manuscript No.
(will be inserted by the editor)

A note on the large-sample behavior of two estimators of the conditional copula under serial data

Taoufik Bouezmarni · Félix Camirand Lemyre · Jean-François Quessy

Abstract As defined by [12], 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 $\mathbb{P}(Y_1 \le y_1, Y_2 \le y_2 \mid X = x) = C_x\{\mathbb{P}(Y_1 \le y_1 | X = x), \mathbb{P}(Y_2 \le y_2 | X = x)\}$. In this note, the weak convergence of two estimators of C_x proposed by [6] is established under strong-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 [14] under the i.i.d. setting. The performance of these estimators in finite sample sizes is investigated.

Keywords α-mixing processes · conditional copula · kernel estimation · weak convergence

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

This research was supported in part by individual grants from the Natural Sciences and Engineering Research Council of Canada (NSERC), by the Canadian Statistical Science Institute (CANSSI) and by the Australian Research Council.

T. Bouezmarni
Département de mathématiques, Université de Sherbrooke, Québec, Canada
Tel.: +1 819 821 8000, ext. 62035
Fax: +1 819 821 7189
E-mail: taoufik.bouezmarni@usherbrooke.ca

F. Camirand Lemyre
Department of mathematics and statistics, University of Melbourne, Parkville, Australia
E-mail: felix.camirand@unimelb.edu.au

J-F. Quessy
Département de mathématiques et d'informatique, Université du Québec à Trois-Rivières, Trois-Rivières, Canada
E-mail: jean-francois.quessy@uqtr.ca

Theorem. In its classical formulation, this result ensures that for any random vector $\mathbf{Y} = (Y_1, \ldots, Y_d)$, there exists a function $C : [0,1]^d \to [0,1]$ called the copula of $\mathbf{Y}$ such that for all $\mathbf{y} = (y_1, \ldots, y_d) \in \mathbb{R}^d$,

$$\mathbb{P}(\mathbf{Y} \leq \mathbf{y}) = C \left\{ \mathbb{P}(Y_1 \leq y_1), \ldots, \mathbb{P}(Y_d \leq y_d) \right\}.$$

When $Y_1, \ldots, Y_d$ are continuous, C is unique.

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 [6], 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 [12]. Formally, letting $H_x(y_1, y_2) = \mathbb{P}(Y_1 \leq y_1, Y_2 \leq y_2 | X = x)$, 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 $(y_1, y_2) \in \mathbb{R}^2$,

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

Two nonparametric estimators of C_x were proposed by [6] and their asymptotic behavior was formally investigated by [14] in the i.i.d. case. The purpose of this note is to extend these large-sample results to the case of time series, since many contexts of applications involve serially dependent observations.

To this end, one adopts a very general framework where the stationary process $\{(Y_{1t}, Y_{2t}, X_t)\}_{t \in \mathbb{Z}}$ satisfies a strong mixing condition. In a certain sense, these results are versions of [2] adapted to the context of conditional copulas. Specifically, let $\mathcal{F}_a^b$ be the σ-field generated by $\{(Y_{1t}, Y_{2t}, X_t)\}_{a \leq t \leq b}$ and define the α-mixing coefficients

$$\alpha(r) = \sup_{k \in \mathbb{Z}} \alpha \left(\mathcal{F}_{-\infty}^k, \mathcal{F}_{k+r}^\infty \right),$$

where

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{(A,B) \in \mathcal{A} \times \mathcal{B}} \left| \mathbb{P}(A \cap B) - \mathbb{P}(A)\, \mathbb{P}(B) \right|.$$

The process $\{(Y_{1t}, Y_{2t}, X_t)\}_{t \in \mathbb{Z}}$ is said to be α-mixing, or strong mixing, if $\alpha(r) \to 0$ as $r \to \infty$. Several parametric time series models satisfy this strong mixing assumption, including ARMA and GARCH models under appropriate restrictions on the parameters involved. For more details, see [10], [4] and [3].

The remaining of the paper is organized as follows. Section 2 establishes the asymptotic behavior of a first estimator of the conditional copula and provides a sketch of the proof. Section 3 mimics Section 2 for a second estimator which aims at reducing the bias. Section 4 presents the results of a numerical study that investigates the performance of the two estimators when computed from serially dependent data. The assumptions needed for the theoretical results to hold, as well as the proofs of some technical results, are to be found in the Appendix.

2 Investigation of a first estimator of C_x

2.1 Description of the estimator

Consider n realizations $(Y_{11}, Y_{21}, X_1), \ldots, (Y_{1n}, Y_{2n}, X_n)$ of a stationary process $\{(Y_{1t}, Y_{2t}, X_t)\}_{t\in\mathbb{Z}}$ that satisfies the strong mixing assumption. In that context, a first estimator of C_x arises naturally upon noting that

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

where $F_{1x}(y) = \mathbb{P}(Y_1 \leq y|X = x)$ and $F_{2x}(y) = \mathbb{P}(Y_2 \leq y|X = x)$. An estimator of the H_x will then provide a plug-in estimation of C_x. In this paper, the local linear kernel smoothing estimator is considered for H_x. To be more specific, let

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

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. This function is defined as

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

where $S_{n,\ell}(x) = (nh)^{-1}\sum_{i=1}^{n}\{(X_i - x)/h\}^{\ell}\,K\{(X_i - x)/h\}$, $\ell \in \{0, 1, 2\}$ and $K(\cdot)$ is a symmetric kernel density (see [5] for details on how to derive this estimator). Hereafter it will be assumed that the kernel is continuous, compactly supported and has bounded second order derivative. From Equation (2), an estimator of C_x is given by

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

where $F_{1xh}(y) = \lim_{w\to\infty} H_{xh}(y, w)$ and $F_{2xh}(y) = \lim_{w\to\infty} H_{xh}(w, y)$ are the conditional empirical marginal distributions. Here and in the sequel, the inverse of a function is understood as its left-continuous generalized inverse.

2.2 Weak convergence

The aim of this subsection is to describe the large-sample behavior of the empirical process $\mathbb{C}_{xh} = \sqrt{nh}(C_{xh} - C_x)$ 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 investigate the asymptotic behaviour of the random function $\mathbb{H}_{xh} = \sqrt{nh}(H_{xh} - H_x)$, where H_{xh} and H_x are defined respectively in Equation (3) and Equation (1). For fixed $(y_1, y_2) \in \mathbb{R}$, the asymptotic normality of $\mathbb{H}_{xh}(y_1, y_2)$ under an α-mixing assumption have been derived in [9]. The following Proposition extends their result to the whole $\mathbb{R}^2$.

Proposition 1 *Suppose that Assumptions $\mathcal{A}_1$–$\mathcal{A}_2$ and $\mathcal{A}_5$ are satisfied, and let f_X denote the density of X. If $n^{13}h^{23} \to \infty$ with $nh^5 \to \kappa^2 < \infty$, then the weighted empirical process $\mathbb{H}_{xh}$ converges weakly in $\ell^\infty(\mathbb{R}^2)$ to a Gaussian limit $\mathbb{H}_x$ such that for $\mathfrak{u}_1 = \int u^2 K(u)\mathrm{d}u$, $\mathfrak{u}_2 = \int K(u)^2 \mathrm{d}u$, $a \wedge b = \min(a,b)$ and $\ddot{H}_w = \partial^2 H_w/\partial w^2$,*

$$\mathrm{E}\left\{\mathbb{H}_x(y_1,y_2)\right\} = \kappa\,\frac{\mathfrak{u}_1}{2}\,\ddot{H}_x(y_1,y_2)$$

and

$$\mathrm{Cov}\left\{\mathbb{H}_x(y_1,y_2), \mathbb{H}_x(y_1',y_2')\right\} = \frac{u_2}{f_X(x)}\sigma_x^2(y_1,y_2,y_1',y_2'),$$

where $\sigma^2_{H_x}(y_1,y_2,y_1',y_2') = H_x(y_1 \wedge y_1', y_2 \wedge y_2') - H_x(y_1,y_2)\,H_x(y_1',y_2')$.

Now the main result of this section can be established.

Proposition 2 *Suppose that the conditions in Proposition 1 are satisfied. Then, if Assumption $\mathcal{A}_3$ holds, the random function $\mathbb{C}_{xh}$ converges weakly in the space $\ell^\infty([0,1]^2)$ to a Gaussian limit $\mathbb{C}_x$ having representation*

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

where $\mathbb{B}_x$ is a Gaussian process on $[0,1]^2$ such that for $\ddot{C}_w = \partial^2 C_w/\partial w^2$:

$$\mathrm{E}\left\{\mathbb{B}_x(u_1,u_2)\right\} = \kappa\,\frac{\mathfrak{u}_1}{2}\ddot{C}_x\left\{F_{1x}^{-1}(u_1), F_{2x}^{-1}(u_2)\right\}$$

and

$$\mathrm{Cov}\left\{\mathbb{B}_x(u_1,u_2), \mathbb{B}_x(u_1',u_2')\right\} = \frac{\mathfrak{u}_2}{f_X(x)}\,\sigma^2_{C_x}(u_1,u_2,u_1',u_2').$$

The limiting representation of $\mathbb{C}_x$ in terms of $\mathbb{B}_x$ stated in Proposition 2 allows to compute the asymptotic bias function $\mathrm{E}\{\mathbb{C}_x(u_1,u_2)\}$, as well as the covariance function $\mathrm{Cov}\{\mathbb{C}_x(u_1,u_2), \mathbb{C}_x(u_1',u_2')\}$. These expressions matches those derived by [14] in the i.i.d case in the special case where the weights in that paper are taken as the local linear ones. In other words, the time-dependency has asymptotically no impact on the limiting distribution of the estimator. This is due to the use of the kernel functions that smooth the covariate space in a shrinking neighborhood of x.

Now sketches of the proofs of Proposition 1 and of Proposition 2 are provided in the next two subsections. Complementary arguments can be found in the Appendix.

2.3 Sketch of the proof of Proposition 1

According for instance to Theorem 1.5.4 of [13], weak convergence in $l^\infty(\mathbb{R}^2)$ is equivalent to finite-dimensional convergence and asymptotic tightness. That the finite-dimensional distributions of $\mathbb{H}_{xh}$ converge to those of $\mathbb{H}_x$ is a consequence of the Cramér-Wold device combined with Theorem 6 of [9]. In order to show the asymptotic tightness of $\mathbb{H}_{xh}$, let

$$\overline{H}_{xh} = (nh)^{-1} \sum_{i=1}^{n} H_{X_i}(y_1, y_2) \mathcal{K}_{xn}\{(X_i - x)/h\},$$

and decompose $\mathbb{H}_{xh} = Z_{xh} + \overline{Z}_{xh}$, where $Z_{xh} = \sqrt{nh}(H_{xh} - \overline{H}_{xh})$ and $\overline{Z}_{xh} = \sqrt{nh}(\overline{H}_{xh} - H_x)$. The asymptotic tightness of $\mathbb{H}_{xh}$ will follow from the tightness of both Z_{xh} and $\overline{Z}_{xh}$. Dealing with $\overline{Z}_{xh}$ first, one uses a Taylor expansion of the function $z \mapsto H_z$ around $z = x$ and the fact that $\sum_{i=1}^{n}(X_i - x)\mathcal{K}_{xn}\{(X_i - x)/h\} = 0$, to deduce that, for some ξ_i between X_i and x:

$$\overline{H}_{xh}(y_1, y_2) - H_x(y_1, y_2) = \frac{1}{2nh} \sum_{i=1}^{n} H_{\xi_i}(y_1, y_2)\, (X_i - x)^2\, \mathcal{K}_{xn}\left(\frac{X_i - x}{h}\right).$$

From Corollary 1 of [8], $S_{n,2}(x) = \mathfrak{u}_2 + o_{\mathrm{a.s}}(1)$, and therefore the dominate convergence theorem entails $\overline{H}_{xh}(y_1, y_2) - H_x(y_1, y_2) = h^2 \ddot{H}_x(y_1, y_2)\mathfrak{u}_2/2 + o_{\mathrm{a.s}}(h^2)$. The asymptotic tightness of $\overline{Z}_{xh}$ is proven, as $\sqrt{nh}h^2 < \infty$.

To show that Z_{xh} is asymptotically tight, consider for a fixed $x \in \mathbb{R}$ the semi-metric $\rho(\mathbf{y}, \mathbf{y}') = |F_{1x}(y_1) - F_{1x}(y_1')| + |F_{2x}(y_2) - F_{2x}(y_2')|$ and define for $\delta > 0$, $f : \mathbb{R}^2 \to \mathbb{R}$ bounded and $T \subseteq \mathbb{R}^2$,

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

The modulus of ρ-continuity 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 according to Theorem 1.5.7 of [13], the process Z_{xn} is asymptotically tight in $\ell^\infty([0,1]^2)$ if and only if for any $\delta > 0$, $\mathfrak{W}_\delta\left(Z_{xn}, \mathbb{R}^2\right)$ converges to zero in probability. To show that it is indeed the case, one proceeds similarly as in Theorem 3 of [1]. Specifically, let $\kappa_\gamma = \lfloor (nh)^{1/2+\gamma} \rfloor$ for some $\gamma \in (0, 1/2)$ and define the rectangles

$$A_\gamma(i, j) = \left[F_{1x}^{-1}\left(\frac{i-1}{\kappa_\gamma}\right), F_{1x}^{-1}\left(\frac{i}{\kappa_\gamma}\right)\right] \times \left[F_{2x}^{-1}\left(\frac{j-1}{\kappa_\gamma}\right), F_{2x}^{-1}\left(\frac{j}{\kappa_\gamma}\right)\right].$$

The collection $A_\gamma(\cdot, \cdot)$ is a partition of $\mathbb{R}^2$ and the ρ-measure of each element is bounded by $2/\kappa_\gamma$. Now for an arbitrary nonempty rectangle $A \in \mathbb{R}^2$, let

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

where $\nu_x(A) = \mathbb{P}\{(Y_{1i}, Y_{2i}) \in A | X_i = x\}$. The definition of the random function $\mathbb{H}_{xh}(A)$ is motivated by the following Lemma whose proof is to be found in the Appendix.

Lemma 1 *Suppose that $\sqrt{n}hh^2 < \infty$ and Assumptions $\mathcal{A}_2$ and $\mathcal{A}_5$ are satisfied. Then, for n sufficiently large, one has for any $\epsilon > 0$ that*

$$\mathbb{P}\left\{\mathfrak{W}_\delta\left(Z_{xn}, \mathbb{R}^2\right) \geq \epsilon\right\} \leq \mathbb{P}\left[\max_{1 \leq i,j \leq \kappa_\gamma} \left|\mathbb{H}_{xh}\left\{A_\gamma(i,j)\right\}\right| \geq \epsilon\right].$$

Now let $\mu_x = \nu_x \otimes \lambda$, where λ is the ρ-measure of A. One then needs to find $\beta > 0$ and $C \in \mathbb{R}$ (that may depend on ϵ and β) such that

$$\mathbb{P}\left[\left|\mathbb{H}_{xh}\{A_\gamma(i,j)\}\right| \geq \epsilon\right] \leq C\left[\mu_x\left\{A_\gamma(i,j)\right\}\right]^{1+\beta}. \tag{4}$$

Such an inequality may be derived from the next lemma whose technical proof is to be found in the Appendix.

Lemma 2 *If Assumptions $\mathcal{A}_1$–$\mathcal{A}_2$ and $\mathcal{A}_5$ are satisfied, one can find a finite constant $\omega > 0$ such that for any rectangle $A \subseteq \mathbb{R}^2$,*

$$\mathrm{E}\left[\left\{\mathbb{H}_{xh}(A)\right\}^6\right] \leq \frac{\omega}{(nh)^3}\left[\mathcal{H}_x(A,n,h) + \mathcal{J}(n,h)\right], \tag{5}$$

where $\mathcal{H}_x(A,n,h) = n\{h\mu_x(A)\}^{1/6} + n^2\{h\mu_x(A)\}^{7/6} + n^3\{h\mu_x(A)\}^{5/2}$ and $\mathcal{J}(n,h) = nh^{1/2} + n^2h^{7/2} + n^3h^{15/2}$.

In view of equations (4) and (5), the Markov inequality entails

$$\mathbb{P}\left(\left|\mathbb{H}_{xh}\{A_\gamma(i,j)\}\right| \geq \epsilon\right) \leq \epsilon^{-6}\,\mathrm{E}\left\{\mathbb{H}_{xh}(A_\gamma(i,j))^6\right\} \leq \epsilon^{-6}\,\omega\,\mu_{x\gamma}^{1+\beta}a_{xh}(\gamma,\beta),$$

where $\mu_{x\gamma} = \mu_x(A_\gamma(i,j))$ and

$$a_{xh}(\gamma,\beta) = \mu_{x\gamma}^{-1-\beta}(nh)^{-3}[\mathcal{H}_x(A_\gamma(i,j),n,h) + \mathcal{J}(n,h)].$$

From the definition of $\mu_{x\gamma}$, one has $(nh)^{-1-2\gamma} \leq \mu_{x\gamma} \leq (nh)^{-1/2-\gamma}$. Therefore, one obtains from direct computations that

$$\mu_{x\gamma}^{-\beta}(nh)^{-3}\mathcal{H}_x(A_\gamma(i,j),n,h) \leq (n^{-11/6}h^{-2})\,n^{4/3(1+\beta)\gamma+4\beta/5} + (n^{-13/12}h^{-23/12})\,n^{2\beta/5+4\gamma(1/6-\beta)/5} + (n^{-1/2})\,n^{2\beta/5+4\gamma(3/2-\beta)/5}$$

and

$$\mu_{x\gamma}^{-\beta}(nh)^{-3}\mathcal{J}(n,h) \leq \left[n^{-1}h^{-3/2} + h^{3/2} + nh^{11/2}\right]n^{4\beta/5+8\gamma(1+\beta)/5}.$$

Since $h \sim n^{-\tau}$ with $1/5 \leq \tau < 13/23$, taking any positive β and γ such that $\beta + 2\gamma < \min\{13/23 - \tau, 1/10\}$ entails $a_{xh}(\gamma,\beta) < 1$. Hence, one can find a constant C (that depends on ϵ, β, γ and τ) such that Equation (4) is satisfied. The asymptotic ρ-equicontinuity follows, for instance, from an extension of Theorem 3 in [1].

2.4 Proof of Proposition 2

Let $\mathbb{D}$ be the space of bivariate distribution functions and define the mapping $\Lambda(H_x)(u_1,u_2) = H_x\{F_{1x}^{-1}(u_1), F_{2x}^{-1}(u_2)\}$. One can then write $\mathbb{C}_{xh} = \sqrt{nh}\{\Lambda(H_{xh}) - \Lambda(H_x)\}$. From [2], one can conclude in view of Assumption $\mathcal{A}_3$ that Λ is Hadamard differentiable with derivative at H_x given for $\widetilde{\Delta}(u_1,u_2) = \Delta\{F_{1x}^{-1}(u_1), F_{2x}^{-1}(u_2)\}$ by

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

From the functional delta method, $\mathbb{C}_{xh}$ converges weakly to

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

where $\mathbb{B}_x(u_1,u_2) = \mathbb{H}_x\left\{F_{1x}^{-1}(u_1), F_{2x}^{-1}(u_2)\right\}$, which completes the proof.

3 Investigation of a second estimator of C_x

3.1 Description of the estimator

As noted by [6], the estimator C_{xh} may be severely biased, especially when the marginal distributions strongly depend on the covariate. For that reason, they proposed a second estimator in order to reduce this effect of the covariate on the margins and hopefully obtain a smaller bias. To this end, define for each $i \in \{1,\ldots,n\}$ the *pseudo-uniformized* observations $(\widetilde{U}_{1i}, \widetilde{U}_{2i}) = (F_{1X_ih_1}(Y_{1i}), F_{2X_ih_2}(Y_{2i}))$, where h_1, h_2 are bandwidth parameters that may differ from h. Then, let

$$G_{xh}(v_1,v_2) = \frac{1}{nh}\sum_{i=1}^{n} \mathcal{K}_{xn}\left(\frac{X_i - x}{h}\right) \mathbb{I}\left\{F_{1X_ih_1}(Y_{1i}) \le v_1, F_{2X_ih_2}(Y_{2i}) \le v_2\right\}$$

and note G_{1xh}, G_{2xh} the marginals of G_{xh}. An estimator of C_x is then

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

3.2 Weak convergence

Let $\mathbb{G}_x$ be the Gaussian process of Proposition 1 when $H_x = C_x$. The weak convergence of $\widetilde{\mathbb{C}}_{xh} = \sqrt{nh}(\widetilde{C}_{xh} - C_x)$ is established next.

Proposition 3 *Suppose Assumptions $\mathcal{A}_1$, $\mathcal{A}_2^\star$, $\mathcal{A}_3$–$\mathcal{A}_5$ are satisfied. If $n^{13}h^{23} \to \infty$, $nh^5 \to \kappa^2 < \infty$, $n\max(h_1^5, h_2^5) < \infty$ and $h/\min(h_1,h_2) < \infty$, then $\widetilde{\mathbb{C}}_{xh}$ converges weakly to a Gaussian limit $\widetilde{\mathbb{C}}_x$ having representation*

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

Like $\mathbb{C}_{xh}$, the limit of $\widetilde{\mathbb{C}}_{xh}$ under strong mixing is the same as that obtained by [14] in the i.i.d. case. In particular, the asymptotic bias and covariance function are the same as those find by these authors.

3.3 Sketch of the 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

$$\widetilde{G}_{xh}(u_1, u_2) = \frac{1}{nh} \sum_{i=1}^{n} \mathcal{K}_{xn} \left(\frac{X_i - x}{h} \right) \mathbb{I}\left(U_i \le u, V_i \le v\right).$$

One can then write for 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). Since the conditional marginal distributions of (U_i, V_i) are uniform on $(0, 1)$, their joint conditional distribution is C_{X_i}. Since Assumptions $\mathcal{A}_1$, $\mathcal{A}_2^\star$, $\mathcal{A}_3$ and $\mathcal{A}_5$ 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 [14], this is closely related to the asymptotic behavior of the processes $\widetilde{Z}_{1xn} = Z_{1xn} - \overline{Z}_{1xn}$ and $\widetilde{Z}_{2xn} = Z_{2xn} - \overline{Z}_{2xn}$, where for $j = 1, 2$ and $z_t = x + t\,C\,h$,

$$Z_{jxn}(t, u) = \sqrt{nh_j}\, F_{jz_t h_j} \left\{ F_{jz_t}^{-1}(u_1) \right\}, \quad \overline{Z}_{jxn}(t, u) = \sqrt{nh_j} \sum_{i=1}^{n} F_{jX_i} \left\{ F_{jz_t}^{-1}(u) \right\}.$$

The key is the following lemma whose proof is deferred to the Supplementary material section.

Lemma 3 *Suppose that Assumptions $\mathcal{A}_1$, $\mathcal{A}_4$ and $\mathcal{A}_5$ are satisfied. Then, as long as $n^{13} h^{23} \to \infty$, $n \max(h_1, h_2)^5 < \infty$ and $h / \min(h_1, h_2) < \infty$, the sequences $\widetilde{Z}_{1xn}$ and $\widetilde{Z}_{2xn}$ are asymptotically tight in $\ell^\infty([-1, 1] \times [0, 1])$.*

Finally, from similar arguments as those in Appendix B.2 of [14], one obtains that $\sqrt{nh}\{\Lambda(G_{xh}) - \Lambda(\widetilde{G}_{xh})\} = o_{\mathbb{P}}(1)$. Hence,

$$\widetilde{\mathbb{C}}_{xh} = \sqrt{nh} \left\{ \Lambda(\widetilde{G}_{xh}) - C_x \right\} + o_{\mathbb{P}}(1),$$

which completes the proof.

4 Sample behavior of the two conditional copula estimators

In order to evaluate the finite sample performance of the estimators C_{xh} and $\widetilde{C}_{xh}$, let $\mathbf{W}_t = (Y_{1t}, Y_{2t}, X_t)$ and consider for some $\theta \in (-1, 1)$ the autoregressive model $\mathbf{W}_t = \theta\, \mathbf{W}_{t-1} + (1 - \theta^2)^{1/2} \boldsymbol{\varepsilon}_t$, where $(\boldsymbol{\varepsilon}_t)_{t \in \mathbb{Z}}$ is a i.i.d. process of innovations from the three-dimensional standard normal distribution with correlation matrix $R = (\rho_{ij})_{i,j=1}^3$. This model entails that $\mathbf{W}_t$ follows a standard Normal with correlation R. Then, the conditional distribution of (Y_{1t}, Y_{2t}) given $X_t = x$ is bivariate Normal with correlation coefficient

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

The conditional copula C_x is that case is therefore the Normal copula with parameter ρ_x; see [11] for more details on this model.

The performance of C_{xh} and $\widetilde{C}_{xh}$ under the above model has been evaluated in the light of the *average integrated squared bias* (AISB) and the *average integrated variance* (AIV) defined by

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

The latter have been estimated from 1 000 replicates under each of the scenario considered; the results are reported in Table 1 for AISB and in Table 2 for AIV. All the simulations have been done using the triweight function $K(y) = 35(1-y^2)^3\,\mathbb{I}(|y| \le 1)/32$.

From Table 1, one notes that $\widetilde{C}_{xh}$ outperforms C_{xh} in terms of AISB when $(\rho_{12}, \rho_{23}, \rho_{13}) = (0.9, 0.8, 0.8)$; an explanation is the fact that $\mathrm{E}(\mathbb{C}_{xh})$ depends in general on F_{1x} and F_{2x}. This explanation is reinforced by the fact that their corresponding AISB are similar under the scenarios where $(\rho_{12}, \rho_{23}, \rho_{13}) \in \{(0.8, 0.1, 0.1), (0.1, 0.1, 0.1)\}$, *i.e.* in cases where the influence on the marginal distributions is rather weak. Also observe that the integrated bias of $\widetilde{C}_{xh}$ stabilises as the bandwidth parameter h takes large values; it is not the case for C_{xh}.

The integrated variance presented in Table 2 is very similar for any values of θ. This is in accordance with the theoretical results that states that the estimators act, asymptotically, as in the i.i.d. case; in the model that was considered, it corresponds to $\theta = 0$. Generally speaking, C_{xh} do slightly better than $\widetilde{C}_{xh}$. Finally note that both AISB and AIV take smaller values when $n = 1\,000$ compared to $n = 250$, as expected.

References

1. Bickel, P.J., Wichura, M.J.: Convergence criteria for multiparameter stochastic processes and some applications. Ann. Math. Statist. **42**, 1656–1670 (1971)
2. Bücher, A., Volgushev, S.: Empirical and sequential empirical copula processes under serial dependence. J. Multivariate Anal. **119**, 61–70 (2013)
3. Carrasco, M., Chen, X.: Mixing and moment properties of various GARCH and stochastic volatility models. Econometric Theory **18**(1), 17–39 (2002)
4. Doukhan, P.: Mixing, *Lecture Notes in Statistics*, vol. 85. Springer-Verlag, New York (1994). Properties and examples

Table 1 Average integrated squared bias ($\times 10^4$) of C_{xh} and $\widetilde{C}_{xh}$, as estimated from 1 000 replicates of a first-order autoregressive Gaussian process. Upper panel: $n = 250$; bottom panel: $n = 1\,000$

		(.9, .8, .8)		(−.9, .8, −.8)		(.8, .1, .1)		(.1, .1, .1)	
θ	h	C_{xh}	$\widetilde{C}_{xh}$	C_{xh}	$\widetilde{C}_{xh}$	C_{xh}	$\widetilde{C}_{xh}$	C_{xh}	$\widetilde{C}_{xh}$
	0.5	0.934	1.359	1.461	1.058	1.192	1.253	0.959	0.974
	1.0	0.125	0.371	1.075	0.291	0.337	0.364	0.313	0.325
0.0	1.5	0.480	0.215	1.668	0.137	0.178	0.202	0.136	0.153
	2.0	1.165	0.154	2.501	0.107	0.111	0.127	0.060	0.069
	2.5	1.852	0.108	3.234	0.088	0.091	0.108	0.062	0.073
	0.5	0.968	1.381	1.418	1.100	1.254	1.334	1.097	1.115
	1.0	0.116	0.370	1.039	0.303	0.344	0.365	0.292	0.303
0.2	1.5	0.447	0.218	1.623	0.140	0.180	0.199	0.134	0.144
	2.0	1.149	0.152	2.459	0.110	0.112	0.132	0.063	0.075
	2.5	1.817	0.105	3.176	0.088	0.093	0.113	0.063	0.080
	0.5	1.044	1.411	1.619	1.263	1.343	1.473	1.336	1.361
	1.0	0.116	0.391	0.996	0.333	0.337	0.377	0.285	0.284
0.4	1.5	0.362	0.233	1.484	0.154	0.192	0.210	0.155	0.158
	2	1.011	0.152	2.261	0.115	0.118	0.138	0.076	0.089
	2.5	1.638	0.108	2.974	0.091	0.098	0.117	0.071	0.086
	0.5	0.026	0.085	0.147	0.061	0.068	0.074	0.046	0.046
	1.0	0.218	0.021	0.443	0.019	0.018	0.021	0.016	0.018
0.0	1.5	0.869	0.012	1.202	0.011	0.010	0.012	0.009	0.013
	2.0	1.646	0.008	2.025	0.008	0.006	0.008	0.004	0.007
	2.5	2.367	0.006	2.749	0.008	0.006	0.008	0.003	0.006
	0.5	0.027	0.082	0.147	0.062	0.074	0.079	0.056	0.058
	1.0	0.206	0.020	0.421	0.02	0.018	0.021	0.015	0.017
0.2	1.5	0.816	0.013	1.135	0.011	0.011	0.013	0.01	0.013
	2.0	1.612	0.008	1.957	0.007	0.006	0.008	0.004	0.006
	2.5	2.323	0.006	2.702	0.007	0.005	0.007	0.003	0.006
	0.5	0.035	0.088	0.153	0.077	0.081	0.086	0.074	0.074
	1.0	0.159	0.021	0.370	0.023	0.019	0.021	0.019	0.021
0.4	1.5	0.683	0.013	0.978	0.012	0.010	0.012	0.013	0.017
	2.0	1.447	0.007	1.786	0.009	0.006	0.008	0.005	0.008
	2.5	2.156	0.005	2.501	0.007	0.005	0.007	0.004	0.008

5. Fan, J., Gijbels, I.: Local polynomial modelling and its applications. Monographs on statistics and applied probability: 66. Boca Raton, Fla. : Chapman & Hall/CRC, 1996. (1996)
6. Gijbels, I., Veraverbeke, N., Omelka, M.: Conditional copulas, association measures and their applications. Comput. Statist. Data Anal. **55**(5), 1919–1932 (2011)
7. Hall, P., Heyde, C.C.: Martingale limit theory and its application. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London (1980). Probability and Mathematical Statistics
8. Masry, E.: Multivariate local polynomial regression for time series: uniform strong consistency and rates. J. Time Ser. Anal. **17**(6), 571–599 (1996)
9. Masry, E., Fan, J.: Local polynomial estimation of regression functions for mixing processes. Scandinavian Journal of Statistics **24**(2), 165–179 (1997)

Table 2 Average integrated variance ($\times 10^4$) of C_{xh} and $\widetilde{C}_{xh}$, as estimated from 1 000 replicates of a first-order autoregressive Gaussian process. Upper panel: $n = 250$; bottom panel: $n = 1\ 000$

		(.9, .8, .8)		(−.9, .8, −.8)		(.8, .1, .1)		(.1, .1, .1)	
θ	h	C_{xh}	$\widetilde{C}_{xh}$	C_{xh}	$\widetilde{C}_{xh}$	C_{xh}	$\widetilde{C}_{xh}$	C_{xh}	$\widetilde{C}_{xh}$
	0.5	3.589	3.778	3.602	3.863	2.924	2.944	6.379	6.440
	1.0	1.649	1.882	1.636	1.894	1.412	1.428	3.287	3.300
0.0	1.5	1.052	1.337	1.087	1.382	0.992	0.995	2.466	2.459
	2.0	0.828	1.173	0.822	1.147	0.842	0.851	1.963	1.987
	2.5	0.671	1.016	0.688	1.054	0.745	0.752	1.850	1.845
	0.5	3.732	3.926	3.742	3.957	2.977	3.031	6.547	6.642
	1.0	1.716	1.967	1.735	1.968	1.473	1.477	3.407	3.394
0.2	1.5	1.115	1.360	1.094	1.401	1.011	1.014	2.495	2.482
	2.0	0.832	1.171	0.832	1.160	0.849	0.853	2.003	2.028
	2.5	0.673	1.026	0.698	1.060	0.757	0.769	1.869	1.882
	0.5	4.069	4.260	4.057	4.196	3.148	3.220	7.204	7.247
	1.0	1.851	2.061	1.853	2.059	1.528	1.544	3.653	3.667
0.4	1.5	1.210	1.484	1.198	1.487	1.077	1.085	2.749	2.755
	2.0	0.900	1.257	0.880	1.219	0.889	0.894	2.214	2.213
	2.5	0.712	1.084	0.741	1.106	0.796	0.810	2.012	2.020
	0.5	0.793	0.825	0.792	0.835	0.617	0.618	1.524	1.532
	1.0	0.385	0.434	0.389	0.444	0.316	0.316	0.799	0.796
0.0	1.5	0.257	0.324	0.246	0.315	0.233	0.233	0.588	0.592
	2.0	0.197	0.272	0.192	0.265	0.198	0.198	0.502	0.503
	2.5	0.162	0.255	0.164	0.249	0.183	0.185	0.453	0.455
	0.5	0.803	0.837	0.800	0.842	0.624	0.628	1.553	1.546
	1.0	0.397	0.443	0.408	0.466	0.319	0.320	0.787	0.789
0.2	1.5	0.263	0.332	0.253	0.325	0.232	0.234	0.598	0.602
	2.0	0.197	0.274	0.194	0.269	0.198	0.201	0.509	0.511
	2.5	0.168	0.256	0.164	0.253	0.185	0.188	0.469	0.470
	0.5	0.874	0.909	0.872	0.903	0.674	0.679	1.639	1.651
	1.0	0.432	0.478	0.437	0.499	0.338	0.337	0.842	0.844
0.4	1.5	0.283	0.357	0.278	0.354	0.245	0.246	0.641	0.644
	2.0	0.206	0.290	0.199	0.283	0.204	0.206	0.548	0.548
	2.5	0.178	0.269	0.175	0.269	0.196	0.197	0.512	0.515

10. Meitz, M., Saikkonen, P.: Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models. Econometric Theory **24**(5), 1291–1320 (2008)
11. Nelsen, R.B.: An introduction to copulas, second edn. Springer Series in Statistics. Springer, New York (2006)
12. Patton, A.J.: Modelling asymmetric exchange rate dependence. Internat. Econom. Rev. **47**(2), 527–556 (2006). DOI 10.1111/j.1468-2354.2006.00387.x. URL http://dx.doi.org/10.1111/j.1468-2354.2006.00387.x
13. van der Vaart, A.W., Wellner, J.A.: Weak convergence and empirical processes. Springer Series in Statistics. Springer-Verlag, New York (1996). With applications to statistics

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

A Assumptions needed in Proposition 1, Proposition 2 and Proposition 3

$\mathcal{A}_1$. The α-mixing coefficients of $\{(Y_{1t}, Y_{2t}, X_t)\}_{t\in\mathbb{Z}}$ are such that $\alpha(r) = O(r^{-a})$ for some $a > 6$.

$\mathcal{A}_2$. The functions $(w, y_1, y_2) \mapsto H_w(y_1, y_2)$, $\dot{H}_w = \partial H_w/\partial w$ and $\ddot{H}_w = \partial^2 H_w/\partial w^2$ exist and are uniformly continuous in $(w, y_1, y_2) \in J_x \times \mathbb{R}^2$, where J_x is an open neighborhood of x.

$\mathcal{A}_2^\star$. The functions $(w, u_1, u_2) \mapsto C_w(u_1, u_2)$, $\dot{C}_w = \partial C_w/\partial w$ and $\ddot{C}_w = \partial^2 C_w/\partial w^2$ exist and are uniformly continuous in $(w, u_1, u_2) \in J_x \times [0, 1]^2$, where J_x is an open neighborhood of x.

$\mathcal{A}_3$. The partial derivatives $C_x^{[1]}(u_1, u_2) = \partial\, C_x(u_1, u_2)/\partial u$ and $C_x^{[2]}(u_1, u_2) = \partial\, C_x(u_1, u_2)/\partial v$ exist and are continuous respectively on $(0, 1) \times [0, 1]$ and $[0, 1] \times (0, 1)$.

$\mathcal{A}_4$. For $j = 1, 2$, the functions $(w, u) \mapsto F_{jw}\{F_{jw}^{-1}(u_1)\}$, $\dot{F}_{jw}\{F_{jw}^{-1}(u_2)\}$ and $\ddot{F}_{jw}\{F_{jw}^{-1}(u_1)\}$ exist and are continuous in $(w, u) \in J_x \times [0, 1]$, where J_x is an open neighborhood of x.

$\mathcal{A}_5$. The density f_X of X exist and is continuous in an open neighbourhood of x. Further, there exist a constant $M > 0$ such that $f_{0,\ell}(u, v) \le M$ for all u, v in a neighborhood of x, where $f_{0,\ell}$ is the density of the random vector (X_0, X_ℓ).

B Proofs

B.1 Proof of Lemma 1

First, define the product space $T_\gamma = T_\gamma^{(1)} \times T_\gamma^{(2)}$, where for $j = 1, 2$,

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

For $y \in \mathbb{R}$, define $\underline{y}_\gamma^{(1)} = \max\{\zeta \in I_\gamma^{(1)} : \zeta \le y\}$, $\overline{y}_\gamma^{(1)} = \min\{\zeta \in I_\gamma^{(1)} : \zeta \le y\}$, $\underline{y}_\gamma^{(2)} = \max\{\zeta \in I_\gamma^{(2)} : \zeta \le y\}$ and $\overline{y}_\gamma^{(2)} = \min\{\zeta \in I_\gamma^{(2)} : \zeta \le y\}$. With this notation,

$$F_{1x}\left(\overline{y}_\gamma^{(1)}\right) - F_{1x}\left(\underline{y}_\gamma^{(1)}\right) \le \frac{1}{\kappa_\gamma} \quad \text{and} \quad F_{2x}\left(\overline{z}_\gamma^{(2)}\right) - F_{2x}\left(\underline{z}_\gamma^{(2)}\right) \le \frac{1}{\kappa_\gamma}.$$

Now observe that for any $\omega = (y, z) \in \mathbb{R}^2$, one has for $\underline{\omega}_\gamma = (\underline{y}_\gamma^{(1)}, \underline{z}_\gamma^{(2)})$ and $\overline{\omega}_\gamma = (\overline{y}_\gamma^{(1)}, \overline{z}_\gamma^{(2)})$, that

$$\begin{aligned} Z_{xn}(\omega) - Z_{xn}(\underline{\omega}_\gamma) &\le Z_{xn}(\overline{\omega}_\gamma) - Z_{xn}(\underline{\omega}_\gamma) \\ &\quad + \frac{2}{\sqrt{nh}} \sum_{i=1}^n \mathrm{E}\left[\mathcal{K}_{xn}\left(\frac{X_i - x}{h}\right) \left\{ H_{X_i}(\overline{\omega}_\gamma) - H_{X_i}(\underline{\omega}_\gamma) \right\} \right]. \end{aligned}$$

It will next be demonstrate that the second term of the left hand side of the previous display is asymptotically negligible. Since Assumption $\mathcal{A}_2$ holds, and because

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

a Taylor expansion allows to write

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\mathcal{K}_{xn}\left(\frac{X_i-x}{h}\right)\left\{H_{X_i}(\overline{\omega}_\gamma)-H_{X_i}(\underline{\omega}_\gamma)\right\}=\sqrt{nh}\left\{H_x(\overline{\omega}_\gamma)-H_x(\underline{\omega}_\gamma)\right\}$$
$$+\ \frac{1}{2\sqrt{nh}}\sum_{i=1}^{n}\mathcal{K}_{xn}\left(\frac{X_i-x}{h}\right)(X_i-x)^2\left\{\ddot{H}_{\zeta_i}(\overline{\omega}_\gamma)-\ddot{H}_{\zeta_i}(\underline{\omega}_\gamma)\right\}, \qquad (6)$$

where ζ_i lies between X_i and x. Now for any bivariate distribution function H with marginal distributions F_1 and F_2, one has for $\omega_1=(y_1,z_1)$ and $\omega_2=(y_2,z_2)$ that $|H(\omega_1)-H(\omega_2)|\leq|F_1(y_1)-F_1(y_2)|+|F_2(z_1)-F_2(z_2)|$. Hence,

$$\sqrt{nh}\left|H_x(\overline{\omega}_\gamma)-H_x(\underline{\omega}_\gamma)\right|\leq\sqrt{nh}\,\rho(\overline{\omega}_\gamma,\underline{\omega}_\gamma).$$

The negligibility of the second term of Equation (6) follows from Theorem 1 of [9] together with Assumption $\mathcal{A}_2$ and the fact that $\sqrt{nh}h^2\to\kappa<\infty$. As a consequence, and from the definition of $\overline{\omega}_\gamma$ and $\underline{\omega}_\gamma$, uniformly in $\omega\in\mathbb{R}^2$:

$$Z_{xn}(\omega)-Z_{xn}(\underline{\omega}_\gamma)\leq Z_{xn}(\overline{\omega}_\gamma)-Z_{xn}(\underline{\omega}_\gamma)+o(1).$$

From similar arguments, one deduces that

$$Z_{xn}(\underline{\omega}_\gamma)-Z_{xn}(\omega)\leq Z_{xn}(\overline{\omega}_\gamma)-Z_{xn}(\underline{\omega}_\gamma)+o(1).$$

Thus, for any $\omega_1,\omega_2\in\mathbb{R}^2$,

$$|Z_{xn}(\omega_1)-Z_{xn}(\omega_2)|\leq\left|Z_{xn}(\overline{\omega_1}_\gamma)-Z_{xn}(\underline{\omega_1}_\gamma)\right|+\left|Z_{xn}(\overline{\omega_2}_\gamma)-Z_{xn}(\underline{\omega_2}_\gamma)\right|$$
$$+\left|Z_{xn}(\underline{\omega_1}_\gamma)-Z_{xn}(\underline{\omega_2}_\gamma)\right|.$$

Since for n sufficiently large, $\rho(\omega_1,\omega_2)<\delta$ entails $\rho(\underline{\omega_1}_\gamma,\underline{\omega_2}_\gamma)<2\delta$, it follows that

$$\mathfrak{W}_\delta(Z_{xn},\mathbb{R}^2)\leq 3\,\mathfrak{W}_{2\delta}(Z_{xn},T_\gamma).$$

It remains to show that for any positive sequence δ_n that decreases to zero as $n\to\infty$ and for any $\epsilon>0$, $\mathbb{P}\left(\mathfrak{W}_{\delta_n}(Z_{xn},T_\gamma)>\epsilon\right)$ tends to zero. To this end, observe that $\mathfrak{W}_{\delta_n}(Z_{xn},T_\gamma)=0$ whenever $\delta_n<2\kappa_\gamma^{-1}$, while $\mathfrak{W}_{\delta_n}(Z_{xn},T_\gamma)\leq\mathfrak{W}_{2\kappa_\gamma}(Z_{xn},T_\gamma)$ otherwise. One can then conclude that

$$\mathbb{P}\left(\mathfrak{W}_{\delta_n}(Z_{xn},T_\gamma)\geq\epsilon\right)\leq\mathbb{P}\left(\max_{1\leq i,j\leq\kappa_\gamma}|\mathbb{H}_{xh}\{A_\gamma(i,j)\}|\geq\epsilon\right).$$

B.2 Proof of Lemma 2

Hereafter, let ϑ_i stand for $\mathbb{I}\{(Y_{1i},Y_{2i})\in A\}-\nu_{X_i}(A)$ and ν_z for $\nu_z(A)$. Before getting to the very heart of the matter, few technical arguments are required. First, from the definition of $\mathcal{K}_{xn}(\cdot)$, one writes,

$$\mathbb{H}_{xh}(A)=(nh)^{-1/2}f_X(x)^{-1}\sum_{i=1}^{n}\vartheta_i\,K\left(\frac{X_i-x}{h}\right)\left[1+R_{xn}\left(\frac{X_i-x}{h}\right)\right],$$

where $R_{xn}(z)=f_X(x)(S_{n,2}(x)-zS_{n,1}(x))/(S_{n,0}(x)S_{n,2}(x)-S_{n,1}(x)^2)-1$. Since $K(\cdot)$ is compactly supported, it can be shown using Assumption $\mathcal{A}_1$ and $\mathcal{A}_5$ together with

Corollary 1 in [8] that there exist a constant $\mathcal{R}(x) > 0$ such that with probability one, $R_{xn}(z) \leq \mathcal{R}(x)$. Hence, it suffices to show that the result is true for

$$\widetilde{\mathbb{H}}_{xh}(A) = (nh)^{-1/2} \sum_{i=1}^{n} \nu_i \, K_{xh}(i),$$

where $K_{xh}(i) = K((X_i - x)/h)$. Second, as condition $\mathcal{A}_2$ holds, for any $i \in \{j : K_{xh}(j) \neq 0\}$,

$$\nu_{X_i}(A) = \nu_x(A) + \dot{\nu}_x(A)(X_i - x) + \frac{1}{2}\ddot{\nu}_{z_i}(A)(X_i - x)^2 \tag{7}$$

where z_i is between X_i and x. Then, from Equation (7), for any $\delta \in (1, 6]$, any integer k and $(i_1, \ldots, i_k) \in \{1, \ldots, n\}^k$,

$$\begin{aligned} \mathrm{E}\left[\left|\prod_{i_j=1}^{k} \vartheta_{i_j} K_{xh}(i_j)\right|^{\delta}\right] &\leq \|K\|_\infty^{k\delta}\, \mathrm{E}\left[\nu_{X_1}(A)\, K_{xh}(1)\right] \\ &= \nu_x \mathrm{E}\left\{K_{xh}(1)\right\} + \dot{\nu}_x \mathrm{E}\left\{(X_1 - x) K_{xh}(1)\ \right\} + \frac{1}{2} E\left\{(X_1 - x)^2 \ddot{\nu}_{z_i} K_{xh}(1)\right\}. \end{aligned}$$

As $K(\cdot)$ is symmetric, and from the dominate convergence, it follows that

$$\mathrm{E}\left[\left|\textstyle\prod_{i_j=1}^{k} \vartheta_{i_j} K_{xh}(i_j)\right|^{\delta}\right] \leq h\nu_{xh}, \tag{8}$$

where $\nu_{xh} = \nu_x + h^2 \mathfrak{u}_2 \ddot{\nu_x}/2$.

As a starting point of the proof of the main result, notice that by stationarity, $\mathrm{E}|\widetilde{\mathbb{H}}_{xh}(A)|^6 \leq 6!(nh)^{-3} \sum_{j=1}^{6} |\mathrm{E}\{T_{hx}^{(j)}\}|$, with

$$T_{hx}^{(j)} = \sum_{1=i_1 \leq \ldots \leq i_j}^{n} \prod_{k=1}^{j} \vartheta_{i_k} K_{xh}(i_k).$$

The goal is now to bound each $T_{hx}^{(j)}$, for $j = 1, \ldots, 6$. That $\overline{T}_{hx}^{(1)} = 0$ follows from $\mathrm{E}(\vartheta_i \mid X_i) = 0$. To deal with the case $j = 2$, one recalls a Lemma due to Davydov which states that any stationary strongly mixing sequence $\{Z_i\}_{i \in \mathbb{Z}}$ satisfies

$$|\mathrm{Cov}(Z_1, Z_{1+\ell})| \leq 8\alpha(\ell)^{1-2/\delta} \left[\mathrm{E}|Z_1|^{\delta}\right]^{2/\delta} \tag{9}$$

(see [7] corol. A2). Therefore, upon setting $\delta = 12/5$, one obtains from Equation (8) and the fact that $\alpha(\ell) \sim \ell^{-a}$ that

$$\begin{aligned} \left|\mathrm{E}\left(T_{hx}^{(2)}\right)\right| &\leq n \sum_{\ell=0}^{n-1} |\mathrm{Cov}\left\{\vartheta_i K_{xh}(i), \vartheta_{i+\ell} K_{xh}(i+\ell)\right\}| \\ &\leq n \sum_{\ell=1}^{n-1} \alpha(\ell)^{1-2/\delta} \left[\mathrm{E}|\vartheta_1 K_{xh}(1)|^{\delta}\right]^{2/\delta} \leq n(h\nu_{xh})^{5/6} \sum_{\ell=1}^{n-1} \alpha(\ell)^{-\frac{a}{6}}. \end{aligned}$$

Since $a > 6$, the sum remains finite as n runs to infinity. Therefore,

$$\left|\mathrm{E}\left(T_{hx}^{(2)}\right)\right| \leq \omega_2\, nh^{5/6} \nu_{xh}^{5/6}. \tag{10}$$

for some $\omega_2 > 0$. For $j = 3$, denote $\mathcal{T}_1 = \{i_1 \leq i_2 \leq i_3 : g_2 \leq g_1\}$ and $\mathcal{T}_2 = \{i_1 \leq i_2 \leq i_3 : g_2 > g_1\}$, where $g_k = i_{k+1} - i_k$ is the gap between two consecutive indices. One then writes $\mathrm{E}(T_{hx}^{(3)}) = W_{hx}^{(1)} + W_{hx}^{(2)}$ with

$$W_{hx}^{(k)} = \sum_{\mathcal{T}_k} \mathrm{E}\left\{\prod_{j=1}^{3} \vartheta_{i_j} K_{xh}(i_j)\right\} = \sum_{\mathcal{T}_k} \mathrm{Cov}\left\{\prod_{j=1}^{k} \vartheta_{i_j} K_{xh}(i_j), \prod_{j=k+1}^{3} \vartheta_{i_j} K_{xh}(i_j)\right\}.$$

Using Davydov's inequality in Equation (9) with $\delta = 3$ in conjunction with Equation (8) together with the fact that $\alpha(\ell) \sim \ell^{-a}$ yields

$$\left|W_{hx}^{(k)}\right| \le n(h\nu_{xh})^{2/3} \sum_{g_k=0}^{n-1} (g_k+1)\alpha(g_k)^{1/3} \le n(h\nu_{xh})^{2/3} \sum_{g_k=1}^{n-1} g_k^{-(a-3)/3}.$$

The fact that $a > 6$ entails $\sum_{g_k=1}^{\infty} g_k^{-(a-3)/3} < \infty$ and therefore

$$\left|\mathrm{E}\left(T_{hx}^{(3)}\right)\right| \le \omega_3\, n(h\nu_{xh})^{2/3}. \tag{11}$$

The case $j = 4$ calls for a special treatment. Let $\mathcal{U}_k = \{i_1 \le \ldots \le i_4 : g_i \le g_k\}$, where once again $g_k = i_{k+1} - i_k$ is the gap between two consecutive indices, and assume wlog that the $\mathcal{U}_k$'s a disjoint (this can be done formally by removing the redundant equalities). Letting

$$U_{hx} = \sum_{\mathcal{U}_2} \mathrm{E}\left\{\prod_{j=1}^{2} \vartheta_{i_j} K_{xh}(i_j)\right\} \mathrm{E}\left\{\prod_{j=3}^{4} \vartheta_{i_j} K_{xh}(i_j)\right\},$$

and since by construction $T_{hx}^{(4)} = \sum_{k=1}^{3} \sum_{\mathcal{U}_k} \prod_{j=1}^{4} \vartheta_{i_j} K_{xh}(i_j)$, notice that

$$\left|\mathrm{E}\left(T_{hx}^{(4)}\right) - U_{hx}\right| \le \sum_{k=1}^{3} \sum_{\mathcal{U}_k} \left|\mathrm{Cov}\left\{\prod_{j=1}^{k} \vartheta_{i_j} K_{xh}(i_j), \prod_{j=k+1}^{4} \vartheta_{i_j} K_{xh}(i_j)\right\}\right|.$$

Equation (9) with $\delta = 6$ and Equation (8) then yields

$$\left|\mathrm{E}\left(T_{hx}^{(4)}\right) - U_{hx}\right| \le 3n(h\nu_{xh})^{1/3} \sum_{\ell=0}^{n-1} (\ell+1)^2 \alpha(\ell)^{2/3} \le \omega_4 n(h\nu_{xh})^{2/3},$$

where the last inequality follows from Assumption $\mathcal{A}_1$. As $|U_{hx}| \le [\mathrm{E}(T_{hx}^{(2)})]^2$, in follows from Equation (10) and the previous display that

$$\left|\mathrm{E}\left(T_{hx}^{(4)}\right)\right| \le [\omega_4 + \omega_2^2][nh^{1/3}\nu_{xh}^{1/3} + n^2 h^{5/3} \nu_{xh}^{5/3}]. \tag{12}$$

Next, let's jump directly to the case $j = 6$ as the case $j = 5$ is similar. Let $\mathcal{V}_k = \{i_1 \le \ldots \le i_6 : g_i \le g_k\}$ and write, for $k = 1, \ldots, 5$,

$$V_{hx}^{(k)} = \sum_{\mathcal{V}_k} \mathrm{E}\left\{\prod_{j=1}^{k} \vartheta_{i_j} K_{xh}(i_j)\right\} \mathrm{E}\left\{\prod_{j=k+1}^{6} \vartheta_{i_j} K_{xh}(i_j)\right\}.$$

Directly, $k \in \{1, 5\}$ entails $V_{hx}^{(k)} = 0$. Otherwise, $k \in \{2, 4\}$ implies from Equation (10) and Equation (12) that

$$\left|V_{hx}^{(k)}\right| \le \left|\mathrm{E}\left(T_{xh}^{(2)}\right)\right| \left|\mathrm{E}\left(T_{xh}^{(4)}\right)\right| \le n^2 h^{7/6} \nu_{xh}^{7/6} + n^3 h^{5/2} \nu_{xh}^{5/2}.$$

Similarly, from Equation (11),

$$\left|V_{hx}^{(3)}\right| \le \left|\mathrm{E}\left(T_{xh}^{(3)}\right)\right|^2 \le n^2 h^{4/3} \nu_{xh}^{2/3}.$$

Finally,

$$\begin{aligned}\left|\mathrm{E}\left(T_{hx}^{(6)}\right) - \sum_{k=2}^{4} V_{hx}^{(k)}\right| &\le \sum_{k=1}^{3} \sum_{\mathcal{V}_k} \left|\mathrm{Cov}\left\{\prod_{j=1}^{k} \vartheta_{i_j} K_{xh}(i_j), \prod_{j=k+1}^{6} \vartheta_{i_j} K_{xh}(i_j)\right\}\right| \\ &\le 5n(h\nu_{xh})^{1/6} \sum_{\ell=0}^{n-1} (\ell+1)^4 \alpha(\ell)^{5/6} \le \omega_6 n(h\nu_{xh})^{1/6},\end{aligned}$$

The last line is obtained from Davydov's inequality with $\delta = 12$ and the fact that $\sum_{\ell=0}^{n-1}(\ell+1)^4\alpha(\ell)^{5/6} \sim \sum_{\ell=1}^{\infty}\ell^{(24-5a)/6} < \omega_6$ as $a > 6$. As a consequence, and since for sufficiently small h, $h\nu_{xh} < 1$, one can find a finite constant $\omega_6' > 0$ such that

$$\left|\mathrm{E}\left(T_{hx}^{(6)}\right)\right| \le \omega_6'\left[nh^{1/6}\nu_{xh}^{1/6} + n^2h^{7/6}\nu_{xh}^{7/6} + n^3h^{5/2}\nu_{xh}^{5/2}\right].$$

Collecting the bounds for the $T_{hx}^{(k)}$'s and noting that $\nu_x \le \mu_x(A)$ proves that

$$\mathrm{E}\left[\{\mathbb{H}_{xh}(A)\}^6\right] \le \frac{\overline{\omega}}{(nh)^3}\left[n\{h\mu_{xh}(A)\}^{1/6} + n^2\{h\mu_{xh}(A)\}^{7/6} + n^3\{h\mu_{xh}(A)\}^{5/2}\right] \tag{13}$$

for some finite constant $\overline{\omega} > 0$, where $\mu_{xh}(A) = \mu_x(A) + h^2\,\mathfrak{u}_2/2\ddot{\nu}_x(A)$. Because of Assumption $\mathcal{A}_2$, $\ddot{\nu}_x$ is uniformly bounded for any rectangle $A \in \mathbb{R}^2$. Therefore, one can find a constant $\upsilon > 0$ such that for any $b \in \mathbb{R}^+$, $\mu_{xh}(A)^b \le (2\upsilon)^b \max\{\mu_x(A), h^2\}^b \le (2\upsilon)^b[\mu_x(A)^b + (h^2)^b]$. Plugging this inequality into Equation (13) completes the proof.

B.3 Proof of Lemma 3

In order to ease readability, we simply write g for h_j. Moreover, as the cases $j = 1$ and $j = 2$ are identical, we drop the index j throughout the section. For any fixed $(t, u) \in \mathcal{I} = [-1, 1] \times [0, 1]$ the asymptotic normality of the random variable $\widetilde{Z}_{xn}(t, u)$ follows from Theorem 6 in [9]. This implies the asymptotic tightness of the random variable $\widetilde{Z}_{xn}(t, u)$ in $\mathbb{R}$. It remains to show the asymptotic tightness of the sequence $\widetilde{Z}_{xn}$ in $\ell^\infty(\mathcal{I})$. To this end, notice that, from the proof of Lemma 2.4, $\widetilde{Z}_{xn}(t, u)$ can be written as

$$\widetilde{Z}_{xn}(t,u) = \frac{f_X(z_t)^{-1}}{\sqrt{ng}}\sum_{i=1}^n\left[\mathbb{I}\left\{Y_i \le F_{z_t}^{-1}(u)\right\} - F_{X_i}\left\{F_{z_t}^{-1}(u)\right\}\right]K\left(\frac{X_i - z_t}{g}\right) \times\left[1 + R_{zn}\left(\frac{X_i - z_t}{g}\right)\right].$$

Since Corollary 1 in [8] implies $R_{zn} = o_{\mathrm{a.s}}(1)$ uniformly in $z \in I_{nx}$, and because $f_X(z)$ is bounded away from 0 for any z in a neighbourhood or x, it follows that $\widetilde{Z}_{xn}$ is asymptotically tight if and only if the same holds for sequence $\widetilde{W}_{xn}(t,u) := \sqrt{ng}[\mathcal{E}_{z_t g}\{F_{z_t}^{-1}(u)\} - \overline{\mathcal{E}}_{z_t g}\{F_{z_t}^{-1}(u)\}]$, where

$$\mathcal{E}_{zg}(y) = \frac{1}{ng}\sum_{i=1}^n \mathbb{I}(Y_i \le y)K\left(\frac{X_i - z}{g}\right),\ \overline{\mathcal{E}}_{zg}(y) = \frac{1}{ng}\sum_{i=1}^n F_{X_i}(y)K\left(\frac{X_i - z}{g}\right).$$

Now let $\rho(t, u, t', u') = |t - t'| + |u - u'|$. For any bounded function $f : \mathcal{I} \to \mathbb{R}$ and subset T of $\mathcal{I}$, define

$$\mathfrak{W}_\delta(f, T) = \sup_{\substack{(t,u),(t',u') \in T \\ \rho(t,u,t',u') < \delta}} |f(t, u) - f(t', u')|.$$

It will be shown below that $\widetilde{W}_{xn}$ is asymptotically ρ-equicontinuous i.e for any $\epsilon > 0$ there is a $\delta > 0$ such that

$$\lim_{n\to\infty}\mathbb{P}\left(\mathfrak{W}_\delta(\widetilde{W}_{xn}, \mathcal{I}) > \epsilon\right) = 0.$$

For $\kappa_\gamma = \lfloor (ng)^{1/2+\gamma}\rfloor$, define grids $I_\gamma = \{0, \frac{1}{\kappa_\gamma}, \ldots, \frac{\kappa_\gamma - 1}{\kappa_\gamma}, 1\}$ and $J_\gamma = \{0, \pm\frac{1}{\kappa_\gamma}, \ldots, \pm\frac{\kappa_\gamma - 1}{\kappa_\gamma}, \pm 1\}$, where $\gamma \in (0, 1/2)$ is a grid parameter to be fixed later, and set $T_\gamma = J_\gamma \times I_\gamma$. For any $(t,u) \in [-1,1] \times [0,1]$, define $(\underline{t}_\gamma, \underline{u}_\gamma)$ and $(\bar{t}_\gamma, \overline{u}_\gamma)$ as in Section B.1. Analogously to that section observe that

$$\widetilde{W}_{xn}(t,u) - \widetilde{W}_{xn}(\underline{t}_\gamma, \underline{u}_\gamma) \le \{\widetilde{W}_{xn}(t, \overline{u}_\gamma) - \widetilde{W}_{xn}(t, \underline{u}_\gamma)\}+$$
$$\{\widetilde{W}_{xn}(t, \underline{u}_\gamma) - \widetilde{W}_{xn}(\underline{t}_\gamma, \underline{u}_\gamma)\} + \sqrt{ng}\left[\overline{\mathcal{E}}_{z_t g}\left\{F_{z_t}^{-1}(\overline{u}_\gamma)\right\} - \overline{\mathcal{E}}_{z_t g}\left\{F_{z_t}^{-1}(\underline{u}_\gamma)\right\}\right]. \qquad (14)$$

Starting with the last summand of equation (14), using a Taylor expansion of F_{X_i} around z_t leads to:

$$\sqrt{ng}\left[\overline{\mathcal{E}}_{z_t g}\left\{F_{z_t}^{-1}(\overline{u}_\gamma)\right\} - \overline{\mathcal{E}}_{z_t g}\left\{F_{z_t}^{-1}(\underline{u}_\gamma)\right\}\right]$$
$$= \sqrt{ng}(\overline{u}_\gamma - \underline{u}_\gamma) + \sqrt{ng}\left[\dot{F}_{z_t}\left\{F_{z_t}^{-1}(\overline{u}_\gamma)\right\} - \dot{F}_{z_t}\left\{F_{z_t}^{-1}(\overline{u}_\gamma)\right\}\right] g\, S_{n,1}(z_t) +$$
$$\frac{\sqrt{ng}g^2}{2}\left(\frac{1}{ng}\sum_{i=1}^n \left[\ddot{F}_{r_{ti}}\{F_{z_t}^{-1}(\overline{u}_\gamma)\} - \ddot{F}_{r_{ti}}\{F_{z_t}^{-1}(\underline{u}_\gamma)\}\right] K\left(\frac{X_i - z_t}{g}\right)\left(\frac{X_i - z_t}{g}\right)^2\right),$$

where r_{ti} lies between z_t and X_i. From Corollary 1 in [8], $S_{n,1}(z) = O_{\mathrm{a.s}}(g + (ng)^{-1/2}\log(n)\}$ and $S_{n,2}(z) = O_{\mathrm{a.s}}(1)$ uniformly in $z \in V(x)$. Therefore, using Assumption $\mathcal{A}_4$ and invoking the dominate convergence theorem, the previous equation is equal to

$$\sqrt{ng}(\overline{u}_\gamma - \underline{u}_\gamma) + o(1)O_{\mathrm{a.s}}(g\log(n)) + o(1)O_{\mathrm{a.s}}(\sqrt{ng}g^2) = o_{\mathrm{a.s}}(1).$$

The last equality follows from the assumptions over the bandwidth parameters, ensuring that $\sqrt{ng}g^2 < \infty$ and $g\log(n) \to 0$, and the fact that the grid definition entails $\sqrt{ng}(\overline{u}_\gamma - \underline{u}_\gamma) = O\{(ng)^{-\gamma}\}$. This yields the negligibility of $\sqrt{ng}\left[\overline{\mathcal{E}}_{z_t g}\left\{F_{z_t}^{-1}(\overline{u}_\gamma)\right\} - \overline{\mathcal{E}}_{z_t g}\left\{F_{z_t}^{-1}(\underline{u}_\gamma)\right\}\right]$.

Next we deal with the term $\widetilde{W}_{xn}(t, \underline{u}_\gamma) - \widetilde{W}_{xn}(\underline{t}_\gamma, \underline{u}_\gamma)$ in equation (14). Let $\widetilde{\mathcal{E}}_{zg} = \mathcal{E}_{zg} - \overline{\mathcal{E}}_{zg}$. In this notation, one writes

$$\widetilde{W}_{xn}(t, \underline{u}_\gamma) - \widetilde{W}_{xn}(\underline{t}_\gamma, \underline{u}_\gamma) = \sqrt{ng}\left[\widetilde{\mathcal{E}}_{z_t g}\left\{F_{z_t}^{-1}(\underline{u}_\gamma)\right\} - \widetilde{\mathcal{E}}_{z_t g}\left\{F_{z_{\underline{t}_\gamma}}^{-1}(\underline{u}_\gamma)\right\}\right]$$
$$+ \sqrt{ng}\left[\widetilde{\mathcal{E}}_{z_t g}\left\{F_{z_{\underline{t}_\gamma}}^{-1}(\underline{u}_\gamma)\right\} - \widetilde{\mathcal{E}}_{z_{\underline{t}_\gamma} g}\left\{F_{z_{\underline{t}_\gamma}}^{-1}(\underline{u}_\gamma)\right\}\right]. \qquad (15)$$

In what follows, it will be shown that $\sqrt{ng}(\widetilde{\mathcal{E}}_{z_t g} - \widetilde{\mathcal{E}}_{z_{\underline{t}_\gamma} g}) = o_{\mathrm{a.s}}(1)$, which will imply the negligibility of the second term of Equation (15). Since $z_t - z_{\underline{t}_\gamma} = Ch(t - \underline{t}_\gamma)$, for any $y \in \mathbb{R}$, one has, for $K^{(1)}(u) = \partial K(u)/\partial u$ that

$$\sqrt{ng}|\widetilde{\mathcal{E}}_{z_t g}(y) - \widetilde{\mathcal{E}}_{z_{\underline{t}_\gamma} g}(y)| \le \frac{1}{\sqrt{ng}}\left|\sum_{i=1}^n K\left(\frac{X_i - z_t}{g}\right) - K\left(\frac{X_i - z_{\underline{t}_\gamma}}{g}\right)\right|$$
$$\le \left\{\sup_{z \in I_x} \frac{1}{ng}\sum_{i=1}^n \left|K^{(1)}\left(\frac{X_i - z}{g}\right)\right|\right\} \times C\sqrt{ng}g^{-1}h(t - \underline{t}_\gamma). \qquad (16)$$

Mimicking the proof of Corollary 1 in [8], one shows that as $K(\cdot)$ as bounded second order derivative, and using Assumptions $\mathcal{A}_1$ and $\mathcal{A}_5$,

$$\sup_{z \in I_x} \frac{1}{ng}\sum_{i=1}^n \left|K^{(1)}\left(\frac{X_i - z}{g}\right)\right| = O_{\mathrm{a.s}}(1).$$

In view of this result, and since $h/g < \infty$, Equation (16) is $O_{\mathrm{a.s}}\left\{(ng)^{-\gamma}g^{-1}h\right\} = o_{\mathrm{a.s}}(1)$. Therefore, $\sqrt{ng}(\widetilde{\mathcal{E}}_{z_t g} - \widetilde{\mathcal{E}}_{z_{\underline{t}_\gamma} g}) = o_{\mathrm{a.s}}(1)$. Equation (15) then becomes

$$\widetilde{W}_{xn}(t, \underline{u}_\gamma) - \widetilde{W}_{xn}(\underline{t}_\gamma, \underline{u}_\gamma) = \sqrt{ng}\left[\widetilde{\mathcal{E}}_{z_t g}\left\{F_{z_t}^{-1}(\underline{u}_\gamma)\right\} - \widetilde{\mathcal{E}}_{z_t g}\left\{F_{z_{\underline{t}_\gamma}}^{-1}(\underline{u}_\gamma)\right\}\right] + o_{\mathrm{a.s}}(1)$$
$$= \sqrt{ng}\left[\widetilde{\mathcal{E}}_{z_{\underline{t}_\gamma} g}\{F_{z_t}^{-1}(\underline{u}_\gamma)\} - \widetilde{\mathcal{E}}_{z_{\underline{t}_\gamma} g}\{F_{z_{\underline{t}_\gamma}}^{-1}(\underline{u}_\gamma)\}\right] + o_{\mathrm{a.s}}(1).$$

Using the same strategy with the first term of Equation (14), one deduces that

$$\widetilde{W}_{xn}(t,\overline{u}_\gamma) - \widetilde{W}_{xn}(t,\underline{u}_\gamma) = \sqrt{ng}\left[\widetilde{\mathcal{E}}_{z_{\underline{t}_\gamma}g}\{F_{z_t}^{-1}(\overline{u}_\gamma)\} - \widetilde{\mathcal{E}}_{z_{\underline{t}_\gamma}g}\{F_{z_{\underline{t}_\gamma}}^{-1}(\underline{u}_\gamma)\}\right] + o_{\text{a.s}}(1).$$

As assumption $\mathcal{A}_4$ implies that the function $z \mapsto F_z^{-1}$ is continuous in a neighborhood of x, and from Assumption $\mathcal{A}_2$, one deduces that for sufficiently large n, one can find a constant $\eta > 0$ such that uniformly in u, $F_{z_t}^{-1}(u) \le \min\{F_{z_{\underline{t}_\gamma}}^{-1}(u+\eta(z_t - z_{\underline{t}_\gamma})), F_{z_{\overline{t}_\gamma}}^{-1}(u+\eta(z_{\overline{t}_\gamma} - z_t))\} \le \min\{F_{z_{\underline{t}_\gamma}}^{-1}(\overline{u}_\gamma), F_{z_{\overline{t}_\gamma}}^{-1}(\overline{u}_\gamma)\}$, since $\max\{z_{\overline{t}_\gamma} - z_t, z_t - z_{\underline{t}_\gamma}\} \le Ch\kappa_\gamma^{-1} \le \kappa_\gamma^{-1}$ for sufficiently small h. Similarly, $F_{z_t}^{-1}(u) \ge \max\{F_{z_{\underline{t}_\gamma}}^{-1}(\underline{u}_\gamma), F_{z_{\overline{t}_\gamma}}^{-1}(\underline{u}_\gamma)\}$. In view of last discussion and of decomposition (14), one concludes that

$$\begin{aligned}\sup_{(t,u)\in\mathcal{I}}&\left|\widetilde{W}_{xn}(t,u) - \widetilde{W}_{xn}(\underline{t}_\gamma,\underline{u}_\gamma)\right| \\ &\le 4\max_{t\in J_\gamma}\sup_{u\in[0,1]}\sqrt{ng}\left|\widetilde{\mathcal{E}}_{z_tg}\left\{F^{-1}(\overline{u}_\gamma)\right\} - \widetilde{\mathcal{E}}_{z_tg}\left\{F_{z_t}^{-1}(\underline{u}_\gamma)\right\}\right| + o_{\text{a.s}}(1).\end{aligned}$$

For $1 \le i \le \kappa_\gamma$, conveniently writing $F_z^{-1}(0) = -\infty$ and $F_z^{-1}(1) = \infty$, let

$$A_\gamma(i,t) = [F_{z_t}^{-1}(\frac{i-1}{\kappa_\gamma}), F_{jz_t}^{-1}(\frac{i}{\kappa_\gamma})].$$

Using similar arguments as in the end of Section B.1, one has for sufficiently large n that

$$\mathfrak{W}_\delta(\widetilde{Z}_{xn},\mathcal{I}) \le 12\max_{t\in J_\gamma}\max_{1\le i\le\kappa_\gamma}\sqrt{ng}|\ \widetilde{\mathcal{E}}_{z_tg}\{A_\gamma(i,t)\}|.$$

For any interval $A = [a,b] \subset \mathbb{R}$, denote $\nu_z(A) = F_z(b) - F_z(a)$. Then, for any $\epsilon > 0$, one uses the Markov Inequality to obtain:

$$\begin{aligned}\mathbb{P}\left\{\mathfrak{W}_\delta\left(\widetilde{Z}_{xn},\mathcal{I}\right) \ge \epsilon\right\} &\le \mathbb{P}\left[\max_{t\in J_\gamma}\max_{1\le i\le\kappa_\gamma}\sqrt{ng}|\ \widetilde{\mathcal{E}}_{z_tg}\{A_\gamma(i,t)\}| \ge \frac{\epsilon}{12}\right] \\ &\le \frac{(ng)^{1+2\gamma}}{\epsilon^6}\left(\max_{t\in J_\gamma}\max_{1\le i\le\kappa_\gamma}\mathrm{E}\Big[\sqrt{ng}\,\widetilde{\mathcal{E}}_{z_tg}\{A_\gamma(i,t)\}\Big]^6\right),\end{aligned}$$

As the assumptions of Lemma 2.4 are satisfied, identical computations as in section B.2 with ν_x being replace with ν_{z_t} enable to find a constant $\omega < \infty$ such that for any interval $A \subset [0,1]$:

$$\mathrm{E}\left[\left\{\sqrt{ng}\,\widetilde{\mathcal{E}}_{z_tg}(A)\right\}^6\right] \le \frac{\omega}{(ng)^3}\left[\mathcal{H}_{z_t}(A,n,g) + \mathcal{J}(n,g)\right],$$

where $\mathcal{H}_{z_t}(A,n,g) = n\{g\nu_{z_t}(A)\}^{1/6} + n^2\{g\nu_{z_t}(A)\}^{7/6} + n^3\{g\nu_{z_t}(A)\}^{5/2}$ and $\mathcal{J}(n,h) = ng^{1/2} + n^2g^{7/2} + n^3g^{15/2}$. The conditions $h/g < \infty$ and $ng^5 < \infty$ ensures that $g \sim n^{-\tau}$ with $1/5 \le \tau < 13/23$. Hence, because $\nu_{z_t}(A_\gamma(i,t)) = (ng)^{-1/2-\gamma}$, taking any positive γ such that $2\gamma < \min\{13/23 - \tau, 1/10\}$ yields

$$\frac{(ng)^{1+2\gamma}}{\epsilon^6}\left(\max_{t\in J_\gamma}\max_{1\le i\le\kappa_\gamma}\mathrm{E}\Big[\sqrt{ng}\,\widetilde{\mathcal{E}}_{z_tg}\{A_\gamma(i,t)\}\Big]^6\right) = o(1).$$

The result is therefore proven.