Online supplement for "*Copula-based regression models with data missing at random*"

Shigeyuki Hamori*– Kobe University

Kaiji Motegi†– Kobe University

Zheng Zhang‡– Renmin University of China

June 30, 2020

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

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

‡*Corresponding author*. Institute of Statistics and Big Data, Renmin University of China. Address: Haidian District, Beijing 100872 China. E-mail: zhengzhang@ruc.edu.cn

Contents

List of Tables

List of Figures

1 Introduction

In this document, we provide technical and numerical details omitted from the main paper Hamori, Motegi, and Zhang (2020). In Section 2, potential extensions to censored and truncated data are briefly discussed (cf. Remark 1 of the main paper). In Section 3, the relationship among the proposed calibration estimator for the copula parameter, $\hat{\boldsymbol{\theta}}_K$, the pseudo-maximum likelihood estimator of Genest, Ghoudi, and Rivest (1995), and the Z-estimator of Tsukahara (2005) are clarified (cf. Remark 2 of the main paper). In Section 4, an alternative approach using vine copulas is briefly discussed (cf. Remark 4 of the main paper). In Section 5, we state the regularity conditions for Proposition 1 of the main paper, which is identical to Theorems 1, 3, and 5 of Hamori, Motegi, and Zhang (2019). In Section 6, full details of the Monte Carlo simulations are presented.

2 A note on censored or truncated data

In the main paper, copula-based regression with data *missing at random (MAR)* is investigated. Data missing at random are a specific type of missing data, and copula-based regression with other types of missing data has not been explored in the literature yet. In particular, censored data (Emura, Lin, and Wang, 2010, De Backer, El Ghouch, and Van Keilegom, 2017) and truncated data (Emura and Wang, 2010, 2012) should be of practical interest. In this section, we provide a brief discussion on these cases.

For censored data, we consider the estimation of a simple bivariate copula, denoted by $c(u_1, u_2; \boldsymbol{\theta}_0)$, between Y_1 and Y_2. Assume that $\{Y_1, Y_2\}$ are censored. The censoring times, denoted by C_1 and C_2, are conditionally independent of Y_1 and Y_2 given a vector of observed covariates $\boldsymbol{X}$. What we observe are *i.i.d.* replications $\{T_{1i}, T_{2i}, O_{1i}, O_{2i}\}$, where $T_{ji} := \mathbf{1}(Y_{ji} \leq C_{ji})$ indicates the censoring status of Y_{ji}, and $O_{ji} := \min\{Y_{ji}, C_{ji}\}$ indicates an observed outcome, for $j \in \{1, 2\}$. The target copula parameter $\boldsymbol{\theta}_0$ can be identified through

$$\begin{aligned}\boldsymbol{\theta}_0 &= \arg\max_{\boldsymbol{\theta}} \mathrm{E}\left[\ln c\left\{F_1(Y_{1i}), F_2(Y_{2i}); \boldsymbol{\theta}\right\}\right] \\ &= \arg\max_{\boldsymbol{\theta}} \mathrm{E}\left[\frac{\mathbf{1}(T_{1i} = T_{2i} = 1)}{\eta(O_{1i}, O_{2i}; \boldsymbol{X}_i)} \ln c\left\{F_1(O_{1i}), F_2(O_{2i}); \boldsymbol{\theta}\right\}\right],\end{aligned}$$

where $\eta(c_1, c_2; \boldsymbol{X}) := \Pr(C_1 > c_1, C_2 > c_2 | \boldsymbol{X})$ is the conditional survival function of censoring variables. Using estimated $\eta(c_1, c_2; \boldsymbol{X})$ and estimated marginal distribution functions $\{F_j\}_{j=1}^2$, an estimator of $\boldsymbol{\theta}_0$ can be obtained.

For truncated data, a pair of data (Y_1, Y_2) is observed *if and only if* $Y_1 \leq Y_2$. Emura and

Wang (2012) study the estimation of the copula between the truncated Y_1 and Y_2. In the absence of missing variables at random, the copula parameter is identified as

$$\boldsymbol{\theta}_0 = \arg\max \mathrm{E}\left[\ln \frac{c\{F_1(Y_{i1}), F_2(Y_{i2})\}}{\int\int_{y_1 \le y_2} c\{F_1(y_1), F_2(y_2)\} dF_1(y_1) dF_2(y_2)}\right].$$

A possible extension in this literature is to assume that both Y_1 and Y_2 are MAR.

3 The calibration estimator $\hat{\boldsymbol{\theta}}_K$ under complete data

Recall from the main paper Hamori, Motegi, and Zhang (2020, Eq. (22)) that the calibration estimator for the copula parameter $\boldsymbol{\theta}_0$ is given by

$$\hat{\boldsymbol{\theta}}_K := \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{N} \mathbf{1}(T_{0i} = \cdots = T_{di} = 1) \hat{q}_K(\boldsymbol{X}_i) \ln c\left\{\hat{F}_{0,K}(Y_i), \hat{F}_{1,K}(W_{1i}), \ldots, \hat{F}_{d,K}(W_{di}); \boldsymbol{\theta}\right\}. \quad (1)$$

In this section, we discuss how $\hat{\boldsymbol{\theta}}_K$ is simplified when there are no missing data:

$$\Pr(T_{ji} = 1) = 1, \quad \forall j \in \{0, 1, \ldots, d\}, \forall i \in \{1, \ldots, N\}.$$

The semiparametric estimation of copula under complete data is well established by Genest, Ghoudi, and Rivest (1995) and Tsukahara (2005). Assuming complete data, Genest, Ghoudi, and Rivest (1995) proposed the pseudo-maximum likelihood estimator for $\boldsymbol{\theta}_0$:

$$\hat{\boldsymbol{\theta}} := \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{N} \ln c\left\{\hat{F}_0(Y_i), \hat{F}_1(W_{1i}), \ldots, \hat{F}_d(W_{di}); \boldsymbol{\theta}\right\}, \quad (2)$$

where $\hat{F}_0, \hat{F}_1, \ldots, \hat{F}_d$ are empirical cumulative distribution functions (CDFs). Keeping the assumption of complete data, Tsukahara (2005) proposed the Z-estimator for $\boldsymbol{\theta}_0$, which is a solution to the estimating equation:

$$\sum_{i=1}^{N} \phi\left\{\hat{F}_0(Y_i), \hat{F}_1(W_{1i}), \ldots, \hat{F}_{d,K}(W_{di}); \boldsymbol{\theta}\right\} = 0,$$

where $\phi(\cdot; \boldsymbol{\theta})$ is an $\mathbb{R}^p$-valued function on $(0,1)^{d+1}$ whose components $\phi_\ell(\cdot; \boldsymbol{\theta})$ satisfy the condition

$$\int \phi_\ell(\boldsymbol{u}; \boldsymbol{\theta}) dC(\boldsymbol{u}; \boldsymbol{\theta}) = 0, \qquad \ell = 1, \ldots, p.$$

Let $\dot{c}(\boldsymbol{\theta}) = (\partial c/\partial \boldsymbol{\theta})(\boldsymbol{\theta})$. When $\phi(\cdot;\boldsymbol{\theta}) = c(\boldsymbol{\theta})^{-1}\dot{c}(\boldsymbol{\theta})$, the resulting Z-estimator coincides with (2). Hence, the pseudo-maximum likelihood estimator of Genest, Ghoudi, and Rivest (1995) is a special case of the Z-estimator of Tsukahara (2005).

Under complete data, the calibration estimator $\hat{\boldsymbol{\theta}}_K$ in (1) reduces to the pseudo-maximum likelihood estimator (2). This fact is demonstrated in Hamori, Motegi, and Zhang (2019, Remark 9). It is not necessarily the case that the calibration estimator coincides with the Z-estimator under complete data, since it is possible that $\phi(\cdot;\boldsymbol{\theta}) \neq c(\boldsymbol{\theta})^{-1}\dot{c}(\boldsymbol{\theta})$.

4 A note on the vine-copula approach

In the main paper, a semiparametric copula with only one or few parameters governs the entire dependence among the $(d+1)$ variables $\{Y, \boldsymbol{W}\}$. It may be too restrictive when d is large, and a potential solution is adopting vine copulas (Joe and Kurowicka, 2011). In this section, a specific procedure for adopting vine copulas is described.

Suppose for concreteness that $d = 2$ with $\boldsymbol{W} = (W_1, W_2)^\top$. By Sklar's (1959) Theorem and the formula that

$$f_{Y,W_1,W_2}(y, w_1, w_2) = f_{Y|W_1,W_2}(y|w_1, w_2) \cdot f_{W_1|W_2}(w_1|w_2) \cdot f_{W_2}(w_2),$$

the joint density function of $\{Y, W_1, W_2\}$ can be decomposed as follows:

$$\begin{aligned} f_{Y,W_1,W_2}(y, w_1, w_2) =& f_0(y) \cdot f_1(w_1) \cdot f_2(w_2) \qquad \text{(marginals)} \\ & \times c_{01}\{F_0(y), F_1(w_1)\} \cdot c_{12}\{F_1(w_1), F_2(w_2)\} \qquad \text{(unconditional pairs)} \\ & \times c_{02|1}\{F_{0|1}(y|w_1), F_{2|1}(w_2|w_1)\}, \qquad \text{(conditional pairs)} \end{aligned}$$

where c_{01} is the copula for (Y, W_1); c_{12} is the copula for (W_1, W_2); $c_{02|1}$ is the conditional copula for (Y, W_2) given W_1; $F_{0|1}$ is the conditional CDF of Y given W_1; $F_{2|1}$ is the conditional CDF of W_2 given W_1.

Hence, Eq. (5) in the main paper is modified as follows.

$$a_0(\boldsymbol{w}) = \arg\min_{a \in \mathbb{R}} \mathrm{E}\left[\frac{T_0}{\pi_0(\boldsymbol{X})} L\{g(Y) - a\} c_{01}\{F_0(Y), F_1(w_1)\} c_{02|1}\{F_{0|1}(Y|w_1), F_{2|1}(w_2|w_1)\}\right]. \quad (3)$$

With the pairwise copulas $c_{01}(\cdot) = c_{01}(\cdot;\theta_{01})$ and $c_{02|1}(\cdot) = c_{02|1}(\cdot;\theta_{02|1})$ being parameterized, the parameters can be estimated by maximizing pseudo-likelihood functions. For example, estimators

for θ_{01} and $\theta_{02|1}$ are given by

$$\hat{\theta}_{01} = \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{N} \mathbf{1}(T_{0i} = T_{1i} = 1)\hat{e}_K(\boldsymbol{X}_i) \ln c_{01}\left\{\hat{F}_0(Y_i), \hat{F}_1(W_{1i}); \boldsymbol{\theta}\right\},$$
$$\hat{\theta}_{02|1} = \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{N} \mathbf{1}(T_{0i} = T_{1i} = T_{2i} = 1)\hat{q}_K(\boldsymbol{X}_i) \ln c_{02|1}\left\{\hat{F}_{0|1}(Y_i|W_{1i}), \hat{F}_{2|1}(W_{2i}|W_{1i}); \boldsymbol{\theta}\right\},$$

where $\hat{e}_K(\boldsymbol{X}_i)$ is the covariate balancing estimator of $\{N\cdot\Pr(T_{0i} = T_{1i} = 1|\boldsymbol{X}_i)\}^{-1}$ whose definition is similar to $\hat{q}_K(\boldsymbol{X}_i)$; $\hat{F}_{0|1}$ and $\hat{F}_{2|1}$ are consistent nonparametric estimators for the conditional CDFs $F_{0|1}$ and $F_{2|1}$, respectively. We motivate an estimator for $F_{0|1}$ based on the following equation.

$$\begin{aligned} F_{0|1}(y|w_1) &= \lim_{h\to 0} \frac{\Pr(Y_i \le y, w_1 - h \le W_{i1} \le w_1 + h)}{\Pr(w_1 - h \le W_{i1} \le w_1 + h)} \\ &= \lim_{h\to 0} \frac{\mathrm{E}[\mathbf{1}(T_{0i} = T_{1i} = 1)\cdot\{\Pr(T_{0i} = T_{1i} = 1|\boldsymbol{X}_i)\}^{-1}\mathbf{1}(Y_i \le y, w_1 - h \le W_{i1} \le w_1 + h)]}{\mathrm{E}[\mathbf{1}(T_{1i} = 1)\cdot\{\pi_1(\boldsymbol{X}_i)\}^{-1}\cdot\mathbf{1}(w_1 - h \le W_{i1} \le w_1 + h)]}. \end{aligned}$$

The estimator of $F_{0|1}$ can be defined as

$$\hat{F}_{0|1}(y|w_1) = \frac{\sum_{i=1}^{N} \mathbf{1}(T_{0i} = T_{1i} = 1)\cdot\hat{r}_K(\boldsymbol{X}_i)\cdot\mathbf{1}(Y_i \le y, w_1 - h \le W_{i1} \le w_1 + h)}{\sum_{i=1}^{N} \mathbf{1}(T_{1i} = 1)\cdot\hat{p}_{1K}(\boldsymbol{X}_i)\cdot\mathbf{1}(w_1 - h \le W_{i1} \le w_1 + h)},$$

where $h > 0$ is a bandwidth shrinking to zero as the sample size increases. An estimator of $F_{2|1}$ can be constructed similarly. Using the estimated vine copulas and (3), the target $a_0(\boldsymbol{w})$ can be estimated.

Although the vine-copula approach is useful for handling high-dimensional regressors, it is beyond the scope of the present paper to investigate its theoretical properties. We leave it as a future task.

5 Regularity conditions for the asymptotic normality of $\hat{\boldsymbol{\theta}}_K$

Proposition 1.(iii) of the main paper states that the proposed calibration estimator for the copula parameter, $\hat{\boldsymbol{\theta}}_K$, is asymptotically normal. Regularity conditions for the asymptotic normality are identical to those for Hamori, Motegi, and Zhang (2019, Theorem 5), and hence omitted from the main paper. In this section, the regularity conditions are restated formally for completeness.

For any function $f(v_0, v_1, ..., v_d; \boldsymbol{\theta})$, define the following derivatives.

$$\begin{aligned}
\partial_{\boldsymbol{\theta}} f(v_0, v_1, \ldots, v_d; \boldsymbol{\theta}) &:= \frac{\partial}{\partial \boldsymbol{\theta}} f(v_0, v_1, \ldots, v_d; \boldsymbol{\theta}), \\
\partial^2_{\boldsymbol{\theta}\boldsymbol{\theta}} f(v_0, v_1, \ldots, v_d; \boldsymbol{\theta}) &:= \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} f(v_0, v_1, \ldots, v_d; \boldsymbol{\theta}), \\
\partial_j f(v_0, v_1, \ldots, v_d; \boldsymbol{\theta}) &:= \frac{\partial}{\partial v_j} f(v_0, v_1, \ldots, v_d; \boldsymbol{\theta}) \quad \text{for } j \in \{0, 1, ..., d\}, \\
\partial^2_{\boldsymbol{\theta} j} f(v_0, v_1, \ldots, v_d; \boldsymbol{\theta}) &:= \frac{\partial^2}{\partial \boldsymbol{\theta} \partial v_j} f(v_0, v_1, \ldots, v_d; \boldsymbol{\theta}) \quad \text{for } j \in \{0, 1, ..., d\}.
\end{aligned}$$

Define

$$\begin{aligned}
&\ell(v_0, v_1, \ldots, v_d; \boldsymbol{\theta}) := \ln c(v_0, v_1, \ldots, v_d; \boldsymbol{\theta}), \\
&U_{0i} := F_0(Y_i), \\
&U_{ji} := F_j(W_{ji}) \ \text{ for } j \in \{1, \ldots, d\}, \\
&\boldsymbol{U}_i := (U_{0i}, U_{1i}, \ldots, U_{di})^\top.
\end{aligned}$$

The regularity conditions for Proposition 1.(iii) of the main paper are as follows.

Assumption 1 (missing at random) $\boldsymbol{T}_i \perp \boldsymbol{O}_{i,\text{mis}} \,|\, \boldsymbol{X}_i$.

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* $\mathrm{E}\left\{u_K(\boldsymbol{X}) u_K(\boldsymbol{X})^\top\right\}$ *is bounded away from zero uniformly in* K*.*

Assumption 4 *The inverse propensity scores* $\{\pi_j(\boldsymbol{x})^{-1}\}_{j=0}^d$ *and* $\eta^{-1}(\boldsymbol{x})$ *are bounded above, i.e., there exists some constant* $\delta < \infty$ *such that* $1 \le \pi_j(\boldsymbol{x})^{-1} \le \delta$ *and* $1 \le \eta^{-1}(\boldsymbol{x}) \le \delta$ *for any* $\boldsymbol{x} \in \mathcal{X}$ *and* $j \in \{0, 1, ..., d\}$*.*

Assumption 5 *There exist* $\boldsymbol{\lambda}_{jK}$*,* $\boldsymbol{\beta}_K \in \mathbb{R}^K$ *and* $\alpha > 0$ *such that* $\sup_{x \in \mathcal{X}} |(\rho')^{-1}\{1/\pi_j(\boldsymbol{x})\} - \boldsymbol{\lambda}_{jK}^\top u_K(\boldsymbol{x})| = O(K^{-\alpha})$ *and* $\sup_{x \in \mathcal{X}} |(\rho')^{-1}\{1/\eta(\boldsymbol{x})\} - \boldsymbol{\beta}_K^\top u_K(\boldsymbol{x})| = O(K^{-\alpha})$ *as* $K \to \infty$*.*

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

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

Assumption 8 $\mathrm{E}\{\partial_{\boldsymbol{\theta}}\ell(U_{0i}, U_{1i}, \ldots, U_{di}; \boldsymbol{\theta})|\boldsymbol{X}_i = \boldsymbol{x}\}$ *is continuously differentiable in* $\boldsymbol{x}$.

Assumption 9 $\boldsymbol{B} := -\mathrm{E}\left\{\partial^2_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell(U_{0i}, U_{1i}, \ldots, U_{di}; \boldsymbol{\theta}_0)\right\}$ *and*

$$\boldsymbol{\Sigma} := \mathrm{Var}\left\{\varphi(\boldsymbol{T}_i, \boldsymbol{X}_i, \boldsymbol{U}_i; \boldsymbol{\theta}_0) + \sum_{j=0}^{d} R_j(T_{ji}, \boldsymbol{X}_i, U_{ji}; \boldsymbol{\theta}_0)\right\}$$

are finite and positive definite, where

$$\begin{aligned}
\varphi(\boldsymbol{T}_i, \boldsymbol{X}_i, \boldsymbol{U}_i; \boldsymbol{\theta}_0) :=\ & \frac{\mathbf{1}(T_{0i} = \cdots = T_{di} = 1)}{\eta(\boldsymbol{X}_i)}\partial_{\boldsymbol{\theta}}\ell(U_{0i}, U_{1i}, \ldots, U_{di}; \boldsymbol{\theta}_0) - \mathrm{E}\left\{\partial_{\boldsymbol{\theta}}\ell(U_{0i}, U_{1i}, \ldots, U_{di}; \boldsymbol{\theta}_0)\right\} \\
& - \left\{\frac{\mathbf{1}(T_{0i} = \cdots = T_{di} = 1)}{\eta(\boldsymbol{X}_i)} - 1\right\}\mathrm{E}\left\{\partial_{\boldsymbol{\theta}}\ell(U_{0i}, U_{1i}, \ldots, U_{di}; \boldsymbol{\theta}_0)|\boldsymbol{X}_i\right\}, \\
R_j(T_{ji}, \boldsymbol{X}_i, U_{0i}; \boldsymbol{\theta}_0) :=\ & \mathrm{E}\left[\partial^2_{\boldsymbol{\theta}_j}\ell(U_{0s}, U_{1s}, \ldots, U_{ds}; \boldsymbol{\theta}_0)\{\phi_j(T_{ji}, \boldsymbol{X}_i, U_{ji}; U_{js}) - U_{js}\}|U_{ji}, \boldsymbol{X}_i, T_{ji}\right], \quad (s \neq i), \\
\phi_j(T_{ji}, \boldsymbol{X}_i, U_{ji}; v) :=\ & \frac{T_{ji}}{\pi_j(\boldsymbol{X}_i)}\mathbf{1}(U_{ji} \leq v) - \left\{\frac{T_{ji}}{\pi_j(\boldsymbol{X}_i)} - 1\right\}\mathrm{E}\left\{\mathbf{1}(U_{ji} \leq v)|\boldsymbol{X}_i\right\}, \quad v \in [0, 1].
\end{aligned}$$

Assumption 10 *(i) For each* $(u_0, u_1, \ldots, u_d) \in (0, 1)^{d+1}$, $\partial^2_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell(u_0, u_1, \ldots, u_d; \boldsymbol{\theta})$ *is continuous in* $\boldsymbol{\theta}$ *in a neighborhood of* $\boldsymbol{\theta}_0$. *(ii)* $\mathrm{E}[\sup_{\boldsymbol{\theta}\in\Theta:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|=o(1)}\|\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}(U_{0i}, U_{1i}, \ldots, U_{di}; \boldsymbol{\theta})\|] < \infty$.

Assumption 11 *For* $j \in \{0, 1, \ldots, d\}$, $\partial_{\boldsymbol{\theta}j}\ell(u_0, u_1, \ldots, u_d; \boldsymbol{\theta}_0)$ *is well defined and continuous in* $(u_0, u_1, \ldots, u_d) \in (0, 1)^{d+1}$. *Furthermore,*

(i) $\|\partial_{\boldsymbol{\theta}}\ell(u_0, u_1, \ldots, u_d; \boldsymbol{\theta}_0)\| \leq$ *constant* $\times \prod_{j=0}^{d}\{u_j(1-u_j)\}^{-a_j}$ *for some* $a_j \geq 0$ *such that* $\mathrm{E}[\prod_{j=0}^{d}\{U_{ji}(1-U_{ji})\}^{-2a_j}] < \infty$;

(ii) $\|\partial^2_{\boldsymbol{\theta}k}\ell(u_0, u_1, \ldots, u_d; \boldsymbol{\theta}_0)\| \leq$ *constant* $\times\{u_k(1-u_k)\}^{-b_k}\prod_{j=0, j\neq k}^{d}\{u_j(1-u_j)\}^{-a_j}$ *for some* $b_k > a_k$ *such that* $\mathrm{E}[\{U_{ki}(1-U_{ki})\}^{\xi_k - b_k}\prod_{j=0, j\neq k}^{d}\{U_{ji}(1-U_{ji})\}^{-a_j}] < \infty$ *for some* $\xi_k \in (0, 1/2)$.

Assumption 2 restricts the covariates to be bounded. This condition is admittedly restrictive but commonly imposed in the nonparametric regression literature, since it simplifies the derivation of the convergence rate under the sup-norm. Assumption 2 can be relaxed if we restrict the tail distribution of $\boldsymbol{X}$. Indeed, Chen, Hong, and Tarozzi (2008, Assumption 3) allows the support of $\boldsymbol{X}$ to be the entire Euclidean space by imposing $\int_{\mathbb{R}^r}(1 + |x|^2)^{\omega} f_X(x)dx < \infty$ for some $\omega > 0$, and applying the weighted sup-norm, defined as $\|g\|_{\infty,\omega} := \sup_{x\in\mathbb{R}^r}|g(x)(1 + |x|^2)^{-\omega/2}|$, to replace the sup-norm used in the present paper. The present paper, however, maintains the stronger assumption of the bounded support in order to simplify the proof and exposition.

Assumption 3, which is also imposed in Newey (1997), essentially requires the sieve basis functions to be orthogonal. Assumption 4, a common condition in the missing data literature,

ensures that a sufficient portion of marginal data are observed. Assumption 5 requires the sieve approximation errors of $(\rho')^{-1}\{\pi_j(\boldsymbol{x})^{-1}\}$ and $(\rho')^{-1}\{\eta(\boldsymbol{x})^{-1}\}$ to shrink at a polynomial rate. This condition is satisfied for a variety of sieve basis functions (Newey, 1997). If $\boldsymbol{X}$ is discrete, then the approximation error is zero for sufficiently large K, satisfying Assumption 5 with $\alpha = \infty$. If $\boldsymbol{X}$ is continuous, then the polynomial rate depends positively on the smoothness of $(\rho')^{-1}\{\pi_j(\boldsymbol{x})^{-1}\}$ and $(\rho')^{-1}\{\eta(\boldsymbol{x})^{-1}\}$ in the continuous components and negatively on the number of the continuous components; indeed, for power series and B-splines, $\alpha = -s/r$, where s is the smoothness of approximand and r is the dimension of $\boldsymbol{X}$.

Assumption 6, another common assumption in nonparametric regression, restricts the smoothing parameter to balance the bias and variance. Assumption 7 is a mild restriction on ρ which is satisfied in all important cases considered in the literature. Assumption 8 controls the approximation error. Assumption 9 guarantees the finiteness of the asymptotic variance. Assumption 10 guarantees the uniform convergence. Assumption 11 allows the score function and its partial derivatives with respect to the first d arguments to blow up at the boundaries, which occurs for many popular copula functions such as Gaussian, Clayton, and t-copulas.

Given these regularity conditions, the asymptotic normality of the estimated copula parameter $\hat{\boldsymbol{\theta}}_K$ can be established.

Proposition 5.1 *Under Assumptions 1-11, we have that*

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_K - \boldsymbol{\theta}_0) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{\xi}_i + o_p(1),$$

where $\boldsymbol{\xi}_i = \boldsymbol{B}^{-1}\{\varphi(\boldsymbol{T}_i, \boldsymbol{X}_i, \boldsymbol{U}_i; \boldsymbol{\theta}_0) + \sum_{j=0}^{d} R_j(T_{ji}, \boldsymbol{X}_i, U_{ji}; \boldsymbol{\theta}_0)\}$.

Proposition 5.1 is a restatement of Proposition 1.(iii) of the main paper and also Hamori, Motegi, and Zhang (2019, Theorem 5). See Hamori, Motegi, and Zhang (2019, Appendix D) for a proof of Proposition 5.1.

6 Full Monte Carlo simulations

In this section, complete details of Monte Carlo simulations are provided. In the main paper Hamori, Motegi, and Zhang (2020, Section 7), the benchmark scenario which has one regressor ($d = 1$) and one covariate ($r = 1$) is considered. In Section 6.1 below, we discuss the benchmark scenario in more detail. In Section 6.2, an extended scenario which has one regressor ($d = 1$) and two covariates ($r = 2$) is presented. In Section 6.3, another extended scenario which has two regressors ($d = 2$) and one covariate ($r = 1$) is presented.

6.1 Benchmark scenario

Let $\boldsymbol{Z}_i = (Z_{1i}, Z_{2i}, Z_{3i})^\top$, and draw $\boldsymbol{Z}_i$ independently and identically from two well-known copulas:

- trivariate Clayton copula with parameter $\theta_0 = 1.333$;
- trivariate Gumbel copula with parameter $\theta_0 = 1.667$.

For both cases, the implied Kendall's tau is $\tau = 0.4$, a moderate level of association. Define the regressand $Y_i = \Phi^{-1}(Z_{1i})$, regressor $W_i = \Phi^{-1}(Z_{2i})$, and covariate $X_i = \Phi^{-1}(Z_{3i})$, where $\Phi^{-1}(\cdot)$ is the inverse distribution function of $\mathcal{N}(0,1)$. Assume that W_i and X_i are observed for all individuals $i \in \{1, \ldots, N\}$, where the sample size is $N \in \{250, 500, 750\}$. Let T_i be a binary indicator which equals 1 if Y_i is observed and 0 if Y_i is missing. The regressand Y_i may be missing with the propensity score function:

$$\Pr(T_i = 1 \,|\, X_i = x) = \frac{1}{1 + \exp(b_0 + b_1 x)}. \tag{4}$$

The logistic function is commonly used to specify the propensity score function in the literature of missing data analysis (see, e.g., Qin, Leung, and Shao, 2002). Suppose that $(b_0, b_1) = (-0.57, 1.5)$, in which case $\mathrm{E}(T_i) = 0.6$ and Y_i is MAR.

Following Noh, El Ghouch, and Bouezmarni (2013), the conditional mean regression is considered here. In this case, the estimated regression curve $\hat{a}(w)$ has the closed-form solution:

$$\hat{a}(w) = \frac{\sum_{i=1}^N T_i \hat{p}_K(X_i) Y_i c\left\{\hat{F}_{0,K}(Y_i), \hat{F}_{1,K}(w); \hat{\theta}_K\right\}}{\sum_{i=1}^N T_i \hat{p}_K(X_i) c\left\{\hat{F}_{0,K}(Y_i), \hat{F}_{1,K}(w); \hat{\theta}_K\right\}}$$

(see Hamori, Motegi, and Zhang, 2020, Remark 3). Hence, the computation of $\hat{a}(w)$ is straightforward once the calibration weight $\hat{p}_K(X_i)$ is computed.

The calibration weights are computed with two alternative specifications for $\rho(\cdot)$:

- Exponential tilting (ET): $\rho(v) = -\exp(-v)$ and hence $\rho'(v) = \exp(-v)$. See Kitamura and Stutzer (1997) and Imbens, Spady, and Johnson (1998).
- Inverse logistic (IL): $\rho(v) = v - \exp(-v)$ and hence $\rho'(v) = 1 + \exp(-v)$.

A practical advantage of these specifications is that $\rho(v)$ is a well-defined increasing, strictly concave function for any $v \in \mathbb{R}$. This property stabilizes numerical optimization especially in

small samples, since the calibration weight must be positive by definition and its formula is indeed $\hat{p}_K(X) = N^{-1}\rho'\{\hat{\boldsymbol{\lambda}}_K^\top u_K(X)\}$. Other possible specifications for $\rho(\cdot)$ would be as follows.

- Empirical likelihood (EL): $\rho(v) = \ln(1+v)$ and hence $\rho'(v) = (1+v)^{-1}$. See Owen (1988).
- Continuously updated GMM (CU): $\rho(v) = -(1-v)^2/2$ and hence $\rho'(v) = 1-v$. See Hansen, Heaton, and Yaron (1996).

For EL, $\rho(v)$ is a well-defined increasing, strictly concave function for $v > -1$, but not for $v \leq -1$. Further, $\rho'(v)$ is undefined at $v = -1$ and negative for $v < -1$. The existence of the ill-behaved region $v \leq -1$ is not a problem in theory or in sufficiently large samples, but may produce an ill-behaved calibration weight $\hat{p}_K(X)$ in small samples. The same argument applies to CU. Hence, ET and IL are more useful than EL and CU in practice.

To compute the calibration weight, the sieve basis function is constructed as power series:

$$u_K(X) = (1, X, \ldots, X^{K-1})^\top. \tag{5}$$

The tuning parameter K is either fixed at $K \in \{2,3,4\}$, or automatically selected from the choice set $\{2,3,4\}$ via the balancing principle of covariate distributions (CB) or the M-fold cross validation (CV) with $M \in \{5,10\}$. In CB, K is chosen to minimize the distance between the estimated distributions of the covariate X in the observed group and the whole group. CV is one of the most well-known approaches for choosing K (Li and Racine, 2007, Hastie, Tibshirani, and Friedman, 2009). See Hamori, Motegi, and Zhang (2020, Section 6) for the procedures of CB and CV.

Since W is assumed to be always observed, the single calibration weight $\hat{p}_K(X_i)$ suffices in this study. The calibration weight with respect to the joint distribution of (Y,W), namely $\hat{q}_K(X_i)$, is identical to $\hat{p}_K(X_i)$ by construction. In the present simulation, a model misspecification is not discussed. When the underlying copula is Clayton (Gumbel), we fit the Clayton (Gumbel) copula to estimate θ_0 and then estimate the target function $a_0(w)$.

For comparison, the equal weight $\hat{p}_K(X_i) = 1/\sum_{i=1}^N T_i$ is also considered. The equal-weight approach, which is essentially equivalent to the list-wise deletion, should fail under the MAR mechanism (4), since by construction it ignores the impact of X on the propensity score.

6.1.1 Performance of point estimation

To evaluate the finite sample performance of the point estimation for $a_0(w)$, draw $J = 1000$ Monte Carlo samples and compute integrated root mean squared errors (IRMSEs) as follows. First, the

RMSE in the j^{th} sample is defined as

$$RMSE_j = \sqrt{\frac{1}{\#\mathcal{W}} \sum_{w \in \mathcal{W}} \{\hat{a}_j(w) - a_0(w)\}^2}, \tag{6}$$

where $\hat{a}_j(w)$ is the estimated target function; $\mathcal{W}$ is the set of w's considered; $\#\mathcal{W}$ is the number of w's considered. Since the marginal distribution of the regressor W_i is $\mathcal{N}(0,1)$, the range $w \in [-3,3]$ should be covered in order to properly evaluate the performance of each estimator. Dividing this interval more finely would make the evaluation more accurate, but it would raise computational burden. To balance the trade-off between evaluation accuracy and computational speed, we use $\mathcal{W} = \{-3.00, -2.95, \ldots, 2.95, 3.00\}$ and hence $\#\mathcal{W} = 121$.

Equation (6) contains the true value $a_0(w)$. Recall that

$$a_0(w) = \arg\min_{a \in \mathbb{R}} \ \mathrm{E}[L\{g(Y) - a\}c\{F_0(Y), F_1(w); \theta_0\}],$$

which can be approximated by

$$a_0(w) \simeq \arg\min_{a \in \mathbb{R}} \ \frac{1}{N} \sum_{i=1}^{N} L\{g(Y_i) - a\}c\{F_0(Y_i), F_1(w); \theta_0\}.$$

Since $L\{g(Y) - a\} = (Y - a)^2$ in the present study, $a_0(w)$ is simply given by

$$a_0(w) = \frac{\sum_{i=1}^{N} Y_i c\{F_0(Y_i), F_1(w); \theta_0\}}{\sum_{i=1}^{N} c\{F_0(Y_i), F_1(w); \theta_0\}}. \tag{7}$$

Substitute (7) into (6) to compute $RMSE_j$. Finally, the IRMSE is defined as

$$IRMSE = \frac{1}{J} \sum_{j=1}^{J} RMSE_j, \qquad J = 1000.$$

The smaller value of IRMSE implies the higher precision in the point estimation of $a_0(w)$.

See Table 1 for the resulting IRMSEs. For each copula and sample size, there are $2 \times 6 = 12$ versions of the calibration approach, since $\rho(\cdot)$ is specified in 2 ways and K is selected in 6 ways. For any copula and sample size, the IRMSE of any version of the calibration approach is always smaller than the IRMSE of the equal-weight approach. This is overwhelming evidence that the calibration approach strictly dominates the equal-weight approach when the regressand Y is MAR. Focus on the Clayton copula with $N = 250$, for example. The IRMSE of the calibration approach

takes the smallest value of 0.144 when the inverse logistic (IL) ρ-function is used and K is selected via the 10-fold cross validation (CV10); it takes the largest value of 0.194 when the IL ρ-function is used and K is fixed at 4. The IRMSE of the equal-weight approach is 0.270, which is clearly worse than the worst case of the calibration approach.

For any of the 12 versions of the calibration approach, the IRMSE shrinks as the sample size grows, which confirms the asymptotic validity of the calibration approach. See the pair $(IL, CV10)$ under the Clayton copula, for example. The IRMSEs are $\{0.144, 0.103, 0.087\}$ for $N \in \{250, 500, 750\}$, respectively.

We next discuss the relative performance between the exponential tilting (ET) and IL ρ-functions. There are $2 \times 3 \times 6 = 36$ cases where they can be compared, since 2 copulas, 3 sample sizes, and 6 ways to select K are considered. In 25 out of the 36 cases (69.4%), IL leads to the smaller IRMSE than ET. In particular, IL always outperforms ET whenever the cross validations are used to select K. In the extended scenarios, we shall observe similar or even stronger evidence for IL (see Sections 6.2 and 6.3). Hence, applied researchers are advised to use IL instead of ET in general.

Focusing on IL, we investigate the relative performance among the 6 selection mechanisms for K. For each copula and sample size, the IRMSE is (nearly) minimized when CV10 is used. CV10 attains the smallest IRMSE in 3 out of the 6 cases, and almost the smallest IRMSE in the other 3 cases. Thus, applied researchers are advised to use the pair (IL, CV10) in order to maximize the accuracy of the point estimation of the regression curve $a_0(w)$.

We plot the true regression curve $a_0(w)$ and the upper and lower 2.5-percentiles of estimated regression curves $\{\hat{a}_j(w)\}_{j=1}^J$ across $J = 1000$ Monte Carlo samples. See Figures 1-2 for the plots under the Clayton and Gumbel copulas, respectively. The calibration approach is implemented with (IL, CV10) as concluded above. For comparison, the conventional equal-weight approach is also considered.

Under the Clayton copula, the upper and lower 2.5-percentiles of $\{\hat{a}_j(w)\}_{j=1}^J$ based on the calibration approach are almost identical to $a_0(w)$ for any $w \in [-3, 3]$ (Panels (a)-(c), Figure 1). The distance between the two percentiles is sufficiently small even for $N = 250$, and it further shrinks toward $a_0(w)$ as N increases. It indicates that the calibration approach performs strikingly well.

For the equal-weight approach, the upper 2.5-percentile is *smaller* than $a_0(w)$ for most $w \in [-3, 3]$, indicating the presence of negative bias (Panels (d)-(f), Figure 1). This bias is a major source of the large IRMSE observed in Table 1. The bias does not vanish as sample size increases, which implies that the equal-weight approach fails when Y is MAR.

Under the Gumbel copula, the upper and lower 2.5-percentiles of $\{\hat{a}_j(w)\}_{j=1}^J$ based on the calibration approach are almost identical to $a_0(w)$ for $w \leq 1$ (Panels (a)-(c), Figure 2). The distance between the two percentiles is relatively large for $w > 1$, suggesting that the upper tail dependence of the Gumbel copula combined with the logistic-type MAR mechanism (4) has a large adverse effect on point estimation. Nevertheless, the two percentiles contain $a_0(w)$ especially when $N = 750$, indicating that reasonably sharp estimation at the upper tail is guaranteed in large samples.

The equal-weight approach utterly fails under the Gumbel copula (Panels (d)-(f), Figure 2). There exists extremely large negative bias for $w > 0$ regardless of the sample size. This result highlights the fatal shortcoming of the equal-weight approach and the remarkable use of the calibration approach under the MAR mechanism.

6.1.2 Performance of variance estimation

In this section, we investigate the finite sample performance of the variance estimation associated with the calibration approach. The inverse logistic $\rho(v) = v - \exp(-v)$ is used to compute calibration weights, since we know from the previous section that it performs best in point estimation. We compute 95% confidence intervals for $a_0(w)$, using the jackknife method, the bootstrap method I, and the bootstrap method II. In the first two methods, the confidence interval is constructed in accordance with the asymptotic normality result (Hamori, Motegi, and Zhang, 2020, Theorem 2), where the asymptotic variance is approximated via either the jackknife or bootstrap method. In the bootstrap method II, the confidence interval is constructed by picking the upper and lower 2.5-percentiles of bootstrapped $\{\hat{a}^{[b]}(w)\}_{b=1}^B$. For the two bootstrap methods, $B = 500$ bootstrap samples are generated. See the main paper Hamori, Motegi, and Zhang (2020, Section 5) for more detailed procedures.

The grid is simplified as $w \in \{-2.5, -1.5, -0.5, 0.5, 1.5, 2.5\}$. Another simplification is that the tuning parameter is fixed at $K \in \{2, 3, 4\}$ and the data-driven selection of K is not considered. Each K is used for computing a point estimator $\hat{a}(w)$, and the same K is used for computing a confidence interval. These simplifications partly stem from the heavy computational burden of the jackknife and bootstrap methods. Another reason is that it is well known that a data-driven selection of K in variance estimation is rather challenging. Nevertheless, the present simple set-up is informative enough to compare the finite sample performance of each method.

Based on each method, compute a 95% confidence interval $CI(w) = [b_\ell(w), b_u(w)]$. Repeat across $J = 1000$ Monte Carlo samples, resulting in $\{CI^{(1)}(w), \ldots, CI^{(J)}(w)\}$. Compute the

coverage probability of the 95% confidence intervals:

$$CP(w) = \frac{1}{J}\sum_{j=1}^{J} \mathbf{1}\{a_0(w) \in CI(w)\}$$

and the mean length of the 95% confidence intervals:

$$\bar{L}(w) = \frac{1}{J}\sum_{j=1}^{J} \{b_u^{(j)}(w) - b_\ell^{(j)}(w)\},$$

where $CI^{(j)}(w) = [b_\ell^{(j)}(w), b_u^{(j)}(w)]$. When evaluating the finite sample performance of the variance estimator, it is important to inspect both $CP(w)$ and $\bar{L}(w)$. There is often a trade-off between the two measures, and a sharp variance estimator is supposed to achieve both $CP(w)$ which is close enough to 0.95 and small enough $\bar{L}(w)$.

See Tables 2-4 for simulation results under the Clayton copula. For almost all values of $N \in \{250, 500, 750\}$, $K \in \{2, 3, 4\}$, and $w \in \{-2.5, \ldots, 2.5\}$, the coverage probability based on any method is close enough to 0.95 as desired. The mean length of the confidence intervals diminishes as N increases for all cases, suggesting the asymptotic validity of the three methods. In most cases, the mean length is sufficiently small for any method considered.

To further compare the relative performance of the three methods, focus on the case with $N = 250$, $w = 2.5$, and $K = 4$. It is a challenging case since the smallest sample size and the edge of the grid are picked. The coverage probabilities of the jackknife, bootstrap I, and bootstrap II are 0.964, 0.937, and 0.952, respectively (Tables 2-4). Hence, we do not observe a clear difference among the three methods in terms of the coverage probability. Their mean lengths, however, are 1.536, 0.789, and 0.780. The jackknife method leads to a much wider confidence interval than the bootstrap methods, suggesting that the latter are more robust to the choice of K in small samples. When $N \geq 500$, the mean length of the jackknife confidence intervals decreases rapidly, and the difference among the three methods becomes less apparent.

See Tables 5-7 for simulation results under the Gumbel copula. The coverage probability is fairly close to 0.95 for $w \leq 0.5$, but sometimes well below 0.95 for $w \geq 1.5$. See, for example, the jackknife method with $N = 250$ and $K = 3$ (Table 5). The coverage probabilities are $\{.897, .922, .946, .914, .795, .560\}$ for $w \in \{-2.5, -1.5, -0.5, 0.5, 1.5, 2.5\}$, respectively. This is not a surprising result, since the point estimation itself exhibits some distortions at the upper tail of the Gumbel copula (recall Figure 2).

As in the Clayton case, the jackknife method is sometimes sensitive to the choice of K in small

samples. See the case with $N = 250$, $K = 4$, and $w = -2.5$ (Table 5). The coverage probability is 0.942, but the mean length of the confidence intervals is 4.140, exhibiting a great deal of instability. The coverage probability and the mean length are $(0.912, 1.054)$ for the bootstrap method I (Table 6), while they are $(0.911, 1.075)$ for the bootstrap method II (Table 7). These results suggest that the bootstrap methods are more robust to the choice of K than the jackknife method.

In many cases, the two bootstrap methods perform roughly as well as each other (Tables 6-7). In some cases, however, the bootstrap method I delivers the sharper inference than the method II. See the case with $K = 4$ and $w = 2.5$. For $N \in \{250, 500, 750\}$, the coverage probabilities are respectively $\{0.574, 0.697, 0.739\}$ for the method I while they are $\{0.254, 0.439, 0.540\}$ for the method II. Evidently, the former has the higher coverage probability than the latter for each sample size. The two methods have similar mean lengths: $\{1.432, 1.155, 0.970\}$ for the method I and $\{1.393, 1.126, 0.943\}$ for the method II. Hence, the bootstrap method I outperforms the bootstrap method II in this case.

In summary, all of the three methods considered operate reasonably well in most cases. In some challenging cases such as the upper tail of the Gumbel copula, the bootstrap method I, which combines the asymptotic normality and bootstrapping, tends to outperform the jackknife method and the bootstrap method II.

6.2 Extended scenario I: Two covariates

In this section, we extend the benchmark scenario by adding another covariate ($r = 2$). Let $\boldsymbol{Z}_i = (Z_{1i}, Z_{2i}, Z_{3i}, Z_{4i})^\top$, and draw $\boldsymbol{Z}_i$ independently and identically from

- the four-variable Clayton copula with parameter $\theta_0 = 1.442$;
- the four-variable Gumbel copula with parameter $\theta_0 = 1.719$.

For both cases, the implied Kendall's tau is $\tau = 0.4$ as in the benchmark scenario. The marginal distributions of $(Y_i, W_i, X_{1i}, X_{2i})$ are all $\mathcal{N}(0, 1)$ as in the benchmark scenario. Assume that W_i and $\boldsymbol{X}_i = (X_{1i}, X_{2i})^\top$ are observed for all individuals. The propensity score function is specified as

$$\Pr(T_i = 1 \,|\, \boldsymbol{X}_i = \boldsymbol{x}) = \frac{1}{1 + \exp(-0.53 + 0.75x_1 + 0.75x_2)}.$$

This specification implies that $\mathrm{E}(T_i) = 0.6$ and Y_i is MAR.

The calibration approach is taken with the exponential tilting (ET) or the inverse logistic (IL) ρ-function. Sieve basis functions are specified as follows.

$$u_3(\boldsymbol{X}) = (1, X_1, X_2)^\top,$$

$$u_6(\boldsymbol{X}) = (1, X_1, X_2, X_1^2, X_2^2, X_1X_2)^\top,$$
$$u_{10}(\boldsymbol{X}) = (1, X_1, X_2, X_1^2, X_2^2, X_1X_2, X_1^3, X_2^3, X_1^2X_2, X_1X_2^2)^\top.$$

The tuning parameter is either fixed at $K \in \{3, 6, 10\}$, or automatically selected from the choice set $\{3, 6, 10\}$ via the covariate balancing approach (CB) or the M-fold cross validation (CV) with $M \in \{5, 10\}$.

Other settings are all analogous to the benchmark scenario. Given the present setting, we inspect the finite sample performance of point estimation. To conserve space, we refrain from discussing variance estimation.

See Table 8 for the IRMSEs after $J = 1000$ Monte Carlo iterations. First, the calibration approach with any $\rho(\cdot)$ and the selection mechanism of K always leads to the smaller IRMSE than the conventional equal-weight approach. Second, the IRMSE of any version of the calibration approach shrinks as the sample size increases. These two results are exactly consistent with the results from the benchmark scenario (Section 6.1.1).

The IL ρ-function produces the smaller IRMSE than ET in 30 out of the 36 cases (83.3%). Further, IL always outperforms ET whenever the cross validations are used to select K. The superiority of IL relative to ET is even more salient than in the benchmark scenario.

Focus on IL and compare the relative performance of the 6 selection mechanisms for K. For each copula and sample size, the IRMSE is (nearly) minimized when CV10 is used. CV10 attains the smallest IRMSE in only 1 out of the 6 cases, but almost the smallest IRMSE in the other 5 cases. Thus, as in the benchmark scenario, the pair (IL, CV10) performs best.

6.3 Extended scenario II: Two regressors

In this section, we extend the benchmark scenario by adding another regressor ($d = 2$). The copulas used here are the same as those in the extended scenario I. The marginal distributions of $(Y_i, W_{1i}, W_{2i}, X_i)$ are all $\mathcal{N}(0, 1)$ as in the previous scenarios. The propensity score is specified as in (4). For the calibration estimator, the sieve basis function is specified as in (5). Since there are two regressors, IRMSEs are computed over a two-dimensional grid $(w_1, w_2) \in \{-3.0, -2.5, \ldots, 2.5, 3.0\} \times \{-3.0, -2.5, \ldots, 2.5, 3.0\}$. Hence, the number of (w_1, w_2) considered is $\#\mathcal{W} = 169$.

Other settings are all analogous to the benchmark scenario. Given the present setting, we inspect the finite sample performance of point estimation. To conserve space, we refrain from discussing variance estimation.

See Table 9 for the IRMSEs after $J = 1000$ Monte Carlo iterations. First, the calibration

approach always leads to the smaller IRMSE than the equal-weight approach. Second, the IRMSE of any version of the calibration approach shrinks as N grows. These results are exactly consistent with the results from the benchmark scenario and the extended scenario I.

The IL ρ-function yields the smaller IRMSE than ET in 25 out of the 36 cases (69.4%). Moreover, IL always outperforms ET whenever the cross validations are used. This result is identical to the result from the benchmark scenario.

Focus on IL and compare the relative performance of the 6 selection mechanisms for K. For each copula and sample size, the IRMSE is (nearly) minimized when CV10 is used. CV10 achieves the smallest IRMSE in 3 out of the 6 cases, and almost the smallest IRMSE in the other 3 cases. Thus, as in the benchmark scenario and the extended scenario I, applied researchers are advised to use the pair $(\text{IL}, \text{CV10})$.

Table 1: IRMSE of $\hat{a}(w)$ based on the calibration approach (benchmark scenario)

Clayton copula, $N = 250$

	$K=2$	$K=3$	$K=4$	CB	CV5	CV10
ET	0.174	0.160	0.170	0.162	0.169	0.173
IL	0.147	0.151	0.194	0.157	0.146	0.144
IRMSE of the equal-weight approach = 0.270						

Gumbel copula, $N = 250$

	$K=2$	$K=3$	$K=4$	CB	CV5	CV10
ET	0.291	0.258	0.282	0.253	0.288	0.289
IL	0.255	0.262	0.350	0.261	0.259	0.256
IRMSE of the equal-weight approach = 0.427						

Clayton copula, $N = 500$

	$K=2$	$K=3$	$K=4$	CB	CV5	CV10
ET	0.133	0.123	0.113	0.120	0.127	0.126
IL	0.104	0.109	0.125	0.104	0.106	0.103
IRMSE of the equal-weight approach = 0.249						

Gumbel copula, $N = 500$

	$K=2$	$K=3$	$K=4$	CB	CV5	CV10
ET	0.226	0.177	0.186	0.173	0.224	0.220
IL	0.181	0.180	0.209	0.184	0.183	0.179
IRMSE of the equal-weight approach = 0.384						

Clayton copula, $N = 750$

	$K=2$	$K=3$	$K=4$	CB	CV5	CV10
ET	0.094	0.105	0.088	0.103	0.096	0.096
IL	0.085	0.090	0.097	0.085	0.088	0.087
IRMSE of the equal-weight approach = 0.241						

Gumbel copula, $N = 750$

	$K=2$	$K=3$	$K=4$	CB	CV5	CV10
ET	0.236	0.145	0.151	0.142	0.215	0.212
IL	0.150	0.144	0.162	0.150	0.152	0.148
IRMSE of the equal-weight approach = 0.370						

This table reports the integrated root mean squared errors (IRMSEs) of the estimated regression curve $\hat{a}(w)$ under the benchmark scenario. Y is MAR and the sample size is $N \in \{250, 500, 750\}$. The IRMSE is computed across the grid $w \in \{-3.00, -2.95, \ldots, 2.95, 3.00\}$ and $J = 1000$ Monte Carlo samples. The calibration approach is taken, and two alternative specifications are used for $\rho(\cdot)$: exponential tilting (ET) $\rho(v) = -\exp(-v)$ and inverse logistic (IL) $\rho(v) = v - \exp(-v)$. The sieve basis function is constructed as $u_K(X) = (1, X, \ldots, X^{K-1})^\top$. The tuning parameter K is either fixed at $K \in \{2, 3, 4\}$, or automatically selected from the choice set $\{2, 3, 4\}$ via covariate balancing (CB) or M-fold cross validation (CV) with $M \in \{5, 10\}$. For comparison, the IRMSEs based on the equal-weight approach are also reported.

Table 2: Performance of confidence intervals based on the jackknife method (Clayton copula)

Coverage probability of 95% CI ($N = 250$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.927	0.955	0.954	0.948	0.940	0.941
3	0.915	0.941	0.946	0.944	0.939	0.938
4	0.947	0.966	0.977	0.968	0.964	0.964

Mean length of 95% CI ($N = 250$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.918	0.521	0.359	0.440	0.531	0.553
3	0.963	0.564	0.378	0.494	0.599	0.624
4	1.398	0.965	0.871	1.260	1.486	1.536

Coverage probability of 95% CI ($N = 500$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.929	0.938	0.947	0.942	0.941	0.942
3	0.929	0.930	0.949	0.943	0.946	0.945
4	0.944	0.946	0.954	0.947	0.950	0.950

Mean length of 95% CI ($N = 500$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.748	0.361	0.251	0.310	0.375	0.390
3	0.797	0.379	0.257	0.330	0.399	0.416
4	0.869	0.488	0.408	0.596	0.714	0.740

Coverage probability of 95% CI ($N = 750$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.940	0.956	0.948	0.946	0.948	0.950
3	0.955	0.939	0.946	0.952	0.950	0.948
4	0.955	0.961	0.954	0.949	0.943	0.942

Mean length of 95% CI ($N = 750$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.629	0.293	0.204	0.254	0.307	0.321
3	0.672	0.304	0.208	0.266	0.323	0.337
4	0.684	0.340	0.270	0.399	0.485	0.504

In this table, the benchmark scenario with the Clayton copula is considered. Y is MAR and the sample size is $N \in \{250, 500, 750\}$. The regression curve $a_0(w)$ is estimated via the calibration approach with inverse logistic $\rho(v) = v - \exp(-v)$, where $w \in \{-2.5, -1.5, -0.5, 0.5, 1.5, 2.5\}$. The sieve basis function is constructed as $u_K(X) = (1, X, \ldots, X^{K-1})^\top$ with $K \in \{2, 3, 4\}$. The 95% confidence interval (CI) for $a_0(w)$ is computed in accordance with the asymptotic normality result, where the asymptotic variance is approximated via the jackknife method. The same K as in the point estimation is used for the jackknife steps. This table reports the coverage probability and the mean length of the 95% CIs across $J = 1000$ Monte Carlo samples.

Table 3: Performance of confidence intervals based on the bootstrap method I (Clayton copula)

Coverage probability of 95% CI ($N = 250$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.900	0.934	0.941	0.933	0.924	0.924
3	0.888	0.939	0.947	0.938	0.943	0.942
4	0.908	0.959	0.954	0.939	0.938	0.937

Mean length of 95% CI ($N = 250$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.748	0.519	0.348	0.420	0.507	0.528
3	0.768	0.564	0.357	0.459	0.555	0.578
4	0.841	0.576	0.430	0.632	0.761	0.789

Coverage probability of 95% CI ($N = 500$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.889	0.950	0.956	0.948	0.935	0.935
3	0.855	0.949	0.947	0.932	0.931	0.932
4	0.893	0.955	0.954	0.954	0.951	0.948

Mean length of 95% CI ($N = 500$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.613	0.359	0.244	0.294	0.356	0.371
3	0.639	0.375	0.249	0.314	0.381	0.398
4	0.668	0.392	0.295	0.439	0.536	0.557

Coverage probability of 95% CI ($N = 750$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.897	0.937	0.935	0.942	0.926	0.931
3	0.897	0.934	0.947	0.939	0.943	0.943
4	0.909	0.952	0.957	0.956	0.963	0.960

Mean length of 95% CI ($N = 750$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.564	0.292	0.200	0.244	0.295	0.308
3	0.583	0.300	0.202	0.254	0.308	0.322
4	0.579	0.307	0.235	0.344	0.419	0.436

In this table, the benchmark scenario with the Clayton copula is considered. Y is MAR and the sample size is $N \in \{250, 500, 750\}$. The regression curve $a_0(w)$ is estimated via the calibration approach with inverse logistic $\rho(v) = v - \exp(-v)$, where $w \in \{-2.5, -1.5, -0.5, 0.5, 1.5, 2.5\}$. The sieve basis function is constructed as $u_K(X) = (1, X, \ldots, X^{K-1})^\top$ with $K \in \{2, 3, 4\}$. The 95% confidence interval (CI) for $a_0(w)$ is computed in accordance with the asymptotic normality result, where the asymptotic variance is approximated by bootstrapping. The same K as in the point estimation is used for $B = 500$ bootstrap iterations. This table reports the coverage probability and the mean length of the 95% CIs across $J = 1000$ Monte Carlo samples.

Table 4: Performance of confidence intervals based on the bootstrap method II (Clayton copula)

Coverage probability of 95% CI ($N = 250$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.890	0.920	0.936	0.938	0.930	0.928
3	0.911	0.898	0.936	0.940	0.943	0.943
4	0.906	0.950	0.962	0.960	0.949	0.952

Mean length of 95% CI ($N = 250$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.720	0.522	0.348	0.419	0.506	0.526
3	0.733	0.569	0.358	0.459	0.555	0.577
4	0.803	0.577	0.430	0.628	0.754	0.780

Coverage probability of 95% CI ($N = 500$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.926	0.935	0.959	0.955	0.941	0.941
3	0.910	0.915	0.941	0.938	0.936	0.934
4	0.940	0.959	0.955	0.949	0.943	0.941

Mean length of 95% CI ($N = 500$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.607	0.360	0.244	0.294	0.356	0.371
3	0.628	0.377	0.250	0.314	0.381	0.398
4	0.659	0.391	0.297	0.441	0.536	0.557

Coverage probability of 95% CI ($N = 750$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.913	0.936	0.937	0.942	0.928	0.931
3	0.893	0.916	0.939	0.946	0.940	0.945
4	0.923	0.952	0.958	0.950	0.950	0.951

Mean length of 95% CI ($N = 750$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.562	0.293	0.200	0.243	0.295	0.307
3	0.581	0.301	0.202	0.254	0.308	0.322
4	0.578	0.307	0.235	0.343	0.417	0.433

In this table, the benchmark scenario with the Clayton copula is considered. Y is MAR and the sample size is $N \in \{250, 500, 750\}$. The regression curve $a_0(w)$ is estimated via the calibration approach with inverse logistic $\rho(v) = v - \exp(-v)$, where $w \in \{-2.5, -1.5, -0.5, 0.5, 1.5, 2.5\}$. The sieve basis function is constructed as $u_K(X) = (1, X, \ldots, X^{K-1})^\top$ with $K \in \{2, 3, 4\}$. The 95% confidence interval (CI) for $a_0(w)$ is computed by picking the upper and lower 2.5-percentiles of $B = 500$ bootstrapped estimates for $a_0(w)$. The same K as in the point estimation is used for the bootstrap iterations. This table reports the coverage probability and the mean length of the 95% CIs across $J = 1000$ Monte Carlo samples.

Table 5: Performance of confidence intervals based on the jackknife method (Gumbel copula)

Coverage probability of 95% CI ($N = 250$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.822	0.874	0.928	0.907	0.755	0.520
3	0.897	0.922	0.946	0.914	0.795	0.560
4	0.942	0.960	0.974	0.939	0.891	0.764

Mean length of 95% CI ($N = 250$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.769	0.541	0.345	0.403	0.774	1.300
3	0.847	0.589	0.377	0.447	0.841	1.321
4	4.140	2.829	1.423	1.793	3.716	3.902

Coverage probability of 95% CI ($N = 500$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.884	0.912	0.941	0.915	0.810	0.631
3	0.913	0.934	0.943	0.923	0.856	0.678
4	0.925	0.946	0.958	0.943	0.890	0.760

Mean length of 95% CI ($N = 500$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.568	0.382	0.240	0.292	0.580	1.119
3	0.588	0.394	0.246	0.304	0.601	1.128
4	1.609	1.052	0.529	0.759	1.630	1.963

Coverage probability of 95% CI ($N = 750$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.881	0.894	0.921	0.934	0.840	0.670
3	0.893	0.905	0.930	0.941	0.878	0.695
4	0.922	0.939	0.956	0.940	0.899	0.753

Mean length of 95% CI ($N = 750$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.470	0.311	0.195	0.238	0.490	1.044
3	0.471	0.311	0.196	0.249	0.493	1.000
4	0.857	0.565	0.305	0.406	0.901	1.409

In this table, the benchmark scenario with the Gumbel copula is considered. Y is MAR and the sample size is $N \in \{250, 500, 750\}$. The regression curve $a_0(w)$ is estimated via the calibration approach with inverse logistic $\rho(v) = v - \exp(-v)$, where $w \in \{-2.5, -1.5, -0.5, 0.5, 1.5, 2.5\}$. The sieve basis function is constructed as $u_K(X) = (1, X, \ldots, X^{K-1})^\top$ with $K \in \{2, 3, 4\}$. The 95% confidence interval (CI) for $a_0(w)$ is computed in accordance with the asymptotic normality result, where the asymptotic variance is approximated via the jackknife method. The same K as in the point estimation is used for the jackknife steps. This table reports the coverage probability and the mean length of the 95% CIs across $J = 1000$ Monte Carlo samples.

Table 6: Performance of confidence intervals based on the bootstrap method I (Gumbel copula)

Coverage probability of 95% CI ($N = 250$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.848	0.898	0.956	0.840	0.726	0.464
3	0.867	0.907	0.944	0.871	0.764	0.467
4	0.912	0.945	0.966	0.904	0.826	0.574

Mean length of 95% CI ($N = 250$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.705	0.519	0.345	0.364	0.680	1.035
3	0.752	0.553	0.367	0.390	0.716	1.033
4	1.054	0.766	0.482	0.553	1.058	1.432

Coverage probability of 95% CI ($N = 500$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.849	0.883	0.935	0.900	0.783	0.567
3	0.879	0.914	0.950	0.911	0.813	0.574
4	0.934	0.952	0.979	0.933	0.890	0.697

Mean length of 95% CI ($N = 500$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.513	0.361	0.235	0.265	0.501	0.889
3	0.528	0.370	0.241	0.278	0.520	0.861
4	0.739	0.513	0.310	0.381	0.785	1.155

Coverage probability of 95% CI ($N = 750$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.883	0.907	0.946	0.919	0.818	0.615
3	0.893	0.915	0.943	0.914	0.833	0.633
4	0.939	0.956	0.964	0.936	0.917	0.739

Mean length of 95% CI ($N = 750$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.430	0.296	0.192	0.224	0.428	0.801
3	0.438	0.301	0.195	0.229	0.432	0.803
4	0.573	0.390	0.232	0.290	0.609	0.970

In this table, the benchmark scenario with the Gumbel copula is considered. Y is MAR and the sample size is $N \in \{250, 500, 750\}$. The regression curve $a_0(w)$ is estimated via the calibration approach with inverse logistic $\rho(v) = v - \exp(-v)$, where $w \in \{-2.5, -1.5, -0.5, 0.5, 1.5, 2.5\}$. The sieve basis function is constructed as $u_K(X) = (1, X, \ldots, X^{K-1})^\top$ with $K \in \{2, 3, 4\}$. The 95% confidence interval (CI) for $a_0(w)$ is computed in accordance with the asymptotic normality result, where the asymptotic variance is approximated by bootstrapping. The same K as in the point estimation is used for $B = 500$ bootstrap iterations. This table reports the coverage probability and the mean length of the 95% CIs across $J = 1000$ Monte Carlo samples.

Table 7: Performance of confidence intervals based on the bootstrap method II (Gumbel copula)

Coverage probability of 95% CI ($N = 250$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.770	0.821	0.940	0.787	0.582	0.280
3	0.768	0.834	0.921	0.822	0.617	0.271
4	0.911	0.945	0.973	0.899	0.778	0.254

Mean length of 95% CI ($N = 250$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.706	0.521	0.346	0.365	0.680	0.981
3	0.756	0.554	0.368	0.392	0.720	0.996
4	1.075	0.778	0.487	0.563	1.079	1.393

Coverage probability of 95% CI ($N = 500$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.781	0.831	0.932	0.855	0.656	0.460
3	0.833	0.880	0.947	0.870	0.715	0.477
4	0.901	0.929	0.977	0.896	0.809	0.439

Mean length of 95% CI ($N = 500$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.514	0.361	0.235	0.265	0.501	0.845
3	0.530	0.370	0.241	0.279	0.520	0.830
4	0.750	0.516	0.311	0.384	0.795	1.126

Coverage probability of 95% CI ($N = 750$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.837	0.867	0.935	0.872	0.711	0.553
3	0.859	0.877	0.943	0.883	0.747	0.553
4	0.899	0.929	0.963	0.887	0.828	0.540

Mean length of 95% CI ($N = 750$)

$K\backslash w$	-2.5	-1.5	-0.5	0.5	1.5	2.5
2	0.432	0.296	0.192	0.224	0.428	0.773
3	0.439	0.301	0.195	0.230	0.433	0.778
4	0.576	0.390	0.231	0.292	0.620	0.943

In this table, the benchmark scenario with the Gumbel copula is considered. Y is MAR and the sample size is $N \in \{250, 500, 750\}$. The regression curve $a_0(w)$ is estimated via the calibration approach with inverse logistic $\rho(v) = v - \exp(-v)$, where $w \in \{-2.5, -1.5, -0.5, 0.5, 1.5, 2.5\}$. The sieve basis function is constructed as $u_K(X) = (1, X, \ldots, X^{K-1})^\top$ with $K \in \{2, 3, 4\}$. The 95% confidence interval (CI) for $a_0(w)$ is computed by picking the upper and lower 2.5-percentiles of $B = 500$ bootstrapped estimates for $a_0(w)$. The same K as in the point estimation is used for the bootstrap iterations. This table reports the coverage probability and the mean length of the 95% CIs across $J = 1000$ Monte Carlo samples.

Table 8: IRMSE of $\hat{a}(w)$ based on the calibration approach (extended scenario I: two covariates)

Clayton copula, $N = 250$

	$K=3$	$K=6$	$K=10$	CB	CV5	CV10
ET	0.150	0.157	0.157	0.157	0.152	0.151
IL	0.140	0.151	0.168	0.152	0.141	0.139
IRMSE of the equal-weight approach = 0.288						

Gumbel copula, $N = 250$

	$K=3$	$K=6$	$K=10$	CB	CV5	CV10
ET	0.274	0.232	0.318	0.256	0.275	0.271
IL	0.230	0.231	0.384	0.263	0.239	0.238
IRMSE of the equal-weight approach = 0.433						

Clayton copula, $N = 500$

	$K=3$	$K=6$	$K=10$	CB	CV5	CV10
ET	0.113	0.120	0.113	0.120	0.116	0.116
IL	0.104	0.108	0.112	0.108	0.102	0.105
IRMSE of the equal-weight approach = 0.270						

Gumbel copula, $N = 500$

	$K=3$	$K=6$	$K=10$	CB	CV5	CV10
ET	0.212	0.158	0.239	0.202	0.211	0.207
IL	0.159	0.156	0.260	0.176	0.163	0.159
IRMSE of the equal-weight approach = 0.392						

Clayton copula, $N = 750$

	$K=3$	$K=6$	$K=10$	CB	CV5	CV10
ET	0.082	0.108	0.100	0.104	0.094	0.093
IL	0.087	0.090	0.089	0.089	0.088	0.088
IRMSE of the equal-weight approach = 0.264						

Gumbel copula, $N = 750$

	$K=3$	$K=6$	$K=10$	CB	CV5	CV10
ET	0.203	0.126	0.219	0.181	0.204	0.205
IL	0.134	0.128	0.200	0.141	0.129	0.131
IRMSE of the equal-weight approach = 0.371						

This table reports the integrated root mean squared errors (IRMSEs) of the estimated regression curve $\hat{a}(w)$ under the extended scenario I (i.e., two covariates). Y is MAR and the sample size is $N \in \{250, 500, 750\}$. The IRMSE is computed across the grid $w \in \{-3.00, -2.95, \ldots, 2.95, 3.00\}$ and $J = 1000$ Monte Carlo samples. The calibration approach is taken, and two alternative specifications are used for $\rho(\cdot)$: exponential tilting (ET) $\rho(v) = -\exp(-v)$ and inverse logistic (IL) $\rho(v) = v - \exp(-v)$. The sieve basis function is constructed as $u_3(\boldsymbol{X}) = (1, X_1, X_2)^\top$, $u_6(\boldsymbol{X}) = (1, X_1, X_2, X_1^2, X_2^2, X_1X_2)^\top$, or $u_{10}(\boldsymbol{X}) = (1, X_1, X_2, X_1^2, X_2^2, X_1X_2, X_1^3, X_2^3, X_1^2X_2, X_1X_2^2)^\top$. The tuning parameter K is either fixed at $K \in \{3, 6, 10\}$, or automatically selected from the choice set $\{3, 6, 10\}$ via covariate balancing (CB) or M-fold cross validation (CV) with $M \in \{5, 10\}$. For comparison, the IRMSEs based on the equal-weight approach are also reported.

Table 9: IRMSE of $\hat{a}(\boldsymbol{w})$ based on the calibration approach (extended scenario II: two regressors)

Clayton copula, $N = 250$

	$K=2$	$K=3$	$K=4$	CB	CV5	CV10
ET	0.222	0.200	0.197	0.197	0.210	0.213
IL	0.175	0.187	0.208	0.193	0.176	0.175
IRMSE of the equal-weight approach = 0.281						

Gumbel copula, $N = 250$

	$K=2$	$K=3$	$K=4$	CB	CV5	CV10
ET	0.186	0.167	0.185	0.173	0.181	0.181
IL	0.170	0.171	0.220	0.179	0.172	0.169
IRMSE of the equal-weight approach = 0.259						

Clayton copula, $N = 500$

	$K=2$	$K=3$	$K=4$	CB	CV5	CV10
ET	0.176	0.152	0.132	0.157	0.159	0.157
IL	0.124	0.129	0.137	0.130	0.125	0.126
IRMSE of the equal-weight approach = 0.247						

Gumbel copula, $N = 500$

	$K=2$	$K=3$	$K=4$	CB	CV5	CV10
ET	0.144	0.120	0.124	0.123	0.134	0.135
IL	0.121	0.124	0.142	0.122	0.122	0.126
IRMSE of the equal-weight approach = 0.225						

Clayton copula, $N = 750$

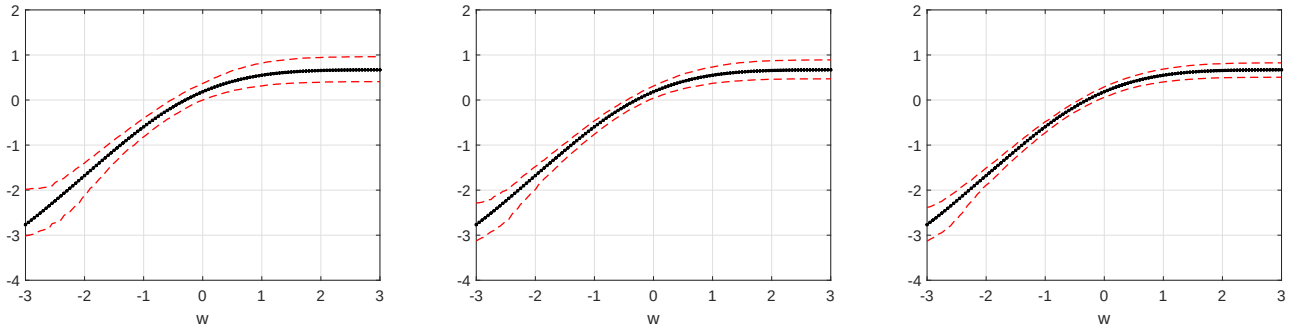
	$K=2$	$K=3$	$K=4$	CB	CV5	CV10
ET	0.123	0.131	0.111	0.133	0.122	0.121
IL	0.101	0.110	0.112	0.106	0.104	0.105
IRMSE of the equal-weight approach = 0.240						

Gumbel copula, $N = 750$

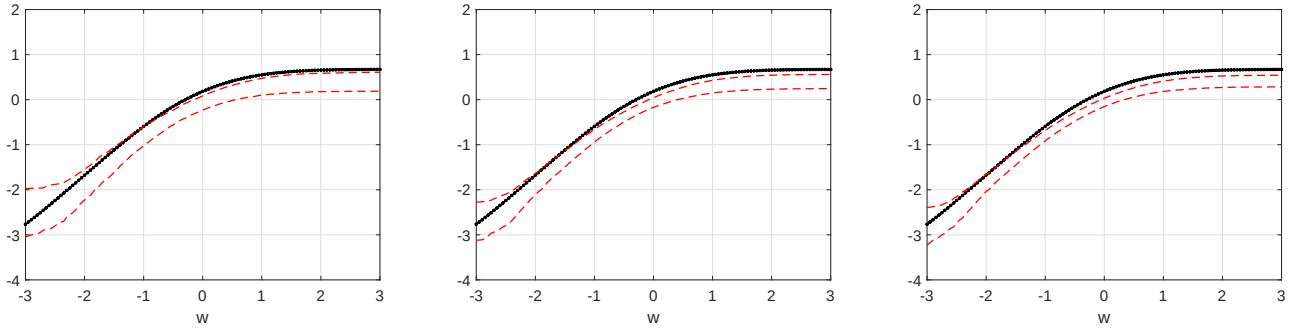
	$K=2$	$K=3$	$K=4$	CB	CV5	CV10
ET	0.145	0.099	0.103	0.101	0.119	0.112
IL	0.101	0.103	0.111	0.102	0.101	0.101
IRMSE of the equal-weight approach = 0.214						

This table reports the integrated root mean squared errors (IRMSEs) of the estimated regression curve $\hat{a}(\boldsymbol{w})$ under the extended scenario II (i.e., two regressors). Y is MAR and the sample size is $N \in \{250, 500, 750\}$. The IRMSE is computed across the grid $(w_1, w_2) \in \{-3.0, -2.5, \ldots, 2.5, 3.0\} \times \{-3.0, -2.5, \ldots, 2.5, 3.0\}$ and $J = 1000$ Monte Carlo samples. The calibration approach is taken, and two alternative specifications are used for $\rho(\cdot)$: exponential tilting (ET) $\rho(v) = -\exp(-v)$ and inverse logistic (IL) $\rho(v) = v - \exp(-v)$. The sieve basis function is constructed as $u_K(X) = (1, X, \ldots, X^{K-1})^\top$. The tuning parameter K is either fixed at $K \in \{2, 3, 4\}$, or automatically selected from the choice set $\{2, 3, 4\}$ via covariate balancing (CB) or M-fold cross validation (CV) with $M \in \{5, 10\}$. For comparison, the IRMSEs based on the equal-weight approach are also reported.

Figure 1: True $a_0(w)$ and the upper and lower 2.5-percentiles of $\{\hat{a}_j(w)\}_{j=1}^J$ (benchmark scenario, Clayton copula)



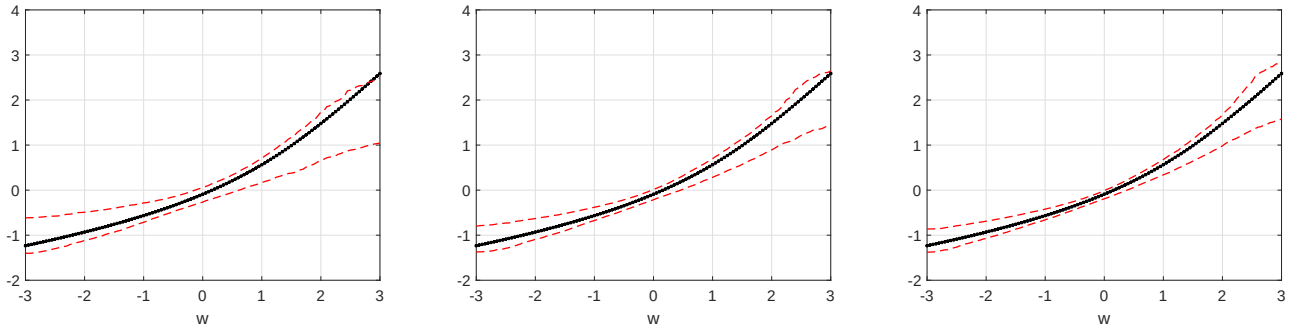
(a) CAL ($N = 250$) (b) CAL ($N = 500$) (c) CAL ($N = 750$)



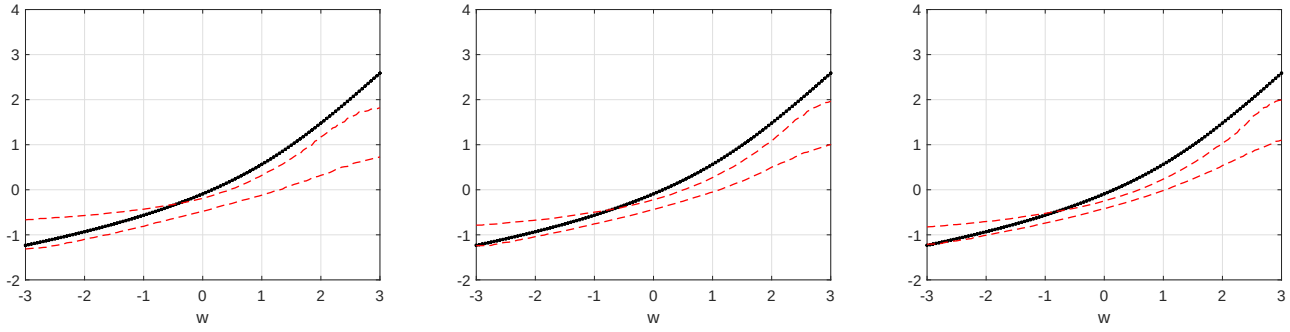
(d) EQW ($N = 250$) (e) EQW ($N = 500$) (f) EQW ($N = 750$)

This figure plots the true regression curve $a_0(w)$ in black, solid lines and the upper and lower 2.5-percentiles of estimated regression curves $\{\hat{a}_j(w)\}_{j=1}^J$ across $J = 1000$ Monte Carlo samples in red, dashed lines. The benchmark scenario with the Clayton copula is considered, and the sample sizes are $N \in \{250, 500, 750\}$. The regression curve is estimated via the calibration approach (CAL) or the equal-weight approach (EQW). For CAL, the inverse logistic $\rho(\cdot)$ and the data-driven selection of K based on the 10-fold cross validation are used, where the choice set is $K \in \{2, 3, 4\}$.

Figure 2: True $a_0(w)$ and the upper and lower 2.5-percentiles of $\{\hat{a}_j(w)\}_{j=1}^J$ (benchmark scenario, Gumbel copula)



(a) CAL ($N = 250$) (b) CAL ($N = 500$) (c) CAL ($N = 750$)



(d) EQW ($N = 250$) (e) EQW ($N = 500$) (f) EQW ($N = 750$)

This figure plots the true regression curve $a_0(w)$ in black, solid lines and the upper and lower 2.5-percentiles of estimated regression curves $\{\hat{a}_j(w)\}_{j=1}^J$ across $J = 1000$ Monte Carlo samples in red, dashed lines. The benchmark scenario with the Gumbel copula is considered, and the sample sizes are $N \in \{250, 500, 750\}$. The regression curve is estimated via the calibration approach (CAL) or the equal-weight approach (EQW). For CAL, the inverse logistic $\rho(\cdot)$ and the data-driven selection of K based on the 10-fold cross validation are used, where the choice set is $K \in \{2, 3, 4\}$.

References

Chen, X., H. Hong, and A. Tarozzi (2008): "Semiparametric Efficiency in GMM Models with Auxiliary Data," *The Annals of Statistics*, 36, 808–843.

De Backer, M., A. El Ghouch, and I. Van Keilegom (2017): "Semiparametric copula quantile regression for complete or censored data," *Electronic Journal of Statistics*, 11, 1660–1698.

Emura, T., C.-W. Lin, and W. Wang (2010): "A goodness-of-fit test for Archimedean copula models in the presence of right censoring," *Computational Statistics & Data Analysis*, 54(12), 3033–3043.

Emura, T., and W. Wang (2010): "Testing quasi-independence for truncation data," *Journal of Multivariate Analysis*, 101(1), 223–239.

——— (2012): "Nonparametric maximum likelihood estimation for dependent truncation data based on copulas," *Journal of Multivariate Analysis*, 110, 171–188.

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 (2019): "Calibration estimation of semiparametric copula models with data missing at random," *Journal of Multivariate Analysis*, 173, 85–109.

——— (2020): "Copula-based regression models with data missing at random," Kobe University and Renmin University of China.

Hansen, L. P., J. Heaton, and A. Yaron (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business & Economic Statistics*, 14, 262–280.

Hastie, T., R. Tibshirani, and J. Friedman (2009): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, second edn.

Imbens, G. W., R. H. Spady, and P. Johnson (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, 66, 333–357.

Joe, H., and D. Kurowicka (2011): *Dependence Modeling: Vine Copula Handbook*. World Scientific.

Kitamura, Y., and M. Stutzer (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65, 861–874.

Li, Q., and J. S. Racine (2007): *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Newey, W. K. (1997): "Convergence rates and asymptotic normality for series estimators," *Journal of Econometrics*, 79(1), 147–168.

Noh, H., A. El Ghouch, and T. Bouezmarni (2013): "Copula-based regression estimation and inference," *Journal of the American Statistical Association*, 108(502), 676–688.

Owen, A. B. (1988): "Empirical Likelihood Ratio Confidence Intervals for a Single Functional,"

Biometrika, 75, 237–249.

Qin, J., D. Leung, and J. Shao (2002): "Estimation with Survey Data under Nonignorable Nonresponse or Informative Sampling," *Journal of the American Statistical Association*, 97, 193–200.

Sklar, A. (1959): "Fonctions de répartition à n dimensions et leurs marges," *Publications de l'Institut de Statistique de L'Université de Paris*, 8, 229–231.

Tsukahara, H. (2005): "Semiparametric estimation in copula models," *The Canadian Journal of Statistics*, 33, 357–375.