ESTIMATION OF A CONDITIONAL COPULA AND ASSOCIATION MEASURES

NOËL VERAVERBEKE[1], MAREK OMELKA[2] AND IRÈNE GIJBELS[3]

[1] Center for Statistics, Hasselt University, Belgium

[2] Department of Probability and Statistics, Faculty of Mathematics and Physics, Charles University Prague, Czech Republic

[3] Department of Mathematics and Leuven Statistics Research Center (LStat), Katholieke Universiteit Leuven, Belgium

ABSTRACT. This paper is concerned with studying the dependence structure between two random variables Y_1 and Y_2 conditionally upon a covariate X. The dependence structure is modelled via a copula function, which depends on the given value of the covariate in a general way. Gijbels *et al.* (2011) suggested two nonparametric estimators of the 'conditional' copula and investigated their numerical performances. In this paper we establish the asymptotic properties of the proposed estimators as well as conditional association measures derived from them. Practical recommendations for their use are then discussed.

Keywords and phrases: asymptotic representation; conditional Kendall's tau; empirical copula process; fixed design; random design; smoothing; weak convergence.

RUNNING HEADLINE: Estimation of a conditional copula

1. Introduction

Suppose the observations are independent identically distributed three-dimensional vectors $(Y_{11}, Y_{21}, X_1)^\mathsf{T}, \ldots, (Y_{1n}, Y_{2n}, X_n)^\mathsf{T}$ from $(Y_1, Y_2, X)^\mathsf{T}$ with cumulative distribution function $H(y_1, y_2, x)$. The contributions in this paper are valid for the case of random design (X is a random variable) and fixed regular design (X is not random), with the design density satisfying some assumptions (see also Section 2.3). In this paper we are interested in studying the dependence structure between the variables Y_1 and Y_2 conditionally upon X. We start by motivating the problem via an example. From the World Factbook of the Central Intelligence Agency (CIA) we retrieved a data set consisting of life expectancies of males and females and under-five mortality rates (referring to the total risk over five years) per thousand life births for 221 countries. In Figure 1 we present the life expectancies of males and females using different symbols for countries for which the logarithm of the under-five mortality rates is: (1) less than 1, (2) between 1 and 1.5 and (3) larger than 1.5. From this figure it is clear that there is a relationship between the life expectancies of males and females, but that the strength of the relationship is not the same for the three categories of countries: the datapoints are more scattered in the second category and less scattered in the last category when compared to the first category (with lowest under-five mortality rate).

The life expectancies of males and females (over all countries) are strongly associated with a Kendall's tau equal to 0.86 and a Pearson correlation coefficient of 0.98. The under-five mortality rate does not only strongly influence the life expectancies, but it is often used as one of the measures characterizing the development status of a given country. Thus, it seems quite natural to explore whether the relationship of life expectancies of males and females is the same for countries with 'low' and 'high' under-five mortality rate. The observation made in Figure 1 is further explored in Figure 2(a) where the life expectancies of males and females are plotted in function of the logarithm of the under-five mortality rate from which it is clear that there is a relationship between life expectancies (of males and of females) and the under-five mortality rate in a country. Gijbels *et al.* (2011) suggested two nonparametric estimators of the conditional dependence function and showed how the corresponding association measures may be constructed. Figure 2(b) presents the resulting estimates of the conditional Kendall's tau association measure (see also Section 3.2 of this paper) of life expectancies of males and females as a function of the under-five mortality rate. While the estimator `tau1` is based on original observations, the estimator `tau2` uses observations which are transformed to the uniform, through (4). For both estimators a bandwidth 0.5 was used. A first observation is that the estimated conditional Kendall's tau clearly changes with the under-five mortality rate (i.e. is far from a constant). Further, the most striking feature of both estimators seems to be a sharp increase in the dependence of life expectancies of males and females for countries with under-five mortality rate between $10^{1.4} \doteq 25$ and $10^{1.8} \doteq 63$. This 'local' feature of the conditional association of life expectancies thus complements Kendall's partial correlation coefficient introduced in Kendall (1942), which equals 0.68 and which gives us only a 'global' measure of association of life expectancies when adjusted for under-five mortality rate.

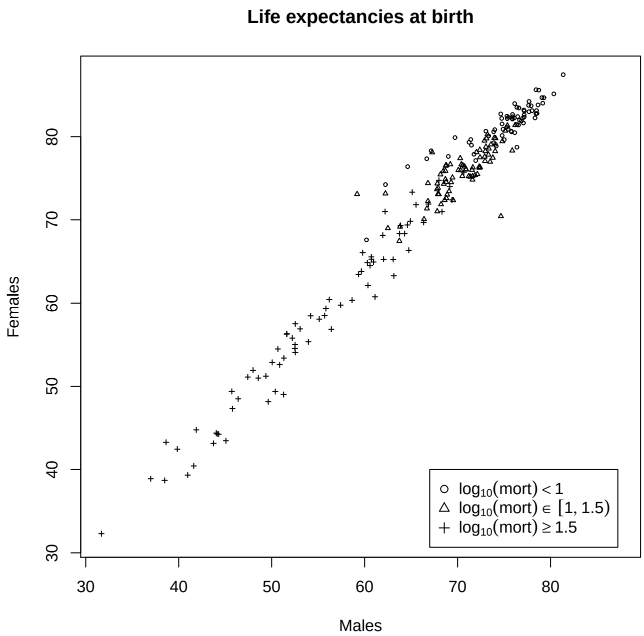

FIGURE 1. Life expectancies of males and females for three different categories of countries (indicated by the symbols "+", "$\triangle$"and "$\circ$") according to the logarithm of the under-five mortality rate (mort).

Let us now formally introduce the setup and the problem. Denote the joint and marginal distribution functions of $(Y_1, Y_2)^\mathsf{T}$, conditionally upon $X = x$, as

$$H_x(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2 \,|\, X = x),$$

$$F_{1x}(y_1) = P(Y_1 \leq y_1 \,|\, X = x), \quad F_{2x}(y_2) = P(Y_2 \leq y_2 \,|\, X = x).$$

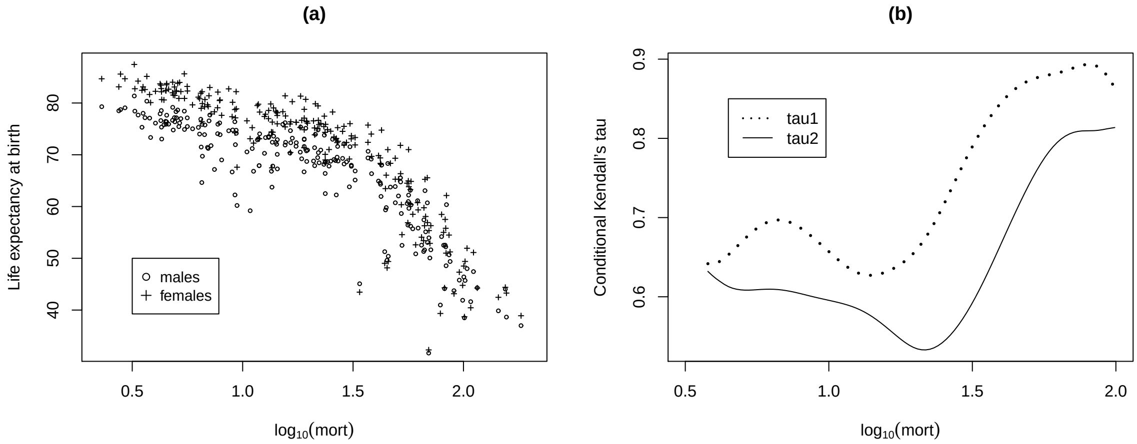

FIGURE 2. Life expectancies of males and females as a function of the logarithm of the under-five mortality rate (mort).

If $F_{1x}(y)$ and $F_{2x}(y)$ are continuous in y, then according to Sklar's theorem (see e.g. Nelsen (2006)) there exists a unique copula C_x which equals

$$C_x(u_1, u_2) = H_x(F_{1x}^{-1}(u_1), F_{2x}^{-1}(u_2)), \tag{1}$$

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

Based on the sample of observations we have the following empirical estimator for $H_x(y_1, y_2)$:

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

where $\{w_{ni}(x, h_n)\}$ is a sequence of weights that smooth over the covariate space and $h_n > 0$ is a bandwidth tending to zero as the sample size increases. The weights do not need to be positive, but throughout the paper we assume that $P\{\min_{1\leq i\leq n} w_{ni}(x, h_n) < 0\}$ tends to zero as n tends to infinity. Further, in (2), $\mathbb{I}\{A\}$ denotes the indicator of an event A. Gijbels *et al.* (2011) suggested the following empirical estimator of the copula C_x $(0 \leq u_1, u_2 \leq 1)$,

$$\begin{aligned} C_{xh}(u_1, u_2) &= H_{xh}\left(F_{1xh}^{-1}(u_1), F_{2xh}^{-1}(u_2)\right) \\ &= \sum_{i=1}^{n} w_{ni}(x, h_n)\, \mathbb{I}\{Y_{1i} \leq F_{1xh}^{-1}(u_1), Y_{2i} \leq F_{2xh}^{-1}(u_2)\}, \end{aligned} \tag{3}$$

where F_{1xh} and F_{2xh} are corresponding marginal distribution functions of H_{xh}. They observed however that this estimator may be strongly biased if any of the marginal distributions is influenced by the covariate. To remove this influence they suggested the following estimator. First, transform the original observations to reduce the effect of the covariate by

$$(\tilde{U}_{1i}, \tilde{U}_{2i})^{\mathsf{T}} = (F_{1X_ig_1}(Y_{1i}), F_{2X_ig_2}(Y_{2i}))^{\mathsf{T}}, \quad i = 1, \ldots, n, \tag{4}$$

where

$$\begin{aligned} F_{1X_ig_1}(y) &= \sum_{j=1}^{n} w_{nj}(X_i, g_{1n})\, \mathbb{I}\{Y_{1j} \leq y\}, \\ F_{2X_ig_2}(y) &= \sum_{j=1}^{n} w_{nj}(X_i, g_{2n})\, \mathbb{I}\{Y_{2j} \leq y\}, \end{aligned}$$

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

$$\tilde{C}_{xh}(u_1, u_2) = \tilde{G}_{xh}\left(\tilde{G}_{1xh}^{-1}(u_1), \tilde{G}_{2xh}^{-1}(u_2)\right), \tag{5}$$

where

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

and $\tilde{G}_{1xh}$ and $\tilde{G}_{2xh}$ are its corresponding marginals.

Gijbels *et al.* (2011) compared the estimators C_{xh} and $\tilde{C}_{xh}$ in a Monte Carlo study and showed that the variances of the estimators are approximately the same, while $\tilde{C}_{xh}$ has in some situations a dramatically lower bias.

In this paper we provide a detailed theoretical study and we compare asymptotic biases and variances of the estimators. We show that while the asymptotic variances of the estimators C_{xh} and $\tilde{C}_{xh}$ are the same, the expression for the asymptotic bias of the estimator $\tilde{C}_{xh}$ consists only of those terms in the asymptotic bias of C_{xh}, which do not include partial derivatives of the conditional marginal distribution functions F_{1x} and F_{2x} with respect to the value of the covariate.

The paper is organised as follows. In Section 2 we state the main theoretical results and discuss the regularity conditions. In Section 3 we show how the results of Section 2 can be applied to find the asymptotic distribution of some conditional measures of association. Further discussions are in Section 4. The proofs of the asymptotic results are given in Appendices A and B in the supplementary material on the journals web site.

2. Main theoretical results

The aim of this section is to establish asymptotic representations for the estimators, to derive weak convergence results, and to evaluate asymptotic bias and variance of the estimators. More precisely, we are interested in the asymptotic properties of the following processes

$$\mathbb{C}_{xn}^{(\mathbb{E})}(u_1,u_2) = \sqrt{nh_n}\big(C_{xh}(u_1,u_2) - C_x(u_1,u_2)\big), \qquad (0 \le u_1,u_2 \le 1), \tag{6}$$

$$\tilde{\mathbb{C}}_{xn}^{(\mathbb{E})}(u_1,u_2) = \sqrt{nh_n}\big(\tilde{C}_{xh}(u_1,u_2) - C_x(u_1,u_2)\big), \qquad (0 \le u_1,u_2 \le 1). \tag{7}$$

All the theoretical results provided in this section are for a fixed but arbitrary value of x, and are uniform with respect to u_1 and u_2.

2.1. The process $\mathbb{C}_{xn}^{(\mathbb{E})}$.

Suppose

$$h_n = O(n^{-1/5}), \qquad n\,h_n \to \infty. \tag{8}$$

Note that (8) allows for $h_n \sim n^{-1/5}$, which is often the optimal rate for bandwidths in nonparametric problems. All other conditions are given in Sections 2.3 and 2.4.

Denote by $C_x^{(1)}(u_1,u_2)$ and $C_x^{(2)}(u_1,u_2)$ the first order partial derivatives of the copula $C_x(u_1,u_2)$ with respect to u_1 and u_2 respectively. The following theorem is proved in Appendix A.

Theorem 1. *Assume* (8), (**R1**)-(**R2**) *and* (**W1**)–(**W6**). *Then it holds uniformly in* $(u_1,u_2) \in [0,1]^2$

$$\mathbb{C}_{xn}^{(\mathbb{E})}(u_1,u_2) = \sqrt{n\,h_n}\sum_{i=1}^{n} w_{ni}(x,h_n)\,\xi_i(u_1,u_2) + o_P(1), \tag{9}$$

where

$$\begin{aligned}\xi_i(u_1,u_2) &= \mathbb{I}\{Y_{1i} \le F_{1x}^{-1}(u_1), Y_{2i} \le F_{2x}^{-1}(u_2)\} - C_x(u_1,u_2)\\ &\quad - C_x^{(1)}(u_1,u_2)\left[\mathbb{I}\{Y_{1i} \le F_{1x}^{-1}(u_1)\} - u_1\right] - C_x^{(2)}(u_1,u_2)\left[\mathbb{I}\{Y_{2i} \le F_{2x}^{-1}(u_2)\} - u_2\right].\end{aligned} \tag{10}$$

Define a process $Z_{xn} = \sqrt{n\,h_n}\sum_{i=1}^{n} w_{ni}(x,h_n)\,\xi_i$, where ξ_i's are given in (10). As (**W5**) holds, typically there exists a finite positive constant V such that

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

Then for all $0 \leq u_1, u_2, v_1, v_2 \leq 1$

$$\operatorname{cov}\left(Z_{xn}(u_1,u_2), Z_{xn}(v_1,v_2)\right) \to V^2 \operatorname{cov}\left(\xi_x(u_1,u_2), \xi_x(v_1,v_2)\right), \quad \text{as } n \to \infty, \tag{11}$$

where

$$\begin{aligned}\xi_x(u_1,u_2) &= \mathbb{I}\{F_{1x}(Y_{1x}) \leq u_1, F_{2x}(Y_{2x}) \leq u_2\} - C_x(u_1,u_2) \\ &\quad - C_x^{(1)}(u_1,u_2)\left[\mathbb{I}\{F_{1x}(Y_{1x}) \leq u_1\} - u_1\right] - C_x^{(2)}(u_1,u_2)\left[\mathbb{I}\{F_{2x}(Y_{2x}) \leq u_2\} - u_2\right].\end{aligned}$$

Thus with the help of (11) it is straightforward to verify the finite dimensional convergence of the process $\{Z_{xn}(u_1,u_2),\, (u_1,u_2) \in [0,1]^2\}$. As the asymptotic tightness of this process is (in a more general setting) verified in Step 1 of the proof of Theorem 1, we deduce that Z_{xn} converges weakly to a Gaussian process Z_x.

Further suppose that there exists H such that $(n\,h_n^5) \to H^2$, with $H \geq 0$. Typically $h_n \sim n^{-1/5}$ so that $H > 0$. In that case, using Taylor expansion and assumption **(R1)** we can approximate the expectation of the limiting process Z_x and find out that (uniformly in (u_1,u_2))

$$\operatorname{E} Z_{xn}(u_1,u_2) = H\,\left[D_K\,\dot{C}_x(u_1,u_2) + \tfrac{E_K}{2}\,B_x(u_1,u_2)\right] + o(1), \tag{12}$$

with D_K and E_K being constants depending on the chosen system of weights $\{w_{ni}\}$ and on the type of the design (see (**W12**) and (**W13**) in Section 2.3) and

$$\begin{aligned}B_x(u_1,u_2) &= \ddot{H}_x(F_{1x}^{-1}(u_1), F_{2x}^{-1}(u_2)) - C_x^{(1)}(u_1,u_2)\,\ddot{F}_{1x}(F_{1x}^{-1}(u_1)) - C_x^{(2)}(u_1,u_2)\,\ddot{F}_{2x}(F_{2x}^{-1}(u_2)) \\ &= \ddot{C}_x(u_1,u_2) + 2\,\dot{C}_x^{(1)}(u_1,u_2)\,\dot{F}_{1x}(F_{1x}^{-1}(u_1)) + 2\,\dot{C}_x^{(2)}(u_1,u_2)\,\dot{F}_{2x}(F_{2x}^{-1}(u_2)) \\ &\quad + C_x^{(1,1)}(u_1,u_2)\,\left[\dot{F}_{1x}(F_{1x}^{-1}(u_1))\right]^2 + C_x^{(2,2)}(u_1,u_2)\,\left[\dot{F}_{2x}(F_{2x}^{-1}(u_2))\right]^2 \\ &\quad + 2\,C_x^{(1,2)}(u_1,u_2)\,\dot{F}_{1x}(F_{1x}^{-1}(u_1))\,\dot{F}_{2x}(F_{2x}^{-1}(u_2)),\end{aligned} \tag{13}$$

where a dot indicates a derivative with respect to the covariate x, e.g. $\dot{F}_z(u_1) = \frac{\partial}{\partial z} F_z(u_1)$, $\ddot{C}_z(u_1,u_2) = \frac{\partial^2}{\partial z^2} C_z(u_1,u_2)$; the symbol $^{(i)}$ indicates a derivative with respect to u_i, e.g. $C_x^{(i,j)}(u_1,u_2) = \frac{\partial^2 C_x(u_1,u_2)}{\partial u_i \partial u_j}$; and $\dot{C}_z^{(i)}(u_1,u_2) = \frac{\partial^2 C_z(u_1,u_2)}{\partial z\, \partial u_i}$, which is a mixture of the above notational rules.

Corollary 1. *If* (11), $(n\,h_n^5) \to H^2$, (**W12**), (**W13**) *and the assumptions of Theorem 1 hold, then the process* $\mathbb{C}_{xn}^{(\mathbb{E})}$ *converges in distribution to a Gaussian process* Z_x*, which can be written* *as*

$$Z_x(u_1,u_2) = V\left\{W_x(u_1,u_2) - C_x^{(1)}(u_1,u_2)W_x(u_1,1) - C_x^{(2)}(u_1,u_2)W_x(1,u_2)\right\} + R_x(u_1,u_2)$$

where W_x *is a bivariate Brownian bridge on* $[0,1]^2$ *with covariance function*

$$\mathrm{E}\left[W_x(u_1,u_2)W_x(v_1,v_2)\right] = C_x(u_1\wedge v_1, u_2\wedge v_2) - C_x(u_1,u_2)C_x(v_1,v_2). \tag{14}$$

and

$$R_x(u_1,u_2) = H\left[D_K\,\dot C_x(u_1,u_2) + \tfrac{E_K}{2}\,B_x(u_1,u_2)\right]. \tag{15}$$

Proof. The proof follows from Theorem 1 and the reasoning given above. □

The constants V, D_K and E_K in general also depend on x, but for simplicity this is not made explicit in the notations.

It should be mentioned that to prove Corollary 1 it is only needed that assumptions (**W12**) and (**W13**) hold without supremum (for $z = x$) and for $a_n = h_n$.

2.2. **The process $\tilde{\mathbb{C}}_{xn}^{(\mathbb{E})}$.** In the following we suppose that for $j = 1,2$

$$\sqrt{n\,h_n}\,g_{jn}^2 = O(1), \quad \tfrac{h_n}{g_{jn}} = O(1), \quad n\,\min(h_n, g_{1n}, g_{2n}) \to \infty. \tag{16}$$

Note that (16) allows for the same rates of h_n as in (8). Further, $h_n \sim n^{-1/5}$ implies that $g_{jn} \sim n^{-1/5}$ for $j = 1,2$ as well. All other conditions are given in Sections 2.3 and 2.4.

Theorem 2. *Assume* (16), **(W1)**–**(W13)** *and* **($\tilde{\mathbf{R}}$1)**–**($\tilde{\mathbf{R}}$3)**, *for* $i = 1,\ldots,n$, *put* $(U_{1i},U_{2i})^{\mathsf T} = (F_{1X_i}(Y_{1i}), F_{2X_i}(Y_{2i}))^{\mathsf T}$, *then uniformly in* (u_1,u_2)

$$\tilde{\mathbb{C}}_{xn}^{(\mathbb{E})}(u_1,u_2) = \sqrt{n\,h_n}\sum_{i=1}^n w_{ni}(x,h_n)\,\tilde\xi_i(u_1,u_2) + o_P(1), \tag{17}$$

where

$$\begin{aligned}\tilde\xi_i(u_1,u_2) &= \mathbb{I}\{U_{1i}\le u_1, U_{2i}\le u_2\} - C_x(u_1,u_2)\\ &\quad - C_x^{(1)}(u_1,u_2)\left[\mathbb{I}\{U_{1i}\le u_1\} - u_1\right] - C_x^{(2)}(u_1,u_2)\left[\mathbb{I}\{U_{2i}\le u_2\} - u_2\right],\end{aligned} \tag{18}$$

Similarly as in Section 2.1 we can state the following corollary.

Corollary 2. *If* (11), $(n\,h_n^5) \to H^2$ *and the assumptions of Theorem 2 hold, then the process* $\tilde{\mathbb{C}}_{xn}^{(\mathbb{E})}$ *converges in distribution to a Gaussian process* $\tilde{Z}_x$*, which can be written as*

$$\tilde{Z}_x(u_1, u_2) = V\left\{W_x(u_1, u_2) - C_x^{(1)}(u_1, u_2)W_x(u_1, 1) - C_x^{(2)}(u_1, u_2)W_x(1, u_2)\right\} + \tilde{R}_x(u_1, u_2)$$

where W_x *is a bivariate Brownian bridge on* $[0, 1]^2$ *with covariance function* (14) *and*

$$\tilde{R}_x(u_1, u_2) = H\,\left[D_K\,\dot{C}_x(u_1, u_2) + \tfrac{E_K}{2}\,\tilde{B}_x(u_1, u_2)\right], \tag{19}$$

with $\tilde{B}_x(u_1, u_2) = \ddot{C}_x(u_1, u_2)$*.*

Thus comparing the limiting processes Z_x and $\tilde{Z}_x$ from Corollary 1 and 2 we see that the only difference is in the bias terms. This difference is a consequence of different random variables that are involved in the Bahadur representations of the processes $\sqrt{n\,h_n}\,(\tilde{C}_{xh} - C_x)$ and $\sqrt{n\,h_n}\,(C_{xh} - C_x)$. The original observations $(Y_{1i}, Y_{2i})^{\mathsf{T}}$ in (10) are replaced by the unobserved $(U_{1i}, U_{2i})^{\mathsf{T}}$ in (18). The key point is that the conditional marginal distributions of $(U_{1i}, U_{2i})^{\mathsf{T}}$ are uniform for each value of the covariate X_i and thus do not depend on the values of the covariate, which results in a much simpler expression for the asymptotic bias given in (19).

Remark 1. There is no guarantee that the asymptotic bias expression for the estimator $\tilde{C}_{xh}$ given by (19) is always closer to zero than that for C_{xh} given in expression (15). Suppose for simplicity that $D_K = 0$, which holds for example for a local linear system of weights (see Section 2.3). Then $B_x(u_1, u_2)$ of (15) may be closer to zero than $\tilde{B}_x(u_1, u_2)$ if the additional terms in (13) turn out to be of opposite sign of the first term $\ddot{C}_x(u_1, u_2)$. For example, suppose that the covariate is standard normal distributed and we are interested in the point $X = 1$. The copula which joins the margins (Y_1, Y_2) is taken to be a Frank copula with the parameter depending on the value of the covariate $X = z$ as $\theta(z) = 5 + \rho\,\sin(\frac{(z-1)\,\pi}{6})$. Further, the margins are taken to be normal with unit variances and mean functions $\mu_1(z) = \mu_2(z) = \sin(z)$.

Consider two values of the parameter ρ. The case $\rho = 1$ represents a situation where the conditional dependence structure is only very mildly affected by the value of the covariate. The plot of the diagonals of the functions B_x and $\tilde{B}_x$ in Figure 3(a) clearly indicates that in terms of bias the estimator $\tilde{C}_{xh}$ is in this situation strongly preferable. This is further

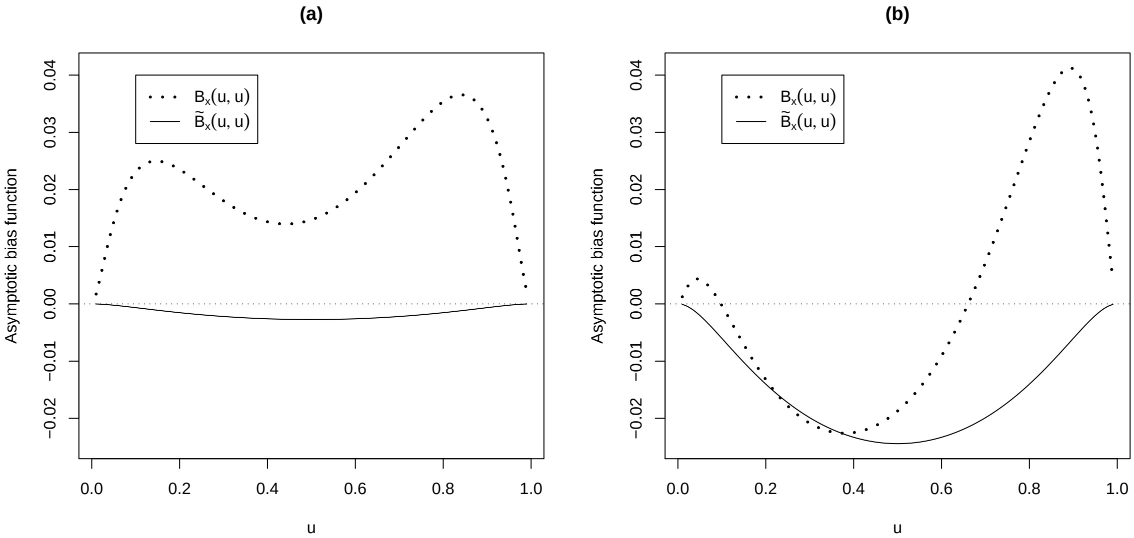

FIGURE 3. Diagonals of the functions B_x and $\tilde{B}_x$ for $\rho = 1$ (a) and $\rho = 7$ (b).

confirmed by calculating $L_2([0,1]^2)$-norms of the functions B_x and $\tilde{B}_x$, which equal 0.014 and 0.001 respectively.

When $\rho = 7$ the conditional dependence structure is strongly influenced by the covariate. Figure 3(b) shows that for this model it is not so easy to judge which estimator should be preferred. At some points B_x is closer to zero and at other points it is the other way around. The $L_2([0,1]^2)$-norms of the functions B_x and $\tilde{B}_x$ now equal 0.011 and 0.012 indicating that the estimator C_{xh} might be slightly preferable if the interest is in estimation of the whole copula function and the mean integrated squared error is taken as the criterion for the quality of the estimate.

Our experience is that it is rather difficult to construct models where the estimator C_{xh} is (more than slightly) preferable to $\tilde{C}_{xh}$. In such models, both conditional marginals as well as the conditional dependence structure have to be strongly dependent on the value of the covariate. Further it must be the case that by a 'lucky coincidence' the additional terms in (13) help to reduce the effect of $\ddot{C}_x$. As this is difficult to predict, one stays on the safe side by using the estimator $\tilde{C}_{xh}$.

Remark 2. The proofs of Theorems 1 and 2 are rather technical and are given in Appendices A and B of the supplementary material posted on the journals web site. While the proof of the weak convergence of the empirical process associated with H_{xh} is quite straightforward, the difficulties in copula estimation arise from the fact that the empirical distribution function estimator H_{xh} is evaluated at the estimated quantiles $F_{1xh}^{-1}(u_1)$ and $F_{2xh}^{-1}(u_2)$ leading to a double stochastic structure of the estimator C_{xh}. Similarly as in Omelka *et al.* (2009) we find the decomposition (A7) useful. Now Theorem 2.11.1 of van der Vaart & Wellner (1996) is used to show that the first term $A_n^{h_n}$ is negligible, while an adaptation of Lemma 4 of Omelka *et al.* (2009) helps to treat the term $C_n^{h_n}$. The proof of Theorem 2 is even more involved as the estimated conditional marginal distributions used in transformation (4) differ for $i = 1, \dots, n$. A crucial point of the proof is that thanks to assumption (**W11**) it is sufficient to consider only such transformations $(F_{1X_i g_1}, F_{2X_i g_2})$ whose X_i's are in a neighbourhood of the point x with radius of order $O(h_n)$. Thus we can prepare everything to make use of the methodology developed in Ghoudi & Rémillard (1998), which helps us to tackle the most difficult term $\bar{A}_n$ in (B6). Combining the above, also the other terms $\bar{B}_n$ and E_n can be proved to be asymptotically negligible giving us (B1). Theorem 1 is then used to conclude the proof of Theorem 2.

2.3. **Some common choices of weights.** As the list of conditions on the weights given in Section 2.4 might be rather discouraging in particular for readers who are less interested in technical details, we comment on several commonly used weight schemes.

Assume for concreteness that a kernel density function K has support $[-1, 1]$ and is symmetric and continuously differentiable. Further suppose that $h_n \sim n^{-1/5}$ and $g_{1n} \sim g_{2n} \sim n^{-1/5}$.

It can be shown that for *Nadaraya-Watson weights* (see Nadaraya (1964) or Watson (1964)), which are defined as

$$w_{ni}(x, h_n) = \frac{K(\frac{X_i - x}{h_n})}{\sum_{j=1}^{n} K(\frac{X_j - x}{h_n})}, \qquad i = 1, \dots, n,$$

assumptions **(W1)**–**(W13)** hold, provided

- **(F1)** $f_X = F_X'$ is continuous and positive at the point x,
- **(F2)** $f_X' = F_X''$ is continuous in a neighbourhood of the point x,

where F_X is the (marginal) distribution function of the covariate X.

Another system of weights, very commonly employed, is a *local linear* [LL] system of weights (see e.g. p. 20 of Fan & Gijbels (1996)), which is given by

$$w_{ni}(x,h_n) = \frac{\frac{1}{n\,h_n}\,K(\frac{X_i-x}{h_n})\,\Big(S_{n,2} - \frac{X_i-x}{h_n}\,S_{n,1}\Big)}{S_{n,0}\,S_{n,2} - S_{n,1}^2}, \qquad i = 1,\ldots,n, \tag{20}$$

where

$$S_{n,j} = \frac{1}{n\,h_n}\sum_{i=1}^{n}\left(\tfrac{X_i-x}{h_n}\right)^j K\left(\tfrac{X_i-x}{h_n}\right), \qquad j = 0,1,2.$$

The nice thing about LL weights is that thanks to $\sum_{i=1}^{n} w_{ni}(x,h_n)(X_i-x) = 0$, it is sufficient to assume only **(F1)**.

In a fixed regular design case (see e.g. Müller (1987)), there exists an absolutely continuous distribution function F_X (with associated density f_X) such that $x_i = F_X^{-1}\left(\frac{i}{n+1}\right)$. In this case the design points are ordered, that is $x_1 \le x_2 \ldots \le x_n$. In this setting *Gasser-Müller* [GM] weights (see Gasser & Müller (1979)) are quite popular. Consider fixed, but arbitrary values $x_0 < x_1$ and $x_{n+1} > x_n$. Then GM weights are defined as

$$w_{ni}(x,h_n) = \tfrac{1}{h_n}\int_{s_i}^{s_{i+1}} K(\tfrac{z-x}{h_n})\,dz, \quad \text{where } s_i = (x_i + x_{i-1})/2, \quad i = 1,\ldots,n. \tag{21}$$

In a fixed regular design case, we conjecture that to verify **(W1)**–**(W13)** it is sufficient to assume **(F1)**.

2.4. **Regularity conditions.** Let us denote $b_n = \max\{h_n, g_{1n}, g_{2n}\}$, $I_x^{(n)} = \{i : w_{ni}(x,b_n) \neq 0\}$ and $J_x^{(n)} = [\min_{i\in I_x^{(n)}} X_i, \max_{i\in I_x^{(n)}} X_i]$. Let a_n stand for a sequence of positive constants such that $(n\,a_n) \to \infty$ and $a_n = O(n^{-1/5})$. The following is a listing of assumptions on the system of weights $\{w_{ni};\, i = 1,\ldots,n\}$ in random design. The conditions for a fixed design, may be derived easily by replacing X_i by x_i and omitting the symbol P in the index.

(W1) $\max_{1\le i\le n} |w_{ni}(x,h_n)| = o_P\left(\frac{1}{\sqrt{n\,h_n}}\right)$, **(W2)** $\sum_{i=1}^{n} w_{ni}(x,h_n) - 1 = o_P\left(\frac{1}{\sqrt{n\,h_n}}\right)$,

(W3) $\sum_{i=1}^{n} w_{ni}(x,h_n)(X_i - x) = O_P\left(\frac{1}{\sqrt{n\,h_n}}\right)$, **(W4)** $\sum_{i=1}^{n} w_{ni}(x,h_n)(X_i - x)^2 = O_P\left(\frac{1}{\sqrt{n\,h_n}}\right)$,

(W5) $\sum_{i=1}^{n} w_{ni}^2(x,h_n) = O_P\left(\frac{1}{n\,h_n}\right)$, **(W6)** $\left(\max_{i\in I_x^{(n)}} X_i - \min_{i\in I_x^{(n)}} X_i\right) = o_P(1)$,

(W7) $\displaystyle\sum_{i=1}^{n} |w_{ni}(x,h_n)| = O_P(1),$ **(W8)** $\displaystyle\sup_{z\in J_x^{(n)}} \left| \sum_{i=1}^{n} w_{ni}(z,g_{jn}) - 1 \right| = o_P\left(g_{jn}^2\right),$

(W9) $\displaystyle\sup_{z\in J_x^{(n)}} \sum_{i=1}^{n} [w_{ni}(z,g_{jn})]^2 = O_P\left(\tfrac{1}{n\,g_{jn}}\right),$ **(W10)** $\displaystyle\sup_{z\in J_x^{(n)}} \sum_{i=1}^{n} [w'_{ni}(z,g_{jn})]^2 = O_P\left(\tfrac{1}{n\,g_{jn}^3}\right),$

(W11) $\displaystyle\exists_{C<\infty}\ \mathrm{P}\left[\sup_{z\in J_x^{(n)}} \max_{1\le i\le n} |w_{ni}(z,h_n)\, \mathbb{I}\{|X_i - z| > C\,h_n\}| > 0\right] = o(1),$

(W12) $\displaystyle\exists_{D_K<\infty}\,\forall_{a_n}\ \sup_{z\in J_x^{(n)}} \left| \sum_{i=1}^{n} w_{ni}(z,a_n)(X_i - z) - a_n^2\, D_K \right| = o_P\left(a_n^2\right),$

(W13) $\displaystyle\exists_{E_K<\infty}\,\forall_{a_n}\ \sup_{z\in J_x^{(n)}} \left| \sum_{i=1}^{n} w_{ni}(z,a_n)(X_i - z)^2 - a_n^2\, E_K \right| = o_P\left(a_n^2\right),$

where $w'_{ni}(z,g_{jn})$ denotes the derivative with respect to z.

Conditions (**W7**)–(**W13**) make a finer control on the behaviour of the weights not only at the point x but also in a (shrinking) neighbourhood of this point. This better control is needed to justify that the transformation (4) is 'painless'. Nevertheless, as argued in the previous section, these conditions hold under usual regularity conditions on the distribution of the covariate X.

Further, we require the conditional copula C_z and the conditional marginals F_{1z} and F_{2z} to satisfy:

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

(R2) The first order partial derivatives $C_x^{(1)}$, $C_x^{(2)}$ with respect to u_1 and u_2 respectively are continuous on $[0,1]^2 \setminus \{(0,0),(0,1),(1,0),(1,1)\}$

($\tilde{\mathbf{R}}$1) $\dot{C}_z(u_1,u_2) = \frac{\partial}{\partial z} C_z(u_1,u_2)$, $\ddot{C}_z(u_1,u_2) = \frac{\partial^2}{\partial z^2} C_z(u_1,u_2)$ exist and are continuous as functions of (z,u_1,u_2), where z takes value in a neighbourhood of x;

($\tilde{\mathbf{R}}$2) The functions $C_z^{(1)}(u_1,u_2)$ and $C_z^{(2)}(u_1,u_2)$ are uniformly continuous in $(z,u_1,u_2) \in U(x) \times ([0,1]^2 \setminus \{(0,0),(0,1),(1,0),(1,1)\})$, where $U(x)$ is a neighbourhood of the point x.

($\tilde{\mathbf{R}}$3) For $j = 1,2$: $F_{jz}(F_{jz}^{-1}(u))$, $\dot{F}_{jz}(F_{jz}^{-1}(u))$, $\ddot{F}_{jz}(F_{jz}^{-1}(u))$ are continuous as functions of (z,u) for z in a neighbourhood of x, where $\dot{F}_{jz}(y) = \frac{\partial}{\partial z} F_{jz}(y)$, $\ddot{F}_{jz}(y) = \frac{\partial^2}{\partial z^2} F_{jz}(y)$.

3. Conditional measures of association

In many situations researchers try to characterize the dependence between variables by just one number. As copulas are invariant to strictly increasing transformations of the marginal distributions, most traditional nonparametric measures of dependence may be expressed as functionals of copulas.

In the following let us observe the vector $(Y_1, Y_2, X)^{\mathsf{T}}$ and put F_1 and F_2 for the (unconditional) marginal distributions of Y_1 and Y_2. Further, let C be the (unconditional marginal) copula corresponding to Y_1 and Y_2. Next, $\hat{C}_{xh}$ will stand either for the estimator C_{xh} or $\tilde{C}_{xh}$ and similarly for the bias function $\hat{R}$. Finally, $(Z_{1i}, Z_{2i})^{\mathsf{T}}$ will be either the original observations $(Y_{1i}, Y_{2i})^{\mathsf{T}}$ or the transformed ones $(\tilde{U}_{1i}, \tilde{U}_{2i})^{\mathsf{T}}$.

3.1. **Blomqvist beta.** Blomqvist (1950) proposed and studied the following simple measure of association

$$\beta = \mathrm{P}\left[\left(Y_1 - F_1^{-1}(0.5)\right)\left(Y_2 - F_2^{-1}(0.5)\right) > 0\right] - \mathrm{P}\left[\left(Y_1 - F_1^{-1}(0.5)\right)\left(Y_2 - F_2^{-1}(0.5)\right) < 0\right],$$

which is often also called *the medial correlation coefficient*. Let C be the copula corresponding to Y_1 and Y_2. Then β can be expressed simply as $\beta = 4\,C(0.5, 0.5) - 1$ (see pp.182–183 of Nelsen (2006)). In the presence of a covariate we can consider Blomqvist beta conditionally on $X = x$ and define it as $\beta_x = 4\,C_x(0.5, 0.5) - 1$. The considered estimator is then

$$\hat{\beta}_{xh} = 4\,\hat{C}_{xh}(0.5, 0.5) - 1,$$

With the help of Corollary 1 (or 2) it is straightforward to show that $\sqrt{n\,h_n}(\hat{\beta}_{xh} - \beta_x)$ is asymptotically normal with the variance equal to $16\,C_x(0.5, 0.5)(1 - C_x(0.5, 0.5))$ and the mean $4\,\hat{R}(0.5, 0.5)$.

3.2. **Kendall's tau.** Kendall's tau is definitely one of the most popular nonparametric measures of association. Gijbels *et al.* (2011) suggested to estimate its conditional version by

$$\hat{\tau}_n(x) = \frac{4}{1 - \sum_{i=1}^{n} w_{ni}^2(x, h_n)} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ni}(x, h_n)\, w_{nj}(x, h_n)\, \mathbb{I}\{Z_{1i} < Z_{1j}, Z_{2i} < Z_{2j}\} - 1.$$

With the help of assumption (**W5**) this equation may be further rewritten as

$$\hat{\tau}_n(x) = \tfrac{4}{1 - \sum_{i=1}^{n} w_{ni}^2(x,h_n)} \iint \hat{C}_{xh}\, d\hat{C}_{xh} - \tfrac{1 + 3\sum_{i=1}^{n} w_{ni}^2(x,h_n)}{1 - \sum_{i=1}^{n} w_{ni}^2(x,h_n)} = 4 \iint \hat{C}_{xh}\, d\hat{C}_{xh} - 1 + O_P(\tfrac{1}{n\,h_n}).$$

Provided that the estimator $\hat{C}_{xh}$ is $\sqrt{nh_n}$-weakly convergent, the asymptotic normality of $\sqrt{nh_n}(\hat{\tau}_n(x) - \tau(x))$ follows by Hadamard differentiability of the functional $C \mapsto \iint C\,dC$ at the point $C = C_x$ (tangentially to the set of continuous functions on $[0,1]^2$). This Hadamard differentiability is verified in Lemma 1 at the end of Section 3. Finally, the mean μ of the asymptotic distribution of $\sqrt{n\,h_n}(\hat{\tau}_n(x) - \tau(x))$ equals

$$\mu = \iint \hat{R}(u_1,u_2)\,dC_x(u_1,u_2) + \iint C_x(u_1,u_2)\,dR(u_1,u_2) = 2\iint \hat{R}(u_1,u_2)\,dC_x(u_1,u_2),$$

where we have used the following formula for integration by parts

$$\int_0^1\int_0^1 C_x(u_1,u_2)\,d\alpha(u_1,u_2)$$
$$= \alpha(1,1) + \int_0^1\int_0^1 \alpha(u_1,u_2)\,dC_x(u_1,u_2) - \int_0^1 \alpha(u,1)\,du - \int_0^1 \alpha(1,u)\,du, \quad (22)$$

which can be taken as a definition of the integral $\iint C_x\,d\alpha$ when the function α is continuous on $[0,1]^2$.

3.3. **Other measures of association.** There are some other measures of association that may be expressed as functionals of a copula (see Chapter 5 of Nelsen (2006)). Among others let us mention

Spearman's rho	$\rho = 12\iint C(u_1,u_2)\,du_1\,du_2 - 3.,$
Gini's coefficient	$\gamma = 2\iint \left[\lvert u_1+u_2-1\rvert - \lvert u_1-u_2\rvert\right]\,dC(u_1,u_2),$
dependence index	$\Phi^2 = 90\iint \left[C(u_1,u_2) - u_1\,u_2\right]^2\,du_1 du_2.$

We can consider conditional versions of any of these measures simply by substituting the conditional copula C_x for C. The asymptotic distribution of the estimators of these conditional measures can be derived similarly as discussed for Kendall's tau.

3.4. **Hadamard differentiability of Kendall's tau.** The following lemma is a slight generalization of Lemma 3.9.17 of van der Vaart & Wellner (1996).

Lemma 1. *The map $\phi : C \to \iint C\,dC$ is Hadamard-differentiable at every point C, which is a copula, tangentially to the set of functions that are continuous on $[0,1]^2$. The derivative is given by*

$$\phi'(\alpha) = \int C\,d\alpha + \int \alpha\,dC,$$

where $\iint C\, d\alpha$ *is defined via integration by parts (see formula* (22)*) if* α *is not of bounded variation.*

Proof. Let $\alpha_t \to \alpha$ (uniformly on $[0,1]^2$) and put $C_t = C + t\,\alpha_t$. As the variation of copula C equals one, we can consider only perturbations such that C_t is of variation bounded by 2.

Expanding

$$\iint C_t\, dC_t = \iint C\, dC + t \iint \alpha_t\, dC + t \iint C\, d\alpha_t + t \iint \alpha_t\, d(C_t - C)$$

gives us

$$\frac{\iint C_t\, dC_t - \iint C\, dC}{t} - \iint \alpha\, dC - \iint C\, d\alpha$$
$$= \iint (\alpha_t - \alpha)\, dC + \iint C\, d(\alpha_t - \alpha) + \iint \alpha_t\, d(C_t - C).$$

The first term on the right-hand side of the above equation converges to zero as $\alpha_t \to \alpha$ uniformly and C is of bounded variation. For the second term we apply the same reasoning after using the integration by parts (see formula (22)). The third term may be further rewritten as

$$\iint \alpha_t\, d(C_t - C) = \iint (\alpha_t - \alpha)\, d(C_t - C) + \iint \alpha\, d(C_t - C).$$

The first term on the right-hand side of the above equation converges to zero, as $\alpha_t \to \alpha$ uniformly and $C_t - C$ is of variation bounded by 3. Finally, we will use the continuity of the function α to bound the second term. For a given $\varepsilon > 0$ let us find partitions $0 = t_0 < t_1 < \ldots < t_{m_1} = 1$ and $0 = s_0 < s_1 < \ldots < s_{m_2} = 1$, such that α varies less than ε on each rectangle $[t_{i-1}, t_i] \times [s_{j-1}, s_j]$. Let $\tilde{\alpha}$ be the discretization that is constant and equal to $\alpha(t_{i-1}, s_{j-1})$ on $[t_{i-1}, t_i] \times [s_{j-1}, s_j]$. Then we can bound

$$\begin{aligned} \left| \int \alpha\, d(C_t - C) \right| &\leq 3\, \|\alpha - \tilde{\alpha}\|_\infty + \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} |\alpha(t_{i-1}, s_{j-1})| \left| \iint_{[t_{i-1}, t_i] \times [s_{j-1}, s_j]} d(C_t - C) \right| \\ &\leq 3\,\varepsilon + 4\, m_1 m_2 \|\alpha\|_\infty \|C_t - C\|_\infty \to 3\,\varepsilon, \end{aligned}$$

where $\|\cdot\|_\infty$ stands for the supremum norm. As we can choose ε arbitrarily small, we have just proved that $\int \alpha\, d(C_t - C)$ converges to zero which concludes the proof of the lemma. $\square$

4. Further discussion

We derived the asymptotic distribution of the estimators of the conditional copula proposed in Gijbels *et al.* (2011). As illustrated in Remark 1 our results clearly indicate why the estimator $\tilde{C}_{xh}$ may have a dramatically smaller bias than the 'straightforward' estimator C_{xh}. On the other hand, the estimator C_{xh} involves the choice of only one smoothing parameter (in contrast to three smoothing parameters needed for $\tilde{C}_{xh}$) and hence might be preferable in situations where the conditional marginal distributions do not change with the covariate.

For a triplet (Y_1, Y_2, X) of random variables several natural questions arise: (1). What is the relation between Y_1 and X?; and between Y_2 and X?; (2). Is there a relation between Y_1 and Y_2?; (3) Is there a relation between Y_1 and Y_2, when X is accounted for?; (4). What is the relation between Y_1 and Y_2 when X is accounted for? In this paper we deal with the last question, but question (1) automatically pops up when we wonder whether the conditional marginal distributions change or not with the covariate. Question (2) is for example asking for estimation of an unconditional copula function; see Omelka *et al.* (2009). Question (3) means that we would look into tests for conditional independence. Further interesting open issues are testing problems related to conditional dependencies, such as for example testing for constancy of the conditional Kendall's tau. These kind of questions are quite challenging and the subject of current research by some of the authors.

A general strategy for conditional copula estimation in not very large samples may be as follows. First, check the scatterplots of the pairs $(X, Y_1)^\mathsf{T}$ and $(X, Y_2)^\mathsf{T}$. If there is no obvious pattern, then the estimator C_{xh} may be used. If this is not the case, we recommend to try to transform the variables Y_1 and Y_2 such that the influence of the covariate on the conditional marginal distributions is suppressed. This might be done in several ways. The transformation (4) is very general and in view of Theorem 2 it cannot be improved if we aim at eliminating the effect of the covariate on the marginals. The price we have to pay is that we have to specify two new bandwidths g_{1n} and g_{2n}. Fortunately, the Monte Carlo simulation results of Gijbels *et al.* (2011) indicate that the rules for bandwidth selection in nonparametric regression may be employed or if h_n is already fixed then using $g_{1n} = g_{2n} = h_n$ for $\tilde{C}_{xh}$ usually results in an estimator which is at least as good as C_{xh}.

Alternatively, in small samples we may try to stabilize the estimator by specifying a simple parametric model for the pairs $(X, Y_1)^\mathsf{T}$ and $(X, Y_2)^\mathsf{T}$. For instance, suppose simple linear regression models

$$Y_{1i} = \alpha_0 + \alpha_1 X_i + \varepsilon_{1i}, \quad Y_{2i} = \beta_0 + \beta_1 X_i + \varepsilon_{2i}, \tag{23}$$

where ε_{1i} and ε_{2i} are independent of X_i. Then it seems natural to replace the original observations $(Y_{1i}, Y_{2i})^{\mathsf{T}}$ with the estimated residuals from models (23). As the estimators of the unknown parameters converge at rate $n^{-1/2}$, the estimator based on the estimated residuals may be shown to be asymptotically equivalent with the one based on the unobserved residuals $(\varepsilon_{1i}, \varepsilon_{2i})^{\mathsf{T}}$ and thus the main effect of the covariate on the marginal distributions is usually removed.

A further step towards the general transformation (4) may be to assume nonparametric models

$$Y_{1j} = m_1(X) + \varepsilon_{1i}, \quad Y_{2j} = m_2(X) + \varepsilon_{2i}, \qquad i = 1, \ldots, n,$$

where $\varepsilon_{1i}, \varepsilon_{2i}$ are independent of X and $m_1(\cdot)$ and $m_2(\cdot)$ are unknown, but sufficiently smooth functions. Let $\hat{m}_1$ and $\hat{m}_2$ be corresponding nonparametric estimators. Then in view of the results of Akritas & van Keilegom (2001) and methods to prove these results, we conjecture that the estimator C_{xh} based on the estimated residuals

$$(\hat{\varepsilon}_{1i}, \hat{\varepsilon}_{2i})^{\mathsf{T}} = \left(Y_{1i} - \hat{m}_1(X_i), Y_{2i} - \hat{m}_2(X_i)\right)^{\mathsf{T}}, \qquad i = 1, \ldots, n \tag{24}$$

will have the same asymptotic properties as the estimator based on the unobserved $(\varepsilon_{1i}, \varepsilon_{2i})^{\mathsf{T}}$. However, such an estimation procedure involves the same number of smoothing parameters and the model does not adjust for heteroscedasticity. Allowing (24) for possible heteroscedasticity would result even in five smoothing parameters.

We are only at the beginning of the theoretical as well as applied research on conditional copulas. Comparing with the well explored field of nonparametric regression several research issues arise. We only mention a few of them.

- Our results are only pointwise in the covariate x. Uniform results are surely of interest. As the distributions of the estimator of the conditional copula at different points are asymptotically independent, we cannot hope that there may be convergence of $\{\mathbb{C}_{xn}^{(\mathbb{E})}, x \in I\}$ to a limiting Gaussian process (with parameters (x, u_1, u_2)). But we may hope that rates for uniform consistency similar as in nonparametric regression may be obtained.

- As usual in statistics, one would like to construct confidence intervals for quantities of interest, e.g. Blomqvist beta or Kendall's tau. Our experience is that it is not so difficult to estimate the variance, but it is not at all clear how to take into consideration possible bias, which is extremely difficult to estimate. One obvious way to kill the bias is to choose a bandwidth h_n such that $n\,h_n^5 \to 0$, implying $H = 0$ and the asymptotic biases in (15) and (19) diminish. However such a choice excludes theoretically optimal bandwidths, and relies on asymptotic (large sample) results.
- The bandwidth selection problem is in fact a completely unexplored area in this setup. Although a lot of work has been done in nonparametric regression, it is not at all straightforward how the methods should be modified or extended for conditional copula estimation.

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 Network P6/03 of the Belgian State (Belgian Science Policy). This work was started while Marek Omelka was a postdoctoral researcher at the Katholieke Universiteit Leuven and the Universiteit Hasselt within the IAP Research Network. The support of Project MSM 0021620839 of Ministry of Education, Youth and Sport of the Czech Republic is also highly appreciated. The first author acknowledges support from research grant MTM 2008-03129 of the Spanish Ministerio de Ciencia e Innovacion. The third author gratefully acknowledges support from the GOA/07/04-project of the Research Fund KULeuven.

Supporting material.

Additional supporting material may be found in the online version of this article:

Appendix A – Proof of Theorem 1

Appendix B – Proof of Theorem 2.

References

Akritas, M. G. & van Keilegom, I. (2001). Non-parametric estimation of the resiudal distribution. *Scand. J. Statist.* **28**, 549–567.

Blomqvist, N. (1950). On a measure of dependence between two random variables. *Ann. Math. Statist.* **21**, 593–600.

Fan, J. & Gijbels, I. (1996). *Local polynomial modelling and its applications.* Chapman & Hall.

Gasser, T. & Müller, H.-G. (1979). Kernel estimates of regression functions. In T. Gasser & M. Rosenblatt, eds., *Lecture notes in mathematics 757.* Springer, New York, pp. 23–68.

Ghoudi, K. & Rémillard, B. (1998). Empirical processes based on pseudo-observations. In B. Szyskowitz, ed., *Asymptotic Methods in Probability and Statistics, A Volume in Honour of Miklós Csörgő.* Elsevier, Amsterdam, pp. 171–197.

Gijbels, I., Veraverbeke, N. & Omelka, M. (2011). Conditional copulas, association measures and their application. *Comput. Stat. Data An.* **55**, 1919–1932.

Kendall, M. G. (1942). Partial rank correlation. *Biometrika* **32**, 277–283.

Müller, H.-G. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *J. Amer. Statist. Assoc.* **82**, 231–238.

Nadaraya, E. A. (1964). On estimating regression. *Theor. Prob. Appl.* **9**, 141–142.

Nelsen, R. B. (2006). *An introduction to copulas.* Springer, New York. Second Edition.

Omelka, M., Gijbels, I. & Veraverbeke, N. (2009). Improved kernel estimation of copulas: Weak convergence and goodness-of-fit testing. *Ann. Statist.* **37**, 3023–3058.

Singh, R. S. (1975). On the Glivenko-Cantelli theorem for weighted empiricals based on random variables. *Ann. Probab.* **3**, 371–374.

van der Vaart, A. W. & Wellner, J. A. (1996). *Weak convergence and empirical processes.* Springer, New York.

van der Vaart, A. W. & Wellner, J. A. (2007). Empirical processes indexed by estimated functions. In E. A. Cator, G. Jongbloed, C. Kraaikamp, H. P. Lopuhaä & J. A. Wellner, eds., *IMS lecture notes monograph series 2007: Asymptotics: Particles, processes and inverse problems*, vol. 55. Institute of Mathematical Statistics, pp. 234–252.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā A* **26**, 359–372.

Corresponding author: Irène Gijbels, Department of Mathematics and Leuven Statistics Research Center (LStat), Katholieke Universiteit Leuven, Celestijnenlaan 200B, Box 2400, B-3001 Leuven (Heverlee), Belgium; E-mail: irene.gijbels@wis.kuleuven.be

Estimation of a conditional copula and association measures

Noël Veraverbeke, Marek Omelka and Irène Gijbels

Supplementary material

Appendix A – Proof of Theorem 1

Some preliminary considerations.

Random design versus fixed design. We prove Theorem 1 for fixed design. This is justified as follows. Let $\varepsilon > 0$ be given. With the help of (**W1**) we find a sequence ε_n of constants going to zero such that for all sufficiently large n

$$\mathrm{P}\left(\max_{1\le i\le n} |w_{ni}(x,h_n)| \le \tfrac{\varepsilon_n}{\sqrt{n\,h_n}}\right) \ge 1 - \tfrac{\varepsilon}{6}. \tag{A1}$$

Let us denote (the sequence of) events

$$A_{1n} = \left[\max_{1\le i\le n} |w_{ni}(x,h_n)| \le \tfrac{\varepsilon_n}{\sqrt{n\,h_n}}\right], \qquad n = 1,2,\ldots.$$

Similarly we construct sequences of events $A_{2n},\ldots,A_{6n}$, which are related to the other conditions (**W2**)–(**W6**) needed in Theorem 1. The proof for the random design can then go conditionally on the sequence of events $A_n = A_{1n} \cap \ldots \cap A_{6n}$, as A_n has the probability greater than $1-\varepsilon$ for all sufficiently large n.

Thus in the following we for simplicity write x_i instead of X_i.

Estimation of a conditional marginal distribution. Using the well known inequality of Singh (1975) and assumption (**W5**) we deduce that for each $\varepsilon > 0$

$$\lim_{n\to\infty} \mathrm{P}\left[\sup_{y\in\mathbb{R}} |F_{jxh}(y) - \mathrm{E}\,F_{jxh}(y)| > \varepsilon\right] = 0, \qquad j = 1,2. \tag{A2}$$

Further by assumptions (**W6**) and **(R1)**

$$\lim_{n\to\infty} \sup_{y\in\mathbb{R}} |\mathrm{E}\,F_{jxh}(y) - F_{jx}(y)| = 0, \qquad j = 1,2. \tag{A3}$$

Combining (A2) and (A3) yields

$$\lim_{n\to\infty} \mathrm{P}\left(\sup_{y\in\mathbb{R}} |F_{jxh}(y) - F_{jx}(y)| > \varepsilon\right) = 0, \qquad j = 1,2. \tag{A4}$$

Thus by assumption (**W1**) and standard arguments we prove that for each $\varepsilon > 0$ (for $j = 1,2$)

$$\begin{aligned}\mathrm{P}\left[\sup_u \left(F_{jxh}^{-1}(u) - F_{jx}^{-1}(u+\varepsilon)\right) \geq 0\right] &\leq \mathrm{P}\left[\sup_u \left(u - F_{jxh}(F_{jx}^{-1}(u+\tfrac{\varepsilon}{2}))\right) \geq 0\right] \\ \leq \ \mathrm{P}&\left[\sup_{y\in\mathbb{R}} (F_{jx}(y) - F_{jxh}(y)) \geq \tfrac{\varepsilon}{2}\right] \to 0, \text{ for } n \to \infty. \end{aligned}\tag{A5}$$

Similarly, we can show

$$\lim_{n\to\infty} \mathrm{P}\left[\inf_u \left(F_{jxh}^{-1}(u) - F_{jx}^{-1}(u-\varepsilon)\right) \leq 0\right] = 0,$$

which together with (A5) yields for each $\varepsilon > 0$

$$\lim_{n\to\infty} \mathrm{P}\left[F_{jx}^{-1}(u-\varepsilon) \leq F_{jxh}^{-1}(u) \leq F_{jx}^{-1}(u+\varepsilon), u \in [0,1]\right] = 1. \tag{A6}$$

Expectation with a substitution of a random function. In the rest of this section the expectations of the form $\mathrm{E}\, f(Y_{i1}, Y_{i2}; F_{1xh}, F_{2xh})$ have to be understood in a way that the functions F_{1xh}, F_{2xh} are fixed (nonrandom) and the expectation is computed only with respect to Y_{i1} and Y_{i2}. Formally

$$\mathrm{E}\, f(Y_{i1}, Y_{i2}; F_{1xh}, F_{2xh}) = \iint f(y_1, y_2; F_{1xh}, F_{2xh})\, dH_{x_i}(y_1, y_2),$$

whenever the integral on the right-hand side exists. The reason for this notation is to simplify the presentation of the proof. This notation as well as the following decomposition apply the ideas of van der Vaart & Wellner (2007).

Decomposition. Let us decompose the copula process $\sqrt{n\,h_n}(C_{xh} - C_x)$ as

$$\sqrt{n\,h_n}\,(C_{xh} - C_x) = A_n^{h_n} + B_n^{h_n} + C_n^{h_n}, \tag{A7}$$

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

$$\begin{aligned} D_n^{h_n}(u_1,u_2) = \sqrt{n\,h_n} \sum_{i=1}^n w_{ni}(x;h_n)\ \Big[&\mathbb{I}\{Y_{1i} \leq F_{1xh}^{-1}(u_1), Y_{2i} \leq F_{2xh}^{-1}(u_2)\} \\ &- \mathbb{I}\{Y_{1i} \leq F_{1x}^{-1}(u_1), Y_{2i} \leq F_{2x}^{-1}(u_2)\}\Big]\ , \end{aligned}\tag{A8}$$

and

$$B_n^{h_n}(u_1,u_2) = \sqrt{n\,h_n}\,\left[\sum_{i=1}^n w_{ni}(x;h_n)\,\mathbb{I}\{Y_{1i} \le F_{1x}^{-1}(u_1), Y_{2i} \le F_{2x}^{-1}(u_2)\} - C_x(u_1,u_2)\right] \quad \text{(A9)}$$

$$C_n^{h_n}(u_1,u_2) = \mathrm{E}\,D_n^{h_n}(u_1,u_2) = \sqrt{n\,h_n}\,\sum_{i=1}^n w_{ni}(x;h_n)\,\big[\mathrm{E}\,\mathbb{I}\{Y_{1i} \le F_{1xh}^{-1}(u_1), Y_{2i} \le F_{2xh}^{-1}(u_2)\}$$
$$-\mathrm{E}\,\mathbb{I}\{Y_{1i} \le F_{1x}^{-1}(u_1), Y_{2i} \le F_{2x}^{-1}(u_2)\}\big]. \quad \text{(A10)}$$

The proof is divided into two steps. In Step 1, we show that the term $A_n^{h_n}$ is asymptotically negligible uniformly in (u_1,u_2). In Step 2 we find the asymptotic representation of the processes $C_n^{h_n}$. This asymptotic representation together with $B_n^{h_n}$ gives us the statement of the theorem.

A1. Step 1 – Asymptotic negligibility of $A_n^{h_n}$. For $(u_1,u_2) \in [0,1]^2$ and G_1, G_2 nondecreasing functions from $\mathbb{R}$ to $[0,1]$ define the stochastic processes

$$Z_{ni}(u_1,u_2,G_1,G_2) = \sqrt{n\,h_n}\,w_{ni}(x,h_n)\,\mathbb{I}\{Y_{1i} \le G_1^{-1}(u_1), Y_{2i} \le G_2^{-1}(u_2)\}, \quad i=1,\ldots,n,$$

and $Z_n = \sum_{i=1}^n Z_{ni}$. Equivalently, we can view the process Z_n as a process indexed by the family of functions from $\mathbb{R}^2$ to $\mathbb{R}$ given by

$$\mathcal{F} = \big\{(w_1,w_2) \mapsto \mathbb{I}\{w_1 \le G_1^{-1}(u_1), w_2 \le G_2^{-1}(u_2)\};$$
$$(u_1,u_2) \in [0,1]^2,\ G_1, G_2 : \mathbb{R} \to [0,1] \text{ nondecreasing}\big\}. \quad \text{(A11)}$$

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

$$D_n^{h_n}(u_1,u_2) = Z_n(f_n) - Z_n(f), \text{ where } f_n = (u_1,u_2,F_{1xh},F_{2xh}),\ f = (u_1,u_2,F_{1x},F_{2x}). \quad \text{(A12)}$$

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

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

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

Lemma 2. *The process $\bar{Z}_n = Z_n - \mathrm{E}\,Z_n$ indexed by $(\mathcal{F},\rho)$ is asymptotically ρ-equicontinuous.*

Proof. It is sufficient to verify the following conditions of Theorem 2.11.1 of van der Vaart & Wellner (1996):

$$\sum_{i=1}^{n} \mathrm{E}\,\|Z_{ni}\|_{\mathcal{F}}^2\, \mathbb{I}\{\|Z_{ni}\|_{\mathcal{F}} > \eta\} \;\to\; 0, \qquad \text{for every } \eta > 0, \tag{A13}$$

$$\sup_{\rho(f,f')<\delta_n} \sum_{i=1}^{n} \mathrm{E}\,\big(Z_{ni}(f) - Z_{ni}(f')\big)^2 \;\to\; 0, \qquad \text{for every } \delta_n \downarrow 0, \tag{A14}$$

$$\int_0^{\delta_n} \sqrt{\log N(\varepsilon, \mathcal{F}, d_n)}\, d\varepsilon \;\xrightarrow{P}\; 0, \qquad \text{for every } \delta_n \downarrow 0, \tag{A15}$$

where $\|.\|_{\mathcal{F}}$ stands for the supremum over the set $\mathcal{F}$ and $N(\varepsilon, \mathcal{F}, d_n)$ is the corresponding covering number of the set $\mathcal{F}$ with a random semimetric d_n given by

$$d_n^2(f,f') = \sum_{i=1}^{n} \big[Z_{ni}(f) - Z_{ni}(f')\big]^2.$$

The first condition (A13) is satisfied as (**W1**) immediately implies $\max_{i=1,\dots,n} \|Z_{ni}\|_{\mathcal{F}} = o_P(1)$. To verify (A14) we make use of assumption (**W5**) and the well known inequality

$$\big|C(u_1,u_2) - C(u_1',u_2')\big| \le |u_1 - u_1'| + |u_2 - u_2'|.$$

Thus with the help of **(R1)** and (**W6**) we can estimate

$$\begin{aligned}
&\sup_{\rho(f,f')<\delta_n} \sum_{i=1}^{n} \mathrm{E}\,\big(Z_{ni}(f) - Z_{ni}(f')\big)^2 \\
&\qquad \le \sup_{\rho(f,f')<\delta_n} n\, h_n \sum_{i=1}^{n} w_{ni}^2(x,h_n) \Big[\Big|F_{1i}(G_1^{-1}(u_1)) - F_{1i}(G_1'^{-1}(u_1'))\Big| \\
&\qquad\qquad\qquad + \Big|F_{2i}(G_2^{-1}(u_2)) - F_{2i}(G_2'^{-1}(u_2'))\Big|\Big] \\
&\qquad \le O(1) \sup_{\rho(f,f')<\delta_n} \max_{i\in I_x^{(n)}} \Big[|F_{1i}(G_1^{-1}(u_1)) - F_{1x}(G_1^{-1}(u_1))| \\
&\qquad\quad + |F_{1x}(G_1^{-1}(u_1)) - F_{1x}(G_1'^{-1}(u_1'))| + |F_{1x}(G_1'^{-1}(u_1')) - F_{1i}(G_1'^{-1}(u_1'))| \\
&\qquad\quad + |F_{2i}(G_2^{-1}(u_2)) - F_{2x}(G_2^{-1}(u_2))| + |F_{2x}(G_2^{-1}(u_2)) - F_{2x}(G_2'^{-1}(u_2'))| \\
&\qquad\quad + |F_{2x}(G_2'^{-1}(u_2')) - F_{2i}(G_2'^{-1}(u_2'))|\Big] \\
&\qquad = O(1)\big(o(1) + \delta_n^2\big) \to 0, \quad \text{as } n \to \infty.
\end{aligned}$$

To show (A15) we use Lemma 2.11.6 of van der Vaart & Wellner (1996). We rewrite

$$\big(Z_{ni}(f) - Z_{ni}(f')\big)^2 = \int \big[\mathbb{I}\{w_1 \leq G_1^{-1}(u_1), w_2 \leq G_2^{-1}(u_2)\} \\ - \mathbb{I}\{w_1 \leq G_1^{\prime -1}(u_1'), w_2 \leq G_2^{\prime -1}(u_2')\}\big]^2 \, d\,\mu_{ni}(w_1, w_2),$$

where $\mu_{ni} = n\,h_n\,w_{ni}^2(x, h_n)\,\delta_{(Y_{1i},Y_{2i})}$ with δ being a Dirac measure. As the set of functions

$$\mathcal{G} = \big\{(w_1, w_2) \mapsto \mathbb{I}\{w_1 \leq a, w_2 \leq b\};\ (a, b) \in \mathbb{R}^2\big\}$$

is a VC-class with envelope $F = 1$ and $\mathcal{F} \subset \mathcal{G}$, the set of functions $\mathcal{F}$ is a VC-class as well with the same envelope. Thus it satisfies the uniform entropy condition (2.11.5) of van der Vaart & Wellner (1996). Moreover,

$$\sum_{i=1}^{n} \mu_{ni} F^2 = n\,h_n \sum_{i=1}^{n} w_{ni}^2(x, h_n) = O(1).$$

We have just verified all the assumptions of Lemma 2.11.6 of van der Vaart & Wellner (1996). Applying this lemma verifies condition (A15). □

Lemma 2 implies that for $\forall \varepsilon, \eta > 0\ \exists \delta > 0$ such that

$$\limsup_{n\to\infty} P\left(\sup_{\rho(f,f')<\delta} |\bar{Z}_n(f) - \bar{Z}_n(f')| > \varepsilon\right) < \eta. \tag{A16}$$

Further uniformly in (u_1, u_2)

$$\begin{aligned}
&\rho^2\left((u_1, u_2, F_{1xh}, F_{2xh}), (u_1, u_2, F_{1x}, F_{2x})\right) \qquad\qquad (A17)\\
&\quad = \big|F_{1x}(F_{1xh}^{-1}(u_1)) - u_1\big| + \big|F_{2x}(F_{2xh}^{-1}(u_2)) - u_2\big| \\
&\quad = \big|F_{1x}(F_{1xh}^{-1}(u_1)) - F_{1xh}(F_{1xh}^{-1}(u_1))\big| + \big|F_{2x}(F_{2xh}^{-1}(u_2)) - F_{2xh}(F_{2xh}^{-1}(u_2))\big| + O(\tfrac{1}{n\,h_n}) \\
&\quad \leq \sup_y |F_{1x}(y) - F_{1xh}(y)| + \sup_y |F_{2x}(y) - F_{2xh}(y)| + O(\tfrac{1}{n\,h_n}) \xrightarrow{P} 0.
\end{aligned}$$

Combining (A12), asymptotic ρ-equicontinuity (A16) and (A17) yields that

$$\sup_{u_1,u_2} \Big|A_n^h(u_1, u_2)\Big| = \sup_{u_1,u_2} \Big|D_n^h(u_1, u_2) - \mathrm{E}\,D_n^h(u_1, u_2)\Big| = o_P(1), \tag{A18}$$

which finishes the first step of the proof.

A2. Step 2 – Asymptotic representation of $C_n^{h_n}$. With the help of (**W3**), (**W4**) and **(R1)** we calculate

$$C_n^{h_n}(u_1,u_2) = \sqrt{n\,h_n}\sum_{i=1}^n w_{ni}(x,h_n)\,\left[H_{x_i}(F_{1xh}^{-1}(u_1),F_{2xh}^{-1}(u_2)) - H_{x_i}(F_{1x}^{-1}(u_1),F_{2x}^{-1}(u_2))\right]$$
$$:= \sqrt{n\,h_n}\,(C_{n1}(u_1,u_2)+C_{n2}(u_1,u_2)+C_{n3}(u_1,u_2))\,, \tag{A19}$$

with

$$C_{n1}(u_1,u_2) = \sum_{i=1}^n w_{ni}(x,h_n)\,\left[H_x(F_{1xh}^{-1}(u_1),F_{2xh}^{-1}(u_2)) - H_x(F_{1x}^{-1}(u_1),F_{2x}^{-1}(u_2))\right],$$
$$C_{n2}(u_1,u_2) = \sum_{i=1}^n w_{ni}(x,h_n)\,(x_i-x)\left[\dot H_x(F_{1xh}^{-1}(u_1),F_{2xh}^{-1}(u_2)) - \dot H_x(F_{1x}^{-1}(u_1),F_{2x}^{-1}(u_2))\right],$$
$$C_{n3}(u_1,u_2) = \sum_{i=1}^n w_{ni}(x,h_n)\,(x_i-x)^2\left[\ddot H_{z_{ih}}(F_{1xh}^{-1}(u_1),F_{2xh}^{-1}(u_2)) - \ddot H_{z_i}(F_{1x}^{-1}(u_1),F_{2x}^{-1}(u_2))\right],$$

where z_{ih}, z_i lie between x_i and x. By assumption **(R1)** and by (A6) we get

$$\max_{i\in I_x^{(n)}}\sup_{u_1,u_2}\left|\ddot H_{z_{ih}}(F_{1xh}^{-1}(u_1),F_{2xh}^{-1}(u_2)) - \ddot H_{z_i}(F_{1x}^{-1}(u_1),F_{2x}^{-1}(u_2))\right| = o_P(1),$$

which together with (**W4**) yields the asymptotic negligibility of the process $\sqrt{n\,h_n}\,C_{n3}$. Similarly we can argue that the process $\sqrt{n\,h_n}\,C_{n2}$ is asymptotically negligible.

Thus using (A19) together with (**W2**) we deduce

$$C_n^{h_n}(u_1,u_2) = \sqrt{n\,h_n}\,\left[C_x\left(F_{1x}(F_{1xh}^{-1}(u_1)),F_{2x}(F_{2xh}^{-1}(u_2))\right) - C_x(u_1,u_2)\right] + o_P(1). \tag{A20}$$

Substituting 1 for u_1 (or u_2) in the decomposition (A7) together with asymptotic negligibility of the process $A_n^{h_n}$ and (A20) yields (uniformly in u for $j=1,2$)

$$\sqrt{n\,h_n}\left[F_{jx}(F_{jxh}^{-1}(u)) - u\right] = -\sqrt{n\,h_n}\left[F_{jxh}(F_{jx}^{-1}(u)) - u\right] + o_P(1). \tag{A21}$$

The process $\mathbb{Y}_{jn}(u) = \sqrt{n\,h_n}\left[F_{jxh}(F_{jx}^{-1}(u)) - u\right]$ 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 modification of Lemma 4 of Omelka *et al.* (2009), (A21) and (A20) implies

$$C_n^{h_n}(u_1,u_2) \;=\; -C_x^{(1)}(u_1,u_2)\mathbb{Y}_{1n}(u_1) - C_x^{(2)}(u_1,u_2)\mathbb{Y}_{2n}(u_2) + o_P(1). \tag{A22}$$

Now, combining (A7), (A18), (A22) gives us the representation (9).

Appendix B – Proof of Theorem 2

Put

$$G_{xh}(u_1,u_2)=\sum_{i=1}^{n} w_{ni}(x,h_n)\,\mathbb{I}\{U_{1i}\le u_1, U_{2i}\le u_2\}.$$

In the following we show that uniformly in (u_1,u_2)

$$\tilde{C}_{xh}(u_1,u_2)=G_{xh}(G_{1xh}^{-1}(u_1),G_{2xh}^{-1}(u_2))+o_P\big(\tfrac{1}{\sqrt{n\,h_n}}\big), \tag{B1}$$

where G_{1xh} and G_{2xh} are the corresponding marginals of G_{xh}. Theorem 2 then follows from Theorem 1 and (B1).

By a similar argument as was used in the proof of Theorem 1 it is sufficient to consider only the fixed design.

B1. Auxiliary results

First we investigate how close a pseudo-observation $\tilde{U}_{ji}$ is to an unobserved U_{ji}. For $j=1,2$ put

$$\mathcal{E}_{jg}^{(n)}(z,u)=F_{jzg_j}(F_{jz}^{-1}(u))-u, \qquad u\in[0,1], \tag{B2}$$

and note that $\mathcal{E}_{jg}^{(n)}(x_i,U_{ji})=\tilde{U}_{ji}-U_{ji}$. It is useful to explore $\mathcal{E}_{jg}^{(n)}(z,u)$ when z is 'close' to x. By assumption (**W11**) (for all sufficiently large n) we can assume that for $z\in J_x^{(n)}$ the weight $w_{ni}(z,h_n)$ is positive, only if $x_i\in[z-C\,h_n, z+C\,h_n]$, where C is a finite constant.

For $j=1,2$ denote

$$Z_n^{(j)}(t,u)=F_{jz_tg_j}(F_{jz_t}^{-1}(u)), \qquad \text{where } z_t=x+t\,C\,h_n,\ (t,u)\in[-1,1]\times[0,1], \tag{B3}$$

and note that for $z\in[x-C\,h_n,x+C\,h_n]$: $\mathcal{E}_{jg}^{(n)}(z,u)=Z_n^{(j)}(\frac{z-x}{C\,h_n},u)-u$.

The following lemma uses the concepts of asymptotic tightness and asymptotic equicontinuity in probability (see p.21 and 37 of van der Vaart & Wellner (1996)) in the space of bounded functions $\ell^\infty([-1,1]\times[0,1])$ that is equipped with the supremum metric ρ_∞.

Lemma 3. *Under assumptions* (16), (**W1**), (**W9**), (**W10**) *and* ($\tilde{\mathbf{R}}\mathbf{3}$) *the processes* $\sqrt{n\,g_{jn}}\,(Z_n^{(j)}-\mathrm{E}\,Z_n^{(j)})$ *are asymptotically tight for* $j=1,2$.

Proof. It is straightforward to verify that for each $(t,u) \in [-1,1] \times [0,1]$ the random variable $\sqrt{n\,g_{jn}}\,(Z_n^{(j)}(t,u) - \mathrm{E}\,Z_n^{(j)}(t,u))$ is asymptotically tight. Thus by Theorem 1.5.7 of van der Vaart & Wellner (1996) it is sufficient to show that the process $\sqrt{n\,g_{jn}}\,(Z_n^{(j)} - \mathrm{E}\,Z_n^{(j)})$ is asymptotically uniformlyρ-equicontinuous in probability, where ρ is a metric such that the metric space $([-1,1] \times [0,1], \rho)$ is totally bounded. For our purposes it is convenient to take the following semimetric

$$\rho^2((t_1,u_1),(t_2,u_2)) = |t_1 - t_2|^2 + |u_1 - u_2|.$$

To prove the asymptotic equicontinuity we again use Theorem 2.11.1 of van der Vaart & Wellner (1996). Note that $Z_n^{(j)} = \sum_{i=1}^n Z_{ni}^{(j)}$, where

$$Z_{ni}^{(j)}(t,u) = w_{ni}(z_t, g_{jn})\,\mathbb{I}\{Y_{ji} \le F_{jz_t}^{-1}(u)\}.$$

As the first and the third assumption of the theorem may be verified similarly as in the proof of Lemma 2, it is sufficient to make a closer inspection of the second assumption. Let $\delta_n \searrow 0$, $f_1 = (t_1,u_1)$, $f_2 = (t_2,u_2)$. With the help of (16), (**W9**), (**W10**) and $(\tilde{\mathbf{R}}\mathbf{3})$ we estimate

$$\begin{aligned}
&\sup_{\rho(f_1,f_2)<\delta_n} n\,g_{jn} \sum_{i=1}^n \mathrm{E}\,\Big(Z_{ni}^{(j)}(f_1) - Z_{ni}^{(j)}(f_2)\Big)^2 \\
&\quad = \sup_{\rho(f_1,f_2)<\delta_n} n\,g_{jn} \sum_{i=1}^n \mathrm{E}\,\Big(w_{ni}(z_{t_1}, g_{jn})\,\mathbb{I}\{Y_{ji} \le F_{jz_{t_1}}^{-1}(u_1)\} \\
&\qquad\qquad\qquad - w_{ni}(z_{t_2}, g_{jn})\,\mathbb{I}\{Y_{ji} \le F_{jz_{t_2}}^{-1}(u_2)\}\Big)^2 \\
&\quad \le \sup_{\rho(f_1,f_2)<\delta_n} 2n\,g_{jn} \Bigg(\sup_{z\in[x-C\,h_n, x+C\,h_n]} \sum_{i=1}^n [w'_{ni}(z,g_{jn})]^2\,C^2\,|t_1-t_2|^2\,h_n^2 \\
&\qquad\qquad + \sup_{z\in[x-C\,h_n, x+C\,h_n]} \sum_{i=1}^n [w_{ni}(z,g_{jn})]^2\,(|u_1-u_2| + o(1))\Bigg) \\
&\quad = O(1) \sup_{\rho(f_1,f_2)<\delta_n} \big(|t_1-t_2|^2 + |u_1-u_2| + o(1)\big) = o(1),
\end{aligned}$$

which finishes the proof of the lemma. □

Remark 3. The asymptotic tightness stated in Lemma 3 implies that for each $\eta \in (0,1)$ there exists a compact subset K_η of $\ell^\infty([-1,1]\times[0,1])$ such that for every $\varepsilon > 0$

$$\liminf_{n\to\infty} P\left[\sqrt{n\,g_{jn}}\left(Z_n^{(j)} - \mathrm{E}\,Z_n^{(j)}\right) \in K_\eta^\varepsilon\right] \geq 1-\eta,$$

where $K_\eta^\varepsilon = \{(t,u) \in [-1,1]\times[0,1] : \rho((t,u),K_\eta) < \varepsilon\}$ is an 'ε-enlargement' around K_η. Thus there exists a decreasing sequence $\varepsilon_n \searrow 0$ such that

$$\liminf_{n\to\infty} P\left[\sqrt{n\,g_{jn}}\left(Z_n^{(j)} - \mathrm{E}\,Z_n^{(j)}\right) \in K_\eta^{\varepsilon_n}\right] \geq 1-\eta.$$

Moreover, as the set K_η is compact, for each $j \in \mathbb{N}$ the set $K_\eta^{\varepsilon_j}$ can be covered with finitely many balls of radius $2\,\varepsilon_j$ with centres in K_η, which are continuous functions. This further implies that $K_\eta^{\varepsilon_j}$ is uniformly bounded and that there exists $\delta > 0$ such that

$$\rho((t_1,u_1),(t_2,u_2)) < \delta \Rightarrow \sup_{f\in K_\eta^{\varepsilon_j}} |f(t_1,u_1) - f(t_2,u_2)| < 5\,\varepsilon_j.$$

Lemma 4. *Under assumption* (16) *and* (**W7**) *we have for each* $\eta \in (0,1)$, *uniformly in* u_1, u_2

$$\sup_{f_1,f_2\in K_\eta^{\varepsilon_n}} \left|\sqrt{n\,h_n}\sum_{i=1}^n w_{ni}(x,h_n)\left[\mathrm{E}\,\mathbb{I}\left\{U_{1i} \leq u_1 + \frac{f_1(\frac{x_i-x}{C\,h_n},U_{1i})}{\sqrt{n\,g_{1n}}},\ U_{2i} \leq u_2 + \frac{f_2(\frac{x_i-x}{C\,h_n},U_{2i})}{\sqrt{n\,g_{2n}}}\right\}\right.\right.$$
$$\left.\left. - C_{x_i}\left(u_1 + \frac{f_1(\frac{x_i-x}{C\,h_n},u_1)}{\sqrt{n\,g_{1n}}},\ u_2 + \frac{f_2(\frac{x_i-x}{C\,h_n},u_2)}{\sqrt{n\,g_{2n}}}\right)\right]\right| = o(1) \tag{B4}$$

Proof. Let $\varepsilon > 0$ be given. Thus we can find $n_\circ$ such that

$$\sup_{f\in K_\eta^{\varepsilon_{n_\circ}}}\ \sup_{|u-v|<\delta}\ \sup_{z\in[-1,1]} |f(z,u) - f(z,v)| < \varepsilon.$$

Further as the sequence of sets $K_\eta^{\varepsilon_n}$ is decreasing and the set $K_\eta^{\varepsilon_1}$ is uniformly bounded, for all sufficiently large n it holds

$$\sup_{f\in K_\eta^{\varepsilon_{n_\circ}}}\ \sup_{z,u} \frac{|f(z,u)|}{\sqrt{n\,\min(g_{1n},g_{2n})}} < \delta.$$

Thus for all sufficiently large n and with the help of (16) and (**W7**) we can bound the left-hand side of (B4) uniformly in u_1, u_2 by

$$\sup_{f_1,f_2\in K_\eta^{\varepsilon_{n_0}}} \Bigg| \sqrt{n\,h_n} \sum_{i=1}^{n} w_{ni}(x,h_n)$$

$$\iint_{|u_1-v_1|\leq\delta, |u_2-v_2|\leq\delta} \Big[\mathbb{I}\Big\{ v_1 \leq u_1 + \tfrac{f_1((x_i-x)/(C\,h_n),\,v_1)}{\sqrt{n\,g_{1n}}},\ v_2 \leq u_2 + \tfrac{f_2((x_i-x)/(C\,h_n),v_2)}{\sqrt{n\,g_{2n}}} \Big\}$$

$$- \mathbb{I}\Big\{ v_1 \leq u_1 + \tfrac{f_1((x_i-x)/(C\,h_n),u_1)}{\sqrt{n\,g_{1n}}},\ v_2 \leq u_2 + \tfrac{f_2((x_i-x)/(C\,h_n),u_2)}{\sqrt{n\,g_{2n}}} \Big\} \Big]\, dC_{x_i}(v_1,v_2) \Bigg|$$

$$\leq \sqrt{n\,h_n} \sum_{i=1}^{n} |w_{ni}(x,h_n)| \left(\tfrac{5\,\varepsilon}{\sqrt{n\,g_{1n}}} + \tfrac{5\,\varepsilon}{\sqrt{n\,g_{2n}}} \right) = \varepsilon\, O(1),$$

which proves the statement of the lemma as the term $O(1)$ does not depend on the particular choice of u_1, u_2 and ε can be taken arbitrarily small. □

B2. Main part of the proof of Theorem 2

Notation. In the rest of the Appendix it will be convenient to generalize the notation already introduced in the proof of Theorem 1. We will often use the expectations of the form $\mathrm{E}\, f(U_{i1}, U_{i2}; F_{1x_ig_1}, F_{2x_ig_2}; G_{1xh}, G_{2xh}; \tilde{G}_{1xh}, \tilde{G}_{2xh})$ which have to be understood in a way that the functions $F_{1x_ig_1}, F_{2x_ig_2}, G_{1xh}, G_{2xh}, \tilde{G}_{1xh}, \tilde{G}_{2xh}$ are considered fixed (nonrandom) and the expectation is computed only with respect to U_{i1} and U_{i2}. Formally

$$\mathrm{E}\, f(U_{i1}, U_{i2}; F_{1x_ig_1}, F_{2x_ig_2}; G_{1xh}, G_{2xh}; \tilde{G}_{1xh}, \tilde{G}_{2xh})$$
$$= \iint f(u_1, u_2; F_{1x_ig_1}, F_{2x_ig_2}; G_{1xh}, G_{2xh}; \tilde{G}_{1xh}, \tilde{G}_{2xh})\, dC_{x_i}(u_1,u_2),$$

whenever the integral on the right-hand side exists.

Decomposition. As the quantity η in Remark 3 can be taken arbitrarily small, in the following we will work conditionally on the event

$$\left[\sqrt{n\,g_{1n}} \left(Z_n^{(1)} - \mathrm{E}\, Z_n^{(1)} \right) \in K_\eta^{\varepsilon_n},\ \sqrt{n\,g_{2n}} \left(Z_n^{(2)} - \mathrm{E}\, Z_n^{(2)} \right) \in K_\eta^{\varepsilon_n} \right], \tag{B5}$$

where the processes $Z_n^{(1)}$ and $Z_n^{(2)}$ were introduced in (B3).

With (B5) in mind we make the following decomposition

$$\sqrt{n\,h_n}\left(\tilde{C}_{xh}(u_1,u_2) - G_{xh}(G_{1xh}^{-1}(u_1), G_{2xh}^{-1}(u_2))\right) = \bar{A}_n(u_1,u_2) + \bar{B}_n(u_1,u_2) + E_n(u_1,u_2), \quad \text{(B6)}$$

where

$$\begin{aligned} A_n(u_1,u_2) &= \sqrt{n\,h_n}\sum_{i=1}^{n} w_{ni}(x;h_n)\,\Big[\mathbb{I}\{\tilde{U}_{1i} \le \tilde{G}_{1xh}^{-1}(u_1), \tilde{U}_{2i} \le \tilde{G}_{2xh}^{-1}(u_2)\} \quad \text{(B7)} \\ &\qquad -\mathbb{I}\{U_{1i} \le \tilde{G}_{1xh}^{-1}(u_1), U_{2i} \le \tilde{G}_{2xh}^{-1}(u_2)\}\Big], \\ \bar{A}_n(u_1,u_2) &= A_n(u_1,u_2) - \mathrm{E}\,A_n(u_1,u_2) \\ B_n(u_1,u_2) &= \sqrt{n\,h_n}\sum_{i=1}^{n} w_{ni}(x;h_n)\,\Big[\mathbb{I}\{U_{1i} \le \tilde{G}_{1xh}^{-1}(u_1), U_{2i} \le \tilde{G}_{2xh}^{-1}(u_2)\} \quad \text{(B8)} \\ &\qquad -\mathbb{I}\{U_{1i} \le G_{1xh}^{-1}(u_1), U_{2i} \le G_{2xh}^{-1}(u_2)\}\Big], \\ \bar{B}_n(u_1,u_2) &= B_n(u_1,u_2) - \mathrm{E}\,B_n(u_1,u_2), \\ E_n(u_1,u_2) &= \mathrm{E}\,A_n(u_1,u_2) + \mathrm{E}\,B_n(u_1,u_2). \quad \text{(B9)} \end{aligned}$$

The proof will be divided into three steps in which we subsequently show that $\bar{A}_n$, $\bar{B}_n$ and E_n are asymptotically negligible uniformly in (u_1,u_2).

Step 1: Treatment of $\bar{A}_n$. First note that

$$\sup_{u_1,u_2} |\bar{A}_n(u_1,u_2)| \le \sup_{u_1,u_2} \left|\bar{D}_n(u_1,u_2)\right|,$$

where $D_n = \sqrt{n\,h_n}(\tilde{G}_{xh} - G_{xh})$ and $\bar{D}_n = D_n - \mathrm{E}\,D_n$.

Recall the definition of $\mathcal{E}_{jg}^{(n)}(x_i,U_{ji})$ given in (B2) and note that we can rewrite the process $\tilde{G}_{xh}$ as

$$\tilde{G}_{xh}(u_1,u_2) = \sum_{i=1}^{n} w_{ni}(x,h_n)\,\mathbb{I}\left\{U_{1i} \le u_1 - \mathcal{E}_{1g}^{(n)}(x_i,U_{1i}),\, U_{2i} \le u_2 - \mathcal{E}_{2g}^{(n)}(x_i,U_{2i})\right\}.$$

By Taylor expansion of $\mathrm{E}\,\mathcal{E}_{jg}^{(n)}(z,u)$, and assumptions (**W8**), (**W12**), (**W13**), **($\tilde{\mathbf{R}}$3)** for any $C<\infty$, we get :

$$\sup_{u\in[0,1]}\ \sup_{|z-x|\le C\,h_n} \left|\mathrm{E}\,\mathcal{E}_{jg}^{(n)}(z,u) - g_{jn}^2\,R^{(j)}(u)\right| = o(g_{jn}^2), \quad \text{(B10)}$$

where

$$R^{(j)}(u) = D_K\, \dot{F}_{jx}(F_{jx}^{-1}(u)) + \tfrac{E_K}{2}\, \ddot{F}_{jx}(F_{jx}^{-1}(u)), \qquad j = 1, 2. \tag{B11}$$

Let us define ${\mathcal{E}'}_{jg}^{(n)}(z,u) = \mathcal{E}_{jg}^{(n)}(z,u) - \mathrm{E}\,\mathcal{E}_{jg}^{(n)}(z,u) + g_{jn}^2\, R^{(j)}(u)$ and

$$\tilde{G}'_{xh}(u_1,u_2) = \sum_{i=1}^{n} w_{ni}(x,h_n)\, \mathbb{I}\left\{U_{1i} \le u_1 - {\mathcal{E}'}_{1g}^{(n)}(x_i,U_{1i}),\, U_{2i} \le u_2 - {\mathcal{E}'}_{2g}^{(n)}(x_i,U_{2i})\right\}.$$

The proof of Step 1 will be divided into two parts. First,we show that $\bar{D}'_n = D'_n - \mathrm{E}\, D'_n = o_P(1)$, where $D'_n = \sqrt{n\, h_n}(\tilde{G}'_{xh} - G_{xh})$. Then we prove that $\bar{D}_n - \bar{D}'_n = o_P(1)$.

Part 1. Process $\bar{D}'_n$. Now we are ready to use the machinery introduced in Ghoudi & Rémillard (1998). Mimicking the proof of Lemma 5.2 of that paper we can show that for given $f_1, f_2 \in C([-1,1]\times[0,1])$

$$\sup_{u_1,u_2} |\gamma_n(u_1,u_2,f_1,f_2) - \mathrm{E}\,\gamma_n(u_1,u_2,f_1,f_2)| = o_P(1),$$

where

$$\gamma_n(u_1,u_2,f_1,f_2) = \sqrt{n\,h_n} \sum_{i=1}^{n} w_{ni}(x,h_n)$$
$$\left[\mathbb{I}\left\{U_{1i} \le u_1 + \frac{f_1(\frac{x_i-x}{C\,h_n},U_{1i})}{\sqrt{n\,g_{1n}}},\, U_{2i} \le u_2 + \frac{f_2(\frac{x_i-x}{C\,h_n},U_{2i})}{\sqrt{n\,g_{2n}}}\right\} - \mathbb{I}\left\{U_{1i} \le u_1, U_{2i} \le u_2\right\}\right].$$

As $\mathcal{E}_{jg}^{(n)}(z_t,u) = Z_n^{(j)}(t,u) - u$, where $z_t = x + t\,C\,h_n$ and $t \in [-1,1]$, and by the properties of the sequence of sets $K_\eta^{\varepsilon_n}$ of Remark 3, we know that with probability arbitrarily close to one and for arbitrarily small ε for all sufficiently large n the processes $\{\sqrt{n\,g_{jn}}(\mathcal{E}_{jg}^{(n)}(z_t,u) - \mathrm{E}\,\mathcal{E}_{jg}^{(n)}(z_t,u)), (t,u) \in [-1,1]\times[0,1]\}$ take values in a set that can be covered with finitely many balls of radius ε and whose centres are continuous functions on $[-1,1]\times[0,1]$. Further, as the functions $R^{(1)}$, $R^{(2)}$ given by (B11) are continuous on $[0,1]$, the same hold true for the processes $\{\sqrt{n\,g_{jn}}\,{\mathcal{E}'}_{jg}^{(n)}(z_t,u), (t,u) \in [-1,1]\times[0,1]\}$. That enable us to mimic the proof of Lemma 4.1 of Ghoudi & Rémillard (1998) and show that

$$\sup_{u_1,u_2} \left|\bar{D}'_n(u_1,u_2)\right| = \sup_{u_1,u_2} \left|D'_n(u_1,u_2) - \mathrm{E}\, D'_n(u_1,u_2)\right| \tag{B12}$$
$$= \sup_{u_1,u_2} \left|\gamma_n(u_1,u_2,\mathcal{E}'^{(n)}_{1g},\mathcal{E}'^{(n)}_{2g}) - \mathrm{E}\,\gamma_n(u_1,u_2,\mathcal{E}'^{(n)}_{1g},\mathcal{E}'^{(n)}_{2g})\right| = o_P(1)$$

Part 2. Process $\bar{D}_n - \bar{D}'_n$. With the help of (B10) for each $\varepsilon > 0$ for all sufficiently large n we can bound (uniformly in u_1, u_2)

$$\begin{aligned} D_n(u_1,u_2) - {} & D'_n(u_1,u_2) \\ &= \sqrt{n\,h_n}\,\left[\tilde{G}_{xh}(u_1,u_2) - \tilde{G}'_{xh}(u_1,u_2)\right] \qquad \text{(B13)} \\ &\le \sqrt{n\,h_n}\,\left[\tilde{G}'_{1xh}(u_1+\varepsilon\,g_{1n}^2) - \tilde{G}'_{1xh}(u_1)\right] + \sqrt{n\,h_n}\,\left[\tilde{G}'_{2xh}(u_2+\varepsilon\,g_{2n}^2) - \tilde{G}'_{2xh}(u_2)\right] \end{aligned}$$

Now with the help of Lemma 4 we get

$$\sup_u \mathrm{E}\,\sqrt{n\,h_n}\,\left|\tilde{G}'_{1xh}(u+\varepsilon\,g_{1n}^2) - \tilde{G}'_{1xh}(u)\right| = \varepsilon\,\sqrt{n\,h_n}\;g_{jn}^2 = \varepsilon\, O(1). \tag{B14}$$

Using (B12), (B14) and the asymptotic ρ-equicontinuity of the processes $\sqrt{n\,h_n}(G_{jxh} - \mathrm{E}\, G_{jxh})$ with respect to the metric $\rho^2(u,v) = |u-v|$ yields that for $j = 1,2$ (uniformly in u_j)

$$\begin{aligned} \sqrt{n\,h_n}\, & \left[\tilde{G}'_{jxh}(u_j+\varepsilon\,g_{1n}^2) - \tilde{G}'_{jxh}(u_j)\right] \\ = {} & \sqrt{n\,h_n}\,\left[G_{jxh}(u_j+\varepsilon\,g_{1n}^2) - G_{jxh}(u_j)\right] - \sqrt{n\,h_n}\;\mathrm{E}\,\left[G_{jxh}(u_j+\varepsilon\,g_{jn}^2) - G_{jxh}(u_j)\right] \\ & + \sqrt{n\,h_n}\;\mathrm{E}\,\left[\tilde{G}'_{jxh}(u_j+\varepsilon\,g_{jn}^2) - \tilde{G}'_{jxh}(u_j)\right] + o_P(1) \qquad \text{(B15)} \\ = {} & o_P(1) + \varepsilon\, O(1). \end{aligned}$$

As ε can be chosen arbitrarily small, combining (B13) and (B15) yields that

$$\sup_{u_1,u_2} \left(D_n(u_1,u_2) - D'_n(u_1,u_2)\right) \le o_P(1).$$

Analogously we can prove that

$$\inf_{u_1,u_2} \left(D_n(u_1,u_2) - D'_n(u_1,u_2)\right) \ge o_P(1),$$

as well as

$$\sup_{u_1,u_2} |\mathrm{E}\, D_n(u_1,u_2) - \mathrm{E}\, D_n'(u_1,u_2)| = o(1), \tag{B16}$$

which finishes Step 1 of the proof of the theorem.

Step 2: Treatment of $\bar{B}_n$. This part of the proof follows from Step 1 of the proof of Theorem 1 (for C_{xh}) provided we prove that for $j = 1, 2$

$$\sup_{u} \left| \tilde{G}_{jxh}^{-1}(u) - G_{jxh}^{-1}(u) \right| = o_P(1). \tag{B17}$$

From the previous step of the proof we know that $\sup_{u_1,u_2} |\bar{D}_n(u_1,u_2)| = o_P(1)$. Putting $u_1 = 1$ (or $u_2 = 1$) then Lemma 4 implies that for $j = 1, 2$ uniformly in u

$$\sqrt{n\,h_n}\,\Big(\tilde{G}_{jxh}(u) - G_{jxh}(u)\Big) = -\sqrt{n\,h_n} \sum_{i=1}^{n} w_{ni}(x,h_n)\,\mathcal{E}_{jg_j}^{(n)}(x_i,u) + o_P(1). \tag{B18}$$

Thus Lemma 3 together with (16), (**W2**), (B10) and (B18) yield that uniformly in u

$$\begin{aligned}
&\sqrt{n\,h_n}\Big(\tilde{G}_{jxh}(u) - G_{jxh}(u)\Big) \\
&\quad = -\sqrt{n\,h_n} \sum_{i=1}^{n} w_{ni}(x,h_n)\, \Big[\mathcal{E}_{jg_j}^{(n)}(x_i,u) - \mathrm{E}\,\mathcal{E}_{jg_j}^{(n)}(x_i,u) + \mathrm{E}\,\mathcal{E}_{jg_j}^{(n)}(x_i,u)\Big] + o_P(1). \\
&\quad = \sqrt{n\,h_n}\, O_P(\tfrac{1}{\sqrt{n\,g_{jn}}}) + \sqrt{n\,h_n}\, O(g_{jn}^2) = O_P(1).
\end{aligned} \tag{B19}$$

The convergence in (B17) now follows from (B19), (**W1**) and the following bound

$$\begin{aligned}
&\sup_u \left|\tilde{G}_{jxh}^{-1}(u) - G_{jxh}^{-1}(u)\right| \le \sup_u |\tilde{G}_{jxh}^{-1}(u) - u| + \sup_u \left|G_{jxh}^{-1}(u) - u\right| \\
&\quad \le \sup_u |\tilde{G}_{jxh}(u) - u| + \sup_u |G_{jxh}(u) - u| + o_P\left(\tfrac{1}{\sqrt{n\,h_n}}\right) \\
&\quad \le \sup_u |\tilde{G}_{jxh}(u) - G_{jxh}(u)| + 2\,\sup_u |G_{jxh}(u) - u| + o_P\left(\tfrac{1}{\sqrt{n\,h_n}}\right) = O_P\left(\tfrac{1}{\sqrt{n\,h_n}}\right).
\end{aligned} \tag{B20}$$

Step 3 : Treatment of E_n. For simplicity of notation for $i = 1, \ldots, n$ and $j = 1, 2$ put $e_j^{(n)}(x_i,u) := \mathcal{E}_{jg}^{(n)}(x_i, \tilde{G}_{jxh}^{-1}(u))$. Lemma 4 yields that

$$\begin{aligned}
E_n(u_1,u_2) &= \mathrm{E}\,A_n(u_1,u_2) + \mathrm{E}\,B_n(u_1,u_2) \\
&= \sqrt{n\,h_n} \sum_{i=1}^{n} w_{ni}(x,h_n)\, \Big[C_{x_i}(\tilde{G}_{1xh}^{-1}(u_1) - e_1^{(n)}(x_i,u_1), \tilde{G}_{2xh}^{-1}(u_2) - e_2^{(n)}(x_i,u_2)) \\
&\qquad\qquad - C_{x_i}(G_{1xh}^{-1}(u_1), G_{2xh}^{-1}(u_2))\Big].
\end{aligned}$$

In the following it is useful to note that thanks to (**W1**)

$$\sqrt{n\,h_n}\,\Big(\tilde{C}_{xh}(u_1,1) - G_{xh}(G_{1xh}^{-1}(u_1), G_{2xh}^{-1}(1))\Big) = \sqrt{n\,h_n}\,\Big(\tilde{G}_{1xh}(\tilde{G}_{1xh}^{-1}(u_1)) - G_{1xh}(G_{1xh}^{-1}(u_1))\Big) = o_p(1).$$

Thus, putting $u_1 = 1$ (or $u_2 = 1$) in (B6) together with assumption (**W2**) and the asymptotic negligibility of processes $\bar{A}_n$, and $\bar{B}_n$ yield that (for $j = 1, 2$)

$$\sup_{u\in[0,1]} \left|\sqrt{n\,h_n}\sum_{i=1}^{n} w_{ni}(x,h_n)\,\Big[\tilde{G}_{jxh}^{-1}(u) - e_j^{(n)}(x_i,u) - G_{jxh}^{-1}(u)\Big]\right| = o_p(1). \tag{B21}$$

Now, let $\varepsilon > 0$ be given. As we can bound E_n by

$$|E_n(u_1,u_2)| \le \sqrt{n\,h_n}\sum_{i=1}^{n}|w_{ni}(x,h_n)|\,\Big[\Big|\tilde{G}_{1xh}^{-1}(u_1) - e_1^{(n)}(x_i,u_1) - G_{1xh}^{-1}(u_1)\Big| + \Big|\tilde{G}_{2xh}^{-1}(u_2) - e_2^{(n)}(x_i,u_2) - G_{2xh}^{-1}(u_2)\Big|\Big],$$

with the help of (16), (**W7**), (B10), (B18), (B20) and Lemma 3 we can find $\delta_\varepsilon > 0$ such that for all sufficiently large n

$$P\left[\sup_{(u_1,u_2)\in A_{\delta_\varepsilon}} |E_n(u_1,u_2)| \ge \varepsilon\right] \le \varepsilon, \tag{B22}$$

where A_{δ_ε} is a union of open δ_ε-neighbourhoods of the points $\{(0,0),(0,1),(1,0),(1,1)\}$. Further making use of Taylor expansion, (B21) and **($\tilde{\mathbf{R}}$2)** we get that uniformly in $(u_1,u_2)\in[0,1]^2\setminus A_{\delta_\varepsilon}$

$$\begin{aligned} E_n(u_1,u_2) \;=\; & \sqrt{n\,h_n}\sum_{i=1}^{n} w_{ni}(x,h_n) \\ & \Big\{C_{x_i}^{(1)}(G_{1xh}^{-1}(u_1), G_{2xh}^{-1}(u_2))\,\Big[\tilde{G}_{1xh}^{-1}(u_1) - e_1^{(n)}(x_i,u_1) - G_{1xh}^{-1}(u_1)\Big] \\ & \quad + C_{x_i}^{(2)}(\tilde{G}_{1xh}^{-1}(u_2), \tilde{G}_{2xh}^{-1}(u_2))\,\Big[\tilde{G}_{2xh}^{-1}(u_2) - e_2^{(n)}(x_i,u_2) - G_{2xh}^{-1}(u_2)\Big] \\ & \quad + o_P\Big(\tfrac{1}{\sqrt{n\,h_n}}\Big)\Big\} \end{aligned}$$

$$
\begin{aligned}
&= \sqrt{n\,h_n}\sum_{i=1}^{n} w_{ni}(x,h_n) \\
&\qquad \Big\{ C_x^{(1)}(G_{1xh}^{-1}(u_1),G_{2xh}^{-1}(u_2)) \Big[\tilde{G}_{1xh}^{-1}(u_1) - e_1^{(n)}(x_i,u_1) - G_{1xh}^{-1}(u_1)\Big] \\
&\qquad\quad + C_x^{(2)}(G_{1xh}^{-1}(u_2),G_{2xh}^{-1}(u_2)) \Big[\tilde{G}_{2xh}^{-1}(u_2) - e_2^{(n)}(x_i,u_2) - G_{2xh}^{-1}(u_2)\Big] \\
&\qquad\quad + o_P\left(\tfrac{1}{\sqrt{n\,h_n}}\right)\Big\} + o_P(1) \\
&= o_P(1),
\end{aligned}
$$

which together with (B22) finishes Step 3 of the proof of the theorem.