A Model-Free Correlation Coefficient for Censored Data

Linlin Dai[1], Tengfei Li[2] and Kani Chen[3] *

[1] *Southwestern University of Finance and Economics*

[2] *University of North Carolina at Chapel Hill*

[3] *Hong Kong University of Science and Technology*

Supplementary Material

S1. Technical Proofs

S1.1 Proof of Theorem 1

Define $F_T(t) = P(T \le t)$ and $G_w(t) = W([t, \infty))$. Recall that $\widetilde{G}(t) = P(T \ge t)$ and $\widetilde{G}_X(t) = P(T \ge t \mid X)$. Here, $G_w(t)$ is left-continuous. Let

$$Q(X, T) := \int \mathrm{var}\{\widetilde{G}_X(t)\} \mathrm{d}W(t).$$

It is clear that (i) holds since $\mathrm{var}\{\mathbb{1}(T \ge t)\} \ge \mathrm{var}[E\{\mathbb{1}(T \ge t) \mid X\}]$ for every t.

*Corresponding author. Email: makchen@ust.hk.

We now prove the statement (ii). If X and T are independent, then $P(T \geq t \mid X) = P(T \geq t)$ or equivalently $\widetilde{G}_X(t) = \widetilde{G}(t)$ almost surely for every $t \in \mathbb{R}$. It follows that $\mathrm{var}\{\widetilde{G}(t)\} = 0$ and thus $\xi(X, T) = 0$.

Conversely, suppose $\xi(X, T) = 0$. There is a Borel set A such that $W(A) = W(\mathbb{R}) < \infty$ and $\mathrm{var}\{\widetilde{G}_X(t)\} = 0$ for every $t \in A$. Recall that $E\{\widetilde{G}_X(t)\} = \widetilde{G}(t)$, we have $\widetilde{G}_X(t) = \widetilde{G}(t)$ almost surely for all $t \in A$. We now show that $A = \mathbb{R}$. In fact, for any $t \in \mathbb{R}$, if $W(\{t\}) > 0$, then $t \in A$. So we assume $W(\{t\}) = 0$, which implies that $G_w(t)$ is right-continuous. Moreover, according to Assumption 1, since $W(\{t\}) = 0$, we must have $\mu(\{t\}) = 0$. Therefore, $\widetilde{G}(t)$ is also right-continuous at t.

If $G_w(s) < G_w(t)$ for all $s > t$, then $W([t, s)) > 0$ for any $s > t$. Hence, $A \cap [t, s) \neq \emptyset$ and there exists a sequence t_n in A such that t_n decreases to t. Since $\widetilde{G}_X(t_n) = \widetilde{G}(t_n)$ almost surely for each n, we have

$$\widetilde{G}_X(t) \geq \lim_{n\to\infty} \widetilde{G}_X(t_n) = \lim_{n\to\infty} \widetilde{G}(t_n) = \widetilde{G}(t).$$

It follows from $E\{\widetilde{G}_X(t)\} = \widetilde{G}(t)$ that $\widetilde{G}_X(t) = \widetilde{G}(t)$ almost surely.

Otherwise, we suppose there is some $s > t$ such that $G_w(s) = G_w(t)$. The largest value of s exists since G_w is left-continuous. If $s = \infty$, then $G_w(s) = G_w(t) = 0$. Assumption 1 implies that for any $t \in \mathbb{R}$ such that $G_w(t) = 0$, we must have $\widetilde{G}(t) = 0$. Therefore, $\widetilde{G}(s) = \widetilde{G}(t) = 0$. According to $E\{\widetilde{G}_X(t)\} = \widetilde{G}(t)$ and $\widetilde{G}_X(t) \geq 0$, we get $\widetilde{G}_X(t) = 0$ almost surely. The second case is that $W(\{s\}) > 0$, which implies $s \in A$. The last case is $W(\{s\}) = 0$, under which $G_w(r) < G_w(s)$ for any $r > s$. Hence, in the last two cases, we have $\widetilde{G}_X(s) = \widetilde{G}(s)$ almost surely. Since $G_w(t) - G_w(s) = W([t, s)) = W((t, s)) + W(\{t\}) = 0$, we have $W((t, s)) = 0$. Therefore, the open set (t, s) is not in the support $\mathcal{S}_W$. Recall that $\mathcal{S}_\mu \subseteq \mathcal{S}_W$. It follows that $(t, s) \subseteq \mathcal{S}_W^c \subseteq \mathcal{S}_\mu^c$. This implies $\mu((t, s)) = \widetilde{G}(t) - \widetilde{G}(s) - \mu(\{t\}) = 0$. According to Assumption 1, we have $\widetilde{G}(t) = \widetilde{G}(s)$ as $\mu(\{t\}) = 0$. As a result, we have $\widetilde{G}_X(t) \geq \widetilde{G}_X(s) = \widetilde{G}(s) = \widetilde{G}(t)$. It follows immediately from $E\{\widetilde{G}_X(t)\} = \widetilde{G}(t)$ that $\widetilde{G}_X(t) = \widetilde{G}(t)$ almost surely. This completes the proof of the claim that $A = \mathbb{R}$. Hence, for any $t \in \mathbb{R}$ and any Borel set $B \subset \mathbb{R}$,

$$
\begin{aligned}
P(\{T \geq t\} \cap \{X \in B\}) &= E\{P(T \geq t \mid X)\mathbb{1}(X \in B)\} \\
&= \widetilde{G}(t)P(X \in B) = P(T \geq t)P(X \in B).
\end{aligned}
$$

This completes the proof that X and T are independent.

Following similar steps in the proof of Corollary A.2 in Chatterjee (2021), since we have the claim that if T is not a constant, the denominator of the expression (2.1) is strictly greater than zero.

We now go ahead to prove the statement (iii). If T is a function of X almost surely, then

$$\int \mathrm{var}\{P(T \geq t \mid X)\}\mathrm{d}W(t) = \int \mathrm{var}\{\mathbb{1}(T \geq t)\}\mathrm{d}W(t).$$

As a result, $\xi(X,T) = 1$.

Conversely, suppose that $\xi(X,T) = 1$. Then,

$$\int [\mathrm{var}\{\mathbb{1}(T \geq t)\} - \mathrm{var}\{\widetilde{G}_X(t)\}]\mathrm{d}W(t) = \int E[\widetilde{G}_X(t)\{1 - \widetilde{G}_X(t)\}]\mathrm{d}W(t) = 0.$$

This implies that $P(E) = 1$, where E is the event

$$\int \widetilde{G}_X(t)\{1 - \widetilde{G}_X(t)\}\mathrm{d}W(t) = 0. \tag{S1.1}$$

Define $\widetilde{G}_x(t) = P(T \geq t \mid X = x)$, $a_x = \sup\{t : \widetilde{G}_x(t) = 1\}$ and $b_x = \inf\{t : \widetilde{G}_x(t) = 0\}$. Then we have $a_x \leq b_x$. Suppose that the event $\{a_X < b_X\} \cap E$ takes place. Since $\widetilde{G}_X(t) \in (0,1)$ for all $t \in (a_X, b_X)$, the condition (S1.1) implies that $W((a_X, b_X)) = 0$. As (a_X, b_X) is an open ball, we have $(a_X, b_X) \subseteq \mathcal{S}_W^c \subseteq \mathcal{S}_\mu^c$. On the other hand, we have $P\{T \in (a_X, b_X) \mid X\} > 0$, which further implies that $P(T \in \mathcal{S}_\mu^c) > 0$. But this is impossible, since $\mathcal{S}_\mu$ is the support of T. Therefore, $P(a_X = b_X) = 1$. So we have $T = a_X$ almost surely. This completes the proof of Theorem 1.

S1.2 Proof of Theorem 2

Recall that $T_{n1} = \int_{\mathbb{R}_+} Q_n(t)w(t)\mathrm{d}t$, $T_{n2} = \int_{\mathbb{R}_+} G_n(t)^2 w(t)\mathrm{d}t$, and $T_{n3} = \int_{\mathbb{R}_+} G_n(t)\hat{S}_c(t)w(t)\mathrm{d}t$. $\hat{\xi}_n(X,T)$ can be written as

$$\hat{\xi}_n(X,T) = \frac{T_{n1} - T_{n2}}{T_{n3} - T_{n2}}. \tag{S1.2}$$

We first come to prove the strong consistency of $\hat{\xi}_n(X,T)$ as below.

(1) Proof of the strong consistency of $\hat{\xi}_n(X,T)$:

Lemma 1 (Bitouzé et al., 1999, Theorem 1)**.** *Let $\{T_i\}_{i=1}^n$ and $\{C_i\}_{i=1}^n$ be independent sequences of independently identically distributed nonnegative random variables with distribution function $F_T(x)$ and $F_C(x)$. Let $\hat{F}_n$ be the Kaplan-Meier estimator of the distribution function $F_T(x)$. Then, there exists a positive constant, D, such that for any positive constant λ,*

$$P(\|(1-F_C)(\hat{F}_n - F_T)\|_\infty > \lambda n^{-1/2}) \leq 2.5\exp(-2\lambda^2 + D\lambda),$$

where $\|(1-F_C)(\hat{F}_n - F_T)\|_\infty = \sup_{t\in\mathbb{R}_+} |\{1-F_C(t)\}\{\hat{F}_n(t) - F_T(t)\}|$.

Recall that $G(t) = P(Y \geq t)$, $\widetilde{G}(t) = P(T \geq t)$, $G_X(t) = P(Y \geq t \mid X)$ and $\widetilde{G}_X(t) = P(T \geq t \mid X)$. Define $\Delta_n = \|G_n - G\|_\infty$. Then, by the Glivenko-Cantelli theorem, $\Delta_n \to 0$ almost surely. Hence, for T_{n1}, since $G_n(t)$ and $G(t)$ are both bounded and $\int_{\mathbb{R}_+} w(t)\mathrm{d}t < \infty$ in Assumption 2, we have

$$\begin{aligned}\left|\int_{\mathbb{R}_+} G_n(t)^2 w(t)\mathrm{d}t - \int_{\mathbb{R}_+} G(t)^2 w(t)\mathrm{d}t\right| &\leq 2\int_{\mathbb{R}_+} |G_n(t) - G(t)| w(t)\mathrm{d}t \\ &\leq 2\Delta_n \int_{\mathbb{R}_+} w(t)\mathrm{d}t \longrightarrow 0 \text{ almost surely, as } n \to \infty.\end{aligned}$$

Therefore, we have that as $n \to \infty$,

$$T_{n2} \longrightarrow T_{02} \equiv \int_{\mathbb{R}_+} G(t)^2 w(t)\mathrm{d}t \text{ almost surely.}$$

For T_{n3}, since $\hat{S}_c(t)$ is bounded by 1, we have

$$\begin{aligned}&\left|\int_{\mathbb{R}_+} G_n(t)\hat{S}_c(t)w(t)\mathrm{d}t - \int_{\mathbb{R}_+} G(t)S_c(t)w(t)\mathrm{d}t\right| \\ &\quad= \left|\int_{\mathbb{R}_+} \{G_n(t) - G(t)\}\hat{S}_c(t)w(t)\mathrm{d}t + \int_{\mathbb{R}_+} G(t)\{\hat{S}_c(t) - S_c(t)\}w(t)\mathrm{d}t\right| \\ &\quad\leq \int_{\mathbb{R}_+} |G_n(t) - G(t)| w(t)\mathrm{d}t + \int_{\mathbb{R}_+} |G(t)\{\hat{S}_c(t) - S_c(t)\}| w(t)\mathrm{d}t. \\ &\quad\leq \Delta_n \int_{\mathbb{R}_+} w(t)\mathrm{d}t + \int_{\mathbb{R}_+} |G(t)\{\hat{S}_c(t) - S_c(t)\}| w(t)\mathrm{d}t. \qquad \text{(S1.3)}\end{aligned}$$

We exchange roles of C and T and by Lemma 1, we have that there exists a positive constant D, such that for any $\lambda > 0$,

$$P\left(\sup_{t\in\mathbb{R}_+}|G(t)\{\hat{S}_c(t) - S_c(t)\}| > \lambda n^{-1/2}\right) \leq 2.5\exp(-2\lambda^2 + D\lambda).$$

Then, for any $\epsilon > 0$, taking $\lambda = n^{1/2}\epsilon$, one has

$$P\left(\sup_{t\in\mathbb{R}_+}|G(t)\{\hat{S}_c(t) - S_c(t)\}| > \epsilon\right) \leq 2.5\exp(-2n\epsilon^2 + Dn^{1/2}\epsilon).$$

This implies that $\sup_{t\in\mathbb{R}_+}|G(t)\{\hat{S}_c(t) - S_c(t)\}|$ converges to zero almost surely. Note that when T is discrete, this convergence still holds by Theorem 2 in Bitouzé et al. (1999). Therefore, according to (S1.3), we have that as $n \to \infty$,

$$T_{n3} \longrightarrow T_{03} \equiv \int_{\mathbb{R}_+} G(t)S_c(t)w(t)\mathrm{d}t \text{ almost surely.}$$

For T_{n1}, we come to prove that

$$T_{n1} \longrightarrow T_{01} \equiv \int_{\mathbb{R}_+} E\{G_X^2(t)\}w(t)\mathrm{d}t \text{ almost surely.}$$

We first show that $E(T_{n1})$ converges to T_{01}. Let $\mathcal{F}$ denote the σ-algebra generated by $\{X_i\}_{i=1}^n$ and the random variables used for breaking ties in the selection of nearest neighbors. Then, by Lemma 11.7 and Lemma 11.8 in Azadkia and Chatterjee (2021), we have

$$
\begin{aligned}
E(T_{n1}) &= n^{-1}\sum_{i=1}^{n}\int_{\mathbb{R}_+} E\{\mathbb{1}(Y_i \geq t)\mathbb{1}(Y_{N(i)} \geq t)\}w(t)\mathrm{d}t \\
&= n^{-1}\sum_{i=1}^{n}\int_{\mathbb{R}_+} E\big[E\{\mathbb{1}(Y_i \geq t)\mathbb{1}(Y_{N(i)} \geq t)|\mathcal{F}\}\big]w(t)\mathrm{d}t \\
&= n^{-1}\sum_{i=1}^{n}\int_{\mathbb{R}_+} E\big\{G_{X_1}(t)G_{X_{N(1)}}(t)\big\}w(t)\mathrm{d}t \\
&= \int_{\mathbb{R}_+} E\big\{G_{X_1}(t)G_{X_{N(1)}}(t)\big\}w(t)\mathrm{d}t \longrightarrow \int_{\mathbb{R}_+} E\{G_X(t)^2\}w(t)\mathrm{d}t = T_{01}.
\end{aligned}
$$

Hence, we have $E(T_{n1}) \to T_{01}$. To proceed, we show the following lemma.

Lemma 2. *There exists some positive constants $C(p,w)$ depending on both p and $C_w = \int_{\mathbb{R}_+} w(t)\mathrm{d}t$, and C_1 such that for any n and $\epsilon > 0$,*

$$P\{|T_{n1} - E(T_{n1})| \geq \epsilon\} \leq C_1 \exp\{-C(p,w)n\epsilon^2\}.$$

Proof. Let $U_1, \cdots, U_n$ be i.i.d. random variables from $U(0,1)$, which is the uniform distribution between $(0,1)$. The random variables $\{U_i\}_{i=1}^n$ are used for breaking ties at random if X_i has more than one neighbors. We attempt to use the bounded difference concentration inequality to complete the proof. So we need to figure out the maximum magnitude of change of T_{n1} when (X_i, Y_i, U_i) is replaced by the alternative (X_i', Y_i', U_i').

Define an equivalence relationship in $\{1, 2, \cdots, n\}$, that is, if $X_i = X_j$ then i and j are equivalent. We call the set of equivalence to be a "cluster" if there are more than one elements and otherwise, the "singleton". And let $\mathcal{A}_c$ be the collection of all clusters and $\mathcal{A}_s$ be the set of all singletons. Then, if i belongs to a cluster, $N(i)$ would be chosen at random by U_i among the indices in that cluster except for i.

Recall $x \wedge y = \min\{x, y\}$. Since

$$T_{n1} = \frac{1}{n}\sum_{i=1}^{n}\int_0^{Y_i \wedge Y_{N(i)}} w(t)\mathrm{d}t = \frac{1}{n}\sum_{i=1}^{n}\{F_W(Y_i) \wedge F_W(Y_{N(i)})\}, \tag{S1.4}$$

where $F_W(t) = \int_0^t w(s)\mathrm{d}s$, we have

$$T_{n1} = \frac{1}{n}\sum_{i=1}^{n} F_W(Y_i \wedge Y_{N(i)}) = \frac{1}{n}\sum_{\mathcal{C}\in\mathcal{A}_c}\sum_{i\in\mathcal{C}} F_W(Y_i \wedge Y_{N(i)}) + \frac{1}{n}\sum_{i\in\mathcal{A}_s} F_W(Y_i \wedge Y_{N(i)}).$$

Let $\mathcal{F}_1$ be the σ-algebra generated by $\{(X_i, Y_i)\}_{i=1}^n$ and $\{U_i\}_{i\in\mathcal{A}_s}$. Then,

$$\begin{aligned}
E(T_{n1} \mid \mathcal{F}_1) &= \frac{1}{n}\sum_{\mathcal{C}\in\mathcal{A}_c}\sum_{i\in\mathcal{C}} E\{F_W(Y_i \wedge Y_{N(i)}) \mid \mathcal{F}_1\} + \frac{1}{n}\sum_{i\in\mathcal{A}_s} F_W(Y_i \wedge Y_{N(i)}) \\
&= \frac{1}{n}\sum_{\mathcal{C}\in\mathcal{A}_c}\sum_{i\in\mathcal{C}} \frac{1}{|\mathcal{C}|-1}\sum_{j\in\mathcal{C}\setminus\{i\}} F_W(Y_i \wedge Y_j) + \frac{1}{n}\sum_{i\in\mathcal{A}_s} F_W(Y_i \wedge Y_{N(i)}) \\
&= \frac{1}{n}\sum_{\mathcal{C}\in\mathcal{A}_c}\sum_{i\in\mathcal{C}} \frac{1}{|\mathcal{C}|-1}\sum_{j\in\mathcal{C}\setminus\{i\}} F_W(Y_i \wedge Y_j) + \frac{1}{n}\sum_{i\in\mathcal{A}_s} F_W(Y_i \wedge Y_{N(i)}) \\
&= \frac{1}{n}\sum_{\mathcal{C}\in\mathcal{A}_c} a(\mathcal{C}) + \frac{1}{n}\sum_{i\in\mathcal{A}_s} F_W(Y_i \wedge Y_{N(i)}),
\end{aligned}$$

where

$$a(\mathcal{C}) = \frac{1}{|\mathcal{C}|-1} \sum_{i\in\mathcal{C}} \sum_{j\in\mathcal{C}\setminus\{i\}} F_W(Y_i \wedge Y_j).$$

Note that T_{n1} conditional on $\mathcal{F}_1$ is a function of all U_i with index $i \notin \mathcal{A}_s$. By replacing U_i with some other value U_i', it is possible that $N(i)$ is changed. And thus the value of T_{n1} will change at most C_w/n. It follows from the McDiarmid's inequality that for any $\epsilon > 0$,

$$P\{|T_{n1} - E(T_{n1} \mid \mathcal{F}_1)| \geq \epsilon \mid \mathcal{F}_1\} \leq 2\exp\{-2C_w^{-2} n\epsilon^2\}.$$

As the right is deterministic, we remove the conditioning on the left-hand side. This implies that $E\{|T_{n1} - E(T_{n1} \mid \mathcal{F}_1)|\} < (2\pi)^{1/2} C_w n^{-1/2}$. Therefore, we have

$$\begin{aligned} &P\big\{|T_{n1} - E(T_{n1})| \geq (2\pi)^{1/2} C_w n^{-1/2} + \epsilon\big\} \\ &\leq P\big\{|T_{n1} - E(T_{n1} \mid \mathcal{F}_1)| \geq \epsilon/2\big\} \\ &\qquad\qquad + P\big[|E(T_{n1} \mid \mathcal{F}_1) - E\{E(T_{n1} \mid \mathcal{F}_1)\}| \geq \epsilon/2\big] \\ &\leq 2e^{-C_w^{-2} n\epsilon^2/2} + P\big[|E(T_{n1}|\mathcal{F}_1) - E\{E(T_{n1} \mid \mathcal{F}_1)\}| \geq \epsilon/2\big]. \end{aligned} \tag{S1.5}$$

So we now need to show the bound for $P(|E(T_{n1} \mid \mathcal{F}_1) - E\{E(T_{n1} \mid \mathcal{F}_1)\}| \geq \epsilon/2)$. We first consider replace X_i with X_i' and remain Y_i and U_i unchanged in $E(T_{n1} \mid \mathcal{F}_1)$. Recall that Lemma 11.4 in Azadkia and Chatterjee (2021) claims that for any j, there is at most $\tilde{C}(p)$ number of X_i in singletons, of which the neighbor is X_j. Here, $\tilde{C}(p)$ is some positive constant depending on p. Now, we consider following cases.

First, suppose that $i \in \mathcal{C}$, where $\mathcal{C}$ is of size greater than 3 and by replacing X_i with X_i', $i \in \mathcal{C}'$, which is another cluster. Since X_i can be at most $\tilde{C}(p)$ number of nearest neighbors of X_j in singletons, the second part in $E(T_{n1} \mid \mathcal{F}_1)$ would change at most $\tilde{C}(p)C_w/n$. On the other hand, for the changes in the cluster part, we can calculate the change in $a(\mathcal{C})$ after the element i is removed. We have

$$\begin{aligned}
|a(\mathcal{C}) - a(\mathcal{C}\backslash\{i\})| &= \left| \frac{1}{|\mathcal{C}|-1} \sum_{k\in\mathcal{C}} \sum_{j\in\mathcal{C}\backslash\{k\}} F_W(Y_k \wedge Y_j) - \frac{1}{|\mathcal{C}|-2} \sum_{k\in\mathcal{C}\backslash\{i\}} \sum_{j\in\mathcal{C}\backslash\{i,k\}} F_W(Y_k \wedge Y_j) \right| \\
&= \left| \frac{1}{|\mathcal{C}|-1} \sum_{k\in\mathcal{C}\backslash\{i\}} F_W(Y_k \wedge Y_i) + \frac{1}{|\mathcal{C}|-1} \sum_{j\in\mathcal{C}\backslash\{i\}} F_W(Y_i \wedge Y_j) \right. \\
&\qquad \left. - \frac{1}{(|\mathcal{C}|-1)(|\mathcal{C}|-2)} \sum_{k\in\mathcal{C}\backslash\{i\}} \sum_{j\in\mathcal{C}\backslash\{i,k\}} F_W(Y_k \wedge Y_j) \right|.
\end{aligned}$$

Hence, $|a(\mathcal{C}) - a(\mathcal{C}\backslash\{i\})|$ is bounded by $3C_w$. Similar results can be found for $|a(\mathcal{C}' \cup \{i\}) - a(\mathcal{C}')|$. As a result, in this case, the change in $E(T_{n1} \mid \mathcal{F}_1)$ is at most $\max\{3C_w/n, \tilde{C}(p)C_w/n\}$.

Secondly, suppose $i \in \mathcal{C}$ with at least 3 elements. After replacement, the index i merges with some singleton and form a new cluster. So, similarly, the change in $a(\mathcal{C})$ is at most $3C_w/n$ and the second part of $E(T_{n1} \mid \mathcal{F}_1)$ would change at most $\tilde{C}(p)C_w/n$. The formation of the new cluster will change the values by at most $2C_w/n$. Therefore, the change in $E(T_{n1} \mid \mathcal{F}_1)$ is bounded by $\max\{3C_w/n, \tilde{C}(p)C_w/n\}$.

The other cases can have a similar conclusion and one can refer to cases (3)–(7) in Lemma 11.9 of Azadkia and Chatterjee (2021). Therefore, by replacing X_i with X_i', the variation in $E(T_{n1} \mid \mathcal{F}_1)$ is bounded by $C(p,w)/n$, where $C(p,w)$ is some constant relies on p and C_w.

We now consider replacing Y_i with Y_i' and remain U_i, X_i fixed. Then, $F_W(Y_i \wedge Y_k)$ changed by at most $2C_w$. If $i \in \mathcal{A}_s$, the $E(T_{n1} \mid \mathcal{F}_1)$ will change at most $2\tilde{C}(p)C_w/n$. If $i \in \mathcal{C}$, the $E(T_{n1} \mid \mathcal{F}_1)$ will change at most $C_1(p,w)/n$ for some constant C_1 depending on p and C_w. At last, if we replace U_i with U_i' for $i \in \mathcal{A}_s$, then the only possible variation is the index $N(i)$, leading to at most $2C_w/n$ changes in $E(T_{n1} \mid \mathcal{F}_1)$. In summary, using again the McDiarmid's inequality, we have that for any $\epsilon > 0$,

$$P\big(|E(T_{n1} \mid \mathcal{F}_1) - E[E(T_{n1} \mid \mathcal{F}_1)]| \geq \epsilon/2\big) \leq 2\exp\{-C(p,w)n\epsilon^2\}. \tag{S1.6}$$

Therefore, by (S1.5) and (S1.6), one has that for some constant C_0 and $C(p)$ depending on p,

$$P\big(|T_{n1} - E(T_{n1})| \geq (2\pi)^{1/2}C_w n^{-1/2} + \epsilon\big) \leq C_0 \exp\{-C(p,w)n\epsilon^2\}.$$

If $\epsilon \geq (2\pi)^{1/2}C_w n^{-1/2}$, then $P(|T_{n1} - E(T_{n1})| \geq 2\epsilon) \leq C_0 \exp\{-C(p,w)n\epsilon^2\}$. Otherwise, one can choose C_0 large enough such that $C_0 \exp\{-C(p,w)n\epsilon^2\} \geq 1$. As a result, $P(|T_{n1} - E(T_{n1})| \geq 2\epsilon) \leq C_0 \exp\{-C(p,w)n\epsilon^2\}$ holds. This completes the proof. $\square$

It follows from Lemma 2 and the result $E(T_{n1}) \to T_{01}$ that $T_{n1} \to T_{01}$ almost surely. Combining with results that both $T_{n2} \to T_{02}$ and $T_{n3} \to T_{03}$ almost surely, we get that

$$\hat{\xi}_n(X,T) \longrightarrow \frac{T_{01} - T_{02}}{T_{03} - T_{02}} = \frac{\int_{\mathbb{R}_+} E\{G_X^2(t)\}w(t)\mathrm{d}t - \int_{\mathbb{R}_+} G(t)^2 w(t)\mathrm{d}t}{\int_{\mathbb{R}_+} G(t)S_c(t)w(t)\mathrm{d}t - \int_{\mathbb{R}_+} G(t)^2 w(t)\mathrm{d}t} \tag{S1.7}$$

almost surely.

It is not yet clear to see that the limit in (S1.7) can be regarded as a measure of the association strength between X and T. Recall that we take $\mathrm{d}W(t) = S_c^2(t)w(t)\mathrm{d}t$ and C is independent with (X,T). By noting that $P(Y \geq t \mid X) = P(T \geq t \mid X)P(C \geq t) = \widetilde{G}_X(t)S_c(t)$, one can easily derive that

$$\begin{aligned} T_{01} &= \int_{\mathbb{R}_+} E\Big\{\frac{G_X^2(t)}{S_c^2(t)}\Big\} S_c^2(t)w(t)\mathrm{d}t = \int_{\mathbb{R}_+} E\Big\{\frac{P(Y \geq t \mid X)^2}{P(C \geq t \mid X)^2}\Big\} S_c^2(t)w(t)\mathrm{d}t \\ &= \int_{\mathbb{R}_+} E\{P(T \geq t \mid X)^2\} S_c^2(t)w(t)\mathrm{d}t = \int_{\mathbb{R}_+} E\{\widetilde{G}_X^2(t)\}\mathrm{d}W(t). \end{aligned} \tag{S1.8}$$

Similarly, we have

$$T_{02} = \int_{\mathbb{R}_+} \widetilde{G}(t)^2 \mathrm{d}W(t) \quad \text{and} \quad T_{03} = \int_{\mathbb{R}_+} \widetilde{G}(t)\mathrm{d}W(t). \tag{S1.9}$$

Combining (S1.8) and (S1.9), we get that the limit in (S1.7) can be rewritten in the form of

$$\begin{aligned}\xi(X,T) &= \frac{\int_{\mathbb{R}_+} E\{\widetilde{G}_X^2(t)\}\mathrm{d}W(t) - \int_{\mathbb{R}_+} \widetilde{G}(t)^2\mathrm{d}W(t)}{\int_{\mathbb{R}_+} \widetilde{G}(t)\mathrm{d}W(t) - \int_{\mathbb{R}_+} \widetilde{G}(t)^2\mathrm{d}W(t)} \\ &= \frac{\int_{\mathbb{R}_+} \mathrm{var}[E\{\mathbb{1}(T\geq t)\mid X\}]\mathrm{d}W(t)}{\int_{\mathbb{R}_+} \mathrm{var}\{\mathbb{1}(T\geq t)\}\mathrm{d}W(t)}.\end{aligned}$$

This completes the proof of strong consistency of $\hat{\xi}_n(X,T)$. Next, we go ahead to prove its asymptotic normality. Before that, we need some additional assumptions.

(2) Proof of the asymptotic normality of $\hat{\xi}_n(X,T)$:

First, we note that, as shown in Lin and Han (2022) (on page 21) and Theorem 4.1 in Azadkia and Chatterjee (2021), under the assumption (A2), by the boundedness of $G_X(t)$, we have

$$\begin{aligned}|T_{01}^* - T_{01}| &= \Big|E\{F_W(Y_1\wedge Y_{N(1)})\} - T_{01}\Big| \\ &= \Big|\int_{\mathbb{R}_+} E\{G_{X_1}(t)G_{X_{N(1)}}(t)\}w(t)\mathrm{d}t - \int_{\mathbb{R}_+} E\{G_X(t)^2\}w(t)\mathrm{d}t\Big| \\ &\leq \int_{\mathbb{R}_+} E\Big|G_{X_{N(1)}}(t) - G_X(t)\Big|w(t)\mathrm{d}t \\ &= O\bigg(\frac{(\log n)^{\{d+\beta+1+\mathbb{1}(d=1)\}}}{n^{1/d}}\bigg).\end{aligned}$$

It follows that $\xi_n^*(X,T)=\xi(X,T)+O(\log(n)^{\{d+\beta+1+\mathbb{1}(d=1)\}}/n^{1/d})$. Recall (S1.2) and together with the Slutsky's Theorem, one can derive that

$$\begin{aligned}\sqrt{n}\{\hat{\xi}_n(X,T)-\xi_n^*(X,T)\} &= \sqrt{n}\left(\frac{T_{n1}-T_{n2}}{T_{n3}-T_{n2}}-\frac{T_{01}^*-T_{02}}{T_{03}-T_{02}}\right)\\ &= K_1\sqrt{n}(T_{n1}-T_{01}^*)+K_2\sqrt{n}(T_{n2}-T_{02})\\ &\quad + K_3\sqrt{n}(T_{n3}-T_{03})+o_p(1), \qquad \text{(S1.10)}\end{aligned}$$

where $K_1=(T_{03}-T_{02})^{-1}$, $K_2=(T_{01}-T_{03})/(T_{03}-T_{02})^2$ and $K_3=(T_{02}-T_{01})/(T_{03}-T_{02})^2$. According to (S1.4), we have

$$\sqrt{n}(T_{n1}-T_{01}^*)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{F_W(Y_i\wedge Y_{N(i)})-T_{01}^*\}. \qquad \text{(S1.11)}$$

Next, we go ahead to deduce the asymptotic expansions for $\sqrt{n}(T_{n2}-T_{02})$ and $\sqrt{n}(T_{n3}-T_{03})$, respectively.

$$\begin{aligned}\sqrt{n}(T_{n2}-T_{02}) &= \sqrt{n}\int_{\mathbb{R}_+}(G_n(t)^2-G(t)^2)w(t)\mathrm{d}t.\\ &= 2\int_{\mathbb{R}_+}\sqrt{n}\{G_n(t)-G(t)\}G(t)w(t)\mathrm{d}t+\Delta_{n1}\\ &= \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\int_{\mathbb{R}_+}\{\mathbb{1}(Y_i\geq t)-G(t)\}G(t)w(t)\mathrm{d}t+\Delta_{n1}, \qquad \text{(S1.12)}\end{aligned}$$

where $\Delta_{n1} = \int_{\mathbb{R}_+} \sqrt{n}\{G_n(t) - G(t)\}^2 w(t)\mathrm{d}t$. By Donsker's theorem and continuous mapping theorem, one can easily show that the sequence $\sqrt{n}\{G_n(t) - G(t)\}$ converges in distribution to a Gaussian process $B^{\circ}(G(t))$ with zero mean and covariance given by $\mathrm{Cov}(B^{\circ}(G(t)), B^{\circ}(G(s))) = \{G(t) \wedge G(s)\} - G(t)G(s)$, where B° is a standard Brownian bridge on the unit interval. Therefore, together with $C_w = \int_{\mathbb{R}_+} w(t)\mathrm{d}t < \infty$, one has $\Delta_{n1} = o_p(1)$.

For T_{n3}, we have

$$
\begin{aligned}
&\sqrt{n}(T_{n3} - T_{03}) \\
&= \sqrt{n} \int_{\mathbb{R}_+} \{G_n(t)\hat{S}_c(t) - G(t)S_c(t)\} w(t)\mathrm{d}t \\
&= \sqrt{n} \left[\int_{\mathbb{R}_+} \{G_n(t) - G(t)\} S_c(t) w(t)\mathrm{d}t + \int_{\mathbb{R}_+} G(t)\{\hat{S}_c(t) - S_c(t)\} w(t)\mathrm{d}t \right] + \Delta_{n2},
\end{aligned}
\tag{S1.13}
$$

where

$$
\Delta_{n2} = \int_{\mathbb{R}_+} \sqrt{n}\{G_n(t) - G(t)\}\{\hat{S}_c(t) - S_c(t)\} w(t)\mathrm{d}t = o_p(1),
$$

according to the uniform consistency of the Kaplan-Meier estimator $\hat{S}_c(t)$ as shown in Theorem 3.4.2 in Fleming and Harrington (2013).

In addition, we introduce some notations for survival analysis. Let $\lambda_c(t)$ denote the hazard function of the censoring variable C. Denote $[n] = \{1, \ldots, n\}$. For $i \in [n]$, we define the individual counting process by $N_i^c(t) = \mathbb{1}(Y_i \le t, \delta_i = 0)$. Then, $M_i(t) = N_i^c(t) - \int_0^t \mathbb{1}(Y_i \ge s)\lambda_c(s)\mathrm{d}s$ is a martingale adapted to the filtration $\mathcal{F}_t$, which is the σ-algebra generated by $\{\mathbb{1}(Y_i \le s, \delta_i = 1), \mathbb{1}(Y_i \le s, \delta_i = 0), X_i : s \le t, i \in [n]\}$; see chap 1. and 2. in Fleming and Harrington (2013). Let $\tau = \sup\{t : P(Y > t) > 0\}$. Then, for any $t > \tau$, it holds that $N_i^c(t) = N_i^c(\tau)$. Hence, $\int_{\mathbb{R}_+} g(t)\mathrm{d}N_i^c(t) = \int_0^\tau g(t)\mathrm{d}N_i^c(t)$; the same holds for $M_i(t)$. By the definition of $\hat{S}_c(t)$ as shown similarly in Li et al. (2016), one can easily obtain that

$$\begin{aligned}
\sqrt{n}\{\hat{S}_c(t) - S_c(t)\} &= -\frac{1}{\sqrt{n}}\sum_{i=1}^n S_c(t)\int_0^t \frac{\mathrm{d}M_i(s)}{P(Y > s)} + o_p(1) \\
&= -\frac{1}{\sqrt{n}}\sum_{i=1}^n S_c(t)\left[\int_0^t \frac{\mathrm{d}N_i^c(s)}{P(Y > s)} - \int_0^t \frac{\mathbb{1}(Y_i \ge s)}{P(Y > s)}\lambda_c(s)\mathrm{d}s\right] + o_p(1) \\
&\equiv -\frac{1}{\sqrt{n}}\sum_{i=1}^n M(Y_i, \delta_i, t)S_c(t) + o_p(1), \qquad \text{(S1.14)}
\end{aligned}$$

where $M(Y_i, \delta_i, t)$ is a zero-mean martingale. Then, plugging (S1.14) in (S1.13),

$$\begin{aligned}
&\sqrt{n}(T_{n3} - T_{03}) = \sqrt{n}\int_{\mathbb{R}_+}\{G_n(t)\hat{S}_c(t) - G(t)S_c(t)\}w(t)\mathrm{d}t \\
&= \sqrt{n}\left[\int_{\mathbb{R}_+}\{G_n(t) - G(t)\}S_c(t)w(t)\mathrm{d}t + \int_{\mathbb{R}_+} G(t)\{\hat{S}_c(t) - S_c(t)\}w(t)\mathrm{d}t\right] + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_{\mathbb{R}_+}\{\mathbb{1}(Y_i \ge t) - G(t) - M(Y_i, \delta_i, t)G(t)\}S_c(t)w(t)\mathrm{d}t + o_p(1). \qquad \text{(S1.15)}
\end{aligned}$$

Let $f_c(t)$ denote the density function of the censoring variable C and then $\lambda_c(t) = f_c(t)/S_c(t)$. Recall $G(t) = P(Y \geq t)$ and "$0/0 = 0$". Then, for any $0 < s \leq t$, $G(t)/G(s) \leq 1$ and $S_c(t)/S_c(s) \leq 1$. Then, according to the boundedness of $G(t)$, $S_c(t)$ and the indicator function, one deduces

$$
\begin{aligned}
&\left| \int_{\mathbb{R}_+} M(Y_i, \delta_i, t) G(t) S_c(t) w(t) \mathrm{d}t \right| \\
&\quad = \left| \int_{\mathbb{R}_+} \left(S_c(t) \int_0^t \frac{G(t)}{G(s)} \mathrm{d}N_i^c(s) - \int_0^t \mathbb{1}(Y_i \geq s) f_c(s) \frac{G(t)}{G(s)} \frac{S_c(t)}{S_c(s)} \mathrm{d}s \right) w(t) \mathrm{d}t \right| \\
&\quad \leq \left| \int_{\mathbb{R}_+} \left(S_c(t) \int_0^t \mathrm{d}N_i^c(s) \right) w(t) \mathrm{d}t \right| + \left| \int_{\mathbb{R}_+} \left(\int_0^t \mathbb{1}(Y_i \geq s) f_c(s) \mathrm{d}s \right) w(t) \mathrm{d}t \right| \leq 2C_w.
\end{aligned}
\tag{S1.16}
$$

In summary, combining (S1.11), (S1.12) and (S1.15), we have that

$$
\begin{aligned}
&\sqrt{n}\{\hat{\xi}_n(X,T) - \xi_n^*(X,T)\} \\
&= K_1 \sqrt{n}(T_{n1} - T_{01}^*) + K_2 \sqrt{n}(T_{n2} - T_{02}) + K_3 \sqrt{n}(T_{n3} - T_{03}) + o_p(1), \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \Bigg[K_1 \{F_W(Y_i \wedge Y_{N(i)}) - T_1^*\} + 2K_2 \int_{\mathbb{R}_+} \{\mathbb{1}(Y_i \geq t) - G(t)\} G(t) w(t) \mathrm{d}t \\
&\qquad + K_3 \int_{\mathbb{R}_+} \{\mathbb{1}(Y_i \geq t) - G(t) - M(Y_i, \delta_i, t) G(t)\} S_c(t) w(t) \mathrm{d}t \Bigg] + o_p(1).
\end{aligned}
\tag{S1.17}
$$

Denote $Z_i = (X_i, T_i, C_i)$ for all $i \in [n]$. Then, we write

$$\sqrt{n}\{\hat{\xi}_n(X,T) - \xi_n^*(X,T)\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h_i(Z_1,\ldots,Z_n) + o_p(1) = W_n + o_p(1). \quad \text{(S1.18)}$$

Notice that $E(W_n) = 0$ and it suffices to develop the self-normalization central limit theorem for W_n.

Let μ_n be the law of $\{W_n - E(W_n)\}/\{\text{var}(W_n)\}^{1/2}$ and ν denote the standard Gaussian law. The Kantorovich-Wasserstein distance between the two probability measures is defined as

$$\delta_{W_n} = \sup\left\{\left|\int h d\mu_n - \int h d\nu\right| : h \text{ is a Lipschitz function, with } ||h||_{Lip} \leq 1\right\}.$$

Note that

(i) for any $i \in [n]$, $h_i(Z_1, \cdots, Z_n)$ is a function of Z_i and its nearest neighbor $Z_{N(i)}$ with nearest neighbor graph constructed by $\{X_i\}_{i=1}^n$;

(ii) for any $i \in [n]$, $h_i(Z_1, \cdots, Z_n)$ is bounded, since $F_W(t)$, $G(t)$, $G_X(t)$ and $S_c(t)$ are all bounded and the inequality (S1.16) holds;

(iii) according to Assumption (A1) and Lemma 3 as below, we have

$$0 < \liminf_{n\to\infty} \text{var}(W_n) \leq \limsup_{n\to\infty} \text{var}(W_n) < \infty.$$

Therefore, leveraging Theorem 3.4 in Chatterjee (2008) with some minor modification as that in Lin and Han (2022) for $Z_1, \ldots, Z_n$, we obtain that $\lim_{n\to\infty} \delta_{W_n} = 0$. Since the convergence in Kantorovich-Wasserstein distance is stronger than weak convergence, one has that

$$\frac{\sqrt{n}\{\xi_n(X,T) - \xi_n^*(X,T)\}}{\sqrt{\mathrm{var}(W_n)}} \longrightarrow N(0,1) \text{ in distribution.}$$

We now prove the following lemma.

Lemma 3. *$\limsup_{n\to\infty} \mathrm{var}(W_n) < \infty$.*

Proof. Recall (S1.17) and (S1.18), we have

$$\begin{aligned}
\mathrm{var}(W_n) &= \frac{1}{n}\mathrm{var}\bigg\{\sum_{i=1}^n h_i(Z_1, \ldots, Z_n)\bigg\} \\
&= \mathrm{var}\bigg[\frac{K_1}{\sqrt{n}}\sum_{i=1}^n \{F_W(Y_i \wedge Y_{N(i)}) - T_{01}^*\} \\
&\qquad + \frac{2K_2}{\sqrt{n}}\sum_{i=1}^n \int_{\mathbb{R}_+} \{\mathbb{1}(Y_i \geq t) - G(t)\}G(t)w(t)\mathrm{d}t \\
&\qquad + \frac{K_3}{\sqrt{n}}\sum_{i=1}^n \int_{\mathbb{R}_+} \{\mathbb{1}(Y_i \geq t) - G(t) - M(Y_i, \delta_i, t)G(t)\}S_c(t)w(t)\mathrm{d}t\bigg] \\
&\equiv \mathrm{var}(A_1 + A_2 + A_3).
\end{aligned}$$

By the Cauchy-Schwarz inequality and noting that K_1, K_2, K_3 are both constant, it suffices to show that $\mathrm{var}(A_1) < \infty$, $\mathrm{var}(A_2) < \infty$, $\mathrm{var}(A_3) < \infty$. Since the terms within the summations of A_2 and A_3 are i.i.d. random variables and each term is bounded, it is straightforward to see that $\mathrm{var}(A_2) < \infty$ and $\mathrm{var}(A_3) < \infty$ hold.

For $\mathrm{var}(A_1)$, using the law of total variance, we can decompose $\mathrm{var}(A_1)$ as $\mathrm{var}(A_1) = E\{\mathrm{var}(A_1 \mid X)\} + \mathrm{var}\{E(A_1 \mid X)\}$, where $X = (X_1, \ldots, X_n)^{\mathrm{T}}$. We first show that $E\{\mathrm{var}(A_1 \mid X)\} < \infty$.

In fact,

$$
\begin{aligned}
\mathrm{var}(A_1 \mid X) = \frac{K_1^2}{n}\mathrm{var}\Big\{\sum_{i=1}^{n} F_W(Y_i \wedge Y_{N(i)}) \mid X\Big\} &= \frac{K_1^2}{n}\sum_{i=1}^{n}\mathrm{var}\{F_W(Y_i \wedge Y_{N(i)}) \mid X\} \\
&+ \frac{K_1^2}{n}\sum_{\substack{j=N(i), i\neq N(j) \\ \text{or } i=N(j), j\neq N(i)}} \mathrm{Cov}\{F_W(Y_i \wedge Y_{N(i)}), F_W(Y_j \wedge Y_{N(j)}) \mid X\} \\
&+ \frac{K_1^2}{n}\sum_{\substack{i\neq j \\ N(i)=N(j)}} \mathrm{Cov}\{F_W(Y_i \wedge Y_{N(i)}), F_W(Y_j \wedge Y_{N(j)}) \mid X\} \\
&+ \frac{K_1^2}{n}\sum_{j=N(i), i=N(j)} \mathrm{Cov}\{F_W(Y_i \wedge Y_{N(i)}), F_W(Y_j \wedge Y_{N(j)}) \mid X\} \\
&+ \frac{K_1^2}{n}\sum_{i,j,N(i),N(j)\ \text{distinct}} \mathrm{Cov}\{F_W(Y_i \wedge Y_{N(i)}), F_W(Y_j \wedge Y_{N(j)}) \mid X\} \\
&= K_1^2(\tilde{T}_1 + \tilde{T}_2 + \tilde{T}_3 + \tilde{T}_4 + \tilde{T}_5).
\end{aligned}
$$

It follows from the independence between $Y_1, \dots, Y_n$ that $E(\tilde{T}_5) = 0$. For $\tilde{T}_1$, we have

$$
\begin{aligned}
E(\tilde{T}_1) &= \frac{1}{n}\sum_{i=1}^{n} E\Big[\mathrm{var}\{F_W(Y_i \wedge Y_{N(i)}) \mid X\}\Big] = E\Big[\mathrm{var}\{F_W(Y_1 \wedge Y_{N(1)}) \mid X\}\Big] \\
&= E\Big[\mathrm{var}\{F_W(Y_1 \wedge \tilde{Y}_1) \mid X_1\}\Big] + o(1), \qquad \text{(S1.19)}
\end{aligned}
$$

where $\tilde{Y}_1$ and Y_1 are independent random variables following the conditional distribution $F_{Y|X=X_1}(t)$. Recall that $G_X(t) = P(Y \geq t \mid X)$. The last equation in (S1.19) holds since it holds that

$$
\int_{\mathbb{R}_+} E\Big(\big|G_{X_{N(1)}}(t) - G_{X_1}(t)\big|\Big) w(t)\mathrm{d}t = O\bigg(\frac{(\log n)^{d+\beta+1+\mathbb{1}(d=1)}}{n^{1/d}}\bigg),
$$

where β is some positive number.

For $\tilde{T}_2$, we have that

$$
\begin{aligned}
E(\tilde{T}_2) &= \frac{1}{n} E\Bigg\{ \sum_{\substack{j=N(i), i\neq N(j) \\ \text{or } i=N(j), j\neq N(i)}} \mathrm{Cov}[F_W(Y_i \wedge Y_{N(i)}), F_W(Y_j \wedge Y_{N(j)}) \mid X] \Bigg\} \\
&= \frac{2}{n} E\Bigg[\sum_{j=N(i), i\neq N(j)} \mathrm{Cov}\{F_W(Y_i \wedge Y_{N(i)}), F_W(Y_j \wedge Y_{N(j)}) \mid X\} \Bigg] \\
&= 2E\Big[\mathrm{Cov}\{F_W(Y_1 \wedge Y_{N(1)}), F_W(Y_{N(1)} \wedge Y_{N(N(1))}) \mid X\}\mathbb{1}\big\{1 \neq N(N(1))\big\}\Big] \\
&= 2E\Big[\mathrm{Cov}\{F_W(Y_1 \wedge \tilde{Y}_1), F_W(\tilde{Y}_1 \wedge \tilde{Y}_1') \mid X_1\}\mathbb{1}\big\{1 \neq N(N(1))\big\}\Big] + o(1),
\end{aligned}
$$

(S1.20)

where $\tilde{Y}_1' \sim F_{Y|X=X_1}$ and the last equation holds due to the fact that $X_{N(N(1))} \to X_1$ almost surely, and thus $G_{X_{N(N(1))}}(t) \to G_{X_1}(t)$ in probability as shown in Lin and Han (2022). Similarly, one can get

$$E(\tilde{T}_4) = E\Big[\mathrm{var}\big\{F_W(Y_1 \wedge \tilde{Y}_1) \mid X_1\big\}\mathbb{1}\big\{1 = N(N(1))\big\}\Big] + o(1). \tag{S1.21}$$

For $\tilde{T}_3$, conditional on X, let $B_1 = B_1(X) \equiv \{j : j \neq 1, N(j) = N(1)\}$, i.e., the set of all indices j such that X_j and X_1 share the same nearest neighbor. Let $\pi(1)$ be the random variable such that for any $j \in B_1$, $P(\pi(1) = j) = 1/|B_1|$. Then, we have

$$\begin{aligned}
E(\tilde{T}_3) &= \frac{1}{n}E\Bigg[\sum_{\substack{i\neq j\\ N(i)=N(j)}} \mathrm{Cov}\{F_W(Y_i \wedge Y_{N(i)}), F_W(Y_j \wedge Y_{N(j)}) \mid X\}\Bigg] \\
&= \frac{1}{n}E\Bigg[\sum_{i=1}^{n} \sum_{j:j\neq i, N(j)=N(i)} \mathrm{Cov}\{F_W(Y_i \wedge Y_{N(i)}), F_W(Y_j \wedge Y_{N(j)}) \mid X\}\Bigg] \\
&= \frac{1}{n}nE\Bigg[\sum_{j\in B_1} \mathrm{Cov}\{F_W(Y_i \wedge Y_{N(i)}), F_W(Y_j \wedge Y_{N(j)}) \mid X\}\Bigg] \\
&= E\Big[|B_1|\mathrm{Cov}\{F_W(Y_1 \wedge \tilde{Y}_1), F_W(\tilde{Y}_1 \wedge \tilde{Y}_1') \mid X_1\}\Big] + o(1),
\end{aligned} \tag{S1.22}$$

By Lemma 20.6 and Theorem 20.16 in Biau and Devroye (2015), the size of $|B_1|$ is always bounded by a constant that depends on d. Hence, together with the boundedness of $F_W(t) = \int_0^t w(t)\mathrm{d}t$, we have $E(\tilde{T}_3) < \infty$. Therefore, combining

(S1.19), (S1.20), (S1.22) and (S1.21), we have $E(\tilde{T}_1) < \infty$, $E(\tilde{T}_2,) < \infty$, $E(\tilde{T}_3) < \infty$ and $E(\tilde{T}_4) < \infty$. As a result,

$$\limsup_{n\to\infty} E(\mathrm{var}[A_1 \mid X]) < \infty. \tag{S1.23}$$

On the other hand, for $\mathrm{var}\{E(A_1 \mid X)\}$, following a similar proof procedure of Lemma 3.1 in Lin and Han (2022), one can easily show that

$$\lim_{n\to\infty} \mathrm{var}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_1(X_i) - E(A_1 \mid X)\right\} = 0, \tag{S1.24}$$

where $g_1(x) = E\{F_W(Y \wedge \tilde{Y}) \mid X = x\}$ for Y and $\tilde{Y}$ independently from $F_{Y|X=x}$. Therefore, (S1.24) together with the Cauchy-Schwarz inequality, one deduces

$$\mathrm{var}\{E(A_1 \mid X)\} = \mathrm{var}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_1(X_i)\right\} + o(1) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{var}\big\{g_1(X_i)\big\} + o(1).$$

Then, it follows from the boundedness of F_W that

$$\limsup_{n\to\infty} \mathrm{var}\{E(A_1 \mid X)\} < \infty. \tag{S1.25}$$

Combining (S1.23) and (S1.25), we have $\limsup_{n\to\infty} \mathrm{var}(A_1) < \infty$. Together with $\mathrm{var}(A_2) < \infty$ and $\mathrm{var}(A_3) < \infty$, this completes the proof of Lemma 3. $\square$

This completes the whole proof of the asymptotic normality of $\hat{\xi}_n(X, T)$.

S1.3 Proof of Theorem 3

We first prove Theorem 3 (i). Define

$$f(x, y, u) = \begin{cases} \mathbb{1}(x > y), & \text{if } x \neq y, \\ u, & \text{if } x = y. \end{cases}$$

Let $U_1, \ldots, U_B$ be i.i.d. Bernoulli random variables of equal probabilities to be 0 or 1. For simplicity, let $\hat{\xi}_n^{(b)} \equiv \hat{\xi}_n^{(b)}(X, T)$ and $\hat{\xi}_n \equiv \hat{\xi}_n(X, T)$. Note that

$$(1+B)^{-1}\left[1 + \sum_{b=1}^{B} \mathbb{1}(\hat{\xi}_n^{(b)} \geq \hat{\xi}_n)\right] \geq (1+B)^{-1}\left[1 + \sum_{b=1}^{B} f(\hat{\xi}_n^{(b)}, \hat{\xi}_n, U_b)\right].$$

Let $\mathcal{E}_n$ be the σ-field generated by events that are invariant under permutations that leave $n+1, n+2, \ldots$ fixed and let $\mathcal{E} = \cap_n \mathcal{E}_n$ be the exchangeable σ-field. Since under H_0, $\hat{\xi}_n, \hat{\xi}_n^{(1)}, \ldots, \hat{\xi}_n^{(B)}$ are exchangeable for each n, it follows from de Finetti's theorem that conditional on $\mathcal{E}$, $\hat{\xi}_n, \hat{\xi}_n^{(1)}, \ldots, \hat{\xi}_n^{(B)}$ are i.i.d. random variables. Accordingly, we have that under H_0, conditional on $\mathcal{E}$,

$$(1+B)^{-1}\left[1 + \sum_{b=1}^{B} f(\hat{\xi}_n^{(b)}, \hat{\xi}_n, U_b)\right]$$

is uniformly distributed over

$$\left\{\frac{1}{1+B}, \frac{2}{1+B}, \cdots, \frac{1+B}{1+B}\right\}.$$

As a result, under H_0,

$$\begin{aligned}
&P_0\left[(1+B)^{-1}\left\{1+\sum_{b=1}^{B}\mathbb{1}\left(\hat{\xi}_n^{(b)} \geq \hat{\xi}_n\right)\right\} \leq \alpha \,\middle|\, \mathcal{E}\right] \\
&\leq P_0\left[(1+B)^{-1}\left\{1+\sum_{b=1}^{B} f(\hat{\xi}_n^{(b)}, \hat{\xi}_n, U_b)\right\} \leq \alpha \,\middle|\, \mathcal{E}\right] \\
&= \frac{\lfloor (1+B)\alpha \rfloor}{1+B} \leq \alpha.
\end{aligned}$$

Then, taking expectation over both sides yields $P_0(T_{\alpha,B}^n = 1) \leq \alpha$.

On the other hand, for the proof of Theorem 3 (ii), let $\Delta_n^{(b)} \equiv \mathbb{1}(\hat{\xi}_n^{(b)} \geq \hat{\xi}_n)$. Then

$$(1+B)^{-1}\left\{1+\sum_{b=1}^{B}\mathbb{1}(\hat{\xi}_n^{(b)} \geq \hat{\xi}_n)\right\} = (1+B)^{-1} + (1+B)^{-1}\sum_{b=1}^{B}\Delta_n^{(b)}.$$

It follows from Theorem 2.1 (i) that $\hat{\xi}_n^{(b)} \xrightarrow{\text{a.s.}} 0, \quad \hat{\xi}_n \xrightarrow{\text{a.s.}} \xi_{H_1}$, where $\xi_{H_1} > 0$ is the value of ξ under the fix alternative. It follows that $\Delta_n^{(b)} = \mathbb{1}(\hat{\xi}_n^{(b)} \geq \hat{\xi}_n) \to 0$ almost surely. Notice that $[\Delta_n^{(b)}]_{b=1}^{B_n}$ are exchangeable for each n. According to Lemma 1.1

in Patterson and Taylor (1985), we have that

$$B_n^{-1} \sum_{b=1}^{B_n} \Delta_n^{(b)} = E\left(\Delta_n^{(1)} \mid \mathcal{B}_n\right), \text{ almost surely,}$$

where $\mathcal{B}_n$ is the σ-field generated by

$$\mathcal{B}_n = \sigma\left(\sum_{b=1}^{B_n} \Delta_n^{(b)}, \sum_{b=1}^{B_{n+1}} \Delta_{n+1}^{(b)}, \ldots\right).$$

Note that $[\mathcal{B}_n]_{n=1}^{\infty}$ is decreasing and $\mathcal{B}_n \to \mathcal{B}_\infty$, where $\mathcal{B}_\infty = \bigcap_{n=1}^{\infty} \mathcal{B}_n$, $0 \leq \Delta_n^{(1)} \leq 1$, and $\Delta_n^{(1)} \to 0$ almost surely. Hence, by Lemma 2(c) in Isaac (1979), we obtain

$$E\left(\Delta_n^{(1)} \mid \mathcal{B}_n\right) \xrightarrow{\text{a.s.}} E\left(0 \mid \mathcal{B}_\infty\right) = 0,$$

which implies

$$B_n^{-1} \sum_{b=1}^{B_n} \Delta_n^{(b)} \xrightarrow{\text{a.s.}} 0.$$

Therefore, we have

$$\lim_{n\to\infty} P_1(T_{\alpha,B}^n = 1) = P_1\left[(1+B)^{-1}\left\{1 + \sum_{b=1}^{B} \mathbb{1}(\hat{\xi}_n^{(b)} \geq \hat{\xi}_n)\right\} \leq \alpha\right] = 1,$$

which completes the proof.

S1.4 Proof of Proposition 1

The proof consists of four steps.

Step 1. Simplification of population quantities under independence between X and T.

Under $H_0\colon X \perp T$, together with the assumption $C \perp (X,T)$, the random variables X, T, C are mutually independent. Hence, we have $X \perp Y$. This yields that $\widetilde{G}_X(t) = \widetilde{G}(t)$ and $G_X(t) = G(t)$ almost surely. Recall the population quantities,

$$T_{01} = \int_{\mathbb{R}_+} E\big\{G_X^2(t)\big\}\, w(t)\,\mathrm{d}t = \int_{\mathbb{R}_+} G(t)^2\, w(t)\,\mathrm{d}t = T_{02}.$$

Since $T_{01} = T_{02}$, the population measure satisfies $\xi(X,T) = 0$.

Recall $T_{01}^* = E\big\{F_W(Y_1 \wedge Y_{N(1)})\big\}$. Since $X \perp Y$, conditional on the σ-algebra $\mathcal{F}$ generated by $\{X_i\}_{i=1}^n$ and the tie-breaking variables, $Y_1, \ldots, Y_n$ are still i.i.d. with survival function $G(t)$. The nearest-neighbour graph $\mathcal{G}_n$ is a deterministic function of $\mathcal{F}$, so conditional on $\mathcal{F}$, Y_i and $Y_{N(i)}$ are independent (since $i \neq N(i)$). Therefore,

$$E\big\{F_W(Y_1 \wedge Y_{N(1)}) \mid \mathcal{F}\big\} = E\big\{F_W(Y_1 \wedge Y_2)\big\} = \int_{\mathbb{R}_+} G(t)^2\, w(t)\,\mathrm{d}t = T_{02},$$

where the second equality uses the independence of Y_1, Y_2. By the tower property, $T_{01}^* = T_{02}$ for all n. Hence, $\xi_n^*(X,T) = (T_{01}^* - T_{02})/(T_{03} - T_{02}) = 0$.

Step 2. Simplification of the asymptotic expansion.

Recall from (S1.10) that

$$\sqrt{n}\,\hat{\xi}_n(X,T) = K_1\sqrt{n}(T_{n1} - T_{01}^*) + K_2\sqrt{n}(T_{n2} - T_{02}) + K_3\sqrt{n}(T_{n3} - T_{03}) + o_p(1),$$

where $K_1 = (T_{03} - T_{02})^{-1}$, $K_2 = (T_{01} - T_{03})/(T_{03} - T_{02})^2$ and $K_3 = (T_{02} - T_{01})/(T_{03} - T_{02})^2$. Under H_0, since $T_{01} = T_{02}$, we obtain $K_1 = D^{-1}$, $K_2 = -D^{-1}$ and $K_3 = 0$. So the Kaplan-Meier martingale term involving $M(Y_i, \delta_i, t)$ in (S1.17) vanishes under H_0. Substituting into the expansion, one has

$$\sqrt{n}\,\hat{\xi}_n(X,T) = \frac{1}{D}\sqrt{n}(T_{n1} - T_{02}) - \frac{1}{D}\sqrt{n}(T_{n2} - T_{02}) + o_p(1) = \frac{\sqrt{n}(T_{n1} - T_{n2})}{D} + o_p(1).$$

Then, using (S1.11) and (S1.12), it follows that

$$\sqrt{n}(T_{n1} - T_{02}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big\{F_W(Y_i \wedge Y_{N(i)}) - T_{02}\big\},$$
$$\sqrt{n}(T_{n2} - T_{02}) = \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\big\{\phi(Y_i) - T_{02}\big\} + o_p(1),$$

where $\phi(y) = \int_0^y G(t)\,w(t)\,\mathrm{d}t$. Therefore,

$$\sqrt{n}\,\hat{\xi}_n(X,T) = \frac{1}{D\sqrt{n}}\sum_{i=1}^{n}\big\{F_W(Y_i \wedge Y_{N(i)}) - 2\phi(Y_i) + T_{02}\big\} + o_p(1). \qquad \text{(S1.26)}$$

Define $\zeta_i = F_W(Y_i \wedge Y_{N(i)}) - 2\phi(Y_i)$. Under H_0,

$$E(\zeta_i) = E\big\{F_W(Y_1 \wedge Y_2)\big\} - 2\,E\big\{\phi(Y_1)\big\} = T_{02} - 2T_{02} = -T_{02}.$$

Hence $\zeta_i + T_{02}$ is centred, and (S1.26) becomes

$$\sqrt{n}\,\hat{\xi}_n(X,T) = \frac{1}{D\sqrt{n}} \sum_{i=1}^{n} \big(\zeta_i - E\zeta_i\big) + o_p(1).$$

The asymptotic variance is therefore

$$\sigma_0^2 = \frac{1}{D^2} \lim_{n\to\infty} \frac{1}{n}\,\mathrm{var}\bigg(\sum_{i=1}^{n} \zeta_i\bigg).$$

Step 3. Variance decomposition conditioning on the nearest-neighbour graph.

Let $\mathcal{G}_n$ denote the associated directed nearest neighbor graph, i.e., $\mathcal{G}_n$ has vertex set $\{1,\cdots,n\}$ and contains a directed edge from i to j (i.e. $i \to j$) whenever X_j is a nearest neighbor of X_i. We apply the law of total variance, conditioning on the nearest-neighbour graph $\mathcal{G}_n$ (equivalently, on $X = (X_1,\ldots,X_n)^\top$); that is

$$\frac{1}{n}\,\mathrm{var}\bigg(\sum_{i=1}^{n} \zeta_i\bigg) = \frac{1}{n}\,E\bigg[\mathrm{var}\bigg(\sum_{i=1}^{n} \zeta_i \,\bigg|\, \mathcal{G}_n\bigg)\bigg] + \frac{1}{n}\,\mathrm{var}\bigg(E\bigg[\sum_{i=1}^{n} \zeta_i \,\bigg|\, \mathcal{G}_n\bigg]\bigg). \tag{S1.27}$$

The second term of (S1.27) vanishes because under H_0, we have for every i and every realisation of $\mathcal{G}_n$:

$$E(\zeta_i \mid \mathcal{G}_n) = E\big\{F_W(Y_1 \wedge Y_2)\big\} - 2\,E\big\{\phi(Y_1)\big\} = -T_{02},$$

which is a constant not depending on $\mathcal{G}_n$. For the first term of (S1.27), we expand the conditional variance. Specifically, conditional on $\mathcal{G}_n$, the $Y_1, \ldots, Y_n$ are i.i.d. and

$$\operatorname{var}\left(\sum_{i=1}^{n} \zeta_i \,\middle|\, \mathcal{G}_n\right) = \sum_{i=1}^{n} \operatorname{var}(\zeta_i \mid \mathcal{G}_n) + \sum_{i \neq k} \operatorname{Cov}(\zeta_i, \zeta_k \mid \mathcal{G}_n). \tag{S1.28}$$

For the first term, since Y_i and $Y_{N(i)}$ are independent conditional on $\mathcal{G}_n$, both distributed as $G(t)$, the individual variances are all equal: $\operatorname{var}(\zeta_i \mid \mathcal{G}_n) = v_0 \quad \text{for all } i$, where v_0 is a constant depending only on $G(t)$ and $w(t)$, to be computed in Step 4.

In addition, for the covariance terms of (S1.28), $\operatorname{Cov}(\zeta_i, \zeta_k \mid \mathcal{G}_n) \neq 0$ only when the index sets $\{i, N(i)\}$ and $\{k, N(k)\}$ overlap. We classify the non-trivial overlaps into three mutually exclusive types:

Type 1 (directed-edge overlap, non-mutual): $k = N(i)$ and $i \neq N(k)$, *or* $i = N(k)$ and $k \neq N(i)$. In either case the edges $(i, N(i))$ and $(k, N(k))$ share exactly one vertex. The covariance equals a constant γ_1.

Type 2 (shared target): $N(i) = N(k) = j$ for some $j \notin \{i, k\}$, with $k \neq N(i)$ and $i \neq N(k)$. The two edges share the common target vertex j. The covariance equals a constant γ_2.

Type 3 (mutual nearest neighbours): $k = N(i)$ and $i = N(k)$. The two edges (i,k) and (k,i) share both vertices. The covariance equals a constant γ_3.

Let $\mathcal{E}(\mathcal{G}_n)$ denote the edge set of $\mathcal{G}_n$ and $n_1(\mathcal{G}_n)$, $n_2(\mathcal{G}_n)$, $n_3(\mathcal{G}_n)$ denote the number of ordered pairs $(i,k) \in \mathcal{E}(\mathcal{G}_n)$ of each type above. Then (S1.28) becomes

$$\frac{1}{n}\operatorname{var}\left(\sum_{i=1}^{n} \zeta_i \,\middle|\, \mathcal{G}_n\right) = v_0 + \frac{n_1(\mathcal{G}_n)}{n}\gamma_1 + \frac{n_2(\mathcal{G}_n)}{n}\gamma_2 + \frac{n_3(\mathcal{G}_n)}{n}\gamma_3.$$

Taking expectations and passing to the limit:

$$\sigma_0^2 = \frac{v_0 + \gamma_1 l_1 + \gamma_2 l_2 + \gamma_3 l_3}{D^2}, \tag{S1.29}$$

where $l_k = \lim_{n\to\infty} E[n_k(\mathcal{G}_n)]/n, k = 1, 2, 3$, to be given in Step 4. These limits exist by Shi et al. (2024) and Henze (1987), and these positive constants only depend on the dimension d.

Step 4. Computation of $v_0, \gamma_1, \gamma_2, \gamma_3, l_1, l_2, l_3$.

Throughout this step, Y_1, Y_2, Y_3 are i.i.d. with survival function $G(t)$. We first show that

$$\operatorname{Cov}\bigl(F_W(Y_1 \wedge Y_2),\ \phi(Y_1)\bigr) = \sigma_\phi^2. \tag{S1.30}$$

Specifically, since Y_1 and Y_2 are independent, we compute

$$E\big\{F_W(Y_1 \wedge Y_2)\,\phi(Y_1)\big\} = \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} G\big(\max(s,t)\big)\,G(s)\,G(t)\,w(s)\,w(t)\,\mathrm{d}s\,\mathrm{d}t. \tag{S1.31}$$

On the other hand,

$$\begin{aligned} E\big\{\phi(Y_1)^2\big\} &= E\Big\{\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \mathbb{1}(Y_1 \geq t)\mathbb{1}(Y_1 \geq s)G(s)\,G(t)\,w(s)\,w(t)\,\mathrm{d}s\,\mathrm{d}t\Big\} \\ &= \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} G\big(\max(s,t)\big)\,G(s)\,G(t)\,w(s)\,w(t)\,\mathrm{d}s\,\mathrm{d}t, \end{aligned}$$

which equals (S1.31). Hence, according to $E\{F_W(Y_1 \wedge Y_2)\} = T_{02}$ and $E\{\phi(Y_1)\} = T_{02}$, we obtain

$$\mathrm{Cov}\big(F_W(Y_1 \wedge Y_2),\ \phi(Y_1)\big) = E\{\phi(Y_1)^2\} - T_{02}^2 = \sigma_\phi^2.$$

Next, we compute v_0.

$$\begin{aligned} v_0 &= \mathrm{var}\big(F_W(Y_1 \wedge Y_2) - 2\phi(Y_1)\big) \\ &= \mathrm{var}\big(F_W(Y_1 \wedge Y_2)\big) + 4\,\mathrm{var}\big(\phi(Y_1)\big) - 4\,\mathrm{Cov}\big(F_W(Y_1 \wedge Y_2),\ \phi(Y_1)\big) \\ &= \sigma_F^2 + 4\sigma_\phi^2 - 4\sigma_\phi^2 = \sigma_F^2. \end{aligned}$$

For γ_1, by (S1.30), Type 1 pairs (e.g., $k = N(i)$, $N(k) \neq i$) lead to

$$\gamma_1 = \mathrm{Cov}\big(F_W(Y_1 \wedge Y_2) - 2\phi(Y_1),\ F_W(Y_2 \wedge Y_3) - 2\phi(Y_2)\big),$$
$$= \mathrm{Cov}\big(F_W(Y_1 \wedge Y_2),\ F_W(Y_2 \wedge Y_3)\big) - 2\,\mathrm{Cov}\big(F_W(Y_1 \wedge Y_2),\ \phi(Y_2)\big)$$

Using $Y_1 \perp Y_3$ and the same calculation as (S1.31),

$$E\big\{F_W(Y_1 \wedge Y_2)\, F_W(Y_2 \wedge Y_3)\big\} = \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} G(s)\, G\big(\max(s,t)\big)\, G(t)\, w(s)\, w(t)\, \mathrm{d}s\, \mathrm{d}t = E\big\{\phi(Y_1)^2\big\}.$$

Thus, $\mathrm{Cov}\big(F_W(Y_1 \wedge Y_2),\ F_W(Y_2 \wedge Y_3)\big) = E\{\phi(Y_1)^2\} - T_{02}^2 = \sigma_\phi^2$. Combining

$$\gamma_1 = \sigma_\phi^2 - 2\sigma_\phi^2 = -\sigma_\phi^2.$$

One can verify that the "reverse" Type 1 case ($i = N(k)$, $k \neq N(i)$) yields the same covariance $\gamma_1 = -\sigma_\phi^2$ by an identical argument.

For γ_2, similar to the calculation of γ_1, Type 2 pairs ($N(i) = N(k) = j$) give

$$\gamma_2 = \mathrm{Cov}\big(F_W(Y_1 \wedge Y_3) - 2\phi(Y_1),\ F_W(Y_2 \wedge Y_3) - 2\phi(Y_2)\big),$$
$$= \mathrm{Cov}\big(F_W(Y_1 \wedge Y_3),\ F_W(Y_2 \wedge Y_3)\big) = \sigma_\phi^2 = -\gamma_1.$$

For γ_3, Type 3 pairs (mutual nearest neighbours, edges (i,k) and (k,i)) give

$$\gamma_3 = \mathrm{Cov}\big(F_W(Y_1 \wedge Y_2) - 2\phi(Y_1),\ F_W(Y_1 \wedge Y_2) - 2\phi(Y_2)\big),$$

where Y_1, Y_2 are i.i.d. Expanding:

$$\begin{aligned}\gamma_3 &= \mathrm{var}\big(F_W(Y_1 \wedge Y_2)\big) - 2\,\mathrm{Cov}\big(F_W(Y_1 \wedge Y_2),\ \phi(Y_2)\big) - 2\,\mathrm{Cov}\big(\phi(Y_1),\ F_W(Y_1 \wedge Y_2)\big)\\ &= \sigma_F^2 - 4\sigma_\phi^2.\end{aligned}$$

Now, we come to give limits l_1, l_2, l_3. For l_2, by Lemma 3.3 in Shi et al. (2024),

$$l_2 = \lim_{n\to\infty} E[n_2(\mathcal{G}_n)]/n = \lim_{n\to\infty} E\left(n^{-1} \sum_{\substack{(i,j,k)\text{ distinct}\\ i\to k, j\to k \in \mathcal{E}(\mathcal{G}_n)}} 1\right) = \mathfrak{o}_d,$$

where

$$\begin{aligned}\mathfrak{o}_d &= \int_{\Gamma_{d;2}} \exp\left[-\lambda\left\{B(\boldsymbol{x}_1, \|\boldsymbol{x}_1\|) \cup B(\boldsymbol{x}_2, \|\boldsymbol{x}_2\|)\right\}\right] \mathrm{d}(\boldsymbol{x}_1, \boldsymbol{x}_2),\\ \Gamma_{d;2} &= \left\{(\boldsymbol{x}_1, \boldsymbol{x}_2) \in (\mathbb{R}^d)^2 : \max(\|\boldsymbol{x}_1\|, \|\boldsymbol{x}_2\|) < \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|\right\}.\end{aligned}$$

For l_3, it follows from Lemma 3.2 in Shi et al. (2024) and Theorem 2 in Devroye

(1988) that

$$l_3 = \lim_{n\to\infty} E[n_3(\mathcal{G}_n)]/n = \lim_{n\to\infty} E\left(n^{-1}\sum_{\substack{(i,j)\text{ distinct}\\ i\to j, j\to i\in\mathcal{E}(\mathcal{G}_n)}} 1\right) = \mathfrak{q}_d,$$

where

$$\mathfrak{q}_d = \left\{2 - I_{3/4}\left(\frac{d+1}{2}, \frac{1}{2}\right)\right\}^{-1}, \quad I_x(a,b) = \frac{\int_0^x t^{a-1}(1-t)^{b-1}\mathrm{d}t}{\int_0^1 t^{a-1}(1-t)^{b-1}\mathrm{d}t}.$$

At last, for l_1, we have

$$\begin{aligned}
l_1 &= \lim_{n\to\infty} E[n_1(\mathcal{G}_n)]/n\\
&= \lim_{n\to\infty} E\left(n^{-1}\sum_{\substack{(i,j,k)\text{ distinct}\\ i\to j, j\to k\in\mathcal{E}(\mathcal{G}_n)}} 1 + n^{-1}\sum_{\substack{(i,j,k)\text{ distinct}\\ i\to k, j\to i\in\mathcal{E}(\mathcal{G}_n)}} 1\right)\\
&= \lim_{n\to\infty} E\left(n^{-1}\sum_{\substack{(i,j)\text{ distinct}\\ i\to j\in\mathcal{E}(\mathcal{G}_n), j\to i\notin\mathcal{E}(\mathcal{G}_n)}} 1 + n^{-1}\sum_{\substack{(i,j)\text{ distinct}\\ i\to j\notin\mathcal{E}(\mathcal{G}_n), j\to i\in\mathcal{E}(\mathcal{G}_n)}} 1\right)\\
&= \lim_{n\to\infty} E\left\{\left(1 - n^{-1}\sum_{\substack{(i,j)\text{ distinct}\\ i\to j, j\to i\in\mathcal{E}(\mathcal{G}_n)}} 1\right) + \left(1 - n^{-1}\sum_{\substack{(i,j)\text{ distinct}\\ i\to j, j\to i\in\mathcal{E}(\mathcal{G}_n)}} 1\right)\right\}\\
&\longrightarrow \big(2 - 2\mathfrak{q}_d\big),
\end{aligned}$$

In summary, substituting $v_0 = \sigma_F^2$, $\gamma_1 = -\sigma_\phi^2$, $\gamma_2 = \sigma_\phi^2$, $\gamma_3 = \sigma_F^2 - 4\sigma_\phi^2$, $l_1 = 2 - 2\mathfrak{q}_d$,

$l_2 = \mathfrak{o}_d$, $l_3 = \mathfrak{q}_d$ into (S1.29),

$$\sigma_0^2\, D^2 = \sigma_F^2 + l_1(-\sigma_\phi^2) + l_2\,\sigma_\phi^2 + l_3(\sigma_F^2 - 4\sigma_\phi^2) = (1+l_3)\,\sigma_F^2 + (l_2 - l_1 - 4l_3)\,\sigma_\phi^2$$
$$= (1+\mathfrak{q}_d)\sigma_F^2 + (\mathfrak{o}_d - 2 - 2\mathfrak{q}_d)\sigma_\phi^2.$$

Therefore,

$$\sigma_0^2 = \frac{(1+\mathfrak{q}_d)\sigma_F^2 + (\mathfrak{o}_d - 2 - 2\mathfrak{q}_d)\sigma_\phi^2}{D^2}\,.$$

This completes the proof.

S1.5 Proof of Corollary 1

First, it follows from (S1.12) and (S1.15) that $\hat{D} = T_{n3} - T_{n2} \to D$ almost surely. For $\hat{\sigma}_\phi^2 = n^{-1}\sum_{i=1}^n \hat{\phi}(Y_i)^2 - T_{n2}^2$, by the continuous mapping theorem, $T_{n2}^2 \to T_{02}^2 = \{E(\phi(Y_1))\}^2$ almost surely. On the other hand, since $C_w = \int_{\mathbb{R}_+} w(t)\,\mathrm{d}t < \infty$, by the Glivenko-Cantelli theorem, we have that for any $y \geq 0$,

$$\big|\hat{\phi}(y) - \phi(y)\big| = \left|\int_0^y \big(G_n(t) - G(t)\big)\, w(t)\,\mathrm{d}t\right| \leq C_w \sup_{t\geq 0}\big|G_n(t) - G(t)\big| \to 0,$$

almost surely. Since $0 \le \phi(y) \le C_w$ for all $y \ge 0$, and consequently $E[\phi(Y_1)^2] \le C_w^2 < \infty$. By the strong law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n}\phi(Y_i)^2 \to E\big[\phi(Y_1)^2\big], \text{ almost surely.} \tag{S1.32}$$

Denote $A_n = n^{-1}\sum_{i=1}^{n}\hat{\phi}(Y_i)^2 - n^{-1}\sum_{i=1}^{n}\phi(Y_i)^2$. Then we have

$$|A_n| = \left|\frac{1}{n}\sum_{i=1}^{n}\big[\hat{\phi}(Y_i)^2 - \phi(Y_i)^2\big]\right| \le \frac{2C_w}{n}\sum_{i=1}^{n}\big|\hat{\phi}(Y_i) - \phi(Y_i)\big| \le 2C_w^2 \sup_{t\ge 0}\big|G_n(t) - G(t)\big| \to 0, \tag{S1.33}$$

almost surely. It follows from (S1.33) and (S1.32) that $n^{-1}\sum_{i=1}^{n}\hat{\phi}(Y_i)^2 \to E(\phi(Y_1)^2)$ almost surely. Therefore, $\hat{\sigma}_\phi^2 \to \sigma_\phi^2$ almost surely.

At last, for the consistency of $\hat{\sigma}_F^2$, define the symmetric kernel function $h : \mathbb{R}^2 \to \mathbb{R}$ as

$$h(y_1, y_2) = F_W(y_1 \wedge y_2)^2 = \left(\int_0^{y_1 \wedge y_2} w(t)\, \mathrm{d}t\right)^2.$$

Then the first term $U_n = \binom{n}{2}^{-1}\sum_{i<j} h(Y_i, Y_j)$ is a U-statistic of order 2 with expectation

$$\theta := E[h(Y_1, Y_2)] = E\big[F_W(Y_1 \wedge Y_2)^2\big].$$

It follows from $F_W(y) \le C_w$ that $0 \le h(y_1, y_2) \le C_w^2 < \infty$ for all y_1, y_2. By the strong law of large numbers for U-statistics, we have $U_n \to \theta$ almost surely. As a result, we have $\hat{\sigma}_F^2 \to \sigma_F^2$ almost surely. The proof is complete.

S2. Additional simulation studies for testing non-additive structures in Section 8.3

To further examine the performance of our CRC test, we consider commonly used models in survival analysis. Specifically, DGP10 consists of nonlinear Cox proportional hazards models, and DGP11 consists of accelerated failure time models (AFT) with heterogeneous errors.

DGP10: (Nonlinear Cox models): T follows the hazard function $\lambda(t|X) = \lambda_0(t)\exp(g(X))$, where $g(X)$ is specified below for each case:

(a) $g(X) = 3\sin(8\pi X)$, $X \sim U(0, 2)$;

(b) $g(X) = 6(X - \lfloor X \rfloor) + 1$, $X \sim U(-3, 3)$;

(c) $g(X) = 3\sin(2\pi e^X)$, $X \sim U(0, 2)$.

The baseline hazard function $\lambda_0(t)$ is assumed to be the Weibull hazard function with shape parameter equals 2 (i.e., $\lambda_0(t) = 2t$). The censoring variable C is generated from $C \sim \text{Exp}(\lambda_c)$.

DGP11: (Heterogeneous AFT models): $\log(T) = 1 + \sigma(X)\epsilon$, where X and ϵ are i.i.d $N(0, 1)$ and $\sigma(X)$ is specified below for each case:

(a) $\sigma(X) = 0.3 + 0.5|X|$;

(b) $\sigma(X) = 1 + \exp\{\cos(2\pi X)\}$;

(c) $\sigma(X) = 1 + \exp\{|\cos(2\pi X)|\} + X^2$.

The censoring variable C is generated from $C \sim \mathrm{Exp}(\lambda_c)$.

Figure S.4 displays the empirical power for all cases in DGPs 10 and 11, separately for censoring rates of 45% and 65%. The corresponding empirical sizes at levels 0.01 and 0.05 are summarized in Tables S.3 and S.4, respectively. These results show that the permutation test based on CRC can effectively detect such complex dependence structures and outperforms the competing methods, while maintaining stable size performance.

S3. Additional simulation and real data results

We present additional simulation results on asymptotic normality in Figures S.1–2 in Section 8.1, and the empirical type-I error rates with 45% and 65% censoring in Tables S.1–2 in Section 8.3. Figure S.3 rejection rates of all tests for cosine function with various frequency. Figure S.5 displays the FDR-adjusted p-values for the ADNI data, and Figure S.6 shows the complete list of the top 100 most significant signals obtained by the CRC method with $w(t)$ taken to be the standard normal distribution in Section 9.

References

Azadkia, M. and Chatterjee, S. (2021). A simple measure of conditional dependence. *Ann. Statist.* **49**, 3070–3102.

Biau, G. and Devroye, L. (2015). *Lectures on the Nearest Neighbor Method,* volume 246. Springer.

Bitouzé, D., Laurent, B. and Massart, P. (1999). A Dvoretzky-Kiefer-Wolfowitz type inequality for the Kaplan-Meier estimator. *Ann. Inst. Henri Poincaré (B) Prob. Stat.* **35**, 735–763.

Chatterjee, S. (2008). A new method of normal approximation. *Ann. Prob.* **36**, 1584–1610.

Chatterjee, S. (2021). A new coefficient of correlation. *J. Am. Stat. Assoc.* **116**, 2009–2022.

Devroye, L. (1988). The expected size of some graphs in computational geometry. *Comput. Math. Appl.* **15**, 53–64.

Fleming, T. R. and Harrington, D. P. (2013). *Counting Processes and Survival Analysis* (Vol. 625), John Wiley & Sons.

Henze, N. (1987). On the fraction of random points with specified nearest-neighbour interrelations and degree of attraction. *Adv. in Appl. Probab.* **19**, 873–895.

Isaac, R. (1979). Markov-dependent-fields and conditional expectations. *Ann. Probab.* **7**, 1088–1091.

Li, J., Zheng, Q., Peng, L. and Huang, Z. (2016). A survival impact index and ultrahigh-dimensional model-free screening with survival outcomes. *Biometrics* **72**, 1145–1154.

Lin, Z. and Han, F. (2022). Limit theorems of Chatterjee's rank correlation. https://arxiv.org/abs/2204.08031

Patterson, R. F. and Taylor, R. L. (1985). Strong laws of large numbers for triangular arrays of exchangeable random variables. *Stochastic Anal. Appl.* **3**, 171–187.

Shi, H., Drton, M. and Han, F. (2024). On Azadkia-Chatterjee's conditional dependence coefficient. *Bernoulli*, **30**, 851–877.

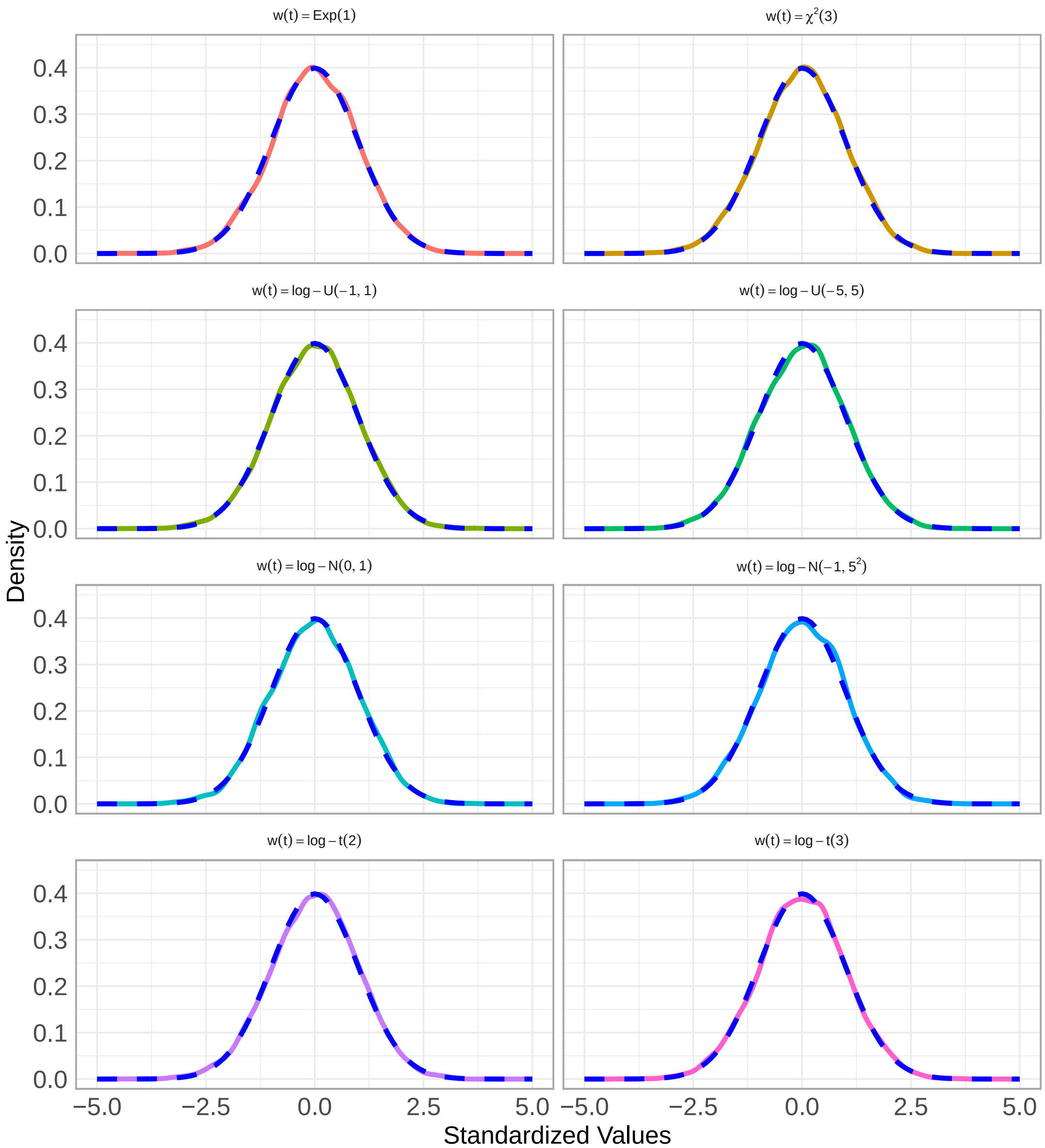

Figure S.1: Density of $N(0,1)$ (dashed lines) and empirical densities (solid lines) of $\hat{\xi}_n(X,T)$ for different $w(t)$ in DGP1 with $\rho = 0.4$, $n = 1000$ and 10000 replications.

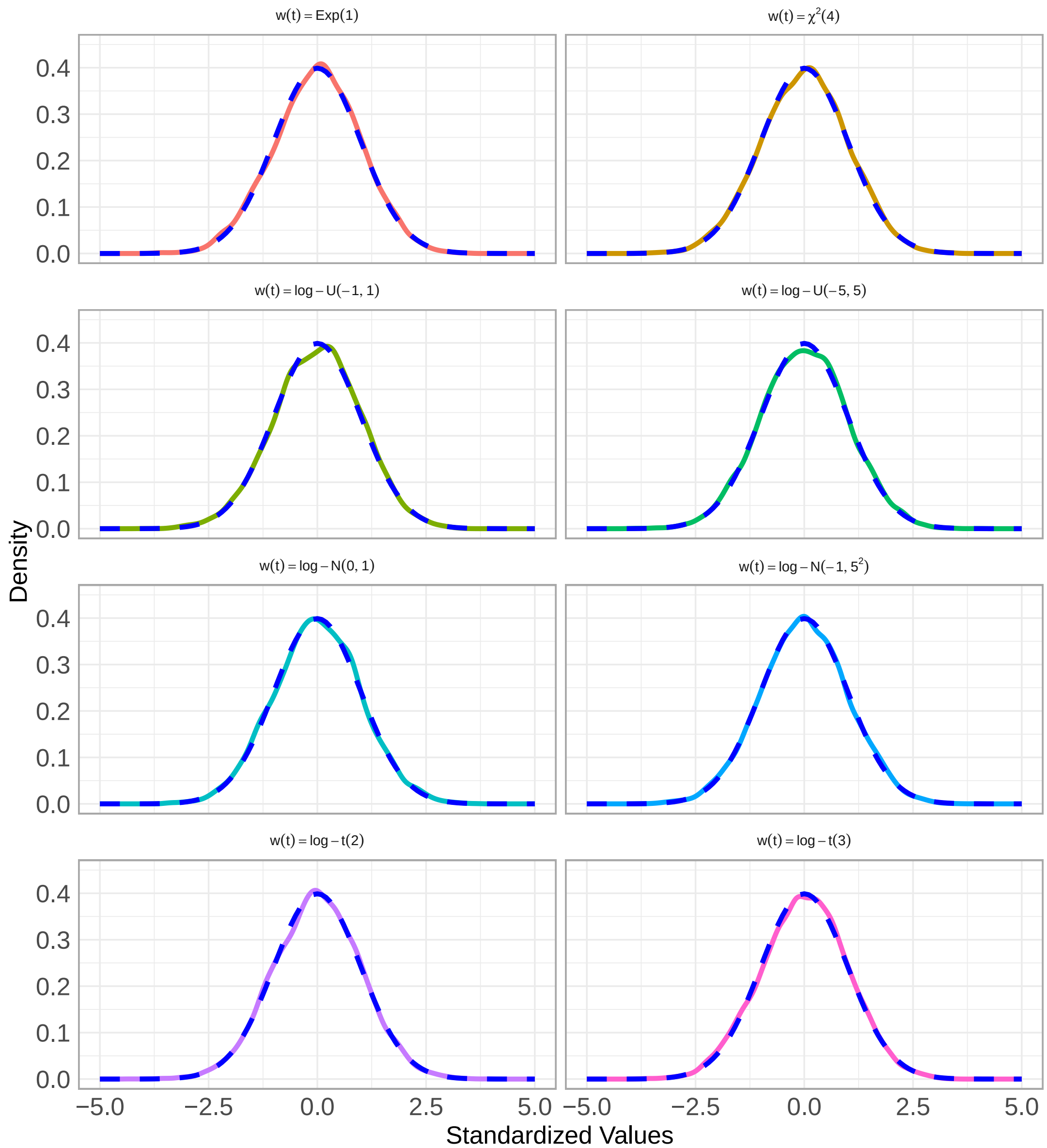

Figure S.2: Density of $N(0,1)$ (dashed lines) and empirical densities (solid lines) of $\hat{\xi}_n(X,T)$ for different $w(t)$ in DGP2 with $\epsilon \sim \chi_2^2$, $n = 1000$ and 10000 replications.

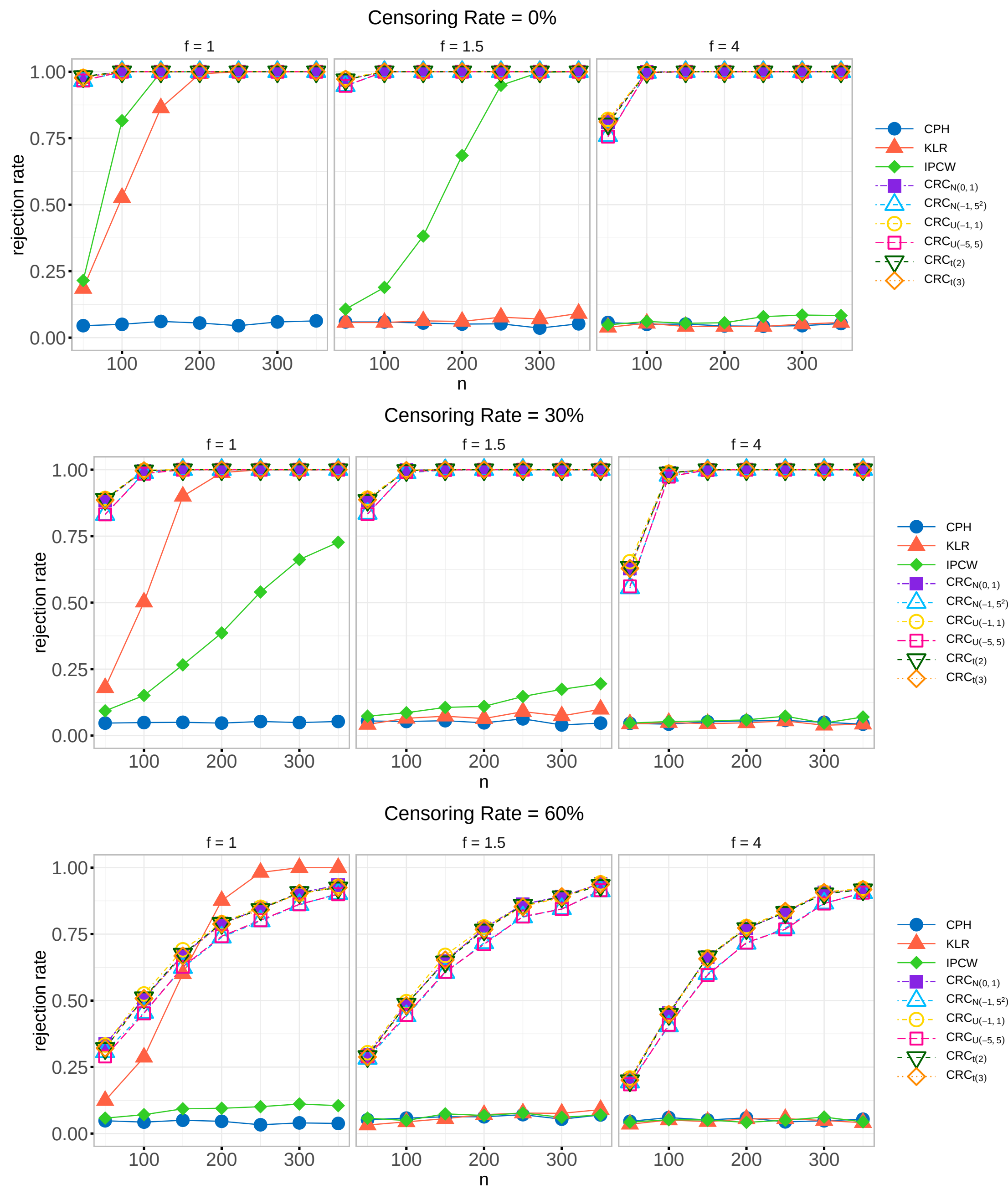

Figure S.3: Rejection rates of all tests for cosine function with the frequency f=1, 1.5, and 4, and the censoring rates $0\%, 30\%$, and 60%, respectively. The dashed lines represent our CRC tests with various $w(t)$.

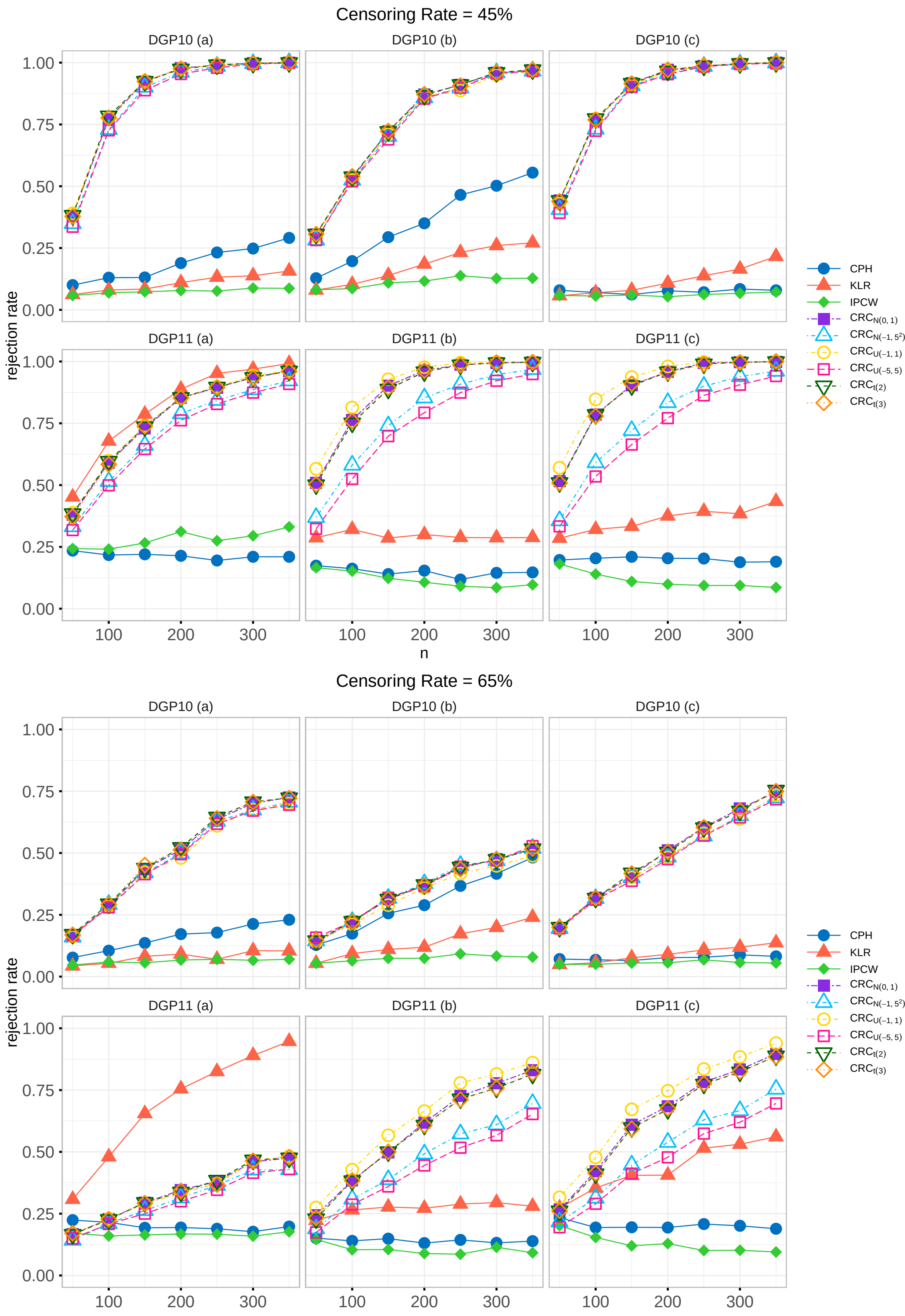

Figure S.4: Rejection rates of all tests in DGPs 10–11 with the censoring rates 45% and 65%, respectively. The dashed lines represent our tests with various $w(t)$.

Table S.1: Empirical type-I error rate with 45% censoring for DGPs 4–9.

						The proposed CRC test with various $w(t)$					
α	DGP	n	CPH	KLR	IPCW	$\mathrm{CRC}_{N(0,1)}$	$\mathrm{CRC}_{N(-1,5^2)}$	$\mathrm{CRC}_{U(-1,1)}$	$\mathrm{CRC}_{U(-5,5)}$	$\mathrm{CRC}_{t(2)}$	$\mathrm{CRC}_{t(3)}$
0.01	4	50	0.014	0.008	0.007	0.013	0.012	0.014	0.016	0.010	0.011
		150	0.011	0.006	0.010	0.009	0.009	0.010	0.008	0.010	0.006
	5	50	0.006	0.004	0.007	0.012	0.010	0.013	0.011	0.012	0.013
		150	0.015	0.016	0.013	0.008	0.006	0.007	0.005	0.007	0.007
	6	50	0.009	0.011	0.016	0.007	0.010	0.007	0.009	0.006	0.007
		150	0.007	0.007	0.004	0.012	0.012	0.011	0.012	0.013	0.010
	7	50	0.011	0.009	0.007	0.009	0.007	0.006	0.009	0.006	0.008
		150	0.006	0.006	0.008	0.013	0.011	0.017	0.009	0.014	0.017
	8	50	0.006	0.009	0.009	0.015	0.015	0.011	0.015	0.013	0.014
		150	0.009	0.007	0.012	0.014	0.007	0.012	0.010	0.012	0.011
	9	50	0.018	0.007	0.008	0.012	0.013	0.009	0.012	0.013	0.010
		150	0.007	0.007	0.009	0.007	0.008	0.009	0.009	0.009	0.012
0.05	4	50	0.060	0.040	0.043	0.053	0.056	0.059	0.056	0.054	0.055
		150	0.058	0.049	0.047	0.048	0.041	0.041	0.041	0.044	0.044
	5	50	0.050	0.040	0.051	0.055	0.054	0.053	0.051	0.053	0.055
		150	0.058	0.052	0.052	0.044	0.049	0.045	0.045	0.046	0.046
	6	50	0.050	0.049	0.064	0.045	0.053	0.047	0.051	0.049	0.046
		150	0.043	0.041	0.041	0.061	0.058	0.063	0.058	0.059	0.061
	7	50	0.064	0.036	0.046	0.051	0.041	0.050	0.046	0.053	0.051
		150	0.044	0.050	0.042	0.048	0.050	0.045	0.049	0.049	0.053
	8	50	0.058	0.048	0.056	0.053	0.054	0.059	0.055	0.053	0.053
		150	0.052	0.047	0.057	0.059	0.049	0.059	0.054	0.055	0.061
	9	50	0.050	0.038	0.049	0.046	0.043	0.046	0.038	0.048	0.046
		150	0.042	0.039	0.047	0.061	0.048	0.063	0.052	0.061	0.061

Table S.2: Empirical type-I error rate with 65% censoring for DGPs 4–9.

						The proposed CRC test with various $w(t)$					
α	DGP	n	CPH	KLR	IPCW	$\text{CRC}_{N(0,1)}$	$\text{CRC}_{N(-1,5^2)}$	$\text{CRC}_{U(-1,1)}$	$\text{CRC}_{U(-5,5)}$	$\text{CRC}_{t(2)}$	$\text{CRC}_{t(3)}$
0.01	4	50	0.013	0.005	0.012	0.012	0.009	0.013	0.009	0.010	0.012
		150	0.010	0.006	0.014	0.012	0.011	0.011	0.011	0.008	0.009
	5	50	0.014	0.006	0.013	0.007	0.009	0.008	0.010	0.008	0.009
		150	0.015	0.014	0.010	0.011	0.012	0.012	0.011	0.010	0.012
	6	50	0.016	0.003	0.008	0.011	0.011	0.010	0.009	0.009	0.009
		150	0.010	0.010	0.004	0.008	0.009	0.008	0.009	0.009	0.010
	7	50	0.010	0.012	0.008	0.006	0.002	0.004	0.004	0.012	0.010
		150	0.010	0.010	0.006	0.012	0.006	0.008	0.014	0.010	0.014
	8	50	0.016	0.005	0.008	0.012	0.013	0.007	0.008	0.007	0.010
		150	0.006	0.006	0.007	0.003	0.007	0.008	0.009	0.005	0.006
	9	50	0.017	0.006	0.011	0.006	0.008	0.010	0.006	0.007	0.008
		150	0.012	0.005	0.007	0.013	0.013	0.011	0.013	0.014	0.011
0.05	4	50	0.067	0.052	0.047	0.056	0.053	0.047	0.049	0.046	0.052
		150	0.056	0.047	0.061	0.054	0.047	0.054	0.047	0.063	0.057
	5	50	0.047	0.039	0.058	0.050	0.054	0.056	0.057	0.055	0.049
		150	0.053	0.051	0.051	0.050	0.054	0.046	0.052	0.048	0.047
	6	50	0.062	0.042	0.057	0.053	0.055	0.050	0.056	0.050	0.050
		150	0.040	0.050	0.039	0.043	0.051	0.043	0.056	0.049	0.047
	7	50	0.044	0.054	0.046	0.060	0.060	0.064	0.068	0.058	0.058
		150	0.050	0.062	0.046	0.056	0.070	0.066	0.070	0.052	0.052
	8	50	0.070	0.044	0.041	0.047	0.047	0.048	0.047	0.047	0.048
		150	0.050	0.050	0.049	0.044	0.051	0.043	0.050	0.046	0.047
	9	50	0.070	0.042	0.047	0.042	0.045	0.036	0.044	0.040	0.037
		150	0.059	0.043	0.047	0.053	0.050	0.057	0.047	0.051	0.053

Table S.3: Empirical type-I error rate with 45% censoring for DGPs 10-11.

						The proposed CRC test with various $w(t)$					
α	DGP	n	CPH	KLR	IPCW	$\text{CRC}_{N(0,1)}$	$\text{CRC}_{N(-1,5^2)}$	$\text{CRC}_{U(-1,1)}$	$\text{CRC}_{U(-5,5)}$	$\text{CRC}_{t(2)}$	$\text{CRC}_{t(3)}$
0.01	10(a)	50	0.012	0.004	0.006	0.011	0.011	0.008	0.011	0.012	0.012
		150	0.017	0.011	0.016	0.011	0.011	0.012	0.012	0.012	0.009
	10(b)	50	0.013	0.006	0.008	0.011	0.009	0.008	0.013	0.007	0.010
		150	0.018	0.014	0.002	0.014	0.008	0.014	0.012	0.013	0.015
	10(c)	50	0.005	0.005	0.010	0.008	0.008	0.008	0.011	0.010	0.007
		150	0.010	0.014	0.008	0.010	0.008	0.011	0.007	0.008	0.010
	11(a)	50	0.010	0.008	0.005	0.010	0.007	0.008	0.010	0.009	0.010
		150	0.010	0.005	0.012	0.013	0.011	0.013	0.014	0.014	0.013
	11(b)	50	0.010	0.011	0.012	0.009	0.010	0.011	0.011	0.009	0.010
		150	0.010	0.008	0.005	0.014	0.014	0.014	0.014	0.013	0.010
	11(c)	50	0.007	0.008	0.006	0.007	0.011	0.007	0.011	0.010	0.008
		150	0.005	0.006	0.013	0.010	0.008	0.010	0.010	0.011	0.009
0.05	10(a)	50	0.050	0.046	0.047	0.055	0.053	0.045	0.051	0.051	0.051
		150	0.044	0.049	0.051	0.056	0.057	0.065	0.062	0.060	0.053
	10(b)	50	0.049	0.043	0.047	0.047	0.051	0.046	0.054	0.051	0.048
		150	0.054	0.053	0.052	0.035	0.032	0.036	0.037	0.037	0.036
	10(c)	50	0.061	0.050	0.041	0.054	0.063	0.058	0.059	0.058	0.056
		150	0.046	0.043	0.050	0.057	0.056	0.053	0.062	0.057	0.060
	11(a)	50	0.050	0.050	0.048	0.034	0.036	0.037	0.035	0.035	0.036
		150	0.042	0.045	0.049	0.054	0.049	0.050	0.052	0.047	0.053
	11(b)	50	0.064	0.045	0.059	0.054	0.061	0.050	0.059	0.054	0.061
		150	0.045	0.040	0.055	0.050	0.058	0.052	0.060	0.053	0.050
	11(c)	50	0.048	0.054	0.035	0.052	0.051	0.049	0.051	0.049	0.051
		150	0.063	0.059	0.043	0.057	0.060	0.059	0.059	0.057	0.056

Table S.4: Empirical type-I error rate with 65% censoring for DGPs 10-11.

						The proposed CRC test with various $w(t)$					
α	DGP	n	CPH	KLR	IPCW	$\mathrm{CRC}_{N(0,1)}$	$\mathrm{CRC}_{N(-1,5^2)}$	$\mathrm{CRC}_{U(-1,1)}$	$\mathrm{CRC}_{U(-5,5)}$	$\mathrm{CRC}_{t(2)}$	$\mathrm{CRC}_{t(3)}$
0.01	10(a)	50	0.019	0.006	0.010	0.009	0.007	0.010	0.012	0.006	0.008
		150	0.009	0.009	0.007	0.013	0.011	0.012	0.014	0.012	0.015
	10(b)	50	0.020	0.011	0.010	0.015	0.014	0.014	0.015	0.012	0.014
		150	0.005	0.007	0.013	0.005	0.004	0.004	0.003	0.007	0.004
	10(c)	50	0.013	0.002	0.005	0.014	0.013	0.013	0.015	0.011	0.010
		150	0.009	0.008	0.005	0.009	0.012	0.008	0.008	0.007	0.005
	11(a)	50	0.008	0.008	0.010	0.008	0.010	0.013	0.011	0.011	0.012
		150	0.011	0.005	0.011	0.009	0.010	0.009	0.009	0.007	0.010
	11(b)	50	0.012	0.011	0.013	0.009	0.014	0.010	0.013	0.009	0.011
		150	0.007	0.010	0.007	0.005	0.003	0.006	0.004	0.006	0.006
	11(c)	50	0.009	0.008	0.009	0.013	0.016	0.016	0.014	0.014	0.013
		150	0.016	0.009	0.011	0.013	0.011	0.010	0.009	0.013	0.010
0.05	10(a)	50	0.053	0.052	0.041	0.065	0.066	0.060	0.065	0.060	0.063
		150	0.057	0.057	0.051	0.038	0.041	0.044	0.041	0.040	0.041
	10(b)	50	0.056	0.048	0.074	0.063	0.065	0.064	0.059	0.059	0.060
		150	0.052	0.044	0.057	0.046	0.053	0.046	0.049	0.050	0.047
	10(c)	50	0.054	0.048	0.045	0.048	0.049	0.048	0.047	0.049	0.050
		150	0.056	0.054	0.042	0.033	0.040	0.034	0.040	0.035	0.030
	11(a)	50	0.058	0.044	0.049	0.052	0.048	0.053	0.058	0.053	0.050
		150	0.069	0.044	0.045	0.046	0.038	0.043	0.039	0.045	0.045
	11(b)	50	0.060	0.044	0.060	0.058	0.052	0.056	0.050	0.054	0.053
		150	0.045	0.042	0.052	0.044	0.039	0.044	0.039	0.038	0.046
	11(c)	50	0.049	0.030	0.049	0.055	0.052	0.059	0.054	0.055	0.055
		150	0.045	0.041	0.053	0.046	0.043	0.046	0.048	0.045	0.043

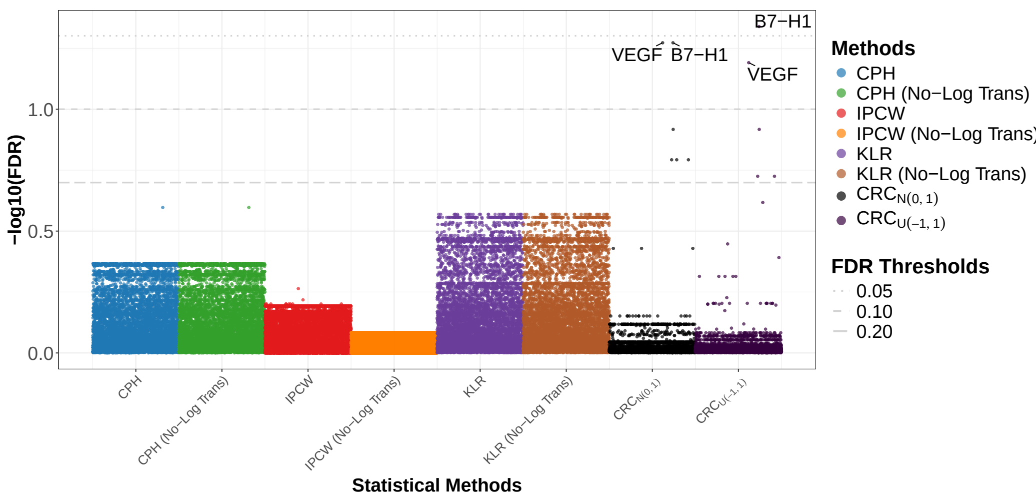

Figure S.5: FDR-adjusted p-values in the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort were analyzed using Cox Proportional Hazards (CPH), inverse-probability-of-censoring weighting (IPCW), and kernel log-rank test (KLR), with and without log-transformed T, comparing with the proposed CRC methods with choices of $w(t)$ as $N(0, 1)$ (normal) and $U(-1, 1)$ (uniform) density functions.

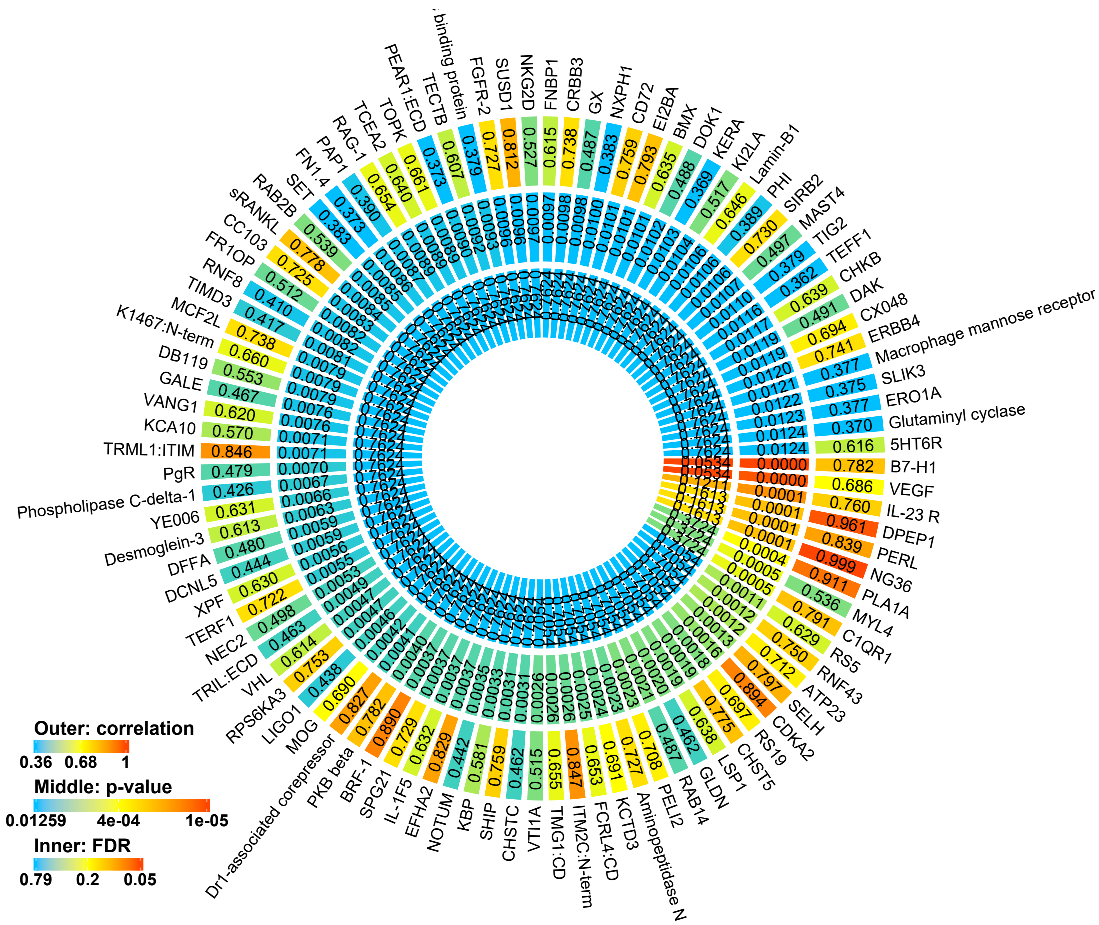

Figure S.6: Correlation coefficients, raw p-values, and false discovery rate (FDR)-adjusted p-values from the proposed CRC method (w(t) = N(0,1), analyzing associations between survival time and protein levels in the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort.