Submitted to the Annals of Statistics

SUPPLEMENT TO "EFFICIENT ESTIMATION OF THE MAXIMAL ASSOCIATION BETWEEN MULTIPLE PREDICTORS AND A SURVIVAL OUTCOME"

BY TZU-JUNG HUANG[1,a], ALEX LUEDTKE[2,b] AND IAN W. MCKEAGUE[3,c]

[1]*Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center,* [a]*thuang2@fredhutch.org*

[2]*Department of Statistics, University of Washington,* [b]*aluedtke@uw.edu*

[3]*Department of Biostatistics, Columbia University,* [c]*im2131@columbia.edu*

CONTENTS

S1. Notation. For convenient reference, below we collect the notation that appears in the main text and will be frequently used in the upcoming proofs.

- M is the martingale part of the single-jump counting process for a censored observation, that is, $dM(s) = 1(X \in ds, \delta = 0) - 1(X \geq s)d\Lambda(s)$ with respect to the filtration

  (S1.1) $$\mathcal{F}_s \equiv \sigma(\{1(X \leq s', \delta = 0), 1(X \geq s'), U : s' \leq s \in \mathcal{T}\}),$$

  where Λ is the cumulative hazard function corresponding to G.

- Based on a fixed and user-defined coarsening c so that $c(U)$, $U \sim P$, is a finitely supported discrete random variable,

  $$M(u, ds) = 1(X \in ds, \delta = 0, c(U) = c(u)) - 1(X \geq s, c(U) = c(u))d\Lambda(s)$$

  is a martingale with respect to the filtration $\mathcal{F}_s$.

- Let $N_n(u, s) = \sum_{i=1}^n 1(X_i \leq s, \delta_i = 0, c(U_i) = c(u))$ and $Y_n(u, s) = \sum_{i=1}^n 1(X_i \geq s, c(U_i) = c(u))$. The conditional cumulative hazard function is estimated by

  $$\hat{\Lambda}_n(\cdot|u) = \int_{-\infty}^{\cdot} \frac{1(Y_n(u, s) > 0)}{Y_n(u, s)} N_n(u, ds).$$

  With $\Lambda(s)$ estimated by $\hat{\Lambda}_n(s|u)$, $\hat{M}(u, ds) = 1(X \in ds, \delta = 0, c(U) = c(u)) - 1(X \geq s, c(U) = c(u))d\hat{\Lambda}_n(s|u)$.

- $\bar{M}$ is the martingale part of the counting process for censored observations with the predictor U that is coarsened at $c(U) = c(u)$; namely, $\bar{M}(u, ds) = N_n(u, ds) - Y_n(u, s)d\Lambda(s)$ is a local square integrable martingale with respect to the aggregated filtration

  (S1.2) $$\bar{\mathcal{F}}_s \equiv \sigma(\{N_n(\cdot, s'), Y_n(\cdot, s'), U_i : i = 1, \ldots, n, s' \leq s \in \mathcal{T}\}).$$

- Starting from Section 3 in the main text and Section S5 in this supplementary, the distribution P is characterized by three features, namely $P = (E_0, Q_u, G)$. For any given k, let $U \equiv U_k$. Based on observations $\{O_1, \ldots, O_j\}$, $\hat{E}_j(u, s)$ is an estimator of $E_0(u, s) \equiv E[\tilde{Y}|U = u, X \geq s]$, and $\mathbb{Q}_j$ is the empirical distribution of any given single predictor whose distribution is Q_u, for $j = q_n, \ldots, n$. Therefore $\hat{P}_{nj} \equiv (\hat{E}_j, \mathbb{Q}_j, \hat{G}_n)$, and $\hat{P}'_j \equiv (\hat{E}_j, \mathbb{Q}_j, G)$, where $\hat{G}_n(\cdot|u)$ is the Kaplan–Meier estimator of the censoring distribution $G(\cdot)$ conditional on $c(U) = c(u)$. Note that $\hat{E}_j(u) = \hat{E}_j(u, \infty)$ to estimate $E[\tilde{Y}|U = u]$, and $\hat{G}_n(\cdot|u)$ reduces to a standard Kaplan–Meier estimator $\hat{G}_n(\cdot)$ when $c(U)$ is a degenerate random variable.
- For a function h mapping from a realization of O to $\mathbb{R}^d$, we use $P[h] \equiv P[h(O)] \equiv \int h(o)dP(o)$. The expectation $\mathbb{E}$ applies to a function of O and $O_1, \ldots, O_n$, regarding O as fixed, in contrast to the expectation under P that applies only to $O \sim P$ and not to any estimator based on $O_1, \ldots, O_n$. In addition, E denotes the expectation of both O and $O_1, \ldots, O_n$. To simplify the notation, moreover, we sometimes omit the subscript P from the functions evaluated under P unless otherwise stated, for instance, Var_P as Var and Cov_P as Cov. Moreover, we use $\mathrm{P}(A)$ to denote the probability of event A concerning the behaviors of the statistics based on O, and $\mathrm{P}(A_n)$ to denote the probability of event A_n concerning those of the statistics derived from $O_1, \ldots, O_n$.

S2. Canonical gradient of the slope parameter. The influence function of the sample Pearson correlation coefficient was first reported by Devlin, Gnanadesikan and Kettenring (1975), who attributed the result to C. L. Mallows. This influence function corresponds to the canonical gradient of the correlation coefficient in a locally nonparametric model. Here we use similar arguments to find the canonical gradient of the slope parameter Γ in this same model. As in Section 3 in the main text, here we focus on the case that there is only a single predictor.

Define a path $\{\tilde{P}_\epsilon = (1-\epsilon)\tilde{P} + \epsilon\delta_{(u,\tilde{y})}, \epsilon \in [0,1]\}$, where $\delta_{(u,\tilde{y})}$ is the Dirac measure putting unit mass at $(u, \tilde{y})$. Then

$$\Gamma(\tilde{P}_\epsilon) = \frac{\mathrm{Cov}_{\tilde{P}_\epsilon}(U, \tilde{Y})}{\mathrm{Var}_{\tilde{P}_\epsilon}(U)},$$

and

$$\begin{aligned}
&\mathrm{Cov}_{\tilde{P}_\epsilon}(U, \tilde{Y}) = \tilde{P}_\epsilon[U\tilde{Y}] - \tilde{P}_\epsilon[U]\tilde{P}_\epsilon[\tilde{Y}] \\
&\quad = (1-\epsilon)\tilde{P}[U\tilde{Y}] + \epsilon[u\tilde{y}] - (1-\epsilon)^2\tilde{P}[U]\tilde{P}[\tilde{Y}] - \epsilon(1-\epsilon)\tilde{P}[U]\tilde{y} - \epsilon(1-\epsilon)\tilde{P}[\tilde{Y}]u - \epsilon^2 u\tilde{y}; \\
&\mathrm{Var}_{\tilde{P}_\epsilon}(U) = \tilde{P}_\epsilon[U^2] - (\tilde{P}_\epsilon[U])^2 = (1-\epsilon)\tilde{P}[U^2] + \epsilon u^2 - (1-\epsilon)^2(\tilde{P}[U])^2 \\
&\qquad\qquad\qquad - 2\epsilon(1-\epsilon)u\tilde{P}[U] - \epsilon^2 u^2,
\end{aligned}$$

where $\tilde{y} = \delta x/G(x)$. Let $D(\tilde{P})$ denote the canonical gradient of Γ at $\tilde{P}$ in a locally nonparametric model. The evaluation of this function at the chosen $(u, \tilde{y})$ is equal to $\frac{d}{d\epsilon}\Gamma(\tilde{P}_\epsilon)\Big|_{\epsilon=0}$, and so

$$D(\tilde{P})(u, \tilde{y}) = \left\{ \frac{(u - \tilde{P}[U])(\tilde{y} - \tilde{P}[\tilde{Y}])}{\mathrm{Var}_{\tilde{P}}(U)} - \frac{\mathrm{Cov}_{\tilde{P}}(U, \tilde{Y})}{\mathrm{Var}^2_{\tilde{P}}(U)}(u - \tilde{P}[U])^2 \right\}.$$

Straightforward calculations show that, for distributions $\tilde{P}_1$ and $\tilde{P}$ for the random variable $(U, \tilde{Y})$, it holds that

$$\int D(\tilde{P})(u,\tilde{y})\, d\tilde{P}_1(u,\tilde{y}) = \Bigg(\frac{\mathrm{Cov}_{\tilde{P}_1}(U,\tilde{Y}) + (\tilde{P}_1[U] - \tilde{P}[U])(\tilde{P}_1[\tilde{Y}] - \tilde{P}[\tilde{Y}])}{\mathrm{Var}_{\tilde{P}}(U)} \\ - \frac{\mathrm{Cov}_{\tilde{P}}(U,\tilde{Y})}{\mathrm{Var}^2_{\tilde{P}}(U)} \left[\mathrm{Var}_{\tilde{P}_1}(U) + (\tilde{P}_1[U] - \tilde{P}[U])^2 \right] \Bigg).$$

Recall that $\tilde{P} \equiv P \circ f_G^{-1}$ for the distribution P that generated each of the variates $O_1, \ldots, O_n$. Suppose also that $\tilde{P}_1 = P_1 \circ f_G^{-1}$ for a distribution P_1 of the random variable O. The above then shows that

$$\mathrm{Rem}_G(P_1, P) \equiv \Gamma(\tilde{P}) - \Gamma(\tilde{P}_1) + \int D(\tilde{P})(u,\tilde{y})\, d\tilde{P}_1(u,\tilde{y})$$

$$= \Bigg\{ \frac{\mathrm{Cov}_{\tilde{P}}(U,\tilde{Y})}{\mathrm{Var}_{\tilde{P}}(U)} - \frac{\mathrm{Cov}_{\tilde{P}_1}(U,\tilde{Y})}{\mathrm{Var}_{\tilde{P}_1}(U)} + \frac{(\tilde{P}_1[U] - \tilde{P}[U])(\tilde{P}_1[\tilde{Y}] - \tilde{P}[\tilde{Y}])}{\mathrm{Var}_{\tilde{P}}(U)} \\ + \frac{\mathrm{Cov}_{\tilde{P}_1}(U,\tilde{Y})}{\mathrm{Var}_{\tilde{P}}(U)} - \frac{\mathrm{Cov}_{\tilde{P}}(U,\tilde{Y})}{\mathrm{Var}^2_{\tilde{P}}(U)} [\mathrm{Var}_{\tilde{P}_1}(U) + (\tilde{P}_1[U] - \tilde{P}[U])^2] \Bigg\}$$

$$= \Bigg\{ \left[\frac{\mathrm{Cov}_{\tilde{P}}(U,\tilde{Y})}{\mathrm{Var}^2_{\tilde{P}}(U)} - \frac{\mathrm{Cov}_{\tilde{P}_1}(U,\tilde{Y})}{\mathrm{Var}_{\tilde{P}_1}(U)\mathrm{Var}_{\tilde{P}}(U)} \right] (\mathrm{Var}_{\tilde{P}}(U) - \mathrm{Var}_{\tilde{P}_1}(U)) \\ - \frac{\mathrm{Cov}_{\tilde{P}}(U,\tilde{Y})}{\mathrm{Var}^2_{\tilde{P}}(U)} (\tilde{P}_1[U] - \tilde{P}[U])^2 + \frac{1}{\mathrm{Var}_{\tilde{P}}(U)} (\tilde{P}_1[U] - \tilde{P}[U])(\tilde{P}_1[\tilde{Y}] - \tilde{P}[\tilde{Y}]) \Bigg\}.$$

Note that the remainder term in the linear expansion (S3.3) is equal to $-\mathrm{Rem}_G(\mathbb{P}_n, P)$. The following lemma shows that this term is asymptotically negligible.

LEMMA S2.1. *Under (A.1) (so that* $\mathrm{Var}_{\tilde{P}}(U) > 0$*), we have* $\mathrm{Rem}_G(\mathbb{P}_n, P) = o_p(n^{-1/2})$.

The proof uses the CLT, the weak law of large numbers and the triangle inequality:

$$\sqrt{n}|\mathrm{Rem}_G(\mathbb{P}_n, P)| \\ \le [\mathrm{Var}_{\tilde{\mathbb{P}}_n}(U) \wedge \mathrm{Var}_{\tilde{P}}(U)]^{-2} \left\{ \left| \sqrt{n}(\mathrm{Cov}_{\tilde{\mathbb{P}}_n}(U,\tilde{Y}) - \mathrm{Cov}_{\tilde{P}}(U,\tilde{Y})) \right| |\mathrm{Var}_{\tilde{\mathbb{P}}_n}(U) - \mathrm{Var}_{\tilde{P}}(U)| \\ + \left| \mathrm{Cov}_{\tilde{P}}(U,\tilde{Y}) \right| \sqrt{n}(\tilde{\mathbb{P}}_n[U] - \tilde{P}[U])^2 + \mathrm{Var}_{\tilde{P}}(U) \left| \sqrt{n}(\tilde{\mathbb{P}}_n[U] - \tilde{P}[U])(\tilde{\mathbb{P}}_n[\tilde{Y}] - \tilde{P}[\tilde{Y}]) \right| \right\}.$$

S3. Slope estimation with a single predictor. Restricting attention to the case of a single predictor U_k, in this section we supplement two other estimators that only works in ad hoc scenarios stated in Sections S3.1 and S3.2, and delineate the framework that enables the development of an asymptotically efficient one-step estimator in Appendix A. For notational simplicity we suppress the subscript k, and just write $\Psi(P)$ and U, both here and in the corresponding proofs in the sequel.

S3.1. *Inefficient estimation with known censoring distribution G.* Let $G(t) \equiv \mathrm{P}(C \ge t)$ be the unknown survival function of C, where throughout we assume $G(\tau) > 0$. For a given survival function G_1, let

$$\Psi_{G_1}(P) \equiv \frac{\mathrm{Cov}_P(U, \delta X / G_1(X))}{\mathrm{Var}_P(U)}.$$

Also let $\tilde{Y} \equiv \delta X/G(X)$. Noting that $E[T] = E[\tilde{Y}]$ and $E[UT] = E[U\tilde{Y}]$, we obtain the useful identity that $\Psi(P) = \Psi_G(P)$. This suggests that, if G is known, then it suffices to try to estimate the value of the inverse-probability-of-censoring weighted parameter Ψ_G evaluated at the distribution that generated the observed data.

To study the parameter Ψ_G, it will be useful to use existing arguments from the uncensored case. To do this, we define $f_G(o) \equiv (u, \tilde{y})$, and then let $\tilde{P}$ denote the pushforward measure $\tilde{P} \equiv P \circ f_G^{-1}$, that is, the distribution of $f_G(O)$ when $O \sim P$. Writing

$$\Gamma(\tilde{P}) \equiv \frac{\mathrm{Cov}_{\tilde{P}}(U, \tilde{Y})}{\mathrm{Var}_{\tilde{P}}(U)},$$

we then see that the parameter $P \mapsto \Gamma(P \circ f_G^{-1})$ is the same as the parameter $P \mapsto \Psi_G(P)$. Therefore, a first-order expansion of the parameter Γ will yield a first-order expansion of the parameter Ψ_G. Noting that Γ is simply the slope parameter in a linear regression in which both the predictor and the outcome are observed, we can develop an expansion of this parameter using an existing result in Devlin, Gnanadesikan and Kettenring (1975), who attributed the result to C. L. Mallows. Let

$$\mathrm{IF}_G^{ipw}(\cdot|P) \colon o \mapsto \frac{(u - P[U])(\tilde{y} - P[T])}{\mathrm{Var}_P(U)} - \frac{\mathrm{Cov}_P(U, T)}{\mathrm{Var}_P^2(U)}(u - P[U])^2,$$

where we note that $\tilde{y}$ is a function of (x, δ) and therefore of o, and the subscript G indicates the implicit dependence of $\tilde{y}$ on G. Using $\tilde{\mathbb{P}}_n$ to denote the empirical distribution of $(U, \tilde{Y})$ when G is known, Mallows' result implies that, for a term $\mathrm{Rem}_G(\mathbb{P}_n, P)$ that is defined in Section S2,

(S3.3)

$$\begin{aligned}
&\Gamma(\tilde{\mathbb{P}}_n) - \Gamma(\tilde{P}) \\
&= (\tilde{\mathbb{P}}_n - \tilde{P}) \left[\left\{ \frac{(U - \tilde{P}[U])(\tilde{Y} - \tilde{P}[\tilde{Y}])}{\mathrm{Var}_{\tilde{P}}(U)} - \frac{\mathrm{Cov}_{\tilde{P}}(U, \tilde{Y})}{\mathrm{Var}_{\tilde{P}}^2(U)}(U - \tilde{P}[U])^2 \right\} \right] + \mathrm{Rem}_G(\mathbb{P}_n, P) \\
&= (\mathbb{P}_n - P) \left[\left\{ \frac{(U - P[U])(\tilde{Y} - P[T])}{\mathrm{Var}_P(U)} - \frac{\mathrm{Cov}_P(U, T)}{\mathrm{Var}_P^2(U)}(U - P[U])^2 \right\} \right] + \mathrm{Rem}_G(\mathbb{P}_n, P) \\
&= (\mathbb{P}_n - P) \left[\mathrm{IF}_G^{ipw}(O|P) \right] + \mathrm{Rem}_G(\mathbb{P}_n, P) \\
&= \mathbb{P}_n \left[\mathrm{IF}_G^{ipw}(O|P) \right] + \mathrm{Rem}_G(\mathbb{P}_n, P),
\end{aligned}$$

where the second equality follows from $P[T] = \tilde{P}[\tilde{Y}]$, $\mathrm{Cov}_P(U, T) = \mathrm{Cov}_{\tilde{P}}(U, \tilde{Y})$ and the fact that U has the same marginal distribution under P and $\tilde{P}$, and the last equality follows from $P\big[\mathrm{IF}_G^{ipw}(O|P)\big] = 0$. For more details of the derivation of IF_G^{ipw}, see Section S2. For concise notations, we hereafter omit the subscript P from the functions evaluated at P unless otherwise stated, which simplifies the presentation of the influence function to

$$\mathrm{IF}_G^{ipw}(\cdot|P) \colon o \mapsto \frac{(u - P[U])(\tilde{y} - P[T])}{\mathrm{Var}(U)} - \frac{\mathrm{Cov}(U, T)}{\mathrm{Var}^2(U)}(u - P[U])^2. \tag{S3.4}$$

Noting that $\tilde{\mathbb{P}}_n = \mathbb{P}_n \circ f_G^{-1}$ and recalling that the parameter $P \mapsto \Gamma(P \circ f_G^{-1})$ is the same as the parameter $P \mapsto \Psi_G(P)$, (S3.3) yields that

$$\Psi_G(\mathbb{P}_n) - \Psi_G(P) = \mathbb{P}_n\big[\mathrm{IF}_G^{ipw}(O|P)\big] + \mathrm{Rem}_G(\mathbb{P}_n, P). \tag{S3.5}$$

S3.2. *Efficiency gains through estimation of G.* We now consider the effect on the inverse-probability-weighted estimator $\Psi_G(\mathbb{P}_n)$ of replacing G by its Kaplan–Meier (K–M) estimator $\hat{G}_n$. This involves replacing $\tilde{Y}$ by the estimated (and observed) synthetic response $Y = \delta X/\hat{G}_n(X)$, and results in the following estimator of $\Psi(P)$ that now just depends on the empirical distribution $\mathbb{P}_n$ of the observed data:

$$\Psi_{\hat{G}_n}(\mathbb{P}_n) = \mathrm{Cov}_{\mathbb{P}_n}(U,Y)/\mathrm{Var}_{\mathbb{P}_n}(U).$$

Importantly, unlike for the estimator presented in Section S3.1, constructing this estimator does not rely on having knowledge of G.

Under the following conditions, Theorem S3.1 below shows that $\Psi_{\hat{G}_n}(\mathbb{P}_n)$ is asymptotically linear and identifies its influence function. The boundedness part of the first condition is stronger than we need, but is made to simplify the proof.

(C.1) The predictor U has bounded support and is non-degenerate.
(C.2) The survival function of the censoring, G, is continuous and $G(\tau) > 0$.
(C.3) There is a positive probability of a subject still being at risk at the end of follow-up: $\mathrm{P}(X \geq \tau) > 0$.

THEOREM S3.1. *Given (C.1)–(C.3),*

$$\Psi_{\hat{G}_n}(\mathbb{P}_n) - \Psi(P) = \mathbb{P}_n\,\mathrm{IF}(O|P) + o_p(n^{-1/2}),$$

where

$$\mathrm{IF}(O|P) \equiv \mathrm{IF}_G^{ipw}(O|P) - \mathrm{IF}_G^{nu}(O|P),$$

the influence function $\mathrm{IF}_G^{ipw}(\,\cdot\,|P)$ *is given in* (S3.4), *and*

$$\mathrm{IF}_G^{nu}(\cdot|P)\colon o \mapsto \frac{1}{\mathrm{Var}(U)} P\left[(u - P[U]) \int_{\mathcal{T}} E[T|U=u, X \geq s]\frac{1(x \in ds, \delta = 0) - 1(x \geq s)d\Lambda(s)}{G(s)}\right]$$

in which Λ is the cumulative hazard function corresponding to G.

Since $P[\mathrm{IF}(\cdot|P)] = 0$, Theorem S3.1 implies that $n^{1/2}[\Psi_{\hat{G}_n}(\mathbb{P}_n) - \Psi(P)] \xrightarrow{d} N(0,\sigma^2)$, where $\sigma^2 = P[\mathrm{IF}^2(\cdot|P)]$. The proof of Theorem S3.1 and the relevant Lemmas S4.1-S4.4 are deferred to Section S4. In particular, Lemma S4.4 shows that IF_G^{nu} is the projection of IF_G^{ipw} onto the nuisance tangent space

$$\mathbf{T}^{nu}(G) = \left\{ \int_{\mathcal{T}} H(s)\, dM(s) \,\middle|\, H\colon \mathcal{T} \to \mathbb{R} \right\}, \tag{S3.6}$$

where H is any measurable function for which the integral has finite variance, M is the martingale part of the single-jump counting process for a censored observation: $dM(s) = 1(X \in ds, \delta = 0) - 1(X \geq s)d\Lambda(s)$ and the filtration is $\mathcal{F}_s = \sigma\{1(X \leq s', \delta = 0), 1(X \geq s'), U,\, s' \leq s \in \mathcal{T}\}$.

REMARK S3.2. Note that $\mathrm{IF}^{car}(\cdot|P)$ in (12) is similar to the earlier definition of $\mathrm{IF}^{nu}(\cdot|P)$, but differs in that the marginal expectation over P is removed, which can lead to an efficiency gain. Indeed, unlike the influence functions for the estimators in the previous two subsections, the above function applied to an observation O_i will make use of the (always) observed U_i, regardless of whether that individual's event time was right-censored, both to estimate the U portion of the covariance between U and T and to evaluate the conditional residual life function on the actually observed U_i, rather than on an "average" U. This can be derived by noting that, under a coarsening at random model, the nuisance tangent space consists of all functions $f \in L^2(P)$ satisfying $E[f(O)|T,U] = 0$ almost surely, where the expectation is over the full data distribution whose coarsening gave rise to P (Theorem 1.3 in van der Laan and Robins, 2003).

S4. Proof of Theorem S3.1. This theorem gives that $\Psi_{\hat{G}_n}(\mathbb{P}_n)$ is asymptotically linear with its influence function $\mathrm{IF}_G^{ipw}(\cdot|P) - \mathrm{IF}_G^{nu}(\cdot|P)$, where $\mathrm{IF}_G^{ipw}(\cdot|P)$ is the influence function of an inverse probability weighted estimator in the setting where G is known, and $\mathrm{IF}_G^{nu}(\cdot|P)$ accounts for the fact that, in truth, G was estimated by the Kaplan–Meier estimator $\hat{G}_n$. Below we follow Lemmas S4.1-S4.3 to validate the substitution of G for $\hat{G}_n$ in the asymptotically linear representation of $\Psi_{\hat{G}_n}(\mathbb{P}_n) - \Psi(P)$, while Lemma S4.4 further indicates that $\mathrm{IF}_G^{nu}(\cdot|P)$ is the projection of $\mathrm{IF}_G^{ipw}(\cdot|P)$ onto the nuisance tangent space $\mathbf{T}^{nu}(G)$.

To obtain the asymptotic linear representation of $\Psi_{\hat{G}_n}(\mathbb{P}_n) - \Psi(P)$, we first have that

$$\begin{aligned}
\Psi_{\hat{G}_n}(\mathbb{P}_n) - \Psi(P) &= [\Psi_{\hat{G}_n}(\mathbb{P}_n) - \Psi_{\hat{G}_n}(P)] + [\Psi_{\hat{G}_n}(P) - \Psi(P)] \\
&= \mathbb{P}_n\big[\,\mathrm{IF}_{\hat{G}_n}^{ipw}(O|P)\big] + [\Psi_{\hat{G}_n}(P) - \Psi(P)] - \mathrm{Rem}_{\hat{G}_n}(\mathbb{P}_n, P) \\
&= \mathbb{P}_n\big[\,\mathrm{IF}_{G}^{ipw}(O|P)\big] + [\Psi_{\hat{G}_n}(P) - \Psi(P)] - \mathrm{Rem}_{\hat{G}_n}(\mathbb{P}_n, P) + o_p(n^{-1/2}),
\end{aligned} \tag{S4.7}$$

where the second equality follows by replacing G by $\hat{G}_n$ in (S3.5), leading to

$$\begin{aligned}
&\Psi_{\hat{G}_n}(\mathbb{P}_n) - \Psi_{\hat{G}_n}(P) \\
&= \mathbb{P}_n\left[\left\{\frac{(U-P[U])(Y-P[T])}{\mathrm{Var}(U)} - \frac{\mathrm{Cov}(U,T)}{\mathrm{Var}^2(U)}(U-P[U])^2\right\}\right] - \mathrm{Rem}_{\hat{G}_n}(\mathbb{P}_n, P) \\
&\equiv \mathbb{P}_n\big[\,\mathrm{IF}_{\hat{G}_n}^{ipw}(O|P)\big] - \mathrm{Rem}_{\hat{G}_n}(\mathbb{P}_n, P),
\end{aligned}$$

and the last equality in (S4.7) holds by the asymptotic equivalence between $\mathbb{P}_n[\mathrm{IF}_{\hat{G}_n}^{ipw}(O|P)]$ and $\mathbb{P}_n[\mathrm{IF}_{G}^{ipw}(O|P)]$ that is a consequence of Lemma 19.24 in van der Vaart (1998) and stated in Lemma S4.2. Moreover, $\mathrm{Rem}_{\hat{G}_n}(\mathbb{P}_n, P)$ is shown asymptotically negligible in Lemma S4.1.

In addition, a first-order Taylor expansion of the middle term on the right-hand side of (S4.7) around G yields the approximation

(S4.8)

$$\begin{aligned}
\Psi_{\hat{G}_n}(P) - \Psi(P) &= \frac{-1}{\mathrm{Var}(U)} P\left[(U-P[U])\frac{\delta X}{G^2(X)}[\hat{G}_n(X) - G(X)]\right] + o_p(n^{-1/2}) \\
&= \frac{-1}{\mathrm{Var}(U)} P\left[(U-P[U])\tilde{Y}\left[\frac{\hat{G}_n(X)}{G(X)} - 1\right]\right] + o_p(n^{-1/2}),
\end{aligned}$$

which can be further expressed in terms of the influence function of $\hat{G}_n$. Lemma S4.3 shows that the influence function of $\hat{G}_n(t)$ is $-G(t)\,\mathrm{IF}_G(O)(t)$, where

$$\mathrm{IF}_G(O)(t) = \int_{-\infty}^{t}\left[\frac{1(X \in ds, \delta = 0)}{P1(X \geq s)} - \frac{1(X \geq s)d\Lambda(s)}{P1(X \geq s)}\right] = \int_{-\infty}^{t}\frac{dM(s)}{P1(X \geq s)} \tag{S4.9}$$

for $t \in \mathcal{T}$; Λ is the cumulative hazard function of G, and $dM(s) \equiv 1(X \in ds, \delta = 0) - 1(X \geq s)d\Lambda(s)$ is a martingale with respect to the filtration generated by the counting process $1(X \leq s, \delta = 0)$. This suggests that $\hat{G}_n(t) - G(t) = -G(t)[n^{-1}\sum_{i=1}^{n}\mathrm{IF}_G(O_i)(t)] + o_p(n^{-1/2})$, which further implies that

$$\frac{\hat{G}_n(t)}{G(t)} - 1 = \frac{-1}{n}\sum_{i=1}^{n}\int_{-\infty}^{t}\frac{dM_i(s)}{P1(X \geq s)} = -\mathbb{P}_n\,\mathrm{IF}_G(O)(t)$$

with dM_i is the replication of dM based on the observation of the i-th subject. Let $\mathbb{L}_n \equiv \mathbb{P}_n\,\mathrm{IF}_G(O)$ and define $\varphi\colon \ell_\tau^\infty \to \mathbb{R}$ by $\varphi(g) = P[(U-P[U])\tilde{Y}g(X)]$, where ℓ_τ^∞ is the space of uniformly bounded functions on $\mathcal{T}$. Then (S4.8) can be expressed in terms of $\mathbb{L}_n$ and φ as follows:

$$\Psi_{\hat{G}_n}(P) - \Psi(P) = \frac{-1}{\mathrm{Var}(U)}\varphi(\mathbb{L}_n) + o_p(n^{-1/2}) = \mathbb{P}_n\left[\frac{-1}{\mathrm{Var}(U)}\varphi(\mathrm{IF}_G(O))\right] + o_p(n^{-1/2}),$$

where the second equality holds by the linearity of $\varphi(g)$ in g.

Evaluating φ at $\mathrm{IF}_G(O)$ in (S4.9), this leads to

$$\begin{aligned}\varphi\left(\mathrm{IF}_G(O)\right) &= P\left[(U - P[U])\tilde{Y}\int_{-\infty}^{X}\frac{dM(s)}{P1(X \geq s)}\right] \\ &= P\left[(U - P[U])\int_{\mathcal{T}} P\left[\tilde{Y}|U, X \geq s\right] P1(X \geq s)\frac{dM(s)}{P1(X \geq s)}\right],\end{aligned}$$

in which

$$\begin{aligned}P\left[\tilde{Y}|U, X \geq s\right] &= \int_{\mathcal{T}}\frac{x}{G(x)}P1(\delta = 1, X \in dx|U, X \geq s) \\ &= \int_{\mathcal{T}}\frac{x1(x \geq s)}{G(x)}\frac{P1(U, C \geq x, T \in dx)}{P1(U, C \geq s, T \geq s)} = \int_{\mathcal{T}}\frac{x1(x \geq s)}{G(x)}\frac{G(x)P1(U, C \geq s, T \in dx)}{G(s)P1(U, C \geq s, T \geq s)} \\ &= \int_{\mathcal{T}}\frac{x}{G(s)}\frac{P1(U, X \geq s, T \in dx)}{P1(U, X \geq s)} = \frac{E[T|U, X \geq s]}{G(s)}.\end{aligned}$$

Therefore we have that

$$\varphi\left(\mathrm{IF}_G(O)\right) = P\left[(U - P[U])\int_{\mathcal{T}} E[T|U, X \geq s]\frac{dM(s)}{G(s)}\right],$$

where the expectation with respect to P only applies to (U, X), not to M, and it implies that the expression in (S4.7) becomes

(S4.10)

$$\begin{aligned}&\Psi_{\hat{G}_n}(\mathbb{P}_n) - \Psi(P) = \mathbb{P}_n\left[\mathrm{IF}_G^{ipw}(O|P)\right] \\ &- \mathbb{P}_n\left[\frac{1}{\mathrm{Var}(U)}P\left[(U - P[U])\int_{\mathcal{T}} E[T|U, X \geq s]\frac{dM(s)}{G(s)}\right]\right] - \mathrm{Rem}_{\hat{G}_n}(\mathbb{P}_n, P) + o_p(n^{-1/2}) \\ &\equiv \mathbb{P}_n\left[\mathrm{IF}_G^{ipw}(O|P)\right] - \mathbb{P}_n\left[\mathrm{IF}_G^{nu}(O|P)\right] - \mathrm{Rem}_{\hat{G}_n}(\mathbb{P}_n, P) + o_p(n^{-1/2}).\end{aligned}$$

LEMMA S4.1. *If (C.1) and (C.2) hold, then* $\mathrm{Rem}_{\hat{G}_n}(\mathbb{P}_n, P) = o_p(n^{-1/2})$.

PROOF. Similar to Lemma S2.1, by the triangle inequality and the equivalence between $\mathbb{P}_n$ and $\tilde{\mathbb{P}}_n$ as they operate on $(U, \tilde{Y})$ marginally or jointly,

$$\begin{aligned}\sqrt{n}|\mathrm{Rem}_{\hat{G}_n}(\mathbb{P}_n, P)| &\leq [\mathrm{Var}_{\mathbb{P}_n}(U) \wedge \mathrm{Var}_{\tilde{P}}(U)]^{-2}\Big\{\left|\mathrm{Cov}_{\mathbb{P}_n}(U, Y) - \mathrm{Cov}_{\mathbb{P}_n}(U, \tilde{Y})\right| \\ &+ \left|\mathrm{Cov}_{\mathbb{P}_n}(U, \tilde{Y}) - \mathrm{Cov}_{\tilde{P}}(U, \tilde{Y})\right|\Big\}\left|\sqrt{n}(\mathrm{Var}_{\mathbb{P}_n}(U) - \mathrm{Var}_{\tilde{P}}(U))\right| \\ &+ \left|\mathrm{Cov}_{\tilde{P}}(U, \tilde{Y})\right|\sqrt{n}\left\{(\mathbb{P}_n - \tilde{P})[U]\right\}^2 \\ &+ \mathrm{Var}_{\tilde{P}}(U)\left|\sqrt{n}(\mathbb{P}_n - \tilde{P})[U]\right|\Big\{\left|\mathbb{P}_n[Y] - \mathbb{P}_n[\tilde{Y}]\right| + \left|(\mathbb{P}_n - \tilde{P})[\tilde{Y}]\right|\Big\}.\end{aligned}$$

Each term above is now shown to be $o_p(1)$. For the first term, note that $\mathrm{Var}_{\tilde{P}}(U) > 0$, $\mathrm{Var}_{\mathbb{P}_n}(U) \xrightarrow{p} \mathrm{Var}_{\tilde{P}}(U) > 0$, $\mathrm{Cov}_{\mathbb{P}_n}(U, Y) - \mathrm{Cov}_{\mathbb{P}_n}(U, \tilde{Y}) = o_p(1)$ by (C.2) and the uniform consistency of $\hat{G}_n$ to G, the convergence of $|\mathrm{Cov}_{\mathbb{P}_n}(U,\tilde{Y}) - \mathrm{Cov}_{\tilde{P}}(U,\tilde{Y})|$ to zero in probability by the weak law of large numbers, and $\sqrt{n}(\mathrm{Var}_{\mathbb{P}_n}(U) - \mathrm{Var}_{\tilde{P}}(U)) = O_p(1)$ by the CLT. The second term is $o_p(1)$ because $\sqrt{n}(\mathbb{P}_n - \tilde{P})[U] = O_p(1)$ and $(\mathbb{P}_n - \tilde{P})[U] \xrightarrow{p} 0$. Similarly, the last term is $o_p(1)$, which is a consequence of $\sqrt{n}(\mathbb{P}_n - \tilde{P})[U] = O_p(1)$, $\mathbb{P}_n[\tilde{Y}] \xrightarrow{p} \tilde{P}[\tilde{Y}]$, and $\mathbb{P}_n[Y] - \mathbb{P}_n[\tilde{Y}] = o_p(1)$ that follows by (C.2) and the uniform consistency of $\hat{G}_n$ as an estimator of G □

LEMMA S4.2. *Suppose (C.1) and (C.2) hold, so U has a finite fourth moment and* $\mathrm{Var}(U) > 0$. *Then*

$$\mathbb{P}_n\big[\mathrm{IF}^{ipw}_{\hat{G}_n}(O|P)\big] = \mathbb{P}_n\big[\mathrm{IF}^{ipw}_{G}(O|P)\big] + o_p(n^{-1/2}).$$

PROOF. Following (C.2), fix an $\tilde{\epsilon}$ such that $0 < \tilde{\epsilon} < G(\tau)$. Let $\mathcal{G}$ be a collection of monotone nonincreasing càdlàg functions $\tilde{G} : \mathcal{T} \to [0,1]$ such that $\tilde{G}(\tau) > \tilde{\epsilon}$, which is seen to be P-Donsker (Example 19.11 in van der Vaart, 1998). The influence function $\psi(\tilde{G}) = \mathrm{IF}^{ipw}_{\tilde{G}}(\cdot|P)$, and it is continuous of $\tilde{G}$ because $\tilde{G}(\tau) > \tilde{\epsilon} > 0$. We further have that $\|\psi(\tilde{G})\|_{P,2} < \infty$ because X is bounded by τ; $\mathrm{Var}(U) > 0$ and the finite fourth moment condition of U given in (C.1) and $\tilde{G}(\tau) > 0$, implying that the class $\psi(\mathcal{G})$ is also P-Donsker by the Donsker preservation properties (Section 9.4 in Kosorok, 2008). Since $\psi(G) \in \psi(\mathcal{G})$, $\psi(\hat{G}_n) \in \psi(\mathcal{G})$ and the uniform consistency of Kaplan–Meier estimator implies that

$$\int (\psi(\hat{G}_n) - \psi(G))^2 dP \xrightarrow{p} 0,$$

it follows that $\sqrt{n}(\mathbb{P}_n - P)(\psi(\hat{G}_n) - \psi(G)) \xrightarrow{p} 0$ by Lemma 19.24 of van der Vaart (1998). Equivalently we have that

$$\mathbb{P}_n\big[\mathrm{IF}^{ipw}_{\hat{G}_n}(O|P)\big] = \mathbb{P}_n\big[\mathrm{IF}^{ipw}_{G}(O|P)\big] + o_p(n^{-1/2})$$

because $P[\psi(G)] = P\big[\mathrm{IF}^{ipw}_{G}(O|P)\big] = 0$ and $P[\psi(\hat{G}_n)] = P\big[\mathrm{IF}^{ipw}_{\hat{G}_n}(O|P)\big] = 0$, the latter implied by the expectation with respect to P not operating on $\hat{G}_n$. □

Standard results concerning the Kaplan–Meier estimator give

LEMMA S4.3. *Given (C.3) and for $t \in \mathcal{T}$, the influence function of $\hat{G}_n(t)$ is*

$$\begin{aligned}\mathrm{IF}_G(O|t) &= -G(t)\int_{-\infty}^{t}\left[\frac{1(X \in ds, \delta = 0)}{P1(X \geq s)} - \frac{1(X \geq s)d\Lambda(s)}{P1(X \geq s)}\right] \\ &= -G(t)\int_{-\infty}^{t}\frac{dM(s)}{P1(X \geq s)}.\end{aligned}$$

LEMMA S4.4. *Given (C.1)–(C.3), the function $\mathrm{IF}^{nu}_{G}(O|P)$ is the projection of $\mathrm{IF}^{ipw}_{G}(O|P)$ onto the nuisance tangent space $\mathbf{T}^{nu}(G)$, where $\mathrm{IF}^{ipw}_{G}(O|P)$ and $\mathrm{IF}^{nu}_{G}(O|P)$ are as defined in Theorem S3.1, and $\mathbf{T}^{nu}(G)$ is as defined in* (S3.6).

PROOF. The projection onto $\mathbf{T}^{nu}(G)$ is given by

$$\Pi(H'(O)|\mathbf{T}^{nu}(G)) = \int_{\mathcal{T}}\big\{P[H'(O)|X = s] - P[H'(O)|X \geq s]\big\}\, dM(s),$$

for any bounded measurable function $H' : \mathcal{T} \times \{0,1\} \times \mathbb{R} \to \mathbb{R}$. The projection of $\mathrm{IF}_G^{ipw}(O|P)$ onto $\mathbf{T}^{nu}(G)$ is therefore

(S4.11)
$$\Pi(\mathrm{IF}_G^{ipw}(O|P)|\mathbf{T}^{nu}(G)) = \int_{\mathcal{T}} \Big\{ P[\mathrm{IF}_G^{ipw}(O|P)|X=s] - P[\mathrm{IF}_G^{ipw}(O|P)|X \geq s] \Big\}\, dM(s)$$
$$= \int_{\mathcal{T}} \left\{ P\left[\frac{(U-P[U])\tilde{Y}}{\mathrm{Var}(U)} \middle| X=s \right] - P\left[\frac{(U-P[U])\tilde{Y}}{\mathrm{Var}(U)} \middle| X \geq s \right] \right\} dM(s).$$

The first term on the right-hand-side of (S4.11) is

$$\frac{1}{\mathrm{Var}(U)} \int_{\mathcal{T}} P[(U-P[U])\tilde{Y}|X=s]dM(s) = 0,$$

and it is easy to see that the second term is

$$\frac{1}{\mathrm{Var}(U)} \int_{\mathcal{T}} P[(U-P[U])\tilde{Y}|X \geq s]dM(s)$$
$$= \frac{1}{\mathrm{Var}(U)} P\left[(U-P[U]) \int_{\mathcal{T}} E[T|U, X \geq s] \frac{dM(s)}{G(s)} \right],$$

applying arguments used for deriving $\mathrm{IF}_G(O)$. Combining these two terms, we can see that the projection of $\mathrm{IF}_G^{ipw}(O|P)$ onto $\mathbf{T}^{nu}(G)$ is equal to $\mathrm{IF}_G^{nu}(O|P)$ that is defined in (S4.10). For further details see the proof of Theorem 2.3 of van der Laan and Robins (2003). □

S5. Proof of lemmas for Theorem 3.1. The first two lemmas in this section are given in order to apply Theorem 2.1 of van der Vaart and Wellner (2007), as done in the third lemma. For notational simplicity, henceforth we suppress the subscript k that is used in Section 3 in the main text, and just write $\Psi(P)$, U and $\mathrm{IF}^*(\cdot)$. First of all, we need the following notation.

By (A.2), we can fix an $\tilde{\epsilon}$ such that $0 < \tilde{\epsilon} < G(\tau)$. Let $\mathcal{G}$ be the collection of monotone nonincreasing càdlàg functions $\tilde{G}\colon \mathcal{T} \to [0,1]$ such that $\tilde{G}(\tau) > \tilde{\epsilon}$, and define

$$\mathcal{G}_0 = \Big\{ \tilde{G} : \tilde{G} \in \mathcal{G};\ \sup_{s \in \mathcal{T}} \Big| \frac{\tilde{G}(s)}{G(s)} - 1 \Big| \leq 1 \Big\}.$$

Let $\mathcal{Q}^*$ be the collection of monotone nondecreasing càdlàg functions $\tilde{Q}\colon \mathbb{R} \to [0,1]$. Note that $\tilde{Q}$ could be the c.d.f. of $U \sim Q_u$. By (A.1) that U is non-degenerate, there exists ν such that $0 < \nu < \mathrm{Var}_{Q_u}(U)$. Define $\mathcal{Q} = \{\tilde{Q} : \tilde{Q} \in \mathcal{Q}^*\,;\ |\tilde{Q}[U]| < \infty\,;\ \nu < \mathrm{Var}_{\tilde{Q}}(U) < \infty\}$. Let $\mathcal{E}$ be the collection of functions $(u,s) \mapsto \tilde{E}(u,s)$ that are uniformly bounded and left-continuous in s, and $\mathcal{H} = \{(u,s) \mapsto (\tilde{Q}(u), \tilde{E}(u,s)) : \tilde{Q} \in \mathcal{Q}; \tilde{E} \in \mathcal{E}\}$. Fix $\epsilon > 0$, and define $\mathcal{H}_0 \equiv \mathcal{Q}_0 \times \mathcal{E}_0 \subset \mathcal{H}$, where

$$\mathcal{Q}_0 = \big\{ \tilde{Q} 1(|\tilde{Q}[U] - Q_u[U]| \leq \epsilon, |\mathrm{Var}_{\tilde{Q}}(U) - \mathrm{Var}_{Q_u}(U)| \leq \epsilon) : \tilde{Q} \in \mathcal{Q} \big\}$$

and $\mathcal{E}_0 = \{\tilde{E} 1(|\tilde{E} - \bar{E}| \leq \epsilon) : \tilde{E} \in \mathcal{E}\}$.

LEMMA S5.1. *Given (A.1), (A.2), (A.3) and (A.4), let $\mathcal{H}_0$ and $\mathcal{G}_0$ be defined above, and let* IF^* *be as defined in the main text. Then*

(S5.1.1) $\mathrm{P}((\hat{E}_n, \mathbb{Q}_n) \in \mathcal{H}_0) \to 1$ *and* $\mathrm{P}(\hat{G}_n \in \mathcal{G}_0) \to 1$*;*
(S5.1.2) $\{\mathrm{IF}^*(\cdot|\tilde{E}, \tilde{Q}, \tilde{G}) : (\tilde{E}, \tilde{Q}) \in \mathcal{H}_0; \tilde{G} \in \mathcal{G}_0\}$ *is P-Donsker.*

PROOF. First note that $\mathbb{Q}_n[U]$ is bounded almost surely, and $\mathrm{Var}_{\mathbb{Q}_n}(U)$ is finite and larger than ν almost surely, by the strong law of large numbers and (A.1) that U is bounded and non-degenerate. In addition, $\hat{E}_n(u,s)$ is uniformly bounded over (u,s) on $\mathbb{R}\times\mathcal{T}$ and left-continuous in s with probability tending to one, by (A.4) that $\mathbb{E}[|\hat{E}_n(u,s)-\bar{E}(u,s)|^2]\to 0$ uniformly and $\bar{E}$ as a left-continuous funciton in s. Thus for sufficiently large n,

$$\begin{aligned}
\mathrm{P}((\hat{E}_n,\mathbb{Q}_n)\notin\mathcal{H}_0) &\le \mathrm{P}\big(\big|\mathbb{Q}_n[U]-Q_u[U]\big|>\epsilon\big)+\mathrm{P}\big(\big|\mathrm{Var}_{\mathbb{Q}_n}(U)-\mathrm{Var}_{Q_u}(U)\big|>\epsilon\big)\\
&\qquad \mathrm{P}\big(\big|\hat{E}_n(u,s)-\bar{E}(u,s)\big|>\epsilon\big)\\
&\le \mathbb{E}\big[\big|\mathbb{Q}_n[U]-Q_u[U]\big|\big]+\mathbb{E}\big[\big|\mathrm{Var}_{\mathbb{Q}_n}(U)-\mathrm{Var}_{Q_u}(U)\big|\big]+\mathbb{E}\big[\big|\hat{E}_n(u,s)-\bar{E}(u,s)\big|^2\big]\to 0
\end{aligned}$$

by the strong law of large numbers, (A.4) along with the above arguments. Note also that $\mathrm{P}(\hat{G}_n\in\mathcal{G}_0)\to 1$ by the uniform consistency of the Kaplan–Meier estimator. This gives (S5.1.1).

Moreover, we see that $\mathcal{G}$ and $\mathcal{Q}^*$ are P-Donsker (Example 19.11 in van der Vaart, 1998), and so are $\mathcal{G}_0$ and $\mathcal{Q}$ because $\mathcal{G}_0\subset\mathcal{G}$ and $\mathcal{Q}\subset\mathcal{Q}^*$ (Theorem 2.10.1 in van der Vaart and Wellner, 1996). Moreover, $\mathcal{H}_0$ is P-Donsker because $\mathcal{Q}_0$ and $\mathcal{E}_0$ are the classes of indicator functions multiplied by some uniformly bounded functions and known as P-Donsker (Corollary 9.32 (i) and (v) in Kosorok, 2008).

Now we start to show (S5.1.2). Based on the properties of the above classes, below we show the class $\mathcal{F}=\{f_1-f_2-f_3: f_1\in\mathcal{F}_1, f_2\in\mathcal{F}_2, f_3\in\mathcal{F}_3\}$ is a P-Donsker class of functions on $\mathcal{T}\times\{0,1\}\times\mathbb{R}$, where

$$\mathcal{F}_1=\left\{(x,\delta,u)\mapsto\frac{(u-\tilde{Q}[U])}{\mathrm{Var}_{\tilde{Q}}(U)}\Big(\frac{\delta x}{\tilde{G}(x)}-\tilde{Q}[\tilde{E}(U)]\Big):(\tilde{E},\tilde{Q})\in\mathcal{H}_0;\,\bar{G}\in\mathcal{G}_0\right\};$$

$$\mathcal{F}_2=\left\{(x,\delta,u)\mapsto\left[\frac{(u-\tilde{Q}[U])}{\mathrm{Var}_{\tilde{Q}}(U)}\right]^2\mathrm{Cov}_{\tilde{Q}}(U,\tilde{E}(U)):(\tilde{E},\tilde{Q})\in\mathcal{H}_0\right\};$$

$$\begin{aligned}\mathcal{F}_3=\Bigg\{(x,\delta,u)\mapsto\frac{(u-\tilde{Q}[U])}{\mathrm{Var}_{\tilde{Q}}(U)}\int_{\mathcal{T}}\tilde{E}(u,s)\big[1(x\in ds,\delta=0)-1(x\ge s)d\tilde{\Lambda}(s)\big]:\\
\tilde{\Lambda}(s)=-\log(\tilde{G}(s));\,(\tilde{E},\tilde{Q})\in\mathcal{H}_0;\,\tilde{G}\in\mathcal{G}_0\Bigg\}.\end{aligned}$$

We can see that $\mathrm{IF}^*(\cdot|\tilde{E},\tilde{Q},\tilde{G})\in\mathcal{F}$, so $\{\mathrm{IF}^*(\cdot|\tilde{E},\tilde{Q},\tilde{G}):(\tilde{E},\tilde{Q})\in\mathcal{H}_0;\tilde{G}\in\mathcal{G}_0\}\subseteq\mathcal{F}$. Note also that $\tilde{\Lambda}$ is a non-decreasing *càdlàg* function on $\mathcal{T}$, which could be a cumulative hazard function corresponding to $\tilde{G}$ if $\tilde{G}$ is a survival function. To show (S5.1.2), it suffices to show $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$ are P-Donsker.

By Corollary 9.32 (iv) in Kosorok (2008), the class $\{(x,\delta,u)\mapsto u-\tilde{Q}[U]:\tilde{Q}\in\mathcal{Q}_0\}$ is P-Donsker. Together with (A.1) that U is non-degenerate, the class $\mathcal{F}_4\equiv\{(x,\delta,u)\mapsto[(u-\tilde{Q}[U])/\mathrm{Var}_{\tilde{Q}}(U)]^r:\tilde{Q}\in\mathcal{Q}_0,\,r\in\{1,2\}\}$ is uniformly bounded and P-Donsker (Corollary 9.32 (iii) in Kosorok, 2008). Note also that $\tilde{Q}[\tilde{E}(U)]$ and $\mathrm{Cov}_{\tilde{Q}}(U,\tilde{E}(U))$ are constant functions of (x,δ,u), so the classes $\mathcal{F}_5\equiv\{(x,\delta,u)\mapsto\tilde{Q}[\tilde{E}(U)]:(\tilde{E},\tilde{Q})\in\mathcal{H}_0\}$ and $\mathcal{F}_6\equiv\{(x,\delta,u)\mapsto\mathrm{Cov}_{\tilde{Q}}(U,\tilde{E}(U)):(\tilde{E},\tilde{Q})\in\mathcal{H}_0\}$ are uniformly bounded and P-Donsker. Moreover, because $\tilde{G}(\tau)>\tilde{\epsilon}>0$ for all $\tilde{G}\in\mathcal{G}$ and $|X|\le\tau$, the aforementioned Donsker preservation properties imply that the class $\mathcal{F}_7\equiv\{(x,\delta,u)\mapsto\delta x/\tilde{G}(x),\tilde{G}\in\mathcal{G}_0\}$ is uniformly bounded and P-Donsker. Therefore by Donsker preservation properties, we have the below classes are P-Donsker:

$$\mathcal{F}_1\subseteq\{f_{11}\cdot(f_{12}-f_{13}):f_{11}\in\mathcal{F}_4,f_{12}\in\mathcal{F}_7,f_{13}\in\mathcal{F}_5\};$$

$$\mathcal{F}_2 \subseteq \{f_{21} \cdot f_{22} : f_{21} \in \mathcal{F}_4, f_{22} \in \mathcal{F}_6\}.$$

Below we show that $\mathcal{F}_3$ is P-Donsker. Let $\mathcal{F}_8 = \{(x, \delta, u) \mapsto \int_{\mathcal{T}} \tilde{E}(u,s)[1(x \in ds, \delta = 0) - 1(x \geq s)d\tilde{\Lambda}(s)] : \tilde{\Lambda}(s) = -\log(\tilde{G}(s)); \tilde{G} \in \mathcal{G}_0; \tilde{E} \in \mathcal{E}_0 \subset \mathcal{H}_0\}$. Let $\{t_1 < t_2 < \ldots < t_m\}$ be an arbitrary partition of $\mathcal{T}$ with uniform increments $\Delta_t = t_{j+1} - t_j$ for all j. Note that $\tilde{E} \in \mathcal{E}_0$ is uniformly bounded: $|\tilde{E}| \leq K_1$ for some positive constant K_1. Let $K' = 4K_1 \sum_{j=1}^m |\tilde{\Lambda}(t_j + \Delta_t)|$. Following that $\tilde{E}$ is also left-continuous in s, any function in $\mathcal{F}_8$ is the scalar multiple (by K') of the uniform limit of the sequence

$$\sum_{j=1}^m \frac{\tilde{E}(u, t_{j+1})}{K'} \Big[1_{[t_j, t_{j+1})}(x)(1-\delta) - \tilde{\Lambda}(t_{j+1})1(x \geq t_{j+1}) + \tilde{\Lambda}(t_j)1(x \geq t_j)\Big]$$

$$= \sum_{j=1}^m \frac{\tilde{E}(u, t_j + \Delta_t)}{K'} \Big[1_{[t_j, t_j+\Delta_t)}(x)(1-\delta) - \tilde{\Lambda}(t_j + \Delta_t)1(x \geq t_j + \Delta_t) + \tilde{\Lambda}(t_j)1(x \geq t_j)\Big]$$

$$= \sum_{j=1}^m \sum_{r=1}^4 \left\{ \frac{\tilde{E}(u, t_j + \Delta_t)}{K'} (-1)^r \tilde{\Lambda}(t_j + \Delta_t 1_{r=3})^{1_{r\geq 3}} \right\} (1-\delta)^{1_{r\leq 2}} 1\big(x \geq t_j + \Delta_t 1_{r\in\{2,3\}}\big)$$

$$\equiv \sum_{j=1}^m \sum_{r=1}^4 \alpha_{jr}(1-\delta)^{1_{r\leq 2}} 1\big(x \geq t_j + \Delta_t 1_{r\in\{2,3\}}\big),$$

where $\sum_{j=1}^m \sum_{r=1}^4 |\alpha_{jr}| \leq 1$ by

$$\sum_{j=1}^m \sum_{r=1}^4 |\alpha_{jr}| \leq \sum_{j=1}^m \sum_{r=1}^4 \left| \frac{\tilde{E}(u, t_j + \Delta_t)}{K'} (-1)^r \tilde{\Lambda}(t_j + \Delta_t 1_{r=3})^{1_{r\geq 3}} \right|$$

$$\leq \sum_{j=1}^m \sum_{r=1}^4 \left| \frac{\tilde{E}(u, t_j + \Delta_t)}{K'} \tilde{\Lambda}(t_j + \Delta_t) \right| \leq \frac{4K_1}{K'} \sum_{j=1}^m \big|\tilde{\Lambda}(t_j + \Delta_t)\big| = 1.$$

Then any function in $\mathcal{F}_8$ is in a class contained in the scalar-multiplied symmetric convex hull of the VC-subgraph class $\{(x, \delta) \mapsto \delta 1(x \geq s) : s \in \mathcal{T}\}$, which is a class of indicator functions. Therefore $\mathcal{F}_8$ is P-Donsker (Theorem 2.10.3 in van der Vaart and Wellner, 1996), and then $\mathcal{F}_3 \subseteq \{f_{31} \cdot f_{32} : f_{31} \in \mathcal{F}_4, f_{32} \in \mathcal{F}_8\}$ is also P-Donsker, (Corollary 9.32 (iii) in Kosorok, 2008). Hence, $\mathcal{F}$ is P-Donsker by Corollary 9.32 (i) of Kosorok (2008), which gives (S5.1.2). □

LEMMA S5.2. *Given (A.1), (A.2), (A.3), (A.4), and let $\mathcal{G}_0$ be defined above and* IF* *be as defined in the main text,* $\sup_{\tilde{G}\in\mathcal{G}_0} P[(\mathrm{IF}^*(\cdot|\hat{E}_n, \mathbb{Q}_n, \tilde{G}) - \mathrm{IF}^*(\cdot|\bar{E}, Q_u, \tilde{G}))^2] \xrightarrow{p} 0$.

PROOF. To give the proof, we apply Chebyshev's inequality, so that it suffices to show the first moments of the relevant mean-squared quantities converge to zero. Recall that P denotes the expectation of a generic variable, one of the conventional notations for empirical processes. For any $\tilde{G} \in \mathcal{G}_0$,

(S5.12)

$$P[\mathrm{IF}^*(\cdot|\hat{E}_n, \mathbb{Q}_n, \tilde{G}) - \mathrm{IF}^*(\cdot|\bar{E}, Q_u, \tilde{G})]^2 \lesssim (i) + (ii) + (iii) + (iv) + (v), \text{ where}$$

$$(i) \equiv P\left[\left\{\left(\frac{(U - \mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{(U - Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\right)\frac{\delta X}{\tilde{G}(X)}\right\}^2\right];$$

$$(ii) \equiv P\left[\left\{\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)}\mathbb{Q}_n[\hat{E}_n(U)] - \frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}Q_u[\bar{E}(U)]\right\}^2\right];$$

$$(iii) \equiv P\left[\left\{\frac{(U-\mathbb{Q}_n[U])^2}{\mathrm{Var}^2_{\mathbb{Q}_n}(U)}\mathrm{Cov}_{\mathbb{Q}_n}(U,\hat{E}_n(U)) - \frac{(U-Q_u[U])^2}{\mathrm{Var}^2_{Q_u}(U)}\mathrm{Cov}_{Q_u}(U,\bar{E}(U))\right\}^2\right];$$

$$(iv) \equiv P\left[\left\{\left(\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\right)\int_{\mathcal{T}}\hat{E}_n(U,s)\tilde{M}(U,ds)\right\}^2\right];$$

$$(v) \equiv P\left[\left\{\frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\int_{\mathcal{T}}\left(\hat{E}_n(U,s)-\bar{E}(U,s)\right)\tilde{M}(U,ds)\right\}^2\right].$$

Then showing S5.2 is equivalent to showing that the quantities (i)–(v) on the right-hand-side of (S5.12) converge to zero in probability. First along with (A.1) that U is non-degenerate, the strong law of large numbers and the continuous mapping theorem give that

$$\left|\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\right| \to 0 \ \text{a.s.} \tag{S5.13}$$

By (S5.13), together with $\tilde{G}(\tau) > \tilde{\epsilon} > 0$ for $\tilde{G} \in \mathcal{G}_0$ and $|X| \leq \tau$ such that $|\delta X/\tilde{G}(X)| \leq \tau/\tilde{\epsilon}$, we have the quantity (i) converging to zero in probability.

The quantity (ii) of (S5.12) is

$$\begin{aligned}\lesssim\, & P\left[\left\{\left[\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\right]Q_u[\bar{E}(U)]\right\}^2\right] \\ &+ P\left[\left\{\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)}\mathbb{Q}_n\left[\hat{E}_n(U)-\bar{E}(U)\right]\right\}^2\right] \\ &+ P\left[\left\{\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)}\left((\mathbb{Q}_n-Q_u)[\bar{E}(U)]\right)\right\}^2\right].\end{aligned} \tag{S5.14}$$

The first quantity of (S5.14) converges to zero in probability by (S5.13) and $Q_u[\bar{E}(U)]$ is bounded. The second quantity of (S5.14) is

$$\begin{aligned}\lesssim\, & P\left[\left\{\left[\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\right]\mathbb{Q}_n\left[\hat{E}_n(U)-\bar{E}(U)\right]\right\}^2\right] \\ &+ P\left[\left\{\frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\mathbb{Q}_n\left[\hat{E}_n(U)-\bar{E}(U)\right]\right\}^2\right],\end{aligned}$$

where by (A.1) that U is bounded and non-degenerate, the dominated convergence theorem implies that the first moment of the first part converges to zero, using (S5.13) and $\sup_{u\in\mathbb{R}}|\hat{E}_n(u) - \bar{E}(u)|$ is bounded in probability implied by (A.4). Therefore applying Chebyshev's inequality, the first part converges to zero in probability, and so does the second part by

$$\begin{aligned}&\mathbb{E}\left\{P\left[\left\{\frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\mathbb{Q}_n\left[\hat{E}_n(U)-\bar{E}(U)\right]\right\}^2\right]\right\} \\ &= P\left[\frac{(U-Q_u[U])^2}{\mathrm{Var}^2_{Q_u}(U)}\frac{1}{n^2}\left\{\sum_{i=1}^n \mathbb{E}\left\{\left[\hat{E}_n(U_i)-\bar{E}(U_i)\right]^2\right\}\right.\right. \\ &\qquad \left.\left.+ \sum_{\{i\neq j: i=1,\ldots,n\}}\sum_{j=1}^n \mathbb{E}\left\{\left[\hat{E}_n(U_i)-\bar{E}(U_i)\right]\left[\hat{E}_n(U_j)-\bar{E}(U_j)\right]\right\}\right\}\right]\end{aligned}$$

$$\lesssim P\left[\frac{1}{n}\mathbb{E}\big\{\big[\hat{E}_n(U_1)-\bar{E}(U_1)\big]^2\big\}+\frac{(n-1)}{n}\mathbb{E}\big\{\big[\hat{E}_n(U_1)-\bar{E}(U_1)\big]\big[\hat{E}_n(U_2)-\bar{E}(U_2)\big]\big\}\right]$$
$$\to 0,$$

where the penultimate line follows (A.1) that implies that $(U-Q_u[U])^2/\mathrm{Var}^2_{Q_u}(U)$ is bounded almost surely, and the convergence to zero holds by the dominated convergence theorem, along with (A.4) that implies that $\sup_{(u,s)}|\hat{E}_n(u,s)-\bar{E}(u,s)|$ is bounded in probability and $\mathbb{E}\{|\hat{E}_n(u)-\bar{E}(u)|\}=o(n^{-1/4})$ for each u. Therefore the second quantity of (S5.14) converges to zero in probability. The last quantity of (S5.14) converges to zero in probability, following (A.1) so that $(U-Q_u[U])/\mathrm{Var}_{Q_u}(U)$ is bounded almost surely and using $(\mathbb{Q}_n-Q_u)[\bar{E}(U)]\xrightarrow{p}0$.

The quantity (iii) of (S5.12) is

$$\text{(S5.15)}\qquad \lesssim P\left[\left\{\left[\frac{(U-\mathbb{Q}_n[U])^2}{\mathrm{Var}^2_{\mathbb{Q}_n}(U)}-\frac{(U-Q_u[U])^2}{\mathrm{Var}^2_{Q_u}(U)}\right]\mathrm{Cov}_{Q_u}(U,\bar{E}(U))\right\}^2\right]$$
$$+P\left[\left\{\frac{(U-\mathbb{Q}_n[U])^2}{\mathrm{Var}^2_{\mathbb{Q}_n}(U)}\left[\mathrm{Cov}_{\mathbb{Q}_n}(U,\hat{E}_n(U))-\mathrm{Cov}_{Q_u}(U,\bar{E}(U))\right]\right\}^2\right].$$

The first quantity of (S5.15) converges to zero in probability by (S5.13) and the fact that $\mathrm{Cov}_{Q_u}(U,\bar{E}(U))$ is bounded (implied by (A.1) and (A.4)). The second quantity of (S5.15) goes to zero in probability, following (A.1) that implies $(U-\mathbb{Q}_n[U])/\mathrm{Var}_{\mathbb{Q}_n}(U)$ is bounded almost surely and

$$\left|\mathrm{Cov}_{\mathbb{Q}_n}(U,\hat{E}_n(U))-\mathrm{Cov}_{Q_u}(U,\bar{E}(U))\right|\le\left|\mathrm{Cov}_{\mathbb{Q}_n}(U,\hat{E}_n(U))-\mathrm{Cov}_{\mathbb{Q}_n}(U,\bar{E}(U))\right|$$
$$+\left|\mathrm{Cov}_{\mathbb{Q}_n}(U,\bar{E}(U))-\mathrm{Cov}_{Q_u}(U,\bar{E}(U))\right|=o_p(1).$$

The above display results from the first term on the right-hand-side converging to zero in probability by $\mathbb{E}\{|\hat{E}_n(u)-\bar{E}(u)|\}=o(n^{-1/4})$ for each u in (A.4) along with (A.1), and the second term converging to zero in probability by the law of large numbers.

The quantity (iv) of (S5.12) is

(S5.16)
$$\lesssim P\left[\left\{\left(\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)}-\frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\right)\int_{\mathcal{T}}\Big(\hat{E}_n(U,s)-\bar{E}(U,s)\Big)\tilde{M}(U,ds)\right\}^2\right]$$
$$+P\left[\left\{\left(\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)}-\frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\right)\int_{\mathcal{T}}\bar{E}(U,s)\tilde{M}(U,ds)\right\}^2\right].$$

To show that the quantity (iv) of (S5.12) converges to zero in probability, it suffices to give the convergence to zero in probability of this upper bound in (S5.16). Note that $\tilde{M}(U,ds)=1(X\in ds,\delta=0)-1(X\ge s)d\tilde{\Lambda}(s|U)$ and

(S5.17)
$$\left|\int_{\mathcal{T}}\Big(\hat{E}_n(U,s)-\bar{E}(U,s)\Big)\tilde{M}(U,ds)\right|^2\le\sup_{(u,s)}\left|\hat{E}_n(u,s)-\bar{E}(u,s)\right|^2\sup_u\big(1+\tilde{\Lambda}(\tau|u)\big)^2$$
$$\lesssim\sup_{(u,s)}\left|\hat{E}_n(u,s)-\bar{E}(u,s)\right|^2,$$

so that

$$E\left\{P\left[\left\{\left(\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)}-\frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\right)\int_{\mathcal{T}}\Big(\hat{E}_n(U,s)-\bar{E}(U,s)\Big)\tilde{M}(U,ds)\right\}^2\right]\right\}$$

$$= \mathbb{E}\bigg\{ P\bigg[\bigg\{ \bigg(\frac{(U - \mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{(U - Q_u[U])}{\mathrm{Var}_{Q_u}(U)} \bigg) \int_{\mathcal{T}} \Big(\hat{E}_n(U,s) - \bar{E}(U,s) \Big) \tilde{M}(U,ds) \bigg\}^2 \bigg] \bigg\}$$

$$\lesssim P\bigg[\mathbb{E}\bigg\{ \bigg(\frac{(U - \mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{(U - Q_u[U])}{\mathrm{Var}_{Q_u}(U)} \bigg)^2 \sup_{(u,s)} \big| \hat{E}_n(u,s) - \bar{E}(u,s) \big|^2 \bigg\} \bigg] \to 0 \text{ a.s.},$$

following the dominated convergence theorem, along with (S5.13), (A.1) that U is bounded and non-degenerate, and (A.4) that $\sup_{(u,s)} |\hat{E}_n(u,s) - \bar{E}(u,s)|$ is bounded in probability. Therefore, we show that the first term in (S5.16) converges to zero in probability. We continue dealing with the second term in (S5.16). Similarly, we first upper-bounds $|\int_{\mathcal{T}} \bar{E}(U,s)\tilde{M}(U,ds)|^2$ by $\sup_{(u,s)} |\bar{E}(u,s)|^2 \sup_u (1 + \tilde{\Lambda}(\tau|u))^2$, which further gives that

$$E\bigg\{ P\bigg[\bigg\{ \bigg(\frac{(U - \mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{(U - Q_u[U])}{\mathrm{Var}_{Q_u}(U)} \bigg) \int_{\mathcal{T}} \bar{E}(U,s)\tilde{M}(U,ds) \bigg\}^2 \bigg] \bigg\}$$

$$= \mathbb{E}\bigg\{ P\bigg[\bigg\{ \bigg(\frac{(U - \mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{(U - Q_u[U])}{\mathrm{Var}_{Q_u}(U)} \bigg) \int_{\mathcal{T}} \bar{E}(U,s)\tilde{M}(U,ds) \bigg\}^2 \bigg] \bigg\}$$

$$\lesssim P\bigg[\mathbb{E}\bigg\{ \bigg(\frac{(U - \mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{(U - Q_u[U])}{\mathrm{Var}_{Q_u}(U)} \bigg)^2 \sup_{(u,s)} \big| \bar{E}(u,s) \big|^2 \bigg\} \bigg] \to 0 \text{ a.s.},$$

following the analogous arguments to those for the first term in (S5.16), along with (A.4) that $\bar{E}$ is uniformly bounded. Therefore, we show that the quantity (iv) of (S5.12) converges to zero in probability.

Now we deal with the last quantity (v) of (S5.12). Applying (S5.17) and the above arguments,

$$E\bigg\{ P\bigg[\bigg\{ \frac{(U - Q_u[U])}{\mathrm{Var}_{Q_u}(U)} \int_{\mathcal{T}} \Big(\hat{E}_n(U,s) - \bar{E}(U,s) \Big) \tilde{M}(U,ds) \bigg\}^2 \bigg] \bigg\}$$

$$= \mathbb{E}\bigg\{ P\bigg[\bigg\{ \frac{(U - Q_u[U])}{\mathrm{Var}_{Q_u}(U)} \int_{\mathcal{T}} \Big(\hat{E}_n(U,s) - \bar{E}(U,s) \Big) \tilde{M}(U,ds) \bigg\}^2 \bigg] \bigg\}$$

$$\lesssim P\bigg[\bigg(\frac{(U - Q_u[U])}{\mathrm{Var}_{Q_u}(U)} \bigg)^2 \mathbb{E}\bigg\{ \sup_{(u,s)} \big| \hat{E}_n(u,s) - \bar{E}(u,s) \big|^2 \bigg\} \bigg] \to 0,$$

where the convergence follows the dominated convergence theorem, (A.1) that U is bounded and non-degenerate, and (A.4) that $\sup_{(u,s)} |\hat{E}_n(u,s) - \bar{E}(u,s)|$ is bounded in probability and $\mathbb{E}\{|\hat{E}_n(u,s) - \bar{E}(u,s)|\} = o(n^{-1/4})$ for each (u,s). Hence, we conclude the proof. □

LEMMA S5.3. *Suppose that (A.1), (A.2), (A.3) and (A.4) hold. Let $\mathcal{H}_0$ be as defined above and* IF* *be as defined in the main text,* $\sup_{(\tilde{E},\tilde{Q}) \in \mathcal{H}_0} P[(\mathrm{IF}^*(\cdot|\tilde{E},\tilde{Q},\hat{G}_n) - \mathrm{IF}^*(\cdot|\tilde{E},\tilde{Q},G))^2] \xrightarrow{p} 0$.

PROOF. For any $(\tilde{E},\tilde{Q}) \in \mathcal{H}_0$,

$$P[\mathrm{IF}^*(\cdot|\tilde{E},\tilde{Q},\hat{G}_n) - \mathrm{IF}^*(\cdot|\tilde{E},\tilde{Q},G)]^2 \tag{S5.18}$$

$$\leq 2\bigg\{ P\bigg[\bigg\{ \frac{(U - \tilde{Q}[U])\delta X}{\mathrm{Var}_{\tilde{Q}}(U)} \bigg(\frac{1}{\hat{G}_n(X|U)} - \frac{1}{G(X)} \bigg) \bigg\}^2 \bigg]$$

$$+ P\bigg[\bigg\{ \frac{(U - \tilde{Q}[U])}{\mathrm{Var}_{\tilde{Q}}(U)} \int_{\mathcal{T}} \tilde{E}(U,s) \Big(\hat{M}(U,ds) - M(U,ds) \Big) \bigg\}^2 \bigg] \bigg\},$$

where $\hat{M}(U,\cdot)$ is the martingale residual with $\hat{\Lambda}_n(\cdot|U)$ corresponding to $\hat{G}_n(\cdot|U)$. Applying Taylor expansion on $(1/\hat{G}_n - 1/G)$, the first term on the right-hand-side of (S5.18) is dominated by

$$P\left[\left\{\frac{(U-\tilde{Q}[U])\delta X}{\mathrm{Var}_{\tilde{Q}}(U)}\left(\frac{1}{G^2(X)}\Big(\hat{G}_n(X|U)-G(X)\Big)\right)\right\}^2\right]$$
$$\leq \sup_{(u,s)\in\mathbb{R}\times\mathcal{T}}\Big(\hat{G}_n(s|u)-G(s)\Big)^2 P\left[\left\{\frac{(U-\tilde{Q}[U])\delta X}{G^2(X)\mathrm{Var}_{\tilde{Q}}(U)}\right\}^2\right] = o_p(1)$$

because $(U-\tilde{Q}[U])\delta X/\mathrm{Var}_{\tilde{Q}}(U) = O_p(1)$ that is implied by (A.1) and $|X|\leq\tau$; $G(\tau)>0$ as stated in (A.2) and the uniform convergence of $\hat{G}_n$.

Then we show that the second term on the right-hand-side of (S5.18) converges to zero in probability. As we have seen that $(U-\tilde{Q}[U])/\mathrm{Var}_{\tilde{Q}}(U) = O_p(1)$ by (A.1), it suffices to show that

$$P\left[\left\{\int_{\mathcal{T}}\tilde{E}(U,s)\Big(\hat{M}(U,ds)-M(U,ds)\Big)\right\}^2\right]\xrightarrow{p}0. \tag{S5.19}$$

The decomposition $\hat{M}(u,ds) = M(u,ds) + 1(X\geq s)(d\Lambda(s) - d\hat{\Lambda}_n(s|u))$ further reduces proving (S5.19) to showing that

$$P\left[\left\{\int_{\mathcal{T}}\tilde{E}(U,s)1(X\geq s)\Big(d\Lambda(s)-d\hat{\Lambda}_n(s|U)\Big)\right\}^2\right]\xrightarrow{p}0. \tag{S5.20}$$

Following $N_n(u,s)$ and $Y_n(u,s)$ as defined in Section S1, we easily see that $\bar{M}(u,ds) = N_n(u,ds) - Y_n(u,s)d\Lambda(s)$ is a local martingale with respect to the aggregated filtration that is defined in (S1.2) from Section S1. Note also that

(S5.21)
$$\hat{\Lambda}_n(t|u)-\Lambda(t) = \int_{-\infty}^{t}\frac{1(Y_n(u,s)>0)}{Y_n(u,s)}\bar{M}(u,ds) + \int_{-\infty}^{t}[1(Y_n(u,s)>0)-1]d\Lambda(s)$$
$$= \int_{-\infty}^{t}\frac{1(Y_n(u,s)>0)}{Y_n(u,s)}\bar{M}(u,ds) - \int_{-\infty}^{t}1(Y_n(u,s)=0)\,d\Lambda(s).$$

Inserting the decomposition in (S5.21) back to (S5.20), along with $(a+b)^2\leq 2\,(a^2+b^2)$, gives that showing (S5.20) is equivalent to showing

$$P\left[\left\{\int_{\mathcal{T}}\tilde{E}(U,s)1(X\geq s)\frac{1(Y_n(U,s)>0)}{Y_n(U,s)}\bar{M}(U,ds)\right\}^2\right]\xrightarrow{p}0\,; \tag{S5.22}$$

$$P\left[\left\{\int_{\mathcal{T}}\tilde{E}(U,s)1(X\geq s)1(Y_n(U,s)=0)d\Lambda(s)\right\}^2\right]\xrightarrow{p}0. \tag{S5.23}$$

Recall that $\mathbb{E}$ denotes the expectation over $O_1,\ldots,O_n$, regarding O as fixed, in contrast to the expectation P that applies to O. Note also that $\mathbb{E}[(\bar{M}(u,ds))^2|\bar{\mathcal{F}}_s] = Y_n(u,s)d\Lambda(s)$. Then for each u, $\tilde{E}(u,s)$ is left-continuous in s and adapted to the filtration $\bar{\mathcal{F}}_s$, so we have the display in (S5.22) by

$$E\left\{P\left[\left\{\int_{\mathcal{T}}\tilde{E}(U,s)1(X\geq s)\frac{1(Y_n(U,s)>0)}{Y_n(U,s)}\bar{M}(U,ds)\right\}^2\right]\right\}$$
$$=\mathbb{E}\left\{P\left[\left\{\int_{\mathcal{T}}\tilde{E}(U,s)1(X\geq s)\frac{1(Y_n(U,s)>0)}{Y_n(U,s)}\bar{M}(U,ds)\right\}^2\right]\right\}$$

$$= P\bigg[\mathbb{E}\bigg\{\bigg[\int_{\mathcal{T}}\tilde{E}(U,s)1(X\geq s)\frac{1(Y_n(U,s)>0)}{Y_n(U,s)}\bar{M}(U,ds)\bigg]^2\bigg\}\bigg]$$
$$= P\bigg[\int_{\mathcal{T}}\tilde{E}^2(U,s)1(X\geq s)\mathbb{E}\bigg\{\frac{1(Y_n(U,s)>0)}{Y_n(U,s)}\bigg\}d\Lambda(s)\bigg]\to 0,$$

where the convergence to zero in the last line follows that $\tilde{E}\in\mathcal{E}_0$ is uniformly bounded, that $\inf_{(u,s)\in\mathbb{R}\times\mathcal{T}}Y_n(u,s)\to\infty$, and the dominated convergence theorem. Similarly, the display in (S5.23) is an immediate consequence of the uniform boundedness of $\tilde{E}$, $1(Y_n(u,s)=0)=o_p(1)$ for each (u,s) and the dominated convergence theorem. Hence, we conclude this proof. □

LEMMA S5.4. *Given (A.1), (A.2), (A.3), (A.4) and* IF* *defined in the main text, we have that*

(S5.4.1) $[\mathbb{P}_n-P][\mathrm{IF}^*(\cdot|\hat{E}_n,\mathbb{Q}_n,G)-\mathrm{IF}^*(\cdot|\bar{E},Q_u,G)]=o_p(n^{-1/2})$;
(S5.4.2) $[\mathbb{P}_n-P][\mathrm{IF}^*(\cdot|\hat{E}_n,\mathbb{Q}_n,\hat{G}_n)-\mathrm{IF}^*(\cdot|\hat{E}_n,\mathbb{Q}_n,G)]=o_p(n^{-1/2})$.

PROOF. As the core of this proof relies on applying Theorem 2.1 of van der Vaart and Wellner (2007), we first relate our notation to theirs. We take the functional $\mathrm{IF}^*(\cdot|\tilde{E},\tilde{Q},\tilde{G})$ to be $f_{\theta,\eta}$ in their notation, where (θ,η) could be either $(\tilde{G},(\tilde{E},\tilde{Q}))$ or $((\tilde{E},\tilde{Q}),\tilde{G})$. Further, if we take $\theta\equiv\tilde{G}$ and $\eta\equiv(\tilde{E},\tilde{Q})$, which corresponds to the situation in (S5.4.1), we have in their notation $\eta_n\equiv(\hat{E}_n,\mathbb{Q}_n)$, $\eta_0\equiv(\bar{E},Q_u)$, $H_0\equiv\mathcal{H}_0$, and $\Theta\equiv\mathcal{G}_0$. Alternatively, if we take $\theta\equiv(\tilde{E},\tilde{Q})$ and $\eta\equiv\tilde{G}$, as in (S5.4.2), then $\eta_n\equiv\hat{G}_n$, $\eta_0\equiv G$, $H_0\equiv\mathcal{G}_0$, and $\Theta\equiv\mathcal{H}_0$. Note that the condition $\mathrm{P}(\eta_n\in H_0)\to 1$ and the P-Donsker condition of their theorem are satisfied by our Lemma S5.1. The main step is to check their condition (3), namely that $\sup_{\theta\in\Theta}\mathrm{P}(f_{\theta,\eta_n}-f_{\theta,\eta_0})^2\to_p 0$, in the instances arising here.

We first show (S5.4.1). For any $\epsilon>0$,

$$\mathrm{P}\Big(\Big|\sqrt{n}[\mathbb{P}_n-P][\mathrm{IF}^*(\cdot|\hat{E}_n,\mathbb{Q}_n,G)-\mathrm{IF}^*(\cdot|\bar{E},Q_u,G)]\Big|>\epsilon\Big)\leq\mathrm{P}\Big((\hat{E}_n,\mathbb{Q}_n)\notin\mathcal{H}_0\Big)$$
$$+\mathrm{P}\Big(G\notin\mathcal{G}_0\Big)+\mathrm{P}\Big((\bar{E},Q_u)\notin\mathcal{H}_0\Big)$$
$$+\mathrm{P}\Big(\sup_{\tilde{G}\in\mathcal{G}_0}\Big|\sqrt{n}[\mathbb{P}_n-P][\mathrm{IF}^*(\cdot|\hat{E}_n,\mathbb{Q}_n,\tilde{G})-\mathrm{IF}^*(\cdot|\bar{E},Q_u,\tilde{G})]\Big|>\epsilon\Big)\to 0,$$

where the first probability on the right-hand-side goes to zero by (S5.1.1) of Lemma S5.1, the second and third probability are trivially zero by the definitions of $\mathcal{G}_0$ and $\mathcal{H}_0$, and the last probability converges to zero by checking their condition (3) using Lemma S5.2.

Similarly, (S5.4.2) holds by

$$\mathrm{P}\Big(\Big|\sqrt{n}[\mathbb{P}_n-P][\mathrm{IF}^*(\cdot|\hat{E}_n,\mathbb{Q}_n,\hat{G}_n)-\mathrm{IF}^*(\cdot|\hat{E}_n,\mathbb{Q}_n,G)]\Big|>\epsilon\Big)$$
$$\leq\mathrm{P}\Big((\hat{E}_n,\mathbb{Q}_n)\notin\mathcal{H}_0\Big)+\mathrm{P}\Big(\hat{G}_n\notin\mathcal{G}_0\Big)+\mathrm{P}\Big(G\notin\mathcal{G}_0\Big)$$
$$+\mathrm{P}\Big(\sup_{(\tilde{E},\tilde{Q})\in\mathcal{H}_0}\Big|\sqrt{n}[\mathbb{P}_n-P][\mathrm{IF}^*(\cdot|\tilde{E},\tilde{Q},\hat{G}_n)-\mathrm{IF}^*(\cdot|\tilde{E},\tilde{Q},G)]\Big|>\epsilon\Big)\to 0,$$

where the first two probabilities on the right-hand-side converge to zero by (S5.1.1) of Lemma S5.1, the third probability is obviously zero by the definition of $\mathcal{G}_0$, and the last probability converges to zero by checking the condition (3) in this instance using Lemma S5.3 □

Before we proceed with the next lemma, we list some properties that will be repeatedly used later:

(S5.24)
$$\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)}-\frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}=\frac{(Q_u-\mathbb{Q}_n)[U]}{\mathrm{Var}_{\mathbb{Q}_n}(U)}+(U-Q_u[U])\Big[\frac{1}{\mathrm{Var}_{\mathbb{Q}_n}(U)}-\frac{1}{\mathrm{Var}_{Q_u}(U)}\Big]$$
$$=o_p(1);$$

(S5.25)
$$\sqrt{n}\Big\{\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)}-\frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\Big\}$$
$$=\frac{\sqrt{n}(Q_u-\mathbb{Q}_n)[U]}{\mathrm{Var}_{\mathbb{Q}_n}(U)}+(U-Q_u[U])\sqrt{n}\Big[\frac{1}{\mathrm{Var}_{\mathbb{Q}_n}(U)}-\frac{1}{\mathrm{Var}_{Q_u}(U)}\Big]=O_p(1),$$

which follow empirical process theories along with (A.1) that U is bounded and non-degenerate. In addition, we observe that $Y_n(u,\tau)\sim\mathrm{Binomial}(n,p_*)$ with $p_*=\mathrm{P}(X\geq\tau,c(U)=c(u))$ that is positive by (A.3). Along with the monotonicity of $Y_n(u,s)$ in $s\in\mathcal{T}$, Hoeffding's inequality gives that as $n\to\infty$, $\mathrm{P}(\inf_{s\in\mathcal{T}}Y_n(u,s)\leq\sqrt{n})=\mathrm{P}(Y_n(u,\tau)\leq\sqrt{n})\leq\exp(-2(\sqrt{n}p_*-1)^2)\to 0$. Therefore, we have that as $n\to\infty$,

(S5.26) $\mathrm{P}(\inf_{s\in\mathcal{T}}Y_n(u,s)>\sqrt{n})\to 1$ for each u;

(S5.27) $\sqrt{n}\mathrm{P}\big(Y_n(u,s)=0\big)\leq\sqrt{n}\mathrm{P}\big(Y_n(u,\tau)=0\big)=\sqrt{n}(1-p_*)^n\to 0,$ for each (u,s).

LEMMA S5.5. *Given (A.1), (A.2), (A.3), (A.4) and* IF^* *as defined in the main text,*
$$P[\mathrm{IF}^*(\cdot|\hat{E}_n,\mathbb{Q}_n,\hat{G}_n)-\mathrm{IF}^*(\cdot|\bar{E},Q_u,\hat{G}_n)+\mathrm{IF}^*(\cdot|\bar{E},Q_u,G)-\mathrm{IF}^*(\cdot|\hat{E}_n,\mathbb{Q}_n,G)]$$
$$=o_p(n^{-1/2}).$$

PROOF. Observe that
$$P[\mathrm{IF}^*(\cdot|\hat{E}_n,\mathbb{Q}_n,\hat{G}_n)]=\frac{P[(U-\mathbb{Q}_n[U])(Y-\mathbb{Q}_n[\hat{E}_n(U)])]}{\mathrm{Var}_{\mathbb{Q}_n}(U)}-\frac{\mathrm{Cov}_{\mathbb{Q}_n}(U,\hat{E}_n(U))}{\mathrm{Var}_{\mathbb{Q}_n}(U)}$$
$$-P[\mathrm{IF}^{CAR}(\cdot|\hat{E}_n,\mathbb{Q}_n,\hat{G}_n)];$$
$$P[\mathrm{IF}^*(\cdot|\hat{E}_n,\mathbb{Q}_n,G)]=\frac{P[(U-\mathbb{Q}_n[U])(\tilde{Y}-\mathbb{Q}_n[\hat{E}_n(U)])]}{\mathrm{Var}_{\mathbb{Q}_n}(U)}-\frac{\mathrm{Cov}_{\mathbb{Q}_n}(U,\hat{E}_n(U))}{\mathrm{Var}_{\mathbb{Q}_n}(U)}$$
$$-P[\mathrm{IF}^{CAR}(\cdot|\hat{E}_n,\mathbb{Q}_n,G)];$$
$$P[\mathrm{IF}^*(\cdot|\bar{E},Q_u,\hat{G}_n)]=\frac{P[(U-Q_u[U])(Y-Q_u[\bar{E}(U)])]}{\mathrm{Var}_{Q_u}(U)}-\frac{\mathrm{Cov}_{Q_u}(U,\bar{E}(U))}{\mathrm{Var}_{Q_u}(U)}$$
$$-P[\mathrm{IF}^{CAR}(\cdot|\bar{E},Q_u,\hat{G}_n)];$$
$$P[\mathrm{IF}^*(\cdot|\bar{E},Q_u,G)]=\frac{P[(U-Q_u[U])(\tilde{Y}-Q_u[\bar{E}(U)])]}{\mathrm{Var}_{Q_u}(U)}-\frac{\mathrm{Cov}_{Q_u}(U,\bar{E}(U))}{\mathrm{Var}_{Q_u}(U)}$$
$$-P[\mathrm{IF}^{CAR}(\cdot|\bar{E},Q_u,G)],$$

which implies that

(S5.28) $P[\mathrm{IF}^*(\cdot|\hat{E}_n,\mathbb{Q}_n,\hat{G}_n)-\mathrm{IF}^*(\cdot|\bar{E},Q_u,\hat{G}_n)+\mathrm{IF}^*(\cdot|\bar{E},Q_u,G)-\mathrm{IF}^*(\cdot|\hat{E}_n,\mathbb{Q}_n,G)]$

$$
\begin{aligned}
&= P\Big[\Big\{\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\Big\}(Y-\tilde{Y})\Big] \\
&\quad - P\left[\frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\int_{\mathcal{T}}\big(\hat{E}_n(U,s)-\bar{E}(U,s)\big)\big[\hat{M}(U,ds)-M(U,ds)\big]\right] \\
&\quad - P\Big[\Big[\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\Big]\int_{\mathcal{T}}\hat{E}_n(U,s)\big[\hat{M}(U,ds)-M(U,ds)\big]\Big].
\end{aligned}
$$

By the decomposition on the left-hand-side of (S5.24), along with $Y=\delta X/\hat{G}_n(X)$ and $\tilde{Y}=\delta X/G(X)$, we have the first quantity on the right-hand-side of (S5.28) as

$$
P\Big[\Big\{\frac{(Q_u-\mathbb{Q}_n)[U]}{\mathrm{Var}_{\mathbb{Q}_n}(U)} + (U-Q_u[U])\Big[\frac{1}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{1}{\mathrm{Var}_{Q_u}(U)}\Big]\Big\}\Big\{\frac{\delta X}{\hat{G}_n(X)} - \frac{\delta X}{G(X)}\Big\}\Big],
$$

which is $o_p(n^{-1/2})$ by (S5.25) and the uniform consistency of $\hat{G}_n$.

Then we deal with the last two terms on the right-hand-side of (S5.28). By the decomposition $\hat{M}(U,ds)-M(U,ds)=1(X\geq s)(d\Lambda(s)-d\hat{\Lambda}_n(s|U))$, the two terms turn into

(S5.29)
$$
\begin{aligned}
&- P\left[\frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\int_{\mathcal{T}}\big(\hat{E}_n(U,s)-\bar{E}(U,s)\big)1(X\geq s)(d\Lambda(s)-d\hat{\Lambda}_n(s|U))\right] \\
&- P\left[\left[\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\right]\int_{\mathcal{T}}\hat{E}_n(U,s)1(X\geq s)(d\Lambda(s)-d\hat{\Lambda}_n(s|U))\right].
\end{aligned}
$$

Now we tackle the first term of (S5.29), and apply similar techniques to the second term. According to the decomposition in (S5.21), the first term of (S5.29) is further expressed as

$$
P\left[\frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\int_{\mathcal{T}}\big(\hat{E}_n(U,s)-\bar{E}(U,s)\big)1(X\geq s)\frac{1(Y_n(U,s)>0)}{Y_n(U,s)}\bar{M}(U,ds)\right] \tag{S5.30}
$$

$$
-P\left[\frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\int_{\mathcal{T}}\big(\hat{E}_n(U,s)-\bar{E}(U,s)\big)1(X\geq s)1(Y_n(U,s)=0)\,d\Lambda(s)\right]. \tag{S5.31}
$$

The quantity (S5.30) is $o_p(n^{-1/2})$ as shown in what follows. First by Jensen's inequality, the second moment of the quantity (S5.30) (multiplied by $\sqrt{n}$) is bounded by

(S5.32)
$$
\begin{aligned}
&E\Big\{P\Big[\Big(\frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\int_{\mathcal{T}}\sqrt{n}\big(\hat{E}_n(U,s)-\bar{E}(U,s)\big)\frac{1(Y_n(U,s)>0)}{Y_n(U,s)}\bar{M}(U,ds)\Big)^2\Big]\Big\} \\
&= P\Big[\frac{(U-Q_u[U])^2}{\mathrm{Var}^2_{Q_u}(U)}\int_{\mathcal{T}}n\mathbb{E}\Big\{\big(\hat{E}_n(U,s)-\bar{E}(U,s)\big)^2\frac{1(Y_n(U,s)>0)}{Y_n^2(U,s)}\mathbb{E}\big[(\bar{M}(U,ds))^2\big|\bar{\mathcal{F}}_s\big]\Big\}\Big] \\
&= P\Big[\frac{(U-Q_u[U])^2}{\mathrm{Var}^2_{Q_u}(U)}\int_{\mathcal{T}}n\mathbb{E}\Big\{\big(\hat{E}_n(U,s)-\bar{E}(U,s)\big)^2\frac{1(Y_n(U,s)>0)}{Y_n(U,s)}\Big\}d\Lambda(s)\Big] \\
&\leq P\Big[\frac{(U-Q_u[U])^2}{\mathrm{Var}^2_{Q_u}(U)}\int_{\mathcal{T}}n\mathbb{E}\Big\{\big(\hat{E}_n(U,s)-\bar{E}(U,s)\big)^2\frac{1}{\sqrt{n}}\Big\}d\Lambda(s)\Big] \to 0.
\end{aligned}
$$

The second line of (S5.32) holds by the fact that $\bar{M}(u,ds)$ is a local martingale with respect to the aggregated filtration $\bar{\mathcal{F}}_s$ (defined in (S1.2) from Section S1), and that $\hat{E}_n$ and Y_n are predictable with respect to $\bar{\mathcal{F}}_s$ (see in the main text for the details of $\hat{E}_n$). Moreover, the inequality in (S5.32) holds by $\inf_{s\in\mathcal{T}}Y_n(u,s)>\sqrt{n}$ with probability tending to one for each

u as given in (S5.26), while the final convergence to zero follows (A.4) that $\mathbb{E}\{|\hat{E}_n(u,s) - \bar{E}(u,s)|\} = o(n^{-1/4})$ for each (u,s), along with using the dominated convergence theorem. Therefore by Chebyshev's inequality, (S5.32) implies that the quantity (S5.30) is $o_p(n^{-1/2})$.

The quantity (S5.31) is $o_p(n^{-1/2})$ because

$$
\begin{aligned}
&\sqrt{n}E\left| - P\left[\frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\int_{\mathcal{T}}\big(\hat{E}_n(U,s)-\bar{E}(U,s)\big)1(X\geq s)1(Y_n(U,s)=0)\,d\Lambda(s)\right]\right| \\
&\leq \sqrt{n}\mathbb{E}\left\{P\left[\frac{|U-Q_u[U]|}{\mathrm{Var}_{Q_u}(U)}\int_{\mathcal{T}}\big|\hat{E}_n(U,s)-\bar{E}(U,s)\big|1(X\geq s)1(Y_n(U,s)=0)d\Lambda(s)\right]\right\} \\
&= \sqrt{n}P\left[\frac{|U-Q_u[U]|}{\mathrm{Var}_{Q_u}(U)}\int_{\mathcal{T}}1(X\geq s)\mathbb{E}\Big\{\big|\hat{E}_n(U,s)-\bar{E}(U,s)\big|1(Y_n(U,s)=0)\Big\}d\Lambda(s)\right] \\
&\leq \sqrt{n}P\left[\frac{|U-Q_u[U]|}{\mathrm{Var}_{Q_u}(U)}\int_{\mathcal{T}}\sqrt{\mathbb{E}\Big\{\big|\hat{E}_n(U,s)-\bar{E}(U,s)\big|^2\Big\}\mathbb{E}\big\{1(Y_n(U,s)=0)\big\}}d\Lambda(s)\right] \to 0,
\end{aligned}
$$

where the last inequality holds by the fact that Λ is nondecreasing and $1(X \geq s) \leq 1$, and the Cauchy–Schwarz inequality. The final convergence to zero follows because (A.4) that $\mathbb{E}\{|\hat{E}_n(u,s) - \bar{E}(u,s)|\} = o(n^{-1/4})$ for each (u,s), and by (S5.27) that

$$\sqrt{n}\mathbb{E}\{1(Y_n(U,s)=0)\} = \sqrt{n}\mathrm{P}(Y_n(U,s)=0) \leq \sqrt{n}\mathrm{P}(Y_n(U,\tau)=0) \to 0,$$

together with (A.1) that U is bounded and non-degenerate and using the dominated convergence theorem. By Chebyshev's inequality, along with the above displays, the quantity (S5.31) is $o_p(n^{-1/2})$.

Similarly, the second term of (S5.29) is expressed as

(S5.33)
$$P\left[\left[\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\right]\int_{\mathcal{T}}\hat{E}_n(U,s)1(X\geq s)\frac{1(Y_n(U,s)>0)}{Y_n(U,s)}\bar{M}(U,ds)\right]$$

(S5.34)
$$-P\left[\left[\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\right]\int_{\mathcal{T}}\hat{E}_n(U,s)1(X\geq s)1(Y_n(U,s)=0)\,d\Lambda(s)\right].$$

The quantity (S5.33) is $o_p(n^{-1/2})$, applying similar arguments to those used for (S5.30). It therefore suffices to show that the second moment of the quantity (S5.33) (multiplied by $\sqrt{n}$) converges to zero in probability. As in (S5.32), we see the second moment of the quantity (S5.33) (multiplied by $\sqrt{n}$) is bounded by

$$
\begin{aligned}
&E\left\{nP\left[\left(\left[\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\right]\int_{\mathcal{T}}\hat{E}_n(U,s)\frac{1(Y_n(U,s)>0)}{Y_n(U,s)}\bar{M}(U,ds)\right)^2\right]\right\} \\
&= P\left[n\mathbb{E}\left\{\left[\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\right]^2\int_{\mathcal{T}}\hat{E}_n^2(U,s)\frac{1(Y_n(U,s)>0)}{Y_n(U,s)}d\Lambda(s)\right\}\right] \to 0,
\end{aligned}
$$

where the convergence to zero follows (S5.25), that $\hat{E}_n(u,s)$ is bounded in probability for each (u,s) (see in the main text for the details of $\hat{E}_n$), and $\inf_{(u,s)\in\mathbb{R}\times\mathcal{T}} Y_n(u,s) \to \infty$. Thus, we have that the quantity (S5.33) is $o_p(n^{-1/2})$, by Chebyshev's inequality. Analogously, we see the quantity (S5.34) is $o_p(n^{-1/2})$ as follows. The expectation of the absolute value of this quantity (multiplied by $\sqrt{n}$) is bounded by

$$\sqrt{n}E\left\{P\left[\left|\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{(U-Q_u[U])}{\mathrm{Var}_{Q_u}(U)}\right|\int_{\mathcal{T}}\big|\hat{E}_n(U,s)\big|1(Y_n(U,s)=0)\,d\Lambda(s)\right]\right\} \to 0,$$

using (S5.25), that $\hat{E}_n(u,s)$ is bounded in probability and as implied by (S5.27) that $1(Y_n(u,s)=0)=o_p(n^{-1/2})$, and the dominated convergence theorem. Along with Chebyshev's inequality, this gives the quantity (S5.34) is $o_p(n^{-1/2})$. Hence, we complete the proof. □

LEMMA S5.6. *Let $\hat{P}'_n=(\hat{E}_n,\mathbb{Q}_n,G)$, IF^* and $\mathrm{IF}^\dagger$ be as respectively defined in the main text. Given (A.1), (A.2), (A.3) and (A.4),*

$$S(\mathbb{P}_n,\hat{P}'_n)-\Psi(P)=[\mathbb{P}_n-P]\big\{\,\mathrm{IF}^*(\cdot|\bar{E},Q_u,G)+\mathrm{IF}^\dagger(\cdot|\bar{E},P)\big\}+o_p(n^{-1/2}).$$

PROOF.

$$\begin{aligned}
S(\mathbb{P}_n,\hat{P}'_n)-\Psi(P)&=\Psi(\hat{P}'_n)+\mathbb{P}_n\,\mathrm{IF}^*(\cdot|\hat{P}'_n)-\Psi(P)\\
&=[\mathbb{P}_n-P]\,\mathrm{IF}^*(\cdot|\hat{P}'_n)+\Big[\Psi(\hat{P}'_n)-\Psi(P)+P\,\mathrm{IF}^*(\cdot|\hat{P}'_n)\Big]\\
&=[\mathbb{P}_n-P]\,\mathrm{IF}^*(\cdot|\hat{P}'_n)+[\mathbb{P}_n-P]\,\mathrm{IF}^\dagger(\cdot|\bar{E},P)+o_p(n^{-1/2})\\
&=[\mathbb{P}_n-P]\big\{\,\mathrm{IF}^*(\cdot|\bar{E},Q_u,G)+\mathrm{IF}^\dagger(\cdot|\bar{E},P)\big\}+[\mathbb{P}_n-P][\mathrm{IF}^*(\cdot|\hat{P}'_n)-\mathrm{IF}^*(\cdot|\bar{E},Q_u,G)]\\
&\quad+o_p(n^{-1/2})\\
&=[\mathbb{P}_n-P]\big\{\,\mathrm{IF}^*(\cdot|\bar{E},Q_u,G)+\mathrm{IF}^\dagger(\cdot|\bar{E},P)\big\}+o_p(n^{-1/2}),
\end{aligned}$$

where the last equality follows by (S5.4.1) of Lemma S5.4. The third equality is shown below.

Recalling the definition in (6) of the main text,

$$\begin{aligned}
\Psi(\hat{P}'_n)-\Psi(P)+P\,\mathrm{IF}^*(\cdot|\hat{P}'_n)&=\frac{\mathrm{Cov}_{\mathbb{Q}_n}(U,\hat{E}_n(U))}{\mathrm{Var}_{\mathbb{Q}_n}(U)}-\frac{\mathrm{Cov}_{Q_u}(U,E[\tilde{Y}|U])}{\mathrm{Var}_{Q_u}(U)}\\
&+\frac{1}{\mathrm{Var}_{\mathbb{Q}_n}(U)}P(U-\mathbb{Q}_n[U])\big(\delta X/G(X)-\mathbb{Q}_n[\hat{E}_n(U)]\big)\\
&-\frac{1}{\mathrm{Var}^2_{\mathbb{Q}_n}(U)}P(U-\mathbb{Q}_n[U])^2\mathrm{Cov}_{\mathbb{Q}_n}(U,\hat{E}_n(U))-P\left[\frac{(U-\mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)}\int_{\mathcal{T}}\hat{E}_n(U,s)dM(s)\right].
\end{aligned}$$

Inserting $\tilde{Y}=\delta X/G(X)$, $\mathbb{Q}_n[U]=n^{-1}\sum_{i=1}^n U_i$ and $\mathbb{Q}_n[\hat{E}_n(U)]=n^{-1}\sum_{i=1}^n\hat{E}_n(U_i)$ back into the third term of the above display implies that

$$\begin{aligned}
&P(U-\mathbb{Q}_n[U])\big(\delta X/G(X)-\mathbb{Q}_n[\hat{E}_n(U)]\big)\\
&=P\Big[\Big\{\Big(U-Q_u[U]\Big)+\Big(Q_u[U]-\frac{1}{n}\sum_{i=1}^n U_i\Big)\Big\}\tilde{Y}\Big]-\frac{1}{n}\sum_{i=1}^n P\Big[\Big(U-\frac{1}{n}\sum_{i=1}^n U_i\Big)\hat{E}_n(U_i)\Big].
\end{aligned}$$

Recall also that P denotes the expectation that applies only to $O\sim P$ and not to any estimator composed by $\{O_1,\ldots,O_n\}$ and let $\tilde{Y}=\tilde{Y}_1$ and $U=U_1$ without loss of generality, so we therefore see that

$$\begin{aligned}
P\Big[\Big(Q_u[U]-\frac{1}{n}\sum_{i=1}^n U_i\Big)\tilde{Y}\Big]&=Q_u[U]P[\tilde{Y}]-\frac{1}{n}P[U_1\tilde{Y}_1]-\frac{1}{n}\sum_{i\neq 1}^n Q_u[U_i]P[\tilde{Y}_1]\\
&=\frac{1}{n}Q_u[U]P[\tilde{Y}]-\frac{1}{n}P[U\tilde{Y}],
\end{aligned}$$

and

$$\frac{1}{n}\sum_{i=1}^n P\Big[\Big(U-\frac{1}{n}\sum_{i=1}^n U_i\Big)\hat{E}_n(U_i)\Big]=\frac{1}{n}P[U_1\hat{E}_n(U_1)]+\frac{1}{n}\sum_{i\neq 1}^n Q_u[U_1]P[\hat{E}_n(U_i)]$$

$$- \frac{1}{n^2}\sum_{i=1}^{n} P[U_i \hat{E}_n(U_i)] - \frac{1}{n^2}\sum_{\{i:\, i=1,\ldots,n;\, i\neq j\}}\sum_{j=1}^{n} Q_u[U_i]P[\hat{E}_n(U_j)] = 0.$$

Hence combining the results in the above three displays, we have that

$$\begin{aligned}
&\frac{1}{\mathrm{Var}_{\mathbb{Q}_n}(U)} P(U - \mathbb{Q}_n[U])\big(\delta X/G(X) - \mathbb{Q}_n[\hat{E}_n(U)]\big) \\
&= \frac{1}{\mathrm{Var}_{\mathbb{Q}_n}(U)}\left\{P\big[(U - Q_u[U])\tilde{Y}\big] + \frac{1}{n}Q_u[U]P[\tilde{Y}] - \frac{1}{n}P[U\tilde{Y}]\right\} \\
&= \frac{\mathrm{Cov}_{Q_u}(U, E[\tilde{Y}|U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)} + o_p(n^{-1/2}),
\end{aligned}$$

following (A.1) and (A.2) so that $Q_u[U]$, $P[\tilde{Y}]$ and $P[U\tilde{Y}]$ are bounded, and $\mathrm{Var}_{\mathbb{Q}_n}(U)$ is bounded away from zero almost surely.

Let

$$\begin{aligned}
&(i) \equiv \frac{\mathrm{Cov}_{\mathbb{Q}_n}(U, \hat{E}_n(U))}{\mathrm{Var}_{\mathbb{Q}_n}(U)};\ (ii) \equiv -\frac{\mathrm{Cov}_{Q_u}(U, E[\tilde{Y}|U])}{\mathrm{Var}_{Q_u}(U)};\ (iii) \equiv \frac{\mathrm{Cov}_{Q_u}(U, E[\tilde{Y}|U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)}; \\
&(iv) \equiv -\frac{1}{\mathrm{Var}^2_{\mathbb{Q}_n}(U)} P(U - \mathbb{Q}_n[U])^2 \mathrm{Cov}_{\mathbb{Q}_n}(U, \hat{E}_n(U)); \\
&(v) \equiv -P\left[\frac{(U - \mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)}\int_{\mathcal{T}} \hat{E}_n(U, s)dM(s)\right].
\end{aligned}$$

so we have

$$\Psi(\hat{P}'_n) - \Psi(P) + P\,\mathrm{IF}^*(\cdot|\hat{P}'_n) = (i) + (ii) + (iii) + (iv) + (v). \tag{S5.35}$$

The quantity (iv) could be simplified as

$$\begin{aligned}
(iv) &= -\frac{1}{\mathrm{Var}^2_{\mathbb{Q}_n}(U)} P(U - Q_u[U] + Q_u[U] - \mathbb{Q}_n[U])^2 \mathrm{Cov}_{\mathbb{Q}_n}(U, \hat{E}_n(U)) \\
&= -\frac{\mathrm{Cov}_{\mathbb{Q}_n}(U, \hat{E}_n(U))}{\mathrm{Var}^2_{\mathbb{Q}_n}(U)}\left[\mathrm{Var}_{Q_u}(U) + P(Q_u[U] - \mathbb{Q}_n[U])^2 - \frac{2}{n}\mathrm{Var}_{Q_u}(U)\right] \\
&= -\frac{\mathrm{Cov}_{\mathbb{Q}_n}(U, \hat{E}_n(U))}{\mathrm{Var}^2_{\mathbb{Q}_n}(U)}\mathrm{Var}_{Q_u}(U) + o_p(n^{-1/2}),
\end{aligned}$$

following the fact that $\sqrt{n}(Q_u[U] - \mathbb{Q}_n[U])^2 = o_p(1)$, $\mathrm{Cov}_{\mathbb{Q}_n}(U, \hat{E}_n(U))/\mathrm{Var}^2_{\mathbb{Q}_n}(U) = O_p(1)$ and $\mathrm{Var}_{Q_u}(U)$ is bounded, by assumptions (A.1) and the details of $\hat{E}_n$ in the main text. Moreover,

$$(v) = -P\left[\frac{(U - \mathbb{Q}_n[U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)}\int_{\mathcal{T}} \hat{E}_n(U, s)P[dM(s)|U]\right] = 0,$$

following $dM(s) = 1(X \in ds, \delta = 0) - 1(X \geq s)d\Lambda(s)$ and $d\Lambda(s) = \mathrm{P}(C \in ds)/\mathrm{P}(C \geq s)$, so that $P[dM(s)|U] = P[1(T \geq s)|U]P[1(C \in ds) - 1(C \geq s)d\Lambda(s)] = 0$, together with the independent censoring assumption.

Inserting the above results into (S5.35), we have that

(S5.36)

$$\Psi(\hat{P}'_n) - \Psi(P) + P\,\mathrm{IF}^*(\cdot|\hat{P}'_n)$$

$$= \frac{\mathrm{Cov}_{\mathbb{Q}_n}(U, \hat{E}_n(U))}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{\mathrm{Cov}_{Q_u}(U, E[\tilde{Y}|U])}{\mathrm{Var}_{Q_u}(U)} + \frac{\mathrm{Cov}_{Q_u}(U, E[\tilde{Y}|U])}{\mathrm{Var}_{\mathbb{Q}_n}(U)}$$

$$- \frac{\mathrm{Cov}_{\mathbb{Q}_n}(U, \hat{E}_n(U))}{\mathrm{Var}^2_{\mathbb{Q}_n}(U)} \mathrm{Var}_{Q_u}(U) + o_p(n^{-1/2})$$

$$= \left[\mathrm{Var}_{\mathbb{Q}_n}(U) - \mathrm{Var}_{Q_u}(U)\right] \frac{1}{\mathrm{Var}_{\mathbb{Q}_n}(U)} \mathrm{Cov}_{Q_u}(U, E[\tilde{Y}|U]) \left[\frac{1}{\mathrm{Var}_{\mathbb{Q}_n}(U)} - \frac{1}{\mathrm{Var}_{Q_u}(U)}\right]$$

$$+ \left[\mathrm{Var}_{\mathbb{Q}_n}(U) - \mathrm{Var}_{Q_u}(U)\right] \left\{\frac{1}{\mathrm{Var}^2_{\mathbb{Q}_n}(U)} - \frac{1}{\mathrm{Var}^2_{Q_u}(U)}\right\}$$

$$\times \left[\mathrm{Cov}_{\mathbb{Q}_n}(U, \hat{E}_n(U)) - \mathrm{Cov}_{Q_u}(U, E[\tilde{Y}|U])\right]$$

$$+ \left[\mathrm{Var}_{\mathbb{Q}_n}(U) - \mathrm{Var}_{Q_u}(U)\right] \frac{1}{\mathrm{Var}^2_{Q_u}(U)} \left[\mathrm{Cov}_{\mathbb{Q}_n}(U, \hat{E}_n(U)) - \mathrm{Cov}_{Q_u}(U, E[\tilde{Y}|U])\right]$$

$$+ o_p(n^{-1/2}).$$

Because $\mathrm{Var}_{\mathbb{Q}_n}(U)$ is bounded away from zero almost surely by (A.1) that U is non-degenerate, along with $\mathrm{Cov}_{Q_u}(U, E[\tilde{Y}|U])$ is bounded by (A.1) and the fact that $E[\tilde{Y}|U = u, X \geq s]$ is uniformly bounded over (u, s), the first quantity on the right-hand-side of (S5.36) is $o_p(n^{-1/2})$, following $\sqrt{n}[\mathrm{Var}_{\mathbb{Q}_n}(U) - \mathrm{Var}_{Q_u}(U)] = O_p(1)$ and $1/\mathrm{Var}_{\mathbb{Q}_n}(U) - 1/\mathrm{Var}_{Q_u}(U) = o_p(1)$.

Note also that using the general properties of $\hat{E}_n$ in (A.4) and the law of large numbers gives

$$\begin{aligned}
&\mathrm{Cov}_{\mathbb{Q}_n}(U, \hat{E}_n(U)) - \mathrm{Cov}_{Q_u}(U, E[\tilde{Y}|U]) = \mathrm{Cov}_{Q_u}(U, \bar{E}(U) - E[\tilde{Y}|U]) \\
&\quad + \mathrm{Cov}_{\mathbb{Q}_n}(U, \hat{E}_n(U) - \bar{E}(U)) + \left[\mathrm{Cov}_{\mathbb{Q}_n}(U, \bar{E}(U)) - \mathrm{Cov}_{Q_u}(U, \bar{E}(U))\right] \\
&= \mathrm{Cov}_{Q_u}(U, \bar{E}(U) - E[\tilde{Y}|U]) + o_p(1).
\end{aligned} \tag{S5.37}$$

By the facts that $\sqrt{n}[\mathrm{Var}_{\mathbb{Q}_n}(U) - \mathrm{Var}_{Q_u}(U)] = O_p(1)$ and $1/\mathrm{Var}^2_{\mathbb{Q}_n}(U) - 1/\mathrm{Var}^2_{Q_u}(U) = o_p(1)$, the second quantity on the right-hand-side of (S5.36) is also $o_p(n^{-1/2})$. In addition, we observe that $\mathrm{Var}_{\mathbb{Q}_n}(U)$ is regular asymptotically linear estimator of $\mathrm{Var}_{Q_u}(U)$ with influence function $o \mapsto (u - Q_u[U])^2$. Combining this fact with $\sqrt{n}[\mathrm{Var}_{\mathbb{Q}_n}(U) - \mathrm{Var}_{Q_u}(U)] = O_p(1)$ and the display in (S5.37), the third quantity on the right-hand-side of (S5.36) turns into

$$\left[\mathrm{Var}_{\mathbb{Q}_n}(U) - \mathrm{Var}_{Q_u}(U)\right] \frac{1}{\mathrm{Var}^2_{Q_u}(U)} \mathrm{Cov}_{Q_u}(U, \bar{E}(U) - E[\tilde{Y}|U]) + o_p(n^{-1/2})$$

$$= \frac{\mathrm{Cov}_{Q_u}(U, \bar{E}(U) - E[\tilde{Y}|U])}{\mathrm{Var}^2_{Q_u}(U)} \left\{[\mathbb{P}_n - P](U - Q_u[U])^2\right\} + o_p(n^{-1/2})$$

$$\equiv \mathbb{P}_n \, \mathrm{IF}^{\dagger}(\cdot|\bar{E}, P) + o_p(n^{-1/2}).$$

Referring to (S5.36) and noting that $P\,\mathrm{IF}^{\dagger}(\cdot|\bar{E}, P) = 0$, we have completed the proof that $\Psi(\hat{P}'_n) - \Psi(P) + P\,\mathrm{IF}^*(\cdot|\hat{P}'_n) = [\mathbb{P}_n - P]\,\mathrm{IF}^{\dagger}(\cdot|\bar{E}, P) + o_p(n^{-1/2})$. □

S6. Proof of lemmas for Theorem 4.1. For any $m_0, m_1 > 0$, let $\mathcal{BV}(\mathcal{T}, m_0, m_1)$ be the collection of functions $f : \mathcal{T} \to [-m_0, m_0]$ with total variation bounded by m_1. The lemma below gives preservation properties of these classes.

LEMMA S6.1. *Fix $m_0, \bar{m}_0, m_1, \bar{m}_1 > 0$ and $\varepsilon \in (0,1)$. For any $f \in \mathcal{BV}(\mathcal{T}, m_0, m_1)$ and $g \in \mathcal{BV}(\mathcal{T}, \bar{m}_0, \bar{m}_1)$, $f+g$ and $f-g$ are contained in $\mathcal{BV}(\mathcal{T}, m_0+\bar{m}_0, m_1+\bar{m}_1)$; fg belongs to $\mathcal{BV}(\mathcal{T}, m_0\bar{m}_0, m_0\bar{m}_1+\bar{m}_0 m_1)$; moreover, if g is such that $\inf_{t\in\mathcal{T}}|g(t)| > \varepsilon$, then f/g is contained in $\mathcal{BV}(\mathcal{T}, m_0/\varepsilon, (m_0\bar{m}_1+\bar{m}_0 m_1)/\varepsilon^2)$.*

PROOF. Let $\|\tilde{f}\|_\nu$ denote the total variation of $s \mapsto \tilde{f}(s), s \in \mathcal{T}$, and let $\|f\|_\infty$ denote $\sup_{t\in\mathcal{T}}|f(t)|$. As the notation indicates, both $\|\cdot\|_\nu$ and $\|\cdot\|_\infty$ are norms, and therefore satisfy the triangle inequality.

As $f \in \mathcal{BV}(\mathcal{T}, m_0, m_1)$ and $g \in \mathcal{BV}(\mathcal{T}, \bar{m}_0, \bar{m}_1)$, we have $\|f\|_\infty < m_0$, $\|f\|_\nu < m_1$, $\|g\|_\infty < \bar{m}_0$ and $\|g\|_\nu < \bar{m}_1$.

To see that $f+g \in \mathcal{BV}(\mathcal{T}, m_0+\bar{m}_0, m_1+\bar{m}_1)$, note that $\|f+g\|_\infty \le \|f\|_\infty + \|g\|_\infty < m_0+\bar{m}_0$ and $\|f+g\|_\nu \le \|f\|_\nu + \|g\|_\nu < m_1+\bar{m}_1$. The same argument shows that $f-g \in \mathcal{BV}(\mathcal{T}, m_0+\bar{m}_0, m_1+\bar{m}_1)$.

To see that $fg \in \mathcal{BV}(\mathcal{T}, m_0\bar{m}_0, m_0\bar{m}_1+\bar{m}_0 m_1)$, note that $\|fg\|_\infty \le \|f\|_\infty\|g\|_\infty < m_0\bar{m}_0$, and, for an arbitrary partition $\{t_1 < t_2 < \ldots < t_m < t_{m+1}\}$ of $\mathcal{T}$,

$$\sum_{j=1}^m |(fg)(t_{j+1}) - (fg)(t_j)| \le \sum_{j=1}^m \Big\{|g(t_{j+1})||f(t_{j+1}) - f(t_j)| + |f(t_j)||g(t_{j+1}) - g(t_j)|\Big\}$$

$$\le \bar{m}_0 \sum_{j=1}^m |f(t_{j+1}) - f(t_j)| + m_0 \sum_{j=1}^m |g(t_{j+1}) - g(t_j)| \le \bar{m}_0\|f\|_\nu + m_0\|g\|_\nu$$

$$< m_0\bar{m}_1 + \bar{m}_0 m_1.$$

For the ratio of two functions, $\|f/g\|_\infty < m_0/\varepsilon$, and, for an arbitrary partition $\{t_1 < t_2 < \ldots < t_m < t_{m+1}\}$ of $\mathcal{T}$,

$$\sum_{j=1}^m |f(t_{j+1})/g(t_{j+1}) - f(t_j)/g(t_j)| = \sum_{j=1}^m \left|\frac{g(t_j)f(t_{j+1})}{g(t_j)g(t_{j+1})} - \frac{g(t_{j+1})f(t_j)}{g(t_{j+1})g(t_j)}\right|$$

$$\le \frac{1}{\varepsilon^2}\sum_{j=1}^m |g(t_j)f(t_{j+1}) - g(t_{j+1})f(t_j)|$$

$$\le \frac{1}{\varepsilon^2}\sum_{j=1}^m \Big\{|g(t_j)||f(t_{j+1}) - f(t_j)| + |f(t_j)||g(t_{j+1}) - g(t_j)|\Big\}$$

$$\le \big(\bar{m}_0\|f\|_v + m_0\|g\|_v\big)/\varepsilon^2 < (m_0\bar{m}_1 + \bar{m}_0 m_1)/\varepsilon^2.$$

□

Note that the present theorem refers to $\hat{E}_n$ given in Remark 4.2 with n replaced by j for $j = q_n, \ldots, n-1$, and $U = U_k$ for a given k. That is, with k included and the sample size of j considered, we are now dealing with

(S6.38)
$$\hat{E}_j(u,s,k) \equiv \mathbb{P}_j[Y1(X\ge s)] + \frac{\mathrm{Cov}_{\mathbb{P}_j}(U_k 1(X\ge s), Y1(X\ge s))}{\mathrm{Var}_{\mathbb{P}_j}(U_k 1(X\ge s))}(u - \mathbb{P}_j[U_k 1(X\ge s)]).$$

By the weak law of large numbers and the continuous mapping theorem, together with the uniform consistency of $\hat{G}_n$ on $\mathcal{T}$, $\hat{E}_n$ is a pointwise consistent estimator of

$$E_0(u,s,k) \equiv P[\tilde{Y}1(X\ge s)] + \frac{\mathrm{Cov}(U_k 1(X\ge s), \tilde{Y}1(X\ge s))}{\mathrm{Var}(U_k 1(X\ge s))}(u - P[U_k 1(X\ge s)]).$$

Note that we suppress the argument s if $s=-\infty$. The following lemma shows that all of $\hat{E}_j$, $j=\{q_n,\dots,n\}$, and E_0 are asymptotically contained in a class of uniformly bounded functions with uniformly bounded total variation.

LEMMA S6.2. *Under the assumptions (A.1), (A.2) and (A.6), there exist positive constants $\tilde{M}_0$ and $\tilde{M}_1$ such that for each k, the function $E_0(\cdot,\cdot,k)$ is contained in the class*

$$\left\{(u,s)\ \mapsto\ a(s)+b(s)u:\ a,b\in\mathcal{BV}(\mathcal{T},\tilde{M}_0,\tilde{M}_1)\right\},$$

and moreover, $\hat{E}_j(\cdot,\cdot,k)$ as defined in (S6.38) is contained in this class with probability tending to one, for $j=q_n,\dots,n$.

PROOF. Let the upper bound of all the $|U_k|$ be $\tilde{M}_u\in(0,\infty)$, which is ensured by the assumption (A.1). And let a constant $\tilde{M}_y>\tau/G(\tau)>0$, following $G(\tau)>0$ in (A.2) such that $\tilde{M}_y\in(0,\infty)$. For some $\varepsilon\in(0,1)$, take $\tilde{M}_0=\tilde{M}_y+\sum_{q=1}^{2}2\varepsilon^{-1}\tilde{M}_u^q\tilde{M}_y$ and $\tilde{M}_1=\tilde{M}_y+\sum_{q=1}^{4}6\varepsilon^{-2}\tilde{M}_u^q\tilde{M}_y$. Note that $\tilde{M}_u$ and $\tilde{M}_y$ do not depend on (j,k,n), so $\tilde{M}_0$ and $\tilde{M}_1$ are independent of (j,k,n).

To have $E_0(\cdot,\cdot,k)$ in the above-defined class, it suffices to have both

$$P[\tilde{Y}1(X\geq s)]-\frac{\mathrm{Cov}(U_k1(X\geq s),\tilde{Y}1(X\geq s))}{\mathrm{Var}(U_k1(X\geq s))}P[U_k1(X\geq s)]$$

and $\mathrm{Cov}(U_k1(X\geq s),\tilde{Y}1(X\geq s))/\mathrm{Var}(U_k1(X\geq s))$ belonging to $\mathcal{BV}(\mathcal{T},\tilde{M}_0,\tilde{M}_1)$. We start by showing that each of the following functions belongs to an appropriate class of uniformly bounded functions with uniformly bounded total variation: $P[\tilde{Y}1(X\geq s)]$, $P[U_k1(X\geq s)]$, $P[(U_k-P[U_k1(X\geq s)])\tilde{Y}1(X\geq s)]$ and $1/P[U_k1(X\geq s)-P[U_k1(X\geq s)]]^2$. Specifically, for each of these functions, we will exhibit an m_0,m_1 such that the function belongs to $\mathcal{BV}(\mathcal{T},m_0,m_1)$.

We see that $s\mapsto P[\tilde{Y}1(X\geq s)]$ is uniformly bounded by $P|\tilde{Y}|=P|\delta X/G(X)|\leq\tau/G(\tau)<\tilde{M}_y$ and has total variation bounded by $\tilde{M}_y$; therefore $s\mapsto P[\tilde{Y}1(X\geq s)]\in\mathcal{BV}(\mathcal{T},\tilde{M}_y,\tilde{M}_y)$. Similarly, $s\mapsto P[U_k1(X\geq s)]$ is uniformly bounded by $\tilde{M}_u$ and has total variation bounded by $\tilde{M}_u$; thus, this function is in $\mathcal{BV}(\mathcal{T},\tilde{M}_u,\tilde{M}_u)$. Moreover, Lemma S6.1 gives that $s\mapsto P[(U_k-P[U_k1(X\geq s)])\tilde{Y}1(X\geq s)]\in\mathcal{BV}(\mathcal{T},2\tilde{M}_u\tilde{M}_y,3\tilde{M}_u\tilde{M}_y)$. Also, $s\mapsto 1/P[U_k1(X\geq s)-P[U_k1(X\geq s)]]^2$ belonging to $\mathcal{BV}(\mathcal{T},1/\varepsilon,3\tilde{M}_u^2/\varepsilon^2)$, using Lemma S6.1, that $s\mapsto P[U_k1(X\geq s)-P[U_k1(X\geq s)]]^2$ in $\mathcal{BV}(\mathcal{T},2\tilde{M}_u^2,3\tilde{M}_u^2)$ and that $P[U_k1(X\geq s)-P[U_k1(X\geq s)]]^2=\mathrm{Var}(U_k1(X\geq s))>\varepsilon>0$ by (A.6). Provided the sufficiently large $\tilde{M}_0$ and $\tilde{M}_1$, the above results and Lemma S6.1 implies that

$$\begin{aligned}
&P[\tilde{Y}1(X\geq s)]-\frac{\mathrm{Cov}(U_k1(X\geq s),\tilde{Y}1(X\geq s))}{\mathrm{Var}(U_k1(X\geq s))}P[U_k1(X\geq s)]\\
&\quad\in\mathcal{BV}\left(\mathcal{T},\tilde{M}_y+\frac{2\tilde{M}_u^2\tilde{M}_y}{\varepsilon},\tilde{M}_y+\frac{6\tilde{M}_u^4\tilde{M}_y}{\varepsilon^2}+\frac{5\tilde{M}_u^2\tilde{M}_y}{\varepsilon}\right)\subset\mathcal{BV}(\mathcal{T},\tilde{M}_0,\tilde{M}_1);\\
&\frac{\mathrm{Cov}(U_k1(X\geq s),\tilde{Y}1(X\geq s))}{\mathrm{Var}(U_k1(X\geq s))}\in\mathcal{BV}\left(\mathcal{T},\frac{2\tilde{M}_u\tilde{M}_y}{\varepsilon},\frac{6\tilde{M}_u^3\tilde{M}_y}{\varepsilon^2}+\frac{3\tilde{M}_u\tilde{M}_y}{\varepsilon}\right)\\
&\quad\subset\mathcal{BV}(\mathcal{T},\tilde{M}_0,\tilde{M}_1).
\end{aligned}$$

Fix $j\in\{q_n,\dots,n\}$. For $\hat{E}_j(\cdot,\cdot,k)$ to belong to the function class given in the lemma, we would need

$$\mathbb{P}_j[Y1(X\geq s)]-\frac{\mathrm{Cov}_{\mathbb{P}_j}(U_k1(X\geq s),Y1(X\geq s))}{\mathrm{Var}_{\mathbb{P}_j}(U_k1(X\geq s))}\mathbb{P}_j[U_k1(X\geq s)]$$

and $\mathrm{Cov}_{\mathbb{P}_j}(U_k 1(X \geq s), Y1(X \geq s))/\mathrm{Var}_{\mathbb{P}_j}(U_k 1(X \geq s))$ belonging to $\mathcal{BV}(\mathcal{T}, \tilde{M}_0, \tilde{M}_1)$. It thus suffices to show that, with probability tending to one, the following functions

$$\mathbb{P}_j[Y1(X \geq s)],\ \mathbb{P}_j[(U_k - \mathbb{P}_j[U_k 1(X \geq s)])Y1(X \geq s)],\ \mathbb{P}_j[U_k 1(X \geq s)],$$

and $1/\mathbb{P}_j[U_k 1(X \geq s) - \mathbb{P}_j[U_k 1(X \geq s)]]^2$ all belong to $\mathcal{BV}(\mathcal{T}, m_0, m_1)$, for suitable m_0 and m_1 that does not depend on j. This will be done by appealing to Lemma S6.1.

Below we show that $|\mathbb{P}_j[Y1(X \geq s)]| \leq \tilde{M}_y$ for any $s \in \mathcal{T}$ and $s \mapsto \mathbb{P}_j[Y1(X \geq s)]$ has total variation bounded by $\tilde{M}_y$ with probability tending to one.

$$|\mathbb{P}_j[Y1(X \geq s)]| \leq \mathbb{P}_j|Y1(X \geq s)| \leq \mathbb{P}_j|Y| \leq \sup_{t \in \mathcal{T}} \left| \frac{1}{\hat{G}_n(t)} - \frac{1}{G(t)} \right| \mathbb{P}_j|\delta X| + \mathbb{P}_j \left| \frac{\delta X}{G(X)} \right|$$

$$\leq \sup_{t \in \mathcal{T}} \left| \frac{1}{\hat{G}_n(t)} - \frac{1}{G(t)} \right| \big\{ (\mathbb{P}_j - P)|\delta X| + P|\delta X| \big\} + (\mathbb{P}_j - P)|\tilde{Y}| + P|\tilde{Y}| < \tilde{M}_y$$

with probability tending to one, using the weak law of large numbers. We also use $|\delta X| \leq \tau$, the uniform consistency of $\hat{G}_n$, $G(\tau) > 0$ by (A.2) and $P|\tilde{Y}| \leq \tau/G(\tau)$. Also, for an arbitrary partition of $\mathcal{T}$, say $\{t_1 < t_2 < \ldots < t_m < t_{m+1}\}$,

$$\sum_{j=1}^{m} \Big| \mathbb{P}_j[Y1(t_j \leq X < t_{j+1})] \Big| \leq \mathbb{P}_j \sum_{j=1}^{m} \Big| Y1(t_j \leq X < t_{j+1}) \Big| = \mathbb{P}_j|Y|$$

$$\leq \sup_{t \in \mathcal{T}} \left| \frac{1}{\hat{G}_n(t)} - \frac{1}{G(t)} \right| \mathbb{P}_j|\delta X| + \mathbb{P}_j \left| \frac{\delta X}{G(X)} \right|$$

$$\leq \sup_{t \in \mathcal{T}} \left| \frac{1}{\hat{G}_n(t)} - \frac{1}{G(t)} \right| \big\{ (\mathbb{P}_j - P)|\delta X| + P|\delta X| \big\} + (\mathbb{P}_j - P)|\tilde{Y}| + P|\tilde{Y}| < \tilde{M}_y$$

with probability tending to one, using the same arguments as the above. Taking a supremum over all partitions of $\mathcal{T}$ shows that the total variation of $s \mapsto \mathbb{P}_j[Y1(X \geq s)]$ is bounded by $\tilde{M}_y$ with probability tending to one. Thus $\mathbb{P}_j[Y1(X \geq s)]$ belongs to $\mathcal{BV}(\mathcal{T}, \tilde{M}_y, \tilde{M}_y)$ with probability tending to one.

Lemma S6.1 implies that, with probability tending to one, $s \mapsto \mathbb{P}_j[U_k 1(X \geq s)] \in \mathcal{BV}(\mathcal{T}, \tilde{M}_u, \tilde{M}_u)$; $s \mapsto \mathbb{P}_j[U_k 1(X \geq s) - \mathbb{P}_j[U_k 1(X \geq s)]]^2 \in \mathcal{BV}(\mathcal{T}, 2\tilde{M}_u^2, 3\tilde{M}_u^2)$, and $s \mapsto \mathbb{P}_j[(U_k - \mathbb{P}_j[U_k 1(X \geq s)])Y1(X \geq s)] \in \mathcal{BV}(\mathcal{T}, 2\tilde{M}_u\tilde{M}_y, 3\tilde{M}_u\tilde{M}_y)$.

Recall that (A.6) assumes $\mathrm{Var}(U_k 1(X \geq s))$ to be uniformly (over k, s) bounded away from zero, that is, $\min_{k,s} \mathrm{Var}(U_k 1(X \geq s)) > \varepsilon$ for some $\varepsilon \in (0,1)$. Then we see that with high probability, $\mathbb{P}_j[U_k 1(X \geq s) - \mathbb{P}_j[U_k 1(X \geq s)]]^2 = \mathrm{Var}_{\mathbb{P}_j}(U_k 1(X \geq s))$ is larger than or equal to ε and then bounded away from zero on $\mathcal{T}$. By Lemma S6.1, we have, with probability tending to one, $1/\mathbb{P}_j[U_k 1(X \geq s) - \mathbb{P}_j[U_k 1(X \geq s)]]^2 \in \mathcal{BV}(\mathcal{T}, 1/\varepsilon, 3\tilde{M}_u^2/\varepsilon^2)$.

Together with all the above results, Lemma S6.1 gives that with probability tending to one,

$$\mathbb{P}_j[Y1(X \geq s)] - \frac{\mathrm{Cov}_{\mathbb{P}_j}(U_k 1(X \geq s), Y1(X \geq s))}{\mathrm{Var}_{\mathbb{P}_j}(U_k 1(X \geq s))} \mathbb{P}_j[U_k 1(X \geq s)]$$

$$\in \mathcal{BV}(\mathcal{T}, \tilde{M}_y + \frac{2\tilde{M}_u^2\tilde{M}_y}{\varepsilon}, \tilde{M}_y + \frac{6\tilde{M}_u^4\tilde{M}_y}{\varepsilon^2} + \frac{5\tilde{M}_u^2\tilde{M}_y}{\varepsilon}) \subset \mathcal{BV}(\mathcal{T}, \tilde{M}_0, \tilde{M}_1);$$

$$\frac{\mathrm{Cov}_{\mathbb{P}_j}(U_k 1(X \geq s), Y1(X \geq s))}{\mathrm{Var}_{\mathbb{P}_j}(U_k 1(X \geq s))} \in \mathcal{BV}(\mathcal{T}, \frac{2\tilde{M}_u\tilde{M}_y}{\varepsilon}, \frac{6\tilde{M}_u^3\tilde{M}_y}{\varepsilon^2} + \frac{3\tilde{M}_u\tilde{M}_y}{\varepsilon})$$

$$\subset \mathcal{BV}(\mathcal{T}, \tilde{M}_0, \tilde{M}_1).$$

□

Let $\tilde\epsilon > 0$; define $\mathcal{G}$ to be the collection of monotone nonincreasing càdlàg functions $\tilde G\colon \mathcal{T} \to [0,1]$ such that $\tilde G(\tau) > \tilde\epsilon$, and let $\tilde M_0, \tilde M_1$ be the constants shown to exist in Lemma S6.2. To simplify the notation, we let $\mathcal{BV}(\mathcal{T}) \equiv \mathcal{BV}(\mathcal{T}, \tilde M_0, \tilde M_1)$. Below we use the notation $\boldsymbol{u} = (u_1, \ldots, u_p)$. For $k \in \mathbb{N}$, $(q, v, w) \in \{0,1,2\}^3$, and $r \in \{0,1,2,3,4\}$, define the function classes

(S6.39)

$$\tilde{\mathcal{F}}_1(k,q,r) = \left\{(x,\delta,\boldsymbol{u}) \;\mapsto\; u_k^r\left(\frac{\delta x}{\tilde G(x)}\right)^q 1(x \ge s) : \tilde G \in \mathcal{G},\ s \in \mathcal{T}\right\};$$

$$\tilde{\mathcal{E}}(k,w) = \left\{(\boldsymbol{u}, s) \;\mapsto\; \Big[a(s) + b(s)u_k\Big]^w : a, b \in \mathcal{BV}(\mathcal{T})\right\};$$

$$\tilde{\mathcal{F}}_2(k,v) = \left\{(x,\delta,\boldsymbol{u}) \;\mapsto\; \Big[\int_{\mathcal{T}} \big[a(s) + b(s)u_k\big] 1(x \ge s)\big\{1(x \in ds, \delta = 0) - \tilde\Lambda(s)\big\}\Big]^v : \right.$$

$$\left. \tilde\Lambda(s) = -\log(\tilde G(s));\ \tilde G \in \mathcal{G};\ a,\ b \in \mathcal{BV}(\mathcal{T})\right\};$$

$$\tilde{\mathcal{F}}(k,q,r,v,w) = \Big[\tilde{\mathcal{F}}_1(k,q,r)\tilde{\mathcal{E}}(k,w)\Big] \cup \Big[\tilde{\mathcal{F}}_1(k,q,r)\tilde{\mathcal{F}}_2(k,v)\Big],$$

where for two function classes $\mathcal{H}_1$ and $\mathcal{H}_2$, we let $\mathcal{H}_1\mathcal{H}_2 = \{h_1(\cdot)h_2(\cdot) : h_1 \in \mathcal{H}_1, h_2 \in \mathcal{H}_2\}$. Also, for $k \in \mathbb{N}$, let $\tilde{\mathcal{F}}(k) = \cup_{q=0}^2 \cup_{r=0}^4 \cup_{v=0}^2 \cup_{w=0}^2 \tilde{\mathcal{F}}(k,q,r,v,w)$. Let $\mathcal{K}_n = \{1, \ldots, p_n\}$. Henceforth we consider a large class $\tilde{\mathcal{F}}_n = \cup_{k \in \mathcal{K}_n} \tilde{\mathcal{F}}(k)$. In view of Lemma S6.2, we have $E_0(\cdot,\cdot,k) \in \tilde{\mathcal{E}}(k,1)$ and with probability tending to one, $\hat E_j(\cdot,\cdot,k) \in \tilde{\mathcal{E}}(k,1)$, $j = q_n, \ldots, n$.

LEMMA S6.3. *Suppose that (A.1) holds and let $\tilde M_u$ be a finite constant such that for all $k \in \mathbb{N}$, $|U_k| \le \tilde M_u$ with P-probability one. For any $k \in \mathbb{N}$ and $\tilde f \in \tilde{\mathcal{E}}(k,1)$, we have that $|\tilde f| \le \tilde M_0(1 + \tilde M_u)$.*

PROOF. This is an immediate consequence of the definition of $\tilde{\mathcal{E}}(k,1)$ and $\mathcal{BV}(\mathcal{T})$, along with the assumption (A.1). □

LEMMA S6.4. *Given $k \in \mathcal{K}_n$, $w \in \{1,2\}$ and the definition of $\tilde{\mathcal{E}}(k,w)$ above, we have $\tilde{\mathcal{E}}(k,w)$ is a Vapnik-Červonenkis (VC)-hull class for sets.*

PROOF. Fix $w = 1$. We start with showing that $\mathcal{BV}(\mathcal{T})$ is a VC-hull class for sets as follows. For any $f \in \mathcal{BV}(\mathcal{T})$ with $|f| \le \tilde M_0$, the Jordan decomposition indicates that $f = f^+ - f^-$, where (f^+, f^-) are the positive and negative parts of f, and both of them are positive, càdlàg and monotonic increasing on $\mathcal{T}$. Therefore, we can see f as the scalar multiple (by $\tilde M_0$) of the limit of the sequence

$$f_m = \sum_{j=1}^m \frac{1}{m}\Big[1\Big(f^+ > \frac{j}{m}\tilde M_0\Big) - 1\Big(f^- > \frac{j}{m}\tilde M_0\Big)\Big].$$

Then $f(\cdot)/\tilde M_0$ is in a class contained in the pointwise sequential closure of the symmetric convex hull of a class of indicator functions $\{1(f^+ > m'), m' \in \mathbb{R}^+\} \cup \{1(f^- > m'), m' \in \mathbb{R}^+\}$, which is a VC-subgraph class because

$$\{1(f^+(\cdot) > m'), m' \in \mathbb{R}^+\} \cup \{1(f^-(\cdot) > m'), m' \in \mathbb{R}^+\}$$
$$\subset \{1(f^+(\cdot) > m'), m' \in \mathbb{R}\} \cup \{1(f^-(\cdot) > m'), m' \in \mathbb{R}\}$$

and the union of the two classes on the right-hand-side forms a VC-subgraph class. Hence, $\mathcal{BV}(\mathcal{T})$ is (a $\tilde{M}_0$-fold rescaling of) a VC-hull class for sets.

For any given k, Lemma S6.3 indicates that any function $\tilde{f} \in \tilde{\mathcal{E}}(k,1)$ is uniformly bounded: $|\tilde{f}| \leq \tilde{M}_0(1+\tilde{M}_u)$. Following similar arguments to the above implies that $\tilde{f}/(\tilde{M}_0(1+\tilde{M}_u))$ is in a class contained in the pointwise sequential closure of the symmetric convex hull of

$$\{1(\tilde{f}^{+}(\cdot) > m'), m' \in \mathbb{R}\} \cup \{1(\tilde{f}^{-}(\cdot) > m'), m' \in \mathbb{R}\},$$

and therefore $\tilde{\mathcal{E}}(k,1)$ is (a $\tilde{M}_0(1+\tilde{M}_u)$-fold rescaling of) a VC-hull class for sets. When $w=2$, $\tilde{\mathcal{E}}(k,2)$ is the square of $\tilde{\mathcal{E}}(k,1)$, so it is also a VC-hull class for sets (Lemmas 2.6.20 in van der Vaart and Wellner, 1996). $\square$

As Lemma S6.2 indicates that for all j, $\hat{E}_j(\cdot,\cdot,k) \in \tilde{\mathcal{E}}(k,1)$ with probability tending to one and $E_0(\cdot,\cdot,k) \in \tilde{\mathcal{E}}(k,1)$, we could include $\mathrm{IF}_k^{CAR}(\cdot|P)$ in $\tilde{\mathcal{F}}_2(k,1)$. The following lemma shows that $\tilde{\mathcal{F}}_1(k,q,r)$, and $\tilde{\mathcal{F}}_2(k,v)$ are VC-hull classes for sets, for any (k,q,r,v).

LEMMA S6.5. *For all $k \in \mathcal{K}_n$, $r \in \{0,1,2,3,4\}$ and $(q,v,w) \in \{0,1,2\}^3$, all of the following are VC-hull classes for sets:*

$$\tilde{\mathcal{F}}_1(k,q,r),\ \tilde{\mathcal{F}}_2(k,v),\ \tilde{\mathcal{F}}_1(k,q,r)\tilde{\mathcal{E}}(k,w),\ \tilde{\mathcal{F}}_1(k,q,r)\tilde{\mathcal{F}}_2(k,v) \text{ and } \tilde{\mathcal{F}}(k,q,r,v,w).$$

Moreover, $\tilde{\mathcal{F}}(k)$ is a VC-hull class for sets.

PROOF. As observed, $\mathcal{G}$ is a VC-hull class for sets because any $\tilde{G} \in \mathcal{G}$ is the pointwise limit of the sequence $\tilde{G}_m = m^{-1}\sum_{j=1}^{m} 1(\tilde{G} \geq m^{-1}j)$, and a bounded VC-major class as well. By Lemma 2.6.19 of van der Vaart and Wellner (1996), $\{x \mapsto (1/\tilde{G}(x))^q : \tilde{G} \in \mathcal{G}\}$ is bounded VC-major. This equivalently implies that $\tilde{\mathcal{F}}_{11}(k,q) = \{(x,\delta,\boldsymbol{u}) \mapsto (1/\tilde{G}(x))^q : \tilde{G} \in \mathcal{G}\}$ is bounded VC-major. Note that $\tilde{\mathcal{F}}_{12}(k,q,r) = \{(x,\delta,\boldsymbol{u}) \mapsto u_k^r(\delta x)^q\}$ is bounded VC-major. Moreover, Lemma 2.6.13 of van der Vaart and Wellner (1996) implies that $\tilde{\mathcal{F}}_{11}(k,q)$ and $\tilde{\mathcal{F}}_{12}(k,q,r)$ are VC-hull classes for sets. Let $\tilde{\mathcal{F}}_{13}(k) = \{(x,\delta,\boldsymbol{u}) \mapsto 1(x \geq s) : s \in \mathcal{T}\}$, which is also a VC-hull class for sets. Therefore, we have $\tilde{\mathcal{F}}_1(k,q,r) = \tilde{\mathcal{F}}_{11}(k,q)\tilde{\mathcal{F}}_{12}(k,q,r)\tilde{\mathcal{F}}_{13}(k)$ as a VC-hull class for sets (Lemma 2.6.20 in van der Vaart and Wellner, 1996).

Below we show that $\tilde{\mathcal{F}}_2(k,v)$ is a VC-hull class for sets. Let $\{t_1 < t_2 < \ldots < t_m\}$ be an arbitrary partition of $\mathcal{T}$ with uniform increments $\Delta_t = t_{j+1} - t_j$ for all j. By Lemma S6.3, any function $\tilde{f} \in \tilde{\mathcal{E}}(k,1)$ is uniformly bounded: $|\tilde{f}| \leq \tilde{M}_0(1+\tilde{M}_u)$. Let $\tilde{M} = 4\tilde{M}_0(1+\tilde{M}_u)\sum_{j=1}^{m}|\tilde{\Lambda}(t_j+\Delta_t)|$. The integral in $\tilde{\mathcal{F}}_2(k,1)$ is the scalar multiple (by $\tilde{M}$) of the limit of the sequence

$$\begin{aligned}
&\sum_{j=1}^{m}\frac{\tilde{f}(t_j)}{\tilde{M}}\Big[1_{[t_j,t_{j+1})}(x)(1-\delta) - \tilde{\Lambda}(t_{j+1})1(x \geq t_{j+1}) + \tilde{\Lambda}(t_j)1(x \geq t_j)\Big] \\
&= \sum_{j=1}^{m}\frac{\tilde{f}(t_j)}{\tilde{M}}\Big[1_{[t_j,t_j+\Delta_t)}(x)(1-\delta) - \tilde{\Lambda}(t_j+\Delta_t)1(x \geq t_j+\Delta_t) + \tilde{\Lambda}(t_j)1(x \geq t_j)\Big] \\
&= \sum_{j=1}^{m}\sum_{r=1}^{4}\left\{\frac{\tilde{f}(t_j)}{\tilde{M}}(-1)^r\tilde{\Lambda}(t_j+\Delta_t 1_{r=3})^{1_{r\geq 3}}\right\}(1-\delta)^{1_{r\leq 2}}1\big(x \geq t_j + \Delta_t 1_{r\in\{2,3\}}\big) \\
&\equiv \sum_{j=1}^{m}\sum_{r=1}^{4}\alpha_{jr}(1-\delta)^{1_{r\leq 2}}1\big(x \geq t_j + \Delta_t 1_{r\in\{2,3\}}\big),
\end{aligned}$$

where $\sum_{j=1}^{m}\sum_{r=1}^{4}|\alpha_{jr}| \leq 1$ by

$$\sum_{j=1}^{m}\sum_{r=1}^{4}|\alpha_{jr}| \leq \sum_{j=1}^{m}\sum_{r=1}^{4}\left|\frac{\tilde{f}(t_j)}{\tilde{M}}(-1)^r\tilde{\Lambda}(t_j+\Delta_t 1_{r=3})^{1_{r\geq 3}}\right| \leq \sum_{j=1}^{m}\sum_{r=1}^{4}\left|\frac{\tilde{f}(t_j)}{\tilde{M}}\tilde{\Lambda}(t_j+\Delta_t)\right|$$
$$\leq \frac{4\tilde{M}_0(1+\tilde{M}_u)}{\tilde{M}}\sum_{j=1}^{m}\left|\tilde{\Lambda}(t_j+\Delta_t)\right| = 1.$$

Then any integral function in $\tilde{\mathcal{F}}_2(k,1)$ is in a class contained in the scalar-multiplied pointwise sequential closure of the symmetric convex hull of the VC-subgraph class $\{(x,\delta)\mapsto \delta 1(x\geq s) : s\in\mathcal{T}\}$, which is a class of indicator functions. Hence $\tilde{\mathcal{F}}_2(k,1)$ is a VC-hull class for sets, and so is $\tilde{\mathcal{F}}_2(k,2)$ because it is the square of $\tilde{\mathcal{F}}_2(k,1)$ (Lemma 2.6.20 in van der Vaart and Wellner, 1996).

Together with the fact that $\tilde{\mathcal{E}}(k,w)$ is shown VC-hull for sets as in Lemma S6.4, repetitively applying Lemma 2.6.20 of van der Vaart and Wellner (1996) further indicates that $\tilde{\mathcal{F}}_1(k,q,r)\tilde{\mathcal{E}}(k,w)$ and $\tilde{\mathcal{F}}_1(k,q,r)\tilde{\mathcal{F}}_2(k,v)$ are VC-hull classes for sets. Thanks to the preservation properties of VC-hull classes for sets, $\tilde{\mathcal{F}}(k,q,r,v,w)$, the union of $\tilde{\mathcal{F}}_1(k,q,r)\tilde{\mathcal{E}}(k,w)$ and $\tilde{\mathcal{F}}_1(k,q,r)\tilde{\mathcal{F}}_2(k,v)$, is a VC-hull class for sets. Analogously, $\tilde{\mathcal{F}}(k)$ is also a VC-hull class for sets. □

By the assumption (A.1) that the U_k are uniformly bounded, there exists a uniform upper bound $\tilde{M}_2 > 0$ for $\sum_{r=0}^{4}|U_k|^r$. As we will now show, the following is an envelope function for $\tilde{\mathcal{F}}(k)$ for all k:

$$F : (x,\delta,\boldsymbol{u}) \mapsto \tilde{M}_2\left\{\sum_{q=0}^{2}\left|\frac{\delta x}{\tilde{\epsilon}}\right|^q\right\}\left\{\sum_{w=0}^{2}(\tilde{M}_0(1+\tilde{M}_u))^w\Big[1+\big[1-\delta-\log(\tilde{\epsilon})\big]^2\Big]\right\}.$$

First we show that F is an envelope function for $\tilde{\mathcal{F}}(k,2,4,2,2)$; similar arguments apply to the other classes involving different values of (q,r,v,w). For any function $f\in\tilde{\mathcal{F}}(k,2,4,2,2)$ depending on $\tilde{G}\in\mathcal{G}$ with $\tilde{\Lambda} = -\log(\tilde{G})$, we see that, for each (δ,x,u_k), $|f(\delta,x,u_k)|$ is bounded by

$$\left|\frac{u_k^4(\delta x)^2}{\tilde{G}^2(\tau)}\right|\left\{[a(\cdot)+b(\cdot)u_k]^2+\Big[\int_{\mathcal{T}}[a(s)+b(s)u_k]\big\{1(x\in ds,\delta=0)-1(x\geq s)\tilde{\Lambda}(s)\big\}\Big]^2\right\}$$
$$\leq \tilde{M}_2\left|\frac{(\delta x)^2}{\tilde{G}^2(\tau)}\right|\left\{(\tilde{M}_0(1+\tilde{M}_u))^2\Big[1+\big[1-\delta+\tilde{\Lambda}(\tau)\big]^2\Big]\right\}$$
$$\leq \tilde{M}_2\left|\frac{\delta x}{\tilde{\epsilon}}\right|^2\left\{(\tilde{M}_0(1+\tilde{M}_u))^2\Big[1+\big[1-\delta-\log(\tilde{\epsilon})\big]^2\Big]\right\} \leq F,$$

where the first inequality holds by seeing that $[a(\cdot)+b(\cdot)u_k]\in\tilde{\mathcal{E}}(k,1)$ so that $|a(\cdot)+b(\cdot)u_k| \leq \tilde{M}_0(1+\tilde{M}_u)$; that $[1-\delta+\tilde{\Lambda}(\tau)]$ is the total variation of the signed measure $1(x\in ds,\delta=0)-1(x\geq s)\tilde{\Lambda}(ds)$, and that

$$\left|\int_{\mathcal{T}}[a(s)+b(s)u_k]\big\{1(x\in ds,\delta=0)-1(x\geq s)\tilde{\Lambda}(s)\big\}\right| \leq \tilde{M}_0(1+\tilde{M}_u)\big[1-\delta+\tilde{\Lambda}(\tau)\big].$$

The second inequality follows by $\tilde{G}(\tau) > \tilde{\epsilon} > 0$.

Because the U_k is uniformly bounded in P-probability, $|\delta X| \leq \tau$ P-almost surely and $\tilde{G}(\tau) > \tilde{\epsilon} > 0$, F is square-integrable: $\|F\|_{Q,2}^2 = \int F^2 dQ < \infty$ for any probability measure Q on the sample space $\mathcal{X}$. Therefore, for all $k\geq 1$, Theorem 2.6.9 of van der Vaart and

Wellner (1996) indicates there exists a universal constant K that does not depend on k and $a \in (0,2)$ so that

$$\sup_Q \log N(\epsilon\|F\|_{Q,2},\ \tilde{\mathcal{F}}(k),\ L_2(Q)) \le K\epsilon^{-a}.$$

Moreover, the above display implies that

$$N(\epsilon\|F\|_{Q,2},\ \tilde{\mathcal{F}}_n,\ L_2(Q)) \le \sum_{k\in\mathcal{K}_n} N(\epsilon\|F\|_{Q,2},\ \tilde{\mathcal{F}}(k),\ L_2(Q)) \le p_n \exp(K\epsilon^{-a}),$$

giving that

$$\sup_Q \log N(\epsilon\|F\|_{Q,2},\ \tilde{\mathcal{F}}_n,\ L_2(Q)) \le \log(p_n) + K\epsilon^{-a}. \tag{S6.40}$$

For $j = 1, \ldots, n$, define the empirical process $\{\mathbb{G}_j(\tilde{f}) : \tilde{f} \in \tilde{\mathcal{F}}_n\}$ pointwise as follows

$$\mathbb{G}_j(\tilde{f}) = \frac{1}{\sqrt{j}} \sum_{i=1}^{j} \left[\tilde{f}(O_i) - P(\tilde{f})\right] = \sqrt{j}(\mathbb{P}_j - P)\tilde{f},$$

where $\mathbb{P}_j$ denotes the empirical distribution of $O_1, \ldots, O_j$. Let $\|\mathbb{G}_j\|_{\tilde{\mathcal{F}}_n} = \sup_{\tilde{f}\in\tilde{\mathcal{F}}_n} |\mathbb{G}_j(\tilde{f})|$. Following (S6.40), Theorem 2.14.1 of van der Vaart and Wellner (1996) gives $\|\mathbb{G}_j\|_{\tilde{\mathcal{F}}_n} \lesssim \sqrt{\log(p_n)}$, so that

$$\sup_{\tilde{f}\in\tilde{\mathcal{F}}_n} \left|(\mathbb{P}_j - P)\tilde{f}\right| \lesssim \sqrt{\log(p_n)/j},$$

where $\lesssim$ means "bounded above up to a universal multiplicative constant that does not depend on (j, n)." We also need the following lemmas.

LEMMA S6.6. *For any sample size n, the event $\mathcal{A}_n$ occurs with probability at least $1 - 1/n$, where $\mathcal{A}_n = \cap_{j=1}^n \mathcal{A}_{nj}$, $\mathcal{A}_{nj} = \{\sup_{\tilde{f}\in\tilde{\mathcal{F}}_n} \left|(\mathbb{P}_j - P)\tilde{f}\right| \lesssim K_{nj}\}$ with $K_{nj} \equiv \sqrt{\log(n \vee p_n)/j}$ and $n \vee p_n = \sup(n, p_n)$.*

PROOF. Follow the same argument as in the proof of Lemma A.4 of Luedtke and van der Laan (2018), except based on the class $\tilde{\mathcal{F}}_n$. □

Note that Lemma S6.6 reduces in the special case of $\cup_{k\in\mathcal{K}_n} \cup_{r=0}^4 \tilde{\mathcal{F}}_1(k, 0, r, 0, 0) \in \tilde{\mathcal{F}}_n$, to $\mathbb{P}_j$ and P being replaced by $\mathbb{Q}_j$ (the empirical distribution of $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_j$) and Q_u, respectively.

Let $K_{nj} = \sqrt{\log(n \vee p_n)/j}$, for $j = q_n, \ldots, n$. For $\tilde{K} \in (0, \infty)$, define

$$\mathcal{B}_n(\tilde{K}) = \Bigg\{ \sup_{t\in\mathcal{T}} \left|\frac{\hat{G}_n(t)}{G(t)} - 1\right| \le \sqrt{\frac{\log n}{n}},\ \sup_{t\in\mathcal{T}} \left|\hat{\Lambda}_n(t) - \Lambda(t)\right| \le \sqrt{\frac{\log n}{n}},\ \inf_{s\in\mathcal{T}} Y_n(s) \ge \sqrt{n},$$

$$\sup_{(k,s)\in\mathcal{K}_n\times\mathcal{T}} \left|\hat{E}_j(U_k, s, k) - E_0(U_k, s, k)\right| \le \tilde{K}K_{nj},\ j = q_n, \ldots, n \Bigg\},$$

where $Y_n(s) = \sum_{i=1}^n 1(X_i \ge s)$ and $\hat{\Lambda}_n(\cdot) = \int_{-\infty}^{\cdot} \left[1(Y_n(s) > 0)/Y_n(s)\right] dN_n(s)$ is the estimator of $\Lambda(\cdot)$. The following lemma concerns the probability of the event $\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K})$.

LEMMA S6.7. *Under the conditions of Theorem 4.1, there exists a $\tilde{K} \in (0, \infty)$ such that* $\mathrm{P}(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K})) \to 1$.

PROOF. Fix $\tilde{K}$. It holds that $\mathrm{P}\left(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K})\right) = \mathrm{P}\left(\mathcal{A}_n\right) - \mathrm{P}\left(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K})^c\right)$. Hence, by Lemma S6.6, we have that:

$$\mathrm{P}\left(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K})\right) \geq 1 - 1/n - \mathrm{P}\left(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K})^c\right). \tag{S6.41}$$

It therefore remains to show that, for any appropriate choice of $\tilde{K}$, $\mathrm{P}\left(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K})^c\right) = o(1)$. To show this, we use that $\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K})^c \subseteq \cup_{j=1}^3 \mathcal{R}_{nj}$, where

$$\begin{aligned}
\mathcal{R}_{n1} &\equiv \left\{\sup_{t\in\mathcal{T}} \left|\frac{\hat{G}_n(t)}{G(t)} - 1\right| > \sqrt{\frac{\log n}{n}}\right\} \cup \left\{\sup_{t\in\mathcal{T}} \left|\hat{\Lambda}_n(t) - \Lambda(t)\right| > \sqrt{\frac{\log n}{n}}\right\}, \\
\mathcal{R}_{n2} &\equiv \left\{\inf_{s\in\mathcal{T}} Y_n(s) < \sqrt{n}\right\}, \\
\mathcal{R}_{n3} &\equiv \mathcal{A}_n \cap \left\{\sup_{t\in\mathcal{T}} \left|\frac{\hat{G}_n(t)}{G(t)} - 1\right| \leq \sqrt{\frac{\log n}{n}}\right\} \\
&\qquad \cap \bigcup_{j=q_n}^{n} \left\{\sup_{(k,s)\in\mathcal{K}_n\times\mathcal{T}} \left|\hat{E}_j(U_k, s, k) - E_0(U_k, s, k)\right| > \tilde{K} K_{nj}\right\}.
\end{aligned}$$

By a union bound, this yields that $\mathrm{P}(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K})^c) \leq \sum_{h=1}^3 P(\mathcal{R}_{nh})$. In the remainder of this proof, we will establish the following three facts

$$\mathrm{P}(\mathcal{R}_{n1}) = o(1),\ \ \mathrm{P}(\mathcal{R}_{n2}) = o(1) \ \text{ and } \ \mathrm{P}(\mathcal{R}_{n3}) = 0. \tag{S6.42}$$

Combining these facts with (S6.41) will then yield the result.

We first show that $\mathrm{P}(\mathcal{R}_{n1}) \to 0$. Observing that $\{\sqrt{n}[\hat{G}_n(t)/G(t) - 1] : t \in \mathcal{T}\}$ converges to a tight Gaussian process, and then the continuous mapping theorem gives that $\sqrt{n}\sup_{t\in\mathcal{T}} |\hat{G}_n(t)/G(t) - 1|$ converges to the supremum of the absolute value of this Gaussian process. Therefore for any sequence $\varepsilon_n \to \infty$, $\mathrm{P}(\sup_{t\in\mathcal{T}} \sqrt{n}|\hat{G}_n(t)/G(t) - 1| > \varepsilon_n\big) \to 0$, in particular, $\varepsilon_n = \sqrt{\log n}$. The identical argument applies to yield that $\mathrm{P}(\sup_{t\in\mathcal{T}} \sqrt{n}|\hat{\Lambda}_n(t) - \Lambda(t)| > \varepsilon_n\big) \to 0$. Consequently, $\mathrm{P}(\mathcal{R}_{n1}) = o(1)$.

To obtain $\mathrm{P}(\mathcal{R}_{n2}) \to 0$, observe that $Y_n(\tau) \sim \text{Binomial}(n, p_*)$ with $p_* = \mathrm{P}(X \geq \tau)$ and $\mathrm{P}(X \geq \tau) > 0$ by (A.5). Along with the monotonicity of $Y_n(s)$ in $s \in \mathcal{T}$, Hoeffding's inequality gives

$$\mathrm{P}(\mathcal{R}_{n2}) \leq \mathrm{P}\left(\inf_{s\in\mathcal{T}} Y_n(s) \leq \sqrt{n}\right) = \mathrm{P}\left(Y_n(\tau) \leq \sqrt{n}\right) \leq \exp(-2(\sqrt{n}p_* - 1)^2) \to 0.$$

In the remainder of the proof, we show that $\mathrm{P}(\mathcal{R}_{n3}) = 0$. It suffices to show that

$$\begin{aligned}
&\mathcal{A}_n \cap \left\{\sup_{t} \left|\frac{\hat{G}_n(t)}{G(t)} - 1\right| \leq \sqrt{\frac{\log n}{n}}\right\} \\
&\subseteq \bigcap_{j=q_n}^{n} \left\{\sup_{(k,s)\in\mathcal{K}_n\times\mathcal{T}} \left|\hat{E}_j(U_k, s, k) - E_0(U_k, s, k)\right| \leq \tilde{K} K_{nj}\right\},
\end{aligned} \tag{S6.43}$$

giving that $\mathcal{R}_{n3} = \emptyset$ and then in turn $\mathrm{P}(\mathcal{R}_{n3}) = 0$.

For the rest of the proof, suppose that $\mathcal{A}_n$ occurs and $\sup_t |\hat{G}_n(t)/G(t) - 1| \leq \sqrt{\log(n)/n}$. To simplify notation, let $Y^s = Y1(X \geq s)$; $\tilde{Y}^s = \tilde{Y}1(X \geq s)$ and $U_k^s = U_k 1(X \geq s)$. Taylor expanding Y^s with respect to $\hat{G}_n$ around G gives

$$\begin{aligned}
\sup_s \Big\{ P|Y^s - \tilde{Y}^s| \Big\} &\le \sum_{r=1}^{\infty} \Big[\sup_t \Big| \frac{\hat{G}_n(t)}{G(t)} - 1 \Big| \Big]^r \sup_s \Big\{ P\big|\tilde{Y}^s\big| \Big\} \\
&\le \sup_s \Big\{ P\big|\tilde{Y}^s\big| \Big\} \sum_{r=1}^{\infty} \Big[\sqrt{\frac{\log n}{n}} \Big]^r \le P\big|\tilde{Y}\big| \sum_{r=1}^{\infty} \Big[\sqrt{\frac{\log n}{n}} \Big]^r \\
&= \sqrt{\frac{\log n}{n}} P\big|\tilde{Y}\big| \sum_{r=1}^{\infty} \Big[\sqrt{\frac{\log n}{n}} \Big]^{r-1} \lesssim \sqrt{\frac{\log n}{n}}, \qquad \text{(S6.44)}
\end{aligned}$$

where the last steps holds by $\sqrt{\log n/n} < 1$ such that $\sum_{r=1}^{\infty} [\sqrt{\log n/n}\,]^{r-1} < \infty$, and $|\tilde{Y}| = |\delta X/G(X)|$ is a bounded random variable. Similarly, the triangle inequality gives

$$\begin{aligned}
\sup_{(k,s)} \big|\mathrm{Cov}(U_k^s, Y^s - \tilde{Y}^s)\big| &\le \sup_{(k,s)} \Big\{ P|U_k^s(Y^s - \tilde{Y}^s)| + P|U_k^s| P|Y^s - \tilde{Y}^s| \Big\} \\
&\le \sup_{(k,s)} \Big\{ P\big|U_k^s \tilde{Y}^s\big| + P|U_k^s| P\big|\tilde{Y}^s\big| \Big\} \sum_{r=1}^{\infty} \Big[\sup_t \Big| \frac{\hat{G}_n(t)}{G(t)} - 1 \Big| \Big]^r \\
&\le \sup_{(k,s)} \Big\{ P\big|U_k^s \tilde{Y}^s\big| + P|U_k^s| P\big|\tilde{Y}^s\big| \Big\} \sum_{r=1}^{\infty} \Big[\sqrt{\frac{\log n}{n}} \Big]^r \\
&\le \sqrt{\frac{\log n}{n}} \max_k \Big\{ P\big|U_k \tilde{Y}\big| + P|U_k| P\big|\tilde{Y}\big| \Big\} \sum_{r=1}^{\infty} \Big[\sqrt{\frac{\log n}{n}} \Big]^{r-1} \lesssim \sqrt{\frac{\log n}{n}}. \qquad \text{(S6.45)}
\end{aligned}$$

By the triangle inequality,

$$\begin{aligned}
&\sup_{(k,s)} \big|\mathrm{Cov}_{\mathbb{P}_j}(U_k^s, Y^s) - \mathrm{Cov}(U_k^s, Y^s)\big| \\
&\quad = \sup_{(k,s)} \Big\{ \big| \mathbb{P}_j[U_k^s Y^s] - \mathbb{P}_j[U_k^s]\mathbb{P}_j[Y^s] - P[U_k^s Y^s] + P[U_k^s]P[Y^s] \big| \Big\} \\
&\quad \le \sup_{(k,s)} \Big\{ \big|(\mathbb{P}_j - P)[U_k^s Y^s]\big| + \big|(\mathbb{P}_j - P)[U_k^s]\big| \big|(\mathbb{P}_j - P)[Y^s]\big| \\
&\qquad\qquad + \big|P[U_k^s]\big| \big|(\mathbb{P}_j - P)[Y^s]\big| + \big|P[Y^s]\big| \big|(\mathbb{P}_j - P)[U_k^s]\big| \Big\},
\end{aligned}$$

and based on the above display, we have

$$\sup_{(k,s)} \big|\mathrm{Cov}_{\mathbb{P}_j}(U_k^s, Y^s) - \mathrm{Cov}(U_k^s, Y^s)\big| \lesssim K_{nj}, \qquad \text{(S6.46)}$$

when $\mathcal{A}_n$ occurs, since the U_k is assumed to be uniformly bounded by (A.1) and $|Y^s|$ is bounded as implied by (A.2) and $|X| \le \tau$, along with $\log(n \vee p_n)/j \le 1$ so that $\log(n \vee p_n)/j \le K_{nj}$, for $j = q_n, \ldots, n$.

Regarding the assumption in (A.6) that $\mathrm{Var}(U_k^s)$ is uniformly bounded away from zero, there exists $\tilde{\zeta} \in (0, \infty)$ so that $\min_{(k,s)} \mathrm{Var}(U_k^s) > \tilde{\zeta} > 0$. Meanwhile, let $\tilde{\eta}$ be the smallest universal positive constant to maintain $\sup_{(k,s)} |\mathrm{Var}_{\mathbb{P}_j}(U_k^s) - \mathrm{Var}(U_k^s)| \le \tilde{\eta} K_{nj}$ that is implied by the occurrence of $\mathcal{A}_n$. Let $q_n \ge \lceil 4\tilde{\eta}^2 \log(n \vee p_n)/\tilde{\zeta}^2 \rceil$ and

$$\begin{aligned}
\mathrm{Var}_{\mathbb{P}_j}(U_k^s) &= \mathrm{Var}(U_k^s) + \mathrm{Var}_{\mathbb{P}_j}(U_k^s) - \mathrm{Var}(U_k^s) \\
&\ge \mathrm{Var}(U_k^s) - \sup_{(k,s)} |\mathrm{Var}_{\mathbb{P}_j}(U_k^s) - \mathrm{Var}(U_k^s)| \ge \mathrm{Var}(U_k^s) - \tilde{\eta} K_{nj}
\end{aligned}$$

$$\geq \min_{(k,s)} \mathrm{Var}(U_k^s) - \tilde{\zeta}/2 > \tilde{\zeta}/2, \ \ \forall\, k,\, s.$$

This yields

$$\sup_{(k,s)} \left| \frac{1}{\mathrm{Var}_{\mathbb{P}_j}(U_k^s)} - \frac{1}{\mathrm{Var}(U_k^s)} \right| \leq \frac{\sup_{(k,s)} |\mathrm{Var}_{\mathbb{P}_j}(U_k^s) - \mathrm{Var}(U_k^s)|}{[\min_{(k,s)} \mathrm{Var}_{\mathbb{P}_j}(U_k^s)][\min_{(k,s)} \mathrm{Var}(U_k^s)]} \lesssim K_{nj}. \tag{S6.47}$$

Using the results in (S6.44)-(S6.47) gives

$$\begin{aligned}
&\sup_{(k,s)} \Big| \big(U_k^s - \mathbb{P}_j[U_k^s]\big)\mathrm{Cov}_{\mathbb{P}_j}(U_k^s, Y^s) - \big(U_k^s - P[U_k^s]\big)\mathrm{Cov}(U_k^s, \tilde{Y}^s) \Big| \\
&\quad \leq \sup_{(k,s)} \Big\{ \big|U_k^s - P[U_k^s]\big| \big|\mathrm{Cov}_{\mathbb{P}_j}(U_k^s, Y^s) - \mathrm{Cov}(U_k^s, Y^s)\big| \\
&\qquad\qquad + \big|(\mathbb{P}_j - P)[U_k^s]\big| \big|\mathrm{Cov}_{\mathbb{P}_j}(U_k^s, Y^s) - \mathrm{Cov}(U_k^s, Y^s)\big| \\
&\qquad\qquad + \big|U_k^s - P[U_k^s]\big| \big|\mathrm{Cov}(U_k^s, Y^s - \tilde{Y}^s)\big| + \big|(\mathbb{P}_j - P)[U_k^s]\big| \big|\mathrm{Cov}(U_k^s, Y^s - \tilde{Y}^s)\big| \\
&\qquad\qquad + \big|(\mathbb{P}_j - P)[U_k^s]\big| \big|\mathrm{Cov}(U_k^s, \tilde{Y}^s)\big| \Big\} \\
&\quad \lesssim K_{nj},
\end{aligned}$$

following the bounded values of $|\tilde{Y}^s|$ and that of $|U_k^s|$ over k, which are implied by (A.1)–(A.2) and $|X| \leq \tau$. As stated, suppose that $\hat{E}_j$ is fitted via the linear regression approach described in Remark 3.5 and as defined in (S6.38). Applying similar arguments and the above results yields

$$\begin{aligned}
&\sup_{(k,s)} \big|\hat{E}_j(U_k, s, k) - E_0(U_k, s, k)\big| \leq \sup_{(k,s)} \Big\{ \big|(\mathbb{P}_j - P)[Y_s]\big| + P|Y_s - \tilde{Y}_s| \\
&\qquad + \frac{1}{\mathrm{Var}(U_{k,s})} \Big| \big(U_{k,s} - \mathbb{P}_j[U_{k,s}]\big)\mathrm{Cov}_{\mathbb{P}_j}(U_{k,s}, Y_s) - \big(U_{k,s} - P[U_{k,s}]\big)\mathrm{Cov}(U_{k,s}, \tilde{Y}_s) \Big| \\
&\qquad + \Big| \frac{1}{\mathrm{Var}_{\mathbb{P}_j}(U_{k,s})} - \frac{1}{\mathrm{Var}(U_{k,s})} \Big| \Big| \big(U_{k,s} - \mathbb{P}_j[U_{k,s}]\big)\mathrm{Cov}_{\mathbb{P}_j}(U_{k,s}, Y_s) \\
&\qquad\qquad\qquad\qquad - \big(U_{k,s} - P[U_{k,s}]\big)\mathrm{Cov}(U_{k,s}, \tilde{Y}_s) \Big| \\
&\qquad + \Big| \big(U_{k,s} - P[U_{k,s}]\big)\mathrm{Cov}(U_{k,s}, \tilde{Y}_s) \Big| \Big| \frac{1}{\mathrm{Var}_{\mathbb{P}_j}(U_{k,s})} - \frac{1}{\mathrm{Var}(U_{k,s})} \Big| \Big\} \\
&\quad \lesssim K_{nj} + \sqrt{\log n / j} \leq \tilde{K} K_{nj},
\end{aligned}$$

for some constant $\tilde{K} \in (1, \infty)$ that does not depend on (j, n), leading to

$$\sup_{(k,s)} \big|\hat{E}_j(U_k, s, k) - E_0(U_k, s, k)\big| \leq \tilde{K} K_{nj}. \tag{S6.48}$$

Hence we have shown that (S6.43) holds and conclude the proof. □

LEMMA S6.8. *Let $\tilde{K}$ be given in Lemma S6.7 and $\mathcal{I}_n \equiv \{(j, k) : j \in \{q_n, \ldots, n\}, k \in \mathcal{K}_n\}$. Suppose the conditions of Theorem 4.1 hold, and introduce the event $\mathcal{C}_n$ that occurs when*

(S6.8.1) $\max_{k \in \mathcal{K}_n} |\mathrm{Var}_{\mathbb{Q}_j}(U_k) - \mathrm{Var}_{Q_u}(U_k)| \lesssim K_{nj}$,

(S6.8.2) $\min_{(j,k)\in\mathcal{I}_n} \mathrm{Var}_{\mathbb{Q}_j}(U_k)$ *is bounded away from zero,*
(S6.8.3) $\max_{k\in\mathcal{K}_n} \left|(\mathbb{Q}_j - Q_u)\left[U_k^r\{E_0(U_k,k)\}^w\right]\right| \lesssim K_{nj}$, $0\le r,w\le 1$,
(S6.8.4) $\max_{k\in\mathcal{K}_n} \left|1/\mathrm{Var}_{\mathbb{Q}_j}(U_k) - 1/\mathrm{Var}_{Q_u}(U_k)\right| \lesssim K_{nj}$,
(S6.8.5) $\sup_{(k,s)\in\mathcal{K}_n\times\mathcal{T}} \left|\hat{E}_j(U_k,s,k)\right| \lesssim K_{nj} + \sup_{(k,s)\in\mathcal{K}_n\times\mathcal{T}} |E_0(U_k,s,k)|$,
(S6.8.6) $\max_{k\in\mathcal{K}_n} |U_k - \mathbb{Q}_j[U_k]| \lesssim K_{nj} + \max_{k\in\mathcal{K}_n} |U_k - Q_u[U_k]|$,

where each of (S6.8.1)–(S6.8.6) relies on appropriately specified constants that do not depend on (j,n). *Then such constants exist such that* $\mathrm{P}(\mathcal{A}_n\cap\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_n)\to 1$.

PROOF. When $\mathcal{A}_n$ occurs, the triangle inequality gives that

$$\begin{aligned}
&\max_{k\in\mathcal{K}_n} |\mathrm{Var}_{\mathbb{Q}_j}(U_k) - \mathrm{Var}_{Q_u}(U_k)| \\
&\quad = \max_{k\in\mathcal{K}_n} \left|(\mathbb{Q}_j[U_k]^2 - Q_u[U_k]^2) - (\mathbb{Q}_j[U_k] + Q_u[U_k])(\mathbb{Q}_j[U_k] - Q_u[U_k])\right| \\
&\quad \lesssim \max_{k\in\mathcal{K}_n} \left|\mathbb{Q}_j[U_k]^2 - Q_u[U_k]^2\right| + \max_{k\in\mathcal{K}_n} \left|\mathbb{Q}_j[U_k] - Q_u[U_k]\right| \lesssim K_{nj}.
\end{aligned}$$

Let $\mathcal{C}_{n1}$ correspond to the event (S6.8.1) using the (j,n)-independent constant implied by the above display, which gives that $\mathrm{P}(\mathcal{A}_n\cap\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_{n1}^c)\to 0$.

By (A.6), there exists $\zeta>0$ such that $\min_{k\in\mathcal{K}_n}\mathrm{Var}_{Q_u}(U_k)\ge\zeta$. Moreover, if $\mathcal{C}_{n1}$ holds, then there exists $\eta>0$ such that for all $n\in\mathbb{N}$ and $(j,k)\in\mathcal{I}_n$, $|\mathrm{Var}_{\mathbb{Q}_j}(U_k)-\mathrm{Var}_{Q_u}(U_k)|\le \eta K_{nj}$. This yields, for all n and all $(j,k)\in\mathcal{I}_n$,

$$\begin{aligned}
\mathrm{Var}_{\mathbb{Q}_j}(U_k) &= \mathrm{Var}_{Q_u}(U_k) + \mathrm{Var}_{\mathbb{Q}_j}(U_k) - \mathrm{Var}_{Q_u}(U_k) \\
&\ge \mathrm{Var}_{Q_u}(U_k) - |\mathrm{Var}_{\mathbb{Q}_j}(U_k) - \mathrm{Var}_{Q_u}(U_k)| \ge \zeta - \eta K_{nj}.
\end{aligned}$$

By the conditions of Theorem 4.1, $\sqrt{q_n}/\log(n\vee p_n)\to\infty$ and let $q_n/\log(n\vee p_n)\ge 4\eta^2/\zeta^2$ for all n sufficiently large. Hence, the above shows that, for all such n, $\mathrm{Var}_{\mathbb{Q}_j}(U_k)\ge\zeta/2$ when $\mathcal{A}_n\cap\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_{n1}$ occurs. Letting $\mathcal{C}_{n2}$ denote the event that $\min_{(j,k)\in\mathcal{I}_n}\mathrm{Var}_{\mathbb{Q}_j}(U_k)\ge \zeta/2$, and we show that $\mathrm{P}(\mathcal{A}_n\cap\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_{n2}^c)\to 0$.

We could regard (S6.8.3) as a consequence of the occurrence of $\mathcal{A}_n$ because the function $u_k\mapsto u_k^r\{E_0(u_k,k)\}^w$ belongs to $\tilde{\mathcal{F}}_n$. Let $\mathcal{C}_{n3}$ correspond to the event (S6.8.3), and $\mathrm{P}(\mathcal{A}_n\cap\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_{n3}^c)\to 0$. Along with the fact that $\min_{k\in\mathcal{K}_n}\mathrm{Var}_{Q_u}(U_k)$ is bounded away from zero that is implied by (A.6), using (S6.8.1) and (S6.8.2) of Lemma S6.8 gives

$$\max_{k\in\mathcal{K}_n}\left|\frac{1}{\mathrm{Var}_{\mathbb{Q}_j}(U_k)} - \frac{1}{\mathrm{Var}_{Q_u}(U_k)}\right| = \frac{\max_{k\in\mathcal{K}_n}|\mathrm{Var}_{\mathbb{Q}_j}(U_k)-\mathrm{Var}_{Q_u}(U_k)|}{[\min_{k\in\mathcal{K}_n}\mathrm{Var}_{\mathbb{Q}_j}(U_k)][\min_{k\in\mathcal{K}_n}\mathrm{Var}_{Q_u}(U_k)]} \lesssim K_{nj}.$$

Let $\mathcal{C}_{n4}$ correspond to the event (S6.8.4), and $\mathcal{C}_{n1}\cap\mathcal{C}_{n2}\subseteq\mathcal{C}_{n4}$, which gives $\mathrm{P}(\mathcal{A}_n\cap\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_{n4}^c)\to 0$.

According to (A.7), $\sup_{(k,s)}|E_0(U_k,s,k)|$ is uniformly bounded by some (k,s,n)-independent constant. Therefore when $\mathcal{B}_n(\tilde{K})$ occurs, we have $\sup_{(k,s)}\left|\hat{E}_j(U_k,s,k) - E_0(U_k,s,k)\right|\le\tilde{K}K_{nj}$ for all j, implying that

$$\sup_{(k,s)}\left|\hat{E}_j(U_k,s,k)\right| \le \tilde{K}K_{nj} + \sup_{(k,s)}\left|E_0(U_k,s,k)\right| \lesssim K_{nj} + \sup_{(k,s)}\left|E_0(U_k,s,k)\right|.$$

Thus, letting $\mathcal{C}_{n5}$ denote the event that

$$\sup_{(k,s)}|\hat{E}_j(U_k,s,k)| \lesssim K_{nj} + \sup_{(k,s)}\left|E_0(U_k,s,k)\right|,\ j=q_n,\ldots,n,$$

we have $\mathrm{P}(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_{n5}^c) \to 0$. Similarly, we have that $\max_k |U_k - Q_u[U_k]|$ is uniformly bounded by some finite constant that does not depend on (k, n), by (A.1). Therefore when $\mathcal{A}_n$ occurs,

$$\begin{aligned}\max_k |U_k - \mathbb{Q}_j[U_k]| &\le \max_k |(\mathbb{Q}_j - Q_u)[U_k]| + \max_k |U_k - Q_u[U_k]| \\ &\lesssim K_{nj} + \max_k |U_k - Q_u[U_k]|.\end{aligned}$$

Let $\mathcal{C}_{n6}$ denote the event that for $j = q_n, \ldots, n$, $\max_k |U_k - \mathbb{Q}_j[U_k]| \lesssim K_{nj} + \max_k |U_k - Q_u[U_k]|$ and $\mathrm{P}(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_{n6}^c) \to 0$. Letting $\mathcal{C}_n \equiv \cap_{q=1}^6 \mathcal{C}_{nq}$, we have shown $\mathrm{P}(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n) \ge 1 - \sum_{q=1}^6 \mathrm{P}(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_{nq}^c) \to 1$. $\square$

For some constant $\tilde{K}$ as given in Lemma S6.7,

$$d_n(P_1, P_2) \equiv P\Big[\max_{k \in \mathcal{K}_n} |\,\mathrm{IF}_k^*(O|P_1) - \mathrm{IF}_k^*(O|P_2)| 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n}\Big].$$

Recall that $\hat{P}_{nj} = (\hat{E}_j, \mathbb{Q}_j, \hat{G}_n)$ and $\hat{P}_j' = (\hat{E}_j, \mathbb{Q}_j, G)$ as defined in Section S1.

LEMMA S6.9. *Suppose the conditions of Theorem 4.1 hold. Then with $\tilde{K} \in (0, \infty)$ given in Lemma S6.7, there exists $K' \in (1, \infty)$ such that $\mathrm{P}(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \mathcal{D}_n(K')) \to 1$, where*

$$\mathcal{D}_n(K') = \big\{d_n(\hat{P}_{nj}, \hat{P}_j') \vee d_n(\hat{P}_j', P) \le K' K_{nj},\ j = q_n, \ldots, n\big\}.$$

PROOF. Let

$$L_{njk} = \left|\frac{(U_k - \mathbb{Q}_j[U_k])}{\mathrm{Var}_{\mathbb{Q}_j}(U_k)} \int_{\mathcal{T}} \hat{E}_j(U_k, s, k)(d\hat{M}(s) - dM(s))\right| 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n}. \tag{S6.49}$$

The triangle inequality first gives the decomposition

(S6.50)

$$\begin{aligned}|\,\mathrm{IF}_k^*(O|\hat{P}_{nj}) - \mathrm{IF}_k^*(O|\hat{P}_j')| 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n} &= \left|\frac{(U_k - \mathbb{Q}_j[U_k])\delta X}{\mathrm{Var}_{\mathbb{Q}_j}(U_k)}\left(\frac{1}{\hat{G}_n(X)} - \frac{1}{G(X)}\right)\right. \\ &\quad \left. + \frac{(U_k - \mathbb{Q}_j[U_k])}{\mathrm{Var}_{\mathbb{Q}_j}(U_k)} \int_{\mathcal{T}} \hat{E}_j(U_k, s, k)(d\hat{M}(s) - dM(s))\right| 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n} \\ &\le \left|\frac{(U_k - \mathbb{Q}_j[U_k])\delta X}{\mathrm{Var}_{\mathbb{Q}_j}(U_k)}\left(\frac{1}{\hat{G}_n(X)} - \frac{1}{G(X)}\right)\right| 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n} + L_{njk}.\end{aligned}$$

Moreover,

(S6.51)

$$\begin{aligned}&\max_{k \in \mathcal{K}_n} \left|\frac{(U_k - \mathbb{Q}_j[U_k])\delta X}{\mathrm{Var}_{\mathbb{Q}_j}(U_k)}\left(\frac{1}{\hat{G}_n(X)} - \frac{1}{G(X)}\right)\right| 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n} \\ &\le 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n} \sum_{r=1}^{\infty} \max_{k \in \mathcal{K}_n} \left|\frac{(U_k - \mathbb{Q}_j[U_k])}{\mathrm{Var}_{\mathbb{Q}_j}(U_k)} \tilde{Y}\right| \left[\sup_t \left|\frac{\hat{G}_n(t)}{G(t)} - 1\right|\right]^r \\ &\le \max_{k \in \mathcal{K}_n} \left|\frac{(U_k - \mathbb{Q}_j[U_k])}{\mathrm{Var}_{\mathbb{Q}_j}(U_k)} \tilde{Y}\right| 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n} \sum_{r=1}^{\infty} \left[\sqrt{\frac{\log n}{n}}\right]^r \lesssim [K_{nj} + 1]\sqrt{\frac{\log n}{n}} \le K'' K_{nj},\end{aligned}$$

where the first inequality holds by Taylor expansion and the triangle inequality, and the second inequality results from $\sup_t |\hat{G}_n(t)/G(t) - 1| \le \sqrt{\log n/n}$ when $\mathcal{B}_n(\tilde{K})$ occurs. The last two steps in the above display hold because there exists some (j,n)-independent constant $K'' \in (1,\infty)$ such that

$$\max_{k\in\mathcal{K}_n}\left|\frac{(U_k - \mathbb{Q}_j[U_k])}{\mathrm{Var}_{\mathbb{Q}_j}(U_k)}\tilde{Y}\right|\sum_{r=1}^{\infty}\left[\sqrt{\frac{\log n}{n}}\right]^r \lesssim [K_{nj}+1]\sqrt{\frac{\log n}{n}} \le K'' K_{nj},$$

following $|\tilde{Y}| = |\delta X/G(X)| < \infty$ and $\max_k\{|U_k - \mathbb{Q}_j[U_k]|/\mathrm{Var}_{\mathbb{Q}_j}(U_k)\} \lesssim K_{nj}+1$, according to (A.1), (S6.8.2) and (S6.8.6) of Lemma S6.8 in view of the occurrence of $\mathcal{C}_n$, and $\sum_{r=1}^{\infty}[\sqrt{\log n/n}]^{r-1} < \infty$. In addition, below we show $\mathrm{P}\big(\max_j\{P[\max_{k\in\mathcal{K}_n} L_{njk}]\} > 1/\sqrt{\log(n\vee p_n)}\big) = o(1)$. By Markov's inequality and Jensen's inequality,

(S6.52)

$$\mathrm{P}\Big(\max_j\Big\{P\Big[\max_{k\in\mathcal{K}_n} L_{njk}\Big]\Big\} > 1/\sqrt{\log(n\vee p_n)}\Big) \le \log(n\vee p_n) E\Big\{\max_j\Big\{P\Big[\max_{k\in\mathcal{K}_n} L_{njk}\Big]\Big\}^2\Big\}$$

$$= \log(n\vee p_n)\mathbb{E}\Big\{\max_j\Big\{P\Big[\max_{k\in\mathcal{K}_n} L_{njk}\Big]\Big\}^2\Big\} \le \log(n\vee p_n)\mathbb{E}\Big\{P\Big[\max_{(j,k)} L_{njk}^2\Big]\Big\}.$$

Recall $N_n(s)$ and $Y_n(s)$ as defined in Section S1 except for removing u; $d\bar{M}(s) \equiv dN_n(s) - Y_n(s)d\Lambda(s)$ is a local martingale with respect to the aggregated filtration

$$\mathcal{F}'_s = \sigma(\{N_n(s'), Y_n(s'), s' \le s \in \mathcal{T}\}, \{\boldsymbol{U}_i\}_{i=1}^n). \tag{S6.53}$$

Applying the decomposition $d\hat{M}(s) - dM(s) = 1(X \ge s)(d\Lambda(s) - d\hat{\Lambda}_n(s))$ to the expression of L_{njk} in (S6.49) and using the inequalities $(a-b)^2 \le 2(a^2+b^2)$ and $1_{\mathcal{A}_n\cap\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_n} \le 1_{\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_n}$ further bound $\mathbb{E}\{P[\max_{(j,k)} L_{njk}^2]\}$ above by the sum (multiplied by 2) of

$$\mathbb{E}\Big\{P\Big[\max_{(j,k)}\Big\{\frac{(U_k - \mathbb{Q}_j[U_k])}{\mathrm{Var}_{\mathbb{Q}_j}(U_k)}\int_{\mathcal{T}}\big[\hat{E}_j(U_k,s,k) - E_0(U_k,s,k)\big]1(X\ge s)\{d\hat{\Lambda}_n(s) - d\Lambda(s)\}\Big\}^2 \times 1_{\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_n}\Big]\Big\}$$

and

$$\mathbb{E}\Big\{P\Big[\max_{(j,k)}\Big\{\frac{(U_k - \mathbb{Q}_j[U_k])}{\mathrm{Var}_{\mathbb{Q}_j}(U_k)}\int_{\mathcal{T}} E_0(U_k,s,k)1(X\ge s)\{d\hat{\Lambda}_n(s) - d\Lambda(s)\}\Big\}^2 1_{\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_n}\Big]\Big\},$$

as respectively given in (S6.54) and (S6.55) below.

First note that $\min_{(j,k)}\mathrm{Var}_{\mathbb{Q}_j}(U_k)$ is bounded away from zero in view of the occurrence of $\mathcal{C}_n$, as shown in (S6.8.2) of Lemma S6.8. Along with (A.1) and that the total variation of $\hat{\Lambda}_n - \Lambda$ over $\mathcal{T}$ is bounded by $|\hat{\Lambda}_n(\tau) + \Lambda(\tau)|$, we have that

$$\begin{aligned}
&\max_{(j,k)}\Big\{\frac{(U_k - \mathbb{Q}_j[U_k])}{\mathrm{Var}_{\mathbb{Q}_j}(U_k)}\int_{\mathcal{T}}\big[\hat{E}_j - E_0\big](U_k,s,k)1(X\ge s)\{d\hat{\Lambda}_n(s) - d\Lambda(s)\}\Big\}^2 1_{\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_n}\\
&\lesssim \sup_{(j,k,s)}\big|\hat{E}_j(U_k,s,k) - E_0(U_k,s,k)\big|^2\big|\hat{\Lambda}_n(\tau) + \Lambda(\tau)\big|^2 1_{\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_n}\\
&\lesssim \sup_{(j,k,s)}\big|\hat{E}_j(U_k,s,k) - E_0(U_k,s,k)\big|^2\big\{\big|\hat{\Lambda}_n(\tau) - \Lambda(\tau)\big|^2 + \Lambda^2(\tau)\big\}1_{\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_n}\\
&\lesssim K_{nq_n}^2\Big[\frac{\log(n)}{n} + \Lambda^2(\tau)\Big],
\end{aligned}$$

where the last line follows the occurrence of $\mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n$ that implies for each j, $\sup_{(k,s)} \big|\hat{E}_j(U_k,s,k) - E_0(U_k,s,k)\big| \le \tilde{K} K_{nj}$ and $\sup_t |\hat{\Lambda}_n(t) - \Lambda(t)| \le \sqrt{\log(n)/n}$, giving that $\sup_{(j,k,s)} \big|\hat{E}_j(U_k,s,k) - E_0(U_k,s,k)\big| \le \tilde{K} K_{nq_n}$ because $K_{nj} = \sqrt{\log(n \vee p_n)/j}$ for $j \ge q_n$ and $\max_j\{K_{nj}\} = K_{nq_n}$. Therefore, we have that

(S6.54)
$$\mathbb{E}\Big\{P\Big[\max_{(j,k)}\Big\{\frac{(U_k - \mathbb{Q}_j[U_k])}{\mathrm{Var}_{\mathbb{Q}_j}(U_k)}\int_{\mathcal{T}}\big[\hat{E}_j(U_k,s,k) - E_0(U_k,s,k)\big]1(X \ge s)\{d\hat{\Lambda}_n(s) - d\Lambda(s)\}\Big\}^2$$
$$\times 1_{\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_n}\Big]\Big\} \lesssim K^2_{nq_n}\Big[\frac{\log(n)}{n} + \Lambda^2(\tau)\Big].$$

Moreover, we observe the decomposition of $\hat{\Lambda}_n - \Lambda$ analogously to (S5.21) without u:
$$d\hat{\Lambda}_n(t) - d\Lambda(t) = 1(Y_n(t) > 0)Y_n(t)^{-1}d\bar{M}(t) - 1(Y_n(t) = 0)\,d\Lambda(t).$$

Note that the occurrence of $\mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n$ eliminates $1(Y_n(t) = 0)$, so using (A.1) and the occurrence of $\mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n$ similarly gives the upper bound:
$$\max_{(j,k)}\Big\{\frac{(U_k - \mathbb{Q}_j[U_k])}{\mathrm{Var}_{\mathbb{Q}_j}(U_k)}\int_{\mathcal{T}} E_0(U_k,s,k)1(X \ge s)\{d\hat{\Lambda}_n(s) - d\Lambda(s)\}\Big\}^2 1_{\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_n}$$
$$\lesssim \max_k\Big\{\int_{\mathcal{T}} E_0(U_k,s,k)1(X \ge s)\frac{1(Y_n(s) > 0)}{Y_n(s)}d\bar{M}(s)\Big\}^2 1_{\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_n}.$$

Therefore,

(S6.55)
$$\mathbb{E}\Big\{P\Big[\max_{(j,k)}\Big\{\frac{(U_k - \mathbb{Q}_j[U_k])}{\mathrm{Var}_{\mathbb{Q}_j}(U_k)}\int_{\mathcal{T}} E_0(U_k,s,k)1(X \ge s)\{d\hat{\Lambda}_n(s) - d\Lambda(s)\}\Big\}^2 1_{\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_n}\Big]\Big\}$$
$$\lesssim \mathbb{E}\Big\{P\Big[\max_k\Big\{\int_{\mathcal{T}} E_0(U_k,s,k)1(X \ge s)\frac{1(Y_n(s) > 0)}{Y_n(s)}d\bar{M}(s)\Big\}^2\Big]1_{\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_n}\Big\}$$
$$\le P\Big[\max_k \mathbb{E}\Big\{\int_{\mathcal{T}} E_0(U_k,s,k)1(X \ge s)\frac{1(Y_n(s) \ge \sqrt{n})}{Y_n(s)}d\bar{M}(s)\Big\}^2\Big] \lesssim \frac{\Lambda(\tau)}{\sqrt{n}},$$
where along with (A.1) that U_k is uniformly bounded on k and (A.7) that E_0 is uniformly bounded, the last inequality holds by using the quadratic variation with respect to the filtration $\mathcal{F}'_s$ that is defined in (S6.53):
$$\mathbb{E}\Big\{\int_{\mathcal{T}} E_0(U_k,s,k)1(X \ge s)\frac{1(Y_n(s) \ge \sqrt{n})}{Y_n(s)}d\bar{M}(s)\Big\}^2$$
$$= \int_{\mathcal{T}}\mathbb{E}\Big\{E_0^2(U_k,s,k)1(X \ge s)\frac{1(Y_n(s) \ge \sqrt{n})}{Y_n(s)}\Big\}d\Lambda(s) \lesssim \frac{\Lambda(\tau)}{\sqrt{n}}.$$
Collecting the results in (S6.52), (S6.54) and (S6.55) leads to
$$\mathrm{P}\Big(\max_j\Big\{P\Big[\max_{k\in\mathcal{K}_n} L_{njk}\Big]\Big\} > 1/\sqrt{\log(n \vee p_n)}\Big)$$
$$\lesssim \log(n \vee p_n)\Big[K^2_{nq_n}\Big[\frac{\log(n)}{n} + \Lambda^2(\tau)\Big] + \frac{\Lambda(\tau)}{\sqrt{n}}\Big] = o(1),$$

following that $\Lambda(\tau)$ is bounded, $K^2_{nq_n} = \log(n \vee p_n)/q_n$, $q_n^{1/4}/\log(n \vee p_n) \to \infty$ and $n/q_n = O(1)$.

Taking the maximum over $k \in \mathcal{K}_n$, applying the triangle inequality, and then taking the expectation with respect to P on both sides of (S6.50), we see that (S6.51) implies that $d(\hat{P}_{nj}, \hat{P}_j') \le K'' K_{nj} + P[\max_{k\in\mathcal{K}_n} L_{njk}]$. Hence, by the above,

$$
\begin{aligned}
&\mathrm{P}\big(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \big\{d(\hat{P}_{nj}, \hat{P}_j') \le K'' K_{nj} + 1/\sqrt{\log(n \vee p_n)},\, j = q_n, \dots, n\big\}\big) \\
&\ge \mathrm{P}\Big(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \Big\{\max_j \Big\{P\Big[\max_{k\in\mathcal{K}_n} L_{njk}\Big]\Big\} \le 1/\sqrt{\log(n \vee p_n)}\Big\}\Big) \\
&= \mathrm{P}\big(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n\big) \\
&\quad - \mathrm{P}\Big(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \Big\{\max_j \Big\{P\Big[\max_{k\in\mathcal{K}_n} L_{njk}\Big]\Big\} > 1/\sqrt{\log(n \vee p_n)}\Big\}\Big) \\
&\ge \mathrm{P}\big(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n\big) - \mathrm{P}\Big(\max_j \Big\{P\Big[\max_{k\in\mathcal{K}_n} L_{njk}\Big]\Big\} > 1/\sqrt{\log(n \vee p_n)}\Big) \to 1,
\end{aligned}
$$

leading to

(S6.56)
$$
\mathrm{P}\big(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \big\{d(\hat{P}_{nj}, \hat{P}_j') \le K'' K_{nj} + 1/\sqrt{\log(n \vee p_n)},\, j = q_n, \dots, n\big\}^c\big) \to 0.
$$

In addition for each j, we have that

$$
\begin{aligned}
&\max_{k\in\mathcal{K}_n} \big|\mathrm{IF}_k^*(O|\hat{P}_j') - \mathrm{IF}_k^*(O|P)\big| 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n} \\
&\le 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n} \Big\{ \frac{1}{\min_k \mathrm{Var}_{\mathbb{Q}_j}(U_k)} \Big(\tilde{Y} \max_k \big|(\mathbb{Q}_j - Q_u)[U_k]\big| \\
&\qquad + \max_k |U_k| \max_k \big|(\mathbb{Q}_j - Q_u)[\hat{E}_j(U_k,k)]\big| \\
&\qquad + \max_k |U_k| Q_u\big[\max_k \big|\hat{E}_j(U_k,k) - E_0(U_k,k)\big|\big] \\
&\qquad + \mathbb{Q}_j\big[\max_k \big|\hat{E}_j(U_k,k)\big|\big] \max_k \big|(\mathbb{Q}_j - Q_u)[U_k]\big| \\
&\qquad + \max_k |Q_u[U_k]| \max_k \big|(\mathbb{Q}_j - Q_u)[\hat{E}_j(U_k,k)]\big| \\
&\qquad + \max_k |Q_u[U_k]| Q_u\big[\max_k \big|\hat{E}_j(U_k,k) - E_0(U_k,k)\big|\big] \Big) \\
&\quad + \max_k \Big| \frac{1}{\mathrm{Var}_{\mathbb{Q}_j}(U_k)} - \frac{1}{\mathrm{Var}_{Q_u}(U_k)} \Big| \max_k \big|(U_k - Q_u[U_k])(\tilde{Y} - Q_u[E_0(U_k,k)])\big| \\
&\quad + \frac{\max_k |\mathrm{Cov}_{\mathbb{Q}_j}(U_k, \hat{E}_j(U_k,k))|}{\min_k \mathrm{Var}^2_{\mathbb{Q}_j}(U_k)} \Big(2 \max_k |U_k| \max_k |(\mathbb{Q}_j - Q_u)[U_k]| \\
&\qquad + \max_k |(\mathbb{Q}_j - Q_u)[U_k]| \max_k |(\mathbb{Q}_j + Q_u)[U_k]| \Big) \\
&\quad + \frac{\max_k (U_k - Q_u[U_k])^2}{\min_k \mathrm{Var}^2_{\mathbb{Q}_j}(U_k)} \Big(\max_k \big|\mathrm{Cov}_{\mathbb{Q}_j}(U_k, \hat{E}_j(U_k,k)) - \mathrm{Cov}_{Q_u}(U_k, E_0(U_k,k))\big| \Big) \\
&\quad + \max_k \Big| \frac{1}{\mathrm{Var}^2_{\mathbb{Q}_j}(U_k)} - \frac{1}{\mathrm{Var}^2_{Q_u}(U_k)} \Big| \max_k |\mathrm{Cov}_{Q_u}(U_k, E_0(U_k,k))| \max_k (U_k - Q_u[U_k])^2 \\
&\quad + \frac{(1+\Lambda(\tau))}{\min \mathrm{Var}_{\mathbb{Q}_j}(U_k)} \Big(\max_k |U_k - Q_u[U_k]| \sup_{(k,s)} |\hat{E}_j(U_k,s,k) - E_0(U_k,s,k)|
\end{aligned}
$$

$$+\max_k |(\mathbb{Q}_j - Q_u)[U_k]| \max_k |E_0(U_k,s,k)|\Big)$$

$$+\max_k \left|\frac{1}{\mathrm{Var}_{\mathbb{Q}_j}(U_k)} - \frac{1}{\mathrm{Var}_{Q_u}(U_k)}\right| \max_k |U_k - Q_u[U_k]| \max_k |E_0(U_k,s,k)|(1+\Lambda(\tau))\Big\}$$

$$\leq K^{'''} K_{nj}$$

for some (j,k,s,n)-independent constant $K^{'''} \in (0,\infty)$ when $\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n$ occurs, by (A.1)–(A.5) and (A.6)–(A.7). This gives

$$\mathrm{P}\big(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \big\{d(\hat{P}_j', P) \leq K^{'''} K_{nj},\ j = q_n,\ldots,n\big\}^c\big) \to 0. \tag{S6.57}$$

Then taking $K' = (K^{''}+1) \vee K^{'''}$ and using (S6.56) and (S6.57), we have $\mathrm{P}(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \mathcal{D}_n(K')^c) \to 0$. Hence the result follows by Lemma S6.8. $\square$

In what follows, we let $\tilde{\sigma}_{nj}^2 \equiv \mathrm{Var}(\mathrm{IF}_{k_j}^*(O|\hat{P}_{nj})|O_1,\ldots,O_j)$.

LEMMA S6.10. *Let $\tilde{K} \in (0,\infty)$ be given in Lemma S6.7. Then, under the conditions of Theorem 4.1, on the event $\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n$ we have $|\hat{\sigma}_{nj}^2 - \tilde{\sigma}_{nj}^2| \lesssim K_{nj}$ almost surely for $j = q_n,\ldots,n$ and n sufficiently large.*

PROOF. Fix $j \in \{q_n,\ldots,n\}$. We first have that

$$|\hat{\sigma}_{nj}^2 - \tilde{\sigma}_{nj}^2| 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n} \leq \bigg\{\Big|\mathbb{P}_j\big[\mathrm{IF}_{k_j}^*(O|\hat{P}_{nj})\big]^2 - P\big[\mathrm{IF}_{k_j}^*(O|\hat{P}_{nj})\big]^2\Big|$$

$$+\Big|\big[\mathbb{P}_j\,\mathrm{IF}_{k_j}^*(O|\hat{P}_{nj})\big]^2 - \big[P\,\mathrm{IF}_{k_j}^*(O|\hat{P}_{nj})\big]^2\Big|\bigg\} 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n}$$

$$\leq \bigg\{\max_{k\in\mathcal{K}_n}\Big|(\mathbb{P}_j - P)\big[\mathrm{IF}_k^*(O|\hat{P}_{nj})\big]^2\Big|$$

$$+\max_{k\in\mathcal{K}_n}\Big|(\mathbb{P}_j - P)\big[\mathrm{IF}_k^*(O|\hat{P}_{nj})\big]\,(\mathbb{P}_j + P)\big[\mathrm{IF}_k^*(O|\hat{P}_{nj})\big]\Big|\bigg\} 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n}.$$

At the conclusion of this proof, we'll show that there exists a constant $\bar{K}$ that does not depend on n such that

$$\max_{(j,k)}\Big|\mathrm{IF}_k^*(o|\hat{P}_{nj}) 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n}\Big| \leq \bar{K} \tag{S6.58}$$

for all observations o in the support of P. This will simplify the above display as

(S6.59)

$$|\hat{\sigma}_{nj}^2 - \tilde{\sigma}_{nj}^2| 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n}$$

$$\lesssim \bigg\{\max_{k\in\mathcal{K}_n}\Big|(\mathbb{P}_j - P)\big[\mathrm{IF}_k^*(O|\hat{P}_{nj})\big]^2\Big| + \max_{k\in\mathcal{K}_n}\Big|(\mathbb{P}_j - P)\big[\mathrm{IF}_k^*(O|\hat{P}_{nj})\big]\Big|\bigg\} 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n},$$

from which we continue showing that $|\hat{\sigma}_{nj}^2 - \tilde{\sigma}_{nj}^2| 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n} \lesssim K_{nj}$.

The proof below proceeds with assuming (without statement) that the event $\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n$ occurs. Recall that $\mathrm{IF}_k^*(\cdot|E,Q,G)$ denotes the function $\mathrm{IF}_k^*(\cdot|P_{E,Q,G})$, where $P_{E,Q,G}$ is a distribution with the conditional residual life function E, the marginal distribution Q of any given single predictor, and the censoring distribution G. For each (k,j), $\mathrm{IF}^*_k(o|\hat{E}_j, Q_u, \hat{G}_n)$ belongs to $\tilde{\mathcal{F}}_n$, observing that $\mathrm{IF}^*_k(o|\hat{E}_j, Q_u, \hat{G}_n)$ is equal to

$$\frac{(u_k - Q_u[U_k])\big(\delta x/\hat{G}_n(x) - Q_u[\hat{E}_j(U_k,k)]\big)}{\mathrm{Var}_{Q_u}(U_k)} - \frac{\mathrm{Cov}_{Q_u}(U_k, \hat{E}_j(U_k,k))}{\mathrm{Var}^2_{Q_u}(U_k)}(u_k - Q_u[U_k])^2$$
$$+ \frac{(u_k - Q_u[U_k])}{\mathrm{Var}_{Q_u}(U_k)} \int_{\mathcal{T}} \hat{E}_j(u_k, s, k)\{1(x \in ds, \delta = 0) - 1(x \geq s) d\hat{\Lambda}_n(s)\},$$

in which the first two terms are contained in $\cup_{q=0}^{1} \cup_{r=0}^{4} \tilde{\mathcal{F}}_1(k,q,r)\tilde{\mathcal{E}}(k,1)$, and the last term is contained in $\cup_{r=0}^{4}\tilde{\mathcal{F}}_1(k,0,r)\tilde{\mathcal{F}}_2(k,1)$. To take advantage of the fact that $\mathrm{IF}^*_k(o|\hat{E}_j, Q_u, \hat{G}_n) \in \mathcal{F}_n$, we replace $\mathrm{IF}^*_k(O|\hat{P}_{nj})$ in (S6.59) by $\mathrm{IF}^*_k(O|\hat{E}_j, Q_u, \hat{G}_n)$ and then (S6.59) turns into

$$\text{(S6.60)} \qquad |\hat{\sigma}^2_{nj} - \tilde{\sigma}^2_{nj}| 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n} \lesssim \Big\{ \max_{k \in \mathcal{K}_n} \Big| (\mathbb{P}_j - P)\big[\mathrm{IF}^*_k(O|\hat{E}_j, Q_u, \hat{G}_n)\big]^2 \Big|$$
$$+ \max_{k \in \mathcal{K}_n} \Big| (\mathbb{P}_j - P)\big[\mathrm{IF}^*_k(O|\hat{E}_j, Q_u, \hat{G}_n)\big] \Big| \Big\} 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n} + K_{nq_n},$$

following that on the event $\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n$,

$$\text{(S6.61)} \qquad \max_{(j,k)} \big| \mathrm{IF}^*_k(o|\hat{P}_{nj}) - \mathrm{IF}^*_k(o|\hat{E}_j, Q_u, \hat{G}_n) \big| \lesssim K_{nq_n},$$

and that there exists a constant $\bar{K}_0$ that does not depend on n such that

$$\text{(S6.62)} \qquad \max_{(j,k)} \big| \mathrm{IF}^*_k(o|\hat{E}_j, Q_u, \hat{G}_n) 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n} \big| \leq \bar{K}_0$$

for all observations o in the support of P. We show (S6.61) and (S6.62) in the sequel.

For (S6.61), first we have that

$$\text{(S6.63)} \quad \big| \mathrm{IF}^*_k(o|\hat{P}_{nj}) - \mathrm{IF}^*_k(o|\hat{E}_j, Q_u, \hat{G}_n) \big|$$
$$= \big| \mathrm{IF}^*_k(o|\hat{E}_j, \mathbb{Q}_j, \hat{G}_n) - \mathrm{IF}^*_k(o|\hat{E}_j, Q_u, \hat{G}_n) \big| \lesssim |(\mathbb{Q}_j - Q_u)[U_k]|\, |\tau/G(\tau)|$$
$$+ |(\mathbb{Q}_j - Q_u)[U_k]|\, |\tau/\hat{G}_n(\tau) - \tau/G(\tau)|$$
$$+ \big|(\mathbb{Q}_j - Q_u)[U_k]\big|\, \big|\mathbb{Q}_j[\hat{E}_j(U_k,k)]\big| + \big|(\mathbb{Q}_j - Q_u)[\hat{E}_j(U_k,k)]\big|\, \big|Q_u[U_k]\big|$$
$$+ \Big| \mathrm{Cov}_{\mathbb{Q}_j}(U_k, \hat{E}_j(U_k,k)) - \mathrm{Cov}_{Q_u}(U_k, \hat{E}_j(U_k,k)) \Big| (u_k - \mathbb{Q}_j[U_k])^2$$
$$+ \Big| (u_k - \mathbb{Q}_j[U_k])^2 - (u_k - Q_u[U_k])^2 \Big| \big| \mathrm{Cov}_{Q_u}(U_k, \hat{E}_j(U_k,k)) \big|$$
$$+ |(\mathbb{Q}_j - Q_u)[U_k]| \, \Big| \int_{\mathcal{T}} \hat{E}_j(u_k, s, k)\{1(x \in ds, \delta = 0) - 1(x \geq s) d\hat{\Lambda}_n(s)\} \Big|.$$

The inequality in (S6.63) holds because for all j, $\min_k \mathrm{Var}_{\mathbb{Q}_j}(U_k)$ is bounded away from zero given the occurrence of $\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n$ as shown in (S6.8.2) of Lemma S6.8 and $\min_k \mathrm{Var}_{Q_u}(U_k) > 0$ by (A.6). We continue expanding (S6.61) in what follows. Together with (A.1) that the support of U_k is uniformly bounded and (A.2) that $G(\tau) > 0$, a Taylor expansion of $\hat{G}_n$ around G and the occurrence of $\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n$ imply that

(S6.64)
$$\left| \frac{\tau}{\hat{G}_n(\tau)} - \frac{\tau}{G(\tau)} \right| \leq \frac{|\tau|}{G(\tau)} \sum_{r=1}^{\infty} \left[\sup_t \left| \frac{\hat{G}_n(t)}{G(t)} - 1 \right| \right]^r \leq \frac{|\tau|}{G(\tau)} \sum_{r=1}^{\infty} \left[\sqrt{\frac{\log n}{n}} \right]^r \lesssim \sqrt{\frac{\log n}{n}}.$$

From (A.1) and (A.7), we have that $\sup_{(k,s)}\left|E_0(U_k,s,k)\right| \le K_1$ for some positive finite constant K_1. Along with (S6.8.5) of Lemma S6.8, it gives that $\max_k\left|\mathbb{Q}_j[\hat{E}_j(U_k,k)]\right| \le \sup_{(k,s)}\left|\hat{E}_j(U_k,s,k)\right| \lesssim K_{nj}+K_1$. Then together with (A.1), these two results lead to

(S6.65)
$$\begin{aligned}
&\left|\mathrm{Cov}_{\mathbb{Q}_j}(U_k,\hat{E}_j(U_k,k)) - \mathrm{Cov}_{Q_u}(U_k,\hat{E}_j(U_k,k))\right| = \left|(\mathbb{Q}_j - Q_u)[U_k\hat{E}_j(U_k,k)]\right| \\
&\quad + \left|\mathbb{Q}_j[\hat{E}_j(U_k,k)]\right|\left|(\mathbb{Q}_j - Q_u)[U_k]\right| + \left|Q_u[U_k]\right|\left|(\mathbb{Q}_j - Q_u)[\hat{E}_j(U_k,k)]\right| \\
&\quad \lesssim \left|(\mathbb{Q}_j - Q_u)[U_k\hat{E}_j(U_k,k)]\right| + (K_{nj}+1)\left|(\mathbb{Q}_j - Q_u)[U_k]\right| + \left|(\mathbb{Q}_j - Q_u)[\hat{E}_j(U_k,k)]\right|
\end{aligned}$$

and

(S6.66)
$$\begin{aligned}
&\left|\int_{\mathcal{T}}\hat{E}_j(u_k,s,k)\{1(x\in ds,\delta=0) - 1(x\ge s)d\hat{\Lambda}_n(s)\}\right| \le \sup_{(k,s)}\left|\hat{E}_j(u_k,s,k)\right|\big(\hat{\Lambda}_n(\tau)+1\big) \\
&\quad \lesssim \big(\hat{\Lambda}_n(\tau)+1\big)(K_{nj}+1).
\end{aligned}$$

Inserting (S6.64)–(S6.66) back along with using (A.1) and $\sup_{(k,s)}\left|\hat{E}_j(U_k,s,k)\right| \lesssim K_{nj} + K_1$ again, (S6.63) turns into

$$\begin{aligned}
&\left|\mathrm{IF}_k^*(o|\hat{P}_{nj}) - \mathrm{IF}_k^*(o|\hat{E}_j,Q_u,\hat{G}_n)\right| = \left|\mathrm{IF}_k^*(o|\hat{E}_j,\mathbb{Q}_j,\hat{G}_n) - \mathrm{IF}_k^*(o|\hat{E}_j,Q_u,\hat{G}_n)\right| \\
&\quad \lesssim \left[\sqrt{\frac{\log n}{n}} + (K_{nj}+1)(\hat{\Lambda}_n(\tau)+1)\right]\left|(\mathbb{Q}_j - Q_u)[U_k]\right| + \left|(\mathbb{Q}_j - Q_u)[\hat{E}_j(U_k,k)]\right| \\
&\qquad + \left|(\mathbb{Q}_j - Q_u)[U_k\hat{E}_j(U_k,k)]\right| \\
&\quad \lesssim \big(\hat{\Lambda}_n(\tau)+1\big)K_{nj} \le \big(\left|\hat{\Lambda}_n(\tau) - \Lambda(\tau)\right| + 2\Lambda(\tau)+1\big)K_{nj}.
\end{aligned}$$

Observing that $\sup_{t\in\mathcal{T}}|\hat{\Lambda}_n(t) - \Lambda(t)| \le \sqrt{\log n}/\sqrt{n} < 1$ on the event $\mathcal{A}_n\cap\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_n$ by Lemma S6.7, it follows from the above display that

$$\max_{(j,k)}\left|\mathrm{IF}_k^*(o|\hat{P}_{nj}) - \mathrm{IF}_k^*(o|\hat{E}_j,Q_u,\hat{G}_n)\right| < 2(\Lambda(\tau)+1)\max_j\{K_{nj}\} \lesssim K_{nq_n},$$

yielding (S6.61). The proof of (S6.62) can be handled using similar arguments that are used to show (S6.58) and will appear later. Then following that $\mathrm{IF}_k^*(\cdot|\hat{E}_j,Q_u,\hat{G}_n)\in\tilde{\mathcal{F}}_n$, (S6.60) leads to

$$|\hat{\sigma}_{nj}^2 - \tilde{\sigma}_{nj}^2|1_{\mathcal{A}_n\cap\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_n} \lesssim K_{nj}.$$

To complete the proof, we now show that (S6.58), and it suffices to show that

$$\limsup_{n\to\infty}\sup_o\max_{(j,k)}|\mathrm{IF}_k^*(o|\hat{P}_{nj})|1_{\mathcal{A}_n\cap\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_n} < \infty.$$

In what follows, we assume without statement the occurrence of $\mathcal{A}_n\cap\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_n$. Note that, for any o,

$$\begin{aligned}
&\max_k\left|\mathrm{IF}_k^*(o|\hat{P}_{nj})\right| \\
&= \max_k\Bigg|\frac{(u_k - \mathbb{Q}_j[U_k])\big(y - \mathbb{Q}_j[\hat{E}_j(U_k,k)]\big)}{\mathrm{Var}_{\mathbb{Q}_j}(U_k)} - \frac{\mathrm{Cov}_{\mathbb{Q}_j}(U_k,\hat{E}_j(U_k,k))}{\mathrm{Var}^2_{\mathbb{Q}_j}(U_k)}(u_k - \mathbb{Q}_j[U_k])^2 \\
&\quad + \frac{(u_k - \mathbb{Q}_j[U_k])}{\mathrm{Var}_{\mathbb{Q}_j}(U_k)}\int_{\mathcal{T}}\hat{E}_j(u_k,s,k)\{1(x\in ds,\delta=0) - 1(x\ge s)d\hat{\Lambda}_n(s)\}\Bigg|
\end{aligned}$$

$$\lesssim \max_k \big|u_k - \mathbb{Q}_j[U_k]\big|\Big[\big|y-\tilde{y}\big| + \big|\tilde{y}\big| + \mathbb{Q}_j\big[\max_k \big|\hat{E}_j(U_k,k)\big|\big]\Big]$$

$$+ \max_k \big|\mathrm{Cov}_{\mathbb{Q}_j}(U_k, \hat{E}_j(U_k,k))\big|\big[\max_k \big|u_k - \mathbb{Q}_j[U_k]\big|\big]^2$$

$$+ \max_k \big|u_k - \mathbb{Q}_j[U_k]\big| \max_k \Big|\int_{\mathcal{T}} \hat{E}_j(u_k,s,k)\{1(x\in ds, \delta=0) - 1(x\geq s)d\hat{\Lambda}_n(s)\}\Big|,$$

where the last inequality holds by (S6.8.2) of Lemma S6.8 and the triangle inequality. When $\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n$ occurs, similar techniques to the arguments for (S6.64) yields that

$$|y-\tilde{y}| \leq |\tilde{y}|\sqrt{\log n}/(\sqrt{n} - \sqrt{\log n}) \lesssim \sqrt{\log n}/(\sqrt{n}-\sqrt{\log n}),$$

following that $|\tilde{y}| = |\delta x/G(x)|$ is bounded by some nonrandom finite constant that does not depend on (j,n) due to (A.2). Meanwhile, by (A.7) and (A.1) that the support of U_k is uniformly bounded, there exist positive finite constants K_0 and K_1 so that $\max_k |u_k - \mathbb{Q}_j[U_k]| \leq K_0$; $\max_k |E_0(U_k,k)| \leq K_1$ and $\sup_{(k,s)} |E_0(u_k,s,k)| \leq K_1$. Therefore when $\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n$ occurs, (S6.8.5)–(S6.8.6) of Lemma S6.8 imply that

$$\max_k |u_k - \mathbb{Q}_j[U_k]|\big[\big|y-\tilde{y}\big| + \big|\tilde{y}\big| + \mathbb{Q}_j\big[\max_k \big|\hat{E}_j(U_k,k)\big|\big]\big]$$

$$\lesssim \sqrt{\log n}/(\sqrt{n}-\sqrt{\log n}) + K_{nj} + 1;$$

$$\max_k \big|\mathrm{Cov}_{\mathbb{Q}_j}(U_k,\hat{E}_j(U_k,k))\big|\big[\max_k \big|u_k - \mathbb{Q}_j[U_k]\big|\big]^2$$

$$\leq \big[\max_k \big|u_k - \mathbb{Q}_j[U_k]\big|\big]^2 \mathbb{Q}_j\Big[\max_k \big|U_k - \mathbb{Q}_j[U_k]\big| \max_k \big|\hat{E}_j(U_k,k)\big|\Big] \lesssim K_{nj} + 1;$$

$$\max_k \big|u_k - \mathbb{Q}_j[U_k]\big| \max_k \Big|\int_{\mathcal{T}} \hat{E}_j(u_k,s,k)\{1(x\in ds,\delta=0) - 1(x\geq s)d\hat{\Lambda}_n(s)\}\Big|$$

$$\leq \max_k \big|u_k - \mathbb{Q}_j[U_k]\big| \sup_{(k,s)} \big|\hat{E}_j(u_k,s,k)\big|\big(1+\hat{\Lambda}_n(\tau)\big)$$

$$\lesssim \big(K_{nj}+K_1\big)\big(1+\hat{\Lambda}_n(\tau)\big) \leq \big(K_{nj}+K_1\big)\big(\big|\hat{\Lambda}_n(\tau) - \Lambda(\tau)\big| + \Lambda(\tau) + 1\big)$$

$$< \big(K_{nj}+K_1\big)\big(2+\Lambda(\tau)\big) \lesssim K_{nj} + 1,$$

where the last line follows that $\sup_{t\in\mathcal{T}} |\hat{\Lambda}_n(t) - \Lambda(t)| \leq \sqrt{\log n}/\sqrt{n} < 1$ on the event $\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n$, using Lemma S6.7. Therefore, the above results ensure that

$$\sup_o \max_{(j,k)} \big|\,\mathrm{IF}_k^*(o|\hat{P}_{nj})\big| 1_{\mathcal{A}_n\cap\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_n} \lesssim \sqrt{\log n}/\big(\sqrt{n}-\sqrt{\log n}\big) + \max_j\{K_{nj}\} + 1,$$

so that $\limsup_{n\to\infty} \sup_o \max_{(j,k)} |\,\mathrm{IF}_k^*(o|\hat{P}_{nj})| 1_{\mathcal{A}_n\cap\mathcal{B}_n(\tilde{K})\cap\mathcal{C}_n}$ is bounded by some (j,n)-independent constant, following that $\max_j\{K_{nj}\} \leq K_{nq_n} \to 0$ by $q_n^{1/4}/\log(n\vee p_n) \to \infty$. □

LEMMA S6.11. *Suppose the conditions of Theorem 4.1 hold. Then there exists an event $\mathcal{E}_n$ that corresponds to the intersection of*

(S6.11.1) $\min_{j\in\{q_n,\ldots,n\}} \hat{\sigma}_{nj}^2$ *is bounded away from zero by a constant;*
(S6.11.2) $|\sigma_{nj}/\hat{\sigma}_{nj} - 1| \lesssim K_{nj}$ *for all* $j = q_n,\ldots,n$, *where* $\sigma_{nj}^2 = \int \mathrm{IF}_{k_j}^*(o|P)^2 dP(o)$,

where each of (S6.11.1) and (S6.11.2) relies on appropriately specified (non-random) constants that do not depend on n, such that

$$\mathrm{P}(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \mathcal{D}_n(K') \cap \mathcal{E}_n) \to 1,$$

with $\tilde{K}, K' \in (0,\infty)$ given in Lemma S6.7 and Lemma S6.9, respectively.

PROOF. When $\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \mathcal{D}_n(K')$ occurs, to show that $\hat{\sigma}^2_{nj}$ is uniformly bounded away from zero, it suffices to show that $|\hat{\sigma}^2_{nj} - \sigma^2_{nj}| \lesssim K_{nj}$ for $j = q_n, \dots, n$ on this event. This gives

$$\hat{\sigma}^2_{nj} \gtrsim \sigma^2_{nj} - K_{nj} \geq \sigma^2_{nj} - \sqrt{\log(n \vee p_n)/q_n} > \zeta'$$

for some universal constant $\zeta' > 0$, where the final inequality is a direct consequence of the conditions for Theorem 4.1: $q_n^{1/4}/\log(n \vee p_n) \to \infty$ and σ^2_{nj} is uniformly bounded away from zero by (A.6).

We now show that $|\hat{\sigma}^2_{nj} - \sigma^2_{nj}| \lesssim K_{nj}$ when $\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \mathcal{D}_n(K')$ occurs. Recall that $\tilde{\sigma}^2_{nj} = \mathrm{Var}(\mathrm{IF}^*_{k_j}(O|\hat{P}_{nj})|O_1, \dots, O_j)$. We first use the triangle inequality to give $|\hat{\sigma}^2_{nj} - \sigma^2_{nj}| \leq |\hat{\sigma}^2_{nj} - \tilde{\sigma}^2_{nj}| + |\tilde{\sigma}^2_{nj} - \sigma^2_{nj}|$. Because $|\hat{\sigma}^2_{nj} - \tilde{\sigma}^2_{nj}| 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \mathcal{D}_n(K')} \lesssim K_{nj}$ is given by Lemma S6.10, it suffices to show that

$$|\tilde{\sigma}^2_{nj} - \sigma^2_{nj}| 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \mathcal{D}_n(K')} \lesssim K_{nj}.$$

Recalling that $\hat{P}'_j = (\hat{E}_j, \mathbb{Q}_j, G)$ and noting that $P\,\mathrm{IF}^*_{k_j}(O|\hat{P}'_j) = 0$ and $P\,\mathrm{IF}^*_{k_j}(O|P) = 0$, Jensen's inequality and the triangle inequality give that

$$\begin{aligned}
&|\tilde{\sigma}^2_{nj} - \sigma^2_{nj}| 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \mathcal{D}_n(K')} \\
&\quad \leq \Big\{ P \big| \mathrm{IF}^*_{k_j}(O|\hat{P}_{nj}) + \mathrm{IF}^*_{k_j}(O|P) \big| \big| \mathrm{IF}^*_{k_j}(O|\hat{P}_{nj}) - \mathrm{IF}^*_{k_j}(O|P) \big| \\
&\qquad + P \big| \mathrm{IF}^*_{k_j}(O|\hat{P}_{nj}) + \mathrm{IF}^*_{k_j}(O|\hat{P}'_j) \big| P \big| \mathrm{IF}^*_{k_j}(O|\hat{P}_{nj}) - \mathrm{IF}^*_{k_j}(O|\hat{P}'_j) \big| \Big\} 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \mathcal{D}_n(K')} \\
&\quad \leq \Big\{ P \big| \mathrm{IF}^*_{k_j}(O|\hat{P}_{nj}) + \mathrm{IF}^*_{k_j}(O|P) \big| \big| \mathrm{IF}^*_{k_j}(O|\hat{P}_{nj}) - \mathrm{IF}^*_{k_j}(O|\hat{P}'_j) \big| \\
&\qquad + P \big| \mathrm{IF}^*_{k_j}(O|\hat{P}_{nj}) + \mathrm{IF}^*_{k_j}(O|P) \big| \big| \mathrm{IF}^*_{k_j}(O|\hat{P}'_j) - \mathrm{IF}^*_{k_j}(O|P) \big| \\
&\qquad + P \big| \mathrm{IF}^*_{k_j}(O|\hat{P}_{nj}) + \mathrm{IF}^*_{k_j}(O|\hat{P}'_j) \big| P \big| \mathrm{IF}^*_{k_j}(O|\hat{P}_{nj}) - \mathrm{IF}^*_{k_j}(O|\hat{P}'_j) \big| \Big\} 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \mathcal{D}_n(K')} \\
&\quad \lesssim P \Big[\max_{k \in \mathcal{K}_n} | \mathrm{IF}^*_k(O|\hat{P}_{nj}) - \mathrm{IF}^*_k(O|\hat{P}'_j) | 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \mathcal{D}_n(K')} \Big] \\
&\qquad + P \Big[\max_{k \in \mathcal{K}_n} | \mathrm{IF}^*_k(O|\hat{P}'_j) - \mathrm{IF}^*_k(O|P) | 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \mathcal{D}_n(K')} \Big],
\end{aligned}$$

where the last inequality holds because $\max_k |\mathrm{IF}^*_k(\cdot|\tilde{P})| 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n}$ is uniformly bounded by some (j,n)-independent constant for $\tilde{P} \in \{\hat{P}_{nj}, \hat{P}'_j, P\}$, using the arguments for (S6.58) in Lemma S6.10 together with (A.1)–(A.5) and (A.6)–(A.7). From the above display, Lemma S6.9 further implies

$$\begin{aligned}
|\tilde{\sigma}^2_{nj} - \sigma^2_{nj}| 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \mathcal{D}_n(K')} &\lesssim P \Big[\max_{k \in \mathcal{K}_n} | \mathrm{IF}^*_k(O|\hat{P}_{nj}) - \mathrm{IF}^*_k(O|\hat{P}'_j) | 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n} \Big] \\
&\quad + P \Big[\max_{k \in \mathcal{K}_n} | \mathrm{IF}^*_k(O|\hat{P}'_j) - \mathrm{IF}^*_k(O|P) | 1_{\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n} \Big] \lesssim K_{nj}.
\end{aligned}$$

This completes the proof of the fact that when $\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \mathcal{D}_n(K')$ occurs, $\hat{\sigma}^2_{nj}$ is uniformly bounded away from zero by a non-random positive lower bound, for $j = q_n, \dots, n$.

When $\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \mathcal{D}_n(K')$ occurs, the statement in (S6.11.2) is an immediate consequence of the already-established (S6.11.1):

$$\left| \frac{\sigma_{nj}}{\hat{\sigma}_{nj}} - 1 \right| \leq \left| \left[\frac{\sigma_{nj}}{\hat{\sigma}_{nj}} - 1 \right] \left[\frac{\sigma_{nj}}{\hat{\sigma}_{nj}} + 1 \right] \right| = \left| \frac{\sigma^2_{nj}}{\hat{\sigma}^2_{nj}} - 1 \right| = \frac{|\hat{\sigma}^2_{nj} - \sigma^2_{nj}|}{\hat{\sigma}^2_{nj}} \lesssim K_{nj}.$$

Hence, we have shown that $\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \mathcal{D}_n(K')$ implies $\mathcal{E}_n$, where $\mathcal{E}_n$ is the event that (S6.11.1) and (S6.11.2) hold with the constants that were shown to exist earlier in this proof. As $\mathrm{P}(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \mathcal{D}_n(K')) \to 1$ (Lemma S6.9), this implies that

$$\mathrm{P}(\mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \mathcal{D}_n(K') \cap \mathcal{E}_n) \to 1.$$

□

We also need the lemma below that concerns the probability of the event $\mathcal{H}_n = \cap_{j=q_n}^{n-1} \mathcal{H}_{nj}$, where

$$\mathcal{H}_{nj} = \left\{ \sup_{t\in\mathcal{T}} \left| \frac{\hat{G}_j(t)}{G(t)} - 1 \right| \leq \frac{1}{S(\tau)} \sqrt{\frac{\log n}{j}} \right\},\ j = q_n, \ldots, n-1$$

with $S(\tau) = \mathrm{P}(T \geq \tau)$. Note that $S(\tau) > 0$, following that $G(\tau) > 0$ by (A.2), that $\mathrm{P}(X \geq \tau) > 0$ by (A.5), and the independent censoring assumption that implies $\mathrm{P}(X \geq \tau) = S(\tau)G(\tau)$.

LEMMA S6.12. *Under the conditions of Theorem 4.1,* $\mathrm{P}(\mathcal{H}_n) \to 1$.

PROOF. Fix $\xi \in (0,1)$. We will use the exponential bound for the Kaplan–Meier estimator that is presented in Theorem 1 of Wellner (2007). For $\lambda > 0$ and some constant $K > 0$, this inequality takes the form

$$\mathrm{P}\Big(\sqrt{j}\big\|S(\hat{G}_j - G)\big\|_\infty > \lambda\Big) \leq 2.5\exp(-2\lambda^2 + K\lambda).$$

Noting that

$$\mathrm{P}\Big(\sqrt{j}\big\|S(\hat{G}_j - G)\big\|_\infty > \lambda\Big) \geq \mathrm{P}\Big(\sqrt{j}\sup_{t\in\mathcal{T}}\big|S(t)(\hat{G}_j(t) - G(t))\big| > \lambda\Big)$$
$$\geq \mathrm{P}\Big(\sqrt{j}S(\tau)\sup_{t\in\mathcal{T}}\big|\hat{G}_j(t) - G(t)\big| > \lambda\Big)$$

and taking $\lambda = \sqrt{\log n}$, we see that

$$\log \mathrm{P}\left(\sup_{t\in\mathcal{T}}\big|\hat{G}_j(t) - G(t)\big| > \frac{1}{S(\tau)}\sqrt{\frac{\log n}{j}}\right) \leq \log(2.5) - 2\log n + K\sqrt{\log n}\,.$$

For all n large enough, $K \leq \xi\sqrt{\log n}$. Combining this with the fact that $S(\tau) > 0$ shows that

$$\log \mathrm{P}\left(\sup_{t\in\mathcal{T}}\big|\hat{G}_j(t) - G(t)\big| > \frac{1}{S(\tau)}\sqrt{\frac{\log n}{j}}\right) \leq \log(2.5) - (2-\xi)\log(n).$$

Hence,

$$\mathrm{P}\left(\bigcup_{j=q_n}^{n-1}\left\{\sup_{t\in\mathcal{T}}\left|\frac{\hat{G}_j(t)}{G(t)} - 1\right| > \frac{1}{S(\tau)}\sqrt{\frac{\log n}{j}}\right\}\right)$$
$$\leq \sum_{j=q_n}^{n-1}\mathrm{P}\left(\sup_{t\in\mathcal{T}}\left|\frac{\hat{G}_j(t)}{G(t)} - 1\right| > \frac{1}{S(\tau)}\sqrt{\frac{\log n}{j}}\right) \leq 2.5\sum_{j=q_n}^{n-1} n^{\xi-2} \leq 2.5n^{\xi-1} \to 0.$$

□

Let $\mathcal{L}_n = \mathcal{A}_n \cap \mathcal{B}_n(\tilde{K}) \cap \mathcal{C}_n \cap \mathcal{D}_n(K') \cap \mathcal{E}_n \cap \mathcal{H}_n$ and note that, from the above lemmas,

$$\mathrm{P}(\mathcal{L}_n) \to 1 \tag{S6.67}$$

when the conditions of Theorem 4.1 hold. The upcoming lemmas give the asymptotic negligibility of (I), (II), (III) and (V) in (16). By (S6.67), it suffices to show the asymptotically negligibility after multiplication by $1_{\mathcal{L}_n}$.

To show the lemmas of asymptotic negligibility, we need additional properties that are given below.

LEMMA S6.13. *Let $\mathcal{X}$ be the sample space, $f_n : \mathcal{X} \to \mathbb{R}$ be a random function that depends on the n observations with $\sup_{o\in\mathcal{X}} |f_n(o)1_{\mathcal{L}_n}| \lesssim \sqrt{\log(n \vee p_n)}/q_n^{1/4}$ with probability tending to one, $O_{j+1,k_j} \equiv (X_{j+1}, \delta_{j+1}, U_{j+1,k_j})$ and $\sigma_{nj}^2 \equiv \mathrm{Var}(\mathrm{IF}^*_{k_j}(O|P))$, for $j = q_n, \ldots, n-1$. Under the conditions of Theorem 4.1,*

$$\left|\frac{1}{\sqrt{n-q_n}}\sum_{j=q_n}^{n-1}\frac{m_j(U_{j+1,k_j} - \mathbb{Q}_j[U_{k_j}])}{\hat{\sigma}_{nj}\mathrm{Var}_{\mathbb{Q}_j}(U_{k_j})} f_n(O_{j+1,k_j})1_{\mathcal{L}_n}\right|$$
$$\lesssim \left|\frac{1}{\sqrt{n-q_n}}\sum_{j=q_n}^{n-1}\frac{m_j(U_{j+1,k_j} - Q_u[U_{k_j}])}{\sigma_{nj}\mathrm{Var}_{Q_u}(U_{k_j})} f_n(O_{j+1}, k_j)1_{\mathcal{L}_n}\right| + o_p(1).$$

PROOF. To show the result, we first observe that

$$\begin{aligned}
&\left|\frac{1}{\sqrt{n-q_n}}\sum_{j=q_n}^{n-1}\frac{m_j(U_{j+1,k_j} - \mathbb{Q}_j[U_{k_j}])}{\hat{\sigma}_{nj}\mathrm{Var}_{\mathbb{Q}_j}(U_{k_j})} f_n(O_{j+1,k_j})1_{\mathcal{L}_n}\right| \\
&\quad\le \left|\frac{1}{\sqrt{n-q_n}}\sum_{j=q_n}^{n-1}\frac{m_j(U_{j+1,k_j} - \mathbb{Q}_j[U_{k_j}])}{\sigma_{nj}\mathrm{Var}_{\mathbb{Q}_j}(U_{k_j})} f_n(O_{j+1,k_j})1_{\mathcal{L}_n}\right| \\
&\qquad+ \left|\frac{1}{\sqrt{n-q_n}}\sum_{j=q_n}^{n-1}\frac{m_j(U_{j+1,k_j} - \mathbb{Q}_j[U_{k_j}])}{\sigma_{nj}\mathrm{Var}_{\mathbb{Q}_j}(U_{k_j})}\left[\frac{\sigma_{nj}}{\hat{\sigma}_{nj}} - 1\right] f_n(O_{j+1,k_j})1_{\mathcal{L}_n}\right|.
\end{aligned}\tag{S6.68}$$

The triangle inequality upper-bounds the first term of the right-hand-side in (S6.68) by

$$\begin{aligned}
&\left|\frac{1}{\sqrt{n-q_n}}\sum_{j=q_n}^{n-1}\frac{m_j(U_{j+1,k_j} - Q_u[U_{k_j}])}{\sigma_{nj}\mathrm{Var}_{Q_u}(U_{k_j})} f_n(O_{j+1,k_j})1_{\mathcal{L}_n}\right| \\
&\quad+ \left|\frac{1}{\sqrt{n-q_n}}\sum_{j=q_n}^{n-1}\frac{m_j(\mathbb{Q}_j - Q_u)[U_{k_j}]}{\sigma_{nj}\mathrm{Var}_{Q_u}(U_{k_j})} f_n(O_{j+1,k_j})1_{\mathcal{L}_n}\right| \\
&\quad+ \left|\frac{1}{\sqrt{n-q_n}}\sum_{j=q_n}^{n-1}\frac{m_j}{\sigma_{nj}}(U_{j+1,k_j} - \mathbb{Q}_j[U_{k_j}])\left[\frac{1}{\mathrm{Var}_{\mathbb{Q}_j}(U_{k_j})} - \frac{1}{\mathrm{Var}_{Q_u}(U_{k_j})}\right] f_n(O_{j+1,k_j})1_{\mathcal{L}_n}\right| \\
&\lesssim \left|\frac{1}{\sqrt{n-q_n}}\sum_{j=q_n}^{n-1}\frac{m_j(U_{j+1,k_j} - Q_u[U_{k_j}])}{\sigma_{nj}\mathrm{Var}_{Q_u}(U_{k_j})} f_n(O_{j+1}, k_j)1_{\mathcal{L}_n}\right| + \sqrt{n-q_n}\frac{\log(n\vee p_n)}{q_n^{3/4}} + o_p(1),
\end{aligned}\tag{S6.69}$$

where the last inequality holds because U_k is uniformly bounded on k by (A.1); that σ_{nj} and $\mathrm{Var}_{Q_u}(U_k)$ are bounded away from zero uniformly over (j,k) by (A.6), and by the occurrence of $\mathcal{A}_n \cap \mathcal{C}_n \supset \mathcal{L}_n$, along with $\sup_o |f_n(o)1_{\mathcal{L}_n}| \lesssim \sqrt{\log(n\vee p_n)}/q_n^{1/4}$ with probability tending to one. Note that $\sqrt{n-q_n}\log(n\vee p_n)/q_n^{3/4} \to 0$ by the conditions of Theorem 4.1.

The second term of the right-hand-side in (S6.68) is bounded by

$$\left|\frac{1}{\sqrt{n-q_n}}\sum_{j=q_n}^{n-1}\frac{m_j(U_{j+1,k_j}-\mathbb{Q}_j[U_{k_j}])}{\sigma_{nj}\mathrm{Var}_{\mathbb{Q}_j}(U_{k_j})}\Big[\frac{\sigma_{nj}}{\hat\sigma_{nj}}-1\Big]f_n(O_{j+1,k_j})1_{\mathcal{L}_n}\right|$$

$$\lesssim \sup_o |f_n(o)1_{\mathcal{L}_n}|\frac{1}{\sqrt{n-q_n}}\sum_{j=q_n}^{n-1}\Big|\frac{\sigma_{nj}}{\hat\sigma_{nj}}-1\Big| \le \sup_o |f_n(o)1_{\mathcal{L}_n}|\sqrt{n-q_n}\frac{\sqrt{\log(n\vee p_n)}}{\sqrt{q_n}}$$

$$\lesssim \sqrt{n-q_n}\frac{\log(n\vee p_n)}{q_n^{3/4}}+o_p(1)\to 0.$$

In above display, the first inequality holds because $|m_j(U_{j+1,k_j}-\mathbb{Q}_j[U_{k_j}])/\mathrm{Var}_{\mathbb{Q}_j}(U_{k_j})|$ is uniformly bounded by some constant that does not depend on (j,n) almost surely for $j=q_n,\dots,n-1$, which is a consequence of (A.1) and (S6.8.2) of Lemma S6.8. The penultimate inequality follows (S6.11.2) of Lemmas S6.11. $\square$

For $j=q_n,\dots,n-1$, we define random functions $\tilde e_{nj}, e_{nj}:\mathbb{R}\times\mathcal{T}\times\mathcal{K}_n\to\mathbb{R}$ by

$$\tilde e_{nj}(u,s,k)=E_0(u,s,k)1(X_{j+1}\ge s)1\big(\inf_s Y_n(s)\ge\sqrt{n}\big),\ \text{and} \tag{S6.70}$$

$$\begin{aligned} e_{nj}(u,s,k)=[E_0(u,s,k)-\hat E_j(u,s,k)]1(X_{j+1}\ge s)\\ \times 1\Big(\sup_{(k,s,u)}|E_0(u,s,k)-\hat E_j(u,s,k)|\lesssim K_{nj},\ \inf_s Y_n(s)\ge\sqrt{n}\Big).\end{aligned} \tag{S6.71}$$

Recall that $N_n(s)$ and $Y_n(s)$ are the aggregated counting process for the censored outcomes and the size of the risk set at time s, as defined in Section S1 (except for removing u), and also note that $d\bar M(s)\equiv dN_n(s)-Y_n(s)d\Lambda(s)$ is a local martingale with respect to the simpler filtration

$$\mathcal{F}'_s=\sigma(\{N_n(s'),\,Y_n(s'):\,s'\le s\in\mathcal{T}\}). \tag{S6.72}$$

Observing the decomposition of $\hat\Lambda_n-\Lambda$ analogously to (S5.21) without u gives that

$$\hat\Lambda_n(t)-\Lambda(t)=\int_{-\infty}^t\frac{1(Y_n(s)>0)}{Y_n(s)}d\bar M(s)-\int_{-\infty}^t 1(Y_n(s)=0)\,d\Lambda(s). \tag{S6.73}$$

In what follows, we will need an exponential inequality for martingales with bounded jumps:

LEMMA S6.14. *Let $W_n(t), t\in\mathcal{T}$ be a martingale with jumps bounded by a constant $K_n>0$, and the quadratic variation $\langle W_n\rangle(t)\le b_n^2$ for a constant $b_n>0$ with respect to the filtration $\sigma(\{N_n(s),Y_n(s):s\le t\},\{\boldsymbol{U}_i\}_{i=1}^n)$, where both K_n and b_n go to zero for sufficiently large n. In particular,*

$$W_n(t)\equiv\int_{-\infty}^t w(s)\frac{1(Y_n(s)\ge\sqrt{n})}{Y_n(s)}d\bar M(s),$$

where the function w is uniformly bounded and left-continuous in s, and adapted to the given filtration. Let ϵ_n be any sequence with values in $(0,1)$ and $\epsilon_n\to 0$; then

$$\mathrm{P}\big(|W_n(\tau)|\ge\epsilon_n\big)\le 2\exp\left(-\frac{\epsilon_n^2}{2(\epsilon_nK_n+b_n^2)}\right).$$

PROOF. Let $\Delta W_n(t)$ be the jump of W_n at time t:

$$\Delta W_n(t) = W_n(t) - W_n(t-) = w(t)\frac{1(Y_n(t) \geq \sqrt{n})}{Y_n(t)}\Delta \bar{M}(t),$$

where $\Delta \bar{M}(t) = \Delta N_n(t) - Y_n(t)\Delta\Lambda(t) = \Delta N_n(t)$, since (A.2) implies that Λ is continuous. Note also that $|\Delta N_n(t)| \leq 1$ because no two individual counting processes that are aggregated in $\{N_n(t) : t \in \mathcal{T}\}$ jump at the same time. Therefore, $|\Delta \bar{M}(t)| = |\Delta N_n(t)| \leq 1$; along with w being uniformly bounded: $|w(t)| \leq K^*$ for all t and a constant K^* that could depend on n,

$$|\Delta W_n(t)| \leq \left| w(t)\frac{1(Y_n(t) \geq \sqrt{n})}{Y_n(t)} \right| \leq \frac{K^*}{\sqrt{n}} \equiv K_n.$$

Meanwhile with respect to the given filtration, the predictable quadratic variation of $W_n(t)$ is

$$\langle W_n \rangle(t) = \int_{-\infty}^{t} w^2(s)\frac{1(Y_n(s) \geq \sqrt{n})}{Y_n(s)} d\Lambda(s) \leq \frac{(K^*)^2\Lambda(t)}{\sqrt{n}} \equiv b_n^2(t).$$

Obviously, $b_n^2(t) \leq b_n^2(\tau) \equiv b_n^2$. By an exponential inequality for martingales with bounded jumps (cf., Lemma 2.1 of van de Geer, 1995),

$$\mathrm{P}\big(|W_n(t)| \geq \epsilon_n \text{ and } \langle W_n \rangle(t) \leq b_n^2 \text{ for some } t \in \mathcal{T} \big) \leq 2\exp\left(-\frac{\epsilon_n^2}{2\big(\epsilon_n K_n + b_n^2\big)} \right).$$

Along with $\langle W_n \rangle(\tau) \leq b_n^2$, the above display gives the same exponential bound on $\mathrm{P}(\, |W_n(\tau)| \geq \epsilon_n \,)$. $\square$

LEMMA S6.15. *For $\mathcal{I}_n = \{(j,k,u) : j \in \{q_n, \ldots, n-1\}, k \in \mathcal{K}_n, u \in [-1,1]\}$, $\tilde{e}_{nj}$ and e_{nj} are as defined in (S6.70) and (S6.71). Under the conditions of Theorem 4.1, with probability tending to one,*

(S6.15.1) $\sup_{(j,k,u)\in\mathcal{I}_n} \left| \int_{\mathcal{T}} \tilde{e}_{nj}(u,s,k)\{d\hat{\Lambda}_n(s) - d\Lambda(s)\} \right| \leq \sqrt{\log(n \vee p_n)}/q_n^{1/4}$;

(S6.15.2) $\sup_{(j,k,u)\in\mathcal{I}_n} \left| \int_{\mathcal{T}} e_{nj}(u,s,k)\{d\hat{\Lambda}_n(s) - d\Lambda(s)\} \right| \lesssim \sqrt{\log(n \vee p_n)}/q_n^{1/4}$, *relying on an appropriately specified (non-random) constant that does not depend on n.*

PROOF. We prove (S6.15.1) and (S6.15.2) sequentially. By the decomposition of $\hat{\Lambda}_n - \Lambda$ in (S6.73), we have that for $(j,k,u) \in \mathcal{I}_n$,

$$\int_{\mathcal{T}} \tilde{e}_{nj}(u,s,k)\{d\hat{\Lambda}_n(s) - d\Lambda(s)\} = \int_{\mathcal{T}} \tilde{e}_{nj}(u,s,k)\frac{1(Y_n(s) > 0)}{Y_n(s)} d\bar{M}(s),$$

where $d\bar{M}(s) = dN_n(s) - Y_n(s)d\Lambda(s)$. Note that by Lemma S6.2 and (S6.70), there exist functions $a_k, b_k \in \mathcal{BV}(\mathcal{T}, \tilde{M}_0, \tilde{M}_1)$ such that $E_0(u,s,k) = a_k(s) + b_k(s)u$ and

$$\tilde{e}_{nj}(u,s,k) = [a_k(s) + b_k(s)u]1(X_{j+1} \geq s)1(Y_n(s) \geq \sqrt{n}), \tag{S6.74}$$

where a_k and b_k are uniformly bounded by $\tilde{M}_0$ on (k,s) and left-continuous in s, inheriting from the properties of E_0 that are assumed in (A.7). Moreover, we see that $a_k(s)$, $b_k(s)$ and $Y_n(s)$ are predictable with respect to $\mathcal{F}'_s$ defined in (S6.72), because they are left-continuous in s and adapted to $\mathcal{F}'_s$. Therefore, (A.1) enables us to suppose that U takes values in $[-1,1]$ without loss of generality, leading to

(S6.75)

$$\sup_{(j,k,u)\in\mathcal{I}_n} \left| \int_{\mathcal{T}} \tilde{e}_{nj}(u,s,k)\frac{1(Y_n(s) > 0)}{Y_n(s)} d\bar{M}(s) \right|$$

$$
\begin{aligned}
&= \sup_{(j,k,u)\in\mathcal{I}_n} \Big| \int_{\mathcal{T}} [a_k(s)+b_k(s)u] 1(X_{j+1}\geq s) 1(Y_n(s)\geq\sqrt{n}) \frac{1(Y_n(s)>0)}{Y_n(s)} d\bar{M}(s) \Big| \\
&\leq \max_{(j,k)} \Big| \int_{\mathcal{T}} a_k(s) \frac{1(X_{j+1}\geq s, Y_n(s)\geq\sqrt{n})}{Y_n(s)} d\bar{M}(s) \Big| \\
&\quad + \max_{(j,k)} \Big| \int_{\mathcal{T}} b_k(s) \frac{1(X_{j+1}\geq s, Y_n(s)\geq\sqrt{n})}{Y_n(s)} d\bar{M}(s) \Big|.
\end{aligned}
$$

So to show the desired result, it suffices to show that each term on the right-hand-side of (S6.75) is bounded above by $\sqrt{\log(n\vee p_n)}/q_n^{1/4}$ with probability tending to one. Here we only tackle the first term on the right-hand-side of (S6.75), and the second term can be handled using nearly identical arguments.

To use Lemma S6.14 for showing the desired result, we define the required notations as follows, especially here we have the martingale W_n further be indexed by (j,k) (with n omitted) and the function $w_k(\cdot)\equiv a_k(\cdot)$, where w now is indexed by k. For $(j,k)\in\{q_n,\dots,n-1\}\times\mathcal{K}_n$, let

$$
W_{jk}(t) \equiv \int_{-\infty}^{t} w_k(s) \frac{1(X_{j+1}\geq s, Y_n(s)\geq\sqrt{n})}{Y_n(s)} d\bar{M}(s),
$$

and $\Delta W_{jk}(t)$ be the jump of W_{jk} at time t:

$$
\Delta W_{jk}(t) = W_{jk}(t) - W_{jk}(t-) = w_k(t) \frac{1(X_{j+1}\geq s, Y_n(t)\geq\sqrt{n})}{Y_n(t)} \Delta\bar{M}(t);
$$

together with $|w_k(t)|\leq\tilde{M}_0$ for all t,

$$
|\Delta W_{jk}(t)| \leq \Big| w_k(t) \frac{1(X_{j+1}\geq s, Y_n(t)\geq\sqrt{n})}{Y_n(t)} \Big| \leq \frac{\tilde{M}_0}{\sqrt{n}} \equiv K_n.
$$

Meanwhile, the predictable quadratic variation of $W_{jk}(t)$ is

$$
\langle W_{jk}\rangle(t) = \Big| \int_{-\infty}^{t} w_k^2(s) \frac{1(X_{j+1}\geq s, Y_n(t)\geq\sqrt{n})}{Y_n(s)} d\Lambda(s) \Big| \leq \frac{\tilde{M}_0^2\Lambda(t)}{\sqrt{n}} \equiv b_n^2(t),
$$

and $b_n^2(t)\leq b_n^2(\tau)\equiv b_n^2$. Let $\epsilon_n=\sqrt{\log(n\vee p_n)}/q_n^{1/4}$; then Lemma S6.14 implies that

$$
\begin{aligned}
\mathrm{P}\big(|W_{jk}(\tau)|\geq\epsilon_n\big) &\leq 2\exp\left(-\frac{\epsilon_n^2}{2\big(\epsilon_n K_n + b_n^2\big)}\right) \\
&= 2\exp\left(-\frac{\sqrt{n}\log(n\vee p_n)}{2q_n^{1/2}} \frac{1}{\big(q_n^{-1/4}\sqrt{\log(n\vee p_n)}\tilde{M}_0 + \Lambda(\tau)\tilde{M}_0^2\big)}\right)
\end{aligned}
$$

for $(j,k)\in\{q_n,\dots,n-1\}\times\mathcal{K}_n$, leading to

$$
\begin{aligned}
\mathrm{P}\Big(\max_{(j,k)}|W_{jk}(\tau)|\geq\epsilon_n\Big) &\leq \sum_{j=q_n}^{n-1}\sum_{k=1}^{p_n} \mathrm{P}\Big(|W_{jk}(\tau)|\geq\epsilon_n\Big) \\
&\leq 2p_n(n-q_n)\exp\left(-\frac{\sqrt{n}\log(n\vee p_n)}{2q_n^{1/2}} \frac{1}{\big(q_n^{-1/4}\sqrt{\log(n\vee p_n)}\tilde{M}_0 + \Lambda(\tau)\tilde{M}_0^2\big)}\right) \\
&\leq 2\exp\big(2\log(n\vee p_n)\big)\exp\big(-n^{1/4}\sqrt{\log(n\vee p_n)}\big) \\
&= 2\exp\Big(\Big[2\sqrt{\log(n\vee p_n)}/n^{1/4}-1\Big] n^{1/4}\sqrt{\log(n\vee p_n)}\Big).
\end{aligned}
$$

Therefore $\limsup_{n\to\infty} \mathrm{P}\big(\max_{(j,k)} |W_{jk}(\tau)| \geq a_n\big) = 0$, following $n/q_n = O(1)$, $\log(n \vee p_n)/q_n^{1/4} \to 0$ and $n^{1/4}\sqrt{\log(n \vee p_n)} \to \infty$. Hence, we complete the proof of (S6.15.1).

Below we present the proof of (S6.15.2). Since the total variation of $\hat{\Lambda}_n - \Lambda$ is bounded by $\hat{\Lambda}_n(\tau) + \Lambda(\tau)$ we have

$$\text{(S6.76)} \qquad \begin{aligned} &\sup_{(j,k,u)\in\mathcal{I}_n} \left| \int_{\mathcal{T}} e_{nj}(u,s,k)\{d\hat{\Lambda}_n(s) - d\Lambda(s)\} \right| \\ &\leq \sup_{(j,k,s,u)} |e_{nj}(u,s,k)| \big|\hat{\Lambda}_n(\tau) + \Lambda(\tau)\big| \\ &\leq \sup_{(j,k,s,u)} |e_{nj}(u,s,k)| \big|\hat{\Lambda}_n(\tau) - \Lambda(\tau)\big| + 2 \sup_{(j,k,s,u)} |e_{nj}(u,s,k)|\Lambda(\tau). \end{aligned}$$

Observing that $\{\sqrt{n}[\hat{\Lambda}_n(t) - \Lambda(t)] : t \in \mathcal{T}\}$ converges to a tight Gaussian process, and then the continuous mapping theorem gives that $\sqrt{n}\sup_{t\in\mathcal{T}} |\hat{\Lambda}_n(t) - \Lambda(t)|$ converges to the supremum of the absolute value of this Gaussian process. Therefore for any sequence $\varepsilon_n \to \infty$, $\mathrm{P}\big(\sup_{t\in\mathcal{T}} \sqrt{n}|\hat{\Lambda}_n(t) - \Lambda(t)| > \varepsilon_n\big) \to 0$, in particular, $\varepsilon_n = \sqrt{\log n}$. Consequently, $\sup_{t\in\mathcal{T}} |\hat{\Lambda}_n(t) - \Lambda(t)| \leq \sqrt{\log n}/\sqrt{n}$ with probability tending to one. Following from (S6.71) that $\sup_{(j,k,s,u)} |e_{nj}(u,s,k)| \leq K_* \max_j K_{nj}$ for some positive constant K_* that does not depend on n, we have that with probability tending to one,

$$\sup_{(j,k,s,u)} |e_{nj}(u,s,k)| \big|\hat{\Lambda}_n(\tau) - \Lambda(\tau)\big| \leq K_* \max_j\{K_{nj}\} \frac{\sqrt{\log n}}{\sqrt{n}} < K_* \max_j\{K_{nj}\}.$$

Hence from (S6.76),

$$\begin{aligned} &\sup_{(j,k,u)\in\mathcal{I}_n} \left| \int_{\mathcal{T}} e_{nj}(u,s,k)\{d\hat{\Lambda}_n(s) - d\Lambda(s)\} \right| < K_*(1 + 2\Lambda(\tau)) \max_j\{K_{nj}\} \\ &\leq K_*(1 + 2\Lambda(\tau)) \frac{\sqrt{\log(n \vee p_n)}}{q_n^{1/4}} \end{aligned}$$

with probability tending to one, which gives (S6.15.2). □

In upcoming lemmas, we show that $|\sum_{j=q_n}^{n-1} D_{n,j+1}|$ converges to zero in probability as n goes to infinity, where for $j = q_n + 1, \ldots, n$,

$$\begin{aligned} D_{nj} &\equiv \frac{1}{\sqrt{n - q_n}} \frac{m_{j-1}(U_{j,k_{j-1}} - Q_u[U_{k_{j-1}}])}{\sigma_{nj-1}\mathrm{Var}_{Q_u}(U_{k_{j-1}})} \int_{\mathcal{T}} \tilde{e}_{n,j-1}(U_{j,k_{j-1}}, s, k_{j-1})\{d\hat{\Lambda}_n(s) - d\Lambda(s)\} \\ &= \frac{1}{\sqrt{n - q_n}} \frac{m_{j-1}(U_{j,k_{j-1}} - Q_u[U_{k_{j-1}}])}{\sigma_{nj-1}\mathrm{Var}_{Q_u}(U_{k_{j-1}})} \int_{\mathcal{T}} \tilde{e}_{n,j-1}(U_{j,k_{j-1}}, s, k_{j-1}) \frac{1(Y_n(s) > 0)}{Y_n(s)} d\bar{M}(s), \end{aligned}$$

with $\tilde{e}_{n,j-1}$ as defined in (S6.70) with j replaced by $j-1$, where the second equality holds by the decomposition of $\hat{\Lambda}_n - \Lambda$ in (S6.73). Following $\tilde{e}_{n,j-1}(u,s,k) = E_0(u,s,k)1(X_j \geq s)1(Y_n(s) \geq \sqrt{n})$, we have a decomposition $D_{nj} = \widetilde{D}_{nj} + \widehat{D}_{nj}$, where

(S6.77)

$$\begin{aligned} \widetilde{D}_{nj} &\equiv \frac{m_{j-1}(U_{j,k_{j-1}} - Q_u[U_{k_{j-1}}])}{\sqrt{n - q_n}\sigma_{nj-1}\mathrm{Var}_{Q_u}(U_{k_{j-1}})} \int_{\mathcal{T}} \big\{E_0(U_{j,k_{j-1}}, s, k_{j-1}) - E_0(U_{j-1,k_{j-2}}, s, k_{j-2})\big\} \\ &\qquad \times \{d\hat{\Lambda}_n(s) - d\Lambda(s)\}; \\ \widehat{D}_{nj} &\equiv \frac{1}{\sqrt{n - q_n}} \frac{m_{j-1}(U_{j,k_{j-1}} - Q_u[U_{k_{j-1}}])}{\sigma_{nj-1}\mathrm{Var}_{Q_u}(U_{k_{j-1}})} \int_{\mathcal{T}} E_0(U_{j-1,k_{j-2}}, s, k_{j-2})1(X_j \geq s) \end{aligned}$$

$$\times \frac{1(Y_n(s) \geq \sqrt{n})}{Y_n(s)} d\bar{M}(s),$$

so that $|\sum_{j=q_n}^{n-1} D_{n,j+1}| \leq |\sum_{j=q_n}^{n-1} \widetilde{D}_{n,j+1}| + |\sum_{j=q_n}^{n-1} \widehat{D}_{n,j+1}|$. To show the desired result, it therefore suffices to show that both $|\sum_{j=q_n}^{n-1} \widetilde{D}_{n,j+1}|$ and $|\sum_{j=q_n}^{n-1} \widehat{D}_{n,j+1}|$ converge to zero in probability, which will be presented in Lemmas S6.16 and S6.17, respectively. Henceforth we state that $(j,s) \in \{q_n, \ldots, n-1\} \times \mathcal{T}$; $(j', s) \in \{q_n, \ldots, n-1\} \times \mathcal{T}$ and $(k,u) \in \mathcal{K}_n \times \mathbb{R}$ in which the ranges will be omitted for succinct presentation in forthcoming displays.

LEMMA S6.16. *Under the conditions of Theorem 4.1, $\sum_{j=q_n}^{n-1} \widetilde{D}_{n,j+1}$ converges to zero in probability as n goes to infinity, where $\widetilde{D}_{nj}$ is as defined in (S6.77) for $j = q_n + 1, \ldots, n$.*

PROOF. Using (A.1), (A.6) and the total variation of $\hat{\Lambda}_n - \Lambda$ over $\mathcal{T}$ is bounded by $|\hat{\Lambda}_n(\tau) + \Lambda(\tau)|$, we have that

$$\begin{aligned}
&\Big|\sum_{j=q_n}^{n-1} \widetilde{D}_{n,j+1}\Big| \\
&\lesssim \sqrt{n-q_n} \max_j \Big|\int_{\mathcal{T}} \{E_0(U_{j,k_{j-1}}, s, k_{j-1}) - E_0(U_{j-1,k_{j-2}}, s, k_{j-2})\}\{d\hat{\Lambda}_n(s) - d\Lambda(s)\}\Big| \\
&\leq \sqrt{n-q_n} \sup_{(j,s)} \Big|E_0(U_{j,k_{j-1}}, s, k_{j-1}) - E_0(U_{j-1,k_{j-2}}, s, k_{j-2})\Big|\big|\hat{\Lambda}_n(\tau) + \Lambda(\tau)\big| \\
&\leq \sqrt{n-q_n}\Big[\big|\hat{\Lambda}_n(\tau) - \Lambda(\tau)\big| + 2\Lambda(\tau)\Big] \sup_{(j,s)} \Big|E_0(U_{j,k_{j-1}}, s, k_{j-1}) - E_0(U_{j-1,k_{j-2}}, s, k_{j-2})\Big|,
\end{aligned}$$

where the last line holds by the triangle inequality.

Since $\sup_{t\in\mathcal{T}} |\hat{\Lambda}_n(t) - \Lambda(t)| \leq \sqrt{\log n}/\sqrt{n}$ with probability tending to one, taking the expectation on the above display gives that

$$\begin{aligned}
&E\Big[\Big|\sum_{j=q_n}^{n-1} \widetilde{D}_{n,j+1}\Big|\Big] \\
&\leq \sqrt{n-q_n}\Big[\frac{\sqrt{\log n}}{\sqrt{n}} + 2\Lambda(\tau)\Big] E\Big[\sup_{(j,s)} \big|E_0(U_{j,k_{j-1}}, s, k_{j-1}) - E_0(U_{j-1,k_{j-2}}, s, k_{j-2})\big|\Big] \to 0,
\end{aligned}$$

where the convergence follows (A.7). Hence we complete the proof. □

LEMMA S6.17. *Under the conditions of Theorem 4.1, $\sum_{j=q_n}^{n-1} \widehat{D}_{n,j+1}$ converges to zero in probability as n goes to infinity, where $\widehat{D}_{nj}$ is as defined in (S6.77) for $j = q_n + 1, \ldots, n$.*

PROOF. First note that, with respect to the filtration $\mathcal{O}_{nj} \equiv \sigma(O_1, \ldots, O_j, \mathcal{F}'_\tau)$ with $\mathcal{F}'_s$ defined in (S6.72), we have that $\widehat{D}_{nj}$ is a martingale difference sequence:

$$\begin{aligned}
E[\widehat{D}_{n,j+1}|\mathcal{O}_{nj}] =& \frac{m_j}{\sqrt{n-q_n}\sigma_{nj}\mathrm{Var}_{Q_u}(U_{k_j})}\Big\{\int_{\mathcal{T}} E_0(U_{j,k_{j-1}}, s, k_{j-1}) \\
&\times \frac{1(X_{j+1} \geq s, Y_n(s) \geq \sqrt{n})}{Y_n(s)} d\bar{M}(s)\Big\} E\big[U_{j+1,k_j} - Q_u[U_{k_j}]\big|\mathcal{O}_{nj}\big] = 0
\end{aligned}$$

for $j = q_n, \ldots, n-1$.

We also provide an upper bound on the conditional variance

$$\sum_{j=q_n}^{n-1} E[\widehat{D}_{n,j+1}^2|\mathcal{O}_{nj}] = \frac{1}{(n-q_n)}\sum_{j=q_n}^{n-1}\frac{1}{\sigma_{nj}^2\mathrm{Var}_{Q_u}^2(U_{k_j})}E\big[(U_{j+1,k_j}-Q_u[U_{k_j}])^2\big|\mathcal{O}_{nj}\big]$$
$$\times\Big\{\int_{\mathcal{T}}E_0(U_{j,k_{j-1}},s,k_{j-1})\frac{1(X_{j+1}\geq s, Y_n(s)\geq\sqrt{n})}{Y_n(s)}d\bar{M}(s)\Big\}^2$$
$$\leq K_0'\sup_{(j,k,u)}\Big\{\int_{\mathcal{T}}E_0(u,s,k)\frac{1(X_{j+1}\geq s, Y_n(s)\geq\sqrt{n})}{Y_n(s)}d\bar{M}(s)\Big\}^2,$$

where $K_0'>0$ is a constant such that for all n

$$\max_j\left\{\frac{E\big[(U_{j+1,k_j}-Q_u[U_{k_j}])^2\big|\mathcal{O}_{nj}\big]}{\sigma_{nj}^2\mathrm{Var}_{Q_u}^2(U_{k_j})}\right\}\leq\max_{(j,k)}\left\{\frac{E\big[(U_{j+1,k}-Q_u[U_k])^2\big|\mathcal{O}_{nj}\big]}{\sigma_{nj}^2\mathrm{Var}_{Q_u}^2(U_k)}\right\}\leq K_0'$$

almost surely; this constant exists by (A.1) and (A.6). We see that

$$\sup_{(j,k,u)}\Big\{\int_{\mathcal{T}}E_0(u,s,k)\frac{1(X_{j+1}\geq s, Y_n(s)\geq\sqrt{n})}{Y_n(s)}d\bar{M}(s)\Big\}^2\leq\frac{\log(n\vee p_n)}{\sqrt{q_n}}.$$

with probability tending to one, which can be seen from (S6.15.1) of Lemma S6.15. We therefore find that the conditional variance $\sum_{j=q_n}^{n-1}E[\widehat{D}_{n,j+1}^2|\mathcal{O}_{nj}]\to 0$ in probability.

By the martingale central theorem given in Theorem 1.2 of Kundu, Majumdar and Mukherjee (2000), $\sum_{j=q_n}^{n-1}\widehat{D}_{n,j+1}$ converges in probability to zero. Indeed, the conditional Lindeberg condition holds trivially: for every $\varepsilon>0$,

$$\sum_{j=q_n}^{n-1}E\big[\widehat{D}_{n,j+1}^2 1\big(\big|\widehat{D}_{n,j+1}\big|>\varepsilon\big)\big|\mathcal{O}_{nj}\big]\leq\sum_{j=q_n}^{n-1}E[\widehat{D}_{n,j+1}^2|\mathcal{O}_{nj}]\xrightarrow{p}0.$$

This completes the proof. □

Recall that for each (j,k),

$$e_{nj}(u,s,k)=[E_0(u,s,k)-\hat{E}_j(u,s,k)]1(X_{j+1}\geq s)1\Big(\sup_{(k,s,u)}|E_0(u,s,k)-\hat{E}_j(u,s,k)|$$
$$\leq\tilde{K}K_{nj},\inf_s Y_n(s)\geq\sqrt{n}\Big).$$

LEMMA S6.18. *Under the conditions of Theorem 4.1,* $\sum_{j=q_n}^{n-1}Q_{n,j+1}$ *converges to zero in probability as* n *goes to infinity, where for* $j=q_n+1,\dots,n$,

$$Q_{nj}\equiv\frac{1}{\sqrt{n-q_n}}\frac{m_{j-1}(U_{j,k_{j-1}}-Q_u[U_{k_{j-1}}])}{\sigma_{nj-1}\mathrm{Var}_{Q_u}(U_{k_{j-1}})}\int_{\mathcal{T}}e_{n,j-1}(U_{j,k_{j-1}},s,k_{j-1})\{d\hat{\Lambda}_n(s)-d\Lambda(s)\},$$

and $e_{n,j-1}$ *is as defined in* (S6.71) *with* j *replaced by* $j-1$.

PROOF. First using (A.1) and (A.6) to bound the middle part of Q_{nj}, together with the expression of $\hat{\Lambda}_n-\Lambda$ in (S6.73), it upper bounds $\big|\sum_{j=q_n}^{n-1}Q_{n,j+1}\big|$ by (up to a constant that does not depend on (j,n))

$$\sqrt{n-q_n}\max_{(j,k)}\Big|\int_{\mathcal{T}}e_{nj}(U_{j+1,k},s,k)1(X_{j+1}\geq s)\frac{1(Y_n(s)\geq\sqrt{n})}{Y_n(s)}d\bar{M}(s)\Big|.$$

Recall that $\mathcal{I}_n = \{(j,k) : j \in \{q_n, \dots, n-1\}, k \in \mathcal{K}_n\}$ and $K_{nq_n} = \sqrt{\log(n \vee p_n)/q_n}$. For each $(j,k) \in \mathcal{I}_n$, Lemma S6.2, along with (A.1) and (A.7), indicates that the integrand in the above martingale integral, namely

$$s \mapsto e_{nj}(U_{j+1,k}, s, k)\sqrt{n-q_n}\frac{1(X_{j+1} \geq s\,, Y_n(s) \geq \sqrt{n})}{Y_n(s)},$$

belongs to $\mathcal{H}_n(\tilde{K})$ with probability tending to one, where $\mathcal{H}_n(\tilde{K})$ is the class of càglàd functions in $\mathcal{BV}(\mathcal{T}, \tilde{K}K_{nq_n}, \tilde{M}_1)$. Therefore to show that $\sum_{j=q_n}^{n-1} Q_{n,j+1} \xrightarrow{p} 0$, it suffices to show that for $\eta > 0$,

$$\mathrm{P}\Bigg(\sup_{h \in \mathcal{H}_n(\tilde{K})}\bigg|\int_{\mathcal{T}} h(s)\,d\bar{M}(s)\bigg| > \eta\Bigg) \to 0$$

under the assumed conditions for Theorem 4.1: $n/q_n = O(1)$ and $\log(n \vee p_n)/q_n^{1/4} \to 0$.

To show the desired result, we use the exponential inequality in Theorem 3.1 of van de Geer (1995). For fixed n, the aggregated counting process $\{N_n(t) : t \in \mathcal{T}\}$ plays the role of $\{N(t) : t \in [0,T]\}$, the compensator $\{A_n(t) : t \in \mathcal{T}\}$ is $\{A(t) : t \in [0,T]\}$, the integrand h in the class $\mathcal{H}_n(\tilde{K})$ corresponds to $g \in \mathscr{G}$, τ is T, $\sigma^2_{n\tau}$ is σ^2_T, and $d^2_{n\tau}(h,0)$ is $d^2_T(g,0)$. Then we have the martingale process $\{\bar{M}(t) = N_n(t) - A_n(t) : t \in \mathcal{T}\}$. In addition, n is now added as a subscript to van de Geer's various constants L, K, ε, b and C_1, C_2, C_3, C_4. The upper bound $H_n(\delta, b_n, B)$ on the δ-bracketing entropy of $\mathcal{H}_n(\tilde{K})$ is as defined in van de Geer (1995), where B is a measurable subset of $\{A_n(\tau) \leq \sigma^2_{n\tau}\}$.

The compensator in our case is

$$A_n(t) = \int_{-\infty}^{t} Y_n(s)\,d\Lambda(s)$$

so that $A_n(\tau) \leq \sigma^2_{n\tau} = n\Lambda(\tau)$. Note also that for any $h \in \mathcal{H}_n(\tilde{K})$,

$$d^2_{n\tau}(h,0) = \frac{1}{2}\int_{-\infty}^{\tau}\big[\exp(h(s)) - \exp(0)\big]^2 dA_n(s) \leq \frac{1}{2}\exp\big(4\tilde{K}K_{nq_n}\big)n\Lambda(\tau) \equiv b_n^2,$$

where we have used the inequality $|e^x - 1| \leq |x|e^{|x|} \leq e^{2|x|}$. Moreover, we may take $H_n(\delta, b_n, B) = c_0\sqrt{n}/\delta$ for some constant $c_0 > 0$ (Example 19.11 of van der Vaart, 1998).

We need to check condition (3.2) of van de Geer (1995), namely that

$$\frac{\varepsilon_n b_n^2}{C_{n1}} \geq \int_{\varepsilon_n b_n^2/(C_{n2}\sigma_{n\tau}) \wedge b_n/8}^{b_n} \sqrt{H_n(x, b_n, B)}\,dx \vee b_n,$$

for appropriate choices of the various constants. For now assume $\varepsilon_n \in (0,1]$, and take $C_{n2} = 8\varepsilon_n b_n/\sqrt{n\Lambda(\tau)}$. Together with the definition of $\sigma_{n\tau}$, this leads to

$$\frac{\varepsilon_n b_n^2}{C_{n2}\sigma_{n\tau}} = \varepsilon_n\frac{\sqrt{n\Lambda(\tau)}}{8\varepsilon_n b_n}b_n^2\frac{1}{\sqrt{n\Lambda(\tau)}} \geq \frac{b_n}{8} \quad \text{and}$$

$$\int_{\varepsilon_n b_n^2/(C_{n2}\sigma_{n\tau}) \wedge b_n/8}^{b_n} \sqrt{H(x, b_n, B)}dx = \sqrt{c_0}n^{1/4}\int_{b_n/8}^{b_n}\frac{1}{\sqrt{x}}dx = 2\sqrt{c_0}\bigg\{1 - \frac{1}{\sqrt{8}}\bigg\}n^{1/4}\sqrt{b_n}.$$

Taking $C_{n1} = \varepsilon_n b_n^{3/2}/\{2\sqrt{c_0}\big[1 - 8^{-1/2}\big]n^{1/4}\}$ shows that van de Geer's condition (3.2) is satisfied. Then, applying her result with $B \subset \{A_n(\tau) \leq \sigma^2_{n\tau}\}$ having probability one, gives

(S6.78)

$$\mathrm{P}\Bigg(\sup_{h \in \mathcal{H}_n(\tilde{K})}\bigg|\int_{\mathcal{T}} h(s)d\bar{M}(s)\bigg| > \varepsilon_n b_n^2\Bigg)$$

$$\leq \mathrm{P}\bigg(\bigg\{\bigg|\int_{\mathcal{T}} h(s)d(N_n - A_n)(s)\bigg| \geq \varepsilon_n b_n^2 \text{ and } d_{n\tau}^2(h,0) \leq b_n^2 \text{ for some } h \in \mathcal{H}_n(\tilde{K})\bigg\} \cap B\bigg)$$

$$\leq C_{n3} \exp\bigg(-\frac{\varepsilon_n^2 b_n^2}{C_{n4}}\bigg),$$

where C_{n3} and C_{n4} are specified below.

From the proof of van de Geer's theorem, we can take $C_{n3} = 2$ and $C_{n4} \geq 2(\varepsilon_n K_n + c_n)$, where

$$c_n = \frac{4\{\exp(-L_n) - 1 - L_n\}}{\{\exp(-L_n) - 1\}^2},$$

$L_n = \tilde{K} K_{nq_n}$ and $K_n = \tilde{K} K_{nq_n}$. Note that $c_n < 0$ since $L_n > 0$, so we can take $C_{n4} = 2\varepsilon_n K_n$. Finally, specifying $\varepsilon_n = n^{-5/4}$, we have

$$\begin{aligned}
\frac{\varepsilon_n^2 b_n^2}{C_{n4}} &= \frac{n^{-5/2} 2^{-1} \exp\big(4\tilde{K} K_{nq_n}\big) n\Lambda(\tau)}{2n^{-5/4} K_n} = \frac{n^{-3/2} \exp\big(4\tilde{K} K_{nq_n}\big)\Lambda(\tau)}{4n^{-5/4} K_n} \\
&= \frac{\exp\big(4\tilde{K}\sqrt{\log(n \vee p_n)}\, q_n^{-1/2}\big)\Lambda(\tau)}{4\tilde{K}\sqrt{\log(n \vee p_n)}\, n^{1/4} q_n^{-1/2}} \\
&= \frac{\exp\big(4\tilde{K}\sqrt{\log(n \vee p_n)}\, q_n^{-1/2}\big)\Lambda(\tau)}{4\tilde{K}} \frac{q_n^{1/4}}{\sqrt{\log(n \vee p_n)}} \frac{q_n^{1/4}}{n^{1/4}} \to \infty,
\end{aligned}$$

under the conditions of our Theorem 4.1: $n/q_n = O(1)$ and $\log(n \vee p_n)/q_n^{1/4} \to 0$, and note that $\tilde{K} > 0$. Thus, from (S6.78), we have

$$\limsup_{n\to\infty} \mathrm{P}\bigg(\sup_{h \in \mathcal{H}_n(\tilde{K})} \bigg|\int_{\mathcal{T}} h(s) d\bar{M}(s)\bigg| > \varepsilon_n b_n^2\bigg) = 0,$$

and since $\varepsilon_n b_n^2 = n^{-5/4} \exp\big(4\tilde{K}\sqrt{\log(n \vee p_n)}\, q_n^{-1/2}\big) n\Lambda(\tau)/2 \to 0$, the proof is complete. $\square$

LEMMA S6.19. *Under the conditions of Theorem 4.1, (I)$1_{\mathcal{L}_n}$ is asymptotically negligible.*

PROOF. Recall that

$$\begin{aligned}
\text{(I)} = \frac{1}{\sqrt{n - q_n}} \sum_{j=q_n}^{n-1} \bigg\{ &\frac{\hat{\sigma}_{nj}^{-1} m_j}{\mathrm{Var}_{\mathbb{Q}_j}(U_{k_j})} (U_{j+1,k_j} - \mathbb{Q}_j[U_{k_j}]) \\
&\times \int_{\mathcal{T}} [E_0(U_{j+1,k_j}, s, k_j) - \hat{E}_j(U_{j+1,k_j}, s, k_j)] d\hat{M}_{j+1}(s) \bigg\}.
\end{aligned}$$

Let $dM_{j+1}(s) = 1(X_{j+1} \in ds, \delta_{j+1} = 0) - 1(X_{j+1} \geq s)d\Lambda(s)$, and $d\hat{M}_{j+1}(s) - dM_{j+1}(s) = -1(X_{j+1} \geq s)\big\{d\hat{\Lambda}_n(s) - d\Lambda(s)\big\}$, so that

$$\begin{aligned}
&\bigg|1_{\mathcal{L}_n} \int_{\mathcal{T}} [E_0(U_{j+1,k_j}, s, k_j) - \hat{E}_j(U_{j+1,k_j}, s, k_j)]\big\{d\hat{M}_{j+1}(s) - dM_{j+1}(s)\big\}\bigg| \\
&\quad = \bigg|1_{\mathcal{L}_n} \int_{\mathcal{T}} [E_0(U_{j+1,k_j}, s, k_j) - \hat{E}_j(U_{j+1,k_j}, s, k_j)] 1(X_{j+1} \geq s)\big\{d\hat{\Lambda}_n(s) - d\Lambda(s)\big\}\bigg|.
\end{aligned}$$

Along with the triangle inequality, the above results imply that

$$\begin{aligned}|(\mathrm{I})1_{\mathcal{L}_n}| \le &\left|\frac{1}{\sqrt{n-q_n}}\sum_{j=q_n}^{n-1}\frac{\hat\sigma_{nj}^{-1}m_j}{\mathrm{Var}_{\mathbb{Q}_j}(U_{k_j})}(U_{j+1,k_j}-\mathbb{Q}_j[U_{k_j}])\right.\\ &\qquad\left.\times 1_{\mathcal{L}_n}\int_{\mathcal{T}}\big[E_0(U_{j+1,k_j},s,k_j)-\hat E_j(U_{j+1,k_j},s,k_j)\big]dM_{j+1}(s)\right|\\ &+\left|\frac{1}{\sqrt{n-q_n}}\sum_{j=q_n}^{n-1}\frac{\hat\sigma_{nj}^{-1}m_j}{\mathrm{Var}_{\mathbb{Q}_j}(U_{k_j})}(U_{j+1,k_j}-\mathbb{Q}_j[U_{k_j}])\right.\\ &\left.\times 1_{\mathcal{L}_n}\int_{\mathcal{T}}\big[E_0(U_{j+1,k_j},s,k_j)-\hat E_j(U_{j+1,k_j},s,k_j)\big]1(X_{j+1}\ge s)\big\{d\hat\Lambda_n(s)-d\Lambda(s)\big\}\right|.\end{aligned}$$

We further apply Lemma S6.13 to the above display, taking

$$\begin{aligned}&f_{n1}(O_{j+1,k_j})=1_{\mathcal{L}_n}\int_{\mathcal{T}}\big[E_0(U_{j+1,k_j},s,k_j)-\hat E_j(U_{j+1,k_j},s,k_j)\big]dM_{j+1}(s),\ \text{with}\\ &\sup_o\big|f_{n1}(o)\big|\le\sup_{(j,k,s,u)}\big|E_0(u,s,k)-\hat E_j(u,s,k)\big|(1+\Lambda(\tau))\lesssim K_{nq_n},\end{aligned}$$

and

$$\begin{aligned}&f_{n2}(O_{j+1,k_j})=1_{\mathcal{L}_n}\int_{\mathcal{T}}\big[E_0(U_{j+1,k_j},s,k_j)-\hat E_j(U_{j+1,k_j},s,k_j)\big]1(X_{j+1}\ge s)\big\{d\hat\Lambda_n(s)-d\Lambda(s)\big\};\\ &\sup_o\big|f_{n2}(o)\big|=\sup_{(j,k,u)}\left|1_{\mathcal{L}_n}\int_{\mathcal{T}}\big[E_0(u,s,k)-\hat E_j(u,s,k)\big]1(X_{j+1}\ge s)\big\{d\hat\Lambda_n(s)-d\Lambda(s)\big\}\right|.\end{aligned}$$

By the definition of e_{nj} in (S6.71) that implies

$$e_{nj}(u,s,k)=1_{\mathcal{L}_n}\big[E_0(u,s,k)-\hat E_j(u,s,k)\big]1(X_{j+1}\ge s), \tag{S6.79}$$

so we see from (S6.15.2) of Lemma S6.15 that $\sup_o\big|f_{n2}(o)\big|\lesssim\sqrt{\log(n\vee p_n)}/q_n^{1/4}$ with probability tending to one. Using Lemma S6.13, it is implied that

$$\begin{aligned}|(\mathrm{I})1_{\mathcal{L}_n}| \lesssim &\left|\frac{1}{\sqrt{n-q_n}}\sum_{j=q_n}^{n-1}\frac{\sigma_{nj}^{-1}m_j}{\mathrm{Var}_{Q_u}(U_{k_j})}(U_{j+1,k_j}-Q_u[U_{k_j}])\right.\\ &\qquad\left.\times 1_{\mathcal{L}_n}\int_{\mathcal{T}}\big[E_0(U_{j+1,k_j},s,k_j)-\hat E_j(U_{j+1,k_j},s,k_j)\big]dM_{j+1}(s)\right|\\ &+\left|\frac{1}{\sqrt{n-q_n}}\sum_{j=q_n}^{n-1}\frac{\sigma_{nj}^{-1}m_j}{\mathrm{Var}_{Q_u}(U_{k_j})}(U_{j+1,k_j}-Q_u[U_{k_j}])1_{\mathcal{L}_n}\right.\\ &\left.\times\int_{\mathcal{T}}\big[E_0(U_{j+1,k_j},s,k_j)-\hat E_j(U_{j+1,k_j},s,k_j)\big]1(X_{j+1}\ge s)\big\{d\hat\Lambda_n(s)-d\Lambda(s)\big\}\right|+o_p(1).\end{aligned}$$

Moreover, note that $1_{\mathcal{L}_n}\le 1\big(\sup_{(j,k,s,u)}|E_0(u,s,k)-\hat E_j(u,s,k)|\lesssim K_{nq_n}\big)$, and then define

$$\begin{aligned}\bar e_{nj}(u,s,k)=&[E_0(u,s,k)-\hat E_j(u,s,k)]1(X_{j+1}\ge s)\\ &\times 1\left(\sup_{(j,k,s,u)}\big|E_0(u,s,k)-\hat E_j(u,s,k)\big|\lesssim K_{nq_n}\right).\end{aligned} \tag{S6.80}$$

Then by (S6.79) and (S6.80), we have that

$$|(\mathrm{I})1_{\mathcal{L}_n}| \lesssim \left| \frac{1}{\sqrt{n-q_n}} \sum_{j=q_n}^{n-1} \frac{m_j(U_{j+1,k_j} - Q_u[U_{k_j}])}{\sigma_{nj}\mathrm{Var}_{Q_u}(U_{k_j})} \int_{\mathcal{T}} \bar{e}_{nj}(U_{j+1,k_j}, s, k_j) dM_{j+1}(s) \right|$$
$$+ \left| \frac{1}{\sqrt{n-q_n}} \sum_{j=q_n}^{n-1} \frac{m_j(U_{j+1,k_j} - Q_u[U_{k_j}])}{\sigma_{nj}\mathrm{Var}_{Q_u}(U_{k_j})} \int_{\mathcal{T}} e_{nj}(U_{j+1,k_j}, s, k_j)\{d\hat{\Lambda}_n(s) - d\Lambda(s)\} \right|$$
$$+ o_p(1),$$

where the middle term converges to zero in probability by Lemma S6.18.

Now we deal with the first term on the right-hand-side of the above inequality. Fix n and for $j \in \{q_n+1, \dots, n\}$,

$$\bar{H}_{nj} \equiv \frac{m_{j-1}(U_{j,k_{j-1}} - Q_u[U_{k_{j-1}}])}{\sqrt{n-q_n}\sigma_{n,j-1}\mathrm{Var}_{Q_u}(U_{k_{j-1}})} \int_{\mathcal{T}} \bar{e}_{n,j-1}(U_{j,k_{j-1}}, s, k_{j-1}) dM_j(s).$$

From (S6.80) and by (A.1),

$$\sup_{(j,s)} |\bar{e}_{nj}(U_{j+1,k_j}, s, k_j)| \le \sup_{(j,k,s,u)} \left| E_0(u,s,k) - \hat{E}_j(u,s,k) \right| \lesssim K_{nq_n} \text{ a.s.} \tag{S6.81}$$

Along with the uniform boundedness of U_k almost surely in (A.1) and that σ_{nj} and $\mathrm{Var}_{Q_u}(U_k)$ are uniformly bounded away from zero in (A.6), (S6.81) implies that there exists a $B_n \equiv K'\sqrt{\log(n \vee p_n)}/\sqrt{(n-q_n)q_n}$ for some constant $K' > 0$ such that $\max_j |\bar{H}_{nj}| \le B_n$ almost surely.

Define the filtration $\mathcal{O}_{nj} \equiv \sigma(O_1, \dots, O_j, \boldsymbol{U}_{j+1})$. We know that $\{(\bar{H}_{nj}, \mathcal{O}_{nj}), j = q_n+1, \dots, n\}$ is a martingale difference sequence because $E|\bar{H}_{nj}| < \infty$; $\bar{H}_{nj}$ is $\mathcal{O}_{nj}$-measurable, and for $j = q_n, \dots, n-1$,

$$\begin{aligned}
&E\left[\bar{H}_{n,j+1} \middle| \mathcal{O}_{nj}\right] \\
&= \frac{1}{\sqrt{n-q_n}} \frac{m_j(U_{j+1,k_j} - Q_u[U_{k_j}])}{\sigma_{nj}\mathrm{Var}_{Q_u}(U_{k_j})} \int_{\mathcal{T}} \bar{e}_{nj}(U_{j+1,k_j}, s, k_j) E\left[dM_{j+1}(s) \middle| \mathcal{O}_{nj}\right] \\
&= \frac{1}{\sqrt{n-q_n}} \frac{m_j(U_{j+1,k_j} - Q_u[U_{k_j}])}{\sigma_{nj}\mathrm{Var}_{Q_u}(U_{k_j})} \int_{\mathcal{T}} \bar{e}_{nj}(U_{j+1,k_j}, s, k_j) E[1(T_{j+1} \ge s)|\boldsymbol{U}_{j+1}] \\
&\qquad \times E\left[1(C_{j+1} \in ds) - 1(C_{j+1} \ge s)d\Lambda(s)\right] \\
&= 0,
\end{aligned}$$

where the second equality holds by the independent censoring assumption, and the last step follows from the definition $d\Lambda(s) = \mathrm{P}(C \in ds)/\mathrm{P}(C \ge s)$. Then for $\varepsilon > 0$,

$$\begin{aligned}
&\mathrm{P}\left(\left| \sum_{j=q_n}^{n-1} \bar{H}_{n,j+1} \right| \ge \varepsilon \right) \le \varepsilon^{-2} E\left[\left(\sum_{j=q_n}^{n-1} \bar{H}_{n,j+1} \right)^2 \right] \\
&= \varepsilon^{-2} \left(\sum_{j=q_n}^{n-1} E\left[\bar{H}_{n,j+1}^2\right] + 2 \sum_{q_n \le i < j \le n-1} E\left[\bar{H}_{n,i+1} E\left[\bar{H}_{n,j+1} \middle| \mathcal{O}_{nj} \right] \right] \right) \\
&= \varepsilon^{-2} \sum_{j=q_n}^{n-1} E\left[\bar{H}_{n,j+1}^2\right] \le \varepsilon^{-2}(n-q_n)B_n^2 \to 0.
\end{aligned}$$

As $|(\mathrm{I})1_{\mathcal{L}_n}| \lesssim |\sum_{j=q_n}^{n-1} \bar{H}_{n,j+1}| + o_p(1)$, we conclude that $(\mathrm{I})1_{\mathcal{L}_n} = o_p(1)$. $\square$

LEMMA S6.20. *Under the conditions of Theorem 4.1, (II)$1_{\mathcal{L}_n}$ is asymptotically negligible.*

PROOF. First we have

$$(\text{II}) = \frac{1}{\sqrt{n-q_n}} \sum_{j=q_n}^{n-1} \frac{m_j(U_{j+1,k_j} - \mathbb{Q}_n[U_{k_j}])}{\hat{\sigma}_{nj} \text{Var}_{\mathbb{Q}_n}(U_{k_j})} \Big[\delta_{j+1} X_{j+1} \Big(\frac{1}{\hat{G}_n(X_{j+1})} - \frac{1}{G(X_{j+1})} \Big) + \int_{\mathcal{T}} E_0(U_{j+1,k_j}, s, k_j) 1(X_{j+1} \geq s) \{ d\hat{\Lambda}_n(s) - d\Lambda(s) \} \Big].$$

By (A.2), we can fix an $\tilde{\epsilon}$ such that $0 < \tilde{\epsilon} < G(\tau)$. Fix $0 < r < 1$, and let $\mathcal{G}_r$ be the collection of monotone nonincreasing càdlàg functions $\tilde{G}\colon \mathcal{T} \to [0,1]$ such that $\tilde{G}(\tau) > \tilde{\epsilon}$, and $\sup_{s \in \mathcal{T}} |\tilde{G}(s)/G(s) - 1| \leq r$. Note that $G \in \mathcal{G}_r \subset \mathcal{G}$ that was defined right before (S6.39), and $\mathrm{P}(\hat{G}_n \in \mathcal{G}_r) \to 1$, using the argument involving $\mathcal{R}_{n1}$ in the proof of Lemma S6.7.

We first give an upper-bound of $|(\text{II})1_{\mathcal{L}_n}|$ by using Lemma S6.13, taking

$$f_n(O_{j+1,k_j}) = \delta_{j+1} X_{j+1} \Big(\frac{1}{\hat{G}_n(X_{j+1})} - \frac{1}{G(X_{j+1})} \Big) + \int_{\mathcal{T}} E_0(U_{j+1,k_j}, s, k_j) 1(X_{j+1} \geq s) \{ d\hat{\Lambda}_n(s) - d\Lambda(s) \},$$

and showing that $\sup_o |f_n(o) 1_{\mathcal{L}_n}| \lesssim \sqrt{\log(n \vee p_n)}/q_n^{1/4}$ below. First, using a Taylor expansion of $z \mapsto \delta x 1_{\mathcal{L}_n}/z$ around $z = G(x)$ gives that

$$\sup_o \Big| \delta x \Big(\frac{1}{\hat{G}_n(x)} - \frac{1}{G(x)} \Big) 1_{\mathcal{L}_n} \Big| \leq \frac{\tau}{G(\tau)} \Big[\sum_{r=1}^{\infty} \Big(\sup_t \Big| \frac{\hat{G}_n(t)}{G(t)} - 1 \Big| \Big)^r \Big] 1_{\mathcal{L}_n}$$
$$\leq \frac{\tau}{G(\tau)} \sum_{r=1}^{\infty} \Big(\sqrt{\frac{\log n}{n}} \Big)^r = \frac{\tau}{G(\tau)} \frac{\sqrt{\log n}}{\sqrt{n} - \sqrt{\log n}} \lesssim \sqrt{\frac{\log n}{n}},$$

where the result follows by the occurrence of $\mathcal{B}_n(\tilde{K}) \supset \mathcal{L}_n$, $\sqrt{\log n/n} \leq 1$ and $G(\tau) > 0$ in (A.2). Along with the above result and (S6.15.1) of Lemma S6.15, the triangle inequality gives that with probability tending to one,

$$\sup_o |f_n(o) 1_{\mathcal{L}_n}| \lesssim \sqrt{\frac{\log n}{n}} + \frac{\sqrt{\log(n \vee p_n)}}{q_n^{1/4}} \leq 2 \frac{\sqrt{\log(n \vee p_n)}}{q_n^{1/4}}.$$

Therefore along with the above results, Lemma S6.13 implies that

$$|(\text{II})1_{\mathcal{L}_n}| \lesssim \Big| \frac{1}{\sqrt{n-q_n}} \sum_{j=q_n}^{n-1} \frac{m_j(U_{j+1,k_j} - Q_u[U_{k_j}])}{\sigma_{nj} \text{Var}_{Q_u}(U_{k_j})} \Big[\delta_{j+1} X_{j+1} \Big(\frac{1}{\hat{G}_n(X_{j+1})} - \frac{1}{G(X_{j+1})} \Big) + \int_{\mathcal{T}} E_0(U_{j+1,k_j}, s, k_j) 1(X_{j+1} \geq s) \{ d\hat{\Lambda}_n(s) - d\Lambda(s) \} \Big] \Big| 1_{\mathcal{L}_n} + o_p(1).$$

According to the definition of $\mathcal{G}_r$, the events contained in $\mathcal{L}_n$ and the definition of $\tilde{e}_{nj}$ as in (S6.70), the above display further leads to

(S6.82)

$$|(\text{II})1_{\mathcal{L}_n}| \leq \Big| \frac{1}{\sqrt{n-q_n}} \sum_{j=q_n}^{n-1} \frac{m_j(U_{j+1,k_j} - Q_u[U_{k_j}])}{\sigma_{nj} \text{Var}_{Q_u}(U_{k_j})} \int_{\mathcal{T}} \tilde{e}_{nj}(U_{j+1,k_j}, s, k_j) \{ d\hat{\Lambda}_n(s) - d\Lambda(s) \} \Big|$$

$$+\frac{1}{\sqrt{n-q_n}}\sup_{\tilde{G}\in\mathcal{G}_r}\Big|\sum_{j=q_n}^{n-1}\frac{m_j(U_{j+1,k_j}-Q_u[U_{k_j}])}{\sigma_{nj}\mathrm{Var}_{Q_u}(U_{k_j})}\delta_{j+1}X_{j+1}\Big(\frac{1}{\tilde{G}(X_{j+1})}-\frac{1}{G(X_{j+1})}\Big)\Big|1_{\mathcal{L}_n}$$

$$+o_p(1),$$

where the first term converges to zero in probability, applying Lemmas S6.16 and S6.17. Therefore it remains to show that the middle term on the right-hand-side converges to zero in probability, using the properties of martingale difference arrays.

Fix n. Define a process $\{\tilde{S}_n(\tilde{G}):\tilde{G}\in\mathcal{G}_r\}$ by $\tilde{S}_n(\tilde{G})\equiv\sum_{j=q_n+1}^{n}V_{nj}(\tilde{G})$, where

$$V_{nj}(\tilde{G})=\frac{m_{j-1}(U_{j,k_{j-1}}-Q_u[U_{k_{j-1}}])}{\sqrt{n-q_n}\sigma_{n,j-1}\mathrm{Var}_{Q_u}(U_{k_{j-1}})}\delta_jX_j\Big(\frac{1}{\tilde{G}(X_j)}-\frac{1}{G(X_j)}\Big).$$

We see that $E|V_{nj}(\tilde{G})|<\infty$ by (A.1), (A.2) and (A.6). Define the filtration $\mathcal{O}_{nj}=\sigma(O_1,\ldots,O_j,\delta_{j+1},X_{j+1})$ and we have that, for each $\tilde{G}$, $E[V_{n,j+1}(\tilde{G})|\mathcal{O}_{nj}]=0$. Therefore $\{(V_{nj}(\tilde{G}),\mathcal{O}_{nj}):j=q_n+1,\ldots,n,\ \tilde{G}\in\mathcal{G}_r\}$ is an array of martingale-differences of adapted processes indexed by $\mathcal{G}_r$. Note that the class $\mathcal{G}_r$ has a finite uniform entropy integral (see Lemma 2.6.13 in van der Vaart and Wellner, 1996). For all $\tilde{G},G'\in\mathcal{G}_r$,

$$\tilde{\sigma}_n^2(\tilde{G},G')\equiv\sum_{j=q_n}^{n-1}E\big[\{V_{n,j+1}(\tilde{G})-V_{n,j+1}(G')\}^2|\mathcal{O}_{nj}\big]$$

$$=\frac{1}{(n-q_n)}\sum_{j=q_n}^{n-1}X_{j+1}^2\Big(\frac{1}{\tilde{G}(X_{j+1})}-\frac{1}{G'(X_{j+1})}\Big)^2E\Big[\frac{(U_{j+1,k_j}-Q_u[U_{k_j}])^2}{\sigma_{nj}^2\mathrm{Var}_{Q_u}^2(U_{k_j})}\Big|\mathcal{O}_{nj}\Big]$$

$$\le\frac{\bar{K}_0^2\tau^2}{\tilde{\epsilon}^4}\frac{1}{(n-q_n)}\sum_{j=q_n}^{n-1}\{\tilde{G}(X_{j+1})-G'(X_{j+1})\}^2,$$

where the inequality holds because $X\le\tau$; for each j, $G(X_{j+1})\ge G(\tau)>\tilde{\epsilon}>0$ by (A.2); for each j and any $\bar{G}\in\mathcal{G}_r$, $\bar{G}(X_{j+1})\ge\bar{G}(\tau)>\tilde{\epsilon}>0$ and $\tilde{G},G'\in\mathcal{G}_r$; for each j, $|U_{j+1,k_j}-Q_u[U_{k_j}]|/[\sigma_{nj}\mathrm{Var}_{Q_u}(U_{k_j})]\le\bar{K}_0$ almost surely by (A.1) and (A.6), for some positive constant $\bar{K}_0$.

Let μ_n be the empirical distribution of $\{O_{q_n},\ldots,O_{n-1}\}$ normalized by $\sup_{s\in\mathcal{T}}\big|\hat{G}_n(s)-G(s)\big|$. Following the notation of Theorem 1 of Bae, Jun and Levental (2010), it gives that for any $\tilde{G},G'\in\mathcal{G}_r$,

$$d_{\mu_n}^{(2)}(\tilde{G},G')^2\equiv\frac{1}{(n-q_n)}\sum_{j=q_n}^{n-1}\frac{\{\tilde{G}(X_{j+1})-G'(X_{j+1})\}^2}{\sup_{s\in\mathcal{T}}|\hat{G}_n(s)-G(s)|}.$$

Therefore, checking condition (3) of Theorem 1 of Bae, Jun and Levental (2010) in our case, we have that for any positive constant L,

$$\mathrm{P}\Big(\sup_{\tilde{G},G'\in\mathcal{G}_r}\frac{\tilde{\sigma}_n^2(\tilde{G},G')}{d_{\mu_n}^{(2)}(\tilde{G},G')^2}\ge L\Big)\le\mathrm{P}\Big(\mathcal{L}_n,\ \sup_{\tilde{G},G'\in\mathcal{G}_r}\frac{\tilde{\sigma}_n^2(\tilde{G},G')}{d_{\mu_n}^{(2)}(\tilde{G},G')^2}\ge L\Big)+\mathrm{P}\Big(\mathcal{L}_n^c\Big)$$

$$\le\frac{1}{L}E\Big[\sup_{\tilde{G},G'\in\mathcal{G}_r}\frac{\tilde{\sigma}_n^2(\tilde{G},G')}{d_{\mu_n}^{(2)}(\tilde{G},G')^2}1_{\mathcal{L}_n}\Big]+\mathrm{P}\Big(\mathcal{L}_n^c\Big)$$

$$\le\frac{\bar{K}_0^2\tau^2}{L\tilde{\epsilon}^4}E\Big[\sup_{s\in\mathcal{T}}\Big|\hat{G}_n(s)-G(s)\Big|1_{\mathcal{L}_n}\Big]+\mathrm{P}\Big(\mathcal{L}_n^c\Big)$$

$$\leq \frac{\bar{K}_0^2\tau^2}{L\tilde{\epsilon}^4} E\left[\sup_{s\in\mathcal{T}}\left|\frac{\hat{G}_n(s)}{G(s)}-1\right|1_{\mathcal{L}_n}\right] + \mathrm{P}\left(\mathcal{L}_n^c\right)$$
$$\leq \frac{\bar{K}_0^2\tau^2}{L\tilde{\epsilon}^4}\sqrt{\frac{\log n}{n}} + \mathrm{P}\left(\mathcal{L}_n^c\right) \to 0 \text{ as } n\to\infty.$$

The Lindeberg condition holds trivially in our case: note that $|V_{n,j+1}(\tilde{G})| \lesssim 1/\sqrt{n-q_n}$, so for any fixed $\epsilon > 0$ and for all n sufficiently large we have

$$\frac{1}{\epsilon}\sum_{j=q_n}^{n-1} E[V_{n,j+1}^2(\tilde{G})1(V_{n,j+1}(\tilde{G}) > \epsilon)] = 0.$$

Now appealing to Theorem 1 of Bae, Jun and Levental (2010), for given $\gamma > 0$ and $\epsilon > 0$, there exists an $\eta > 0$ for which

$$\limsup_{n\to\infty} \mathrm{P}\left(\sup_{\tilde{G}\in\mathcal{G}_r\,:\,d_{\mu_n}^{(2)}(\tilde{G},G)\leq\eta} |\tilde{S}_n(\tilde{G})| > 5\gamma\right) \leq 3\epsilon. \tag{S6.83}$$

Note by the arguments in the proof of Lemma S6.7, $\mathrm{P}(\hat{G}_n \in \mathcal{G}_r) \to 1$, and

$$d_{\mu_n}^{(2)}(\hat{G}_n, G) \leq \left\{\sup_{s\in\mathcal{T}}\left|\hat{G}_n(s) - G(s)\right|\right\}^{1/2} \leq \left\{\sup_{s\in\mathcal{T}}\left|\frac{\hat{G}_n(s)}{G(s)} - 1\right|\right\}^{1/2} \to 0$$

with probability tending to one. Hence, $\hat{G}_n \in \{\tilde{G}\in\mathcal{G}_r : d_{\mu_n}^{(2)}(\tilde{G},G)\leq\eta\}$ with probability tending to one, and so

$$\limsup_{n\to\infty}\mathrm{P}\left(|\tilde{S}_n(\hat{G}_n)| > 5\gamma\right) \leq \limsup_{n\to\infty} \mathrm{P}\left(\sup_{\tilde{G}\in\mathcal{G}_r\,:\,d_{\mu_n}^{(2)}(\tilde{G},G)\leq\eta} |\tilde{S}_n(\tilde{G})| > 5\gamma\right) \leq 3\epsilon.$$

As $\epsilon > 0$ was arbitrary, $\limsup_{n\to\infty}\mathrm{P}\left(|\tilde{S}_n(\hat{G}_n)| > 5\gamma\right) = 0$ and, as $\gamma > 0$ was arbitrary, this shows that $\tilde{S}_n(\hat{G}_n) = o_p(1)$. The argument following (S6.82) then shows that (II)$1_{\mathcal{L}_n} = o_p(1)$. □

To prove the next lemma, we need to develop a decomposition involving three types of martingale differences. The filtrations for these martingale differences are

$$\begin{aligned}
\mathcal{O}_{nj}^{\dagger} &\equiv \sigma\big(O_1,\ldots,O_j,U_{j+1,k_j}\big);\\
\mathcal{O}_{nj}^{*} &\equiv \sigma\big(O_1,\ldots,O_{j-1},k_j,m_j,E\big[\tilde{Y}\big|U_{j+1,k_j}\big],\mathrm{Var}_{\mathbb{Q}_j}(U_{k_j}),U_{j,k_{j-1}}\big);\\
\tilde{\mathcal{O}}_{nj} &\equiv \sigma(O_1,\ldots,O_j,\boldsymbol{U}_{j+1}),
\end{aligned} \tag{S6.84}$$

$j = q_n,\ldots,n-1$. Define

$$\begin{aligned}
h_{nj}(u) \equiv &\left[\frac{(u-\mathbb{Q}_j[U_{k_j}])}{\mathrm{Var}_{\mathbb{Q}_j}(U_{k_j})} - \frac{(u-Q_u[U_{k_j}])}{\mathrm{Var}_{Q_u}(U_{k_j})}\right]1\Big(\max_k\big|(\mathbb{Q}_j - Q_u)[U_k]\big| \lesssim K_{nj},\\
&\max_k\big|\mathrm{Var}_{\mathbb{Q}_j}^{-1}(U_k) - \mathrm{Var}_{Q_u}^{-1}(U_k)\big| \lesssim K_{nj}\Big),
\end{aligned} \tag{S6.85}$$

where in $\lesssim$ the implicit constants are independent of (j,n). The martingale differences to be used in the proof are then defined by

$$H_{nj}^{\dagger} \equiv \frac{1}{\sqrt{n-q_n}}\frac{m_{j-1}}{\sigma_{n,j-1}}h_{n,j-1}(U_{j,k_{j-1}})\big\{\tilde{Y}_j - E\big[\tilde{Y}\big|U_{j,k_{j-1}}\big]\big\}, \tag{S6.86}$$

$$H^*_{nj} \equiv \frac{1}{\sqrt{n-q_n}} \frac{m_{j-1}}{\sigma_{n,\,j-1}} h_{n,\,j-1}(U_{j,k_{j-1}}) E\big[\tilde{Y}\big|U_{j,k_{j-1}}\big], \text{ and}$$

$$\tilde{H}_{nj} \equiv \frac{1}{\sqrt{n-q_n}} \frac{m_{j-1}}{\sigma_{n,\,j-1}} h_{n,\,j-1}(U_{j,k_{j-1}}) \int_{\mathcal{T}} E_0(U_{j,k_{j-1}}, s, k_{j-1}) dM_j(s).$$

Note that $(H^\dagger_{nj}, \mathcal{O}^\dagger_{nj})$, $(H^*_{nj}, \mathcal{O}^*_{nj})$, $(\tilde{H}_{nj}, \tilde{\mathcal{O}}_{nj})$, $j = q_n + 1, \ldots, n$, are martingale difference sequences. In particular,

$$\text{(S6.87)} \quad \sqrt{n-q_n}\, E\big[H^\dagger_{n,\,j+1}\big|\mathcal{O}^\dagger_{nj}\big] = E\Big[\frac{m_j}{\sigma_{n,j}} h_{n,j}(U_{j+1,k_j})\Big[\tilde{Y}_{j+1} - E\big[\tilde{Y}\big|U_{j+1,k_j}\big]\Big]\Big|\mathcal{O}^\dagger_{nj}\Big]$$

$$= \frac{m_j}{\sigma_{nj}} h_{n,j}(U_{j+1,k_j})\Big[E\big[\tilde{Y}_{j+1}\big|U_{j+1,k_j}\big] - E\big[\tilde{Y}\big|U_{j+1,k_j}\big]\Big] = 0,$$

and

$$\text{(S6.88)} \quad \sqrt{n-q_n} E\big[H^*_{n,\,j+1}\big|\mathcal{O}^*_{nj}\big] = E\Big[\frac{m_j}{\sigma_{n,j}} h_{n,j}(U_{j+1,k_j}) E\big[\tilde{Y}\big|U_{j+1,k_j}\big]\Big|\mathcal{O}^*_{nj}\Big]$$

$$= \frac{m_j E\big[\tilde{Y}\big|U_{j+1,k_j}\big]}{\sigma_{nj}} \Big\{ \frac{E[(Q_u - \mathbb{Q}_j)[U_{k_j}]|\mathcal{O}^*_{nj}]}{\mathrm{Var}_{\mathbb{Q}_j}(U_{k_j})}$$

$$+ E\big[U_{j+1,k_j} - Q_u[U_{k_j}]\big|\mathcal{O}^*_{nj}\big]\Big[\frac{1}{\mathrm{Var}_{\mathbb{Q}_j}(U_{k_j})} - \frac{1}{\mathrm{Var}_{Q_u}(U_{k_j})}\Big]\Big\}$$

$$= 0,$$

where the first step of (S6.88) holds by

$$\text{(S6.89)} \quad \frac{(U_{j+1,k_j} - \mathbb{Q}_j[U_{k_j}])}{\mathrm{Var}_{\mathbb{Q}_j}(U_{k_j})} - \frac{(U_{j+1,k_j} - Q_u[U_{k_j}])}{\mathrm{Var}_{Q_u}(U_{k_j})}$$

$$= \frac{(Q_u - \mathbb{Q}_j)[U_{k_j}]}{\mathrm{Var}_{\mathbb{Q}_j}(U_{k_j})} + (U_{j+1,k_j} - Q_u[U_{k_j}])\Big[\frac{1}{\mathrm{Var}_{\mathbb{Q}_j}(U_{k_j})} - \frac{1}{\mathrm{Var}_{Q_u}(U_{k_j})}\Big].$$

In addition,

(S6.90)

$$\sqrt{n-q_n} E\big[\tilde{H}_{n,\,j+1}\big|\tilde{\mathcal{O}}_{nj}\big] = \frac{m_j}{\sigma_{nj}} h_{nj}(U_{j+1,k_j}) \int_{\mathcal{T}} E_0(U_{j+1,k_j}, s, k_j) E\big[dM_{j+1}(s)\big|\tilde{\mathcal{O}}_{nj}\big]$$

$$= \frac{m_j}{\sigma_{nj}} h_{nj}(U_{j+1,k_j}) \int_{\mathcal{T}} E_0(U_{j+1,k_j}, s, k_j) E[1(T_{j+1} \geq s)|\boldsymbol{U}_{j+1}]$$

$$\times E\,[1(C_{j+1} \in ds) - 1(C_{j+1} \geq s) d\Lambda(s)]$$

$$= 0,$$

where the first step holds by the independent censoring assumption, and the second step follows from the definition $d\Lambda(s) = \mathrm{P}(C \in ds)/\mathrm{P}(C \geq s)$.

LEMMA S6.21. *Under the conditions of Theorem 4.1, (III)*$1_{\mathcal{L}_n}$ *is asymptotically negligible.*

PROOF. Note that $dM_{j+1}(s) = 1(X_{j+1} \in ds, \delta_{j+1} = 0) - 1(X_{j+1} \geq s) d\Lambda(s)$; we re-express (III) (from (16) in the main text) as

$$\frac{1}{\sqrt{n-q_n}} \sum_{j=q_n}^{n-1} m_j \Big[\frac{1}{\hat{\sigma}_{nj}} - \frac{1}{\sigma_{nj}} + \frac{1}{\sigma_{nj}}\Big]\Big[\frac{(U_{j+1,k_j} - \mathbb{Q}_j[U_{k_j}])}{\mathrm{Var}_{\mathbb{Q}_j}(U_{k_j})} - \frac{(U_{j+1,k_j} - Q_u[U_{k_j}])}{\mathrm{Var}_{Q_u}(U_{k_j})}\Big]$$

$$\times \Big[\tilde{Y}_{j+1} - \int_{\mathcal{T}} E_0(U_{j+1,k_j}, s, k_j) dM_{j+1}(s)\Big].$$

Along with the fact that U_k is uniformly bounded in (A.1), that σ_{nj} is uniformly bounded away from zero in (A.6), (S6.8.2) and (S6.8.4) of Lemma S6.8, and (S6.11.2) of Lemma S6.11, the above display further gives that

$$\begin{aligned}
&\Big|\frac{1}{\sqrt{n-q_n}}\sum_{j=q_n}^{n-1} m_j \Big[\frac{1}{\hat{\sigma}_{nj}} - \frac{1}{\sigma_{nj}}\Big]\Big[\frac{(U_{j+1,k_j} - \mathbb{Q}_j[U_{k_j}])}{\mathrm{Var}_{\mathbb{Q}_j}(U_{k_j})} - \frac{(U_{j+1,k_j} - Q_u[U_{k_j}])}{\mathrm{Var}_{Q_u}(U_{k_j})}\Big] \\
&\qquad\qquad \times \Big[\tilde{Y}_{j+1} - \int_{\mathcal{T}} E_0(U_{j+1,k_j}, s, k_j) dM_{j+1}(s)\Big] 1_{\mathcal{L}_n}\Big| \\
&\lesssim \Big[\frac{\tau}{G(\tau)} + (1+\Lambda(\tau)) \sup_{(k,s,u)} \big|E_0(u,s,k)\big|\Big] \frac{\log(n \vee p_n)}{\sqrt{n-q_n}} \sum_{j=q_n}^{n-1} \frac{1}{j} \\
&\le \Big[\frac{\tau}{G(\tau)} + (1+\Lambda(\tau)) \sup_{(k,s,u)} \big|E_0(u,s,k)\big|\Big] \frac{\log(n \vee p_n)}{\sqrt{n-q_n}} \log\Big(\frac{n}{q_n}\Big) \to 0,
\end{aligned}$$

where the convergence to zero follows by $G(\tau) > 0$ in (A.2), that $\sup_{(k,s,u)} \big|E_0(u,s,k)\big|$ is bounded in (A.7), and the conditions: $n/q_n = O(1)$ and $q_n^{1/4}/\log(n \vee p_n) \to \infty$.

Combining all the above results, we have that

$$|(\mathrm{III})1_{\mathcal{L}_n}| \le \Big|\frac{1}{\sqrt{n-q_n}}\sum_{j=q_n}^{n-1} \frac{m_j}{\sigma_{nj}} h_{nj}(U_{j+1,k_j}) \Big[\tilde{Y}_{j+1} - \int_{\mathcal{T}} E_0(U_{j+1,k_j}, s, k_j) dM_{j+1}(s)\Big]\Big|.$$

Therefore, we have that

$$|(\mathrm{III})1_{\mathcal{L}_n}| \le \Big|\sum_{j=q_n}^{n-1} H^{\dagger}_{n,j+1}\Big| + \Big|\sum_{j=q_n}^{n-1} H^{*}_{n,j+1}\Big| + \Big|\sum_{j=q_n}^{n-1} \tilde{H}_{n,j+1}\Big|, \tag{S6.91}$$

where $H^{\dagger}_{nj}$, H^{*}_{nj} and $\tilde{H}_{nj}$ are defined in (S6.86). To complete the proof, it suffices to show that the three terms on the right-hand-side of (S6.91) are $o_p(1)$.

First note that for all n,

$$\max_{j \in \{q_n, \dots, n-1\}} |h_{nj}(U_{j+1,k_j})| \lesssim \frac{\sqrt{\log(n \vee p_n)}}{\sqrt{q_n}} \quad \text{a.s.} \tag{S6.92}$$

from the decomposition in (S6.89), the definition of h_{nj} in (S6.85), and (A.1); here the implicit constant in $\lesssim$ does not depend on n. Also, note that $\tilde{Y}_{j+1} - E[\tilde{Y}|U_{j+1,k_j}]$ is bounded almost surely using (A.1) and (A.2). Then, with σ_{nj} uniformly bounded away from zero in (A.6), (S6.92) further implies that

$$\max_j |H^{\dagger}_{nj}| \lesssim \sqrt{\frac{\log(n \vee p_n)}{q_n(n-q_n)}} \equiv B_n,$$

where in $\lesssim$ the implicit constant is also independent of n. Then for $\varepsilon > 0$,

$$\begin{aligned}
&\mathrm{P}\Big(\Big|\sum_{j=q_n}^{n-1} H^{\dagger}_{n,j+1}\Big| \ge \varepsilon\Big) \le \varepsilon^{-2} E\Big[\Big(\sum_{j=q_n}^{n-1} H^{\dagger}_{n,j+1}\Big)^2\Big] \\
&= \varepsilon^{-2}\Big(\sum_{j=q_n}^{n-1} E\big[(H^{\dagger}_{n,j+1})^2\big] + 2\sum_{q_n \le i < j \le n-1} E\Big[H^{\dagger}_{n,i+1} E\Big[H^{\dagger}_{n,j+1}\Big|\mathcal{O}_{nj}\Big]\Big]\Big)
\end{aligned}$$

$$= \varepsilon^{-2} \sum_{j=q_n}^{n-1} E\big[\big(H^{\dagger}_{n,j+1}\big)^2\big] \lesssim \varepsilon^{-2}(n-q_n)B_n^2 \to 0.$$

This shows the first term on the right-hand-side of (S6.91) converges to zero in probability. The second and the last terms on the right-hand-side of (S6.91) can be handled, using similar arguments. □

LEMMA S6.22. *Under the conditions of Theorem 4.1, (V)$1_{\mathcal{L}_n}$ is asymptotically negligible.*

PROOF. It is trivial to see that (V)$1_{\mathcal{L}_n} = o_p(1)$ under the null. To verify it under the alternative, we first have

$$|(\mathrm{V})1_{\mathcal{L}_n}| = \left| \frac{1}{\sqrt{n-q_n}} \sum_{j=q_n}^{n-1} \frac{m_j}{\hat\sigma_{nj}} 1_{\mathcal{L}_n} [\Psi_{k_j}(P) - \Psi(P)] \right|$$

$$\le \max_{j\in\{q_n,\dots,n-1\}} \left| \frac{\sigma_{nj}}{\hat\sigma_{nj}} 1_{\mathcal{L}_n} \right| \frac{1}{\sqrt{n-q_n}} \sum_{j=q_n}^{n-1} \frac{1}{\sigma_{nj}} |\Psi_{k_j}(P) - \Psi(P)| 1_{\mathcal{L}_n}.$$

Because σ_{nj} is assumed to be bounded away from zero in (A.6) and $\max_j |\sigma_{nj} 1_{\mathcal{L}_n}/\hat\sigma_{nj}|$ is bounded above using (S6.11.2) of Lemma S6.11, it suffices to show that

$$\frac{1}{\sqrt{n-q_n}} \sum_{j=q_n}^{n-1} |\Psi_{k_j}(P) - \Psi(P)| 1_{\mathcal{L}_n} = o_p(1).$$

Recall that

$$k_j = \arg\max_{k\in\mathcal{K}_n} \left| \frac{\mathrm{Cov}_{\mathbb{P}_j}(U_k, \delta X/\hat G_j(X))}{\mathrm{Var}_{\mathbb{P}_j}(U_k)} \right| \text{ and } m_j = \mathrm{sgn}\left[\frac{\mathrm{Cov}_{\mathbb{P}_j}(U_k, \delta X/\hat G_j(X))}{\mathrm{Var}_{\mathbb{P}_j}(U_k)} \right],$$

and let

$$k_0 = \arg\max_{k\in\mathcal{K}_n} \left| \frac{\mathrm{Cov}(U_k, T)}{\mathrm{Var}(U_k)} \right| \text{ and } m_0 = \mathrm{sgn}\left[\frac{\mathrm{Cov}(U_{k_0}, T)}{\mathrm{Var}(U_{k_0})} \right].$$

Because $m\mathrm{Cov}_{\mathbb{P}_j}(U_k, \delta X/\hat G_j(X))/\mathrm{Var}_{\mathbb{P}_j}(U_k)$ is maximized at (k_j, m_j), we observe that

$$0 \ge m_0 \frac{\mathrm{Cov}_{\mathbb{P}_j}(U_{k_0}, \delta X/\hat G_j(X))}{\mathrm{Var}_{\mathbb{P}_j}(U_{k_0})} - m_j \frac{\mathrm{Cov}_{\mathbb{P}_j}(U_{k_j}, \delta X/\hat G_j(X))}{\mathrm{Var}_{\mathbb{P}_j}(U_{k_j})} \tag{S6.93}$$

$$= \left[m_0 \frac{\mathrm{Cov}(U_{k_0}, T)}{\mathrm{Var}(U_{k_0})} - m_j \frac{\mathrm{Cov}(U_{k_j}, T)}{\mathrm{Var}(U_{k_j})} \right] + m_0 \left[\frac{\mathrm{Cov}_{\mathbb{P}_j}(U_{k_0}, \delta X/\hat G_j(X))}{\mathrm{Var}_{\mathbb{P}_j}(U_{k_0})} \right.$$

$$\left. - \frac{\mathrm{Cov}(U_{k_0}, T)}{\mathrm{Var}(U_{k_0})} \right] - m_j \left[\frac{\mathrm{Cov}_{\mathbb{P}_j}(U_{k_j}, \delta X/\hat G_j(X))}{\mathrm{Var}_{\mathbb{P}_j}(U_{k_j})} - \frac{\mathrm{Cov}(U_{k_j}, T)}{\mathrm{Var}(U_{k_j})} \right]$$

$$\ge \Psi(P) - \Psi_{k_j}(P) - 2\max_{k\in\mathcal{K}_n} |\Psi_{\hat G_j, k}(\mathbb{P}_j) - \Psi_{G,k}(P)|,$$

where $\Psi_k(P) = \Psi_{G,k}(P) \equiv \mathrm{Cov}(U_k, \delta X/G(X))/\mathrm{Var}(U_k)$, according to $\mathrm{Cov}(U_k, T) = \mathrm{Cov}(U_k, \delta X/G(X))$. Moreover,

(S6.94)

$$|\Psi_{\hat G_j, k}(\mathbb{P}_j) - \Psi_{G,k}(P)| 1_{\mathcal{L}_n} = \left| \frac{\mathrm{Cov}_{\mathbb{P}_j}(U_k, \delta X/\hat G_j(X))}{\mathrm{Var}_{\mathbb{P}_j}(U_k)} - \frac{\mathrm{Cov}(U_k, \delta X/G(X))}{\mathrm{Var}(U_k)} \right| 1_{\mathcal{L}_n}$$

$$
\begin{aligned}
&= \frac{1}{\mathrm{Var}_{\mathbb{P}_j}(U_k)} 1_{\mathcal{L}_n} \Big| \mathrm{Cov}_{\mathbb{P}_j}(U_k, \delta X/\hat{G}_j(X)) - \mathrm{Cov}_{\mathbb{P}_j}(U_k, \delta X/G(X)) \\
&\qquad\qquad\qquad\qquad + \mathrm{Cov}_{\mathbb{P}_j}(U_k, \delta X/G(X)) - \mathrm{Cov}(U_k, \delta X/G(X)) \Big| \\
&\quad + 1_{\mathcal{L}_n} \Big| \mathrm{Cov}(U_k, \delta X/G(X)) \Big| \left| \frac{1}{\mathrm{Var}_{\mathbb{P}_j}(U_k)} - \frac{1}{\mathrm{Var}(U_k)} \right| \\
&\le 1_{\mathcal{L}_n} \frac{|\mathrm{Cov}_{\mathbb{P}_j}(U_k, \delta X/\hat{G}_j(X) - \delta X/G(X))|}{\mathrm{Var}_{\mathbb{P}_j}(U_k)} \\
&\quad + 1_{\mathcal{L}_n} \frac{|\mathrm{Cov}_{\mathbb{P}_j}(U_k, \delta X/G(X)) - \mathrm{Cov}(U_k, \delta X/G(X))|}{\mathrm{Var}_{\mathbb{P}_j}(U_k)} \\
&\quad + 1_{\mathcal{L}_n} |\mathrm{Cov}(U_k, \delta X/G(X))| \left| \frac{1}{\mathrm{Var}_{\mathbb{P}_j}(U_k)} - \frac{1}{\mathrm{Var}(U_k)} \right|,
\end{aligned}
$$

where the second equality holds by using the identity $a_n b_n - ab = (a_n - a)b_n + (b_n - b)a$, and the ensuing step follows by the triangle inequality.

Below we further tackle each term in the upper bound of $|\Psi_{\hat{G}_j,k}(\mathbb{P}_j) - \Psi_{G,k}(P)|1_{\mathcal{L}_n}$ from (S6.94). To address the first term,

$$
\begin{aligned}
\text{(S6.95)}\quad & 1_{\mathcal{L}_n} \frac{|\mathrm{Cov}_{\mathbb{P}_j}(U_k, \delta X/\hat{G}_j(X) - \delta X/G(X))|}{\mathrm{Var}_{\mathbb{P}_j}(U_k)} \\
&\le 1_{\mathcal{L}_n} \frac{1}{\mathrm{Var}_{\mathbb{P}_j}(U_k)} \left[\mathbb{P}_j \left| U_k \tilde{Y} \sum_{r=1}^{\infty} \Big(\sup_t \Big| \frac{\hat{G}_j(t)}{G(t)} - 1 \Big| \Big)^r \right| \right. \\
&\qquad\qquad\qquad\qquad \left. + \mathbb{P}_j |U_k| \mathbb{P}_j \left| \tilde{Y} \sum_{r=1}^{\infty} \Big(\sup_t \Big| \frac{\hat{G}_j(t)}{G(t)} - 1 \Big| \Big)^r \right| \right] \\
&\le \frac{1}{\mathrm{Var}_{\mathbb{P}_j}(U_k)} \left[\mathbb{P}_j \left| U_k \tilde{Y} \sum_{r=1}^{\infty} \left(\sqrt{\frac{\log n}{j}} \right)^r \right| + \mathbb{P}_j |U_k| \mathbb{P}_j \left| \tilde{Y} \sum_{r=1}^{\infty} \left(\sqrt{\frac{\log n}{j}} \right)^r \right| \right] \\
&\le \frac{1}{\mathrm{Var}_{\mathbb{P}_j}(U_k)} \left[\sum_{r=1}^{\infty} \left(\sqrt{\frac{\log n}{q_n}} \right)^r \right] \left[\mathbb{P}_j |U_k \tilde{Y}| + \mathbb{P}_j |U_k| \mathbb{P}_j |\tilde{Y}| \right] \\
&\lesssim \left(\frac{\sqrt{\log n}}{\sqrt{q_n} - \sqrt{\log n}} \right) \max_{j,k} \left[\mathbb{P}_j \left| U_k \frac{\delta X}{G(X)} \right| + \mathbb{P}_j |U_k| \mathbb{P}_j \left| \frac{\delta X}{G(X)} \right| \right] \to 0 \text{ a.s.},
\end{aligned}
$$

where the first inequality holds by the triangle inequality and Taylor expansion with respect to $\hat{G}_j$ around G; the second inequality results from $\sup_t \big|\hat{G}_j(t)/G(t) - 1\big| \le \sqrt{\log n/j}$ given by the occurrence of $\mathcal{L}_n$ and Lemma S6.12; the third inequality holds by $j \ge q_n$; the last inequality follows from the fact that $\mathrm{Var}_{\mathbb{P}_j}(U_k) = \mathrm{Var}_{\mathbb{Q}_j}(U_k)$ that we have showed bounded away from zero in (S6.8.2) of Lemma S6.8, and the final convergence to zero results from the uniform boundedness of $\mathbb{P}_j|U_k \delta X/G(X)|$, $\mathbb{P}_j|U_k|$ and $\mathbb{P}_j|\delta X/G(X)|$ almost surely, which is implied by (A.1), $X \le \tau$, $G(\tau) > 0$ in (A.2) and the condition $q_n^{1/4}/\log(n \vee p_n) \to \infty$. This gives the first term on the right-hand-side of (S6.94) is $o_p(1)$ uniformly in (j,k).

To tackle the second term on the right-hand-side of (S6.94), again using the identity $a_n b_n - ab = (a_n - a)b_n + (b_n - b)a$ and the triangle inequality gives that for some positive finite

(j,n)-independent constant ζ_1',

(S6.96)

$$
\begin{aligned}
&1_{\mathcal{L}_n}\frac{|\mathrm{Cov}_{\mathbb{P}_j}(U_k,\delta X/G(X))-\mathrm{Cov}(U_k,\delta X/G(X))|}{\mathrm{Var}_{\mathbb{P}_j}(U_k)}\\
&\le \frac{1_{\mathcal{L}_n}}{\min_{(j,k)}\mathrm{Var}_{\mathbb{P}_j}(U_k)}\max_k\Big\{\big|(\mathbb{P}_j-P)[U_k\delta X/G(X)]\big|+\big|(\mathbb{Q}_j-Q_u)[U_k]P[\delta X/G(X)]\big|\\
&\qquad\qquad +\big|(\mathbb{P}_j-P)[\delta X/G(X)]\mathbb{Q}_j[U_k]\big|\Big\}\\
&\le \frac{\zeta}{\tilde\varepsilon}\Big[1+\big|P[\delta X/G(X)]\big|+\max_{(j,k)}\big\{\big|\mathbb{Q}_j[U_k]\big|\big\}\Big]K_{nj}\equiv \zeta_1'K_{nj},
\end{aligned}
$$

where the second inequality holds given the occurrence of $\mathcal{A}_n\supset\mathcal{L}_n$ for a sufficiently large constant ζ that does not depend on (j,n), and that $\min_{(j,k)}\mathrm{Var}_{\mathbb{P}_j}(U_k)=\min_{(j,k)}\mathrm{Var}_{\mathbb{Q}_j}(U_k)$ is bounded away from zero by (S6.8.2) of Lemma S6.8, so that $\min_{(j,k)}\mathrm{Var}_{\mathbb{P}_j}(U_k)>\tilde\varepsilon$ for some sufficiently small positive constant $\tilde\varepsilon$ that is independent of (j,n). Similarly for some positive (j,n)-independent constant ζ_2', the third term on the right-hand-side of (S6.94) is

$$
\begin{aligned}
\text{(S6.97)}\qquad &1_{\mathcal{L}_n}\big|\mathrm{Cov}(U_k,\delta X/G(X))\big|\left|\frac{1}{\mathrm{Var}_{\mathbb{P}_j}(U_k)}-\frac{1}{\mathrm{Var}(U_k)}\right|\\
&=1_{\mathcal{L}_n}\big|\mathrm{Cov}(U_k,\delta X/G(X))\big|\frac{|\mathrm{Var}_{\mathbb{Q}_j}(U_k)-\mathrm{Var}_{Q_u}(U_k)|}{\mathrm{Var}_{Q_u}(U_k)\mathrm{Var}_{\mathbb{Q}_j}(U_k)}\\
&\le 1_{\mathcal{L}_n}\max_k\big|\mathrm{Cov}(U_k,\delta X/G(X))\big|\frac{\max_k|\mathrm{Var}_{\mathbb{Q}_j}(U_k)-\mathrm{Var}_{Q_u}(U_k)|}{\min_k\mathrm{Var}_{Q_u}(U_k)\min_{(j,k)}\mathrm{Var}_{\mathbb{Q}_j}(U_k)}\\
&\le \zeta_2'K_{nj},
\end{aligned}
$$

where the last step holds by the presence of $1_{\mathcal{L}_n}$ in which $\min_{(j,k)}\mathrm{Var}_{\mathbb{Q}_j}(U_k)$ is bounded away from zero by (S6.8.2) of Lemma S6.8, along with that $\max_k\big|\mathrm{Cov}(U_k,\delta X/G(X))\big|$ is assumed to be bounded using (A.1) and (A.2), and $\min_k\mathrm{Var}_{Q_u}(U_k)$ is assumed to be bounded away from zero in (A.6).

Let $\zeta'=\zeta_1'+\zeta_2'$. Collecting the above results in (S6.94)–(S6.97) gives that $|\Psi_{\hat G_j,k}(\mathbb{P}_j)-\Psi_{G,k}(P)|$ is bounded above by $\zeta'K_{nj}$ and then $\zeta'\sqrt{\log(n\vee p_n)/q_n}$ for $j\ge q_n$. Inserting this result back into (S6.93) leads to

$$|\Psi_{k_j}(P)-\Psi(P)|1_{\mathcal{L}_n}\le 2\zeta'K_{nj}\le 2\zeta'K_{nq_n}.$$

Together with $\Psi(P)\equiv\max_k|\Psi_k(P)|$ in (2), the above display implies that on the event $\mathcal{L}_n$,

$$\text{(S6.98)}\qquad 0\le\Psi(P)-|\Psi_{k_j}(P)|\le|\Psi_{k_j}(P)-\Psi(P)|<\epsilon K_{nq_n},$$

where $\epsilon>2\zeta'$ is chosen in connection with (A.8). Recall that k_0 is the label of the predictor that attains $\Psi(P)$ under the alternatives, so it is easy to see that $k_0\in\mathcal{K}_n^*$, where $\mathcal{K}_n^*$ contains the predictors that have stronger association with T than the other predictors in $\mathcal{K}_n$, as indicated in (A.8). Therefore (S6.98) implies that $k_j\in\mathcal{K}_n^*$, because $\Psi(P)-|\Psi_{k_j}(P)|$ is then under the threshold specified in (A.8). By (A.8), we conclude that

$$\frac{1}{\sqrt{n-q_n}}\sum_{j=q_n}^{n-1}|\Psi_{k_j}(P)-\Psi(P)|1_{\mathcal{L}_n}=O_p(\sqrt{n-q_n}\,\mathrm{Diam}(\mathcal{K}_n^*))=o_p(1).$$

□

S7. Supplementary results for simulation studies and real data application. In Tables S.1-S.2 we report the names of the identified features based on the various competing approaches, separately for the Subtype B and C datasets. For the stabilized one-step estimator, in Subtype B the most correlated binary feature is the interaction of $hxb2.677.K.1mer$ and $hxb2.460.sequon\,actual.1mer$; the most correlated count feature is the interaction of $sequons.total.gp120$ and $sequons.total.v5$. In other words, the presence of specific amino acid (coded by K) at position 677 and the presence of some enzymatic processes starting at position 460 are found to have a synergistic influence on IC_{50}. Similarly, the change in IC_{50} appears to be simultaneously affected by the total numbers of observed chemical reactions in the region of gp120 and of V5.

TABLE S.1

The most correlated features identified by the competing methods, according to data type. The interactions are coded as in Table S.2, with $\alpha \times \beta$ denoting the interaction between α and β.

	Method	Feature	
		Binary predictors	Count predictors
Subtype B	Bonferroni Cox	NA	$\gamma_5 \times \zeta$
	Bonferroni One-Step	$\alpha_4 \times \alpha_8$	$\gamma_3 \times \delta$
	Stabilized One-Step	$\alpha_5 \times \beta_2$	$\gamma_2 \times \gamma_4$
Subtype C	Bonferroni Cox	$\alpha_1 \times \alpha_3$	NA
	Bonferroni One-Step	$\alpha_2 \times \alpha_6$	NA
	Stabilized One-Step	$\alpha_7 \times \beta_1$	$\gamma_1 \times \epsilon$

TABLE S.2

Coding of feature names used in Table S.1 based on source code for Magaret et al. (2019).

$\alpha_1: hxb2.279.D.1mer$	$\alpha_2: hxb2.389.G.1mer$
$\alpha_3: hxb2.459.G.1mer$	$\alpha_4: hxb2.130.K.1mer$
$\alpha_5: hxb2.677.K.1mer$	$\alpha_6: hxb2.462.N.1mer$
$\alpha_7: hxb2.363.S.1mer$	$\alpha_8: hxb2.132.T.1mer$
$\beta_1: hxb2.142.sequon\,actual.1mer$	$\beta_2: hxb2.460.sequon\,actual.1mer$
$\gamma_1: sequons.total.env$	$\gamma_2: sequons.total.gp120$
$\gamma_3: sequons.total.loop.e$	$\gamma_4: sequons.total.v5$
$\gamma_5: sequons.total.vrc01$	
$\delta: cysteines.total.gp120$	$\epsilon: length.env$
$\zeta: length.loop.e$	

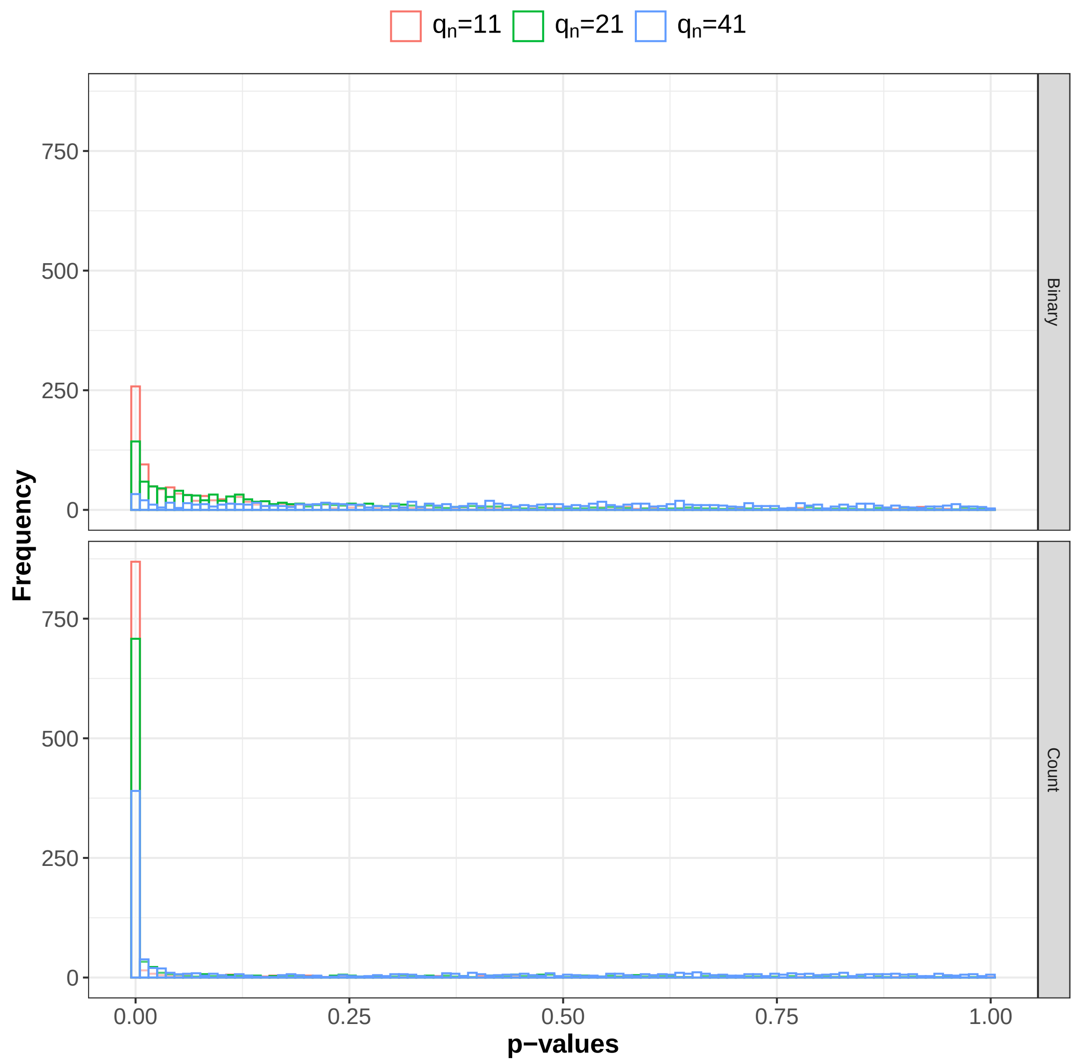

Fig S.1: Histogram of p-values obtained by applying the stabilized one-step estimator to 1000 random orderings of the Subtype B data for various values of q_n, separated according to binary and count predictors.

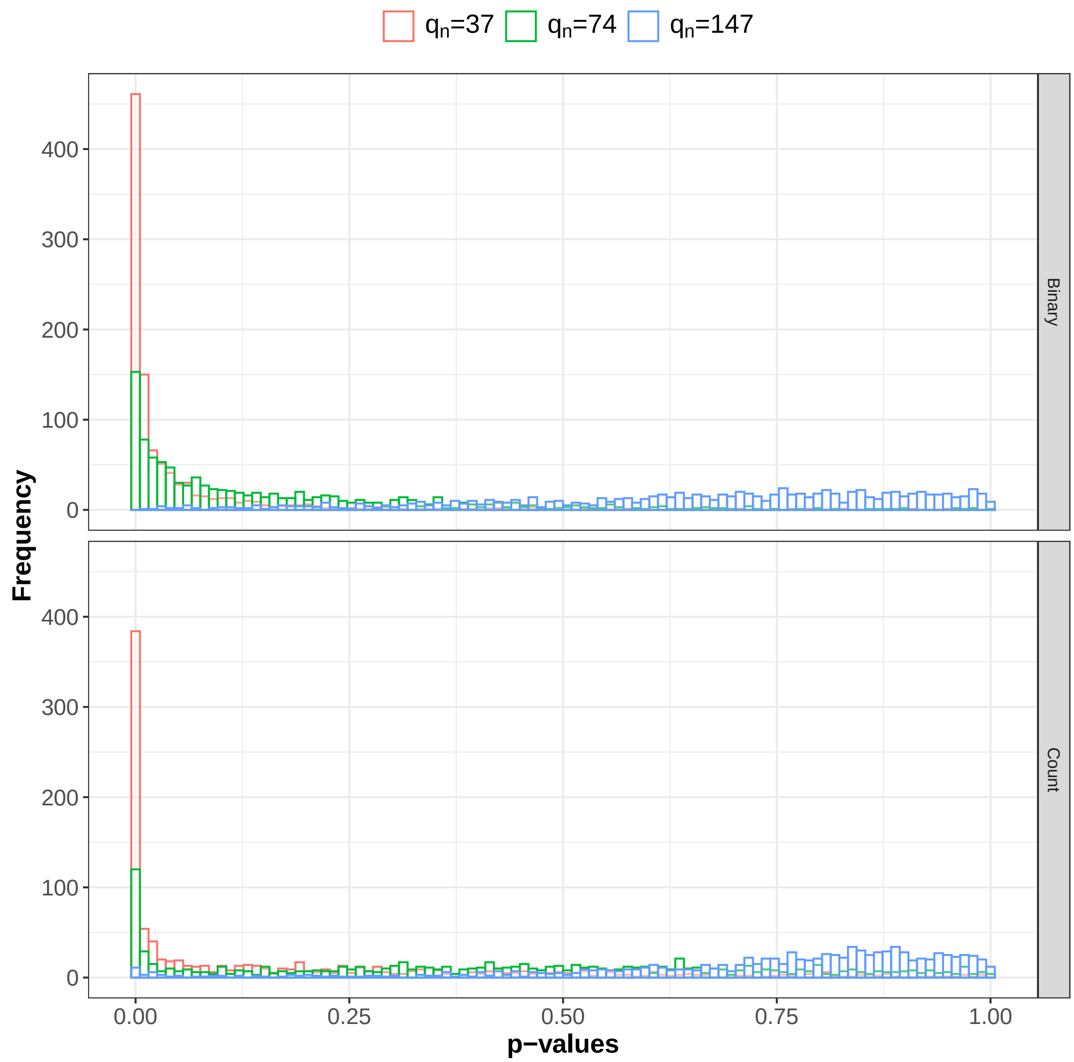

Fig S.2: As in Figure S.1, except for the data on virus subtype C.

REFERENCES

BAE, J., JUN, D. and LEVENTAL, S. (2010). The uniform CLT for martingale difference arrays under the uniformly integrable entropy. *Bulletin of the Korean Mathematical Society* **47** 39–51.

DEVLIN, S. J., GNANADESIKAN, R. and KETTENRING, J. R. (1975). Robust estimation and outlier detection with correlation coefficients. *Biometrika* **62** 531–545.

KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. *Springer Series in Statistics*. Springer.

KUNDU, S., MAJUMDAR, S. and MUKHERJEE, K. (2000). Central Limit Theorems revisited. *Statistics & Probability Letters* **47** 265–275.

LUEDTKE, A. R. and VAN DER LAAN, M. J. (2018). Parametric-rate inference for one-sided differentiable parameters. *Journal of American Statistical Association* **113** 780–788.

MAGARET, C. A., BENKESER, D. C., WILLIAMSON, B. D., BORATE, B. R., CARPP, L. N., GEORGIEV, I. S., SETLIFF, I., DINGENS, A. S., SIMON, N., CARONE, M., SIMPKINS, C., MONTEFIORI, D., ALTER, G., YU, W.-H., JURASKA, M., EDLEFSEN, P. T., KARUNA, S., MGODI, N. M., EDUGUPANTI, S. and GILBERT, P. B. (2019). Prediction of VRC01 neutralization sensitivity by HIV-1 gp160 sequence features. *PLoS Comput Biol* **15** e1006952.

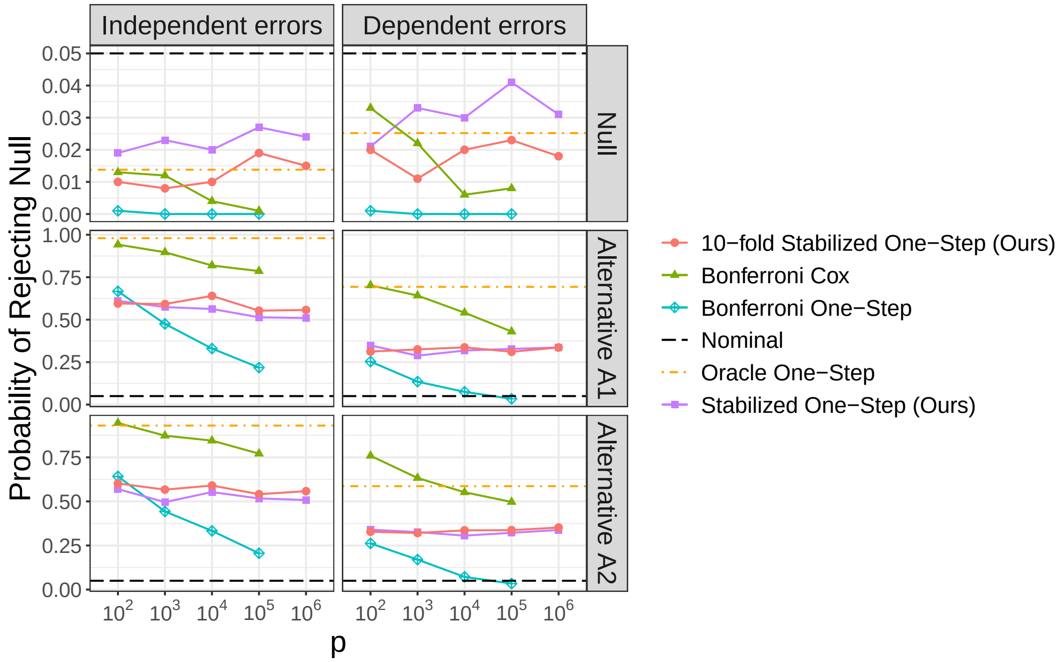

Fig S.3: As in Figure 1, except under heavy censoring (30%). The Bonferroni Cox method has good power, but it is computationally expensive when $p \geq 10^3$ and computationally infeasible when $p = 10^6$. The power of Bonferroni Cox decreases rapidly with increasing p, whereas our procedures maintain good power even with increasing p.

VAN DE GEER, S. (1995). Exponential inequalities for martingales with application to maximum likelihood estimation for counting processes. *The Annals of Statistics* **23** 1779–1801.

VAN DER LAAN, M. J. and ROBINS, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. Springer, New York, NY, USA.

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer, New York, NY, USA.

VAN DER VAART, A. W. and WELLNER, J. A. W. (2007). Empirical processes indexed by estimated functions. *IMS Lecture Notes–Monograph Series Asymptotics: Particles, Processes and Inverse Problems* **55** 234–252.

WELLNER, J. A. (2007). On an exponential bound for the Kaplan-Meier estimator. *Lifetime Data Analysis* **13** 481–496.

S8. R code.

README document

The R scripts below were used in "Efficient Estimation of the Maximal Association between Multiple Predictors and a Survival Outcome" by Tzu-Jung Huang, Alex Luedtke and Ian W. McKeague. The scripts provide p-values and confidence intervals based on the Stabilized One-Step Estimator ('SOSE') and competing methods (as mentioned in the Competing Methods section of the article). The simulation results can be obtained by running the main script (displayed in the chunk *example_sim_maxCorrSurv*) using auxiliary functions collected in the chunks *data_management*, *code_maxCorrSurv* and *sim_data_acquisition* (in order).

To deliver a quick demonstration on a standalone local computer, the number of predictors (p) is set as *1000*; the number of MC simulation runs (*num_r*) is set as *1* and the number of random data ordering (*num_rdo*) is set as *1* in the following example simulation. To obtain the simulation results, use *p = c(100, 1000, 1e4, 1e5, 1e6)*, *num_r = 1000*, and *num_rdo = 1* or *10*. The use of parallel computing is recommended.

In what follows, the three chunks *data_management*, *code_maxCorrSurv* and *sim_data_acquisition* are displayed; the code of conducting the example simulation is provided, and the confidence intervals and p-values by method and model are presented in a table. Interested readers can access and download R files at https://github.com/tzujunghuang/maxCorr_SurvivalOutcome.

Required functions

data_management

```
num_digits = 7

mostfrequent_percent = function(col){
  tbl = table(col)
  res = cbind(tbl, round(prop.table(tbl),2))
  colnames(res) = c('Count','Percentage')
  res[which.max(res[,2]), 2]
}

check_data_type = function(col){
  if (length(unique(col)) < 2) { data_type = 'void'
  } else if (length(unique(col)) == 2 & mostfrequent_percent(col) >= 0.95) {
    data_type = 'void binary'
  } else if (length(unique(col)) == 2 & mostfrequent_percent(col) < 0.95) {
    data_type = 'binary'
  } else if (length(unique(col)) > 2 & all((col == floor(col)) & (col >= 0))) {
    data_type = 'count/categorical'
  } else if (length(unique(col)) > 2 & all(col > floor(col))) {
    data_type = 'continuous'
  }
  return(data_type)
}

check_void = function(df, maineffect=TRUE){
  df_info = data.frame(names(df), apply(df, 2, function(x){check_data_type(x)}),
                       apply(df, 2, function(x){mostfrequent_percent(x)}))

  if (maineffect) {
    names(df_info) = c('GeneticVariable', 'DataType', 'LargestProb_of_Success')
    # 65.58%
  } else{
```

```
    names(df_info) = c('Interaction', 'DataType', 'LargestProb_of_Success')
    # 59.94%
  }

  df_info$Void = 1*grepl('void', as.vector(df_info$DataType))
  n_binary = dim(df_info[grepl('binary', df_info$DataType) |
                        grepl('void', df_info$DataType),])[1]
  n_void = dim(df_info[grepl('void', df_info$DataType),])[1]
  return( list(df_info = df_info,
               void_percent = round(100*n_void/n_binary, 2)) )
}

colRound <- function(mat, num_digits=7, block_size=1e5) {

  num_portions = ceiling(dim(mat)[2]/block_size) - 1

  if (num_portions == 0) {
      results = round(mat, num_digits)
  } else {
      results = do.call(cbind, lapply(0:num_portions, function(i) {
                  if (i < num_portions) {
                      sub_mat = mat[,(1 + i*block_size):((i + 1)*block_size)]
                  } else {
                      sub_mat = mat[,(1 + i*block_size):dim(mat)[2]]  }

                  round(sub_mat, num_digits) })) }
  return(results)
}

colSubtract <- function(mat, colMeans_mat, block_size=1e4) {

  num_portions = ceiling(dim(mat)[2]/block_size) - 1

  if (num_portions == 0) {
    n0 = dim(mat)[1];
    results = (mat - matrix(rep(colMeans_mat, n0), byrow = TRUE, nrow = n0))
  } else {
    results <- do.call(cbind, lapply(0:num_portions, function(i) {
      print(i)
      if (i < num_portions) {
        n0 = dim(mat)[1];
        sub_mat = mat[,(1 + i*block_size):((i + 1)*block_size)]
        sub_mean_mat = matrix(rep(colMeans_mat[(1 + i*block_size):
                                              ((i + 1)*block_size)], n0),
                              byrow = TRUE, nrow = n0)
        result = sub_mat - sub_mean_mat
      } else {
        n0 = dim(mat)[1];
        sub_mat = mat[,(1 + i*block_size):dim(mat)[2]]
        sub_mean_mat = matrix(rep(colMeans_mat[(1 + i*block_size):dim(mat)[2]],
                                  n0), byrow = TRUE, nrow = n0)
        result = sub_mat - sub_mean_mat }
      return( result ) })) }
```

```
  return( results )
}

colIF_ipw <- function(mat, colMeans_mat, colVars_mat, colCovs_mat, m,
                      central_Y_Ghat = Y_Ghat - mean_YGhat, block_size = 1e4) {

  num_portions = ceiling(dim(mat)[2]/block_size) - 1

  if (num_portions == 0) {
    n0 = dim(mat)[1]
    mu_selectU_0 = matrix(rep(colMeans_mat, n0), byrow = TRUE, nrow = n0)
    var_selectU_0 = matrix(rep(colVars_mat, n0), byrow = TRUE, nrow = n0)
    cov_selectU0_Y = matrix(rep(colCovs_mat, n0), byrow = TRUE, nrow = n0)

    a = (mat - mu_selectU_0)*central_Y_Ghat / var_selectU_0
    b = ((mat - mu_selectU_0)^2)*cov_selectU0_Y / (var_selectU_0)^2
    results = matrix(rep(m,n0), byrow = TRUE, nrow = n0)*(a - b)

  } else {
    results <- do.call(cbind, lapply(0:num_portions, function(i) {
      if (i < num_portions) {
        n0 = dim(mat)[1]
        sub_mat = mat[,(1 + i*block_size):((i + 1)*block_size)]
        sub_mu_selectU_0 = matrix(rep(colMeans_mat[(1 + i*block_size):
                                                   ((i + 1)*block_size)], n0),
                                  byrow = TRUE, nrow = n0)
        sub_var_selectU_0 = matrix(rep(colVars_mat[(1 + i*block_size):
                                                   ((i + 1)*block_size)], n0),
                                   byrow = TRUE, nrow = n0)
        sub_cov_selectU0_Y = matrix(rep(colCovs_mat[(1 + i*block_size):
                                                    ((i + 1)*block_size)], n0),
                                    byrow = TRUE, nrow = n0)

        a = (sub_mat - sub_mu_selectU_0)*central_Y_Ghat / sub_var_selectU_0
        b = (((sub_mat - sub_mu_selectU_0)^2)*sub_cov_selectU0_Y /
          (sub_var_selectU_0)^2)
        result = matrix(rep(m[(1 + i*block_size):((i + 1)*block_size)], n0),
                        byrow = TRUE, nrow = n0)*(a - b)

      } else {
        n0 = dim(mat)[1]
        sub_mat = mat[,(1 + i*block_size):dim(mat)[2]]
        sub_mu_selectU_0 = matrix(rep(colMeans_mat[(1 + i*block_size):
                                                   dim(mat)[2]], n0),
                                  byrow = TRUE, nrow = n0)
        sub_var_selectU_0 = matrix(rep(colVars_mat[(1 + i*block_size):
                                                   dim(mat)[2]], n0),
                                   byrow = TRUE, nrow = n0)
        sub_cov_selectU0_Y = matrix(rep(colCovs_mat[(1 + i*block_size):
                                                    dim(mat)[2]], n0),
                                    byrow = TRUE, nrow = n0)

        a = (sub_mat - sub_mu_selectU_0)*central_Y_Ghat / sub_var_selectU_0
```

```
        b = (((sub_mat - sub_mu_selectU_0)^2)*sub_cov_selectU0_Y /
          (sub_var_selectU_0)^2)
        result = matrix(rep(m[(1 + i*block_size):dim(mat)[2]], n0),
                        byrow = TRUE, nrow = n0)*(a - b) }

      return( result ) })) }

  return( results )
}

colIF_CAR <- function(mat, colMeans_mat, colVars_mat, mart_val, m,
                      block_size=1e4) {

  num_portions = ceiling(dim(mat)[2]/block_size) - 1

  if (num_portions == 0) {
    n0 = dim(mat)[1]
    mu_selectU_0 = matrix(rep(colMeans_mat, n0), byrow = TRUE, nrow = n0)
    var_selectU_0 = matrix(rep(colVars_mat, n0), byrow = TRUE, nrow = n0)
    results = ( matrix(rep(m,n0), byrow = TRUE, nrow = n0)
                * (mat - mu_selectU_0)*mart_val / var_selectU_0 )

  } else {
    results <- do.call(cbind, lapply(0:num_portions, function(i) {
      if (i < num_portions) {
        n0 = dim(mat)[1]
        sub_mat = mat[,(1 + i*block_size):((i + 1)*block_size)]
        sub_mu_selectU_0 = matrix(rep(colMeans_mat[(1 + i*block_size):
                                            ((i + 1)*block_size)], n0),
                                  byrow = TRUE, nrow = n0)
        sub_var_selectU_0 = matrix(rep(colVars_mat[(1 + i*block_size):
                                            ((i + 1)*block_size)], n0),
                                   byrow = TRUE, nrow = n0)
        result = ( matrix(rep(m[(1 + i*block_size):((i + 1)*block_size)], n0),
                          byrow = TRUE, nrow = n0)
                   * (sub_mat - sub_mu_selectU_0)*mart_val / sub_var_selectU_0 )

      } else {
        n0 = dim(mat)[1]
        sub_mat = mat[,(1 + i*block_size):dim(mat)[2]]
        sub_mu_selectU_0 = matrix(rep(colMeans_mat[(1 + i*block_size):
                                            dim(mat)[2]], n0),
                                  byrow = TRUE, nrow = n0)
        sub_var_selectU_0 = matrix(rep(colVars_mat[(1 + i*block_size):
                                            dim(mat)[2]], n0),
                                   byrow = TRUE, nrow = n0)
        result = ( matrix(rep(m[(1 + i*block_size):dim(mat)[2]], n0),
                          byrow = TRUE, nrow = n0) *
                   (sub_mat - sub_mu_selectU_0)*mart_val / sub_var_selectU_0 ) }
      return( result ) })) }

  return( results )
}
```

```r
my_colMeans <- function(mat, block_size=1e5) {

  num_portions = ceiling(dim(mat)[2]/block_size) - 1

  if (num_portions == 0) {
    colmean = t(colMeans(mat, na.rm = TRUE))
  } else {
    colmean = do.call(cbind, lapply(0:num_portions, function(i) {
      if (i < num_portions) {
        sub_mat = mat[,(1 + i*block_size):((i + 1)*block_size)]
      } else {
        sub_mat = mat[,(1 + i*block_size):dim(mat)[2]]  }

      t(colMeans(sub_mat, na.rm = TRUE)) }))  }
  return(colmean)
}

colStandardization <- function(mat, colSD, block_size=1e4) {

  num_portions = ceiling(dim(mat)[2]/block_size) - 1

  if (num_portions == 0) {
    n0 = dim(mat)[1]
    results = as.matrix((mat - matrix(rep(my_colMeans(mat), n0), nrow = n0,
                                      byrow = TRUE)) /
                        matrix(rep(colSD, n0), nrow = n0, byrow = TRUE))
  } else {
    results <- do.call(cbind, lapply(0:num_portions, function(i) {
                 if (i < num_portions) {
                   sub_mat = mat[,(1 + i*block_size):((i + 1)*block_size)]
                   sub_colSD = colSD[(1 + i*block_size):((i + 1)*block_size)]
                 } else {
                   sub_mat = mat[,(1 + i*block_size):dim(mat)[2]]
                   sub_colSD = colSD[(1 + i*block_size):dim(mat)[2]] }

                 nn = dim(sub_mat)[1]
                 result = (sub_mat - matrix(rep(my_colMeans(sub_mat), nn),
                                        nrow = nn, byrow = TRUE)) /
                   matrix(rep(sub_colSD, nn), nrow = nn, byrow = TRUE)
                 as.matrix(result) })) }
  return(results)
}

colVars <- function(mat, sd_use, block_size=1e4, na.rm=TRUE) {

  if (is.null(dim(mat))) {
    results = t(rep(0, length(mat)))
  } else {
    if (dim(mat)[1] <= 1) {
      results = t(rep(0, length(mat)))
    } else {
      num_portions = ceiling(dim(mat)[2]/block_size) - 1
```

```
      if (num_portions == 0) {
        nn = ifelse(na.rm, colSums(!is.na(mat)), nrow(mat))
        colVar = ((colMeans((mat)^2, na.rm = na.rm) -
                   (colMeans(mat, na.rm = na.rm))^2) * nn / (nn - 1))
        if (sd_use) { results = t(sqrt(colVar)) } else { results = t(colVar) }

      } else {
        results = do.call(cbind, lapply(0:num_portions, function(i) {
                      if (i < num_portions) {
                        sub_mat = mat[,(1 + i*block_size):((i + 1)*block_size)]
                      } else {
                        sub_mat = mat[,(1 + i*block_size):dim(mat)[2]]  }

                      nn = ifelse(na.rm, colSums(!is.na(sub_mat)),
                                  nrow(sub_mat))
                      colVar = (colMeans((sub_mat)^2, na.rm = na.rm) -
                            (colMeans(sub_mat, na.rm = na.rm))^2)*nn / (nn - 1)

                      if (sd_use) { result = sqrt(colVar) } else {
                        result = colVar }
                      t(result) })) }
      }
    }
  return(results)
}

colCovs <- function(mat, y, block_size=1e5, na.rm=TRUE) {

  num_portions = ceiling(dim(mat)[2]/block_size) - 1

  if (num_portions == 0) {
      colCov = t(cov(mat, y))
  } else {
      colCov = do.call(cbind, lapply(0:num_portions, function(i) {
                  if (i < num_portions) {
                      sub_mat = mat[,(1 + i*block_size):((i + 1)*block_size)]
                  } else {
                      sub_mat = mat[,(1 + i*block_size):dim(mat)[2]]  }

                  t(cov(sub_mat, y)) }) ) }
  return(colCov)
}
```

code_maxCorrSurv

```
num_digits = 7

NumericalStudy <- setRefClass( "NumericalStudy",
 fields = list(
    input_data = "list",
    n = "numeric",
    p = "numeric"
  ),
```

```
methods = list(
 initialize = function(input_data = dat){
   input_data <<- input_data
   n <<- dim(input_data$U)[1]
   p <<- dim(input_data$U)[2]
 },

 KM_weight_func = function(obs){
   # obs is a vector of observation/individual indices
   data_km = data.frame(X = input_data$X[obs], delta = input_data$delta[obs])
   data_km = data_km[order(data_km$X),]
   prod = c(1, cumprod( sapply(1:(n - 1), function(i){
     ((n - i)/(n - i + 1))^(data_km$delta[i]) }) ) )
   kmwts = sapply(1:n, function(i){ data_km$delta[i]*prod[i]/(n - i + 1) } )
   return( kmwts )
 },

 KM_SurF = function(t, obs){
   # t is a scalar; obs is a vector of observation/individual indices
   data_km = data.frame(X = input_data$X[obs], delta = input_data$delta[obs])
   km = survfit(Surv(X, 1-delta)~1, data = data_km)
   rm(data_km)

   survest = cbind(km$time, km$surv)
   if ( length(which(survest[, 1] <= t)) > 0 ) {
       return( survest[max(which(survest[, 1] <= t)), 2] )
   } else { return( 1 ) }
 },

 KM_SurF_self = function(obs, quar_trunc){
   # obs is a vector of observation/individual indices
   data_km = data.frame(X = input_data$X[obs], delta = input_data$delta[obs])
   tau = quantile(data_km$X, probs = quar_trunc)
   tau_surv = KM_SurF(tau, obs)

   km = survfit(Surv(X, 1-delta)~1, data = data_km)
   rm(data_km)

   km$surv[km$time > tau] = tau_surv
   survest = cbind(km$time, km$surv)
   return( survest )
 },

 Inverse_weight_func = function(x0, delta0, obs0, quar_trunc, err_msg){
   tau = quantile(x0, probs = quar_trunc)
   tau_surv = KM_SurF(tau, obs = obs0)

   if (length(x0) == 1) {
     inverse_weight = ifelse(x0 < tau, KM_SurF(x0, obs = obs0), tau_surv)
   } else {
     if (length(unique(x0)) == length(x0)) {
       KM_table = KM_SurF_self(obs = obs0, quar_trunc)
     } else {
```

```
        KM_table0 = data.frame(KM_SurF_self(obs = obs0, quar_trunc))
        names(KM_table0) = c("time", "surv_prob")
        n_occur = data.frame(table(x0)); names(n_occur) = c("time", "freq")
        n_freq = n_occur$freq; n1 = n_freq[n_freq > 1];
        n1[is.nan(n1) | is.na(n1) | is.null(n1)] = 0

        tryCatch({ temp = data.frame(cbind(
          rep(KM_table0[KM_table0$time %in% n_occur$time[n_freq > 1],]$time,
              n1),
          rep(KM_table0[KM_table0$time %in% n_occur$time[n_freq > 1],]$surv_prob,
              n1))) },
          error = function(err_msg) {
            message("Original error message:");
            message(paste(err_msg,"\n", sep = ''))
            save( list(x0 = x0, delta0 = delta0),
                  file = paste('errordata_', cr, '_', r, '.Rdata', sep = '') )
            stop(err_msg) })
        names(temp) = names(KM_table0)
        # times_once = as.numeric( as.character(n_occur$time[n_freq == 1]) )
        KM_table = as.matrix(rbind(temp,
            KM_table0[KM_table0$time %in% n_occur$time[n_freq == 1],])) }

      diff = setdiff(x0, KM_table[,1])
      if (length(diff) > 0) {
        for (item in diff) { KM_table = rbind(KM_table,
                  c(item, KM_table[max(which(item >= KM_table[,1])), 2])) } }
      KM_table[KM_table[,1] > tau, 2] = tau_surv
      inverse_weight = as.vector(KM_table[order(KM_table[,1]),][,2]) }

    if (length(inverse_weight) != length(x0)) {
      save( list(x0 = x0, delta0 = delta0),
            file = paste('errordata_', cr, '_', r, '.Rdata', sep = ''))
      stop('Dimension Problem!')
    } else { return( inverse_weight ) }
  },

  Est_Psi0_d = function(U_index, obs, quar_trunc){
    # U_index is a scalar or a vector of predictor indices;
    # obs is a vector of observation/individual indices
    data0 = cbind(input_data$X[obs], input_data$delta[obs],
                  input_data$U[obs, U_index])
    data0 = data0[order(data0[,1]),]

    x_0 = data0[,1]; delta_0 = data0[,2]
    selectU_0 = data0[,3:dim(data0)[2]]
    rm(data0)

    inverse_weight = Inverse_weight_func(x0 = x_0, delta0 = delta_0, obs0 = obs,
                                         quar_trunc,
                      err_msg = 'Error in Est_Psi0_d Inverse_weight KM_table')
    Y_Ghat = (x_0*delta_0) / inverse_weight;
    Y_Ghat[is.nan(Y_Ghat) | is.na(Y_Ghat)] = 0
    cov_Y_selectU = colCovs(selectU_0, Y_Ghat);
```

```r
    var_selectU = colVars(selectU_0, sd_use = FALSE)
    est = cov_Y_selectU / var_selectU
    return( est )
  },

  IF_star_self = function(m, U_index, obs, quar_trunc){
    # m is a scalar; U_index is a scalar or a vector of predictor indices;
    # obs is the vector of old/whole individuals indices
    data0 = cbind(input_data$X[obs], input_data$delta[obs],
                  input_data$U[obs, U_index])
    data0 = data0[order(data0[,1]),]
    X = data0[,1]; delta = data0[,2]; selectU_0 = data0[,3:dim(data0)[2]]
    rm(data0)
    tau = quantile(X, probs = quar_trunc)
    inverse_weight = Inverse_weight_func(x0 = X, delta0 = delta, obs0 = obs,
                                         quar_trunc,
                     err_msg = 'Error in IF_star_self Inverse_weight KM_table')

    if ( is.matrix(selectU_0) ) {
       mu_selectU_0 = matrix(rep(my_colMeans(selectU_0), length(obs)),
                             byrow = TRUE, nrow = length(obs))
       selectU_0_colVars = colVars(selectU_0, sd_use = FALSE)
    } else { mu_selectU_0 = mean(selectU_0)
             selectU_0_colVars = var(selectU_0) }

    var_selectU_0 = matrix(rep(selectU_0_colVars, length(obs)), byrow = TRUE,
                           nrow = length(obs))
    gc()

    Y_Ghat = (X*delta) / inverse_weight;
    Y_Ghat[is.nan(Y_Ghat) | is.na(Y_Ghat)] = 0
    mean_YGhat = mean(Y_Ghat);

    if ( is.matrix(selectU_0) ) { cov_mat = colCovs(selectU_0, Y_Ghat)
    } else { cov_mat = cov(selectU_0, Y_Ghat) }

    a = ((selectU_0 - mu_selectU_0)*(Y_Ghat - mean_YGhat)) / var_selectU_0
    b = ((selectU_0 - mu_selectU_0)^2)*matrix(rep(cov_mat, length(obs)),
                                              nrow = length(obs),
                                              byrow = TRUE) / (var_selectU_0)^2
    IF_ipw = matrix(rep(m,length(obs)), nrow = length(obs),
                    byrow = TRUE)*(a - b)
    gc(); rm(a); rm(b)

    time_comparison1 = outer(X, X, '=='); time_comparison2 = outer(X, X, '>=')
    event_nums = colSums(time_comparison1*(1 - delta));
    risk_set = colSums(time_comparison2)
    Y_Ghat_mat = ( time_comparison2 * (X*delta) / inverse_weight )
    Y_Ghat_mat[is.nan(Y_Ghat_mat) | is.na(Y_Ghat_mat)] = 0
    mean_YGhat_mat = colMeans(Y_Ghat_mat)

    ### Slow when obs is large
    if ( is.matrix(selectU_0) ) {
```

```
 Ehat_pointwise_val = sapply( 1:length(obs), function(i){
   a = mean_YGhat_mat[i]; n_s = length(obs) - (i - 1);
   p_s = length(U_index)
   if (i == length(obs)) {
       selectU = t(as.matrix(selectU_0[length(obs),]))
   } else{ selectU = selectU_0[i:length(obs),] }

   tryCatch({ mean_selectU = my_colMeans(selectU);
       var_selectU = colVars(selectU, sd_use = FALSE) },
       error = function(msg='Error in IF_star_self Ehat_pointwise_val') {
       message("Original error message:");
       message(paste(msg,"\n", sep = ''))
       save( input_data, file = paste('errordata_in_selectU_for', i, '_in',
                                      r, '.Rdata', sep = '') )
       stop(msg) })

   cov_Y_selectU = my_colMeans( (selectU - matrix(rep(mean_selectU, n_s),
                                              byrow = TRUE, ncol = p_s))
                                * matrix(rep(Y_Ghat_mat[i:length(obs),i],
                                             p_s), ncol = p_s) )
   b = cov_Y_selectU / var_selectU; b[is.nan(b) | is.na(b)] = 0
   c = selectU_0[i,] - mean_selectU
   rm(selectU); rm(mean_selectU); rm(var_selectU); rm(cov_Y_selectU)
   return( a + c*b ) } )
 Ehat_pointwise_val[is.nan(Ehat_pointwise_val) |
                    is.na(Ehat_pointwise_val)] = 0

} else {

 Ehat_pointwise_val = sapply( 1:length(obs), function(i){
    a = mean_YGhat_mat[i]; n_s = length(obs) - (i - 1);
    p_s = length(U_index)
    if (i == length(obs)) {
      selectU = selectU_0[length(obs)]
    } else{ selectU = selectU_0[i:length(obs)] }

    mean_selectU = mean(selectU);
    if (length(selectU) > 1) { var_selectU = var(selectU);
      cov_Y_selectU = mean( (selectU - matrix(rep(mean_selectU, n_s),
                                          byrow = TRUE, ncol = p_s))
                            * matrix(rep(Y_Ghat_mat[i:length(obs),i],
                                         p_s), ncol = p_s) )
      b = cov_Y_selectU / var_selectU
      b[is.nan(b) | is.na(b)] = 0
    } else { var_selectU = 'NA'; cov_Y_selectU = 0; b = 0 }

    c = selectU_0[i] - mean_selectU
    rm(selectU); rm(mean_selectU); rm(var_selectU); rm(cov_Y_selectU)
    return( a + c*b ) } )
 Ehat_pointwise_val[is.nan(Ehat_pointwise_val) |
                    is.na(Ehat_pointwise_val)] = 0

}
```

```
  hazard = event_nums/risk_set; hazard[is.nan(hazard) | is.na(hazard)] = 0
  mart_X = (X <= tau & delta == 0) - (X <= tau)*hazard
  val_1 = ( matrix(t(Ehat_pointwise_val), nrow = length(obs))*mart_X )
  IF_CAR = ( matrix(rep(m,length(obs)), nrow = length(obs), byrow = TRUE)
             * (selectU_0 - mu_selectU_0) * val_1 / var_selectU_0 )
  return( round((IF_ipw - IF_CAR), 3) )
},

IF_star = function(m, U_index, obs_all, obs0, obs1, quar_trunc){
  # m is a scalar; U_index is a scalar or a vector of predictor indices;
  # (obs_all, obs0, obs1) are the vectors of
  # whole, old and new individual indices
  data_all = cbind(input_data$X[obs_all], input_data$delta[obs_all])
  data_all = data_all[order(data_all[,1]),]
  x_all = data_all[,1]; delta_all = data_all[,2]
  rm(data_all)

  data0 = cbind(input_data$X[obs0], input_data$delta[obs0],
                input_data$U[obs0, U_index])
  data0 = data0[order(data0[,1]),]
  x_0 = data0[,1]; delta_0 = data0[,2]
  selectU_0 = matrix(data0[,3:dim(data0)[2]], nrow = length(obs0),
                     ncol = length(U_index))
  rm(data0)

  dup_vals_xall = x_all[duplicated(x_all)]; dup_vals_x0 = x_0[duplicated(x_0)]
  extra_dup_vals = setdiff(dup_vals_xall, dup_vals_x0)
  if (length(dup_vals_xall) > length(dup_vals_x0) &
      length(dup_vals_x0) == 0) {
    indices = order(x_all)[!duplicated(x_all) & x_all %in% x_0]
  } else if (length(dup_vals_xall) > length(dup_vals_x0) &
             length(dup_vals_x0) > 0) {
    indices_to_drop = order(x_all)[duplicated(x_all) &
                                   x_all %in% extra_dup_vals]
    indices = setdiff(order(x_all)[x_all %in% x_0], indices_to_drop)
  } else if (length(dup_vals_xall) == length(dup_vals_x0)) {
    indices = order(x_all)[x_all %in% x_0] }

  tau = quantile(x_all, probs = quar_trunc)
  tau_surv = KM_SurF(tau, obs = obs_all)

  data1 = list(X = input_data$X[obs1], delta = input_data$delta[obs1],
               U = input_data$U[obs1, U_index])
  X = data1$X; delta = data1$delta;
  selectU_1 = matrix(data1$U, nrow = length(obs1), ncol = length(U_index))
  rm(data1)

  if (length(obs1) == 1) {
      KM_SurF_terms = ifelse(X < tau, KM_SurF(X, obs = obs_all), tau_surv)
  } else {
      KM_SurF_terms = sapply(X, KM_SurF, obs = obs_all)
      KM_SurF_terms[X >= tau] = tau_surv }
```

```
mu_selectU_0 = matrix(rep(my_colMeans(selectU_0), length(obs1)),
                      nrow = length(obs1), byrow = TRUE)
selectU_0_colVars = colVars(selectU_0, sd_use = FALSE)
var_selectU_0 = matrix(rep(selectU_0_colVars, length(obs1)),
                       nrow = length(obs1), byrow = TRUE)

inverse_weight = Inverse_weight_func(x0 = x_all, delta0 = delta_all,
                                     obs0 = obs_all, quar_trunc,
                      err_msg = 'Error in IF_star Inverse_weight KM_table')
Y_Ghat_0 = (x_all*delta_all) / inverse_weight;
Y_Ghat_0[is.nan(Y_Ghat_0) | is.na(Y_Ghat_0)] = 0
Y_Ghat_0 = Y_Ghat_0[indices]
mean_YGhat_0 = mean(Y_Ghat_0)

time_comparison1 = outer(x_all, X, '==')
delta_matrix = 1 - delta_all
time_comparison2 = outer(x_all, X, '>=')
tryCatch({
  event_nums = colSums(time_comparison1*delta_matrix);
  risk_set = colSums(time_comparison2)
  inv_weight_mat = inverse_weight
  Y_Ghat_mat = time_comparison2*x_all*delta_all / inv_weight_mat
  Y_Ghat_mat[is.nan(Y_Ghat_mat) | is.na(Y_Ghat_mat)] = 0
  Y_Ghat_mat = Y_Ghat_mat[indices,]
  mean_YGhat_mat = colMeans(Y_Ghat_mat)

  cov_mat_0 = colCovs(selectU_0, Y_Ghat_0)
  Y_Ghat_1 = delta*X/KM_SurF_terms;
  Y_Ghat_1[is.nan(Y_Ghat_1) | is.na(Y_Ghat_1)] = 0
},
error = function(msg='Error in IF_star time_comparison dimension') {
  message("Original error message:"); message(paste(msg,"\n", sep = ''))
  save( input_data, file = paste('errdata_', r, '.Rdata', sep = '') )
  stop(msg) })

a = (((selectU_1 - mu_selectU_0) * (Y_Ghat_1 - mean_YGhat_0)) /
  var_selectU_0)
b = ((((selectU_1 - mu_selectU_0)^2) *
       matrix(rep(cov_mat_0, length(obs1)), nrow = length(obs1),
              byrow = TRUE)) / (var_selectU_0)^2)
IF_ipw = matrix(rep(m,length(obs1)), nrow = length(obs1),
                byrow = TRUE)*(a - b)
rm(a); rm(b)

### Slow when obs1 is large;
Ehat_pointwise_val = sapply(1:length(obs1), function(i){
   a = mean_YGhat_mat[i]
   selectU = matrix(selectU_0[time_comparison2[indices,i],],
                    ncol = length(U_index))
   n_s = dim(selectU)[1]; p_s = dim(selectU)[2]

   if (n_s == 0) { NaN
   } else {
```

```
        mean_selectU = my_colMeans(selectU);
        var_selectU = colVars(selectU, sd_use = FALSE)
        cov_Y_selectU = my_colMeans((selectU - matrix(rep(mean_selectU, n_s),
                                                   byrow = TRUE, ncol = p_s))
         * matrix(rep(Y_Ghat_mat[time_comparison2[indices,i],i], p_s),
                  ncol = p_s))
        b = cov_Y_selectU / var_selectU; b[is.nan(b) | is.na(b)] = 0
        u = matrix(selectU_1[i,], byrow = TRUE, ncol = length(U_index))
        c = u - matrix(rep(mean_selectU, dim(u)[1]), byrow = TRUE,
                       ncol = dim(u)[2])
        rm(u)
        return( a + c*b ) }})

  Ehat_pointwise_val[is.nan(Ehat_pointwise_val) |
                     is.na(Ehat_pointwise_val)] = 0
  rm(selectU_0)

  hazard = event_nums/risk_set; hazard[is.nan(hazard) | is.na(hazard)] = 0
  mart_X = (X <= tau & delta == 0) - (X <= tau)*hazard
  val_1 = ( matrix(t(Ehat_pointwise_val), nrow = length(obs1),
                   ncol = length(U_index))
            * matrix(rep(mart_X, length(U_index)), nrow = length(obs1),
                     ncol = length(U_index)) )
  IF_CAR = ( matrix(rep(m,length(obs1)), nrow = length(obs1), byrow = TRUE)
             * (selectU_1 - mu_selectU_0) * val_1 / var_selectU_0 )
  rm(selectU_1)
  return( round((IF_ipw - IF_CAR), 3) )
},

Stab_onestep_est = function(all_obs, chunk_size, elln, est_index, alpha,
                            num_top = 1, quar_trunc){
  # chunk_size, elln, alpha are scalars.
  mt = rowMeans( sapply(0:(ceiling((n - elln)/chunk_size) - 1), function(i){
              old_obs = all_obs[1:(elln + i*chunk_size)]
              if (i < ceiling((n - elln)/chunk_size) - 1) {
                 new_obs = all_obs[(elln + i*chunk_size + 1):
                                   (elln + (i + 1)*chunk_size)]
              } else {
                 new_obs = all_obs[(elln + i*chunk_size + 1):n]  }

              cors0 = Est_Psi0_d(U_index = 1:p, obs = old_obs, quar_trunc);
              sgn0 = rep(1, p)

              ### EXPENSIVE!!!
              if (p <= 1e4) {
                IF_star_mat = IF_star_self(m = sgn0, U_index = 1:p,
                                           obs = old_obs, quar_trunc)
                utility = (cors0 + my_colMeans(IF_star_mat))
              } else { utility = cors0 }

              if (num_top < p) {
                k0 = order(abs(utility), decreasing = TRUE)[seq(num_top)]
              } else { k0 = seq(p) }
```

```
          cor_k0 = cors0[k0]
          sgn_k0 = 2*(utility[k0] >= 0) - 1

          Uk0 = input_data$U[old_obs, k0, drop = FALSE]

          mu_Uk0 = colMeans(Uk0)
          sd_Uk0 = apply(Uk0, 2, sd)

          if (est_index == 'subsamples') {
            obs_all_val = old_obs; obs0_val = old_obs
          } else if (est_index == 'partial subsamples') {
            obs_all_val = all_obs; obs0_val = old_obs
          } else if (est_index == 'whole samples') {
            obs_all_val = all_obs; obs0_val = all_obs }

    curr_sigma_inv0 = 1/sd(rowMeans(IF_star(m = sgn_k0, U_index = k0,
                                            obs_all = obs_all_val,
                                            obs0 = obs0_val,
                                            obs1 = old_obs, quar_trunc)))
    est0 = mean( (mean(sgn_k0*cor_k0)
                 + rowMeans(IF_star(m = sgn_k0, U_index = k0,
                                    obs_all = obs_all_val,
                                    obs0 = obs0_val,
                                    obs1 = new_obs, quar_trunc)))
                 * curr_sigma_inv0 )
    return( c(est0, curr_sigma_inv0) )  } ) )
  est = mt[1]/mt[2]
  se = 1/(mt[2] * sqrt(n - elln))
  ci = c( est - qnorm(1 - alpha/2)/(mt[2] * sqrt(n - elln)),
          est + qnorm(1 - alpha/2)/(mt[2] * sqrt(n - elln)) )
  rej = 1*( ci[1] > 0 | ci[2] < 0 )
  p_val = 2*pnorm(abs(est/se), mean = 0, sd = 1, lower.tail = FALSE,
                  log.p = FALSE)


  k_est = sapply(n:n, function(i){
    old_obs = all_obs[1:n]

    #cors0 = Est_Psi0_d(U_index=1:p, obs=old_obs, quar_trunc)
    cors0 = .self$Est_Psi0_d(U_index = 1:p, obs = old_obs, quar_trunc)
    sgn0 = rep(1, p)

    ### EXPENSIVE!!!
    if (p <= 1e4) {
      # IF_star_mat = IF_star_self(m = sgn0, U_index = 1:p,
      #                            obs = old_obs, quar_trunc)
      IF_star_mat = .self$IF_star_self(m = sgn0, U_index = 1:p,
                                       obs = old_obs, quar_trunc)
      utility = (cors0 + my_colMeans(IF_star_mat))
    } else { utility = cors0 }

    if (num_top < p) {
      k0 = order(abs(utility), decreasing = TRUE)[seq(num_top)]
```

```
      } else { k0 = seq(p) }

      return( k0 ) } )

    if (is.na(mt[2])) {
      save(input_data, file = paste('NAdata_stabOSE_model', model, '_n', n,
                                    '_elln', elln, '_r', r, '.Rdata',
                                    sep = '') )
    }

    return( list(rej = rej, Sn = abs(est/se), est = est, se = se, ci = ci,
                 p_val = p_val, k_est = k_est) )
},

Oracle_onestep_est = function(alpha, quar_trunc, idx_taken=1){
  # Using 1:2 here is to compile the vectorized syntax in codes,
  # even if we only need the results of the first predictor
  cors0 = Est_Psi0_d(U_index = 1:2, obs = 1:n, quar_trunc)
  IF_star_mat = IF_star_self(m = rep(1,2), U_index = 1:2, obs = 1:n,
                             quar_trunc)
  sigma_IF_star = colVars(IF_star_mat, sd_use = TRUE)[idx_taken]
  est = (cors0 + my_colMeans(IF_star_mat))[idx_taken]
  rej = 1*( sqrt(n)*abs(est/sigma_IF_star) > qnorm(1 - alpha/2, 0, 1,
                                                   lower.tail = TRUE,
                                                   log.p = FALSE) )
  p_val = 2*pnorm( sqrt(n)*abs(est/sigma_IF_star), mean = 0, sd = 1,
                   lower.tail = FALSE, log.p = FALSE )

  if (is.nan(est) | is.na(sigma_IF_star)) {
    save(input_data, file = paste('NAdata_singleOSE_model', model, '_n', n,
                                  '_r', r, '.Rdata', sep = '') )
  }
  return( list( rej = rej, Sn = sqrt(n)*abs(est/sigma_IF_star), est = est,
                se = sigma_IF_star, p_val = p_val, k_est = 1 ) )
},

Naive_onestep_est = function(alpha, quar_trunc, num_top=1){

  cors0 = Est_Psi0_d(U_index = 1:p, obs = 1:n, quar_trunc); sgn0 = rep(1, p)

  ### EXPENSIVE!!!
  if (p <= 1e4) {
    IF_star_mat0 = IF_star_self(m = sgn0, U_index = 1:p, obs = 1:n,
                                quar_trunc)
    utility = (cors0 + my_colMeans(IF_star_mat0))
  } else { utility = cors0 }

  if (num_top < p) {
    k0 = order(abs(utility), decreasing = TRUE)[seq(num_top)]
  } else { k0 = seq(p) }

  cor_k0 = cors0[k0]
  sgn_k0 = 2*(utility[k0] >= 0) - 1
```

```
    IF_star_mat = IF_star_self(m = sgn_k0, U_index = k0, obs = 1:n, quar_trunc)
    est = (cor_k0 + mean(IF_star_mat))
    sigma_IF_star = sd(IF_star_mat)
    rej = 1*( sqrt(n)*abs(est/sigma_IF_star)
             > qnorm(1 - alpha/2, 0, 1, lower.tail = TRUE, log.p = FALSE) )
    abs_test_stat_val = sqrt(n)*abs(est/sigma_IF_star)
    p_val = min(1, 2*pnorm(abs_test_stat_val, mean = 0, sd = 1,
                           lower.tail = FALSE, log.p = FALSE))

    return( list( rej = rej, Sn = abs_test_stat_val,
                  est = est, se = sigma_IF_star, p_val = p_val, k_est = k0 ) )
  },

  Bonf_onestep_ests = function(alpha, quar_trunc){
    cors0 = Est_Psi0_d(U_index = 1:p, obs = 1:n, quar_trunc)
    IF_star_mat = IF_star_self(m = rep(1,p), U_index = 1:p, obs = 1:n,
                               quar_trunc)
    est = (cors0 + my_colMeans(IF_star_mat))
    sigma_IF_star = colVars(IF_star_mat, sd_use = TRUE)
    rej = 1*( sqrt(n)*max(abs(est/sigma_IF_star))
                        > qnorm(1 - alpha/(2*p), 0, 1,
                                lower.tail = TRUE, log.p = FALSE) )
    abs_test_stat_vals = sqrt(n)*abs(est/sigma_IF_star)
    p_vals = 2*pnorm(abs_test_stat_vals, mean = 0, sd = 1,
                     lower.tail = FALSE, log.p = FALSE)
    p_val = min(1, p*2*pnorm(max(abs_test_stat_vals), mean = 0, sd = 1,
                             lower.tail = FALSE, log.p = FALSE))

    return( list( rej = rej, Sn = sqrt(n)*max(abs(est/sigma_IF_star)),
                  est = est[which.max(abs_test_stat_vals)],
                  se = sigma_IF_star[which.max(abs_test_stat_vals)],
                  p_val = p_val, k_est = which.max(abs_test_stat_vals),
                  k_ests = which(p_vals <= alpha/p) ) )
  },

  Bonf_Cox_ests = function(alpha){ # alpha is a scalar
    data_est = cbind(input_data$X, input_data$delta, input_data$U)
    data_est = data_est[order(data_est[,1]),] %>% as.data.frame()
    colnames(data_est) = c('X', 'delta', paste('V', 1:p, sep = ''))

    result = do.call(rbind, lapply(1:p, function(idx){
      cox_fit = coxph(Surv(X, delta) ~ data_est[,paste('V',idx, sep = '')],
                      data_est, control = coxph.control(eps = 1e-05))
      pval = summary(cox_fit)$coefficients[,5]
      return( c(idx, pval) ) }))

    return( list( rej = 1*(min(result[,2]) <= alpha/p),
                  p_val = min(1, p*min(result[,2])),
                  k_est = result[which.min(result[,2]), 1],
                  k_ests = result[result[,2] <= alpha/p, 1] ) )
  }
 )
)
```

sim_data_acquisition

```
num_digits = 7

# Simulate data, according to the distributions described in the article
Simulate_data = function(n, p, model, censoring_rate, rho,
                         censoring_dist_par = c(0.07, 0.15, 0.28, 0.43),
                         just.first100 = FALSE, mc_true_par = FALSE,
                         block_size = 1e4){

  if (rho == 0) {
    U = matrix(rnorm(n*p), ncol = p)
  } else {
    a = uniroot( function(x){ 1 - rho - (x - sqrt((1 - x^2)/(p - 1)))^2 },
                 c(0,1), tol = .Machine$double.eps )$root
    b = sqrt((1 - a^2)/(p - 1))
    p0 = ifelse(just.first100, 100, p)
    tmp = matrix(rnorm(n*p0), ncol = p0)
    U = matrix( rep(rowSums(tmp), p0), ncol = p0 )*b + tmp[,1:p0]*(a - b)
    rm(tmp) }

  p0 = dim(U)[2]

  # Simulate data of error structure
  if (grepl('IE', model, fixed = TRUE)) {
    error_IE = rnorm(n, 0, 1)
  } else {
    error_DE = rnorm(n, 0, 0.7*(abs(U[,1]) + 0.7)) }

  # Simulate data of survival time based on the specified AFT model
  if (model == 'N.IE') {
    log_T = error_IE
  } else if (model == 'A1.IE') {
    log_T = 0.25*U[,1] + error_IE;
  } else if (model == 'A2.IE') {
    log_T = c( as.matrix(U[,1:10]) %*% matrix(c(rep(0.15,5),rep(-0.1,5)),
                                               ncol = 1) ) + error_IE
  } else if (model == 'N.DE') {
    log_T = error_DE
  } else if (model == 'A1.DE') {
    log_T = 0.25*U[,1] + error_DE
  } else if (model == 'A2.DE') {
    log_T = c( as.matrix(U[,1:10]) %*% matrix(c(rep(0.15,5),rep(-0.1,5)),
                                               ncol = 1) ) + error_DE
  } else stop('Invalid model choice.')

  if (!(mc_true_par)) {
    # Simulate data of censoring time
    if (censoring_rate == '0%') {
      C = 1e5
    }
    else{
      for (i in 1:length(censoring_dist_par)) {
        assign( paste('par', i, sep = ''), censoring_dist_par[i] ) }
```

```
      cr_par = par1*(censoring_rate == '10%') + par2*(censoring_rate == '20%') +
               par3*(censoring_rate == '30%') + par4*(censoring_rate == '40%')
      C = log(rexp(n, cr_par)) }

    X = round(pmin(log_T, C), num_digits);
    delta = 1*(log_T <= C);

    U_colSDs = colVars(U, sd_use = TRUE, na.rm = TRUE)
    U = colStandardization(U, U_colSDs)
    input_data = list(X = X, delta = delta, U = U)
  } else {
    U_colSDs = colVars(U, sd_use = TRUE, na.rm = TRUE)
    U = colStandardization(U, U_colSDs)
    log_T = round(log_T, num_digits)
    input_data = list(timeT = log_T, U = U)
  }
  return( input_data )
}
```

Simulation example

example_sim_maxCorrSurv

```
##### Configuration
num_r = 1; num_rdo = 1
# Estimation Index
est_index_val = 'whole samples'
# Censoring rate: '10%' or '30%'
cr = '10%'
# Methods to run
meths_vec = c('BONF_COX', 'Naive_OSE', 'BONF_OSE', 'OOSE', 'SOSE')
# List of (n,p) values to run simulation
np_list = list(c(500, 1000))
# List of correlation values between predictors
rho_val = 0.75
# Tau
quar = 0.9
# Significance level
alpha_val = 0.05
# elln
elln_part = c(2); dim_elln = length(elln_part)
# List of data generating distributions to use when running simulations
model_list = c('N.IE','A1.IE','A2.IE','N.DE','A1.DE','A2.DE')

# Number of digits used here
num_digits = 7; options(digits = num_digits)


# Make sure we have every package needed
package_list = c('survival', 'MASS', 'Matrix', 'mvtnorm', 'dplyr', 'here')
for (package in package_list) {
  require(package, character.only = TRUE);
  library(package, character.only = TRUE) }
```

```
sim = data.frame(method = NA, model = NA, n = NA, p = NA, rho = NA,
                 quar = NA, elln = NA, est = NA, se = NA,
                 lb_ci = NA, ub_ci = NA, p_val = NA)

set.seed(2023)
np = np_list[[1]]
n = np[1]; p = np[2]; rho = rho_val

for (r in 1:num_r) {
  for (model in model_list) {

    dat = Simulate_data(n, p, model, censoring_rate = cr, rho = rho,
                        censoring_dist_par = c(0.07, 0.15, 0.28, 0.43))
    obj0 <- NumericalStudy$new(input_data = dat)

    out = list()
    for (meth in meths_vec) {
      if (meth == 'SOSE') {

        SOSE_est = do.call(rbind, lapply(seq(num_rdo), function(rdo_idx){
          inds = sample(1:n)
          dat0 = list(X = dat$X[1:n][inds], delta = dat$delta[1:n][inds],
                      U = dat$U[1:n,1:p][inds,])
          obj1 <- NumericalStudy$new(input_data = dat0)

          do.call(rbind, lapply(1:length(elln_part), function(d_idx){
            d = elln_part[d_idx]; elln = ceiling(n/d)
            chunk_size = ceiling((n - elln)/10)
            SOSE_est1 = obj1$Stab_onestep_est(all_obs = 1:n, chunk_size,
                                              elln, est_index = est_index_val,
                                              alpha = alpha_val, num_top = 1,
                                              quar_trunc = quar)
            return( t( c( SOSE_est1$ci, SOSE_est1$est, SOSE_est1$se,
                          SOSE_est1$p_val, elln ) ) ) } ) )
        } ) )
        out$SOSE = SOSE_est

      } else if (meth == 'OOSE') {

        OOSE_est = obj0$Oracle_onestep_est(alpha = alpha_val,
                                           quar_trunc = quar, idx_taken = 1)
        out$OOSE = c( OOSE_est$est, OOSE_est$se, OOSE_est$p_val )

      } else if (meth == 'Naive_OSE') {

        Naive_OSE_est = obj0$Naive_onestep_est(alpha = alpha_val,
                                               quar_trunc = quar, num_top = 1)
        out$Naive_OSE = c( Naive_OSE_est$est, Naive_OSE_est$se,
                           Naive_OSE_est$p_val )

      } else if (meth == 'BONF_COX') {
```

```
        BONF_COX_est = obj0$Bonf_Cox_ests(alpha = alpha_val)
        out$BONF_COX = c( BONF_COX_est$p_val )

      } else if (meth == 'BONF_OSE') {

        BONF_OSE_est = obj0$Bonf_onestep_ests(alpha = alpha_val,
                                              quar_trunc = quar)
        out$BONF_OSE = c( BONF_OSE_est$est, BONF_OSE_est$se,
                          BONF_OSE_est$p_val )

      } else stop('Invalid method.') }

    sim = rbind(sim, do.call(rbind,lapply(meths_vec, function(meth) {
      if ( meth == 'SOSE' ) {
        result = data.frame( method = rep(meth, dim_elln),
                             model = rep(model, dim_elln),
                             n = rep(n, dim_elln), p = rep(p, dim_elln),
                             rho = rep(rho, dim_elln),
                             quar = rep(quar, dim_elln),
                             elln = out[[meth]][,6],
                             est = out[[meth]][,3], se = out[[meth]][,4],
                             lb_ci = out[[meth]][,1], ub_ci = out[[meth]][,2],
                             p_val = out[[meth]][,5] )

      } else if ( meth == 'BONF_COX' ) {
        result = data.frame( method = meth, model = model, n = n, p = p,
                             rho = rho, quar = quar, elln = NA,
                             est = NA, se = NA, lb_ci = NA, ub_ci = NA,
                             p_val = out[[meth]][1] )
      } else {
        result = data.frame( method = meth, model = model, n = n, p = p,
                             rho = rho, quar = quar, elln = NA,
                             est = out[[meth]][1], se = out[[meth]][2],
                             lb_ci = NA, ub_ci = NA,
                             p_val = out[[meth]][3] )
      }
      return( result ) })))
  }
}

sim = sim[-1,]
rownames(sim) = 1:nrow(sim)

##### sim contains the simulation output. Each row contains:
# method: Method run
# model: Data generating distributions
# n: Sample size
# p: Covariate dimension
# rho: Designed correlation of covariates
# quar: The quantile of outcomes used for truncation
# elln: Value of elln used by our method
# est: Estimate of the parameter
# se: Standard error of the statistic
```

```
# lb_ci: Lower bound of the confidence interval of the parameter
# ub_ci: Upper bound of the confidence interval of the parameter
# p-value: P-value of rejecting the null of no correlation
```

Results

TABLE S.3 Confidence intervals (C.I.) and p-values by method and model in the example simulation

Method	Model	n	p	C.I.	P-value
BONF_COX	N.IE	500	1000	NA	1.000
Naive_OSE	N.IE	500	1000	NA	0.001
BONF_OSE	N.IE	500	1000	NA	1.000
OOSE	N.IE	500	1000	NA	0.168
SOSE	N.IE	500	1000	(-0.025, 0.231)	0.114
BONF_COX	A1.IE	500	1000	NA	0.000
Naive_OSE	A1.IE	500	1000	NA	0.000
BONF_OSE	A1.IE	500	1000	NA	0.006
OOSE	A1.IE	500	1000	NA	0.000
SOSE	A1.IE	500	1000	(0.084, 0.344)	0.001
BONF_COX	A2.IE	500	1000	NA	0.001
Naive_OSE	A2.IE	500	1000	NA	0.000
BONF_OSE	A2.IE	500	1000	NA	0.017
OOSE	A2.IE	500	1000	NA	0.004
SOSE	A2.IE	500	1000	(0.064, 0.309)	0.003
BONF_COX	N.DE	500	1000	NA	1.000
Naive_OSE	N.DE	500	1000	NA	0.078
BONF_OSE	N.DE	500	1000	NA	1.000
OOSE	N.DE	500	1000	NA	0.641
SOSE	N.DE	500	1000	(-0.057, 0.279)	0.194
BONF_COX	A1.DE	500	1000	NA	0.000
Naive_OSE	A1.DE	500	1000	NA	0.000
BONF_OSE	A1.DE	500	1000	NA	0.003
OOSE	A1.DE	500	1000	NA	0.000
SOSE	A1.DE	500	1000	(0.035, 0.3)	0.014
BONF_COX	A2.DE	500	1000	NA	0.002
Naive_OSE	A2.DE	500	1000	NA	0.001
BONF_OSE	A2.DE	500	1000	NA	0.860
OOSE	A2.DE	500	1000	NA	0.018
SOSE	A2.DE	500	1000	(-0.028, 0.319)	0.099