Flexible modeling based on copulas in nonparametric median regression

Peer-reviewed author version

BRAEKERS, Roel & VAN KEILEGOM, Ingrid (2009) Flexible modeling based on copulas in nonparametric median regression. In: JOURNAL OF MULTIVARIATE ANALYSIS, 100(6). p. 1270-1281.

DOI: 10.1016/j.jmva.2008.11.009
Handle: http://hdl.handle.net/1942/9517

Flexible modeling based on copulas in nonparametric median regression

Roel Braekers[1]
and
Ingrid Van Keilegom[2]

September 23, 2008

Abstract

Consider the model $Y = m(X) + \varepsilon$, where $m(\cdot) = \text{med}(Y|\cdot)$ is unknown but smooth. It is often assumed that ε and X are independent. However, in practice this assumption is in many cases violated. In this paper we propose to model the dependence between ε and X by means of a copula model, i.e. $(\varepsilon, X) \sim \mathcal{C}_\theta(F_\varepsilon(\cdot), F_X(\cdot))$, where $\mathcal{C}_\theta$ is a copula function depending on an unknown parameter θ, and F_ε and F_X are the marginals of ε and X. Since many parametric copula families contain the independent copula as a special case, the so-obtained regression model is more flexible than the 'classical' regression model.

We estimate the parameter θ via a pseudo-likelihood method and prove the asymptotic normality of the estimator, based on delicate empirical process theory. We also study the estimation of the conditional distribution of Y given X. The procedure is illustrated by means of a simulation study, and the method is applied to data on food expenditures in households.

Keywords: Conditional distribution; Copulas; Empirical processes; Median regression; Nonparametric regression; Quantiles; Weak convergence.

MSC2000 Classification: Primary: 62G08, Secondary: 62G05, 62G20, 62F12, 62E20

Running head: Copula based modeling in regression

[1]Corresponding Author, Interuniversity Institute for Biostatistics and statistical Bioinformatics, Universiteit Hasselt, Agoralaan 1, 3590 Diepenbeek, Belgium, and Katholieke Universiteit Leuven, Belgium, E-mail: roel.braekers@uhasselt.be

[2]Institute of Statistics, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium, E-mail: vankeilegom@stat.ucl.ac.be

1 Introduction

Consider the model

$$Y = m(X) + \varepsilon, \tag{1.1}$$

where $m(\cdot) = \text{med}(Y|\cdot)$ is the median regression function. The aim of this paper is to propose and study a flexible way to model the dependence between ε and X. This goal will be achieved by using copulas to model the joint distribution of ε and X.

Let us start with motivating this goal. When no assumption is imposed on the relation between ε and X (except that $\text{med}(\varepsilon|X) = 0$), the estimation of the conditional distribution $F(y|x) = P(Y \leq y|X = x)$ can be done by using a kernel estimator of the type $\sum_{i=1}^n W_{ni}(x, h_n) I(Y_i \leq y)$, where $W_{ni}(x, h_n)$ is an appropriate weight function depending on a bandwidth h. This estimator has the advantage of making no model assumption, but the disadvantage of only using local information around the point x. One way to overcome this drawback is to assume that the error term ε is independent of the covariate X, in which case the conditional distribution $F(y|x)$ can be estimated by $n^{-1} \sum_{i=1}^n I(Y_i - \hat{m}(X_i) \leq y - \hat{m}(x))$, where $\hat{m}(\cdot)$ is e.g. a kernel estimator of $m(\cdot)$. This estimator has been studied in Akritas and Van Keilegom (2001) and has the advantage of much better exploiting the available data, since it is a global empirical distribution, instead of a local one. On the other hand, the assumption of independence between ε and X is often not satisfied in practice. See e.g. Einmahl and Van Keilegom (2008a,b), where two procedures are developed for testing this independence assumption.

For these reasons, we propose an intermediate model, which combines the flexibility of the completely nonparametric model, and the efficient use of the data of the model assuming independence. The model assumes that $Y = m(X) + \varepsilon$, where $m(X) = \text{med}(Y|X)$ and the relationship between ε and X is given by

$$(\varepsilon, X) \sim \mathcal{C}_\theta(F_\varepsilon(\cdot), F_X(\cdot)), \tag{1.2}$$

where $F_\varepsilon(y) = P(\varepsilon \leq y)$, $F_X(x) = P(X \leq x)$, and $\mathcal{C}_\theta$ is a copula function belonging to a parametric family $\{\mathcal{C}_\theta : \theta \in \Theta\}$, where Θ is a compact subset of $\mathbb{R}^k$. The true, but unknown, value of θ is denoted by θ_0. Since many copula families contain the independent copula as a special case, the so-obtained regression model is more flexible and robust than the model assuming independence between ε and X, and is on the other hand more efficient than the completely nonparametric model.

Under this regression model we are interested in estimating the conditional distribution $F(y|x)$. The motivation for studying this function has many roots. First of all, one might be interested in the estimation of the conditional distribution itself for a given value of the predictor X. Second, any function or functional of the conditional distribution $F(\cdot|x)$ can be obtained once $F(\cdot|x)$ has been properly estimated. Examples include the conditional quantile function of Y given X, the Lorenz curve or Gini index, any conditional moment (skewness, kurtosis, ...), the extreme value index, etc. The conditional distribution $F(y|x)$ can be rewritten as

$$\begin{aligned} F(y|x) &= P\left(m(X)+\varepsilon \leq y|X=x\right) = P\left(\varepsilon \leq y-m(x)|X=x\right) \\ &= \mathcal{C}_\theta^2\left(F_\varepsilon\left(y-m(x)\right), F_X(x)\right), \end{aligned} \tag{1.3}$$

where $\mathcal{C}_\theta^2(u,v) = \frac{\partial}{\partial v}\mathcal{C}_\theta(u,v)$, $(u,v) \in [0,1]^2$, is the partial derivative of the copula function $\mathcal{C}_\theta$ with respect to its second component. Hence, $F(y|x)$ can be estimated once we have estimators for θ, the marginal distributions of ε and X, and the regression function $m(\cdot)$. Note that due to the relation between the error variable ε and the covariate X, the conditional distribution of the response Y is also influenced by the distribution of the covariate X.

Copula models have become a useful and important tool in modeling dependencies between random variables. They have been used in a large variety of areas in statistics, like in survival analysis (see e.g. Wang and Wells (2000), Braekers and Veraverbeke (2005) and Chen and Fan (2007) for some of the more recent contributions in this field), in risk theory (Frees and Valdez (1998), Charpentier and Segers (2007), Genest and Segers (2008), among others) and in econometrics (see e.g. Hu (2006)). See also Joe (1997) and Nelsen (1999) for two books devoted to this topic.

Note that the copula $\mathcal{C}$ has to be chosen in such a way that $\text{med}(\varepsilon|X) = 0$. It is easy to see that this constraint is equivalent to imposing that

$$\text{med}(\varepsilon) = 0 \qquad \text{and} \qquad \text{med}(U|V) = 1/2, \tag{1.4}$$

where U and V are uniform random variables on $[0,1]$ satisfying $(U,V) \sim C_\theta$. This equivalence is important, as it allows to decompose the constraint $\text{med}(\varepsilon|X) = 0$ into a constraint on the marginal distribution of ε and a constraint on the copula function. On the other hand, if we would have taken a mean regression model, then the copula would have to satisfy $E(\varepsilon|X) = 0$, and this cannot be decomposed in a constraint only on the marginals, and a constraint only on the copula function. Hence, the median regression model offers important advantages over the mean regression model in this context.

The paper is organized as follows. In Section 2, we define the estimators of the parameter vector θ_0 and of the conditional distribution $F(y|x)$. Section 3 is devoted to the asymptotic properties of the two estimators. In Section 4 we investigate the finite sample properties of the estimators in a simulation study, whereas data on food expenditures in households are analyzed in Section 5. Finally, in the Appendix we give the proofs of the asymptotic results of Section 3.

2 The proposed estimators

Let (X_i, Y_i), $i = 1, \ldots, n$, be i.i.d. data coming from the model defined by (1.1), (1.2) and (1.4). We develop in this section an estimator of the association parameter θ_0 and of the conditional distribution $F(y|x)$. In order to estimate θ_0, we first need to estimate the marginal distributions of X and ε. Define

$$\hat{F}_X(x) = n^{-1} \sum_{i=1}^{n} I(X_i \leq x),$$

and

$$\hat{F}_\varepsilon(y) = n^{-1} \sum_{i=1}^{n} I(\hat{\varepsilon}_i \leq y).$$

Here, $\hat{\varepsilon}_i = Y_i - \hat{m}(X_i)$, $i = 1, \ldots, n$, and for any x in the support R_X of X,

$$\hat{m}(x) = \tilde{F}^{-1}(0.5|x) = \inf\{y : \tilde{F}(y|x) \geq 0.5\},$$

where $\tilde{F}(y|x) = \sum_{i=1}^{n} W_{ni}(x, h_n) I(Y_i \leq y)$ is a weighted empirical distribution function (see Stone 1977). The weights are the Nadaraya-Watson kernel weights, defined as

$$W_{ni}(x, h_n) = \frac{K\left(\frac{x - X_i}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right)}, \ i = 1, \ldots, n,$$

with K a probability density function (kernel) and $h = h_n$ a sequence of positive constants tending to zero as n tends to infinity (bandwidth sequence). The estimator $\hat{F}_\varepsilon$ has been proposed and studied in detail in Akritas and Van Keilegom (2001).

We now estimate θ_0 by using a pseudo-likelihood approach, as in Genest, Ghoudi and Rivest (1995) and Tsukahara (2005). Other approaches are possible. See e.g. Chen, Fan and Tsyrennikov (2006) for a sieve maximum likelihood procedure to estimate jointly the association parameter θ_0 and the marginals of ε and X. Suppose that the copula $\mathcal{C}_\theta$ is absolutely continuous with density $\mathcal{C}_\theta^{12}(u,v) = \frac{\partial^2}{\partial u \partial v}\mathcal{C}_\theta(u,v)$, $(u,v) \in [0,1]^2$, and that this density is differentiable with respect to the components of θ. Denote $\mathcal{C}_\theta^{12'}(u,v) = \left(\frac{\partial}{\partial \theta_1}\mathcal{C}_\theta^{12}, \ldots, \frac{\partial}{\partial \theta_k}\mathcal{C}_\theta^{12}\right)(u,v)$. We estimate θ_0 by

$$\hat{\theta} = \text{argmax}_{\theta \in \Theta}\, l(\theta), \tag{2.1}$$

where $l(\theta)$ is the following pseudo-loglikelihood function :

$$l(\theta) = \sum_{i=1}^{n} \log \mathcal{C}_\theta^{12}\left(\hat{F}_\varepsilon^*(\hat{\varepsilon}_i), \hat{F}_X^*(X_i)\right),$$

with $\hat{F}_\varepsilon^* = n/(n+1)\hat{F}_\varepsilon$ and $\hat{F}_X^* = n/(n+1)\hat{F}_X$. This is equivalent to finding the value $\hat{\theta}$ of θ which solves the equation

$$\sum_{i=1}^{n} \frac{\mathcal{C}_\theta^{12'}\left(\hat{F}_\varepsilon^*(\hat{\varepsilon}_i), \hat{F}_X^*(X_i)\right)}{\mathcal{C}_\theta^{12}\left(\hat{F}_\varepsilon^*(\hat{\varepsilon}_i), \hat{F}_X^*(X_i)\right)} = 0. \tag{2.2}$$

Combining the previous estimators and using equation (1.3), we define an estimator of the conditional distribution $F(y|x)$ by

$$\hat{F}(y|x) = \mathcal{C}_{\hat{\theta}}^{2}\left(\hat{F}_\varepsilon\left(y - \hat{m}(x)\right), \hat{F}_X(x)\right). \tag{2.3}$$

Remark 2.1. Note that in the above estimation procedure we have estimated the conditional median $m(\cdot)$ and the marginal distribution F_ε of ε in a nonparametric way. One could however also replace them by parametric or semiparametric estimators. This does not change the basic idea of using copulas to model the dependence between the error and the covariate, and has as far as we know, never been proposed in the literature.

Remark 2.2. If the goal of the analysis would be to estimate the median regression function $m(x)$ instead of estimating the conditional distribution $F(y|x)$, then one could update the original completely nonparametric estimator $\hat{m}(x)$ by a new, copula based, estimator given by $\hat{F}^{-1}(0.5|x)$. We do not study this estimator in this paper, but its asymptotic properties could be derived in a fairly easy way, starting from the properties of $\hat{F}(y|x)$.

Remark 2.3. The proposed copula model can also be interpreted by looking at the quantile function $F^{-1}(\cdot|x)$ for a given x. It is easily seen that under the assumed model, we have for any $0 < p < 1$,

$$F^{-1}(p|x) = m(x) + F_\varepsilon^{-1}\Big((C^2_{\theta,F_X(x)})^{-1}(p)\Big),$$

where $(C^2_{\theta,F_X(x)})^{-1}(p) = z$ if and only if $C^2_\theta(z, F_X(x)) = p$. Hence, for a fixed value of x, the difference between two quantiles is completely driven by the choice of the copula function (and the marginals of ε and X). The nice feature of the above formula of $F^{-1}(p|x)$ is that it is monotone in p, or in other words, the quantile curves will never cross. See also Cosma, Scaillet and von Sachs (2007) for other nonparametric estimation methods that are shape preserving.

3 Asymptotic properties

We will develop the asymptotic theory of the proposed estimators $\hat{\theta}$ and $\hat{F}(y|x)$ by making use of the results in Chen, Linton and Van Keilegom (2003), who developed generic conditions under which a parameter estimator that is defined via an estimating equation depending on some nonparametric nuisance functions, is consistent and asymptotically normal. Define

$$G_n(\theta, H, F_X) = n^{-1}\sum_{i=1}^n g(X_i, Y_i, \theta, H, F_X) \tag{3.1}$$

$$G(\theta, H, F_X) = E[g(X, Y, \theta, H, F_X)], \tag{3.2}$$

where

$$g(x, y, \theta, H, F_X) = \frac{C_\theta^{12\prime}(H(x,y), F_X(x))}{C_\theta^{12}(H(x,y), F_X(x))}, \tag{3.3}$$

$H(x, y) = F_\varepsilon(y - m(x))$, $C^1_\theta(u, v)$, $C^2_\theta(u, v)$ and $C'_\theta(u, v)$ denote respectively the derivative of $C_\theta(u, v)$ with respect to u, v and the vector θ, and higher order derivatives of $C_\theta(u, v)$ are defined in a similar way.

The true value θ_0 of θ then satisfies $G(\theta_0, H, F_X) = 0$, and $\hat{\theta} = \text{argmin}_{\theta\in\Theta}\|G_n(\theta, \hat{H}^*, \hat{F}_X^*)\|$, where $\|\cdot\|$ denotes the Euclidean norm, and $\hat{H}^*(x, y) = \hat{F}^*_\varepsilon(y - \hat{m}(x))$.

The following regularity conditions are needed for the results below.

(A1) (i) h satisfies $nh^{3+\delta}(\log h^{-1})^{-1} \to \infty$ for some $\delta > 0$ and $nh^4 \to 0$.

(ii) K has compact support, is symmetric and is twice continuously differentiable.

(iii) R_X is a closed interval in $I\!R$.

(A2) Except for a finite number of values of u, v and θ, the function $(u, v, \theta) \to \mathcal{C}_\theta^{12'}(u, v)/\mathcal{C}_\theta^{12}(u, v)$ is twice continuously differentiable with respect to u and v, and once with respect to the components of θ, and all these derivatives are continuous in (u, v, θ).

(A3) (i) F_X is three times continuously differentiable and $\inf_x f_X(x) > 0$.
(ii) $F(y|x)$ is twice continuously differentiable with respect to x and y, all derivatives up to order two are continuous in (x, y) and are bounded uniformly in (x, y).

(A4) (i) For all $\delta > 0$, there exists $\epsilon > 0$ such that $\inf_{\|\theta-\theta_0\|>\delta} \|G(\theta, H, F_X)\| \geq \epsilon$.
(ii) $\Gamma = \frac{\partial}{\partial\theta} G(\theta, H, F_X)|_{\theta=\theta_0}$ is of full rank.

Theorem 3.1 *Assume (A1)-(A4). Then,*

$$n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Gamma^{-1} V \Gamma^{-1}),$$

where

$$V = Var\Big\{g(X, Y, \theta_0, H, F_X) + v(X, Y, \theta_0)\Big\}$$

and

$$\begin{aligned} v(x, y, \theta) &= E\left[\frac{\partial}{\partial u} d_\theta(u, F_X(X))\Big|_{u=F_\varepsilon(\varepsilon)} \Big\{I\Big(y - m(x) \leq \varepsilon\Big) - F_\varepsilon(\varepsilon) + \varphi(x, y, \varepsilon)\Big\}\right. \\ &\qquad \left. + \frac{\partial}{\partial v} d_\theta(F_\varepsilon(\varepsilon), v)\Big|_{v=F_X(X)} \Big\{I(x \leq X) - F_X(X)\Big\}\right] \\ &\quad + E\left[\frac{\partial}{\partial u} d_\theta(u, F_X(x))\Big|_{u=F_\varepsilon(\varepsilon)} f_\varepsilon(\varepsilon)\Big| X = x\right] \frac{I(y - m(x) \leq 0) - \frac{1}{2}}{f(m(x)|x)}, \\ d_\theta(u, v) &= \frac{C_\theta^{12'}(u, v)}{C_\theta^{12}(u, v)}, \\ \varphi(x, y, e) &= -f_\varepsilon(e) \frac{I(y - m(x) \leq 0) - \frac{1}{2}}{f(m(x)|x)}. \end{aligned}$$

Theorem 3.2 *Assume (A1)-(A4). Then, for any $x \in R_X$, the process*

$$(nh)^{1/2}(\hat{F}(y|x) - F(y|x))$$

$(-\infty < y < \infty)$ converges weakly to a Gaussian process $W(y|x)$ with zero mean and covariance function given by

$$Cov(W(y_1|x), W(y_2|x)) = f(y_1|x) f(y_2|x) \frac{\|K\|_2^2}{4 f^2(m(x)|x) f_X(x)},$$

where $\|K\|_2^2 = \int K^2(u) du$.

Remark 3.1. Note that the asymptotic variance of $\hat{\theta}$ does not show up in the asymptotic variance of $\hat{F}(y|x)$, since $\hat{\theta}$ is estimated at a faster rate of convergence than $\hat{F}(y|x)$. The copula model is however indirectly present in the above variance formula, via the conditional density of Y given X.

It is also interesting to compare the asymptotic variance of $\hat{F}(y|x)$ with that of the completely nonparametric estimator $\tilde{F}(y|x)$ of $F(y|x)$, given by

$$\tilde{F}(y|x) = \frac{1}{\sum_{i=1}^{n} K\Big(\frac{x-X_i}{h}\Big)} \sum_{i=1}^{n} K\Big(\frac{x-X_i}{h}\Big) I(Y_i \leq y).$$

The asymptotic variance of $(nh)^{1/2}\tilde{F}(y|x)$ equals $\frac{\|K\|_2^2}{f_X(x)} F(y|x)(1-F(y|x))$. Hence, when $y = m(x)$, both estimators have the same asymptotic variance. Figure 1 shows the asymptotic variance of both estimators for a number of choices of the density $f(\cdot|x)$. Note that for the normal and Student-t case (graphs on first row of Figure 1), the new estimator outperforms the completely nonparametric one for all y, whereas for the Weibull and χ^2 distribution this is not true for a certain range of y-values. This can be explained by the fact that the variance of $\hat{F}(y|x)$ is driven by the density $f(\cdot|x)$, whereas the variance of $\tilde{F}(y|x)$ depends on the value of the distribution $F(y|x)$. These two functions behave quite differently especially when the density is asymmetric. Overall, the new estimator shows however a significant improvement in the variance.

4 Simulations

In this section we carry out a simulation study to investigate the finite sample properties of the estimators $\hat{\theta}$ and $\hat{F}(y|x)$. Note that, as explained in Section 1, the copula function $\mathcal{C}_\theta$ needs to satisfy $\text{med}(U|V) = 1/2$, where $(U,V) \sim \mathcal{C}_\theta$. The independent copula $\Pi(u,v) = uv$ satisfies this property, but most other common copula functions do not. Therefore, we first need to find a way to create parametric families of copulas satisfying this property. This can either be done by constructing a copula family 'by hand' or by adjusting (or transforming) an existing family of copulas. Here, we work out the second method, and we propose four examples of possible transformations. Many other transformations can be thought of, and the transformation to use in practice will in fact depend on the type of dependence structure to be expected. In the first two constructions, we linearly redistribute, for a given value of v, the probability mass of the conditional distribution $F_{U|V}(u|v)$ in such a way that $F_{U|V}(1/2|v)$ becomes equal to $1/2$, while in the

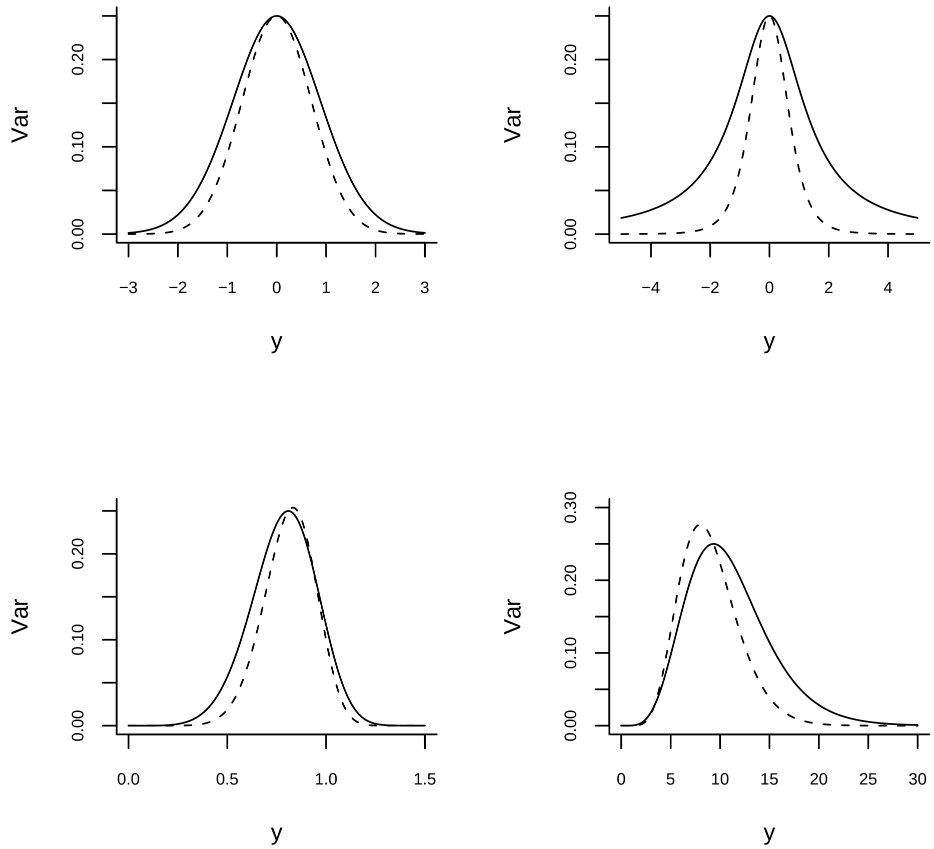

Figure 1: *Graphs of the asymptotic variance functions of $\hat{F}(y|x)$ (dashed curve) and $\tilde{F}(y|x)$ (full curve), divided by their common factor ($\|K\|_2^2/f_X(x)$). The upper left corner corresponds to the case where $f(y|x)$ is a standard normal density, the upper right corner is for a Student-t density with 2 degrees of freedom, the lower left corner for a Weibull density with parameters (2,5) and the lower right corner for a χ^2 density with 10 degrees of freedom.*

last two constructions the u-axis is transformed (first linearly, then quadraticly). This idea is visually represented in Figure 2.

The first transformation is given by the following ‘symmetric’ copula :

$$\mathcal{C}_{1\theta}(u,v) = \begin{cases} \displaystyle\int_0^v \frac{\mathcal{C}_\theta^2(F_{1Y}^{-1}(u),v^*)}{2\mathcal{C}_\theta^2\,(0.5,v^*)}dv^* & 0 \leq u \leq 0.5, 0 \leq v \leq 1 \\ \displaystyle v - \int_0^v \frac{\mathcal{C}_\theta^2(1-F_{1Y}^{-1}(u),v^*)}{2\mathcal{C}_\theta^2\,(0.5,v^*)}dv^* & 0.5 < u \leq 1, 0 \leq v \leq 1 \end{cases}$$

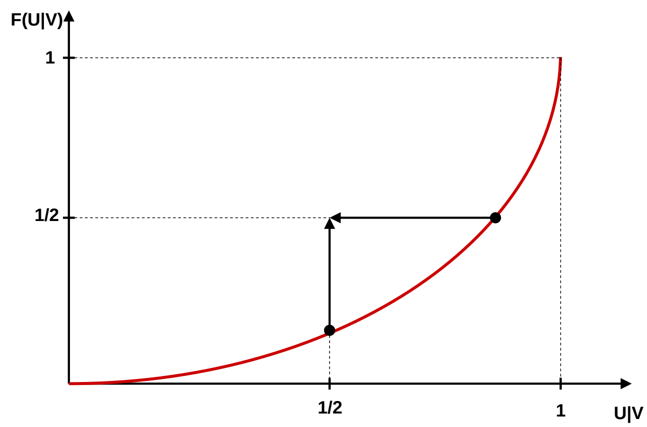

Figure 2: *Visual representation of the transformation of the probability mass (vertical transformation) and of the u-axis (horizontal transformation) in order to obtain a copula for which med*$(U|V) = 1/2$.

with

$$F_{1Y}(y) = \begin{cases} \int\limits_0^1 \frac{\mathcal{C}_\theta^2(y, v^*)}{2\mathcal{C}_\theta^2\,(0.5, v^*)} dv^* & 0 \leq y \leq 0.5 \\ 1 - \int\limits_0^1 \frac{\mathcal{C}_\theta^2(1-y, v^*)}{2\mathcal{C}_\theta^2\,(0.5, v^*)} dv^* & 0.5 < y \leq 1 \end{cases}$$

Note that with this transformation, the conditional density of $U|V = v$ is symmetric around 0.5 for every v. A second transformation is given by

$$\mathcal{C}_{2\theta}(u, v) = \begin{cases} \int\limits_0^v \frac{\mathcal{C}_\theta^2(F_{2Y}^{-1}(u), v^*)}{2\mathcal{C}_\theta^2\,(0.5, v^*)} dv^* & 0 \leq u \leq 0.5, 0 \leq v \leq 1 \\ \frac{v}{2} + \int\limits_0^v \frac{\mathcal{C}_\theta^2(F_{2Y}^{-1}(u), v^*) - \mathcal{C}_\theta^2(0.5, v^*)}{2(1 - \mathcal{C}_\theta^2\,(0.5, v^*))} dv^* & 0.5 < u \leq 1, 0 \leq v \leq 1 \end{cases}$$

with

$$F_{2Y}(y) = \begin{cases} \int\limits_0^1 \frac{\mathcal{C}_\theta^2(y, v^*)}{2\mathcal{C}_\theta^2\,(0.5, v^*)} dv^* & 0 \leq y \leq 0.5 \\ 0.5 + \int\limits_0^1 \frac{\mathcal{C}_\theta^2(y, v^*) - \mathcal{C}_\theta^2(0.5, v^*)}{2(1 - \mathcal{C}_\theta^2\,(0.5, v^*))} dv^* & 0.5 < y \leq 1 \end{cases}$$

In this second construction, we redistributed the probability mass in the conditional distribution of $U|V = v$ to get an equal mass on both sides of $u = 0.5$. In a third construction, we linearly transform the u-axis and get for $0 \leq v \leq 1$,

$$\mathcal{C}_{3\theta}(u,v) = \begin{cases} \int_0^v \mathcal{C}_\theta^2 \left(2F_{3Y}^{-1}(u)(\mathcal{C}_{\theta,v^*}^2)^{-1}(0.5), v^*\right) dv^* & 0 \leq u \leq 0.5 \\ \int_0^v \mathcal{C}_\theta^2 \left(1 - 2((\mathcal{C}_{\theta,v^*}^2)^{-1}(0.5) - 1)(F_{3Y}^{-1}(u) - 1), v^*\right) dv^* & 0.5 < u \leq 1 \end{cases}$$

with

$$F_{3Y}(y) = \begin{cases} \int_0^1 \mathcal{C}_\theta^2 \left(2y(\mathcal{C}_{\theta,v^*}^2)^{-1}(0.5), v^*\right) dv^* & 0 \leq y \leq 0.5 \\ \int_0^1 \mathcal{C}_\theta^2 \left(1 - 2((\mathcal{C}_{\theta,v^*}^2)^{-1}(0.5) - 1)(y - 1), v^*\right) dv^* & 0.5 < y \leq 1 \end{cases}$$

The fourth construction is similar to the third one, but is based on a bounded second order interpolation instead of a piecewise linear transformation of the u-axis :

$$\mathcal{C}_{4\theta}(u,v) = \int_0^v \mathcal{C}_\theta^2 \left(G(F_{4Y}^{-1}(u), v^*), v^*\right) dv^* \text{ with } F_{4Y}(y) = \int_0^1 \mathcal{C}_\theta^2 \left(G(F_{4Y}^{-1}(u), v^*), v^*\right) dv^*$$

and $G(y,v) = \max\left(0, \min\left(2y(y-0.5) - 4y(y-1)(\mathcal{C}_{\theta,v})^{-1}(0.5), 1\right)\right)$. In Figure 3, we generate a sample of 1500 data points for the four different copula constructions. As underlying copula we use the Frank copula with $\theta = 5$. We note that for constructions 2 and 3 the data cloud splits into two separate groups, unlike for constructions 1 and 4. Consequently, the bivariate density function is not continuous at $u = 0.5$ for these constructions. To show that the four constructed copulas are indeed copulas, we use the definition of a copula given by Nelsen (1999). Each of the constructed functions has uniform marginals by construction. To show that these functions are 2-increasing on any rectangle $[u_1, u_2] \times [v_1, v_2]$ with $u_1, u_2, v_1, v_2 \in [0, 1]$, $u_1 \leq u_2$ and $v_1 \leq v_2$, we note that either $[u_1, u_2] \times [v_1, v_2]$ lies within $[0, 0.5] \times [0, 1]$ $(u_1 \leq u_2 \leq 0.5)$ or $[0.5, 1] \times [0, 1]$ $(0.5 \leq u_1 \leq u_2)$, or across the line $u = 0.5$ $(u_1 \leq 0.5 \leq u_2)$. In the first and second setting, the constructed functions are 2-increasing due to the monotonicity of the underlying copula. In the third setting, we divide the rectangle $[u_1, u_2] \times [v_1, v_2]$ into two areas $[u_1, 0.5] \times [v_1, v_2]$ and $[0.5, u_2] \times [v_1, v_2]$, and it is easy to show that the constructed functions are 2-increasing on both areas separately.

We now use the above copula constructions to carry out a small simulation study. We assume that the error ε is standard normally distributed and that the covariate X is uniformly distributed on $[0, 1]$. Furthermore, let $m(x) = 5.5 - 4x + 3x^2$. We model

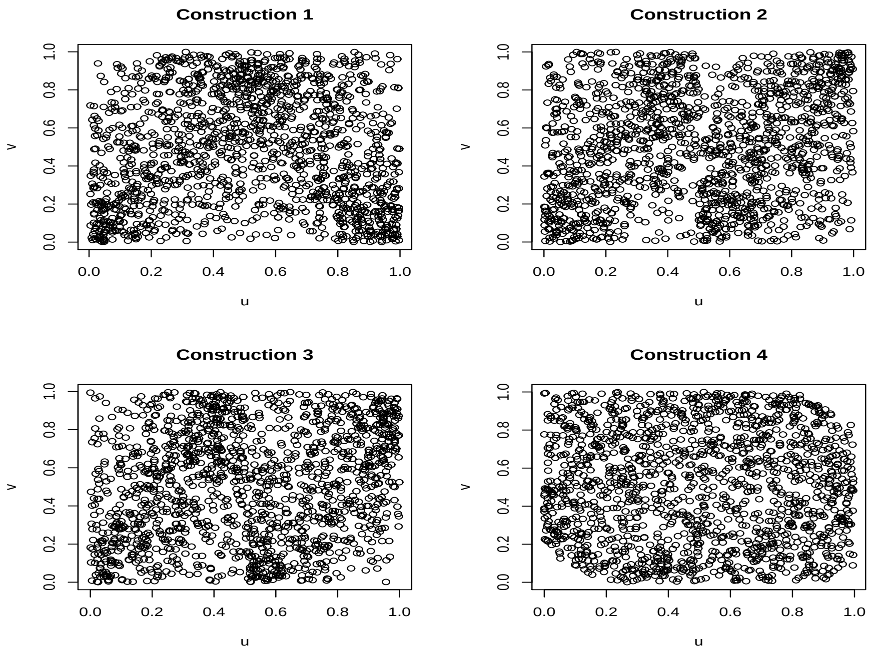

Figure 3: *A simulated data cloud* ($n = 1500$) *generated from the different copula constructions with an underlying Frank copula* ($\theta = 5$).

the association between the error ε and the covariate X by using the first three copula constructions outlined above with both the Frank and the Plackett copula as underlying copula. The fourth copula construction is not considered in the simulations, since it leads to much heavier computations. On the other hand, we will illustrate this copula construction for the analysis of a data set, considered in the next section. For each sample, we estimate the conditional median $m(x)$ by the estimator $\hat{m}(x)$ studied in Section 2, where we take a biquadratic kernel function $K(x) = \frac{15}{16}(1-x^2)^2 I(|x| \leq 1)$ and use a cross validation criterion to determine the best bandwidth. This nonparametric estimator does not depend on the form of the relation between ε and X, and therefore consistently estimates $m(x)$, even when the model is not fully satisfied.

In Tables 1 and 2 we present the results of the simulation study. Table 1 gives for each combination of the copula constructions and underlying copulas, the bias and variance of the estimator $\hat{\theta}$, while Table 2 gives the bias and variance of the estimator $\hat{F}(4.5|0.45)$. The results are shown for different sample sizes. The number of simulated samples is in each setting equal to 500. When the underlying copula is the Frank copula, we take $\theta = 5$, while for the Plackett copula we set $\theta = 25$. In both scenarios, we note that we clearly stay away from the independent copula. Table 1 shows that the bias of $\hat{\theta}$ is larger for the copula constructions 2 and 3 than for the 'symmetric' construction 1. As we expect the bias diminishes when the sample size increases. The bias always has a negative sign, which means that the estimated values of θ are shrinked towards zero. Table 2 indicates that the bias of the estimator $\hat{F}(4.5|0.45)$ is small in all cases. This suggests that the shrinkage which is present for the estimation of θ, has only little effect on the estimation of the conditional distribution.

Sample size	Construction 1	Construction 2	Construction 3
		Frank family ($\theta = 5$)	
$n = 100$	-0.9301 (2.1502)	-3.5619 (3.1948)	-3.0748 (3.2799)
$n = 500$	-0.3366 (0.3409)	-2.4372 (1.1214)	-2.3802 (1.1508)
$n = 1000$	-0.1791 (0.1871)	-2.1701 (0.7634)	-2.0186 (0.5946)
		Plackett family ($\theta = 25$)	
$n = 100$	-11.9732 (63.6874)	-17.7490 (50.4729)	-20.9438 (35.1630)
$n = 500$	-5.7858 (25.8923)	-14.7173 (32.7931)	-20.5509 (7.7449)
$n = 1000$	-4.1806 (16.1749)	-13.0706 (32.8799)	-19.7610 (5.8461)

Table 1: *The bias (variance) of the estimator* $\hat{\theta}$.

Sample size	Construction 1	Construction 2	Construction 3
		Frank family ($\theta = 5$)	
$n = 100$	0.0053 (0.0090)	-0.0303 (0.0083)	-0.0238 (0.0099)
$n = 500$	-0.0032 (0.0016)	-0.0361 (0.0017)	-0.0317 (0.0014)
$n = 1000$	-0.0020 (0.0009)	-0.0414 (0.0007)	-0.0286 (0.0007)
		Plackett family ($\theta = 25$)	
$n = 100$	0.0058 (0.0073)	-0.0603 (0.0073)	-0.0221 (0.0094)
$n = 500$	-0.0004 (0.0015)	-0.0630 (0.0012)	-0.0319 (0.0014)
$n = 1000$	-0.0011 (0.0010)	-0.0624 (0.0008)	-0.0322 (0.0007)

Table 2: *The bias (variance) of the estimator* $\hat{F}(4.5|0.45)$.

Finally, we have carried out some simulations in which we assume that the function

$m(x)$ belongs to the class of quadratic regression functions, and estimated the regression coefficients from parametric L_1-regression. The bias of the estimator $\hat{\theta}$ turned out to be considerably smaller than in Table 1, which suggests that the bias in Table 1 is mainly caused by the fact that $m(x)$ is estimated nonparametrically.

5 Example: food expenditures in households

In this section, we illustrate the developed estimation methods on a data set on food expenditures in Dutch households. The data are extracted from the Data Archive of the Journal of Applied Econometrics; see also Adang and Melenberg (1995). As in Einmahl and Van Keilegom (2008a,b), we look at the expenditures on food and the total expenditures accumulated over the year from October 1986 through September 1987 for two person households. The sample size is $n = 159$. Two models are considered. First, we regress the response Y_1 = share of food expenditure on X = log(total expenditure) and second, we study the relationship between Y_2 = log(food expenditure) and X. Einmahl and Van Keilegom (2008a,b) have tested the independence between the error $\varepsilon_j = Y_j - m_j(X)$ $(j = 1, 2)$ and X in a completely nonparametric way by using two different test procedures. The results of their tests suggest that in the first model the error depends on X, whereas in the second it does not.

Here, we will analyze these data by assuming that the dependence between ε_j and X follows a given copula model, i.e. we assume that $Y_j = m_j(X) + \varepsilon_j$, where $m_j(X) = \text{med}(Y_j|X)$ and

$$(\varepsilon_j, X) \sim \mathcal{C}_{\theta_j}(F_{\varepsilon_j}(\cdot), F_X(\cdot)) \quad (j = 1, 2). \tag{5.1}$$

We use the Frank and Plackett copula, combined with the copula transformation $\mathcal{C}_{4\theta_j}$, given in the previous section. Under these models, we like to estimate the parameter θ_j, but also we like to test whether ε_j and X are independent.

Under the assumed copula model, testing whether the error ε is independent of X, is equivalent to testing whether the association parameter θ equals the value of θ that corresponds to the independent copula ($\theta = 0$ for the Frank copula, and $\theta = 1$ for the Plackett copula). In this paper we use the following likelihood ratio type test statistic :

$$LR = 2l(\hat{\theta}) - 2l(\theta_{H_0})$$

where $l(\theta)$ is the pseudo-loglikelihood function given in Section 2. We note that under H_0 (i.e. when ε and X are independent), the second term in this expression is zero. Furthermore we see that, under H_0, the association parameter θ is embedded in the interior of the parameter space. Therefore, we expect that the distribution of the test statistic LR converges to a χ^2 distribution with one degree of freedom. Based on this approximated distribution, we calculate the p-value of the test. Table 3 gives the results of this analysis. The bandwidth for calculating $m(\cdot)$ is determined by means of a cross-validation procedure, as in the simulation section. The table shows that for the response Y_2, the independence between ε_2 and X is strongly accepted, since the p-values are close to one for both the transformed Frank and Plackett copula. On the other hand, for the response Y_1 the situation is less clear. The p-values are non-significant, but are somewhat borderline for the Frank copula. More research is needed here to understand this dependence.

	Response 1			Response 2		
	$\hat{\theta}_1$	LR	p-value	$\hat{\theta}_2$	LR	p-value
Frank	2.573	2.436	0.119	4.14×10^{-8}	2.36×10^{-8}	0.999
Plackett	3.257	1.791	0.181	0.999	-2.04×10^{-14}	1

Table 3: *Estimated values of θ, values of the LR-test statistic and associated p-values for model (5.1).*

A crucial element that needs more investigation is the choice of the copula model and of the copula transformation. More insight in the nature of the dependence between ε and X, coming from economical studies on expenditure behavior (the so-called Engel curves) will be useful here. Furthermore it would be interesting to construct and do inference for a goodness-of-fit test for the parametric copula family to which our unknown copula $\mathcal{C}_{\theta_j}$ belongs. In the case where the errors $\varepsilon_1, \ldots, \varepsilon_n$ would be observable, this problem has been studied by Fermanian (2005), Scaillet (2007), among others. See also Genest and Rémillard (2008) for a bootstrap approximation in general semiparametric models, that could be used to approximate the distribution of the test statistic. In the present context where $\varepsilon_1, \ldots, \varepsilon_n$ need to be estimated, it would be interesting to work out the asymptotic theory and to test the finite sample behavior of appropriately modified versions of the above mentioned tests.

Appendix: Proofs

Proof of Theorem 3.1. Throughout this proof, we will denote the true distributions of ε and X by $F_{\varepsilon 0}$ and F_{X0} respectively. We will make use of Theorem 2 in Chen, Linton and Van Keilegom (2003) (CLV hereafter), which gives generic conditions under which $\hat{\theta}$ is asymptotically normal. First of all, we need to show that $\hat{\theta} - \theta_0 = o_P(1)$. For this, we verify the conditions of Theorem 1 in CLV. Condition (1.1) holds by definition of $\hat{\theta}$, while the second and third condition are guaranteed by assumptions (A2) and (A4). Finally, conditions (1.4) and (1.5) are weaker than conditions (2.4) and (2.5), respectively, of Theorem 2 of CLV, which we will verify below. So, the conditions of Theorem 1 are verified, except for conditions (1.4) and (1.5) which we postpone to later. Next, we verify conditions (2.1)–(2.6) of Theorem 2 in CLV. Condition (2.1) is, as for condition (1.1), valid by construction of the estimator $\hat{\theta}$, while condition (2.2) follows from assumptions (A2) and (A4). Straightforward calculations show that

$$
\begin{aligned}
&\Gamma_2(\theta, H_0, F_{X0})[H - H_0, F_X - F_{X0}] \\
&= \lim_{\tau \to 0} \tau^{-1} \Big[G(\theta, H_0 + \tau(H - H_0), F_{X0} + \tau(F_X - F_{X0})) - G(\theta, H_0, F_{X0}) \Big] \\
&= E\Big[\frac{\partial}{\partial u} d_\theta(u, F_{X0}(X)) \Big|_{u = F_{\varepsilon 0}(\varepsilon)} (H - H_0)(X, Y) \\
&\qquad + \frac{\partial}{\partial v} d_\theta(F_{\varepsilon 0}(\varepsilon), v) \Big|_{v = F_{X0}(X)} (F_X - F_{X0})(X) \Big],
\end{aligned}
$$

and hence,

$$
\begin{aligned}
&G(\theta, H, F_X) - G(\theta, H_0, F_{X0}) - \Gamma_2(\theta, H_0, F_{X0})[H - H_0, F_X - F_{X0}] \\
&= E\Big\{ d_\theta(H(X, Y), F_X(X)) - d_\theta(F_{\varepsilon 0}(\varepsilon), F_{X0}(X)) \\
&\qquad - \frac{\partial}{\partial u} d_\theta(u, F_{X0}(X)) \Big|_{u = F_{\varepsilon 0}(\varepsilon)} (H - H_0)(X, Y) \\
&\qquad - \frac{\partial}{\partial v} d_\theta(F_{\varepsilon 0}(\varepsilon), v) \Big|_{v = F_{X0}(X)} (F_X - F_{X0})(X) \Big\}.
\end{aligned}
\tag{A.1}
$$

Hence, using a Taylor expansion of order two, it follows that the norm of (A.1) is bounded by a constant times $\|(H - H_0, F_X - F_{X0})\|^2_{\mathcal{H}} := \max(E|H(X,Y) - H_0(X,Y)|^2, E|F_X(X) - F_{X0}(X)|^2)$. This shows the first part of condition (2.3). For the second part, it follows from the proof of Theorem 2 in CLV that it suffices to show that

$$
\begin{aligned}
&\Big\| \Gamma_2(\hat{\theta}, H_0, F_{X0})[\hat{H}^* - H_0, \hat{F}_X^* - F_{X0}] - \Gamma_2(\theta_0, H_0, F_{X0})[\hat{H}^* - H_0, \hat{F}_X^* - F_{X0}] \Big\| \\
&= o_P(1) \|\hat{\theta} - \theta_0\|,
\end{aligned}
$$

and this follows from the differentiability conditions on the copula function C and from the mean value theorem. Next, for condition (2.4), define the classes

$$\begin{aligned}\mathcal{H}_1 = \Big\{(x,y) \to h_1\Big(y - g_m(x)\Big) : h_1 \text{ is monotone and maps onto } [0,1], \qquad \text{(A.2)}\\ g_m \in C_M^{1+\delta}(R_X)\Big\},\\ \mathcal{H}_2 = \{x \to h_2(x) : h_2 \text{ is monotone and maps onto } [0,1]\}, \qquad \text{(A.3)}\end{aligned}$$

and $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$, where $C_M^{1+\delta}(R_X)$ (for some $\delta > 0$) is defined by the set of all differentiable functions $g : R_X \to I\!R$ with $\|g\|_{1+\delta} \leq M$, where

$$\|g\|_{1+\delta} := \max\{\sup_x |g(x)|, \sup_x |g'(x)|\} + \sup_{x_1,x_2} \frac{|g'(x_1) - g'(x_2)|}{|x_1 - x_2|^\delta},$$

and $2\|m\|_{1+\delta} \leq M < \infty$. Then,

$$P\Big(\Big\{(x,y) \to \hat{H}^*(x,y)\Big\} \in \mathcal{H}_1\Big) \to 1,$$

and $P(\{x \to \hat{F}_X^*(x)\} \in \mathcal{H}_2) \to 1$, since $\hat{F}_\varepsilon^*$ and $\hat{F}_X^*$ are monotone and $P(\hat{m} \in C_M^{1+\delta}(R_X)) \to 1$ (see e.g. Akritas and Van Keilegom (2001)). Moreover,

$$\|(\hat{H}^* - H_0, \hat{F}_X^* - F_{X0})\|_{\mathcal{H}}^2 = O_P((nh_n)^{-1}\log n) = o_P(n^{-1/2}).$$

For condition (2.5) note that by Theorem 3 in CLV, it suffices to show that

$$\begin{aligned}&|g_j(x,y,\theta_1,H_1,F_{X1}) - g_j(x,y,\theta_2,H_2,F_{X2})|\\ &\leq b_j(x,y)\Big\{\|\theta_1 - \theta_2\| + \|(H_1 - H_2, F_{X1} - F_{X2})\|_{\mathcal{H}}\Big\}, \qquad \text{(A.4)}\end{aligned}$$

where $g_j(x,y,\theta,H,F_X)$ $(j = 1,\ldots,k)$ is the j-th component of the function g defined in (3.3), and $E[b_j(X,Y)^2] < \infty$, and that

$$\int_0^\infty \sqrt{\log N(\lambda, \mathcal{H}, \|\cdot\|_{\mathcal{H}})}\, d\lambda < \infty, \qquad \text{(A.5)}$$

where $N(\lambda, \mathcal{H}, \|\cdot\|_{\mathcal{H}})$ is the covering number, defined by the minimum number of balls of radius λ (with respect to the norm $\|\cdot\|_{\mathcal{H}}$) needed to cover $\mathcal{H}$. Condition (A.4) follows easily from the assumptions on the copula C. For (A.5), note that for any $\lambda > 0$,

$$\begin{aligned}N(\lambda, \mathcal{H}, \|\cdot\|_{\mathcal{H}}) &\leq N_{[]}(2\lambda, \mathcal{H}, \|\cdot\|_{\mathcal{H}})\\ &\leq N_{[]}(2\lambda, \mathcal{H}_1, \|\cdot\|_{L_2}) \times N_{[]}(2\lambda, \mathcal{H}_2, \|\cdot\|_{L_\infty}),\end{aligned}$$

where the first inequality follows from page 84 in Van der Vaart and Wellner (1996) (VdVW hereafter). Consider first

$$\log N_{[]}(\lambda, \mathcal{H}_2, \|\cdot\|_{L_\infty}) \leq \frac{K}{\lambda},$$

by Theorem 2.7.5 in VdVW. Next, for $\mathcal{H}_1$ we know by Corollary 2.7.2 in VdVW that

$$\log N_{[]}(\lambda^2, C_M^{1+\delta}(R_X), \|\cdot\|_{L_\infty}) \le \frac{K}{\lambda^{2/(1+\delta)}}.$$

Let $g_{m1}^L \le g_{m1}^U, \ldots, g_{ms}^L \le g_{ms}^U$ be $s = O(\exp(K/\lambda^{2/(1+\delta)}))$ λ^2-brackets for the functions $g_m \in C_M^{1+\delta}(R_X)$, and let $h_1^L \le h_1^U, \ldots, h_r^L \le h_r^U$ be $r = O(\exp(K/\lambda))$ monotone λ-brackets for the set of monotone functions h that map onto $[0,1]$. Then, for an arbitrary function $(x,y) \to h(y - g_m(x))$ in $\mathcal{H}_1$, there exist $1 \le j \le s$ and $1 \le \ell \le r$ such that

$$h_\ell^L\Big(y - g_{mj}^U(x)\Big) \le h\Big(y - g_m(x)\Big) \le h_\ell^U\Big(y - g_{mj}^L(x)\Big)$$

for all x, y. Moreover,

$$\begin{aligned}
&E\Big[\Big\{h_\ell^U\Big(Y - g_{mj}^L(X)\Big) - h_\ell^L\Big(Y - g_{mj}^U(X)\Big)\Big\}^2\Big] \\
&\le 2E\Big[\Big\{h_\ell^U\Big(Y - g_{mj}^L(X)\Big) - h_\ell^U\Big(Y - g_{mj}^U(X)\Big)\Big\}^2\Big] \\
&\quad +2E\Big[\Big\{h_\ell^U\Big(Y - g_{mj}^U(X)\Big) - h_\ell^L\Big(Y - g_{mj}^U(X)\Big)\Big\}^2\Big]. \qquad \text{(A.6)}
\end{aligned}$$

The first term above is bounded above by (since $\sup_e h_\ell^U(e) \le 1$)

$$\begin{aligned}
&2E\Big[h_\ell^U\Big(Y - g_{mj}^L(X)\Big) - h_\ell^U\Big(Y - g_{mj}^U(X)\Big)\Big] \\
&\le 2\int\int \Big[h_\ell^U\Big(y - g_{mj}^L(x)\Big) - h_\ell^U\Big(y - g_{mj}^U(x)\Big)\Big] f(y|x)\, dy\, dF_X(x) \\
&= 2\int\int h_\ell^U(z)[f(z + g_{mj}^L(x)|x) - f(z + g_{mj}^U(x)|x)]\, dz\, dF_X(x) \\
&< K\lambda^2,
\end{aligned}$$

for some $K > 0$. Next, consider the second term of (A.6), which can be written as

$$\begin{aligned}
&2\int [h_\ell^U(e) - h_\ell^L(e)]^2 f_{Y - g_{mj}^U(X)}(e)\, de \\
&\le 2\sup_e |h_\ell^U(e) - h_\ell^L(e)|^2 \le K\lambda^2,
\end{aligned}$$

for some $K > 0$. This shows that

$$\int_0^\infty \sqrt{N(\lambda, \mathcal{H}, \|\cdot\|_{\mathcal{H}})}\, d\lambda < \infty.$$

It remains to show condition (2.6) in CLV. Consider

$$\begin{aligned}
&G_n(\theta_0, H_0, F_{X0}) + \Gamma_2(\theta_0, H_0, F_{X0})[\hat{H}^* - H_0, \hat{F}_X^* - F_{X0}] \\
&= n^{-1}\sum_{i=1}^n d_{\theta_0}(F_{\varepsilon 0}(\varepsilon_i), F_X(X_i)) \\
&\quad +E\left[\frac{\partial}{\partial u} d_{\theta_0}(u, F_{X0}(X))\Big|_{u = F_{\varepsilon 0}(\varepsilon)} (\hat{H}^* - H_0)(X, Y)\right. \\
&\quad \left. +\frac{\partial}{\partial v} d_{\theta_0}(H_0(\varepsilon), v)\Big|_{v = F_{X0}(X)} (\hat{F}_X^* - F_{X0})(X)\right]. \qquad \text{(A.7)}
\end{aligned}$$

Using the i.i.d. representation of $\hat{H}^* - H_0$, which can be easily derived from the proof of Theorem 1 in Akritas and Van Keilegom (2001), we have that (where $e = y - m(x)$)

$$(\hat{H}^* - H_0)(x,y) = \hat{F}^*_\varepsilon(y - \hat{m}(x)) - F_{\varepsilon 0}(y - m(x))$$
$$= n^{-1}\sum_{i=1}^n \{I(\varepsilon_i \le e) - F_{\varepsilon 0}(e) + \varphi(X_i, Y_i, e)\} - f_{\varepsilon 0}(e)\{\hat{m}(x) - m(x)\} + o_P(n^{-1/2}),$$

uniformly in x, y. Note that

$$\hat{m}(x) - m(x)$$
$$= -f^{-1}(m(x)|x)\Big[\hat{F}(m(x)|x) - \frac{1}{2}\Big] + O_P((nh)^{-1}\log n)$$
$$= -f^{-1}(m(x)|x)f_X^{-1}(x)\Big[n^{-1}\sum_{i=1}^n K_h(x - X_i)\Big\{I(Y_i \le m(x)) - \frac{1}{2}\Big\}\Big] + O_P((nh)^{-1}\log n),$$

uniformly in x, where $K_h(u) = h^{-1}K(u/h)$. Now define

$$S(x,e) = \frac{\partial}{\partial u} d_{\theta_0}(u, F_{X0}(x))|_{u=F_{\varepsilon 0}(e)} f_{\varepsilon 0}(e) f^{-1}(m(x)|x) f_X^{-1}(x).$$

Then, it can be easily seen that

$$n^{-1}\sum_{i=1}^n E\Big[S(X,\varepsilon)K_h(X - X_i)\Big\{I(Y_i \le m(X)) - \frac{1}{2}\Big\}\Big]$$
$$= n^{-1}\sum_{i=1}^n f_X(X_i)E[S(X,\varepsilon)|X = X_i]\Big\{I(Y_i \le m(X_i)) - \frac{1}{2}\Big\} + o_P(n^{-1/2}).$$

Hence, (A.7) can be written as

$$n^{-1}\sum_{i=1}^n \Bigg(d_{\theta_0}(F_{\varepsilon 0}(\varepsilon_i), F_{X0}(X_i))$$
$$+E\left[\frac{\partial}{\partial u} d_{\theta_0}(u, F_{X0}(X))\Big|_{u=F_{\varepsilon 0}(\varepsilon)}\Big\{I(\varepsilon_i \le \varepsilon) - F_{\varepsilon 0}(\varepsilon) + \varphi(X_i, Y_i, \varepsilon)\Big\}\right.$$
$$\left.+\frac{\partial}{\partial v} d_{\theta_0}(F_{\varepsilon 0}(\varepsilon), v)\Big|_{v=F_{X0}(X)}\Big\{I(X_i \le X) - F_{X0}(X)\Big\}\right]$$
$$+E\left[\frac{\partial}{\partial u} d_{\theta_0}(u, F_{X0}(X))_{u=F_{\varepsilon 0}(\varepsilon)} f_{\varepsilon 0}(\varepsilon)\Big|X = X_i\right] f^{-1}(m(X_i)|X_i)\Big\{I(\varepsilon_i \le 0) - \frac{1}{2}\Big\}\Bigg)$$
$$+o_P(n^{-1/2}).$$

The result now follows.

Proof of Theorem 3.2. Write

$$\hat{F}(y|x) - F(y|x)$$
$$= C^2_{\hat{\theta}}(\hat{F}_\varepsilon(y - \hat{m}(x)), \hat{F}_X(x)) - C^2_{\theta_0}(F_\varepsilon(y - m(x)), F_X(x))$$
$$= C^{2'}_{\xi}(\hat{F}_\varepsilon(y - \hat{m}(x)), \hat{F}_X(x))\,[\hat{\theta} - \theta_0] + C^{22}_{\theta_0}(\hat{F}_\varepsilon(y - \hat{m}(x)), \hat{\eta})\,[\hat{F}_X(x) - F_X(x)]$$
$$+C^{21}_{\theta_0}(F_\varepsilon(y - m(x)), F_X(x))\,[\hat{F}_\varepsilon(y - \hat{m}(x)) - F_\varepsilon(y - m(x))] + O_P((nh)^{-1}\log n),$$

where $\hat{\xi}$ is between θ_0 and $\hat{\theta}$, and $\hat{\eta}$ is between $F_X(x)$ and $\hat{F}_X(x)$. The first and second term above are $O_P(n^{-1/2}) = o_P((nh)^{-1/2})$, whereas the third term can be written as

$$
\begin{aligned}
&\frac{f(y|x)}{f_\varepsilon(y-m(x))}[\hat{F}_\varepsilon(y-\hat{m}(x)) - F_\varepsilon(y-m(x))] \\
&= \frac{f(y|x)}{f_\varepsilon(y-m(x))}\Big\{[\hat{F}_\varepsilon(y-\hat{m}(x)) - F_\varepsilon(y-\hat{m}(x))] + [F_\varepsilon(y-\hat{m}(x)) - F_\varepsilon(y-m(x))]\Big\} \\
&= -f(y|x)[\hat{m}(x) - m(x)] + o_P((nh)^{-1/2}) \\
&= -\frac{f(y|x)}{f(m(x)|x)f_X(x)}\left[(nh)^{-1}\sum_{i=1}^n K\Big(\frac{x-X_i}{h}\Big)\Big\{I(Y_i \le m(x)) - \frac{1}{2}\Big\}\right] + o_P((nh)^{-1/2}),
\end{aligned}
$$

which follows from the fact that $\sup_x |\hat{m}(x) - m(x)| = O_P((nh)^{-1/2}(\log n)^{1/2})$ and $\sup_e |\hat{F}_\varepsilon(e) - F_\varepsilon(e)| = O_P(n^{-1/2}(\log n)^{1/2}) = o_P((nh)^{-1/2})$ (see Theorem 2 in Akritas and Van Keilegom (2001)). The result now follows, since $\mathrm{Var}[(nh)^{1/2}\{\hat{m}(x) - m(x)\}] = \frac{\|K\|_2^2}{4f^2(m(x)|x)} + o(1)$.

Acknowledgement

Both authors gratefully acknowledge financial support from IAP research networks P5/24 and P6/03 of the Belgian Government (Belgian Science Policy). The second author also acknowledges support from ERC grant nr. 203650. Both authors also like to thank Johan Segers for useful discussions that helped to improve the paper.

References

Adang, P.J.M. and Melenberg, B. (1995). Nonnegativity constraints and intratemporal uncertainty in multi-good life-cycle models. *J. Appl. Econometrics*, **10**, 1-15.

Akritas, M.G. and Van Keilegom, I. (2001). Nonparametric estimation of the residual distribution. *Scand. J. Statist.*, **28**, 549–568.

Braekers, R. and Veraverbeke, N. (2005). A copula-graphic estimator for the conditional survival function under dependent censoring. *Canad. J. Statist.*, **33**, 429-447.

Charpentier, A. and Segers, J. (2007). Lower tail dependence for Archimedean copulas : characterizations and pitfalls. *Insurance : Math. Economics*, **40**, 525-532.

Chen, X. and Fan, Y. (2007). A model selection test for bivariate failure-time data. *Econometric Theory*, **23**, 414-439.

Chen, X., Fan, Y. and Tsyrennikov, V. (2006). Efficient estimation of semiparametric multivariate copula models. *J. Amer. Statist. Assoc.*, **101**, 1228-1240.

Chen, X., Linton, O. and Van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica*, **71**, 1591-1608.

Cosma, A., Scaillet, O. and von Sachs, R. (2007). Multivariate wavelet-based shape preserving estimation for dependent observations. *Bernoulli*, **13**, 301-329.

Einmahl, J. and Van Keilegom, I. (2008a). Tests for independence in nonparametric regression. *Statist. Sinica*, **18**, 601-616.

Einmahl, J. and Van Keilegom, I. (2008b). Specification tests in nonparametric regression. *J. Econometrics*, **143**, 88-102.

Fermanian, J.-D. (2005). Goodness-of-fit tests for copulas. *J. Multiv. Anal.*, **95**, 119-152.

Frees, E.W. and Valdez, E.A. (1998). Understanding relationships using copulas. *N. Amer. Actuar. J.*, **2**, 1-25.

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

Genest, C. and Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Pointcaré* (to appear).

Genest, C. and Segers, J. (2008). Rank-based inference for bivariate extreme-value copulas. *Ann. Statist.* (under revision).

Hu, L. (2006). Dependence patterns across financial markets : a mixed copula approach. *Appl. Financ. Economics*, **16**, 717-729.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall.

Nelsen, R.B. (1999). *An Introduction to Copulas*. Springer, New York.

Scaillet, O. (2007). Kernel based goodness-of-fit tests for copulas with fixed smoothing parameters. *J. Multiv. Anal.*, **98**, 533-543.

Stone, C.J. (1977). Consistent nonparametric regression. *Ann. Statist.*, **5**, 595-645.

Tsukahara, H. (2005). Semiparametric estimation in copula models. *Canad. J. Statist.*, **33**, 357-375.

Van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

Wang, W. and Wells, M.T. (2000). Model selection and semiparametric inference for bivariate failure-time data. *J. Amer. Statist. Assoc.*, **95**, 62-72.