Flexible Copula Density Estimation with Penalized Hierarchical B-Splines

GÖRAN KAUERMANN

Department of Statistics, Ludwig-Maximilians-University Munich

CHRISTIAN SCHELLHASE

Department for Business Administration and Economics, Bielefeld University

DAVID RUPPERT

School of Operations Research and Information Engineering, Cornell University

1st February 2013

Abstract

The paper introduces a new method for flexible spline fitting for copula density estimation. Spline coefficients are penalized to achieve a smooth fit. To weaken the curse of dimensionality, instead of a full tensor spline basis, a reduced tensor product based on sparse grids Zenger (1991) is used. To achieve uniform margins of the copula density, linear constraints are placed on the spline coefficients and quadratic programming is used to fit the model. Simulations and practical examples accompany the presentation.

Keywords: Copula Density Estimation, Hierarchical B-splines, Penalized Spline Smoothing, Quadratic Programming.

1 Introduction

Copulas allow for stochastic modelling of multivariate distributions beyond the classical normal distribution. The idea traces back to Sklar (1959), though Hoeffding (1940)

might be consulted as earlier reference, see Nelsen (2006). Copulas have experienced general interest in the last years, primarily in the area of finance, see for instance McNeil *et al.* (2005), though the idea has been applied in other contexts as well, see for example Bogaerts & Lesaffre (2008) or Song *et al.* (2009) for bio-statistical applications or Danaher & Smith (2011) for the use of copulas in marketing. A general overview and survey of recent contributions in copula modelling is found e.g. in Härdle & Okhrin (2009) or, from a more personal viewpoint, in Embrechts (2009); see also Kolev *et al.* (2006). A comprehensive collection of new approaches in copula estimation is provided in Jaworski *et al.* (2010). This includes, inter alia, hierarchical modelling of Archimedean copulas as suggested in Okhrin *et al.* (2009) and Savu & Trede (2010). Lambert (2007) uses Bayesian spline smoothing for estimating the generator function of a Archimedean copula. Joe (1996) pursues the use of so called pair-copulas, where multiple interaction is reduced to bivariate copula modelling, see also Bedford & Cooke (2002) or Czado (2010).

While the above literature on copula estimation is vast and extensive, this does not apply to non- and semi-parametric routines for copula estimation which is tackled in this paper. This is surprising at a first glance but can in our opinion be explained with the following two reasons. First, a copula has the property that its univariate margins are uniform. Such side constraints are however difficult to accommodate in available non-parametric estimation routines. Secondly, copulas have the potential to work in high dimensional problems, while classical non-parametric techniques suffer from the so called curse of dimensionality if the dimension exceeds two (or three). Our approach presented in this paper solves the first problem by directly including constraints on the margins in the optimization routine. It turns out that the requirement of uniform margins can be easily formulated as linear constraints on spline coefficients. Moreover, we tackle the second problem, the curse of dimensionality, by making use of so-called sparse grids. This means instead of a full tensor product of splines as basis, a reduced form is used to achieve numerical feasibility in dimensions beyond two (or three).

Considering the literature on non-parametric copula estimation we refer to kernel density methods proposed in Gijbels & Mielniczuk (1990) which are further discussed in Fermanian & Scaillet (2003), Fermanian *et al.* (2004) and Chen & Huang (2007). In

these papers, the copula itself is fitted using a smoothed version of the empirical copula. Omelka *et al.* (2009) modify the estimate by correcting the "corner" bias of the kernel density estimates. More recently the use of wavelet based estimation has been suggested by Morettin *et al.* (2010) for copula estimation or Genest *et al.* (2009) for copula density estimation, see also for a more theoretical investigation Autin *et al.* (2010). As an alternative to wavelets, the use of Bernstein polynomials has been proposed in Sancetta & Satchell (2004); see also Qu *et al.* (2009) and Pfeifer *et al.* (2009). Instead of Bernstein polynomials one may also use linear B-splines, as pursued in this paper, see also Shen *et al.* (2008). Replacing the copula density itself by a piecewise constant function has been pursued by Qu *et al.* (2009) or in Qu & Yin (2012). The use of Wavelets, piecewise constraints, Bernstein polynomials and B-splines allows to accommodate the constraint that univariate marginal distributions are uniform. In practice, however, none of these methods do directly extend to higher dimensions due to the above-mentioned curse of dimensionality. That is to say, numerically it is hardly feasible to apply the routines to more than two (or three) dimensions, so that the major focus in all cited papers lies on the bivariate case. In our paper, we make use of B-splines to model the copula density itself. To do so, we replace the copula density by a (linear) combination of tensor products of univariate B-splines on $[0, 1]$. This idea builds upon Marx & Eilers (2005); see also Koo (1996). With simple linear constraints on the spline coefficients we can guarantee that the univariate margins of the copula are uniform, that is the spline estimate itself is a copula density. To achieve smoothness of the fitted copula, we impose a penalty on the spline coefficients as suggested by Eilers & Marx (1996), see also Ruppert et al. (2003, 2009).

With the spline approach suggested in the paper, we are, however, still faced with the problem of the curse of dimensionality. This implies that the dimensionality of the spline basis increases exponentially with the dimension of the variables and, in fact, can reach the order of a million even for 4 or 5 dimensional random vectors. To adapt the spline approach to higher dimensions, we make use of so called "sparse grids". The idea traces back to Zenger (1991) and is extensively discussed and motivated in Bungartz & Griebel (2004); see also Garcke (2006). Sparse grids make use of hierarchical B-splines as discussed, for instance, in Forsey & Bartels (1988). The idea is to represent a B-

splines basis by B-splines of lower dimension, that is, built upon fewer knots. Figure 1 shows how a linear B-spline [plot (a)] can be represented by B-splines constructed at fewer knots [plots (b) to (d)]. More details are provided in the paper. The idea of sparse grids is now to replace the full tensor product by a reduced form including only products of hierarchical splines up to a limited hierarchy order. This reduces the numerical effort tremendously and allows us to weaken the curse of dimensionality. Practically it means we are able to fit 4 (or even 5) dimensional copulas with a fully semi-parametric approach.

The novel contributions of the paper are (a) copula density estimation which guarantees uniform margins and allows for fast numerical fitting by imposing simple linear constraints on the parameters and (b) proposing the use of sparse grids in the field of nonparametric copula estimation which allows to weaken the curse of dimensionality to fit models in 3, 4 or 5 dimensions. The paper is organized as follows. In Section 2 we introduce the estimation routine with hierarchical B-splines and sparse grids. At the end of Section 2, we discuss the numerical implementation including the incorporation of constraints on the marginal densities. In Section 3, we investigate the performance of our copula estimator using simulations and two examples. A conclusion finalizes the paper with technical details transferred to the Appendix.

2 Penalized B-Spline Estimation of a Copula Density

2.1 B-Spline Density Basis

Following Sklar's (1959) theorem, we can write the distribution of the p dimensional random vector $X = (X_1, \ldots, X_p)$ as

$$F(x_1, \ldots, x_p) = C\{F_1(x_1), \ldots, F_p(x_p)\}, \tag{1}$$

where $C(., \ldots, .)$ is the copula corresponding to $F(\cdot)$. We assume that copula $C(., \ldots, .)$ is a distribution function on the p-dimensional cube $[0, 1]^p$, with uniform marginal distributions and copula density $c(., \ldots, .)$ which is related to the density $f(x_1, \ldots, x_p)$

through

$$f(x_1, \ldots, x_p) = c\left\{F_1(x_1), \ldots, F_p(x_p)\right\} \prod_{j=1}^{p} f_j(x_j). \tag{2}$$

Our intention is to estimate the copula density $c(.)$ itself, either assuming the marginal distribution $F_j(x_j)$ to be known or being estimated separately. Let therefore $u_j = F(x_j)$ so that $c(u_1, \ldots, u_p)$ is a density on $[0,1]^p$ with the p margin-constraints

$$\int_{\substack{p \\ \times \\ i \neq j}} c(u_1, \ldots, u_p) \prod_{i \neq j}^{p} \mathrm{d}u_i = 1, \text{ for } j = 1, \ldots, p. \tag{3}$$

We estimate $c(\cdot)$ in a flexible semi-parametric way by taking the p constraints (3) into account. To do this, we will approximate $c(\cdot)$ by a mixture of basis densities. Let therefore $\phi_k(u)$ be a regular linear univariate B-spline normalized to be a density, i.e., $\int \phi_k(u)\, \mathrm{d}u = 1$ with $u \in [0,1]$ and denote with $\Phi(\cdot) = \{\phi_l(\cdot),\ l = 1, \ldots, K\}$ the univariate B-spline density basis of dimension K, see Figure 1 (a). We construct the full tensor product as $\mathbf{\Phi}(u_1, \ldots, u_p) = \bigotimes_{j=1}^{p} \Phi(u_j)$ and reexpress $\mathbf{\Phi}(\cdot)$ by letting $\mathbf{k} = (k_1, \ldots, k_p)$ be a p-tuple with $\mathbf{k} \in \mathcal{K} = \{1, \ldots, K\}^p$. The components of $\mathbf{\Phi}(\cdot)$ are then

$$\phi_{\mathbf{k}}(u_1, \ldots, u_p) = \phi_{k_1, \ldots, k_p}(u_1, \ldots, u_p) = \prod_{j=1}^{p} \phi_{k_j}(u_j),$$

where $k_j \in \{1, \ldots, K\}$ for $j = 1, \ldots, p$. The idea is now to approximate the copula density through the B-splines such that

$$c(u_1, \ldots, u_p) \approx \sum_{\mathbf{k} \in \mathcal{K}} b_{\mathbf{k}} \phi_{\mathbf{k}}(u_1, \ldots, u_p) =: c(u_1, \ldots, u_p; \mathbf{b}). \tag{4}$$

The goodness of the approximation depends thereby on the richness of the basis, that is, on the number of elements in $\mathcal{K}$. We discuss this point later. The elements of $\mathbf{b} = (b_{\mathbf{k}}, \mathbf{k} \in \mathcal{K})$ are subsequently called the spline basis coefficients and with each single basis spline being a density itself we obtain with conditions

$$\sum_{\mathbf{k} \in \mathcal{K}} b_{\mathbf{k}} = 1, \quad c(\mathbf{u}; \mathbf{b}) \geq 0 \tag{5}$$

that $c(\mathbf{u}; \mathbf{b})$ in (4) is a density. For simplicity we ignore at this point that $c(\cdot; \mathbf{b})$ is not guaranteed to be copula density in that univariate margins are not guaranteed to be uniform. We will come back to this condition later.

To construct the likelihood for spline coefficients $\mathbf{b}$, assume we have a random sample $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})$ with $i = 1, \ldots, n$ from which we construct $\mathbf{u}_i = (u_{i1}, \ldots, u_{ip})$ through $u_{ij} = \hat{F}_j(x_{ij})$. Here, $\hat{F}_j(.)$ is a $\sqrt{n}$ consistent estimate of the marginal distribution function, which in the simplest case is just the empirical distribution function and hence nu_{ij} are the ranks. Based on $\mathbf{u}_i, i = 1, \ldots, n$, the log likelihood for $\mathbf{b}$ is then

$$l(\mathbf{b}) = \sum_{i=1}^{n} \log \left\{ \sum_{\mathbf{k} \in \mathcal{K}} b_{\mathbf{k}} \phi_{\mathbf{k}}(u_{i1}, \ldots, u_{ip}) \right\}, \tag{6}$$

which needs to be maximized subject to the constraints (5). The accuracy of the spline approximation in (4) improves for large K, but the corresponding fit will suffer from estimation variability due to over-parameterization of the data. Entertaining the ideas of penalized splines (see also Ruppert *et al.* 2003), we impose a penalty on spline coefficients $b_{\mathbf{k}}$ to achieve a smooth fit for $c(\cdot)$. Eilers & Marx (1996) suggest to penalize r-th order differences for the B-spline coefficients. This easily extends to the multivariate setting as shown in Marx & Eilers (2005). Let $L \in \mathbb{R}^{(K-r)\times K}$ be a difference matrix of order r, e.g. for $r = 1$ we get

$$L = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix},$$

and let $W = \text{diag}(w_1, \ldots, w_K)$ be the weight matrix linking a regular B-spline basis to a B-spline density basis, i.e. w_l is the integral from 0 to 1 of the l-th regular B-spline. With matrix L we can now penalize differences in neighbouring spline coefficients and define the penalty matrix $P = WL^T LW$; see also Wand & Ormerod (2008) and Ruppert *et al.* (2003). This penalty applies only to a single dimension. To achieve smoothness of the fitted copula density for all variables, we use the Kronecker product yielding the entire penalty matrix

$$\mathbf{P}(\boldsymbol{\lambda}) = \sum_{j=1}^{p} \lambda_j \mathbf{P}_j.$$

with $\mathbf{P}_j = \left(\bigotimes_{l=1}^{j-1} I_K \right) \otimes P \otimes \left(\bigotimes_{l=j+1}^{p} I_K \right)$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)$ where I_K is the K dimensional identity matrix and $\bigotimes_{l=1}^{j-1}$ denotes component-by-component tensor products (where $\bigotimes_{l=1}^{0} I_K = 1 = \bigotimes_{l=p+1}^{p} I_K$). The coefficient λ_j is the penalty parameter for

the j-th variable which needs to be selected in a data driven manner, as discussed later. Incorporating the penalty into the log likelihood gives the penalized log likelihood

$$l_p(\mathbf{b}, \boldsymbol{\lambda}) = l(\mathbf{b}) - \frac{1}{2}\mathbf{b}^T\mathbf{P}(\boldsymbol{\lambda})\mathbf{b}, \tag{7}$$

which is maximized for given $\boldsymbol{\lambda}$ with respect to $\mathbf{b}$. Note that $\boldsymbol{\lambda}$ determines the amount of smoothness for the fitted coefficients and setting $\boldsymbol{\lambda} = 0$ gives the unpenalized ML estimate.

2.2 Hierarchical B-splines and Sparse Grids

The modelling approach proposed above becomes numerically infeasible if the dimension p exceeds 2 or 3, since the dimension of the tensor product basis grows exponentially in p. To illustrate this curse of dimensionality, Table 1 gives the dimension of a full tensor product based on a linear B-spline basis of dimension $K = 2^d + 1$ for different dimensions of $\mathbf{u}$ ranging from $p = 2$ to $p = 5$. Even for a $p = 3$ dimensional vector u and $K = 17$, one ends up with nearly 5000 parameters, which is at the limit of numerical feasibility. We therefore suggest reducing the spline dimension for numerical purposes by not taking a full tensor product but, instead, using a reduced form to guarantee numerical feasibility. Our approach makes use of Zenger's (1991) so called 'sparse grids'. To apply the idea we first transform the univariate B-spline density into its hierarchical form. Let the linear univariate B-spline density basis be built upon $2^d + 1$ equidistant knots $\tau_k = k2^{-d}, k = 0, \ldots, 2^d$. The basis has dimension $K = 2^d + \mathbf{1}$ and is denoted subsequently as $\Phi_{(d)}(u) = \left\{\phi_{(d)l}(u), l = 1, \ldots, K\right\}$. We can reexpress this basis in hierarchical terms as derived in Forsey and Bartels (1988, 1995); see also Garcke (2006). Let $\mathcal{I}_0 = \{1, \mathbf{2}\}$ and $\mathcal{I}_h = \{2j, \text{for } 1 \leq j \leq 2^{h-1}\}$ for $h = 1, \ldots, d$ denote hierarchical index sets. The hierarchical B-spline basis linearly equivalent to $\Phi_{(d)}(u)$ is then defined through

$$\tilde{\Phi}_{(d)}(u) = \left\{\phi_{(h)l}(u), l \in \mathcal{I}_h, h = 0, \ldots, d\right\} = \left\{\Phi_{(h)\mathcal{I}_h}, h = 0, \ldots, d\right\}. \tag{8}$$

Figure 1 illustrates the hierarchical spline in plots (b) to (d) with B-spline basis $\phi_{(0)1}(.), \phi_{(0)2}(.)$ for (b), $\phi_{(1)2}(.)$ for (c) and $\phi_{(2)2}(.), \phi_{(2)4}(.)$ for (d). It is not difficult to show that both bases, (a) and (b) to (d), span the same space so that $\Phi_{(d)}(u) = \tilde{\Phi}_{(d)}(u)\tilde{A}$

for some invertible $K \times K$ matrix $\tilde{A}$. We now reformulate the penalized likelihood (7) by replacing the B-spline bases in (4) with their hierarchical form. To do this, let the complete tensor product based on the hierarchical B-spline basis $\tilde{\Phi}_{(d)}(\cdot)$ be denoted with

$$\tilde{\mathbf{\Phi}}_{(d)}(u_1, \ldots, u_p) = \bigotimes_{j=1}^{p} \tilde{\Phi}_{(d)}(u_j) = \mathbf{\Phi}(u)\tilde{\mathbf{A}}^{-1}$$

and $\tilde{\mathbf{A}}^{-1} = \bigotimes_{j=1}^{p} \tilde{A}^{-1}$. Let $\tilde{\mathbf{b}} = \tilde{\mathbf{A}}^{-1}\mathbf{b}$ denote the corresponding spline coefficient vector for basis $\tilde{\mathbf{\Phi}}_{(d)}(\cdot)$. The penalized likelihood (7) can then be rewritten in terms of $\tilde{\mathbf{b}}$ taking the form

$$\tilde{l}_p(\tilde{\mathbf{b}}, \boldsymbol{\lambda}) = \tilde{l}(\tilde{\mathbf{b}}) - \frac{1}{2}\tilde{\mathbf{b}}^T\tilde{\mathbf{P}}(\boldsymbol{\lambda})\tilde{\mathbf{b}}$$

with $\tilde{l}(\tilde{\mathbf{b}}) = \sum_{i=1}^{n} \log\left\{\tilde{\mathbf{\Phi}}_{(d)}(\mathbf{u}_i)\tilde{\mathbf{b}}\right\}$ and $\tilde{\mathbf{P}}(\boldsymbol{\lambda}) = \sum_{j=1}^{p} \lambda_j \tilde{\mathbf{P}}_j$ where

$$\tilde{\mathbf{P}}_j = \left(\bigotimes_{l=1}^{j-1} \tilde{I}_{(d)}\right) \otimes \{(\tilde{A}^{-1})^T P \tilde{A}^{-1}\} \otimes \left(\bigotimes_{l=j+1}^{p} \tilde{I}_{(d)}\right)$$

and $\tilde{I}_{(d)} = (W\tilde{A}^{-1})^T(W\tilde{A}^{-1})$.

The parameterization with hierarchical B-splines allows us to tackle the curse of dimensionality by making use of a so-called sparse grid approach. The underlying idea is to consider spline tensor products up to a cumulated hierarchy order D only. Figure 2 illustrates the idea for dimension $p = 2$ and $D = 2$ using a linear B-spline basis. To be specific, we define the sparse grid tensor product as

$$\tilde{\mathbf{\Phi}}_{(d)}^{(D)}(u_1, \ldots, u_p) = \left(\bigotimes_{j=1}^{p} \Phi_{(h_j)\mathcal{I}_{h_j}}(u_j), \sum_{j=1}^{p} h_j \leq D\right). \tag{9}$$

The upper index D refers to the maximum hierarchy level and the lower index d is the hierarchy level of the marginal hierarchical B-spline basis. Note that $d \leq D \leq pd$ is a useful range for D and $\tilde{\mathbf{\Phi}}_{(d)}^{(pd)}(\cdot) = \tilde{\mathbf{\Phi}}_{(d)}(\cdot)$. The reduction of the basis reduces the numerical effort tremendously as can be seen from Table 1 where we show the dimension of $\tilde{\mathbf{\Phi}}_{(d)}$ and $\tilde{\mathbf{\Phi}}_{(d)}^{(D)}$ for various values of d and D. For $p = 3$ and $d = D = 4$ (i.e. $K = 2^d + 1$) we get a 297 dimensional basis instead of 4913 dimensional. Note that the reduced basis is created by extracting columns of the complete tensor product basis. This means we can write

$$\tilde{\mathbf{\Phi}}_{(d)}^{(D)}(u_1, \ldots, u_p) = \tilde{\mathbf{\Phi}}_{(d)}(u_1, \ldots, u_p)\mathbf{J}_{(d)}^{(D)}$$

where $\mathbf{J}_{(d)}^{(D)}$ is an indicator matrix with entries 0 and a single entry 1 per column for extracting the matching columns of $\tilde{\mathbf{\Phi}}_{(d)}$. Note that with this definition $\mathbf{J}_{(d)}^{(pd)}$ is the identity matrix. Let $\tilde{\mathbf{b}}^{(D)}$ denote the basis coefficients corresponding to the sparse splines basis. We define the sparse penalized log likelihood by extracting the corresponding elements from the complete penalty matrix, that is

$$\tilde{l}_p^{(D)}(\tilde{\mathbf{b}}^{(D)}, \boldsymbol{\lambda}) = \tilde{l}^{(D)}(\tilde{\mathbf{b}}^{(D)}) - \frac{1}{2}\tilde{\mathbf{b}}^{(D)^T}\tilde{\mathbf{P}}^{(D)}(\boldsymbol{\lambda})\tilde{\mathbf{b}}^{(D)} \tag{10}$$

with obvious definition for $\tilde{l}^{(D)}(\tilde{\mathbf{b}}^{(D)})$ and $\tilde{\mathbf{P}}^{(D)}(\boldsymbol{\lambda}) = \mathbf{J}_{(d)}^{(D)^T}\tilde{\mathbf{P}}(\boldsymbol{\lambda})\mathbf{J}_{(d)}^{(D)}$. Note that since $\tilde{\mathbf{b}}^{(pd)} = \tilde{\mathbf{b}}$ we have $\tilde{l}^{(pd)}(\cdot) = \tilde{l}(\cdot)$.

Now that we have reduced the basis dimension to make copula density estimation feasible even beyond the bivariate case, it remains to tackle the question of how well we can approximate an arbitrary copula density $c(u)$ by a sparse grid representation $c(\mathbf{u}, \mathbf{b}^{(D)}) = \mathbf{\Phi}_{(d)}^{(D)}(\mathbf{u})\mathbf{b}^{(D)}$.

Lemma 1 *Assuming, that $c(\cdot)$ is continuously differentiable in $[0,1]$, the cumulated hierarchy D of the sparse grid representation $c(\mathbf{u}, \mathbf{b}^{(D)}) = \mathbf{\Phi}_{(d)}^{(D)}(\mathbf{u})\mathbf{b}^{(D)}$ determines the order of the approximation error by*

$$c(\mathbf{u}) = c(\mathbf{u}, \mathbf{b}^{(D)}) + O(2^{-D}). \tag{11}$$

A proof is given in the Appendix. So, not surprisingly, accuracy and numerical feasibility are in competition.

Lemma 2 *Based on the assumptions of Lemma 1, the bias and variance of the estimated penalized Maximum Likelihood estimate $\hat{\tilde{\mathbf{b}}}^{(D)}$ based on* (10) *are equal to*

$$E(\hat{\tilde{\mathbf{b}}}^{(D)}) = \tilde{\mathbf{b}}_0^{(D)} + \left\{\mathbf{H}_p^{(D)}(\tilde{\mathbf{b}}_0^{(D)}, \boldsymbol{\lambda})\right\}^{-1}\tilde{\mathbf{P}}^{(D)}(\boldsymbol{\lambda})\tilde{\mathbf{b}}_0^{(D)} \tag{12}$$

$$Var(\hat{\tilde{\mathbf{b}}}^{(D)}) = \left\{\mathbf{H}_p^{(D)}(\tilde{\mathbf{b}}_0^{(D)}, \boldsymbol{\lambda})\right\}^{-1}\mathbf{H}_p^{(D)}\left(\tilde{\mathbf{b}}_0^{(D)}, \boldsymbol{\lambda} = 0\right)\left\{\mathbf{H}_p^{(D)}(\tilde{\mathbf{b}}_0^{(D)}, \boldsymbol{\lambda})\right\}^{-1}. \tag{13}$$

The proofs are given in the Appendix.

2.3 Constraints on the Parameters and Penalization

Until now we have not incorporated the constraints that univariate margins of the copula density $c(\mathbf{u})$ are uniform. To have the estimate $c(\mathbf{u}, \hat{\tilde{\mathbf{b}}}^{(D)})$ be a proper copula

density we need to impose uniform, univariate margins. First, we need to calculate the marginal density from $\tilde{\mathbf{\Phi}}_{(d)}^{(D)}(u_1,\ldots,u_p)\tilde{\mathbf{b}}^{(D)}$. Looking for example at Figure 2 we can appreciate that the univariate margins are represented with the univariate spline basis $\tilde{\Phi}_{(d)}^{(D)}(u_j)$ and the corresponding marginal basis coefficient vector $\tilde{\mathbf{b}}_{(j)}^{(D)}$, say, with elements being calculated as the sum over a set of elements of $\tilde{\mathbf{b}}^{(D)}$. In the bivariate case this results from summing up row-wise (for u_2) or column-wise (for u_1) the corresponding spline coefficients in the basis representation shown in Figure 2. We give exact formulae how to calculate $\tilde{\mathbf{b}}_{(j)}^{(D)}$ from $\tilde{\mathbf{b}}^{(D)}$ in the Appendix. Note that this is a simple linear calculation and hence fast and straight forward, so that the marginal density is numerically easy to obtain. To guarantee that the marginal density is uniform, we now simply impose the constraints on the coefficients evaluated at the knots τ_k

$$\tilde{\Phi}_{(d)}(\tau_k)\hat{\tilde{b}}_{(j)}^{(D)} = 1, k = 1,\ldots,K,\ j = 1,\ldots,p. \tag{14}$$

We need two further constraints to have $c(\mathbf{u},\tilde{\mathbf{b}}^{(\mathbf{D})})$ being a density. First, the fitted curve $c(\mathbf{u};\hat{\tilde{\mathbf{b}}}^{(D)}) := \tilde{\mathbf{\Phi}}_{(d)}^{(D)}(u_1,\ldots,u_p)\hat{\tilde{\mathbf{b}}}^{(D)}$ is required to be a density. Since all columns in the hierarchical basis $\tilde{\mathbf{\Phi}}_{(d)}^{(D)}$ are B-spline densities over $u_1\ldots,u_p$ we therefore need to guarantee that the sum of the components of $\hat{\tilde{\mathbf{b}}}^{(D)}$ equals 1, i.e.,

$$\mathbf{1}^T\hat{\tilde{\mathbf{b}}}^{(D)} = 1. \tag{15}$$

We also need that the fitted density is nonnegative which yields the additional constraint

$$c(u_1,\ldots,u_p;\hat{\tilde{\mathbf{b}}}^{(D)}) \geq 0,\ u_j \in [0,1],\ j = 1,\ldots,p. \tag{16}$$

The constraints (14), (15) and (16) can be accommodated as side conditions in a quadratic programming tool to maximize the likelihood (10). We made use of the implemented version in `R` in the `quadprog` package. As a starting value for $\tilde{\mathbf{b}}$, we use a uniform distribution on the the cube $[0,1]^p$. This is easily obtained with the hierarchical B-spline basis. The knots are placed equidistantly. The entire procedure is implemented in the `R` package `pencopula` to be provided on the `CRAN` server (see *http://cran.r-project.org/*).

Note that (14) and (15) are simple equations. To satisfy constraint (16), we require the condition to hold at the $(2^d+1)^p$ equidistant knots locations of the tensor product

B-spline density basis. If p and d increase, the number of conditions and hence the computational effect of the quadratic program increase enormously, e.g. a full tensor product for $p = 4$ and $d = 4$ contains 83521 entries. With the following trick, we can reduce the calculation time without any loss of accuracy. The idea is, when calculating the constraints, to omit knot locations of the full tensor product where the density itself is high. This is incorporated in the algorithm in two ways. First, in the initial step we omit knot locations for the calculation of the constraint (16) which are close to the observations. In the subsequent steps, when density estimates in the iteration are available, we omit knot locations with a high value of the fitted density. Such reduction of the constraints accelerates the computation of the quadratic programming step. The final thing to adjust is the amount of penalization. In practice, we need to choose λ in a data-driven manner and in principle we need to select a separate λ_j for each dimension. To limit the numerical effort, however, we let $\lambda_1 = \lambda_2 = \cdots = \lambda_p$ and minimize the corrected Akaike information criterion (Hurvich & Tsai 1989, see also Burnham & Anderson 2010) defined as

$$\mathrm{AIC}_c(\boldsymbol{\lambda}) = -2\tilde{l}(\hat{\tilde{\mathbf{b}}}^{(D)}, \boldsymbol{\lambda}) + 2\mathrm{df}(\boldsymbol{\lambda}) + \frac{2\mathrm{df}(\boldsymbol{\lambda})(\mathrm{df}(\boldsymbol{\lambda}) + 1)}{n - \mathrm{df}(\boldsymbol{\lambda}) - 1} \tag{17}$$

where $\mathrm{df}(\boldsymbol{\lambda})$ is the degree of the model defined through

$$\mathrm{df}(\boldsymbol{\lambda}) = \mathrm{tr}\left[\left\{\tilde{\mathbf{H}}_p^{(D)}(\hat{\tilde{\mathbf{b}}}^{(D)}, \boldsymbol{\lambda})\right\}^{-1} \tilde{\mathbf{H}}_p^{(D)}\left(\hat{\tilde{\mathbf{b}}}^{(D)}, \boldsymbol{\lambda} = 0\right)\right].$$

where $\tilde{H}_p^{(D)}(.)$ is the second order derivative of the likelihood, see formula (22) in Appendix for details.

3 Examples and Simulations

3.1 Simulation

To get an impression of the performance of the routine, we simulated data from a given copula $c_0(\cdot)$, say, using the `copula` package in `R`; see Yan (2007). We thereby simulate data from different copulas in three correlation scenarios with Kendall's tau $\tau = 0.25$, $\tau = 0.5$ and $\tau = 0.75$, respectively. With respect to the copulas, we simulate data from (i) a Clayton copula, (ii) a Frank copula, (iii) a Gumbel copula and two

different t-copulas, (iv) a t-copula with 3 degrees of freedom, and (v) a t-copula with 4 degrees of freedom, each with sample size n=500. We simulate data in $p = 2, 3$ and 4 dimensions.

We fit the simulated data following our procedure and the performance is validated by analyzing the simulation mean of the corrected Akaike information criterion AIC_c of non-parametric estimators, denoted by $\widehat{AIC}_{np}$. The results are based on 200 simulations for $p = 2, 3$ and 100 simulations for $p = 4$ and shown in Table 2 for different values of d, the spline dimension, and D, the hierarchy order. The optimal smoothing parameter λ is selected with a simple grid search. Note that $d = 3$, $D = 6$ as well as $d = 4$, $D = 8$ refer to a full tensor product for $p = 2$. For comparison, we fit the data with a kernel density estimator using the quadratic Epanechnikov-kernel and optimal bandwidth selected with likelihood cross-validation. For fitting we use the R package `np` (see Hayfield & Racine 2008). The corresponding AIC_c is denoted as $\widehat{AIC}_{kernel}$, where we use the multivariate analogon of the univariate Akaike information criterion by Loader (1999). Furthermore, we fit the data with Bernstein polynomials as basis functions but without any penalization (see Sancetta & Satchell 2004). We use quadratic programming with the same side constraints as in our routine, that is imposing uniform margins. As basis dimensions of the Bernstein polynomial we use $3, 4, 5, \ldots, 10$. To avoid over-fitting we select the dimension of the basis again by the use of the corrected Akaike information criterion AIC_c. The corresponding AIC_c is denoted as $\widehat{AIC}_{bern}$. As an ultimate benchmark, we calculated the AIC_c value for the true copula from which we simulated the data but with their parameter replaced by its Maximum Likelihood fitted value, as implemented in `R` using the `copula` package. This value is denoted as $\widehat{AIC}_{true}$

Let us now look at the results in Table 2. First we investigate the two dimensional setting, i.e. $p = 2$, which is visualized in Figure 3 by plotting the distance to the optimal AIC_c for the different competitors. We start with the low correlation case, i.e. $\tau = 0.25$. The results of the full tensor product kernel $d = 4, D = 8$ yield optimal results for each copula scenario. Furthermore the sparse grid ($d = 3, D = 3$ and $d = 4, D = 4$) is slightly less efficient for this scenario, but shows comparably distances to the optimal AIC_c as the optimal full tensor product does. The kernel density

approach shows the largest difference to the optimal AIC_c in this case. Also, the Bernstein polynomials are outperformed with respect to the difference to the optimal AIC_c in this case. The picture changes slightly when looking at the stronger correlation $\tau = 0.5$. Again, the full tensor product for $d = 4, D = 8$ yields the best results with respect to the distance to optimal AIC_c followed by the the full tensor product for $d = 3, D = 6$ with slightly increased differences. Moreover the sparse grid ($d = 3, D = 3$ and $d = 4, D = 4$) performs weaker but still better than the kernel approach and the Bernstein polynomials, which have the highest distance to optimal AIC_c. In the case of the correlation $\tau = 0.75$, the full tensor product for d=4, D=8 yields the best results, again, whereas the small tensor product for d=3, D=3 is less efficient than the kernel estimator and comparable to the Bernstein polynomials. The sparse grid for d=4, D=4 outperforms the kernel approach, up to one exception, and is better than the Bernstein polynomials.

Next we look at dimension $p = 3$. The results are visualized in Figure 4. Note that for $p = 3$ all cases of our approach are sparse grids and the full tensor pruduct with e.g. $d = 4, D = 12$ would be numerically demanding, see also Table 1. Generally, for the small correlation case $\tau = 0.25$ (top) we see a tendency that the sparse grid fit outperforms both, the Bernstein polynomial fit and the kernel based fit. Looking at sparse grids using $d = 3, D = 6$ and $d = 4, D = 8$, we obtain the smallest distance to the optimal AIC_c. A similar picture is seen for the correlation case $\tau = 0.5$ and $\tau = 0.75$. The spars grids using $d = 3, D = 3$ and $d = 4, D = 4$ show comparable differences to optimal AIC_c, while for $\tau = 0.75$ the sparse grid using $d = 3, D = 6$ results with the smallest distance to optimal AIC_c.

Finally, considering the four dimensional case $p = 4$, we simulate from the Clayton, Frank and t-copula with 4 degrees of freedom. For the low correlation case $\tau = 0.25$ we observe the lowest distances to the optimal AIC_c for the sparse grids and the Bernstein polynomials and the kernel approach are outperformed. Looking at the stronger correlation $\tau = 0.5$ we observe a similar behaviour. The picture changes for $\tau = 0.75$, where the kernel estimator results with minimal distance to optimal AIC_c. Overall we can conclude that the sparse grid behaves mostly competitive, in particular for dimensions beyond 2.

Finally, looking at the computing time we list in Table 3 the CPU time for the sparse grid approach for different values of $d(= D)$ and dimensions $p = 2, 3, 4$. Again, though the computing time increases with p, calculation is still feasible for dimension $p = 4$.

3.2 Examples

Finally, we illustrate the applicability of the procedure with two examples. In both examples, we use t-distribution as univariate margins with maximum-likelihood theory estimated parameters. We present the results with smoothing parameter λ, chosen by AIC_c in Table 4.

First, we look at monthly interest rate data from the `R` package `Ecdat` using the data set `Capm`. The raw data are monthly risk-free interest rates which could be used to fit a Capital Asset Pricing Model (CAPM) We have jittered the data somewhat and created a bivariate sample by computing lagged rates and changes in rates. The data and the contour plot of the sparse grid-based fitted copula (left) and the corresponding copula density (right) are plotted in Figure 5, for $d = 5$ and $D = 5$. Note that the copula distribution function on $\bigtimes_{j=1}^{p}[0, 1]$ is easily calculated by taking the integrated B-spline densities weighted with the spline coefficients. The density shows a strong positive association between the lagged rate and the volatility of the rate change. Specifically, the density is high where the lagged rate and the magnitude of the rate change are either both small or both large. For comparison, we fitted the copula for different spline dimensions and also with a full tensor product and list the results in Table 4 (left). We show the maximum likelihood $\hat{l}$ and the Akaike Information criterion. Moreover we fit classical copula families and rotated versions of extreme value copula families to the data with maximum-likelihood theory estimated parameters. Also, we use Bernstein polynomials to construct the copula and choose the dimension of them by the Akaike Information Criterion. The results are shown in Table 4. Apparently, none of the parametric models are close to the results of the non-parametric approach and among the latter, the penalization spline estimators outperform the Bernstein polynomial estimators, using the AIC_c as the criterion.

As a second example, we investigate three daily world currency indices from January 3rd, 2000 until May 6th, 2011. The dataset includes values of $n = 2854$ business days

compared to the US-dollar. The data set includes the Australian dollar (AUS), the Euro (EUR) and the Japanese yen (YEN). We analyze the log-return from day t to day $t+1$. We present the results for this data set in Table 4 (right). For comparison we also fit parametric copula models to the data, also listed in Table 4. Note, a full tensor product for $p = 3$ is constructed with $d = 3, D = 9$ or $d = 4, D = 12$, but at least for $d = 4, D = 12$ the approach is not feasible due to the curse of dimensionality. Therefore, we fit the data with a compromise between the smallest sparse grid and the full tensor product, using $d = 3, D = 6$ and $d = 4, D = 8$. The greater sparse grids with $d = 3, D = 6$ and $d = 4, D = 8$ result with higher log-likelihood compared with the cases $d = 3, D = 3$ and $d = 4, D = 4$. The fitted t-Copula results with minimal AIC_c, but the AIC_c values of our approach are comparable and result with a higher log-likelihood. Our approach allows to analyze the bivariate margins of this estimated three dimensional copula. The contour plot of the fitted bivariate margins (left) with minimal AIC_c and the corresponding copula density (right) are plotted in Figure 6 with $d = 4$ and $D = 8$. We observe different dependencies among the bivariate margins. Obviously, the high peaks in $(0, 0)$ and $(1, 1)$ in the bivariate marginal copula of the Euro and the Australian dollar (Figure 6, right in the top row) indicate dependence between these currencies in the observation period, both currencies have risen or have fallen if one of them have risen or have fallen. The bivariate marginal copula of the Euro and the Japanese yen (Figure 6, right in the middle row) shows a different dependency. The bivariate marginal copula of the Australian dollar and the Japanese yen (Figure 6, right in the middle row) shows more complex behaviour, which is mirrored in the non-parametric fit.

4 Discussion

We propose in the paper how to fit copula densities with penalized B-splines. Our approach thereby accommodates side constraints like uniform univariate margins so that the fitted density is a copula density itself. The use of a reduced tensor product basis allows to extend the approach to higher dimensions by maintaining numerical feasibility. Apparently, the approach does not circumvent the curse of dimensionality,

but it shifts it a little bit so that calculation on 3, 4 (or 5) dimensions is possible. Moreover, we show (see Table 4), that the choices of d and D are not crucial, if they are chosen large enough to avoid substantial bias. The approach can be extended to higher dimensions by making use of further techniques as for instance pair copula estimation. Generally, the semi-parametric approach suggested in the paper contributes to the weakly development field of non- and semi-parametric copula estimation.

Acknowledgment

Göran Kauermann & Christian Schellhase acknowledge financial support provided by the Deutsche Forschungsgesellschaft. David Ruppert was supported by the National Science Foundation.

References

Autin, F., Pennec, E. Le, & Tribouley, K. 2010. Thresholding methods to estimate copula density. *J. Multivariate Anal.*, **101**(1), 200 – 222.

Bedford, Tim, & Cooke, Roger M. 2002. Vines: A New Graphical Model for Dependent Random Variables. *Ann. Statist.*, **30**(4), 1031–1068.

Bogaerts, Kris, & Lesaffre, Emmanuel. 2008. Modeling the association of bivariate interval-censored data using the copula approach. *Stat. Med.*, **27**(30), 6379–6392.

Bungartz, Hans-Joachim, & Griebel, Michael. 2004. Sparse grids. *Acta Numer.*, **13**, 147–269.

Burnham, K., & Anderson, D. R. 2010. *Model selection and multimodel inference - A practical Information-Theoretic Approach*. Springer, Berlin Heidelberg.

Chen, S. X., & Huang, T.-M. 2007. Nonparametric estimation of copula functions for dependence modelling. *Canad. J. Statist.*, **35**(2), 265–282.

Czado, Claudia. 2010. Pair-Copula Constructions of Multivariate Copulas. *Pages 93–109 of:* Bickel, P., Diggle, P., Fienberg, S., Gather, U., Olkin, I., Zeger, S., Jaworski,

Piotr, Durante, Fabrizio, Härdle, Wolfgang Karl, & Rychlik, Tomasz (eds), *Copula Theory and Its Applications*. Lecture Notes in Statistics, vol. 198. Springer, Berlin Heidelberg.

Danaher, P. J., & Smith, M. S. 2011. Modeling multivaraite distributions using copulas: applications in marketing. *Marketing Science*, **30**(1), 4–21.

Eilers, Paul H. C., & Marx, Brian D. 1996. Flexible Smoothing with B-splines and Penalties. *Statist. Sci.*, **11**(2), 89–121.

Embrechts, P. 2009. Copulas: A personal view. *Journal of Risk and Insurance*, **76**(3), 639–650.

Fermanian, J.-D., & Scaillet, O. 2003. Nonparametric estimation of copulas for time series. *Journal of Risk*, **5**(4).

Fermanian, J.-D., Radulovic, D., & Wegkamp, M.J. 2004. Weak convergence of empirical copula processes. *Bernoulli*, **10**(5), 847–860.

Forsey, David R., & Bartels, Richard H. 1988. Hierarchical B-spline refinement. *Pages 205–212 of: SIGGRAPH '88: Proceedings of the 15th annual conference on Computer graphics and interactive techniques*. New York, NY, USA: ACM.

Forsey, David R., & Bartels, Richard H. 1995. Surface fitting with hierarchical splines. *ACM Transactions on Graphics*, **14**(2), 134–161.

Garcke, Jochen. 2006. *Sprase Grid Tutorial*. Tech. rept. Centre for mathematics and its applications, Mathematical Sciences Institute, Australien National University, Canberra.

Genest, C., Masiello, E., & Tribouley, K. 2009. Estimating copula densities through wavelets. *Insurance Math. Econom.*, **44**(2), 170–181.

Gijbels, I., & Mielniczuk, J. 1990. Estimation the Density of a Copula Function. *Comm. Statist. Theory Methods*, **19**(2), 445–464.

Härdle, W., & Okhrin, O. 2009. De copulis non est disputandum - Copulae: an Overview. *AStA Adv. Stat. Anal.*, **94**(1), 1–31.

Hayfield, Tristen, & Racine, Jeffrey S. 2008. Nonparametric Econometrics: The np Package. *Journal of Statistical Software*, **27**(5).

Hoeffding, W. 1940. Masstabinvariante Korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, **5**, 179–233.

Hurvich, Clifford M., & Tsai, Chih-Ling. 1989. Regression and time series model selection in small samples. *Biometrika*, **76**(2), 297–307.

Jaworski, Piotr, Durante, F., Härdle, W.K., & Rychlik, T. 2010. *Copula Theory and Its Applications*. Lecture notes in statistics ; 198 : Proceedings. Springer.

Joe, Harry. 1996. Families of m-Variate Distributions with Given Margins and m(m-1)/2 Bivariate Dependence Parameters. *Lecture Notes-Monograph Series*, **28**, 120–141.

Kauermann, G., Krivobokova, T., & Fahrmeir, L. 2009. Some asymptotic results on generalized penalized spline smoothing. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **71**(2), 487–503.

Kolev, N., Anjos, U., & Mendes, B.V.M. 2006. Copulas: a review and recent developments. *Stoch. Models*, **22**(4), 617–660.

Koo, J.-Y. 1996. Bivariate B-splines for Tensor logspline density estimation. *Comput. Statist. Data Anal.*, **21**(1), 31–42.

Lambert, Philippe. 2007. Archimedean copula estimation using Bayesian splines smoothing techniques. *Comput. Statist. Data Anal.*, **51**(12), 6307–6320.

Loader, C. 1999. *Local Regression and Likelihood*. Springer.

Marx, B.D, & Eilers, P. H. C. 2005. Multidimensional penalized signal regression. *Technometrics*, **47**, 13–22.

McNeil, A., Frey, R., & Embrechts, P. 2005. *Quantitative Risk Management*. Princeton University Press, Princeton Series in Finance.

Morettin, P.A., Toloi, C.M.C., Chiann, C., & Miranda, J. C. 2010. Wavelet smoothed empirical copula estimations. *Brazilian Review of Finance*, **8**(3), 263–281.

Nelsen, R. 2006. *An introduction to copulas*. second edn. Berlin: Springer.

Okhrin, O., Okhrin, Y., & Schmid, W. 2009. *Properties of the hierarchical Archimedean copulas*. Tech. rept. Humboldt Universität zu Berlin, SFB 649 Discussion Paper 2009-014.

Omelka, M., Gijbels, I., & Veraverbeke, N. 2009. Improved kernel estimation of copulas: Weak convergence and goodness-of-fit testing. *Ann. Statist.*, **37**(5B), 3023–3058.

Pfeifer, D, Straßburger, D, & Philipps, J. 2009. Modelling and simulation of dependence structures in nonlife insurance with Bernstein copulas. *Preprint*.

Qu, Leming, & Yin, Wotao. 2012. Copula density estimation by total variation penalized likelihood with linear equality constraints. *Comput. Statist. Data Anal.*, **56**(2), 384 – 398.

Qu, Leming, Qian, Yi, & Xie, Hui. 2009. Copula density estimation by total variation penalized likelihood. *Comm. Statist. Simulation Comput.*, **38**(9), 1891–1908.

Ruppert, D., Wand, M.P., & Carroll, R.J. 2003. *Semiparametric Regression*. Cambridge University Press.

Ruppert, D., Wand, M. P., & Carroll, R. J. 2009. Semiparametric regression during 2003 – 2007. *Electron. J. Stat.*, **3**, 1193–1256.

Sancetta, Alessio, & Satchell, Stephen. 2004. Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory*, **20**(3), 535–562.

Savu, Cornelia, & Trede, Mark. 2010. Hierarchies of Archimedean copulas. *Quant. Finance*, **10**(3), 295–304.

Shen, X., Zhu, Y., & Song, L. 2008. Linear B-spline copulas with applications to nonparametric estimation of copulas. *Comput. Statist. Data Anal.*, **52**(7), 3806–3819.

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

Song, P., Mingyao, L., & Yuan, Y. 2009. Joint regression analysis of correlated Data Using Gaussian Copulas. *Biometrics*, **65**, 60–68.

Wand, M. P., & Ormerod, J. T. 2008. On semiparametric regression with O'Sullivan penalised splines. *Aust. N. Z. J. Stat.*, **50**(2), 179–198.

Yan, Jun. 2007. Enjoy the joy of copulas: with a package copula. *Journal of Statistical Software*, **21**(4), 1–21.

Zenger, Christoph. 1991. Sparse Grids. *Notes Numer. Fluid Mech. Multidiscip. Des.*, **31**, 241–251.

Göran Kauermann, Department of Statistics, Ludwig-Maximilians-University Munich, Ludwigstraße 33, D-80539 München.
E-mail: Goeran.Kauermann@stat.uni-muenchen.de

Appendix: Proofs of Lemmas and technical details

Proof of Lemma 1. Let $c^{(D)}(\mathbf{u};\mathbf{b}) = \tilde{\mathbf{\Phi}}^{(D)}_{(d)}(\mathbf{u})\tilde{\mathbf{b}}^{(D)}$ denote the sparse grid B-spline representation of the true copula density $c(\mathbf{u})$. We assume that $c(\mathbf{u})$ is continuously differentiable, and we denote with $\tilde{\mathbf{b}}_0^{(D)}$ the true parameter in the sense that $c^{(D)}(\mathbf{u};\tilde{\mathbf{b}}_0^{(D)})$ and $c(\mathbf{u})$ have smallest Kullback-Leibler distance with $\tilde{\mathbf{b}}_0^{(D)}$ fulfilling constraint (15). This implies that vector $\tilde{\mathbf{b}}_0^{(D)}$ minimizes the Lagrange function

$$E\left\{\log\big(c^{(D)}(\mathbf{u};\tilde{\mathbf{b}}^{(D)})\big)\right\} + \rho(\mathbf{1}^T\tilde{\mathbf{b}}^{(D)} - 1) \tag{18}$$

with ρ as the Lagrange multiplier. Differentiation of (18) with respect to $\tilde{\mathbf{b}}^{(D)}$ yields

$$\int \tilde{\mathbf{\Phi}}^T(\mathbf{u})\frac{c(\mathbf{u})}{c^{(D)}(\mathbf{u};\tilde{\tilde{\mathbf{b}}}^{(D)})}\mathrm{d}\mathbf{u} = \rho\mathbf{1} \tag{19}$$

where $\rho = 1$ results from multiplying (19) from the left hand side with $\tilde{\mathbf{b}}_0^{(D)^T}$. Using definition (9), we find the components of $\tilde{\mathbf{\Phi}}(\mathbf{u})$ to have the form $\prod_{j=1}^p \phi_{(h_j)l_j}(u_j)$ with

$l_j \in \mathcal{I}_{(h_j)}$ and $\sum_{j=1}^p h_j \leq D$ where $h_j \geq 0$. We naturally assume that $D \geq d$ and define with $r(\mathbf{u}) = c(\mathbf{u})/c^{(D)}(\mathbf{u}; \tilde{\mathbf{b}}_0^{(D)})$ the ratio of the true and the approximate copula density. With (19) we get for a single component in $\tilde{\mathbf{\Phi}}(\mathbf{u})$

$$\begin{aligned} 1 &= \int_0^1 \phi_{(h_1)l_1}(u_1) \left\{ \int_{\times_{j=2}^p [0,1]} \prod_{j=2}^p \phi_{(h_j)l_j}(u_j) r(u_1, \ldots, u_p) \mathrm{d}u_2 \ldots \mathrm{d}u_p \right\} \mathrm{d}u_1 \\ &= \int_{U_{(h_1)l_1}} \phi_{(h_1)l_1}(u_1) r_1(u_1) \mathrm{d}u_1 \end{aligned} \tag{20}$$

where $r_1(u_1)$ denotes the bracketed term in (20) and $U_{(h_j)l_1}$ is the support of basis $\phi_{(h_1)l_1}$. Following the mean value theorem for integration we find a value $\tilde{u}_{(h_1)l_1} \in U_{(h_1)l_1}$ so that $r_1(\tilde{u}_{(h_1)l_1}) = 1$. This allows to recursively apply the same argument to the bracketed term in (20). Let $h_1^{(D)} = D - \sum_{j=2}^p h_j$, then condition (20) holds for all $h_1 \leq h_1^{(D)}$. Since $\left\{ \phi_{(h_1)\mathcal{I}_{h_1}}(u_1), h_1 \leq h_1^{(D)} \right\}$ spans the linear space of $\Phi_{\left(h_1^{(D)}\right)}(u_1)$ of the non hierarchical B-spline basis of order h_1 we obtain that for all $u_1 \in [0,1]$ there exists a $\tilde{u}_1$ with $|u_1 - \tilde{u}_1| \leq 2^{-h_1^{(D)}}$ and $r_1(\tilde{u}_1) = 1$. Applying the same argument recursively we get the final result that for all $\mathbf{u} \in [0,1]^p$ there exists a $\tilde{\mathbf{u}}$ with $||\mathbf{u} - \tilde{\mathbf{u}}|| \leq 2^{-D}$ and $c(\mathbf{u}) = c^{(D)}(\tilde{\mathbf{u}}; \mathbf{b}^{(D)})$. With simple Taylor approximation we therefore obtain

$$c(\mathbf{u}) = c^{(D)}(\tilde{\mathbf{u}}; \mathbf{b}^{(D)}) + O(2^{-D}).$$

Proof of Lemma 2. Let $\hat{\tilde{\mathbf{b}}}$ denote the penalized Maximum Likelihood estimate based on (10) and let $\tilde{\mathbf{s}}_p^{(D)}(\tilde{\mathbf{b}}^{(D)}, \boldsymbol{\lambda})$ and $\tilde{\mathbf{H}}_p^{(D)}(\tilde{\mathbf{b}}^{(D)}, \boldsymbol{\lambda})$ be the first and second order derivatives of $\tilde{l}_p^{(D)}(\tilde{\mathbf{b}}^{(D)}, \boldsymbol{\lambda})$, respectively, i.e.,

$$\tilde{\mathbf{s}}_p^{(D)}(\tilde{\mathbf{b}}^{(D)}, \boldsymbol{\lambda}) = \sum_{i=1}^n \frac{\mathbf{\Phi}_{(d)}^{(D)}(u_i)}{c^{(D)}(u_i, \tilde{\mathbf{b}}^{(D)})} - \tilde{\mathbf{P}}^{(D)}(\boldsymbol{\lambda}) \tilde{\mathbf{b}}^{(D)} \tag{21}$$

$$\tilde{\mathbf{H}}_p^{(D)}(\tilde{\mathbf{b}}^{(D)}, \boldsymbol{\lambda}) = -\sum_{i=1}^n \frac{\mathbf{\Phi}_{(d)}^{(D)}(u_i) \mathbf{\Phi}_{(d)}^{(D)^T}(u_i)}{c^{(D)}(u_i, \tilde{\mathbf{b}}^{(D)})} - \tilde{\mathbf{P}}^{(D)}(\boldsymbol{\lambda}) \tag{22}$$

where $c^{(D)}(u_i, \tilde{\mathbf{b}}^{(D)}) = \mathbf{\Phi}_{(d)}^{(D)}(u_i) \tilde{\mathbf{b}}^{(D)}$. Denote with $\tilde{\mathbf{b}}_0^{(D)}$ the 'true' spline coefficient vector, in the sense that the true copula density $c(\mathbf{u})$ and $c^{(D)}(\mathbf{u}, \tilde{\mathbf{b}}_0^{(D)})$ have smallest Kullback-Leibler distance. This defines $\tilde{\mathbf{b}}_0^{(D)}$ implicitly through $E\left\{ \tilde{\mathbf{s}}_p(\tilde{\mathbf{b}}_0^{(D)}, \boldsymbol{\lambda} = 0) \right\} = 0$. For the solution of $\tilde{s}_p^{(D)}(\hat{\tilde{\mathbf{b}}}^{(D)}, \boldsymbol{\lambda}) = 0$, we get with simple regular expansion techniques

(see e.g., Kauermann *et al.* (2009))

$$\hat{\tilde{\mathbf{b}}}^{(D)} - \tilde{\mathbf{b}}_0^{(D)} = -\boldsymbol{H}_p^{(D)^{-1}}(\tilde{\mathbf{b}}_0^{(D)}, \boldsymbol{\lambda})\mathbf{s}_p^{(D)}(\tilde{\mathbf{b}}_0^{(D)}, \boldsymbol{\lambda}) + \ldots$$

which allows us to derive asymptotic statements about the estimates for $n \to \infty$ and D fixed. In fact, applying the central limit theorem we can derive asymptotic normality of $\hat{\tilde{\mathbf{b}}}^{(D)}$ with mean and variance asymptotically equal to

$$\begin{aligned} \mathrm{E}(\hat{\tilde{\mathbf{b}}}^{(D)}) &= \tilde{\mathbf{b}}_0^{(D)} + \left\{\mathbf{H}_p^{(D)}(\tilde{\mathbf{b}}_0^{(D)}, \boldsymbol{\lambda})\right\}^{-1} \tilde{\mathbf{P}}^{(D)}(\boldsymbol{\lambda})\tilde{\mathbf{b}}_0^{(D)} \\ \mathrm{Var}(\hat{\tilde{\mathbf{b}}}^{(D)}) &= \left\{\mathbf{H}_p^{(D)}(\tilde{\mathbf{b}}_0^{(D)}, \boldsymbol{\lambda})\right\}^{-1} \mathbf{H}_p^{(D)}\left(\tilde{\mathbf{b}}_0^{(D)}, \boldsymbol{\lambda} = 0\right)\left\{\mathbf{H}_p^{(D)}(\tilde{\mathbf{b}}_0^{(D)}, \boldsymbol{\lambda})\right\}^{-1}. \end{aligned}$$

Calculation of Marginal Spline Coefficients. Let the marginal hierarchical basis $\tilde{\Phi}_{(d)}(u)$ in (8) be indexed by $\{\tilde{\phi}_{(d)l}(\cdot), l = 1, \ldots, K\}$, and let $\tilde{h} = (\tilde{h}_l, l = 1, \ldots, K)$ denote the hierarchy level of $\tilde{\phi}_{(d)l}(u)$, that is, $\tilde{\phi}_{(d)l}(u)$ is element of $\tilde{\Phi}_{(\tilde{h}_l)\mathcal{I}_{\tilde{h}_l}}$. For instance, looking at Figure 1 (or 2), the hierarchy levels for the hierarchical bases built from (b), (c), and (d) are 0, 1, and 2, respectively. The sparse grid basis $\tilde{\mathbf{\Phi}}_{(d)}^{(D)}(\mathbf{u})$ in (9) can now be indexed as

$$\left\{\prod_{j=1}^{p} \tilde{\phi}_{(d)l_j}(u_j), \sum_{j=1}^{p} \tilde{h}_{l_j} \leq D,\ l_j = 1, \ldots, K\right\}$$

and accordingly we index the spline coefficient vector with $\tilde{\mathbf{b}}^{(D)} = (\tilde{b}_{l_1,\ldots,l_p}^{(D)}; \sum_{j=1}^{p} \tilde{h}_{l_j} \leq D)$. As a result, the marginal density for u_j is as follows. Let $\mathrm{d}u_{-j}$ denote the integral measure $\prod_{m \neq j} \mathrm{d}u_m$, then

$$\int_{\substack{p \\ \times \\ i \neq j}\,[0,1]} \tilde{\mathbf{\Phi}}_{(d)}^{(D)}(u_1, \ldots, u_p)\tilde{\mathbf{b}}^{(D)}\mathrm{d}u_{-j} = \sum_{l_j=1}^{K} \tilde{\phi}_{(d)l_j}(u_j)\tilde{b}_{(j)l_j}^{(D)} =: \tilde{\Phi}_{(d)}(u_j)\tilde{b}_{(j)}^{(D)} \tag{23}$$

with $\tilde{\Phi}_{(d)}(\cdot)$ as hierarchical marginal basis defined in (8). The elements of coefficient vector $\tilde{b}_{(j)}^{(d)}$ result from the $p-1$ dimensional sum

$$b_{(j)l_j}^{(D)} = \sum_{l_{-j}:\sum_{m \neq j} l_m \leq D} \tilde{b}_{l_1,\ldots,l_p}^{(D)} \tag{24}$$

where l_{-j} denote the sum over all l_m with $m \neq j$. Note that (24) is a simple linear calculation.

Table 1: Dimension of tensor product basis $\tilde{\mathbf{\Phi}}_{(d)}(u_1, \ldots, u_p)$ (full tensor product) and reduced sparse hierarchical basis $\tilde{\mathbf{\Phi}}_{(d)}^{(D)}(u_1, \ldots, u_p)$ with D set equal to d for $q = 1$, i.e., linear B-splines.

$d = D$	basis	$p = 2$	$p = 3$	$p = 4$	$p = 5$
3 ($K = 9$)	tensor prod. ($D = dp$)	81	729	6,561	59,049
	sparse ($D = d$)	37	123	368	1,032
4 ($K = 17$)	tensor prod. ($D = dp$)	289	4,913	83,521	1,419,857
	sparse ($D = d$)	81	297	961	2,882
5 ($K = 33$)	tensor prod. ($D = dp$)	1,089	35,937	1,185,921	39,135,393
	sparse ($D = d$)	177	705	2,441	7,763

Table 2: Reported is the mean (sd) of the AIC_c. The optimal results are set in bold.

	copula	true	$d=3$		$d=4$		Bernstein	kernel
		AIC_{true}	$D=3$	$D=6$	$D=4$	$D=8$	AIC_{bern}	Epanechnikov AIC_{kernel}
$p=2$	(i) Clayton $\tau=0.25$	-107.01 (22.38)	-69.85 (16.67)	-71.66 (17.40)	-70.47 (17.34)	**-72.22 (17.89)**	-65.09 (15.51)	19.33 (23.05)
	(i) Clayton $\tau=0.50$	-427.26 (38.25)	-288.87 (23.01)	-330.31 (32.40)	-321.50 (29.39)	**-339.61 (46.75)**	-262.14 (22.89)	-262.74 (38.64)
	(i) Clayton $\tau=0.75$	-1085.49 (48.37)	-573.86 (19.11)	-834.62 (90.55)	-760.01 (26.80)	**-894.93 (71.08)**	-528.94 (18.89)	-804.61 (73.78)
	(ii) Frank $\tau=0.25$	-72.70 (15.94)	-60.93 (14.33)	-61.19 (14.37)	**-61.97 (15.06)**	-61.19 (14.55)	-58.37 (16.03)	46.34 (25.67)
	(ii) Frank $\tau=0.50$	-315.72 (28.86)	-276.43 (26.21)	-279.60 (32.83)	-277.65 (25.89)	**-290.42 (28.25)**	-260.35 (23.65)	-186.22 (35.77)
	(ii) Frank $\tau=0.75$	-879.89 (37.05)	-599.72 (18.03)	-787.67 (47.10)	-777.90 (32.42)	**-839.25 (42.72)**	-545.22 (19.39)	-736.07 (53.25)
	(iii) Gumbel $\tau=0.25$	-94.19 (19.38)	-60.51 (15.25)	-62.30 (15.62)	-60.92 (15.51)	**-62.66 (15.87)**	-57.76 (14.86)	35.44 (21.04)
	(iii) Gumbel $\tau=0.50$	-374.33 (34.29)	-276.38 (25.09)	-302.60 (30.24)	-295.31 (29.20)	**-309.89 (31.91)**	-258.99 (24.66)	-221.33 (34.27)
	(iii) Gumbel $\tau=0.75$	-989.99 (45.75)	-556.91 (19.26)	-802.29 (89.95)	-760.61 (31.41)	**-856.90 (54.82)**	-548.24 (21.61)	-755.31 (64.27)
	(iv) tcop $df=3, \tau=0.25$	-119.29 (25.59)	-69.17 (20.39)	-74.25 (21.27)	-71.48 (20.95)	**-75.44 (21.52)**	-62.36 (17.93)	12.32 (25.16)
	(iv) tcop $df=3, \tau=0.50$	-390.74 (43.55)	-272.07 (29.82)	-307.75 (37.40)	-300.02 (36.21)	**-319.51 (39.84)**	-256.34 (29.66)	-223.74 (42.62)
	(iv) tcop $df=3, \tau=0.75$	-975.41 (57.08)	-532.09 (23.16)	-694.03 (121.08)	-738.45 (40.19)	**-841.85 (70.15)**	-533.29 (26.21)	-657.23 (120.00)
	(v) tcop $df=4, \tau=0.25$	-102.52 (21.86)	-62.97 (17.10)	-66.56 (17.63)	-64.36 (17.46)	**-67.53 (17.80)**	-59.75 (15.71)	25.60 (25.18)
	(v) tcop $df=4, \tau=0.50$	-376.86 (38.80)	-275.60 (29.48)	-304.79 (34.86)	-297.81 (33.47)	**-312.76 (42.46)**	-260.91 (28.51)	-214.26 (41.80)
	(v) tcop $df=4, \tau=0.75$	-964.11 (51.21)	-541.45 (21.27)	-730.18 (108.93)	-746.95 (35.66)	**-844.46 (55.32)**	-541.21 (24.58)	-683.92 (87.83)
$p=3$	(i) Clayton $\tau=0.25$	-273.40 (37.94)	-174.14 (25.85)	-180.14 (28.15)	-177.79 (26.69)	**-185.16 (29.68)**	-151.76 (23.89)	14.45 (37.46)
	(i) Clayton $\tau=0.50$	-974.26 (67.70)	-624.11 (39.54)	**-714.50 (47.27)**	-662.98 (113.29)	-693.25 (123.01)	-508.62 (33.75)	-579.69 (60.68)
	(i) Clayton $\tau=0.75$	-2332.59 (85.64)	-1191.32 (31.19)	-1573.81 (177.50)	-1413.24 (113.48)	-1517.58 (113.23)	-890.73 (28.35)	**-1666.51 (108.51)**
	(ii) Frank $\tau=0.25$	-192.99 (27.84)	-163.18 (25.81)	-173.13 (26.95)	-165.01 (26.25)	**-176.21 (27.80)**	-143.22 (25.94)	58.60 (38.50)
	(ii) Frank $\tau=0.50$	-747.08 (48.07)	-625.93 (40.03)	-697.03 (43.33)	-657.44 (49.28)	**-709.70 (74.74)**	-533.13 (32.25)	-474.06 (52.78)
	(ii) Frank $\tau=0.75$	-1922.29 (56.24)	-1240.49 (28.77)	**-1667.60 (47.82)**	-1438.24 (115.65)	-1545.39 (112.00)	-916.99 (26.61)	-1609.71 (99.79)
	(iii) Gumbel $\tau=0.25$	-247.64 (35.96)	-159.77 (25.31)	-168.13 (24.44)	-163.37 (26.03)	**-174.40 (25.86)**	-139.78 (23.91)	27.76 (35.72)
	(iii) Gumbel $\tau=0.50$	-876.30 (59.10)	-613.29 (39.78)	-693.67 (45.92)	-648.01 (53.34)	**-729.04 (55.94)**	-518.21 (34.66)	-525.27 (55.88)
	(iii) Gumbel $\tau=0.75$	-2159.20 (78.43)	-1161.39 (35.10)	**-1623.68 (42.98)**	-1401.33 (160.28)	-1534.66 (118.77)	-907.00 (32.71)	-1619.74 (100.70)
	(iv) tcop $df=3, \tau=0.25$	-299.08 (37.95)	-171.96 (26.35)	-185.16 (26.53)	-178.98 (27.53)	**-195.71 (29.29)**	-141.52 (23.23)	12.05 (37.86)
	(iv) tcop $df=3, \tau=0.50$	-896.82 (62.32)	-588.61 (41.44)	-691.52 (54.46)	-647.94 (48.92)	**-724.21 (99.85)**	-499.84 (38.24)	-473.40 (67.63)
	(iv) tcop $df=3, \tau=0.75$	-2102.88 (82.47)	-1114.81 (36.63)	**-1568.85 (61.45)**	-1394.85 (73.43)	-1536.88 (96.52)	-880.99 (30.61)	-1293.54 (201.34)
	(v) tcop $df=4, \tau=0.25$	-261.47 (36.83)	-158.86 (26.85)	-170.10 (32.20)	-163.74 (27.83)	**-176.65 (34.63)**	-137.83 (25.17)	39.48 (36.90)
	(v) tcop $df=4, \tau=0.50$	-859.93 (64.03)	-589.89 (44.97)	-683.12 (58.14)	-641.70 (49.90)	**-718.80 (71.10)**	-505.96 (41.39)	-448.67 (67.88)
	(v) tcop $df=4, \tau=0.75$	-2066.19 (85.21)	-1126.39 (39.17)	**-1579.61 (61.89)**	-1381.68 (112.71)	-1538.19 (119.57)	-888.38 (34.19)	-1353.48 (172.46)
$p=4$	(i) Clayton $\tau=0.25$	-462.05 (56.02)	-278.72 (36.68)	-	**-288.16 (37.57)**	-	-164.85 (37.33)	24.82 (54.77)
	(i) Clayton $\tau=0.50$	-1576.78 (95.39)	-886.38 (47.02)	-	**-916.73 (89.81)**	-	-716.17 (44.24)	-885.47 (81.09)
	(i) Clayton $\tau=0.75$	-3661.15 (138.07)	-1555.99 (146.84)	-	-1686.40 (136.45)	-	-1157.98 (39.73)	**-2502.95 (130.87)**
	(ii) Frank $\tau=0.25$	-346.10 (34.68)	-276.96 (30.91)	-	**-285.13 (32.66)**	-	-162.40 (31.57)	80.21 (46.67)
	(ii) Frank $\tau=0.50$	-1232.39 (61.67)	**-959.32 (50.03)**	-	-940.42 (125.61)	-	-765.26 (42.94)	-773.73 (72.53)
	(ii) Frank $\tau=0.75$	-3037.17 (89.62)	-1610.96 (49.12)	-	-1746.43 (145.76)	-	-1179.66 (44.09)	**-2539.12 (145.88)**
	(v) tcop $df=4, \tau=0.25$	-456.45 (54.97)	-263.87 (34.45)	-	**-280.39 (37.33)**	-	-155.55 (32.49)	78.12 (58.53)
	(v) tcop $df=4, \tau=0.50$	-1409.14 (87.22)	-895.12 (52.63)	-	**-914.93 (121.91)**	-	-717.53 (41.36)	-657.75 (110.54)
	(v) tcop $df=4, \tau=0.75$	-3225.57 (106.42)	-1477.17 (194.46)	-	-1593.85 (181.12)	-	-1134.68 (38.14)	**-2003.64 (187.47)**

Table 3: Elapsed `system.time` for a Frank copula with $N = 500$ observations.

$d = D$	$p = 2$	$p = 3$	$p = 4$
3 ($K = 9$)	1.063	2.020	13.652
4 ($K = 17$)	4.017	11.081	175.251

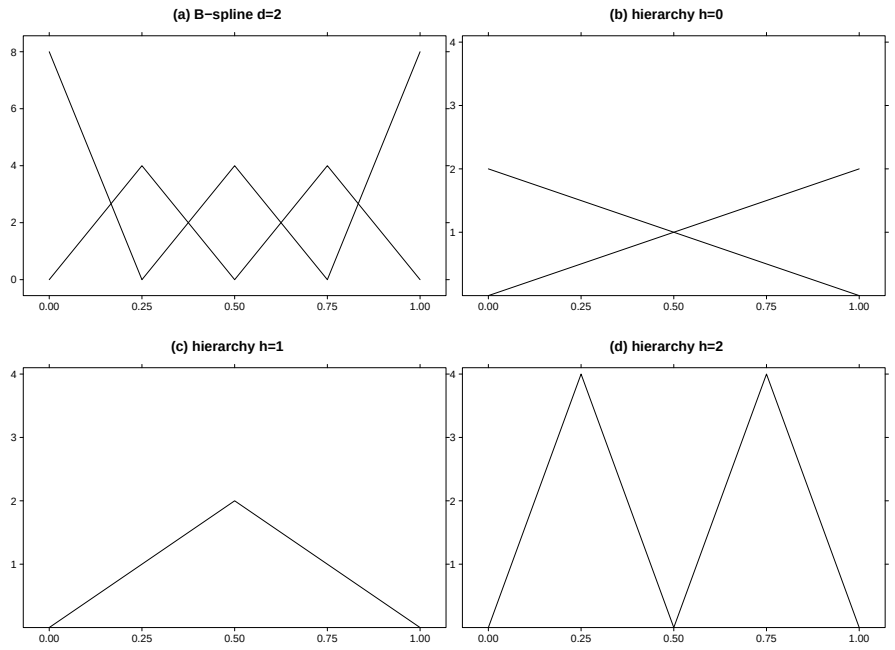


Figure 1: (a) B-spline density basis and corresponding hierarchical B-spline density basis ((b),(c),(d)) with different hierarchy levels.

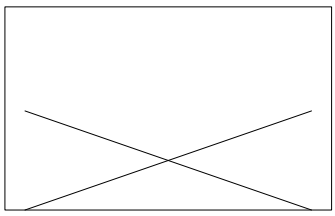
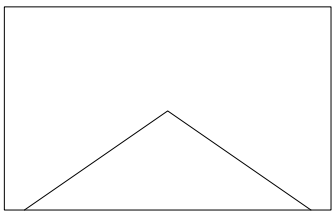
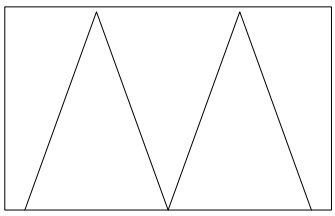
	 $\Phi_{(0)\mathcal{I}_0}(u_1)$	 $\Phi_{(1)\mathcal{I}_1}(u_1)$	 $\Phi_{(2)\mathcal{I}_2}(u_1)$
$\Phi_{(0)\mathcal{I}_0}(u_2)$	$\Phi_{(0)\mathcal{I}_0}(u_1) \otimes \Phi_{(0)\mathcal{I}_0}(u_2)$	$\Phi_{(1)\mathcal{I}_1}(u_1) \otimes \Phi_{(0)\mathcal{I}_0}(u_2)$	$\Phi_{(2)\mathcal{I}_2}(u_1) \otimes \Phi_{(0)\mathcal{I}_0}(u_2)$
$\Phi_{(1)\mathcal{I}_1}(u_2)$	$\Phi_{(0)\mathcal{I}_0}(u_1) \otimes \Phi_{(1)\mathcal{I}_1}(u_2)$	$\Phi_{(1)\mathcal{I}_1}(u_1) \otimes \Phi_{(1)\mathcal{I}_1}(u_2)$	**omitted**
$\Phi_{(2)\mathcal{I}_2}(u_2)$	$\Phi_{(0)\mathcal{I}_0}(u_1) \otimes \Phi_{(2)\mathcal{I}_2}(u_2)$	**omitted**	**omitted**

Figure 2: Representation of $\tilde{\mathbf{\Phi}}^{(2)}_{(2)}(u_1, u_2)$ for two dimensions ($p = 2$).

Table 4: Results for various combinations of d and D for data examples in Section 3.2, compared with results fitting maximum likelihood based optimal parameters for classical copula families and Bernstein polynomials choosing the dimension by the Akaike Information Criterion. Clayton$_x$ and Gumbel$_x$ are copulas rotated by x degrees.

		`Capm` data		exchange rate data	
d	D	log-likelihood $\hat{l}$	AIC_c	log-likelihood $\hat{l}$	AIC_c
3	3	40.343	-51.162	873.980	-1610.068
3	6	50.932	-55.714	1007.578	-1725.735
4	4	43.983	-52.412	978.359	-1707.725
4	8	57.361	-57.077	1117.326	-1774.491
5	5	46.202	-53.209	-	-
5	10	60.598	-58.556	-	-
Frank		2.811	-3.654	2.707	-3.412
Normal		3.990	-5.972	27.654	-53.307
t-Copula		23.127	-44.239	896.181	-1784.283
Clayton		19.008	-36.007	83.410	-164.819
Clayton$_{90}$		17.906	-33.804	-	-
Clayton$_{180}$		0.783	0.441	-	-
Clayton$_{270}$		1.690	-1.371	-	-
Gumbel		1.391	-0.775	31.649	-61.296
Gumbel$_{90}$		17.578	-33.148	-	-
Gumbel$_{180}$		0.000	2.008	-	-
Gumbel$_{270}$		1.476	-0.944	-	-
Bernstein		34.417	-36.833	886.640	-1523.279

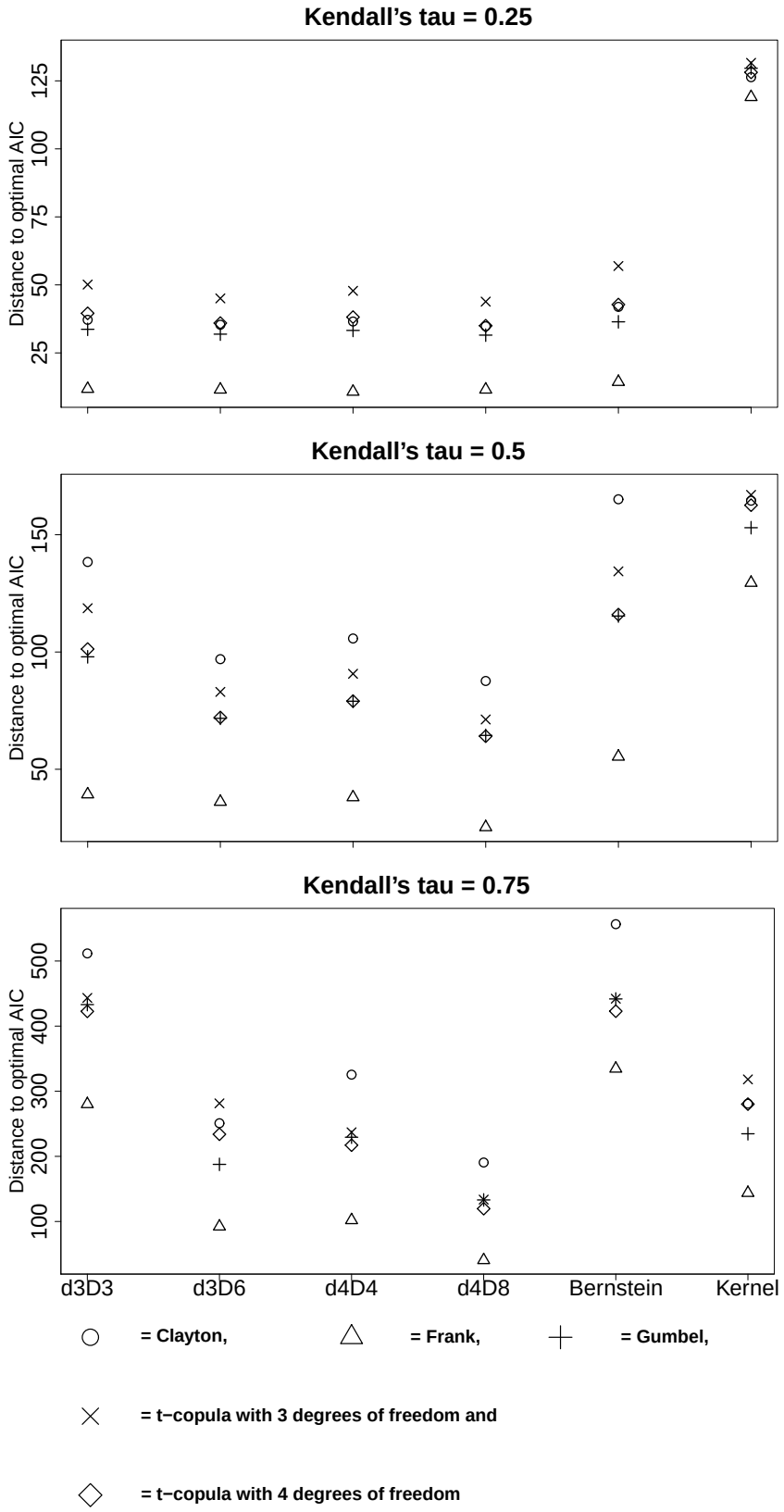


Figure 3: Simulated AIC difference $\widehat{AIC} - AIC_{true}$ for $p = 2$. From left to right: $\widehat{AIC}_{np} - AIC_{true}$ for $d = 3, D = 3$ and $d = 3, D = 6$ and $d = 4, D = 4$ and $d = 4, D = 8$, respectively, $\widehat{AIC}_{bernstein} - AIC_{true}$ and finally $\widehat{AIC}_{kernel} - AIC_{true}$

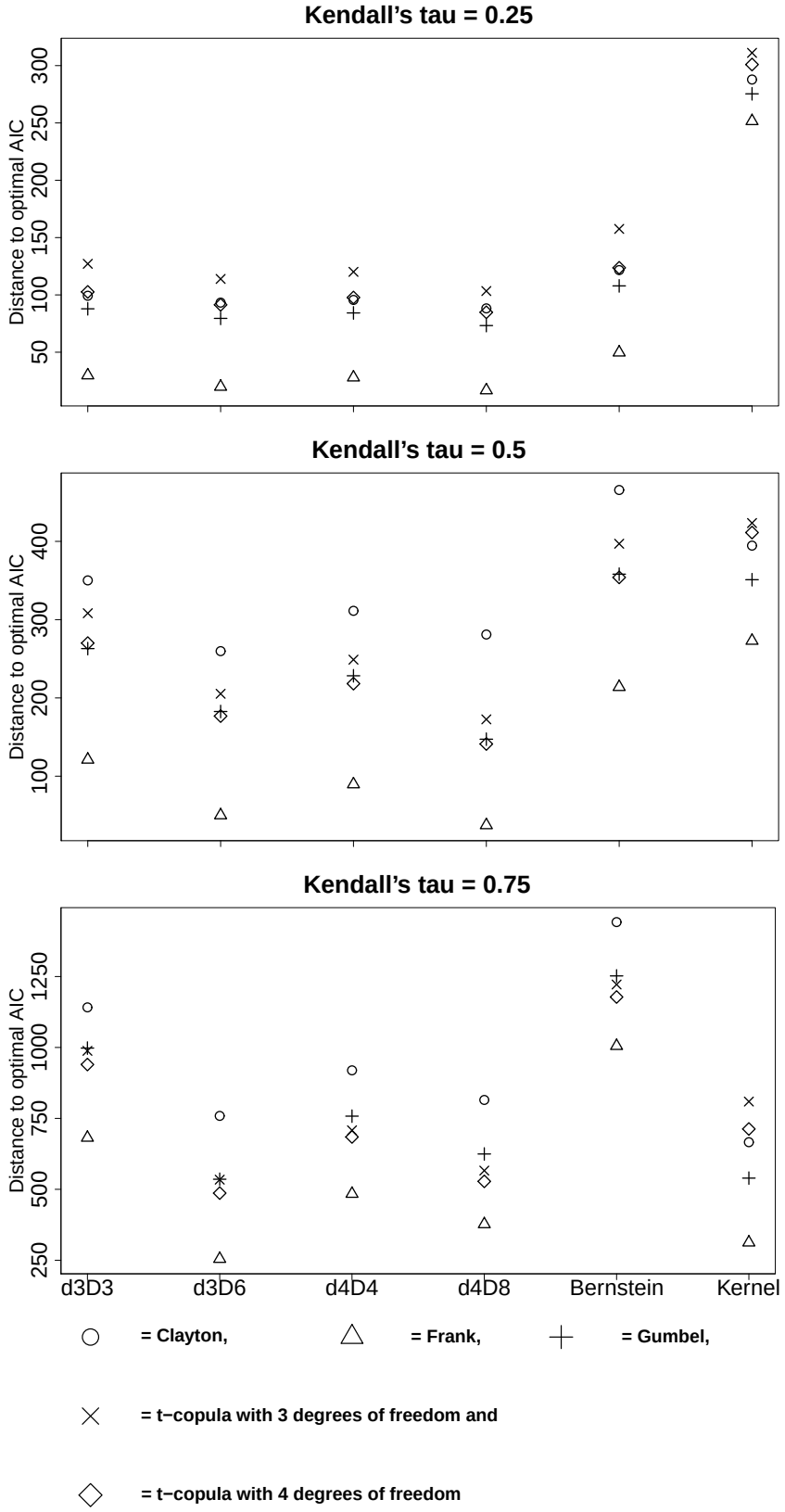


Figure 4: Simulated AIC difference $\widehat{AIC} - AIC_{true}$ for $p = 3$. From left to right: $\widehat{AIC}_{np} - AIC_{true}$ for $d = 3, D = 3$ and $d = 3, D = 6$ and $d = 4, D = 4$ and $d = 4, D = 8$, respectively, $\widehat{AIC}_{bernstein} - AIC_{true}$ and finally $\widehat{AIC}_{kernel} - AIC_{true}$

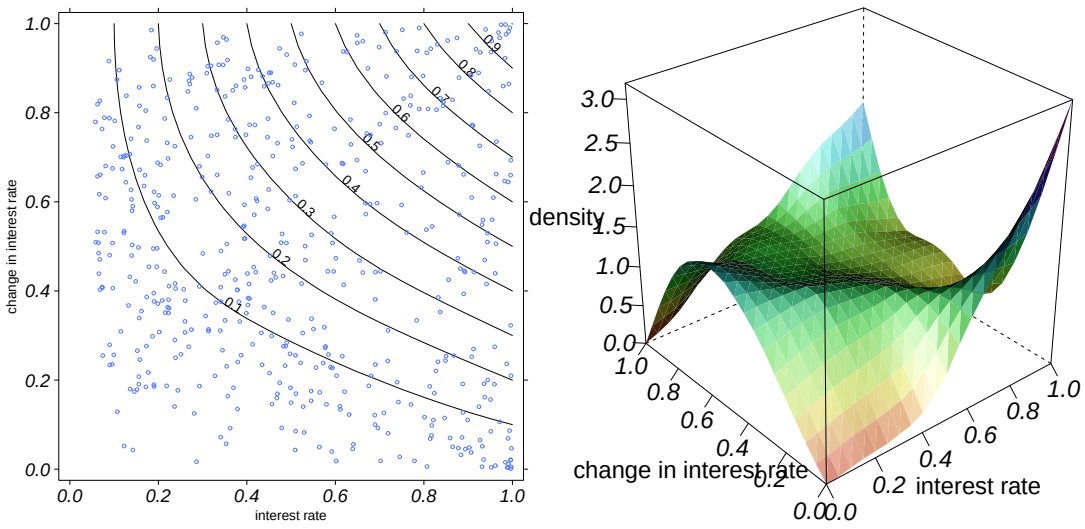


Figure 5: Copula (left) and copula density (right) for the interest rate data from the data set `Capm` in the `R` package `Ecdat` with $d = 5$ and $D = 5$.

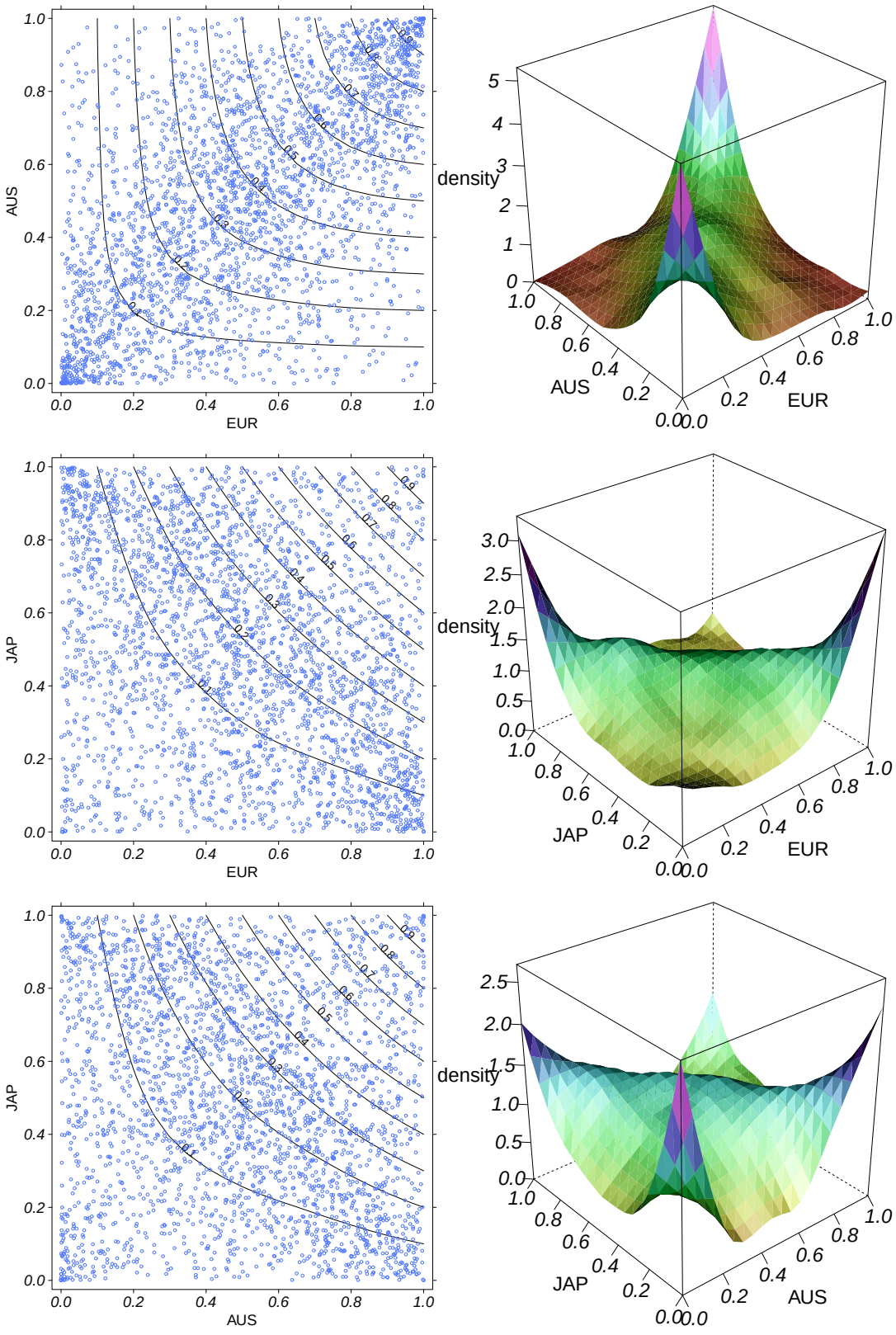


Figure 6: Bivariate marginal copula (left) and copula density (right) between Euro (EUR), Australian Dollar (AUS) and Japanese Yen (JAP) compared to the US-dollar from January 3rd, 2000 until May 6th, 2011 with $d = 4$ and $D = 8$.