}_n + (\boldsymbol{U} - \mathbb{P}_n \boldsymbol{U})^T \boldsymbol{\beta}_n - (U_{j_0} - \mathbb{P}_n U_{j_0}) C_{j_0}^T \boldsymbol{\beta}_n / V_{j_0}\right)(U_{j_0} - \mathbb{P}_n U_{j_0}) \\ &\quad + \mathbb{P}_n \frac{(U_{j_0} - P U_{j_0}) \tilde{Y} \mathbb{L}_n(X)}{G_0(X)} + o_p(1).\end{aligned} \tag{S3.7}$$

By $\boldsymbol{\beta}_n \to \boldsymbol{\beta}_0$ along with the definitions of $\widetilde{\mathbb{M}}_{n,j}$ and Ψ_j for any fixed j, (S3.7) further gives the desired result. □

Lemma 3.3. *Suppose that (A.1)–(A.4) hold. The empirical process $\mathbb{W}_n$ converges to a mean-zero Gaussian process $\mathbb{W}$ with covariance function σ_W, where for $(a_1, \ldots, a_p) \in \mathbb{R}^p$,*

$$\sigma_W(s,t) = \sum_{j=1}^p \sum_{k=1}^p a_j a_k \sigma_M(j,k) + \sum_{j=1}^p a_j \sigma_{ML}(j,s) + \sum_{j=1}^p a_j \sigma_{ML}(j,t) + \sigma_L(s,t)$$

with $\sigma_M(j,k)$, $\sigma_{ML}(j,t)$ and $\sigma_L(s,t)$ provided in the proof, for any j,k,s,t.

Proof. Recall that $\mathbb{W}_n = \{\mathbb{W}_n(t), t \in \mathcal{T}\}$, where

$$\begin{aligned}\mathbb{W}_n(t) &= \mathbb{L}_n(t) + \sum_{j=1}^p a_j \widetilde{\mathbb{M}}_{n,j} \\ &= \mathbb{G}_n[\,\phi_t(X)\gamma_0(X)(1-\delta) + \gamma_1(X,t)\delta - \gamma_2(X,t) - G_0(t) \\ &\qquad + \sum_{j=1}^p a_j(\tilde{\varepsilon}_n + (\boldsymbol{U} - E\boldsymbol{U})^T \boldsymbol{\beta}_0 - (U_j - EU_j) C_j^T \boldsymbol{\beta}_0 / V_j)(U_j - EU_j)] + o_p(1).\end{aligned}$$

Let $\boldsymbol{U}_{ij}$ denote the i-th subject's observation of U_j. The empirical process $\mathbb{W}_n$ can be approximated by triangular array:

$$\{h_{ni}(t) = \sum_{j=1}^p a_j f_{ni,j} + g_{ni}(t), i = 1, \ldots, n, t \in \mathcal{T}\},$$

where

$$f_{ni,j} = \frac{1}{\sqrt{n}}\left(\tilde{\varepsilon}_{ni} + (\boldsymbol{U}_i - E\boldsymbol{U})^T \boldsymbol{\beta}_0 - (\boldsymbol{U}_{ij} - EU_j)\frac{C_j^T \boldsymbol{\beta}_0}{V_j}\right)(\boldsymbol{U}_{ij} - EU_j)$$

and

$$g_{ni}(t) = \frac{1}{\sqrt{n}}[\phi_t(X_i)\gamma_0(X_i)(1-\delta_i) + \gamma_1(X_i,t)\delta_i - \gamma_2(X_i,t) - G_0(t)].$$

It is easy to see that $Ef_{ni,j} = 0$ for all i, j, and $Eg_{ni}(t) = 0$ for any $t \in \mathcal{T}$ (Stute (1995)). It implies that we can directly formulate $\mathbb{W}_n(t) = \sum_{i=1}^n h_{ni}(t) + o_p(1)$, and $\mathbb{L}_n(t) = \sum_{i=1}^n g_{ni}(t)$, respectively. Below, we check required conditions and apply Pollard's functional central limit theorem to establish the weak convergence of $\mathbb{W}_n$.

Condition (A)

We start with verifying the manageability of triangular array. Let

$$\tilde{\mathcal{H}}_n = \{(h_{n1}(t), h_{n2}(t), \ldots, h_{nn}(t)) \in \mathbb{R}^n, t \in \mathcal{T}\}$$

whose envelope function is $\tilde{\boldsymbol{H}}_n = (H_{n1}, H_{n2}, \ldots, H_{nn}) \in \mathbb{R}^n$. For each i,

$$H_{ni} = \sum_{j=1}^p a_j F_{ni,j} + G_{ni} \;\text{ with }\; F_{ni,j} = |f_{ni,j}| \;\text{ and }\; G_{ni} = \sup_{t\in\mathcal{T}} |g_{ni}(t)|. \tag{S3.8}$$

Let $\odot$ denote the operation of point-wise vector product. For any non-negative vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)^T \in \mathbb{R}^n$, we can create a class

$$\boldsymbol{\xi} \odot \tilde{\mathcal{H}}_n = \{(\xi_1 h_{n1}(t), \xi_2 h_{n2}(t), \ldots, \xi_n h_{nn}(t)) \in \mathbb{R}^n, t \in \mathcal{T}\}.$$

Let $\|\cdot\|$ denote L_2 norm, and $\|\cdot\|_{Q,2}$ denote $L_2(Q)$-norm, which is the norm of the class of square-integrable functions under a finitely discrete probability measure Q. Let $D(q, \mathcal{K})$ denote the packing number of class $\mathcal{K}$ (the maximal number of points that can fit in $\mathcal{K}$ while maintaining a distance greater than q (measured by a pre-specified norm) between all points). Our triangular array of processes $\{h_{ni}(t), i = 1, \ldots, n, t \in \mathcal{T}\}$ is manageable (with respect to the envelopes $\tilde{\boldsymbol{H}}_n$) if we can find a deterministic function λ *(capacity bound)* such that

(1) $\int_0^1 \sqrt{\log \lambda(x)}dx < \infty$.

(2) $D(\zeta\|\boldsymbol{\xi} \odot \tilde{\boldsymbol{H}}_n\|, \boldsymbol{\xi} \odot \tilde{\mathcal{H}}_n) \le \lambda(\zeta)$ for $0 < \zeta \le 1$, $\boldsymbol{\xi} \in \mathbb{R}^n$ of non-negative weights, all $n \ge 1$.

Let $\boldsymbol{u}_j$ be the j-th element of $\boldsymbol{u} \in \mathbb{R}^p$. We define functions $f_{n,j}\colon \mathcal{X} \to \mathbb{R}$ and $g_{n,t}\colon \mathcal{T} \times \{0,1\} \to \mathbb{R}$, where

$$f_{n,j}(x,d,\boldsymbol{u}) = \frac{1}{\sqrt{n}}\left[\left(\left(\frac{dx}{G_0(x)} - \alpha_0 - \boldsymbol{u}^T\boldsymbol{\beta}_0\right) + (\boldsymbol{u} - E\boldsymbol{U})^T\boldsymbol{\beta}_0 - (\boldsymbol{u}_j - EU_j)\frac{C_j^T\boldsymbol{\beta}_0}{V_j}\right)(\boldsymbol{u}_j - EU_j)\right],$$

and

$$g_{n,t}(x,d) = \frac{1}{\sqrt{n}}[\phi_t(x)\gamma_0(x)(1-d) + \gamma_1(x,t)d - \gamma_2(x,t) - G_0(t)].$$

We create another function class (changing with sample size n) $\mathcal{H}_n = \{h_{n,t}, t \in \mathcal{T}\}$, where the t-indexed function $h_{n,t}\colon \mathcal{X} \to \mathbb{R}$ is defined by

$$h_{n,t}(x,d,\boldsymbol{u}) = \sum_{j=1}^{p} a_j f_{n,j}(x,d,\boldsymbol{u}) + g_{n,t}(x,d) \text{ such that } h_{n,t}(X_i,\delta_i,\boldsymbol{U}_i) = h_{ni}(t).$$

Moreover, its envelope function $H_n : \mathcal{X} \to \mathbb{R}$, where $H_n(X_i,\delta_i,\boldsymbol{U}_i) = \sqrt{n}H_{ni}$. For any $\boldsymbol{\xi} \in \mathbb{R}^n$, it is easy to see that $\mathcal{H}_n \supseteq \boldsymbol{\xi} \odot \tilde{\mathcal{H}}_n$.

Let $N(q,\mathcal{K})$ denote the covering number of class $\mathcal{K}$ (the minimal number of closed balls of radius q (measured by a pre-specified norm) required to cover any class $\mathcal{K}$). Condition (1) for manageability could be fulfilled if we let

$$\lambda(x) = \lim_{n\to\infty} \sup \sup_{Q} N(x\|H_n\|_{Q,2}/2, \mathcal{H}_n),$$

and if the class $\mathcal{H}_n$ satisfies the bounded uniform entropy integral (BUEI) condition

$$\lim_{n\to\infty} \sup \sup_{Q} \int_0^1 \sqrt{\log N(x\|H_n\|_{Q,2}/2, \mathcal{H}_n)}dx < \infty, \tag{S3.9}$$

where $\sup_Q$ means that the supremum is taken over all finitely discrete probability measures. To verify the BUEI condition in (S3.9), it suffices to show $\mathcal{H}_n$ is a BUEI class, for all $n \geq 1$. Let $h^{\star}_{n,t}\colon \mathcal{X} \to \mathbb{R}$ and $h^{\divideontimes}_{n,t}\colon \mathcal{T} \times \{0,1\} \to \mathbb{R}$, where

$$h^{\star}_{n,t}(x,d,\boldsymbol{u}) = \sum_{j=1}^{p} a_j f_{n,j}(x,d,\boldsymbol{u}) + \frac{1}{\sqrt{n}}[\phi_t(x)\gamma_0(x)(1-d) + \gamma_1(x,t)d]$$

and

$$h^{\divideontimes}_{n,t}(x,d) = \frac{-1}{\sqrt{n}}[\gamma_2(x,t) + G_0(t)],$$

such that we can further decompose $h_{n,t} = h^{\star}_{n,t} + h^{\divideontimes}_{n,t}$. Let $\mathcal{H}^{\star}_n = \{h^{\star}_{n,t}, t \in \mathcal{T}\}$ and $\mathcal{H}^{\divideontimes}_n = \{h^{\divideontimes}_{n,t}, t \in \mathcal{T}\}$. We can easily see for all $n \geq 1$, $\mathcal{H}^{\star}_n$ and $\mathcal{H}^{\divideontimes}_n$ are both VC classes because (1) the collection $\{(-\infty, t], t \in \mathcal{T}\}$ is a VC class (VC index=2), and (2) both $h^{\star}_{n,t}$ and $h^{\divideontimes}_{n,t}$ are monotone in t. Since VC class belongs to BUEI class, then $\mathcal{H}^{\star}_n$ and $\mathcal{H}^{\divideontimes}_n$ are both BUEI classes. The preservation property of BUEI class implies $\mathcal{H}_n$ is a BUEI class (Kosorok (2008)), such that for all $n \geq 1$,

$$\sup_Q \int_0^1 \sqrt{\log N(x\|H_n\|_{Q,2}/2, \mathcal{H}_n)}dx < \infty.$$

Hence, the BUEI condition in (S3.9) holds for $\mathcal{H}_n$. Subsequently, we verify Condition (2) for manageability as follows. For any $\boldsymbol{\xi} \in \mathbb{R}^n$, let $\|\cdot\|_{Q_{\boldsymbol{\xi}},2}$ denote $L_2(Q_{\boldsymbol{\xi}})$-norm, where $Q_{\boldsymbol{\xi}}$ is a finitely discrete probability measure:

$$Q_{\boldsymbol{\xi}} = (n\|\boldsymbol{\xi}\|)^{-1} \sum_{i=1}^n \xi_i^2 1(X_i, \delta_i, \boldsymbol{U}_i).$$

Thus for $0 < \zeta \leq 1$, $\boldsymbol{\xi} \in \mathbb{R}^n$ of non-negative weights and $n \geq 1$, we have

$$\begin{aligned}\zeta\|\boldsymbol{\xi} \odot \tilde{\boldsymbol{H}}_n\| &= \zeta[\sum_{i=1}^n \xi_i^2 H_{ni}^2]^{1/2} = \zeta[\sum_{i=1}^n n^{-1}\xi_i^2 H_n^2(X_i, \delta_i, \boldsymbol{U}_i)]^{1/2} \\ &\geq \zeta[\sum_{i=1}^n (n\|\boldsymbol{\xi}\|)^{-1}\xi_i^2 H_n^2(X_i, \delta_i, \boldsymbol{U}_i)1(X_i, \delta_i, \boldsymbol{U}_i)]^{1/2} = \zeta\|H_n\|_{Q_{\boldsymbol{\xi}},2}.\end{aligned}$$

Arguments used in Section 8.1.2 (Kosorok (2008), Chap. 8) indicate the relationship between packing number $D(q, \mathcal{K})$ and covering number $N(q, \mathcal{K})$ for each $q > 0$ and any class $\mathcal{K}$ with respect to the same norm :

$$N(q, \mathcal{K}) \leq D(q, \mathcal{K}) \leq N(q/2, \mathcal{K}).$$

If we let $q = \zeta\|\boldsymbol{\xi} \odot \tilde{\boldsymbol{H}}_n\|$, then this relationship implies for the class $\boldsymbol{\xi} \odot \tilde{\mathcal{H}}_n$,

$$D(\zeta\|\boldsymbol{\xi} \odot \tilde{\boldsymbol{H}}_n\|, \boldsymbol{\xi} \odot \tilde{\mathcal{H}}_n) \leq N(\zeta\|\boldsymbol{\xi} \odot \tilde{\boldsymbol{H}}_n\|/2, \boldsymbol{\xi} \odot \tilde{\mathcal{H}}_n).$$

Since we have perceived $\mathcal{H}_n \supseteq \boldsymbol{\xi} \odot \tilde{\mathcal{H}}_n$ and $\zeta\|\boldsymbol{\xi} \odot \tilde{\boldsymbol{H}}_n\| \geq \zeta\|H_n\|_{Q_{\boldsymbol{\xi}},2}$, it leads to

$$N(\zeta\|\boldsymbol{\xi} \odot \tilde{\boldsymbol{H}}_n\|/2, \boldsymbol{\xi} \odot \tilde{\mathcal{H}}_n) \leq N(\zeta\|H_n\|_{Q_{\boldsymbol{\xi}},2}/2, \mathcal{H}_n).$$

The above two equations further reveal that

$$D(\zeta\|\boldsymbol{\xi} \odot \tilde{\boldsymbol{H}}_n\|, \boldsymbol{\xi} \odot \tilde{\mathcal{H}}_n) \leq \sup_Q N(\zeta\|H_n\|_{Q,2}/2, \mathcal{H}_n). \tag{S3.10}$$

Let $\lambda(\zeta) = \limsup_{n\to\infty} \sup_Q N(\zeta\|H_n\|_{Q,2}/2, \mathcal{H}_n)$. By (S3.10), we can conclude

$$D(\zeta\|\boldsymbol{\xi} \odot \tilde{\boldsymbol{H}}_n\|, \boldsymbol{\xi} \odot \tilde{\mathcal{H}}_n) \leq \lambda(\zeta),$$

for $0 < \zeta \leq 1$, $\boldsymbol{\xi} \in \mathbb{R}^n$ of non-negative weights, and all $n \geq 1$. Note that λ does not depend on n.

Condition (B)

Since $E\mathbb{W}_n(t) = 0$ for any t, we can obtain that for $s, t \in \mathcal{T}$,

$$\begin{aligned}\sigma_W(s,t) &= \lim_{n\to\infty} E\mathbb{W}_n(t)\mathbb{W}_n(s) = \lim_{n\to\infty}\sum_{i=1}^{n} Eh_{ni}(t)h_{ni}(s)\\ &= \sum_{j=1}^{p}\sum_{k=1}^{p}\tilde{a}_j\tilde{a}_k\sigma_M(j,k) + \sum_{j=1}^{p}\tilde{a}_j\sigma_{ML}(j,s) + \sum_{j=1}^{p}\tilde{a}_j\sigma_{ML}(j,t) + \sigma_L(s,t).\end{aligned}$$

where $(\sigma_M(j,k))_{j,k=1,\dots,p}$ is the covariance matrix of the mean-zero normal random vector $\boldsymbol{M}$, $\sigma_L(s,t)$ is the covariance function of the Gaussian process $\mathbb{L}$ at any s as well as t, and $\sigma_{ML}(j,t)$ is the covariance function of the joint Gaussian process $(\boldsymbol{M}, \mathbb{L})$ for any j, t.

Recall that $\tilde{Y} = \delta X/G_0(X)$ and $\tilde{\varepsilon} = \tilde{Y} - \alpha_0 - \boldsymbol{U}^T\boldsymbol{\beta}_0$. Specifically, $(\sigma_M(j,k))_{j,k=1,\dots,p}$ can be given by the covariance matrix of the random vector with components

$$(\tilde{\varepsilon} + (\boldsymbol{U} - E\boldsymbol{U})^T\boldsymbol{\beta}_0 - (U_j - EU_j)C_j^T\boldsymbol{\beta}_0/V_j)(U_j - EU_j), \tag{S3.11}$$

for $j = 1, \dots, p$. The dominated convergence theorem ensures that $\sigma_L(s,t)$ can be provided by the covariance function of a stochastic process at locations s and t, where the stochastic process is

$$\{\phi_t(X)\gamma_0(X)(1-\delta) + \gamma_1(X,t)\delta - \gamma_2(X,t) - G_0(t),\ t \in \mathcal{T}\}. \tag{S3.12}$$

Moreover, we can obtain $\sigma_{ML}(j,t)$ by the cross covariance between the component $(\tilde{\varepsilon} + (\boldsymbol{U} - E\boldsymbol{U})^T\boldsymbol{\beta}_0 - (U_j - EU_j)C_j^T\boldsymbol{\beta}_0/V_j)(U_j - EU_j)$ and the process in (S3.12) at location t. The finite fourth moment of $\boldsymbol{U}$ (implied by (A.1)) and the square-integrability of $\tilde{\varepsilon}$ (implied by (A.2)), along with the results in Stute (1995) (based on (A.4)), ensure the existence of $\sigma_W(s,t)$ for any $s, t \in \mathcal{T}$.

Condition (C)

According to the definition of $\tilde{\boldsymbol{H}}_n$, we first express $\sum_{i=1}^{n} EH_{ni}^2$ as

$$\begin{aligned}&\sum_{i=1}^{n} E(\sum_{j=1}^{p} a_j|f_{ni,j}| + \sup_{t\in\mathcal{T}}|g_{ni}(t)|)^2\\ &= \sum_{i=1}^{n}\left[\sum_{j,k=1}^{p} a_ja_kE|f_{ni,j}f_{ni,k}| + E(\sup_{t\in\mathcal{T}}|g_{ni}^2(t)|) + 2\sum_{j=1}^{p} a_jE(|f_{ni,j}|\sup_{t\in\mathcal{T}}|g_{ni}(t)|)\right],\end{aligned}$$

which is bounded by

$$\begin{aligned}
\max_i\{ [\sum_{j,k=1}^{p} a_j a_k E(\tilde{\varepsilon}_{ni} + (\boldsymbol{U}_i - E\boldsymbol{U})^T\boldsymbol{\beta}_0 - (\boldsymbol{U}_{ij} - EU_j)\frac{C_j^T\boldsymbol{\beta}_0}{V_j})(\tilde{\varepsilon}_{ni} + (\boldsymbol{U}_i \\
- E\boldsymbol{U})^T\boldsymbol{\beta}_0 - (\boldsymbol{U}_{ik} - EU_k)\frac{C_k^T\boldsymbol{\beta}_0}{V_k})(\boldsymbol{U}_{ij} - EU_j)(\boldsymbol{U}_{ik} - EU_k)] \\
+ E(\sup_{t\in\mathcal{T}}[\phi_t(X_i)\gamma_0(X_i)(1-\delta_i) + \gamma_1(X_i,t)\delta_i - \gamma_2(X_i,t) - G_0(t)]^2) \\
+ 2\sum_{j=1}^{p} a_j E(|(\tilde{\varepsilon}_{ni} + (\boldsymbol{U}_i - E\boldsymbol{U})^T\boldsymbol{\beta}_0 - (\boldsymbol{U}_{ij} - EU_j)\frac{C_j^T\boldsymbol{\beta}_0}{V_j})(\boldsymbol{U}_{ij} - EU_j)| \\
\sup_{t\in\mathcal{T}}|\phi_t(X_i)\gamma_0(X_i)(1-\delta_i) + \gamma_1(X_i,t)\delta_i - \gamma_2(X_i,t) - G_0(t)|) \}.
\end{aligned} \tag{S3.13}$$

We can further show that the first term in (S3.13) is finite because (A.1)–(A.2) imply the square integrability $\tilde{\varepsilon}$ and U_jU_k for all j,k. By (A.3)–(A.4), the restriction $X_i \le \tau < \tau_H$ leads to the uniform boundedness of the second term in (S3.13) over $\mathcal{T}$, for all i. By (A.1)–(A.4), it is easy to see the third term is finite as well. Hence, we verify $\limsup_{n\to\infty}\sum_{i=1}^n EH_{ni}^2 < \infty$.

Condition (D)

Recall that $H_n(X_i,\delta_i,\boldsymbol{U}_i) = \sqrt{n}H_{ni}$ and the definition of H_{ni} in (S3.8). For each $\eta > 0$,

$$\sum_{i=1}^{n} EH_{ni}^2 1(H_{ni} > \eta) = n^{-1}\sum_{i=1}^{n} EH_n^2(X_i,\delta_i,\boldsymbol{U}_i)1(H_n(X_i,\delta_i,\boldsymbol{U}_i) > \eta\sqrt{n}), \tag{S3.14}$$

where

$$\begin{aligned}
H_n(X_i,\delta_i,\boldsymbol{U}_i) &= \sqrt{n}[\sum_{j=1}^{p} a_j F_{ni,j} + G_{ni}] \\
&= \sum_{j=1}^{p} a_j |(\tilde{\varepsilon}_{ni} + (\boldsymbol{U}_i - E\boldsymbol{U})^T\boldsymbol{\beta}_0 - (\boldsymbol{U}_{ij} - EU_j)\frac{C_j^T\boldsymbol{\beta}_0}{V_j})(\boldsymbol{U}_{ij} - EU_j)| \\
&\quad + \sup_{t\in\mathcal{T}}|\phi_t(X_i)\gamma_0(X_i)(1-\delta_i) + \gamma_1(X_i,t)\delta_i - \gamma_2(X_i,t) - G_0(t)|.
\end{aligned}$$

Note that (A.1)–(A.2) imply $\tilde{\varepsilon}$ and U_jU_k are square-integrable for all j,k, and (A.4) gives that $\phi_t(X_i)\gamma_0(X_i)(1-\delta_i)+\gamma_1(X_i,t)\delta_i-\gamma_2(X_i,t)-G_0(t)$ is uniformly bounded over $\mathcal{T}$ for all i. Therefore, we have $H_n(X_i,\delta_i,\boldsymbol{U}_i)$ is bounded for all but finite many i for all $n \ge 1$. As $n \to \infty$, (S3.14) tends to zero since the numerator is a finite sum but the denominator diverges. Hence, we show Condition (D) (the analogy of the Lindeberg condition) satisfied.

Condition (E)

For every $s, t \in \mathcal{T}$, $\rho_n(s,t) = (\sum_{i=1}^n E|h_{ni}(t) - h_{ni}(s)|^2)^{1/2}$, such that

$$\rho_n^2(s,t) = \sum_{i=1}^n E|h_{ni}(t) - h_{ni}(s)|^2 = \sum_{i=1}^n E|g_{ni}(t) - g_{ni}(s)|^2.$$

Since $g_{ni}(t) = g_i(t)/\sqrt{n} + o(1)$, where $g_i(t) = [\phi_t(X_i)\gamma_0(X_i)(1-\delta_i) + \gamma_1(X_i,t)\delta_i - \gamma_2(X_i,t) - G_0(t)]$ and $\{g_i\}$ are i.i.d., then Condition (E) is trivially satisfied, according to Pollard (1990). □

Lemma 3.4. *Suppose that (A.1)–(A.4) hold and $\boldsymbol{\beta}_0 \neq \mathbf{0}$. The function Ψ_j is continuous on $\mathbb{R} \times \ell_\tau^\infty \times \mathcal{P}$, for all j.*

Proof. For $\alpha > 0$ and $A \in \mathcal{A}$, denote the Euclidean norm by $\|\cdot\|$ and we define the distance

$$d(\boldsymbol{x}, A) = \inf\{\|\boldsymbol{x} - \boldsymbol{a}\| : \boldsymbol{a} \in A\}$$

and $A_\alpha = \{\boldsymbol{x} : d(\boldsymbol{x}, A) \leq \alpha\}$ if $A \neq \emptyset$; otherwise, $A_\alpha = \emptyset$. For any probability measure $Q \in \mathcal{P}$, we can further define the Prokhorov metric between P and Q as

$$d_p(P,Q) = \inf\{\alpha > 0 : P(A) \leq Q(A_\alpha) + \alpha \text{ and } Q(A) \leq P(A_\alpha) + \alpha, \forall A \in \mathcal{A}\}.$$

For any given $\tilde{\epsilon} > 0$, suppose that there exists a probability measure $Q \in \mathcal{P}$ that satisfies $d_p(P,Q) < \tilde{\epsilon}$. Since $\tilde{\epsilon}$ can be arbitrarily small, it implies that there is a positive sequence $\alpha_n \downarrow 0$, such that $P(A) \leq Q(A_{\alpha_n}) + \alpha_n$ and $Q(A) \leq P(A_{\alpha_n}) + \alpha_n$, for all n. We can easily see A_α is closed, and therefore so is A_{α_n}. Let $\bar{A} = \cap_n A_{\alpha_n}$, where $\bar{A}$ is closed and $\bar{A}$ is exactly the closure of A. It follows that $P(A) \leq Q(\bar{A})$ and $Q(A) \leq P(\bar{A})$, which leads to $P(A) = Q(A)$ for all closed sets A. Hence, we can conclude that $P = Q$ by inner regularity.

Recall that

$$\Psi_j(m, h, Q) = m + Q\left[\frac{(U_j - EU_j)\tilde{Y}h(X)}{G_0(X)}\right],$$

where we should point out that $Q[\cdot]$ is the expected value of a random variable with respect to the probability measure Q and EU_j denotes the expectation of U_j with respect to $P \in \mathcal{P}$. To show the continuity of Ψ_j on $\mathbb{R} \times \ell_\tau^\infty \times \mathcal{P}$, it suffices to prove that the second term of Ψ_j is continuous on $\ell_\tau^\infty \times \mathcal{P}$. For any $\epsilon > 0$, there exists $\tilde{\epsilon} > 0$ such that $\sup_{t\in\mathcal{T}} |\tilde{\mathbb{L}}(t) - \mathbb{L}(t)| < \tilde{\epsilon}$ and $d_p(\tilde{P}, P) < \tilde{\epsilon}$, where

$\tilde{\mathbb{L}}, \mathbb{L} \in \ell_\tau^\infty$ and $\tilde{P}, P \in \mathcal{P}$. It follows that

$$\begin{aligned}
&\left|\tilde{P}\frac{(U_j - EU_j)\tilde{Y}\tilde{\mathbb{L}}(X)}{G_0(X)} - P\frac{(U_j - EU_j)\tilde{Y}\mathbb{L}(X)}{G_0(X)}\right| \\
&\quad\le \left|(\tilde{P} - P)\frac{(U_j - EU_j)\tilde{Y}\tilde{\mathbb{L}}(X)}{G_0(X)}\right| + \left|P\frac{(U_j - EU_j)\tilde{Y}(\tilde{\mathbb{L}}(X) - \mathbb{L}(X))}{G_0(X)}\right| \\
&\quad\le \left|(\tilde{P} - P)\frac{(U_j - EU_j)\tilde{Y}\tilde{\mathbb{L}}(X)}{G_0(X)}\right| + P\left|\frac{(U_j - EU_j)\tilde{Y}(\tilde{\mathbb{L}}(X) - \mathbb{L}(X))}{G_0(X)}\right| \\
&\quad\le \left|(\tilde{P} - P)\frac{(U_j - EU_j)\tilde{Y}\tilde{\mathbb{L}}(X)}{G_0(X)}\right| + \sup_{t\in\mathcal{T}}|\tilde{\mathbb{L}}(t) - \mathbb{L}(t)|P\left|\frac{(U_j - EU_j)\tilde{Y}}{G_0(X)}\right|.
\end{aligned} \tag{S3.15}$$

Recall that (A.1) and (A.4) imply that U_j is bounded, and G_0 is bounded away from zero, respectively. Also recall that $X \in \mathcal{T} = (-\infty, \tau]$, and let $\tilde{\mathbb{L}} \in \ell_\tau^\infty$ (the space of stochastically bounded functions on $\mathcal{T}$). Then, we can show the first term in the last inequality from (S3.15) disappearing because of $\tilde{P} = P$ by inner regularity. Accompanying the square-integrability of U_jU_k for any j, k, the finite second moment of $\tilde{\varepsilon}$, and non-zero $G_0(t)$ for all $t \in \mathcal{T}$ (indicated by (A.1), (A.2) and (A.4), respectively), it leads to

$$P\left|\frac{(U_j - PU_j)\tilde{Y}}{G_0(X)}\right| \le M,$$

where M is a constant. Hence, it implies that

$$\left|\tilde{P}\frac{(U_j - EU_j)\tilde{Y}\tilde{\mathbb{L}}(X)}{G_0(X)} - P\frac{(U_j - EU_j)\tilde{Y}\mathbb{L}(X)}{G_0(X)}\right| \le \tilde{\epsilon}\cdot M.$$

Let $\epsilon \ge \tilde{\epsilon}\cdot M$, and the proof of continuity is completed. □

Lemma 3.5. *Suppose that (A.1)–(A.4) hold; that $\boldsymbol{\beta}_0 \ne \mathbf{0}$, and that j_0 is unique when $\boldsymbol{\beta}_0 \ne \mathbf{0}$. We have that*

$$\sqrt{n}(\hat{\theta}_n - \theta_n) \xrightarrow{d} \frac{\Psi_{j_0}(M_{j_0}, \mathbb{L}, P)}{V_{j_0}}.$$

Following notations in Theorem 1, it leads to

$$\sqrt{n}(\hat{\theta}_n - \theta_n) \xrightarrow{d} \frac{M_{j_0} + \varphi_{j_0}(\mathbb{L})}{V_{j_0}}, \text{ where } \varphi_{j_0}(\mathbb{L}) = E\left[\frac{(U_{j_0} - EU_{j_0})T\mathbb{L}(T)}{G_0(T)}\right].$$

Proof. Since Lemma 3.3 gives that $(\tilde{\mathbb{M}}_n, \mathbb{L}_n)$ converges weakly to $(\boldsymbol{M}, \mathbb{L})$ on $\mathbb{R}^p \times \ell_\tau^\infty$ a.s., and $\mathbb{P}_n$ converges a.s. to P, then we could have $(\tilde{\mathbb{M}}_n, \mathbb{L}_n, \mathbb{P}_n) \xrightarrow{d} (\boldsymbol{M}, \mathbb{L}, P)$ on $\mathbb{R}^p \times \ell_\tau^\infty \times \mathcal{P}$ a.s. It can further indicate that $(\mathbb{M}_{n,j_0}, \mathbb{L}_n, \mathbb{P}_n) \xrightarrow{d} (M_{j_0}, \mathbb{L}, P)$ on $\mathbb{R} \times \ell_\tau^\infty \times \mathcal{P}$ a.s. Recall that Lemma 3.2 gives

$$\sqrt{n}(\hat{\theta}_n - \theta_n) S^2_{\hat{j}_n} = \Psi_{j_0}(\mathbb{M}_{n,j_0}, \mathbb{L}_n, \mathbb{P}_n) + o_p(1).$$

Accompanying the continuity of Ψ_{j_0} shown in Lemma 3.4, therefore we can use continuous mapping theorem to develop that

$$\Psi_{j_0}(\mathbb{M}_{n,j_0}, \mathbb{L}_n, \mathbb{P}_n) \xrightarrow{d} \Psi_{j_0}(M_{j_0}, \mathbb{L}, P).$$

Along with the fact that $S^2_{\hat{j}_n}$ converges to V_{j_0} a.s. by $\hat{j}_n \xrightarrow{a.s.} j_0$ (shown in Lemma 3.6) and SLLN, Slutsky's lemma implies that,

$$\sqrt{n}(\hat{\theta}_n - \theta_n) \xrightarrow{d} \frac{\Psi_{j_0}(M_{j_0}, \mathbb{L}, P)}{V_{j_0}} = M_{j_0} + E\left[\frac{(U_{j_0} - EU_{j_0})T\mathbb{L}(T)}{G_0(T)}\right],$$

where the last equality follows from techniques of conditional expectation and the dominated convergence theorem when (A.1)–(A.4) hold. □

Lemma 3.6 (The oracle property)**.** *Suppose that (A.1)–(A.4) hold; that* $\boldsymbol{\beta}_0 \neq \mathbf{0}$*, and that* j_0 *is unique when* $\boldsymbol{\beta}_0 \neq \mathbf{0}$*. We have* $\hat{j}_n$ *converges to* j_0 *a.s.*

Proof. Recall that $\boldsymbol{U}_{ij}$ denotes the i-th subject's U_j. Based on a marginal AFT model with respect to U_j, we can have mean squared errors $\hat{R}_j = \mathbb{P}_n(Y - \hat{\alpha}_j - \hat{\beta}_j U_j)^2$, where $(\hat{\alpha}_j, \hat{\beta}_j)$ denotes the KSV estimator of parameters in this marginal AFT model and can be written as $(\mathbb{P}_n Y - \hat{\beta}_j \mathbb{P}_n U_j,\ \mathbb{P}_n(U_j - \mathbb{P}_n U_j)Y/S_j^2)$. Therefore for all j, we have

$$\hat{R}_j = S_Y^2 - \mathbb{P}_n(U_j - \mathbb{P}_n U_j)Y/S_j^2,$$

and the above display indicates that the following two arguments are equivalent:

$$\arg\max_j \left|\frac{\mathbb{P}_n(U_j - \mathbb{P}_n U_j)Y}{S_Y S_j}\right| \text{ and } \arg\min_j \hat{R}_j. \tag{S3.16}$$

Equation (S3.16) reveals that

$$\hat{j}_n = \arg\max_j \left|\frac{\mathbb{P}_n(U_j - \mathbb{P}_n U_j)Y}{S_Y S_j}\right| = \arg\min_j \hat{R}_j.$$

We first need to prove: for all j,

$$\mathbb{P}_n U_j Y = \mathbb{P}_n U_j \tilde{Y} \text{ a.s., and } \mathbb{P}_n Y = \mathbb{P}_n \tilde{Y} \text{ a.s.} \tag{S3.17}$$

To construct the first equality in (S3.17), we re-express $\mathbb{P}_n U_j Y$ as

$$\mathbb{P}_n \left[\frac{U_j \delta X}{G_0(X)}\right] - \mathbb{P}_n \left[\frac{U_j \delta X}{G_0(X)} \left(\frac{\hat{G}_n(X) - G_0(X)}{\hat{G}_n(X)}\right)\right],$$

which can be defined as $\mathbb{P}_n U_j \tilde{Y} - r_1$, and gives us that $|\mathbb{P}_n U_j Y - \mathbb{P}_n U_j \tilde{Y}| = |r_1|$. It is easy to see the remainder term $|r_1|$ bounded by

$$\frac{\sup_{t \le \tau} |\hat{G}_n(t) - G_0(t)|}{\hat{G}_n(\tau)} \mathbb{P}_n \left|\frac{U_j \delta X}{G_0(X)}\right|,$$

where this upper bound doesn't diverge since in (A.4) we assume non-zero $G_0(t)$, for all $t \le \tau$. Along with (A.3)–(A.4), SLLN and the square-integrability of $\tilde{\varepsilon}$ as well as $U_j U_k$ for all j, k (implied by (A.1)–(A.2)) give that

$$\mathbb{P}_n \left|\frac{U_j \delta X}{G_0(X)}\right| = \mathbb{P}_n |U_j \tilde{Y}| \overset{a.s.}{\to} E|U_j T|,$$

where $E|U_j T|$ can be shown as a finite constant by (A.1)–(A.2). Accompanying the strong uniform consistency of Kaplan–Meier estimator (Stute and Wang (1993)), it implies that the upper bound of $|r_1|$ converges to zero a.s., and so does $|r_1|$, leading to the first equality in (S3.17). We can also ensure the second equality in (S3.17) by similar arguments. Along with the square-integrability of $\tilde{\varepsilon}$ and $U_j U_k$ for all j, k in (A.1)–(A.2), SLLN implies that

$$\mathbb{P}_n U_j \tilde{Y} \overset{a.s.}{\to} EU_j \tilde{Y} = EU_j T \text{ and } \mathbb{P}_n \tilde{Y} \overset{a.s.}{\to} E\tilde{Y} = ET,$$

where the equalities $EU_j \tilde{Y} = EU_j T$ and $E\tilde{Y} = ET$ follow from the arguments of conditional expectation and (A.3). Combined with $\mathbb{P}_n U_j \overset{a.s.}{\to} EU_j$, the above display further indicates that

$$\mathbb{P}_n (U_j - \mathbb{P}_n U_j) \tilde{Y} \overset{a.s.}{\to} EU_j T - EU_j ET = \text{Cov}(U_j, T). \tag{S3.18}$$

Because SLLN implies $\mathbb{P}_n U_j^2 \overset{a.s.}{\to} EU_j^2$ and $\mathbb{P}_n U_j \overset{a.s.}{\to} EU_j$, it is also easy to see that

$$S_j^2 \overset{a.s.}{\to} V_j. \tag{S3.19}$$

Applying continuous mapping theorem on (S3.18) and (S3.19), we obtain that

$$\hat{\beta}_j = \frac{\mathbb{P}_n(U_j - \mathbb{P}_n U_j)\tilde{Y}}{S_j^2} \xrightarrow{a.s.} \frac{\text{Cov}(U_j, T)}{V_j} \text{ for each } j,$$

so that

$$\widehat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)^T \xrightarrow{a.s.} \left(\frac{\text{Cov}(U_1, T)}{V_1}, \ldots, \frac{\text{Cov}(U_p, T)}{V_p}\right)^T. \tag{S3.20}$$

Let $\widehat{\boldsymbol{R}} = (\hat{R}_1, \ldots, \hat{R}_p)^T$ and $\mathbf{1}_p$ denote a p-variate vector $(1, \ldots, 1)^T$. When $\boldsymbol{\beta}_0 \neq \mathbf{0}$ such that $\text{Var}(\boldsymbol{U}^T\boldsymbol{\beta}_0) > 0$, using continuous mapping theorem on (S3.19) and (S3.20) leads to

$$\begin{aligned}\frac{S_Y^2\mathbf{1}_p - \widehat{\boldsymbol{R}}}{\text{Var}(\boldsymbol{U}^T\boldsymbol{\beta}_0)} &= \left(\frac{\hat{\beta}_1^2 S_1^2}{\text{Var}(\boldsymbol{U}^T\boldsymbol{\beta}_0)}, \ldots, \frac{\hat{\beta}_p^2 S_p^2}{\text{Var}(\boldsymbol{U}^T\boldsymbol{\beta}_0)}\right) \xrightarrow{a.s.} \left(\frac{\text{Cov}^2(U_1, T)}{V_1\text{Var}(\boldsymbol{U}^T\boldsymbol{\beta}_0)}, \ldots, \frac{\text{Cov}^2(U_p, T)}{V_p\text{Var}(\boldsymbol{U}^T\boldsymbol{\beta}_0)}\right)^T \\ &= (\text{Corr}^2(U_1, T), \ldots, \text{Corr}^2(U_p, T))^T.\end{aligned}$$

Note that $j_0 = \arg\max_j |\text{Corr}(U_j, T)|$, which is equivalent to $j_0 = \arg\max_j \text{Corr}^2(U_j, T)$. Since we have shown that $\hat{j}_n$ can also be the argument to maximize $(S_Y^2 - \hat{R}_j)/\text{Var}(\boldsymbol{U}^T\boldsymbol{\beta}_0)$ among all j's, then $\hat{j}_n \xrightarrow{a.s.} j_0$, given that j_0 is unique. □

Lemma 3.7. *Suppose that (A.1)–(A.4) hold and* $\boldsymbol{\beta}_0 = \mathbf{0}$. *The joint limiting distribution of* $\sqrt{n}\widehat{\boldsymbol{\theta}}$ *and* $n(S_Y^2\mathbf{1}_p - \widehat{\boldsymbol{R}})$ *can be derived as*

$$\begin{pmatrix} (M_1 + \varphi_1(\mathbb{L}) + C_1^T\boldsymbol{b}_0)/V_1, & \ldots, & (M_p + \varphi_p(\mathbb{L}) + C_p^T\boldsymbol{b}_0)/V_p \\ (M_1 + \varphi_1(\mathbb{L}) + C_1^T\boldsymbol{b}_0)^2/V_1, & \ldots, & (M_p + \varphi_p(\mathbb{L}) + C_p^T\boldsymbol{b}_0)^2/V_p \end{pmatrix}^T,$$

where C_j *as well as* V_j *are as previously defined, for any fixed* j.

Proof. To prove this lemma, the first step is to derive the limiting distribution of $\sqrt{n}\widehat{\boldsymbol{\theta}} = (\sqrt{n}\hat{\theta}_1, \ldots, \sqrt{n}\hat{\theta}_p)^T$, where $\hat{\theta}_j$ is the KSV estimator of the regression coefficient in a marginal AFT model with the predictor U_j and the outcome Y. The second step provides the joint limiting distribution of $\sqrt{n}\widehat{\boldsymbol{\theta}}$ and $n(S_Y^2\mathbf{1}_p - \widehat{\boldsymbol{R}}) = (n(S_Y^2 - \hat{R}_1), \ldots, n(S_Y^2 - \hat{R}_p))^T$, where $\hat{R}_j$ is defined as before, for all j. We begin with re-expressing $\sqrt{n}\hat{\theta}_j$ as $\sqrt{n}\mathbb{P}_n(U_j - \mathbb{P}_n U_j)Y/S_j^2$, which can be further written as, for all j,

$$\frac{\sqrt{n}}{S_j^2}\mathbb{P}_n(U_j - \mathbb{P}_n U_j)\tilde{Y} + \frac{1}{S_j^2}\mathbb{P}_n\left[\frac{(U_{j_0} - EU_{j_0})\tilde{Y}\mathbb{L}_n(X)}{G_0(X)}\right] + o_p(1). \tag{S3.21}$$

Since $\tilde{\varepsilon}_n = \tilde{Y} - \alpha_0 - \boldsymbol{U}^T\boldsymbol{\beta}_n$, the linear property of sample covariance implies that

$$\mathbb{P}_n(U_j - \mathbb{P}_n U_j)\tilde{Y} = \mathbb{P}_n(U_j - \mathbb{P}_n U_j)\boldsymbol{U}^T\boldsymbol{\beta}_n + \mathbb{P}_n(U_j - \mathbb{P}_n U_j)\tilde{\varepsilon}_n, \tag{S3.22}$$

where we can further have

$$\begin{aligned}\mathbb{P}_n(U_j - \mathbb{P}_n U_j)\boldsymbol{U}^T\boldsymbol{\beta}_n = &(\mathbb{P}_n - P)U_j\boldsymbol{U}^T\boldsymbol{\beta}_n + PU_j\boldsymbol{U}^T\boldsymbol{\beta}_n - (\mathbb{P}_n - P)U_j\mathbb{P}_n\boldsymbol{U}^T\boldsymbol{\beta}_n \\ &- PU_j(\mathbb{P}_n - P)\boldsymbol{U}^T\boldsymbol{\beta}_n - PU_jP\boldsymbol{U}^T\boldsymbol{\beta}_n,\end{aligned}$$

and

$$\mathbb{P}_n(U_j - \mathbb{P}_n U_j)\tilde{\varepsilon}_n = (\mathbb{P}_n - P)(\tilde{\varepsilon}_n(U_j - PU_j) - \mathbb{P}_n\tilde{\varepsilon}_n(\mathbb{P}_n - P)U_j. \tag{S3.23}$$

Let $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$. Along with $C_j^T\boldsymbol{\beta}_n = PU_j\boldsymbol{U}^T\boldsymbol{\beta}_n - PU_jP\boldsymbol{U}^T\boldsymbol{\beta}_n$, (S3.21)–(S3.23) lead to

$$\begin{aligned}\sqrt{n}\hat{\theta}_j = &\frac{(\mathbb{G}_nU_j\boldsymbol{U}^T - PU_j\mathbb{G}_n\boldsymbol{U}^T - \mathbb{G}_nU_j\mathbb{P}_n\boldsymbol{U}^T)\boldsymbol{\beta}_n}{S_j^2} - \frac{\mathbb{P}_n\tilde{\varepsilon}_n\mathbb{G}_nU_j}{S_j^2} + \frac{\mathbb{G}_n\tilde{\varepsilon}_n(U_j - PU_j)}{S_j^2} \\ &+ \frac{1}{S_j^2}\mathbb{P}_n\left[\frac{(U_{j_0} - EU_{j_0})\tilde{Y}\mathbb{L}_n(X)}{G_0(X)}\right] + \frac{\sqrt{n}C_j^T\boldsymbol{\beta}_n}{S_j^2} + o_p(1).\end{aligned}$$

When $\boldsymbol{\beta}_0 = \mathbf{0}$, then $\sqrt{n}\boldsymbol{\beta}_n = b_0$ such that

$$\begin{aligned}\sqrt{n}\hat{\theta}_j = &\frac{(\mathbb{G}_nU_j\boldsymbol{U}^T - PU_j\mathbb{G}_n\boldsymbol{U}^T - \mathbb{G}_nU_j\mathbb{P}_n\boldsymbol{U}^T)\boldsymbol{b}_0}{\sqrt{n}\,S_j^2} - \frac{\mathbb{P}_n\tilde{\varepsilon}_n\mathbb{G}_nU_j}{S_j^2} + \frac{\mathbb{G}_n\tilde{\varepsilon}_n(U_j - PU_j)}{S_j^2} \\ &+ \frac{1}{S_j^2}\mathbb{P}_n\left[\frac{(U_j - EU_j)\tilde{Y}\mathbb{L}_n(X)}{G_0(X)}\right] + \frac{C_j^T\boldsymbol{b}_0}{S_j^2} + o_p(1).\end{aligned} \tag{S3.24}$$

Since the first two terms in (S3.24) are $o_p(1)$ by SLLN, along with the definition of $\mathbb{M}_{n,j}$, we can have

$$\begin{aligned}\sqrt{n}\hat{\theta}_j &= \frac{1}{S_j^2}\left\{\mathbb{G}_n\tilde{\varepsilon}_n(U_j - PU_j) + \mathbb{P}_n\frac{(U_j - EU_j)\tilde{Y}\mathbb{L}_n(X)}{G_0(X)}\right\} + \frac{C_j^T\boldsymbol{b}_0}{S_j^2} + o_p(1) \\ &= \frac{1}{S_j^2}\left\{\mathbb{M}_{n,j} + \mathbb{P}_n\frac{(U_j - EU_j)\tilde{Y}\mathbb{L}_n(X)}{G_0(X)}\right\} + \frac{C_j^T\boldsymbol{b}_0}{S_j^2} + o_p(1).\end{aligned}$$

Using previous arguments along with (A.1)–(A.4), it further leads to

$$\sqrt{n}\widehat{\boldsymbol{\theta}} \xrightarrow{d} \left(\frac{M_1 + \varphi_1(\mathbb{L}) + C_1^T\boldsymbol{b}_0}{V_1}, \ldots, \frac{M_p + \varphi_p(\mathbb{L}) + C_p^T\boldsymbol{b}_0}{V_p}\right)^T.$$

To complete the second step, we re-express $n(S_j^2\mathbf{1}_p - \widehat{\boldsymbol{R}})$ as $(\sqrt{n}\widehat{\boldsymbol{\theta}}) \odot (\sqrt{n}\widehat{\boldsymbol{\theta}}) \odot (S_1^2, \ldots, S_p^2)^T$, where $\odot$ denotes the Hadamard product. Hence when $\boldsymbol{\beta}_0 = \mathbf{0}$, the joint distribution of $\sqrt{n}\widehat{\boldsymbol{\theta}}$ and $n(S_Y^2\mathbf{1}_p - \widehat{\boldsymbol{R}})$

can be obtained as

$$\begin{pmatrix} \sqrt{n}\widehat{\boldsymbol{\theta}} \\ n(S_Y^2\mathbf{1}_p - \widehat{\boldsymbol{R}}) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} ((M_1+\varphi_1(\mathbb{L})+C_1^T\boldsymbol{b}_0)/V_1, & \ldots, & (M_p+\varphi_p(\mathbb{L})+C_p^T\boldsymbol{b}_0)/V_p)^T \\ ((M_1+\varphi_1(\mathbb{L})+C_1^T\boldsymbol{b}_0)^2/V_1, & \ldots, & (M_p+\varphi_p(\mathbb{L})+C_p^T\boldsymbol{b}_0)^2/V_p)^T \end{pmatrix}.$$

□

Lemma 3.8 (McKeague and Qian, (2015)). *Let $\boldsymbol{z}$ be a p-dimensional random vector and $f:\mathbb{R}^{2p}\to\mathbb{R}^p$ a function such that $f(\boldsymbol{z},\cdot)$ is continuous for every $\boldsymbol{z}\in\mathbb{R}^p$, and $f(\boldsymbol{z},\boldsymbol{b})_j \neq f(\boldsymbol{z},\boldsymbol{b})_k$ a.s. for all $j\neq k$ and $\boldsymbol{b}\in\mathbb{R}^p$. Then, $J(\boldsymbol{b})\equiv \arg\max_{j=1,\ldots,p} f(\boldsymbol{z},\boldsymbol{b})_j$ is unique a.s. Also, if $\boldsymbol{b}_l\to\boldsymbol{b}_0$, then $J(\boldsymbol{b}_l)=J(\boldsymbol{b}_0)$ for l sufficiently large a.s.*

Lemma 3.9. *Suppose that (A.1)–(A.4) hold and that $\boldsymbol{\beta}_0=\mathbf{0}$.*

$$\sqrt{n}(\hat{\theta}_n-\theta_n) \xrightarrow{d} (M_J+\varphi_J(\mathbb{L}))/V_J + (C_J/V_J - C_{j(\boldsymbol{b}_0)}/V_{j(\boldsymbol{b}_0)})^T\boldsymbol{b}_0,$$

where J, $j(\boldsymbol{b}_0)$, C_j and V_j are as defined in Theorem 1, for each j.

Proof. It is easy to perceive $f(\boldsymbol{z},\cdot)$ we defined is continuous with respect to $\boldsymbol{z}$. Also, $\boldsymbol{Z} = (Z_1,\ldots,Z_p)^T$ is a random vector and $|\mathrm{Corr}(U_j,U_k)|<1$ for $j\neq k$, so it indicates that $f(\boldsymbol{Z},\boldsymbol{b}_0)_j \neq f(\boldsymbol{Z},\boldsymbol{b}_0)_k$ for any $j\neq k$ a.s., where $f(\boldsymbol{Z},\boldsymbol{b}_0)_j = (Z_j + C_j^T\boldsymbol{b}_0)^2/V_j$. Thus, we can point out that $J = J(\boldsymbol{b}_0) = \arg\max_{j=1,\ldots,p} f(\boldsymbol{Z},\boldsymbol{b}_0)_j$ is unique a.s. Since $\hat{j}_n = \arg\min_j \hat{R}_j$ (equivalent to $\arg\max_j n(S_Y^2 - \hat{R}_j)$) and it is uniquely determined, then we can say that $\boldsymbol{h}(n(S_Y^2\mathbf{1}_p - \widehat{\boldsymbol{R}}))$ is continuous. Moreover in the case of $\boldsymbol{\beta}_0=\mathbf{0}$, we also see that

$$\sqrt{n}\hat{\theta}_n = \sqrt{n}\widehat{\boldsymbol{\theta}}\boldsymbol{h}(n(S_Y^2\mathbf{1}_p - \widehat{\boldsymbol{R}}));\ \ \sqrt{n}\theta_n = \frac{\sqrt{n}C_{j(\boldsymbol{b}_0)}^T\boldsymbol{\beta}_n}{V_{j(\boldsymbol{b}_0)}} \equiv \frac{C_{j(\boldsymbol{b}_0)}^T\boldsymbol{b}_0}{V_{j(\boldsymbol{b}_0)}}.$$

Hence, the desired limiting distribution of $\sqrt{n}\hat{\theta}_n$ can be derived by applying continuous mapping theorem on the joint distribution of $\sqrt{n}\widehat{\boldsymbol{\theta}}$ and $n(S_Y^2\mathbf{1}_p - \widehat{\boldsymbol{R}})$ derived in Lemma 3.7. □

S4 Proof for Theorem 2

Before entering the core of the proof for Theorem 2, we clarify the large sample behavior of the maximally selected studentized statistic $\mathbb{T}_n$ in Lemma 4.1 below. Together with the conditions of the threshold λ_n, the results in this lemma would play a crucial role in designing adaptive resampling.

Lemma 4.1. *Suppose that the threshold λ_n satisfies $\lambda_n = o(\sqrt{n})$ and $\lambda_n\to\infty$, and that (A.1)–(A.4) hold. We have $1(|\mathbb{T}_n|>\lambda_n) \xrightarrow{p} 1(\boldsymbol{\beta}_0\neq\mathbf{0})$.*

Proof. Recall that S_j^2 is the sample variance of U_j for all j and $\mathbb{T}_n = \sqrt{n}\hat{\theta}_n/\hat{\sigma}_n$, where $\hat{\sigma}_n^2 = \mathbb{P}_n(Y - \hat{\alpha}_n - \hat{\theta}_n U_{\hat{j}_n})^2/S_{\hat{j}_n}^2$. We start the proof with verifying that $\hat{\sigma}_n$ is asymptotically bounded above and below. Let $(\hat{\alpha}_j, \hat{\theta}_j)$ denote the estimated intercept and the estimated regression coefficient of U_j in the marginal AFT model that only contains one active predictor U_j for the outcome Y. By (A.1)–(A.4), SLLN and the uniform consistency of Kaplan–Meier estimator, we can show $\hat{\theta}_j \overset{a.s.}{\to} \theta_j \equiv \text{Cov}(U_j, \boldsymbol{U})^T\boldsymbol{\beta}_0/\text{Var}(U_j)$ and $\hat{\alpha}_j \overset{a.s.}{\to} \alpha_0 + E\boldsymbol{U}^T\boldsymbol{\beta}_0 - \theta_j EU_j$, for all j. This further leads to

$$\mathbb{P}_n(Y - \hat{\alpha}_j - \hat{\theta}_n U_j)^2 \overset{a.s.}{\to} E(\tilde{Y} - \alpha_0 - E\boldsymbol{U}^T\boldsymbol{\beta}_0 - (U_j - EU_j)\theta_j)^2 = E(\tilde{\varepsilon} - (U_j - EU_j)\theta_j)^2.$$

Along with $S_j^2 \overset{a.s.}{\to} \text{Var}(U_j) > 0$ for all j, the continuous mapping theorem implies that

$$\frac{\mathbb{P}_n(Y - \hat{\alpha}_j - \hat{\theta}_j U_j)^2}{S_j^2} \overset{a.s.}{\to} \frac{E(\tilde{\varepsilon} - (U_j - EU_j)\theta_j)^2}{\text{Var}(U_j)}.$$

For all j, we ensure that $E(\tilde{\varepsilon} - (U_j - EU_j)\theta_j)^2 < \infty$ by $\text{Var}(U_j) > 0$ and the square-integrability of $\tilde{\varepsilon}$ and U_j. Therefore, $\max_{j=1,\ldots,p}\{\mathbb{P}_n(Y - \hat{\alpha}_j - \hat{\theta}_j U_j)^2/S_j^2\}$ converges to a finite constant. Since $\hat{\sigma}_n \le [\max_{j=1,\ldots,p}\{\mathbb{P}_n(Y - \hat{\alpha}_j - \hat{\theta}_j U_j)^2/S_j^2\}]^{1/2}$, it implies that $\hat{\sigma}_n$ is asymptotically bounded above. Since it is obvious that

$$E(\tilde{\varepsilon} - (U_j - EU_j)\theta_j)^2/\text{Var}(U_j) > 0 \text{ for all } j,$$

then we see that $[\min_{j=1,\ldots,p}\{\mathbb{P}_n(Y - \hat{\alpha}_j - \hat{\theta}_j U_j)^2/S_j^2\}]^{1/2}$ converges to a non-zero finite constant. Because $\hat{\sigma}_n \ge [\min_{j=1,\ldots,p}\{\mathbb{P}_n(Y - \hat{\alpha}_j - \hat{\theta}_j U_j)^2/S_j^2\}]^{1/2}$, we therefore show that $\hat{\sigma}_n$ is asymptotically bounded below. Together with results in Theorem 1, we then prove that $|\mathbb{T}_n| \overset{a.s.}{\to} \infty$ when $\boldsymbol{\beta}_0 \neq \boldsymbol{0}$ and $|\mathbb{T}_n| = O_p(1)$ when $\boldsymbol{\beta}_0 = \boldsymbol{0}$.

To prove this lemma, it suffices to show that the probabilities in the following equation converge to zero:

$$\begin{aligned} E|1(|\mathbb{T}_n| > \lambda_n) - 1(\boldsymbol{\beta}_0 \neq \boldsymbol{0})| &= E|1(|\mathbb{T}_n| \le \lambda_n) - 1(\boldsymbol{\beta}_0 = \boldsymbol{0})| \\ &= P(|\mathbb{T}_n| > \lambda_n, \boldsymbol{\beta}_0 = \boldsymbol{0}) + P(|\mathbb{T}_n| \le \lambda_n, \boldsymbol{\beta}_0 \neq \boldsymbol{0}) \\ &= P(|\mathbb{T}_n| > \lambda_n | \boldsymbol{\beta}_0 = \boldsymbol{0})1(\boldsymbol{\beta}_0 = \boldsymbol{0}) + P(|\mathbb{T}_n| \le \lambda_n | \boldsymbol{\beta}_0 \neq \boldsymbol{0})1(\boldsymbol{\beta}_0 \neq \boldsymbol{0}). \end{aligned} \tag{S4.1}$$

We can see that the first probability in (S4.1) converges to zero because $\lambda_n \to \infty$ along with $|\mathbb{T}_n| = O_p(1)$ when $\boldsymbol{\beta}_0 = \boldsymbol{0}$. Meanwhile, the second probability converges to zero because $\lambda_n = o(\sqrt{n})$ and $0 < |\mathbb{T}_n|/\sqrt{n} = O_p(1)$ when $\boldsymbol{\beta}_0 \neq \boldsymbol{0}$. □

More notations for the bootstrap version of estimators are introduced below. Let $\mathbb{P}_n^*$ be the nonparametric bootstrap of $\mathbb{P}_n$. Replacing P by $\mathbb{P}_n$ and $\mathbb{P}_n$ by $\mathbb{P}_n^*$, $\mathbb{G}_n^* = \sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n)$ is the bootstrapped empirical process, where $\mathbb{P}_n^*$, $\mathbb{P}_n$ and P only operate on functions defined on the sample space $\mathcal{X}$. The notation $\hat{\theta}_n^*$, $\hat{j}_n^*$ and $\hat{\theta}_j^*$ means that the bootstrap version of $\hat{\theta}_n$, $\hat{j}_n$ and $\hat{\theta}_j$, respectively. The bootstrapped Kaplan–Meier estimator is denote by $\hat{G}_n^*$. Note that under the operation of $\mathbb{P}_n^*$ or $\mathbb{G}_n^*$, we use $\hat{G}_n^*$ to replace $\hat{G}_n$ and $\hat{G}_n$ to replace G_0, respectively. All of the bootstrapped estimators are based on n i.i.d. observations taken from $\mathbb{P}_n$. Let E^* denote the expectation conditional on the data, and P^* be the corresponding probability measure.

To justify the claimed results, we first verify the following statements: (1) $1(|\mathbb{T}_n^*| > \lambda_n$ or $|\mathbb{T}_n| > \lambda_n) \xrightarrow{P^*} 1(\boldsymbol{\beta}_0 \neq \mathbf{0})$ and (2) $1(|\mathbb{T}_n^*| \leq \lambda_n$ and $|\mathbb{T}_n| \leq \lambda_n) \xrightarrow{P^*} 1(\boldsymbol{\beta}_0 = \mathbf{0})$ conditionally (on the data) in probability. Afterward, we prove Lemma 4.2 and 4.3, and obtain the desired results along with statements (1) and (2). To show statements (1) and (2), it suffices to give

$$\begin{aligned} &E^*|1(|\mathbb{T}_n^*| > \lambda_n) - 1(\boldsymbol{\beta}_0 \neq \mathbf{0})| = P^*(|\mathbb{T}_n^*| > \lambda_n, \boldsymbol{\beta}_0 = \mathbf{0}) + P^*(|\mathbb{T}_n^*| \leq \lambda_n, \boldsymbol{\beta}_0 \neq \mathbf{0}) \\ &= P^*(|\mathbb{T}_n^*| > \lambda_n | \boldsymbol{\beta}_0 = \mathbf{0}) 1(\boldsymbol{\beta}_0 = \mathbf{0}) + P^*(|\mathbb{T}_n^*| \leq \lambda_n | \boldsymbol{\beta}_0 \neq \mathbf{0}) 1(\boldsymbol{\beta}_0 \neq \mathbf{0}) \to 0 \end{aligned} \tag{S4.2}$$

in probability, implying that $1(|\mathbb{T}_n^*| > \lambda_n) \xrightarrow{p^*} 1(\boldsymbol{\beta}_0 \neq \mathbf{0})$ and $1(|\mathbb{T}_n^*| \leq \lambda_n) \xrightarrow{p^*} 1(\boldsymbol{\beta}_0 = \mathbf{0})$ conditionally (on the data) in probability. The convergence in (S4.2) follows from below arguments. Using Lemma 3.9 and the condition that $\lambda_n \to \infty$ as $n \to \infty$, we can have $P^*(|\mathbb{T}_n^*| > \lambda_n | \boldsymbol{\beta}_0 = \mathbf{0}) \to 0$ in probability. Besides, it is also easy to see $|\theta_n| \to |C_{j_0}^T \boldsymbol{\beta}_0| / V_{j_0}$ when $\boldsymbol{\beta}_0 \neq \mathbf{0}$ and j_0 is unique. Along with the condition that $\lambda_n = o(\sqrt{n})$ and that $\hat{\sigma}_n^*$ converges to a finite constant conditionally (on the data) in probability, we can use Lemma 3.5 and Lemma 4.2 (shown later) to prove

$$\begin{aligned} &P^*(|\mathbb{T}_n^*| \leq \lambda_n | \boldsymbol{\beta}_0 \neq \mathbf{0}) = P^*(\sqrt{n}|(\hat{\theta}_n^* - \hat{\theta}_n) + (\hat{\theta}_n - \theta_n) + \theta_n| \leq \lambda_n \hat{\sigma}_n^* | \boldsymbol{\beta}_0 \neq \mathbf{0}) \\ &\leq P^*(|\theta_n| \leq n^{-1/2} \lambda_n \hat{\sigma}_n^* + |\hat{\theta}_n^* - \hat{\theta}_n| + |\hat{\theta}_n - \theta_n| \, | \boldsymbol{\beta}_0 \neq \mathbf{0}) \to 0 \end{aligned}$$

in probability. Since $1(|\mathbb{T}_n^*| > \lambda_n) \xrightarrow{p^*} 1(\boldsymbol{\beta}_0 \neq \mathbf{0})$ and $1(|\mathbb{T}_n^*| \leq \lambda_n) \xrightarrow{p^*} 1(\boldsymbol{\beta}_0 = \mathbf{0})$ conditionally (on the data) in probability, along with $1(|\mathbb{T}_n| > \lambda_n) \to 1(\boldsymbol{\beta}_0 \neq \mathbf{0})$ in probability, we can justify statements (1) and (2), using Slutsky's lemma.

Before stating necessary lemmas, we express the bootstrapped marginal regression coefficient as follows, which will appear in Lemma 4.2. For $j = 1, \ldots, p$,

$$\begin{aligned} \sqrt{n}\hat{\theta}_j^* &= \frac{\sqrt{n}[\mathbb{P}_n^* U_j Y - (\mathbb{P}_n^* U_j)(\mathbb{P}_n^* Y)]}{[\mathbb{P}_n^* U_j^2 - (\mathbb{P}_n^* U_j)^2]} = \frac{\mathbb{G}_n^* U_j Y - \mathbb{G}_n^* U_j \mathbb{P}_n^* Y - \mathbb{P}_n U_j \mathbb{G}_n^* Y + \sqrt{n}[\mathbb{P}_n U_j Y - \mathbb{P}_n U_j \mathbb{P}_n Y]}{[\mathbb{P}_n^* U_j^2 - (\mathbb{P}_n^* U_j)^2]} \\ &= \frac{\mathbb{G}_n^* U_j Y - \mathbb{G}_n^* U_j \mathbb{P}_n^* Y - \mathbb{P}_n U_j \mathbb{G}_n^* Y + \sqrt{n}\hat{\theta}_j[\mathbb{P}_n U_j^2 - (\mathbb{P}_n U_j)^2]}{[\mathbb{P}_n^* U_j^2 - (\mathbb{P}_n^* U_j)^2]}. \end{aligned}$$

Lemma 4.2. *Suppose that (A.1)–(A.4) hold; that* $\boldsymbol{\beta}_0 \neq \mathbf{0}$ *and that* $j_0 = j(\boldsymbol{\beta}_0)$ *is unique when*

$\boldsymbol{\beta}_0 \neq \mathbf{0}$. *We can have* $\hat{j}_n^* \xrightarrow{p^*} j_0$ *conditionally (on the data) a.s., and* $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \xrightarrow{d} (M_{j_0} + \varphi_{j_0}(\mathbb{L}))/V_{j_0}$ *conditionally (on the data) in probability.*

Proof. Let $S_Y^{*2} = \mathbb{P}_n^* Y^2 - (\mathbb{P}_n^* Y)^2$ and $S_j^{*2} = \mathbb{P}_n^* U_j^2 - (\mathbb{P}_n^* U_j)^2$. When $\boldsymbol{\beta}_0 \neq \mathbf{0}$, SLLN and Slutsky's lemma imply that,

$$S_j^{*2}\hat{\theta}_j^* = n^{-1/2}[\mathbb{G}_n^* U_j Y - \mathbb{G}_n^* U_j \mathbb{P}_n^* Y - \mathbb{P}_n U_j \mathbb{G}_n^* Y] + \hat{\theta}_j S_j^2 \xrightarrow{p^*} C_j^T \boldsymbol{\beta}_0 \text{ a.s.},$$

implying that $\hat{\theta}_j^* \xrightarrow{p^*} C_j^T \boldsymbol{\beta}_0 / V_j$ a.s., for $j = 1, \ldots, p$. Using a similar fashion to expressing the mean squared error, the corresponding bootstrap version can be written as $\hat{R}_j^* = S_Y^{*2} - \hat{\theta}_j^{*2} S_j^{*2}$, leading to that

$$\hat{j}_n^* = \arg\min_j \hat{R}_j^* = \arg\max_j \frac{S_Y^{*2} - \hat{R}_j^*}{\mathrm{Var}(\boldsymbol{U}^T\boldsymbol{\beta}_0)} = \arg\max_j \frac{\hat{\theta}_j^{*2} S_j^{*2}}{\mathrm{Var}(\boldsymbol{U}^T\boldsymbol{\beta}_0)}.$$

Moreover, Slutsky's lemma and SLLN indicate

$$\frac{\hat{\theta}_j^{*2} S_j^{*2}}{\mathrm{Var}(\boldsymbol{U}^T\boldsymbol{\beta}_0)} \xrightarrow{p^*} \mathrm{Corr}^2(U_j, \boldsymbol{U}^T\boldsymbol{\beta}_0) \text{ a.s., for } j = 1, \ldots, p.$$

Along with the condition that j_0 is unique when $\boldsymbol{\beta}_0 \neq \mathbf{0}$, it implies that

$$P^*(\hat{j}_n^* \neq j_0) = P^*\left(\bigcup_{j:j\neq j_0}\left\{\frac{\hat{\theta}_{j_0}^{*2} S_{j_0}^{*2}}{\mathrm{Var}(\boldsymbol{U}^T\boldsymbol{\beta}_0)} \leq \frac{\hat{\theta}_j^{*2} S_j^{*2}}{\mathrm{Var}(\boldsymbol{U}^T\boldsymbol{\beta}_0)}\right\}\right) \leq \sum_{j:j\neq j_0} P^*\left(\frac{\hat{\theta}_{j_0}^{*2} S_{j_0}^{*2}}{\mathrm{Var}(\boldsymbol{U}^T\boldsymbol{\beta}_0)} \leq \frac{\hat{\theta}_j^{*2} S_j^{*2}}{\mathrm{Var}(\boldsymbol{U}^T\boldsymbol{\beta}_0)}\right)$$

converging to zero a.s. Let $\hat{\varepsilon}_n = Y - \hat{\alpha}_n - \hat{\theta}_n U_{\hat{j}_n}$. Recall that $\mathbb{P}_n \hat{\varepsilon}_n = 0$ and the definition of $\hat{\theta}_n^*$, we can have that

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) S_{\hat{j}_n^*}^{*2} &= \sqrt{n}[\mathbb{P}_n^* U_{\hat{j}_n^*} Y - \mathbb{P}_n^* U_{\hat{j}_n^*} \mathbb{P}_n^* Y - \hat{\theta}_n S_{\hat{j}_n^*}^{*2}] \\
&= \sqrt{n}[\mathbb{P}_n^* U_{\hat{j}_n^*} \hat{\varepsilon}_n - \mathbb{P}_n^* U_{\hat{j}_n^*} \mathbb{P}_n^* \hat{\varepsilon}_n - \hat{\theta}_n(\mathbb{P}_n^* U_{\hat{j}_n^*}^2 - (\mathbb{P}_n^* U_{\hat{j}_n^*})^2 - \mathbb{P}_n^* U_{\hat{j}_n^*} U_{\hat{j}_n} + \mathbb{P}_n^* U_{\hat{j}_n^*} \mathbb{P}_n^* U_{\hat{j}_n})] \\
&= \mathbb{G}_n^* U_{\hat{j}_n^*} \hat{\varepsilon}_n - \mathbb{G}_n^* \hat{\varepsilon}_n \mathbb{P}_n U_{\hat{j}_n^*} - \mathbb{G}_n^* U_{\hat{j}_n^*} \mathbb{P}_n^* \hat{\varepsilon}_n - \sqrt{n}\hat{\theta}_n[\mathbb{P}_n^* U_{\hat{j}_n^*}^2 - (\mathbb{P}_n^* U_{\hat{j}_n^*})^2 - \mathbb{P}_n^* U_{\hat{j}_n^*} U_{\hat{j}_n} \\
&\quad + \mathbb{P}_n^* U_{\hat{j}_n^*} \mathbb{P}_n^* U_{\hat{j}_n}] + o_{p^*}(1) \text{ a.s.} \\
&= \mathbb{G}_n^* \hat{\varepsilon}_n (U_{\hat{j}_n^*} - PU_{\hat{j}_n^*}) - \mathbb{G}_n^* \hat{\varepsilon}_n (\mathbb{P}_n - P) U_{\hat{j}_n^*} - \mathbb{G}_n^* U_{\hat{j}_n^*} (\mathbb{P}_n^* - \mathbb{P}_n)\hat{\varepsilon}_n \\
&\quad + \sqrt{n}\hat{\theta}_n[(\mathbb{P}_n^* U_{\hat{j}_n^*})^2 - \mathbb{P}_n^* U_{\hat{j}_n^*}^2 + \mathbb{P}_n^* U_{\hat{j}_n^*} U_{\hat{j}_n} - \mathbb{P}_n^* U_{\hat{j}_n^*} \mathbb{P}_n^* U_{\hat{j}_n}] + o_{p^*}(1) \text{ a.s.,}
\end{aligned} \tag{S4.3}$$

where the third equality follows from $\mathbb{P}_n U_{\hat{j}_n} \hat{\varepsilon}_n = 0$; $\hat{j}_n^* \xrightarrow{p^*} j_0$ a.s.; $\hat{j}_n \to j_0$ a.s., and the last equality follows from $\mathbb{P}_n \hat{\varepsilon}_n = 0$. In the last equality in (S4.3), all the terms can be shown as $o_{p^*}(1)$ a.s. by similar arguments and SLLN, except for the first term. The next to show is the first term in (S4.3) converges in distribution to some weak limit conditionally (on the data) in probability.

According to Lemma 3.6 (under (A.1)–(A.4)), we can easily see that $\hat{\theta}_n \xrightarrow{p} \theta_0 \equiv C_{j_0}^T \boldsymbol{\beta}_0 / V_{j_0}$ and $\hat{\alpha}_n \xrightarrow{p} \alpha_0 + P\boldsymbol{U}^T\boldsymbol{\beta}_0 - \theta_0 PU_{j_0}$. Let $\bar{\varepsilon}_n = \tilde{\varepsilon}_n + (\boldsymbol{U} - P\boldsymbol{U})^T\boldsymbol{\beta}_0 - \theta_0(U_{j_0} - PU_{j_0})$. The first term on the right-hand side (r.h.s.) of (S4.3) can be decomposed as

$$\mathbb{G}_n^* \hat{\varepsilon}_n [(U_{\hat{j}_n^*} - PU_{\hat{j}_n^*}) - (U_{j_0} - PU_{j_0})] + \mathbb{G}_n^*(\hat{\varepsilon}_n - \bar{\varepsilon}_n)(U_{j_0} - PU_{j_0}) + \mathbb{G}_n^* \bar{\varepsilon}_n (U_{j_0} - PU_{j_0}). \tag{S4.4}$$

In (S4.4), the first term is $o_{p^*}(1)$ a.s. because for any $\epsilon > 0$,

$$P^*(\mathbb{G}_n^* \hat{\varepsilon}_n [(U_{\hat{j}_n^*} - PU_{\hat{j}_n^*}) - (U_{j_0} - PU_{j_0})] > \epsilon) \leq P^*(\hat{j}_n^* \neq j_0) \to 0 \text{ a.s.}$$

The second term in (S4.4) can be reformatted as

$$\begin{aligned}
&(\mathbb{P}_n^* - \mathbb{P}_n)[(U_{j_0} - PU_{j_0})\boldsymbol{U}^T\boldsymbol{b}_0] - [\hat{\alpha}_n - (\alpha_0 + P\boldsymbol{U}^T\boldsymbol{\beta}_0 - \theta_0 PU_{j_0})]\mathbb{G}_n^*(U_{j_0} - PU_{j_0}) \\
&\quad - (\hat{\theta}_n - \theta_0)\mathbb{G}_n^* U_{j_0}(U_{j_0} - PU_{j_0}) + \hat{\theta}_n \mathbb{G}_n^*[(U_{j_0} - U_{\hat{j}_n})(U_{j_0} - PU_{j_0})] \\
&\quad + \mathbb{G}_n^*(U_{j_0} - PU_{j_0})(Y - \tilde{Y}).
\end{aligned} \tag{S4.5}$$

Because $E^*[\hat{G}_n^*(t)] = \hat{G}_n(t)$ for all $t \in \mathcal{T}$ (Lo (1993)), along with first order Taylor expanding with respect to $\hat{G}_n$, the last term in (S4.5) reduces to

$$\mathbb{P}_n^* \left[\frac{(U_{j_0} - PU_{j_0})\tilde{Y}\mathbb{L}_n^*(X)}{\hat{G}_n(X)} \right] + o_{p^*}(1) \text{ a.s.,}$$

where $\mathbb{L}_n^* \colon \mathcal{X} \mapsto \ell_\tau^\infty$ is a bootstrapped empirical process

$$\{\mathbb{G}_n^*[\phi_t(X)\gamma_0(X)(1-\delta) + \gamma_1(X,t)\delta - \gamma_2(X,t) - G_0(t)], t \in \mathcal{T}\}.$$

We use $\mathbb{L}_n^*$ to approximate $\{\sqrt{n}[\hat{G}_n^*(t) - \hat{G}_n(t)], t \in \mathcal{T}\}$ with ϕ_t, γ_0, γ_1 and γ_2 stated in Lemma 3.1. By the consistency of $(\hat{\alpha}_n, \hat{\theta}_n)$, bootstrap consistency of the sample mean and

$$P^*(\mathbb{G}_n^*[(U_{\hat{j}_n} - U_{j_0})(U_{j_0} - PU_{j_0})] > \epsilon) \leq 1(\hat{j}_n \neq j_0) \to 0 \text{ a.s.,}$$

equation (S4.5) reduces to

$$\mathbb{P}_n^* \left[\frac{(U_{j_0} - PU_{j_0})\tilde{Y}\mathbb{L}_n^*(X)}{\hat{G}_n(X)} \right] + o_{p^*}(1) \text{ in probability.} \tag{S4.6}$$

Parallel to $\mathbb{M}_{n,j} = \mathbb{G}_n \tilde{\varepsilon}_n (U_j - \mathbb{P}_n U_j)$ for $j = 1, \ldots, p$, let

$$\mathbb{M}^*_{n,j} = \mathbb{G}^*_n \tilde{\varepsilon}_n (U_j - PU_j). \tag{S4.7}$$

Since $\theta_0 = C^T_{j_0} \boldsymbol{\beta}_0 / V_{j_0}$ implying that $\bar{\varepsilon}_n = \tilde{\varepsilon}_n$, then we can express the remaining term in (S4.4) $\mathbb{G}^*_n \bar{\varepsilon}_n (U_{j_0} - PU_{j_0})$ as $\mathbb{M}^*_{n,j_0}$. By the definition of Ψ_j in (S3.1) and $EU_j = PU_j$ for all j along with the uniform consistency of $\hat{G}_n$, (S4.3)–(S4.6) lead to

$$\begin{aligned} \sqrt{n}(\hat{\theta}^*_n - \hat{\theta}_n) S^{*2}_{\hat{j}^*_n} &= \mathbb{M}^*_{n,j_0} + \mathbb{P}^*_n \left[\frac{(U_{j_0} - EU_{j_0}) \tilde{Y} \mathbb{L}^*_n(X)}{G_0(X)} \right] + o_{p^*}(1) \\ &= \Psi_{j_0}(\mathbb{M}^*_{n,j_0}, \mathbb{L}^*_n, \mathbb{P}^*_n) + o_{p^*}(1) \text{ in probability.} \end{aligned}$$

Note that $S^{*2}_{\hat{j}^*_n} \xrightarrow{p^*} V_{j_0}$ in probability. Together with bootstrap consistency of Kaplan–Meier estimator based on Efron's resampling plan (Efron (1981), Akritas (1986)), we obtain the desired result, using similar arguments for the proofs of Lemmas 3.3–3.5 and Theorem 3.6.1 of van der Vaart and Wellner (van der Vaart and Wellner (1996), Chap. 3). □

Lemma 4.3. *Suppose that (A.1)–(A.4) hold; that $\boldsymbol{\beta}_0 = \mathbf{0}$, and that $j(\boldsymbol{b}_0)$ is unique when $\boldsymbol{\beta}_0 = \mathbf{0}$. Then, $\mathbb{Q}^*_n(\boldsymbol{b}_0)$ converges to the limiting distribution of $\sqrt{n}(\hat{\theta}_n - \theta_n)$ conditionally (on the data) in probability.*

Proof. Following previous arguments, we can have

$$\sqrt{n}\hat{\theta}_j = (\mathbb{M}_{n,j} + \mathbb{D}_{n,j} + n^{-1} \sum_{i=1}^{n} (\boldsymbol{U}_{ij} - \bar{\boldsymbol{U}}_{\cdot j}) \boldsymbol{U}^T_i \boldsymbol{b}_0) / S^2_j, \tag{S4.8}$$

where

$$\begin{aligned} \mathbb{M}_{n,j} &= \mathbb{G}_n \tilde{\varepsilon}_n (U_j - \mathbb{P}_n U_j)\,; \\ \mathbb{D}_{n,j} &= \sqrt{n}\mathbb{P}_n (U_j - \mathbb{P}_n U_j)(Y - \tilde{Y}) = \mathbb{P}_n[(U_j - PU_j)\tilde{Y}\mathbb{L}_n(X)/G_0(X)] + o_p(1). \end{aligned}$$

According to the definition of Ψ_j, (S4.8) implies that

$$\sqrt{n}\hat{\theta}_j = \frac{\Psi_j(\mathbb{M}_{n,j}, \mathbb{L}_n, \mathbb{P}_n) + \widehat{\mathrm{Cov}}(U_j, \boldsymbol{U}^T \boldsymbol{b}_0)}{S^2_j}.$$

Let $\mathbb{M}_n$ be a p-dimensional vector with the j-th components given by $\mathbb{M}_{n,j}$. Let $\mathbb{J}_n(\boldsymbol{b})$ denote a

p-dimensional vector with the j-th component defined by

$$\mathbb{J}_{n,j}(\boldsymbol{b}) = (\Psi_j(\mathbb{M}_{n,j}, \mathbb{L}_n, \mathbb{P}_n) + \widehat{\text{Cov}}(U_j, \boldsymbol{U}^T\boldsymbol{b}))^2/S_j^2,$$

and $J(\boldsymbol{b})$ is a p-dimensional vector whose j-th component is $J_j(\boldsymbol{b}) = |\text{Corr}(U_j, \boldsymbol{U}^T\boldsymbol{b})|$. Moreover, we define a $p \times p$ matrix $\mathbb{A}_n(\boldsymbol{b})$ whose (j,k)-th component is provided by

$$(\Psi_j(\mathbb{M}_{n,j}, \mathbb{L}_n, \mathbb{P}_n) + \widehat{\text{Cov}}(U_j, \boldsymbol{U}^T\boldsymbol{b}))/S_j^2 - C_k/V_k.$$

In addition, let $\mathbb{H}_n(\boldsymbol{b})$ and $H(\boldsymbol{b})$ be p-dimensional vectors of zeros, except with a 1 at the entry that maximizes $\mathbb{J}_n(\boldsymbol{b})$ and $J(\boldsymbol{b})$, respectively. We can have that

$$\begin{aligned}\mathbb{Q}_n(\boldsymbol{b}) &= (\mathbb{M}_{n,J_n(\boldsymbol{b})} + \mathbb{D}_{n,J_n(\boldsymbol{b})} + \mathbb{P}_n(U_{J_n(\boldsymbol{b})} - \mathbb{P}_n U_{J_n(\boldsymbol{b})})\boldsymbol{U}^T\boldsymbol{b})/S_{J_n(\boldsymbol{b})}^2 - C_{j(\boldsymbol{b})}^T\boldsymbol{b}/V_{j(\boldsymbol{b})} \\ &= \mathbb{H}_n(\boldsymbol{b})^T\mathbb{A}_n(\boldsymbol{b})H(\boldsymbol{b}).\end{aligned}$$

We define $\mathbb{J}(\boldsymbol{b})$, $\mathbb{A}(\boldsymbol{b})$ and $\mathbb{H}(\boldsymbol{b})$ as processes (not indexed by n) with the same form as $\mathbb{J}_n(\boldsymbol{b})$, $\mathbb{A}_n(\boldsymbol{b})$ and $\mathbb{H}_n(\boldsymbol{b})$, except with $\mathbb{M}_{n,j}$ replaced by M_j; $\mathbb{L}_n$ replaced by $\mathbb{L}$; $\mathbb{P}_n$ replaced by P, and the sample variance or covariances replaced by their population versions. According to Theorem 1 (under (A.1)–(A.4)), it implies that when $\boldsymbol{\beta}_0 = \mathbf{0}$,

$$\sqrt{n}(\hat{\theta}_n - \theta_n) = \mathbb{Q}_n(\boldsymbol{b}_0) = \mathbb{H}_n(\boldsymbol{b}_0)^T\mathbb{A}_n(\boldsymbol{b}_0)H(\boldsymbol{b}_0) \xrightarrow{d} \mathbb{H}(\boldsymbol{b}_0)^T\mathbb{A}(\boldsymbol{b}_0)H(\boldsymbol{b}_0). \tag{S4.9}$$

Recall the bootstrap version of $\mathbb{M}_{n,j}$ defined in (S4.7) Let $\mathbb{A}_n^*(\boldsymbol{b})$ and $\mathbb{J}_n^*(\boldsymbol{b})$ denote the bootstrap versions of $\mathbb{A}_n(\boldsymbol{b})$ and $\mathbb{J}_n(\boldsymbol{b})$, respectively, where the (j,k)-th component of $\mathbb{A}_n^*(\boldsymbol{b})$ is given by

$$\frac{\Psi_j^*(\mathbb{M}_{n,j}^*, \mathbb{L}_n^*, \mathbb{P}_n^*) + \widehat{\text{Cov}}^*(U_j, \boldsymbol{U}^T\boldsymbol{b})}{S_j^{*2}} - \frac{\widehat{\text{Cov}}(U_k, \boldsymbol{U}^T\boldsymbol{b})}{S_k^2},$$

and the j-th component of $\mathbb{J}_n^*(\boldsymbol{b})$ is provided by

$$\mathbb{J}_{n,j}^*(\boldsymbol{b}) = [\Psi_j^*(\mathbb{M}_{n,j}^*, \mathbb{L}_n^*, \mathbb{P}_n^*) + \widehat{\text{Cov}}^*(U_j, \boldsymbol{U}^T\boldsymbol{b})]^2/S_j^{*2}.$$

The above display enables us to derive that, together with similar arguments used to close the proof of Lemma 4.2,

$$(\hat{H}_n(\boldsymbol{b}_0), \mathbb{A}_n^*(\boldsymbol{b}_0), \mathbb{J}_n^*(\boldsymbol{b}_0)) \xrightarrow{d} (H(\boldsymbol{b}_0), \mathbb{A}(\boldsymbol{b}_0), \mathbb{J}(\boldsymbol{b}_0))$$

conditionally (on the data) in probability, where $\hat{H}_n(\boldsymbol{b})$ denotes the sample version of $H(\boldsymbol{b})$. Moreover, we can observe that

$$\sqrt{n}\hat{\theta}_j^* = \frac{\Psi_j(\mathbb{M}_{n,j}^*, \mathbb{L}_n^*, \mathbb{P}_n^*) + \widehat{\text{Cov}}^*(U_j, \boldsymbol{U}^T\boldsymbol{b}_0)}{S_j^{*2}} + o_{p^*}(1) \text{ a.s., for all } j,$$

Hence, parallel arguments to obtain (S4.9) imply that

$$\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) = \mathbb{Q}_n^*(\boldsymbol{b}_0) = \mathbb{H}_n^*(\boldsymbol{b}_0)^T \mathbb{A}_n^*(\boldsymbol{b}_0)\hat{H}_n(\boldsymbol{b}_0) \xrightarrow{d} \mathbb{H}(\boldsymbol{b}_0)^T \mathbb{A}(\boldsymbol{b}_0) H(\boldsymbol{b}_0)$$

conditionally (on the data) in probability, where $\mathbb{H}_n^*(\boldsymbol{b})$ is a p-dimensional vector of zeros, except with a 1 at the entry that maximizes $\mathbb{J}_n^*(\boldsymbol{b})$. □

S5 Screening performance of ARTS

In this section we report the results of a simulation study to assess the screening performance of ARTS when applying the cut-off to all marginal regression estimates. That is, we conduct ARTS and obtain a pair of cut-off points $(\kappa_{\alpha/2}, \kappa_{(1-\alpha/2)})$, the $\alpha/2$ and $(1-\alpha/2)$ quantiles of the limiting distribution of $\sqrt{n}\hat{\theta}_n$, where $\alpha = 5\%$. We declare a predictor active if the point estimate of its slope parameter falls outside the interval $\mathcal{I}_\alpha \equiv [\kappa_{\alpha/2}/\sqrt{n}, \kappa_{(1-\alpha/2)}/\sqrt{n}]$.

We will assess screening performance in terms of false discovery rate (FDR), false negative rate (FNR) and false positive rate (FPR). Their empirical versions are given by

$$\widehat{FDR} = \frac{\#\{j : \hat{\beta}_j \notin \mathcal{I}_\alpha, \beta_j = 0\}}{\#\{j : \hat{\beta}_j \notin \mathcal{I}_\alpha\}}, \ \widehat{FNR} = \frac{\#\{j : \hat{\beta}_j \in \mathcal{I}_\alpha, \beta_j \neq 0\}}{\#\{j : \beta_j \neq 0\}} \text{ and } \widehat{FPR} = \frac{\#\{j : \hat{\beta}_j \notin \mathcal{I}_\alpha, \beta_j = 0\}}{\#\{j : \beta_j = 0\}},$$

respectively, where β_j is the marginal slope parameter based on U_j and has the point estimate $\hat{\beta}_j$ as defined in Lemma 3.6.

We generate 1000 samples of size $n = 100$ from Model (S1) given by $T = \sum_{j=1}^p \beta_j U_j + \varepsilon$, where $\beta_1 = \beta_2 = \beta_3 = 1.2$, $\beta_4 = \beta_5 = 0.8$, $\beta_6 = \beta_7 = \beta_8 = -0.8$, $\beta_9 = \beta_{10} = -0.5$ and $\beta_j = 0$, $j \geq 11$. The components of $\boldsymbol{U}$ and ε are independent standard normal. We also applied the Benjamini–Hochberg procedure (BH, Benjamini and Hochberg (1995)) and the Holm–Bonferroni procedure (HB, Holm (1979)) to the p-values based on marginal Z-tests of $\beta_j = 0$, $j = 1, \ldots, p$, using a nominal FDR or significance level of 5%. The results of the BONF-AFT procedure (Bonf) are also provided for comparison. The average values of $\widehat{FDR}$, $\widehat{FNR}$ and $\widehat{FPR}$ are provided in Table 1, along with their counterparts obtained from BH, HB and Bonf. See Section 6.2 of the main paper for discussion of the results.

References

Akritas, M. G. (1986). Bootstrapping the Kaplan-Meier estimator. *Journal of American Statistical Association*, **81**, 1032–1038.

$\widehat{FDR}$		$p=50$				$p=100$			
	Cens	ARTS	BH	HB	Bonf	ARTS	BH	HB	Bonf
	10%	0.029	0.037	0.018	0.018	0.043	0.050	0.024	0.024
	20%	0.045	0.034	0.025	0.026	0.043	0.043	0.028	0.028
	40%	0.043	0.032	0.035	0.034	0.040	0.044	0.041	0.040
$\widehat{FNR}$		$p=50$				$p=100$			
	Cens	ARTS	BH	HB	Bonf	ARTS	BH	HB	Bonf
	10%	0.793	0.711	0.809	0.812	0.843	0.780	0.849	0.849
	20%	0.859	0.800	0.877	0.877	0.907	0.854	0.906	0.906
	40%	0.884	0.825	0.876	0.877	0.920	0.873	0.907	0.907
$\widehat{FPR}$		$p=50$				$p=100$			
	Cens	ARTS	BH	HB	Bonf	ARTS	BH	HB	Bonf
	10%	0.004	0.004	0.001	0.001	0.003	0.002	0.001	0.001
	20%	0.005	0.003	0.001	0.001	0.002	0.001	0.000	0.000
	40%	0.005	0.003	0.001	0.001	0.003	0.001	0.001	0.001

Table 1: Average $\widehat{FDR}$, $\widehat{FNR}$ and $\widehat{FPR}$ based on 1000 samples of $n = 100$ generated from Model (S1) with 50 and 100 independent $\mathcal{N}(0, 1)$ predictors, for various censoring rates.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **57**, 289–300.

Efron, B. (1981). Censored data and the bootstrap. *Journal of American Statistical Association*, **76**, 312–319.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**, 65–70.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics. Springer, New York, NY, USA.

Lo, A. Y. (1993). A Bayesian bootstrap for censored data. *The Annals of Statistics*, **21**, 100–123.

McKeague, I. W. and Qian, M. (2015). An adaptive resampling test for detecting the presence of significant predictors (with discussion). *Journal of the American Statistical Association*, **110**, 1422–1433.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. NSF-CBMS Regional Conference Series in Probability and Statistics: Vol 2. Institute of Mathematical Statistics.

Stute, W. (1995). The central limit theorem under random censorship. *The Annals of Statistics*, **23**, 422–439.

Stute, W. and Wang, J.-L. (1993). The strong law under random censorship. *The Annals of Statistics*, **21**, 1591–1607.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer, New York, NY, USA.