The Korean Economic Review
Volume 32, Number 2, Winter 2016, 239-268.

AMH Copula ML Estimation for the Sample Selection Model

Hosin Song*

In this paper, we propose a copula ML estimation method for the sample selection model using the Ali-Mikhail-Haq (AMH) copula. The proposed AMH copula ML estimation is compared with the well-known bivariate ML estimation and Heckman's two-step estimation. Monte Carlo experiments are conducted to compare their performance in terms of the mean squared error (MSE) depending on the following 2 conditions: (i) whether the imposed distributional assumption is correct, and (ii) whether some regressors of the participation and outcome equation are correlated. The results of the experiments show that the estimation results for the proposed method can be better than those of the two well-known methods, particularly when the imposed distributional assumption is incorrect and some regressors of the two equations are correlated. Hence, the proposed method can be a practically useful alternative for the sample selection model.

JEL Classification: C14, C15, C18, C24
Keywords: sample selection, Ali-Mikhail-Haq copula ML, bivariate ML, Heckman's two-step estimation

I. Introduction

The bivariate normal maximum likelihood (BN ML) estimation method by Heckman (1974, 1978) and the two-step estimation method, also by Heckman (1979), are well-known parametric methods for the sample selection model where the random sample assumption fails. Heckman (1974, 1978) proposed the ML estimation method by imposing bivariate normal assumptions on the two error terms of the participation equation and the outcome equation. Heckman (1979) also proposed to use the two-step estimation method to eliminate the sample selection bias introduced by the usual least squares (LS) estimation of the outcome

Received: April 30, 2015. Revised: Nov. 24, 2015. Accepted: Aug. 22, 2016.
* Associate Professor, Department of Economics, Ewha Womans University, Seoul 120-750, Korea, Email address: hsong@ewha.ac.kr. The previous version of the paper was presented at the 25th annual meeting of the Midwest Econometrics Group at the Federal Reserve Bank of St. Louis.

equation when the two errors are correlated. Such a two-step estimation procedure is referred to as "Heckit". Both parametric methods have been the most extensively used to estimate the sample selection model, and these depend on the critical assumption that the two error terms are jointly normally distributed.

The joint distribution of the two error terms in the selection model can be approximated in flexible way by using copula. In this paper, the Ali-Mikhail-Haq (AMH) copula ML method is proposed for the sample selection model as an exemplary copula ML, and its validity in terms of MSE is assessed compared to BN ML and Heckit. To assess the validity of the proposed method over the two well-known estimation methods, some Monte Carlo experiments are conducted. Two conditions are considered for the Monte Carlo experiments. One is whether the imposed assumption on the joint distribution is correct or not, and the other is whether some regressors of the two equations are correlated or not. The correlation among the regressors of the two equations is considered in terms of the near multicollinearity problem in Heckit as pointed out by Nawata (1993) and Nawata and Nagase (1996). The results of the Monte Carlo experiments show that the proposed estimation method can be a good alternative to estimate the selection model based on the mean squared error (MSE). In particular, when the imposed distributional assumption is not correct and some regressors of the two equations are correlated, the proposed method shows the smallest MSE.

Many studies have been published on the copula. Chen, Fan and Tsyrennikov (2006) used the copula to explore efficient sieve estimation. They showed that a plug-in copula maximum likelihood estimator for all smooth functionals with unknown marginal distributions is semiparametrically efficient. Zimmer and Trivedi (2006) extended the bivariate copula to the trivariate copula by allowing for two dependence parameters. Oh and Patten (2013) proposed a copula-based simulated method of moments estimation and established the consistency and asymptotic normality of their proposed estimator. In particular, the copula has been actively used in finance and insurance analysis. For example, see Patten (2006), Fan and Gu (2003) and Hu (2006). There have also been many publications about the sample selection model since Heckman (1974, 1976, 1978, 1979).[1] Lee (1983) extended the sample selection model by trying to allow for more flexible marginal distribution, which turned out to be the introduction of the Gaussian copula. Recently, semiparametric and semi-nonparametric works have been actively explored in the selection model literature. See Gallant and Nychka (1987), Vella (1992, 1993), Chen (1997), Honore, Kyriazidou and Udry (1997), and Das, Newey and Vella (2003).

However, there are only a few publications about the selection model using

[1] Cragg (1971) also considered diverse limited dependent variable models, one of which was the same as the selection model.

copula. For example, see Lee (1983) and Smith (2003). The main reason for the scarcity of copula sample selection model literature may be related to the fact that it is not trivial to obtain the joint distributions of the two unobserved error terms via copula. That is a stark contrast with the widespread use of copula for financial market analysis. Usually, financial variables such as asset prices are observed, and thus it is relatively very easy to handle them by using copula. Smith (2003) proposed to use a copula approach for the sample selection model. He tried several copulas to estimate the sample selection model. But, his estimation was concentrated on the copula parameter and Kendall's tau.

This paper is similar to Smith (2003) in that a parametric copula ML is considered. However, we focus on the performance of the proposed copula ML estimation method, and conduct comparisons against two other well-known parametric estimation methods. For that goal, unlike Smith (2003), our approach employs some Monte Carlo experiments instead of using actual observations since we can control the specification of the model depending on the true data generating process in the experiments. The results of the Monte Carlo experiments suggest that the proposed AMH copula ML method can be a practically useful alternative method to the BN ML and Heckit methods to estimate the sample selection model since it exhibits good performance in terms of MSE, particularly when researchers' estimation models are misspecified.

The remaining of the paper is organized as follows. In Section 2, the sample selection model and two well-known estimation methods are addressed. Moreover, the AMH copula ML estimation method is proposed for the selection model. In Section 3, Monte Carlo experiments are conducted to confirm the validity of the proposed method. In Section 4, some concluding remarks are provided.

As to the notations, a bold letter denotes a parameter vector, "$\xrightarrow{d}$" denotes the convergence in distribution, "$\sim$" denotes "distributed to", and the distribution function of a random variable X is denoted by $F_X(\cdot)$. Moreover, uppercase letters denote random variables, while lowercase letters denote realizations.

II. Sample Selection Model

One typical example of the sample selection model is the female worker's wage model. The female worker's wage is the outcome variable, which is observed only when the female person participates in the labor market. In this paper, we consider a typical bivariate error related sample selection model in Heckman (1979). It consists of a participation equation and an outcome equation. The participation equation can be defined as follows.

$$Y_1 = I(Y_1^* > 0) \quad \text{where} \quad Y_1^* = X_1'\boldsymbol{\beta} + V_1, \tag{1}$$

Y_1^* is a latent variable, and $I(\cdot)$ is the indicator function. $Y_1 = 1$ indicates participation, while $Y_1 = 0$ indicates no participation. The outcome Y_2 is observed for the participant, but not for the non-participant. Hence, the outcome equation can be defined as follows.

$$Y_2 = \begin{cases} Y_2^* & \text{if} \quad Y_1 = 1 \\ - & \text{if} \quad Y_1 = 0 \end{cases} \tag{2}$$

where $Y_2^* = X_2'\delta + V_2$ and Y_2^* is a latent variable, and "$-$" indicates that the outcome is missing.[2] Both equations (1)-(2) involve the latent variables Y_1^* and Y_2^*, hence we need some assumptions for the covariate X_j and the error term V_j for each $j = 1,2$.

Assumption 1. *(i) Equations (1) and (2) hold. (ii)* V_1 *and* V_2 *are independent of* (X_1, X_2). *(iii) Each error term* $V_j, j = 1,2$, *is continuously distributed with its density* f_{V_j}.

There are two well-known estimation methods for the selection model. One is the BN ML estimation by Heckman (1974, 1978), and the other is Heckman's two-step estimation, the so called "Heckit", also by Heckman (1979). Both estimation methods depend on the bivariate normal distribution assumption for the two error terms (V_1, V_2).

2.1. Two Well-known Estimation Methods for the Sample Selection Model

Throughout the paper, we suppress the conditioning of X_1 and X_2 for notational convenience.

Assumption 2. *The joint distribution of the two error terms* (V_1, V_2) *given* (X_1, X_2) *is the following bivariate normal distribution.*

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \right).$$

Under Assumptions 1 and 2, the two well-known estimation methods are proposed by Heckman (1978) and Heckman (1979). Both will be addressed briefly in the following two subsubsections.

[2] One may denote "$-$" as zero.

2.1.1. Bivariate Normal (BN) ML Method

Under Assumptions 1-2, the log likelihood function is

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}) \tag{3}$$

where

$$\ell_i = \Big[(1-y_{1i})\log\Phi(-x'_{1i}\boldsymbol{\beta}) + y_{1i}\left(\log\Phi\left(\frac{x'_{1i}\beta + \rho\frac{1}{\sigma_2}(y_{2i}-x'_{2i}\boldsymbol{\delta})}{\sqrt{1-\rho^2}}\right) + \log\phi\left(\frac{y_{2i}-x'_{2i}\boldsymbol{\delta}}{\sigma_2}\right) - \log\sigma_2\right)\Big]$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\delta}', \rho, \sigma_2)'$. Hence, the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{BN}$ is

$$\hat{\boldsymbol{\theta}}_{BN} = \arg\max_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta})$$

where log $L(\boldsymbol{\theta})$ is defined in (3). It is well known that under Assumptions 1-2 and usual ML regularity conditions, $\hat{\boldsymbol{\theta}}_{BN}$ is consistent and asymptotically normal, that is,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{BN} - \boldsymbol{\theta}_{0,BN}) \xrightarrow{d} N(0, V_{BN}) \tag{4}$$

where $V = E[\frac{\partial \ell_1(\boldsymbol{\theta}_{0,BN})}{\partial \boldsymbol{\theta}} \frac{\partial \ell_1(\boldsymbol{\theta}_{0,BN})}{\partial \boldsymbol{\theta}'}]^{-1}$ and $\boldsymbol{\theta}_{0,BN} = (\boldsymbol{\beta}'_0, \boldsymbol{\delta}'_0, \rho_0, \sigma_{0,2})'$. The consistent estimator of the asymptotic variance V is $\hat{V} = (n^{-1}\sum_{i=1}^{n} \frac{\partial \ell_i(\hat{\boldsymbol{\theta}}_{BN})}{\partial \boldsymbol{\theta}} \frac{\partial \ell_i(\hat{\boldsymbol{\theta}}_{BN})}{\partial \boldsymbol{\theta}'})^{-1}$ where $\frac{\partial \ell_i(\hat{\boldsymbol{\theta}}_{BN})}{\partial \boldsymbol{\theta}} = (\frac{\partial \ell_i(\hat{\boldsymbol{\theta}}_{BN})}{\partial \boldsymbol{\beta}'}, \frac{\partial \ell_i(\hat{\boldsymbol{\theta}}_{BN})}{\partial \boldsymbol{\delta}'}, \frac{\partial \ell_i(\hat{\boldsymbol{\theta}}_{BN})}{\partial \rho}, \frac{\partial \ell_i(\hat{\boldsymbol{\theta}}_{BN})}{\partial \sigma_2})'$.

2.1.2. Two-step Estimation Method: Heckit

It is worthwhile to note the following well-known remark.

Remark 1. *Assumption 2 implies* $E[V_2 \mid V_1 = v_1] = \sigma_{12} v_1$.

The proof of the remark is trivial, and thus it is left as an exercise for readers. The remark leads to the following conditional expectation, which motivates the two-step estimation "Heckit". Let $X = (X'_1, X'_2)'$

$$E[Y_2 \mid X, Y_1^* > 0] = X'_2\boldsymbol{\delta} + \sigma_{12}\lambda(X'_1\boldsymbol{\beta}) \tag{5}$$

where $\lambda(t) \equiv \phi(t)/\Phi(t)$, and $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal distribution function and density function, respectively. It follows from (5) that

$$Y_{2i} = X_{2i}'\delta + \sigma_{12}\lambda(X_{1i}'\beta) + \eta_i \tag{6}$$

where $E(\eta_i \mid X_i, Y_{1i} = 1) = 0$. Suppose that the parameter of our interest is δ. Then, the two-step estimation can be applied to obtain a consistent estimator of $\gamma \equiv (\delta', \sigma_{12})'$. In the first stage, $\hat{\beta}_{probit}$ can be obtained by using a probit ML estimation. In the second stage, $\hat{\gamma} = (\hat{\delta}'_{Heckit}, \hat{\sigma}_{12,Heckit})'$ can be obtained by the LS estimation of Y_{2i} on X_{2i} and $\lambda(X_{1i}'\hat{\beta}_{probit})$. The standard error for $\hat{\gamma}$ should be corrected since it depends on the first stage estimator $\hat{\beta}_{probit}$. For the corrected asymptotic normality of the Heckit estimator, see Heckman (1979).

2.2. Ali-Mikhail-Haq (AMH) Copula ML Estimation

The well-known Sklar's theorem (1959) states that there is a unique copula $C(\cdot,\cdot)$ such that

$$F_{V_1V_2}(v_1, v_2) = C(F_{V_1}(v_1), F_{V_2}(v_2)), \tag{7}$$

where $F_{V_1V_2}(\cdot,\cdot)$ is the joint distribution function of the random vector (V_1, V_2), while $F_{V_1}(\cdot)$ and $F_{V_2}(\cdot)$ are strictly increasing marginal distribution functions of continuous random variables V_1 and V_2.[3] Thus, (7) can be defined as

$$F_{V_1V_2}(F_{V_1}^{-1}(u_1), F_{V_2}^{-1}(u_2)) = C(u_1, u_2), \tag{8}$$

where $u_1 = F_{V_1}(v_1)$ and $u_2 = F_{V_2}(v_2)$ since $U_1 = F_{V_1}(V_1)$ and $U_2 = F_{V_2}(V_2)$ are uniformly distributed on [0, 1].

In general, copula $C(u_1, u_2)$ is parameterized to be $C(u_1, u_2; \alpha)$ where the copula parameter α can be a scalar or vector.[4] Once a specific copula is chosen, some parametric distributional assumptions can be imposed on the marginal distributions V_1 and V_2 to implement the copula ML estimation method. Then, the usual parametric ML estimation method can be applied. The likelihood function is

[3] For details of Sklar's theorem, see subsection 2.3 in Nelsen (2006).
[4] In this paper α is a scalar.

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} \{\Pr[Y_{1i}^* \le 0]\}^{1-y_{1i}} \left\{ \int_0^\infty f_{Y_1^* Y_2^*}(y_{1i}^*, y_{2i}) dy_{1i}^* \right\}^{y_{1i}} \\
&= \prod_{i=1}^{n} \{F_{V_1}(-x_{1i}'\boldsymbol{\beta})\}^{1-y_{1i}} \left\{ f_{V_2}(y_{2i} - x_{2i}'\boldsymbol{\delta}) \left[1 - \frac{\partial C(u_{1i}, u_{2i}; \alpha)}{\partial u_{2i}} \right] \right\}^{y_{1i}}
\end{aligned} \tag{9}
$$

where $u_{1i} = F_{V_1}(-x_{1i}'\boldsymbol{\beta})$, $u_{2i} = F_{V_2}(y_{2i} - x_{2i}'\boldsymbol{\delta})$, and $C(u_1, u_2; \alpha)$ is a copula which is involved with two marginal distributions F_{V_1} and F_{V_2}. Hence, $C(u_{1i}, u_{2i}; \alpha) = F_{V_1 V_2}(-x_{1i}'\boldsymbol{\beta}, y_{2i} - x_{2i}'\boldsymbol{\delta})$. Note that $\int_0^\infty f_{Y_1^* Y_2^*}(y_1^*, y_2) dy_1^*$ denotes the likelihood of the event $\{Y_1^* > 0, Y_2 = y_2\}$.[5] The last equation in (9) follows from the following lemma.

Lemma 1. $\int_{-x_1'\boldsymbol{\beta}}^{\infty} f_{V_1 V_2}(v_1, y_2 - x_2'\boldsymbol{\delta}) dv_1 = f_{V_2}(y_2 - x_2'\boldsymbol{\delta})[1 - C_{u_2}(u_1; \alpha)]$ *where* $u_1 = F_{V_1}(v_1)$, $u_2 = F_{V_2}(y_2 - x_2'\boldsymbol{\delta})$, *and* $C_{u_2}(u_1, u_2; \alpha) \equiv \frac{\partial C(u_1, u_2; \alpha)}{\partial u_2}$.

The proof of Lemma 1 is provided in the Appendix. In this paper, we propose to use the AMH copula since it is easy to handle. Moreover, as Kumar (2010) mentioned, the AMH copula is the only one among 22 Archimedean copulas in Nelsen (2006) whose parameter lies on a closed interval $[-1, 1]$ and measures both, positive and negative, dependence.[6] The AMH copula is considered here as one exemplary case for the selection model since there are too many copulas to choose from. It does not mean that AMH copula is the best choice either. AMH copula ML is just one promising example. Readers may consider using other copulas for the selection model depending on the knowledge of the dependence structure. But, those issues are not crucial here since the purpose of this paper is to see the performance of AMH copula ML method for the selection model compared to the two well-known models depending on whether the specification chosen by a researcher is correct or not.

Ali, Mikhail and Haq (1978) proposed searching for bivariate distributions, whose survival odds ratios satisfying

$$
\frac{1 - F_{XY}(x, y)}{F_{XY}(x, y)} = \frac{1 - F_X(x)}{F_X(x)} + \frac{1 - F_Y(y)}{F_Y(y)} + (1 - \alpha) \frac{1 - F_X(x)}{F_X(x)} \cdot \frac{1 - F_Y(y)}{F_Y(y)}
$$

for some constant $\alpha \in [-1, 1]$. Therefore, the AMH copula is defined as follows.

[5] As one referee commented, its probability is zero. Hence, the terminology "likelihood" is more appropriate here than the probability.

[6] Smith (2003) also considered AMH copula.

$$C(u_1, u_2; \alpha) = \frac{u_1 u_2}{1-\alpha(1-u_1)(1-u_2)} \quad \text{for} \quad \alpha \in [-1,1]. \tag{10}$$

There are some remarks worth noting regarding AMH copula. First of all, consider the derivative of the AMH copula with respect to u_2.

$$C_{u_2}(u_1, u_2; \alpha) \equiv \frac{\partial C(u_1, u_2; \alpha)}{\partial u_2} = \frac{u_1[1-\alpha(1-u_1)]}{[1-\alpha(1-u_1)(1-u_2)]^2}. \tag{11}$$

The derivative (11) is used to construct the likelihood function (9). Another important fact is that (11) is the conditional distribution of U_1 given $U_2 = u_2$ as shown in the proof of Lemma 1. The conditional distribution can be used to obtain a random drawing from the copula in the experiments in section 3.[7] The parameter α reflects the dependence between U_1 and U_2. If $\alpha = 0$, then U_1 and U_2 are independent. Hence, so are V_1 and V_2. The last remark about the AMH copula is that it is an Archimedean copula. Archimedean copulas are a particular family of copulas satisfying the following condition:

$$C(u_1, u_2) = \varphi^{[-1]}(\varphi(u_1) + \varphi(u_2))$$

where φ is a generator function, and $\varphi^{[-1]}$ is a pseudo-inverse function.[8] Since AMH copula's generator function is strict, $\varphi^{[-1]}(t) = \varphi^{-1}(t)$ for any $t \in [0,1]$.[9] One well-known feature of an Archimedean copula is that the Kendall's rank correlation can be calculated directly from the generator function.[10]

$$\text{Kendall's rank correlation} = 1 + 4\int_0^1 \frac{\varphi(s)}{\varphi'(s)} ds. \tag{12}$$

For the detailed features of AMH copula as a member of Archimedean copula, see Kumar (2006).

For the AMH copula ML, we need to make some parametric distributional assumptions for the marginals F_{V_1} and F_{V_2}. Let $V_1 \sim \mathcal{N}(0,1)$ and $V_2 \sim \mathcal{N}(0,\sigma^2)$. That is $f_{V_1}(v) = \phi(v)$ and $f_{V_2}(v) = \sigma^{-1}\phi((y_2 - x_2'\delta)/\sigma)$. Therefore, $u_1 = \Phi(-x_1'\beta)$ and $u_2 = \Phi((y_2 - x_2'\delta)/\sigma)$. Then, the likelihood function (9) is

[7] The method is called the "inverse conditional distribution method". For details, see pages 40-41 in Nelsen (2006).

[8] The generator function φ should be a continuous, strictly decreasing function from $[0,1]$ to $[0,\infty]$ with $\varphi(1) = 0$. For the definition of pseudo-inverse, see Definition 4.1.1. in Nelsen (2006).

[9] AMH copula's generator function is $\varphi(t) = \ln((1-\alpha(1-t))/t)$.

[10] See Genest and MacKay (1986) and Nelsen (2006).

$$L(\boldsymbol{\theta})=\prod_{i=1}^{n}\Phi(-x_{1i}'\boldsymbol{\beta})^{1-y_{1i}}\left[\frac{1}{\sigma}\phi((y_{2i}-x_{2i}'\boldsymbol{\delta})/\sigma\right]^{y_{1i}}\left[1-\frac{u_{1i}[1-\alpha(1-u_{1i})]}{[1-\alpha(1-u_{1i})(1-u_{2i})]^2}\right]^{y_{1i}}$$

where $u_{1i}=\Phi(-x_{1i}'\boldsymbol{\beta})$, $u_{2i}=\Phi((y_{2i}-x_{2i}'\boldsymbol{\delta})/\sigma)$ and $\boldsymbol{\theta}=(\boldsymbol{\beta}',\boldsymbol{\delta}',\sigma,\alpha)'$. Hence, the log likelihood function is

$$\log L(\boldsymbol{\theta})=\sum_{i=1}^{n}\{(1-y_{1i})\log\Phi(-x_{1i}'\boldsymbol{\beta})+y_{1i}(\log[\phi((y_{2i}-x_2'\boldsymbol{\delta})/\sigma)]-\log\sigma)$$
$$+y_{1i}\log\left(1-\frac{u_{1i}[1-\alpha(1-u_{1i})]}{[1-\alpha(1-u_{1i})(1-u_{2i})]^2}\right)\Bigg\} \tag{13}$$
$$\text{where } u_{1i}=\Phi(-x_{1i}'\boldsymbol{\beta}),\ u_{2i}=\Phi((y_{2i}-x_{2i}'\boldsymbol{\delta})/\sigma) \text{ and } \boldsymbol{\theta}=(\boldsymbol{\beta}',\boldsymbol{\delta}',\sigma,\alpha)' \tag{14}$$

Recall that we need to restrict the copula parameter α to be in $[-1,1]$. To do that, we consider using the one-to-one bounded transformation $\alpha\equiv\frac{2}{\pi}\arctan(a)$ in the simplex method by Nelder and Mead (1965).[11] Recall $-1\leq\alpha\leq1$ for any $a\in\mathbb{R}$. Hence, searching in terms of α via a makes the estimation more convenient. Therefore, the AMH copula maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{CML}$ is

$$\hat{\boldsymbol{\theta}}_{CML}=\arg\max_{\boldsymbol{\theta}}\log L(\boldsymbol{\theta}) \text{ which is defined in (13).}$$

Under ML regularity conditions including correct distributional assumption, AMH CML estimator is consistent and asymptotically normal.[12] Hence, it is easily obtained that $\sqrt{n}(\hat{\boldsymbol{\theta}}_{CML}-\boldsymbol{\theta}_{0,CML})\xrightarrow{d}\mathcal{N}(0,V_{CML})$ where $V_{CML}=E[\frac{\partial\ell_1(\boldsymbol{\theta}_{0,CML})}{\partial\boldsymbol{\theta}}\frac{\partial\ell_1(\boldsymbol{\theta}_{0,CML})}{\partial\boldsymbol{\theta}'}]^{-1}$ since $\hat{\boldsymbol{\theta}}_{CML}$ is the usual ML estimator of the parameter vector $\boldsymbol{\theta}_{0,CML}=(\boldsymbol{\beta}_0',\boldsymbol{\delta}_0',\sigma_0,\alpha_0)'$.[13]

III. The Performance of the AMH CML Estimator

In this section the performance of the proposed method is compared with the other two well-known parametric estimation methods addressed in the previous

[11] Throughout our experiments, we found that in practice the simplex method worked better than the usual derivative methods such as the Gauss-Newton procedure in that it had a much higher success rate of finding the estimator than the derivative search throughout the Monte Carlo experiments. It was more apparent under misspecification.

[12] The consistency of usual M-estimators can be applied.

[13] Derivatives of $\ell_i(\boldsymbol{\theta})\equiv\log f_i(\boldsymbol{\theta})$ w.r.t. $\boldsymbol{\theta}$ are provided in the Appendix for readers who may be interested in them.

section. Since the analytical comparison of their asymptotic variances can not be determined in general, we conduct some numerical simulations to find the usefulness of the AMH copula maximum likelihood (CML) method even though it may be limited.[14] See the Appendix for the limitation of the asymptotic variance comparison. Hence, Monte Carlo experiments are conducted to find the validity of the proposed AMH CML method relative to the two well-known methods for the sample selection model.

Suppose that we are interested in the parameter δ in the outcome equation. We compare the performance of the proposed AHM CML method and the two well-known methods BN ML and Heckit by comparing the mean squared error (MSE) of the estimation results from 1000 independent replications. Two environments are considered for Monte Carlo experiments. One is the case where some of the three models (AMH CML, BN ML, Heckit) are correctly specified, and the other is the case where none of them is correct.

For the former, two cases are considered. The first one is the case where the true joint distribution of two errors is the bivariate normal distribution so that BN ML or Heckit are correctly specified. The second one is the case where the true joint distribution of two errors is the AMH copula distribution and their marginal distributions are standard normal. In this case, only the AMH ML model is correctly specified.

For the latter, we consider the following two cases. The first case is that the true joint distribution is the Farlie-Gumbel-Morgenstern (FGM) copula and their marginal distributions are standard normal. Note that none of the three models is correct even though the specification of the marginal distribution is correct. The second case is that the true joint distribution is the FGM copula and the marginal distribution of each error is standard logistic distribution. Note that none of the three models is correct under this true DGP.

In each case, we consider four situations depending on the strength of the correlation among the regressors of the participation and outcome equation. The reason to consider the correlation is that there are near multicollinearity problems in Heckman's two-step estimation in the presence of correlation as addressed by Nawata (1993) and Nawata and Nagase (1996).

Throughout all the experiments, suppose that Assumption 1 holds. Specifically, the following assumptions are common to all experiments in this section.

(i) $Y_1 = I(Y_1^* \geq 0)$ where $Y_1^* = \beta_0 + \beta_1 X_1 + V_1$, and $\boldsymbol{\beta}_0 = (\beta_0, \beta_1) = (-1,1)$.

(ii) $Y_2 = Y_2^* I(Y_1 = 1)$ where $Y_2^* = \delta_0 + \delta_1 X_2 + V_2$ and $\boldsymbol{\delta}_0 = (\delta_0, \delta_1) = (20,1)'$.

(iii) $X_1 = \frac{(1-\tau)\xi + \tau X_2}{\sqrt{\tau^2 + (1-\tau)^2}}$ where $\xi \sim \mathcal{N}(0,1)$, $\tau \in [0,1]$ and ξ is independent of X_2.

---

[14] Similar arguments as to the analytical comparison are pointed out by Smith (2003).

(iv) $X_2 \sim \mathcal{U}[0,4]$ where $\mathcal{U}[0,4]$ denotes the uniform distribution on the interval $[0,4]$.

Note $\tau = 0$ implies $X_1 = \xi$, while $\tau = 1$ implies $X_1 = X_2$. Also, note the correlation between X_1 and X_2 is $\frac{\tau}{\sqrt{\tau^2+(1-\tau)^2}}$ which is between 0 and 1.[15] We will try $\tau = 0, 0.5, 0.9$ and 1.

In each experiment with different DGP, the three estimation models are considered as follows.

- For BN ML, do (3) where $\boldsymbol{\theta}_{BN} = (\boldsymbol{\beta}', \boldsymbol{\delta}', \rho, \sigma_2)'$.
- For Heckit, do the probit ML to estimate the participation equation, and estimate the outcome equation (6). Hence, the related parameter is $\boldsymbol{\theta}_{Heckit} = (\boldsymbol{\beta}', \gamma')' = (\boldsymbol{\beta}', \delta', \sigma_{12})'$.
- For AMH copula ML, estimate the model in (13)-(14). Assume that a researcher chooses the following parametric specifications for the marginal distributions: $V_1 \sim \mathcal{N}(0,1)$ and $V_2 \sim \mathcal{N}(0,\sigma^2)$. This assumption on the margins is needed for the comparison between AMH ML and other two well-known methods. Here, $\boldsymbol{\theta}_{CML} = (\boldsymbol{\beta}', \boldsymbol{\delta}', \sigma, \alpha)'$.

In each experiment, the performance of the AMH copula ML and other two well-known methods are examined in terms of the mean squared error (MSE). We focus on the common parameter $(\boldsymbol{\beta}', \boldsymbol{\delta}')$ in the three models. In particular, our primary interest lies in the outcome equation parameter $\boldsymbol{\delta}$. The MSE of the estimator $\hat{\delta}_k$ is defined as $E[(\hat{\delta}_k - \delta_k)^2]$, $k = 0,1$. Therefore, the MSE from the experiment can be estimated as $\frac{1}{R}\sum_{r=1}^{R}(\hat{\delta}_{k,r} - \delta_k)^2$, $k = 0,1$, where $\hat{\delta}_{k,r}$ is the estimator of δ_k from the rth replication, and R is the number of replications. The estimation results will be focused on $MSE(\hat{\delta}_k)$, $k = 0,1$, instead of $MSE(\hat{\beta}_k)$ since there is no substantial difference in $MSE(\hat{\beta}_k)$ results of the three models.[16]

3.1. When some of the Three Models are Correctly Specified

3.1.1. Case 1: When the True Joint Distribution is the Bivariate Normal Distribution

In this case, the BN ML and Heckit model are correctly specified, while the AMH ML model is not correctly specified. For the experiment, the two error terms

[15] This idea follows from Nawata (1993).

[16] For the readers who may be interested in the participation equation parameter $\boldsymbol{\beta}$ and copula parameter α, the estimation results of $MSE(\hat{\beta}_k)$ and $MSE(\hat{\alpha})$ are reported in the Appendix. (Tables 5-9.)

(V_1, V_2) are randomly drawn from the following bivariate normal distribution.

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right). \tag{15}$$

For each replication sample, independent 1000 random drawings of (V_1, V_2) are obtained. Hence, $(Y_{1i}, Y_{2i}, X_{1i}, X_{2i}), i = 1, \ldots, 1000$, are generated according to the procedure (i)-(iv) depending on the value of τ. The true DGP of (V_1, V_2) in (15) is unknown to a researcher. The researcher implements the aforementioned three methods to estimate the selection model by treating the replication sample as an actual sample. 1000 replications are conducted.[17]

Table 1 shows the MSE's of the three models' estimators when the true joint distribution is the bivariate normal distribution. Note that the estimation model can be said to be correctly specified only when either the BN ML or Heckit method is chosen. If the AMH CML model is chosen, the specification will not be correct. It is shown that the BN ML method has the smallest MSE among the three estimators in all values of τ. This result is expected since the ML estimator achieves the Cramer-Rao lower bound when the chosen specification is correct. The smallest MSE of the BN ML estimator is achieved irrespective of the value of τ, which indicates the extent of the dependence between X_1 and X_2. Heckit also has MSE as good as the BN ML when there is no linear correlation between X_1 and X_2. Its MSE is still less than that of the AMH CML when $\tau = 0.5$, but the AMH CML estimator's MSE becomes less than that of the Heckit as the correlation strengthens. Such findings are apparent when $\tau = 0.9$ or $\tau = 1$. That is, the performance of the Heckit is not as good as that of the AMH CML in terms of MSE when X_1 and X_2 are strongly correlated, which is reflected in large variance of the Heckit estimators.

Theoretically, only both BN ML and Heckit estimators are consistent estimators under the DGP of this experiment. But, the Heckit estimator seems to suffer from the multicollinearity problem when $\tau = 0.9$ or $\tau = 1$ in spite of the correct specification. In particular, it suffers much more when $n = 500$ as shown in Table 10 in the Appendix. In that sense, the AMH CML estimation can be a good alternative for the Heckit estimation method when regressors are heavily correlated.

[17] For the readers who may be interested, the results $MSE(\hat{\delta})$'s are provided in the Appendix when the sample size is 500.

[Table 1] Case 1: Estimation of δ when the true joint distribution is bivariate normal

	$\tau=0$		$\tau=0.5$		$\tau=0.9$		$\tau=1$	
parameter	δ_0	δ_1	δ_0	δ_1	δ_0	δ_1	δ_0	δ_1
				BN ML				
MSE	0.0277	0.0027	0.0226	0.0021	0.0475	0.0049	0.0533	0.0055
Bias	0.0023	-0.0012	0.0067	-0.0019	0.0330	-0.0101	0.0395	-0.0122
Variance	0.0277	0.0027	0.0226	0.0021	0.0464	0.0048	0.0518	0.0053
				Heckit				
MSE	0.0289	0.0027	0.0284	0.0024	0.0940	0.0090	0.0997	0.0096
Bias	0.0002	-0.0012	0.0010	-0.0006	-0.0006	-0.0001	-0.0004	-0.0002
Variance	0.0289	0.0027	0.0284	0.0024	0.0940	0.0090	0.0997	0.0096
				AMH				
MSE	0.0308	0.0028	0.0338	0.0032	0.0694	0.0070	0.0757	0.0077
Bias	0.1061	-0.0007	0.1321	-0.0397	0.2005	-0.0634	0.2096	-0.0662
Variance	0.0196	0.0028	0.0163	0.0016	0.0292	0.0030	0.0318	0.0033

3.1.2. Case 2: When the True Joint Distribution is the AMH Copula Distribution with Normal Marginal Distributions

In this experiment, the true data generating process (DGP) is the AMH copula where their marginal distributions are normal distributions. Specifically, suppose the true joint distribution of the two error terms (V_1, V_2) is

$$F_{V_1V_2}(v_1,v_2)=\frac{\Phi(v_1)\Phi(v_2/\sigma_0)}{1-\alpha_0(1-\Phi(v_1))(1-\Phi(v_2/\sigma_0))} \tag{16}$$

where $\Phi(\cdot)$ is the standard normal distribution function. Set $\alpha_0=0.7$, and $\sigma_0=1$ for the experiment. Note that the true marginal distribution of $V_1 \sim \mathcal{N}(0,1)$, and that of $V_2 \sim \mathcal{N}(0,1)$. The random drawing $(\tilde{U}_1, \tilde{U}_2)$ from the AMH copula can be generated by the inverse conditional distribution method in Nelsen (2006), which is described in detail in the Appendix. Then, the random drawing of the two error terms $(\tilde{V}_1, \tilde{V}_2)$ can be generated from the joint distribution (16) by using the inversion method $\tilde{V}_1=\Phi^{-1}(\tilde{U}_1)$ and $\tilde{V}_2=\Phi^{-1}(\tilde{U}_2)$. Independent 1000 random drawings are generated for each replication sample, and 1000 replications are conducted for the experiment.[18]

In this experiment case, the estimation model is correctly specified only when the AMH CML estimation method is chosen.[19] Hence, the AMH CML estimator would have the smallest MSE, which is confirmed in Table 2. The MSE of the

[18] For the readers who may be interested, the results $MSE(\hat{\delta})$ are provided in the Appendix when the sample size is 500.

[19] If α is zero, then the joint distribution (16) becomes the bivariate normal joint distribution. But, recall that α is set to be nonzero in this experiment.

AMH CML estimator is shown to be the smallest among the three for every τ. As for the comparison between BN ML and Heckit, the MSE of Heckit is slightly less than that of BN ML for $\tau = 0.5, 0.9, 1$. So, in this case the BN ML estimator suffers from the misspecification more than the Heckit does. But, the AMH CML shows smaller MSE than the Heckit for every τ, which is expected since only AMH CML estimator is consistent estimators under this case. The AMH CML estimator still shows good MSE performance even when $n = 500$ except for $\tau = 0$. See Table 11 in the Appendix.

[Table 2] Case 2: Estimation of δ when the true joint distribution is the AMH copula with normal margins

	$\tau=0$		$\tau=0.5$		$\tau=0.9$		$\tau=1$	
parameter	δ_0	δ_1	δ_0	δ_1	δ_0	δ_1	δ_0	δ_1
				BN ML				
MSE	0.0335	0.0030	0.0304	0.0026	0.1116	0.0108	0.1158	0.0112
Bias	0.0609	0.0005	0.0170	-0.0017	0.0709	-0.0206	0.0632	-0.0185
Variance	0.0298	0.0030	0.0301	0.0026	0.1065	0.0103	0.1118	0.0108
				Heckit				
MSE	0.0334	0.0030	0.0294	0.0025	0.0997	0.0096	0.1050	0.0101
Bias	0.0650	0.0005	0.0325	-0.0054	0.0926	-0.0272	0.0868	-0.0256
Variance	0.0292	0.0030	0.0283	0.0025	0.0912	0.0089	0.0974	0.0094
				AMH				
MSE	0.0315	0.0029	0.0263	0.0022	0.0606	0.0059	0.0494	0.0049
Bias	0.0172	0.0003	0.0222	-0.0063	0.0919	-0.0284	0.0623	-0.0198
Variance	0.0312	0.0029	0.0258	0.0022	0.0522	0.0051	0.0456	0.0046

3.2. When no Model is Correctly Specified

Recall that the assumptions regarding (X_1, X_2) and the parameters are the same as those addressed in the beginning of the section 3. Only the data generating process on the two errors (V_1, V_2) are different in each case. In this subsection, we consider two cases where none of the three models is correct.

3.2.1. Case 3: When the True Joint Distribution is Farlie-Gumbel-Morgenstern (FGM) Copula with Normal Marginal Distributions

In this experiment, the true data generating process (DGP) is the FGM copula where their marginal distributions are standard normal distributions. Specifically, the true joint distribution for the two error terms (V_1, V_2) is

$$F_{V_1V_2}(v_1, v_2) = \Phi(v_1)\Phi(v_2)(1 + \alpha(1 - \Phi(v_1))(1 - \Phi(v_2))). \tag{17}$$

For this experiment, set $\alpha = 0.7$. Note that $C(u_1, u_2; \alpha) = u_1 u_2 (1 + \alpha(1 - u_1)(1 - u_2))$ is a Farlie-Gumbel-Morgenstern(FGM) copula satisfying $-1 \le \alpha \le 1$.

The random drawings ($\tilde{U}_1, \tilde{U}_2$) from the FGM copula can be obtained via the inverse conditional distributional method.[20] Given the random drawing ($\tilde{U}_1, \tilde{U}_2$), let $\tilde{V}_1 = \Phi(\tilde{U}_1)^{-1}$ and $\tilde{V}_2 = \Phi(\tilde{U}_2)^{-1}$. Hence, the ($\tilde{V}_1, \tilde{V}_2$) is a random drawing from the joint distribution (17).

By using independently randomly generated ($\tilde{V}_{1i}, \tilde{V}_{2i}$) from the joint distribution (17) and the set-up in (i)-(iv), we can obtain the replication sample for the experiment. For each replication, obtain an independent pair ($\tilde{V}_{1i}, \tilde{V}_{2i}$) given (X_{1i}, X_{2i}), $i = 1, \ldots, 1000$. Thus (Y_{1i}, Y_{2i}) are obtained conditional on (X_{1i}, X_{2i}), $i = 1, \ldots, 1000$. In each replication, the three estimation methods are implemented. 1000 replications are independently conducted.[21]

Note that any estimation model cannot be said to be correctly specified when the estimation method is chosen out of the three. Table 3 shows the results of MSE from this experiment. The three estimation methods show similar results when $\tau = 0$. When $\tau = 0.5$, the AMH CML method has slightly smaller MSE than the other two methods. When $\tau = 0.9$ or $\tau = 1$, AMH CML shows the smallest MSE among the three methods. Hence, AMH CML can be a good alternative method for the selection model when the chosen model is misspecified and some correlation between covariates of the two equations is strong.

[Table 3] Case 3: Estimation of δ when the true joint distribution is FGM copula with normal margins

	$\tau = 0$		$\tau = 0.5$		$\tau = 0.9$		$\tau = 1$	
parameter	δ_0	δ_1	δ_0	δ_1	δ_0	δ_1	δ_0	δ_1
				BN ML				
MSE	0.0301	0.0031	0.0302	0.0026	0.0849	0.0083	0.0884	0.0087
Bias	0.0155	0.0005	0.0013	-0.0015	0.0335	-0.0114	0.0292	-0.0101
Variance	0.0299	0.0031	0.0302	0.0026	0.0838	0.0082	0.0876	0.0086
				Heckit				
MSE	0.0301	0.0032	0.0299	0.0026	0.0971	0.0093	0.1003	0.0097
Bias	0.0186	0.0005	0.0036	-0.0020	0.0029	-0.0023	0.0013	-0.0018
Variance	0.0297	0.0031	0.0299	0.0026	0.0971	0.0093	0.1003	0.0097
				AMH				
MSE	0.0304	0.0031	0.0242	0.0022	0.0397	0.0041	0.0412	0.0043
Bias	-0.0116	0.0003	0.0180	-0.0079	0.0559	-0.0191	0.0578	-0.0197
Variance	0.0303	0.0031	0.0239	0.0021	0.0366	0.0037	0.0378	0.0039

[20] The procedure for random drawings from FGM copula is explained in the Appendix.

[21] For the readers who may be interested, the results $MSE(\hat{\delta})$ are provided in the Appendix when the sample size is 500.

3.2.2. Case 4: When the True Joint Distribution is the FGM with Logistic Marginal Distributions

Suppose the true joint distribution for the two error terms (V_1, V_2) is

$$F_{V_1V_2}(v_1, v_2) = \Lambda(v_1)\Lambda(v_2)(1+\alpha(1-\Lambda(v_1))(1-\Lambda(v_2))) \tag{18}$$

where $\Lambda(v) = 1/(1+\exp(-v))$ is the standard logistic distribution function. The DGP of (U_1, U_2) from the FGM copula is the same as in the previous subsection. But, the DGP of (V_1, V_2) from the joint distribution (18) is different from the joint distribution (17) since the marginal distributions are different.

The random drawing ($\tilde{V}_1, \tilde{V}_2$) from (18) can be done as follows. Specifically, $\tilde{V}_1 = \Lambda(\tilde{U}_1)^{-1}$ and $\tilde{V}_2 = \Lambda(\tilde{U}_2)^{-1}$ where ($\tilde{U}_1, \tilde{U}_2$) is a random drawing from the FGM copula. 1000 replications are conducted for the experiment, and 1000 observations are generated for each replication.[22]

Note that no estimation model among the three can be said to be correctly specified in this experiment. In particular, both the joint and marginal distributions of the two errors are misspecified when choosing one out of the three methods. In spite of both misspecifications, Table 4 shows that the AMH CML has the smallest MSE among the three estimators for every value of τ. Hence, the AMH CML estimation method seems to be a good alternative for the selection model when a researcher is not sure about the true joint distributions of the errors.[23]

It is also noticeable that MSE's of Heckit are very large compared to those of BN ML and AMH CML when the correlation among the regressors is high, i.e., $\tau = 0.9$ and 1. When the sample size is 500, the MSE of Heckit becomes higher than those from the experiment with $n = 1000$.

---

[22] For readers who may be interested, the results $MSE(\hat{\delta})$ are provided in the Appendix when the sample size is 500.

[23] Of course, the MSE from AMH CML of this experiment is much higher than that of the experiment in 3.1.2 since the latter chooses the correct specification. But, the MSE results in this experiment is relatively good compared with the two well-known methods.

[Table 4] Case 4: Estimation of δ when the true joint distribution is the FGM with logistic marginal distributions

	$\tau=0$		$\tau=0.5$		$\tau=0.9$		$\tau=1$	
parameter	δ_0	δ_1	δ_0	δ_1	δ_0	δ_1	δ_0	δ_1
				BN ML				
MSE	0.1302	0.0081	0.1828	0.0107	0.5671	0.0406	0.6275	0.0449
Bias	0.0220	-0.0011	-0.0470	0.0061	-0.0021	-0.0041	0.0862	-0.0267
Variance	0.1298	0.0081	0.1806	0.0107	0.5671	0.0406	0.6200	0.0442
				Heckit				
MSE	0.1211	0.0081	0.1491	0.0093	1.4483	0.0956	1.9647	0.1284
Bias	0.0395	-0.0009	-0.0016	-0.0025	0.0025	-0.0054	-0.0362	0.0037
Variance	0.1195	0.0081	0.1491	0.0093	1.4483	0.0956	1.9634	0.1284
				AMH				
MSE	0.1146	0.0078	0.0968	0.0058	0.1116	0.0081	0.1076	0.0079
Bias	-0.1824	-0.0008	-0.1952	0.0211	-0.2174	0.0415	-0.2204	0.0425
Variance	0.0813	0.0078	0.0586	0.0054	0.0644	0.0064	0.0590	0.0061

IV. Concluding Remarks

In this paper, the AMH CML estimation method is proposed for the sample selection model, and numerically compared with most frequently used two well-known methods for the selection model.[24] Some Monte Carlo experiments were conducted to compare the performance of the proposed AMH CML and the two well-known methods: BN ML and Heckit.

The experiment results show that the performance of the BN ML and Heckit estimators yield smaller MSE than the AMH CML estimator for $\tau=0$ or 0.5 when the true joint distribution is the bivariate normal distribution. However, the AMH CML estimator's MSE is less than that of Heckit for $\tau=0.9$ or $\tau=1$, even though the true joint distribution is the bivariate normal distribution. The results suggest that AMH CML can be a better alternative than Heckit even though Heckit is correct specification.

We find that when the true joint distribution is the AMH copula with normal marginal distributions, the AMH CML has the smallest MSE among the three for any value of τ. These results are expected since the AMH CML is the correctly specified model.

Interesting results are found when the imposed distributional assumption is far from the correct joint distribution and some of the regressors of the two equations are correlated. The AMH CML estimator has the smallest MSE among the three

[24] The difficulty of analytical comparison of their asymptotic variances and MSE's are discussed in the Appendix.

methods for any value of τ when the true joint distribution is the FGM with standard logistic marginal distributions (case 4). It also has the smallest MSE except for $\tau = 0$ when the true joint distribution is the FGM copula with normal marginal distributions (case 3). These results suggest that the AMH CML method can be a good alternative to estimate the sample selection model in terms of MSE when the specification is not correct. The proposed AMH CML method can be very useful if considering the fact that many researchers usually do not exactly know about the true DGP, and the correlation among the regressors of two equations is frequently present in practice.

Admittedly, the ideal copula ML approach for the sample selection model is to approximate the joint distribution by choosing the correct copula with correct marginal distributions. In that respect, the approach of this paper has some limitations since the proposed method might have different results depending on the choice of the copula. But, as seen in Nelsen (2006), there are too many copulas to pick in practice. Hence, it is not plausible to choose the correct copula in advance to begin the analysis of the selection model. Furthermore, there is still another issue with choosing the correct marginal distributions for the two error terms. The best solution to these problems is to verify whether the chosen copula and marginal distributions are correct after the estimation is implemented. One solution is to exploit the idea of the integrated conditional moments as in Bierens and Song (2012), and Bierens and Song (2014). This can be accomplished by comparing two conditional joint distributions of $(Y_1, Y_2)|(X_1, X_2)$ and $(\tilde{Y}_1, \tilde{Y}_2)|(X_1, X_2)$, where $(Y_1, Y_2)|(X_1, X_2)$ are actual observations from the true conditional joint distribution and $(\tilde{Y}_1, \tilde{Y}_2)|(X_1, X_2)$ are simulated observations from the chosen copula and marginal distributions. If they are the same, then the chosen copula estimator would be the true joint distribution. This topic can be another research avenue for the copula approach for the sample selection model.

In addition, two extensions can be considered. One involves a semi-parametric approach for the environment where the distribution of one error term is known, but that of the other error term is unknown. The other approach involves taking a semi-nonparametric approach for the environment where no distributional assumption is imposed on any error term. For the semi-nonparametric approach, we need some identification conditions.[25] These two extensions will be the directions for future research.

[25] In this case, we need to consider nonparametric identification of the marginal distribution of the error term in the participation equation. For example, see Manski (1998), and Bierens (2014).

Appendix

Proof of Lemma 1

$$
\begin{aligned}
&\int_{-x_1'\beta_1}^{\infty} f_{V_1V_2}(v_1, y_2 - x_2'\delta)dv_1 \\
&= \int_{-\infty}^{\infty} f_{V_1V_2}(v_1, y_2 - x_2'\delta)dv_1 - \int_{-\infty}^{-x_1'\beta} f_{V_1V_2}(v_1, y_2 - x_2'\delta)dv_1 \\
&= f_{V_2}(y_2 - x_2'\delta) - \int_{-\infty}^{-x_1'\beta} f_{V_1|V_2}(v_1 \mid y_2 - x_2'\delta) f_{V_2}(y_2 - x_2'\delta)dv_1 \\
&= f_{V_2}(y_2 - x_2'\delta)\left[1 - \int_{-\infty}^{-x_1'\beta} f_{V_1|V_2}(v_1 \mid y_2 - x_2'\delta)dv_1\right] \\
&= f_{V_2}(y_2 - x_2'\delta)[1 - F_{V_1|V_2}(-x_1'\beta \mid \mathrm{y}_2 - x_2'\delta)] \\
&= f_{V_2}(y_2 - x_2'\delta)[1 - \Pr(V_1 \le -x_1'\beta \mid V_2 = \mathrm{y}_2 - x_2'\delta)] \\
&= f_{V_2}(y_2 - x_2'\delta)[1 - \Pr(F_{V_1}(V_1) \le F_{V_1}(-x_1'\beta) \mid F_{V_2}(V_2) = F_{V_2}(\mathrm{y}_2 - x_2'\delta))] \\
&= f_{V_2}(y_2 - x_2'\delta)[1 - \Pr(U_1 \le F_{V_1}(-x_1'\beta) \mid U_2 = F_{V_2}(\mathrm{y}_2 - x_2'\delta))] \\
&= f_{V_2}(y_2 - x_2'\delta)[1 - \Pr(U_1 \le u_1 \mid U_2 = u_2)] \\
&= f_{V_2}(y_2 - x_2'\delta)\left[1 - \frac{\partial C(u_1, u_2; \alpha)}{\partial u_2}\right] \\
&= f_{V_2}(y_2 - x_2'\delta)[1 - C_{u_2}(u_1, u_2; \alpha)]
\end{aligned}
$$

where $U_1 = F_{V_1}(V_1)$, $U_2 = F_{V_2}(V_2)$, $u_1 = F_{V_2}(-x_1'\beta)$, $u_2 = F_{V_2}(y_2 - x_2'\delta)$, and $C_{u_2}(u_1, u_2; \alpha) \equiv \Pr(U_1 \le u_1 \mid U_2 = u_2)$. The seventh equality follows from the fact that each F_{V_j}, $j = 1,2$, follows a uniform distribution on $[0,1]$, and the ninth equality follows from the fact given below.

$$
\begin{aligned}
&\Pr(U_1 \le u_1 \mid U_2 = u_2) \\
&= \lim_{\Delta u_2 \to 0} \frac{\Pr(U_1 \le u_1, U_2 \le u_2) - \Pr(U_1 \le u_1, U_2 \le u_2 - \Delta u_2)}{\Pr(U_2 \le u_2) - \Pr(U_2 \le u_2 - \Delta u_2)} \\
&= \lim_{\Delta u_2 \to 0} \frac{C(u_1, u_2; \alpha) - C(u_1, u_2 - \Delta u_2; \alpha)}{u_2 - (u_2 - \Delta u_2)} \\
&= \lim_{\Delta u_2 \to 0} \frac{C(u_1, u_2; \alpha) - C(u_1, u_2 - \Delta u_2; \alpha)}{\Delta u_2} \\
&= \frac{\partial C(u_1, u_2; \alpha)}{\partial u_2} \equiv C_{u_2}(u_1, u_2; \alpha).
\end{aligned}
$$

■

$\partial \ell_i(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ when using the Ali-Mikhail-Haq copula

$$\ell_i(\boldsymbol{\theta}) = (1-y_{1i}) \log \Phi(-x_{1i}'\boldsymbol{\beta}) - \frac{y_{1i}}{2}\left(\log(2\pi) + 2\log\sigma_2 + \left(\frac{y_{2i}-x_{2i}'\boldsymbol{\delta}}{\sigma_2}\right)^2 \right)$$
$$+ y_{1i} \log(1 - C_{u_2}(u_{1i}, u_{2i}, \alpha)) \tag{19}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\delta}', \sigma_2, \alpha)'$, $u_{1i} = \Phi(-x_{1i}'\boldsymbol{\beta})$, $u_{2i} = \Phi(\frac{y_{2i}-x_{2i}'\boldsymbol{\delta}}{\sigma_2})$, and $C_{u_2}(u_{1i}, u_{2i}; \alpha) = \frac{u_{1i}[1-\alpha(1-u_{1i})]}{[1-\alpha(1-u_{1i})(1-u_{2i})]^2}$. Then, the first derivatives are as follows.

$$\frac{\partial \ell_i}{\partial \boldsymbol{\beta}} = -x_{1i}(1-y_{1i})\lambda(-x_{1i}'\boldsymbol{\beta}) + \frac{x_{1i}y_{1i}\phi(-x_{1i}'\boldsymbol{\beta})}{1-\frac{u_{1i}[1-\alpha(1-u_{1i})]}{[1-\alpha(1-u_{1i})(1-u_{2i})]^2}} \times \left(\frac{1-\alpha(1-2u_{1i})}{[1-\alpha(1-u_{1i})(1-u_{2i})]^2} \right.$$
$$\left. - \frac{2u_{1i}[1-\alpha(1-u_{1i})]\alpha(1-u_{2i})}{[1-\alpha(1-u_{1i})(1-u_{2i})]^3} \right) \tag{20}$$

where $\lambda(\cdot) = \phi(\cdot) / \Phi(\cdot)$.

$$\frac{\partial \ell_i}{\partial \boldsymbol{\delta}} = \frac{x_{2i}y_{1i}}{\sigma_2}\left[\frac{(y_{2i}-x_{2i}'\boldsymbol{\delta})}{\sigma_2} - \frac{\phi(\frac{y_{2i}-x_{2i}'\boldsymbol{\delta}}{\sigma_2})}{1-\frac{u_{1i}[1-\alpha(1-u_{1i})]}{[1-\alpha(1-u_{1i})(1-u_{2i})]^2}} \left(\frac{2\alpha u_{1i}(1-u_{1i})[1-\alpha(1-u_{1i})]}{[1-\alpha(1-u_{1i})(1-u_{2i})]^3} \right) \right] \tag{21}$$

since $\phi'(x) = -x\phi(x)$.

$$\frac{\partial \ell_i}{\partial \sigma_2} = -\frac{y_{1i}}{\sigma_2}\left(1 - \left(\frac{y_{2i}-x_{2i}'\boldsymbol{\delta}}{\sigma_2}\right)^2 \right) - 2y_{1i}\frac{(y_{2i}-x_{2i}'\boldsymbol{\delta})}{\sigma_2^2}\phi\left(\frac{y_{2i}-x_{2i}'\boldsymbol{\delta}}{\sigma_2}\right)$$
$$\times \frac{1}{1-\frac{u_{1i}[1-\alpha(1-u_{1i})]}{[1-\alpha(1-u_{1i})(1-u_{2i})]^2}} \times \frac{\alpha u_{1i}(1-u_{1i})[1-\alpha(1-u_{1i})]}{[1-\alpha(1-u_{1i})(1-u_{2i})]^3}. \tag{22}$$

$$\frac{\partial \ell_i}{\partial \alpha} = \frac{y_{1i}}{1-\frac{u_{1i}[1-\alpha(1-u_{1i})]}{[1-\alpha(1-u_{1i})(1-u_{2i})]^2}} \times \left(\frac{u_{1i}(1-u_{1i})}{[1-\alpha(1-u_{1i})(1-u_{2i})]^2} \right.$$
$$\left. - \frac{2u_{1i}[1-\alpha(1-u_{1i})](1-u_{1i})(1-u_{2i})}{[1-\alpha(1-u_{1i})(1-u_{2i})]^3} \right). \tag{23}$$

All derivations above are based on the following facts:

- $$\frac{\partial C(u_1,u_2;\alpha)}{\partial u_2}=\frac{u_1[1-\alpha(1-u_1)(1-u_2)]-u_1u_2\alpha(1-u_1)}{[1-\alpha(1-u_1)(1-u_2)]^2}=\frac{u_1[1-\alpha(1-u_1)]}{[1-\alpha(1-u_1)(1-u_2)]^2}$$
- $$\frac{\partial C(u_1,u_2;\alpha)}{\partial u_1}=\frac{u_2[1-\alpha(1-u_2)]}{[1-\alpha(1-u_1)(1-u_2)]^2}$$
- $$\frac{\partial^2 C(u_1,u_2;\alpha)}{\partial u_1\partial u_2}=\frac{\partial}{\partial u_1}\left(\frac{\partial C(u_1,u_2;\alpha)}{\partial u_2}\right)=\frac{1-\alpha(1-2u_1)}{[1-\alpha(1-u_1)(1-u_2)]^2}-\frac{2u_1[1-\alpha(1-u_1)]\alpha(1-u_2)}{[1-\alpha(1-u_1)(1-u_2)]^3}$$
- $$\frac{\partial^2 C(u_1,u_2;\alpha)}{\partial u_2^2}=\frac{\partial}{\partial u_2}\left(\frac{\partial C(u_1,u_2;\alpha)}{\partial u_2}\right)=-\frac{2\alpha u_1(1-u_1)[1-\alpha(1-u_1)]}{[1-\alpha(1-u_1)(1-u_2)]^3}$$
- $$\frac{\partial^2 C(u_1,u_2;\alpha)}{\partial u_2\partial \sigma_2}=\frac{\partial}{\partial u_2}\left(\frac{\partial C(u_1,u_2;\alpha)}{\partial u_2}\right)=\frac{\partial}{\partial \sigma_2}C_{u_2}(u_1,u_2;\alpha)=\frac{\partial C_{u_2}(u_1,u_2;\alpha)}{\partial u_2}\frac{\partial u_2}{\partial \sigma_2}=\frac{\partial^2 C(u_1,u_2;\alpha)}{\partial u_2^2}\frac{\partial u_2}{\partial \sigma_2}$$
- $$\frac{\partial u_2}{\partial \sigma_2}=\phi((y_2-x_2'\delta)/\sigma_2)(-1)(y_2-x_2'\delta)/\sigma_2^2=-\frac{(y_2-x_2'\delta)\phi((y_2-x_2'\delta)/\sigma_2)}{\sigma_2^2}$$
- $$\frac{\partial^2 C(u_1,u_2;\alpha)}{\partial u_2\partial \alpha}=\frac{\partial}{\partial \alpha}\left(\frac{\partial C(u_1,u_2;\alpha)}{\partial u_2}\right)=\frac{-u_1(1-u_1)}{[1-\alpha(1-u_1)(1-u_2)]^2}+\frac{2u_1[1-\alpha(1-u_1)](1-u_1)(1-u_2)}{[1-\alpha(1-u_1)(1-u_2)]^3}$$

Random drawing ($\tilde U_1,\tilde U_2$) from AMH copula

Note $\frac{\partial C(u_1,u_2)}{\partial u_2}=\Pr(U_1\le u_1\,|\,U_2=u_2)$ is the conditional distribution which is shown in the proof of Lemma 1. Hence, $\Pr(U_1\,|\,U_2=u_2)$ follows $\mathcal{U}[0,1]$. Moreover, it follows from AMH copula that

$$\frac{\partial C(u_1,u_2)}{\partial u_2}=\frac{u_1[1-\alpha(1-u_1)]}{[1-\alpha(1-u_1)(1-u_2)]^2}.$$

To randomly draw ($\tilde U_1,\tilde U_2;\alpha$) from a joint distribution $C(u_1,u_2,\alpha)=\frac{u_1u_2}{1-\alpha(1-u_1)(1-u_2)}$, we can use the following procedure.

1. Independently randomly draw ($\overline{u}_{10},u_{20}$) from $\mathcal{U}[0,1]$.
2. Set $\tilde U_2=u_{20}$.
3. Now find $\tilde U_1$ satisfying $\frac{\partial C(\tilde U_1,u_{20};\alpha)}{\partial u_2}=\overline{u}_{10}$, that is, $\tilde U_1=C_{\tilde u_2}^{-1}(\overline{u}_{10})$. Specifically,

numerically find $\tilde{U}_1 \in [0,1]$ which satisfies $\frac{\tilde{U}_1[1-\alpha(1-\tilde{U}_1)]}{[1-\alpha(1-\tilde{U}_1)(1-\tilde{U}_2)]^2} - \overline{u}_{10} = 0$ given $\alpha, \overline{u}_{10}, \tilde{U}_2$.

The numerical procedure such as "fsolve" in Matlab can be used to find $\tilde{U}_1$.

4. Hence, the random drawing $(\tilde{U}_1, \tilde{U}_2)$ from AMH copula is obtained.

Note the order of the conditioning to generate random drawings from the joint distribution $C(u_1, u_2; \alpha)$ does not matter.

Random drawing $(\tilde{U}_1, \tilde{U}_2)$ from FGM copula

Note FGM copula is $C(u_1, u_2; \alpha) = u_1 u_2 (1 + \alpha(1-u_1)(1-u_2))$.

1. Independently randomly generate $(\overline{u}_{10}, u_{20})$ from $\mathcal{U}[0,1]$.
2. Set $\tilde{U}_2 = u_{20}$.
3. Obtain $\tilde{U}_1 = C^{-1}_{\tilde{u}_2}(\overline{u}_{10})$ satisfying $\overline{u}_{10} = C_{\tilde{u}_2}(\tilde{U}_1) = \frac{\partial C(\tilde{U}_1, \tilde{U}_2)}{\partial u_2}$. That is, find $\tilde{u}_1 \in [0,1]$ which satisfies $\tilde{U}_1(1 + \alpha(1-\tilde{U}_1)(1-2\tilde{U}_2)) - \overline{u}_{10} = 0$ given $\alpha, \overline{u}_{10}, \tilde{U}_2$. Hence,

$$\alpha(1-2\tilde{U}_2)\tilde{U}_1^2 - (1+\alpha(1-2\tilde{U}_2))\tilde{U}_1 + \overline{u}_{10} = 0 . \tag{24}$$

$$\tilde{U}_1 = \frac{(1+\alpha(1-2\tilde{U}_2)) - \sqrt{(1+\alpha(1-2\tilde{U}_2))^2 - 4\alpha(1-2\tilde{U}_2)\overline{u}_{10}}}{2\alpha(1-2\tilde{U}_2)} .$$

Note that we can have explicit random drawings for FGM copula if you have any independent random drawings $(\overline{u}_{10}, u_{20})$ from $\mathcal{U}[0,1]$.

4. Hence, the random drawing $(\tilde{U}_1, \tilde{U}_2)$ from FGM copula is

$$\tilde{U}_2 = u_{20} \text{ , and}$$

$$\tilde{U}_1 = \frac{(1+\alpha(1-2u_{20})) - \sqrt{(1+\alpha(1-2u_{20}))^2 - 4\alpha(1-2u_{20})\overline{u}_{10}}}{2\alpha(1-2u_{20})} .$$

Note that we do not need to solve (24) numerically in FGM copula unlike AMH copula.

Comparison of the asymptotic variances of BN ML and AMH CML estimators

For example, consider the comparison of BN ML estimator and AMH CML estimator. Recall that the consistency requires that the distributional assumption chosen by a researcher should be true. Otherwise, the estimator cannot be consistent. Considering the possible inconsistency from the incorrect specification, the asymptotic normality of $\hat{\boldsymbol{\theta}}_{BN}$ can be written as follow.

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{BN}-\boldsymbol{\theta}^*_{BN})=\begin{bmatrix}\sqrt{n}(\hat{\boldsymbol{\beta}}_{BN}-\boldsymbol{\beta}^*_{BN})\\ \sqrt{n}(\hat{\boldsymbol{\delta}}_{BN}-\boldsymbol{\delta}^*_{BN})\\ \sqrt{n}(\hat{\rho}_{BN}-\rho^*_{BN})\\ \sqrt{n}(\hat{\sigma}_{BN}-\sigma^*_{2,BN})\end{bmatrix}\xrightarrow{d} N(0,V^*_{BN}) \tag{25}$$

where $\boldsymbol{\theta}^*_{BN}$ is a pseudo-true parameter which is the sum of the true parameter $\boldsymbol{\theta}_{0,BN}$ and a possible bias. $V^*_{BN}=A^{*\ -1}_{BN}B^{*\ -1}_{BN}A^{*\ -1}_{BN}$ where $A^*_{BN}=E[\frac{\partial^2\ell_1^{(BN)}(\boldsymbol{\theta}^*_{BN})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}]$, $B^*_{BN}=E[\frac{\partial\ell_1^{(BN)}(\boldsymbol{\theta}^*_{BN})}{\partial\boldsymbol{\theta}}\frac{\partial\ell_1^{(BN)}(\boldsymbol{\theta}^*_{BN})}{\partial\boldsymbol{\theta}'}]$, and

$$\begin{aligned}\ell_i^{(BN)}(\boldsymbol{\theta})=[&(1-y_{1i})\log\Phi(-x'_{1i}\boldsymbol{\beta})\\&+y_{1i}\left(\log\Phi\left(\frac{x'_{1i}\boldsymbol{\beta}+\rho\frac{1}{\sigma_2}(y_{2i}-x'_{2i}\boldsymbol{\delta})}{\sqrt{1-\rho^2}}\right)+\log\phi\left(\frac{y_{2i}-x'_{2i}\boldsymbol{\delta}}{\sigma_2}\right)-\log\sigma_2\right)\Bigg]\end{aligned}$$

where $\boldsymbol{\theta}=(\boldsymbol{\beta}',\boldsymbol{\delta}',\rho,\sigma_2)'$.

Similarly, the same arguments apply to the AMH CML estimator.

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{CML}-\boldsymbol{\theta}^*_{CML})=\begin{bmatrix}\sqrt{n}(\hat{\boldsymbol{\beta}}_{CML}-\boldsymbol{\beta}^*_{CML})\\ \sqrt{n}(\hat{\boldsymbol{\delta}}_{CML}-\boldsymbol{\delta}^*_{CML})\\ \sqrt{n}(\hat{\alpha}_{CML}-\alpha^*_{CML})\\ \sqrt{n}(\hat{\sigma}_{CML}-\sigma^*_{CML})\end{bmatrix}\xrightarrow{d}\mathcal{N}(0,V^*_{CML})$$

where $\boldsymbol{\theta}^*_{CML}$ is a pseudo-true parameter which is the sum of the true parameter $\boldsymbol{\theta}_{0,CML}$ and a possible bias. $V^*_{CML}=A^{*\ -1}_{CML}B^*_{CML}A^{*\ -1}_{CML}$ where $A^*_{CML}=E[\frac{\partial^2\ell_1^{(CML)}(\boldsymbol{\theta}^*_{CML})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}]$, $B^*_{CML}=E[\frac{\partial\ell_1^{(CML)}(\boldsymbol{\theta}^*_{CML})}{\partial\boldsymbol{\theta}}\frac{\partial\ell_1^{(CML)}(\boldsymbol{\theta}^*_{CML})}{\partial\boldsymbol{\theta}'}]$, and

$$\begin{aligned}\ell_1^{(CML)}(\boldsymbol{\theta})=&(1-y_{1i})\log\Phi(-x'_{1i}\boldsymbol{\beta})+y_{1i}(\log[\phi((y_{2i}-x'_{2i}\boldsymbol{\delta})/\sigma)]-\log\sigma)\\&+y_{1i}\log\left(1-\frac{u_{1i}[1-\alpha(1-u_{1i})]}{[1-\alpha(1-u_{1i})(1-u_{2i})]^2}\right)\end{aligned}$$

where $u_{1i} = \Phi(-x'_{1i}\boldsymbol{\beta})$, $u_{2i} = \Phi((y_{2i} - x'_{2i}\boldsymbol{\delta})/\sigma)$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\delta}', \sigma, \alpha)'$.

In general, we cannot determine whether the block of $(\boldsymbol{\beta}, \boldsymbol{\delta})$ in V^*_{CML} minus the same block in V^*_{BN} is positive semi-definite or negative semi-definite. For the comparison of the MSE, we also need to consider the bias $\boldsymbol{\theta}^* - \boldsymbol{\theta}_0$, which makes the comparison be indeterminate in general.

By using the inverse of the partitioned matrices and $\boldsymbol{\theta}^*_{BN} = \boldsymbol{\theta}_{0,BN}$, we can obtain the asymptotic variance of $\begin{bmatrix} \sqrt{n}(\hat{\boldsymbol{\beta}}_{BN} - \boldsymbol{\beta}_0) \\ \sqrt{n}(\hat{\boldsymbol{\delta}}_{BN} - \boldsymbol{\delta}_0) \end{bmatrix}$.

$$\begin{bmatrix} \sqrt{n}(\hat{\boldsymbol{\beta}}_{BN} - \boldsymbol{\beta}_0) \\ \sqrt{n}(\hat{\boldsymbol{\delta}}_{BN} - \boldsymbol{\delta}_0) \end{bmatrix} \xrightarrow{d} \mathcal{N}(0, V_{(\boldsymbol{\beta},\boldsymbol{\delta}),BN})$$

where

$$V_{(\boldsymbol{\beta},\boldsymbol{\delta}),BN} = [C_{1,BN} - C'_{2,BN} C^{-1}_{3,BN} C_{2,BN}]^{-1},$$

$$C_{1,BN} = \begin{bmatrix} \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \beta} \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \boldsymbol{\beta}'} & \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \boldsymbol{\beta}} \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \boldsymbol{\delta}'} \\ \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \boldsymbol{\delta}} \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \boldsymbol{\beta}'} & \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \boldsymbol{\delta}} \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \boldsymbol{\delta}'} \end{bmatrix},$$

$$C'_{2,BN} = \begin{bmatrix} \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \boldsymbol{\beta}} \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \rho} & \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \boldsymbol{\beta}} \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \sigma_2} \\ \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \boldsymbol{\delta}} \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \rho} & \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \boldsymbol{\delta}} \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \sigma_2} \end{bmatrix}, \text{ and}$$

$$C_{3,BN} = \begin{bmatrix} \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \rho} \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \rho} & \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \rho} \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \sigma} \\ \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \sigma_2} \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \rho} & \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \sigma_2} \frac{\partial \ell_1^{(BN)}(\boldsymbol{\theta}_{0,BN})}{\partial \sigma_2} \end{bmatrix}$$

where $\boldsymbol{\theta}_{0,BN} = (\boldsymbol{\beta}'_0, \boldsymbol{\delta}'_0, \rho_0, \sigma_{0,2})'$.

The similar results can be obtained for the AMH CML estimator.

$$\begin{bmatrix} \sqrt{n}(\hat{\boldsymbol{\beta}}_{CML} - \boldsymbol{\beta}_0) \\ \sqrt{n}(\hat{\boldsymbol{\delta}}_{CML} - \boldsymbol{\delta}_0) \end{bmatrix} \xrightarrow{d} \mathcal{N}(0, V_{(\boldsymbol{\beta},\boldsymbol{\delta}),CML})$$

where

$$V_{(\boldsymbol{\beta},\boldsymbol{\delta}),CML} = [C_{1,CML} - C'_{2,CML} C^{-1}_{3,CML} C_{2,CML}]^{-1},$$

$$C_{1,CML} = \begin{bmatrix} \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \boldsymbol{\beta}} \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \boldsymbol{\beta}'} & \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \boldsymbol{\beta}} \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \boldsymbol{\delta}'} \\ \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \boldsymbol{\delta}} \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \boldsymbol{\beta}'} & \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \boldsymbol{\delta}} \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \boldsymbol{\delta}'} \end{bmatrix},$$

$$C'_{2,CML} = \begin{bmatrix} \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \boldsymbol{\beta}} \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \rho} & \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \boldsymbol{\beta}} \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \sigma} \\ \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \boldsymbol{\delta}} \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \alpha} & \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \boldsymbol{\delta}} \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \sigma} \end{bmatrix},$$

$$C_{3,CML} = \begin{bmatrix} \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \alpha} \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \alpha} & \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \alpha} \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \sigma} \\ \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \sigma} \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \alpha} & \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \sigma} \frac{\partial \ell_1^{(CML)}(\boldsymbol{\theta}_{0,CML})}{\partial \sigma} \end{bmatrix}, \text{ and}$$

$$\ell_1^{(CML)}(\boldsymbol{\theta}) = (1-y_{1i})\log\Phi(-x'_{1i}\boldsymbol{\beta}) + y_{1i}(\log[\phi((y_{2i} - x'_{2i}\boldsymbol{\delta})/\sigma)] - \log\sigma)$$
$$+ y_{1i}\log\left(1 - \frac{u_{1i}[1-\alpha(1-u_{1i})]}{[1-\alpha(1-u_{1i})(1-u_{2i})]^2}\right).$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\delta}', \sigma, \alpha)'$ and $\boldsymbol{\theta}_{0,CML} = (\boldsymbol{\beta}'_0, \boldsymbol{\delta}'_0, \sigma_0, \alpha_0)'$.

Even in a hypothetical situations where both specifications are correct, we cannot determine in general whether $V_{(\boldsymbol{\beta},\boldsymbol{\delta}),CML} - V_{(\boldsymbol{\beta},\boldsymbol{\delta}),BN}$ is positive semi-definite or negative semi-definite.

Estimation results of $\boldsymbol{\beta}$ in the three models

[Table 5] (Case 1) Estimation of $\boldsymbol{\beta}$ when the true joint distribution is a bivariate normal distribution

	$\tau=0$		$\tau=0.5$		$\tau=0.9$		$\tau=1$	
parameter	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
				BN ML				
MSE	0.0038	0.0046	0.0075	0.0034	0.0099	0.0041	0.0102	0.0043
				Heckit				
MSE	0.0038	0.0046	0.0075	0.0034	0.0100	0.0041	0.0102	0.0042
				AMH				
MSE	0.0039	0.0005	0.0076	0.0035	0.0101	0.0042	0.0102	0.0043

[Table 6] (Case 2) Estimation of β when the true joint distribution is the AMH copula distribution

	$\tau=0$		$\tau=0.5$		$\tau=0.9$		$\tau=1$	
parameter	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
				BN ML				
MSE	0.0033	0.0051	0.0072	0.0036	0.0101	0.0041	0.0098	0.0041
				Heckit				
MSE	0.0033	0.0051	0.0072	0.0036	0.0102	0.0041	0.0097	0.0040
				AMH				
MSE	0.0033	0.0052	0.0071	0.0036	0.0103	0.0042	0.0097	0.0040

[Table 7] (Case 3) Estimation of β when the true joint distribution is FGM copula with normal marginal distributions

	$\tau=0$		$\tau=0.5$		$\tau=0.9$		$\tau=1$	
parameter	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
				BN ML				
MSE	0.0033	0.0049	0.0075	0.0037	0.0097	0.0041	0.0099	0.0041
				Heckit				
MSE	0.0033	0.0049	0.0075	0.0037	0.0097	0.0040	0.0098	0.0040
				AMH				
MSE	0.0033	0.0050	0.0075	0.0037	0.0097	0.0041	0.0098	0.0040

[Table 8] (Case 4) Estimation of β when the true joint distribution is the FGM copula with logistic marginal distributions

	$\tau=0$		$\tau=0.5$		$\tau=0.9$		$\tau=1$	
parameter	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
				BN ML				
MSE	0.1646	0.1716	0.1661	0.1601	0.1744	0.1691	0.1751	0.1695
				Heckit				
MSE	0.1644	0.1708	0.1646	0.1590	0.1740	0.1685	0.1750	0.1691
				AMH				
MSE	0.1640	0.1713	0.1668	0.1619	0.1763	0.1713	0.1772	0.1718

Estimation results of α in AMH CML

[Table 9] Estimation of α in AMH copula ML

	$\tau=0$	$\tau=0.5$	$\tau=0$	$\tau=0.5$
		Case 1		
MSE	0.0611	0.0446	0.0881	0.1036
Bias	0.2172	0.1626	0.0184	-0.0018
Variance	0.0140	0.0182	0.0878	0.1036
		Case 2		
MSE	0.1133	0.1535	0.3093	0.2268
Bias	-0.0794	-0.3386	-0.4360	-0.1694
Variance	0.1070	0.0389	0.1192	0.1981
		Case 3		
MSE	0.0999	0.0987	0.2206	0.2200
Bias	-0.0969	-0.1568	-0.2567	-0.2607
Variance	0.0905	0.0741	0.1547	0.1520
		Case 4		
MSE	0.0930	0.0541	0.0575	0.0491
Bias	0.0752	0.0799	0.0802	0.0863
Variance	0.0874	0.0477	0.0511	0.0417

Estimation results of δ when $n=500$

[Table 10] (Case 1: $n=500$) Estimation of δ when the true joint distribution is bivariate normal

	$\tau=0$		$\tau=0.5$		$\tau=0.9$		$\tau=1$	
parameter	δ_0	δ_1	δ_0	δ_1	δ_0	δ_1	δ_0	δ_1
				BN ML				
MSE($n=500$)	0.0573	0.0053	0.0521	0.0046	0.1013	0.0101	0.1019	0.0102
				Heckit				
MSE($n=500$)	0.0583	0.0053	0.0651	0.0053	0.1832	0.0173	0.1927	0.0183
				AMH				
MSE($n=500$)	0.0549	0.0055	0.0554	0.0050	0.0954	0.0095	0.0974	0.0098

[Table 11] (Case 2: $n = 500$) Estimation of δ when the true joint distribution is the AMH copula with normal margins

	$\tau = 0$		$\tau = 0.5$		$\tau = 0.9$		$\tau = 1$	
parameter	δ_0	δ_1	δ_0	δ_1	δ_0	δ_1	δ_0	δ_1
				BN ML				
MSE($n = 500$)	0.0571	0.0057	0.0652	0.0054	0.1935	0.0185	0.1855	0.0178
				Heckit				
MSE($n = 500$)	0.0559	0.0058	0.0628	0.0053	0.2054	0.0195	0.2111	0.0201
				AMH				
MSE($n = 500$)	0.0540	0.0058	0.0545	0.0046	0.0661	0.0069	0.0676	0.0071

[Table 12] (Case 3: $n = 500$) Estimation of δ when the true joint distribution is FGM copula with normal margins

	$\tau = 0$		$\tau = 0.5$		$\tau = 0.9$		$\tau = 1$	
parameter	δ_0	δ_1	δ_0	δ_1	δ_0	δ_1	δ_0	δ_1
				BN ML				
MSE($n = 500$)	0.0600	0.0061	0.0625	0.0053	0.1558	0.0151	0.1570	0.0153
				Heckit				
MSE($n = 500$)	0.0590	0.0061	0.0613	0.0052	0.2027	0.0192	0.2058	0.0196
				AMH				
MSE($n = 500$)	0.0526	0.0062	0.0475	0.0043	0.0607	0.0065	0.0629	0.0068

[Table 13] (Case 4: $n = 500$) Estimation of δ when the true joint distribution is the FGM copula with logistic marginal distributions

	$\tau = 0$		$\tau = 0.5$		$\tau = 0.9$		$\tau = 1$	
parameter	δ_0	δ_1	δ_0	δ_1	δ_0	δ_1	δ_0	δ_1
				BN ML				
MSE($n = 500$)	0.2776	0.0157	0.3678	0.0218	0.7829	0.0576	0.7712	0.0570
				Heckit				
MSE($n = 500$)	0.2430	0.0158	0.3081	0.0190	3.2285	0.2102	4.2885	0.2762
				AMH				
MSE($n = 500$)	0.1809	0.0152	0.1686	0.0123	0.1667	0.0145	0.1582	0.0136

References

Ali, M. M., N. N. Mikhail and M. S. Haq (1978), "A Class of Bivariate Distributions Including the Bivariate Logistic," *Journal Multivariate Analysis*, 8, 405-412.

Bierens, H. J. and H. Song (2012), "Semi-Nonparametric Estimation of Independently and Identically Repeated First-Price Auctions via an Integrated Simulated Moments Method," *Journal of Econometrics*, 168, 108-119.

______________________ (2014), "Semi-Nonparametric Modeling and Estimation of First-Price Auctions Models with Auction-Specific Heterogeneity," mimeo.

Bierens, H. J. (2014), "Consistency and Asymptotic Normality of Sieve ML Estimators Under Low-Level Conditions," *Econometric Theory*, 30, 1021-1076.

Chen, S. (1997), "Semiparametric Estimation of the Type-3 Tobit Model," *Journal of Econometrics*, 80, 1-34.

Chen, X., Y. Fan and V. Tsyrennikov (2006), "Efficient Estimation of Semiparametric Multivariate Copula Models," *Journal of the American Statistical Association*, 101, 1228-1240.

Cragg, J. G. (1971), "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods," *Econometrica*, 39, 829-844.

Das, M., W. K. Newey and F. Vella (2003), "Nonparametric Estimation of Sample Selection Models," *Review of Economic Studies*, 70, 33-58.

Fan, J. and J. Gu (2003), "Semiparametric Estimation of Value at Risk," *Econometrics Journal*, 6, 261-290.

Gallant, A. R., and D. W. Nychka (1987), "Semi-nonparametric Maximum Likelihood Estimation," *Econometrica*, 55, 363-390.

Genest, C. and J. Mackay (1986), "The Joys of Copulas: Bivariate Distributions with Uniform Marginals," *American Statistician*, 40, 280-283.

Heckman, J. J. (1974), "Shadow Prices, Market Wages, and Labor Supply," *Econometrica*, 42, 679-694.

____________ (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *The Annals of Economic and Social Measurement*, 5, 475-492.

____________ (1978), "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46, 931-959.

____________ (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153-161.

Honore, B. E., E. Kyriazidou and C. Udry (1997), "Estimation of Type-3 Tobit Models Using Symmetric Trimming and Pairwise Comparisons," *Journal of Econometrics*, 76, 107-128.

Hu, L. (2006), "Dependence Patterns Across Financial Markets: A Mixed Copula Approach," *Applied Financial Economics*, 16, 717-729.

Kumar, P. (2010), "Probability Distributions and Estimation of Ali-Mikhail-Haq Copula," *Applied Mathematical Sciences*, 4, 657-666.

Lee, L. (1983), "Generalized Econometric Models with Selectivity," *Econometrica*, 51, 507-512.

Manski, C. F. (1988), "Identification of Binary Response Models," *Journal of the American Statistical Association*, 83, 729-738.

Nawata, K. (1993), "A Note on the Estimation of Models with Sample-selection Biases," *Economics Letters*, 42, 15-24.

Nawata, K. and N. Nagase (1996), "Estimation of Sample Selection Bias Models," *Econometric Reviews*, 104, 387-400.

Nelder, J. A. and R. Mead (1965), "A Simplex Method for Function Minimization," *The Computer Journal*, 7, 30813.

Nelsen, R. B. (2006), "An Introduction to Copulas," *Springer*.

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.

Patton, A. J. (2006), "Modelling Asymmetric Exchange Rate Dependence," *International Economic Review*, 47, 527-556.

Sklar, A. (1959) "Fonctions de rpartition n dimensions et leurs marges," *Publ Inst Statist Univ Paris*, 8, 229-231.

Smith, M. D. (2003), "Modeling Sample Selection Using Archimedean Copulas," *Econometrics Journal*, 6, 99-123.

Vella, F. (1992), "Simple Test for Sample Selection Bias in Censored and Discrete Choice Models," *Journal of Applied Econometrics*, 7, 413-421.

________ (1993), "A Simple Estimator for Simultaneous Models with Censored Endogenous Regressors," *International Economic Review*, 34, 441-457.

Zimmer, D. M. and P. K. Trivedi (2006), "Using Trivariate Copulas to Model Sample Selection and Treatment Effects: Application of Family Health Care Demand," *Journal of Business & Economic Statistics*, 24, 63-76.