Calibration Estimation of Semiparametric Copula Models with Data Missing at Random

Shigeyuki Hamori$^1$  Kaiji Motegi$^1$  Zheng Zhang$^2$

$^1$Kobe University  
$^2$Renmin University of China

Institute of Statistics and Big Data  
Renmin University of China

August 28, 2018
Introduction

Copula Model
-- Flexible modelling of multivariate distributions
-- Popular in business, economics, finance, etc.

We unite them for the first time in the literature.

Missing Data
-- Common problem in all fields
-- Deep literature in statistics
-- Binary phenomenon (to observe or not to observe)

Treatment Effect
-- Estimate unobserved outcome
-- Central topic in econometrics

Calibration estimation

Analogy
How to fit copula models when there are missing data?

A naïve approach is **listwise deletion (LD)**:

1. Keep individuals with all $d$ components being observed, and discard all other individuals.

2. Treat the individuals with complete data in an equal way.
LD leads to a **consistent** estimator for the copula parameter of interest if the missing mechanism is **missing completely at random (MCAR)**.

LD leads to an **inconsistent** estimator if the missing mechanism is **missing at random (MAR)**.

Under MAR, target variables $\mathbf{Y}_i = [Y_{1i}, \ldots, Y_{di}]^\top$ and their missing status are independent of each other given observed covariates $\mathbf{X}_i = [X_{1i}, \ldots, X_{ri}]^\top$.

LD treats individuals with complete data all equally, and it does not use the information of $\mathbf{X}_i$. That can cause **substantial bias** under MAR.
How to obtain a consistent estimator for the copula parameter when the missing mechanism is MAR?

A key step is the estimation of propensity score function (i.e. conditional probability of observing data given covariates).

Direct estimation of propensity score is notoriously challenging, whether it is performed parametrically or nonparametrically.

Parametric approaches are haunted by misspecification problems, while nonparametric approaches are often unstable.
We apply Chan, Yam, and Zhang’s (2016) **calibration estimation** to a missing data problem for the first time in the literature.

The calibration estimator for the copula parameter satisfies **consistency** and **asymptotic normality** under some assumptions including *i.i.d.* data and the MAR condition.

We also derive a consistent estimator for the asymptotic covariance matrix.

Our simulation results indicate that the calibration estimator dominates listwise deletion, parametric approach, and nonparametric approach.
1) Introduction
2) Review of Copula Models
3) Review of Missing Data
4) Set-up of Main Problem
5) Theory of Calibration Estimation
6) Data-Driven Selection of Tuning Parameter $K$
7) Monte Carlo Simulations
8) Conclusions
Suppose that there are $N$ individuals and $d$ components:

$$Y_i = [Y_{1i}, Y_{2i}, \ldots, Y_{di}]^\top \quad (i = 1, \ldots, N).$$

Suppose that we want to estimate the $d$-dimensional joint distribution of $Y_i$, assuming i.i.d. What can we do?

There are potential problems about estimating the joint distribution directly.

<table>
<thead>
<tr>
<th></th>
<th>Parametric</th>
<th>Nonparametric</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d = \text{small}$</td>
<td>Misspecification</td>
<td>(Curse of dimensionality)</td>
</tr>
<tr>
<td>$d = \text{large}$</td>
<td>Misspecification, Parameter proliferation</td>
<td>Curse of dimensionality</td>
</tr>
</tbody>
</table>
Copula models accomplish flexible specification with a small number of parameters.

Copula models follow a two-step procedure.

- **Step 1**: Model the marginal distribution of each of the $d$ components separately.
- **Step 2**: Combine the $d$ marginal distributions to recover a joint distribution.

<table>
<thead>
<tr>
<th>Step 1</th>
<th>Step 2</th>
<th>Name of model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parametric</td>
<td>Parametric</td>
<td>Parametric copula model</td>
</tr>
<tr>
<td>Nonparametric</td>
<td>Parametric</td>
<td>Semiparametric copula model (Target of our paper)</td>
</tr>
<tr>
<td>Nonparametric</td>
<td>Nonparametric</td>
<td>Nonparametric copula model</td>
</tr>
</tbody>
</table>
The copula approach is justified by Sklar’s (1959) theorem.

Sklar’s theorem ensures the existence of a unique copula function \( C : (0, 1)^d \rightarrow (0, 1) \) that recovers a true joint distribution.

**Theorem (Sklar, 1959)**

Let \( \{ \mathbf{Y}_i \} \) be i.i.d. random vectors with a joint distribution \( F : \mathbb{R}^d \rightarrow (0, 1) \). Assume that the marginal distribution of \( Y_{ji} \), written as \( F_j : \mathbb{R} \rightarrow (0, 1) \), is continuous for \( j \in \{ 1, \ldots, d \} \). Then, there exists a unique function \( C : (0, 1)^d \rightarrow (0, 1) \) such that

\[
F(y_1, \ldots, y_d) = C(F_1(y_1), \ldots, F_d(y_d)),
\]

or in terms of probability density functions,

\[
f(y_1, \ldots, y_d) = c(f_1(y_1), \ldots, f_d(y_d)).
\]
Suppose that \( \{Y_{1i}, \ldots, Y_{di}\}_{i=1}^{N} \) are target variables.

Define the missing indicator:

\[
T_{ji} = \begin{cases} 
1 & \text{if } Y_{ji} \text{ is observed}, \\
0 & \text{if } Y_{ji} \text{ is missing}.
\end{cases}
\]

Suppose that \( \{X_{1i}, \ldots, X_{mi}\}_{i=1}^{N} \) are observable covariates.

**Missing Completely at Random (MCAR):**

\[
\{T_{1i}, \ldots, T_{di}\} \perp \{Y_{1i}, \ldots, Y_{di}\}.
\]

**Missing at Random (MAR):**

\[
\{T_{1i}, \ldots, T_{di}\} \perp \{Y_{1i}, \ldots, Y_{di}\} \mid \{X_{1i}, \ldots, X_{ri}\}.
\]
An illustrative example on health survey ($d = r = 1$):

$$Y_{1i} = \text{weight of individual } i; \quad X_{1i} = I(\text{Individual } i \text{ is female}).$$

- **MCAR** requires that $P[\text{Individual } i \text{ reports his/her weight}]$ should be independent of both weight and gender of individual $i$.

- **MAR** requires that:
  - $P[\text{A man reports his weight}]$ should be independent of his weight.
  - $P[\text{A woman reports her weight}]$ should be independent of her weight.
Review of Missing Data

Probably too restrictive, since men and women may have different willingness to report their weights.

More plausible than MCAR, since MAR controls for gender.

MAR may be still restrictive, since men (or women) with different weights may have different willingness to report their weights.

MNAR is most general since it controls for both gender and weight, but MNAR is hard to handle technically.
Review of Missing Data

- It is well known that listwise deletion (LD) leads to consistent inference under MCAR.

- It is also well known that LD leads to inconsistent inference under MAR.

- Correct inference under MAR has been extensively studied since the seminal work of Rubin (1976).

- The present paper assumes MAR and elaborates the estimation of semiparametric copula models, which has not been done in the literature.
Semiparametric copula models are estimated in two steps:

**Step 1:** Estimate the marginal distributions \{F_1, \ldots, F_d\} nonparametrically via

\[
F_j(y) = \mathbb{P}(Y_{ji} \leq y) = \mathbb{E}[I(Y_{ji} \leq y)].
\]

**Step 2:** Estimate the true copula parameter \(\theta_0\) via

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

If data were all observed, then we could simply replace the population quantities with sample counterparts.

When data are Missing at Random, we need to assign some weights based on **propensity score functions**.
Set-up of Main Problem

- Define propensity score functions:

\[ \pi_j(\mathbf{x}) = \mathbb{P}(T_{ji} = 1 \mid \mathbf{X}_i = \mathbf{x}), \quad j \in \{1, \ldots, d\}. \]

- Define \( p_j(\mathbf{x}) = 1/\pi_j(\mathbf{x}) \).

- Step 1 is rewritten as

\[
\begin{align*}
F_j(y) &= \mathbb{E}[I(Y_{ji} \leq y)] = \mathbb{E}[\mathbb{E}[I(Y_{ji} \leq y) \mid \mathbf{X}_i]] \\
&= \mathbb{E}\left[ \mathbb{E}\left[ \frac{T_{ji}}{\pi_j(\mathbf{X}_i)} \mathbf{X}_i \right] \times \mathbb{E}[I(Y_{ji} \leq y) \mid \mathbf{X}_i] \right] \\
&= \mathbb{E}\left[ \mathbb{E}\left[ \frac{T_{ji}}{\pi_j(\mathbf{X}_i)} \times I(Y_{ji} \leq y) \mid \mathbf{X}_i \right] \right] \quad (\because \text{MAR}) \\
&= \mathbb{E}\left[ \frac{T_{ji}}{\pi_j(\mathbf{X}_i)} \times I(Y_{ji} \leq y) \right] \quad (\because \text{MAR}) \\
&= \mathbb{E}[T_{ji} \times p_j(\mathbf{X}_i) \times I(Y_{ji} \leq y)].
\end{align*}
\]
Set-up of Main Problem

- We have derived:

$$F_j(y) = \mathbb{E}[I(T_{ji} = 1) \times p_j(x_i) \times I(Y_{ji} \leq y)].$$

- Horvitz and Thompson’s (1952) inverse probability weighting (IPW) estimator for $F_j$ is written as

$$\tilde{F}_j(y) = \frac{1}{N} \sum_{i=1}^{N} I(T_{ji} = 1)p_j(x_i)I(Y_{ji} \leq y).$$

- If $p_j(x)$ were known, then it would be straightforward to compute the IPW estimator.

- $p_j(x)$ is unknown in reality.
Set-up of Main Problem

- Define a propensity score function:
  \[ \eta(x) = \mathbb{P}(T_{1i} = 1, \ldots, T_{di} = 1 \mid X_i = x). \]

- Define \( q(x) = 1/\eta(x). \)

- Using MAR and LIE, Step 2 is rewritten as
  \[ \theta_0 = \arg \max_{\theta \in \Theta} \mathbb{E} [I(T_{1i} = 1, \ldots, T_{di} = 1) q(X_i) \log c(F_1(Y_{1i}), \ldots, F_d(Y_{di}); \theta)]. \]

- The IPW estimator for \( \theta_0 \) is given by
  \[ \tilde{\theta} = \arg \max_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} I(T_{1i} = 1, \ldots, T_{di} = 1) q(X_i) \log c(\tilde{F}_1(Y_{1i}), \ldots, \tilde{F}_d(Y_{di}); \theta). \]

- \( q(x) \) is unknown in reality.
Direct estimation of propensity score functions is challenging, whether it is performed parametrically or nonparametrically.

In the treatment effect literature, Chan, Yam, and Zhang (2016) propose an alternative approach that bypasses a direct estimation of propensity score.

They construct *calibration weights* by balancing the moments of observed covariates among treatment, control, and whole groups.

The present paper applies their method to a missing data problem for the first time.
Theory of Calibration Estimation

- Under MAR, the **moment matching condition** holds:

\[
\mathbb{E} \left[ I(T_{ji} = 1)p_{j,K_j}(X_i)u_{K_j}(X_i) \right] = \mathbb{E}[u_{K_j}(X_i)], \quad j \in \{1, \ldots d\}
\]

for any integrable function \(u_{K_j} : \mathbb{R}^r \to \mathbb{R}^{K_j}\) called an **approximation sieve**. A common choice is, say,

\[
u_{K_j}(X_i) = [1, X_i, X_i^2, X_i^3]^\top \quad (r = 1, K_j = 4).
\]

- A sample counterpart is written as

\[
\frac{1}{N} \sum_{i=1}^{N} I(T_{ji} = 1) \times p_{j,K_j}(X_i) \times u_{K_j}(X_i) = \frac{1}{N} \sum_{i=1}^{N} u_{K_j}(X_i).
\]

- Multiple values of \(\{p_{j,K_j}(X_1), \ldots, p_{j,K_j}(X_N)\}\) satisfy the moment matching condition. Among them, we choose the one closest to a **uniform** weight given some distance measure.
Theory of Calibration Estimation

Why do we want the **uniform** weight?

1. If there are no missing data, then the uniform weight leads to a natural estimator
   \[ \hat{F}_j(y) = (N + 1)^{-1} \sum_{i=1}^{N} I(Y_{ji} \leq y). \]

2. It is well known that volatile weights cause instability in the Horvitz-Thompson IPW estimator.

Let \( \rho : \mathbb{R} \to \mathbb{R} \) be any strictly concave function.

Define a concave objective function:

\[
G_{j,K_j}(\lambda) = \frac{1}{N} \sum_{i=1}^{N} \left[ I(T_{ji} = 1) \rho(\lambda^\top u_{K_j}(X_i)) - \lambda^\top u_{K_j}(X_i) \right], \quad \lambda \in \mathbb{R}^{K_j}.
\]

Compute

\[
\hat{\lambda}_{j,K_j} = \arg \max_\lambda G_{j,K_j}(\lambda).
\]
Compute calibration weights for marginal distributions:

\[ \hat{p}_{j,K_j}(X_i) = \rho'(\hat{\lambda}_{j,K_j}^\top u_{K_j}(X_i)). \]

Estimate the marginal distribution of the \( j \)-th component by

\[ \hat{F}_j(y) = \frac{1}{N} \sum_{i=1}^{N} I(T_{ji} = 1)\hat{p}_{j,K_j}(X_i)I(Y_{ji} \leq y). \]

Step 2 (maximum likelihood) can be handled in the same way. We put calibration weights \( \hat{q}_K(X_i) \) on individual log-likelihoods.
Theory of Calibration Estimation

**Theorem (Consistency and Asymptotic Normality)**

Impose a set of assumptions including what follows:

- The missing mechanism is missing at random (MAR).
- \( \{Y_i, T_i, X_i\} \) are i.i.d. across individuals \( i \in \{1, \ldots, N\} \).
- \( \pi_1(\cdot), \ldots, \pi_d(\cdot), \) and \( \eta(\cdot) \) are sufficiently smooth.
- \( K_j(N) \to \infty \) as \( N \to \infty \), and the rate of divergence is sufficiently slow. (The same assumption applies for \( K(N) \).)

Then, consistency and asymptotic normality follow.

1. \( \hat{\theta} \xrightarrow{p} \theta_0 \) as \( N \to \infty \).
2. \( \sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V) \) as \( N \to \infty \).
The asymptotic covariance matrix $V$ is expressed as

$$V = B^{-1} \Sigma B^{-1}.$$ 

We can construct consistent estimators for $B$ and $\Sigma$, and hence

$$\hat{V} = \hat{B}^{-1} \hat{\Sigma} \hat{B}^{-1} \xrightarrow{p} V.$$ 

See the main paper for a complete set of assumptions, proofs of the consistency and asymptotic normality, and the construction of $V$ and $\hat{V}$. 
A practical question is how to select \( \{K_1, \ldots, K_d, K\} \), the dimensions of approximation sieves.

Focus on \( K_j \) for concreteness. We select an optimal \( K_j^* \) from a viewpoint of **covariate balancing**.

A natural estimator of the joint distribution of \( X \), based on the whole group of individuals, is

\[
\hat{F}_X(x) = \frac{1}{N} \sum_{i=1}^{N} I(X_{1i} \leq x_1, \ldots, X_{ri} \leq x_r).
\]

An alternative estimator based on the observed group of individuals is

\[
\hat{F}_{X,K_j}(x) = \sum_{i=1}^{N} T_{ji} \hat{p}_{j,K_j}(X_i) \cdot I(X_{1i} \leq x, \ldots, X_{ri} \leq x_r).
\]
On one hand, we should select $K_j$ that makes $\hat{F}_X$ and $\hat{F}_{X,K_j}$ as close as possible. 

On the other hand, we should impose a penalty against raising $K_j$ in order to keep parsimony.

Hence we propose to select $K_j^* \in \{1, 2, \ldots, K\}$ that minimizes a penalized $L^2$-distance:

$$d_{K_j}(\hat{F}_X, \hat{F}_{X,K_j}) = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{\hat{F}_X(X_i) - \hat{F}_{X,K_j}(X_i)}{(1 - K_j^2/N)^2} \right)^2.$$

The proposed approach is analogous to the generalized cross-validation of Li and Racine (2007, Sec. 15.2).
Monte Carlo Simulations: DGP

- Target variables are $Y_i = [Y_{1i}, Y_{2i}]^\top (d = 2)$.
- Consider a scalar covariate $X_i$.
- Define $U_i = [U_{1i}, U_{2i}, U_{3i}]^\top = [F_1(Y_{1i}), F_2(Y_{2i}), F_X(X_i)]^\top$.
  - $F_1(\cdot)$ is the marginal distribution of $Y_{1i}$, and we use $N(0, 1)$.
  - $F_2(\cdot)$ is the marginal distribution of $Y_{2i}$, and we use $N(0, 1)$.
  - $F_X(\cdot)$ is the marginal distribution of $X_i$, and we use $N(0, 1)$.
- The inverse distribution functions $F_1^{-1}(\cdot), F_2^{-1}(\cdot), \text{ and } F_X^{-1}(\cdot)$ are known and tractable.
Monte Carlo Simulations: DGP

- **Step 1**: Draw $U_i \overset{i.i.d.}{\sim} Gumbel_3(\gamma_0)$ with $\gamma_0 = 4$ (Kendall’s $\tau = 0.75$).

- **Step 2**: Recover $Y_{1i} = F_1^{-1}(U_{1i})$, $Y_{2i} = F_2^{-1}(U_{2i})$, and $X_i = F_X^{-1}(U_{3i})$.

- **Step 3**: Assume that $\{Y_{11}, \ldots, Y_{1N}\}$ are all observed. Make some of $\{Y_{21}, \ldots, Y_{2N}\}$ missing according to

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

  We use $(a, b) = (-0.420, 0.400)$, which implies MAR with $\mathbb{E}[T_{2i}] = 0.6$.

- **Step 4**: Repeat Steps 1-3 $J = 1000$ times with sample size $N \in \{250, 500\}$. 
Approach #1: Listwise deletion.

Step 1: Estimate the marginal distribution of the $j$-th component by

$$
\hat{F}_j(y) = \frac{1}{N^* + 1} \sum_{i=1}^{N} I(T_{1i} = 1, T_{2i} = 1)I(Y_{ji} < y),
$$

where $N^* = \sum_{i=1}^{N} I(T_{1i} = 1, T_{2i} = 1)$ is the number of individuals with complete data.
**Step 2:** Compute the maximum likelihood estimator $\hat{\gamma}$ by

$$
\max_{\gamma \in (1, \infty)} \frac{1}{N^*} \sum_{i=1}^{N} I(T_{1i} = 1, T_{2i} = 1) \log c_2 \left( \hat{F}_1(Y_{1i}), \hat{F}_2(Y_{2i}); \gamma \right),
$$

where

$$
c_2(u_1, u_2; \gamma) = \exp \left[ - \left\{ \sum_{k=1}^{2} (\log u_k)^\gamma \right\}^{\frac{1}{\gamma}} \right]
$$

is the probability density function of $Gumbel_2(\gamma)$.

This likelihood function is correctly specified since any bivariate marginal distribution of $Gumbel_3(\gamma)$ is $Gumbel_2(\gamma)$.
Monte Carlo Simulations: Estimation

- **Approach #2**: Parametric estimation.
- Consider a correctly specified model for the propensity score:
  \[
  \pi_2(x; a, b) = \frac{1}{1 + \exp(a + bx)}.
  \]
- We estimate \((a, b)\) via
  \[
  \max \sum_{i=1}^{N} \left[ T_{2i} \log \pi_2(X_i; a, b) + (1 - T_{2i}) \log (1 - \pi_2(X_i; a, b)) \right].
  \]
Monte Carlo Simulations: Estimation

- Compute

\[ \hat{p}_2(X_i) = \hat{q}(X_i) = \frac{1}{\pi_2(X_i; \hat{a}, \hat{b})}. \]

- Estimate marginal distributions by

\[ \hat{F}_j(y) = \frac{1}{N} \sum_{i=1}^{N} I(T_{ji} = 1)\hat{p}_j(X_i)I(Y_{ji} < y). \]

- Estimate the copula parameter \( \gamma \) via

\[ \max_{\gamma \in (1, \infty)} \frac{1}{N} \sum_{i=1}^{N} I(T_{1i} = 1, T_{2i} = 1)\hat{q}(X_i) \log c_2 \left( \hat{F}_1(Y_{1i}), \hat{F}_2(Y_{2i}); \gamma \right). \]
For comparison, consider a misspecified model:

\[
\pi_2(x; b) = \frac{1}{1 + \exp(bx)}.
\]

This model is misspecified since the true value is

\[a = -0.420 \neq 0.\]

The remaining procedure is the same.
Monte Carlo Simulations: Estimation

- Define the approximation sieve: 
  \[ u_K(X_i) = [1, X_i, \ldots, X_i^{K-1}]^\top. \]
- Define 
  \[ \pi_{2K}(X_i; \lambda) = \frac{1}{1 + \exp[-\lambda^\top u_K(X_i)]}. \]
- Estimate \( \lambda \) via 
  \[
  \max \sum_{i=1}^{N} \left[ T_{2i} \log \pi_{2K}(X_i; \lambda) + (1 - T_{2i}) \log (1 - \pi_{2K}(X_i; \lambda)) \right].
  \]
- Compute 
  \[ \hat{p}_{2K}(X_i) = \hat{q}_K(X_i) = 1/\pi_{2K}(X_i; \hat{\lambda}). \]
- We use \( K \in \{3, 4\} \) and the data-driven selection of \( K \) with upper bound \( \bar{K} = 5 \).
Approach #4: Calibration estimation.

Calibration weights are computed with Exponential Tilting:

\[ \rho(v) = -\exp(-v). \]

We use \( K \in \{3, 4\} \) and the data-driven selection of \( K \) with upper bound \( \bar{K} = 5 \).

The procedure is as explained.
Monte Carlo Simulations: Results

<table>
<thead>
<tr>
<th></th>
<th>$N = 250$</th>
<th>$N = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Truth: $\gamma_0 = 4.000$</strong></td>
<td><strong>Bias, Stdev, RMSE</strong></td>
<td><strong>Bias, Stdev, RMSE</strong></td>
</tr>
<tr>
<td>Listwise Deletion</td>
<td>-0.243, 0.348, 0.424</td>
<td>-0.256, 0.244, 0.353</td>
</tr>
<tr>
<td>Param (Correct)</td>
<td>-0.313, 0.323, 0.450</td>
<td>-0.153, 0.223, 0.271</td>
</tr>
<tr>
<td>Param (Misspec)</td>
<td><strong>-2.610</strong>, 0.204, 2.618</td>
<td><strong>-2.620</strong>, 0.124, 2.623</td>
</tr>
<tr>
<td>Nonparam ($K = 3$)</td>
<td>-0.318, 0.315, 0.448</td>
<td>-0.165, 0.233, 0.285</td>
</tr>
<tr>
<td>Nonparam ($K = 4$)</td>
<td><strong>-0.645</strong>, 0.923, 1.126</td>
<td><strong>-0.682</strong>, 1.081, 1.278</td>
</tr>
<tr>
<td>Nonparam (CB)</td>
<td>-0.323, 0.308, 0.446</td>
<td>-0.140, 0.224, 0.264</td>
</tr>
<tr>
<td>Calibration ($K = 3$)</td>
<td><strong>-0.140</strong>, 0.324, 0.353</td>
<td><strong>-0.093</strong>, 0.234, 0.252</td>
</tr>
<tr>
<td>Calibration ($K = 4$)</td>
<td><strong>-0.129</strong>, 0.342, 0.366</td>
<td><strong>-0.094</strong>, 0.243, 0.260</td>
</tr>
<tr>
<td>Calibration (CB)</td>
<td><strong>-0.142</strong>, 0.333, 0.362</td>
<td><strong>-0.100</strong>, 0.240, 0.260</td>
</tr>
</tbody>
</table>
Monte Carlo Simulations: Results

- Listwise deletion produces bias under MAR as expected.
- The parametric approach produces enormous bias if a propensity score model is misspecified.
- The nonparametric approach exhibits considerable instability across $K$.
- The proposed calibration approach achieves strikingly sharp and stable performance.
- The data-driven selection of $K$ performs well for both the nonparametric and calibration estimators.
Conclusions

- We investigate the estimation of **semiparametric copula models under missing data** for the first time in the literature.
- There is analogy between missing data and **average treatment effects** since both of them are binary in nature.
- We apply Chan, Yam, and Zhang’s (2016) **calibration estimation** to missing data for the first time in the literature.
- The calibration estimator satisfies **consistency** and **asymptotic normality** under the *i.i.d.* assumption and the **Missing at Random (MAR)** condition.
- Our simulation results indicate that the calibration estimator dominates listwise deletion, parametric approach, and nonparametric approach.
References


