CTE$_2$: Conditional tail expectation treatment effect

Zi Miao, Fudan University

Liujun Chen, University of Science and Technology of China

Deyuan Li*, Fudan University

Supplementary Material

The supplementary material includes detailed proofs of Theorems 1–5 from the main paper, along with auxiliary explanations, lemmas, and propositions. Additionally, we present supplementary simulation results that complement the findings in the main paper. In the rest of the paper, $\lceil u \rceil$ means the smallest integer greater than or equal to u.

S1 Estimation of propensity score function

For the estimation of $\Pi(\boldsymbol{z})$, following Hirano et al. (2003) and Deuber et al. (2024), we set

$$\widehat{\Pi}(\boldsymbol{z}) := \frac{1}{1 + \exp\{-H_{h_n}(\boldsymbol{z})^T \widehat{\boldsymbol{\pi}}_{\boldsymbol{n}}\}},$$

where $H_{h_n} = (H_{h_n,j})_{j=1,2,\ldots,h_n} : R^r \to R^{h_n}$ is a vector of sieve basis functions, and

$$\widehat{\boldsymbol{\pi}}_{\boldsymbol{n}} := \arg\max_{\boldsymbol{\pi} \in R^{h_n}} \sum_{i=1}^{n} D_i \log L(H_{h_n}(\boldsymbol{Z_i})^T \boldsymbol{\pi}) + (1 - D_i) \log(1 - L(H_{h_n}(\boldsymbol{Z_i})^T \boldsymbol{\pi})),$$

*Corresponding author. E-mail: deyuanli@fudan.edu.cn

and $L(u) = 1/(1 + e^{-u})$.

For H_{h_n}, the span of these basis functions contains all polynomials up to order m if $h_n > (m + 1)^r$. Define

$$\zeta(h_n) = \sup_{\boldsymbol{z}\in\mathcal{Z}} \|H_{h_n}(\boldsymbol{z})\|.$$

If H_{h_n} consists of orthonormal polynomials, it follows that $\zeta(h_n) = O(h_n)$, see, e.g. Newey (1994) and Deuber et al. (2024).

For the estimation of $\Pi(\boldsymbol{z})$ in simulations with $n = 1000, 2000$, we set $h_n = \lfloor 5 \times n^{1/11} \rfloor$ in Sections 3.1–3.2 under the one-dimensional covariate setting, which satisfies Assumption (g). In Section S7, under the two-dimensional covariate setting, we set $m = \lfloor 2 \times n^{1/11} \rfloor$.

S2 Assumption (g)

Assumption (g) consists of the following two assumptions.

Assumption (g1). Assume that as $n \to \infty$,

$$\sup_{\boldsymbol{z}\in\mathcal{Z}} |\widehat{\Pi}(\boldsymbol{z}) - \Pi(\boldsymbol{z})| = o_p(n^{-1/4}).$$

Assumption (g2). We assume that the following limits exist:

$$\widetilde{H}_1 = \lim_{n\to\infty} \frac{n}{k}\mathbb{E}\left[\frac{\Pr\{Y(1) > U_{Y(1)}(n/k)|\boldsymbol{Z}\}}{\Pi(\boldsymbol{Z})}\right],$$
$$\widetilde{H}_0 = \lim_{n\to\infty} \frac{n}{k}\mathbb{E}\left[\frac{\Pr\{Y(0) > U_{Y(0)}(n/k)|\boldsymbol{Z}\}}{1-\Pi(\boldsymbol{Z})}\right],$$
$$\widetilde{G}_1 = \lim_{n\to\infty} \frac{n}{k}\mathbb{E}\left[\frac{1}{\Pi(\boldsymbol{Z})}\mathbb{E}[S_{1,n}^2|\boldsymbol{Z}]\right],$$
$$\widetilde{G}_0 = \lim_{n\to\infty} \frac{n}{k}\mathbb{E}\left[\frac{1}{1-\Pi(\boldsymbol{Z})}\mathbb{E}[S_{0,n}^2|\boldsymbol{Z}]\right],$$
$$\widetilde{J}_1 = \lim_{n\to\infty} \frac{n}{k}\mathbb{E}\left[\frac{1}{\Pi(\boldsymbol{Z})}\mathbb{E}[S_{1,n}|\boldsymbol{Z}]\right],$$
$$\widetilde{J}_0 = \lim_{n\to\infty} \frac{n}{k}\mathbb{E}\left[\frac{1}{1-\Pi(\boldsymbol{Z})}\mathbb{E}[S_{0,n}|\boldsymbol{Z}]\right],$$

where $S_{j,n} = \gamma_j \log\left(\frac{k}{n(1-F_{Y(j)}(Y(j)))}\right)\mathcal{I}\{Y(j) > U_{Y(j)}(n/k)\}$ for $j = 0, 1$.

Remark S1. *Assumption (g1) can be obtained by the following sufficient conditions through Assumptions (a) and Lemma G.3 in Deuber et al. (2024):*

- *Assume that $\boldsymbol{Z}$ has a compact support $\mathcal{Z}$, and $\boldsymbol{Z}$ is continuous and there exists $c' > 0$ such that $c' \leq f_{\boldsymbol{Z}}(\boldsymbol{z}) \leq 1/c'$ holds uniformly on $\boldsymbol{z} \in \mathcal{Z}$.*

- *$\Pi(\boldsymbol{z})$ is s-order continuously differentiable with $s \geq 4r$ and all derivatives are bounded, where r is the dimension of $\boldsymbol{Z}$. Also, we assume that $0 \leq h_n = O(n^h)$, and $\frac{\zeta(h_n)^2 h_n}{\sqrt{n}} \to 0$. In addition, s, h satisfy that $\zeta(h_n)h_n^{-s/2r} = o(n^{-1/4})$.*

Assumption (g2) represents the moment conditions for the asymptotic *distributions of* $\widehat{\gamma}_j$*.*

S3 Preliminary lemmas

The following lemma shows the asymptotic properties of $\widehat{\gamma}_1$ and $\widehat{\gamma}_0$, which can be obtained by combining the results of Theorems 1, 2 and Lemma G.6 in Deuber et al. (2024).

Lemma S1. *Suppose that Assumptions (a), (b), (e) and (g) hold. Moreover, assume that* $k = O(n^{1/2})$ *and* $\sqrt{k}\{|A_1(n/k)| + |A_0(n/k)|\} \to 0$*, as* $n \to \infty$*. Then, we have that,*

$$\sqrt{k}(\widehat{\gamma}_1 - \gamma_1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\psi_{i,1,n} - \gamma_1 \phi_{i,1,n}) + o_p(1) =: \Gamma_{n1} + o_p(1),$$
$$\sqrt{k}(\widehat{\gamma}_0 - \gamma_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\psi_{i,0,n} - \gamma_0 \phi_{i,0,n}) + o_p(1) =: \Gamma_{n0} + o_p(1),$$

where

$$\psi_{i,0,n} = \sqrt{\frac{n}{k}} \left(\frac{1 - D_i}{1 - \Pi(\boldsymbol{Z_i})} S_{i,0,n} - \frac{k\gamma_0}{n} \right),$$
$$\psi_{i,1,n} = \sqrt{\frac{n}{k}} \left(\frac{D_i}{\Pi(\boldsymbol{Z_i})} S_{i,1,n} - \frac{k\gamma_1}{n} \right),$$
$$\phi_{i,0,n} = \sqrt{\frac{n}{k}} \frac{1 - D_i}{1 - \Pi(\boldsymbol{Z_i})} T_{i,0,n},$$
$$\phi_{i,1,n} = \sqrt{\frac{n}{k}} \frac{D_i}{\Pi(\boldsymbol{Z_i})} T_{i,1,n},$$

and for $j = 0, 1$,

$$S_{i,j,n} = \gamma_j \log\left(\frac{k}{n(1 - F_{Y(j)}(Y_i(j)))}\right) \mathcal{I}\{Y_i(j) > U_{Y(j)}(n/k)\},$$

$$T_{i,j,n} = \mathcal{I}\{Y_i(j) > U_{Y(j)}(n/k)\} - k/n.$$

Moreover, we have that,

$$(\Gamma_{n1}, \Gamma_{n0}) \xrightarrow{d} (\Gamma_1, \Gamma_0) \sim N(0, B\widetilde{\Sigma}B^T),$$

where

$$\widetilde{\Sigma} := \begin{pmatrix} \widetilde{G}_1 & 0 & \widetilde{J}_1 & 0 \\ 0 & \widetilde{G}_0 & 0 & \widetilde{J}_0 \\ \widetilde{J}_1 & 0 & \widetilde{H}_1 & 0 \\ 0 & \widetilde{J}_0 & 0 & \widetilde{H}_0 \end{pmatrix}.$$

Remark S2. *In the proof of Theorem 1 in Deuber et al. (2024), we can show that* $\sqrt{k}G_n^4 = o_p(1)$ *under the condition* $k = O(n^{1/2})$. *This condition simplifies the asymptotic analysis, which aids in the potential variance estimation. In addition, the condition* $k = O(n^{1/2})$ *removes the need for Assumption 4, B.1 (iii) and the last three limit expressions in (iv) of Deuber et al. (2024).*

The following lemma establishes the Lipschitz continuity of $\widetilde{R}^{(1)}(x, y)$ and $\widetilde{R}^{(0)}(x, y)$, which aids in establishing the continuity of the associated Gaussian processes.

Lemma S2. *The functions* $\widetilde{R}^{(1)}(x,y)$ *and* $\widetilde{R}^{(0)}(x,y)$ *are Lipschitz continuous with respect to* $(x,y) \in [0,\infty)^2$.

Proof of Lemma S2. Without loss of generality, we focus on $\widetilde{R}^{(1)}(x,y)$, as the analysis for $\widetilde{R}^{(0)}(x,y)$ is analogous.

Note that

$$
\begin{aligned}
&|\widetilde{R}^{(1)}(x_1,y_1) - \widetilde{R}^{(1)}(x_2,y_2)| \\
\leq &|\widetilde{R}^{(1)}(x_1,y_1) - \widetilde{R}^{(1)}(x_1,y_2)| + |\widetilde{R}^{(1)}(x_1,y_2) - \widetilde{R}^{(1)}(x_2,y_2)| \\
\leq &|\widetilde{R}^{(1)}(\infty,y_1) - \widetilde{R}^{(1)}(\infty,y_2)| + |\widetilde{R}^{(1)}(x_1,\infty) - \widetilde{R}^{(1)}(x_2,\infty)|.
\end{aligned}
$$

Then by the condition that $\Pi(\boldsymbol{z}) \in (c, 1-c)$, we have that for any $(x_1,y_1) \in [0,\infty)^2$, $(x_2,y_2) \in [0,\infty)^2$,

$$
\begin{aligned}
&|\widetilde{R}^{(1)}(x_1,y_1) - \widetilde{R}^{(1)}(x_2,y_2)| \\
\leq &\lim_{n\to\infty} \frac{n}{k}\mathbb{E}\left[\frac{1}{\Pi(\boldsymbol{Z})}\mathcal{I}\{1-F_X(X) < kx_1/n, 1-F_{Y(1)}(Y(1)) \in [ky_1/n, ky_2/n]\}\right] \\
&+ \lim_{n\to\infty} \frac{n}{k}\mathbb{E}\left[\frac{1}{\Pi(\boldsymbol{Z})}\mathcal{I}\{1-F_X(X) \in [kx_1/n, kx_2/n], 1-F_{Y(1)}(Y(1)) < ky_2/n\}\right] \\
\leq &\frac{1}{c}\lim_{n\to\infty} \frac{n}{k}\mathbb{E}[\mathcal{I}\{1-F_{Y(1)}(Y(1)) \in [ky_1/n, ky_2/n]\}] \\
&\quad + \frac{1}{c}\lim_{n\to\infty} \frac{n}{k}\mathbb{E}[\mathcal{I}\{1-F_X(X) \in [kx_1/n, kx_2/n]\}] \\
= &\frac{1}{c}\left|y_1 - y_2\right| + \frac{1}{c}\left|x_1 - x_2\right|.
\end{aligned}
$$

Thus, Lemma S2 holds. $\square$

S4 Variance Estimation

In Section 2.4, we propose the variance estimator for $\widehat{\mathrm{CTE}}_2(p)$. Specifically, we estimate $\widetilde{\Sigma}$ by

$$\widehat{\Sigma} := \begin{pmatrix} \widehat{G}_1 & 0 & \widehat{J}_1 & 0 \\ 0 & \widehat{G}_0 & 0 & \widehat{J}_0 \\ \widehat{J}_1 & 0 & \widehat{H}_1 & 0 \\ 0 & \widehat{J}_0 & 0 & \widehat{H}_0 \end{pmatrix},$$

where

$$\begin{aligned}
\widehat{H}_1 &:= \frac{1}{k}\sum_{i=1}^{n} \frac{D_i}{\widehat{\Pi}(\boldsymbol{Z_i})^2}\mathcal{I}\{Y_i > \widehat{Q}_{Y(1)}(1-k/n)\}, \\
\widehat{H}_0 &:= \frac{1}{k}\sum_{i=1}^{n} \frac{1-D_i}{(1-\widehat{\Pi}(\boldsymbol{Z_i}))^2}\mathcal{I}\{Y_i > \widehat{Q}_{Y(0)}(1-k/n)\}, \\
\widehat{G}_1 &:= \frac{1}{k}\sum_{i=1}^{n}(\log(Y_i) - \log(\widehat{Q}_{Y(1)}(1-k/n)))^2 \frac{D_i}{\widehat{\Pi}(\boldsymbol{Z_i})^2}\mathcal{I}\{Y_i > \widehat{Q}_{Y(1)}(1-k/n)\}, \\
\widehat{G}_0 &:= \frac{1}{k}\sum_{i=1}^{n}(\log(Y_i) - \log(\widehat{Q}_{Y(0)}(1-k/n)))^2 \frac{1-D_i}{(1-\widehat{\Pi}(\boldsymbol{Z_i}))^2}\mathcal{I}\{Y_i > \widehat{Q}_{Y(0)}(1-k/n)\}, \\
\widehat{J}_1 &:= \frac{1}{k}\sum_{i=1}^{n}(\log(Y_i) - \log(\widehat{Q}_{Y(1)}(1-k/n))) \frac{D_i}{\widehat{\Pi}(\boldsymbol{Z_i})^2}\mathcal{I}\{Y_i > \widehat{Q}_{Y(1)}(1-k/n)\}, \\
\widehat{J}_0 &:= \frac{1}{k}\sum_{i=1}^{n}(\log(Y_i) - \log(\widehat{Q}_{Y(0)}(1-k/n))) \frac{1-D_i}{(1-\widehat{\Pi}(\boldsymbol{Z_i}))^2}\mathcal{I}\{Y_i > \widehat{Q}_{Y(0)}(1-k/n)\}.
\end{aligned}$$

Then we can obtain the following variance estimator of σ^2:

$$\widehat{\sigma}^2 := \widehat{v}_\kappa^T \widehat{B}\widehat{\Sigma}\widehat{B}^T \widehat{v}_\kappa,$$

where

$$\widehat{B} := \begin{pmatrix} 1 & 0 & -\widehat{\gamma}_1 & 0 \\ 0 & 1 & 0 & -\widehat{\gamma}_0 \end{pmatrix}, \quad \widehat{v}_\kappa := \begin{pmatrix} \min\{1, \widehat{\kappa}\} \\ -\min\{1, 1/\widehat{\kappa}\} \end{pmatrix}$$

with $\widehat{\kappa} = \widehat{\theta}_p^{(1)}/\widehat{\theta}_p^{(0)}$.

S5 Multivariate Causal Tail Empirical Process

In this section, we define multivariate causal tail empirical process (MCTEP) and explore its asymptotic properties. These results provide the theoretical foundation for the proofs of the main theorems in our paper.

For $i = 1, 2, \ldots, n$, define

$$\begin{aligned} H_{n,i} = &\frac{1}{\sqrt{k}} \frac{1 - D_i}{1 - \Pi(\boldsymbol{Z_i})} \delta_{\left\{\left(\frac{n}{k}(1-F_X(X_i)), \frac{n}{k}(1-F_{Y(0)}(Y_i(0))), 0\right)\right\}} \\ &+ \frac{1}{\sqrt{k}} \frac{D_i}{\Pi(\boldsymbol{Z_i})} \delta_{\left\{\left(\frac{n}{k}(1-F_X(X_i)), \frac{n}{k}(1-F_{Y(1)}(Y_i(1))), 1\right)\right\}} \\ &+ \frac{1}{\sqrt{k}} \delta_{\left\{\left(\frac{n}{k}(1-F_X(X_i)), \frac{n}{k}(1-F_{Y(0)}(Y_i(0))), 2\right)\right\}} \\ &+ \frac{1}{\sqrt{k}} \delta_{\left\{\left(\frac{n}{k}(1-F_X(X_i)), \frac{n}{k}(1-F_{Y(1)}(Y_i(1))), 3\right)\right\}}, \end{aligned}$$

where $\delta_{\boldsymbol{U}}(\cdot)$ is the Dirac measure satisfying

$$\delta_{\boldsymbol{U}}(\mathcal{S}) = \begin{cases} 1, & \text{if } \boldsymbol{U} \in \mathcal{S}, \\ 0, & \text{otherwise.} \end{cases}$$

For any function f, define $H_{n,i}(f) = \int f dH_{n,i}$. For fixed $T > 0$, and for all $0 < x, y \leq T, a \in \{0, 1, 2, 3\}$, we define that

$$f_{x,y,a} = \frac{I_{[0,x)\times[0,y)\times\{a\}}}{y^\eta},$$

$$f^{(1)}_{x,a} = \frac{I_{[0,x)\times[0,\infty]\times\{a\}}}{x^\eta},$$

$$f^{(2)}_{y,a} = \frac{I_{[0,\infty]\times[0,y)\times\{a\}}}{y^\eta},$$

where $\eta \in [0, 1/2)$ is fixed.

Define

$$T_n^{(a)}(x,y) = \begin{cases} \frac{1}{\sqrt{k}} \sum_{i=1}^n \frac{1-D_i}{1-\Pi(\boldsymbol{Z_i})} \mathcal{I}\left\{1 - F_X(X_i) < \frac{kx}{n}, 1 - F_{Y(0)}(Y_i(0)) < \frac{ky}{n}\right\}, & \text{if } a = 0, \\ \frac{1}{\sqrt{k}} \sum_{i=1}^n \frac{D_i}{\Pi(\boldsymbol{Z_i})} \mathcal{I}\left\{1 - F_X(X_i) < \frac{kx}{n}, 1 - F_{Y(1)}(Y_i(1)) < \frac{ky}{n}\right\}, & \text{if } a = 1, \\ \frac{1}{\sqrt{k}} \sum_{i=1}^n \mathcal{I}\left\{1 - F_X(X_i) < \frac{kx}{n}, 1 - F_{Y(0)}(Y_i(0)) < \frac{ky}{n}\right\}, & \text{if } a = 2, \\ \frac{1}{\sqrt{k}} \sum_{i=1}^n \mathcal{I}\left\{1 - F_X(X_i) < \frac{kx}{n}, 1 - F_{Y(1)}(Y_i(1)) < \frac{ky}{n}\right\}, & \text{if } a = 3, \end{cases}$$

and

$$R_n^{(a)}(x,y) = \begin{cases} \frac{n}{k} \Pr\left\{1 - F_X(X) < \frac{kx}{n}, 1 - F_{Y(0)}(Y(0)) < \frac{ky}{n}\right\}, & \text{if } a = 0 \text{ or } 2, \\ \frac{n}{k} \Pr\left\{1 - F_X(X) < \frac{kx}{n}, 1 - F_{Y(1)}(Y(1)) < \frac{ky}{n}\right\}, & \text{if } a = 1 \text{ or } 3. \end{cases}$$

Then we have that

$$T_n^{(a)}(x,y) = y^\eta H_{n,i}(f_{x,y,a}), \quad R_n^{(a)}(x,y) = E[T_n^{(a)}(x,y)],$$

$$T_n^{(a)}(x,\infty) = x^\eta H_{n,i}(f^{(1)}_{x,a}), \quad R_n^{(a)}(x,\infty) = E[T_n^{(a)}(x,\infty)],$$

$$T_n^{(a)}(\infty,y) = y^\eta H_{n,i}(f^{(2)}_{y,a}), \quad R_n^{(a)}(\infty,y) = E[T_n^{(a)}(\infty,y)].$$

We define a mean zero Gaussian process $W^{(a)}(x,y)$, where $(x,y) \in [0,\infty]^2 \backslash \{(\infty,\infty)\}, a \in \{0,1,2,3\}$. For any (a,x,y), $W^{(a)}(x,y)$ follows a normal distribution with mean zero. The covariance structure is given by

$$\mathrm{Cov}(W^{(a)}(x,y), W^{(b)}(x',y')) = \begin{cases} \widetilde{R}^{(0)}(x \wedge x', y \wedge y'), & \text{if } a = b = 0, \\ \widetilde{R}^{(1)}(x \wedge x', y \wedge y'), & \text{if } a = b = 1, \\ R^{(0)}(x \wedge x', y \wedge y'), & \text{if } a = b = 2, \\ R^{(1)}(x \wedge x', y \wedge y'), & \text{if } a = b = 3, \\ R^{(0)}(x \wedge x', y \wedge y'), & \text{if } (a,b) = (0,2),(2,0) \\ R^{(1)}(x \wedge x', y \wedge y'), & \text{if } (a,b) = (1,3),(3,1), \\ R_{X,Y(0),Y(1)}(x \wedge x', y', y), & \text{if } (a,b) = (3,2), \\ R_{X,Y(0),Y(1)}(x \wedge x', y, y'), & \text{if } (a,b) = (2,3), \\ R_{X,Y(0),Y(1)}(x \wedge x', y, y'), & \text{if } (a,b) = (0,3), \\ R_{X,Y(0),Y(1)}(x \wedge x', y', y), & \text{if } (a,b) = (3,0), \\ R_{X,Y(0),Y(1)}(x \wedge x', y', y), & \text{if } (a,b) = (1,2), \\ R_{X,Y(0),Y(1)}(x \wedge x', y, y'), & \text{if } (a,b) = (2,1), \\ 0, & \text{if } (a,b) = (0,1),(1,0). \end{cases}$$

Then we have that for any fixed a, $W^{(a)}(x,y)$ is a continuous Gaussian process with regard to (x,y) by Corollary 1.11 in Adler (1990). In the rest of the paper, we denote

$$W_{\widetilde{R}^{(0)}}(x,y) := W^{(0)}(x,y), \quad W_{\widetilde{R}^{(1)}}(x,y) := \ W^{(1)}(x,y),$$

$$W_{R^{(0)}}(x,y) := W^{(2)}(x,y), \quad W_{R^{(1)}}(x,y) := \ W^{(3)}(x,y).$$

The following lemma, analogous to Lemma 3.2 in Einmahl et al. (2006), shows the boundedness of $W^{(a)}(x,y)$ with proper weighting function.

Lemma S3. *For any $T>0$, $\eta\in[0,1/2)$ and $a\in\{0,1,2,3\}$, we have that with probability 1,*

$$\sup_{0<x<\infty, 0<y\leq T}\frac{|W^{(a)}(x,y)|}{y^\eta}<\infty, \qquad \sup_{0<x\leq T, 0<y<\infty}\frac{|W^{(a)}(x,y)|}{x^\eta}<\infty.$$

Proof of Lemma S3. We only give the proof of the first statement of Lemma S3 with $a=1$, since the proofs in other cases are analogous.

Let $\lambda>0$ be a constant. We intend to derive upper bounds of the following probabilities:

$$J_1 := \Pr\left\{\sup_{x,y\in(0,T]}\frac{|W_{\widetilde{R}^{(1)}}(x,y)|}{y^\eta}>\lambda\right\},$$
$$J_2 := \Pr\left\{\sup_{x\geq T, y\in(0,T]}\frac{|W_{\widetilde{R}^{(1)}}(x,y)|}{y^\eta}>\lambda\right\}.$$

We start with J_1. Define

$$\mathcal{A}_{l,m} = \left\{(x,y): \frac{T}{2^{l+1}}\leq x\leq T, \frac{T}{2^{m+1}}\leq y\leq\frac{T}{2^m}\right\},$$
$$\mathcal{A}_m = \left\{(x,y): 0<x\leq T, \frac{T}{2^{m+1}}\leq y\leq\frac{T}{2^m}\right\}.$$

It follows that $\mathcal{A}_{0,m}\subseteq\mathcal{A}_{1,m}\ldots$, and $\lim_{l\to\infty}\mathcal{A}_{l,m}=\mathcal{A}_m$, $(0,T]\times(0,T]=$

$\cup_{m=0}^{\infty}\mathcal{A}_m$. Moreover, we have that,

$$\begin{aligned}
&\Pr\left\{\sup_{x,y\in\mathcal{A}_{l,m}}\frac{|W_{\widetilde{R}^{(1)}}(x,y)|}{y^\eta}>\lambda\right\}\\
\leq&\Pr\left\{\sup_{x,y\in\mathcal{A}_{l,m}}|W_{\widetilde{R}^{(1)}}(x,y)|\geq\lambda\left(\frac{T}{2^{m+1}}\right)^\eta\right\}\\
\leq&4\Pr\left\{\left|W_{\widetilde{R}^{(1)}}\left(T,\frac{T}{2^m}\right)\right|\geq\lambda\left(\frac{T}{2^{m+1}}\right)^\eta\right\}\\
=&4\Pr\left\{\left(\widetilde{R}^{(1)}\left(T,\frac{T}{2^m}\right)\right)^{1/2}|Z|\geq\lambda\left(\frac{T}{2^{m+1}}\right)^\eta\right\}\\
\leq&4\Pr\left\{|Z|\geq\left(\frac{2^m c}{T}\right)^{1/2}\lambda\left(\frac{T}{2^{m+1}}\right)^\eta\right\}\\
\leq&8\exp\left(-\frac{2^{m(1-2\eta)}c\lambda^2}{T^{1-2\eta}2^{2\eta+1}}\right),
\end{aligned}$$

where $Z\sim N(0,1)$. The second inequality can be similarly derived from Lemma 3.2 in Einmahl et al. (2006) or an adaptation of Lemma 1.2 in Orey and Pruitt (1973). The equality in the third line follows from the fact that $W_{\tilde{R}^{(1)}}(x,y)$ is equal to $(\widetilde{R}^{(1)}(x,y))^{1/2}|Z|$ in distribution. The third inequality holds by Assumption (a). The last inequality can be derived from Mill's ratio.

Since the upper bound above is irrelevant to l, we have that

$$\begin{aligned}
\Pr\left\{\sup_{x,y\in\mathcal{A}_m}\frac{|W_{\widetilde{R}^{(1)}}(x,y)|}{y^\eta}>\lambda\right\}=&\Pr\left\{\lim_{l\to\infty}\sup_{x,y\in\mathcal{A}_{l,m}}\frac{|W_{\widetilde{R}^{(1)}}(x,y)|}{y^\eta}>\lambda\right\}\\
=&\lim_{l\to\infty}\Pr\left\{\sup_{x,y\in\mathcal{A}_{l,m}}\frac{|W_{\widetilde{R}^{(1)}}(x,y)|}{y^\eta}>\lambda\right\}\\
\leq&\,8\exp\left(-\frac{2^{m(1-2\eta)}c\lambda^2}{T^{1-2\eta}2^{2\eta+1}}\right).
\end{aligned}$$

The first equality follows from the fact that

$$\left\{\omega : \lim_{l\to\infty} \sup_{x,y\in\mathcal{A}_{l,m}} \frac{|W_{\widetilde{R}^{(1)}}(x,y)(\omega)|}{y^{\eta}} > \lambda\right\} = \left\{\omega : \sup_{x,y\in\mathcal{A}_{m}} \frac{|W_{\widetilde{R}^{(1)}}(x,y)(\omega)|}{y^{\eta}} > \lambda\right\}.$$

The second equality follows from the monotone convergence theorem. Consequently, we obtain that

$$\begin{aligned} J_1 = & \Pr\left\{\sup_{x,y\in(0,T]} \frac{|W_{\widetilde{R}^{(1)}}(x,y)|}{y^{\eta}} > \lambda\right\} \\ \leq & \sum_{m=0}^{\infty} \Pr\left\{\sup_{x,y\in\mathcal{A}_m} \frac{|W_{\widetilde{R}^{(1)}}(x,y)|}{y^{\eta}} > \lambda\right\} \\ \leq & \sum_{m=0}^{\infty} 8\exp\left(-\frac{2^{m(1-2\eta)}c^2\lambda^2}{T^{1-2\eta}2^{2\eta+1}}\right). \end{aligned}$$

Next, we derive the upper bound of J_2. Define

$$\mathcal{B}_{b,m} = \left\{(x,y) : T \leq x \leq 2^b T, \frac{T}{2^{m+1}} \leq y \leq \frac{T}{2^m}\right\},$$

$$\mathcal{B}_{m} = \lim_{b\to\infty}\mathcal{B}_{b,m} = \left\{(x,y) : T \leq x < \infty, \frac{T}{2^{m+1}} \leq y \leq \frac{T}{2^m}\right\}.$$

It follows that $\mathcal{B}_{1,m} \subseteq \mathcal{B}_{2,m}\ldots$, and $\lim_{b\to\infty}\mathcal{B}_{b,m} = \mathcal{B}_m$, $[T,\infty)\times(0,T] = \cup_{m=0}^{\infty}\mathcal{B}_m$. Similarly, we have that

$$\begin{aligned} & \Pr\left\{\sup_{x,y\in\mathcal{B}_{b,m}} \frac{|W_{\widetilde{R}^{(1)}}(x,y)|}{y^{\eta}} > \lambda\right\} \\ \leq & \Pr\left\{\sup_{x,y\in\mathcal{B}_{b,m}} |W_{\widetilde{R}^{(1)}}(x,y)| \geq \lambda\left(\frac{T}{2^{m+1}}\right)^{\eta}\right\} \\ \leq & 4\Pr\left\{\left|W_{\widetilde{R}^{(1)}}\left(2^b T, \frac{T}{2^m}\right)\right| \geq \lambda\left(\frac{T}{2^{m+1}}\right)^{\eta}\right\} \\ \leq & 4\Pr\left\{|Z| \geq \left(\frac{2^m c}{T}\right)^{1/2}\lambda\left(\frac{T}{2^{m+1}}\right)^{\eta}\right\} \\ \leq & 8\exp\left(-\frac{2^{m(1-2\eta)}c\lambda^2}{T^{1-2\eta}2^{2\eta+1}}\right). \end{aligned}$$

where $Z \sim N(0,1)$. It follows that

$$\begin{aligned}
&\Pr\left\{\sup_{x,y\in\mathcal{B}_m}\frac{|W_{\widetilde{R}^{(1)}}(x,y)|}{y^\eta}>\lambda\right\}\\
=&\lim_{b\to\infty}\Pr\left\{\sup_{x,y\in\mathcal{B}_{b,m}}\frac{|W_{\widetilde{R}^{(1)}}(x,y)|}{y^\eta}>\lambda\right\}\\
\leq&8\exp\left(-\frac{2^{m(1-2\eta)}c\lambda^2}{T^{1-2\eta}2^{2\eta+1}}\right),
\end{aligned}$$

and

$$\begin{aligned}
J_2=&\Pr\left\{\sup_{x\geq T,y\in(0,T]}\frac{|W_{\widetilde{R}^{(1)}}(x,y)|}{y^\eta}>\lambda\right\}\\
\leq&\sum_{m=0}^{\infty}\Pr\left\{\sup_{x,y\in\mathcal{B}_m}\frac{|W_{\widetilde{R}^{(1)}}(x,y)|}{y^\eta}>\lambda\right\}\\
\leq&\sum_{m=0}^{\infty}8\exp\left(-\frac{2^{m(1-2\eta)}c\lambda^2}{T^{1-2\eta}2^{2\eta+1}}\right).
\end{aligned}$$

Combining the upper bounds of J_1 and J_2, we obtain that,

$$\Pr\left\{\sup_{x\geq T,y\in(0,T]}\frac{|W_{\widetilde{R}^{(1)}}(x,y)|}{y^\eta}>\lambda\right\}\leq\sum_{m=0}^{\infty}8\exp\left(-\frac{2^{m(1-2\eta)}c\lambda^2}{T^{1-2\eta}2^{2\eta+1}}\right).$$

Thus, for any $\varepsilon>0$, there exists $\lambda>0$ such that,

$$\sum_{m=0}^{\infty}8\exp\left(-\frac{2^{m(1-2\eta)}c\lambda^2}{T^{1-2\eta}2^{2\eta+1}}\right)<\varepsilon.$$

The proof of Lemma S3 is then complete. □

The following lemma establishes the convergence of the MCTEP. For convenient presentation, all the processes are defined on the same probability space, via the Skorohod constructions.

Lemma S4. *Suppose that condition* (2.2), *Assumptions (a) and (d) hold. Let* $T>0$ *be a constant. There exists a sequence of Gaussian processes*

$\{W_n^{(a)}(x,y)\}$, *satisfying* $W_n^{(a)}(x,y) \stackrel{d}{=} W^{(a)}(x,y)$, *such that for any* $\eta \in [0,1/2)$ *and* $a \in \{0,1,2,3\}$,

$$\sup_{x,y\in(0,T]}\left|\frac{\sqrt{k}\{T_n^{(a)}(x,y)-R_n^{(a)}(x,y)\}-W_n^{(a)}(x,y)}{y^\eta}\right|\xrightarrow{P}0,$$
$$\sup_{x\in(0,T]}\left|\frac{\sqrt{k}\{T_n^{(a)}(x,\infty)-x\}-W_n^{(a)}(x,\infty)}{x^\eta}\right|\xrightarrow{P}0,$$
$$\sup_{y\in(0,T]}\left|\frac{\sqrt{k}\{T_n^{(a)}(\infty,y)-y\}-W_n^{(a)}(\infty,y)}{y^\eta}\right|\xrightarrow{P}0.$$

Proof of Lemma S4. We only provide the proofs for $a=1$. The proofs for $a \in \{0,2,3\}$ are similar and thus omitted. Thus, we only prove that, for any $\eta \in [0,1/2)$,

$$\sqrt{k}(T_n^{(1)}(x,y)-R_n^{(1)}(x,y))/y^\eta \xrightarrow{d} W_{\widetilde{R}^{(1)}}(x,y)/y^\eta, x,y\in(0,T], \quad \text{(S5.1)}$$
$$\sqrt{k}(T_n^{(1)}(x,\infty)-R_n^{(1)}(x,\infty))/x^\eta \xrightarrow{d} W_{\widetilde{R}^{(1)}}(x,\infty)/x^\eta, x\in(0,T], \quad \text{(S5.2)}$$
$$\sqrt{k}(T_n^{(1)}(\infty,y)-R_n^{(1)}(\infty,y))/y^\eta \xrightarrow{d} W_{\widetilde{R}^{(1)}}(\infty,y)/y^\eta, y\in(0,T]. \quad \text{(S5.3)}$$

We only provide the proofs for (S5.1). The proofs for (S5.2) and (S5.3) are similar and thus omitted.

Let $\mathcal{F}_{2,1}$ be the set of all forms of $f_{x,y,1}$. Then, (S5.1) is equivalent to the weak convergence of the process

$$\left\{\sum_{i=1}^n [H_{n,i}(f)-\mathbb{E}H_{n,i}(f)], f\in\mathcal{F}_{2,1}\right\}$$

to the related Gaussian process. We equip $\mathcal{F}_{2,1}$ with the following semimetric d, defined by

$$\begin{aligned}
d^2(f_{x,y,1}, f_{u,v,1}) =& \mathbb{E}\left[\frac{W^{(1)}(x,y)}{y^\eta} - \frac{W^{(1)}(u,v)}{v^\eta}\right]^2 \\
=& \mathbb{E}\left[\frac{W_{\widetilde{R}^{(1)}}(x,y)}{y^\eta} - \frac{W_{\widetilde{R}^{(1)}}(u,v)}{v^\eta}\right]^2 \\
=& \frac{v^{2\eta}\widetilde{R}^{(1)}(x,y) - 2y^\eta v^\eta \widetilde{R}^{(1)}(u\wedge x, v\wedge y) + y^{2\eta}\widetilde{R}^{(1)}(u,v)}{(yv)^{2\eta}}.
\end{aligned}$$

By Theorem 2.11.9 of van der Vaart and Wellner (1996), we divide the proof of (S5.1) into three steps. In the first step, we prove $(\mathcal{F}_{2,1}, d)$ is totally bounded. In the second step, we prove the asymptotic tightness of the process $\{\sum_{i=1}^n [H_{n,i}(f) - \mathbb{E}H_{n,i}(f)], f \in \mathcal{F}_{2,1}\}$. In the third step, we prove the weak convergence of the finite-dimensional distributions.

Step 1. We prove $(\mathcal{F}_{2,1}, d)$ is totally bounded.

For any $\delta > 0$, assume that $|x-u| \le \delta, |y-v| \le \delta$. We distinguish eight cases below.

1. If $v \le \delta, y \ge v$, we have that

$$\begin{aligned}
d^2(f_{x,y,1}, f_{u,v,1}) \le& \frac{\widetilde{R}^{(1)}(x,y)}{y^{2\eta}} + \frac{2\widetilde{R}^{(1)}(u,v)}{v^{2\eta}} + \frac{\widetilde{R}^{(1)}(u,v)}{v^{2\eta}} \\
\le& \frac{1}{c}(y^{1-2\eta} + 3v^{1-2\eta}) \le \frac{1}{c}5\delta^{1-2\eta}
\end{aligned}$$

by Assumption (a).

2. If $v \le \delta, y \le v$, we have

$$
\begin{aligned}
d^2(f_{x,y,1}, f_{u,v,1}) &\le \frac{\widetilde{R}^{(1)}(x,y)}{y^{2\eta}} + \frac{2\widetilde{R}^{(1)}(x,y)}{y^{2\eta}} + \frac{\widetilde{R}^{(1)}(u,v)}{v^{2\eta}} \\
&\le \frac{1}{c}(3y^{1-2\eta} + v^{1-2\eta}) \le \frac{1}{c} 4\delta^{1-2\eta}.
\end{aligned}
$$

3. If $v > \delta, x \ge u, y \ge v$, due to the fact that

$$
\widetilde{R}^{(1)}(x,y) \le \widetilde{R}^{(1)}(u,v) + \widetilde{R}^{(1)}([u,x]\times[0,\infty]) + \widetilde{R}^{(1)}([0,\infty]\times[v,y]) \le \widetilde{R}^{(1)}(u,v) + \frac{2\delta}{c},
$$

we obtain that

$$
\begin{aligned}
d^2(f_{x,y,1}, f_{u,v,1}) &\le \frac{\widetilde{R}^{(1)}(u,v)(y^\eta - v^\eta)^2}{(yv)^{2\eta}} + \frac{2\delta v^{2\eta}}{c(yv)^{2\eta}} \\
&\le \frac{1}{c} v^{1-4\eta}(y^\eta - v^\eta)^2 + \frac{2\delta^{1-2\eta}}{c} \\
&\le \frac{1}{c}(v^{1-4\eta} y^{2\eta-2}(y-v)^2 + 2\delta^{1-2\eta}) \\
&\le \frac{1}{c}(v^{-1-2\eta}(y-v)^2 + 2\delta^{1-2\eta}) \le \frac{3\delta^{1-2\eta}}{c}.
\end{aligned}
$$

4. If $v > \delta, x < u, y \ge v$, we have that

$$
\widetilde{R}^{(1)}(x,y) \le \widetilde{R}^{(1)}(x,v) + \frac{\delta}{c}, \quad \widetilde{R}^{(1)}(u,v) \le \widetilde{R}^{(1)}(x,v) + \frac{\delta}{c}.
$$

It follows that

$$
\begin{aligned}
d^2(f_{x,y,1}, f_{u,v,1}) &\le \frac{\widetilde{R}^{(1)}(x,v)(y^\eta - v^\eta)^2}{(yv)^{2\eta}} + \frac{\delta v^{2\eta}}{c(yv)^{2\eta}} + \frac{\delta y^{2\eta}}{c(yv)^{2\eta}} \\
&\le \frac{1}{c} v^{1-4\eta}(y^\eta - v^\eta)^2 + \frac{2\delta^{1-2\eta}}{c} \\
&\le \frac{1}{c}(v^{1-4\eta} y^{2\eta-2}(y-v)^2 + 2\delta^{1-2\eta}) \\
&\le \frac{1}{c}(v^{-1-2\eta}(y-v)^2 + 2\delta^{1-2\eta}) \le \frac{3\delta^{1-2\eta}}{c}.
\end{aligned}
$$

5. If $v > \delta, x < u, \delta \le y < v$, we have that

$$\widetilde{R}^{(1)}(u,v) \le \widetilde{R}^{(1)}(x,y) + \frac{2\delta}{c},$$

and hence

$$\begin{aligned} d^2(f_{x,y,1}, f_{u,v,1}) &\le \frac{\widetilde{R}^{(1)}(x,y)(y^\eta - v^\eta)^2}{(yv)^{2\eta}} + \frac{2\delta y^{2\eta}}{c(yv)^{2\eta}} \\ &\le \frac{1}{c} y^{1-4\eta}(v^\eta - y^\eta)^2 + \frac{2\delta^{1-2\eta}}{c} \\ &\le \frac{1}{c}(y^{1-4\eta} v^{2\eta-2}(v-y)^2 + 2\delta^{1-2\eta}) \\ &\le \frac{1}{c}(y^{-1-2\eta}(v-y)^2 + 2\delta^{1-2\eta}) \le \frac{3\delta^{1-2\eta}}{c}. \end{aligned}$$

6. If $v > \delta, x < u, y < v, y < \delta$, we have $\delta < v < 2\delta$, and

$$\frac{\widetilde{R}^{(1)}(u,v)}{v^{2\eta}} \le \frac{v^{1-2\eta}}{c} < \frac{2\delta^{1-2\eta}}{c}, \quad \widetilde{R}^{(1)}(x,y) \le \frac{y}{c}.$$

It follows that

$$d^2(f_{x,y,1}, f_{u,v,1}) \le \frac{1}{c}(y^{1-2\eta} + 2y^{1-2\eta} + 2\delta^{1-2\eta}) \le \frac{5\delta^{1-2\eta}}{c}.$$

7. If $v > \delta, x \ge u, \delta \le y < v$, we have that

$$\widetilde{R}^{(1)}(x,y) \le \widetilde{R}^{(1)}(u,y) + \frac{\delta}{c}, \quad \widetilde{R}^{(1)}(u,v) \le \widetilde{R}^{(1)}(u,y) + \frac{\delta}{c}.$$

Thus,

$$\begin{aligned} d^2(f_{x,y,1}, f_{u,v,1}) &\le \frac{\widetilde{R}^{(1)}(u,y)(y^\eta - v^\eta)^2}{(yv)^{2\eta}} + \frac{\delta y^{2\eta}}{c(yv)^{2\eta}} + \frac{\delta v^{2\eta}}{c(yv)^{2\eta}} \\ &\le \frac{1}{c} y^{1-4\eta}(v^\eta - y^\eta)^2 + \frac{2\delta^{1-2\eta}}{c} \\ &\le \frac{1}{c}(y^{1-4\eta} v^{2\eta-2}(v-y)^2 + 2\delta^{1-2\eta}) \\ &\le \frac{1}{c}(y^{-1-2\eta}(v-y)^2 + 2\delta^{1-2\eta}) \le \frac{3\delta^{1-2\eta}}{c}. \end{aligned}$$

8. If $v > \delta, x \geq u, y < v, y < \delta$, we have that

$$\widetilde{R}^{(1)}(u,v) \leq \widetilde{R}^{(1)}(x,v) \leq \widetilde{R}^{(1)}(x,y) + \frac{\delta}{c} \leq \frac{2\delta}{c},$$

and hence

$$\begin{aligned} d^2(f_{x,y,1}, f_{u,v,1}) &\leq \frac{1}{c}(y^{1-2\eta} + 2y^{1-2\eta} + 2\delta^{1-2\eta}) \\ &\leq \frac{5\delta^{1-2\eta}}{c}. \end{aligned}$$

From the above analysis, and under the condition $1 - 2\eta > 0$, we conclude that for any $\epsilon > 0$, there exists $\delta > 0$ such that for all $|x-u| < \delta$ and $|y-v| < \delta$, we have that $d^2(f_{x,y,1}, f_{u,v,1}) < \epsilon$. Therefore, $(\mathcal{F}_{2,1}, d)$ is totally bounded, since $[0,T]^2$ is totally bounded in Euclidean space.

Step 2. We prove the asymptotic tightness of $\{\sum_{i=1}^n [H_{n,i}(f) - \mathbb{E}H_{n,i}(f)], f \in \mathcal{F}_{2,1}\}$.

For any $\epsilon > 0$, define the bracketing number $N_{[]}(\epsilon, \mathcal{F}_{2,1})$ as the minimal number of sets N_ϵ in a partition of $\mathcal{F}_{2,1} = \cup_{j=1}^{N_\epsilon} F_{2,1}^{j,\epsilon}$, such that for each partitioning set $F_{2,1}^{j,\epsilon}$, we have

$$\sum_{i=1}^n \mathbb{E}^* \sup_{f,g \in \mathcal{F}_{2,1}^{j,\epsilon}} |H_{n,i}(f) - H_{n,i}(g)|^2 \leq \epsilon^2,$$

where $\mathbb{E}^*$ denotes the outer expectation (Chapter 1 in van der Vaart and Wellner (1996)). By Theorem 2.11.9 in van der Vaart and Wellner (1996), we only need to prove that for any $\lambda > 0$,

$$\sum_{i=1}^n \mathbb{E}^* \|H_{n,i}\|_{\mathcal{F}_{2,1}} \mathcal{I}\{\|H_{n,i}\|_{\mathcal{F}_{2,1}} > \lambda\} \to 0, \tag{S5.4}$$

and for every $\delta_n \to 0$

$$\int_0^{\delta_n} \sqrt{\log N_{[]}(\epsilon, \mathcal{F}_{2,1})} d\epsilon \to 0, \tag{S5.5}$$

where $\|H_{n,i}\|_{\mathcal{F}_{2,1}} = \sup_{f \in \mathcal{F}_{2,1}} |H_{n,i}(f)|$.

Define $(U_i, V_i) = (1 - F_X(X_i), 1 - F_{Y(1)}(Y_i(1)))$. Under Assumption (a), for $\eta \in (0, 1/2)$, we have that

$$\begin{aligned}
&\sup_{x,y \in (0,T]} \frac{1}{\sqrt{k}} \frac{D_i}{\Pi(\boldsymbol{Z_i})} \mathcal{I}\{1 - F_X(X_i) < (k/n)x, 1 - F_{Y(1)}(Y_i) < (k/n)y\}/y^{\eta} \\
\leq& \frac{1}{c\sqrt{k}} \frac{1}{(\frac{n}{k}(1 - F_{Y(1)}(Y_i(1))))^{\eta}} = \frac{1}{c\sqrt{k}} \frac{1}{(\frac{n}{k} V_i)^{\eta}}.
\end{aligned}$$

Then for $\eta \in (0, 1/2)$, we have that

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbb{E}^* \|H_{n,i}\|_{\mathcal{F}_{2,1}} \mathcal{I}\{\|H_{n,i}\|_{\mathcal{F}_{2,1}} > \lambda\} \\
\leq& \frac{n}{c\sqrt{k}} \mathbb{E}\left(\frac{1}{\left(\frac{nV_1}{k}\right)^{\eta}} \mathcal{I}\left\{ \frac{n}{k} V_1 \leq (\sqrt{k}\lambda c)^{-1/\eta} \right\} \right) \\
=& \frac{n}{c\sqrt{k}} \left(\frac{n}{k}\right)^{-\eta} \int_0^{(\sqrt{k}\lambda c)^{-1/\eta} \frac{k}{n}} x^{-\eta} dx \\
=& \frac{1}{c(1-\eta)} \frac{n}{\sqrt{k}} \frac{k}{n} k^{\frac{\eta-1}{2\eta}} (\lambda c)^{\frac{\eta-1}{\eta}} = \frac{(\lambda c)^{\frac{\eta-1}{\eta}}}{c(1-\eta)} k^{1-\frac{1}{2\eta}} \to 0.
\end{aligned}$$

For $\eta = 0$, since $\mathbb{E}^* \|H_{n,i}\|_{\mathcal{F}_{2,1}} \leq \frac{1}{c\sqrt{k}}$ by $\|H_{n,i}\|_{\mathcal{F}_{2,1}} \leq \frac{1}{c\sqrt{k}}$, it follows that for any $\lambda > 0$, by taking $k > \lambda^{-2}$, we have that $\mathcal{I}\{\|H_{n,i}\|_{\mathcal{F}_{2,1}} > \lambda\} = 0$ (for $i = 1, \ldots, n$).

Thus, we conclude that (S5.4) holds for $\eta \in [0, 1/2)$.

Next, we prove (S5.5). Without loss of generality, we set $T = 1$. Let $\epsilon > 0$ be a small constant, to be chosen later. Define

$$\xi = \epsilon^{\frac{3}{1-2\eta}}, \quad \theta = 1 - \epsilon^3, \quad \tilde{\theta} = 1 - \xi.$$

Note that, we can choose ϵ sufficiently small and C sufficiently large, such that,

$$\frac{4}{c^2(1-2\eta)}\xi^{1-2\eta} \leq \epsilon^2, \tag{S5.6}$$

$$\frac{\tilde{\theta}^l(1-\tilde{\theta})}{\theta^{2\eta(m+1)}} = \epsilon^{\frac{3}{1-2\eta}}\tilde{\theta}^l\theta^{-2\eta(m+1)} \leq C\epsilon^{\frac{3}{1-2\eta}}\xi^{-2\eta} = C\epsilon^{\frac{3-6\eta}{1-2\eta}} = C\epsilon^3, \tag{S5.7}$$

$$(1-\epsilon^3)^{-2\eta} \leq (1-\epsilon^3)^{-1} < 1/2, \quad \frac{1}{\sqrt{1-\epsilon^3}} - 1 \leq C\epsilon^3, \tag{S5.8}$$

for $l = 1, 2, \ldots, \lceil \log\xi/\log\tilde{\theta}\rceil, m = 1, 2, \ldots, \lceil \log\xi/\log\theta\rceil$.

Define

$$\mathcal{F}^{(1)}(\xi) = \{f_{x,y,1} \in \mathcal{F}_{2,1} : y \leq \xi\},$$

$$\mathcal{F}^{(2)}(\xi) = \{f_{x,y,1} \in \mathcal{F}_{2,1} : y > \xi, x \leq \xi\},$$

$$\mathcal{F}(l,m) = \{f_{x,y,1} \in \mathcal{F}_{2,1} : \tilde{\theta}^{l+1} \leq x \leq \tilde{\theta}^l, \theta^{m+1} \leq y \leq \theta^m\}.$$

Then, we have that,

$$\mathcal{F}_{2,1} = \mathcal{F}^{(1)}(\xi) \cup \mathcal{F}^{(2)}(\xi) \cup \{\cup_{m=0}^{\lceil \log\xi/\log\theta\rceil} \cup_{l=0}^{\lceil \log\xi/\log\tilde{\theta}\rceil} \mathcal{F}(l,m)\}.$$

For $\mathcal{F}^{(1)}$, we have that

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbb{E}^{*} \sup_{f,g \in \mathcal{F}^{(1)}(\xi)} (H_{n,i}(f) - H_{n,i}(g))^2 \\
=& n\mathbb{E}^{*} \sup_{f,g \in \mathcal{F}^{(1)}(\xi)} (H_{n,1}(f) - H_{n,1}(g))^2 \\
\leq& 4n\mathbb{E}^{*} \sup_{f \in \mathcal{F}^{(1)}(\xi)} H_{n,1}^2(f) \\
\leq& 4n\mathbb{E}^{*} \sup_{x,y>0, y\leq \xi} H_{n,1}^2(f) \\
\leq& \frac{4n}{kc^2} \mathbb{E} \left(\frac{n}{k}V_1\right)^{-2\eta} \mathcal{I}\left\{\frac{nV_1}{k} < \xi\right\} \\
=& \frac{4n}{kc^2} \int_0^{\frac{k\xi}{n}} \left(\frac{n}{k}x\right)^{-2\eta} dx = \frac{4}{c^2(1-2\eta)} \xi^{1-2\eta} \leq \epsilon^2,
\end{aligned}$$

where the last inequality follows from (S5.6).

For $\mathcal{F}^{(2)}$, we have that

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbb{E}^{*} \sup_{f,g \in \mathcal{F}^{(2)}(\xi)} (H_{n,i}(f) - H_{n,i}(g))^2 \\
=& n\mathbb{E}^{*} \sup_{f,g \in \mathcal{F}^{(2)}(\xi)} (H_{n,1}(f) - H_{n,1}(g))^2 \\
\leq& 4n\mathbb{E}^{*} \sup_{f \in \mathcal{F}^{(2)}(\xi)} H_{n,1}^2(f) \\
\leq& 4n\mathbb{E}^{*} \sup_{y>\xi, 0<x\leq \xi} H_{n,1}^2(f) \\
\leq& \frac{4n}{kc^2} \xi^{(-2\eta)} \mathbb{E}\mathcal{I}\left\{\frac{nU_1}{k} < \xi\right\} \\
=& \frac{4n}{kc^2} \xi^{(-2\eta)} \int_0^{\frac{k\xi}{n}} dx = \frac{4}{c^2} \xi^{1-2\eta} \leq \epsilon^2,
\end{aligned}$$

where the last inequality follows from (S5.6).

For $\mathcal{F}(l,m)$, note that

$$
\begin{aligned}
&\mathbb{E}\Bigg(\mathcal{I}\{U_1 < k\tilde{\theta}^{l+1}/n, V_1 < k\theta^{m+1}/n\}\left(\frac{1}{\theta^{\eta m}} - \frac{1}{\theta^{\eta(m+1)}}\right) \\
&\quad + \Big(\mathcal{I}\{U_1 < k\tilde{\theta}^{l+1}/n, V_1 < k\theta^{m+1}/n\} - \mathcal{I}\{U_1 < k\tilde{\theta}^{l}/n, V_1 < k\theta^{m}/n\}\Big)\frac{1}{\theta^{\eta(m+1)}}\Bigg)^2 \\
=&\mathbb{E}\Bigg(\mathcal{I}\{U_1 < k\tilde{\theta}^{l+1}/n, V_1 < k\theta^{m+1}/n\}\left(\frac{1}{\theta^{\eta(m+1)}} - \frac{1}{\theta^{\eta m}}\right)^2\Bigg) \\
&\quad + \mathbb{E}\Bigg(\Big(\mathcal{I}\{U_1 < k\tilde{\theta}^{l}/n, V_1 < k\theta^{m}/n\} - \mathcal{I}\{U_1 < k\tilde{\theta}^{l+1}/n, V_1 < k\theta^{m+1}/n\}\Big)\frac{1}{\theta^{2\eta(m+1)}}\Bigg).
\end{aligned}
$$

It follows that,

$$
\begin{aligned}
&\sum_{i=1}^{n}\mathbb{E}^* \sup_{f,g\in\mathcal{F}(l,m)}(H_{n,i}(f) - H_{n,i}(g))^2 \\
\leq & n\mathbb{E}^*\left(\sup_{f\in\mathcal{F}(l,m)} H_{n,1}(f) - \inf_{f\in\mathcal{F}(l,m)} H_{n,1}(f)\right)^2 \\
\leq & \frac{n}{c^2k}\mathbb{E}\left(\frac{\mathcal{I}\{U_1 < k\tilde{\theta}^{l}/n, V_1 < k\theta^{m}/n\}}{(\theta^{m+1})^{\eta}} - \frac{\mathcal{I}\{U_1 < k\tilde{\theta}^{l+1}/n, V_1 < k\theta^{m+1}/n\}}{(\theta^{m})^{\eta}}\right)^2 \\
= & \frac{n}{c^2k}\mathbb{E}\Bigg(\mathcal{I}\{U_1 < k\tilde{\theta}^{l+1}/n, V_1 < k\theta^{m+1}/n\}\left(\frac{1}{\theta^{\eta(m+1)}} - \frac{1}{\theta^{\eta m}}\right)^2\Bigg) \\
&\quad + \mathbb{E}\Bigg(\Big(\mathcal{I}\{U_1 < k\tilde{\theta}^{l}/n, V_1 < k\theta^{m}/n\} - \mathcal{I}\{U_1 < k\tilde{\theta}^{l+1}/n, V_1 < k\theta^{m+1}/n\}\Big)\frac{1}{\theta^{2\eta(m+1)}}\Bigg) \\
= & \frac{n}{kc^2}\Bigg(C^{(1)}\left(\frac{k}{n}\tilde{\theta}^{l+1}, \frac{k}{n}\theta^{m+1}\right)\frac{1}{\theta^{2\eta m}}\left(\frac{1}{\theta^{\eta}} - 1\right)^2 \\
&\quad + \left[C^{(1)}\left(\frac{k}{n}\tilde{\theta}^{l}, \frac{k}{n}\theta^{m}\right) - C^{(1)}\left(\frac{k}{n}\tilde{\theta}^{l+1}, \frac{k}{n}\theta^{m+1}\right)\right]\frac{1}{\theta^{2\eta(m+1)}}\Bigg) \\
\leq & \frac{n}{c^2k}\left[\frac{k\theta^m}{n\theta^{2\eta m}}\left(\frac{1}{\theta^{\eta}} - 1\right)^2 + \frac{k\theta^m}{n\theta^{2\eta(m+1)}}(1-\theta) + \frac{k\tilde{\theta}^l(1-\tilde{\theta})}{n\theta^{2\eta(m+1)}}\right].
\end{aligned}
$$

By using (S5.7) and (S5.8), we have that

$$
\sum_{i=1}^{n}\mathbb{E}^* \sup_{f,g\in\mathcal{F}(l,m)}(H_{n,i}(f) - H_{n,i}(g))^2 \leq \frac{2(\theta^{-1/2}-1)^2 + 2(1-\theta) + C\epsilon^3}{c^2} \leq \epsilon^2.
$$

Consequently, we obtain the partition $\mathcal{F}_{2,1} = \mathcal{F}^{(1)}(\xi) \cup \mathcal{F}^{(2)}(\xi) \cup \{\cup_{m=0}^{[\log \xi / \log \theta]} \cup_{l=0}^{[\log \xi / \log \tilde{\theta}]} \mathcal{F}(l,m)\}$ and the covering number of this partition is bounded by the order of $\epsilon^{-4-3/(1-2\eta)}$. Therefore (S5.5) holds, and asymptotic tightness follows.

Step 3. We prove the weak convergence of the finite-dimensional distributions of the process $\{\sum_{i=1}^{n}[H_{n,i}(f) - \mathbb{E}H_{n,i}(f)], f \in \mathcal{F}_{2,1}\}$.

Note that, for sufficient large n and any $f \in \mathcal{F}_{2,1}$,

$$\mathcal{I}\{|H_{n,i}(f)| > \lambda\} = 0, \quad i = 1, 2, \ldots, n.$$

Hence, we have that

$$\sum_{i=1}^{n} \mathbb{E}[|H_{n,i}(f) - \mathbb{E}H_{n,i}(f)|^2 \mathcal{I}\{|H_{n,i}(f)| > \lambda\}] \to 0,$$

for all $\lambda > 0$. Then, by applying the Lindeberg-Feller central limit theorem, we can prove that, the finite-dimensional distributions of our process converge weakly. $\square$

S6 Proofs of main results

Proof of Lemma 1. Using the property of conditional expectation and Assumption (a), we obtain that

$$
\begin{aligned}
&\frac{1}{p}\mathbb{E}\left[\frac{YD\mathcal{I}\{X>Q_X(1-p)\}}{\Pi(\boldsymbol{Z})}\right],\\
=&\frac{1}{p}\mathbb{E}\left[\mathbb{E}\left[\frac{Y(1)D\mathcal{I}\{X>Q_X(1-p)\}}{\Pi(\boldsymbol{Z})}\middle|Y(1),Z\right]\right]\\
=&\frac{1}{p}\mathbb{E}\left[\frac{Y(1)\mathcal{I}\{X>Q_X(1-p)\}}{\Pi(\boldsymbol{Z})}\mathbb{E}[D|Y(1),Z]\right]\\
=&\frac{1}{p}\mathbb{E}\left[\frac{Y(1)\mathcal{I}\{X>Q_X(1-p)\}}{\Pi(\boldsymbol{Z})}\mathbb{E}[D|Z]\right]\\
=&\frac{1}{p}\mathbb{E}\left[Y(1)\mathcal{I}\{X>Q_X(1-p)\}\right]\\
=&\theta_p^{(1)}.
\end{aligned}
$$

For $\theta_p^{(0)}$, the proof is similar and thus omitted. Thus, Lemma 1 holds. $\square$

Before proving the asymptotic normality of $\widehat{\theta}_{k/n}^{(j)}$ in Theorem 1, we establish the asymptotic behavior of the following pseudo estimators:

$$
\overline{\theta}_{k/n}^{(1)}=\frac{1}{k}\sum_{i=1}^{n}\frac{Y_iD_i}{\Pi(\boldsymbol{Z_i})}\mathcal{I}\{X_i>X_{n-k,n}\},\quad \overline{\theta}_{k/n}^{(0)}=\frac{1}{k}\sum_{i=1}^{n}\frac{Y_i(1-D_i)}{1-\Pi(\boldsymbol{Z_i})}\mathcal{I}\{X_i>X_{n-k,n}\}.
$$

Define

$$
\begin{aligned}
\widetilde{\theta}_{kx/n}^{(1)}&=\frac{1}{kx}\sum_{i=1}^{n}\frac{D_iY_i}{\Pi(\boldsymbol{Z_i})}\mathcal{I}\{X_i>U_X(n/(kx))\},\\
\widetilde{\theta}_{kx/n}^{(0)}&=\frac{1}{kx}\sum_{i=1}^{n}\frac{(1-D_i)Y_i}{1-\Pi(\boldsymbol{Z_i})}\mathcal{I}\{X_i>U_X(n/(kx))\}.
\end{aligned}
$$

Then, we have that,

$$\overline{\theta}^{(j)}_{k/n} = e_n \widetilde{\theta}^{(j)}_{ke_n/n}, \quad j = 0, 1,$$

where

$$e_n := \frac{n}{k} \{1 - F_X(X_{n-k,n})\} \xrightarrow{P} 1.$$

We intend to establish the asymptotic behavior of $\widetilde{\theta}^{(j)}_{kx/n}$, on the interval $[1/2, 2]$. Then by applying the result for $x = e_n$, we can obtain the asymptotic normality of $\overline{\theta}^{(j)}_{k/n}$. The asymptotic behavior of $\widetilde{\theta}^{(1)}_{kx/n}$ and $\widetilde{\theta}^{(0)}_{kx/n}$ are established in the following proposition.

Proposition S1. *Suppose that conditions* (2.1) *and* (2.2)*, Assumptions (a) and (d) hold with* $\gamma_j \in (0, 1/2)$*. Then, we have that, as* $n \to \infty$*, for* $j = 0, 1$,

$$\sup_{1/2 \le x \le 2} \left| \frac{\sqrt{k}}{U_{Y(j)}(n/k)} (\widetilde{\theta}^{(j)}_{kx/n} - \theta^{(j)}_{kx/n}) + \frac{1}{x} \int_0^\infty W_{\widetilde{R}^{(j)},n}(x, s) ds^{-\gamma_j} \right| \xrightarrow{P} 0.$$

Proof. We only provide the proof for $j = 1$. The proof for $j = 0$ is similar and thus omitted. Simple calculation yields

$$x\widetilde{\theta}^{(1)}_{kx/n} = -U_{Y(1)}(n/k) \int_0^\infty T^{(1)}_n(x, s_n(y)) dy^{-\gamma_1},$$

where

$$s_n(y) = (n/k)[1 - F_{Y(1)}(U_{Y(1)}(n/k) y^{-\gamma_1})], \quad y > 0.$$

For any fixed $T > 0$, we have

$$
\begin{aligned}
&\sup_{1/2\le x\le 2}\left|\frac{\sqrt{k}}{U_{Y(1)}(n/k)}(x\widetilde{\theta}^{(1)}_{kx/n} - x\theta^{(1)}_{kx/n}) + \int_0^\infty W_{\widetilde{R}^{(1)},n}(x,s)ds^{-\gamma_1}\right| \\
= &\sup_{1/2\le x\le 2}\left|\int_0^\infty \sqrt{k}[T_n^{(1)}(x,s_n(y)) - R_n^{(1)}(x,s_n(y))]dy^{-\gamma_1} - \int_0^\infty W_{\widetilde{R}^{(1)},n}(x,y)dy^{-\gamma_1}\right| \\
\le &\sup_{1/2\le x\le 2}\left|\int_T^\infty W_{\widetilde{R}^{(1)},n}(x,y)dy^{-\gamma_1}\right| + \sup_{1/2\le x\le 2}\left|\int_T^\infty \sqrt{k}[T_n^{(1)}(x,s_n(y)) - R_n^{(1)}(x,s_n(y))]dy^{-\gamma_1}\right| \\
&+ \sup_{1/2\le x\le 2}\left|\int_0^T \sqrt{k}[T_n^{(1)}(x,s_n(y)) - R_n^{(1)}(x,s_n(y))] - W_{\widetilde{R}^{(1)},n}(x,y)dy^{-\gamma_1}\right| \\
=: &\, I_1(T) + I_{2,n}(T) + I_{3,n}(T).
\end{aligned}
$$

We intend to prove that, for any $\epsilon > 0$, there exist $T_0 = T_0(\epsilon), n_0 = n_0(T_0)$, such that for $n > n_0$,

$$
\Pr\{I_1(T) > \epsilon\} \le \epsilon, \quad \Pr\{I_{2,n}(T) > \epsilon\} \le \epsilon, \quad \Pr\{I_{3,n}(T) > \epsilon\} \le \epsilon.
$$

For $I_1(T)$, by Lemma S3, we have that for any $\epsilon > 0$, there exists $T_1 = T_1(\epsilon)$ such that

$$
\Pr\{\sup_{1/2\le x\le 2, 0<y<\infty} |W_{\widetilde{R}^{(1)},n}(x,y)| > T_1^{\gamma_1}\epsilon\} < \epsilon.
$$

So for any $T > T_1$,

$$
\Pr\{I_1(T) > \epsilon\} < \epsilon.
$$

Next, we handle $I_{2,n}(T)$. Let $\widetilde{P}, \widetilde{P}_n$ be the probability measure and empirical probability measure defined by $(1 - F_X(X), 1 - F_{Y(1)}(Y(1)))$ and $\{1 - F_X(X_i), 1 - F_{Y(1)}(Y_i(1))\}_{i=1}^n$, respectively.

Without loss of generality, we assume that X is the first component of $\boldsymbol{Z}$. For any x, y, define the (random) function

$$f_{(x,y)} := \frac{D}{\Pi(\boldsymbol{Z})}\mathcal{I}\{1 - F_X(X) < kx/n, 1 - F_{Y(1)}(Y) < ks_n(y)/n\}.$$

Consider the following function class $\widetilde{\mathcal{F}}_n$:

$$\widetilde{\mathcal{F}}_n := \left\{f_{(x,y)}|y > T,\, 0.5 \leq x \leq 2\right\}.$$

with envelope

$$G_n = \frac{1}{c}\mathcal{I}\{1 - F_X(X) \leq 2k/n\}.$$

It follows that,

$$\begin{aligned}
\Pr\{I_{2,n}(T) > \epsilon\} &= \Pr\left\{\sup_{1/2\leq x\leq 2}\left|\int_T^\infty \sqrt{k}[T_n^{(1)}(x, s_n(y)) - R_n^{(1)}(x, s_n(y))]dy^{-\gamma_1}\right| > \epsilon\right\} \\
&\leq \Pr\left\{\sup_{1/2\leq x\leq 2, y>T} |\sqrt{k}[T_n^{(1)}(x, s_n(y)) - R_n^{(1)}(x, s_n(y))]| > \epsilon T^{\gamma_1}\right\} \\
&= \Pr\left\{\sup_{f\in\widetilde{\mathcal{F}}_n} |\sqrt{n}(\widetilde{P}_n - \widetilde{P})(f)| > \epsilon T^{\gamma_1}\sqrt{k/n}\right\} \\
&\leq \frac{\sqrt{n}\mathbb{E}^*[\sup_{f\in\widetilde{\mathcal{F}}_n} |\sqrt{n}(\widetilde{P}_n - P)(f)|]}{\epsilon T^{\gamma_1}\sqrt{k}},
\end{aligned}$$

where the last inequality above comes from Markov inequality. By Theorem 2.14.2 in van der Vaart and Wellner (1996), we have the following inequality:

$$\mathbb{E}^*\|\sup_{f\in\widetilde{\mathcal{F}}_n} |\sqrt{n}(P_n - P)(f)|\| \lesssim \int_0^1 \sqrt{1 + \log N_{[]}(\epsilon\|G_n\|_{P,2}, \widetilde{\mathcal{F}}_n, L_2(P))}d\epsilon\|G_n\|_{P,2},$$

where

$$\|G_n\|_{P,2} = (\mathbb{E}G_n^2)^{1/2} = \left(\frac{2k}{c^2 n}\right)^{1/2}.$$

Thus, if the Dudley entropy integral $\int_0^1 \sqrt{1+\log N_{[]}(\epsilon\|G_n\|_{P,2},\widetilde{\mathcal{F}}_n,L_2(P))}d\epsilon$ exists and is bounded, then there exists constants $C_1>0, C_1'>0$, such that

$$\mathbb{E}^*[\sup_{f\in\widetilde{\mathcal{F}}_n}|\sqrt{n}(P_n-P)(f)|]\leq C_1\|G_n\|_{P,2},\quad \Pr\{I_{2,n}(T)>\epsilon\}\leq\frac{C_1'}{\epsilon T_1^{\gamma}},$$

and hence there exists $T_2(\epsilon)$ such that for $T>T_2(\epsilon)$,

$$\Pr\{I_{2,n}(T)>\epsilon\}\leq\epsilon. \tag{S6.9}$$

Now, we prove that the Dudley entropy integral $\int_0^1 \sqrt{1+\log N_{[]}(\epsilon\|G_n\|_{P,2},\widetilde{\mathcal{F}}_n,L_2(P))}d\epsilon$ exists and is bounded.

Note that

$$\begin{aligned}
&\mathbb{E}|f_{(x_1,y_1)}-f_{(x_2,y_2)}|\\
\leq&\mathbb{E}\left[\frac{D}{\Pi(\boldsymbol{Z})}\mathcal{I}\left\{U\leq\frac{k(x_1\vee x_2)}{n},\frac{k(s_n(y_1)\wedge s_n(y_2))}{n}\leq V\leq\frac{k(s_n(y_1)\vee s_n(y_2))}{n}\right\}\right]\\
&+\mathbb{E}\left[\frac{D}{\Pi(\boldsymbol{Z})}\mathcal{I}\left\{\frac{k(x_1\wedge x_2)}{n}\leq U\leq\frac{k(x_1\vee x_2)}{n},V\leq\frac{k(s_n(y_1)\vee s_n(y_2))}{n}\right\}\right]\\
\leq&\frac{k}{n}(|R_n^{(1)}(2,s_n(y_1))-R_n^{(1)}(2,s_n(y_2))|\\
&+|R_n^{(1)}(x_2,s_n(y_1)\vee s_n(y_2))-R_n^{(1)}(x_1,s_n(y_1)\vee s_n(y_2))|)\\
\leq&\frac{k}{n}(|s_n(y_1)-s_n(y_2)|+|x_1-x_2|).
\end{aligned}$$

Since $s_n(y)$ is monotonically increasing and continuous for y, we have that if $|x_1-x_2|+|y_1-y_2|\to0$, then $|s_n(x_1)-s_n(x_2)|+|y_1-y_2|\to0$. Therefore, $\mathbb{E}(f_{(x,y)})$ is continuous at (x,y). Moreover, note that, $\mathbb{E}(f_{(x,y)})\leq 1/c$. Thus, for any sufficiently small $\epsilon>0$, we can choose two sequences of points

$\{x_i\}_{i=1}^{m_1}, \{y_i\}_{i=1}^{m_2}$ such that

$$0.5 = x_1 < x_2 < \cdots < x_{m_1} = 2, \quad T = y_1 < y_2 < \cdots < y_{m_2} = \infty,$$

and for all $i_1 = 2, 3, \ldots, m_1, i_2 = 2, 3, \ldots, m_2$, we have

$$0 \leq \mathbb{E}[f_{(x_{i_1},\infty)} - f_{(x_{i_1-1},\infty)}] < \frac{2k\epsilon}{cn}, \quad 0 \leq \mathbb{E}[f_{(2,y_{i_2})} - f_{(2,y_{i_2-1})}] < \frac{2k\epsilon}{cn}.$$

The integers m_1, m_2 can be chosen to satisfy that $m_1 \leq 2/\epsilon, m_2 \leq 2/\epsilon$ by $\mathbb{E}^*[\sup_{f\in\widetilde{\mathcal{F}}_n} |f_{(x,y)}|] \leq \mathbb{E}[G_n] = \frac{2k}{cn}$. Then, we have that

$$\log N_{[]}(\epsilon\|G_n\|_{P,2}, \widetilde{\mathcal{F}}_n, L_2(P)) \leq -C_2 \log \epsilon,$$

where C_2 is a positive constant that only depends on c. So the Dudley entropy integral

$$\int_0^1 \sqrt{1 + \log N_{[]}(\epsilon\|G_n\|_{P,2}, \widetilde{\mathcal{F}}_n, L_2(P))} d\epsilon$$

exists and is bounded for n. Therefore, (S6.9) holds.

Next, we consider $I_{3,n}(T)$. Note that

$$\begin{aligned}
&\Pr\{I_{3,n}(T) > \epsilon\} \\
&\leq \Pr\left\{\sup_{1/2\leq x\leq 2}\left|\int_0^T \sqrt{k}[T_n^{(1)}(x, s_n(y)) - R_n^{(1)}(x, s_n(y))] - W_{\widetilde{R}^{(1)},n}(x, s_n(y))dy^{-\gamma_1}\right| > \epsilon/2\right\} \\
&\quad + \Pr\left\{\sup_{1/2\leq x\leq 2}\left|\int_0^T W_{\widetilde{R}^{(1)},n}(x, s_n(y)) - W_{\widetilde{R}^{(1)},n}(x, y)dy^{-\gamma_1}\right| > \epsilon/2\right\} \\
&=: p_{31} + p_{32}.
\end{aligned}$$

For p_{31}, for any $T > 0$, there exists $n_1 = n_1(T)$ such that for any $n > n_1$, we have $s_n(T) < T + 1$. Then for $n > n_1$, and any $\eta_0 \in (\gamma_1, 1/2)$, we have that

$$p_{31} \leq \Pr\left\{\sup_{1/2\leq x\leq 2, 0<s\leq T+1}\left|\frac{\sqrt{k}(T_n^{(1)}(x,s) - R_n^{(1)}(x,s)) - W_{\widetilde{R}^{(1)},n}(x,s)}{s^{\eta_0}}\right|\left|\int_0^T s_n(y)^{\eta_0}dy^{-\gamma_1}\right| > \epsilon/2\right\}$$

By expression (14) of Lemma 3 in Cai et al. (2015), we have that as $n \to \infty$,

$$\left|\int_0^T s_n(y)^{\eta_0}dy^{-\gamma_1}\right| \to \frac{\gamma_1}{\eta_0 - \gamma_1}T^{\eta_0-\gamma_1}.$$

Then combined with Lemma S4, there exists $n_2(T) > n_1(T)$ such that for any $n > n_2(T)$, we have $p_{31} < \epsilon/2$.

For p_{32}, there exists $\lambda_0 = \lambda_0(\eta_0, \epsilon)$, such that

$$\Pr\left\{\sup_{1/2\leq x\leq 2, 0<y<\infty}\frac{|W_{\widetilde{R}^{(1)},n}(x,y)|}{y^{\eta_0}} \geq \lambda_0\right\} \leq \epsilon/3.$$

By Lemma 3 in Cai et al. (2015) and the continuity of $W_{\widetilde{R}^{(1)},n}(x,y)$ on $[1/2, 2] \times (0, \infty)$ with $g = W_{\widetilde{R}^{(1)}}, S = T$ and $S_0 = T + 1$, we conclude that, there exists $n_3 = n_3(T)$ such that for $n > n_3$, we have $p_{32} < \epsilon/2$. Thus there exists $n_4 = n_3(T) + n_2(T)$ such that, for $n > n_4$, we have $\Pr\{I_{3,n}(T) > \epsilon\} < \epsilon$.

To conclude, there exists $T_0(\epsilon) = T_1(\epsilon) + T_2(\epsilon)$ and $n_{\epsilon,T_0} = \sum_{s=0}^{3} n_s(T)$ such that, for $T = T_0(\epsilon)$ and $n > n_{\epsilon,T_0}$, we have $\Pr\{I_1(T) > \epsilon\} < \epsilon, \Pr\{I_{2,n}(T) > \epsilon\} < \epsilon$, and $\Pr\{I_{3,n}(T) > \epsilon\} < \epsilon$. Hence, Proposition S1 holds. $\square$

Next proposition shows asymptotic properties of $\overline{\theta}^{(j)}_{k/n}$ for $j = 0, 1$.

Proposition S2. *Suppose that Assumptions (a)-(d), (f) hold with* $\gamma_0, \gamma_1 \in (0, 1/2)$. *Then for* $j = 0, 1$,

$$\sqrt{k}\left(\frac{\overline{\theta}^{(j)}_{k/n}}{\theta^{(j)}_{k/n}} - 1\right) \xrightarrow{d} \Theta_j,$$

where

$$\Theta_j = (\gamma_j - 1)W_{R^{(j)}}(1, \infty) - \left\{\int_0^\infty R^{(j)}(1, s^{-1/\gamma_j})ds\right\}^{-1} \int_0^\infty W_{\widetilde{R}^{(j)}}(1, s)ds^{-\gamma_j}.$$

Proof. We only give the proof for $j = 1$, since the proof for $j = 0$ is analogous. Define

$$\Theta_{j,n} = (\gamma_1 - 1)W_{R^{(j)},n}(1, \infty) - \left\{\int_0^\infty R^{(j)}(1, s^{-1/\gamma_j})ds\right\}^{-1} \int_0^\infty W_{\widetilde{R}^{(j)},n}(1, s)ds^{-\gamma_j},$$

then $\Theta_{j,n} \stackrel{d}{=} \Theta_j$.

By Lemma S4, $T_n^{(3)}(e_n, \infty) = 1$ and the continuity of $W_{R^{(1)}}(1, \infty)$, we have that $\sqrt{k}(e_n - 1) = -W_{R^{(1)},n}(1, \infty) + o_p(1)$, and thus

$$\lim_{n\to\infty} \Pr\{|e_n - 1| > k^{-1/4}\} = 0.$$

By $\overline{\theta}^{(1)}_{k/n} = e_n \widetilde{\theta}^{(1)}_{ke_n/n}$ with probability 1, we have that

$$\begin{aligned}
&\frac{\sqrt{k}}{U_{Y(1)}(n/k)}(e_n\widetilde{\theta}^{(1)}_{ke_n/n} - \theta^{(1)}_{k/n}) - \Theta_{1,n}\int_0^\infty R^{(1)}(1, s^{-1/\gamma_1})ds \\
&= \left\{ \frac{\sqrt{k}}{U_{Y(1)}(n/k)}(e_n\widetilde{\theta}^{(1)}_{ke_n/n} - e_n\theta^{(1)}_{ke_n/n}) + \int_0^\infty W_{\widetilde{R}^{(1)},n}(1,s)ds^{-\gamma_1} \right\} \\
&\quad + \left\{ \frac{\sqrt{k}}{U_{Y(1)}(n/k)}(e_n\theta^{(1)}_{ke_n/n} - \theta^{(1)}_{k/n}) - W_{R^{(1)},n}(1,\infty)(\gamma_1 - 1)\int_0^\infty R^{(1)}(1, s^{-1/\gamma_1})ds \right\} \\
&=: J_1 + J_2.
\end{aligned}$$

For J_1, we have that with probability tending to 1,

$$\begin{aligned}
|J_1| \le & \sup_{|x-1|<k^{-1/4}} \left| \frac{\sqrt{k}}{U_{Y(1)}(n/k)}(x\widetilde{\theta}^{(1)}_{kx/n} - x\theta^{(1)}_{kx/n}) + \int_0^\infty W_{\widetilde{R}^{(1)},n}(x,s)ds^{-\gamma_1} \right| \\
& + \sup_{|x-1|<k^{-1/4}} \left| \int_0^\infty (W_{\widetilde{R}^{(1)},n}(x,s) - W_{\widetilde{R}^{(1)},n}(1,s))ds^{-\gamma_1} \right| \\
& =: J_{11} + J_{12}.
\end{aligned}$$

By Proposition S1, we have that $J_{11} = o_p(1)$. For J_{12}, for any $\epsilon > 0, 0 < \delta < 1$, and $\eta \in (\gamma_1, 1/2)$, we have

$$
\begin{aligned}
&\Pr\left\{\sup_{|x-1|<k^{-1/4}}\left|\int_0^\infty W_{\widetilde{R}^{(1)},n}(x,s)-W_{\widetilde{R}^{(1)},n}(1,s)ds^{-\gamma_1}\right|>\epsilon\right\}\\
\leq&\Pr\left\{\sup_{|x-1|<k^{-1/4}}\left|\int_0^\delta W_{\widetilde{R}^{(1)},n}(x,s)-W_{\widetilde{R}^{(1)},n}(1,s)ds^{-\gamma_1}\right|>\epsilon/2\right\}\\
&+\Pr\left\{\sup_{|x-1|<k^{-1/4}}\left|\int_\delta^\infty W_{\widetilde{R}^{(1)},n}(x,s)-W_{\widetilde{R}^{(1)},n}(1,s)ds^{-\gamma_1}\right|>\epsilon/2\right\}\\
\leq&\Pr\left\{\sup_{1/2\leq x\leq 2,s>0}\frac{\left|W_{\widetilde{R}^{(1)},n}(x,s)\right|}{s^\eta}>\frac{\epsilon(\eta-\gamma_1)}{4\gamma_1}\delta^{\gamma_1-\eta}\right\}\\
&+\Pr\left\{\sup_{|x-1|<k^{-1/4},s>0}|W_{\widetilde{R}^{(1)},n}(x,s)-W_{\widetilde{R}^{(1)},n}(1,s)|\delta^{-\gamma_1}>\epsilon/2\right\}\\
=:&p_{11}+p_{12}.
\end{aligned}
$$

For p_{11}, for fixed ϵ, applying Lemma S3, there exists $\delta(\epsilon)$ such that for any $\delta<\delta(\epsilon)$, we have $p_{11}<\epsilon$.

For p_{12}, as $n\to\infty$,

$$
\sup_{|x-1|<k^{-1/4},s>0}|W_{\widetilde{R}^{(1)},n}(x,s)-W_{\widetilde{R}^{(1)},n}(1,s)|\xrightarrow{P}0,
$$

by Corollary 1.11 in Adler (1990). So for fixed δ, there exists $n(\delta)$ such that for any $n>n(\delta)$, we have $p_{12}<\epsilon$.

Thus, we conclude that, $J_{12}=o_P(1)$ and hence $J_1=o_P(1)$.

For J_2, by relationship (27) in Cai et al. (2015), we have

$$
\sup_{1/2\leq x\leq 2}\sqrt{k}\left|\int_0^\infty R_n^{(1)}(x,s_n(y))-R^{(1)}(x,y)dy^{-\gamma_1}\right|\to 0.
$$

Hence,

$$\frac{\theta^{(1)}_{k/n}}{U_{Y(1)}(n/k)} = -\int_0^\infty R_n^{(1)}(1, s_n(y))dy^{-\gamma_1} = -\int_0^\infty R^{(1)}(1, y)dy^{-\gamma_1} + o(1/\sqrt{k}), \qquad \text{(S6.10)}$$

$$\frac{e_n\theta^{(1)}_{ke_n/n}}{U_{Y(1)}(n/k)} = -\int_0^\infty R_n^{(1)}(e_n, s_n(y))dy^{-\gamma_1} = -\int_0^\infty R^{(1)}(e_n, y)dy^{-\gamma_1} + o(1/\sqrt{k}).$$

By the following homogenous property of $R^{(1)}(x, y)$:

$$\int_0^\infty R^{(1)}(x, y)dy^{-\gamma_1} = x^{1-\gamma_1}\int_0^\infty R^{(1)}(1, y)dy^{-\gamma_1},$$

we have that

$$e_n\theta^{(1)}_{ke_n/n} = e_n^{1-\gamma_1}\theta^{(1)}_{k/n} + o_p\left(\frac{U_{Y(1)}(n/k)}{\sqrt{k}}\right).$$

Then

$$\begin{aligned}
&\frac{\sqrt{k}}{U_{Y(1)}(n/k)}(e_n\theta^{(1)}_{ke_n/n} - \theta^{(1)}_{k/n}) \\
=&\sqrt{k}(e_n^{1-\gamma_1} - 1)\frac{\theta^{(1)}_{k/n}}{U_{Y(1)}(n/k)} + o_p(1) \\
=&(\gamma_1 - 1)W_{R^{(1)},n}(1, \infty)\int_0^\infty R^{(1)}(1, s^{-1/\gamma_1})ds + o_p(1).
\end{aligned}$$

So we have that $J_2 \xrightarrow{P} 0$.

Thus Proposition S2 holds. □

Combining with Propositions S1 and S2, we can continue the proof of Theorem 1.

Proof of Theorem 1. Recall that $\overline{\theta}^{(1)}_{k/n} = \frac{1}{k}\sum_{i=1}^n \frac{Y_iD_i}{\Pi(\boldsymbol{Z_i})}\mathcal{I}\{X_i > X_{n-k,n}\}$, and write

$$\begin{aligned}
&\frac{\sqrt{k}}{U_{Y(1)}(n/k)}(\widehat{\theta}^{(1)}_{k/n}-\theta^{(1)}_{k/n})\\
=&\frac{\sqrt{k}}{U_{Y(1)}(n/k)}(\widehat{\theta}^{(1)}_{k/n}-\overline{\theta}^{(1)}_{k/n})+\frac{\sqrt{k}}{U_{Y(1)}(n/k)}(\overline{\theta}^{(1)}_{k/n}-\theta^{(1)}_{k/n}).
\end{aligned}$$

Note that

$$\begin{aligned}
&\frac{\sqrt{k}}{U_{Y(1)}(n/k)}(\widehat{\theta}^{(1)}_{k/n}-\overline{\theta}^{(1)}_{k/n})\\
=&\frac{\sqrt{k}}{U_{Y(1)}(n/k)}\frac{1}{k}\sum_{i=1}^{n}Y_iD_i\mathcal{I}\{X_i>X_{n-k,n}\}\left(\frac{1}{\widehat{\Pi}(\boldsymbol{Z_i})}-\frac{1}{\Pi(\boldsymbol{Z_i})}\right)\\
=&\frac{\sqrt{k}}{U_{Y(1)}(n/k)}\frac{1}{k}\sum_{i=1}^{n}\frac{Y_iD_i\mathcal{I}\{X_i>X_{n-k,n}\}}{\Pi(\boldsymbol{Z_i})}\left(\frac{\Pi(\boldsymbol{Z_i})-\widehat{\Pi}(\boldsymbol{Z_i})}{\widehat{\Pi}(\boldsymbol{Z_i})}\right).
\end{aligned}$$

Then, we obtain that

$$\begin{aligned}
&\left|\frac{\sqrt{k}}{U_{Y(1)}(n/k)}(\widehat{\theta}^{(1)}_{k/n}-\overline{\theta}^{(1)}_{k/n})\right|\\
\leq&\sup_{\boldsymbol{z}\in\mathcal{Z}}\left|\frac{\Pi(\boldsymbol{z})-\widehat{\Pi}(\boldsymbol{z})}{\widehat{\Pi}(\boldsymbol{z})}\right|\frac{\sqrt{k}}{U_{Y(1)}(n/k)}\frac{1}{k}\sum_{i=1}^{n}\frac{Y_iD_i\mathcal{I}\{X_i>X_{n-k,n}\}}{\Pi(\boldsymbol{Z_i})}\\
=&\sup_{\boldsymbol{z}\in\mathcal{Z}}\left|\frac{\Pi(\boldsymbol{z})-\widehat{\Pi}(\boldsymbol{z})}{\widehat{\Pi}(\boldsymbol{z})}\right|\frac{\sqrt{k}}{U_{Y(1)}(n/k)}\times\overline{\theta}^{(1)}_{k/n}\\
=&o_p(n^{-1/4}\sqrt{k})=o_p(1)
\end{aligned}$$

by Remark S1, the following asymptotic relationship

$$\frac{\theta^{(1)}_p}{U_{Y(1)}(1/p)}\to\int_0^\infty R^{(1)}(1,y^{-1/\gamma_1})dy,$$

and

$$\frac{\sqrt{k}}{U_{Y(1)}(n/k)}(\overline{\theta}^{(1)}_{k/n}-\theta^{(1)}_{k/n})=O_p(1)$$

by Proposition S2.

It follows that,

$$\frac{\sqrt{k}}{U_{Y(1)}(n/k)}(\widehat{\theta}^{(1)}_{k/n}-\theta^{(1)}_{k/n}) = \frac{\sqrt{k}}{U_{Y(1)}(n/k)}(\overline{\theta}^{(1)}_{k/n}-\theta^{(1)}_{k/n}) + o_p(1),$$

$$\begin{aligned}\sqrt{k}\left(\frac{\widehat{\theta}^{(1)}_{k/n}}{\theta^{(1)}_{k/n}}-1\right) &= \sqrt{k}\left(\frac{\overline{\theta}^{(1)}_{k/n}}{\theta^{(1)}_{k/n}}-1\right)+\sqrt{k}\frac{\widehat{\theta}^{(1)}_{k/n}-\overline{\theta}^{(1)}_{k/n}}{\theta^{(1)}_{k/n}} \\ &= \sqrt{k}\left(\frac{\overline{\theta}^{(1)}_{k/n}}{\theta^{(1)}_{k/n}}-1\right)+o_p(1).\end{aligned}$$

Therefore, we have that

$$\frac{\sqrt{k}}{U_{Y(1)}(n/k)}(\widehat{\theta}^{(1)}_{k/n}-\theta^{(1)}_{k/n}) = \frac{\sqrt{k}}{U_{Y(1)}(n/k)}(\overline{\theta}^{(1)}_{k/n}-\theta^{(1)}_{k/n}) + o_p(1),$$

and

$$\sqrt{k}\left(\frac{\widehat{\theta}^{(1)}_{k/n}}{\theta^{(1)}_{k/n}}-1\right)\xrightarrow{d}\Theta_1$$

by Propositions S1 and S2.

For the variance of Θ_1, we have that

$$\begin{aligned}\mathrm{Var}(\Theta_1) &= (\gamma_1-1)^2+\left\{\int_0^\infty R^{(1)}(1,s^{-1/\gamma_1})ds\right\}^{-2}\mathbb{E}\left[\left(\int_0^\infty W_{\widetilde{R}^{(1)}}(1,s)ds^{-\gamma_1}\right)^2\right] \\ &\quad -2(\gamma_1-1)\left\{\int_0^\infty R^{(1)}(1,s^{-1/\gamma_1})ds\right\}\mathrm{Cov}\left[W_{R^{(1)}}(1,\infty),\int_0^\infty W_{\widetilde{R}^{(1)}}(1,s)ds^{-\gamma_1}\right] \\ &= (\gamma_1-1)^2+\left\{\int_0^\infty R^{(1)}(1,s^{-1/\gamma_1})ds\right\}^{-2}\int_0^\infty\int_0^\infty \tilde{R}^{(1)}(1,s\wedge t)ds^{-\gamma_1}dt^{-\gamma_1} \\ &\quad -2(\gamma_1-1)\left\{\int_0^\infty R^{(1)}(1,s^{-1/\gamma_1})ds\right\}\int_0^\infty R^{(1)}(1,s)ds^{-\gamma_1} \\ &= \gamma_1^2-1-\left\{\int_0^\infty R^{(1)}(1,s^{-1/\gamma_1})ds\right\}^{-2}\int_0^\infty \widetilde{R}^{(1)}(1,s)ds^{-2\gamma_1}.\end{aligned}$$

Since the proof in $\widehat{\theta}^{(0)}_{k/n}$ is similar, the proof of Theorem 1 is complete.

□

Proof of Theorem 2. Write

$$\frac{\widehat{\theta}_p^{(j)}}{\theta_p^{(j)}} = \frac{d_n^{\widehat{\gamma}_j}}{d_n^{\gamma_j}} \frac{\widehat{\theta}_{k/n}^{(j)}}{\theta_{k/n}^{(j)}} \frac{d_n^{\gamma_j} \theta_{k/n}^{(j)}}{\theta_p^{(j)}} =: L_{1j} L_{2j} L_{3j}.$$

For L_{1j}, we have that

$$\begin{aligned}\frac{\sqrt{k}}{\log(d_n)}(\log(L_{1j})) &= \frac{\sqrt{k}}{\log(d_n)}(\log(d_n)(\widehat{\gamma}_j - \gamma_j)) \\ &= \Gamma_{nj} + o_p(1)\end{aligned}$$

by Lemma S1. Hence,

$$\frac{\sqrt{k}}{\log(d_n)}(L_{1j} - 1) = \Gamma_{nj} + o_p(1).$$

For L_{2j}, we have $L_{2j} = 1 + O_p(1/\sqrt{k})$ by Theorem 1.

For L_{3j}, under the assumption that $\sqrt{k} A_j(n/k) \to 0$, and applying

$$\frac{U_{Y(j)}(1/p)}{U_{Y(j)}(n/k) d_n^{\gamma_j}} - 1 = O(A_j(n/k))$$

by Assumption (b), we obtain that

$$\frac{U_{Y(j)}(1/p)}{U_{Y(j)}(n/k) d_n^{\gamma_j}} - 1 = o(1/\sqrt{k}).$$

Also, similar to the previous analysis in (S6.10) in Proposition S2 for $p \le k/n$, we have that

$$\frac{\theta_p^{(j)}}{U_{Y(j)}(1/p)} - \int_0^\infty R^{(j)}(1, s^{-1/\gamma_j}) ds = o(1/\sqrt{k}).$$

Then we obtain that

$$L_{3j} = \frac{\theta_{k/n}^{(j)}/U_{Y(j)}(n/k)}{\theta_p^{(j)}/U_{Y(j)}(1/p)} \frac{U_{Y(j)}(n/k) d_n^{\gamma_j}}{U_{Y(j)}(1/p)} = 1 + o(1/\sqrt{k}).$$

Finally, we obtain that

$$\begin{aligned}\frac{\widehat{\theta}_p^{(j)}}{\theta_p^{(j)}} - 1 &= L_{1j}L_{2j}L_{3j} - 1\\ &= \frac{\log(d_n)}{\sqrt{k}}\Gamma_{nj} + O_p\left(\frac{1}{\sqrt{k}}\right) + o_p\left(\frac{\log(d_n)}{\sqrt{k}}\right) + o_p\left(\frac{1}{\sqrt{k}}\right)\\ &= \frac{\log(d_n)}{\sqrt{k}}\Gamma_{nj} + o_p\left(\frac{\log(d_n)}{\sqrt{k}}\right).\end{aligned}$$

Then

$$\frac{\sqrt{k}}{\log(d_n)}\left(\frac{\widehat{\theta}_p^{(j)}}{\theta_p^{(j)}} - 1\right) = \Gamma_{nj} + o_p(1),$$

and

$$\left(\frac{\sqrt{k}}{\log(d_n)}\left(\widehat{\theta}_p^{(1)}/\theta_p^{(1)} - 1\right), \frac{\sqrt{k}}{\log(d_n)}\left(\widehat{\theta}_p^{(0)}/\theta_p^{(0)} - 1\right)\right) \xrightarrow{d} (\Gamma_1, \Gamma_0).$$

The proof of Theorem 2 is complete. $\square$

Proof of Theorem 3. Note that under Assumption (h),

$$\begin{aligned}&\beta_n(\widehat{\mathrm{CTE}}_2(p) - \mathrm{CTE}_2(p))\\ =&\frac{\sqrt{k}}{\log(d_n)}\frac{\theta_p^{(1)}}{\max\{\theta_p^{(1)},\theta_p^{(0)}\}}\frac{\widehat{\theta}_p^{(1)} - \theta_p^{(1)}}{\theta_p^{(1)}} - \frac{\sqrt{k}}{\log(d_n)}\frac{\theta_p^{(0)}}{\max\{\theta_p^{(1)},\theta_p^{(0)}\}}\frac{\widehat{\theta}_p^{(0)} - \theta_p^{(0)}}{\theta_p^{(0)}}\\ \xrightarrow{d}&\min\{1,\kappa\}\Gamma_1 - \min\{1,1/\kappa\}\Gamma_0.\end{aligned}$$

Also, for $\kappa \neq 1$, we have that

$$\begin{aligned}&\frac{\sqrt{k}}{\log(d_n)}\left(\frac{\widehat{\mathrm{CTE}}_2(p)}{\mathrm{CTE}_2(p)} - 1\right)\\ =&\frac{\sqrt{k}}{\log(d_n)}\frac{\widehat{\theta}_p^{(1)} - \theta_p^{(1)}}{\theta_p^{(1)}}\frac{\theta_p^{(1)}}{\theta_p^{(1)} - \theta_p^{(0)}} - \frac{\sqrt{k}}{\log(d_n)}\frac{\widehat{\theta}_p^{(0)} - \theta_p^{(0)}}{\theta_p^{(0)}}\frac{\theta_p^{(0)}}{\theta_p^{(1)} - \theta_p^{(0)}}\\ \xrightarrow{d}&\frac{\kappa}{\kappa - 1}\Gamma_1 - \frac{1}{\kappa - 1}\Gamma_0.\end{aligned}$$

Thus, Theorem 3 holds. $\square$

Proof of Theorem 4. For $p \le k/n$ and $j = 0, 1$, by the proof of Theorem 2 in Cai et al. (2015), we can conclude that Assumptions (a)-(d) still hold for $(X, Y(1)^+)$ and $(X, Y(0)^+)$. Furthermore, for any sufficiently small $\epsilon > 0$, we have

$$\begin{aligned}\frac{\theta_p^{(j)}}{\theta_p^{(j+)}} &= 1 + \frac{\mathbb{E}\{Y(j)|X > U_X(1/p)\}}{\theta_p^{(j+)}} \\ &= 1 + O\left\{\frac{p^{-1+(1-\tau_j)(1-\gamma_j)}}{U_{Y(j)}(1/p)}\right\} \\ &= 1 + O(p^{-\tau_j(1-\gamma_j)}).\end{aligned}$$

Then we obtain that

$$\theta_{k/n}^{(j+)}/\theta_{k/n}^{(j)} = 1 + o(1/\sqrt{k}), \quad \theta_p^{(j+)}/\theta_p^{(j)} = 1 + o(1/\sqrt{k})$$

by Assumption (j). Therefore, Theorems 1–3 still hold for $\widehat{\theta}_{k/n}^{(j+)}$ and $\widehat{\theta}_p^{(j+)}$, combined with the fact that the extreme value indices for $Y(1)^+$ and $Y(0)^+$ are the same as those for $Y(1)$ and $Y(0)$, respectively.

Thus, Theorem 4 holds. $\square$

The proof of Theorem 5 is similar to Theorem 4 in Deuber et al. (2024) and thus omitted.

S7 Additional simulations and figures

We extend our framework to accommodate 2-dimensional covariates by defining the covariate vector as $\boldsymbol{Z} = (X, V) \in \mathbb{R}^2$. Set propensity score function as $\Pi(\boldsymbol{z}) = \Pr\{D = 1 | \boldsymbol{Z} = \boldsymbol{z}\}$.

Similar to Section 3 of the main paper, we first generate the three-dimensional random vector (X, W_1, W_0) from either a t-copula or Gumbel copula distribution. Let $V \sim \mathrm{Unif}[2, 5]$ be independent of $(X, Y(1), Y(0))$, and let $U \sim \mathrm{Unif}[0, 1]$ be a uniform random variable independent of (X, W_1, W_0, V).

We set treatment assignment as $D = I\{U \leq \Pi(\boldsymbol{Z})\}$, where the propensity score function is given by

$$\Pi(\boldsymbol{z}) = \frac{e^{-0.1x+0.2v}}{1 + e^{-0.1x+0.2v}} - 0.1.$$

For $Y(1)$ and $Y(0)$, we set them as the same as in Models (1) and (2), i.e.,

(1-m) $Y(1) = 6F_{t_5}^{-1}((1 + W_1)/2)$, $Y(0) = 2F_{t_5}^{-1}((1 + W_0)/2)$.

(2-m) $Y(1) = 8/(-\log(W_1))^{1/6}$, $Y(0) = 4/(-\log(W_0))^{1/6}$.

For $i = 1, 2$, denote Model (i-t-m) and Model (i-g-m) as Model (i-m) in which (X, W_1, W_0) generated from t-copula and Gumbel copula, respectively. In Model (1-m), $Y(1)/6, Y(0)/2$ follow the absolute t_5-distribution, and $\kappa = 3$. In Model (2-m), $Y(1)/8, Y(0)/4$ follow the Fréchet(6) distribution, and $\kappa = 2$. Our target is to estimate $\mathrm{CTE}_2(p)$ at extreme levels $p = 0.005, 0.001, 0.0005$.

Figures S1-S2 show the simulation results in terms of boxplots and rMSEs in Section 3.1 for Models (1-g) and (2-g). The simulation results in terms of boxplots and rMSEs in Section 3.2 and Section S7 are shown in Figure S5-S8 and S11-S14, respectively. The additional results show patterns consistent with those in Figures 1-2 of the main paper and demonstrate the performance of our estimator under varying tail dependence structures.

Figures S15 and S16 show the empirical coverage probabilities of CIs for $\widehat{\mathrm{CTE}}_2(p)$ across different values of k in Models (1-m) and (2-m). In general, the empirical coverage probabilities of CIs for $\widehat{\mathrm{CTE}}_2(p)$ in Models (1-m) and (2-m) are lower than those in Models (1)-(2), with the performance in Models (1-t-m) and (2-t-m) being particularly inferior to that of Model (2-t). Besides the differences between the t-copula and the Gumbel copula, and those between the absulute t-distribution and the Fréchet distribution, one additional factor contributing to this phenomenon is the increased estimation error of the propensity score function $\Pi(\boldsymbol{z})$ in the multicovariate setting. Specifically, the two-dimensional Z can increase the complexity and difficulty of accurately estimating $\Pi(\boldsymbol{z})$ via sieve method. This may adversely affect the performance of our proposed $\widehat{\mathrm{CTE}}_2(p)$, thus leading to the lower empirical coverage probabilities of the corresponding CIs.

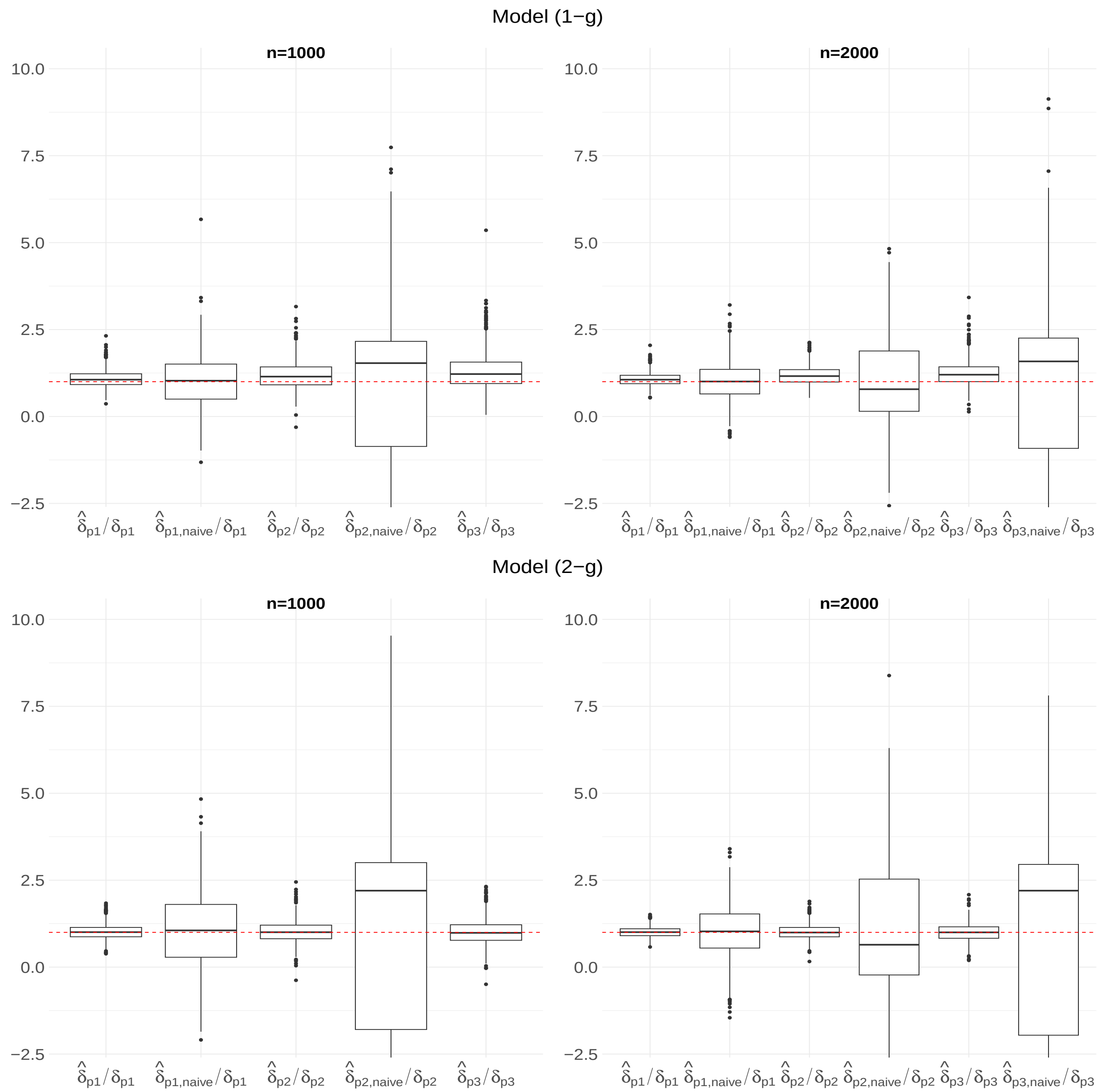

Figure S1: Boxplots for simulations under Models (1-g) and (2-g). Here, δ_p, $\widehat{\delta}_p$ and $\widehat{\delta}_{p,\text{naive}}$ denote $\text{CTE}_2(p)$, $\widehat{\text{CTE}}_2(p)$ and $\widehat{\text{CTE}}_{2,\text{naive}}(p)$, respectively. The values of p considered are $(p1, p2, p3) = (0.005, 0.001, 0.0005)$.

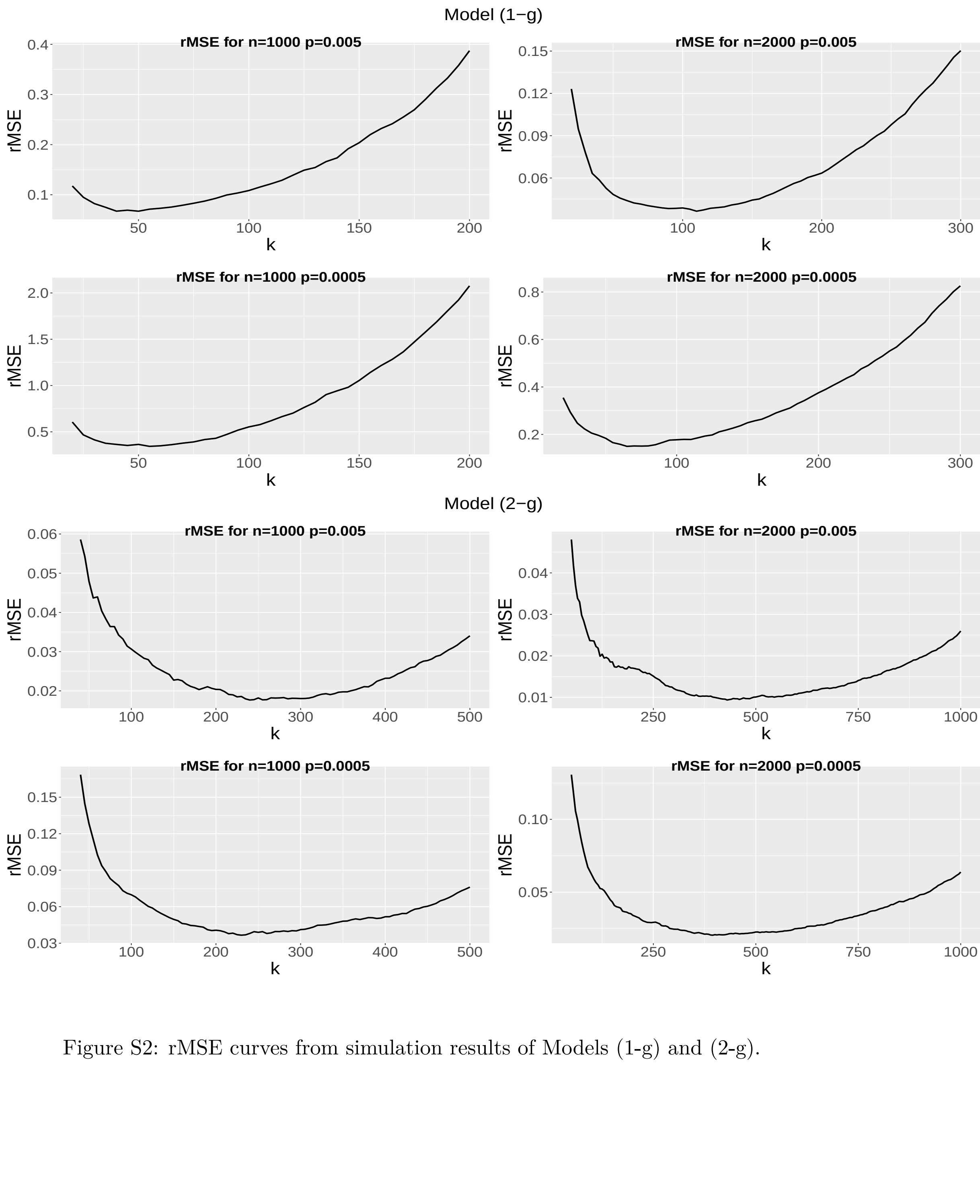

Figure S2: rMSE curves from simulation results of Models (1-g) and (2-g).

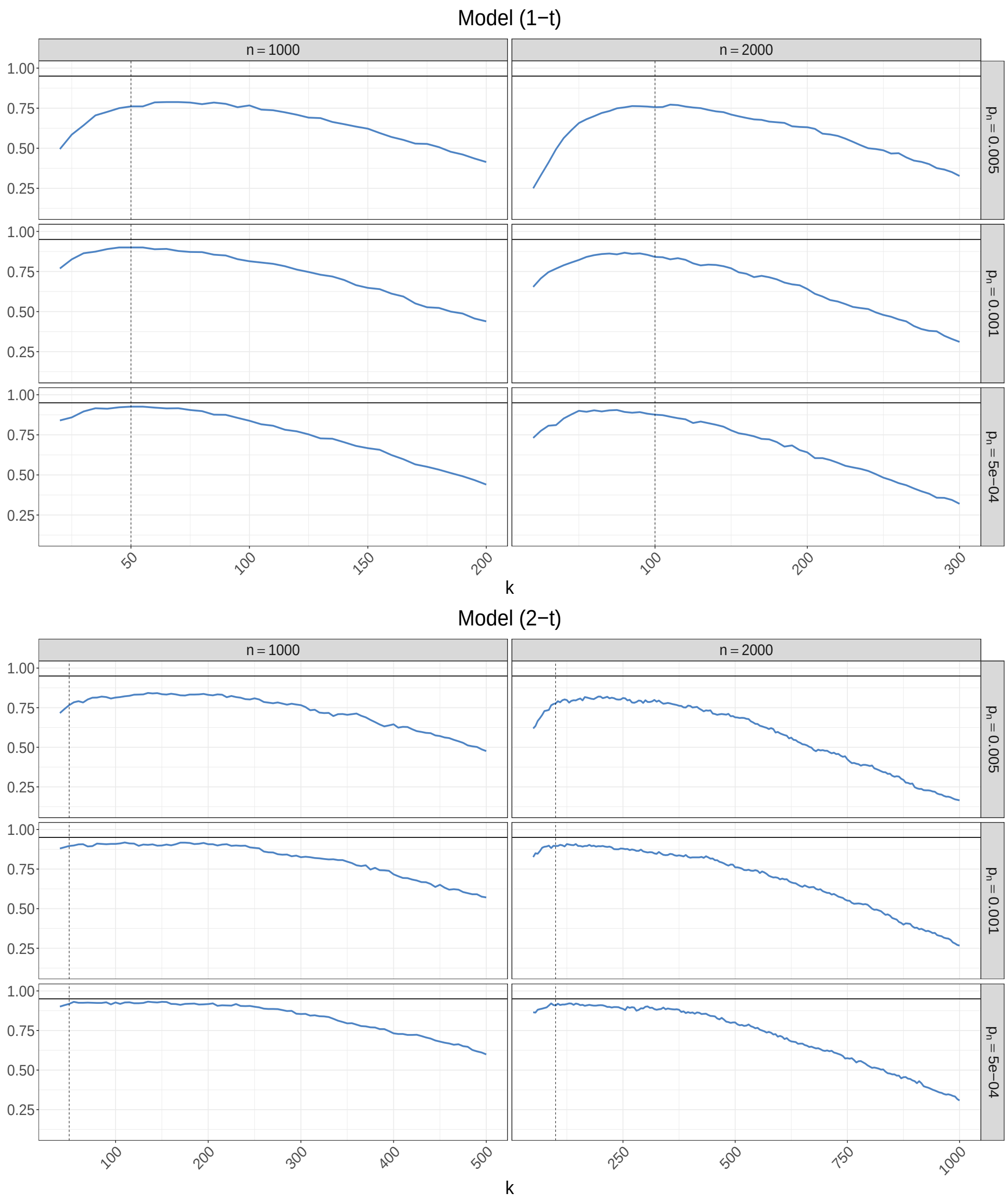

Figure S3: Plots of empirical coverage of CIs under Models (1-t) and (2-t). The vertical dashed line indicates $k = 0.05n$.

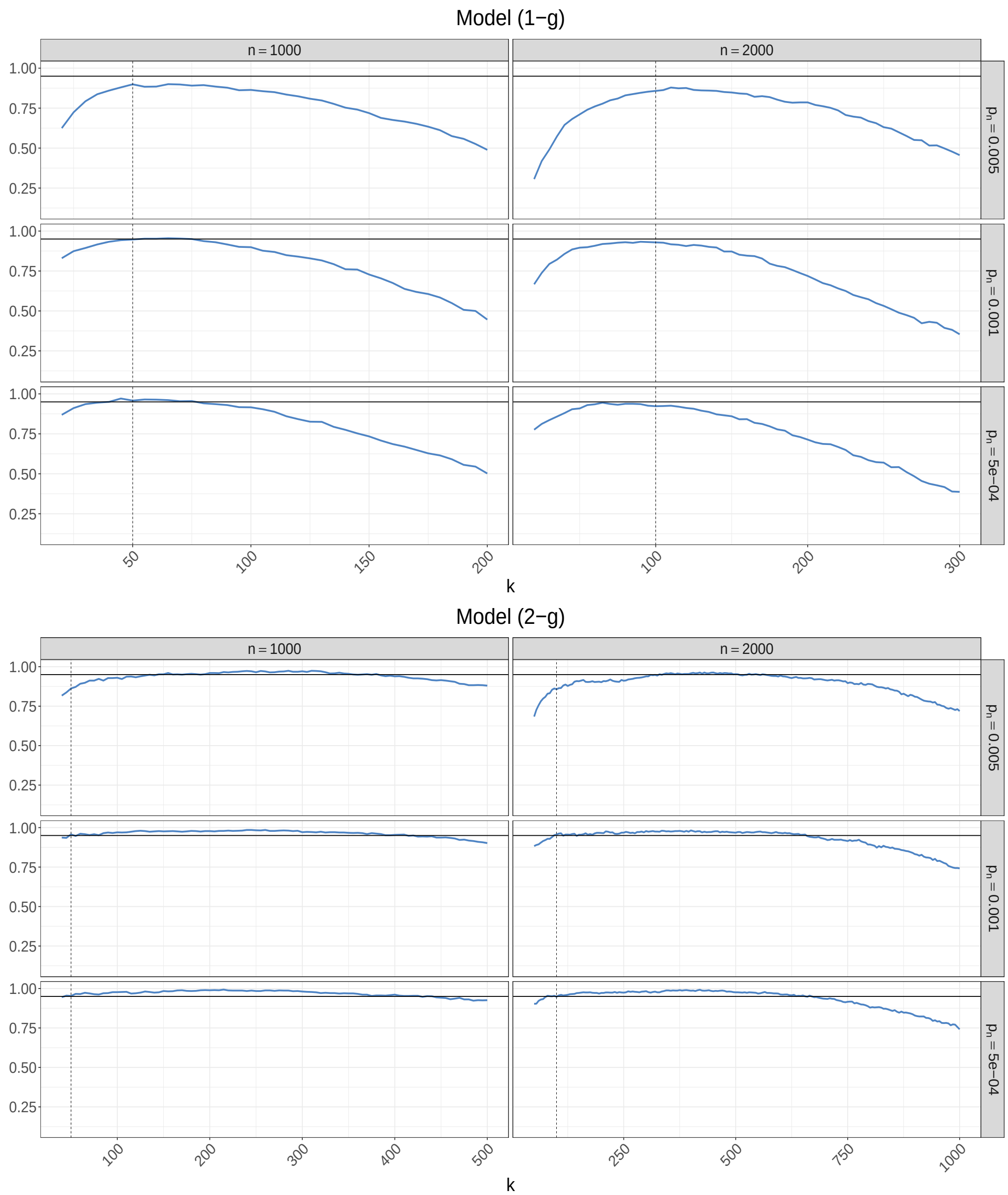

Figure S4: Plots of empirical coverage of CIs under Models (1-g) and (2-g). The vertical dashed line indicates $k = 0.05n$.

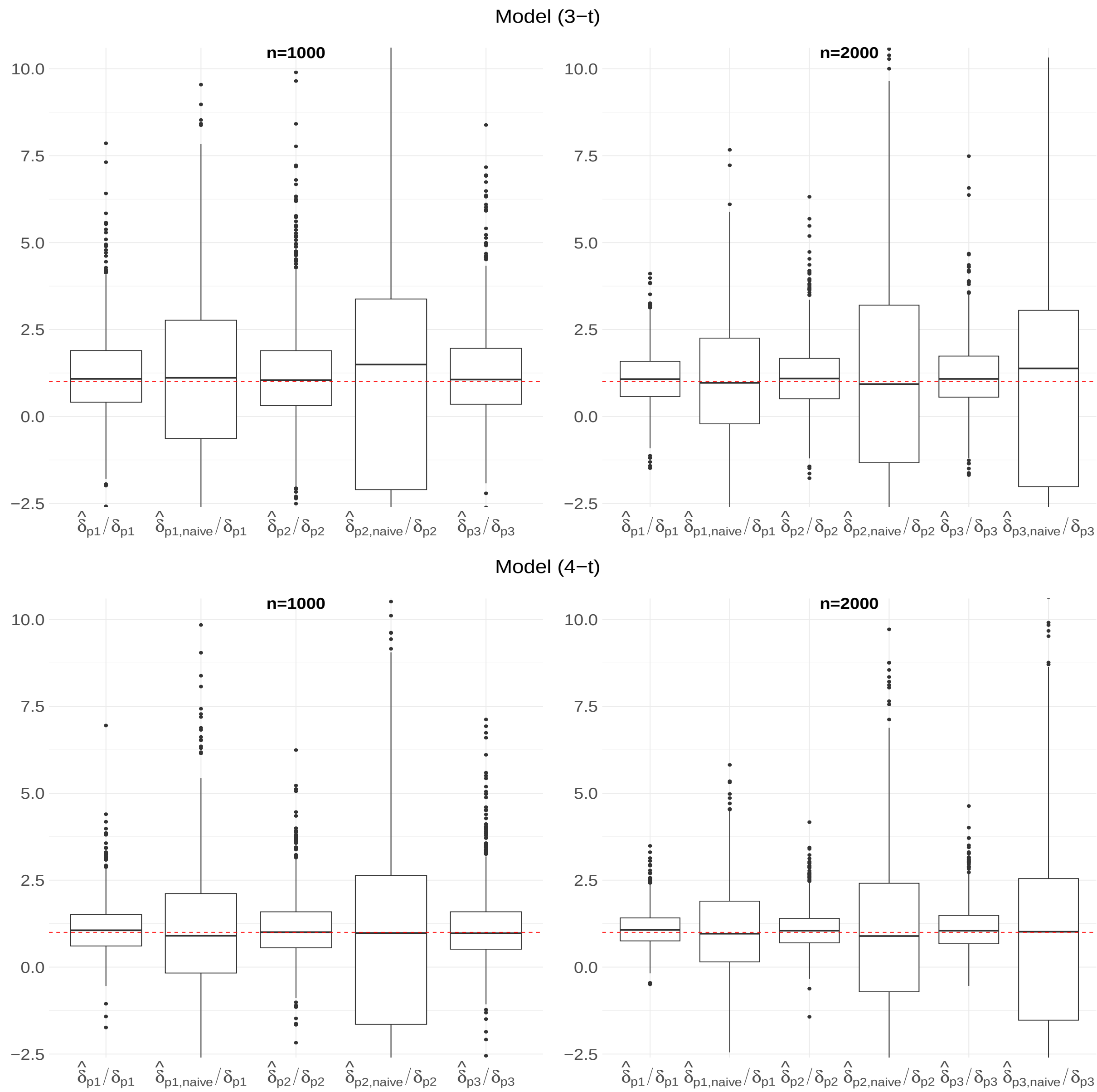

Figure S5: Boxplots for simulations under Models (3-t) and (4-t). Here, δ_p, $\widehat{\delta}_p$ and $\widehat{\delta}_{p,\mathrm{naive}}$ denote $\mathrm{CTE}_2(p)$, $\widehat{\mathrm{CTE}}_2(p)$ and $\widehat{\mathrm{CTE}}_{2,\mathrm{naive}}(p)$, respectively. The values of p considered are $(p1, p2, p3) = (0.005, 0.001, 0.0005)$.

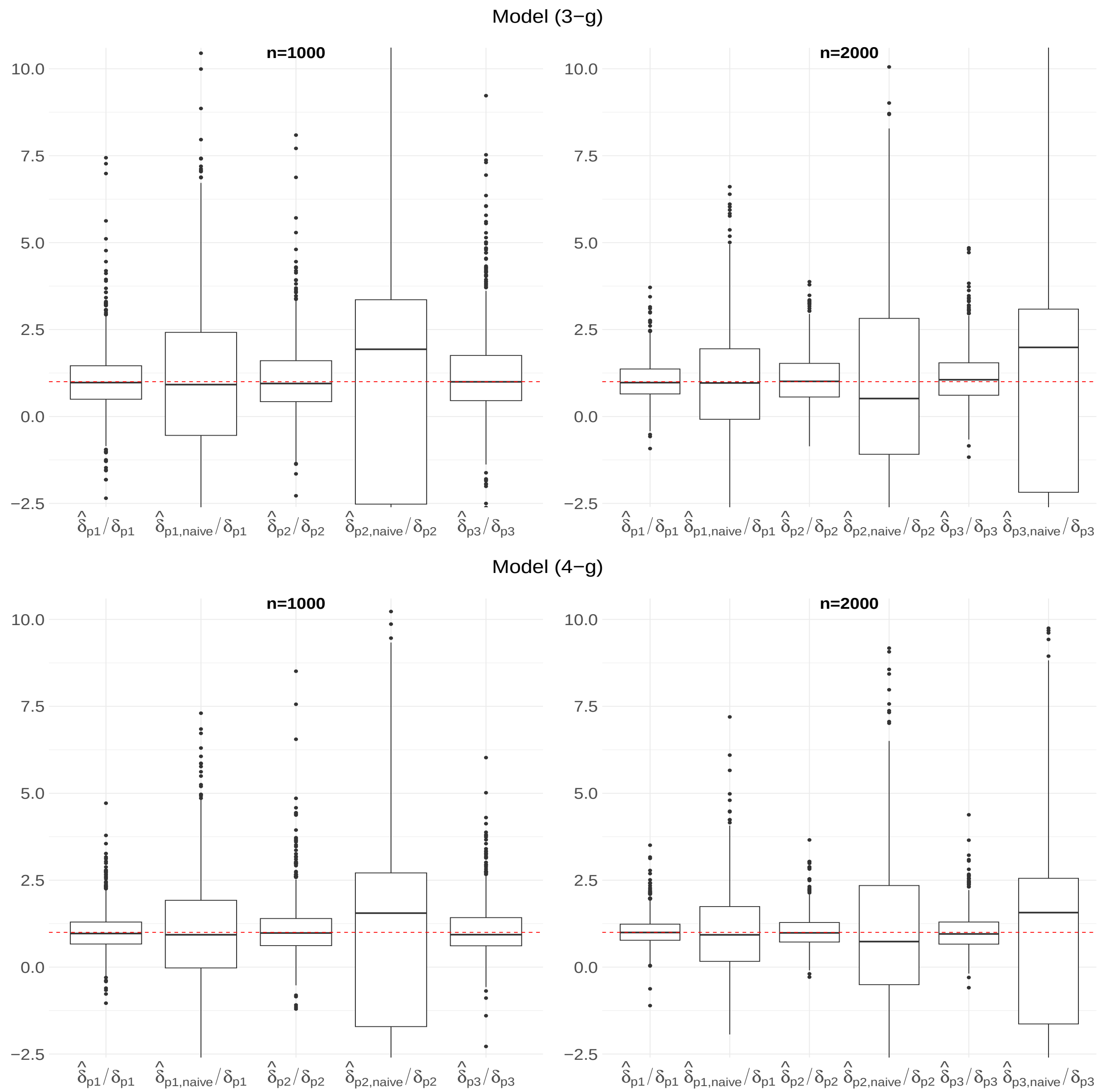

Figure S6: Boxplots for simulations under Models (3-g) and (4-g). Here, δ_p, $\widehat{\delta}_p$ and $\widehat{\delta}_{p,\text{naive}}$ denote $\text{CTE}_2(p)$, $\widehat{\text{CTE}}_2(p)$ and $\widehat{\text{CTE}}_{2,\text{naive}}(p)$, respectively. The values of p considered are $(p1, p2, p3) = (0.005, 0.001, 0.0005)$.

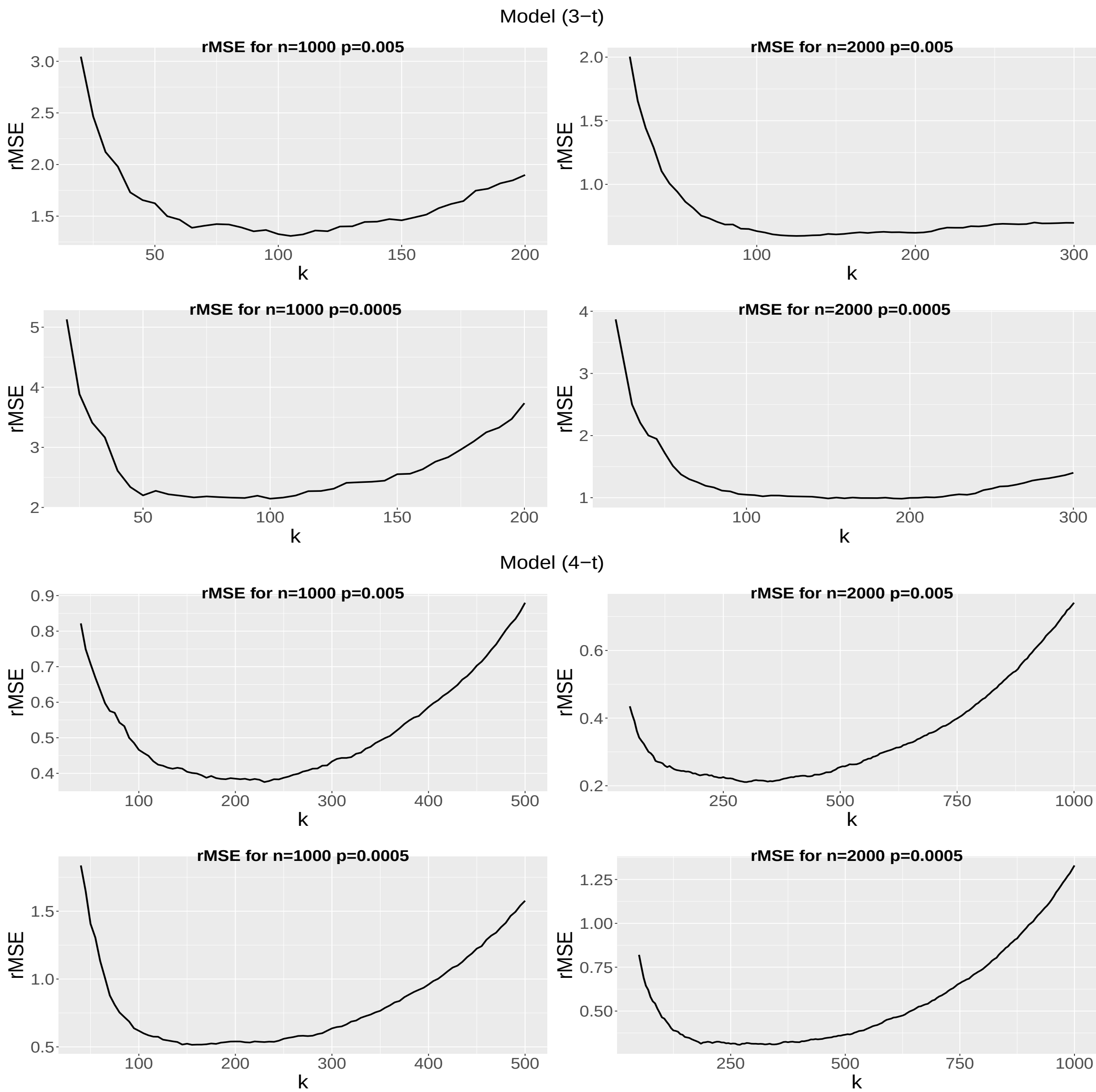

Figure S7: rMSE curves from simulation results of Models (3-t) and (4-t).

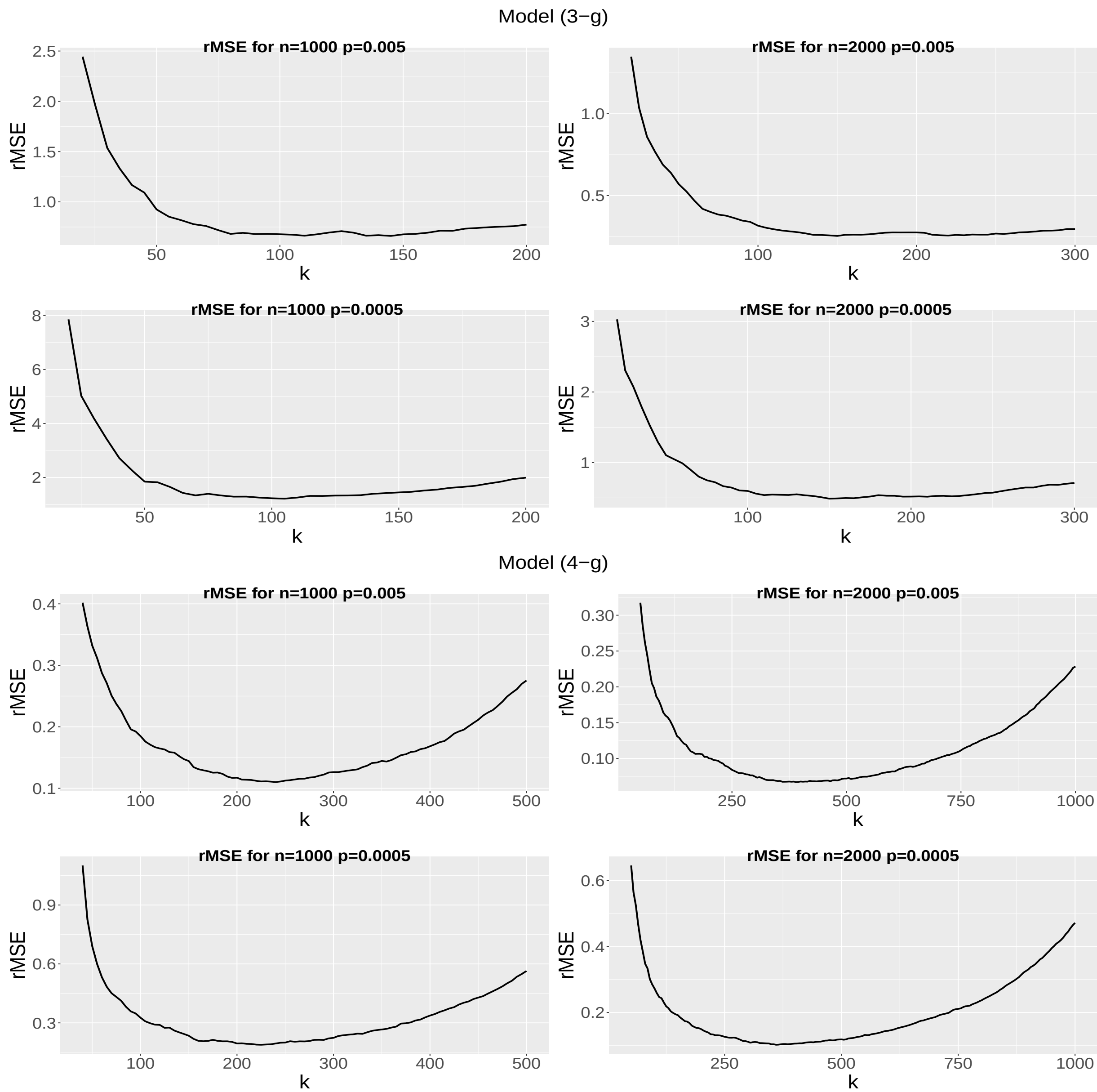

Figure S8: rMSE curves from simulation results of Models (3-g) and (4-g).

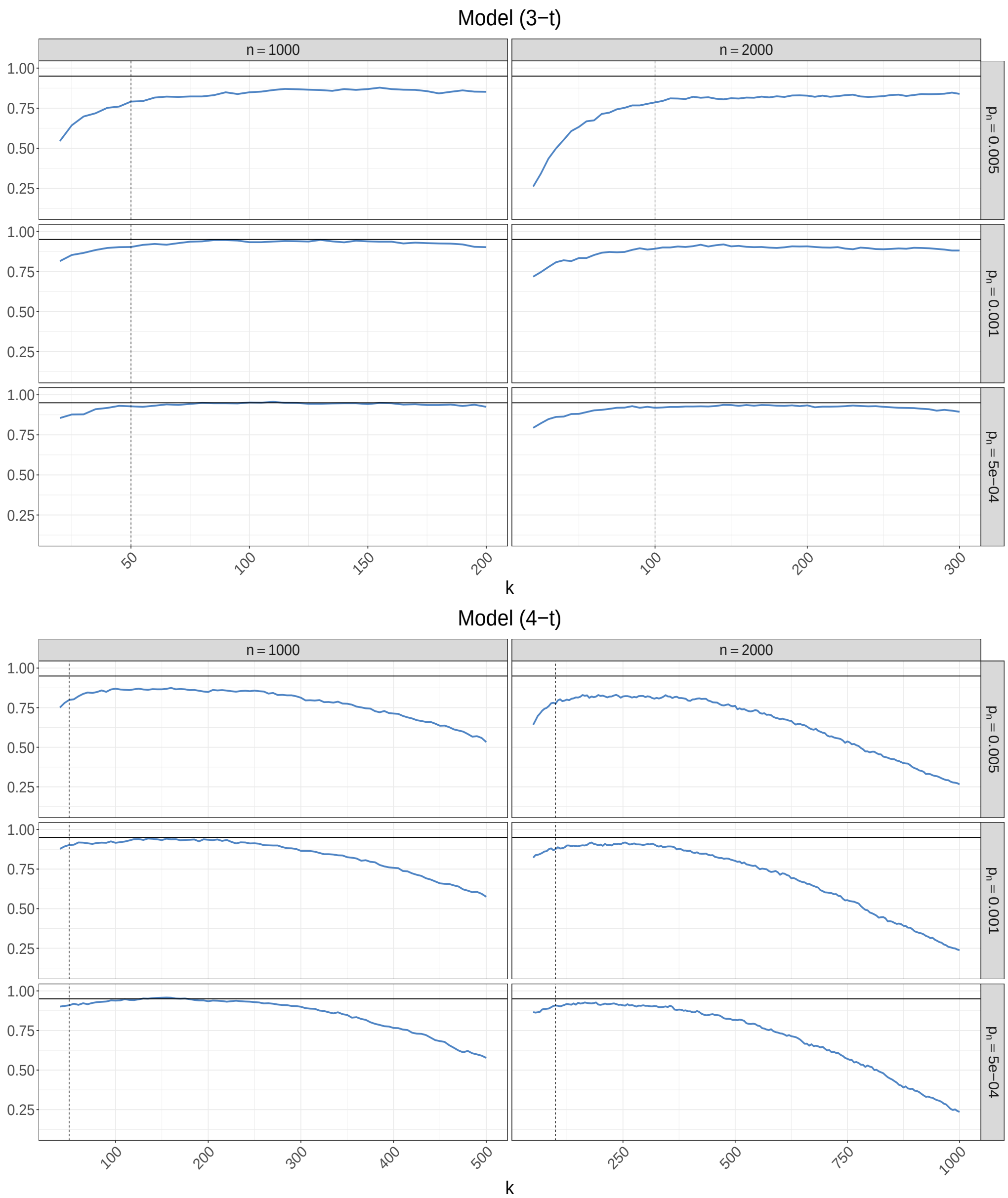

Figure S9: Plots of empirical coverage of CIs under Models (3-t) and (4-t). The vertical dashed line indicates $k = 0.05n$.

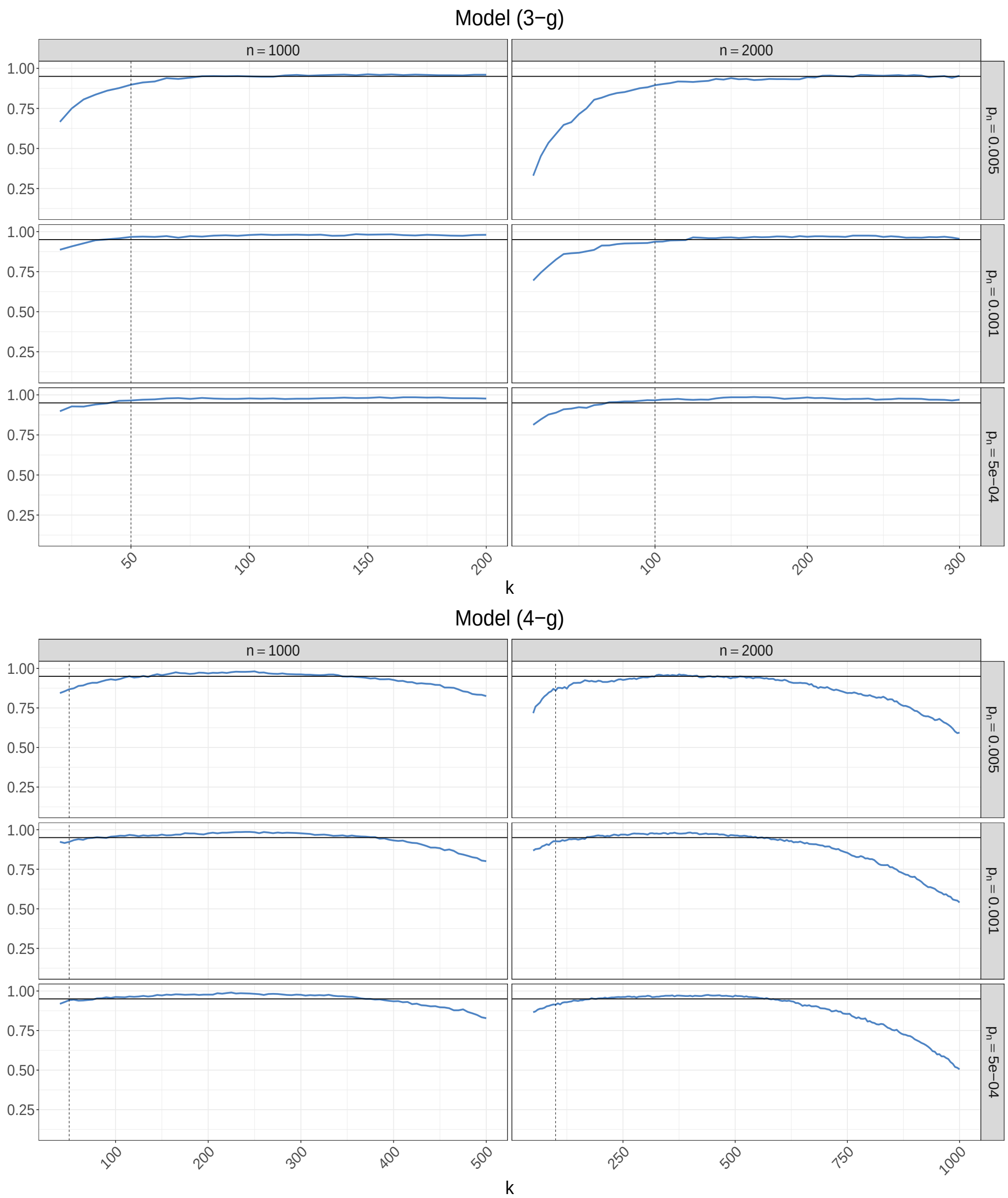

Figure S10: Plots of empirical coverage of CIs under Models (3-g) and (4-g). The vertical dashed line indicates $k = 0.05n$.

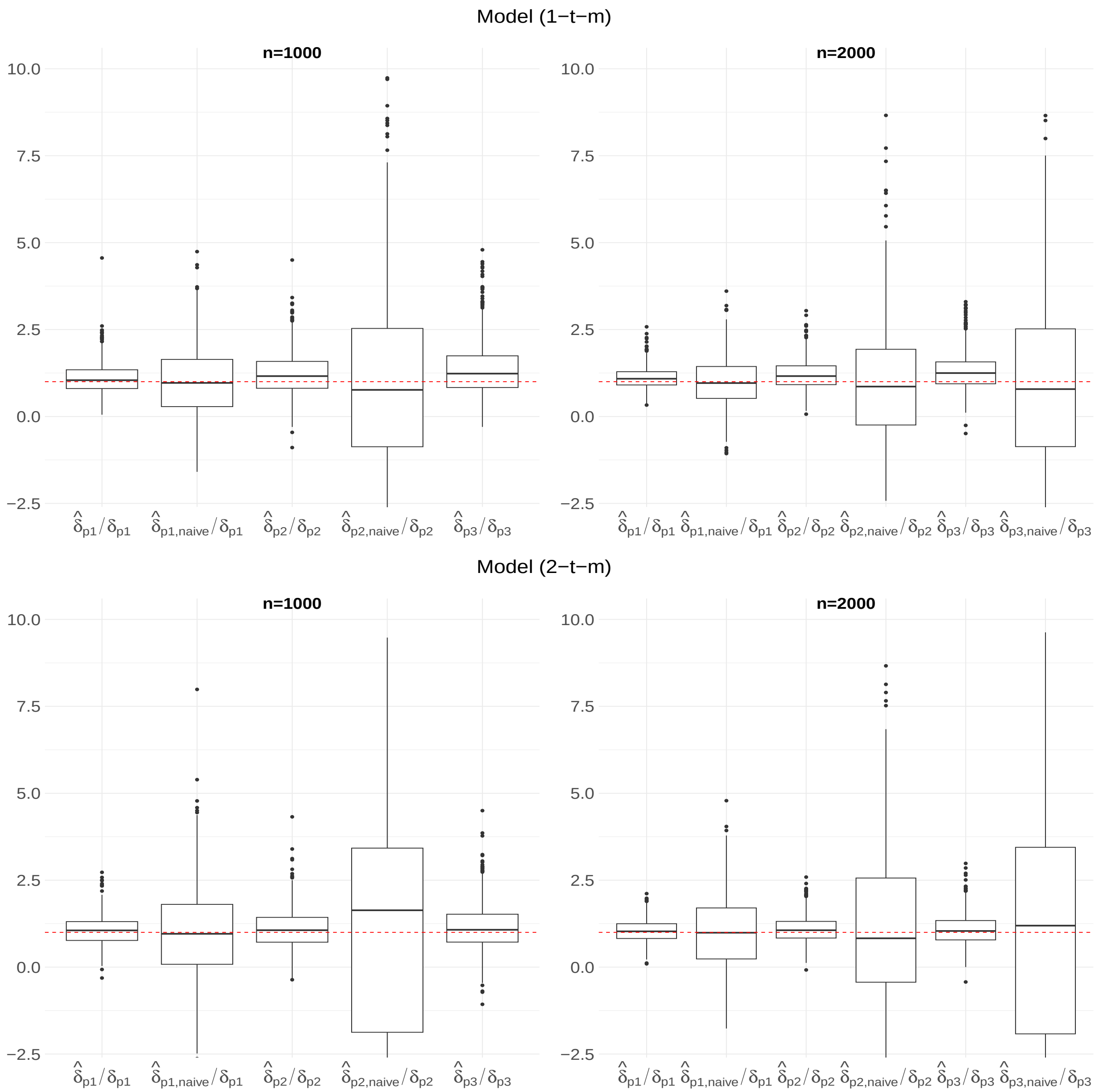

Figure S11: Boxplots for simulations under Models (1-t-m) and (2-t-m). Here, δ_p, $\widehat{\delta}_p$ and $\widehat{\delta}_{p,\text{naive}}$ denote $\text{CTE}_2(p)$, $\widehat{\text{CTE}}_2(p)$ and $\widehat{\text{CTE}}_{2,\text{naive}}(p)$, respectively. The values of p considered are $(p1, p2, p3) = (0.005, 0.001, 0.0005)$.

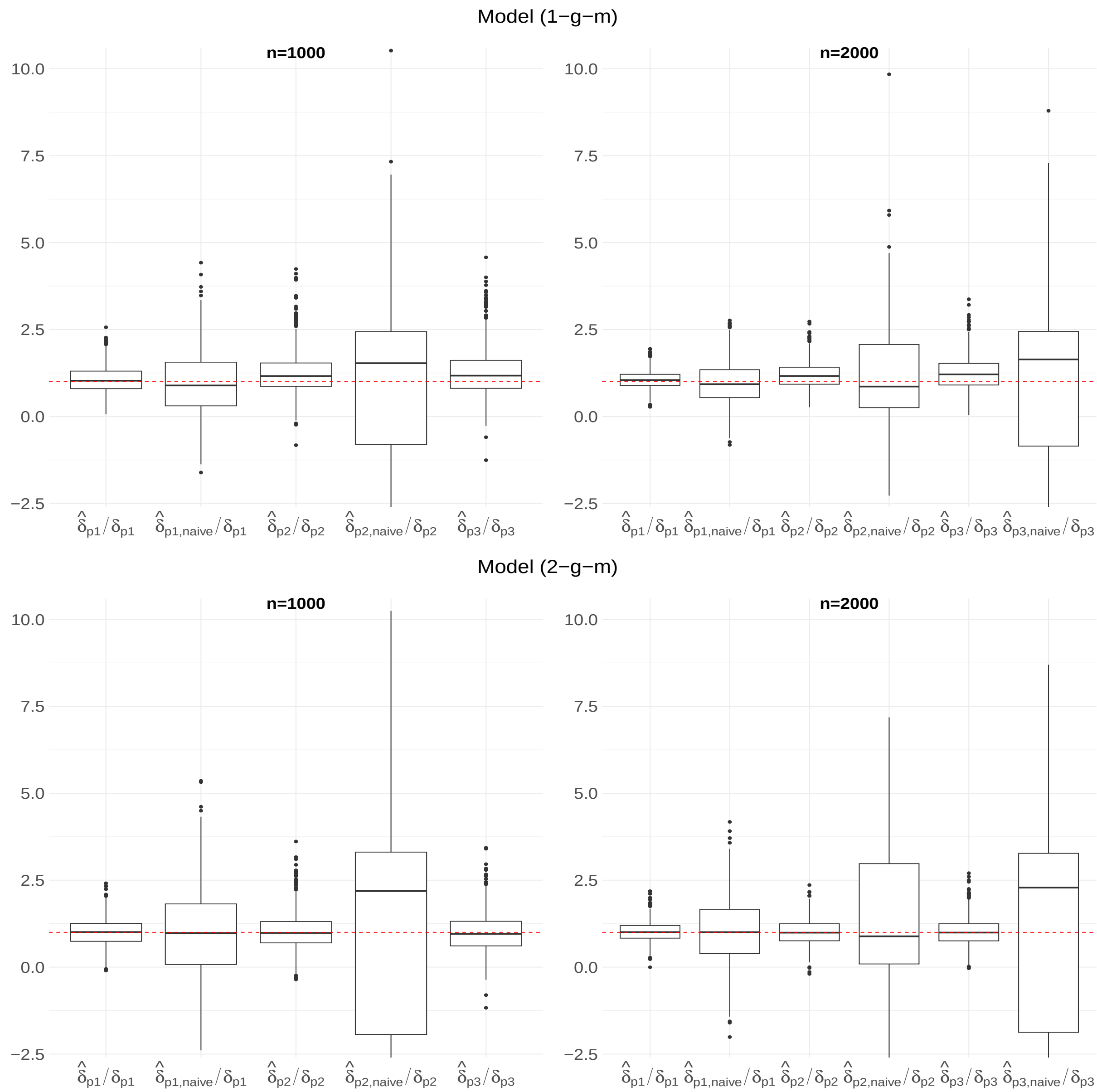

Figure S12: Boxplots for simulations under Models (1-g-m) and (2-g-m). Here, δ_p, $\widehat{\delta}_p$ and $\widehat{\delta}_{p,\text{naive}}$ denote $\text{CTE}_2(p)$, $\widehat{\text{CTE}}_2(p)$ and $\widehat{\text{CTE}}_{2,\text{naive}}(p)$, respectively. The values of p considered are $(p1, p2, p3) = (0.005, 0.001, 0.0005)$.

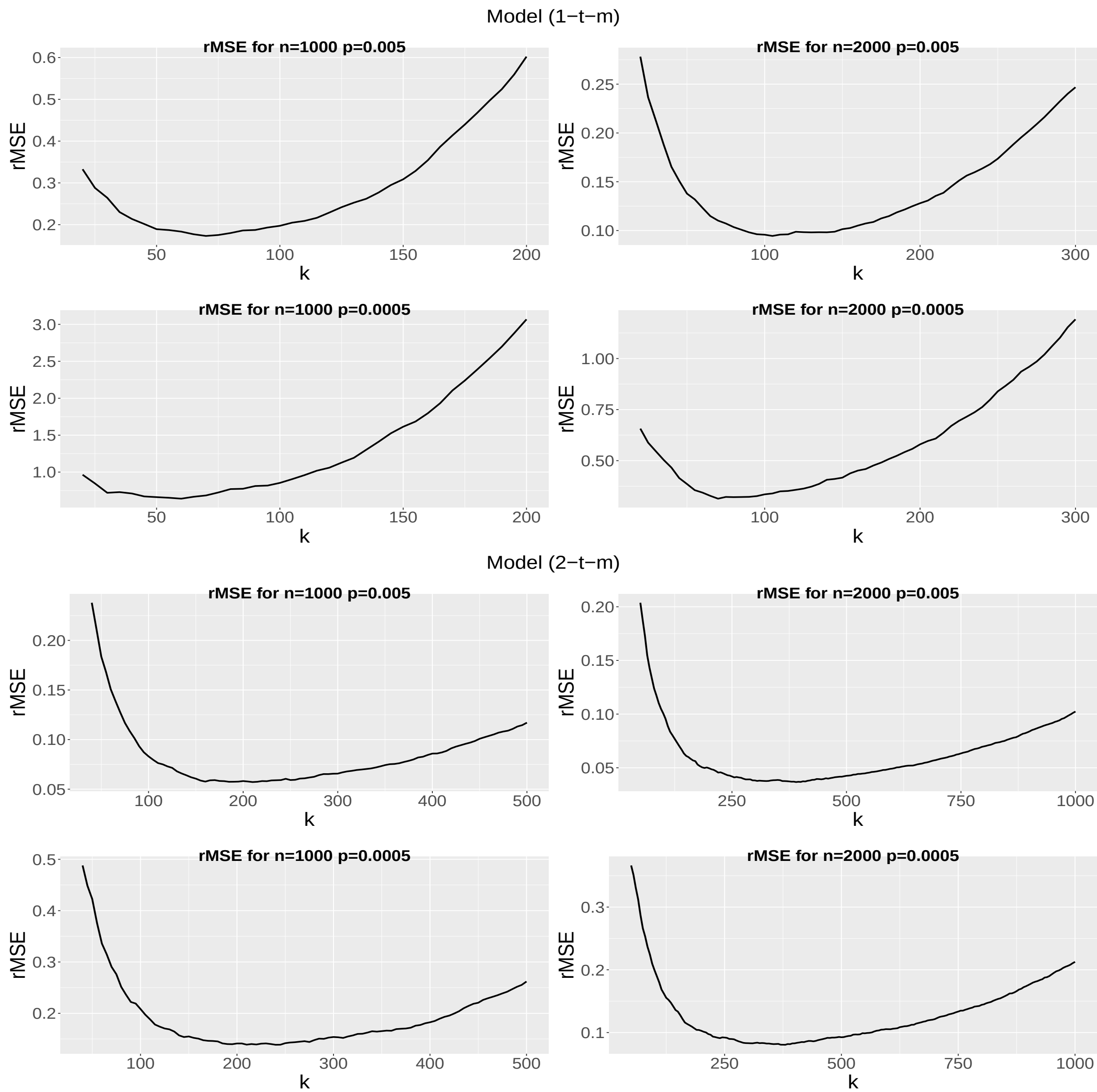

Figure S13: rMSE curves from simulation results of Models (1-t-m) and (2-t-m).

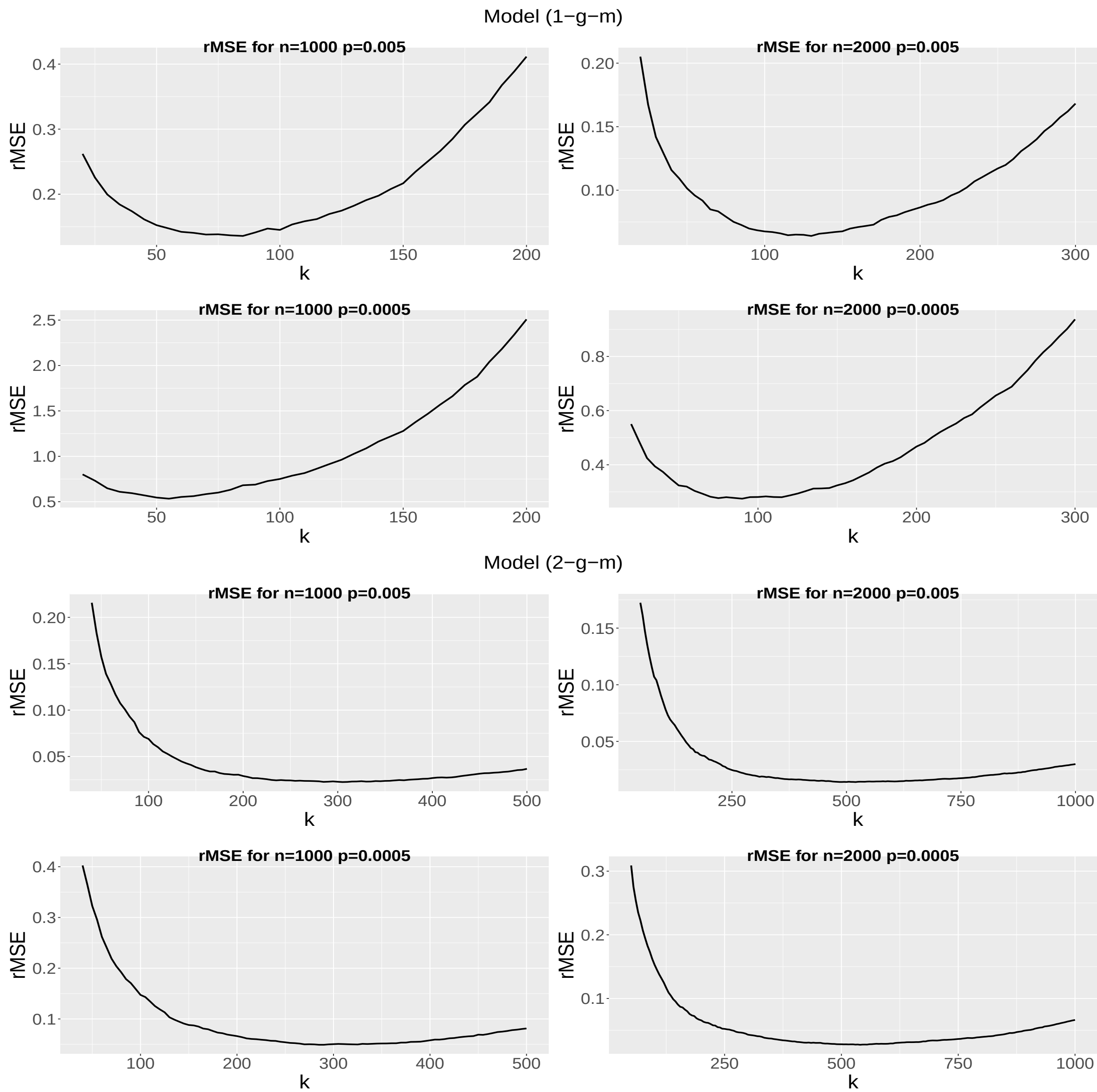

Figure S14: rMSE curves from simulation results of Models (1-g-m) and (2-g-m).

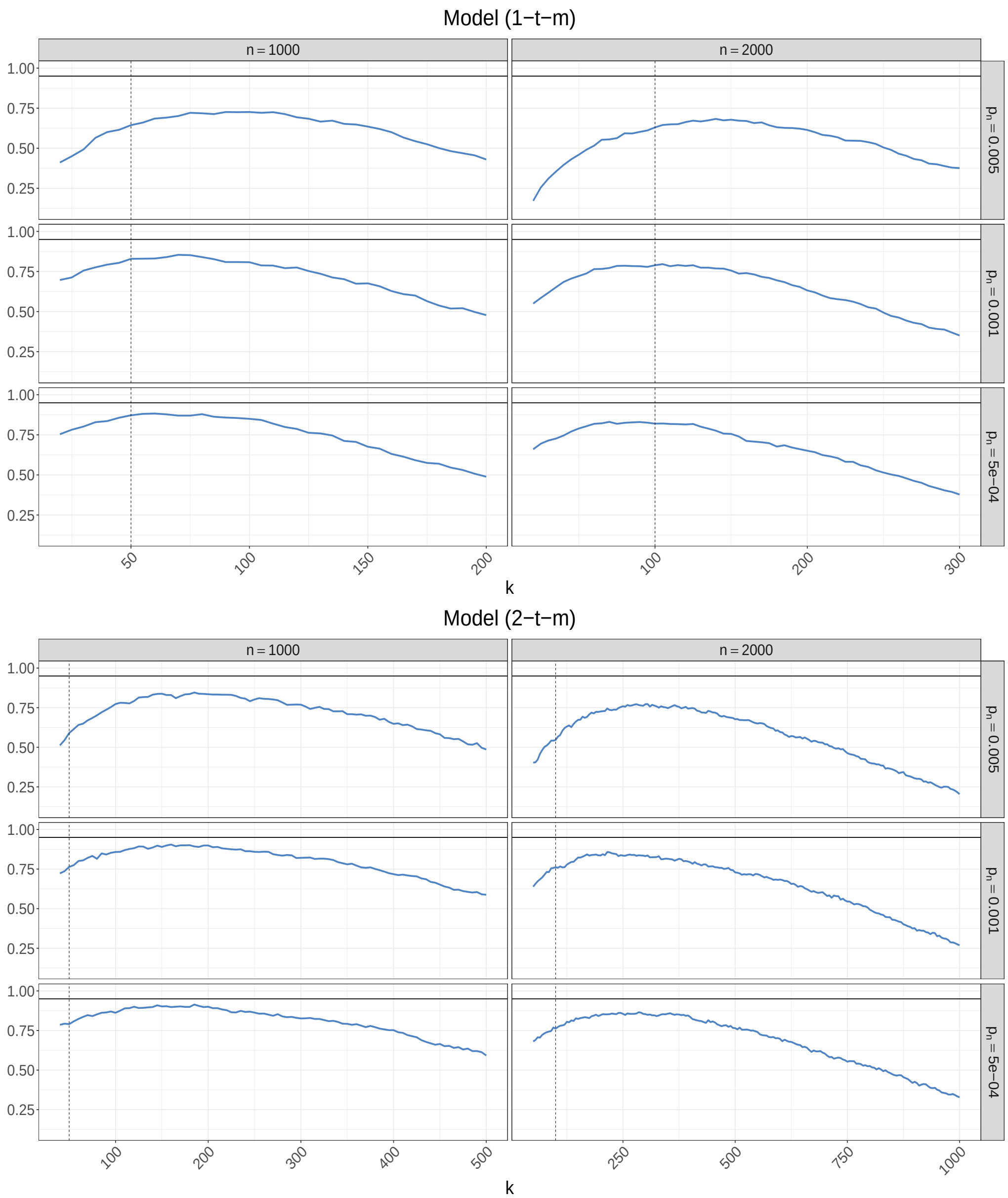

Figure S15: Plots of empirical coverage of CIs under Models (1-t-m) and (2-t-m). The vertical dashed line indicates $k = 0.05n$.

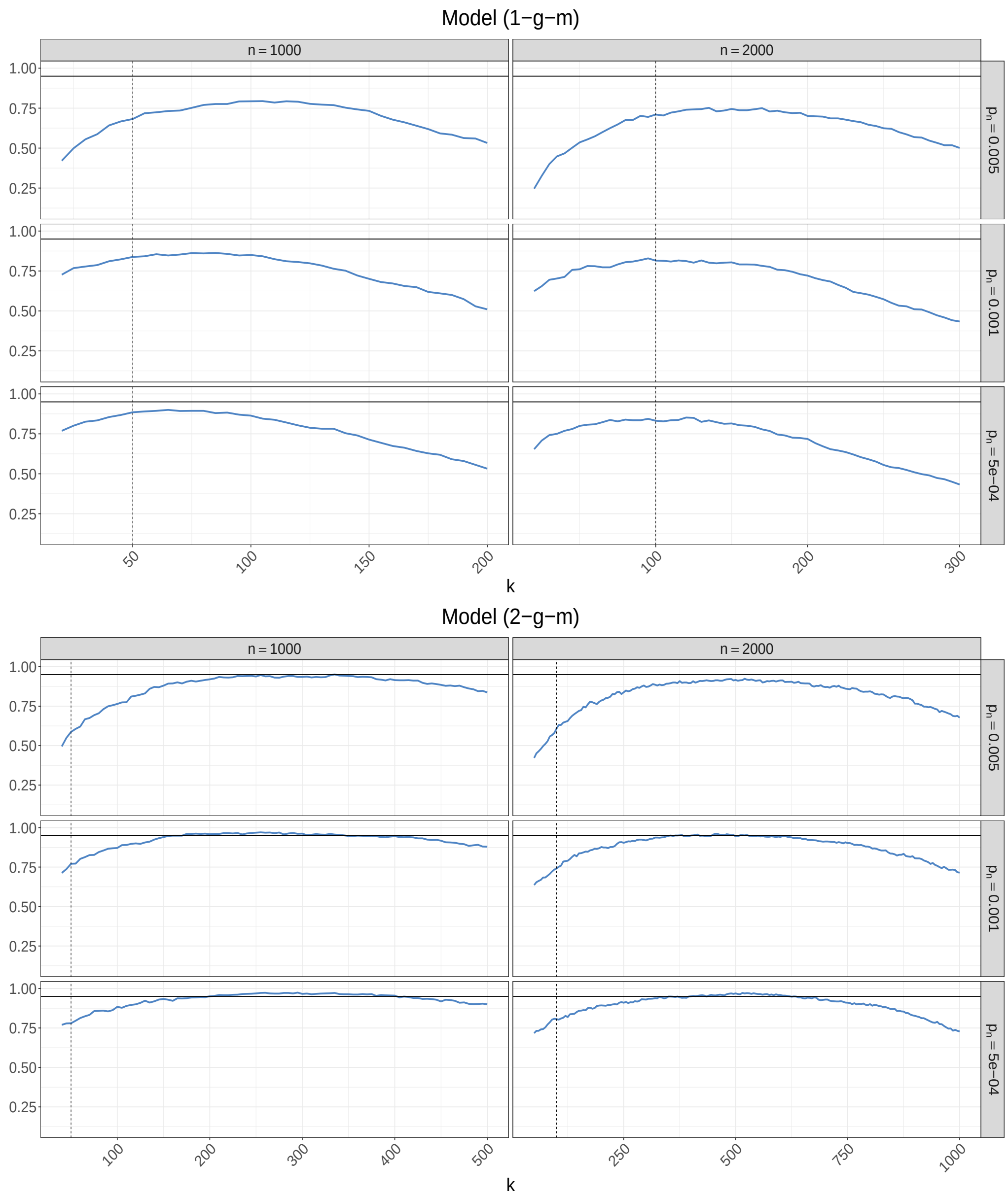

Figure S16: Plots of empirical coverage of CIs under Models (1-g-m) and (2-g-m). The vertical dashed line indicates $k = 0.05n$.

Bibliography

Adler, R. J. (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. Hayward: Institute of Mathematical Statistics.

Cai, J. J., Einmahl, J. H. J., Haan, L. and Zhou, C. (2015). Estimation of the marginal expected shortfall: the mean when a related variable is extreme. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **77**(2): 417–442.

Deuber, D., Li, J., Engelke, S. and Maathuis, M. H. (2024). Estimation and inference of extremal quantile treatment effects for heavy-tailed distributions. *Journal of the American Statistical Association*, **119**(547): 2206–2216.

Einmahl, J. H. J., de Haan, L. and Li, D. (2006). Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. *Annals of Statistics*, **34**(4): 1987–2014.

Hirano, K., Imbens, G. W. and Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, **71**(4): 1161–1189.

Newey, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica*, **62**(6): 1349–1382.

Orey, S. and Pruitt, W. E. (1973). Sample functions of the n-parameter wiener process. *The Annals of Probability*, **1**(1): 138–163.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer.

School of Management, Fudan University

E-mail: 24110690006@m.fudan.edu.cn

Faculty of Business for Science and Technology, School of Management, University of Science and Technology of China

E-mail: ljchen22@ustc.edu.cn

ORCID: https://orcid.org/0009-0001-5302-3527

School of Management, Fudan University

E-mail: deyuanli@fudan.edu.cn

ORCID: https://orcid.org/0000-0003-2155-2552