Iranian Journal of Fuzzy Systems

Volume 17, Number 6, (2020), pp. 29-38

University of Sistan and Baluchestan

Copula-based measurement error models

A. Sheikhi[1] and R. Mesiar[2]

[1] *Department of Statistics, Faculty of Mathematics and Computer, Shahid Bahonar University, Kerman, Iran*
[2] *Department of Mathematics and Descriptive Geometry, Faculty of Civil Engineering, STU Bratislava, Slovakia*

sheikhy.a@uk.ac.ir, radko.mesiar@stuba.sk

Abstract

In this work, we study the joint distribution function as well as the copula of $(X-Z,Y)$ where the random vector (X,Y,Z) is characterized by a copula $C_{X,Y,Z}$. We use this copula to analyze a measurement error model. Some theoretical results, several examples as well as a simulation study are proposed for illustration.

Keywords: Copula, noise, perturbation of copula, random vector, measurement error.

1 Introduction

One of the most important assumptions in regression analysis is that all the observations are correctly observed. However, in many applications this assumption is violated and the predictor variable cannot be directly or precisely measured, and instead, its surrogate measurement is observed with an error. The difference between the value we observe and the true value of the predictor variable is termed as measurement error and such models are called measurement error model. In these models, estimation based on the standard techniques leads to inconsistent estimates. Also, measurement error is known to cause biased parameter estimates (Carroll et al. (2006) [4]). Lack of power is a direct consequence of these misestimates and ignoring measurement error often leads to incorrect inferences about parameters (see, for example, Shalabh and Kuma, 2009 [8]). There exist many motivations to use measurement error models. In epidemiological studies, risk factors such as nutrient intake, systolic blood pressure or air pollutants are often subject to measurement error (see e.g., Song and Wang (2104) [11] and Kim et al. (2016) [15]). In econometric, indices such as trade balance, income, GDP, etc. are measured with errors. We refer to Singh et al. (2012) [11], Bingley and Martinello (2017) [3] and references therein for more application of these models.

Assuming that $(X_1, X_2, ..., X_d)$ is a random vector with the joint multivariate distribution function $F_{X_1,X_2,...X_d}$ and the univariate marginals F_{X_i}, $i = 1, 2, ..., d$, a copula is a multivariate function $C : [0,1]^d \to [0,1]$, that joins the multivariate distribution function to its marginals as

$$F_{X_1,X_2,...X_d}(x_1, x_2, ..., x_d) = C(F_{X_1}(x_1), F_{X_2}(x_2), ..., F_{X_d}(x_d)), \qquad (1)$$

which has the grounded, uniformly marginal and $d-$increasing properties. See e.g. Joe (1997) [13], Nelsen (2006) [21] and Durante and Sempi (2016) [11] for more details.

In the special case $d = 2$, the three basic copula are the upper Fréchet-Hoeffding, the lower Fréchet-Hoeffding and the product copula which are denoted by $M_2(u,v) = \min(u,v)$, $W_2(u,v) = \max(0, u+v-1)$ and $\Pi_2(u,v) = uv$, respectively. A generalization of these copulas can be defined as their convex combination as

$$C_2^F(u,v) = \alpha M_2(u,v) + (1-\alpha-\beta)\Pi_2(u,v) + \beta W_2(u,v),$$

where $\alpha, \beta \in [0,1]$ and $\alpha+\beta \le 1$. The upper Fréchet-Hoeffding copula corresponds to $\alpha = 1$ which depicts the complete positive dependence and the lower Fréchet-Hoeffding copula arises when $\beta = 1$, viz, the complete negative dependence.

Corresponding Author: A. Sheikhi
Received: February 2020; Revised: April 2020; Accepted: June 2020.

Also, the product copula corresponds to $\alpha = \beta = 0$ which reveals the independence (Fréchet, 1958). One special case of this family of copulas when $\alpha = 0$ is known as the Fréchet-Mardia copulas defined as

$$C_2^{FM}(u,v) = \beta M_2(u,v) + (1-\beta)\,\Pi_2(u,v), \tag{2}$$

where $\beta \in [0,1]$. Another important copula family which includes the product copula is the Farlie-Gumbel-Morgenstern (FGM) family of copulas having the form

$$C_2^{FGM}(u,v) = uv[1+\theta(1-u)(1-v)], \tag{3}$$

where $\theta \in [-1,1]$. Evidently, (3) reduces to the product copula if $\theta = 0$.

Also, the perturbed copula which deals on the effect of a perturbation on a copula, has been disused by many researchers. We refer to Durante et al. (2013) [9] who provided a family of perturbed copulas based on a perturbation of a given copula and Mesiar et al. (2019) [20] who obtained the perturbed copula $C_{X+Z,Y}$. Sheikhi et al. (2020) [24] investigated the the copula function $C_{X+Z,Y+Z}$ as well as the joint distribution function of the copula-induced random vector $(X+Z, Y+Z)$. Also, see Mesiar et al. (2015) [19] and Komorník et al. (2016) [18] for constructions of copulas using perturbations.

On the other hand, the copula-based regression models have been investigated in the literature. Kolev et al (2009) [17] have presented a survey in copula-based regression models. Chen and Hansen (2017) [5] have discussed a copula regression models for discrete and mixed bivariate responses. Crane and Hoek (2008) [7] have stated the copula-based regression equations based on the conditional expectation, see also, [14], [22] and [27].

The main aim of this work was to accomplish a copula-based measurement error analysis. In Section 2, we introduce the distribution of the difference of two copula-induced random variables as well as the the distribution function and the copula of $(X-Z,Y)$. We apply the results of section 2 to investigate a copula-based measurement error regression equation and section 4 provides some numerical analysis using simulation studies. Finally, some concluding remarks are added.

2 The copula-induced subtract perturbation

Consider two continuous random variables X and Y connected by a copula. The distribution of their difference $X-Y$ is desired. It is known that when X, Y are independent random variables the distribution of their difference, $T = X - Y$, can be obtained as

$$F_T(t) = \int_{-\infty}^{\infty} F_X(t+y)dF_Y(y) \quad y \in \mathbb{R}, \tag{4}$$

where F_X and F_Y are the CDF of X and Y respectively. When the random variables are not independent and their joint distribution is associated to a copula, then the following lemma provides the distribution function of their difference.

Lemma 2.1. *If the random vector (X,Y) follows a distribution function $F_{X,Y}(x,y)$ with the connecting copula function $C_{X,Y}(x,y)$ then the distribution function of $T = X - Y$ is*

$$F_T(t) = \int D_2C_{X,Y}(F_X(y+t), F_Y(y))\,\mathrm{d}F_Y(y), \tag{5}$$

where $D_2C_{X,Y}(u,v) = \frac{\partial C_{X,Y}(u,v)}{\partial v}$ if its derivative exists, and zero otherwise. Moreover, the integral with no bounds, denotes the integral on the entire domain.

Proof. From the total probability theorem and $P(X < x|Y = y) = D_2C_{X,Y}(F_X(x), F_Y(y))$, see e.g., Nelsen (2006) [21], we have

$$F_T(t) = P(X-Y<t) = \int P(X<y+t|Y=y)\,\mathrm{d}F_Y(y) = \int D_2C_{X,Y}(F_X(y+t),F_Y(y))\,\mathrm{d}F_Y(y).$$

which yields the assertion. □

Alternatively, changing the roles of x and y, (5) is equivalent to

$$F_T(t) = \int D_1C_{X,Y}(F_X(x), F_Y(t+x))dF_X(x),$$

where $D_1C_{X,Y}(u,v) = \frac{\partial C_{X,Y}(u,v)}{\partial u}$.

The following two examples play important roles in the sequel. Note that though the computations below are rather trivial, we give them here for the readers's convenience (similarly, in the case of the other examples).

Example 2.2. *Let X and Y be random variables uniformly distributed on $[0,1]$ linked by the Farlie-Gumbel-Morgenstern copula (3), with parameter $\theta \in [-1,1]$, then by taking partial derivation of (3), with respect to $F_Y(y)$, we obtain*

$$D_2C_{X,Y}(F_X(y+t), F_Y(y)) = (y+t) + \theta(y+t)(1-y-t)(1-2y),$$

and by substituting $D_2C_{X,Y}(F_X(y+t), F_Y(y))$ in (5) the following two cases arise:
1- If $-1 \leq t < 0$ then

$$F_{X-Y}(t) = \int_{-t}^{1} [(y+t) + \theta(y+t)(1-y-t)(1-2y)]dy = \frac{-\theta t^4 + 3(\theta+1)t^2 + 2(\theta+3)t + 3}{6}.$$

2- If $0 \leq t \leq 1$ then

$$F_{X-Y}(t) = \int_{0}^{1-t} [(y+t) + \theta(y+t)(1-y-t)(1-2y)]dy + \int_{1-t}^{1} dy = \frac{\theta t^4 - 3(\theta+1)t^2 + 2(\theta+3)t + 3}{6},$$

which yields

$$F_{X-Y}(t) = \begin{cases} \frac{-\theta t^4+3(\theta+1)t^2+2(\theta+3)t+3}{6} & \text{if} \quad -1 \leq t < 0 \\ \frac{\theta t^4-3(\theta+1)t^2+2(\theta+3)t+3}{6} & \text{if} \quad 0 \leq t \leq 1 \end{cases}. \tag{6}$$

We have to emphasize that if $t = 0$ then both two parts of (6) are same and equal to $F_{X-Y}(t) = \int_0^1 [(y+t) + \theta(y+t)(1-y-t)(1-2y)]dy = \frac{1}{2}$.

A special case of (6) occurs when $\theta = 0$ and yields the distribution of the difference of two independent uniform random variables which is a triangular-type distribution.

Example 2.3. *Let X and Y be random variables uniformly distributed on $[0,1]$ linked by Fréchet-Mardia copula (2), with parameter $\beta \in [0,1]$, then similar to Example 2.2, from (2) we have*

$$D_2C_{X,Y}(u,v) = \begin{cases} \beta + (1-\beta)u & \text{if} \ \ u \geq v-1 \\ (1-\beta)u & \text{if} \ \ u \leq v-1 \end{cases}.$$

So, from (5) we have the following two cases:
1- If $-1 \leq t < 0$ then

$$F_{X-Y}(t) = \int_{-t}^{\frac{1-t}{2}} [(1-\beta)(t+y)]dy + \int_{\frac{1-t}{2}}^{1} [\beta + (1-\beta)(t+y)]dy = \frac{t^2(1-\beta) + t(2-\beta) + 1}{2}.$$

2- If $0 \leq t \leq 1$ then

$$F_{X-Y}(t) = \int_{0}^{\frac{1-t}{2}} [(1-\beta)(t+y)]dy + \int_{\frac{1-t}{2}}^{1-t} [\beta + (1-\beta)(t+y)]dy + \int_{1-t}^{1} dy = \frac{-t^2(1-\beta) + t(2-\beta) + 1}{2},$$

and hence the distribution function of $X - Y$ is

$$F_{X-Y}(t) = \begin{cases} \frac{t^2(1-\beta)+t(2-\beta)+1}{2} & \text{if} \quad -1 \leq t < 0 \\ \frac{-t^2(1-\beta)+t(2-\beta)+1}{2} & \text{if} \quad 0 \leq t \leq 1 \end{cases}. \tag{7}$$

It is worth noting that, these two parts of (7) are equal to $\frac{1}{2}$ when $t = 0$. Also, by substituting $\beta = 1$ in Example 2.3 we obtain the distribution of the difference of X and Y where they are linked by the lower Fréchet-Hoeffding copula as $F_{X-Y}(t) = \frac{1+t}{2}, \ \ -1 \leq t \leq 1$. Moreover, for $\beta = 0$, (7) gives the CDF of the $X - Y$, which is again a triangular-type distribution.

In the rest of the paper, we will consider X, Y and Z are connected by a 3-copula $C_{X,Y,Z}$ where X, Y and Z are continuous random variables.

Theorem 2.4. *If $C_{X,Y,Z}(u,v,w)$ is the copula function of the random vector (X,Y,Z), then the joint distribution and the copula function of the vector $(X - Z, Y)$ are given respectively by*

$$F_{X-Z,Y}(t,y) = \int D_3C_{X,Y,Z}(F_X(t+z), F_Y(y), F_Z(z)) \, \mathrm{d}F_Z(z), \tag{8}$$

and

$$C_{X-Z,Y}(u,v) = \int D_3C_{X,Y,Z}[F_X(F_{X-Z}^{-1}(u) + F_Z^{-1}(w)), v, w] \, \mathrm{d}w, \tag{9}$$

where $D_3C_{X,Y,Z}(u,v,w) = \frac{\partial C_{X,Y,Z}(u,v,w)}{\partial w}$ if this derivative exists, and $D_3C_{X,Y,Z}(u,v,w) = 0$ otherwise.

Proof. Using the theorem of total probability we have

$$\begin{aligned} F_{X-Z,Y}(t,y) &= P(X-Z<t, Y<y) = \int P(X<t+z, Y<y|Z=z)\,\mathrm{d}F_Z(z) \\ &= \int D_3C_{X,Y,Z}(F_X(t+z), F_Y(y), F_Z(z))\,\mathrm{d}F_Z(z), \end{aligned}$$

which is (8). Now, defining $w = F_Z(z)$ and $v = F_Y(t)$, the perturbed copula $C_{X-Z,Y}$ is given by

$$C_{X-Z,Y}(u,v) = \int D_3C_{X,Y,Z}[F_X(F^{-1}_{X-Z}(u) + F^{-1}_Z(w)), v, w]\,\mathrm{d}w,$$

which finishes the proof. □

A special case of the theorem (2.4) arises when the perturbed random variable Z is independent of X and Y, and is presented in the following corollary.

Corollary 2.5. *If X, Y, Z are connected by a copula $C_{X,Y,Z}$ in which, $C_{X,Y,Z}(u,v,w) = wC_{X,Y}(u,v)$ then the joint distribution function and the copula function of $X-Z, Y$ are respectively given by*

$$F_{X-Z,Y}(t,y) = \int\int_{-\infty}^{y} D_2C_{X,Y}(F_X(t+z), F_Y(u))\mathrm{d}u\mathrm{d}F_Z(z). \tag{10}$$

and

$$C_{X-Z,Y}(u,v) = \int\int_{0}^{v} D_2C_{X,Y}\left(F_X(F^{-1}_{X-Z}(u)+z), r\right)\,\mathrm{d}r\mathrm{d}F_Z(z), \tag{11}$$

where $D_2C_{X,Y}(u,v) = \frac{\partial C_{X,Y}(u,v)}{\partial v}$ if this derivative exists, and $D_2C_{X,Y}(u,v) = 0$ otherwise.

Proof. Similar to the proof of Theorem (2.4), since the random variable Z is independent of X and Y, we have

$$P(X<t+z, Y<y|Z=z) = P(X<t+z, Y<y) = \int_{-\infty}^{y} D_2C_{X,Y}(F_X(t+z), F_Y(u))\mathrm{d}u.$$

□

Example 2.6. *Let X, Y and Z be random variables uniformly distributed on $[0,1]$ with the connected copula $C_{X,Y,Z}(u,v,w) = wM_2(u,v)$ then the joint distribution function and the copula function of $(X-Z,Y)$ are respectively*

$$F_{X-Z,Y}(t,y) = \begin{cases} \frac{(1+t)^2}{2} & \text{if } t\le 0,\ y\ge 1+t \\ \frac{(1+t)^2}{2} - \frac{(1+t-y)^2}{2} & \text{if } t\le 0,\ y\le 1+t \\ y & \text{if } t\ge 0,\ y\le t \\ y - \frac{(y-t)^2}{2} & \text{if } t\ge 0,\ y\ge t, \end{cases} \tag{12}$$

and

$$C_{X-Z,Y}(u,v)) = \begin{cases} u & \text{if } 0\le u\le \frac{1}{2},\ \sqrt{2u}\le v \\ v\sqrt{2u} - \frac{v^2}{2} & \text{if } 0\le u\le \frac{1}{2},\ \sqrt{2u}\ge v \\ v & \text{if } \frac{1}{2}\le u\le 1,\ 1-\sqrt{2(1-u)}\ge v \\ v - \frac{(\sqrt{2(1-u)}+v-1)^2}{2} & \text{if } \frac{1}{2}\le u\le 1,\ 1-\sqrt{2(1-u)}\le v. \end{cases} \tag{13}$$

For detailed computations see Appendix.

Figure 1 presents the joint distribution function as well as the copula function of $(X-Z,Y)$ which show non-overlapping as well as the continuity of both of 12 and 13. We note that the copula function (13) is the same as the copula of $(X+Z,Y)$ obtained by Mesiar et al. (2019) under the same assumptions as in Example 2.6; this is because $X-Z$ is a strictly increasing function of $X+Z$ in this example.

Example 2.7. *Let X, Y and Z be exponential random variables in which $X, Y \sim exp(\lambda)$ and $Z\sim exp(\varepsilon)$ with the connected copula $C_{X,Y,Z}(u,v,w) = wM_2(u,v)$ then the joint distribution function and the copula of $(X-Z,Y)$ are given respectively by*

$$F_{X-Z,Y}(t,y) = \begin{cases} \frac{\lambda}{\lambda+\varepsilon}[e^{\varepsilon t}(1-e^{-\lambda y-\varepsilon y})] & \text{if } t\le 0 \\ 1-e^{-\lambda y} & \text{if } t\ge 0,\ y\le t \\ 1-e^{-\lambda t} + \frac{\lambda}{\lambda+\varepsilon}(e^{-\lambda t} - e^{-\lambda y-\varepsilon(y-t)}) & \text{if } t\ge 0,\ y\ge t, \end{cases} \tag{14}$$

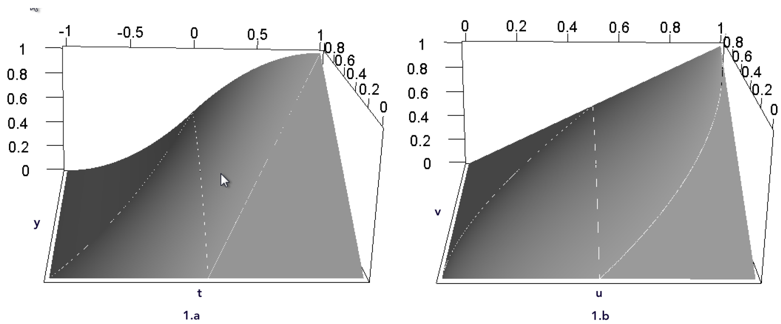

Figure 1: The CDF and the copula function of $(X-Z,Y)$ from Example 2.6.

and

$$C_{X-Z,Y}(u,v)) = \begin{cases} u[1-(1-v)^{\frac{\lambda+\varepsilon}{\lambda}}] & if \;\; 0 \le u \le \frac{\lambda}{\lambda+\varepsilon} \\ v & if \;\; \frac{\lambda}{\lambda+\varepsilon} \le u \le 1 \;\; and \;\; \frac{1-v}{1-u} \le \frac{\lambda+\varepsilon}{\varepsilon} \\ u+\frac{\lambda}{\lambda+\varepsilon}[(1-v)^{\frac{\lambda+\varepsilon}{\lambda}}(1-u)^{\frac{\lambda+\varepsilon}{\varepsilon}}]^{\frac{-\varepsilon}{\lambda}} & if \;\; \frac{\lambda}{\lambda+\varepsilon} \le u \le 1 \;\; and \;\; \frac{1-v}{1-u} \ge \frac{\lambda+\varepsilon}{\varepsilon} \end{cases} . \quad (15)$$

For detailed computations see Appendix.

For $\lambda = 1$ and $\varepsilon = 2$, Figure 2.a presents the joint distribution function of $(X-Z,Y)$ in the range of $-3 \le t \le 3$ and $0 \le y \le 3$. Also, its copula function is depicted in Fig 2.b.

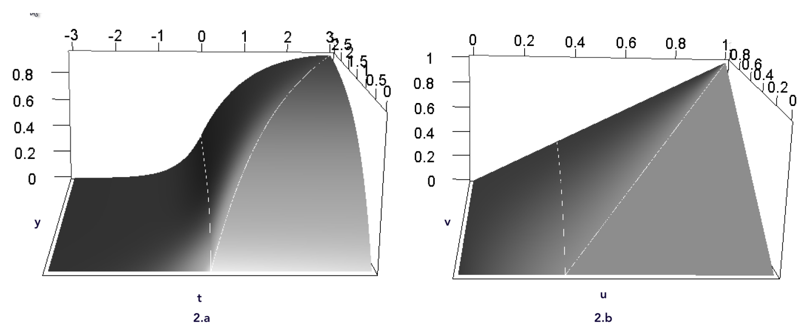

Figure 2: The CDF and the copula function of $(X-Z,Y)$ from Example 2.7 .

3 Copula-based measurement error models

Consider the random variables X and Y that are linked by a copula $C_{X,Y}$ and the random variable Z that is independent of X and Y. In this section we use the results of the previous section to develop a copula-based measurement error model.

Consider the regression model

$$\begin{aligned} Y &= \alpha + \beta X^* + \delta \\ X &= X^* + Z, \end{aligned} \quad (16)$$

where X^* is the latent predictor variable and X is its surrogate variable which is measured with the error term Z. This types of models are called *Measurement Error Model.* A reasonable assumption in these models is that the error term Z is independent of the observed variables X and Y (see, e.g., [4] and [29]). In what follows we assume that X, Y and Z are connected by a copula $C_{X,Y,Z}(u,v,w) = wC_{X,Y}(u,v)$. Also, for the sake of the notational simplicity we adopt $T = X^* = X - Z$. The following corollary enables us to get the desired regression function.

Corollary 3.1. *Under the assumptions of Corollary 2.5, the regression equation of model (16) can be obtained as*

$$r_{Y|T}(t) = \int y f_{Y|T}(y|t)dy = \int yD_{12}C_{X-Z,Y}(F_{X-Z}(t), F_Y(y))dF_Y(y) \quad (17)$$

where $D_{12}C_{X,Y}(u,v)=\frac{\partial^2 C_{X,Y}(u,v)}{\partial u \partial v}$ if this derivative exists, and $D_{12}C_{X,Y}(u,v)=0$ otherwise.

Proof. From (16) we obtain $Y=\alpha+\beta(X-Z)+\varepsilon$ and we have

$$r_{Y|T}(t)=E_{Y|T}(y|t)=\int y\frac{\partial}{\partial y}P(Y\le y|T=t)dy=\int y\frac{\partial}{\partial y}D_1C_{T,Y}(F_T(t),F_Y(y))dy=\int yD_{12}C_{T,Y}(F_T(t),F_Y(y))dF_Y(y),$$

which proves the claim. □

We note that (17) is not necessarily a linear regression function, although in some special cases it yields a linear equation. For an instance see Theorem (2.1) of Sungur (2006) [27]. The following example considers a more general case of Example 2.6. Without loss of generality we assume that $\varepsilon\le 1$. The proof for $\varepsilon\ge 1$ is analogous.

Example 3.2. *Let X and Y be random variables uniformly distributed on $[0,1]$ and random variable Z has uniform distribution on $[0,\varepsilon]$, $0<\varepsilon\le 1$ and $C_{X,Y,Z}(u,v,w)=uM_2(v,w)$, then from (16) we can obtain $Y=\alpha+\beta(X-Z)+\varepsilon$ and defining $T=X-Z$, similar to the proof of Example 2.6 we obtain the joint distribution of T and Y as*

$$F_{T,Y}(t,y)=\begin{cases}0 & if\ \ t\le-\varepsilon\\ \frac{(\varepsilon+t)^2-(\varepsilon+t-y)^2}{2\varepsilon} & if\ -\varepsilon\le t\le 0\le y\le\varepsilon+t\\ \frac{(\varepsilon+t)^2}{2\varepsilon} & if\ \ -\varepsilon\le t\le 0\le y\ge\varepsilon+t\\ y & if\ \begin{array}{l}0\le t\le 1-\varepsilon,\ 0\le y\le t\ \ or\\ 1-\varepsilon\le t\le 1,\ 0\le y\le t\end{array}\\ y-\frac{(y-t)^2}{2\varepsilon} & if\ \begin{array}{l}0\le t\le 1-\varepsilon,\ t\le y\le t+\varepsilon\ \ or\\ 1-\varepsilon\le t\le 1,\ \ t\le y\le 1\end{array}\end{cases}.$$

Then, taking its derivative and obtaining the marginal density of $f_T(t)$, the conditional density of Y given $T=t$ is obtained as

$$f_{Y|T}(y|t)=\begin{cases}\frac{1}{t+\varepsilon} & if\ -\varepsilon\le t\le 0\le y\le\varepsilon+t\\ \frac{1}{\varepsilon} & if\ \ 0\le t\le 1-\varepsilon,\ t\le y\le t+\varepsilon\ \ t\\ \frac{1}{1-t} & if\ 1-\varepsilon\le t\le 1,\ \ t\le y\le 1\\ 0 & o.w.\end{cases}.$$

Hence, taking its expectation, we readily have

$$E_{Y|T}(y|t)=\begin{cases}\frac{\varepsilon+t}{2} & if\ -\varepsilon\le t\le 0\\ \frac{\varepsilon+2t}{2} & if\ \ 0\le t\le 1-\varepsilon\\ \frac{1+t}{2} & if\ \ 1-\varepsilon\le t\le 1\end{cases},$$

Consequently, the regression equation of model (16) can be written as

$$r_{Y|T}(t)=\begin{cases}\frac{\varepsilon+t}{2} & if\ -\varepsilon\le t\le 0\\ \frac{\varepsilon+2t}{2} & if\ \ 0\le t\le 1-\varepsilon\\ \frac{1+t}{2} & if\ 1-\varepsilon\le t\le 1\end{cases}. \tag{18}$$

We may note that if $\varepsilon\to 0$, then (18) reduces to $r_{Y|T}(t)=t,\ 0\le t\le 1$ as reported by Crane and Hock (2008)[7]. Also, if $\varepsilon=1$ then $r_{Y|T}(t)=\frac{1+t}{2},\ -1\le t\le 1$.

Example 3.3. *Let X, Y and Z be exponential random variables in which $X,Y\sim exp(\lambda)$ and $Z\sim exp(\varepsilon)$ with the connected copula $C_{X,Y,Z}(u,v,w)=wM_2(u,v)$ then using (14) we obtain the joint density of $X-Z,Y$ as*

$$f_{X-Z,Y}(t,y)=\lambda\varepsilon e^{-\lambda y-\varepsilon(y-t)},\qquad t\le 0\ \ or\ \ t\ge 0,\ y\ge t.$$

So, the conditional density is

$$f_{Y|X-Z}(y|t)=\begin{cases}(\lambda+\varepsilon)e^{-(\lambda+\varepsilon)y} & if\ \ t\le 0,y\ge 0\\ (\lambda+\varepsilon)e^{-(\lambda+\varepsilon)y+(\lambda+\varepsilon)t} & if\ \ t\ge 0,\ y\ge t\end{cases}.$$

So, its expectation and hence the regression equation of model (16) will be

$$r_{Y|T}(t)=\begin{cases}\frac{1}{\lambda+\varepsilon} & if\ t\le 0\\ t+\frac{1}{\lambda+\varepsilon} & if\ \ t\ge 0\end{cases}. \tag{19}$$

4 Simulation results

In this section we present a simulation study to evaluate the results of the previous section. To visualize the results of Example 3.2, the simulations were based on model (16) when the minimal copula couples the random variables X and Y which are uniformly distributed on $[0, 1]$ and random variable Z has uniform distribution on $[0, \varepsilon]$, for $\varepsilon = 0.01,\ 0.25,\ 0.5$ and 1. Taking in mind that $X = Y$, a.s., we take a sample size of 200 from these distributions and repeat this simulation 10 times and calculate the average of the simulated data. Using (18), Figure 3 depicts the measurement error regression equations for various values of ε. The Left top panel shows the simulated data as well as the regression function when the random variable Z is simulated from $U[0, \varepsilon]$ with $\varepsilon = 1$. The right top panel, the left bottom panel and the right bottom panel respectively show the simulated data and their regression function for $\varepsilon = 0.5,\ 0.25$ and 0.01.

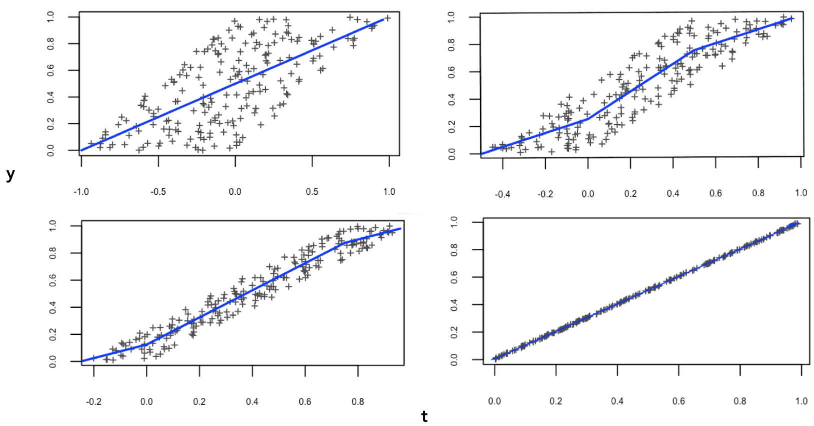

Figure 3: Simulated data and their fitted regression equation from Example 3.2.

Again, in order to investigate the results of Example 3.3, we generate data from exponential distributions with different values of λ and ε. Figure 4 presents these simulated data and their fitted regression equation. In the left top, right top, left middle and the right middle panel we simulate Z from $exp(\varepsilon)$ with $\varepsilon = 10,\ 1,\ 0.1$ and 0.01 respectively; while in all of these four panels the simulated values X and Y came from the exponential distribution with $\lambda = 1$. The left bottom of Figure 4 shows these simulated data and the fitted regression equation for $\lambda = 10$ and $\varepsilon = 1$ and the right bottom is for $\lambda = 100$ and $\varepsilon = 100$.

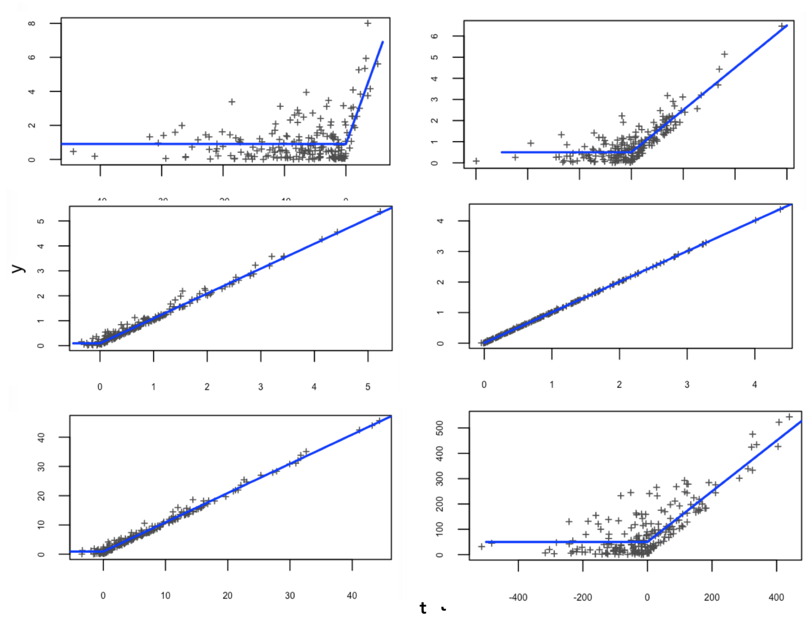

Figure 4: Simulated data and their fitted regression equation from Example 3.3.

5 Conclusion

In this work we have studied the joint distribution and the copula function of the random vector $(X-Z, Y)$ when X, Y and Z are associated by a ternary copula $C_{X,Y,Z}$. Some examples of the uniform and the exponential distributions and some connecting copulas such as the three basic copulas, Farlie-Gumbel-Morgenstern copulas and the Fréchet-Mardia copulas were presented. Applying these joint distribution functions and copulas we carry out measurement error regression models when there are connected copulas which couple variables. These regression equations are multi-function linear regression functions which reduce to a linear regression in some special case and so they are the general cases of what was obtained by [7] and [27].

Since in this work we only consider that the random variables X, Y and Z are related by a copula function $C_{X,Y,Z}(u,v,w) = wM_2(u,v)$, this work has a great potential to improve in many circumstances. One may extend this results when the random variables X, Y and Z are connected via some other copulas. Also, these results may improve in a multiple regression scenario or when there are more than one variables which are measured with error.

Another interesting subject in this field is considering another copula families. Since there is a longitudinal literature in measurement error models with elliptical distributions, our next ongoing research is to consider the Archimedean copulas, and especially the Gaussian copulas.

Acknowledgement

The authors wish to thank the anonymous reviewers for the comments and suggestions that did lead to a significant improvement of the original manuscript. The work of the second author was supported by the project VEGA 1/0891/17.

References

[1] P. Arbenz, P. Embrechts, G. Puccetti, *The AEP algorithm for the fast computation of the distribution of the sum of dependent random variables*, Bernoulli, **17**(2) (2011), 562-591.

[2] P. Arbenz, P. Embrechts, G. Puccetti, *The GAEP algorithm for the fast computation of the distribution of a function of dependent random variables*, Stochastics, **84**(5-6) (2012), 569-597.

[3] P. Bingley, A. Martinello, *Measurement error in income and schooling and the bias of linear estimators*, Journal of Labor Economics, **35**(4) (2017), 1117-1148.

[4] R. Carroll, J. Ruppert, L. Stefanski, C. Crainiceanu, *Measurement error in nonlinear models: A modern perspective*, 2nd edition, CRC Press, Boca Raton, Fla, 2006.

[5] Y. Chen, T. Hanson, *Copula regression models for discrete and mixed bivariate responses*, Journal of Statistical Theory and Practice, **11**(4) (2017), 515-530.

[6] U. Cherubini, F. Gobbi, S. Mulinacci, *Convolution copula econometrics*, Springer-Briefs in Statistics, 2016.

[7] G. J. Crane, J. Van der Hoek, *Conditional expectation formulae for copulas*, Australian and New Zealand Journal of Statistics, **50**(1) (2008), 53-67.

[8] G. J. Crane, J. Van der Hoek, *Consistent estimation of regression parameters under replicated ultrastructural model with non-normal errors*, Journal of Statistical Computation and Simulation, **79**(3) (2009), 251-274.

[9] F. Durante, J. Fernández-Sánchez, M. Úbeda-Flores, *Bivariate copulas generated by perturbations*, Fuzzy Sets and System, **228** (2013), 137-144.

[10] F. Durante, C. Sempi, *A note on the compatibility of bivariate copulas*, Communications in Statistics-Theory and Methods, **43**(9) (2014), 1918-1923.

[11] F. Durante, C. Sempi, *Principles of copula theory*, CRC Chapman and Hall: Boca Raton, FL, USA, 2016.

[12] I. Gijbels, K. Herrmann, *On the distribution of sums of random variables with copula-induced dependence*, Insurance: Mathematics and Economics, **59**(C) (2014), 27-44.

[13] H. Joe, *Multivariate models and dependence concepts*, Chapman and Hall, 1997.

[14] D. Kim, J. Kim, *Analysis of directional dependence using asymmetric copula-based regression models*, Journal of Statistical Computation and Simulation, **84**(9) (2010), 1990-2010.

[15] S. Kim, Y. Li, D. Spiegelman, *A semiparametric copula method for Cox models with covariate measurement error*, Lifetime Data Analysis, **22**(1) (2016), 1-16.

[16] A. Kolesařová, R. Mesiar, C. Sempi, *Measure-preserving transformations, copula and compatibility*, Mediterranean Journal of Mathematics, **5**(3) (2008), 325-339.

[17] N. Kolev, P. Delhi, *Copula-based regression models: A survey*, Journal of Statistical Planning and Inference, **139**(11) (2009), 3847-3856.

[18] J. Komorník, M. Komorníková, J. Kalická, *Dependence measures for perturbations of copulas*, Journal of Statistical Theory and Applications, **15**(2) (2016), 153-160.

[19] R. Mesiar, M. Komorníková, J. Komorník, *Perturbation of bivariate copula*, Fuzzy Sets and System, **268** (2015), 127-140.

[20] R. Mesiar, A. Sheikhi, M. Komorníková, *Random noise and perturbation of copulas*, Kybernetika, **55**(2) (2019), 422-434.

[21] R. B. Nelsen, *An introduction to copulas*, Second Edition. Springer Series in Statistics, Springer-Verlag, New York, 2006.

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

[23] J. M. Sarabia, E. G. Deniz, F. Prieto, V. Jordá, *Aggregation of dependent risks in mixtures of exponential distributions and extensions*, ASTIN Bulletin: The Journal of the IAA, **48**(3) (2018), 1079-1107.

[24] A. Sheikhi, V. Amirzadek, R. Mesiar, *A comprehensive family of copulas to modeling bivariate random noise and perturbation*, Fuzzy Set and Systems, to appear. https://doi.org/10.1016/j.fss.2020.04.010.

[25] A. Sklar, *Fonctions de répartition à n dimensions et leurs marges*, Publications de lÍnstitut de statistique de lÚniversité de Paris, **8** (1959), 229-231.

[26] S. Sukhbir, K. Jain, S. Sharma, *Using stochastic prior information in consistent estimation of regression coefficients in replicated measurement error model*, Journal of Multivariate Analysis, **111** (2012), 198-212.

[27] E. Sungur, *Some observations on copula regression functions*, Communications in Statistics-Theory and Methods, **34**(9) (2005), 1967-1978.

[28] R. Wang, *Sum of arbitrarily dependent random variables*, Electronic Journal of Probability, **19**(84) (2014), 1-18.

[29] T. Wansbeek, E. Meijer, *Measurement error and latent variables in econometrics*, Washington D.C.: Advanced Textbooks in Economics, **37** (2000).

[30] R. C. Williamson, R. C. Downs, *Probabilistic arithmetic I: Numerical methods for calculating convolutions and dependency bounds*, International Journal of Approximate Reasoning, **4** (2014), 89-158.

[31] S. Xiao, C. Wang, *Proportional hazards model with covariate measurement error and instrumental variables*, Journal of the American Statistical Association, **109**(508) (2014), 1636-646.

6 Appendix

Computation for Example 2.6

Using (10) we have the following cases to obtain the joint CDF of $X-Z,Y$:

1. If $t \leq 0$ and $y \geq 1+t$, then

$$F_{X-Z,Y}(t,y) = \int_0^{1+t}\int_{x-t}^{1} \mathrm{d}z\mathrm{d}x = \frac{(1+t)^2}{2}.$$

2. If $t \leq 0$ and $y \leq 1+t$, then

$$F_{X-Z,Y}(t,y) = \int_0^y \int_{x-t}^1 \mathrm{d}z\mathrm{d}x = \frac{(1+t)^2}{2} - \frac{(1+t-y)^2}{2}.$$

3. If $t \geq 0$ and $y \leq t$, then

$$F_{X-Z,Y}(t,y) = \int_0^y \int_0^1 \mathrm{d}z\mathrm{d}x = y.$$

4. If $t \geq 0$ and $y \geq t$, then

$$F_{X-Z,Y}(t,y) = \int_0^t \int_0^1 \mathrm{d}z\mathrm{d}x + \int_t^y \int_{x-t}^1 \mathrm{d}z\mathrm{d}x = y - \frac{(y-t)^2}{2},$$

which are (12). Now, since X and Z are independent, the distribution of $X-Z$ follows from the triangular distribution and hence the inverse of its distribution function will be $F^{-1}_{X-Z}(u) = \begin{cases} \sqrt{2u}-1 & \text{if } 0 \leq u \leq \frac{1}{2} \\ 1-\sqrt{2(1-u)} & \text{if } \frac{1}{2} \leq u \leq 1 \end{cases}$.

Substituting $F^{-1}_{X-Z}(u)$ and $F^{-1}_Y(v)$ into (12) readily yields (13).

Computation for Example 2.7

Similar to the proof of example (3), we have the following cases:

1. If $t \leq 0$, then

$$F_{X-Z,Y}(t,y) = \int_0^s \int_{x-t}^\infty \lambda\varepsilon e^{-\lambda x - \varepsilon z}\mathrm{d}z\mathrm{d}x = \frac{\lambda}{\lambda+\varepsilon}(e^{\varepsilon t} - e^{-\lambda y - \varepsilon(y-t)}).$$

2. If $t \geq 0$ and $y \leq t$, then

$$F_{X-Z,Y}(t,y) = \int_0^s \int_0^\infty \lambda\varepsilon e^{-\lambda x - \varepsilon z}\mathrm{d}z\mathrm{d}x = 1 - e^{-\lambda y}.$$

3. If $t \geq 0$ and $y \geq t$, then

$$F_{X-Z,Y}(t,y) = \int_0^t \int_0^\infty \lambda\varepsilon e^{-\lambda x - \varepsilon z}\mathrm{d}z\mathrm{d}x + \int_t^s \int_{x-t}^\infty \lambda\varepsilon e^{-\lambda x - \varepsilon z}\mathrm{d}z\mathrm{d}x = 1 - e^{-\lambda t} + \frac{\lambda}{\lambda+\varepsilon}(e^{-\lambda t} - e^{-\lambda y - \varepsilon(y-t)}),$$

which are (14). Now, since X and Z are independent, the distribution of $X-Z$ follows from the double exponential distribution with the joint distribution function

$$F_{X-Z}(t) = \begin{cases} \frac{\lambda}{\lambda+\varepsilon} e^{\frac{t}{\varepsilon}} & \text{if } t \leq 0 \\ 1 - \frac{\lambda}{\lambda+\varepsilon} e^{\frac{-t}{\lambda}} & \text{if } t \geq 0 \end{cases}.$$

Hence, (15) easily obtains by substituting the inverse of $F_{X-Z}(t)$ and the inverse of $F_Y(y)$ into (14).

Iranian Journal of Fuzzy Systems
Volume 17, Number 6 (2020), pp. 207

University of Sistan and Baluchestan

Copula-based measurement error models

A. Sheikhi and R. Mesiar

مدل‌های خطای اندازه‌گیری برمبنای Copula (عضو رابط)

چکیده. در این کار، تابع توزیع مشترک، همچنین عضو رابط (x-z, y) را که بردار تصادفی (X, Y, Z) توسط یک عضو رابط $C_{x, y, z}$ مشخص می‌شود، مورد مطالعه قرار می‌دهیم. این عضو رابط را جهت آنالیز یک مدل خطای اندازه‌گیری به کار می‌بریم. برای روشن شدن مطلب، برخی از نتایج نظری، مثال-های متعدد و همچنین یک بررسی شبیه‌سازی پیشنهاد شده‌است.