Estimation of conditional copulas: revisiting asymptotic results in the i.i.d. case and extension to serial dependence

Félix Camirand Lemyre[a], Taoufik Bouezmarni[a], Jean-François Quessy[b]

[a]*Département de mathématiques, Université de Sherbrooke, Québec, Canada*
[b]*Département de mathématiques et d'informatique, Université du Québec à Trois-Rivières, Trois-Rivières, Canada*

Abstract

Let $(Y_1, Y_2, X) \in \mathbb{R}^3$ be a random vector and consider the conditional distribution $H_x(y_1, y_2) = \mathbb{P}(Y_1 \leq y_1, Y_2 \leq y_2 | X = x)$. The conditional copula in that context is the dependence structure that one extracts from H_x via the celebrated Sklar's theorem. Specifically, if the conditional marginal distributions F_{1x} and F_{2x} of H_x are continuous, then there exists a unique conditional copula $C_x : [0,1]^2 \to [0,1]$ such that $H_x(y_1, y_2) = C_x\{F_{1x}(y_1), F_{2x}(y_2)$ for all $(y_1, y_2) \in \mathbb{R}^2$. Then, C_x contains all the information on the form, and in particular on the strength, of the dependence between Y_1 and Y_2 for a given value of the covariate X. This paper considers the estimation of C_x when serially dependent copies of (Y_1, Y_2, X) are available, extending recent results obtained by Veraverbeke et al. (2011) and Gijbels et al. (2011). Such observations naturally appear in financial contexts. It is shown that under appropriate conditions, the limiting behavior of suitably standardized versions match those in the i.i.d. case. An application is given on measuring causality between Standard & Poor's 500 index data and volume.

Keywords: α-mixing processes, conditional copula, empirical copula process, functional delta method, weak convergence

Email addresses: `felix.camirand.lemyre@usherbrooke.ca` (Félix Camirand Lemyre), `taoufik.bouezmarni@usherbrooke.ca` (Taoufik Bouezmarni), `jean-francois.quessy@uqtr.ca` (Jean-François Quessy)

1. Introduction

Consider a random pair (Y_1, Y_2) from a joint distribution H having continuous marginal distributions F_1 and F_2. Then a celebrated theorem due to Sklar (1959) ensures the existence of a unique copula $C : [0,1]^2 \to [0,1]$ such that $H(y_1, y_2) = C\{F_1(y_1), F_2(y_2)\}$ holds for all $(y_1, y_2) \in \mathbb{R}^2$; see Nelsen (2006) for details on the theory of copulas. Based on $(Y_{11}, Y_{21}), \ldots, (Y_{1n}, Y_{2n})$ i.i.d. H, the nonparametric estimation of C is usually performed by computing the empirical copula. A version that is asymptotically equivalent to that originally proposed by Rüschendorf (1976) is

$$C_n(u_1, u_2) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\left\{Y_{1i} \le F_{n1}^{-1}(u_1), Y_{2i} \le F_{n2}^{-1}(u_2)\right\},$$

where F_{n1} and F_{n2} are the marginal empirical distribution functions.

The asymptotic behavior of the empirical copula process $\mathbb{C}_n = \sqrt{n}(C_n - C)$ has been investigated by Deheuvels (1979) under independence, *i.e.* when $C(u, v) = uv$ is the independence copula. General weak convergence in the space $D([0,1]^2)$ of càdlàg functions equipped with the Skorohod topology was investigated by Gaenßler & Stute (1987); van der Vaart & Wellner (1996) show weak convergence in the space $\ell^\infty([a,b]^2)$ of bounded functions on $[a,b]^2$ for $0 < a < b < 1$. The result was extended to the space $\ell^\infty([0,1]^2)$ by Fermanian et al. (2004) while assuming the existence and continuity of the partial derivatives $C^{[1]}(u_1, u_2) = \partial C(u_1, u_2)/\partial u_1$ and $C^{[2]}(u_1, u_2) = \partial C(u_1, u_2)/\partial u_2$ on $[0,1]^2$. In that case, $\mathbb{C}_n$ converges weakly with respect to the supremum distance to a limit process of the form

$$\mathbb{C}(u_1, u_2) = \alpha(u_1, u_2) - C^{[1]}(u_1, u_2)\,\alpha(u_1, 1) - C^{[2]}(u_1, u_2)\,\alpha(1, u_2), \qquad (1)$$

where α is a continuous and centered Gaussian process such that

$$\mathrm{E}\left\{\alpha(u_1, u_2)\,\alpha(u_1', u_2')\right\} = C\left(u_1 \wedge u_1', v_1 \wedge v_1'\right) - C(u_1, u_2)\,C(u_1', u_2').$$

Here and in the sequel, $a \wedge b = \min(a, b)$ for $a, b \in \mathbb{R}$. As shown by Segers (2012), these requirements on the partial derivatives of C are not satisfied for many extensively used copula models. Fortunately, this author obtained that the result still holds if $C^{[1]}$ and $C^{[2]}$ exist and are continuous on the sets $(0,1) \times [0,1]$ and $[0,1] \times (0,1)$, respectively. The extension of this result to serially dependent data has been considered recently by Bücher & Volgushev (2013) and Bücher & Ruppert (2013).

One is often interested in the behavior of a random couple conditional on the value taken by some covariate. Specifically, let (Y_1, Y_2, X) be a random vector taking values in $\mathbb{R}^3$ and for a fixed $x \in \mathbb{R}$, consider the conditional joint distribution of (Y_1, Y_2) given $X = x$, that is

$$H_x(y_1, y_2) = \mathbb{P}\left(Y_1 \leq y_1, Y_2 \leq y_2 | X = x\right). \tag{2}$$

The conditional marginal distributions of Y_1 and Y_2 given $X = x$ obtains from H_x via $F_{1x}(y) = \lim_{w\to\infty} H_x(y, w)$ and $F_{2x}(y) = \lim_{w\to\infty} H_x(w, y)$. If F_{1x} and F_{2x} are continuous, then Sklar's theorem ensures that there exists a unique copula $C_x : [0, 1]^2 \to [0, 1]$ such that $H_x(y_1, y_2) = C_x\{F_{1x}(y_1), F_{2x}(y_2)\}$. Conversely, the copula associated to the bivariate conditional distribution H_x can be extracted from the formula

$$C_x(u_1, u_2) = H_x\left\{F_{1x}^{-1}(u_1), F_{2x}(u_2)\right\}. \tag{3}$$

The bivariate function C_x is called the *conditional copula*. The latter contains all the dependence feature of (Y_1, Y_2) given a fixed value taken by the covariate. For that reason, it is an important task to be able to estimate C_x.

The estimation of C_x from i.i.d. observations has been considered by Veraverbeke et al. (2011) and Gijbels et al. (2011). Specifically, assuming the availability of independent random triplets $\mathbf{W}_1, \ldots, \mathbf{W}_n$, where $\mathbf{W}_i = (Y_{1i}, Y_{2i}, X_i)$, two empirical versions of C_x were proposed. First consider the estimator of the joint conditional distribution H_x given by

$$H_{xh}(y_1, y_2) = \sum_{i=1}^{n} w_{hi}(x)\, \mathbb{I}\left(Y_{1i} \leq y_1, Y_{2i} \leq y_2\right), \tag{4}$$

where $w_{h1}, \ldots, w_{hn}$ are weight functions that smooth the covariate space and h is a bandwidth parameter. The latter typically depends on the sample size, but the subscript n is omitted in the sequel for notational simplicity. The conditional empirical marginal distributions are simply $F_{1xh}(y) = \lim_{w\to\infty} H_{xh}(y, w)$ and $F_{2xh}(y) = \lim_{w\to\infty} H_{xh}(w, y)$. From representation (3), a natural plug-in estimator of C_x is given by

$$\begin{aligned} C_{xh}(u_1, u_2) &= H_{xh}\left\{F_{1xh}^{-1}(u_1), F_{2xh}^{-1}(u_2)\right\} \\ &= \sum_{i=1}^{n} w_{hi}(x)\, \mathbb{I}\left\{Y_{1i} \leq F_{1xh}^{-1}(u_1), Y_{2i} \leq F_{2xh}^{-1}(u_2)\right\}, \end{aligned} \tag{5}$$

where for $j = 1, 2$, $F_{jxh}^{-1}(u) = \inf\{y \in \mathbb{R} : F_{jxh}(y) \geq u\}$ is the left-continuous generalized inverse of F_{jxh}. As noted by Gijbels et al. (2011), the estimator C_{xh} may be severely biased if any of the two marginal distributions is strongly influenced by the covariate. For that reason, they proposed a second estimator that aims at removing 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

$$\widetilde{G}_{xh}(v_1, v_2) = \sum_{i=1}^{n} w_{hi}(x)\, \mathbb{I}\left\{F_{1X_ih_1}(Y_{1i}) \leq v_1, F_{2X_ih_2}(Y_{2i}) \leq v_2\right\} \tag{6}$$

and estimate C_x with

$$\widetilde{C}_{xh}(u_1, u_2) = \widetilde{G}_{xh}\left\{\widetilde{G}_{1xh}^{-1}(u_1), \widetilde{G}_{2xh}^{-1}(u_2)\right\}, \tag{7}$$

where $\widetilde{G}_{1xh}$ and $\widetilde{G}_{2xh}$ are the marginals extracted from $\widetilde{G}_{xh}$. The investigation of the asymptotic properties of the conditional empirical copula processes

$$\mathbb{C}_{xh} = \sqrt{nh}\big(C_{xh} - C_x\big) \quad \text{and} \quad \widetilde{\mathbb{C}}_{xh} = \sqrt{nh}\big(\widetilde{C}_{xh} - C_x\big)$$

as random elements in the space $\ell^\infty([0,1]^2)$ of bounded functions on $[0,1]^2$ endowed with the supremum norm have been done by Veraverbeke et al. (2011). Specifically, they obtained that $\mathbb{C}_{xh}$ and $\widetilde{\mathbb{C}}_{xh}$ both converge weakly to Gaussian processes whose stochastic behavior differ only on their respective bias. The main goals of this paper are to

(i) re-derive the above-mentioned asymptotic results of Veraverbeke et al. (2011) in a simpler way using the functional delta method;

(ii) extend these results to serially dependent observations, *i.e.* time series.

The paper is organized as followed. Section 2 revisits some results on classical and conditional empirical copula processes in the light of the functional delta method. Section 3 extends these results to conditional empirical copula processes under serial dependence. Some simulation results that aim at studying the accuracy of these estimators are provided in Section 4. An application is given in Section 5 on measuring causality in the bivariate time series of returns and volume of the Standard & Poor's 500 index. The proofs of the main results are to be found in Section 6 and auxiliary results needed for these theoretical results to hold are relegated to an Appendix.

2. Revisiting conditional empirical copulas in the i.i.d. case

2.1. A Hadamard functional

Let $\mathbb{D}$ be the space of bivariate distribution functions on $\mathbb{R}^2$ and consider the mapping $\Phi : \mathbb{D} \to \ell^\infty([0,1]^2)$ defined by

$$\Phi(H) = H \circ \left(F_1^{-1}, F_2^{-1}\right), \tag{8}$$

where $F_1(y) = \lim_{w\to\infty} H(y,w)$ and $F_2(y) = \lim_{w\to\infty} H(w,y)$. Then if F_1 and F_2 are continuous, the unique copula of H writes $C = \Phi(H)$. The Hadamard differentiability of Φ has been first investigated in van der Vaart & Wellner (1996) under strong assumptions on the partial derivatives of C. These assumptions were recently relaxed by Bücher & Volgushev (2013) and match those identified by Segers (2012).

Definition 2.1. *In the sequel, $\mathcal{D}$ is the class of copulas C such that the partial derivatives $C^{[1]}(u_1,u_2) = \partial C(u_1,u_2)/\partial u_1$ and $C^{[2]}(u_1,u_2) = \partial C(u_1,u_2)/\partial u_2$ exist and are continuous respectively on the sets $(0,1)\times[0,1]$ and $[0,1]\times(0,1)$.*

If the copula C of H belongs to $\mathcal{D}$, then Bücher & Volgushev (2013) show that Φ is Hadamard differentiable with derivative at H given by

$$\Phi'_H(\Delta)(u_1,u_2) = \widetilde{\Delta}(u_1,u_2) - C^{[1]}(u_1,u_2)\,\widetilde{\Delta}(u_1,1) - C^{[2]}(u_1,u_2)\,\widetilde{\Delta}(1,u_2),$$

with $\widetilde{\Delta}(u_1,u_2) = \Delta\{F_1^{-1}(u_1), F_2^{-1}(u_2)\}$. As a consequence, if H_n is an estimator of H such that $\mathbb{H}_n = \sqrt{n}(H_n - H)$ converges weakly to some limit $\mathbb{H}$, an application of the functional delta method yields for $C_n = \Phi(H_n)$ that

$$\mathbb{C}_n = \sqrt{n}\,(C_n - C) \rightsquigarrow \mathbb{C} = \Phi'_H\,(\mathbb{H})\,.$$

When $(Y_{11},Y_{21}),\ldots,(Y_{1n},Y_{2n})$ are i.i.d., H_n is usually the bivariate empirical distribution function. From classical results, $\mathbb{H}$ is a centered Gaussian process with covariance function given for each $(y_1,y_2),(y_1',y_2') \in \mathbb{R}^2$ by

$$\Gamma_H(y_1,y_2,y_1',y_2') = H(y_1 \wedge y_1', y_2 \wedge y_2') - H(y_1,y_2)\,H(y_1',y_2'). \tag{9}$$

It follows that the limit $\mathbb{C} = \Phi'_H(\mathbb{H})$ of $\mathbb{C}_n$ is a centered Gaussian process on $[0,1]^2$ of the form given in equation (1). One then recovers the result obtained by Segers (2012).

2.2. Asymptotic behavior of $\mathbb{C}_{xh}$

As already mentioned, the asymptotic behavior of $\mathbb{C}_{xh}$ has been investigated by Veraverbeke et al. (2011). Here, this result will be re-derived upon noting that $C_{xh} = \Phi(H_{xh})$, where the functional Φ is defined in equation (8). When the weights in the definition of H_{xh} are nonnegative and sum to one, H_{xh} is a distribution function and thus belongs to the space $\mathbb{D}$ of bivariate distribution functions on $\mathbb{R}^2$. In that case, the asymptotic behavior of $\mathbb{C}_{xh}$ will be a consequence of the large-sample behavior of $\mathbb{H}_{xh} = \sqrt{nh}(H_{xh} - H_x)$. To this end, one needs to introduce the following family of distributions.

Definition 2.2. *In the sequel, $\mathcal{F}_x$ is the class of distribution functions H on $\mathbb{R}^3$ whose conditional distribution $H_z(y_1, y_2) = \partial H(y_1, y_2, z)/\partial z$ is such that*

$$\dot{H}_z(y_1, y_2) = \frac{\partial}{\partial z} H_z(y_1, y_2) \quad \textit{and} \quad \ddot{H}_z(y_1, y_2) = \frac{\partial^2}{\partial z^2} H_z(y_1, y_2)$$

are uniformly continuous for each z in a neighborhood of x.

Now the following proposition can be deduced from Veraverbeke et al. (2011).

Proposition 2.3. *Let $\mathbf{W}_1, \ldots, \mathbf{W}_n$, where $\mathbf{W}_i = (Y_{1i}, Y_{2i}, X_i)$, be i.i.d. from a distribution function H that belongs to $\mathcal{F}_x$. Also assume that the weights in the definition of H_{xh} satisfy conditions W_1–W_5 described in Appendix A. If $nh \to \infty$ and $nh^5 \to K < \infty$ as $n \to \infty$, then the empirical process $\mathbb{H}_{xh}$ converges weakly to a Gaussian process $\mathbb{H}_x$ having covariance function*

$$\mathrm{Cov}\left\{\mathbb{H}_x(y_1, y_2), \mathbb{H}_x(y_1', y_2')\right\} = K_5\, \Gamma_{H_x}(y_1, y_2, y_1', y_2')$$

and asymptotic bias

$$B_{H_x}(y_1, y_2) = K \left\{ K_3\, \dot{H}_x(y_1, y_2) + \frac{K_4}{2}\, \ddot{H}_x(y_1, y_2) \right\}. \tag{10}$$

As a straightforward application of the functional delta method, one obtains from the conclusion of Proposition 2.3 that as long as C_x belongs to the class $\mathcal{D}$, the empirical process $\mathbb{C}_{xh}$ converges weakly to $\mathbb{C}_x = \Phi'_{H_x}(\mathbb{H}_x)$. One then deduces the asymptotic representation

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

where $\alpha_x(u_1,u_2) = \mathbb{H}_x\{F_{1x}^{-1}(u_1), F_{2x}^{-1}(u_2)\}$ is a C_x-Brownian bridge. Because Φ'_{H_x} is a linear functional, the asymptotic bias of $\mathbb{C}_x$ obtains via $\mathrm{E}\{\Phi'_{H_x}(\mathbb{H}_x)\} = \Phi'_{H_x}\{\mathrm{E}(\mathbb{H}_x)\} = \Phi'_{H_x}(B_x)$. From equation (10), one obtains the expression already derived by Veraverbeke et al. (2011), namely

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

where $\widetilde{B}_{H_x}(u_1,u_2) = B_{H_x}\{F_{1x}^{-1}(u_1), F_{2x}^{-1}(u_2)\}$.

Remark 2.4. *When H_{xh} fails to be a distribution function, a version $\mathbb{H}'_{xh}$ is obtained upon setting negative weights to zero and by re-scaling the resulting weights in order that they sum to one. Then, as long as the difference between $\mathbb{H}_{xh}$ and $\mathbb{H}'_{xh}$ is asymptotically negligible, the large-sample behavior of $\mathbb{C}_{xh}$ obtains as a consequence of that of $\mathbb{H}'_{xh}$. As pointed out in Omelka et al. (2013), this holds if*

$$\sqrt{nh} \sup_{z\in J_x^{(n)}} \left| \sum_{i=1}^n w_{hi}(z) - 1 \right| \quad \textit{and} \quad \sup_{z\in J_x^{(n)}} \left| \sum_{i=1}^n w_{hi}(z)\, \mathbb{I}\{w_{hi}(z) < 0\} \right|$$

both converge in probability to zero, where $J_x^{(n)} = [\min_{i\in I_x^{(n)}} X_i, \max_{i\in I_x^{(n)}} X_i]$ and $I_x^{(n)} = \{i : w_{hi}(x) > 0\}$.

2.3. Asymptotic behavior of $\widetilde{\mathbb{C}}_{xh}$

Obtaining the asymptotic behavior of $\widetilde{\mathbb{C}}_{xh}$ is more tricky because it is based on the *pseudo-uniformized* observations $(\widetilde{U}_{11},\widetilde{U}_{21}),\ldots,\widetilde{U}_{1n},\widetilde{U}_{2n})$. As a first step, consider a version of $\widetilde{G}_{xh}$ computed from $(U_{11},U_{21}),\ldots,(U_{1n},U_{2n})$, where $U_{1i} = F_{1X_i}(Y_{1i})$ and $U_{2i} = F_{2X_i}(Y_{2i})$, namely

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

The weak limit of $\mathbb{G}_{xh} = \sqrt{nh}(G_{xh} - C_x)$ obtains from Proposition 2.3 with H replaced by C. Therefore, under the conditions of Proposition 2.3, $\mathbb{G}_{xh}$ converges weakly to $\mathbb{G}_x$, where $\mathbb{G}_x$ is a Gaussian process with covariance function $\mathrm{Cov}\{\mathbb{G}_x(u_1,u_2), \mathbb{G}_x(u'_1,u'_2)\} = K_5\,\Gamma_{C_x}(u_1,u_2,u'_1,u'_2)$ and asymptotic bias $\mathrm{E}\{\mathbb{G}_x(u_1,u_2)\} = B_{C_x}(u_1,u_2)$. If in addition, C_x belongs to $\mathcal{D}$, then the empirical process $\widetilde{\mathbb{C}}^\star_{xh} = \sqrt{nh}\{\Phi(G_{xh}) - \Phi(C_x)\}$ converges weakly to

$\widetilde{\mathbb{C}}_x^\star = \Phi'_{C_x}(\mathbb{G}_x)$. Since $B_{C_x}(u,1) = B_{C_x}(1,u) = 0$ for any $u \in [0,1]$, one deduces $\mathrm{E}\{\Phi'_{C_x}(\mathbb{G}_x)\} = \Phi'_{C_x}(B_{C_x}) = B_{C_x}$ as the bias of $\widetilde{\mathbb{C}}_x^\star$.

It remains to show that the processes $\widetilde{\mathbb{C}}_{xh}^\star$ and $\widetilde{\mathbb{C}}_{xh}$ are asymptotically equivalent. It is indeed the case if one can show that the *pseudo uniformized* observations are close to the *truely* uniformized observations. This is the subject of the next proposition whose consequence is that one recovers the result stated in Veraverbeke et al. (2011).

Proposition 2.5. *Suppose that for $j = 1,2$, the functions $F_{jz}\{F_{jz}^{-1}(u)\}$, $\dot{F}_{jz}\{F_{jz}^{-1}(u)\}$ and $\ddot{F}_{jz}\{F_{jz}^{-1}(u)\}$ are continuous in (z,u) for z in a neighborhood of x. Further assume that $nh_1 \to \infty$ and $nh_2 \to \infty$ as $n \to \infty$, and that $h/\min(h_1,h_2) < \infty$ with $nh_1^5 < \infty$ and $nh_2^5 < \infty$. Then under the conditions of Proposition 2.3 and if in addition, conditions W_6–W_{10} in Appendix A hold, the empirical process $\widetilde{\mathbb{C}}_{xh}$ converges weakly to $\Phi'_{C_x}(\mathbb{G}_x)$.*

3. Asymptotic behavior of $\mathbb{C}_{xh}$ and $\widetilde{\mathbb{C}}_{xh}$ under serial dependence

3.1. α-mixing processes

Consider a sequence $(\eta_i)_{i\in\mathbb{Z}}$ of random vectors defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The σ-field generated by $(\eta_i)_{a\le i\le b}$, $a,b \in \mathbb{Z}\cup\{-\infty,+\infty\}$ is denoted by $\mathcal{F}_a^b$. As defined, *e.g.* in Rio (2000), the α-mixing coefficients of $(\eta_i)_{i\in\mathbb{Z}}$ are given by $\alpha_\eta(0) = 1/2$ and for $r \ge 1$,

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

where for two σ-fields $\mathcal{A},\mathcal{B} \subset \mathcal{F}$,

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

The process $(\eta_i)_{i\in\mathbb{Z}}$ is said to be α-mixing if $\alpha_\eta(r) \to 0$ as $r \to \infty$. The α-mixing condition is also called *strong mixing assumption* in the literature. Many well-known stochastic processes satisfy this property, for example the ARMA and ARCH processes are $\alpha-$mixing under some regular conditions on their parameters. For more details, see Meitz & Saikkonen (2002), Doukhan (1994) and Carrasco & Chen (2002).

The following two subsections establish the asymptotic behavior of $\mathbb{C}_{xh}$ and $\widetilde{\mathbb{C}}_{xh}$ under the strong mixing assumption.

3.2. The process $\mathbb{C}_{xh}$

Similarly as in the i.i.d. case, the asymptotic behavior of $\mathbb{C}_{xh}$ under serial dependence can be derived as a consequence of the large-sample behavior of the associated weighted empirical process $\mathbb{H}_{xh} = \sqrt{nh}(H_{xh} - H_x)$. As pointed out in the next theorem, one obtains under certain conditions on the weights that $\mathbb{H}_{xh}$ converges weakly, under a strong mixing assumption, to a Gaussian process whose representation coincides with that of $\mathbb{H}_x$ appearing in Proposition 2.3 in the i.i.d. situation.

Theorem 3.1. *Let $(\mathbf{W}_i)_{i\in\mathbb{Z}}$, where $\mathbf{W}_i = (Y_{1i}, Y_{2i}, X_i)$, be a stationary sequence with α-mixing coefficients such that $\alpha_{\mathbf{W}}(r) = O(r^{-a})$ for some $a > 6$. Assume that their common joint distribution H belongs to the class $\mathcal{F}_x$ and suppose that conditions $W_1^\star$, W_2–W_5 and W_{11}–W_{13} on the weights are fulfilled. Then if $nh \to \infty$ and $nh^5 \to K < \infty$ as $n \to \infty$, the process $\mathbb{H}_{xh}$ converges weakly to the same Gaussian process $\mathbb{H}_x$ as in the i.i.d. case.*

Interestingly, the asymptotic covariance structure of $\mathbb{H}_{xh}$ under a strong mixing assumption matches the asymptotic covariance structure of $\mathbb{H}_{xh}$ found in the i.i.d. setting. In other words, the impact of time-dependency, compared to the serially independent case, is asymptotically negligible. This is due to the use of the weight functions that smooth the covariate space in a shrinking neighborhood of x as n goes to infinity. Note however that the stronger assumptions $W_1^\star$ and W_{11}–W_{13} on the weight functions were needed in order to tackle moments of order six involved by time-dependency. These conditions are fulfilled, among others, by the Nadaraya–Watson and local linear weights. The latter are given respectively by

$$w_{hi}^{\mathrm{NW}}(x) = \frac{K\left(\frac{X_i - x}{h}\right)}{S_{n,0}} \quad \text{and} \quad w_{hi}^{\mathrm{LL}}(x) = \frac{K\left(\frac{X_i - x}{h}\right)\left\{S_{n,2} - \left(\frac{X_i - x}{h}\right) S_{n,1}\right\}}{S_{n,0}S_{n,2} - S_{n,1}^2},$$

where K is a symmetric and continuously differentiable kernel density function on $[-1, 1]$ and for $j \in \{0, 1, 2\}$,

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

The following corollary to Theorem 3.1 is a straightforward consequence of the Hadamard differentiability of the functional Φ defined in (8) and of the weak convergence of $\mathbb{H}_{xh}$ under a strong mixing assumption.

Corollary 3.2. *If the conditions of Theorem 3.1 are met and C_x belongs to $\mathcal{D}$, then the empirical process $\mathbb{C}_{xh}$ converges weakly to $\mathbb{C}_x = \Phi'_{H_x}(\mathbb{H}_x)$. Hence, $\mathbb{C}_{xh}$ has the same asymptotic behavior as in the i.i.d. case.*

3.3. The process $\widetilde{\mathbb{C}}_{xh}$

From the definitions of $\widetilde{G}_{xh}$ and G_{xh}, one can write

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

Provided the conditions of Corollary 3.2 are satisfied, one deduces that the first summand on the right converges weakly to $\widetilde{\mathbb{C}}_x = \Phi'_{C_x}(\mathbb{G}_x)$. Therefore, the key result for the weak convergence of $\widetilde{\mathbb{C}}_{xh}$ is the asymptotic negligibility of the second summand; this is established next. Then under the conditions of Proposition 2.3 and if in addition, conditions W_6–W_{10} in Appendix A hold, the empirical process $\widetilde{\mathbb{C}}_{xh}$ converges weakly to $\Phi'_{C_x}(\mathbb{G}_x)$.

Theorem 3.3. *Let $(\mathbf{W}_i)_{i\in\mathbb{Z}}$, where $\mathbf{W}_i = (Y_{1i}, Y_{2i}, X_i)$, be a stationary sequence with associated α-mixing coefficients such that $\alpha_{\mathbf{W}}(r) = O(r^{-a})$ for some $a > 6$. Suppose that for $j = 1, 2$, the functions $F_{jz}\{F_{jz}^{-1}(u)\}$, $\dot{F}_{jz}\{F_{jz}^{-1}(u)\}$ and $\ddot{F}_{jz}\{F_{jz}^{-1}(u)\}$ are continuous in (z, u) for z in a neighborhood of x. Further assume that $nh_1 \to \infty$ and $nh_2 \to \infty$ as $n \to \infty$, $\sqrt{nh_1}h_1^2 < \infty$, $\sqrt{nh_2}h_2^2 < \infty$ and that $h/\min(h_1, h_2) < \infty$. In addition, assume that conditions $W_1^\star$, W_2–W_{13} on the weights are satisfied. If the distribution function of $(F_{1X}(Y_1), F_{2X}(Y_2), X)$ belongs to $\mathcal{F}_x$, then*

$$\sup_{(u_1,u_2)\in[0,1]^2} \sqrt{nh}\left|\Phi(\widetilde{G}_{xh})(u_1, u_2) - \Phi(G_{xh})(u_1, u_2)\right|$$

converges in probability to zero as $n \to \infty$.

4. Sample behavior of the conditional copula estimators

In order to evaluate the sample performance of the estimators C_{xh} and $\widetilde{C}_{xh}$, Veraverbeke et al. (2011) compared their asymptotic bias under various scenarios. They observed that generally speaking, $\widetilde{C}_{xh}$ has a significantly smaller bias than C_{xh}. The aim of this section is to perform a similar investigation in the case of serially dependent data. To this end, consider the general autoregressive data generating process

$$\mathbf{W}_i = \theta\,\mathbf{W}_{i-1} + (1 - \theta^2)^{1/2}\boldsymbol{\varepsilon}_i,$$

where for each $i \in \mathbb{N}$, $\mathbf{W}_i = (Y_{1i}, Y_{2i}, X_i)$ and $\boldsymbol{\varepsilon}_i$ are i.i.d. from some distribution on $\mathbb{R}^3$ having mean vector zero. In that model, the parameter $\theta \in [0, 1)$ controls the strength of the serial dependence between successive observations; it is a situation where the data are α-mixing.

For the results that will be presented next, one restricts to the case when $\boldsymbol{\varepsilon}_i$ are standard normal with some correlation matrix $R = (R_{ij})_{i,j=1}^3$ where R_{ij} is the correlation coefficient between Y_i and Y_j . In that case, it is well known that the joint conditional distribution of (Y_{1i}, Y_{2i}) given $X_i = x$ is the bivariate Normal distribution with correlation coefficient

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

As a consequence, the conditional copula of (Y_{1i}, Y_{2i}) given $X_i = x$ is simply the so-called Gaussian copula with parameter ρ_x. A convenient way to quantify dependence in a bivariate random vector is via its corresponding value of Kendall's tau, which writes in terms of the underlying copula C as

$$\tau(C) = 4 \int_0^1 \int_0^1 C(u_1, u_2)\, \mathrm{d}C(u_1, u_2) - 1. \tag{11}$$

For a given conditional copula C_x, the conditional Kendall's tau is simply $\tau_x = \tau(C_x)$. For the conditional Gaussian copula, $\tau_x = (2/\pi) \sin^{-1}(\rho_x)$. Four variants of the model have been considered; they are described in Table 1.

Table 1: The four models considered in the simulation study

Model	R_{12}	R_{13}	R_{23}	ρ_x	τ_x
M_1	.9	.8	.8	.72	.52
M_2	-.9	.8	-.8	-.72	-.52
M_3	.8	.1	.1	.80	.59
M_4	.1	.1	.1	.09	.06

The relative performance of C_{xh} and $\widetilde{C}_{xh}$ for the estimation of C_x has been evaluated in the light of the *average integrated squared bias* (AISB) and the *average integrated variance* (AIV). To be specific, if $\widehat{C}$ is some estimator of C_x, then

$$\mathrm{AISB}(\widehat{C}) = \int_0^1 \int_0^1 \left\{ \mathrm{E}(\widehat{C}(u_1, u_2)) - C_x(u_1, u_2) \right\}^2 \mathrm{d}u_1 \mathrm{d}u_2$$

and

$$\mathrm{AIV}(\widehat{C}) = \int_0^1 \int_0^1 \left\{ \widehat{C}^2(u_1,u_2) - \Big(\mathrm{E}(\widehat{C}(u_1,u_2))\Big)^2 \right\} \mathrm{d}u_1 \mathrm{d}u_2.$$

The latter have been estimated from 1 000 replicates under each of the scenario considered with $n = 250$. The results about the estimated values of AISB and AIV for the two conditional copula estimators are reported in Figures 1–4 respectively for models M_1–M_4.

First observe that, interestingly, the integrated variance is very similar for any values of θ. This is in accordance with the theoretical results of Section 3 that states that the estimators act, asymptotically, as in the i.i.d. case. Considering model M_1 first, one notes that the $\widetilde{C}_{xh}$ outperforms C_{xh} in term of AISB. This difference might be explained by the fact that $\mathrm{E}(\mathbb{C}_{xh})$ depends in general on F_{1x} and F_{2x}, and therefore is affected in some way by the dependence in the conditional marginals. Also observe that the integrated bias of $\widetilde{C}_{xh}$ stabilizes as the bandwidth parameter h takes large values; it is not the case for C_{xh}. Finally note that C_{xh} do slightly better than $\widetilde{C}_{xh}$ in term of AIV. Under model M_2, one has $R_{13} = -R_{23}$ and this entails that $\mathrm{E}\{\mathbb{C}_x(u,u)\} = \mathrm{E}\{\widetilde{\mathbb{C}}_x(u,u)\}$. As a consequence, the terms involving the marginal distributions cancel. For M_3 and M_4, the two estimators perform quite similarly in term of bias. This may be explained by the fact that here, the covariate has a weak influence on the marginal distributions. However, there is a slight advantage of C_{xh} in term of variance for small values of the bandwidth h.

5. Empirical application: measuring causality in financial data

The concept of causality was introduced by Wiener (1956) and Granger (1969) in order to study the dynamic relationship between time series. Whenever a test for the non-causality hypothesis is rejected, it is important to appropriately quantify the strength of this causality. In this regard, measures based on linear models were proposed by Geweke (1982) and Geweke (1984), while measure using the Kullback–Leibler information criterion were investigated by Gouriéroux et al. (1987) while assuming the normality of the data. A nonparametric copula-based measure was recently proposed by Taamouti et al. (2014). However, the above-mentioned measures of causality are *global* in the sense that they do not allow to quantify the *local* strength of causality.

A way to measure local causality in time series is by the mean of the conditional Kendall's measure of association, namely $\tau_x = \tau(C_x)$, where $\tau(C)$ is defined in (11). To be specific, we consider $(\mathbf{Z}_i)_{i\in\mathbb{Z}}$, where $\mathbf{Z}_i = (Z_{1i}, Z_{2i})$. Then, define the process $(\mathbf{W}_i)_{i\in\mathbb{Z}}$, where $\mathbf{W}_i = (Z_{1i}, Z_{2,i-1}, Z_{1,i-1})$. The conditional Kendall's tau of $(Z_{1i}, Z_{2,i-1})$ given $Z_{1,i-1}$ is a nonparametric measure of the local Granger causality from Z_2 to Z_1. In other words, based on $\mathbf{W}_1, \ldots, \mathbf{W}_n$, one considers $\widehat{\tau}_x$ arising while replacing C_x in $\tau(C_x)$ by one of the two conditional copula estimators described in this work.

An illustration of this new local measure of causality will be given on the time series of daily returns and volume of the Standard and Poor's 500 (S&P500) Index. The relationship between these two indices has been the subject of extensive theoretical and empirical researches; see Bouezmarni et al. (2012) for more details. The data set that will be analyzed comes from Yahoo Finance and consists of $n = 3\ 032$ daily observations taken between January 1997 and January 2009. Specifically, one considers the continuously compounded changes in prices (returns) and trading volume (volume growth rate). Also, according to the tests of stationarity reported in Bouezmarni et al. (2012), the first difference of logarithmic price and volume, rather than their level, are considered. Consequently, the upcoming causality relations have to be interpreted in terms of growth rates. First note that the null hypothesis of Granger non-causality from returns to volume was clearly rejected by the tests proposed by Bouezmarni et al. (2012) and Su & White (2008). Moreover, a non-linear feedback effect from volume to returns was detected. In this therefore of interest to study these relationships in the light of the proposed local causality measure.

First, we consider the causality measure from volume to return. In this case, the return (resp. volume) plays the role of Z_1 (resp. Z_2) in the process $(\mathbf{W}_i)_{i\in\mathbb{Z}}$ given above. Figure A.5 reports the resulting value of $\widehat{\tau}_x$ as a function of x. Note that the confidence bands are based on a multiplier bootstrap method under the independence assumption, see Camirand (2013) for more details. In this figure, one can see that the conditional Kendall's tau that measures the causality from volume to return clearly exceeds the confidence bands in the central region of the return when $\widehat{\tau}_x$ is based on $\widetilde{C}_{xh}$; this is less obvious when $\widehat{\tau}_x$ is based on C_{xh}.

Second, we consider the causality measure from return to volume. In this case, the volume (resp. return) plays the role of Z_1 (resp. Z_2). Figure A.6 illustrates the resulting value of $\widehat{\tau}_x$ as a function of x. In this figure, the conditional Kendall's tau that measures the causality from return to volume clearly exceeds the confidence bands in the right region of the volume when $\widehat{\tau}_x$ is based on C_{xh}. Also, one can see that it clearly exceeds the confidence bands in the central region of the volume when $\widehat{\tau}_x$ is based on $\widetilde{C}_{xh}$.

6. Proofs of the main results

6.1. Proof of Theorem 3.1

The proof divides into two steps, namely

(i) it is first shown that the finite-dimensional distributions of the process $Z_{xn} = \mathbb{H}_{xh} - \mathrm{E}(\mathbb{H}_{xh})$ are asymptotically Normal;

(ii) the tightness of the sequence Z_{xn} is established.

Finite-dimensional distributions. It will be shown that the random variable $Z_{xn}(y_1, y_2)$ is asymptotically normally distributed for fixed and arbitrary $y_1, y_2 \in \mathbb{R}$. The arguments that will be presented next for that purpose can easily be adapted to show the joint weak convergence of any finite-dimensional components by mean of the Cramér–Wold device.

First define $\sigma_z^2(y_1, y_2) = \Gamma_{H_z}(y_1, y_2, y_1, y_2)$, where Γ_H is given in (9). Then for each $i \in \{1, \ldots, n\}$, let $\vartheta_i(y_1, y_2) = \mathbb{I}(Y_{1i} \leq y_1, Y_{2i} \leq y_2) - H_{x_i}(y_1, y_2)$ and observe that

$$\begin{aligned}\mathrm{Var}\left\{Z_{xn}(y_1, y_2)\right\} &= nh \sum_{i=1}^n \left\{w_{hi}(x)\right\}^2 \sigma_{x_i}^2(y_1, y_2) \\ &\quad + 2 \sum_{i=1}^n \sum_{\ell=1}^{n-i} w_{hi}(x)\, w_{h,i+\ell}(x)\, \mathrm{Cov}\left\{\vartheta_i(y_1, y_2), \vartheta_{i+\ell}(y_1, y_2)\right\} \\ &= \Lambda_{n1}(y_1, y_2) + \Lambda_{n2}(y_1, y_2).\end{aligned}$$

Since $H \in \mathcal{F}_x$, one has $\sigma_{x_i}(y_1, y_2) = \sigma_x(y_1, y_2) + o(1)$ for any $i \in I_{nx}$, and then in view of Condition W_4, $\Lambda_{n1}(y_1, y_2) \to K_5\, \sigma_x(y_1, y_2)$ as $n \to \infty$. For the second summand, one follows a similar idea as in Li & Racine (2007). To this end, note that the α-mixing coefficients are such that

$$\left|\mathrm{Cov}\left\{\vartheta_i(y_1, y_2), \vartheta_{i+\ell}(y_1, y_2)\right\}\right| \leq \alpha_{\mathbf{W}}(\ell) \leq 1.$$

Then, for $\pi_n = \lfloor h^{-1/2} \rfloor$, one can write

$$\begin{aligned}
\Lambda_{n2}(y_1, y_2) &= nh \sum_{i=1}^{n} \sum_{\ell=1}^{\pi_n} w_{hi}(x)\, w_{h,i+\ell}(x) \operatorname{Cov}\{\vartheta_i(y_1, y_2), \vartheta_{i+\ell}(y_1, y_2)\} \\
&\quad + nh \sum_{i=1}^{n} \sum_{\ell=\pi_n+1}^{n-i} w_{hi}(x)\, w_{h,i+\ell}(x) \operatorname{Cov}\{\vartheta_i(y_1, y_2), \vartheta_{i+\ell}(y_1, y_2)\} \\
&\le nh \left(\pi_n \max_{1\le\ell\le\pi_n} \sum_{i=1}^{n-\pi_n} w_{hi}(x) w_{h,i+\ell}(x) + \max_{1\le\ell\le n} w_{hi}(x) \sum_{\ell=\pi_n+1}^{n} \alpha_{\mathbf{W}}(\ell) \right) \\
&= O\left(h^{1/2} + \sum_{\ell=\pi_n+1}^{n} \alpha_{\mathbf{W}}(\ell) \right),
\end{aligned}$$

where the last equality follows from assumptions $W_1^\star$ and W_{11}. From the assumptions on h and on the α-mixing coefficients, one concludes that this last expression is indeed $o(1)$. Thus, as $n \to \infty$, $\operatorname{Var}\{Z_{xn}(y_1, y_2)\} \to K_5\, \sigma_x^2(y_1, y_2)$.

Now in order to show the asymptotic normality, a blocking technique described for instance in Billingsley (1968) will be used. To this end, write $n = r_n(b_n + \ell_n)$ and assume without loss of generality that r_n, b_n and ℓ_n are integers such that $b_n \sim n^{1-\epsilon+\delta}$ and $\ell_n \sim n^{1-\epsilon}$ for some $\delta, \epsilon > 0$. Then, introduce for each $i \in \{1, \ldots, n\}$ the sets $\mathcal{S}_{n1}(i) = \{j \in \mathbb{N} : (i-1)(b_n + \ell_n) + 1 \le j \le (i-1)(b_n + \ell_n) + b_n\}$ and $\mathcal{S}_{n2}(i) = \{j \in \mathbb{N} : i(b_n + \ell_n) + 1 - \ell_n \le j \le i(b_n + \ell_n)\}$. With this notation, one can write $Z_{xn} = U_n + W_n$, where $U_n = \sqrt{nh} \sum_{i=1}^{r_n} U_{ni}$ and $W_n = \sqrt{nh} \sum_{i=1}^{r_n} W_{ni}$, with

$$U_{ni} = \sum_{j \in S_{n1}(i)} w_{hj}(x)\, \vartheta_j(y_1, y_2) \quad \text{and} \quad W_{ni} = \sum_{j \in S_{n2}(i)} w_{hj}(x)\, \vartheta_j(y_1, y_2).$$

For any $\kappa > 0$, one has

$$\mathbb{P}(|W_n| > \kappa) \le \sum_{i=1}^{r_n} \mathbb{P}\left(|W_{ni}| > \frac{\kappa}{\sqrt{nh}\, r_n} \right) \le \sum_{i=1}^{r_n} \frac{r_n^4 (nh)^2}{\kappa^4} \operatorname{E}\left(|W_{ni}|^4 \right).$$

Now consider the following instrumental Lemma.

Lemma 6.1. *Assume $W_1^\star$, W_{11} and W_{12} are satisfied and suppose $\sum_{r\ge0}(r+1)^2 \alpha_{\mathbf{W}}(r) < \infty$. Let $d \in \{\ell_n, b_n\}$, and let $\mathcal{S}_n(k)$ be the corresponding set of indices (either $S_{n1}(k)$ or $S_{n2}(k)$) of length d and suppose $d\,h \to \infty$. For* *any $y_1, y_2 \in \mathbb{R}$, write $S_d(k) = \sum_{i \in \mathcal{S}_n(k)} \vartheta_i(y_1, y_2) w_{hi}(x)$. Then there exists a constant C_α such that*

$$\sum_{k=1}^{r_n} \mathrm{E}\left\{S_d(k)^4\right\} \leq C_\alpha n^{-3} h^{-2} d.$$

Because $\alpha_{\mathbf{W}}(r) = O(r^{-a})$ for some $a > 6$, one can apply Lemma 6.1 with $d = \ell_n$ whenever $\ell_n\, h \to \infty$ in order to bound $\sum_{i=1}^{r_n} \mathrm{E}(|W_{ni}|^4)$. Hence, one concludes that there exists a constant C_α such that

$$\mathbb{P}(|W_n| > \kappa) \leq \frac{C_\alpha\, r_n^4\, \ell_n}{\kappa^4\, n} \to 0 \quad \text{whenever } 3\epsilon < 4\delta. \tag{12}$$

In order to deal with U_n, let $U'_{n1}, \ldots, U'_{nr_n}$ be independent random variables having the same conditional distribution function as $U_{n1}, \ldots, U_{nr_n}$. Based on Billingsley (1968), p. 376, one can show that the respective characteristic functions of U_n and $U'_n = \sqrt{nh} \sum_{i=1}^{r_n} U'_{n,i}$ differ by at most $16\, r_n\, \alpha_{\ell_n} \to 0$ whenever $(a+1)\epsilon < a + \delta$. As a consequence, U_n and U'_n are asymptotically equivalent. By straightforward computations,

$$\mathrm{Var}\left(\sqrt{nh} \sum_{i=1}^{r_n} U'_{ni}\right) = K_5\, \sigma_x^2(y_1, y_2) + o(1).$$

Next, using Lemma 6.1 again with $d = b_n$ such that $b_n\, h \to \infty$,

$$\begin{aligned} \sum_{i=1}^{r_n} \mathrm{E}\left\{\sqrt{nh}\, (U'_{ni})^2 \mathbb{I}\left(\sqrt{nh} (U'_{ni})^2 > \kappa\right)\right\} &\leq \sum_{i=1}^{r_n} \left(\frac{nh}{\kappa}\right)^2 \mathrm{E}\left\{(U'_{ni})^4\right\} \\ &= O\left(\frac{b_n}{n}\right). \end{aligned}$$

This last expression is $o(1)$ as long as $\epsilon > \delta$. Thus if τ is such that $h \sim n^{-\tau}$, the choice of $\epsilon = 4/5 \min\{a/(a+1), 1 - \tau\}$ and $\delta = 4/5\epsilon$ leads the Lyapunov ratio to converge to 0.

Asymptotic tightness. For $\mathbf{y} = (y_1, y_2)$ and $\mathbf{y}' = (y'_1, y'_2)$ in $\mathbb{R}^2$ and a fixed $x \in \mathbb{R}$, consider the semi-metric

$$\rho(\mathbf{y}, \mathbf{y}') = |F_{1x}(y_1) - F_{1x}(y'_1)| + |F_{2x}(y_2) - F_{2x}(y'_2)|.$$

Now for a bounded function $f : \mathbb{R}^2 \to \mathbb{R}$ and a subset T of $\mathbb{R}^2$, let

$$\mathfrak{W}_\delta(f, T) = \sup_{\substack{\mathbf{y},\mathbf{y}' \in T; \\ \rho(\mathbf{y},\mathbf{y}')<\delta}} |f(\mathbf{y}) - f(\mathbf{y}')| \,, \quad \text{where } \delta > 0.$$

Hence, $\mathfrak{W}_\cdot(Z_{xn}, \mathbb{R}^2)$ corresponds to the modulus of ρ-continuity of Z_{xn}. Now for $\gamma \in (0, 1/2)$ and $\kappa_{n,\gamma} = \lfloor (nh)^{1/2+\gamma} \rfloor$, define for each $j \in \{1, 2\}$ the grid

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

and let $T_{n,\gamma} = I_{n,\gamma}^{(1)} \times I_{n,\gamma}^{(2)}$. For $\mathbf{y} = (y_1, y_2) \in \mathbb{R}^2$, define $\underline{y}_{j,\gamma} = \max\{z \in I_{n,\gamma}^{(j)} : z \leq y_j\}$ and $\overline{y}_{j,\gamma} = \min\{z \in I_{n,\gamma}^{(j)} : z \geq y_j\}$, so that $F_{jx}(\overline{y}_{j,\gamma} r) - F_{jx}(\underline{y}_{j,\gamma}) \leq 1/\kappa_{n,\gamma}$. Now observe that for any $\mathbf{y} \in \mathbb{R}^2$,

$$\begin{aligned} Z_{xn}(\mathbf{y}) - Z_{xn}(\underline{\mathbf{y}}_\gamma) &\leq Z_{xn}(\overline{\mathbf{y}}_\gamma) - Z_{xn}(\underline{\mathbf{y}}_\gamma) \\ &\quad + 2\sqrt{nh} \sum_{i=1}^n w_{hi}(x) \left\{ H_{x_i}(\overline{\mathbf{y}}_\gamma) - H_{x_i}(\underline{\mathbf{y}}_\gamma) \right\} \\ &= Z_{xn}(\overline{\mathbf{y}}_\gamma) - Z_{xn}(\underline{\mathbf{y}}_\gamma) + 2\sqrt{nh}\, \rho\left(\overline{\mathbf{y}}_\gamma, \underline{\mathbf{y}}_\gamma\right) + o(1). \end{aligned}$$

Indeed, the assumption $H \in \mathcal{F}_x$ allows the Taylor expansion
$\sqrt{nh} \sum_{i=1}^n w_{hi}(x) \left\{ H_{x_i}(\overline{\mathbf{y}}_\gamma) - H_{x_i}(\underline{\mathbf{y}}_\gamma) \right\}$

$$\begin{aligned} &= \sqrt{nh}[H_x(\overline{\mathbf{y}}_\gamma) - H_x(\underline{\mathbf{y}}_\gamma)] + [\dot{H}_x(\overline{\mathbf{y}}_\gamma) - \dot{H}_x(\underline{\mathbf{y}}_\gamma)]\sqrt{nh} \sum_{i=1}^n w_{hi}(x)\{x_i - x\} \\ &\quad + \sqrt{nh} \sum_{i=1}^n w_{hi}(x)\{x_i - x\}^2 [\ddot{H}_{z_i}(\overline{\mathbf{y}}_\gamma) - \ddot{H}_{z_i}(\underline{\mathbf{y}}_\gamma)] \end{aligned}$$

where z_i lies between x_i and x. From Conditions W_2–W_3 and the fact that $H \in \mathcal{F}_x$, the latter is bounded by

$$\sqrt{nh}\, \rho\left(\overline{\mathbf{y}}_\gamma, \underline{\mathbf{y}}_\gamma\right) + nh^5 \left(K_2 + K_3\right) \times o(1),$$

where one uses the general inequality on distribution functions

$$|H_x(\mathbf{y}) - H_x(\mathbf{y}')| \leq |F_{1x}(y_1) - F_{1x}(y_2')| + |F_{2x}(y_2) - F_{2x}(y_2')| \,.$$

As a consequence, uniformly in $\mathbf{y} \in \mathbb{R}^2$,

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

From similar arguments, one deduces

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

Thus for any $\mathbf{y}, \mathbf{z} \in \mathbb{R}^2$,

$$\begin{aligned} |Z_{xn}(\mathbf{y}) - Z_{xn}(\mathbf{z})| \leq \; & \left|Z_{xn}(\overline{\mathbf{y}}_\gamma) - Z_{xn}(\underline{\mathbf{y}}_\gamma)\right| + \left|Z_{xn}(\overline{\mathbf{z}}_\gamma) - Z_{xn}(\underline{\mathbf{z}}_\gamma)\right| \\ & + \left|Z_{xn}(\underline{\mathbf{y}}_\gamma) - Z_{xn}(\underline{\mathbf{z}}_\gamma)\right|. \end{aligned}$$

Since for n sufficiently large, $\rho(\mathbf{y}, \mathbf{z}) < \delta$ implies $\rho(\underline{\mathbf{y}}_\gamma, \underline{\mathbf{z}}_\gamma) < 2\delta$, it follows that $\mathfrak{W}_\delta(Z_{xn}, \mathbb{R}^2) \leq 3\, \mathfrak{W}_{2\delta}(Z_{xn}, T_n)$. It remains to show that for any positive sequence $\delta_n \downarrow 0$ and any $\epsilon > 0$,

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

Observe that $\mathfrak{W}_{\delta_n}(Z_{xn}, T_n) = 0$ whenever $\delta_n < 2\kappa_{n,\gamma}^{-1}$, while $\mathfrak{W}_{\delta_n}(Z_{xn}, T_n) \leq \mathfrak{W}_{2\kappa_{n,\gamma}}(Z_{xn}, T_n)$ otherwise. Then, for any $1 \leq i, j \leq \kappa_{n,\gamma}$, define the intervals

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

One can then conclude that

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

where for an arbitrary nonempty rectangle $A \in \mathbb{R}^2$,

$$\mathbb{H}_{xh}(A) = \sqrt{nh} \sum_{i=1}^n w_{hi}(x) \left\{\mathbb{I}\left((Y_{1i}, Y_{2i}) \in A\right) - \nu_{x_i}(A)\right\},$$

with $\nu_x(A) = \mathbb{P}(\mathbf{Y}_x \in A)$, $\mathbf{Y}_x \sim H_x$. Define $\mu_x = \nu_x \otimes \lambda$, where λ denotes the ρ-measure of A. In order to derive an extension of Theorem 3 of Bickel & Wichura (1971), one needs to find $\beta > 1$ and $C \in \mathbb{R}$ (that may depend on ϵ and β) such that

$$\mathbb{P}\left(|\mathbb{H}_{xh}(A_{n,\gamma}(i,j))| \geq \epsilon\right) \leq C\mu_x(A_{n,\gamma}(i,j))^\beta.$$

The following Lemma provides a moment inequality for $\mathbb{H}_{xh}(A)$.

Lemma 6.2. *Let $(\mathbf{W})_{i\in\mathbb{Z}}$, where $\mathbf{W}_i = (Y_{1i}, Y_{2i}, X_i)$, be a stationary sequence with α-mixing coefficients satisfying $\alpha_{\mathbf{W}}(r) = O(r^{-a})$ for some $a > 6$ and distribution function $H \in \mathcal{F}_x$. Suppose the bandwidth parameter h is such that $nh^{1+r_a} < \infty$, where r_a denotes the largest even integer such that $r_a < a$. Moreover, assume Conditions $W_1^\star$, W_2–W_5 and W_{11}–W_{13} are satisfied. Then for any rectangle $A \subseteq \mathbb{R}^2$ such that $\mu_x(A) \geq (nh)^{-1-\delta}$ for some positive $\delta < 1$, there exists $C_\alpha(a, x) < \infty$ such that*

$$\begin{aligned}
\mathrm{E}\left\{\mathbb{H}_{xh}(A)^{r_a}\right\} &\leq \frac{C_\alpha(r_a, x)}{(nh)^{r_a/2}} \sum_{p=1}^{r_a/2} \{\mu_x(A) + h^2\}^p (nh)^p \\
&\leq \frac{C_\alpha(a, x)}{(nh)^{r_a/2}} \left\{ \sum_{p=1}^{r_a/2} n^p h^{3p} \sum_{k=1}^{p} \binom{p}{k} \mu_x(A)^k h^{-2k} + 1 \right\}.
\end{aligned}$$

Starting from the Markov inequality,

$$\mathbb{P}\left(|\mathbb{H}_{xh}(A_{n,\gamma}(i,j))| \geq \epsilon\right) \leq \epsilon^{-r_a} \mathrm{E}\left\{\mathbb{H}_{xh}(A_{n,\gamma}(i,j))^{r_a}\right\}.$$

Next note that for any $1 \leq i, j \leq \kappa_{n,\gamma}$, the rectangle $A = A_{n,\gamma}(i,j)$ satisfies

$$(nh)^{-1-2\gamma} \leq \mu_x(A) \leq (nh)^{-1/2-\gamma}.$$

Upon taking $\beta < 1/4(r_a + 2)$ and γ such that $n^{2\gamma} h < 1$ in Lemma 6.2, this yields for any $p < \beta$ and n sufficiently large,

$$\sum_{k=1}^{p} \binom{p}{k} \mu_x(A)^{k-\beta} h^{2p-2k} (nh)^{-r_a/2+p} < 1.$$

Moreover, if $\beta(1 + 2\gamma) < r_a/2$, then the same holds true when $p > \beta$. Thus, if $\beta < \min\{(r_a + 2)/4, r_a/2(1 + 2\gamma)\}$ and for a suitable choice of γ,

$$\sum_{p=1}^{r_a/2} \sum_{k=1}^{p} \binom{p}{k} \mu_x(A)^{k-\beta} h^{2(p-k)} (nh)^{-r_a/2+p} \leq \frac{r_a}{2}.$$

Finally,

$$\begin{aligned}
\sum_{p=1}^{r_a/2} h^{2p} \mu_x(A)^{-\beta} (nh)^{-r_a/2+p} &\leq \max_{p\in\{1,\ldots,r_a/2\}} \left\{ h^{2p} (nh)^{(1+2\gamma)\beta} (nh)^{-r_a/2+p} \right\} \\
&\sim n^{-2p\tau + (1+2\gamma)(1-\tau)\beta - (1-\tau)(r_a/2-p)} \\
&\sim n^{p(1-3\tau) + (1+2\gamma)(1-\tau)\beta - (1-\tau)r_a/2}.
\end{aligned}$$

If $\tau < 1/3$, the maximum is attained at $p = r_a/2$, which yields the bound $\beta < r_a/4$ to ensure the negligibility of the previous display since $\tau > 1/5$. Else, if $\tau > 1/3$ the maximum is at $p = 1$ and imposes $\beta < r_a/2$. Since $a > 6$, $r_a \geq 6$ and all above quantities allows a choice of β greater than 1. As a consequence one can find a constant C that depends on $\epsilon, \beta, r_a, \gamma$ and τ such that

$$\mathbb{P}\left(|\mathbb{H}_{xh}(A_{n,\gamma}(i,j))| \geq \epsilon\right) \leq C\mu_x(A_{n,\gamma}(i,j))^\beta.$$

The asymptotic ρ-equicontinuity follows from an application of an extension to Theorem 3 in Bickel & Wichura (1971) or Lemma 2 in Balacheff & Dupont (1980). Since the sequences $Z_{xn}(\mathbf{y})$ are asymptotically tight in $\mathbb{R}^2$ and because $\mathbb{R}^2$ is totally bounded for ρ, the tightness of the sequence Z_{xn} in $\ell^\infty(\mathbb{R}^2)$ can be deduced from Theorem 1.5.7 in van der Vaart & Wellner (1996), which completes the proof.

6.2. Proof of Theorem 3.3

As pointed out by Veraverbeke et al. (2011), the asymptotic negligibility of $\sqrt{nh}\{\Phi(G_{xh}) - \Phi(\widetilde{G}_{xh})\}$ is closely related to the asymptotic behavior of the processes of pseudo-observations $\widetilde{Z}_{1xn} = Z_{1xn} - \mathrm{E}(Z_{1xn})$ and $\widetilde{Z}_{2xn} = Z_{2xn} - \mathrm{E}(Z_{2xn})$, where for $j = 1, 2$,

$$Z_{jxn}(t,u) = \sqrt{nh_j}\, F_{jz_th_j}\left\{F_{jz_t}^{-1}(u)\right\}, \quad z_t = x + tCh.$$

The key result is the following lemma whose proof is to be found in the Appendix B.

Lemma 6.3. *Let $(\mathbf{W}_i)_{i\in\mathbb{Z}}$, where $\mathbf{W}_i = (Y_{1i}, Y_{2i}, X_i)$, be a stationary sequence with associated α-mixing coefficients such that $\alpha_{\mathbf{W}}(r) = O(r^{-a})$ for some $a > 6$. Suppose that for $j = 1, 2$, the functions $F_{jz}\{F_{jz}^{-1}(u)\}$, $\dot{F}_{jz}\{F_{jz}^{-1}(u)\}$ and $\ddot{F}_{jz}\{F_{jz}^{-1}(u)\}$ are continuous in (z, u) for z in a neighborhood of x. Moreover assume $nh_1^5 < \infty$ and $nh_2^5 < \infty$. Finally, suppose assumptions $W_1^\star$, W_6–W_{13} are satisfied. Then the sequences $\widetilde{Z}_{1xn}$ and $\widetilde{Z}_{2xn}$ are asymptotically tight in $l^\infty([-1,1]\times[0,1])$.*

Since the assumptions in Lemma 6.3 are satisfied, one can conclude that $\widetilde{Z}_{1xn}$ and $\widetilde{Z}_{2xn}$ are asymptotically tight in the space $\ell^\infty([-1,1]\times[0,1])$. The asymptotic negligibility of $\sqrt{nh}\{\Phi(G_{xh}) - \Phi(\widetilde{G}_{xh})\}$ will then follow from similar arguments as in Appendix B.2 of Veraverbeke et al. (2011).

References

BALACHEFF, S. & DUPONT, G. (1980). Normalite asymptotique des processus empiriques tronques et des processus de rang. *Lecture Notes in Mathematics* , 19–45.

BICKEL, P. J. & WICHURA, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *The Annals of Mathematical Statistics* , 1656–1670.

BILLINGSLEY, P. (1968). *Convergence of probability measures.* New York: John Wiley & Sons Inc.

BOUEZMARNI, T., ROMBOUTS, J. & TAAMOUTI, A. (2012). A nonparametric copula based test for conditional independence with applications to granger causality. *Journal of Business & Economic Statistics* **30**, 275–287.

BÜCHER, A. & RUPPERT, M. (2013). Consistent testing for a constant copula under strong mixing based on the tapered block multiplier technique. *J. Multivariate Anal.* **116**, 208–229.

BÜCHER, A. & VOLGUSHEV, S. (2013). Empirical and sequential empirical copula processes under serial dependence. *J. Multivariate Anal.* .

CAMIRAND, F. (2013). *Test d'indépendance conditionnel basé sur la copule conditionnelle.* Master's thesis, Université de Sherbrooke.

CARRASCO, M. & CHEN, X. (2002). Mixing and moment properties of various garch and stochastic volatility models. *Econometric Theory* **18**, 17–39.

DEHEUVELS, P. (1979). La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. *Acad. Roy. Belg. Bull. Cl. Sci. (5)* **65**, 274–292.

DOUKHAN, P. (1994). Mixing: Properties and examples .

FERMANIAN, J.-D., RADULOVIĆ, D. & WEGKAMP, M. H. (2004). Weak convergence of empirical copula processes. *Bernoulli* **10**, 847–860.

GAENSSLER, P. & STUTE, W. (1987). *Seminar on empirical processes.* Basel: Birkhäuser Verlag.

GEWEKE, J. (1982). Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association* **77**, 304–313.

GEWEKE, J. (1984). Measures of conditional linear dependence and feedback between time series. *Journal of the American Statistical Association* **79**, 907–915.

GIJBELS, I., VERAVERBEKE, N. & OMELKA, M. (2011). Conditional copulas, association measures and their applications. *Comput. Statist. Data Anal.* **55**, 1919–1932.

GOURIÉROUX, C., MONFORT, A. & RENAULT, É. (1987). Kullback causality measures. *Annales dÉconomie et de Statistique* **6/7**, 369–410.

GRANGER, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–459.

LI, Q. & RACINE, J. S. (2007). *Nonparametric econometrics: Theory and practice*. Princeton University Press.

MEITZ, M. & SAIKKONEN, P. (2002). Ergodicity, mixing, and existence of moments of a class of markov models with applications to garch and acd models. *Econometric Theory* **24**, 1291–1320.

NELSEN, R. B. (2006). *An introduction to copulas*. Springer Series in Statistics. New York: Springer, 2nd ed.

OMELKA, M., VERAVERBEKE, N. & GIJBELS, I. (2013). Bootstrapping the conditional copula. *J. Statist. Plann. Inference* **143**, 1–23.

RIO, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants*. Mathématiques et Applications. Springer.

RÜSCHENDORF, L. (1976). Asymptotic distributions of multivariate rank order statistics. *Ann. Statist.* **4**, 912–923.

SEGERS, J. (2012). Weak convergence of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli* **18**, 764–782.

SKLAR, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8**, 229–231.

Su, L. & White, H. (2008). A nonparametric hellinger metric test for conditional independence. *Econometric Theory* **24**, 829–864.

Taamouti, A., Bouezmarni, T. & El Ghouch, A. (2014). Nonparametric estimation and inference for conditional density based granger causality measures. *Journal of Econometrics* **180**, 251–264.

van der Vaart, A. W. & Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. New York: Springer-Verlag. With applications to statistics.

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

Wiener, N. (1956). The theory of prediction. *in E. F. Beckenback, ed., The Theory of Prediction, McGraw-Hill, New York, chapter 8.* .

Appendix A. Conditions on the weights

The following conditions are needed in order to establish Proposition 2.3.

W_1. $\max_{1\le i\le n} |w_{hi}(x)| = o_{\mathbb{P}}\left((nh)^{-1/2}\right)$;

W_2. $\sum_{i=1}^n w_{hi}(x)(X_i - x) = h^2 K_3 + o_{\mathbb{P}}\left((nh)^{-1/2}\right)$ for some $K_3 < \infty$;

W_3. $\sum_{i=1}^n w_{hi}(x)(X_i - x)^2 = h^2 K_4 + o_{\mathbb{P}}\left((nh)^{-1/2}\right)$ for some $K_4 > 0$;

W_4. $nh \sum_{i=1}^n \{w_{hi}(x)\}^2 = K_5 + o_{\mathbb{P}}(1)$ for some $K_5 > 0$;

W_5. $\max_{i\in I_{nx}} X_i - \min_{i\in I_{nx}} X_i = o_{\mathbb{P}}(1)$, where $I_{nx} = \{j \in \{1,\dots,n\} : w_{hj}(x) > 0\}$.

Additionally, one has to assume the following conditions in Proposition 2.5.

W_6. $\sup_{z\in J_x^{(n)}} \sum_{i=1}^n |w'_{g_j i}(z)| = O_{\mathbb{P}}(g_j^{-1})$, where $J_x^{(n)} = [\min_{i\in I_{nx}} X_i, \max_{i\in I_{nx}} X_i]$;

W_7. $\sup_{z\in J_x^{(n)}} \sum_{i=1}^n \{w'_{g_j i}(z)\}^2 = O_{\mathbb{P}}(n^{-1} g_j^{-3})$;

W_8. For some finite constant C,

$$\mathbb{P}\left(\sup_{z\in J_x^{(n)}} \max_{1\le i\le n} \left|w_{g_j i}(z)\, \mathbb{I}\left(|x_i - x| > Ch > 0\right)\right|\right) = o_{\mathbb{P}}(1);$$

W_9. There exists $D_K < \infty$ such that for all a_n,

$$\sup_{z \in J_x^{(n)}} \left| \sum_{i=1}^{n} w_{a_n i}(z)(x_i - z) - a_n^2 D_k \right| = o_{\mathbb{P}}(a_n^2);$$

W_{10}. There exists $E_K < \infty$ such that for all a_n,

$$\sup_{z \in J_x^{(n)}} \left| \sum_{i=1}^{n} w_{a_n i}(z)(x_i - z)^2 - a_n^2 E_k \right| = o_{\mathbb{P}}(a_n^2).$$

The following version of condition W_1 is needed to calculate the variance of $\mathbb{H}_{xh}$:

$W_1^\star$. $\max_{1 \le i \le n} |w_{hi}(x)| = O_{\mathbb{P}}\left((nh)^{-1}\right)$.

In order to establish moment inequalities of order r, one needs that for any integer $1 \le k \le r$, any choice of $L_1, \ldots, L_k \in \mathbb{N}$ such that $L_1 + \ + L_k = r$, and for some positive sequence υ_n satisfying $n - \upsilon_n \to \infty$:

W_{11}. $$\sup_{z \in J_x} \max_{1 \le \ell_2 < \cdots < \ell_k \le \upsilon_n} \sum_{i=1}^{n-\ell_k} w_{hi}(x)^{L_1} \prod_{j=2}^{k} w_{h,i+\ell_j}(z)^{L_j} = O_{\mathbb{P}}\left(\frac{h^{k-1}}{(nh)^{r-1}}\right);$$

W_{12}. $$\sup_{z \in J_x} \sum_{i=1}^{n} w_{hi}(z)^{L_1} = O_{\mathbb{P}}\left((nh)^{-L_1+1}\right);$$

W_{13}. $$\sup_{z \in J_x} \max_{1 \le \ell_2 < \cdots < \ell_k \le \upsilon_n} \sum_{i=1}^{n-\ell_k} (x_i - z) w_{hi}(x)^{L_1} \prod_{j=2}^{k} w_{h,i+\ell_j}(z)^{L_j} = O_{\mathbb{P}}\left(\frac{h^{k+1}}{(nh)^{r-1}}\right).$$

Appendix B. Proofs of Lemma 6.1, Lemma 6.2 and Lemma 6.3

Appendix B.1. Proof of Lemma 6.1

Fix $y_1, y_2 \in \mathbb{R}$ and for simplicity let ϑ_i stand for $\vartheta_i(y_1, y_2)$. First decompose $S_d(k)^4 = \sum_{j=1}^{5} S_d^{(j)}(k)$ where

$S_d^{(1)}(k) = \sum_{i \in \mathcal{S}_n(k)} \vartheta_i^4 w_{hi}(x)^4$, $S_d^{(4)} = 6 \sum_{\substack{i_1, i_2, i_3 \in \in \mathcal{S}_n(k) \\ i_1 \ne i_2 \ne i_3}} \vartheta_{i_1}^2 \vartheta_{i_2} \vartheta_{i_3} w_{hi_1}(x)^2 w_{hi_2}(x) w_{hi_3}(x)$

$S_d^{(2)}(k) = 4 \sum_{\substack{i_1,i_2 \in \mathcal{S}_n(k)\\ i_1 \neq i_2}} \vartheta_{i_1}^3 \vartheta_{i_2} w_{hi_1}(x)^3 w_{hi_2}(x)$, $S_d^{(5)}(k) = \sum_{\substack{i_1 \cdot i_4 \in \mathcal{S}_n(k)\\ i_1 \neq \ldots \neq i_4}} \prod_{k=1}^4 \vartheta_{i_k} w_{hi_k}(x)$

$S_d^{(3)}(k) = 6 \sum_{\substack{i_1,i_2 \in \mathcal{S}_n(k)\\ i_1 \neq i_2}} \vartheta_{i_1}^2 \vartheta_{i_2}^2 w_{hi_1}(x)^2 w_{hi_2}(x)^2.$

First, in view of condition W_{12}, one can find a constant K_{12} such that $\sum_{i=1}^n w_{hi}(x)^4 \leq K_{12} n^{-3} h^{-3}$. Since $d > h^{-1}$, one writes

$$\sum_{k=1}^{r_n} \mathrm{E}\left\{S_d^{(1)}(k)\right\} \leq K_{12} n^{-3} h^{-3} \leq K_{12} n^{-3} h^{-2} d.$$

Next, split $S_d^{(2)}(k)$ in $S_d^{(2,<)}(k)$ and $S_d^{(2,>)}(k)$ according to the cases $i_1 < i_2$ and $i_1 > i_2$. Once again, for any $1 \leq \ell \leq d - i$, $|\mathrm{E}\{\vartheta_i^3 \vartheta_{i+\ell}\}| \leq \alpha_{\mathbf{W}}(\ell) \leq 1$. One writes for $\pi_n = \lfloor 1/\sqrt{h} \rfloor$

$$\begin{aligned}
&\sum_{k=1}^{r_n} \mathrm{E}\left\{S_d^{(2,<)}(k)\right\} \\
&\leq \sum_{i=1}^{n} \sum_{\ell=1}^{\pi_n} \mathrm{E}\left\{\vartheta_i^3 \vartheta_{i+\ell}\right\} w_{hi}(x)^3 w_{h,i+\ell}(x) + \sum_{i=1}^{n} \sum_{\ell=\pi_n+1}^{d-i} \mathrm{E}\left\{\vartheta_i^3 \vartheta_{i+\ell}\right\} w_{hi}(x)^3 w_{h,i+\ell}(x) \\
&\leq \pi_n \max_{1 \leq \ell \leq \pi_n} \sum_{i=1}^{n-\ell} w_{hi}(x)^3 w_{h,i+\ell}(x) + \sum_{i=1}^{n} w_{hi}(x)^3 \times \max_{1 \leq i \leq n} w_{hi}(x) \sum_{\ell=\pi_n+1}^{d} \alpha_{\mathbf{W}}(\ell) \\
&\sim \ o\left(n^{-3} h^{-2} d\right)
\end{aligned}$$

where the last equality follows from assumption $W_1^\star$, W_{11}-W_{12} together with the assumption on the mixing coefficients $\alpha_{\mathbf{W}}(r)$. Similarly,

$$\sum_{k=1}^{r_n} \mathrm{E}\, S_d^{(2,>)}(k) \sim o\left(n^{-3} h^{-2} d\right).$$

Next decompose $S_d^{(3)}(k) = A_d^{(3)}(k) + \{S_d^{(3)}(k) - A_d^{(3)}(k)\}$ where

$$\begin{aligned}
A_d^{(3)}(k) &= 6 \sum_{\substack{i_1,i_2 \in \mathcal{S}_n(k)\\ i_1 \neq i_2}} \mathrm{E}\{\vartheta_{i_1}^2\} \mathrm{E}\{\vartheta_{i_2}^2\} w_{hi_1}(x)^2 w_{hi_2}(x)^2 \\
&= 12 \sum_{\substack{i_1,i_2 \in \mathcal{S}_n(k)\\ i_1 < i_2}} \mathrm{E}\{\vartheta_{i_1}^2\} \mathrm{E}\{\vartheta_{i_2}^2\} w_{hi_1}(x)^2 w_{hi_2}(x)^2.
\end{aligned}$$

Taking care of the first term, with assumption W_{11} in mind, one writes

$$\begin{aligned}\sum_{k=1}^{r_n} A_d^{(3)}(k) &\leq \sum_{i=1}^{n} \sum_{\ell=1}^{d\wedge(n-i)} \mathrm{E}\{\vartheta_{i_1}^2\}\mathrm{E}\{\vartheta_{i+\ell}^2\} w_{hi_1}(x)^2 w_{h,i+\ell}(x)^2 \\ &\leq d \times \left\{ \max_{1\leq \ell \leq d} \sum_{i=1}^{n-\ell} w_{hi_1}(x)^2 w_{h,i+\ell}(x)^2 \right\} \sim O(dn^{-3}h^{-2}).\end{aligned}$$

Next, since $|\mathrm{E}\{\vartheta_i^2\vartheta_{i+\ell}^2\} - \mathrm{E}\{\vartheta_i^2\}\mathrm{E}\{\vartheta_{i+\ell}^2\}| \leq \alpha_{\mathbf{W}}(\ell)$,

$$\begin{aligned}\sum_{k=1}^{r_n} S_d^{(3)}(k) - A_d^{(3)}(k) &\leq \pi_n \sum_{i=1}^{n} w_{hi}(x)^2 w_{h,i+\ell}(x)^2 \\ &\quad + \left\{ \max_{1\leq i\leq n} w_{hi} \right\}^2 \left\{ \sum_{i=1}^{n} w_{hi}(x)^2 \right\} \sum_{\ell=\pi_n+1}^{d} \alpha_{\mathbf{W}}(\ell) \\ &\sim O(n^{-3}h^{-3/2}) + O(n^{-3}h^{-3})o(1) = o(n^{-3}h^{-2}d)\end{aligned}$$

where the last display is obtained from conditions $W_1^\star$,W_{11}–W_{12}, and the (assumed) boundedness of $\sum_{\ell=1}^{\infty} \alpha_{\mathbf{W}}(\ell)$. Using a similar strategy, with the help of those conditions and the fact that $\sum_{\ell=1}^{\infty}(\ell+1)\alpha_{\mathbf{W}}(\ell) < \infty$ one obtains

$$\sum_{k=1}^{r_n} S_d^{(4)}(k) \sim o(n^{-3}h^{-2}d).$$

Finally observe that

$$\mathrm{E}\, S_d^{(5)}(k) = 4! \sum_{\substack{i_1 \cdot i_4 \in \mathcal{S}_n(k) \\ i_1 < \ldots < i_4}} \mathrm{E}\left\{ \textstyle\prod_{j=1}^{4} \vartheta_{i_j} w_{hi_j}(x) \right\}.$$

Denote for each $j = 1, 2, 3$ the sets $\mathcal{S}_n^{(j)}(k) = \{i_1 < \ldots < i_4 \in S_n(k) : \max(g_1, g_2, g_3) \leq g_j\}$, where $g_j = i_{j+1} - i_j$ is the gap between two consecutive indices. A useful observation is that $\mathrm{E}\, S_d^{(5)}(k)$ is bounded by $A_d^{(5,1)}(k) + A_d^{(5,2)}(k) + A_d^{(5,3)}(k)$, where $A_d^{(5,j)}(k) = \sum_{\mathcal{S}_n^{(j)}(k)} \mathrm{E}\{\prod_{j=1}^{4} \vartheta_{i_j} w_{hi_j}(x)\}$.

Since $|\mathrm{E}\{\vartheta_i\vartheta_{i+g}\}| \leq \alpha_{\mathbf{W}}(g)$, one writes

$$\begin{aligned}\sum_{k=1}^{r_n} A_d^{(5,1)}(k) &\leq \pi_n \left\{ \max_{1\leq g_2,g_3\leq g_1\leq \pi_n} \sum_{i=1}^{n-g_1-g_2-g_3} w_{hi}(x) \prod_{k=1}^{3} w_{h,i+g_k}(x) \right\} \\ &\quad + \left\{ \max_{1\leq i\leq n} w_{hi}(x) \right\}^3 \sum_{g_1=\pi_n+1}^{d} (g_1+1)^2 \alpha_{\mathbf{W}}(g_1) \\ &= O(n^{-3}h^{-3/2}) + O(n^{-3}h^{-3})o(1) \sim o(n^{-3}h^{-2}d)\end{aligned}$$

where the last equality follows from assumption $W_1^\star$ and W_{11} together with the assumption over the finiteness of $\sum_{g_1=\pi_n+1}^{d}(g_1+1)^2\alpha_{\mathbf{W}}(g_1)$. Using the same conditions one deduces

$$\sum_{k=1}^{r_n} A_r^{(5,3)}(k) \sim o(n^{-3}h^{-2}d).$$

Finally, as for $S_d^{(3)}(k)$, decompose $A_d^{(5,2)}(k) = D_r^{(5,2)}(k) + \{A_r^{(5,2)}(k) - D_r^{(5,2)}(k)\}$ where

$$D_r^{(5,2)}(k) = \sum_{\mathcal{S}_n^{(2)}(k)} \mathrm{E}\{\vartheta_{i_1}\vartheta_{i_2}\}\mathrm{E}\{\vartheta_{i_3}\vartheta_{i_4}\} \prod_{j=1}^{4} w_{hi_j}(x).$$

Roughly,

$$\begin{aligned}\sum_{r=1}^{r_n} D_r^{(5,2)}(k) &\leq d \sum_{i=1}^{n} \sum_{\ell=1}^{d\wedge(n-i)} \mathrm{E}\left\{\vartheta_i\vartheta_{i+\ell}\right\} w_{hi}(x) w_{h,i+\ell}(x) \max_{1\leq i\leq n} w_{hi}(x)^2 \sum_{g_3=1}^{d} \alpha_{\mathbf{W}}(g_3) \\ &\leq \left\{ \max_{1\leq \ell\leq d} \sum_{i=1}^{n-\ell} w_{hi}(x) w_{h,i+\ell}(x) \right\} \times O(dn^{-2}h^{-2}) \\ &= O(dn^{-3}h^{-2})\end{aligned}$$

using $W_1^\star$, W_{11} and the assumption over the mixing coefficients $\alpha_{\mathbf{W}}(r)$. Moreover, for any $i_1 < i_2 < i_3 < i_4$, $|\mathrm{E}\{\vartheta_{i_1}\vartheta_{i_2}\vartheta_{i_3}\vartheta_{i_4}\} - \mathrm{E}\{\vartheta_{i_1}\vartheta_{i_2}\}\mathrm{E}\{\vartheta_{i_3}\vartheta_{i_4}\}| \leq \alpha_{\mathbf{W}}(g_2)$. Similar development as for $S_d^{(3)}(k) - A_d^{(3)}(k)$ leads

$$\sum_{r=1}^{r_n} \{A_d^{(5,2)}(k) - D_r^{(5,2)}(k)\} \sim o(n^{-3}h^{-2}d).$$

Wrapping up every terms of the decomposition of $\sum_{k=1}^{r_n} \mathrm{E}\{S_d(k)^4\}$, one concludes that there exist a finite constant C_α that depends on the coefficients $\alpha_{\mathbf{W}}(r)$ trough the sums $\sum_{\ell=1}^{\infty}(1+\ell)^j \alpha_{\mathbf{W}}(\ell)$ for $j = 1, 2, 3$ and on the choice of weight system through conditions $W_1^\star, W_{11}$ and W_{12}, such that

$$\sum_{k=1}^{r_n} \mathrm{E}\{S_d(k)^4\} \leq C_\alpha n^{-3} h^{-2} d.$$

Appendix B.2. Proof of Lemma 6.2

First note that since $H \in \mathcal{F}_x$, whenever x_i falls in a neighborhood of x,

$$\nu_{x_i}(A) = \nu_x(A) + \dot{\nu}_x(A)\{x_i - x\} + \ddot{\nu}_{z_i}(A)\{x_i - x\}^2 \tag{B.1}$$

where z_i is between x_i and x. For simplicity let ϑ_i stand for $\vartheta_i(A)$ and ν_z for $\nu_z(A)$. The proof begins with a proposition.

Proposition Appendix B.1. *Suppose Conditions $W_1^\star$, W_{11}–W_{13} are satisfied, and assume the mixing coefficients of the sequence $\mathbf{W}_i = (Y_{1i}, Y_{2i}, X_i)$ satisfy $\alpha_{\mathbf{W}}(r) = O(r^{-a})$ for some $a > 6$. For any integers $L_1, L_2 \geq 1$, write $S_n(L_1, L_2) = \sum_{\ell=1}^{n-1} \sum_{i=1}^{n-\ell} \vartheta_i^{L_1} \vartheta_{i+\ell}^{L_2} w_{hi}(x)^{L_1} w_{h,i+\ell}(x)^{L_2}$. Then there exist a constant $K_\alpha > 0$ such that*

$$\mathrm{E}\{S_n(L_1, L_2)\}$$

$$= K_\alpha \begin{cases} (nh)^{-l_1-l_2+1}\left[\{\nu_x + h^2\} + \pi_n^{-r_a} + nh\{\nu_x + h^2\}^2\right] & \text{if } \min(L_1, L_2) > 1 \\ (nh)^{-l_1-l_2+1}\left[\{\nu_x + h^2\} + \pi_n^{-r_a}\right] & \text{else }. \end{cases}$$

where $\pi_n = \lfloor h^{-1} \rfloor$.

Proof. First decompose $S_n(L_1, L_2) = \{S_n(L_1, L_2) - A_n(L_1, L_2)\} + A_n(L_1, L_2)$, where

$$A_n(L_1, L_2) = \sum_{\ell=1}^{n-1} \sum_{i=1}^{n-\ell} \mathbb{E}\left(\vartheta_i^{L_1}\right) \mathbb{E}\left(\vartheta_{i+\ell}^{L_2}\right) w_{hi}(x)^{L_1} w_{h,i+\ell}(x)^{L_2}.$$

Notice that on one side $|\mathrm{E}\{\vartheta_i^{L_1} \vartheta_{i+\ell}^{L_2}\} - \mathbb{E}\left(\vartheta_i^{L_1}\right) \mathbb{E}\left(\vartheta_{i+\ell}^{L_2}\right)| \leq \alpha_{\mathbf{W}}(\ell)$ while on the other side $|\mathrm{E}\{\vartheta_i^{L_1} \vartheta_{i+\ell}^{L_2}\} - \mathbb{E}\left(\vartheta_i^{L_1}\right) \mathbb{E}\left(\vartheta_{i+\ell}^{L_2}\right)| \leq \nu_{x_i}$. Set $\upsilon_n = \lfloor n/2 \rfloor$. For n sufficiently large, $\pi_n < \upsilon_n$ and one writes

$$
\begin{aligned}
&\mathrm{E}\{S_n(L_1,L_2)-A_n(L_1,L_2)\}\\
&\quad= \sum_{\ell=1}^{\pi_n}\sum_{i=1}^{n-\ell}[\mathrm{E}\{\vartheta_i^{L_1}\vartheta_{i+\ell}^{L_2}\}-\mathbb{E}\left(\vartheta_i^{L_1}\right)\mathbb{E}\left(\vartheta_{i+\ell}^{L_2}\right)]w_{hi}(x)^{L_1}w_{h,i+\ell}(x)^{L_2}\\
&\qquad+\sum_{\ell=\pi_n+1}^{n-1}\sum_{i=1}^{n-\ell}[\mathrm{E}\{\vartheta_i^{L_1}\vartheta_{i+\ell}^{L_2}\}-\mathbb{E}\left(\vartheta_i^{L_1}\right)\mathbb{E}\left(\vartheta_{i+\ell}^{L_2}\right)]w_{hi}(x)^{L_1}w_{h,i+\ell}(x)^{L_2}\\
&\quad\leq \pi_n\left\{\max_{1\leq\ell\leq\pi_n}\sum_{i=1}^{n-\ell}\nu_{x_i}w_{hi}(x)^{L_1}w_{h,i+\ell}(x)^{L_2}\right\}\\
&\qquad+\sum_{\ell=\pi_n+1}^{\upsilon_n}\alpha_{\mathbf{W}}(\ell)\left\{\max_{\pi_n+1\leq\ell\leq\upsilon_n}\sum_{i=1}^{n-\ell}w_{hi}(x)^{L_1}w_{h,i+\ell}(x)^{L_2}\right\}\\
&\qquad+\sum_{\upsilon_n}^{n-1}\alpha_{\mathbf{W}}(\ell)(\upsilon_n+1)\left\{\max_{1\leq i\leq n}w_{hi}(x)\right\}^{L_1+L_2}\\
&\quad= O\left((nh)^{-L_1-L_2-1}\{\pi_n h\{\nu_x+\epsilon_A h^2\}+\pi_n^{-r_a+1}h+\upsilon_n^{-r_a+2}h^{-1}\}\right)
\end{aligned}
$$

where $\epsilon_A = \dot{\nu}_x(A)+\ddot{\nu}_x(A)$ piked from equation (B.1) together with assumption W_{11}, W_{13} and $W_1^\star$, since $\sum_{\ell=\gamma_n}^n \alpha_{\mathbf{W}}(\ell)\sim O(\gamma_n^{r_a+1})$ whenever $\gamma_n\to\infty$. From $H\in\mathcal{F}_x$, ϵ_A remains uniformly bounded and can simply be ignored. From $\pi_n h\to 1$, there exist a constant C_α such that $\mathrm{E}\{S_n(L_1,L_2)-A_n(L_1,L_2)\}\leq C_\alpha(nh)^{-L_1-L_2-1}(\{\nu_x+h^2\}+\pi_n^{-r_a})$.

Next, notice that $A_n(1,L_2)=A_n(L_1,1)=0$ since $\mathrm{E}\{\vartheta_i\}=0$. Otherwise, using W_{12} and W_{13} coupled with equation (B.1), one deduces there exist a constant C' such that $A_n(L_1,L_2)\leq C'\{\nu_x+h^2\}^2(nh)^{-L_1-L_2}$. Setting $K_\alpha=\max\{C_\alpha,C'\}$ concludes the proof. ☐

Denote the set of indices

$$
\mathcal{B}_p^{(r)}=\left\{\ L_1,...,L_r\in\{1,...,p\}:L_1+...+L_r=p\ \right\}.
$$

In the following, using an induction type of argument one shows that for any $p\leq r_a$, for any $r\leq p$ and for any $(L_1,...,L_r)\in B_p^{(r)}$ there exist a constant $C(r,p)$ such that

$$
\begin{aligned}
&\textstyle\sum_{i_1<...<i_r}\mathbb{E}\left(\prod_{k=1}^r\vartheta_{i_k}^{L_k}w_{hi_k}(x)^{L_k}\right)\\
&\qquad\leq C(r,p)\left\{\sum_{k=1}^{r^\star}\{\mu_x+h^2\}^k(nh)^{-p+k}+\pi_n^{-r_a}(nh)^{-p+1}\right\}
\end{aligned}
\qquad\text{(B.2)}
$$

Step 1. The case $p = 2$, $p = 3$ and $p = 4$.

For $p = 2$ and $r = 1$, the set $B_2^{(1)}$ contains only one indice (2). Moreover since Condition W_{12} holds one can find a constant $C(1, 2)$ such that:

$$\begin{aligned}\sum_{i=1}^{n} \mathbb{E}\left(\vartheta_i^2\right) w_{hi}(x)^2 &\leq C(1,2)\{\nu_x + h^2\}(nh)^{-1} \\ &\leq C(1,2)\{\mu_x + h^2\}(nh)^{-1}.\end{aligned}$$

For $r = 2$, notice that the set $B_2^{(2)}$ only contains the pair $(1, 1)$. One can than use Proposition Appendix B.1 with $L_1 = L_2 = 1$ to find a constant $C(2, 2)$ such that

$$\begin{aligned}\sum_{i_1<i_2} \mathbb{E}\left(\vartheta_{i_1}(A)\vartheta_{i_2}(A)\right) w_{hi_1}(x) w_{hi_2}(x) &\leq \{\nu_x(A) + h^2\}(nh)^{-1} + \pi_n^{-r_a}(nh)^{-1} \\ &\leq \{\mu_x(A) + h^2\}(nh)^{-1} + \pi_n^{-r_a}(nh)^{-1}.\end{aligned}$$

The equation (B.2) is therefore true for $p = 2$. For $p = 3$ and $r = 1$, using similar arguments :

$$\sum_{i=1}^{n} \mathbb{E}\left(\vartheta_i(A)^3\right) w_{hi}(x)^2 \leq C(1,3)\{\mu_x(A) + h^2\}(nh)^{-2}$$

For $r = 2$, the set $B_3^{(2)}$ contains two pairs $(1, 2)$ and $(2, 1)$ which can be treated using twice Proposition (Appendix B.1) with $L_1 = 1$ and $L_2 = 2$, and $L_1 = 2$ and $L_2 = 1$ respectively. One therefore concludes that

$$\sum_{i_1<i_2} \mathbb{E}\left(\vartheta_{i_1}(A)\vartheta_{i_2}(A)^2\right) w_{hi_1}(x) w_{hi_2}(x)^2 \leq \frac{C(2,3)}{(nh^2)}\{\mu_x(A) + h^2 + \pi_n^{-r_a}\}$$

and

$$\sum_{i_1<i_2} \mathbb{E}\left(\vartheta_{i_1}(A)^2\vartheta_{i_2}(A)\right) w_{hi_1}(x) w_{hi_2}(x)^2 \leq \frac{C(2,3)}{(nh^2)}\{\mu_x(A) + h^2 + \pi_n^{-r_a}\}.$$

Finally the set $B_3^{(3)}$ contains only one pair $(1, 1, 1)$. Here and in the sequel denote $\ell_j = i_{j+1} - i_j$ and write $\sum_{i_1:i_k}^{\ell_j}$ for the sum over all the indices $i_1 < ... < i_k$ such that $\ell_a \leq \ell_j$ for $a = 1, ..., k - 1$. For $\upsilon_n = \lfloor n/3 \rfloor$, observe that

$\sum_{i_1<i_2<i_3} \mathbb{E}\left(\prod_{k=1}^3 \vartheta_{i_k} w_{hi_k}(x)\right) \leq \sum_{a=1}^2 \sum_{i_1:i_3}^{\ell_a} \mathbb{E}\left(\prod_{k=1}^3 \vartheta_{i_k} w_{hi_k}(x)\right)$
$\leq \sum_{a=1}^2 \left\{\pi_n\{\nu_x + h^2\} + \sum_{\ell_a=\pi_n+1}^{\upsilon} (\ell_a+1)\alpha_{\mathbf{W}}(\ell_a)\right\}$

$$\times \max_{\ell_1,\ell_2\leq \upsilon_n} \sum_{i=1}^{n-\ell_1-\ell_2} w_{hi}(x) w_{h,i+\ell_1}(x) w_{h,i+\ell_2}(x) + \sum_{\ell_a=\upsilon_n}^{n-1} (\ell_a+1)\alpha_{\mathbf{W}}(\ell_a) \max_{1\leq i\leq n} w_{hi}(x)^3$$
$$\leq C(3,3)(nh)^{-2}\left\{\{\mu_x(A)+h^2\} + \pi_n^{-r_a}\right\}$$

where the last bound comes from assumption W_{11}-W_{13} combined with equation (B.1), and the fact that $\sum_{\ell=\gamma_n}^{n-1} (\ell+1)\alpha_{\mathbf{W}}(\ell_a) = O(\gamma_n^{-r_a+2})$ when $\gamma_n \to \infty$. Equation (B.2) is therefore true for $p = 3$. For $p = 4$ and $r = 1, 3, 4$ one proves Equation (B.2) similarly. The case $r = 2$ is different. In that case the set $B_4^{(2)}$ only contains the pair $(2, 2)$. Using Proposition Appendix B.1 with $L_1 = 2$ and $L_2 = 2$:

$\sum_{i_1<i_2} \mathbb{E}\left(\vartheta_{i_1}^2 \vartheta_{i_2}^2\right) w_{hi_1}(x)^2 w_{hi_2}(x)^2$

$$\leq C(2,4)(nh)^{-3}\left\{[\nu_x + h^2] + \pi_n^{-r_a} + (nh)\{\nu_x+h^2\}^2\right\}$$
$$\leq C(2,4)\sum_{k=1}^{2}\{\mu_x+h^2\}^k (nh)^{-4+k} + (nh)^{-3}\pi_n^{-r_a}$$

which completes to show (B.2) for p=4.

Step 2. Induction hypothesis.

Suppose for any $r \leq p-1$ and for any $(L_1, ..., L_r) \in B_{p-1}^{(r)}(q)$ equation (B.2) is satisfied.

Step 3. We now show that the result is still true for any $r \leq p$ and for any $(L_1, ..., L_r) \in B_p^{(r)}(q)$.

Once again one splits the sum according to which gap is the largest which leads to consider

$$\sum_{i_1<\ldots<i_r} \mathrm{E}\,(\prod_{k=1}^{r} \vartheta_{i_k}^{L_k} w_{hi_k}(x)^{L_k}) = \sum_{g=1}^{r-1}\{T_{1n}(g) + T_{2n}(g)\}$$

where

$$T_{1n}(g) = \sum_{i_1:i_r}^{\ell_g} \mathrm{E}\{\prod_{k=1}^{r} \vartheta_{i_k}^{L_k} w_{hi_k}(x)^{L_k}\} - T_{2n}(g)$$
$$T_{2n}(g) = \sum_{i_1:i_r}^{\ell_g} \mathrm{E}\{\prod_{k=1}^{g} \vartheta_{i_k}^{L_k} w_{hi_k}(x)^{L_k}\}\mathrm{E}\{\prod_{k=g+1}^{r} \vartheta_{i_k}^{L_k} w_{hi_k}(x)^{L_k}\}.$$

Dealing with $T_{1n}(g)$ first, using $W_1^\star$ and W_{11} together with equation (B.1), for $\upsilon_n = n - (nh)^{1-\delta}$ (where δ is defined in the statement of the lemma):

$$T_{1n}(g) \le \left\{\pi_n\{\nu_x + h^2\} + \textstyle\sum_{\ell=\pi_n+1}^{\upsilon_n}(\ell+1)^{r-2}\alpha_{\mathbf{W}}(\ell)\right\}$$
$$\times \max_{1\le\ell_2<\ell_3<\ldots<\ell_r\le\upsilon_n} \sum_{i=1}^{n-\ell_r} w_{hi}(x)^{L_1} \prod_{k=2}^{r} w_{hi+\ell_k}(x)^{L_k}$$
$$+ \sum_{\ell=\upsilon_n+1}^{n-1} (\ell+1)^{r-2}\alpha_{\mathbf{W}}(\ell) \times (n-\upsilon_n)\left\{\max_{1\le i\le n} w_{hi}(x)\right\}^p$$
$$\sim (nh)^{-p+1}\left\{\{h^{r-1}\pi_n\mu_x + h^2\} + h^{r-1}\pi_n^{-r_a+r-1} + (nh)^{1-\delta}(nh)^{-1}\upsilon_n^{-r_a+r-1}\right\}$$
$$\le (nh)^{-p+1}\left\{\{\mu_x + h^2\} + \pi_n^{-r_a}\right\}$$

where the last equality follows from the fact that $\pi_n h \sim O(1)$ and $\mu_x \ge (nh)^{-1-\delta}$. Next write $\bar{l}_g = \sum_{i=1}^{g} l_g$. Using the induction hypothesis twice for $(r,p) = (g, \bar{l}_g)$ and $(r,p) = (r-g, p-\bar{l}_g)$ respectively, one has

$$T_{2n}(g) \sim \left\{\textstyle\sum_{k=1}^{g^\star}\{\mu_x + h^2\}^k (nh)^{-\bar{l}_g+k} + \pi_n^{-r_a}(nh)^{-\bar{l}_g+1}\right\}$$
$$\times \left\{\sum_{k=1}^{(r-g)^\star}\{\mu_x + h^2\}^k (nh)^{-(p-\bar{l}_g)+k} + \pi_n^{-r_a}(nh)^{-(p-\bar{l}_g)}\right\}$$
$$\sim \sum_{k=2}^{r^\star}\{\mu_x + h^2\}^k (nh)^{-p+k} + \pi_n^{-r_a}(nh)^{-(p-\bar{l}_g)+1}\sum_{k=1}^{g^\star}\{\mu_x + h^2\}^k (nh)^{-\bar{l}_g+k}$$
$$+ \pi_n^{-r_a}(nh)^{-\bar{l}_g+1}\sum_{k=1}^{(r-g)^\star}\{\mu_x + h^2\}^k (nh)^{-(p-\bar{l}_g)+k} + \pi_n^{-2r_a}(nh)^{-p+2}$$
$$\sim \sum_{k=1}^{r^\star}\{\mu_x + h^2\}^k (nh)^{-p+k} + \pi_n^{-r_a}(nh)^{-p+1}$$

since $nh^{1+r_a} < \infty$ which implies $\pi_n^{-r_a} < (nh)^{-1}$. Thus there exist a constant $C(r,p)$ such that

$$T_{1n}(g) + T_{2n}(g) \le C(r,p) \left\{ \sum_{k=1}^{r^\star} \{\mu_x(A) + h^2\}^k (nh)^{-p+k} + \pi_n^{-r_a}(nh)^{-p+1} \right\}.$$

Equation (B.2) is therefore satisfy.This completes Step 3. The proof follows from the decomposition

$$\mathbb{E}\left(\mathbb{H}_{xh}(A)^{r_a}\right) = (nh)^{r_a/2} \sum_{r=1}^{r_a} r! \sum_{(l_1,\ldots,l_r)\in B_{r_a}^{(r)}} \sum_{i_1<\ldots<i_r} \mathbb{E}\left(\prod_{k=1}^{r} \vartheta_{i_k}(A)^{l_k} w_{hi}(x)^{l_k} \right).$$

Since $\mu_x + h^2 \ge h^2$ one can integrate the term $\pi_n^{-r_a}(nh)^{-p+1}$ into the first term at the cost of perhaps enlarging the constant. This completes the proof.

Appendix B.3. Proof of Lemma 6.3

For any fixed $(t,u) \in \mathcal{I} = [-1,1] \times [0,1]$ the asymptotic normality of the random variable $\widetilde{Z}_{jxn}(t,u)$ follows from similar arguments as in the proof of Theorem 3.1. This implies the asymptotic tightness of the random variable $\widetilde{Z}_{jxn}(t,u)$ in $\mathbb{R}$. It remains to show the asymptotic tightness of the sequence $\widetilde{Z}_{jxn}$ in $l^\infty(\mathcal{I})$. To this end, let $\rho(t,u,t',u') = |t-t'| + |u-u'|$ and for a bounded function $f : \mathcal{I} \to \mathbb{R}$ and a subset T of $\mathcal{I}$, define

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

We know show the asymptotic ρ-equicontinuity of $\widetilde{Z}_{jxn}$ i.e for any $\epsilon > 0$ there is a $\delta > 0$ such that

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

As in the proof of 3.1, for $\kappa_{n,\gamma}^{(j)} = \lfloor (nh_j)^{1/2+\gamma} \rfloor$ define grids

$$I_{n,\gamma}^{(1)} = \left\{ 0, \pm\frac{1}{\kappa_{n,\gamma}^{(j)}}, \ldots, \pm\frac{\kappa_{n,\gamma}^{(j)} - 1}{\kappa_{n,\gamma}^{(j)}}, \pm 1 \right\} \qquad I_{n,\gamma}^{(2)} = \left\{ 0, \frac{1}{\kappa_{n,\gamma}^{(j)}}, \ldots, \frac{\kappa_{n,\gamma}^{(j)} - 1}{\kappa_{n,\gamma}^{(j)}}, 1 \right\}$$

where $\gamma \in (0, 1/2)$ is a grid parameter to be fixed later, and set $T_n = I^{(1)}_{n,\gamma} \times I^{(2)}_{n,\gamma}$. For any $(t, u) \in [-1, 1] \times [0, 1]$, define $(\underline{t}_\gamma, \underline{u}_\gamma)$ and $(\overline{t}_\gamma, \overline{u}_\gamma)$ as in Section 6.1. Analogously to that section observe that

$$\begin{aligned}
\widetilde{Z}_{jxn}(t,u)-\widetilde{Z}_{jxn}(\underline{t}_\gamma,\underline{u}_\gamma) &= \{\widetilde{Z}_{jxn}(t,u)-\widetilde{Z}_{jxn}(t,\underline{u}_\gamma)\}+\{\widetilde{Z}_{jxn}(t,\underline{u}_\gamma)-\widetilde{Z}_{jxn}(\underline{t}_\gamma,\underline{u}_\gamma)\} \\
&\le \{\widetilde{Z}_{jxn}(t,\overline{u}_\gamma)-\widetilde{Z}_{jxn}(t,\underline{u}_\gamma)\}+\{\widetilde{Z}_{jxn}(t,\underline{u}_\gamma)-\widetilde{Z}_{jxn}(\underline{t}_\gamma,\underline{u}_\gamma)\} \\
&\quad +\mathrm{E}\{Z_{jxn}(t,\overline{u}_\gamma)-Z_{jxn}(t,\underline{u}_\gamma)\}. \qquad \text{(B.3)}
\end{aligned}$$

Starting with the last term, using a Taylor expansion of F_{jx_i} around z_t:

$$\begin{aligned}
\mathrm{E}\{Z_{jxn}(t,\overline{u}_\gamma)-Z_{jxn}(t,\underline{u}_\gamma)\} &= \sqrt{nh_j}\{\overline{u}_\gamma-\underline{u}_\gamma\} \\
&+\{\dot{F}_{jz_t}(F^{-1}_{jz_t}(\overline{u}_\gamma))-\dot{F}_{jz_t}(F^{-1}_{jz_t}(\underline{u}_\gamma))\}\left[\sqrt{nh_{jn}}\sum_{i=1}^n w_{h_j i}(z_t)(x_i-z_t)\right] \\
&+1/2\sqrt{nh_j}\sum_{i=1}^n\{\ddot{F}_{jr_{ti}}(F^{-1}_{jz_t}(\overline{u}_\gamma))-\ddot{F}_{jr_{ti}}(F^{-1}_{jz_t}(\underline{u}_\gamma))\}w_{h_j i}(z_t)(x_i-z_t)^2
\end{aligned}$$

where r_{ti} lies between z_t and x_i. Assumptions W_9 and W_{10} together with the (assumed) uniform continuity of the functions $\dot{F}_{jz}(F^{-1}_{jz})$ and $\ddot{F}_{jz}(F^{-1}_{jz})$ yields

$$\{\dot{F}_{jz_t}(F^{-1}_{jz_t}(\overline{u}_\gamma))-\dot{F}_{jz_t}(F^{-1}_{jz_t}(\underline{u}_\gamma))\}\sqrt{nh_j}\sum_{i=1}^n w_{h_j i}(z_t)(x_i-z_t)=o(1)O(\sqrt{nh_j}h_j^2)$$

and

$$\sqrt{nh_j}\sum_{i=1}^n\{\ddot{F}_{jr_{ti}}(F^{-1}_{jz_t}(\overline{u}_\gamma))-\ddot{F}_{jr_{ti}}(F^{-1}_{jz_t}(\underline{u}_\gamma))\}w_{h_j i}(z_t)(x_i-z_t)^2=o(1)O(\sqrt{nh_j}h_j^2).$$

From the boundedness of $\sqrt{nh_j}h_j^2$, one deduces the negligibility of the last two equations. From the grid definition, $\sqrt{nh_j}\{\overline{u}_\gamma-\underline{u}_\gamma\} = O((nh_j)^{-\gamma})$, which leads to the neglibility of $\mathrm{E}\{Z_{jxn}(t,\overline{u}_\gamma)-Z_{jxn}(t,\underline{u}_\gamma)\}$.

Next we deal with the term $\widetilde{Z}_{jxn}(t,\underline{u}_\gamma)-\widetilde{Z}_{jxn}(\underline{t}_\gamma,\underline{u}_\gamma)$ in equation B.3. Denote $\mathbb{F}_{jzh_j} = \sqrt{nh_j}\{F_{jzh_j} - \mathrm{E}\,F_{jzh_j}\}$ and notice that $\mathbb{F}_{jzh_j}\circ F^{-1}_{jz} = \widetilde{Z}_{jxn}$. One writes

$$\begin{aligned}
\widetilde{Z}_{jxn}(t,\underline{u}_\gamma)-\widetilde{Z}_{jxn}(\underline{t}_\gamma,\underline{u}_\gamma) &= \left[\mathbb{F}_{jz_th_j}\{F^{-1}_{jz_t}(\underline{u}_\gamma)\}-\mathbb{F}_{jz_th_j}\{F^{-1}_{jz_{\underline{t}_\gamma}}(\underline{u}_\gamma)\}\right] \\
&+\left[\mathbb{F}_{jz_th_j}\{F^{-1}_{jz_{\underline{t}_\gamma}}(\underline{u}_\gamma)\}-\mathbb{F}_{jz_{\underline{t}_\gamma}h_j}\{F^{-1}_{jz_{\underline{t}_\gamma}}(\underline{u}_\gamma)\}\right].
\end{aligned}$$

In view of assumption W_6 and the fact that $z_t - z_{\underline{t}_\gamma} = Ch(t - \underline{t}_\gamma)$, for any $y \in \mathbb{R}$:

$$
\begin{aligned}
\sqrt{nh_j}|F_{jz_th_j}(y) - F_{jz_{\underline{t}_\gamma}h_j}(y)| &= \sqrt{nh_j}\left|\sum_{i=1}^n \mathbb{I}\{Y_{ji} \le y\}[\,w_{h_ji}(z_t) - w_{h_ji}(z_{\underline{t}_\gamma})\,]\right| \\
&\le \sqrt{nh_j}\left\{\sup_{z\in I_x}\sum_{i=1}^n |w'_{hi}(z)|\right\} \times h(t - \underline{t}_\gamma) \\
&= O\left((nh_j)^{-\gamma}h_j^{-1}h\right).
\end{aligned}
$$

Since $h/\min(h_1,h_2) < \infty$ the latter is o(1) uniformly in y. From similar arguments one deduces that $\sup_{y,t}|\mathbb{F}_{jz_th_j}(y) - \mathbb{F}_{jz_{\underline{t}_\gamma}h_j}(y)| = o(1)$. It follows that

$$
\begin{aligned}
\widetilde{Z}_{jxn}(t,\underline{u}_\gamma) - \widetilde{Z}_{jxn}(\underline{t}_\gamma,\underline{u}_\gamma) &= \mathbb{F}_{jz_th_j}\{F^{-1}_{jz_t}(\underline{u}_\gamma)\} - \mathbb{F}_{jz_th_j}\{F^{-1}_{jz_{\underline{t}_\gamma}}(\underline{u}_\gamma)\} + o(1) \\
&= \mathbb{F}_{jz_{\underline{t}_\gamma}h_j}\{F^{-1}_{jz_t}(\underline{u}_\gamma)\} - \mathbb{F}_{jz_{\underline{t}_\gamma}h_j}\{F^{-1}_{jz_{\underline{t}_\gamma}}(\underline{u}_\gamma)\} + o(1).
\end{aligned}
$$

Using the same strategy with the first term of equation (B.3), one deduces that this term is

$$
\widetilde{Z}_{jxn}(t,\overline{u}_\gamma) - \widetilde{Z}_{jxn}(t,\underline{u}_\gamma) = \mathbb{F}_{jz_{\underline{t}_\gamma}h_j}\{F^{-1}_{jz_t}(\overline{u}_\gamma)\} - \mathbb{F}_{jz_{\underline{t}_\gamma}h_j}\{F^{-1}_{jz_{\underline{t}_\gamma}}(\underline{u}_\gamma)\} + o(1).
$$

From the continuity of the function $z \mapsto F^{-1}_{jz}$ in a neighborhood of x and similar arguments as the ones used previously,

$$
\begin{aligned}
|\mathbb{F}_{jz_{\underline{t}_\gamma}h_j}\{F^{-1}_{jz_t}(\overline{u}_\gamma)\} - \mathbb{F}_{jz_{\underline{t}_\gamma}h_j}\{F^{-1}_{jz_{\underline{t}_\gamma}}(\underline{u}_\gamma)\}| &\le |\mathbb{F}_{jz_{\underline{t}_\gamma}h_j}\{F^{-1}_{jz_{\overline{t}_\gamma}}(\overline{u}_\gamma)\} - \mathbb{F}_{jz_{\underline{t}_\gamma}h_j}\{F^{-1}_{jz_{\underline{t}_\gamma}}(\overline{u}_\gamma)\}| \\
&\quad + |\mathbb{F}_{jz_{\underline{t}_\gamma}h_j}\{F^{-1}_{jz_{\underline{t}_\gamma}}(\overline{u}_\gamma)\} - \mathbb{F}_{jz_{\underline{t}_\gamma}h_j}\{F^{-1}_{jz_{\underline{t}_\gamma}}(\underline{u}_\gamma)\}| + o(1).
\end{aligned}
$$

Wrapping up the discussions around decomposition (B.3), one concludes that

$$
\begin{aligned}
\sup_{(t,u)\in\mathcal{I}}|\widetilde{Z}_{jxn}(t,u) - \widetilde{Z}_{jxn}(\underline{t}_\gamma,\underline{u}_\gamma)| &\le 2\sup_{t\in I^{(1)}_{n,\gamma},u\in[0,1]}|\mathbb{F}_{jz_th_j}\{F^{-1}_{jz_t}(\overline{u}_\gamma)\} - \mathbb{F}_{jz_th_j}\{F^{-1}_{jz_t}(\underline{u}_\gamma)\}| \\
&\quad + 2\sup_{t\in[-1,1],u\in I^{(2)}_{n,\gamma}}|\mathbb{F}_{jz_{\underline{t}_\gamma}h_j}\{F^{-1}_{jz_{\overline{t}_\gamma}}(u)\} - \mathbb{F}_{jz_{\underline{t}_\gamma}h_j}\{F^{-1}_{jz_{\underline{t}_\gamma}}(u)\}| + o(1).
\end{aligned}
$$

For $t_k = \frac{k}{\kappa^{(j)}_{n,\gamma}}$, denote $A^{(j)}_{n,\gamma}(i,t) = [F^{-1}_{jz_t}\{\frac{i-1}{\kappa^{(j)}_{n,\gamma}}\}, F^{-1}_{jz_t}\{\frac{i}{\kappa^{(j)}_{n,\gamma}}\}]$ and $B^{(j)}_{n,\gamma}(i,k) = [F^{-1}_{jz_{t_k}}\{\frac{i}{\kappa^{(j)}_{n,\gamma}}\}, F^{-1}_{jz_{t_{k+1}}}\{\frac{i}{\kappa^{(j)}_{n,\gamma}}\}]$. Moreover, write

$$
\mathcal{G}_{n,\gamma} = \{0,1,\ldots,\kappa^{(j)}_{n,\gamma}\} \times \{0,\pm1,\ldots,\pm\kappa^{(j)}_{n,\gamma}\}.
$$

Then from similar arguments as in the end of Section 6.1, for sufficiently large n:

$$\mathfrak{W}_\delta(\widetilde{Z}_{jxn},\mathcal{I}) \le 6 \max_{(i,k)\in\mathcal{G}_{n,\gamma}} |\mathbb{F}_{jz_{t_k}h_j}\{A^{(j)}_{n,\gamma}(i,t_k)\}| + 6 \max_{(i,k)\in\mathcal{G}_{n,\gamma}} |\mathbb{F}_{jz_{t_k}h_j}\{B^{(j)}_{n,\gamma}(i,t_k)\}|.$$

For any intervale $A=[a,b]\subset\mathbb{R}$, denote $\nu_{jz}(A)=F_{jz}(b)-F_{jz}(a)$. On one hand, $\nu_{jz_{t_k}}(A^{(j)}_{n,\gamma}(i,t_k))=(nh)^{-1/2-\gamma}$. On the other hand,

$$\begin{aligned}\nu_{jz_{t_k}}(B^{(j)}_{n,\gamma}(i,k)) &= |u-F_{jz_{t_k}}\{F^{-1}_{jz_{t_{k+1}}}(u)\}| \\ &= |\dot{F}_{jz_{t_{k+1}}}\{F^{-1}_{jz_{t_{k+1}}}(u)\}(z_{t_k}-z_{t_{k+1}})+\frac{1}{2}\ddot{F}_{jz_{t^\star}}\{F^{-1}_{jz_{t_{k+1}}}(u)\}(z_{t_k}-z_{t_{k+1}})^2|\end{aligned}$$

where $t^\star\in[t_k,t_{k+1}]$. Since $z_{t_k}-z_{t_{k+1}}=h(nh_j)^{1/2+\gamma}$, the $\nu_{jz_{t_k}}$-measure of the set $B^{(j)}_{n,\gamma}(i,k)$ is smaller than the $\nu_{jz_{t_k}}$-measure of the set $A^{(j)}_{n,\gamma}(i,t_k)$. One then argues that for n sufficiently large, for any $(i,k)\in\mathcal{G}_{n,\gamma}$, either $B^{(j)}_{n,\gamma}(i,k)\subset A^{(j)}_{n,\gamma}(i-1,t_k)$ or $B^{(j)}_{n,\gamma}(i,k)\subset A^{(j)}_{n,\gamma}(i,t_k)$. Thus for any $\epsilon>0$:

$$\begin{aligned}\mathbb{P}\left(\mathfrak{W}_\delta(\widetilde{Z}_{jxn},\mathcal{I})\ge\epsilon\right) &\le \mathbb{P}\left(12\max_{(i,k)\in\mathcal{G}_{n,\gamma}}|\mathbb{F}_{jz_{t_k}h_j}\{A^{(j)}_{n,\gamma}(i,t_k)\}|\ge\frac{\epsilon}{12}\right)\\ &\le \sum_{(i,k)\in\mathcal{G}_{n,\gamma}}\mathbb{P}\left(|\mathbb{F}_{jz_{t_k}h_j}\{A^{(j)}_{n,\gamma}(i,t_k)\}|\ge\frac{\epsilon}{12}\right)\\ &\le (nh_j)^{1+2\gamma}\max_{(i,k)\in\mathcal{G}_{n,\gamma}}\mathbb{P}\left(|\mathbb{F}_{jz_{t_k}h_j}\{A^{(j)}_{n,\gamma}(i,t_k)\}|\ge\frac{\epsilon}{12}\right)\end{aligned}$$

Using Markov inequality coupled with Lemma 2 with $\mu_x=\nu_{jz_{t_k}}$ and $r_a=6$, for some $\beta>0$:

$$\begin{aligned}\mathbb{P}\left(|\mathbb{F}_{jz_{t_k}h_j}\{A^{(j)}_{n,\gamma}(i,t_k)\}|\ge\epsilon\right) &\le \epsilon^{-6}\mathbb{E}\left((\mathbb{F}_{jz_{t_k}h_j}\{A^{(j)}_{n,\gamma}(i,t_k)\})^6\right)\\ &\le C_\alpha(r_a,z_t)\sum_{\ell=1}^{3}\{\nu_{jz_{t_k}}(A^{(j)}_{n,\gamma}(i,t_k))+h_j^2\}^\ell(nh_j)^{\ell-3}\\ &= C_\alpha(r_a,z_t)(nh_j)^{-1-2\gamma-\beta}S_{n,\gamma}(\beta)\end{aligned}$$

where $S_{n,\gamma}(\beta)=\sum_{\ell=1}^3\{(nh_j)^{-1/2-\gamma}+h_j^2\}^\ell(nh_j)^{\ell-2+\beta+2\gamma}$. It will now be shown that $S_{n,\gamma}(\beta)<1$ for some proper choice of β. Consider first the case when $h_j\sim n^{-\tau}$ with $\tau\le\frac{1}{3+2\gamma}$. This implies $h_j^2>(nh_j)^{-1/2-\gamma}$. It that case :

$$\begin{aligned}S_{n,\gamma}(\beta) &\le 8h_j^6(nh_j)^{1+2\gamma+\beta}+4h_j^4(nh_j)^{2\gamma+\beta}+2h_j^2(nh_j)^{-1+2\gamma+\beta}\\ &\sim n^{-6\tau+(1-\tau)(1+2\gamma+\beta)}+n^{-4\tau+(1-\tau)(2\gamma+\beta)}+n^{-2\tau+(1-\tau)(-1+2\gamma+\beta)}.\end{aligned}$$

Thus, taking $\beta < 1/2 - 2\gamma$ leads $S_{n,\gamma}(\beta) < 1$ for some small $\gamma < 1/4$ since $\tau \geq 1/5$. Next consider the case when $\tau > \frac{1}{3+2\gamma}$ which implies $h_j^2 < (nh_j)^{-1/2-\gamma}$. In that case:

$$S_{n,\gamma}(\beta) \leq 8(nh_j)^{-1/2-\gamma+\beta} + 4(nh_j)^{-1+\beta} + 2(nh_j)^{-3/2+\gamma+\beta}.$$

The choice of $\beta < 1/2 - 2\gamma$ still implies $S_{n,\gamma}(\beta) < 1$. The conclusion is that $\mathbb{P}\left(\mathfrak{W}_\delta(\widetilde{Z}_{jxn}, \mathcal{I}) \geq \epsilon\right) \leq (nh)^{-\beta}$. The lemma is therefore proven.

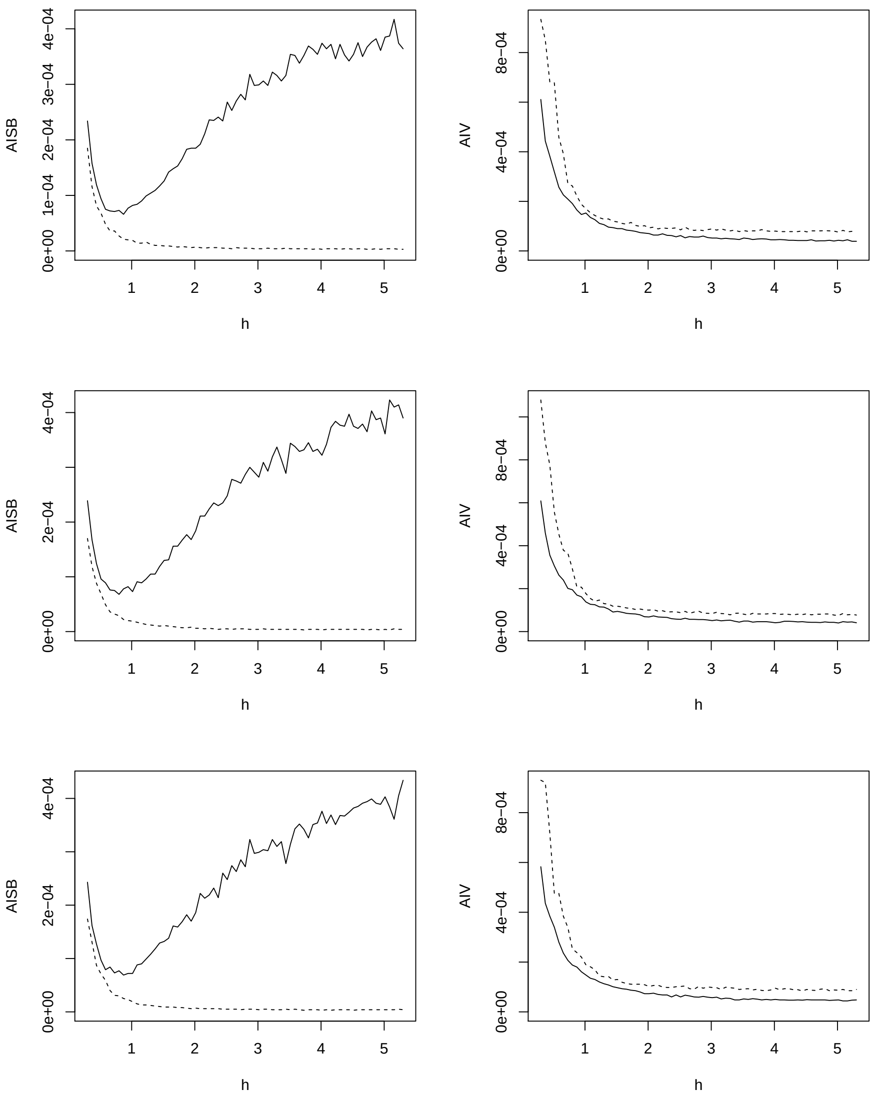

Figure 1: Average integrated squared bias (left) and average integrated variance (right) of C_{xh} (line) and $\tilde{C}_{xh}$ (dashed line) as a function of h under model M_1 when $\theta = 0$ (upper panels), $\theta = .25$ (middle panels) and $\theta = .5$ (bottom panels)

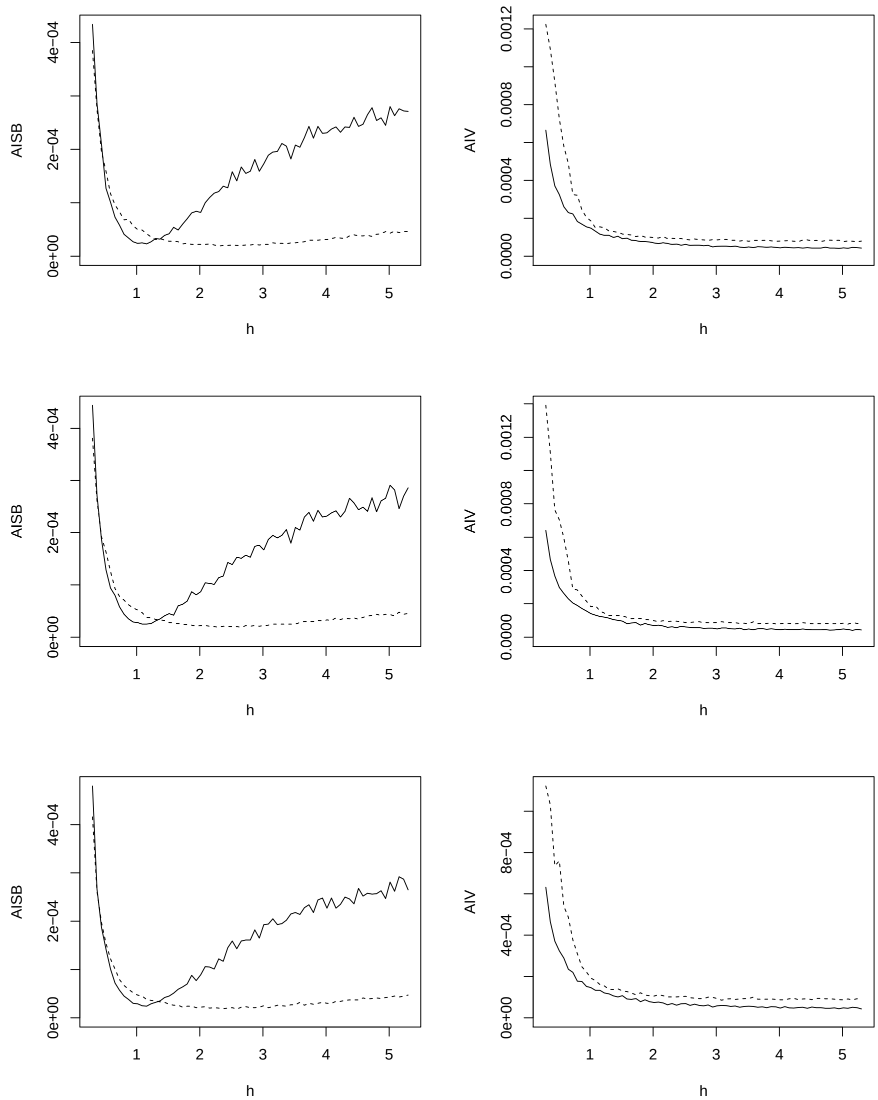

Figure 2: Average integrated squared bias (left) and average integrated variance (right) of C_{xh} (line) and $\tilde{C}_{xh}$ (dashed line) as a function of h under model M_2 when $\theta = 0$ (upper panels), $\theta = .25$ (middle panels) and $\theta = .5$ (bottom panels)

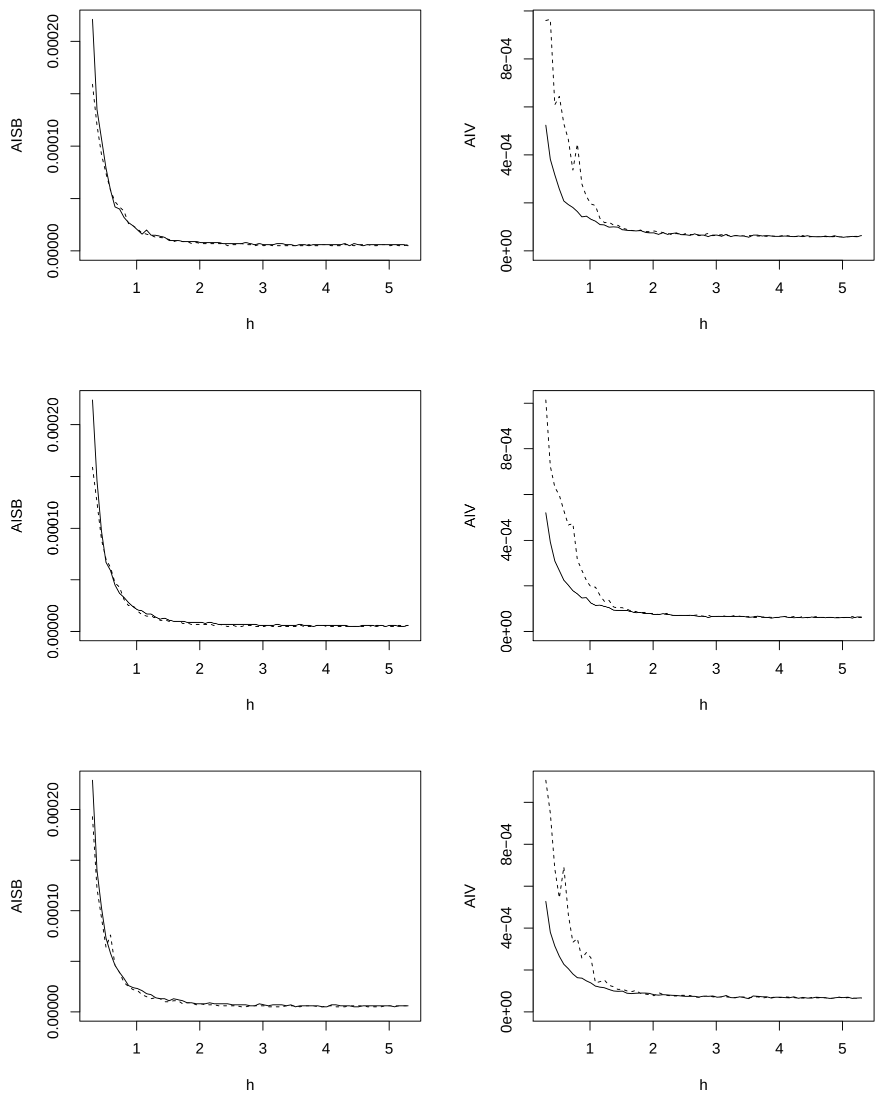

Figure 3: Average integrated squared bias (left) and average integrated variance (right) of C_{xh} (line) and $\tilde{C}_{xh}$ (dashed line) as a function of h under model M_3 when $\theta = 0$ (upper panels), $\theta = .25$ (middle panels) and $\theta = .5$ (bottom panels)

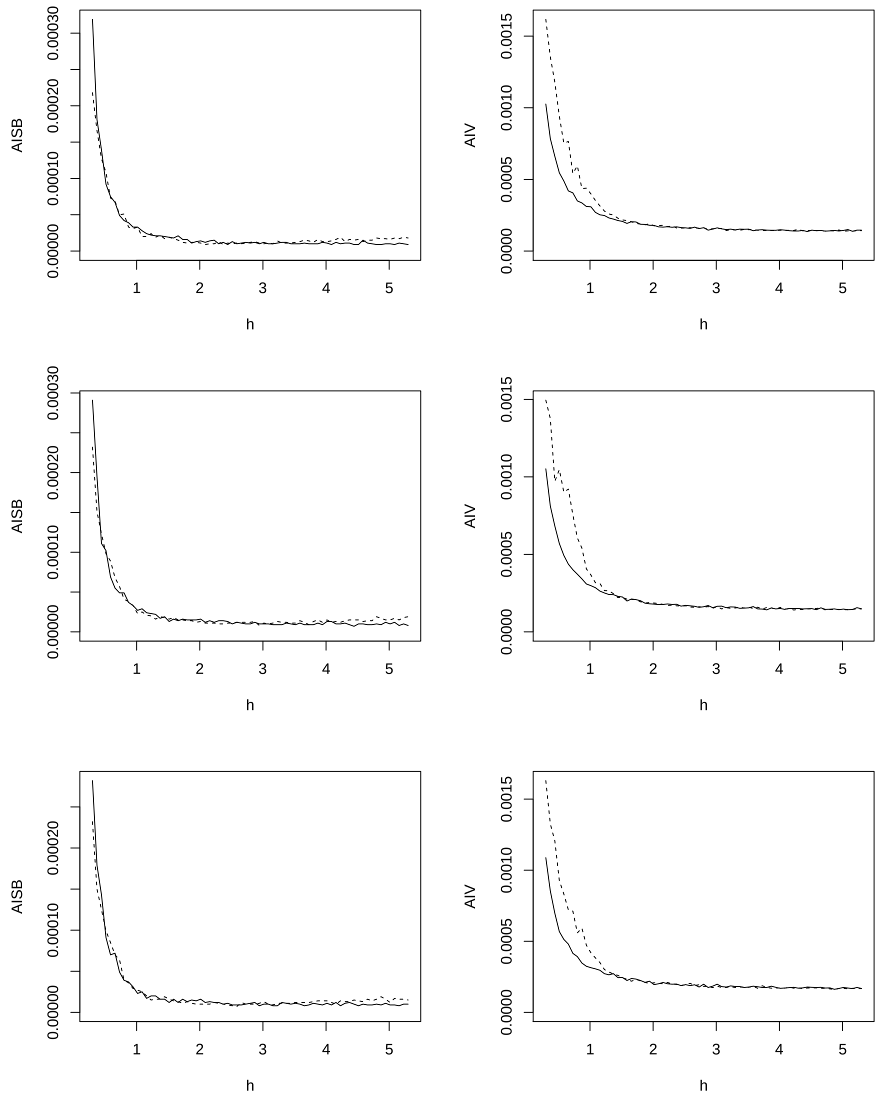

Figure 4: Average integrated squared bias (left) and average integrated variance (right) of C_{xh} (line) and $\tilde{C}_{xh}$ (dashed line) as a function of h under model M_4 when $\theta = 0$ (upper panels), $\theta = .25$ (middle panels) and $\theta = .5$ (bottom panels)

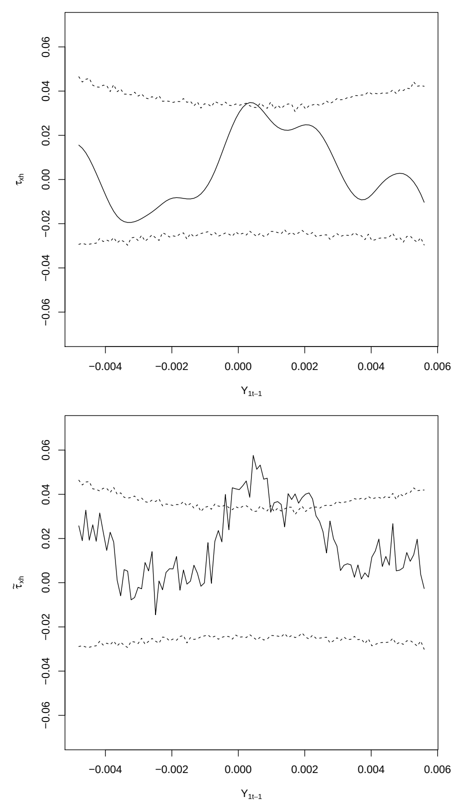

Figure 5: Estimated values of Kendall's tau of $(Z_{1i}, Z_{2,i-1})$ given $Z_{1,i-1} = x$ as a function of x, where Z_{1i} and Z_{2i} are respectively the return and the volume at time i. Upper panel: estimation based on C_{xh}; bottom panel: estimation based on $\widetilde{C}_{xh}$

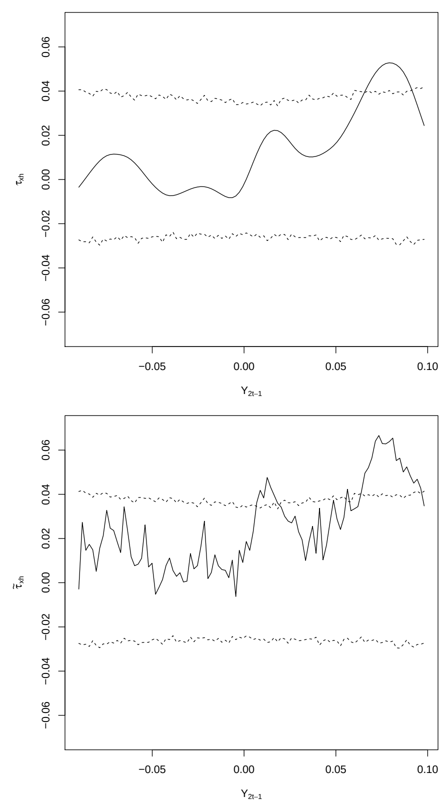

Figure 6: Estimated values of Kendall's tau of $(Z_{1i}, Z_{2,i-1})$ given $Z_{1,i-1} = x$ as a function of x, where Z_{1i} and Z_{2i} are respectively the volume and the return at time i. Upper panel: estimation based on C_{xh}; bottom panel: estimation based on $\widetilde{C}_{xh}$