GeD spline estimation of multivariate Archimedean copulas

Dimitrina S. Dimitrova [a], Vladimir K. Kaishev [b,*], Spiridon I. Penev [c]

[a] *Cass Business School, 106 Bunhill Row, London, EC1Y8TZ, UK*

[b] *Cass Business School, 106 Bunhill Row, London, EC1Y8TZ, UK*

[c] *The University of New South Wales, Sydney, 2052 NSW, Australia*

Abstract

A new multivariate Archimedean copula estimation method is proposed in a non-parametric setting. The method uses the so called Geometrically Designed splines (GeD splines) to represent the cdf of a random variable W_θ, obtained through the probability integral transform of an Archimedean copula with parameter θ. Sufficient conditions for the GeD spline estimator to posses the properties of the underlying theoretical cdf, $K(\theta, t)$, of W_θ, are given. The latter conditions allow for defining a three-step estimation procedure for solving the resulting non-linear regression problem with linear inequality constraints. In the proposed procedure, finding the number and location of the knots and the coefficients of the unconstrained GeD spline estimator and solving the constraint least-squares optimisation problem, are separated. Thus, the resulting spline estimator $\hat{K}(\hat{\theta}, t)$ is used to recover the generator and the related Archimedean copula by solving an ordinary differential equation. The proposed method is truly multivariate, it brings about numerical efficiency and as a result can be applied with large volumes of data and for dimensions $d \geq 2$, as illustrated by the numerical examples presented.

Key words: Archimedean copula, generator, Kendall's process, B-spline, geometrically designed regression splines, shape preserving
PACS: 02.50.Sk, 02.50.Tt, 02.60.Ed

[*] Corresponding author. Tel. +44 (0) 20 7040 8453; fax +44 (0) 20 7040 8572, e-mail: v.kaishev@city.ac.uk

1 Introduction

Recently, considerable attention has been paid to the problem of inference about copulas. The monographs by [3], [20] and [16] summarize to some extent the activities in this area. In broad terms, a copula function is a multivariate distribution function with uniform marginals. It is used as a linking block between the joint cumulative distribution function (cdf) $F(x_1, \ldots, x_d)$ of a vector of random variables $\mathbf{X} = (X_1, \ldots, X_d)$ and its marginal cdf's $F_1(X_1), \ldots, F_d(X_d)$. This probabilistic interpretation of copulas is justified by the famous *Sklar's theorem* which states that, under some mild conditions, there exists a unique copula function $C(u_1, \ldots, u_d)$ such that

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)) \tag{1}$$

holds. For the joint density $f(x_1, \ldots, x_d)$ of $\mathbf{X}$ one easily gets from (1) that

$$f(x_1, \ldots, x_d) = c(F_1(x_1), \ldots, F_d(x_d)) \prod_{j=1}^{d} f_j(x_j),$$

where $f_j(\cdot), j = 1, \ldots, d$ are the marginal densities and $c(\cdot, \ldots, \cdot)$ denotes the copula density

$$c(u_1, \ldots, u_d) = \frac{\partial^d C(u_1, \ldots, u_d)}{\partial u_1 \ldots \partial u_d}, u_i \in (0,1), i = 1, \ldots, d.$$

In general, estimation of the joint cdf $F(x_1, \ldots, x_d)$ in (1) involves estimation of both the copula C and the marginals $F_j(\cdot)$, $j = 1, \ldots, d$. Depending on the degree to which the copula and the marginals are assumed to be known, parametric or non-parametric estimation methods have been developed. In terms of parametric inference, the Maximum Likelihood could be adopted whenever feasible. An alternative, simpler approach, called Inference Function for Margins (IFM), (see [16]) involves a two-step procedure where one first estimates the parameters of the marginal distributions and then substitutes them to maximize the likelihood of the copula. On the other hand, if, in contrast to the copula, the marginals can not be specified parametrically, a semiparametric approach has been suggested. This approach goes back at least to 1995 (see [22] and [7]). In these papers, the marginals are estimated by the empirical distribution functions. After substituting them in the formula for the copula density, one tries to maximise the resulting rank-based log-likelihood. The approach has gained popularity in practice under the name of "pseudo-likelihood". It enjoys consistency and asymptotic normality although asymptotic efficiency is not granted in general. Quite recently, the paper [2] showed that plug-in *sieve* MLE works and produces asymptotically efficient estimators for the parametric part. They also show that prior restrictions on the marginal distributions can be incorporated in order to achieve efficiency gains when constraints hold. The non-parametric estimation has also been paid due attention. In [15] the copula density estimation has been dealt with whereas, [21] consider estimation of both the copula and its density by using Bernstein polynomials to smooth out the empirical copula.

A particular class of copulas, called Archimedean copulas, have recently gained considerable popularity as a dependence modelling tool. It involves a non-parametric component $\phi(\cdot)$, called generator, which is a function of one variable and completely describes the dependency structure of the entire d-dimensional vector $\mathbf{X}$. This brings about essential simplification with respect to the inference for Archimedean copulas. To see this, recall that the Archimedean copula is defined as (see e.g. [20], Theorem 4.6.2)

$$C(u_1, \dots, u_d) = \phi^{-1}(\phi(u_1) + \cdots + \phi(u_d)) \tag{2}$$

where the *generator function* $\phi(\cdot)$ is a continuous, strictly decreasing convex function on (0,1) such that $\phi(1) = 0$ and $\phi^{-1}(.)$ is completely monotonic, i.e.

$$\frac{(-1)^i d^i}{dx^i} \phi^{-1}(x) \geq 0, \ \ i = 1, ..., d. \tag{3}$$

If $\phi(0_+) = \infty$ the generator is strict, otherwise if $\phi(0_+) < \infty$ it is called non-strict. Major measures of association, such as Kendall's tau (τ) and Spearman's rho (ρ), do not depend on the marginals and can be directly expressed through the generator (see e.g. [20]). From (2), it can be seen that the generator is only determined up to a multiplicative positive constant. Thus, as seen from (2), in order to estimate an Archimedean copula one needs to be able to estimate the generator $\phi(\cdot)$, based on a sample of observations on $\mathbf{X}$. The solution of this estimation problem substantially depends on the parametrization of $\phi(\cdot)$. A summary of the existing most popular Archimedean copulas and their generators (including the Clayton, Ali-Mikhail-Haq, Gumber, Frank and many other families) can be found in [20] and [3]. These generators typically give a limited description of the dependence structure between the random variables $X_1, \dots, X_d$, since they are characterized via one (or two) dimensional parameter θ. Although estimation in this case is simpler, there is a scope for more richly parameterized generators which allow for better flexibility in modelling the copula C. An interesting approach is illustrated in the discussion [8] where the authors demonstrate several ways to generate bivariate Archimedean copula models via smooth transformations of existing generators. The recent paper by [23] is yet another attempt that tries to make the generator more flexible via local interpolation of existing "textbook" germs of generators. However, the latter approach still has some limitations in its flexibility since it would in general demonstrate a bias towards the chosen germs of generators.

In practice, one would like to be given a flexible family of generators without too much of a bias induced in their choice. A possible general approach to increasing the flexibility of the generator ϕ and hence of C, is to use spline functions in their representation. An attempt in this direction is the paper by [18] who uses a penalized smoothing spline to represent the function $\lambda(\cdot) = \frac{\phi(\cdot)}{\phi'(\cdot)}$ and then estimates its parameters via a MCMC algorithm in a Bayesian framework. The author has noticed that approximating $\lambda(\cdot)$ instead of $\phi(\cdot)$ is more convenient since $\lambda(\cdot)$ is uniquely determined, regardless of the multiplicative positive constant selected in the definition of the generator. However, the proposed smoothing splines involve high number (usually 20-30 according to [18]) equidistant knots and a penalty parameter, which leads to a high dimension of the estimation space and hence to increasing complexity of the subsequent MCMC Bayesian parameter estimation. Another drawback of this approach is that one needs to perform thousands of simulations from the posterior distribution of the parameters and average them in order to produce a resulting model. This can be prohibitively time consuming, especially for large data samples and parameter dimensions. An alternative approach could be to express the generator $\phi(\cdot)$ directly as a spline function of a fixed degree and to consider its coefficients and knots as unknown parameters θ. Then, the requirements for the spline $\phi_\theta(\cdot)$ to be a generator would translate in some shape-preserving constraints on its unknown parameters. In this case, it could be argued that the resulting copula density is a parametric density and hence, the constrained maximum-likelihood method could be applied to achieve *asymptotic efficiency* when estimating the generator. In order to illustrate the corresponding details of such a maximum-likelihood approach, let us consider the case of $d = 2$. Then, (2) becomes

$$C_\theta(u_1, u_2) = \phi_\theta^{-1}(\phi_\theta(u_1) + \phi_\theta(u_2))$$

and we have

$$c_\theta(u_1, u_2) = \frac{\partial^2 C_\theta(u_1, u_2)}{\partial u_1 \partial u_2} = -\frac{\phi_\theta^{''}(C_\theta(u_1, u_2))\phi_\theta^{'}(u_1)\phi_\theta^{'}(u_2)}{\phi_\theta^{'}(C_\theta(u_1, u_2))^3}, u_1, u_2 \in [0, 1],$$

where the derivatives of $\phi_\theta(u)$ are with respect to u. Given a sample of n i.i.d. copies of the dependent uniform $(0, 1)$ random variables (u_{1i}, u_{2i}) , $i = 1, ..., n$, the likelihood function is

$$L(\theta) = \prod_{i=1}^{n} c_\theta(u_{1i}, u_{2i}) \tag{4}$$

and it (or its *logarithm, the log-likelihood function*) has to be maximized with respect to the parameter vector θ, under some *shape-preserving* constraints on θ. Unfortunately, under this direct free-knot spline approximation approach, the actual maximization in (4) is very difficult to implement numerically and is computationally very expensive since this becomes a multi-extrema constrained, non-linear optimization problem in possibly high dimension involving inversion of the spline generator. Moreover, there is a singularity of the $\phi'_\theta(\cdot)\big/\phi_\theta(\cdot)$ values at 0 and at 1. These difficulties lead to a prohibitive computational burden when the sample size n and/or the number of knots increase. The latter case is typical if a smoothing penalty is introduced in order to avoid oversmoothing. In addition, it must be said that although the problem formally may seem as a parametric one, the dimension of the parameter containing the spline coefficients (and possibly the knots) typically increases with the sample size so that the spline fit is more a non-parametric than a parametric likelihood fit. The parameters involved in the spline-based likelihood function do not have any statistical meaning.

The difficulties in calculating the "parametric" spline-based "maximum likelihood" estimator of $\phi_\theta(\cdot)$ motivates alternative approaches to estimating Archimedean copulas. In this paper, we propose a minimum-distance type Archimedean copula estimation method which utilizes ideas from Computer Aided Geometric Design.

The structure of the paper is as follows. In Section 2, we introduce our new approach to the Archimedean copula estimation problem which is based on the application of the so called Geometrically Designed splines (see [13] and [14]). In Section 3, some bivariate and higher dimensional numerical examples are presented and discussed. Section 4 concludes the paper.

2 GeD spline estimation of Archimedean copulas

In this section, we formulate a new approach to the Archimedean copula estimation problem. It utilizes the so called Geometrically Designed Regression Splines (abbreviated as GeD Splines or GeDS) that have first been developed by Kaishev et al. (2006 a,b) [13] and [14] for the context of unconstrained, variable-knot spline regression estimation. We consider the following Archimedean copula estimation problem.

Let $(X_{11}, \ldots, X_{d1}), \ldots, (X_{1n}, \ldots, X_{dn})$ be $n \geq 2$ independent observations of the random vector $\mathbf{X} = (X_1, \dots, X_d)$, with a joint distribution function given as in (1), but assuming that its copula $C_\theta(u_1, \dots, u_d)$ is a parametrized Archimedean copula, defined as in (2) with a generator $\phi_\theta(.)$ and that $F_i(x_i)$, $i = 1, \dots, d$, are some continuous marginals. We will equivalently use the notation $(U_{11}, \dots, U_{d1}), \dots, (U_{1n}, \dots, U_{dn})$ to denote the probability integral transforms $U_i = F_i(X_i)$, $i = 1, \dots, d$, of the original observations. Then, we can consider the random variable

$$W_\theta = C_\theta(U_1, \dots, U_d) = \phi_\theta^{-1}(\phi_\theta(U_1) + \cdots + \phi_\theta(U_d))$$

with cdf $K(\theta, t) = P(W_\theta \le t)$.

A simple argument which appears in [10] shows that if we define

$$\begin{aligned} W_{j,n} &= \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(U_{1i} \le U_{1j}, \dots, U_{di} \le U_{dj}) \\ &= \frac{1}{n} \sum_{i=1}^{n} \prod_{l=1}^{d} \mathbf{1}(F_l^{-1}(U_{li}) \le F_l^{-1}(U_{lj})) \\ &= \frac{1}{n} \sum_{i=1}^{n} \prod_{l=1}^{d} \mathbf{1}(X_{li} \le X_{lj}), j = 1, \dots, n, \end{aligned}$$

where $\mathbf{1}(\cdot)$ is the indicator function of the event $(\cdot)$, then $W_{j,n}$ can be considered pseudo-observations of the random variable W_θ. In what follows, we will demonstrate that the proposed GeD spline estimator of $C_\theta(u_1, \dots, u_d)$, depends solely on the pseudo-observations $W_{j,n}$ $j = 1, \dots, n$. On the other hand, as shown above $W_{j,n}$, $j = 1, \dots, n$ are directly expressed in terms of the observations $(X_{1j}, \dots, X_{dj})$, $j = 1, \dots, n$ and only depend on their relative order, i.e. on their ranks. Therefore, knowledge of the marginals is not required for the development of the proposed GeD spline Archimedean copula estimation procedure.

The empirical version of the cdf, $K(\theta, t)$, of W_θ is then defined as

$$K_n(t) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}(W_{n,(j)} \le t),$$

where $W_{n,(j)}$, $j = 1, \dots, n$, are the ordered values of $W_{j,n}$, $j = 1, \dots, n$. Let us point out that due to the discretisation effect in the definition of $W_{j,n}, j = 1, \dots, n$, some of the resulting $W_{n,(j)}$ values will coincide. It should also be noted that the pseudo-observations are dependent and hence can not be viewed as a random sample from $K(\theta, t)$. Despite this, the function $K_n(t)$ indeed happens to be a consistent estimator of $K(\theta, t)$ (see [1]). The empirical process

$$\mathbf{K}_n(t) = \sqrt{n}\{K_n(t) - K(\theta, t)\},$$

called the *Kendall's Process*, has been explored in the two-dimensional case ($d = 2$), by [10]. In the general case ($d \geq 2$), under some mild conditions, $\mathbf{K}_n(t)$ has been shown by [1] to converge to a zero mean Gaussian process with a certain covariance function. These authors have also established the following useful representation

$$K(\theta, t) = t + \sum_{i=1}^{d-1} (-1)^i \frac{\{\phi_\theta(t)\}^i}{i!} \frac{d^i}{dx^i} \phi_\theta^{-1}(x)|_{x=\phi_\theta(t)}, \tag{5}$$

where it is assumed that

$$\frac{\{\phi_\theta(t)\}^i}{i!} \frac{d^i}{dx^i} \phi_\theta^{-1}(x)|_{x=\phi_\theta(t)} \to 0 \text{ as } t \to 0_+ \text{ for all } i = 1, ..., d-1. \tag{6}$$

The following properties of the cdf $K(\theta, t)$ follow from the properties of $\phi_\theta(.)$ (see also [20], Chapter 4):

1) $K(\theta, 0) = 0, K(\theta, 1) = 1$
2) $K(\theta, t) > t, t \in (0, 1)$
3) $K'(\theta, t) > 0, t \in (0, 1)$

Let us note that the inequality in 2), which holds for any $d \geq 2$, follows from (5), noting that, for $d \geq 2$, each term in the summation is positive by the requirement (3) and therefore, $K(\theta, t) - t > 0, t \in (0, 1)$. This important inequality seems not to have been recorded in the literature for the case $d > 2$. Obviously, we also have that $0 < t < K(\theta, t)_{d=2} \leq K(\theta, t)_{d=3} \leq \dots$ which suggests that $K(\theta, t)$ becomes more 'rectangular' as the dimension d increases.

2.1 Estimating the cdf $K(\theta, t)$

Our approach to estimating the Archimedean copula $C_\theta(u_1, \dots, u_d)$ is to approximate $K(\theta, t)$ with a spline function, $K_\alpha(t; t_{k,m})$ of order m (degree $m-1$), defined on the set of $2m + k$ knots

$$t_{k,m} = \{0, \dots, 0 < t_{m+1} < \dots < t_{m+k} < 1, \dots, 1\}$$

i.e., to assume that $K(\theta, t)$ admits the representation

$$K_\alpha(t; t_{k,m}) = \alpha' N_m(t) = \sum_{i=1}^{p} \alpha_i N_{i,m}(t), t \in [0, 1]$$

where $\alpha = (\alpha_1, \ldots, \alpha_p)'$ is a vector of unknown coefficients and $N_m(t) = (N_{1,m}(t), \ldots, N_{p,m}(t))'$ are $p = m + k$ B-splines of order m on the set of knots $t_{k,m}$. The extended vector of unknown parameters θ, includes the coefficients α, the number of internal knots, k, and their locations $t_{m+1}, \ldots, t_{m+k}$, i.e., $\theta = (\alpha_1, \ldots, \alpha_p, k, t_{m+1}, \ldots, t_{m+k})$. Then, based on the empirical cdf $K_n(t)$, we define the estimator of θ as a minimum distance estimator, i.e., one that minimizes the distance between $K_n(t)$ and $K_\alpha(t; t_{k,m})$, for a suitably chosen distance measure. In this paper, we chose to minimise the weighted L_2-distance between $K_n(.)$ and $K_\alpha(., t_{k,m})$:

$$\int_0^1 n\{K_n(t) - K_\alpha(t, t_{k,m})\}^2 dK_n(t), \tag{7}$$

under the constraints imposed on the spline estimator, $K_\alpha(t; t_{k,m})$, by the properties 1)-3) of $K(\theta, t)$. Thus, it is not difficult to see that the requirements 1)-3) with respect to $K_\alpha(t; t_{k,m})$ translate into constraints on the unknown parameters θ. This is made more precise by the following Lemma, which gives (sufficient only) conditions for the spline $K_\alpha(t; t_{k,m})$ to reproduce the properties 1)-3) possessed by the underlying cdf $K(\theta, t)$.

Lemma 2.1 The spline $K_\alpha(t; t_{k,m})$ satisfies Properties 1)-3) if the following constraints hold:

1) $\alpha_1 = 0 < \alpha_2 < \alpha_3 < \cdots < \alpha_p = 1$
2) $\alpha_i > \xi_i, \quad i = 2, \ldots, p - 1,$

where $\xi_i = \frac{t_{i+1} + \cdots + t_{i+m-1}}{m-1}, \quad i = 2, \ldots, p-1, \quad \alpha_1 = \xi_1 = 0, \quad \alpha_p = \xi_p = 1$ are the so called *Greville abscissae*, defined on the set of knots $t_{k,m}$.

Proof It is not difficult to see that the first part of Property 1), i.e. $K_\alpha(0; t_{k,m}) = \sum_{i=1}^{p} \alpha_i N_{i,m}(0) = 0$ holds since $\alpha_1 = 0$, the B-splines $N_{2,m}(t), \ldots, N_{m,m}(t)$ defined on $t_{k,m}$, vanish at $t = 0$ and the remaining B-splines $N_{m+1,m}(t) = \cdots = N_{p,m}(t) = 0$ by definition (see [4], Chapters 9,10). Similarly, $K_\alpha(1; t_{k,m}) = 1$ follows noting that $N_{m+k-1,m}(1) = \cdots = N_{k+1,m}(1) = 0$, $N_{p,m}(1) = 1$ and $\alpha_p = 1$. Property 2) follows noting that $\alpha_i > \xi_i, i = 2, \ldots, p - 1$ imply

$$K_\alpha(t; t_{k,m}) = \sum_{i=1}^{p} \alpha_i N_{i,m}(t) > \sum_{i=1}^{p} \xi_i N_{i,m}(t) = t,$$

where in the last equality we have used the identity

$$\sum_{i=1}^{p} \xi_i N_{i,m}(t) = t$$

referred to as the linear precision property of B-splines.

Finally, Property 3) holds, since up to a multiplicative positive constant,

$$\Delta\alpha_i = \begin{cases} (\alpha_i - \alpha_{i-1})/(t_{i+m-1} - t_i) \text{ if } t_i < t_{i+m-1} \\ \Delta\alpha_{i-1} \text{ if } t_i = t_{i+m-1} \end{cases},$$

$i = 2, \ldots, p$ are the coefficients of the derivative of the spline $K_\alpha(t; t_{k,m})$, which itself is a spline of order $m-1$ and $\Delta\alpha_i > 0$ means that $K'_\alpha(t; t_{k,m}) > 0, t > 0$ holds. □

The problem of estimating an Archimedean copula can now be formulated as consisting of two subproblems. The first one is to find a minimum L_2-distance spline estimator $\hat{K}(\hat{\theta}, t)$ of $K(\theta, t)$, following (7). The second one is, by using $\hat{K}(\hat{\theta}, t)$, to estimate the generator $\phi_\theta(.)$ and its related Archimedean copula $C_\theta(u_1, \ldots, u_d)$. The first of these two problems can now be specified as follows. Given the pseudo-observations $W_{j,n}$, $j = 1, \ldots, n$, find

$$\min_\theta \sum_{j=1}^{n} \{K_n(W_{n,(j)}) - \sum_{i=1}^{p} \alpha_i N_{i,m}(W_{n,(j)})\}^2 \tag{8}$$

subject to the constraints

$$0 = \alpha_1 < \alpha_2 < \cdots < \alpha_p = 1 \tag{9}$$

$$\alpha_i > \xi_i, i = 2, \ldots, p-1, \tag{10}$$

$$0 < t_{m+1} < \cdots < t_{m+k} < 1, \tag{11}$$

where ξ_i are the *Greville abscissae.* We note that (8) is just an equivalent way of writing (7).

The constraints (9) and (10) are a consequence of Lemma 2.1. The constraints (11) are obvious.

Let us note that, in general, (8) is a non-linear least-squares optimization problem and (9), (10) and (11) are linear inequality constraints on the parameter vector θ. It is known that even unconstrained free-knot least-squares splines are virtually impossible to find (see e.g. [4]). For a detailed account on the related difficulties we refer to [19]. The constraints (9), (10) and (11), make the minimization in (8) even more problematic.

2.2 The GeD spline Archimedean copula estimation procedure

In order to overcome the difficulties mentioned above, we propose the following three-step Archimedean copula estimation procedure. The first two steps deal with solving the constrained minimization problem (8). In the third step, the generator and its related Archimedean copula are estimated.

Step 1) Ignoring constraints (9) and (10), find a set of knots $t^*_{k,m}$ and spline coefficients α^*, such that $\hat{K}_{\alpha^*}\left(t; t^*_{k,m}\right)$ is a variable-knot least square GeD spline estimate of $K_n(t)$, $t \in [0,1]$, i.e $\hat{K}_{\alpha^*}\left(t; t^*_{k,m}\right)$ is a (sub)optimal solution to (8).

Step 2) If α^* does not satisfy the constraints (9) and (10), then for the fixed optimal knots $t^*_{k,m}$, from step 1, re-solve (8) with respect to α subject to (9) and (10) to obtain the constrained (sub)optimal solution, $\hat{K}_{\hat{\alpha}}\left(t; t^*_{k,m}\right)$ of (8). Otherwise, $\hat{K}_{\alpha^*}\left(t; t^*_{k,m}\right)$ coincides with $\hat{K}_{\hat{\alpha}}\left(t; t^*_{k,m}\right)$ and one proceeds with step 3.

Step 3) Substitute the estimated cdf, $\hat{K}_{\hat{\alpha}}\left(t; t^*_{k,m}\right)$, from step 2, for $K(\theta, t)$ in the expression (5), due to [1] and solve the ordinary differential equation (5) in order to express the estimator of the generator, $\hat{\phi}_{\hat{\theta}}(t)$, in terms of $\hat{K}_{\hat{\alpha}}\left(t; t^*_{k,m}\right)$. Then, using the definition (2) obtain an estimate of the Archimedean copula $\hat{C}_{\hat{\theta}}(u_1, \ldots, u_d)$.

In order to construct the GeD spline estimator $\hat{K}_{\alpha^*}\left(t; t^*_{k,m}\right)$ of step 1, the method developed by Kaishev et al. (2006 a, b) [13] and [14] for the unconstrained regression context can be used. An essential ingredient of this method is the very close relationship between a spline regression function and its so called *control polygon*, with vertices whose y-coordinates are the regression coefficients and the x-coordinates are the Greville abscissae. The method involves a two-stage procedure. In the first stage, a variable-knot, least-squares linear spline fit to the data set $\left\{K_n(W_{n,(j)}), W_{n,(j)}\right\}_{j=1}^{n}$ is constructed. This fit is viewed as the initial position of the control polygon of a smoother higher order ($m > 2$) spline curve. In the second stage, the optimal set of knots $t^*_{k,m}$ of this higher order ($m > 2$) smooth spline curve, $\hat{K}_{\alpha}\left(t; t^*_{k,m}\right)$ is found, so that it preserves the shape of the initial control polygon and then this curve is fitted to the data, $\left\{K_n(W_{n,(j)}), W_{n,(j)}\right\}_{j=1}^{n}$ to adjust its position (i.e., to find α^*) in the unconstrained LS sense. In this way, it is ensured that the m-th order smooth LS fit $\hat{K}_{\alpha^*}\left(t; t^*_{k,m}\right)$ follows the shape of the initial control polygon, and hence the shape of the data. This procedure simultaneously produces quadratic, cubic, or higher order splines and the LS fit with the minimum residual sum of squares is chosen as the final fit which recovers best the underlying unknown cdf $K(\theta, t)$. The two stages of this approach have been given a formal interpretation as certain optimization problems with respect to the variables $k, t_{k,m}, \alpha$ and m (see [13]). Hence, the approach produces a solution which does not necessarily coincide with the globally optimal unconstraint solution to (8), under the free-knot non-linear optimization approach. As illustrated by the numerical examples presented in Kaishev et al. (2006 a, b) [13] and [14], it produces LS spline fits which are characterized by a small number of non-coalescent knots and very low mean square errors. Thus, the unconstrained GeD spline regression fits are shown to be nearly optimal and to enjoy some very good large sample properties, such as asymptotic normality. The latter allow for the construction of asymptotic confidence intervals illustrated in [13].

Step 1 of the GeD spline Archimedean copula estimation procedure proposed above has similar remarkable numerical efficiency (even for very large sample size n) typical for the GeD spline regression method. As illustrated in Section 4, due to the intrinsic shape preserving properties of the unconstrained GeD spline fit, $\hat{K}_{\alpha^*}\Big(t; t^*_{k,m}\Big)$, from step 1), in most cases, especially for large data samples ($n \geq 500$), directly meets the constraints (9), (10) and (11) and step 2 can be omitted. In general, as the optimal number of knots k and their locations $t^*_{k,m}$ found in step 1 are assumed fixed, step 2 is a linear least squares problem with respect to the α-as, involving only the simple linear constraints (9) and (10), and as a result it is not numerically expensive. The implementation of step 3 is somewhat more involved, since it requires the solution with respect to $\phi_\theta(t)$ of the ordinary differential equations (5). However, as illustrated in Section 4, it is again extremely numerically efficient and takes a few seconds on a standard PC. The two-dimensional ($d = 2$) and multidimensional cases ($d > 2$) have been given separate treatment, which we provide in the next section.

2.3 Recovering the generator and its related copula

In the two-dimensional Archimedean copula case, $d = 2$, (5) simplifies to

$$K(\theta, t) = t - \frac{\phi_\theta(t)}{\phi'_\theta(t)},$$

and step 3 of the Archimedean copula GeD spline estimation procedure yields directly the following estimator of the generator

$$\hat{\phi}_{\hat{\theta}}(t) = exp\Big(\int_0^t \frac{1}{s - \hat{K}_{\hat{\alpha}}(s; t^*_{k,m})} ds\Big). \tag{12}$$

The Archimedean copula estimator, $\hat{C}_{\hat{\theta}}(u_1, u_2)$ is then easily obtained using the *Mathematica* built-in function FindRoot in order to invert the estimated generator $\hat{\phi}_{\hat{\theta}}(t)$, following definition (2).

In the general, multivariate case ($d > 2$) we consider first the three-dimensional Archimedean copula estimation, $d = 3$. In this case, following step 3 and applying the change of variables $\eta_\theta(t) = t - \frac{\phi_\theta(t)}{\phi'_\theta(t)}$, equation (5) can be rewritten as the first order differential equation

$$\hat{K}_{\hat{\alpha}}\Big(t; t^*_{k,m}\Big) = \eta_\theta(t) - \frac{1}{2}(t - \eta_\theta(t))\eta'_\theta(t), \tag{13}$$

with initial condition $\eta_\theta(0_+) = 0$. In the case $d = 4$, following step 3, equation (5) can be rewritten in terms of $\eta_\theta(t)$ as

$$\begin{gathered}\hat{K}_{\hat{\alpha}}\Big(t; t^*_{k,m}\Big) = \eta_\theta(t) - \frac{1}{2}(t - \eta_\theta(t))\eta'_\theta(t) + \\ \frac{1}{6}(t - \eta_\theta(t))^2\eta''_\theta(t) - \frac{1}{6}(t - \eta_\theta(t))\eta'_\theta(t)\Big(1 + \eta'_\theta(t)\Big),\end{gathered} \tag{14}$$

with initial conditions $\eta_\theta(0_+) = 0$, $\eta'_\theta(0_+) = 0$. It should be noted that condition (6) is essential in order to remove the singularity in the point $t = 0$. Despite the deceivingly simple form of equation (13), which is transformed by the substitution $t - \eta_\theta(t) = \zeta(t)$ into an Abel's equation of the second kind, it seems impossible to solve it analytically and obtain an explicit expression for $\hat{\phi}_{\hat{\theta}}(t)$. Equation (14) is even more difficult to solve analytically. However, numerical solutions of both (13) and (14) are easily obtained with the *Mathematica* system, applying the NDSolve built-in function. There are no principle difficulties to put through this approach, even for dimensions $d > 4$, but we abstain from doing this here. The corresponding solution $\hat{\eta}_{\hat{\theta}}(t)$ can be used to obtain the estimator for the generator

$$\hat{\phi}_{\hat{\theta}}(t) = exp\Big(\int_0^t \frac{1}{s - \hat{\eta}_{\hat{\theta}}(s)} ds\Big). \tag{15}$$

The Archimedean copula estimator can then be obtained as

$$\hat{C}_{\hat{\theta}}(u_1, \ldots, u_d) = \hat{\phi}^{-1}_{\hat{\theta}}(\hat{\phi}_{\hat{\theta}}(u_1) + \cdots + \hat{\phi}_{\hat{\theta}}(u_d)), \tag{16}$$

whereby, for the inversion of $\hat{\phi}_{\hat{\theta}}(\cdot)$ we use the *Mathematica* built-in function FindRoot, which is a reliable one-dimensional root-finder. The three-step Archimedean copula estimation procedure described above has been implemented in *Mathematica 6.0* and its numerical performance is illustrated in Section 4.

Let us note that our GeD spline Archimedean copula estimation procedure yields an estimator for the Kendall's tau for any set of $d \geq 2$ random variables. When $d = 2$, this is the classical definition of Kendall's tau. For $d > 2$, this definition is extended in [17] to the generalised Kendall's tau as

$$\tau_d = \frac{2^d \int_0^1 (1 - K(\theta, t))dt - 1}{2^{d-1} - 1}$$

(compare also [11] and note the typographical error in [1], p. 198). Since estimating Kendall's tau is an important problem in its own right, we give below an explicit expression for the resulting GeD estimator.

Proposition 2.2 The GeD-spline estimator $\hat{K}_{\hat{\alpha}}(t; t^*_{k,m})$ implies an estimator of Kendall's tau by

$$\hat{\tau}_d = \frac{2^d(1 - \frac{1}{m}\sum_{i=1}^p \hat{\alpha}_i(t^*_{i+m} - t^*_i)) - 1}{2^{d-1} - 1}, \quad d \geq 2. \tag{17}$$

Proof It suffices to notice that

$$\int_0^1 (1 - \hat{K}(\hat{\theta}, t))dt = 1 - \sum_{i=1}^p \hat{\alpha}_i \int_0^1 N_{i,m}(t)dt = 1 - \sum_{i=1}^p \hat{\alpha}_i \frac{t^*_{i+m} - t^*_i}{m}.$$

3 Numerical results

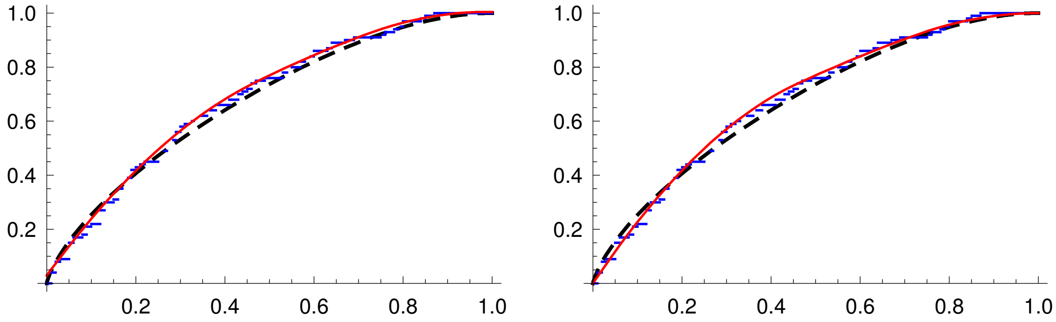
In this section, we illustrate the numerical performance of the Archimedean copula estimation procedure developed in Section 2 on several examples.

3.1 The two-dimensional case ($d = 2$)

We start with an example, considered by [18], in which $n = 100$ data points $\{K_n(W_{n,(j)}), W_{n,(j)}\}_{j=1}^{100}$ are simulated from two-dimensional Frank copula with Kendall's $\tau = 0.3$ (parameter of the generator $\theta = 2.92$ (see [20])).

Applying the proposed three-step GeD spline Archimedean copula estimation procedure, on step 1 we obtain the quadratic GeD spline estimate $\hat{K}_{\alpha^*}\left(t; t^*_{2,3}\right)$, with $\alpha^* = (0.0286, 0.5473, 0.8464, 1.0080, 1.0041)$ and knots $t^*_{2,3} = \{0, 0, 0, 0.4619, 0.7240, 1, 1, 1 \}$, presented in the left panel of Fig. 1. Step 2 of the

Fig. 1. *Simulated data points, $n = 100$, from a two-dimensional Frank copula, the true underlying cdf K (dashed line), the unconstrained GeDS estimate $\hat{K}_{\alpha^*}\left(t; t^*_{2,3}\right)$ (continuous line in the left panel) and the constrained GeDS estimate $\hat{K}_{\hat{\alpha}}\left(t; t^*_{2,3}\right)$ (continuous line in the right panel). The stepwise function represents $K_n(t)$.*

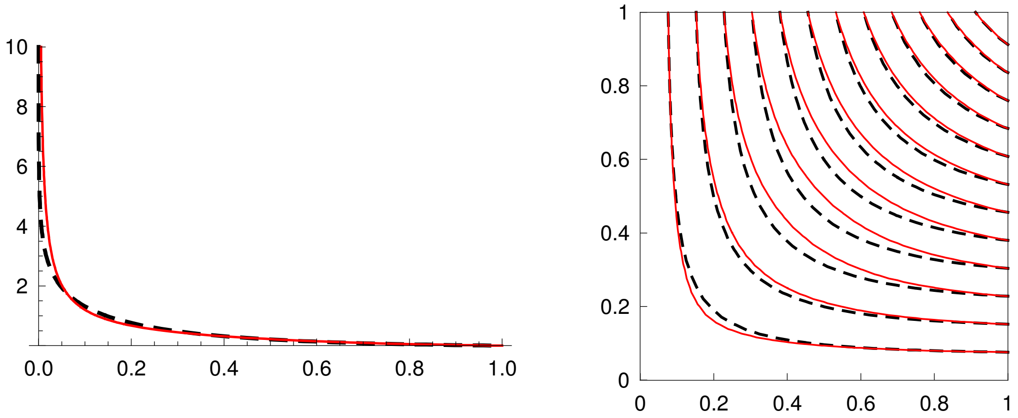


procedure yields $\hat{K}_{\hat{\alpha}}\left(t; t^*_{2,3}\right)$ with $\hat{\alpha}$ =(0, 0.5671, 0.8404, 0.9999,1), given in the right panel of Fig. 1. As can be seen on this example with $n = 100$ data points, α^* and $\hat{\alpha}$ for the unconstrained and the constrained spline approximations, are very close. This effect becomes more pronounced for larger data sets as will be illustrated in Fig. 6. Another advantage of the resulting quadratic GeDS approximation $\hat{K}_{\hat{\alpha}}\left(t; t^*_{2,3}\right)$, is the small number of internal knots, $k = 2$, and B-spline coefficients, $p = 5$, compared to 20 equidistant internal knots, 24 B-spline coefficients and a penalty parameter estimated by [18] via a complex procedure involving thousands time consuming iterations. The presented GeD spline approximation $\hat{K}_{\hat{\alpha}}\left(t; t^*_{2,3}\right)$ is obtained for 0.78 seconds on a standard PC (Pentium IV, 1.6Ghz, 512 RAM).

In the left panel of Fig. 2, we illustrate the true and estimated Frank copula generator, obtained on step 3 of the proposed Archimedean copula GeD spline estimation procedure, using (12). Contour plots of the true Frank copula and its estimated version, resulting from step 3, are presented in the right panel of Fig. 2. The contour plots of the estimated copula, $\hat{C}_{\hat{\theta}}(u_1, u_2)$, are obtained, applying the FindRoot *Mathematica* built-in function to invert the estimated generator, $\hat{\phi}_{\hat{\theta}}(t)$. As can be seen from Fig. 2, both the generator and the copula are recovered with a good accuracy with a very few parameters, $dim\theta = 5 + 1 + 2 = 8$. Note that the actual dimension of θ, reflecting the number of free parameters, is 6, since the first and the last B-spline coefficients are fixed to 0 and 1 respectively.

We have also compared the performance of our GeD method with other competing procedures. Direct comparison has been done with the non-parametric estimation method (NP) of Genest and Rivest [10] which has also been used by Lambert [18] as a benchmark for comparison with his method. As a result we are able to also compare indirectly with Lambert's Bayesian spline smoothing method. As in Lambert [18], data was simulated for both Frank and Clayton copulas with $d = 2, \tau = 0.3$ ($\theta = 2.92$ and $\theta = 0.86$, respectively ([20])). For

Fig. 2. *Left panel: The true Frank copula generator with Kendall's* $\tau = 0.3$ *(dashed line) and its GeD spline estimate (continuous line); Right panel: Contour plots of the true Frank copula (dashed line) and the estimated copula (continuous line).*



each of 100 simulation runs, we calculated the distance measure

$$\frac{1}{card(\Upsilon)} \sum_{v \in \Upsilon} |\varphi(v) - \tilde{\varphi}(v)| \tag{18}$$

where φ denotes the true generator and $\tilde{\varphi}$ corresponds to either the NP estimate or the GeD estimate. In both cases, Υ denotes the set of values at which $K_n(t)$ jumps. It can be seen from the boxplots presented in Fig. 3 that the GeD spline estimator outperforms the NP estimator. We have also indirectly compared our method with that of Lambert [18]. In Fig. 2 of [18] (see section 5 therein) the author has presented box plot comparisons of his method against the nonparametric method of Genest and Rivest, based on the distance (18) in terms of the function, λ, instead of φ. Let us note that the definition of this measure in section 5 and Fig. 2 of [18] is wrong and φ should be replaced by λ, as communicated to us by the author. Comparison of our Fig. 4 with the box plots presented in Fig. 2 of [18] indicates that the GeD spline estimator performs at least as well as the Bayesian spline smoothing estimator of Lambert in terms of accuracy. It should be noted, however, that in terms of computational efficiency and time, our procedure significantly outperforms that of [18]. The small differences in the results for the common benchmark NP estimator could be attributed to the different random number generators used by us in *Mathematica* and by Lambert in R. It is also interesting to observe that the average number of knots allocated automatically by our procedure in the 100 simulation runs, is equal to five (i.e. $p = 8$ B-spline coefficients) for both the Frank and Clayton copula. This also compares quite favourably to the 20 equidistant knots on (0,1) (i.e. 24 regression coefficients plus a smoothing parameter), used by [18].

Our second example aims at illustrating the performance of the procedure for reasonably small, $n = 50$, and large, $n = 500$, data samples. Data points

Fig. 3. *Simulated data: boxplots of* $\frac{1}{card(\Upsilon)}\sum_{v\in\Upsilon}|\varphi(v)-\tilde{\varphi}(v)|$ *for the Frank and for the Clayton copulas. The labels in the abscissa indicate which estimate* $\tilde{\varphi}(v)$ *was used: 1: GeD spline estimate. 2: NP estimate by Genest and Rivest.*

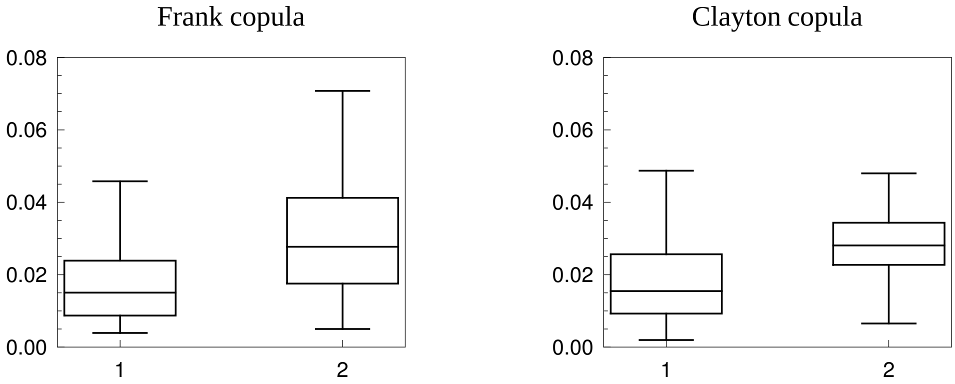
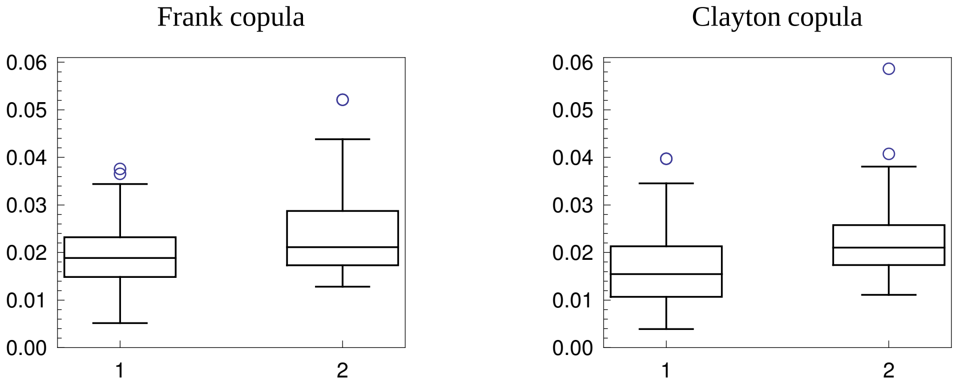


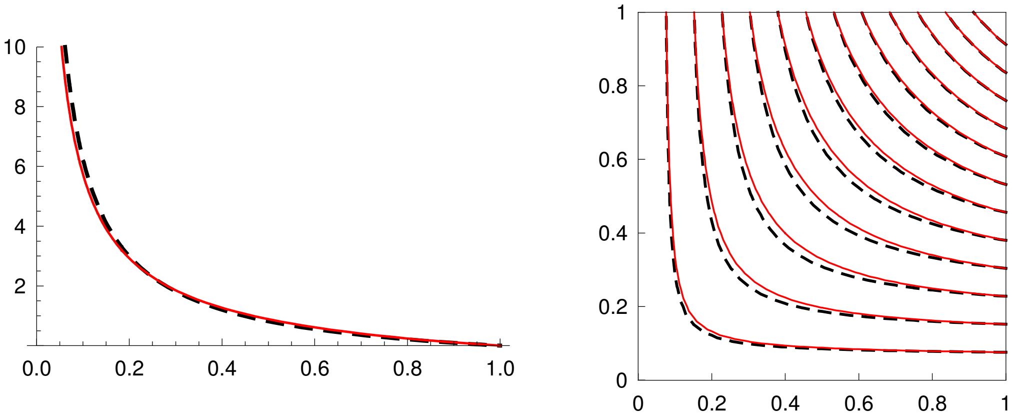
Fig. 4. *Simulated data: boxplots of* $\frac{1}{card(\Upsilon)}\sum_{v\in\Upsilon}|\lambda(v)-\tilde{\lambda}(v)|$ *for the Frank and for the Clayton copulas. The labels in the abscissa indicate which estimate* $\tilde{\lambda}(v)$ *was used: 1: GeD spline estimate. 2: NP estimate by Genest and Rivest.*



are simulated from a two-dimensional Clayton copula with Kendall's $\tau = 0.3$ (parameter of the generator $\theta = 0.86$). Based on $n = 50$ data points, in step 1 we obtain the quadratic GeD spline estimate $\hat{K}_{\alpha^*}\left(t; t^*_{1,3}\right)$ with one internal knot, $t^*_4 = 0.5194$ and α^* =(0.0324, 0.5503, 1.0235, 1.0051). The constrained GeD spline estimate $\hat{K}_{\hat{\alpha}}\left(t; t^*_{1,3}\right)$ obtained in step 2 has B-spline coefficients $\hat{\alpha} = (0, 0.5762, 0.9999, 1)$. As can be seen, even with $n = 50$ data points, α^* and $\hat{\alpha}$ for the unconstrained and the constrained spline approximations, are still reasonably close. The resulting GeD spline approximation $\hat{K}_{\hat{\alpha}}\left(t; t^*_{1,3}\right)$ is obtained for 0.31 seconds.

In the left panel of Fig. 5, the true and estimated Clayton copula generator in the case of $n = 50$ data points are presented. Contour plots of the true

Fig. 5. *Left panel: The true Clayton copula generator with Kendall's* $\tau = 0.3$ *(dashed line) and its GeD spline estimate (continuous line); Right panel: Contour plots of the true Clayton copula (dashed line) and the estimated copula (continuous line).*



Clayton copula and its estimated version, resulting from step 3, are given in the right panel of Fig. 5. Here again both the generator and the copula are recovered with a very good accuracy and only 4 free parameters (α_2, α_3, k, t_4).

For the case of $n = 500$, the estimated unconstrained cdf $\hat{K}_{\alpha^*}\left(t; t^*_{3,3}\right)$ has α^* =(0, 0.2744, 0.6306, 0.8257, 0.9999,1) and $t^*_{3,3} = \{$ 0, 0, 0, 0.2727, 0.4538, 0.6925, 1, 1, 1 $\}$. In this case, $\hat{K}_{\alpha^*}\left(t; t^*_{3,3}\right)$ is obtained for 2.50 seconds and it coincides with the constrained cdf $\hat{K}_{\hat{\alpha}}\left(t; t^*_{3,3}\right)$. The corresponding true Clayton copula generator is estimated by $\hat{\phi}_{\hat{\theta}}(t)$ with higher accuracy, compared to the case $n = 50$. Substituting $\hat{\alpha}$ and $t^*_{3,3}$ in (17) delivers an estimate of Kendall's tau $\hat{\tau}_2 = 0.31$ in this example.

3.2 The multivariate case ($d > 2$)

Our multivariate examples illustrate the performance of the method for dimension $d > 2$. In particular, we consider here the cases $d = 3$ and $d = 4$. In order to highlight the numerical efficiency of the proposed methodology we have used samples of size $n = 2000$ which were simulated from a three- and four-dimensional Clayton copula with Kendall's $\tau = 0.3$.

In the case $d = 3$, the quadratic constrained GeD spline estimate $\hat{K}_{\hat{\alpha}}\left(t; t^*_{3,3}\right)$ with three internal knots, $t^*_{3,3} = \{$ 0, 0, 0, 0.2084, 0.3564, 0.5827, 1, 1, 1$\}$ and B-spline coefficients $\hat{\alpha} = (0, 0.3155, 0.6663, 0.8888, 0.9999, 1)$, is presented in the left panel of Fig. 6. Using the spline estimate, $\hat{K}_{\hat{\alpha}}\left(t; t^*_{3,3}\right)$, the differential equation (13) is solved and its solution, $\hat{\eta}_{\hat{\theta}}(t)$, is used to estimate the underlying

Fig. 6. *Left panel: simulated data points,* $n = 2000$*, from a three-dimensional Clayton copula, the true underlying cdf* K *(dashed line), the constrained GeDS estimate* $\hat{K}_{\hat{\alpha}}\left(t; t^*_{3,3}\right)$ *(continuous line).The stepwise function represents* $K_n(t)$*. Right panel: The true Clayton copula generator (dashed line) and its GeD spline estimate (continuous line).*

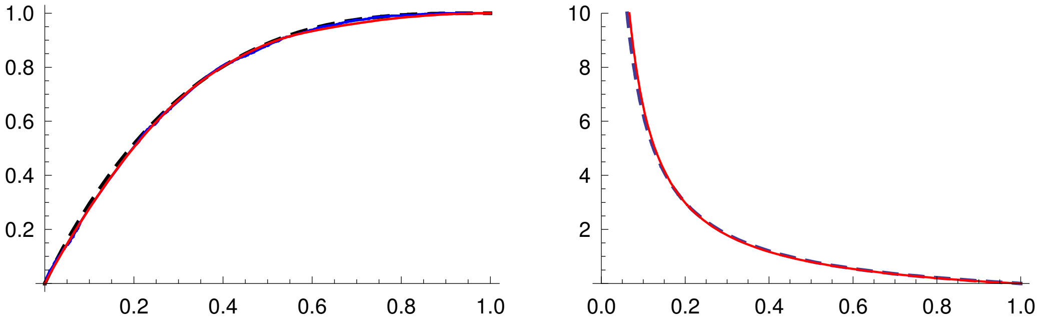
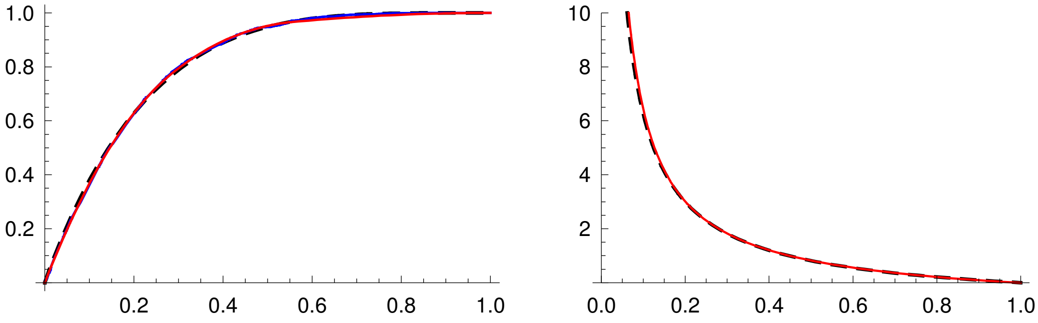


Fig. 7. *Left panel: simulated data points,* $n = 2000$*, from a four-dimensional Clayton copula, the true underlying cdf* K *(dashed line), the constrained GeDS estimate* $\hat{K}_{\hat{\alpha}}\left(t; t^*_{3,3}\right)$ *(continuous line).The stepwise function represents* $K_n(t)$*. Right panel: The true Clayton copula generator (dashed line) and its GeD spline estimate (continuous line).*



generator, following (15). The true and estimated Clayton copula generator are presented in the right panel of Fig. 6. As can be seen, the true underlying generator is almost perfectly recovered. It has to be noted that the GeD spline approximation $\hat{K}_{\hat{\alpha}}\left(t; t^*_{3,3}\right)$ is obtained for 10.61 seconds and solving (13) takes 0.05 seconds. The estimate of Kendall's tau using (17) is $\hat{\tau}_3 = 0.32$.

For the case of $d = 4$ and $n = 2000$, the graph of the estimated quadratic constrained cdf $\hat{K}_{\hat{\alpha}}\left(t; t^*_{4,3}\right)$, where $\hat{\alpha} = (0, 0.3351, 0.698, 0.8652, 0.9559, 0.9999, 1)$ and $t^*_{4,3} = \{0, 0, 0, 0.1598, 0.290, 0.4193, 0.5558, 1, 1, 1\}$, is plotted in the left panel of Fig. 7. The spline estimate $\hat{K}_{\hat{\alpha}}\left(t; t^*_{4,3}\right)$ is obtained for 13.20 seconds and the corresponding solution of the differential equation (14) is obtained in 0.06 seconds. In the right panel of Fig. 7, the corresponding true and estimated Clayton copula generators are presented. As can be seen, the estimate $\hat{\phi}_{\hat{\theta}}(t)$, can not be visually distinguished from the true underlying Clayton copula generator. The estimate of Kendall's tau using (17) is $\hat{\tau}_4 = 0.29$.

Analysing the presented examples, it has to be noted that with the increase of the dimension, d, the shape of the cdf $K(\theta, t)$ becomes more rectangular and therefore, more knots are needed in order to approximate it using splines. The constraints (9) and (10) come into play when the number of data points is relatively small, as seen from Fig. 1. In addition, in the two-dimensional case, $d = 2$, for small data samples, $n \leq 200$, one may attempt to solve the optimization problem (8) subject to the constraints (9), (10) and (11) directly using a non-linear optimization procedure in order to find the globally optimal spline estimate of $K(\theta, t)$. For example, using the NMinimize built-in *Mathematica* function, in the case of $n = 50$ data points simulated from a two-dimensional Clayton copula, the globally optimal constrained solution to (8) is found in 24.54 seconds and $\hat{\alpha}^{opt} = (0, 0.4008, 0.9999, 1)$, $t_4^{opt} = 0.3479, RSS^{*}opt = 0.1511$. This optimal solution does not differ significantly from the the quadratic GeD spline estimate $\hat{\alpha} = (0, 0.5762, 0.9999, 1)$, $t_4^{*} = 0.5194, RSS = 0.1552$, obtained in 0.31 seconds using the proposed three-step GeD spline Archimedean copula estimation method. The two spline estimates, $\hat{K}_{\hat{\alpha}^{opt}}\Big(t; t_{1,3^{opt}}\Big)$ and $\hat{K}_{\hat{\alpha}}\Big(t; t_{1,3}^{*}\Big)$, and their coresponding generators are hard to distinguish from one another. However, the globally optimal solution to (8) is obtained at a much higher computational cost.

In the case of $n = 100$ data points simulated from a two-dimensional Frank copula, the globally optimal constrained solution to (8) is found in 65.34 seconds, compared to 0.78 seconds using the proposed GeD spline Archimedean copula estimation method and the two estimates are again very close. The direct approach of solving (8) becomes infeasible for $d > 2$, i.e. when higher number of internal knots is required, and for large data samples, e.g. $n \geq 500$.

4 Comments and conclusions

The proposed method of estimating multivariate Archimedean copulas has been demonstrated to efficiently recover the underlying generator even for dimensions $d > 2$. To the the best of our knowledge this multivariate feature of the proposed procedure is unique and opens the scope for truly multivariate applications of Archimedean copulas in a variety of practical areas. Its extremely good numerical efficiency makes the method applicable for estimating dependence based on large volumes of data combined with high dimensions.

In principle, it should be possible to construct a new goodness-of-fit test of the null hypothesis $H_0 : C(u_1, \ldots, u_d)$ belongs to a particular family of Archimedean copulas (e.g. Frank, Clayton), based on Kolmogorov-Smirnov or Cramér-von Mises type statistics, using the GeD spline estimator $\hat{K}(\hat{\theta}, t)$ or $\hat{C}_{\hat{\theta}}(u_1, \ldots, u_d)$. These new goodness-of-fit tests can be viewed as alternatives to the existing methods based on K_n, such as that of [12]. Further details of how this can be done are outside the scope of this paper and are subject of an ongoing investigation.

Acknowledgements

The authors gratefully acknowledge the support received from the School of Mathematics and Statistics, The University of New South Wales, Sydney and the Faculty of Actuarial Science and Insurance, Cass Business School, City University, London which allowed them to exchange visits and work on this project.

The authors would also like to thank the two referees for the very constructive comments and suggestions which helped to improve the quality of the paper.

References

[1] Barbe, P., Genest, C., Ghoudi, K. and Rémillard, B. (1996). On Kendall's Process. *Journal of Multivariate Analysis*, **58**, 197-229.

[2] Chen, X., Fan, Y. and Tsyrennikov, V. (2006). Efficient estimation of semiparametric multivariate copula models. *JASA*, **101**, Issue 475, 1228-1240.

[3] Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula Methods in Finance*. Wiley & Sons, Chichester.

[4] De Boor, C. (2001). *A Practical guide to Splines.* Revised Edition, New York: Springer.

[5] Dechevsky, L. and Penev, S. (1998). On Shape-preserving wavelet estimators of cumulative distribution functions and densities. *Stochastic Analysis and Applications*, **16**, 3, 423-462.

[6] Fermanian, J.-D. (2005). Goodness-of-fit tests for copulas. *Journal of Multivariate Analysis*, **95**, 119-152.

[7] Genest, C., Ghoudi, K. and Rivest, L.-P. (1995). A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions. *Biometrika*, **82**, 3, 543-552.

[8] Genest, C., Ghoudi, K. and Rivest, L.-P. (1998). Discussion to "Understanding Relationships Using Copulas" by E. Frees and E. Valdez. *North American Actuarial Journal*, **2**, 1, 143-149.

[9] Genest, C. and Rémillard, B. (2007). Validity of the parametric bootstrap for goodnes-of-fit teting in semiparametric models. *Ann. Inst. Henri Poincaré*, **37**, in press.

[10] Genest, C. and Rivest, L.-P. (1993). Statistical Inference for bivariate Archinedean copulas. *JASA*, **88**, Issue 423, 1034-1043.

[11] Genest, C. and Rivest, L.-P. (2001). On the multivariate probability integral transformation. *Statistics & Probability Letters*, **53**, 391-399.

[12] Genest, C., Quessy, J-F., and Rémillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics*, **33**, 337-366.

[13] Kaishev, V., Dimitrova, D., Haberman, S. and Verrall, R. (2006). Geometrically desgined, variable knot regression splines: Asymptotics and inference. Statistical Research Paper No28, Cass Business School, London.

[14] Kaishev, V., Dimitrova, D., Haberman, S. and Verrall, R. (2006). Geometrically designed, variable knot regression splines: Variation diminishing optimality of knots. Statistical Research Paper No29, Cass Business School, London.

[15] Hall, P. and Neumeyer, N. (2006). Estimating a bivariate density when there are extra data on one or both components, *Biometrika*, **93**, 2, 439-450.

[16] Joe, H. (1977). *Multivariate Models and Dependence Concepts.* Chapman & Hall, London.

[17] Joe, H. (1990.) Multivariate concordance. *Journal of Multivariate Analysis*, **35**, 12-30.

[18] Lambert, P. (2007). Archimedean copula estimation using Bayesain splines smoothing techniques. *Computational Statistics & Data Analysis*, **51**, 12, 6307-6320.

[19] Lindstrom, M.J. (1999). Penalized estimation of free-knot splines. *J. Computational and Graphical Statistics*, 8, 2, 333-352.

[20] Nelsen, R.B. (1999). *An introduction to copulas.* Springer, New York.

[21] Sancetta, A and Satchell, S.(2004). The Bernstein copula and its appplications to modeling and approximations of multivariate distributions. *Econometric Theory*, **20**, 535-562.

[22] Shih, J.H. and Louis, T.A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, **51**, 4, 1384-1399.

[23] Vandenhende, F. and Lambert, P. (2005). Local dependence estimation using semiparametric Archimedean copulas. *The Canadian Journal of Statistics*, **33**, 3, 377-388.