Calibration Estimation of Semiparametric Copula Models with Data Missing at Random

Shigeyuki Hamori* Kaiji Motegi† Zheng Zhang‡

March 28, 2018

Abstract

This paper investigates the estimation of semiparametric copula models with data missing at random. The two-step maximum likelihood estimation of Genest, Ghoudi, and Rivest (1995) is infeasible due to the presence of missing data. We propose a class of calibration estimators for the nonparametric marginal distributions and the copula parameters of interest by balancing the empirical moments of covariates between non-missing and complete groups. Our proposed estimators do not require the estimation of missing mechanism, and enjoy stable performance even when sample size is small. We prove that our estimators satisfy consistency and asymptotic normality. We also provide a consistent estimator for the asymptotic variance. Simulation results highlight the dominance of our proposed method relative to existing alternatives.

Keywords: Semiparametric copula model; Missing at random; Covariate balancing.

*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`

1 Introduction

Copula models are a compelling tool for analyzing complex interdependence of multiple variables. A key characteristic of copula models is that, as Sklar (1959) proved, any multivariate joint distribution can be recovered by inputting univariate marginal distributions to a correct copula function. The copula approach is capable of capturing a wide range of interdependence among variables with relatively small computational burden.[1] There is a vast and growing literature applying copula models to economic and financial data.[2]

A popular class of copula models is *semiparametric models*, which comprise nonparametric marginal distributions and parametric copula functions. Genest, Ghoudi, and Rivest (1995) proposes the widely used *two-step maximum likelihood estimator* for the copula parameter.[3]

Most papers in the copula literature, including Genest, Ghoudi, and Rivest (1995), assume complete data. It remains unclear how to run copula models when there are missing data. Indeed, missing data frequently appear in a broad range of empirical research. In survey analysis, for example, respondents may refuse to report their personal information such as age, education, gender, race, salary, and weight. In financial econometrics, missing data are a perverse phenomenon since different countries have different holidays. There may also be unexpected market closures due to circuit breakers, technical maintenance, or terrorist attacks.

A primitive way of dealing with missing data is *listwise deletion*, which picks individuals with complete data and treats them all equally. The listwise deletion delivers consistent inference if data are *missing completely at random* (MCAR), where target variables $\boldsymbol{Y}_i$ and their missing status $\boldsymbol{T}_i$ are independent of each other. In practice the MCAR condition is

[1] See Trivedi and Zimmer (2007) for a general overview of copula models.

[2] Recent applications using copula models include Aloui, Aïssa, and Nguyen (2013), Oh and Patton (2013), Salvatierra and Patton (2015), Marra and Wyszynski (2016), and Oh and Patton (2016, 2017a,b). See Patton (2009, 2012, 2013) and Fan and Patton (2014) for extensive surveys.

[3] Chen and Fan (2005) proposes pseudo-likelihood ratio tests for model selection. Chen and Fan (2006) studies the estimation of a class of copula-based semiparametric stationary Markov models.

often violated, and in such a case the listwise deletion can deliver heavily biased estimators. It is thus desired to work under a more general assumption called *missing at random* (MAR), originally put forward by Rubin (1976), where $\boldsymbol{Y}_i$ and $\boldsymbol{T}_i$ are independent of each other given some observed covariates $\boldsymbol{X}_i$.

Inverse probability weighting, a popular approach of handling MAR data, typically requires a construction of propensity score function (cf. Horvitz and Thompson, 1952, Zhao and Lipsitz, 1992, Hirano, Imbens, and Ridder, 2003, Imbens, Newey, and Ridder, 2005, Chen, Hong, and Tarozzi, 2008). Propensity score functions are unknown in practice and need to be estimated either parametrically or non-parametrically. A major drawback of the parametric approach is that estimators may have severe bias if propensity score functions are misspecified. The non-parametric approach, such as kernel regression, is free of misspecification and hence more robust than the parametric approach. The nonparametric estimation, however, often has a poor finite sample performance due to extreme weights across individuals.

In the literature of causal inference with binary treatments, Chan, Yam, and Zhang (2016) recently proposed a novel estimation technique that is relevant to missing data analysis. They construct a class of nonparametric calibration weights by balancing the moments of covariates among treated, controlled, and combined groups. Their method bypasses an explicit specification of a propensity score function. Moreover, calibration weights satisfy certain moment constraints in both finite sample and large sample so that extreme weights are unlikely to occur. As a result, the calibration estimation attains significantly better finite sample performance than other nonparametric approximation methods.

As is well known, causal inference with binary treatments can be regarded as a specific form of missing data problems since we can observe one and only one of potential outcomes. Being motivated by such an intimate connection, we extend the two-step maximum likelihood approach of Genest, Ghoudi, and Rivest (1995) by adapting the calibration procedure of Chan, Yam, and Zhang (2016) in order to study semiparametric copula models with data missing at random. Under the *i.i.d.* condition, our estimator satisfies consistency and asymptotic normality.[4] We also present a consistent estimator for the asymptotic variance of our estimator.

Another contribution of this article lies in our simulation design. We perform Monte Carlo simulations in a way that ensures the MAR condition and the tractability of the true unconditional distribution of target variables. A key step for meeting those two conditions at the same time is that we draw target variables $\boldsymbol{Y}_i$ and covariates $\boldsymbol{X}_i$ jointly from Clayton or Gumbel copulas. Then $\boldsymbol{Y}_i$ and $\boldsymbol{X}_i$ are related with each other as the MAR condition requires, and the unconditional distribution of $\boldsymbol{Y}_i$ is again Clayton or Gumbel with the same parameter. To our best knowledge, there is no existing literature discussing how to simulate data with a given copula structure when data are missing at random.

The simulation results highlight the dominance of our proposed estimator relative to existing alternatives. First, the listwise deletion leads to severe bias under the MAR condition. In particular, we reveal that there is a positive bias under the Clayton copula and a negative bias under the Gumbel copula if a missing mechanism is specified via a logistic function. We provide a precise reason for those facts for the first time in the literature. Second, the parametric approach of estimating the propensity score suffers from substantial bias whenever the propensity score model is misspecified. Third, the nonparametric approach of Hirano, Imbens, and Ridder (2003) exhibits serious instability due to frequent appearances of extreme weights. Our proposed estimator achieves a remarkably sharp and robust performance compared with the other methods.

The remainder of this paper is organized as follows. In Section 2 we explain our notations and basic set-up. In Section 3 we propose our estimator and study its large sample properties. In Section 4 we present a nonparametric consistent estimator for the asymptotic variance of our estimator. In Section 5 we perform Monte Carlo simulations. In Section 6 we provide some concluding remarks. Details of notations and assumptions are presented in Technical Appendices. Proofs of propositions, theorems, and lemmas are collected in the

[4] The *i.i.d.* assumption is admittedly a restrictive one that rules out time series applications. Non-*i.i.d.* data with missing observations, however, are a notoriously challenging problem.

supplemental material Hamori, Motegi, and Zhang (2018).

2 Notations and Basic Framework

Let $d \geq 2$ be a fixed positive integer that signifies the dimension of target variables $\boldsymbol{Y}_i$. Suppose that $\{\boldsymbol{Y}_i = (Y_{1i}, \ldots, Y_{di})^\top\}_{i=1}^N$ are *i.i.d.* random vectors following the distribution $F^0(y_1, ..., y_d)$. The marginal distributions of $F^0(y_1, ..., y_d)$, denoted by $\{F_j^0, j = 1, ..., d\}$, are assumed to be continuous and differentiable. Sklar's (1959) characterization theorem ensures the existence of a unique copula C^0 such that $F^0(y_1, ..., y_d) = C^0(F_1^0(y_1), ..., F_d^0(y_d))$. We assume copula function $C^0(u_1, ..., u_d)$ has continuous partial derivatives, then

$$f^0(y_1, ..., y_d) = c^0(F_1^0(y_1), \cdots, F_d^0(y_d)) \prod_{j=1}^{d} f_j^0(y_j) \ , \tag{2.1}$$

where f^0, f_j^0, and c^0 are the density functions of F^0, F_j^0, and C^0, respectively.

Estimation of copula models has been studied extensively. In particular, Genest, Ghoudi, and Rivest (1995) pioneered the estimation of semiparametric copula models, where the copula function belongs to a parametric family (i.e. $C^0(F_1^0(y_1), ..., F_d^0(y_d)) = C(F_1^0(y_1), ..., F_d^0(y_d); \theta_0)$ for some $\theta_0 \in \mathbb{R}^p$), while the marginal distributions $\{F_j^0\}_{j=1}^d$ are left unknown.[5] They proposed the widely used *two-step maximum likelihood estimator* for the target parameter θ_0:

$$\tilde{\theta} = \arg\max_{\theta \in \Theta} \left\{ \frac{1}{N} \sum_{i=1}^{N} \log c(\tilde{F}_1(Y_{1i}), ..., \tilde{F}_d(Y_{di}); \theta) \right\} \ , \tag{2.2}$$

where $c(u_1, ..., u_d; \theta)$ is the density of $C(u_1, ..., u_d; \theta)$, Θ is a compact set of $\mathbb{R}^p$ containing the true value θ_0, and $\tilde{F}_j(y) = (N+1)^{-1} \sum_{i=1}^N I(Y_{ji} \leq y)$ is a rescaled empirical marginal distribution.

The existing literature on copula models, including Genest, Ghoudi, and Rivest (1995),

[5] See also Oakes (1994), Shih and Louis (1995), and Chen and Fan (2005, 2006) for more results on semiparametric copula models.

assumes complete data. That is a strong assumption since missing data can arise in virtually any field of application. A main goal of this paper is to generalize the two-step maximum likelihood estimator in (2.2) to deal with missing data.

Let $\boldsymbol{T}_i = (T_{1i}, \ldots, T_{di})^\top \in \{0,1\}^d$ be a binary random vector indicating the missing status of the i^{th} individual, namely,

$$T_{ji} = 0 \text{ if } Y_{ji} \text{ is missing}; \qquad T_{ji} = 1 \text{ if } Y_{ji} \text{ is observed}.$$

If $\boldsymbol{T}_i$ and $\boldsymbol{Y}_i$ are independent of each other, then it is called *missing completely at random* (MCAR). Under the MCAR condition, an elementary approach of *listwise deletion*, which merely picks individuals with complete observations and puts equal weights on them, is well known to deliver consistent inference. The MCAR condition, however, is an unrealistically strong assumption that is violated in many applications.

In this paper we impose a more realistic assumption called *missing at random* (MAR), which was put forward by Rubin (1976). Let $\boldsymbol{X}_i = (X_{1i}, \ldots, X_{ri})^\top$ be a vector of covariates that are observable for all individuals $i \in \{1, \ldots, N\}$. The MAR condition assumes that $\boldsymbol{T}_i$ and $\boldsymbol{Y}_i$ are independent of each other *given* $\boldsymbol{X}_i$.

Assumption 2.1 (Missing at Random)**.** $\{T_{1i}, ..., T_{di}\} \perp \{Y_{1i}, ..., Y_{di}\} | \boldsymbol{X}_i$ *for any* $i \in \{1, \ldots, N\}$.

The MAR condition has been popularly used in econometrics and statistics to identify the parameter of interest (see e.g. Robins and Rotnitzky, 1995, Little and Rubin, 2002, Chen, Hong, and Tarozzi, 2008, Tan, 2011). The MAR condition does not require the *unconditional* independence between $\boldsymbol{T}_i$ and $\boldsymbol{Y}_i$. In many applications $\boldsymbol{T}_i$ and $\boldsymbol{Y}_i$ are unconditionally correlated with each other through $\boldsymbol{X}_i$, and that violates MCAR but not MAR.

3 Weighted Two-Step Estimation

We assume throughout the paper that the true copula parameter θ_0 is a unique solution to

$$\theta_0 = \arg\max_{\theta\in\Theta} \mathbb{E}\left[\log c(F_1^0(Y_{1i}), ..., F_d^0(Y_{di}); \theta)\right] \ .$$

Using Assumption 2.1 and the law of iterated expectations, we can express θ_0 as follows:

$$\begin{aligned}
\theta_0 &= \arg\max_{\theta\in\Theta} \mathbb{E}\left[\mathbb{E}\left[\log c(F_1^0(Y_{1i}), ..., F_d^0(Y_{di}); \theta)\middle|\boldsymbol{X}_i\right]\right] \\
&= \arg\max_{\theta\in\Theta} \mathbb{E}\left[\mathbb{E}\left[\log c(F_1^0(Y_{1i}), ..., F_d^0(Y_{di}); \theta)\middle|\boldsymbol{X}_i\right] \cdot \mathbb{E}\left[\frac{1}{\eta(\boldsymbol{X}_i)} I(T_{1i}=1, ..., T_{di}=1)\middle|\boldsymbol{X}_i\right]\right] \\
&= \arg\max_{\theta\in\Theta} \mathbb{E}\left[\mathbb{E}\left[\frac{1}{\eta(\boldsymbol{X}_i)} I(T_{1i}=1, ..., T_{di}=1) \log c(F_1^0(Y_{1i}), ..., F_d^0(Y_{di}); \theta)\middle|\boldsymbol{X}_i\right]\right] \\
&= \arg\max_{\theta\in\Theta} \mathbb{E}\left[\frac{1}{\eta(\boldsymbol{X}_i)} I(T_{1i}=1, ..., T_{di}=1) \log c(F_1^0(Y_{1i}), ..., F_d^0(Y_{di}); \theta)\right] \ , \qquad (3.1)
\end{aligned}$$

where $\eta(\boldsymbol{X}_i) = \mathbb{P}(T_{1i}=1, ..., T_{di}=1|\boldsymbol{X}_i)$ is called a *propensity score function.*

In view of (3.1), we can propose the *weighted two step maximum likelihood estimator* for θ_0 as follows:

Step 1 Estimate the marginal distributions $\{F_j\}_{j=1}^d$, denoted by $\{\hat{F}_j\}_{j=1}^d$.

Step 2 Estimate the inverse probability weights $(N\eta(\boldsymbol{X}))^{-1}$, denoted by $\hat{q}(\boldsymbol{X})$, and compute $\hat{\theta}$ via a sample version of (3.1):

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \sum_{i=1}^{N} \hat{q}(\boldsymbol{X}_i) I(T_{1i}=1, ..., T_{di}=1) \log c(\hat{F}_1(Y_{1i}), ..., \hat{F}_d(Y_{di}); \theta) \ .$$

Step 1 is elaborated in Section 3.1, where we present a class of calibration estimators for the marginal distributions $\{F_j\}_{j=1}^d$. Step 2 is elaborated in Section 3.2.

3.1 Estimation of Marginal Distributions

3.1.1 Existing Estimation

Under Assumption 2.1, for $j \in \{1, ..., d\}$, marginal distribution F_j^0 can be represented by

$$\begin{aligned} F_j^0(y) &= \mathbb{E}[I(Y_{ji} \leq y)] = \mathbb{E}\left[\mathbb{E}[I(Y_{ji} \leq y)|\boldsymbol{X}_i]\right] = \mathbb{E}\left[\mathbb{E}[I(Y_{ji} \leq y)|\boldsymbol{X}_i] \cdot \mathbb{E}\left[\frac{T_{ji}}{\pi_j(\boldsymbol{X}_i)}\middle|\boldsymbol{X}_i\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[I(Y_{ji} \leq y) \cdot \frac{T_{ji}}{\pi_j(\boldsymbol{X}_i)}\middle|\boldsymbol{X}_i\right]\right] = \mathbb{E}\left[\frac{T_{ji}}{\pi_j(\boldsymbol{X}_i)} I(Y_{ji} \leq y)\right] , \end{aligned} \tag{3.2}$$

where $\pi_j(x) \triangleq \mathbb{P}(T_{ji} = 1|\boldsymbol{X}_i = x)$ is the *propensity score function.* If $\pi_j(x)$ were known, then it would be straightforward to estimate F_j via a sample analogue of (3.2):

$$\tilde{F}_j(y) \triangleq \frac{1}{N}\sum_{i=1}^{N} \frac{T_{ji}}{\pi_j(\boldsymbol{X}_i)} I(Y_{ji} \leq y) .$$

This estimator is known as inverse probability weighting (IPW) estimator (cf. Horvitz and Thompson, 1952). Since $\pi_j(x)$ is unknown in practice, it is typically estimated either parametrically (cf. Zhao and Lipsitz, 1992, Robins, Rotnitzky, and Zhao, 1994, Bang and Robins, 2005) or nonparametrically (cf. Hahn, 1998, Hirano, Imbens, and Ridder, 2003, Imbens, Newey, and Ridder, 2005, Chen, Hong, and Tarozzi, 2008). Parametric methods are easy to implement, but will lead to erroneous results if the model is misspecified. Nonparametric methods such as kernel or sieve regression offer asymptotically robust estimators since they do not require the model assumption on the propensity score, but their small sample performance is notoriously poor.

3.1.2 Calibration Weighting Estimator

A key property of the propensity score $\pi_j(\boldsymbol{X})$ is that for all integrable function $u(\boldsymbol{X})$:

$$\mathbb{E}\left[T_{ji} \times \frac{1}{\pi_j(\boldsymbol{X}_i)} \times u(\boldsymbol{X}_i)\right] = \mathbb{E}\left[u(\boldsymbol{X}_i)\right] \ , \ j \in \{1, ..., d\} \ . \tag{3.3}$$

The propensity score π_j balances all moments of the covariates between the non-missing group and the whole group, and it is characterized by the infinite moments condition (3.3). Without an explicit estimation of the unknown propensity score functions, the calibration weights $\{\hat{p}_{jK}(\boldsymbol{X})\}_{j=1}^d$ are supposed to satisfy a sample analogy of (3.3):

$$\sum_{i=1}^N T_{ji}\hat{p}_{ji}(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i) = \frac{1}{N}\sum_{i=1}^N u_K(\boldsymbol{X}_i)\ ,\ j \in \{1, ..., d\}\ , \tag{3.4}$$

where $u_K(\boldsymbol{X}) = (u_{K1}(\boldsymbol{X}), ..., u_{KK}(\boldsymbol{X}))^\top$ is the known basis function with dimension $K \in \mathbb{N}$. The functions $u_K(\boldsymbol{X})$ are called the *approximation sieve* and can be used to approximate any suitable functions $u(\boldsymbol{X})$ arbitrarily well as $K \to \infty$ (cf. Chen, 2007). This idea was first put forward by Chan, Yam, and Zhang (2016) in the context of causal inference.

We now define $\hat{p}_{ji}(\boldsymbol{X}_i)$. Let $D(v, v_0)$ be a known distance measure that is continuously differentiable in $v \in \mathbb{R}$, non-negative, strictly convex in v, and $D(v_0, v_0) = 0$. We define the calibration weights by solving the following constrained optimization problem:

$$\begin{cases} \min \quad \sum_{i=1}^N T_{ji} D(Np_{ji}, 1)\ , \\ \text{subject to} \quad \sum_{i=1}^N T_{ji}p_{ji}u_K(\boldsymbol{X}_i) = \frac{1}{N}\sum_{i=1}^N u_K(\boldsymbol{X}_i)\ , \end{cases} \tag{3.5}$$

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

The choice of uniform design weights in (3.5) is based on a few observations. First, if there are no missing data, then we can estimate $F_j^0(y)$ by the empirical distribution $(N+1)^{-1}\sum_{i=1}^N I(Y_{ji} \le y)$, which assigns equal weights for each individual. Second, there is no need to estimate $\pi_j(x)$ when the uniform design weights are used. Third, by minimizing the aggregate distance from uniform weights, the dispersion of the resulting weights is well controlled and we can avoid extreme weights. It is well known that extreme weights cause instability in the IPW estimator when the propensity score function is misspecified.

The primal problem (3.5) is a convex separable programming with linear constraints.

The dual problem, by contrast, is an *unconstrained* convex maximization problem. The latter enhances the speed and stability of numerical optimization algorithms (cf. Tseng and Bertsekas, 1987). Hence we solve for the dual problem to compute calibration weights.

Let $f(v) = D(1-v, 1)$, and $f'(v) = \partial f(v)/\partial v$. When $T_{ji} = 1$, the dual solution of (3.5) is given by

$$\hat{p}_{jK}(\boldsymbol{X}_i) \triangleq \frac{1}{N}\rho'(\hat{\lambda}_{jK}^{\top} u_K(\boldsymbol{X}_i)) \ , \tag{3.6}$$

where ρ' is the first derivative of a strictly concave function

$$\rho(v) = f((f')^{-1}(v)) + v - v(f')^{-1}(v) \tag{3.7}$$

and $\hat{\lambda}_{jK} \in \mathbb{R}^K$ maximizes the following concave objective function

$$\hat{G}_{jK}(\lambda) \triangleq \frac{1}{N}\sum_{i=1}^{N}[T_{ji}\rho(\lambda^{\top} u_K(\boldsymbol{X}_i)) - \lambda^T u_K(\boldsymbol{X}_i)] \ . \tag{3.8}$$

In view of the first-order condition of the dual problem, it is straightforward to verify that the solution to the dual problem satisfies the linear constraints in primal problem (3.5).

The relationship between $\rho(v)$ and $f(v) = D(1-v, 1)$ is given in the supplemental material Hamori, Motegi, and Zhang (2018), where we show that the strict convexity of $D(\cdot, 1)$ is equivalent to the strict concavity of $\rho(\cdot)$. Since the primal and dual problems lead to the same solution and the latter is simpler to solve, we shall express the calibration estimator in terms of $\rho(v)$ hereafter.

The calibration weights have close connections with generalized empirical likelihood. When $\rho(v) = -\exp(-v)$, the weights are equivalent to the implied weights of exponential tilting (Kitamura and Stutzer, 1997, Imbens, Spady, and Johnson, 1998). When $\rho(v) = \log(1+v)$, the weights correspond to empirical likelihood (Owen, 1988, Qin and Lawless, 1994). When $\rho(v) = -(1-v)^2/2$, the weights are the implied weights of the continuous updating estimator of generalized method of moments (Hansen, Heaton, and Yaron, 1996) and also minimize the squared distance function. When $\rho(v) = v - \exp(-v)$, the weights are equivalent to the inverse of a logistic function.

The following result states that our calibration weights will converge to the inverse propensity score uniformly and in L^2, and also gives the convergence rates. The proof follows from Lemmas 4.1 and 4.2 in the supplemental material Hamori, Motegi, and Zhang (2018).

Proposition 3.1. *Under Assumptions B.1, B.2, B.3, B.5, B.6 listed in Technical Appendix B, we have that for* $j \in \{1, ..., d\}$

$$\sup_{x \in \mathcal{X}} |N\hat{p}_{jK}(x) - \pi_j(x)^{-1}| = O_p\left(K^{-\frac{s}{2r}+1} + \sqrt{\frac{K^3}{N}}\right) \ ,$$

and

$$\int_{\mathcal{X}} |N\hat{p}_{jK}(x) - \pi_j(x)^{-1}|^2 dF_X(x) = O_p\left(K^{-\frac{s}{r}+1} + \frac{K^2}{N}\right) \ .$$

Therefore, the calibration estimator of the marginal distribution F_j^0 is defined by

$$\hat{F}_j(y) \triangleq \sum_{i=1}^{N} T_{ji}\hat{p}_{jK}(\boldsymbol{X}_i)I(Y_{ji} \le y) \ . \tag{3.9}$$

The following result states that $\hat{F}_j$ is a $\sqrt{N}$-consistent estimator of F_j^0, and gives the asymptotic behavior of $\sqrt{N}(\hat{F}_j - F_j^0)$ which will be used later. The proof is presented in the supplemental material.

Proposition 3.2. *Under Assumptions 2.1, B.1-B.6 listed in Technical Appendix B, we have that for all* $j \in \{1, ..., d\}$,

$$\sqrt{N}\{\hat{F}_j(y) - F_j^0(y)\} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\psi_j(\boldsymbol{Y}_i, \boldsymbol{X}_i, T_{ji}; y) + o_p(1) \ , \ \forall y \in \mathbb{R} \ ,$$

where

$$\psi_j(\boldsymbol{Y}_i, \boldsymbol{X}_i, T_{ji}; y) \triangleq \frac{T_{ji}}{\pi_j(\boldsymbol{X}_i)} I(Y_{ji} \leq y) - \frac{T_{ji}}{\pi_j(\boldsymbol{X}_i)} \cdot \mathbb{E}[I(Y_{ji} \leq y)|\boldsymbol{X}_i] + \mathbb{E}[I(Y_{ji} \leq y)|\boldsymbol{X}_i] - F_j^0(y) \, .$$

3.2 Estimation of Target Parameter

In this section, we construct calibration weights which lead to consistent estimators for the inverse probability $(N\eta(\boldsymbol{X}_i))^{-1}$. We then obtain a consistent estimator for the target parameter θ_0 in accordance with (3.1).

Note that the MAR condition implies

$$\mathbb{E}\left[I(T_{1i} = 1, ..., T_{di} = 1) \times \frac{1}{\eta(\boldsymbol{X}_i)} \times u(\boldsymbol{X}_i)\right] = \mathbb{E}[u(\boldsymbol{X}_i)] \tag{3.10}$$

for all integrable function $u(\boldsymbol{X})$. Similar to the construction of calibration weights $\{\hat{p}_{jK}(\boldsymbol{X})\}$ in (3.5), we define another calibration weights $\hat{q}_K(\boldsymbol{X})$ by solving the following constraint optimization problem

$$\begin{cases} \min \quad \sum_{i=1}^N I(T_{1i} = 1, ..., T_{di} = 1) D(Nq_i, 1) \, , \\ \text{subject to} \quad \sum_{i=1}^N I(T_{1i} = 1, ..., T_{di} = 1) q_i u_K(\boldsymbol{X}_i) = \frac{1}{N} \sum_{i=1}^N u_K(\boldsymbol{X}_i) \, . \end{cases} \tag{3.11}$$

Similar to obtain (3.6), the dual solution of (3.11) is given by

$$\hat{q}_K(\boldsymbol{X}_i) = \frac{1}{N} \rho' \left(\hat{\beta}_K^\top u_K(\boldsymbol{X}_i)\right) \text{ for } i \text{ such that } T_{1i} = \cdots = T_{di} = 1 \, , \tag{3.12}$$

where $\hat{\beta}_K$ maximizes the following concave objective function:

$$\hat{H}_K(\beta) = \frac{1}{N} \sum_{i=1}^N I(T_{1i} = 1, ..., T_{di} = 1) \rho \left(\beta^\top u_K(\boldsymbol{X}_i)\right) - \frac{1}{N} \sum_{i=1}^N \beta^\top u_K(\boldsymbol{X}_i) \, . \tag{3.13}$$

Similar to Proposition 3.1, we can also derive the following result for the calibration weights

$\hat{q}_K(x)$:

Proposition 3.3. *Under Assumptions B.1, B.2, B.3, B.5, B.6 listed in Technical Appendix B, we have that for*

$$\sup_{x\in\mathcal{X}} |N\hat{q}_K(x) - \eta(x)^{-1}| = O_p\left(K^{-\frac{s}{2r}+1} + \sqrt{\frac{K^3}{N}}\right) ,$$

and

$$\int_{\mathcal{X}} \left|N\hat{q}_K(x) - \eta(x)^{-1}\right|^2 dF_X(x) = O_p\left(K^{-\frac{s}{r}+1} + \frac{K^2}{N}\right) .$$

Finally, our *weighted two-step maximum likelihood estimator* of θ_0 is defined by

$$\hat{\theta} = \arg\max_{\theta\in\Theta}\left\{\sum_{i=1}^{N} I(T_{1i}=1,...,T_{di}=1)\hat{q}_K(\boldsymbol{X}_i)\log c(\hat{F}_1(Y_{1i}),...,\hat{F}_d(Y_{di});\theta)\right\} . \quad (3.14)$$

The following theorem states the consistency and asymptotic normality of our proposed estimator. The proof is presented in the supplemental material.

Theorem 3.4. *Under Assumptions 2.1, B.1-B.6 listed in Technical Appendix B, we have*

1. $\hat{\theta} \xrightarrow{p} \theta_0$;
2. *Furthermore, if additional Assumptions B.7-B.13 hold, then*

$$\sqrt{N}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0, V_0) ,$$

where $V_0 = B^{-1}\Sigma B^{-1}$ *with* B *and* Σ *being defined in* (B.1) *and* (B.2), *respectively.*

It can easily be verified that, if there are no missing data (i.e. $\pi_j(x) = \eta(x) = 1$), then V_0 reduces to the asymptotic variance of the two step maximum likelihood estimator derived by Genest, Ghoudi, and Rivest (1995) and Chen and Fan (2005).

4 Variance Estimation

As shown in Theorem 3.4, the asymptotic variance of $\sqrt{N}(\hat{\theta}-\theta_0)$ is given by $V_0 = B^{-1}\Sigma B^{-1}$. In order to estimate V_0, it suffices to consistently estimate both B and Σ.

4.1 Estimation of B

Using Assumption 2.1, B can be rewritten as

$$B = -\mathbb{E}\left[\frac{I(T_{1i}=1,...,T_{di}=1)}{\eta(\boldsymbol{X}_i)} l_{\theta\theta}(U_{1i},...,U_{di};\theta_0)\right] ,$$

where $l_{\theta\theta}(u_1,...,u_d;\theta) = \frac{\partial^2}{\partial\theta\partial\theta'}\log c(u_1,...,u_d;\theta)$ and $U_{ji} = F_j^0(Y_{ji})$, $j \in \{1,...,d\}$. By Propositions 3.2 and 3.3, we know that $N\hat{q}_K(x)$ is a consistent estimate for $\eta^{-1}(x)$, and $\hat{F}_j$ is a consistent estimate for F_j^0. Hence we define the plug-in estimator of B by

$$\hat{B} = -\sum_{i=1}^{N} I(T_{1i}=1,...,T_{di}=1)\hat{q}_K(\boldsymbol{X}_i) l_{\theta\theta}(\hat{U}_{1i},...,\hat{U}_{di};\hat{\theta}) , \tag{4.1}$$

where $\hat{U}_{ji} \triangleq \hat{F}_j(Y_{ji}) = \sum_{s=1}^{N} T_{js}\hat{p}_{jK}(\boldsymbol{X}_s)I(Y_{js} \le Y_{ji})$.

4.2 Estimation of Σ

Under Assumption 2.1, Σ can be written as

$$\Sigma = \mathbb{E}\left[\frac{I(T_{1i}=1,...,T_{di}=1)}{\eta(\boldsymbol{X}_i)}\left(\varphi(\boldsymbol{T}_i,\boldsymbol{X}_i,\boldsymbol{U}_i;\theta_0) + \sum_{j=1}^{d} W_j(T_{ji},\boldsymbol{X}_i,U_{ji};\theta_0)\right)^2\right] , \tag{4.2}$$

where $\varphi(\boldsymbol{T}_i,\boldsymbol{X}_i,\boldsymbol{U}_i;\theta)$ and $W_j(T_{ji},\boldsymbol{X}_i,U_{ji};\theta)$ are defined in Section A of Appendix. Similar to the estimation of B, we can define the estimators of $\varphi(\boldsymbol{T}_i,\boldsymbol{X}_i,\boldsymbol{U}_i;\theta_0)$ and $W_j(T_{ji},\boldsymbol{X}_i,U_{ji};\theta_0)$ by

$$\widehat{\varphi}(\boldsymbol{T}_i,\boldsymbol{X}_i,\boldsymbol{U}_i;\theta_0) \triangleq I(T_{1i}=1,...,T_{di}=1)N\hat{q}_K(\boldsymbol{X}_i) l_{\theta}(\hat{U}_{1i},...,\hat{U}_{di};\hat{\theta})$$

$$- I(T_{1i}=1,...,T_{di}=1)N\hat{q}_K(\boldsymbol{X}_i)\cdot\widehat{\mathbb{E}}[l_\theta(U_{1i},...,U_{di};\theta_0)|\boldsymbol{X}_i]$$

$$+ \widehat{\mathbb{E}}[l_\theta(U_{1i},...,U_{di};\theta_0)|\boldsymbol{X}_i] - \widehat{\mathbb{E}}[l_\theta(U_{1i},...,U_{di};\theta_0)] \ , \tag{4.3}$$

and

$$\widehat{W}_j(T_{ji},\boldsymbol{X}_i,U_{ji};\theta_0) \triangleq \sum_{s=1}^N I(T_{1s}=1,...,T_{ds}=1)\hat{q}_K(\boldsymbol{X}_s)l_{\theta j}(\hat{U}_{1s},...,\hat{U}_{ds};\hat{\theta})\left\{\widehat{\phi}_j(T_{ji},\boldsymbol{X}_i,U_{ji};\hat{U}_{js})-\hat{U}_{js}\right\} \ , \tag{4.4}$$

where $\hat{\mathbb{E}}[l_\theta(U_{1i},...,U_{di};\theta_0)|\boldsymbol{X}_i]$ is the least square estimator of $l_\theta(\hat{U}_{1i},...,\hat{U}_{di};\hat{\theta})$ based on the basis $u_K(\boldsymbol{X})$:

$$\widehat{\mathbb{E}}[l_\theta(U_{1i},...,U_{di};\theta_0)|\boldsymbol{X}_i] \triangleq \left\{\sum_{i=1}^N I(T_{1i}=1,...,T_{di}=1)l_\theta(\hat{U}_{1i},...,\hat{U}_{di};\hat{\theta})u_K(\boldsymbol{X}_i)\right\}^\top \cdot\left\{\sum_{i=1}^N I(T_{1i}=1,...,T_{di}=1)u_K(\boldsymbol{X}_i)u_K^\top(\boldsymbol{X}_i)\right\}^{-1} u_K(\boldsymbol{X}_i) \ ,$$

and

$$\widehat{\mathbb{E}}[l_\theta(U_{1i},...,U_{di};\theta_0)] \triangleq \sum_{i=1}^N I(T_{1i}=1,...,T_{di}=1)\hat{q}_K(\boldsymbol{X}_i)l_\theta(\hat{U}_{1i},...,\hat{U}_{di};\hat{\theta}) \ ,$$

and

$$\widehat{\phi}_j(T_{ji},\boldsymbol{X}_i,U_{ji};v) \triangleq T_{ji}\{N\hat{p}_{jK}(\boldsymbol{X}_i)\}I(\hat{U}_{ji}\le v) - T_{ji}\{N\hat{p}_{jK}(\boldsymbol{X}_i)\}\cdot\widehat{\mathbb{E}}[I(U_{ji}\le v)|\boldsymbol{X}_i,T_{ji}=1] + \widehat{\mathbb{E}}[I(U_{ji}\le v)|\boldsymbol{X}_i,T_{ji}=1] \ ,$$

and

$$\widehat{\mathbb{E}}[I(U_{ji}\le v)|\boldsymbol{X}_i,T_{ji}=1] = \left\{\sum_{i=1}^N I(T_{ji}=1)I(U_{ji}\le v)u_K(\boldsymbol{X}_i)\right\}^\top\left\{\sum_{i=1}^N I(T_{ji}=1)u_K(\boldsymbol{X}_i)u_K^\top(\boldsymbol{X}_i)\right\}^{-1} u_K(\boldsymbol{X}_i) \ . \tag{4.5}$$

Finally, the estimates of the asymptotic variance are defined by

$$\widehat{V} \triangleq \widehat{B}^{-1}\widehat{\Sigma}\widehat{B}^{-1} \ .$$

The following result states that our proposed variance estimator is consistent, and the proof is left to the supplemental material.

Theorem 4.1. *Under Assumptions 2.1, B.1-B.13 listed in Appendix,* $\|\widehat{V} - V_0\| \xrightarrow{p} 0$.

5 Monte Carlo Simulations

In this section, we perform Monte Carlo simulations in order to evaluate the finite sample properties of the calibration estimator and other existing estimators. See Section 5.1 for a simulation design and Section 5.2 for results.

5.1 Simulation Design

5.1.1 Data Generating Processes

Suppose that $\boldsymbol{Y}_i = [Y_{1i}, Y_{2i}]^\top$ are bivariate target variables (i.e. $d = 2$) and X_i is a scalar covariate (i.e. $r = 1$). We specify the joint distribution of $\boldsymbol{Z}_i = [\boldsymbol{Y}_i^\top, X_i]^\top$ via two Archimedean copulas that are widely used in empirical applications. The first one is the trivariate Clayton copula with a scalar parameter α_0, written as $C_3(\alpha_0)$. The cumulative distribution function of $C_3(\alpha_0)$ is given by

$$C(u_1, u_2, u_3; \alpha_0) = \left(\sum_{j=1}^{k} u_j^{-\alpha_0} - k + 1 \right)^{-1/\alpha_0}, \quad \alpha_0 > 0,$$

where $k = 3$ is the dimension of the copula. The inputs are $u_1 = F_1(y_1)$, $u_2 = F_2(y_2)$, and $u_3 = F_X(x)$, where $F_j(\cdot)$ is the marginal distribution function of Y_{ji} and $F_X(\cdot)$ is the marginal distribution function of X_i. We use the standard Gaussian distribution for F_1, F_2, and F_X.

Since the standard Gaussian distribution has a tractable inverse distribution function, it is straightforward to draw $\{Y_{1i}, Y_{2i}, X_i\}$ by first generating (U_{1i}, U_{2i}, U_{3i}) from the copula and then transforming them to $Y_{1i} = F_1^{-1}(U_{1i})$, $Y_{2i} = F_2^{-1}(U_{2i})$, and $X_i = F_X^{-1}(U_{3i})$.

The second copula is the trivariate Gumbel copula with a scalar parameter γ_0, written as $G_3(\gamma_0)$. The cumulative distribution function of $G_3(\gamma_0)$ is given by

$$C(u_1, u_2, u_3;\, \gamma_0) = \exp\left[-\left\{\sum_{j=1}^{k}(-\log u_j)^{\gamma_0}\right\}^{1/\gamma_0}\right], \quad \gamma_0 > 1.$$

As in the Clayton case, we assume that the marginal distributions of Y_{1i}, Y_{2i}, and X_i are standard Gaussian.

As implied by Genest, Nešlehová, and Ben Ghorbal (2011, Examples 1 and 2), Kendall's τ is given by $\tau = \alpha_0/(\alpha_0+2)$ for $C_3(\alpha_0)$ and $\tau = 1-1/\gamma_0$ for $G_3(\gamma_0)$. We consider two cases that $\tau \in \{0.4, 0.7\}$. Hence we set the true copula parameters to be $(\alpha_0, \gamma_0) = (1.333, 1.667)$ for $\tau = 0.4$ and $(\alpha_0, \gamma_0) = (4.667, 3.333)$ for $\tau = 0.7$.

We next specify missing mechanisms. Assume for simplicity that $\{Y_{1i}\}$ is always observed and only $\{Y_{2i}\}$ can be missing. Specifically, suppose that $\mathbb{P}(T_{1i} = 1 \,|\, X_i = x_i) = 1$ and

$$\mathbb{P}(T_{2i} = 1 \,|\, X_i = x_i) = \frac{1}{1+\exp[a+bx_i]}. \tag{5.1}$$

It is common in the missing data literature to use the logistic function to specify missing probability (see e.g. Qin, Leung, and Shao, 2002).

To choose the key parameters (a, b), note that

$$\mu \equiv \mathbb{E}[T_{2i}] = \mathbb{E}[\mathbb{E}[T_{2i} \,|\, X_i = x_i]] = \mathbb{E}[\mathbb{P}(T_{1i} = 1 \,|\, X_i = x_i)] = \mathbb{E}\left[(1+\exp[a+bx_i])^{-1}\right]$$

and

$$\sigma^2 \equiv \mathbb{V}[T_{2i}] = \mathbb{E}[T_{2i}^2] - \mu^2 = \mathbb{E}[T_{2i}] - \mu^2 = \mu - \mu^2.$$

μ measures the marginal probability of observing Y_{2i}, while σ^2 measures the marginal variance of T_{2i}. Analytical solutions of μ and σ^2 are rather hard to derive, but it is straightforward to approximate them numerically by $\hat{\mu}_N = (1/N)\sum_{i=1}^{N}(1+\exp[a+bX_i])^{-1}$ and $\hat{\sigma}_N^2 = \hat{\mu}_N - \hat{\mu}_N^2$. Hence, given a feasible pair of target quantities (μ, σ^2), we can find a corresponding pair of parameters (a, b) via a numerical search.[6] We consider four cases:

Case A $(a, b) = (-1.385, 0.000)$ so that $(\mu, \sigma^2) = (0.800, 0.160)$.

Case B $(a, b) = (-1.430, 0.400)$ so that $(\mu, \sigma^2) = (0.800, 0.160)$.

Case C $(a, b) = (-0.405, 0.000)$ so that $(\mu, \sigma^2) = (0.600, 0.240)$.

Case D $(a, b) = (-0.420, 0.400)$ so that $(\mu, \sigma^2) = (0.600, 0.240)$.

$(\mu, \sigma^2) = (0.800, 0.160)$ for both Case A and Case B so that 20% of $\{Y_{2i}\}$ are missing on average. The crucial difference between them is that $b = 0$ (i.e. MCAR) in Case A while $b \neq 0$ (i.e. MAR) in Case B. Under MCAR, the missing probability of Y_{2i} does not depend on X_i although $\boldsymbol{Y}_i$ is (nonlinearly) related with X_i via the copula. Under MAR, the missing probability of Y_{2i} does depend on X_i and $\boldsymbol{Y}_i$ is related with X_i via the copula.

Similar structures apply for Cases C and D with a larger missing probability. $(\mu, \sigma^2) = (0.600, 0.240)$ so that 40% of $\{Y_{2i}\}$ are missing on average. The missing mechanism is MCAR in Case C and MAR in Case D. The latter case is expected to be the most challenging case for estimation, since we have relatively many missing data with the MAR mechanism.

We draw $J = 1000$ Monte Carlo samples with sample size $N \in \{250, 500, 1000\}$. The procedure is summarized as follows. First, draw (U_{1i}, U_{2i}, U_{3i}) from the Clayton or Gumbel copula. Second, transform them to $Y_{1i} = F_1^{-1}(U_{1i})$, $Y_{2i} = F_2^{-1}(U_{2i})$, and $X_i = F_X^{-1}(U_{3i})$. Third, make some of $\{Y_{2i}\}$ missing according to Cases A-D.

[6] Note that some pairs of (μ, σ^2) are infeasible since there exists a (nonlinear) dependence between μ and σ^2. Having $(\mu, \sigma^2) = (0.9, 0.2)$, for example, is infeasible since $\mu = 0.9$ implies that 90% of $\{T_{21}, \dots, T_{2N}\}$ are 1 and 10% are 0 so that the corresponding value of σ^2 must be much smaller than 0.2 (i.e. approximately 0.09).

5.1.2 Estimation

#1. Listwise Deletion The first approach is semiparametric estimation with listwise deletion. For each component $j \in \{1, 2\}$, we estimate the marginal distribution by $\hat{F}_j(y) = (N^* + 1)^{-1} \sum_{i=1}^{N} I(T_{1i} = 1, T_{2i} = 1) I(Y_{ji} < y)$, where $N^* = \sum_{i=1}^{N} I(T_{1i} = 1, T_{2i} = 1)$ is the number of individuals with complete data. We then compute the maximum likelihood estimator for the copula parameter based on the complete data. Taking the Clayton copula as an example, the maximum likelihood estimator is defined as follows. (The Gumbel case can be treated analogously.)

$$\hat{\alpha} = \arg \max_{\alpha \in (0,\infty)} \left\{ \sum_{i=1}^{N} I(T_{1i} = 1, T_{2i} = 1) \log c_2 \left(\hat{F}_1(Y_{1i}), \hat{F}_2(Y_{2i}); \alpha \right) \right\}, \tag{5.2}$$

where

$$c_2(u_1, u_2; \alpha) = (1 + \alpha)(u_1 u_2)^{-\alpha-1}(u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-\frac{1}{\alpha}-2}$$

is the probability density function of the bivariate Clayton copula $C_2(\alpha)$. There is not a misspecification problem in the second step since there is a well-known property that the bivariate marginal distribution of $C_3(\alpha_0)$ is indeed $C_2(\alpha_0)$.

The listwise deletion has two major characteristics. First, prior to the first step, Y_{1i} is discarded whenever Y_{2i} is missing. Second and more importantly, equal weights are assigned for all N^* individuals with complete data, and the information of X_i is not used at all (i.e. $p_2(X_i) = q(X_i) = 1/N^*$). The latter feature causes bias in the resulting estimator when data are MAR.

#2. Parametric Estimation The second approach estimates the propensity score function $\pi_2(x) = \mathbb{P}(T_{2i} = 1 | X_i = x)$ parametrically. Define

$$\pi_2(x; a, b) = \frac{1}{1 + \exp(a + bx)}, \tag{5.3}$$

then the log-likelihood function of $\{T_{2i}, X_i\}_{i=1}^N$ is given by

$$l(a,b) = \sum_{i=1}^{N} \left[T_{2i} \log \pi_2(X_i; a, b) + (1 - T_{2i}) \log \left(1 - \pi_2(X_i; a, b)\right) \right].$$

We compute the maximum likelihood estimator $(\hat{a}, \hat{b})$ that maximizes $l(a, b)$, and then calculate $\hat{p}_2(X_i) = \hat{q}(X_i) = [N \times \pi_2(X_i; \hat{a}, \hat{b})]^{-1}$. Marginal distributions are estimated as

$$\hat{F}_j(y) = \sum_{i=1}^{N} I(T_{ji} = 1)\hat{p}_j(X_i) I(Y_{ji} < y) \tag{5.4}$$

and the copula parameter is estimated as

$$\hat{\alpha} = \arg \max_{\alpha \in (0,\infty)} \left\{ \sum_{i=1}^{N} I(T_{1i} = 1, T_{2i} = 1)\hat{q}(X_i) \log c_2 \left(\hat{F}_1(Y_{1i}), \hat{F}_2(Y_{2i});\, \alpha \right) \right\}. \tag{5.5}$$

Note that (5.3) is correctly specified relative to the true propensity score function (5.1). For comparison, we also use a misspecified model

$$\pi_2(x; a, b) = \frac{1}{1 + \exp(bx)}. \tag{5.6}$$

Model (5.6) is misspecified since $a \neq 0$ in each of Cases A-D. We are supposed to have consistent estimators when (5.3) is used and inconsistent estimators when (5.6) is used.

#3. Nonparametric Estimation The third approach estimates the propensity score function $\pi_2(x) = \mathbb{P}(T_{2i} = 1 | X_i = x)$ non-parametrically based on Hirano, Imbens, and Ridder (2003). Define the logistic function $h(v) = [1 + \exp(-v)]^{-1}$ and $\pi_{2K}(X_i; \lambda) = h(\lambda^\top u_K(X_i))$, where $u_K(X_i)$ is the approximation sieve also used in the calibration estimation. The log-likelihood function of $\{T_{2i}, X_i\}_{i=1}^N$ is written as

$$l(\lambda) = \sum_{i=1}^{N} \left[T_{2i} \log \pi_{2K}(X_i; \lambda) + (1 - T_{2i}) \log \left(1 - \pi_{2K}(X_i; \lambda)\right) \right].$$

Compute the maximum likelihood estimator $\hat{\lambda}$ that maximizes $l(\lambda)$, and calculate $\hat{p}_{2K}(X_i) = \hat{q}_K(X_i) = [N \times \pi_{2K}(X_i; \hat{\lambda})]^{-1}$. Then use (5.4) and (5.5) to complete the procedure.

The difference between the parametric and nonparametric approaches is that the former requires an explicit specification of propensity score functions while the latter does not. The nonparametric approach, however, requires a selection of K (i.e. the dimension of the approximation sieve). We use $u_K(X_i) = [1, X_i, X_i^2]^\top$ (i.e. $K = 3$) and $u_K(X_i) = [1, X_i, X_i^2, X_i^3]^\top$ (i.e. $K = 4$) in order to see how results change across different values of K.

#4. Calibration Estimation The fourth approach is our proposed calibration estimation. For each component j, we estimate the marginal distribution by (5.4), where the computation of $\hat{p}_{jK}(X_i)$ is detailed in Section 3.1.2. We then compute the maximum likelihood estimator for the copula parameter from (5.5), where the computation of $\hat{q}_K(X_i)$ is detailed in Section 3.2. As in the nonparametric approach, we use $u_K(X_i) = [1, X_i, X_i^2]^\top$ (i.e. $K = 3$) and $u_K(X_i) = [1, X_i, X_i^2, X_i^3]^\top$ (i.e. $K = 4$).

To conclude this section, it is worth emphasizing a clever aspect of our simulation design. $\boldsymbol{Y}_i$ and X_i, on one hand, should be associated with each other since we are interested in data missing at random. The unconditional joint distribution of $\boldsymbol{Y}_i$, on the other hand, should be known and tractable so that we can perform well-defined Monte Carlo experiments. To our best knowledge, the only way of meeting those two requirements simultaneously is to draw $\boldsymbol{Z}_i = [\boldsymbol{Y}_i^\top, X_i]^\top$ jointly from a trivariate copula whose bivariate marginal distribution is known and tractable. The Clayton and Gumbel copulas lie in such a useful class.[7] $\boldsymbol{Y}_i$ and X_i are associated with each other by construction, and the unconditional distribution of $\boldsymbol{Y}_i$ is characterized by the same copula with the same parameter. The same procedure can be applied whatever the dimensions of target variables and covariates are. Since the present paper is the earliest work on copula models with data missing at random, our simulation design itself is an innovation to the literature.

[7] The Frank copula shares the same property although this paper does not cover it to save space.

5.2 Simulation Results

5.2.1 Preliminary Results

Before discussing complete results with $J = 1000$ Monte Carlo samples, it is instructive to pick a representative Monte Carlo sample in order to explain why the listwise deletion fails and the calibration estimation succeeds.[8] We first focus on the Clayton copula with $\alpha_0 = 4.667$, Case D (MAR), and $N = 500$. See Figure 1 for a scatter plot of (Y_{1i}, Y_{2i}) in a Monte Carlo sample. In this specific sample, Y_{2i} is observed for 291 individuals and missing for 209 individuals. When the listwise deletion is executed, we get $\hat{\alpha} = 5.085$ so that there is a substantial *positive* bias of $5.085 - 4.667 = 0.418$. When the calibration estimation with $K = 4$ is executed, we get $\hat{\alpha} = 4.625$ so that there is a sufficiently small bias of $4.625 - 4.667 = -0.042$.

We provide an intuitive reason why the listwise deletion results in the *positive* bias under the Clayton copula. Recall that (Y_{1i}, Y_{2i}, X_i) are jointly drawn from the Clayton copula, which has a lower-tail dependence and upper-tail independence. A small value of X_i, therefore, 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}). In view of (5.1), the smaller (larger) X_i implies the higher (lower) probability of observing Y_{2i} under Case D. In fact, we can see in Figure 1 that (Y_{1i}, Y_{2i}) having large negative values are often observed whereas (Y_{1i}, Y_{2i}) having large positive values are often missing. As a result, the observed pairs exhibit a deceivingly strong association.

The listwise deletion literally accepts the strong association observed, since it just puts uniform weights $\hat{q}(X_i) = 1/291 = 0.003$ for all individuals with complete data (see Panel (a) of Figure 2). This is a precise reason why the listwise deletion *over-estimates* the association between Y_{1i} and Y_{2i}. The calibration estimation, by contrast, puts non-uniform weights $\hat{q}(X_i)$ across the individuals. Panel (b) of Figure 2 indeed indicates that the weights

[8] To save space, we do not discuss the parametric and nonparametric approaches until Section 5.2.2.

are generally increasing in X_i so that the weak association between Y_{1i} and Y_{2i} at the upper tail is more emphasized and the strong association at the lower tail is less emphasized. That adjustment leads to an unbiased estimator as desired.

We now replace the Clayton copula with the Gumbel copula with $\gamma_0 = 3.333$, keeping the other settings the same. See Figure 3 for a scatter plot of (Y_{1i}, Y_{2i}) given a Monte Carlo sample. In this specific sample, Y_{2i} is observed for 317 individuals and missing for 183 individuals. When the listwise deletion is executed, we get $\hat{\gamma} = 3.068$ so that there is a substantial *negative* bias of -0.265. When the calibration estimation is executed, we get $\hat{\gamma} = 3.272$ so that there is a sufficiently small bias of -0.061.

Repeating a similar logic as in the Clayton case, it is straightforward to see why the listwise deletion results in the *negative* bias under the Gumbel copula. The Gumbel copula has a lower-tail independence and upper-tail dependence. A large value of X_i, therefore, tends to be accompanied by jointly large values of (Y_{1i}, Y_{2i}) whereas a small value of X_i is not necessarily accompanied by jointly small values of (Y_{1i}, Y_{2i}). Given the missing mechanism (5.1), (Y_{1i}, Y_{2i}) having large negative values are often observed whereas (Y_{1i}, Y_{2i}) having large positive values are often missing (Figure 3). As a result, the observed pairs exhibit a deceivingly weak association. The listwise deletion literally accepts the weak association observed so that it *under-estimates* the association between Y_{1i} and Y_{2i}. The calibration estimation, by contrast, leads to an unbiased estimator by putting non-uniform weights that are generally increasing in X_i.[9]

In summary, it is a new finding that the listwise deletion results in positive bias under Clayton copulas and negative bias under Gumbel copulas given the logistic missing probability. We have provided the exact reason for those facts and shown that the calibration estimation successfully balances the covariates among observed, missing, and whole groups.

[9] A figure that displays the weights of the listwise deletion and the calibration estimation is omitted to save space, but available upon request.

5.2.2 Complete Results

We now report the bias, standard deviation, and root mean squared error (RMSE) of each estimator given the full Monte Carlo samples. See Table 1 for the Clayton copula with $\alpha_0 = 1.333$; Table 2 for Clayton with $\alpha_0 = 4.667$; Table 3 for Gumbel with $\gamma_0 = 1.667$; Table 4 for Gumbel with $\gamma_0 = 3.333$.

First, the listwise deletion leads to diminishing bias in Cases A and C (MCAR) and substantial bias in Cases B and D (MAR). In the latter cases, we observe positive bias under the Clayton copula and negative bias under the Gumbel copula. The magnitude of bias is larger in Case D than in Case B since there exist more missing data in Case D. All those results are reasonable and consistent with Section 5.2.1.

Second, the parametric estimator with the correctly specified model performs well as expected. When there is the weaker association between the target variables, the parametric estimator produces negligibly small bias even for the smallest sample size of $N = 250$, and the bias gets even smaller as sample size grows up (see Tables 1 and 3). Bias under the stronger association is larger than the bias under the weaker association, but it is still diminishing as sample size grows up (see Tables 2 and 4). See, for example, Case D in Table 4. The bias is -0.188 for $N = 250$, -0.093 for $N = 500$, and -0.051 for $N = 1000$.

Third, the parametric estimator with the misspecified model always suffers from large bias. See, for example, Table 2 with $N = 1000$. The bias is -2.445 in Case A, -2.220 in Case B, -0.866 in Case C, and -0.709 in Case D. The magnitude of bias is larger in Cases A-B than in Cases C-D, since there is more serious misspecification in the former cases. To see that, recall from (5.6) that the misspecified model imposes $a = 0$ while the true values of a are -1.385 in Case A, -1.430 in Case B, -0.405 in Case C, and -0.420 in Case D.

Fourth, the nonparametric estimator of Hirano, Imbens, and Ridder (2003) often has substantial bias for each missing mechanism and sample size. See, for instance, Table 2 with $N = 1000$. When $K = 3$, the bias is -1.082 in Case A, -0.518 in Case B, -2.803 in Case C, and -2.181 in Case D. We observe similar results for $K = 4$. Those are not surprising results since the nonparametric estimator of Hirano, Imbens, and Ridder (2003) is well known to be unstable due to the frequent occurrence of extreme weights.

Fifth and most importantly, the calibration estimator performs as well as or even better than the parametric estimator with the correctly specified model. See Case D in Table 4, for example. The bias of the calibration estimator with $K = 3$ is -0.079 for $N = 250$, -0.055 for $N = 500$, and -0.035 for $N = 1000$. For each sample size, the calibration estimator has smaller bias than the parametric estimator in absolute values. Moreover, the two estimators have similar standard deviations so that the calibration estimator dominates the parametric estimator in terms of RMSE. The same result follows if we use $K = 4$ instead of $K = 3$.

An intuitive reason why the calibration estimator is much more stable than the nonparametric estimator of Hirano, Imbens, and Ridder (2003) is that calibration weights are designed to be as close as possible to the uniform weights subject to the moment matching conditions. Overall, the calibration estimator dominates the other estimators across various copulas, sample sizes, and missing mechanisms.

6 Discussions and Conclusions

Copula models are a useful tool for capturing complex interdependence of multiple variables. A popular class of copula models is semiparametric models, which consist of nonparametric marginal distributions and parametric copula functions. Genest, Ghoudi, and Rivest (1995) proposed the two-step maximum likelihood estimator for semiparametric copula models. While there exists a vast literature on copula models, most papers including Genest, Ghoudi, and Rivest (1995) and Chen and Fan (2005) assume complete data.

In this article, we study the estimation of semiparametric copula models with data missing at random. We extend the two-step maximum likelihood approach of Genest, Ghoudi, and Rivest (1995) by adapting the calibration procedure developed in Chan, Yam, and Zhang (2016). Under the MAR condition, our estimator satisfies consistency and asymptotic normality.

We also present a consistent estimator for the asymptotic variance of our estimator.

The large sample properties of the proposed estimator established in the previous sections allow for a wide range of values for K. In practice, we need to choose an appropriate K in finite sample computation. We suggest choosing $u_K(\boldsymbol{X})$ to be the first and possibly higher moments of candidate covariates. When the covariate distributions of the missing group and the whole group differ only by a mean shift, matching the first moment of $\boldsymbol{X}$ would be sufficient for removing imbalance. When the variances differ, we can also match the second moment. Matching moments of covariate distributions are intuitive to non-statisticians. For thorough discussions of the choice of K using graphical method and cross-validation, we refer to Section 6 of Chan, Yam, and Zhang (2016).

Another contribution of this article lies in the simulation design. We provide a technique that generates data with a given copula structure while satisfying the missing at random condition. Simulation results highlight the dominance of our proposed estimator relative to the listwise deletion, parametric estimators, and nonparametric estimators based on Hirano, Imbens, and Ridder (2003). An interesting finding is that, given the logistic-type missing mechanism and the MAR condition, the listwise deletion results in positive bias under the Clayton copula and negative bias under the Gumbel copula. We have provided logical explanations for those results, and shown that our estimator removes bias by properly balancing the moments of covariates.

Acknowledgements

We thank Qu Feng, Eric Ghysels, and seminar participants at Nanyang Technological University for helpful comments. This work is supported by JSPS KAKENHI Grant Numbers 17H00983 and 16K17104. The second author is grateful for the financial supports of Japan Center for Economic Research, Kikawada Foundation, Mitsubishi UFJ Trust Scholarship Foundation, and Nomura Foundation.

References

Aloui, R., M. S. B. Aïssa, and D. K. Nguyen (2013): "Conditional Dependence Structure between Oil Prices and Exchange Rates: A Copula-GARCH Approach," *Journal of International Money and Finance*, 32, 719–738.

Bang, H., and J. M. Robins (2005): "Doubly Robust Estimation in Missing Data and Causal Inference Models," *Biometrics*, 61, 962–973.

Chan, K. C. G., S. C. P. Yam, and Z. Zhang (2016): "Globally Efficient Non-Parametric Inference of Average Treatment Effects by Empirical Balancing Calibration Weighting," *Journal of the Royal Statistical Society, Series B*, 78, 673–700.

Chen, X. (2007): "Large sample sieve estimation of semi-nonparametric models," *Handbook of econometrics*, 6, 5549–5632.

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

——— (2006): "Estimation of Copula-Based Semiparametric Time Series Models," *Journal of Econometrics*, 130, 307–335.

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

Fan, Y., and A. J. Patton (2014): "Copulas in Econometrics," *Annual Review of Economics*, 6, 179–200.

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.

Genest, C., J. Nešlehová, and N. Ben Ghorbal (2011): "Estimators Based on Kendall's Tau in Multivariate Copula Models," *Australian & New Zealand Journal of Statistics*, 53, 157–177.

Hahn, J. (1998): "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, 66, 315–331.

Hamori, S., K. Motegi, and Z. Zhang (2018): "Supplemental Material for "Calibration Estimation of Semiparametric Copula 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.

Hirano, K., G. W. Imbens, and G. Ridder (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71, 1161–1189.

Horvitz, D. G., and D. J. Thompson (1952): "A Generalization of Sampling Without Replacement from a Finite Universe," *Journal of the American Statistical Association*, 47, 663–685.

Imbens, G. W., W. Newey, and G. Ridder (2005): "Mean-square-error Calculations for Average Treatment Effects," IEPR Working Paper 05.34.

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

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

Little, R. J. A., and D. B. Rubin (2002): *Statistical Analysis with Missing Data*. Wiley-Interscience, second edn.

Marra, G., and K. Wyszynski (2016): "Semi-Parametric Copula Sample Selection Models for Count Responses," *Computational Statistics and Data Analysis*, 104, 110–129.

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

Oakes, D. (1994): "Multivariate Survival Distributions," *Journal of Nonparametric Statistics*, 3, 343–354.

Oh, D. H., and A. J. Patton (2013): "Simulated Method of Moments Estimation for Copula-Based Multivariate Models," *Journal of the American Statistical Association*, 108, 689–700.

——— (2016): "High-Dimensional Copula-Based Distributions with Mixed Frequency Data," *Journal of Econometrics*, 193, 349–366.

——— (2017a): "Modeling Dependence in High Dimensions With Factor Copulas," *Journal of Business and Economic Statistics*, 35, 139–154.

——— (2017b): "Time-Varying Systemic Risk: Evidence From a Dynamic Copula Model of CDS Spreads," *Journal of Business and Economic Statistics*, forthcoming.

Owen, A. B. (1988): "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75, 237–249.

Patton, A. J. (2009): "Copula-Based Models for Financial Time Series," in *Handbook of Financial Time Series*, ed. by T. G. Andersen, R. A. Davis, J. P. Kreiss, and T. Mikosch, pp. 767–785. Springer-Verlag.

——— (2012): "A Review of Copula Models for Economic Time Series," *Journal of Multivariate Analysis*, 110, 4–18.

——— (2013): "Copula Methods for Forecasting Multivariate Time Series," in *Handbook of Economic Forecasting*, ed. by E. Graham, and A. Timmermann, vol. 2B, chap. 16, pp. 899–960. Elsevier B. V.

Qin, J., and J. Lawless (1994): "Empirical Likelihood and General Estimating Equations," *Annals of Statistics*, 22, 300–325.

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.

Robins, J. M., and A. Rotnitzky (1995): "Semiparametric Efficiency in Multivariate Regression Models with Missing Data," *Journal of the American Statistical Association*, 90, 122–129.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994): "Estimation of Regression Coefficients When Some Regressors are not Always Observed," *Journal of the American Statistical Association*, 89, 846–866.

Robinson, P. M. (1988): "Root-N-consistent semiparametric regression," *Econometrica: Journal of the Econometric Society*, pp. 931–954.

Rubin, D. B. (1976): "Inference and Missing Data," *Biometrika*, 63, 581–592.

Salvatierra, I. D. L., and A. J. Patton (2015): "Dynamic Copula Models and High Frequency Data," *Journal of Empirical Finance*, 30, 120–135.

Shih, J. H., and T. A. Louis (1995): "Inferences on the Association Parameter in Copula Models for Bivariate Survival Data," *Biometrics*, 51, 1384–1399.

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.

Stone, C. J. (1982): "Optimal global rates of convergence for nonparametric regression," *The annals of statistics*, pp. 1040–1053.

——— (1994): "The use of polynomial splines and their tensor products in multivariate function estimation," *The Annals of Statistics*, pp. 118–171.

Tan, Z. (2011): "Efficient Restricted Estimators for Conditional Mean Models with Missing Data," *Biometrika*, 98, 663–684.

Trivedi, P. K., and D. M. Zimmer (2007): *Copula Modeling: An Introduction for Practitioners*. Now Publishers, Foundations and Trends in Econometrics.

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

Zhao, L. P., and S. Lipsitz (1992): "Designs and Analysis of Two-Stage Studies," *Statistics in Medicine*, 11, 769–782.

Technical Appendices

A Notations

For each $j \in \{1, ..., d\}$, we define the following notations:

$$
\begin{aligned}
&U_{ji} \triangleq F_j^0(Y_{ji}) \ , \ \boldsymbol{U}_i \triangleq (U_{1i}, ..., U_{di})^\top \ , \ l(v_1, ..., v_d; \theta) \triangleq \log c(v_1, ..., v_d; \theta) \ , \\
&l_\theta(v_1, ..., v_d; \theta) \triangleq \frac{\partial}{\partial \theta} l(v_1, ..., v_d; \theta) \ , \ l_{\theta\theta}(v_1, ..., v_d; \theta) \triangleq \frac{\partial^2}{\partial \theta \partial \theta'} l(v_1, ..., v_d; \theta) \ , \\
&l_j(v_1, ..., v_d; \theta) \triangleq \frac{\partial}{\partial v_j} l(v_1, ..., v_d; \theta) \ , \ l_{\theta j}(v_1, ..., v_d; \theta) \triangleq \frac{\partial^2}{\partial \theta \partial v_j} l(v_1, ..., v_d; \theta) \ .
\end{aligned}
$$

The following notations are used for describing the asymptotic variance:

$$
\begin{aligned}
\boldsymbol{\varphi}(\boldsymbol{T}_i, \boldsymbol{X}_i, \boldsymbol{U}_i; \theta_0) \triangleq & \frac{I(T_{1i} = 1, ..., T_{di} = 1)}{\eta(\boldsymbol{X}_i)} l_\theta(U_{1i}, ..., U_{di}; \theta_0) - \frac{I(T_{1i} = 1, ..., T_{di} = 1)}{\eta(\boldsymbol{X}_i)} \mathbb{E}[l_\theta(U_{1i}, ... U_{di}; \theta_0) | \boldsymbol{X}_i] \\
& + \mathbb{E}[l_\theta(U_{1i}, ... U_{di}; \theta_0) | \boldsymbol{X}_i] - \mathbb{E}[l_\theta(U_{1i}, ... U_{di}; \theta_0)] \ , \qquad \text{(A.1)}
\end{aligned}
$$

$$
\phi_j(T_{ji}, \boldsymbol{X}_i, U_{ji}; v) \triangleq \frac{T_{ji}}{\pi_j(\boldsymbol{X}_i)} I(U_{ji} \le v) - \frac{T_{ji}}{\pi_j(\boldsymbol{X}_i)} \mathbb{E}[I(U_{ji} \le v) | \boldsymbol{X}_i] + \mathbb{E}[I(U_{ji} \le v) | \boldsymbol{X}_i] \ , \ v \in [0, 1] \ , \qquad \text{(A.2)}
$$

$$
\boldsymbol{W}_j(T_{ji}, \boldsymbol{X}_i, U_{ji}; \theta_0) \triangleq \mathbb{E}\left[l_{\theta j}(U_{1s}, ..., U_{ds}; \theta_0) \{\phi_j(T_{ji}, \boldsymbol{X}_i, U_{ji}; U_{js}) - U_{js}\} | U_{ji}, \boldsymbol{X}_i, T_{ji}\right] \ (s \ne i) \ . \qquad \text{(A.3)}
$$

B Assumptions

We first introduce the smoothness classes of functions used in the nonparametric estimation; see e.g. Stone (1982, 1994), Robinson (1988), Newey (1997) and Chen (2007). Suppose that $\mathcal{X}$ is the Cartesian product of r-compact intervals. Let $0 < \delta \le 1$. A fucntion f on $\mathcal{X}$ is said to satisfy a Hölder condition with exponent δ if there is a positive constant L such that $\|f(\boldsymbol{x}_1) - f(\boldsymbol{x}_2)\| \le L\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^\delta$ for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X}$. Given a r-tuple $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_r)$ of nonnegative integer, denote $[\boldsymbol{\alpha}] = \alpha_1 + \cdots + \alpha_r$ and let $D^{\boldsymbol{\alpha}}$ denote the differential operator defined by $D^{\boldsymbol{\alpha}} = \frac{\partial^{[\boldsymbol{\alpha}]}}{\partial x_1^{\alpha_1} \cdots \partial x_r^{\alpha_r}}$, where $\boldsymbol{x} = (x_1, ..., x_r)$.

Definition Let s be a nonnegative integer and $s := s_0 + \delta$. The function f on $\mathcal{X}$ is said to be s-smooth if it is s times continuously differentiable on $\mathcal{X}$ and $D^{\boldsymbol{\alpha}} f$ satisfies a Hölder condition with exponent δ for all $\boldsymbol{\alpha}$ with $[\boldsymbol{\alpha}] = s_0$.

Assumptions B.1-B.6 are needed for showing $\sqrt{N}$-consistent estimate of the marginal distributions $\{F_j^0\}_{j=1}^d$ and the asymptotic normality of $\sqrt{N}(\hat{F}_j - F_j^0)$. Assumptions B.1-B.6 are similar to Conditions 3-8 of Chan, Yam, and Zhang (2016).

Assumption B.1. *The support of the covariate $\boldsymbol{X}$, which is denoted by $\mathcal{X}$, is a Cartesian product of r compact intervals.*

Assumption B.2. *For any $j \in \{1, ..., d\}$, the function $\pi_j(x)$ is bounded below, i.e. there exists some constant η_0 such that*

$$0 < \eta_0 \leq \pi_j(x) \leq 1 \quad \forall x \in \mathcal{X} \ , \ j \in \{1, ..., d\} \ .$$

Assumption B.3. *For any $j \in \{1, ..., d\}$, the function $\pi_j(x)$ are s-times continuously differentiable, where $s > 9r$.*

Assumption B.4. *For any $j \in \{1, ..., d\}$, the conditional distribution functions $\mathbb{P}(Y_{ji} \leq y | \boldsymbol{X}_i = x)$ is $\bar{s}$-smooth in x with $\bar{s} > 0$.*

Assumption B.5. *$\rho \in C^3(\mathbb{R})$ is a strictly concave function defined on $\mathbb{R}$, i.e., $\rho'' < 0$, and the range of ρ' contains $[1, 1/\eta_0]$ which is a subset of the positive real line.*

Assumption B.6. *$K(N) = O(N^{\nu})$, where $\frac{1}{s/r-2} < \nu < \frac{1}{7}$.*

Assumptions B.1 restricts the covariates to be bounded. This condition is restrictive but commonly imposed in the nonparametric regression literature. Assumption B.2 ensures that there are always sufficient portion of observed marginal data. This condition is typically required in the missing data literature. The smoothness conditions of Assumptions B.3 and B.4 are used to control the approximation error, and again are common in the nonparametric literature. Assumption B.5 is a mild assumption on ρ which is chosen by the statisticians and includes all the important special cases considered in the literature, such as exponential tilting, empirical likelihood, quadratic weighting, inverse logistic. Assumption B.6 restricts the smoothing parameter to balance the bias against the variance. This condition is typical in the nonparametric regression literature.

Additional Assumptions B.7-B.12 are needed for showing the $\sqrt{N}$-consistency and asymptotic normality of $\sqrt{N}(\hat{\theta} - \theta_0)$. Assumptions B.9, B.11, B.12, and B.13 are also maintained in Chen and Fan (2005).

Assumption B.7. *The function $\eta(x)$ is bounded below, i.e. there exists some constant η_0 such that*

$$0 < \eta_0 \leq \eta(x) \leq 1 \quad \forall x \in \mathcal{X} \ , \ j \in \{1, ..., d\} \ .$$

Assumption B.8. *The function $\eta(x)$ is s-times continuously differentiable, where $s > 9r$.*

Assumption B.9.

1. *For any $\boldsymbol{u} \in (0,1)^d$, $l(\boldsymbol{u}; \theta)$ is a continuous function of θ.*
2. *$\mathbb{E}[\sup_{\theta \in \Theta} |l(U_{1i}, ..., U_{di}; \theta)|] < \infty$.*

Assumption B.10. *The function $\mathbb{E}[l_\theta(U_{1i}, ..., U_{di}; \theta) | \boldsymbol{X} = x]$ is $\bar{s}$-smooth in x with $\bar{s} > 0$.*

Assumption B.11. *The following matrices are finite and positive definite:*

$$B \triangleq -\mathbb{E}\left[l_{\theta\theta}(U_{1i}, ..., U_{di}; \theta_0)\right] \ , \tag{B.1}$$

$$\Sigma \triangleq Var\left(\varphi(\boldsymbol{T}_i, \boldsymbol{X}_i, \boldsymbol{U}_i; \theta_0) + \sum_{j=1}^{d} W_j(T_{ji}, \boldsymbol{X}_i, U_{ji}; \theta_0)\right) \ . \tag{B.2}$$

Assumption B.12.

1. *For every* $(u_1, ..., u_d) \in (0,1)^d$, *the function* $l_{\theta\theta}(u_1, ..., u_d; \theta)$ *is continuous with respect to* θ *in a neighborhood of* θ_0*;*

2. $\mathbb{E}\left[\sup_{\theta\in\Theta:\|\theta-\theta_0\|=o(1)} \|l_{\theta\theta}(U_{1i}, ..., U_{di}; \theta)\|\right] < \infty$.

Assumption B.13. *For* $j \in \{1, ..., d\}$, $l_{\theta j}(u_1, ..., u_d; \theta_0)$ *is well defined and continuous in* $(u_1, ..., u_d) \in (0,1)^d$. *Furthermore, we assume*

1. $\|l_\theta(u_1, ..., u_d; \theta_0)\| \leq$ *constant* $\times \prod_{j=1}^d \{v_j(1-v_j)\}^{-a_j}$ *for some* $a_j \geq 0$ *such that*

$$\mathbb{E}\left[\prod_{j=1}^d \{U_{ji}(1-U_{ji})\}^{-2a_j}\right] < \infty \ ;$$

2. $\|l_{\theta k}(u_1, ..., u_d; \theta_0)\| \leq$ *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*

$$\mathbb{E}\left[\{U_{ki}(1-U_{ki})\}^{\xi_k - b_k} \prod_{j=1, j\neq k}^d \{U_{ji}(1-U_{ji})\}^{-a_j}\right] < \infty$$

   for some $\xi_k \in (0, 1/2)$.

Assumptions B.7-B.9 are required for showing the consistent estimation of the target parameter. Assumption B.7 ensures that there are sufficient portion of complete data. The smoothness condition of Assumption B.8 is used to control the approximation error. Assumption B.9 is used for ensuring the uniform convergence of criteria functions, which is a standard condition for M-estimator. Assumptions B.10-B.13 are needed for ensuring our estimator has asymptotic normal behavior. Assumption B.10 is used to control the approximation error. Assumption B.11 is used to guarantee the finiteness of the asymptotic variance. Assumption B.12 is used to guarantee the uniform convergence. Assumption B.13 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.

Table 1: Simulation Results on Clayton Copula with $\alpha_0 = 1.333$

$N = 250$				
(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
Listwise Deletion	0.041, 0.212, 0.216	0.090, 0.230, 0.247	0.062, 0.247, 0.254	0.135, 0.265, 0.297
Param (Correct)	0.026, 0.209, 0.211	0.026, 0.195, 0.197	0.036, 0.232, 0.235	0.034, 0.220, 0.222
Param (Misspec)	-0.294, 0.154, 0.332	-0.265, 0.152, 0.305	-0.063, 0.209, 0.218	-0.047, 0.209, 0.214
Nonparam ($K = 3$)	-0.063, 0.431, 0.435	0.079, 1.147, 1.149	-0.469, 0.555, 0.727	-0.327, 0.747, 0.815
Nonparam ($K = 4$)	-0.008, 1.079, 1.079	0.287, 1.926, 1.947	-0.418, 1.714, 1.764	-0.208, 1.337, 1.354
Calib. Est. ($K = 3$)	0.039, 0.203, 0.207	0.026, 0.201, 0.203	0.035, 0.221, 0.224	0.035, 0.215, 0.218
Calib. Est. ($K = 4$)	0.036, 0.208, 0.211	0.034, 0.207, 0.210	0.037, 0.227, 0.230	0.036, 0.222, 0.224

$N = 500$				
(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
Listwise Deletion	0.021, 0.151, 0.152	0.061, 0.163, 0.174	0.032, 0.188, 0.190	0.118, 0.182, 0.216
Param (Correct)	0.012, 0.146, 0.146	0.025, 0.144, 0.146	0.006, 0.167, 0.168	0.018, 0.155, 0.156
Param (Misspec)	-0.297, 0.106, 0.315	-0.271, 0.106, 0.291	-0.066, 0.141, 0.156	-0.055, 0.149, 0.159
Nonparam ($K = 3$)	-0.046, 0.524, 0.526	0.027, 0.851, 0.851	-0.438, 0.687, 0.814	-0.321, 0.364, 0.485
Nonparam ($K = 4$)	0.069, 1.951, 1.952	0.170, 1.666, 1.675	-0.285, 2.141, 2.160	-0.294, 0.646, 0.710
Calib. Est. ($K = 3$)	0.021, 0.141, 0.143	0.020, 0.145, 0.146	0.021, 0.161, 0.163	0.016, 0.153, 0.154
Calib. Est. ($K = 4$)	0.013, 0.144, 0.145	0.016, 0.141, 0.142	0.020, 0.157, 0.158	0.021, 0.153, 0.155

$N = 1000$				
(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
Listwise Deletion	0.005, 0.103, 0.103	0.047, 0.108, 0.118	0.020, 0.122, 0.123	0.102, 0.130, 0.165
Param (Correct)	0.006, 0.104, 0.104	0.009, 0.101, 0.101	0.007, 0.112, 0.112	0.011, 0.111, 0.112
Param (Misspec)	-0.298, 0.074, 0.307	-0.273, 0.076, 0.283	-0.075, 0.103, 0.127	-0.071, 0.103, 0.124
Nonparam ($K = 3$)	-0.058, 0.606, 0.609	0.017, 0.389, 0.389	-0.393, 0.469, 0.612	-0.324, 0.280, 0.429
Nonparam ($K = 4$)	-0.013, 1.199, 1.199	0.286, 1.761, 1.784	-0.389, 1.332, 1.387	-0.331, 0.336, 0.472
Calib. Est. ($K = 3$)	0.003, 0.102, 0.102	0.006, 0.099, 0.099	0.006, 0.115, 0.115	0.014, 0.113, 0.114
Calib. Est. ($K = 4$)	0.008, 0.099, 0.100	0.010, 0.100, 0.101	0.012, 0.115, 0.115	0.007, 0.110, 0.110

Bias, standard deviation, and root mean squared error of each estimator after $J = 1000$ Monte Carlo trials. Cases A and B imply MCAR and MAR, respectively, with relatively low missing probability (i.e. $\mathbb{E}[T_{2i}] = 0.8$). Cases C and D imply MCAR and MAR, respectively, with relatively high missing probability (i.e. $\mathbb{E}[T_{2i}] = 0.6$).

Table 2: Simulation Results on Clayton Copula with $\alpha_0 = 4.667$

$N = 250$				
(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
Listwise Deletion	-0.042, 0.504, 0.506	0.183, 0.557, 0.586	-0.061, 0.608, 0.611	0.362, 0.671, 0.763
Param (Correct)	-0.111, 0.504, 0.516	-0.099, 0.485, 0.495	-0.230, 0.530, 0.578	-0.120, 0.527, 0.541
Param (Misspec)	-2.481, 0.284, 2.497	-2.258, 0.296, 2.278	-1.062, 0.476, 1.163	-0.800, 0.479, 0.932
Nonparam ($K = 3$)	-1.097, 1.940, 2.229	-0.428, 3.109, 3.138	-2.844, 1.768, 3.349	-2.283, 1.498, 2.731
Nonparam ($K = 4$)	-0.895, 4.356, 4.447	0.214, 5.442, 5.446	-2.504, 3.541, 4.337	-2.048, 2.452, 3.195
Calib. Est. ($K = 3$)	-0.096, 0.501, 0.510	-0.064, 0.485, 0.489	-0.161, 0.528, 0.552	-0.109, 0.525, 0.536
Calib. Est. ($K = 4$)	-0.086, 0.509, 0.516	-0.047, 0.477, 0.479	-0.142, 0.523, 0.542	-0.123, 0.545, 0.559

$N = 500$				
(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
Listwise Deletion	-0.039, 0.381, 0.383	0.147, 0.382, 0.410	-0.016, 0.430, 0.430	0.372, 0.463, 0.594
Param (Correct)	-0.076, 0.343, 0.351	-0.035, 0.344, 0.346	-0.153, 0.375, 0.405	-0.076, 0.386, 0.394
Param (Misspec)	-2.467, 0.211, 2.476	-2.227, 0.210, 2.237	-0.928, 0.349, 0.991	-0.769, 0.329, 0.837
Nonparam ($K = 3$)	-1.004, 2.615, 2.801	-0.581, 1.310, 1.433	-2.744, 1.749, 3.254	-2.253, 1.050, 2.486
Nonparam ($K = 4$)	-1.062, 3.148, 3.322	0.133, 4.598, 4.600	-2.442, 3.306, 4.110	-1.980, 1.503, 2.486
Calib. Est. ($K = 3$)	-0.079, 0.355, 0.363	-0.032, 0.336, 0.337	-0.118, 0.365, 0.383	-0.063, 0.368, 0.374
Calib. Est. ($K = 4$)	-0.052, 0.350, 0.354	-0.050, 0.346, 0.350	-0.090, 0.368, 0.378	-0.082, 0.367, 0.376

$N = 1000$				
(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
Listwise Deletion	-0.021, 0.260, 0.260	0.167, 0.288, 0.333	-0.011, 0.310, 0.311	0.406, 0.325, 0.521
Param (Correct)	-0.044, 0.252, 0.256	-0.024, 0.241, 0.242	-0.090, 0.256, 0.271	-0.044, 0.257, 0.260
Param (Misspec)	-2.445, 0.151, 2.449	-2.220, 0.149, 2.225	-0.866, 0.264, 0.905	-0.709, 0.243, 0.749
Nonparam ($K = 3$)	-1.082, 1.083, 1.531	-0.518, 1.600, 1.682	-2.803, 1.302, 3.091	-2.181, 0.940, 2.375
Nonparam ($K = 4$)	-1.060, 2.627, 2.832	0.079, 3.723, 3.724	-2.374, 3.229, 4.008	-2.013, 1.235, 2.361
Calib. Est. ($K = 3$)	-0.040, 0.248, 0.251	-0.023, 0.243, 0.244	-0.062, 0.257, 0.264	-0.056, 0.259, 0.265
Calib. Est. ($K = 4$)	-0.021, 0.241, 0.242	-0.042, 0.246, 0.250	-0.049, 0.260, 0.265	-0.032, 0.265, 0.267

Bias, standard deviation, and root mean squared error of each estimator after $J = 1000$ Monte Carlo trials. Cases A and B imply MCAR and MAR, respectively, with relatively low missing probability (i.e. $\mathbb{E}[T_{2i}] = 0.8$). Cases C and D imply MCAR and MAR, respectively, with relatively high missing probability (i.e. $\mathbb{E}[T_{2i}] = 0.6$).

Table 3: Simulation Results on Gumbel Copula with $\gamma_0 = 1.667$

$N = 250$				
(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
Listwise Deletion	0.027, 0.117, 0.120	-0.023, 0.111, 0.113	0.037, 0.136, 0.141	-0.058, 0.127, 0.139
Param (Correct)	-0.009, 0.104, 0.104	-0.012, 0.111, 0.112	-0.016, 0.125, 0.126	-0.024, 0.121, 0.124
Param (Misspec)	-0.633, 0.009, 0.634	-0.631, 0.012, 0.631	-0.520, 0.070, 0.525	-0.524, 0.066, 0.528
Nonparam ($K = 3$)	-0.138, 0.115, 0.180	-0.112, 0.136, 0.176	-0.449, 0.150, 0.473	-0.473, 0.196, 0.512
Nonparam ($K = 4$)	-0.184, 0.191, 0.265	-0.180, 0.196, 0.266	-0.492, 0.163, 0.519	-0.552, 0.152, 0.573
Calib. Est. ($K = 3$)	0.015, 0.108, 0.109	0.013, 0.108, 0.109	0.015, 0.122, 0.123	0.009, 0.124, 0.124
Calib. Est. ($K = 4$)	0.010, 0.105, 0.106	0.014, 0.115, 0.116	0.012, 0.123, 0.124	0.008, 0.123, 0.123

$N = 500$				
(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
Listwise Deletion	0.012, 0.080, 0.081	-0.033, 0.077, 0.084	0.014, 0.094, 0.095	-0.070, 0.088, 0.113
Param (Correct)	-0.001, 0.077, 0.077	-0.002, 0.076, 0.076	-0.005, 0.089, 0.089	-0.013, 0.084, 0.085
Param (Misspec)	-0.633, 0.006, 0.633	-0.629, 0.008, 0.629	-0.527, 0.045, 0.529	-0.523, 0.049, 0.525
Nonparam ($K = 3$)	-0.149, 0.101, 0.180	-0.115, 0.122, 0.168	-0.415, 0.145, 0.440	-0.459, 0.178, 0.492
Nonparam ($K = 4$)	-0.183, 0.196, 0.268	-0.164, 0.216, 0.272	-0.442, 0.194, 0.482	-0.556, 0.149, 0.575
Calib. Est. ($K = 3$)	0.010, 0.079, 0.080	0.012, 0.077, 0.078	0.008, 0.088, 0.088	0.004, 0.088, 0.088
Calib. Est. ($K = 4$)	0.007, 0.079, 0.079	0.007, 0.077, 0.077	0.008, 0.087, 0.087	0.003, 0.089, 0.089

$N = 1000$				
(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
Listwise Deletion	0.008, 0.058, 0.058	-0.039, 0.054, 0.067	0.012, 0.066, 0.067	-0.080, 0.060, 0.100
Param (Correct)	-0.000, 0.055, 0.055	-0.000, 0.054, 0.054	0.001, 0.062, 0.062	-0.002, 0.063, 0.063
Param (Misspec)	-0.632, 0.005, 0.632	-0.628, 0.006, 0.628	-0.527, 0.028, 0.528	-0.526, 0.028, 0.527
Nonparam ($K = 3$)	-0.119, 0.664, 0.674	-0.072, 0.828, 0.831	-0.401, 0.119, 0.419	-0.444, 0.592, 0.740
Nonparam ($K = 4$)	-0.056, 1.279, 1.280	-0.068, 1.113, 1.115	-0.298, 1.548, 1.576	-0.568, 0.127, 0.582
Calib. Est. ($K = 3$)	0.004, 0.057, 0.057	0.003, 0.056, 0.056	0.006, 0.061, 0.061	-0.002, 0.061, 0.061
Calib. Est. ($K = 4$)	0.004, 0.054, 0.055	0.004, 0.056, 0.056	0.005, 0.062, 0.062	0.001, 0.063, 0.063

Bias, standard deviation, and root mean squared error of each estimator after $J = 1000$ Monte Carlo trials. Cases A and B imply MCAR and MAR, respectively, with relatively low missing probability (i.e. $\mathbb{E}[T_{2i}] = 0.8$). Cases C and D imply MCAR and MAR, respectively, with relatively high missing probability (i.e. $\mathbb{E}[T_{2i}] = 0.6$).

Table 4: Simulation Results on Gumbel Copula with $\gamma_0 = 3.333$

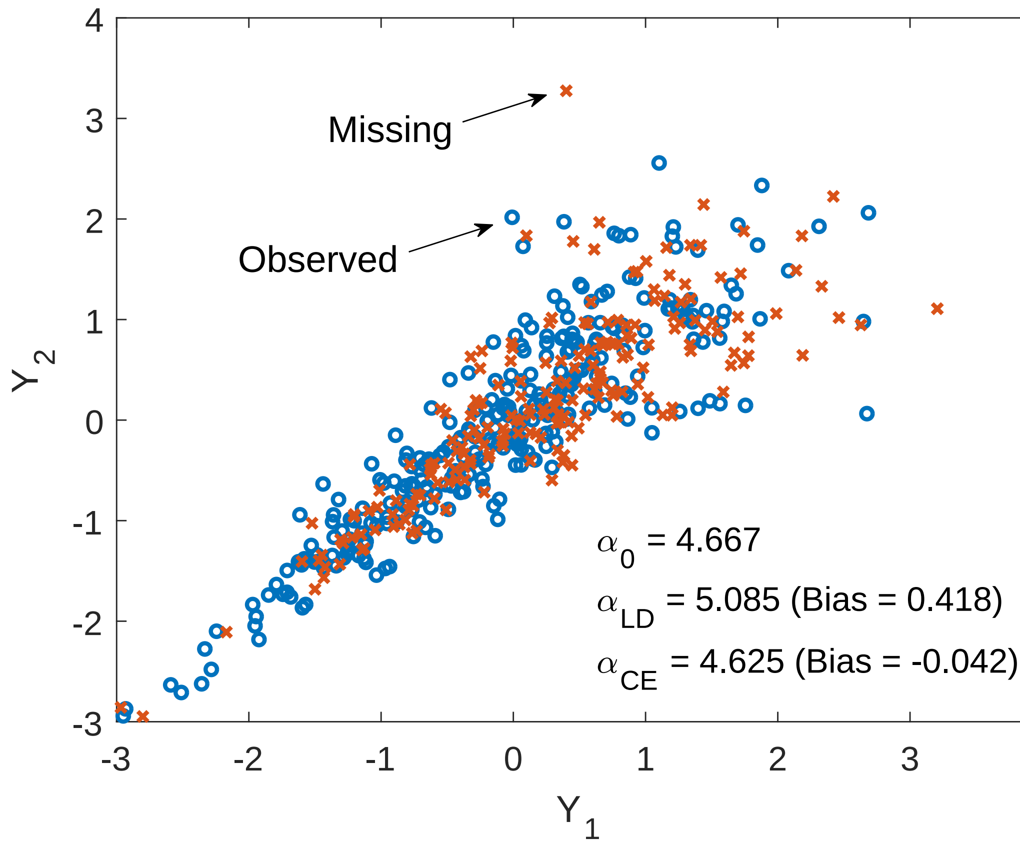
$N = 250$				
(a, b)	A. $(-1.385, 0.000)$	B. $(-1.430, 0.400)$	C. $(-0.405, 0.000)$	D. $(-0.420, 0.400)$
	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Listwise Deletion	0.020, 0.269, 0.270	-0.108, 0.252, 0.274	0.016, 0.305, 0.306	-0.173, 0.286, 0.334
Param (Correct)	-0.095, 0.228, 0.247	-0.124, 0.233, 0.264	-0.150, 0.272, 0.310	-0.188, 0.250, 0.313
Param (Misspec)	-2.264, 0.014, 2.264	-2.259, 0.015, 2.259	-1.964, 0.206, 1.974	-1.964, 0.216, 1.976
Nonparam ($K = 3$)	-1.163, 0.270, 1.194	-0.997, 0.326, 1.049	-1.908, 0.302, 1.931	-1.932, 0.403, 1.974
Nonparam ($K = 4$)	-1.264, 0.397, 1.325	-1.230, 0.501, 1.328	-2.014, 0.299, 2.036	-2.115, 0.317, 2.138
Calib. Est. ($K = 3$)	0.006, 0.250, 0.250	-0.008, 0.237, 0.237	-0.020, 0.266, 0.267	-0.079, 0.273, 0.284
Calib. Est. ($K = 4$)	-0.015, 0.254, 0.254	-0.022, 0.240, 0.241	-0.028, 0.266, 0.267	-0.080, 0.278, 0.290

$N = 500$				
(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
Listwise Deletion	0.011, 0.190, 0.190	-0.116, 0.173, 0.209	0.010, 0.217, 0.217	-0.221, 0.201, 0.298
Param (Correct)	-0.042, 0.174, 0.179	-0.049, 0.175, 0.182	-0.050, 0.196, 0.202	-0.093, 0.193, 0.214
Param (Misspec)	-2.263, 0.010, 2.263	-2.256, 0.011, 2.256	-1.992, 0.110, 1.995	-1.982, 0.122, 1.986
Nonparam ($K = 3$)	-1.178, 0.213, 1.198	-1.003, 0.252, 1.034	-1.848, 0.229, 1.862	-1.907, 0.392, 1.947
Nonparam ($K = 4$)	-1.278, 0.378, 1.333	-1.210, 0.451, 1.292	-1.961, 0.293, 1.983	-2.113, 0.294, 2.133
Calib. Est. ($K = 3$)	-0.001, 0.182, 0.182	-0.023, 0.175, 0.176	-0.023, 0.190, 0.192	-0.055, 0.189, 0.197
Calib. Est. ($K = 4$)	0.010, 0.184, 0.184	-0.012, 0.169, 0.169	-0.019, 0.199, 0.200	-0.069, 0.190, 0.202

$N = 1000$				
(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
Listwise Deletion	0.003, 0.129, 0.129	-0.116, 0.124, 0.170	0.010, 0.154, 0.154	-0.219, 0.145, 0.263
Param (Correct)	-0.019, 0.126, 0.127	-0.023, 0.131, 0.133	-0.027, 0.136, 0.138	-0.051, 0.132, 0.142
Param (Misspec)	-2.262, 0.007, 2.262	-2.255, 0.008, 2.255	-1.996, 0.071, 1.997	-1.994, 0.070, 1.995
Nonparam ($K = 3$)	-1.153, 0.684, 1.341	-0.972, 0.983, 1.382	-1.776, 1.004, 2.040	-1.910, 0.380, 1.948
Nonparam ($K = 4$)	-1.193, 0.892, 1.490	-0.933, 2.165, 2.357	-1.870, 0.481, 1.931	-2.125, 0.294, 2.145
Calib. Est. ($K = 3$)	-0.002, 0.123, 0.123	-0.017, 0.125, 0.126	-0.014, 0.136, 0.137	-0.035, 0.133, 0.137
Calib. Est. ($K = 4$)	-0.007, 0.120, 0.120	-0.009, 0.123, 0.123	-0.008, 0.136, 0.136	-0.029, 0.133, 0.136

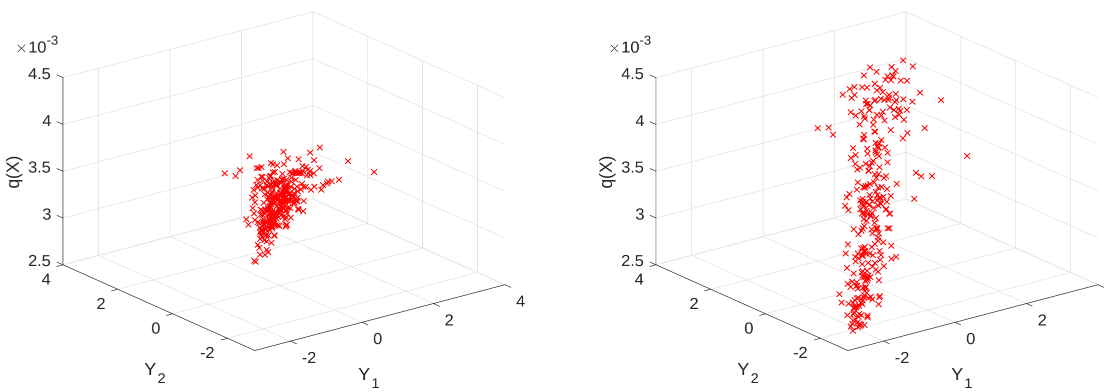
Bias, standard deviation, and root mean squared error of each estimator after $J = 1000$ Monte Carlo trials. Cases A and B imply MCAR and MAR, respectively, with relatively low missing probability (i.e. $\mathbb{E}[T_{2i}] = 0.8$). Cases C and D imply MCAR and MAR, respectively, with relatively high missing probability (i.e. $\mathbb{E}[T_{2i}] = 0.6$).

Figure 1: Observed and Missing Groups of (Y_{1i}, Y_{2i}) under Clayton Copula



This figure visualizes a representative Monte Carlo sample on the Clayton copula with $\alpha_0 = 4.667$, Case D (MAR), and $N = 500$. "o" signifies 291 individuals with Y_{2i} observed, while "x" signifies 209 individuals with Y_{2i} missing. When the listwise deletion is executed, there is a substantial *positive* bias of 0.418. When the calibration estimation with $K = 4$ is executed, there is a sufficiently small bias of -0.042.

Figure 2: Weights $\hat{q}(X_i)$ for Individuals with Complete Data (Clayton Copula)

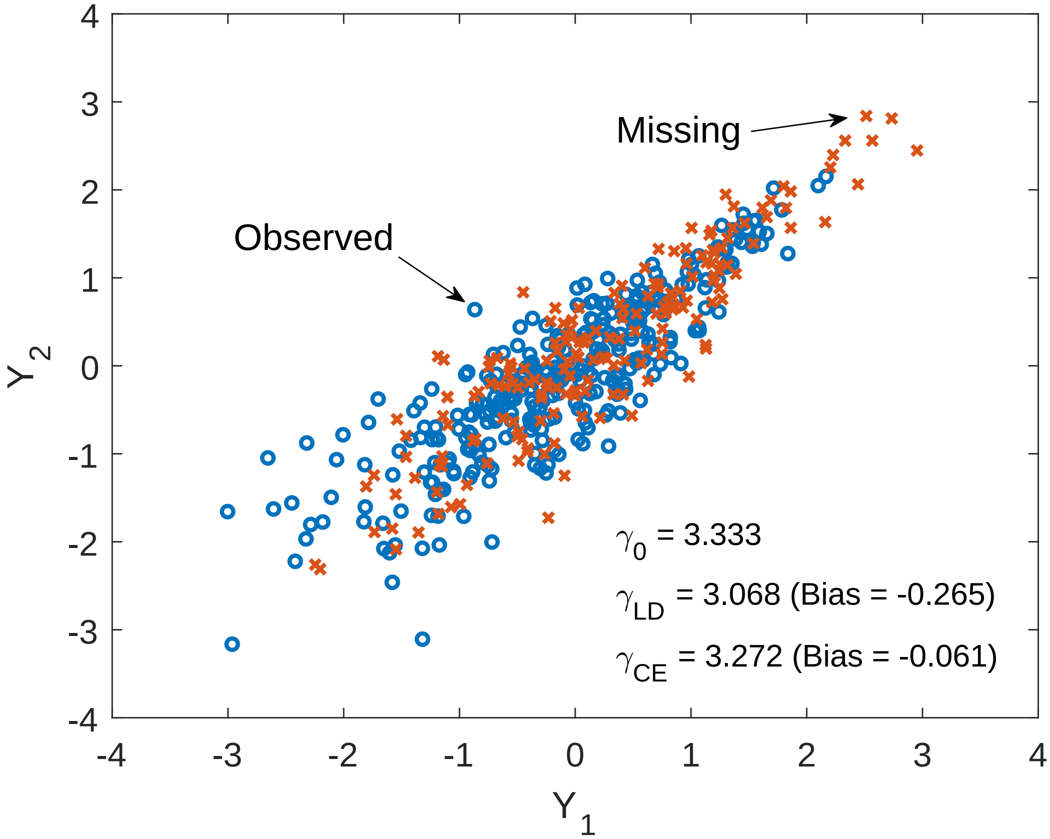


(a) Listwise Deletion

(b) Calibration Estimation

In this figure we plot the key weights $\hat{q}(X_i)$ assigned for the 291 individuals with complete data, continuing the representative Monte Carlo sample in Figure 1. When the listwise deletion is executed, $\hat{q}(X_i) = 1/291 = 0.003$ for any individual. When the calibration estimation with $K = 4$ is executed, $\hat{q}(X_i)$ is generally increasing in X_i so that the weak association between Y_{1i} and Y_{2i} at the upper tail is more emphasized and the strong association at the lower tail is less emphasized.

Figure 3: Observed and Missing Groups of (Y_{1i}, Y_{2i}) under Gumbel Copula



This figure visualizes a representative Monte Carlo sample on the Gumbel copula with $\gamma_0 = 3.333$, Case D (MAR), and $N = 500$. "o" signifies 317 individuals with Y_{2i} observed, while "x" signifies 183 individuals with Y_{2i} missing. When the listwise deletion is executed, there is a substantial *negative* bias of -0.265. When the calibration estimation with $K = 4$ is executed, there is a sufficiently small bias of -0.061.