Metrika (2019) 82:823–841
https://doi.org/10.1007/s00184-019-00711-y

On the large-sample behavior of two estimators of the conditional copula under serially dependent data

Taoufik Bouezmarni[1,2] · **Félix Camirand Lemyre**[1] · **Jean-François Quessy**[3]

Received: 15 December 2018 / Published online: 28 February 2019

Abstract
The conditional copula of a random pair (Y_1, Y_2) given the value taken by some covariate $X \in \mathbb{R}$ is the function $C_x : [0, 1]^2 \to [0, 1]$ such that $\mathbb{P}(Y_1 \leq y_1, Y_2 \leq y_2|X = x) = C_x\{\mathbb{P}(Y_1 \leq y_1|X = x), \mathbb{P}(Y_2 \leq y_2|X = x)\}$. In this note, the weak convergence of the two estimators of C_x proposed by Gijbels et al. (Comput Stat Data Anal 55(5):1919–1932, 2011) is established under α-mixing. It is shown that under appropriate conditions on the weight functions and on the mixing coefficients, the limiting processes are the same as those obtained by Veraverbeke et al. (Scand J Stat 38(4):766–780, 2011) under the i.i.d. setting. The performance of these estimators in small sample sizes is investigated with simulations.

Keywords α-mixing processes · Conditional copula · Local linear kernel estimation · Weak convergence

This research was supported in part by individual grants from the Natural Sciences and Engineering Research Council of Canada (NSERC), by the Fonds de recherche du Québec—Nature et technologies (FRQNT) and by Australian Research Council Discovery Project DP140100125.

✉ Félix Camirand Lemyre
felix.camirand.lemyre@usherbrooke.ca

Taoufik Bouezmarni
taoufik.bouezmarni@usherbrooke.ca

Jean-François Quessy
jean-francois.quessy@uqtr.ca

1 Département de mathématiques, Université de Sherbrooke, Sherbrooke, Canada

2 Centre interuniversitaire de recherche en économie quantitative (CIREQ), Montréal, Canada

3 Département de mathématiques et d'informatique, Université du Québec à Trois-Rivières, Trois-Rivières, Canada

1 Introduction

Copulas have become a popular tool for modeling the dependence between the components of a random vector. The starting point of copula theory is Sklar's Theorem. In its classical formulation, this result ensures that for any random pair (Y_1, Y_2), there exists a function $C : [0, 1]^2 \to [0, 1]$ called the copula of (Y_1, Y_2) such that for all $(y_1, y_2) \in \mathbb{R}^2$,

$$\mathbb{P}(Y_1 \leq y_1, Y_2 \leq y_2) = C\left\{\mathbb{P}(Y_1 \leq y_1), \mathbb{P}(Y_2 \leq y_2)\right\}.$$

When Y_1 and Y_2 are continuous, then C is unique. The d-variate extension of Sklar's Theorem is straightforward. See Nelsen (2006) for more details on copulas.

Recently, some works concentrated on capturing the influence of a covariate $X \in \mathbb{R}$ on the dependence structure of a random pair. A motivating example is given in Gijbels et al. (2011), where the relationship between the life expectancy of men (Y_1) and women (Y_2) with respect to the gross domestic product (X) is studied. Such an investigation relies on an extension of Sklar's Theorem to the case of conditional dependence as initiated by Patton (2006). Formally, letting $H_x(y_1, y_2) = \mathbb{P}(Y_1 \leq y_1, Y_2 \leq y_2 | X = x)$ be the joint conditional distribution, the dependence between Y_1 and Y_2 conditional on $X = x$ is characterized by the conditional copula $C_x : [0, 1]^2 \to [0, 1]$ such that for all $(y_1, y_2) \in \mathbb{R}^2$,

$$H_x(y_1, y_2) = C_x\left\{\mathbb{P}(Y_1 \leq y_1 | X = x), \mathbb{P}(Y_2 \leq y_2 | X = x)\right\}. \tag{1}$$

In that context, the estimation of the conditional dependence structure may be meaningful. To this end, Gijbels et al. (2011) proposed two nonparametric estimators of C_x; their asymptotic behavior was formally investigated by Veraverbeke et al. (2011) in the i.i.d. case.

The purpose of this note is to extend these large-sample results to the case of time series. To this end, a general framework is adopted where the process $\{(Y_{1t}, Y_{2t}, X_t)\}_{t\in\mathbb{Z}}$ is stationary and satisfies a strong mixing condition. Specifically, define $\mathcal{F}_a^b$ as the σ-field generated by $\{(Y_{1t}, Y_{2t}, X_t)\}_{a\leq t\leq b}$ and let

$$\alpha_r = \sup_{k\in\mathbb{Z}} \left\{ \sup_{(A,B)\in\mathcal{F}_{-\infty}^k \times \mathcal{F}_{k+r}^\infty} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\,\mathbb{P}(B)| \right\}$$

be the α-mixing coefficients. Then the process $\{(Y_{1t}, Y_{2t}, X_t)\}_{t\in\mathbb{Z}}$ is said to be α-mixing, or strongly mixing, if $\alpha_r \to 0$ as $r \to \infty$. Several parametric time series models are α-mixing, including ARMA and GARCH processes under appropriate parametric restrictions. More details can be found in Doukhan (1994), Meitz and Saikkonen (2008) and Carrasco and Chen (2002). The setup adopted in this paper is therefore very general and convenient in contexts that involve serially dependent observations, *e.g.* financial data.

The remaining of this note is organized as follows. Section 2 describes the two estimators introduced by Gijbels et al. (2011) and establishes their weak convergence under α-mixing; it is shown that these limits are the same as those derived by Veraverbeke et al. (2011) under the i.i.d. setting. Section 3 investigates the performance of the two estimators in small and moderate sample sizes under serially dependent data. The proofs of the main results are given in the “Appendix”.

2 Two estimators of C_x and their behavior under α-mixing

In the sequel, consider n realizations $(Y_{11}, Y_{21}, X_1), \ldots, (Y_{1n}, Y_{2n}, X_n)$ of a stationary and strongly mixing process $\{(Y_{1t}, Y_{2t}, X_t)\}_{t\in\mathbb{Z}}$. The goal is to estimate the copula C_x of the conditional distribution of (Y_{1t}, Y_{2t}) given $X_t = x$.

2.1 A first estimator

As long as the marginal conditional distributions $F_{1x}(y) = \mathbb{P}(Y_1 \le y|X = x)$ and $F_{2x}(y) = \mathbb{P}(Y_2 \le y|X = x)$ are continuous, an estimator of C_x will arise upon noting that Eq. (1) can be rewritten as

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

where, here and in the sequel, the inverse of a function is understood as its left-continuous generalized inverse. An estimator of H_x will then provide a plug-in estimator of C_x. One possibility is to estimate H_x with a local linear kernel smoothing estimator, namely

$$H_{xh}(y_1, y_2) = \frac{1}{nh}\sum_{i=1}^{n} \mathcal{K}_{xn}\left(\frac{X_i - x}{h}\right)\mathbb{I}\left(Y_{1i} \le y_1, Y_{2i} \le y_2\right), \tag{3}$$

where $h = h_n$ is a bandwidth parameter that typically depends on the sample size, and $\mathcal{K}_{xn}$ is the *local linear* weight function. Specifically, for a given symmetric positive kernel density K on $\mathbb{R}$,

$$\mathcal{K}_{xn}(z) = K(z)\left\{\frac{S_{n,2}(x) - z\, S_{n,1}(x)}{S_{n,0}(x)\, S_{n,2}(x) - S_{n,1}^2(x)}\right\},$$

where for $\ell \in \{0, 1, 2\}$,

$$S_{n,\ell}(x) = \frac{1}{nh}\sum_{i=1}^{n}\left(\frac{X_i - x}{h}\right)^{\ell} K\left(\frac{X_i - x}{h}\right).$$

See Fan and Gijbels (1996) for details on how to derive this estimator. From Eq. (2), a plug-in estimator of C_x is therefore given by

$$C_{xh}(u_1, u_2) = H_{xh}\left\{F_{1xh}^{-1}(u_1), F_{2xh}^{-1}(u_2)\right\}, \tag{4}$$

where F_{1xh} and F_{2xh} are the empirical marginal conditional distributions, i.e.

$$F_{1xh}(y) = \lim_{w\to\infty} H_{xh}(y, w) \quad \text{and} \quad F_{2xh}(y) = \lim_{w\to\infty} H_{xh}(w, y).$$

Now the goal is to describe the large-sample behavior of the empirical process $\mathbb{C}_{xh} = \sqrt{nh}(C_{xh} - C_x)$ viewed as a random element in the space $\ell^\infty([0, 1]^2)$ of bounded functions defined on $[0, 1]^2$. The first step toward this goal is to obtain the asymptotic behaviour of the weighted process $\mathbb{H}_{xh} = \sqrt{nh}(H_{xh} - H_x)$ under α-mixing sequences, where H_x and H_{xh} are defined respectively in Eq. (1) and Eq. (3). Note that for fixed $(y_1, y_2) \in \mathbb{R}^2$, the asymptotic normality of $\mathbb{H}_{xh}(y_1, y_2)$ under α-mixing has been derived by Masry and Fan (1997). Proposition 1 extends this result to the whole of $\mathbb{R}^2$ under the following assumptions. In the sequel, J_x is an open neighborhood of x.

($\mathcal{S}$) The process $\{(Y_{1t}, Y_{2t}, X_t)\}_{t\in\mathbb{Z}}$ is stationary and its α-mixing coefficients satisfy $\alpha_r = O(r^{-a})$ for some $a > 6$. Also, one can find $M > 0$ such that the joint density $f_X^{(\ell_1,\ldots,\ell_5)}$ of $(X_0, X_{\ell_1}, \ldots, X_{\ell_5})$ satisfies $f_X^{(\ell_1,\ldots,\ell_5)}(x_0, \ldots, x_{\ell_5}) \le M$ for all $x_0, \ldots, x_{\ell_5} \in J_x$.

($\mathcal{H}$) The following functions are uniformly continuous for $(w, y_1, y_2) \in J_x \times \mathbb{R}^2$:

$$\dot{H}_w(y_1, y_2) = \frac{\partial}{\partial w} H_w(y_1, y_2) \quad \text{and} \quad \ddot{H}_w(y_1, y_2) = \frac{\partial^2}{\partial w^2} H_w(y_1, y_2).$$

($\mathcal{LL}$) The kernel K is continuous, vanishes outside of a compact support and its second order derivative is bounded. Also, the density f_X of X exists, is twice differentiable and strictly positive at x.

($\mathcal{N}$) As $n \to \infty$, $nh^5 \to \kappa^2 < \infty$ and $n^{8-\delta}h^{21} \to \infty$ for some $\delta > 0$.

The result may now be stated. To this end, define for an arbitrary bivariate distribution function H and $(y_1, y_2), (y_1', y_2') \in \mathbb{R}^2$,

$$\sigma_H^2(y_1, y_2, y_1', y_2') = H\left\{\min(y_1, y_1'), \min(y_2, y_2')\right\} - H(y_1, y_2)\, H(y_1', y_2'). \quad (5)$$

Proposition 1 *Under Assumptions ($\mathcal{S}$), ($\mathcal{H}$), ($\mathcal{LL}$) and ($\mathcal{N}$), $\mathbb{H}_{xh} = \sqrt{nh}(H_{xh} - H_x)$ converges weakly in the space $\ell^\infty(\mathbb{R}^2)$ to a Gaussian limit $\mathbb{H}_x$ with mean*

$$\mathrm{E}\{\mathbb{H}_x(y_1, y_2)\} = \frac{\kappa}{2}\left(\int_{\mathbb{R}} z^2 K(z)\,\mathrm{d}z\right) \ddot{H}_x(y_1, y_2)$$

and covariance function defined for $(y_1, y_2), (y_1', y_2') \in \mathbb{R}^2$ by

$$\mathrm{Cov}\left\{\mathbb{H}_x(y_1, y_2), \mathbb{H}_x(y_1', y_2')\right\} = \frac{1}{f_X(x)}\left(\int_{\mathbb{R}} \{K(z)\}^2\,\mathrm{d}z\right) \sigma_{H_x}^2(y_1, y_2, y_1', y_2').$$

To establish the weak convergence of C_{xh}, an additional assumption on the partial derivatives of the conditional copula C_x is needed.

($\mathcal{C}_x$) The first order partial derivatives

$$C_x^{[1]}(u_1,u_2)=\frac{\partial}{\partial u_1}C_x(u_1,u_2) \quad \text{and} \quad C_x^{[2]}(u_1,u_2)=\frac{\partial}{\partial u_2}C_x(u_1,u_2)$$

are continuous respectively on $(0,1)\times[0,1]$ and $[0,1]\times(0,1)$.

Proposition 2 *Under Assumptions* ($\mathcal{S}$), ($\mathcal{H}$), ($\mathcal{LL}$), ($\mathcal{N}$) *and* ($\mathcal{C}_x$), $\mathbb{C}_{xh}=\sqrt{nh}(C_{xh}-C_x)$ *converges weakly in the space* $\ell^\infty([0,1]^2)$ *to the Gaussian limit*

$$\mathbb{C}_x(u_1,u_2)=\mathbb{B}_x(u_1,u_2)-C_x^{[1]}(u_1,u_2)\,\mathbb{B}_x(u_1,1)-C_x^{[2]}(u_1,u_2)\,\mathbb{B}_x(1,u_2),$$

where $\mathbb{B}_x$ *is a Gaussian process on* $[0,1]^2$ *with mean*

$$\mathrm{E}\left\{\mathbb{B}_x(u_1,u_2)\right\}=\frac{\kappa}{2}\left(\int_{\mathbb{R}}z^2K(z)\,\mathrm{d}z\right)\ddot{H}_x\left\{F_{1x}^{-1}(u_1),\,F_{2x}^{-1}(u_2)\right\}$$

and covariance function defined for $(u_1,u_2),(u_1',u_2')\in[0,1]^2$ *by*

$$\mathrm{Cov}\left\{\mathbb{B}_x(u_1,u_2),\mathbb{B}_x(u_1',u_2')\right\}=\frac{1}{f_X(x)}\left(\int_{\mathbb{R}}\{K(z)\}^2\,\mathrm{d}z\right)\sigma^2_{C_x}(u_1,u_2,u_1',u_2'),$$

and where for $j\in\{1,2\}$, $C_x^{[j]}$ *is defined as* 0 *on the set* $\{(u_1,u_2)\in[0,1]^2:u_j\in\{0,1\}\}$.

The asymptotic bias and covariance function of $\mathbb{C}_x$ can be computed from its representation in term of $\mathbb{B}_x$ as stated in Proposition 2. Indeed, it can be shown that these expressions match those discovered by Veraverbeke et al. (2011) under the i.i.d setting and in the special case of local linear weights. In other words, somewhat surprisingly, the time-dependency has no impact on the asymptotic behavior of the estimator C_{xh}. This feature can be explained by the fact that the kernel function smooths the covariate space in a shrinking neighborhood of x.

Note that Assumptions ($\mathcal{H}$) and ($\mathcal{LL}$) are the same as those in Veraverbeke et al. (2011) under serial independence. However, additional and/or modified assumptions were needed compared to them. Firstly, Assumption ($\mathcal{S}$) was needed in order to take into account the serial nature of the process; this condition is stronger than α-mixing. Secondly, while $nh\to\infty$ and $nh^5<\infty$ were assumed in Veraverbeke et al. (2011), the requirement that $n^{8-\delta}h^{21}\to\infty$ as $n\to\infty$ for some $\delta>0$ was also needed in Assumption ($\mathcal{N}$) to control the impact of time dependencies on the large sample behaviour of the local linear system of weights.

2.2 A second estimator

As noted by Gijbels et al. (2011), the estimator C_{xh} may be severely biased, especially when the marginal conditional distributions F_{1x} and F_{2x} strongly depend on the covariate. For that reason, these authors proposed a second estimator in order to reduce this effect of the covariate on the margins and hopefully obtain a smaller bias. To describe it, define for each $i \in \{1, \ldots, n\}$ the *pseudo-uniformized* observations $(\widetilde{U}_{1i}, \widetilde{U}_{2i}) = (F_{1X_i h_1}(Y_{1i}), F_{2X_i h_2}(Y_{2i}))$, where h_1 and h_2 are bandwidth parameters that may differ from h. Then, for $(v_1, v_2) \in [0, 1]^2$, define

$$G_{xh}(v_1, v_2) = \frac{1}{nh} \sum_{i=1}^{n} \mathcal{K}_{xn} \left(\frac{X_i - x}{h} \right) \mathbb{I} \left\{ F_{1X_i h_1}(Y_{1i}) \leq v_1, F_{2X_i h_2}(Y_{2i}) \leq v_2 \right\}.$$

Letting $G_{1xh}(y) = \lim_{w \to \infty} G_{xh}(y, w)$ and $G_{2xh}(y) = \lim_{w \to \infty} G_{xh}(w, y)$, an estimator of C_x is given by

$$\widetilde{C}_{xh}(u_1, u_2) = G_{xh} \left\{ G_{1xh}^{-1}(u_1), G_{2xh}^{-1}(u_2) \right\}. \tag{6}$$

The weak convergence of $\widetilde{\mathbb{C}}_{xh} = \sqrt{nh}(\widetilde{C}_{xh} - C_x)$ is established in Proposition 3. Before stating the result, consider the following modified versions of Assumption $(\mathcal{H})$ and Assumption $(\mathcal{N})$.

($\mathcal{H}^\star$) For $j \in \{1, 2\}$, the functions $F_{jw}\{F_{jw}^{-1}(u)\}$, $\dot{F}_{jw}\{F_{jw}^{-1}(u)\}$ and $\ddot{F}_{jw}\{F_{jw}^{-1}(u)\}$ are uniformly continuous for $(w, u) \in J_x \times [0, 1]$ and $u \mapsto \dot{F}_{jw}\{F_{jw}^{-1}(u)\}$ is Lipchitz continuous for $z \in J_x$. In addition, the following functions are uniformly continuous for $(w, u_1, u_2) \in J_x \times [0, 1]^2$:

$$\dot{C}_w(u_1, u_2) = \frac{\partial}{\partial w} C_w(u_1, u_2) \quad \text{and} \quad \ddot{C}_w(u_1, u_2) = \frac{\partial^2}{\partial w^2} C_w(u_1, u_2).$$

($\mathcal{N}^\star$) As $n \to \infty$, $nh^5 \to \kappa^2 < \infty$, $n^{8-\delta} h^{21} \to \infty$ for some $\delta > 0$, $\max(nh_1^5, nh_2^5) < \infty$ and $\max(h/h_1, h/h_2) < \infty$.

Proposition 3 *Under Assumptions* $(\mathcal{S})$, $(\mathcal{H}^\star)$, $(\mathcal{LL})$, $(\mathcal{N}^\star)$ *and* $(\mathcal{C}_x)$, $\widetilde{\mathbb{C}}_{xh} = \sqrt{nh}(\widetilde{C}_{xh} - C_x)$ *converges weakly in the space* $\ell^\infty([0, 1]^2)$ *to the Gaussian limit*

$$\widetilde{\mathbb{C}}_x(u_1, u_2) = \mathbb{G}_x(u_1, u_2) - C_x^{[1]}(u_1, u_2)\, \mathbb{G}_x(u_1, 1) - C_x^{[2]}(u_1, u_2)\, \mathbb{G}_x(1, u_2),$$

where $\mathbb{G}_x$ *is a Gaussian process with mean*

$$\mathrm{E}\left\{\mathbb{G}_x(u_1, u_2)\right\} = \frac{\kappa}{2} \left(\int_{\mathbb{R}} z^2 K(z)\, \mathrm{d}z \right) \ddot{C}_x(u_1, u_2)$$

and covariance function defined for $(u_1, u_2), (u_1', u_2') \in [0, 1]^2$ *by*

$$\mathrm{Cov}\left\{\mathbb{G}_x(u_1, u_2), \mathbb{G}_x(u_1', u_2')\right\} = \frac{1}{f_X(x)} \left(\int_{\mathbb{R}} \{K(z)\}^2\, \mathrm{d}z \right) \sigma_{C_x}^2(u_1, u_2, u_1', u_2'),$$

and where for $j \in \{1, 2\}$, $C_x^{[j]}$ *is defined as* 0 *on the set* $\{(u_1, u_2) \in [0, 1]^2 : u_j \in \{0, 1\}\}$.

Note that $\mathbb{G}_x$ is the Gaussian process identified in Proposition 1 in the special when $H_x = C_x$. Similarly as $\mathbb{C}_{xh}$, one can deduce that the limit of $\widetilde{\mathbb{C}}_{xh}$ under α-mixing is the same as that obtained by Veraverbeke et al. (2011) in the i.i.d. case. Similarly as for the first estimator, additional conditions compared to Veraverbeke et al. (2011) were needed. Indeed, the result holds under the additional Assumption ($\mathcal{S}$) about the serial structure of the process, the Lipchitz continuity of $u \mapsto \dot{F}_{jz}\{F_{jz}^{-1}(u)\}$, as well as $n^{8-\delta}h^{21} \to \infty$ for some $\delta > 0$ and $\max(h/h_1, h/h_2) < \infty$.

3 Behavior in small samples of the two conditional copula estimators

To evaluate the performance of the estimators C_{xh} and $\widetilde{C}_{xh}$ described in (4) and (6), respectively, consider the first-order vector autoregressive process

$$\begin{pmatrix} Y_{1t} \\ Y_{2t} \\ X_t \end{pmatrix} = \begin{pmatrix} \theta\, Y_{1,t-1} \\ \theta\, Y_{2,t-1} \\ \theta\, X_{t-1} \end{pmatrix} + \begin{pmatrix} \sqrt{1-\theta^2}\,\varepsilon_{1t} \\ \sqrt{1-\theta^2}\,\varepsilon_{2t} \\ \sqrt{1-\theta^2}\,\varepsilon_{3t} \end{pmatrix},$$

where $|\theta| < 1$ and $(\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})_{t\in\mathbb{Z}}$ is a process of i.i.d. innovations from the three-dimensional standard normal distribution with correlation matrix $\Sigma \in \mathbb{R}^{3\times 3}$. This model entails that (Y_{1t}, Y_{1t}, X_t) follows a standard Normal with correlation Σ for each $t \in \mathbb{Z}$. From well-known results on the multivariate normal distribution, one deduces that the conditional distribution of (Y_{1t}, Y_{2t}) given $X_t = x$ is the bivariate standard Normal with correlation coefficient

$$\rho_x = \frac{\Sigma_{12} - \Sigma_{13}\,\Sigma_{23}}{\sqrt{1-\Sigma_{13}^2}\sqrt{1-\Sigma_{23}^2}}.$$

As a consequence, C_x in that case is the Normal copula with parameter ρ_x, i.e. for Φ being the cumulative distribution function of the standard univariate normal distribution and h_{ρ_x} the bivariate standard normal density with correlation ρ_x,

$$C_x(u_1, u_2) = \int_{-\infty}^{\Phi^{-1}(u_2)} \int_{-\infty}^{\Phi^{-1}(u_1)} h_{\rho_x}(s_1, s_2)\, \mathrm{d}s_1\, \mathrm{d}s_2.$$

It can be seen easily that Assumption ($\mathcal{S}$) holds for this model.

The performance of C_{xh} and $\widetilde{C}_{xh}$ under this vector autoregressive process has been investigated in the light of the average integrated squared bias (AISB) and average integrated variance (AIV) defined respectively by

$$\mathrm{AISB}(\widehat{C}) = \int_{[0,1]^2} \left[\mathrm{E}\left\{\widehat{C}(u_1, u_2)\right\} - C_x(u_1, u_2)\right]^2 \mathrm{d}u_1 \mathrm{d}u_2$$

and

$$\mathrm{AIV}(\widehat{C}) = \int_{[0,1]^2} \left[\left\{ \widehat{C}(u_1, u_2) \right\}^2 - \left\{ \mathrm{E}\left(\widehat{C}(u_1, u_2) \right) \right\}^2 \right] \mathrm{d}u_1 \mathrm{d}u_2.$$

For each of the scenarios that have been considered, $\mathrm{AISB}(C_{xh})$, $\mathrm{AISB}(\widetilde{C}_{xh})$, $\mathrm{AIV}(C_{xh})$ and $\mathrm{AIV}(\widetilde{C}_{xh})$ have been estimated from 1000 Monte–Carlo experiments based on the triweight kernel, i.e. the density

$$K(y) = \frac{35}{32}(1 - y^2)^3\, \mathbb{I}(|y| \leq 1).$$

The latter satisfies Assumption $(\mathcal{LL})$. For the second estimator, the bandwidth parameters have been set such that $h = h_1 = h_2$.

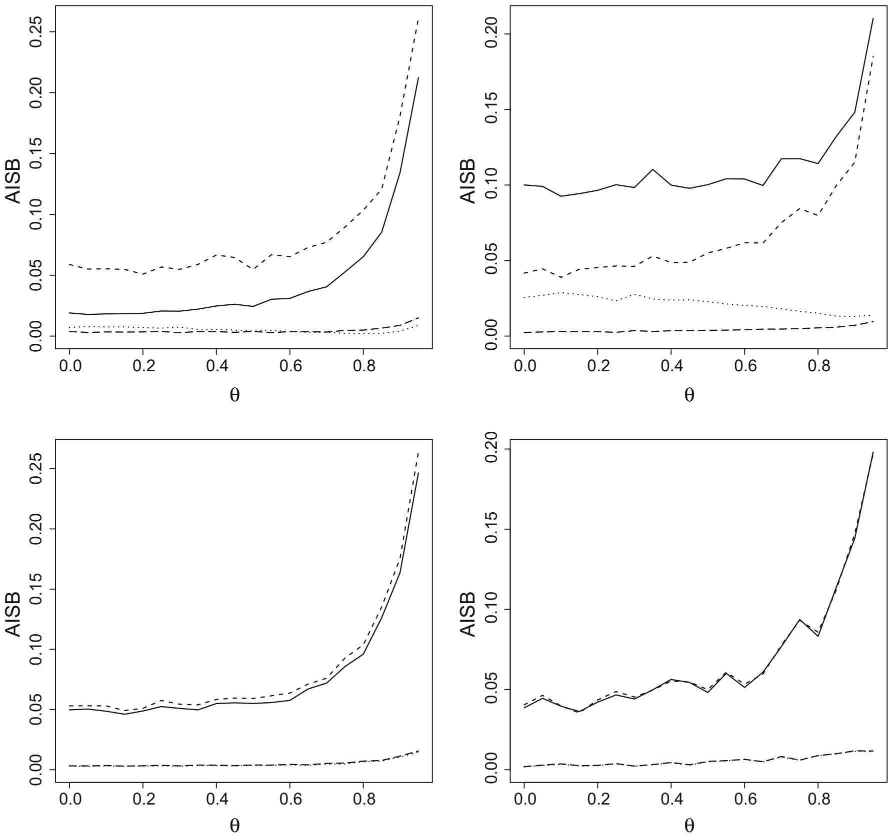

Fig. 1 Average integrated squared bias (AISB) ($\times 10^4$), as estimated from 1000 replicates, of C_{xh} (solid line for $n = 250$ and dotted line for $n = 1000$) and $\widetilde{C}_{xh}$ (small dashed line for $n = 250$ and long dashed line for $n = 1000$) as a function of $\theta \in [0, 1)$ under four scenarios of the first-order autoregressive process, namely $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (.9, .8, .8)$ (top left), $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (-.9, .8, -.8)$ (top right), $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (.8, .1, .1)$ (bottom left) and $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (.1, .1, .1)$ (bottom right)

The first investigation concerns the effect of the level of serial dependence, as controlled by the parameter θ, on the performance of the two conditional copula estimators. The curves of AISB as a function of $\theta \in [0, 1)$ when $n \in \{250, 1000\}$ are to be found in Fig. 1, while the curves of AIV are in Fig. 2. Four scenarios have been considered, namely $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (0.9, 0.8, 0.8), (-.9, .8, -.8), (.8, .1, .1), (.1, .1, .1)$; the corresponding values of the conditional correlation coefficient are $\rho_x = 0.26, -0.26, 0.79, 0.09$. Note that the bandwidth parameter has been set to $h = 1$; results with other values of h, not presented here, show very similar results.

First note that both AISB and AIV take smaller values when $n = 1000$ compared to $n = 250$, as expected. The influence of the serial parameter θ on the accuracy of the two estimators is apparent when $n = 250$, mainly for values close to 1, in which case the estimators are less accurate. However, this influence of serial dependence almost disappears when $n = 1000$, in accordance with the theoretical results that states that the estimators behave like in the i.i.d. case asymptotically. Differ-

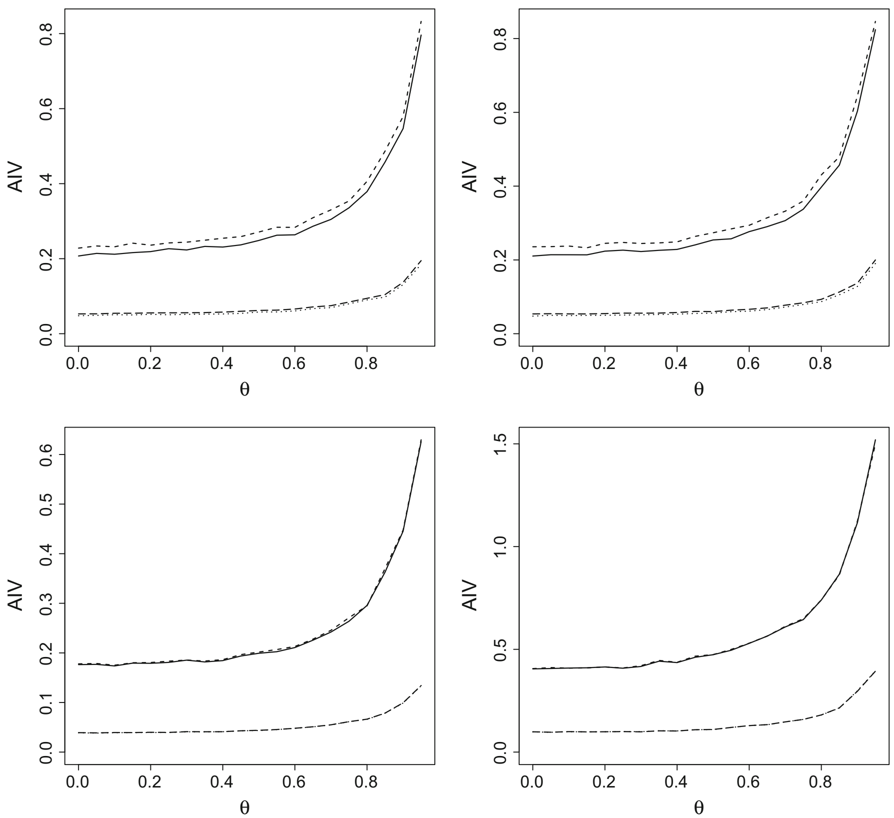

Fig. 2 Average integrated variance (AIV) ($\times 10^4$), as estimated from 1000 replicates, of C_{xh} (solid line for $n = 250$ and dotted line for $n = 1000$) and $\widetilde{C}_{xh}$ (small dashed line for $n = 250$ and long dashed line for $n = 1000$) as a function of $\theta \in [0, 1)$ under four scenarios of the first-order autoregressive process, namely $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (.9, .8, .8)$ (top left), $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (-.9, .8, -.8)$ (top right), $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (.8, .1, .1)$ (bottom left) and $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (.1, .1, .1)$ (bottom right)

ences in terms of bias are visible when $n = 250$ and under the scenario where $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (0.9, 0.8, 0.8)$, in which case C_{xh} is more accurate than $\widetilde{C}_{xh}$, and when $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (-.9, .8, -.8)$, where it is $\widetilde{C}_{xh}$ that is better. Otherwise, the two estimators are very similar.

Figures 3 and 4 explore the influence of the choice of the bandwidth parameter h on the performance of the two estimators in terms of AISB and AIV, respectively. The results are presented only when $n = 1000$ and under the vector autoregressive process when the serial parameter is set to $\theta = 0.4$. As already argued, the results are very similar for other levels of serial dependence.

First note that $\widetilde{C}_{xh}$ outperforms C_{xh} in terms of AISB when $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (0.9, 0.8, 0.8)$ and $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (-.9, .8, -.8)$. An explanation of this behavior is the fact that F_{1x} and F_{2x} depend strongly on the covariate under these scenarios, which in turn influences $\mathrm{E}(\mathbb{C}_{xh})$. The AISB are similar when $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) \in$

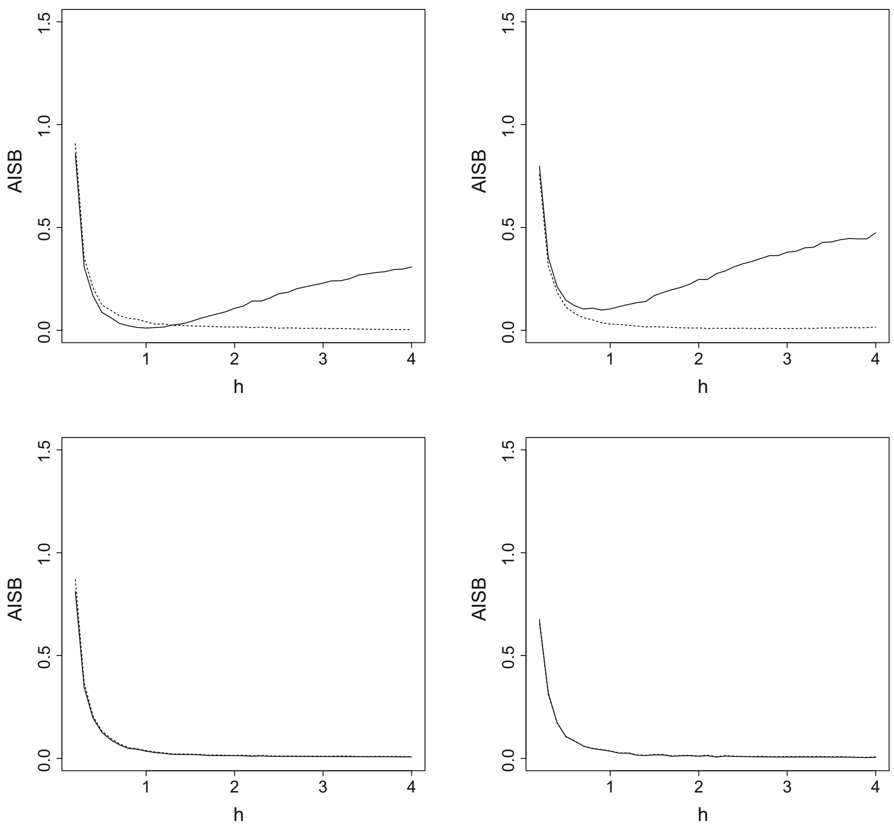

Fig. 3 Average integrated squared bias (AISB) ($\times 10^4$), as estimated from 1000 replicates, of C_{xh} (solid line) and $\widetilde{C}_{xh}$ (dashed line) as a function of the bandwidth parameter h when $n = 1000$ under four scenarios of the first-order autoregressive process with $\theta = 0.4$, namely $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (.9, .8, .8)$ (top left), $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (-.9, .8, -.8)$ (top right), $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (.8, .1, .1)$ (bottom left) and $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (.1, .1, .1)$ (bottom right)

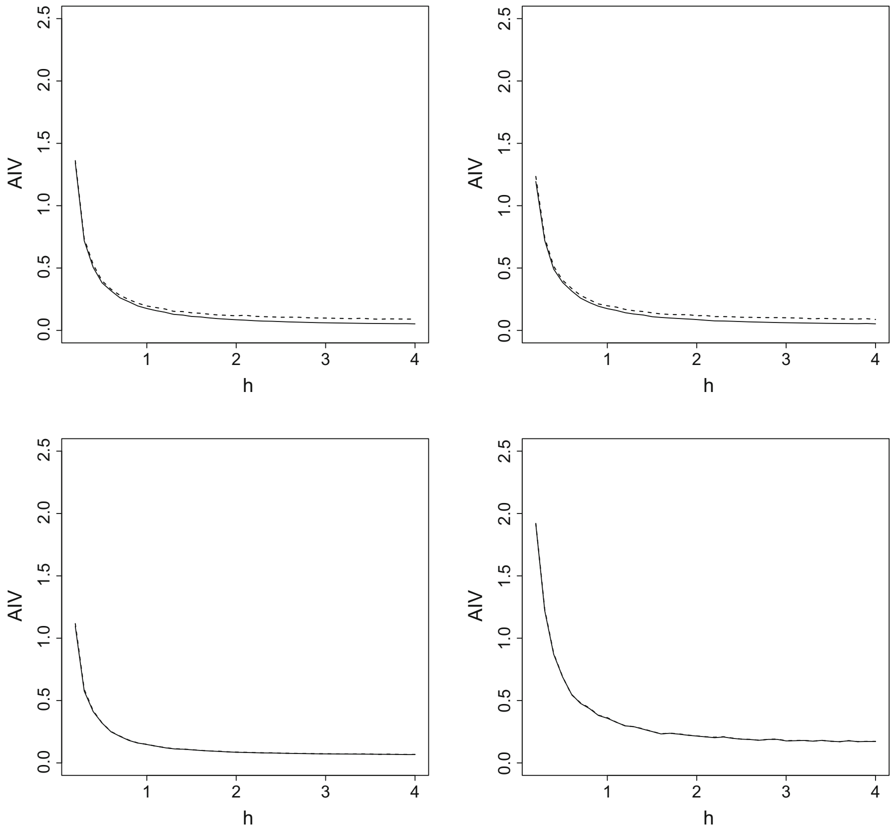

Fig. 4 Average integrated varaince (AIV) ($\times 10^4$), as estimated from 1000 replicates, of C_{xh} (solid line) and $\widetilde{C}_{xh}$ (dashed line) as a function of the bandwidth parameter h when $n = 1000$ under four scenarios of the first-order autoregressive process with $\theta = 0.4$, namely $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (.9, .8, .8)$ (top left), $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (-.9, .8, -.8)$ (top right), $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (.8, .1, .1)$ (bottom left) and $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (.1, .1, .1)$ (bottom right)

$\{(0.8, 0.1, 0.1), (0.1, 0.1, 0.1)\}$, which correspond to cases where the influence on the marginal distributions is rather weak. In the light of AIV, the two estimators are very similar under the four scenarios.

In the case of $\widetilde{C}_{xh}$, both the AISB and the AIV decrease as a function of the bandwidth parameter, but tend to stabilise when, say, $h \geq 2$. The behavior of C_{xh} is similar to that of $\widetilde{C}_{xh}$ in term of AIV. However, under the first two scenarios, i.e. when $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (0.9, 0.8, 0.8)$ and $(\Sigma_{12}, \Sigma_{23}, \Sigma_{13}) = (-.9, .8, -.8)$, one should choose a bandwidth parameter close to $h = 1$, since otherwise the bias can become quite large.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Appendix A: Proof of the main theoretical results

This section is devoted to the proof of Proposition 1, Proposition 2 and Proposition 3. Some arguments given in this section rely on Lemma 1, Lemma 2, Lemma 3 and Lemma 4 whose proofs can be found in the technical report by Bouezmarni et al. (2019). One of these results is Lemma 1 stated below; the latter is helpful to demonstrate the main results in Proposition 1–3. Basically, this result identifies the random behaviour of the local linear system of weights under α-mixing as the sample size gets large.

Lemma 1 *Under Assumptions ($\mathcal{S}$), ($\mathcal{LL}$) and ($\mathcal{N}$), one has almost surely that as $n \to \infty$,*

$$\sup_{z \in J_x} \sup_{u : K(u) > 0} \frac{1}{K(u)} \left| \mathcal{K}_{zn}(u) - \frac{K(u)}{f_X(z)} \right| \to 0.$$

A.1: Proof of Proposition 1

According for instance to Theorem 1.5.4 of van der Vaart and Wellner (1996), weak convergence in $\ell^\infty(\mathbb{R}^2)$ is equivalent to the finite-dimensional convergence combined with the asymptotic tightness. That the finite-dimensional distributions of $\mathbb{H}_{xh}$ converge to those of $\mathbb{H}_x$ under α-mixing is a consequence of Theorem 6 of Masry and Fan (1997) and of the Cramér–Wold device. In particular, one deduces

$$\mathrm{E}\left\{\mathbb{H}_x(y_1, y_2)\right\} = \frac{\kappa}{2} \left(\int_{\mathbb{R}} z^2 K(z)\,\mathrm{d}z \right) \ddot{H}_x(y_1, y_2)$$

and for $\sigma^2_{H_x}$ defined in (5),

$$\mathrm{Cov}\left\{\mathbb{H}_x(y_1, y_2), \mathbb{H}_x(y_1', y_2')\right\} = \frac{1}{f_X(x)} \left(\int_{\mathbb{R}} \{K(z)\}^2\,\mathrm{d}z \right) \sigma^2_{H_x}(y_1, y_2, y_1', y_2').$$

In order to show the asymptotic tightness of $\mathbb{H}_{xh}$, define $Z^\star_{xh} = \sqrt{nh}\left(\bar{H}_{xh} - H_x\right)$ and $Z_{xh} = \sqrt{nh}\left(H_{xh} - \bar{H}_{xh}\right)$, where

$$\bar{H}_{xh}(y_1, y_2) = \frac{1}{nh} \sum_{i=1}^n \mathcal{K}_{xn}\left(\frac{X_i - x}{h}\right) H_{X_i}(y_1, y_2).$$

One can then write $\mathbb{H}_{xh} = Z^\star_{xh} + Z_{xh}$, so that the asymptotic tightness of $\mathbb{H}_{xh}$ will follow from that of both $Z^\star_{xh}$ and Z_{xh}. For $Z^\star_{xh}$, note that a Taylor expansion of order two allows to write that for some ζ_i between X_i and x,

$$H_{X_i}(y_1, y_2) = H_x(y_1, y_2) + (X_i - x)\, \dot{H}_x(y_1, y_2) + \frac{1}{2}\,(X_i - x)^2\, \ddot{H}_{\zeta_i}(y_1, y_2).$$

Using the fact that

$$\frac{1}{nh}\sum_{i=1}^{n} \mathcal{K}_{xn}\left(\frac{X_i - x}{h}\right) = 1 \quad \text{and} \quad \frac{1}{nh}\sum_{i=1}^{n} (X_i - x)\, \mathcal{K}_{xn}\left(\frac{X_i - x}{h}\right) = 0,$$

one deduces from straightforward computations that

$$Z^{\star}_{xh}(y_1, y_2) = \frac{1}{\sqrt{nh}}\sum_{i=1}^{n} \frac{1}{2}\,(X_i - x)^2\, \mathcal{K}_{xn}\left(\frac{X_i - x}{h}\right) \ddot{H}_{\zeta_i}(y_1, y_2).$$

Since Assumptions ($\mathcal{S}$), ($\mathcal{L}$) and ($\mathcal{N}$) are satisfied, and because $z \mapsto \ddot{H}_z$ is uniformly continuous in a neighbourhood of x (see Condition ($\mathcal{H}$)), one can invoke Lemma 1 and write

$$\begin{aligned}
Z^{\star}_{xh}(y_1, y_2) &= \frac{1}{\sqrt{nh}}\sum_{i=1}^{n} \frac{1}{2}\,(X_i - x)^2 \left\{ \mathcal{K}_{xn}\left(\frac{X_i - x}{h}\right) \right. \\
&\quad \left. -\frac{1}{f_X(x)}\, K\left(\frac{X_i - x}{h}\right) \right\} \ddot{H}_{\zeta_i}(y_1, y_2) \\
&\quad + \frac{1}{2\, f_X(x)}\, \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} (X_i - x)^2\, K\left(\frac{X_i - x}{h}\right) \ddot{H}_{\zeta_i}(y_1, y_2) \\
&= \frac{1 + o_{as}(1)}{2\, f_X(x)}\, \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} (X_i - x)^2\, K\left(\frac{X_i - x}{h}\right) \ddot{H}_{\zeta_i}(y_1, y_2) \\
&= \frac{\ddot{H}_x(y_1, y_2) + o_{as}(1)}{2\, f_X(x)}\, \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} (X_i - x)^2\, K\left(\frac{X_i - x}{h}\right) \\
&= \frac{\ddot{H}_x(y_1, y_2) + o_{as}(1)}{2\, f_X(x)} \sqrt{nh^5}\, S_{n,2}(x).
\end{aligned}$$

Now according to Corollary 1 of Masry (1996), Assumptions ($\mathcal{S}$), ($\mathcal{LL}$) and ($\mathcal{N}$) ensure that

$$S_{n,2}(x) = f_X(x) \int_{\mathbb{R}} z^2 K(z)\, \mathrm{d}z + o_{as}(1).$$

Since Assumption ($\mathcal{N}$) ensures that $nh^5 \to \kappa^2 < \infty$ as $n \to \infty$, one can conclude that

$$Z^{\star}_{xh}(y_1, y_2) = \frac{\kappa \ddot{H}_x(y_1, y_2)}{2} \int_{\mathbb{R}} z^2 K(z)\, \mathrm{d}z \; + o_{as}(1). \tag{7}$$

In view of Assumption ($\mathcal{H}$), one can conclude that $Z^{\star}_{xh}$ is asymptotically tight.

Now to show that Z_{xh} is also asymptotically tight, consider for a fixed $x \in \mathbb{R}$ and $\mathbf{y} = (y_1, y_2)$, $\mathbf{y}' = (y_1', y_2')$, the semi-metric

$$\rho(\mathbf{y}, \mathbf{y}') = \left|F_{1x}(y_1) - F_{1x}(y_1')\right| + \left|F_{2x}(y_2) - F_{2x}(y_2')\right|$$

and define for $\delta > 0$, $f : \mathbb{R}^2 \to \mathbb{R}$ bounded and $T \subseteq \mathbb{R}^2$,

$$\mathfrak{W}_\delta(f, T) = \sup_{\mathbf{y},\mathbf{y}' \in T;\, \rho(\mathbf{y},\mathbf{y}') < \delta} \left| f(\mathbf{y}) - f(\mathbf{y}')\right|.$$

The modulus of ρ-equicontinuity of Z_{xn} is then given by $\mathfrak{W}_{\cdot}(Z_{xn}, \mathbb{R}^2)$. For a fixed $\mathbf{y} \in \mathbb{R}^2$, the random variable $Z_{xn}(\mathbf{y})$ is asymptotically tight in $\mathbb{R}$, so to prove that the process Z_{xn} is asymptotically tight in $\ell^\infty([0, 1]^2)$, it suffices to show (see Theorem 1.5.7 in van der Vaart and Wellner (1996)) that for every $\epsilon > 0$,

$$\lim_{\delta \downarrow 0} \lim_{n \to \infty} \mathbb{P}\left\{\mathfrak{W}_\delta(Z_{xn}, \mathbb{R}^2) > \epsilon\right\} = 0. \tag{8}$$

Note that if the available observations were serially independent, one could proceed as in Veraverbeke et al. (2011) and use the empirical process machinery developed in van der Vaart and Wellner (1996) to show that (8) holds by checking few conditions on the process Z_{xh}. Under a mixing assumption, however, these arguments cannot be used anymore. To overcome this problem, and following for example Bücher and Kojadinovic (2016) (see the proof of Lemma A.3 therein), a possibility is to proceed in the spirit of Theorem 3 of Bickel and Wichura (1971) and to study the increments of the process Z_{xn} on blocks. Specifically, for an arbitrary non-empty rectangle $A \in \mathbb{R}^2$, define

$$\mathbb{H}_{xh}(A) = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \mathcal{K}_{xn}\left(\frac{X_i - x}{h}\right)\left[\mathbb{I}\{(Y_{1i}, Y_{2i}) \in A\} - \nu_{X_i}(A)\right],$$

where $\nu_x(A) = \mathbb{P}\{(Y_{1i}, Y_{2i}) \in A | X_i = x\}$. The next result whose proof can be found in Bouezmarni et al. (2019) provides a bound on the moment of order six of $\mathbb{H}_{xh}(A)$.

Lemma 2 *Under Assumptions* ($\mathcal{S}$), ($\mathcal{H}$) *and* ($\mathcal{LL}$), *one can find a finite constant* $\omega > 0$ *such that for all* b *satisfying* $0 < b < \min\{(a-6)/a, 2/5\}$,

$$\mathrm{E}\left\{|\mathbb{H}_{xh}(A)|^6\right\} \le \omega \left\{\frac{\nu_x(A)}{n^2h^2} + \frac{\nu_x(A)^{2-\frac{4}{a}}}{nh} + \nu_x(A)^{3-\frac{6}{a}} + \mathcal{J}_n(h, b)\right\},$$

where $\mathcal{J}_n(h, b) = h^4 h^{2b} + h^b (nh)^{-1} + h^{5b}(nh)^{-2}$.

Now for $\gamma \in (0, 1/2)$, define the product space $T_\gamma = T_\gamma^{(1)} \times T_\gamma^{(2)}$, where for $\kappa_\gamma = \lfloor (nh)^{1/2+\gamma} \rfloor$,

$$T_\gamma^{(j)} = \left\{ F_{jx}^{-1}(0), F_{jx}^{-1}\left(\frac{1}{\kappa_\gamma}\right), \ldots, F_{jx}^{-1}(1) \right\}, \quad j \in \{1, 2\}.$$

Lemma 3 *Under Assumptions* ($\mathcal{S}$), ($\mathcal{LL}$) *and* ($\mathcal{N}$)*, one has for n sufficiently large that for any* $\epsilon > 0$ *and* $\delta > 2\kappa_\gamma^{-1}$,

$$\mathbb{P}\left\{ \mathfrak{W}_\delta(Z_{xn}, \mathbb{R}^2) > \epsilon \right\} \leq \mathbb{P}\left\{ \mathfrak{W}_{2\delta}(Z_{xn}, T_\gamma) > \frac{\epsilon}{3} \right\}.$$

Lemma 3 entails that (8) will hold if for any $\epsilon > 0$,

$$\lim_{\delta \downarrow 0} \lim_{n \to \infty} \mathbb{P}\left\{ \mathfrak{W}_\delta(Z_{xn}, T_\gamma) > \epsilon \right\} = 0. \tag{9}$$

According to Problem 2.1.5 in van der Vaart and Wellner (1996), Eq. (9) holds if and only if for any sequence $\delta_n \downarrow 0$,

$$\lim_{n \to \infty} \mathbb{P}\left\{ \mathfrak{W}_{\delta_n}(Z_{xn}, T_\gamma) > \epsilon \right\} = 0.$$

In order to show that it is the case, one proceeds as in Bücher and Kojadinovic (2016) and uses Lemma 2 of Balacheff and Dupont (1980). To this end, define $\lambda_x(B_1 \times B_2) = \mathbb{P}(Y_1 \in B_1 | X = x) \times \mathbb{P}(Y_2 \in B_2 | X = x)$ for any $B_1, B_2 \subset \mathbb{R}$. At this point, letting $\mu_x = \nu_x + \lambda_x$, note that for any rectangle A_γ whose corner points are all distinct and lie in T_γ,

$$\left(\frac{1}{nh}\right)^{1+2\gamma} \leq \mu_x(A_\gamma) \leq 2.$$

As a consequence, the Markov inequality and Lemma 2 entails that for any $\eta > 0$ and β such that $0 < \beta < (a - 5)/a$,

$$
\begin{aligned}
\mathbb{P}\left\{\left|\mathbb{H}_{xh}(A_\gamma)\right| > \eta\right\} &\leq \frac{\omega}{\eta^6}\left\{\frac{\mu_x(A_\gamma)}{n^2h^2} + \frac{\mu_x(A_\gamma)^{2-\frac{4}{a}}}{nh} + \mu_x(A_\gamma)^{3-\frac{6}{a}} + \mathcal{J}_n(h,b)\right\} \\
&\leq \frac{\omega}{\eta^6}\,\mu_x(A_\gamma)^{1+\beta}\left\{\frac{\mu_x(A_\gamma)^{-\beta}}{n^2h^2}\right. \\
&\quad \left. + \frac{\mu_x(A_\gamma)^{\frac{1}{a}}}{nh} + \mu_x(A_\gamma)^{1-\frac{1}{a}} + \mu_x(A_\gamma)^{-1-\beta}\mathcal{J}_n(h,b)\right\} \\
&\leq \frac{\omega}{\eta^6}\,\mu_x(A_\gamma)^{1+\beta}\left\{\frac{(nh)^{(1+2\gamma)\beta}}{n^2h^2} + 4\right. \\
&\quad \left. + (nh)^{(1+2\gamma)(1+\beta)}\mathcal{J}_n(h,b)\right\}.
\end{aligned}
$$

From Assumption ($\mathcal{N}$), $nh^5 \to \kappa^2$, so that nh^5 is bounded above by some positive and finite constant cst as $n \to \infty$. It follows that

$$
(nh)^{(1+2\gamma)(1+\beta)}h^4h^{2b} \leq \mathrm{cst}(h^{-4})^{\beta+2\gamma\beta+2\gamma}h^{2b} = \mathrm{cst}(h^2)^{b-2(\beta+2\gamma\beta+2\gamma)}.
$$

In addition, since $h < 1$ and $nh > 1$ for n sufficiently large, one has

$$
\begin{aligned}
(nh)^{(1+2\gamma)(1+\beta)}\left\{\frac{h^{5b}}{(nh)^2} + \frac{h^b}{nh}\right\} &\leq \mathrm{cst}\,(nh)^{2\gamma+\beta+2\gamma\beta}h^b \\
&\leq \mathrm{cst}\,h^{b-4(2\gamma+\beta+2\gamma\beta)}.
\end{aligned}
$$

It follows that for any $\beta, \gamma \in (0, b/16)$ and n sufficiently large,

$$
(nh)^{(1+2\gamma)(1+\beta)}\mathcal{J}_n(h,b) < 1. \tag{10}
$$

One can then write

$$
\mathbb{P}\left\{\left|\mathbb{H}_{xh}(A_\gamma)\right| > \eta\right\} \leq \frac{6\omega}{\eta^6}\,\mu_x(A_\gamma)^{1+\beta}. \tag{11}
$$

Now let $\widetilde{\mu}_x$ be the finite positive measure such that for $(y_1, y_2) \in T_\gamma$, $\widetilde{\mu}_x(\{(y_1, y_2)\})$ vanishes if $F_{1x}(y_1) = 0$ or $F_{2x}(y_2) = 0$, and $\widetilde{\mu}_x(\{(y_1, y_2)\}) = \mu_x(]\underline{y}_1, y_1]\times]\underline{y}_2, y_2])$ otherwise, where $\underline{y}_j = \max\{\xi \in T_\gamma^{(j)} : \xi < y_j\}$. Since $\mu_x(A_\gamma) = \widetilde{\mu}(A_\gamma \cap T_\gamma)$, the inequality in (11) may be expressed equivalently as

$$
\mathbb{P}\left\{\left|\mathbb{H}_{xh}(A_\gamma)\right| > \eta\right\} \leq \frac{6\omega}{\eta^6}\,\widetilde{\mu}_x(A_\gamma \cap T_\gamma)^{1+\beta}.
$$

For $\delta_n \downarrow 0$, define $\delta_n' \downarrow 0$ in such a way that for each $n \in \mathbb{N}$, $\delta_n' \in \{1, 1/2, 1/3, 1/4, \ldots\}$ and $\delta_n' \geq \max(\delta_n, \kappa_\gamma^{-1})$. From Lemma 2 of Balacheff and Dupont (1980), one deduces by a straightforward reparametrization that there exists a constant $\vartheta = \vartheta(\epsilon, \beta) > 0$ such that

$$\begin{aligned}
&\mathbb{P}\left\{\mathfrak{W}_{\delta_n}(Z_{xn}, T_\gamma) > \epsilon\right\} \\
&\quad \le \mathbb{P}\left\{\mathfrak{W}_{\delta_n'}(Z_{xn}, T_\gamma) > \epsilon\right\} \\
&\quad \le \vartheta(\epsilon, \beta)\, \widetilde{\mu}_x(T_\gamma) \left\{ \sup_{\substack{y<y' \in T_\gamma^{(1)} \\ F_{1x}(y') - F_{1x}(y) \le 3\delta_n'}} \widetilde{\mu}_x\left(]y, y'] \cap T_\gamma^{(1)} \times T_\gamma^{(2)}\right) \right. \\
&\qquad \left. + \sup_{\substack{y<y' \in T_\gamma^{(2)} \\ F_{2x}(y') - F_{2x}(y) \le 3\delta_n'}} \widetilde{\mu}_x\left(T_\gamma^{(1)} \times]y, y'] \cap T_\gamma^{(2)}\right) \right\}^{\beta} .
\end{aligned}$$

One can then conclude that as $n \to \infty$,

$$\begin{aligned}
\mathbb{P}\left\{\mathfrak{W}_{\delta_n}(Z_{xn}, T_\gamma) > \epsilon\right\} \le \vartheta(\epsilon, \beta)\, \mu_x(\mathbb{R}^2) &\left\{ \sup_{\substack{y<y' \in T_\gamma^{(1)} \\ F_{1x}(y') - F_{1x}(y) \le 3\delta_n'}} \mu_x\left(]y, y'] \times \mathbb{R}\right) \right. \\
&\left. + \sup_{\substack{y<y' \in T_\gamma^{(2)} \\ F_{2x}(y') - F_{2x}(y) \le 3\delta_n'}} \mu_x\left(\mathbb{R} \times]y, y']\right) \right\}^{\beta} ,
\end{aligned}$$

so that $\mathbb{P}\left\{\mathfrak{W}_{\delta_n}(Z_{xn}, T_\gamma) > \epsilon\right\} \to 0$. This finally entails that (9) is satisfied, which in turn ensures that (8) holds true. The proof is therefore complete.

A.2: Proof of Proposition 2

Let $V_{1i} = F_{1x}(Y_{1i})$, $V_{2i} = F_{2x}(Y_{2i})$, and consider

$$J_{xh}(u, v) = \frac{1}{nh} \sum_{i=1}^{n} \mathcal{K}_{xn}\left(\frac{X_i - x}{h}\right) \mathbb{I}(V_{1i} \le u, V_{2i} \le v),$$

$I_{1xh}(y) = J_{xh}(y, 1)$ and $I_{2xh}(y) = J_{xh}(1, y)$. As $C_{xh}(u, v) = J_{xh}\{I_{1xh}^{-1}(u), I_{2xh}^{-1}(u)\}$, then $\mathbb{C}_{xh} = \sqrt{nh}\{J_{xh}(I_{1xh}^{-1}, I_{2xh}^{-1}) - C_x\}$. Now define $\mathbb{D}$ as the space of bivariate distribution functions J on $[0, 1]^2$ whose marginal cumulative distribution functions I_1 and I_2 satisfy $I_1(0) = I_2(0) = 0$, and consider the mapping $\Lambda : \mathbb{D} \to \mathbb{D}$ such that for $J \in \mathbb{D}$,

$$\Lambda(J)(u_1, u_2) = J\left\{I_1^{-1}(u_1), I_2^{-1}(u_2)\right\}.$$

With this notation,

$$\mathbb{C}_{xh} = \sqrt{nh}\left\{\Lambda(J_{xh}) - C_x\right\} = \sqrt{nh}\left\{\Lambda(J_{xh}) - \Lambda(C_x)\right\}.$$

Also, let $\mathbb{D}_0 = \{\alpha \in C([0,1]^2) : \alpha(1,1) = 0 \text{ and } \alpha(z_1, z_2) = 0 \text{ if } \min(z_1, z_2) = 0\}$, where $C([0,1]^2)$ is the space of continuous functions on $[0,1]^2$. From Theorem 2.4 in Bücher and Volgushev (2013), one has in view of Assumption $(\mathcal{C}_x)$ that Λ is Hadamard differentiable at C_x tangentially to $\mathbb{D}_0$, with derivative given for $\Delta \in \mathbb{D}_0$ by

$$\Lambda'_{C_x}(\Delta)(u_1, u_2) = \Delta(u_1, u_2) - C_x^{[1]}(u_1, u_2)\,\Delta(u_1, 1) - C_x^{[2]}(u_1, u_2)\,\Delta(1, u_2).$$

It is a consequence of Lemma 1 that under Assumptions $(\mathcal{S})$, $(\mathcal{LL})$ and $(\mathcal{N})$ it holds for sufficiently large n that $\mathcal{K}_{xn} \geq 0$ almost surely, yielding $J_{xh} \in \mathbb{D}$. Moreover, under conditions $(\mathcal{S})$, $(\mathcal{H})$, $(\mathcal{LL})$ and $(\mathcal{N})$, it can easily be shown that $(V_{11}, V_{21}, X_1), \ldots, (V_{1n}, V_{2n}, X_n)$ fulfill the requirements of Proposition 1. Therefore, one deduces that $\sqrt{nh}(J_{xh} - C_x)$ converges weakly in $l^\infty([0,1]^2)$ to $\mathbb{B}_x$, where $\mathbb{B}_x(u_1, u_2) = \mathbb{H}_x\{F_{1x}^{-1}(u_1), F_{2x}^{-1}(u_2)\} \in \mathbb{D}_0$. Hence, from the functional delta method, one can then conclude that $\mathbb{C}_{xh}$ converges weakly to

$$\mathbb{C}_x = \Lambda'_{C_x}(\mathbb{B}_x) = \mathbb{B}_x(u_1, u_2) - C_x^{[1]}(u_1, u_2)\,\mathbb{B}_x(u_1, 1) - C_x^{[2]}(u_1, u_2)\,\mathbb{B}_x(1, u_2).$$

A.3: Proof of Proposition 3

Consider a version of G_{xh} based on $(U_1, V_1, X_1), \ldots, (U_n, V_n, X_n)$, where $U_i = F_{1X_i}(Y_{1i})$ and $V_i = F_{2X_i}(Y_{2i})$, namely

$$\widetilde{G}_{xh}(u, v) = \frac{1}{nh}\sum_{i=1}^{n} \mathcal{K}_{xn}\left(\frac{X_i - x}{h}\right)\mathbb{I}\left(U_i \leq u, V_i \leq v\right).$$

One can then write for the functional Λ defined in the proof of Proposition 2 that

$$\widetilde{\mathbb{C}}_{xh} = \sqrt{nh}\left\{\Lambda(\widetilde{G}_{xh}) - C_x\right\} + \sqrt{nh}\left\{\Lambda(G_{xh}) - \Lambda(\widetilde{G}_{xh})\right\}.$$

The first summand is a special case of Proposition 2 with (Y_{1i}, Y_{2i}, X_i) replaced by (U_i, V_i, X_i). Because the conditional marginal distributions of (U_i, V_i) are uniform on $(0,1)$, their joint conditional distribution is C_{X_i}. Since Assumptions $(\mathcal{S})$, $(\mathcal{H}^\star)$, $(\mathcal{LL})$ and $(\mathcal{C}_x)$ are satisfied, Proposition 2 ensures that $\sqrt{nh}\{\Lambda(\widetilde{G}_{xh}) - C_x\}$ converges weakly to $\Lambda'_{C_x}(\mathbb{G}_x) = \widetilde{\mathbb{C}}_x$. It remains to show that $\sqrt{nh}\{\Lambda(G_{xh}) - \Lambda(\widetilde{G}_{xh})\}$ is asymptotically negligible. As pointed out by Veraverbeke et al. (2011), this is closely related to the asymptotic behavior of the processes $\widetilde{Z}_{jxn} = Z_{jxn} - \bar{Z}_{jxn}$, $j \in \{1, 2\}$, where for $z_t = x + t\,C\,h$ and $C = \inf\{z > 0 : K(z) = 0\}$,

$$Z_{jxn}(t, u) = \sqrt{nh_j}\, F_{jz_t h_j} \circ F_{jz_t}^{-1}(u) \quad \text{and} \quad \bar{Z}_{jxn}(t, u) = \sqrt{nh_j} \sum_{i=1}^{n} F_{jX_i} \circ F_{jz_t}^{-1}(u).$$

The key is the following lemma whose proof is to be found in Bouezmarni et al. (2019).

Lemma 4 *Under Assumptions ($\mathcal{S}$), ($\mathcal{H}^\star$), ($\mathcal{LL}$) and ($\mathcal{N}^\star$), the sequences $\widetilde{Z}_{1xn}$ and $\widetilde{Z}_{2xn}$ are asymptotically tight in $\ell^\infty([-1, 1] \times [0, 1])$.*

Finally, from arguments similar as those in Appendix B.2 of Veraverbeke et al. (2011), one obtains that $\sqrt{nh}\{\Lambda(G_{xh}) - \Lambda(\widetilde{G}_{xh})\} = o_{\mathbb{P}}(1)$, and thus $\widetilde{\mathbb{C}}_{xh} = \sqrt{nh}\{\Lambda(\widetilde{G}_{xh}) - C_x\} + o_{\mathbb{P}}(1)$.

References

Balacheff S, Dupont G (1980) Normalité asymptotique des processus empiriques tronqués et des processus de rang (cas multidimensionnel mélangeant). In: Nonparametric asymptotic statistics (Proc. Conf., Rouen, 1979) (French), Lecture Notes in Math., vol 821, Springer, Berlin, pp 19–45

Bickel PJ, Wichura MJ (1971) Convergence criteria for multiparameter stochastic processes and some applications. Ann Math Stat 42:1656–1670

Bouezmarni T, Camirand Lemyre F, Quessy JF (2019) Supplementary material for the paper "On the large-sample behavior of two estimators of the conditional copula under serially dependent data". Tech. Rep. 2019-166, Département de mathématiques, Université de Sherbrooke, Sherbrooke, QC, Canada

Bücher A, Kojadinovic I (2016) A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing. Bernoulli 22(2):927–968

Bücher A, Volgushev S (2013) Empirical and sequential empirical copula processes under serial dependence. J Multivar Anal 119:61–70

Carrasco M, Chen X (2002) Mixing and moment properties of various GARCH and stochastic volatility models. Econom Theory 18(1):17–39

Doukhan P (1994) Mixing, Lecture Notes in Statistics, vol 85. Springer, New York (properties and examples)

Fan J, Gijbels I (1996) Local polynomial modelling and its applications. Monographs on statistics and applied probability: 66. Chapman & Hall/CRC, Boca Raton

Gijbels I, Veraverbeke N, Omelka M (2011) Conditional copulas, association measures and their applications. Comput Stat Data Anal 55(5):1919–1932

Masry E (1996) Multivariate local polynomial regression for time series: uniform strong consistency and rates. J Time Ser Anal 17(6):571

Masry E, Fan J (1997) Local polynomial estimation of regression functions for mixing processes. Scand J Stat 2:165

Meitz M, Saikkonen P (2008) Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models. Econom Theory 24(5):1291–1320

Nelsen RB (2006) An introduction to copulas, 2nd edn. Springer Series in Statistics, Springer, New York

Patton AJ (2006) Modelling asymmetric exchange rate dependence. Int Econom Rev 47(2):527–556

van der Vaart AW, Wellner JA (1996) Weak convergence and empirical processes. Springer Series in Statistics, Springer, New York (with applications to statistics)

Veraverbeke N, Omelka M, Gijbels I (2011) Estimation of a conditional copula and association measures. Scand J Stat 38(4):766–780