Calibration Estimation of Semiparametric Copula Models with Data Missing at Random

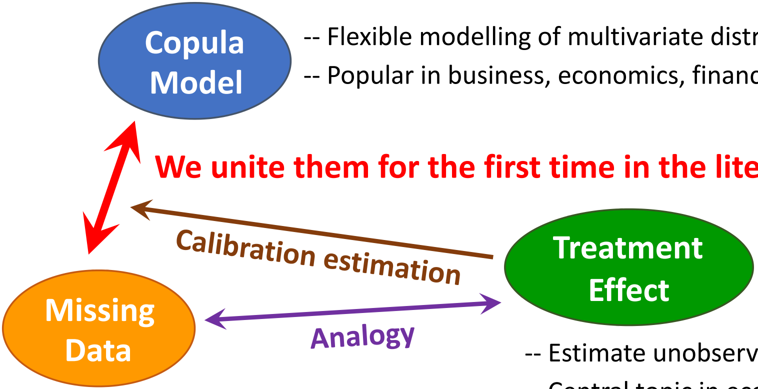
Shigeyuki Hamori[1] Kaiji Motegi[1] Zheng Zhang[2]

[1]Kobe University
[2]Renmin University of China

Econometrics Workshop
Keio University

July 10, 2018

Introduction



Copula Model:

-- Flexible modelling of multivariate distributions

-- Popular in business, economics, finance, etc.

Treatment Effect:

-- Estimate unobserved outcome

-- Central topic in econometrics

Missing Data:

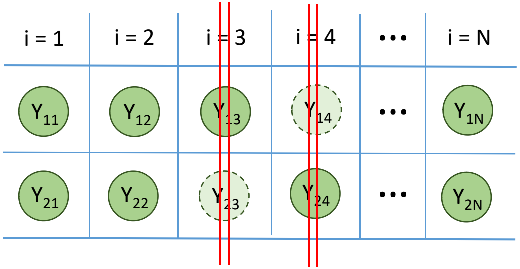
-- Common problem in all fields

-- Deep literature in statistics

-- **Binary phenomenon** (to observe or not to observe)

Introduction

- 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.



Introduction

- 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 $\boldsymbol{Y}_i = [Y_{1i}, \ldots, Y_{di}]^\top$ and their missing status are independent of each other given observed covariates $\boldsymbol{X}_i = [X_{1i}, \ldots, X_{mi}]^\top$.
- LD treats individuals with complete data all equally, and it does not use the information of $\boldsymbol{X}_i$. That can cause **substantial bias** under MAR.

Introduction

- How to obtain a consistent estimator for the copula parameter when the missing mechanism is MAR?
- As is well known in the literature of missing data and average treatment effects, 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 often **lack stability**.

Introduction

- Chan, Yam, and Zhang (2016, JRSS-B) propose an alternative approach called **calibration estimation** in the literature of **average treatment effects**.
- The calibration estimator is derived by balancing covariates among treatment, control, and whole groups. It does **not** require a direct estimation of propensity score.
- We apply the calibration estimation to a missing data problem for the first time in the literature.

Introduction

- 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.
- We perform Monte Carlo simulations. Our simulation results indicate that the calibration estimator **dominates** listwise deletion, parametric approach, and nonparametric approach.

Table of Contents

1) Introduction
2) Review of Copula Models
3) Review of Missing Data
4) Set-up of Main Problem
5) Theory of Calibration Estimation
6) Monte Carlo Simulations
7) Conclusions

Review of Copula Models

- Suppose that there are N individuals and d components:

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

- Suppose that we want to estimate the d-dimensional joint distribution of $\boldsymbol{Y}_i$, assuming $i.i.d.$ What can we do?
- There are potential problems about estimating the joint distribution directly.

	Parametric	Nonparametric
d = small	Misspecification	(Curse of dimensionality)
d = large	Misspecification Parameter proliferation	Curse of dimensionality

Review of Copula Models

- 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.

Step 1 →	Step 2	Name of model
Parametric →	Parametric	Parametric copula model
Nonparametric →	**Parametric**	Semiparametric copula model (Target of our paper)
Nonparametric →	Nonparametric	Nonparametric copula model

Review of Copula Models

- The copula approach is justified by Sklar's (1959) theorem.
- Sklar's theorem ensures the existence of a unique copula function $C : (0,1)^d \to (0,1)$ that recovers a true joint distribution.

Theorem (Sklar, 1959)

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

$$F(y_1, \ldots, y_d) = C\left(F_1(y_1), \ldots, F_d(y_d)\right),$$

or in terms of probability density functions,

$$f(y_1, \ldots, y_d) = c\left(f_1(y_1), \ldots, f_d(y_d)\right).$$

Review of Copula Models

- A well-known example of copula function: bivariate **Clayton** copula with a scalar parameter $\alpha > 0$.
- Cumulative distribution function is

$$C_2(u_1, u_2;\, \alpha) = (u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-\frac{1}{\alpha}},$$

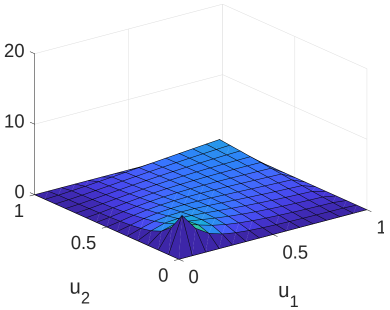
where $u_1 = F_1(y_1) \in (0,1)$ and $u_2 = F_2(y_2) \in (0,1)$ are marginal distribution functions of Y_{1i} and Y_{2i}, respectively.

- Probability density function is

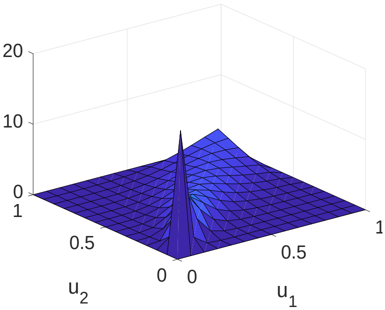
$$c_2(u_1, u_2;\, \alpha) = (1+\alpha)(u_1 u_2)^{-\alpha-1}(u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-\frac{1}{\alpha}-2}.$$

- **Kendall's rank correlation coefficient** is $\tau = \alpha/(\alpha+2)$.
- Larger α implies stronger association between Y_{1i} and Y_{2i}.

Review of Copula Models



p.d.f. ($\alpha = 1.333$ or $\tau = .4$)



p.d.f. ($\alpha = 4.667$ or $\tau = .7$)

- Association is **stronger in the lower tail** than in the upper tail.
- Such an asymmetry matches many economic and financial phenomena (e.g. stock market contagion).
- Larger α implies stronger association between Y_{1i} and Y_{2i}.

Review of Missing Data

- 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.
- There are three well-known layers of missing mechanism.
  1. Missing Completely at Random (MCAR).
  2. Missing at Random (MAR).
  3. Missing Not at Random (MNAR).

Review of Missing Data

- Each concept is defined as follows.

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

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

2. **Missing at Random (MAR)**:

$$\{T_{1i}, \ldots, T_{di}\} \perp \{Y_{1i}, \ldots, Y_{di}\} \mid \{X_{1i}, \ldots, X_{mi}\}.$$

3. **Missing Not at Random (MNAR)**:

$$\{T_{1i}, \ldots, T_{di}\} \not\perp \{Y_{1i}, \ldots, Y_{di}\} \mid \{X_{1i}, \ldots, X_{mi}\}.$$

Review of Missing Data

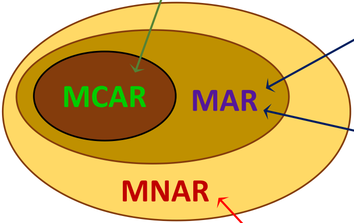
- An illustrative example on health survey ($d = m = 1$):

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

- MCAR requires that $\mathbb{P}$[Individual i reports his/her weight] should be independent of **both** weight **and** gender of individual i.
- MAR requires that:
  - $\mathbb{P}$[A man reports his weight] should be independent of his weight.
  - $\mathbb{P}$[A woman reports her weight] should be independent of her weight.
- MNAR allows for the following situations:
  - $\mathbb{P}$[A man reports his weight] depends on his weight.
  - $\mathbb{P}$[A woman reports her weight] depends on 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.

Set-up of Main Problem

- Semiparametric copula models are estimated in two steps:
- **Step 1**: Estimate the marginal distributions $\{F_1, \dots, F_d\}$ 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 $\boldsymbol{\theta}_0$ via

$$\boldsymbol{\theta}_0 = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \ \mathbb{E}\left[\log c(F_1(Y_{1i}), \dots, F_d(Y_{di}); \boldsymbol{\theta})\right].$$

- If data were all observed, then we could simply replace the population means with sample means.
- 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:

$$\pi_j(\boldsymbol{x}) = \mathbb{P}(T_{ji} = 1 \,|\, \boldsymbol{X}_i = \boldsymbol{x}), \quad j \in \{1, \ldots, d\}.$$

- Define $p_j(\boldsymbol{x}) = 1/\pi_j(\boldsymbol{x})$.
- Step 1 is rewritten as

$$\begin{aligned}
F_j(y) &= \mathbb{E}[I(Y_{ji} \leq y)] = \mathbb{E}[\mathbb{E}[I(Y_{ji} \leq y)|\boldsymbol{X}_i]] \qquad (\because \textbf{LIE}) \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{T_{ji}}{\pi_j(\boldsymbol{X}_i)}\middle|\boldsymbol{X}_i\right] \times \mathbb{E}\left[I(Y_{ji} \leq y)|\boldsymbol{X}_i\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{T_{ji}}{\pi_j(\boldsymbol{X}_i)} \times I(Y_{ji} \leq y)\middle|\boldsymbol{X}_i\right]\right] \qquad (\because \textbf{MAR}) \\
&= \mathbb{E}\left[\frac{T_{ji}}{\pi_j(\boldsymbol{X}_i)} \times I(Y_{ji} \leq y)\right] \qquad (\because \textbf{LIE}) \\
&= \mathbb{E}[T_{ji} \times p_j(\boldsymbol{X}_i) \times I(Y_{ji} \leq y)].
\end{aligned}$$

Set-up of Main Problem

- We have derived:

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

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

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

- If $p_j(\boldsymbol{x})$ were known, then it would be straightforward to compute the IPW estimator.
- $p_j(\boldsymbol{x})$ is **unknown** in reality.

Set-up of Main Problem

- Define a propensity score function:

$$\eta(\boldsymbol{x}) = \mathbb{P}(T_{1i} = 1, \ldots, T_{di} = 1 \,|\, \boldsymbol{X}_i = \boldsymbol{x}).$$

- Define $q(\boldsymbol{x}) = 1/\eta(\boldsymbol{x})$.
- Using MAR and LIE, Step 2 is rewritten as

$$\boldsymbol{\theta}_0 = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathbb{E}\left[I(T_{1i} = 1, ..., T_{di} = 1) q(\boldsymbol{X}_i) \log c(F_1(Y_{1i}), ..., F_d(Y_{di}); \boldsymbol{\theta})\right].$$

- The IPW estimator for $\boldsymbol{\theta}_0$ is given by

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

- $q(\boldsymbol{x})$ is **unknown** in reality.

Set-up of Main Problem

- Estimation of propensity score functions has been a major issue in the literature of missing data and treatment effects.
- Many papers attempt a **direct** estimation of $p_j(\boldsymbol{x})$ and $q(\boldsymbol{x})$, either parametrically or nonparametrically.

- **Parametric approaches**: Zhao and Lipsitz (1992), Robins, Rotnitzky, and Zhao (1994), and Bang and Robins (2005).
- The parametric approaches are notoriously sensitive to misspecification (cf. Lawless, Kalbfleisch, and Wild, 1999).

- **Nonparametric approaches**: Hahn (1998), Hirano, Imbens, and Ridder (2003), Imbens, Newey, and Ridder (2005), and Chen, Hong, and Tarozzi (2008).
- The nonparametric approaches have a notoriously poor performance in finite sample.

Theory of Calibration Estimation

- 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:

$$\mathbb{E}\left[I(T_{ji}=1)p_j(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i)\right] = \mathbb{E}[u_K(\boldsymbol{X}_i)], \quad j \in \{1, \dots d\}$$

for any integrable function $u_K : \mathbb{R}^m \to \mathbb{R}^K$ called an **approximation sieve**. A common choice is, say,

$$u_K(X_i) = [1, X_i, X_i^2, X_i^3]^\top \qquad (m = 1, K = 4).$$

- A sample counterpart is written as

$$\frac{1}{N}\sum_{i=1}^{N} I(T_{ji}=1) \times p_j(\boldsymbol{X}_i) \times u_K(\boldsymbol{X}_i) = \frac{1}{N}\sum_{i=1}^{N} u_K(\boldsymbol{X}_i).$$

- There are multiple values of $\{p_j(\boldsymbol{X}_1), \dots, p_j(\boldsymbol{X}_N)\}$ that 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.
- Primal problem is a **constrained** optimization problem: minimize the distance s.t. the moment matching condition.
- Dual problem is written as an **unconstrained** optimization problem.
- To express the dual problem, let $\rho : \mathbb{R} \to \mathbb{R}$ be any strictly concave function.

Theory of Calibration Estimation

- Define a concave objective function:

$$G_{jK}(\boldsymbol{\lambda}) = \frac{1}{N}\sum_{i=1}^{N}\left[I(T_{ji}=1)\rho(\boldsymbol{\lambda}^\top u_K(\boldsymbol{X}_i)) - \boldsymbol{\lambda}^\top u_K(\boldsymbol{X}_i)\right], \quad \boldsymbol{\lambda} \in \mathbb{R}^K.$$

- Compute

$$\hat{\boldsymbol{\lambda}}_{jK} = \arg\max_{\boldsymbol{\lambda}} G_{jK}(\boldsymbol{\lambda}).$$

- Compute calibration weights for marginal distributions:

$$\hat{p}_{jK}(\boldsymbol{X}_i) = \rho'(\hat{\boldsymbol{\lambda}}_{jK}^\top u_K(\boldsymbol{X}_i)).$$

- Estimate the marginal distribution of the j-th component by

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

Theory of Calibration Estimation

- Arbitrariness of ρ arises from the arbitrariness of the distance measure. Functional forms often used in the nonparametric literature include:
    - Exponential Tilting: $\rho(v) = -\exp(-v)$.
    - Empirical Likelihood: $\rho(v) = \log(1+v)$.
    - Quadratic: $\rho(v) = -0.5(1-v)^2$.
    - Inverse Logistic: $\rho(v) = v - \exp(-v)$.

Theory of Calibration Estimation

- Step 2 (likelihood maximization) can be handled analogously.
- Under MAR, the **moment matching condition** holds:

$$\mathbb{E}\left[I(T_{1i}=1,\ldots,T_{di}=1)q(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i)\right]=\mathbb{E}[u_K(\boldsymbol{X}_i)],$$

  where

$$q(\boldsymbol{X}_i)\equiv\frac{1}{\eta(\boldsymbol{X}_i)}=\frac{1}{\mathbb{P}(T_{1i}=1,\ldots,T_{di}=1\,|\,\boldsymbol{X}_i=\boldsymbol{x})}.$$

- Find calibration weights that satisfy the moment matching condition and are closest to the uniform weight given some distance measure.

Theory of Calibration Estimation

- Define a concave objective function:

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

- Compute

$$\hat{\boldsymbol{\beta}}_K = \arg\max_{\boldsymbol{\beta}} \; H_K(\boldsymbol{\beta}).$$

- Compute a calibration weight for the likelihood:

$$\hat{q}_K(\boldsymbol{X}_i) = \rho'(\hat{\boldsymbol{\beta}}_K^\top u_K(\boldsymbol{X}_i)).$$

- Compute the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ via

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

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).*
- $\{\boldsymbol{Y}_i, \boldsymbol{T}_i, \boldsymbol{X}_i\}$ *are* $i.i.d.$ *across individuals* $i \in \{1, \dots, N\}$.
- $\pi_1(\cdot)$, ..., $\pi_d(\cdot)$, *and* $\eta(\cdot)$ *are* s*-times continuously differentiable with sufficiently large* s.
- $K(N) \to \infty$ *as* $N \to \infty$, *and the rate of divergence is sufficiently slow.*

Then, consistency and asymptotic normality follow.

1. $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$ *as* $N \to \infty$.
2. $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{V})$ *as* $N \to \infty$.

Theory of Calibration Estimation

- The asymptotic covariance matrix $\boldsymbol{V}$ is expressed as

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

- We can construct consistent estimators for $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$, and hence

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

- See the main paper for a complete set of assumptions, proofs of the consistency and asymptotic normality, and the construction of $\boldsymbol{V}$ and $\hat{\boldsymbol{V}}$.

Monte Carlo Simulations: DGP

- Target variables are $\boldsymbol{Y}_i = [Y_{1i}, Y_{2i}]^\top$ $(d = 2)$.
- Consider a scalar covariate X_i.
- Define $\boldsymbol{U}_i = [U_{1i}, U_{2i}, U_{3i}]^\top = [F_1(Y_{1i}), F_2(Y_{2i}), F_X(X_i)]^\top$.
  - $F_1(\cdot)$ 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 $F_1^{-1}(\cdot)$, $F_2^{-1}(\cdot)$, and $F_X^{-1}(\cdot)$ are known and tractable.

Monte Carlo Simulations: DGP

- **Step 1**: Draw $\boldsymbol{U}_i \overset{i.i.d.}{\sim} Clayton_3(\alpha_0)$ with $\alpha_0 = 4.667$.
- **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 $\{Y_{11}, \dots, Y_{1N}\}$ are all observed. Make some of $\{Y_{21}, \dots, Y_{2N}\}$ missing according to

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

  where (a, b) are to be chosen below. Having $b = 0$ implies MCAR, while having $b \neq 0$ implies MAR.
- **Step 4**: Repeat Steps 1-3 $J = 1000$ times with sample size $N = 1000$.

Monte Carlo Simulations: DGP

- We consider two cases for (a, b):
  - **MCAR**: $(a, b) = (-0.405, 0.000) \Longrightarrow \mathbb{E}[T_{2i}] = 0.6.$
  - **MAR**: $(a, b) = (-0.420, 0.400) \Longrightarrow \mathbb{E}[T_{2i}] = 0.6.$
- In both cases, 40% of $\{Y_{21}, ..., Y_{2N}\}$ are missing on average.
- Missing mechanisms are different – MCAR vs. MAR.

Monte Carlo Simulations: Estimation

- **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.

Monte Carlo Simulations: Estimation

- **Step 2**: Compute the maximum likelihood estimator $\hat{\alpha}$ by

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

  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 $Clayton_2(\alpha)$.

- Bias should not arise under MCAR but should arise under MAR.

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

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

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 α via

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

Monte Carlo Simulations: Estimation

- For comparison, consider a misspecified model:

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

- This model is misspecified since $a \neq 0$ for both of the MCAR and MAR cases.
- The remaining procedure is the same.

Monte Carlo Simulations: Estimation

- **Approach #3**: Nonparametric estimation of Hirano, Imbens, and Ridder (2003).
- The approximation sieve $u_K(X_i)$ is chosen to be

$$u_K(X_i) = [1, X_i, X_i^2, X_i^3]^\top \quad (K = 4).$$

- Define

$$\pi_{2K}(X_i; \boldsymbol{\lambda}) = \frac{1}{1 + \exp[-\boldsymbol{\lambda}^\top u_K(X_i)]}.$$

- Estimate $\boldsymbol{\lambda}$ via

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

- Compute

$$\hat{p}_{2K}(X_i) = \hat{q}_K(X_i) = \frac{1}{\pi_{2K}(X_i; \hat{\boldsymbol{\lambda}})}.$$

Monte Carlo Simulations: Estimation

- The nonparametric approach with $K = 2$ is essentially identical to the parametric approach with the correctly specified model.
- This is just a coincidence given that the true missing mechanism obeys a logistic function.
- In theory, $K \to \infty$ as $N \to \infty$ and the nonparametric approach leads to a consistent estimator for any missing mechanism.
- Finite sample performance is another question.

Monte Carlo Simulations: Estimation

- **Approach #4**: Calibration estimation.
- Calibration weights are computed with Exponential Tilting:

$$\rho(v) = -\exp(-v).$$

- The approximation sieve $u_K(X_i)$ is chosen to be

$$u_K(X_i) = [1, X_i, X_i^2, X_i^3]^\top \quad (K = 4).$$

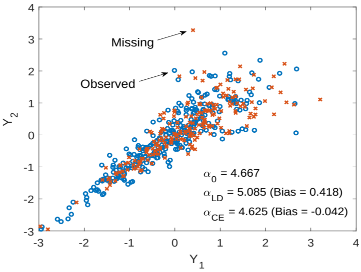
- The procedure is as explained.

Monte Carlo Simulations: Results

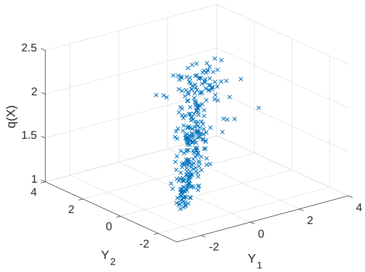
	MCAR	MAR
Truth: $\alpha_0 = 4.667$	Bias, Stdev, RMSE	Bias, Stdev, RMSE
Listwise Deletion	-0.019, 0.302, 0.303	**0.366**, 0.320, 0.486
Param (Correct)	-0.100, 0.280, 0.297	-0.046, 0.262, 0.266
Param (Misspec)	**-0.865**, 0.265, 0.904	**-0.731**, 0.234, 0.768
Nonparam ($K = 4$)	**-0.749**, 1.793, 1.943	**-0.615**, 2.186, 2.271
Calib. Est. ($K = 4$)	**-0.067**, **0.271**, **0.279**	**-0.041**, **0.261**, **0.264**

- Listwise deletion does not cause bias under MCAR, but does cause bias under MAR.
- Parametric approach results in bias if the model is misspecified.
- Nonparametric approach has a poor finite sample performance.
- **Calibration estimation performs well whether data are MCAR or MAR**.

Monte Carlo Simulations: Results



A MC Sample ($N = 500$)



Calibration Weights

- Clayton has lower-tail dependence and upper-tail independence.
- Missing data arise more often at the upper tail and hence the association appears to be stronger than what it is.
- Hence listwise deletion results in **positive** bias.
- CE avoids bias by putting larger weights at the upper tail.

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 observing or not observing data is a binary phenomenon.
- Chan, Yam, and Zhang (2016) propose the **calibration estimation** for average treatment effects.
- We apply the calibration estimation to missing data for the first time in the literature.

Conclusions

- The calibration estimator satisfies **consistency** and **asymptotic normality** under some assumptions including $i.i.d.$ data and the **missing at random (MAR)** mechanism.
- We also derive a consistent estimator for the asymptotic covariance matrix.
- In view of the simulation results, the calibration estimator **dominates** listwise deletion, parametric approach, and nonparametric approach.

References

- 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., H. Hong, and A. Tarozzi (2008). Semiparametric Efficiency in GMM Models with Auxiliary Data. *The Annals of Statistics, 36*, 808-843.
- Hahn, J. (1998). On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects. *Econometrica, 66*, 315-331.

References

- 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. K. Newey, and G. Ridder (2005). Mean-square-error Calculations for Average Treatment Effects. IEPR Working Paper No. 05.34.
- Lawless, J. F., J. D. Kalbfleisch, and C. J. Wild (1999). Semiparametric Methods for Response-Selective and Missing Data Problems in Regression. *Journal of the Royal Statistical Society, Series B, 61*, 413-438.

References

- 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.
- Rubin, D. B. (1976). Inference and Missing Data. *Biometrika, 63*, 581-592.
- 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.
- Zhao, L. P. and S. Lipsitz (1992). Designs and Analysis of Two-Stage Studies. *Statistics in Medicine, 11*, 769-782.