

Flexible and Dynamic Modeling of Dependencies via Copulas

Irène Gijbels, Klaus Herrmann, and Dominik Sznajder

Abstract In this chapter we first review recent developments in the use of copulas for studying dependence structures between variables. We discuss and illustrate the concepts of unconditional and conditional copulas and association measures, in a bivariate setting. Statistical inference for conditional and unconditional copulas is discussed, in various modeling settings. Modeling the dynamics in a dependence structure between time series is of particular interest. For this we present a semiparametric approach using local polynomial approximation for the dynamic time parameter function. Throughout the chapter we provide some illustrative examples. The use of the proposed dynamical modeling approach is demonstrated in the analysis and forecast of wind speed data.

1 Introduction

When the aim is to model the dependence structure between d random variables, denoted by Y_1, \dots, Y_d , we can distinguish between several approaches. In a regression approach, one is interested in how a variable of primary interest, say Y_d , and called the response variable, is influenced on average by Y_1, \dots, Y_{d-1} , the covariates. A general regression model is of the form

$$Y_d = g(Y_1, \dots, Y_{d-1}) + \varepsilon,$$

where $g : \mathbb{R}^{d-1} \rightarrow \mathbb{R}$ is a $(d - 1)$ -dimensional function of the covariates, and where the error term ε has conditional mean $E(\varepsilon|Y_1, \dots, Y_{d-1})$ equal to zero. Consequently, the conditional mean function of Y_d given the covariates $Y_1 = y_1, \dots, Y_{d-1} = y_{d-1}$, with $(y_1, \dots, y_{d-1}) \in \mathbb{R}^{d-1}$ equals $E(Y_d|Y_1 = y_1, \dots, Y_{d-1} = y_{d-1}) = g(y_1, \dots, y_{d-1})$. For the mean regression

I. Gijbels (✉) • K. Herrmann • D. Sznajder

Department of Mathematics and Leuven Statistics Research Centre (LStat), KU Leuven, Celestijnenlaan 200B, Box 2400, B-3001 Leuven (Heverlee), Belgium

e-mail: Irene.Gijbels@wis.kuleuven.be; Klaus.Herrmann@wis.kuleuven.be;

Dominik.Sznajder@wis.kuleuven.be

© Springer International Publishing Switzerland 2015

A. Antoniadis et al. (eds.), *Modeling and Stochastic Learning for Forecasting in High Dimensions*, Lecture Notes in Statistics 217,

DOI 10.1007/978-3-319-18732-7_7

function g one can either assume some known parametric form, or leaving its form fully unspecified (and to be determined from the data), in respectively parametric and nonparametric regression. An alternative is a semiparametric modeling in which the influence of some covariates might be modeled parametrically, whereas the influence on average of other covariates on Y_d might be described via an unknown function (a nonparametric functional part). In a general approach, the dependencies between the various components in the random vector (Y_1, \dots, Y_d) are fully described by the joint distribution of (Y_1, \dots, Y_d) , i.e. by the d -variate cumulative distribution function of (Y_1, \dots, Y_d) denoted by $P\{Y_1 \leq y_1, \dots, Y_d \leq y_d\}$. From this one can calculate for example the conditional mean function $E(Y_d | Y_1 = y_1, \dots, Y_{d-1} = y_{d-1})$.

Denote the marginal cumulative distribution function of Y_j , by F_j , for $j = 1, \dots, d$. According to Sklar's Theorem [34] there exists a copula function C defined on $[0, 1]^d$ such that

$$P\{Y_1 \leq y_1, \dots, Y_d \leq y_d\} = C(F_1(y_1), \dots, F_d(y_d)) \quad \forall (y_1, \dots, y_d) \in \mathbb{R}^d.$$

In case the marginal distribution functions, F_1, \dots, F_d , are continuous, the copula function C is unique. See [28]. The copula function *couple*s the joint distribution function to its univariate margins F_1, \dots, F_d . The dependence structure between the components of (Y_1, \dots, Y_d) is fully characterized by the copula function C .

Note that a copula function is nothing but a joint distribution function on $[0, 1]^d$ with uniform margins. Based on a joint distribution function, we can study conditional distribution functions derived from it, as well as characteristics of these (e.g. moments, medians, ...). Translated into the copula context this leads to various concepts for describing dependence structures.

The aim of this chapter is to first provide a review of copula modeling concepts, and statistical inference for them in various settings (parametric, semiparametric and nonparametric). This is done in a setting of independent data in Sects. 2 and 3. For simplicity of presentation, we restrict throughout the chapter to the setting of bivariate copulas (the case $d = 2$). In Sect. 4 we turn to the dynamical modeling of the dependence between time series, extending existing approaches of local polynomial fitting to this setting. We conclude the chapter by an illustration of the use of the method in a practical forecasting application in Sect. 5.

2 Global Dependencies and Unconditional Copulas

2.1 Population Concepts

Consider two random variables Y_1 and Y_2 , with joint distribution function H , and continuous marginal distributions functions F_1 and F_2 respectively. There then exists

a (unique) bivariate copula function C , such that

$$H(y_1, y_2) = P\{Y_1 \leq y_1, Y_2 \leq y_2\} = C(F_1(y_1), F_2(y_2)) \quad (y_1, y_2) \in \mathbb{R}^2. \quad (1)$$

It is common practice to measure the strength of the relationship between Y_1 and Y_2 via a so-called association measure. There are various statistical association measures. Among the most well-known are: Pearson's correlation coefficient, Kendall's tau, Spearman's rho, Gini's coefficient, and Blomqvist's beta. See [4, 16, 17, 27] and [25], among others. Pearson's correlation coefficient, defined as $\text{Cov}(Y_1, Y_2) / \sqrt{\text{Var}(Y_1) \text{Var}(Y_2)}$, only exists if the second order moments of both margins Y_1 and Y_2 , exist, and equals ± 1 in case Y_2 is a (perfect) linear transform of Y_1 . Gini's coefficient and Blomqvist's beta, are often used in economical sciences (for example, as a measure of inequality of income or wealth). Several well-known association measures can be expressed as functionals of the copula function. Denote by (Y'_1, Y'_2) and (Y''_1, Y''_2) , two independent copies of (Y_1, Y_2) . For the following association measures we give their definitions followed by an expression in terms of the copula function (for some measures alternative expressions in terms of C exist).

- Kendall's tau:

$$\begin{aligned} \tau_{Y_1, Y_2} &= P\{(Y_1 - Y'_1)(Y_2 - Y'_2) > 0\} - P\{(Y_1 - Y'_1)(Y_2 - Y'_2) < 0\} \\ &= 4 \iint_{[0,1]^2} C(u_1, u_2) dC(u_1, u_2) - 1. \end{aligned} \quad (2)$$

- Spearman's rho:

$$\begin{aligned} \rho_{Y_1, Y_2} &= 3 [P\{(Y_1 - Y'_1)(Y_2 - Y''_2) > 0\} - P\{(Y_1 - Y'_1)(Y_2 - Y''_2) < 0\}] \\ &= 12 \iint_{[0,1]^2} C(u_1, u_2) du_1 du_2 - 3. \end{aligned}$$

- Gini's coefficient:

$$\begin{aligned} \gamma_{Y_1, Y_2} &= 2E(|F_1(Y_1) + F_2(Y_2) - 1| - |F_1(Y_1) - F_2(Y_2)|) \\ &= 2 \iint_{[0,1]^2} (|u_1 + u_2 - 1| - |u_1 - u_2|) dC(u_1, u_2). \end{aligned}$$

- Blomqvist's beta:

$$\beta_{Y_1, Y_2} = 2P\{(Y_1 - F_1^{-1}(0.5))(Y_2 - F_2^{-1}(0.5)) > 0\} - 1 = 4C\left(\frac{1}{2}, \frac{1}{2}\right) - 1,$$

where $F_j^{-1}(0.5)$ is the median of the margin F_j , for $j = 1, 2$.

See [28] for an overview of association measures, their basic properties, and their interrelationships.

When talking about unraveling the dependence structure between Y_1 and Y_2 , it is transparent from (1) that informations about the copula function C as well as about the two margins F_1 and F_2 are of importance. We illustrate the impact of these elements with some examples in Sect. 2.2.

2.2 Illustration: Examples

The copula function and the marginal distributions together determine the joint distribution function (see (1)), and consequently all of the population characteristics. This is reflected in, among others, the typical observed scatter plots and the different values for the association measures.

As an illustration we consider the following examples:

Example 1.

$$Y_1 \sim N(0, 4), \quad Y_2 \sim \text{Exp}(2)$$

$$C(u_1, u_2) = \left(\max(u_1^{-\theta} + u_2^{-\theta} - 1, 0) \right)^{-\frac{1}{\theta}}, \quad \theta = 1,$$

where the copula belongs to the Clayton copula family.

Example 2.

$$Y_1 \sim \text{Exp}(2), \quad Y_2 \sim \text{Beta}(1, 4)$$

and with the same copula as in Example 1.

Example 3.

$$Y_1 \sim \text{Student}(5, 3), \quad Y_2 \sim \text{Ex}(0.2)$$

$$C(u_1, u_2) = \exp \left(- \left((-\log u_1)^\theta + (-\log u_2)^\theta \right)^{\frac{1}{\theta}} \right), \quad \theta = 3,$$

where Student(ν, μ) is a noncentral Student distribution with ν degrees of freedom and noncentrality parameter μ , and where the copula belongs to the Gumbel copula family.

Example 4.

$$Y_1 \sim \text{Exp}(2), \quad Y_2 \sim \text{Beta}(1, 4)$$

and the same copula as in Example 3.

In Fig. 1 we present a typical sample $((Y_{11}, Y_{21}), \dots, (Y_{1n}, Y_{2n}))$ from (Y_1, Y_2) (with n the sample size) from the above models (left column of pictures) together with their pseudo-observations, defined as $(F_1(Y_{1i}), F_2(Y_{2i}))$, for $i = 1, \dots, n$, (right column of pictures). Due to the probability integral transformation, the pseudo-observations $F_j(Y_{ji})$ are uniformly distributed (for each $j = 1, 2$). For Examples 1 and 2 the marginal distribution functions are different, but the dependence structure (the copula) is the same. In the left panels of rows one and two of Fig. 1 we depict scatter plots based on random samples of sizes 700 and 400 from, respectively, Examples 1 and 2. The scatter plots look quite different for the samples from the two models, but notice the similarity between the plots for the pseudo-observations (right panels). We can observe a mild positive dependence everywhere with a higher concentration in the lower tails (the lower left corner of the plots). In Example 3, both the margins and the copula are different, and the scatter plots based on a sample of size $n = 700$ from that model looks very different from the previous examples. Here, there is clearly more positive dependence visible, and we also notice heavier right tail characteristics. When comparing the plots for samples from Examples 3 and 4 (rows three and four in Fig. 1) we can observe similarities in the scatter plots of the pseudo-observations (the right panels), due to the fact that these examples share the same underlying copula. Furthermore, looking at scatter plots of typical observations from Examples 2 and 4, we can see the impact of changing a copula while keeping the same marginal distributions.

In Table 1 we present the values of some association measures for the four examples. Note the equality of the measures for Examples 1 and 2 on the one hand, and for Examples 3 and 4, on the other hand, since these examples share the same copula (i.e. have the same dependence structure). The values of the association measures only depend on the underlying copula function and not on the marginal distributions.

2.3 Statistical Inference

Suppose now that $((Y_{11}, Y_{21}), \dots, (Y_{1n}, Y_{2n}))$ is a sample of size n of independent observations from (Y_1, Y_2) , and the interest is in estimating the copula function in (1). Once an estimator for the copula function is available, the way is open to obtain estimates for association measures that can be expressed as functionals of the copula, as those for example listed in Sect. 2.1.

According to available information on either the copula and/or the margins we distinguish between different situations in the modeling aspects. In the fully parametric setting the copula function is assumed to be known, up to some parameters, and the same for the distribution of the margins. Other settings are listed in Table 2. We briefly review statistical inference under each of these settings.

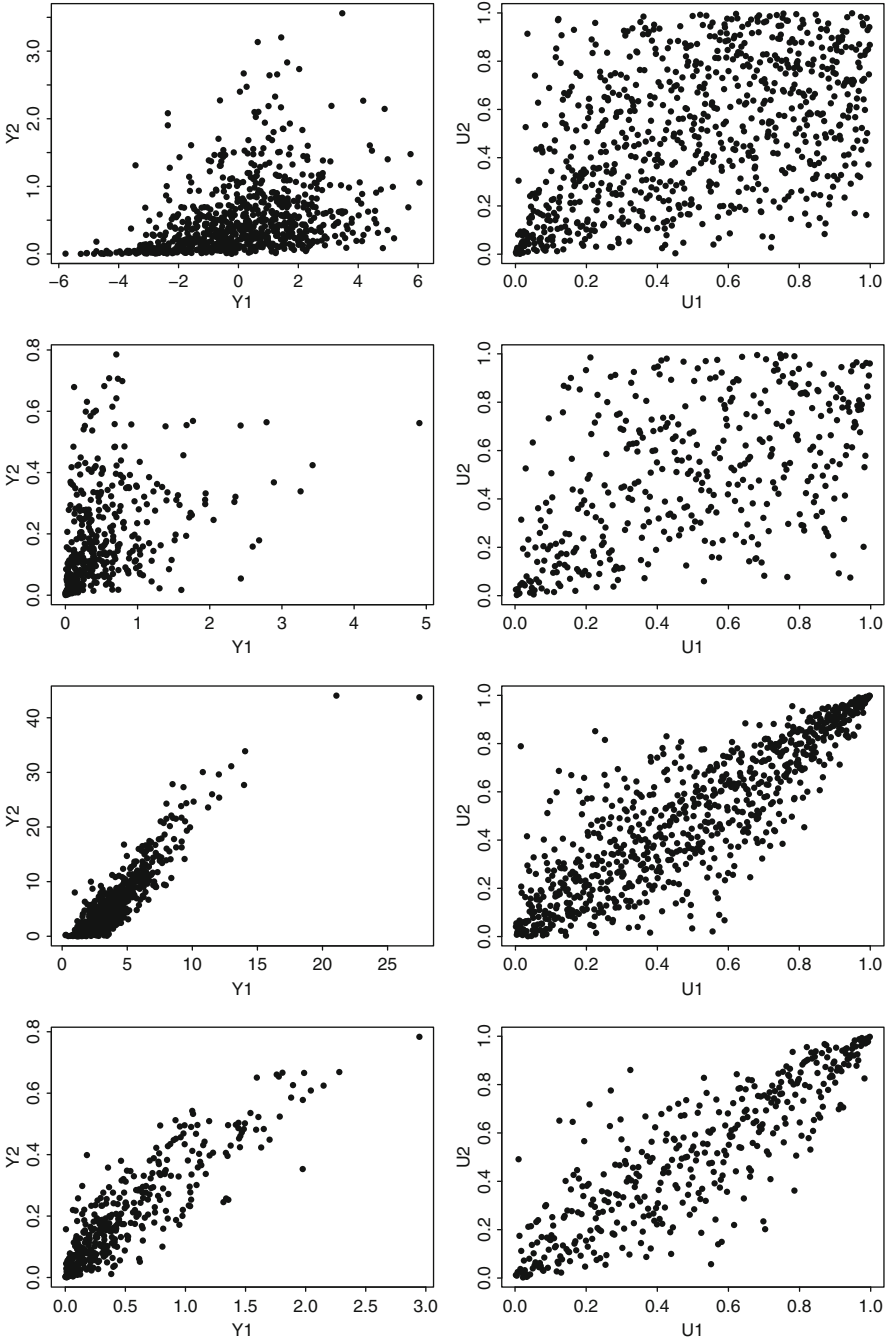


Fig. 1 Typical samples for Examples 1–4 (top row to bottom row); left columns: scatter plot of the observations, right columns: scatter plots of the pseudo-observations

Table 1 Association measures for Examples 1–4

Example	Kendall’s tau	Spearman’s rho	Gini’s index	Blomqvist’s beta
1 & 2	0.333	0.478	0.382	0.333
3 & 4	0.667	0.849	0.725	0.670

Table 2 Situations for statistical inference

Copula	Margins	Approach
Parametric	Parametric	Fully parametric
Parametric	Nonparametric	Semiparametric
Nonparametric	Parametric	Semiparametric
Nonparametric	Nonparametric	Fully nonparametric

2.3.1 Fully Parametric Approach

In a fully parametric approach one starts by assuming a specific parametric model for the copula function as well as for the margins. More formally, suppose that the copula $C(\cdot, \cdot) = C(\cdot, \cdot; \theta_C)$, and that $F_j(\cdot) = F(\cdot; \theta_j)$, for $j = 1, 2$, where θ_C, θ_j , for $j = 1, 2$ are the respective parameter vectors, taking values in parameter spaces Θ_C, Θ_1 and Θ_2 respectively. These parameters spaces can have nonempty intersections, in other words, the parameter vectors θ_C and θ_j can have common elements.

Assume for simplicity that the density of the copula function exists, i.e. the second order partial derivative of the copula function exists

$$c(u_1, u_2; \theta_C) = \frac{\partial^2 C(u_1, u_2; \theta_C)}{\partial u_1 \partial u_2} \quad \forall (u_1, u_2) \in [0, 1]^2 .$$

If in addition the corresponding densities f_j of F_j , for $j = 1, 2$, exist, then the joint density of (Y_1, Y_2) is given by (see (1))

$$h(y_1, y_2) = c(F_1(y_1), F_2(y_2)) f_1(y_1) f_2(y_2) \quad \forall (y_1, y_2) \in \mathbb{R}^2 .$$

Keeping this in mind, the logarithm of the likelihood function then equals

$$\ell_n(\theta_1, \theta_2, \theta_C) = \sum_{i=1}^n \log(c(F_1(Y_{1i}; \theta_1), F_2(Y_{2i}; \theta_2); \theta_C) f_1(Y_{1i}; \theta_1) f_2(Y_{2i}; \theta_2)) , \tag{3}$$

which needs to be maximized with respect to $(\theta_1, \theta_2, \theta_C)$. Denote this maximizer by $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_C)$.

An estimator for, for example, the associated Kendall’s tau is then obtained via (2)

$$\hat{\tau}_{Y_1, Y_2} = 4 \iint_{[0, 1]^2} C(u_1, u_2; \hat{\theta}_C) dC(u_1, u_2; \hat{\theta}_C) - 1 .$$

2.3.2 Semiparametric Approach

Suppose now that the margins F_1 and F_2 cannot be parametrized, but are fully unknown. For the copula function, on the contrary, we still believe that a parametric model $C(\cdot, \cdot; \theta_C)$ is a reasonable assumption. In this case, we thus need to estimate the margins from the available data. The log-likelihood in (3), where the margins F_1 and F_2 are unknown, is replaced by the pseudo log-likelihood

$$\ell_n(\theta_C) = \sum_{i=1}^n \log(c(F_{1n}(Y_{1i}), F_{2n}(Y_{2i}); \theta_C)) \quad (4)$$

where the unknown margins F_1 and F_2 are replaced by the estimates

$$F_{1n}(y_1) = \frac{1}{n+1} \sum_{i=1}^n I\{Y_{1i} \leq y_1\} \quad F_{2n}(y_2) = \frac{1}{n+1} \sum_{i=1}^n I\{Y_{2i} \leq y_2\} ,$$

with $I\{y \in A\}$ the indicator function on a set A , i.e. $I\{y \in A\} = 1$, if $y \in A$ and $I\{y \in A\} = 0$, if $y \notin A$. In the empirical estimates, it is recommended to use the modified factor $(n+1)^{-1}$ instead of the usual factor n^{-1} , because by using $(n+1)^{-1}$ the values $F_{jn}(Y_{ji})$ are in the set $\{\frac{1}{n+1}, \dots, \frac{n}{n+1}\}$ instead of in the set $\{\frac{1}{n}, \dots, \frac{n-1}{n}, 1\}$, and hence by using this modified factor, one stays away from both boundary points, 0 as well as 1. Note that this modification has no effect on the (asymptotic) properties of the resulting estimates. See, for example, [15].

The pseudo log-likelihood estimate $\hat{\theta}_C$ of θ_C is then obtained by maximizing the pseudo log-likelihood in (4) with respect to θ_C .

See [33] for a study on semiparametric efficient estimation in case of Gaussian copulas with unknown margins.

2.3.3 Fully Nonparametric Approach

We now turn to the fully nonparametric approach where one can neither for C nor for the margins (F_1 and F_2) propose an appropriate parametric model. Hence C , F_1 and F_2 are fully unknown.

Nonparametric estimation of a copula goes back to the early seventies. In the paper [10] the empirical copula estimator was introduced and studied. The basic idea behind this estimator is very simple. As can be seen from (1), $C(\cdot, \cdot)$ is in fact nothing else but the joint cumulative distribution function after the margins have been transformed, via the probability integral transformation

$$U_1 = F_1(Y_1) \quad \text{and} \quad U_2 = F_2(Y_2) .$$

In other words, $C(\cdot, \cdot)$ is the joint cumulative distribution function of (U_1, U_2) . If independent observations $((U_{11}, U_{21}), \dots, (U_{1n}, U_{2n}))$ from (U_1, U_2) would be

available, then the cumulative distribution function could be estimated via the usual bivariate empirical distribution function $\frac{1}{n} \sum_{i=1}^n I\{U_{1i} \leq u_1, U_{2i} \leq u_2\}$. Since we have no observations from $F_1(Y_1)$ and $F_2(Y_2)$ we simply replace the unobserved $U_{ji} = F_j(Y_{ji})$ by a pseudo-observation, its ‘slightly modified’ rank $F_{nj}(Y_{ji})$ in the original sample, and obtain the empirical copula estimator

$$C_n(u_1, u_2) = \frac{1}{n} \sum_{i=1}^n I\{\tilde{U}_{1i} \leq u_1, \tilde{U}_{2i} \leq u_2\} \quad \tilde{U}_{1i} = F_{1n}(Y_{1i}) \quad \tilde{U}_{2i} = F_{2n}(Y_{2i}) . \tag{5}$$

This estimator was also studied further in [14] and [32], among others.

Obviously, the estimator in (5) is a step function, which might not be very desirable, when the function $C(\cdot, \cdot)$ is continuous or even differentiable.

Nonparametric estimation methods that lead to smooth estimators for the copula C have been derived. Among these are kernel estimators. See, for example, [6, 18] and [29]. Kernel estimators of the copula C are essentially obtained by replacing the non-smooth indicator function $I\{\tilde{U}_{1i} \leq u_1, \tilde{U}_{2i} \leq u_2\} = I\{\tilde{U}_{1i} \leq u_1\} I\{\tilde{U}_{2i} \leq u_2\}$ in (5) by a smooth kernel function. The non-smooth function $I\{\tilde{U}_{ji} \leq u_j\}$ is replaced by the smooth function $K\left(\frac{u_j - \tilde{U}_{ji}}{h_n}\right)$, where $K(y) = \int_{-1}^y k(t)dt$ is the kernel distribution function associated with the kernel k , a symmetric density function, with support the interval $[-1, 1]$, and $h_n > 0$ is a bandwidth parameter. The bandwidth parameter determines the size of the neighbourhood in which the jump in the indicator function is ‘smoothed out’. An example of a kernel function k is the Epanechnikov kernel $k(x) = \frac{3}{4}(1 - x^2)I\{|x| \leq 1\}$. As an illustration we depict in Fig. 2 the indicator function $I\{0.6 \leq u\}$ as well as a smooth version of it, namely $K\left(\frac{u-0.6}{h_n}\right)$, with K based on the Epanechnikov kernel, for two different values of the bandwidth h_n . The larger the bandwidth, the larger the neighbourhood over which the jump is ‘smeared out’.

Such a simple replacement of the indicator part by a smooth part, leads to the kernel estimator

$$\frac{1}{n} \sum_{i=1}^n K\left(\frac{u_1 - \tilde{U}_{1i}}{h_n}\right) K\left(\frac{u_2 - \tilde{U}_{2i}}{h_n}\right) .$$

Since the copula C is defined on the interval $[0, 1]^2$ (a compact support) special attention however is needed to obtain a kernel estimator that shows the same nice asymptotic properties at the boundaries of $[0, 1]^2$ as in the interior of that support. The aim is to obtain kernel estimators for which the convergence rate is the same on the interior of the square $[0, 1]^2$ as well as on the edges of it. This can be done for example, by using a reflection type of method, which consists of reflecting each pseudo-observation $(\tilde{U}_{1i}, \tilde{U}_{2i})$ with respect to all four corner points, and all four edges of the interval $[0, 1]^2$, resulting into an augmented data set of size $9n$, from

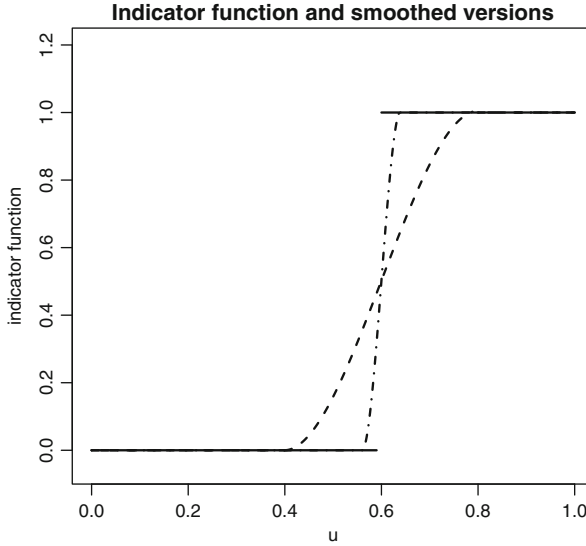


Fig. 2 The indicator function $I\{0.6 \leq u\}$ and its smooth version $K\left(\frac{u-0.6}{h_n}\right)$, using the Epanechnikov kernel, and bandwidths $h_n = 0.04$ (dashed-dotted curve) and $h_n = 0.20$ (dashed curve)

which the kernel estimator is defined. See Fig. 3 for an illustration of a data point and the eight points resulting from reflections of the given point with respect to all corners and edges of the unit square $[0, 1]^2$.

This leads to the kernel Mirror-Reflection type estimator introduced and studied in [18]:

$$\hat{C}_n^{\text{MR}}(u_1, u_2) = \frac{1}{n} \sum_{i=1}^n \sum_{\ell=1}^9 \left[K\left(\frac{u_1 - \tilde{U}_{1i}^{(\ell)}}{h_n}\right) - K\left(\frac{-\tilde{U}_{1i}^{(\ell)}}{h_n}\right) \right] \times \left[K\left(\frac{u_2 - \tilde{U}_{2i}^{(\ell)}}{h_n}\right) - K\left(\frac{-\tilde{U}_{2i}^{(\ell)}}{h_n}\right) \right],$$

where

$$\begin{aligned} & \{(\tilde{U}_{1i}^{(\ell)}, \tilde{U}_{2i}^{(\ell)}), i = 1, \dots, n, \ell = 1, \dots, 9\} \\ & = \{(\pm\tilde{U}_{1i}, \pm\tilde{U}_{2i}), (\pm\tilde{U}_{1i}, 2 - \tilde{U}_{2i}), (2 - \tilde{U}_{1i}, \pm\tilde{U}_{2i}), (2 - \tilde{U}_{1i}, 2 - \tilde{U}_{2i}), i = 1, \dots, n\}. \end{aligned}$$

and K is the integral of the considered kernel k , as mentioned above.

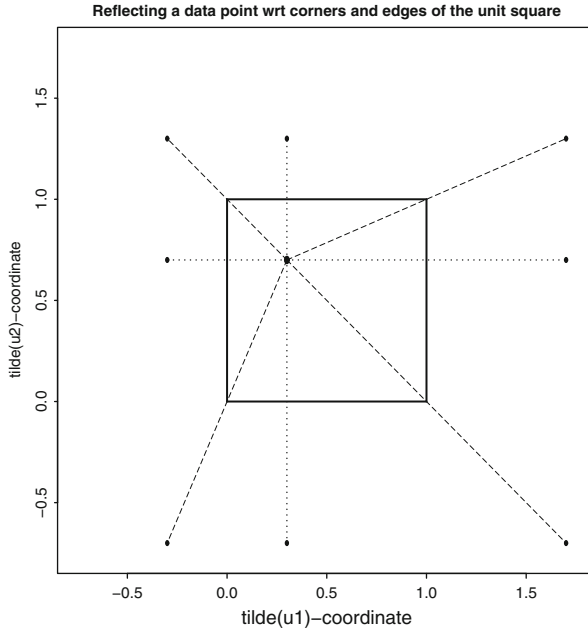


Fig. 3 A data point (indicated by a “•” in the unit square), and the reflected points (indicated by a “•”) with respect to all corners (see the *dashed lines*) and edges (see the *dotted lines*) of the unit square

An alternative approach to deal with the boundary issue is by using local linear fitting, and its implicit boundary kernel. This was done by [6], and resulted in the kernel Local Linear estimator:

$$\hat{C}_n^{LL}(u_1, u_2) = \frac{1}{n} \sum_{i=1}^n K_{u_1, h_n} \left(\frac{u_1 - \tilde{U}_{1i}}{h_n} \right) K_{u_2, h_n} \left(\frac{u_2 - \tilde{U}_{2i}}{h_n} \right),$$

where K_{u, h_n} is the integral of the modified boundary kernel

$$k_{u, h}(x) = \frac{k(x) (a_2(u, h) - a_1(u, h)x)}{a_0(u, h)a_2(u, h) - a_1^2(u, h)} I_{\left\{ \frac{u-1}{h} < x < \frac{u}{h} \right\}},$$

where

$$a_\ell(u, h) = \int_{\frac{u-1}{h}}^{\frac{u}{h}} t^\ell k(t) dt \quad \text{for } \ell = 0, 1, 2 .$$

Note that nonparametric methods such as the above kernel methods, involve the choice of a bandwidth parameter. This issue is not discussed here. See for example [29] (Section 3.2 in that paper), and references therein, for some discussion on

bandwidth choice. One could also use other smoothing methods than the kernel method. We do not discuss these.

3 Local Dependencies and Conditional Copulas

3.1 Population Concepts

Suppose now that the interest is in the relationship between the random variables Y_1 and Y_2 but that both random variables are possibly influenced by another random variable, say X . A first interest is then in studying the conditional dependence between Y_1 and Y_2 given a specific value for X , say $X = x$. Instead of simply looking at the joint distribution between Y_1 and Y_2 , as in Sect. 2, we now focus on the joint distribution of (Y_1, Y_2) *conditionally* upon $X = x$:

$$H_x(y_1, y_2) = P\{Y_1 \leq y_1, Y_2 \leq y_2 | X = x\}.$$

Applying Sklar's theorem to this conditional joint distribution function results into

$$H_x(y_1, y_2) = C_x(F_{1x}(y_1), F_{2x}(y_2)) \quad (y_1, y_2) \in \mathbb{R}^2, \quad (6)$$

where

$$F_{1x}(y_1) = P\{Y_1 \leq y_1 | X = x\} \quad \text{and} \quad F_{2x}(y_2) = P\{Y_2 \leq y_2 | X = x\},$$

denote the marginal cumulative distributions functions of Y_1 and Y_2 , respectively, *conditionally* upon $X = x$. The main difference between (1) and (6) is that the copula function C_x changes with the fixed value of X ($X = x$), as well as the margins F_{jx} (for $j = 1, 2$). We refer to C_x as the *conditional copula* function. This notion was first considered by [30] in the specific context of modeling the dynamics of exchange rates, where the conditioning variable is related to time. See also Sect. 4.

Analogously to the case of the (unconditional) copula C in Sect. 2, the strength of the dependence relationship between Y_1 and Y_2 , but now *conditionally* upon the given value of $X = x$, can be measured using an association measure. For simplicity of presentation, we just focus on the Kendall's tau association measure. Denote by (Y'_1, Y'_2, X') an independent copy of (Y_1, Y_2, X) . Then the *conditional* Kendall's tau function is defined as

$$\begin{aligned} \tau(x) &= P\{(Y_1 - Y'_1)(Y_2 - Y'_2) > 0 | X = X' = x\} \\ &\quad - P\{(Y_1 - Y'_1)(Y_2 - Y'_2) < 0 | X = X' = x\} \\ &= 4 \iint_{[0,1]^2} C_x(u_1, u_2) dC_x(u_1, u_2) - 1. \end{aligned} \quad (7)$$

For not making the notation too involved, we dropped the superscript $\{Y_1, Y_2\}$ to indicate that we are interested in the dependence structure between Y_1 and Y_2

(conditionally upon $X = x$). See for example [20], for illustrations and examples of use of conditional association measures.

Note from (6) that the conditional dependence of Y_1 and Y_2 (conditionally given $X = x$) may be different for different values taken by X , i.e. the dependence structure, as well as its strength, may change with the value taken by the third variable X . This possible change in the strength of the relationship is reflected in the fact that Kendall's tau is now a function of x .

It is often noticed that in applications, one uses the following simplification: the dependence structure itself, captured by the copula function, does not change with the specific value that X takes, and the dependence on x only comes in via the conditional margins, i.e.

$$H_x(y_1, y_2) = C(F_{1x}(y_1), F_{2x}(y_2)) \quad (y_1, y_2) \in \mathbb{R}^2.$$

In, for example, the literature on C-vine and D-vine copulas this assumption is inherently present. See [24] and [3], among others. See also the chapter (and its discussion) by [36], in which the dependence structure is assumed to stay constant in time.

3.2 Illustration: Examples

We now illustrate the concepts of Sect. 3.1 with some examples. A first example, Example 5, is an extension of Examples 1 and 2 of Sect. 2.2: instead of a bivariate Clayton copula we start from a three-variate Frank copula. In a second example, Example 6, we modify Example 3 of Sect. 2.2 by allowing the parameter of the copula and the marginal distribution of the second component to change with the random variable X . More precisely, the examples are as follows.

Example 5.

$$Y_1 \sim N(0, 4), \quad Y_2 \sim \text{Exp}(2), \quad X \sim \text{Beta}(1, 4)$$

$$C(u_1, u_2, u_3) = -\frac{1}{\theta} \log \left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)(e^{-\theta u_3} - 1)}{(e^{-\theta} - 1)^2} \right), \quad \theta = 3.$$

Example 6.

$$Y_1 \sim \text{Student}(5, 3), \quad Y_2|X \sim \text{Exp}(0.2(10X + 1)), \quad X \sim U(0, 1)$$

$$C_x(u_1, u_2) = \exp \left(- \left((-\log u_1)^{\theta(x)} + (-\log u_2)^{\theta(x)} \right)^{\frac{1}{\theta(x)}} \right)$$

$$\theta(x) = 2 \sin(2\pi x) + 3,$$

with X independent from (Y_1, Y_2) .

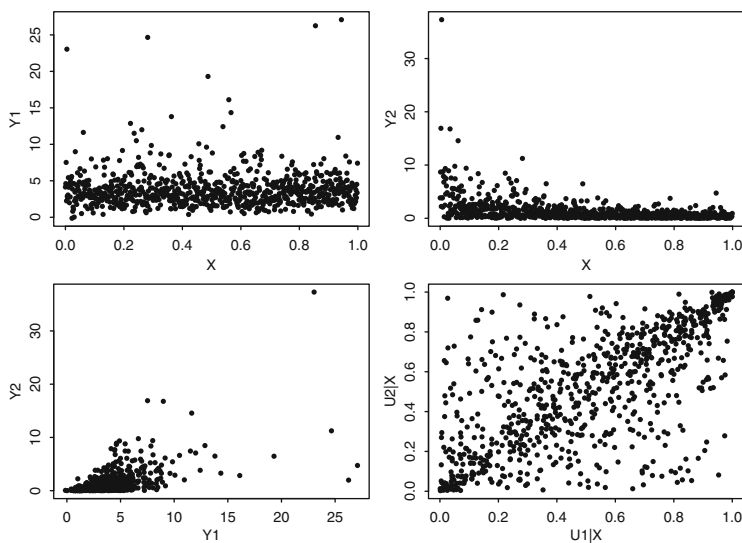


Fig. 4 Scatter plots based on a typical sample of size $n = 800$ from Example 6

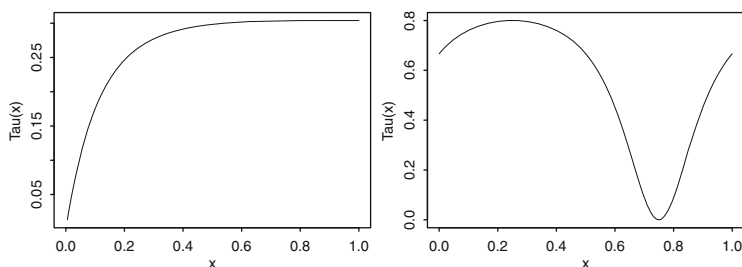


Fig. 5 Conditional Kendall's tau function for Examples 5 (left panel) and 6 (right panel)

In Fig. 4 we present the pairwise scatter plots for a typical sample of size $n = 800$ from Example 6, revealing the independence between Y_1 and X , but the dependence between Y_2 and X . In the bottom right panel of Fig. 4 we plot $F_{1X_i}(Y_{1i})$ versus $F_{2X_i}(Y_{2i})$, for each $i = 1, \dots, n$. Each of these $F_{jX_i}(Y_{ji})$ should be (close to) uniformly distributed. In Fig. 5 we plot the conditional Kendall's tau function for Examples 5 and 6. Note that in Example 5 there is a mild positive dependence between Y_1 and Y_2 , conditionally upon $X = x$, but that the dependence increases with x . In Example 6 the dependence switches from very positive via independence back to strongly positive dependence.

3.3 Statistical Inference

Suppose now that $((Y_{11}, Y_{21}, X_1), \dots, (Y_{1n}, Y_{2n}, X_n))$ is a sample of size n of independent observations from (Y_1, Y_2, X) . The interest is in estimating the conditional copula function C_x . As in Sect. 2, one can distinguish between different modeling settings, depending on what is known on possible appropriate parametric forms for C_x on the one hand and F_{jx} on the other hand. For convenience of the reader, we discuss similar modeling settings as in Sect. 2, but in a different order.

3.3.1 Fully Parametric Approach

In a fully parametric approach, we model $C_x(\cdot, \cdot)$ via $C(\cdot, \cdot; \theta_C(x))$ where $\theta_C(x)$ is a known parametric function of x , for example a polynomial of degree p : $\theta_C(x) = \theta_{C,1} + \theta_{C,2}x + \dots + \theta_{C,p}x^p$. We denote the corresponding parameter vector by $\theta_C = (\theta_{C,1}, \dots, \theta_{C,p})$. Similarly, the conditional margins can be modeled via

$$F_{jx}(\cdot) = F_j(\cdot; \theta_j(x)) \quad \text{with a parametric function } \theta_j(x) .$$

For example, the functions $\theta_j(x)$ could be polynomial functions or any other given parametric functional form. Denote the resulting parameter vectors by θ_1 and θ_2 respectively. For example, if $\theta_1(x)$ is a cubic function of x , then the dimension of θ_1 is 4.

With these parametrizations, we are again in a setting that is quite similar to that of Sect. 2.3.1. Indeed, considering the second order partial derivative of $C_x(u_1, u_2)$ with respect to its arguments, we obtain

$$c_x(u_1, u_2) = \frac{\partial^2 C(u_1, u_2; \theta_C(x))}{\partial u_1 \partial u_2} \quad \forall (u_1, u_2) \in [0, 1]^2 ,$$

which due to the structure can be written as $c(u_1, u_2; \theta_C(x))$. Analogously denote the marginal densities by $f_1(\cdot; \theta_1(x))$ and $f_2(\cdot; \theta_2(x))$.

A data point (Y_{1i}, Y_{2i}, X_i) contributes to the likelihood with the factor

$$c(F_1(Y_{1i}; \theta_1(X_i)), F_2(Y_{2i}; \theta_2(X_i)); \theta_C(X_i)) f_1(Y_{1i}; \theta_1(X_i)) f_2(Y_{2i}; \theta_2(X_i)) .$$

Finally, we get to the logarithm of the likelihood function

$$\begin{aligned} & \tilde{\ell}_n(\theta_1, \theta_2, \theta_C) \\ &= \sum_{i=1}^n \log (c(F_1(Y_{1i}; \theta_1(X_i)), F_2(Y_{2i}; \theta_2(X_i)); \theta_C(X_i)) f_1(Y_{1i}; \theta_1(X_i)) f_2(Y_{2i}; \theta_2(X_i))) \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=1}^n \log (c (F_1(Y_{1i}; \theta_1(X_i)), F_2(Y_{2i}; \theta_2(X_i)); \theta_C(X_i))) \\
 &\quad + \sum_{i=1}^n \log (f_1(Y_{1i}; \theta_1(X_i))f_2(Y_{2i}; \theta_2(X_i))) , \tag{8}
 \end{aligned}$$

which needs to be maximized with respect to $(\theta_1, \theta_2, \theta_C)$. From that point on we proceed as in Sect. 2.3.1. Denote the maximizer of (8) by $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_C)$.

An estimator for, for example, the associated conditional Kendall’s tau function is then obtained by substituting $C_x(\cdot, \cdot) = C(\cdot, \cdot; \theta_C(x))$ in (7) by its estimator $C(\cdot, \cdot; \hat{\theta}_C(x))$, where $\hat{\theta}_C(x)$ is obtained by replacing the parameter vector θ_C in the parametric form of $\theta_C(x)$ by its maximum likelihood estimator $\hat{\theta}_C$. For example, in case $\theta_C(x) = \theta_{C,1} + \theta_{C,2}x + \dots + \theta_{C,p}x^p$, this is $\hat{\theta}_C(x) = \hat{\theta}_{C,1} + \hat{\theta}_{C,2}x + \dots + \hat{\theta}_{C,p}x^p$ based on $\hat{\theta}_C = (\hat{\theta}_{C,1}, \dots, \hat{\theta}_{C,p})$. So, the estimator for the conditional Kendall’s tau is then

$$\hat{\tau}(x) = 4 \iint_{[0,1]^2} C(u_1, u_2; \hat{\theta}_C(x)) dC(u_1, u_2; \hat{\theta}_C(x)) - 1 .$$

3.3.2 Fully Nonparametric Approach

An alternative expression for (6) is

$$C_x(u_1, u_2) = H_x(F_{1x}^{-1}(u_1), F_{2x}^{-1}(u_2)) \quad (u_1, u_2) \in [0, 1]^2 , \tag{9}$$

where $F_{jx}^{-1}(\cdot)$ denotes the quantile function corresponding to $F_{jx}(\cdot)$, for $j = 1, 2$.

From (9) it is transparent that we need to find estimators for the conditional joint cumulative distribution function $H_x(\cdot, \cdot)$ as well as for the (quantiles of the) conditional margins $F_{1x}(\cdot)$ and $F_{2x}(\cdot)$. Since these are conditional quantities, some smoothing in the domain of X is needed. Nonparametric estimation of a conditional distribution function using kernel methods has been well-studied in the literature. See for example [23] and [38], among others. A general estimator is obtained by ‘smearing out’ the mass n^{-1} that is in the expression for a bivariate (unconditional) empirical distribution function, in the covariate domain, using a weight function:

$$\hat{H}_x(y_1, y_2) = \sum_{i=1}^n w_{ni}(x, b_n) I \{Y_{1i} \leq y_1, Y_{2i} \leq y_2\} , \tag{10}$$

with $w_{ni}(x, b_n) \geq 0$, a sequence of weights that ‘smooths’ over the covariate space. Herein $b_n > 0$ is a sequence of bandwidths. Since $H_x(y_1, y_2)$ is a distribution function, the weights need to tend to 1 when y_1 and y_2 tend to infinity. This is achieved by ensuring that the weights are such that $\sum_{i=1}^n w_{ni}(x, b_n) = 1$ (either

exactly or asymptotically, as $n \rightarrow \infty$). There are many scenario's of appropriate weight functions available in the literature. The simplest set of weights is given by the Nadaraya-Watson type of weights defined as

$$w_{ni}(x, b_n) = \frac{k_{b_n}(X_i - x)}{\sum_{j=1}^n k_{b_n}(X_j - x)},$$

with $k_{b_n}(\cdot) = \frac{1}{b_n}k(\cdot/b_n)$ a rescaled version of $k(\cdot)$. Alternative scenario's of weights are local linear weights, Gasser-Müller weights, etc.

From (10) we obtain estimators for the conditional marginal distribution functions $F_{jx}(\cdot)$ by simply letting the other argument in the estimated joint cumulative distribution function tend to infinity:

$$\hat{F}_{jx}(y) = \sum_{i=1}^n w_{ni}(x, b_{jn})I\{Y_{ji} \leq y\} \quad j = 1, 2, \tag{11}$$

where the bandwidth sequences $b_{1n} > 0$ and $b_{2n} > 0$ (for estimation of the conditional margins) do not need to be the same and/or do not need to be the same as this for the joint estimation. For practical simplicity one can take $b_n = b_{1n} = b_{2n}$.

Asymptotic properties for kernel type estimators of conditional distributions functions have been established in [35, 37] and [39], among others. For a recent contribution in the area, see [38].

Remark that in (10) and (11) one can again replace the indicator function by a smooth function, if differentiability properties of the resulting estimators are of importance.

From the estimators $\hat{F}_{jx}(\cdot)$ in (11), we obtain estimators for the quantile functions $F_{jx}^{-1}(\cdot)$. From the estimators for $H_x(\cdot, \cdot)$ and $F_{jx}^{-1}(\cdot)$ one then derives an estimator for $C_x(\cdot, \cdot)$ by replacing in (9) the former quantities by their estimators. In the literature these and improved estimators are studied, also in more complex frameworks (of multivariate or functional covariates). See, for example, [39] and [19].

3.3.3 Semiparametric Approach

There are at least a few semiparametric approaches, depending on the particular modeling setting. We just discuss some major approaches.

Firstly, assume that the conditional margins are fully known, i.e. $F_{jx}(\cdot)$, for $j = 1, 2$, are fully known for all x in the domain of X . Suppose that the conditional copula function $C_x(\cdot, \cdot)$ depends on x through a parameter function $\theta_C(x)$, i.e. $C_x(\cdot, \cdot) = C(\cdot, \cdot; \theta_C(x))$, but contrary to Sect. 3.3.1 the function $\theta_C(x)$ is fully unknown. This setting has been studied by [22] and [2], among others. In the sequel we drop the subscript C in θ_C and $\theta_C(x)$ for simplifying the notation.

The parametric copula family $C(\cdot, \cdot; \theta)$ that serves as a starting point here (and in fact also in Sect. 3.3.1, and before) of course has some restrictions on the parameter

space Θ for θ . For example, for a Gaussian copula: $\theta \in (-1, 1)$; for a Clayton copula $\theta \in (0, \infty)$. Such restrictions on the parameter space of the parametric copula $C(\cdot, \cdot; \theta)$ should in fact not be ignored. The same holds when looking at a conditional copula modeled via $C_x(u_1, u_2) = C(u_1, u_2; \theta(x))$, with corresponding copula density $c(u_1, u_2; \theta(x))$.

Since the function $\theta(x)$ is fully unknown, we are going to approximate this function locally, i.e. in the neighbourhood of x by, for example, a polynomial of degree p say. But since a polynomial takes on values in \mathbb{R} , we need to take care of the restrictions on the parameter space Θ of the parametric copula family $C(\cdot, \cdot; \theta)$. One therefore transforms the function $\theta(x)$, which takes on values in Θ , via a given transformation $\psi(\cdot)$, into the function

$$\eta(x) = \psi(\theta(x)) ,$$

which takes on value in \mathbb{R} .

In the sequel, we assume that the inverse transformation $\psi^{-1}(\cdot)$ exists, such that we can obtain $\theta(\cdot)$ from $\eta(\cdot)$:

$$\theta(x) = \psi^{-1}(\eta(x)) , \tag{12}$$

which takes values in Θ .

Consider now independent observations $((Y_{11}, Y_{21}, X_1), \dots, (Y_{1n}, Y_{2n}, X_n))$ from (Y_1, Y_2, X) . A data point (Y_{1i}, Y_{2i}, X_i) then contributes to the (pseudo) log-likelihood with

$$\log c(F_{1X_i}(Y_{1i}), F_{2X_i}(Y_{2i}); \theta(X_i)) = \log c(F_{1X_i}(Y_{1i}), F_{2X_i}(Y_{2i}); \psi^{-1}(\eta(X_i))) . \tag{13}$$

For a data point X_i in a neighbourhood of x , we then can approximate $\eta(X_i)$ using a Taylor expansion, by

$$\begin{aligned} \eta(X_i) &\approx \eta(x) + \eta'(x)(X_i - x) + \dots + \frac{\eta^{(p)}(x)(X_i - x)^p}{p!} \\ &\equiv \beta_0 + \beta_1(X_i - x) + \dots + \beta_p(X_i - x)^p , \end{aligned}$$

where we denoted

$$\beta_r = \beta_r(x) = \frac{\eta^{(r)}(x)}{r!} \quad r = 0, \dots, p .$$

If X_i is near x , then the contribution in (13) to the (pseudo) log-likelihood, can be approximated by

$$\begin{aligned} &\log c(F_{1X_i}(Y_{1i}), F_{2X_i}(Y_{2i}); \theta(X_i)) \\ &\approx \log c(F_{1X_i}(Y_{1i}), F_{2X_i}(Y_{2i}); \psi^{-1}(\beta_0 + \beta_1(X_i - x) + \dots + \beta_p(X_i - x)^p)) . \end{aligned}$$

This approximation is only valid for X_i near x , and this is taken care off by multiplying this contribution in the log-likelihood by a weight factor $k_{h_n}(\cdot) = \frac{1}{h_n}k(\frac{\cdot}{h_n})$, with $k(\cdot)$ as before, and $h_n > 0$ a bandwidth parameter.

This leads to the local log-likelihood

$$\ell_n(\boldsymbol{\beta}) = \sum_{i=1}^n \log c(F_{1X_i}(Y_{1i}), F_{2X_i}(Y_{2i}); \psi^{-1} \{\beta_0 + \beta_1(X_i - x) + \dots + \beta_p(X_i - x)^p\}) \times k_{h_n}(X_i - x), \tag{14}$$

which is a localized version of the (pseudo) log-likelihood function in the parametric setting. Maximization of this local log-likelihood with respect to $\boldsymbol{\beta}$ leads to the estimated vector $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \dots, \hat{\beta}_p)$, and hence, in particular, an estimator for $\eta(x)$ is $\hat{\beta}_0$. From (12) an estimator for $\theta(x)$ is

$$\hat{\theta}(x) = \psi^{-1}(\hat{\beta}_0) .$$

By maximizing the local log-likelihood (14) in a grid of x -values, one obtains estimates of the unknown parameter function $\theta(\cdot)$ in a grid of points.

We next turn to the setting where also the conditional margins are fully unknown, i.e. $F_{jx}(\cdot)$, for $j = 1, 2$, are fully unknown for all x in the domain of X . For X_i in a neighbourhood of x , we then can replace the contribution

$$\log c(F_{1X_i}(Y_{1i}), F_{2X_i}(Y_{2i}); \psi^{-1} \{\beta_0 + \beta_1(X_i - x) + \dots + \beta_p(X_i - x)^p\}) k_{h_n}(X_i - x)$$

in the log-likelihood function by

$$\log c(F_{1x}(Y_{1i}), F_{2x}(Y_{2i}); \psi^{-1} \{\beta_0 + \beta_1(X_i - x) + \dots + \beta_p(X_i - x)^p\}) k_{h_n}(X_i - x) ,$$

and next we substitute the unknown quantities $F_{1x}(Y_{1i})$ and $F_{2x}(Y_{2i})$ by nonparametric estimators, such as these provided in (11). This then leads to the local log-likelihood

$$\tilde{\ell}_n(\boldsymbol{\beta}) = \sum_{i=1}^n \log c(\hat{F}_{1x}(Y_{1i}), \hat{F}_{2x}(Y_{2i}); \psi^{-1} \{\beta_0 + \beta_1(X_i - x) + \dots + \beta_p(X_i - x)^p\}) \times k_{h_n}(X_i - x) ,$$

which needs to be maximized with respect to $\boldsymbol{\beta}$.

This estimation method is called a local polynomial maximum pseudo log-likelihood estimation method. Properties of the resulting estimator have been studied in [1]. That paper also contains a brief discussion on some practical bandwidth selection methods, including a rule-of-thumb type of bandwidth selector and a cross-validation procedure. For a general treatment of the use of local polynomial

modeling in a maximum likelihood framework, see for example [13]. For a general discussion on the choice of the degree of the polynomial approximation see [12].

4 Dynamics of a Dependence Structure and Copulas

When introducing copulas to the modeling of time series data different approaches are possible. Following, for example, [9] copulas can be used to model the inter-temporal dependence within one time series by specifying the transition probabilities in a Markov process. See [8] for recent developments and further references for such settings. In this section we focus on a different approach.

4.1 Dynamical Modeling of a Dependence Structure

Alternatively copulas can be used to model the spatial dependence of a bivariate stochastic process $\mathbf{Y}_t = (Y_{1,t}, Y_{2,t})$, $t \in \mathbb{Z}$. As time series analysis is naturally formulated conditionally upon the history of the process we revert to the conditional copula concept, now enlarging the conditioning from one variable as in Sect. 3 to the entire past of the process. This is done in a mathematical rigorously way by conditioning upon the sigma-algebras generated by the past of the time series. We opted for a more layman’s term presentation here. As first introduced by [30] such a setting allows time dependent variation in the joint distribution of $(Y_{1,t}, Y_{2,t})$ conditionally upon $\mathbf{W}_t = (\mathbf{Y}_{t-k})_{k>0}$ via

$$P\{Y_{1,t} \leq y_1, Y_{2,t} \leq y_2 | \mathbf{W}_t = \mathbf{w}_t\} = C_t(F_{1,t}(y_1), F_{2,t}(y_2)) , \tag{15}$$

where $F_{j,t}(y_j) = P\{Y_{j,t} \leq y_j | \mathbf{W}_t = \mathbf{w}_t\}$, $j = 1, 2$, and $C_t(\cdot, \cdot) = C(\cdot, \cdot | \mathbf{W}_t = \mathbf{w}_t)$ is the conditional copula implied by Sklar’s Theorem.

The conditional modeling in (15) can be readily combined with, for example, a GARCH(r, s) error structure for the two involved time series $\{Y_{1,t}\}$ and $\{Y_{2,t}\}$. See [5], for example, for a standard reference to Generalized Autoregressive Conditional Heteroskedasticity (GARCH) type of modeling. More precisely, we model the marginal time series as

$$Y_{j,t} = \mu_j + \varepsilon_{j,t}, \quad \text{where} \quad \varepsilon_{j,t} = \sigma_{j,t} \eta_{j,t} , \tag{16}$$

$$\sigma_{j,t}^2 = \alpha_j + \sum_{\ell=1}^r \beta_{j,\ell} \sigma_{j,t-\ell}^2 + \sum_{m=1}^s \gamma_{j,m} \varepsilon_{j,t-m}^2 , \tag{17}$$

where $\alpha_j, \beta_{j,\ell}, \gamma_{j,m} \geq 0$, $\mu_j \in \mathbb{R}$, $r, s \in \mathbb{N}_0$ and $(\eta_{j,t})_{t \in \mathbb{Z}}$ is white noise with zero mean and unit variance. GARCH models are designed to account for the time varying and clustering volatility of shocks observed frequently, but not exclusively, in financial time series. This is accomplished by relating the time t variance of $\varepsilon_{j,t}$ to the lagged r realized variances as well as to the lagged s realized shocks, where it is noteworthy that the dependence on the lagged shocks makes the variance itself stochastic (random). Combining (15) and (16) the marginal time series can now be fused into a joint model by specifying its conditional joint distribution

$$\mathbf{Y}_t | \mathbf{W}_t = \mathbf{w}_t \sim C_t(F_{1,t}, F_{2,t}),$$

where the conditional mean and variance of $F_{j,t}$, for $j = 1, 2$, are only determined by information from the past (up to time point $t - 1$) and are given as μ_j and $\sigma_{j,t}^2$. This framework combines autocorrelated shocks with a flexible modeling of the conditional joint distribution. A detailed review of copula models in economic time series can be found in [31]. For the purpose of this chapter, we focus on applying a semiparametric approach as discussed in Sect. 3.3.3 to the time series framework, with the difference that in the semiparametric approach described here, we model the marginal time series via parametric GARCH models. For simplicity, let $t \in \mathbb{N}_0$, and denote by $(Y_{1,t}, Y_{2,t})_{t=1}^T$ the available sample of size T ($T \in \mathbb{N}_0$). Furthermore, denote the observed (standardized) time points by t/T , so that all observational points t/T are in the interval $[0, 1]$. The conditional copula is chosen to be time dependent through an unknown parameter function $\theta_C(t^*)$ for $t^* \in [0, 1]$. To obey restrictions in the parameter space we again consider a suitable one-to-one transformation ψ such that $\eta(t^*) = \psi(\theta_C(t^*)) \in \mathbb{R}$ and recover θ_C via $\theta_C(t^*) = \psi^{-1}(\eta(t^*))$. For a sample $(Y_{1,t}, Y_{2,t})_{t=1}^T$ we can write the log-likelihood of the overall joint density by successive conditioning in terms of the contributions of the bivariate densities $\mathbf{Y}_t | \mathbf{W}_t = \mathbf{w}_t$ to the log-likelihood as

$$\begin{aligned} \ell_T &= \sum_{t=1}^T \log(c(F_{1,t}(Y_{1,t}), F_{2,t}(Y_{2,t}); \theta_C(t/T))) + \sum_{j=1}^2 \sum_{t=1}^T \log(f_{j,t}(Y_{j,t})) \\ &= \ell_{T,C} + \ell_{T,1} + \ell_{T,2}. \end{aligned}$$

See also (8) in Sects. 3.3.1 and 3.3.3. Following the so-called inference of margins approach (a two-steps procedure), see [26], maximizing ℓ_T can be accomplished by first separately maximizing $\ell_{T,1}$ and $\ell_{T,2}$, under our semiparametric setting by a standard parametric maximum likelihood estimation method, and then maximizing $\ell_{T,C}$ taking estimates of the previous step into account by replacing the distribution functions $F_{j,t}$ by their respective estimates $\hat{F}_{j,t}$ (for $j = 1, 2$). In order to maximize $\ell_{T,C}$ we extend the local constant fitting approach of [22] to the local polynomial approach discussed in Sect. 3.3.2. Asymptotic normality of the resulting estimator in case of local constant fitting can be found in [22]. The study of the asymptotic properties of the local polynomial dynamic copula estimator, presented in this

section, is part of future research. Consider a fixed point $t^* \in [0, 1]$, and denote by $k_{h_n}(\cdot)$ a rescaled kernel with bandwidth h_n , as before. The local log-likelihood for the problem considered is then given by

$$\begin{aligned} &\ell_{T,C}(\boldsymbol{\beta}) \\ &= \sum_{t=1}^T \log c \left(\hat{F}_{1,t}(Y_{1,t}), \hat{F}_{2,t}(Y_{2,t}); \psi^{-1} \right. \\ &\quad \left. \times \left\{ \beta_0 + \beta_1 \left(\frac{t}{T} - t^* \right) + \dots + \beta_p \left(\frac{t}{T} - t^* \right)^p \right\} \right) \times k_{h_n} \left(\frac{t}{T} - t^* \right), \end{aligned} \quad (18)$$

which needs to be maximized with respect to $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)$.

4.2 Illustrative Example

We illustrate the presented methodology by simulating from a bivariate GARCH(1,1) model, where the conditional marginal distributions are set to be Normal distributions, and the conditional copula is assumed to be a student t copula where the degrees of freedom are fixed to 4. The remaining free copula parameter is a time varying parameter function $\theta_C(t) = 2 \sin(0.95\pi(2B(t; 2, 3) - 1)/6)$, with $B(t; a, b)$ the cumulative distribution function of the Beta distribution. See [11] for a reference on (student) t copulas. The remaining parameters are set to $\mu_1 = \mu_2 = 0$, $\alpha_1 = \alpha_2 = 0.1$, $\beta_{1,1} = \beta_{2,1} = 0.4$ and $\gamma_{1,1} = \gamma_{2,1} = 0.4$.

Figure 6 show the first and second marginal time series for a simulated trajectory of the process, highlighting the volatility clustering inherent in the process. In Fig. 7 we show simulated scatter plots of the unconditional distribution at three different time points. As expected the time varying conditional dependence structure carries

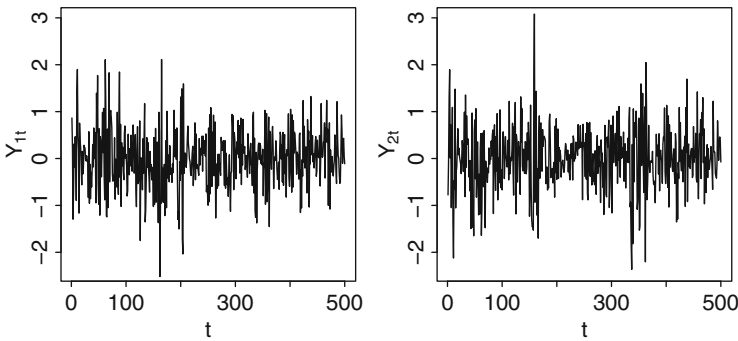


Fig. 6 Marginal time series $Y_{1,t}$ (left) and $Y_{2,t}$ (right) of a simulated bivariate copula-GARCH(1,1) model, $t = 1, \dots, 500$

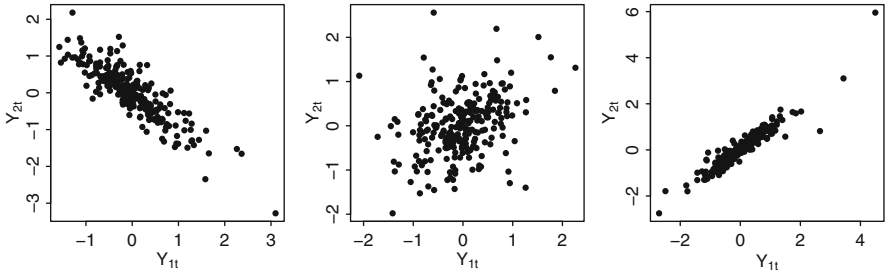


Fig. 7 Simulated scatter plot of $(Y_{1,50}, Y_{2,50})$ (left panel); $(Y_{1,250}, Y_{2,250})$ (middle panel) and $(Y_{1,450}, Y_{2,450})$ (right panel). The scatter plots are based on 250 independently simulated trajectories

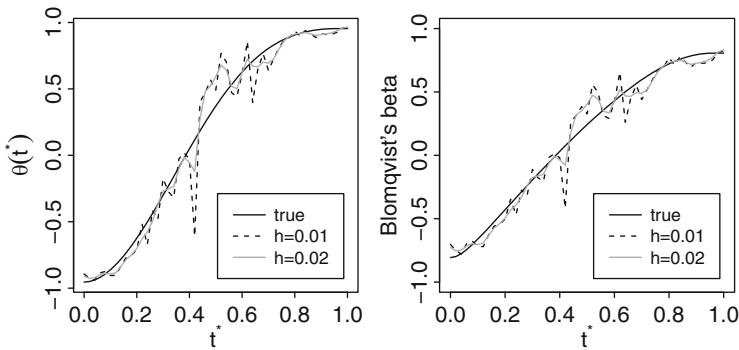


Fig. 8 True (solid curve) and estimated curve using bandwidths $h_n = 0.01$ (dashed) and $h_n = 0.02$ (dotted). Left panel: true and estimated $\theta(t^*)$. Right panel: true and estimated conditional Blomqvist's beta. Calculations are based on a simulated sample of size $T = 500$

over to the distribution of $(Y_{1,t}, Y_{2,t})$, displaying a negative dependence at early time stages, and gradually switching to a positive dependence later on (moving from the left panel to the right panel).

To illustrate the local log-likelihood estimation of $\theta(\cdot)$ we first obtain estimates $\hat{\mu}_1, \hat{\mu}_2, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_{1,1}, \hat{\beta}_{2,1}, \hat{\gamma}_{1,1}$ and $\hat{\gamma}_{2,1}$ by fitting a GARCH(1,1) model to each of the marginal time series individually. From the estimates we can then recover, for $j = 1, 2$, the conditional variances $\hat{\sigma}_{j,t}^2$ to find $\hat{F}_{j,t}(Y_{j,t}) = \Phi((Y_{j,t} - \hat{\mu}_j)/\hat{\sigma}_{j,t})$, where Φ denotes the standard normal distribution function. To perform the local log-likelihood estimation in (18) we settle for a local approximation with a polynomial of degree $p = 1$, i.e. performing local linear fitting. For a fixed point $t^* \in [0, 1]$ we then maximize $\ell_{T,C}(\beta_0, \beta_1)$ as given in (18), where we choose $\psi^{-1} : \mathbb{R} \rightarrow (-1, 1)$, with $\psi^{-1}(x) = \tanh(x)$.

In the left panel of Fig. 8 we show the true and estimated $\theta(t^*)$ when using different bandwidths h_n in the estimation procedure. In the right panel of Fig. 8 we also plot the true and estimated conditional Blomqvist's beta as a function of time, using the same bandwidths as for the estimation of $\theta(\cdot)$.

5 Dynamic Modeling via Copulas: Application in Forecasting

We consider wind speed data obtained by weather stations in Kennewick (southern Washington) and Vansycle/Butler Grade (north-eastern Oregon), in the USA. The raw data¹ consist of average wind speeds (in miles per hour = mph) for intervals of 5 min. For the analysis here we restrict to the period between April 1 and July 1, 2013. A more detailed description and analysis of these data can be found in [21]. In this section, we illustrate how the discussed methods can be used for estimation and for forecasting of wind speeds.

To fit the time series data we extend the model described in (16) and (17) to include an autoregressive component, leading to an AR(q)-GARCH(r, s) model, where (16) is replaced by (for $j = 1, 2$)

$$Y_{j,t} = \mu_j + \sum_{\ell=1}^q \phi_{j,\ell} Y_{j,t-\ell} + \varepsilon_{j,t}, \quad \text{where} \quad \varepsilon_{j,t} = \sigma_{j,t} \eta_{j,t}, \quad (19)$$

and (17) is kept. Herein $\phi_{j,\ell} \in \mathbb{R}$, $q \in \mathbb{N}_0$. As in Sect. 4 the conditional marginal distributions are Normal, and the marginal time series are coupled by a time dependent copula to form a bivariate model.

As wind speed forecasts with a two-hour forecast horizon are needed to meet practical demands (see [21]), we transform the raw data into the hourly averages, shown in Fig. 9, and denote the sample by $(Y_{1,t}, Y_{2,t})_{t=1}^{2,184}$. In the next step we fit an AR(1)-GARCH(1, 1) model, described by (19) and (17), to the marginal time series in an one hour rolling window type fashion as follows. The first estimates

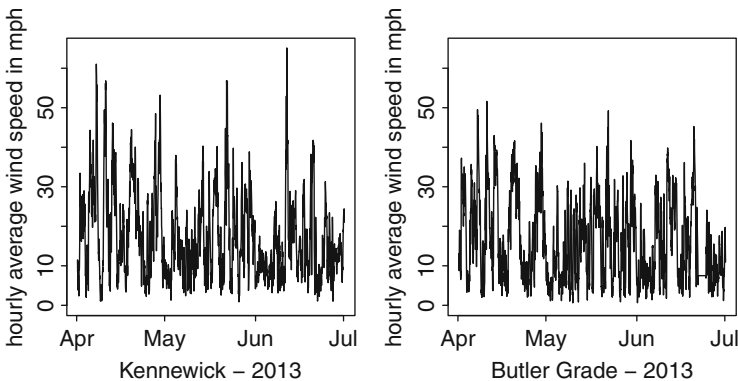


Fig. 9 Hourly averages of wind speed (mph) at Kennewick (*left*) and at Butler Grade (*right*) from April 1 to July 1, 2013

¹Datasets can be obtained from the web site of the Bonneville Power Administration under <http://transmission.bpa.gov/Business/Operations/Wind/MetData.aspx>.

$(\hat{\mu}_j, \hat{\phi}_{j,1}, \hat{\alpha}_j, \hat{\beta}_{j,1}, \hat{\gamma}_{j,1}), j = 1, 2$, are obtained by fitting the AR-GARCH model to $(Y_{j,t})_{t=1}^{744}$. The second estimates are then based on the shifted data $(Y_{j,t})_{t=2}^{745}$ and so forth. We repeat this process 240 times, yielding estimates covering a span of 10 days. By keeping the number of observations fixed at 744 all estimates are effectively based on data of the last respective 31 days. The so obtained estimates are plotted against the window shift in Figs. 10 and 11. While the mean and AR coefficients (Fig. 10) are very comparable between both sites, the GARCH parameters show a differing pattern: while the baseline variance and lagged realized shock coefficients are generally higher in Kennewick than in Butler Grade (left and right panels of Fig. 11), the situation is reversed considering the dependence on lagged realized variances. See the middle panel of Fig. 11. The two weather stations are on different altitudes, the difference being around 130 m. This could may be explain the differences noticed.

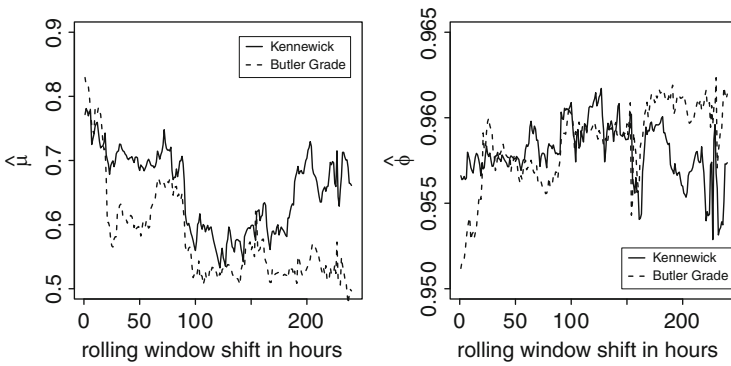


Fig. 10 Estimates of μ_j (left panel) and of $\phi_{j,1}$ (right panel), for $j = 1, 2$, for the 240 rolling windows, based on 744 observations each

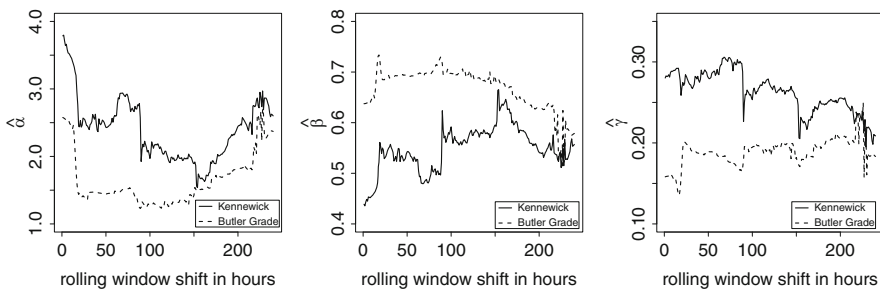


Fig. 11 Estimates of $\alpha_{j,1}$ (left panel), of $\beta_{j,1}$ (middle panel) and of $\gamma_{j,1}$ (right panel), for $j = 1, 2$, for the 240 rolling windows, based on 744 observations each

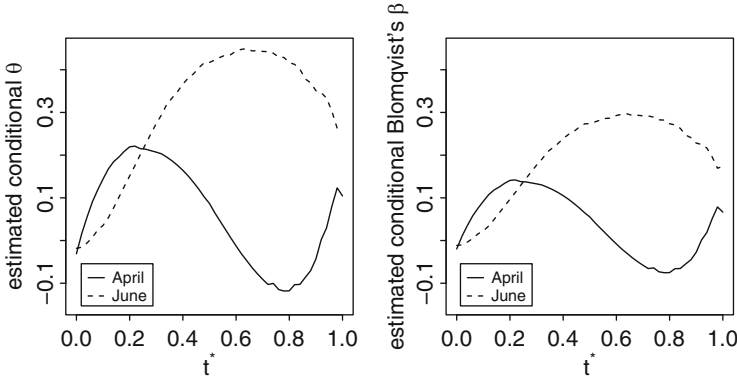


Fig. 12 Estimates of $\theta(t^*)$ (left) and the conditional Blomqvist's beta (right), by local-linear fitting with $h = 0.2$, for a student t copula with 4 degrees of freedom based on the observations $(Y_{1,t}, Y_{2,t})_{t=1}^{744}$ (for the April period) and on the observations $(Y_{1,t}, Y_{2,t})_{t=1,441}^{2,184}$ (for the June period). Solid line: April period; Dotted line: June period

The conditional dependence structure between the marginal time series is modeled by a student t copula with 4 fixed degrees of freedom, where the remaining parameter θ is allowed to vary as a smooth function of time, as explained in Sect. 4. In Fig. 12 (left panel) we show the estimated conditional copula parameter function for observations $(Y_{j,t})_{t=1}^{744}$ (the solid curve) using the previously estimated AR-GARCH parameters, and applying local linear fitting with bandwidth $h = 0.2$ (see Sect. 4). We repeat the procedure also based on the very last 744 observations $(Y_{j,t})_{t=1,441}^{2,184}$ (i.e. the June period) and show the results in the left panel of Fig. 12 (the dotted curve). We also present the corresponding results for the estimated conditional Blomqvist's beta in the right panel of Fig. 12. As can be seen, the dependence structure (between the observations from the two stations) varies within the periods (April and June), but also seems to be different for the two periods examined (early spring and summer period).

Turning towards forecasting we compute for each set of rolling window estimates the conditional copula parameter at $t^* = 1$, i.e. $\hat{\theta}(744)$ for the first window, $\hat{\theta}(745)$ for the second and so on. This yields successive estimates $(\hat{\beta}_0, \hat{\beta}_1)$ that we use to predict the one hour ahead forecast and, respectively, the two hours ahead forecast of $\theta(t)$ by $\tilde{\theta}(T + k) = \psi^{-1}(\hat{\beta}_0 + k\hat{\beta}_1/744)$, with $k = 1$, respectively $k = 2$, where ψ^{-1} is the link function as in Sect. 4 and T denotes the last time point of the respective rolling window sample. The obtained 2 h ahead forecast of θ is shown in the left panel of Fig. 13.

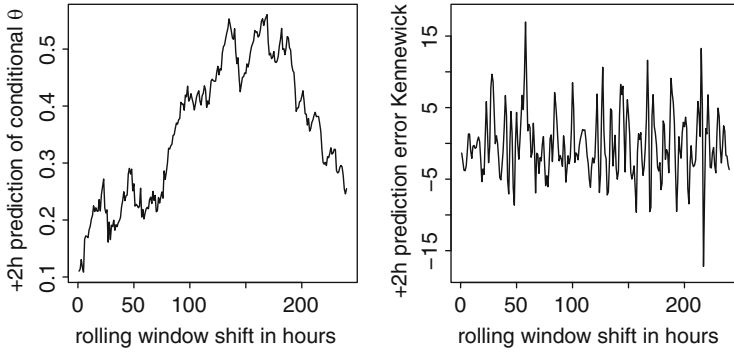


Fig. 13 *Left*: Two hours ahead prediction of $\theta(\cdot)$ for the rolling windows, based on 744 observations. *Right*: Realized minus two hours ahead predicted wind speeds in Kennewick for the rolling windows

Concerning the marginal time series we use the point estimates of the parameters as one and two hours ahead forecasts of the parameters. This allows to compute point forecasts of the wind speed by (see (19))

$$\begin{aligned} \tilde{Y}_{j,T+1} &= E[Y_{j,T+1} | Y_{j,T-\ell}, \ell \geq 0] = \mu_j + \phi_{j,1} Y_{j,T} \\ \tilde{Y}_{j,T+2} &= E[Y_{j,T+2} | Y_{j,T-\ell}, \ell \geq 0] = \mu_j + \phi_{j,1} (\mu_j + \phi_{j,1} Y_{j,T}) \end{aligned}$$

for $j = 1, 2$.

To assess the forecast quality we compute the square root of the mean squared error of the predicted to the realized values for the 240 forecasts, denoted by RMSE. For one hours ahead forecasts we obtain: RMSE = 2.9485 for the Kennewick station, and RMSE = 3.0026 for the Butler Grade station. For the two hours ahead forecasts these are: RMSE = 4.4725 for Kennewick and RMSE = 4.8190 for Butler Grade. The right panel of Fig. 13 depicts the differences of realized to predicted two hours ahead forecasts at Kennewick.

Having predictions of the marginal time series, as well as the conditional dependence between them allows to go beyond point forecasting and to predict their joint behaviour. In Fig. 14 we show contours of the predicted two hours ahead joint distribution for the first rolling window, and then 5.5 days later, in respectively the left and right side panels. As shown by the figure, not only the mean of the distribution, represented by the wind speed point forecasts, but also the shape changes as implied by the prediction of $\theta(\cdot)$.

To further visualize the impact of the time varying association we forecast the probability of $Y_{1,T+2}$ and $Y_{2,T+2}$ staying jointly below their conditional $\alpha \times 100\%$ quantiles. Within a copula framework this probability equals $C(\alpha, \alpha; \theta(T + 2))$. See [7] for an application in economics. For example based on the first rolling

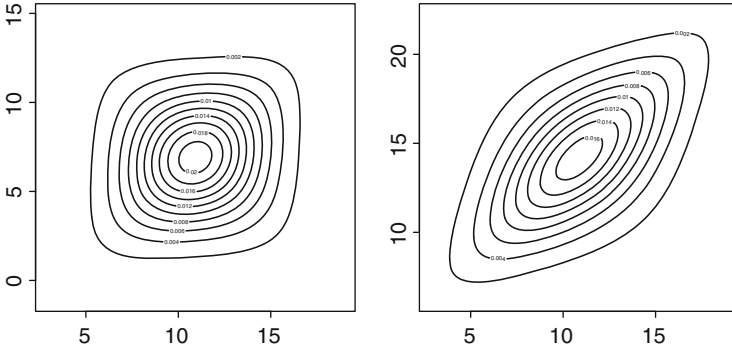


Fig. 14 Two hours ahead prediction of the conditional joint density of $(Y_{1,746}, Y_{2,746})$ (left) and the conditional joint density of $(Y_{1,877}, Y_{2,877})$ (right) based on the first rolling window, respectively on rolling window 132

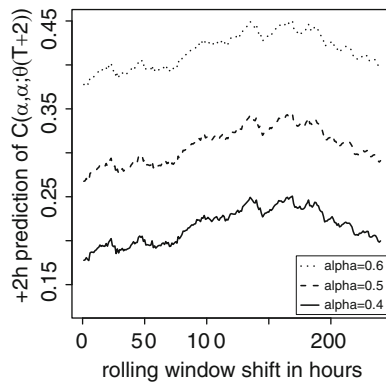


Fig. 15 Two hours ahead prediction of $C(\alpha, \alpha; \theta(T + 2))$ for the rolling windows, based on 744 observations

window we compute the conditional two hours ahead 40 % quantiles as 10.1487 (for Kennewick) and 6.1633 (for Butler Grade). This yields a prediction of $P\{Y_{1,746} \leq 10.1487, Y_{2,746} \leq 6.1633\} = 0.1777$. Figure 15 shows the obtained predictions for different values of α .

Acknowledgements The authors thank the organizers of the “Second workshop on Industry Practices for Forecasting” (WIPFOR 2013) for a very simulating meeting. This research is supported by IAP Research Network P7/06 of the Belgian State (Belgian Science Policy), and the project GOA/12/014 of the KU Leuven Research Fund. The third author is Postdoctoral Fellow of the Research Foundation – Flanders, and acknowledges support from the foundation.

References

1. Abegaz, F., Gijbels, I., & Veraverbeke, N. (2012). Semiparametric estimation of conditional copulas. *Journal of Multivariate Analysis, Special Issue on "Copula Modeling and Dependence"*, 110, 43–73.
2. Acar, E. F., Craiu, R. V., & Yao, F. (2011). Dependence calibration in conditional copulas: A nonparametric approach. *Biometrics*, 67, 445–453.
3. Acar, E. F., Genest, C., & Nešlehová, J. (2012). Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis, Special Issue on "Copula Modeling and Dependence"*, 110, 74–90.
4. Blomqvist, N. (1950). On a measure of dependence between two random variables. *The Annals of Mathematical Statistics*, 21, 593–600.
5. Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307–327.
6. Chen, S. C., & Huang, T.-M. (2007). Nonparametric estimation of copula functions for dependence modelling. *The Canadian Journal of Statistics*, 35, 265–282.
7. Cherubini, U., & Luciano, E. (2001). Value-at-risk trade-off and capital allocation with copulas. *Economic Notes*, 30, 235–256.
8. Cherubini, U., Mulinacci, S., Gobbi, F., & Romagnoli S. (2011). *Dynamic copula methods in finance*. New York: Wiley.
9. Darsow, W. F., Nguyen, B., & Olsen, E. T. (1992). Copulas and Markov processes. *Illinois Journal of Mathematics*, 36, 600–642.
10. Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. *Académie Royale de Belgique, Bulletin de la Classe des Sciences, 5e Série*, 65, 274–292.
11. Demarta, S., & McNeil, A. J. (2005). The t copula and related copulas. *International Statistical Review*, 73, 111–129.
12. Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications*. London: Chapman and Hall.
13. Fan, J., Farmen, M., & Gijbels, I. (1998). Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society, Series B*, 60, 591–608.
14. Fermanian, J.-D., Radulović, D., & Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10, 847–860.
15. Genest, C., Ghoudi, K., & Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82, 543–552.
16. Gini, C. (1909). Concentration and dependency ratios (in Italian). English translation in *Rivista di Politica Economica*, 87, 769–789 (1997).
17. Gini, C. (1912). Variability and mutability (in Italian, 156 p.). Bologna: C. Cuppini. (Reprinted in E. Pizetti & T. Salvemini (Eds.), *Memorie di metodologica statistica*. Rome: Libreria Eredi Virgilio Veschi (1955))
18. Gijbels, I., & Mielniczuk, J. (1990). Estimating the density of a copula function. *Communications in Statistics – Theory and Methods*, 19, 445–464.
19. Gijbels, I., Omelka, M., & Veraverbeke, N. (2012). Multivariate and functional covariates and conditional copulas. *Electronic Journal of Statistics*, 6, 1273–1306.
20. Gijbels, I., Veraverbeke, N., & Omelka, M. (2011). Conditional copulas, association measures and their applications. *Computational Statistics & Data Analysis*, 55, 1919–1932.
21. Gneiting, T., Larson, K., Westrick, K., Genton, M., & Aldrich, E. (2006). Calibrated probabilistic forecasting at the stateline wind energy center: The regime-switching-space-time method. *Journal of the American Statistical Society*, 101, 968–979.
22. Hafner, C., & Reznikova, O. (2010). Efficient estimation of a semiparametric dynamic copula model. *Computational Statistics & Data Analysis*, 54, 2609–2627.
23. Hall, P., Wolff, R. C. L., & Yao, Q. (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association*, 94, 154–163.

24. Hobæk Haff, I., Aas, K., & Frigessi, A. (2010). On the simplified pair-copula construction – Simply useful or too simplistic? *Journal of Multivariate Analysis*, *101*, 1296–1310.
25. Hollander, M., & Wolfe, D. A. (1999). *Nonparametric statistical methods* (2nd ed.). New York: Wiley.
26. Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, *94*, 401–419.
27. Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association*, *53*, 814–861.
28. Nelsen, R. B. (2006). *An introduction to copulas* (Lecture notes in statistics, 2nd ed.). New York: Springer.
29. Omelka, M., Gijbels, I., & Veraverbeke, N. (2009). Improved kernel estimation of copulas: Weak convergence and goodness-of-fit testing. *The Annals of Statistics*, *37*, 3023–3058.
30. Patton, A. (2006). Modeling asymmetric exchange rate dependence. *International Economical Review*, *47*, 527–556.
31. Patton, A. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis*, *110*, 4–18.
32. Segers, J. (2012). Weak convergence of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli*, *18*, 764–782.
33. Segers, J., van den Akker, R., & Werker, B. J. M. (2014). Semiparametric gaussian copula models: Geometry and efficient rank-based estimation. *Annals of Statistics*, *42*(5), 1911–1940.
34. Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de L'Institut de Statistique de L'Université de Paris*, *8*, 229–231.
35. Stute, W. (1986). On almost sure convergence of conditional empirical distribution functions. *The Annals of Statistics*, *14*, 891–901.
36. Tatsu, J., Pinson, P., Madsen, H. (2015). Space-time trajectories of wind power generation: Parametrized precision matrices under a Gaussian copula approach. *Lecture Notes in Statistics 217: Modeling and Stochastic Learning for Forecasting in High Dimension*, 267–296.
37. Van Keilegom, I., & Veraverbeke, N. (1997). Estimation and bootstrap with censored data in fixed design nonparametric regression. *The Annals of the Institute of Statistical Mathematics*, *49*, 467–491.
38. Veraverbeke, N., Gijbels, I., & Omelka, M. (2014). Pre-adjusted nonparametric estimation of a conditional distribution function. *Journal of the Royal Statistical Society, Series B*, *76*, 399–438.
39. Veraverbeke, N., Omelka, M., & Gijbels, I. (2011). Estimation of a conditional copula and association measures. *The Scandinavian Journal of Statistics*, *38*, 766–780.