



Research Article

Marija Tepegjzova*, Jing Zhou, Gerda Claeskens, and Claudia Czado

Nonparametric C- and D-vine-based quantile regression

<https://doi.org/10.1515/demo-2022-0100>

received August 14, 2021; accepted January 14, 2022

Abstract: Quantile regression is a field with steadily growing importance in statistical modeling. It is a complementary method to linear regression, since computing a range of conditional quantile functions provides more accurate modeling of the stochastic relationship among variables, especially in the tails. We introduce a nonrestrictive and highly flexible nonparametric quantile regression approach based on C- and D-vine copulas. Vine copulas allow for separate modeling of marginal distributions and the dependence structure in the data and can be expressed through a graphical structure consisting of a sequence of linked trees. This way, we obtain a quantile regression model that overcomes typical issues of quantile regression such as quantile crossings or collinearity, the need for transformations and interactions of variables. Our approach incorporates a two-step ahead ordering of variables, by maximizing the conditional log-likelihood of the tree sequence, while taking into account the next two tree levels. We show that the nonparametric conditional quantile estimator is consistent. The performance of the proposed methods is evaluated in both low- and high-dimensional settings using simulated and real-world data. The results support the superior prediction ability of the proposed models.

Keywords: vine copulas, conditional quantile function, nonparametric pair-copulas

MSC 2020: 62H05, 62G08, 62G05

1 Introduction

As a robust alternative to the ordinary least squares regression, which estimates the conditional mean, quantile regression [38] focuses on the conditional quantiles. This method has been studied extensively in statistics, economics, and finance. The pioneer literature by Koenker [35] investigated linear quantile regression systematically. It presented properties of the estimators including asymptotic normality and consistency, under various assumptions such as independence of the observations, independent and identically distributed (i.i.d.) errors with continuous distribution, and predictors having bounded second moment. Subsequent extensions of linear quantile regression have been intensively studied, see for example adapting quantile regression in the Bayesian framework [66], for longitudinal data [34], time-series models [63], high-dimensional models with l_1 -regularizer [7], nonparametric estimation by kernel

* **Corresponding author: Marija Tepegjzova**, Department of Mathematics, Technische Universität München, Boltzmannstraße 3, 85748, Garching, Germany, e-mail: m.tepegjzova@tum.de

Jing Zhou: ORStat and Leuven Statistics Research Centre, KU Leuven, Naamsestraat 69-box 3555, Leuven, Belgium; Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Krijgslaan 281-S9, Gent, Belgium, e-mail: jing.zhou@kuleuven.be, Jing.Zhou@UGent.be

Gerda Claeskens: ORStat and Leuven Statistics Research Centre, KU Leuven, Naamsestraat 69-box 3555, Leuven, Belgium, e-mail: gerda.claeskens@kuleuven.be

Claudia Czado: Department of Mathematics and Munich Data Science Institute, Technische Universität München, Boltzmannstraße 3, 85748, Garching, Germany, e-mail: cczado@ma.tum.de

weighted local linear fitting [65], by additive models [37,20], etc. The theoretical analysis of the above-mentioned extensions is based on imposing additional assumptions such as samples that are i.i.d. (see for example Belloni and Chernozhukov [7], Yu and Jones [65]), or that are generated by a known additive function (see for example Koenker [37,34]). Such assumptions, which guarantee the performance of the proposed methods for certain data structures, cause concerns in applications due to the uncertainty of the real-world data structures. Bernard and Czado [8] addressed other potential concerns such as quantile crossings and model-misspecification, when the dependence structure of the response variables and the predictors does not follow a Gaussian copula. Flexible models without assuming homoscedasticity, or a linear relationship between the response and the predictors are of interest. Recent research on dealing with this issue includes quantile forests [2,42,44] inspired by the earlier work of random forests [9] and modeling conditional quantiles using copulas (see also Chen *et al.* [13], Noh *et al.* [50,51]).

Vine copulas in the context of conditional quantile prediction have been investigated by Kraus and Czado [41] using drawable vine copulas (D-vines) and by Chang and Joe [11] and, most recently, Zhu *et al.* [67] using restricted regular vines (R-vines). The approach of Kraus and Czado [11] is based on first finding the locally optimal regular vine structure among all predictors and then adding the response to each selected tree in the vine structure as a leaf, as also followed by Bauer and Czado [5] in the context of non-Gaussian conditional independence testing. The procedure in Chang and Joe [11] allows for a recursive determination of the response quantiles, which is restricted through the prespecified dependence structure among predictors. The latter might not be the one maximizing the conditional response likelihood, which is the main focus in regression setup. The approach of Kraus and Czado [41] is based on optimizing the conditional log-likelihood and selecting predictors sequentially until no improvement of the conditional log-likelihood is achieved. This approach based on the conditional response likelihood is more appropriate to determine the associated response quantiles. Furthermore, the intensive simulation study in Kraus and Czado [41] showed the superior performance of the D-vine copulas-based quantile regression compared to various quantile regression methods, i.e. linear quantile regression [38], boosting additive quantile regression [35,37,20], nonparametric quantile regression [43], and semiparametric quantile regression [51]. In parallel to our work, Zhu *et al.* [67] proposed an extension of this D-vine-based forward regression to a restricted R-vine forward regression with comparable performance to the D-vine regression. Thus, the D-vine quantile regression will be our benchmark model.

We extend the method of Kraus and Czado [41] in two ways: (1) our approach is applicable to both C-vine and D-vine copulas; (2) a two-step ahead construction is introduced, instead of the one-step ahead construction. Since the two-step ahead construction is the main difference between our method and Kraus and Czado [41], we further explain the second point in more detail. Our proposed method proceeds by adding predictors to the model sequentially. However, in contrast to Kraus and Czado [41] with only one variable ahead, our new approach proposes to look up two variables ahead for selecting the variable to be added in each step. The general idea of this two-step ahead algorithm is as follows: in each step of the algorithm, we study combinations of two variables to find the variable, which in combination with the other improves the conditional log-likelihood the most. Thus, in combination with a forward selection method, this two-step ahead algorithm allows us to construct nonparametric quantile estimators that improve the conditional log-likelihood in each step and, most importantly, take possible future improvements into account. Our method is applicable to both low and high-dimensional data. By construction, quantile crossings are avoided. All marginal densities and copulas are estimated nonparametrically, allowing more flexibility than parametric specifications. Kraus and Czado [41] addressed the necessity and possible benefit of the nonparametric estimation of bivariate copulas in the quantile regression framework. This construction permits a large variety of dependence structures, resulting in a well-performing conditional quantile estimator. Moreover, extending to the C-vine copula class, in addition to the D-vine copulas, provides greater flexibility.

The article is organized as follows. Section 2 introduces the general setup, the concept of C-vine and D-vine copulas, and the nonparametric approach for estimating copula densities. Section 3 describes the vine-based approach for quantile regression. The new two-step ahead forward selection algorithms are described in Section 4. We investigate in Proposition 4.1 the consistency of the conditional quantile

estimator for given variable orders. The finite sample performance of the vine-based conditional quantile estimator is evaluated in Section 5 by several quantile-related measurements in various simulation settings. We apply the newly introduced algorithms to low- and high-dimensional real data in Section 6. In Section 7, we conclude and discuss possible directions of future research.

2 Theoretical background

Consider the random vector $\mathbf{X} = (X_1, \dots, X_d)^T$ with observed values $\mathbf{x} = (x_1, \dots, x_d)^T$, joint distribution and density function F and f , marginal distribution and density functions F_{X_j} and f_{X_j} for X_j , $j = 1, \dots, d$. Sklar's theorem [57] allows us to represent any multivariate distribution in terms of its marginals F_{X_j} and a copula C encoding the dependence structure. In the continuous case, C is unique and satisfies $F(\mathbf{x}) = C(F_{X_1}(x_1), \dots, F_{X_d}(x_d))$ and $f(\mathbf{x}) = c(F_{X_1}(x_1), \dots, F_{X_d}(x_d))[\prod_{j=1}^d f_{X_j}(x_j)]$, where c is the density function of the copula C . To characterize the dependence structure of \mathbf{X} , we transform each X_j to a uniform variable U_j by applying the probability integral transform, i.e. $U_j := F_{X_j}(X_j)$, $j = 1, \dots, d$. Then the random vector $\mathbf{U} = (U_1, \dots, U_d)^T$ with observed values $(u_1, \dots, u_d)^T$ has a copula as a joint distribution denoted as C_{U_1, \dots, U_d} with associated copula density function c_{U_1, \dots, U_d} . While the catalogue of bivariate parametric copula families is large, this is not true for $d > 2$. Therefore, conditioning was applied to construct multivariate copulas using only bivariate copulas as building blocks. Joe [28] gave the first pair copula construction for d dimensions in terms of distribution functions, whereas Bedford and Cooke [6] independently developed constructions in terms of densities together with a graphical building plan, called a regular vine tree structure. It consists of a set of linked trees T_1, \dots, T_d (edges in tree T_j become nodes in tree T_{j+1}) satisfying a proximity condition, which allows us to identify all possible constructions. Each edge of the trees is associated with a pair copula $C_{U_i, U_j | U_D}$, where D is a subset of indices not containing i, j . In this case, the set $\{i, j\}$ is called the conditioned set, while D is the conditioning set. A joint density using the class of vine copulas is then the product of all pair copulas identified by the tree structure evaluated at appropriate conditional distribution functions $F_{X_j | X_D}$ and the product of the marginal densities f_{X_j} , $j = 1, \dots, d$. A detailed treatment of vine copulas together with estimation methods and model choice approaches are given, for example in the studies of Joe [30] and Czado [17].

Since we are interested in simple copula-based estimation methods for conditional quantiles, we restrict to two subclasses of the regular vine tree structure, namely, the D- and C-vine structure. We show later that these structures allow us to express conditional distribution and quantiles in closed form. In the D-vine tree structure all trees are paths, i.e. all nodes have degree ≤ 2 . Nodes with degree 1 are called leaf nodes. A C-vine structure occurs, when all trees are stars with a root node in the centre. The right and left panels of Figure 1 illustrate a D-vine and a C-vine tree sequence in four dimensions, respectively.

For these sub classes we can easily give the corresponding vine density [17, Chapter 4]. For a D-vine density we have a permutation s_1, \dots, s_d of $1, \dots, d$ such that

$$f(x_1, \dots, x_d) = \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} C_{U_{s_i}, U_{s_{i+j}} | U_{s_{i+1}}, \dots, U_{s_{i+j-1}}} (F_{X_{s_i} | X_{s_{i+1}}, \dots, X_{s_{i+j-1}}} (x_{s_i} | x_{s_{i+1}}, \dots, x_{s_{i+j-1}}), \\ F_{X_{s_{i+j}} | X_{s_{i+1}}, \dots, X_{s_{i+j-1}}} (x_{s_{i+j}} | x_{s_{i+1}}, \dots, x_{s_{i+j-1}})) \cdot \prod_{k=1}^d f_{X_{s_k}} (x_{s_k}), \quad (1)$$

while for a C-vine density the following representation holds

$$f(x_1, \dots, x_d) = \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} C_{U_{s_j}, U_{s_{j+i}} | U_{s_1}, \dots, U_{s_{j-1}}} (F_{X_{s_j} | X_{s_1}, \dots, X_{s_{j-1}}} (x_{s_j} | x_{s_1}, \dots, x_{s_{j-1}}), \\ F_{X_{s_{j+i}} | X_{s_1}, \dots, X_{s_{j-1}}} (x_{s_{j+i}} | x_{s_1}, \dots, x_{s_{j-1}})) \cdot \prod_{k=1}^d f_{X_{s_k}} (x_{s_k}). \quad (2)$$

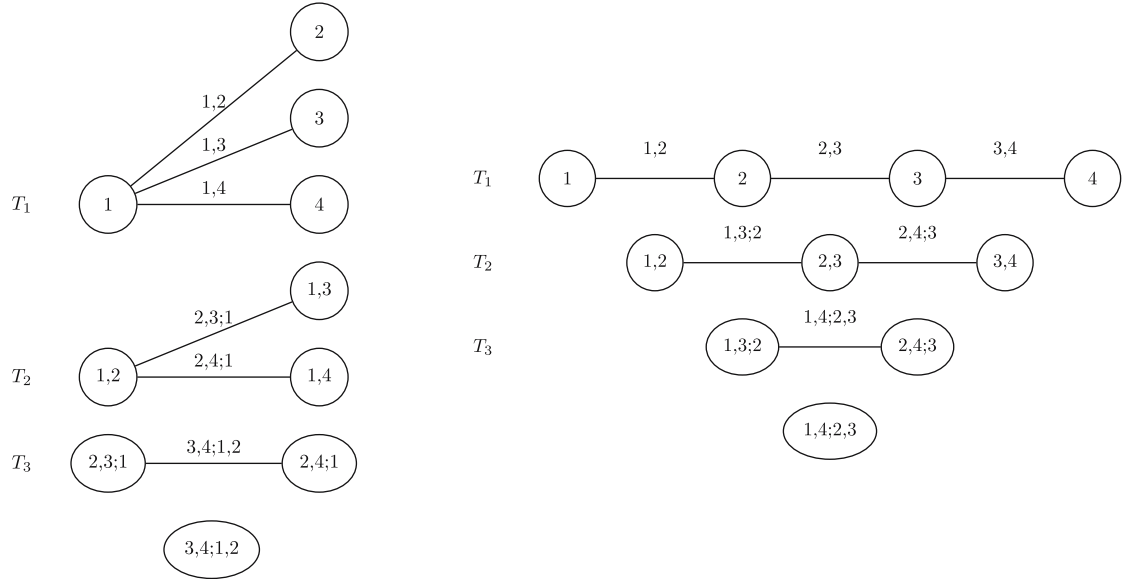


Figure 1: C-vine tree sequence (left panel) and a D-vine tree sequence (right panel) in four dimensions.

To determine the needed conditional distribution $F_{X_j|X_D}$ in Eqs. (1) and (2) for appropriate choices of j and D , the recursion discussed in Joe [28] is available. Using $u_j = F_{X_j|X_D}(x_j|x_D)$ for $j = 1, \dots, d$ we can express them as compositions of h-functions. These are defined in general as $h_{U_i|U_j;U_D}(u_i|u_j;u_D) = \frac{\partial}{\partial u_j} C_{U_i,U_j;U_D}(u_i, u_j;u_D)$. Additionally, we made in Eqs. (1) and (2) the simplifying assumption ([17], Section 5.4) that is, the copula function $C_{U_i,U_j;U_D}$ does not depend on the specific conditioning value of u_D , i.e. $C_{U_i,U_j;U_D}(u_i, u_j;u_D) = C_{U_i,U_j;U_D}(u_i, u_j)$. The dependence on u_D in Eqs. (1) and (2) is completely captured by the arguments of the pair copulas. This assumption is often made for tractability reasons in higher dimensions (Haff et al. [27] and Stoeber et al. [58]). It implies further that the h-function satisfies $h_{U_i|U_j;U_D}(u_i|u_j;u_D) = \frac{\partial}{\partial u_j} C_{U_i,U_j;U_D}(u_i, u_j) = C_{U_i|U_j;U_D}(u_i|u_j)$ and is independent of u_D .

2.1 Nonparametric estimators of the copula densities and h-functions

There are many methods to estimate a bivariate copula density c_{U_i,U_j} nonparametrically. Examples are the transformation estimator [12], the transformation local likelihood estimator [22], the tapered transformation estimator [61], the beta kernel estimator [12], and the mirror-reflection estimator [24]. Among the above-mentioned kernel estimators, the transformation local likelihood estimator [22] was found by Nagler et al. [47] to have an overall best performance. The estimator is implemented in the R packages `kdecopula` [45] and `rvinecopulib` [49] using Gaussian kernels. We review its construction in Appendix A. To satisfy the copula definition, it is scaled to have uniform margins.

As mentioned above the simplifying assumption implies that $h_{U_i|U_j;U_D}(u_i|u_j;u_D)$ is independent of specific values of u_D . Thus, it is sufficient to show how the h-function $h_{U_i|U_j} = C_{U_i|U_j}(u_i|u_j)$ can be estimated nonparametrically. For this, we use as estimator

$$\hat{C}_{U_i|U_j}(u_i|u_j) = \int_0^{u_j} \hat{c}_{U_i,U_j}(\tilde{u}_i, u_j) d\tilde{u}_i,$$

where \hat{c}_{U_i,U_j} is one of the aforementioned nonparametric estimators of the bivariate copula density of (U_i, U_j) for which it holds that \hat{c}_{U_i,U_j} integrates to 1 and has uniform margins.

3 Vine-based quantile regression

In the general regression framework, the predictive ability of a set of variables $\mathbf{X} = (X_1, \dots, X_p)^T$ for the response $Y \in \mathbb{R}$ is studied. The main interest of vine-based quantile regression is to predict the $\alpha \in (0, 1)$ quantile $q_\alpha(x_1, \dots, x_p) = F_{Y|X_1, \dots, X_p}^{-1}(\alpha|x_1, \dots, x_p)$ of the response variable Y given \mathbf{X} by using a copula-based model of $(Y, \mathbf{X})^T$. As shown in Kraus and Czado [41], this can be expressed as

$$F_{Y|X_1, \dots, X_p}^{-1}(\alpha|x_1, \dots, x_p) = F_Y^{-1}\left(C_{V|U_1, \dots, U_p}^{-1}\left(\alpha|F_{X_1}(x_1), \dots, F_{X_p}(x_p)\right)\right), \quad (3)$$

where $C_{V|U_1, \dots, U_p}$ is the conditional distribution function of $V = F_Y(Y)$ given $U_j = F_{X_j}(X_j) = u_j$ for $j = 1, \dots, p$, with corresponding density $c_{V|U_1, \dots, U_p}$, and C_{V, U_1, \dots, U_p} denotes the $(p + 1)$ -dimensional copula associated with the joint distribution of $(Y, \mathbf{X})^T$. In view of Section 1, we have $d = p + 1$. An estimate of $q_\alpha(x_1, \dots, x_p)$ can be obtained using estimated marginal quantile functions $\hat{F}_Y^{-1}, \hat{F}_{X_j}^{-1}$, $j = 1, \dots, p$, and the estimated conditional distribution function $\hat{C}_{V|U_1, \dots, U_p}^{-1}$ giving $\hat{q}_\alpha(x_1, \dots, x_p) = \hat{F}_Y^{-1}\left(\hat{C}_{V|U_1, \dots, U_p}^{-1}\left(\alpha|\hat{F}_{X_1}(x_1), \dots, \hat{F}_{X_p}(x_p)\right)\right)$.

In general, C_{V, U_1, \dots, U_p} can be any $(p + 1)$ -dimensional multivariate copula, however, for certain vine structures the corresponding conditional distribution function $C_{V|U_1, \dots, U_p}$ can be obtained in closed form not requiring numerical integration. For D-vine structures this is possible and has been already used in the studies of Kraus and Czado [41]. Tepegjuzova [59] showed that this is also the case for certain C-vine structures. More precisely, the copula C_{V, U_1, \dots, U_p} with D-vine structure allows us to express $C_{V|U_1, \dots, U_p}$ in a closed form if and only if the response V is a leaf node in the first tree of the tree sequence. For a C-vine structure we need, that the node containing the response variable V in the conditioned set is not a root node in any tree. Additional flexibility in using such D- and C-vine structures is achieved by allowing for nonparametric pair-copulas as building blocks.

The order of the predictors within the tree sequences itself is a free parameter with direct impact on the target function $C_{V|U_1, \dots, U_p}$ and thus, on the corresponding prediction performance of $q_\alpha(x_1, \dots, x_p)$. For this, we recall the concept of a node order for C- and D-vine copulas introduced in the study by Tepegjuzova [59]. A D-vine copula denoted by C_D has order $O_D(C_D) = (V, U_{i_1}, \dots, U_{i_p})$, if the response V is the first node of the first tree T_1 and U_{i_k} is the $(k + 1)$ th node of T_1 , for $k = 1, \dots, p$. A C-vine copula C_C has order $O_C(C_C) = (V, U_{i_1}, \dots, U_{i_p})$, if U_{i_1} is the root node in the first tree T_1 , $U_{i_2}U_{i_1}$ is the root node in the second tree T_2 , and $U_{i_k}U_{i_{k-1}}; U_{i_1}, \dots, U_{i_{k-2}}$ is the root node in the k th tree T_k for $k = 3, \dots, p - 1$.

Now our goal is to find an optimal order of D- or C-vine copula model with regard to a fit measure. This measure has to allow us to quantify the explanatory power of a model. One such measure is the estimated conditional copula log-likelihood function as a fit measure. For N i.i.d. observations $\mathbf{v} := (v^{(1)}, \dots, v^{(N)})^T$ and $\mathbf{u}_j := (u_j^{(1)}, \dots, u_j^{(N)})^T$, for $j = 1, \dots, p$ of the random vector $(V, U_1, \dots, U_p)^T$ we fit a C- or D-vine copula with order $O(\hat{C}) = (V, U_1, \dots, U_p)$ using nonparametric pair copulas. We denote this copula by \hat{C} , then the fitted conditional log-likelihood can be determined as

$$\begin{aligned} cll(\hat{C}, \mathbf{v}, (\mathbf{u}_1, \dots, \mathbf{u}_p)) &= \sum_{n=1}^N \ln \hat{c}_{V|U_1, \dots, U_p}(v^{(n)}|u_1^{(n)}, \dots, u_p^{(n)}) \\ &= \sum_{n=1}^N \left[\ln \hat{c}_{V, U_1}(v^{(n)}, u_1^{(n)}) + \sum_{j=2}^p \ln \hat{c}_{V, U_j|U_1, \dots, U_{j-1}}\left(\hat{C}_{V|U_1, \dots, U_{j-1}}(v^{(n)}|u_1^{(n)}, \dots, \right. \right. \\ &\quad \left. \left. u_{j-1}^{(n)}), \hat{C}_{U_j|U_1, \dots, U_{j-1}}(u_j^{(n)}|u_1^{(n)}, \dots, u_{j-1}^{(n)})\right) \right]. \end{aligned}$$

Penalizations for model complexity when parametric pair copulas are used can be added as shown in the study by Tepegjuzova [59]. To define an appropriate penalty in the case of using nonparametric pair copulas is an open research question (see also Section 7).

4 Forward selection algorithms

Having a set of p predictors, there are $p!$ different orders that uniquely determine $p!$ C-vines and $p!$ D-vines. Fitting and comparing all of them are computationally inefficient. Thus, the idea is to have an algorithm that will sequentially choose the elements of the order, so that at every step the resulting model for the prediction of the conditional quantiles has the highest conditional log-likelihood. Building upon the idea of Kraus and Czado [41] for the one-step ahead D-vine regression, we propose an algorithm which allows for more flexibility and which is less greedy, given the intention to obtain a globally optimal C- or D-vine fit. The algorithm builds the C- or D-vine step by step, starting with an order consisting of only the response variable V . Each step adds one of the predictors to the order based on the improvement of the conditional log-likelihood, while taking into account the possibility of future improvement, i.e. extending our view two steps ahead in the order. As discussed in Section 2.1, the pair copulas at each step are estimated nonparametrically in contrast to the parametric approach of Kraus and Czado [41]. We present the implementation for both C-vine and D-vine-based quantile regression in a single algorithm, in which the user decides whether to fit a C-vine or D-vine model based on the background knowledge of dependency structures in the data. Implementation for a large data set is computationally challenging; therefore, randomization is introduced to guarantee computational efficiency in high dimensions.

4.1 Two-step ahead forward selection algorithm for C- and D-vine-based quantile regression

Input and data preprocessing: Consider N i.i.d observations $\mathbf{y} := (y^{(1)}, \dots, y^{(N)})$ and $\mathbf{x}_j := (x_j^{(1)}, \dots, x_j^{(N)})$ for $j = 1, \dots, p$, from the random vector $(Y, X_1, \dots, X_p)^T$. The input data are on the x-scale, but in order to fit bivariate copulas we need to transform it to the u-scale using the probability integral transform. Since the marginal distributions are unknown we estimate them, i.e. F_Y and F_{X_j} , for $j = 1, \dots, p$, are estimated using a univariate nonparametric kernel density estimator with the R package `kde1d` [48]. This results in the pseudo copula data $\hat{v}^{(n)} := \hat{F}_Y(y^{(n)})$ and $\hat{u}_j^{(n)} := \hat{F}_{X_j}(x_j^{(n)})$, for $n = 1, \dots, N$, $j = 1, \dots, p$. The normalized marginals (z-scale) are defined as $Z_j := \Phi^{-1}(U_j)$ for $j = 1, \dots, p$, and $Z_V := \Phi^{-1}(V)$, where Φ denotes the standard normal distribution function.

Step 1: To reduce computational complexity, we perform a pre-selection of the predictors based on Kendall's τ . This is motivated by the fact that Kendall's τ is rank-based, therefore, invariant with respect to monotone transformations of the marginals and can be expressed in terms of pair copulas. Using the pseudo copula data $(\hat{\mathbf{v}}, \hat{\mathbf{u}}_j) = \{\hat{v}^{(n)}, \hat{u}_j^{(n)} | n = 1, \dots, N\}$, estimates $\hat{\tau}_{VU_j}$ of the Kendall's τ values between the response V , and all possible predictors U_j for $j = 1, \dots, p$, are obtained. For a given $k \leq p$, the k largest estimates of $|\hat{\tau}_{VU_j}|$ are selected and the corresponding indices q_1, \dots, q_k are identified such that $|\hat{\tau}_{VU_{q_1}}| \geq |\hat{\tau}_{VU_{q_2}}| \geq \dots \geq |\hat{\tau}_{VU_{q_k}}| \geq |\hat{\tau}_{VU_{q_{k+1}}}| \geq \dots \geq |\hat{\tau}_{VU_{q_p}}|$. The parameter k is a hyper-parameter and therefore, subject to tuning. To obtain a parsimonious model, we suggest a k corresponding to 5–20% of the total number of predictors. The k candidate predictors and the corresponding candidate index set of step 1 are defined as U_{q_1}, \dots, U_{q_k} and $K_1 = \{q_1, \dots, q_k\}$, respectively. For all $c \in K_1$ and $j \in \{1, \dots, p\} \setminus \{c\}$ the candidate two-step ahead C- or D-vine copulas are defined as the three-dimensional copulas $C_{c,j}^1$ with order $\mathcal{O}(C_{c,j}^1) = (V, U_c, U_j)$. The first predictor is added to the order based on the conditional log-likelihood of the candidate two-step ahead C- or D-vine copulas, $C_{c,j}^1$, given as

$$cll(C_{c,j}^1, \hat{\mathbf{v}}, (\hat{\mathbf{u}}_c, \hat{\mathbf{u}}_j)) = \sum_{n=1}^N \left[\log \hat{c}_{V, U_c}(\hat{v}^{(n)}, \hat{u}_c^{(n)}) + \log \hat{c}_{V, U_j | U_c}(\hat{h}_{V | U_c}(\hat{v}^{(n)} | \hat{u}_c^{(n)}), \hat{h}_{U_j | U_c}(\hat{u}_j^{(n)} | \hat{u}_c^{(n)})) \right].$$

For each candidate predictor U_c , the maximal two-step ahead conditional log-likelihood at step 1, cll_c^1 , is defined as $cll_c^1 := \max_{j \in \{1, \dots, p\} \setminus \{c\}} cll(C_{c,j}^1, \hat{\mathbf{v}}, (\hat{\mathbf{u}}_c, \hat{\mathbf{u}}_j))$, $\forall c \in K_1$. Finally, based on the maximal two-step ahead

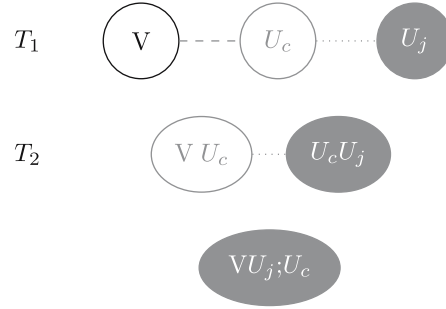


Figure 2: V is fixed as the first node of T_1 and the first candidate predictor to be included in the model, U_c (grey), is chosen based on the conditional log-likelihood of the two-step ahead copula including the predictor U_j (grey filled).

conditional log-likelihood at step 1, cll_c^1 , the index t_1 is chosen as $t_1 := \arg \max_{c \in K_1} cll_c^1$, and the corresponding candidate predictor U_{t_1} is selected as the first predictor added to the order. An illustration of the vine tree structure of the candidate two-step ahead copulas $C_{c,j}^1$, in the case of fitting a D-vine model, with order $O_D(C_{c,j}^1) = (V, U_c, U_j)$ is given in Figure 2. Finally, the current optimal fit after the first step is the C-vine or D-vine copula, C_1 with order $O(C_1) = (V, U_{t_1})$.

Step r : After $r - 1$ steps, the current optimal fit is the C- or D-vine copula C_{r-1} with order $O(C_{r-1}) = (V, U_{t_1}, \dots, U_{t_{r-1}})$. At each previous step i , the order of the current optimal fit is sequentially updated with the predictor U_{t_i} for $i = 1, \dots, r - 1$. At the r th step, the next predictor candidate is to be included. To do so, the set of potential candidates is narrowed based on a partial correlation measure. Defining a partial Kendall's τ is not straightforward and requires the notion of a partial copula, which is the average over the conditional copula given the values of the conditioning values (e.g. see Gijbels and Matherne [23] and references therein). In addition, the computation in the case of multivariate conditioning is very demanding and still an open research problem. Therefore, we took a pragmatic view and base our candidate selection on partial correlation. Due to the assumption of Gaussian margins inherited to the Pearson's partial correlation, the estimates are computed on the z -scale. Estimates of the empirical Pearson's partial correlation, $\hat{\rho}_{Z_V, Z_j; Z_{t_1}, \dots, Z_{t_{r-1}}}$, between the normalized response variable V and available predictors U_j for $j \in \{1, 2, \dots, p\} \setminus \{t_1, \dots, t_{r-1}\}$ are obtained. Similar to the first step, a set of candidate predictors of size k is selected based on the largest values of $|\hat{\rho}_{Z_V, Z_j; Z_{t_1}, \dots, Z_{t_{r-1}}}|$ and the corresponding indices q_1, \dots, q_k . The k candidate predictors and the corresponding candidate index set of step r are defined as U_{q_1}, \dots, U_{q_k} and the set $K_r = \{q_1, \dots, q_k\}$, respectively. For all $c \in K_r$ and $j \in \{1, 2, \dots, p\} \setminus \{t_1, \dots, t_{r-1}, c\}$, the candidate two-step ahead C- or D-vine copulas are defined as the copulas $C_{c,j}^r$ with order $O(C_{c,j}^r) = (V, U_{t_1}, \dots, U_{t_{r-1}}, U_c, U_j)$. There are $k(p - r)$ different candidate two-step ahead C- or D-vine copulas $C_{c,j}^r$ (since we have k candidates for the one-step ahead extension U_c , and for each, $p - (r - 1) - 1$ two step ahead extensions U_j). Their corresponding conditional log-likelihood functions are given as

$$\begin{aligned}
& cll(C_{c,j}^r, \hat{\mathbf{v}}, (\hat{\mathbf{u}}_{t_1} \dots \hat{\mathbf{u}}_{t_{r-1}}, \hat{\mathbf{u}}_c, \hat{\mathbf{u}}_j)) \\
&= cll(C_{r-1}, \hat{\mathbf{v}}, (\hat{\mathbf{u}}_{t_1} \dots \hat{\mathbf{u}}_{t_{r-1}})) \\
&+ \sum_{n=1}^N \log \hat{c}_{V U_c; U_{t_1}, \dots, U_{t_{r-1}}}(\hat{C}_{V|U_{t_1}, \dots, U_{t_{r-1}}}(\hat{\mathbf{v}}^{(n)} | \hat{\mathbf{u}}_{t_1}^{(n)}, \dots, \hat{\mathbf{u}}_{t_{r-1}}^{(n)}), \hat{C}_{U_c|U_{t_1}, \dots, U_{t_{r-1}}}(\hat{\mathbf{u}}_c^{(n)} | \hat{\mathbf{u}}_{t_1}^{(n)}, \dots, \hat{\mathbf{u}}_{t_{r-1}}^{(n)})) \\
&+ \sum_{n=1}^N \log \hat{c}_{V U_j; U_{t_1}, \dots, U_{t_{r-1}}, U_c}(\hat{C}_{V|U_{t_1}, \dots, U_{t_{r-1}}, U_c}(\hat{\mathbf{v}}^{(n)} | \hat{\mathbf{u}}_{t_1}^{(n)}, \dots, \hat{\mathbf{u}}_{t_{r-1}}^{(n)}, \hat{\mathbf{u}}_c^{(n)}), \\
&\quad \hat{C}_{U_j|U_{t_1}, \dots, U_{t_{r-1}}, U_c}(\hat{\mathbf{u}}_j^{(n)} | \hat{\mathbf{u}}_{t_1}^{(n)}, \dots, \hat{\mathbf{u}}_{t_{r-1}}^{(n)}, \hat{\mathbf{u}}_c^{(n)})).
\end{aligned}$$

The r th predictor is then added to the order based on the maximal two-step ahead conditional log-likelihood at Step r , cll_c^r , defined as

$$cll_c^r := \max_{j \in \{1, 2, \dots, p\} \setminus \{t_1, \dots, t_{r-1}, c\}} cll(C_{c,j}^r, \hat{\mathbf{v}}, (\hat{\mathbf{u}}_{t_1} \dots \hat{\mathbf{u}}_{t_{r-1}}, \hat{\mathbf{u}}_c, \hat{\mathbf{u}}_j)), \quad \forall c \in K_r. \quad (4)$$

The index t_r is chosen as $t_r := \arg \max_{c \in K_r} cll_c^r$, and the predictor U_{t_r} is selected as the r th predictor of the order. An illustration of the vine tree structure of the candidate two-step ahead copulas $C_{c,j}^r$ for a D-vine model with order $O_D(C_{c,j}^r) = (V, U_{t_1}, \dots, U_{t_{r-1}}, U_c, U_j)$ is given in Figure 3. At this step, the current optimal fit is the C-vine or D-vine copula C_r , with order $O(C_r) = (V, U_{t_1}, \dots, U_{t_r})$. The iterative procedure is repeated until all predictors are included in the order of the C- or D-vine copula model.

4.1.1 Additional variable reduction in higher dimensions

The above search procedure requires calculating $p - r$ conditional log-likelihoods for each candidate predictor at a given step r . This leads to calculating a total of $(p - r)k$ conditional log-likelihoods, where k is the number of candidates. For p large, this procedure would cause a heavy computational burden. Hence, the idea is to reduce the number of conditional log-likelihoods calculated for each candidate predictor. This is achieved by reducing the size of the set, over which the maximal two-step ahead conditional log-likelihood cll_c^r in Eq. (4), is computed. Instead of over the set $\{1, 2, \dots, p\} \setminus \{t_1, \dots, t_{r-1}, c\}$, the maximum can be taken over an appropriate subset. This subset can then be chosen either based on the largest Pearson's partial correlations in absolute value denoted as $|\hat{\rho}_{Z_V, Z_j; Z_{t_1}, \dots, Z_{t_{r-1}}, Z_c}|$, by random selection, or a combination of the two. The selection method and the size of reduction are user-decided.

4.2 Consistency of the conditional quantile estimator

The conditional quantile function on the original scale in Eq. (3) requires the inverse of the marginal distribution function of Y . Following Kraus and Czado [41] and Noh *et al.* [50], the marginal cumulative distribution functions F_Y and F_{X_j} , $j = 1, \dots, p$, are estimated nonparametrically to reduce the bias caused by model misspecification. Examples of nonparametric estimators for the marginal distributions F_Y and F_{X_j} , are the continuous kernel smoothing estimator [52] and the transformed local likelihood estimator in the univariate case [21]. Using a Gaussian kernel, the above two estimators of the marginal distribution are uniformly strong consistent. When all inverses of the h-functions are also estimated nonparametrically, we

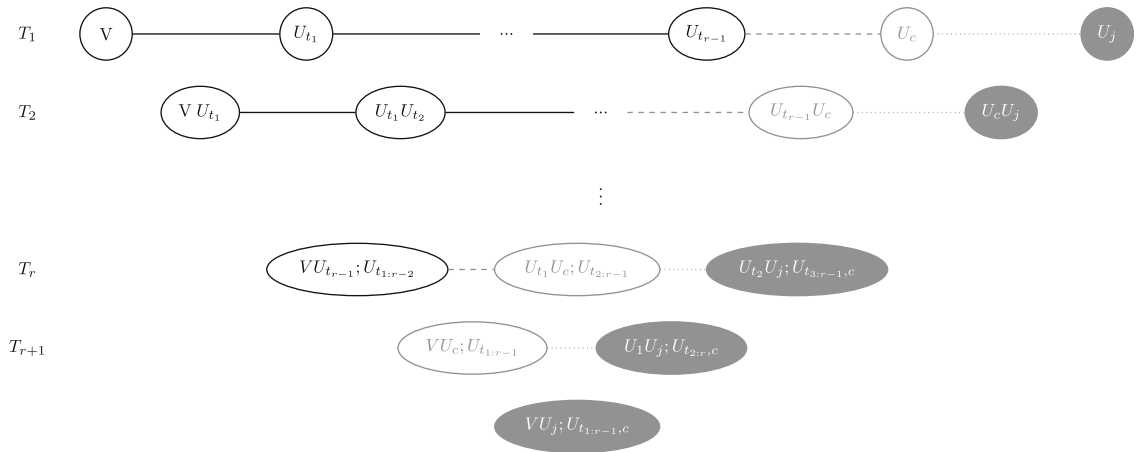


Figure 3: In step r , the current optimal fit, C_{r-1} (black), is extended by one more predictor, U_c (grey), to obtain the new current optimal fit C_r (black and grey), based on the conditional log-likelihood of the two-step ahead copula $C_{c,j}^r$ which also includes the predictor U_j (grey filled) (In the figure, we use the shortened notation $U_{t_1, \dots, t_{r-1}}$ instead of writing $U_{t_1}, \dots, U_{t_{r-1}}$ and we use $U_{t_1, r-1, c}$ instead of $U_{t_1}, \dots, U_{t_{r-1}}, U_c$).

establish the consistency of the conditional quantile estimator $\hat{F}_{Y|X_1, \dots, X_p}^{-1}$ in Proposition 4.1 for fixed variable orders. By showing the uniform consistency, Proposition 4.1 gives an indication of the performance of the conditional quantile estimator $\hat{F}_{Y|X_1, \dots, X_p}^{-1}$ for fixed variable orders, while combining the consistent estimators of F_Y , F_{X_j} , and bivariate copula densities. Under the consistency guarantee, the numerical performance of $\hat{F}_{Y|X_1, \dots, X_p}^{-1}$ investigated by extensive simulation studies is presented in Section 5.

Proposition 4.1. *Let the inverse of the marginal distribution functions F_Y and F_{X_j} , $j = 1, \dots, p$ be uniformly continuous and estimated nonparametrically, and let the inverse of the h-functions expressing the conditional quantile estimator $C_{V|U_1, \dots, U_p}^{-1}$ be uniformly continuous and estimated nonparametrically in the interior of the support of bivariate copulas, i.e. $[\delta, 1 - \delta]^2$, $\delta \rightarrow 0_+$.*

1. *If estimators of the inverse of marginal functions \hat{F}_Y^{-1} , $\hat{F}_{X_j}^{-1}$, $j = 1, \dots, p$, are uniformly strong consistent on the support $[\delta, 1 - \delta]$, $\delta \rightarrow 0_+$, and the estimators of the inverse of h-functions composing the conditional quantile estimator $C_{V|U_1, \dots, U_p}^{-1}$ are uniformly strong consistent, then the estimator $\hat{F}_{Y|X_1, \dots, X_p}^{-1}(\alpha|x_1, \dots, x_p)$ is also uniformly strong consistent.*
2. *If estimators of the inverse of marginal functions \hat{F}_Y^{-1} , $\hat{F}_{X_j}^{-1}$, $j = 1, \dots, p$, are at least weak consistent, and the estimators of the inverse of h-functions are also at least weak consistent, then the estimator $\hat{F}_{Y|X_1, \dots, X_p}^{-1}(\alpha|x_1, \dots, x_p)$ is weak consistent.*

For more details about uniform continuous functions see the studies by Bartle and Sherbert ([4] Section 5.4), Kolmogorov and Fomin ([39], p. 109, Def. 1). For a definition of strong uniform consistency or convergence with probability one, see the studies by Ryzin [54], Silverman [56], and Durrett ([19], p. 16), while for a definition for weak consistency or convergence in probability, see the studies by Durrett ([19], p. 53). The strong uniform consistency result in Proposition 1 requires additionally that all estimators of \hat{F}_Y^{-1} , $\hat{F}_{X_j}^{-1}$, for $j = 1, \dots, p$, are strong uniformly consistent on a truncated compact interval $[\delta, 1 - \delta]$, $\delta \rightarrow 0_+$. Although not directly used in the proof of Proposition 4.1 in Appendix B, the truncation is an essential condition for guaranteeing the strong uniform consistency of all estimators of the inverse of the marginal distributions (i.e. estimators of quantile functions), see the studies by Cheng [15,14], Van Keilegom and Veraverbeke [60].

5 Simulation study

The proposed two-step ahead forward selection algorithms for C- and D-vine-based quantile regression, from Section 4.1, are implemented in the statistical language R [53]. The D-vine one-step ahead algorithm is implemented in the R package `vinereg` [46]. In the simulation study from R [41], it is shown that the D-vine one-step ahead forward selection algorithm performs better or similar, compared to other state-of-the-art quantile methods, boosting additive quantile regression [20,36], nonparametric quantile regression [43], semi-parametric quantile regression [51], and linear quantile regression [38]. Thus, we use the one-step ahead algorithm as the benchmark competitive method in the simulation study. We set up the following simulation settings given below. Each setting is replicated for $R = 100$ times. In each simulation replication, we randomly generate N_{train} samples used for fitting the appropriate nonparametric vine-based quantile regression models. Additionally, another $N_{\text{eval}} = \frac{1}{2}N_{\text{train}}$ sample for settings (a)–(f) and $N_{\text{eval}} = N_{\text{train}}$ for settings (g) and (h) are generated for predicting conditional quantiles from the models. settings (a)–(f) are designed to test quantile prediction accuracy of nonparametric C- or D-vine quantile regression in cases where $p \leq N$; hence, we set $N_{\text{train}} = 1,000$ or 300. settings (g) and (h) test quantile prediction accuracy in cases where $p > N$; hence, we set $N_{\text{train}} = 100$.

(a) Simulation setting M5 from Kraus and Czado [41]:

$$Y = \sqrt{|2X_1 - X_2 + 0.5|} + (-0.5X_3 + 1)(0.1X_4^3) + \sigma\varepsilon,$$

with $\varepsilon \sim N(0, 1)$, $\sigma \in \{0.1, 1\}$, $(X_1, X_2, X_3, X_4)^T \sim N_4(0, \Sigma)$, and the (i, j) th component of the covariance matrix given as $(\Sigma)_{i,j} = 0.5^{|i-j|}$.

(b) $(Y, X_1, \dots, X_5)^T$ follows a mixture of two six-dimensional T copulas with degrees of freedom equal to 3 and mixture probabilities 0.3 and 0.7. Association matrices R_1, R_2 and marginal distributions are recorded in Table 1.

(c) Linear and heteroscedastic [11]: $Y = 5(X_1 + X_2 + X_3 + X_4) + 10(U_1 + U_2 + U_3 + U_4)\varepsilon$, where $(X_1, X_2, X_3, X_4)^T \sim N_4(0, \Sigma)$, $\Sigma_{i,j} = 0.5^{|i \neq j|}$, $\varepsilon \sim N_4(0, 0.5)$, and $U_j, j = 1, \dots, 4$ are obtained from the X_j 's by the probability integral transform.

(d) Nonlinear and heteroscedastic [11]: $Y = U_1U_2e^{1.8U_3U_4} + 0.5(U_1 + U_2 + U_3 + U_4)\varepsilon$, where $U_j, j = 1, \dots, 4$ are probability integral transformed from $N_4(0, \Sigma)$, $\Sigma_{i,j} = 0.5^{|i \neq j|}$, and $\varepsilon \sim N(0, 0.5)$.

(e) R-vine copula [17]: $(V, U_1, \dots, U_4)^T$ follows an R-vine distribution with pair copulas given in Table 2.

(f) D-vine copula [59]: $(V, U_1, \dots, U_5)^T$ follows a D-vine distribution with pair copulas given in Table 3.

(g) Similar to setting (a),

$$Y = \sqrt{|2X_1 - X_2 + 0.5|} + (-0.5X_3 + 1)(0.1X_4^3) + (X_5, \dots, X_{110})(0, \dots, 0)^T + \sigma\varepsilon,$$

where $(X_1, \dots, X_{110})^T \sim N_{110}(0, \Sigma)$ with the (i, j) th component of the covariance matrix $(\Sigma)_{i,j} = 0.5^{|i-j|}$, $\varepsilon \sim N(0, 1)$, and $\sigma \in \{0.1, 1\}$.

(h) Similar to (g), $Y = (X_1^3, \dots, X_{110}^3)\beta + \varepsilon$, where $(X_1, \dots, X_{110})^T \sim N_{110}(0, \Sigma_A)$ with the (i, j) th component of the covariance matrix $(\Sigma_A)_{i,j} = 0.8^{|i-j|}$, $(X_{11}, \dots, X_{110})^T \sim N_{100}(0, \Sigma_B)$ with $(\Sigma_B)_{i,j} = 0.4^{|i-j|}$. The first ten entries

Table 1: Association matrices of the multivariate t-copula and marginal distributions for setting (b)

$R_1 =$	$\begin{pmatrix} 1 & 0.6 & 0.5 & 0.6 & 0.7 & 0.1 \\ 0.6 & 1 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 & 0.5 \\ 0.6 & 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0.7 & 0.5 & 0.5 & 0.5 & 1 & 0.5 \\ 0.1 & 0.5 & 0.5 & 0.5 & 0.5 & 1 \end{pmatrix}$	$R_2 =$	$\begin{pmatrix} 1 & -0.3 & -0.5 & -0.4 & -0.5 & -0.1 \\ -0.3 & 1 & 0.5 & 0.5 & 0.5 & 0.5 \\ -0.5 & 0.5 & 1 & 0.5 & 0.5 & 0.5 \\ -0.4 & 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ -0.5 & 0.5 & 0.5 & 0.5 & 1 & 0.5 \\ -0.1 & 0.5 & 0.5 & 0.5 & 0.5 & 1 \end{pmatrix}$		
Y	X₁	X₂	X₃	X₄	X₅
$N(0, 1)$	t_4	$N(1, 4)$	t_4	$N(1, 4)$	t_4

Table 2: Pair copulas of the R-vine $C_{V, U_1, U_2, U_3, U_4}$, with their family parameter and Kendall's τ for setting (e)

Tree	Edge	Conditioned; Conditioning	Family	Parameter	Kendall's τ
1	1	$U_1, U_3;$	Gumbel	3.9	0.74
1	2	$U_2, U_3;$	Gauss	0.9	0.71
1	3	$V, U_3;$	Gauss	0.5	0.33
1	4	$V, U_4;$	Clayton	4.8	0.71
2	1	$V, U_1; U_3$	Gumbel(90)	6.5	-0.85
2	2	$V, U_2; U_3$	Gumbel(90)	2.6	-0.62
2	3	$U_3, U_4; V$	Gumbel	1.9	0.48
3	1	$U_1, U_2; V, U_3$	Clayton	0.9	0.31
3	2	$U_2, U_4; V, U_3$	Clayton(90)	5.1	-0.72
4	1	$U_1, U_4; V, U_2, U_3$	Gauss	0.2	0.13

Table 3: Pair copulas of the D-vine $C_{V,U_1,U_2,U_3,U_4,U_5}$, with their family parameter and Kendall's τ for setting (f)

Tree	Edge	Conditioned; Conditioning	Family	Parameter	Kendall's τ
1	1	$V, U_1;$	Clayton	3.00	0.60
1	2	$U_1, U_2;$	Joe	8.77	0.80
1	3	$U_2, U_3;$	Gumbel	2.00	0.50
1	4	$U_3, U_4;$	Gauss	0.20	0.13
1	5	$U_4, U_5;$	Indep.	0.00	0.00
2	1	V, U_2, U_1	Gumbel	5.00	0.80
2	2	$U_1, U_3; U_2$	Frank	9.44	0.65
2	3	$U_2, U_4; U_3$	Joe	2.78	0.49
2	4	$U_3, U_5; U_4$	Gauss	0.20	0.13
3	1	$V, U_3; U_1, U_2$	Joe	3.83	0.60
3	2	$U_1, U_4; U_2, U_3$	Frank	6.73	0.55
3	3	$U_2, U_5; U_3, U_4$	Gauss	0.29	0.19
4	1	$V, U_4; U_1, U_2, U_3$	Clayton	2.00	0.50
4	2	$U_1, U_5; U_2, U_3, U_4$	Gauss	0.09	0.06
5	1	$V, U_5; U_1, U_2, U_3, U_4$	Indep.	0.00	0.00

of β are a descending sequence between (2, 1.1) with increment of 0.1, respectively, and the rest are equal to 0. We assume $\varepsilon \sim N(0, \sigma)$ and $\sigma \in \{0.1, 1\}$.

Since the true regression quantiles are difficult to obtain in most settings, we consider the averaged check loss [41,40] and the interval score [11,25], instead of the out-of-sample mean averaged square error in Kraus and Czado [41], to evaluate the performance of the estimation methods. For a chosen $\alpha \in (0, 1)$, the averaged check loss is defined as

$$\widehat{\text{CL}}_\alpha = \frac{1}{R} \sum_{r=1}^R \left\{ \frac{1}{N_{\text{eval}}} \sum_{n=1}^{N_{\text{eval}}} \{ \gamma_\alpha(Y_{r,n}^{\text{eval}} - \hat{q}_\alpha(X_{r,n}^{\text{eval}})) \} \right\}, \quad (5)$$

where γ_α is the check loss function.

The interval score, for the $(1 - \alpha) \times 100\%$ prediction interval, is defined as

$$\begin{aligned} \widehat{\text{IS}}_\alpha = & \frac{1}{R} \sum_{r=1}^R \left\{ \frac{1}{N_{\text{eval}}} \sum_{n=1}^{N_{\text{eval}}} \left\{ (\hat{q}_{\alpha/2}(X_{r,n}^{\text{eval}}) - \hat{q}_{1-\alpha/2}(X_{r,n}^{\text{eval}})) \right. \right. \\ & + \frac{2}{\alpha} (\hat{q}_{1-\alpha/2}(X_{r,n}^{\text{eval}}) - Y_{r,n}^{\text{eval}}) I\{Y_{r,n}^{\text{eval}} \leq \hat{q}_{1-\alpha/2}(X_{r,n}^{\text{eval}})\} \\ & \left. \left. + \frac{2}{\alpha} (Y_{r,n}^{\text{eval}} - \hat{q}_{\alpha/2}(X_{r,n}^{\text{eval}})) I\{Y_{r,n}^{\text{eval}} > \hat{q}_{\alpha/2}(X_{r,n}^{\text{eval}})\} \right\} \right\}, \quad (6) \end{aligned}$$

and smaller interval scores are better.

For settings (a)–(f), the estimation procedure for the two-step ahead C- or D-vine quantile regression follows exactly Section 4.1 where the candidate sets at each step include all possible remaining predictors. The additional variable reduction described in Section 4.1.1 is not applied; thus, we calculate all possible conditional log-likelihoods in each step. On the contrary, due to computational burden in settings (g) and (h), we set the number of candidates to be $k = 5$ and the additional variable reduction from Section 4.1.1 is applied. The chosen subset contains 20% of all possible choices, where 10% are predictors having the highest Pearson's partial correlation with the response and the remaining 10% are chosen randomly from the remaining predictors. Performance of the C- and D-vine two-step ahead quantile regression is compared with the C- and D-vine one-step ahead quantile regression. The performance of the competitive methods, evaluated by the averaged check loss at 5, 50, and 95% quantile levels and interval score for the 95%

prediction interval, is recorded in Tables 4 and 5. All densities are estimated nonparametrically for a fair comparison. Table 4 shows that the C- and D-vine two-step ahead regression models outperform the C- and D-vine one-step ahead regression models in five out of seven settings, except settings (b) and (e), in which all models perform quite similar to each other. Again, when comparing regression models within the same vine copula class, the C-vine two-step ahead regression models outperform the C-vine one-step ahead models in five out of seven settings. Similarly, the D-vine two-step ahead models outperform the D-vine one-step ahead models in six out of seven scenarios, except setting (b). In scenarios where there is no significant improvement through the second step, both one-step and two-step ahead approaches perform very similar. All of that implies that the two-step ahead vine-based quantile regression greatly improves the performance of the one-step ahead quantile regression. Table 5 indicates that in the high-dimensional settings, where the two-step ahead quantile regression was used in combination with the additional variable selection from Section 4.1.1, in three out of four simulation settings, the two-step ahead models outperform the one-step ahead models. In setting (g), we can see that all models show similar performance. In setting (g) with standard deviation $\sigma = 0.1$, the D-vine one-step ahead model outperforms the other models, while in setting (g) with $\sigma = 1$, the D-vine two-step ahead model shows a better performance. In setting (h), we see a significant improvement in the two-step ahead models compared to the one-step ahead models. For both $\sigma = 0.1$ and $\sigma = 1$, the best performing model is the C-vine two-step ahead model. These results indicate that the newly proposed method improves the accuracy of the one-step ahead quantile

Table 4: Out-of-sample predictions $\widehat{IS}_{0.05}$, $\widehat{CL}_{0.05}$, $\widehat{CL}_{0.5}$, $\widehat{CL}_{0.95}$ for settings (a)–(f) with $N_{\text{train}} = 300$ and $N_{\text{train}} = 1,000$

Setting	Model	$\widehat{IS}_{0.05}$	$\widehat{CL}_{0.05}$	$\widehat{CL}_{0.5}$	$\widehat{CL}_{0.95}$	$\widehat{IS}_{0.05}$	$\widehat{CL}_{0.05}$	$\widehat{CL}_{0.5}$	$\widehat{CL}_{0.95}$
		$N_{\text{train}} = 300$				$N_{\text{train}} = 1,000$			
$\sigma = 0.1$	(a) D-vine One-step	55.54	0.66	0.16	0.51	55.89	0.67	0.15	0.50
	D-vine Two-step	43.33	0.47	0.10	0.41	40.74	0.45	0.09	0.37
	** C-vine One-step	53.51	0.64	0.16	0.49	54.52	0.66	0.15	0.49
	C-vine Two-step	42.01	0.45	0.10	0.40	40.04	0.44	0.09	0.37
$\sigma = 1$	(a) D-vine One-step	154.35	1.63	0.45	1.62	162.12	1.70	0.43	1.66
	D-vine Two-step	148.53	1.57	0.45	1.56	156.77	1.63	0.42	1.62
	** C-vine One-step	151.60	1.61	0.45	1.60	160.78	1.68	0.43	1.65
	C-vine Two-step	148.41	1.56	0.45	1.56	156.79	1.63	0.42	1.62
(b)	D-vine One-step	118.75	1.29	0.42	1.30	125.33	1.37	0.40	1.36
	* D-vine Two-step	119.10	1.30	0.42	1.30	125.24	1.36	0.40	1.36
	C-vine One-step	119.08	1.30	0.41	1.30	125.12	1.36	0.40	1.36
	C-vine Two-step	118.90	1.30	0.42	1.30	125.30	1.36	0.40	1.36
(c)	D-vine One-step	2908.90	30.54	8.55	30.42	3064.78	31.69	8.15	31.47
	** D-vine Two-step	2853.52	30.21	8.70	29.95	3041.95	31.61	8.20	31.26
	C-vine One-step	2859.23	30.24	8.59	29.95	3046.52	31.64	8.18	31.25
	C-vine Two-step	2850.10	30.19	8.64	29.84	3042.46	31.62	8.20	31.23
(d)	D-vine One-step	86.40	0.92	0.24	0.91	91.11	0.96	0.22	0.95
	** D-vine Two-step	83.54	0.90	0.24	0.88	89.56	0.96	0.22	0.92
	C-vine One-step	84.99	0.91	0.24	0.90	90.40	0.96	0.22	0.94
	C-vine Two-step	83.33	0.90	0.24	0.87	89.47	0.96	0.22	0.92
(e)	D-vine One-step	10.59	0.11	0.03	0.11	10.49	0.11	0.03	0.11
	* D-vine Two-step	10.32	0.10	0.03	0.11	10.26	0.09	0.02	0.11
	C-vine One-step	10.23	0.11	0.03	0.10	10.02	0.10	0.02	0.10
	C-vine Two-step	10.35	0.10	0.03	0.11	10.33	0.10	0.02	0.11
(f)	D-vine One-step	13.79	0.16	0.04	0.14	13.70	0.16	0.04	0.14
	** D-vine Two-step	8.44	0.09	0.02	0.08	8.28	0.09	0.02	0.08
	C-vine One-step	12.62	0.14	0.04	0.13	12.23	0.13	0.04	0.13
	C-vine Two-step	9.09	0.10	0.02	0.09	8.93	0.09	0.02	0.08

Lower values, indicating better performance, are highlighted in bold. With ** we denote the scenarios in which there is an improvement through the second step and with * we denote scenarios in which the models perform similar.

Table 5: Out-of-sample predictions $\widehat{IS}_{0.5}$, $\widehat{CL}_{0.05}$, $\widehat{CL}_{0.5}$, $\widehat{CL}_{0.95}$ for settings (g)–(h) with $N_{\text{train}} = 100$

Model	$\widehat{IS}_{0.05}$	$\widehat{CL}_{0.05}$	$\widehat{CL}_{0.5}$	$\widehat{CL}_{0.95}$	$\widehat{IS}_{0.05}$	$\widehat{CL}_{0.05}$	$\widehat{CL}_{0.5}$	$\widehat{CL}_{0.95}$
	(g), $\sigma = 0.1^*$				(g), $\sigma = 1^{**}$			
D-vine One-step	19.63	0.26	0.25	0.23	53.38	0.69	0.67	0.65
D-vine Two-step	20.48	0.26	0.26	0.25	52.17	0.68	0.65	0.63
C-vine One-step	19.73	0.25	0.25	0.24	53.62	0.69	0.67	0.65
C-vine Two-step	19.79	0.25	0.25	0.25	52.35	0.67	0.65	0.64
	(h), $\sigma = 0.1^{**}$				(h), $\sigma = 1^{**}$			
D-vine One-step	558.36	6.92	6.98	7.04	554.18	6.87	6.93	6.99
D-vine Two-step	529.51	6.46	6.62	6.78	531.30	6.64	6.64	6.64
C-vine One-step	514.08	6.05	6.43	6.81	512.96	6.39	6.41	6.44
C-vine Two-step	479.66	5.87	6.00	6.12	483.92	6.05	6.05	6.05

Lower values, indicating better performance, are highlighted in bold. With ** we denote the scenarios in which there is an improvement through the second step and with * we denote scenarios in which the models perform similar.

regression in high dimensions, even with an attempt to ease the computational complexity of the two-step ahead model with a low number of candidates, compared to the number of predictors.

The proposed two-step algorithms, as compared to the one-step algorithms, are computationally more intensive. We present the averaged computation time over $R = 100$ replications on 100 paralleled cores (Xeon Gold 6140 CPUs@2.6 GHz) in settings (g) and (h) where $p > N_{\text{train}}$, for the one step ahead and the two-step ahead approach. The high-dimensional settings have similar computational times since the computational intensity depends on the number of pair copula estimations and the number of candidates, which are the same for settings (g) and (h). Hence, we only report the averaged computational times for settings (g) and (h). The average computation time in minutes for the one-step ahead (C- and D-vine) approach is 83.01, in contrast to 200.28 by the two-step ahead (C- and D-vine) approach. With the variable reduction from Section 4.1.1, the two-step algorithms double the time consumption of the one-step algorithms in exchange for prediction accuracy.

6 Real data examples

We test the proposed methods on two real data sets, i.e. the Concrete data set from Yeh [64] corresponding to $p \leq N$, and the riboflavin data set from Bühlmann and van de Geer [10] corresponding to $p > N$. For both, performance of the four competitive algorithms is evaluated by the averaged check loss defined in Eq. (5) at 5, 50, and 95% quantile levels, and the 95% prediction interval score defined in Eq. (6), by randomly splitting the data set into training and evaluation sets 100 times.

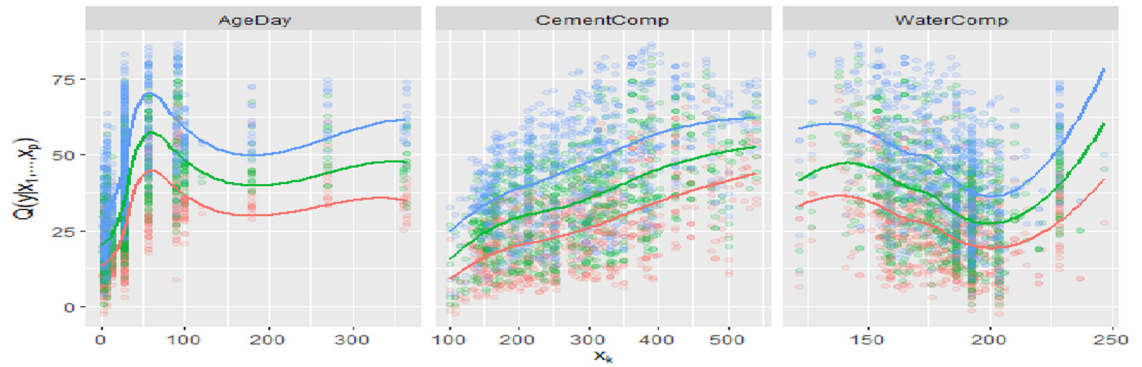
6.1 Concrete data set

The concrete data set was initially used in Yeh [64], and is available at the UCI Machine Learning Repository [18]. The data set has in total 1,030 samples. Our objective is quantile predictions of the concrete compressive strength, which is a highly nonlinear function of age and ingredients. The predictors are age (AgeDay, counted in days) and seven physical measurements of the concrete ingredients (given in kg in a m^3 mixture): cement (CementComp), blast furnace slag (BlastFur), fly ash (FlyAsh), water (WaterComp), superplasticizer, coarse aggregate (CoarseAggre), and fine aggregate (FineAggre). We randomly split the data set into a training set with 830 samples and an evaluation set with 200 samples; the random splitting is

Table 6: Concrete data set: out-of-sample predictions $\widehat{IS}_{0.5}$, $\widehat{CL}_{0.05}$, $\widehat{CL}_{0.5}$, $\widehat{CL}_{0.95}$

Model	$\widehat{IS}_{0.05}$	$\widehat{CL}_{0.05}$	$\widehat{CL}_{0.5}$	$\widehat{CL}_{0.95}$
D-vine One-step	1032.32	10.75	2.76	10.52
D-vine Two-step	987.10	10.54	2.78	9.82
C-vine One-step	976.75	10.65	2.70	9.45
C-vine Two-step	967.00	10.52	2.64	9.45

The best performing model is highlighted in bold.

**Figure 4:** Marginal effect plots for the three most influential predictors on the concrete compressive strength for α values of 0.05 (red colour), 0.5 (green colour), and 0.95 (blue colour).

repeated 100 times. Performance of the proposed C- and D-vine two-step ahead quantile regression, compared to the C- and D-vine one-step ahead quantile regression, is evaluated by several measurements reported in Table 6 after 100 repetitions of fitting the models. It is not unexpected that the results of the four algorithms are more distinct than most simulation settings, given the small number of predictors. However, there is an improvement in the performance of the two-step ahead approach compared to the one-step ahead approach for both C- and D-vine-based models. Also, the C-vine model seems more appropriate for modeling the dependency structure in the data set. Finally, out of all models, the C-vine two-step ahead algorithm is the best performing algorithm in terms of out-of-sample predictions $\widehat{IS}_{0.5}$, $\widehat{CL}_{0.05}$, $\widehat{CL}_{0.5}$, $\widehat{CL}_{0.95}$ on the Concrete data set, as seen in Table 6.

In Figure 4, the marginal effect plots based on the fitted quantiles, from the C-vine two-step model, for the three most influential predictors are given. The marginal effect of a predictor is its expected impact on the quantile estimator, where the expectation is taken over all other predictors. This is estimated using all fitted conditional quantiles and smoothed over the predictors considered.

6.2 Riboflavin data set

The riboflavin data set, available in the R package *hdi*, aims at quantile predictions of the log-transformed production rate of *Bacillus subtilis* using log-transformed expression levels of 4,088 genes. To reduce the computational burden, we perform a pre-selection of the top 100 genes with the highest variance [10], resulting in a subset with $p = 100$ log-transformed gene expressions and $N = 71$ samples. Random splitting of the subset into training set with 61 samples and evaluation set with 10 samples, is repeated for 100 times. For the C- and D-vine two-step ahead quantile regression the number of candidates is set to $k = 10$. Additionally, to further reduce the computational burden the additional variable selection from Section 4.1.1 is applied with the chosen subset containing 25% of all possible choices, where 15% are predictors

Table 7: Out-of-sample predictions $\widehat{IS}_{0.5}$, $\widehat{CL}_{0.05}$, $\widehat{CL}_{0.5}$, $\widehat{CL}_{0.95}$

Model	$\widehat{IS}_{0.05}$	$\widehat{CL}_{0.05}$	$\widehat{CL}_{0.5}$	$\widehat{CL}_{0.95}$
D-vine One-step	33.83	0.44	0.42	0.41
D-vine Two-step	30.57	0.44	0.38	0.33
C-vine One-step	34.52	0.49	0.43	0.38
C-vine Two-step	28.59	0.41	0.36	0.30

The best performing model is highlighted in bold.

Table 8: Ten most influential gene expressions on the conditional quantile function, ranked based on their position in the order

Model/Position	1	2	3	4	5	6	7	8	9	10
D-vine One-step	GGT	YCIC	MTA	RPSE	YVAK	THIK	ANSB	SPOVB	YVZB	YQJB
D-vine Two-step	MTA	RPSE	THIK	YMFE	YCIC	sigM	PGM	YACC	YVQF	YKPB
C-vine One-step	GGT	YCIC	MTA	RPSE	HIT	BFMBAB	PHRC	YBAE	PGM	YHEF
C-vine Two-step	MTA	RPSE	THIK	YCIC	YURU	PGM	sigM	YACC	YKRM	ASNB

having the highest partial correlation with the log-transformed *Bacillus subtilis* production rate and the remaining 10% are chosen randomly from the remaining predictors. Performance of competitive quantile regression models is reported in Table 7, where we see that the proposed C-vine two-step ahead quantile regression is the best performing model and outperforms both the D-vine one-step ahead quantile regression from Kraus and Czado [41] and the C-vine one-step ahead quantile regression to a large extent. Furthermore, the second best performing method is the D-vine two-step ahead model which, while performing slightly worse than the C-vine two-step ahead model, also significantly outperforms both the C-vine and D-vine one-step ahead models. Since the predictors entering the C- and D-vine models yield a descending order of the predictors contributing to maximizing the conditional log-likelihood, the order indicates the influence of the predictors to the response variable. It is often of practical interest to know which gene expressions are of the highest importance for prediction. Since we repeat the random splitting of the subset for $R = 100$ times, the importance of the gene expressions is ranked sequentially by choosing

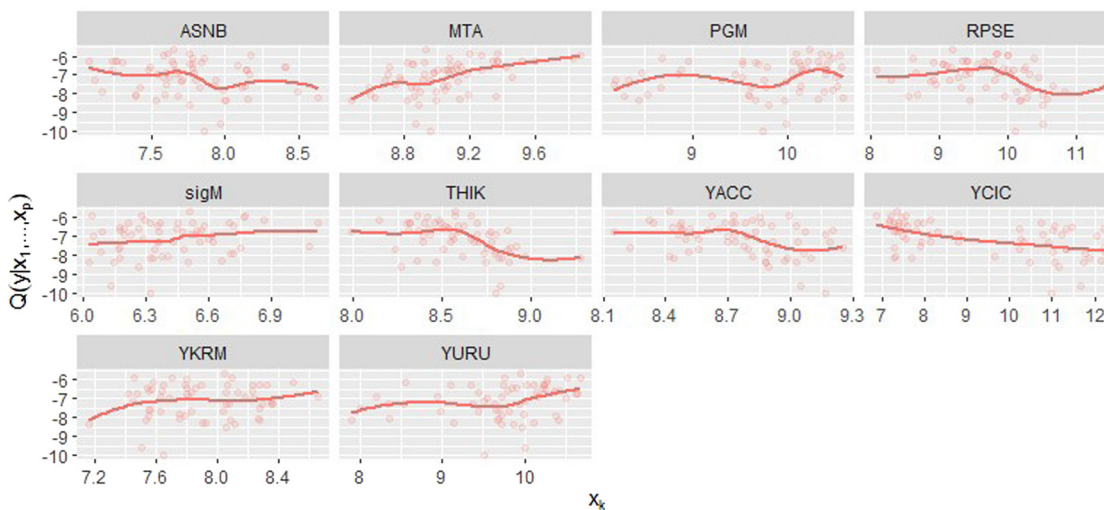


Figure 5: Marginal effect plots for the ten most influential predictors on the log-transformed *Bacillus subtilis* production rate for $\alpha = 0.5$.

the one with the highest frequency of each element in the order excluding the gene expressions chosen in the previous steps. For instance, the most important gene expression is chosen as the one most frequently ranked first; the second most important gene is chosen as the one most frequently chosen as the second element in the order, excluding the most important gene selected in the previous step. The top ten most influential gene expressions using the C- and D-vine one- or two-step ahead models are recorded in Table 8. Figure 5 shows the marginal effect plots based on the fitted quantiles, from the C-vine two-step model, for the ten most influential predictors on the log-transformed *Bacillus subtilis* production rate.

7 Summary and discussion

In this article, we introduce a two-step ahead forward selection algorithm for nonparametric C- and D-vine copula-based quantile regression. Inclusion of future information, obtained through considering the next tree in the two-step ahead algorithm, yields a significantly less greedy sequential selection procedure in comparison to the already existing one-step ahead algorithm for D-vine-based quantile regression in Kraus and Czado [41]. We extend the vine-based quantile regression framework to include C-vine copulas, providing an additional choice for the dependence structure. Furthermore, for the first time, nonparametric bivariate copulas are used to construct vine copula-based quantile regression models. The nonparametric estimation overcomes the problem of possible family misspecification in the parametric estimation of bivariate copulas and allows for even more flexibility in dependence estimation. Additionally, under mild regularity conditions, the nonparametric conditional quantile estimator is shown to be consistent.

The extensive simulation study, including several different settings and data sets with different dimensions, strengths of dependence, and tail dependencies, shows that the two-step ahead algorithm outperforms the one-step ahead algorithm in most scenarios. The results for the Concrete and Riboflavin data sets are especially interesting, as the C-vine two-step ahead algorithm has a significant improvement compared to the other algorithms. These findings provide strong evidence for the need of modeling the dependence structure following a C-vine copula. In addition, the two-step ahead algorithm allows controlling the computational intensity independently of the data dimensions through the number of candidate predictors and the additional variable selection discussed in Section 5. Thus, fitting vine-based quantile regression models in high dimensions becomes feasible. As seen in several simulation settings, there is a significant gain by introducing additional dependence structures other than the D-vine-based quantile regression. A further research area is developing similar forward selection algorithms for R-vine tree structures while optimizing the conditional log-likelihood.

At each step of the vine building stage, we compare equal-sized models with the same number of variables. The conditional log-likelihood is suited for such a comparison. Other questions might come in handy, such as choosing between C-vine, D-vine, or R-vine information criteria. When maximum likelihood estimation is employed at all stages, the selection criteria of Akaike (AIC) [1], the Bayesian information criterion (BIC) [55], and the focused information criterion (FIC) [16] might be used immediately. Ko *et al.* [33] studied FIC and AIC specifically for the selection of parametric copulas. The copula information criterion in the spirit of the Akaike information criterion by Gronneberg and Hjort [26] can be used for selection among copula models with empirically estimated margins, while Ko and Hjort [32] studied such a criterion for parametric copula models. We plan a deeper investigation of the use of information criteria for nonparametrically estimated copulas and for vines in particular. Such a study is beyond the scope of this article but could be interesting to study stopping criteria too for building vines.

Nonparametrically estimated vines are offering considerable flexibility. Their parametric counterparts, on the other hand, are enjoying simplicity. An interesting route for further research is to combine parametric and nonparametric components in the construction of the vines in an efficient way to bring the most benefit, which should be made tangible through some criteria such that guidance can be provided about which components should be modelled nonparametrically and which others are best modelled parametrically. For some types of models, such choice between a parametric and a nonparametric model has been

investigated by Jullum and Hjort [31] via the focused information criterion. This and alternative methods taking the effective degrees of freedom into account are worth further investigating for vine copula models.

Acknowledgements: The authors would like to thank the editor and the two referees for their comments, which helped to improve the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft [DFG CZ 86/6-1], the Research Foundation Flanders and KU Leuven internal fund C16/20/002. The resources and services used in this work were provided by the VSC (Flemish Supercomputer centre), funded by the Research Foundation-Flanders (FWO) and the Flemish Government.

Conflict of interest: Prof. Claudia Czado is an Honorary Board Member of “Dependence Modeling” although had no involvement in the final decision.

Appendix

A Construction of the transformation local likelihood estimator of the copula density

Let the $N \times 2$ transformed sample matrix be

$$D = (S, T), \quad (\text{A.1})$$

where the transformed samples $D_n = (S_n = \Phi^{-1}(U_i^{(n)}), T_n = \Phi^{-1}(U_j^{(n)}))$, $n = 1, \dots, N$, and Φ denotes the cumulative distribution function of a standard Gaussian distribution. The logarithm of the density $f_{S,T}$ of the transformed samples (S_n, T_n) , $n = 1, \dots, N$ is approximated locally by a bivariate polynomial expansion P_{a_m} of order m with intercept $\tilde{a}_{m,0}$ such that the approximation is denoted by

$$\tilde{f}_{S,T}(\Phi^{-1}(u_i^{(n)}), \Phi^{-1}(u_j^{(n)})) = \exp\{\tilde{a}_{m,0}(\Phi^{-1}(u_i^{(n)}), \Phi^{-1}(u_j^{(n)}))\}.$$

The transformation local likelihood estimator is then defined as

$$\tilde{c}(u_i^{(n)}, u_j^{(n)}) = \frac{\tilde{f}_{S,T}(\Phi^{-1}(u_i^{(n)}), \Phi^{-1}(u_j^{(n)}))}{\phi(\Phi^{-1}(u_i^{(n)}))\phi(\Phi^{-1}(u_j^{(n)}))}. \quad (\text{A.2})$$

To get the local polynomial approximation, we need a kernel function \mathbf{K} with 2×2 bandwidth matrix \mathbf{B}_N . For some pairs (\check{s}, \check{t}) close to (s, t) , $\log f_{ST}(\check{s}, \check{t})$ is assumed to be well approximated, locally, by for instance a polynomial with $m = 1$ (log-linear)

$$P_{a_1}(\check{s} - s, \check{t} - t) = a_{1,0}(s, t) + a_{1,1}(s, t)(\check{s} - s) + a_{1,2}(s, t)(\check{t} - t),$$

or $m = 2$ (log-quadratic)

$$P_{a_2}(\check{s} - s, \check{t} - t) = a_{2,0}(s, t) + a_{2,1}(s, t)(\check{s} - s) + a_{2,2}(s, t)(\check{t} - t) \\ + a_{2,3}(s, t)(\check{s} - s)^2 + a_{2,4}(s, t)(\check{t} - t)^2 + a_{2,5}(s, t)(\check{s} - s)(\check{t} - t).$$

The coefficient vector of the polynomial expansion P_{a_m} is denoted by $a_m(s, t)$, where $a_1(s, t) = (a_{1,0}(s, t), a_{1,1}(s, t), a_{1,2}(s, t))$ for the log-linear approximation and $a_2(s, t) = (a_{2,0}(s, t), \dots, a_{2,5}(s, t))$ for the log-quadratic. The estimated coefficient vector $\tilde{a}_m(s, t)$ is obtained by a maximization problem in Eq. (A.3)

$$\tilde{a}_m(s, t) = \arg \max_{a_m} \left\{ \sum_{n=1}^N \mathbf{K} \left(\mathbf{B}_N^{-1/2} \begin{pmatrix} s - S_n \\ t - T_n \end{pmatrix} \right) P_{a_m}(S_n - s, T_n - t) \right. \\ \left. - N \left\{ \int \int_{\mathbb{R}^2} \mathbf{K} \left(\mathbf{B}_N^{-1/2} \begin{pmatrix} s - \check{s} \\ t - \check{t} \end{pmatrix} \right) \exp(P_{a_m}(\check{s} - s, t - \check{t})) d\check{s} d\check{t} \right\} \right\}. \quad (\text{A.3})$$

While it is well known that kernel estimators suffer from the curse of dimensionality, in the vine construction only two-dimensional functions need to be estimated, this thus avoids problems with high-dimensionality. We next explain as in the study by Geenens et al. [22] how a bandwidth selection is obtained. Consider the principal component decomposition for the $N \times 2$ sample matrix $D = (S, T)$ in Eq. (A.1), such that the $N \times 2$ matrix (Q, R) follows

$$(Q, R)^T = WD^T, \quad (\text{A.4})$$

where each row of W is an eigenvector of $D^T D$. We obtain an estimator of f_{ST} through the density estimator of f_{QR} , which can be estimated based on a diagonal bandwidth matrix $\text{diag}(h_Q^2, h_R^2)$. Selecting the bandwidths h_Q uses samples $Q_n, n = 1, \dots, N$ as

$$h_Q = \arg \min_{h>0} \left\{ \int_{-\infty}^{\infty} \{\tilde{f}_Q^{(p)}\}^2 dq - \frac{2}{N} \sum_{n=1}^N \tilde{f}_{Q(-n)}^{(p)}(\hat{Q}_n) \right\}, \quad (\text{A.5})$$

where $\tilde{f}_Q^{(p)}$ ($p = 1, 2$) are the local polynomial estimators for f_Q , and $\tilde{f}_{Q(-n)}^{(p)}$ is the “leave-one-out” version of $\tilde{f}_Q^{(p)}$ computed by leaving out Q_n . The procedure of selecting h_R is similar. The bandwidth matrix for the bivariate copula density is then given by $\mathbf{B}_N = K_N^{(p)} W^{-1} \text{diag}(h_Q^2, h_R^2) W^{-1}$, where $K_N^{(p)}$ takes $N^{1/45}$ to ensure an asymptotic optimal bandwidth order for the local log-quadratic case ($p = 2$), see the study by Geenens et al. ([22], Section 4) for details. Selection for the k-nearest-neighbour-type bandwidth is similar. The k-nearest-neighbour bandwidths denoted as h'_Q and h'_R are obtained by restricting the minimization in Eq. (A.5) in the interval $(0, 1)$, i.e.,

$$h'_Q = \arg \min_{h'_Q \in (0,1)} \left\{ \int_{-\infty}^{\infty} \{\tilde{f}_Q^{(p)}\}^2 dq - \frac{2}{N} \sum_{n=1}^N \tilde{f}_{Q(-n)}^{(p)}(\hat{Q}_n) \right\}.$$

Estimating f_{QR} at any (q, r) is obtained by using its $k = K_N^{(p)} \cdot h'_Q \cdot N$ nearest neighbours where $K_N^{(p)}$ takes $N^{-4/45}$ for $p = 2$. The R package `rvinecopulib` only implemented the bandwidth in Eq. (A.5) for the quadratic case with $p = 2$.

B Proof of Proposition 4.1

Proof. We first show statement 1. By Eq. (3), the estimator $\hat{F}_{Y|X_1, \dots, X_p}^{-1}(\alpha|x_1, \dots, x_p) = \hat{F}_Y^{-1}(\hat{C}_{V|U_1, \dots, U_p}^{-1}(\alpha|\hat{u}_1, \dots, \hat{u}_p))$, where $\hat{u}_j = \hat{F}_j(x_j)$, $j = 1, \dots, p$ denote variables on the u-scale. To avoid heavy notation, N referring to the sample size will be omitted here. Then [56,62], to show the uniformly strong consistency of $\hat{F}_Y^{-1}(\hat{C}_{V|U_1, \dots, U_p}^{-1}(\alpha|\hat{u}_1, \dots, \hat{u}_p))$, we show

$$\sup_{\alpha} |\hat{F}_Y^{-1}(\hat{C}_{V|U_1, \dots, U_p}^{-1}(\alpha|\hat{u}_1, \dots, \hat{u}_p)) - F_Y^{-1}(C_{V|U_1, \dots, U_p}^{-1}(\alpha|u_1, \dots, u_p))| \rightarrow 0 \text{ a. s.}$$

To improve the readability and simplify the notation in the proof, we first introduce some shorthand notations. Define

$$\begin{aligned} D_{C,1} &= \hat{F}_Y^{-1}(\hat{C}_{V|U_1, \dots, U_p}^{-1}(\alpha|\hat{u}_1, \dots, \hat{u}_p)), \quad D_{C,2} = F_Y^{-1}(\hat{C}_{V|U_1, \dots, U_p}^{-1}(\alpha|\hat{u}_1, \dots, \hat{u}_p)), \\ D_{C,3} &= \hat{F}_Y^{-1}(C_{V|U_1, \dots, U_p}^{-1}(\alpha|u_1, \dots, u_p)), \quad D_{C,4} = F_Y^{-1}(C_{V|U_1, \dots, U_p}^{-1}(\alpha|u_1, \dots, u_p)), \end{aligned}$$

and the two differences $D_C = D_{C,1} - D_{C,3}$ and $D_F = D_{C,3} - D_{C,4}$.

For all $\alpha > 0$,

$$\begin{aligned}
1 &\geq P(\sup_{\alpha} |D_{C,1} - D_{C,4}| \leq \varepsilon) = P(\sup_{\alpha} |D_{C,1} - D_{C,3} + D_{C,3} - D_{C,4}| \leq \varepsilon) = P(\sup_{\alpha} |D_C + D_F| \leq \varepsilon) \\
&\geq P(\sup_{\alpha} \{|D_C| + |D_F|\} \leq \varepsilon) \geq P(\sup_{\alpha} |D_C| + \sup_{\alpha} |D_F| \leq \varepsilon) \\
&\geq P\left(\left(\sup_{\alpha} |D_C| \leq \frac{3}{4}\varepsilon\right) \cap \left(\sup_{\alpha} |D_F| \leq \frac{1}{4}\varepsilon\right)\right) \\
&= P\left(\left(\sup_{\alpha} |D_C| \leq \frac{3}{4}\varepsilon\right) \mid \left(\sup_{\alpha} |D_F| \leq \frac{1}{4}\varepsilon\right)\right) \cdot P\left(\sup_{\alpha} |D_F| \leq \frac{1}{4}\varepsilon\right).
\end{aligned} \tag{A.6}$$

Denote the event $A = \sup_{\alpha} |D_F| \leq \frac{1}{4}\varepsilon$, $P(A) = 1$ holds by the uniform strong consistency of the estimator of F_Y^{-1} . Next, we show that the conditional probability in Eq. (A.6) is equal to 1.

$$\begin{aligned}
P\left(\left(\sup_{\alpha} |D_C| \leq \frac{3}{4}\varepsilon\right) \mid A\right) &= P\left(\sup_{\alpha} |D_{C,1} - D_{C,2} + D_{C,2} - D_{C,3} + D_{C,3} - D_{C,4}| \leq \frac{3}{4}\varepsilon \mid A\right) \\
&\geq P\left(\sup_{\alpha} |D_{C,1} - D_{C,2}| + \sup_{\alpha} |D_{C,2} - D_{C,3}| + \sup_{\alpha} |D_{C,3} - D_{C,4}| \leq \frac{3}{4}\varepsilon \mid A\right).
\end{aligned}$$

This conditional probability is equal to 1, since the first and second supremum are less than or equal to $\frac{1}{4}\varepsilon$ by conditioning on A and due to the uniform consistency of \hat{F}_Y^{-1} . The last supremum is less than or equal to $\frac{1}{4}\varepsilon$ by Bartle and Joichi ([3], Thm. 2) on almost uniform convergence, applied to the continuous inverse distribution function F_Y^{-1} , and taking the measurable space to be the probability space. First, $P(\sup_{\alpha} |(\hat{C}_{V|U_1, \dots, U_p}^{-1}(\alpha|\hat{u}_1, \dots, \hat{u}_p)) - (C_{V|U_1, \dots, U_p}^{-1}(\alpha|u_1, \dots, u_p))| \leq \frac{1}{4}\varepsilon) = 1$, which can be argued similar to Eq. (A.6) using the uniform consistency and continuity of the inverse of the h-functions. Next, Eq. (A.6) states $P(\sup_{\alpha} |D_{C,1} - D_{C,4}| \leq \varepsilon) = 1$. We conclude that $\hat{F}_Y^{-1}(\hat{C}_{V|U_1, \dots, U_p}^{-1}(\alpha|\hat{u}_1, \dots, \hat{u}_p))$ is uniformly strong consistent.

To prove the weak consistency in 2, by [62,56], we only need to show $P(|D_{C,1} - D_{C,4}| \leq \varepsilon) \rightarrow 1$. Using the same technique as in Eq. (A.6) and a similar argument for proving statement 2 of Proposition 4.1 with Theorem 2 on convergence in measure in [3], the weak consistency can be obtained. \square

References

- [1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. Petrov & F. Csáki (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest: Akadémiai Kiadó.
- [2] Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2), 1148–1178.
- [3] Bartle, R. G., & Joichi, J. T. (1961). The preservation of convergence of measurable functions under composition. *Proceedings of the American Mathematical Society*, 12(1), 122–126.
- [4] Bartle, R. G., & Sherbert, D. R. (2000). *Introduction to real analysis*. New York: Wiley.
- [5] Bauer, A., & Czado, C. (2016). Pair-copula Bayesian networks. *Journal of Computational and Graphical Statistics*, 25(4), 1248–1271.
- [6] Bedford, T., & Cooke, R. M. (2002). Vines—a new graphical model for dependent random variables. *The Annals of Statistics*, 30(4), 1031–1068.
- [7] Belloni, A., & Chernozhukov, V. (2011). ℓ_1 -penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics*, 39(1), 82–130.
- [8] Bernard, C., & Czado, C. (2015). Conditional quantiles and tail dependence. *Journal of Multivariate Analysis*, 138, 104–126.
- [9] Breiman, L. (2001). Random forests, machine learning 45. *Journal of Clinical Microbiology*, 2(30), 199–228.
- [10] Bühlmann, P., & van de Geer, S. (2011). *Statistics for high-dimensional data: Methods, theory and applications*. Berlin/Heidelberg, Germany: Springer Science & Business Media.
- [11] Chang, B., & Joe, H. (2019). Prediction based on conditional distributions of vine copulas. *Computational Statistics & Data Analysis*, 139, 45–63.
- [12] Charpentier, A., Fermanian, J.-D., & Scaillet, O. (2007). The estimation of copulas: Theory and practice. In J. Rank (Ed.), *Copulas: From theory to application in finance* (pp. 35–64). London: Risk Books.

- [13] Chen, X., Koenker, R., & Xiao, Z. (2009). Copula-based nonlinear quantile autoregression. *The Econometrics Journal*, 12, S50–S67.
- [14] Cheng, C. (1995). Uniform consistency of generalized kernel estimators of quantile density. *The Annals of Statistics*, 23(6), 2285–2291.
- [15] Cheng, K.-F. (1984). On almost sure representation for quantiles of the product limit estimator with applications. *Sankhyā: The Indian Journal of Statistics, Series A*, 46, 426–443.
- [16] Claeskens, G., & Hjort, N. (2003). The focused information criterion. *Journal of the American Statistical Association*, 98, 900–916. With discussion and a rejoinder by the authors.
- [17] Czado, C. (2019). *Analyzing dependent data with vine copulas: A practical guide with R. Lecture notes in statistics*. Basel, Switzerland: Springer International Publishing.
- [18] Dua, D., & Graff, C. (2017). *UCI machine learning repository*. Irvine: University of California, School of Information and Computer Sciences.
- [19] Durrett, R. (2010). *Probability: Theory and examples*. Cambridge, UK: Cambridge University Press.
- [20] Fenske, N., Kneib, T., & Hothorn, T. (2011). Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression. *Journal of the American Statistical Association*, 106(494), 494–510.
- [21] Geenens, G. (2014). Probit transformation for kernel density estimation on the unit interval. *Journal of the American Statistical Association*, 109(505), 346–358.
- [22] Geenens, G., Charpentier, A., & Paindaveine, D. (2017). Probit transformation for nonparametric kernel estimation of the copula density. *Bernoulli*, 23(3), 1848–1873.
- [23] Gijbels, I., & Matherne, M. (2021). Study of partial and average conditional Kendall's tau. *Dependence Modeling*, 9, 82–120.
- [24] Gijbels, I., & Mielniczuk, J. (1990). Estimating the density of a copula function. *Communications in Statistics-Theory and Methods*, 19(2), 445–464.
- [25] Grønneberg, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378.
- [26] Gronneberg, S., & Hjort, N. L. (2014). The copula information criteria. *Scandinavian Journal of Statistics*, 41(2), 436–459.
- [27] Haff, I. H., Aas, K., & Frigessi, A. (2010). On the simplified pair-copula construction—simply useful or too simplistic? *Journal of Multivariate Analysis*, 101(5), 1296–1310.
- [28] Joe, H. (1996). Families of m -variate distributions with given margins and $m(m - 1)/2$ bivariate dependence parameters. *Lecture Notes-Monograph Series*, 28, 120–141.
- [29] Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. Boca Raton, Florida, USA: CRC Press.
- [30] Joe, H. (2014). *Dependence modeling with copulas*. Boca Raton, Florida, USA: CRC Press.
- [31] Jullum, M., & Hjort, N. L. (2017). Parametric or nonparametric: The FIC approach. *Statistica Sinica*, 27(3), 951–981.
- [32] Ko, V., & Hjort, N. L. (2019). Copula information criterion for model selection with two-stage maximum likelihood estimation. *Econometrics and Statistics*, 12, 167–180.
- [33] Ko, V., Hjort, N. L., & Hobæ Haff, I. (2019). Focused information criteria for copulas. *Scandinavian Journal of Statistics*, 46(4), 1117–1140.
- [34] Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1), 74–89.
- [35] Koenker, R. (2005a). *Quantile regression*. Cambridge, UK: Cambridge University Press.
- [36] Koenker, R. (2005b). *Quantile regression. Econometric Society Monographs*. Cambridge University Press.
- [37] Koenker, R. (2011). Additive models for quantile regression: Model selection and confidence bands. *Brazilian Journal of Probability and Statistics*, 25(3), 239–262.
- [38] Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, 46, 33–50.
- [39] Kolmogorov, A. N., & Fomin, S. (1970). *Introductory real analysis*. Hoboken, New Jersey, USA: Prentice-Hall.
- [40] Komunjer, I. (2013). Quantile Prediction. In Y. Ait-Sahalia & L. Peter Hansen (Eds.), *Chapter 17 in Handbook of financial econometrics*. Amsterdam, Netherlands: Elsevier.
- [41] Kraus, D., & Czado, C. (2017). D-vine copula based quantile regression. *Computational Statistics & Data Analysis*, 110, 1–18.
- [42] Li, A. H., & Martin, A. (2017). Forest-type regression with general losses and robust forest. In International Conference on Machine Learning (pp. 2091–2100). PMLR.
- [43] Li, Q., Lin, J., & Racine, J. S. (2013). Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. *Journal of Business & Economic Statistics*, 31(1), 57–65.
- [44] Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, 7(Jun), 983–999.
- [45] Nagler, T. (2018). kdecopula: An R package for the kernel estimation of bivariate copula densities. *Journal of Statistical Software*, 84(7), 1–22.
- [46] Nagler, T. (2019). *vinereg: D-Vine Quantile Regression*. R package version 0.7.0.
- [47] Nagler, T., Schellhase, C., & Czado, C. (2017). Nonparametric estimation of simplified vine copula models: Comparison of methods. *Dependence Modeling*, 5, 99–120.
- [48] Nagler, T., & Vatter, T. (2019a). kde1d: Univariate Kernel Density Estimation. R package version 1.0.2.

- [49] Nagler, T., & Vatter, T. (2019b). *rvinecopulib: High Performance Algorithms for Vine Copula Modeling*. R package version 0.5.1.1.0.
- [50] Noh, H., Ghouch, A. E., & Bouezmarni, T. (2013). Copula-based regression estimation and inference. *Journal of the American Statistical Association*, 108(502), 676–688.
- [51] Noh, H., Ghouch, A. E., & Van Keilegom, I. (2015). Semiparametric conditional quantile estimation through copula-based multivariate models. *Journal of Business & Economic Statistics*, 33(2), 167–178.
- [52] Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3), 1065–1076.
- [53] R CoreTeam. (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- [54] Ryzin, J. V. (1969). On strong consistency of density estimates. *The Annals of Mathematical Statistics*, 40(5), 1765–1772.
- [55] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464.
- [56] Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *The Annals of Statistics*, 6, 177–184.
- [57] Sklar, M. (1959). *Fonctions de repartition an dimensions et leurs marges* (Vol. 8, pp. 229–231). Paris: Publ. Inst. Statist. Univ.
- [58] Stoeber, J., Joe, H., & Czado, C. (2013). Simplified pair copula constructions—limitations and extensions. *Journal of Multivariate Analysis*, 119, 101–118.
- [59] Tepegjuzova, M. (2019). *D- and C-vine quantile regression for large data sets*. Masterarbeit, Technische Universität München, Garching b. München.
- [60] Van Keilegom, I., & Veraverbeke, N. (1998). Bootstrapping quantiles in a fixed design regression model with censored data. *Journal of Statistical Planning and Inference*, 69(1), 115–131.
- [61] Wen, K., & Wu, X. (2015). An improved transformation-based kernel estimator of densities on the unit interval. *Journal of the American Statistical Association*, 110(510), 773–783.
- [62] Wied, D., & Weißbach, R. (2012). Consistency of the kernel density estimator: A survey. *Statistical Papers*, 53(1), 1–21.
- [63] Xiao, Z., & Koenker, R. (2009). Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *Journal of the American Statistical Association*, 104(488), 1696–1712.
- [64] Yeh, I.-C. (1998). Modeling of strength of high-performance concrete using artificial neural networks. *Cement and Concrete Research*, 28(12), 1797–1808.
- [65] Yu, K., & Jones, M. (1998). Local linear quantile regression. *Journal of the American statistical Association*, 93(441), 228–237.
- [66] Yu, K., & Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54(4), 437–447.
- [67] Zhu, K., Kurowicka, D., & Nane, G. F. (2021). Simplified R-vine-based forward regression. *Computational Statistics & Data Analysis*, 155, 107091.