Supplemental material for "*Calibration estimation of semiparametric copula models with data missing at random*"

Shigeyuki Hamori*– Kobe University
Kaiji Motegi†– Kobe University
Zheng Zhang‡– Renmin University of China

December 7, 2018

*Graduate School of Economics, Kobe University. E-mail: hamori@econ.kobe-u.ac.jp

†*Corresponding author*. Graduate School of Economics, Kobe University. Address: 2-1 Rokkodai-cho, Nada, Kobe, Hyogo 657-8501 Japan. E-mail: motegi@econ.kobe-u.ac.jp

‡Institute of Statistics and Big Data, Renmin University of China. E-mail: zhengzhang@ruc.edu.cn

Contents

List of Tables

1 Introduction

In this supplemental material, we present technical and numerical details omitted in the main paper Hamori, Motegi, and Zhang (2018). We begin with a brief review of the key notation:

$$U_{ji} := F_j^0(Y_{ji}) \ , \ \boldsymbol{U}_i := (U_{1i}, ..., U_{di})^\top \ , \ \ell(v_1, ..., v_d; \theta) := \ln c(v_1, ..., v_d; \theta) \ ,$$
$$\ell_\theta(v_1, ..., v_d; \theta) := \frac{\partial}{\partial\theta}\ell(v_1, ..., v_d; \theta) \ , \ \ell_{\theta\theta}(v_1, ..., v_d; \theta) := \frac{\partial^2}{\partial\theta\partial\theta^\top}\ell(v_1, ..., v_d; \theta) \ ,$$
$$\ell_j(v_1, ..., v_d; \theta) := \frac{\partial}{\partial v_j}\ell(v_1, ..., v_d; \theta) \ , \ \ell_{\theta j}(v_1, ..., v_d; \theta) := \frac{\partial^2}{\partial\theta\partial v_j}\ell(v_1, ..., v_d; \theta) \ ,$$

where Y_{ji} signifies a target variable of individual $i \in \{1, \ldots, N\}$ and component $j \in \{1, \ldots, d\}$. T_{ji} is a binary variable that equals 1 if Y_{ji} is observed and 0 if Y_{ji} is missing.

A vector of the target variables $\boldsymbol{Y}_i = (Y_{1i}, \ldots, Y_{di})^\top$ can formally be divided into two parts $\boldsymbol{Y}_i = (\boldsymbol{Y}_{i,obs}^\top, \boldsymbol{Y}_{i,mis}^\top)^\top$, where the component $\boldsymbol{Y}_{i,obs} \in \mathbb{R}^{d_{obs}}$ is assumed to be always observed while the component $\boldsymbol{Y}_{i,mis} \in \mathbb{R}^{d_{mis}}$ may be missing, and $d = d_{obs} + d_{mis}$. As in the main paper, we simplify notation by assuming that $\boldsymbol{Y}_{i,obs} = \emptyset$ and $\boldsymbol{Y}_{i,mis} = \boldsymbol{Y}_i$, which means that all components of $\boldsymbol{Y}_i$ are possibly missing. Then $d_{obs} = 0$, $d_{mis} = d$, and $0 < \Pr(T_{ji} = 1) < 1$ for all $j \in \{1, \ldots, d\}$. This simplification assumption does not cause any loss of generality.

The rest of this supplemental material is organized as follows. In Section 2, we list and relabel all assumptions stated in the main paper for clarity. In Section 3, we derive a dual solution to the primal problem (7) of the main paper.

In Section 4, we prove Theorem 2 of the main paper, which concerns the weak convergence of estimated marginal distributions $\{\hat{F}_1, \ldots, \hat{F}_d\}$.

In Section 5, we show that the asymptotic covariance matrix V_0 reduces to a well-known form when there are no missing data. In Section 6, we provide a proof of Theorem 6 of the main paper, which establishes the consistency of our variance estimator $\hat{V}$.

In Section 7, we report complete results of the Monte Carlo simulations.

2 Assumptions

In this section, we list all assumptions stated in the main paper Hamori, Motegi, and Zhang (2018) in order to make this supplemental material self-contained. See the main paper for detailed discussions on each assumption.

Assumption 1 (Missing at Random) $\boldsymbol{T}_i \perp \boldsymbol{Y}_{i,mis} | \boldsymbol{X}_i$ *for any* $i \in \{1, \ldots, N\}$.

Assumption 2 *The support of the covariate* $\boldsymbol{X}$*, which is denoted by* $\mathcal{X}$*, is a Cartesian product of* r *compact intervals.*

Assumption 3 *The smallest eigenvalue of* $\mathrm{E}\left[u_{K_j}(\boldsymbol{X})u_{K_j}(\boldsymbol{X})^\top\right]$ *is bounded away from zero uniformly in* K_j*.*

Assumption 4 *For any* $j \in \{1, \ldots, d\}$*, the inverse propensity score* $\pi_j(x)^{-1}$ *is bounded above, i.e., there exists some constant* η_1 *such that* $1 \leq \pi_j(x)^{-1} \leq \eta_1 < \infty$*,* $\forall x \in \mathcal{X}$*.*

Assumption 5 *There exists* λ_{jK} *in* $\mathbb{R}^{K_j}$ *and* $\alpha > 0$ *such that* $\sup_{x\in\mathcal{X}} |(\rho')^{-1}(1/\pi_{jK}(x)) - \lambda_{jK}^\top u_{K_j}(x)| = O(K_j^{-\alpha})$ *as* $K_j \to \infty$*.*

Assumption 6 $\zeta(K_j)^2 K_j^4/N \to 0$ *and* $\sqrt{N}K_j^{-\alpha} \to 0$*.*

Assumption 7 $\rho(\cdot)$ *is a strictly concave function defined on* $\mathbb{R}$ *and three times continuously differentiable, and the range of* ρ' *contains* $[1, \eta_1]$*.*

Assumption 8 *For any* $j \in \{1, \ldots, d\}$*, the conditional distribution function* $F_j(y|x) := \Pr(Y_{ji} \leq y|\boldsymbol{X}_i = x)$ *is continuously differentiable in* x *and is Lipschitz continuous in* y*.*

Assumption 9 *The smallest eigenvalue of* $\mathrm{E}\left[u_{K_\eta}(\boldsymbol{X})u_{K_\eta}(\boldsymbol{X})^\top\right]$ *is bounded away from zero uniformly in* K_η*.*

Assumption 10 *There exists some constant* $\eta_1 > 0$ *such that* $1 \leq \eta(x)^{-1} \leq \eta_1 < \infty$*,* $\forall x \in \mathcal{X}$*.*

Assumption 11 *There exist* $\beta_K \in \mathbb{R}^{K_\eta}$ *and* $\alpha > 0$ *such that* $\sup_{x\in\mathcal{X}} |(\rho')^{-1}(1/\eta(x)) - \beta_K^\top u_{K_\eta}(x)| = O({K_\eta}^{-\alpha})$ *as* $K_\eta \to \infty$*.*

Assumption 12 $\zeta(K_\eta)^2 {K_\eta}^4/N \to 0$ *and* $\sqrt{N}{K_\eta}^{-\alpha} \to 0$*.*

Assumption 13 *Let* $U_{ji} := F_j^0(Y_{ji})$ *and* $\ell(v_1, \ldots, v_d; \theta) := \ln c(v_1, \ldots, v_d; \theta)$*.*

1. $\ell(v_1, \ldots, v_d; \theta)$ *is a continuous function of* θ*.*

2. $\mathrm{E}[\sup_{\theta\in\Theta} |\ell(U_{1i}, \ldots, U_{di}; \theta)|] < \infty$*.*

Assumption 14 $\mathrm{E}[\ell_\theta(U_{1i}, \ldots, U_{di}; \theta)|\boldsymbol{X}_i = x]$ *is continuously differentiable in* x*.*

Assumption 15 *$B := -\mathrm{E}\left[\ell_{\theta\theta}(U_{1i},\ldots,U_{di};\theta_0)\right]$ and*

$$\Sigma := \mathrm{Var}(\varphi(\boldsymbol{T}_i,\boldsymbol{X}_i,\boldsymbol{U}_i;\theta_0) + \sum_{j=1}^{d} W_j(T_{ji},\boldsymbol{X}_i,U_{ji};\theta_0))$$

are finite and positive definite, where

$$\begin{aligned}
\varphi(\boldsymbol{T}_i,\boldsymbol{X}_i,\boldsymbol{U}_i;\theta_0) &:= \frac{I(T_{1i}=\cdots=T_{di}=1)}{\eta(\boldsymbol{X}_i)}\ell_\theta(U_{1i},\ldots,U_{di};\theta_0) - \frac{I(T_{1i}=\cdots=T_{di}=1)}{\eta(\boldsymbol{X}_i)}\mathrm{E}[\ell_\theta(U_{1i},\ldots,U_{di};\theta_0)|\boldsymbol{X}_i] \\
&\quad + \mathrm{E}[\ell_\theta(U_{1i},\ldots,U_{di};\theta_0)|\boldsymbol{X}_i] - \mathrm{E}[\ell_\theta(U_{1i},\ldots,U_{di};\theta_0)], \\
W_j(T_{ji},\boldsymbol{X}_i,U_{ji};\theta_0) &:= \mathrm{E}\left[\ell_{\theta j}(U_{1s},\ldots,U_{ds};\theta_0)\{\phi_j(T_{ji},\boldsymbol{X}_i,U_{ji};U_{js}) - U_{js}\}|U_{ji},\boldsymbol{X}_i,T_{ji}\right] \quad (s \neq i), \\
\phi_j(T_{ji},\boldsymbol{X}_i,U_{ji};v) &:= \frac{T_{ji}}{\pi_j(\boldsymbol{X}_i)}I(U_{ji}\leq v) - \frac{T_{ji}}{\pi_j(\boldsymbol{X}_i)}\mathrm{E}[I(U_{ji}\leq v)|\boldsymbol{X}_i] + \mathrm{E}[I(U_{ji}\leq v)|\boldsymbol{X}_i],\ v\in[0,1].
\end{aligned}$$

Assumption 16 *(i) For each $(u_1,\ldots,u_d)\in(0,1)^d$, $\ell_{\theta\theta}(u_1,\ldots,u_d;\theta)$ is continuous with respect to θ in a neighborhood of θ_0. (ii) $\mathrm{E}[\sup_{\theta\in\Theta:\|\theta-\theta_0\|=o(1)}\|\ell_{\theta\theta}(U_{1i},\ldots,U_{di};\theta)\|]<\infty$.*

Assumption 17 *For $j\in\{1,\ldots,d\}$, $\ell_{\theta j}(u_1,\ldots,u_d;\theta_0)$ is well defined and continuous in $(u_1,\ldots,u_d)\in(0,1)^d$. Furthermore,*

1. *$\|\ell_\theta(u_1,\ldots,u_d;\theta_0)\| \leq constant \times \prod_{j=1}^{d}\{v_j(1-v_j)\}^{-a_j}$ for some $a_j\geq 0$ such that $\mathrm{E}[\prod_{j=1}^{d}\{U_{ji}(1-U_{ji})\}^{-2a_j}]<\infty$;*
2. *$\|\ell_{\theta k}(u_1,\ldots,u_d;\theta_0)\| \leq constant \times \{v_k(1-v_k)\}^{-b_k}\prod_{j=1,j\neq k}^{d}\{v_j(1-v_j)\}^{-a_j}$ for some $b_k>a_k$ such that $\mathrm{E}[\{U_{ki}(1-U_{ki})\}^{\xi_k-b_k}\prod_{j=1,j\neq k}^{d}\{U_{ji}(1-U_{ji})\}^{-a_j}]<\infty$ for some $\xi_k\in(0,1/2)$.*

3 Duality of primal problem (7) of the main paper

Recall the primal problem presented in the main paper Hamori, Motegi, and Zhang (2018, Eq. (7)):

$$\begin{cases} \min & \sum_{i=1}^{N} T_{ji}D(Np_{ji},1), \\ \text{subject to} & \sum_{i=1}^{N} T_{ji}p_{ji}u_{K_j}(\boldsymbol{X}_i) = \frac{1}{N}\sum_{i=1}^{N} u_{K_j}(\boldsymbol{X}_i), \end{cases} \tag{1}$$

where $K_j\to\infty$ as $N\to\infty$ yet with $K_j/N\to 0$.

In this section, we derive a dual solution to (1) by using the method of Tseng and Bertsekas (1987). (A dual solution to Eq. (9) of the main paper can be derived similarly, and hence the derivation is omitted.) For each $j\in\{1,...,d\}$, we define $E_{K_j\times N} := \big(u_{K_j}(\boldsymbol{X}_1),\ldots,u_{K_j}(\boldsymbol{X}_N)\big)$.

Also define $s_{ji} := 1 - T_{ji}Np_{ji}$ and $s_j := (s_{j1}, \ldots, s_{jN})^\top$ for each $j \in \{1, ..., d\}$ and $i \in \{1, \ldots, N\}$. Let $f(v) := D(1-v) = D(1-v, 1)$, then we can rewrite the problem (1) as

$$\min_{s_j} \sum_{i=1}^{N} T_{ji} f(s_{ji}) \text{ subject to } E_{K_j \times N} \cdot s_j = 0.$$

For every $i \in \{1, \ldots, N\}$, we define the conjugate convex function (Tseng and Bertsekas, 1987) of $T_{ji}f(\cdot)$ to be

$$\begin{aligned} g_{ji}(z_{ji}) =& \sup_{s_{ji}} \{z_{ji}s_{ji} - T_{ji}f(s_{ji})\} = \sup_{p_{ji}} \{-T_{ji}Np_{ji}z_{ji} + z_{ji} - T_{ji}f(1 - T_{ji}Np_{ji})\} \\ =& \sup_{p_{ji}} \{-T_{ji}Np_{ji}z_{ji} + z_{ji} - T_{ji}f(1 - Np_{ji})\} \\ =& - T_{ji}Np^*_{ji}z_{ji} + z_{ji} - T_{ji}f(1 - Np^*_{ji}) \ , \end{aligned}$$

where the third equality follows by $T_{ji}f(1 - T_{ji}Np_{ji}) = T_{ji}f(1 - Np_{ji})$, and p^*_{ji} satisfies the first order condition:

$$-T_{ji}z_{ji} = -T_{ji}f'(1 - Np^*_{ji}) \ \Rightarrow \ p^*_{ji} = \frac{1}{N}\left\{1 - (f')^{-1}(z_{ji})\right\} \ ;$$

then we can have

$$\begin{aligned} g_{ji}(z_{ji}) =& - T_{ji}z_{ji}\left\{1 - (f')^{-1}(z_{ji})\right\} + z_{ji} - T_{ji}f\left((f')^{-1}(z_{ji})\right) \\ =& - T_{ji}\left\{f\left((f')^{-1}(z_{ji})\right) + z_{ji} - z_{ji}(f')^{-1}(z_{ji})\right\} + z_{ji} \\ =& - T_{ji}\rho(z_{ji}) + z_{ji} \ , \end{aligned}$$

where

$$\rho(z) := f\left((f')^{-1}(z)\right) + z - z(f')^{-1}(z).$$

By Tseng and Bertsekas (1987), the dual problem of (1) is therefore

$$\begin{aligned} \min_{\lambda_j} \sum_{i=1}^{N} g_{ji}(\lambda_j^\top E_i) =& \min_{\lambda_j} \sum_{i=1}^{N} g_{ji}(\lambda_j^\top u_{K_j}(\boldsymbol{X}_i)) = \min_{\lambda_j} \sum_{i=1}^{N} \left\{-T_{ji}\rho\left(\lambda_j^\top u_{K_j}(\boldsymbol{X}_i)\right) + \lambda_j^\top u_{K_j}(\boldsymbol{X}_i)\right\} \\ =& - \max_{\lambda_j} \sum_{i=1}^{N} \left\{T_{ji}\rho\left(\lambda_j^\top u_{K_j}(\boldsymbol{X}_i)\right) - \lambda_j^\top u_{K_j}(\boldsymbol{X}_i)\right\} = - \max_{\lambda_j} \hat{G}_{jK}(\lambda_j) \ , \end{aligned}$$

where E_i is the i-th column of $E_{K_j \times N}$, i,e., $E_i = u_{K_j}(\boldsymbol{X}_i)$.

Since $D(\cdot)$ is strictly convex, i.e., $D''(v) > 0$, and $f''(v) = D''(1-v)$, $f(\cdot)$ is also strictly convex and $f'(\cdot)$ is strictly increasing. Note that

$$\rho(v) = f((f')^{-1}(v)) + v - v(f')^{-1}(v) \;\Leftrightarrow\; \rho\left(f'(v)\right) = f(v) + f'(v) - vf'(v).$$

Take derivatives with respect to v for both sides of the latter equation to get:

$$\rho'\left(f'(v)\right) f''(v) = f'(v) + f''(v) - f'(v) - vf''(v) = (1-v)f''(v).$$

Since $f''(v) > 0$, it follows that $\rho'(f'(v)) = 1 - v$. Take derivatives with respect to v for both sides to get $\rho''\left(f'(v)\right) f''(v) = -1$. It implies that

$$\rho''(v) = -\frac{1}{f''\left((f')^{-1}(v)\right)} < 0.$$

Therefore, the convexity of $D(\cdot)$ is equivalent to the concavity of $\rho(\cdot)$.

4 Proof of Theorem 2

To prove Theorem 2, we introduce some preliminary notation and results. Let

$$\begin{aligned}
\Sigma_{jK} &:= \mathrm{E}\left[\pi_j(\boldsymbol{X})\rho''((\lambda_{jK}^*)^\top u_{K_j}(\boldsymbol{X}))u_{K_j}(\boldsymbol{X})u_{K_j}^\top(\boldsymbol{X})\right] \ , \\
\Psi_{jK}(y) &:= -\mathrm{E}\left[F_j(y|\boldsymbol{X}_i)\pi_j(\boldsymbol{X}_i)\rho''\left((\lambda_{jK}^*)^\top u_{K_j}(\boldsymbol{X}_i)\right)u_{K_j}(\boldsymbol{X}_i)\right] \ , \\
Q_{jK}(x,y) &:= \Psi_{jK}(y)^\top \Sigma_{jK}^{-1} \cdot u_{K_j}(x).
\end{aligned}$$

Note that $-Q_{jK}(x,y)$ is a weighted L^2-projection (with the weighted measure $-\rho''((\lambda_{jK}^*)^\top u_{K_j}(x))dF_X(x)$) of $F_j(y|x)$ on the space linearly spanned by $u_{K_j}(x)$. Because of Assumption 3, we can assume the sieve bases $\{u_{K_j}(\boldsymbol{X})\}$ are orthonormalized, namely,

$$\mathrm{E}\left[u_{K_j}(\boldsymbol{X})u_{K_j}(\boldsymbol{X})^\top\right] = I_{K_j}. \tag{2}$$

We also introduce the following notation:

$$\begin{aligned}
\tilde{\Sigma}_{jK} &:= \frac{1}{N}\sum_{i=1}^{N} T_{ji}\rho''((\tilde{\lambda}_{jK})^\top u_{K_j}(\boldsymbol{X}_i))u_{K_j}(\boldsymbol{X}_i)u_{K_j}^\top(\boldsymbol{X}_i) \ , \\
\tilde{\Psi}_{jK}(y) &:= -\int_{\mathcal{X}} F_j(y|x)\cdot \pi_j(x)\cdot \rho''(\tilde{\lambda}_{jK}^\top u_{K_j}(x))u_{K_j}(x)dF_X(x) \ , \\
\tilde{Q}_{jK}(x,y) &:= \tilde{\Psi}_{jK}(y)^\top \tilde{\Sigma}_{jK} u_{K_j}(x) \ ,
\end{aligned}$$

where $\tilde{\lambda}_{jK}$ lies on the line joining λ_{jK}^* and $\hat{\lambda}_{jK}$ such that the following Mean Value Theorem holds:

$$\begin{aligned}0 =&\hat{G}_{jK}(\hat{\lambda}_{jK}) = \frac{1}{N}\sum_{i=1}^{N} T_{ji}\rho'\left(\hat{\lambda}_{jK}^\top u_{K_j}(\boldsymbol{X}_i)\right) u_{K_j}(\boldsymbol{X}_i) - \frac{1}{N}\sum_{i=1}^{N} u_{K_j}(\boldsymbol{X}_i)\\ =&\frac{1}{N}\sum_{i=1}^{N} T_{ji}\rho'\left((\lambda_{jK}^*)^\top u_{K_j}(\boldsymbol{X}_i)\right) u_{K_j}(\boldsymbol{X}_i) - \frac{1}{N}\sum_{i=1}^{N} u_{K_j}(\boldsymbol{X}_i)\\ &+ \frac{1}{N}\sum_{i=1}^{N} T_{ji}\rho''\left(\tilde{\lambda}_{jK}^\top u_{K_j}(\boldsymbol{X}_i)\right) u_{K_j}(\boldsymbol{X}_i) u_{K_j}(\boldsymbol{X}_i)^\top \cdot \left\{\hat{\lambda}_{jK} - \lambda_{jK}^*\right\}.\end{aligned}$$

Hence

$$\hat{\lambda}_{jK} - \lambda_{jK}^* = -\tilde{\Sigma}_{jK}^{-1} \cdot u_{K_j}(\boldsymbol{X}_i) \cdot \left\{\frac{1}{N}\sum_{i=1}^{N} T_{ji}\rho'\left((\lambda_{jK}^*)^\top u_{K_j}(\boldsymbol{X}_i)\right) - 1\right\}. \tag{3}$$

For each $j \in \{1, \dots, d\}$, we have the following decomposition:

$$\begin{aligned}&\sqrt{N}\left\{\hat{F}_j(y) - F_j^0(y) - \frac{1}{N}\sum_{i=1}^{N}\psi_j(Y_{ji}, \boldsymbol{X}_i, T_{ji}; y)\right\}\\ =& \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{N T_{ji}\hat{p}_{jK}(\boldsymbol{X}_i) I(Y_{ji} \le y) - F_j^0(y) - \left(\frac{T_{ji}}{\pi_j(\boldsymbol{X}_i)} I(Y_{ji} \le y) - \frac{F_j(y|\boldsymbol{X}_i)}{\pi_j(\boldsymbol{X}_i)}(T_{ji} - \pi_j(\boldsymbol{X}_i)) - F_j^0(y)\right)\right\}\\ =& \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\left(N\hat{p}_{jK}(\boldsymbol{X}_i) - N p_{jK}^*(\boldsymbol{X}_i)\right) T_{ji} I(Y_{ji} \le y) - \int_{\mathcal{X}} F_j(y|x)\pi_j(x)(N\hat{p}_{jK}(x) - N p_{jK}^*(x)) dF_X(x)\right\} && (4)\\ &+ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\left(N p_{jK}^*(\boldsymbol{X}_i) - \frac{1}{\pi_j(\boldsymbol{X}_i)}\right) T_{ji} I(Y_{ji} \le y) - \mathrm{E}\left[F_j(y|\boldsymbol{X}_i)\pi_j(\boldsymbol{X}_i)\left(N p_{jK}^*(\boldsymbol{X}_i) - \frac{1}{\pi_j(\boldsymbol{X}_i)}\right)\right]\right\} && (5)\\ &+ \sqrt{N}\mathrm{E}\left[F_j(y|\boldsymbol{X})\pi_j(\boldsymbol{X})\left(N p_{jK}^*(\boldsymbol{X}) - \frac{1}{\pi_j(\boldsymbol{X})}\right)\right] && (6)\\ &+ \sqrt{N}\left\{\int_{\mathcal{X}} F_j(y|x)\pi_j(x)(N\hat{p}_{jK}(x) - N p_{jK}^*(x)) dF_X(x) - \frac{1}{N}\sum_{i=1}^{N}\left[T_{ji}\rho'\left((\lambda_{jK}^*)^\top u_{K_j}(\boldsymbol{X}_i)\right) - 1\right]\tilde{Q}_{jK}(\boldsymbol{X}_i, y)\right\} && (7)\\ &+ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[T_{ji}\rho'((\lambda_{jK}^*)^\top u_{K_j}(\boldsymbol{X}_i)) - 1](\tilde{Q}_{jK}(\boldsymbol{X}_i, y) - Q_{jK}(\boldsymbol{X}_i, y)) && (8)\\ &+ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{[T_{ji}\rho'((\lambda_{jK}^*)^\top u_{K_j}(\boldsymbol{X}_i)) - 1] Q_{jK}(\boldsymbol{X}_i, y) + \frac{F_j(y|\boldsymbol{X}_i)}{\pi_j(\boldsymbol{X}_i)}(T_{ji} - \pi_j(\boldsymbol{X}_i))\right\}. && (9)\end{aligned}$$

We shall show the terms (4)-(9) are of $o_p(1)$ uniformly in $y \in \mathbb{R}$.

***For the term* (4):** Note that

$$(4) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\left(N\hat{p}_{jK}(\boldsymbol{X}_i) - Np^*_{jK}(\boldsymbol{X}_i)\right)T_{ji}I(Y_{ji}\le y) - \left(N\hat{p}_{jK}(\boldsymbol{X}_i) - Np^*_{jK}(\boldsymbol{X}_i)\right)\pi_j(\boldsymbol{X}_i)F_j(y|\boldsymbol{X}_i)\right\} \quad (10)$$

$$+ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\left(N\hat{p}_{jK}(\boldsymbol{X}_i) - Np^*_{jK}(\boldsymbol{X}_i)\right)\pi_j(\boldsymbol{X}_i)F_j(y|\boldsymbol{X}_i) - \int_{\mathcal{X}} F_j(y|x)\pi_j(x)(N\hat{p}_{jK}(x) - Np^*_{jK}(x))dF_X(x)\right\}. \quad (11)$$

Consider the term (10). Given the σ-algebra $\sigma(\boldsymbol{X}_i, i\ge 1)$, $\hat{p}_{jK}(x)$ is a deterministic function of x, and the summands $\{(N\hat{p}_{jK}(\boldsymbol{X}_i) - Np^*_{jK}(\boldsymbol{X}_i))T_{ji}I(Y_{ji}\le y) - (N\hat{p}_{jK}(\boldsymbol{X}_i) - Np^*_{jK}(\boldsymbol{X}_i))\pi_j(\boldsymbol{X}_i)F_j(y|\boldsymbol{X}_i)\}_{i=1}^N$ are conditionally $i.i.d.$ with conditional mean zero in accordance with Assumption 1. For any fixed $y\in\mathbb{R}$, by computing the conditional second moment of (10) and using Lemma 2 of the main paper, we can obtain

$$\begin{aligned}
&\mathrm{E}\left[|(10)|^2|\sigma(\boldsymbol{X}_t, t\ge 1)\right]\\
=&\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}\left[|\left(N\hat{p}_{jK}(\boldsymbol{X}_i) - Np^*_{jK}(\boldsymbol{X}_i)\right)T_{ji}I(Y_{ji}\le y) - \left(N\hat{p}_{jK}(\boldsymbol{X}_i) - Np^*_{jK}(\boldsymbol{X}_i)\right)\pi_j(\boldsymbol{X}_i)F_j(y|\boldsymbol{X}_i)|^2|\sigma(\boldsymbol{X}_t, t\ge 1)\right]\\
\le&\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}\left[|\left(N\hat{p}_{jK}(\boldsymbol{X}_i) - Np^*_{jK}(\boldsymbol{X}_i)\right)T_{ji}I(Y_{ji}\le y)|^2|\sigma(\boldsymbol{X}_t, t\ge 1)\right]\\
\le&\frac{1}{N}\sum_{i=1}^{N}\left(N\hat{p}_{jK}(\boldsymbol{X}_i) - Np^*_{jK}(\boldsymbol{X}_i)\right)^2 = O_p(K_j/N) = o_p(1),
\end{aligned}$$

then by Chebyshev's inequality the term (10) is of $o_p(1)$ for fixed $y\in\mathbb{R}$. We next show that (10) is of $o_p(1)$ uniformly in $y\in\mathbb{R}$. Denote

$$f(\boldsymbol{X}_i, T_{ji}, Y_{ji}; y) := \left(N\hat{p}_{jK}(\boldsymbol{X}_i) - Np^*_{jK}(\boldsymbol{X}_i)\right)T_{ji}I(Y_{ji}\le y) - \left(N\hat{p}_{jK}(\boldsymbol{X}_i) - Np^*_{jK}(\boldsymbol{X}_i)\right)\pi_j(\boldsymbol{X}_i)F_j(y|\boldsymbol{X}_i).$$

Given the σ-algebra $\sigma(\boldsymbol{X}_t, t\ge 1)$, for large enough N and fixed $\delta>0$, the following L^2-continuity (conditional version) holds:

$$\begin{aligned}
&\left\{\mathrm{E}\left[\sup_{|y_1-y_2|\le\delta}|f(\boldsymbol{X}_i, T_{ji}, Y_{ji}; y_1) - f(\boldsymbol{X}_i, T_{ji}, Y_{ji}; y_2)|^2\middle|\sigma(\boldsymbol{X}_t, t\ge 1)\right]\right\}^{1/2}\\
=&\left\{\left(N\hat{p}_{jK}(\boldsymbol{X}_i) - Np^*_{jK}(\boldsymbol{X}_i)\right)^2\cdot\mathrm{E}\left[\{T_{ji}(I(Y_{ji}\le y_1) - I(Y_{ji}\le y_2)) - \pi_j(\boldsymbol{X}_i)(F_j(y_1|\boldsymbol{X}_i) - F_j(y_2|\boldsymbol{X}_i))\}^2\middle|\sigma(\boldsymbol{X}_t, t\ge 1)\right]\right\}^{1/2}\\
=&\left\{\left(N\hat{p}_{jK}(\boldsymbol{X}_i) - Np^*_{jK}(\boldsymbol{X}_i)\right)^2\times\left(\pi_j(\boldsymbol{X}_i)\cdot|F_j(y_1|\boldsymbol{X}_i) - F_j(y_2|\boldsymbol{X}_i)| - \pi_j(\boldsymbol{X}_i)^2\cdot|F_j(y_1|\boldsymbol{X}_i) - F_j(y_2|\boldsymbol{X}_i)|^2\right)\right\}^{1/2}\\
\le&\left\{\left(N\hat{p}_{jK}(\boldsymbol{X}_i) - Np^*_{jK}(\boldsymbol{X}_i)\right)^2\times\pi_j(\boldsymbol{X}_i)\cdot|F_j(y_1|\boldsymbol{X}_i) - F_j(y_2|\boldsymbol{X}_i)|\right\}^{1/2}\\
\le&\sup_{x\in\mathcal{X}}|N\hat{p}_{jK}(x) - Np^*_{jK}(x)|\cdot\sqrt{L}\cdot|y_1-y_2|^{1/2}\le|y_1-y_2|^{1/2},
\end{aligned}$$

where L denotes the Lipschitz's coefficient of $F_j(y|x)$ by Assumption 8, and the last inequality follows from Lemma 2. By the law of iterated expectation, the following L^2-continuity (unconditional version) holds for large enough N:

$$\left\{\mathrm{E}\left[\sup_{|y_1-y_2|\leq\delta}|f(\boldsymbol{X}_i,T_{ji},Y_{ji};y_1)-f(\boldsymbol{X}_i,T_{ji},Y_{ji};y_2)|^2\right]\right\}^{1/2}\leq|y_1-y_2|^{1/2}\text{ for fixed }\delta>0,$$

therefore, Condition (5.3) of Andrews (1994) is satisfied. By Theorems 4 and 5 of Andrews (1994), the function class $\left\{\frac{1}{\sqrt{N}}\sum_{i=1}^N f(\boldsymbol{X}_i,T_{ji},Y_{ji};y);y\in\mathbb{R}\right\}$ is *stochastically equicontinuous*. Therefore,

$$(10)=o_p(1)\text{ holds uniformly in }y\in\mathbb{R}.$$

Next we consider (11). Using Mean Value Theorem twice, we can decompose (11) as follows:

$$\begin{aligned}
(11)=&\frac{1}{\sqrt{N}}\sum_{i=1}^N\left[\pi_j(\boldsymbol{X}_i)F_j(y|\boldsymbol{X}_i)\rho''(\tilde{\lambda}_{jK}^\top u_{K_j}(\boldsymbol{X}_i))u_{K_j}(\boldsymbol{X}_i)^\top-\int_{\mathcal{X}}F_j(y|x)\pi_j(x)\rho''(\tilde{\lambda}_{jK}^\top u_{K_j}(x))u_{K_j}(x)^\top dF_X(x)\right](\hat{\lambda}_{jK}-\lambda_{jK}^*)\\
=&\left\{\frac{1}{\sqrt{N}}\sum_{i=1}^N\left[\pi_j(\boldsymbol{X}_i)F_j(y|\boldsymbol{X}_i)\rho''((\lambda_{ji}^*)^\top u_{K_j}(\boldsymbol{X}_i))u_{K_j}(\boldsymbol{X}_i)^\top-\int_{\mathcal{X}}F_j(y|x)\pi_j(x)\rho''((\lambda_{jK}^*)^\top u_{K_j}(x))u_{K_j}(x)^\top dF_X(x)\right]\right.\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^N\pi_j(\boldsymbol{X}_i)F_j(y|\boldsymbol{X}_i)\rho'''(\xi_{j3}(\boldsymbol{X}_i))(\tilde{\lambda}_{jK}-\lambda_{jK}^*)^\top u_{K_j}(\boldsymbol{X}_i)u_{K_j}(\boldsymbol{X}_i)^\top\\
&\left.-\sqrt{N}\int_{\mathcal{X}}F_j(y|x)\pi_j(x)\rho'''(\xi_{j3}(x))(\tilde{\lambda}_{jK}-\lambda_{jK}^*)^\top u_{K_j}(x)u_{K_j}(x)^\top dF_X(x)\right\}(\hat{\lambda}_{jK}-\lambda_{jK}^*)\\
=&\left(W_{jK}^{(1)}(y)+W_{jK}^{(2)}(y)+W_{jK}^{(3)}(y)\right)^\top(\hat{\lambda}_K-\lambda_K^*)\,,
\end{aligned}$$

where

$$\begin{aligned}
W_{jK}^{(1)}(y)&:=\frac{1}{\sqrt{N}}\sum_{i=1}^N\left[\pi_j(\boldsymbol{X}_i)F_j(y|\boldsymbol{X}_i)\rho''((\lambda_K^*)^\top u_{K_j}(\boldsymbol{X}_i))u_{K_j}(\boldsymbol{X}_i)-\int_{\mathcal{X}}F_j(y|x)\pi_j(x)\rho''((\lambda_{jK}^*)^\top u_{K_j}(x))u_{K_j}(x)dF_X(x)\right],\\
W_{jK}^{(2)}(y)&:=\frac{1}{\sqrt{N}}\left[\sum_{i=1}^N T_{ji}I(Y_{ji}\leq y)\rho'''(\xi_{j3}(\boldsymbol{X}_i))u_{K_j}(\boldsymbol{X}_i)u_{K_j}(\boldsymbol{X}_i)^\top\right](\tilde{\lambda}_{jK}-\lambda_{jK}^*)\,,\\
W_{jK}^{(3)}(y)&:=-\sqrt{N}\int_{\mathcal{X}}\pi_j(x)F_j(y|x)\rho'''(\xi_{j3}(x))u_{K_j}(x)u_{K_j}(x)^\top dF_X(x)\cdot(\tilde{\lambda}_{jK}-\lambda_{jK}^*)\,,
\end{aligned}$$

and $\tilde{\lambda}_{jK}$ lies on the line joining λ_{jK}^* and $\hat{\lambda}_{jK}$, $\xi_{j3}(x)$ lies between $\tilde{\lambda}_{jK}^\top u_{K_j}(x)$ and $(\lambda_{jK}^*)^\top u_{K_j}(x)$.

Consider $W_{jK}^{(1)}(y)$. Noting $\mathrm{E}[W_{jK}^{(1)}(y)]=0$ and computing the second moment of $W_{jK}^{(1)}(y)$, we have:

$$
\begin{aligned}
&\mathrm{E}\left[|W_{jK}^{(1)}(y)|^2\right]\\
=&\mathrm{E}[\pi_j(\boldsymbol{X}_i)^2F_j(y|\boldsymbol{X}_i)^2\rho''((\lambda_{jK}^*)^\top u_{K_j}(\boldsymbol{X}_i))^2u_{K_j}(\boldsymbol{X}_i)^\top u_{K_j}(\boldsymbol{X}_i)]\\
&-\mathrm{E}[F_j(y|\boldsymbol{X}_i)\pi_j(\boldsymbol{X}_i)\rho''((\lambda_{jK}^*)^\top u_{K_j}(\boldsymbol{X}_i))u_{K_j}^\top(\boldsymbol{X}_i)]\cdot\mathrm{E}[F_j(y|\boldsymbol{X}_i)\pi_j(\boldsymbol{X}_i)\rho''((\lambda_{jK}^*)^\top u_{K_j}(\boldsymbol{X}_i))u_{K_j}^\top(\boldsymbol{X}_i)]\\
\leq&\mathrm{E}[\pi_j(\boldsymbol{X}_i)^2F_j(y|\boldsymbol{X}_i)^2\rho''((\lambda_{jK}^*)^\top u_{K_j}(\boldsymbol{X}_i))^2u_{K_j}(\boldsymbol{X}_i)^\top u_{K_j}(\boldsymbol{X}_i)]\\
\leq&\sup_{\gamma\in\Gamma_1}|\rho''(\gamma)|^2\cdot\mathrm{E}[\|u_{K_j}(\boldsymbol{X})\|^2]=O(K_j),\ \text{(use (2))}
\end{aligned}
$$

where the second inequality follows from the fact $(\lambda_{jK}^*)^\top u_{K_j}(x)\in\Gamma_1:=[\underline{\gamma}-1,\overline{\gamma}+1]$, $\forall x\in\mathcal{X}$ when K_j is large enough (see the proof of Lemma 1 of the main paper). Then by Chebyshev's inequality, we have that for each fixed $y\in\mathbb{R}$,

$$
\|W_{jK}^{(1)}(y)\|=O_p(\sqrt{K_j}). \tag{12}
$$

We next consider $\sup_{y\in\mathbb{R}}\|W_{jK}^{(3)}(y)\|$. By Lemma 2 of the main paper, we have that

$$
\begin{aligned}
&\sup_{y\in\mathbb{R}}\|W_{jK}^{(3)}(y)\|^2\\
=&N\cdot\sup_{y\in\mathbb{R}}\left\{(\tilde{\lambda}_{jK}-\lambda_{jK}^*)^\top\int_{\mathcal{X}}\pi_j(x)F_j(y|x)\rho'''(\xi_{j3}(x))u_{K_j}(x)u_{K_j}(x)dF_X(x)\right.\\
&\qquad\left.\times\int_{\mathcal{X}}\pi_j(x)F_j(y|x)\rho'''(\xi_{j3}(x))u_{K_j}(x)u_{K_j}(x)^\top dF_X(x)\times(\tilde{\lambda}_{jK}-\lambda_{jK}^*)\right\}\\
\leq&N\cdot\left(\sup_{x\in\mathcal{X}}|\rho'''(\xi_{j3}(x))|\cdot\sup_{x\in\mathcal{X}}\pi_j(x)\cdot\sup_{(x,y)\in\mathcal{X}\times\mathbb{R}}F_j(y|x)\right)^2\cdot(\tilde{\lambda}_{jK}-\lambda_{jK}^*)^\top\left\{\int_{\mathcal{X}}u_{K_j}(x)u_{K_j}(x)^\top dF_X(x)\right\}^2(\tilde{\lambda}_{jK}-\lambda_{jK}^*)\\
\leq&N\cdot O_p(1)\cdot\|\tilde{\lambda}_{jK}-\lambda_{jK}^*\|^2=O_p(K_j)\ ,
\end{aligned}
$$

where $\xi_{j3}(x)$ lies between $\tilde{\lambda}_{jK}^\top u_{K_j}(x)$ and $(\lambda_{jK}^*)^\top u_{K_j}(x)$. From the proof of Lemma 2, we know that $\sup_{x\in\mathcal{X}}|\rho'''(\xi_{3j}(x))|=O_p(1)$. Note $\tilde{\lambda}_{jK}$ lies on the line joining $\hat{\lambda}_{jK}$ and λ_{jK}^*, by Lemma 2 of the main paper, we have $\|\tilde{\lambda}_{jK}-\lambda_{jK}^*\|=O_p\left(\sqrt{K_j/N}\right)$. Therefore,

$$
\sup_{y\in\mathbb{R}}\|W_{jK}^{(3)}(y)\|=O(\sqrt{K_j}). \tag{13}
$$

Finally, we compute the probability order of $\sup_{y\in\mathbb{R}}\|W_{jK}^{(2)}(y)\|$.

$$
\sup_{y\in\mathbb{R}}\|W_{jK}^{(2)}(y)\|^2=N\cdot\sup_{y\in\mathbb{R}}\left\{(\tilde{\lambda}_{jK}-\lambda_{jK}^*)^\top\left[\frac{1}{N}\sum_{i=1}^N T_{ji}I(Y_{ji}\leq y)\rho'''(\xi_{j3}(\boldsymbol{X}_i))u_{K_j}(\boldsymbol{X}_i)u_{K_j}(\boldsymbol{X}_i)^\top\right]\right.
$$

$$
\begin{aligned}
&\times \left[\frac{1}{N}\sum_{i=1}^{N} T_{ji} I(Y_{ji} \le y)\rho'''(\xi_{j3}(\boldsymbol{X}_i)) u_{K_j}(\boldsymbol{X}_i) u_{K_j}(\boldsymbol{X}_i)^\top\right](\tilde{\lambda}_{jK} - \lambda_{jK}^*)\Bigg\} \\
\le & N \cdot \sup_{x\in\mathcal{X}} |\rho'''(\xi_{j3}(x))|^2 \cdot \left\{(\tilde{\lambda}_{jK} - \lambda_{jK}^*)^\top \left[\frac{1}{N}\sum_{i=1}^{N} u_{K_j}(\boldsymbol{X}_i) u_{K_j}(\boldsymbol{X}_i)^\top\right]^2 (\tilde{\lambda}_{jK} - \lambda_{jK}^*)\right\} \\
\le & N \cdot \sup_{x\in\mathcal{X}} |\rho'''(\xi_{j3}(x))|^2 \cdot \|\tilde{\lambda}_{jK} - \lambda_{jK}^*\|^2 \cdot \lambda_{\max}^2 \left(\frac{1}{N}\sum_{i=1}^{N} u_{K_j}(\boldsymbol{X}_i) u_{K_j}(\boldsymbol{X}_i)^\top\right) \\
\le & N \cdot O_p(1) \cdot O_p(K_j/N) \cdot O_p(1) = O_p(K_j),
\end{aligned}
$$

where $\lambda_{\max}\left(N^{-1}\sum_{i=1}^{N} u_{K_j}(\boldsymbol{X}_i) u_{K_j}(\boldsymbol{X}_i)^\top\right)$ denotes the largest eigenvalue of $N^{-1}\sum_{i=1}^{N} u_{K_j}(\boldsymbol{X}_i) u_{K_j}(\boldsymbol{X}_i)^\top$, and it converges to $\lambda_{\max}(\mathrm{E}[u_{K_j}(\boldsymbol{X}) u_{K_j}(\boldsymbol{X})^\top]) < \infty$. Therefore,

$$
\sup_{y\in\mathbb{R}} \|W_{jK}^{(2)}(y)\| = O_p(\sqrt{K_j}). \tag{14}
$$

Then in light of Lemma 2 of the main paper and Assumption 6, we have that for fixed $y \in \mathbb{R}$,

$$
\begin{aligned}
(11) \le & \left(\|W_{jK}^{(1)}(y)\| + \|W_{jK}^{(2)}(y)\| + \|W_{jK}^{(3)}(y)\|\right) \cdot \|\hat{\lambda}_{jK} - \lambda_{jK}^*\| \\
\le & \left(O_p(\sqrt{K_j}) + O_p(\sqrt{K_j}) + O_p(\sqrt{K_j})\right) \cdot O_p\left(\sqrt{\frac{K_j}{N}}\right) = O_p\left(\sqrt{\frac{K_j^2}{N}}\right) = o_p(1).
\end{aligned}
$$

Using a similar argument of showing (10) is of $o_p(1)$ uniformly in $y \in \mathbb{R}$, we can obtain

$$
(11) = o_p(1) \text{ holds uniformly in } y \in \mathbb{R}.
$$

Hence,

$$
(4) = o_p(1) \text{ holds uniformly in } y \in \mathbb{R}.
$$

***For term* (5):** For fixed $y \in \mathbb{R}$, by computing the second moment of (5) and using Lemma 1 of the main paper, we obtain

$$
\begin{aligned}
& \mathrm{E}\left[|(5)|^2\right] \\
=& \mathrm{E}\left[\left(\left(Np_{jK}^*(\boldsymbol{X}_i) - \frac{1}{\pi_j(\boldsymbol{X}_i)}\right) T_{ji} I(Y_{ji} \le y) - \mathrm{E}\left[\left(Np_{jK}^*(\boldsymbol{X}_i) - \frac{1}{\pi_j(\boldsymbol{X}_i)}\right) F_j(y|\boldsymbol{X}_i)\pi_j(\boldsymbol{X}_i)\right]\right)^2\right] \\
\le & \mathrm{E}\left[\left(Np_{jK}^*(\boldsymbol{X}_i) - \frac{1}{\pi_j(\boldsymbol{X}_i)}\right)^2 \cdot T_{ji}^2 \cdot I(Y_{ji} \le y)^2\right] \\
\le & \mathrm{E}\left[\left(Np_{jK}^*(\boldsymbol{X}_i) - Np_j(\boldsymbol{X}_i)\right)^2\right] = O(K_j^{-2\alpha}) = o(1),
\end{aligned}
$$

which implies that for fixed $y \in \mathbb{R}$, (5) is of $o_p(1)$. Using a similar argument of showing (10) is of $o_p(1)$ uniformly in $y \in \mathbb{R}$, we can obtain that (5) is of $o_p(1)$ uniformly in $y \in \mathbb{R}$.

For term **(6):** By Lemma 1, Assumption 6, we can derive

$$
\begin{aligned}
&\sup_{y\in\mathbb{R}}\left|\sqrt{N}\mathrm{E}\left[F_j(y|\boldsymbol{X})\pi_j(\boldsymbol{X})\left(Np^*_{jK}(\boldsymbol{X})-\frac{1}{\pi_j(\boldsymbol{X})}\right)\right]\right| \\
\leq &\sqrt{N}\mathrm{E}[|Np^*_{jK}(\boldsymbol{X})-\pi_j(\boldsymbol{X})^{-1}|^2]^{\frac{1}{2}} = \sqrt{N}\cdot O\left(K_j^{-\alpha}\right) = o(1).
\end{aligned}
$$

For term **(7):** Using Mean Value Theorem, we have

$$
\begin{aligned}
&\sqrt{N}\int_{\mathcal{X}} F_j(y|x)\pi_j(x)(N\hat{p}_{jK}(x)-Np^*_{jK}(x))dF_X(x) \\
=&\sqrt{N}\int_{\mathcal{X}} F_j(y|x)\pi_j(x)\rho''(\tilde{\lambda}_{jK}^\top u_{K_j}(x))u_{K_j}(x)^\top dF_X(x)(\hat{\lambda}_{jK}-\lambda^*_{jK}) \\
=&-\sqrt{N}\tilde{\Psi}_{jK}^\top(y)(\hat{\lambda}_{jK}-\lambda^*_{jK}) \\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^N[T_{ji}\rho'((\lambda^*_{ji})^\top u_{K_j}(\boldsymbol{X}_i))-1]\tilde{\Psi}_{jK}^\top(y)\tilde{\Sigma}_{jK}^{-1}u_{K_j}(\boldsymbol{X}_i) \text{ (using (3))} \\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^N[T_{ji}\rho'((\lambda^*_{ji})^\top u_{K_j}(\boldsymbol{X}_i))-1]\tilde{Q}_K(\boldsymbol{X}_i;y).
\end{aligned}
$$

Therefore, the term (7) is exactly zero.

For term **(8):** We can decompose (8) as follows:

$$
\begin{aligned}
(8) =& \frac{1}{\sqrt{N}}\sum_{i=1}^N[T_{ji}\rho'((\lambda^*_{jK})^\top u_{K_j}(\boldsymbol{X}_i))-1]\left\{\tilde{\Psi}_{jK}^\top(y)\tilde{\Sigma}_{jK}^{-1}-\Psi_{jK}^\top(y)\Sigma_{jK}^{-1}\right\}u_{K_j}(\boldsymbol{X}_i) \\
=&\left\{\tilde{\Psi}_{jK}^\top(y)\tilde{\Sigma}_{jK}^{-1}-\Psi_{jK}^\top(y)\Sigma_{jK}^{-1}\right\}\sqrt{N}\hat{G}'_{jK}(\lambda^*_{jK}) \\
=&\left\{\tilde{\Psi}_{jK}(y)-\Psi_{jK}(y)\right\}^\top\tilde{\Sigma}_{jK}^{-1}\sqrt{N}\hat{G}'_{jK}(\lambda^*_{K})+\Psi_{jK}^\top(y)\left\{\tilde{\Sigma}_{jK}^{-1}-\Sigma_{jK}^{-1}\right\}\sqrt{N}\hat{G}'_{jK}(\lambda^*_{jK}).
\end{aligned} \tag{15}
$$

Consider the first term in (15). In the proof of Lemma 2 of the main paper, we proved that

$$
\|\hat{G}'_{jK}(\lambda^*_{jK})\| = O_p\left(\sqrt{\frac{K_j}{N}}\right). \tag{16}
$$

By Mean Value Theorem, we have

$$\tilde{\Psi}_{jK}(y) - \Psi_{jK}(y) = -\int_{\mathcal{X}} F_j(y|x)\pi_j(x)\left[\rho''(\tilde{\lambda}_{jK}^\top u_{K_j}(x)) - \rho''((\lambda_{jK}^*)^\top u_{K_j}(x))\right] u_{K_j}(x) dF_X(x) = \frac{W_{jK}^{(3)}(y)}{\sqrt{N}}.$$

Note that the matrix $\tilde{\Sigma}_{jK}$ is negative definite with probability approaching one, $\lambda_{\min}(\tilde{\Sigma}_{jK}^{-1}) = \lambda_{\max}(\tilde{\Sigma}_{jK})^{-1} < 0$ and $\lambda_{\max}(\tilde{\Sigma}_{jK})$ is bounded away from zero with probability approaching one, then $|\lambda_{\min}(\tilde{\Sigma}_{jK}^{-1})| = O_p(1)$. Using (13) and (16), we have

$$\begin{aligned}
&\sup_{y\in\mathbb{R}} |(\tilde{\Psi}_{jK}(y) - \Psi_{jK}(y))^\top \tilde{\Sigma}_{jK}^{-1} \sqrt{N} \hat{G}_{jK}'(\lambda_{jK}^*)| \\
\leq &\sup_{y\in\mathbb{R}} \|(W_{jK}^{(3)}(y))^\top \tilde{\Sigma}_{jK}^{-1}\| \|\hat{G}_{jK}'(\lambda_{jK}^*)\| \\
= &\sup_{y\in\mathbb{R}} \sqrt{(W_{jK}^{(3)}(y))^\top \left(\tilde{\Sigma}_{jK}^{-1}\right)^2 W_{jK}^{(3)}(y)} \cdot \|\hat{G}_{jK}'(\lambda_{jK}^*)\| \\
\leq &\sup_{y\in\mathbb{R}} \sqrt{\lambda_{\min}^2(\tilde{\Sigma}_{jK}^{-1}) W_{jK}^{(3)}(y)^\top \cdot I_{K_j} \cdot W_{jK}^{(3)}(y)} \cdot \|\hat{G}_{jK}'(\lambda_{jK}^*)\| \\
\leq &O_p(1) O_p(K_j^{\frac{1}{2}}) O_p\left(\sqrt{\frac{K_j}{N}}\right) = O_p\left(\sqrt{\frac{K_j^2}{N}}\right).
\end{aligned} \tag{17}$$

For the second term in (15),

$$\begin{aligned}
&\sup_{y\in\mathbb{R}} |\Psi_{jK}^\top(y)(\tilde{\Sigma}_{jK}^{-1} - \Sigma_{jK}^{-1})\sqrt{N}\hat{G}_{jK}'(\lambda_{jK}^*)| \\
= &\sup_{y\in\mathbb{R}} \sqrt{N} |\hat{G}_{jK}'(\lambda_{jK}^*)^\top \tilde{\Sigma}_{jK}^{-1}(\Sigma_{jK} - \tilde{\Sigma}_{jK})\Sigma_{jK}^{-1}\Psi_{jK}(y)| \\
\leq &\sqrt{N}\|\hat{G}_{jK}'(\lambda_{jK}^*)\| \cdot \sup_{y\in\mathbb{R}} \|\Psi_{jK}(y)\| \cdot \|\tilde{\Sigma}_{jK}^{-1}(\Sigma_{jK} - \tilde{\Sigma}_{jK})\Sigma_{jK}^{-1}\| \\
= &\sqrt{N}\|\hat{G}_{jK}'(\lambda_{jK}^*)\| \cdot \sup_{y\in\mathbb{R}} \|\Psi_{jK}(y)\| \cdot \mathrm{tr}\left(\tilde{\Sigma}_{jK}^{-1}(\Sigma_{jK} - \tilde{\Sigma}_{jK})\Sigma_{jK}^{-1}\Sigma_{jK}^{-1}(\Sigma_{jK} - \tilde{\Sigma}_{jK})\tilde{\Sigma}_{jK}^{-1}\right)^{\frac{1}{2}} \\
= &\sqrt{N}\|\hat{G}_{jK}'(\lambda_{jK}^*)\| \cdot \sup_{y\in\mathbb{R}} \|\Psi_{jK}(y)\| \cdot \mathrm{tr}\left(\Sigma_{jK}^{-1}\Sigma_{jK}^{-1}(\Sigma_{jK} - \tilde{\Sigma}_{jK})\tilde{\Sigma}_{jK}^{-1}\tilde{\Sigma}_{jK}^{-1}(\Sigma_{jK} - \tilde{\Sigma}_{jK})\right)^{\frac{1}{2}} \\
\leq &\sqrt{N}\|\hat{G}_{jK}'(\lambda_{jK}^*)\| \cdot \sup_{y\in\mathbb{R}} \|\Psi_{jK}(y)\| \cdot \left|\lambda_{\max}(\Sigma_{jK}^{-1}\Sigma_{jK}^{-1})\right|^{\frac{1}{2}} \cdot \mathrm{tr}\left((\Sigma_{jK} - \tilde{\Sigma}_{jK})\tilde{\Sigma}_{jK}^{-1}\tilde{\Sigma}_{jK}^{-1}(\Sigma_{jK} - \tilde{\Sigma}_{jK})\right)^{\frac{1}{2}} \\
= &\sqrt{N}\|\hat{G}_{jK}'(\lambda_{jK}^*)\| \cdot \sup_{y\in\mathbb{R}} \|\Psi_{jK}(y)\| \cdot \left|\lambda_{\max}(\Sigma_{jK}^{-1}\Sigma_{jK}^{-1})\right|^{\frac{1}{2}} \cdot \mathrm{tr}\left(\tilde{\Sigma}_{jK}^{-1}\tilde{\Sigma}_{jK}^{-1}(\Sigma_{jK} - \tilde{\Sigma}_{jK})(\Sigma_{jK} - \tilde{\Sigma}_{jK})\right)^{\frac{1}{2}} \\
\leq &\sqrt{N}\|\hat{G}_{jK}'(\lambda_{jK}^*)\| \cdot \sup_{y\in\mathbb{R}} \|\Psi_{jK}(y)\| \cdot \left|\lambda_{\max}(\Sigma_{jK}^{-1}\Sigma_{jK}^{-1})\right|^{\frac{1}{2}} \cdot \left|\lambda_{\max}(\tilde{\Sigma}_{jK}^{-1}\tilde{\Sigma}_{jK}^{-1})\right|^{\frac{1}{2}} \cdot \mathrm{tr}\left((\Sigma_{jK} - \tilde{\Sigma}_{jK})(\Sigma_{jK} - \tilde{\Sigma}_{jK})\right)^{\frac{1}{2}}
\end{aligned} \tag{18}$$

$$=\sqrt{N}\|\hat{G}'_{jK}(\lambda^*_{jK})\| \cdot \sup_{y\in\mathbb{R}} \|\Psi_{jK}(y)\| \cdot \left|\lambda_{\min}(\Sigma_{jK}^{-1})\right| \cdot \left|\lambda_{\min}(\tilde{\Sigma}_{jK}^{-1})\right| \cdot \left\|\Sigma_{jK} - \tilde{\Sigma}_{jK}\right\|,$$

where the second and third inequalities follow from the fact that $\text{tr}(AB) \le \lambda_{\max}(B)\text{tr}(A)$ for any symmetric B and positive semidefinite matrix A.

We now estimate the probability order of $\left\|\Sigma_{jK} - \tilde{\Sigma}_{jK}\right\|$. Using Mean Value Theorem and triangle inequality, we have

$$\begin{aligned}
&\left\|\Sigma_{jK} - \tilde{\Sigma}_{jK}\right\| \qquad (19)\\
\le&\left\|\mathrm{E}\left[\pi_j(\boldsymbol{X})\rho''((\lambda^*_{jK})^\top u_{K_j}(\boldsymbol{X}))u_{K_j}(\boldsymbol{X})u_{K_j}(\boldsymbol{X})^\top\right] - \frac{1}{N}\sum_{i=1}^N T_{ji}\rho''((\lambda^*_{jK})^\top u_{K_j}(\boldsymbol{X}_i))u_{K_j}(\boldsymbol{X}_i)u_{K_j}(\boldsymbol{X}_i)^\top\right\|\\
&+\left\|\frac{1}{N}\sum_{i=1}^N T_{ji}\rho'''(\xi_{j3}(\boldsymbol{X}_i))u_{K_j}(\boldsymbol{X}_i)u_{K_j}(\boldsymbol{X}_i)^\top \cdot (\tilde{\lambda}_{jK}-\lambda^*_{jK})^\top u_{K_j}(\boldsymbol{X}_i)\right\|\\
\le&O_p\left(\zeta(K_j)\sqrt{\frac{K_j}{N}}\right) + \sup_{x\in\mathcal{X}}|\rho'''(\xi_{j3}(x))| \cdot \left\|\frac{1}{N}\sum_{i=1}^N u_{K_j}(\boldsymbol{X}_i)u_{K_j}(\boldsymbol{X}_i)^\top\right\| \cdot \|\tilde{\lambda}_{jK}-\lambda^*_{jK}\| \cdot \zeta(K_j)\\
=&O_p\left(\zeta(K_j)\sqrt{\frac{K_j}{N}}\right) + O_p(1)\cdot O_p\left(\sqrt{K_j}\right)\cdot O_p\left(\zeta(K_j)\sqrt{\frac{K_j}{N}}\right) = O_p\left(\zeta(K_j)\sqrt{\frac{K_j^2}{N}}\right),
\end{aligned}$$

where $\xi_{j3}(x)$ lies between $(\lambda^*_{jK})^\top u_{K_j}(x)$ and $\tilde{\lambda}_{jK}^\top u_{K_j}(x)$. Note that

$$\sup_{y\in\mathbb{R}}\|\Psi_{jK}(y)\| \le \mathrm{E}[|\rho''((\lambda^*_{jK})^\top u_{K_j}(\boldsymbol{X}))|\|u_{K_j}(\boldsymbol{X})\|] \le O(1)\cdot \mathrm{E}[\|u_{K_j}(\boldsymbol{X}))\|^2]^{\frac{1}{2}} = O(\sqrt{K_j}). \qquad (20)$$

Combining (18)-(20), we can obtain that

$$\begin{aligned}
&\sup_{y\in\mathbb{R}}\|\Psi_{jK}^\top(y)(\tilde{\Sigma}_{jK}^{-1}-\Sigma_{jK}^{-1})\sqrt{N}\hat{G}'_{jK}(\lambda^*_{jK})\|\\
=&\sqrt{N}O_p\left(\sqrt{\frac{K_j}{N}}\right)O(\sqrt{K_j})O_p(1)O_p(1)O_p\left(\zeta(K_j)\sqrt{\frac{K_j^2}{N}}\right) = O_p\left(\zeta(K_j)\sqrt{\frac{K_j^4}{N}}\right),
\end{aligned}$$

then with (15), (17) and Assumption 6, we have

$$(8) = O_p\left(\sqrt{\frac{K_j^2}{N}}\right) + O_p\left(\zeta(K_j)\sqrt{\frac{K_j^4}{N}}\right) = o_p(1) \text{ holds uniformly in } y\in\mathbb{R}.$$

For term (9)**:** Computing the second moment of (9) yields:

$$\begin{aligned}
&\mathrm{E}\left[\left([T_{ji}\rho'((\lambda_{jK}^*)^\top u_{K_j}(\boldsymbol{X}_i))-1]Q_{jK}(\boldsymbol{X}_i,y)+\frac{F_j(y|\boldsymbol{X}_i)}{\pi_j(\boldsymbol{X}_i)}(T_{ji}-\pi_j(\boldsymbol{X}_i))\right)^2\right]\\
\leq{}& 2\cdot\mathrm{E}\left[\left(\rho'((\lambda_{jK}^*)^\top u_{K_j}(\boldsymbol{X}_i))-\frac{1}{\pi_j(\boldsymbol{X}_i)}\right)^2 T_{ji}^2 Q_{jK}(\boldsymbol{X}_i,y)^2\right]+2\cdot\mathrm{E}\left[\frac{(F_j(y|\boldsymbol{X}_i)+Q_{jK}(\boldsymbol{X}_i,y))^2}{\pi_j(\boldsymbol{X}_i)^2}(T_{ji}-\pi_j(\boldsymbol{X}_i))^2\right].
\end{aligned}\tag{21}$$

Note that $Q_{jK}(\boldsymbol{X},y)$ can be rewritten as:

$$Q_{jK}(\boldsymbol{X},y)=\beta_K(y)^\top u_{K_j}(\boldsymbol{X})$$

where

$$\begin{aligned}
\beta_K(y):={}&\arg\min_{\beta}\mathrm{E}\left[\pi_j(\boldsymbol{X})\left(-\rho''\left(\left(\lambda_{jK}^*\right)^\top u_{K_j}(\boldsymbol{X})\right)\right)\left\{F_j(y|\boldsymbol{X})-\beta^\top u_{K_j}(\boldsymbol{X})\right\}^2\right]\\
={}&-\mathrm{E}\left[\pi_j(\boldsymbol{X})\rho''\left(\left(\lambda_{jK}^*\right)^\top u_{K_j}(\boldsymbol{X})\right)u_{K_j}(\boldsymbol{X})u_{K_j}(\boldsymbol{X})^\top\right]^{-1}\mathrm{E}\left[\pi_j(\boldsymbol{X})\rho''\left(\left(\lambda_{jK}^*\right)^\top u_{K_j}(\boldsymbol{X})\right)F_j(y|\boldsymbol{X})\right].
\end{aligned}$$

Thus $Q_{jK}(\boldsymbol{X},y)$ is the weighted L^2-projection (with the weighted measure $-\pi_j(x)\rho''((\lambda_{jK}^*)^\top u_{K_j}(x))dF_X(x)$) of $F_j(y|\boldsymbol{X})$ on the space spanned by $u_{K_j}(\boldsymbol{X})$, and the weighting function $-\pi_j(x)\rho''\left(\left(\lambda_{jK}^*\right)^\top u_{K_j}(x)\right)$ is uniformly bounded away from zero and infinity. Therefore, by definition, for any $\beta\in\mathbb{R}^{K_j}$,

$$\begin{aligned}
&\mathrm{E}\left[\{F_j(y|\boldsymbol{X})-Q_{jK}(\boldsymbol{X},y)\}^2\right]\\
=&\mathrm{E}\left[\frac{\pi_j(\boldsymbol{X})\left(-\rho''\left(\left(\lambda_{jK}^*\right)^\top u_{K_j}(\boldsymbol{X})\right)\right)}{\pi_j(\boldsymbol{X})\left(-\rho''\left(\left(\lambda_{jK}^*\right)^\top u_{K_j}(\boldsymbol{X})\right)\right)}\left\{F_j(y|\boldsymbol{X})-\beta_K(y)^\top u_{K_j}(\boldsymbol{X})\right\}^2\right]\\
\leq&\frac{1}{\eta_1}\cdot\frac{1}{\inf_{\gamma\in\Gamma_1}|\rho''(\gamma)|}\cdot\mathrm{E}\left[\pi_j(\boldsymbol{X})\left(-\rho''\left(\left(\lambda_{jK}^*\right)^\top u_{K_j}(\boldsymbol{X})\right)\right)\left\{F_j(y|\boldsymbol{X})-\beta_K(y)^\top u_{K_j}(\boldsymbol{X})\right\}^2\right]\\
\leq&\frac{1}{\eta_1}\cdot\frac{1}{\inf_{\gamma\in\Gamma_1}|\rho''(\gamma)|}\cdot\mathrm{E}\left[\pi_j(\boldsymbol{X})\left(-\rho''\left(\left(\lambda_{jK}^*\right)^\top u_{K_j}(\boldsymbol{X})\right)\right)\left\{F_j(y|\boldsymbol{X})-\beta^\top u_{K_j}(\boldsymbol{X})\right\}^2\right]\\
\leq&\frac{1}{\eta_1}\cdot\frac{\sup_{\gamma\in\Gamma_1}|\rho''(\gamma)|}{\inf_{\gamma\in\Gamma_1}|\rho''(\gamma)|}\cdot\mathrm{E}\left[\left\{F_j(y|\boldsymbol{X})-\beta^\top u_{K_j}(\boldsymbol{X})\right\}^2\right].
\end{aligned}$$

Taking infimum over $\beta\in\mathbb{R}^{K_j}$ in above expression yields:

$$\mathrm{E}\left[\left\{F_j(y|\boldsymbol{X})-\beta_K(y)^\top u_{K_j}(\boldsymbol{X})\right\}^2\right]\leq\frac{1}{\eta_1}\cdot\frac{\sup_{\gamma\in\Gamma_1}|\rho''(\gamma)|}{\inf_{\gamma\in\Gamma_1}|\rho''(\gamma)|}\cdot\sup_{\beta\in\mathbb{R}^{K_j}}\mathrm{E}\left[\left\{F_j(y|\boldsymbol{X})-\beta^\top u_{K_j}(\boldsymbol{X})\right\}^2\right]\to 0,$$

as $K_j \to \infty$. By (21), Lemmas 1-2, Assumption 6, and the result above, we obtain that

$$\begin{aligned}
&\mathrm{E}\left[\left([T_{ji}\rho'((\lambda_{jK}^*)^\top u_{K_j}(\boldsymbol{X}_i)) - 1]Q_{jK}(\boldsymbol{X}_i, y) + \frac{F_j(y|\boldsymbol{X}_i)}{\pi_j(\boldsymbol{X}_i)}(T_{ji} - \pi_j(\boldsymbol{X}_i))\right)^2\right] \\
\leq &2 \cdot \sup_{x\in\mathcal{X}}\left(\rho'((\lambda_{jK}^*)^\top u_{K_j}(x)) - \frac{1}{\pi_j(x)}\right)^2 \cdot \mathrm{E}\left[Q_{jK}(\boldsymbol{X}_i, y)^2\right] + \frac{2}{\eta_1^2} \cdot \mathrm{E}\left[\{F_j(y|\boldsymbol{X}) - Q_{jK}(\boldsymbol{X}, y)\}^2\right] \\
\leq &O(\zeta(K_j)^2 K_j^{-2\alpha}) \cdot O(1) + o(1) = o(1).
\end{aligned}$$

By Chebyshev's inequality, we have that for every $y \in \mathbb{R}$,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N\left\{[T_{ji}\rho'((\lambda_{jK}^*)^\top u_{K_j}(\boldsymbol{X}_i)) - 1]Q_{jK}(\boldsymbol{X}_i, y) + \frac{F_j(y|\boldsymbol{X}_i)}{\pi_j(\boldsymbol{X}_i)}(T_{ji} - \pi_j(\boldsymbol{X}_i))\right\} \xrightarrow{p} 0. \tag{22}$$

By Assumption 8,

$$[T_{ji}\rho'((\lambda_{jK}^*)^\top u_{K_j}(\boldsymbol{X}_i)) - 1]Q_{jK}(\boldsymbol{X}_i, y) + \frac{F_j(y|\boldsymbol{X}_i)}{\pi_j(\boldsymbol{X}_i)}(T_{ji} - \pi_j(\boldsymbol{X}_i))$$

is Lipschitz continuous in $y \in \mathbb{R}$. Then Condition (5.3) of Andrews (1994) is satisfied. By Theorems 4 and 5 of Andrews (1994), the convergence result (22) holds uniformly in $y \in \mathbb{R}$. Hence (9) is of $o_p(1)$ uniformly in $y \in \mathbb{R}$.

Therefore, we can obtain that

$$\sup_{y\in\mathbb{R}}\left|\sqrt{N}\left\{\hat{F}_j(y) - F_j^0(y) - \frac{1}{N}\sum_{i=1}^N \psi_j(Y_{ji}, \boldsymbol{X}_i, T_{ji}; y)\right\}\right| = o_p(1).$$

5 Simplified asymptotic variance under complete data

If there are no missing data (i.e., $T_{1i} = \cdots = T_{di} = 1$), then $\pi_j(\boldsymbol{X}_i) = 1$ and $\eta(\boldsymbol{X}_i) = 1$. In such a case, the notation in the asymptotic variance V_0, appearing in Assumption 15, simplifies to

$$\varphi(\boldsymbol{T}_i, \boldsymbol{X}_i, \boldsymbol{U}_i; \theta_0) = \ell_{\theta\theta}(U_{1i}, ..., U_{di}; \theta_0),$$

$$\phi_j(T_{ji}, \boldsymbol{X}_i, U_{ji}) = I(U_{ji} \leq v),$$

and

$$W_j(T_{ji}, \boldsymbol{X}_i, U_{ji}; \theta_0) = \mathrm{E}\left[\ell_{\theta j}(U_{1s}, ..., U_{ds}; \theta_0)\{\phi_j(T_{ji}, \boldsymbol{X}_i, U_{ji}; U_{js}) - U_{js}\}|U_{ji}, \boldsymbol{X}_i, T_{ji}\right]$$

$$= \mathrm{E}\left[\ell_{\theta j}(U_{1s}, ..., U_{ds}; \theta_0) \left\{ I(U_{ji} \leq U_{js}) - U_{js} \right\} |U_{ji}\right], \ s \neq i.$$

Then

$$\Sigma = Var\left\{\varphi(\boldsymbol{T}_i, \boldsymbol{X}_i, \boldsymbol{U}_i; \theta_0) + \sum_{j=1}^{d} W_j(U_{ji}; \theta_0)\right\} = Var\left\{\ell_\theta(U_{1i}, ..., U_{di}; \theta_0) + \sum_{j=1}^{d} W_j(U_{ji}; \theta_0)\right\}$$
$$= Var\left\{\ell_\theta(U_{1i}, ..., U_{di}; \theta_0) + \sum_{j=1}^{d} \mathrm{E}\left[\ell_{\theta j}(U_{1s}, ..., U_{ds}; \theta_0) \left\{ I(U_{ji} \leq U_{js}) - U_{js} \right\} |U_{ji}\right]\right\}, \ s \neq i \ ,$$

and $B = -\mathrm{E}[\ell_{\theta\theta}(U_{1i}, ..., U_{di}; \theta_0)]$. Hence the asymptotic variance $V_0 = B^{-1}\Sigma B^{-1}$ reduces to that of Genest, Ghoudi, and Rivest (1995) and Chen and Fan (2005).

6 Proof of Theorem 6

It suffices to show $\|\widehat{B} - B\| \xrightarrow{p} 0$, and $\|\Sigma - \widehat{\Sigma}\| \xrightarrow{p} 0$ can be similarly established. Recall from (D.3) of the main paper that

$$\sup_{\theta \in \Theta: \|\theta - \theta_0\| = o(1)} \left\| \sum_{i=1}^{N} I(T_{1i} = 1, ..., T_{di} = 1)\hat{q}_K(\boldsymbol{X}_i)\ell_{\theta\theta}(\hat{F}_1(Y_{1i}), ..., \hat{F}_d(Y_{di}); \theta) - \mathrm{E}[\ell_{\theta\theta}(U_{1i}, ..., U_{di}; \theta_0)] \right\| = o_p(1). \tag{23}$$

Rewrite (23) slightly as

$$\sup_{\theta \in \Theta: \|\theta - \theta_0\| = o(1)} \left\| \sum_{i=1}^{N} I(T_{1i} = 1, ..., T_{di} = 1)\hat{q}_K(\boldsymbol{X}_i)\ell_{\theta\theta}(\hat{U}_{1i}, ..., \hat{U}_{di}; \theta) - \mathrm{E}\left[\frac{I(T_{1i} = 1, ..., T_{di} = 1)}{\eta(\boldsymbol{X}_i)}\ell_{\theta\theta}(U_{1i}, ..., U_{di}; \theta_0)\right] \right\| = o_p(1). \tag{24}$$

Since $\|\hat{\theta} - \theta_0\| = O_p(N^{-1/2})$, we have that for any $\epsilon > 0$ and any fixed constant $C > 0$,

$$\Pr(\|\hat{\theta} - \theta_0\| > C \cdot N^{-1/3}) \to 0 \quad \text{as} \quad N \to \infty. \tag{25}$$

Therefore, by (24) and (25),

$$\Pr(\|\widehat{B} - B\| > \epsilon)$$
$$= \Pr\left(\left\| \sum_{i=1}^{N} I(T_{1i} = 1, ..., T_{di} = 1)\hat{q}_K(\boldsymbol{X}_i)\ell_{\theta\theta}(\hat{U}_{1i}, ..., \hat{U}_{di}; \hat{\theta}) - \mathrm{E}\left[\frac{I(T_{1i} = 1, ..., T_{di} = 1)}{\eta(\boldsymbol{X}_i)}\ell_{\theta\theta}(U_{1i}, ..., U_{di}; \theta_0)\right] \right\| > \epsilon\right)$$
$$\leq \Pr\left(\left\| \sum_{i=1}^{N} I(T_{1i} = 1, ..., T_{di} = 1)\hat{q}_K(\boldsymbol{X}_i)\ell_{\theta\theta}(\hat{U}_{1i}, ..., \hat{U}_{di}; \hat{\theta}) \right.\right.$$
$$\left.\left. - \mathrm{E}\left[\frac{I(T_{1i} = 1, ..., T_{di} = 1)}{\eta(\boldsymbol{X}_i)}\ell_{\theta\theta}(U_{1i}, ..., U_{di}; \theta_0)\right] \right\| > \epsilon; \|\hat{\theta} - \theta_0\| \leq C \cdot N^{-1/3}\right)$$

$$+ \Pr(\|\hat{\theta} - \theta_0\| > C \cdot N^{-1/3})$$
$$\xrightarrow{N\to\infty} 0 \ ,$$

which implies that $\|\widehat{B} - B\| \xrightarrow{p} 0$.

7 Complete simulation results

7.1 Benchmark scenario

Recall the benchmark simulation experiment in the main paper Hamori, Motegi, and Zhang (2018), where we only reported partial results with respect to copula parameters and missing mechanisms. For the copula parameters, the main paper focused on the Clayton copula with $\alpha_0 = 6.000$ (Kendall's $\tau = 0.75$) and the Gumbel copula with $\gamma_0 = 4.000$ ($\tau = 0.75$). In the supplemental material, we add Clayton with $\alpha_0 = 1.636$ ($\tau = 0.45$) and Gumbel with $\gamma_0 = 1.818$ ($\tau = 0.45$). The smaller value of τ implies that there is a weaker association among (Y_{1i}, Y_{2i}, X_i). For the missing mechanisms, the main paper focused on Case C (i.e., MCAR with $\mathrm{E}[T_{2i}] = 0.6$) and Case D (i.e., MAR with $\mathrm{E}[T_{2i}] = 0.6$). In the supplemental material, we add Case A (i.e., MCAR with $\mathrm{E}[T_{2i}] = 0.8$) and Case B (i.e., MAR with $\mathrm{E}[T_{2i}] = 0.8$). The average amount of missing observations is smaller in Cases A-B than in Cases C-D.

See Tables 1-6 for complete results on the benchmark scenario. The smaller value of Kendall's τ generally has a positive impact on the performance of each estimator. See, for instance, the calibration estimator assisted by the data-driven selection of K_2^* in Case D with the Clayton copula and $N = 250$ (Table 1). The bias, standard deviation, and root mean squared error (RMSE) are $\{0.036, 0.253, 0.256\}$ for $\alpha_0 = 1.636$ ($\tau = 0.45$) and $\{-0.259, 0.653, 0.703\}$ for $\alpha_0 = 6.000$ ($\tau = 0.75$). It is a reasonable result since the weaker association among (Y_{1i}, Y_{2i}, X_i) should imply the greater amount of information given sample size N.

The smaller missing probability also has a positive impact on the performance of each estimator. See, for instance, the calibration estimator assisted by the data-driven K_2^* under the Clayton copula with $\alpha_0 = 6.000$ and $N = 250$ (Table 1). The bias, standard deviation, and RMSE are $\{-0.140, 0.618, 0.634\}$ in Case B (MAR with $\mathrm{E}[T_{2i}] = 0.8$) and $\{-0.259, 0.653, 0.703\}$ in Case D (MAR with $\mathrm{E}[T_{2i}] = 0.6$). It is also a reasonable result since the smaller missing probability implies the greater amount of information available given N.

In the main paper, we confirmed that the listwise-deletion estimator is inconsistent under MAR. We also noted that the listwise deletion results in *positive* bias under the Clayton copula and *negative* bias under the Gumbel copula. Those results are essentially a consequence of our simulation design and the tail-dependence properties of the two copulas. In this supplemental material, we provide a precise reason for the positive bias under Clayton and the negative bias under Gumbel.

Consider the Clayton copula, which has a lower-tail dependence and upper-tail independence. When (Y_{1i}, Y_{2i}, X_i) are jointly drawn from Clayton, a small value of X_i tends to be accompanied by jointly small values of (Y_{1i}, Y_{2i}), whereas a large value of X_i is not necessarily accompanied by jointly large values of (Y_{1i}, Y_{2i}). Recall that the true propensity score function of Y_{2i} is

$$\pi_2(x_i) = \Pr(T_{2i} = 1 \mid X_i = x_i) = \frac{1}{1 + \exp[a + bx_i]}. \tag{26}$$

In view of (26), the smaller (larger) X_i implies the higher (lower) probability of observing Y_{2i} under MAR. As a result, observed pairs of (Y_{1i}, Y_{2i}) exhibit a deceivingly strong association. The listwise deletion literally accepts the strong association observed, since it just puts uniform weights for all individuals with complete data. This is why the listwise deletion *over-estimates* the association between Y_{1i} and Y_{2i} under the Clayton copula. Analogously, it is natural to see the negative bias under the Gumbel copula, which has a lower-tail independence and upper-tail dependence.

7.2 Benchmark scenario with B-splines

Recall that the nonparametric estimator of Hirano, Imbens, and Ridder (2003) and the proposed calibration estimator require an approximation sieve $u_{K_2}(X)$. In the main paper and Section 7.1 of this supplemental material, the approximation sieve is specified to be power series $u_{K_2}(X) = (1, X, X^2, \ldots, X^{K_2 - 1})^\top$. In this section, we use B-splines in order to check if simulation results change.

B-splines require a choice of *order* $R \geq 2$ and *knots* (see Hastie, Tibshirani, and Friedman, 2009, Section 5 for an overview of B-splines). Popularly used orders are $R \in \{2, 3, 4\}$ (i.e., linear, quadratic, and cubic B-splines), and we use all those values. We use the minimum of covariate X as left outside knots; the $\{25\%, 50\%, 75\%\}$ quantiles of X as inside knots; the maximum of X as right outside knots. The dimension of the resulting base functions $u_{K_2}(X)$ is $K_2 = R + Q$, where Q is the number of inside knots. Since $R \in \{2, 3, 4\}$ and $Q = 3$ in our case, we have that $K_2 \in \{5, 6, 7\}$. It is possible to use more or less quantiles as inside knots, but we leave it a future task to explore such a possibility.

See Tables 7-10 for results. Recall from Tables 1-6 that the nonparametric estimator has a serious sensitivity to a choice of K_2 and that the data-driven selection of K_2^* alleviates the sensitivity substantially. Tables 7-10 indicate that using B-splines is another compelling way to alleviate the sensitivity of the nonparametric estimator. The bias, standard deviation, and RMSE of the nonparametric estimator are all similar across $R \in \{2, 3, 4\}$ in each case, and they are as sharp as the performance of the nonparametric estimator with power series and data-driven selection of K_2^*.

When Kendall's $\tau = 0.45$, the calibration estimator based on the B-spline with any $R \in \{2, 3, 4\}$ performs as well as the calibration estimator with power series and data-driven K_2^*. When Kendall's $\tau = 0.75$, the calibration estimators based on B-splines exhibit some sensitivity to the order R, and their performances are sometimes worse than the performance of the calibration estimator with power series and data-driven K_2^*. See, for example, Case D with the Clayton copula, $\alpha_0 = 6.000$, and $N = 500$ (Tables 2 and 8). The bias, standard deviation, and RMSE are $\{-0.118, 0.458, 0.473\}$ for the B-spline with $R = 2$; $\{-0.285, 0.565, 0.633\}$ for $R = 3$; $\{-0.429, 0.567, 0.711\}$ for $R = 4$; $\{-0.142, 0.472, 0.493\}$ for the power series with data-driven K_2^*. In this example, the linear B-spline performs slightly better than the power series with data-driven K_2^* while the quadratic and cubic B-splines perform worse than that. This result suggests that using B-splines might have a negative impact on the small sample performance of the calibration estimator when there exists a strong association among target variables and covariates.

In large sample, a choice of sieve basis functions should not matter. Focusing on the same example as above, the performance of the calibration estimator with the quadratic B-spline indeed keeps improving as sample size increases: $\{-0.328, 0.709, 0.782\}$ for $N = 250$, $\{-0.285, 0.565, 0.633\}$ for $N = 500$, and $\{-0.226, 0.466, 0.518\}$ for $N = 1000$ (Table 8). The same goes for the linear and cubic B-splines.

7.3 Extended scenarios

Recall the extended simulation experiments in Section 6.2 of the main paper. The first scenario concerns a misspecified missing mechanism, while the second scenario concerns two covariates. We provide complete simulation results on the first scenario in Section 7.3.1 and the second scenario in Section 7.3.2. In both scenarios, the approximation sieve is specified to be power series. It is left as a future task to consider B-splines under those scenarios.

7.3.1 Misspecified missing mechanism

Recall that the true propensity score function is specified as

$$\pi_2(x_i, y_{1i}) := \Pr(T_{2i} = 1 \mid X_i = x_i, Y_{1i} = y_{1i}) = \frac{1}{1 + \exp[-0.42 + 0.2x_i + 0.2y_{1i}]}. \tag{27}$$

In the main paper, we performed the nonparametric estimation of Hirano, Imbens, and Ridder (2003) and the calibration estimation by constructing approximation sieves based on the power series of X_i only:

$$u_{K_2}(X_i) = (1, X_i, X_i^2, \ldots, X_i^{K_2-1})^\top.$$

We then used fixed $K_2 \in \{2, 3, 4\}$ or the data-driven selection of K_2^* based on the covariate balancing principle.

In this supplemental material, we consider two extra estimators that are expected to have sharp performances by construction. The first one is the parametric estimator whose propensity score model is correctly specified relative to the true propensity score (27). The second estimator is the calibration estimator whose approximation sieve consists of power series of X_i and Y_{1i}. When $K_2 = 10$, the approximation sieve is defined as

$$u_{10}(X_i, Y_{1i}) = (1, X_i, Y_{1i}, X_i^2, Y_{1i}^2, X_i Y_{1i}, X_i^3, Y_{1i}^3, X_i^2 Y_{1i}, X_i Y_{1i}^2)^\top.$$

When $K_2 \leq 10$, $u_{K_2}(X_i, Y_{1i})$ consists of the first K_2 elements of $u_{10}(X_i, Y_{1i})$. For the second estimator, we use $K_2 = 3$ (i.e., only the first moments of X_i and Y_{1i}), $K_2 = 6$ (i.e., the second moments added), and $K_2 = 10$ (i.e., the third moments added). We also use the data-driven K_2^* with a choice set $K_2 \in \{1, \ldots, 10\}$.

Further, we only reported results on the Clayton copula with $\alpha_0 = 6.000$ in the main paper. Here we report complete results in Tables 11-14, where we cover Clayton with $\alpha_0 \in \{1.636, 6.000\}$ and Gumbel with $\gamma_0 \in \{1.818, 4.000\}$.

The simulation results in Tables 11-14 indicate that the calibration estimator based on the power series of X_i performs as well as the two extra competitors in terms of bias and standard deviation. See, for example, the Gumbel copula with $\gamma_0 = 4.000$ and $N = 1000$ (Table 14). The bias, standard deviation, and RMSE are $\{-0.074, 0.171, 0.186\}$ when the parametric approach is taken; $\{-0.041, 0.166, 0.171\}$ when the power series of X_i and data-driven K_2^* are used; $\{-0.057, 0.169, 0.178\}$ when the power series of (X_i, Y_{1i}) and data-driven K_2^* are used. Hence, in our set-up, the extra impact of Y_{1i} on the propensity score is well captured by the power series of X_i only. That is not a surprising result since Y_{1i} and X_i are associated with each other through the copula.

7.3.2 Two covariates

Recall that there are $r = 2$ covariates $\boldsymbol{X}_i = (X_{1i}, X_{2i})^\top$ in the second extended scenario. In the main paper, we only reported results on the Clayton copula with $\alpha_0 = 6.000$. In this supplemental material, we report complete results in Tables 15-16, where we cover Clayton with $\alpha_0 \in \{1.636, 6.000\}$ and Gumbel with $\gamma_0 \in \{1.818, 4.000\}$.

The results are consistent with the benchmark scenario with a single covariate. See, for example, the Clayton copula with $\alpha_0 = 6.000$ and $N = 1000$ (Table 15). First, the nonparametric estimators with $K_2 \in \{3, 6, 10\}$ exhibit serious sensitivity, and we observe an extremely large bias of -1.579 when $K_2 = 10$. Second, the data-driven selection of K_2^* stabilizes the performance of the nonparametric estimator successfully; the bias, standard deviation, and RMSE are $\{-0.108, 1.265, 1.270\}$. Third, the calibration estimators with $K_2 \in \{3, 6, 10\}$ have stable and sharp performances. Fourth, the bias, standard deviation, and RMSE of the calibration estimator with data-driven K_2^* are $\{-0.098, 0.346, 0.360\}$, which is a remarkably sharp result. When the data-driven approach is taken, the nonparametric and calibration estimators lead to an equally small bias, but the latter leads to much smaller variance. We thus confirm that the advantage of the calibration estimator observed in the single-covariate scenario is preserved under the two-covariate scenario.

Table 1: Benchmark simulation results on Clayton copula ($N = 250$)

$\alpha_0 = 1.636$ (Kendall's $\tau = 0.45$)				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Listwise deletion	0.031, 0.249, 0.251	0.076, 0.248, 0.260	0.050, 0.302, 0.306	0.154, 0.296, 0.333
Param (correct)	0.021, 0.237, 0.238	0.039, 0.232, 0.235	0.025, 0.263, 0.264	0.044, 0.251, 0.255
Param (misspec)	−0.426, 0.163, 0.456	−0.374, 0.166, 0.409	−0.108, 0.240, 0.263	−0.091, 0.237, 0.254
Nonparam ($K_2 = 3$)	0.031, 0.244, 0.246	0.024, 0.233, 0.234	0.028, 0.259, 0.261	0.031, 0.251, 0.253
Nonparam ($K_2 = 4$)	−0.055, 0.714, 0.716	−0.021, 0.711, 0.712	−0.031, 0.745, 0.745	−0.011, 0.710, 0.710
Nonparam (CB)	0.028, 0.227, 0.229	0.034, 0.233, 0.236	0.018, 0.257, 0.257	0.037, 0.244, 0.247
Calibration ($K_2 = 3$)	0.038, 0.236, 0.239	0.025, 0.225, 0.226	0.024, 0.266, 0.267	0.035, 0.252, 0.255
Calibration ($K_2 = 4$)	0.035, 0.237, 0.239	0.040, 0.223, 0.227	0.032, 0.267, 0.269	0.034, 0.250, 0.253
Calibration (CB)	0.024, 0.228, 0.230	0.026, 0.226, 0.227	0.026, 0.263, 0.264	0.036, 0.253, 0.256

$\alpha_0 = 6.000$ (Kendall's $\tau = 0.75$)				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Listwise deletion	−0.082, 0.660, 0.665	0.134, 0.667, 0.681	−0.115, 0.754, 0.763	0.420, 0.840, 0.940
Param (correct)	−0.230, 0.612, 0.653	−0.118, 0.594, 0.606	−0.442, 0.663, 0.797	−0.248, 0.650, 0.696
Param (misspec)	−3.662, 0.328, 3.677	−3.362, 0.345, 3.380	−1.687, 0.647, 1.807	−1.367, 0.610, 1.497
Nonparam ($K_2 = 3$)	−0.181, 0.636, 0.661	−0.164, 0.610, 0.632	−0.369, 0.649, 0.747	−0.248, 0.612, 0.661
Nonparam ($K_2 = 4$)	−0.472, 2.118, 2.170	−0.505, 1.753, 1.824	−0.630, 1.985, 2.083	−0.495, 1.916, 1.978
Nonparam (CB)	−0.182, 0.590, 0.618	−0.145, 0.590, 0.607	−0.307, 0.659, 0.727	−0.167, 0.670, 0.691
Calibration ($K_2 = 3$)	−0.207, 0.605, 0.640	−0.168, 0.592, 0.615	−0.365, 0.648, 0.744	−0.266, 0.605, 0.661
Calibration ($K_2 = 4$)	−0.196, 0.608, 0.639	−0.137, 0.585, 0.601	−0.304, 0.646, 0.714	−0.274, 0.649, 0.705
Calibration (CB)	−0.186, 0.601, 0.629	−0.140, 0.618, 0.634	−0.317, 0.643, 0.717	−0.259, 0.653, 0.703

This table reports bias, standard deviation, and root mean squared error (RMSE) of each estimator with respect to $\alpha_0 \in \{1.636, 6.000\}$ after $J = 1000$ Monte Carlo trials. Cases A and B imply MCAR and MAR, respectively, with relatively low missing probability (i.e., $\mathrm{E}[T_{2i}] = 0.8$). Cases C and D imply MCAR and MAR, respectively, with relatively high missing probability (i.e., $\mathrm{E}[T_{2i}] = 0.6$). "Param (correct)" signifies the parametric estimator based on a correctly specified propensity score model. "Param (misspec)" signifies the parametric estimator based on a misspecified propensity score model. For the nonparametric estimator of Hirano, Imbens, and Ridder (2003) and the proposed calibration estimator, approximation sieves are constructed from the power series of X_i. The dimension of the approximation sieve is either fixed at $K_2 \in \{3, 4\}$ or automatically selected from $K_2 \in \{1, \ldots, 5\}$ based on the covariate balancing (CB) principle.

Table 2: Benchmark simulation results on Clayton copula ($N = 500$)

$\alpha_0 = 1.636$ (Kendall's $\tau = 0.45$)				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Listwise deletion	0.022, 0.176, 0.177	0.068, 0.181, 0.194	0.023, 0.204, 0.206	0.130, 0.213, 0.250
Param (correct)	0.012, 0.162, 0.163	0.010, 0.158, 0.158	0.003, 0.182, 0.182	0.014, 0.174, 0.175
Param (misspec)	−0.417, 0.115, 0.432	−0.378, 0.117, 0.396	−0.116, 0.169, 0.205	−0.094, 0.168, 0.192
Nonparam ($K_2 = 3$)	0.024, 0.165, 0.167	0.018, 0.165, 0.165	0.016, 0.178, 0.179	0.023, 0.169, 0.171
Nonparam ($K_2 = 4$)	−0.121, 0.731, 0.741	−0.067, 0.744, 0.747	−0.139, 0.838, 0.849	−0.086, 0.941, 0.945
Nonparam (CB)	0.016, 0.157, 0.158	0.008, 0.159, 0.159	0.024, 0.181, 0.183	0.027, 0.171, 0.173
Calibration ($K_2 = 3$)	0.016, 0.164, 0.165	0.018, 0.164, 0.165	0.012, 0.181, 0.181	0.005, 0.175, 0.175
Calibration ($K_2 = 4$)	0.015, 0.164, 0.165	0.013, 0.154, 0.154	0.013, 0.179, 0.180	0.024, 0.178, 0.179
Calibration (CB)	0.023, 0.165, 0.166	0.016, 0.155, 0.155	0.023, 0.182, 0.184	0.018, 0.181, 0.182

$\alpha_0 = 6.000$ (Kendall's $\tau = 0.75$)				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Listwise deletion	−0.055, 0.468, 0.471	0.214, 0.476, 0.522	−0.046, 0.539, 0.541	0.453, 0.594, 0.747
Param (correct)	−0.174, 0.423, 0.457	−0.092, 0.430, 0.439	−0.283, 0.464, 0.544	−0.152, 0.454, 0.479
Param (misspec)	−3.629, 0.245, 3.637	−3.353, 0.235, 3.362	−1.570, 0.456, 1.635	−1.273, 0.420, 1.340
Nonparam ($K_2 = 3$)	−0.150, 0.454, 0.478	−0.100, 0.431, 0.442	−0.239, 0.482, 0.538	−0.152, 0.434, 0.460
Nonparam ($K_2 = 4$)	−0.663, 2.380, 2.471	−0.690, 2.374, 2.472	−0.801, 1.945, 2.104	−0.723, 2.358, 2.467
Nonparam (CB)	−0.114, 0.437, 0.451	−0.066, 0.435, 0.440	−0.185, 0.481, 0.515	−0.125, 0.452, 0.469
Calibration ($K_2 = 3$)	−0.120, 0.430, 0.446	−0.098, 0.417, 0.428	−0.245, 0.456, 0.517	−0.161, 0.450, 0.477
Calibration ($K_2 = 4$)	−0.094, 0.431, 0.441	−0.104, 0.412, 0.425	−0.175, 0.460, 0.492	−0.165, 0.456, 0.485
Calibration (CB)	−0.115, 0.431, 0.446	−0.096, 0.419, 0.430	−0.213, 0.462, 0.509	−0.142, 0.472, 0.493

This table reports bias, standard deviation, and root mean squared error (RMSE) of each estimator with respect to $\alpha_0 \in \{1.636, 6.000\}$ after $J = 1000$ Monte Carlo trials. Cases A and B imply MCAR and MAR, respectively, with relatively low missing probability (i.e., $\mathrm{E}[T_{2i}] = 0.8$). Cases C and D imply MCAR and MAR, respectively, with relatively high missing probability (i.e., $\mathrm{E}[T_{2i}] = 0.6$). "Param (correct)" signifies the parametric estimator based on a correctly specified propensity score model. "Param (misspec)" signifies the parametric estimator based on a misspecified propensity score model. For the nonparametric estimator of Hirano, Imbens, and Ridder (2003) and the proposed calibration estimator, approximation sieves are constructed from the power series of X_i. The dimension of the approximation sieve is either fixed at $K_2 \in \{3, 4\}$ or automatically selected from $K_2 \in \{1, \ldots, 5\}$ based on the covariate balancing (CB) principle.

Table 3: Benchmark simulation results on Clayton copula ($N = 1000$)

$\alpha_0 = 1.636$ (Kendall's $\tau = 0.45$)				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Listwise deletion	0.019, 0.120, 0.122	0.057, 0.121, 0.134	0.010, 0.139, 0.139	0.129, 0.147, 0.196
Param (correct)	0.007, 0.117, 0.117	0.003, 0.111, 0.111	0.006, 0.129, 0.129	0.010, 0.129, 0.130
Param (misspec)	−0.416, 0.083, 0.424	−0.373, 0.085, 0.383	−0.118, 0.119, 0.168	−0.095, 0.118, 0.152
Nonparam ($K_2 = 3$)	0.006, 0.123, 0.123	0.013, 0.114, 0.115	0.002, 0.127, 0.127	0.014, 0.128, 0.128
Nonparam ($K_2 = 4$)	−0.211, 0.650, 0.683	−0.179, 0.832, 0.851	−0.217, 0.644, 0.679	−0.169, 0.786, 0.804
Nonparam (CB)	0.006, 0.117, 0.118	0.005, 0.116, 0.116	0.005, 0.127, 0.127	0.012, 0.125, 0.126
Calibration ($K_2 = 3$)	0.006, 0.112, 0.112	0.007, 0.116, 0.116	0.011, 0.130, 0.131	0.010, 0.126, 0.126
Calibration ($K_2 = 4$)	0.008, 0.113, 0.113	0.013, 0.119, 0.120	0.008, 0.128, 0.128	0.008, 0.120, 0.120
Calibration (CB)	0.010, 0.115, 0.115	0.003, 0.115, 0.115	0.008, 0.126, 0.126	−0.003, 0.122, 0.122

$\alpha_0 = 6.000$ (Kendall's $\tau = 0.75$)				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Listwise deletion	−0.036, 0.327, 0.329	0.218, 0.324, 0.390	−0.052, 0.379, 0.382	0.472, 0.408, 0.624
Param (correct)	−0.106, 0.297, 0.316	−0.053, 0.303, 0.307	−0.172, 0.337, 0.378	−0.089, 0.322, 0.334
Param (misspec)	−3.608, 0.177, 3.612	−3.320, 0.175, 3.325	−1.470, 0.338, 1.509	−1.194, 0.330, 1.239
Nonparam ($K_2 = 3$)	−0.102, 0.338, 0.353	−0.053, 0.312, 0.316	−0.151, 0.364, 0.394	−0.057, 0.322, 0.327
Nonparam ($K_2 = 4$)	−1.011, 2.249, 2.466	−0.846, 2.228, 2.383	−0.921, 2.153, 2.342	−0.972, 2.470, 2.654
Nonparam (CB)	−0.069, 0.299, 0.307	−0.036, 0.308, 0.310	−0.116, 0.347, 0.366	−0.062, 0.333, 0.338
Calibration ($K_2 = 3$)	−0.095, 0.309, 0.324	−0.056, 0.299, 0.304	−0.145, 0.307, 0.339	−0.110, 0.322, 0.340
Calibration ($K_2 = 4$)	−0.070, 0.304, 0.312	−0.051, 0.302, 0.306	−0.130, 0.324, 0.349	−0.092, 0.327, 0.340
Calibration (CB)	−0.088, 0.313, 0.325	−0.045, 0.306, 0.309	−0.122, 0.319, 0.342	−0.076, 0.315, 0.324

This table reports bias, standard deviation, and root mean squared error (RMSE) of each estimator with respect to $\alpha_0 \in \{1.636, 6.000\}$ after $J = 1000$ Monte Carlo trials. Cases A and B imply MCAR and MAR, respectively, with relatively low missing probability (i.e., $\mathrm{E}[T_{2i}] = 0.8$). Cases C and D imply MCAR and MAR, respectively, with relatively high missing probability (i.e., $\mathrm{E}[T_{2i}] = 0.6$). "Param (correct)" signifies the parametric estimator based on a correctly specified propensity score model. "Param (misspec)" signifies the parametric estimator based on a misspecified propensity score model. For the nonparametric estimator of Hirano, Imbens, and Ridder (2003) and the proposed calibration estimator, approximation sieves are constructed from the power series of X_i. The dimension of the approximation sieve is either fixed at $K_2 \in \{3, 4\}$ or automatically selected from $K_2 \in \{1, \ldots, 5\}$ based on the covariate balancing (CB) principle.

Table 4: Benchmark simulation results on Gumbel copula ($N = 250$)

$\gamma_0 = 1.818$ (Kendall's $\tau = 0.45$)				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Listwise deletion	0.027, 0.129, 0.132	−0.030, 0.127, 0.130	0.029, 0.154, 0.157	−0.068, 0.138, 0.153
Param (correct)	−0.015, 0.126, 0.127	−0.012, 0.118, 0.118	−0.017, 0.127, 0.129	−0.025, 0.140, 0.142
Param (misspec)	−0.779, 0.010, 0.779	−0.775, 0.012, 0.775	−0.646, 0.076, 0.650	−0.638, 0.087, 0.643
Nonparam ($K_2 = 3$)	−0.010, 0.124, 0.124	−0.011, 0.123, 0.123	−0.020, 0.135, 0.137	−0.034, 0.129, 0.133
Nonparam ($K_2 = 4$)	−0.101, 0.261, 0.280	−0.135, 0.284, 0.315	−0.096, 0.262, 0.279	−0.125, 0.269, 0.296
Nonparam (CB)	−0.006, 0.122, 0.123	−0.014, 0.122, 0.122	−0.012, 0.137, 0.138	−0.037, 0.136, 0.141
Calibration ($K_2 = 3$)	0.012, 0.123, 0.124	0.020, 0.125, 0.126	0.018, 0.138, 0.139	0.014, 0.145, 0.145
Calibration ($K_2 = 4$)	0.016, 0.124, 0.125	0.017, 0.129, 0.131	0.018, 0.132, 0.133	0.002, 0.137, 0.137
Calibration (CB)	0.018, 0.125, 0.126	0.006, 0.121, 0.122	0.015, 0.136, 0.137	0.008, 0.145, 0.145

$\gamma_0 = 4.000$ (Kendall's $\tau = 0.75$)				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Listwise deletion	0.013, 0.318, 0.318	−0.160, 0.308, 0.347	0.033, 0.374, 0.375	−0.243, 0.348, 0.424
Param (correct)	−0.151, 0.285, 0.323	−0.185, 0.278, 0.334	−0.231, 0.307, 0.384	−0.313, 0.323, 0.450
Param (misspec)	−2.926, 0.015, 2.926	−2.921, 0.016, 2.921	−2.593, 0.249, 2.605	−2.610, 0.204, 2.618
Nonparam ($K_2 = 3$)	−0.169, 0.292, 0.337	−0.205, 0.294, 0.359	−0.211, 0.293, 0.361	−0.318, 0.315, 0.448
Nonparam ($K_2 = 4$)	−0.525, 0.961, 1.095	−0.572, 0.953, 1.111	−0.587, 0.960, 1.125	−0.645, 0.923, 1.126
Nonparam (CB)	−0.150, 0.289, 0.325	−0.195, 0.288, 0.348	−0.198, 0.311, 0.369	−0.323, 0.308, 0.446
Calibration ($K_2 = 3$)	−0.020, 0.300, 0.301	−0.055, 0.294, 0.299	−0.075, 0.323, 0.332	−0.140, 0.324, 0.353
Calibration ($K_2 = 4$)	−0.022, 0.291, 0.292	−0.059, 0.308, 0.313	−0.059, 0.336, 0.341	−0.129, 0.342, 0.366
Calibration (CB)	−0.024, 0.316, 0.317	−0.062, 0.313, 0.319	−0.062, 0.347, 0.352	−0.142, 0.333, 0.362

This table reports bias, standard deviation, and root mean squared error (RMSE) of each estimator with respect to $\gamma_0 \in \{1.818, 4.000\}$ after $J = 1000$ Monte Carlo trials. Cases A and B imply MCAR and MAR, respectively, with relatively low missing probability (i.e., $\mathrm{E}[T_{2i}] = 0.8$). Cases C and D imply MCAR and MAR, respectively, with relatively high missing probability (i.e., $\mathrm{E}[T_{2i}] = 0.6$). "Param (correct)" signifies the parametric estimator based on a correctly specified propensity score model. "Param (misspec)" signifies the parametric estimator based on a misspecified propensity score model. For the nonparametric estimator of Hirano, Imbens, and Ridder (2003) and the proposed calibration estimator, approximation sieves are constructed from the power series of X_i. The dimension of the approximation sieve is either fixed at $K_2 \in \{3, 4\}$ or automatically selected from $K_2 \in \{1, \ldots, 5\}$ based on the covariate balancing (CB) principle.

Table 5: Benchmark simulation results on Gumbel copula ($N = 500$)

$\gamma_0 = 1.818$ (Kendall's $\tau = 0.45$)				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Listwise deletion	0.018, 0.091, 0.092	−0.039, 0.087, 0.096	0.022, 0.099, 0.102	−0.083, 0.098, 0.128
Param (correct)	−0.003, 0.086, 0.086	−0.005, 0.087, 0.087	−0.006, 0.098, 0.099	−0.007, 0.096, 0.097
Param (misspec)	−0.778, 0.007, 0.778	−0.774, 0.008, 0.774	−0.650, 0.051, 0.652	−0.647, 0.052, 0.649
Nonparam ($K_2 = 3$)	−0.004, 0.091, 0.091	−0.008, 0.089, 0.089	−0.004, 0.100, 0.100	−0.015, 0.098, 0.099
Nonparam ($K_2 = 4$)	−0.165, 0.315, 0.356	−0.161, 0.312, 0.350	−0.141, 0.296, 0.328	−0.155, 0.306, 0.343
Nonparam (CB)	−0.004, 0.087, 0.087	−0.004, 0.087, 0.087	−0.006, 0.093, 0.094	−0.012, 0.099, 0.100
Calibration ($K_2 = 3$)	0.004, 0.092, 0.092	0.006, 0.089, 0.089	0.010, 0.097, 0.097	−0.001, 0.101, 0.101
Calibration ($K_2 = 4$)	0.010, 0.090, 0.090	0.009, 0.087, 0.087	0.004, 0.095, 0.095	−0.003, 0.094, 0.095
Calibration (CB)	0.011, 0.086, 0.087	0.009, 0.092, 0.092	0.006, 0.097, 0.097	−0.003, 0.095, 0.095

$\gamma_0 = 4.000$ (Kendall's $\tau = 0.75$)				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Listwise deletion	−0.009, 0.227, 0.227	−0.149, 0.223, 0.268	0.007, 0.266, 0.266	−0.256, 0.244, 0.353
Param (correct)	−0.066, 0.211, 0.221	−0.096, 0.213, 0.233	−0.113, 0.232, 0.258	−0.153, 0.223, 0.271
Param (misspec)	−2.925, 0.010, 2.925	−2.918, 0.011, 2.918	−2.618, 0.137, 2.622	−2.620, 0.124, 2.623
Nonparam ($K_2 = 3$)	−0.075, 0.203, 0.216	−0.093, 0.242, 0.259	−0.113, 0.253, 0.277	−0.165, 0.233, 0.285
Nonparam ($K_2 = 4$)	−0.614, 1.101, 1.261	−0.727, 1.159, 1.368	−0.631, 1.104, 1.271	−0.682, 1.081, 1.278
Nonparam (CB)	−0.061, 0.207, 0.216	−0.079, 0.204, 0.218	−0.080, 0.227, 0.240	−0.140, 0.224, 0.264
Calibration ($K_2 = 3$)	−0.022, 0.216, 0.217	−0.025, 0.210, 0.211	−0.042, 0.232, 0.236	−0.093, 0.234, 0.252
Calibration ($K_2 = 4$)	−0.009, 0.210, 0.210	−0.033, 0.212, 0.214	−0.037, 0.234, 0.237	−0.094, 0.243, 0.260
Calibration (CB)	−0.008, 0.212, 0.212	−0.039, 0.213, 0.216	−0.045, 0.240, 0.244	−0.100, 0.240, 0.260

This table reports bias, standard deviation, and root mean squared error (RMSE) of each estimator with respect to $\gamma_0 \in \{1.818, 4.000\}$ after $J = 1000$ Monte Carlo trials. Cases A and B imply MCAR and MAR, respectively, with relatively low missing probability (i.e., $\mathrm{E}[T_{2i}] = 0.8$). Cases C and D imply MCAR and MAR, respectively, with relatively high missing probability (i.e., $\mathrm{E}[T_{2i}] = 0.6$). "Param (correct)" signifies the parametric estimator based on a correctly specified propensity score model. "Param (misspec)" signifies the parametric estimator based on a misspecified propensity score model. For the nonparametric estimator of Hirano, Imbens, and Ridder (2003) and the proposed calibration estimator, approximation sieves are constructed from the power series of X_i. The dimension of the approximation sieve is either fixed at $K_2 \in \{3, 4\}$ or automatically selected from $K_2 \in \{1, \dots, 5\}$ based on the covariate balancing (CB) principle.

Table 6: Benchmark simulation results on Gumbel copula ($N = 1000$)

$\gamma_0 = 1.818$ (Kendall's $\tau = 0.45$)				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Listwise deletion	0.005, 0.062, 0.062	−0.053, 0.060, 0.080	0.006, 0.075, 0.075	−0.087, 0.071, 0.113
Param (correct)	0.003, 0.060, 0.060	0.000, 0.065, 0.065	0.001, 0.069, 0.069	−0.004, 0.066, 0.066
Param (misspec)	−0.777, 0.005, 0.777	−0.773, 0.006, 0.773	−0.649, 0.034, 0.650	−0.648, 0.035, 0.649
Nonparam ($K_2 = 3$)	−0.002, 0.067, 0.067	0.000, 0.063, 0.063	−0.001, 0.073, 0.073	−0.007, 0.070, 0.070
Nonparam ($K_2 = 4$)	−0.197, 0.344, 0.396	−0.192, 0.371, 0.418	−0.178, 0.405, 0.442	−0.146, 0.463, 0.486
Nonparam (CB)	0.000, 0.059, 0.059	0.001, 0.060, 0.060	−0.003, 0.065, 0.065	−0.005, 0.068, 0.068
Calibration ($K_2 = 3$)	0.005, 0.062, 0.063	0.002, 0.062, 0.062	0.001, 0.067, 0.067	−0.000, 0.068, 0.068
Calibration ($K_2 = 4$)	0.003, 0.061, 0.061	0.008, 0.061, 0.061	0.008, 0.068, 0.069	−0.005, 0.070, 0.070
Calibration (CB)	0.005, 0.059, 0.060	0.004, 0.062, 0.063	0.001, 0.066, 0.066	0.001, 0.069, 0.069

$\gamma_0 = 4.000$ (Kendall's $\tau = 0.75$)				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Listwise deletion	0.001, 0.158, 0.158	−0.152, 0.157, 0.219	−0.007, 0.186, 0.186	−0.274, 0.173, 0.324
Param (correct)	−0.023, 0.162, 0.163	−0.038, 0.150, 0.155	−0.047, 0.168, 0.174	−0.081, 0.158, 0.177
Param (misspec)	−2.924, 0.007, 2.924	−2.917, 0.008, 2.917	−2.621, 0.088, 2.622	−2.623, 0.082, 2.625
Nonparam ($K_2 = 3$)	−0.051, 0.251, 0.256	−0.052, 0.180, 0.188	−0.046, 0.178, 0.184	−0.074, 0.172, 0.188
Nonparam ($K_2 = 4$)	−0.783, 1.296, 1.514	−0.889, 1.374, 1.636	−0.797, 1.239, 1.474	−0.785, 1.233, 1.462
Nonparam (CB)	−0.020, 0.148, 0.149	−0.032, 0.148, 0.151	−0.041, 0.162, 0.167	−0.073, 0.164, 0.179
Calibration ($K_2 = 3$)	−0.015, 0.150, 0.151	−0.020, 0.151, 0.153	−0.026, 0.158, 0.160	−0.058, 0.168, 0.177
Calibration ($K_2 = 4$)	−0.006, 0.154, 0.154	−0.022, 0.147, 0.149	−0.029, 0.167, 0.170	−0.046, 0.163, 0.169
Calibration (CB)	−0.005, 0.155, 0.155	−0.016, 0.156, 0.157	−0.018, 0.161, 0.162	−0.064, 0.172, 0.183

This table reports bias, standard deviation, and root mean squared error (RMSE) of each estimator with respect to $\gamma_0 \in \{1.818, 4.000\}$ after $J = 1000$ Monte Carlo trials. Cases A and B imply MCAR and MAR, respectively, with relatively low missing probability (i.e., $\mathrm{E}[T_{2i}] = 0.8$). Cases C and D imply MCAR and MAR, respectively, with relatively high missing probability (i.e., $\mathrm{E}[T_{2i}] = 0.6$). "Param (correct)" signifies the parametric estimator based on a correctly specified propensity score model. "Param (misspec)" signifies the parametric estimator based on a misspecified propensity score model. For the nonparametric estimator of Hirano, Imbens, and Ridder (2003) and the proposed calibration estimator, approximation sieves are constructed from the power series of X_i. The dimension of the approximation sieve is either fixed at $K_2 \in \{3, 4\}$ or automatically selected from $K_2 \in \{1, \ldots, 5\}$ based on the covariate balancing (CB) principle.

Table 7: Benchmark simulation results on $C_3(1.636)$ with B-splines (Kendall's $\tau = 0.45$)

$N = 250$				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Nonparam ($R = 2$)	0.030, 0.233, 0.234	0.036, 0.235, 0.238	0.032, 0.258, 0.260	0.030, 0.254, 0.256
Nonparam ($R = 3$)	0.024, 0.237, 0.238	0.032, 0.232, 0.234	0.037, 0.262, 0.264	0.015, 0.242, 0.243
Nonparam ($R = 4$)	0.035, 0.235, 0.237	0.023, 0.233, 0.234	0.019, 0.265, 0.266	0.041, 0.252, 0.255
Calibration ($R = 2$)	0.021, 0.229, 0.230	0.030, 0.229, 0.231	0.018, 0.245, 0.245	0.040, 0.252, 0.255
Calibration ($R = 3$)	0.021, 0.227, 0.227	0.037, 0.237, 0.240	0.023, 0.258, 0.259	0.009, 0.253, 0.253
Calibration ($R = 4$)	0.018, 0.229, 0.230	0.025, 0.242, 0.243	0.006, 0.264, 0.264	−0.005, 0.254, 0.254

$N = 500$				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Nonparam ($R = 2$)	0.012, 0.169, 0.169	0.012, 0.165, 0.165	0.006, 0.176, 0.176	0.017, 0.171, 0.172
Nonparam ($R = 3$)	0.010, 0.161, 0.161	0.023, 0.160, 0.161	0.005, 0.185, 0.185	0.015, 0.169, 0.169
Nonparam ($R = 4$)	0.012, 0.164, 0.165	0.010, 0.157, 0.158	0.014, 0.176, 0.177	0.017, 0.181, 0.182
Calibration ($R = 2$)	0.019, 0.167, 0.168	0.017, 0.162, 0.163	0.015, 0.180, 0.181	0.013, 0.172, 0.172
Calibration ($R = 3$)	0.007, 0.169, 0.169	0.016, 0.160, 0.161	0.011, 0.183, 0.184	−0.002, 0.181, 0.181
Calibration ($R = 4$)	0.003, 0.164, 0.164	−0.005, 0.161, 0.161	−0.020, 0.177, 0.178	−0.020, 0.184, 0.185

$N = 1000$				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Nonparam ($R = 2$)	0.001, 0.115, 0.115	0.009, 0.113, 0.113	0.004, 0.125, 0.125	0.005, 0.118, 0.119
Nonparam ($R = 3$)	0.009, 0.120, 0.121	0.013, 0.114, 0.115	0.005, 0.127, 0.127	0.008, 0.121, 0.121
Nonparam ($R = 4$)	0.004, 0.116, 0.116	0.007, 0.117, 0.118	0.010, 0.131, 0.132	0.008, 0.125, 0.125
Calibration ($R = 2$)	0.009, 0.119, 0.119	0.013, 0.117, 0.118	0.007, 0.126, 0.127	0.016, 0.123, 0.124
Calibration ($R = 3$)	0.003, 0.118, 0.118	0.010, 0.114, 0.115	−0.004, 0.129, 0.129	−0.010, 0.131, 0.131
Calibration ($R = 4$)	−0.002, 0.122, 0.122	−0.009, 0.114, 0.114	−0.031, 0.135, 0.138	−0.031, 0.130, 0.133

This table reports bias, standard deviation, and RMSE of the nonparametric estimator of Hirano, Imbens, and Ridder (2003) and the proposed calibration estimator with respect to the Clayton copula parameter $\alpha_0 = 1.636$ after $J = 1000$ Monte Carlo trials. Cases A and B imply MCAR and MAR, respectively, with $\mathrm{E}[T_{2i}] = 0.8$. Cases C and D imply MCAR and MAR, respectively, with $\mathrm{E}[T_{2i}] = 0.6$. Linear, quadratic, and cubic B-spline base functions are used when $R \in \{2, 3, 4\}$, respectively. The minimum of covariate X is used as left outside knots; the $\{25\%, 50\%, 75\%\}$ quantiles of X are used as inside knots; the maximum of X is used as right outside knots.

Table 8: Benchmark simulation results on $C_3(6.000)$ with B-splines (Kendall's $\tau = 0.75$)

$N = 250$				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Nonparam ($R = 2$)	−0.213, 0.622, 0.658	−0.122, 0.607, 0.619	−0.325, 0.605, 0.687	−0.186, 0.627, 0.654
Nonparam ($R = 3$)	−0.158, 0.651, 0.670	−0.124, 0.609, 0.622	−0.286, 0.649, 0.709	−0.181, 0.630, 0.656
Nonparam ($R = 4$)	−0.119, 0.605, 0.617	−0.124, 0.597, 0.610	−0.229, 0.646, 0.686	−0.192, 0.612, 0.642
Calibration ($R = 2$)	−0.164, 0.613, 0.635	−0.126, 0.610, 0.623	−0.299, 0.632, 0.700	−0.191, 0.612, 0.641
Calibration ($R = 3$)	−0.187, 0.580, 0.609	−0.139, 0.598, 0.614	−0.370, 0.714, 0.804	−0.328, 0.709, 0.782
Calibration ($R = 4$)	−0.222, 0.620, 0.659	−0.267, 0.649, 0.702	−0.520, 0.708, 0.878	−0.568, 0.768, 0.955

$N = 500$				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Nonparam ($R = 2$)	−0.099, 0.418, 0.429	−0.088, 0.421, 0.430	−0.244, 0.451, 0.513	−0.113, 0.459, 0.473
Nonparam ($R = 3$)	−0.102, 0.424, 0.436	−0.068, 0.431, 0.436	−0.198, 0.455, 0.496	−0.098, 0.465, 0.475
Nonparam ($R = 4$)	−0.098, 0.443, 0.454	−0.072, 0.437, 0.443	−0.165, 0.455, 0.484	−0.132, 0.460, 0.479
Calibration ($R = 2$)	−0.116, 0.425, 0.441	−0.125, 0.426, 0.444	−0.196, 0.466, 0.505	−0.118, 0.458, 0.473
Calibration ($R = 3$)	−0.104, 0.449, 0.461	−0.086, 0.447, 0.455	−0.273, 0.531, 0.597	−0.285, 0.565, 0.633
Calibration ($R = 4$)	−0.183, 0.454, 0.489	−0.196, 0.468, 0.507	−0.406, 0.550, 0.684	−0.429, 0.567, 0.711

$N = 1000$				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Nonparam ($R = 2$)	−0.083, 0.314, 0.325	−0.062, 0.292, 0.298	−0.136, 0.323, 0.351	−0.069, 0.337, 0.344
Nonparam ($R = 3$)	−0.067, 0.298, 0.306	−0.034, 0.297, 0.299	−0.105, 0.330, 0.346	−0.073, 0.316, 0.324
Nonparam ($R = 4$)	−0.057, 0.304, 0.309	−0.039, 0.318, 0.321	−0.092, 0.338, 0.350	−0.045, 0.322, 0.325
Calibration ($R = 2$)	−0.077, 0.306, 0.316	−0.053, 0.295, 0.300	−0.131, 0.338, 0.363	−0.084, 0.326, 0.337
Calibration ($R = 3$)	−0.084, 0.304, 0.316	−0.074, 0.298, 0.307	−0.218, 0.391, 0.447	−0.226, 0.466, 0.518
Calibration ($R = 4$)	−0.126, 0.307, 0.332	−0.158, 0.322, 0.359	−0.351, 0.421, 0.548	−0.343, 0.432, 0.552

This table reports bias, standard deviation, and RMSE of the nonparametric estimator of Hirano, Imbens, and Ridder (2003) and the proposed calibration estimator with respect to the Clayton copula parameter $\alpha_0 = 6.000$ after $J = 1000$ Monte Carlo trials. Cases A and B imply MCAR and MAR, respectively, with $\mathrm{E}[T_{2i}] = 0.8$. Cases C and D imply MCAR and MAR, respectively, with $\mathrm{E}[T_{2i}] = 0.6$. Linear, quadratic, and cubic B-spline base functions are used when $R \in \{2, 3, 4\}$, respectively. The minimum of covariate X is used as left outside knots; the $\{25\%, 50\%, 75\%\}$ quantiles of X are used as inside knots; the maximum of X is used as right outside knots.

Table 9: Benchmark simulation results on $G_3(1.818)$ with B-splines (Kendall's $\tau = 0.45$)

$N = 250$				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Nonparam ($R = 2$)	−0.014, 0.123, 0.124	−0.009, 0.124, 0.125	−0.025, 0.133, 0.135	−0.041, 0.131, 0.137
Nonparam ($R = 3$)	−0.017, 0.121, 0.122	−0.023, 0.120, 0.123	−0.011, 0.136, 0.136	−0.030, 0.136, 0.140
Nonparam ($R = 4$)	−0.005, 0.121, 0.121	−0.016, 0.126, 0.127	−0.014, 0.139, 0.140	−0.030, 0.136, 0.140
Calibration ($R = 2$)	0.019, 0.123, 0.124	0.017, 0.122, 0.123	0.013, 0.137, 0.137	0.007, 0.143, 0.143
Calibration ($R = 3$)	0.017, 0.125, 0.126	0.010, 0.121, 0.121	0.024, 0.140, 0.142	0.010, 0.145, 0.146
Calibration ($R = 4$)	0.021, 0.128, 0.129	0.021, 0.127, 0.129	0.018, 0.142, 0.144	0.005, 0.151, 0.151

$N = 500$				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Nonparam ($R = 2$)	−0.003, 0.083, 0.083	−0.010, 0.085, 0.086	−0.010, 0.094, 0.095	−0.009, 0.099, 0.099
Nonparam ($R = 3$)	−0.003, 0.084, 0.084	−0.002, 0.083, 0.083	−0.007, 0.097, 0.097	−0.007, 0.101, 0.101
Nonparam ($R = 4$)	−0.005, 0.086, 0.087	−0.003, 0.086, 0.086	−0.002, 0.098, 0.098	−0.010, 0.094, 0.095
Calibration ($R = 2$)	0.012, 0.088, 0.089	0.008, 0.087, 0.087	0.006, 0.098, 0.098	0.002, 0.097, 0.097
Calibration ($R = 3$)	0.010, 0.086, 0.087	0.009, 0.089, 0.089	0.014, 0.097, 0.098	−0.001, 0.094, 0.094
Calibration ($R = 4$)	0.015, 0.087, 0.088	0.007, 0.087, 0.087	0.017, 0.101, 0.102	0.011, 0.103, 0.104

$N = 1000$				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Nonparam ($R = 2$)	−0.001, 0.063, 0.063	−0.003, 0.061, 0.061	−0.003, 0.071, 0.071	−0.006, 0.068, 0.068
Nonparam ($R = 3$)	0.002, 0.061, 0.061	−0.001, 0.066, 0.066	−0.002, 0.068, 0.068	−0.010, 0.068, 0.069
Nonparam ($R = 4$)	0.001, 0.060, 0.060	−0.000, 0.069, 0.069	0.002, 0.070, 0.070	−0.005, 0.068, 0.068
Calibration ($R = 2$)	0.007, 0.060, 0.061	0.003, 0.062, 0.062	0.003, 0.068, 0.068	0.001, 0.070, 0.070
Calibration ($R = 3$)	0.010, 0.064, 0.064	0.004, 0.063, 0.063	0.011, 0.070, 0.071	0.002, 0.069, 0.069
Calibration ($R = 4$)	0.009, 0.061, 0.062	0.014, 0.060, 0.061	0.015, 0.067, 0.069	0.013, 0.070, 0.071

This table reports bias, standard deviation, and RMSE of the nonparametric estimator of Hirano, Imbens, and Ridder (2003) and the proposed calibration estimator with respect to the Gumbel copula parameter $\gamma_0 = 1.818$ after $J = 1000$ Monte Carlo trials. Cases A and B imply MCAR and MAR, respectively, with $\mathrm{E}[T_{2i}] = 0.8$. Cases C and D imply MCAR and MAR, respectively, with $\mathrm{E}[T_{2i}] = 0.6$. Linear, quadratic, and cubic B-spline base functions are used when $R \in \{2, 3, 4\}$, respectively. The minimum of covariate X is used as left outside knots; the $\{25\%, 50\%, 75\%\}$ quantiles of X are used as inside knots; the maximum of X is used as right outside knots.

Table 10: Benchmark simulation results on $G_3(4.000)$ with B-splines (Kendall's $\tau = 0.75$)

$N = 250$				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Nonparam ($R = 2$)	−0.152, 0.296, 0.332	−0.203, 0.275, 0.342	−0.201, 0.315, 0.373	−0.316, 0.299, 0.435
Nonparam ($R = 3$)	−0.171, 0.275, 0.323	−0.211, 0.306, 0.372	−0.191, 0.323, 0.375	−0.270, 0.333, 0.429
Nonparam ($R = 4$)	−0.133, 0.301, 0.329	−0.188, 0.319, 0.370	−0.148, 0.331, 0.363	−0.275, 0.327, 0.427
Calibration ($R = 2$)	−0.033, 0.306, 0.308	−0.056, 0.297, 0.302	−0.038, 0.344, 0.346	−0.141, 0.336, 0.365
Calibration ($R = 3$)	−0.032, 0.303, 0.305	−0.056, 0.308, 0.313	−0.085, 0.346, 0.356	−0.178, 0.388, 0.427
Calibration ($R = 4$)	−0.063, 0.316, 0.322	−0.099, 0.337, 0.351	−0.177, 0.385, 0.424	−0.281, 0.459, 0.538

$N = 500$				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Nonparam ($R = 2$)	−0.066, 0.212, 0.222	−0.087, 0.217, 0.234	−0.094, 0.227, 0.246	−0.148, 0.228, 0.272
Nonparam ($R = 3$)	−0.061, 0.213, 0.222	−0.076, 0.210, 0.223	−0.085, 0.229, 0.244	−0.140, 0.228, 0.268
Nonparam ($R = 4$)	−0.067, 0.219, 0.229	−0.098, 0.240, 0.259	−0.089, 0.230, 0.246	−0.139, 0.230, 0.269
Calibration ($R = 2$)	−0.020, 0.213, 0.213	−0.039, 0.214, 0.218	−0.034, 0.234, 0.236	−0.087, 0.230, 0.246
Calibration ($R = 3$)	−0.020, 0.209, 0.210	−0.033, 0.210, 0.212	−0.050, 0.251, 0.256	−0.136, 0.268, 0.300
Calibration ($R = 4$)	−0.049, 0.227, 0.232	−0.067, 0.228, 0.237	−0.120, 0.257, 0.284	−0.175, 0.287, 0.336

$N = 1000$				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Nonparam ($R = 2$)	−0.029, 0.148, 0.150	−0.036, 0.152, 0.156	−0.033, 0.168, 0.171	−0.076, 0.169, 0.185
Nonparam ($R = 3$)	−0.024, 0.155, 0.156	−0.038, 0.158, 0.163	−0.038, 0.162, 0.167	−0.061, 0.164, 0.175
Nonparam ($R = 4$)	−0.027, 0.153, 0.155	−0.035, 0.153, 0.157	−0.041, 0.162, 0.168	−0.063, 0.160, 0.172
Calibration ($R = 2$)	−0.004, 0.148, 0.148	−0.016, 0.154, 0.154	−0.025, 0.166, 0.168	−0.053, 0.168, 0.176
Calibration ($R = 3$)	−0.009, 0.150, 0.150	−0.021, 0.158, 0.159	−0.040, 0.181, 0.185	−0.108, 0.226, 0.251
Calibration ($R = 4$)	−0.025, 0.154, 0.156	−0.052, 0.165, 0.173	−0.094, 0.184, 0.207	−0.103, 0.200, 0.224

This table reports bias, standard deviation, and RMSE of the nonparametric estimator of Hirano, Imbens, and Ridder (2003) and the proposed calibration estimator with respect to the Gumbel copula parameter $\gamma_0 = 4.000$ after $J = 1000$ Monte Carlo trials. Cases A and B imply MCAR and MAR, respectively, with $\mathrm{E}[T_{2i}] = 0.8$. Cases C and D imply MCAR and MAR, respectively, with $\mathrm{E}[T_{2i}] = 0.6$. Linear, quadratic, and cubic B-spline base functions are used when $R \in \{2, 3, 4\}$, respectively. The minimum of covariate X is used as left outside knots; the $\{25\%, 50\%, 75\%\}$ quantiles of X are used as inside knots; the maximum of X is used as right outside knots.

Table 11: Simulation results on $C_3(1.636)$ with misspecified missing mechanism

	$N = 250$	$N = 500$	$N = 1000$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Param (correct)	0.040, 0.268, 0.271	0.022, 0.179, 0.180	0.005, 0.125, 0.125
Nonparam ($K_2 = 2$)	0.043, 0.255, 0.258	0.024, 0.177, 0.178	0.028, 0.126, 0.129
Nonparam ($K_2 = 3$)	0.046, 0.252, 0.256	0.025, 0.178, 0.179	0.016, 0.127, 0.128
Nonparam ($K_2 = 4$)	−0.046, 0.975, 0.976	−0.129, 0.583, 0.597	−0.191, 0.775, 0.798
Nonparam (CB)	0.042, 0.244, 0.248	0.020, 0.176, 0.177	0.017, 0.128, 0.129
Calibration ($K_2 = 2$)	0.034, 0.248, 0.250	0.032, 0.178, 0.181	0.019, 0.126, 0.128
Calibration ($K_2 = 3$)	0.045, 0.248, 0.252	0.014, 0.183, 0.184	0.014, 0.127, 0.128
Calibration ($K_2 = 4$)	0.025, 0.252, 0.253	0.021, 0.171, 0.172	0.020, 0.124, 0.126
Calibration (CB)	0.023, 0.250, 0.251	0.013, 0.181, 0.182	0.020, 0.125, 0.126
Calibration-Y ($K_2 = 3$)	0.002, 0.247, 0.247	−0.004, 0.175, 0.175	−0.013, 0.122, 0.123
Calibration-Y ($K_2 = 6$)	0.029, 0.249, 0.251	0.016, 0.175, 0.176	0.004, 0.122, 0.122
Calibration-Y ($K_2 = 10$)	0.049, 0.255, 0.259	0.019, 0.178, 0.179	−0.004, 0.132, 0.132
Calibration-Y (CB)	0.038, 0.242, 0.244	0.020, 0.175, 0.176	0.007, 0.124, 0.124

In this table, the true copula is the trivariate Clayton copula with $\alpha_0 = 1.636$ (Kendall's $\tau = 0.45$). We report bias, standard deviation, and RMSE of each estimator with respect to α_0 after $J = 1000$ Monte Carlo trials. The missing mechanism of Y_{2i} is specified as $\pi_2(X_i, Y_{1i}) = [1 + \exp(-0.42 + 0.2X_i + 0.2Y_{1i})]^{-1}$. This implies MAR with $\mathrm{E}[T_{2i}] = 0.6$. "Param (correct)" is the parametric estimator based on a correctly specified propensity score model. "Nonparam" is the nonparametric estimator of Hirano, Imbens, and Ridder (2003), where the approximation sieve consists of power series of X_i only: $u_{K_2}(X_i) = (1, X_i, X_i^2, \dots, X_i^{K_2-1})^\top$. "Calibration" is the calibration estimator whose approximation sieve consists of power series of X_i only. "Calibration-Y" is the calibration estimator whose approximation sieve consists of power series of X_i and Y_{1i}: $u_{10}(X_i, Y_{1i}) = (1, X_i, Y_{1i}, X_i^2, Y_{1i}^2, X_iY_{1i}, X_i^3, Y_{1i}^3, X_i^2Y_{1i}, X_iY_{1i}^2)^\top$ for $K_2 = 10$. When $K_2 \le 10$, $u_{K_2}(X_i, Y_{1i})$ consists of the first K_2 elements of $u_{10}(X_i, Y_{1i})$. "CB" signifies that the data-driven K_2^* is chosen among $K_2 \in \{1, \dots, \bar{K}_2\}$ based on the covariate balancing principle. $\bar{K}_2 = 5$ for "Nonparam" and "Calibration", while $\bar{K}_2 = 10$ for "Calibration-Y".

Table 12: Simulation results on $C_3(6.000)$ with misspecified missing mechanism

	$N = 250$	$N = 500$	$N = 1000$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Param (correct)	−0.238, 0.666, 0.707	−0.162, 0.470, 0.497	−0.103, 0.320, 0.336
Nonparam ($K_2 = 2$)	−0.231, 0.637, 0.677	−0.121, 0.464, 0.479	−0.049, 0.339, 0.342
Nonparam ($K_2 = 3$)	−0.230, 0.622, 0.663	−0.134, 0.471, 0.490	−0.058, 0.338, 0.343
Nonparam ($K_2 = 4$)	−0.378, 2.650, 2.677	−0.686, 2.031, 2.144	−0.827, 2.395, 2.534
Nonparam (CB)	−0.181, 0.641, 0.666	−0.140, 0.454, 0.475	−0.059, 0.327, 0.332
Calibration ($K_2 = 2$)	−0.290, 0.619, 0.683	−0.195, 0.458, 0.498	−0.125, 0.339, 0.362
Calibration ($K_2 = 3$)	−0.231, 0.641, 0.682	−0.148, 0.452, 0.475	−0.080, 0.313, 0.322
Calibration ($K_2 = 4$)	−0.218, 0.637, 0.674	−0.123, 0.463, 0.479	−0.068, 0.322, 0.330
Calibration (CB)	−0.212, 0.628, 0.663	−0.147, 0.453, 0.476	−0.069, 0.323, 0.330
Calibration-Y ($K_2 = 3$)	−0.315, 0.643, 0.716	−0.225, 0.472, 0.523	−0.168, 0.327, 0.368
Calibration-Y ($K_2 = 6$)	−0.273, 0.642, 0.697	−0.179, 0.471, 0.504	−0.101, 0.314, 0.330
Calibration-Y ($K_2 = 10$)	−0.246, 0.634, 0.680	−0.163, 0.448, 0.477	−0.074, 0.328, 0.336
Calibration-Y (CB)	−0.232, 0.639, 0.680	−0.160, 0.460, 0.487	−0.078, 0.314, 0.324

In this table, the true copula is the trivariate Clayton copula with $\alpha_0 = 6.000$ (Kendall's $\tau = 0.75$). We report bias, standard deviation, and RMSE of each estimator with respect to α_0 after $J = 1000$ Monte Carlo trials. The missing mechanism of Y_{2i} is specified as $\pi_2(X_i, Y_{1i}) = [1 + \exp(-0.42 + 0.2X_i + 0.2Y_{1i})]^{-1}$. This implies MAR with $\mathrm{E}[T_{2i}] = 0.6$. "Param (correct)" is the parametric estimator based on a correctly specified propensity score model. "Nonparam" is the nonparametric estimator of Hirano, Imbens, and Ridder (2003), where the approximation sieve consists of power series of X_i only: $u_{K_2}(X_i) = (1, X_i, X_i^2, \ldots, X_i^{K_2-1})^\top$. "Calibration" is the calibration estimator whose approximation sieve consists of power series of X_i only. "Calibration-Y" is the calibration estimator whose approximation sieve consists of power series of X_i and Y_{1i}: $u_{10}(X_i, Y_{1i}) = (1, X_i, Y_{1i}, X_i^2, Y_{1i}^2, X_iY_{1i}, X_i^3, Y_{1i}^3, X_i^2Y_{1i}, X_iY_{1i}^2)^\top$ for $K_2 = 10$. When $K_2 \leq 10$, $u_{K_2}(X_i, Y_{1i})$ consists of the first K_2 elements of $u_{10}(X_i, Y_{1i})$. "CB" signifies that the data-driven K_2^* is chosen among $K_2 \in \{1, \ldots, \bar{K}_2\}$ based on the covariate balancing principle. $\bar{K}_2 = 5$ for "Nonparam" and "Calibration", while $\bar{K}_2 = 10$ for "Calibration-Y".

Table 13: Simulation results on $G_3(1.818)$ with misspecified missing mechanism

	$N = 250$	$N = 500$	$N = 1000$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Param (correct)	−0.024, 0.142, 0.144	−0.007, 0.096, 0.096	−0.002, 0.070, 0.070
Nonparam ($K_2 = 2$)	−0.029, 0.140, 0.143	−0.010, 0.098, 0.099	−0.003, 0.071, 0.071
Nonparam ($K_2 = 3$)	−0.027, 0.136, 0.139	−0.015, 0.101, 0.102	−0.006, 0.076, 0.077
Nonparam ($K_2 = 4$)	−0.130, 0.278, 0.307	−0.140, 0.293, 0.325	−0.175, 0.504, 0.533
Nonparam (CB)	−0.022, 0.131, 0.133	−0.006, 0.097, 0.097	0.000, 0.068, 0.068
Calibration ($K_2 = 2$)	−0.009, 0.135, 0.136	−0.004, 0.100, 0.100	−0.010, 0.070, 0.070
Calibration ($K_2 = 3$)	0.009, 0.142, 0.142	0.003, 0.097, 0.097	−0.001, 0.069, 0.069
Calibration ($K_2 = 4$)	0.002, 0.134, 0.134	0.003, 0.099, 0.099	0.000, 0.069, 0.069
Calibration (CB)	0.010, 0.137, 0.138	0.005, 0.094, 0.094	0.003, 0.070, 0.070
Calibration-Y ($K_2 = 3$)	0.002, 0.142, 0.142	−0.003, 0.100, 0.100	−0.007, 0.071, 0.071
Calibration-Y ($K_2 = 6$)	0.010, 0.134, 0.134	0.003, 0.096, 0.096	−0.001, 0.067, 0.067
Calibration-Y ($K_2 = 10$)	0.025, 0.138, 0.140	0.005, 0.104, 0.104	−0.004, 0.082, 0.082
Calibration-Y (CB)	0.003, 0.140, 0.140	0.010, 0.103, 0.104	−0.003, 0.069, 0.069

In this table, the true copula is the trivariate Gumbel copula with $\gamma_0 = 1.818$ (Kendall's $\tau = 0.45$). We report bias, standard deviation, and RMSE of each estimator with respect to γ_0 after $J = 1000$ Monte Carlo trials. The missing mechanism of Y_{2i} is specified as $\pi_2(X_i, Y_{1i}) = [1 + \exp(-0.42 + 0.2X_i + 0.2Y_{1i})]^{-1}$. This implies MAR with $\mathrm{E}[T_{2i}] = 0.6$. "Param (correct)" is the parametric estimator based on a correctly specified propensity score model. "Nonparam" is the nonparametric estimator of Hirano, Imbens, and Ridder (2003), where the approximation sieve consists of power series of X_i only: $u_{K_2}(X_i) = (1, X_i, X_i^2, \ldots, X_i^{K_2-1})^\top$. "Calibration" is the calibration estimator whose approximation sieve consists of power series of X_i only. "Calibration-Y" is the calibration estimator whose approximation sieve consists of power series of X_i and Y_{1i}: $u_{10}(X_i, Y_{1i}) = (1, X_i, Y_{1i}, X_i^2, Y_{1i}^2, X_iY_{1i}, X_i^3, Y_{1i}^3, X_i^2Y_{1i}, X_iY_{1i}^2)^\top$ for $K_2 = 10$. When $K_2 \leq 10$, $u_{K_2}(X_i, Y_{1i})$ consists of the first K_2 elements of $u_{10}(X_i, Y_{1i})$. "CB" signifies that the data-driven K_2^* is chosen among $K_2 \in \{1, \ldots, \bar{K}_2\}$ based on the covariate balancing principle. $\bar{K}_2 = 5$ for "Nonparam" and "Calibration", while $\bar{K}_2 = 10$ for "Calibration-Y".

Table 14: Simulation results on $G_3(4.000)$ with misspecified missing mechanism

	$N = 250$	$N = 500$	$N = 1000$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Param (correct)	−0.327, 0.304, 0.447	−0.175, 0.229, 0.288	−0.074, 0.171, 0.186
Nonparam ($K_2 = 2$)	−0.304, 0.302, 0.429	−0.157, 0.232, 0.280	−0.070, 0.173, 0.186
Nonparam ($K_2 = 3$)	−0.331, 0.303, 0.449	−0.142, 0.220, 0.262	−0.066, 0.189, 0.201
Nonparam ($K_2 = 4$)	−0.600, 0.899, 1.080	−0.710, 1.117, 1.323	−0.815, 1.279, 1.517
Nonparam (CB)	−0.297, 0.316, 0.434	−0.139, 0.226, 0.265	−0.062, 0.166, 0.177
Calibration ($K_2 = 2$)	−0.206, 0.348, 0.405	−0.138, 0.247, 0.283	−0.098, 0.174, 0.200
Calibration ($K_2 = 3$)	−0.142, 0.326, 0.355	−0.088, 0.228, 0.245	−0.044, 0.160, 0.166
Calibration ($K_2 = 4$)	−0.157, 0.340, 0.375	−0.063, 0.236, 0.244	−0.043, 0.161, 0.167
Calibration (CB)	−0.137, 0.330, 0.357	−0.087, 0.241, 0.256	−0.041, 0.166, 0.171
Calibration-Y ($K_2 = 3$)	−0.174, 0.344, 0.386	−0.131, 0.247, 0.280	−0.095, 0.172, 0.196
Calibration-Y ($K_2 = 6$)	−0.135, 0.324, 0.351	−0.084, 0.244, 0.258	−0.053, 0.165, 0.173
Calibration-Y ($K_2 = 10$)	−0.162, 0.369, 0.403	−0.131, 0.265, 0.296	−0.109, 0.225, 0.249
Calibration-Y (CB)	−0.140, 0.323, 0.352	−0.083, 0.236, 0.250	−0.057, 0.169, 0.178

In this table, the true copula is the trivariate Gumbel copula with $\gamma_0 = 4.000$ (Kendall's $\tau = 0.75$). We report bias, standard deviation, and RMSE of each estimator with respect to γ_0 after $J = 1000$ Monte Carlo trials. The missing mechanism of Y_{2i} is specified as $\pi_2(X_i, Y_{1i}) = [1 + \exp(-0.42 + 0.2X_i + 0.2Y_{1i})]^{-1}$. This implies MAR with $\mathrm{E}[T_{2i}] = 0.6$. "Param (correct)" is the parametric estimator based on a correctly specified propensity score model. "Nonparam" is the nonparametric estimator of Hirano, Imbens, and Ridder (2003), where the approximation sieve consists of power series of X_i only: $u_{K_2}(X_i) = (1, X_i, X_i^2, \ldots, X_i^{K_2-1})^\top$. "Calibration" is the calibration estimator whose approximation sieve consists of power series of X_i only. "Calibration-Y" is the calibration estimator whose approximation sieve consists of power series of X_i and Y_{1i}: $u_{10}(X_i, Y_{1i}) = (1, X_i, Y_{1i}, X_i^2, Y_{1i}^2, X_iY_{1i}, X_i^3, Y_{1i}^3, X_i^2Y_{1i}, X_iY_{1i}^2)^\top$ for $K_2 = 10$. When $K_2 \leq 10$, $u_{K_2}(X_i, Y_{1i})$ consists of the first K_2 elements of $u_{10}(X_i, Y_{1i})$. "CB" signifies that the data-driven K_2^* is chosen among $K_2 \in \{1, \ldots, \bar{K}_2\}$ based on the covariate balancing principle. $\bar{K}_2 = 5$ for "Nonparam" and "Calibration", while $\bar{K}_2 = 10$ for "Calibration-Y".

Table 15: Simulation results on Clayton copula with two covariates

$\alpha_0 = 1.636$ (Kendall's $\tau = 0.45$)			
	$N = 250$	$N = 500$	$N = 1000$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Nonparam ($K_2 = 3$)	0.038, 0.248, 0.251	0.015, 0.179, 0.180	0.011, 0.121, 0.121
Nonparam ($K_2 = 6$)	0.028, 0.257, 0.259	0.020, 0.184, 0.185	0.000, 0.149, 0.149
Nonparam ($K_2 = 10$)	−0.068, 1.440, 1.441	−0.252, 1.152, 1.179	−0.403, 0.960, 1.041
Nonparam (CB)	0.032, 0.254, 0.256	0.001, 0.185, 0.185	−0.003, 0.136, 0.136
Calibration ($K_2 = 3$)	0.010, 0.253, 0.253	−0.001, 0.180, 0.180	−0.007, 0.128, 0.128
Calibration ($K_2 = 6$)	0.032, 0.239, 0.241	0.020, 0.168, 0.169	0.014, 0.128, 0.129
Calibration ($K_2 = 10$)	0.045, 0.267, 0.271	0.004, 0.176, 0.176	0.010, 0.126, 0.126
Calibration (CB)	0.022, 0.256, 0.257	−0.002, 0.173, 0.173	0.014, 0.129, 0.130

$\alpha_0 = 6.000$ (Kendall's $\tau = 0.75$)			
	$N = 250$	$N = 500$	$N = 1000$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Nonparam ($K_2 = 3$)	−0.266, 0.626, 0.681	−0.150, 0.461, 0.484	−0.086, 0.324, 0.335
Nonparam ($K_2 = 6$)	−0.252, 0.688, 0.733	−0.160, 0.531, 0.554	−0.153, 0.476, 0.500
Nonparam ($K_2 = 10$)	−0.929, 2.725, 2.879	−1.127, 4.028, 4.183	−1.579, 3.013, 3.402
Nonparam (CB)	−0.318, 0.671, 0.743	−0.106, 1.153, 1.158	−0.108, 1.265, 1.270
Calibration ($K_2 = 3$)	−0.308, 0.655, 0.724	−0.253, 0.458, 0.523	−0.178, 0.330, 0.375
Calibration ($K_2 = 6$)	−0.260, 0.619, 0.671	−0.147, 0.446, 0.470	−0.084, 0.331, 0.342
Calibration ($K_2 = 10$)	−0.265, 0.646, 0.698	−0.143, 0.455, 0.477	−0.076, 0.331, 0.340
Calibration (CB)	−0.301, 0.619, 0.688	−0.200, 0.458, 0.500	−0.098, 0.346, 0.360

This table reports bias, standard deviation, and RMSE of each estimator with respect to $\alpha_0 \in \{1.636, 6.000\}$ after $J = 1000$ Monte Carlo trials. There are two covariates $\boldsymbol{X}_i = (X_{1i}, X_{2i})^\top$, and the missing mechanism of Y_{2i} is specified as $\pi_2(\boldsymbol{X}_i) = [1+\exp(-0.42+0.2X_{1i}+0.2X_{2i})]^{-1}$. This implies MAR with $\mathrm{E}[T_{2i}] = 0.6$. The approximation sieve of dimension $K_2 = 10$ is constructed as $u_{10}(\boldsymbol{X}) = (1, X_1, X_2, X_1^2, X_2^2, X_1X_{2i}, X_1^3, X_2^3, X_1^2X_2, X_1X_2^2)^\top$. When $K_2 \leq 10$, $u_{K_2}(\boldsymbol{X})$ is defined to be the first K_2 elements of $u_{10}(\boldsymbol{X})$. We use $K_2 = 3$ (i.e., only the first moments of $\boldsymbol{X}$), $K_2 = 6$ (i.e., the second moments added), and $K_2 = 10$ (i.e., the third moments added). "CB" means that the data-driven K_2^* is chosen from the choice set $K_2 \in \{1, \dots, 10\}$ based on the covariate balancing principle.

Table 16: Simulation results on Gumbel copula with two covariates

$\gamma_0 = 1.818$ (Kendall's $\tau = 0.45$)			
	$N = 250$	$N = 500$	$N = 1000$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Nonparam ($K_2 = 3$)	−0.026, 0.134, 0.136	−0.009, 0.098, 0.098	−0.002, 0.070, 0.070
Nonparam ($K_2 = 6$)	−0.041, 0.140, 0.146	−0.023, 0.109, 0.112	−0.015, 0.102, 0.104
Nonparam ($K_2 = 10$)	−0.305, 0.364, 0.475	−0.388, 0.373, 0.538	−0.459, 0.381, 0.596
Nonparam (CB)	−0.065, 0.170, 0.182	−0.066, 0.182, 0.193	−0.076, 0.186, 0.201
Calibration ($K_2 = 3$)	−0.006, 0.141, 0.141	−0.008, 0.100, 0.100	−0.012, 0.069, 0.070
Calibration ($K_2 = 6$)	0.001, 0.136, 0.136	0.008, 0.096, 0.096	0.000, 0.065, 0.065
Calibration ($K_2 = 10$)	0.013, 0.143, 0.144	0.011, 0.103, 0.104	0.001, 0.077, 0.077
Calibration (CB)	−0.021, 0.135, 0.136	−0.021, 0.101, 0.103	−0.023, 0.075, 0.078

$\gamma_0 = 4.000$ (Kendall's $\tau = 0.75$)			
	$N = 250$	$N = 500$	$N = 1000$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Nonparam ($K_2 = 3$)	−0.340, 0.308, 0.459	−0.157, 0.228, 0.277	−0.092, 0.160, 0.185
Nonparam ($K_2 = 6$)	−0.320, 0.374, 0.492	−0.204, 0.356, 0.410	−0.135, 0.385, 0.408
Nonparam ($K_2 = 10$)	−1.312, 1.279, 1.832	−1.566, 1.348, 2.066	−1.708, 1.575, 2.323
Nonparam (CB)	−0.372, 0.407, 0.551	−0.322, 0.630, 0.708	−0.403, 1.335, 1.394
Calibration ($K_2 = 3$)	−0.205, 0.329, 0.387	−0.127, 0.241, 0.273	−0.101, 0.174, 0.201
Calibration ($K_2 = 6$)	−0.131, 0.320, 0.345	−0.089, 0.243, 0.259	−0.057, 0.173, 0.182
Calibration ($K_2 = 10$)	−0.173, 0.361, 0.400	−0.121, 0.283, 0.308	−0.107, 0.231, 0.254
Calibration (CB)	−0.204, 0.324, 0.383	−0.161, 0.258, 0.303	−0.166, 0.250, 0.300

This table reports bias, standard deviation, and RMSE of each estimator with respect to $\gamma_0 \in \{1.818, 4.000\}$ after $J = 1000$ Monte Carlo trials. There are two covariates $\boldsymbol{X}_i = (X_{1i}, X_{2i})^\top$, and the missing mechanism of Y_{2i} is specified as $\pi_2(\boldsymbol{X}_i) = [1+\exp(-0.42+0.2X_{1i}+0.2X_{2i})]^{-1}$. This implies MAR with $\mathrm{E}[T_{2i}] = 0.6$. The approximation sieve of dimension $K_2 = 10$ is constructed as $u_{10}(\boldsymbol{X}) = (1, X_1, X_2, X_1^2, X_2^2, X_1X_{2i}, X_1^3, X_2^3, X_1^2X_2, X_1X_2^2)^\top$. When $K_2 \leq 10$, $u_{K_2}(\boldsymbol{X})$ is defined to be the first K_2 elements of $u_{10}(\boldsymbol{X})$. We use $K_2 = 3$ (i.e., only the first moments of $\boldsymbol{X}$), $K_2 = 6$ (i.e., the second moments added), and $K_2 = 10$ (i.e., the third moments added). "CB" means that the data-driven K_2^* is chosen from the choice set $K_2 \in \{1, \ldots, 10\}$ based on the covariate balancing principle.

References

ANDREWS, D. W. (1994): "Empirical process methods in econometrics," *Handbook of econometrics*, 4, 2247–2294.

CHEN, X., AND Y. FAN (2005): "Pseudo-Likelihood Ratio Tests for Semiparametric Multivariate Copula Model Selection," *The Canadian Journal of Statistics*, 33, 389–414.

GENEST, C., K. GHOUDI, AND L.-P. RIVEST (1995): "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions," *Biometrika*, 82, 543–552.

HAMORI, S., K. MOTEGI, AND Z. ZHANG (2018): "Calibration Estimation of Semiparametric Copula Models with Data Missing at Random," Working Paper at Kobe University and Renmin University of China.

HASTIE, T., R. TIBSHIRANI, AND J. FRIEDMAN (2009): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, second edn.

HIRANO, K., G. W. IMBENS, AND G. RIDDER (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71, 1161–1189.

TSENG, P., AND D. P. BERTSEKAS (1987): "Relaxation Methods for Problems with Strictly Convex Separable Costs and Linear Constraints," *Mathematical Programming*, 38, 303–321.