INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION
PAPER

2012/10

COPULA-BASED REGRESSION ESTIMATION AND INFERENCE

NOH, N., EL GHOUCH, A. and T. BOUEZMARNI

Copula-Based Regression Estimation and Inference

Hohsuk Noh *
Université catholique de Louvain

Anouar El Ghouch †
Université catholique de Louvain

Taoufik Bouezmarni
Université de Sherbrooke

March 27, 2012

Abstract

In this paper we investigate a new approach of estimating a regression function based on copulas. The main idea behind this approach is to write the regression function in terms of a copula and marginal distributions. Once the copula and the marginal distributions are estimated we use the plug-in method to construct the new estimator. Because various methods are available in the literature for estimating both a copula and a distribution, this idea provides a rich and flexible alternative to many existing regression estimators. We provide some asymptotic results related to this copula-based regression modeling when the copula is estimated via profile likelihood and the marginals are estimated nonparametrically. We also study the finite sample performance of the estimator and illustrate its usefulness by analyzing data from air pollution studies.

1 Introduction

Let $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ be a random vector of dimension $d \geq 1$ and Y be a random variable with cumulative distribution function (c.d.f.) F_0 and density function f_0. Y is our response variable and $\boldsymbol{X}$ is our set of covariates. We denote by F_j the c.d.f. of X_j and we denote by f_j its corresponding

*H. Noh acknowledges financial support from IAP research network P6/03 of the Belgian Government (Belgian Science Policy).

†A. El Ghouch acknowledges financial support from IAP research network P6/03 of the Belgian Government (Belgian Science Policy), and from the contract 'Projet d'Actions de Recherche Concertées' (ARC) 11/16-039 of the 'Communauté française de Belgique', granted by the 'Académie universitaire Louvain'.

density. For a given $\boldsymbol{x} = (x_1, \ldots, x_d)^\top$ we will use $\boldsymbol{F}(\boldsymbol{x})$ as a shortcut for $(F_1(x_1), \ldots, F_d(x_d))$. From the inspiring work of Sklar (1959), the c.d.f. of $(Y, \boldsymbol{X}^\top)^\top$ evaluated at $(y, \boldsymbol{x}^\top)$ can be expressed as $C(F_0(y), \boldsymbol{F}(\boldsymbol{x}))$, where C is the copula distribution of $(Y, \boldsymbol{X}^\top)^\top$, that is, the function from $[0,1]^{d+1}$ to $[0,1]$ defined by $C(u_0, u_1, \ldots, u_d) = P(F_0(Y) \leq u_0, F_1(X_1) \leq u_1, \ldots, F_d(X_d) \leq u_d)$. This is nothing but a joint distribution function with margins that are uniform over the unit interval $[0,1]$. Using copulas allows a complete separation of dependence modeling from the marginal distributions and by specifying a copula we summarize all the dependencies between margins, see Nelsen (2006) for more about this subject. From the definition of copula function, the conditional density of Y given $\boldsymbol{X}^\top$ is given by $f_0(y)\dfrac{c(F_0(y), \boldsymbol{F}(\boldsymbol{x}))}{c_{\boldsymbol{X}}(\boldsymbol{F}(\boldsymbol{x}))}$, where $c(u_0, \boldsymbol{u}) \equiv c(u_0, u_1, \ldots, u_d) = \dfrac{\partial^{d+1} C(u_0, u_1, \ldots, u_d)}{\partial u_0 \partial u_1 \ldots \partial u_d}$ is the copula density corresponding to C and $c_{\boldsymbol{X}}(\boldsymbol{u}) \equiv c_{\boldsymbol{X}}(u_1, \ldots, u_d) = \dfrac{\partial^d C(1, u_1, \ldots, u_d)}{\partial u_1 \ldots \partial u_d}$ is the copula density of $\boldsymbol{X}$. Obviously, the conditional mean, $m(\boldsymbol{x})$, of Y given $\boldsymbol{X} = \boldsymbol{x}$ can be written as

$$m(\boldsymbol{x}) = \mathbb{E}(Y w(F_0(Y), \boldsymbol{F}(\boldsymbol{x}))) = \frac{e(\boldsymbol{F}(\boldsymbol{x}))}{c_{\boldsymbol{X}}(\boldsymbol{F}(\boldsymbol{x}))}, \tag{1}$$

where $w(u_0, \boldsymbol{u}) = c(u_0, \boldsymbol{u}) / c_{\boldsymbol{X}}(\boldsymbol{u})$ and

$$e(\boldsymbol{u}) = \mathbb{E}(Y c(F_0(Y), \boldsymbol{u})) = \int_0^1 F_0^{-1}(u_0) c(u_0, \boldsymbol{u}) du_0. \tag{2}$$

The equality (1) shows that, given the marginals, one can obtain the mean regression function relating Y to $\boldsymbol{X}$ directly from the copula density, or equivalently the copula distribution of $(Y, \boldsymbol{X}^\top)^\top$. It also implies that the conditional mean is "just" a weighted mean with weights induced by the unknown "conditional" copula function w defined above. This relation is not new and has been already applied in Sungur (2005), Leong and Valdez (2005) and Crane and Van Der Hoek (2008) to compute the mean regression function corresponding to several well known copula families (Gaussian, t, Farlie-Gumbel-Morgenstern (FGM), Iterated FGM, Archimedean, etc.) with single ($d = 1$) and multiple covariate(s). To illustrate the idea, we briefly cite two examples :

- If the copula density of (Y, X_1) belongs to the FGM family with a parameter θ, i.e. $c(u_0, u_1) = 1 + \theta(1 - 2u_0)(1 - 2u_1)$, then we have

$$m(x_1) = \mathbb{E}(Y) + \theta(2F_1(x_1) - 1) \int F_0(y)(1 - F_0(y)) dy. \tag{3}$$

  A similar formula holds for the multiple covariate case; see Leong and Valdez (2005) .

- Let $\boldsymbol{\rho} = (corr(Y, X_1), \ldots, corr(Y, X_d))^\top$ and $\Sigma_{\boldsymbol{X}}$ denote the correlation matrix of $\boldsymbol{X}$. If the

copula of $(Y, \boldsymbol{X}^\top)^\top$ is Gaussian, then we have

$$m(\boldsymbol{x}) = \mathbb{E}[F_0^{-1}(\Phi(\boldsymbol{u}^\top \Sigma_{\boldsymbol{X}}^{-1}\boldsymbol{\rho} + \sqrt{1 - \boldsymbol{\rho}^\top \Sigma_{\boldsymbol{X}}^{-1}\boldsymbol{\rho}}Z))], \tag{4}$$

where $\boldsymbol{u} = (\Phi^{-1}(F_1(x_1)), \ldots, \Phi^{-1}(F_d(x_d)))^\top$, $Z \sim \mathcal{N}(0,1)$ and Φ is the standard normal cumulative distribution function.

Note that in the single covariate case we have, $c_{\boldsymbol{X}}(\boldsymbol{u}) \equiv c_{X_1}(u_1) = 1$ for all $u_1 \in [0,1]$. In such a case the weight function w coincides with the copula density c and (1) reduces to $m(x_1) = e(F_1(x_1)) = \mathbb{E}(Yc(F_0(Y), F_1(x_1)))$. Also, if the covariates are mutually independent then $c_{\boldsymbol{X}}(\boldsymbol{u}) = 1$ and $m(\boldsymbol{x})$ coincides with $e(\boldsymbol{F}(\boldsymbol{x}))$. In other words, $e(\boldsymbol{F}(\boldsymbol{x}))$, the numerator of $m(\boldsymbol{x})$ in (1), is the mean regression function of Y given $\boldsymbol{X}$ assuming independence between the covariates or, equivalently, assuming that the conditional density of $Y|\boldsymbol{X}$ is $f_0(y)c(F_0(y), \boldsymbol{F}(\boldsymbol{x}))$. Thus, in term of copulas, the mean regression function is the ratio of a numerator that only captures the mean dependence between Y and $\boldsymbol{X}$ and a denominator that captures the dependence within $\boldsymbol{X}$.

The equality (1) can also be used as an estimating equation. In fact, if $\hat{w}$, $\hat{F}_0$ and $\hat{F}_j$ are any given estimators for w, F_0 and F_j, respectively, then m can obviously be estimated by

$$\hat{m}(\boldsymbol{x}) = \int_{-\infty}^{\infty} y\hat{w}(\hat{F}_0(y), \hat{\boldsymbol{F}}(\boldsymbol{x}))d\hat{F}_0(y), \tag{5}$$

where $\hat{\boldsymbol{F}}(\boldsymbol{x}) = (\hat{F}_1(x_1), \ldots, \hat{F}_d(x_d))^\top$. To the best of our knowledge, such an approach has never been proposed or investigated in the literature in neither single nor multiple covariate case. To estimate w, one needs an estimator for the copula densities c and $c_{\boldsymbol{X}}$. The copula density $c_{\boldsymbol{X}}$ can be obtained form c by integration. In fact,

$$c_{\boldsymbol{X}}(\boldsymbol{u}) = \int_{-\infty}^{\infty} f_0(y)c(F_0(y), \boldsymbol{u}))dy = \int_0^1 c(u_0, \boldsymbol{u})du_0 \tag{6}$$

Therefore, given an estimator $\hat{c}$ for c, one can easily estimate $c_{\boldsymbol{X}}$ using the plug-in method and then estimate m by (5).

Since, in the literature, there are many different methods available for estimating a copula and a c.d.f., $\hat{m}(\boldsymbol{x})$ defines a new large class of interesting estimators. Depending on the method to estimate the components in (5), $\hat{m}(\boldsymbol{x})$ can be a nonparametric or a semiparametric or a fully parametric estimator. For example, using a nonparametric estimators for c, F_0 and F_j, $j = 1 \ldots, d$, leads to a fully nonparametric estimator. Nonparametric methods for estimating c include kernel smoothing estimators (see for example Gijbels and Mielniczuk (1990), Charpentier et al. (2006) and Chen and

Huang (2007)) and Bernstein estimator (see Bouezmarni et al. (2010)) to cite only two examples. In spite of the great flexibility of nonparametric methods, they are typically affected by the curse of dimensionality and they come with the difficult problem of selecting a good smoothing parameter. On the other hand imposing a parametric structure on both the copula and marginal distributions can lead to severely biased and inconsistent (fully parametric) estimator in case of misspecification. For this reason and in order to avoid, as much as possible, these problems, we consider here a semiparametric approach where the copula is modeled parametrically but the marginal distributions are modeled nonparametrically. As it is shown in the next sections, the proposed method has many interesting properties both from theoretical and practical point of view. Especially, the asymptotic properties are easy to obtain, the numerical calculations can be done directly using existing packages and, unlike many semiparametric methods, no iteration procedure is needed to guarantee the consistency. Also, the asymptotic variance can be estimated without any extra complications.

The plan of the paper is as follows. Section 2 presents the general theoretical framework of the method with the necessary notations and assumptions. In Section 3, we establish the asymptotic representation of the proposed estimator in the univariate and multivariate case. From this representation we establish the asymptotic distribution and the asymptotic variance of the estimator. In Section 4, we study theoretical properties of the estimator under misspecification. In Section 5 we provide a simulation exercise to evaluate the performance and investigate the finite sample properties (under correct and misspecified copula model). Finally, we analyze data from air pollution studies to illustrate the usefulness of the proposed estimator in Section 6. Proofs appear in the Appendix.

2 Theoretical Background

Let $(Y_i, \boldsymbol{X}_i^\top)$, $i = 1, \ldots, n$, be an independent and identically distributed (i.i.d.) sample of n observations generated from the distribution of $(Y, \boldsymbol{X}^\top)$. For each i, let $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d})^\top$ and let f_0 (F_0) and f_j (F_j) be the density (c.d.f.) of Y_i and $X_{i,j}$, respectively. Clearly, the shape and the performance of our estimator $\hat{m}$ in (5) will heavily depend on the methods of estimation for c, F_0 and F_j. In this work, F_0 is estimated empirically by

$$\hat{F}_0(y) = \frac{1}{n} \sum_{i=1}^{n} I(Y_i \leq y).$$

Estimating the other c.d.f.'s F_j, $j = 1, \ldots, d$, can also be done empirically via $\hat{F}_j$. However, this results in a nonsmooth estimate $\hat{m}(\boldsymbol{x})$ as it is illustrated in Figure 1, where we show the resulting estimator

using $\hat{F}_1$ in the univariate linear case. To get a more visually attractive regression curve, one should smooth the empirical c.d.f. The simple way to do that is to use a kernel smoothing method. Let $k(\cdot)$ be a function which is a symmetric probability density function and $h \equiv h_n \to 0$ be a bandwidth parameter. Then, a kernel smoothing estimator of F_j is given by

$$\tilde{F}_j(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - X_{i,j}}{h}\right),$$

where $K(x) = \int_{-\infty}^{x} k(t)dt$. The estimator $\tilde{F}_j$ is asymptotically equivalent to $\hat{F}_j$, in the sense that, if $nh^4 \to 0$, then $\tilde{F}_j$ satisfies the following assumption.

Assumption A:

$$\tilde{F}_j(x) = n^{-1}\sum_{i=1}^{n} I(X_{i,j} \le x) + o_p(n^{-1/2}) \quad \text{for } j = 1, \ldots, d.$$

Evidently, this assumption also holds for the empirical c.d.f. as well as for the rescaled empirical c.d.f. $(n/(n+1))\hat{F}_j$.

Before delving into the asymptotic analysis of $\hat{m}$, we run a small simulation study to examine the effect of the methods of estimating F_1 on the regression fit. We generate (Y, X_1) using FGM copula according to the data generating procedure DGP.S.b described in Section 5. Table 1 shows the empirical integrated mean squared errors (IMSE), see (11) below, together with the empirical integrated biases (IBIAS) and empirical integrated variances (IVAR) of $\hat{m}$ based on 1000 replications. We compare the performance of $\hat{m}$ using the three estimators of F_1: $\hat{F}_1$ the empirical c.d.f., $\tilde{F}_{opt}$ the kernel smoothing estimates with the mean square optimal bandwidth, i.e.

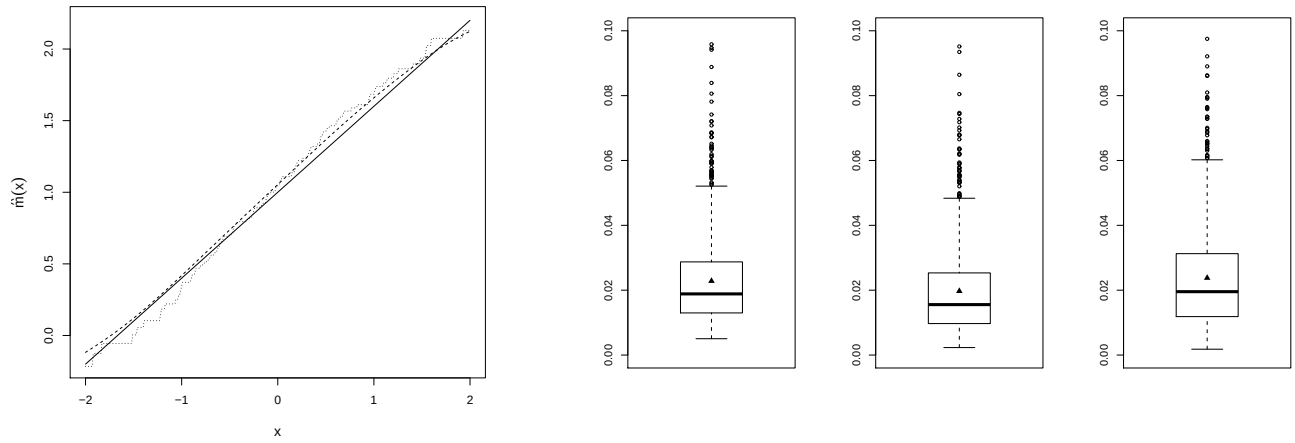
$$h_{opt}(x_1) = \left[\frac{2f_1(x_1)\int tk(t)K(t)dt}{(\int t^2k(t)dt)^2\{f_1'(x_1)\}^2}\right]^{1/3} n^{-1/3},$$

and $\tilde{F}_{cv}$ the kernel smoothing estimates with a bandwidth chosen via the cross-validation method. Compared to the empirical distribution estimate, we see that the kernel smoothing estimate gives better results if the optimal bandwidth is used. When the bandwidth is chosen by the data, its performance is similar to the non-smooth estimator $\hat{F}_1$. This latter leads to a less biased regression estimator but with slightly large variance. Figure 1, which shows the boxplots of the empirical integrated squared errors (ISE), also supports this observation. In this simulation the copula was estimated using maximum pseudo-likelihood method as described below.

Table 1: IBIAS, IVAR and IMSE ($\times 100$) of $\hat{m}$ depending on the method of estimating F_1

$\hat{F}_1$			$\tilde{F}_{opt}$			$\tilde{F}_{cv}$		
IBIAS	IVAR	IMSE	IBIAS	IVAR	IMSE	IBIAS	IVAR	IMSE
0.773	1.511	2.282	0.661	1.311	1.971	1.014	1.364	2.377

Figure 1: Estimated regression functions using $\hat{F}_1$ (dotted line) and $\tilde{F}_1$ with h_{opt} (dashed line); the black line is the true regression function. Boxplots of the empirical integrated squared errors of $\hat{m}$ with $\hat{F}_1$, $\tilde{F}_{opt}$ and $\tilde{F}_{cv}$, respectively (from left to right).



As mentioned in the introduction, in this work we adopt a semiparametric approach. We assume that c belongs to a given parametric family $\mathcal{C} = \{c(.;\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \Theta\}$, where Θ is a compact subset of $\mathbb{R}^p$. We denote by $\boldsymbol{\theta}_0$ the true (but unknown) copula parameter, thus the copula density of $(Y, \boldsymbol{X}^\top)^\top$, $c(.)$, coincides with $c(.;\theta_0)$. Let $\hat{\boldsymbol{\theta}}$ be any estimator of $\boldsymbol{\theta}_0$ satisfying the following assumption.

Assumption B:

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = n^{-1} \sum_{i=1}^{n} \boldsymbol{\eta}_i + o_p(n^{-1/2}),$$

where $\boldsymbol{\eta}_i = \boldsymbol{\eta}(F_0(Y_i), \boldsymbol{F}(\boldsymbol{X}_i); \boldsymbol{\theta}_0)$ is a p-dimensional random vector such that $\mathbb{E}\boldsymbol{\eta} = \mathbf{0}$ and $\mathbb{E}||\boldsymbol{\eta}||^2 < \infty$. $\boldsymbol{F}(\boldsymbol{X}_i) = (F_1(X_{i,1}), \ldots, F_d(X_{i,d}))$.

Many existing estimators for $\boldsymbol{\theta}_0$ in the literature satisfy this assumption. One of the most promising estimator among them is the (semiparametric) maximum pseudo-likelihood (PL) estimator $\hat{\boldsymbol{\theta}}_{PL}$, which

is defined as the maximizer of

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log c\left(\frac{n}{n+1}\hat{F}_0(Y_i), \frac{n}{n+1}\hat{\boldsymbol{F}}(\boldsymbol{X}_i); \boldsymbol{\theta}\right),$$

where $\hat{\boldsymbol{F}}(\boldsymbol{X}_i) = (\hat{F}_1(X_{i,1}), \ldots, \hat{F}_d(X_{i,d}))^\top$. $\hat{\boldsymbol{\theta}}_{PL}$ was studied by Genest et al. (1995), Silvapulle et al. (2004), Tsukahara (2005) and Kojadinovic and Yan (2011). Kojadinovic and Yan (2011) compared $\hat{\boldsymbol{\theta}}_{PL}$ with two method-of-moment estimators and found that the PL estimator performs best overall in terms of mean squared error. A similar conclusion was drawn in Silvapulle et al. (2004) regarding the comparison of $\hat{\boldsymbol{\theta}}_{PL}$ with two other estimators based on the maximum likelihood. The conditions under which $\hat{\boldsymbol{\theta}}_{PL}$ satisfies Assumption B can be found in Tsukahara (2005). For the PL estimator, we have,

$$\eta(u_0, \boldsymbol{u}; \boldsymbol{\theta}_0) = J^{-1}(\boldsymbol{\theta}_0) \times K(u_0, \boldsymbol{u}; \boldsymbol{\theta}_0), \tag{7}$$

where

$$J(\boldsymbol{\theta}_0) = \int_{[0,1]^{d+1}} \left(\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \log c(u_0, \boldsymbol{u}; \boldsymbol{\theta}_0)\right) dC(u_0, \boldsymbol{u}; \boldsymbol{\theta}_0)$$

and $K(u_0, \boldsymbol{u}; \boldsymbol{\theta}_0)$ is p-dimensional function vector whose k-th element is

$$\frac{\partial}{\partial\theta_k} \log c(u_0, \boldsymbol{u}; \boldsymbol{\theta}_0) + \sum_{j=0}^{d} \int_{[0,1]^{d+1}} (I(u_j \le v_j) - v_j) \left(\frac{\partial^2}{\partial\theta_k \partial u_j} \log c(u_0, \boldsymbol{u}; \boldsymbol{\theta}_0)\right) dC(v_0, \boldsymbol{v}; \boldsymbol{\theta}_0).$$

Before continuing, we need to introduce the following notations.

Notations:

- $c_j = \dfrac{\partial c}{\partial u_j}$, for $j = 0, \ldots, d$ and $c_{\boldsymbol{X},j} = \dfrac{\partial c_{\boldsymbol{X}}}{\partial u_j}$ and $e_j = \dfrac{\partial e}{\partial u_j}$, for $j = 1, \ldots, d$.
- $\boldsymbol{e}'(\boldsymbol{u}) = (e_1(\boldsymbol{u}), \ldots, e_d(\boldsymbol{u}))^\top$ and $\boldsymbol{c}'_{\boldsymbol{X}}(\boldsymbol{u}) = (c'_{\boldsymbol{X},1}(\boldsymbol{u}), \ldots, c'_{\boldsymbol{X},d}(\boldsymbol{u}))^\top$
- $\dot{\boldsymbol{c}} = \left(\dfrac{\partial c}{\partial\theta_1}, \ldots, \dfrac{\partial c}{\partial\theta_p}\right)^\top$, $\dot{\boldsymbol{c}}_{\boldsymbol{X}} = \left(\dfrac{\partial c_{\boldsymbol{X}}}{\partial\theta_1}, \ldots, \dfrac{\partial c_{\boldsymbol{X}}}{\partial\theta_p}\right)^\top$ and $\dot{\boldsymbol{e}} = \left(\dfrac{\partial e}{\partial\theta_1}, \ldots, \dfrac{\partial e}{\partial\theta_p}\right)^\top$, where $e(\boldsymbol{u}; \boldsymbol{\theta}) = \mathbb{E}(Y c(F_0(Y), \boldsymbol{u}; \boldsymbol{\theta}))$.

Additionally, when no confusion is possible, hereafter we will suppress $\boldsymbol{\theta}_0$ in all our notations. For example, instead of writing $c(\cdot, \boldsymbol{\theta}_0)$ and $e(\cdot, \boldsymbol{\theta}_0)$, we will simply write $c(\cdot)$ and $e(\cdot)$, respectively. Finally, to facilitate the asymptotic analysis, we need the following assumptions.

Assumption C:

(C1) (i) $\mathbb{E}|Y| < \infty$ and (ii) $yF_0(y) \to 0$ as $y \to -\infty$

(C2) $\dot{\boldsymbol{c}}$ and c_j, $j = 0, \ldots, d$, are continuous on $[0,1] \times \Pi_{j=1}^{d}[F_j(x_j) - \epsilon, F_j(x_j) + \epsilon]$ for some $\epsilon > 0$.

(C3) $\mathbb{E}[Y c_j(F_0(Y), \boldsymbol{F}(\boldsymbol{x}); \boldsymbol{\theta}_0)]^2 < \infty$, $j = 0, \ldots, d$, and $\mathbb{E}[Y \frac{\partial c}{\partial \theta_k}(F_0(Y), \boldsymbol{F}(\boldsymbol{x}); \boldsymbol{\theta}_0)]^2 < \infty$, $k = 1, \ldots, p$.

(C4) $\mathbb{E}[c_j(F_0(Y), \boldsymbol{F}(\boldsymbol{x}); \boldsymbol{\theta}_0)]^2 < \infty$, $j = 0, \ldots, d$, and $\mathbb{E}[\frac{\partial c}{\partial \theta_k}(F_0(Y), \boldsymbol{F}(\boldsymbol{x}); \boldsymbol{\theta}_0)]^2 < \infty$, $k = 1, \ldots, p$.

3 Main Results

Now we are ready to present the main results. We will show the asymptotic i.i.d. representation of the proposed estimator, which implies that the estimator follows a normal distribution asymptotically.

3.1 Single covariate case ($d = 1$)

First, consider the more simple case where there is only one covariate X_1. In this case,

$$m(x_1) = e(F_1(x_1); \boldsymbol{\theta}_0) = \mathbb{E}[Y c(F_0(Y), F_1(x_1); \boldsymbol{\theta}_0)],$$

can be estimated by

$$\hat{m}(x_1) = \hat{e}(\tilde{F}(x_1); \hat{\boldsymbol{\theta}}) := n^{-1} \sum_{i=1}^{n} Y_i c(\hat{F}_0(Y_i), \tilde{F}_1(x_1); \hat{\boldsymbol{\theta}}).$$

The following theorem gives an asymptotic i.i.d. representation of this estimator. Its proof is given in in the Appendix.

Theorem 3.1 *Under Assumptions (C1), (C2) and (C3), if $\tilde{F}_1$ satisfies Assumption A and $\hat{\boldsymbol{\theta}}$ satisfies Assumption B, then we have*

$$\begin{aligned}\hat{m}(x_1) - m(x_1) = n^{-1} \sum_{i=1}^{n} [&\zeta(F_1(X_{i,1}), F_1(x_1)) \times e_1(F_1(x_1)) - \gamma_0(F_0(Y_i), F_1(x_1)) \\ &+ \boldsymbol{\eta}_i^{\top} \times \dot{\boldsymbol{e}}(F_1(x_1))] + o_p(n^{-1/2}),\end{aligned}$$

where $\zeta(u, v) = I(u \le v) - v$ and $\gamma_0(u_0, \boldsymbol{u}) \equiv \gamma_{F_0}(u_0, \boldsymbol{u}; \boldsymbol{\theta}_0) = \int \zeta(u_0, F_0(y)) c(F_0(y), \boldsymbol{u}; \boldsymbol{\theta}_0) dy$.

Theorem 3.1 implies that $\sqrt{n}(\hat{m}(x_1) - m(x_1))$ follows asymptotically a normal distribution with mean 0 and variance $\sigma^2(x_1) = \mathbb{V}ar(E_i(x_1))$, with $E_i(x_1) = \zeta(F_1(X_{i,1}), F_1(x_1)) \times e_1(F_1(x_1)) - \gamma_0(F_0(Y_i), F_1(x_1)) + \boldsymbol{\eta}^{\top}(F_0(Y_i), F_1(X_{i,1})) \times \dot{\boldsymbol{e}}(F_1(x_1))$. By plug-in principle, a natural estimator of $\sigma^2(x_1)$ is given by $\hat{\sigma}^2(x_1) = n^{-1} \sum_{i=1}^{n} (\hat{E}_i(x_1) - n^{-1} \sum_{i=1}^{n} \hat{E}_i(x_1))^2$, where $\hat{E}_i(x_1)$ is the same as $E_i(x_1)$ but with $\hat{F}_0$, $\tilde{F}_1$ and $\hat{\boldsymbol{\theta}}$ instead of F_0, F_1 and $\boldsymbol{\theta}_0$, respectively. The validity (consistency) of this approach is investigated numerically in Section 5.

3.2 Multiple covariate case ($d \geq 2$)

In the general case ($d \geq 2$), the regression function is given by

$$m(\boldsymbol{x}) = \frac{e(\mathbf{F}(\boldsymbol{x}); \boldsymbol{\theta}_0)}{c_{\boldsymbol{X}}(\boldsymbol{F}(\boldsymbol{x}); \boldsymbol{\theta}_0)}. \tag{8}$$

Estimating the numerator of $m(\boldsymbol{x})$ can be done as in the single covariate case by $\hat{e}(\tilde{\boldsymbol{F}}(\boldsymbol{x})) := n^{-1} \sum_{i=1}^{n} Y_i c(\hat{F}_0(Y_i), \tilde{\boldsymbol{F}}(\boldsymbol{x}); \hat{\boldsymbol{\theta}})$, where $\tilde{\boldsymbol{F}}(\boldsymbol{x}) = (\tilde{F}_1(x_1), \ldots, \tilde{F}_d(x_d))$. Following the proof of Theorem 3.1, one can easily check that, under Assumptions A, B, (C1), (C2) and (C3),

$$\hat{e}(\tilde{\boldsymbol{F}}(\boldsymbol{x})) - e(\mathbf{F}(\mathbf{x})) = n^{-1} \sum_{i=1}^{n} E_i(\boldsymbol{F}(\boldsymbol{x}); \boldsymbol{\theta}_0) + o_p(n^{-1/2}), \tag{9}$$

where $E_i(\boldsymbol{u}; \boldsymbol{\theta}_0) \equiv E_{F_0(Y_i), \boldsymbol{F}(\boldsymbol{X}_i)}(\boldsymbol{u}; \boldsymbol{\theta}_0) = \boldsymbol{\zeta}^\top(\boldsymbol{F}(\boldsymbol{X}_i), \boldsymbol{u}) \times \boldsymbol{e}'(\boldsymbol{u}) - \gamma_0(F_0(Y_i), \boldsymbol{u}) + \boldsymbol{\eta}_i^\top \times \dot{\boldsymbol{e}}(\boldsymbol{u})$, with $\boldsymbol{\zeta}(\boldsymbol{v}, \boldsymbol{u}) = (\zeta(v_1, u_1), \ldots, \zeta(v_d, u_d))^\top$. From equation (6), a natural estimator of the denominator of (8) is $\int_0^1 c(u_0, \tilde{\boldsymbol{F}}(\boldsymbol{x}); \hat{\boldsymbol{\theta}}) du_0$. This is a "good" estimator if we are interested only in $c_{\boldsymbol{X}}(\boldsymbol{F}(\boldsymbol{x}))$. However, this is not our estimation target and we are interested in the ratio given by (8) instead. In such a situation, it is beneficial for reducing the estimation error of the ratio to have the estimation procedure of the denominator mimic the one of the numerator. Because of this reason, using the fact that $c_{\mathbf{X}}(\mathbf{u}) = \mathbb{E}[c(F_0(Y), \mathbf{u})]$, see (6), we propose to estimate $c_{\boldsymbol{X}}(\boldsymbol{F}(\boldsymbol{x}))$ by $\hat{c}_{\boldsymbol{X}}(\tilde{\boldsymbol{F}}(\boldsymbol{x})) = n^{-1} \sum_{i=1}^{n} c(\hat{F}_0(Y_i), \tilde{\boldsymbol{F}}(\mathbf{x}); \hat{\boldsymbol{\theta}})$. Thus, our estimator of $m(\boldsymbol{x})$ is given by

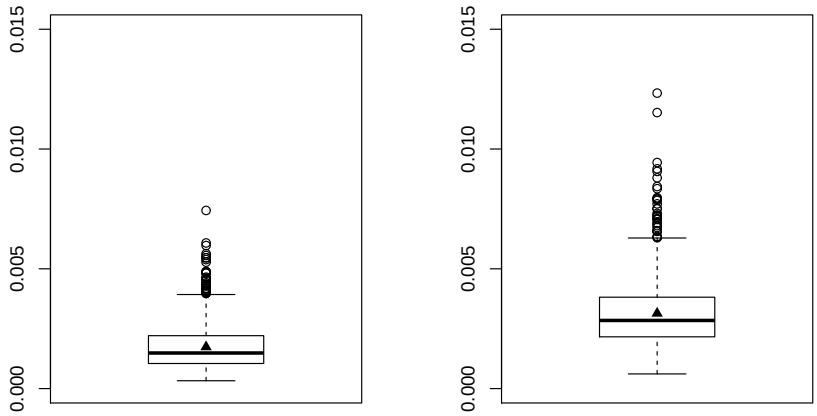
$$\hat{m}(\boldsymbol{x}) = \frac{\hat{e}(\tilde{\boldsymbol{F}}(\boldsymbol{x}))}{\hat{c}_{\boldsymbol{X}}(\tilde{\boldsymbol{F}}(\boldsymbol{x}))} = \sum_{i=1}^{n} Y_i \frac{c(\hat{F}_0(Y_i), \tilde{\boldsymbol{F}}(\boldsymbol{x}); \hat{\boldsymbol{\theta}})}{\sum_{i=1}^{n} c(\hat{F}_0(Y_i), \tilde{\boldsymbol{F}}(\boldsymbol{x}); \hat{\boldsymbol{\theta}})}.$$

Table 2 shows the results of a small Monte Carlo study designed to compare this estimator with the "naive" one, i.e. $\tilde{m}(\boldsymbol{x}) = \hat{e}(\tilde{\boldsymbol{F}}(\boldsymbol{x})) / \int_0^1 c(u_0, \tilde{\boldsymbol{F}}(\boldsymbol{x}); \hat{\boldsymbol{\theta}}) du_0$. Using 1000 random samples generated from DGP.M.a ($d = 2$) described in Section 5, we compute the empirical IMSE, IBIAS and IVAR for $\tilde{m}$ and $\hat{m}$. We see that this latter performs clearly better than the "naive" one, both in terms of bias and variance. This observation is also confirmed by Figure 2, which shows the boxplots of the empirical ISE's of the two estimators.

Table 2: IBIAS, IVAR and IMSE ($\times 1000$) of $\hat{m}$ and $\tilde{m}$

	IBIAS	IVAR	IMSE	IBIAS	IVAR	IMSE	
$\hat{m}$	0.172	1.567	1.737	1.178	1.963	3.139	$\tilde{m}$

Figure 2: Boxplots of the empirical ISE's of $\hat{m}$ (left) and $\tilde{m}$ (right). The triangular dots indicate the average.



The asymptotic representation of $\hat{c}_{\boldsymbol{X}}(\tilde{\boldsymbol{F}}(\boldsymbol{x}))$ follows by using similar arguments as in the proof of Theorem 3.1. In fact, under Assumptions A, B, (C2) and (C4), one can easily check that

$$\hat{c}_{\boldsymbol{X}}(\tilde{\boldsymbol{F}}(\boldsymbol{x})) - c_{\boldsymbol{X}}(\boldsymbol{F}(\boldsymbol{x})) = n^{-1}\sum_{i=1}^{n} C_i(\boldsymbol{F}(\boldsymbol{x});\boldsymbol{\theta}_0) + o_p(n^{-1/2}), \tag{10}$$

where $C_i(\boldsymbol{u};\boldsymbol{\theta}_0) \equiv C_{\boldsymbol{F}(\boldsymbol{X}_i)}(\boldsymbol{u};\boldsymbol{\theta}_0) = \boldsymbol{\zeta}^\top(\boldsymbol{F}(\boldsymbol{X}_i),\boldsymbol{u}) \times \boldsymbol{c}'_{\boldsymbol{X}}(\boldsymbol{u}) + \boldsymbol{\eta}_i^\top \times \dot{\boldsymbol{c}}_{\boldsymbol{X}}(\boldsymbol{u})$.

Remark This result also shows that, up to $o_p(n^{-1/2})$, $\hat{c}_{\boldsymbol{X}}(\tilde{\boldsymbol{F}}(\boldsymbol{x}))$ is asymptotically equivalent to $\int_0^1 c(u_0, \tilde{\boldsymbol{F}}(\boldsymbol{x});\hat{\boldsymbol{\theta}})du_0$. This means that, up to the first order approximation, the effect of introducing $\hat{F}_0$ in our estimating procedure is asymptotically negligible. However, as explained above, for finite sample size, this is beneficial to reduce the resulting error in the ratio estimation.

Finally, combining (9) with (10) leads to our main result.

Theorem 3.2 *Under Assumption C, if $\tilde{F}$ satisfies Assumption A and $\hat{\boldsymbol{\theta}}$ satisfies Assumption B, then we have*

$$\hat{m}(\boldsymbol{x}) - m(\boldsymbol{x}) = n^{-1}\sum_{i=1}^{n} \frac{1}{c_{\boldsymbol{X}}(\boldsymbol{F}(\boldsymbol{x}))}\left[E_i(\boldsymbol{F}(\boldsymbol{x})) - m(\boldsymbol{x})C_i(\boldsymbol{F}(\boldsymbol{x}))\right] + o_p(n^{-1/2}).$$

As in the univariate case, this result directly leads to the asymptotic normality of $\sqrt{n}(\hat{m}(\boldsymbol{x}) - m(\boldsymbol{x}))$. The asymptotic variance of $\sqrt{n}(\hat{m}(\boldsymbol{x}) - m(\boldsymbol{x}))$, which is given by

$$\mathbb{V}ar\left(\frac{1}{c_{\boldsymbol{X}}(\boldsymbol{F}(\boldsymbol{x}))}\left[E_1(\boldsymbol{F}(\boldsymbol{x})) - m(\boldsymbol{x})C_1(\boldsymbol{F}(\boldsymbol{x}))\right]\right),$$

can be estimated using the plug-in method. As a consequence, one can easily construct pointwise confidence intervals for m. The validity of this approach is investigated in Section 5.

4 Consideration of Misspecification

If the copula family is known, then, as it will be shown in the simulation study, the proposed estimator is highly accurate. However, in practice, the copula shape is unknown and needs typically to be selected by the data. Any selection procedure has the possibility that it may select the wrong copula family in practice. Like any parametric or semiparametric method, using a misspecified (copula) model will lead to an inconsistent estimator. In this section we are interested in the following question: what is the effect (cost) of using a misspecified copula model on the resulting regression estimator? Let $\mathcal{C} = \{c(.;\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \Theta\}$ be any parametric family of copula densities. In the previous sections we assumed that there exists the true parameter $\boldsymbol{\theta}_0$ such that $c(.;\boldsymbol{\theta}_0)$ coincides with the true copula density, $c(\cdot)$, of $(Y, \boldsymbol{X}^\top)^\top$. Under a possibly misspecified model, such $\boldsymbol{\theta}_0$ may not exist. Instead, we can define $\boldsymbol{\theta}^*$ to be the unique minimum within the set Θ of

$$I(\boldsymbol{\theta}) = \int_{[0,1]^{d+1}} \ln\left(\frac{c(u_0, \boldsymbol{u})}{c(u_0, \boldsymbol{u}; \boldsymbol{\theta})}\right) dC(u_0, \boldsymbol{u}).$$

This is the classical Kullback-Leibler information criterion expressed in terms of copulas densities instead of the traditional densities. From the proof of Theorem 3.2, we have the following theorem regarding the asymptotic behavior of $\hat{m}(\boldsymbol{x})$ when the copula family is misspecified.

Theorem 4.1 *Under Assumption C, with $\boldsymbol{\theta}^*$ instead of $\boldsymbol{\theta}_0$, if $\tilde{F}$ satisfies Assumption A and $\hat{\boldsymbol{\theta}}$ satisfies Assumption B, with $\boldsymbol{\theta}^*$ instead of $\boldsymbol{\theta}_0$, then*

$$\begin{aligned}\hat{m}(\boldsymbol{x}) - m(\boldsymbol{x}) =& m(\boldsymbol{x}; \boldsymbol{\theta}^*) - m(\boldsymbol{x}) \\ &+ n^{-1} \sum_{i=1}^{n} \frac{1}{c_{\boldsymbol{X}}(\boldsymbol{F}(\boldsymbol{x}); \boldsymbol{\theta}^*)} \left[E_i(\boldsymbol{F}(\boldsymbol{x}); \boldsymbol{\theta}^*) - m(\boldsymbol{x}; \boldsymbol{\theta}^*) C_i(\boldsymbol{F}(\boldsymbol{x}); \boldsymbol{\theta}^*)\right] \\ &+ o_p(n^{-1/2}),\end{aligned}$$

where $m(x, \boldsymbol{\theta}^*)$ is the mean regression function under the assumption that the joint distribution of $(Y, \boldsymbol{X}^\top)^\top$ is $C(F_0(Y), \boldsymbol{F}(\boldsymbol{X}); \boldsymbol{\theta}^*)$.

Clearly, when $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$, this theorem reduces to Theorem 3.2. This result also shows that a misspecified copula brings out a bias in the estimation of $m(\boldsymbol{x})$, which is asymptotically nothing but the difference between the true regression function and its best approximation (in terms of likelihood) among the regression function family $\{m(\boldsymbol{x}; \boldsymbol{\theta}) := \mathbb{E}(Y c(F_0(Y), \boldsymbol{F}(\boldsymbol{x}); \boldsymbol{\theta}))/c_{\boldsymbol{X}}(\boldsymbol{F}(\boldsymbol{x}); \boldsymbol{\theta}),\ \boldsymbol{\theta} \in \Theta\}$.

Remark Let

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \log c\left(\frac{n}{n+1}\hat{F}_0(Y_i), \frac{n}{n+1}\hat{\boldsymbol{F}}(\boldsymbol{X}_i); \boldsymbol{\theta}\right)$$

be the maximum pseudo-likelihood estimator. By the classical maximum likelihood theory under misspecification, see White (1982), and following the proof of Theorem 1 in Tsukahara (2005), we have verified that $\hat{\boldsymbol{\theta}}$ satisfies Assumption B with $\boldsymbol{\eta}$ as given by (7) but with $\boldsymbol{\theta}^*$ instead of $\boldsymbol{\theta}_0$.

5 Simulations

The objective of this section is first to check whether the asymptotic theory for $\hat{m}(\boldsymbol{x})$ works both when the copula is well-specified (Theorem 3.2) and when the copula is misspecified (Theorem 4.1). The second objective is to compare our semiparametric estimator with some competitors both when the true copula family is known and when the copula shape is adaptively selected using the data. To this ends, we consider the following data generating procedures (DGP):

- **DGP S.a** $(F_0(Y), F_1(X_1)) \sim$ Gaussian copula with parameter $\rho_1 = corr(Y, X_1)$; $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$.
  - The resulting regression function is $m(x_1) = \mu_Y + \sigma_Y \rho_1 \Phi^{-1}(F_1(x_1))$, where Φ is the c.d.f. of a standard normal distribution.
  - X_1 is generated from $\mathcal{N}(\mu_{X_1}, \sigma_{X_1}^2)$.
- **DGP S.b** $(F_0(Y), F_1(X_1)) \sim$ FGM copula with parameter θ; $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$.
  - The resulting regression function is $m(x_1) = \left(\mu_Y - \frac{\theta}{\sqrt{\pi}}\sigma_Y\right) + 2\frac{\theta}{\sqrt{\pi}}\sigma_Y F_1(x_1)$.
  - X_1 is generated from the c.d.f. $F_{X_1}(x_1) = 1 - \exp(-\exp(x_1))$.
- **DGP S.c** $(F_0(Y), F_1(X_1)) \sim$ Student t copula with parameters ρ and df; $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$.
  - The resulting regression function is

    $$m(x_1) = \mathbb{E}\left\{\sigma_Y \Phi^{-1}(\Phi_{df}(\rho a + \sqrt{df(1-\rho^2)(1+a^2/df)/(df+1)}T)) + \mu_Y\right\},$$

    where $a = \Phi_{df}^{-1}(F_{X_1}(x_1))$. Φ_{df} is the c.d.f. of a univariate Student t distribution with degrees of freedom df and T is a univariate Student t random variable with degrees of freedom $df + 1$.
  - X_1 is generated from $\mathcal{N}(\mu_{X_1}, \sigma_{X_1}^2)$.

- **DGP M.a** $(F_0(Y), F_1(X_1), \ldots, F_d(X_d)) \sim$ Gaussian copula with correlation matrix $\Sigma = \begin{bmatrix} 1 & \boldsymbol{\rho}^\top \\ \boldsymbol{\rho} & \Sigma_{\boldsymbol{X}} \end{bmatrix}$ where $\boldsymbol{\rho} = (corr(Y, X_1), \ldots, corr(Y, X_d))^\top$ and $\Sigma_{\boldsymbol{X}}$ is the $d \times d$ correlation matrix of $\boldsymbol{X}$; $Y \sim \mathcal{U}(0,1)$.
  - The resulting regression function is $m(\boldsymbol{x}) = \Phi\left(\sum_{j=1}^{d} \frac{a_j}{\sqrt{1+(1-\boldsymbol{\rho}^\top \boldsymbol{a})^2}} \Phi^{-1}(F_j(x_j))\right)$, where $\boldsymbol{a} = (a_1, \ldots, a_d)^\top := \Sigma_{\boldsymbol{X}}^{-1} \boldsymbol{\rho}$.
  - X_j is generated from $\mathcal{N}(\mu_{X_j}, \sigma^2_{X_j})$, $j = 1, \ldots, d$.
- **DGP M.b** $Y = \Psi(-0.3X_1 + 0.9X_2 + 0.3X_3) + \sigma\epsilon$, where Ψ is the c.d.f. of the standard Cauchy distribution.
  - $\boldsymbol{X} = (X_1, X_2, X_3)^\top$ is multivariate normal with mean $\mathbf{0}$ and $\mathbb{C}ov(X_i, X_j) = 0.5^{|i-j|}$.
  - $\epsilon \sim \mathcal{N}(0,1)$ independent of $\boldsymbol{X}$.

The parameters of the copula and of the marginal distributions that we use for each DGP are given in Table 3.

Table 3: Parameters of the copula and of the marginal distributions for each DGP.

DGP	copula parameter	marginal parameter	m
S.a	$\rho_1 = 0.6$	$\mu_{X_1} = 0, \mu_Y = 1$ $\sigma_{X_1} = \sigma_Y = 1$	$1 + 0.6x_1$
S.b	$\theta = 0.8$	$\mu_Y = 0, \sigma_Y = 1$	$0.8/\sqrt{\pi} - 1.6/\sqrt{\pi}\exp(-\exp(x_1))$
S.c	$\rho = 0.6$, $df = 3, 5, 100$	$\mu_{X_1} = 0, \mu_Y = 1$ $\sigma_{X_1} = \sigma_Y = 1$	no simple form
M.a ($d = 2$)	$\Sigma = \begin{pmatrix} 1 & -0.5 & 0.9 \\ -0.5 & 1 & -0.4 \\ 0.9 & -0.4 & 1 \end{pmatrix}$	$\mu_{X_1} = \mu_{X_2} = 0$ $\sigma_{X_1} = \sigma_{X_2} = 1$	$\Phi(-0.154x_1 + 0.771x_2)$
M.a ($d = 3$)	$\Sigma = \begin{pmatrix} 1 & 0.23 & 0.90 & 0.67 \\ 0.23 & 1 & 0.51 & 0.26 \\ 0.90 & 0.51 & 1 & 0.49 \\ 0.67 & 0.26 & 0.49 & 1 \end{pmatrix}$	$\mu_{X_1} = \mu_{X_2} = \mu_{X_3} = 0$, $\sigma_{X_1} = \sigma_{X_2} = \sigma_{X_3} = 1$	$\Phi(-0.3x_1 + 0.9x_2 + 0.3x_3)$

5.1 Verifying asymptotic results

In order to verify the asymptotic results in Section 2 and 3, we draw Quantile-Quantile (Q-Q) plots of $\hat{m}(\boldsymbol{x})$ and calculate empirical coverage probabilities (ECP) of the $(1-\alpha)$-confidence intervals of $m(\boldsymbol{x})$ according to Theorem 3.2 and 4.1. ECP means the proportion of confidence intervals that contain the true value of the regression function $m(\boldsymbol{x})$. By seeing whether the ECP's are close to the nominal level $(1-\alpha)$, we can check that the proposed estimator $\hat{m}(\boldsymbol{x})$ is asymptotically normal. Also, we can validate the asymptotic i.i.d. representation for $\hat{m}(\boldsymbol{x})$ and the appropriateness of the proposed variance estimator of $\hat{m}(\boldsymbol{x})$. We calculate the estimator $\hat{m}(\boldsymbol{x})$ from $N = 1000$ random samples of size $n = 50, 100, 200$ and 400. This is done for different values of significance level α and points of interest $\boldsymbol{x}$. Only selected and representative results are shown to save space and avoid repetition.

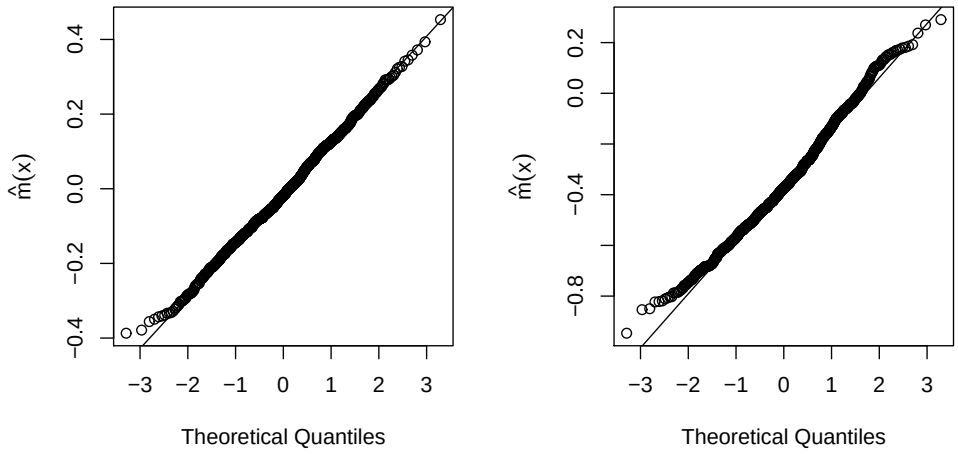
When the copula is well-specified

Table 4 presents the ECP of the proposed 95% confidence intervals for $m(\boldsymbol{x})$ with data generated from DGP S.a, DGP S.b and DGP M.a ($d = 2$). We calculate ECP when $\boldsymbol{x}$ is an interior point (median or mean) and a boundary point. As can be seen in Table 4, the ECP's for all the settings lie mostly in the range of 0.93 and 0.96. The only significant exception is the ECP (0.907) obtained with DGP S.a at the boundary region for $n = 50$. However, even in this case, the ECP increases with sample size to reach the nominal confidence level 0.95 when $n = 400$. On the whole, even for small sample size, the ECP's are quite close to their nominal confidence level, which demonstrates the validity of our i.i.d. representation and the consistency of our asymptotic variance estimator. Figure 3, which shows the QQ-plots for our estimator with DGP S.b when $n = 50$, also confirms this finding. Those plots clearly indicate the accuracy of the normal approximation to the asymptotic distribution of $\hat{m}(\boldsymbol{x})$. Note that this results are obtained under the ideal situation when the copula family is correctly specified. To estimate the copula parameter we make use of the R package *copula* (see Yan (2007)).

Table 4: Coverage probabilities of the proposed confidence interval for $m(x)$, $\alpha = 0.05$.

	DGP	$n = 50$	$n = 100$	$n = 200$	$n = 400$
Interior	S.a	0.943	0.938	0.939	0.953
	S.b	0.953	0.952	0.951	0.949
	M.a ($d = 2$)	0.946	0.943	0.937	0.943
Boundary	S.a	0.907	0.933	0.956	0.950
	S.b	0.968	0.957	0.962	0.955
	M.a ($d = 2$)	0.937	0.963	0.965	0.933

Figure 3: Q-Q plots of $\hat{m}(x)$ at an interior point (Left) and at a boundary point (Right). $n = 50$, DGP S.b.



When the copula is misspecified

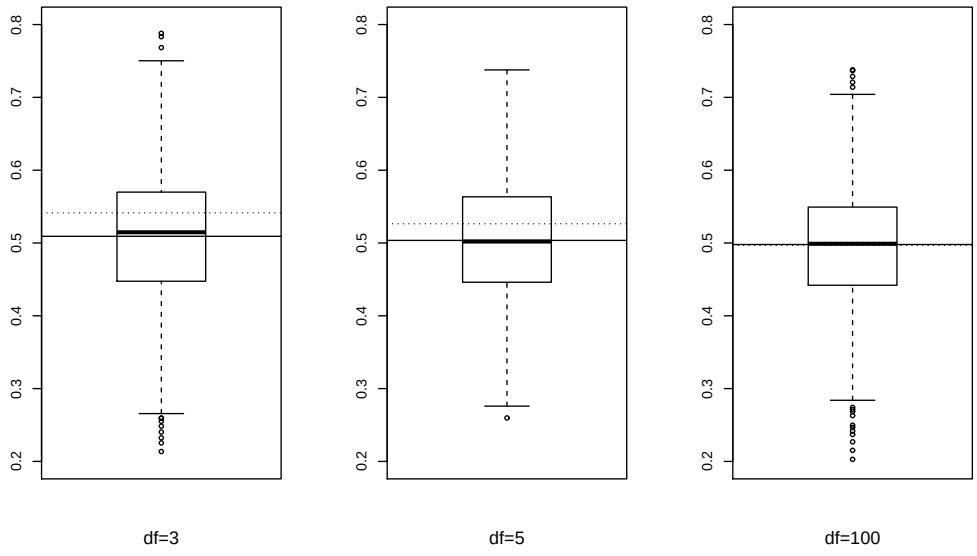
To verify the asymptotic behavior of our estimator under misspecification, we generate data from Student t copula according to DGP S.c but, in our estimation procedure, we use Gaussian copula. To see how misspecification influences the resulting regression function estimator, we vary the Student degrees of freedom df in $\{3, 5, 100\}$. The 'pseudo'-true regression function is $m(x_1, \rho^*) = 1 + \rho^* x_1$ with $\rho^* = 0.5831, 0.5901, 0.5963$ for $df = 3, 5, 100$, respectively. Table 5 shows the ECP's of confidence intervals for $m(x_1, \rho^*)$ at $x_1 = F_1^{-1}(0.2) = -0.8416$, calculated by the asymptotic representation in Theorem 4.1. From Table 5, we see that the ECP's are close to the nominal confidence level regardless of the degree of freedom. Figure 4 shows the boxplots of the estimators $\hat{m}(x_1)$ obtained from 1000

random samples of size 200. We see that the estimator concentrates around $m(x_1, \rho^*) \neq m(x_1)$. As the *df* increases, the difference between $m(x_1, \rho^*)$ and $m(x_1)$ decreases and our estimator becomes consistent. This confirms the asymptotic theory.

Table 5: Coverage probabilities of the proposed confidence interval for $m(x_1; \rho^*)$, $\alpha = 0.05$.

df	$n = 50$	$n = 100$	$n = 200$
3	0.941	0.944	0.945
5	0.943	0.941	0.953
100	0.942	0.936	0.945

Figure 4: Boxplots of $\hat{m}(x_1)$ at $x_1 = F_1^{-1}(0.2) = -0.8416$ for different degree of freedom. The horizontal solid line represents $m(x_1, \rho^*)$ and the dotted line represents $m(x_1)$.



5.2 Robustness and Comparison with other methods

In this subsection we compare our semiparametric estimator both with a parametric competitor and with a nonparametric competitor. We consider four kinds of estimators for comparison.

- $\hat{m}_{tc}$: our estimator when the true copula family is used.
- $\hat{m}_{uc}$: our estimator when the copula density family is selected by the data. The detail about this is given below.

- $\hat{m}_{ls}$: parametric regression estimator based on the classical least square method.
- $\hat{m}_{ll}$: nonparametric regression estimator (local linear).

As a comparison criterion we calculate the empirical Integrated Mean Squared Error (IMSE) given by

$$\begin{aligned} \text{IMSE} &= \frac{1}{N}\sum_{j=1}^{N} \text{ISE}(\hat{m}^{(j)}) := \frac{1}{N}\sum_{j=1}^{N}\left[\frac{1}{I}\sum_{i=1}^{I}\left(\hat{m}^{(j)}(\boldsymbol{x}_i) - m(\boldsymbol{x}_i)\right)^2\right] \\ &= \frac{1}{I}\sum_{i=1}^{I}\left(m(\boldsymbol{x}_i) - \bar{\hat{m}}(\boldsymbol{x}_i)\right)^2 + \frac{1}{I}\sum_{i=1}^{I}\left[\frac{1}{N}\sum_{j=1}^{N}(\hat{m}^{(j)}(\boldsymbol{x}_i) - \bar{\hat{m}}(\boldsymbol{x}_i))^2\right] \\ &\equiv \text{IBIAS}^2 + \text{IVAR}. \end{aligned} \tag{11}$$

where $\{x_i,\ i = 1, \ldots, I\}$ is the fixed evaluation set which corresponds to a random sample of size $I = 500$ from the distribution of $\boldsymbol{X}$, $\hat{m}^{(j)}(\cdot)$ is the estimated regression function from the j-th data sample and $\bar{\hat{m}}(x_i) = N^{-1}\sum_{j=1}^{N}\hat{m}^{(j)}(x_i)$.

Single covariate case

In order to compute the estimator $\hat{m}_{uc}$, we should decide which copula family to use and then estimate its parameters. In our simulations, we use AIC criterion to select one bivariate copula family among ten candidates: two are elliptical (Gaussian and Student t) and eight are Archimedean (Clayton, Gumbel, Frank, Joe, Clayton-Gumbel, Joe-Gumbel, Joe-Clayton and Joe-Frank). See, e.g., Brechmann and Schepsmeier (2011) for the definition of all these copulas.

The data was generated from FGM copula according to DGP S.b. The regression function is given by $m(x_1) = 0.8/\sqrt{\pi} - 1.6/\sqrt{\pi}\exp(-\exp(x_1))$. To calculate $\hat{m}_{tc}$ we use the FGM copula. Note that the true copula is not included in the list of ten candidate copulas families cited above and so the misspecified copula based estimator $\hat{m}_{uc}$ is expected to behave badly compared to $\hat{m}_{tc}$. As a parametric regression model we use the true regression function $a\exp(-\exp(bx_1)) + c$. To calculate $\hat{m}_{ls}$, we estimate a, b and c by the non-linear least squared method using the R package *nlrwr*. For details, we refer to Ritz and Streibig (2008). We also make use of the R package *np* to calculate $\hat{m}_{ll}$; see Hayfield and Racine (2008). The bandwidth parameter is selected via the cross-validation method.

The obtained results (see Table 6) of this study are better than what we expected. In fact, in terms of mean squared error, our estimator beats not only the local linear estimator but also the least squared estimator even when the copula distribution is unknown and selected (incorrectly) by the data. There are two reasons that may explain such a result. The first is that the classical least

square estimator suffers here from the fact that the error variance is quite large and varies with x_1 (see figure 5). It seems that our proposed method is not affected by this problem. The second is the fact that two completely different copula distributions can lead to the same or very similar regression models. For example, in the FGM copula regression model (see $m(x_1)$ in DGP S.b), if $X_1 \sim \mathcal{U}(0,1)$ then $m(x_1)$ becomes linear in x_1 as in the Gaussian regression model (see $m(x_1)$ in DGP S.a). In Table 7 we provide the IBIAS's, the IVAR's and the IMSE's of the four estimators. We observe that the variance is, by far, the dominant component of the mean squared error for all the estimators. The copula based estimators ($\hat{m}_{tc}$ and $\hat{m}_{uc}$) have less variation compared to their competitors. The fact that our estimator is more precise and more stable can also be seen in Figure 6 that shows the boxplots of the ISE's.

Table 6: $100 \times IMSE$ for $\hat{m}_{ls}$ (least square), $\hat{m}_{tc}$ (true copula), $\hat{m}_{uc}$ (unknown copula) and $\hat{m}_{ll}$ (local linear).

DGP	n	$\hat{m}_{ls}$	$\hat{m}_{tc}$	$\hat{m}_{uc}$	$\hat{m}_{ll}$
	50	4.841	3.007	4.230	15.04
S.b	100	2.517	1.599	2.178	7.500
	200	1.452	0.869	1.257	5.282

Figure 5: Scatter plot of a random sample of size 100 generated from DGP S.b (Left panel) and DGP M.a ($d = 2$) (Right panel). The solid line represents the true regression curve.

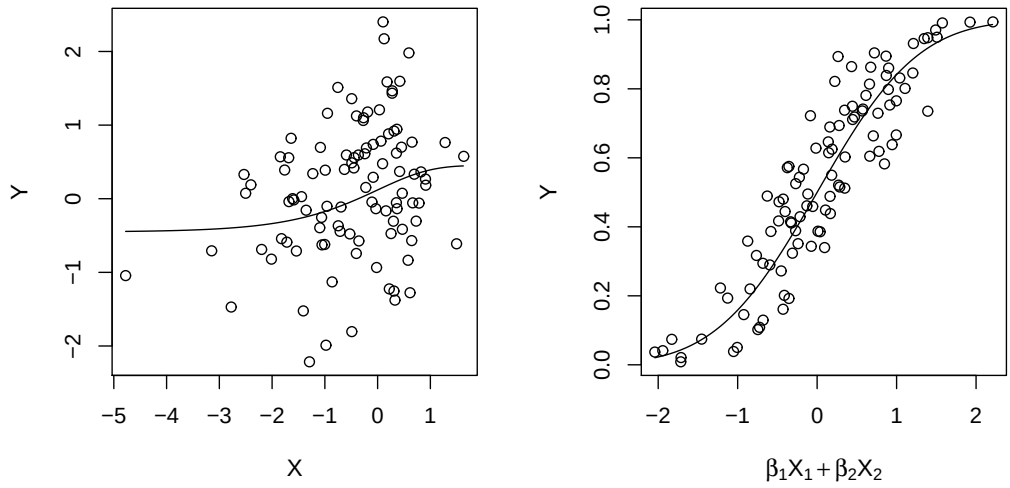
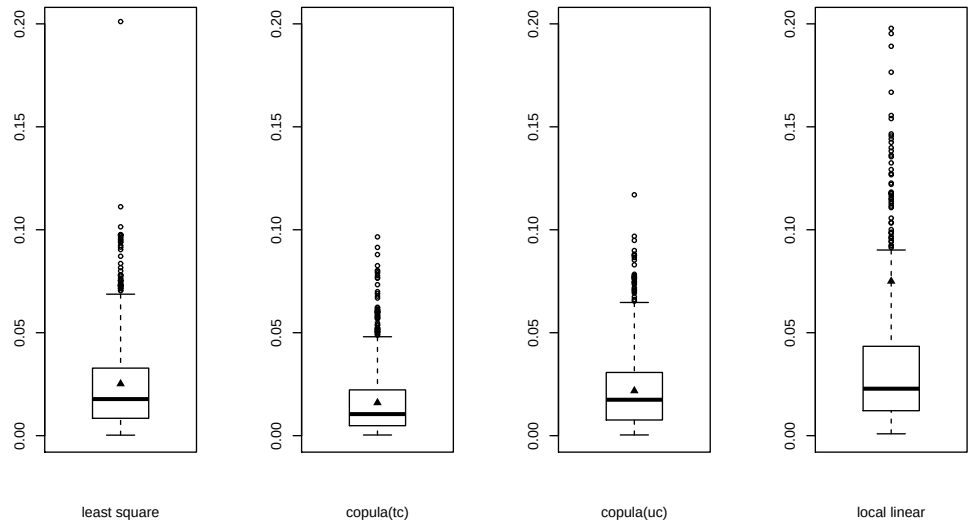


Table 7: IBIAS, IVAR and IMSE ($\times 100$) of the four estimators. $n = 100$.

	$\hat{m}_{ls}$	$\hat{m}_{tc}$	$\hat{m}_{uc}$	$\hat{m}_{ll}$
IBIAS	0.079	0.053	0.088	0.186
IVAR	2.440	1.546	2.092	7.320
IMSE	2.517	1.599	2.178	7.500

Figure 6: Boxplots of ISE's for four estimators. The triangular dots represent the average. $n = 100$.



Multiple covariate case

Different from the one covariate case, when the number of the covariates is large, it may be difficult to choose an appropriate copula family and its corresponding parameters in practice. Moreover, the set of high-dimensional copulas available in the literature is limited to very special and restrictive copula families such as elliptical copulas and Archimedean copulas. For this reason, the strategy that we advocate and adopt here is to make use of the recent available work about pair-copula decomposition. The main idea is to decompose a multivariate copula to a cascade of bivariate copulas so that we can take advantage of the relative simplicity of bivariate copula selection and estimation. To be more specific, we briefly describe such an approach for the case of three-variate vector $\boldsymbol{X} = (X_1, X_2, X_3)^T$.

By applying Sklar's theorem recursively one can write (for example)

$$\begin{aligned} c(F_0(y), F_1(x_1), F_2(x_2), F_3(x_3)) =& c_{\mathbf{X}}(F_1(x_1), F_2(x_2), F_3(x_3)) \times c_{01}(F_0(y), F_1(x_1)) \times \\ & c_{02|1}(F_{0|1}(y|x_1), F_{2|1}(x_2|x_1)|x_1) \times \\ & c_{03|12}(F_{0|12}(y|x_1, x_2), F_{3|12}(x_3|x_1, x_2)|x_1, x_2), \end{aligned} \tag{12}$$

where c_{01}, $c_{02|1}$ and $c_{03|12}$ are the copula densities associated with the distributions of (Y, X_1), $(Y, X_2)|X_1$ and $(Y, X_3)|(X_1, X_2)$, respectively. Similarly, $c_{\boldsymbol{X}}$ can be, for example, decomposed as $c_{\boldsymbol{X}}(F_1(x_1), F_2(x_2), F_3(x_3)) = c_{12}(F_1(x_1), F_2(x_2)) \times c_{23}(F_2(x_2), F_3(x_3)) \times c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2)$. If we assume that all the conditional copulas depend on the conditioning variables only through the conditional distributions, e.g. $c_{02|1}(F_{0|1}(y|x_1), F_{2|1}(x_2|x_1)|x_1) = c_{02|1}(F_{0|1}(y|x_1), F_{2|1}(x_2|x_1))$, then it leads to the so-called simplified pair-copula decomposition. Because any bivariate copula family could be used as a building block for this decomposition, the simplified pair-copula decomposition provides high flexibility and the ability to cover a wide range of complex dependencies. In our simulation, we consider all the possible pair-copula decompositions with ten candidate bivariate copulas and choose one decomposition (vine structure) which maximizes the AIC criterion. Hobæk Haff et al. (2010) discussed the conditions under which such a simplification is possible and found that this is not a severe restriction in many situations. For more about vines see the recent book by Kurowicka and Joe (2010). The problem of selecting an appropriate simplified decomposition and an appropriate parametric shape of each pair-copula and the estimation of the copula parameters are discussed in, e.g., Aas et al. (2009), Kurowicka and Joe (2010), Hobæk Haff (2012) and the references given there.

Remark Observe that the equality (12) holds without any restrictions in the copula c. As a consequence, by (1), the regression function can also be express as

$$\begin{aligned} m(x_1, x_2, x_3) =& \mathbb{E}(Y \times c_{01}(F_0(Y), F_1(x_1)) \times c_{02|1}(F_{0|1}(Y|x_1), F_{2|1}(x_2|x_1)|x_1) \times \\ & c_{03|12}(F_{0|12}(Y|x_1, x_2), F_{3|12}(x_3|x_1, x_2)|x_1, x_2)). \end{aligned}$$

Assuming that the pair simplifications holds, then one can use this equality to define a new class of estimation method. This approach will not be investigated in the current work.

We generate data according to DGP M.a ($d = 2$) and M.a ($d = 3$). Figure 5 shows the scatter plot of Y versus $-0.154X_1 + 0.771X_2$ using one random sample generated from DGP M.a ($d = 2$). To calculate $\hat{m}_{uc}$, the pair-copulas in each candidate decomposition are also selected using the AIC

criterion as in the one covariate case. We estimate the copula parameters by the maximum pseudo-likelihood method. All the computations are done via the R package *CDVine* (see Brechmann and Schepsmeier (2011)). As a parametric competitor we consider the nonlinear least square estimators from the model $m(x_1, x_2) = \Phi(\beta_1 x_1 + \beta_2 x_2)$ when $d = 2$ and $m(x_1, x_2, x_3) = \Phi(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)$ when $d = 3$. Table 8 presents the IMSE's for each estimator. As expected, for small ($n = 50$) and moderate ($n = 100$) sample sizes the least square estimator $\hat{m}_{ls}$ performs significantly better than our semiparametric copula estimators. However, the difference (in IMSE) between $\hat{m}_{ls}$ and our copula estimators decreases as the sample size increases. Because additional variability comes into the estimation from the selection of copula family for each bivariate copula, $\hat{m}_{uc}$ has larger IMSE than $\hat{m}_{tc}$ but the difference is relatively moderate. This is because the Gaussian copula admits D-vine decompositions with Gaussian pair-copulas; see Hobæk Haff et al. (2010). We observe that the estimator $\hat{m}_{uc}$ does better than the local linear estimator on the whole and especially at boundary regions (the details are not shown here). The same remarks remain valid in three covariate case. In this case, it is important to note that the difference between $\hat{m}_{tc}$ and $\hat{m}_{uc}$ becomes bigger than in two covariate case especially when the sample size is small. This is because the number of bivariate copula to be selected and estimated by the data in the decomposition increases with the number of the covariates (six instead of three in our case). However, when $d = 3$, $\hat{m}_{uc}$ still remains significantly better than the local linear estimator, which is known to suffer from the curse of dimensionality.

Table 8: $100 \times IMSE$ for $\hat{m}_{ls}$ (least square), $\hat{m}_{tc}$ (true copula), $\hat{m}_{uc}$ (unknown copula) and $\hat{m}_{ll}$ (local linear).

DGP	n	$\hat{m}_{ls}$	$\hat{m}_{tc}$	$\hat{m}_{uc}$	$\hat{m}_{ll}$
M.a ($d = 2$)	50	0.062	0.176	0.210	0.728
	100	0.032	0.079	0.097	0.215
	200	0.015	0.039	0.047	0.092
M.a ($d = 3$)	50	0.030	0.195	0.252	0.588
	100	0.015	0.095	0.116	0.184
	200	0.007	0.047	0.055	0.079

The case where data comes from a regression model

So far, we considered the case where the data comes from one of the well-known copula families. In this subsection, we investigate the case where data comes from a single index model according to DGP M.b. In this case, we have no information about the true copula and to our knowledge, the copula density of $(F_0(Y), F_1(X_1), F_2(X_2), F_3(X_3))$ cannot fit into any pair-copula decomposition with our candidate list of bivariate copulas. To calculate $\hat{m}_{uc}$ we follow the same procedure as described in the previous subsection. In this setting, our estimator $\hat{m}_{uc}$ is likely to be affected by the misspecification in the dependence structure induced by the pair-copula decomposition and the bivariate copula selection. We compare the estimator $\hat{m}_{uc}$ with $\hat{m}_{ls}$ the classical (nonlinear) least square estimator of the parametric model $\Psi(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)$, $\hat{m}_{ll}$ the local linear estimator and $\hat{m}_{si}$ the (semiparametric) single index estimator. The results in Table 9 show that our estimator has a very nice performance compared to the other ones in terms of IMSE. It ranks the second next to the least square estimator. Surprisingly, the single index estimator does worse than the local linear estimator. Increasing the error variance makes the MSE's larger for all the estimators. Our estimator seems to be more affected by the increase in the error variance than the parametric one.

Table 9: $100 \times IMSE$ for $\hat{m}_{ls}$ (least square), $\hat{m}_{uc}$ (unknown copula), $\hat{m}_{si}$ (single index) and $\hat{m}_{ll}$ (local linear).

σ	n	$\hat{m}_{ls}$	$\hat{m}_{uc}$	$\hat{m}_{si}$	$\hat{m}_{ll}$
0.1	50	0.066	0.332	2.172	0.857
	100	0.032	0.181	1.190	0.308
	200	0.015	0.111	0.662	0.151
0.2	50	0.271	0.851	3.312	1.667
	100	0.129	0.460	1.893	0.727
	200	0.061	0.273	1.056	0.407

6 Real Data Analysis

To illustrate our method, we analyze data from air pollution studies. The data consists of measurements of daily ozone concentration (Y = log(ozone)), solar radiation (X_1 = rad), daily maximum temperature (X_2 = temp) and wind speed(X_3 = wind) on $n = 111$ days from May to September 1973

in New York. To estimate $m(X_1, X_2, X_3) = \mathbb{E}(Y \,|\, X_1, X_2, X_3)$, we consider the following:

(LL) Local linear estimator with the bandwidth selected by cross-validation,

(LS) Least square estimator $\hat{\beta}_0 + \hat{\beta}_1 \text{rad} + \hat{\beta}_2 \text{temp} + \hat{\beta}_3 \text{wind} + \hat{\beta}_4 \text{wind}^2$,

(AD) Additive model estimator $\hat{g}_1(\text{rad}) + \hat{g}_2(\text{temp}) + \hat{g}_3(\text{wind})$,

(SI) Single index model estimator $\hat{g}(\hat{\beta}_1 \text{rad} + \hat{\beta}_2 \text{temp} + \hat{\beta}_3 \text{wind})$,

(CO) Our copula-based estimator with the copula selected by the data using pair-copula decomposition.

Model (LS) was considered by Crawley (2005), who found that the quadratic model fit the data well. (AD),(SI) and (CO) are semiparametric models. We use a smoothing spline to fit the additive model (AD) using R package *gam* and a local linear estimator to fit the single index model (SI) using R package *np*. Finally, as a evaluation measure of each estimator, we compute the following two versions of cross-validation error:

$$\text{CV}_1(\hat{m}) = \text{median}_{1 \le i \le n} |Y_i - \hat{m}_{-i}(\boldsymbol{X}_i)| \quad \text{and} \quad \text{CV}_2(\hat{m}) = \sum_{i=1}^{n} (Y_i - \hat{m}_{-i}(\boldsymbol{X}_i))^2 \tag{13}$$

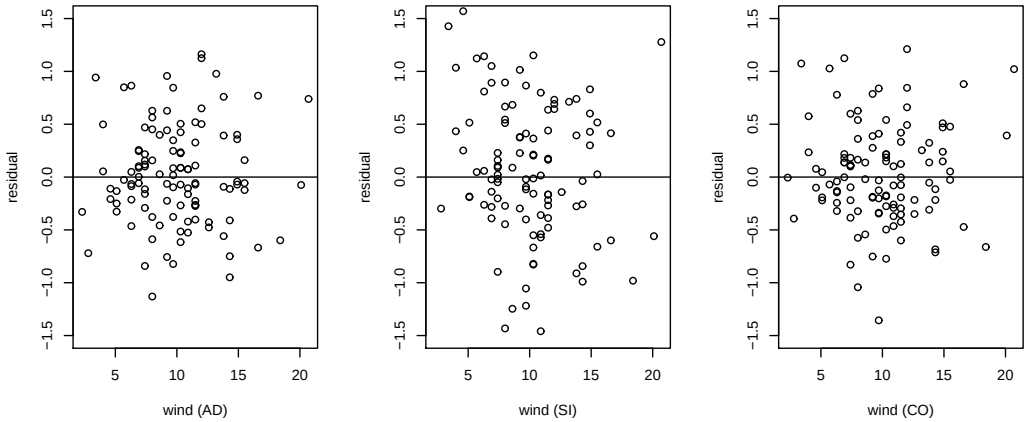
where $\hat{m}_{-i}(\boldsymbol{X}_i)$ is the estimate of $m(\boldsymbol{X}_i)$ from the dataset $\{(Y_j, \boldsymbol{X}_j);\ j \neq i,\ j = 1, \ldots, n\}$. CV_1 is a robustified version of the standard cross-validation error CV_2.

Table 10 suggests that all the estimators except (SI) show more or less similar performances. Clearly the single index structure is not appropriate for this data and the additive structure seems to be more adequate. In terms of the CV criterions, our estimator shows very good performance. The residual plots between X_{3i} (wind) and $Y_i - \hat{m}_{-i}(\boldsymbol{X}_i)$ for (AD),(SI) and (CO), see Figure 7, also supports this remarks. We observe that the residuals from (S2) show decreasing trend as X_3 (wind) increases while the residuals from (S1) and (S3) show random pattern. This example clearly shows the flexibility of our estimator and its ability to adapt to the underlying regression structure of the data.

Table 10: Cross-validation error of each estimator for ozone data

	(LL)	(LS)	(AD)	(SI)	(CO)
CV_1	0.2815	0.3183	0.2917	0.4465	0.2768
CV_2	0.1701	0.2174	0.2203	0.3446	0.2065

Figure 7: Residual plots for (AD), (SI) and (CO)



7 Conclusion and Future Works

This paper proposes a semiparametric regression estimation method based on the copula. The estimator combines parametric copulas with empirical marginal distributions. This method is flexible, easy to implement and robust to the curse of dimensionality problems. We derive some asymptotic properties of the proposed estimator. In the simulations, we show the finite sample performance of the estimator. Further research on this work can be done on different angles. For the multiple covariate case, we have considered all the possible vine structures and choose the best one in terms of AIC, but this becomes computationally infeasible when the number of covariates is large. Recently, new methods are proposed for selecting the appropriate vine structure in an efficient way, see for example Weiss and Padberg (2011). It would be interesting to incorporate such a selection scheme into our regression modeling framework. Additionally, since copula has some advantages in modeling tail dependence, it would be interesting to see whether copula regression framework benefits from those advantages in the estimation when the data has a specific tail dependence.

Appendix

Proof of Theorem 3.1

Using Taylor expansion, we have

$$\hat{m}(x_1) = n^{-1}\sum_{i=1}^{n} Y_i c(F_0(Y_i), F_1(x_1); \boldsymbol{\theta}_0) + V_1 + V_2 + V_3,$$

where

$$\begin{aligned} V_1 &= n^{-1}\sum_{i=1}^{n} Y_i(\hat{F}_0(Y_i) - F_0(Y_i))\, c_0(\hat{u}_{i,0}, \hat{u}_1; \tilde{\boldsymbol{\theta}}),\ V_2 = n^{-1}\sum_{i=1}^{n} Y_i(\hat{F}_1(x_1) - F_1(x_1))\, c_1(\hat{u}_{i,0}, \hat{u}_1; \tilde{\boldsymbol{\theta}}), \\ V_3 &= n^{-1}\sum_{i=1}^{n} Y_i(\hat{\theta} - \theta_0)^\top \dot{\boldsymbol{c}}(\hat{u}_{0,i}, \hat{u}_1; \tilde{\boldsymbol{\theta}}) \end{aligned}$$

with $\hat{u}_{i,0} = F_0(Y_i) + t(\hat{F}_0(Y_i) - F_0(Y_i))$, $\hat{u}_1 = F_1(x_1) + t(\hat{F}_1(x_1) - F_1(x_1))$ and $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 + t(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ for some $t \in [0,1]$. Using Taylor expansion again, V_1 can be represented as

$$V_1 = n^{-1}\sum_{i=1}^{n} Y_i(\hat{F}_0(Y_i) - F_0(Y_i))\, c_0(F_0(Y_i), F_1(x_1); \boldsymbol{\theta}_0) + R_1,$$

where

$$R_1 = n^{-1}\sum_{i=1}^{n} Y_i(\hat{F}_0(Y_i) - F_0(Y_i))[c_0(\hat{u}_{i,0}, \hat{u}_1; \tilde{\boldsymbol{\theta}}) - c_0(F_0(Y_i), F_1(x_1); \boldsymbol{\theta}_0)].$$

By Assumption (C1)(i) and (C2) and the compactness of $\boldsymbol{\Theta}$, we have

$$\begin{aligned} |R_1| &\le \sup_i |\hat{F}_0(Y_i) - F_0(Y_i)| \sup_i |c_0(\hat{u}_{i,0}, \hat{u}_1; \tilde{\boldsymbol{\theta}}) - c_0(F_0(Y_i), F_1(x_1); \boldsymbol{\theta}_0)|\, n^{-1}\sum_{i=1}^{n} |Y_i| \\ &= O_p(n^{-1/2}) o_p(1) O_p(1) = o_p(n^{-1/2}), \end{aligned}$$

which leads to $V_1 = \tilde{V}_1 + o_p(n^{-1/2})$ with $\tilde{V}_1 = n^{-1}\sum_{i=1}^{n} Y_i(\hat{F}_0(Y_i) - F_0(Y_i)) c_0(F_0(Y_i), F_1(x_1))$. Similarly, we know that $V_2 = \tilde{V}_2 + o_p(n^{-1/2})$ with $\tilde{V}_2 = n^{-1}\sum_{i=1}^{n} Y_i(\tilde{F}_1(x) - F_1(x)) c_1(F_0(Y_i), F_1(x_1))$, and also that $V_3 = \tilde{V}_3 + o_p(n^{-1/2})$ with $\tilde{V}_3 = n^{-1}\sum_{i=1}^{n} Y_i(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \dot{\boldsymbol{c}}(F_0(Y_i), F_1(x_1))$ using $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = O_p(n^{-1/2})$ in Assumption B. Summing up the results until now, we conclude

$$\hat{m}(x_1) = n^{-1}\sum_{i=1}^{n} Y_i c(F_0(Y_i), F_1(x_1)) + \tilde{V}_1 + \tilde{V}_2 + \tilde{V}_3 + o_p(n^{-1/2}) \tag{14}$$

By Assumption (A) and (C4),

$$\tilde{V}_2 = n^{-1}\sum_{i=1}^{n} \zeta(F_1(X_{i,1}), F_1(x_1)) \times e_1(F_1(x_1)) + o_p(n^{-1/2}) \tag{15}$$

Note that $\tilde{V}_1$ is a V-statistic with the kernel

$$h_1(Y_i, Y_j) = \frac{1}{2}[Y_i(I(Y_j \leq Y_i) - F_0(Y_i))c_0(F_0(Y_i), F_1(x_1)) + Y_j(I(Y_i \leq Y_j) - F_0(Y_j))c_0(F_0(Y_j), F_1(x_1))].$$

By Assumption (C4), using the fact that $\mathbb{E}h_1(Y_i, Y_j) = 0$ and the classical V-statistic techniques (see e.g. Serfling (1980)), we obtain that

$$\tilde{V}_1 = n^{-1} \sum_{i=1}^{n} \lambda(Y_i, F_1(x_1)) + o_p(n^{-1/2}), \tag{16}$$

where $\lambda(t, u_1) = \mathbb{E}[Y(I(t \leq Y) - F_0(Y))c_0(F_0(Y), u_1)]$. Further, $\tilde{V}_3$ is also a V-statistic with the kernel

$$h_3((X_{i,1}, Y_i)^\top, (X_{j,1}, Y_j)^\top) = \frac{1}{2}[Y_i \boldsymbol{\eta}_j^\top \times \dot{\boldsymbol{c}}(F_0(Y_i), F_1(x_1)) + Y_j \boldsymbol{\eta}_i^\top \times \dot{\boldsymbol{c}}(F_0(Y_j), F_1(x_1))].$$

By Assumption (B) and (C4) we obtain that

$$\tilde{V}_3 = n^{-1} \sum_{i=1}^{n} \boldsymbol{\eta}_i^\top \times \dot{\boldsymbol{e}}(F_1(x_1)) + o_p(n^{-1/2}). \tag{17}$$

From (14)-(17), we know that

$$\hat{m}(x_1) = n^{-1} \sum_{i=1}^{n} [Y_i c(F_0(Y_i), F_1(x_1)) + \lambda(Y_i, F_1(x_1)) + \zeta(F_1(X_{i,1}), F_1(x_1)) \times e_1(F_1(x_1)) + \boldsymbol{\eta}^\top(F_0(Y_i), F_1(X_{i,1})) \times \dot{\boldsymbol{e}}(F_1(x_1))] + o_p(n^{-1/2}),$$

This combined with the fact $\lambda(t, u_1) = \mathbb{E}[Y c(F_0(Y), u_1)] - t c(F_0(t), u_1) - \int (I(t \leq y) - F_0(y)) c(F_0(y), u_1) dy$ from Assumption (C1) concludes the proof.

References

K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44:182–198, 2009.

T. Bouezmarni, J. V. K. Rombouts, and A. Taamouti. Asymptotic properties of the bernstein density copula estimator for α-mixing data. *Journal of Multivariate Analysis*, 101:1 – 10, 2010.

E. Brechmann and U. Schepsmeier. Modeling dependence with C-and D-vine copulas: The R-package CDVine. Technical report, 2011.

A. Charpentier, J.-D. Fermanian, and O. Scaillet. *Copulas: From theory to application in finance*, chapter The Estimation of Copulas: Theory and Practice. Risk Publications, London, 2006.

S. X. Chen and T.-M. Huang. Nonparametric estimation of copula functions for dependence modelling. *Canadian Journal of Statistics*, 35:265–282, 2007.

G. J. Crane and J. Van Der Hoek. Conditional Expectation Formulae for Copulas. *Australian and New Zealand Journal of Statistics*, 50:53–67, 2008.

M. J. Crawley. *Statistics: An Introduction using R.* John Wiley & Sons, Ltd., 2005.

C. Genest, K. Ghoudi, and L. Rivest. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82:543–552, 1995.

I. Gijbels and J. Mielniczuk. Estimating the density of a copula function. *Communications in Statistics - Theory and Methods*, 19:445–464, 1990.

T. Hayfield and J. S. Racine. Nonparametric econometrics: The np package. *Journal of Statistical Software*, 27, 2008.

I. Hobæk Haff. Parameter estimation for pair-copula constructions. *Bernoulli*, to appear, 2012.

I. Hobæk Haff, K. Aas, and A. Frigessi. On the simplified pair-copula construction - Simply useful or too simplistic? *Journal of Multivariate Analysis*, 101:1296–1310, 2010.

I. Kojadinovic and J. Yan. A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Statistics and Computing*, 21:17–30, 2011.

D. Kurowicka and H. Joe, editors. *Dependence Modeling: Vine Copula Handbook*. World Scientific Books. World Scientific Publishing Co. Pte. Ltd., 2010.

Y. K. Leong and E. A. Valdez. Claims Prediction with Dependence using Copula Models. Technical report, 2005.

R. B. Nelsen. *An introduction to copulas*. Springer Series in Statistics. Springer, New York, 2006.

C. Ritz and J. C. Streibig. *Nonlinear Regression in R.* Springer, 2008.

R. J. Serfling. *Approximation theorems of mathematical statistics*. John Wiley & Sons Inc., 1980. Wiley Series in Probability and Mathematical Statistics.

P. Silvapulle, G. Kim, and M. J. Silvapulle. Robustness of a semiparametric estimator of a copula. Econometric Society 2004 Australasian Meetings 317, Econometric Society, 2004.

M. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8: 229–231, 1959.

E. A. Sungur. Some Observations on Copula Regression Functions. *Communications in Statistics-Theory and Methods*, 34:1967–1978, 2005.

H. Tsukahara. Semiparametric estimation in copula models. *Canadian Journal of Statistics*, 33: 357–375, 2005.

G. Weiss and M. Padberg. Automated vine copula calibration using genetic algorithms. Technical report, TU Dortmund University, 2011.

H. White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50:1–25, 1982.

J. Yan. Enjoy the Joy of Copulas : With a Package copula. *Journal of Statistical Software*, 21, 2007.