

Dimitrova, D. S., Kaishev, V. & Penev, S. (2007). GeD spline estimation of multivariate Archimedean copulas (Report No. Actuarial Research Paper No. 179). London, UK: Faculty of Actuarial Science & Insurance, City University London.



**CITY UNIVERSITY  
LONDON**

[City Research Online](#)

**Original citation:** Dimitrova, D. S., Kaishev, V. & Penev, S. (2007). GeD spline estimation of multivariate Archimedean copulas (Report No. Actuarial Research Paper No. 179). London, UK: Faculty of Actuarial Science & Insurance, City University London.

**Permanent City Research Online URL:** <http://openaccess.city.ac.uk/2311/>

### **Copyright & reuse**

City University London has developed City Research Online so that its users may access the research outputs of City University London's staff. Copyright © and Moral Rights for this paper are retained by the individual author(s) and/ or other copyright holders. All material in City Research Online is checked for eligibility for copyright before being made available in the live archive. URLs from City Research Online may be freely distributed and linked to from other web pages.

### **Versions of research**

The version in City Research Online may differ from the final published version. Users are advised to check the Permanent City Research Online URL above for the status of the paper.

### **Enquiries**

If you have any enquiries about any aspect of City Research Online, or if you wish to make contact with the author(s) of this paper, please email the team at [publications@city.ac.uk](mailto:publications@city.ac.uk).



Cass Business School  
City of London

Cass means business

## Faculty of Actuarial Science and Insurance

# GeD Spline Estimation of Multivariate Archimedean Copulas.

Dimitrina Dimitrova, Vladimir  
Kaishev & Spiridon Penev.

Actuarial Research Paper  
No. 179

April 2007

ISBN 978-1-905752-08-9

Cass Business School  
106 Bunhill Row  
London EC1Y 8TZ  
T +44 (0)20 7040 8470  
[www.cass.city.ac.uk](http://www.cass.city.ac.uk)

“Any opinions expressed in this paper are my/our own and not necessarily those of my/our employer or anyone else I/we have discussed them with. You must not copy this paper or quote it without my/our permission”.

# GeD spline estimation of multivariate Archimedean copulas

D. Dimitrova , V. Kaishev, \*      S. Penev †

12th April 2007

---

\*Cass Business School, London

†The University of New South Wales, Sydney

## 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), recently introduced by Kaishev et al. (2006 a,b) [10] and [11], to represent the cdf of a random variable  $W_\theta$ , obtained through the probability integral transform of an Archimedean copula with parameter  $\theta$ . Sufficient conditions for the GeD spline estimator to possess the properties of the underlying theoretical cdf,  $K(\theta, t)$ , of  $W_\theta$ , 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.

Keywords: *Archimedean copula, generator, Kendall's process, B-spline, geometrically designed regression splines, shape preserving.*

## 1 Introduction

Recently, considerable attention has been paid to the problem of inference about copulas. The monographs by [3], [16] and [13] 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)  $F(x_1, \dots, x_d)$  of a vector of random variables  $\mathbf{X} = (X_1, \dots, X_d)$  and its marginal cdf's  $F_1(X_1), \dots, F_d(X_d)$ . 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  $C(u_1, \dots, u_d)$  such that

$$(1) \quad F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$$

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

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

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

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

In general, estimation of the joint cdf  $F(x_1, \dots, x_d)$  in (1) involves estimation of both the copula  $C$  and the marginals  $F_j(\cdot), j = 1, \dots, d$ . 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 [13]) 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 and implemented quite recently by [2]. The authors show 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 of the marginals has also been paid due attention. In [12] the copula density estimation has been dealt with whereas, [17] 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. [16], Theorem 4.6.2)

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

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

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

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 ( $\tau$ ) and Spearman's rho ( $\rho$ ), do not depend on the marginals and can be directly expressed through the generator (see e.g. [16]). 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 [16] 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  $\theta$ . 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  $C$ . The recent paper by [18] is, to our knowledge, the only one that tries to make the generator more flexible via local interpolation of existing "textbook" germs of generators. However, this 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 approach to increasing the flexibility of the generator  $\phi$  and hence of  $C$ , is to use spline functions in their representation. An attempt in this direction is the paper by [14] 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 [14]) 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  $\theta$ . 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  $(u_{1i}, u_{2i})$ ,  $i = 1, \dots, n$ , the likelihood function is

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

and it (or its *logarithm, the log-likelihood function*) has to be maximized with respect to the parameter vector  $\theta$ , under some *shape-preserving* constraints on  $\theta$ . 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  $\phi'_\theta(\cdot)/\phi_\theta(\cdot)$  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 [10] and [11]). Section 3 considers briefly a goodness-of-fit testing procedure which could be applied to the proposed spline estimator. In Section 4 some bivariate and higher dimensional numerical examples are presented and discussed.

## 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) [10] and [11] for the context of unconstrained, variable-knot spline regression estimation. We consider the following Archimedean copula estimation problem.

Let  $(X_{11}, \dots, X_{d1}), \dots, (X_{1n}, \dots, X_{dn})$  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  $\phi_\theta(\cdot)$  and that,  $F_i(x_i)$ ,  $i = 1, \dots, d$ , are some known marginals. Under the lat-

ter assumption of known 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

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

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

A simple argument which appears in [7] shows that if we define

$$W_j = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(U_{1i} \leq U_{1j}, \dots, U_{di} \leq U_{dj}), j = 1, \dots, n,$$

where  $\mathbf{1}(\cdot)$  is the indicator function of the event  $(\cdot)$ , then  $W_j$  can be considered pseudo observations of the random variable  $W_\theta$ . The empirical version of its cdf,  $K_n(t)$ , is then defined as

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

where  $W_{(j)}$ ,  $j = 1, \dots, n$ , are the ordered values of  $W_j$ ,  $j = 1, \dots, n$ . In fact,  $W_j$ ,  $j = 1, \dots, n$ , can be viewed as a sample of observations from  $K(\theta, t)$ . 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 [7]. 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

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

where it is assumed that

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

The following properties of the cdf  $K(\theta, t)$  follow from the properties of  $\phi_\theta(\cdot)$  (see also [16], 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  $\alpha = (\alpha_1, \dots, \alpha_p)'$  is a vector of unknown coefficients and  $N_m(t) = (N_{1,m}(t), \dots, N_{p,m}(t))'$  are  $p = m + k$  B-splines of order  $m$  on the set of knots  $t_{k,m}$ . The extended vector of unknown parameters  $\theta$ , includes the coefficients  $\alpha$ , the number of internal knots,  $k$ , and their locations  $t_{m+1}, \dots, t_{m+k}$ , i.e.,  $\theta = (\alpha_1, \dots, \alpha_p, k, t_{m+1}, \dots, t_{m+k})$ . Then, based on the empirical cdf  $K_n(t)$  we define the estimator of  $\theta$  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. Note that the properties 1)-3) of  $K(\theta, t)$  must also be met by its spline estimator  $K_\alpha(t; t_{k,m})$ . 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  $\theta$ . 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:*

$$1) \alpha_1 = 0 < \alpha_2 < \alpha_3 < \dots < \alpha_p = 1.$$

$$2) \alpha_i > \xi_i, i = 1, \dots, p.$$

where  $\xi_i = \frac{t_{i+1} + \dots + t_{i+m-1}}{m-1}$ ,  $i = 1, \dots, p$  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  $N_{2,m}(t), \dots, N_{m,m}(t)$  defined on  $t_{k,m}$ , vanish at  $t = 0$  and the remaining B-splines  $N_{m+1,m}(t) = \dots = N_{p,m}(t) = 0$  by definition (see [4], Chapers 9,10). Similarly,  $K_\alpha(1; t_{k,m}) = 1$  follows noting that  $N_{m+k-1,m}(1) = \dots = N_{k+1,m}(1) = 0$ ,  $N_{p,m}(1) = 1$  and  $\alpha_p = 1$ . Property 2) follows noting that  $\alpha_i > \xi_i, i = 1, \dots, p$  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,

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

$i = 2, \dots, p$  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 distance spline estimator  $\hat{K}(\hat{\theta}, t)$  of  $K(\theta, t)$ . The second one is, by using  $\hat{K}(\hat{\theta}, t)$ , to estimate the generator  $\phi_\theta(\cdot)$  and its related Archimedean copula  $C_\theta(u_1, \dots, u_d)$ . The first of these two problems can now be specified as follows. Given the sample of observations,  $W_j, j = 1, \dots, n$ , find

$$(7) \quad \min_{\theta} \sum_{j=1}^n \left\{ \frac{j}{n} - \sum_{i=1}^p \alpha_i N_{i,m}(W_{(j)}) \right\}^2$$

subject to the constraints

$$(8) \quad 0 = \alpha_1 < \alpha_2 < \dots < \alpha_p = 1$$

$$(9) \quad \alpha_i > \xi_i, i = 1, \dots, p$$

$$(10) \quad 0 < t_{m+1} < \dots < t_{m+k} < 1,$$

where  $\xi_i$  are the *Greville abscissae*.

The constraints (8) and (9) are a consequence of Lemma 2.1. The constraints (10) are obvious.

Let us note that, in general, (7) is a non-linear least-squares optimization problem and (8), (9) and (10) are linear inequality constraints on the parameter vector  $\theta$ . 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 [15]. The constraints (8), (9) and (10), make the minimization in (7) 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 (7). In the third step, the generator and its related Archimedean copula are estimated.

*Step 1)* Ignoring constraints (8) and (9), find a set of knots  $t_{k,m}^*$  and spline coefficients  $\alpha^*$ , such that  $\hat{K}_{\alpha^*}(t; t_{k,m}^*)$  is a variable-knot least square

GeD spline estimate of  $K_n(t)$ ,  $t \in [0, 1]$ , i.e.  $\hat{K}_{\alpha^*}(t; t_{k,m}^*)$  is a (sub)optimal solution to (7).

*Step 2)* If  $\alpha^*$  does not satisfy the constraints (8) and (9), then for the fixed optimal knots  $t_{k,m}^*$ , from step 1, re-solve (7) with respect to  $\alpha$  subject to (8) and (9) to obtain the constraint (sub)optimal solution,  $\hat{K}_{\hat{\alpha}}(t; t_{k,m}^*)$  of (7). Otherwise,  $\hat{K}_{\alpha^*}(t; t_{k,m}^*)$  coincides with  $\hat{K}_{\hat{\alpha}}(t; t_{k,m}^*)$  and one proceeds with step 3.

*Step 3)* Substitute the estimated cdf,  $\hat{K}_{\hat{\alpha}}(t; t_{k,m}^*)$ , 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  $\hat{K}_{\hat{\alpha}}(t; t_{k,m}^*)$ . Then, using the definition (2) obtain an estimate of the Archimedean copula  $\hat{C}_{\hat{\theta}}(u_1, \dots, u_d)$ .

In order to construct the GeD spline estimator  $\hat{K}_{\alpha^*}(t; t_{k,m}^*)$  of step 1, the method developed by Kaishev et al. (2006 a, b) [10] and [11] 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  $\{\frac{j}{n}, W_{(j)}\}_{j=1}^n$  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,  $\hat{K}_{\alpha}(t; t_{k,m}^*)$  is found, so that it preserves the shape of the initial control polygon and then this curve is fitted to the data,  $\{\frac{j}{n}, W_{(j)}\}_{j=1}^n$  to adjust its position (i.e., to find  $\alpha^*$ ) in the unconstrained LS sense. In this way, it is ensured that the  $m$ -th order smooth LS fit  $\hat{K}_{\alpha^*}(t; t_{k,m}^*)$  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 [10]). Hence, the approach produces a solution which does not necessarily coincide with the globally optimal unconstrained solution to (7), under the free-knot non-linear optimization approach. As illustrated by the numerical examples presented in Kaishev et al. (2006 a,

b) [10] and [11], 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 (see the examples in Kaishev et al., (2006 a, b) [10], [11] 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 [10].

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,  $\hat{K}_{\alpha^*}(t; t_{k,m}^*)$ , from step 1), in most cases, especially for large data samples ( $n \geq 500$ ), directly meets the constraints (8), (9) and (10) 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  $\alpha$ -as, involving only the simple linear constraints (8) and (9), 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 5, 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

$$(11) \quad \phi_\theta(t) = \exp\left(\int_0^t \frac{1}{s - \hat{K}_{\hat{\alpha}}(s; t_{k,m}^*)} ds\right).$$

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

$$(12) \quad \hat{K}_{\hat{\alpha}}(t; t_{k,m}^*) = \eta_\theta(t) - \frac{1}{2}(t - \eta_\theta(t))\eta'_\theta(t),$$

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

$$(13) \quad \hat{K}_{\hat{\alpha}}(t; t_{k,m}^*) = \eta_\theta(t) - \frac{1}{2}(t - \eta_\theta(t))\eta'_\theta(t) + \frac{1}{6}(t - \eta_\theta(t))^2\eta''_\theta(t) - \frac{1}{6}(t - \eta_\theta(t))\eta'_\theta(t)(1 + \eta'_\theta(t)),$$

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 (12), 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 (13) is even more difficult to solve analytically. However, numerical solutions of both (12) and (13) 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

$$(14) \quad \hat{\phi}_{\hat{\theta}}(t) = \exp\left(\int_0^t \frac{1}{s - \hat{\eta}_{\hat{\theta}}(s)} ds\right).$$

The Archimedean copula estimator can then be obtained as

$$(15) \quad \hat{C}_{\hat{\theta}}(u_1, \dots, u_d) = \hat{\phi}_{\hat{\theta}}^{-1}(\hat{\phi}_{\hat{\theta}}(u_1) + \dots + \hat{\phi}_{\hat{\theta}}(u_d)),$$

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* and its numerical performance is illustrated in Section 4.



Let us note that our GeD spline Archimedean copula estimation procedure yields the estimator for the Kendall's tau

$$\hat{\tau} = \frac{2^d \tilde{K}(\tilde{\theta}, t) - 1}{2^{d-1} - 1}$$

(compare [8] and note the typographical error in [1], p. 198) where  $\tilde{K}(\tilde{\theta}, t)$  is a spline of degree  $m + 1$  obtained by integrating the GeDS Archimedean copula estimator  $\hat{K}(\hat{\theta}, t)$  over the interval  $(0, 1)$ .

The next section is devoted to some goodness of fit considerations.

### 3 Goodness of fit testing

In principle, it should be possible to perform a goodness-of-fit test of the null hypothesis  $H_0 : C(u_1, u_2, \dots, u_d)$  is an Archimedean copula versus the alternative that it is not an Archimedean copula. General theoretical treatment of this problem is very difficult and, to the best of our knowledge, is still considered as one of the important open problems in the area of inference about copulas. However, using the approach we have outlined, it should be possible to use the bootstrap approach to test the important hypothesis  $H_0$ . The idea is to compare the distance between the  $\hat{C}_{\hat{\theta}}(u_1, \dots, u_d)$  estimator from (15) and the empirical copula estimator (the Deheuvel's estimator)

$$\hat{C}_n(u_1, \dots, u_d) = \frac{1}{n} \sum_{i=1}^n \prod_{k=1}^d \mathbf{1}(U_{ik} \leq u_k)$$

using some sup-norm or a weighted norm. For example, the following statistic considered in [9] could be used:

$$T_n = \sqrt{n} \max_{t_l=0,1,\dots,n-1, l=1,2,\dots,d} \max_{i=0,1} \left| \hat{C}_n\left(\frac{t_1}{n}, \dots, \frac{j}{n}, \dots, \frac{t_d}{n}\right) - \hat{C}_{\hat{\theta}}\left(\frac{t_1}{n}, \dots, \frac{j+i}{n}, \dots, \frac{t_d}{n}\right) \right|$$

When the observed value of  $T_n$  is greater than a threshold, the null hypothesis should be rejected.

In order to implement this test and find appropriate  $p$  values for the test statistics  $T_n$ , one can apply the following algorithm, suggested by [9] for a related problem.

*Step 1* Estimate  $\theta$  following (7) based on the data sample.

*Step 2* Generate  $N$  random samples of size  $n$  from  $\hat{C}_{\hat{\theta}}(\cdot)$  and, for each of these samples, estimate  $\theta$  by following (7), and determine the value of  $T_n$ .

*Step 3* Estimate the critical value of the test, at level  $\alpha$  based on  $T_n$ , by computing  $T_{(\lfloor(1-\alpha)N\rfloor)}$ , and estimate the  $p$  value corresponding to the observed value of the statistic  $T_n$ , using  $\frac{1}{N} \{j : T_{(j)} \geq T_n\}$ , where  $T_{(1)} \leq \dots \leq T_{(N)}$  denote the ordered values of the test statistics calculated in Step 2.

The properties of this procedure are currently under investigation.

## 4 Numerical results

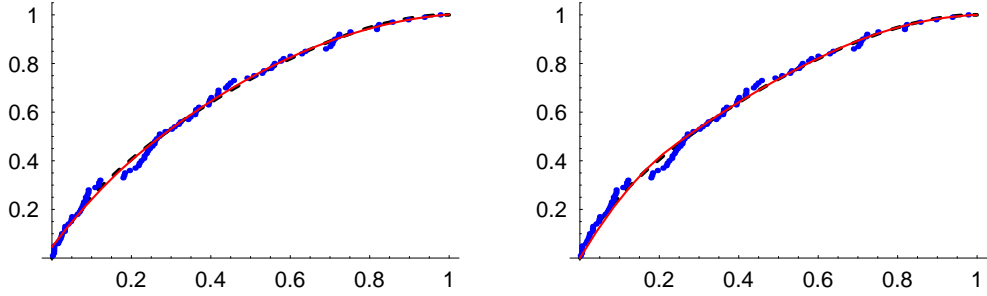
In this section, we illustrate the numerical performance of the Archimedean copula estimation procedure developed in Sections 2 and 3 on several examples.

### 4.1 The two-dimensional case ( $d = 2$ )

We start with an example, considered by [14], in which  $n = 100$  data points  $\{\frac{j}{100}, W_{(j)}\}_{j=1}^{100}$  are simulated from a two-dimensional Frank copula with Kendall's  $\tau = 0.3$  (parameter of the generator  $\theta = 2.92$ ). Applying the proposed three step GeD spline Archimedean copula estimation procedure, on step 1 we obtain the quadratic GeD spline estimate  $\hat{K}_{\alpha^*}(t; t_{2,3}^*)$ , with  $\alpha^* = (0.0459, 0.296, 0.7121, 0.977, 1.0018)$  and knots  $t_{2,3}^* = \{0, 0, 0, 0.2344, 0.6367, 1, 1, 1\}$ , presented in the left panel of Fig. 1. Step 2 of the procedure yields  $\hat{K}_{\hat{\alpha}}(t; t_{2,3}^*)$  with  $\hat{\alpha} = (0, 0.3211, 0.6974, 0.9873, 1)$ , given in the right panel of Fig. 1. As can be seen on this example with  $n = 100$  data points,  $\alpha^*$  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. 5. Another advantage of the resulting quadratic GeDS approximation  $\hat{K}_{\hat{\alpha}}(t; t_{2,3}^*)$ , 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 [14] via a complex procedure involving thousands time consuming iterations. The presented GeD spline approximation  $\hat{K}_{\hat{\alpha}}(t; t_{2,3}^*)$  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

Figure 1: Simulated data points,  $n = 100$ , from a two-dimensional Frank copula, the true underlying cdf  $K(2.92, t)$  (dashed line), the unconstrained GeDS estimate  $\hat{K}_{\alpha^*}(t; t_{2,3}^*)$  (continuous line in the left panel) and the constrained GeDS estimate  $\hat{K}_{\hat{\alpha}}(t; t_{2,3}^*)$  (continuous line in the right panel).



GeD spline estimation procedure, using (11). 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 very good accuracy with a very few parameters,  $\dim\theta = 5 + 1 + 2 = 8$ . Note that the actual dimension of  $\theta$ , reflecting the number of free parameters, is 6, since the first and the last B-spline coefficients are fixed to 0 and 1 respectively.

Our second example aims at illustrating the performance of the procedure for reasonably small,  $n = 50$ , and large,  $n = 500$ , data samples. Data points  $\{\frac{j}{n}, W_{(j)}\}_{j=1}^n$  are simulated from a two-dimensional Clayton copula with Kendall's  $\tau = 0.3$  (parameter of the generator  $\theta = 0.86$ ). The quadratic GeD spline estimate  $\hat{K}_{\alpha^*}(t; t_{1,3}^*)$  with one internal knot,  $t_4^* = 0.6094$  and  $\alpha^* = (-0.0615, 0.7608, 0.9697, 1.0093)$ , obtained in step 1, based on  $n = 50$  data points, is presented in the left panel of Fig. 3. In step 2, the constrained GeD spline estimate  $\hat{K}_{\hat{\alpha}}(t; t_{1,3}^*)$ , given in the right panel of Fig. 3, is obtained and its B-spline coefficients are  $\hat{\alpha} = (0, 0.7103, 0.9999, 1)$ . As can be seen, even with  $n = 50$  data points,  $\alpha^*$  and  $\hat{\alpha}$  for the unconstrained and the constrained spline approximations, are still reasonably close. The resulting GeD spline approximation  $\hat{K}_{\hat{\alpha}}(t; t_{1,3}^*)$  is obtained for 0.31 seconds.

In the left panel of Fig. 4, the true and estimated Clayton copula generator in the case of  $n = 50$  data points are presented. Contour plots of the true

Figure 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).*

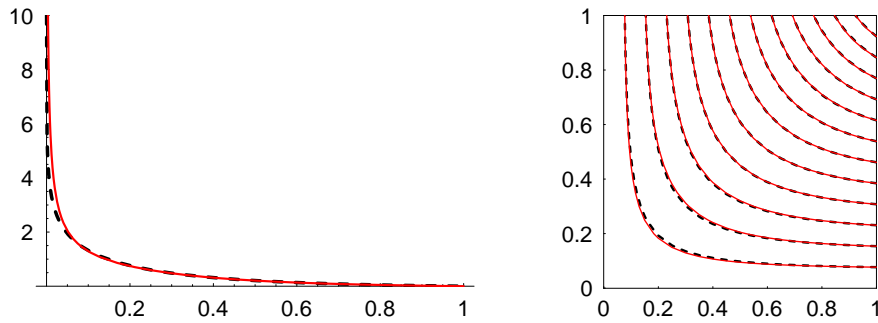


Figure 3: *Simulated data points,  $n = 50$ , from a two-dimensional Clayton copula, the true underlying cdf  $K(0.86, t)$  (dashed line), the unconstrained GeDS estimate  $\hat{K}_{\alpha^*}(t; t_{1,3}^*)$  (continuous line in the left panel) and the constrained GeDS estimate  $\hat{K}_{\hat{\alpha}}(t; t_{1,3}^*)$  (continuous line in the right panel).*

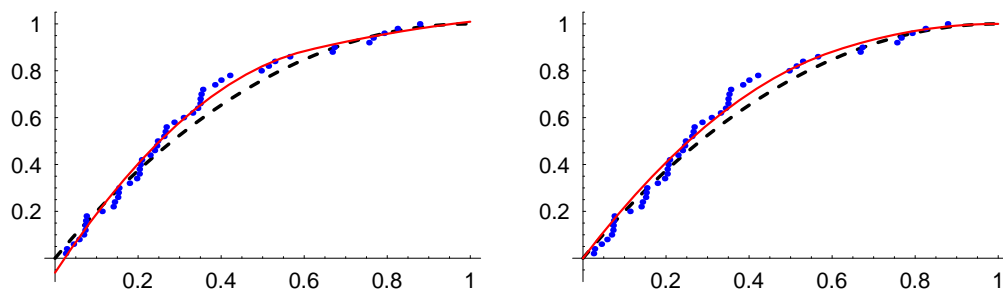
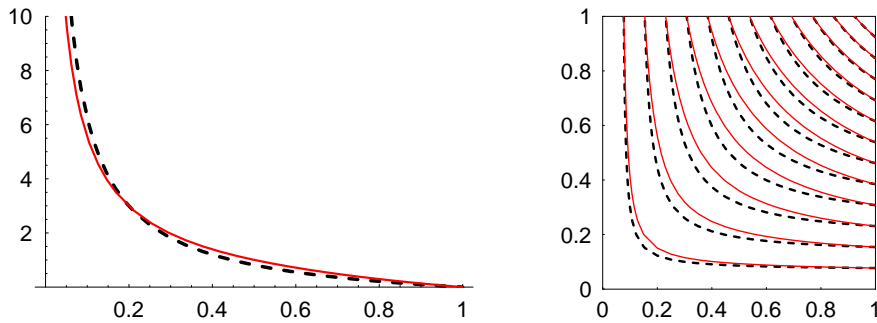


Figure 4: *Left panel: The true Clayton copula generator with Kendall's  $\tau = 0.3$  (dashed line) and its GeD spline estimate (continuous line); Right panel: Contour plots of the true Clayton copula (dashed line) and the estimated copula (continuous line).*



Clayton copula and its estimated version, resulting from step 3, are given in the right panel of Fig. 4. Here again both the generator and the copula are recovered with good accuracy and only 4 free parameters ( $\alpha_2, \alpha_3, k, t_4$ ).

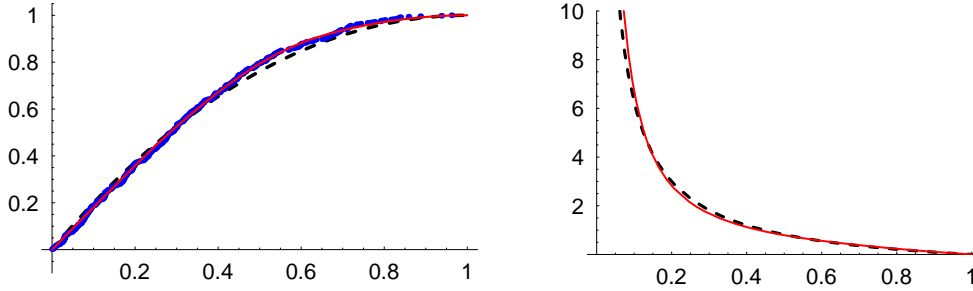
For the case of  $n = 500$ , the graph of the estimated unconstrained cdf  $\hat{K}_{\alpha^*}(t; t_{3,3}^*)$  with  $\alpha^* = (0, 0.2248, 0.6075, 0.8607, 0.9999, 1)$  and  $t_{3,3}^* = \{0, 0, 0, 0.2447, 0.4463, 0.6532, 1, 1, 1\}$ , is plotted in the left panel of Fig. 5. In this case,  $\hat{K}_{\alpha^*}(t; t_{3,3}^*)$  is obtained for 2.70 seconds and it coincides with the constrained cdf  $\hat{K}_{\hat{\alpha}}(t; t_{3,3}^*)$ . In the right panel of Fig. 5, the corresponding true and estimated Clayton copula generator are presented and as can be seen, the accuracy of the estimate,  $\hat{\phi}_{\hat{\theta}}(t)$ , has improved compared to the case  $n = 50$ .

## 4.2 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  $\{\frac{j}{n}, W_{(j)}\}_{j=1}^n$  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  $\hat{K}_{\hat{\alpha}}(t; t_{3,3}^*)$  with three internal knots,  $t_{3,3}^* = \{0, 0, 0, 0.1953, 0.3618, 0.5861, 1, 1, 1\}$  and B-spline coefficients  $\hat{\alpha} = (0, 0.2909, 0.6711, 0.8953, 0.9999, 1)$ , is presented in the left panel of Fig. 6. Using the spline estimate,  $\hat{K}_{\hat{\alpha}}(t; t_{3,3}^*)$ , the differential

Figure 5: *Left panel: simulated data points,  $n = 500$ , from a two-dimensional Clayton copula, the true underlying cdf  $K(0.86, t)$  (dashed line), the unconstrained GeDS estimate  $\hat{K}_{\alpha^*}(t; t_{3,3}^*)$  (continuous line); Right panel: The true Clayton copula generator (dashed line) and its GeD spline estimate (continuous line).*



equation (12) is solved and its solution,  $\hat{\eta}_{\hat{\theta}}(t)$ , is used to estimate the underlying generator, following (14). 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  $\hat{K}_{\hat{\alpha}}(t; t_{3,3}^*)$  is obtained for 10.61 seconds and solving (12) takes 0.05 seconds.

For the case of  $d = 4$  and  $n = 2000$ , the graph of the estimated quadratic constrained cdf  $\hat{K}_{\hat{\alpha}}(t; t_{4,3}^*)$ , where  $\hat{\alpha} = (0, 0.3406, 0.6902, 0.8534, 0.9536, 0.9999, 1)$  and  $t_{4,3}^* = \{0, 0, 0, 0.155, 0.2887, 0.426, 0.5556, 1, 1, 1\}$ , is plotted in the left panel of Fig. 7. The spline estimate  $\hat{K}_{\hat{\alpha}}(t; t_{4,3}^*)$  is obtained for 13.20 seconds and the corresponding solution of the differential equation (13) 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.

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 (8) and (9) come into play when the number of data points is relatively small, as seen from Fig. 3. In addition, in the two-dimensional case,  $d = 2$ , for small data samples,  $n \leq 200$ , one may attempt to solve the optimization problem (7) subject to the con-

Figure 6: *Left panel: simulated data points,  $n = 2000$ , from a three-dimensional Clayton copula, the true underlying cdf  $K(0.86, t)$  (dashed line), the constrained GeDS estimate  $\hat{K}_{\hat{\alpha}}(t; t_{3,3}^*)$  (continuous line); Right panel: The true Clayton copula generator (dashed line) and its GeD spline estimate (continuous line).*

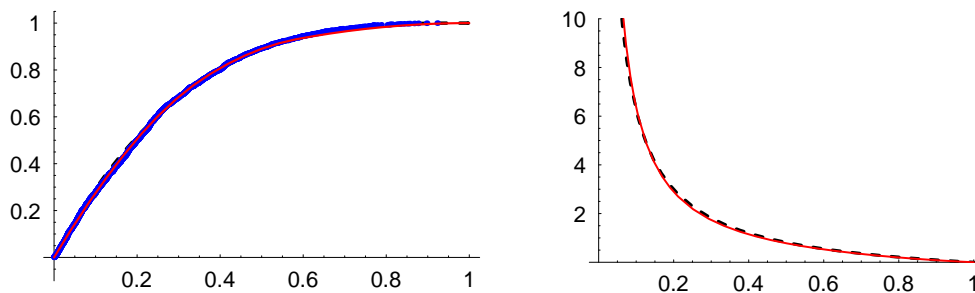


Figure 7: *Left panel: simulated data points,  $n = 2000$ , from a four-dimensional Clayton copula, the true underlying cdf  $K(0.86, t)$  (dashed line), the constrained GeDS estimate  $\hat{K}_{\hat{\alpha}}(t; t_{3,3}^*)$  (continuous line); Right panel: The true Clayton copula generator (dashed line) and its GeD spline estimate (continuous line).*

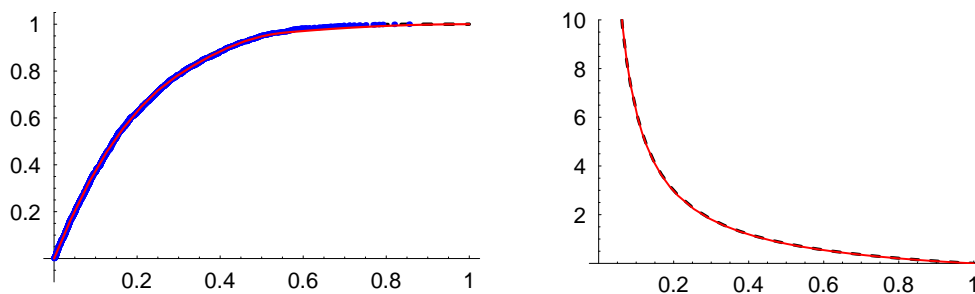
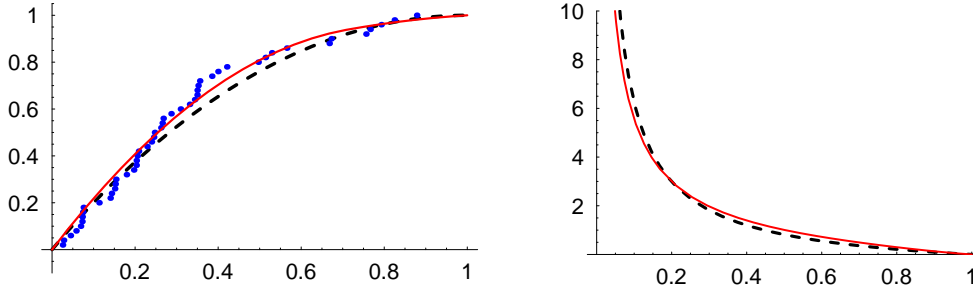


Figure 8: *Left panel: simulated data points,  $n = 50$ , from a two-dimensional Clayton copula, the true underlying cdf  $K(0.86, t)$  (dashed line), the globally optimal estimate  $\hat{K}_{\hat{\alpha}^{opt}}(t; t_{1,3}^{opt})$  (continuous line); Right panel: The true Clayton copula generator (dashed line) and its globally optimal spline estimate (continuous line).*



straints (8), (9) and (10) 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 (7) is found in 82.73 seconds and  $\hat{\alpha}^{opt} = (0, 0.8267, 0.9844, 1)$ ,  $t_4^{opt} = 0.7118$ . This optimal solution does not differ significantly from the the quadratic GeD spline estimate  $\hat{\alpha} = (0, 0.7103, 0.9999, 1)$ ,  $t_4^* = 0.6094$ , obtained in 0.31 seconds using the proposed three-step GeD spline Archimedean copula estimation method. The closeness of the two spline estimates,  $\hat{K}_{\hat{\alpha}^{opt}}(t; t_{1,3}^{opt})$  and  $\hat{K}_{\hat{\alpha}}(t; t_{1,3}^*)$ , can also be seen comparing the right panel of Fig. 3 with the left panel of Fig. 8. The corresponding two spline estimates of the generator are hard to distinguish from one another, as seen comparing the left panel of Fig. 4 with the right panel of Fig. 8. However, the globally optimal solution to (7) 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 (7) is found in 569.97 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 (7) becomes infeasible for  $d > 2$ , i.e. when higher number of internal knots is required, and for large data samples, e.g.  $n \geq 500$ .



## 5 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.

## 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*, Vol. 101, Number 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. and Rivest, L-P. (1993) Statistical Inference for bivariate Archimedean copulas. *JASA*, **88**, Issue 423, 1034-1043.
- [8] Genest, C. and Rivest, L-P. (2001) On the multivariate probability integral transformation. *Statistics & Probability Letters*, **53**, 391-399.

- [9] 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.
- [10] Kaishev, V., Dimitrova, D., Haberman, S. and Verrall, R. (2006) Geometrically designed, variable knot regression splines: Asymptotics and inference, Submitted to *JRSS(B)*.
- [11] Kaishev, V., Dimitrova, D., Haberman, S. and Verrall, R. (2006) Geometrically designed, variable knot regression splines: Variation diminishing optimality of knots, Submitted to *Computational Statistics & Data Analysis*.
- [12] 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.
- [13] Joe, H. (1977) *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.
- [14] Lambert, P. (2007) Archimedean copula estimation using Bayesian splines smoothing techniques. To appear in *Computational Statistics & Data Analysis*.
- [15] Lindstrom, M.J. (1999) Penalized estimation of free-knot splines. *J. Computational and Graphical Statistics*, 8, 2, 333-352.
- [16] Nelsen, R.B. (1999) *An introduction to copulas*. Springer, New York.
- [17] Sancetta, A and Satchell, S.(2004) The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory* **20**, 535-562.
- [18] Vandenhende, F. and Lambert, P. (2005), Local dependence estimation using semiparametric Archimedean copulas, *The Canadian Journal of Statistics*, **33**, 3, 377-388.

## FACULTY OF ACTUARIAL SCIENCE AND INSURANCE

### Actuarial Research Papers since 2001

Report Number	Date	Publication Title	Author
135.	February 2001.	On the Forecasting of Mortality Reduction Factors. ISBN 1 901615 56 1	Steven Haberman Arthur E. Renshaw
136.	February 2001.	Multiple State Models, Simulation and Insurer Insolvency. ISBN 1 901615 57 X	Steve Haberman Zoltan Butt Ben Rickayzen
137.	September 2001	A Cash-Flow Approach to Pension Funding. ISBN 1 901615 58 8	M. Zaki Khorasane
138.	November 2001	Addendum to "Analytic and Bootstrap Estimates of Prediction Errors in Claims Reserving". ISBN 1 901615 59 6	Peter D. England
139.	November 2001	A Bayesian Generalised Linear Model for the Bornhuetter- Ferguson Method of Claims Reserving. ISBN 1 901615 62 6	Richard J. Verrall
140.	January 2002	Lee-Carter Mortality Forecasting, a Parallel GLM Approach, England and Wales Mortality Projections. ISBN 1 901615 63 4	Arthur E. Renshaw Steven Haberman.
141.	January 2002	Valuation of Guaranteed Annuity Conversion Options. ISBN 1 901615 64 2	Laura Ballotta Steven Haberman
142.	April 2002	Application of Frailty-Based Mortality Models to Insurance Data. ISBN 1 901615 65 0	Zoltan Butt Steven Haberman
143.	Available 2003	Optimal Premium Pricing in Motor Insurance: A Discrete Approximation.	Russell J. Gerrard Celia Glass
144.	December 2002	The Neighbourhood Health Economy. A Systematic Approach to the Examination of Health and Social Risks at Neighbourhood Level. ISBN 1 901615 66 9	Les Mayhew
145.	January 2003	The Fair Valuation Problem of Guaranteed Annuity Options : The Stochastic Mortality Environment Case. ISBN 1 901615 67 7	Laura Ballotta Steven Haberman
146.	February 2003	Modelling and Valuation of Guarantees in With-Profit and Unitised With-Profit Life Insurance Contracts. ISBN 1 901615 68 5	Steven Haberman Laura Ballotta Nan Want
147.	March 2003.	Optimal Retention Levels, Given the Joint Survival of Cedent and Reinsurer. ISBN 1 901615 69 3	Z. G. Ignatov Z.G., V. Kaishev R.S. Krachunov
148.	March 2003.	Efficient Asset Valuation Methods for Pension Plans. ISBN 1 901615707	M. Iqbal Owadally
149.	March 2003	Pension Funding and the Actuarial Assumption Concerning Investment Returns. ISBN 1 901615 71 5	M. Iqbal Owadally

150.	Available August 2004	Finite time Ruin Probabilities for Continuous Claims Severities	D. Dimitrova Z. Ignatov V. Kaishev
151.	August 2004	Application of Stochastic Methods in the Valuation of Social Security Pension Schemes. ISBN 1 901615 72 3	Subramaniam Iyer
152.	October 2003.	Guarantees in with-profit and Unitized with profit Life Insurance Contracts; Fair Valuation Problem in Presence of the Default Option <sup>1</sup> . ISBN 1-901615-73-1	Laura Ballotta Steven Haberman Nan Wang
153.	December 2003	Lee-Carter Mortality Forecasting Incorporating Bivariate Time Series. ISBN 1-901615-75-8	Arthur E. Renshaw Steven Haberman
154.	March 2004.	Operational Risk with Bayesian Networks Modelling. ISBN 1-901615-76-6	Robert G. Cowell Yuen Y, Khuen Richard J. Verrall
155.	March 2004.	The Income Drawdown Option: Quadratic Loss. ISBN 1 901615 7 4	Russell Gerrard Steven Haberman Bjorn Hojgarrd Elena Vigna
156.	April 2004	An International Comparison of Long-Term Care Arrangements. An Investigation into the Equity, Efficiency and sustainability of the Long-Term Care Systems in Germany, Japan, Sweden, the United Kingdom and the United States. ISBN 1 901615 78 2	Martin Karlsson Les Mayhew Robert Plumb Ben D. Rickayzen
157.	June 2004	Alternative Framework for the Fair Valuation of Participating Life Insurance Contracts. ISBN 1 901615-79-0	Laura Ballotta
158.	July 2004.	An Asset Allocation Strategy for a Risk Reserve considering both Risk and Profit. ISBN 1 901615-80-4	Nan Wang
159.	December 2004	Upper and Lower Bounds of Present Value Distributions of Life Insurance Contracts with Disability Related Benefits. ISBN 1 901615-83-9	Jaap Spreeuw
160.	January 2005	Mortality Reduction Factors Incorporating Cohort Effects. ISBN 1 90161584 7	Arthur E. Renshaw Steven Haberman
161.	February 2005	The Management of De-Cumulation Risks in a Defined Contribution Environment. ISBN 1 901615 85 5.	Russell J. Gerrard Steven Haberman Elena Vigna
162.	May 2005	The IASB Insurance Project for Life Insurance Contracts: Impact on Reserving Methods and Solvency Requirements. ISBN 1-901615 86 3.	Laura Ballotta Giorgia Esposito Steven Haberman
163.	September 2005	Asymptotic and Numerical Analysis of the Optimal Investment Strategy for an Insurer. ISBN 1-901615-88-X	Paul Emms Steven Haberman
164.	October 2005.	Modelling the Joint Distribution of Competing Risks Survival Times using Copula Functions. ISBN 1-901615-89-8	Vladimir Kaishev Dimitrina S, Dimitrova Steven Haberman
165.	November 2005.	Excess of Loss Reinsurance Under Joint Survival Optimality. ISBN1-901615-90-1	Vladimir K. Kaishev Dimitrina S. Dimitrova
166.	November 2005.	Lee-Carter Goes Risk-Neutral. An Application to the Italian Annuity Market. ISBN 1-901615-91-X	Enrico Biffis Michel Denuit

167.	November 2005	Lee-Carter Mortality Forecasting: Application to the Italian Population. ISBN 1-901615-93-6	Steven Haberman Maria Russolillo
168.	February 2006	The Probationary Period as a Screening Device: Competitive Markets. ISBN 1-901615-95-2	Jaap Spreeuw Martin Karlsson
169.	February 2006	Types of Dependence and Time-dependent Association between Two Lifetimes in Single Parameter Copula Models. ISBN 1-901615-96-0	Jaap Spreeuw
170.	April 2006	Modelling Stochastic Bivariate Mortality ISBN 1-901615-97-9	Elisa Luciano Jaap Spreeuw Elena Vigna.
171.	February 2006	Optimal Strategies for Pricing General Insurance. ISBN 1901615-98-7	Paul Emms Steve Haberman Irene Savoulli
172.	February 2006	Dynamic Pricing of General Insurance in a Competitive Market. ISBN1-901615-99-5	Paul Emms
173.	February 2006	Pricing General Insurance with Constraints. ISBN 1-905752-00-8	Paul Emms
174.	May 2006	Investigating the Market Potential for Customised Long Term Care Insurance Products. ISBN 1-905752-01-6	Martin Karlsson Les Mayhew Ben Rickayzen
175.	December 2006	Pricing and Capital Requirements for With Profit Contracts: Modelling Considerations. ISBN 1-905752-04-0	Laura Ballotta
176.	December 2006	Modelling the Fair Value of Annuities Contracts: The Impact of Interest Rate Risk and Mortality Risk. ISBN 1-905752-05-9	Laura Ballotta Giorgia Esposito Steven Haberman
177.	December 2006	Using Queuing Theory to Analyse Completion Times in Accident and Emergency Departments in the Light of the Government 4-hour Target. ISBN 978-1-905752-06-5	Les Mayhew David Smith
178.	April 2007	In Sickness and in Health? Dynamics of Health and Cohabitation in the United Kingdom. ISBN 978-1-905752-07-2	Martin Karlsson Les Mayhew Ben Rickayzen
179.	April 2007	GeD Spline Estimation of Multivariate Archimedean Copulas. ISBN 978-1-905752-08-9	Dimitrina Dimitrova Vladimir Kaishev Spiridon Penev

### Statistical Research Papers

Report Number	Date	Publication Title	Author
1.	December 1995.	Some Results on the Derivatives of Matrix Functions. ISBN 1 874 770 83 2	P. Sebastiani
2.	March 1996	Coherent Criteria for Optimal Experimental Design. ISBN 1 874 770 86 7	A.P. Dawid P. Sebastiani
3.	March 1996	Maximum Entropy Sampling and Optimal Bayesian Experimental Design. ISBN 1 874 770 87 5	P. Sebastiani H.P. Wynn

4.	May 1996	A Note on D-optimal Designs for a Logistic Regression Model. ISBN 1 874 770 92 1	P. Sebastiani R. Settimi
5.	August 1996	First-order Optimal Designs for Non Linear Models. ISBN 1 874 770 95 6	P. Sebastiani R. Settimi
6.	September 1996	A Business Process Approach to Maintenance: Measurement, Decision and Control. ISBN 1 874 770 96 4	Martin J. Newby
7.	September 1996.	Moments and Generating Functions for the Absorption Distribution and its Negative Binomial Analogue. ISBN 1 874 770 97 2	Martin J. Newby
8.	November 1996.	Mixture Reduction via Predictive Scores. ISBN 1 874 770 98 0	Robert G. Cowell.
9.	March 1997.	Robust Parameter Learning in Bayesian Networks with Missing Data. ISBN 1 901615 00 6	P. Sebastiani M. Ramoni
10.	March 1997.	Guidelines for Corrective Replacement Based on Low Stochastic Structure Assumptions. ISBN 1 901615 01 4.	M.J. Newby F.P.A. Coolen
11.	March 1997	Approximations for the Absorption Distribution and its Negative Binomial Analogue. ISBN 1 901615 02 2	Martin J. Newby
12.	June 1997	The Use of Exogenous Knowledge to Learn Bayesian Networks from Incomplete Databases. ISBN 1 901615 10 3	M. Ramoni P. Sebastiani
13.	June 1997	Learning Bayesian Networks from Incomplete Databases. ISBN 1 901615 11 1	M. Ramoni P. Sebastiani
14.	June 1997	Risk Based Optimal Designs. ISBN 1 901615 13 8	P. Sebastiani H.P. Wynn
15.	June 1997.	Sampling without Replacement in Junction Trees. ISBN 1 901615 14 6	Robert G. Cowell
16.	July 1997	Optimal Overhaul Intervals with Imperfect Inspection and Repair. ISBN 1 901615 15 4	Richard A. Dagg Martin J. Newby
17.	October 1997	Bayesian Experimental Design and Shannon Information. ISBN 1 901615 17 0	P. Sebastiani. H.P. Wynn
18.	November 1997.	A Characterisation of Phase Type Distributions. ISBN 1 901615 18 9	Linda C. Wolstenholme
19.	December 1997	A Comparison of Models for Probability of Detection (POD) Curves. ISBN 1 901615 21 9	Wolstenholme L.C
20.	February 1999.	Parameter Learning from Incomplete Data Using Maximum Entropy I: Principles. ISBN 1 901615 37 5	Robert G. Cowell
21.	November 1999	Parameter Learning from Incomplete Data Using Maximum Entropy II: Application to Bayesian Networks. ISBN 1 901615 40 5	Robert G. Cowell
22.	March 2001	FINEX : Forensic Identification by Network Expert Systems. ISBN 1 901615 60X	Robert G. Cowell
23.	March 2001.	Wren Learning Bayesian Networks from Data, using Conditional Independence Tests is Equivalent to a Scoring Metric ISBN 1 901615 61 8	Robert G Cowell
24.	August 2004	Automatic, Computer Aided Geometric Design of Free-Knot, Regression Splines. ISBN 1-901615-81-2	Vladimir K Kaishev, Dimitrina S. Dimitrova, Steven Haberman Richard J. Verrall
25.	December 2004	Identification and Separation of DNA Mixtures Using Peak Area Information. ISBN 1-901615-82-0	R.G. Cowell S.L. Lauritzen

- |            |                |   |  |
|------------|----------------|---|--|
| <b>26.</b> | November 2005. | The Quest for a Donor : Probability Based Methods Offer Help. ISBN 1-90161592-8   | J Mortera,<br>P.F.Mostad<br>T. Egeland.,<br>R.G. Cowell<br>V. Bosnes<br>Ø. Braaten     |
| <b>27.</b> | February 2006  | Identification and Separation of DNA Mixtures Using Peak Area Information. (Updated Version of Research Report Number 25). ISBN 1-901615-94-4 | R.G.Cowell<br>S.L.Lauritzen<br>J Mortera,  |
| <b>28.</b> | October 2006   | Geometrically Designed, Variable Knot Regression Splines : Asymptotics and Inference. ISBN 1-905752-02-4                                      | Vladimir K Kaishev,<br>Dimitrina S.Dimitrova,<br>Steven Haberman<br>Richard J. Verrall |
| <b>29.</b> | October 2006   | Geometrically Designed, Variable Knot Regression Splines : Variation Diminishing Optimality of Knots. ISBN 1-905752-03-2                      | Vladimir K Kaishev,<br>Dimitrina S.Dimitrova,<br>Steven Haberman<br>Richard J. Verrall |

*Papers can be downloaded from*

*<http://www.cass.city.ac.uk/arc/actuarialreports.html>*

# Faculty of Actuarial Science and Insurance

---

## Actuarial Research Club

The support of the corporate members

- CGNU Assurance
- English Matthews Brockman
- Government Actuary's Department

is gratefully acknowledged.

Copyright 2007 © Faculty of Actuarial Science and Insurance,  
Cass Business School  
106 Bunhill Row,  
London EC1Y 8TZ.

ISBN 978-1-905752-08-9