Les Cahiers du GERAD ISSN: 0711–2440

Semi-parametric copula-based models under non-stationarity

B. Nasri, B. Rémillard,
T. Bouezmarni

G–2017–102

November 2017
Revised: September 2018

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2018
– Bibliothèque et Archives Canada, 2018

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2018
– Library and Archives Canada, 2018

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

Semi-parametric copula-based models under non-stationarity

Bouchra Nasri [a]
Bruno Rémillard [a]
Taoufik Bouezmarni [b]

[a] *GERAD, and Department of Decision Sciences, HEC Montréal, Montréal (Québec) Canada, H3T 2A7*

[b] *Department of Statistics, Université de Sherbrooke, Sherbrooke, (Québec) Canada, J1K 2R1*

bouchra.nasri@hec.ca
bruno.remillard@hec.ca
taoufik.bouezmarni@usherbrooke.ca

November 2017
Revised: September 2018
Les Cahiers du GERAD
G–2017–102

Abstract: In this paper, we consider non-stationary response variables and covariates, where the marginal distributions and the associated copula may be time-dependent. We propose estimators for the unknown parameters and we establish the limiting distribution of the estimators of the copula and the conditional copula, together with a parametric bootstrap method for constructing confidence bands around the estimator and for testing the adequacy of the model. We also consider three examples of functionals of the copula-based model under non-stationarity: conditional quantiles, conditional mean, and conditional expected shortfall. The asymptotic distribution of the estimation errors is shown to be Gaussian, and bootstrapping methods are proposed to estimate their asymptotic variances. The finite sample performance of our estimators is investigated through Monte Carlo experiments, and we show three examples of implementation of the proposed methodology.

Keywords: copula, covariates, non-stationarity, conditional distribution

1 Introduction

In many applications, the relationship between a response variable and covariates is often assumed to be linear or a known function of a linear combination of the covariates. Recently, copula-based models have been introduced to model the general dependance between the response and covariates in a stationary setting. For example, Noh et al. (2013) proposed an estimator of the conditonal mean, while Noh et al. (2015), Kraus and Czado (2017), and Rémillard et al. (2017) were interested in conditional quantiles.

However, in practice, data are typically non-stationary. For example, in hydrology, because of climate change, the distribution of an hydrological time series is likely to change over time and/or according to some time-dependent covariates. In the literature, some attempts have been made to deal with non-stationarity for copula-based models. Recently, Jiang et al. (2015) modeled the dependence between bivariate response variables by letting the copula parameters depend on covariates. Ahn and Palmer (2016) did something similar. They proposed to use pseudo-mle for the estimation of parameters but they did not study the convergence of the proposed estimators. Note that in the stationary case, their model is a particular case of the so-called single-index copula (Fermanian and Lopez, 2015) where the copula between the multivariate responses is indexed by a parameter depending on covariates.

However, in this paper, we propose to model the joint distribution of the response and covariates, which is more general than the single-index copula setting. Furthermore, in view of applications, we consider only a univariate response but our results can be extended to multivariate responses, which could be used for considering spatial dependence. This problem will be investigated in a forthcoming paper. More precisely, the main objective of the present paper is to propose a flexible model by allowing three extensions of the copula-based method. First, we assume a parametric model for the distribution of the response variable over time; it makes more sense in the non-stationary setting and it can be useful in some applications such as computing conditional quantiles for extreme cases. Second, we consider a combination of time-dependent and i.i.d. covariates. The distribution of the i.i.d. covariates is estimated nonparametrically by using the empirical distribution functions, while parametric estimators are used to fit the distribution of the time-dependent covariates. To distinguish between these two types of covariates, one can use for example change-point tests (Rémillard, 2012, 2013; Holmes et al., 2013). Third, we model the dependence between the variable of interest and the covariates by fitting a (time-dependent) parametric family of copulas.

The paper is organized as follows: in the next section, we establish the limiting distribution of the estimators of the copula and the conditional copula, together with a parametric bootstrap method for constructing confidence bands around the estimator and for testing the adequacy of the model. The proofs of these results are given in Appendix A. Next, in Section 3, we consider three examples of functionals of the copula-based model under non-stationarity : conditional quantiles, conditional mean, and conditional expected shortfall. The asymptotic distribution of the estimation errors is shown to be Gaussian, and bootstrapping methods are proposed to estimate their asymptotic variances.

The finite sample performance of our estimators is investigated in Section 4 through Monte Carlo experiments, while in Section 5, the usefulness of our method is illustrated with one simulated dataset and two case studies, one from hydro-climatology and the other one from finance. Section 6 provides a conclusion.

2 Estimation of joint and conditional distributions

For $t \in \{1, \ldots, n\}$, $\boldsymbol{X}_t = (X_{t1}, \ldots, X_{td})$ is a covariate vector of dimension $d \geq 1$, and Y_t is the response variable of interest. In what follows, we assume that $(Y_1, \mathbf{X}_1), \ldots, (Y_n, \mathbf{X}_n)$ are independent observations, where $(Y_t, X_{t1}, \ldots, X_{td})$ has continuous margins $\left(\mathcal{G}_t, F_t^{(1)}, \ldots, F_t^{(d)}\right)$ and copula C_t with density c_t, for any $t \in \{1, \ldots, n\}$. Recall that according to Sklar (1959), and since the margins are continuous, for any $t \in \{1, \ldots, n\}$ there exists a unique copula C_t such that for all $y \in \mathbb{R}$ and for all $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$, one has

$$H_t(y, \mathbf{x}) = P(Y_t \leq y, \mathbf{X}_t \leq \mathbf{x}) = C_t\{\mathcal{G}_t(y), \mathbf{F}_t(\mathbf{x})\}, \tag{1}$$

where $\mathbf{F}_t(\mathbf{x}) = \left(F_t^{(1)}(x_1), \ldots, F_t^{(d)}(x_d)\right)^\top$, and C_t is the joint distribution function of $(U_t, \mathbf{V}_t)$, with $U_t = \mathcal{G}_t(Y_t)$ and $\mathbf{V}_t = \mathbf{F}_t(\mathbf{X}_t)$. Further denote by $\mathcal{K}_t(y, \mathbf{x})$ the conditional distribution function of Y_t given $\boldsymbol{X}_t = \mathbf{x}$. Then

$$\mathcal{K}_t(y, \mathbf{x}) = \mathrm{P}(Y_t \leq y | \mathbf{X}_t = \mathbf{x}) = \mathcal{D}_t\{\mathcal{G}_t(y), \mathbf{F}_t(\mathbf{x})\}, \quad y \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^d, \tag{2}$$

where $\mathcal{D}_t(u, \mathbf{v})$ is the conditional distribution function of U_t given $\mathbf{V}_t = \mathbf{v}$ (Rémillard, 2013), defined by

$$\mathcal{D}_t(u, \mathbf{v}) = \frac{\partial v_1 \cdots \partial v_d C_t\left(u, v_1, \ldots, v_d\right)}{\partial v_1 \cdots \partial v_d C_t\left(1, v_1, \ldots, v_d\right)}, \quad u \in (0,1), \mathbf{v} = (v_1, \ldots, v_d) \in (0,1)^d.$$

We also assume that $\mathcal{G}_t = G_{\boldsymbol{\beta}(t,\boldsymbol{b})}$, for a given parametric family of distribution $\{G_{\boldsymbol{\beta}(t,\boldsymbol{b})}; \boldsymbol{b} \in \mathcal{B}\}$ where $\mathcal{B}$ is a parametric space of $\boldsymbol{b}$ and $\boldsymbol{\beta}(t, \boldsymbol{b})$ is a function of time. This modelling hypothesis is needed in order to take care of the non-stationarity and to estimate quantiles associated to extremal values, i.e., when α is close to 0 or 1. We also assume that for any $t \in \{1, \ldots, n\}$, the copula C_t belongs to a parametric family of copula $\{C_{\boldsymbol{\phi}(t,\boldsymbol{\theta})} : \boldsymbol{\theta} \in \mathcal{O}\}$ having a continuous density $c_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_0)}$ on $(0,1)^{1+d}$ for any $\boldsymbol{\phi}$, where $\mathcal{O}$ is a parameter space of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}(t, \boldsymbol{\theta})$ is a function of time.

Suppose for the moment that $\boldsymbol{b}_n$, $\mathbf{F}_{n,t} = \left(F_{t,n}^{(1)}, \ldots, F_{t,n}^{(d)}\right)^\top$, and $\boldsymbol{\theta}_n$ are consistent estimator of $\boldsymbol{b}$, $\mathbf{F}_t$, and $\boldsymbol{\theta}$ respectively; particular cases of consistent estimators are proposed in Subsections 2.1–2.3. Then it follows from (1) that a consistent estimator $H_{n,t}(y, \mathbf{x})$ of $H_t(y, \mathbf{x})$ is given by

$$H_{n,t}(y, \mathbf{x}) = C_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_n)}\{G_{\boldsymbol{\beta}(t,\boldsymbol{b}_n)}(y), \mathbf{F}_{n,t}(\mathbf{x})\}, \quad (y, \mathbf{x}) \in \mathbb{R} \times \mathbb{R}^d. \tag{3}$$

Next, for a given t and $\mathbf{x}$, it follows from (2) that a consistent estimator $\mathcal{K}_{n,t}(y, \mathbf{x})$ of $\mathcal{K}_t(y, \mathbf{x})$ is given by

$$\mathcal{K}_{n,t}(y, \mathbf{x}) = \mathcal{D}_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_n)}\{G_{\boldsymbol{\beta}(t,\boldsymbol{b}_n)}(y), \mathbf{F}_{n,t}(\mathbf{x})\}, \quad y \in \mathbb{R}. \tag{4}$$

We now proposed specific consistent estimators $\boldsymbol{b}_n$, $\boldsymbol{\theta}_n$ and $\mathbf{F}_n$.

2.1 Estimation of b

There is no loss of generality in assuming that for any $t \in \{1, \ldots, n\}$, $\boldsymbol{\beta}(t, \mathbf{b}) = \mathbf{B}_t\mathbf{b}$, where $\mathbf{b} \in \mathcal{B}$ is an unknown parameter, $\mathcal{B}$ is the parameter space for $\mathbf{b}$, and $\mathbf{B}_t$ is a given matrix. For example, $\boldsymbol{\beta}$ can be a polynomial function (linear, quadratic, etc.) or a piecewise parametric function, e.g., a B-spline function (Stone and Koo, 1986; Hastie and Tibshirani, 1990), as long as there is no smoothing parameter, in order to have $n^{1/2}$ consistency. In particular, a linear function $\boldsymbol{\beta}(t, b)$ can be written as $\boldsymbol{\beta}(t, b) = b_0 + b_1 t = \mathbf{B}_t\mathbf{b}$ with $\mathbf{b} = (b_0, b_1)^\top$ and $\mathbf{B}_t = (1, t)$. To estimate the parameter $\mathbf{b}$ of the function $\boldsymbol{\beta}(t, \mathbf{b})$, we suggest to use a maximum likelihood estimator, i.e.,

$$\mathbf{b}_n = \arg\max_{\mathbf{b} \in \mathcal{B}} \sum_{t=1}^{n} \log\left\{g_{\boldsymbol{\beta}(t,\mathbf{b})}(Y_t)\right\}. \tag{5}$$

From now on, we assume that the density $g_{\boldsymbol{\beta}}$ satisfies the smoothness conditions R1–R3 of Serfling (1980, p.144–145), meaning that it is thrice continuously differentiable and locally bounded by integrable functions with respect to $\boldsymbol{\beta}$, with (column) gradient and Hessian matrix denoted by $\dot{g}_t$ and $\ddot{g}_t$ at $\boldsymbol{\beta}(t, \mathbf{b}_0)$. Further assume that the sequence $\mathbf{B}_t^\top \frac{\dot{g}_t(Y_t)}{g_t(Y_t)}$ satisfies Lindeberg's condition (Billingsley, 1995, p.359),

$$\mathbf{S}_n = n^{-1} \sum_{t=1}^{n} \frac{\mathbf{B}_t^\top \dot{g}_t(Y_t) \dot{g}_t(Y_t)^\top \mathbf{B}_t}{g_t^2(Y_t)} \tag{6}$$

converges in probability to $\mathbf{S}$, and $n^{-1}\sum_{t=1}^{n} \frac{\mathbf{B}_t^\top \ddot{g}_t(Y_t)\mathbf{B}_t}{g_t(Y_t)}$ converges in probability to 0, as $n \to \infty$. Finally, set $\mathfrak{b}_n = n^{1/2}(\mathbf{b}_n - \mathbf{b}_0)$, where $\mathbf{b}_0$ is the true value of the parameter.

As a result of these hypotheses, a Taylor expansion as in Serfling (1980, p.145) yields

$$\mathbf{S}_n \mathfrak{b}_n = n^{-1/2} \sum_{t=1}^{n} \frac{\mathbf{B}_t^\top \dot{g}_t(Y_t)}{g_t(Y_t)} + o_P(1). \tag{7}$$

In particular, Lindeberg's CLT can be applied to deduce that $\mathfrak{b}_n$ converges in law to $\mathfrak{b} \sim N\left(0, \mathbf{S}^{-1}\right)$, where $\mathbf{S}^{-1}$ is the Moore-Penrose inverse of $\mathbf{S}$. As a by-product, the last result together with the Delta method (van der Vaart and Wellner, 1996) yields that for any given $t \in \{1, \ldots, n\}$, and uniformly in $y \in \mathbb{R}$,

$$\mathbb{G}_{n,t}(y) = n^{1/2}\{G_{\boldsymbol{\beta}(t,\mathbf{b}_n)}(y) - \mathcal{G}_t(y)\} \rightsquigarrow \mathbb{G}_t(y) = \dot{G}_t(y)^\top \mathbf{B}_t \mathfrak{b}, \tag{8}$$

with $\dot{G}_t(y) = \int_{-\infty}^{y} \dot{g}_t(z)dz$.

2.2 Estimation of $\mathbf{F}_t$

Here, the marginal distributions are estimated according to whether they depend on time or not. First, let $\mathcal{L}$ be the set of indices $j \in \{1, \ldots, d\}$ such that $F_t^{(j)}$ is independent of $t \in \{1, \ldots, n\}$. For any $j \in \mathcal{L}$, $F^{(j)}$ is estimated by $F_n^{(j)}(x_j) = \dfrac{1}{n+1}\displaystyle\sum_{t=1}^{n} \mathbb{I}(X_{tj} \leq x_j)$, which is, up to a factor, the empirical cumulative distribution function of $X_{1j}, \ldots, X_{nj}$. Here, $\mathbb{I}$ stands for the indicator function. It is well-known that $\mathbb{F}_{n,t}^{(j)}(x_j) = n^{1/2}\left\{F_n^{(j)}(x_j) - F^{(j)}(x_j)\right\}$ converges in law to $\mathbb{B}_j\{F^{(j)}(x_j)\}$, where $\mathbb{B}_j$ is a Brownian bridge, i.e., a continuous centered Gaussian process with covariance function $\mathrm{Cov}\{\mathbb{B}_j(s), \mathbb{B}_j(u)\} = \min(s,u) - su,\ s,u \in [0,1]$.

Next, for any $j \notin \mathcal{L}$, the distribution is time-dependent, and we assume that $F_t^{(j)} = F^{(j)}_{\boldsymbol{\gamma}^{(j)}(t)}$, where $\boldsymbol{\gamma}^{(j)}(t) = \boldsymbol{\gamma}^{(j)}(t, \mathbf{q}^{(j)}) = \mathbf{Q}_t^{(j)}\mathbf{q}^{(j)}$, for some given matrix function $\mathbf{Q}_t^{(j)}$. We set $\boldsymbol{\gamma}_n^{(j)}(t) = \mathbf{Q}_t^{(j)}\mathbf{q}_n^{(j)}$, where $\mathbf{q}_n^{(j)}$ is a consistent estimate of $\mathbf{q}^{(j)} \in \mathfrak{Q}$, where $\mathfrak{Q}$ is a space parameter. To simplify notations, we still use $\mathbf{F}_{n,t}$ to denote the estimation of $\mathbf{F}_t$. For the estimation of $\mathbf{q}^{(j)}$ we propose to use the MLE defined by

$$\mathbf{q}_n^{(j)} = \arg\max_{\mathbf{q}\in\mathfrak{Q}} \sum_{t=1}^{n} \log\left\{f^{(j)}_{\boldsymbol{\gamma}^{(j)}(t,\mathbf{q})}(X_{tj})\right\}, \tag{9}$$

where $f_t^{(j)} = f^{(j)}_{\boldsymbol{\gamma}^{(j)}(t,\mathbf{q}^{(j)})}$ is the density of $F_t^{(j)}$.

We assume that the density $f^{(j)}_{\boldsymbol{\gamma}^{(j)}(t,\mathbf{q}^{(j)})}$ with respect to $\boldsymbol{\gamma}^{(j)}$ satisfy the smoothness conditions R1–R3 of Serfling (1980, p.144–145) defined in Section 2.1, with (column) gradient and Hessian matrix denoted by $\dot{f}_t^{(j)}$ and $\ddot{f}_t^{(j)}$ at $\boldsymbol{\gamma}^{(j)}(t, \mathbf{q}_0^{(j)})$. Further assume that the sequence $\left\{\mathbf{Q}_t^{(j)}\right\}^\top \frac{\dot{f}_t^{(j)}(X_{tj})}{f_t^{(j)}(X_{tj})}$ satisfies Lindeberg's condition,

$$\boldsymbol{\mathcal{V}}_{\boldsymbol{n,j}} = n^{-1}\sum_{t=1}^{n} \frac{\left[\mathbf{Q}_t^{(j)}\right]^\top \dot{f}_t^{(j)}(X_{tj})\dot{f}_t^{(j)}(X_{tj})^\top \mathbf{Q}_t^{(j)}}{\left\{f_t^{(j)}(X_{tj})\right\}^2} \tag{10}$$

converges in probability to $\boldsymbol{\mathcal{V}}_{\boldsymbol{j}}$, and $n^{-1}\sum_{t=1}^{n} \frac{\left[\mathbf{Q}_t^{(j)}\right]^\top \ddot{f}_t^{(j)}(X_{tj})\mathbf{Q}_t^{(j)}}{f_t^{(j)}(X_{tj})}$ converges in probability to 0, as $n \to \infty$.

As a result of these hypotheses, one obtains that

$$\boldsymbol{\mathcal{V}}_{\boldsymbol{n,j}}\mathfrak{q}_n^{(j)} = n^{-1/2}\sum_{t=1}^{n} \frac{\left[\mathbf{Q}_t^{(j)}\right]^\top \dot{f}_t^{(j)}(X_{tj})}{f_t^{(j)}(X_{tj})} + o_P(1), \tag{11}$$

where $\mathfrak{q}_n^{(j)} = n^{1/2}\left(\mathbf{q}_n^{(j)} - \mathbf{q}_0^{(j)}\right)$. In particular, Lindeberg's CLT can be applied to deduce that $\mathfrak{q}_n^{(j)}$ converges in law to $\mathfrak{q}^{(j)} \sim N\left(0, \boldsymbol{\mathcal{V}}_{\boldsymbol{j}}^{-\mathbf{1}}\right)$, where $\boldsymbol{\mathcal{V}}_{\boldsymbol{j}}^{-\mathbf{1}}$ is the Moore-Penrose inverse of $\boldsymbol{\mathcal{V}}_{\boldsymbol{j}}$. As a by-product, the last result together with the Delta method yields that for any given $t \in \{1, \ldots, n\}$, and uniformly in $x_j \in \mathbb{R}$,

$$\mathbb{F}_{n,t}^{(j)}(x_j) = n^{1/2}\left\{F_{n,t}^{(j)}(x_j) - F_t^{(j)}(x_j)\right\} \rightsquigarrow \dot{F}_t^{(j)}(x_j)^\top \mathbf{Q}_t^{(j)}\mathfrak{q}^{(j)} \quad \text{for } j \notin \mathcal{L}, \tag{12}$$

with $\dot{F}_t^{(j)}(y) = \int_{-\infty}^{y} \dot{f}_t^{(j)}(z)dz$.

To write the results in a uniform way, define

$$\mathbb{F}_{n,t}(\mathbf{x}) = \left(\mathbb{F}_{n,1}^{(1)}(x_1), \ldots, \mathbb{F}_{n,t}^{(d)}(x_d)\right)^\top, \quad \mathbb{F}_t(\mathbf{x}) = \left(\mathbb{F}_1^{(1)}(x_1), \ldots, \mathbb{F}_t^{(d)}(x_d)\right)^\top,$$

where $\mathbb{F}_t^{(j)} = \mathcal{A}_t^{(j)}(x_j)\mathbb{A}^{(j)}(x_j)$, with

$$\mathcal{A}_t^{(j)}(x_j) = \begin{cases} 1, & \text{if } j \in \mathcal{L}, \\ \dot{F}_t^{(j)}(x_j)^\top \mathbf{Q}_t^{(j)}, & \text{if } j \notin \mathcal{L}, \end{cases} \tag{13}$$

$$\mathbb{A}^{(j)}(x_j) = \begin{cases} \mathbb{B}_j \circ F^{(j)}(x_j), & \text{if } j \in \mathcal{L}, \\ \mathfrak{q}^{(j)}, & \text{if } j \notin \mathcal{L}. \end{cases} \tag{14}$$

Then, under the smoothness assumptions, $\mathbb{F}_{n,t} = \mathcal{A}_t\mathbb{A}_n \rightsquigarrow \mathcal{A}_t\mathbb{A} = \mathbb{F}_t$ for any $t \in \{1, \ldots, n\}$, where $\mathcal{A}_t$ is the diagonal matrix with elements $\mathcal{A}_t^{(j)}$ and

$$\mathbb{A}_n = (\mathbb{A}_n^{(1)}, \ldots, \mathbb{A}_n^{(d)}) \quad \text{with} \quad \mathbb{A}_n^{(j)}(x_j) = \begin{cases} \mathbb{F}_{n,t}^{(j)}(x_j), & \text{if } j \in \mathcal{L}, \\ \mathfrak{q}_n^{(j)}, & \text{if } j \notin \mathcal{L}. \end{cases}$$

It follows from their representations as sums of functions of the independent variables $(Y_t, \mathbf{X}_t)$ that $(\mathfrak{b}_n, \mathbb{A}_n)$ converges jointly in law to $(\mathfrak{b}, \mathbb{A})$

2.3 Estimation of $\boldsymbol{\theta}$

As for the parameters β and $F_t^{(j)}$ with $j \notin \mathcal{L}$, we consider similar assumptions on the parameter of the copula family, namely that $\boldsymbol{\phi}(t, \boldsymbol{\theta}) = \mathbf{h}_t\boldsymbol{\theta}$, for a given matrix function $\mathbf{h}_t$ and $\boldsymbol{\theta}$ belongs to a parameter space denoted $\mathcal{O}$.

Next, recall that $U_t = G_{\boldsymbol{\beta}(t,\boldsymbol{b}_0)}(Y_t)$ and $\mathbf{V}_t = \mathbf{F}_t(\mathbf{X}_t)$, with joint law $C_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_0)}$ on $[0,1]^{d+1}$, are not observable. For the estimation of the copula parameter $\boldsymbol{\theta}$, we will use an hybrid technique, using possibly parametric and nonparametric estimates for the marginal distributions, depending if the components of $\mathbf{F}_t$ are time-dependent or not. Since the distribution of Y_t is time-dependent, we estimate U_t by the parametric estimator $U_{n,t} = G_{\boldsymbol{\beta}(t,\boldsymbol{b}_n)}(Y_t)$, as in the so-called IFM method (Xu, 1996; Joe and Xu, 1996), while the estimate of $\mathbf{V}_t$ is $V_{n,t} = \mathbf{F}_{n,t}(\mathbf{X}_t)$, for any $t \in \{1, \ldots, n\}$. Based on the pseudo-observations $(U_{n,t}, \mathbf{V}_{n,t})$, $t \in \{1, \ldots, n\}$, we then use the pseudo MLE

$$\boldsymbol{\theta}_n = \arg\max_{\boldsymbol{\theta} \in \mathcal{O}} \sum_{t=1}^{n} \log\left\{c_{\boldsymbol{\phi}(t,\boldsymbol{\theta})}(U_{n,t}, \mathbf{V}_{n,t})\right\}. \tag{15}$$

It follows from an easy adaptation of Genest et al. (1995) and Shih and Louis (1995) that $\boldsymbol{\theta}_n$ is a $n^{1/2}$-consistent estimator of $\boldsymbol{\theta}$, if $c_{\boldsymbol{\phi}}$ is smooth enough. The exact convergence result is given in Theorem 1 below. Assume that $c_{\boldsymbol{\phi}}$ is twice continuously differentiable with respect to $u, \mathbf{v}$, and that with respect to $\boldsymbol{\phi}$, $c_{\boldsymbol{\phi}}$ satisfies the smoothness conditions R1–R3 of Serfling (1980), defined in Section 2.1. For simplicity, set $c_t = c_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_0)}$, and denote by $\dot{c}_t$ (resp. $\ddot{c}_t$) the gradient (Hessian matrix) of c_t with respect to ϕ.

$$\mathcal{W}_n = n^{-1/2} \sum_{t=1}^{n} \frac{\mathbf{h}_t^\top \dot{c}_t(U_t, \mathbf{V}_t)}{c_t(U_t, \mathbf{V}_t)}, \tag{16}$$

$$\mathcal{I}_n = n^{-1} \sum_{t=1}^{n} \frac{\mathbf{h}_t^\top \dot{c}_t(U_t, \mathbf{V}_t)\dot{c}_t^\top(U_t, \mathbf{V}_t)\mathbf{h}_t}{c_t^2(U_t, \mathbf{V}_t)}. \tag{17}$$

Further set

$$J_n(\mathbf{x}) = \int_0^1 \left[n^{-1} \sum_{t=1}^{n} \mathbf{h}_t^\top \left\{ \nabla_{\mathbf{v}} \dot{c}_t\{u, \mathbf{F}_t(\mathbf{x})\} - \frac{\dot{c}_t\{u, \mathbf{F}_t(\mathbf{x})\} \nabla_{\mathbf{v}} c_t\{u, \mathbf{F}_t(\mathbf{x})\}^\top}{c_t\{u, \mathbf{F}_t(\mathbf{x})\}} \right\} \mathcal{A}_t(\mathbf{x}) f_t(\mathbf{x}) \right] du,$$

where $f_t(\mathbf{x}) = \prod_{j=1}^d f_t^{(j)}(x_j)$, and

$$\mathbf{K}_n = n^{-1}\sum_{t=1}^n \mathbf{h}_t^\top \left[\int_{(0,1)^{d+1}} \left\{\partial_u \dot{c}_t(u,\mathbf{v}) - \frac{\dot{c}_t(u,\mathbf{v})\partial_u c_t(u,\mathbf{v})}{c_t(u,\mathbf{v})}\right\} \mathcal{G}_t \circ \mathcal{G}_t^{-1}(u)^\top du d\mathbf{v}\right] \boldsymbol{B}_t.$$

Theorem 1 *Suppose that Lindeberg's condition is satisfied for the sequence* $\frac{\mathbf{h}_t^\top \dot{c}_t(U_t,\mathbf{V}_t)}{c_t(U_t,\mathbf{V}_t)}$, $t \in \{1,\ldots,n\}$. *Also assume that uniformly on* $\mathbf{x}$, $(J_n(\mathbf{x}), \mathbf{K}_n, \mathcal{I}_n)$ *converges in probability to* $(J(\mathbf{x}), \mathbf{K}, \mathcal{I})$, *with* $\mathcal{I}$ *invertible, and* $n^{-1}\sum_{t=1}^n \frac{\mathbf{h}_t^\top \ddot{c}_t(U_t,\mathbf{V}_t)\mathbf{h}_t}{c_t^2(U_t,\mathbf{V}_t)}$ *converges in probability to* 0. *Set* $\boldsymbol{\Theta}_n = n^{1/2}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0)$. *Then, as* $n \to \infty$, $(\mathfrak{b}_n, \mathbb{A}_n, \mathcal{W}_n, \boldsymbol{\Theta}_n)$ *converges in law to* $(\mathfrak{b}, \mathbb{A}, \mathcal{W}, \boldsymbol{\Theta})$, *where* $\mathcal{W} \sim N(0, \mathcal{I})$ *and*

$$\boldsymbol{\Theta} = \mathcal{I}^{-1}\left\{\mathcal{W} + \int_{\mathbb{R}^d} J(\mathbf{x})\mathbb{A}(\mathbf{x})d\mathbf{x} + \mathbf{K}\mathfrak{b}\right\}.$$

Moreover, $\mathfrak{b}$, $\mathbb{A}$ *and* $\mathcal{W}$ *are independent Gaussian processes.*

2.4 Convergence of the joint distribution and conditional distribution estimators

For $t \in \{1,\ldots,n\}$, $y \in \mathbb{R}$, and $\mathbf{x} \in \mathbb{R}^d$, set $\mathbb{H}_{n,t}(y,\mathbf{x}) = n^{1/2}\{H_{n,t}(y,\mathbf{x}) - H_t(y,\mathbf{x})\}$, and $\mathbb{K}_{n,t}(y,\mathbf{x}) = n^{1/2}\{\mathcal{K}_{n,t}(y,\mathbf{x}) - \mathcal{K}_t(y,\mathbf{x})\}$. To simplify notations, set $\dot{C}_t(u,\mathbf{v}) = \nabla_{\boldsymbol{\phi}} C_{\boldsymbol{\phi}(t)}(u,\mathbf{v})\big|_{\boldsymbol{\phi}=\boldsymbol{\phi}(t)}$ and $\dot{\mathcal{D}}_t(u,\mathbf{v}) = \nabla_{\boldsymbol{\phi}} \mathcal{D}_{\boldsymbol{\phi}(t)}(u,\mathbf{v})\big|_{\boldsymbol{\phi}=\boldsymbol{\phi}(t)}$. Throughout this section, we assume that these derivatives are continuous. From now on, the convergence in law means convergence in law in the space of continuous functions, equipped with the supremum norm. The next theorem states the weak convergence of the processes $\mathbb{H}_{n,t}$ and $\mathbb{K}_{n,t}$.

Theorem 2 *Suppose that the smoothness conditions hold, together with the assumptions of Theorem 1. Then, as* $n \to \infty$, $\mathbb{H}_{n,t}$ *converges in law to a continuous centered Gaussian process* $\mathbb{H}_t$, *where for any* $y \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^d$,

$$\mathbb{H}_t(y,\mathbf{x}) = \boldsymbol{\Theta}^\top \mathbf{h}_t^\top \dot{C}_t\{\mathcal{G}_t(y), \mathbf{F}_t(\mathbf{x})\} + \mathbb{G}_t(y)\partial_u C_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_0)}\{\mathcal{G}_t(y), \mathbf{F}_t(\mathbf{x})\} + \mathbb{F}_t(\mathbf{x})^\top \nabla_{\mathbf{v}} C_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_0)}\{\mathcal{G}_t(y), \mathbf{F}_t(\mathbf{x})\}. \quad (18)$$

Furthermore, given $\mathbf{X}_t = \mathbf{x}$, *as* $n \to \infty$, $\mathbb{K}_{n,t}$ *converges in law to a continuous centered Gaussian process* $\mathbb{K}_t$, *where for any* $y \in \mathbb{R}$,

$$\mathbb{K}_t(y,\mathbf{x}) = \boldsymbol{\Theta}^\top \mathbf{h}_t^\top \dot{\mathcal{D}}_t\{\mathcal{G}_t(y), \mathbf{F}_t(\mathbf{x})\} + \mathbb{G}_t(y)\partial_u \mathcal{D}_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_0)}\{\mathcal{G}_t(y), \mathbf{F}_t(\mathbf{x})\} + \mathbb{F}_t(\mathbf{x})^\top \nabla_{\mathbf{v}} \mathcal{D}_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_0)}\{\mathcal{G}_t(y), \mathbf{F}_t(\mathbf{x})\}. \quad (19)$$

2.5 Parametric bootstrap

Let $\mathbf{b}_n, \boldsymbol{\theta}_n, \mathbf{F}_{n,t}$ be the estimators of $(\mathbf{b}_0, \boldsymbol{\theta}_0, \mathbf{F}_t)$, $t \in \{1,\ldots,n\}$.

Algorithm 1 *For each* $k \in \{1,\ldots,N\}$, *repeat the following steps:*

- *generate* $(U_t^\star, \mathbf{V}_t^\star) \sim C_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_n)}$, *and compute* $Y_t^\star = G^{-1}_{\boldsymbol{\beta}(t,\mathbf{b}_n)}(U_t^*)$, $\mathbf{X}_t^\star = \mathbf{F}_{n,t}^{-1}(V_t^\star)$, $t \in \{1,\ldots,n\}$;
- *estimate* $\mathbf{b}_0$ *and* $\mathbf{F}_t$ *by* $\boldsymbol{b}_n^\star$ *and* $\mathbf{F}_{n,t}^\star$, *using* $(Y_1^\star, X_1^\star), \ldots, (Y_n^\star, \mathbf{X}_n^\star)$.
- *calculate the pseudo-observations* $U_{n,t}^\star = G_{\boldsymbol{\beta}(t,\mathbf{b}_n^\star)}(Y_t^\star)$, *and* $\mathbf{V}_{n,t}^\star = \mathbf{F}_{n,t}^\star(\mathbf{X}_t^\star)$, $t \in \{1,\ldots,n\}$;
- *estimate* $\boldsymbol{\theta}_0$ *by using* $\left(U_{n,1}^\star, \mathbf{V}_{n,1}^\star\right), \ldots, \left(U_{n,n}^\star, \mathbf{V}_{n,n}^\star\right)$;
- *compute* $H_{n,t}^\star(y,\mathbf{x}) = C_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_n^\star)}\left\{G_{\boldsymbol{\beta}(t,\mathbf{b}_n^\star)}(y), \mathbf{F}_{n,t}^\star(\mathbf{x})\right\}$;
- *compute* $\mathcal{K}_{n,t}^\star(y,\mathbf{x}) = \mathcal{D}_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_n^\star)}\left\{G_{\boldsymbol{\beta}(t,\mathbf{b}_n^\star)}(y), \mathbf{F}_{n,t}^\star(\mathbf{x})\right\}$;
- *set* $\mathbb{H}_{n,t}^{(k)}(y,\mathbf{x}) = n^{1/2}\left\{H_{n,t}^\star(y,\mathbf{x}) - H_{n,t}(y,\mathbf{x})\right\}$, $\mathbb{K}_{n,t}^{(k)}(y,\mathbf{x}) = n^{1/2}\left\{\mathcal{K}_{n,t}^\star(y,\mathbf{x}) - \mathcal{K}_{n,t}(y,\mathbf{x})\right\}$, *and* $W_{n,t}^{(k)} = \mathcal{D}_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_n^\star)}\left(U_{n,t}^\star, \mathbf{V}_{n,t}^\star\right)$, $t \in \{1,\ldots,n\}$.

The next result shows the validity of the proposed bootstrap procedure.

Theorem 3 *Under the smoothness conditions and the assumptions of Theorem 1, as* $n \to \infty$, $\mathbb{H}_{n,t}^{(1)}, \ldots, \mathbb{H}_{n,t}^{(N)}$ *converge to independent copies of* $\mathbb{H}_t$ *and* $\mathbb{K}_{n,t}^{(1)}, \ldots, \mathbb{K}_{n,t}^{(N)}$ *converge to independent copies of* $\mathbb{K}_t$.

As a by-product of this algorithm, we obtain a formal goodness-of-fit test of the model.

2.6 Goodness-of-fit test

A visual assessment of the adequacy of the model is provided by the graph of the pseudo-observations $W_{n,t} = \mathcal{K}_{n,t}(Y_t, \mathbf{X}_t) = \mathcal{D}_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_n)}(U_{n,t}, \mathbf{V}_{n,t})$ over time, since $W_{1,n}, \ldots, W_{n,n}$ are approximately independent and uniformly distributed over $(0,1)$ if the model is correct. This is due to the fact that the values $W_t = \mathcal{K}_t(Y_t, \mathbf{X}_t) = \mathcal{D}_{\boldsymbol{\phi}(t)}(U_t, \mathbf{V}_t)$ are independent and uniformly distributed over $(0,1)$. In addition, it follows from Rosenblatt (1952) that W_t is independent of $\mathbf{X}_t$ for all $t \in \{1, \ldots, n\}$. As a result, a formal test of goodness-of-fit can be based on functionals of the empirical process

$$\mathbb{D}_n(u) = n^{1/2} \left\{ \frac{1}{n} \sum_{t=1}^{n} \mathbb{I}(W_{n,t} \leq u) - u \right\}, \quad u \in [0,1].$$

The derivation of the limiting process $\mathbb{D}$, which is a continuous centered Gaussian process, is quite complicated but it can be done using the tools developed in Ghoudi and Rémillard (1998). The exact form of the limit $\mathbb{D}$ is given in Corollary 1 below. Note that we can use a bootstrapping technique to generate independent asymptotic copies of the limiting process $\mathbb{D}$. In fact, we can use the bootstrapped values $W_{n,t}^{(k)}$, $t \in \{1, \ldots, n\}$, $k \in \{1, \ldots, N\}$, generated in Algorithm 1. This way, one can define, for any $k \in \{1, \ldots, N\}$,

$$\mathbb{D}_n^{(k)}(u) = n^{1/2} \left\{ \frac{1}{n} \sum_{t=1}^{n} \mathbb{I}\left(W_{n,t}^{(k)} \leq u\right) - u \right\}, \quad u \in [0,1].$$

More precisely, if N is large enough, say $N = 1000$, and $\mathcal{T}$ is a continuous functional over the set of continuous functions on $[0,1]$, then an approximate P-value for the statistic $\mathcal{S}_n = \mathcal{T}(\mathbb{D}_n)$ is given by

$$\frac{1}{N} \sum_{k=1}^{N} \mathbb{I}\left(\mathcal{S}_n^{(k)} > \mathcal{S}_n\right),$$

provided we reject the null hypothesis when $\mathcal{S}_n$ is large enough. Here, for any $k \in \{1, \ldots, N\}$, $\mathcal{S}_n^{(k)} = \mathcal{T}\left(\mathbb{D}_n^{(k)}\right)$. For example, if $w_{1,n} < \cdots < w_{n,n}$ are the ordered values of the pseudo-observations $W_{1,n}, \ldots, W_{n,n}$, one could take the Kolmogorov-Smirnov test statistic given by

$$\mathrm{KS_n} = \mathcal{T}(\mathbb{D}_\mathrm{n}) = \sup_{\mathrm{u}\in[0,1]} |\mathbb{D}_\mathrm{n}(\mathrm{u})| = \mathrm{n}^{1/2} \max\left\{ \max_{\mathrm{t}\in\{1,\ldots,\mathrm{n}\}} \left| \mathrm{w_{n,t}} - \frac{\mathrm{t}}{\mathrm{n}} \right|, \max_{\mathrm{t}\in\{1,\ldots,\mathrm{n}\}} \left| \mathrm{w_{n,t}} - \frac{(\mathrm{t}-1)}{\mathrm{n}} \right| \right\}, \tag{20}$$

or the Cramér-von Mises test statistic

$$\mathrm{CVM_n} = \mathcal{T}(\mathbb{D}_\mathrm{n}) = \int_0^1 \{\mathbb{D}_\mathrm{n}(\mathrm{u})\}^2 \mathrm{du} = \frac{1}{12\mathrm{n}} + \sum_{\mathrm{t}=1}^{\mathrm{n}} \left\{ \mathrm{w_{n,t}} - \frac{(2\mathrm{t}-1)}{2\mathrm{n}} \right\}^2. \tag{21}$$

Let $c_t^{(d)}(\mathbf{v}) = \int_0^1 c_t(z, \mathbf{v}) dz$ be the density of the copula associated with X_t. To find the asymptotic behavior of $\mathbb{D}_n$, one needs to define the following functions for any $u \in (0,1)$:

$$\begin{aligned}
\mathbb{Z}_n(u) &= n^{-1} \sum_{t=1}^{n} \mathbf{B}_t^\top \int_{(0,1)^d} \dot{G}_{\boldsymbol{\beta}(t)} \left\{ G_{\boldsymbol{\beta}(t)}^{-1}(u) \right\} c_t\{\Gamma_t(u, \mathbf{v}), \mathbf{v}\} d\mathbf{v}, \\
\mathbb{X}_n(u) &= n^{-1} \sum_{t=1}^{n} \mathcal{A}_t^\top \{\mathbf{F}_t^{-1}(\mathbf{v})\} \nabla_\mathbf{v} \mathcal{D}_t\{\Gamma_t(t, \mathbf{v}), \mathbf{v}\} c_t\{\Gamma_t(u, \mathbf{v}), \mathbf{v}\} d\mathbf{v}, \\
\mathbb{C}_n(u) &= n^{-1} \sum_{t=1}^{n} \mathbf{h}_t^\top \int_{(0,1)^d} \dot{\mathcal{D}}_t\{\Gamma_t(u, \mathbf{v}), \mathbf{v}\} \mathfrak{d}_t(\mathbf{v}) d\mathbf{v}.
\end{aligned}$$

Further define the process $\mathbb{W}_n(u) = n^{-1/2} \sum_{t=1}^{n} \{\mathbb{I}(W_t \leq u) - u\}$, $u \in [0,1]$.

Corollary 1 *Assume that the smoothness conditions and the assumptions of Theorem 1 hold. Further assume that for any $u \in (0,1)$, the functions $\mathbb{Z}_n, \mathbb{C}_n, \mathbb{X}_n$ converge to continuous functions $\mathbb{Z}$, $\mathbb{C}$ and $\mathbb{X}$ respectively. Then, as $n \to \infty$, $\mathbb{D}_n$ converges in law to a continuous centered Gaussian process $\mathbb{D} = \mathbb{W} - \mu(\mathbb{K})$, where $\mathbb{W}$ is a Brownian bridge and $\mu(\mathbb{K}) = \boldsymbol{\Theta}^\top \mathbb{C} + \mathfrak{b}^\top \mathbb{Z} + \mathbb{A}^\top \mathbb{X}$.*

It then follows from Theorem 3, Corollary 1 and Genest and Rémillard (2008) that Algorithm 1 can also be used to bootstrap $\mathbb{D}$.

3 Applications to functionals of the conditional distribution

In what follows, we consider three examples of functionals of the conditional distribution: conditional quantiles, conditional mean, and conditional expected shortfalls. The proposed estimators are described next, together with their asymptotic behavior and bootstrapping methods.

3.1 Estimation of conditional quantiles

The associated quantile function at level α, denoted by $\mathcal{Q}_t(\alpha, \mathbf{x})$, is given by the left-continuous inverse of $\mathcal{K}_t$ viz.

$$\mathcal{Q}_t(\alpha, \mathbf{x}) = \inf\{y \in \mathbb{R} : \mathcal{K}_t(y, \mathbf{x}) \geq \alpha\}, \quad \alpha \in (0, 1). \tag{22}$$

Recently, the connection between marginal distribution functions and copulas have been used to find an explicit expression for the conditional quantile function (Kraus and Czado, 2017; Rémillard et al., 2017). As a result, the conditional quantile function $\mathcal{Q}_t$ depends only on the margins $\mathcal{G}_t$, $\mathbf{F}_t$ and the copula C_t viz.

$$\mathcal{Q}_t(\alpha, \mathbf{x}) = \mathcal{G}_t^{-1}\left[\Gamma_t\{\alpha, \mathbf{F}_t(\mathbf{x})\}\right], \tag{23}$$

where $\Gamma_t(\alpha, \mathbf{v})$ is the quantile of order α of the distribution function $\mathcal{D}_t(u, \mathbf{v})$, $u \in [0, 1]$, with $\mathbf{v} \in (0, 1)^d$.

Recently, Formula (23) was used by Nasri (2017) and Kraus and Czado (2017) for estimating the conditional quantile copula. They suggested to estimate the marginal distributions by nonparametric methods and assumed a parametric model for the copula. Expressions of conditional quantile functions for some well-known copula families were also provided. However, these approaches were proposed for the i.i.d. case and did not allow for the time-dependence of the distribution of the response, the distributions of covariates, or the copula function.

Next, let $G^{-1}_{\boldsymbol{\beta}(t,\boldsymbol{b}_n)}(\cdot)$, $\Gamma_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_n)}(\cdot, \mathbf{v})$ and $\mathcal{Q}_{n,t}(\cdot, \mathbf{x})$ be the (left-continuous) inverse functions of $G_{\boldsymbol{\beta}(t,\boldsymbol{b}_n)}(\cdot)$, $\mathcal{D}_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_n)}(\cdot, \mathbf{v})$ and $\mathcal{K}_{n,t}(\cdot, \mathbf{x})$ respectively. It then follows from (4) and (23) that a consistent estimator of $\mathcal{Q}_t(\alpha, \mathbf{x})$ is given by

$$\mathcal{Q}_{n,t}(\alpha, \mathbf{x}) = G^{-1}_{\boldsymbol{\beta}(t,\boldsymbol{b}_n)}\left[\Gamma_{\boldsymbol{\phi}(t,\boldsymbol{\theta_n})}\{\alpha, \mathbf{F}_{n,t}(\mathbf{x})\}\right], \quad \alpha \in (0, 1). \tag{24}$$

Set $\mathbb{Q}_{n,t}(u, \mathbf{x}) = n^{1/2}\{\mathcal{Q}_{n,t}(u, \mathbf{x}) - \mathcal{Q}_t(u, \mathbf{x})\}$, $u \in [0, 1]$, and $\mathbf{x} \in \mathbb{R}^d$. The next corollary states the weak convergence of the process $\mathbb{Q}_{n,t}$.

Corollary 2 *Suppose that the smoothness conditions hold, together with the assumptions of Theorem 1. Then, given $\mathbf{X}_t = \mathbf{x}$, and $\mathbf{v} = \mathbf{F}_t(\mathbf{x})$, as $n \to \infty$, $\mathbb{Q}_{n,t}$ converges in law to $\mathbb{Q}_t$, where $\mathbb{Q}_t(u, \mathbf{x}) = -\frac{\mathbb{K}_t\{\mathcal{Q}_t(u,\mathbf{x}),\mathbf{x}\}}{h_t\{\mathcal{Q}_t(u,\mathbf{x}),\mathbf{x}\}}$, $u \in (0, 1)$, and $h_t(y, \mathbf{x})$ is the density of $\mathcal{K}_t(y, \mathbf{x})$.*
In particular, we can show that for any $[a, b] \subset (0, 1)$, $n^{1/2} \sup_{u\in[a,b]} |\mathcal{Q}_{n,t}(u, \mathbf{x}) - \mathcal{Q}_t(u, \mathbf{x})|$ converges in law to $\sup_{u\in[a,b]} \left|\frac{\mathbb{K}_t\{\mathcal{Q}_t(u,\mathbf{x}),\mathbf{x}\}}{h_t\{\mathcal{Q}_t(u,\mathbf{x}),\mathbf{x}\}}\right|$.

Remark 1 *Using Algorithm 1 and Corollary 2, one can construct a 95% confidence interval about $\mathcal{Q}_t(\alpha, \mathbf{x})$. Set $y_{n,t} = \mathcal{Q}_{n,t}(\alpha, \mathbf{x})$, choose N large enough (say $N = 10000$), and let s_N be the standard deviation of the bootstrapped values $\mathbb{K}^{(k)}_{n,t}(y, \mathbf{x})$, $k \in \{1, \ldots, N\}$. Then a 95% interval for $\mathcal{Q}_t(\alpha, \mathbf{x})$ is $\mathcal{Q}_{n,t}(\alpha, \mathbf{x}) \pm 1.96\frac{s_N}{n^{1/2}h_{n,t}}$, where*

$$h_{n,t} = g_{\boldsymbol{\beta}_n(t)}(y_{n,t})\mathfrak{d}_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_n)}\left\{G_{\boldsymbol{\beta}_n(t)}(y_{n,t}), \mathbf{v}_n\right\} = g_{\boldsymbol{\beta}_n(t)}(y_{n,t}), \mathfrak{d}_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_n)}\left\{\Gamma_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_n)}(\alpha, \mathbf{v}_n), \mathbf{v}_n\right\},$$

and where $g_{\boldsymbol{\beta}}$ is the density of $G_{\boldsymbol{\beta}}$. Note that $\mathcal{K}_{n,t}(y_{n,t}, \mathbf{x}) = \alpha$, $t \in \{1, \ldots, n\}$. Furthermore, a uniform confidence bands about $\mathcal{Q}_t(\alpha, \mathbf{x})$ for $\alpha \in [\alpha_0, \alpha_1]$, with $0 < \alpha_0 \leq \alpha_1 < 1$ can also be constructed.

3.2 Conditional expectation estimator

We cannot use the estimator proposed by Noh et al. (2013) in our non-stationary setting because they used a weighted mean that cannot estimates $E(Y_t|\mathbf{X}_t = \mathbf{x})$ for a fixed t. However, there is another estimator that can be used. It is based on a functional of the conditional distribution. To this end, suppose that $Z \sim K$, where K is a distribution function. Assume that $E\{|Z|\} < \infty$. Then, one can write

$$E(Z) = \int_0^\infty \{1 - K(z)\}dz - \int_{-\infty}^0 K(z)dz. \tag{25}$$

Based on (25), one has

$$m_t(\mathbf{x}) = E(Y_t|\mathbf{X}_t = \mathbf{x}) = \int_0^\infty \{1 - \mathcal{K}_t(y, \mathbf{x})\}dy - \int_{-\infty}^0 \mathcal{K}_t(y, \mathbf{x})dy, \tag{26}$$

while the proposed estimator is

$$m_{n,t}(\mathbf{x}) = \int_0^\infty \{1 - \mathcal{K}_{n,t}(y, \mathbf{x})\}dy - \int_{-\infty}^0 \mathcal{K}_{n,t}(y, \mathbf{x})dy. \tag{27}$$

Corollary 3 *Suppose that the smoothness conditions hold, together with the assumptions of Theorem 1. Assume also that Y has a finite moment of order 2. Then, for given t and $\mathbf{X}_t = \mathbf{x}$, as $n \to \infty$,*

$$\mathbb{M}_{n,t}(\mathbf{x}) = n^{1/2}\{m_{n,t}(\mathbf{x}) - m_t(\mathbf{x})\} = -\int_{-\infty}^{+\infty} \mathbb{K}_{n,t}(y, \mathbf{x})dy$$

converges in law to a continuous centered Gaussian variable $\mathbb{M}_t(\mathbf{x})$, where

$$\mathbb{M}_t(\mathbf{x}) = -\int_{-\infty}^{+\infty} \mathbb{K}_t(y, \mathbf{x})dy. \tag{28}$$

Proof. The proof is based on the fact that for any $M > 0$, $\int_{-M}^{+M} \mathbb{K}_{n,t}(y, \mathbf{x})dy$ converges in law to $\int_{-M}^{+M} \mathbb{K}_t(y, \mathbf{x})dy$, being a continuous functional of $\mathbb{K}_{n,t}$ and that $\int_{-\infty}^{-M} \mathbb{K}_{n,t}(y, \mathbf{x})dy$ and $\int_M^{+\infty} \mathbb{K}_{n,t}(y, \mathbf{x})dy$, $\int_{-\infty}^{-M} \mathbb{K}_t(y, \mathbf{x})dy$ and $\int_M^{+\infty} \mathbb{K}_t(y, \mathbf{x})dy$ can all be made arbitrarily small in probability by choosing M large enough. The rest of the proof is similar to the one of Proposition 3.1 in Genest and Rémillard (2004) and uses Lemma F.1 in Genest et al. (2017). □

Remark 2 *Using Algorithm 1 and Corollary 3, one can construct a 95% confidence interval about $m_t(\mathbf{x})$. Choose N large enough (say $N = 10000$), and let s_N be the standard deviation of the bootstrapped values $\mathbb{M}_{n,t}^{(k)}(\mathbf{x}) = -\int_{-\infty}^{+\infty} \mathbb{K}_{n,t}^{(k)}(y, \mathbf{x})dy$, $k \in \{1, \ldots, N\}$. Then a 95% interval for $m_t(\mathbf{x})$ is $m_{n,t}(\mathbf{x}) \pm 1.96\frac{s_N}{\sqrt{n}}$.*

3.3 Conditional expected shortfall

Suppose $Z \sim K$, where K is a continuous distribution function with quantile function q. Assume that $E\{|Z|\} < \infty$. Then the expected shortfall of level $\alpha \in (0, 1)$ can be defined by $E(Z|Z < q_\alpha)$. It is then easy to show that

$$E(Z|Z < q_\alpha) = q_\alpha - \frac{1}{\alpha}\int_{-\infty}^{q_\alpha} K(z)dz. \tag{29}$$

Based on (29), for any $\alpha \in (0, 1)$, the conditional expected shortfall is given by

$$\mathrm{ES}_t(\alpha, \mathbf{x}) = \mathcal{Q}_t(\alpha, \mathbf{x}) - \frac{1}{\alpha}\int_{-\infty}^{\mathcal{Q}_t(\alpha,\mathbf{x})} \mathcal{K}_t(y, \mathbf{x})dy, \tag{30}$$

while the proposed estimator is

$$\mathrm{ES}_{n,t}(\alpha, \mathbf{x}) = \mathcal{Q}_{n,t}(\alpha, \mathbf{x}) - \frac{1}{\alpha}\int_{-\infty}^{\mathcal{Q}_{n,t}(\alpha,\mathbf{x})} \mathcal{K}_{n,t}(y, \mathbf{x})\}dy. \tag{31}$$

Corollary 4 *Suppose that the smoothness conditions hold, together with the assumptions of Theorem 1. Assume also that Y has a finite moment of order 2. Then, for given t and $\mathbf{X}_t = \mathbf{x}$, as $n \to \infty$, $\mathbb{E}_{n,t}(\alpha, \mathbf{x}) = n^{1/2}\{\mathrm{ES}_{n,t}(\alpha, \mathbf{x}) - \mathrm{ES}_t(\alpha, \mathbf{x})\}$ converges in law to a continuous centered Gaussian variable $\mathbb{E}_t(\alpha, \mathbf{x})$, where*

$$\mathbb{E}_t(\alpha, \mathbf{x}) = -\frac{1}{\alpha}\int_{-\infty}^{\mathcal{Q}_t(\alpha,\mathbf{x})} \mathbb{K}_t(y, \mathbf{x})dy. \tag{32}$$

Remark 3 *Using Algorithm 1 and Corollary 4, one can construct a 95% confidence interval about $\mathrm{ES}_t(\alpha, \mathbf{x})$, $\alpha \in (0, 1)$. Set $y_{n,t} = \mathcal{Q}_{n,t}(\alpha, \mathbf{x})$, choose N large enough (say $N = 10000$), and let s_N be the standard deviation of the bootstrapped values $\mathbb{E}^{(k)}_{n,t}(\alpha, \mathbf{x}) = -\frac{1}{\alpha}\int_{-\infty}^{y_{n,t}} \mathbb{K}^{(k)}_{n,t}(y, \mathbf{x})dy$, $k \in \{1, \ldots, N\}$. Then a 95% interval for $\mathrm{ES}_t(\alpha, \mathbf{x})$ is $\mathrm{ES}_{n,t}(\alpha, \mathbf{x}) \pm 1.96\frac{s_N}{\sqrt{n}}$.*

Remark 4 *One might also be interested in $E(Z|Z > q_\alpha)$. In this case, the associated formula is*

$$\mathcal{Q}_t(\alpha, \mathbf{x}) + \frac{1}{1-\alpha}\int_{\mathcal{Q}_t(\alpha,\mathbf{x})}^{\infty} \{1 - \mathcal{K}_t(y, \mathbf{x})\}dy, \tag{33}$$

the estimation is $\mathcal{Q}_{n,t}(\alpha, \mathbf{x}) + \frac{1}{1-\alpha}\int_{\mathcal{Q}_{n,t}(\alpha,\mathbf{x})}^{\infty}\{1 - \mathcal{K}_{n,t}(y, \mathbf{x})\}dy$, and the estimation error converges to

$$-\frac{1}{1-\alpha}\int_{\mathcal{Q}_t(\alpha,\mathbf{x})}^{\infty} \mathbb{K}_t(y, \mathbf{x})dy.$$

The details are left to the reader.

4 Simulation study

In this section we consider six Monte Carlo experiments for assessing the level and power of the proposed goodness-of-fit tests based on Kolmogorov-Smirnov and Cramér-von Mises type statistics defined in Section 2.6. We generated random samples of size $n \in \{100, 250\}$ from four bivariate copula families: Clayton, Gumbel, Gaussian and Student (with $\nu = 5$). In the first four experiments, we considered a "linear" case, i.e., $\tau_t = 1/\left(1 + e^{-H_t^{(1)}\theta^{(1)}}\right)$, where $H_t^{(1)}\theta^{(1)} = .4055t/n$, so that $\tau_1 = .5004$ and $\tau_n = 0.6$. In the last two experiments, we considered a "quadratic" case, i.e., $\tau_t = 1/\left(1 + e^{-H_t^{(2)}\theta^{(2)}}\right)$ where $H_t^{(2)}\theta^{(2)} = -0.4055 + 2.0637t/n - 2.5055t^2/n$. Both graphs of τ_t are displayed in Figure 1.

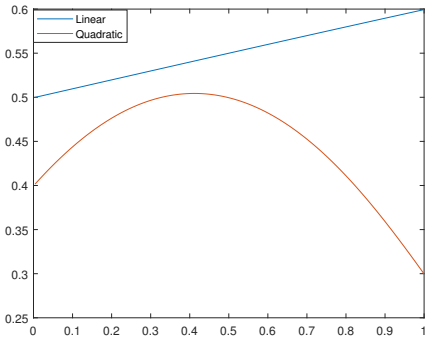

Figure 1: Kendall's tau for the linear and quadratic models with $n = 250$.

The same four families were used under the null hypothesis. For example, we generated data from Clayton copula and tested four null hypothesis: Clayton, Gumbel, Gaussian and Student copula. We repeat this by generating data from Gumbel, Gaussian and Student copulas. Furthermore, for each experiment 1000 replications and in each replication, 100 bootstrap samples ($N = 100$) were used to compute the p-value of the test statistics.

For the first two experiments (linear model with $n \in \{100, 250\}$), the marginal distributions were assumed to be unknown but constant, in order to focus only on the copula family under the null hypothesis. Next, for the last four experiments, we considered Gaussian time-dependent margins for Y ($\mu_t = -0.5t/n$, $\sigma = 0.2$) and X ($\mu_t = 0.5 + 2t/n$, $\sigma = 0.1$), and both the linear and quadratic cases for the copula families.

The results of the six experiments are displayed in Table 1. As seen for these results, the levels are not significantly different from the target level of 5% . Also the power of the goodness-of-fit test is satisfactory, even with $n = 100$. As expected, the power increases with the sample size. Furthermore, it follows from Experiments 1–4 that for the linear case, the proposed goodness-of-fit test perform quite well whether the margins are constant or linear time-dependent. However, in the quadratic case, the power of the test decreases in general since there are more parameters to estimate.

Finally, it seems that most of the time, the goodness-of-fit test based on the Cramér-von Mises statistic is more powerful that the one based on the Kolmogorov-Smirnov statistic.

Table 1: Percentage of rejection for the Kolmogorov-Smirnov test statistic (KS) and Cramér-von Mises statistic (CVM) for a target level of 5% for the six Monte Carlo experiments. The levels of the tests are displayed in bold.

	Copula family under H_1							
	Clayton		Gumbel		Gaussian		Student	
H_0	KS	CVM	KS	CVM	KS	CVM	KS	CVM
Exp. 1: Linear case with unknown constant margins ($n = 100$)								
Clayton	**4.5**	**4.1**	30.2	37.9	17.7	20.8	16.3	21.4
Gumbel	47.9	62.7	**4.2**	**4.0**	9.2	12.0	8.9	12.3
Gaussian	35.7	45.9	12.8	17.3	**5.3**	**4.5**	9.2	11.6
Student	19.8	24.3	7.3	7.7	4.0	4.1	**3.3**	**3.0**
Exp. 2: Linear case with unknown constant margins ($n = 250$)								
Clayton	**4.4**	**3.6**	72.7	84.6	47.3	58.1	42.2	53.6
Gumbel	92.4	98.2	**4.9**	**5.5**	21.1	29.4	18.8	24.6
Gaussian	81.5	92.1	30.2	40.4	**6.0**	**5.7**	15.9	24.0
Student	60.5	73.5	19.2	23.6	5.0	4.3	**4.2**	**3.8**
Exp. 3: Linear case with estimated Gaussian margins ($n = 100$)								
Clayton	**4.3**	**4.7**	29.8	38.6	22.4	28.0	20.4	26.8
Gumbel	54.0	66.9	**4.5**	**5.0**	10.3	11.4	12.3	14.8
Gaussian	43.3	52.3	16.9	19.3	**4.8**	**4.3**	10.7	13.1
Student	26.8	33.7	9.2	11.1	4.1	4.2	**5.0**	**3.4**
Exp. 4: Linear case with estimated Gaussian margins ($n = 250$)								
Clayton	**5.8**	**5.2**	67.1	78.0	44.6	55.1	37.8	46.7
Gumbel	91.6	96.5	**5.6**	**5.8**	21.7	30.2	21.6	29.6
Gaussian	84.1	91.7	35.5	43.7	**5.2**	**5.6**	15.0	21.0
Student	65.2	76.5	21.9	25.9	5.4	6.2	**3.9**	**3.7**
Exp. 5: Quadratic case with estimated Gaussian margins ($n = 100$)								
Clayton	**5.3**	**5.4**	18.7	23.5	17.7	19.3	12.2	14.3
Gumbel	36.2	44.6	**4.7**	**4.3**	10.7	13.2	9.0	11.4
Gaussian	25.5	33.2	9.8	13.8	**4.6**	**5.1**	5.4	7.1
Student	18.0	21.8	8.5	8.2	4.2	4.2	**3.6**	**2.7**
Exp. 6: Quadratic case with estimated Gaussian margins ($n = 250$)								
Clayton	**3.7**	**3.7**	43.4	54.4	30.4	37.6	23.8	30.7
Gumbel	78.4	89.3	**4.9**	**4.7**	17.2	23.8	16.1	20.7
Gaussian	60.3	72.4	20.4	27.2	**4.1**	**4.4**	8.8	11.1
Student	32.1	40.6	7.4	10.8	3.9	3.8	**3.2**	**2.2**

Remark 5 *When estimating the parameter ν of the Student copula, we restricted its value to be lower than* 25*, since the Student copula converges to the Gaussian copula when ν tends to infinity.*

We did not consider the Frank copula family in these simulations because there is no explicit formula expressing its parameter in terms of Kendall's tau. Since the latter varies with time, it means that one has to perform n numerical inversions to obtain the associated parameter for a given τ_t, $t \in \{1, \ldots, n\}$. To have

an idea of the computing time required to estimate the parameters when $n = 250$, we refer the reader to the last line of Table 3. Performing the goodness-of-fit test for the Frank copula family in the linear case described previously requires about 20 minutes, while it takes 13 seconds for the Gaussian copula family. Note also that performing 1000 goodness-of-fit tests with $N = 100$ bootstrap samples when $n = 250$ requires about 116000 seconds on a laptop with 6 cores, even if we were using parallel computing. The construction of Table 1 below required 3 weeks.

5 Applications

In this section, we illustrate the proposed methodology with one simulated data set suggested in Dette et al. (2014), and two real data sets. The first application is basically a discussion on the choice of a copula when the relation between the variable of interest and a covariate is quadratic, so it is assumed that the copula and the margins are constant. However, in the last two examples, either the margins or the copula vary with time.

5.1 Dette et al. (2014)'s example

In this paper, the authors tried to model the dependance between X_t and Y_t, where $Y_t = (X_t - 0.5)^2 + \epsilon_t$, with $\epsilon_t \sim N(0, \sigma^2)$ and $\sigma = 0.1$, using a copula-based mean regression model as describe in Section 3.2. They assumed that the copula and the margins were constant over time. Then, they looked for a copula family which could capture this kind of dependance. However, they limited their search to the basic families (Elliptical and Archimedean), and their rotations. This is indeed true that these families fail to reproduce non monotonic dependance.

In the literature, there are many copula families for which $E(Y|X = x)$ is not monotonic in x. Here is an example: consider the chi-square copula introduced by Quessy et al. (2016). A chi-square copula with parameters $a_1, a_2 \in \mathbb{R}$, and $\rho \in [-1, 1]$ is the copula associated with the random variables $(Z_1 - a_1)^2$ and $(Z_2 - a_2)^2$, where Z_1 and Z_2 are joint standard Gaussian variables with correlation ρ. As an experiment, we simulated 1000 pairs of variables (X_t, Y_t) from the above regression model and we estimated the parameters for the chi-square copula. Rounding the parameters, we got $a_1 = 2.6$, $a_2 = 0$, and $\rho = 0.99$.

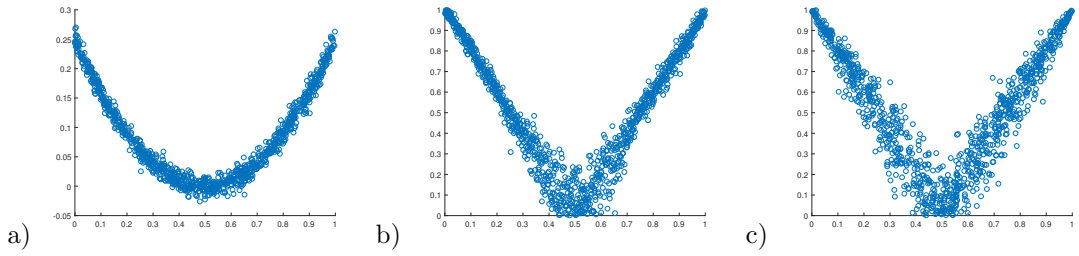

Figure 2: Panel a: Simulated data for the regression model; panel b: pseudo-observations for the regression model; panel c: simulated data from a chi-square copula with parameters $a_1 = 2.6$, $a_2 = 0$, and $\rho = 0.99$.

Next, we applied this copula to the simulated dataset defined previously. By using the conditional expectation estimator defined in Section 3.2, one can check from Figure 3 that the quadratic dependance can be indeed reproduced quite well with a copula-based model.

5.2 Estimation of conditional quantile of maximum annual streamflow

For Southeastern Canada extreme streamflow characterization, daily amounts in (m^3/s) of streamflow have been extracted from all of the 312 stations located in Ontario (ON), Québec (QC), Newfoundland and Labrador (NFL), Nova Scotia (NS), New Brunswick (NB), and Prince Edward Island (PEI) provinces. Data comes from Environment Canada's HYDAT interface (ftp://arccf10.tor.ec.gc.ca/wsc/software/HYDAT/) and the center of Québec water expertise (https://www.cehq.gouv.qc.ca/). We kept only stations

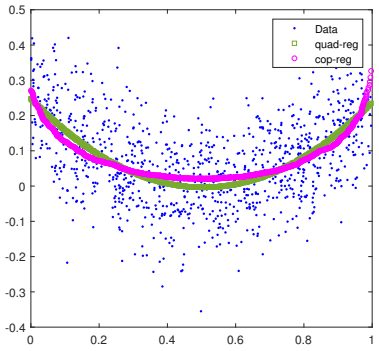

Figure 3: Quadratic regression curve (squares) and conditional expectation estimator (circles) for the simulated dataset.

with: (i) record length of at least 30 full years to ensure good quality fitting; (ii) non serial dependencies; (iii) some form of nonstationarity. Finally, 33 stations were chosen, of which 17 are located in ON, 8 in QC, 4 in NS, 3 in NL, and 1 in NB. Figures 4 and 5 illustrate respectively the geographic location of all selected stations and the variations of daily annual maximum streaflows (AMS) at some chosen station from each province for case illustration. For this study, we tested the dependence between the AMS and several climatic indices covariables: Atlantic Multidecadal Oscillation (AMO), Pacific Decadal Oscillation (PDO), El Niño–Southern Oscillation (ENSO), etc. (see, e.g. Nasri et al. (2017) for definitions). Climate indices data come from NOAA website (`http://www.esrl.noaa.gov`). These covariates are considered at the same day for which the maximum of streamflow is observed for each station. Ultimately, only the covariate that have the most significant dependence with the variables of interest was kept, which is AMO. AMO is found to have the most significant dependence with AMS time series and it indicates the fluctuation in the sea temperature in the North Atlantic ocean. Figure 6 illustrates the variations of the AMO index at the chosen stations. The change-point in the AMO time series is studied in all of the 33 stations using the same tests cited above for the change-point in AMS series. The results show that generally, in all selected stations, there is indeed a change-point, which is coherent with the existing literature (Knudsen et al., 2011). The change-point in dependence (copula) between AMO and AMS is also investigated by using the test proposed in Holmes et al. (2013) and the results show that, in most of the stations, the dependence is constant over time.

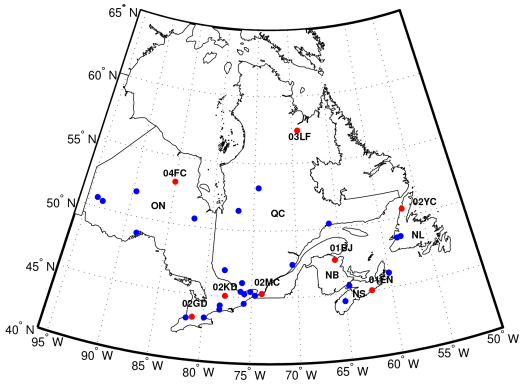

Figure 4: Study region in Southeastern Canada and locations of 33 stations in this study region. Red highlighted stations are targeted ones for case illustration.

Before estimating the conditional quantile functions for each selected station, we have to choose first the marginal distributions for the covariate (AMO) and the dependent variable (AMS), and then we can select the best copula function fitting the data. For the marginal distributions, several cumulative functions are compared, including Normal, 2-parameter Weibull, GEV, lognormal and Gamma. This comparison is done by using the Akaike information criterion (AIC)(Akaike, 1974). The variation in time of the parameters of the selected marginal distributions is described by using the B-spline function structure (De Boor, 2001;

Nasri et al., 2017). Note that a B-spline is a polynomial piecewise function which depends on a number of knots and a degree of function. When the number of knots is equal to one, B-spline is a polynomial function. As for copulas, several dependence models were compared, including Frank, Clayton, Gumbel, Gaussian and Student, in order to find the best model. The choice is done by calculating the p-value of the goodness-of fit test statistics $S_n(B)$ and $S_n(C)$ developed in Genest et al. (2009).

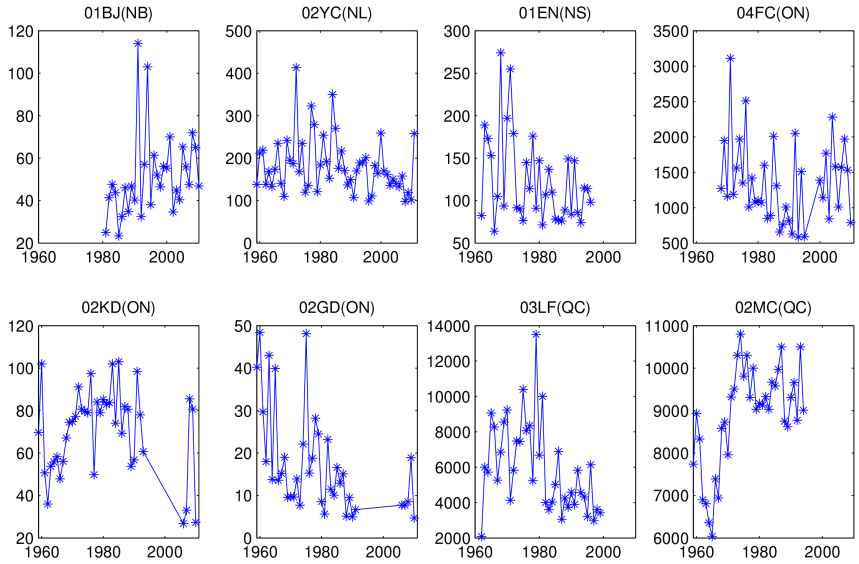

Figure 5: Variation of maximum daily annual streamflows (m^3/s) for the stations selected for case illustration. Here, we have three stations from Ontario, two from Quebec and one from NS, NL and NB.

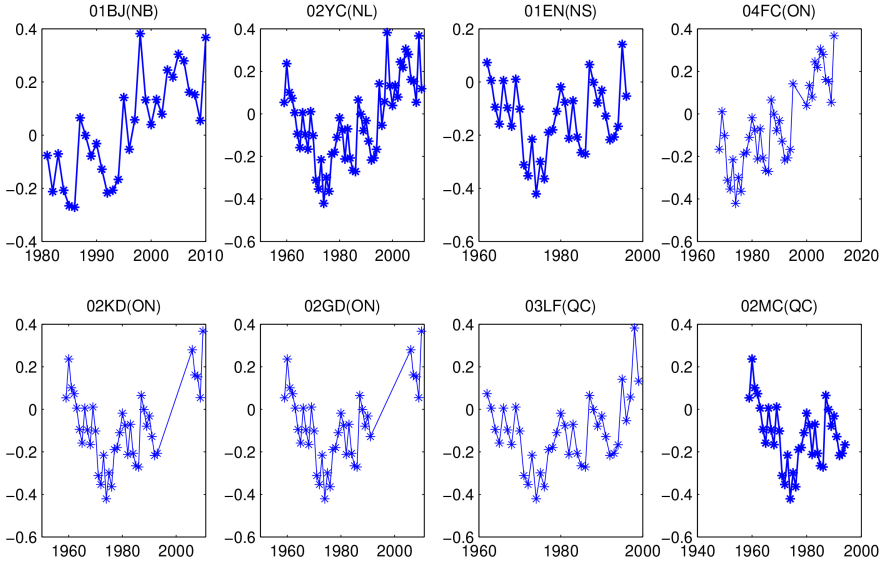

Figure 6: Variation of AMO index for the stations selected for case illustration. Here, we have three stations from Ontario, two from Quebec and one from NS, NL and NB.

As described in the methodology section, we have to choose the margins for the studied variables, the best B-spline function linking their parameters and time, and also the best copula model. For all stations, the Gaussian distribution is selected for the AMO, while the GEV is chosen for the response variable. The results show that only the location parameter for the Gaussian and GEV varies over time. This variation is modeled generally by two piecewise linear functions. For the copula models, the Gaussian copula was selected most of the time but for some stations in the Southern region, where the Student copula was chosen. Recall

that the conditional quantile of a Student copula, with parameters $\nu > 0$ and $\rho \in (0, 1)$, is given by

$$\Gamma(u, v) = t_\nu \left[\rho t_\nu^{-1}(u) + t_{\nu+1}^{-1}(\alpha) \sqrt{\frac{\nu + \{t_\nu^{-1}(u)\}^2}{\nu + 1} (1 - \rho^2)} \right], \quad \alpha \in (0, 1),$$

where t_ν is the c.d.f. of a Student distribution with ν degrees of freedom. Figure 7 illustrates these results for all stations. Using the best selected copula models together with the margins, a nonstationary conditional quantile is estimated for each station, for levels $\alpha \in \{0.5, 0.9\}$, corresponding to return periods of $T = 2$ years and $T = 10$ years. Quantiles for high returns periods (e.g. $T = 50$, $T = 100$) could also be estimated. The 95% confidence band is constructed for those quantiles. Figure 8 displays the relevant results for the chosen stations. It can be seen that generally, for higher levels, the nonstationary conditional quantiles take much larger values that the stationary unconditional quantiles. For example, the results for station "02KD" in ON shows that the stationary unconditional median is equal to 65 (m^3/s). However, the nonstationary conditional quantile can reach 83 (m^3/s). The same behaviour can be observed for other stations, confirming the importance of considering nonstationary conditional quantiles for better water resource management practices.

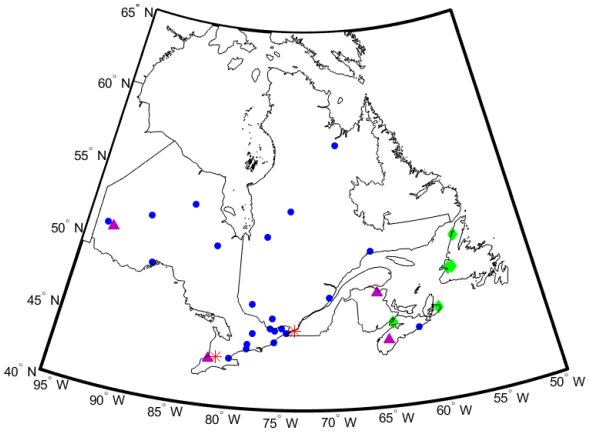

Figure 7: All stations depicted by blue circles have a Gaussian dependence structure with negative Kendall's tau and one or two piecewise linear dependence functions between time and location parameter. Red stars highlighted stations have two piecewise quadratic dependence between time and location parameter, while the purple triangles display a positive Kendall's tau, and the green squares have a Student copula dependence structure with two piecewise linear dependance functions.

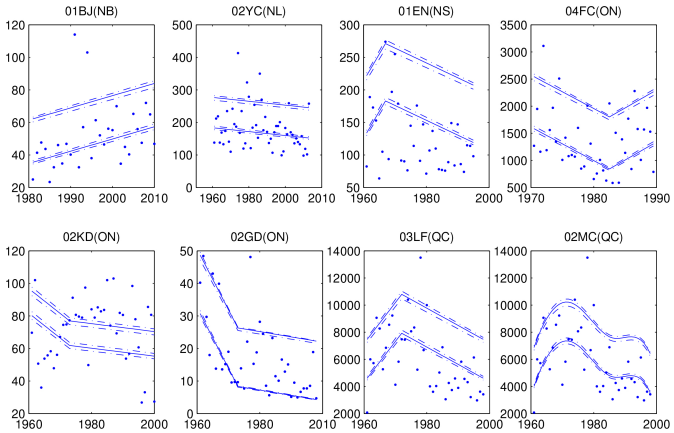

Figure 8: 0.5 and 0.9 nonstationary conditional quantile results with their corresponding uniform confidence bands for the selected stations.

5.3 Bivariate HMM models

There have been a huge literature on the dependence between individual stocks and the market, the most popular being of the form $Y_t = \alpha + \beta X_t + \epsilon_t$, where Y_t is the return of an individual stock at period t, and X_t is the return of the market where the stock is traded. The coefficient β is interpreted as the relative risk of the stock compared to the market, leading to a linear dependence between the stock returns and market returns. However this is quite limitative if the dependence is nonlinear. This is why it might be much better to look at copula-based methods.

Here we consider the monthly log-returns of Apple and the Nasdaq index from December 1996 to October 2017 and we fit a dynamic model to each time series separately. They are displayed in Figure 9. We first tried to fit GARCH models with Gaussian and unknown innovations but these models were rejected, using the tests proposed in Ghoudi and Rémillard (2013, 2018). In order to fit this dataset better, we propose to use regime switching models for the dynamics of each time series. Recall that in a regime-switching model (X_t, τ_t), the regimes $\tau_t \in \{1, \ldots, \ell\}$ are not observed and they are modeled by a finite Markov chain with transition matrix Q. Then, given the regimes $\tau_1 = i_1, \ldots, \tau_n = i_n$, the variables $X_1, \ldots, X_n$ are independent with distribution functions $F^{(i_1)}, \ldots, F^{(i_n)}$, where $F^{(j)}$ and $f^{(j)}$ are respectively the cdf and the density under regime $j \in \{1, \ldots, \ell\}$.

According to Rémillard et al. (2017), if $\eta_{t-1}(j)$ is the probability of being in regime $j \in \{1, \ldots, \ell\}$ at time $t-1$ given the past observations $X_1, \ldots, X_{t-1}$, then the conditional distribution of X_t given the past has distribution function $F_t(x) = \sum_{i=1}^{\ell} F^{(i)}(x) W_{t-1}(i)$, with density $f_t(x) = \sum_{i=1}^{\ell} f^{(i)}(x) W_{t-1}(i)$, where $W_{t-1}(i) = \sum_{j=1}^{\ell} \eta_{t-1}(j) Q_{ji}$ is the probability of being in regime j at time t given the past observations. It then follows that the sequence $V_t = F_t(X_t)$ are i.i.d. uniform random variables.

Here, to simplify the presentation, we choose Gaussian distributions with mean μ and standard deviation σ depending on the regimes. Using the goodness-of-fit test proposed in Rémillard et al. (2017), we find that the optimal number of regimes for both financial time series is two, with p-values of 13.4% and 9.7 % respectively, using 1000 bootstrap samples. The estimated parameters for both time series are given in Table 2.

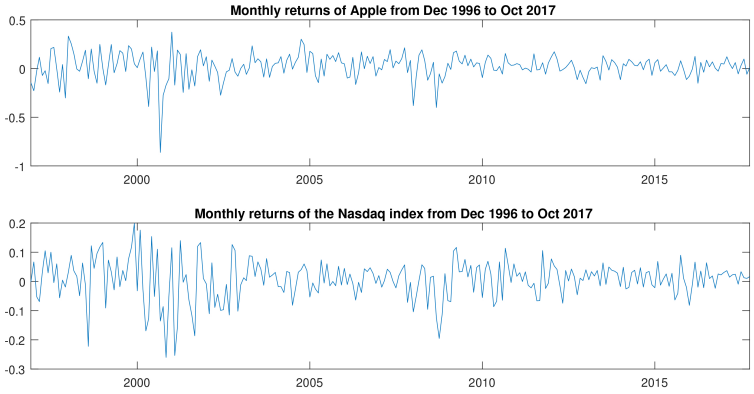

Figure 9: Variation of Apple and Nasdaq index over time.

Table 2: Estimated parameters for Apple and the Nasdaq index using Gaussian hidden Markov models. Here, ν is the stationary distribution of the regimes.

	Apple		Nasdaq	
Parameter	Regime 1	Regime 2	Regime 1	Regime 2
μ	0.0052	0.0304	-0.0098	0.0132
σ	0.1917	0.0776	0.1067	0.0412
ν	0.2940	0.7060	0.2820	0.7180
Q	$\begin{pmatrix} 0.9635 & 0.0365 \\ 0.0152 & 0.9848 \end{pmatrix}$		$\begin{pmatrix} 0.9607 & 0.0393 \\ 0.0154 & 0.9846 \end{pmatrix}$	

From now on, let X denotes the monthly returns of the Nasdaq index and let Y denotes de monthly returns of Apple. Let G_t and F_t be the conditional distributions of Y_t and X_t given the past observations, as defined previously. Further set $U_t = G_t(Y_t)$ and $V_t = F_t(X_t)$. As proposed in Nasri and Rémillard (2017), let C_t be the copula associated with (U_t, V_t). Since F_t and G_t are not known exactly, (U_t, V_t) is never observed and one has to use the pseudo-observations $u_{t,n} = G_{t,n}(Y_t)$ and $v_{t,n} = F_{t,n}(Y_t)$, where $F_{t,n}$ and $G_{t,n}$ are computed with the parameters of Table 2. As shown in Nasri and Rémillard (2017), the pseudo-observations $(u_{t,n}, v_{t,n})$ can be used as if (U_t, V_t) were observed, as long as one takes the ranks to estimate the copula and its parameters.

Next, to check whether the copula is time-dependent, we performed a change-point test as in Rémillard (2013). We got a p-value of 3.4%, so we do conclude that the copula is time-dependent. Since the simplest model to implement is a linear case for τ, i.e., $\tau_t = 1 \Big/ \left(1 + e^{-\theta_0 - \theta_1 t/n}\right)$, this is what we try first, applying this to five copula families: Clayton, Gumbel, Frank, Gaussian and Student. We did not considered the Frank family for the Monte Carlo experiments described in Section 4, but for a goodness-of-fit test, one can consider it even if it takes a long time to compute. The results of the goodness-of-fit tests for the five copula families are displayed in Table 3, together with the computation time. Clearly, the best model is the Gaussian copula for which one gets $\theta_0 = -1.1601$ and $\theta_1 = 1.4751$. The corresponding values of ρ_t are displayed in Figure 10, while the conditional 5% expected shortfall and the conditional expectation of the return of Apple given the Nasdaq index for November 2017 are displayed in Figure 11.

Table 3: P-values in percentage calculated with $N = 100$ bootstrap samples.

	Copula family				
Statistic	Clayton	Gumbel	Frank	Gaussian	Student
Kolmogorov-Smirnov	0	25	26	91	8
Cramér-von Mises	0	33	14	65	10
Computation time (sec)	14	20	1183	13	159

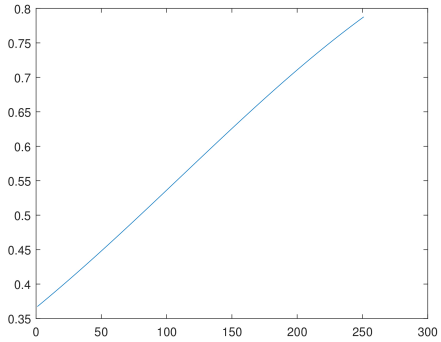

Figure 10: Graph of ρ_t for the Gaussian copula.

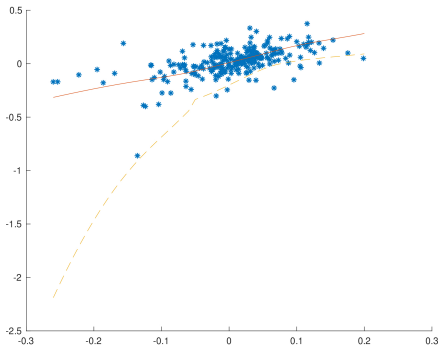

Figure 11: Copula-based conditional expectation estimator (plain line) and 5% conditional expected shortfall (dashed line) for the next month returns of Apple given Nasdaq returns.

6 Conclusion

Statistical risk management is of great importance in finance, hydrology and many other fields of applications. For example, in the last two decades, one has witnessed the development of a number of statistical modeling approaches for extreme quantile functions in the presence of non-stationarity of the variable of interest and/or the covariates.

In this paper, we proposed a copula-based approach to estimate non-stationary marginal, copula and condition distribution functions. Doing aso we were able to propose estimators for conditional quantile functions, conditiona expectations, and conditional expected shortfalls. More precisely, we assumed a parametric model for the time-dependent copula function and the time-dependent marginal distribution of the response variable, while we used a combination of nonparametric and time-dependent parametric models for the marginal distributions of the covariates, depending whether they are stationary or not. Under some smoothness conditions, the asymptotic normality of the proposed estimators is obtained. Also, we proposed a bootstrap procedure in order to construct uniform confidence bands around functionals of the conditional distribution, As a by-product, we obtained a formal goodness-of-fit test, and through Monte Carlo experiments, we were able to show that the proposed tests are powerful enough. Finally, to illustrate the approach, we considered three applications, one with simulated data, one using a hydro-climatic dataset, and one with a financial dataset. Finally, note that the proposed methodology can easily be adapted to a large number of covariates, by using vine copula structures as in Kraus and Czado (2017) and Rémillard et al. (2017).

A Proofs of the main results

In this section, we study the asymptotic behavior of the estimators, and we prove the validity of both the bootstrap algorithm and the goodness-of-fit test.

A.1 Proof of Theorem 1

Recall that $\mathcal{G}_t = G_{\boldsymbol{\beta}(t)}$. Under the smoothness assumptions of the copula density $c_{\boldsymbol{\phi}}$, and $\boldsymbol{\phi}(t, \boldsymbol{\theta})$, one gets

$$
\begin{aligned}
0 &= n^{-1/2} \sum_{t=1}^{n} \frac{\mathbf{h}_t^\top \dot{c}_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_n)}(U_{n,t}, \mathbf{V}_{n,t})}{c_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_n)}(U_{n,t}, \mathbf{V}_{n,t})} \\
&= \mathcal{W}_n + n^{-1} \sum_{t=1}^{n} \mathbf{h}_t^\top \left\{ \frac{\nabla_{\mathbf{v}} \dot{c}_t(U_t, \mathbf{V}_t)}{c_t(U_t, \mathbf{V}_t)} - \frac{\dot{c}_t(U_t, \mathbf{V}_t) \nabla_{\mathbf{v}} c_t(U_t, \mathbf{V}_t)^\top}{c_t^2(U_t, \mathbf{V}_t)} \right\} \mathbb{F}_{n,t}(\mathbf{X}_t) \\
&\quad + n^{-1} \sum_{t=1}^{n} \mathbf{h}_t^\top \left\{ \frac{\partial_u \dot{c}_t(U_t, \mathbf{V}_t)}{c_t(U_t, \mathbf{V}_t)} - \frac{\dot{c}_t(U_t, \mathbf{V}_t) \partial_u c_t(U_t, \mathbf{V}_t)}{c_t^2(U_t, \mathbf{V}_t)} \right\} \mathbb{G}_{n,t}(Y_t) - \mathcal{I}_n \boldsymbol{\Theta}_n + o_P(1) \\
&= \mathcal{W}_n + \int_0^1 \int_{\mathbb{R}^d} \left[n^{-1} \sum_{t=1}^{n} \mathbf{h}_t^\top \left\{ \nabla_{\mathbf{v}} \dot{c}_t\{u, \mathbf{F}_t(\mathbf{x})\} - \frac{\dot{c}_t\{u, \mathbf{F}_t(\mathbf{x})\} \nabla_{\mathbf{v}} c_t\{u, \mathbf{F}_t(\mathbf{x})\}^\top}{c_t\{u, \mathbf{F}_t(\mathbf{x})\}} \right\} \mathcal{A}_t(\mathbf{x}) f_t(\mathbf{x}) \right] du \mathbb{A}_n(\mathbf{x}) d\mathbf{x} \\
&\quad + \left[n^{-1} \sum_{t=1}^{n} \mathbf{h}_t^\top \int_{(0,1)^{d+1}} \left\{ \partial_u \dot{c}_t(u, \mathbf{v}) - \frac{\dot{c}_t(u, \mathbf{v}) \partial_u c_t(u, \mathbf{v})}{c_t(u, \mathbf{v})} \right\} \mathcal{G}_t \circ \mathcal{G}_t^{-1}(u)^\top \boldsymbol{B}_t du d\mathbf{v} \right] \mathfrak{b}_n \\
&\quad - \mathcal{I} \boldsymbol{\Theta}_n + o_P(1) \\
&= \mathcal{W}_n + \int_{\mathbb{R}^d} J_n(\mathbf{x}) \mathbb{A}_n(\mathbf{x}) d\mathbf{x} + \mathbf{K}_n \mathfrak{b}_n - \mathcal{I} \boldsymbol{\Theta}_n + o_P(1).
\end{aligned}
$$

Then, as $n \to \infty$, $(\mathfrak{b}_n, \mathbb{A}_n, \mathcal{W}_n, \boldsymbol{\Theta}_n)$ converges in law to $(\mathfrak{b}, \mathbb{A}, \mathcal{W}, \boldsymbol{\Theta})$, the limit of $\boldsymbol{\Theta}_n$ can be deduced. Note in passing that $E(\mathfrak{b}\mathcal{W}^\top) = E(\mathbf{q}^{(j)}\mathfrak{b}^\top) = E(\mathbf{q}^{(j)}\mathcal{W}^\top) = 0$, for any $j \notin \mathcal{L}$, while for any $j \in \mathcal{L}$ and any $v_j \in [0, 1]$, $E\{\mathfrak{b}\mathbb{B}_j^\top(v_j)) = E\{\mathbb{B}_j(v_j)\mathcal{W}^\top\} = 0$. This results will be useful in the proof of Theorem 3. □

A.2 Proof of Theorem 2

The asymptotic behavior of the parametric quantile process, follows readily from the Delta method (van der Vaart and Wellner, 1996). It is also similar to the proofs in Rémillard et al. (2017). □

A.3 Proof of Theorem 3

First, we need an auxiliary result, which is an easy extension of the work of Genest and Rémillard (2008) to the non-stationary case. To this end, set

$$\mathcal{U}_n = n^{-1/2} \sum_{t=1}^{n} \mathbf{B}_t^\top \dot{G}_t(Y_t^\star) \frac{\partial_u c_t(U_t^\star, V_t^\star)}{c_t(U_t^\star, V_t^\star)}, \tag{34}$$

$$\mathbf{M}_n = n^{-1} \sum_{t=1}^{n} \mathbf{B}_t^\top \dot{G}_t(Y_t^\star) \frac{\partial_u c_t(U_t^\star, V_t^\star) \partial_u c_t(U_t^\star, V_t^\star) \dot{G}_t(Y_t^\star)^\top \mathbf{B}_t}{c_t^2(U_t^\star, V_t^\star)}, \tag{35}$$

$$\mathcal{E}_n^{(j)} = n^{-1/2} \sum_{t=1}^{n} (\mathbf{Q}_t^{(j)})^\top \{\dot{F}_t^{(j)}\}^\top (X_{tj}^\star) \frac{\partial_{v_j} c_t(U_t^\star, V_t^\star)}{c_t(U_t^\star, V_t^\star)}, \quad j \notin \mathcal{L}, \tag{36}$$

$$\mathbf{E}_n^{(j)} = n^{-1} \sum_{t=1}^{n} (\mathbf{Q}_t^{(j)})^\top \dot{F}_t^{(j)}(X_{tj}^\star) \frac{\partial_{v_j} c_t(U_t^\star, V_t^\star) \partial_{v_j} c_t(U_t^\star, V_t^\star) \dot{F}_t^{(j)}(X_{tj}^\star)^\top \mathbf{Q}_t^{(j)}}{c_t^2(U_t^\star, V_t^\star)}, \quad j \notin \mathcal{L}. \tag{37}$$

For the proof of the theorem, we need the following auxiliary lemma.

Lemma 1 *Let $(Y_1^\star, \mathbf{X}_1^\star), \ldots, (Y_n^\star, \mathbf{X}_n^\star)$ be an independent copy of $(Y_1, \mathbf{X}_1), \ldots, (Y_n, \mathbf{X}_n)$, and denote by $\mathcal{P}_n$ the joint distribution of the latter. Next, define*

$$\mathbf{F}_{\boldsymbol{\gamma}(t,\mathbf{q})}^{(j)}(x_j) = \begin{cases} F^{(j)}(x_j), & j \in \mathcal{L}, \\ F_{\boldsymbol{\gamma}^{(j)}(t,\mathbf{q}^{(j)})}^{(j)}(x_j), & j \notin \mathcal{L}. \end{cases}$$

Further let $\mathcal{P}_n^\star$ be the distribution defined by the log-likelihood ratio

$$\begin{aligned} \ell_n &= \log\left[\frac{d\mathcal{P}_n^\star}{d\mathcal{P}_n}\{(Y_1^\star, \mathbf{X}_1^\star), \ldots, (Y_n^\star, \mathbf{X}_n^\star)\}\right] \\ &= \sum_{t=1}^{n} \log\left[\frac{g_{\boldsymbol{\beta}(t,\boldsymbol{b}_n)}(Y_t^\star)}{g_{\boldsymbol{\beta}(t,\boldsymbol{b}_0)}(Y_t^\star)}\right] + \sum_{j \notin \mathcal{L}} \sum_{t=1}^{n} \log\left[\frac{f_{\boldsymbol{\gamma}^{(j)}(t,\mathbf{q}_n)}^{(j)}(X_{tj}^\star)}{f_{\boldsymbol{\gamma}^{(j)}(t,\mathbf{q}_0)}^{(j)}(X_{tj}^\star)}\right] \\ &\quad + \sum_{t=1}^{n} \log\left[\frac{c_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_n)}\left\{G_{\boldsymbol{\beta}(t,\boldsymbol{b}_n)}(Y_t^\star), F_{\boldsymbol{\gamma}^{(j)}(t,\mathbf{q}_n^{(j)})}^{(j)}(X_{tj}^\star)\right\}}{c_{\boldsymbol{\phi}(t,\boldsymbol{\theta}_0)}(U_t^\star, \mathbf{V}_t^\star)}\right], \end{aligned} \tag{38}$$

where $U_t^\star = G_{\boldsymbol{\beta}(t,\boldsymbol{b}_0)}(Y_t^\star)$ and $\mathbf{V}_t^\star = \mathbf{F}_t(\mathbf{X}_t^\star)$, $t \in \{1, \ldots, n\}$. Finally, let $\mathbf{S}_n^\star$, $\mathfrak{b}_n^\star$, $\mathcal{V}_n^\star$, $\mathbf{q}_n^\star$, $\mathcal{W}_n^\star$, $\mathcal{I}_n^\star$ be defined by (6), (7), (10), (11), (16), and (17), using $(Y_1^\star, \mathbf{X}_1^\star), \ldots, (Y_n^\star, \mathbf{X}_n^\star)$. Under the smoothness assumptions and conditions of Theorem 1, $(\mathfrak{b}_n, \boldsymbol{\Theta}_n, \mathbb{A}_n, \mathfrak{b}_n^\star, \boldsymbol{\Theta}_n^\star, \mathbb{A}_n^\star, \ell_n)$ converges in law to $(\mathfrak{b}, \boldsymbol{\Theta}, \mathbb{A}, \mathfrak{b}^\perp, \boldsymbol{\Theta}^\perp, \mathbb{A}^\perp, \ell)$, where $(\mathfrak{b}^\perp, \boldsymbol{\Theta}^\perp, \mathbb{A}^\perp)$ is an independent copy of $(\mathfrak{b}, \boldsymbol{\Theta}, \mathbb{A})$, and

$$\begin{aligned} \ell &= \mathfrak{b}^\top \mathbf{S} \mathfrak{b}^\perp - \frac{1}{2} \mathfrak{b}^\top \mathbf{S} \mathfrak{b} + \sum_{j \notin \mathcal{L}} \left\{ (\mathfrak{q}^{(j)})^\top \mathcal{V}^{(j)} (\mathfrak{q}^{(j)})^\perp - \frac{1}{2} (\mathfrak{q}^{(j)})^\top \mathcal{V}^{(j)} \mathfrak{q}^{(j)} \right\} + \boldsymbol{\Theta}^\top \mathcal{W}^\perp - \frac{1}{2} \boldsymbol{\Theta} \mathcal{I} \boldsymbol{\Theta} \\ &\quad + \mathfrak{b}^\top \mathcal{U}^\perp - \frac{1}{2} \mathfrak{b}^\top \mathbf{M} \mathfrak{b} + \sum_{j \notin \mathcal{L}} \left\{ (\mathfrak{q}^{(j)})^\top (\mathcal{E}^{(j)})^\perp - \frac{1}{2} (\mathfrak{q}^{(j)})^\top \mathbf{E}^{(j)} \mathfrak{q}^{(j)} \right\} \\ &\quad - \mathfrak{b}^\top E\left(\mathcal{U} \mathcal{W}^\top\right) \boldsymbol{\Theta} - (\mathfrak{q}^{(j)})^\top \sum_{j \notin \mathcal{L}} E\left(\mathcal{E}^{(j)} \mathcal{W}^\top\right) \boldsymbol{\Theta} - (\mathfrak{q}^{(j)})^\top \sum_{j \notin \mathcal{L}} E\left(\mathcal{E}^{(j)} \mathcal{U}^\top\right) \mathfrak{b}. \end{aligned} \tag{39}$$

Furthermore, $E(\mathcal{U}\mathfrak{b}^\top) = E(\mathcal{E}^{(j)}\mathfrak{b}^\top) = 0$ and $E(\mathcal{U}\mathbb{A}^\top) = E(\mathcal{E}^{(j)}\mathbb{A}^\top)0$, so $E(e^\ell) = 1$.

By construction, it is obvious that $(\mathfrak{b}_n, \boldsymbol{\Theta}_n, \mathbb{A}_n, \mathfrak{b}_n^\star, \boldsymbol{\Theta}_n^\star, \mathbb{A}_n^\star)$ converges in law to $(\mathfrak{b}, \boldsymbol{\Theta}, \mathbb{A}, \mathfrak{b}^\perp, \boldsymbol{\Theta}^\perp, \mathbb{A}^\perp)$, where $(\mathfrak{b}^\perp, \boldsymbol{\Theta}^\perp, \mathbb{A}^\perp)$ is an independent copy of $(\mathfrak{b}, \boldsymbol{\Theta}, \mathbb{A})$. Next, it follows from the results of Sections 2.1–2.4 and the proofs in Genest and Rémillard (2008) that

$$
\begin{aligned}
\ell_n &= \mathfrak{b}_n^\top \mathbf{S}_n^\star \mathfrak{b}_n^\star - \frac{1}{2}\mathfrak{b}_n^\top \mathbf{S}_n^\star \mathfrak{b}_n + \sum_{j\notin\mathcal{L}}\left[\{\mathfrak{q}_n^{(j)}\}^\top\{\mathcal{V}_n^{(j)}\}^\star\{\mathfrak{q}_n^{(j)}\}^\star - \frac{1}{2}\{\mathfrak{q}_n^{(j)}\}^\top\{\mathcal{V}_n^{(j)}\}^\star\{\mathfrak{q}_n^{(j)}\}\right] \\
&\quad + \boldsymbol{\Theta}_n^\top \mathcal{W}_n^\star - \frac{1}{2}\boldsymbol{\Theta}_n \mathcal{I}_n^\star \boldsymbol{\Theta}_n + \mathfrak{b}_n^\top \mathcal{U}_n^\star - \frac{1}{2}\mathfrak{b}_n^\top M_n^\star \mathfrak{b}_n \\
&\quad + \sum_{j\notin\mathcal{L}}\left[\{\mathfrak{q}_n^{(j)}\}^\top\{\mathcal{E}_n^{(j)}\}^\star - \frac{1}{2}\{\mathfrak{q}_n^{(j)}\}^\top\{\mathbf{E}_n^{(j)}\}^\star\mathfrak{q}_n^{(j)}\right] - \mathfrak{b}_n^\top E\left\{\mathcal{U}_n^\star(\mathcal{W}_n^\star)^\top\right\}\boldsymbol{\Theta}_n \\
&\quad - \sum_{j\notin\mathcal{L}}(\mathfrak{q}_n^{(j)})^\top E\left\{(\mathcal{E}_n^{(j)})^\star(\mathcal{W}_n^\star)^\top\right\}\boldsymbol{\Theta}_n - \sum_{j\notin\mathcal{L}}(\mathfrak{q}_n^{(j)})^\top E\left\{(\mathcal{E}_n^{(j)})^\star(\mathcal{U}_n^\star)^\top\right\}\mathfrak{b}_n + o_P(1).
\end{aligned}
$$

Hence, ℓ_n converges in law to ℓ given by (39). It is easy to check that $E(\mathcal{U}\mathfrak{b}^\top) = E(\mathcal{E}^{(j)}\mathfrak{b}^\top) = 0$ and $E(\mathcal{U}\mathbb{A}^\top) = E(\mathcal{E}^{(j)}\mathbb{A}^\top)0$, so $E(e^\ell) = 1$. This completes the proof of the lemma.

We can now complete the proof of Theorem 3. As in Genest and Rémillard (2008), the idea of the proof is to use LeCam's third Lemma (van der Vaart and Wellner, 1996). Using Theorem 2 and Lemma 1, we get

$$
(\mathfrak{b}_n, \boldsymbol{\Theta}_n, \mathcal{W}_n, \mathbb{A}_n, \mathcal{U}_n, \mathcal{E}_n, \mathfrak{b}_n^\star, \boldsymbol{\Theta}_n^\star, \mathcal{W}_n^\star, \mathbb{A}_n^\star, \mathcal{U}_n^\star, \mathcal{E}_n^\star) \rightsquigarrow \left(\mathfrak{b}, \boldsymbol{\Theta}, \mathcal{W}, \mathbb{A}, \mathcal{U}, \mathcal{E}, \mathfrak{b}^\perp, \boldsymbol{\Theta}^\perp, \mathcal{W}^\perp, \mathbb{A}^\perp, \mathcal{U}^\perp, \mathcal{E}^\perp\right),
$$

where $\left(\mathfrak{b}^\perp, \boldsymbol{\Theta}^\perp, \mathcal{W}^\perp, \mathbb{A}^\perp, \mathcal{U}^\perp, \mathcal{E}^\perp\right)$ is an independent copy of $(\mathfrak{b}, \boldsymbol{\Theta}, \mathcal{W}, \mathbb{A}, \mathcal{U}, \mathcal{E})$. Since ℓ_n converges in law to ℓ given by (39), and $(U_1^\star, \mathbf{V}_1^\star), \ldots, (U_n^\star, \mathbf{V}_n^\star)$ have joint distribution $\mathcal{P}_n^\star$, it follows from Lecam's third lemma that for any bounded continuous function Ξ of the variables $(\mathfrak{b}_n, \boldsymbol{\Theta}_n, \mathcal{W}_n, \mathbb{A}_n, \mathcal{U}_n, \mathcal{E}_n, \mathfrak{b}_n^\star, \boldsymbol{\Theta}_n^\star, \mathcal{W}_n^\star, \mathbb{A}_n^\star, \mathcal{U}_n^\star, \mathcal{E}_n^\star)$, $E\left\{\Xi\left(\mathfrak{b}_n, \boldsymbol{\Theta}_n, \mathcal{W}_n, \mathbb{A}_n, \mathcal{U}_n, \mathcal{E}_n, \mathfrak{b}_n^\star, \boldsymbol{\Theta}_n^\star, \mathcal{W}_n^\star, \mathbb{A}_n^\star, \mathcal{U}_n^\star, \mathcal{E}_n^\star\right)\right\}$ converges as $n \to \infty$ to

$$
E\left\{\Xi\left(\mathfrak{b}, \boldsymbol{\Theta}, \mathcal{W}, \mathbb{A}, \mathcal{U}, \mathcal{E}, \mathfrak{b}^\star, \boldsymbol{\Theta}^\star, \mathcal{W}^\star, \mathbb{A}^\star, \mathcal{U}^\star, \mathcal{E}^\star\right)\right\} = E\left\{e^\ell \Xi\left(\mathfrak{b}, \boldsymbol{\Theta}, \mathcal{W}, \mathbb{A}, \mathcal{U}, \mathcal{E}, \mathfrak{b}^\perp, \boldsymbol{\Theta}^\perp, \mathcal{W}^\perp, \mathbb{A}^\perp, \mathcal{U}^\perp, \mathcal{E}^\perp\right)\right\}.
$$

Now, using representation (39) and the fact that the joint distributions are Gaussian, one obtains easily that $\mathfrak{b}^\star = \mathfrak{b} + \mathfrak{b}^\perp$, for any $j \notin \mathcal{L}$, $(\mathbf{q}^{(j)})^\star = \mathbf{q}^{(j)} + (\mathbf{q}^{(j)})^\perp$. Next, take $j \in \mathcal{L}$. It follows that if $\mathbb{B}_{nj}^\star(v_j) = \frac{1}{n+1}\sum_{t=1}\mathbb{I}(V_{tj} \le v_j)$, $v_j \in [0,1]$, then $\mathbb{B}_{nj}^\star \rightsquigarrow \mathbb{B}_j^\star = \mathbb{B}_j^\perp$. As a result, $n^{1/2}(v_{nj} - v_j) = n^{1/2}\left\{F_{n,t}^{(j)}(x_j) - F_t(x_j)\right\} = \mathbb{B}_{nj}(v_j)$, so

$$
n^{1/2}\left\{(\mathbf{F}_{n,t}^{(j)})^\star(\mathbf{x}_j) - v_j\right\} = \mathbb{B}_{nj}^\star(v_{nj}) + \mathbb{B}_{nj}(v_j) \rightsquigarrow \mathbb{B}_j(v_j) + \mathbb{B}_j^\perp(v_j).
$$

Furthermore, $\boldsymbol{\Theta}^\star = \boldsymbol{\Theta} + \boldsymbol{\Theta}^\top$. Combining these results, one gets that $\mathbb{A}_n^\star \rightsquigarrow \mathbb{A} + \mathbb{A}^\perp$. To complete the proof, set $\mathbf{v} = \mathbf{F}_t(\mathbf{x})$. It follows from the Delta Method, Theorem 2, and the previous calculations that if $u = G_{\boldsymbol{\beta}(t)}(y)$, then, using (19), one obtains

$$
\begin{aligned}
n^{1/2}\left\{\mathcal{K}_{n,t}^\star(y,\mathbf{x}) - \mathcal{K}_t(y,\mathbf{x})\right\} &= \dot{\mathcal{D}}_t(u,v)^\top \mathbf{h}_t \boldsymbol{\Theta}_n^\star + \nabla_{\mathbf{v}}\mathcal{D}_t(u,\mathbf{v})\mathcal{A}_t(\mathbf{x})\mathbb{A}_n^\star(\mathbf{x}) \\
&\quad + \mathfrak{d}_t\{u,\mathbf{v}\}\nabla_{\boldsymbol{\beta}} G_{\boldsymbol{\beta}(t)}(y)^\top \mathbf{B}_t \mathfrak{b}_n^\star + o_P(1) \\
&\rightsquigarrow \dot{\mathcal{D}}_t(u,v)^\top \mathbf{h}_t\left(\boldsymbol{\Theta}^\perp + \boldsymbol{\Theta}\right) + \nabla_{\mathbf{v}}\mathcal{D}_t(u,\mathbf{v})\mathcal{A}_t(\mathbf{x})\left\{\mathbb{A}^\perp(\mathbf{x}) + \mathbb{A}(\mathbf{x})\right\} \\
&\quad + \mathfrak{d}_t(u,\mathbf{v})\nabla_{\boldsymbol{\beta}} G_{\boldsymbol{\beta}(t)}(y)^\top \mathbf{B}_t\left(\mathfrak{b}^\perp + \mathfrak{b}\right) \\
&= \mathbb{K}_t^\perp(y,\mathbf{x}) + \mathbb{K}_t(y,\mathbf{x}),
\end{aligned}
$$

where $\mathbb{K}_t^\perp$ is an independent copy of $\mathbb{K}_t$. As a result, $\mathbb{K}_{n,t}^{(1)}, \ldots, \mathbb{K}_{n,t}^{(N)}$ converge to independent copies of $\mathbb{K}_t$. Similarly, using (18), one gets that $\mathbb{H}_{n,t}^{(1)}, \ldots, \mathbb{H}_{n,t}^{(N)}$ converge to independent copies of $\mathbb{H}_t$. □

A.4 Proof of Corollary 1

It follows from Theorem 2 that

$$
\mathbb{K}_t(y,\mathbf{x}) = \boldsymbol{\Theta}^\top \mathbf{h}_t^\top \dot{\mathcal{D}}_t\{G_{\boldsymbol{\beta}(t)}(y), \mathbf{v}\} + \mathbb{G}_t(y)\mathfrak{d}_{\boldsymbol{\phi}(t)}\{G_{\boldsymbol{\beta}(t)}(y), \mathbf{v}\} + \mathbb{F}_t(\mathbf{x})^\top \nabla_{\mathbf{v}}\mathcal{D}_{\boldsymbol{\phi}(t)}\{G_{\boldsymbol{\beta}(t)}(y), \mathbf{v}\}.
$$

The rest of the proof is an adaptation of Ghoudi and Rémillard (1998, Theorem 2.1) with $r \equiv 1$. The only new assumptions are the ones guaranteeing the convergence of $\mu_n(\mathbb{K})(u) = \frac{1}{n}\sum_{t=1}^n \mathbb{K}_t\{\mathcal{Q}_t(u,\mathbf{X}_t), \mathbf{X}_t)\}$ to $\mu(\mathbb{K})(u)$. □

References

Ahn, K.-H. and Palmer, R. N. (2016). Use of a nonstationary copula to predict future bivariate low flow frequency in the Connecticut river basin. Hydrological Processes, 30(19):3518–3532.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6):716–723.

Billingsley, P. (1995). Probability and Measure. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition. A Wiley-Interscience Publication.

De Boor, C. (2001). A practical guide to spline. Springer Series in Statistics.

Dette, H., Van Hecke, R., and Volgushev, S. (2014). Some comments on copula-based regression. Journal of the American Statistical Association, 109(507):1319–1324.

Fermanian, J.-D. and Lopez, O. (2015). Single-index copulae. ArXiv e-prints.

Genest, C., Ghoudi, K., and Rivest, L. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika, 82:543–52.

Genest, C., Nešlehová, J. G., and Rémillard, B. (2017). Asymptotic behavior of the empirical multilinear copula process under broad conditions. Journal of Multivariate Analysis, 159:82–110.

Genest, C. and Rémillard, B. (2004). Tests of independence or randomness based on the empirical copula process. Test, 13:335–369.

Genest, C. and Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. Annales de l'Institut Henri Poincare. Probabilites et Statistiques, 44:337–366.

Genest, C., Rémillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. Insurance: Mathematics and Economics, 44:199–213.

Ghoudi, K. and Rémillard, B. (1998). Empirical processes based on pseudo-observations. In Asymptotic Methods in Probability and Statistics (Ottawa, ON, 1997), pages 171–197. North-Holland, Amsterdam.

Ghoudi, K. and Rémillard, B. (2013). Comparison of specification tests for GARCH models. Computational Statistics & Data Analysis, 76:291–300.

Ghoudi, K. and Rémillard, B. (2018). Serial independence tests for innovations of conditional mean and variance models. TEST, 27(1):3–26.

Hastie, T. J. and Tibshirani, R. J. (1990). Generalized additive models, volume 43. CRC press.

Holmes, M., Kojadinovic, I., and Quessy, J.-F. (2013). Nonparametric tests for change-point detection à la gombay and horváth. Journal of Multivariate Analysis, 115:16–32.

Jiang, C., Xiong, L., Xu, C.-Y., and Guo, S. (2015). Bivariate frequency analysis of nonstationary low-flow series based on the time-varying copula. Hydrological Processes, 29(6):1521–1534.

Joe, H. and Xu, J. (1996). The estimation method of inference functions for margins for multivariate models. Technical Report 166, Department of Statistics, University of British Columbia.

Knudsen, M. F., Seidenkrantz, M.-S., Jacobsen, B. H., and Kuijpers, A. (2011). Tracking the atlantic multidecadal oscillation through the last 8,000 years. Nature communications, 2:178.

Kraus, D. and Czado, C. (2017). D-vine copula based quantile regression. Computational Statistics & Data Analysis, 110:1–18.

Nasri, B. (2017). Méthodes d'estimation des quantiles conditionnels en hydro-climatologie. PhD thesis, INRS-ETE.

Nasri, B., Bouezmarni, T., St-Hilaire, A., and Ouarda, T. (2017). Non-stationary hydrologic frequency analysis using b-spline quantile regression. Journal of Hydrology, 554:532–544.

Nasri, B. and Rémillard, B. (2017). Copula-based dynamic models for multivariate time series. Technical report, GERAD.

Noh, H., Ghouch, A. E., and Bouezmarni, T. (2013). Copula-based regression estimation and inference. Journal of the American Statistical Association, 108(502):676–688.

Noh, H., Ghouch, A. E., and Keilegom, I. V. (2015). Semiparametric conditional quantile estimation through copula-based multivariate models. Journal of Business & Economic Statistics, 33(2):167–178.

Quessy, J.-F., Rivest, L.-P., and Toupin, M.-H. (2016). On the family of multivariate chi-square copulas. Journal of Multivariate Analysis, 152:40–60.

Rémillard, B. (2012). Specification tests for dynamic models using multipliers. Technical report, SSRN Working Paper Series No. 2028558.

Rémillard, B. (2013). Statistical Methods for Financial Engineering. Chapman and Hall/CRC Financial Mathematics Series. Taylor & Francis.

Rémillard, B., Hocquard, A., Lamarre, H., and Papageorgiou, N. A. (2017). Option pricing and hedging for discrete time regime-switching model. Modern Economy, 8:1005–1032.

Rémillard, B., Nasri, B., and Ben-Abdellatif, M. (2017). Replication methods for financial indexes. Submitted.

Rémillard, B., Nasri, B., and Bouezmarni, T. (2017). On copula-based conditional quantile estimators. Statistics & Probability Letters, 128:14–20.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. The Annals of Mathematical Statistics, 23:470–472.

Serfling, R. J. (1980). Approximation Theorems of Mathematical Statistics. John Wiley & Sons Inc., New York. Wiley Series in Probability and Mathematical Statistics.

Shih, J. and Louis, T. A. (1995). Inference on association parameter in copula models for bivariate survival data. Biometrics, 26:183–214.

Sklar, A. (1959). Fonctions de répartition à n dimentions et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris, 8:229–231.

Stone, C. J. and Koo, C.-Y. (1986). Additive splines in statistics. In Proceedings of the statistical computing section, volume 45, page 48.

van der Vaart, A. W. and Wellner, J. A. (1996). Weak convergence and empirical processes. Springer Series in Statistics. Springer-Verlag, New York.

Xu, J. J. (1996). Statistical Modelling and Inference for Multivariate and Longitudinal Discrete Response Data. PhD thesis, University of British Columbia.