

Multivariate and functional covariates and conditional copulas

Irène Gijbels

*Department of Mathematics and Leuven Statistics Research Center (LStat),
Katholieke Universiteit Leuven, Belgium
e-mail: Irene.Gijbels@wis.kuleuven.be*

Marek Omelka

*Department of Probability and Statistics, Faculty of Mathematics and Physics,
Charles University in Prague, Czech Republic
e-mail: omelka@karlin.mff.cuni.cz*

and

Noël Veraverbeke

*Center for Statistics, Hasselt University, Belgium
and
Unit for BMI, North-West University, Potchefstroom, South Africa
e-mail: noel.veraverbeke@uhasselt.be*

Abstract: In this paper the interest is to estimate the dependence between two variables conditionally upon a covariate, through copula modelling. In recent literature nonparametric estimators for conditional copula functions in case of a univariate covariate have been proposed. The aim of this paper is to nonparametrically estimate a conditional copula when the covariate takes on values in more complex spaces. We consider multivariate covariates and functional covariates. We establish weak convergence, and bias and variance properties of the proposed nonparametric estimators. We also briefly discuss nonparametric estimation of conditional association measures such as a conditional Kendall's tau. The case of functional covariates is of particular interest and challenge, both from theoretical as well as practical point of view. For this setting we provide an illustration with a real data example in which the covariates are spectral curves. A simulation study investigating the finite-sample performances of the discussed estimators is provided.

AMS 2000 subject classifications: Primary 62G05, 62H20; secondary 62G20.

Keywords and phrases: Asymptotic representation, empirical copula process, functional covariates, multivariate covariates, small ball probability, random design, smoothing.

Received August 2011.

1. Introduction

Assume that $(\mathcal{X}_1, Y_{11}, Y_{21}), \dots, (\mathcal{X}_n, Y_{1n}, Y_{2n})$ is a sample of n independent and identically distributed triples of random variables. The random variables Y_{1i}

and Y_{2i} are real and the \mathcal{X}_i 's are random elements with values in a space \mathcal{E} that will be specified later.

Researchers are often interested in the dependence structure of the bivariate outcome $(Y_1, Y_2)^\top$ when the value of the covariate is fixed at a given level, say $\mathcal{X} = \chi$. Consider the following example in the food industry. Spectral analysis is used to obtain information on chemical components in a sample of food, for example in meat. Apart from the spectral curves one also measures the percentages of, for example, fat, protein and water. One of the points of interest is to find out how the percentages of fat and protein relate to each other and how strong they are related. The interest in this paper is to discover how this dependency changes with the form of the spectral curve. This is an example of a situation in which \mathcal{X} is of a functional type. See Section 6 for a more detailed description and for the analysis of a real data example using the methodology developed in this paper.

Suppose that the conditional distribution of $(Y_1, Y_2)^\top$ given $\mathcal{X} = \chi$ exists and denote the corresponding conditional joint distribution function by

$$H_\chi(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2 \mid \mathcal{X} = \chi).$$

If the marginals of H_χ denoted as

$$F_{1\chi}(y_1) = P(Y_1 \leq y_1 \mid \mathcal{X} = \chi), \quad F_{2\chi}(y_2) = P(Y_2 \leq y_2 \mid \mathcal{X} = \chi),$$

are continuous, then according to Sklar's theorem (see e.g. Nelsen, 2006 [15]) there exists a unique copula C_χ which equals

$$C_\chi(u_1, u_2) = H_\chi(F_{1\chi}^{-1}(u_1), F_{2\chi}^{-1}(u_2)),$$

where $F_{1\chi}^{-1}(u) = \inf\{y : F_{1\chi}(y) \geq u\}$ is the conditional quantile function of Y_1 given $\mathcal{X} = \chi$ and $F_{2\chi}^{-1}$ is the conditional quantile function of Y_2 given $\mathcal{X} = \chi$. The conditional copula C_χ fully describes the conditional dependence structure of $(Y_1, Y_2)^\top$ given $\mathcal{X} = \chi$.

The estimation of C_χ is the subject of the current research. Gijbels et al. (2011) [12] and Veraverbeke et al. (2011) [19] investigated nonparametric estimation of C_χ when the covariate is real, that is $\mathcal{E} = \mathbb{R}$. Semiparametric estimation of conditional copulas in this case has been studied in Hafner and Reznikova (2010) [13], Acar et al. (2011) [2] and Abegaz et al. (2012) [1]. In this paper we introduce a nonparametric estimator of C_χ when the covariate space is more complex.

The estimators are of a similar type as these in Gijbels et al. (2011) [12], but their theoretical study and practical use can be quite different depending on the complexity of the covariate space. See Sections 3–6. An estimator of the joint conditional distribution function H_χ is

$$H_{\chi h}(y_1, y_2) = \sum_{i=1}^n w_{ni}(\chi, h_n) \mathbb{I}\{Y_{1i} \leq y_1, Y_{2i} \leq y_2\},$$

where $\{w_{ni}(\chi, h_n)\}$ is a sequence of weights that smooth over the covariate space \mathcal{E} and $h = \{h_n > 0\}$ is a bandwidth sequence tending to zero as the

sample size increases. Then analogously as in Gijbels et al. (2011) [12] one can suggest the following empirical estimator of the copula C_X

$$C_{Xh}(u_1, u_2) = H_{Xh} \left(F_{1Xh}^{-1}(u_1), F_{2Xh}^{-1}(u_2) \right), \quad 0 \leq u_1, u_2 \leq 1, \quad (1)$$

where F_{1Xh} and F_{2Xh} are the corresponding marginal distribution functions of H_{Xh} , i.e., $F_{1Xh}(y_1) = H_{Xh}(y_1, +\infty)$ and $F_{2Xh}(y_2) = H_{Xh}(+\infty, y_2)$. As demonstrated by Gijbels et al. (2011) [12] and Veraverbeke et al. (2011) [19] it is often advisable to remove the influence of the covariate on the marginal distributions before the estimation of C_X . This can be done in various ways. One can for instance assume a regression model linking the covariate X with the response Y_1 (Y_2) and then replace the original observations (Y_{1i}, Y_{2i}) with the estimated residuals.

Gijbels et al. (2011) [12] suggested a very general way of removing the influence of the covariate on the marginal distributions which can be described as follows. First, estimate the unobserved marginally uniform observations $(U_{1i}, U_{2i})^T = (F_{1X_i}(Y_{1i}), F_{2X_i}(Y_{2i}))^T$ with

$$(\tilde{U}_{1i}, \tilde{U}_{2i})^T = (F_{1X_i g_1}(Y_{1i}), F_{2X_i g_2}(Y_{2i}))^T, \quad i = 1, \dots, n, \quad (2)$$

where $g_1 = \{g_{1n}\} \searrow 0$ and $g_2 = \{g_{2n}\} \searrow 0$. Second, use the transformed observations $(\tilde{U}_{1i}, \tilde{U}_{2i})^T$ in a similar way as the original observations and construct

$$\tilde{C}_{Xh}(u_1, u_2) = \tilde{G}_{Xh} \left(\tilde{G}_{1Xh}^{-1}(u_1), \tilde{G}_{2Xh}^{-1}(u_2) \right), \quad (3)$$

where

$$\tilde{G}_{Xh}(u_1, u_2) = \sum_{i=1}^n w_{ni}(\chi, h_n) \mathbb{I}\{\tilde{U}_{1i} \leq u_1, \tilde{U}_{2i} \leq u_2\},$$

and \tilde{G}_{1Xh} and \tilde{G}_{2Xh} are the corresponding marginals: $\tilde{G}_{1Xh}(u_1) = \tilde{G}_{Xh}(u_1, 1)$ and $\tilde{G}_{2Xh}(u_2) = \tilde{G}_{Xh}(1, u_2)$.

In case of a univariate real-valued covariate a thorough study comparing the performances of the two types of estimators C_{Xh} and \tilde{C}_{Xh} , defined in respectively (1) and (3), can be found in Gijbels et al. (2011) [12] revealing the following practical recommendation: the estimator C_{Xh} is generally preferable when the covariate does not influence the marginal distributions; in the opposite situation, it is safer to use the estimator \tilde{C}_{Xh} . None of the two estimators is however uniformly (in all situations) outperforming the other one. Moreover, the effect of the covariate on the marginal distributions and/or on the conditional copula itself is often unknown. It is therefore worthwhile to study both estimators.

In this paper we establish weak convergence results of the processes associated with the estimators C_{Xh} and \tilde{C}_{Xh} , for a multivariate covariate (in Section 2) and for a functional covariate (in Section 3). Nonparametric estimators of conditional association measures are very briefly discussed in Section 4. In a simulation study in Section 5 the finite-sample performances of the estimators, in both the multivariate and functional covariate case, are investigated. In Section 6 the

discussed methods are used to analyze a real dataset with a functional covariate. Some further discussions are provided in Section 7. The proofs and technical details are given in the Appendix.

2. Multivariate covariate – $\mathcal{E} = \mathbb{R}^d$

Assume that $(\mathbf{X}_i, Y_{1i}, Y_{2i})$ is a sample of i.i.d. triples, where $\mathbf{X}_i = (X_{i1}, \dots, X_{id})^\top$ is a d -dimensional continuous covariate with a density $f_{\mathbf{X}}$. Analogously as in Veraverbeke et al. (2011) [19] one can consider also a fixed design, but then one should think of $f_{\mathbf{X}}$ as a design density.

As $\mathcal{E} = \mathbb{R}^d$ the role of the χ is now played by $\mathbf{x} = (x_1, \dots, x_d)^\top$. A common system of weights $\{w_{ni}\}$ is based on the quantity $|\mathbf{B}|^{-1/2} K_{\mathbf{B}}(\mathbf{X}_i - \mathbf{x})$, where K is a d -variate kernel, \mathbf{B} is the bandwidth matrix with determinant $|\mathbf{B}|$, and $K_{\mathbf{B}}(\mathbf{y}) = K(\mathbf{B}^{-1/2}\mathbf{y})$. For simplicity of presentation we will suppose that $\mathbf{B} = h_n^2 \mathbf{I}_d$, where \mathbf{I}_d is the d -dimensional identity matrix. Then for example the Nadaraya-Watson weights are defined as

$$w_{ni}(\mathbf{x}, h_n) = \frac{K_{\mathbf{B}}(\mathbf{X}_i - \mathbf{x})}{\sum_{j=1}^n K_{\mathbf{B}}(\mathbf{X}_j - \mathbf{x})}, \quad i = 1, \dots, n. \quad (4)$$

The definition of local linear weights can be found in e.g. Ruppert and Wand (1994) [16].

In this paper we consider only pointwise convergence. Therefore, let $\chi = \mathbf{x}$ be a fixed element in $\mathcal{E} = \mathbb{R}^d$, and define the empirical copula processes

$$\mathbb{C}_{\mathbf{x}n}^{(\mathcal{E})} = \sqrt{nh_n^d} (C_{\mathbf{x}h} - C_{\mathbf{x}}), \quad \tilde{\mathbb{C}}_{\mathbf{x}n}^{(\mathcal{E})} = \sqrt{nh_n^d} (\tilde{C}_{\mathbf{x}h} - C_{\mathbf{x}}). \quad (5)$$

Denote by $C_{\mathbf{x}}^{(1)}$, $C_{\mathbf{x}}^{(2)}$ the partial derivatives of the copula function $C_{\mathbf{x}}$ with respect to u_1 and u_2 respectively. In order to establish the weak convergence results for the processes in (5) we need some regularity conditions, which we discuss first.

2.1. Regularity assumptions

Regularity conditions needed for the following theorems can be formulated in an analogous way as in Veraverbeke et al. (2011) [19]. Technically speaking, the extension from the univariate to the multivariate covariate case is rather straightforward, and therefore we only very briefly discuss this case. Generally speaking, one has to replace h_n with h_n^d and take into consideration that the covariate is now a vector. The complete list of conditions is to be found in Appendix A.

Similarly as it was argued in Veraverbeke et al. (2011) [19] the assumptions about the weights are satisfied for the Nadaraya-Watson weighting scheme provided that $f_{\mathbf{X}}(\mathbf{x}) > 0$ and $\frac{\partial f_{\mathbf{X}}(\mathbf{z})}{\partial \mathbf{z}}$ is continuous in the neighbourhood of the point \mathbf{x} . For local linear weights the latter assumption can even be omitted.

When dealing with the $d > 1$ covariate case however, smoothing operations are needed over the d -dimensional space. This implies that the estimation procedure suffers from the curse of dimensionality, as any nonparametric smoothing technique would do in this case. For a further discussion, see Section 7.

2.2. Results for the estimator C_{xh} of type (1)

Suppose that

$$h_n = O(n^{-1/(4+d)}), \quad n h_n^d \rightarrow \infty. \tag{6}$$

Theorem 1. *Assume (6), (R1)-(R2) and (W1). Then the following linear asymptotic representation*

$$\mathbb{C}_{\mathbf{x}n}^{(E)}(u_1, u_2) = \sqrt{n h_n^d} \sum_{i=1}^n w_{ni}(\mathbf{x}, h_n) \xi_i(u_1, u_2) + o_P(1), \tag{7}$$

holds uniformly in $(u_1, u_2) \in [0, 1]^2$, where

$$\begin{aligned} \xi_i(u_1, u_2) &= \mathbb{I}\{Y_{1i} \leq F_{1\mathbf{x}}^{-1}(u_1), Y_{2i} \leq F_{2\mathbf{x}}^{-1}(u_2)\} - C_{\mathbf{x}}(u_1, u_2) \\ &\quad - C_{\mathbf{x}}^{(1)}(u_1, u_2) [\mathbb{I}\{Y_{1i} \leq F_{1\mathbf{x}}^{-1}(u_1)\} - u_1] \\ &\quad - C_{\mathbf{x}}^{(2)}(u_1, u_2) [\mathbb{I}\{Y_{2i} \leq F_{2\mathbf{x}}^{-1}(u_2)\} - u_2]. \end{aligned} \tag{8}$$

With the help of the asymptotic representation (7) it is usually possible to describe the limiting process of $\mathbb{C}_{\mathbf{x}n}^{(E)}$.

The choice of the weights has some impact on the variance-covariance structure of the limiting process, through the asymptotic behaviour of the quantity $n h_n^d \sum_{i=1}^n w_{ni}^2(\mathbf{x}, h_n)$. Thanks to assumptions (W1) there typically exists, in commonly-used weight systems, a finite positive constant V such that

$$n h_n^d \sum_{i=1}^n w_{ni}^2(\mathbf{x}, h_n) = V^2 + o_P(1). \tag{9}$$

For example, for Nadaraya-Watson weights defined in (4), $V^2 = \mu_K / f_{\mathbf{x}}(\mathbf{x})$, where $\mu_K = \int_{\mathbb{R}^d} K^2(\mathbf{u}) d\mathbf{u}$.

In order to derive the asymptotic bias of the normalized empirical process, we need to study the expectation of the leading term in the linear asymptotic representation in (7). Suppose that there exists H such that $n h_n^{4+d} \rightarrow H^2$, with $H \geq 0$, using Taylor expansion and assumption (R1) one can find that (uniformly in (u_1, u_2))

$$\begin{aligned} n h_n^d \sum_{i=1}^n w_{ni}(\mathbf{x}, h_n) \mathbb{E} \xi_i(u_1, u_2) \\ = H \left[\mathbf{D}_K^T \dot{C}_{\mathbf{x}}(u_1, u_2) + \text{tr} \left\{ \frac{\mathbf{E}_K}{2} B_{\mathbf{x}}(u_1, u_2) \right\} \right] + o_P(1), \end{aligned} \tag{10}$$

with \mathbf{D}_K being a vector and \mathbf{E}_K being a matrix of constants depending on the chosen system of weights $\{w_{ni}\}$ and on the type of the design (see the Assumptions **(W3)** in the Appendix) and

$$\begin{aligned} B_{\mathbf{x}}(u_1, u_2) &= \ddot{H}_{\mathbf{x}}(F_{1\mathbf{x}}^{-1}(u_1), F_{2\mathbf{x}}^{-1}(u_2)) - C_{\mathbf{x}}^{(1)}(u_1, u_2) \ddot{F}_{1\mathbf{x}}(F_{1\mathbf{x}}^{-1}(u_1)) \\ &\quad - C_{\mathbf{x}}^{(2)}(u_1, u_2) \ddot{F}_{2\mathbf{x}}(F_{2\mathbf{x}}^{-1}(u_2)) \\ &= \ddot{C}_{\mathbf{x}}(u_1, u_2) + 2\dot{C}_{\mathbf{x}}^{(1)}(u_1, u_2) \dot{F}_{1\mathbf{x}}(F_{1\mathbf{x}}^{-1}(u_1))^{\top} \\ &\quad + 2\dot{C}_{\mathbf{x}}^{(2)}(u_1, u_2) \dot{F}_{2\mathbf{x}}(F_{2\mathbf{x}}^{-1}(u_2))^{\top} \\ &\quad + C_{\mathbf{x}}^{(1,1)}(u_1, u_2) \dot{F}_{1\mathbf{x}}(F_{1\mathbf{x}}^{-1}(u_1)) \dot{F}_{1\mathbf{x}}(F_{1\mathbf{x}}^{-1}(u_1))^{\top} \\ &\quad + C_{\mathbf{x}}^{(2,2)}(u_1, u_2) \dot{F}_{2\mathbf{x}}(F_{2\mathbf{x}}^{-1}(u_2)) \dot{F}_{2\mathbf{x}}(F_{2\mathbf{x}}^{-1}(u_2))^{\top} \\ &\quad + C_{\mathbf{x}}^{(1,2)}(u_1, u_2) \dot{F}_{1\mathbf{x}}(F_{1\mathbf{x}}^{-1}(u_1)) \dot{F}_{2\mathbf{x}}(F_{2\mathbf{x}}^{-1}(u_2))^{\top} \\ &\quad + C_{\mathbf{x}}^{(2,1)}(u_1, u_2) \dot{F}_{2\mathbf{x}}(F_{2\mathbf{x}}^{-1}(u_2)) \dot{F}_{1\mathbf{x}}(F_{1\mathbf{x}}^{-1}(u_1))^{\top}, \end{aligned} \quad (11)$$

where a dot indicates a derivative with respect to the covariate \mathbf{x} , e.g. $\dot{F}_{\mathbf{z}}(u_1) = \frac{\partial}{\partial \mathbf{z}} F_{\mathbf{z}}(u_1)$, $\ddot{C}_{\mathbf{z}}(u_1, u_2) = \frac{\partial^2}{\partial \mathbf{z} \partial \mathbf{z}^{\top}} C_{\mathbf{z}}(u_1, u_2)$; the symbol $^{(i)}$ indicates a derivative with respect to u_i , e.g. $C_{\mathbf{x}}^{(i,j)}(u_1, u_2) = \frac{\partial^2 C_{\mathbf{x}}(u_1, u_2)}{\partial u_i \partial u_j}$; and $\dot{C}_{\mathbf{z}}^{(i)}(u_1, u_2) = \frac{\partial^2 C_{\mathbf{z}}(u_1, u_2)}{\partial \mathbf{z} \partial u_i}$, which is a mixture of the above notational rules. Finally, the product $\mathbf{E}_K B_{\mathbf{x}}$ in (10) is a matrix product.

We are now ready to state the weak convergence result for the process $\mathbb{C}_{\mathbf{x}n}^{(\varepsilon)}$ in (5).

Corollary 1. *If (9), $nh_n^{4+d} \rightarrow H^2$, **(W3)** and the assumptions of Theorem 1 hold, then the process $\mathbb{C}_{\mathbf{x}n}^{(\varepsilon)}$ converges in distribution to a Gaussian process $Z_{\mathbf{x}}$*

$$\begin{aligned} Z_{\mathbf{x}}(u_1, u_2) &= V \left\{ W_{\mathbf{x}}(u_1, u_2) - C_{\mathbf{x}}^{(1)}(u_1, u_2) W_{\mathbf{x}}(u_1, 1) - C_{\mathbf{x}}^{(2)}(u_1, u_2) W_{\mathbf{x}}(1, u_2) \right\} \\ &\quad + R_{\mathbf{x}}(u_1, u_2), \end{aligned} \quad (12)$$

where V is a constant depending on the asymptotic properties of the weights $\{w_{ni}\}$, as defined in (9), $W_{\mathbf{x}}$ is a bivariate Brownian bridge on $[0, 1]^2$ with covariance function

$$\mathbb{E}[W_{\mathbf{x}}(u_1, u_2) W_{\mathbf{x}}(v_1, v_2)] = C_{\mathbf{x}}(u_1 \wedge v_1, u_2 \wedge v_2) - C_{\mathbf{x}}(u_1, u_2) C_{\mathbf{x}}(v_1, v_2),$$

and where $R_{\mathbf{x}}(u_1, u_2)$ is the (deterministic) mean function

$$R_{\mathbf{x}}(u_1, u_2) = H \left[\mathbf{D}_K^{\top} \dot{C}_{\mathbf{x}}(u_1, u_2) + \text{tr} \left\{ \frac{\mathbf{E}_K}{2} B_{\mathbf{x}}(u_1, u_2) \right\} \right]. \quad (13)$$

2.3. Results for the estimator \tilde{C}_{xh} of type (3)

For this estimator we need to specify the relation of the three bandwidths that are used. In the following we suppose that for $j = 1, 2$

$$\sqrt{nh_n^d g_{jn}^2} = O(1), \quad \frac{h_n}{g_{jn}} = O(1), \quad n^{1/d} \min(h_n, g_{1n}, g_{2n}) \rightarrow \infty. \quad (14)$$

The next theorem establishes a linear asymptotic representation for the normalized empirical copula process in (5) and subsequently the weak convergence result for $\tilde{C}_{\mathbf{x}n}^{(E)}$.

Theorem 2. Assume (14), (W1)–(W3) and ($\tilde{R}1$)–($\tilde{R}3$), then uniformly in $(u_1, u_2) \in [0, 1]^2$

$$\tilde{C}_{\mathbf{x}n}^{(E)}(u_1, u_2) = \sqrt{nh_n^d} \sum_{i=1}^n w_{ni}(\mathbf{x}, h_n) \tilde{\xi}_i(u_1, u_2) + o_P(1),$$

where

$$\begin{aligned} \tilde{\xi}_i(u_1, u_2) &= \mathbb{I}\{U_{1i} \leq u_1, U_{2i} \leq u_2\} - C_{\mathbf{x}}(u_1, u_2) \\ &\quad - C_{\mathbf{x}}^{(1)}(u_1, u_2) [\mathbb{I}\{U_{1i} \leq u_1\} - u_1] \\ &\quad - C_{\mathbf{x}}^{(2)}(u_1, u_2) [\mathbb{I}\{U_{2i} \leq u_2\} - u_2] \end{aligned} \tag{15}$$

and $(U_{1i}, U_{2i}) = (F_{1\mathbf{x}_i}(Y_{1i}), F_{2\mathbf{x}_i}(Y_{2i}))$.

Moreover, if (9) holds and $(nh_n^{4+d}) \rightarrow H^2$, then the process $\tilde{C}_{\mathbf{x}n}^{(E)}$ also converges in distribution to a Gaussian process $Z_{\mathbf{x}}$, as in (12), but now with the (deterministic) mean function equal to

$$R_{\mathbf{x}}(u_1, u_2) = H \left[\mathbf{D}_K^T \dot{C}_{\mathbf{x}}(u_1, u_2) + \text{tr} \left\{ \frac{\mathbf{E}_K}{2} \ddot{C}_{\mathbf{x}}(u_1, u_2) \right\} \right]. \tag{16}$$

Similarly as in Veraverbeke et al. (2011) [19] one can see from Corollary 1 and Theorem 2 that both estimators of the conditional copula have the same asymptotic variances (provided the same bandwidth h_n is used), but that the bias of the estimator $\tilde{C}_{\mathbf{x}h}$ does not involve terms coming from the dependence of the marginal distributions on the covariate. See expressions (11), (12), (13) and (16).

3. Functional covariate \mathcal{X}

In this section we study the weak convergence of the, properly normalized, empirical copula processes in case of a functional covariate.

Assume now that $(\mathcal{X}_i, Y_{1i}, Y_{2i})$ is a sample of i.i.d. triples, where the \mathcal{X}_i s are random elements with values in a functional space \mathcal{E} , equipped with a semi-metric d . Very commonly $\mathcal{E} = \mathcal{L}_2([a, b])$, where $\mathcal{L}_2([a, b])$ is the class of square-integrable functions on the interval $[a, b]$ with $d(\chi, \chi') = \sqrt{\int_a^b (\chi(t) - \chi'(t))^2 dt}$. In the following we will take \mathcal{E} to be a separable Banach space endowed with a norm $\|\cdot\|$. As argued in Ferraty et al. (2007) [10] the space is still very general and the separability avoids measurability problems.

Also here we first introduce some notations and state some primary regularity conditions.

3.1. Notations and regularity conditions

Let us consider the Nadaraya-Watson weights

$$w_{ni}(\chi, h_n) = \frac{K\left(\frac{\|\mathcal{X}_i - \chi\|}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{\|\mathcal{X}_j - \chi\|}{h_n}\right)}, \quad i = 1, \dots, n, \quad (17)$$

where K is a given (univariate) kernel. Another system of weights that naturally extends to the functional covariate case is that of the k -nearest neighbour weights; see Burba et al. (2009) [7]. On the other hand there is no unique generalization of local linear weights to the functional covariate case. For possible suggestions see e.g. Baíllo and Grané (2009) [3], Barrientos-Marin et al. (2010) [4] and Berlinet et al. (2011) [5]. For simplicity we concentrate only on Nadaraya-Watson weights in the following, but with some further effort it might be shown that the results also hold for k -nearest neighbour weights as defined in Burba et al. (2009) [7].

Analogously as in Ferraty et al. (2007) [10] define

$$\phi_{\mathcal{X},y}^{F_j}(s) = \mathbb{E}_{\mathcal{X}} [F_{j\mathcal{X}}(y) - F_{j\chi}(y) \mid \|\mathcal{X} - \chi\| = s],$$

Note that the function $\phi_{\mathcal{X},y}^{F_j}$ quantifies the expected difference $F_{j\mathcal{X}}(y) - F_{j\chi}(y)$ when \mathcal{X} is forced to be at a distance s from the point χ .

In an analogous way define

$$\phi_{\mathcal{X},y_1,y_2}^H(s) = \mathbb{E}_{\mathcal{X}} [H_{\mathcal{X}}(y_1, y_2) - H_{\chi}(y_1, y_2) \mid \|\mathcal{X} - \chi\| = s],$$

and similarly for $\phi_{\mathcal{X},u_1,u_2}^C$.

A very important role is played by the so called *small ball probability function*

$$\varphi_{\mathcal{X}}(h) = \mathbb{P}(\|\mathcal{X} - \chi\| < h).$$

Further, for $s \in [0, 1]$ put

$$\tau_{\mathcal{X},h}(s) = \frac{\varphi_{\mathcal{X}}(hs)}{\varphi_{\mathcal{X}}(h)} = \mathbb{P}(\|\mathcal{X} - \chi\| < sh \mid \|\mathcal{X} - \chi\| < h). \quad (18)$$

In this functional covariate case the appropriate normalization factor for the empirical processes turns out to be $\sqrt{n\varphi_{\mathcal{X}}(h_n)}$. We consider the normalized empirical processes

$$\mathbb{C}_{\mathcal{X}^n}^{(\mathbb{E})} = \sqrt{n\varphi_{\mathcal{X}}(h_n)} (C_{\mathcal{X}h} - C_{\mathcal{X}}), \quad \tilde{\mathbb{C}}_{\mathcal{X}^n}^{(\mathbb{E})} = \sqrt{n\varphi_{\mathcal{X}}(h_n)} (\tilde{C}_{\mathcal{X}h} - C_{\mathcal{X}}), \quad (19)$$

which is in analogy with (5). Indeed, note that this normalization factor is consistent with that in Section 2 as for $\mathcal{E} = \mathbb{R}^d$ one gets (see e.g. Ferraty and Vieu, 2002 [9], Chapter 13).

$$\varphi_{\mathbf{x}}(h) = \frac{\pi^{d/2} h^d}{\Gamma(\frac{d}{2} + 1)} f_X(\mathbf{x}) + o(h^d), \quad \text{where } \Gamma(\cdot) \text{ is the gamma function,}$$

and hence the order of the normalization factor $\sqrt{n\varphi_{\mathcal{X}}(h_n)}$ in (19) coincides with that of $\sqrt{nh_n^d}$ in (5).

We now state the weak convergence results for the processes in (19).

3.2. Results for the estimator $C_{\chi h}$ of type (1)

The following theorem is proved in Appendix B.

Theorem 3. Assume **(F0)**–**(F4)** and **(R2)** of Appendix B. Then $\mathbb{C}_{\chi n}^{(\mathbb{E})}$ has a linear asymptotic representation

$$\mathbb{C}_{\chi n}^{(\mathbb{E})}(u_1, u_2) = \sqrt{n \varphi_{\chi}(h_n)} \sum_{i=1}^n w_{ni}(\chi, h_n) \xi_i(u_1, u_2) + o_P(1),$$

with $\xi_i(u_1, u_2)$ as defined in (8) (with χ replacing \mathbf{x}).

Moreover if $n\varphi_{\chi}(h_n)h_n^{2\beta} \rightarrow H^2$, then the process $\mathbb{C}_{\chi n}^{(\mathbb{E})}$, defined in (19), converges in distribution to a Gaussian process Z_{χ} , as defined in (12) but now with $V = \sqrt{M_2}/M_1$ and the mean function R_{χ} given by

$$\begin{aligned} R_{\chi}(u_1, u_2) &= \frac{\widetilde{M}_{\beta}}{M_1} \left(\vartheta_H(F_{1\chi}^{-1}(u_1), F_{2\chi}^{-1}(u_2)) - C_{\chi}^{(1)}(u_1, u_2) \vartheta_H(F_{1\chi}^{-1}(u_1), \infty) \right. \\ &\quad \left. - C_{\chi}^{(2)}(u_1, u_2) \vartheta_H(\infty, F_{2\chi}^{-1}(u_2)) \right). \end{aligned} \quad (20)$$

The quantities M_1, M_2 and \widetilde{M}_{β} are defined in respectively (B2), (B6) and (B5) in Appendix B.

If $n\varphi_{\chi}(h_n)h_n^{2\beta} \rightarrow 0$ then the asymptotic bias given by the function R_{χ} diminishes. On the other hand if $n\varphi_{\chi}(h_n)h_n^{2\beta} \rightarrow \infty$, then the bias dominates the variance. Thus the assumption $n\varphi_{\chi}(h_n)h_n^{2\beta} = O(1)$ gives the optimal rate of convergence. But it should be said that the practical impact of such a result appears as rather limited given the current state-of-the-art of the research area. The general problem is that a random element with values in a functional space is generally a very complex object and it is not clear yet what are reasonable assumptions about its distribution.

To be more explicit, it is already well-understood that the variability of kernel estimators with functional covariate depends on the behaviour of the small ball probability function $\varphi_{\chi}(h)$ near zero. For many standard processes, see e.g. Ferraty and Vieu (2006) [11], Ferraty et al. (2007) [10], and Hall et al. (2009) [14], one gets $\varphi_{\chi}(h) = o(h^d)$, where d might be arbitrarily high. Roughly speaking this means that such a process is near χ more sparse than a continuous multivariate covariate of arbitrarily high dimension near a point with a positive density. This somehow violates the philosophy of functional data analysis, which states that switching from multivariate to functional variables helps to deal with the curse of dimensionality. In fact the small ball probability function $\varphi_{\chi}(h)$ is directly linked with the concentration properties of a functional variable \mathcal{X} . Obviously this small ball probability depends on the choice of the semi-metric (and consequently norm). Appropriate choices of semi-metrics can thus lead to an increase in the concentration properties of the functional variable \mathcal{X} . The co-called *projection type semi-metrics* are of this type, and a general procedure for construction can be found in Lemma 13.6 of Ferraty and Vieu (2006) [11].

Next, the lack of sense what is reasonable to assume about the distribution of \mathcal{X} also affects our ability to check assumption **(F3)** in Appendix B about the behaviour of the function $\phi_{\mathcal{X}, y_1, y_2}^H(s)$ near zero. Note that if $\mathcal{E} = \mathbb{R}$ then

$$\begin{aligned}\phi_{\mathcal{X}, y_1, y_2}^H(s) &\sim \frac{1}{2} (H_{x-s}(y_1, y_2) - H_x(y_1, y_2)) + \frac{1}{2} (H_{x+s}(y_1, y_2) - H_x(y_1, y_2)) \\ &\sim \frac{s^2}{2} \ddot{H}_x(y_1, y_2),\end{aligned}$$

thus one would expect that $\beta = 2$ is a reasonable assumption. On the other hand in nonparametric regression with a functional covariate Ferraty et al. (2007) [10] and (2010) [8] explicitly assume (when translating to the context of estimation of distribution functions) that $\partial \phi_{\mathcal{X}, y_1, y_2}^H(s) / \partial s$ at $s = 0$ is non-vanishing and finite, which corresponds to the case $\beta = 1$.

As it is usually rather difficult to find the highest possible β in **(F3)** even if the model is specified, researchers often assume Lipschitz continuity in the covariate of the quantity of interest. Let us suppose that there exists $\gamma > 0$ such that

$$\exists C < \infty \forall \mathcal{X}' \in \mathcal{E} \quad \sup_{y_1, y_2} |H_{\mathcal{X}'}(y_1, y_2) - H_{\mathcal{X}}(y_1, y_2)| \leq C [d(\mathcal{X}, \mathcal{X}')]^\gamma.$$

Then we can arrive at Theorem 3 with $R_{\mathcal{X}}(u_1, u_2) = 0$ provided that

$$n \varphi_{\mathcal{X}}(h_n) h_n^{2\gamma} \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

3.3. Results for the estimator $\tilde{C}_{\mathcal{X}h}$ of type (3)

Similarly as in the case of a univariate covariate, it seems to be advisable to try to remove the effect of the covariate on the marginal distributions.

To guarantee an analogous result as stated in Theorem 2 we need regularity assumptions, listed in Appendix B. Most of them just guarantee that the assumptions made in the previous section hold in some sense uniformly on a neighbourhood of the point of interest.

Note that assumption **(F5)** in Appendix B is specific for a functional covariate, as for a univariate or multivariate covariate it is satisfied. Somebody may find this assumption rather restrictive as the unit ball in a Banach space is totally bounded if and only if the space has a finite dimension. But on the other hand, there are many interesting sets of functions that satisfy this assumption. For instance all the sets of functions for which we can find finite covering or bracketing numbers (see e.g. van der Vaart and Wellner (1996) [18]) are totally bounded. Note that a condition similar to **(F5)** is also assumed in Ferraty et al. (2010) [8]. Their assumption (C7) (together with (C6)) implies that for a given $\varepsilon > 0$ for all sufficiently large n there exists a finite cover such that the diameter of each element of the cover is at most ε for each of the sets $B(\mathcal{X}/h_n, 1) = \{\mathcal{X}' \in \mathcal{E}, \|\mathcal{X}' - \mathcal{X}/h_n\| \leq 1\}$.

The next theorem states the weak convergence result for the second normalized empirical process in (19).

Theorem 4. Assume **(F0)**, **(F1)**–**(F5)** and **(R2)** of Appendix B. Then the linear asymptotic representation

$$\tilde{C}_{\mathcal{X}^n}^{(\varepsilon)}(u_1, u_2) = \sqrt{n \varphi_{\mathcal{X}}(h_n)} \sum_{i=1}^n w_{ni}(\mathcal{X}, h_n) \tilde{\xi}_i(u_1, u_2) + o_P(1),$$

holds, with $\tilde{\xi}_i$ as in (15) (with \mathcal{X} replacing \mathbf{x}).

If moreover $n\varphi_{\mathcal{X}}(h_n)h_n^{2\beta} \rightarrow H^2$, then the process $\tilde{C}_{\mathcal{X}^n}^{(\varepsilon)}$, defined in (19), converges in distribution to a Gaussian process $Z_{\mathcal{X}}$, as defined in (12) with $V = \sqrt{M_2}/M_1$ and the drift function $R_{\mathcal{X}}$ given by

$$R_{\mathcal{X}}(u_1, u_2) = \frac{\tilde{M}_{\beta}}{M_1} \vartheta_C(u_1, u_2).$$

Remark 1. Comparing Theorems 3 and Theorems 4 one can see that asymptotic variances of both estimators are the same. The comparison of the biases is less straightforward here than for a univariate or multivariate covariate. To get some insight into the problem, recall that $C_{\mathcal{X}}^{(j)} = \partial C_{\mathcal{X}}/\partial u_j$ and $C_{\mathcal{X}}^{(i,j)} = \partial^2 C_{\mathcal{X}}/\partial u_i \partial u_j$. Further suppose that for $j = 1, 2$ uniformly in (u_1, u_2)

$$\begin{aligned} \mathbb{E} \left[\left(C_{\mathcal{X}}^{(j)}(u_1, u_2) - C_{\mathcal{X}}^{(j)}(u_1, u_2) \right) (F_{j\mathcal{X}}(F_{j\mathcal{X}}^{-1}(u_j)) - u_j) \mid \|\mathcal{X} - \chi\| = s \right] \\ = s^{\beta} \vartheta_{C^{(j)} F_j}(u_1, u_2) + o(s^{\beta}), \end{aligned}$$

$$\begin{aligned} \mathbb{E} \left[C_{\mathcal{X}}^{(j,j)}(u_1, u_2) (F_{j\mathcal{X}}(F_{j\mathcal{X}}^{-1}(u_j)) - u_j)^2 \mid \|\mathcal{X} - \chi\| = s \right] \\ = s^{\beta} \vartheta_{C^{(j,j)} F_j^2}(u_1, u_2) + o(s^{\beta}), \end{aligned}$$

and finally

$$\begin{aligned} \mathbb{E} \left[C_{\mathcal{X}}^{(1,2)}(u_1, u_2) (F_{1\mathcal{X}}(F_{1\mathcal{X}}^{-1}(u_1)) - u_1) (F_{2\mathcal{X}}(F_{2\mathcal{X}}^{-1}(u_2)) - u_2) \mid \|\mathcal{X} - \chi\| = s \right] \\ = s^{\beta} \vartheta_{C^{(1,2)} F_1 F_2}(u_1, u_2) + o(s^{\beta}), \end{aligned}$$

where the functions $\vartheta_{C^{(j)} F_j}(u_1, u_2)$, $\vartheta_{C^{(j,j)} F_j^2}$ and $\vartheta_{C^{(1,2)} F_1 F_2}$ are continuous.

Then one can derive that the function $R_{\mathcal{X}}$ of (20) can be expressed as

$$R_{\mathcal{X}} = \frac{\tilde{M}_{\beta}}{M_1} \left(\vartheta_C + \vartheta_{C^{(1)} F_1} + \vartheta_{C^{(2)} F_2} + \frac{1}{2} \vartheta_{C^{(1,1)} F_1^2} + \vartheta_{C^{(1,2)} F_1 F_2} + \frac{1}{2} \vartheta_{C^{(2,2)} F_2^2} \right).$$

Thus similarly as for a univariate or multivariate covariate one can see that the bias structure of the estimator $\tilde{C}_{\mathcal{X}^n}$ is much simpler as it does not include the several terms coming from the effect of the covariate on the marginal distributions. That is why we generally recommend trying to reduce the effect of the covariate on the marginal distributions.

Remark 2. We can get rid of assumption $(\tilde{\mathbf{F}}4)$ if we slightly undersmooth. More precisely instead of $(\tilde{\mathbf{F}}4)$ suppose that:

$(\tilde{\mathbf{F}}4')$ There exists $\eta > 0$ such that the bandwidths h_n , g_{n1} and g_{n2} satisfy

$$\frac{\varphi_X(h_n)(\log n)^{1+\eta}}{\varphi_X(g_{jn})} \rightarrow 0, \quad \frac{n \varphi_X(g_{jn}) g_{nj}^{2\beta}}{(\log n)^{1+\eta}} = O(1), \quad n \varphi_X(h_n) \rightarrow \infty. \quad (21)$$

Then it can be proved that Theorem 4 holds with $R_X \equiv 0$.

Note however that (21) implies $n \varphi_X(h_n) h_n^{2\beta} = o(1)$. Hence, this means that no optimal bandwidth, as in (B3), can be used.

4. Conditional association measures derived from the conditional copulas

Similar to the unconditional case, we can measure the association between Y_1 and Y_2 , given $\mathcal{X} = \chi$ by the conditional Kendall's tau

$$\tau_{Y_1, Y_2}(\chi) = 4 \iint C_\chi(u_1, u_2) dC_\chi(u_1, u_2) - 1,$$

and the conditional Spearman's rho

$$\rho_{Y_1, Y_2}(\chi) = 12 \iint C_\chi(u_1, u_2) du_1 du_2 - 3,$$

expressed with the help of the conditional copula C_χ .

With the nonparametric estimators of the conditional copula as defined in Section 1 we can estimate the conditional Kendall's tau by

$$\hat{\tau}_n(\chi) = \frac{4}{A(\chi)} \sum_{i=1}^n \sum_{j=1}^n w_{ni}(\chi, h_n) w_{nj}(\chi, h_n) \mathbb{I}\{Y_{1i} < Y_{1j}, Y_{2i} < Y_{2j}\} - 1, \quad (22)$$

where $A(\chi) = 1 - \sum_{i=1}^n w_{ni}^2(\chi, h_n)$ and the (Y_{1i}, Y_{2i}) can be replaced with $(\tilde{U}_{1i}, \tilde{U}_{2i})$ defined in (2) if the relationship could be blurred with the effect of the covariate on the marginal distribution. We will denote this estimator as $\tilde{\tau}_n(\chi)$. The use of the conditional Kendall's tau estimators $\hat{\tau}_n(\chi)$ and $\tilde{\tau}_n(\chi)$ will be illustrated in the simulation section (Section 5) and in the real data application (Section 6).

In a similar fashion a nonparametric estimator for the conditional Spearman's rho is

$$\hat{\rho}_n(\chi) = 12 \iint C_{\chi h}(u_1, u_2) du_1 du_2 - 3 = 12 \sum_{i=1}^n w_{ni}(\chi, h_n) (1 - \hat{U}_{1i})(1 - \hat{U}_{2i}) - 3,$$

with $\hat{U}_{1i} = F_{1\chi h}(Y_{1i})$ and $\hat{U}_{2i} = F_{2\chi h}(Y_{2i})$.

5. Simulation study

To complement the theoretical results we provide here a simulation study to illustrate the finite-sample performance of the estimator $C_{\mathcal{X}h}$ of (1) and the estimator $\tilde{C}_{\mathcal{X}h}$ of (3). We include also the ‘benchmark’ estimator that corresponds to the estimator $C_{\mathcal{X}h}$ calculated from the (unobserved) $(U_{1i}, U_{2i})^\top = (F_{1\mathcal{X}_i}(Y_{1i}), F_{2\mathcal{X}_i}(Y_{2i}))^\top$. This estimator corresponds to the situations when conditional marginal distributions are known and/or the covariate does not affect the marginal distributions.

For brevity of presentation while addressing various aspects of the studied problem, we focus on estimation of the conditional copula at some fixed covariate value for the multivariate covariate case, and on estimation on the conditional Kendall’s tau function for the functional covariate case.

5.1. Multivariate covariate

In this section we are interested in estimation of a conditional copula for a two- or three-dimensional covariate at a given point $\mathbf{x} = (x_1, \dots, x_d)^\top$. The performance of the estimators is evaluated using the average (over all simulations) of the integrated squared error

$$\int_0^1 \int_0^1 \left[\hat{C}_{\mathbf{x}h}(u_1, u_2) - C_{\mathbf{x}}(u_1, u_2) \right]^2 du_1 du_2,$$

where $\hat{C}_{\mathbf{x}h}$ stands for an estimator of a conditional copula.

The model for the marginals is given by $Y_1 = \mu_1(\mathbf{X}) + \varepsilon_1$ and $Y_2 = \mu_2(\mathbf{X}) + \varepsilon_2$, where the covariates $\mathbf{X} = (X_1, \dots, X_d)^\top$ are supposed to be independent standard normal variables, and with (for $d = 2$ of 3)

$$\mu_1(\mathbf{X}) = X_1 - X_2 - (d - 2)X_3, \quad \mu_2(\mathbf{X}) = \cos(X_2 - 1) - \sin(X_1 - 1) + (d - 2)X_3.$$

Further, ε_1 and ε_2 are standard normal random variables with the joint conditional distribution function for $\mathbf{X} = \mathbf{x}$ (with $\mathbf{x} = (x_1, \dots, x_d)$) given by

$$P(\varepsilon_1 \leq e_1, \varepsilon_2 \leq e_2 \mid \mathbf{X} = \mathbf{x}) = C_{\theta(\mathbf{x})}(\Phi(e_1), \Phi(e_2)),$$

where Φ is the distribution function of a standard normal random variable and $C_{\theta(\mathbf{x})}$ is a Frank copula with the copula parameter depending on the point \mathbf{x} and given by

$$\theta(\mathbf{x}) = 5 + 7 \left(\frac{1}{1 + \exp\{2x_1^2 - 4x_1 - x_2 + (d-2)(x_3-1)^2 - 1\}} - \frac{1}{2} \right).$$

To calculate the estimators we use $\mathbf{B} = h_n^2 \mathbf{I}_d$ and a multiplicative kernel $K(\mathbf{y}) = \prod_{i=1}^d k(y_i)$ with k being a triweight kernel function $k(z) = \frac{35}{32} (1 - z^2)^3 \mathbb{I}\{|z| \leq 1\}$. The local linear weights of Ruppert and Wand (1994) [16] are used.

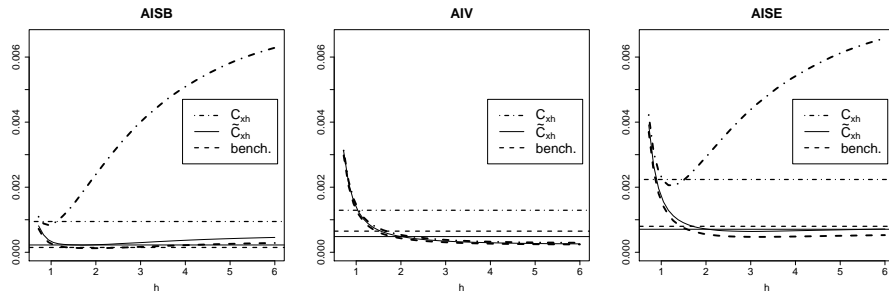


FIG 1. Copula estimation for $d = 2$.

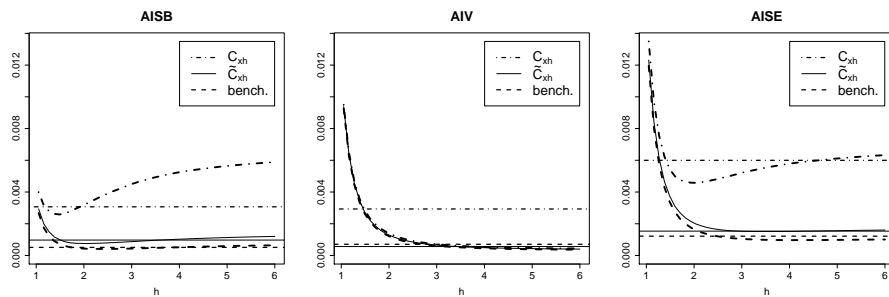


FIG 2. Copula estimation for $d = 3$.

We estimate the conditional copula at the point $\mathbf{x} = (1, 1)$, for $d = 2$, and at the point $\mathbf{x} = (1, 1, 1)$ for $d = 3$. Results based on 1000 simulated samples of size $n = 300$ are in Figures 1 and 2 respectively, where we present (from left to right) the average of the integrated squared bias (AISB), the average of the integrated variance (AIV) and the average of the integrated squared error (AISE), plotted as functions of the bandwidth h . The dotted-dashed curve shows the result for the estimator $C_{\mathbf{x}h}$ of (1), the solid curve for the estimator $\tilde{C}_{\mathbf{x}h}$ with $g_1 = g_2 = 2h$ and finally the dashed curve stands for the benchmark estimator (where we use the information on the marginals to transform these). Using the same line types we depict as horizontal lines the values of AISB, AIV and AISE when a plug-in bandwidth choice (mimicking the procedure described in Section 2.3 of Gijbels et al. (2011) [12]) is applied for each of the three displayed estimators.

Note that in the considered simulation models the marginals clearly depend on the covariates (through regression models) and hence in this situation adjusting for the effect of the covariates on the marginals is quite crucial to obtain a good performing estimator. Once the effect of the covariates is removed from the marginals, the curves of the integrated squared error of the estimator $\tilde{C}_{\mathbf{x}h}$ and of the benchmark estimator become flat and even large bandwidths give still very reasonable results. So by removing the effect of the covariates on the marginals, the choice of the bandwidth seems to have a lesser impact for the estimator $\tilde{C}_{\mathbf{x}h}$, as compared to the estimator $C_{\mathbf{x}h}$.

Note that the estimator $\tilde{C}_{\mathbf{x}h}$ is quite close to the benchmark estimator. The results of $\tilde{C}_{\mathbf{x}h}$ could be further slightly improved when one uses more specific models instead of the general approach given in (2) to adjust the marginals for the effect of the covariate. For instance additive models would be completely appropriate for this purpose in this setting. See also Section 7.

The results for $d = 2$ and $d = 3$ are similar, with somewhat larger errors for the $d = 3$ case, and the estimator $\tilde{C}_{\mathbf{x}h}$ that is not so close to the benchmark estimator as for $d = 2$. With increasing dimension of the covariate it is more difficult to adjust the marginals for the effect of the covariate.

5.2. Functional covariate

In this section we consider the following general model:

$$Y_1 = \mu_1(\mathcal{X}) + \varepsilon_1, \quad Y_2 = \mu_2(\mathcal{X}) + \varepsilon_2,$$

where ε_1 and ε_2 are standard normal random variables with the joint conditional distribution function of $(\varepsilon_1, \varepsilon_2)$ for $\mathcal{X} = \chi$ given by

$$P(\varepsilon_1 \leq e_1, \varepsilon_2 \leq e_2 \mid \mathcal{X} = \chi) = C_{\theta(\chi)}(\Phi(e_1), \Phi(e_2)),$$

where $C_{\theta(\chi)}$ is a Frank copula with the parameter $\theta(\chi)$ depending on the functional covariate $\mathcal{X}(t)$ that is observed at the equispaced points $0 = t_0 < t_1 < \dots < t_{100} = \pi$. Further elements of the different simulation models are listed in Table 1. Herein, the random variables A_1, A_2, A_3 and A_4 are independent random variables uniformly distributed on $[0, 1]$.

The aim of this section is also to provide illustrations of various aspects of the application in Section 6. We therefore focus on estimation of the conditional

TABLE 1
Simulation models for the functional covariate case

A	functional covariate $\mu_1(\mathcal{X})$ and $\mu_2(\mathcal{X})$ copula parameter function	$\mathcal{X}(t) = A_1 \cos(2t) + A_2 \sin(4t) + A_3(t^2 - \pi t + \frac{2\pi^2}{9}) + A_4 \pi$ $\mu_1(\mathcal{X}) = \int_0^\pi t \cos(t) [\mathcal{X}'(t)]^2 dt$ $\mu_2(\mathcal{X}) = \int_0^\pi \sin(t) [\mathcal{X}'(t)]^2 dt$ $\theta(\mathcal{X}) = 2 \left(\int_0^\pi [\mathcal{X}'(t)]^2 dt \right)^{1/2}$
B	functional covariate $\mu_1(\mathcal{X})$ and $\mu_2(\mathcal{X})$ real-valued covariates copula parameter function	as in Model A as in Model A $X_1 = \int_0^\pi \mathcal{X}(t) dt$ and $X_2 = \left(\int_0^\pi [\mathcal{X}'(t)]^2 dt \right)^{1/2}$ as in Model A
C	functional covariate $\mu_1(\mathcal{X})$ and $\mu_2(\mathcal{X})$ real-valued covariates copula parameter function	as in Model A $\mu_1(\mathcal{X}) = 2 \sin(\frac{1}{2} X_2)$ and $\mu_2(\mathcal{X}) = 0$ X_1 and X_2 as in Model B as in Model A
D	functional covariate $\mu_1(\mathcal{X})$ and $\mu_2(\mathcal{X})$ real-valued covariates copula parameter function	$\mathcal{X}(t) = A_1 \cos(2t) + A_2 \sin(2t) + A_3 \pi$ $\mu_1(\mathcal{X}) = 2 \sin(\frac{1}{2} X_2)$ and $\mu_2(\mathcal{X}) = 2 \cos(\frac{1}{2} X_2)$ X_1 and X_2 as in Model B $\theta(\mathcal{X}) = 4 \left(\int_0^\pi [\mathcal{X}'(t)]^2 dt \right)^{1/2}$

Kendall's tau at all points of the sample. We measure the performance of an estimator $\hat{\tau}(\chi)$ of $\tau(\chi)$ via the 'integrated squared error'

$$\text{ISE} = \int_{\mathcal{E}} [\hat{\tau}(\chi) - \tau(\chi)]^2 dF_n(\chi) = \frac{1}{n} \sum_{i=1}^n [\hat{\tau}(\chi_i) - \tau(\chi_i)]^2, \quad (23)$$

where F_n stands for the empirical distribution of the functional covariate in a given sample. The quantity ISE in (23) is then further averaged across all 1 000 simulations. Note that ISE can be decomposed into the 'squared integrated bias'

$$\text{SIB} = \left[\int_{\mathcal{E}} (\hat{\tau}(\chi) - \tau(\chi)) dF_n(\chi) \right]^2 = \left[\frac{1}{n} \sum_{i=1}^n \hat{\tau}(\chi_i) - \frac{1}{n} \sum_{i=1}^n \tau(\chi_i) \right]^2 = [\bar{\hat{\tau}} - \bar{\tau}]^2$$

and the 'integrated variance'

$$\text{IV} = \int_{\mathcal{E}} [\hat{\tau}(\chi) - \tau(\chi) - (\bar{\hat{\tau}} - \bar{\tau})]^2 dF_n(\chi) = \frac{1}{n} \sum_{i=1}^n [\hat{\tau}(\chi_i) - \tau(\chi_i) - (\bar{\hat{\tau}} - \bar{\tau})]^2.$$

An estimator for the conditional Kendall's tau is given in (22).

5.2.1. Simulation results for Model A

Simulation model A is inspired by the data generation process used in Ferraty et al. (2010) [8]. In a typical sample from this model the conditional Kendall's tau ranges from 0.1 to 0.6 with the average value of 0.4.

We consider the following four estimators:

- **orig.**: The original observations (Y_{1i}, Y_{2i}) are used in (22);
- **unif.**: The pseudo-observations $(\tilde{U}_{1i}, \tilde{U}_{2i})$ of (2) are used in (22);
- **nonp.regr.**: The residuals from the nonparametric regression of Y_1 and Y_2 on \mathcal{X} are used in (22);
- **bench.**: The unobserved $(\varepsilon_1, \varepsilon_2)$ are used in (22).

To calculate the estimators, a triweight kernel function given by $k(z) = \frac{35}{32} (1 - z^2)^3 \mathbb{I}\{|z| \leq 1\}$ and Nadaraya-Watson weights (17) as well as k-nearest neighbour weights, as in Burba et al. (2009) [7], are used.

As distance function we consider the L_2 -norm of a difference of observed functions (denoted by d_2) and the L_2 -norm of a difference of the first derivatives of observed functions (denoted by $d_2^{(1)}$). Note that the latter seems to be more appropriate here taking into account the structure of Model A.

For the nonparametric regression fit needed for the estimator **nonp.regr.** we use the R software functions *funopare.kernel.cv* (for the NW-weight system) and *funopare.knn.gcv* (for the k-nearest neighbour weights) that are available at the website <http://www.math.univ-toulouse.fr/staph/npfda/>. Finally, the bandwidths returned by the function *funopare.kernel.cv* are used as g_{n1} and g_{n2} in the calculation of the estimator **unif.** based on $(\tilde{U}_{1i}, \tilde{U}_{2i})$.

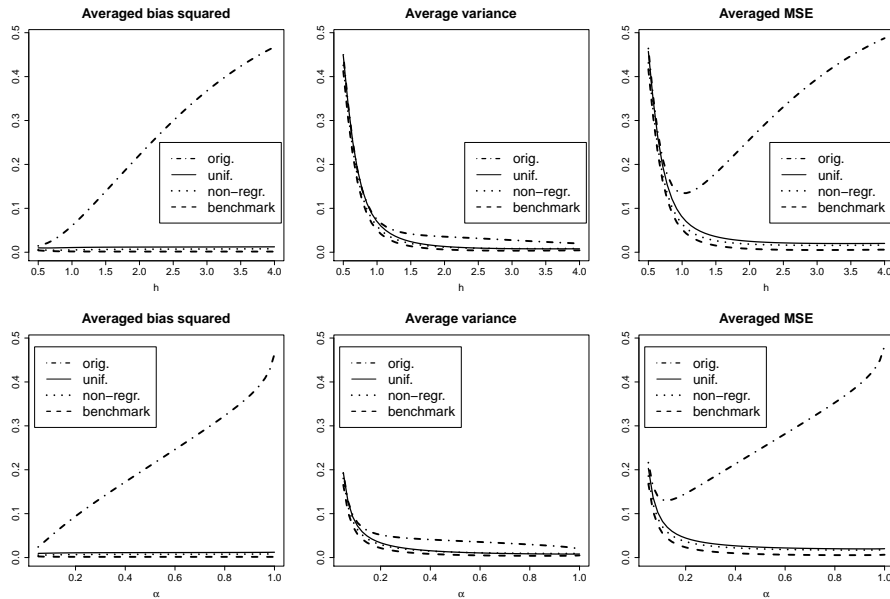


FIG 3. Conditional Kendall's tau estimators for a functional covariate when the L_2 -norm of the first derivative is used.

In the top (respectively bottom) panel of Figure 3, the simulation results for sample size $n = 200$ using the distance function based on the first derivative ($d_2^{(1)}$), are presented for the NW-weights (respectively the k-nearest neighbour weights; where $\alpha = \frac{k}{n}$). The results for the two type of weights are very similar. Note also the similarity with the results for a multivariate covariate. The estimators `unif.` and `nonp.regr.` are both very close to the benchmark estimator. The estimator `nonp.regr.` is doing slightly better than the estimator `unif.`, as nonparametric regression is, for this model, the appropriate method to adjust the marginals for the effect of the covariate.

The results for the standard L_2 -distance d_2 are summarized in Figure 4. In this case, both estimators `unif.` and `nonp.regr.` are doing considerable worse than the benchmark estimator. The reason is that the distance d_2 is not so well suited as $d_2^{(1)}$ for adjusting the marginals for the effect of the covariate. A closer inspection of the results also reveals that the benchmark estimator is doing worse for the distance d_2 than for the (more appropriate) $d_2^{(1)}$.

5.2.2. Conditioning on a functional covariate or on a summarizing real-valued covariate

When dealing with a complex covariate one might wonder whether it is not possible to replace the covariate with a simpler quantity that is easier to handle. The success of this strategy depends on how well this simpler quantity substitutes the complex covariate.

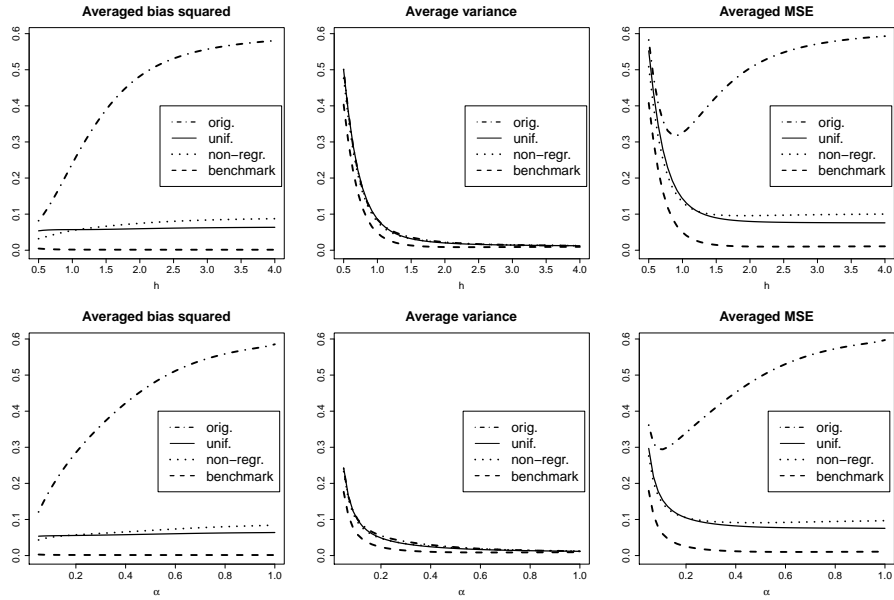


FIG 4. Conditional Kendall's tau estimators for a functional covariate when the L_2 -norm is used.

To illustrate this we consider Models B and C, which go back to Model A, but where now we define two scalar summaries X_1 and X_2 of the functional covariate \mathcal{X} (from Model A). For a real-valued covariate, one can simply use the approach of Gijbels et al. (2011) [12] and Veraverbeke et al. (2011) [19].

When calculating the estimator of the conditional Kendall's tau based either on X_1 or X_2 we take Nadaraya-Watson (or k-nearest neighbour) weights based on the same kernel as for the other estimators. For brevity we only present results for the \tilde{C} -type estimators, and for nearest neighbour weights.

Note that in Models A and B the dependence of the copula parameter on the functional covariate is fully described through X_2 (but not at all through X_1), but that this is not the case for the dependence of the marginals on the functional covariate. This is in contrast with Models C and D where the dependence of the marginals and the copula parameter is fully captured through the real-valued covariate X_2 . These models can thus be used to investigate how much we loose if the conditioning is done upon a functional covariate whereas it could have been done simply via a real-valued covariate. Also for Model D the conditional Kendall's tau ranges from 0.1 to 0.6 with the average value of 0.4.

The results for Model B are presented in Figure 5. The dotted-dashed curve is the result for the estimator based on X_1 and the dotted curve for the estimator based on X_2 . As in Figures 3 and 4 the solid curve represents the estimator $\tilde{C}_{\mathcal{X}h}$ and the dashed curve stands for the benchmark estimator (both based on $d_2^{(1)}$). From Figure 5 it is seen that the X_1 -based estimator is performing very bad,

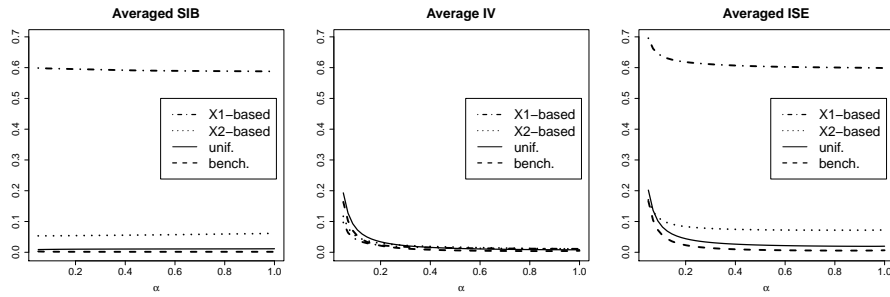


FIG 5. Conditional Kendall's tau estimators for Model B, when conditioning on the scalars X_1 or X_2 , or on the functional covariate.

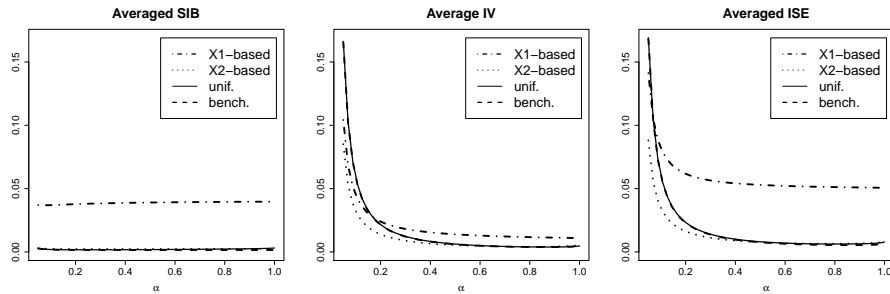


FIG 6. Conditional Kendall's tau estimators, for Model C, when conditioning on the scalars X_1 or X_2 , or on the functional covariate.

due to the enormous bias of the estimator. The X_2 -based estimator is doing much better, but also does not really have a satisfactory performance. In this model none of the two real-valued covariates can fully describe the dependence structure.

Figure 6 shows the results for Model C. Note that the X_2 -based estimator is slightly preferable to the *unif.* estimator (as well as to the benchmark estimator) only for small α . With increasing α the differences between the estimators diminishes. The reason behind this is that for larger values of α , the covariates with similar values of X_2 are also close when the distance is measured by $d_2^{(1)}$ for the functional covariate.

In contrast to this, for Model D, similar values of X_2 can be given by curves of the functional covariate that are far away when measured by the distance $d_2^{(1)}$. The results for Model D are in Figure 7. Note that the X_2 -based estimator is doing even slightly better than the benchmark estimator. Although in this case, conditioning on the functional covariate is not necessary, we see that by doing so the conditional Kendall's tau estimator still behaves reasonably well, and little harm has been done by considering the functional covariate instead of the real-valued covariate X_2 .

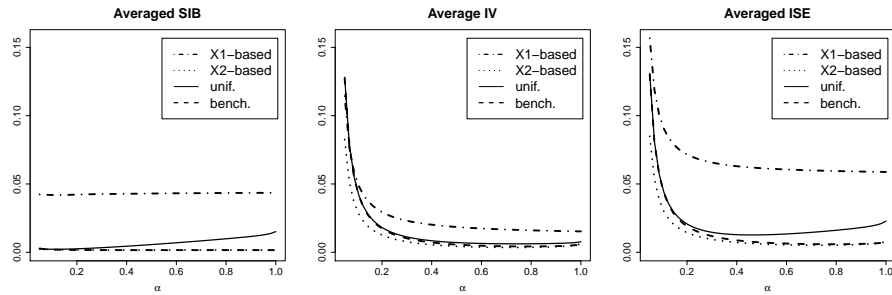


FIG 7. Conditional Kendall's tau estimators, for Model D, when conditioning on the scalars X_1 or X_2 , or on the functional covariate.

6. Application to real data

The `tecator` dataset contains 215 spectra of light absorbance as functions of the wavelength, observed on finely chopped pieces of meat. These data were first studied by Borggaard and Thodberg (1992) [6]. The original data come from a quality control problem in the food industry and can be found at website <http://lib.stat.cmu.edu/datasets/tecator>. For each finely chopped pure meat sample a 100 channel spectrum of absorbances ($-\log_{10}$ of the transmittance measured by a spectrometer) and the contents of moisture (water), fat and protein were measured. The 215 spectral curves (100 discretely observed values each; over a common grid of wavelength points) are presented in Figure 8(a). To each spectral curve corresponds a three-dimensional vector – percentage of (fat, protein, water) in each piece of meat. In this analysis we concentrate on the relationship between fat and protein.

For simplicity we summarize the degree of dependence by the conditional Kendall's tau (see Section 4). To calculate the estimator in (22) we use Nadaraya-Watson weights defined by (17). The kernel function is chosen to be $K(u) = 1 - u^2$ for $u \in [0, 1]$ and zero otherwise. For a given χ the bandwidth h_n is taken to be such that the ball of radius h_n contains 60 observations (note that this corresponds to a k-nearest neighbour type of bandwidth; see Burba et al. (2009) [7]). For calculating $(\tilde{U}_{1i}, \tilde{U}_{2i})$ defined in (2) we set $g_{n1} = g_{n2}$ such that ball of radius g_{n1} contains 30 observations.

We consider two distance functions on the covariate space, denoted by d_2 and $d_2^{(2)}$. While, the first one is the standard L_2 -norm of a difference between two spectral curves, the second one is the L_2 -norm based on a difference of the second derivatives of two spectral curves. We include the distance function $d_2^{(2)}$ as it is often used in the functional data framework analysis of this data set, see e.g. Ferraty and Vieu (2002) [9], (2006) [11], Ferraty et al. (2007) [10], (2010) [8] and Burba et al. (2009) [7] among others.

In Figure 9(a) a scatterplot of the percentages of fat and of protein for the 215 meat samples are plotted. This reveals that there is a strong negative relationship between protein and fat percentage with Kendall's tau equal to -0.69 . In

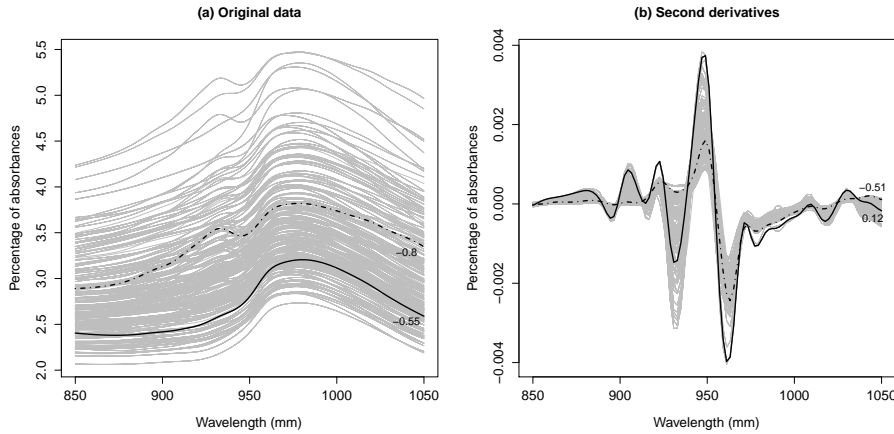


FIG 8. The spectrometric curves data (a), and the second derivatives of these curves (b), with indication of two particular curves: with the smallest (respectively largest) estimated conditional Kendall's tau: dotted-dashed (respectively solid) curves.

Figure 9(b) we plot $(\frac{n}{n+1}F_{n1}(Y_{1i}), \frac{n}{n+1}F_{n2}(Y_{2i}))$, where F_{1n} (F_{2n}) are the (unconditional) empirical distribution functions of the Y_{i1} 's (Y_{i2} 's) observations, i.e. the original data transformed in an unconditional way to uniform margins. Figures 9(a) and (b) reveal a similar strong negative relationship between protein and fat percentage. A question that rises is: How different is this relationship for different (types of) spectral curves? Or also: How does this relationship change when conditioning on spectral curves that are close in the sense of one of the two considered semi-metrics? Answering these questions can help in identifying clusters of food with a similar dependence between fat and protein.

We first consider the distance function d_2 (based on the standard L_2 -norm). For each of the spectral curves \mathcal{X}_i we estimate the conditional Kendall's tau $\hat{\tau}_n(\mathcal{X}_i)$ ($\tilde{\tau}_n(\mathcal{X}_i)$) by formula (22). Further, we take the minimal envelope (that is the pointwise minimum) of the spectral curves as the reference curve. This seems to be an appropriate reference curve in this example given the somewhat layered appearance of the spectral curves. In Figure 10(a) the conditional Kendall's tau estimates ($\hat{\tau}_n(\mathcal{X}_i)$ and $\tilde{\tau}_n(\mathcal{X}_i)$) are depicted, as a function of the distance from the minimal envelope (reference) curve. The dashed horizontal line represents the standard (unconditional) Kendall's tau (-0.69) measuring the global association between fat and protein.

Note that both estimates $\hat{\tau}_n$ and $\tilde{\tau}_n$ are very close (although $\tilde{\tau}_n$ is for most of the data points slightly more negative) which indicates that the marginal distributions are almost unaffected by the covariate when the L_2 -norm is considered. This is also confirmed by Figure 9(c) where the adjusted observations $(\tilde{U}_{1i}, \tilde{U}_{2i})$ (defined in (2)) are plotted. The relationships of the original observations (Figures 9(a) and (b)) and of the adjusted data (conditionally transformed margins) in Figure 9(c), seem to follow a very similar pattern. Finally, the main

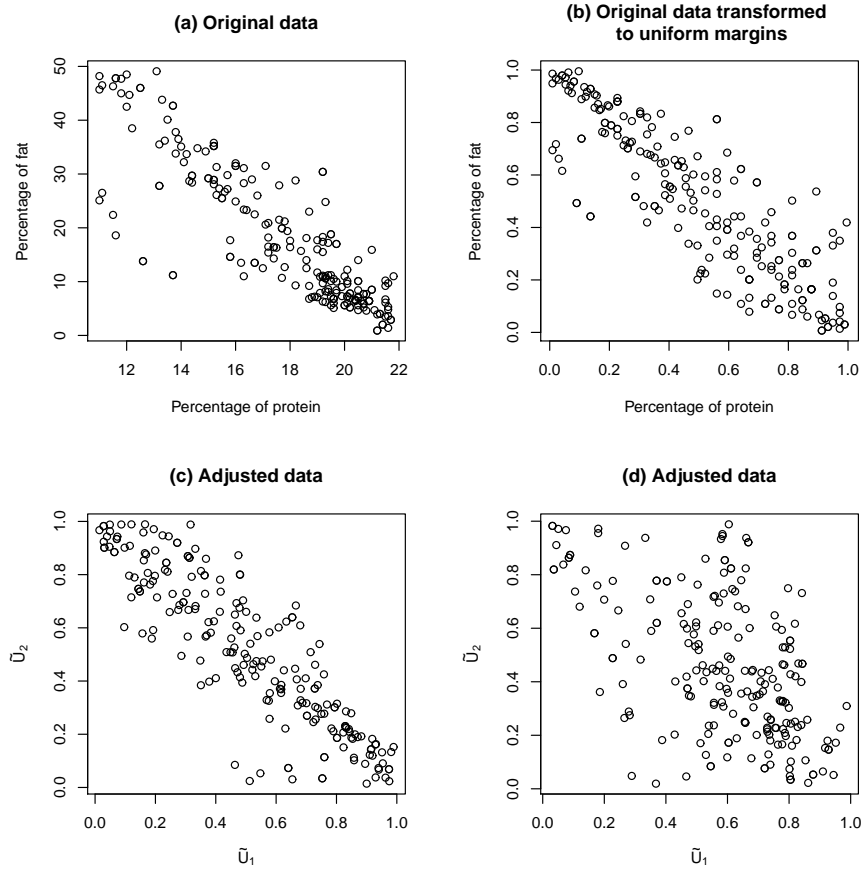


FIG 9. Protein vs. Fat: (a) Original observations; (b) Original data (unconditionally) transformed to uniform margins; Transformed data $(\tilde{U}_{1i}, \tilde{U}_{2i})$, defined in (2), when d_2 is used – (c); when $d_2^{(2)}$ is used – (d).

message of this analysis is that (see Figure 10(a)) the conditional Kendall’s tau drops (from about -0.6 to about -0.8) when moving from the minimal envelope to the curves with distances slightly above one and then comes back to about -0.65 . When focusing for example on the $\hat{\tau}_n$ estimated values one can see that the conditional Kendall’s tau ranges from -0.80 to -0.55 , where the curves that correspond to these extreme values are indicated in Figure 8(a) as bold solid and dotted-dashed curves. One can think of these as spectra for meat in which the (negative) relationship between fat and protein is most (respectively less) pronounced.

Now, let us switch to the distance function $d_2^{(2)}$ that is based on the L_2 -norm of the second derivatives. The zero function is used here as a reference function. Thus in Figure 10(b) the estimates of the conditional Kendall’s tau ($\hat{\tau}_n$ and $\tilde{\tau}_n$) are plotted against the L_2 -norm of the second derivative of a given spectral

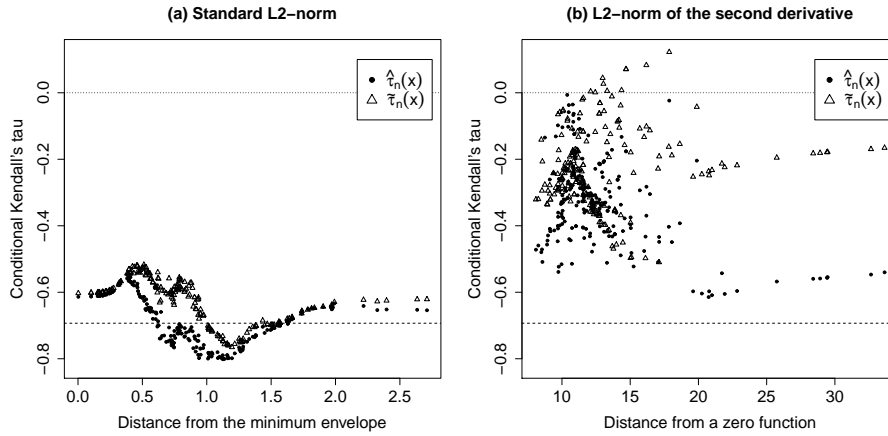


FIG 10. Conditional Kendall's tau when d_2 is used - (a); when $d_2^{(2)}$ is used - (b).

curve. Although the message of Figure 10(b) is not so clear as the message of Figure 10(a) there are several interesting things to note. First, the estimates of the conditional Kendall's tau are much closer to zero for $d_2^{(2)}$ than for d_2 . This indicates that the second derivative of a spectral curve explains a significant part of the dependence of protein and fat. In Figure 8(b) we depict the (discrete) second-order derivatives of the original spectra. In this figure we indicate the two curves for which the estimator $\tilde{\tau}_n$ achieves the minimal and maximal value (respectively -0.51 and 0.12). One can think of these as (derivative) spectra for meat in which the relationship between fat and protein is largest negative, respectively largest positive.

Further analysis reveals that the adjusted estimator $\tilde{\tau}_n$ is for approximately 74 percent of the observations higher than $\hat{\tau}_n$, with the difference being as high as 0.45. It is also interesting to look at Figure 9(d) that depicts the adjusted observations $(\tilde{U}_{1i}, \tilde{U}_{2i})$ (based on the distance function $d_2^{(2)}$). When comparing Figures 9(c) and (d) one can see that when adjusting for the effect of the covariate on the marginal distributions using $d_2^{(2)}$, there is no longer such an obvious pattern of negative relationship between fat and protein as in the case when d_2 is used. This finding points in the same direction: the second derivative of a spectral curve explains a significant part of the dependence of protein and fat.

In conclusion, the L_2 -distance based on the second derivative of the spectral curves proves to be more informative than the standard L_2 -distance not only when fitting marginal distributions of either fat or protein (cfr previous analyses in the literature), but also when one is interested in the relationship between fat and protein. Focusing on the spectral curves that are close in d_2 -sense, results in estimates of the conditional Kendall's tau ranging from -0.5 to -0.8 with the average value $\frac{1}{n} \sum_{i=1}^n \hat{\tau}_n(\chi_i) \doteq -0.69$ (or $\frac{1}{n} \sum_{i=1}^n \tilde{\tau}_n(\chi_i) \doteq -0.62$). When focusing on the spectral curves that are close in $d_2^{(2)}$ -sense estimates of the

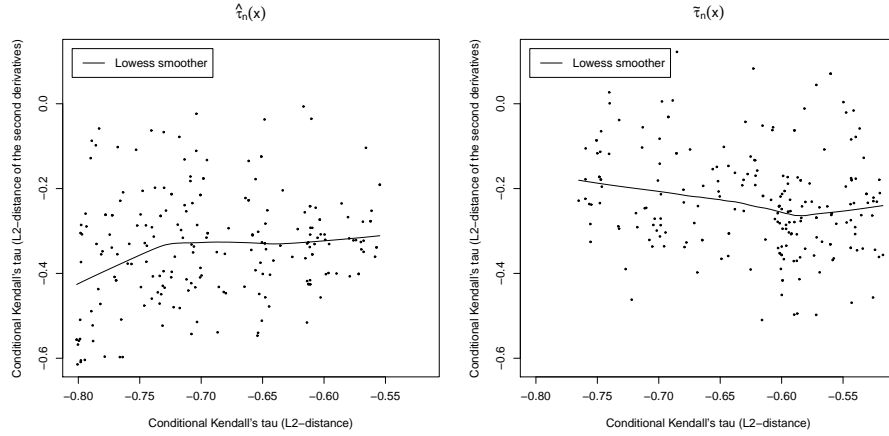


FIG 11. For each given spectral curve \mathcal{X}_i , estimates of $\hat{\tau}_n(\mathcal{X}_i)$ based on the distance function d_2 plotted versus $\hat{\tau}_n(\mathcal{X}_i)$ based on the distance function $d_2^{(2)}$. right panel: similar plot but for the estimator $\tilde{\tau}_n(\cdot)$.

conditional Kendall’s tau range for $\hat{\tau}_n$ from -0.61 to 0 (with the average value -0.33) and for $\tilde{\tau}_n$ from -0.51 to 0.12 (with the average value -0.23).

A referee pointed out that the need for having to choose a reference (spectral) curve in the above analysis might, in other examples, be less evident. One can avoid the choice of reference curves, by simply presenting estimates of the conditional Kendall’s tau for given $\mathcal{X} = \mathcal{X}_i$ (for each of the $i = 215$ spectral curves). For the two distance functions (d_2 and $d_2^{(2)}$) this leads to two estimated values for each given spectral curve. Plotting these two numbers against each other for all spectral curves results into Figure 11. The left panel shows the results for the estimator $\hat{\tau}_n$ and the right one for the estimator $\tilde{\tau}_n$. For a visual impression we also include the Lowess smoother on each scatter plot.

7. Further discussion

In this paper we introduced estimators of a conditional copula function when the covariate is either a multivariate vector or a functional covariate. From these, estimators of conditional association measures can then be easily obtained.

The proposed estimators are defined in terms of kernel weights, requiring the choice of a bandwidth parameter. A study of theoretical optimal choices of bandwidths would start from the theoretical bias and variance properties of the estimators. These would then induce a discussion on practical choices of bandwidth parameters, such as plug-in type of bandwidths. In Section 5 we illustrate the impact of the choice of a bandwidth parameter on the performance of the estimators in the current context. We also implemented a plug-in type of bandwidth selector. However, a detailed study on optimal choices of bandwidth parameters in (conditional) copula estimation, is mostly lacking so far.

When dealing with multivariate covariates of dimension d , one needs to smooth in d dimensions. This involves working with local neighbourhoods in higher dimensions, and hence one cannot avoid to face the problem of curse of dimensionality (large sample sizes are needed in high dimensions). The presented methods lead to good results for moderate sample sizes (a few hundred) for dimensions up to 3. In multivariate nonparametric regression the curse of dimensionality problem is dealt with by restricting the class of models to for example additive models or single-index models. This could be done in the current setting by assuming that the multivariate covariate influences Y_1 as well as Y_2 through an additive regression model, or via a single-index model. Modeling the conditional copula function could then be done in a semiparametric way (see e.g. Abegaz et al. (2012) [1]), restricting the copula parameter function to be modelled also in an additive way. This is the subject of future research.

In case of a functional covariate, there is also the issue of the choice of distance function (or norm) to measure the distance between two curves. In the application in Section 6 we worked out analysis for two different distance functions, revealing that: (i) a comparison between the analyses based on different distances can lead to interesting findings; and (ii) some distance functions might appear more natural than others in a given study.

Acknowledgement

The authors are grateful to the Editor, the Associate Editor and two referees for their very valuable comments which led to a considerable improvement of the manuscript. This work was supported by the IAP Research Networks P6/03 and P7/13 of the Belgian State (Belgian Science Policy). The first author gratefully acknowledges support from the projects GOA/07/04 and GOA/12/014 of the Research Fund KULeuven, as well as support from the the FWO-project G.0328.08N of the Flemish Science Foundation. The work of the second author was supported by the grant GACR P201/11/P290. The third author is extraordinary professor at the North-West University, Potchefstroom, South Africa. He also acknowledges support from research grant MTM 2008-03129 of the Spanish Ministerio de Ciencia e Innovacion.

Appendix A: Multivariate covariate

Regularity assumptions for Theorems 1 and 2

The regularity assumptions needed in Theorems 1 and 2 are as follows.

Let us denote $b_n = \max\{h_n, g_{1n}, g_{2n}\}$, $I_{\mathbf{x}}^{(n)} = \{i : w_{ni}(\mathbf{x}, b_n) \neq 0\}$ and $J_{\mathbf{x}}^{(n)} = \text{Conv}\{\mathbf{X}_i, i \in I_{\mathbf{x}}^{(n)}\}$ where Conv stands for the convex hull. Let c_n stand for a sequence of positive constants such that $n c_n \rightarrow \infty$ and $c_n = O(n^{-1/(4+d)})$. The following is a listing of assumptions on the system of weights $\{w_{ni}; i = 1, \dots, n\}$ in random design. The conditions for a fixed design, may be derived easily by replacing \mathbf{X}_i by \mathbf{x}_i and omitting the symbol P in the index.

Assumptions (W1)

Let λ_d be the d -dimensional Lebesgue measure. Assume that $\lambda_d(J_{\mathbf{x}}^{(n)}) = o_P(1)$ and

$$\begin{aligned} \max_{1 \leq i \leq n} |w_{ni}(\mathbf{x}, h_n)| &= o_P\left(\frac{1}{\sqrt{n} h_n^d}\right), \quad \sum_{i=1}^n w_{ni}(\mathbf{x}, h_n) - 1 = o_P\left(\frac{1}{\sqrt{n} h_n^d}\right), \\ \sum_{i=1}^n w_{ni}^2(\mathbf{x}, h_n) &= O_P\left(\frac{1}{n h_n^d}\right), \quad \sum_{i=1}^n w_{ni}(\mathbf{x}, h_n)(\mathbf{X}_i - \mathbf{x}) = O_P\left(\frac{1}{\sqrt{n} h_n^d}\right). \end{aligned}$$

Assumptions (W2)

Let $w'_{ni}(\mathbf{z}, g_{jn})$ denote the derivative with respect to \mathbf{z} , $\|\cdot\|$ stands for the Euclidean norm and \otimes^2 stands for the outer product (that is $\mathbf{x} \otimes^2 = \mathbf{x} \mathbf{x}^\top$). Assume that

$$\begin{aligned} \sum_{i=1}^n |w_{ni}(\mathbf{x}, h_n)| &= O_P(1), \quad \sup_{\mathbf{z} \in J_{\mathbf{x}}^{(n)}} \left| \sum_{i=1}^n w_{ni}(\mathbf{z}, g_{jn}) - 1 \right| = o_P(g_{jn}^2), \\ \sum_{i=1}^n w_{ni}(\mathbf{x}, h_n)(\mathbf{X}_i - \mathbf{x}) \otimes^2 &= O_P\left(\frac{1}{\sqrt{n} h_n^d}\right), \\ \sup_{\mathbf{z} \in J_{\mathbf{x}}^{(n)}} \sum_{i=1}^n [w_{ni}(\mathbf{z}, g_{jn})]^2 &= O_P\left(\frac{1}{n g_{jn}^d}\right), \quad \sup_{\mathbf{z} \in J_{\mathbf{x}}^{(n)}} \sum_{i=1}^n \|w'_{ni}(\mathbf{z}, g_{jn})\|^2 = O_P\left(\frac{1}{n g_{jn}^{2+d}}\right). \end{aligned}$$

Assumptions (W3)

$$\begin{aligned} \exists_{C < \infty} \mathbb{P} \left[\sup_{\mathbf{z} \in J_{\mathbf{x}}^{(n)}} \max_{1 \leq i \leq n} |w_{ni}(\mathbf{z}, h_n)| \mathbb{I}\{\|\mathbf{X}_i - \mathbf{z}\| > C h_n\} > 0 \right] &= o(1), \\ \exists_{\mathbf{D}_K \in \mathbb{R}^d, \|\mathbf{D}_K\| < \infty} \forall_{c_n} \sup_{\mathbf{z} \in J_{\mathbf{x}}^{(n)}} \left| \sum_{i=1}^n w_{ni}(\mathbf{z}, c_n)(\mathbf{X}_i - \mathbf{z}) - c_n^2 \mathbf{D}_K \right| &= o_P(c_n^2), \\ \exists_{\mathbf{E}_K \in \mathbb{R}^{d \times d}, \|\mathbf{E}_K\| < \infty} \forall_{c_n} \sup_{\mathbf{z} \in J_{\mathbf{x}}^{(n)}} \left| \sum_{i=1}^n w_{ni}(\mathbf{z}, c_n)(\mathbf{X}_i - \mathbf{z}) \otimes^2 - c_n^2 \mathbf{E}_K \right| &= o_P(c_n^2). \end{aligned}$$

Assumptions about conditional distributions

We require the conditional copula $C_{\mathbf{z}}$ and the conditional marginal distribution functions $F_{1\mathbf{z}}$ and $F_{2\mathbf{z}}$ to satisfy:

- (R1) The functions $\dot{H}_{\mathbf{z}}(F_{1\mathbf{x}}^{-1}(u_1), F_{2\mathbf{x}}^{-1}(u_2))$ and $\ddot{H}_{\mathbf{z}}(F_{1\mathbf{x}}^{-1}(u_1), F_{2\mathbf{x}}^{-1}(u_2))$ are uniformly continuous in (\mathbf{z}, u_1, u_2) , where \mathbf{z} takes value in a neighbourhood of \mathbf{x} .

- (R2) For $j = 1, 2$, the j th first-order partial derivative of $C_{\mathbf{x}}$ (i.e. derivative with respect to u_j) exists and is continuous on the set $\{(u_1, u_2) \in [0, 1]^2 : 0 < u_j < 1\}$.
- (R̃1) $\dot{C}_{\mathbf{z}}(u_1, u_2) = \frac{\partial}{\partial \mathbf{z}} C_{\mathbf{z}}(u_1, u_2)$, $\ddot{C}_{\mathbf{z}}(u_1, u_2) = \frac{\partial^2}{\partial \mathbf{z} \partial \mathbf{z}^\top} C_{\mathbf{z}}(u_1, u_2)$ exist and are continuous as functions of (\mathbf{z}, u_1, u_2) , where \mathbf{z} takes value in a neighbourhood of \mathbf{x} .
- (R̃2) For $j = 1, 2$, the j -th first-order partial derivative of $C_{\mathbf{z}}$ exists and is continuous on the set $U(\mathbf{x}) \times \{(u_1, u_2) \in [0, 1]^2 : 0 < u_j < 1\}$, where $U(\mathbf{x})$ is a neighbourhood of the point \mathbf{x} .
- (R̃3) For $j = 1, 2$: $F_{j\mathbf{z}}(F_{j\mathbf{z}}^{-1}(u))$, $\dot{F}_{j\mathbf{z}}(F_{j\mathbf{z}}^{-1}(u))$, $\ddot{F}_{j\mathbf{z}}(F_{j\mathbf{z}}^{-1}(u))$ are continuous as functions of (\mathbf{z}, u) for \mathbf{z} in a neighbourhood of \mathbf{x} , where $\dot{F}_{j\mathbf{z}}(y) = \frac{\partial}{\partial \mathbf{z}} F_{j\mathbf{z}}(y)$, $\ddot{F}_{j\mathbf{z}}(\mathbf{y}) = \frac{\partial^2}{\partial \mathbf{z} \partial \mathbf{z}^\top} F_{j\mathbf{z}}(y)$.

Appendix B: Functional covariate

We first state the regularity assumptions needed in Theorems 3 and 4 respectively.

Regularity assumptions for Theorem 3

- (F0) The functional H_z is continuous at the point χ uniformly in (y_1, y_2) , that is

$$\sup_{\chi' \in \mathcal{E}, \|\chi'\| \leq 1} \sup_{y_1, y_2} |H_{\chi+\delta\chi'}(y_1, y_2) - H_{\chi}(y_1, y_2)| \rightarrow 0, \quad \text{as } \delta \rightarrow 0_+.$$

- (F1) It is assumed that

$$\forall s \in [0, 1], \tau_{\chi, h}(s) \rightarrow \tau_{\chi, 0}(s), \quad \text{as } h \rightarrow 0_+. \tag{B1}$$

- (F2) Suppose that the (univariate) kernel K is supported on $[0, 1]$ and has a continuous derivative on $[0, 1)$, $K'(s) \leq 0$ and

$$M_1 = \left(K(1) - \int_0^1 K'(s) \tau_{\chi, 0}(s) ds \right) > 0. \tag{B2}$$

- (F3) The function ϕ_{χ, y_1, y_2}^H satisfies uniformly in (y_1, y_2)

$$\phi_{\chi, y_1, y_2}^H(s) = s^\beta \vartheta_H(y_1, y_2) + o(s^\beta), \quad \text{as } s \rightarrow 0_+,$$

for some $\beta > 0$, where $\vartheta_H(F_{1\chi}^{-1}(u_1), F_{2\chi}^{-1}(u_2))$, viewed as a function of the arguments (u_1, u_2) , is continuous on $[0, 1]^2$.

- (F4) The bandwidth h_n satisfies

$$n \varphi_\chi(h_n) \rightarrow \infty, \quad n \varphi_\chi(h_n) h_n^{2\beta} = O(1). \tag{B3}$$

As $K'(s) \leq 0$, assumption (B2) might seem at first sight superfluous. But as reviewed e.g. in Ferraty et al. (2007) [10] for a broad class of processes \mathcal{X} the small ball probability function is of order

$$\varphi_{\mathcal{X}}(h) \sim C h^{-\alpha} e^{-C/h^\gamma}, \quad \text{where } \alpha > 0, \gamma > 0. \quad (\text{B4})$$

This further implies that $\tau_{\mathcal{X},0}$ equals the Dirac mass at 1, which gives

$$\int_0^1 K'(s) \tau_{\mathcal{X},0}(s) ds = 0.$$

Thus in order to guarantee that $M_1 > 0$ the assumption $K(1) > 0$ is usually explicitly stated. As noted in Ferraty and Vieu (2006) [11] when the small ball probability function is of order (B4), the rate of convergence of the nonparametric estimator is only logarithmic which compromises the whole estimating procedure.

Next to M_1 it is useful to define the following constants

$$\widetilde{M}_\beta = \left(K(1) - \int_0^1 (s^\beta K(s))' \tau_{\mathcal{X},0}(s) ds \right), \quad (\text{B5})$$

$$M_2 = \left(K^2(1) - \int_0^1 (K^2(s))' \tau_{\mathcal{X},0}(s) ds \right). \quad (\text{B6})$$

Regularity assumptions for Theorem 4

($\tilde{\mathbf{F}}1$) There exists an open neighbourhood $U(\chi)$ of χ such that $\sup_{\chi' \in U(\chi)} \frac{\varphi_{\chi'}(t)}{\varphi_{\chi}(t)}$ is a bounded function of t in a neighbourhood of 0. Moreover the functions $\tau_{\mathcal{X},h}(s)$ and $\tau_{\mathcal{X},0}(s)$ defined in (18) and (B1) satisfy

$$\sup_{\chi' \in U(\chi)} \sup_{0 \leq s \leq 1} |\tau_{\mathcal{X}',h}(s) - \tau_{\mathcal{X}',0}(s)| = o(1), \quad \text{for } h \rightarrow 0_+.$$

($\tilde{\mathbf{F}}2$) Suppose that the (univariate) kernel K is supported on $[0, 1]$, has a continuous derivative on $[0, 1)$, $K'(s) \leq 0$ and there exists an open neighbourhood $U(\chi)$ of the point χ such that

$$\inf_{\chi' \in U(\chi)} M_{1\chi'} > 0, \quad \text{where } M_{1\chi'} = \left(K(1) - \int_0^1 K'(s) \tau_{\mathcal{X}',0}(s) ds \right).$$

($\tilde{\mathbf{F}}3$) There exists an open neighbourhood $U(\chi)$ of the point χ such that the functional $\phi_{\mathcal{X}',u_1,u_2}^C$ satisfies uniformly in (\mathcal{X}', u_1, u_2)

$$\phi_{\mathcal{X}',u_1,u_2}^C(s) = s^\beta \vartheta_C(\mathcal{X}', u_1, u_2) + o(s^\beta), \quad \text{as } s \rightarrow 0_+,$$

for some $\beta > 0$, where $\vartheta_C(\mathcal{X}', u_1, u_2)$, viewed as a function of (\mathcal{X}', u_1, u_2) , is continuous on $U(\chi) \times [0, 1]^2$. Analogously for $j = 1, 2$ uniformly in y

$$\phi_{\mathcal{X}',y}^{F_j}(s) = s^\beta \vartheta_{F_j}(\mathcal{X}', y) + o(s^\beta), \quad \text{as } s \rightarrow 0_+,$$

where $\vartheta_{F_j}(\mathcal{X}', F_{j\mathcal{X}'}^{-1}(u))$, viewed as a function of (\mathcal{X}', u) , is continuous on $U(\chi) \times [0, 1]$.

($\tilde{\mathbf{F}}4$) The bandwidths h_n , g_{n1} and g_{n2} satisfy

$$n \varphi_{\mathcal{X}}(h_n) g_{jn}^{2\beta} = O(1), \quad \frac{h_n}{g_{jn}} = O(1), \quad n \varphi_{\mathcal{X}}(\min(h_n, g_{1n}, g_{2n})) \rightarrow \infty.$$

($\tilde{\mathbf{F}}5$) The set $B(0, 1) = \{\mathcal{X}' \in \mathcal{E}, \|\mathcal{X}'\| \leq 1\}$ is totally bounded (that is, for every $\varepsilon > 0$, there exists a finite cover such that the diameter of each element of the cover is at most ε).

Estimation of a conditional marginal distribution

It will be useful to summarize (and in some aspects slightly extend) some of the results about nonparametric estimation of a conditional distributional function in this functional covariate setting.

Assume for this purpose that (\mathcal{X}_i, Y_i) is a sample of i.i.d. pairs and $F_{\mathcal{X}h}(y)$ is an estimator of the conditional distribution of Y given $\mathcal{X} = \chi$, defined as

$$F_{\mathcal{X}h}(y) = \frac{R_{n1}(y)}{R_{n2}},$$

with

$$R_{n1}(y) = \frac{1}{n \varphi_{\mathcal{X}}(h_n)} \sum_{i=1}^n K\left(\frac{\|\mathcal{X}_i - \mathcal{X}\|}{h_n}\right) \mathbb{I}\{Y_i \leq y\} \tag{B7}$$

and

$$R_{n2} = R_{n1}(\infty) = \frac{1}{n \varphi_{\mathcal{X}}(h_n)} \sum_{i=1}^n K\left(\frac{\|\mathcal{X}_i - \mathcal{X}\|}{h_n}\right). \tag{B8}$$

Then by a standard decomposition

$$F_{\mathcal{X}h}(y) - \frac{\mathbb{E}R_{n1}(y)}{\mathbb{E}R_{n2}} = \frac{1}{\mathbb{E}R_{n2}} (R_{n1}(y) - \mathbb{E}R_{n1}(y)) - \frac{R_{n1}(y)}{R_{n2} \mathbb{E}R_{n2}} (R_{n2} - \mathbb{E}R_{n2}). \tag{B9}$$

By the same arguments as will be given later in Step 1 one can show that the following process

$$\mathbb{R}_{n1}(y) = \sqrt{n \varphi_{\mathcal{X}}(h_n)} [R_{n1}(y) - \mathbb{E}R_{n1}(y)], \quad y \in \mathbb{R},$$

where

$$R_{n1}(y) = \sum_{i=1}^n K\left(\frac{\|\mathcal{X}_i - \mathcal{X}\|}{h_n}\right) \mathbb{I}\{Y_i \leq y\},$$

is asymptotically tight. This tightness together with assumption (**F0**) also implies

$$R_{n2} = \mathbb{E}R_{n2} + o_P(1) = \frac{1}{\varphi_{\mathcal{X}}(h_n)} \mathbb{E}K\left(\frac{\|\mathcal{X}_1 - \mathcal{X}\|}{h_n}\right) + o_P(1) = M_1 + o_P(1), \tag{B10}$$

where M_1 was defined in (B2). For the last equation see the calculations done in the proofs of Lemma 1 and Lemma 2 of Ferraty et al. (2007) [10]. Analogously one can derive with the help of an analogy to assumption (F3) that

$$\frac{1}{\varphi_\chi(h_n)} \mathbb{E} \left[K \left(\frac{\|\mathcal{X}_1 - \chi\|}{h_n} \right) [\mathbb{I}\{Y_1 \leq y\} - F_\chi(y)] \right] = O(h_n^\beta). \quad (\text{B11})$$

Finally, the asymptotic tightness of \mathbb{R}_{n1} , (B7), (B8), (B9), (B10) and (B11) imply that

$$\sup_{y \in \mathbb{R}} |F_{\chi h}(y) - F_\chi(y)| = O_P \left(\frac{1}{n \varphi_\chi(h_n)} \right) + O(h_n^\beta). \quad (\text{B12})$$

Thus by a similar argument as in Veraverbeke et al. (2011) [19] one can show that for each $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[F_\chi^{-1}(u - \varepsilon) \leq F_{\chi h}^{-1}(u) \leq F_\chi^{-1}(u + \varepsilon), u \in [0, 1] \right] = 1. \quad (\text{B13})$$

B.1. Proof of Theorem 3

Although the structure of the proof follows that of the proof of Theorem 1 in Veraverbeke et al. (2011) [19], contrary to that proof it is no longer useful to condition on the value of the covariate. The issue is that such conditioning would require tools as a Taylor expansion in the neighbourhood of χ that is no longer straightforwardly available in this functional context.

For $(u_1, u_2) \in [0, 1]^2$ put

$$R_{n3}(u_1, u_2) = \frac{1}{n \varphi_\chi(h_n)} \sum_{i=1}^n K \left(\frac{\|\mathcal{X}_i - \chi\|}{h_n} \right) \mathbb{I}\{Y_{1i} \leq F_{1\chi h}^{-1}(u_1), Y_{2i} \leq F_{2\chi h}^{-1}(u_2)\},$$

and note that

$$C_{\chi h}(u_1, u_2) = \frac{R_{n3}(u_1, u_2)}{R_{n2}}, \quad (\text{B14})$$

where R_{n2} was defined in (B8).

Process R_{n3}

Let us decompose the process R_{n3} as

$$\sqrt{n \varphi_\chi(h_n)} R_{n3} = A_n^{h_n} + B_n^{h_n} + C_n^{h_n}, \quad (\text{B15})$$

where $A_n^{h_n} = D_n^{h_n} - \mathbb{E}D_n^{h_n}$, with

$$D_n^{h_n}(u_1, u_2) = \frac{1}{\sqrt{n \varphi_\chi(h_n)}} \sum_{i=1}^n K \left(\frac{\|\mathcal{X}_i - \chi\|}{h_n} \right) \left[\mathbb{I}\{Y_{1i} \leq F_{1\chi h}^{-1}(u_1), Y_{2i} \leq F_{2\chi h}^{-1}(u_2)\} \right. \\ \left. - \mathbb{I}\{Y_{1i} \leq F_{1\chi}^{-1}(u_1), Y_{2i} \leq F_{2\chi}^{-1}(u_2)\} \right],$$

and

$$\begin{aligned}
 B_n^{h_n}(u_1, u_2) &= \\
 &= \frac{1}{\sqrt{n} \varphi_X(h_n)} \sum_{i=1}^n K\left(\frac{\|X_i - X\|}{h_n}\right) \mathbb{I}\{Y_{1i} \leq F_{1X}^{-1}(u_1), Y_{2i} \leq F_{2X}^{-1}(u_2)\} \quad (\text{B16})
 \end{aligned}$$

$$\begin{aligned}
 C_n^{h_n}(u_1, u_2) &= \mathbb{E} D_n^{h_n}(u_1, u_2) \quad (\text{B17}) \\
 &= \sqrt{\frac{n}{\varphi_X(h_n)}} \mathbb{E} \left\{ K\left(\frac{\|X_1 - X\|}{h_n}\right) \left[\mathbb{I}\{Y_{11} \leq F_{1Xh}^{-1}(u_1), Y_{21} \leq F_{2Xh}^{-1}(u_2)\} \right. \right. \\
 &\quad \left. \left. - \mathbb{I}\{Y_{11} \leq F_{1X}^{-1}(u_1), Y_{21} \leq F_{2X}^{-1}(u_2)\} \right] \right\}.
 \end{aligned}$$

In the following two steps: 1. We show that the term $A_n^{h_n}$ is asymptotically negligible uniformly in (u_1, u_2) ; 2. We find the asymptotic representation of the processes $C_n^{h_n}$. Together with (B16) this will give us the asymptotic representation for the process R_{n3} .

Step 1 – Asymptotic negligibility of $A_n^{h_n}$

The process $A_n^{h_n}$ can be viewed as an empirical process Z_n indexed by the family of functions from $\mathcal{E} \times \mathbb{R}^2$ to \mathbb{R} given by

$$\begin{aligned}
 \mathcal{F}_n = \left\{ (X', w_1, w_2) \mapsto \frac{1}{\sqrt{\varphi_X(h_n)}} K\left(\frac{\|X' - X\|}{h_n}\right) \mathbb{I}\{w_1 \leq G_1^{-1}(u_1), w_2 \leq G_2^{-1}(u_2)\}; \right. \\
 \left. (u_1, u_2) \in [0, 1]^2, G_1, G_2 : \mathbb{R} \rightarrow [0, 1] \text{ nondecreasing} \right\}.
 \end{aligned}$$

Thus each function $f \in \mathcal{F}_n$ may be formally identified by (u_1, u_2, G_1, G_2) . The introduction of the process Z_n is motivated by the fact that

$$D_n^{h_n}(u_1, u_2) = Z_n(f_n) - Z_n(f),$$

where

$$f_n = (u_1, u_2, F_{1Xh}, F_{2Xh}), \quad f = (u_1, u_2, F_{1X}, F_{2X}).$$

Finally, let us equip the index set \mathcal{F}_n with a semimetric ρ defined as

$$\rho^2(f, f') = \left| F_{1X}(G_1^{-1}(u_1)) - F_{1X}(G_1'^{-1}(u_1')) \right| + \left| F_{2X}(G_2^{-1}(u_2)) - F_{2X}(G_2'^{-1}(u_2')) \right|$$

and note that the semimetric space (\mathcal{F}_n, ρ) is totally bounded.

Note that by Lemma 2.6.18(vi) of van der Vaart and Wellner (1996) [18] the set \mathcal{F}_n is a VC-class (Vapnik–Chervonenkis class) of functions with VC-index independent of n and the envelope $F_n = \frac{1}{\sqrt{\varphi_X(h_n)}} K\left(\frac{\|X' - X\|}{h_n}\right)$ which satisfies

$$\mathbb{E} F_n^2 = O(1), \quad \forall_{\eta > 0} \mathbb{E} \left[F_n^2 \mathbb{I}\{F_n > \eta \sqrt{n}\} \right] \rightarrow 0, \text{ as } n \rightarrow \infty. \quad (\text{B18})$$

Further with the help of **(F0)** and the identity $EY = E\{E[Y|X]\}$ one can for a given sequence $\{\delta_n\}$ of positive numbers going to zero estimate

$$\begin{aligned}
& \sup_{\rho(f,f') < \delta_n} E(f(\mathcal{X}_1, Y_{11}, Y_{12}) - f'(\mathcal{X}_1, Y_{11}, Y_{12}))^2 \\
& \leq \sup_{\rho(f,f') < \delta_n} \frac{1}{\varphi_X(h_n)} EK^2 \left(\frac{\|\mathcal{X}_1 - X\|}{h_n} \right) \left[|F_{1\mathcal{X}_1}(G_1^{-1}(u_1)) - F_{1\mathcal{X}_1}(G_1'^{-1}(u_1'))| \right. \\
& \quad \left. + |F_{2\mathcal{X}_1}(G_2^{-1}(u_2)) - F_{2\mathcal{X}_1}(G_2'^{-1}(u_2'))| \right] \\
& \leq O(1) \sup_{\rho(f,f') < \delta_n} \left[|F_{1X}(G_1^{-1}(u_1)) - F_{1X}(G_1'^{-1}(u_1'))| \right. \\
& \quad \left. + |F_{2X}(G_2^{-1}(u_2)) - F_{2X}(G_2'^{-1}(u_2'))| + o(1) \right] \\
& = O(1) (o(1) + \delta_n^2) \rightarrow 0, \quad \text{as } n \rightarrow \infty. \tag{B19}
\end{aligned}$$

Now thanks to **(B18)** and **(B19)** one can apply Theorem 2.11.22 of van der Vaart and Wellner (1996) [18] which yields the asymptotic uniform ρ -equicontinuity of the empirical process indexed by the functions from \mathcal{F}_n . Further with the help of **(B12)** one can mimic the corresponding computation done in Veraverbeke et al. (2011) [19] and show that uniformly in (u_1, u_2)

$$\rho^2((u_1, u_2, F_{1Xh}, F_{2Xh}), (u_1, u_2, F_{1X}, F_{2X})) = o_P(1),$$

which further implies the asymptotic negligibility of the process $A_n^{h_n}$.

Step 2 – Asymptotic representation of $C_n^{h_n}$

Similarly as in the proofs of Lemma 1 and Lemma 2 of Ferraty et al. (2007) [10] one can make use of **(B13)** and assumption **(F3)** and derive

$$\begin{aligned}
C_n^{h_n}(u_1, u_2) &= \sqrt{\frac{n}{\varphi_X(h_n)}} EK \left(\frac{\|\mathcal{X}_1 - X\|}{h_n} \right) \left[H_{\mathcal{X}_1}(F_{1\mathcal{X}h}^{-1}(u_1), F_{2\mathcal{X}h}^{-1}(u_2)) \right. \\
& \quad \left. - H_{\mathcal{X}_1}(F_{1X}^{-1}(u_1), F_{2X}^{-1}(u_2)) \right] \\
&= \sqrt{\frac{n}{\varphi_X(h_n)}} EK \left(\frac{\|\mathcal{X}_1 - X\|}{h_n} \right) \left[H_X(F_{1Xh}^{-1}(u_1), F_{2Xh}^{-1}(u_2)) \right. \\
& \quad \left. - H_X(F_{1X}^{-1}(u_1), F_{2X}^{-1}(u_2)) \right] + \sqrt{n \varphi_X(h_n)} o(h^\beta). \tag{B20}
\end{aligned}$$

With the help of **(F4)** the last term on the right-hand side of **(B20)** is of order $o(1)$. Further, thanks to the last equation in **(B10)** it is sufficient to deal with the process

$$\sqrt{n \varphi_X(h_n)} \left[C_X \left(F_{1X}(F_{1Xh}^{-1}(u_1)), F_{2X}(F_{2Xh}^{-1}(u_2)) \right) - C_X(u_1, u_2) \right]. \tag{B21}$$

Substituting 1 for u_1 (or u_2) in decomposition **(B15)** together with the asymptotic negligibility of the process $A_n^{h_n}$ and **(B21)** yields (uniformly in u_j for $j = 1, 2$)

$$\begin{aligned}
 &M_1 \sqrt{n \varphi_X(h_n)} \left[F_{jX}(F_{jXh}^{-1}(u_j)) - u_j \right] \\
 &= -\frac{1}{\sqrt{n \varphi_X(h_n)}} \sum_{i=1}^n K\left(\frac{\|X_i - X\|}{h_n}\right) [\mathbb{I}\{Y_{ji} \leq F_{jX}^{-1}(u_j)\} - u_j] + o_P(1). \quad (\text{B22})
 \end{aligned}$$

Denote by \mathbb{Y}_{jn} the process on the right-hand side of equation (B13). This process is asymptotically ρ -equicontinuous with $\rho^2(u, v) = |u - v|$ and the expectation of this process can be made arbitrarily small by taking u close to either 0 or 1 and n sufficiently large. Thus a straightforward application of the results of Section 3 of Segers (2012) [17], (B20), (B21) and (B22) imply

$$C_n^{h_n}(u_1, u_2) = -C_X^{(1)}(u_1, u_2)\mathbb{Y}_{1n}(u_1) - C_X^{(2)}(u_1, u_2)\mathbb{Y}_{2n}(u_2) + o_P(1). \quad (\text{B23})$$

Now, combining **Step 1** and **Step 2** yields the representation

$$\sqrt{n \varphi_X(h_n)} R_{n3} = B_n^{h_n} - C_X^{(1)} \mathbb{Y}_{1n} - C_X^{(2)} \mathbb{Y}_{2n} + o_P(1) \quad (\text{B24})$$

Thus (B10), (B14), (B16), and (B24) implies

$$\begin{aligned}
 \sqrt{n \varphi_X(h_n)} (C_{Xh} - C_X) &= \frac{\sqrt{n \varphi_X(h_n)} R_{n3}}{R_{n2}} - C_X \sqrt{n \varphi_X(h_n)} R_{n2} \\
 &= \frac{1}{R_{n2}} \left(B_n^{h_n} - C_X^{(1)} \mathbb{Y}_{1n} - C_X^{(2)} \mathbb{Y}_{2n} - C_X \sqrt{n \varphi_X(h_n)} R_{n2} \right) + o_P(1),
 \end{aligned}$$

which together with (B8) finally gives the statement of Theorem 3 about the asymptotic representation of the process $C_{Xn}^{(E)}$.

Proof of Theorem 4

The proof differs from the proof of Theorem 2 in Veraverbeke et al. (2011) [19] in an analogous way as the proof of Theorem 3 of this paper differs from the proof of Theorem 1 in Veraverbeke et al. (2011) [19]. The details are omitted for the reason of brevity of the paper.

References

[1] ABEGAZ, F., GIJBELS, I. and VERAVERBEKE, N. (2012). Semiparametric estimation of conditional copulas. *J. Multivariate Anal.* **110** 43–73. [MR2927509](#)

[2] ACAR, E. F., CRAIU, R. V. and YAO, F. (2011). Dependence calibration in conditional copulas: a nonparametric approach. *Biometrics* **67** 445–453. [MR2829013](#)

- [3] BAÍLLO, A. and GRANÉ, A. (2009). Local linear regression for functional predictor and scalar response. *J. Multivariate Anal.* **100** 102–111. [MR2460480](#)
- [4] BARRIENTOS-MARIN, J., FERRATY, F. and VIEU, P. (2010). Locally modelled regression and functional data. *J. Nonparametr. Stat.* **22** 617–632. [MR2682211](#)
- [5] BERLINET, A., ELAMINE, A. and MAS, A. (2011). Local linear regression for functional data. *Ann. Inst. Statist. Math.* **63** 1047–1075. [MR2822967](#)
- [6] BORGGAARD, C. and THODBERG, H. H. (1992). Optimal minimal neural interpretation of spectra. *Analytical Chemistry* **64** 545–551.
- [7] BURBA, F., FERRATY, F. and VIEU, P. (2009). k -Nearest Neighbour method in functional nonparametric regression. *J. Nonparametr. Stat.* **21** 453–469. [MR2571722](#)
- [8] FERRATY, F., VAN KEILEGOM, I. and VIEU, P. (2010). On the Validity of the Bootstrap in Non-Parametric Functional Regression. *Scand. J. Statist.* **37** 286–306. [MR2682301](#)
- [9] FERRATY, F. and VIEU, P. (2002). The functional nonparametric model and application to spectrometric data. *Computation. Stat.* **17** 545–564. [MR1952697](#)
- [10] FERRATY, F., MAS, A. and VIEU, P. (2007). Nonparametric regression on functional data: inference and practical aspects. *Aust. N. Z. J. Stat.* **49** 267–286. [MR2396496](#)
- [11] FERRATY, F. and VIEU, P. (2006). *Nonparametric functional data analysis. Theory and practice*. Springer, New York. [MR2229687](#)
- [12] GIJBELS, I., VERAVERBEKE, N. and OMELKA, M. (2011). Conditional copulas, association measures and their application. *Comput. Stat. Data An.* **55** 1919–1932. [MR2765054](#)
- [13] HAFNER, C. M. and REZNIKOVA, O. (2010). Efficient estimation of a semi-parametric dynamic copula model. *Comput. Stat. Data An.* **54** 2609–2627. [MR2720464](#)
- [14] HALL, P., MÜLLER, H. G. and YAO, F. (2009). Estimation of functional derivatives. *Ann. Statist.* **37** 3307–3329. [MR2549561](#)
- [15] NELSEN, R. B. (2006). *An Introduction to Copulas*. Springer, New York Second Edition. [MR2197664](#)
- [16] RUPPERT, D. and WAND, M. P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.* **22** 1346–1370. [MR1311979](#)
- [17] SEGERS, J. (2012). Weak convergence of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli* **18** 764–782.
- [18] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. [MR1385671](#)
- [19] VERAVERBEKE, N., OMELKA, M. and GIJBELS, I. (2011). Estimation of a conditional copula and association measures. *Scand. J. Statist.* **38** 766–780. [MR2859749](#)