Sur des estimateurs et des tests non-paramétriques pour des distributions et copules conditionnelles

par

Félix Camirand Lemyre

Thèse présentée au Département de mathématiques

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, septembre 2016

Le 2 septembre 2016

le jury a accepté la thèse de Monsieur Félix Camirand Lemyre
dans sa version finale

Membres du jury

Professeur Taoufik Bouezmarni
Directeur de recherche
Département de mathématiques

Professeur Jean-François Quessy
Codirecteur de recherche
Département de mathématiques et d'informatiques
Université du Québec à Trois-Rivières

Professeur Bernard Colin
Membre interne
Département de mathématiques

Professeure Elif Acar Membre externe
Department of Statistics
University of Manitoba

Professeur Éric Marchand
Président-Rapporteur
Département de mathématiques

SOMMAIRE

Pour modéliser un vecteur aléatoire en présence d'une co-variable, on peut d'abord faire appel à la fonction de répartition conditionnelle. En effet, cette dernière contient toute l'information ayant trait au comportement du vecteur étant donné une valeur prise par la co-variable. Il peut aussi être commode de séparer l'étude du comportement conjoint du vecteur de celle du comportement individuel de chacune de ses composantes. Pour ce faire, on utilise la copule conditionnelle, qui caractérise complètement la dépendance conditionnelle régissant les différentes associations entre les variables. Dans chacun des cas, la mise en oeuvre d'une stratégie d'estimation et d'inférence s'avère une étape essentielle à leur utilisant en pratique. Lorsqu'aucune information n'est disponible *a priori* quant à un choix éventuel de modèle, il devient pertinent d'opter pour des méthodes non-paramétriques.

Le premier article de cette thèse, co-écrit par Jean-François Quessy et moi-même, propose une façon de ré-échantillonner des estimateurs non-paramétriques pour des distributions conditionnelles. Cet article a été publié dans la revue *Statistics and Computing*. En autres choses, nous y montrons comment obtenir des intervalles de confiance pour des statistiques s'écrivant en terme de la fonction de répartition conditionnelle.

Le second article de cette thèse, co-écrit par Taoufik Bouezmarni, Jean-François Quessy et moi-même, s'affaire à étudier deux estimateurs non-paramétriques de la copule condi-

tionnelles, proposés par Gijbels *et al.* (2011), en présence de données sérielles. Cet article a été soumis dans la revue *Statistics and Probability Letters*. Nous identifions la distribution asymptotique de chacun de ces estimateurs pour des données α-mélangeantes.

Le troisième article de cette thèse, co-écrit par Taoufik Bouezmarni, Jean-François Quessy et moi-même, propose une nouvelle façon d'étudier les relations de causalité entre deux séries chronologiques. Cet article a été soumis dans la revue *Electronic Journal of Statistics*. Dans cet article, nous utilisons la copule conditionnelle pour caractériser une version locale de la causalité au sens de Granger. Puis, nous proposons des mesures de causalité basées sur la copule conditionnelle.

Le quatrième article de cette thèse, co-écrit par Taoufik Bouezmarni, Anouar El Ghouch et moi-même, propose une méthode qui permette d'estimer adéquatement la copule conditionnelle en présence de données incomplètes. Cet article a été soumis dans la revue *Scandinavian Journal of Statistics*. Les propriétés asymptotiques de l'estimateur proposé y sont aussi étudiées.

Finalement, la dernière partie de cette thèse contient un travail inédit, qui porte sur la mise en oeuvre de tests statistiques permettant de déterminer si deux copules conditionnelles sont concordantes. En plus d'y présenter des résultats originaux, cette étude illustre l'utilité des techniques de ré-échantillonnage développées dans notre premier article.

REMERCIEMENTS

Mes premiers remerciements s'adressent à mes directeurs de thèse, les professeurs Taoufik Bouezmarni et Jean-François Quessy, pour leur précieux soutien accordé durant ces quelques années. Je leur suis très reconnaissant de l'encadrement qu'ils m'ont patiemment prodigué, et de toutes ces opportunités qu'ils ont généreusement placées le long de mon chemin. Qu'ils en soient ici chaleureusement remerciés.

J'exprime aussi ma gratitude à mon collaborateur Anouar El Ghouch de l'Université catholique de Louvain, pour son accueil amical durant mes deux voyages en Belgique.

Les examinateurs ayant accepté de siéger sur le jury de cette thèse doivent ici trouver la manifestation de ma profonde reconnaissance.

Je remercie également mes collègues et amis pour ce parfum de folie insufflé indéfectiblement à chacune de mes journées.

À mes parents, mon frère et ma soeur, compagnons et compagnes de route, je ne peux trouver de mots assez forts pour vous témoigner de toute ma reconnaissance pour ces moments partagés avec vous, à travers les nombreux aléas de mon parcours.

Finalement, je tiens à remercier le *Conseil de recherche en sciences naturelles et en génie du Canada* (CRSNG) pour la bourse de doctorat qu'il m'a accordée, ainsi que le département de mathématiques de l'Université de Sherbrooke, pour ces belles années.

Félix Camirand Lemyre

Sherbrooke, Juillet 2016

TABLE DES MATIÈRES

INTRODUCTION

0.1 Considérations générales sur les distributions et les copules conditionnelles

Il est souvent d'intérêt d'étudier le comportement d'un vecteur aléatoire $\mathbf{Y} \in \mathbb{R}^d$ en présence d'une co-variable $X \in \mathbb{R}$. Dans le cas particulier où $d = 1$, plusieurs chercheurs se sont penchés sur l'influence de la co-variable sur la moyenne de Y, ce qui équivaut à étudier son espérance conditionnelle. Dans ce domaine, une approche classique consiste à supposer un modèle de régression du type $Y = m(X) + \epsilon$, avec ϵ une variable aléatoire de moyenne nulle et de variance finie. Selon le contexte de la modélisation, certains n'émettront aucune restriction sur la fonction $m(\cdot)$, d'autres présumeront que cette dernière est un polynôme dont les coefficients sont inconnus.

Lorsque le contexte ne permet pas de préjuger d'une telle relation entre le vecteur d'intérêt et la co-variable, il est commode de faire appel à la fonction de répartition du vecteur $\mathbf{Y} = (Y_1, \ldots, Y_d)$ conditionnellement à $X = x$, c'est-à-dire

$$F_x(y_1, \ldots, y_d) = \mathrm{P}\left(Y_1 \leq y_1, \ldots, Y_d \leq y_d | X = x\right).$$

En effet, la fonction $F_x : \mathbb{R}^d \to [0, 1]$ caractérise complètement le comportement stochastique du vecteur aléatoire $\mathbf{Y}$ lorsque $X = x$. À partir de F_x, on extrait les distributions marginales de chacune des j composantes par l'entremise de la relation

$$F_{jx}(y_j) = \lim_{y_k \to \infty, k \neq j} F_x(\mathbf{y}).$$

Plusieurs moments s'expriment en fonction de F_x. Par exemple, lorsque $d = 1$, la moyenne et la variance conditionnelles peuvent s'exprimer respectivement par

$$\mathrm{E}(Y|X = x) = \int_{\mathbb{R}^+} \{1 - F_x(y) + F_x(-y)\} \, \mathrm{d}y \tag{1}$$

et

$$\mathrm{Var}(Y|X = x) = \int_{\mathbb{R}^2} \{F_x(y \wedge y') - F_x(y) F_x(y')\} \, \mathrm{d}y \, \mathrm{d}y'. \tag{2}$$

Dans le cas bidimensionnel, c'est-à-dire quand $d = 2$, la covariance conditionnelle s'écrit

$$\mathrm{Cov}(Y_1, Y_2|X = x) = \int_{\mathbb{R}^2} \{F_x(y, y') - F_{1x}(y) \, F_{2x}(y')\} \, \mathrm{d}y \, \mathrm{d}y'. \tag{3}$$

À noter que les Équations (2) et (3) sont une conséquence de l'identité de Hoeffding ; pour plus de détails sur cette identité, voir par exemple Hoeffding (1994) et Lehmann (1966). Maintenant, soient des *copies* du vecteur aléatoire $(\mathbf{Y}, X)$, c'est-à-dire un échantillon de la forme $(\mathbf{Y}_1, X_1), \ldots, (\mathbf{Y}_n, X_n)$. À partir de ces observations, il est possible d'estimer la fonction inconnue F_x par un estimateur $\widehat{F}_x$. Une fois qu'un tel estimateur est disponible, on estime les moments décrits aux Équations (1)–(3) en remplaçant simplement F_x par $\widehat{F}_x$.

Dans cette thèse, une attention particulière sera portée à la dépendance conditionnelle vue à travers le prisme de la théorie des copules. En effet, on sait depuis le fameux Théorème de Sklar (1959) que le comportement stochastique d'un vecteur aléatoire dans le cas $d \geq 2$ est caractérisé par les effets de ses lois marginales et d'une fonction copule qui explique tous les liens de dépendance entre ses composantes. Plus précisément, le Théorème de Sklar assure que lorsque les distributions marginales conditionnelles de F_x sont continues, alors il existe une unique fonction $C_x : [0, 1]^d \to [0, 1]$ telle que

$$F_x(y_1, \ldots, y_d) = C_x \{F_{1x}(y_1), \ldots, F_{dx}(y_d)\}.$$

La fonction $C_x : [0,1]^d \to [0,1]$ s'appelle la *copule conditionnelle* ; elle contient toute l'information concernant la dépendance entre les composantes du vecteur $\mathbf{Y}$ conditionnellement à $X = x$. Inversement, on peut extraire la copule conditionnelle associée à une distribution conditionnelle donnée. À cette fin, soit la fonctionnelle

$$\Lambda^{\mathrm{C}}(\delta) = \delta\left\{\delta_1^{-1}(u_1), \ldots, \delta_d^{-1}(u_d)\right\}$$

définie sur l'espace des fonctions de répartition d-dimensionnelles δ, où

$$\delta_i(y_i) = \lim_{y_j \to \infty, j \neq i} \delta(\mathbf{y}) \quad \text{et} \quad \delta_i^{-1}(u) = \inf_{y \in \mathbb{R}} \left\{\delta_i(y) \geq u\right\}.$$

La copule conditionnelle associée à F_x est alors

$$C_x(u_1, \ldots, u_d) = \Lambda^{\mathrm{C}}(F_x) = F_x\left\{F_{1x}^{-1}(u_1), \ldots, F_{dx}^{-1}(u_d)\right\}. \tag{4}$$

À partir de cette formule, un estimateur naturel pour C_x est $\widehat{C}_x = \Lambda^{\mathrm{C}}(\widehat{F}_x)$, où $\widehat{F}_x$ est un estimateur de la fonction de répartition conditionnelle F_x. Ainsi,

$$\widehat{C}_x(u_1, \ldots, u_d) = \widehat{F}_x\left\{\widehat{F}_{1x}^{-1}(u_1), \ldots, \widehat{F}_{dx}^{-1}(u_d)\right\}.$$

Cette relation entre la copule conditionnelle et la distribution conditionnelle permettra de déduire les propriétés asymptotiques de cet estimateur de C_x à partir de celles de $\widehat{F}_x$. Cet aspect est traité en détails au Chapitre 3.

À noter que quelques auteurs ont récemment proposé diverses façons d'estimer la copule conditionnelle, et ce dans différents contextes. Par exemple, des méthodes paramétriques, qui par nature nécessitent de spécifier une connexion fonctionnelle entre la co-variable X et la copule C_x, ont été explorées par Jondeau & Rockinger (2006) et Patton (2006). Une approche non-paramétrique pour estimer ce lien fonctionnel a été adoptée par Acar *et coll.* (2011) dans un contexte où les lois marginales sont connues ; sa généralisation au cas de marges inconnues est traitée par Abegaz *et coll.* (2011). Enfin, des méthodes entièrement non-paramétriques sont explorées dans Gijbels *et coll.* (2011) et Veraverbeke *et coll.* (2011).

Un des intérêts principaux de l'approche par copule est que plusieurs mesures de dépendance populaires s'écrivent comme des fonctionnelles de la copule. Deux exemples notoires concernent les mesures de dépendance de Kendall et de Spearman. À l'origine, l'indice de dépendance de Kendall, ou tau de Kendall, est défini comme la différence entre les probabilités de concordance et de discordance. Ainsi, dans le cas conditionnel, on le calcule à partir de deux triplets de variables aléatoires indépendantes (Y_1, Y_2, X) et (Y_1', Y_2', X') de même loi via la formule

$$\begin{aligned} \tau_x &= \mathrm{P}\left\{ \left(Y_1 - Y_1'\right)\left(Y_2 - Y_2'\right) > 0 | X = X' = x \right\} \\ &\quad - \mathrm{P}\left\{ \left(Y_1 - Y_1'\right)\left(Y_2 - Y_2'\right) < 0 | X = X' = x \right\}. \end{aligned}$$

À partir de cette définition, on montre qu'en fait, la valeur de τ_x s'exprime en fonction de la copule conditionnelle par la relation

$$\tau_x = 4 \int_{[0,1]^2} C_x(u_1, u_2) \, \mathrm{d}C_x(u_1, u_2) - 1. \tag{5}$$

De son côté, le rho de Spearman conditionnel est défini comme la correlation conditionnelle entre $F_{1x}(Y_1)$ et $F_{2x}(Y_2)$ étant donné que $X = x$. Plus précisément,

$$\rho_x = \mathrm{Corr}\left\{ F_{1x}(Y_1), F_{2x}(Y_2) \right\}.$$

À l'instar de τ_x, l'indice de dépendance ρ_x est une fonctionnelle de C_x car

$$\rho_x = 12 \int_{[0,1]^2} C_x(u_1, u_2) \, \mathrm{d}u_1 \, \mathrm{d}u_2 - 3. \tag{6}$$

Des estimateurs pour τ_x et pour ρ_x sont donc disponibles en remplaçant C_x par $\widehat{C}_x$ dans les Équations (5) et (6).

Les distributions et les copules conditionnelles jouent un rôle crucial dans l'exploration du comportement conjoint d'un vecteur aléatoire en présence d'une co-variable. En fait, que ce soit la copule conditionnelle, les mesures de distribution conditionnelles ou de dépendance conditionnelles, toutes ont en commun de s'écrire en fonction de F_x. Ces dernières peuvent donc être estimées à partir d'une version empirique de F_x. Ainsi, le choix d'une stratégie d'estimation pour la distribution conditionnelle est au coeur d'une telle investigation statistique. Lorsque le contexte ne permet pas d'identifier un modèle paramétrique pour F_x, cette estimation peut s'effectuer de façon non-paramétrique à l'aide de méthodes dites *à noyau* ; cette technique d'estimation est en quelque sorte à la base de ce travail de thèse. D'abord, dans un cadre d'estimation non conditionnelle, c'est-à-dire sans co-variable, un estimateur non-paramétrique de la fonction de répartition F est donné par la *fonction de répartition empirique*, à savoir

$$F_n(\mathbf{y}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left(\mathbf{Y}_i \leq \mathbf{y}\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left(Y_{1i} \leq y_1, \ldots, Y_{di} \leq y_d\right).$$

On voit qu'un poids uniforme de $1/n$ est accordé à chaque observation. En présence d'une co-variable, puisque l'intérêt réside dans le comportement de $\mathbf{Y}$ au cas précis où $X = x$, une stratégie consiste à octroyer plus de poids aux observations $(\mathbf{Y}_i, X_i)$ pour lesquelles X_i se situe *près* de x. Pour ce faire, on utilise des pondérations

$$w_{n1}(x,h), \ldots, w_{nn}(x,h)$$

qui modulent le poids accordé à chacune des observations. Ici et dans la suite, $h = h_n$ est un paramètre de lissage qui dépend généralement de la taille d'échantillon n. Un estimateur de F_x est alors donné par la fonction de répartition empirique conditionnelle

$$F_{xh}(\mathbf{y}) = \sum_{i=1}^{n} \mathbb{I}\left(\mathbf{Y}_i \leq \mathbf{y}\right) w_{ni}(x,h).$$

Dans la littérature, on retrouve plusieurs possibilités concernant le choix des poids. Parmi ceux-ci, on retrouve les populaires poids de Nadaraya–Watson définis par

$$w_{ni}^{\mathrm{NW}}(x,h) = K(\widetilde{X}_{i,h}) \Big/ \sum_{i=1}^{n} K(\widetilde{X}_{i,h}),$$

où $\widetilde{X}_{i,h} = (X_i - x)/h$ et K est une densité sur $[-1,1]$ qui est symétrique en zéro et continument différentiable. On peut également utiliser les poids locaux-linéaires, à savoir

$$w_{ni}^{\mathrm{LL}}(x,h) = \frac{K(\widetilde{X}_{i,h})\left(S_{n,2} - \widetilde{X}_{i,h}\, S_{n,1}\right)}{S_{n,0}\, S_{n,2} - S_{n,1}^2},$$

où pour $j \in \{0,1,2\}$,

$$S_{n,j} = \sum_{i=1}^{n} \widetilde{X}_{i,h}^{j}\, K(\widetilde{X}_{i,h}).$$

Un choix populaire pour K est la fonction *triweight* définie par

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

Cette fonction est présentée à la figure 1.

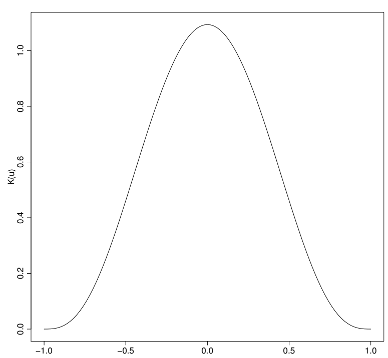


Figure 1 – Graphique de la fonction $K(u) = (35/32)(1-u^2)^3\, \mathbb{I}(|u| \leq 1)$

Les différents outils statistiques développés dans cette thèse sont présentés suivant un système de poids non spécifiés $w_{n1}(x,h), \ldots, w_{nn}(x,h)$ qui satisfait une liste de conditions qui varient selon le résultat désiré. Toutefois, à noter que tous les résultats obtenues dans cette thèse sont toujours valides pour les poids Nadaraya–Watson et locaux-linéaires.

Ainsi, avec l'aide de cet estimateur pour F_x, les estimateurs de la moyenne et la variance conditionnelle s'écrivent respectivement

$$\widehat{\mathrm{E}(Y|X=x)} = \sum_{i=1}^{n} Y_i\, w_{ni}(x,h)$$

et

$$\widehat{\mathrm{Var}(Y|X=x)} = \sum_{i=1}^{n} Y_i^2 w_{ni}(x,h) - \left\{ \sum_{i=1}^{n} Y_i^2 w_{ni}(x,h) \right\}^2 .$$

On peut aussi obtenir de façon analogue une forme explicite d'un estimateur de $\mathrm{Cov}(Y_1, Y_2|X = x)$.

0.2 Présentation et détails sur l'article présenté au Chapitre 1

Le Chapitre 1 développe des méthodes de ré-échantillonnage spécifiquement adaptées au cas de distributions conditionnelles. En fait, les propriétés asymptotiques des estimateurs basés sur F_{xh} peuvent souvent être déduites à partir de celles de cette fonction de répartition empirique conditionnelle. En établissant la loi limite de la fonction aléatoire $\mathbb{F}_{xh} = \sqrt{nh}(F_{xh} - F_x)$, des intervalles de confiance pour une fonctionnelle $\Lambda(F_x)$ peuvent, du moins en principe, être déduits. En présence de triplets indépendants $(\mathbf{Y}_1, X_1), \ldots, (\mathbf{Y}_n, X_n)$, un résultat de convergence au sens faible du processus $\mathbb{F}_{xh}$ dans l'espace $\ell^\infty(\mathbb{R}^d)$ des fonctions continues et bornées sur $\mathbb{R}^d$ a été obtenu par Veraverbeke *et coll.* (2011). Ainsi, pourvu que $\sqrt{nh}\, h^2 \to K < \infty$, et en imposant que la fonction $(x, \mathbf{y}) \to F_x(\mathbf{y})$ satisfasse quelques conditions de régularité, $\mathbb{F}_{xh}$ converge vers un processus de la forme $\mathbb{F}_x = \alpha_x + B_x$, où α_x est un processus gaussien centré de fonction de covariance

$$\mathrm{Cov}\left\{\mathbb{F}_x(\mathbf{y}), \mathbb{F}_x(\mathbf{y}')\right\} = K_4 \left\{F_x(\mathbf{y} \wedge \mathbf{y}') - F_x(\mathbf{y})F_x(\mathbf{y}')\right\}$$

et la fonction B_x est un biais asymptotique qui s'exprime par

$$B_x(\mathbf{y}) = K \left\{ K_2\, \dot{F}_x(\mathbf{y}) + \frac{K_3}{2}\, \ddot{F}_x(\mathbf{y}) \right\},$$

où

$$\dot{F}_x(\mathbf{y}) = \frac{\partial}{\partial x} F_x(\mathbf{y}) \quad \text{et} \quad \ddot{F}_x(\mathbf{y}) = \frac{\partial^2}{\partial x^2} F_x(\mathbf{y}).$$

Les constantes K_2, K_3 et K_4 proviennent de conditions imposées aux poids ; elles apparaissent dans les hypothèses W_2–W_4 que l'on retrouve à l'Annexe A. À noter que le terme de biais est la conséquence de l'utilisation des pondérateurs $w_{n1}(x,h),\ldots,w_{nn}(x,h)$ qui répartissent le poids de manière lisse autour de x.

Le résultat asymptotique que l'on vient de décrire s'avère peu utile lorsque l'on désire construire des intervalles de confiance. En effet, la loi limite de F_{xh} fait intervenir la fonction F_x. Sauf que justement, si on a estimé F_x, c'est que celle-ci est inconnue ! On se retrouve ainsi devant une impasse. Pour contourner ce problème, il faut utiliser des méthodes de ré-échantillonnage appropriées. C'est exactement ce qui est développé dans l'article intitulé *Multiplier bootstrap methods for conditional distributions*, co-écrit par Jean-François Quessy et moi-même, et publié dans la revue *Statistics and Computing.*

D'abord, supposons que l'on cherche à approximer le comportement limite du processus $\mathbb{F}_n = \sqrt{n}(F_n - F)$, où F_n est la fonction de répartition empirique basée sur $\mathbf{Y}_1,\ldots,\mathbf{Y}_n$. Pour ce faire, on peut recourir au *bootstrap* traditionnel. Cette méthode consiste à se doter d'un échantillon *bootstrap* $\mathbf{Y}_1^\star,\ldots,\mathbf{Y}_n^\star$ tiré au hasard avec remise dans une urne contenant $\mathbf{Y}_1,\ldots,\mathbf{Y}_n$; ainsi, $\mathbf{Y}_i^\star \sim F_n$. L'approximation *bootstrap* de $\mathbb{F}_n$ est alors

$$\mathbb{F}_n^{\text{boot}}(\mathbf{y}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \mathbb{I}\left(\mathbf{Y}_i^\star \le \mathbf{y}\right) - F_n(\mathbf{y}) \right\}.$$

Une alternative au bootstrap consiste à recourir à la méthode du multiplicateur. Cette technique est bien développée dans un cadre non conditionnel, c'est-à-dire sans covariable. En outre, elle est très efficace et relativement simple à utiliser. L'idée est basée sur des variables aléatoires *multiplicateurs* $\xi_1,\ldots,\xi_n$ qui sont i.i.d. de moyenne nulle, de variance unitaire et telles qu'il existe $\delta > 0$ pour lequel $\mathrm{P}(|\xi_i|^{2+\delta} > \epsilon) \to 0$ pour tout

$\epsilon > 0$. La version multiplicateur de $\mathbb{F}_n$ est alors

$$\mathbb{F}_n^{\text{mult}}(\mathbf{y}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\xi_i - \bar{\xi}_n\right) \mathbb{I}\left(\mathbf{Y}_i \leq \mathbf{y}\right),$$

où $\bar{\xi}_n = (\xi_1 + \cdots + \xi_n)/n$. Des résultats classiques présentés par exemple dans le livre de Kosorok (2008) montrent que $\mathbb{F}_n^{\text{mult}}$ converge faiblement, conditionnellement aux données, vers un processus gaussien identique à celui de $\mathbb{F}$ et indépendant de celui-ci.

Toutedois, le ré-échantillonnage du processus $\mathbb{F}_{xh}$ n'est pas évident. En effet, plus de précautions sont nécessaires dus à l'emploi de pondérateurs $w_{n1}(x,h), \ldots, w_{nn}(x,h)$. À ce sujet, un ré-échantillonnage de type *bootstrap* a été proposé par Aerts *et coll.* (1994) dans un contexte d'estimation de centiles. Afin d'assurer la validité de cette méthode, un paramètre de lissage auxiliaire $g = g_n$ asymptotiquement plus grand que h est nécessaire. Ainsi, cette méthode consiste à se doter d'un échantillon $(\mathbf{Y}_1^\star, \ldots, \mathbf{Y}_n^\star)$ où, cette fois-ci, $\mathbf{Y}_i^\star \sim F_{x_i g}$. Ainsi, conditionnellement aux données, $\mathrm{P}(\mathbf{Y}_i^\star = \mathbf{Y}_j) = w_{nj}(x_i, g)$. Une version *bootstrap* de $\mathbb{F}_x$ est alors $\widehat{\mathbb{F}}_x^{\text{boot}} = \sqrt{nh_n}(\widehat{F}_x^{\text{boot}} - F_{xg})$, où

$$\widehat{F}_x^{\text{boot}}(\mathbf{y}) = \sum_{i=1}^{n} \mathbb{I}\left(\mathbf{Y}_i^\star \leq \mathbf{y}\right) w_{ni}(x,h).$$

Dans l'article Lemyre & Quessy (2016), deux versions de type *multiplicateur* sont proposées. Leur validité asymptotique est formellement établie. Entre autres choses, on montre comment déduire des intervalles de confiance pour des fonctionnelles statistiques générales de la forme $\Lambda(F_x)$. Ceci couvre un très large spectre, car cela inclue en particulier les moyenne, variance et covariance conditionnelles, de même que la copule conditionnelle.

0.3 Présentation et détails sur l'article présenté au Chapitre 2

Le Chapitre 2 traite du comportement asymptotique de deux estimateurs de la copule conditionnelle en présence de données sérielles. À partir d'un échantillon i.i.d. $(Y_{11}, Y_{21}, X_1), \ldots, (Y_{1n}, Y_{2n}, X_n)$, on a déjà vu comment estimer la copule conditionnelle C_x à partir de F_{xh}. Cet estimateur s'écrit

$$C_{xh}(u_1, u_2) = \Lambda^{\mathrm{C}}(F_{xh})(u_1, u_2) = \sum_{i=1}^{n} \mathbb{I}\left\{Y_{1i} \leq F_{1xh}^{-1}(u_1), Y_{2i} \leq F_{2xh}^{-1}(u_2)\right\} w_{ni}(x, h). \qquad (7)$$

Il a d'abord été étudié par Gijbels *et coll.* (2011) dans un contexte où les observations sont supposées i.i.d. Tel que l'ont mentionné par ces auteurs, cet estimateur peut être fortement biaisé, notamment lorsque la co-variable influence considérablement les distributions marginales conditionnelles. Cette constatation les a menés à proposer un second estimateur dont le but est de corriger ce problème. Pour le décrire, on procède d'abord à une transformation des observations. Ainsi, pour $i \in \{1, \ldots, n\}$, on introduit les *pseudo-observations uniformisées* $(\widetilde{U}_{1i}, \widetilde{U}_{2i}) = (F_{1X_ih_1}(Y_{1i}), F_{2X_ih_2}(Y_{2i}))$, où h_1 et h_2 sont des paramètres de lissage qui peuvent être différents de h. On définit ensuite

$$G_{xh}(u_1, u_2) = \sum_{i=1}^{n} \mathbb{I}\left(\widetilde{U}_{1i} \leq u_1, \widetilde{U}_{2i} \leq u_2\right) w_{ni}(x, h).$$

Le deuxième estimateur de C_x est alors $\widetilde{C}_{xh} = \Lambda^{\mathrm{C}}(G_{xh})$, c'est-à-dire que

$$\widetilde{C}_{xh}(u_1, u_2) = \sum_{i=1}^{n} \mathbb{I}\left\{\widetilde{U}_{1i} \leq G_{1xh}^{-1}(u_1), \widetilde{U}_{2i} \leq G_{2xh}^{-1}(u_2)\right\} w_{ni}(x, h). \qquad (8)$$

Veraverbeke *et coll.* (2011) montrent que les processus $\mathbb{C}_{xh} = \sqrt{nh}(C_{xh} - C_x)$ et $\widetilde{\mathbb{C}}_{xh} = \sqrt{nh}(\widetilde{C}_{xh} - C_x)$ convergent faiblement dans l'espace $\ell^\infty([0,1]^2)$ vers des processus gaussien $\mathbb{C}_x$ et $\widetilde{\mathbb{C}}_x$. Ces processus limites diffèrent uniquement par leurs biais respectifs, car ils possèdent la même fonction de covariance.

Puisque les résultats asymptotiques de Veraverbeke *et coll.* (2011) sont basés sur l'hypothèse de l'indépendance entre les observations, il serait intéressant de les généraliser dans un cadre de séries chronologiques. Dans l'article intitulé *On the asymptotic behavior of two estimators of the conditional copula based on time series* et soumis pour publication à *Statistics and Probability Letters*, le but est justement d'étendre la portée de ces résultats dans un cadre sériel. On suppose ainsi que les observations $(Y_{11}, Y_{21}, X_1), \ldots, (Y_{1n}, Y_{2n}, X_n)$ présentent une certaine forme de dépendance sérielle au sens où pour un délai donné $\ell \in \mathbb{N}$, les triplets (Y_{1i}, Y_{2i}, X_i) et $(Y_{1,i+\ell}, Y_{2,i+\ell}, X_{i+\ell})$ ne sont pas nécessairement indépendants.

Le cadre adopté pour ce travail est très général car il suppose que les observations proviennent d'un processus stochastique stationnaire $\{(Y_{1t}, Y_{2t}, X_t)\}_{t\in\mathbb{Z}}$ qui est α-*mélangeant.* Pour ce type de processus, la dépendance entre (Y_{1i}, Y_{2i}, X_i) et $(Y_{1,i+\ell}, Y_{2,i+\ell}, X_{i+\ell})$, à mesure que l'horizon $\ell \in \mathbb{N}$ augmente, décroit. De façon plus formelle, pour la σ-algèbre $\mathcal{F}_a^b$ engendrée par les variables aléatoires $\{(Y_{1t}, Y_{2t}, X_t)\}_{a\leq t\leq b}$, on définit alors les coefficients α-mélangeants par

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

où

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

Le processus $\{(Y_{1t}, Y_{2t}, X_t)\}_{t\in\mathbb{Z}}$ est dit α-mélangeant si $\alpha(\ell) \to 0$ lorsque $\ell \to \infty$. Dans l'article, il est démontré formellement qu'en présence d'un tel modèle, et sous certaines conditions supplémentaires sur $\alpha(\ell)$, les processus $\mathbb{C}_{xh}$ et $\widetilde{\mathbb{C}}_{xh}$ convergent vers les mêmes processus gaussiens $\mathbb{C}_x$ et $\widetilde{\mathbb{C}}_x$ obtenus dans le cas i.i.d. Pour ce faire, on montre d'abord la convergence du processus $\mathbb{F}_{xh}$. Puisque les données ne sont plus supposées i.i.d., les arguments standards en théorie des processus empiriques ne s'appliquent plus, ce qui complique considérablement l'argumentaire. En utilisant les propriétés de la fonctionnelle Λ^{C}, on en déduit la convergence de $\mathbb{C}_{xh}$. Enfin, on établit la convergence du processus $\widetilde{\mathbb{C}}_{xh}$

en démontrant que la différence entre les pseudo-observations uniformisées $(\widetilde{U}_{1i}, \widetilde{U}_{2i})$ et les vraies observations uniformisées $(F_{1X_i}(Y_{1i}), F_{2X_i}(Y_{2i}))$ est asymptotiquement négligeable.

0.4 Présentation et détails sur l'article présenté au Chapitre 3

Au Chapitre 3, on étudie des mesures non-paramétriques de causalité locale et on développe des tests de non causalité pour des données sérielles. Que ce soit en économie ou en finance, on s'intéresse souvent à la relation de causalité qui existe entre deux séries chronologiques $(Y_t)_{t\in\mathbb{Z}}$ et $(Z_t)_{t\in\mathbb{Z}}$. À cette fin, la notion de causalité telle qu'introduite par Wiener (1956) et Granger (1969) est généralement employée. Plus précisément, la causalité est définie en terme de la prédictibilité d'une variable aléatoire Y en fonction du passé de Z, étant donné le passé de Y. Autrement dit, ce concept de causalité cherche à savoir s'il est utile de connaître le passé de Z pour prédire le présent de Y, ayant déjà l'information sur le passé de Y. Pour traduire cette question de façon statistique, on note les observations disponibles jusqu'au temps t pour Y et Z par $\mathbf{Y}_t$ et $\mathbf{Z}_t$. On dira que Z *cause* Y au sens de Granger si Y_t et $\mathbf{Z}_{t-1}$ ne sont pas indépendantes conditionnellement à $\mathbf{Y}_{t-1}$. L'hypothèse nulle de non causalité peut donc se formuler par

$$\mathcal{H}_0 : Y_t \text{ et } \mathbf{Z}_{t-1} \text{ sont conditionnellement indépendants étant donné } \mathbf{Y}_{t-1}.$$

Plusieurs auteurs se sont affairés à développer des tests de non causalité, notamment à partir de tests standards d'indépendance conditionnelle. Parmi ceux-ci, on retrouve par exemple Florens & Fougere (1996). Dans un contexte où les observations sont i.i.d., de telles méthodologies ont été développées aussi par Song (2009), Huang (2010), Bergsma (2013), Su & Spindler (2013) et Linton & Gozalo (2014). Une généralisation de ces tests pour des données α-mélangeantes a été proposée par Su & White (2008, 2012). De récentes avancées ont également, entres autres, été effectuées par Bouezmarni *et coll.* (2012), Wang & Hong (2013) et Bouezmarni & Taamouti (2014).

Lorsque l'hypothèse $\mathcal{H}_0$ est rejetée, l'étape suivante consiste souvent à quantifier la force du lien de causalité existant entre les variables en jeu. À ce jour, relativement peu de mesures de causalité ont été explorées, à part celles proposées par Geweke (1982) et Geweke (1984). Certaines sont basées sur des modèles paramétriques, voir par exemple Polasek (1994), Polasek (2002) et Dufour & Taamouti (2010). Afin d'éviter le problème de commettre une erreur quant au choix du modèle, des mesures de causalité non-paramétriques basées sur l'information de Kullback–Leibler ont été proposées dans l'article de Taamouti *et coll.* (2014). Toutefois, quoique commode dans plusieurs situations, ces mesures ne décrivent que la relation de dépendance *globale* liant Y_t et $\mathbf{Z}_{t-1}$ conditionnellement à $\mathbf{Y}_{t-1}$. En effet, elles ne permettent pas de décrire convenablement cette relation lorsque la nature même du lien entre Y_t et $\mathbf{Z}_{t-1}$ change en fonction de la valeur prise par $\mathbf{Y}_{t-1}$.

Pour savoir si Z cause Y, on peut faire appel au coefficient de corrélation partielle. Ce dernier se calcule en supposant une relation auto-régressive d'ordre un pour Y et Z, à savoir que $Y_t = \alpha_1 Y_{t-1} + \epsilon_{1t}$ et $Z_t = \alpha_2 Y_t + \epsilon_{2t}$, où ϵ_{1t} et ϵ_{2t} sont des innovations de loi $\mathcal{N}(0, \sigma^2)$. Le coefficient de corrélation partiel est alors défini par la corrélation entre ϵ_{1t} et ϵ_{2t}. Sa version empirique est le coefficient de corrélation empirique basé sur les résidus de ce modèle obtenus en estimant α_1 et α_2.

Le problème potentiel quant à l'utilisation d'une mesure de causalité *globale* sera illustré à l'aide de l'étude de la causalité entre le rendement et le volume d'échanges quotidiens de l'indice Standard & Poor 500. En se basant sur le test de stationnarité effectué par Bouezmarni *et coll.* (2012), on considère les différences des logarithmes de deux journées consécutives du rendement Y et du volume Z. Les données pour l'année 2015 sont présentées à la Figure 0.4. Pour ces données, la valeur du coefficient de corrélation partiel empirique est $-0,024 \times 10^{-4}$; sur cette base, l'hypothèse de non-causalité n'est pas rejetée, car la p-valeur du test est 0,36. Toutefois, une telle conclusion est trompeuse dans la mesure où la relation entre Y_t et Z_{t-1} change en fonction de la valeur prise par Y_{t-1}. Ainsi, en considérant d'un côté le sous-échantillon constitué des triplets (Y_t, Z_t, Y_{t-1}) pour lesquelles $Y_{t-1} > 0$, la valeur de la corrélation partielle empirique est maintenant 0,072, ce qui est significativement différent de zéro car la p-valeur associée est 0,039 $<$ 0,05. D'un autre côté, le sous-échantillon formé des triplets tels que $Y_{t-1} < 0$ amène une valeur significativement négative, à savoir -0,095, à laquelle est associée une p-valeur de 0,01 $<$ 0,05.

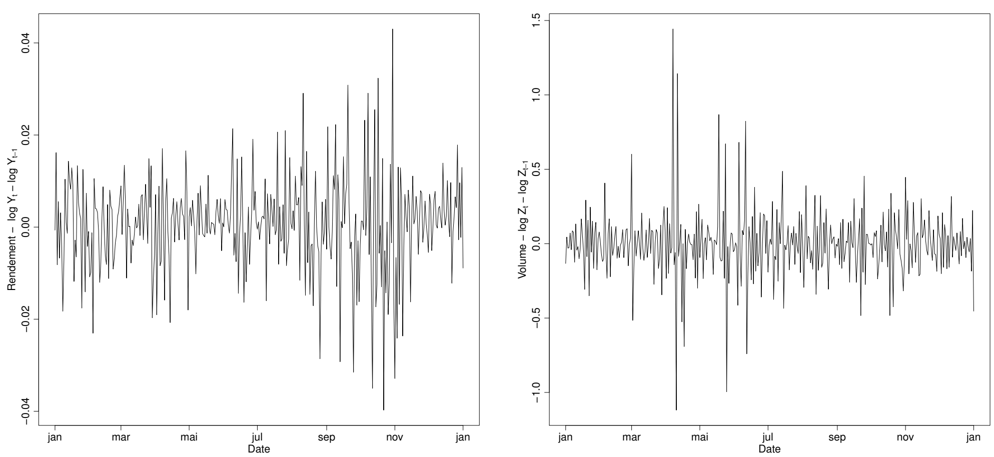


Figure 2 – Évolution des valeurs du logarithme de deux journées consécutives du rendement (à gauche) et du volume (à droite) pour les données du S&P 500 entre Janvier 2015 et Janvier 2016

Pour obtenir une analyse plus raffinée, c'est-à-dire qui tienne compte des fluctuations locales de dépendance, on a défini des mesures de causalité locale. Cette approche est ainsi plus conforme à la réalité. Le fruit de ce travail se retrouve dans l'article intitulé *Nonparametric measures of local causality and tests of local non-causality in time series*, co-écrit par Jean-François Quessy, Taoufik Bouezmarni et moi-même, et soumis à l'*Electronic Journal of Statistics*. L'idée qui y est développée consiste à mesurer la dépendance entre Y_t et Z_{t-1} conditionnellement à $Y_{t-1} = x$. L'intérêt de ces mesures est qu'elles sont définies comme fonctionnelles de la copule conditionnelle C_x. Pour décrire plus formellement notre approche, soit la fonction de répartition de causalité locale

$$H_x^{Z\to Y}(y,z) = \mathrm{P}\left(Y_t \leq y, Z_{t-1} \leq z | Y_{t-1} = x\right). \tag{9}$$

En considérant que les distributions marginales $F_{1x}(y) = \mathrm{P}(Y_t \leq y|Y_{t-1} = x)$ et $F_{2x}(z) = \mathrm{P}(Z_{t-1} \leq z|Y_{t-1} = x)$ sont continues, le Théorème de Sklar (1959) garantit l'existence d'une unique copule de causalité $C_x^{Z\to Y} : [0,1]^2 \to [0,1]$ telle que

$$H_x^{Z\to Y}(y,z) = C_x^{Z\to Y}\left\{F_{1x}(y), F_{2x}(z)\right\}.$$

On adapte ensuite les estimateurs présentés aux équations (7) et (8) au contexte de l'étude de la causalité locale. Les propriétés asymptotiques de ces estimateurs sont alors obtenues comme un cas particulier de l'article du Chapitre 3. Ensuite, la normalité asymptotique d'estimateurs non-paramétriques de mesures de causalité locale est formellement établie. On montre également comment des intervalles de confiance pour ces mesures peuvent être construits à partir d'un estimateur de la variance limite.

0.5 Présentation et détails sur l'article présenté au Chapitre 4

Le Chapitre 4 concerne l'estimation de la copule conditionnelle lorsqu'une des variables est censurée. En fait, il arrive souvent que les méthodes d'estimations développées pour des observations i.i.d. soient inapplicables en pratique, dû à la structure parfois complexe sous laquelle se présentent les données. Une telle complication survient lorsque les données ne sont pas complètement observées. Ce genre de phénomène survient régulièrement dans les cas où les données proviennent d'une étude clinique ou en gestion du risque. Pour bien illustrer de quoi il en ressort, on considère l'exemple de l'analyse du délai entre le début d'un traitement et le décès de patients atteints d'une maladie. Pour un groupe initial d'individus, il peut arriver que certains sujets quittent l'étude en cours de route, ou encore soient encore en vie à la fin de l'étude. Dans les deux cas, l'événement d'intérêt n'est pas mesuré, mais c'est plutôt un temps qui précède l'événement qui est enregistré. Lorsque cette situation se présente, on dit qu'il y a de la *censure*. Dans ce qui suit, on étudie une façon d'estimer la copule conditionnelle du couple (Y_1, Y_2) sachant $X = x$ lorsqu'une des deux variables d'intérêt est censurée, disons Y_1.

On définit premièrement les variables de censure $\mathfrak{C}_1, \ldots, \mathfrak{C}_n$ comme des variables aléatoires indépendantes telles que $G_{x_i}(t) = \mathrm{P}(\mathfrak{C}_i \leq t | X = x_i)$. On pose ensuite

$$T_i = \min\left(Y_{1i}, \mathfrak{C}_i\right) \quad \text{et} \quad \delta_i = \mathbb{I}\left(Y_{1i} \leq \mathfrak{C}_i\right).$$

La variable T_i représente la valeur enregistrée pour l'individu i ; il peut s'agir de la censure ou du décès. La variable δ_i indique s'il y a eu censure ou non. En pratique, on dispose donc des observations $(T_1, Y_{21}, \delta_1, X_1), \ldots, (T_n, Y_{2n}, \delta_n, X_n)$. L'article intitulé *Estimation of a conditional copula when a variable is subject to random right censoring*, co-écrit par Taoufik Bouezmarni, Anouar El Ghouch et moi-même, propose une méthode non-paramétrique pour estimer la copule conditionnelle en présence de ce type de censure. D'une certaine manière, cet estimateur est une adaptation naturelle de l'estimateur C_{xh} proposé par Gijbels *et coll.* (2011) au contexte de données censurées. Une analyse semblable a été effectuée récemment par Gribkova & Lopez (2015), mais pour l'estimation de la copule dans un cadre non conditionnel.

Puisque l'analyse du cadre conditionnel est développée de façon similaire, on présente ici un bref exposé des résultats lorsqu'il n'y a pas de co-variable. D'abord, suivant une idée originalement proposée par Robins & Rotnitzky (1992), l'estimation non-paramétrique de la fonction de répartition univariée $F_1(t) = \mathrm{P}(Y_{1i} \leq t)$ peut s'effectuer en accordant un poids supplémentaire, égal à l'inverse de la probabilité de non censure, aux observations non censurées. Cette approche se justifie en notant premièrement que comme

$$\mathrm{E}\left\{\mathbb{I}(T_i \leq t)\,\frac{\delta_i}{1-\mathrm{P}(\mathfrak{C}_i \leq T_i)}\right\} = F_1(t),$$

un estimateur potentiel de F_1 est

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(T_i \leq t)\,\frac{\delta_i}{1-\mathrm{P}(\mathfrak{C}_i \leq T_i)}\,.$$

Toutefois, cet estimateur n'est pas calculable car la distribution des censures $C_1, \ldots, C_n$ est inconnue. Heureusement, celle-ci peut s'estimer à l'aide de l'estimateur de Kaplan–Meier, à savoir

$$G_n^{(\mathrm{c})}(t) = 1 - \prod_{T_{(i)} \leq t}\left(1 - \frac{i}{n-i+1}\right)^{1-\delta_{[i]}},$$

où $T_{(1)} \leq \cdots \leq T_{(n)}$ sont les observations ordonnées et $\delta_{[i]}$ est la variable indicatrice de censure associée à $T_{(i)}$. L'estimateur tel que proposé par Robins & Rotnitzky (1992) s'écrit alors

$$F_{1n}^{(\mathrm{c})}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left(T_{1i} \leq t\right)\,\frac{\delta_i}{1-G_n^{(\mathrm{c})}(Y_{1i})}. \tag{10}$$

Il a été montré dans l'article Satten & Datta (2001) que cet estimateur est équivalent à l'estimateur de Kaplan–Meier pour la variable Y_1, c'est-à-dire que

$$F_{1n}^{(\mathrm{c})}(t) = 1 - \prod_{T_{(i)} \leq t}\left(1 - \frac{i}{n-i+1}\right)^{\delta_{[i]}}.$$

Cette façon de ré-écrire l'estimateur de Kaplan–Meier suggère d'estimer la fonction de répartition conjointe de (Y_1, Y_2) lorsque seule Y_1 est censurée par

$$F_n^{(\mathrm{c})}(t, y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left(T_i \leq t, Y_{2i} \leq y\right)\,\frac{\delta_i}{1-G_n^{(\mathrm{c})}(Y_{1i})}. \tag{11}$$

Un estimateur de la copule de (Y_1, Y_2) est alors obtenu via

$$C_n^{(\mathrm{c})}(u_1, u_2) = F_n^{(\mathrm{c})}\left\{F_{1n}^{(\mathrm{c})-1}(u_1), F_{2n}^{(\mathrm{c})-1}(u_2)\right\},$$

où $F_{1n}^{(c)}(t) = F_n^{(c)}(t, \infty)$ et $F_{2n}^{(c)}(y) = F_n^{(c)}(\infty, y)$. Cet estimateur a été proposé par Gribkova & Lopez (2015) dans un contexte un peu plus général. De plus, puisque la variable Y_2 est complètement observée, on pourrait remplacer $F_{2n}^{(c)}$ par la fonction de répartition empirique de Y_2. À noter qu'il est possible de généraliser l'estimateur $F_n^{(c)}$ au cas où Y_1 et Y_2 sont toutes deux censurées ; pour plus de détails à ce sujet, voir Gribkova & Lopez (2015).

Pour estimer la copule conditionnelle en présence de censure, nous avons adapté cette idée à la présence d'une co-variable. Ainsi, un estimateur non-paramétrique de la fonction de répartition conjointe conditionnelle est d'abord proposé. On obtient alors un estimateur de la copule conditionnelle à partir de celui-ci. La convergence des processus associés est formellement démontrée sous certaines conditions imposées aux paramètres de lissage.

0.6 Présentation et détails sur la contribution présentée au Chapitre 5

On a vu que la copule conditionnelle décrit complètement la dépendance conditionnelle entre les composantes d'un vecteur aléatoire en présence d'une co-variable. De ce fait, il devient pertinent de s'interroger sur la façon dont la valeur que prend X influence l'association entre deux variables aléatoires Y_1 et Y_2. Par exemple, peut-on affirmer que Y_1 et Y_2 sont moins dépendantes lorsque $X = x$ que lorsque $X = y$? Cette question appelle à la notion *d'ordre de concordance* entre deux copules. Formellement, soient C_x et C_y, les copules conditionnelles qui caractérisent la dépendance entre Y_1 et Y_2 lorsque $X = x$ et $X = y$, respectivement. Selon Nelsen (2006), on dit que C_x et C_y sont concordantes lorsque

$$C_x(u_1, u_2) \leq C_y(u_1, u_2) \quad \text{pour tout } (u_1, u_2) \in [0, 1]^2.$$

Lorsque tel est le cas, on écrira simplement $C_x \preceq C_y$. On dira alors que les composantes du couple (Y_1, Y_2) sont moins dépendantes lorsque $X = x$ que lorsque $X = y$. Toutefois, $C_x \not\preceq C_y$ n'implique pas nécessairement $C_y \preceq C_x$. Ainsi, l'ordre de concordance $\preceq$ défini sur l'ensemble des copules est un ordre partiel.

La plupart des mesures de dépendances utilisées en pratique préservent l'ordre de concordance. Par exemple, on vérifie aisément que $C_x \preceq C_y$ implique $\tau_x \leq \tau_y$ et $\rho_x \leq \rho_y$. Or, l'inverse est généralement faux. Pour illustrer, prenons la copule normale de paramètre $\varrho \in [-1, 1]$ définie par

$$C_\varrho^{\mathrm{N}}(u_1, u_2) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \varphi_\varrho(z_1, z_2)\, \mathrm{d}z_2\, \mathrm{d}z_1,$$

où φ_ϱ est la densité de la loi normale bivarié standard de corrélation ϱ et Φ^{-1} est la fonction centile de la loi $\mathbb{N}(0, 1)$. On considère aussi la copule de Clayton

$$C_\gamma^{\mathrm{CL}}(u_1, u_2) = \left(u_1^{-\gamma} + u_2^{-\gamma} - 1\right)^{-1/\gamma}, \quad \gamma > 0.$$

Il existe des relations intéressantes entre les paramètres ϱ et γ et la mesure d'association de Kendall. Ainsi, on a

$$\tau_\varrho^{\mathrm{N}} = \frac{2}{\pi} \arcsin(\varrho) \quad \text{et} \quad \tau_\theta^{\mathrm{CL}} = \frac{\theta}{\theta + 2}\,. \tag{12}$$

En posant $C_x = C_{0{,}59}^{\mathrm{N}}$ et $C_y = C_2^{\mathrm{CL}}$, ces formules permettent d'établir que $\tau_x = 0.4 < \tau_y = 0.5$. Or, $C_x \not\preceq C_y$ car, par exemple, $C_x(0{,}9, 0{,}9) = 0{,}838$ et $C_y(0{,}9, 0{,}9) = 0{,}825$.

Le Chapitre 5 propose des méthodes statistiques pour tester formellement les hypothèses

$$\mathcal{H}_0 : C_x \preceq C_y \quad \text{et} \quad \mathcal{H}_0 : C_x \not\preceq C_y.$$

À cette fin, des tests universellement convergents sont proposés. De plus, puisque $C_x \preceq C_y$ implique $\rho_x \leq \rho_y$ et $\tau_x \leq \tau_y$, on considère aussi des tests pour $\mathcal{H}_0$ basés sur le rho de Spearman et le tau de Kendall. Comme l'hypothèse nulle ne présuppose aucun modèle spécifique, il devient avantageux de recourir à des méthodes non-paramétriques pour la construction des tests. On considère donc des estimateurs basés sur des méthodes à noyau. Aussi, pour déterminer les régions de rejets des différentes statistiques de test, on a recours à une des méthodes de ré-échantillonnage du multiplicateur adaptées aux distributions conditionnelles telles que décrites au Chapitre 1.

CHAPITRE 1

Multiplier bootstrap methods for conditional distributions

Résumé

La méthode du multiplicateur est une alternative efficace et facile à implémenter au bootstrap traditionnel. Cette technique a fait l'objet de plusieurs utilisations fructueuses dans différents domaines de la statistique. Dans cet article, des méthodes de ré-échantillonnage basées sur des multiplicateurs sont proposées dans le cadre général de l'étude d'un vecteur aléatoire $\mathbf{Y} \in \mathbb{R}^d$ conditionnellement à une co-variable $X \in \mathbb{R}$. En fait, deux versions dites « du multiplicateur » seront proposées pour la fonction de distribution conditionnelle empirique, et leur validité théorique sera formellement établie. De plus, puisque la méthode du multiplicateur se conjugue de manière élégante à la méthode delta fonctionnelle, la théorie entourant l'estimation de fonctionnelles statistiques est développée en conséquence. Nous traitons en outre de la façon de construire des intervalles de confiance pour la moyenne et la variance conditionnelle, pour le coefficient de corrélation conditionnelle, pour de tau de Kendall ainsi que la copule conditionnelle. Nous considérons aussi

des schèmes d'inférence composite pour la fonction de distribution conditionnelle univariée et multi-variée. Enfin, nous examinons les performances des différentes méthodes développées au travers d'une étude de simulation.

Abstract

The multiplier bootstrap is a fast and easy-to-implement alternative to the standard bootstrap ; it has been used successfully in many statistical contexts. In this paper, resampling methods based on multipliers are proposed in a general framework where one investigates the stochastic behavior of a random vector $\mathbf{Y} \in \mathbb{R}^d$ conditional on a covariate $X \in \mathbb{R}$. Specifically, two versions of the multiplier bootstrap adapted to empirical conditional distributions are introduced as alternatives to the conditional bootstrap and their asymptotic validity is formally established. As the method walks hand-in-hand with the functional delta method, theory around the estimation of statistical functionals is developed accordingly ; this includes the interval estimation of conditional mean and variance, conditional correlation coefficient, Kendall's dependence measure and copula. Composite inference about univariate and joint conditional distributions is also considered. The sample behavior of the new bootstrap schemes and related estimation methods are investigated via simulations and an illustration on real data is provided.

1.1 Introduction

Suppose that one is interested in the stochastic behavior of a random vector $\mathbf{Y} \in \mathbb{R}^d$ given some covariate X taking values in $\mathbb{R}$. This setup may occur in many statistical applications. For example, when $d = 2$, Gijbels *et al.* (2011) noted that the relationship between the life expectancy of men (Y_1) and women (Y_2) strongly depends on the gross domestic product (X). When $d = 1$, a popular approach is to use nonparametric regression, where typically one assumes a relationship of the form $Y = m(X) + \sigma(X)\,\varepsilon$, where m and σ are smooth functions to be estimated and ε is a standardized random variable[1]. The kernel estimation of m can be traced back to the works of Watson (1964), Priestley & Chao (1972), Benedetti (1977) and Gasser & Müller (1979). An alternative and more general approach for capturing this relationship is to consider the whole conditional distribution of $\mathbf{Y}$ given $X = x$ either via its cdf $F_x(\mathbf{y}) = \mathrm{P}(\mathbf{Y} \leq \mathbf{y} | X = x)$ or its related d-dimensional conditional density f_x, provided it exists. Since many inferential procedures are based on statistical functionals of the conditional distribution function, the estimation of F_x is of a particular interest. For example, the conditional mean $m(x)$ in the above-mentioned regression model can be recovered from F_x.

Suppose a fixed design where $(\mathbf{Y}, X)$ is observed at $X \in \{x_1, \ldots, x_n\}$. From the sample of independent vectors $(\mathbf{Y}_1, x_1), \ldots, (\mathbf{Y}_n, x_n)$, a nonparametric kernel estimator of F_x is given by

$$F_{xh}(\mathbf{y}) = \sum_{i=1}^{n} \mathbb{I}\left(\mathbf{Y}_i \leq \mathbf{y}\right) w_{ni}(x, h_n), \tag{1.1}$$

where $w_{n1}(x, h_n), \ldots, w_{nn}(x, h_n)$ is a sequence of weights that smooth the covariate space and $h = h_n > 0$ is a bandwidth parameter that depends on the sample size ; typically, w_{ni} depends on x_i. It will be assumed in the sequel that $w_{n1}(x, h_n) + \cdots + w_{nn}(x, h_n) = 1$. The asymptotic behavior of F_{xh} as the sample size n goes to infinity essentially depends on assumptions imposed to the weights. Specifically, under Assumptions W_1–W_5 listed in Appendix E.1 and as long as $nh_n \to \infty$ and $nh_n^5 \to K^2 < \infty$, one has for a fixed $x \in \mathbb{R}$ that the empirical function $\mathbb{F}_{xh} = \sqrt{nh_n}(F_{xh} - F_x)$ converges weakly to a Gaussian limit of the form $\mathbb{F}_x = \alpha_x + B_x$, where α_x is a centered Gaussian process on $\mathbb{R}^d$ and B_x is an asymptotic bias function that results from the kind of smoothing that is made around x. More details on this result are given at the beginning of Section 2.

1. For the sake of clarity, this sentence has been modified from its original version published in *Statistics and Computing*.

As will be seen, the limit process $\mathbb{F}_x$ is cumbersome and more importantly, depends on the unknown form of F_x. Therefore, a crucial issue in view of the development of formal inferential procedures for conditional distributions and related statistical functionals is to be able to replicate the stochastic behavior of $\mathbb{F}_{xh}$. This task is not as straightforward as in the unconditional case, where the standard bootstrap can generally do the job. Among the bootstrap strategies that have been proposed in a nonparametric regression context, one can cite Härdle & Bowman (1988) and Hardle & Marron (1991) using non-pivotal techniques and Hall (1992) based on pivotal quantities ; see also the more recent contribution by McMurry & Politis (2008), where a bias correction is suggested.

In this article, a more general perspective is adopted in which one aims at resampling the whole conditional distribution function. Specifically, two methods based on suitably adapted versions of the multiplier bootstrap procedure are proposed. The newly introduced techniques are easily implementable and feature the fact that each element of the original sample contributes to every bootstrap samples. In addition, the fact that a distribution function point-of-view is taken allows for the bootstrapping of many statistical quantities of interest, including smooth functions in nonparametric regression, as well as conditional measures of association and estimation of conditional copulas. The resampling strategies proposed here will be compared to a conditional bootstrap method described by Omelka *et al.* (2013) in the context of conditional copula estimation ; the latter method was itself inspired by Aerts *et al.* (1994) for the estimation of regression quantiles. Specifically, this paper's goals are to

(i) provide extensions of the multiplier bootstrap for conditional distributions ;

(ii) develop the theory in conjunction with the functional delta method in order to provide a general methodology for interval estimation of Hadamard differentiable functionals ;

(iii) see how to infer about conditional distributions, *i.e.* construct confidence bands and design tests for composite hypotheses ;

(iv) investigate the sample properties and usefulness of the methodologies.

The paper is organized as follows. The two proposed multiplier methods for $\mathbb{F}_{xh}$ are introduced and their asymptotic validity is established in Section 2. A general framework for resampling Hadamard differentiable statistical functionals is developed in Section 3 ; details are also given for the estimation of the conditional mean and variance in non-parametric regression, as well as some conditional measures of association and copula. How to make inference about conditional distributions is explained in Section 4. The sampling properties of the newly introduced bootstraps and related estimation methods are carefully investigated in Section 5 with the help of Monte Carlo simulations under various scenarios. In Section 6, the methodology is illustrated on real data. Assumptions and complementary computations are relegated to two appendices and the proofs of the main results are to be found in the online resource.

1.2 Resampling conditional distribution functions

1.2.1 Asymptotic behavior of F_{xh}

For the remaining of the paper, a fixed design where $(\mathbf{Y}, X)$ is observed at $X \in \{x_1, \ldots, x_n\}$ is assumed. Specifically, one has a random sample of independent vectors $(\mathbf{Y}_1, x_1)$, ..., $(\mathbf{Y}_n, x_n)$. The methodologies developed in this work are also valid under random designs, *i.e.* when X is a random variable. The adaptation to a random design would simply consist in replacing x_i by X_i and using O_{P} and o_{P} instead of O and o arguments in the assumptions stated in Appendix E.1.

As already mentioned in the Introduction, the random function $\mathbb{F}_{xh} = \sqrt{nh_n}(F_{xh} - F_x)$ converges weakly to a limit process of the form $\mathbb{F}_x = \alpha_x + B_x$ as long as $nh_n \to \infty$, $nh_n^5 \to$

$K^2 < \infty$ and Assumptions W_1–W_5 are satisfied. As identified by Veraverbeke *et al.* (2011), α_x is a centered Gaussian process on $\mathbb{R}^d$ such that for each $\mathbf{y}, \mathbf{y}' \in \mathbb{R}^d$,

$$\Gamma_x(\mathbf{y}, \mathbf{y}') \;=\; \text{Cov}\left\{\alpha_x(\mathbf{y}), \alpha_x(\mathbf{y}')\right\} = K_4 \left\{F_x(\mathbf{y} \wedge \mathbf{y}') - F_x(\mathbf{y})\, F_x(\mathbf{y}')\right\},$$

where $K_4 \in (0, \infty)$ is defined in Assumption W_4 and $\mathbf{r}_1 \wedge \mathbf{r}_2$ is the componentwise minimum between $\mathbf{r}_1, \mathbf{r}_2 \in \mathbb{R}^d$. Moreover, by a Taylor expansion, one can show that the asymptotic bias is given for $\dot{F}_x = \partial F_x / \partial x$ and $\ddot{F}_x = \partial^2 F_x / \partial^2 x$ by

$$B_x(\mathbf{y}) \;=\; \lim_{n\to\infty} \sqrt{nh_n} \sum_{i=1}^{n} \left\{F_{x_i}(\mathbf{y}) - F_x(\mathbf{y})\right\} w_{ni}(x, h_n) = K \left\{K_2\, \dot{F}_x(\mathbf{y}) + K_3\, \ddot{F}_x(\mathbf{y})\right\},$$

where $K_2, K_3 \in (0, \infty)$ are given in Assumptions W_2–W_3. The expression is valid under the uniform continuity in $(z, \mathbf{y})$ of $\dot{F}_z(\mathbf{y})$, $\ddot{F}_z(\mathbf{y})$ in a neighborhood of x.

The next three subsections describe procedures for the replication of $\mathbb{F}_{xh}$. Essentially, the goal is to define empirical functions $\widehat{\mathbb{F}}_x^{(1)}, \ldots, \widehat{\mathbb{F}}_x^{(B)}$ that are independent and asymptotically equivalent to the limit $\mathbb{F}_x = \alpha_x + B_x$. These replicates of $\mathbb{F}_{xh}$ will depend on two bandwidth parameters, namely $h = h_n$ and $g = g_n$; the latter is required for the purpose of resampling. In order to ease notation, however, these subscripts are omitted in the sequel. The first of these methods is a version of the standard bootstrap adapted to conditional distributions; this idea was already explored by Omelka *et al.* (2013) in the context of conditional copula estimation. The other two methods are based on the multiplier bootstrap.

1.2.2 Conditional bootstrap

An extension of the standard bootstrap for the estimation of regression quantiles was proposed by Aerts *et al.* (1994). In this procedure, a bootstrap sample $(\mathbf{Y}_1^\star, \ldots, \mathbf{Y}_n^\star)$ is drawn, where $\mathbf{Y}_i^\star \sim F_{x_i g_n}$, *i.e.* conditionally on the data, $\mathrm{P}(\mathbf{Y}_i^\star = \mathbf{Y}_j) = w_{nj}(x_i, g_n)$. Here,

$g = g_n$ is a bandwidth parameter such that as $n \to \infty$, $g_n \to 0$ and $n^{1-\delta} g_n^5 \to \infty$ for some $\delta > 0$; hence, g_n is asymptotically larger than h_n since $h_n/g_n = o(1)$. A bootstrap version of $\mathbb{F}_{xh}$ is obtained by defining $\widehat{\mathbb{F}}_x^{\text{boot}} = \sqrt{nh_n}(\widehat{F}_x^{\text{boot}} - F_{xg})$, where

$$\widehat{F}_x^{\text{boot}}(\mathbf{y}) = \sum_{i=1}^{n} \mathbb{I}\left(\mathbf{Y}_i^\star \leq \mathbf{y}\right) w_{ni}(x, h_n).$$

The empirical function F_x^{boot} can be written in terms of the original sample $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ upon introducing for $i, j \in \{1, \ldots, n\}$ the Bernoulli random variables $M_{ig}^{(j)}$ such that $M_i^{(j)} = 1$ if $\mathbf{Y}_i$ is selected at the j-th drawing, and $M_i^{(j)} = 0$ otherwise. One can then write

$$\begin{aligned} \widehat{F}_x^{\text{boot}}(\mathbf{y}) &= \sum_{i=1}^{n} \sum_{j=1}^{n} M_j^{(i)} \, \mathbb{I}\left(\mathbf{Y}_j \leq \mathbf{y}\right) w_{ni}(x, h_n) \\ &= \sum_{j=1}^{n} \mathbb{I}\left(\mathbf{Y}_j \leq \mathbf{y}\right) \left\{ \sum_{i=1}^{n} M_j^{(i)} \, w_{ni}(x, h_n) \right\} = \sum_{j=1}^{n} \mathbb{I}\left(\mathbf{Y}_j \leq \mathbf{y}\right) M_{jx}^{\text{boot}}, \quad (1.2) \end{aligned}$$

where $M_{1x}^{\text{boot}}, \ldots, M_{nx}^{\text{boot}}$ are the random variables

$$M_{ix}^{\text{boot}} = \sum_{j=1}^{n} M_i^{(j)} w_{nj}(x, h_n), \quad i \in \{1, \ldots, n\}. \qquad (1.3)$$

The bootstrap version of $\mathbb{F}_{xh}$ can then be written

$$\widehat{\mathbb{F}}_x^{\text{boot}}(\mathbf{y}) \;=\; \sqrt{nh_n} \sum_{i=1}^{n} \mathbb{I}(\mathbf{Y}_i \leq \mathbf{y}) \left\{ M_{ix}^{\text{boot}} - w_{ni}(x, g_n) \right\}. \qquad (1.4)$$

The validity of this bootstrap method is formally established by Omelka *et al.* (2013) under the same assumptions that ensure the weak convergence of $\mathbb{F}_{xh}$ to $\mathbb{F}_x$, in addition to Assumptions W_1'–W_3' and W_1''–W_6''.

A multiplier version of this bootstrap method will be proposed in subsection 2.4, where it will be helpful to note that

$$\begin{aligned} \sum_{i=1}^{n} M_{ix}^{\text{boot}} &= \sum_{i=1}^{n} \left\{ \sum_{j=1}^{n} M_i^{(j)} \, w_{nj}(x, h_n) \right\} \\ &= \sum_{j=1}^{n} w_{nj}(x, h_n) \left\{ \sum_{i=1}^{n} M_i^{(j)} \right\} = \sum_{j=1}^{n} w_{nj}(x, h_n) = 1. \end{aligned}$$

Starting from (1.5), one obtains easily

$$\widehat{\mathbb{F}}_x^{\mathrm{boot}}(\mathbf{y}) = \sqrt{nh_n} \sum_{i=1}^n \{\mathbb{I}(\mathbf{Y}_i \leq \mathbf{y}) - F_{xg}(\mathbf{y})\} \, M_{ix}^{\mathrm{boot}}. \tag{1.5}$$

1.2.3 Independent and identically distributed multipliers

Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be i.i.d. from a d-variate distribution function F. In that case, F may be estimated by the empirical distribution function F_n. A way to replicate the asymptotic behavior of $\mathbb{F}_n = \sqrt{n}(F_n - F)$ is to use the so-called *multiplier method*. The strategy consists in generating, independently of the data, a random vector $(\xi_1, \ldots, \xi_n)$ of independent multiplier random variables. Many versions of the method exist depending on the nature of the multiplier variables ; for details, see Kosorok (2008). A special case occurs when $\xi_1, \ldots, \xi_n$ are positive and such that $\mathrm{E}(\xi_i) = \mathrm{Var}(\xi_i) = 1$ for each $i \in \{1, \ldots, n\}$. The multiplier bootstrap version of $\mathbb{F}_n$ is then

$$\widehat{\mathbb{F}}(\mathbf{y}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{\mathbb{I}(\mathbf{Y}_i \leq \mathbf{y}) - F_n(\mathbf{y})\} \, \xi_i.$$

The above multiplier method will be adapted to the conditional setup. The first step consists in replicating the centered part α_x of $\mathbb{F}_x$; in the second step, the deterministic bias B_x will be consistently estimated. Formally, proceed as in the unconditional case and consider a vector $(\xi_1, \ldots, \xi_n)$ of positive and independent random variables with unit mean and variance. Further assume that for each $i \in \{1, \ldots, n\}$, one has for any $\delta > 0$ that $\mathrm{E}(\xi_i^{2+\delta}) < \infty$. Upon noting that α_x is the weak limit of the centered process

$$\alpha_{xh} = \sqrt{nh_n}\{F_{xh} - \mathrm{E}(F_{xh})\},$$

define its multiplier version by

$$\widehat{\alpha}_x(\mathbf{y}) = \sqrt{nh_n} \sum_{i=1}^n \{\mathbb{I}\,(\mathbf{Y}_i \leq \mathbf{y}) - F_{xh}(\mathbf{y})\} \, w_{ni}(x, h_n) \, \xi_i.$$

In the sequel, $\mathrm{P}^\star$ is the multiplier bootstrap probability measure and $\mathrm{E}^\star$ is the expectation with respect to the multiplier random variables only. In other words, $\mathrm{P}^\star$ and $\mathrm{E}^\star$ are computed conditional on the data.

Proposition 1 *Suppose $\dot{F}_z$, $\ddot{F}_z$ are uniformly continuous in $(z, \mathbf{y})$ and Assumptions W_1, W_4–W_6 are satisfied. If $nh_n \to \infty$ and $nh_n^5 \to K^2 < \infty$ as $n \to \infty$, then $(\alpha_{xh}, \widehat{\alpha}_x)$ converges in multiplier bootstrap probability measure* $\mathrm{P}^\star$ [P]*–almost surely to $(\alpha_x, \widetilde{\alpha}_x)$, where $\widetilde{\alpha}_x$ is an independent copy of the limit α_x of α_{xh}.*

In the second part of the proof of Proposition 1, it is shown that $\widehat{\alpha}_x$ has the same asymptotic covariance structure as $\mathbb{F}_x$. However, observe that $\mathrm{E}^\star\{\widehat{\alpha}_x(\mathbf{y})\} = 0$ for all $\mathbf{y} \in \mathbb{R}^d$. Hence, there remains the problem of estimating the asymptotic bias. To this end, let $g = g_n$ be a bandwidth parameter satisfying the assumptions described in subsection 2.2. In view of equation (1.2), an estimator of the bias function B_x is given by

$$\widehat{B_x}(\mathbf{y}) = \sqrt{nh_n} \sum_{i=1}^{n} \left\{F_{xig}(\mathbf{y}) - F_{xg}(\mathbf{y})\right\} w_{ni}(x, h_n). \tag{1.6}$$

Proposition 2 *Suppose $\dot{F}_z$, $\ddot{F}_z$ are uniformly continuous in $(z, \mathbf{y})$ and Assumptions W_2–W_3, W_1'–W_3' and W_1''–W_6'' are satisfied. Then* [P]*–almost surely,*

$$\sup_{\mathbf{y} \in \mathbb{R}^d} \left| \widehat{B_x}(\mathbf{y}) - B_x(\mathbf{y}) \right| \overset{n \to \infty}{\longrightarrow} 0.$$

Propositions 1–2 suggest to define $\widehat{\mathbb{F}}_x^{\mathrm{mult1}} = \widehat{\alpha}_x + \widehat{B_x}$ as an appropriate multiplier version for $\mathbb{F}_{xh}$. Its asymptotic validity follows easily under Assumptions W_1–W_6, W_1'–W_3' and W_1''–W_6''. By straightforward computations,

$$\widehat{\mathbb{F}}_x^{\mathrm{mult1}}(\mathbf{y}) = \sqrt{nh_n} \sum_{i=1}^{n} \mathbb{I}(\mathbf{Y}_i \le \mathbf{y})\, M_{ix}^{\mathrm{mult1}},$$

where for $\bar{\xi}_x = w_{n1}(x,h_n)\,\xi_1 + \cdots + w_{nn}(x,h_n)\,\xi_n$,

$$M_{ix}^{\text{mult1}} \;=\; \left(\xi_i - \bar{\xi}_x\right) w_{ni}(x,h_n) + \sum_{j=1}^n w_{ni}(x_j,g_n)\, w_{nj}(x,h_n) - w_{ni}(x,g_n).$$

1.2.4 Multipliers that depend on the covariate

A second multiplier approach is inspired by the conditional bootstrap described in subsection 2.2. The basic idea is to define multiplier variables whose first two moments match those of $M_{1x}^{\text{boot}}, \ldots, M_{nx}^{\text{boot}}$ defined in (1.3); these random variables appear in the definition of F_{xh}^{boot} in equation (1.2). One can show by straightforward computations that $\mathrm{E}^\star(M_{ix}^{\text{boot}}) = p_{ix}$ and $\mathrm{Var}^\star(M_{ix}^{\text{boot}}) = v_{ix} - v_{iix}$, where

$$\begin{aligned}
p_{ix} &= \sum_{j=1}^n w_{ni}(x_j,g_n)\, w_{nj}(x,h_n), \\
v_{ix} &= \sum_{j=1}^n w_{ni}(x_j,g_n)\, \{w_{nj}(x,h_n)\}^2, \\
v_{ii'x} &= \sum_{j=1}^n w_{ni}(x_j,g_n)\, w_{ni'}(x_j,g_n)\, \{w_{nj}(x,h_n)\}^2.
\end{aligned}$$

Then, consider the *covariate-dependent* multiplier random vector $(\xi_{1x}, \ldots, \xi_{nx})$ of positive and independent random variables such that $\mathrm{E}^\star(\xi_{ix}) = p_{ix}$ and $\mathrm{Var}^\star(\xi_{ix}) = v_{ix} - v_{iix}$. Replacing M_{ix}^{boot} by ξ_{ix} in the conditional bootstrap version of $\mathbb{F}_{xh}$ in (1.5), one obtains the alternative multiplier bootstrap version

$$\begin{aligned}
\widehat{\mathbb{F}}_x^{\text{mult2}}(\mathbf{y}) &= \sqrt{nh_n} \sum_{i=1}^n \{\mathbb{I}(\mathbf{Y}_i \le \mathbf{y}) - F_{xg}(\mathbf{y})\}\, \xi_{ix} \\
&= \sqrt{nh_n} \sum_{i=1}^n \mathbb{I}(\mathbf{Y}_i \le \mathbf{y})\, \{\xi_{ix} - \xi_{\cdot x}\, w_{ni}(x,g_n)\},
\end{aligned}$$

where $\xi_{\cdot x} = \xi_{1x} + \cdots + \xi_{nx}$. Note that as long as Assumption W_1 holds, $v_{ii'x} = v_{ix} \times o((nh_n)^{-1/2})$, so that v_{iix} is asymptotically negligible compared to v_{ix}. One can therefore consider that $\mathrm{Var}^\star(\xi_{ix}) = v_{ix} - v_{iix} \approx v_{ix}$.

Proposition 3 *Suppose $\dot{F}_z$, $\ddot{F}_z$ are uniformly continuous in $(z, \mathbf{y})$, Assumptions W_1–W_6, W_1'–W_3' and W_1''–W_6'' are fulfilled, and $\sqrt{nh_n} \max_{1 \le i \le n} \xi_{ix} = o_{\mathrm{P}^\star}(1)$ [P]–almost surely. Then $(\mathbb{F}_{xh}, \widehat{\mathbb{F}}_x^{\mathrm{mult2}})$ converges in multiplier bootstrap probability measure $\mathrm{P}^\star$ [P]–almost surely to $(\mathbb{F}_x, \widetilde{\mathbb{F}}_x)$, where $\widetilde{\mathbb{F}}_x$ is an independent copy of $\mathbb{F}_x$.*

The requirement that $\sqrt{nh_n} \max_{1 \le i \le n} \xi_{ix} = o_{\mathrm{P}^\star}(1)$ in Proposition 3 can be translated into a moment condition on ξ_{ix}. Indeed, since from Assumption W_1, $p_{ix} = o((nh_n)^{-1/2})$, this is true whenever for all $\epsilon > 0$, one can find $\nu > 1$ such that

$$\mathrm{P}^\star \left(\sqrt{nh_n} \max_{1 \le i \le n} \xi_{ix} > \epsilon \right) \le \left(\frac{nh_n}{\epsilon^2} \right)^{\nu/2} \sum_{i=1}^{n} \mathrm{E}^\star \left\{ (\xi_{ix} - p_{ix})^\nu \right\}.$$

The following lemma indicates that the condition is satisfied for Gamma distributed multiplier variables.

Lemma 1 *Let ξ_{ix} be Gamma with parameters $\alpha = p_{ix}^2 / v_{ix}$ and $\beta = v_{ix}/p_{ix}$. Then if Assumption W_1 and Assumption W_4 hold, one has [P]–almost surely that*

$$\sqrt{nh_n} \max_{1 \le i \le n} \xi_{ix} = o_{\mathrm{P}^\star}(1).$$

1.2.5 Unification of the three resampling techniques

Note that the three resampling methods encountered so far yield bootstrap versions of the form

$$\widehat{\mathbb{F}}_x(\mathbf{y}) = \sqrt{nh_n} \sum_{i=1}^{n} \mathbb{I}(\mathbf{Y}_i \le \mathbf{y})\, L_{ix}, \tag{1.7}$$

where $L_{1x}, \ldots, L_{nx}$ are random variables depending on the resampling scheme that is used. For the conditional bootstrap, $L_{ix} = M_{ix}^{\mathrm{boot}} - w_{ni}(x, g_n)$; for the first multiplier method, $L_{ix} = M_{ix}^{\mathrm{mult1}}$; for the second multiplier method, $L_{ix} = \xi_{ix} - \xi_{\cdot x}\, w_{ni}(x, g_n)$. In a certain way, these random variables are chosen such that asymptotically, the first two moments of $\widehat{\mathbb{F}}_x$, *i.e.* the bias function and the covariance structure, match those of the limit $\mathbb{F}_x$ of $\mathbb{F}_{xh}$. From computations in subsections 1.1.2 and 1.3.2 of the Supplementary material, this is indeed the case for the first and second multiplier methods.

In the sequel, $\widehat{\mathbb{F}}_x$ will refer to one of the three above-mentioned resampling methods. In order to further ease notation, and at the same time provide easily implementable formulas based on products of vectors, define

$$\begin{aligned}
\mathbf{w}_x &= (w_1(x,h_n),\ldots,w_n(x,h_n))\,, \\
\mathbf{L}_x &= (L_{1x},\ldots,L_{nx})\,, \\
\mathcal{I}(\mathbf{y}) &= (\mathbb{I}(\mathbf{Y}_1 \le \mathbf{y}),\ldots,\mathbb{I}(\mathbf{Y}_n \le \mathbf{y}))\,, \quad \mathbf{y} \in \mathbb{R}^d.
\end{aligned}$$

With this notation, one can write

$$F_{xh}(\mathbf{y}) = \mathcal{I}(\mathbf{y})\,\mathbf{w}_x^\top \quad \text{and} \quad \widehat{\mathbb{F}}_x(\mathbf{y}) = \sqrt{nh_n}\,\mathcal{I}(\mathbf{y})\,\mathbf{L}_x^\top.$$

1.3 Hadamard differentiable statistical functionals

In many applications, one is interested in a real-valued quantity that can be expressed as a functional of F_x; this is called a statistical functional. In this section, the theory is developed for both linear and nonlinear mappings that are Hadamard differentiable; several examples of application are then detailed. Recall that a map $\Lambda : D \to E$, where D and E are normed spaces, is said to be Hadamard differentiable at $\delta \in D$ tangentially to $D_0 \subset D$ if there exists a continuous linear mapping $\Lambda'_\delta : D_0 \to E$ such that for $t_n \to 0$ and $\Delta_n \to \Delta$,

$$\lim_{n\to\infty} \left\| \frac{\Lambda\left(\delta + t_n\,\Delta_n\right) - \Lambda\left(\delta\right)}{t_n} - \Lambda'_\delta(\Delta) \right\| = 0,$$

where $\|\cdot\|$ is a norm on E; see van der Vaart & Wellner (1996) for more details. In the sequel, D will be a subset of the space $\ell^\infty(\mathbb{R}^d)$ of bounded functions defined on $\mathbb{R}^d$.

1.3.1 Linear functionals

The simplest of cases occurs when the parameter of interest can be written as a linear functional of F_x. To describe this context, let $\Lambda : D \to \mathbb{R}$, $D \subset \ell^\infty(\mathbb{R}^d)$, be a linear functional in the sense that for any $\delta_1, \delta_2 \in D$ and $r_1, r_2 \in \mathbb{R}$, one has

$$\Lambda(r_1\,\delta_1 + r_2\,\delta_2) = r_1\,\Lambda(\delta_1) + r_2\,\Lambda(\delta_2).$$

The parameter to be estimated is $\theta_x = \Lambda(F_x)$, for which a plug-in estimator is given by $\theta_{xh} = \Lambda(F_{xh})$. An application of the Continuous Mapping Theorem entails

$$\Theta_{xh} = \sqrt{nh_n}(\theta_{xh} - \theta_x) = \Lambda(\mathbb{F}_{xh}) \rightsquigarrow \Theta_x = \Lambda(\mathbb{F}_x),$$

where here and in the sequel, $\rightsquigarrow$ means "converges weakly to". Because the limit $\mathbb{F}_x$ is Gaussian and Λ is linear, Lemma 3.9.8 in van der Vaart & Wellner (1996) ensures that $\Theta_x \sim \mathbb{N}(\mu_x, \sigma_x^2)$, where from (1.2) and the linearity of Λ,

$$\mu_x = \Lambda(B_x) \;=\; K\left\{K_2\,\Lambda(\dot{F}_x) + K_3\,\Lambda(\ddot{F}_x)\right\} = K\left(K_2\,\dot{\theta}_x + K_3\,\ddot{\theta}_x\right),$$

with $\dot{\theta}_x = \partial\theta_x/\partial x$ and $\ddot{\theta}_x = \partial^2\theta_x/\partial x^2$; the asymptotic variance is $\sigma_x^2 = \Lambda^\star(\Gamma_x)$, where Γ_x is given in (1.2) and for $\delta \in \ell^\infty(\mathbb{R}^d \times \mathbb{R}^d)$, $\Lambda^\star(\delta) = \Lambda\{\Lambda(\delta(s,\cdot))\}$.

As long as the resampling version $\widehat{\mathbb{F}}_x$ of $\mathbb{F}_{xh}$ described in (1.7) belongs to D, the behavior of Θ_x can be replicated asymptotically by letting $\widehat{\Theta}_x = \Lambda(\widehat{\mathbb{F}}_x)$. Since Λ is linear, one obtains the simple formulas

$$\theta_{xh} = \Lambda(\mathcal{I})\,\mathbf{w}_x^\top \quad \text{and} \quad \widehat{\Theta}_x = \sqrt{nh_n}\,\Lambda(\mathcal{I})\,\mathbf{L}_x^\top, \tag{1.8}$$

where it is understood that $\Lambda(\mathcal{I})$ is taken componentwise. In practice, the resampling procedure is repeated B times in order to obtain asymptotically independent copies $\widehat{\mathbb{F}}_x^{(1)}, \ldots, \widehat{\mathbb{F}}_x^{(B)}$ of $\mathbb{F}_{xh}$. Then, based on the asymptotic normality of θ_{xh}, an approximate confidence interval of level $1-\alpha$ for θ_x is given by

$$\mathcal{CI}_{\theta_x}^{\text{Norm}} = \left[\left(\theta_{xh} - \frac{\widehat{\mu}_x}{\sqrt{nh_n}}\right) \pm z_{\frac{\alpha}{2}}\,\frac{\widehat{\sigma}_x}{\sqrt{nh_n}}\right], \tag{1.9}$$

where z_α is the upper α-percentile of the standard Normal distribution and $\widehat{\mu}_x$, $\widehat{\sigma}_x^2$ are respectively the sample mean and the sample variance of multiplier versions $\widehat{\Theta}_x^{(1)}, \ldots, \widehat{\Theta}_x^{(B)}$ of θ_{xh}. An alternative approach is to build an interval from the empirical percentiles $\widehat{R}_q$ of order $q \in (0,1)$ of these multiplier versions, namely

$$\mathcal{CI}_{\theta_x}^{\text{Perc}} = \left[\theta_{xh} - \frac{\widehat{R}_{1-\alpha/2}}{\sqrt{nh_n}},\, \theta_{xh} - \frac{\widehat{R}_{\alpha/2}}{\sqrt{nh_n}}\right]. \tag{1.10}$$

The empirical performance of $\mathcal{CI}_{\theta_x}^{\text{Norm}}$ and $\mathcal{CI}_{\theta_x}^{\text{Perc}}$ will be investigated in Section 5 via simulations.

1.3.2 General statistical functionals

The case of a nonlinear functional $\Lambda : D \to \mathbb{R}$ calls for a special treatment compared to linear statistical functionals. First, if Λ admits a Hadamard derivative Λ'_{F_x} at F_x tangentially to some $D_0 \subset D$, the functional delta method ensures that provided the limit $\mathbb{F}_x$ of $\mathbb{F}_{xh}$ belongs to D_0, then

$$\Theta_{xh} = \sqrt{nh_n}(\theta_{xh} - \theta_x) \rightsquigarrow \Theta_x = \Lambda'_{F_x}(\mathbb{F}_x).$$

Because Hadamard derivatives are linear mappings (see van der Vaart & Wellner (1996)), Θ_x is a Normal random variable with mean

$$\Lambda'_{F_x}(B_x) = K\left\{K_2\,\Lambda'_{F_x}(\dot{F}_x) + K_3\,\Lambda'_{F_x}(\ddot{F}_x)\right\}.$$

A strategy to replicate Θ_{xh} would be to simply consider $\Lambda'_{F_x}(\widehat{\mathbb{F}}_x)$; however, this approach is impracticable because F_x is usually unknown. A first, say *direct* approach, is to estimate Λ'_{F_x} with an estimator $\widehat{\Lambda'_{F_x}}$ that is uniformly consistent in the sense that for any $\delta \in D$,

$$\left\|\widehat{\Lambda'_{F_x}}(\delta) - \Lambda'_{F_x}(\delta)\right\| = o_{\mathrm{P}}(1).$$

Then, an asymptotically valid replicate of Θ_{xh} is

$$\widehat{\Theta}_x^{(1)} = \widehat{\Lambda'_{F_x}}(\widehat{\mathbb{F}}_x) = \sqrt{nh_n}\,\widehat{\Lambda'_{F_x}}(\mathfrak{I})\,\mathbf{L}_x^\top.$$

Since $\widehat{\Lambda'_{F_x}}(\mathfrak{I})$ has to be computed only once from the data, B replicates of Θ_{xh} can be obtained very quickly. An *indirect* way can be followed by considering

$$\widehat{\Theta}_x^{(2)} = \sqrt{nh_n}\left\{\Lambda\left(F_{xh} + \frac{\widehat{\mathbb{F}}_x}{\sqrt{nh_n}}\right) - \Lambda(F_{xh})\right\}.$$

Note that the space D on which Λ is defined must be chosen in order that $F_{xh}+\widehat{\mathbb{F}}_x/\sqrt{nh_n} \in D$. To establish the asymptotic validity of the method, one can invoke Theorem 3.9.4 of van der Vaart & Wellner (1996) and deduce that

$$\widehat{\Theta}_x^{(2)} = \Lambda'_{F_x}(\widehat{\mathbb{F}}_x) + o_{\mathrm{P}}(1) \rightsquigarrow \Lambda'_{F_x}(\mathbb{F}_x).$$

This approach has the advantage of avoiding the estimation of Λ'_{F_x}. However, unlike $\widehat{\Theta}_x^{(1)}$, there is no simple formula for $\widehat{\Theta}_x^{(2)}$. Nevertheless, in order to ease the computations, note that

$$F_{xh} + \frac{\widehat{\mathbb{F}}_x}{\sqrt{nh_n}} = \mathfrak{I}(\cdot)\,(\mathbf{L}_x + \mathbf{w}_x)^\top.$$

The next three subsections focus on specific examples of statistical functionals in the light of the general methodologies developed in 3.1 and 3.2. Subsection 3.3 is devoted to the estimation of condition mean and variance in nonparametric regression ; subsection 3.4 considers conditional measures of association based on correlation and on Kendall's tau ; subsection 3.5 treats the functional estimation of conditional copulas. All the formulas for the estimator θ_{xh} and for the multiplier versions $\widehat{\Theta}_x^{(1)}$ and $\widehat{\Theta}_x^{(2)}$ are postponed to Appendix A.2.

1.3.3 Nonparametric regression

Consider the general regression problem of estimating the conditional mean and variance of $Y \in \mathbb{R}$ given $X = x$, namely $m(x) = \mathrm{E}(Y|X = x)$ and $v(x) = \mathrm{Var}(Y|X = x)$. This setup encapsulates the model $Y = m(X) + \sqrt{v(X)}\,\varepsilon$, where $Y \in \mathbb{R}$ and ε is a random variable independent of Y such that $\mathrm{E}(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = 1$.

Without any *a priori* knowledge on the real functions m and $v > 0$, one is interested in the nonparametric estimation of the statistical functionals $\theta_x^{\mathrm{m}} = m(x)$ and $\theta_x^{\mathrm{v}} = v(x)$. Letting $F_x(y) = \mathrm{P}(Y \leq y | X = x)$, $y \in \mathbb{R}$, one can write $\theta_x^{\mathrm{m}} = \Lambda^{\mathrm{m}}(F_x)$ and $\theta_x^{\mathrm{v}} = \Lambda^{\mathrm{v}}(F_x)$, where for any distribution function $\delta \in \ell^\infty(\mathbb{R})$ and $\bar{\delta}(y) = \lim_{\gamma \to \infty} \delta(\gamma) - \delta(y)$,

$$\Lambda^{\mathrm{m}}(\delta) = \int_0^\infty \left\{\bar{\delta}(y) - \delta(-y)\right\} \mathrm{d}y\,, \text{ and } \Lambda^{\mathrm{v}}(\delta) = \int_{\mathbb{R}^2} \{\delta(y_1 \wedge y_2) - \delta(y_1)\,\delta(y_2)\}\, \mathrm{d}y_1\, \mathrm{d}y_2.$$

The functional Λ^{m} follows from a well-known result in mathematical statistics that states that the mean of $Y \sim F$ is $\mathrm{E}(Y) = \int_{-\infty}^0 F(y)\,\mathrm{d}y + \int_0^\infty \{1 - F(y)\}\,\mathrm{d}y$; the formula for Λ^{v} deduces from Hoeffding's equality that states that for $(Y_1, Y_2) \sim F$,

$$\mathrm{Cov}(Y_1, Y_2) = \int_{\mathbb{R}^2} \{F(y_1, y_2) - F(y_1, \infty) F(\infty, y_2)\}\, \mathrm{d}y_1\, \mathrm{d}y_2.$$

See Hoeffding (1994) and Lehmann (1966) for more details on this identity.

Nonparametric estimators of θ_x^{m} and θ_x^{v} based on (Y_1, x_1), ..., (Y_n, x_n) are therefore given by $\theta_{xh}^{\mathrm{m}} = \Lambda^{\mathrm{m}}(F_{xh})$ and $\theta_{xh}^{\mathrm{v}} = \Lambda^{\mathrm{v}}(F_{xh})$. Since Λ^{m} is linear, one only has to compute $\Lambda^{\mathrm{m}}(\mathfrak{I}) = (Y_1, \ldots, Y_n)$ since $\mathbb{I}(Y_i \leq y)$ is the distribution function of the single observation Y_i. Thus, a straightforward application of (1.8) yields

$$\theta_{xh}^{\mathrm{m}} = (Y_1, \ldots, Y_n)\, \mathbf{w}_x^\top, \qquad \widehat{\Theta}_x^{\mathrm{m}} = \sqrt{nh_n}\, (Y_1, \ldots, Y_n)\, \mathbf{L}_x^\top.$$

The estimator θ_{xh}^{m} has the form of the usual kernel estimation of a conditional mean. Moreover, $\Theta_{xh}^{\mathrm{m}} = \sqrt{nh_n}(\theta_{xh}^{\mathrm{m}} - \theta_x^{\mathrm{m}})$ is asymptotically normal with mean $\mu_x^{\mathrm{m}} = K\{K_2\, m'(x) + K_3\, m''(x)\}$. Also, as long as $\delta(s, \infty) = \delta(\infty, s) = 0$, $\Lambda^{\mathrm{m},\star}(\delta) = \int_{\mathbb{R}^2} \delta(s,t)\,\mathrm{d}t\,\mathrm{d}s$ and from Hoeffding's identity, one can show that the asymptotic variance is $(\sigma_x^{\mathrm{m}})^2 = K_4\, \theta_x^{\mathrm{v}}$.

For the estimator of the variance, one can show from direct computation that $\theta_{xh}^{\mathrm{v}} = \Lambda^{\mathrm{v}}(F_{xh}) = \Lambda^{\mathrm{v}}(\mathfrak{I})\, \mathbf{w}_x^\top$, where $\Lambda^{\mathrm{v}}(\mathfrak{I}) = ((Y_1 - \theta_{xh}^{\mathrm{m}})^2, \ldots, (Y_n - \theta_{xh}^{\mathrm{m}})^2)$. Alternatively, this formula obtains upon noting that F_{xh} is the distribution function of the population that assigns mass $w_{ni}(x, h_n)$ at point Y_i, so that θ^{v}_{xh} is the variance of that population. This estimator is similar to the one described, for instance, in Hušková & Meintanis (2009). However, the asymptotic treatment and the resampling of θ^{v}_{xh} are more involved since Λ^{v} is nonlinear. If D_0 is the space of integrable continuous functions, it is shown in Appendix A.2.1 that the Hadamard derivative of Λ^{v} at δ is

$$(\Lambda^{\text{v}})'_\delta(\Delta) = 2\int_{\mathbb{R}^2} \{\mathbb{I}(y_1 \le y_2) - \delta(y_2)\}\,\Delta(y_1)\,\mathrm{d}y_1\,\mathrm{d}y_2.$$

The consistent estimation of $(\Lambda^{\text{v}})'_\delta$ required for the use of the *direct* method based on $\widehat{\Theta}^{(1)}_x$ is quite easy here, since one only has to set $\widehat{(\Lambda^{\text{v}})'_{F_x}} = (\Lambda^{\text{v}})'_{F_{xh}}$.

1.3.4 Conditional measures of association

Consider the conditional correlation coefficient ρ_x that measures the strength of the linear relationship between the components of a bivariate random vector $(Y_1, Y_2) \in \mathbb{R}^2$ given $X = x$. Let $D = \{\delta \in \ell^\infty(\mathbb{R}^2) : 0 < \Lambda^v(\delta_1) < \infty,\ 0 < \Lambda^v(\delta_2) < \infty\}$, where for $\delta \in D$, one has $\delta_1(y_1) = \lim_{y_2\to\infty} \delta(y_1, y_2)$ and $\delta_2(y_2) = \lim_{y_1\to\infty} \delta(y_1, y_2)$. From Hoeffding's identity, $\mathrm{Cov}(Y_1, Y_2|X = x) = \Lambda^{\mathrm{Cov}}(F_x)$, where $\Lambda^{\mathrm{Cov}}(\delta) = \int_{\mathbb{R}^2} \{\delta(y_1, y_2) - \delta_1(y_1)\,\delta_2(y_2)\}\,\mathrm{d}y_1\,\mathrm{d}y_2$. One can then write $\rho_x = \Lambda^\rho(F_x)$, where for any $\delta \in D$,

$$\Lambda^\rho(\delta) = \frac{\Lambda^{\mathrm{Cov}}(\delta)}{\{\Lambda^{\text{v}}(\delta_1)\,\Lambda^{\text{v}}(\delta_2)\}^{1/2}}\,.$$

Here, Λ^{v} is the variance functional defined in subsection 1.3.3. The Hadamard derivative of Λ^ρ at $\delta \in \ell^\infty(\mathbb{R}^2)$ can be seen to be given for $\Delta \in \ell^\infty(\mathbb{R}^2)$ by

$$(\Lambda^\rho)'_\delta(\Delta) \quad = \quad \Lambda^\rho(\delta)\left\{\frac{(\Lambda^{\mathrm{Cov}})'_\delta(\Delta)}{\Lambda^{\mathrm{Cov}}(\delta)} - \frac{(\Lambda^{\text{v}})'_{\delta_1}(\Delta_1)}{2\,\Lambda^{\text{v}}(\delta_1)} - \frac{(\Lambda^{\text{v}})'_{\delta_2}(\Delta_2)}{2\,\Lambda^{\text{v}}(\delta_2)}\right\},$$

where

$$(\Lambda^{\mathrm{Cov}})'_\delta(\Delta) = \int_{\mathbb{R}^2} \{\Delta(y_1, y_2) + \delta_1(y_1)\,\Delta(\infty, y_2) + \delta_2(y_2)\,\Delta(y_1, \infty)\}\,\mathrm{d}y_1\,\mathrm{d}y_2$$

is the Hadamard derivative of Λ^{Cov} at δ. Similarly as the conditional variance, let $\widehat{(\Lambda^{\rho})'_{F_x}} = (\Lambda^{\rho})'_{F_{xh}}$ in order to estimate $(\Lambda^{\rho})'_{\delta}$ consistently, as required by the *direct* resampling method using $\widehat{\Theta}_x^{(1)}$.

When nonlinear dependence between the components of a bivariate pair is suspected, an alternative to the use of correlation is to consider Kendall's tau. In the conditional context, the latter is defined by the difference between the conditional probabilities of concordance and discordance, namely

$$\tau_x \;=\; \mathrm{P}\left\{(Y_1 - Y_1')\,(Y_2 - Y_2') > 0 | X = X' = x\right\} \mathrm{P}\left\{(Y_1 - Y_1')\,(Y_2 - Y_2') < 0 | X = X' = x\right\},$$

where (Y_1, Y_2, X) and (Y_1', Y_2', X') are identically distributed independent triplets. Another representation for Kendall's tau is $\tau_x = \Lambda^{\tau}(F_x)$, where

$$\Lambda^{\tau}(\delta) = -1 + 4 \int_{\mathbb{R}^2} \delta(y_1, y_2)\, \mathrm{d}\delta(y_1, y_2).$$

In the formula above, it is assumed that δ is of bounded variation. From Veraverbeke *et al.* (2011), the functional Λ^{τ} is Hadamard differentiable at δ tangentially to the set of continuous functions on $\overline{\mathbb{R}^2}$. Its Hadamard derivative at δ is

$$(\Lambda^{\tau})'_{\delta}(\Delta) = 4 \int_{\mathbb{R}^2} \Delta(y_1, y_2)\, \mathrm{d}\delta(y_1, y_2) + \; 4 \int_{\mathbb{R}^2} \delta(y_1, y_2)\, \mathrm{d}\Delta(y_1, y_2).$$

Again, let $\widehat{(\Lambda^{\tau})'_{F_x}} = (\Lambda^{\tau})'_{F_{xh}}$.

1.3.5 Copula

Sometimes, a statistical functional of interest is a function ; this happens in the estimation of the copula associated to a distribution function. Specifically, assuming that F_x has marginal distributions $F_{1x}, \ldots, F_{dx}$ that are continuous, Sklar's Theorem ensures that there exists a unique $C_x : [0,1]^d \to [0,1]$ such that

$$F_x(y_1, \ldots, y_d) = C_x\left\{F_{1x}(y_1), \ldots, F_{dx}(y_d)\right\}.$$

The function C_x is called the conditional copula and contains all the information about the dependence between the components of $\mathbf{Y}$ given $X = x$. Noting that

$$C_x(u_1, \ldots, u_d) = F_x\left\{F_{1x}^{-1}(u_1), \ldots, F_{dx}^{-1}(u_d)\right\},$$

a plug-in estimator of the conditional copula is given by $C_{xh} = \Lambda^{\mathrm{C}}(F_{xh})$, where $\Lambda^{\mathrm{C}} : D \to \ell^\infty([0,1]^d)$ is defined for $\mathbf{u} = (u_1, \ldots, u_d) \in [0,1]^d$ by

$$\Lambda^{\mathrm{C}}(\delta)(\mathbf{u}) = \delta\left\{\delta_1^{-1}(u_1), \ldots, \delta_d^{-1}(u_d)\right\},$$

with $\delta_j(y_j) = \lim_{y_k \to \infty, k \neq j} \delta(\mathbf{y})$. This estimator was suggested by Veraverbeke *et al.* (2011). From Bücher & Volgushev (2013), Λ^{C} is Hadamard differentiable with derivative at F_x given by

$$(\Lambda^{\mathrm{C}})'_{F_x}(\Delta)(\mathbf{u}) = \Delta\left\{\mathbf{F}_x^{-1}(\mathbf{u})\right\} - \sum_{j=1}^d C_x^{(j)}(\mathbf{u})\,\Delta_j\left\{F_{jx}^{-1}(u_j)\right\},$$

where

$$\mathbf{F}_x^{-1}(\mathbf{u}) = (F_{1x}^{-1}(u_1), \ldots, F_{dx}^{-1}(u_d)), \qquad C_x^{(j)}(\mathbf{u}) = \partial C_x(\mathbf{u})/\partial u_j$$

and $\Delta_j(y_j) = \lim_{y_k \to \infty, k \neq j} \Delta_j(\mathbf{y})$.

The estimation of $(\Lambda^{\mathrm{C}})'_{F_x}$ requires the estimation of $\mathbf{F}_x^{-1}$ and of the partial derivatives of C_x; this can be done by replacing $\mathbf{F}_x^{-1}$ by $\mathbf{F}_{xh}^{-1} = (F_{1xh}^{-1}, \ldots, F_{dxh}^{-1})$ and by using finite-difference estimators $\widehat{C_x^{(1)}}, \ldots, \widehat{C_x^{(d)}}$ based on C_{xh}. This yields

$$\widehat{(\Lambda^{\mathrm{C}})'_{F_x}}(\Delta) = \Delta\left\{\mathbf{F}_{xh}^{-1}(\mathbf{u})\right\} - \sum_{j=1}^d \widehat{C_x^{(j)}}(\mathbf{u})\,\Delta_j\left\{F_{jxh}^{-1}(u_j)\right\}.$$

1.4 Composite inference

1.4.1 Testing the equality of two marginal distributions

An important step for the investigation of the stochastic behavior of a random pair $\mathbf{Y} = (Y_1, Y_2)$ given $X = x$ is to seek for significant differences between its marginal distributions F_{1x} and F_{2x}. In other words, one wants to test for the null hypothesis $\mathcal{H}_0 : F_{1x} = F_{2x}$. To this end, consider the Kolmogorov–Smirnov statistic

$$S_{xh}^{\mathrm{KS}} = \sqrt{nh_n} \sup_{y \in \mathbb{R}} \left| F_{1xh}(y) - F_{2xh}(y) \right| .$$

Under the null hypothesis, one can write

$$S_{xh}^{\mathrm{KS}} = \sup_{y \in \mathbb{R}} \left| \mathbb{F}_{xh}(y, \infty) - \mathbb{F}_{xh}(\infty, y) \right| .$$

Hence, under Assumptions W_1–W_5, a simple application of the Continuous Mapping Theorem entails that

$$S_{xh}^{\mathrm{KS}} \rightsquigarrow \sup_{y \in \mathbb{R}} |\mathbb{F}_x(y, \infty) - \mathbb{F}_x(\infty, y)|,$$

where $\mathbb{F}_x(y, \infty) - \mathbb{F}_x(\infty, y)$ is a Gaussian process with mean $B_x(y, \infty) - B_x(\infty, y)$ and covariance function given by $K_4\{2\,F_{1x}(y \wedge y') - F_x(y, y') - F_x(y', y)\}$.

The null hypothesis $\mathcal{H}_0 : F_{1x} = F_{2x}$ is rejected at level α when $S_{xh}^{\mathrm{KS}} > Q_\alpha$, where Q_α is the positive real number Q_α such that

$$\lim_{n \to \infty} \mathrm{P}\left(S_{xh}^{\mathrm{KS}} > Q_\alpha \right) = \alpha.$$

While Q_α cannot be determined explicitly, the form of the limit distribution of S_{xh}^{KS} suggests that Q_α be consistently estimated by the empirical α-th upper percentile $\widehat{Q}_\alpha$ of $\sup_{y \in \mathbb{R}} |\widehat{\mathbb{F}}_x^{(b)}(y, \infty) - \widehat{\mathbb{F}}_x^{(b)}(\infty, y)|$, where $b \in \{1, \ldots, B\}$.

1.4.2 Testing the influence of a covariate

Consider testing $\mathcal{H}_0 : F_x = F_{x'}$ against $\mathcal{H}_1 : F_x \neq F_{x'}$ for some fixed $x \neq x'$. A natural way to do this is to measure a functional distance between F_{xh} and $F_{x'h}$. This can be done using, *e.g.*, the Kolmogorov–Smirnov or Cramér–von Mises functional. As the latter generally yields more powerful tests, consider the test statistic

$$S_{xx'h}^{\mathrm{CvM}} = nh_n \int_{\mathbb{R}^d} \left\{ F_{xh}(\mathbf{y}) - F_{x'h}(\mathbf{y}) \right\}^2 \mathrm{d}\mathbf{y}.$$

Under $\mathcal{H}_0$ and as long as Assumptions W_1–W_5 hold,

$$S^{\mathrm{CvM}}_{xx'h} = \int_{\mathbb{R}^d} \{\mathbb{F}_{xh}(\mathbf{y}) - \mathbb{F}_{x'h}(\mathbf{y})\}^2 \, \mathrm{d}\mathbf{y} \rightsquigarrow S^{\mathrm{CvM}}_{xx'} = \int_{\mathbb{R}^d} \{\mathbb{F}_{x}(\mathbf{y}) - \mathbb{F}_{x'}(\mathbf{y})\}^2 \, \mathrm{d}\mathbf{y}.$$

Clearly, $\mathbb{F}_x - \mathbb{F}_{x'}$ is a Gaussian process and from straightforward computations, its mean and covariance functions are given respectively by $B_x - B_{x'}$ and $\Gamma_x(\mathbf{y}, \mathbf{y}') + \Gamma_{x'}(\mathbf{y}, \mathbf{y}')$, where one uses the fact that the processes $\mathbb{F}_x$ and $\mathbb{F}_{x'}$ are independent. Next, define

$$\widehat{S}^{\mathrm{CvM}}_{xx'} = \int_{\mathbb{R}^d} \{\widehat{\mathbb{F}}_x(\mathbf{y}) - \widehat{\mathbb{F}}_{x'}(\mathbf{y})\}^2 \mathrm{d}\mathbf{y}.$$

An asymptotically valid α-level test will consist in rejecting $\mathcal{H}_0$ in favor of $\mathcal{H}_1$ if $S^{\mathrm{CvM}}_{xx'h}$ exceeds the empirical α-th upper percentile of B such bootstrap replicates. Formulas for $S^{\mathrm{CvM}}_{xx'h}$ and $\widehat{S}^{\mathrm{CvM}}_{xx'}$ are given in Appendix A.2.5 in the case when $d = 1$.

1.5 Investigation of the sample properties of the proposed methods

1.5.1 Parameters of the simulations

The three bootstrap methods described in Section 2 will be investigated here with the help of simulations. Specifically, computation of empirical coverage probabilities of confidence intervals for the estimation of a conditional mean, correlation coefficient and Kendall's tau will be made, and the size/power of the test for $\mathcal{H}_0 : F_x = F_{x'}$ will be studied. All the estimations are based on 10 000 replicates and the sample sizes considered are $n \in \{250, 500, 1\ 000\}$.

For the first multiplier bootstrap method, $\xi_1, \ldots, \xi_n$ are exponential with mean one. For the second multiplier method, one has for each $i \in \{1, \ldots, n\}$ that ξ_{ix} is Gamma distributed with parameters p_{ix}^2/v_{ix} and v_{ix}/p_{ix}, so that $\mathrm{E}(\xi_{ix}) = p_{ix}$ and $\mathrm{Var}(\xi_{ix}) =$

v_{ix}. As stated in Lemma 1, the assumption $\sqrt{nh_n}\max_{1\le i\le n}\xi_{ix} = o_{\mathrm{P}^\star}(1)$ required in Proposition 3 is satisfied in that case.

Many simulations not reported here suggest that the methods perform better when the multiplier random variables are standardized. Specifically, for the first multiplier, one uses $\xi_i^\star = \xi_i/\bar{\xi}_x$ instead of ξ_i ; for the second multiplier, $\xi_{ix}^\star = \xi_{ix}/\xi_{\cdot x}$ replaces ξ_{ix}. These changes are asymptotically negligible since $\bar{\xi}_x$ and $\xi_{\cdot x}$ converge to one almost surely, so that the conclusions of Proposition 1 and Proposition 3 are still valid.

The inference procedures that will be investigated require the choice of a system of weights $w_{1n}, \ldots, w_{nn}$ and of bandwidth parameters h_n and g_n. As shown by Omelka *et al.* (2013), the assumptions listed in Appendix E.1 are satisfied for the extensively used Nadaraya–Watson and local linear weight functions. The simulation results that will be reported here were obtained for the local linear weights defined for $a_n \in \{h_n, g_n\}$ by

$$w_{ni}^{\mathrm{LL}}(x, a_n) = \frac{K\left(\frac{x_i - x}{a_n}\right)\left\{S_{n,2} - \left(\frac{x_i - x}{a_n}\right)S_{n,1}\right\}}{S_{n,0}\,S_{n,2} - S_{n,1}^2},$$

where $K(u) = 35(1-u^2)^3\,\mathbb{I}(|u| \le 1)/32$ is the triweight kernel and for $\ell \in \{0,1,2\}$,

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

Because the asymptotic results presented in this paper require the weights to be non-negative and to sum to one, negative weights are taken to be zero and the remaining weights are simply re-scaled ; this operation is asymptotically negligible. Three methods have been considered for the choice of h_n,, namely

(i) the plug-in bandwidth selection rule of Gijbels *et al.* (2011) ;

(ii) $h_n = 1.06 \times \sqrt{\mathrm{Var}(X)} \times n^{-1/5}$, where $\mathrm{Var}(X)$ is the variance of $\{x_1, \ldots, x_n\}$;

(iii) $h_n = \mathrm{IQR}(X) \times n^{-1/5}$, where $\mathrm{IQR}(X)$ is the inter-quartile range of $\{x_1, \ldots, x_n\}$.

Once h_n is selected, one follows the recommendation of Härdle & Bowman (1988) and put $g_n = 1.5\,h_n\,n^{0.1}$. Based on many experiments, the results are quite similar for the three strategies above. Nevertheless, the third option based on the inter-quartile range shows a slight advantage over the other two, and hence only the results obtained with this method will be presented next.

1.5.2 Interval estimation of a conditional mean

Consider a model where $Y \in \mathbb{R}$ is influenced by a covariate X through the regression $Y = m(X) + \varepsilon$, where X is uniform on $(-1, 1)$ and $\varepsilon \sim \mathcal{N}(0, 1/16)$. For the simulations in Table 1.1, two link functions were considered, namely $m(x) = x$ and $m(x) = 0.4 \sin(3\pi x/4)$. For each $x \in \{-1/2, 1/2\}$, the empirical coverage probability has been estimated for interval estimates based on the asymptotic Normal distribution and on percentiles ; for details, see equations (1.9) and (1.10). The bootstrap method of Härdle & Bowman (1988) was also investigated ; the latter consists in resampling from the residuals of $Y = m(X) + \varepsilon$.

Generally speaking, the results of the three resampling methods discussed in this paper are good for both 90% and 95% level intervals ; they are also good for the method of Härdle & Bowman (1988) based on residuals. The coverage probabilities tend to be slightly closer to their theoretical values for larger values of n. An exception occurs with the multiplier method based on independent multipliers when $x = -1/2$ and for the two models considered. Finally note that both the intervals based on the asymptotic normality and on percentiles perform well here.

Results not reported here indicate that the lengths of the confidence intervals obtained from the conditional bootstrap and the second multiplier method are quite similar ; those of the first multiplier and of the method based on residuals are shorter, especially when n is small. This might explain the inaccurate empirical coverage probabilities of the first multiplier. These differences in lengths tend to decrease as n increases.

1.5.3 Interval estimation of a conditional correlation

Let (Y_1, Y_2, X) be trivariate standard Normal with correlation matrix $R = (R_{\ell\ell'}) \in \mathbb{R}^{3\times 3}$. In that case, it is well known that the joint conditional distribution of (Y_1, Y_2) given $X = x$ is the bivariate Normal with partial correlation coefficient

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

The conditional correlation coefficient in that context is therefore unaffected by the value at which the covariate is evaluated. For the simulations in Table 1.2, three variants of R were considered, namely (i) $R^{(1)}$, where $R^{(1)}_{12} = .5$ and $R^{(1)}_{13} = R^{(1)}_{23} = .3$, (ii) $R^{(2)}$, where $R^{(2)}_{12} = -.5$ and $R^{(2)}_{13} = -R^{(2)}_{23} = .3$, and (iii) $R^{(3)}$, where $R^{(3)}_{12} = R^{(3)}_{13} = R^{(3)}_{23} = .0$. For these three scenarios, one has respectively $\rho_x = .45, -.45, .0$.

The results in Table 1.2 concern the empirical coverage probability for interval estimates based on the asymptotic normality when $x = 1/2$. Unlike the estimation of a conditional mean, the results for confidence intervals based on percentiles were not very good and hence are not presented here. One can conjecture that the rate of convergence toward the limiting normal distribution is very quick, so that the method based on the asymptotic normality works best. Both the *direct* method based on $\widehat{\Theta}^{(1)}_x$ and the *indirect* method based on $\widehat{\Theta}^{(2)}_x$ have been investigated.

The results are generally good, especially when $n = 500$ and $n = 1\,000$. For the multiplier method based on independent multipliers, however, the estimated coverage probabilities are quite far from their expected values under model $R^{(1)}$, while the problem seems to resolve slowly as the sample size increases ; the problem is less apparent for the direct method. Observe also that the *direct* and *indirect* methods based respectively on $\widehat{\Theta}^{(1)}_x$ and $\widehat{\Theta}^{(2)}_x$ are equivalently good when the resampling is based either on the conditional bootstrap or on the second multiplier. Finally note that comments similar as those for the estimation of a conditional mean apply here about the lengths of the intervals.

1.5.4 Interval estimation of conditional Kendall's tau

Let (Y_1, Y_2, X) follow an Archimedean distribution on $[0,1]^3$, *i.e.* $F(y_1, y_2, x) = \mathrm{P}(Y_1 \leq y_1, Y_2 \leq y_2, X \leq x) = \varphi^{-1}\{\varphi(y_1) + \varphi(y_2) + \varphi(x)\}$, where $\varphi : [0,1] \to \mathbb{R}^+$ is a univariate decreasing and convex Archimedean generator. As shown by Mesfioui & Quessy (2008), the joint distribution of (Y_1, Y_2) given $X = x$ has marginal distributions $F_{1x}(y) = F_{2x}(y) = (\varphi^{-1})'\{\varphi(y) + \varphi(x)\}\varphi'(x)$ and a copula that stays in the Archimedean class, but whose generator is

$$\varphi_x(t) = \varphi\left\{(\varphi')^{-1}\left(\frac{\varphi'(x)}{t}\right)\right\} - \varphi(x).$$

For Clayton's copula whose generator is $\varphi_\theta^{\mathrm{C}\ell}(t) = (t^{-\theta} - 1)/\theta$, $\theta > -1$, computations in Mesfioui & Quessy (2008) show that the conditional generator is $\varphi_x = \varphi_{\widetilde{\theta}}^{\mathrm{C}\ell}$, where $\widetilde{\theta} = \theta/(\theta+1)$, so that this copula is family-invariant with respect to conditioning. A simple computation shows that $\tau_x(\theta) = \theta/(3\theta+2) = \tau(\theta)/\{2\tau(\theta)+1\}$, where $\tau(\theta) = \theta/(\theta+2)$ is Kendall's tau associated to $\varphi_\theta^{\mathrm{C}\ell}$. Thus, the conditional Kendall's tau is the same whatever the value of x. For Frank's copula with parameter $\theta \in \mathbb{R}$, φ_x corresponds to Ali–Mikhail–Haq's generator with parameter $1 - e^{-\theta x}$. From there one deduces

$$\tau_x(\theta) = 1 - \frac{2\left(1 - e^{-\theta x} - \theta x e^{-2\theta x}\right)}{3\left(1 - e^{-\theta x}\right)^2}.$$

More details on the above-mentioned copulas and on much more dependence models can be found in the excellent monograph by Nelsen (2006).

The results in Table 1.3 concern the Clayton and Frank copulas in the case when Kendall's tau for (Y_1, Y_2) is set to $\tau = 1/2$, and where $x \in \{1/3, 2/3\}$. For Clayton's copula, this means that $\theta = 2$, so that $\tau_x = 1/4$ for any x; for Frank's copula, $\theta \approx 5.75$, so that $\tau_x \approx 0.256$ when $x = 1/3$ and $\tau_x \approx 0.320$ when $x = 2/3$. These results can be interpreted very similarly as those in Table 1.2 for the estimation of ρ_x. The only notable difference is for the multiplier method based on independent multipliers, which performs better here.

1.5.5 Test of $\mathcal{H}_0 : F_x = F_{x'}$

Consider testing the null hypothesis $\mathcal{H}_0 : F_x = F_{x'}$ against $\mathcal{H}_1 : F_x \neq F_{x'}$ in the case $d = 1$. For the simulations that are presented, X is uniformly distributed on $(0, 1)$ and $Y|X = x$ follows a distribution F with some parameter depending on X through the function $\zeta(x) = 1 + (1/3)(4x - 1)(4x - 3)$, so that $\zeta(0) = \zeta(1) = 2$ and $\zeta(1/4) = \zeta(3/4) = 1$. In the first scenario considered, the distribution of $Y|X = x$ is exponential with mean $\zeta(x)$, while in the second scenario, the law of $Y|X = x$ is Normal with mean $\zeta(x)$ and variance 1/2. In the simulation results reported in Table 1.4, $x = 1/4$ and $x' \in \{1/2, 5/8, 3/4, 7/8\}$; the case $x' = 3/4$ is a situation under the null hypothesis.

First observe that the test based on $S^{\text{CvM}}_{xx'h}$ is good at keeping its nominal 5% level under the null hypothesis, *i.e* when $x = 3/4$, for the three considered resampling methods. As one could expect from the asymptotic theory, the power are very similar among the three methods for a given alternative hypothesis. Nevertheless, one can observe a mild advantage of the multiplier method based on independent multipliers, although it seems that its type I error is slightly over 5%. The power increases as the sample size increases as an effect of the consistency of test based on the Cramér–von Mises statistic. Finally note that discrepancies between F_x and $F_{x'}$ are more easily detected under the Normal distribution than under the exponential distribution.

1.6 Illustration on real data

According to the *World Factbook*, a commonly used indicator of the level of health in a country is its infant mortality rate, which consists in the number of deaths of infants under one-year-old per thousands live births in a given year. The influence of this index on life expectancies will be investigated in the light of the methods previously described.

Specifically, the data set that will be analyzed consists of triplets $(Y_{1i}, Y_{2i}, x_i)_{i=1}^n$ for $n = 224$ countries, where for a given country i, Y_{1i} and Y_{2i} are the life expectancies of males and females, respectively, and the value x_i of the covariate is the log-infant mortality rate for that country. A similar investigation was made by Veraverbeke *et al.* (2011), where the relationship between the life expectancies was investigated with respect to the values of the under-five mortality rate.

A first step is to investigate how the mean of life expectancies of males and females are influenced by the log-infant mortality rate. The curves of the point-wise nonparametric estimation of $m_1(x) = \mathrm{E}(Y_1|X = x)$ and $m_2(x) = \mathrm{E}(Y_2|X = x)$ for $x \in \{1.00, 1.05, \ldots, 4.50\}$ are presented in Figure 1.1. The confidence bands are based on point-wise confidence intervals using the asymptotic Normal distribution and the second multiplier strategy. One can see that the shape of the regression curves for the life expectancy of males and females are quite similar for small and large values of x (say $x \leq 1.5$ and $x \geq 3$). Marked differences between these curves occur around $x = 2$.

Table 1.5 reports the estimation of the conditional means and standard deviations for three values of x corresponding to the first quartile, the median and the third quartile of $x_1, \ldots, x_{224}$; these values are respectively $Q_1 = 1.8$, $M_e = 2.6$ and $Q_3 = 3.7$. Confidence intervals based on the asymptotic Normal distribution are also provided. In view of (1.9) and from an application of the standard delta method, a confidence interval for the conditional standard deviation $\sqrt{v(x)}$ is

$$\left[\left(\sqrt{\theta^{\mathrm{v}}_{xh}} - \frac{\widehat{\mu}_x}{2\sqrt{nh_n\theta^{\mathrm{v}}_{xh}}}\right) \pm z_{\alpha/2}\,\frac{\widehat{\sigma}_x}{2\sqrt{nh_n\theta^{\mathrm{v}}_{xh}}}\right],$$

where θ^{v}_{xh} is the estimation of $v(x)$ and $\widehat{\mu}_x$, $\widehat{\sigma}_x$ are the bootstrap mean and variance, respectively. As suggested by Figure 1.1, the conditional means of the life expectancy of

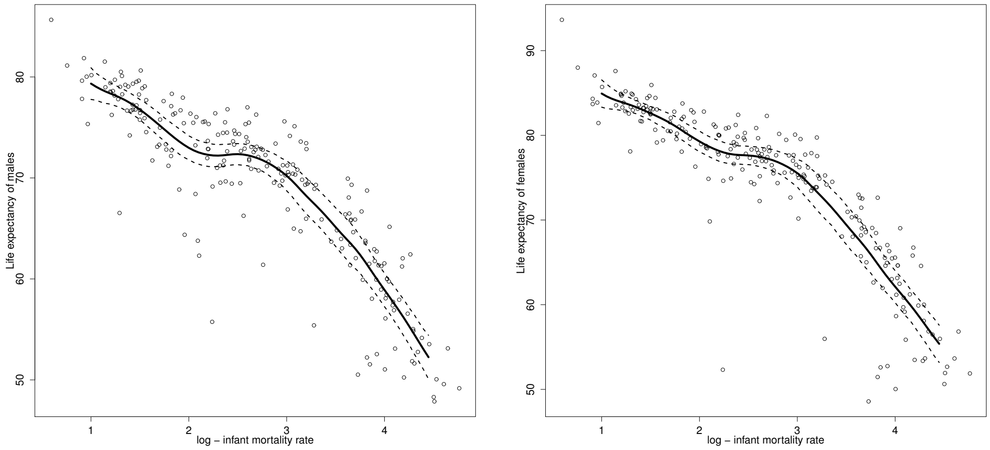


Figure 1.1 – Estimation of the mean of the life expectancies of males (left panel) and females (right panel) conditional to the log-infant mortality rate. The dotted lines represent point-wise 95% confidence intervals based on the asymptotic Normal distribution using the second multiplier bootstrap

males are significantly lower than that of females, especially when $x = Q_1, M_e$. On the other hand, the conditional variances are not significantly different for the three values of the covariate.

Figure 1.1 and Table 1.1 strongly suggest that the distribution of $Y_1|X = x$ significantly differs from that of $Y_2|X = x$, at least for the three values of x that have been considered, namely $x \in \{Q_1, M_e, Q_3\}$. It is indeed clearly the case if one looks at Figure 1.2, where the curve $|F_{1xh}(y) - F_{2xh}(y)|$ as a function of y is reported together with the critical value $\widehat{Q}_{.05}$ as estimated following the procedure in Section 1.4.1 using the first multiplier method. Hence, the tests based on the test statistic S_{xh}^{KS} conclude to a rejection of the null hypothesis $F_{1x} = F_{2x}$.

The fact that the log-infant mortality rate has a clear influence on the marginal behavior

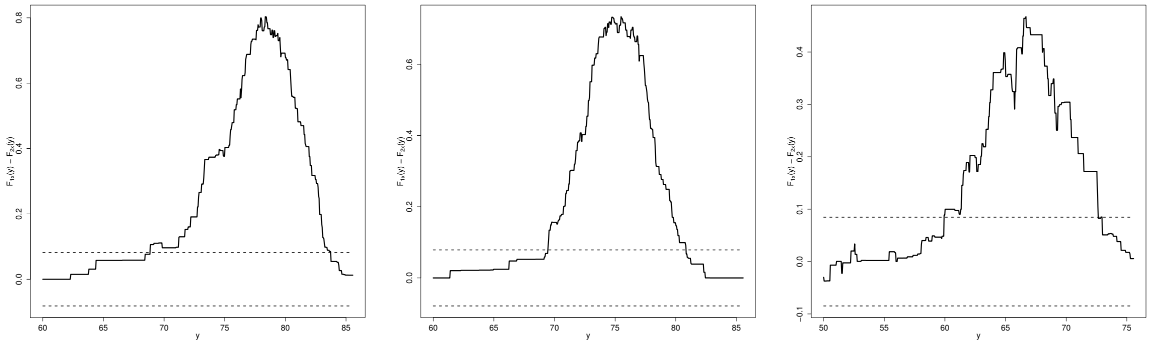


Figure 1.2 – Curves of the difference (in absolute values) between the distribution functions of the life expectancies of males and females conditional on the log-infant mortality rate when $x = Q_1$ (left panel), $x = M_e$ (middle panel) and $x = Q_3$ (right panel), together with the estimated critical value $\widehat{Q}_{.05}$ computed under the null hypothesis $F_{1x} = F_{2x}$ (dotted lines)

of life expectancies of males and females suggests that it may also play a role on the link between Y_1 and Y_2. It seems to be indeed the case looking at Figure 1.3, where the curves of the estimation of ρ_x and τ_x are reported for $x \in \{1.00, 1.05, \ldots, 4.50\}$. Confidence intervals based on the asymptotic Normal distribution using the second multiplier bootstrap are also provided. For both the conditional correlation ρ_x and Kendall's tau τ_x, the lowest association between Y_1 and Y_2 occurs when x takes values somewhere between 2.5 and 3.

The procedure described in Section 1.4.2 allows to detect significant differences in the conditional distribution of F_x and $F_{x'}$ for fixed values of x and x'. For selected values of x, the test based on $S^{\mathrm{CvM}}_{xx'h}$ has been performed until values $x_L < x$ and $x_U > x$ are reached so that the null hypothesis is rejected for x outside $[x_L, x_U]$. For the life expectancies of males, one obtains $[x_L, x_U] = [1.57, 2.13]$ when $x = Q_1$, $[x_L, x_U] = [2.04, 2.88]$ when $x = M_e$, and $[x_L, x_U] = [3.48, 3.86]$ when $x = Q_3$. For females, $[x_L, x_U] = [1.57, 2.18]$ when

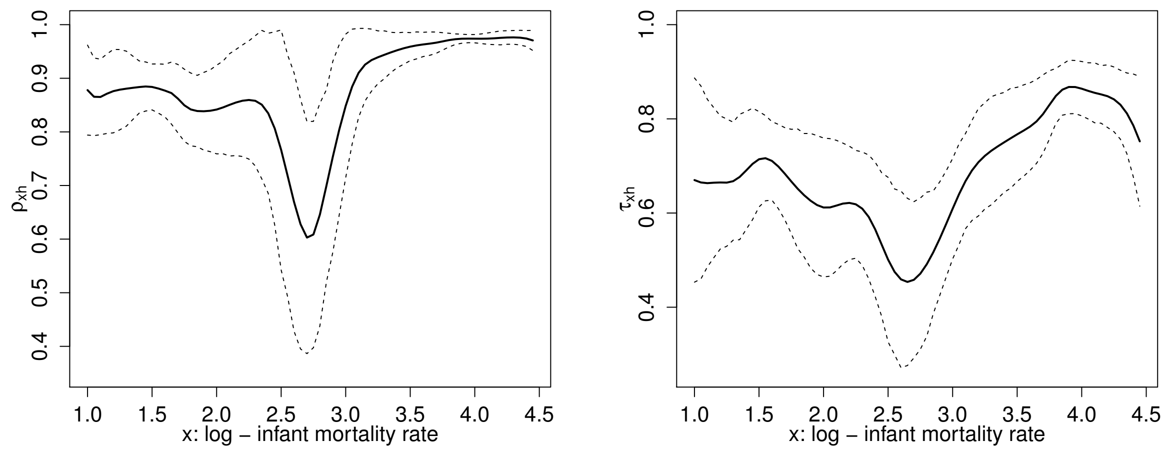


Figure 1.3 – Estimation of ρ_x (left panel) and τ_x (right panel) for the life expectancies of males and females conditional to the log-infant mortality rate. The dotted lines are point-wise 95% confidence intervals based on the asymptotic Normal distribution using the second multiplier bootstrap

$x = Q_1$, $[x_L, x_U] = [2.27, 2.88]$ when $x = M_e$, and $[x_L, x_U] = [3.48, 3.86]$ when $x = Q_3$. The critical values of the test were computed using the first multiplier bootstrap ; results not presented here using the second multiplier were very similar.

1.7 Conclusion

In this paper, two resampling methods adapted to conditional distributions have been introduced. They can be seen as alternative to the conditional bootstrap based on weighted resampling of the original observations. Their asymptotic validity has been formally established. These methods are easy-to-compute and are particularly well-adapted when combined with the theory of Hadamard differentiable statistical functionals.

The sample properties of the newly introduced methods have also been investigated for the construction of confidence intervals. Generally speaking, in the case of asymptotically Normal statistical functional, the confidence intervals based on this limit distribution are more accurate than those based on bootstrap percentiles. In the case of nonlinear functionals, the direct approach based on $\widehat{\Theta}_x^{(1)}$ is clearly quicker than that based on $\widehat{\Theta}_x^{(2)}$. As the results for both methods are equivalently good when combined either with the conditional bootstrap or the second multiplier method, a general recommendation would be to use the direct method with one of these two resampling schemes.

In the light of the simulations results that have been presented, the first multiplier method based on i.i.d. multipliers often yields less accurate interval estimates, especially for small sample sizes. In order to improve its performance in terms of coverage probabilities, an idea would be to account for the additional variability introduced by the estimated bias, as is done by Calonico *et al.* (2014). Other possibilities of future works would be to extend the validity of the introduced multiplier bootstraps to serially dependent data. Also, adaptation of the methodology for the testing of composite goodness-of-fit hypotheses could lead to very quick procedures compared to the use of the parametric bootstrap. A more challenging task would be to conceive inference procedures that are valid uniformly for all $x \in \mathcal{X} \subset \mathbb{R}$, for example when testing the global influence of the covariate.

Tableau 1.1 – Empirical coverage probability, based on 10 000 replicates, for nominal 90% and 95% confidence intervals for $m(x) = \mathrm{E}(Y|X = x)$ based on four resampling methods ; the intervals are based on the asymptotic normality (Norm) or on bootstrap percentiles (Perc)

$m(x)$	Interval type	Resampling method	$n = 250$ 90%	$n = 250$ 95%	$n = 500$ 90%	$n = 500$ 95%	$n = 1\ 000$ 90%	$n = 1\ 000$ 95%
x, $x = -1/2$	Norm	Bootstrap	89.0	94.0	90.5	94.6	90.3	95.7
		Multiplier 1	86.1	92.3	87.0	92.0	87.2	93.0
		Multiplier 2	87.3	94.3	89.3	95.2	91.2	95.1
		Residuals	88.7	94.1	90.2	94.3	89.8	95.6
	Perc	Bootstrap	88.4	93.7	89.9	94.5	89.2	94.5
		Multiplier 1	87.8	92.3	87.0	92.0	86.1	92.2
		Multiplier 2	87.3	93.6	88.4	94.9	89.9	94.9
		Residuals	87.7	93.4	89.4	94.1	89.8	94.8
x, $x = 1/2$	Norm	Bootstrap	90.5	95.6	90.6	95.7	89.2	94.0
		Multiplier 1	87.4	93.3	88.5	93.6	88.4	94.3
		Multiplier 2	90.7	95.4	88.9	94.6	90.8	94.8
		Residuals	91.2	95.5	90.7	95.5	89.8	94.3
	Perc	Bootstrap	89.6	95.4	90.2	95.3	88.2	92.7
		Multiplier 1	87.6	92.5	88.4	92.9	87.2	92.2
		Multiplier 2	90.7	94.8	86.7	93.8	89.1	94.7
		Residuals	89.8	94.8	89.2	94.8	89.0	94.1
$0.4\ \sin(3\pi x/4)$, $x = -1/2$	Norm	Bootstrap	89.2	94.4	89.7	94.9	90.4	95.6
		Multiplier 1	86.4	92.8	86.2	92.0	87.5	93.1
		Multiplier 2	87.5	94.9	90.1	95.8	91.2	95.0
		Residuals	86.0	92.4	88.1	93.8	87.8	93.8
	Perc	Bootstrap	90.4	95.2	91.6	94.9	91.6	96.2
		Multiplier 1	88.1	94.0	87.7	93.3	88.5	93.9
		Multiplier 2	90.1	95.3	91.9	96.6	92.5	95.5
		Residuals	88.2	93.3	89.2	94.4	90.4	94.9
$0.4\ \sin(3\pi x/4)$, $x = 1/2$	Norm	Bootstrap	91.3	96.0	90.3	95.9	88.9	94.1
		Multiplier 1	87.7	93.7	88.3	93.6	88.9	94.3
		Multiplier 2	90.8	95.4	90.1	95.4	90.2	95.2
		Residuals	88.1	94.1	88.9	94.5	87.4	92.5
	Perc	Bootstrap	92.6	96.8	92.0	96.8	90.1	94.9
		Multiplier 1	89.1	94.8	90.2	95.1	89.7	94.7
		Multiplier 2	92.6	95.6	90.8	95.0	91.9	95.6
		Residuals	90.2	95.1	89.7	94.8	89.5	94.2

Tableau 1.2 – Empirical coverage probability, based on 10 000 replicates, for nominal 90% and 95% confidence intervals for ρ_x based on the asymptotic normality ; three resampling methods used, combined either with the *direct* ($\widehat{\Theta}_x^{(1)}$) or *indirect* ($\widehat{\Theta}_x^{(2)}$) approach

R	Approach	Resampling method	$n = 250$ 90%	$n = 250$ 95%	$n = 500$ 90%	$n = 500$ 95%	$n = 1\ 000$ 90%	$n = 1\ 000$ 95%
$R^{(1)}$	$\widehat{\Theta}_x^{(1)}$	Bootstrap	87.2	93.2	90.3	94.6	90.8	95.4
		Multiplier 1	85.5	90.2	86.8	93.1	86.8	93.2
		Multiplier 2	88.4	93.0	88.4	93.5	90.2	94.9
	$\widehat{\Theta}_x^{(2)}$	Bootstrap	87.0	92.9	90.3	94.6	91.1	95.4
		Multiplier 1	85.9	91.1	87.8	93.6	87.3	93.6
		Multiplier 2	89.1	93.2	89.1	93.8	90.6	95.0
$R^{(2)}$	$\widehat{\Theta}_x^{(1)}$	Bootstrap	88.8	93.3	89.0	94.3	90.1	95.0
		Multiplier 1	84.9	90.1	86.3	91.4	86.0	91.7
		Multiplier 2	85.7	91.8	85.6	91.8	90.4	95.6
	$\widehat{\Theta}_x^{(2)}$	Bootstrap	88.6	92.9	89.1	94.1	90.4	95.0
		Multiplier 1	86.0	90.7	87.0	91.8	86.8	92.2
		Multiplier 2	86.8	92.4	86.4	91.9	90.9	95.9
$R^{(3)}$	$\widehat{\Theta}_x^{(1)}$	Bootstrap	88.9	94.0	89.4	93.9	92.5	96.5
		Multiplier 1	86.8	92.5	88.7	94.0	89.5	95.1
		Multiplier 2	89.1	93.1	89.6	95.1	90.0	94.8
	$\widehat{\Theta}_x^{(2)}$	Bootstrap	88.9	93.8	89.3	93.7	92.5	96.5
		Multiplier 1	87.8	92.6	89.5	94.3	89.6	95.4
		Multiplier 2	89.9	94.0	90.1	95.4	90.8	95.0

Tableau 1.3 – Empirical coverage probability, based on 10 000 replicates, for nominal 90% and 95% confidence intervals for τ_x based on the asymptotic normality ; three resampling methods used, combined either with the *direct* ($\widehat{\Theta}_x^{(1)}$) or *indirect* ($\widehat{\Theta}_x^{(2)}$) approach

Copula	Approach	Resampling method	$n = 250$ 90%	$n = 250$ 95%	$n = 500$ 90%	$n = 500$ 95%	$n = 1\ 000$ 90%	$n = 1\ 000$ 95%
Clayton, $x = 1/3$	$\widehat{\Theta}_x^{(1)}$	Bootstrap	90.9	95.4	91.9	95.3	91.3	94.7
		Multiplier 1	87.1	92.8	89.0	92.9	88.8	93.5
		Multiplier 2	90.3	95.2	91.4	95.0	91.2	95.6
	$\widehat{\Theta}_x^{(2)}$	Bootstrap	91.3	95.6	92.1	95.4	91.5	95.2
		Multiplier 1	88.3	93.4	89.6	93.5	89.5	94.0
		Multiplier 2	91.1	95.8	92.0	95.2	91.5	95.9
Clayton, $x = 2/3$	$\widehat{\Theta}_x^{(1)}$	Bootstrap	90.4	94.3	90.9	94.7	90.5	94.2
		Multiplier 1	87.5	93.5	88.5	93.9	89.8	94.9
		Multiplier 2	87.6	93.2	90.2	95.3	91.1	95.5
	$\widehat{\Theta}_x^{(2)}$	Bootstrap	90.5	94.4	90.9	94.8	90.5	94.2
		Multiplier 1	88.2	93.9	88.6	94.2	90.5	95.1
		Multiplier 2	88.3	93.6	90.6	95.5	91.3	95.5
Frank, $x = 1/3$	$\widehat{\Theta}_x^{(1)}$	Bootstrap	91.1	95.7	90.4	95.4	90.8	96.2
		Multiplier 1	85.0	92.2	88.6	93.1	90.4	95.0
		Multiplier 2	90.9	96.5	90.2	94.8	89.6	95.3
	$\widehat{\Theta}_x^{(2)}$	Bootstrap	91.1	95.5	90.6	95.9	90.8	96.3
		Multiplier 1	85.8	92.8	89.2	93.7	90.8	95.3
		Multiplier 2	92.1	96.7	90.7	95.0	90.4	95.3
Frank, $x = 2/3$	$\widehat{\Theta}_x^{(1)}$	Bootstrap	90.4	95.6	90.7	95.6	89.7	94.4
		Multiplier 1	85.3	90.8	87.2	92.5	87.3	93.2
		Multiplier 2	91.0	96.1	89.4	94.9	89.6	94.9
	$\widehat{\Theta}_x^{(2)}$	Bootstrap	90.5	95.8	90.8	95.9	89.9	94.5
		Multiplier 1	86.3	91.0	87.9	92.7	88.0	93.6
		Multiplier 2	91.8	96.3	89.6	95.1	89.7	95.1

Tableau 1.4 – Power of the test based on $S^{\text{CvM}}_{xx'h}$ with $x = 1/4$ under the exponential and Normal distribution with mean $\zeta(x) = 1 + (1/3)(4x - 1)(4x - 3)$; $\alpha = 0.05$

x'	Resampling method	$(Y\|X = x) \sim \exp\{\zeta(x)\}$ $n = 250$	$n = 500$	$n = 1\ 000$	$(Y\|X = x) \sim N(\zeta(x), 1/2)$ $n = 250$	$n = 500$	$n = 1\ 000$
$\frac{1}{2}$	Bootstrap	28.1	48.8	73.1	43.1	65.2	86.5
	Multiplier 1	34.2	50.8	73.5	47.4	66.2	88.4
	Multiplier 2	27.9	48.5	73.3	42.5	65.4	87.4
$\frac{5}{8}$	Bootstrap	16.5	27.4	44.1	24.6	42.4	63.6
	Multiplier 1	19.5	29.5	44.7	30.7	42.9	65.6
	Multiplier 2	18.3	25.9	45.2	25.5	41.8	63.1
$\frac{3}{4}$	Bootstrap	4.9	4.4	4.8	4.7	4.3	4.6
	Multiplier 1	6.7	6.8	6.2	6.3	5.9	5.4
	Multiplier 2	3.2	3.8	4.8	5.8	4.9	5.0
$\frac{7}{8}$	Bootstrap	23.9	38.1	59.2	55.6	81.6	96.4
	Multiplier 1	24.7	40.9	59.0	57.6	83.0	96.9
	Multiplier 2	24.0	40.1	57.9	57.4	83.1	97.1

Tableau 1.5 – Nonparametric point estimation and 95% confidence intervals based on the asymptotic normality and the two multiplier methods for $m(x)$ and $\sqrt{v(x)}$ in the regression model $Y = m(X) + \sqrt{v(X)}\,\varepsilon$ for the life expectancies of males and females conditional to the log-infant mortality rate

		Model $Y_1 = m(X) + \sqrt{v(X)}\,\varepsilon$			Model $Y_2 = m(X) + \sqrt{v(X)}\,\varepsilon$		
x	θ_x	θ_{xh}	Multiplier 1	Multiplier 2	θ_{xh}	Multiplier 1	Multiplier 2
Q_1	$m(x)$	74.2	[73.0, 75.5]	[73.2, 75.3]	80.6	[79.5, 81.7]	[79.9, 81.3]
	$\sqrt{v(x)}$	2.1	[0.6,3.5]	[1.0,3.1]	1.8	[0.0,5.0]	[0.2,3.1]
M_e	$m(x)$	72.2	[71.1, 73.3]	[71.4, 73.0]	77.4	[76.3, 78.5]	[76.8, 78.0]
	$\sqrt{v(x)}$	1.7	[0.0,3.5]	[0.6,2.4]	1.8	[0.0,5.1]	[0.4,2.4]
Q_3	$m(x)$	63.6	[61.6, 65.5]	[62.2, 64.9]	67.4	[65.1, 69.7]	[65.6, 69.2]
	$\sqrt{v(x)}$	2.2	[0.9,4.1]	[1.3,3.7]	2.5	[1.3,5.1]	[1.3,5.0]

CHAPITRE 2

On the asymptotic behavior of two estimators of the conditional copula based on time series

Résumé

Telle que définie par Patton (2006), la copule conditionnelle extraite d'un vecteur aléatoire (Y_1, Y_2) conditionnellement à $X = x \in \mathbb{R}$ est la fonction $C_x : [0,1]^2 \to [0,1]$ satisfaisant $\mathrm{P}(Y_1 \leq y_1, Y_2 \leq y_2) = C_x\{\mathrm{P}(Y_1 \leq y_1 | X = x), \mathrm{P}(Y_2 \leq y_2 | X = x)\}$. Dans cette note, la convergence faible de deux estimateurs de C_x proposés par Gijbels *et al.* (2011) est obtenue en présence de mélange fort. Nous montrons que, suivant certaines conditions sur le système de pondérateurs utilisés ainsi que sur les coefficients de mélange, la représentation du processus limite est la même que celle obtenue dans l'article Veraverbeke *et al.* (2011) alors que les données étaient supposées i.i.d. La performance de ces estimateurs est examinée au moyen d'une étude de simulations.

Abstract

As defined by Patton (2006), 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 $\mathrm{P}(Y_1 \le y_1, Y_2 \le y_2) = C_x\{\mathrm{P}(Y_1 \le y_1 | X = x), \mathrm{P}(Y_2 \le y_2 | X = x)\}$. In this note, the weak convergence of two estimators of C_x proposed by Gijbels *et al.* (2011) 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 Veraverbeke *et al.* (2011) under the i.i.d. setting. The performance of these estimators in finite sample sizes is investigated.

2.1 Introduction

Copulas have become a popular tool for modeling the dependence between the components of a random vector. The starting point of copula theory is Sklar's Theorem. In its classical formulation, this result ensures that for any random 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 $\mathbf{y} = (y_1, \ldots, y_d) \in \mathbb{R}^d$,

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

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 Gijbels *et al.* (2011), where the relationship between the life expectancy of men (Y_1) and women (Y_2) with respect to the gross domestic product (X) is studied. Such an investigation relies on an extension of Sklar's Theorem to the case of conditional dependence as initiated by Patton (2006). Formally, letting $H_x(y_1, y_2) = \mathrm{P}(Y_1 \le y_1, Y_2 \le y_2 | X = x)$, 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 $(y_1, y_2) \in \mathbb{R}^2$,

$$H_x(y_1, y_2) = C_x \left\{ \mathrm{P}(Y_1 \leq y_1 | X = x), \mathrm{P}(Y_2 \leq y_2 | X = x) \right\}. \tag{2.1}$$

Two nonparametric estimators of C_x were proposed by Gijbels *et al.* (2011) and their asymptotic behavior was formally investigated by Veraverbeke *et al.* (2011) in the i.i.d. case. The purpose of this note is to extend these large-sample results to the case of time series, 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 Bücher & Volgushev (2013) 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

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

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 Meitz & Saikkonen (2002), Doukhan (1994) and Carrasco & Chen (2002).

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 are listed in an appendix and the detailed proofs of the main results are to be found in the Supplementary materials.

2.2 Investigation of a first estimator of C_x

2.2.1 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

$$C_x(u_1, u_2) = H_x \left\{ F_{1x}^{-1}(u_1), F_{2x}^{-1}(u_2) \right\}, \tag{2.2}$$

where $F_{1x}(y) = \mathrm{P}(Y_1 \leq y | X = x)$ and $F_{2x}(y) = \mathrm{P}(Y_2 \leq y | X = x)$. An estimator of H_x will then provide a plug-in estimation of C_x. Specifically, let

$$H_{xh}(y_1, y_2) = \sum_{i=1}^{n} w_{ni}(x, h) \, \mathbb{I} \left(Y_{1i} \leq y_1, Y_{2i} \leq y_2 \right), \tag{2.3}$$

where $w_{n1}, \ldots, w_{nn}$ are nonnegative weight functions that smooth the covariate space and $h = h_n$ is a bandwidth parameter that typically depends on the sample size. Hereafter, it is assumed that the $w_{ni}(x, h)$'s sum to 1. From Equation (4.1), 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.

2.2.2 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 behavior of $\mathbb{H}_{xh} = \sqrt{nh}(H_{xh} - H_x)$, where H_{xh} and H_x are defined respectively in Equation (2.3) and Equation (4.2).

Proposition 1 *Suppose that Assumptions $\mathcal{A}_1$–$\mathcal{A}_2$, W_1–W_5 and W_{11}–W_{13} are satisfied. If $nh \to \infty$ and $nh^5 \to K^2 < \infty$, then the empirical process $\mathbb{H}_{xh}$ converges weakly in the space $\ell^\infty(\mathbb{R}^2)$ to a Gaussian limit $\mathbb{H}_x$ such that*

$$\mathrm{E}\left\{\mathbb{H}_x(y_1, y_2)\right\} = K\left\{K_2\, \dot{H}_x(y_1, y_2) + K_3\, \ddot{H}_x(y_1, y_2)\right\}$$

and for $a \wedge b = \min(a, b)$,

$$\mathrm{Cov}\left\{\mathbb{H}_x(y_1, y_2), \mathbb{H}_x(y_1', y_2')\right\} = K_4\left\{H_x(y_1 \wedge y_1', y_2 \wedge y_2') - H_x(y_1, y_2)\, H_x(y_1', y_2')\right\},$$

where the constants K_2–K_4 are defined in Assumptions W_2–W_4.

It is worth noting that the asymptotic covariance structure of $\mathbb{H}_{xh}$ under the strong mixing assumption is the same as that obtained by Veraverbeke *et al.* (2011) in the i.i.d. setting. In other words, the influence of time-dependency is asymptotically negligible here. An explanation is that the weight functions smooth the covariate space in a shrinking neighborhood of x as n goes to infinity. Note however that compared to the i.i.d. context, the additional assumptions W_{11}–W_{13} on the weight functions are needed in order to tackle moments of order six entailed by time-dependency.

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 empirical process $\mathbb{C}_{xh}$ 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

$$\mathrm{E}\left\{\mathbb{B}_x(u_1,u_2)\right\} = K\left[K_2\,\dot{C}_x(u_1,u_2) + K_3\,\ddot{C}_x\left\{F_{1x}^{-1}(u_1), F_{2x}^{-1}(u_2)\right\}\right]$$

and

$$\mathrm{Cov}\left\{\mathbb{B}_x(u_1,u_2), \mathbb{B}_x(u_1',u_2')\right\} = K_4\left\{C_x(u_1\wedge u_1', u_2\wedge u_2') - C_x(u_1,u_2)\,C_x(u_1',u_2')\right\}.$$

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 are identical to those derived by Veraverbeke *et al.* (2011) in the i.i.d case.

Now sketches of the proofs of Proposition 1 and of Proposition 2 are provided in the next two subsections. The full arguments can be found in the Supplementary material section.

2.2.3 Sketch of the proof of Proposition 1

In the sequel, expectations of the form $\mathrm{E}\{f(Y_{1i},Y_{2i},X_i)\}$ are taken conditionally on X_i, *i.e.*

$$\mathrm{E}\left\{f(Y_{1i},Y_{2i},X_i)\right\} = \int_{\mathbb{R}^2} f\left(y_1,y_2,X_i\right)\mathrm{d}H_{X_i}(y_1,y_2).$$

Since Assumptions W_2, W_3 and W_5 hold, one deduces from Veraverbeke *et al.* (2011) that $\sqrt{nh}\{\mathrm{E}(H_{xh}) - H_x\} = \mathrm{E}(\mathbb{H}_x) + o(1)$. Therefore, one only needs to show that the process

$$Z_{xn}(y_1,y_2) = \sqrt{nh}\left\{H_{xh}(y_1,y_2) - \mathrm{E}(H_{xh}(y_1,y_2))\right\}$$

is asymptotically gaussian and that its limiting covariance function matches that of $\mathbb{H}_x$. According for instance to Theorem 1.5.4 of van der Vaart & Wellner (1996), one needs to show that the finite-dimensional distributions of Z_{xn} are asymptotically jointly Normal and that Z_{xn} is asymptotically tight.

That the finite-dimensional distributions of Z_{xn} are asymptotically jointly Normal is established in the Supplementary material section, where the arguments in the proof of Theorem 27.4 of Billingsley (1968), which apply to serially dependent data, are adapted to the conditional setup.

In order to show that the covariance function of Z_{xn} converges to that of $\mathbb{H}_x$, one follows an idea similar to the one developed by Li & Racine (2007). Specifically, for $\mathbf{y} = (y_1, y_2)$ and $\mathbf{y}' = (y_1', y_2')$, one has

$$\mathrm{Cov}\left\{Z_{xn}(\mathbf{y}), Z_{xn}(\mathbf{y}')\right\} = \Lambda_{n1}(\mathbf{y},\mathbf{y}') + \Lambda_{n2}(\mathbf{y},\mathbf{y}') + \Lambda_{n3}(\mathbf{y},\mathbf{y}'),$$

where for $\vartheta_i(\mathbf{y}) = \mathbb{I}(Y_{1i} \leq y_1, Y_{2i} \leq y_2) - H_{X_i}(y_1, y_2)$, $i \in \{1,\ldots,n\}$ and $L_{i,\ell}(\mathbf{y},\mathbf{y}') = \mathrm{Cov}\{\vartheta_i(\mathbf{y}), \vartheta_{i+\ell}(\mathbf{y}')\}$,

$$\begin{aligned}
\Lambda_{n1}(\mathbf{y},\mathbf{y}') &= nh \sum_{i=1}^{n} \{w_{ni}(x,h)\}^2 \, L_{i,0}(\mathbf{y},\mathbf{y}'), \\
\Lambda_{n2}(\mathbf{y},\mathbf{y}') &= \sum_{i=1}^{n} \sum_{\ell=1}^{\lfloor h^{-1/2} \rfloor} w_{ni}(x,h)\, w_{n,i+\ell}(x,h) \left\{L_{i,i+\ell}(\mathbf{y},\mathbf{y}') + L_{i,i+\ell}(\mathbf{y}',\mathbf{y})\right\}, \\
\Lambda_{n3}(\mathbf{y},\mathbf{y}') &= \sum_{i=1}^{n} \sum_{\ell=\lfloor h^{-1/2} \rfloor + 1}^{n} w_{ni}(x,h)\, w_{n,i+\ell}(x,h) \left\{L_{i,i+\ell}(\mathbf{y},\mathbf{y}') + L_{i,i+\ell}(\mathbf{y}',\mathbf{y})\right\}.
\end{aligned}$$

Since $L_{i,0}(\mathbf{y},\mathbf{y}') = H_{X_i}(y_1 \wedge y_1', y_2 \wedge y_2') - H_{X_i}(y_1, y_2)\, H_{X_i}(y_1, y_2')$, Assumption $\mathcal{A}_2$ ensures that $L_{i,0}(\mathbf{y},\mathbf{y}') = H_x(y_1 \wedge y_1', y_2 \wedge y_2') - H_x(y_1, y_2)\, H_x(y_1, y_2') + o_{\mathrm{P}}(1)$ for any $i \in I_{nx}$. From Assumption W_4,

$$\lim_{n\to\infty} \Lambda_{n1}(\mathbf{y},\mathbf{y}') = K_4 \left\{H_x(y_1 \wedge y_1', y_2 \wedge y_2') - H_x(y_1, y_2)\, H_x(y_1, y_2')\right\}.$$

Next,using the fact that $|L_{i,\ell}| \leq 1$ and from Assumption W_{11} in the special case when $k = 1$ and $v_n = \lfloor h^{-1/2} \rfloor$, one obtains

$$\Lambda_{n2}(\mathbf{y}, \mathbf{y}') \leq 2nh \lfloor h^{-1/2} \rfloor \left\{ \max_{1 \leq \ell \leq \lfloor h^{-1/2} \rfloor} \sum_{i=1}^{n} w_{ni}(x,h)\, w_{n,i+\ell}(x,h) \right\} = O_{\mathrm{P}}(h^{-1/2}).$$

Finally, the strong mixing assumption entails that for all $\mathbf{y}, \mathbf{y}' \in \mathbb{R}^2$,

$$|\mathrm{Cov}\{\vartheta_i(\mathbf{y}), \vartheta_{i+\ell}(\mathbf{y}')\}| \leq \alpha(\ell).$$

From the fact that the weight functions sum to 1 and in view of Assumption W_1,

$$\Lambda_{n3}(\mathbf{y}, \mathbf{y}') \leq nh \left\{ \sum_{i=1}^{n} w_{ni}(x,h) \right\} \left\{ \max_{1 \leq i \leq n} w_{ni}(x,h) \right\} \sum_{\ell = \lfloor h^{-1/2} \rfloor + 1}^{n} \alpha(\ell) = o_{\mathrm{P}}(1),$$

where the last equality is a consequence of Assumption A_1 that ensures that $\alpha(\ell) = O(\ell^{-a})$ for some $a > 6$. One can finally conclude that

$$\lim_{n \to \infty} \mathrm{Cov}\left\{ Z_{xn}(\mathbf{y}), Z_{xn}(\mathbf{y}') \right\} = \mathrm{Cov}\left\{ \mathbb{H}_x(\mathbf{y}), \mathbb{H}_x(\mathbf{y}') \right\}.$$

Now in order to show the asymptotic tightness of Z_{xh}, 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 van der Vaart & Wellner (1996), the process Z_{xn} is asymptotically tight in $\ell^\infty([0,1]^2)$ if and only if for any $\delta > 0$, $\mathfrak{W}_\delta(Z_{xn}, \mathbb{R}^2)$ converges to zero in probability. To show that it is indeed the case, one proceeds similarly as in Theorem 3 of Bickel & Wichura (1971). 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

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

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

Lemma 1 *Suppose that $\sqrt{nh}h^2 < \infty$ and Assumptions $\mathcal{A}_2$ and W_2–W_3 are satisfied. Then, for n sufficiently large, one has for any $\epsilon > 0$ that*

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

Now let $\mu_x = \nu_x \otimes \lambda$, where λ is the ρ-measure of A. One then needs to find $\beta > 1$ and $C \in \mathbb{R}$ (that may depend on ϵ and β) such that

$$\mathrm{P}\left[|\mathbb{H}_{xh}\{A_\gamma(i,j)\}| \geq \epsilon \right] \leq C \left[\mu_x \left\{ A_\gamma(i,j) \right\} \right]^\beta . \tag{2.4}$$

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

Lemma 2 *If Assumptions $\mathcal{A}_1$–$\mathcal{A}_2$, W_1 and W_{11}–W_{13} are satisfied, one can find a finite constant $\omega > 0$ such that for any rectangle $A \subseteq \mathbb{R}^2$,*

$$\mathrm{E}\left\{ |\mathbb{H}_{xh}(A)|^6 \right\} \leq \omega \sum_{k=1}^{3} \left\{ \mu_x(A) + h^2 \right\}^k (nh)^{-3+k}. \tag{2.5}$$

In view of equations (2.4) and (2.5), the Markov inequality entails

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

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

$$\begin{aligned} a_{xh}(\gamma,\beta) &= \mu_{x\gamma}^{1-\beta}\left\{(nh)^{-2}+n^{-1}+h^4\right\}+\mu_{x\gamma}^{2-\beta}\left\{(nh)^{-1}+h^2\right\} \\ &\quad +\mu_{x\gamma}^{3-\beta}+\mu_{x\gamma}^{-\beta}\left\{n^{-2}+n^{-1}h^3+h^6\right\}. \end{aligned} \tag{2.6}$$

From the definition of $\mu_{x\gamma}$, one has $(nh)^{-1-2\gamma} \le \mu_{x\gamma} \le (nh)^{-1/2-\gamma}$. Hence, a meticulous inspection of Equation (2.6) shows that for any small γ and any β close to 1, one has $a_{xh}(\gamma,\beta) < 1$ since h satisfies $nh^5 < \infty$. Hence, one can find a constant C (that depends on ϵ, β, γ and τ) such that Equation (2.4) is satisfied. The asymptotic ρ-equicontinuity follows, for instance, from an extension of Theorem 3 in Bickel & Wichura (1971).

2.2.4 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 Bücher & Volgushev (2013), 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.

2.3 Investigation of a second estimator of C_x

2.3.1 Description of the estimator

As noted by Gijbels *et al.* (2011), 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

$$G_{xh}(v_1, v_2) = \sum_{i=1}^{n} w_{ni}(x, h)\, \mathbb{I}\left\{F_{1X_ih_1}(Y_{1i}) \leq v_1, F_{2X_ih_2}(Y_{2i}) \leq v_2\right\}$$

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

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

2.3.2 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 $\mathcal{A}_1$, $\mathcal{A}_2^\star$, $\mathcal{A}_3$–$\mathcal{A}_4$ and W_1–W_{13} are satisfied. If*

$$n \min(h_1, h_2) \to \infty, \quad n \max(h_1^5, h_2^5) < \infty \quad \textit{and} \quad h/\min(h_1, h_2) < \infty,$$

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 Veraverbeke *et al.* (2011) in the i.i.d. case. In particular, the asymptotic bias and covariance function are the same as those found by these authors.

2.3.3 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

$$\widetilde{G}_{xh}(u_1, u_2) = \sum_{i=1}^{n} w_{ni}(x, h) \, \mathbb{I}\left(U_i \leq u, V_i \leq v\right).$$

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 2.1 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 $\mathcal{A}_1$, $\mathcal{A}_2^\star$, W_1–W_5 and W_{11}–W_{13} are satisfied, Proposition 2.1 ensures that $\sqrt{nh}\{\Lambda(\widetilde{G}_{xh}) - C_x\}$ converges weakly to $\Lambda'_{C_x}(\mathbb{G}_x) = \widetilde{\mathbb{C}}_x$.

It remains to show that $\sqrt{nh}\{\Lambda(G_{xh}) - \Lambda(\widetilde{G}_{xh})\}$ is asymptotically negligible. As pointed out by Veraverbeke *et al.* (2011), this is closely related to the asymptotic behavior of the processes

$$\widetilde{Z}_{1xn} = Z_{1xn} - \mathrm{E}(Z_{1xn}) \quad \text{and} \quad \widetilde{Z}_{2xn} = Z_{2xn} - \mathrm{E}(Z_{2xn}),$$

where for $j = 1, 2$ and $z_t = x + t\,C\,h$,

$$Z_{jxn}(t, u) = \sqrt{nh_j}\, F_{jz_t h_j} \left\{ F_{jz_t}^{-1}(u_1) \right\}.$$

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

Lemma 3 *Suppose that Assumptions $\mathcal{A}_1$, $\mathcal{A}_4$, W_1, W_6–W_{13} are satisfied. Then, as long as $nh_1^5 < \infty$ and $nh_2^5 < \infty$, 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 Veraverbeke *et al.* (2011), one obtains that $\sqrt{nh}\{\Lambda(G_{xh}) - \Lambda(\widetilde{G}_{xh})\} = o_{\mathrm{P}}(1)$. Hence,

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

which completes the proof.

2.4 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 Nelsen (2006) 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

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

The latter have been estimated from 1 000 replicates under each of the scenario considered ; the results are reported in Table 2.1 for AISB and in Table 2.2 for AIV. All the simulations have been done using the Local–Linear weights defined for $\widetilde{X}_i = (X_i - x)/h$ by

$$w_{ni}(x,h) = K(\widetilde{X}_i) \left(\frac{S_{n,2} - \widetilde{X}_i \, S_{n,1}}{S_{n,0}\, S_{n,2} - S_{n,1}^2} \right),$$

where $K(y) = 35(1-y^2)^3\, \mathbb{I}(|y| \leq 1)/32$ is the triweight function and $S_{n,\ell} = \sum_{i=1}^n \widetilde{X}_i^\ell \, K(\widetilde{X}_i)$, $\ell \in \{0,1,2\}$. Since negative weights can occur, they are taken to be zero in that case. By arguments similar as those in Li & Racine (2007), one can show that Assumptions W_1–W_{13} are satisfied whenever Assumption $\mathcal{A}_1$ on the alpha-mixing coefficients is satisfied.

From the entries in Table 2.1, one notes that $\widetilde{C}_{xh}$ outperforms C_{xh} in terms of AISB when $(\rho_{12}, \rho_{23}, \rho_{13}) \in \{(0.9, 0.8, 0.8), (-0.9, 0.8, -0.8)\}$; 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 $\widetilde{C}_{xh}$ stabilizes as the bandwidth parameter h takes large values ; it is not the case for C_{xh}.

As can be seen in Table 2.2, the integrated variance 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} does 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.

Tableau 2.1 – 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

Tableau 2.2 – 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

CHAPITRE 3

Nonparametric measures of local causality and tests of local non-causality in time series

Résumé

L'étude des relations de causalité entre deux séries chronologiques $(Y_t, Z_t)_{t\in\mathbb{Z}}$ revêt un interêt particulier en économie et en finance. Dans ce domaine, on fait souvent appel à la notion de causalité de Granger. Dans le cas de modèles Markoviens de premier ordre, cette notion est définie en terme de la fonction de répartition conditionnelle du vecteur (Y_t, Z_{t-1}) étant donné Y_{t-1}. À ce jour, les mesures de causalité existantes sont globales, en ce sens où, si la nature de la relation entre Y_t et Z_{t-1} change en fonction des valeurs prises par Y_{t-1}, ces dernières seront incapables de le détecter. Pour résoudre ce problème, nous proposons dans cet articles de mesures *locales* de la causalité de Granger. Ces mesures sont basées sur la copule conditionnelle du vecteur (Y_t, Z_{t-1}) étant donné $Y_{t-1} = x$. En exploitant les résultats asymptotiques de deux estimateurs de la copule conditionnelle pour des données sérielles, la normalité asymptotique des estimateurs non-paramétriques proposés y est établie et des intervalles de confiance y sont construits. Des tests de non-causalité sont aussi développés. Les performances des méthodes proposées sont examinées par l'entremise de simulations numériques, et leurs utilités sont illustrées à travers l'analyse du prix et du volume d'échange du Standard & Poors 500.

Abstract

The study of the causal relationships in a process $(Y_t, Z_t)_{t\in\mathbb{Z}}$ is a subject of a particular interest in finance and economy. A widely-used approach is to consider the notion of Granger causality, which in the case of first order Markovian processes is based on the joint distribution function of (Y_t, Z_{t-1}) given Y_{t-1}. The Granger causality measures proposed so far are global in the sense that if the relationship between Y_t and Z_{t-1} changes with the value taken by Y_{t-1}, this will not be captured. To circumvent this limitation, this paper proposes *local* Granger causality indices based on the conditional copula of (Y_t, Z_{t-1}) given $Y_{t-1} = x$. Exploiting the asymptotic behavior of two kernel-based conditional copula estimators for α-mixing processes, the asymptotic normality of nonparametric estimators of these local Granger indices is deduced and confidence intervals are built. Tests of local non-causality are developed as well. The efficiency of the proposed methods is investigated via simulations and their usefulness is illustrated on the bivariate time series of Standard & Poor's 500 prices and trading volumes.

3.1 Introduction

The concept of causality as originally introduced by Wiener (1956) and Granger (1969) is helpful for studying the dynamic relationships in multivariate time series. This notion is defined in terms of predictability at horizon one of a random variable (or random vector)

Y from its past and the past of another random variable (or vector) Z. Specifically, assume that data are available for the process $(Y_t, Z_t)_{t\in\mathbb{Z}}$, and let $\mathbf{Y}_{t-1}$, $\mathbf{Z}_{t-1}$ be the observations up to time $t-1$ on Y and Z, respectively. According to Granger (1969), the causality from Z to Y *one period ahead* is defined as follows : Z is said to cause Y if $\mathbf{Z}_{t-1}$ can help predict Y_t, conditional on $\mathbf{Y}_{t-1}$.

Many works considered testing the null hypothesis of non-causality. For example, testing causality has been investigated for multivariate ARMA models by Boudjellaba *et al.* (1992) and Boudjellaba *et al.* (1994). Because the Granger non-causality is a form of conditional independence, tests can be deduced from standard conditional independence tests ; see Florens & Fougere (1996), for instance. In the context of i.i.d. data, such procedures were derived by Song (2009), Huang (2010), Bergsma (2013), Su & Spindler (2013) and Linton & Gozalo (2014), among others. Generalizations to the case of time series have been investigated by Su & White (2008) and Su & White (2012) under α-mixing and by de Matos & Fernandes (2007) and Su & White (2008) under β-mixing. See also the recent contributions by Bouezmarni *et al.* (2012), Wang & Hong (2013) and Bouezmarni & Taamouti (2014).

When the hypothesis of non-causality is rejected, one may be interested in measuring the strength of this causal relationship. The first causality measures were proposed by Geweke (1982) and Geweke (1984) using the mean-squared forecast errors, and by Gouriéroux *et al.* (1987) based on the Kullback–Leibler information. Causality indices under parametric models were later investigated by Polasek (1994) and Polasek (2002), and by Dufour & Taamouti (2010) under ARMA models, where measures for short and long run were proposed. Mainly inspired by the fact that these measures suffer from model misspecification, nonparametric indices where proposed by Taamouti *et al.* (2014) using the Kullback–Leibler information and nonparametric density copula estimators. Recently, Zhang *et al.* (2016) investigated causality measures at multiple horizons for exchange rate and commodity prices.

It is worth mentioning that all of the above-cited papers focus on characterizing the global relationship between Y_t and $\mathbf{Z}_{t-1}$, conditional on $\mathbf{Y}_{t-1}$. Unfortunately, if the nature of the link between Y_t and $\mathbf{Z}_{t-1}$ changes with the value taken by $\mathbf{Y}_{t-1}$, this feature will not necessarily be captured by global measures. A possible solution to this issue is to compute the partial correlation coefficient. However, doing so implicitly assumes a linear relationship and the measure depends on the marginal behavior. In other words, such an approach would suffer from the same drawbacks as the classical Pearson correlation coefficient.

To circumvent these limitations, this paper proposes nonparametric local Granger causality indices for measuring the strength of the relationship in $(Y_t, \mathbf{Z}_{t-1})$ given a particular value taken by Y_{t-1}. In order to simplify the presentation, a focus is put on Markovian models of order one. In that particular case, one considers the dependence structure of (Y_t, Z_{t-1}) given $Y_{t-1} = x$ as captured by its associated conditional copula. This approach allows for the definition of nonparametric measures of local causality that do not suffer from the drawbacks that arise when using partial correlations. Specifically, let $\{(Y_t, Z_t)\}_{t\in\mathbb{Z}}$ be a stationary process and define the local causality distribution function

$$H_x^{Z\to Y}(y,z) = \mathrm{P}\left(Y_t \le y, Z_{t-1} \le z | Y_{t-1} = x\right). \tag{3.1}$$

Then if the conditional marginal distributions $F_{1x}(y) = \mathrm{P}\left(Y_t \le y | Y_{t-1} = x\right)$ and $F_{2x}(z) = \mathrm{P}\left(Z_{t-1} \le z | Y_{t-1} = x\right)$ are continuous, Sklar's Theorem guarantees the existence of a unique copula $C_x^{Z\to Y} : [0,1]^2 \to [0,1]$ such that

$$H_x^{Z\to Y}(y,z) = C_x^{Z\to Y}\left\{F_{1x}(y), F_{2x}(z)\right\}.$$

The bivariate function $C_x^{Z\to Y}$ will be called the *local causality copula* and corresponds to the dependence structure of (Y_t, Z_{t-1}) given $Y_{t-1} = x$.

The first goal of this paper is to describe nonparametric estimators of $C_x^{Z\to Y}$ in a general framework of serially dependent bivariate data. The weak convergence of suitably standardized versions of these conditional copula estimators is deduced from general results by Bouezmarni *et al.* (2016) under some conditions on the α-mixing coefficients of $\{(Y_t, Z_t)\}_{t\in\mathbb{Z}}$. Then, local causality indices are defined for measuring the strength of the causal relationship in a bivariate time series and nonparametric estimators based on the empirical conditional copulas are proposed. Their asymptotic normality is established from the functional delta method. Tests for the null hypothesis of local non-causality are developed as well.

The paper is organized as follows. In Section 2, two estimators of the local causality copula are described and their asymptotic behavior is obtained in the light of results by Bouezmarni *et al.* (2016). In Section 3, general local causality indices are defined and the large-sample behavior of nonparametric estimators is investigated. A consistent estimator of the asymptotic variance is also proposed, leading to interval estimations of the local causality measures. In Section 4, tests for the null hypothesis of local non-causality are developed. Section 5 investigates the sampling properties of point and interval estimators of causality indices based on the Spearman and Kendall measures of association. The efficiency of tests of local non-causality is studied as well. An illustration on financial data is provided in Section 6. Technical details and the assumptions required in Section 2 and 3 are relegated to the Appendix.

3.2 Estimation of the local causality copula

3.2.1 Two estimators of $C_x^{Z\to Y}$

Let $(Y_1, Z_1), \ldots, (Y_{n+1}, Z_{n+1})$ be a realization of a stationary process $\{(Y_t, Z_t)\}_{t\in\mathbb{Z}}$. In that context, an estimator of the joint conditional distribution in (4.2) is

$$H_{xh}^{Z\to Y}(y,z) = \sum_{i=2}^{n+1} \mathcal{K}\left(\frac{Y_{i-1}-x}{h}\right) \mathbb{I}\left(Y_i \le y, Z_{i-1} \le z\right),$$

where $\mathcal{K} = \mathcal{K}_n$ is a non-negative kernel-based weight that may depend on $Y_1, \ldots, Y_n$ and $h = h_n$ is a bandwidth parameter. Hereafter, it is assumed that

$$\sum_{i=1}^{n} \mathcal{K}\left(\frac{Y_i - x}{h}\right) = 1.$$

A first estimator of the local causality copula arises upon noting that $C_x^{Z\to Y}$ can be extracted from $H_x^{Z\to Y}$ via

$$C_x^{Z\to Y}(u,v) = H_x^{Z\to Y}\left\{F_{1x}^{-1}(u), F_{2x}^{-1}(v)\right\}.$$

From this representation, a natural plug-in estimator of $C_x^{Z\to Y}$ is given by

$$C_{xh}^{Z\to Y}(u,v) = H_{xh}^{Z\to Y}\left\{F_{1xh}^{-1}(u), F_{2xh}^{-1}(v)\right\}, \tag{3.2}$$

where F_{1xh}^{-1} and F_{2xh}^{-1} are the left-continuous generalized inverses of
$F_{1xh}(y) = \lim_{z\to\infty} H_{xh}^{Z\to Y}(y,z)$ and $F_{2xh}(z) = \lim_{y\to\infty} H_{xh}^{Z\to Y}(y,z)$, respectively.

As noted by Veraverbeke *et al.* (2011) and Gijbels *et al.* (2011) in the i.i.d. case, the plug-in estimator $C_{xh}^{Z\to Y}$ may be severely biased, especially when the conditional marginal distributions strongly depend on the covariate. For that reason, a second estimator that aims at reducing this possible effect of the covariate on the margins is proposed. To this end, define for each $i \in \{1, \ldots, n\}$ the *pseudo-uniformized* observations $(\widetilde{U}_i, \widetilde{V}_i) =$

$(F_{1Y_ih_1}(Y_{i+1}), F_{2Y_ih_2}(Z_i))$, where h_1 and h_2 are bandwidth parameters that may differ from h. Then, let

$$G_{xh}^{Z\to Y}(y,z) = \sum_{i=1}^{n} \mathcal{K}\left(\frac{Y_i - x}{h}\right) \mathbb{I}\left(\widetilde{U}_i \le y, \widetilde{V}_i \le z\right).$$

An alternative estimator of $C_x^{Z\to Y}$ is then given by

$$\widetilde{C}_{xh}^{Z\to Y}(u,v) = G_{xh}^{Z\to Y}\left\{G_{1xh}^{-1}(u), G_{2xh}^{-1}(v)\right\}, \tag{3.3}$$

where $G_{1xh}(y) = \lim_{z\to\infty} G_{xh}^{Z\to Y}(y,z)$ and $G_{2xh}(z) = \lim_{y\to\infty} G_{xh}^{Z\to Y}(y,z)$.

3.2.2 Weak convergence

This section describes the asymptotic behavior of the processes

$$\mathbb{C}_{xh}^{Z\to Y} = \sqrt{nh}\left(C_{xh}^{Z\to Y} - C_x^{Z\to Y}\right) \quad \text{and} \quad \widetilde{\mathbb{C}}_{xh}^{Z\to Y} = \sqrt{nh}\left(\widetilde{C}_{xh}^{Z\to Y} - C_x^{Z\to Y}\right).$$

These large-sample results are derived under the assumption that the stationary process $\{(Y_t, Z_t)\}_{t\in\mathbb{Z}}$ is α-mixing. Specifically, for each $r \in \mathbb{N}\cup\{0\}$, define its associated α-mixing coefficient of lag r by

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

where $\mathcal{F}_a^b$ is the σ-field generated by $\{(Y_t, Z_t)\}_{a\le t\le b}$ and

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

Then $\alpha(r) \to 0$ as $r \to \infty$, which means that $\{(Y_t, Z_t)\}_{t\in\mathbb{Z}}$ is α-mixing.

Because the estimation of the local causality copula is based on the trivariate process $(Y_t, Z_{t-1}, Y_{t-1})_{t\in\mathbb{Z}}$, where the third component is the conditioning variable, or covariate, the setup in this paper is a special case of that investigated by Bouezmarni *et al.* (2016). In this latter work, the estimation of conditional copulas from general three-dimensional

α-mixing processes is considered. In particular, one can deduce from their Proposition 2.1 the weak convergence of the process

$$\mathbb{H}_{xh}^{Z\to Y} = \sqrt{nh}\left(H_{xh}^{Z\to Y} - H_x^{Z\to Y}\right).$$

This weak convergence takes place in the space $\ell^\infty(\mathbb{R}^2)$ of bounded functions in $\mathbb{R}^2$. The result is formally stated next. In the sequel, $\dot{H}_x^{Z\to Y}$ and $\ddot{H}_x^{Z\to Y}$ denote the first and second derivatives of $H_x^{Z\to Y}$ with respect to x, *i.e.*

$$\dot{H}_x^{Z\to Y}(y,z) = \frac{\partial}{\partial x} H_x^{Z\to Y}(y,z) \quad \text{and} \quad \ddot{H}_x^{Z\to Y}(y,z) = \frac{\partial^2}{\partial x^2} H_x^{Z\to Y}(y,z).$$

Proposition 1 *Suppose that Assumptions $\mathcal{A}_1$–$\mathcal{A}_2$, W_1–W_5 and W_{11}–W_{13} are satisfied. If $nh\to\infty$ and $nh^5 \to K^2 < \infty$, the empirical process $\mathbb{H}_{xh}^{Z\to Y}$ converges weakly in the space $\ell^\infty(\mathbb{R}^2)$ to a Gaussian limit $\mathbb{H}_x^{Z\to Y}$ such that for K_2–K_4 defined in Assumptions W_2–W_4,*

$$\mathrm{E}\left\{\mathbb{H}_x^{Z\to Y}(y,z)\right\} = K\left\{K_2\,\dot{H}_x^{Z\to Y}(y,z) + K_3\,\ddot{H}_x^{Z\to Y}(y,z)\right\}$$

and

$$\begin{aligned}\mathrm{Cov}\left\{\mathbb{H}_x^{Z\to Y}(y,z), \mathbb{H}_x^{Z\to Y}(y',z')\right\} &= K_4\,H_x^{Z\to Y}\left\{\min(y,y'),\min(z,z')\right\} \\ &\quad - K_4\,H_x^{Z\to Y}(y,z)\,H_x^{Z\to Y}(y',z').\end{aligned}$$

Because the estimator $C_{xh}^{Z\to Y}$ can be expressed in terms of a Hadamard differentiable functional of $H_{xh}^{Z\to Y}$ (see Bücher & Volgushev (2013)), the weak convergence of $\mathbb{C}_{xh}^{Z\to Y}$ in the space $\ell^\infty([0,1]^2)$ of bounded functions in $[0,1]^2$ can be deduced from an application of the functional delta method (see van der Vaart & Wellner (1996) for more details). Before stating the result, define

$$C_x^{[1]}(u,v) = \frac{\partial}{\partial u} C_x^{Z\to Y}(u,v) \quad \text{and} \quad C_x^{[2]}(u,v) = \frac{\partial}{\partial v} C_x^{Z\to Y}(u,v).$$

Corollary 1 *Under the conditions of Proposition 1 and if in addition Assumption $\mathcal{A}_3$ holds, $\mathbb{C}_{xh}^{Z\to Y}$ converges weakly in the space $\ell^\infty([0,1]^2)$ to*

$$\mathbb{C}_x^{Z\to Y}(u,v) = \alpha_x^{Z\to Y}(u,v) - C_x^{[1]}(u,v)\,\alpha_x^{Z\to Y}(u,1) - C_x^{[2]}(u,v)\,\alpha_x^{Z\to Y}(1,v),$$

where in terms of the process $\mathbb{H}_x^{Z\to Y}$ defined in Proposition 1,

$$\alpha_x^{Z\to Y}(u,v) = \mathbb{H}_x^{Z\to Y}\{F_{1x}^{-1}(u), F_{2x}^{-1}(v)\}.$$

In view of the bias function given in Proposition 1, a consequence of Corollary 1 is that the asymptotic bias of $\mathbb{C}_{xh}^{Z\to Y}$ is

$$\begin{aligned}\mathbb{B}_x(u,v) &= K\Big[K_2\,\dot{C}_x^{Z\to Y}(u,v) + K_3\,\ddot{H}_x^{Z\to Y}\left\{F_{1x}^{-1}(u), F_{2x}^{-1}(v)\right\} \\ &\qquad - K_3\,C_x^{[1]}(u,v)\ddot{F}_{1x}\left\{F_{1x}^{-1}(u)\right\} - K_3\,C_x^{[2]}(u,v)\ddot{F}_{2x}\left\{F_{2x}^{-1}(v)\right\}\Big].\end{aligned}$$

The covariance function can be derived as well, but its expression is cumbersome.

As noted by Bouezmarni *et al.* (2016) in the general case, the asymptotic behavior of $\mathbb{C}_{xh}^{Z\to Y}$ under α-mixing is the same as that under serial independence. In other words, the impact of time-dependency is asymptotically negligible. This behavior is a consequence of Assumption $\mathcal{A}_1$ on the α-mixing coefficients, combined with the use of a kernel function that smooths the covariate space in a shrinking neighborhood of x as n goes to infinity. Note however that compared to the i.i.d. setting, Assumptions W_1 and W_{11}–W_{13} on the weight functions are needed in order to tackle moments of order six entailed by time-dependency.

Before stating the result on the weak convergence of $\widetilde{\mathbb{C}}_{xh}^{Z\to Y}$, introduce the Gaussian process $\mathbb{G}_x^{Z\to Y}$ that arises as the limit of $\mathbb{H}_{xh}^{Z\to Y}$ when (Y_1, Z_1), ..., (Y_{n+1}, Z_{n+1}) is replaced by $(U_1, V_1), \ldots, (U_n, V_n)$, where $U_i = F_{1Y_i}(Y_{i+1})$ and $V_i = F_{2Y_i}(Z_i)$. This situation corresponds to a case where the marginal conditional distributions are known. One then deduces from Proposition 1 with $H_x^{Z\to Y} = C_x^{Z\to Y}$ that the bias of $\mathbb{G}_x^{Z\to Y}$ is given for $(u,v)\in[0,1]^2$ by

$$\widetilde{\mathbb{B}}_x(u,v) = K\left\{K_2\,\dot{C}_x^{Z\to Y}(u,v) + K_3\,\ddot{C}_x^{Z\to Y}(u,v)\right\}$$

and for $(u,v),(u',v')\in[0,1]^2$, its covariance function is

$$\begin{aligned}\mathrm{Cov}\left\{\mathbb{G}_x^{Z\to Y}(u,v), \mathbb{G}_x^{Z\to Y}(u',v')\right\} &= K_4\,C_x^{Z\to Y}\left\{\min(u,u'), \min(v,v')\right\} \\ &\quad - K_4\,C_x^{Z\to Y}(u,v)\,C_x^{Z\to Y}(u',v').\end{aligned}$$

The next proposition is deduced from Proposition 3.1 of Bouezmarni *et al.* (2016).

Proposition 2 *Suppose that Assumptions $\mathcal{A}_1$, $\mathcal{A}_3$–$\mathcal{A}_5$ and W_1–W_{13} are satisfied. If $n \min(h_1, h_2) \to \infty$, $n \max(h_1^5, h_2^5) < \infty$ and $h/\min(h_1, h_2) < \infty$, $\widetilde{\mathbb{C}}_{xh}$ converges weakly in the space $\ell^\infty([0,1]^2)$ to the Gaussian process*

$$\widetilde{\mathbb{C}}_x^{Z\to Y}(u,v) = \mathbb{G}_x^{Z\to Y}(u,v) - C_x^{[1]}(u,v)\, \mathbb{G}_x^{Z\to Y}(u,1) - C_x^{[2]}(u,v)\, \mathbb{G}_x^{Z\to Y}(1,v).$$

One can show that the asymptotic bias of $\widetilde{\mathbb{C}}_{xh}$ is $\widetilde{\mathbb{B}}_x$. It can also be seen from Proposition 1 and Proposition 2 that $\mathbb{C}_{xh}$ and $\widetilde{\mathbb{C}}_{xh}$ share the same covariance structure. However, as pointed out in Gijbels *et al.* (2011), they have a different bias function in general.

3.3 Measuring local causality

3.3.1 Theoretical measures of local causality

Measuring the strength of the causal relationship from Z to Y can be done using functionals of the local causality copula. Specifically, let $\Lambda : \ell^\infty([0,1]^2) \to \mathbb{R}$ be such that $\Lambda(\Pi) = 0$, $\Lambda(M) = 1$ and $\Lambda(W) = -1$, where $\Pi(u,v) = uv$, $M(u,v) = \min(u,v)$ and $W(u,v) = \max(u+v-1, 0)$ are respectively the independence, perfect positive dependence and perfect negative dependence copulas. A measure of local causality from Z to Y at x based on Λ is then

$$\theta_{\Lambda,x}^{Z\to Y} = \Lambda\left(C_x^{Z\to Y}\right). \tag{3.4}$$

This measure has the desirable property of being marginal-free. Among the possibilities, one can define local causality indices based on the popular Spearman and Kendall functionals given respectively for $\delta \in \ell^\infty([0,1]^2)$ by

$$\Lambda_\rho(\delta) = 12 \int_{[0,1]^2} \delta(u,v)\, \mathrm{d}u\mathrm{d}v - 3 \quad \text{and} \quad \Lambda_\tau(\delta) = 4 \int_{[0,1]^2} \delta(u,v)\, \mathrm{d}\delta(u,v) - 1.$$

The corresponding measures of local causality will be referred to $\rho_x^{Z\to Y}$ and $\tau_x^{Z\to Y}$ in the sequel.

3.3.2 Nonparametric estimators

The estimation of the local causality index $\theta_{\Lambda,x}^{Z\to Y}$ defined in Equation (3.4) can be based on the empirical local causality copulas $C_{xh}^{Z\to Y}$ and $\widetilde{C}_{xh}^{Z\to Y}$. Specifically, two estimators of $\theta_{\Lambda,x}^{Z\to Y}$ are given by

$$\theta_{\Lambda,xh}^{Z\to Y} = \Lambda\left(C_{xh}^{Z\to Y}\right) \quad \text{and} \quad \widetilde{\theta}_{\Lambda,xh}^{Z\to Y} = \Lambda\left(\widetilde{C}_{xh}^{Z\to Y}\right).$$

The next result establishes the asymptotic normality of

$$\Theta_{\Lambda,xh}^{Z\to Y} = \sqrt{nh}\left(\theta_{\Lambda,xh}^{Z\to Y} - \theta_{\Lambda,x}^{Z\to Y}\right) \quad \text{and} \quad \widetilde{\Theta}_{\Lambda,xh}^{Z\to Y} = \sqrt{nh}\left(\widetilde{\theta}_{\Lambda,xh}^{Z\to Y} - \theta_{\Lambda,x}^{Z\to Y}\right).$$

Proposition 3 *Assume that Λ is Hadamard differentiable with derivative at g given by Λ'_g. Then, let*

$$\sigma_{\Lambda,x}^2 = \mathrm{Var}\left\{\Lambda'_{C_x^{Z\to Y}}\left(\mathbb{C}_x^{Z\to Y}\right)\right\} = \mathrm{Var}\left\{\Lambda'_{C_x^{Z\to Y}}\left(\widetilde{\mathbb{C}}_x^{Z\to Y}\right)\right\}. \tag{3.5}$$

(i) Under the conditions of Corollary 1, $\Theta_{\Lambda,xh}^{Z\to Y}$ converges in law to the Normal distribution with mean $\mu_{\Lambda,x} = \Lambda'_{C_x^{Z\to Y}}(\mathbb{B}_x)$ and variance $\sigma_{\Lambda,x}^2$.

(ii) Under the conditions of Proposition 2, $\widetilde{\Theta}_{\Lambda,xh}^{Z\to Y}$ converges in law to the Normal distribution with mean $\widetilde{\mu}_{\Lambda,x} = \Lambda'_{C_x^{Z\to Y}}(\widetilde{\mathbb{B}}_x)$ and variance $\sigma_{\Lambda,x}^2$.

Note that the functionals associated to the Kendall and Spearman causality measures are Hadamard differentiable with derivatives given respectively by

$$\begin{aligned}
(\Lambda_\rho)'_g(\delta) &= 12\int_{[0,1]^2} \delta(u,v)\,\mathrm{d}u\mathrm{d}v, \\
(\Lambda_\tau)'_g(\delta) &= 4\int_{[0,1]^2} \left\{\delta(u,v)\,\mathrm{d}g(u,v) + g(u,v)\,\mathrm{d}\delta(u,v)\right\}.
\end{aligned}$$

Hence, the conclusions of Proposition 3 apply in these cases.

3.3.3 Estimation of the asymptotic variance

If the goal is to build a confidence interval for a local causality measure, Proposition 3 cannot be used directly since the asymptotic variance $\sigma^2_{\Lambda,x}$ in Equation (3.5) is unknown. In order to motivate the form of the estimator of $\sigma^2_{\Lambda,x}$ that will be introduced, first consider a context where one wants to estimate a conditional mean $\mu_x = \mathrm{E}(Y|X = x)$. Based on observations $(Y_1, X_1), \ldots, (Y_n, X_n)$, a natural estimator is given by the weighted mean

$$\mu_{xh} = \sum_{i=1}^{n} \mathcal{K}\left(\frac{X_i - x}{h}\right) Y_i.$$

Then, under Assumptions W_1–W_5, it can be shown that $\sqrt{nh}(\mu_{xh} - \mu_x)$ is asymptotically Normal with variance $K_4\sigma^2_x$, where $\sigma^2_x = \mathrm{Var}(Y|X = x)$. It can also be shown that a consistent estimator of σ^2_x is given by

$$\sigma^2_{xh} = \sum_{i=1}^{n} \mathcal{K}\left(\frac{X_i - x}{h}\right) (Y_i - \mu_{xh})^2 .$$

Hence, a consistent estimator of the asymptotic variance of μ_{xh} is given by $\widehat{K}_4\sigma^2_{xh}$, where in view of Assumption W_4,

$$\widehat{K}_4 = nh \sum_{i=1}^{n} \left\{\mathcal{K}\left(\frac{X_i - x}{h}\right)\right\}^2 .$$

Now in order to adapt the idea to the context of local causality indices, recall that $\sigma^2_{\Lambda,x}$ is the variance of $\Lambda'_{C_x^{Z\to Y}}(\mathbb{C}_x^{Z\to Y})$. Then, observe that the process $\mathbb{C}_x^{Z\to Y}$ can be seen as the weak limit of

$$\sqrt{nh} \sum_{i=2}^{n+1} \mathcal{K}\left(\frac{Y_{i-1} - x}{h}\right) L_{x,i},$$

where for each $(u, v) \in [0, 1]^2$,

$$\begin{aligned} L_{x,i}(u, v) \quad &= \quad \mathbb{I}\left\{Y_i \le F_{1x}^{-1}(u), Z_{i-1} \le F_{2x}^{-1}(v)\right\} - C_x^{Z\to Y}(u, v) \\ &\quad - C_x^{[1]}(u, v) \left[\mathbb{I}\left\{Y_i \le F_{1x}^{-1}(u)\right\} - u\right] \\ &\quad - C_x^{[2]}(u, v) \left[\mathbb{I}\left\{Z_{i-1} \le F_{2x}^{-1}(v) - v\right\}\right] . \end{aligned}$$

Hence, since Hadamard derivatives are linear functionals (see van der Vaart & Wellner (1996)), $\Lambda'_{C_x^{Z\to Y}}(\mathbb{C}_x^{Z\to Y})$ can be taken as the limit of $\sqrt{nh}\,\lambda_{xh}$, where in terms of $\lambda_{x,i} = \Lambda'_{C_x^{Z\to Y}}(L_{x,i})$,

$$\lambda_{xh} = \sum_{i=2}^{n+1} \mathcal{K}\left(\frac{Y_{i-1}-x}{h}\right)\lambda_{x,i}.$$

In view of the above discussion, an estimator of $\sigma^2_{\Lambda,x}$ could therefore be based on $\lambda_{x,2},\ldots,$ $\lambda_{x,n+1}$. However, since the marginal conditional distributions F_{1x}, F_{2x} and the partial derivatives $C_x^{[1]}$, $C_x^{[2]}$ are unknown, consider instead the version

$$\begin{aligned}\widehat{L}_{x,i}(u,v) &= \mathbb{I}\left\{Y_i \le F_{1xh}^{-1}(u), Z_{i-1} \le F_{2xh}^{-1}(v)\right\} \\ &\quad - \widehat{C}_x^{[1]}(u,v)\,\mathbb{I}\left\{Y_i \le F_{1xh}^{-1}(u)\right\} \\ &\quad - \widehat{C}_x^{[2]}(u,v)\,\mathbb{I}\left\{Z_{i-1} \le F_{2xh}^{-1}(v)\right\},\end{aligned}$$

where $\widehat{C}_x^{[1]}$ and $\widehat{C}_x^{[2]}$ are estimators of the partial derivatives of $C_x^{\mathrm{Z}\to\mathrm{Y}}$ that are uniformly consistent in the sense that for any $\varepsilon > 0$,

$$\sup_{\substack{u\in[\varepsilon,1-\varepsilon]\\ v\in[0,1]}} \left|\widehat{C}_x^{[1]}(u,v) - C_x^{[1]}(u,v)\right| \quad \text{and} \quad \sup_{\substack{v\in[\varepsilon,1-\varepsilon]\\ u\in[0,1]}} \left|\widehat{C}_x^{[2]}(u,v) - C_x^{[2]}(u,v)\right|$$

converge in probability to zero. Then, letting $\widehat{\lambda}_{x,i} = \Lambda'_{C_x^{Z\to Y}}(\widehat{L}_{x,i})$, the proposed estimator of $\sigma^2_{\Lambda,x}$ is

$$\widehat{\sigma}^2_{\Lambda,x} = \widehat{K}_4 \sum_{i=2}^{n+1} \mathcal{K}\left(\frac{Y_{i-1}-x}{h}\right)\left(\widehat{\lambda}_{x,i} - \widehat{\lambda}_{xh}\right)^2,$$

where

$$\widehat{\lambda}_{xh} = \sum_{i=2}^{n+1} \mathcal{K}\left(\frac{Y_{i-1}-x}{h}\right)\widehat{\lambda}_{x,i}.$$

The consistency $\widehat{\sigma}^2_{\Lambda,x}$ is stated next.

Proposition 4 *Assume that Λ is Hadamard differentiable with derivative at g given by Λ'_g. Moreover, assume there exists a constant $D > 0$ such that*

$$\mathrm{P}\left(\max_{j=1,2}\sup_{u,v\in[0,1]}\left|\widehat{C}_x^{[j]}(u,v)\right| > D\right) \to 0.$$

Under the conditions of Corollary 1, $\widehat{\sigma}^2_{\Lambda,x}$ is a consistent estimator of $\sigma^2_{\Lambda,x}$.

When the functional Λ is linear, its Hadamard derivative is free of g, *i.e.* $\Lambda'_g = \Lambda'$ for all g; this happens in particular for the Spearman functional. In most cases, however, Λ'_g needs to be estimated. One can then replace $\Lambda'_{C_x^{Z\to Y}}$ by an estimator $\widehat{\Lambda'_{C_x^{Z\to Y}}}$ in the above procedure as long as for any $\delta \in \ell^\infty([0,1]^2)$,

$$\left|\widehat{\Lambda'_{C_x^{Z\to Y}}}(\delta) - \Lambda'_{C_x^{Z\to Y}}(\delta)\right| = o_{\mathrm{P}}(1).$$

This modification has no impact on the conclusion of Proposition 4. For example, in the case of the Kendall functional, $(\Lambda_\tau)'_{C_x^{Z\to Y}}$ is estimated by

$$\widehat{(\Lambda_\tau)'}_{C_x^{Z\to Y}}(\delta) = 4\int_{[0,1]^2} \left\{\delta(u,v)\,\mathrm{d}C_{xh}(u,v) + C_{xh}(u,v)\,\mathrm{d}\delta(u,v)\right\}.$$

3.3.4 Confidence intervals

Proposition 3 and Proposition 4 can now be combined to build confidence intervals. Neglecting the possible bias, an approximate confidence interval of level $1-\alpha$ for $\theta_{\Lambda,x}^{Z\to Y}$ based on $\theta_{\Lambda,xh}^{Z\to Y}$ is given by

$$\mathcal{CI}_\alpha(\theta_{\Lambda,x}^{Z\to Y}) = \theta_{\Lambda,xh}^{Z\to Y} \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{\widehat{\sigma}_{\Lambda,x}}{\sqrt{nh}},$$

where Φ is the cdf of the standard Normal distribution. The confidence interval is similar when $\theta_{\Lambda,x}^{Z\to Y}$ is estimated by $\widetilde{\theta}_{\Lambda,xh}^{Z\to Y}$.

Strictly speaking, this confidence interval is asymptotically of level $1-\alpha$ if and only if the large-sample bias of $\theta_{\Lambda,xh}^{Z\to Y}$ vanishes. Since this bias term is generally difficult to estimate, a strategy would be to choose a bandwidth h such that $nh^5 \to 0$ in order that the biases tend to zero asymptotically. However, based on many numerical experiments, it is usually safer to simply neglect it, since it is often close to zero.

3.4 Testing for local non-causality

Saying that there is no local causality relationship from Z to Y at x in the process $\{(Y_t, Z_t)\}_{t\in\mathbb{Z}}$ means that Y_t and Z_{t-1} are conditionally independent with respect to $Y_{t-1} = x$. In other words, the local causality copula corresponds to the independence copula in that case, *i.e.* $C_x^{Z\to Y}(u,v) = \Pi(u,v) = u\,v$. The following result is a consequence of Corollary 1 and of Proposition 2. Before stating it, let $\Omega_\Lambda : \ell^\infty([0,1]^4) \to \mathbb{R}$ be the functional such that for $\eta \in \ell^\infty([0,1]^4)$,

$$\Omega_\Lambda(\eta) = \Lambda'_{C_x^{Z\to Y}} \left[\Lambda'_{C_x^{Z\to Y}} \left\{ \eta(\cdot,\cdot,u',v') \right\} \right].$$

In other words, the operator inside the brackets is computed with respect to the first two arguments of η.

Proposition 5 *Suppose that the conditions in Corollary 1 and in Proposition 1 are satisfied respectively for $\mathbb{C}_{xh}^{Z\to Y}$ and $\widetilde{\mathbb{C}}_{xh}^{Z\to Y}$. Then under the null hypothesis of non-causality from Z to Y, $\sqrt{nh}\,\theta_{\Lambda,xh}^{Z\to Y}$ and $\sqrt{nh}\,\widetilde{\theta}_{\Lambda,xh}^{Z\to Y}$ are asymptotically Normal with variance $\sigma^2_{\Lambda,x} = K_4\,\Omega_\Lambda(\gamma)$, where for $(u,v),(u',v') \in [0,1]^2$,*

$$\gamma(u,u',v,v') = \{\min(u,u') - uu'\}\{\min(v,v') - vv'\}.$$

Proposition 5 can be exploited to test the null hypothesis of local non-causality from Z to Y. In that case, the null and alternative hypotheses are

$$\mathbb{H}_0 : \theta_{\Lambda,x}^{Z\to Y} = 0 \quad \text{and} \quad \mathbb{H}_1 : \theta_{\Lambda,x}^{Z\to Y} \neq 0.$$

Neglecting the asymptotic bias, a test based on the statistic $\theta_{\Lambda,xh}^{Z\to Y}$ will reject the null hypothesis of local non-causality whenever

$$\sqrt{nh}\left|\theta_{\Lambda,xh}^{Z\to Y}\right| > \widehat{K}_4\,\Omega_\Lambda(\gamma)\,\Phi^{-1}\left(1 - \frac{\alpha}{2}\right).$$

The test based on $\widetilde{\theta}_{\Lambda,xh}^{Z\to Y}$ is similar. For the Spearman and Kendall functionals, $\Omega_{\Lambda_\rho}(\gamma) = 1$ and $\Omega_{\Lambda_\tau}(\gamma) = 4/9$. This is a consequence of the fact that

$$\int_{[0,1]^4} \{\min(u,u') - uu'\}\{\min(v,v') - vv'\}\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}u'\,\mathrm{d}v' = 1.$$

3.5 Simulation study

3.5.1 Preliminaries

The aim of this section is to investigate how well the method introduced in this work perform. A particular attention will be given to the procedures based on the Spearman and Kendall functionals. By straightforward computations, explicit expressions for their empirical versions are given by

$$\begin{aligned}
\rho_{xh}^{Z\to Y} &= 12\sum_{i=2}^{n+1}\mathcal{K}\left(\frac{Y_{i-1}-x}{h}\right)\{1-F_{1xh}(Y_i)\}\{1-F_{2xh}(Z_{i-1})\}-3,\\
\widetilde{\rho}_{xh}^{Z\to Y} &= 12\sum_{i=1}^{n}\mathcal{K}\left(\frac{Y_i-x}{h}\right)\left\{1-G_{1xh_1}(\widetilde{U}_i)\right\}\left\{1-G_{2xh_2}(\widetilde{V}_i)\right\}-3,\\
\tau_{xh}^{Z\to Y} &= 4\sum_{i,j=2}^{n+1}\mathcal{K}\left(\frac{Y_{i-1}-x}{h}\right)\mathcal{K}\left(\frac{Y_{j-1}-x}{h}\right)\mathbb{I}\left(Y_j\le Y_i, Z_{j-1}\le Z_{i-1}\right)-1,\\
\widetilde{\tau}_{xh}^{Z\to Y} &= 4\sum_{i,j=1}^{n}\mathcal{K}\left(\frac{Y_i-x}{h}\right)\mathcal{K}\left(\frac{Y_j-x}{h}\right)\mathbb{I}\left(\widetilde{U}_j\le\widetilde{U}_i, \widetilde{V}_i\le\widetilde{V}_j\right)-1.
\end{aligned}$$

Following Gijbels *et al.* (2011), mildly modified versions of Kendall's measure were used, where the summation is taken over all $i\neq j$ and then standardized with

$$1-\sum_{i=1}^{n}\left\{\mathcal{K}\left(\frac{Y_i-x}{h}\right)\right\}^2.$$

The interval estimation of causality measures requires the estimation of the partial derivatives $C_x^{[1]}$ and $C_x^{[2]}$. In the upcoming simulations, one considers the finite difference estimator given by

$$\widehat{C}_x^{[1]}(u,v)=\sqrt{nh}\left\{C_{xh}^{Z\to Y}(u_h^\star,v)-C_{xh}^{Z\to Y}(u,v)\right\},$$

where $u_h^\star=u+\min\{1/\sqrt{nh},1-u\}$. The estimator $\widehat{C}_x^{[2]}$ is defined similarly. This particular choice fulfills the assumption stated in Proposition 4. The results that will be reported have been obtained using the triweight function $L(y) = 35(1-y^2)^3\,\mathbb{I}(|y| \leq 1)/32$, leading to the local linear kernel

$$\mathcal{K}(y) = L(y)\left(\frac{S_{n,2} - y\, S_{n,1}}{S_{n,0}\, S_{n,2} - S_{n,1}^2}\right),$$

where for $a \in \{h, h_1, h_2\}$,

$$S_{n,\ell} = \sum_{i=1}^{n} \left(\frac{Y_i - x}{a}\right)^{\ell} L\left(\frac{Y_i - x}{a}\right), \quad \ell \in \{0, 1, 2\}.$$

Negative weights are taken to be zero and the remaining weights are re-scaled in order that they sum to one. Using similar arguments as those in Li & Racine (2007), one can show that Assumptions W_1–W_{13} are satisfied whenever the alpha-mixing coefficients are of order $O(r^{-a})$ for $a > 6$. Other experiments with the Nadaraya–Watson kernel show very similar results, so they are not presented here. For the selection of the bandwidth parameters h, h_1 and h_2, several methods have been considered in the case $h = h_1 = h_2$, namely

(1) the plug-in bandwidth selection rule of Gijbels *et al.* (2011);

(2) the minimization of the estimated integrated squared error;

(3) setting $h = S(Y_1, \ldots, Y_n) \times n^{-1/5}$, where $S(\cdot)$ is either the variance, the inter-quartile range or the range between the 5th and 95th percentiles.

Based on many experiments, the third method based on the inter-quartile range performs better for the tests based on Kendall's functional, while for Spearman's functional, it is the third method using the 5th and 95th percentiles that is the best. Also, it is worth mentioning that the performance of Spearman's functional is more sensible to the choice of h compared to the Kendall functional, at least for the scenarios that were considered.

3.5.2 Accuracy of the local causality estimators

The performance of the local causality measures $\rho_{xh}^{Z\to Y}$, $\widetilde{\rho}_{xh}^{Z\to Y}$, $\tau_{xh}^{Z\to Y}$ and $\widetilde{\tau}_{xh}^{Z\to Y}$ will be investigated in the light of their bias and variance. To this end, a vector autoregressive model of order one has been considered. Specifically, time series have been simulated based on the stationary process

$$\begin{pmatrix} Y_t \\ Z_t \end{pmatrix} = \Sigma \begin{pmatrix} Y_{t-1} \\ Z_{t-1} \end{pmatrix} + \boldsymbol{\epsilon}_t,$$

where $\Sigma \in \mathbb{R}^{2\times 2}$ is such that $\Sigma_{11} = \Sigma_{22} = (\theta_2 - \theta_1\theta_3)/(1-\theta_3^2)$ and $\Sigma_{21} = \Sigma_{12} = (\theta_1 - \theta_2\theta_3)/(1-\theta_3^2)$ and $\boldsymbol{\epsilon}_t$ is distributed as a centered and symmetric bivariate Normal distribution with variance $\sigma_\epsilon^2 = 1 - \Sigma_{11}^2 - \Sigma_{21}^2 - 2\theta_3\Sigma_{11}\Sigma_{21}$ and correlation $\rho_\epsilon = \theta_3(1 - \Sigma_{11}^2 - \Sigma_{21}^2) - 2\Sigma_{11}\Sigma_{21}$. With this particular choice of parameters, $(Y_t, Z_t, Y_{t-1}, Z_{t-1})$ is centered Normal with covariance matrix

$$\Upsilon = \begin{bmatrix} 1 & \theta_3 & \theta_1 & \theta_2 \\ \theta_3 & 1 & \theta_2 & \theta_1 \\ \theta_1 & \theta_2 & 1 & \theta_3 \\ \theta_2 & \theta_1 & \theta_3 & 1 \end{bmatrix}.$$

In that case, the local causality is controlled by a Normal copula with parameter $\varrho = (\theta_1 - \theta_2\,\theta_3)/(\sqrt{1-\theta_2^2}\sqrt{1-\theta_3^2})$, *i.e.* $C_x^{\mathrm{Z}\to\mathrm{Y}}(u,v) = \Phi_\varrho\{\Phi^{-1}(u), \Phi^{-1}(v)\}$, where Φ_ϱ is the cdf of the bivariate standard Normal distribution with correlation $\varrho \in [-1,1]$. The results reported in Figure 3.1 and Figure 3.2 are based on 1 000 replicates of this model when $x = 1/2$.

First observe that $\widetilde{\rho}_{xh}^{Z\to Y}$ outperforms $\rho_{xh}^{Z\to Y}$ both in terms of bias and variance when $(\theta_1,\theta_2,\theta_3) \in \{(.4,-.25,.3),(-.4,.25,.3),(.2,.44,.44)\}$; their performance are however similar when $(\theta_1,\theta_2,\theta_3) = (.0,.3,.0)$. This might be explained by the influence of the conditional marginal distributions F_{1x} and F_{2x} on $\mathrm{E}(\rho_{xh}^{Z\to Y})$, which is quite low for the latter model. The same comment can be made about $\tau_{xh}^{Z\to Y}$ and $\widetilde{\tau}_{xh}^{Z\to Y}$. Finally, it is worth noting that the bias and variance of $\widetilde{\rho}_{xh}^{Z\to Y}$ and $\widetilde{\tau}_{xh}^{Z\to Y}$ stabilize as the bandwidth parame-

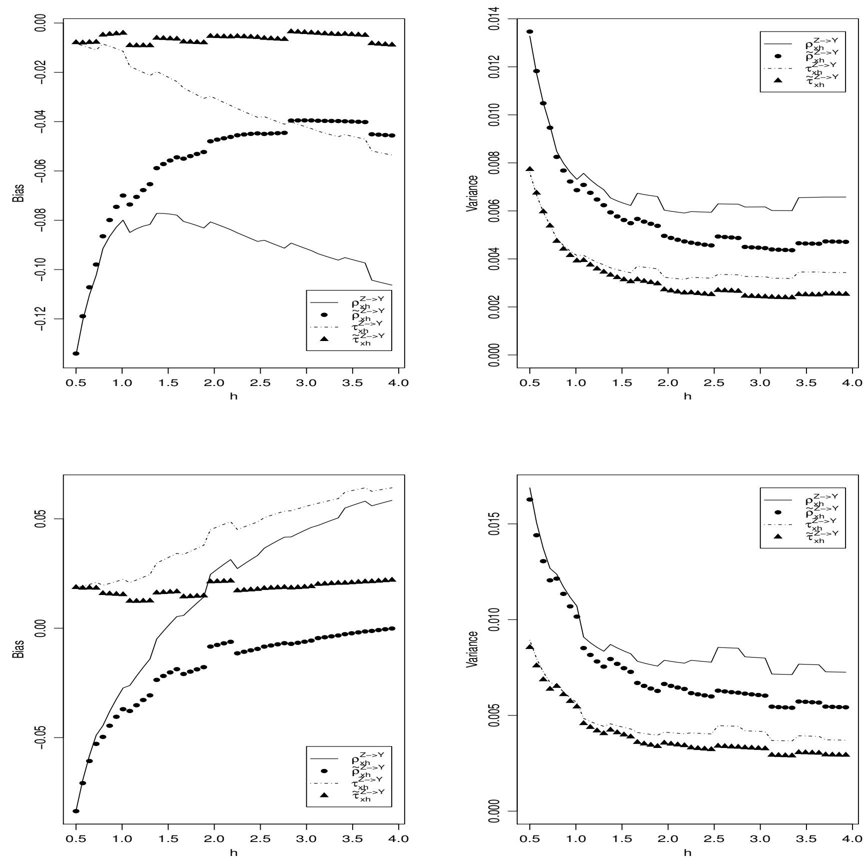


Figure 3.1 – Estimated Bias (left panels) and Variance (right panels) of $\rho_{xh}^{Z\to Y}$, $\widetilde{\rho}_{xh}^{Z\to Y}$, $\tau_{xh}^{Z\to Y}$ and $\widetilde{\tau}_{xh}^{Z\to Y}$ as a function of the bandwidth parameter h under Gaussian vector autoregressive processes when $n = 250$. Upper panels : $(\theta_1, \theta_2, \theta_3) = (.4, -.25, .3)$; bottom panels : $(\theta_1, \theta_2, \theta_3) = (-.4, .25, .3)$.

ter h increases. However, the biases of $\rho_{xh}^{Z\to Y}$ and $\tau_{xh}^{Z\to Y}$ are more sensible to the values of h.

3.5.3 Coverage probability of interval estimations

A general D-Vine structure for bivariate processes was recently suggested by Beare & Seo (2015). Specifically, let $C_1, \ldots, C_5$ be such that C_1 is the copula of (Y_t, Z_t),

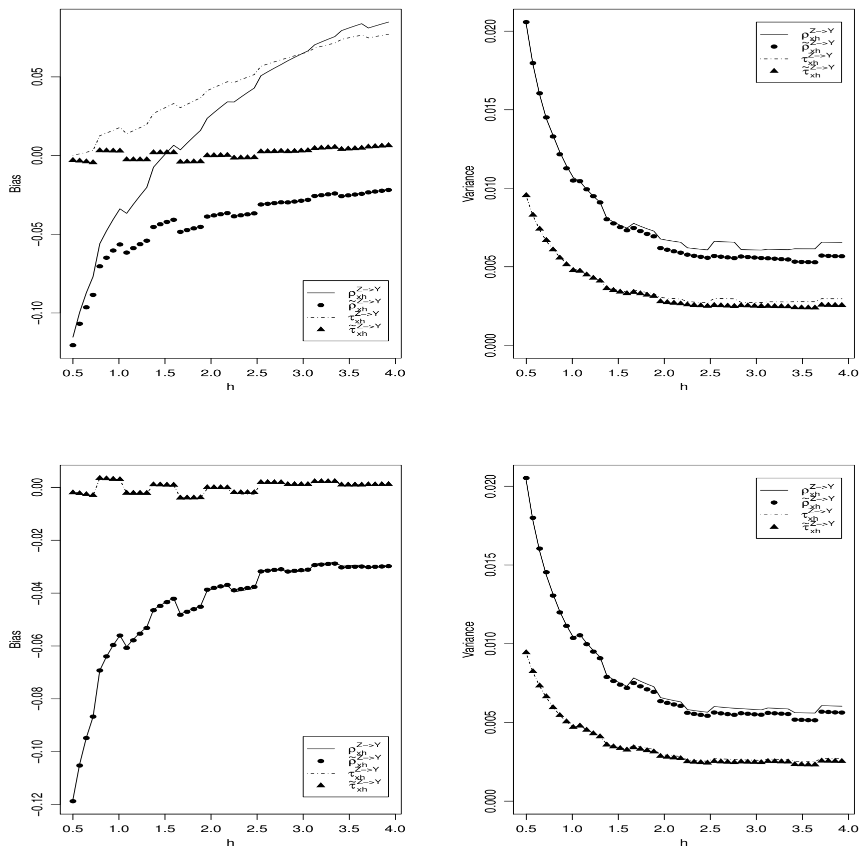


Figure 3.2 – Estimated Bias (left panels) and Variance (right panels) of $\rho_{xh}^{Z\to Y}$, $\widetilde{\rho}_{xh}^{Z\to Y}$, $\tau_{xh}^{Z\to Y}$ and $\widetilde{\tau}_{xh}^{Z\to Y}$ as a function of the bandwidth parameter h under Gaussian vector autoregressive processes when $n = 250$. Upper panels : $(\theta_1, \theta_2, \theta_3) = (.2, .44, .44)$; bottom panels : $(\theta_1, \theta_2, \theta_3) = (.0, .3, .0)$.

C_2 is the copula of (Y_{t-1}, Y_t), C_3 is the conditional copula of (Y_{t-1}, Z_t) given $Y_t = x$, C_4 is the conditional copula of (Y_t, Z_{t-1}) given $Y_{t-1} = x$ and C_5 is the conditional copula of (Z_t, Z_{t-1}) given $Y_{t-1} = x, Y_t = x'$. In this setup, C_4 plays the role of the local causality copula.

Here, the coverage probabilities of interval estimations based on $\rho_{xh}^{Z\to Y}$, $\widetilde{\rho}_{xh}^{Z\to Y}$, $\tau_{xh}^{Z\to Y}$ and $\widetilde{\tau}_{xh}^{Z\to Y}$ will be estimated in the case when C_4 is the Normal copula parameterized in such a way that $\tau_x^{Z\to Y} \in \{0, .1, .2, .3, .4\}$; this is done easily using the relationship $\varrho = \sin(\pi\, \tau_x^{Z\to Y}/2)$. The copulas C_1, C_2, C_3 and C_5 are also Normal and are parameterized in terms of their respective values τ_1, τ_2, τ_3 and τ_5 of Kendall's tau. The results on the coverage probability in the case $x = 0.5$ are to be found in Table 3.1.

Generally speaking, the coverage probabilities tend to be closer to their 95% nominal level as n increases. An exception occurs for $\rho_{xh}^{Z\to Y}$ when $(\tau_1, \tau_2, \tau_3, \tau_5) = (.5, .3, .75, .05)$. In that case, since the coverage probabilities of $\widetilde{\rho}_{xh}^{Z\to Y}$ behave well, this may be due to the influence of the conditional marginal distributions on the asymptotic bias. Finally note that the coverage probabilities are very similar for all levels of conditional dependence as measured by $\tau_x^{Z\to Y}$.

3.5.4 Power of the tests of local non-causality

Consider testing the null hypothesis $\mathbb{H}_0$ of the local non-causality from Z to Y at x, *i.e.* the conditional independence between Y_t and Z_{t-1} given $Y_{t-1} = x$. To this end, one considers again the D-Vine structure for $(Y_t, Z_t)_{t\in\mathbb{Z}}$ described in Subsection 3.5.3. Here, C_4 is taken to be either the Normal or the Clayton copula ; the latter is defined by $C_\theta^{\mathrm{CL}}(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$, $\theta > 0$. Both the Normal and the Clayton copulas are parameterized in such a way that $\tau_x^{Z\to Y} = (3/2)\{\Phi(x) - 1/2\}^2$. From well-known relationships between Kendall's tau and the parameters of the Normal and Clayton copulas, this corresponds to

$$\varrho = \sin\left[\frac{3\,\{\Phi(x) - 1/2\}^2}{2\pi}\right] \quad \text{and} \quad \theta = \frac{6\,\{\Phi(x) - 1/2\}^2}{2 - 3\,\{\Phi(x) - 1/2\}^2}\,.$$

The values of x are chosen in order that $\tau_x^{Z\to Y} \in \{0, .1, .25\}$. Here again, C_1, C_2, C_3 and C_5 are Normal copulas parameterized in terms of their respective values τ_1, τ_2, τ_3 and τ_5 of Kendall's tau ; the results in Table 3.2 concerns the case when C_4 is Normal, while Table 3.3 concerns the Clayton copula.

Looking at Tables 3.2–3.3, one can say that generally speaking, the four tests are good at maintaining their nominal level under the null hypothesis of non-causality. An exception occurs for the test based on $\rho_{xh}^{Z\to Y}$ when $(\tau_1, \tau_2, \tau_3, \tau_5) \in \{(.5, .3, .3, .05), (5, .3, .75, .05)\}$, when $C_x^{Z\to Y}$ is the Normal or the Clayton copula. It can be seen that the nominal levels in these situations are rather far from 5%. This might be due to the fact that the procedure neglects the bias. Since $\widetilde{\rho}_{xh}^{Z\to Y}$ performs well under these models, it is an indication that this bias is due to the conditional marginal distributions.

The ability of the tests to reject the null hypothesis under departures from non-causality is good and increases with the sample size, as expected. Interestingly, when $n = 250$, the two tests based on Kendall's tau are more powerful than those based on Spearman's rho, while the latter are better for larger sample sizes. It is especially true when $\tau_x^{Z\to Y} = .1$ and $n = 1\ 000$. Note however that these conclusions for $\rho_{xh}^{Z\to Y}$ might be influenced by the fact that the latter hardly keeps its nominal level when $(\tau_1, \tau_2, \tau_3, \tau_5) \in \{(.5, .3, .3, .05), (5, .3, .75, .05)\}$.

3.6 Illustration on financial data

The following illustration is based on the bivariate time series of the 1 512 daily observations taken between January 2010 and January 2016 for the compounded changes in prices (returns) and trading volume of the Standard and Poor's 500 (S&P500) Index. The relationship between these two indices has been extensively studied, both from a theoretical and from an empirical perspective. According to the tests of stationarity reported in Bouezmarni *et al.* (2012), one will work instead with the first difference in logarithmic returns (Y) and with the first difference in logarithmic volume (Z). As a consequence, the upcoming conclusions will have to be interpreted in terms of growth rates.

The causality from Z to Y is then investigated from the sample (Y_2, Z_1, Y_1), ..., $(Y_{1511}, Z_{1510}, Y_{1510})$. For these data, the value of the partial correlation coefficient of (Y_t, Z_{t-1}) given Y_{t-1} is -0.024×10^{-4}, leading to the conclusion of a global non-causality (p-value = 0.36). However, as mentioned in the introduction, such a conclusion can be misleading when the relationship between Y_t and Z_{t-1} changes according to the value taken by Y_{t-1}. This is exactly what happens here. For example, if one considers the sub-sample for which $Y_{t-1} > 0$, then the partial correlation coefficient is 0.072, which is significantly different from zero (p-value = 0.039). On the other hand, the subsample for which $Y_{t-1} < 0$ leads to a partial correlation coefficient of -0.095 (p-value = 0.01).

In order to take into account the levels of Y_{t-1}, a solution is to rely on local causality indices as introduced in Section 3. Figure 3.3 reports the values of $\rho_{xh}^{Z\to Y}$, $\widetilde{\rho}_{xh}^{Z\to Y}$, $\tau_{xh}^{Z\to Y}$ and $\widetilde{\tau}_{xh}^{Z\to Y}$ as a function of x, for x ranging between the 10^{th} and the 90^{th} percentile of the Y. In addition, the 95% point-wise confidence intervals as computed from the method in Section 3 are given. Clearly, the values taken by the local causality indices depend on x.

In Figure 3.4, the same curves are given, this time with the 95% point-wise critical values of the test of local non-causality are given. It can be seen that the curves based on $\rho_{xh}^{Z\to Y}$ and $\widetilde{\rho}_{xh}^{Z\to Y}$ are below the lower bound when Y_{t-1} is less than 0.002 while it exceeds the upper bound for $Y_{t-1} > 0.004$. This is in accordance with the conclusions of the tests of non-causality based on the partial correlation coefficient.

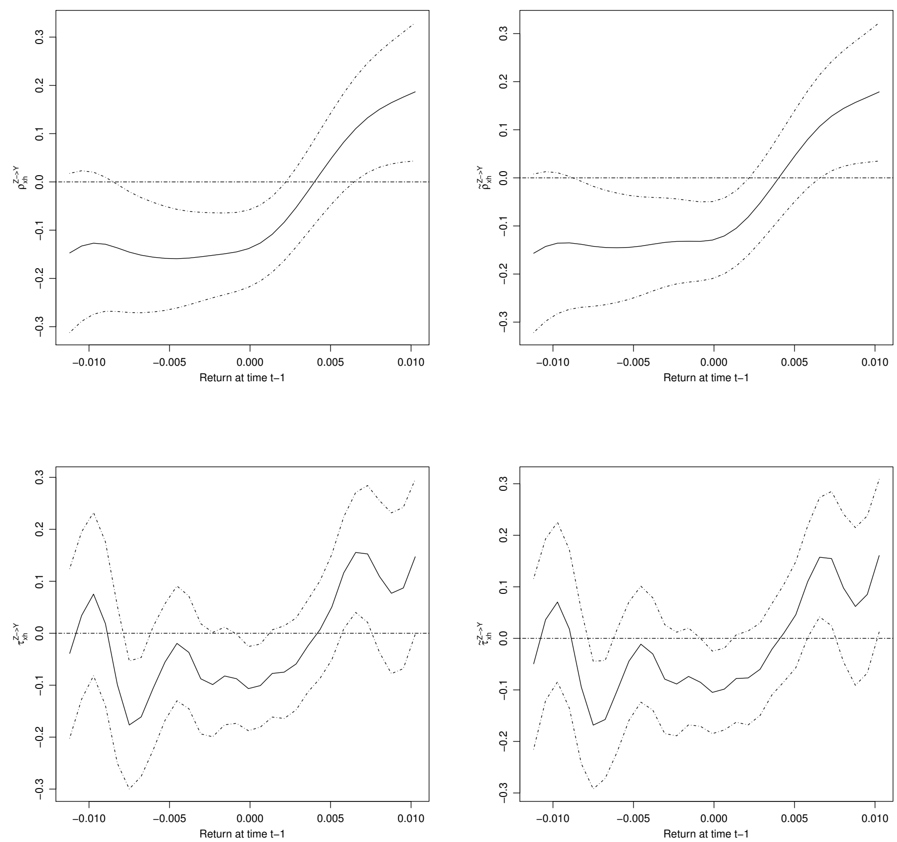


Figure 3.3 – Curves of $\rho_{xh}^{Z\to Y}$ (upper left), $\widetilde{\rho}_{xh}^{Z\to Y}$ (upper right), $\widetilde{\tau}_{xh}^{Z\to Y}$ (bottom left) and $\widetilde{\tau}_{xh}^{Y\to Z}$ (bottom right) as a function of x, together with 95% point-wise confidence bands, for the Standard and Poor's 500 bivariate time series

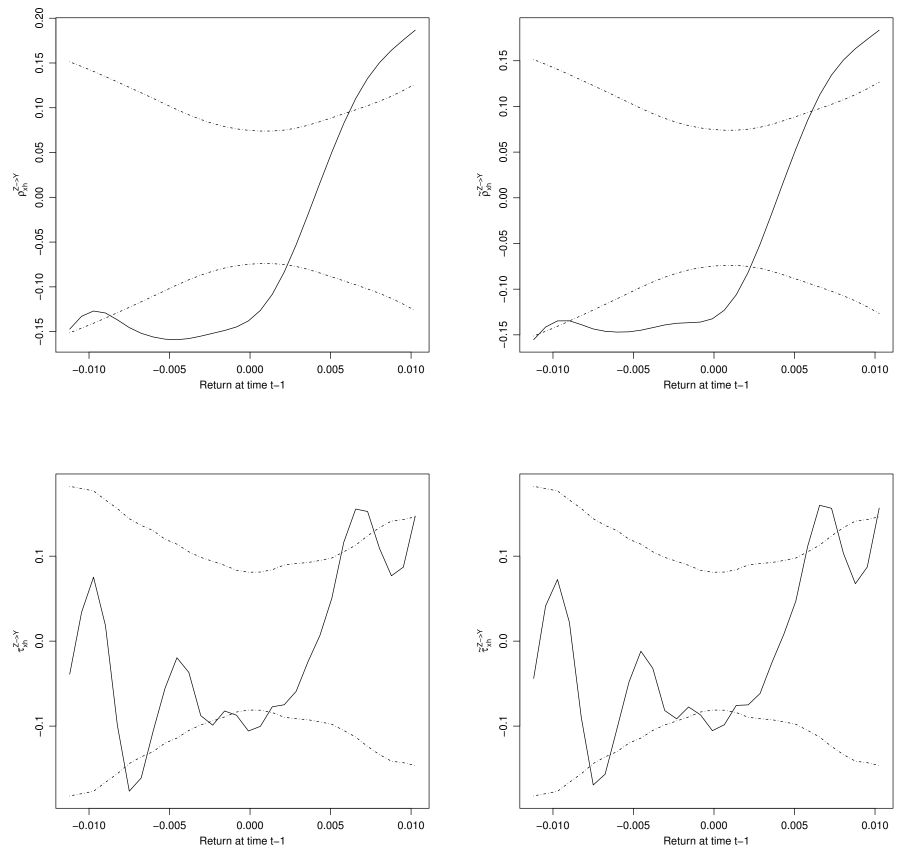


Figure 3.4 – Curves of $\rho_{xh}^{Z\to Y}$ (upper left), $\widetilde{\rho}_{xh}^{Z\to Y}$ (upper right), $\widetilde{\tau}_{xh}^{Z\to Y}$ (bottom left) and $\widetilde{\tau}_{xh}^{Y\to Z}$ (bottom right) as a function of x, together with the 95% point-wise critical values of the test of local non-causality, for the Standard and Poor's 500 bivariate time series

Tableau 3.1 – Coverage probabilities, as estimated from 1 000 replicates, of 95% confidence intervals for the local causality measures $\rho_x^{Z\to Y}$ and $\tau_x^{Z\to Y}$ under bivariate D-vine time series in which $C_x^{Z\to Y}$ is the Normal copula

$\tau_x^{Z\to Y}$	Estimator	$(\tau_1,\tau_2,\tau_3,\tau_5)=(.3,.1,.1,.1)$			$(\tau_1,\tau_2,\tau_3,\tau_5)=(.05,.05,.05,.05)$		
		$n=250$	$n=500$	$n=1\ 000$	$n=250$	$n=500$	$n=1\ 000$
0	$\rho_{xh}^{Z\to Y}$	90.6	94.0	95.2	89.6	92.7	93.5
	$\widetilde{\rho}_{xh}^{Z\to Y}$	88.7	92.3	93.2	88.9	92.3	93.9
	$\tau_{xh}^{Z\to Y}$	94.0	94.0	95.0	93.1	94.6	94.6
	$\widetilde{\tau}_{xh}^{Z\to Y}$	93.8	94.2	94.6	93.3	94.4	94.3
0.1	$\rho_{xh}^{Z\to Y}$	90.4	94.0	95.2	89.9	93.8	94.7
	$\widetilde{\rho}_{xh}^{Z\to Y}$	90.1	93.5	94.5	89.6	93.4	94.7
	$\tau_{xh}^{Z\to Y}$	92.9	93.8	94.4	93.0	94.1	94.9
	$\widetilde{\tau}_{xh}^{Z\to Y}$	92.3	93.9	94.5	92.1	94.1	94.9
0.2	$\rho_{xh}^{Z\to Y}$	90.5	93.6	94.5	90.1	93.4	93.9
	$\widetilde{\rho}_{xh}^{Z\to Y}$	89.6	93.2	93.9	90.0	93.0	93.6
	$\tau_{xh}^{Z\to Y}$	93.5	94.3	95.0	93.2	95.1	95.2
	$\widetilde{\tau}_{xh}^{Z\to Y}$	93.2	94.1	95.4	93.9	95.1	95.0
0.3	$\rho_{xh}^{Z\to Y}$	90.8	93.4	94.4	91.1	92.5	93.7
	$\widetilde{\rho}_{xh}^{Z\to Y}$	90.5	92.7	94.2	90.7	92.2	93.9
	$\tau_{xh}^{Z\to Y}$	93.6	93.7	94.4	93.9	93.6	93.9
	$\widetilde{\tau}_{xh}^{Z\to Y}$	94.3	93.7	94.0	93.9	93.1	93.8
0.4	$\rho_{xh}^{Z\to Y}$	89.4	92.7	93.9	90.1	92.0	94.2
	$\widetilde{\rho}_{xh}^{Z\to Y}$	89.8	92.0	94.0	90.2	92.1	93.9
	$\tau_{xh}^{Z\to Y}$	94.2	94.7	94.8	93.5	94.4	94.8
	$\widetilde{\tau}_{xh}^{Z\to Y}$	94.3	94.4	95.4	93.9	94.6	95.0

		$(\tau_1,\tau_2,\tau_3,\tau_5)=(.5,.3,.3,.05)$			$(\tau_1,\tau_2,\tau_3,\tau_5)=(5,.3,.75,.05)$		
		$n=250$	$n=500$	$n=1\ 000$	$n=250$	$n=500$	$n=1\ 000$
0	$\rho_{xh}^{Z\to Y}$	94.0	94.3	93.7	77.3	82.1	84.3
	$\widetilde{\rho}_{xh}^{Z\to Y}$	89.8	92.4	93.2	89.6	92.4	93.6
	$\tau_{xh}^{Z\to Y}$	93.5	94.3	94.5	93.6	94.7	95.4
	$\widetilde{\tau}_{xh}^{Z\to Y}$	93.7	94.2	94.1	93.3	93.7	95.4
0.1	$\rho_{xh}^{Z\to Y}$	94.7	93.7	93.2	75.2	81.1	83.2
	$\widetilde{\rho}_{xh}^{Z\to Y}$	90.6	91.4	94.0	89.3	92.2	93.9
	$\tau_{xh}^{Z\to Y}$	95.0	93.9	95.7	92.3	94.3	93.5
	$\widetilde{\tau}_{xh}^{Z\to Y}$	94.5	93.5	95.8	92.6	93.9	94.1
0.2	$\rho_{xh}^{Z\to Y}$	93.7	94.3	94.4	71.1	75.5	77.0
	$\widetilde{\rho}_{xh}^{Z\to Y}$	89.8	92.6	92.8	90.8	92.6	93.7
	$\tau_{xh}^{Z\to Y}$	93.7	92.7	94.2	93.6	92.9	94.3
	$\widetilde{\tau}_{xh}^{Z\to Y}$	93.6	93.0	93.9	92.8	92.4	94.5
0.3	$\rho_{xh}^{Z\to Y}$	93.1	93.9	94.6	65.8	67.2	68.8
	$\widetilde{\rho}_{xh}^{Z\to Y}$	90.0	92.2	93.9	90.2	92.7	94.2
	$\tau_{xh}^{Z\to Y}$	91.6	93.4	94.3	93.7	92.7	93.5
	$\widetilde{\tau}_{xh}^{Z\to Y}$	91.5	94.3	94.8	93.3	92.1	92.7
0.4	$\rho_{xh}^{Z\to Y}$	91.6	94.1	95.1	50.3	54.8	55.2
	$\widetilde{\rho}_{xh}^{Z\to Y}$	88.6	91.9	93.5	90.3	91.8	93.3
	$\tau_{xh}^{Z\to Y}$	92.4	93.8	93.0	94.4	94.2	94.9
	$\widetilde{\tau}_{xh}^{Z\to Y}$	92.0	93.9	93.7	93.0	92.9	93.4

Tableau 3.2 – Percentages of rejection of the null hypothesis of local non-causality, as estimated from 1 000 replicates, for the tests at level $\alpha = 0.05$ based on $\rho_{xh}^{Z\to Y}$, $\widetilde{\rho}_{xh}^{Z\to Y}$, $\tau_{xh}^{Z\to Y}$ and $\widetilde{\tau}_{xh}^{Z\to Y}$ under bivariate D-vine time series in which $C_x^{Z\to Y}$ is the Normal copula

		$(\tau_1,\tau_2,\tau_3,\tau_5) = (.3,.1,.1,.1)$			$(\tau_1,\tau_2,\tau_3,\tau_5) = (.05,.05,.05,.05)$		
$\tau_x^{Z\to Y}$	Estimator	$n = 250$	$n = 500$	$n = 1\,000$	$n = 250$	$n = 500$	$n = 1\,000$
0	$\rho_{xh}^{Z\to Y}$	5.5	6.4	4.6	5.7	5.2	4.0
	$\widetilde{\rho}_{xh}^{Z\to Y}$	6.0	5.8	4.3	6.1	5.0	4.4
	$\tau_{xh}^{Z\to Y}$	4.3	4.9	4.4	5.0	4.7	5.0
	$\widetilde{\tau}_{xh}^{Z\to Y}$	5.3	5.2	4.9	4.8	4.9	5.2
0.1	$\rho_{xh}^{Z\to Y}$	17.9	36.8	69.1	16.1	35.2	65.9
	$\widetilde{\rho}_{xh}^{Z\to Y}$	17.0	35.9	65.8	16.8	34.5	65.5
	$\tau_{xh}^{Z\to Y}$	20.2	25.2	41.3	19.7	26.6	41.9
	$\widetilde{\tau}_{xh}^{Z\to Y}$	18.8	25.2	41.2	18.9	26.0	41.1
0.25	$\rho_{xh}^{Z\to Y}$	27.9	71.8	98.5	27.0	72.6	98.0
	$\widetilde{\rho}_{xh}^{Z\to Y}$	26.2	71.0	98.4	25.0	70.9	98.1
	$\tau_{xh}^{Z\to Y}$	40.3	63.0	86.5	39.3	64.6	87.2
	$\widetilde{\tau}_{xh}^{Z\to Y}$	38.9	62.4	85.2	37.8	61.5	86.1

		$(\tau_1,\tau_2,\tau_3,\tau_5) = (.5,.3,.3,.05)$			$(\tau_1,\tau_2,\tau_3,\tau_5) = (.5,.3,.75,.05)$		
		$n = 250$	$n = 500$	$n = 1\,000$	$n = 250$	$n = 500$	$n = 1\,000$
0	$\rho_{xh}^{Z\to Y}$	5.4	8.3	10.1	12.5	9.8	8.2
	$\widetilde{\rho}_{xh}^{Z\to Y}$	5.7	4.2	4.6	5.3	6.0	5.0
	$\tau_{xh}^{Z\to Y}$	5.0	6.6	5.6	5.4	6.5	5.4
	$\widetilde{\tau}_{xh}^{Z\to Y}$	5.2	7.0	6.7	5.7	5.5	4.8
0.1	$\rho_{xh}^{Z\to Y}$	26.6	49.8	81.9	4.5	14.2	38.6
	$\widetilde{\rho}_{xh}^{Z\to Y}$	14.3	33.6	66.7	15.9	35.4	67.7
	$\tau_{xh}^{Z\to Y}$	20.1	28.4	41.7	17.9	22.9	39.8
	$\widetilde{\tau}_{xh}^{Z\to Y}$	19.0	27.3	40.3	20.6	25.9	43.8
0.25	$\rho_{xh}^{Z\to Y}$	34.0	77.5	98.2	15.4	51.6	92.9
	$\widetilde{\rho}_{xh}^{Z\to Y}$	26.1	69.9	96.7	29.1	70.2	98.3
	$\tau_{xh}^{Z\to Y}$	41.3	61.8	85.0	38.3	60.6	85.0
	$\widetilde{\tau}_{xh}^{Z\to Y}$	38.2	61.4	83.5	39.2	63.2	86.7

Tableau 3.3 – Percentages of rejection of the null hypothesis of local non-causality, as estimated from 1 000 replicates, for the tests at level $\alpha = 0.05$ based on $\rho_{xh}^{Z\to Y}$, $\widetilde{\rho}_{xh}^{Z\to Y}$, $\tau_{xh}^{Z\to Y}$ and $\widetilde{\tau}_{xh}^{Z\to Y}$ under bivariate D-vine time series in which $C_x^{Z\to Y}$ is the Clayton copula

		$(\tau_1,\tau_2,\tau_3,\tau_5) = (.3, .1, .1, .1)$			$(\tau_1,\tau_2,\tau_3,\tau_5) = (.05, .05, .05, .05)$		
$\tau_x^{Z\to Y}$	Estimator	$n = 250$	$n = 500$	$n = 1\ 000$	$n = 250$	$n = 500$	$n = 1\ 000$
0	$\rho_{xh}^{Z\to Y}$	5.3	5.7	4.3	5.4	5.0	4.0
	$\widetilde{\rho}_{xh}^{Z\to Y}$	5.6	5.3	3.9	6.0	4.9	4.0
	$\tau_{xh}^{Z\to Y}$	4.6	5.0	4.6	4.4	5.3	4.6
	$\widetilde{\tau}_{xh}^{Z\to Y}$	5.3	5.9	4.9	4.7	5.1	4.9
0.1	$\rho_{xh}^{Z\to Y}$	18.5	37.9	72.1	16.3	35.1	66.5
	$\widetilde{\rho}_{xh}^{Z\to Y}$	16.5	34.6	68.3	16.1	34.2	65.7
	$\tau_{xh}^{Z\to Y}$	20.6	25.2	41.6	21.5	25.3	40.3
	$\widetilde{\tau}_{xh}^{Z\to Y}$	20.7	25.7	41.2	20.3	26.4	39.5
0.25	$\rho_{xh}^{Z\to Y}$	28.4	73.4	98.1	26.6	73.9	97.6
	$\widetilde{\rho}_{xh}^{Z\to Y}$	25.5	71.6	97.8	25.6	72.6	97.8
	$\tau_{xh}^{Z\to Y}$	41.1	62.9	84.8	40.4	62.8	87.0
	$\widetilde{\tau}_{xh}^{Z\to Y}$	40.0	61.7	84.4	40.4	62.6	86.7
		$(\tau_1,\tau_2,\tau_3,\tau_5) = (.5, .3, .3, .05)$			$(\tau_1,\tau_2,\tau_3,\tau_5) = (.5, .3, .75, .05)$		
		$n = 250$	$n = 500$	$n = 1\ 000$	$n = 250$	$n = 500$	$n = 1\ 000$
0	$\rho_{xh}^{Z\to Y}$	5.5	8.5	10.3	12.9	10.3	8.5
	$\widetilde{\rho}_{xh}^{Z\to Y}$	5.9	4.3	4.9	4.8	6.3	4.4
	$\tau_{xh}^{Z\to Y}$	5.0	6.3	5.0	5.4	6.7	5.2
	$\widetilde{\tau}_{xh}^{Z\to Y}$	4.7	5.9	4.8	5.5	6.6	5.5
0.1	$\rho_{xh}^{Z\to Y}$	27.5	53.0	83.6	5.8	14.5	38.0
	$\widetilde{\rho}_{xh}^{Z\to Y}$	14.7	35.4	66.4	16.3	34.4	67.0
	$\tau_{xh}^{Z\to Y}$	21.9	29.7	43.8	19.3	21.9	41.3
	$\widetilde{\tau}_{xh}^{Z\to Y}$	19.3	26.7	41.4	21.8	24.4	42.7
0.25	$\rho_{xh}^{Z\to Y}$	34.2	78.4	98.5	14.6	52.4	90.9
	$\widetilde{\rho}_{xh}^{Z\to Y}$	25.3	69.1	97.3	27.5	70.6	97.4
	$\tau_{xh}^{Z\to Y}$	40.6	64.0	86.6	36.0	58.6	84.3
	$\widetilde{\tau}_{xh}^{Z\to Y}$	39.1	62.4	84.7	38.7	59.3	85.7

CHAPITRE 4

Estimation of a conditional copula when a variable is subject to random right censoring

Résumé

Dans cet article, nous étudions la structure de dépendance régissant le comportement aléatoire du vecteur (Y_1, Y_2) étant donné une co-variable X lorsqu'une variable est censurée à droite. Cette structure de dépendance est décrite par une copule conditionnelle, définie comme étant la fonction $C_x : [0,1]^2 \to [0,1]$ satisfaisant $\mathrm{P}(Y_1 \leq y_1, Y_2 \leq y_2) = C_x\{\mathrm{P}(Y_1 \leq y_1 | X = x), \mathrm{P}(Y_2 \leq y_2 | X = x)\}$. Dans cet article, nous proposons une façon d'estimer C_x lorsque la variable Y_1 est censurée. Nous étudions ensuite le comportement asymptotique de l'estimateur proposé. Nous conduisons aussi quelques simulations numériques afin examiner les propriétés à taille d'échantillon finie de l'estimateur proposé. Enfin, l'utilité de la méthode est illustrée à travers l'étude d'un exemple traitant de patients opérés en raison d'un mélanome malin.

Abstract

This paper is concern with studying the dependence structure of a random pair (Y_1, Y_2) conditionally upon a covariate X when a variable, say Y_1, is subject to random right censoring. The dependence structure is described by a conditional copula, defined as the function $C_x : [0,1]^2 \to [0,1]$ such that $\mathrm{P}(Y_1 \le y_1, Y_2 \le y_2) = C_x\{\mathrm{P}(Y_1 \le y_1 | X = x), \mathrm{P}(Y_2 \le y_2 | X = x)\}$. In this paper, we propose a procedure to estimate the conditional copula when the variable Y_1 is censored. We establish the asymptotic properties of the proposed estimator. Its finite sample behavior is then investigate in a numerical study. The methodology is illustrated through a real data example featuring patients with malignant melanoma.

keyword Survival analysis, conditional copula, functional delta method, Kendall's tau, Spearman's rho, weak convergence

4.1 Introduction

Copulas have become a popular tool to model dependence. Recently, many works in this field have been concerned with capturing the influence of a covariate $X \in \mathbb{R}$ on the dependance structure of a vector interest $(Y_1, Y_2) \in \mathbb{R}$. An example is given in Gijbels *et al.* (2011), where a copula function is used to illustrate how the relationship between the life expectancy of men (Y_1) and women (Y_2) varies with the gross domestic product (X). To describe this copula function, consider the conditional joint distribution of (Y_1, Y_2) given $X = x$, for a real number x, given by

$$F_x(y_1, y_2) = \mathrm{P}\left(Y_1 \le y_1, Y_2 \le y_2 | X = x\right).$$

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

$$\mathbb{C}_x(u_1, u_2) = F_x\left\{F_{1x}^{-1}(u_1), F_{2x}^{-1}(u_2)\right\}. \tag{4.1}$$

The bivariate function $\mathbb{C}_x$ is called the *conditional copula* and 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 $\mathbb{C}_x$.

The topic of modeling and estimating conditional copula models have recently gained momentum, since the pioneering work of Patton (2006). For example, models specifying a functional connection between the covariate and a parametric copula were studied in Jondeau & Rockinger (2006) and Patton (2006). A nonparametric estimation procedure for this functional connection was proposed in Acar *et al.* (2011) while Abegaz *et al.* (2012) have considered an extension of this method to the case of unknown conditional marginal distributions. Assuming the availability of an i.i.d sample, a nonparametric approach has been investigated in Veraverbeke *et al.* (2011) and Gijbels *et al.* (2011), and a bootstrap method suitable for this estimation procedure was developed in Omelka *et al.* (2013).

However, all of the previously-mentioned estimation strategies rely on the full knowledge of the random variables $(Y_{11}, Y_{21}, X_1), \ldots, (Y_{1n}, Y_{2n}, X_n)$ and therefore reveal unsatisfactory when the data are incomplete. Amongst others, the right censoring scheme is a source of incompleteness that frequently appears in medical studies and clinical trials. On this matter, this occurs in a dataset, to be detailed later, featuring patients with malignant melanoma followed after their surgery for skin tumor.

The purpose of this work is to propose a proper methodology designed to estimate the conditional copula when a variable is subject to random right censoring. In that case, the true event time is not recorded for one component but instead a smaller time is observed. To be more specific, assume Y_2 is completely observed, and that the observed random variables are $(T_i, Y_{2i}, X_i, \delta_i)$, where

$$T_i = \min(Y_{1i}, C_i) \quad \delta_i = \mathbb{I}\{Y_{1i} \leq C_i\},\ i = 1, ..., n,$$

and $C_1, \ldots, C_n$ are independent and non-negative censoring variables. Hereafter, the conditional distribution function of C_i given $X = x_i$ will be denoted $G_{x_i}(t) = \mathrm{P}\{C_i \leq t \mid X = x_i\}$. Assuming that (Y_{1i}, Y_{2i}) are independent from C_i conditional upon x_i implies that $F_{2x_i} - H_{x_i} = (F_{2x_i} - F_{x_i})(1 - G_{x_i})$, where $H_{x_i}(t, y) = \mathrm{P}\{T_i \leq t, Y_{2i} \leq y\}$ is the distribution function of the survival times with marginals H_{1x_i} and H_{2x_i}.

This paper is organized as follows. In Section 4.2, we propose an estimator for the conditional copula in presence of censoring. This estimator relies on a nonparametric estimator for the joint conditional distribution, which is also an original contribution of this paper. In Section 3, we investigate the asymptotic propreties of these estimators by providing an asymptotic i.i.d. representation for the conditional distribution estimator, and by identifying the weak limit of properly re-scaled version of these estimators. A simulation study showing the performance of the conditional copula estimation procedure is presented in Section 4.4. In Section 4.5, we apply this methodology to the melanoma dataset to illustrate the influence of tumor thickness on the relationship between the survival time and the age of a patient after surgery. All the assumptions and conditions required for the theoretical validity of the results presented in Section 4.3 are provided in Section 4.6. The proofs are given in the Appendix.

4.2 An inverse-conditional-probability-of-censoring estimator for C_x

In the sequel, a fixed design where (Y_1, Y_2, C, X) is realized at $X \in \{x_1, \ldots, x_n\}$ is assumed. To be more specific, a random sample of independent and identically distributed vectors $(Y_{11}, Y_{21}, C_1, x_1), \ldots, (Y_{1n}, Y_{2n}, C_n, x_n)$ is observed. The theoretical developments presented in this work are also valid under random designs, *i.e.* when X is a random variable. The results travel from a fixed design to a random design by simply replacing x_i by X_i and using O_{P} and o_{P} instead of O and o arguments in the assumptions stated in the appendix.

As previously mentioned, the estimation of $\mathbb{C}_x$ from i.i.d. observations has been considered by Veraverbeke *et al.* (2011) and Gijbels *et al.* (2011). Specifically, from independent (and fully observed) random variables $(Y_{11}, Y_{21}, x_1), \ldots, (Y_{1n}, Y_{2n}, x_n)$, consider the estimator of the joint conditional distribution F_x given by

$$F_{xh}(y_1, y_2) = \sum_{i=1}^{n} w_{ni}(x, h) \, \mathbb{I}\left(Y_{1i} \leq y_1, Y_{2i} \leq y_2\right),$$

where $w_{n1}(x, h), \ldots, w_{nn}(x, h)$ are non-negative kernel-based weight functions that smooth the covariate space and $h = h_n$ is a bandwidth parameter that typically depends on the sample size. Popular choices for theses functions are the Nadaraya-Watson and Local-Linear weights given respectively by

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

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 conditional empirical marginal distributions extracted from F_{xh} are simply

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

From representation (4.1), a natural plug-in estimator of $\mathbb{C}_x$ is given by

$$\begin{aligned} \mathbb{C}_{xh}(u_1,u_2) &= F_{xh}\left\{F_{1xh}^{-1}(u_1), F_{2xh}^{-1}(u_2)\right\} \\ &= \sum_{i=1}^{n} w_{ni}(x,h)\,\mathbb{I}\left\{Y_{1i} \le F_{1xh}^{-1}(u_1), Y_{2i} \le F_{2xh}^{-1}(u_2)\right\}, \end{aligned} \tag{4.2}$$

where for $j=1,2$, $F_{jxh}^{-1}(u) = \inf\{y \in \mathbb{R} : F_{jxh}(y) \ge u\}$ is the left-continuous generalized inverse of F_{jxh}.

The goal of this section is to propose an estimator for the conditional copula in order to take into account the presence of censoring on the variable Y_1. To do this, we need an estimator for the conditional distribution function F_x. In the unconditional context (i.e without a covariate), the nonparametric estimation of the bivariate distribution of (Y_1, Y_2) in presence of censoring have been studied by many authors, see for example Dabrowska (1988), Akritas (1994) and Akritas & Keilegom (2003). However, to the best of our knowledge, the nonparametric estimation of F_x have never been addressed and hence is an original contribution of the present paper.

To built our estimator for F_x, we use a similar idea as the one originally exposed in Robins & Rotnitzky (1992). To compensate for the presence of censoring, each uncensored observation receives an extra weight equal to its inverse probability of failiure. This idea is motivated by the fact that

$$\mathrm{E}\left\{\mathbb{I}(T \le t, Y_2 \le y)\frac{\delta}{1-G_x(T-)} \mid X = x\right\} = F_x(t,y).$$

Hence, if G_x is known, one could estimate F_x with

$$\sum_{i=1}^{n} w_{ni}(x,h)\mathbb{I}(T_i \le t, Y_2 \le y)\frac{\delta_i}{1-G_x(T_i-)}. \tag{4.3}$$

As G_x is unknown, we simply replace it with the conditional Kaplan-Meier estimator for the censoring variable C. This estimator is given by :

$$G_{xg}(t) = 1 - \prod_{T_{(i)} \leq t} \left\{ 1 - \frac{w_{n[i]}(x, g)}{1 - \sum_{k=1}^{i-1} w_{n[k]}(x, g)} \right\}^{1-\delta_{[i]}},$$

where $T_{(1)} \leq \ldots \leq T_{(n)}$ are the ordered T_i, and $\delta_{[i]}$ and $w_{n[i]}(x, h)$ are respectively the corresponding δ_i and $w_{ni}(x, h)$. Here, $g = g_n$ is an auxiliary bandwidth parameter that may differ from h. The resulting estimator for the conditional distribution function F_x is then given by

$$F_{xh}^{(\mathrm{rc})}(t, y) = \sum_{i=1}^{n} \mathbb{I}\{T_i \leq t, Y_{2i} \leq y\} \frac{w_{ni}(x, h)\delta_i}{1 - G_{xg}(T_i-)}.$$

Although $F_{xh}^{(\mathrm{rc})}$ also depends on the bandwidth g, the latter is omitted for notational simplicity. This estimator can be seen as a conditional bivariate analogue to the inverse-probability-of-censoring estimator proposed in Robins & Rotnitzky (1992). Notice that when no censoring occurs, $F_{xh}^{(\mathrm{rc})}(t, y)$ is equal to F_{xh}.

The marginal distributions of $F_{xh}^{(\mathrm{rc})}$ are simply

$$F_{1xh}^{(\mathrm{rc})}(t) = \lim_{y \to \infty} F_{xh}^{(\mathrm{rc})}(t, y) \quad \text{and} \quad F_{2xh}^{(\mathrm{rc})}(t) = \lim_{y \to \infty} F_{xh}^{(\mathrm{rc})}(t, y).$$

When $g = h$, we can show that $F_{1xh}^{(\mathrm{rc})}(t)$ coincides with the conditional Kaplan-Meier estimator for the survival time Y_1 introduced in Beran (1981) (see D.4).

Following Equation (4.1), a plug-in estimator for the conditional copula could be define as

$$F_{xh}^{(\mathrm{rc})}(u, v) = F_{xh}^{(\mathrm{rc})} \left\{ F_{1xh}^{(\mathrm{rc})-1}(u), F_{2xh}^{(\mathrm{rc})-1}(v) \right\}$$

However, this estimator does not properly take advantage of the fact that Y_2 is completely observed. Instead of using $F_{2xh}^{(\mathrm{rc})}$, consider the estimator

$$F_{2xh}(y) = \sum_{i=1}^{n} \mathbb{I}\left(Y_{2i} \leq y\right) w_{ni}(x, h).$$

Note that this expression does not, in general, coincide with $F_{2xh}^{(\mathrm{rc})}$. Nevertheless, F_{2xh} uses all of the available knowledge related to F_{2x}. Then, one proceeds to the estimation of $\mathbb{C}_x$ with

$$\begin{aligned}\mathbb{C}_{xh}^{(\mathrm{rc})}(u,v) &= F_{xh}^{(\mathrm{rc})}\{F_{1xh}^{(\mathrm{rc})-1}(u), F_{2xh}^{-1}(v)\} \\ &= \sum_{i=1}^{n} \mathbb{I}\{T_i \le F_{1xh}^{(\mathrm{rc})-1}(u), Y_{2i} \le F_{2xh}^{-1}(v)\} \frac{w_{ni}(x,h)\delta_i}{1 - G_{xg}(T_i-)}.\end{aligned}$$

When all the survival times are observed, it follows that $\mathbb{C}_{xh}^{(\mathrm{rc})}$ is equal to $\mathbb{C}_{xh}$. Moreover, upon setting all the weight functions $w_{ni}(x,\cdot)$ equal to n^{-1}, we retrieve a very similar estimator as the one proposed in Gribkova & Lopez (2015) to estimate the (unconditional) copula. The only difference is in the estimation of the second marginal distribution.

4.3 Main theoretical results

The aim of this section is to investigate the large sample behaviour of the processes

$$\mathbb{F}_{xh}^{(\mathrm{rc})} = \sqrt{nh}(F_{xh}^{(\mathrm{rc})} - F_x) \quad \text{and} \quad \mathbb{B}_{xh}^{(\mathrm{rc})} = \sqrt{nh}(\mathbb{C}_{xh}^{(\mathrm{rc})} - C_x).$$

To this end, we first provide in Theorem 1 an asymptotic i.i.d representation for $\mathbb{F}_{xh}^{(\mathrm{rc})}$. Then, based on this representation, we obtain a weak convergence result for $\mathbb{F}_{xh}^{(\mathrm{rc})}$ in Corollary 1. Finally, the weak limit of $\mathbb{F}_{xh}^{(\mathrm{rc})}$ allows us to establish in Proposition 1 the weak convergence of $\mathbb{B}_{xh}^{(\mathrm{rc})}$.

4.3.1 Asymptotic i.i.d representation for $\mathbb{F}_{xh}^{(\mathbf{rc})}$

For any distribution function L, let τ_L be the right endpoint of its support, i.e $\inf\{t : L(t) = 1\}$, and write $\overline{\tau}_x = \min\{\tau_{F_{1x}}, \tau_{G_x}\}$. It is a well known problem in life time analysis that the tail support of the distribution of a random variable may not be identifiable due to right censoring (see Stute (1994)). This occurs when the support of the censoring variable is included in the support of the variable of interest, i.e when $\tau_{G_x} < \tau_{F_x}$. As a consequence, we cannot hope to infer on the conditional distribution beyond τ_x. Nevertheless, we next establish the asymptotic behavior of $\mathbb{F}_{xh}^{(\mathrm{rc})}$ over any closed subset included in $[0, \tau_x[\times\mathbb{R}$.

Hereafter, the sub-distribution function of the uncensored observations will be denoted by $H_{x_i}^u(t,y) = \mathrm{P}(T_i \le t, Y_{2i} \le y, \delta_i = 1 \mid X = x_i)$. Also, we write $H_{1x_i}(t) = \lim_{y\to\infty} H_{x_i}(t,y)$, $H_{1x_i}^u(t) = \lim_{y\to\infty} H_{x_i}^u(t,y)$ and $H_{1x_i}^c(t) = \mathrm{P}(T_i \le t, \delta_i = 0 \mid X = x_i)$.

To identify the asymptotic i.i.d representation for $\mathbb{F}_{xh}^{(\mathrm{rc})}$, we introduce the following random functions

$$\mathcal{J}_{ix}^{(1)}(t,y) = \frac{\mathbb{I}\left(T_i \le t, Y_{2i} \le y, \delta_i = 1\right)}{1 - G_x(T_i)} - F_x(t,y)$$

and

$$\begin{aligned}\mathcal{J}_{ix}^{(2)}(t,y) = &\int_0^t \frac{\mathbb{I}\left(T_i \le v\right) - H_{1x}(v)}{\{1 - H_{1x}(v)\}^2} \upsilon_x(t,y,v)\mathrm{d}H_{1x}^c(v) \\ &+ \int_0^t \frac{\mathbb{I}\left(T_i \le v, \delta_i = 0\right) - H_{1x}^c(v)}{\{1 - H_{1x}(v)\}\{1 - G_x(v)\}} h_x^u(v,y)\mathrm{d}v \\ &- \int_0^t \frac{\mathbb{I}\left(T_i \le v, \delta = 0\right) - H_{1x}^c(v)}{\{1 - H_{1x}(v)\}^2} \upsilon_x(t,y,v)\mathrm{d}H_{1x}(v),\end{aligned}$$

where $h_x^u(v,y) = \frac{\partial}{\partial v} H_x^u(v,y)$ and $\upsilon_x(t,y,v) = \int_v^t \frac{h_x^u(z,y)\mathrm{d}z}{1-G_x(z)} = F_x(t,y) - F_x(v,y)$. The assumptions required in the next theorem can be found in Section 4.6.

Theorem 1 *Suppose that* $\frac{nh^5}{\log(n)} = O(1)$, $\max(g,h) \to 0$, $\frac{ng^5}{\log(n)} = O(1)$ *and* $\frac{h}{g} = O(1)$. *Assume Conditions* W_1*–*W_5 *are satisfied, and suppose that Assumptions* $(\mathcal{C}_1)$ *to* $(\mathcal{C}_6)$ *are fulfilled for* F_x,H_x,H_x^u, H_{1x}^c *and* G_x. *For any* $0 < \mathfrak{t} < \overline{\tau}_x$, *write* $\mathcal{T}_{\mathfrak{t}} = [0, \mathfrak{t}] \times \mathbb{R}$. *Then, uniformly in* $(t,y) \in \mathcal{T}_{\mathfrak{t}}$, *we have*

$$\mathbb{F}_{xh}^{(rc)} = \sqrt{nh} \sum_{i=1}^n \left\{ w_{ni}(x,h)\mathcal{J}_{ix}^{(1)} + w_{ni}(x,g)\mathcal{J}_{ix}^{(2)} \right\} + o_{\mathrm{P}}(1).$$

In some way, Theorem 1 decomposes the random function $\mathbb{F}_{xh}^{(\mathrm{rc})}$ into two components. The first component can be associated to the estimation of a conditional distribution function provided that the conditional probability of censoring is known. In other words, it appears as a mildly modified and properly re-scaled version of the random function presented in Equation (4.3). From this perspective, the second component appears as a consequence of estimating the conditional probability of censoring.

We note that when no censoring occurs, then $\mathfrak{J}_{ix}^{(2)}(t,y) = 0$ and $\mathfrak{J}_{ix}^{(1)}(t,y) = \mathbb{I}(Y_{1i} \leq t, Y_{2i} \leq y) - F_x(t,y)$. Therefore,

$$\begin{aligned} F_{xh}^{(rc)}(t,y) - F_x(t,y) &= \sum_{i=1}^{n} w_{ni}(x,h)\mathfrak{J}_{ix}^{(1)}(t,y) \\ &= F_{xh}(t,y) - F_x(t,y). \end{aligned}$$

Next, as y goes to infinity, one has $F_{xh}^{(rc)}(t,y) - F_x(t,y) = F_{1xh}^{(rc)}(t,y) - F_{1x}(t,y)$. Also, when $g = h$, $F_{1xh}^{(rc)}$ is equal to the conditional Kaplan-Meier estimator introduced in Beran (1981). Hence, the random function $\sqrt{nh}\{F_{1xh}^{(rc)} - F_{1x}\}$ reduces to the conditional Kaplan-Meier process studied in Van Keilegom & Veraverbeke (1997). As expected, it is shown, in 2, that

$$\begin{aligned} &\lim_{y\to\infty} \left\{ \sum_{i=1}^{n} w_{ni}(x,h)\mathfrak{J}_{ix}^{(1)}(t,y) + \mathfrak{J}_{ix}^{(2)}(t,y) \right\} \\ &= \sum_{i=1}^{n} w_{ni}(x,h)\{1 - F_{1x}(t)\} \left[\int_0^t \frac{\mathbb{I}(T_i \leq v) - H_{1x}(v)}{\{1 - H_{1x}(v)\}^2} \mathrm{d}H_{1x}^{u}(v) \right. \\ &\qquad + \frac{\mathbb{I}(T_i \leq t, \delta_i = 1) - H_{1x}^{u}(t)}{1 - H_{1x}(t)} \\ &\qquad \left. - \int_0^t \frac{\mathbb{I}(T_i \leq v, \delta_i = 1) - H_{1x}^{u}(v)}{\{1 - H_{1x}(v)\}^2} \mathrm{d}H_{1x}(v) \right], \end{aligned}$$

which coincides with the asymptotic i.i.d representation given in Van Keilegom & Veraverbeke (1997).

4.3.2 Weak convergence of $\mathbb{F}_{xh}^{(\mathbf{rc})}$

In view of Theorem 1, the large sample behavior of $\mathbb{F}_{xh}^{(\mathrm{rc})}$ will essentially depend on the conditions imposed on the weight functions and on the bandwidth parameters h and g. To establish its weak limit, we consider the mean zero gaussian processes $\mathbb{J}_x^{(1)}$ and $\mathbb{J}_x^{(2)}$ with covariance function

$$\mathrm{Cov}\{\mathbb{J}_x^{(1)}(t,y),\mathbb{J}_x^{(1)}(t',y')\} \;=\; K_4\left[\int_0^{t\wedge t'}\frac{f_x(v,y\wedge y')}{1-G_x(v)}\mathrm{d}v - F_x(t,y)F_x(t',y')\right]$$

and

$$\mathrm{Cov}\{\mathbb{J}_x^{(2)}(t,y),\mathbb{J}_x^{(2)}(t',y')\} \;=\; K_4\int_0^{t\wedge t'}\frac{\upsilon_x(t\wedge t',y,v)\upsilon_x(t\wedge t',y',v)}{\{1-H_{1x}(v)\}^2}\mathrm{d}H_{1x}^u(v),$$

where the constant K_4 is defined in Assumption W_4 and $f_x(t,y)=\frac{\partial}{\partial t}F_x(t,y)$. Moreover, let

$$\begin{aligned} b_x^{(1)}(t,y) \;=\;& K_2\dot{F}_x(t,y)+\frac{K_3}{2}\ddot{F}_x(t,y) \\ & -\int_0^t\frac{K_2\dot{G}_x(v)+\frac{K_3}{2}\ddot{G}_x(v)}{1-G_x(v)}f_x(v,y)\mathrm{d}v \end{aligned}$$

and

$$\begin{aligned} b_x^{(2)}(t,y) \;=\;& \int_0^t\frac{K_2\dot{H}_{1x}(v)+\frac{K_3}{2}\ddot{H}_{1x}(v)}{1-H_{1x}(v)}\upsilon_x(t,y,v)\mathrm{d}H_x^c(v) \\ & +\int_0^t\frac{\upsilon_x(t,y,v)}{\{1-H_{1x}(v)\}}\mathrm{d}\left\{K_2\dot{G}_x^u(v)+\frac{K_3}{2}\ddot{H}_x^c(v)\right\}. \end{aligned}$$

In the latter, the constants K_2–K_3 are given in Assumption W_2–W_3. On one hand, the deterministic function $b_x^{(1)}$ will appears as the asymptotic bias of the process $\sum_{i=1}^n w_{ni}(x,h)\mathcal{J}_{ix}^{(1)}$. Therefore, it would have been present even if G_{xg} was replaced with G_x in the definition of $F_{xh}^{(\mathrm{rc})}$. On the other hand, $b_x^{(2)}$ will emerge as the bias related to the estimation of the conditional probability of censoring.

Corollary 1 *Suppose that the assumptions of Theorem 1 are met. For any $0 < \mathfrak{t} < \overline{\tau}_x$, write $\mathcal{T}_{\mathfrak{t}} = [0, \mathfrak{t}] \times \mathbb{R}$.*

(a) If $hg^{-1} \to 0$, and in addition if $\sqrt{nh}g^2 \to K$ for some $K > 0$, then $\mathbb{F}_{xh}^{(rc)}$ converges weakly in $l^\infty(\mathcal{T}_{\mathfrak{t}})$ to $\mathbb{J}_x^{(1)} + Kb_x^{(2)}$ over $\mathcal{T}_{\mathfrak{t}}$.

(b) If $g = h$, and in addition if $\sqrt{nh^5} \to K$ for some $K > 0$, then $\mathbb{F}_{xh}^{(rc)}$ converges weakly to a gaussian process having the representation $\mathbb{J}_x := \mathbb{J}_x^{(1)} + \mathbb{J}_x^{(2)} + K\{b_x^{(1)} + b_x^{(2)}\}$ over $\mathcal{T}_{\mathfrak{t}}$, with

$$\begin{aligned}
&\mathrm{Cov}\{\mathbb{J}_x^{(1)}(t,y), \mathbb{J}_x^{(2)}(t',y')\} \\
&\quad = \int_0^{t'} \frac{\upsilon_x(t',y',v)}{\{1 - H_{1x}(v)\}^2}\{F_x(t\wedge v, y) - H_{1x}(v)F_x(t,y)\}\,\mathrm{d}H_x^c(v) \\
&\qquad - F_x(t,y)\int_0^{t'} \frac{H_x^c(v)}{\{1-H_{1x}(v)\}\{1-G_x(v)\}} h_x^u(v,y')\mathrm{d}v \\
&\qquad + F_x(t,y)\int_0^{t'} \frac{H_x^c(v)}{\{1-H_{1x}(v)\}^2}\upsilon_x(t',y',v)\mathrm{d}H_{1x}(v).
\end{aligned}$$

Remark 1 *In view of Part (a) of Corollary 1, the impact of estimating the conditional probability of censoring is negligible, provided that the bandwidth g is asymptotically larger than h. However, it is still required that $ng^5 < \infty$, which means that g must not exceed $n^{-1/5}$. Also, because $h/g \to 0$, we have $h \sim o(n^{-1/5})$. Therefore, choosing g larger than h excludes the optimal bandwidth parameter order for h in terms of mean squared error.*

When the probability of censoring is 0, the term $\mathbb{J}_x^{(2)}$ is not present in the weak limit of $\mathbb{F}_{xh}^{(\mathrm{rc})}$. Therefore, the asymptotic covariance function of $\mathbb{F}_{xh}^{(\mathrm{rc})}$ becomes

$$\begin{aligned}
\mathrm{Cov}\{\mathbb{J}_x^{(1)}(t,y), \mathbb{J}_x^{(1)}(t',y')\} &= K_4\left[\int_0^{t\wedge t'} f_x(v, y\wedge y')\mathrm{d}v - F_x(t,y)F_x(t',y')\right] \\
&= K_4\left[F_x(v, y\wedge y') - F_x(t,y)F_x(t',y')\right],
\end{aligned}$$

which corresponds to the asymptotic variance of the process $\sqrt{nh}(F_{xh} - F_x)$ in the context of complete i.i.d data (see e.g. Veraverbeke *et al.* (2011)). Moreover, the bias reduces to

$$K_2\dot{F}_x(t,y) + \frac{K_3}{2}\ddot{F}_x(t,y),$$

which matches the asymptotic bias of the process $\sqrt{nh}(F_{xh} - F_x)$.

Remark 2 *As mentioned in Remark 1, using two different bandwidth parameters in the estimation of the conditional distribution excludes the theoretical optimal order for* h*. Nevertheless, this implies that the bias related to the estimation of* F_x *provided* G_x *is known, namely* $b_x^{(1)}$*, becomes neglible. Hence, in some cases, one might obtain a bias reduction at the cost of excluding the optimal order for* h*. Note however that the same dilemma traditionally occurs in nonparametric density estimation, referring to the decision to under-smooth or not.*

4.3.3 Weak convergence of $\mathbb{B}_{xh}^{(\mathbf{rc})}$

The next result states the weak limit of the conditional copula estimator under random censoring.

Proposition 1 *Suppose that the assumptions in Theorem 1 are satisfied, and assume Condition* $(\mathcal{D})$*, given in the appendix, regarding the partial derivatives of the conditional copula, is satisfied. For any* $0 < \mathfrak{t} < \overline{\tau}_x$ *and by denoting* $\widetilde{\mathfrak{T}}_{\mathfrak{t}} = [0, H_{1x}^u(\mathfrak{t})] \times [0, 1]$*, we have,*
(a) If $hg^{-1} \to 0$*, and* $\sqrt{nh}g^2 \to K$ *for some* $K > 0$*, then* $\mathbb{B}_{xh}^{(rc)}$ *converges weakly in* $l^\infty(\widetilde{\mathfrak{T}}_{\mathfrak{t}})$ *to a gaussian process with the following representation :*

$$\alpha_x^{(rc)}(u,v) - C_x^{[1]}(u,v)\alpha_x^{(rc)}(u,1) - C_x^{[2]}(u,v)\alpha_x^{(rc)}(1,v),$$

where $\alpha_x^{(rc)}(u,v) = \mathbb{J}_x^{(1)}\{F_{1x}^{-1}(u), F_{2x}^{-1}(v)\} + b_x^{(2)}\{F_{1x}^{-1}(u), F_{2x}^{-1}(v)\}$*.*
(b) If $h = g$ *and* $\sqrt{nh^5} \to K$ *for some* $K > 0$*, then* $\mathbb{B}_{xh}^{(rc)}$ *converges weakly in* $l^\infty(\widetilde{\mathfrak{T}}_{\mathfrak{t}})$ *to a gaussian process with the following representation*

$$\beta_x^{(rc)}(u,v) - C_x^{[1]}(u,v)\beta_x^{(rc)}(u,1) - C_x^{[2]}(u,v)\beta_x^{(rc)}(1,v),$$

where $\beta_x^{(rc)}(u,v) = \mathbb{J}_x\{F_{1x}^{-1}(u), F_{2x}^{-1}(v)\}$*.*

As pointed out in Section 4.3.2, when all the survival times are completely observed, the term $\mathbb{J}_x^{(2)}$ reduces to 0. In this case, the covariance structure of the limit process $\alpha_x^{(\mathrm{rc})}$ matches the one of the conditional copula process $\sqrt{nh}(\mathbb{C}_{xh} - \mathbb{C}_x)$ found in Veraverbeke *et al.* (2011).

4.4 Simulation study

The nonparametric estimation of the conditional copula involves a choice for the weight functions $w_{n1}(x,h),\ldots,w_{nn}(x,h)$ that fulfills the required assumptions listed in Section 4.6.2. It is shown in Omelka *et al.* (2013) that the requirements W_1–W_5 are satisfied, among others, by the Nadaraya–Watson and local linear weights, presented in Section 4.2.

The simulation results that will be reported here have been obtained using the local linear weights with the triweight function $K(y) = 35(1-y^2)^3\,\mathbb{I}(|y|\leq 1)/32$. When it happens, negative weights are taken to be zero and the remaining weights are simply re-scaled in order that they sum to one. As pointed out in Omelka *et al.* (2013), this modification is asymptotically negligible. Finally note that all the numerical experiments were also run using the Nadaraya–Watson kernel. As the results were very similar, they are not presented here.

The primary aim of this section is to evaluate the performance of the proposed conditional copula estimator with respect to the percentage of censoring, the influence of the covariate on the dependance and the effect of the sample size. This performance is evaluated by considering the *average squared bias* (ASB) and the *average variance* (AV). To be specific, if $\widehat{\mathbb{C}}_x$ is some estimator of $\mathbb{C}_x$, then

$$\mathrm{ASB}(\widehat{\mathbb{C}}_x) = \frac{1}{K^2}\sum_{i,j=1}^{K}\left\{\mathrm{E}(\widehat{\mathbb{C}}_x(u_i,u_j)) - \mathbb{C}_x(u_i,u_j)\right\}^2$$

and

$$\mathrm{AV}(\widehat{\mathbb{C}}_x) = \frac{1}{K^2}\sum_{i,j=1}^{K}\mathrm{E}\left\{\widehat{\mathbb{C}}_x^2(u_i,u_j) - \left(\mathrm{E}(\widehat{\mathbb{C}}_x(u_i,u_j))\right)^2\right\}.$$

The latter have been estimated from 1 000 replicates under each of the scenario considered for $x = 0.5$ with $n = 250$ and $n = 1\,000$ and $K = 15$.

Also, the nonparametric estimation of $\mathbb{C}_x$ requires a choice for either one or two bandwidth parameters. Indeed, an interesting aspect of Proposition 1 is that the limiting distribution of the copula process $\mathbb{B}_{xh}^{(\mathrm{rc})}$ differs in the case where $g = h$ and $\frac{h}{g} \to 0$. The secondary aim of this section is to evaluate the impact of using a single or two bandwidth parameters in the estimation of $\mathbb{C}_x$. In the following, we denote by $\mathbb{C}_{xh}^{(rc,1)}$ and $\mathbb{C}_{xh}^{(rc,2)}$ the estimators resulting from the choices $g = h$ and $g \neq h$ respectively. Upon setting $g = h \times 0.25 \log(n)$ for $\mathbb{C}_{xh}^{(rc,2)}$, their performance are compared for different values of h.

The covariate is simulated as a standard normal and the estimation of the conditional copula is evaluate at $x = 0.5$. The copula which joins the marginals is either a normal copula $\mathbb{C}_\varrho^{\mathrm{N}}$ or a Clayton copula $\mathbb{C}_\gamma^{\mathrm{CL}}$. Theses are defined for $-1 < \varrho < 1$ and $\theta > 0$ by

$$\mathbb{C}_\varrho^{\mathrm{N}}(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \varphi_\varrho(y, z)\, \mathrm{d}z\, \mathrm{d}y,$$

and

$$\mathbb{C}_\gamma^{\mathrm{CL}}(u, v) = \left(u^{-\gamma} + v^{-\gamma} - 1\right)^{-1/\gamma},$$

where φ_ϱ is the bivariate standard Normal density with correlation ϱ and Φ is the standard Normal distribution. Their parameters will be set to vary with X in the following way. In fact, since the copula parameters are sometimes hard to interpret, it is convenient to quantify the dependence in a bivariate random vector using its corresponding value of Kendall's tau. For any copula $\mathbb{C}$, its associate Kendall's tau can be written as

$$\mathfrak{T}(\mathbb{C}) = 4 \int_0^1 \int_0^1 \mathbb{C}(u_1, u_2)\, \mathrm{d}\mathbb{C}(u_1, u_2) - 1. \tag{4.4}$$

For a given conditional copula $\mathbb{C}_x$, the conditional Kendall's tau is simply $\mathfrak{T}_x = \mathfrak{T}(\mathbb{C}_x)$. Both copulas are parameterized in such a way that $\mathfrak{T}_x = \mathrm{const} \times \{\Phi(x) - 0.2\}^2$. From the relationships between Kendall's tau and the parameters of the Normal and Clayton copulas, this can be done by setting

$$\varrho(x) = \sin\left(\frac{\mathfrak{T}_x \times \pi}{2}\right) \qquad \text{and} \qquad \gamma(x) = \frac{2\,\mathfrak{T}_x}{1 - \mathfrak{T}_x}.$$

This means that the triplet (Y_{1i}, Y_{2i}, x_i) is obtained by considering either $\varrho(x_i)$ or $\gamma(x_i)$ combined with the corresponding copula. The constant is chosen so that $\mathfrak{T}_x \in \{0, .1, .25, .35\}$.

As discussed at the beginning of Section 4.3.1, the tail support of the distribution of a random variable may not be identifiable due to right censoring when the support of the censoring variable is included in the support of the variable of interest, i.e when $\tau_{C_x} < \tau_{F_x}$. To evaluate the impact on the estimation of $\mathbb{C}_x$, the cases where $\tau_{C_x} = \tau_{F_x}$ and $\tau_{C_x} < \tau_{F_x}$ are examined separately in the following two sections.

4.4.1 $\tau_{C_x} = \tau_{F_x}$

Here, we have considered the case where $\tau_{C_x} = \tau_{F_x} = \infty$. To do this, the marginal distributions of Y_{1i} and Y_{2i} are generated from the exponential distribution with mean given by

$$\lambda_{x_i} = a\{1 + \Phi(x_i) + \Phi(x_i)^2\},$$

i.e

$$F_{jx_i}(y) = \mathrm{P}\{Y_{ji} \leq y \mid X = x_i\} = 1 - e^{-\frac{y}{\lambda_{x_i}}}.$$

The censoring variable C_i is also picked as an exponential but with mean $c\{1 + \Phi(x_i) + \Phi(x_i)^2\}$. Hence, the probability of censoring conditional on $X = x$, denoted θ thereafter, is simply $\frac{a}{a+c}$. The results are reported for $a = 5$ and $\theta \in \{.2, .4, .6\}$ in Table 4.1 and 4.2.

4.4.2 $\tau_{C_x} < \tau_{F_x}$

Here, we have considered $\tau_{F_x} = \infty$ and $\tau_{C_x} < \infty$. In that case, the marginal distributions of Y_{1i} and Y_{2i} are generated from the exponential distribution with mean λ_{x_i}. The censoring variable C_i is generated from the uniform distribution over $[0, \text{const} \times \lambda_{x_i}]$. We can show that in this case, the percentage of censoring is given by

$$\theta = \frac{1}{\text{const}} \times \left(1 - e^{-\text{const}}\right).$$

The constant is chosen so that $\theta \in \{0.2, 0.4\}$. We have also cover the scenario $\theta = 0$, which corresponds to the situation when all the survival times are observed. The results are reported in Table 4.3 and 4.4.

4.4.3 Comments on the simulations results

From the obtained results it can be seen that, globally, when the association between Y_1 and Y_2 increases, the bias increases whereas the variance decreases slightly. Another interesting finding is that there is no significant difference between the results obtained with a single bandwidth ($g = h$) and double bandwidth ($g \neq h$), except for a large sample size ($n = 1000$). In this case, double bandwidth reduces the bias substantially without increasing the variance of the resulting estimator. The results from the Normal and Clayton copula are quite similar. Also, and as expected, increasing the percentage of censoring decreases the performances of the copula estimator both in terms of bias and variance. The opposite is observed regarding the effect of the sample size. A larger (smaller) bandwidth is needed when censoring (sample size) increase. Notice that we obtain much more accurate information on the conditional copula when $\tau_{C_x} = \tau_{F_x}$. When $\tau_{C_x} < \tau_{F_x}$, one needs a large sample size to get accurate estimates otherwise the results should be interpreted with care especially when the percentage of censoring is high. Finally, as for any kernel based estimator, we can see that a large bandwidth, typically, leads to a larger bias and smaller variance. This becomes clear with large sample size (see the results for $n = 1000$).

4.5 Illustrative example

In this section we consider a dataset that was used in Andersen *et al.* (1993). This data contains information on 205 patients with malignant melanoma that were followed for a period up to 15 years. The main variable of interest is Y_1 : the survival times after surgery for skin tumour. Other measured quantities include Y_2 : the age of the patient when the surgery occurs and X : the tumour thickness in mm. 134 patients were alive by the end of the follow-up period and 14 patients died of causes unrelated to melanoma. These patients are censored ($status = 0$) at their last observed duration time or death time. All the reaming patients died from melanoma and so they are uncensored ($status = 1$). The typical objective of such studies is to asses the effect of risk factors (like age and tumour thickness) on survival time. This is done usually by constructing a regression model with Y_1 as response and Y_2 and X as covariates. Before attempting to model the relations between these variables, it may be helpful to measure the strength of the relationship between them using model-free tools.

Kendall's tau is a popular coefficient that measure the concordance-discordance between two random variables. This coefficient lies in $[-1, 1]$ and is equal to zero for independent random variables. In contrast to the well-known Pearson correlation coefficient, Kendall's tau does not require knowledge of the parametric form of the marginal distributions. For more details, see Nelsen (2006). A conditional version of this coefficient was suggested by Gijbels *et al.* (2011). In terms of copula, the population version of the conditional Kendall's tau of (Y_1, Y_2) given $X = x$ is

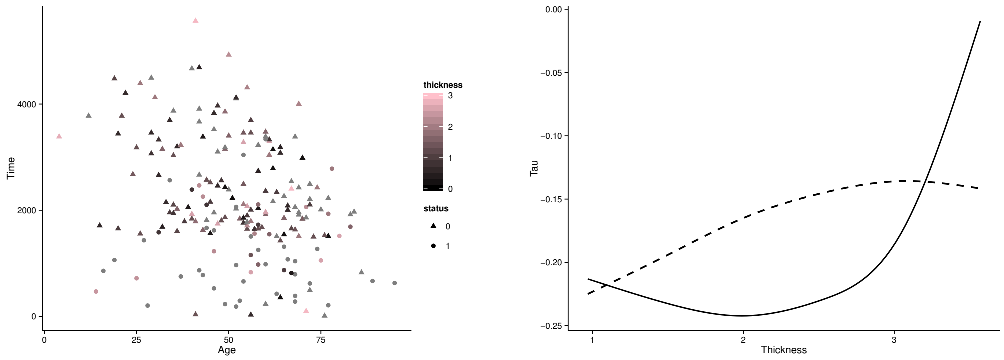
$$\mathfrak{T}(\mathbb{C}_x) = 4 \int_0^1 \int_0^1 \mathbb{C}_x(u_1, u_2)\, \mathrm{d}\mathbb{C}_x(u_1, u_2) - 1. \tag{4.5}$$

A natural way to estimate this coefficient is to replace the unknown quantity $\mathbb{C}_x$ in the above expression by its nonparametric estimator $\hat{\mathbb{C}}_x$ given by (4.4). This can be expressed as

$$\mathfrak{T}(\hat{\mathbb{C}}_x) = 4 \int_0^{H_{1x}^u(\mathfrak{t})} \int_0^1 \hat{\mathbb{C}}_x(u_1, u_2) d\hat{\mathbb{C}}_x(u_1, u_2) - 1. \tag{4.6}$$

Except the case when $H_{1x}^u(\mathfrak{t}) = 1$, the truncation in the integral above is needed because $\hat{\mathbb{C}}_x$ is inconsistent outside $[0, H_{1x}^u(\mathfrak{t})] \times [0, 1]$, see Proposition 1 above. Unfortunately the quantity $H_{1x}^u(\mathfrak{t})$ is unknown and there is no obvious way to estimate it without imposing some restrictive assumptions on the data generating process. In practice one may consider (4.6) without the truncation, but then the results should be interpreted with care.

Figure 1 (a) shows the scatter plot of the observed survival times on the y axis and age values on the x axis using different symbols for censored/uncensored observations and different colors for tumour thickness. From this figure it can be seen that there is a relationship between time and age : the survival time has tendency to decrease with increasing age. This tendency is not very strong as the estimated unconditional (global) Kendall's tau is only of -0.13. Figure 1 (b) shows the estimated conditional Kendall's tau between time and age given thickness. The dashed curve corresponds to the estimator, say $\mathfrak{T}(\hat{\mathbb{C}}_x^{ic})$, obtained ignoring censoring, i.e. we consider all observed times as exact, and the solid curve is the estimator $\mathfrak{T}(\hat{\mathbb{C}}_x)$ obtained using our method that takes into account censoring. For both estimators a bandwidth $h = 0.93$ was used. We can see that while the estimated conditional Kendall's tau coefficients remain negative their magnitude changes with thickness. When the latter increases, the absolute value of $\mathfrak{T}(\hat{\mathbb{C}}_x)$ slightly increases to reach its maximum value of 0.242 when tumor thickness is $2mm$ and then it starts decreasing rapidly to reach 0. So unlike the global Kendall's tau which measures only the "average" association between time and age, $\mathfrak{T}(\hat{\mathbb{C}}_x)$ gives us a more precise picture about this association accounting for the effect of tumor thickness. From the figure it seems that, except for large tumor thickness, the "uncorrected" estimator $\mathfrak{T}(\hat{\mathbb{C}}_x^{ic})$ underestimates the strength of association between time and age.



(a) (b)

Figure 4.1 – (a) Scatter plot of time versus age (1 ≡ uncensored, 0 ≡ censored); (b) Conditional Kendall's as function of thickness estimated (i) taking into account censoring (solid curve) and (ii) considering censored observations as exact survival times (dashed curve).

4.6 Assumptions

4.6.1 (sub-)distribution functions

Smoothness conditions over $F_x, H_x, H_x^u, H_{1x}^c$ and G_x are needed in the proof of Theorem 1. We formulate the conditions for a general (sub-)distribution function L_x.

($\mathcal{C}_1$) $\dot{L}_x(t,y) = \frac{\partial}{\partial x} L_x(t,y)$ exists and is continuous over $V(x) \times \mathcal{T}_\mathrm{t}$, where $V(x)$ is a neighborhood of x;

($\mathcal{C}_2$) $L_x^{(1)}(t,y) = \frac{\partial}{\partial t} L_x(t,y)$ and $L_x^{(2)}(t,y) = \frac{\partial}{\partial y} L_x(t,y)$ exist and are continuous over $V(x) \times \mathcal{T}_\mathrm{t}$;

($\mathcal{C}_3$) $\ddot{L}_x(t,y) = \frac{\partial^2}{\partial x^2} L_x(t,y)$ exist and is continuous over $V(x) \times \mathcal{T}_\mathrm{t}$;

($\mathcal{C}_4$) $L_x^{(1,1)}(t,y) = \frac{\partial^2}{\partial t^2} L_x(t,y)$,$L_x^{(1,2)}(t,y) = \frac{\partial^2}{\partial t \partial y} L_x(t,y)$ and $L_x^{(2,2)}(t,y) = \frac{\partial^2}{\partial y^2} L_x(t,y)$ exist and are continuous over $V(x) \times \mathcal{T}_\mathrm{t}$;

($\mathcal{C}_5$) $\dot{L}_x^{(1)}(t,y) = \frac{\partial^2}{\partial t \partial x} L_x(t,y)$ and $\dot{L}_x^{(2)}(t,y) = \frac{\partial^2}{\partial y \partial x} L_x(t,y)$ exist and are continuous over $V(x) \times \mathcal{T}_\mathrm{t}$;

($\mathcal{C}_6$) $\dot{L}_x^{(1,2)}(t,y) = \frac{\partial^3}{\partial t \partial y \partial x} L_x(t,y)$ and $\ddot{L}_x^{(1,2)}(t,y) = \frac{\partial^4}{\partial t \partial y \partial x^2} L_x(t,y)$ exist and are continuous over $V(x) \times \mathcal{T}_{\mathfrak{t}}$;

The following assumption is needed to guaranty the weak convergence of $\mathbb{B}_{xh}^{(\mathrm{rc})}$.

($\mathcal{D}$). The partial derivatives $\mathbb{C}_x^{[1]}(u,v) = \partial\, \mathbb{C}_x(u,v)/\partial u$ and $\mathbb{C}_x^{[2]}(u,v) = \partial\, \mathbb{C}_x(u,v)/\partial v$ exist and are continuous on $(0,1) \times [0,1]$ and $[0,1] \times (0,1)$ respectively.

4.6.2 Weight functions

Assumptions W_1–W_5 below are requierd to establish the asymptotic behavior of $\sqrt{nh}\{F_{xh}^{\mathrm{rc}} - F_x\}$ stated in Section 4.3.

W_1. $\sqrt{nh} \max_{1 \le i \le n} |w_{ni}(x,h)| = o(1)$;

W_2. $\sqrt{nh} \left| \sum_{i=1}^{n} w_{ni}(x,h)(x_i - x) - h^2 K_2 \right| = o(1)$ for some $K_2 = K_2(x) \in (0,\infty)$;

W_3. $\sqrt{nh} \left| \sum_{i=1}^{n} w_{ni}(x,h)(x_i - x)^2/2 - h^2 K_3 \right| = o(1)$ for some $K_3 = K_3(x) \in (0,\infty)$;

W_4. $nh \sum_{i=1}^{n} \{w_{ni}(x,h)\}^2 - K_4 = O(1)$ for some $K_4 = K_4(x) \in (0,\infty)$;

W_5. $\max_{i \in I_{nx}} x_i - \min_{i \in I_{nx}} x_i = o(1)$, where $I_{nx} = \{i : w_{ni}(x,h) > 0\}$.

Tableau 4.1 – Average integrated square bias (AISB $\times 10^4$) estimated from 1 000 replicates of for $\mathbb{C}_{xh}^{(\mathrm{rc})}$ with $n = 250$ and $n = 1\ 000$ in the case $\tau_{C_x} = \tau_{F_x} = \infty$. Upper pannel : Normal Copula. Bottom pannel : Clayton Copula.

θ	h	$\mathfrak{T}_x = .1$				$\mathfrak{T}_x = .25$				$\mathfrak{T}_x = .35$			
		$n = 250$		$n = 1000$		$n = 250$		$n = 1000$		$n = 250$		$n = 1000$	
		$g = h$	$g \neq h$	$g = h$	$g \neq h$	$g = h$	$g \neq h$	$g = h$	$g \neq h$	$g = h$	$g \neq h$	$g = h$	$g \neq h$
20%	0.9	1.420	0.510	0.223	0.023	1.403	0.481	0.232	0.028	1.515	0.536	0.286	**0.051**
	1.2	0.356	0.167	**0.014**	**0.009**	0.361	0.178	**0.022**	**0.028**	0.471	0.264	**0.057**	0.057
	1.5	0.138	0.116	0.021	0.030	0.138	0.118	0.057	0.073	0.237	0.212	0.102	0.115
	1.8	0.049	0.050	0.047	0.049	**0.083**	**0.083**	0.106	0.111	**0.188**	**0.185**	0.164	0.168
	2.1	**0.038**	**0.037**	0.085	0.087	0.091	0.090	0.161	0.167	0.216	0.214	0.225	0.231
	2.4	0.049	0.048	0.095	0.096	0.114	0.114	0.189	0.195	0.254	0.251	0.265	0.270
	2.7	0.063	0.064	0.120	0.121	0.159	0.163	0.234	0.240	0.296	0.299	0.319	0.324
40%	0.9	1.707	1.464	0.532	0.050	1.525	1.282	0.535	0.045	1.553	1.304	0.609	**0.068**
	1.2	0.973	0.639	0.078	**0.032**	0.859	0.537	**0.072**	**0.039**	0.948	0.604	**0.119**	0.074
	1.5	0.535	0.405	**0.054**	0.035	0.435	0.311	**0.072**	0.067	0.530	**0.392**	0.149	0.131
	1.8	0.390	0.374	0.060	0.056	0.318	0.302	0.114	0.111	**0.430**	0.409	0.214	0.204
	2.1	0.329	0.321	0.079	0.077	0.300	0.286	0.145	0.151	0.457	0.432	0.270	0.265
	2.4	0.291	0.280	0.088	0.084	**0.273**	**0.258**	0.190	0.186	0.448	0.425	0.340	0.328
	2.7	**0.259**	**0.249**	0.090	0.087	0.291	0.277	0.210	0.208	0.487	0.466	0.384	0.375
60%	0.9	10.841	8.653	1.551	1.059	9.378	7.289	1.271	0.788	8.699	6.667	1.186	0.683
	1.2	5.339	4.948	0.923	0.764	4.297	3.935	0.699	**0.538**	3.915	3.566	**0.662**	**0.480**
	1.5	4.465	4.420	0.881	0.733	3.342	3.283	0.715	0.566	2.959	2.888	0.743	0.571
	1.8	4.441	4.230	0.752	0.758	3.329	3.130	0.618	0.590	2.968	2.762	0.723	0.662
	2.1	3.699	3.680	**0.687**	**0.691**	2.764	2.713	0.627	0.601	2.602	2.530	0.798	0.739
	2.4	3.705	3.641	0.791	0.743	**2.622**	**2.538**	0.733	0.659	**2.423**	**2.315**	0.950	0.844
	2.7	3.607	3.502	0.824	0.787	2.673	2.561	0.782	0.721	2.550	2.424	1.053	0.961
20%	0.9	1.324	0.442	0.185	0.024	1.567	0.618	0.214	**0.034**	1.563	0.432	0.486	**0.083**
	1.2	0.325	0.142	**0.031**	**0.014**	0.191	0.170	**0.061**	0.076	0.368	0.305	**0.163**	0.191
	1.5	0.141	0.085	0.047	0.049	**0.161**	**0.135**	0.174	0.180	**0.323**	**0.325**	0.318	0.347
	1.8	0.061	**0.055**	0.104	0.106	0.178	0.179	0.283	0.292	0.417	0.419	0.488	0.529
	2.1	**0.060**	0.060	0.121	0.123	0.248	0.252	0.346	0.353	0.561	0.567	0.622	0.634
	2.4	0.085	0.084	0.156	0.157	0.302	0.304	0.441	0.450	0.656	0.662	0.772	0.786
	2.7	0.097	0.096	0.223	0.226	0.368	0.370	0.551	0.560	0.792	0.797	0.915	0.930
40%	0.9	2.746	1.376	0.608	0.158	2.282	1.086	0.527	0.087	2.250	1.266	0.431	**0.118**
	1.2	0.864	0.600	0.072	**0.039**	0.910	0.606	**0.130**	**0.085**	0.925	0.506	**0.228**	0.191
	1.5	0.681	0.503	**0.059**	0.054	0.498	0.378	0.152	0.154	0.528	**0.485**	0.318	0.324
	1.8	0.369	0.330	0.101	0.098	0.363	0.339	0.254	0.258	**0.551**	0.532	0.488	0.497
	2.1	**0.247**	**0.226**	0.118	0.116	**0.338**	**0.331**	0.297	0.303	0.616	0.611	0.587	0.598
	2.4	0.295	0.284	0.131	0.130	0.402	0.393	0.372	0.378	0.716	0.710	0.714	0.724
	2.7	0.350	0.337	0.178	0.178	0.487	0.471	0.455	0.464	0.873	0.864	0.850	0.867
60%	0.9	11.450	7.530	1.866	1.264	9.481	6.093	1.555	1.000	9.294	5.838	2.023	0.767
	1.2	6.922	5.600	0.976	0.885	5.650	4.892	0.871	0.709	4.649	3.984	**0.785**	**0.662**
	1.5	4.734	4.604	**0.787**	**0.729**	4.327	3.823	**0.726**	**0.676**	3.532	3.051	0.789	0.697
	1.8	3.968	3.849	0.841	0.814	2.822	2.675	0.756	0.711	2.421	2.351	0.946	0.891
	2.1	3.569	3.428	0.955	0.972	**2.428**	**2.359**	0.877	0.823	**2.281**	**2.173**	1.072	1.026
	2.4	**3.431**	**3.385**	0.957	0.920	2.712	2.603	0.997	0.956	2.446	2.321	1.265	1.212
	2.7	4.136	4.099	0.997	0.969	3.166	3.043	1.034	0.983	2.918	2.818	1.372	1.306

Tableau 4.2 – Average integrated variance (AIV $\times 10^4$) estimated from 1 000 replicates of for $\mathbb{C}_{xh}^{(\mathrm{rc})}$ with $n = 250$ and $n = 1\ 000$ in the case $\tau_{C_x} = \tau_{F_x} = \infty$. Upper pannel : Normal Copula. Bottom pannel : Clayton Copula.

θ	h	$\mathfrak{T}_x = .1$				$\mathfrak{T}_x = .25$				$\mathfrak{T}_x = .35$			
		$n = 250$		$n = 1000$		$n = 250$		$n = 1000$		$n = 250$		$n = 1000$	
		$g = h$	$g \neq h$	$g = h$	$g \neq h$	$g = h$	$g \neq h$	$g = h$	$g \neq h$	$g = h$	$g \neq h$	$g = h$	$g \neq h$
20%	0.9	25.033	9.210	14.408	3.398	26.750	9.235	15.656	3.515	27.752	9.117	16.448	3.559
	1.2	10.312	3.793	4.302	0.945	10.633	3.426	4.534	0.827	10.724	3.072	4.662	0.730
	1.5	5.537	4.444	1.934	0.807	5.385	4.176	1.968	0.723	5.187	3.902	1.959	0.640
	1.8	2.878	2.882	0.708	0.711	2.556	2.560	0.614	0.619	2.264	2.266	0.543	0.548
	2.1	2.691	2.693	0.645	0.648	2.378	2.382	0.569	0.571	2.069	2.075	0.497	0.499
	2.4	2.504	2.514	0.614	0.616	2.194	2.202	0.539	0.542	1.920	1.927	0.469	0.472
	2.7	**2.438**	**2.449**	**0.589**	**0.591**	**2.106**	**2.115**	**0.509**	**0.511**	**1.812**	**1.819**	**0.446**	**0.448**
40%	0.9	19.754	16.672	19.448	1.997	20.449	17.054	21.114	1.857	20.733	17.127	22.155	1.703
	1.2	12.547	6.187	5.979	1.536	12.789	5.728	6.301	1.392	12.750	5.243	6.474	1.264
	1.5	8.735	5.478	4.680	1.334	8.591	4.989	4.923	1.228	8.350	4.530	5.038	1.120
	1.8	4.810	4.798	1.172	1.166	4.347	4.339	1.052	1.051	3.932	3.923	0.952	0.956
	2.1	4.419	4.425	2.225	1.105	3.987	3.996	2.241	1.003	3.573	3.582	2.212	0.902
	2.4	4.256	4.257	1.012	1.012	3.831	3.835	0.916	0.918	3.448	3.457	0.815	0.821
	2.7	**4.160**	**4.165**	**0.993**	**0.993**	**3.727**	**3.732**	**0.894**	**0.898**	**3.317**	**3.324**	**0.800**	**0.804**
60%	0.9	39.515	30.505	15.744	6.265	40.965	30.816	16.583	6.046	41.665	30.771	17.002	5.784
	1.2	17.887	15.903	8.380	4.062	17.286	15.089	8.564	3.782	16.588	14.231	8.591	3.513
	1.5	14.052	13.855	6.879	3.610	13.212	13.017	6.934	3.323	12.337	12.156	6.879	3.052
	1.8	12.913	11.926	3.225	3.169	11.996	10.901	2.952	2.917	11.334	10.160	2.701	2.684
	2.1	11.578	11.564	3.018	2.959	10.467	10.472	2.714	2.681	9.569	9.577	2.435	2.428
	2.4	10.745	10.602	2.844	2.822	9.693	9.577	2.552	2.540	8.816	8.714	2.286	2.283
	2.7	**10.262**	**10.160**	**2.665**	**2.641**	**9.278**	**9.225**	**2.373**	**2.360**	**8.480**	**8.454**	**2.124**	**2.124**
20%	0.9	23.844	8.186	12.115	2.329	28.510	12.680	15.334	1.093	28.287	5.322	23.461	2.235
	1.2	10.258	3.809	6.478	0.967	5.766	4.588	3.260	0.844	8.048	4.297	5.865	0.754
	1.5	6.459	3.218	0.790	0.793	6.420	2.850	0.708	0.708	2.583	3.839	1.908	0.616
	1.8	4.062	2.967	0.714	0.716	2.582	2.591	0.631	0.634	2.283	2.287	1.854	0.556
	2.1	2.649	2.651	0.645	0.647	2.318	2.324	0.570	0.570	2.050	2.065	0.496	0.499
	2.4	2.499	2.500	0.617	0.618	2.138	2.146	0.527	0.530	1.873	1.872	0.459	0.459
	2.7	**2.401**	**2.408**	**0.588**	**0.590**	**2.065**	**2.068**	**0.516**	**0.517**	**1.809**	**1.815**	**0.447**	**0.449**
40%	0.9	29.534	14.413	19.150	6.339	26.838	11.294	20.597	4.213	26.139	13.168	17.852	2.914
	1.2	12.588	7.344	4.891	1.605	16.057	9.181	8.617	1.456	16.231	5.245	8.918	1.334
	1.5	9.609	5.360	3.532	1.328	8.377	4.862	1.205	1.207	5.740	4.483	1.097	1.093
	1.8	5.864	4.788	2.257	1.154	5.476	4.285	1.066	1.068	3.867	3.876	0.970	0.968
	2.1	4.466	4.469	1.079	1.078	3.979	3.963	0.973	0.973	3.525	3.534	0.885	0.889
	2.4	4.187	4.179	1.030	1.029	3.732	3.736	0.909	0.906	3.395	3.405	0.809	0.809
	2.7	**4.029**	**4.033**	**0.981**	**0.983**	**3.583**	**3.580**	**0.874**	**0.878**	**3.202**	**3.206**	**0.782**	**0.786**
60%	0.9	41.210	22.473	15.480	6.182	40.648	20.576	15.916	5.707	43.935	23.550	27.011	4.244
	1.2	23.991	15.355	6.393	4.213	20.477	15.180	8.445	3.795	19.801	14.016	7.162	3.434
	1.5	15.315	13.351	4.689	3.593	15.274	12.114	4.426	3.228	14.634	11.134	6.734	2.975
	1.8	12.811	11.739	3.189	3.106	11.655	10.466	2.904	2.884	9.899	9.875	2.664	2.652
	2.1	11.306	11.293	2.996	2.929	9.956	9.926	2.614	2.608	8.995	8.988	2.362	2.353
	2.4	10.645	10.561	2.833	2.795	9.248	9.253	2.542	2.533	8.522	8.470	2.272	2.256
	2.7	**9.772**	**9.695**	**2.731**	**2.701**	**8.950**	**8.961**	**2.454**	**2.413**	**8.279**	**8.243**	**2.170**	**2.163**

Tableau 4.3 – Average integrated square bias (AISB $\times 10^4$) estimated from 1 000 replicates of for $\mathbb{C}_{xh}^{(\mathrm{rc})}$ with $n = 250$ and $n = 1\ 000$ in the case $\tau_{C_x} < \tau_{F_x}$. Upper pannel : Normal Copula. Bottom pannel : Clayton Copula.

θ	h	$\mathfrak{T}_x = .1$				$\mathfrak{T}_x = .25$				$\mathfrak{T}_x = .35$			
		$n = 250$		$n = 1000$		$n = 250$		$n = 1000$		$n = 250$		$n = 1000$	
		$g = h$	$g \neq h$	$g = h$	$g \neq h$	$g = h$	$g \neq h$	$g = h$	$g \neq h$	$g = h$	$g \neq h$	$g = h$	$g \neq h$
0%	0.9	0.230	0.230	**0.005**	**0.005**	0.228	0.228	**0.013**	**0.013**	0.284	0.284	0.032	**0.032**
	1.2	0.089	0.089	0.013	0.013	0.116	0.116	0.035	0.035	0.198	0.198	0.059	0.059
	1.5	0.047	0.047	0.043	0.043	**0.073**	**0.073**	0.087	0.087	**0.158**	**0.158**	0.117	0.117
	1.8	0.030	0.030	0.075	0.075	0.083	0.083	0.143	0.143	0.167	0.167	0.183	0.183
	2.1	**0.029**	**0.029**	0.126	0.126	0.105	0.105	0.209	0.209	0.200	0.200	0.251	0.251
	2.4	0.046	0.046	0.148	0.148	0.130	0.130	0.254	0.254	0.235	0.235	0.300	0.300
	2.7	0.082	0.082	0.189	0.189	0.195	0.195	0.314	0.314	0.301	0.301	0.358	0.358
20%	0.9	1.454	0.729	0.608	0.036	1.437	0.696	0.644	0.039	1.543	0.761	0.739	0.065
	1.2	0.464	0.184	0.091	**0.010**	0.466	0.192	0.085	**0.027**	0.584	0.276	0.138	**0.057**
	1.5	0.150	0.125	**0.022**	0.030	0.147	0.126	**0.058**	0.072	0.245	0.221	**0.103**	0.114
	1.8	0.061	0.060	0.049	0.051	**0.093**	**0.092**	0.108	0.113	**0.197**	**0.194**	0.168	0.172
	2.1	**0.049**	**0.047**	0.086	0.087	0.103	0.100	0.159	0.163	0.228	0.224	0.225	0.229
	2.4	0.059	0.058	0.097	0.098	0.121	0.120	0.190	0.194	0.264	0.261	0.269	0.272
	2.7	0.072	0.073	0.121	0.122	0.167	0.169	0.233	0.238	0.305	0.306	0.320	0.324
40%	0.9	7.057	3.481	6.590	0.773	6.437	2.904	6.698	0.605	6.289	2.718	6.979	0.551
	1.2	4.061	2.231	1.035	0.642	3.590	1.823	0.909	**0.515**	3.547	1.744	0.931	**0.491**
	1.5	2.029	1.718	0.798	**0.597**	1.608	1.339	0.740	0.526	1.548	1.284	0.824	0.559
	1.8	1.829	1.701	**0.603**	0.615	1.461	1.357	**0.596**	0.574	1.445	1.343	**0.722**	0.669
	2.1	1.630	1.539	0.605	0.600	1.357	1.272	0.641	0.610	1.439	1.347	0.799	0.743
	2.4	1.689	1.563	0.642	0.611	1.374	1.260	0.712	0.658	1.448	**1.332**	0.933	0.852
	2.7	**1.538**	**1.438**	0.643	0.626	**1.321**	**1.237**	0.741	0.701	**1.440**	1.354	0.999	0.929
0%	0.9	0.212	0.212	**0.008**	**0.008**	0.215	0.215	**0.035**	**0.035**	0.298	0.298	**0.088**	**0.088**
	1.2	0.065	0.065	0.022	0.022	**0.115**	**0.115**	0.096	0.096	**0.263**	**0.263**	0.210	0.210
	1.5	**0.036**	**0.036**	0.070	0.070	0.133	0.133	0.207	0.207	0.335	0.335	0.374	0.374
	1.8	0.049	0.049	0.142	0.142	0.215	0.215	0.345	0.345	0.464	0.464	0.576	0.576
	2.1	0.064	0.064	0.180	0.180	0.297	0.297	0.429	0.429	0.617	0.617	0.697	0.697
	2.4	0.114	0.114	0.226	0.226	0.392	0.392	0.551	0.551	0.744	0.744	0.867	0.867
	2.7	0.130	0.130	0.305	0.305	0.468	0.468	0.674	0.674	0.893	0.893	1.024	1.024
20%	0.9	1.610	0.533	0.354	0.047	1.841	0.452	0.632	**0.040**	2.768	0.560	0.364	**0.087**
	1.2	0.421	0.160	0.108	**0.016**	0.255	0.181	**0.062**	0.078	0.466	**0.329**	**0.166**	0.191
	1.5	0.129	0.090	**0.049**	0.050	**0.162**	**0.144**	0.169	0.173	**0.343**	**0.329**	0.301	0.346
	1.8	0.081	**0.065**	0.107	0.109	0.180	0.187	0.280	0.287	0.424	0.425	0.510	0.522
	2.1	**0.070**	0.068	0.127	0.128	0.252	0.252	0.342	0.349	0.566	0.570	0.611	0.623
	2.4	0.097	0.096	0.157	0.158	0.316	0.319	0.442	0.451	0.663	0.665	0.765	0.778
	2.7	0.107	0.107	0.222	0.224	0.380	0.382	0.546	0.556	0.798	0.802	0.903	0.916
40%	0.9	8.009	4.098	5.463	1.269	8.886	3.196	4.665	0.943	8.132	2.789	4.685	0.759
	1.2	3.588	2.221	1.990	0.706	3.255	1.710	1.865	**0.691**	2.935	1.569	1.852	**0.690**
	1.5	2.634	1.901	0.875	**0.671**	2.095	1.520	0.840	0.684	1.720	**1.367**	**0.976**	0.805
	1.8	1.888	1.535	**0.750**	0.723	1.517	1.384	**0.869**	0.824	**1.541**	1.406	1.038	0.989
	2.1	**1.444**	**1.332**	0.760	0.749	**1.341**	**1.277**	0.885	0.849	1.522	1.440	1.151	1.108
	2.4	1.649	1.531	0.756	0.752	1.485	1.400	0.998	0.974	1.634	1.548	1.288	1.250
	2.7	1.776	1.674	0.840	0.812	1.649	1.571	1.038	1.014	1.872	1.791	1.403	1.357

Tableau 4.4 – Average integrated variance (AIV $\times 10^4$) estimated from 1 000 replicates of for $\mathbb{C}_{xh}^{(\mathrm{rc})}$ with $n = 250$ and $n = 1\ 000$ in the case $\tau_{C_x} < \tau_{F_x}$. Upper pannel : Normal Copula. Bottom pannel : Clayton Copula.

θ	h	$\mathfrak{T}_x = .1$				$\mathfrak{T}_x = .25$				$\mathfrak{T}_x = .35$			
		$n = 250$		$n = 1000$		$n = 250$		$n = 1000$		$n = 250$		$n = 1000$	
		$g = h$	$g \neq h$	$g = h$	$g \neq h$	$g = h$	$g \neq h$	$g = h$	$g \neq h$	$g = h$	$g \neq h$	$g = h$	$g \neq h$
0%	0.9	3.581	3.581	0.849	0.849	3.164	3.164	0.750	0.750	2.771	2.771	0.649	0.649
	1.2	2.703	2.703	0.691	0.691	2.387	2.387	0.589	0.589	2.078	2.078	0.500	0.500
	1.5	2.442	2.442	0.583	0.583	2.109	2.109	0.509	0.509	1.790	1.790	0.436	0.436
	1.8	2.124	2.124	0.511	0.511	1.839	1.839	0.434	0.434	1.558	1.558	0.369	0.369
	2.1	1.933	1.933	0.467	0.467	1.659	1.659	0.398	0.398	1.394	1.394	0.334	0.334
	2.4	1.844	1.844	0.455	0.455	1.565	1.565	0.388	0.388	1.337	1.337	0.327	0.327
	2.7	**1.820**	**1.820**	**0.423**	**0.423**	**1.542**	**1.542**	**0.354**	**0.354**	**1.282**	**1.282**	**0.300**	**0.300**
20%	0.9	24.896	13.329	24.072	4.462	26.604	13.807	26.315	4.695	27.605	13.997	27.771	4.815
	1.2	12.322	3.652	10.927	0.918	12.878	3.294	11.853	0.800	13.125	2.950	12.432	0.700
	1.5	5.445	4.353	1.911	0.787	5.294	4.083	1.946	0.704	5.099	3.811	1.936	0.620
	1.8	2.832	2.831	0.682	0.685	2.498	2.500	0.590	0.594	2.197	2.200	0.520	0.524
	2.1	2.631	2.641	0.628	0.630	2.316	2.326	0.553	0.555	2.013	2.021	0.481	0.484
	2.4	2.438	2.436	0.599	0.601	2.125	2.126	0.526	0.528	1.855	1.858	0.457	0.460
	2.7	**2.390**	**2.391**	**0.574**	**0.576**	**2.055**	**2.057**	**0.495**	**0.498**	**1.765**	**1.767**	**0.431**	**0.433**
40%	0.9	44.135	19.743	63.494	4.405	47.572	20.341	69.854	4.422	49.689	20.560	74.010	4.372
	1.2	26.871	9.619	13.628	1.718	28.587	9.369	14.703	1.524	29.528	9.043	15.368	1.364
	1.5	8.899	6.801	9.151	1.501	8.618	6.285	9.786	1.333	8.284	5.806	10.156	1.188
	1.8	5.181	5.136	1.325	1.293	4.574	4.536	1.142	1.123	4.068	4.038	1.010	1.000
	2.1	4.700	4.698	1.278	1.252	4.118	4.120	1.105	1.088	3.630	3.641	0.963	0.953
	2.4	4.587	4.577	1.189	1.173	3.998	3.999	1.017	1.013	3.528	3.533	0.876	0.876
	2.7	**4.440**	**4.411**	**1.178**	**1.156**	**3.860**	**3.842**	**0.998**	**0.986**	**3.378**	**3.366**	**0.856**	**0.851**
0%	0.9	3.605	3.605	0.884	0.884	3.197	3.197	0.769	0.769	2.773	2.773	0.658	0.658
	1.2	2.790	2.790	0.708	0.708	2.417	2.417	0.611	0.611	2.084	2.084	0.523	0.523
	1.5	2.332	2.332	0.577	0.577	1.996	1.996	0.496	0.496	1.717	1.717	0.425	0.425
	1.8	2.154	2.154	0.513	0.513	1.827	1.827	0.438	0.438	1.550	1.550	0.372	0.372
	2.1	1.926	1.926	0.467	0.467	1.626	1.626	0.398	0.398	1.382	1.382	0.336	0.336
	2.4	1.798	1.798	0.446	0.446	1.476	1.476	0.370	0.370	1.246	1.246	0.311	0.311
	2.7	**1.764**	**1.764**	**0.419**	**0.419**	**1.463**	**1.463**	**0.354**	**0.354**	**1.244**	**1.244**	**0.301**	**0.301**
20%	0.9	26.717	9.075	17.445	4.482	31.666	7.874	26.904	3.450	42.212	8.880	19.740	3.477
	1.2	12.285	3.736	11.903	0.937	8.020	4.481	6.845	0.823	11.665	5.456	7.110	0.731
	1.5	5.279	3.113	0.774	0.775	5.148	2.772	0.682	0.683	5.023	2.496	3.182	0.602
	1.8	5.078	2.892	0.688	0.690	3.727	2.527	0.613	0.615	2.192	2.203	0.535	0.537
	2.1	2.557	2.559	0.626	0.629	2.263	2.268	0.545	0.547	1.973	1.975	0.475	0.478
	2.4	2.430	2.432	0.598	0.598	2.073	2.072	0.516	0.516	1.805	1.807	0.445	0.447
	2.7	**2.344**	**2.347**	**0.568**	**0.571**	**1.997**	**2.003**	**0.496**	**0.497**	**1.739**	**1.743**	**0.433**	**0.434**
40%	0.9	47.333	22.684	52.828	10.644	56.433	19.019	53.310	8.996	55.804	15.543	56.266	5.505
	1.2	21.724	7.633	27.012	1.795	25.893	6.950	29.195	2.762	25.443	6.473	29.539	1.450
	1.5	13.852	5.634	9.070	1.513	12.944	4.986	9.597	1.323	10.548	4.486	9.952	1.181
	1.8	9.272	5.075	2.425	1.300	5.661	4.508	3.542	1.148	5.270	4.026	2.299	1.029
	2.1	4.781	4.750	2.330	1.219	4.180	4.170	1.096	1.085	3.714	3.688	0.971	0.964
	2.4	4.441	4.421	1.207	1.188	3.934	3.912	1.029	1.016	3.564	3.569	0.887	0.879
	2.7	**4.291**	**4.294**	**1.154**	**1.136**	**3.681**	**3.683**	**0.978**	**0.971**	**3.246**	**3.235**	**0.854**	**0.850**

CHAPITRE 5

Tests de concordance

5.1 Introduction

Au chapitre 1, l'étude de l'influence du niveau de santé d'un pays, tel que décrit par le logarithme du tau de mortalité infantile (X), sur la relation existant entre l'espérance vie des hommes (Y_1) et celle des femmes (Y_2) a illustré l'importance de tenir compte de l'effet que peut avoir une co-variable sur la distribution d'un vecteur aléatoire. Par exemple, la figure 1.3 suggère que l'association entre les variables Y_1 et Y_2 est moindre lorsque le tau de mortalité infantile se situe autour de 15 que lorsque ce dernier est trois fois plus élevé. Toutefois, la validation d'une telle analyse se devrait d'être établie dans le cadre plus formel d'un test d'hypothèses.

Pour déterminer si les composantes d'un vecteur aléatoire (Y_1, Y_2) sont plus ou moins dépendantes selon la valeur que prend la co-variable X, il est d'usage de recourir à la notion d'*ordre de concordance*, définie en terme des copules conditionnelles C_x et C_y décrivant la dépendance du vecteur (Y_1, Y_2) lorsque $X = x$ et $X = y$ respectivement.

Rappelons brièvement que C_x et C_y sont dites concordantes, noté $C_x \preceq C_y$, lorsque

$$C_x(u,v) \leq C_y(u,v) \quad \text{pour tout} \quad (u,v) \in [0,1]^2.$$

Dans ce cas, nous disons que les composantes du couple (Y_1, Y_2) sont moins dépendantes lorsque $X = x$ que lorsque $X = y$.

Tel que mentionné dans l'introduction, lorsque C_x et C_y sont concordantes, il s'ensuit que leur tau de Kendall associé, donnés par les équations

$$\tau_x = 4 \int C_x(u,v) \mathrm{d}C_x(u,v) - 1 \qquad \text{et} \qquad \tau_y = 4 \int C_y(u,v) \mathrm{d}C_y(u,v) - 1$$

satisfont $\tau_x \leq \tau_y$. Ainsi, les méthodes de ré-échantillonnage introduites au chapitre 1 permettent d'ores et déjà d'entrevoir une stratégie permettant de tester l'hypothèse de concordance entre C_x et C_y. En effet, il s'agirait de vérifier si la borne supérieure d'un intervalle de confiance pour τ_x dépasse la borne inférieure d'un intervalle de confiance pour τ_y. Or, un tel test ne sera pas toujours universellement convergent, puisque, tel qu'illustré dans l'introduction, il est possible d'avoir à la fois $\tau_x \leq \tau_y$ et $C_x \npreceq C_y$.

Ce chapitre procure les outils théoriques permettant de tester formellement l'hypothèse de concordance entre deux copules conditionnelles. Des tests universellement convergents suivant lesquels l'hypothèse alternative exprime la négation de l'hypothèse nulle seront présentés. Des tests basés sur des mesures de dépendance seront aussi proposés. Puisque la concordance ne requiert aucune hypothèse paramétrique sur la copule, les méthodes présentées seront non-paramétriques. Les hypothèses requises afin d'assurer la validité des résultats théoriques présentés dans ce chapitre sont placées en annexe.

5.2 Tests universellement convergents

5.2.1 Formulation du test d'hypothèses

Lorsque deux copules conditionnelles C_x et C_y sont concordantes, leur différence $\mathcal{J}_{xy}(u,v) = C_x(u,v) - C_y(u,v)$ est une fonction uniformément négative. À l'opposé, si $C_x \npreceq C_y$, il existe au moins un couple $(u_0, v_0) \in [0,1]^2$ pour lequel $\mathcal{J}_{xy}(u_0, v_0) > 0$. Un test d'hypothèses universel pour vérifier la concordance entre C_x et C_y s'exprime donc sous la forme

$$\mathcal{H}_0 : \mathcal{J}_{xy} \leq 0 \qquad \text{v.s} \qquad \mathcal{H}_1 : \mathcal{J}_{xy} > 0.$$

Considérons une fonctionnelle croissante $\Psi : l^\infty([0,1]^2) \mapsto \mathbb{R}$ satisfaisant $\Psi(a\delta) = a\,\Psi(\delta)$ pour tout $a \in \mathbb{R}^+$ et $\delta \in l^\infty([0,1]^2)$. Introduisons aussi la fonctionnelle $\Psi^+(g) = \Psi(g \vee 0)$, où $(g \vee 0)(u,v) = g(u,v)\mathbb{I}\{g(u,v) > 0\}$. Remarquons que, lorsque C_x et C_y sont continues, les hypothèses nulles et alternatives sont respectivement équivalentes à $\Psi^+(J_{xy}) = 0$ et $\Psi^+(J_{xy}) > 0$. Une fonctionnelle possédant ces propriétés s'avère donc un critère adéquat pour discerner $\mathcal{H}_0$ de $\mathcal{H}_1$. On vérifie aiséement que tel est le cas pour la fonctionnelle de Cramer von mises Ψ^{c} et celle de Kolmogorov Ψ^{k}, se définissant comme :

$$\Psi^{\mathrm{c}}(\delta) = \sqrt{\int_{[0,1]^2} \delta(u,v)^2 \mathrm{d}u\mathrm{d}v} \quad \text{et} \qquad \Psi^{\mathrm{k}}(\delta) = \sup_{[0,1]^2} |\delta(u,v)|.$$

5.2.2 Statistiques de test

Puisque les hypothèses $\mathcal{H}_0$ et $\mathcal{H}_1$ s'appuient sur la fonction $\mathcal{J}_{xy} = C_x - C_y$, il apparaît naturel de définir une statistique de test à l'aide de versions non paramétriques de C_x et de C_y. À ce sujet, deux façon d'estimer une copule conditionnelle en présence d'observations i.i.d $(Y_{11}, Y_{21}, X_1), \ldots, (Y_{1n}, Y_{2n}, X_n)$ ont été proposées par Gijbels *et al.* (2011). Ces méthodes sont décrites dans le chapitre d'introduction de cette thèse. Ainsi, deux estimateurs de $\mathcal{J}_{xy}$ sont obtenus en considérant

$$\mathcal{J}_{xyh}(u,v) = C_{xh}(u,v) - C_{yh}(u,v) \quad \text{et} \quad \widetilde{\mathcal{J}}_{xyh}(u,v) = \widetilde{C}_{xh}(u,v) - \widetilde{C}_{yh}(u,v),$$

où les estimateurs C_{xh} et C_{yh} sont décrits par l'équation (7) alors que $\widetilde{C}_{xh}$ et $\widetilde{C}_{yh}$ s'obtiennent par l'entremise de l'équation (8). Dans ce chapitre, on émet l'hypothèse que les paramètres de lissage h, h_1 et h_2 satisfont

$$n\min(h,h_1,h_2) \to \infty, \quad n\max(h^5,h_1^5,h_2^5) < \infty, \quad \text{et} \quad h/\min(h_1,h_2) < \infty.$$

Pour détecter une violation de $\mathcal{H}_0$, il suffit donc de rejeter cette dernière suivant la réalisation de grandes valeurs de $\Psi^+(\mathcal{J}_{xyh})$ ou $\Psi^+(\widetilde{\mathcal{J}}_{xyh})$. Dans ce qui suit, pour alléger la lecture, on notera simplement $\bar{C}_{xh}$ et $\bar{C}_{yh}$ pour désigner soit C_{xh} et C_{yh}, ou $\widetilde{C}_{xh}$ et C_{yh}. Les fonctions reliées à ces estimateurs seront à leur tour munies d'une barre.

Pour déterminer le seuil de rejet de $\mathcal{H}_0$, un examen sommaire des propriétés asymptotiques de la variable aléatoire $\bar{\mathbb{S}}^{\Psi}_{xyh} = \sqrt{nh}\Psi^+(\bar{\mathcal{J}}_{xyh})$ est requis. En premier lieu, remarquons que, pour $\bar{\mathbb{J}}_{xyh} = \sqrt{nh}(\bar{\mathcal{J}}_{xyh} - \mathcal{J}_{xy})$, on a

$$\mathrm{P}\left(\bar{\mathbb{S}}^{\Psi}_{xyh} > Q\right) = \mathrm{P}\left[\Psi^+\left\{\bar{\mathbb{J}}_{xyh} + \sqrt{nh}\mathcal{J}_{xy}\right\} > Q\right].$$

En second lieu, lorsque l'hypothèse nulle est vraie, il s'ensuit que $\mathcal{J}_{xy} \leq 0$. Alors, en vertu de la propriété de croissance de Ψ,

$$\mathrm{P}\left[\Psi^+\left\{\bar{\mathbb{J}}_{xyh} + \sqrt{nh}\mathcal{J}_{xy}\right\} > Q\right] \leq \mathrm{P}\left\{\Psi^+\left(\bar{\mathbb{J}}_{xyh}\right) > Q\right\}.$$

Suivant une valeur z incluse dans le support de la co-variable X, on obtient de l'article Veraverbeke *et al.* (2011) que le processus de copule conditionnelle $\bar{\mathbb{C}}_{zh}$ converge vers le processus $\bar{\mathbb{C}}_z = \alpha_z + \bar{B}_z$, où α_z dénote un processus gaussien centré et $\bar{B}_z$ désigne son biais asymptotique (il se trouve que seul le biais asymptotique change suivant les choix $\bar{C}_{zh} = C_{zh}$ et $\bar{C}_{zh} = \widetilde{C}_{zh}$). Les expressions de la fonction de covariance de α_z et du biais $\bar{B}_z$ peuvent être trouvées dans l'article Veraverbeke *et al.* (2011), ou, de façon équivalente, dans la formulation des propositions 2 et 3 du chapitre 2. Ensuite, pour un choix arbitraire de $x \neq y$, il est mentionné dans l'article Veraverbeke *et al.* (2011) que les processus de copule conditionnelle $\bar{\mathbb{C}}_{xh}$ et $\bar{\mathbb{C}}_{yh}$ sont asymptotiquement indépendants. Puisque la différence de deux processus gaussiens indépendents est aussi gaussienne, cette dernière discussion permet de conclure que le processus $\bar{\mathbb{J}}_{xyh}$ converge vers le processus $\bar{\mathbb{J}}_{xy} = \bar{\mathbb{C}}_x - \bar{\mathbb{C}}_y$. Ainsi, en vertu du théorème des applications continues,

$$\lim_{n\to\infty} \mathrm{P}\left\{\Psi^+\left(\bar{\mathbb{J}}_{xyh}\right) > Q\right\} = \mathrm{P}\left\{\Psi^+\left(\bar{\mathbb{J}}_{xy}\right) > Q\right\}.$$

En désignant le quantile d'ordre $1-\alpha$ de la distribution de la variable aléatoire $\Psi^+\left(\bar{\mathbb{J}}_{xy}\right)$ par $\bar{Q}^{\psi}_{xy}(\alpha)$, un test asymptotiquement de niveau α pour $\mathcal{H}_0$ consisterait à en rejeter cette dernière lorsque $\Psi^+\left(\bar{\mathbb{J}}_{xyh}\right) > \bar{Q}^{\psi}_{xy}(\alpha)$. Or, le quantile $\bar{Q}^{\psi}_{xy}(\alpha)$ étant généralement inconnu, la section suivante présente une méthode pour résoudre ce problème.

5.2.3 Ré-échantillonnage sous $\mathcal{H}_0$

Il est possible d'estimer le quantile $\bar{Q}^{\psi}_{xy}(\alpha)$ suivant la méthode de ré-échantillonnage dite « du multiplicateur conditionnel» (*covariate dependent multiplier*) décrite dans le chapitre 1. Rappelons brièvement que, dans ce chapitre, il est proposé de ré-échantillonner l'estimateur C_{zh} à partir de multiplicateurs $(\xi_{1z}, \ldots, \xi_{nz})$ définis comme étant des variables aléatoires positives dont la moyenne et la variance sont décrites par les expressions :

$$\mathrm{E}^\star(\xi_{iz}) = p_{ix} = \sum_{j=1}^n w_{ni}(X_j, g) w_{nj}(z,h) \quad \text{et} \quad \mathrm{Var}^\star(\xi_{iz}) = v_{ix} = \sum_{j=1}^n w_{ni}(X_j, g) w_{nj}(z,h)^2.$$

Dans la dernière équation, $g = g_n$ désigne un autre paramètre de lissage satisfaisant $n \to \infty$ $g_n \to 0$ and $n^{1-\delta} g_n^5 \to \infty$ pour un certain $\delta > 0$. De plus, ci-après, on notera $\mathrm{P}^\star$ la mesure de probabilité bootstrap, et $\mathrm{E}^\star$ et $\mathrm{Var}^\star$ l'espérance et la variance calculées suivant $\mathrm{P}^\star$. Enfin, pour $\xi_{\cdot z} = \xi_{1z} + \ldots + \xi_{nz}$, on considère

$$\alpha^\star_{zh}(u,v) = \sqrt{nh} \sum_{i=1}^n \{\xi_{iz} - \xi_{\cdot z} w_{ni}(z,g)\} \mathbb{I}\left\{Y_{1i} \leq F^{-1}_{1zh}(u), Y_{2i} \leq F^{-1}_{2zh}(v)\right\}.$$

Alors, une version multiplicateur du processus $\mathbb{C}_{zh}$ est obtenue via

$$\mathbb{C}^{\star}_{zh}(u,v) = \alpha^{\star}_{zh}(u,v) - \widehat{C}^{[1]}_{z}(u,v)\alpha^{\star}_{zh}(u,1) - \widehat{C}^{[2]}_{z}(u,v)\alpha^{\star}_{zh}(1,v),$$

où $\widehat{C}^{[1]}_{z}$ et $\widehat{C}^{[2]}_{z}$ désignent des estimateurs des dérivées partielles de la copule C_z.

Il est toutefois impossible d'obtenir directement des versions multiplicateur du processus $\widetilde{\mathbb{C}}_{zh}$ à partir des méthodes décrites dans le chapitre 1, puisque ce dernier est composé des *pseudo-observations uniformisées* $(\widetilde{U}_{1i}, \widetilde{U}_{2i})$. Néanmoins, il s'avère que seule une modification mineure est nécessaire. À cette fin, on introduit deux paramètres de lissage $g_1 = g_{1n}$ et $g_2 = g_{2n}$ satisfaisant, lorsque $n \to \infty$, et pour un certain choix de $\delta_1, \eta > 0$:

$$\frac{h_n \log^{1+\delta_1} n}{g_{jn}} \to 0 \qquad \text{et} \qquad n h_n g_{jn}^{4+\eta} \to \ , \; j = 1, 2.$$

En considérant une version des *pseudo-observations uniformisées* $(\widetilde{U}^b_{1i}, \widetilde{U}^b_{2i})$ calculées en remplaçant les paramètres de lissages h_1 et h_2 par g_1 et g_2, on définit

$$\widetilde{\alpha}^{\star}_{zh}(u,v) = \sqrt{nh} \sum_{i=1}^{n} \{\xi_{iz} - \xi_{\cdot z} w_{ni}(z,g)\} \mathbb{I}\left(\widetilde{U}^b_{1i} \leq u, \widetilde{U}^b_{2i} \leq v\right).$$

Alors, une version multiplicateur du processus $\widetilde{\mathbb{C}}_{zh}$ est obtenue de la façon suivante :

$$\widetilde{\mathbb{C}}^{\star}_{zh}(u,v) = \widetilde{\alpha}^{\star}_{zh}(u,v) - \widehat{C}^{[1]}_{z}(u,v)\widetilde{\alpha}^{\star}_{zh}(u,1) - \widehat{C}^{[2]}_{z}(u,v)\widetilde{\alpha}^{\star}_{zh}(1,v).$$

Afin d'assurer la validité théorique de cette approche, nous avons montré la proposition suivante.

Proposition 1 *Supposons que les hypothèses W_1–W_8, W'_1–W'_3 et W''_1–W''_6, $\mathcal{A}_1$, $\mathcal{A}^{\star}_2$ et $\mathcal{A}_3$ soient satisfaites. Supposons également que les multiplicateurs vérifient la condition*

$$\sqrt{nh_n} \max_{1 \leq i \leq n} \xi_{iz} = o_{\mathrm{P}^{\star}}(1).$$

Alors, $(\widetilde{\mathbb{C}}_{zh}, \widetilde{\mathbb{C}}^{\star}_{zh})$ convergent faiblement dans $l^{\infty}([0,1]^2)$ suivant la mesure de probabilité $\mathrm{P}^{\star}$ [P]*-presque surement vers $(\widetilde{\mathbb{C}}_z, \widetilde{\mathbb{C}}^{\star}_z)$, où $\widetilde{\mathbb{C}}^{\star}_z$ est une copie indépendante de $\widetilde{\mathbb{C}}_z$.*

D'une certaine manière, la Proposition 1 étend au second estimateur de la copule conditionnelle les résultats obtenus dans le chapitre 1 à la section 3.5.

Forts de ces méthodes de ré-échantillonnage pour les estimateurs C_{zh} et $\widetilde{C}_{zh}$, nous présentons dans ce qui suit une procédure pour en déduire une approximation du seuil critique de rejet de $\mathcal{H}_0$. Le processus obtenu par la méthode du multiplicateur associée au processus $\bar{\mathbb{C}}_{zh}$ sera notée $\bar{\mathbb{C}}^{\star}_{zh}$.

Remarquons que $\bar{Q}^{\psi}_{xy}(\alpha)$ est défini en fonction de la distribution limite de la variable aléatoire

$$\bar{\mathbb{J}}_{xyh} = \mathbb{C}_{xh} - \mathbb{C}_{yh} = \sqrt{nh}(\bar{C}_{xh} - C_x) - \sqrt{nh}(\bar{C}_{yh} - C_y).$$

Une approximation multiplicateur de $\bar{\mathbb{J}}_{xyh}$ peut être déduite de la discussion précédente, en considérant simplement $\bar{\mathbb{J}}^{\star}_{xyh} = \bar{\mathbb{C}}^{\star}_{xh} - \bar{\mathbb{C}}^{\star}_{yh}$. La validité d'une telle approche est assurée par la section 3.5 du chapitre 1 ainsi que par la Proposition 1. Alors, en se dotant d'un échantillon bootstrap $\bar{\mathbb{J}}^{\star(1)}_{xyh}, \ldots, \bar{\mathbb{J}}^{\star(B)}_{xyh}$, on définit une version multiplicateur de $\bar{Q}^{\psi}_{xy}(\alpha)$ en prenant le quantile empirique d'ordre $1-\alpha$ de l'ensemble $\{\Psi^+(\bar{\mathbb{J}}^{\star(1)}_{xyh}), \ldots, \Psi^+(\bar{\mathbb{J}}^{\star(B)}_{xyh})\}$.

5.3 Tests basés sur des fonctionnelles statistiques

5.3.1 Formulation du test d'hypothèses

Dans certains cas, il s'avère plus facile de détecter une déviation de l'hypothèse nulle en utilisant un test non universellement convergent basé sur des fonctionnelles statistiques. Afin de mettre en oeuvre de tels tests, on considère $\Lambda : l^{\infty}([0,1]^2) \to \mathbb{R}$ une fonctionnelle préservant d'ordre de concordance, *i.e* pour tout $\delta, \delta' \in l^{\infty}([0,1]^2)$, $\delta \preceq \delta'$ implique $\Lambda(\delta) \leq \Lambda(\delta')$. Notamment, cette propriété est satisfaite pour la fonctionnelle de Spearman

Λ_ρ et celle de Kendall Λ_τ définies, pour $\delta \in \ell^\infty([0,1]^2)$, par :

$$\Lambda_\rho(\delta) = 12 \int_{[0,1]^2} \delta(u,v)\, \mathrm{d}u\mathrm{d}v - 3 \quad \text{and} \quad \Lambda_\tau(\delta) = 4 \int_{[0,1]^2} \delta(u,v)\, \mathrm{d}\delta(u,v) - 1.$$

À la lumière de cette propriété, il semble raisonnable de vérifier la concordance de C_x et C_y en utilisant le critère $T_{xy}^\Lambda = \Lambda(C_x) - \Lambda(C_y)$ par l'entremise du test d'hypothèses

$$\mathcal{H}_0^\Lambda : T_{xy}^\Lambda \leq 0 \qquad \text{v.s} \qquad \mathcal{H}_1^\Lambda : T_{xy}^\Lambda > 0.$$

Cette approche est justifiée du fait que la réalisation de la contre hypothèse $\mathcal{H}_1^\Lambda$ implique $C_x \not\preceq C_y$. Or, bien que la concordance de C_x et C_y entraîne la véracité de $\mathcal{H}_0^\Lambda$, il est possible d'observer à la fois $C_x \not\preceq C_y$ et $\Lambda(C_x) \leq \Lambda(C_y)$. Une telle situation est d'ailleurs détaillée dans l'introduction.

5.3.2 Statistiques de test

Puisque T_{xy}^Λ s'écrit en fonction de C_x et C_y, un estimateur non paramétrique est obtenu en y substituant les copules conditionnelles par leur version empirique, nommément $\bar{C}_{xh}$ et $\bar{C}_{yh}$. L'estimateur résultant se présente alors sous la forme :

$$\bar{T}_{xyh}^\Lambda = \Lambda(\bar{C}_{xh}) - \Lambda(\bar{C}_{yh}).$$

Un test pour $\mathcal{H}_0^\Lambda$ sera donc établi en rejetant cette dernière pour de grandes valeurs de $\bar{T}_{xyh}^\Lambda$. De manière similaire à la section précédente, nous déterminons la valeur critique du test en étudiant le comportement asymptotique des variables aléatoires $\bar{\mathbb{U}}_{xyh}^\Lambda = \sqrt{nh}\bar{T}_{xyh}^\Lambda$ et $\bar{\mathbb{T}}_{xyh}^\Lambda = \sqrt{nh}(\bar{T}_{xyh}^\Lambda - T_{xy}^\Lambda)$. D'abord,

$$\mathrm{P}\left(\bar{\mathbb{T}}_{xyh}^\Lambda > Q\right) = \mathrm{P}\left(\bar{\mathbb{T}}_{xyh}^\Lambda + \sqrt{nh}T_{xy}^\Lambda > Q\right).$$

Ensuite, en raison du fait que l'hypothèse $\mathcal{H}_0^\Lambda$ entraîne $T_{xy}^\Lambda < 0$, il en résulte que

$$\mathrm{P}\left(\bar{\mathbb{T}}_{xyh}^\Lambda + \sqrt{nh}T_{xy}^\Lambda > Q\right) \leq \mathrm{P}\left(\bar{\mathbb{T}}_{xyh}^\Lambda > Q\right).$$

Il se trouve que la distribution asymptotique de la variable aléatoire $\bar{\mathbb{T}}^{\Lambda}_{xyh}$ peut être décrite à partir de celles des processus $\bar{\mathbb{C}}_{xh}$ et $\bar{\mathbb{C}}_{yh}$ lorsque la fonctionnelle $\delta \mapsto \Lambda(\delta)$ admet une dérivée d'Hadamard en $\delta = C_x$ et $\delta = C_y$. En effet, suivant les développements présentés à la section 3 du chapitre 1, on montre d'abord que, pour un certain z inclus dans le support de la co-variable, $\sqrt{nh}\{\Lambda(\bar{C}_{zh}) - \Lambda(C_z)\}$ converge en loi vers une variable aléatoire gaussienne s'écrivant comme $\Lambda'_{C_z}(\bar{\mathbb{C}}_z)$, où Λ'_{C_z} désigne la dérivée d'Hadamard de Λ en C_z. Pour de plus amples informations concernant la dérivée d'Hadamard, se référer au début de la section 3 du chapitre 1 de cette thèse, ou encore à la section 3.9 du livre van der Vaart & Wellner (1996). Ensuite, puisque les processus $\bar{\mathbb{C}}_{xh}$ et $\bar{\mathbb{C}}_{yh}$ sont asymptotiquement indépendants, il en résulte que $\bar{\mathbb{T}}^{\Lambda}_{xyh}$ converge en loi vers une variable aléatoire normale de moyenne et de variance donnée par les expressions :

$$\bar{\mu}^{\Lambda}_{xy} = \Lambda'_{C_x}\left\{\mathrm{E}\left(\bar{\mathbb{C}}_x\right)\right\} - \Lambda'_{C_y}\left\{\mathrm{E}\left(\bar{\mathbb{C}}_y\right)\right\}, \text{ et } \bar{v}^{\Lambda}_{xy} = \mathrm{Var}\left\{\Lambda'_{C_x}(\bar{\mathbb{C}}_x)\right\} + \mathrm{Var}\left\{\Lambda'_{C_y}(\bar{\mathbb{C}}_y)\right\}. \quad (5.1)$$

En désignant par $\bar{Q}^{\Lambda}_{xy}(\alpha)$ le quantile d'ordre $1-\alpha$ d'une distribution normale de moyenne $\bar{\mu}^{\Lambda}_{xy}$ et de variance $\bar{v}^{\Lambda}_{xy}$, un test asymptotiquement de niveau α pour $\mathcal{H}^{\Lambda}_0$ consisterait à en rejeter $\mathcal{H}^{\Lambda}_0$ lorsque $\bar{\mathbb{T}}_{xyh}$ dépasse le seuil critique $\bar{Q}^{\Lambda}_{xy}(\alpha)$.

À nouveau, puisque ce seuil critique repose sur des quantités généralement inconnues en pratique, nous faisons appel à la technique de ré-échantillonnage du multiplicateur conditionnel.

5.3.3 Ré-échantillonnage sous $\mathcal{H}^{\Lambda}_0$

Suivant la méthodologie décrite dans la section 3 du chapitre 1 de cette thèse, nous définissons une version multiplicateur de $\bar{\mathbb{T}}^{\Lambda}_{xyh}$ comme :

$$\bar{\mathbb{T}}^{\Lambda,\star}_{xyh} = \Lambda'_{C_x}\left(\bar{\mathbb{C}}^{\star}_{xh}\right) - \Lambda'_{C_y}\left(\bar{\mathbb{C}}^{\star}_{yh}\right).$$

Afin d'estimer le seuil de rejet à l'aide de B répliques bootstrap $\bar{\mathbb{T}}_{xyh}^{\Lambda,\star,(1)}, \ldots, \bar{\mathbb{T}}_{xyh}^{\Lambda,\star,(B)}$, deux stratégies sont envisageables. La première s'appuie sur la normalité asymptotique de $\bar{\mathbb{T}}_{xyh}^{\Lambda}$. En estimant la moyenne et la variance présentées à l'équation (5.1) par

$$\bar{\mu}_{xyh}^{\Lambda,\star} = \frac{1}{B} \sum_{i=1}^{B} \bar{\mathbb{T}}_{xyh}^{\Lambda,\star,(i)} \qquad \bar{v}_{xyh}^{\Lambda,\star} = \frac{1}{B} \sum_{i=1}^{B} \left(\bar{\mathbb{T}}_{xyh}^{\Lambda,\star,(i)} - \bar{\mu}_{xyh}^{\Lambda,\star} \right)^2,$$

un estimateur de $\bar{Q}_{xy}^{\Lambda}(\alpha)$ est obtenu en calculant

$$\bar{\mu}_{xyh}^{\Lambda,\star} + z_{1-\alpha} \left(v_{xyh}^{\Lambda,\star} \right)^{1/2},$$

où $z_{1-\alpha}$ désigne le quantile d'ordre $1-\alpha$ d'une distribution gaussienne de moyenne 0 et de variance 1. La seconde estime le seuil rejet à l'aide du quantile d'ordre $1-\alpha$ de l'échantillon bootstrap.

Remarque 1 *Il arrive que la dérivée d'Hadamard Λ'_{C_z} dépende de quantités inconnues en pratique. C'est le cas notamment pour la fonctionnelle de Kendall dont la dérivée d'Hadamard en C_z est donnée, pour $\delta \in l^\infty([0,1]^2)$, par l'expression*

$$(\Lambda_\tau)'_{C_z}(\delta) = 4 \int \delta(u,v)\, \mathrm{d}C_z(u,v) + 4 \int C_z(u,v)\, \mathrm{d}\delta(u,v).$$

Or, en vertu d'une remarque formulée à la section 3.2 du chapitre 1, il suffit, pour conserver la validité de la méthode de ré-échantillonnage, de disposer d'un estimateur uniformément consistent $\widehat{\Lambda}'_{C_z}$ (voir la section 3.2 du chapitre 1 pour plus de détails). Pour la fonctionnelle de Kendall, l'estimateur résultant de la substitution de C_z par $\bar{C}_{zh}$ satisfait à ce critère.

5.4 Étude de simulations

5.4.1 Choix des différents paramètres

Pour le choix des multiplicateurs nécessaires au ré-échantillonnage, nous considérons, pour $z \in \{x,y\}$ et $i \in \{1,\ldots,n\}$, la variable aléatoire ξ_{iz} distribuée suivant une loi

Gamma de paramètres $\alpha = p_{iz}^2/v_{iz}$ et $\beta = v_{iz}/p_{iz}$ de telle sorte que $\mathrm{E}(\xi_{iz}) = p_{iz}$ et $\mathrm{Var}(\xi_{iz}) = v_{iz}$. Ce choix satisfait aux conditions assurant la validité théorique de cette méthode, en vertu du Lemme 1 présenté à la section 2.4 du chapitre 1. De plus, suivant les recommandations formulées à la lumière de l'étude de simulation conduite au chapitre 1, nous remplaçons ξ_{iz} par $\xi_{iz}^* = \xi_{iz}/\xi_{\cdot z}$. Tel que mentionné dans cette section, cette modification n'a asymptotiquement aucun impact sur la validité des résultats théoriques en jeu.

Les différents tests statistiques décrits dans ce chapitre requièrent le choix d'un système de poids $w_{n1}, \ldots, w_{nn}$ ainsi que de plusieurs paramètres de lissage. Pour les simulations ici présentées, nous avons considéré le système de poids Local-linéaire combiné avec un noyau triweight, dont les définitions respectives sont présentées dans l'introduction de cette thèse. Puisque les résultats théoriques nécessitent que les poids soient non-négatifs et somment à un, les poids négatifs, s'il s'en trouve, sont mis égaux à 0 et les poids restants sont divisés par la somme totale des poids positifs. Tel que souligné dans l'article Omelka *et al.* (2013), cette opération est asymptotiquement négligeable. En ce qui a trait au paramètre de lissage, nous avons à nouveau suivi à la recommandation formulée au chapitre 1 et ainsi considéré $h = \mathrm{EIQ}(X) \times n^{-1/5}$ et $g = 1.5 \times h \times n^{0.1}$, où $\mathrm{EIQ}(X)$ désigne l'étendue inter-quartile de l'ensemble $\{X_1, \ldots, X_n\}$. Puis, les autres paramètres sont fixés en prenant simplement $h_1 = h_2 = h$, et $g_1 = g_2 = .25\, h \log n$.

Certains de ces tests requièrent aussi des estimateurs pour les dérivées partielles $C_x^{[1]}$ et $C_x^{[2]}$. Dans les simulations ici présentées, nous avons considéré l'estimateur dit « à différence finie » défini par l'équation :

$$\widehat{C}_x^{[1]}(u,v) = \sqrt{nh}\left\{C_{xh}^{Z\to Y}(u_h^\star, v) - C_{xh}^{Z\to Y}(u,v)\right\}$$

où $u_h^\star = u + \min\{1/\sqrt{nh}, 1-u\}$. L'estimateur $\widehat{C}_x^{[2]}(u,v)$ est défini de façon similaire.

Dans ce qui suit, nous comparerons l'habileté de différentes procédures de tests à maintenir la probabilité de recouvrement sous l'hypthèse de concordance entre deux copules conditionnelles, ainsi que leur capacité à rejeter l'hypothèse nulle lorsque la contre-hypothèse est réalisée. Quatre différents tests seront considérés, suivant les fonctionnelles de Cramer von-mises (CVM) , Kolmogorov (K), Spearman (SP) et Kendall (KD). Tel que présenté dans ce chapitre, pour une fonctionnelle donnée, deux types d'estimateurs sont possibles pour la statistique de test associée, soit celui basé sur $C_{\cdot h}$ ou celui s'appuyant sur $\widetilde{C}_{\cdot h}$. Lorsqu'il s'agit du second, on munira l'acronyme correspondant d'un tilde.

En ce qui a trait aux fonctionnelles de Spearman et Kendall, le choix d'un estimateur pour le seuil critique de rejet est aussi requis. Une étude de simulation, qui n'est pas présentée dans cette thèse, suggère que les deux estimateurs décrits à la section 3 de ce chapitre présentent des comportements très similaires, avec une légère préférence pour celui basé sur la normalité asymptotique. Pour cette raison, seuls les résultats pour ce dernier seront ici présentés.

Enfin, dans chacune des deux sections suivantes, les résultats sont calculés depuis 1 000 échantillons i.i.d $(Y_{11}, Y_{21}, X_1), \ldots, (Y_{1n}, Y_{2n}, X_n)$ simulés suivant différents modèles pour des tailles d'échantillons $n \in \{250, 500, 1000\}$.

5.4.2 Comportement des tests sous l'hypothèse nulle de concordance

L'objectif de cette section est de comparer les performances des différents tests quant à leur capacité à maintenir leur probabilité de recouvrement sous l'hypothèse nulle. Dans cette optique, deux scénarios seront exposés. De la même façon qu'à la section 5.3 du chapitre 1, on considère un triplet (Y_1, Y_2, X) de distribution gaussienne centrée ayant pour matrice de corrélation $R = R_{\ell,\ell'} \in \mathbb{R}^{3\times3}$. Il s'ensuit que, conditionnellement à $X = x$ et $X = y$, le vecteur (Y_1, Y_2) suit également une loi normale bi-variée centrée caractérisée par le coefficient de corrélation

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

Ainsi, les copules conditionnelles associées C_x et C_y sont égales, et donc concordante peu importe le choix de $x \neq y$. Trois choix de paramètres ont été pris en compte, soient $(R_{12}, R_{23}, R_{12}) \in \{(0.5, 0.3, 0.3), (\text{-}0.5, 0.3, \text{-}0.3), (0.1, 0.1, 0.1)\}$. On vérifie, à l'aide de la relation donnée à l'équation précédente, et en utilisant la relation liant le paramètre de la copule normale au tau de Kendall décrite par l'équation (12), que les valeurs du tau de Kendall associés à ces choix de paramètres sont respectivement $0.3, -0.3$ et 0.09. Les résultats sont rapportés dans la table 5.1

Pour le deuxième scénario, on considère un vecteur aléatoire (Y_1, Y_2, X) dont les composantes sont distribués selon une loi normale centrée de variance 1 et dont la dépendance est, cette fois, régie par la copule de Clayton tri-variée de paramètre θ. Suivant la section 5.4 du chapitre 1, les copules C_x et C_y extraites du vecteur (Y_1, Y_2) conditionnellement à $X = x$ et $X = y$ sont aussi des copules de Clayton de paramètre $\theta_x = \theta_y = \frac{\theta}{\theta+1}$. À nouveau, on observe l'égalité $C_x = C_y$. Les résultats sont présentés aux tables 5.1 et 5.2 pour $\theta \in \{0.28, 1, 2\}$ qui conduisent, suivant à nouveau l'équation (12), aux valeurs $\tau_x \in \{0.1, 0.2, 0.25\}$. Pour chacun de ces deux scénarios, la concordance sera évaluée pour $x = -0.5$ et $y \in \{0, 1\}$.

De façon générale, les différents tests maintiennent bien leur probabilité de recouvrement, et les résultats s'améliorer lorsque la taille d'échantillon augmente. Une exception survient cependant dans le second scénario, suivant le cas $\theta = 2$ et $y = 1$, pour les tests basés sur les fonctionnelle de Cramer von mises et de Kolmogorov, où le seuil de rejet se situe près de 8% lorsque $n = 1000$.

Tableau 5.1 – Probabilité de recouvrement pour $\alpha = 5\%$. Premier scénario.

	Statistique	$(R_{12}, R_{13}, R_{23}) = (0.5, 0.5, 0.3)$			$(R_{12}, R_{13}, R_{23}) = (-0.5, 0.5, -0.3)$			$(R_{12}, R_{13}, R_{23}) = (0.1, 0.1, 0.1)$		
	de test	$n = 250$	$n = 500$	$n = 1000$	$n = 250$	$n = 500$	$n = 1000$	$n = 250$	$n = 500$	$n = 1000$
$y = 0$	CVM	94.1	93.9	93.6	95.9	94.4	94.7	96.9	96.9	96.1
	$\widetilde{\text{CVM}}$	93.5	94.1	92.8	95.2	93.3	94.6	95.6	95.7	94.9
	K	95.2	95.3	92.4	96.4	94.2	95.4	98.2	97.1	96.9
	$\widetilde{\text{K}}$	95.1	95.2	92.4	96.4	94.4	94.3	97.9	97.2	96.5
	SP	95.1	95.1	94.8	97.0	94.8	94.2	96.9	96.5	94.9
	$\widetilde{\text{SP}}$	94.7	95.0	94.5	95.7	93.4	94.6	95.5	95.5	95.0
	KD	94.2	94.3	94.2	96.1	93.3	93.6	94.9	95.5	94.6
	$\widetilde{\text{KD}}$	94.1	94.6	94.2	94.8	92.8	93.8	95.0	94.7	95.2
$y = 1$	CVM	93.4	94.9	93.8	94.9	93.8	94.3	93.9	95.9	95.7
	$\widetilde{\text{CVM}}$	92.8	94.6	93.5	93.4	93.6	94.2	92.9	94.6	95.0
	K	94.4	93.9	92.2	94.1	94.4	92.0	96.2	95.5	92.4
	$\widetilde{\text{K}}$	93.8	93.1	92.7	94.7	94.4	92.2	95.5	95.6	94.8
	SP	94.5	95.4	93.9	95.6	94.1	92.9	94.4	94.5	92.9
	$\widetilde{\text{SP}}$	94.6	95.9	94.3	94.8	93.6	94.4	93.2	94.5	94.6
	KD	94.7	95.3	94.0	95.2	94.1	93.5	93.5	94.4	93.4
	$\widetilde{\text{KD}}$	94.4	95.1	94.3	95.2	94.1	94.7	93.2	94.1	94.9

Tableau 5.2 – Probabilité de recouvrement pour $\alpha = 5\%$. Second scénario.

	Statistique	$\theta = 0.28$			$\theta = 1$			$\theta = 2$		
	de test	$n = 250$	$n = 500$	$n = 1000$	$n = 250$	$n = 500$	$n = 1000$	$n = 250$	$n = 500$	$n = 1000$
$y = 0$	CVM	96.1	95.7	95.8	96.3	96.0	95.0	93.8	93.5	93.9
	$\widetilde{\text{CVM}}$	95.0	95.9	95.9	95.6	95.8	94.9	92.4	93.6	93.7
	K	97.5	97.6	96.3	96.4	95.9	94.8	94.2	94.7	91.7
	$\widetilde{\text{K}}$	97.2	97.6	96.2	96.2	96.2	95.6	93.2	94.2	92.1
	Sp	95.8	94.2	94.5	95.7	95.1	93.7	94.7	93.7	95.1
	$\widetilde{\text{SP}}$	94.6	94.5	94.4	95.6	94.8	93.6	93.9	94.2	95.0
	KD	94.7	93.7	94.0	95.3	94.1	92.9	94.2	93.8	94.5
	$\widetilde{\text{KD}}$	94.0	93.9	93.9	94.8	94.1	93.0	93.5	93.9	93.7
$y = 1$	CVM	95.2	96.0	95.7	95.4	95.6	95.3	93.8	93.3	92.1
	$\widetilde{\text{CVM}}$	94.3	95.8	95.7	94.5	94.4	95.0	93.7	92.9	91.8
	K	95.4	95.3	95.1	94.3	95.7	94.0	92.0	93.3	91.9
	$\widetilde{\text{K}}$	95.1	95 .3	95.6	94.3	95.9	94.1	92.1	92.9	91.4
	SP	94.4	94.6	93.4	94.5	93.6	93.8	95.2	94.4	93.5
	$\widetilde{\text{SP}}$	94.0	93.9	93.4	94.2	93.2	93.7	94.4	94.5	93.3
	KD	94.7	94.8	93.7	94.1	93.8	94.0	95.5	95.1	93.4
	$\widetilde{\text{KD}}$	94.6	94.6	94.0	93.8	93.3	94.0	95.1	94.8	93.3

5.4.3 Puissance des différents tests

Cette section vise à vérifier la capacité des huit tests proposés à détecter que l'hypothèse nulle n'est pas vérifiée. Pour ce faire, mettons en situation un vecteur aléatoire (Y_1, Y_2, X) pour lequel X est distribué selon une loi uniforme sur $[0,1]$. La copule conditionnelle associée au vecteur (Y_1, Y_2) sachant $X = x$ est donnée soit par la copule de Normale C_ϱ^{N} ou la copule de Clayton C_γ^{CL} dont les paramètres ϱ et γ sont fixés suivant la valeur de x. Puisque les paramètres sont parfois difficilement interprétables, nous avons fixé leur valeur d'une telle manière que le tau de Kendall conditionnel associé à la copule en question satisfasse à la relation $\tau_x = \kappa(x + x^2)$. En vertu de l'équation (12), nous aurons alors

$$\varrho = \sin\left\{\kappa \times \frac{\pi}{2}(x + x^x)\right\} \quad \text{et} \quad \gamma = \frac{\kappa(x + x^2)}{1 - \kappa(x + x^2)}.$$

Les variables aléatoires Y_1 et Y_2 sont distribuées suivant une loi normale de moyenne $a(x+1)$ et de variance 1.

Dans cette étude, nous avons considéré la concordance entre $x \in \{0.4, 0.6, 0.8\}$ et $y = 0.2$ pour des valeurs de $\kappa \in \{0.4, 0.6\}$ et $a \in \{0, 5\}$. Les résultats sont présentés aux tables 5.3 et 5.4.

Conformément à ce qui était attendu, on peut voir dans ces tables que la puissance de chacun des huit tests considérés augmente avec la taille d'échantillon et la valeur prise par κ. De plus, les tests sont aussi de plus en plus puissants au fur et à mesure que x s'éloigne de y. On constate que les tests les plus performants sont généralement ceux basés sur les fonctionnelles de Spearman et de Kendall, avec un léger avantage pour le premier. Enfin, lorsque la co-variable influence les distributions marginales conditionnelles de Y_1 et Y_2, *i.e* lorsque $a = 5$, les versions des statistiques de test basée sur des *Pseudo-observations uniformisée* (munies d'un tilde) sont généralement meilleures que celle utilisant les observations non transformées. Cela est peut être causé par la présence d’un biais plus important, dans ce dernier cas, dû à l’influence de la co-variable dans les distributions marginales.

Tableau 5.3 – Pourcentage de rejet avec $\alpha = 5\%$ le premier modèle. Sans influence de la co-variable dans les distributions marginales conditionnelles.

	Statistique de test	$n = 250$			$n = 500$			$n = 1000$		
		$x = 0.4$	$x = 0.6$	$x = 0.8$	$x = 0.4$	$x = 0.6$	$x = 0.8$	$x = 0.4$	$x = 0.6$	$x = 0.8$
	CVM	18.4	55.1	87.0	29.3	81.4	99.4	49.1	97.2	100.0
	$\widetilde{\mathrm{CVM}}$	19.1	57.8	87.9	29.6	82.6	99.5	50.6	97.4	100.0
	K	15.6	47.9	76.0	23.9	69.9	97.0	41.1	92.4	100.0
$\kappa = 0.4$	$\widetilde{K}$	15.1	47.3	77.7	23.5	70.6	97.9	41.1	92.0	100.0
	SP	23.3	71.0	98.6	36.0	91.1	100.0	57.8	99.2	100.0
	$\widetilde{\mathrm{SP}}$	23.8	70.8	97.5	35.7	91.4	100.0	58.4	99.4	100.0
	KD	23.7	71.0	98.5	35.5	91.1	100.0	57.7	99.3	100.0
	$\widetilde{\mathrm{KD}}$	23.3	71.0	98.6	36.0	91.1	100.0	57.8	99.2	100.0
	CVM	27.7	76.7	96.1	47.2	97.3	100.0	77.2	100.0	100.0
	$\widetilde{\mathrm{CVM}}$	28.4	79.0	97.3	47.7	97.2	100.0	77.8	100.0	100.0
	K	23.2	68.4	93.1	37.9	92.5	100.0	65.7	99.9	100.0
$\kappa = 0.6$	$\widetilde{K}$	23.7	67.5	93.2	37.5	92.2	99.9	66.3	99.8	100.0
	SP	42.8	96.6	100.0	63.5	99.9	100.0	88.8	100.0	100.0
	$\widetilde{\mathrm{SP}}$	42.0	94.9	100.0	62.9	99.9	100.0	88.8	100.0	100.0
	KD	42.7	96.9	100.0	64.2	99.9	100.0	88.7	100.0	100.0
	$\widetilde{\mathrm{KD}}$	42.8	96.6	100.0	63.5	99.9	100.0	88.8	100.0	100.0
	CVM	14.4	46.9	77.7	24.9	74.9	98.2	52.2	95.5	99.9
	$\widetilde{\mathrm{CVM}}$	20.6	58.4	88.0	31.7	83.2	99.5	53.5	97.3	100.0
	K	15.1	41.5	70.1	25.3	67.3	93.8	49.6	89.8	99.9
$\kappa = 0.4$	$\widetilde{\mathrm{K}}$	15.0	46.8	75.3	23.7	68.3	97.5	43.0	91.1	100.0
	SP	22.8	70.2	98.2	35.9	91.3	100.0	58.7	99.3	100.0
	$\widetilde{\mathrm{SP}}$	28.0	74.4	98.4	39.9	92.8	100.0	62.3	99.6	100.0
	KD	18.3	67.0	98.1	32.4	90.7	100.0	61.9	99.5	100.0
	$\widetilde{\mathrm{KD}}$	22.8	70.2	98.2	35.9	91.3	100.0	58.7	99.3	100.0
	CVM	20.8	68.8	91.4	39.6	94.9	99.9	76.1	99.9	100.0
	$\widetilde{\mathrm{CVM}}$	30.6	79.2	97.0	49.1	97.0	100.0	79.7	100.0	100.0
	K	21.1	63.2	87.3	35.9	89.6	99.7	70.0	99.3	100.0
$\kappa = 0.6$	$\widetilde{\mathrm{K}}$	21.3	66.9	92.3	36.7	92.9	100.0	64.6	99.9	100.0
	SP	42.6	96.4	100.0	63.5	99.9	100.0	87.7	100.0	100.0
	$\widetilde{\mathrm{SP}}$	47.6	96.4	100.0	66.6	99.9	100.0	89.2	100.0	100.0
	KD	36.5	95.7	100.0	59.1	99.9	100.0	90.1	100.0	100.0
	$\widetilde{\mathrm{KD}}$	42.6	96.4	100.0	63.5	99.9	100.0	87.7	100.0	100.0

Tableau 5.4 – Pourcentage de rejet avec $\alpha = 5\%$ le second modèle. Sans influence de la co-variable dans les distributions marginales conditionnelles

	Statistique de test	$n = 250$			$n = 500$			$n = 1000$		
		$x = 0.4$	$x = 0.6$	$x = 0.8$	$x = 0.4$	$x = 0.6$	$x = 0.8$	$x = 0.4$	$x = 0.6$	$x = 0.8$
$\kappa = 0.4$	CVM	21.3	55.3	85.2	29.0	79.3	99.1	47.2	96.2	100.0
	$\widetilde{\text{CVM}}$	21.8	57.0	85.6	28.6	81.0	99.0	46.6	96.0	100.0
	K	18.5	45.7	78.4	25.0	71.2	98.2	41.1	93.1	99.9
	$\widetilde{\text{K}}$	18.2	46.3	78.1	24.6	71.7	98.5	41.5	92.5	100.0
	SP	26.4	71.3	96.4	35.4	90.3	99.9	55.6	98.1	100.0
	$\widetilde{\text{SP}}$	25.1	67.3	94.9	33.2	88.2	99.9	54.5	98.1	100.0
	KD	26.1	71.0	97.0	35.0	90.5	99.9	55.3	98.2	100.0
	$\widetilde{\text{KD}}$	26.4	71.3	96.4	35.4	90.3	99.9	55.6	98.1	100.0
$\kappa = 0.6$	CVM	29.2	75.3	96.6	43.1	96.6	100.0	73.5	100.0	100.0
	$\widetilde{\text{CVM}}$	29.8	78.9	96.3	42.5	96.7	100.0	73.8	100.0	100.0
	K	26.2	69.8	93.5	37.8	93.9	100.0	64.5	99.6	100.0
	$\widetilde{\text{K}}$	25.8	69.0	93.2	38.2	94.8	100.0	64.6	99.3	100.0
	SP	43.5	97.4	100.0	61.5	100.0	100.0	83.3	100.0	100.0
	$\widetilde{\text{SP}}$	39.1	94.3	100.0	57.5	99.6	100.0	81.6	100.0	100.0
	KD	42.7	97.2	100.0	61.2	100.0	100.0	83.6	100.0	100.0
	$\widetilde{\text{KD}}$	43.5	97.4	100.0	61.5	100.0	100.0	83.3	100.0	100.0
$\kappa = 0.4$	CVM	14.8	45.7	74.4	23.4	71.0	96.9	49.8	92.8	99.9
	$\widetilde{\text{CVM}}$	23.0	57.8	87.6	29.9	81.9	99.1	48.2	96.4	100.0
	K	14.5	38.5	65.1	21.0	64.2	93.7	44.8	85.7	99.9
	$\widetilde{\text{K}}$	17.2	46.4	77.3	23.6	71.0	97.7	39.3	92.3	100.0
	SP	25.9	70.9	96.9	35.5	89.6	99.9	55.2	98.1	100.0
	$\widetilde{\text{SP}}$	28.3	72.0	96.6	38.9	90.6	99.9	58.8	98.2	100.0
	KD	23.3	68.3	95.9	34.1	89.6	99.9	61.4	98.4	100.0
	$\widetilde{\text{KD}}$	25.9	70.9	96.9	35.5	89.6	99.9	55.2	98.1	100.0
$\kappa = 0.6$	CVM	21.2	64.0	91.8	38.3	93.4	99.9	71.3	99.1	100.0
	$\widetilde{\text{CVM}}$	31.2	80.0	96.8	46.0	96.4	100.0	75.4	100.0	100.0
	K	20.1	57.0	88.4	31.9	88.3	99.8	63.2	99.2	100.0
	$\widetilde{\text{K}}$	25.3	68.5	93.2	37.3	92.9	100.0	64.7	99.1	100.0
	SP	43.4	97.0	100.0	60.9	100.0	100.0	83.0	100.0	100.0
	$\widetilde{\text{SP}}$	45.8	96.3	100.0	63.3	99.9	100.0	84.6	100.0	100.0
	KD	37.7	96.4	100.0	59.5	99.6	100.0	85.7	100.0	100.0
	$\widetilde{\text{KD}}$	43.4	97.0	100.0	60.9	100.0	100.0	83.0	100.0	100.0

CHAPITRE 6

Conclusion

Cette thèse était dédiée à l'étude d'estimateurs et de tests non-paramétriques pour des distributions et des copules conditionnelles. Dans un premier temps, des méthodes de ré-échantillonnage adéquates et adaptées aux estimateurs à noyaux de distributions conditionnelles ont été étudiées. Ces méthodes ont permis d'obtenir des intervalles de confiance pour des mesures de dépendance conditionnelle. Ensuite, nous avons montré la convergence de deux estimateurs de la copule conditionnelle, proposés par Gijbels *et al.* (2011), en présence de données sérielles. Ces résultats nous ont mené à définir de nouvelles mesures de la causalité locale, ainsi que des intervalles de confiance pour ces mesures. Nous avons aussi proposé une façon d'estimer la distribution conditionnelle et la copule conditionnelle lorsqu'une variable est censurée. Enfin, nous avons adapté les méthodes de ré-échantillonnage proposées dans le premier article pour mettre en oeuvre un test d'hypothèses pour vérifier si deux copules conditionnelles sont concordantes ou non.

Les travaux présentés dans cette thèse ont, chacun à leur façon, contribué modestement au développement de méthodes d'estimation et d'analyse en présence d'une co-variable. Ces derniers s'inscrivent dans la mouvance d'un domaine en plein essor, au sein duquel plusieurs avenues sont encore à ce jour inexplorées. Parmi celles-ci, quelques une se trouvent dans la lignée des résultats ici présentés. En voici quelques exemples.

D'abord, les méthodes de ré-échantillonnages développées dans le premier article permettent d'envisager la mise en oeuvre de tests d'indépendance et d'adéquation pour les distributions et copules conditionnelles. À plus large spectre, la question de déterminer si la copule conditionnelle dépend de la valeur prise par la co-variable constitue, à ce jour, encore un défi de taille, bien que les tests de concordance permettent, d'une certaine manière, de poser un premier diagnostique.

Ensuite, il serait bien sûr souhaitable d'adapter les stratégies « du multiplicateur » à la copule conditionnelle dans le cadre où les données proviendraient de séries chronologiques. Ceci permettrait, par exemple, de prendre en considération le biais asymptotique des estimateurs des mesures de causalité proposées dans le troisième article.

Enfin, les résultats présentés dans le quatrième article constituent un premier pas à l'étude de mesures de dépendance conditionnelle basées sur la copule conditionnelle. Tel que mentionné dans cet article, le principal obstacle à cet objectif se pose lorsque le support de la variable de censure est inclus dans celui de la variable d'intérêt, ce qui rend inaccessible une partie du domaine de la copule conditionnelle. Il faut alors songer à une façon d'estimer ce point de troncation à partir duquel aucune information sur la variable d'intérêt n'est disponible.

Annexe A

Multiplier bootstrap methods for conditional distributions

A.1 Assumptions on the weights

The following notation will be adopted in the sequel. For a sequence of random variables $(Z_n)_{n\in\mathbb{N}}$ and a sequence of real numbers $(a_n)_{n\in\mathbb{N}}$, one writes $Z_n = O_e(a_n)$ if there exist finite constants C_1, C_2 and $\eta > 0$ such that $\mathrm{P}(Z_n/a_n \geq C_1) \leq C_2 \exp(-n^\eta/C_1)$. Also, let $J_{nx} = [\min_{i\in I_{nx}} x_i, \max_{i\in I_{nx}} x_i]$, where $I_{nx} = \{j \in \{1,\dots,n\} : w_{nj}(x,h_n) > 0\}$. The assumptions on the weights listed below are for fixed designs ; the adaptation to random designs generally consists in replacing o by $o_{\text{a.s.}}$ and O by $O_{\text{a.s.}}$, except for W_4, W_6, W_2' and W_2'', where O has to be replaced by O_e. This stronger requirement ensures the validity of Lemma 3 and Lemma 4 in Omelka *et al.* (2013) in the context of a random design. These lemmas will prove useful in the proofs presented in the online resource.

Assumptions W_1–W_5 below are needed to establish the asymptotic behavior of $\mathbb{F}_{xh}$ stated in subsection 2.1 ; they are also needed, as well as W_6, in the proof of Proposition 1.

W_1. $\sqrt{nh_n} \max_{1 \le i \le n} |w_{ni}(x, h_n)| = o(1)$;

W_2. $\sqrt{nh_n} \left| \sum_{i=1}^{n} w_{ni}(x, h_n)(x_i - x) - h_n^2 K_2 \right| = o(1)$ for some $K_2 = K_2(x) \in (0, \infty)$;

W_3. $\sqrt{nh_n} \left| \sum_{i=1}^{n} w_{ni}(x, h_n)(x_i - x)^2/2 - h_n^2 K_3 \right| = o(1)$ for some $K_3 = K_3(x) \in (0, \infty)$;

W_4. $nh_n \sum_{i=1}^{n} \{w_{ni}(x, h_n)\}^2 - K_4 = O(1)$ for some $K_4 = K_4(x) \in (0, \infty)$;

W_5. $\max_{i \in I_{nx}} x_i - \min_{i \in I_{nx}} x_i = o(1)$;

W_6. $(nh_n)^2 \sum_{i=1}^{n} \{w_{ni}(x, h_n)\}^4 = O(n^{-\delta})$ for some $\delta > 0$.

The next assumptions are needed in order to establish the consistency of $\widehat{B_x}$ and the validity of the two multiplier bootstrap methods.

W_1'. $\left| \sum_{i=1}^{n} w_{ni}'(x, g_n) \right| = o(1)$;

W_2'. $n^{\delta} \sum_{i=1}^{n} \{w_{ni}'(x, g_n)\}^2 = O(1)$ for some $\delta > 0$;

W_3'. $\left| \sum_{i=1}^{n} w_{ni}'(x, g_n)(x_i - x) - 1 \right| = o(1)$;

W_1''. $\sup_{z \in J_{nx}} \left| \sum_{i=1}^{n} w_{ni}''(z, g_n) \right| = o(1)$;

W_2''. $n^{\delta} \sup_{z \in J_{nx}} \sum_{i=1}^{n} \{w_{ni}''(x, g_n)\}^2 = O(1)$ for some $\delta > 0$;

W_3''. $\sup_{z \in J_{nx}} \left| \sum_{i=1}^{n} w_{ni}''(z, g_n)(x_i - x) \right| = o(1)$;

W_4''. $\sup_{z \in J_{nx}} \left| \sum_{i=1}^{n} \{w_{ni}''(x, g_n)(x_i - x)\}^2 - 1 \right| = o(1)$;

W_5''. $g_n^2 \sup_{z \in J_{nx}} \sum_{i=1}^{n} |w_{ni}''(z, g_n)| = O(1)$;

W_6''. One can find constants $C_1, C_2 < \infty$ and $\alpha > 0$ such that for all $z_1, z_2 \in J_{nx}$,

$$\max_{i \in I_{nx}} |w_{ni}''(z_1, g_n) - w_{ni}''(z_2, g_n)| \le C_1 \, g_n^{-C_2} |z_1 - z_2|^{\alpha}.$$

A.2 Complementary computations

A.2.1 Hadamard derivative of the variance functional

Let $\mathcal{D}$ be the space of distribution functions. Then, for any $\delta \in \mathcal{D}$, define $\delta_t = \delta + t\Delta_t \in \ell^\infty(\mathbb{R})$ such that $\Delta_t \to \Delta \in \mathbb{D}_0$ uniformly as $t \to 0$. Then,

$$
\begin{aligned}
&\{\delta_t(y_1 \wedge y_2) - \delta_t(y_1)\delta_t(y_2)\} - \{\delta(y_1 \wedge y_2) - \delta(y_1)\delta(y_2)\} \\
&\quad = t\Delta_t(y_1 \wedge y_2) - t\{\Delta_t(y_1)\delta(y_2) - \Delta_t(y_2)\delta(y_1)\} - t^2\Delta_t(y_1)\Delta_t(y_2) \\
&\quad = t\{\mathbb{I}(y_1 \le y_2) - \delta(y_2)\}\Delta_t(y_1) + t\{\mathbb{I}(y_2 \le y_1) - \delta(y_1)\}\Delta_t(y_2) - t^2\Delta_t(y_1)\Delta_t(y_2).
\end{aligned}
$$

It then follows that as $t \to 0$,

$$
\begin{aligned}
\frac{\Lambda^v(\delta_t) - \Lambda^v(\delta)}{t} &= 2\int_{\mathbb{R}^2} \{\mathbb{I}(y_1 \le y_2) - \delta(y_2)\}\Delta_t(y_1)\mathrm{d}y_1\mathrm{d}y_2 - t\left\{\int_{\mathbb{R}} \Delta_t(y_1)\mathrm{d}y_1\right\}^2 \\
&= 2\int_{\mathbb{R}^2} \{\mathbb{I}(y_1 \le y_2) - \delta(y_2)\}\Delta_t(y_1)\mathrm{d}y_1\mathrm{d}y_2 - t\{\Lambda^m(\Delta_t)\}^2 \\
&\to 2\int_{\mathbb{R}^2} \{\mathbb{I}(y_1 \le y_2) - \delta(y_2)\}\Delta(y_1)\mathrm{d}y_1\mathrm{d}y_2.
\end{aligned}
$$

A.2.2 Conditional variance and correlation

First introduce the notation $\mathbf{A}^{(j)} = (Y_{j1} - \theta^{\mathrm{m}}_{xjh}, \dots, Y_{jn} - \theta^{\mathrm{m}}_{xjh})$ and $\mathbf{B}^{(jj')} = \mathrm{diag}((\mathbf{A}^{(j)})^\top \mathbf{A}^{(j')})$ for $j, j' \in \{1, 2\}$. For the variance functional, one has

$$
\widehat{(\Lambda^{\mathrm{v}})'_{F_x}}(\Delta) = (\Lambda^{\mathrm{v}})'_{F_{xh}}(\Delta) = 2\sum_{i=1}^n w_{ni}(x, h_n)\int_{\mathbb{R}^2} \{\mathbb{I}(y_2 \le y_1) - \mathbb{I}(Y_i \le y_1)\}\Delta(y_2)\,\mathrm{d}y_1\mathrm{d}y_2,
$$

and then long but straightforward computations yield $\widehat{\Theta}^{(1)}_x = \widehat{\Theta}^{(2)}_x = \sqrt{nh_n}\,\mathbf{B}^{(11)}\,\mathbf{L}_x$. For the correlation functional, one can show that $\Lambda^{\mathrm{Cov}}(F_{xh}) = \mathbf{B}^{(12)}\,\mathbf{w}_x^\top$, $\Lambda^{\mathrm{v}}(F_{1xh}) = \mathbf{B}^{(11)}\,\mathbf{w}_x^\top$ and $\Lambda^{\mathrm{v}}(F_{2xh}) = \mathbf{B}^{(22)}\,\mathbf{w}_x^\top$, so that one can write

$$
\rho_{xh} = \mathbf{B}^{(12)}\,\mathbf{w}_x^\top / \{\mathbf{B}^{(11)}\,\mathbf{w}_x^\top\}^{1/2}\{\mathbf{B}^{(22)}\,\mathbf{w}_x^\top\}^{1/2}.
$$

By similar computations, one can show that

$$\widehat{\Theta}_x^{(1)} = (\Lambda^\rho)'_{F_{xh}}(\widehat{\mathbb{F}}_x) = \rho_{xh}\left(\frac{\mathbf{B}^{(12)}\,\mathbf{L}_x^\top}{\mathbf{B}^{(12)}\,\mathbf{w}_x^\top} - \frac{\mathbf{B}^{(11)}\,\mathbf{L}_x^\top}{2\mathbf{B}^{(11)}\,\mathbf{w}_x^\top} - \frac{\mathbf{B}^{(22)}\,\mathbf{L}_x^\top}{2\mathbf{B}^{(22)}\,\mathbf{w}_x^\top}\right)$$

and

$$\widehat{\Theta}_x^{(2)} = \sqrt{nh_n}\left\{\frac{\mathbf{B}^{(12)}\,(\mathbf{w}_x+\mathbf{L}_x)^\top}{\{\mathbf{B}^{(11)}\,(\mathbf{w}_x+\mathbf{L}_x)^\top\,\mathbf{B}^{(22)}\,(\mathbf{w}_x+\mathbf{L}_x)^\top\}^{1/2}} - \frac{\mathbf{B}^{(12)}\,\mathbf{w}_x^\top}{\{\mathbf{B}^{(11)}\,\mathbf{w}_x^\top\mathbf{B}^{(22)}\,\mathbf{w}_x^\top\}^{1/2}}\right\}.$$

A.2.3 Kendall's tau

For $\mathbf{Y}_i = (Y_{1i}, Y_{2i})$, $i \in \{1,\ldots,n\}$, let $(\mathcal{K}_{ii'}) = \mathbb{I}(\mathbf{Y}_i \leq \mathbf{Y}_{i'})$. Then, one obtains from a direct computation that $\tau_{xh} = \Lambda^\tau(F_{xh}) = -1 + 4\,\mathbf{w}_x\,\mathcal{K}\,\mathbf{w}_x^\top$. Next, since $\mathrm{d}F_{xh}$ puts mass $w_{ni}(x,h_n)$ at $\mathbf{Y}_i$, one obtains

$$\begin{aligned}(\Lambda^\tau)'_{F_{xh}}(\Delta) &= 4\int_{\mathbb{R}^2}\Delta(y_1,y_2)\,\mathrm{d}F_{xh}(y_1,y_2) + 4\int_{\mathbb{R}^2}F_{xh}(y_1,y_2)\,\mathrm{d}\Delta(y_1,y_2)\\ &= 4\sum_{i=1}^n w_{ni}(x,h_n)\left\{\Delta(Y_{1i},Y_{2i}) + \int_{Y_{1i}}^\infty\int_{Y_{2i}}^\infty \mathrm{d}\Delta(y_1,y_2)\right\}.\end{aligned}$$

If $\lim_{\gamma\to\infty}\Delta(\gamma,y) = \lim_{\gamma\to\infty}\Delta(y,\gamma) = 0$, the last equality reduces to

$$(\Lambda^\tau)'_{F_{xh}}(\Delta) = 8\sum_{i=1}^n w_{ni}(x,h_n)\,\Delta(Y_{1i},Y_{2i}).$$

Hence,

$$\widehat{\Theta}_x^{(1)} = (\Lambda^\tau)'_{F_{xh}}(\widehat{\mathbb{F}}_x) = 8\sqrt{nh_n}\,\mathbf{w}_x\,\mathcal{K}\,\mathbf{L}_x^\top.$$

Also, one can show that $\Lambda^\tau(\widehat{F}_x) = -1 + 4(\mathbf{w}_x+\mathbf{L}_x)\,\mathcal{K}\,(\mathbf{w}_x+\mathbf{L}_x)^\top$, and then

$$\begin{aligned}\widehat{\Theta}_x^{(2)} &= \sqrt{nh_n}\left\{\left(-1+4(\mathbf{w}_x+\mathbf{L}_x)\,\mathcal{K}\,(\mathbf{w}_x+\mathbf{L}_x)^\top\right) - \left(4\,\mathbf{w}_x\,\mathcal{K}\,\mathbf{w}_x^\top - 1\right)\right\}\\ &= 4\sqrt{nh_n}\left(\mathbf{w}_x\,\mathcal{K}\,\mathbf{L}_x^\top + \mathbf{L}_x\,\mathcal{K}\,\mathbf{w}_x^\top + \mathbf{L}_x\,\mathcal{K}\,\mathbf{L}_x^\top\right).\end{aligned}$$

A.2.4 Copula

The empirical conditional copula is $C_{xh}(\mathbf{u}) = \Lambda^{\mathrm{C}}(F_{xh}) = \mathcal{I}^{\mathrm{C}}(\mathbf{u})\,\mathbf{w}_x^\top$, where $(\mathcal{I}^{\mathrm{C}})_i = \mathbb{I}\left\{\mathbf{Y}_i \leq \mathbf{F}_{xh}^{-1}(\cdot)\right\}$. One also obtains $\widehat{\Theta}_x^{(1)} = \widehat{\Theta}_x^{(1)}(\mathbf{u}) = \widetilde{\mathcal{I}}^{\mathrm{C}}(\mathbf{u})\,\mathbf{L}_x^\top$, where

$$(\widetilde{\mathcal{I}}^{\mathrm{C}}(\mathbf{u}))_i \;\; = \;\; \mathbb{I}\left\{\mathbf{Y}_i \leq \mathbf{F}_{xh}^{-1}(\mathbf{u})\right\} - \sum_{j=1}^{d} \widehat{C_x^{(j)}}(\mathbf{u})\,\mathbb{I}\left\{Y_{ji} \leq F_{xjh}^{-1}(u_j)\right\}.$$

The formula for $\widehat{\Theta}_x^{(2)}$ is more involved.

A.2.5 Testing a covariate's influence

Let $(\mathcal{Y}_{ii'})_{i,i'=1}^{n} = Y_i \wedge Y_{i'}$. Then by straightforward computations, one can show that

$$S_{xx'h}^{\mathrm{CvM}} = nh_n\left(\mathbf{w}_x - \mathbf{w}_{x'}\right)\,\mathcal{Y}\,\left(\mathbf{w}_x - \mathbf{w}_{x'}\right)^\top$$

and

$$\widehat{S}_{xx'}^{\mathrm{CvM}} = nh_n\left(\mathbf{L}_x - \mathbf{L}_{x'}\right)\,\mathcal{Y}\,\left(\mathbf{L}_x - \mathbf{L}_{x'}\right)^\top.$$

Annexe B

On the asymptotic behavior of two estimators of the conditional copula based on time series

B.1 Assumptions needed in Proposition 1, Proposition 1 and Proposition 5

B.1.1 Distributional assumptions

$\mathcal{A}_1$. 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.

$\mathcal{A}_3$. The partial derivatives

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

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

$\mathcal{A}_2^{\star}$. 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.

$\mathcal{A}_4$. 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.

B.1.2 Assumptions on the weights

W_1. $\max_{1 \le i \le n} |w_{ni}(x, h)| = O_{\mathrm{P}}\left((nh)^{-1}\right);$

W_2. $\sum_{i=1}^n w_{ni}(x, h)(X_i - x) = h^2 K_2 + o_{\mathrm{P}}\left((nh)^{-1/2}\right)$ for some $K_2 < \infty$;

W_3. $\sum_{i=1}^n w_{ni}(x, h)(X_i - x)^2/2 = h^2 K_3 + o_{\mathrm{P}}\left((nh)^{-1/2}\right)$ for some $K_3 < \infty$;

W_4. $nh \sum_{i=1}^n \{w_{ni}(x, h)\}^2 = K_4 + o_{\mathrm{P}}(1)$ for some $K_4 > 0$;

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

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

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

W_8. For some finite constant C,

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

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

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

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

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

In order to establish moment inequalities of order r, one needs that for any integer $1 \leq k \leq 6$, any choice of $L_1, \ldots, L_k \in \mathbb{N}$ such that $L_1 + \cdots + L_k = 6$, and for some positive sequence v_n satisfying $n - v_n \to \infty$:

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

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

W_{13}. $\sup_{z \in J_x} \max_{1 \leq \ell_2 < \cdots < \ell_k \leq v_n} \sum_{i=1}^{n-\ell_k} (X_i - z) w_{ni}(z,h)^{L_1} \prod_{j=2}^{k} w_{n,i+\ell_j}(z,h)^{L_j} = O_{\mathrm{P}}\left(\frac{h^{k+1}}{(nh)^{r-1}}\right)$.

B.2 Supplementary material

B.2.1 Asymptotic normality of Z_{xn}

It will be shown that for any arbitrary $y, z \in \mathbb{R}$, the random variable $Z_{xn}(y,z)$ is asymptotically normal. The arguments can easily be adapted to show the joint weak convergence of $Z_{nx}(y_1, z_1), \ldots, Z_{nx}(y_K, z_K)$ by invoking the Cramér–Wold device.

To prove the asymptotic normality of $Z_{xn}(y,z)$, 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 for for some $\delta, \epsilon > 0$, r_n, b_n and ℓ_n are integers such that $b_n \sim n^{1-\epsilon+\delta}$, $\ell_n \sim n^{1-\epsilon}$ and $\ell_n > h^{-1}$. Then, introduce for each $i \in \{1, \ldots, n\}$ the following sets :

$$\begin{aligned} \mathcal{U}_i &= \{j \in \mathbb{N} : (i-1)(b_n + \ell_n) + 1 \leq j \leq (i-1)(b_n + \ell_n) + b_n\}\,; \\ \mathcal{W}_i &= \{j \in \mathbb{N} : i(b_n + \ell_n) + 1 - \ell_n \leq j \leq i(b_n + \ell_n)\}\,. \end{aligned}$$

Letting

$$U_{ni} = \sum_{j \in \mathcal{U}_i} w_{nj}(x,h)\,\vartheta_j(y,z) \quad \text{and} \quad W_{ni} = \sum_{j \in \mathcal{W}_i} w_{nj}(x,h)\,\vartheta_j(y,z),$$

one can write $Z_{xn}(y,z) = U_n + W_n$, where

$$U_n = \sqrt{nh}\sum_{i=1}^{r_n} U_{ni} \quad \text{and} \quad W_n = \sqrt{nh}\sum_{i=1}^{r_n} W_{ni}.$$

It will next be demonstrate that for an appropriate choice of ϵ and δ, the random variable W_n is asymptotically negligible while U_n is asymptotically normal.

Firstly, computation presented in Section B.3 shows, that as long as $\ell_n > h^{-1}$, and since $\mathcal{A}_1$ holds, there exists a constant C_α that depends on the α-mixing coefficients and on the weight functions such that for sufficiently large n,

$$\sum_{i=1}^{r_n} \mathrm{E}\left(|W_{ni}|^4\right) \leq C_\alpha n^{-3} h^{-2} \ell_n.$$

Hence, for any $\kappa > 0$, one has that

$$\begin{aligned} \mathrm{P}(|W_n| > \kappa) &\leq \sum_{i=1}^{r_n} \mathrm{P}\left(|W_{ni}| > \frac{\kappa}{\sqrt{nh}\,r_n}\right) \\ &\leq \frac{r_n^4 (nh)^2}{\kappa^4} \sum_{i=1}^{r_n} \mathrm{E}\left(|W_{ni}|^4\right) \\ &\leq \frac{C_\alpha\, r_n^4\, \ell_n}{\kappa^4\, n} \sim n^{3\epsilon - 4\delta}. \end{aligned}$$

This last expression tends to zero as $n \to \infty$ whenever $3\epsilon < 4\delta$, which would entail $\mathrm{P}(|W_n| > \kappa) \to 0$.

Secondly, in order to deal with U_n, let $U'_{n1}, \ldots, U'_{nr_n}$ be independent random variables sharing the same conditional distribution 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 $16\,r_n\,\alpha_{\ell_n}$. Since assumption $\mathcal{A}_1$ ensures $\alpha(\ell) \sim \ell^{-a}$ for some $a > 6$, one obtains that

$$16 r_n \alpha_{\ell_n} \sim n^{(a+1)\epsilon - a - \delta}.$$

Therefore $16\,r_n\,\alpha_{\ell_n} \to 0$ whenever $(a+1)\epsilon < a + \delta$. As a consequence, U_n and U_n' are asymptotically equivalent. Then, it suffices to show that U_n' is asymptotically normal to prove the asymptotic normality of U_n. To do this, it will next be establish that the random variable U_n satisfies the Lyapunov condition required for the central limit theorem. By a straightforward computation, $\mathrm{Var}(\sqrt{nh}\sum_{i=1}^{r_n} U_{ni}') = K_4\,\sigma_x^2(y,z) + o(1)$. Moreover, it is shown in Section B.3 that

$$\sum_{i=1}^{r_n} \mathrm{E}\left\{(U_{ni}')^4\right\} = O\left(\frac{b_n}{n^3 h^2}\right).$$

For any $\kappa > 0$, one then has

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

As long as $\epsilon > \delta$, this last expression is $o(1)$, leading to a Lyapunov ratio that converges to zero. Thus, if τ is such that $h \sim n^{-\tau}$, letting $\epsilon = 4/5\min\{a/(a+1), 1-\tau\}$ and $\delta = 4/5\epsilon$ ensures the asymptotic normality of Z_{xh}.

B.3 Computation of $\sum_{k=1}^{r_n} \mathrm{E}\,W_{nk}^4$ and $\sum_{k=1}^{r_n} \mathrm{E}\,U_{nk}'^4$

Fix $y, z \in \mathbb{R}$ and for simplicity let ϑ_i stand for $\vartheta_i(y,z)$. For readability of the presentation, we write $\sum_{\mathcal{W}_k}$ when all the indices involved in the summation are in $\mathcal{W}_k$. The computations to be exposed next are valid provided assumptions $\mathcal{A}_1$, W_{11}–W_{12} are satisfied, as well as $\ell_n > h^{-1}$.

First decompose $W_{nk}^4 = \sum_{j=1}^5 W_{nk}^{(j)}$, where

$$W_{nk}^{(1)} = \sum_{\mathcal{W}_k} \vartheta_i^4 w_{ni}(x,h)^4, \quad W_{nk}^{(4)} = 6 \sum_{\substack{\mathcal{W}_k \\ i_1 \neq i_2 \neq i_3}} \vartheta_{i_1}^2 \vartheta_{i_2} \vartheta_{i_3} w_{ni_1}(x,h)^2 w_{ni_2}(x,h) w_{ni_3}(x,h)$$

$$W_{nk}^{(2)} = 4 \sum_{\substack{\mathcal{W}_k \\ i_1 \neq i_2}} \vartheta_{i_1}^3 \vartheta_{i_2} w_{ni_1}(x,h)^3 w_{ni_2}(x,h), \qquad W_{nk}^{(5)} = \sum_{\substack{\mathcal{W}_k \\ i_1 \neq \ldots \neq i_4}} \prod_{k=1}^4 \vartheta_{i_k} w_{ni_k}(x,h)$$

$$W_{nk}^{(3)} = 6 \sum_{\substack{\mathcal{W}_k \\ i_1 \neq i_2}} \vartheta_{i_1}^2 \vartheta_{i_2}^2 w_{ni_1}(x,h)^2 w_{ni_2}(x,h)^2.$$

The goal is now to bound $\sum_{k=1}^{r_n} \mathrm{E}\, W_{nk}^{(j)}$ for each j. We begin with $j = 1$. In view of condition W_{12}, $\sum_{i=1}^n w_{ni}(x,h)^4 \leq K_{12} n^{-3} h^{-3}$ for some $K_{12} < \infty$. As $\ell_n > h^{-1}$, one obtains that

$$\sum_{k=1}^{r_n} \mathrm{E}\, W_{nk}^{(1)} \leq K_{12} n^{-3} h^{-3} \leq K_{12} n^{-3} h^{-2} \ell_n.$$

To deal with $j = 2$, split $W_{nk}^{(2)}$ in $W_{nk}^{(2,<)}$ and $W_{nk}^{(2,>)}$ according to the cases $i_1 < i_2$ and $i_1 > i_2$. Next, as the random variables $\vartheta_1, \ldots, \vartheta_n$ satisfy a strong mixing assumption, it is useful to note that for any $1 \leq \ell \leq \ell_n - i$,

$$|\mathrm{E}(\vartheta_i^3 \vartheta_{i+\ell})| \leq \alpha(\ell) \leq 1$$

Letting π_h denote the integer part of $1/\sqrt{h}$, one writes

$$\begin{aligned} \sum_{k=1}^{r_n} \mathrm{E}\, W_{nk}^{(2,<)} \quad \leq \quad & \sum_{i=1}^{n} \sum_{l=1}^{\pi_h} \mathrm{E}\, \left(\vartheta_i^3 \vartheta_{i+\ell}\right) w_{ni}(x,h)^3 w_{n,i+\ell}(x,h) \\ & + \sum_{i=1}^{n} \sum_{l=\pi_h+1}^{l_n - i} \mathrm{E}\, \left(\vartheta_i^3 \vartheta_{i+\ell}\right) w_{ni}(x,h)^3 w_{n,i+\ell}(x,h). \end{aligned}$$

The first summand is bounded by

$$\pi_h \max_{1 \leq \ell \leq \pi_h} \sum_{i=1}^{n-\ell} w_{ni}(x,h)^3 w_{h,i+\ell}(x).$$

As W_{11} holds and $\ell_n > \pi_h$, the latter is $O(n^{-3}h^{-2}\ell_n)$. The second summand is bounded by

$$\left\{\max_{1\leq i\leq n} w_{ni}(x,h)\right\}\left\{\sum_{i=1}^{n} w_{ni}(x,h)^3\right\}\sum_{\ell=\pi_h+1}^{d}\alpha(\ell),$$

which can be bounded by

$$O(n^{-3}h^{-2}\ell_n)\times\sum_{\ell=\pi_h+1}^{d}\alpha(\ell)$$

since W_1 and W_{12} are satisfied together with $\ell_n > h^{-1}$. As assumption $\mathcal{A}_1$ holds, $\sum_{\ell=\pi_h+1}^{d}\alpha(\ell)\to 0$ and therefore the latter display is $o(n^{-3}h^{-2}\ell_n)$. As a consequence,

$$\sum_{k=1}^{r_n}\mathrm{E}\,W_{nk}^{(2,<)} = O(n^{-3}h^{-2}\ell_n).$$

Proceeding similarly for $W_{nk}^{(2,>)}$ yields

$$\sum_{k=1}^{r_n}\mathrm{E}\,W_{nk}^{(2)} = O(n^{-3}h^{-2}\ell_n)$$

In the case $j=3$, one decomposes $W_{nk}^{(3)} = \overline{W}_{nk}^{(3)} + (W_{nk}^{(3)} - \overline{W}_{nk}^{(3)})$, where

$$\overline{W}_{nk}^{(3)} = 2\sum_{i\in\mathcal{W}_k}\sum_{\ell=1}^{\ell_n-i}\mathrm{E}(\vartheta_i^2)\mathrm{E}(\vartheta_{i+\ell}^2)w_{ni}(x,h)w_{n,i+\ell}(x,h).$$

As $|\vartheta_i|\leq 1$, one deduces that

$$\sum_{k=1}^{r_n}\overline{W}_{nk}^{(3)} \;\leq\; \sum_{i=1}^{n}\sum_{\ell=1}^{\ell_n\wedge(n-i)} w_{ni}(x,h)^2 w_{n,i+\ell}(x,h)^2 = O(\ell_n n^{-3}h^{-2}),$$

where the last equality follows from assumption W_{11}. Next, since

$$|\mathrm{E}(\vartheta_i^2\vartheta_{i+\ell}^2) - \mathrm{E}(\vartheta_i^2)\mathrm{E}(\vartheta_{i+\ell}^2)|\leq\alpha(\ell),$$

one uses the same " splitting " strategy as previously to obtain that $\sum_{k=1}^{r_n}(W_{nk}^{(3)} - \overline{W}_{nk}^{(3)})$ is $O(n^{-3}h^{-2}\ell_n)$. We go directly to the case $j=5$, as the case $j=4$ can be dealt similarly but the computations are more involved in the former. Denote for each index $m=1,2,3$

the sets $\mathcal{W}_{nk}^{(m)} = \{i_1 < ... < i_4 \in \mathcal{W}_{nk} : \max(g_1, g_2, g_3) \leq g_m\}$, where $g_m = i_{m+1} - i_m$ is the gap between two consecutive indices. The introduction of these sets of indices is justified by the fact that $\mathrm{E}\, W_{nk}^{(5)} \leq 4!(T_{nk}^{(1)} + T_{nk}^{(2)} + T_{nk}^{(3)})$, where

$$T_{nk}^{(m)} = \sum_{\mathcal{W}_{nk}^{(m)}} \mathrm{E}\left\{\prod_{\ell=1}^{4} \vartheta_{i_\ell} w_{ni_\ell}(x,h)\right\}.$$

As previously, one then re-decompose $T_{nk}^{(m)} = \overline{T}_{nk}^{(m)} + (T_{nk}^{(m)} - \overline{T}_{nk}^{(m)})$, where

$$\overline{T}_{nk}^{(m)} = \sum_{\mathcal{W}_{nk}^{(m)}} \mathrm{E}\left(\prod_{\ell=l}^{m} \vartheta_{i_l}\right) \mathrm{E}\left(\prod_{l'=m+1}^{4} \vartheta_{i_{l'}}\right) \prod_{\ell=1}^{4} w_{ni_\ell}(x,h).$$

As the expectation of ϑ_i is zero, both $\overline{T}_{nk}^{(1)}$ and $\overline{T}_{nk}^{(3)}$ are 0. Using the fact that $|\mathrm{E}(\vartheta_{i_1}\vartheta_{i_2}\vartheta_{i_3}\vartheta_{i_4}) - \mathrm{E}(\vartheta_{i_1}\vartheta_{i_2})\mathrm{E}(\vartheta_{i_3}\vartheta_{i_4})| \leq \alpha(g_2)$, one obtains that

$$\begin{aligned}
\sum_{k=1}^{r_n} \overline{T}_{nk}^{(2)} &\leq \ell_n \left\{\max_{1\leq \ell \leq \ell_n} \sum_{i=1}^{n-\ell} w_{ni}(x,h) w_{n,i+\ell}(x,h)\right\} \left\{\max_{1\leq i\leq n} w_{ni}(x,h)\right\}^2 \sum_{g_2=1}^{\ell_n} \alpha(g_2) \\
&= O(\ell_n n^{-3} h^{-2}) \sum_{g_2=1}^{\ell_n} \alpha(g_2),
\end{aligned}$$

where the last inequality follows from condition W_1 and W_{11}. As the assumption over the alpha mixing coefficients implies that $\sum_{g_2=1}^{\infty} \alpha(g_2) < \infty$, it follows that the latter equation is $O(\ell_n n^{-3} h^{-2})$. Finally, as

$$\left|\mathrm{E}(\prod_{\ell=1}^{4} \vartheta_{i_\ell}) - \mathrm{E}(\prod_{\ell=l}^{j} \vartheta_{i_l})\mathrm{E}(\prod_{l'=j+1}^{4} \vartheta_{i_{l'}})\right| \leq \alpha(g_j),$$

one has

$$
\begin{aligned}
T_{nk}^{(j)} - \overline{T}_{nk}^{(j)} &= [\sum_{\substack{\mathcal{W}_{nk}^{(j)} \\ g_j \le \pi_h}} + \sum_{\substack{\mathcal{W}_{nk}^{(j)} \\ g_j > \pi_h}}] \left\{ \mathrm{E}\left(\prod_{\ell=l}^{j} \vartheta_{i_l}\right) - \mathrm{E}\left(\prod_{\ell=l}^{j} \vartheta_{i_l}\right) \mathrm{E}\left(\prod_{l'=j+1}^{4} \vartheta_{i_{l'}}\right)\right\} \\
&\quad \times \prod_{\ell=1}^{4} w_{ni_\ell}(x,h) \\
&\le \sum_{\substack{\mathcal{W}_{nk}^{(j)} \\ g_j \le \pi_h}} \prod_{\ell=1}^{4} w_{ni_\ell}(x,h) + \sum_{\substack{\mathcal{W}_{nk}^{(j)} \\ g_j > \pi_h}} \alpha(g_j) \prod_{\ell=1}^{4} w_{ni_\ell}(x,h).
\end{aligned}
$$

In view of last equation, one directly obtains

$$
\begin{aligned}
\sum_{k=1}^{r_n} (T_{nk}^{(j)} - \overline{T}_{nk}^{(j)}) &\le \pi_h \left\{ \max_{1 \le h_1 < h_2 < h_3 < \pi_h} \sum_{i=1}^{n-h_3} w_{ni}(x,h) \prod_{k=1}^{3} w_{n,i+h_k}(x,h) \right\} \\
&\quad + \left\{ \max_{1 \le i \le n} w_{ni}(x,h) \right\}^3 \sum_{g_j = \pi_h + 1}^{\ell_n} (g_j + 1)^2 \alpha(g_j)
\end{aligned}
$$

From assumption W_1 and W_{11}, the first summand is $O(n^{-3}h^{-1}\pi_h)$. Moreover, as the assumption $\mathcal{A}_1$ over the alpha-mixing coefficient implies the finiteness of $\sum_{g_j=1}^{\infty}(g_j + 1)^2\alpha(g_j)$, one deduces that the previous equation is $o(n^{-3}h^{-2}\ell_n)$. Wrapping up all the terms, one obtains the desired result, i.e $\sum_{k=1}^{r_n} \mathrm{E}\, W_{nk}^4 = O(n^{-3}h^{-2}\ell_n)$. Similar arguments leads to $\sum_{k=1}^{r_n} \mathrm{E}\, U_{nk}'^4 = O(n^{-3}h^{-2}b_n)$.

B.3.1 Proof of Lemma 2.3

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

$$
\begin{aligned}
T_\gamma^{(1)} &= \left\{ F_{1x}^{-1}(0), F_{1x}^{-1}\left(\frac{1}{\kappa_\gamma}\right), \ldots, F_{1x}^{-1}(1) \right\}, \\
T_\gamma^{(2)} &= \left\{ F_{2x}^{-1}(0), F_{2x}^{-1}\left(\frac{1}{\kappa_\gamma}\right), \ldots, F_{2x}^{-1}(1) \right\}.
\end{aligned}
$$

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

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

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

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

Since Assumption $\mathcal{A}_2$ holds, a Taylor expansion allows to write

$$\begin{aligned} &\sqrt{nh}\sum_{i=1}^n w_{ni}(x,h)\left\{H_{X_i}(\overline{\omega}_\gamma) - H_{X_i}(\underline{\omega}_\gamma)\right\} \\ &\quad= \sqrt{nh}\left\{H_x(\overline{\omega}_\gamma) - H_x(\underline{\omega}_\gamma)\right\} + \left\{\dot{H}_x(\overline{\omega}_\gamma) - \dot{H}_x(\underline{\omega}_\gamma)\right\}\sqrt{nh}\sum_{i=1}^n w_{ni}(x,h)(X_i - x) \\ &\qquad + \sqrt{nh}\sum_{i=1}^n w_{ni}(x,h)(X_i - x)^2\left\{\ddot{H}_{\zeta_i}(\overline{\omega}_\gamma) - \ddot{H}_{\zeta_i}(\underline{\omega}_\gamma)\right\}, \end{aligned} \tag{B.1}$$

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 $|H(\omega_1) - H(\omega_2)| \leq |F_1(y_1) - F_1(y_2)| + |F_2(z_1) - F_2(z_2)|$. From Assumptions W_2–W_3 and $\mathcal{A}_2$, the right-hand side of equation (B.1) is bounded by $\sqrt{nh}\,\rho(\overline{\omega}_\gamma, \underline{\omega}_\gamma) + o(1)$. As a consequence, uniformly in $\omega \in \mathbb{R}^2$,

$$Z_{xn}(\omega) - Z_{xn}(\underline{\omega}_\gamma) \leq Z_{xn}(\overline{\omega}_\gamma) - Z_{xn}(\underline{\omega}_\gamma) + o(\sqrt{nh}h^2).$$

The negligibility of the remainder term $o(\sqrt{nh}h^2)$ is ensured by the fact that $\sqrt{nh}h^2 < \infty$. From similar arguments, one deduces

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

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

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

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 $\epsilon > 0$, $\mathrm{P}\left(\mathfrak{W}_{\delta_n}(Z_{xn}, T_\gamma) > \epsilon\right)$ 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

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

B.3.2 Proof of Lemma 2.4

First note that since condition $\mathcal{A}_2$ holds, for any $i \in I_{nx}$,

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

where z_i is between X_i and x. For simplicity let ϑ_i stand for $\vartheta_i(A)$ and ν_z for $\nu_z(A)$. Moreover, throughout this section, set $\nu_{xh} = \nu_x + h^2(\dot{\nu}_x + \ddot{\nu}_x)$. As a starting point, write $\mathcal{S}_j = \{\mathbf{L} = (L_1, \ldots, L_j) \in \{1, \ldots, 6\}^j : L_1 + \ldots + L_j = 6\}$ and notice that $|\mathbb{H}_{xh}(A)|^6 = (nh)^3 \sum_{j=1}^{6} \sum_{\mathbf{L} \in \mathcal{S}_j} T_{hx}^{(j)}(\mathbf{L})$, with

$$T_{hx}^{(j)}(\mathbf{L}) = \sum_{\substack{i_1 \ldots i_j = 1 \\ i_1 \neq \ldots \neq i_j}}^{n} \prod_{k=1}^{j} \vartheta_{i_k}^{L_k} w_{ni_k}(x,h)^{L_k}.$$

The goal is now to bound each $T_{hx}^{(j)}(\mathbf{L})$, for $j = 1, \ldots, 6$. We begin with $j = 1$. In this case, $\mathcal{S}_1 = \{(6)\}$. One then obtains from equation (B.2) together with assumption $\mathcal{A}_2$

that

$$\begin{aligned}\mathrm{E}\, T_{hx}^{(1)}(1) \quad &\leq \textstyle\sum_{i=1}^{n} \nu_{X_i} w_{ni}(x,h)^6 = \nu_x \sum_{i=1}^{n} w_{ni}(x,h)^6 \\ &\quad + \dot{\nu}_x \textstyle\sum_{i=1}^{n} (X_i - x) w_{ni}(x,h)^6 + \frac{1}{2} \sum_{i=1}^{n} (X_i - x)^2 \ddot{\nu}_{z_i} w_{ni}(x,h)^6.\end{aligned}$$

Hence, in view of assumption W_{12}–W_{13} together with condition $\mathcal{A}_2$, the previous equation can be bounded by $\overline{\omega}_1 (nh)^{-5}\{\nu_x + h^2(\dot{\nu}_x + \ddot{\nu}_x)\}$ for some $\overline{\omega}_1 > 0$. For $j = 2$, one first splits $T_{hx}^{(2)}(\mathbf{L})$ into $T_{hx}^{(2,<)}$ and $T_{hx}^{(2,>)}$ according to the cases $i_1 < i_2$ and $i_2 < i_1$. Starting with $T_{hx}^{(2,<)}$, one decomposes $T_{hx}^{(2,<)}(\mathbf{L}) = \overline{T}_{hx}^{(2)}(\mathbf{L}) + \{T_{hx}^{(2,<)}(\mathbf{L}) - \overline{T}_{hx}^{(2,<)}(\mathbf{L})\}$, where

$$\overline{T}_{hx}^{(2,<)}(\mathbf{L}) = \sum_{i_1 < i_2} \mathrm{E}\,(\vartheta_{i_1}^{L_1}) \mathrm{E}\,(\vartheta_{i_2}^{L_2}) w_{ni_1}(x,h)^{L_1} w_{ni_2}(x,h)^{L_2}.$$

Note that $\mathbf{L} \in \{(1,5),(5,1)\}$ implies $\overline{T}_{hx}^{(2,<)}(\mathbf{L}) = 0$. Otherwise, as

$$\overline{T}_{hx}^{(2,<)}(\mathbf{L}) \leq \sum_{i_1} \nu_{X_{i_1}} w_{ni_1}(x,h)^{L_1} \times \sum_{i_2} \nu_{X_{i_2}} w_{ni_2}(x,h)^{L_2},$$

one uses equation (B.2) together with assumptions W_{12}–W_{13} to deduce that

$$\overline{T}_{hx}^{(2,<)}(\mathbf{L}) \leq \omega_2 (nh)^{2-L_1-L_2} \{\nu_x + h^2(\dot{\nu}_x + \ddot{\nu}_x)\}^2 = \omega_2 (nh)^{2-L_1-L_2} \nu_{xh}^2 \tag{B.3}$$

for some $\omega_2 < \infty$. Next, let υ_n be the integer part of $\frac{n}{2}$. One then has

$$\begin{aligned}
&T_{hx}^{(2,<)}(\mathbf{L}) - \overline{T}_{hx}^{(2,<)}(\mathbf{L}) \\
&\quad = \sum_{\ell=1}^{\pi_h} \sum_{i=1}^{n-\ell} \left[\mathrm{E}(\vartheta_i^{L_1} \vartheta_{i+\ell}^{L_2}) - \mathrm{E}(\vartheta_i^{L_1}) \mathrm{E}(\vartheta_{i+\ell}^{L_2})\right] w_{ni}(x,h)^{L_1} w_{n,i+\ell}(x)^{L_2} \\
&\qquad + \sum_{\ell=\pi_h+1}^{n-1} \sum_{i=1}^{n-\ell} \left[\mathrm{E}(\vartheta_i^{L_1} \vartheta_{i+\ell}^{L_2}) - \mathbb{E}\left(\vartheta_i^{L_1}\right) \mathbb{E}\left(\vartheta_{i+\ell}^{L_2}\right)\right] w_{ni}(x,h)^{L_1} w_{n,i+\ell}(x)^{L_2} \\
&\quad \leq \pi_h \left\{ \max_{1 \leq \ell \leq \pi_h} \sum_{i=1}^{n-\ell} \nu_{X_i} w_{ni}(x,h)^{L_1} w_{n,i+\ell}(x,h)^{L_2} \right\} \\
&\qquad + \left\{ \sum_{\ell=\pi_h+1}^{\upsilon_n} \alpha(\ell) \right\} \left\{ \max_{\pi_h+1 \leq \ell \leq \upsilon_n} \sum_{i=1}^{n-\ell} w_{ni}(x,h)^{L_1} w_{h,i+\ell}(x)^{L_2} \right\} \\
&\qquad + \left\{ \sum_{\ell=\upsilon_n+1}^{n-1} \alpha(\ell) \right\} \left\{ \max_{\upsilon_n < \ell \leq n-1} \sum_{i=1}^{n-\ell} w_{ni}(x,h)^{L_1} w_{h,i+\ell}(x)^{L_2} \right\}.
\end{aligned}$$

In view of equation (B.2), one uses assumption W_{11} and W_{13} to deduce that the first summand is $O(\pi_h h \nu_{xh} (nh)^{1-L_1-L_2}) = O(\nu_{xh}(nh)^{1-L_1-L_2})$. As assumption $\mathcal{A}_1$ holds, it follows that $\sum_{\ell=\gamma_n}^{n} \alpha(\ell) \sim O(\gamma_n^{-5})$ whenever $\gamma_n \to \infty$. Hence, from assumption W_{11}, the second summand is $O(h^6(nh)^{-L_1-L_2+1})$. Finally, as third summand is bounded by

$$\left\{ \sum_{\ell=\pi_h+1}^{\upsilon_n} \alpha(\ell) \right\} \left\{ \sum_{i=1}^{n} w_{ni}(x,h)^{L_1} \right\} \left\{ \max_{1\le i\le n} w_{ni}(x,h) \right\}^{L_2},$$

one uses assumption W_1 and W_{12} to derive that

$$T_{hx}^{(2,<)}(\mathbf{L}) - \overline{T}_{hx}^{(2,<)}(\mathbf{L}) \le \omega_2'(nh)^{1-L_1-L_2}[\nu_{xh} + h^6] \tag{B.4}$$

for some $\omega_2' > 0$. Wrapping last discussion, one concludes from equation (B.3) and (B.4) that

$$T_{hx}^{(2,<)}(\mathbf{L}) \le (nh)^{1-L_1-L_2} \left\{ \omega_2 (nh)\nu_{xh}^2 + \omega_2'(\nu_{xh} + h^6) \right\}.$$

Identical arguments yields the same bound for $T_{hx}^{(2,>)}(\mathbf{L})$. Therefore, upon setting $\overline{\omega}_2 = 2(\omega_2 + \omega_2')$, one obtains :

$$\begin{aligned} T_{hx}^{(2)}(\mathbf{L}) &\le \overline{\omega}_2\, (nh)^{1-L_1-L_2} \left\{ (nh)\nu_{xh}^2 + \nu_{xh} + h^6 \right\} \qquad \text{(B.5)} \\ &= \overline{\omega}_2\, (nh)^{-5} \left\{ (nh)\nu_{xh}^2 + \nu_{xh} + h^6 \right\}, \end{aligned}$$

where the last equality follows from the fact that $L_1 + L_2 = 6$. The case $j = 3$ calls for a special treatment. Denote for each $k = 1, 2$ the sets $\mathcal{T}_k = \{i_1 < i_2 < i_3 : \max(g_1, g_2) \le g_k\}$, where $g_k = i_{k+1} - i_k$ is the gap between two consecutive indices. Next, let $\mathcal{P}$ be the set of all permutations σ over the indices $\{1, 2, 3\}$, i.e

$$\mathcal{P} = \{\sigma : \{1,2,3\} \to \{1,2,3\} : \sigma(1) \ne \sigma(2) \ne \sigma(3)\}.$$

The introduction of this notation is justified by the fact that $\mathrm{E}T_{hx}^{(3)}(\mathbf{L}) \le \sum_{\sigma\in\mathcal{P}} W_{hx}^{(1,\sigma)}(\mathbf{L}) + W_{hx}^{(2,\sigma)}(\mathbf{L})$, where

$$W_{hx}^{(k,\sigma)}(\mathbf{L}) = \mathrm{E} \sum_{\mathcal{T}_k} \prod_{j=1}^{3} w_{ni_{\sigma(j)}}(x,h)^{L_{\sigma(j)}} \vartheta_{i_{\sigma(j)}}^{L_{\sigma(j)}}.$$

Note that the permutations play a similar role as the cases $i_1 < i_2$ and $i_2 < i_1$ required for the analysis of $T_{hx}^{(2)}$. However, as the treatment of $W_{hx}^{(k,\sigma)}$ is the same for all permutations, we consider only the case $(\sigma(1), \sigma(2), \sigma(3)) = (1, 2, 3)$ and we omit the σ in the notations.

Similarly as previously, one decomposes $W_{hx}^{(k)}(\mathbf{L}) = \overline{W}_{hx}^{(k)}(\mathbf{L}) + \{W_{hx}^{(k)}(\mathbf{L}) - \overline{W}_{hx}^{(k)}(\mathbf{L})\}$ with

$$\overline{W}_{hx}^{(k)}(\mathbf{L}) = \sum_{\mathcal{T}_k} \mathrm{E}\left(\prod_{j=1}^{k} \vartheta_{i_j}^{L_j}\right) \mathrm{E}\left(\prod_{j=k+1}^{3} \vartheta_{i_j}^{L_j}\right) \prod_{j=1}^{3} w_{ni_j}(x,h)^{L_j}.$$

We deal only with the term $\overline{W}_{hx}^{(1)}$, as the other case is identical. Note that $\overline{W}_{hx}^{(1)}(\mathbf{L}) = 0$ whenever $\mathbf{L} \in \{(1, b, 5-b) : 1 \leq b \leq 4\}$. Otherwise, observe that

$$\overline{W}_{hx}^{(1)}(\mathbf{L}) = \sum_{g_1=1}^{n-1} \sum_{g_2=1}^{g_1} \mathrm{E}(\vartheta_{i_2}^{L_2} \vartheta_{i_3}^{L_3}) w_{ni_2}(x,h)^{L_2} w_{ni_3}(x,h)^{L_3} \sum_{i_1=1}^{n-g_1} \mathrm{E}(\vartheta_{i_1}^{L_1}) w_{ni_1}(x,h)^{L_1}.$$

Proceeding as previously, one uses equation (B.2) and assumption W_{12}-W_{13} to obtain :

$$\sum_{i_1=1}^{n-g_1} \mathrm{E}(\vartheta_{i_1}^{L_1}) w_{ni_1}(x,h)^{L_1} \leq \omega_1 (nh)^{-L_1+1} \nu_{xh}.$$

Moreover, one deduces from equation (B.5) that

$$\sum_{g_1=1}^{n-1} \sum_{g_2=1}^{g_1} \mathrm{E}(\vartheta_{i_2}^{L_2} \vartheta_{i_3}^{L_3}) w_{ni_2}(x,h)^{L_2} w_{ni_3}(x,h)^{L_3}$$

$$\leq \overline{\omega}_2 (nh)^{1-L_2-L_3} \{(nh)\nu_{xh}^2 + \nu_{xh} + h^6\},$$

leading to

$$\overline{W}_{hx}^{(1)}(\mathbf{L}) \leq \omega_1 \overline{\omega}_2 (nh)^{-4} \{(nh)\nu_{xh}^3 + \nu_{xh}^2 + \nu_{xh} h^6\} \tag{B.6}$$

since the L_i's sum to 6. Finally,

$$
\begin{aligned}
&W^{(1)}_{hx}(\mathbf{L}) - \overline{W}^{(1)}_{hx}(\mathbf{L}) \\
&\quad = \sum_{\substack{\mathcal{T}_1 \\ g_1 \le \pi_h}} \left\{ \mathrm{E}\left(\prod_{j=1}^{3} \vartheta_{i_j}^{L_j} \right) - \mathrm{E}\left(\vartheta_{i_1}^{L_1}\right) \mathrm{E}\left(\vartheta_{i_2}^{L_2} \vartheta_{i_3}^{L_3}\right) \right\} \prod_{j=1}^{3} w_{ni_j}(x,h)^{L_j} \\
&\qquad + \sum_{\substack{\mathcal{T}_1 \\ g_1 > \pi_h}} \left\{ \mathrm{E}\left(\prod_{j=1}^{3} \vartheta_{i_j}^{L_j} \right) - \mathrm{E}\left(\vartheta_{i_1}^{L_1}\right) \mathrm{E}\left(\vartheta_{i_2}^{L_2} \vartheta_{i_3}^{L_3}\right) \right\} \prod_{j=1}^{3} w_{ni_j}(x,h)^{L_j} \\
&\quad \le \pi_h^2 \left\{ \max_{1 \le g_2 < g_1 \le \pi_h} \sum_{i_1=1}^{n-\ell} \nu_{X_{i_1}} \prod_{j=1}^{3} w_{ni_j}(x,h)^{L_j} \right\} \\
&\qquad + \sum_{\substack{\mathcal{T}_1 \\ \pi_h \le g_1 \le \upsilon_n}} \alpha(g_1) \prod_{j=1}^{3} w_{ni_j}(x,h)^{L_j} + \sum_{\substack{\mathcal{T}_1 \\ g_1 \ge \upsilon_n}} \alpha(g_1) \prod_{j=1}^{3} w_{ni_j}(x,h)^{L_j}.
\end{aligned}
$$

In view of equation (B.2), assumption W_{11} and W_{13}, the first summand is $O\{\nu_{xh}(nh)^{-5}\}$ since $\pi_h^2 h^2 \to 1$. The second summand is bounded by

$$
\begin{aligned}
&\sum_{g_1=\pi_h}^{\upsilon_n} (g_1+1)\alpha(g_1) \max_{1<\ell_1<\ell_2<\pi_h} \sum_{i=1}^{n} w_{ni}(x,h)^{L_1} w_{n,i+\ell_1}(x,h)^{L_2} w_{n,i+\ell+2}(x,h)^{L_3} \\
&\qquad\qquad + \sum_{g_1=\upsilon_n}^{n} (g_1+1)\alpha(g_1) \times \upsilon_n \max_{1 \le i \le n} w_{ni}(x,h)^3.
\end{aligned}
$$

As condition $\mathcal{A}_1$ is satisfied, $\sum_{g_1=\gamma_n}^{n}(g_1+1)\alpha(g_1) = O(\gamma_n^{-4})$ provided $\gamma_n \to \infty$. Thus, from W_1 and W_{11}, one obtains that the previous equation is $O\{\pi_h^{-4}h^2(nh)^{-5} + \upsilon_n^{-3}(nh)^{-6}\} = O\{(nh)^{-5}h^6\}$. It follows that

$$
W^{(1)}_{hx}(\mathbf{L}) - \overline{W}^{(1)}_{hx}(\mathbf{L}) = O\left\{(nh)^{-5}(\nu_{xh} + h^6)\right\}. \tag{B.7}
$$

Hence, one deduces from equation (B.6) and (B.7) that there exist a constant $\kappa_3 > 0$ such that

$$
W^{(1)}_{hx}(\mathbf{L}) \le \omega_3 (nh)^{-5}\{(nh)^2\nu_{xh}^3 + (nh)\nu_{xh}^2 + (nh^7)\nu_x + h^6\}.
$$

For sufficiently large n, $nh^7 < 1$ since $nh^5 < \infty$ and $h \to 0$. Therefore the factor nh^7 in front of ν_x can be omitted. As the case $W^{(2)}_{hx}(\mathbf{T})$ is totally identical, one concludes that

$$
T^{(3)}_{xh}(\mathbf{L}) \le \varpi_3 (nh)^{-5}\{(nh)^2\nu_{xh}^3 + (nh)\nu_{xh}^2 + \nu_x + h^6\}
$$

for some constant $\overline{\omega}_3$. Similar but long computations yields to the same bound but a possibly different constant for the cases $j = 4, 5, 6$, i.e

$$T_{xh}^{(j)}(\mathbf{L}) \leq \overline{\omega}_j (nh)^{-5}\{(nh)^2\nu_{xh}^3 + (nh)\nu_{xh}^2 + \nu_x + h^6\} \qquad \overline{\omega}_j < \infty,\ j = 4, 5, 6.$$

Collecting the bounds for each $T_{xh}^{(j)}(\mathbf{L})$ allows to conclude that there exist a global constant $\omega < \infty$ such that

$$\mathrm{E}|\mathbb{H}_{xh}|^6 \leq \omega \sum_{k=1}^{3} \{\nu_x + h^2(\dot{\nu}_x + \ddot{\nu}_x)\}^k (nh)^{-3+k} + (nh)^{-2}h^6.$$

Since Assumption $\mathcal{A}_2$ holds, $\dot{\nu}_x(A)$ and $\ddot{\nu}_x(A)$ are uniformly bounded. Moreover, since $\nu_x(A) \leq \mu_x(A)$,

$$\mathrm{E}\,|\mathbb{H}_{xh}(A)|^6 \leq (\omega + 1) \sum_{k=1}^{3} \left\{\mu_x(A) + h^2\right\}^k (nh)^{-3+k}.$$

B.3.3 Proof of Lemma 3.2

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 similar arguments as in the proof of Proposition 2.1. 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, 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

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

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

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

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.3.1. Analogously to that section observe that

$$
\begin{aligned}
\widetilde{Z}_{xn}(t,u) &- \widetilde{Z}_{xn}(\underline{t}_\gamma, \underline{u}_\gamma) \\
&= \{\widetilde{Z}_{xn}(t,u) - \widetilde{Z}_{xn}(t, \underline{u}_\gamma)\} + \{\widetilde{Z}_{xn}(t, \underline{u}_\gamma) - \widetilde{Z}_{xn}(\underline{t}_\gamma, \underline{u}_\gamma)\} \\
&\leq \{\widetilde{Z}_{xn}(t,\overline{u}_\gamma) - \widetilde{Z}_{xn}(t, \underline{u}_\gamma)\} + \{\widetilde{Z}_{xn}(t, \underline{u}_\gamma) - \widetilde{Z}_{xn}(\underline{t}_\gamma, \underline{u}_\gamma)\} \\
&\quad + \mathrm{E}\{Z_{xn}(t, \overline{u}_\gamma) - Z_{xn}(t, \underline{u}_\gamma)\}. \qquad \text{(B.8)}
\end{aligned}
$$

Starting with the last term of equation (B.8), using a Taylor expansion of F_{x_i} around z_t leads to :

$$
\begin{aligned}
\mathrm{E}\{Z_{xn}(t, \overline{u}_\gamma) - Z_{xn}(t, \underline{u}_\gamma)\} &= \sqrt{ng}(\overline{u}_\gamma - \underline{u}_\gamma) \\
&+ \left[\dot{F}_{z_t}\{F_{z_t}^{-1}(\overline{u}_\gamma)\} - \dot{F}_{z_t}\{F_{z_t}^{-1}(\underline{u}_\gamma)\}\right] \sqrt{ng} \sum_{i=1}^{n} w_{ni}(g, z_t)(X_i - z_t) \\
&+ \frac{1}{2}\sqrt{ng} \sum_{i=1}^{n} \left[\ddot{F}_{r_{ti}}\{F_{z_t}^{-1}(\overline{u}_\gamma)\} - \ddot{F}_{r_{ti}}\{F_{z_t}^{-1}(\underline{u}_\gamma)\}\right] w_{ni}(g, z_t)(X_i - z_t)^2,
\end{aligned}
$$

where r_{ti} lies between z_t and X_i. From assumptions W_9, W_{10} and $\mathcal{A}_4$, the previous equation is equal to

$$
\sqrt{ng}(\overline{u}_\gamma - \underline{u}_\gamma) + o(1)O(\sqrt{ng}g^2) = o(1).
$$

The last equality follows from the assumptions over the bandwidth parameters, ensuring that $\sqrt{ng}g^2 < \infty$, and the fact that the grid definition entails $\sqrt{ng}(\overline{u}_\gamma - \underline{u}_\gamma) = O\{(ng)^{-\gamma}\}$. This yields the negligibility of $\mathrm{E}\{Z_{xn}(t, \overline{u}_\gamma) - Z_{xn}(t, \underline{u}_\gamma)\}$.

Next we deal with the term $\widetilde{Z}_{xn}(t, \underline{u}_\gamma) - \widetilde{Z}_{xn}(\underline{t}_\gamma, \underline{u}_\gamma)$ in equation (B.8). Denote $\mathbb{F}_{zg} =$

$\sqrt{ng}(F_{zg} - \mathrm{E}\, F_{zg})$ and notice that $\mathbb{F}_{ztg}\{F_{z_t}^{-1}(u)\} = \widetilde{Z}_{xn}(t,u)$. Therefore, one writes

$$\begin{aligned}\widetilde{Z}_{xn}(t,\underline{u}_\gamma) - \widetilde{Z}_{xn}(\underline{t}_\gamma,\underline{u}_\gamma) &= \left[\mathbb{F}_{ztg}\left\{F_{z_t}^{-1}(\underline{u}_\gamma)\right\} - \mathbb{F}_{ztg}\left\{F_{z_{\underline{t}_\gamma}}^{-1}(\underline{u}_\gamma)\right\}\right] \\ &\quad + \left[\mathbb{F}_{ztg}\left\{F_{z_{\underline{t}_\gamma}}^{-1}(\underline{u}_\gamma)\right\} - \mathbb{F}_{jz_{\underline{t}_\gamma}g}\left\{F_{z_{\underline{t}_\gamma}}^{-1}(\underline{u}_\gamma)\right\}\right].\end{aligned}$$

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

$$\begin{aligned}&\sqrt{ng}|F_{ztg}(y) - F_{jz_{\underline{t}_\gamma}g}(y)| \\ &\qquad = \sqrt{ng}\left|\sum_{i=1}^{n} \mathbb{I}(Z_i \le y)\left\{w_{ni}(g,z_t) - w_{ni}(g,z_{\underline{t}_\gamma})\right\}\right| \\ &\qquad \le \sqrt{ng}\sup_{z\in I_x}\sum_{i=1}^{n}|w'_{ni}(z,g)| \times h(t-\underline{t}_\gamma) = O\left\{(ng)^{-\gamma}g^{-1}h\right\}.\end{aligned}$$

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

$$\begin{aligned}\widetilde{Z}_{xn}(t,\underline{u}_\gamma) - \widetilde{Z}_{xn}(\underline{t}_\gamma,\underline{u}_\gamma) &= \mathbb{F}_{ztg}\{F_{z_t}^{-1}(\underline{u}_\gamma)\} - \mathbb{F}_{ztg}\{F_{z_{\underline{t}_\gamma}}^{-1}(\underline{u}_\gamma)\} + o(1) \\ &= \mathbb{F}_{z_{\underline{t}_\gamma}g}\{F_{z_t}^{-1}(\underline{u}_\gamma)\} - \mathbb{F}_{z_{\underline{t}_\gamma}g}\{F_{z_{\underline{t}_\gamma}}^{-1}(\underline{u}_\gamma)\} + o(1).\end{aligned}$$

Using the same strategy with the first term of equation (B.8), one deduces that

$$\widetilde{Z}_{xn}(t,\overline{u}_\gamma) - \widetilde{Z}_{xn}(t,\underline{u}_\gamma) = \mathbb{F}_{z_{\underline{t}_\gamma}g}\{F_{z_t}^{-1}(\overline{u}_\gamma)\} - \mathbb{F}_{z_{\underline{t}_\gamma}g}\{F_{z_{\underline{t}_\gamma}}^{-1}(\underline{u}_\gamma)\} + o(1).$$

As assumption $\mathcal{A}_4$ implies that the function $z \mapsto F_z^{-1}$ is continuous in a neighborhood of x, one deduces that

$$\begin{aligned}&\left|\mathbb{F}_{z_{\underline{t}_\gamma}g}\left\{F_{z_t}^{-1}(\overline{u}_\gamma)\right\} - \mathbb{F}_{z_{\underline{t}_\gamma}g}\left\{F_{z_{\underline{t}_\gamma}}^{-1}(\underline{u}_\gamma)\right\}\right| \\ &\qquad \le \left|\mathbb{F}_{z_{\underline{t}_\gamma}g}\left\{F_{z_{\overline{t}_\gamma}}^{-1}(\overline{u}_\gamma)\right\} - \mathbb{F}_{z_{\underline{t}_\gamma}g}\left\{F_{z_{\underline{t}_\gamma}}^{-1}(\overline{u}_\gamma)\right\}\right| \\ &\qquad\qquad + \left|\mathbb{F}_{z_{\underline{t}_\gamma}g}\left\{F_{z_{\underline{t}_\gamma}}^{-1}(\overline{u}_\gamma)\right\} - \mathbb{F}_{z_{\underline{t}_\gamma}g}\left\{F_{z_{\underline{t}_\gamma}}^{-1}(\underline{u}_\gamma)\right\}\right| + o(1).\end{aligned}$$

In view of last discussion and of decomposition (B.8), one concludes that

$$\sup_{(t,u)\in\mathcal{I}}\left|\widetilde{Z}_{xn}(t,u)-\widetilde{Z}_{xn}(\underline{t}_\gamma,\underline{u}_\gamma)\right|$$

$$\leq\ 2\sup_{(t,u)\in T_n}\left|\mathbb{F}_{z_t g}\left\{F_{z_t}^{-1}(\overline{u}_\gamma)\right\}-\mathbb{F}_{z_t g}\left\{F_{z_t}^{-1}(\underline{u}_\gamma)\right\}\right|$$

$$+2\sup_{(t,u)\in T_n}\left|\mathbb{F}_{z_{\underline{t}_\gamma} g}\left\{F_{z_{\overline{t}_\gamma}}^{-1}(u)\right\}-\mathbb{F}_{z_{\underline{t}_\gamma} g}\left\{F_{z_{\underline{t}_\gamma}}^{-1}(u)\right\}\right|+o(1).$$

For $t_k=\frac{k}{\kappa_\gamma}$, denote

$$A_\gamma(i,t)=[F_{z_t}^{-1}(\frac{i-1}{\kappa_\gamma}),F_{jz_t}^{-1}(\frac{i}{\kappa_\gamma})]\ \text{ and }\ B_\gamma(i,k)=[F_{z_{t_k}}^{-1}(\frac{i}{\kappa_\gamma}),F_{z_{t_{k+1}}}^{-1}(\frac{i}{\kappa_\gamma})].$$

Moreover, write $\mathcal{G}_\gamma=\{0,1,\ldots,\kappa_\gamma^{-1}\}\times\{0,\pm1,\ldots,\pm\kappa_\gamma^{-1}\}$. Then from similar arguments as in the end of Section B.3.1, for sufficiently large n :

$$\mathfrak{W}_\delta(\widetilde{Z}_{xn},\mathcal{I})\ \leq\ 6\max_{(i,k)\in\mathcal{G}_\gamma}|\mathbb{F}_{z_{t_k}g}\{A_\gamma(i,t_k)\}|+6\max_{(i,k)\in\mathcal{G}_\gamma}|\mathbb{F}_{z_{t_k}g}\{B_\gamma(i,t_k)\}|.$$

For any interval $A=[a,b]\subset\mathbb{R}$, denote $\nu_z(A)=F_z(b)-F_z(a)$. On one hand, $\nu_{z_{t_k}}\{A_\gamma(i,t_k)\}=(ng)^{-1/2-\gamma}$. On the other hand,

$$\begin{aligned}\nu_{z_{t_k}}\{B_\gamma(i,k)\} &= \left|u-F_{z_{t_k}}\left\{F_{z_{t_{k+1}}}^{-1}(u)\right\}\right|\\ &= \left|\dot{F}_{z_{t_{k+1}}}\left\{F_{z_{t_{k+1}}}^{-1}(u)\right\}(z_{t_k}-z_{t_{k+1}})\right.\\ &\quad\left.+\frac{1}{2}\ddot{F}_{z_{t^\star}}\left\{F_{z_{t_{k+1}}}^{-1}(u)\right\}(z_{t_k}-z_{t_{k+1}})^2\right|,\end{aligned}$$

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

$$\begin{aligned}\mathrm{P}\left\{\mathfrak{W}_\delta\left(\widetilde{Z}_{xn},\mathcal{I}\right)\geq\epsilon\right\} &\leq \mathrm{P}\left[\max_{(i,k)\in\mathcal{G}_\gamma}\left|\mathbb{F}_{z_{t_k}g}\left\{A_\gamma(i,t_k)\right\}\right|\geq\frac{\epsilon}{12}\right]\\ &\leq \sum_{(i,k)\in\mathcal{G}_\gamma}\mathrm{P}\left[\left|\mathbb{F}_{z_{t_k}g}\left\{A_\gamma(i,t_k)\right\}\right|\geq\frac{\epsilon}{12}\right]\\ &\leq \frac{(ng)^{1+2\gamma}}{\epsilon^6}\left(\max_{(i,k)\in\mathcal{G}_\gamma}\mathrm{E}\big[\mathbb{F}_{z_{t_k}g}\left\{A_\gamma(i,t_k)\right\}\big]^6\right),\end{aligned}$$

where the last line follows from the use Markov inequality. As the assumptions of Lemma 2.4 are satisfied, identical computations as in section B.3.2 with ν_x being replace with $\nu_{z_{t_k}}$ enables to find a constant $\omega < \infty$ such that for any interval $A \in [0,1]$:

$$\mathrm{E}\{\mathbb{F}_{z_{t_k} g}(A)^6\} \leq \omega \sum_{k=1}^{3} \{\nu_{z_{t_k}}(A) + g^2\}^k (ng)^{-3+k} + (ng)^{-2} g^6.$$

Since $\nu_{z_{t_k}}\{A_\gamma(i,t_k)\} = (ng)^{-1/2-\gamma}$ and $\sqrt{ng}g^2 < \infty$, it follow that for sufficiently large n, $g^2 > (ng)^{-1/2-\gamma}$. The previous equation is therefore bounded by $8\omega(n^{-2} + n^{-1}g^2 + g^6)$. In follows that

$$\begin{aligned} \mathrm{P}\left\{\mathfrak{W}_\delta(\widetilde{Z}_{xn}, \mathfrak{I}) \geq \epsilon\right\} &\leq \frac{8\kappa}{\epsilon^6}(ng)^{1+2\gamma}(n^{-2} + n^{-1}g^2 + g^6) \\ &= \frac{8\kappa}{\epsilon^6}(n^{-1+2\gamma}g^{1+2\gamma} + n^{2\gamma}g^{3+2\gamma} + n^{1+2\gamma}g^{7+2\gamma}) \end{aligned}$$

Since $\sqrt{ng}g^2 < \infty$ implies $g = O(n^{-\tau})$ with $\tau \geq 1/5$, it follows that the latter is $o(1)$ upon taking $\gamma \in (0, r_\tau)$ with $r_\tau = \min\{1/2, \frac{3\tau}{2(1-\tau)}, \frac{7\tau-1}{2(1-\tau)}\}$. The lemma is therefore proven.

Annexe C

Nonparametric measures of local causality and tests of local non-causality in time series

C.1 Assumptions needed in Proposition 1, Corollary 1 and Proposition 1

C.1.1 Conditions on the weights

For simplicity, one uses in the sequel the notation

$$w_{ni}(x,h) = \mathcal{K}\left(\frac{Y_i - x}{h}\right) \quad \text{and} \quad w'_{ni}(x,h) = \frac{\partial}{\partial x} w_{ni}(x,h). \tag{C.1}$$

In addition, let

$$I_{xn} = \{j \in \{1,\dots,n\} : w_{nj}(x,h) > 0\} \quad \text{and} \quad J_{xn} = \left\{\min_{i \in I_{nx}} Y_i,\ \max_{i \in I_{nx}} Y_i\right\}.$$

The following assumptions are needed to establish Proposition 1, Corollary 1 and Proposition 1.

W_1. $\max_{1\le i\le n} |w_{ni}(x,h)| = O_{\mathrm{P}}\left((nh)^{-1}\right)$;

W_2. $\sum_{i=1}^{n} w_{ni}(x,h)(Y_i - x) = h^2 K_2 + o_{\mathrm{P}}\left((nh)^{-1/2}\right)$ for some $K_2 < \infty$;

W_3. $\sum_{i=1}^{n} w_{ni}(x,h)(Y_i - x)^2/2 = h^2 K_3 + o_{\mathrm{P}}\left((nh)^{-1/2}\right)$ for some $K_3 < \infty$;

W_4. $nh\sum_{i=1}^{n} \{w_{ni}(x,h)\}^2 = K_4 + o_{\mathrm{P}}(1)$ for some $K_4 > 0$;

W_5. $\max_{i\in I_{xn}} Y_i - \min_{i\in I_{xn}} Y_i = o_{\mathrm{P}}(1)$;

W_6. $\sup_{\xi\in J_{xn}} \sum_{i=1}^{n} |w'_{ni}(\xi,h)| = O_{\mathrm{P}}(h^{-1})$;

W_7. $\sup_{\xi\in J_{xn}} \sum_{i=1}^{n} \{w'_{ni}(\xi,h)\}^2 = O_{\mathrm{P}}(n^{-1}h^{-3})$;

W_8. For some finite constant C,

$$\mathrm{P}\left(\sup_{\xi\in J_{xn}} \max_{1\le i\le n} |w_{ni}(\xi,h)\,\mathbb{I}\left(|Y_i - x| > Ch\right)| > 0\right) = o_{\mathrm{P}}(1);$$

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

$$\sup_{\xi\in J_{xn}} \left|\sum_{i=1}^{n} w_{ni}(\xi,a_n)(Y_i - \xi) - a_n^2\, D_K\right| = o_{\mathrm{P}}(a_n^2);$$

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

$$\sup_{\xi\in J_{xn}} \left|\sum_{i=1}^{n} w_{ni}(\xi,a_n)(Y_i - z)^2 - a_n^2\, E_K\right| = o_{\mathrm{P}}(a_n^2).$$

W_{11}. $\sup_{\xi\in J_{xn}} \sum_{i=1}^{n} \{w_{ni}(\xi,h)\}^{L_1} = O_{\mathrm{P}}\left((nh)^{-L_1+1}\right)$;

W_{12}. For any $2 \le k \le 6$, define $V_k = \{1 \le \ell_2 < \cdots < \ell_k\}$ and $\boldsymbol{\ell} = (\ell_2, \ldots, \ell_k)$. Then, for any positive integers $L_1 + \cdots + L_k = r \le 6$:

$$\sup_{\xi\in J_{xn}} \max_{\boldsymbol{\ell}\in V_k} \sum_{i=1}^{n-\ell_k} \{w_{ni}(\xi,h)\}^{L_1} \prod_{j=2}^{k} \left\{w_{n,i+\ell_j}(\xi,h)\right\}^{L_j} = O_{\mathrm{P}}\left(\frac{h^{k-1}}{(nh)^{r-1}}\right)$$

and

$$\sup_{\xi \in J_{xn}} \max_{\boldsymbol{\ell} \in V_k} \sum_{i=1}^{n-\ell_k} (Y_i - \xi) \{w_{ni}(\xi, h)\}^{L_1} \prod_{j=2}^{k} \{w_{n,i+\ell_j}(\xi, h)\}^{L_j} = O_{\mathrm{P}}\left(\frac{h^{k+1}}{(nh)^{r-1}}\right).$$

C.1.2 Additional conditions

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

$\mathcal{A}_2$. $H_\xi^{Z \to Y}(y, z)$, $\dot{H}_\xi^{Z \to Y}(y, z)$ and $\ddot{H}_\xi^{Z \to Y}(y, z)$ exist and are continuous for $(\xi, y, z) \in J_{xn} \times \mathbb{R}^2$.

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

$\mathcal{A}_4$. For $j = 1, 2$, $F_{j\xi}\{F_{j\xi}^{-1}(u)\}$, $\dot{F}_{j\xi}\{F_{j\xi}^{-1}(u)\}$ and $\ddot{F}_{j\xi}\{F_{j\xi}^{-1}(u)\}$ exist and are continuous for $(\xi, u) \in J_{xn} \times [0, 1]$.

$\mathcal{A}_5$. $C_\xi^{Z \to Y}(u, v)$, $\dot{C}_\xi^{Z \to Y}(u, v)$ and $\ddot{C}_\xi^{Z \to Y}(u, v)$ exist and are continuous for $(\xi, u, v) \in J_{xn} \times [0, 1]^2$.

C.2 Proof of proposition 4

First note that $\sigma^2_{\Lambda,x} = \mathrm{Var}\{\Lambda'_{C_x^{\mathrm{Z}\to\mathrm{Y}}}(\mathbb{C}_x^{\mathrm{Z}\to\mathrm{Y}})\}$ can be written in terms of the covariance function of $\mathbb{C}_x^{\mathrm{Z}\to\mathrm{Y}}$. Specifically, letting $\Omega = \Lambda'_{C_x^{\mathrm{Z}\to\mathrm{Y}}} \circ \Lambda'_{C_x^{\mathrm{Z}\to\mathrm{Y}}}$, one obtains from the linearity of $\Lambda'_{C_x{}^{\mathrm{Z}\to\mathrm{Y}}}$ that

$$\begin{aligned} \sigma^2_{\Lambda,x} &= \mathrm{Cov}\left\{\Lambda'_{C_x^{\mathrm{Z}\to\mathrm{Y}}}(\mathbb{C}_x^{\mathrm{Z}\to\mathrm{Y}}), \Lambda'_{C_x^{\mathrm{Z}\to\mathrm{Y}}}(\mathbb{C}_x^{\mathrm{Z}\to\mathrm{Y}})\right\} \\ &= \Omega\left[\mathrm{Cov}\left\{\mathbb{C}_x^{Z\to Y}(u, v), \mathbb{C}_x^{\mathrm{Z}\to\mathrm{Y}}(u', v')\right\}\right]. \end{aligned}$$

Moreover, one retrieves from the work of Bouezmarni *et al.* (2016) that the covariance function of $\mathbb{C}_x^{\mathrm{Z}\to\mathrm{Y}}$ can itself be expressed as

$$K_4 \, \Upsilon \circ \Upsilon \left\{ C_x^{\mathrm{Z}\to\mathrm{Y}}(u \wedge u', v \wedge v') - C_x^{\mathrm{Z}\to\mathrm{Y}}(u,v) C_x^{\mathrm{Z}\to\mathrm{Y}}(u',v') \right\},$$

where for any $\delta \in \ell^\infty([0,1]^2)$ and $\Delta \in \ell^\infty([0,1]^4)$,

$$\begin{aligned} \Upsilon(\delta)(u,v) &= \delta(u,v) - C_x^{[1]}(u,v)\delta(u,1) - C_x^{[2]}(u,v)\delta(1,v), \\ \Upsilon \circ \Upsilon(\Delta) &= \Upsilon \left[\Upsilon \left\{ \Delta(\cdot,\cdot), u', v' \right\} \right]. \end{aligned}$$

The functional Υ is a linear functional that corresponds to the Hadamard derivative of the copula mapping $H_x^{\mathrm{Z}\to\mathrm{Y}} \mapsto H_x^{\mathrm{Z}\to\mathrm{Y}} \circ \{F_{1x}^{-1}, F_{2x}^{-1}\}$ (more details can be found in Bouezmarni *et al.* (2016)). Proceeding as in Appendix E.1, first define $w_{ni}(x,h) = \mathcal{K}\{(Y_i - x)/h\}$ for each $i \in \{2, \ldots, n+1\}$. From the conditions on the estimators of the partial derivatives of $C_x^{\mathrm{Z}\to\mathrm{Y}}$, one can readily show that for each $i \in \{1, \ldots, n\}$,

$$\sup_{(u,v)\in[0,1]^2} \left| \widehat{L}_{x,i}(u,v) - L'_{x,i}(u,v) \right| = o_{\mathrm{P}}(1),$$

where

$$L'_{x,i}(u,v) = \Upsilon \left[\mathbb{I} \left\{ Y_i \le F_{1xh}^{-1}(u), Z_{i-1} \le F_{2xh}^{-1}(v) \right\} \right].$$

Hence, by the definition of $\widehat{\sigma}^2_{\Lambda,x}$, one can write

$$
\begin{aligned}
\frac{\widehat{\sigma}^2_{\Lambda,x}}{\widehat{K}_4} &= \sum_{i=2}^{n+1} w_{ni}(x,h) \left\{ \Upsilon\left(\widehat{L}_{x,i}\right) - \sum_{j=2}^{n+1} w_{nj}(x,h)\,\Upsilon\left(\widehat{L}_{x,j}\right) \right\}^2 \\
&= \sum_{i=2}^{n+1} w_{ni}(x,h) \left\{ \Upsilon\left(L'_{x,i}\right) - \sum_{j=2}^{n+1} w_{nj}(x,h)\,\Upsilon\left(L'_{x,j}\right) \right\}^2 + o_{\mathrm{P}}(1) \\
&= \sum_{i=2}^{n+1} w_{ni}(x,h) \left\{ \Upsilon\left(L'_{x,i}\right) \right\}^2 - \left\{ \sum_{j=2}^{n+1} w_{nj}(x,h)\,\Upsilon\left(L'_{x,j}\right) \right\}^2 + o_{\mathrm{P}}(1) \\
&= \Omega \left\{ \Upsilon \circ \Upsilon \left\{ \sum_{i=2}^{n+1} w_{ni}(x,h)\, \mathbb{I}\left\{ Y_i \le F^{-1}_{1xh}(u \wedge u'), Z_{i-1} \le F^{-1}_{2xh}(v \wedge v') \right\} \right\} \right] \\
&\quad - \Omega \left\{ \Upsilon \circ \Upsilon \left\{ \sum_{j=2}^{n+1} w_{nj}(x,h)\, \mathbb{I}\left\{ Y_j \le F^{-1}_{1xh}(u), Z_{j-1} \le F^{-1}_{2xh}(v) \right\} \right\}^2 \right] + o_{\mathrm{P}}(1) \\
&= \Omega \left[\Upsilon \circ \Upsilon \left\{ C^{\mathrm{Z}\to\mathrm{Y}}_{xh}(u \wedge u', v \wedge v') - C^{\mathrm{Z}\to\mathrm{Y}}_{xh}(u,v) C^{\mathrm{Z}\to\mathrm{Y}}_{xh}(u',v') \right\} \right] + o_{\mathrm{P}}(1).
\end{aligned}
$$

Since $C^{\mathrm{Z}\to\mathrm{Y}}_{xh}$ is consistent for $C^{\mathrm{Z}\to\mathrm{Y}}_{x}$ and the fact that $\widehat{K}_4 = K_4 + o_{\mathrm{P}}(1)$, one can conclude from the Continuous Mapping Theorem that $\widehat{\sigma}^2_{\Lambda,x} \to \sigma^2_{\Lambda,x} = K_4\,\Omega(\gamma)$.

Annexe D

Estimation of a conditional copula when a variable is subject to random right censoring

D.1 Proofs of main results

In this section, all the expectations of the form $\mathrm{E}\{f(T_i, Y_{2i}, C_i)\}$ have to be understood as taken conditional upon $X = x_i$. Formally, for any $1 \leq i \leq n$,

$$\mathrm{E}\{f(T_i, Y_{2i}, C_i)\} = \int \int f(t, y, c) \mathrm{d}F_{x_i}(t, y) \mathrm{d}G_{x_i}(c),$$

whenever the left-hand side of the integral exists.

D.1.1 I.i.d representation for $\sqrt{nh}(F_{xh}^{(\mathbf{rc})} - F_x)$

We start by observing that $F_{xh}^{(\text{rc})} = \sum_{i=1}^n \delta_i f(t, y, T_{1i}, Y_{2i}, G_{xg})$, where for any $(t, y) \in \mathcal{T}_t$, $(v, v') \in \mathbb{R}^2$ and for any function $G : \mathbb{R} \to [0, 1)$:

$$f(t, y, v, v', G) = \frac{\mathbb{I}(v \leq t, v' \leq y)}{1 - G(v)}.$$

To provide an i.i.d representation for $\sqrt{nh}(F_{xh}^{(\text{rc})} - F_x)$, we apply the ideas of van der Vaart & Wellner (2007). To this end, we introduce the operator $\mathbb{E}(\cdot)$ defined over the set of random variables of the form $\delta_1 f(t, y, T_1, Y_{21}, G), \ldots, \delta_n f(t, y, T_n, Y_{2n}, G)$ such that

$$\mathbb{E}\left\{\delta_i f(t, y, T_i, Y_{2i}, G)\right\} = \int f(t, y, v, v', G) \mathrm{d}H_{x_i}^u(v, v'),$$

whenever the right-hand side of the integral exists. Observe that when the function G is fixed (non random), it follows that

$$\mathbb{E}\left\{\delta_i f(t, y, T_i, Y_{2i}, G)\right\} = \mathrm{E}\left\{\delta_i f(t, y, T_i, Y_{2i}, G)\right\}.$$

Then, we consider the following decomposition :

$$\sqrt{nh}(F_{xh}^{(\text{rc})} - F_x) = \{A_{xh} - \mathbb{E}(A_{xh})\} + B_{xh} + \mathbb{E}(A_{xh}),$$

where

$$A_{xh}(t, y) = \sqrt{nh} \sum_{i=1}^n \mathbb{I}(T_i \leq t, Y_{2i} \leq y)\, w_{ni}(x, h)\delta_i \left\{\frac{1}{1 - G_{xg}(T_i)} - \frac{1}{1 - G_x(T_i)}\right\}$$

and

$$\begin{aligned} B_{xh}(t, y) &= \sqrt{nh}\left\{\sum_{i=1}^n \mathbb{I}(T_i \leq t, Y_{2i} \leq y) w_{ni}(x, h) \frac{\delta_i}{1 - G_x(T_i)} - F_x(t, y)\right\} \\ &= \sqrt{nh} \sum_{i=1}^n w_{ni}(x, h) \mathcal{J}_{ix}^{(1)}(t, y). \end{aligned}$$

Firstly, we remark that B_{xh} is, by construction, an i.i.d representation. Secondly, we show, in D.1.1, that $A_{xh} - \mathbb{E}(A_{xh})$ is asymptotically negligible. Finally, in D.1.1, we provide an i.i.d representation for $\mathbb{E}(A_{xh})$.

Asymptotic negligibility of $A_{xh} - \mathbb{E}(A_{xh})$

Recall that $\mathcal{T}_{\mathfrak{t}} = [0, \mathfrak{t}] \times \mathbb{R}$, where $0 < \mathfrak{t} < \tau_{H_x}$. Let $\epsilon > 0$ such that $G_x(\mathfrak{t}) + \epsilon < 1$, and set $\mathfrak{t}_\epsilon = G_x(\mathfrak{t}) + \epsilon$. Introduce the set of functions $\mathcal{G}_{t_\epsilon} := \{G : \mathbb{R} \to [0,1]$ nondecreasing and $G(\mathfrak{t}) < \mathfrak{t}_\epsilon\}$. For $(t, y, G) \in \mathcal{T}_{\mathfrak{t}} \times \mathcal{G}_{\mathfrak{t}_\epsilon}$, define the stochastic processes

$$Z_{hi}(t, y, G) = \sqrt{nh}\,\mathbb{I}(T_i \le t, Y_{2i} \le y) w_{ni}(x, h) \frac{\delta_i}{1 - G(T_i)}$$

and $Z_{xh} = \sum_{i=1}^{n} Z_{hi}$. The process Z_{xh} can be viewed as a process indexed by the family of functions from $\mathbb{R}^2 \times \{0, 1\} \to \mathbb{R}$ given by

$$\mathcal{F} = \left\{ (v, v', w) \mapsto \mathbb{I}(v \le t, v' \le y) \frac{w}{1 - G(v)}, (t, y) \in \mathcal{T}_{\mathfrak{t}},\, G \in \mathcal{G}_{\mathfrak{t}_\epsilon} \right\}.$$

Hence, each function $f \in \mathcal{F}$ may be formally identified by a triplet (t, y, G). The introduction of the process Z_{xh} is motivated by the fact that $A_{xh}(t, h) = Z_{xh}(t, y, G_{xg}) - Z_{xh}(t, y, G_x)$. While the ϵ*–enlargement* in the definition of the class $\mathcal{G}_{\mathfrak{t}_\epsilon}$ might appear overdone, it is however required to guaranty that G_{xg} asymptotically fits into $\mathcal{G}_{\mathfrak{t}_\epsilon}$.

Finally, we equip the index set $\mathcal{F}$ with a semimetric ρ_{F_x} defined for $f = (t, y, G)$ and $f' = (t', y', G')$ as

$$\rho_{F_x}(f, f') = |F_{1x}(t) - F_{1x}(t')| + |F_{2x}(y) - F_{2x}(y')| + \sup_{z \in [0, \mathfrak{t}]} |G(z) - G'(z)|.$$

Notice that $(\mathcal{F}, \rho)$ is totally bounded for ρ_{F_x}, as $\mathbb{I}\{v \le t, v' \le y\} \frac{w}{1-G(v)} \le \frac{\mathbb{I}\{v \le \mathfrak{t}\}}{1-G(v)} \le \frac{1}{1-\mathfrak{t}_\epsilon} < \infty$. Moreover, $\mathfrak{F} := \frac{1}{1-\mathfrak{t}_\epsilon}$ is an enveloppe function for $\mathcal{F}$.

In fact, as Assumptions W_1-W_4-W_6 and $\mathcal{C}_1$ are satisfied, we conclude from 1 that the process $\overline{Z}_{xh} := Z_{xh} - \mathrm{E}Z_{xh}$ indexed by $(\mathcal{F}, \rho_{F_x})$ is asymptotically ρ_{F_x}–equicontinuous.

This implies that for any $\eta > 0$ and $\eta' > 0$, there exists $\delta > 0$ such that

$$\limsup_{n\to\infty} \mathrm{P}^{\star}\left\{\sup_{\rho_{F_x}(f,f')<\delta} |\overline{Z}_n(f') - \overline{Z}_n(f)| > \eta\right\} < \eta'.$$

For a particular choice of weight system $w_{n1}(x,\cdot),\ldots,w_{nn}(x,\cdot)$, it is shown in Van Keilegom & Veraverbeke (1996) that as long as conditions $(\mathcal{C}_1)$ and $(\mathcal{C}_3)$ are satisfied for H_x, G_x and H_x^c, then there exist constants $C_1, C_2 > 0$ that depends on the weight functions such that whenver $\epsilon > \max\{C_1 g^2, C_2 \frac{1}{\sqrt{ng}}\}$, we have

$$\mathrm{P}\left\{\sup_{t\in[0,\mathfrak{t}]} |G_{xg}(t) - G_x(t)| > \epsilon\right\} \leq C_3 n g \epsilon e^{-C_4 n g \epsilon^2} \tag{D.1}$$

for some constants C_3 and C_4 that rely on $\mathfrak{t}$ and the weights. It can be shown for instance by using Lemma 3 in Omelka *et al.* (2013) that this result still holds for general $w_{n1},\ldots,w_{nn}$, at the cost of perhaps enlarging the constants provided this weight system satisfy assumptions W_1-W_5. Hence, we obtain

$$\sup_{t\in[0,\mathfrak{t}]} |G_{xg}(t) - G_x(t)| = O((ng)^{-1/2}(\log n)^{1/2}) \qquad \mathrm{P} - \text{a.s.} \tag{D.2}$$

In view of Equation (D.2), we conclude that for any $\epsilon > 0$, $G_{xg} \in \mathcal{G}_\epsilon$ with probability 1. Moreover, as $\rho_{F_x}\{(t,y,G_{xg}),(t,y,G_x)\} = \sup_{t\in[0,\mathfrak{t}]} |G_{xg}(t) - G_x(t)|$, we deduce that for sufficiently large n :

$$\sup_{(t,y)\in\mathcal{T}_\mathfrak{t}} \left|\sqrt{nh}\,[A_{xh}(t,y) - \mathbb{E}\{A_{xh}(t,y)\}]\right| \leq \sup_{\rho_{F_x}(f,f')<\delta} |\overline{Z}_n(f') - \overline{Z}_n(f)| \xrightarrow{\mathrm{P}} 0.$$

which concludes the proof.

Asymptotic representation of $\mathbb{E}(A_{xh})$

Note that the random variable $D_{xh}(t,y) = \mathbb{E}\{A_{xh}(t,y)\}$ can be rewritten as

$$\begin{aligned} D_{xh}(t,y) &= \sqrt{nh}\int_0^t\int_0^y \frac{G_{xg}(v)-G_x(v)}{\{1-G_x(v)\}\{1-G_{xh}(v)\}}\sum_{i=1}^n w_{ni}(x,h)\mathrm{d}H^u_{x_i}(v,v') \\ &= \sqrt{nh}\int_0^t \frac{G_{xg}(v)-G_x(v)}{\{1-G_x(v)\}\{1-G_{xg}(v)\}}\sum_{i=1}^n w_{ni}(x,h)\int_0^y \mathrm{d}H^u_{x_i}(v,v') \\ &= \sqrt{hg^{-1}}\int_0^t \frac{\mathbb{G}_{xg}(v)}{\{1-G_x(v)\}\{1-G_{xg}(v)\}}\sum_{i=1}^n w_{ni}(x,h)h^u_{x_i}(v,y)\mathrm{d}v, \end{aligned}$$

where $h^u_{x_i}(v,y) = \frac{\partial}{\partial v}H^u_{x_i}(v,y)$ and $\mathbb{G}_{xg} = \sqrt{ng}\{G_{xg}-G_x\}$. Next, we write

$$\widetilde{D}_{xh}(t,y) = \sqrt{hg^{-1}}\int_0^t \frac{\mathbb{G}_{xg}(v)}{\{1-G_x(v)\}^2}\sum_{i=1}^n w_{ni}(x,h)h^u_{x_i}(v,y)\mathrm{d}v$$

and we introduce for any function $g \in l^\infty([0,\mathfrak{t}])$ the functional $\Lambda_x : l^\infty([0,\mathfrak{t}]) \to l^\infty(\mathcal{T}_\mathfrak{t})$ defined as

$$\Lambda_x(g)(t,y) = \int_0^t \frac{g(v)}{\{1-G_x(v)\}^2}h^u_x(v,y)\mathrm{d}v.$$

The asymptotic representation of D_{xh} will follow from the representation of $\sqrt{hg^{-1}}\Lambda_x(\mathbb{G}_{xg})$ and the asymptotic negligibility of the two terms $\widetilde{D}_{xh} - \sqrt{hg^{-1}}\Lambda_x(\mathbb{G}_{xg})$ and $D_{xh} - \widetilde{D}_{xh}$.

For the asymptotic representation of $\sqrt{hg^{-1}}\Lambda_x(\mathbb{G}_{xg})$, let

$$\begin{aligned} g_{ix}(t) = \{1-G_x(t)\}\Bigg[&\int_0^t \frac{\mathbb{I}(T_i\le v)-H_{1x}(v)}{\{1-H_{1x}(v)\}^2}\mathrm{d}H^c_x(v) \\ &+\frac{\mathbb{I}(T_i\le t,\delta_i=0)-H^c_x(t)}{1-H_{1x}(t)} \\ &-\int_0^t \frac{\mathbb{I}(T_i\le v,\delta_i=0)-H^c_x(v)}{\{1-H_{1x}(v)\}^2}\mathrm{d}H_{1x}(v)\Bigg]. \end{aligned}$$

It can be shown for instance by using Lemma 3 of Omelka *et al.* (2013) that the result stated in Theorem 2.1 in Van Keilegom & Veraverbeke (1997) still holds as long as conditions $(\mathcal{C}_1)$–$(\mathcal{C}_5)$ are satisfied for G_x and G^u_x, provided assumptions W_1–W_5 on the weight functions are fulfilled. Hence, we conclude that, uniformly in $t \in [0, \mathfrak{t}]$, $\sqrt{hg^{-1}}\mathbb{G}_{xg} = \sqrt{nh}\sum_{i=1}^{n} w_{ni}(x, g)g_{ix} + o_{\text{a.s}}(1)$ as long as $\frac{ng^5}{\log(n)} < \infty$. As the map $\Lambda_x(\cdot)$ is linear and continuous, the continuous mapping theorem implies that

$$\sqrt{hg^{-1}}\Lambda_x\left(\mathbb{G}_{xg}\right) = \sqrt{nh}\sum_{i=1}^{n} w_{ni}(x, g)\Lambda_x(g_{ix}) + o_{\text{a.s}}(1).$$

Furthermore, from switching the order of integration and further computations, we show that $\Lambda_x(g_{ix})(t, y) = \mathcal{J}_{ix}^{(2)}(t, y)$.

To show the asymptotic negligibility of $\widetilde{D}_{xh} - \sqrt{hg^{-1}}\Lambda_x(\mathbb{G}_{xg})$, we observe that for any $(t, y) \in \mathcal{T}_{\mathfrak{t}}$,

$$\begin{aligned}\widetilde{D}_{xh}(t, y) - \sqrt{hg^{-1}}\Lambda_x(\mathbb{G}_{xg})(t, y) &= \int_0^t \frac{G_{xg}(v) - G_x(v)}{\{1 - G_x(v)\}^2} R_{xh}(v, y)\mathrm{d}v \\ &\leq \frac{\sup_{v\in[0,\mathfrak{t}]} |G_{xg}(v) - G_x(v)|}{\{1 - G_x(\mathfrak{t})\}^2} \int_0^t |R_{xh}(v, y)|\mathrm{d}v,\end{aligned}$$

where

$$R_{xh}(v, y) = \sqrt{nh}\sum_{i=1}^{n} w_{ni}(x, h)\{h_x^u(v, y) - h_{x_i}^u(v, y)\}.$$

Condition $(\mathcal{C}_6)$ together with W_5 allows the Taylor expansion

$$\begin{aligned}R_{xh}(v, y) &= \dot{h}_x^u(v, y)\sqrt{nh}\sum_{i=1}^{n}(x - x_i)w_{ni}(x, h) \\ &\quad -\frac{1}{2}\sqrt{nh}\sum_{i=1}^{n}(x - x_i)^2 w_{ni}(x, h)\ddot{h}_{z_i}^u(v, y),\end{aligned}$$

where z_i lies between x_i and x. Hence, from Assumptions W_2, W_3, W_6 and $(\mathcal{C}_6)$, we obtain that

$$\sqrt{nh}\int_0^t |R_{xh}(u, v)|\mathrm{d}v = O(1).$$

In view of the previous discussion, we use Equation (D.2) to obtain that uniformly in $(t, y) \in \mathcal{T}_{\mathfrak{t}}$, $\widetilde{D}_{xh} - \sqrt{hg^{-1}}\Lambda_x(\mathbb{G}_{xg}) = o_{\text{a.s.}}(1)O(1)$.

Finally, notice that

$$
\begin{aligned}
&D_{xh} - \widetilde{D}_{xh} \\
&\quad = \sqrt{nh}\int_0^t \frac{\{G_{xh}(v) - G_x(v)\}^2}{\{1 - G_x(v)\}^2\{1 - G_{xh}(v)\}} \sum_{i=1}^n w_{ni}(x,h) h^u_{x_i}(v,y)\mathrm{d}v \\
&\quad \leq \sqrt{nh}\frac{\sup_{v\in[0,\mathfrak{t}]}\{G_{xg}(v) - G_x(v)\}^2}{\{1 - G_x(\mathfrak{t})\}^2\{1 - G_{xg}(\mathfrak{t})\}} \int_0^t |\sum_{i=1}^n w_{ni}(x,h) h^u_{x_i}(v,y)|\mathrm{d}v.
\end{aligned}
$$

Hence, using Equation (D.2), we deduce that the latter is $o_{\text{a.s.}}(1)$.

D.1.2 Weak convergence of $\sqrt{nh}\{F^{(\mathbf{rc})}_{xh} - F_x\}$

From Theorem 1, we have

$$
\sqrt{nh}(F^{(\mathrm{rc})}_{xh} - F_x) = \sqrt{nh}\sum_{i=1}^n \left\{ w_{ni}(x,h)\mathfrak{J}^{(1)}_{ix} + w_{ni}(x,g)\mathfrak{J}^{(2)}_{ix} \right\} + o_{\mathrm{P}}(1).
$$

We start by showing that the sequence $\sqrt{nh}\sum_{i=1}^n w_{ni}(x,h)\mathfrak{J}^{(1)}_{ix}$ is asymptotically gaussian. Then, regarding the assumptions over the bandwidth h and g, the asymptotic representation of

$$
\sqrt{nh}\sum_{i=1}^n w_{ni}(x,g)\mathfrak{J}^{(2)}_{ix} \quad \text{and} \quad \sqrt{nh}\sum_{i=1}^n \{w_{ni}(x,h)\mathfrak{J}^{(1)}_{ix} + w_{ni}(x,g)\mathfrak{J}^{(2)}_{ix}\}
$$

are discussed.

Weak convergence of $\sqrt{nh}\sum_{i=1}^n w_{ni}(x,h)\mathfrak{J}^{(1)}_{ix}$

Let's prove that the process $\mathbb{J}^{(1)}_{xh} := \sqrt{nh}\sum_{i=1}^n \mathfrak{J}^{(1)}_{ix}$ converges to the gaussian process $\mathbb{J}^{(1)}_x + Kb^{(1)}_x$ over $\mathcal{T}_{\mathfrak{t}}$.

First, the tightness of the sequence $\mathbb{J}^{(1)}_{xh}$ can be checked using similar arguments as in D.1.1. Second, from direct computation,

$$
\begin{aligned}
&\mathrm{E}\left\{\mathbb{J}^{(1)}_{xh}(t,y)\right\}\\
&= \sqrt{nh}\sum_{i=1}^{n} w_{ni}(x,h)\mathrm{E}\left\{\mathrm{E}\left(\delta_i \mid Y_{1i}, Y_{2i}\right)\frac{\mathbb{I}(Y_{1i}\le t, Y_{2i}\le y)}{1-G_x(Y_{1i})} - F_x(t,y)\right\}\\
&= \sqrt{nh}\sum_{i=1}^{n} w_{ni}(x,h)\mathrm{E}\left\{\frac{1-G_{x_i}(Y_{1i})}{1-G_x(Y_{1i})}\mathbb{I}(Y_{1i}\le t, Y_{2i}\le y) - F_x(t,y)\right\}\\
&= \sqrt{nh}\sum_{i=1}^{n} w_{ni}(x,h)\{F_{x_i}(t,y) - F_x(t,y)\} + o(1)\\
&+ \sqrt{nh}\sum_{i=1}^{n} w_{ni}(x,h)(x_i-x)\mathrm{E}\left\{\frac{-\dot{G}_x(Y_{1i})}{1-G_x(Y_{1i})}\mathbb{I}(Y_{1i}\le t, Y_{2i}\le y)\right\}\\
&+ \frac{1}{2}\sqrt{nh}\sum_{i=1}^{n} w_{ni}(x,h)(x_i-x)^2\mathrm{E}\left\{\frac{-\ddot{G}_x(Y_{1i})}{1-G_x(Y_{1i})}\mathbb{I}(Y_{1i}\le t, Y_{2i}\le y)\right\}, \qquad \text{(D.3)}
\end{aligned}
$$

where last equality follows from a Taylor expansion of the function $z \mapsto 1-G_z$ around $z=x$ together with the fact that from Assumptions $\mathcal{C}_1$ and $\mathcal{C}_3$, G_x,$\dot{G}_x$ and $\ddot{G}_x$ are uniformly continuous over $\mathcal{T}_{\mathrm{t}}$. We calculate the first term. In fact,

$$
\begin{aligned}
&\sqrt{nh}\sum_{i=1}^{n} w_{ni}(x,h)\{F_{x_i}(t,y) - F_x(t,y)\}\\
&\qquad = \sqrt{nh}\sum_{i=1}^{n} w_{ni}(x,h)(x_i-x)\dot{F}_x(t,y)\\
&\qquad\quad + \frac{1}{2}\ddot{F}_x(t,y)\sqrt{nh}\sum_{i=1}^{n} w_{ni}(x,h)(x_i-x)^2 + o(1), \qquad \text{(D.4)}
\end{aligned}
$$

where last equality follows from a Taylor expansion of the function $z \mapsto F_z$ around $z=x$ together with the fact that F_x,$\dot{F}_x$ and $\ddot{F}_x$ are uniformly continuous over $V(x)\times\mathcal{T}_{\mathrm{t}}$, where $V(x)$ is some neighborhood of x, as staten from Assumptions $\mathcal{C}_1$–$\mathcal{C}_3$. From Assumptions W_2 and W_3, Equation (D.4) is equal to

$$
\sqrt{nh}h^2\left\{K_2\dot{F}_x(t,y) + \frac{K_3}{2}\ddot{F}_x(t,y) + o(1)\right\}.
$$

where the constants K_2 and K_3 are defined via Assumptions W_2 and W_3. As $\sqrt{nh}h^2 \to K$, Equation (D.4) reduces to

$$K\left\{K_2\dot{F}_x(t,y) + \frac{K_3}{2}\ddot{F}_x(t,y)\right\} + o(1).$$

Next, for the second and third terms of Equation (D.3) , we denote

$$B_z^{(1)}(t,y) = \int_0^t \frac{-\dot{G}_x(v)}{1-G_x(v)} f_z(v,y)\mathrm{d}v$$

and

$$B_z^{(2)}(t,y) = \int_0^t \frac{-\ddot{G}_x(v)}{1-G_x(v)} f_z(v,y)\mathrm{d}v.$$

The second and third terms of Equation (D.3) can re-written as

$$\begin{aligned}\sqrt{nh}\sum_{i=1}^n w_{ni}(x,h)(x_i-x)B_{x_i}^{(1)}(t,y) + \sqrt{nh}\sum_{i=1}^n w_{ni}(x,h)(x_i-x)^2 B_{x_i}^{(2)}(t,y)&\\ = K\left\{K_2 B_x^{(1)}(t,y)K_3 + o(1)B_x^{(2)}(t,y)\right\} + o(1).&\qquad \text{(D.5)}\end{aligned}$$

Putting together Equations (D.3),(D.4) and (D.5) concludes to show that

$$\mathrm{E}\left\{\mathbb{J}_{xh}^{(1)}\right\} = Kb_x^{(1)} + o(1).$$

Next, as Assumption W_4 is satisfied, we use a similar strategy to obtain

$$\begin{aligned}&\mathrm{Cov}\{\mathbb{J}_{xh}^{(1)}(t,y), \mathbb{J}_{xh}^{(1)}(t',y')\}\\ &\quad = nh\sum_{i=1}^n w_{ni}(x,h)^2\left[\mathrm{E}\left\{\frac{\mathbb{I}(T_i\le t\wedge t', Y_{2i}\le y\wedge y')\delta_i}{\{1-G_{x_i}(T_i)\}^2}\right\}\right.\\ &\qquad\qquad \left. -\mathrm{E}\left\{\frac{\mathbb{I}(T_i\le t, Y_{2i}\le y)\delta_i}{1-G_{x_i}(T_i)}\right\}\mathrm{E}\left\{\frac{\mathbb{I}(T_i\le t', Y_{2i}\le y')\delta_i}{1-G_{x_i}(T_i)}\right\}\right]\\ &\quad = K_4\left\{\int_0^{t\wedge t'}\frac{f_x(v,y\wedge y')}{1-G_x(v)}\mathrm{d}v - F_x(t,y)F_x(t',y')\right\} + o(1).\end{aligned}$$

In view of the tightness of $\mathbb{J}_{xh}^{(1)}$, the fact that

$$\mathrm{Cov}\{\mathbb{J}_{xh}^{(1)}(t,y), \mathbb{J}_{xh}^{(1)}(t',y')\} \to \mathrm{Cov}\{\mathbb{J}_x^{(1)}(t,y), \mathbb{J}_x^{(1)}(t',y')\},$$

and using Theorem 2.11.1 of van der Vaart & Wellner (1996), we conclude that $\mathbb{J}_{xh}^{(1)}$ converges weakly to a gaussian process whose representation matches the one of $\mathbb{J}_x^{(1)} + Kb_x^{(1)}$.

Asymptotic representation of $\sqrt{nh}\sum_{i=1}^n w_{ni}(x,g)\mathcal{J}_{ix}^{(2)}$

From the definition of g_{ix} and of $\Lambda_x(\cdot)$ in D.1.1, we ca re-write

$$\sum_{i=1}^n w_{ni}(x,g)\mathcal{J}_{ix}^{(2)} = \sum_{i=1}^n w_{ni}(x,g)\Lambda_x(g_{ix}) = \Lambda_x\left\{\sum_{i=1}^n w_{ni}(x,g)g_{ix}\right\}.$$

Now because $\Lambda_x(\cdot)$ is linear, we deduce that

$$\mathrm{E}\left[\Lambda_x\left\{\sum_{i=1}^n w_{ni}(x,g)g_{ix}\right\}\right] = \Lambda_x\left[\mathrm{E}\left\{\sum_{i=1}^n w_{ni}(x,g)g_{ix}\right\}\right].$$

From Van Keilegom & Veraverbeke (1997), using Assumptions W_2–W_3 together with $\mathcal{C}_1$ and $\mathcal{C}_3$, be obtain that

$$\begin{aligned}\mathrm{E}\left\{\sum_{i=1}^n w_{ni}(x,g)g_{ix}(t)\right\} &= g^2\{1-G_x(t)\}\int_0^t \frac{K_2\dot{H}_{1x}(s)+\frac{K_3}{2}\ddot{H}_{1x}(s)}{\{1-H_{1x}(s)\}^2}\mathrm{d}H_{1x}^c(s)\\ &\quad + \frac{\mathrm{d}\{K_2\dot{H}_x^c(s)+\frac{K_3}{2}\ddot{H}_x^c(s)\}}{\{1-H_{1x}(s)\}} + o(g^2).\end{aligned}$$

Hence, integrating by parts leads to :

$$\Lambda_x\left[\mathrm{E}\left\{\sum_{i=1}^n w_{ni}(x,g)g_{ix}\right\}\right](t,y) = g^2\mathrm{E}\left\{\mathbb{J}_x^{(2)}(t,y)\right\} + o(g^2). \tag{D.6}$$

Furthermore, provided Assumption W_4 is satisfied, Van Keilegom & Veraverbeke (1997) also gives us

$$\begin{aligned}&\mathrm{Cov}\left\{\textstyle\sum_{i=1}^n w_{ni}(x,g)g_{ix}(t), \sum_{i=1}^n w_{ni}(x,g)g_{ix}(t')\right\}\\ &\qquad = \frac{K_4}{ng}\{1-G_x(t)\}\{1-G_x(t')\}\int_0^{t\wedge t'}\frac{\mathrm{d}H_x^c(s)}{\{1-H_{1x}(s)\}^2} + o(\frac{1}{ng}).\end{aligned}$$

Then,

$$\mathrm{Cov}\left[\Lambda_x\left\{\sum_{i=1}^n w_{ni}(x,g)g_{ix}\right\}(t,y),\Lambda_x\left\{\sum_{i=1}^n w_{ni}(x,g)g_{ix}\right\}(t',y')\right]$$

$$\begin{aligned} &= \int_0^t\int_0^{t'}\frac{\mathrm{Cov}\{g_{ix}(v),g_{ix}(v')}{\{1-G_x(v)\}^2\{1-G_x(v')\}^2}h_x^u(v,y)h_x^u(v',y')\mathrm{d}v\mathrm{d}v' \\ &= \frac{K_4}{ng}\int_0^{t\wedge t'}\int_s^t\int_s^{t'}\frac{h_x^u(v,y)h_x^u(v',y')\mathrm{d}v\mathrm{d}v'}{\{1-G_x(v)\}\{1-G_x(v')\}}\frac{\mathrm{d}H_{1x}^c(s)}{\{1-H_{1x}(s)\}^2}+o\left(\frac{1}{ng}\right). \end{aligned} \quad \text{(D.7)}$$

From the fact that $\int_s^t\frac{h_x^u(v,y)}{\{1-G_x(v)\}}\mathrm{d}v=\int_s^t f_x(v,y)\mathrm{d}v=F_x(t,y)-F_x(s,y)$, the non negligible term of Equation (D.7) is equal to $\mathrm{Cov}\{\mathbb{J}_x^{(2)}(t,y),\mathbb{J}_x^{(2)}(t',y')\}$. Since $\sqrt{nh}g^2<\infty$, and $\frac{h}{g}=O(1)$, we deduce from Equations (D.6) and (D.7) that the sequence $\sum_{i=1}^n w_{ni}(x,g)\mathcal{J}_{ix}^{(2)}$ is asymptotically tight on $\mathcal{T}_{\mathfrak{t}}$.

Asymptotic normality of $\sqrt{nh}\sum_{i=1}^n w_{ni}(x,h)\{\mathcal{J}_{ix}^{(1)}+\mathcal{J}_{ix}^{(2)}\}$

Upon setting $\mathcal{J}_{xh}:=\sqrt{nh}\sum_{i=1}^n w_{ni}(x,h)\{\mathcal{J}_{ix}^{(1)}+\mathcal{J}_{ix}^{(2)}\}$, we compute

$$\begin{aligned} \mathrm{Cov}\{\mathcal{J}_{xh}(t,s),\mathcal{J}_{xh}(t',s')\} &= nh\sum_{i=1}^n w_{ni}(x,h)^2\mathrm{Cov}\{\mathcal{J}_{ix}^{(1)}(t,s),\mathcal{J}_{ix}^{(1)}(t',s')\} \\ &+ nh\sum_{i=1}^n w_{ni}(x,h)^2\mathrm{Cov}\{\mathcal{J}_{ix}^{(1)}(t,s),\mathcal{J}_{ix}^{(2)}(t',s')\} \\ &+ nh\sum_{i=1}^n w_{ni}(x,h)^2\mathrm{Cov}\{\mathcal{J}_{ix}^{(1)}(t',s'),\mathcal{J}_{ix}^{(2)}(t,s)\} \\ &+ nh\sum_{i=1}^n w_{ni}(x,h)^2\mathrm{Cov}\{\mathcal{J}_{ix}^{(2)}(t,s),\mathcal{J}_{ix}^{(2)}(t',s')\}. \end{aligned}$$

We obtain

$\sum_{i=1}^{n} w_{ni}(x,h)^2 \text{Cov}\{\mathfrak{J}_{ix}^{(1)}(t,y), \mathfrak{J}_{ix}^{(2)}(t',y')\} = o_{\text{P}}(\frac{1}{nh})$

$$-\frac{K_4}{nh}\{F_{2x}(y') - F_x(t',y')\}\int_0^{t'} \frac{F_x(t\wedge v, y) - H_{1x}(v)F_x(t,y)}{\{1-H_{1x}(v)\}\{1-G_x(v)\}}\text{d}H_{1x}^c(v)$$
$$-\frac{K_4}{nh}F_x(t,y)\int_0^{t'} \frac{1-H_{1x}^u(v)}{\{1-H_{1x}(v)\}\{1-G_x(v)\}}h_x^u(v,y')\text{d}v - F_x(t,y)F_x(t',y')$$
$$+\frac{K_4}{nh}F_x(t,y)\{F_{2x}(y') - F_x(t',y')\}\int_0^{t'} \frac{H_{1x}^u(v)}{\{1-H_{1x}(v)\}\{1-G_x(v)\}}\text{d}H_{1x}(v).$$

Hence, $\text{Cov}\{\mathfrak{J}_{xh}(t,s), \mathfrak{J}_{xh}(t',s')\} = \text{Cov}\{\mathfrak{J}_x(t,s), \mathfrak{J}_x(t',s')\} + o(1)$. The weak convergence follows from similar arguments as the ones expose in D.1.1.

D.2 Weak convergence of $\mathbb{B}_{xh}^{(\mathbf{rc})}$

This section follows along the lines of Theorem 1 in Veraverbeke *et al.* (2011). First, one decompose

$$\sqrt{nh}(\mathbb{C}_{xh}^{(\text{rc})} - C_x) = \{\widetilde{A}_{xh} - \mathbb{E}\,(\widetilde{A}_{xh})\} + \{\widetilde{A}'_{xh} - \mathbb{E}(\widetilde{A}'_{xh})\} + \widetilde{B}_{xh} + \mathbb{E}\,(\widetilde{A}_{xh}) + \mathbb{E}\,(\widetilde{A}'_{xh}),$$

where

$$\begin{aligned}\widetilde{A}_{xh}(u,v) &= A_{xh}\left\{F_{1xh}^{(\text{rc})-1}(u), F_{2xh}^{-1}(u)\right\},\\ \widetilde{A}'_{xh}(u,v) &= \sqrt{nh}\sum_{i=1}^{n} w_{ni}(x,h)\Big[\mathfrak{J}_{ix}^{(1)}\{F_{1xh}^{(\text{rc})-1}(u), F_{2xh}^{-1}(v)\}\\ &\qquad -\mathfrak{J}_{ix}^{(1)}\{F_{1x}^{-1}(u), F_{2x}^{-1}(v)\}\Big],\end{aligned}$$

and

$$\widetilde{B}_{xh}(u,v) = \sqrt{nh}\sum_{i=1}^{n} w_{ni}(x,h)\mathfrak{J}_{ix}^{(1)}\{F_{1x}^{-1}(u), F_{2x}^{-1}(v)\}$$

It will now be shown that both $\{\widetilde{A}_{xh} - \mathbb{E}\,(\widetilde{A}_{xh})\}$ and $\{\widetilde{A}'_{xh} - \mathbb{E}\,(\widetilde{A}'_{xh})\}$ are asymptotically negligible. Then, it will be establish that $\mathbb{E}\,(\widetilde{A}_{xh})$ asymptotically equals

$$\sqrt{nh}\sum_{i=1}^{n} w_{ni}(g,h)\mathfrak{J}_{ix}^{(2)}(F_{1x}^{-1}(u), F_{2x}^{-1}(v)\}.$$

Finally, the asymptotic representation of $\mathbb{E}(\widetilde{A}'_{xh})$ will be given.

D.2.1 Asymptotic negligibility of $\sqrt{nh}\{\widetilde{A}_{xh}-\mathbb{E}\,(\widetilde{A}_{xh})\}$ and $\sqrt{nh}\{\widetilde{A}'_{xh}-\mathbb{E}\,(\widetilde{A}'_{xh})\}$

First, a weak consequence of Corollary 1 is that uniformly in $(t,y) \in \mathfrak{T}_{\mathfrak{t}}$:

$$F^{(\mathrm{rc})}_{1xh}(t) - F_{1x}(t) = o_{\mathrm{P}}(1) \quad \text{and} \quad F_{2xh}(y) - F_{2x}(y) = o_{\mathrm{P}}(1).$$

Second, it follows from assumption W_1 that, for any sufficiently small $\epsilon > 0$,

$$\begin{aligned} \lim_{n\to\infty} \mathrm{P}\left\{F^{-1}_{1x}(u-\epsilon) \le F^{(\mathrm{rc})-1}_{1xh}(u) \le F^{-1}_{1x}(u+\epsilon), u \in [0, H_{1x}(\mathfrak{t})]\right\} &= 1, \\ \lim_{n\to\infty} \mathrm{P}\left\{F^{-1}_{2x}(v-\epsilon) \le F^{-1}_{2xh}(v) \le F^{-1}_{1x}(v+\epsilon), v \in [0,1]\right\} &= 1. \qquad \text{(D.8)} \end{aligned}$$

Now we recall the definition of Z_{xh} in Lemma 1, and we note that

$$\widetilde{A}_{xh}(u,v) = Z_{xh}\{F^{(\mathrm{rc})-1}_{1xh}(u), F^{-1}_{2xh}(v), G_{xg}\} - Z_{xh}\{F^{(\mathrm{rc})-1}_{1xh}(u), F^{-1}_{2xh}(v), G_x\}$$

and

$$\widetilde{A}'_{xh}(u,v) = Z_{xh}\{F^{(\mathrm{rc})-1}_{1xh}(u), F^{-1}_{2xh}(v), G_x\} - Z_{xh}\{F^{-1}_{1xh}(u), F^{-1}_{2x}(v), G_x\}.$$

For $f_1 = (F^{(\mathrm{rc})-1}_{1xh}(u), F^{-1}_{2xh}(v), G_{xh})$, $f_2 = (F^{(\mathrm{rc})-1}_{1xh}(u), F^{-1}_{2xh}(v), G_x)$ and $f_3 = (F^{-1}_{1x}(u), F^{-1}_{2x}(v), G_x)$, it follows from Equations (D.2) and (D.8) that

$$\rho_{F_x}(f_1,f_2) \to 0 \quad \text{and} \quad \rho_{F_x}(f_2,f_3) \to 0.$$

The negligibility of $\widetilde{A}_{xh} - \mathbb{E}(A_{xh})$ and $\widetilde{A}'_{xh} - \mathbb{E}(A'_{xh})$ is then ensured by Lemma 1 and identical arguments as the ones use in the end of D.1.1.

D.2.2 Asymptotic representation of $\mathbb{E}(\widetilde{A}_{xh})$

For any small $\epsilon > 0$ such that $H_{1x}(\mathfrak{t} + \epsilon) < 1$, one obtains from Equation (D.8) that uniformly in $u \in [0, H_{1x}(\mathfrak{t})]$ and with probability 1, $F_{1xh}^{(rc)-1}(u) \in [0, \mathfrak{t} + \epsilon]$. Thus, we get from D.1.1 that uniformly in $(u, v) \in \widetilde{\mathfrak{T}}_{\mathfrak{t}}$,

$$\mathbb{E}(\widetilde{A}_{xh}) \;=\; \sqrt{nh} \sum_{i=1}^{n} w_{ni}(x, g) \mathcal{J}_{xh}^{(2)} \left\{ F_{1xh}^{(rc)-1}(u), F_{2xh}^{-1}(v) \right\} + o_{\mathrm{P}}(1).$$

Now we let $t_{xh} = F_{1xh}^{(rc)-1}(u)$, $t_x = F_{1x}^{-1}(u)$, and we define y_{xh} and y_x analogously. Recall, from D.1.1, that

$$\begin{aligned} \mathcal{J}_{xh}^{(2)}(t_{xh}, y_{xh}) &= \Lambda_x(g_{ix})(t_{xh}, y_{xh}) \\ &= \int_0^{t_{xh}} \frac{g_{ix}(v)}{\{1 - G_x(v)\}^2} h_x^u(v, y_{xh}). \end{aligned}$$

Hence, from the mean value theorem, we have

$$\mathcal{J}_{xh}^{(2)}(t_{xh}, y_{xh}) - \mathcal{J}_{xh}^{(2)}(t_x, y_{xh}) \;=\; \frac{g_{ix}(t_{xh}^{\star})}{\{1 - G_x(t_{xh}^{\star})\}^2} h_x^u(t_{xh}^{\star}, y_{xh})(t_{xh} - t_x),$$

for some $t_{xh}^{\star}$ between t_{xh} and t_x. Further, since the g_{ix}'s are bounded random variables, and since $\frac{h}{g} = O(1)$, we can show that

$$\sup_{t,y \in \mathfrak{T}_{\mathfrak{t}}} \left| \sqrt{nh} \sum_{i=1}^{n} w_{ni}(x, g) \frac{g_{ix}(t)}{\{1 - G_x(t)\}^2} h_x^u(t, y) \right| = O_{\mathrm{P}}(1).$$

Thus, Equation (D.8) implies that

$$\sqrt{nh} \sum_{i=1}^{n} w_{ni}(x, g) \left\{ \mathcal{J}_{xh}^{(2)}(t_{xh}, y_{xh}) - \mathcal{J}_{xh}^{(2)}(t_x, y_{xh}) \right\} = o_{\mathrm{P}}(1).$$

Similarly, we have

$$\sqrt{nh} \sum_{i=1}^{n} w_{ni}(x, g) \left\{ \mathcal{J}_{xh}^{(2)}(t_x, y_{xh}) - \mathcal{J}_{xh}^{(2)}(t_x, y_x) \right\} = o_{\mathrm{P}}(1).$$

D.2.3 Asymptotic representation of $\mathbb{E}(A'_{xh})$

From the definition of $\mathbb{E}(\cdot)$, we can write

$$\mathbb{E}(A'_{xh})(u,v) = \sqrt{nh}\sum_{i=1}^{n} w_{ni}(x,h)F_{x_i}\left\{F_{1xh}^{(rc)-1}(u), F_{2xh}^{-1}(v)\right\} - C_x(u,v).$$

Conditions $\mathcal{C}_1$, $\mathcal{C}_3$ and W_2–W_3 allow to mimick the proof of Theorem 1 in Veraverbeke *et al.* (2011) to obtain, uniformly in $(u,v) \in \widetilde{\mathfrak{T}}_{\mathrm{t}}$, that

$$\mathbb{E}(A'_{xh})(u,v) = \sqrt{nh}\left(C_x\left[F_{1x}\{F_{1xh}^{(rc)-1}(u)\}, F_{2x}\{F_{2xh}^{-1}(v)\}\right] - C_x(u,v)\right).$$

Substituting v by 1 in the previous Equation with the asymptotic negligibility of $\widetilde{A}'_{xh} - \mathbb{E}A'_{xh}$ leads to

$$\sqrt{nh}\left[F_{1x}\left\{F_{1xh}^{(rc)-1}(u)\right\} - u\right] = -\sqrt{nh}\left[F_{1xh}^{(rc)}\left\{F_{1x}^{-1}(u)\right\}\right\} - u\right].$$

Morevoer, from Veraverbeke *et al.* (2011), we have

$$\sqrt{nh}\left[F_{2x}\left\{F_{2xh}^{-1}(v)\right\} - v\right] = -\sqrt{nh}\left[F_{2xh}\left\{F_{2x}^{-1}(v)\right\}\right\} - v\right].$$

Then, following the end of the proof of Theorem 1 in Veraverbeke *et al.* (2011) yields, uniformly over $\widetilde{\mathfrak{T}}_{\mathrm{t}}$, to

$$\begin{aligned}\mathbb{E}(A'_{xh})(u,v) &= -\sqrt{nh}\mathbb{C}_x^{[1]}(u,v)\left[F_{1xh}^{(rc)}\left\{F_{1x}^{-1}(u)\right\}\right\} - u\right] \\ &\quad -\sqrt{nh}\mathbb{C}_x^{[2]}(u,v)\left[F_{2xh}\left\{F_{2x}^{-1}(v)\right\}\right\} - v\right].\end{aligned}$$

D.3 Auxiliary Lemma

The following lemma is required in D.1.1 to establish the i.i.d representation for $\mathbb{F}_{xh}^{(\mathrm{rc})}$.

Lemma 1 *Recall the definition of $Z_{xh}, \mathcal{F}$ and ρ_{F_x} at the beginning of D.1.1. Suppose Assumptions W_1,W_4 and W_6 are satisfied, and that the maps $z \mapsto F_{1z}$ and $z \mapsto F_{2z}$ are uniformly continuous for all z in a neighborhood of x. Then, process $\overline{Z}_{xh} := Z_{xh} - \mathrm{E}Z_{xh}$ indexed by $(\mathcal{F}, \rho_{F_x})$ is asymptotically ρ_{F_x}–equicontinuous.*

Proof : From Theorem 2.11.1 of van der Vaart & Wellner (1996), we can conclude that Z_{xh} is tight if the following requirements hold :

$(\mathcal{R}_1)$ $\sum_{i=1}^{n} \mathrm{E}\left(\|Z_{hi}\|_{\mathcal{F}}^2\right) \mathbb{I}\left\{\|Z_{hi}\|_{\mathcal{F}} > \eta\right\} \to 0$ for any $\eta > 0$;

$(\mathcal{R}_2)$ $\sup_{\rho_{F_x}(f,f')<\delta_n} \sum_{i=1}^{n} \mathrm{E}\left[\left\{Z_{hi}(f) - Z_{hi}(f')\right\}^2\right] \to 0$ for every $\delta_n \downarrow 0$.

$(\mathcal{R}_3)$ $\int_0^{\delta_n} \left\{\log N(\epsilon, \mathcal{F}, d_n)\right\}^{1/2} \mathrm{d}\epsilon \xrightarrow{\mathrm{P}} 0$ for every $\delta_n \downarrow 0$, where $N(\epsilon, \mathcal{F}, d_n)$ is the covering number of the set $\mathcal{F}$ with respect to the random semi-metric

$$d_n^2(f, f') = \sum_{i=1}^{n} \left\{Z_{hi}(f) - Z_{hi}(f')\right\}^2 .$$

In the latter, $\|\cdot\|_{\mathcal{F}}$ stands for the supremum norm over $\mathcal{F}$.

$(\mathcal{R}_1)$: Because $Z_{hi}(f) = 0$ whenever $Y_{1i} > t$, and since G is nondecreasing, we have

$$Z_{hi}(f) \quad \leq \quad \sqrt{nh}\frac{w_{ni}(x,h)}{1 - G(\mathsf{t})} \leq \sqrt{nh}\frac{w_{ni}(x,h)}{1 - \mathsf{t}_\epsilon}.$$

From Assumption W_1, it follows that the latter is $o(1)$. Hence, for any $\eta > 0$, one can find $N > 0$ such that for all $n \geq N$: $\max_{1 \leq i \leq n} Z_{hi}(f) < \eta$, proving that requirement $\mathcal{R}_1$ is fulfiled.

$(\mathcal{R}_2)$: Assume wlog that $t \leq t'$ and $y \leq y'$. It is useful to note that when $\delta_i = 1$, it follows that $\mathbb{I}(T_i \leq t)\delta_i = \mathbb{I}(Y_{1i} \leq t)$ whereas $\mathbb{I}(T_i \leq t)\delta_i = 0$ if $\delta_i = 0$. As a consequence, $\mathbb{I}(T_i \leq t)\delta_i \leq \mathbb{I}(Y_{1i} \leq t)$. Then,

$$
\begin{aligned}
&Z_{hi}(f') - Z_{hi}(f) \\
&\quad = \sqrt{nh}\mathbb{I}\left(t < T_i \le t', y < Y_{2i} \le y'\right) \frac{\delta_i w_{ni}(x,h)}{1 - G'(T_i)} \\
&\qquad + \sqrt{nh}\mathbb{I}\left(T_i \le t, Y_{2i} \le y\right) \delta_i w_{ni}(x,h) \frac{G(T_i) - G'(T_i)}{\{1 - G(T_i)\}\{1 - G'(T_i)\}} \\
&\quad \le \sqrt{nh} \frac{w_{ni}(x,h)}{1 - \mathfrak{t}_\epsilon} \left\{ \mathbb{I}\left(t < Y_{1i} \le t', y < Y_{2i} \le y'\right) \delta_i + \|G - G'\|_{[0,\mathfrak{t}]} \right\}.
\end{aligned}
$$

From the last equation, one directly obtains

$$
\begin{aligned}
&\sup_{\rho_{F_x}(f,f') < \delta_n} \textstyle\sum_{i=1}^n \mathrm{E}\left\{ (Z_{hi}(f') - Z_{hi}(f))^2 \right\} \\
&\quad \le \sup_{\rho_{F_x}(f,f') < \delta_n} nh \sum_{i=1}^n \frac{w_{ni}(x,h)^2}{(1 - \mathfrak{t}_\epsilon)^2} \left\{ |F_{1x_i}(t') - F_{1x_i}(t)| + |F_{2x_i}(y') - F_{2x_i}(y)| \right\} \\
&\qquad + nh \sum_{i=1}^n \frac{w_{ni}(x,h)^2}{(1 - \mathfrak{t}_\epsilon)^2} \left(\|G - G'\|_{[0,\mathfrak{t}]} + \|G - G'\|^2_{[0,\mathfrak{t}]} \right) \\
&\quad \le \sup_{\rho_{F_x}(f,f') < \delta_n} \max_{i \in I_{nx}} \left[|F_{1x_i}(t') - F_{1x_i}(t)| + |F_{2x_i}(y') - F_{2x_i}(y)| \right] \\
&\qquad \times nh \sum_{i=1}^n \frac{w_{ni}(x,h)^2}{(1 - \mathfrak{t}_\epsilon)^2} + nh \sum_{i=1}^n \frac{w_{ni}(x,h)^2}{(1 - \mathfrak{t}_\epsilon)^2} \left(\|G - G'\|_{[0,\mathfrak{t}]} + \|G - G'\|^2_{[0,\mathfrak{t}]} \right).
\end{aligned}
$$

From Assumption W_4, we obtain that $nh \sum_{i=1}^n w_{ni}(x,h)^2 = O(1)$. Moreover, as Assumption W_6 holds together with the uniform continuity of the maps $z \mapsto F_{1z}$ and $z \mapsto F_{2z}$, we deduce that the latter display is bounded by $O(1)\{2\delta_n + o(1)\}$. Hence, requirement $\mathcal{R}_2$ is fulfilled as $\delta_n \to 0$.

($\mathcal{R}_3$) : To show the last requirement, the goal is to apply Lemma 2.11.6 of van der Vaart & Wellner (1996). To do this, three conditions must be verified. First, we rewrite

$$
\{Z_{hi}(f') - Z_{hi}(f)\}^2 = \int \left\{ \frac{\mathbb{I}(v \le t, v' \le y)}{1 - G(v)} - \frac{\mathbb{I}(v \le t', v' \le y')}{1 - G'(v)} \right\}^2 \mathrm{d}\mu_{ni},
$$

where $\mu_{ni} = nh w_{ni}(x,h)^2 \delta_i \mathbb{I}(v = Y_{1i}, v' = Y_{2i})$. Hence, the process Z_n is *measurelike* with respect to the random measure μ_{ni} (see van der Vaart & Wellner (1996), Section 2.11).

Second, as Assumption W_4 holds, $\sum_{i=1}^{n} \int \mathfrak{F} \mathrm{d}\mu_{ni} = nh \sum_{i=1}^{n} \frac{w_{ni}(x,h)^2}{1-\mathrm{t}_\epsilon} = O(1)$. Third, it is required to show that the class $\mathcal{F}$ satisfy the uniform entropy condition (2.11.5) of van der Vaart & Wellner (1996). To show that it is indeed the case, let

$$\mathcal{F}_1 = \{(v, v', w) \mapsto \mathbb{I}\{v \leq t, v' \leq y\}, (t, y) \in \mathcal{T}_{\mathrm{t}}, w \in \{0, 1\}\}$$

and $\mathcal{F}_2$ be the class of monotone and bounded functions over $[0, \frac{1}{1-\mathrm{t}_\epsilon}]$. Now we observe that $\mathcal{F} \subset \mathcal{F}_1\mathcal{F}_2 = \{f = f_1 f_2, f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2\}$. As $\mathcal{F}_1$ is a VC-class and $\mathcal{F}_2$ is a VC-hull class for sets with enveloppe functions respectively $\mathfrak{F}_1 = 1$ and $\mathfrak{F}_2 = \frac{1}{1-\mathrm{t}_\epsilon}$, an application Lemma 2.6.20 of van der Vaart & Wellner (1996) allows to conclude that $\mathcal{F}_1\mathcal{F}_2$ is VC-hull class for sets with envelope function $\mathfrak{F}_1 \times \mathfrak{F}_2 = \mathfrak{F}$. Therefore, the uniform entropy condition is fullfiled. As a result, the conclusion Lemma 2.11.6 of van der Vaart & Wellner (1996) applies which proves the requirement $\mathcal{R}_3$.

D.4 Auxiliary results

The following Proposition establishes that $F_{1xh}^{(\mathrm{rc})}$ coincide with the conditional Kaplan-Meier estimator for F_{1x} when the bandwidth parameters g and h in the definition of $F_{1xh}^{(\mathrm{rc})}$ are equal.

Proposition 2 *Consider the conditional Kaplan-Meier estimator given by :*

$$\begin{aligned} F_{1xh}^{(KM)}(t) &= 1 - \left\{ \prod_{T_{(i)} \leq t} \left(1 - \frac{w_{n[i]}(x,h)}{1 - F_{1xh}(T_{(i)}-)} \right)^{\delta_{[i]}} \right\} \\ &= 1 - \left\{ \prod_{T_{(i)} \leq t} \left(1 - \frac{w_{n[i]}(x,h)}{1 - \sum_{k=1}^{i-1} w_{n[k]}(x,h)} \right)^{\delta_{[i]}} \right\}. \end{aligned}$$

Then, for any $t \in \mathbb{R}$, $F_{1xh}^{(KM)}(t) = F_{1xh}^{(rc)}(t)$ when the bandwidth parameters g and h in the definition of $F_{1xh}^{(rc)}$ are equal.

Proof : First set $T_{(0)} = 0$ and $T_{n+1} = \infty$. Second, in order to proove the result, it suffices to show that for any $0 \le k \le n+1$, $F_{1xh}^{(\mathrm{rc})}(T_{(k)}) = F_{1xh}^{(\mathrm{KM})}(T_{(k)})$. To do this, we use the following induction argument.

Basis : Trivially, $F_{1xh}^{(\mathrm{rc})}(0) = F_{1xh}^{(\mathrm{KM})}(0) = 0$. Moreover,

$$F_{1xh}^{(\mathrm{rc})}(T_{(1)}) = \frac{\delta_{[1]} w_{n[1]}(x,h)}{\big(1 - w_{n[1]}(x,h)\big)^{1-\delta_{[1]}}} = \delta_{[1]} w_{n[1]}(x,h)$$

and

$$F_{1xh}^{(\mathrm{KM})}(T_{(1)}) = 1 - \big(1 - w_{n[1]}(x,h)\big)^{\delta_{[1]}} = \delta_{[1]} w_{n[1]}(x,h).$$

Hence, the Basis step is verified.

Induction step : Assuming that the equality $F_{1xh}^{(\mathrm{rc})}(t) = F_{1xh}^{(\mathrm{KM})}(t)$ holds for $t = T_{(0)}$ up to $t = T_{(k)}$, let's show that the equality is verified for $t = T_{(k+1)}$. From direct computations,

$$\begin{aligned} &F_{1xh}^{(\mathrm{KM})}(T_{(k+1)}) \\ &\quad = 1 - \prod_{i=1}^{k} \left(1 - \frac{w_{n[i]}(x,h)}{1 - F_{1xh}(T_{(i-1)})}\right)^{\delta_{[i]}} \left(1 - \frac{w_{n[k+1]}(x,h)}{1 - F_{1xh}(T_{(k)})}\right)^{\delta_{[k+1]}} \\ &\quad = 1 + \{F_{1xh}^{(\mathrm{rc})}(T_{(k)}) - 1\} \left(1 - \frac{w_{n[k+1]}(x,h)}{1 - F_{1xh}(T_{(k)})}\right)^{\delta_{[k+1]}}, \end{aligned}$$

where the latter equation follows from the induction hypothesis. If $\delta_{[k+1]} = 0$, we use the induction hypothesis to obtain

$$F_{1xh}^{(\mathrm{KM})}(T_{(k+1)}) = F_{1xh}^{(\mathrm{KM})}(T_{(k)}) = F_{1xh}^{(\mathrm{rc})}(T_{(k)}) = F_{1xh}^{(\mathrm{rc})}(T_{(k+1)}).$$

Otherwise, if $\delta_{[k+1]} = 1$, then

$$\begin{aligned} F_{1xh}^{(\mathrm{KM})}(T_{(k+1)}) &= F_{1xh}^{(\mathrm{rc})}(T_{(k)}) + \{1 - F_{1xh}^{(\mathrm{rc})}(T_{(k)})\} \left(\frac{w_{n[k+1]}(x,h)}{1 - F_{1xh}(T_{(k+1)})}\right) \\ &= F_{1xh}^{(\mathrm{rc})}(T_{(k)}) + \{1 - F_{1xh}^{(\mathrm{KM})}(T_{(k)})\} \left(\frac{w_{n[k+1]}(x,h)}{1 - F_{1xh}(T_{(k+1)})}\right). \end{aligned}$$

From the identity $\{1 - F^{(\mathrm{KM})}_{1xh}(T_{(k)})\}\{1 - G_{1xh}(T_{(k)})\} = 1 - F_{1xh}(T_{(k)})$, we deduce that

$$F^{(\mathrm{KM})}_{1xh}(T_{(k+1)}) \quad = \quad F^{(\mathrm{rc})}_{1xh}(T_{(k)}) + \frac{w_{n[k+1]}(x,h)}{1 - G_{1xh}(T_{(k)})}.$$

The proof follows from the fact that $G_{1xh}(T_{(k)}) = G_{1xh}(T_{(k+1)})$ since $\delta_{[k+1]} = 1$.

The next Lemma shows that, upon setting $y = \infty$, the i.i.d representation for $\mathbb{F}^{(\mathrm{rc})}_{xh}$ found in Theorem 1 is the same as the one obtained in Van Keilegom & Veraverbeke (1997) in the case where the bandwidths g and h required for $F^{(\mathrm{rc})}_{xh}$ are equal.

Lemma 2 *The following identity holds :*

$$\begin{aligned}
&\lim_{y\to\infty}\{\mathcal{J}^{(1)}_{ix}(t,y) + \mathcal{J}^{(2)}_{ix}(t,y)\} \\
&\quad = \{1 - F_{1x}(t)\}\Bigg[\int_0^t \frac{\mathbb{I}(T_i \le v) - H_{1x}(v)}{\{1 - H_{1x}(v)\}^2}\mathrm{d}H^u_{1x}(v) \\
&\qquad\qquad + \frac{\mathbb{I}\{T_i \le t, \delta_i = 1\} - H^u_{1x}(t)}{1 - H_{1x}(t)} \\
&\qquad\qquad - \int_0^t \frac{\mathbb{I}(T_i \le v, \delta_i = 1) - H^u_{1x}(v)}{\{1 - H_{1x}(v)\}^2}\mathrm{d}H_{1x}(v)\Bigg].
\end{aligned}$$

Proof : First, since $h^u_x(v,\infty) = \mathrm{d}H^u_{1x}(v)$, we deduce that

$$v_x(t,\infty,v) = \int_v^t \frac{h^u_x(v,\infty)\mathrm{d}v}{1 - G_x(v)} = F_{1x}(t) - F_{1x}(v). \tag{D.9}$$

In order to show the result, we first deal with the terms that contains the T_i's in the sum $\lim_{y\to\infty}\{\mathcal{J}^{(1)}_{ix}(t,y) + \mathcal{J}^{(2)}_{ix}(t,y)\}$. In view of Equation (D.9), we have

$$\frac{\mathbb{I}\{T_i \le t, \delta_i = 1\}}{1 - G_x(T_i)} \tag{D.10}$$

$$+ \int_0^t \frac{\mathbb{I}(T_i \le v)}{\{1 - H_{1x}(v)\}^2}[F_{1x}(t) - F_{1x}(v)]\mathrm{d}H^c_{1x}(v) \tag{D.11}$$

$$+ \int_0^t \frac{\mathbb{I}(T_i \le v, \delta_i = 0)}{\{1 - H_{1x}(v)\}\{1 - G_x(v)\}}\mathrm{d}H^u_{1x}(v) \tag{D.12}$$

$$- \int_0^t \frac{\mathbb{I}(T_i \le v, \delta_i = 0)}{\{1 - H_{1x}(v)\}^2}[F_{1x}(t) - F_{1x}(v)]\mathrm{d}H_{1x}(v). \tag{D.13}$$

Beginning with (D.11), from the relationship $1 - H_{1x}(v) = \{1 - F_{1x}(v)\}\{1 - G_x(v)\}$, we show that

$$\begin{aligned}
\text{(D.11)} \quad &= -\{1 - F_{1x}(t)\} \int_0^t \frac{\mathbb{I}(T_i \le v)}{\{1 - H_{1x}(v)\}^2} \mathrm{d}H_{1x}^c(v) \\
&\qquad + \int_0^t \frac{\mathbb{I}(T_i \le v)}{\{1 - H_{1x}(v)\}^2} \{1 - F_{1x}(v)\} \mathrm{d}H_{1x}^c(v) \\
&= -\{1 - F_{1x}(t)\} \int_0^t \frac{\mathbb{I}(T_i \le v)}{\{1 - H_{1x}(v)\}^2} \mathrm{d}H_{1x}^c(v) \\
&\qquad + \int_0^t \frac{\mathbb{I}(T_i \le v)}{\{1 - H_{1x}(v)\}\{1 - G_x(v)\}} \mathrm{d}H_{1x}^c(v).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\text{(D.11)} \quad &= -\{1 - F_{1x}(t)\} \left[\int_0^t \frac{\mathbb{I}(T_i \le v)}{\{1 - H_{1x}(v)\}^2} \mathrm{d}H_{1x}(v) - \int_0^t \frac{\mathbb{I}(T_i \le v)}{\{1 - H_{1x}(v)\}^2} \mathrm{d}H_{1x}^u(v) \right] \\
&\quad + \int_0^t \frac{\mathbb{I}(T_i \le v)}{\{1 - H_{1x}(v)\}\{1 - G_x(v)\}} \mathrm{d}H_{1x}(v) \\
&\qquad - \int_0^t \frac{\mathbb{I}(T_i \le v)}{\{1 - H_{1x}(v)\}\{1 - G_x(v)\}} \mathrm{d}H_{1x}^u(v).
\end{aligned}$$

because

$$H_{1x}^c(v) = H_{1x}(v) - H_{1x}^u(v).$$

Now, from the fact that $\mathbb{I}(T_i \le v, \delta_i = 0) = \mathbb{I}(T_i \le v) - \mathbb{I}(T_i \le v, \delta_i = 1)$, we obtain

$$\begin{aligned}
\text{(D.12)} \quad &= \int_0^t \frac{\mathbb{I}(T_i \le v)}{\{1 - H_{1x}(v)\}\{1 - G_x(v)\}} \mathrm{d}H_{1x}^u(v) \\
&\qquad - \int_0^t \frac{\mathbb{I}(T_i \le v, \delta_i = 1)}{\{1 - H_{1x}(v)\}\{1 - G_x(v)\}} \mathrm{d}H_{1x}^u(v).
\end{aligned}$$

Then, from the identity $H^u_{1x}(v) = \int_0^v \{1 - G_x(t-)\} \mathrm{d}F_{1x}(t)$, integration by parts leads to

$$
\begin{aligned}
\text{(D.12)} &= \int_0^t \frac{\mathbb{I}(T_i \leq v)}{\{1 - H_{1x}(v)\}\{1 - G_x(v)\}} \mathrm{d}H^u_{1x}(v) + \frac{\mathbb{I}\{T_i \leq t, \delta_i = 1\}}{1 - G_x(t)} \\
&\quad - \frac{\mathbb{I}\{T_i \leq t, \delta_i = 1\}}{1 - G_x(T_i)} - \int_0^t \frac{\mathbb{I}\{T_i \leq v, \delta_i = 1\}}{\{1 - H_{1x}(v)\}^2} \mathrm{d}H_{1x}(v) \\
&= \int_0^t \frac{\mathbb{I}(T_i \leq v)}{\{1 - H_{1x}(v)\}\{1 - G_x(v)\}} \mathrm{d}H^u_{1x}(v) \\
&\quad + \{1 - F_{1x}(t)\} \frac{\mathbb{I}\{T_i \leq t, \delta_i = 1\}}{1 - H_{1x}(t)} - \frac{\mathbb{I}\{T_i \leq t, \delta_i = 1\}}{1 - G_x(T_i)} \\
&\quad - \int_0^t \frac{\mathbb{I}\{T_i \leq v, \delta_i = 1\}}{\{1 - H_{1x}(v)\}\{1 - G_x(v)\}} \mathrm{d}H_{1x}(v).
\end{aligned}
$$

Proceeding similarly, we obtain that

$$
\begin{aligned}
\text{(D.13)} &= \{1 - F_{1x}(t)\} \left[\int_0^t \frac{\mathbb{I}(T_i \leq v)}{\{1 - H_{1x}(v)\}^2} \mathrm{d}H_{1x}(v) - \int_0^t \frac{\mathbb{I}(T_i \leq v, \delta_i = 1)}{\{1 - H_{1x}(v)\}^2} \mathrm{d}H_{1x}(v) \right] \\
&\quad - \int_0^t \frac{\mathbb{I}(T_i \leq v)}{\{1 - H_{1x}(v)\}\{1 - G_x(v)\}} \mathrm{d}H_{1x}(v) \\
&\quad + \int_0^t \frac{\mathbb{I}(T_i \leq v, \delta_i = 1)}{\{1 - H_{1x}(v)\}\{1 - G_x(v)\}} \mathrm{d}H_{1x}(v).
\end{aligned}
$$

Then, adding these terms leads to

$$
\begin{aligned}
\text{(D.10)} + \ldots + \text{(D.13)} &= \{1 - F_{1x}(t)\} \left[\int_0^t \frac{\mathbb{I}(T_i \leq v)}{\{1 - H_{1x}(v)\}^2} \mathrm{d}H^u_{1x}(v) \right. \\
&\quad \left. + \frac{\mathbb{I}\{T_i \leq t, \delta_i = 1\}}{1 - H_{1x}(t)} - \int_0^t \frac{\mathbb{I}(T_i \leq v, \delta_i = 1)}{\{1 - H_{1x}(v)\}^2} \mathrm{d}H_{1x}(v) \right],
\end{aligned}
$$

which concludes the proof.

Annexe E

Test de concordance

E.1 Définition des conditions requises

Dans cette section, nous avons recours aux notations suivantes. Si on dénote une suite de variables aléatoires par $(Z_n)_{n\in\mathbb{N}}$, et une suite de nombres réels par $(a_n)_{n\in\mathbb{N}}$, nous écrivons $Z_n = O_e(a_n)$ lorsqu'il existe des constantes C_1, C_2 et $\eta > 0$ telles que $\mathrm{P}(Z_n/a_n \geq C_1) \leq C_2 \exp(-n^\eta/C_1)$. De plus, définissons $J_{nx} = [\,\min_{i\in I_{nx}} x_i, \max_{i\in I_{nx}} x_i\,]$, où $I_{nx} = \{j \in \{1,\ldots,n\} : w_{nj}(x,h_n) > 0\}$.

Les hypothèses sur les fonctions de poids énumérées ci-après sont écrites en adoptant le cadre d'un design fixe. L'écriture de ces conditions en design aléatoire consiste généralement à en remplacer o par $o_{\text{a.s.}}$ et O par $O_{\text{a.s.}}$, à l'exception de W_4, W_6, W_2' et W_2'', hypothèses pour lesquelles O doit être remplacé par O_e. Ces conditions un peu plus fortes sont nécessaires pour assurer la validité de versions « design aléatoire » du lemme 3 et du lemme 4 de l'article Omelka *et al.* (2013). Ces lemmes seront utilisés dans la prochaine démonstration.

Les conditions suivantes permettent de garantir la convergence faible des processus $\mathbb{C}_{xh}$ et $\widetilde{\mathbb{C}}_{xh}$

W_1. $\sqrt{na_n} \max_{1 \le i \le n} |w_{ni}(x, h_n)| = o(1)$;

W_2. $\sup_{z \in J_{nx}} \sqrt{na_n} \left| \sum_{i=1}^n w_{ni}(z, a_n)(x_i - x) - a_n^2 K_2 \right| = o(1)$ pour un certain $K_2 = K_2(x) \in (0, \infty)$;

W_3. $\sup_{z \in J_{nx}} \sqrt{na_n} \left| \sum_{i=1}^n w_{ni}(z, a_n)(x_i - z)^2/2 - h_n^2 K_3 \right| = o(1)$ pour un certain $K_3 = K_3(x) \in (0, \infty)$;

W_4. $\sup_{z \in J_{nx}} na_n \sum_{i=1}^n \{w_{ni}(z, a_n)\}^2 - K_4 = O(1)$ pour un certain $K_4 = K_4(x) \in (0, \infty)$;

W_5. $\max_{i \in I_{nx}} x_i - \min_{i \in I_{nx}} x_i = o(1)$;

W_6. Il existe une constante C telle que :

$$\mathrm{P}\left(\sup_{z \in J_x^{(n)}} \max_{1 \le i \le n} |w_{ni}(z, a_n) \mathbb{I}(|X_i - x| > Ca_n)| > 0 \right) = o_{\mathrm{P}}(1);$$

W_7. $\sup_{z \in J_{nx}} \sum_{i=1}^n |w'_{ni}(z, a_n)| = O_{\mathrm{P}}(a_n^{-1})$

Les conditions suivantes permettent de montrer la validité des stratégies de ré-échantillonnage basées sur des multiplicateurs.

W_8. $(nh_n)^2 \sum_{i=1}^n \{w_{ni}(x, h_n)\}^4 = O(n^{-\delta})$ pour un certain $\delta > 0$.

W_1'. $\left| \sum_{i=1}^n w'_{ni}(x, g_n) \right| = o(1)$;

W_2'. $n^\delta \sum_{i=1}^n \{w'_{ni}(x, g_n)\}^2 = O(1)$ pour un certain $\delta > 0$;

W_3'. $\left| \sum_{i=1}^n w'_{ni}(x, g_n)(x_i - x) - 1 \right| = o(1)$;

W_1''. $\sup_{z \in J_{nx}} \left| \sum_{i=1}^n w''_{ni}(z, g_n) \right| = o(1)$;

W_2''. $n^\delta \sup_{z\in J_{nx}} \sum_{i=1}^n \{w_{ni}''(x,g_n)\}^2 = O(1)$ pour un certain $\delta > 0$;

W_3''. $\sup_{z\in J_{nx}} \left| \sum_{i=1}^n w_{ni}''(z,g_n)(x_i - x) \right| = o(1)$;

W_4''. $\sup_{z\in J_{nx}} \left| \sum_{i=1}^n \{w_{ni}''(x,g_n)(x_i - x)\}^2 - 1 \right| = o(1)$;

W_5''. $g_n^2 \sup_{z\in J_{nx}} \sum_{i=1}^n |w_{ni}''(z,g_n)| = O(1)$;

W_6''. Il existe des constantes $C_1, C_2 < \infty$ et $\alpha > 0$ telles que, pour tout $z_1, z_2 \in J_{nx}$,

$$\max_{i\in I_{nx}} |w_{ni}''(z_1,g_n) - w_{ni}''(z_2,g_n)| \leq C_1\, g_n^{-C_2} |z_1 - z_2|^\alpha.$$

E.1.1 Hypothèses distributionnelles

$\mathcal{A}_1$. Les dérivées partielles $C_x^{[1]}(u,v) = \partial\, C_x(u,v)/\partial u$ et $C_x^{[2]}(u,v) = \partial\, C_x(u,v)/\partial v$ existent et sont continues sur $(0,1)\times[0,1]$ et $[0,1]\times(0,1)$.

$\mathcal{A}_2$. Les fonctions $(z,y_1,y_2) \mapsto H_z(y_1,y_2)$, $\dot{H}_z = \partial H_z/\partial z$ et $\ddot{H}_z = \partial^2 H_z/\partial z^2$ existent et sont uniformément continues en $(z,y_1,y_2) \in J_x \times \mathbb{R}^2$, où J_x désigne un voisinage de x.

$\mathcal{A}_2^\star$. Les fonctions $(z,u_1,u_2) \mapsto C_z(u_1,u_2)$, $\dot{C}_z = \partial C_z/\partial z$ et $\ddot{C}_z = \partial^2 C_z/\partial z^2$ existent et sont uniformément continues sur $(z,u_1,u_2) \in J_x \times [0,1]^2$, où J_x désigne un voisinage de x.

$\mathcal{A}_3$. Pour $j = 1,2$, les fonctions $(z,u) \mapsto F_{jz}\{F_{jz}^{-1}(u)\}$, $\dot{F}_{jz}\{F_{jz}^{-1}(u)\}$ et $\ddot{F}_{jz}\{F_{jz}^{-1}(u)\}$ existent et sont continues sur $(z,u) \in J_x \times [0,1]$, où J_x désigne un voisinage de x.

$\mathcal{A}_4$. Pour $j = 1,2$ et $\epsilon > 0$, il existe des constantes $C > 0$ et $\eta > 0$ et un voisinage $U(x)$

tels que

$$\sup_{z_1,z_2\in U(x)} \sup_{u,v\in[\epsilon,1-\epsilon]} |\ddot{F}_{jz_1}\{F_{jz_2}^{-1}(u)\} - \ddot{F}_{jz_1}\{F_{jz_2}^{-1}(v)\}| \leq C|u-v|^\eta.$$

E.2 Preuve de la proposition 1

Dans ce qui suit, nous présentons une preuve de la proposition 1 en adoptant le cadre d'un design fixe. La preuve en design aléatoire pourrait être obtenue à la suite d'arguments similaires en conditionnant d'abord par rapport à la co-variable. De plus, il est à noter que les espérances se présentant sous la forme $\mathrm{E}^\star\{f(\xi_i, Y_{1i}, Y_{2i}, x_i)\}$ suivant la mesure de probabilité bootstrap $\mathrm{P}^\star$ sont calculées en rapport aux multiplicateurs ξ_i seulement, ou, autrement dit, conditionnellement aux données.

E.2.1 Espérance

D'abord, posons

$$G_{xh}^b(u,v) = \sum_{i=1}^n \mathbb{I}\left(\widetilde{U}_{1i}^b \leq u, \widetilde{U}_{2i}^b \leq v\right) w_{ni}(x,h).$$

En utilisant cette notation, nous ré-écrivons :

$$\begin{aligned} \widetilde{\alpha}_{xh}^\star(u,v) &= \sqrt{nh}\sum_{i=1}^n (\xi_{ix} - \xi_{\cdot x})\mathbb{I}\left(\widetilde{U}_{1i}^b \leq u, \widetilde{U}_{2i}^b \leq v\right) \\ &= \sqrt{nh}\sum_{i=1}^n \xi_{ix}\mathbb{I}\left(\widetilde{U}_{1i}^b \leq u, \widetilde{U}_{2i}^b \leq v\right) - \xi_{\cdot x} G_{xg}^b(u,v). \end{aligned}$$

Puisque $\sum_{i=1}^n p_{ix} = 1$, on en déduit que

$$\mathrm{E}^\star\left\{\widetilde{\alpha}_{xh}^\star(u,v)\right\} = \sqrt{nh}\sum_{i=1}^n p_{ix}\mathbb{I}\left(U_{1i}^b \leq u, \widetilde{U}_{2i}^b \leq v\right) - G_{xg}^b(u,v).$$

En vertu de la définition de p_{ix}, nous obtenons alors que

$$\begin{aligned}\sum_{i=1}^{n} p_{ix}\mathbb{I}\left(U_{1i}^{b} \leq u, \widetilde{U}_{2i}^{b} \leq v\right) &= \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ni}(x_j, g) w_{nj}(x,h)\mathbb{I}\left(U_{1i}^{b} \leq u, \widetilde{U}_{2i}^{b} \leq v\right) \\ &= \sum_{j=1}^{n}\left\{\sum_{i=1}^{n} w_{ni}(x_j, g)\mathbb{I}\left(U_{1i}^{b} \leq u, \widetilde{U}_{2i}^{b} \leq v\right)\right\} w_{nj}(x,h) \\ &= \sum_{j=1}^{n} G_{x_j g}^{b}(u,v) w_{nj}(x,h).\end{aligned}$$

Ainsi, le résultat escompté peut être obtenu en montrant que :

$$\sqrt{nh}\sum_{i=1}^{n} w_{ni}(x,h)\{G_{x_j g}^{b}(u,v) - G_{xg}^{b}(u,v)\} = KB_{C_x}(u,v) + o_{\text{a.s.}}(1). \tag{E.1}$$

Pour ce faire, il s'avère qu'il suffit d'examiner la différence entre les couples transformés $(\widetilde{U}_{1i}^{b}, \widetilde{U}_{2i}^{b})$ et les *vraies* observations uniformisées $(U_{1i}, U_{2i}) = (F_{1X_i}(Y_{1i}), F_{2X_i}(Y_{2i}))$. En vertu des hypothèses relatives aux paramètres de lissage g_1 et g_2, et puisque les conditions $\mathcal{A}_3$, W_2–W_4 sont satisfaites, nous déduisons du Théorème 2 de l'article Omelka *et al.* (2013) que pour $j = 1, 2$:

$$\max_{i \in I_{nx}} |\widetilde{U}_{1i}^{b} - U_{1i}| = O_{\text{a.s.}}(g_j^2). \tag{E.2}$$

En utilisant un développement de Taylor, on montre que

$$G_{x_i g}^{b}(u,v) - G_{xg}^{b}(u,v) = \dot{G}_{xg}^{b}(u,v)(x_i - x) + \frac{1}{2}\ddot{G}_{z_i g}^{b}(u,v)(x_i - x)^2,$$

où z_i est un réel situé entre x_i et x. En vertu des hypothèses W_2'', W_5'', W_6'', et avec l'aide de l'équation (E.2), il est possible de combiner le lemme 6 au lemme 4 de l'article Omelka *et al.* (2013) pour le choix $d_{i,n} = w_{ni}''(z_i, g_n)$ pour obtenir :

$$\ddot{G}_{z_i g}^{b}(u,v) = \sum_{j=1}^{n} C_{X_j}(u,v) w_{nj}''(z_i, g) + o_{\text{a.s.}}(1)$$

uniformément en $u, v \in [0,1]$. De plus, en ayant recours à un développement de Taylor de C_{X_j} autour du point z_i, il est possible d'écrire, pour un certain z_{ij} entre x_j et z_i, que

$$\begin{aligned}
\mathrm{E}\left\{\ddot{G}^b_{z_i g}(\mathbf{y})\right\} &= C_{z_i}(u,v)\sum_{j=1}^n w''_{nj}(z_i,g) + \dot{C}_{z_i}(u,v)\sum_{j=1}^n w''_{nj}(z_i,g)(x_j - z_i) \\
&\quad + \sum_{j=1}^n w''_{nj}(z_i,g)\ddot{F}_{z_{ij}}(\mathbf{y})(x_j - z_i)^2/2 \\
&= \ddot{C}_{z_i}(u,v) + o_{\mathrm{a.s.}}(1).
\end{aligned}$$

La dernière égalité est obtenue par l'entremise des conditions W''_1, W''_2 et W''_4. Ainsi, $\ddot{G}^b_{z_i g}(u,v) = \ddot{C}_{z_i}(u,v) + o_{\mathrm{a.s.}}(1)$. Aussi, de façon similaire, les hypothèses W'_1–W'_3 nous permettent de conclure que $\dot{G}^b_{xg}(u,v) = \dot{C}_x(u,v) + o_{\mathrm{a.s.}}(1)$. En rassemblant les résultats ici discutés, et en ayant recours aux conditions W_2 et W_3, on en déduit que l'équation (E.1) est vraie, ce qui montre le résultat attendu.

E.2.2 Convergence faible

Soit $\mathcal{F}$ la classe des fonctions indicatrices $\mathbb{I}\{\cdot \leq v, \cdot \leq$ sur $\mathbb{R}^2$, et dénotons par $\|\cdot\|_{\mathcal{F}}$ la norme suprémum définie sur l'ensemble $\mathcal{F}$. Pour $Z_{hi}(u,v) = \sqrt{nh_n}\,\mathbb{I}(U^b_{1i} \leq u, \widetilde{U}^b_{2i} \leq v)\{\xi_{ix} - \mathrm{E}^\star(\xi_{ix})\}$ et $T_{hi}(u,v) = \sqrt{nh_n}\,\{\xi_{ix} - \mathrm{E}^\star(\xi_{ix})\}G^b_{xg}(u,v)$, on introduit la fonction aléatoire

$$Z_{xh} = \sum_{i=1}^n Z_{hi} \quad \text{and} \quad T_{xh} = \sum_{i=1}^n T_{hi}.$$

On décompose alors simplement $\widetilde{\alpha}^\star_{xh} = Z_{xh} + T_{xh}$. Montrons d'abord que Z_{xh} est asymptotiquement tendu. En vertu du Théorème 2.11.1 de van der Vaart & Wellner (1996), il est possible de conclure que Z_{xh} est tendu si les conditions suivantes sont satisfaites [P]–presque sûrement :

($\mathcal{R}_1$) Pour tout $\eta > 0$:

$$\sum_{i=1}^n \mathrm{E}^\star\left(\|Z_{hi}\|^2_{\mathcal{F}}\right)\mathbb{I}\left\{\|Z_{hi}\|_{\mathcal{F}} > \eta\right\} \to 0;$$

($\mathcal{R}_2$) Pour toute suite décroissante $\delta_n > 0$ convergeant à 0 :

$$\sup_{|u-u'|+|v-v'|<\delta_n} \sum_{i=1}^{n} \mathrm{E}^\star \left\{ \left(Z_{hi}(u,v) - Z_{hi}(u',v')\right)^2 \right\} \to 0;$$

($\mathcal{R}_3$) Pour toute suite décroissante $\delta_n > 0$ convergeant à 0 :

$$\int_0^{\delta_n} \{\log N(\epsilon, \mathcal{F}, d_n)\}^{1/2} \, \mathrm{d}\epsilon \xrightarrow{\mathrm{P}^\star} 0,$$

où $N(\epsilon, \mathcal{F}, d_n)$ est le « nombre de recouvrement » (covering number) de $\mathcal{F}$ calculé suivant la semi-métrique aléatoire

$$d_n^2(u,v,u',v') = \sum_{i=1}^{n} \{Z_{hi}(u,v) - Z_{hi}(u',v')\}^2 .$$

D'abord, la condition $\mathcal{R}_1$ est trivialement satisfaite, en raison du fait que

$$\max_{1\le i\le n} \xi_{ix} = o_{\mathrm{P}^\star} \left\{ (nh_n)^{-\frac{1}{2}} \right\}.$$

Ensuite, en supposant sans perte de généralité que $u' \le u$ et $v' \le v$, on obtient :

$$\begin{aligned}
&\textstyle\sum_{i=1}^{n} \mathrm{E}^\star \left\{ \left(Z_{hi}(\mathbf{y}) - Z_{hi}(\mathbf{y}')\right)^2 \right\} \\
&\quad \le \; nh_n \sum_{i=1}^{n} v_{ii}(x) \left\{ \mathbb{I}(u' < \widetilde{U}_{1i}^b \le u) + \mathbb{I}(v' < U_{2i}^b \le v) \right\} \\
&\quad = \; nh_n \sum_{i,k=1}^{n} \left\{ \mathbb{I}(u' < \widetilde{U}_{1i}^b \le u) + \mathbb{I}(v' < U_{2i}^b \le v) \right\} w_{ni}(x_k, g_n) \{w_{nk}(x,h_n)\}^2 \\
&\quad = \; nh_n \sum_{k=1}^{n} \left| G_{1x_k g}^b(u) - G_{1x_k g}^b(u') \right| \{w_{nk}(x,h_n)\}^2 \\
&\qquad + nh \sum_{k=1}^{n} \left| G_{2x_k g}^b(v) - G_{2x_k g}^b(v') \right| \{w_{nk}(x,h_n)\}^2. \qquad (\text{E.3})
\end{aligned}$$

De la même façon qu'à la section E.2.1, en combinant successivement le lemme 6 au lemme 3 de Omelka *et al.* (2013), on peut montrer que $G_{jzg}^b(u) = u + o_{\mathrm{a.s.}}(1)$. En insérant ce résultat au sein de l'équation (E.3), on obtient :

$$\sum_{i=1}^{n} \mathrm{E}^{\star} \left\{ (Z_{hi}(\mathbf{y}) - Z_{hi}(\mathbf{y}')) \right\}^2 \leq \left(nh_n \sum_{k=1}^{n} \{ w_{nk}(x, h_n) \}^2 \right) \times \left\{ |u - u'| + |v - v'| + o_{\mathrm{a.s.}}(1) \right\}.$$

À la lumière de l'hypothèse W_4, nous concluons donc que la condition $\mathcal{R}_2$ est satisfaite. Enfin, la condition $\mathcal{R}_3$ peut être vérifiée à la manière des développements présentés dans la preuve du théorème 1 de Veraverbeke *et al.* (2011). Ainsi, Z_{xh} est asymptotiquement tendu. De plus, puisque $G_{xg} \leq b$, on peut aisément conclure que T_{xh} est aussi asymptotiquement tendu en utilisant à nouveau le fait que $\max_{1 \leq i \leq n} \xi_{ix} = o_{\mathrm{P}^{\star}}((nh_n)^{-1/2})$, combiné à l'hypothèse W_4. Ainsi, $\widetilde{\alpha}^{\star}_{xh}$ satisfait au critère de tension asymptotique.

Afin de conclure que $\widetilde{\alpha}^{\star}_{xh}$ est asymptotiquement gaussien, il ne reste plus qu'à montrer que la variance de ce processus est, asymptotiquement, la même que celle de $\widetilde{\alpha}^{\star}_{x}$. Pour ce faire, remarquons que

$$\begin{aligned}
\mathrm{Var} \left\{ \widetilde{\alpha}^{\star}_{xh}(u, v) \right\} &= nh \sum_{i=1}^{n} \left\{ \mathbb{I}(\widetilde{U}^b_{1i} \leq u, \widetilde{U}^b_{2i} \leq v) - G^b_{xg}(u, v) \right\}^2 v_{ix} \\
&= nh \sum_{k=1}^{n} \sum_{i=1}^{n} \left\{ \mathbb{I}(\widetilde{U}^b_{1i} \leq u, \widetilde{U}^b_{2i} \leq v) - G^b_{xg}(u, v) \right\}^2 w_{ni}(x_k, g) w_{nk}(x, h)^2 \\
&= nh \sum_{k=1}^{n} \left\{ G^b_{x_k g}(u, v) - 2 G^b_{x_k g}(u, v) G^b_{xg}(u, v) + G^b_{xg}(u, v)^2 \right\} w_{nk}(x, h)^2.
\end{aligned}$$

Puisque $G^b_{zg} = C_z + o_{\mathrm{a.s.}}(1) = C_x + o_{\mathrm{a.s.}}(1) + o(1)$ uniformément pour $z \in J_{nx}$, on en déduit que la dernière équation se réduit à l'expression suivante :

$$\left\{ nh \sum_{k=1}^{n} w_{nk}(x, h)^2 \right\} \left[C_x(u, v) \{ 1 - C_x(u, v) \} + o_{\mathrm{a.s.}}(1) + o(1) \right].$$

Ainsi, le résultat attendu est obtenu en utilisant la condition W_4.

Bibliographie

Abegaz *et al.* (2012). ABEGAZ, F., GIJBELS, I. & VERAVERBEKE, N. (2012). Semiparametric estimation of conditional copulas. *J. Multivariate Anal.* **110**, 43–73.

Acar *et al.* (2011). ACAR, E. F., CRAIU, R. V. & YAO, F. (2011b). Dependence calibration in conditional copulas : a nonparametric approach. *Biometrics* **67**, 445–453.

Aerts *et al.* (1994). AERTS, M., JANSSEN, P. & VERAVERBEKE, N. (1994). Bootstrapping regression quantiles. *J. Nonparametr. Statist.* **4**, 1–20.

Akritas (1994). AKRITAS, M. G. (1994). Nearest neighbor estimation of a bivariate distribution under random censoring. *The Annals of Statistics* , 1299–1327.

Akritas & Keilegom (2003). AKRITAS, M. G. & KEILEGOM, I. V. (2003). Estimation of bivariate and marginal distributions with censored data. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)* **65**, 457–471.

Andersen *et al.* (1993). ANDERSEN, P. K., BORGAN, GILL, R. D. & KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer Series in Statistics. Springer Verlag, 2nd ed.

Beare & Seo (2015). BEARE, B. K. & SEO, J. (2015). Vine copula specifications for stationary multivariate Markov chains. *J. Time Series Anal.* **36**, 228–246.

Benedetti (1977). BENEDETTI, J. K. (1977). On the nonparametric estimation of regression functions. *Journal of the Royal Statistical Society. Series B (Methodological)* **39**, 248–253.

Beran (1981). BERAN, R. (1981). Nonparametric regression with randomly censored survival data. Tech. rep., California, Berkeley, 19.

Bergsma (2013). BERGSMA, W. (2013). Nonparametric testing of conditional independence by means of the partial copula. Working paper. Department of Statistics, London School of Economics.

Bickel & Wichura (1971). Bickel, P. J. & Wichura, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42**, 1656–1670.

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

Boudjellaba *et al.* (1992). Boudjellaba, H., Dufour, J.-M. & Roy, R. (1992). Testing causality between two vectors in multivariate autoregressive moving average models. *J. Amer. Statist. Assoc.* **87**, 1082–1090.

Boudjellaba *et al.* (1994). Boudjellaba, H., Dufour, J.-M. & Roy, R. (1994). Simplified conditions for noncausality between vectors in multivariate ARMA models. *J. Econometrics* **63**, 271–287.

Bouezmarni *et al.* (2016). Bouezmarni, T., Camirand-Lemyre, F. & Quessy, J.-F. (2016). On the asymptotic behavior of two estimators of the conditional copula based on time series. *Technical report, Université de Sherbrooke* .

Bouezmarni *et al.* (2012). Bouezmarni, T., Rombouts, J. V. K. & Taamouti, A. (2012). Nonparametric copula-based test for conditional independence with applications to Granger causality. *J. Bus. Econom. Statist.* **30**, 275–287.

Bouezmarni & Taamouti (2014). Bouezmarni, T. & Taamouti, A. (2014). Nonparametric tests for conditional independence using conditional distributions. *J. Nonparametr. Stat.* **26**, 697–719.

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

Calonico *et al.* (2014). Calonico, S., Cattaneo, M. D. & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* **82**, 2295–2326.

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

Dabrowska (1988). DABROWSKA, D. M. (1988). Kaplan-meier estimate on the plane. *The Annals of Statistics* , 1475–1489.

de Matos & Fernandes (2007). DE MATOS, J. A. & FERNANDES, M. (2007). Testing the Markov property with high frequency data. *Journal of Econometrics* **141**, 44–64.

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

Dufour & Taamouti (2010). DUFOUR, J.-M. & TAAMOUTI, A. (2010). Short and long run causality measures : Theory and inference. *Journal of Econometrics* **154**, 42–58.

Florens & Fougere (1996). FLORENS, J.-P. & FOUGERE, D. (1996). Noncausality in continuous time. *Econometrica* **64**, 1195–1212.

Gasser & Müller (1979). GASSER, T. & MÜLLER, H.-G. (1979). Kernel estimation of regression functions. In *Smoothing techniques for curve estimation (Proc. Workshop, Heidelberg, 1979)*, vol. 757 of *Lecture Notes in Math.* Springer, Berlin, pp. 23–68.

Geweke (1982). GEWEKE, J. (1982). Measurement of linear dependence and feedback between multiple time series. *J. Amer. Statist. Assoc.* **77**, 304–324. With discussion and with a reply by the author.

Geweke (1984). GEWEKE, J. F. (1984). Measures of conditional linear dependence and feedback between time series. *J. Amer. Statist. Assoc.* **79**, 907–915.

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

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

Granger (1969). GRANGER, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica : Journal of the Econometric Society* , 424–438.

Gribkova & Lopez (2015). GRIBKOVA, S. & LOPEZ, O. (2015). Non-parametric copula estimation under bivariate censoring. *Scandinavian Journal of Statistics* , 925.

Hall (1992). HALL, P. (1992). Efficient bootstrap simulation. *Exploring the limits of bootstrap. John Wiley and Sons, Inc., New York, New York* , 127–143.

Härdle & Bowman (1988). HÄRDLE, W. & BOWMAN, A. W. (1988). Bootstrapping in nonparametric regression : local adaptive smoothing and confidence bands. *J. Amer. Statist. Assoc.* **83**, 102–110.

Hardle & Marron (1991). HARDLE, W. & MARRON, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* **19**, 778–796.

Hoeffding (1994). HOEFFDING, W. (1994). *The collected works of Wassily Hoeffding.* Springer Series in Statistics : Perspectives in Statistics. Springer-Verlag, New York. Edited and with a preface by N. I. Fisher and P. K. Sen.

Huang (2010). HUANG, T.-M. (2010). Testing conditional independence using maximal nonlinear conditional correlation. *Ann. Statist.* **38**, 2047–2091.

Hušková & Meintanis (2009). HUŠKOVÁ, M. & MEINTANIS, S. G. (2009). Goodness-of-fit tests for parametric regression models based on empirical characteristic functions. *Kybernetika* **45**, 960–971.

Jondeau & Rockinger (2006). JONDEAU, E. & ROCKINGER, M. (2006). The copula-garch model of conditional dependencies : An international stock market application. *Journal of international money and finance* **25**, 827–853.

Kosorok (2008). KOSOROK, M. (2008). *Introduction to empirical processes and semiparametric inference.* New York : Springer.

Lehmann (1966). LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37**, 1137–1153.

Lemyre & Quessy (2016). LEMYRE, F. C. & QUESSY, J.-F. (2016). Multiplier bootstrap methods for conditional distributions. *Statistics and Computing* , 1–17.

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

Linton & Gozalo (2014). LINTON, O. & GOZALO, P. (2014). Testing conditional independence restrictions. *Econometric Rev.* **33**, 523–552.

McMurry & Politis (2008). MCMURRY, T. L. & POLITIS, D. N. (2008). Bootstrap confidence intervals in nonparametric regression with built-in bias correction. *Statistics & Probability Letters* **78**, 2463 – 2469.

Meitz & Saikkonen (2002). 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.

Mesfioui & Quessy (2008). MESFIOUI, M. & QUESSY, J.-F. (2008). Dependence structure of conditional Archimedean copulas. *J. Multivariate Anal.* **99**, 372–385.

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

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

Patton (2006). PATTON, A. J. (2006). Modelling asymmetric exchange rate dependence. *Internat. Econom. Rev.* **47**, 527–556.

Polasek (1994). POLASEK, W. (1994). Temporal causality measures based on AIC. In *Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling : an Informational Approach, Vol. 1 (Knoxville, TN, 1992).* Kluwer Acad. Publ., Dordrecht.

Polasek (2002). Polasek, W. (2002). Bayesian causality meaures for multiple ARCH models using marginal likelihoods. *Technical report, Institute of Statistics and Econometrics, Univeristy of Basel., Switzerland.* .

Priestley & Chao (1972). Priestley, M. B. & Chao, M. T. (1972). Non-parametric function fitting. *J. Roy. Statist. Soc. Ser. B* **34**, 385–392.

Robins & Rotnitzky (1992). Robins, J. M. & Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology*. Springer, pp. 297–331.

Satten & Datta (2001). Satten, G. A. & Datta, S. (2001). The kaplan–meier estimator as an inverse-probability-of-censoring weighted average. *The American Statistician* **55**, 207–210.

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

Song (2009). Song, K. (2009). Testing conditional independence via Rosenblatt transforms. *Ann. Statist.* **37**, 4011–4045.

Stute (1994). Stute, W. (1994). The bias of kaplan-meier integrals. *Scandinavian Journal of Statistics* **21**, 475–484.

Su & Spindler (2013). Su, L. & Spindler, M. (2013). Nonparametric testing for asymmetric information. *J. Bus. Econom. Statist.* **31**, 208–225.

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

Su & White (2012). Su, L. & White, H. (2012). Conditional independence specification testing for dependent processes with local polynomial quantile regression. *Advances in Econometrics* **29**, 355–434.

Taamouti *et al.* (2014). TAAMOUTI, A., BOUEZMARNI, T. & EL GHOUCH, A. (2014). Nonparametric estimation and inference for conditional density based Granger causality measures. *J. Econometrics* **180**, 251–264.

van der Vaart & Wellner (1996). 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.

van der Vaart & Wellner (2007). VAN DER VAART, A. W. & WELLNER, J. A. (2007). Empirical processes indexed by estimated functions. *Lecture Notes-Monograph Series*, 234–252.

Van Keilegom & Veraverbeke (1996). VAN KEILEGOM, I. & VERAVERBEKE, N. (1996). Uniform strong convergence results for the conditional kaplan-meier estimator and its quantiles. *Communications in Statistics–Theory and Methods* **25**, 2251–2265.

Van Keilegom & Veraverbeke (1997). VAN KEILEGOM, I. & VERAVERBEKE, N. (1997). Estimation and bootstrap with censored data in fixed design nonparametric regression. *Annals of the Institute of Statistical Mathematics* **49**, 467–491.

Veraverbeke *et al.* (2011). VERAVERBEKE, N., OMELKA, M. & GIJBELS, I. (2011). Estimation of a conditional copula and association measures. *Scand. J. Stat.* **38**, 766–780.

Wang & Hong (2013). WANG, X. & HONG, Y. (2013). Characteristic function based testing for conditional independence : a nonparametric regression approach. Working paper. Dept. of Economics, Cornell University.

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

Wiener (1956). WIENER, N. (1956). The theory of prediction. *Modern mathematics for engineers* **1**, 125–139.

Zhang *et al.* (2016). Zhang, H. J., Dufour, J.-M. & Galbraith, J. W. (2016). Exchange rates and commodity prices : Measuring causality at multiple horizons. *Journal of Empirical Finance* **36**, 100 – 120.