Copula Goodness-of-fit Tests: A Comparative Study**

Daniel Berg

The Norwegian Computing Center and University of Oslo.
E-mail: daniel@danielberg.no

Henrik Bakken

The Norwegian Computing Center and The Norwegian University of Science and Technology.
E-mail: Bakken.Henrik@bcg.com

[Revised May 12th 2006]

Abstract. Copulae is one of the main ways of modelling dependence. However, to check whether the dependency structure of a data set is appropriately modelled by a chosen copula, there is no recommended method agreed upon. Several copula goodness-of-fit approaches have been proposed lately. Many reduce the multivariate problem to a univariate one and then apply a univariate test. This makes the approach numerically efficient even for large dimensional problems. In this paper we fit three such approaches into the same framework to ease comparison. Generalizations and improvements are attempted. We supplement our analysis with a full multivariate approach. We examine properties, strengths and weaknesses of each approach. In addition we compare the power of the approaches at distinguishing tail heaviness and skewness properties. Results show that there are several alternatives for the bivariate case while for higher dimensions the approach proposed by Berg and Bakken (2005) stands out as having superior power and being the most flexible approach with regards to weighting of particular regions of the copula. Concluding remarks and recommendations are made.

Keywords: Copula, Goodness-of-fit, Probability Integral Transform

1. Introduction

Copulae have proved to be a very useful tool in the analysis of dependency structures. The concept of copulae was introduced by Sklar (1959), but was first used in financial applications by Embrechts et al. (1999). Since then we have seen a tremendous increase of copula related research and applications.

The limitation of the copula approach is the lack of a recommended way of checking whether the dependency structure of a data set is appropriately modelled by the chosen copula. Information criterions, such as Akaike's Information Criterion (AIC), are not able to provide any understanding about the power of the decision rule employed. Goodness-of-fit (GOF) approaches on the other hand, are able to reject or fail to reject a parametric copula and are thus preferred.

Several copula GOF approaches have been proposed in literature. Genest and Rivest (1993) have developed an empirical method to identify the best copula in the Archimedean case. Fermanian (2003) approximates the underlying probability density function by kernel smoothing of the empirical density. Diebold et al. (1998), Diebold et al. (1999), Hong (2000), Berkowitz (2001), Thompson (2002) and Breymann et al. (2003) focus on the probability integral transform (PIT) of Rosenblatt (1952) in the evaluation of density models. Berg and Bakken (2005) also focus on the PIT and a transformation of the PIT data. Panchenko (2005) focuses on positive definite bilinear forms while Genest et al. (2006) utilize the Kendall's process and the empirical copula in the Archimedean case. Some approaches are full multivariate approaches while several approaches reduce the multivariate problem to a univariate problem, and then apply some univariate test. The latter approach leads to numerically efficient approaches for high dimensional problems.

**This research was supported by the Norwegian Research Council, grant number 154079/420.

In this paper we consider three such dimension reduction approaches, namely the ones by Breymann et al. (2003), Berg and Bakken (2005) and Genest et al. (2006). We supplement our analysis by considering the full multivariate approach by Panchenko (2005). Power at detecting tail heaviness and skewness properties will be examined and compared. The objective of this paper is to provide an overview of the copula GOF approaches proposed lately, comparing them and illustrating various properties, strengths and weaknesses.

The paper is organized as follows. In Section 2 we recall some basic theory that will be useful, such as copulae, the probability integral transform and univariate test statistics. Section 3 examines the four GOF approaches. In Section 4 we present results while Section 5 concludes.

2. Basic theory

2.1. Copula Basics

The definition of a d-dimensional copula is a multivariate distribution C, with uniform margins $U(0,1)$. Sklar (1959)'s theorem states that every multivariate distribution F with margins $F_1, \ldots, F_d$ can be written as

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)), \tag{2.1}$$

for some copula C. If we have a random vector $\mathbf{X} = (X_1, \ldots, X_d)$, the copula of their joint distribution function may be extracted from Equation (2.1):

$$C(u_1, \ldots, u_d) = F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)), \tag{2.2}$$

where the F_i^{-1}'s are the cumulative distribution functions (cdf) of the margins.

For the implicit copula of an absolutely continuous joint distribution function F, with strictly continuous marginal distribution functions $F_1, \ldots, F_d$, the copula density is given by

$$c(\mathbf{u}) = \frac{f(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \cdots f_d(F_d^{-1}(u_d))}. \tag{2.3}$$

Hence,

$$c(F_1(x_1), \ldots, F_d(x_d)) = \frac{f(x_1, \ldots, x_d)}{f_1(x_1) \cdots f_d(x_d)}. \tag{2.4}$$

This means that a general d-dimensional density can be written as

$$f(x_1, \ldots, x_d) = c(F_1(x_1), \ldots, F_d(x_d)) \cdot f_1(x_1) \cdots f_d(x_d) \tag{2.5}$$

for some copula density $c(\cdot)$.

The most attractive properties of copulae, making it so popular for financial applications, is the decoupling of the copula and the margins and the invariance to strictly increasing transformations. For a thorough analysis of copulae, see Joe (1997) or Nelsen (1999).

2.2. Empirical Distributions

For copula GOF testing we are interested in the fit of the copula alone. We do not wish to introduce any distributional assumptions for the margins. Instead we use empirical margins to transform the observed data set into the observed copula.

The empirical marginal cdf for n observations $X_{1i}, \ldots, X_{ni}$ of a variable X_i is

$$\widehat{F}_i(x) = \frac{1}{n+1} \sum_{j=1}^{n} I(X_{ji} \leq x) \qquad i = 1, \ldots, d, \tag{2.6}$$

where $I(\cdot)$ is the indicator function returning 1 if $X_{ji} \leq x$ and 0 otherwise. Here $n+1$ is used for division to keep the empirical cdf lower than 1. The empirical distribution converges towards the actual distribution function as $n \to \infty$. We can then define the empirical uniform values $u_{ji} = \widehat{F}_i(x_{ji}), i = 1, \ldots, d, j = 1, \ldots, n$.

Using the empirical marginal cdf's, the empirical copula is given by

$$\begin{aligned} \widehat{C}(\mathbf{u}) &= \frac{1}{n+1} \sum_{j=1}^{n} I\left(\widehat{F}_1(X_{j1}) \le u_1, \ldots, \widehat{F}_d(X_{jd}) \le u_d\right) \\ &= \frac{1}{n+1} \sum_{j=1}^{n} I\left(U_{j1} \le u_1, \ldots, U_{jd} \le u_d\right). \end{aligned} \tag{2.7}$$

where $\mathbf{u} = (u_1, \ldots, u_d)$. The empirical copula is the observed frequency of $P(U_1 < u_1, \ldots, U_d < u_d)$.

2.3. The Probability Integral Transform

The PIT transforms a set of dependent variables into a new set of independent $U(0,1)$ variables, given the multivariate distribution. The PIT is a universally applicable way of creating a set of iid $U(0,1)$ variables from any data set with known distribution. Given a test for multivariate, independent uniformity, this transformation can be used to test the fit of any assumed model. The concept was first introduced by Rosenblatt (1952) and can be interpreted as the inverse of simulation.

DEFINITION 2.1 (PROBABILITY INTEGRAL TRANSFORM). *Let* $\mathbf{X} = (X_1, \ldots, X_d)$ *denote a random vector with marginal distributions* $F_i(x_i) = P(X_i \le x_i)$ *and conditional distributions* $F_{i|1\ldots i-1}(X_i \le x_i | X_1 = x_1, \ldots, X_{i-1} = x_{i-1})$ *for* $i = 1, \ldots, d$. *The PIT of* $\mathbf{X}$ *is defined as* $T(\mathbf{X}) = (T_1(X_1), \ldots, T_d(X_d))$ *where* $T_i(X_i)$ *is defined as follows:*

$$\begin{aligned} T_1(X_1) &= P(X_1 \le x_1) = F_1(x_1), \\ T_2(X_2) &= P(X_2 \le x_2 | X_1 = x_1) = F_{2|1}(x_2|x_1), \\ &\vdots \\ T_d(X_d) &= P(X_d \le x_d | X_1 = x_1, \ldots, X_{d-1} = x_{d-1}) = F_{d|1\ldots d-1}(x_d | x_1, \ldots, x_{d-1}). \end{aligned}$$

The random variables $Z_i = T_i(X_i),\ i = 1, \ldots, d$ *are uniformly and independently distributed on* $[0,1]^d$.

A recent application of the PIT is multivariate GOF tests. A data set is first PIT under a null hypothesis, and then a test of multivariate independence is performed. The null hypothesis may be e.g. a parametric copula family. The parameters of this copula needs to be estimated as a part of the PIT. Genest et al. (1995) gives an introduction to parameter estimation for copulae.

An advantage with the PIT in this setting is that the null- and alternative hypotheses are the same, regardless of the distribution before the PIT. The PIT also enables weighting in a simple way since the data, under $\mathcal{H}_0$, is always iid $U(0,1)$. Hong and Li (2002) report Monte Carlo evidence of multivariate tests using PIT variables outperforming tests using the original random variables. Chen et al. (2004) believe that a similar conclusion also applies to GOF tests for copulae. A disadvantage with the PIT is that it depends on the permutation of the variables. However, as long as the permutation is decided randomly, the results will not be influenced in any particular direction and will thus be consistent.

For more details on the PIT see e.g. Rosenblatt (1952), D'Agostino and Stephens (1986) or Breymann et al. (2003).

2.4. Univariate Goodness-of-fit Test Statistics

There are several univariate GOF test statistics to choose among. Two main categories emerge, based either on the max-function or integration, and either on the empirical cdf, $\widehat{F}$, or the empirical probability distribution function (pdf), $\widehat{f}$. For a thorough treatment of univariate GOF test statistics, see e.g. D'Agostino and Stephens (1986).

Suppose we have a random vector $\mathbf{W} = (w_1, \ldots, w_n)$ which is iid $U(0,1)^n$. Suppose further that the cdf of $\mathbf{W}$ is $F(w)$ and the pdf of $\mathbf{W}$ is $f(w)$. Since $\mathbf{W}$ is iid $U(0,1)^n$ we have that $F(w) = w$ and $f(w) = 1$.

2.4.1. *CDF Test Statistics*

We examine the following cdf statistics: Kolmogorov-Smirnov (KS), Cramér-von Mises (CvM), Anderson-Darling (AD) and Kolmogorov-Smirnov Anderson-Darling (KSAD). These statistics may be referred to by different names elsewhere. We define them as follows:

$$\mathcal{T}^{KS} = \sup_w \left|\sqrt{n}(\widehat{F}(w) - F(w))\right|, = \sup_w \left|\sqrt{n}(\widehat{F}(w) - w)\right|, \tag{2.8}$$

$$\mathcal{T}^{CvM} = n \int_0^1 \left(\widehat{F}(w) - F(w)\right)^2 \mathrm{d}F(w) = n \int_0^1 \left(\widehat{F}(w) - w\right)^2 \mathrm{d}w, \tag{2.9}$$

$$\mathcal{T}^{AD} = n \int_0^1 \frac{\left(\widehat{F}(w) - F(w)\right)^2}{F(w)(1-F(w))} \mathrm{d}F(w) = n \int_0^1 \frac{\left(\widehat{F}(w) - w\right)^2}{w(1-w)} \mathrm{d}w, \tag{2.10}$$

$$\mathcal{T}^{KSAD} = \sup_w \left|\sqrt{n}\frac{\widehat{F}(w)-F(w)}{F(w)(1-F(w))}\right| = \sup_w \left|\sqrt{n}\frac{\widehat{F}(w)-w}{w(1-w)}\right|, \tag{2.11}$$

where $\widehat{F}$ is the empirical cdf.

The empirical versions of these statistics are derived in Appendix B. They are:

$$\widehat{\mathcal{T}}^{KS} = \sqrt{n} \max_{i=0,1;0<j\leq n} \left\{\left|\widehat{F}\left(\tfrac{j}{n+1}\right) - \tfrac{j+i}{n+1}\right|\right\}, \tag{2.12}$$

$$\widehat{\mathcal{T}}^{CvM} = \frac{n}{3} + \frac{n}{n+1}\sum_{j=1}^n \widehat{F}\left(\frac{j}{n+1}\right)^2 - \frac{n}{(n+1)^2}\sum_{j=1}^n (2j+1)\widehat{F}\left(\frac{j}{n+1}\right), \tag{2.13}$$

$$\widehat{\mathcal{T}}^{AD} = \frac{n}{n+1}\sum_{j=1}^n \frac{\left(\widehat{F}\left(\frac{j}{n+1}\right) - \frac{j}{n+1}\right)^2}{\frac{j}{n+1}\left(1-\frac{j}{n+1}\right)}, \tag{2.14}$$

$$\widehat{\mathcal{T}}^{KSAD} = \sqrt{n} \max_{i=0,1;0<j\leq n-1 \text{ or } i=0;j=n} \left\{\left|\frac{\widehat{F}\left(\frac{j}{n+1}\right) - \frac{j}{n+1}}{\frac{j+i}{n+1}\left(1-\frac{j+i}{n+1}\right)}\right|\right\}. \tag{2.15}$$

The KS statistic is known to be most sensitive around the median of the distribution and relatively insensitive to deviations in the tails. The latter due to lower empirical cdf variance (Aslan and Zech, 2002). The CvM statistic is a member of the Cramér-von Mises family. This statistic is likely to be very stable since areas of high probability are emphasized. It is however insensitive to tail deviance. The AD statistic is another a member of the Cramér-von Mises family. The normalization means that this statistic does not have any bias, neither at the center nor at the tails of the distribution. It strongly weights deviations near $w = 0$ and $w = 1$. This is justified by the small experimental deviations here due to the constraints $\widehat{F}(w) - w = 0$ at $w = 0$ and $w = 1$. For general $U(0,1)^d$ tests, the AD statistic is the best for cdfs since it has no bias, in contrast to the CvM and KS statistics. The KSAD statistic is an altered KS statistic. This statistic does not have the bias toward the center as the KS statistic has. Thus it will be a non-biased maximum average statistic.

2.4.2. *PDF Test Statistics*

When using PIT variables, the density function $f(w)$ equals unity everywhere w is defined. Thus, the χ^2 statistic coincides with the L2 norm statistic. Thus, we only consider the latter:

$$\mathcal{T}^{L2} = n\int_0^1 \left(\widehat{f}(w) - f(w)\right)^2 \mathrm{d}w = n\int_0^1 \left(\widehat{f}(w) - 1\right)^2 \mathrm{d}w, \tag{2.16}$$

To approximate $\widehat{f}$ we consider two approaches, kernel density estimation (KDE) (for an introduction to KDE see e.g. Azzalini and Bowman (1997) or Peterson (2004)) and binning. For the KDE approach, the integral function of the L2 norm test statistic needs to be discretized to enable numerical calculation.

For the binning approach, the interval $[0, 1]$ is divided into disjoint subsets $A_1, \dots, A_p$ and the empirical estimate of the density, the number of observed z's in subset A_i per observation n, is given as $\widehat{P}(w \in A_i) = \#\big[w_j \in A_i, j = 1, \dots, n\big]/n$. This subset is chosen as an evenly spaced grid since the variables are uniform under the null hypothesis. For both approaches, $q = 25$ segments are chosen for discretization. This is the number of segments used in Chen et al. (2004).

The empirical versions of the L2 norm pdf statistics are

$$\widehat{\mathcal{T}}^{L2KDE} = \frac{n}{q}\sum_{i=1}^{q}\left(\widehat{f}\left(\frac{i-1/2}{q}\right) - 1\right)^2, \tag{2.17}$$

$$\widehat{\mathcal{T}}^{L2BIN} = n\sum_{i=1}^{q}\left(\widehat{P}\left(w \in \left[\frac{i-1}{q}, \frac{i}{q}\right]\right) - \frac{1}{q}\right)^2. \tag{2.18}$$

The discretization of the integrals is done by assuming $\widehat{f}(w) = \widehat{f}((i - 1/2)/q)$ for $w \in [(i - 1)/q, i/q]$.

3. Copula Goodness-of-fit Testing

For univariate distributions, the GOF assessment can be performed by e.g. the well-known Anderson-Darling (Anderson and Darling, 1954) test, or less quantitatively using a QQ-plot. In the multivariate domain there are fewer alternatives. In financial applications, economic theory sheds little light on the dependence structure between financial assets. Multivariate normality is often assumed a priori. Empirical studies shows, however, that more appropriate dependence structures are available (Chen et al., 2004; Dobrić and Schmid, 2005).

GOF approaches for copulae is a special case of the more general problem of testing multivariate density models, but is complicated due to the unspecified marginal distributions. Empirical margins are used and this introduces infinitely many nuisance parameters. This complicates the deduction of the asymptotic distribution properties for the approaches and p-values are commonly found by simulation. This is computationally very intensive and much effort is invested trying to solve this issue.

A simple way to build GOF approaches for multivariate random variables is to consider multi-dimensional chi-square approaches, as in Pollard (1979), D'Agostino and Stephens (1986) and Snedecor and Cochran (1986). The problem with these approaches, as with all binned approaches based on gridding the probability space, is that they will not be feasible for high dimensional problems since the need for data would be too great. Another issue with binned approaches is that the grouping of the data is not trivial. Grouping too coarsely destroys valuable information and the ability to contrast distributions becomes very limited. On the other hand, too small groups leads to highly irregular empirical cdf's due to the limited amount of data. For these reasons we will not consider multivariate binned approaches.

Further, a GOF approach is most useful for high-dimensional problems since these copulae are harder to conceptualize and because the consequences of poor model choice is often much greater in higher dimensional problems, e.g. risk assessments for high dimensional financial portfolios. For these reasons we will not consider multidimensional KDE approaches such as the one proposed by Fermanian (2003). A multidimensional KDE for high dimensional problems would simply be too computationally demanding.

The class of dimension reduction approaches is a better alternative. We consider the approaches proposed by Breymann et al. (2003), Berg and Bakken (2005) and Genest et al. (2006). We supplement our analysis with the full multivariate approach proposed by Panchenko (2005). We base all four approaches on the PIT and some generalizations and improvements are attempted.

3.1. Dimension Reduction Approaches

We now introduce the three dimension reduction approaches. The testing procedure is given in Algorithm 3.1.

3.1.1. Breymann, Dias and Embrechts' Approach

The approach proposed by Breymann et al. (2003) coincides with the approach proposed by Malevergne and Sornette (2003) when the latter is based on PIT data. It also coincides with the second approach in Chen et al. (2004).

Let $\mathbf{Z} = (Z_1, \ldots, Z_d)$ be the iid $U(0,1)^d$ variables obtained from applying the PIT to a multivariate data set $\mathbf{X} = (X_1, \ldots, X_d)$. The dimension reduction is then performed as follows:

$$Y_G = \sum_{i=1}^{d} \Phi^{-1}(Z_i)^2, \tag{3.1}$$

where $\Phi^{-1}(\cdot)$ is the inverse Gaussian cdf. Since the Z_i's are iid $U(0,1)$ under $\mathcal{H}_0$, the variables $\Phi^{-1}(Z_i)$ are iid $\mathcal{N}(0,1)$. Hence Y_G is χ^2_d distributed and we define

$$W_G = F_{\chi^2_d}(Y_G), \tag{3.2}$$

which is an iid $U(0,1)^n$ vector under $\mathcal{H}_0$. We have now reduced the multivariate problem to a univariate one, and the approach G is defined as the cdf of W_G:

$$F_G(w) = P\left(W_G \le w\right), \qquad w \in [0,1]. \tag{3.3}$$

Under the null hypothesis $F_G(w) = w$ and the density function $f_g(w) = 1$. Given n observations of the d-dimensional vector $\mathbf{Z}$, the empirical version, $\widehat{F}_G(w)$, equals:

$$\widehat{F}_G(w) = \frac{1}{n+1} \sum_{j=1}^{n} I(W_{Gj} \le w), \qquad w = \frac{1}{n+1}, \ldots, \frac{n}{n+1}. \tag{3.4}$$

Breymann et al. (2003) apply the Anderson-Darling test statistic to F_G. Malevergne and Sornette (2003) apply various cdf statistics to F_G while Chen et al. (2004) proposed to use the L2 norm KDE statistic on f_g. We have generalized this so any univariate cdf or pdf test statistic may be applied to F_G and f_g, respectively.

This approach is computationally very efficient. However, it has its weaknesses. First of all the approach is not consistent, meaning that the resulting test statistic is not strictly increasing for every deviation from the null hypothesis. Some deviations may be overlooked. The projection from a multivariate problem to a univariate problem through Y_G is what causes this inconsistency. Another feature with this approach that may be considered a weakness is the fact that it weights the tails of the copula. If we are not particularly interested in the fit in the tails, such a tail weighting may be undesirable.

3.1.2. Berg and Bakken's Approach

The approach proposed by Berg and Bakken (2005) extends the approach in Breymann et al. (2003). However, the consistency issue is solved by transforming the PIT data before the dimension reduction. The deviance measure is also decoupled from the weighting functionality. Any weight function may be used, enabling the weighting of any region of the copula.

Let $\mathbf{Z} = (Z_1, \ldots, Z_d)$ be the iid $U(0,1)^d$ variables obtained from applying the PIT to a multivariate data set $\mathbf{X} = (X_1, \ldots, X_d)$. Define a new vector $\mathbf{Z}^*$ as

$$Z_i^* = P(r_i \le \widetilde{Z}_i | r_1, \ldots, r_{i-1}) = \left(1 - \left(\frac{1 - \widetilde{Z}_i}{1 - r_{i-1}}\right)^{d-(i-1)}\right), \tag{3.5}$$

for $i = 1, \ldots, d$, where $\widetilde{\mathbf{Z}} = (\widetilde{Z}_1, \ldots, \widetilde{Z}_d)$ is the sorted counterpart of $\mathbf{Z}$ and r_i is rank variable i[1] from $\mathbf{Z}$. The dimension reduction is then performed using $\mathbf{Z}^*$:

$$Y_B = \sum_{i=1}^{d} \gamma(Z_i; \boldsymbol{\alpha}) \cdot \Phi^{-1}(Z_i^*)^2, \tag{3.6}$$

[1]Rank variables are the observed variables, ordered ascendingly.

where γ is a weight function used for weighting $\Phi^{-1}(Z_i^*)^2$ depending on its corresponding value Z_i, and $\boldsymbol{\alpha}$ is the set of weight parameters. Further let $F_{Y_B}(\cdot)$ be the cdf of Y_B, i.e. the cdf of a linear combination of squared normal variables, found either numerically (see e.g. Farebrother (1990)) or by simulation. We then define

$$W_B = F_{Y_B}(Y_B), \tag{3.7}$$

which is an iid $U(0,1)^n$ vector under $\mathcal{H}_0$. The approach B is then defined as the cdf of W_B:

$$F_B(w) = P\left(W_B \leq w\right), \qquad w \in [0,1]. \tag{3.8}$$

Under the null hypothesis $F_B(w) = w$ and the density function $f_b(w) = 1$. The empirical version, $\widehat{F}_B(w)$, equals:

$$\widehat{F}_B(w) = \frac{1}{n+1}\sum_{j=1}^{n} I(W_{Bj} \leq w), \qquad w = \frac{1}{n+1}, \ldots, \frac{n}{n+1}. \tag{3.9}$$

This approach is computationally quite fast. It is slower than the approach G, due to the distribution of the linear combination of squared normal variables. Further, the weighting functionality adds valuable flexibility. Any weight function can be applied to any region of the copula. The fit of different specific regions of the copula may be of special interest for different applications and may hence be weighted more than the rest, e.g. the tails for credit risk analysis. In this paper we apply the approach B with no weight and with power tail weight, $\gamma(Z_i;\alpha) = (Z_i - \frac{1}{2})^{\alpha}, \quad \alpha \in (2,4,\ldots)$.

The transformation Z^* in Equation (3.5) enables a consistent dimension reduction, without loosing any information. The rationale behind the transformation can be explained as follows. To solve the consistency problem of the approach G, we wish to find the probability, under $\mathcal{H}_0$, that the variable with rank i, given the variables with rank $1,\ldots,i-1$, will be smaller than or equal to the observed variable with rank i, $\widetilde{Z}_i$. Hence, we wish to find $P(r_i \leq \widetilde{Z}_i | r_1,\ldots,r_{i-1}) = 1 - P(r_i > \widetilde{Z}_i | r_1,\ldots,r_{i-1})$. The only way r_i can be greater than $\widetilde{Z}_i$ is if all remaining $d-(i-1)$ variables are greater than $\widetilde{Z}_i$. Since the remaining $d-(i-1)$ variables are independent, the probability of all being greater than $\widetilde{Z}_i$ is the product over the probabilities of each $r_k,\ k \in [i,d]$, being greater than $\widetilde{Z}_i$:

$$P(\widetilde{Z}_i < r_k < 1 | r_k > r_{i-1}) = \frac{P(r_k > \widetilde{Z}_i \cap r_k > r_{i-1})}{P(r_k > r_{i-1})} = \frac{P(r_k > \widetilde{Z}_i)}{P(r_k > r_{i-1})} = \frac{1-\widetilde{Z}_i}{1-r_{i-1}}, \quad k \in [i,d].$$

3.1.3. Genest, Quessy and Rémillard's Procedure

Genest et al. (2006) propose an approach for Archimedean copulae based on Kendall's process (for an introduction to Kendall's process, see Barbe et al. (1996)). They utilize the empirical copula cdf for the dimension reduction. We first shortly introduce the approach as it is presented in Genest et al. (2006), and then we attempt some generalizations and improvements.

Original Approach

Given F as defined in Equation (2.1), the $\mathcal{H}_0$ copula C, with parameters θ and the observed multivariate data set $\mathbf{X}$, Genest et al. (2006) define

$$F_K(w) = P(F(\mathbf{X}) \leq w) = P(C(\mathbf{U}) \leq w), \qquad w \in [0,1]. \tag{3.10}$$

Its density function is given by $f_k(w) = \partial F_K(w)/\partial w$. Under the null hypothesis, $F_K(w) = F_{K,0}(w)$, where $F_{K,0}(w)$ is copula specific and must be derived for all copulae used. This can be seen by rewriting $F_K(w)$ as

$$F_K(w) = \int_0^1 \ldots \int_0^1 I(C(u_1,\ldots,u_d) \leq w)\ c(u_1,\ldots,u_d)\ \mathrm{d}u_1 \ldots \mathrm{d}u_d, \tag{3.11}$$

where $I(\cdot)$ is the regular indicator function and $c(u_1,\ldots,u_d)$ is the copula density, which is copula specific. The empirical version of $F_K(w)$, $\widehat{F}_K(w)$, equals

$$\widehat{F}_K(w)=\frac{1}{n}\sum_{j=1}^{n}I(\widehat{F}(\mathbf{x}_j)\leq w)=\frac{1}{n}\sum_{j=1}^{n}I(\widehat{C}(\mathbf{u}_j)\leq w), \tag{3.12}$$

where $\mathbf{u}_j$ is the j-th observation of the d-dimensional vector $\mathbf{u}$, $\widehat{F}(\mathbf{x}_j)$ is from Equation (2.6) and $\widehat{C}(\mathbf{u})$ is the empirical copula cdf as defined in Equation (2.7).

Genest et al. (2006) specify some hypotheses based on $\sqrt{n}(\widehat{F}_K(w)-F_K(w,\boldsymbol{\theta}))$, known as Kendall's process. They further derive two univariate test statistics that they apply to $F_K(w)$, a CvM and a KS statistic. Implementation of these test statistics is shown in Genest et al. (2006). The distribution of these test statistics are, unfortunately, dependent on the null hypothesis copula. Thus, there is no simple way of obtaining the critical values for a given level of significance α. A parametric bootstrap or Monte Carlo simulation can be used.

Altered Approach

In its original form $F_{K,0}(w)$ needs to be derived for every copula, and applications to copulae that are not Archimedean is, if possible, difficult. Using a PIT approach implies that $F_{K,0}(w)$ only needs to be derived for one copula and the approach can be used for all copulae that we can PIT. We thus alter the approach to be based on PIT data. We also improve the approach slightly by including the last observation in the approach. This is done by altering $\widehat{F}_K(w)$ to use the denominator $n+1$ instead of n to avoid $\widehat{F}_K(w)$ reaching 1. This enables the inclusion of the last observation in the calculation of the CvM test statistic (see Genest et al. (2006, page 11)). An issue that will also be addressed is that $\lim_{w\to 0} f_k(w)\to\infty$, and this discontinuity is not desirable.

As before let $\mathbf{Z}$ be the uniformly and independently distributed variables on $[0,1]^d$, obtained from applying the PIT to the multivariate data set $\mathbf{X}$. The approach K is then defined as the cdf of the copula:

$$F_K(w)=P(C(\mathbf{Z})\leq w),\qquad w\in[0,1], \tag{3.13}$$

and its density function equals $f_k(w)=\partial F_K(w)/\partial w$.

The empirical version of $F_K(w)$, $\widehat{F}_K(w)$, equals:

$$\widehat{F}_K(w)=\frac{1}{n+1}\sum_{j=1}^{n}I\left(\widehat{C}(\mathbf{z}_j)\leq w\right),\qquad w=\frac{1}{n+1}\ldots,\frac{n}{n+1}, \tag{3.14}$$

where $\mathbf{z}_j$ is observation number j of the d-dimensional vector $\mathbf{Z}$.

For Archimedean copulae, $F_K(w)$ is on the form:

$$F_K(w,\theta)=w+\sum_{i=1}^{d-1}\frac{(-1)^i}{i!}\varphi(w)^i\frac{\mathrm{d}^i}{\mathrm{d}x^i}\varphi^{-1}(x)\bigg|_{x=\varphi(w)}, \tag{3.15}$$

where $\varphi(w)$ is the copula generator function (see e.g. Nelsen (1999) for the definition of the copula generator function) and $\varphi^{-1}(x)$ is the inverse copula generator function.

Since $\mathbf{Z}$ is independent under the null hypothesis, $F_K(w)$ will be on the form:

$$F_K(w)=P(C(\mathbf{Z})\leq w)=w+w\sum_{i=1}^{d-1}\frac{(-1)^i}{i!}\ln(w)^i=1-\frac{\Gamma_l(-\ln(w),d)}{\Gamma(d)}, \tag{3.16}$$

where $\Gamma_l(\cdot)$ is the lower incomplete gamma function, defined as $\Gamma_l(x,a)=\int_0^x t^{a-1}\exp(-t)\mathrm{d}t$. This can be proved by insertion of e.g. the Clayton or the Gumbel generator function into Equation (3.15), with the parameter value corresponding to independence.

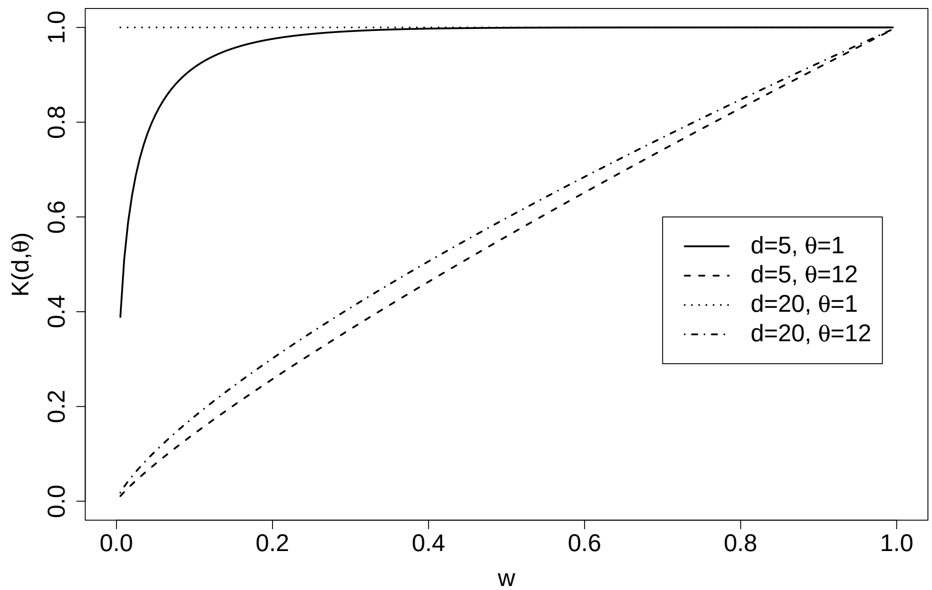


Figure 3.1. $F_K(w)$ for the Gumbel copula.

For large d, the distribution of Equation (3.16) quickly approaches a unit step function at $w = 0$ and makes $F_K(w)$ useless for testing. See Figure 3.1 for $\theta = 1$ for illustration. To counter this problem and obtain a more linear univariate distribution, the PIT data set can be transformed to a distribution with upper tail dependence (to increase the probability of $(U_{j1} \geq u_1, \ldots, U_{jd} \geq u_d)$) and a known $F_{K,0}(w)$. The most apparent choice is the Gumbel copula, since it is on a simple parametric form and is an Archimedean copula with upper tail dependence[2]. Using the transformed data, $F_K(w)$ becomes usable. Figure 3.1 shows how $F_K(w)$ approaches the uniform distribution for increasing θ. Note that $\theta = 1$ corresponds to the independent copula.

Next we need to find an expression for $F_{K,0}$ for a Gumbel dependence structure. $F_{K,Gumbel}(w)$ is found by insertion of $\varphi(w) = (-\ln w)^\theta$ and $\varphi^{-1}(x) = \exp\left(-x^{1/\theta}\right)$ into Equation (3.15). It can be shown (Appendix A.2) that the i'th derivative of $\phi^{-1}(x)$ can be written as:

$$\frac{\mathrm{d}^i \varphi^{-1}(x)}{\mathrm{d}x^i} = \xi_i(x) \frac{\mathrm{d}\varphi^{-1}(x)}{\mathrm{d}x}, \tag{3.17}$$

where

$$\xi_i(x) = \xi_2(x) \cdot \xi_{i-1}(x) + \xi_{i-1}^{(1)}(x), \qquad i \geq 2, \tag{3.18}$$

$$\xi_i^{(j)}(x) = \xi_{i-1}^{(j+1)}(x) + \sum_{m=0}^{j} \binom{j}{m} \xi_2^{(m)}(x) \xi_{i-1}^{(j-m)}(x), \qquad j \geq 0, \tag{3.19}$$

[2]This transformation is, however, not trivial. The popular algorithm for compound constructions of copulae suggested by Marshall and Olkin (1988), for simulating from an Archimedean copula, can not be used since it uses $d+1$ variables to simulate a d-dimensional copula. Here, a rather slow simulation procedure from Embrechts et al. (2003) was applied (see Appendix A.1 for an algorithm). In addition, a general expression for $F_{K,Gumbel}(w)$ is not on a simple form and must be found using a complicated, recursive calculation procedure.

and

$$\xi_2^{(j)}(x) = \frac{(-1)^j}{x^{j+1}} \left(x^{\frac{1}{\theta}} \prod_{m=0}^{j} \left(j - \frac{1}{\theta} \right) - \left(1 - \frac{1}{\theta} \right) j! \right), \qquad j \geq 0. \tag{3.20}$$

Here, superscript (j) indicates the j'th derivative. It can be shown (Appendix A.2) that by computing the first $d-3$ derivatives of $\xi_2(x)$, all $\xi_i,\ i = 2, \ldots, d-1$, can be found.

The issue of $\lim_{w\to 0} f_k(w) \to \infty$ is countered by replacing the approach K with a slightly altered approach M. By applying the inverse function $F_K^{-1}(w)$ to the empirical results $\widehat{F}_K(w)$, a uniform distribution is obtained. Let the function $F_K^{-1}(x)$ be a function that solves $F_K(w) = x$ with respect to w. This function must be evaluated numerically. The approach M can be considered to be a generalization of the approach K and is defined as:

$$F_M(w) = F_K(F_K^{-1}(w)), \quad w \in [0, 1]. \tag{3.21}$$

Under the null hypothesis $F_M(w) = w$ and the density function $f_m(w) = 1$. The empirical version, $\widehat{F}_M(w)$, equals:

$$\widehat{F}_M(w) = \frac{1}{n+1} \sum_{j=1}^{n} I(F_K(\widehat{C}(z_j)) \leq w), \qquad w = \frac{1}{n+1}, \ldots, \frac{n}{n+1}. \tag{3.22}$$

The expression for $\widehat{F}_M(w)$ was derived by noting that

$$\begin{aligned} \widehat{F}_M(w) &= \widehat{F}_K(F_K^{-1}(w)) = \frac{1}{n+1} \sum_{j=1}^{n} I(\widehat{C}(z_j) \leq F_K^{-1}(w)) \\ &= \frac{1}{n+1} \sum_{j=1}^{n} I(F_K(\widehat{C}(z_j)) \leq w). \end{aligned}$$

The approach M is quite slow due to the Gumbel transform and the numerical inversion of F_K. Further, the approach implicitly weight small values, i.e. values close to zero, the left tail. This can be proven as follows. Consider the independent random vector $\mathbf{z}$. The empirical copula of this independent vector is defined as $\widehat{C}(\mathbf{z}) = \prod_{i=1}^{d} z_i$. Suppose we add a small perturbation, Δ, to an arbitrary value z_k:

$$\widehat{C}(\Delta\mathbf{z}) = z_1 \cdots (z_k + \Delta) \cdots z_d = \widehat{C}(\mathbf{z}) + \Delta \prod_{i=1, i\neq k}^{d} z_i = \widehat{C}(\mathbf{z}) + \Delta \frac{\widehat{C}(\mathbf{z})}{z_k}. \tag{3.23}$$

Now the ratio $\widehat{C}(\Delta\mathbf{z})/\widehat{C}(\mathbf{z}) = 1 + \Delta/z_k$ is a decreasing function of z_k. Hence, the approaches K and M have a left tail bias. What this means is that a deviation from $\mathcal{H}_0$ will have a greater effect on the approach K/M for small values close to zero than for high values close to one. However, this bias can be reduced by increasing the upper tail dependency, i.e. by increasing θ in the infliction of a Gumbel dependency structure. As $\theta \longrightarrow \infty$ we obtain perfect dependence and the perturbation Δ will have no effect.

3.2. Full Multivariate Approach

To supplement the three dimension reduction approaches above, we also consider the full multivariate approach proposed by Panchenko (2005). The dimension reduction approaches is a two stage approach, first the problem is reduced to a univariate problem, second a univariate test statistic is applied. In contrast, the full multivariate approach is a test in itself, testing the entire data set in one step. Thus any further unification of the framework, beyond the use of PIT data, is not sensible.

Panchenko (2005) proposes an approach based on positive definite bilinear forms. Let f_1 and f_2 be two integrable functions and define the bilinear form as

$$\langle f_1|\kappa_d|f_2\rangle = \int\int \kappa_d(\mathbf{x}_1,\mathbf{x}_2) f_1(\mathbf{x}_1) f_2(\mathbf{x}_2) \mathrm{d}\mathbf{x}_1 \mathrm{d}\mathbf{x}_2, \tag{3.24}$$

where $\mathbf{x}_1,\ \mathbf{x}_2 \in \mathbb{R}^d$ are two random vectors and $\kappa_d(\cdot,\cdot)$ is a positive definite symmetric kernel such as the Gaussian kernel:

$$\kappa_d(\mathbf{x}_1,\mathbf{x}_2) = \exp\left\{-\|\mathbf{x}_1 - \mathbf{x}_2\|^2/(2dh^2)\right\}, \tag{3.25}$$

where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^d$ and $h > 0$ is a bandwidth. The Gaussian kernel is chosen for convenience. In general, other positive definite kernels may be chosen.

Next, define the squared distance Q between f_1 and f_2 as

$$Q = \langle f_1 - f_2|\kappa_d|f_1 - f_2\rangle\,. \tag{3.26}$$

It follows from a theorem (Diks and Tong, 1999) that Q becomes zero only when $f_1(\cdot)$ and $f_2(\cdot)$ are equal. In fact, $\langle f_1|\kappa_d|f_2\rangle$ is an inner product of f_1 and f_2, which can be used as a measure of distance between f_1 and f_2. Using the properties of an inner product, Q can be decomposed as follows

$$Q = Q_{11} - 2Q_{12} + Q_{22}, \tag{3.27}$$

where $Q_{kp} = \langle f_k|\kappa_d|f_p\rangle$. Panchenko (2005) estimates each term of the above decomposition using V-statistics (see Denker and Keller (1983) for an introduction to U- and V-statistics):

$$\widehat{Q}_{kp} = \frac{1}{n^2}\sum_{j_k=1}^{n}\sum_{j_p=1}^{n} \kappa_d(\mathbf{X}_k^{j_k}, \mathbf{X}_p^{j_p}), \tag{3.28}$$

where $\mathbf{X}_k^j$ denotes observation number j of the random vector $\mathbf{X}_k$.

This approach is quite slow since it loops through all $n \times d$ observations. Further, it does not weight any region of the copula. It is a simple approach in the sense that one does not have to bother with which univariate statistic to use. This can however also be regarded a lack of flexibility and thus a drawback.

3.3. Summary

We have examined four copula GOF approaches, Breymann et al. (2003), Berg and Bakken (2005) and Genest et al. (2006), all dimension reduction approaches, and finally the full multivariate approach by Panchenko (2005). We denote the approaches G, B, K and Q, respectively. We altered the approach K to fit it into our framework and to generalize, introducing the approach M.

The approaches solve the dimension reduction in different ways and thus the deviances from $\mathcal{H}_0$ are computed differently. When it comes to numerical efficiency there is no doubt that the approach G is by far the quickest approach. The approach B is somewhat slower while approaches M and Q are quite slow compared to the first two. The approaches also weight different regions of the copula. The approach G implicitly weights the tails of the copula, the approach B allows any weighting of any region of the copula while the approach K or M weights the left tail of the copula. The weighting is incorporated in the dimension reduction technique and is thus independent of which univariate test statistic we use. The full multivariate approach Q does not allow for any weighting.

Breymann et al. (2003) chose to neglect the issue of empirical margins mentioned in Section 3 and applied a straightforward version of the AD statistic. However, preliminary tests we have undertaken shows significant differences in the p-values obtained from their approach and through simulation. Chen et al. (2004) derives an asymptotic expression for their L2 norm KDE statistic. However, we are highly uncertain about the validity of this result. Simulation experiments do not confirm their asymptotic distribution and an error was found in the proof of the mean of the distribution.

We simulate the distribution of the test statistics by repeatedly looping through the entire testing procedure under $\mathcal{H}_0$. This ensures that we use the correct distribution when computing p-values. We simulate $N = 10000$ times. The minimum number of simulations required to obtain a certain confidence level can be deduced using well-known techniques.

The testing procedure is the same for the three dimension reduction approaches, only differing in the dimension reduction technique, i.e. in the calculation of $F_T(w)$, where T is G, B or M. For the full multivariate approach the testing procedure is somewhat different.

Algorithm 3.1 Testing procedure for the dimension reduction approaches

1: Construct the copula $\mathbf{Z}$ by applying the PIT to the observed data set $\mathbf{X}$, given a $\mathcal{H}_0$ copula.
2: Compute $\widehat{F}_T(w)$ according to Equations (3.4), (3.9) or (3.22) respectively.
3: Compute some univariate test statistic $\widehat{\mathcal{T}}$ using $\widehat{F}_T(w)$ according to Equations (2.8)-(2.16).
4: Repeatedly (N times) perform steps 1-4, using a simulated observed data set $\mathbf{X}^*$, simulated from the $\mathcal{H}_0$ distribution. The resulting N values of $\widehat{\mathcal{T}}^*$ form the distribution of $\mathcal{T}$.
5: Compute the p-value, $p = \frac{1+\sum_{i=1}^{N} I(\widehat{\mathcal{T}}_i^* \geq \widehat{\mathcal{T}})}{N+1}$.

Algorithm 3.2 Testing procedure for the full multivariate approach proposed by Panchenko (2005).

1: Construct the copula $\mathbf{Z}$ by applying the PIT to the observed data set $\mathbf{X}$, given a $\mathcal{H}_0$ copula.
2: Compute $\widehat{Q}$ according to Equations (3.26)-(3.28).
3: Repeatedly (N times) perform steps 1-2, using a simulated observed data set $\mathbf{X}^*$, simulated from the $\mathcal{H}_0$ distribution, with parameters estimated from the original observed data set $\mathbf{X}$. The resulting N values of $\widehat{Q}^*$ form the distribution of Q.
4: Compute the p-value, $p = \frac{1+\sum_{i=1}^{N} I(\widehat{Q}_i^* \geq \widehat{Q})}{N+1}$.

4. Results

We assess the statistical power of the approaches G, B, M and Q by performing mixing tests. The approach K in its original form is not included since we are not able to derive an expression for $F_{K,0}(w)$ for the Gaussian copula. The mixing tests give us an impression of the approaches' ability to detect tail heaviness and skewness properties. The ability to distinguish the Gaussian from the Student's t copula indicates the power at detecting tail heaviness, while the ability to distinguish the Gaussian from the Clayton- and survival Clayton copula indicates the power at detecting skewness. The mixing tests are performed by mixing a Gaussian copula ($\rho = 0.5$) with an alternative copula to construct a mixed copula C^{Mix}:

$$C^{Mix} = (1-\beta) \cdot C^{Ga} + \beta \cdot C^{Alt}, \quad \beta \in [0,1],$$

where β is the mixing parameter, C^{Ga} denotes the Gaussian copula and C^{Alt} denotes the alternative copula. In this paper the alternative copulae considered are:

- C^{St}: the Student's t copula ($\rho = 0.5,\ \nu = 4$),
- C^{Cl}: the one-parameter Clayton copula ($\delta = 1.0$),
- C^{sCl}: the one-parameter survival Clayton copula ($\delta = 1.0$).

For $\beta = 0$, C^{Mix} is a Gaussian copula while for $\beta = 1$, C^{Mix} is the alternative copula. For $0 < \beta < 1$ we sample from the Gaussian copula with probability $(1-\beta)$ and from the alternative copula with probability β. Our null hypothesis is that the mixed copula is a Gaussian copula. We PIT C^{Mix} under this null hypothesis and compute $\widehat{\mathcal{T}}$ and the corresponding p-value. This is repeated 500 times in order to obtain rejection rates and corresponding power curves. This relatively low number of repetitions is why the power curves in Figures 4.1, 4.2 and 4.3 are not smoother.

First, we examined the effect of dimension and number of observations. For the approach B, the power of the approach increases with number of observations n and with dimension d, as expected. n has the same effect on all four approaches and all univariate test statistics. Next, also as expected, d has the same effect for all approaches, most prominent for G and B. For M and Q this effect is not so clear.

Next, we examined the various univariate test statistics. The KSAD statistic performed remarkably well in several cases. However, when examining all approaches and combinations of n, d and C^{Alt}, the KSAD statistic seems quite unstable in the sense that the rejection rate in some cases does not increase smoothly with β and in some few cases it performs quite poor. The AD statistic also performs well, in all cases. In addition, the AD statistic seems more stable than the KSAD statistic. Thus, the AD statistic seems to be the best performing statistic for our use. Our beliefs in Section 2.4 are thus confirmed. We only consider the AD statistic in the remains of the paper.

Finally, we turn to the performance of the various copula GOF approaches. We examine the power, using the AD statistic, for the approaches G, B and M. The results are displayed in Figures 4.1, 4.2 and 4.3. We see that it varies with n, d and C^{Alt}, which approach has the best power. However, some general conclusions can be made:

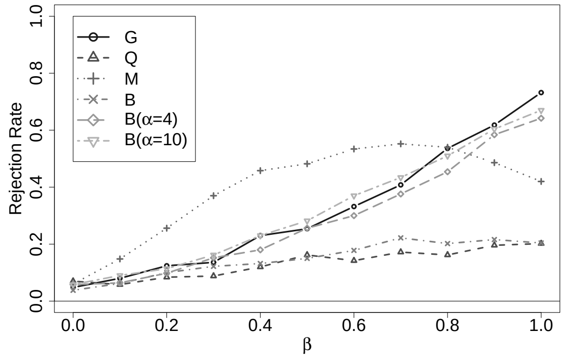
- For $C^{Alt} = C^{St}$ the approach G performs very well. We believe this is due to its implicit tail weighting. Similarly, we see that the approach B with power tail weight performs very well, increasingly well as the tail weight is increased.
- For $C^{Alt} = C^{Cl}$ and $C^{Alt} = C^{sCl}$ the approaches M and Q perform very well for $d = 2$. For $d > 2$ they do not perform that well while B performs very well. G does not perform so well for these alternative copulae, this is due to the inconsistency of this approach.
- For $C^{Alt} = C^{sCl}$ and $d = 5$ and 10, the approach M performs very poor. This is probably due to the implicit left tail weighting. As we know, the survival Clayton copula has upper tail dependence.
- For some combinations of n, d and C^{Alt}, mainly for low values of n, and at some β levels, the rejection rate for the approach M decreases as β is increased. So it seems like the approach M suffers more from few observations than the other approaches. We have no explanation for this phenomenon as of now.
- For $d = 5$ and $d = 10$ the approach B outperforms the other tests.

5. Conclusion

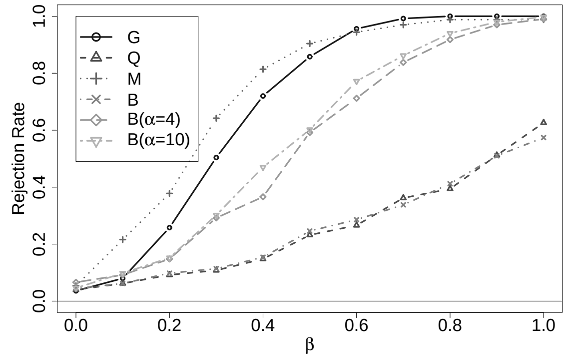
We have compared the copula GOF approaches by Breymann et al. (2003), Berg and Bakken (2005), Genest et al. (2006) and Panchenko (2005). All four are fitted into a PIT framework and some generalizations and improvements are made. Three of the approaches project the multivariate problem to a univariate problem, and then apply a univariate test. We have examined the most popular univariate tests. Further we have examined the power of the GOF approaches at distinguishing tail heaviness and skewness properties.

The best performing approach varies with the alternative copula in the mixing process, as well as the dimension and the number of observations. The results are similar for all univariate statistics. Some concluding remarks:

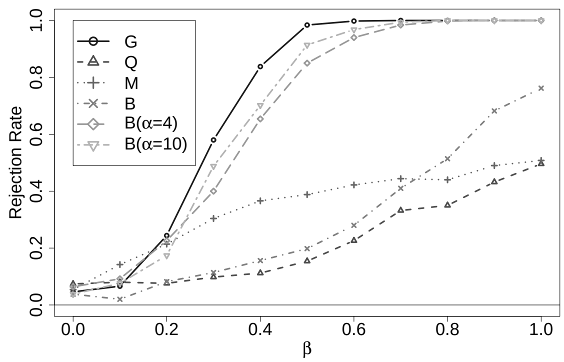
- G: Tail weight. Quick to compute and very well suited for tail heaviness testing. Performs rather poor for skewness testing. Not consistent.
- M: Left tail weight. Performs very well in the bivariate case with a high number of observations, not so well for higher dimensions and seems unstable for few observations. Performs very poor at detecting upper tail dependence for higher dimensions. Computationally demanding in our framework.



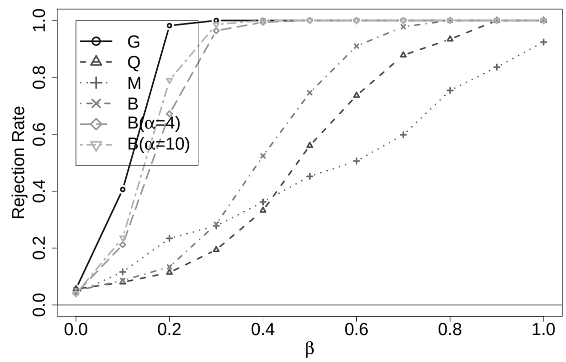
(a) $C^{Alt} = C^{St}$, $d = 2$, $n = 500$.



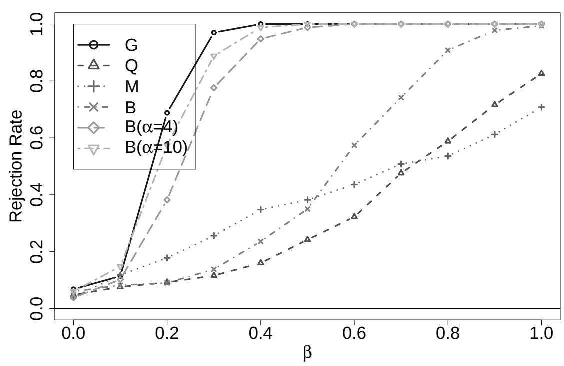
(b) $C^{Alt} = C^{St}$, $d = 2$, $n = 2500$.



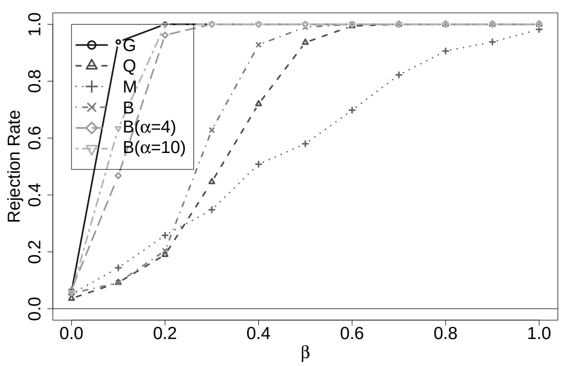
(c) $C^{Alt} = C^{St}$, $d = 5$, $n = 500$.



(d) $C^{Alt} = C^{St}$, $d = 5$, $n = 2500$.

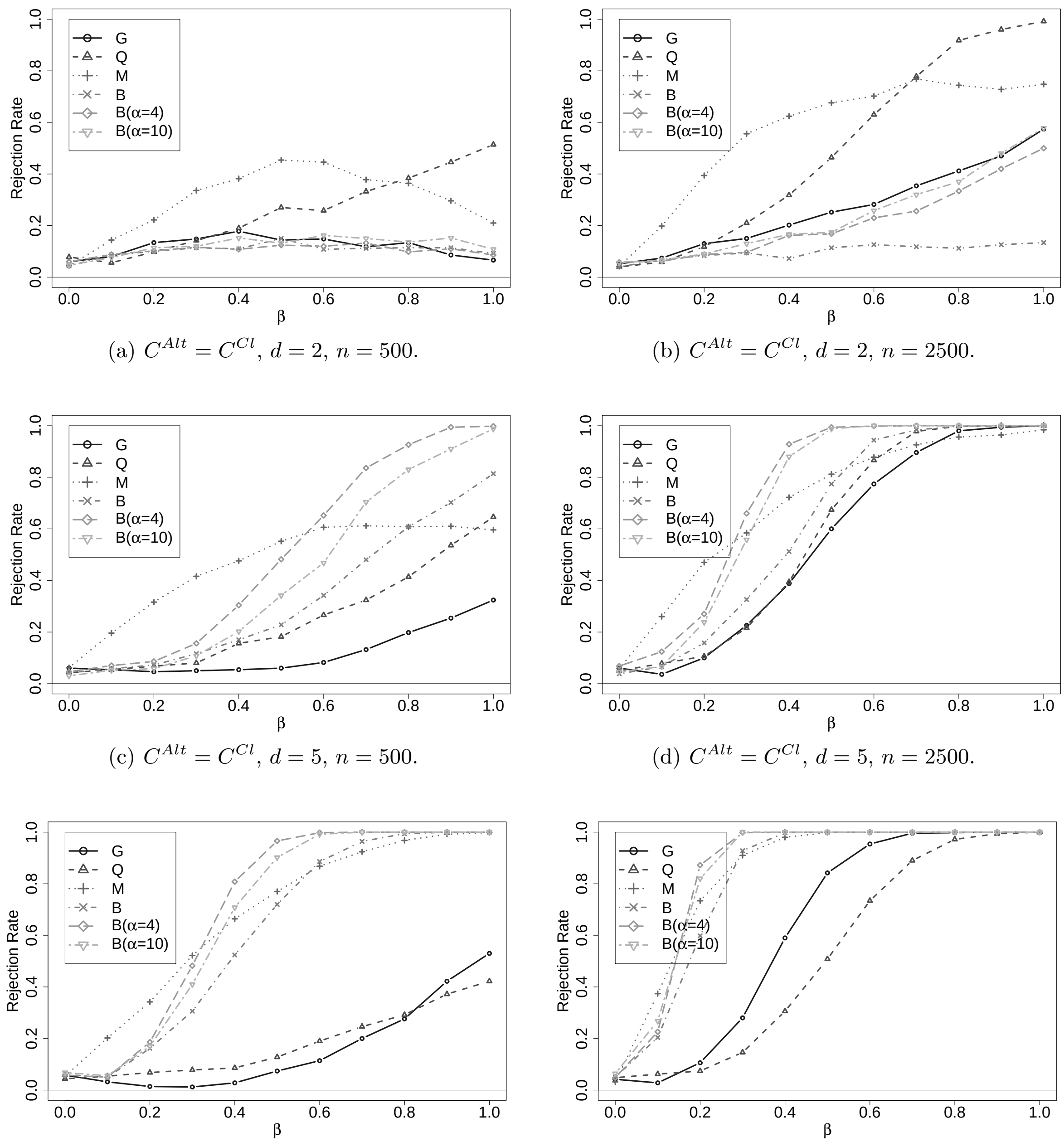


(e) $C^{Alt} = C^{St}$, $d = 10$, $n = 500$.



(f) $C^{Alt} = C^{St}$, $d = 10$, $n = 2500$.

Figure 4.1. Power curves for all copula GOF approaches, $d = (2, 5, 10)$, $n = (500, 2500)$, $C^{Alt} = C^{St}$. For G, B and M the AD statistic is used. On the x-axis we see the mixing parameter β, while on the y-axis we see the portion of times the Gaussian copula (i.e. the $\mathcal{H}_0$ copula) is rejected, with a 5% significance level.



(a) $C^{Alt} = C^{Cl}$, $d = 2$, $n = 500$.

(b) $C^{Alt} = C^{Cl}$, $d = 2$, $n = 2500$.

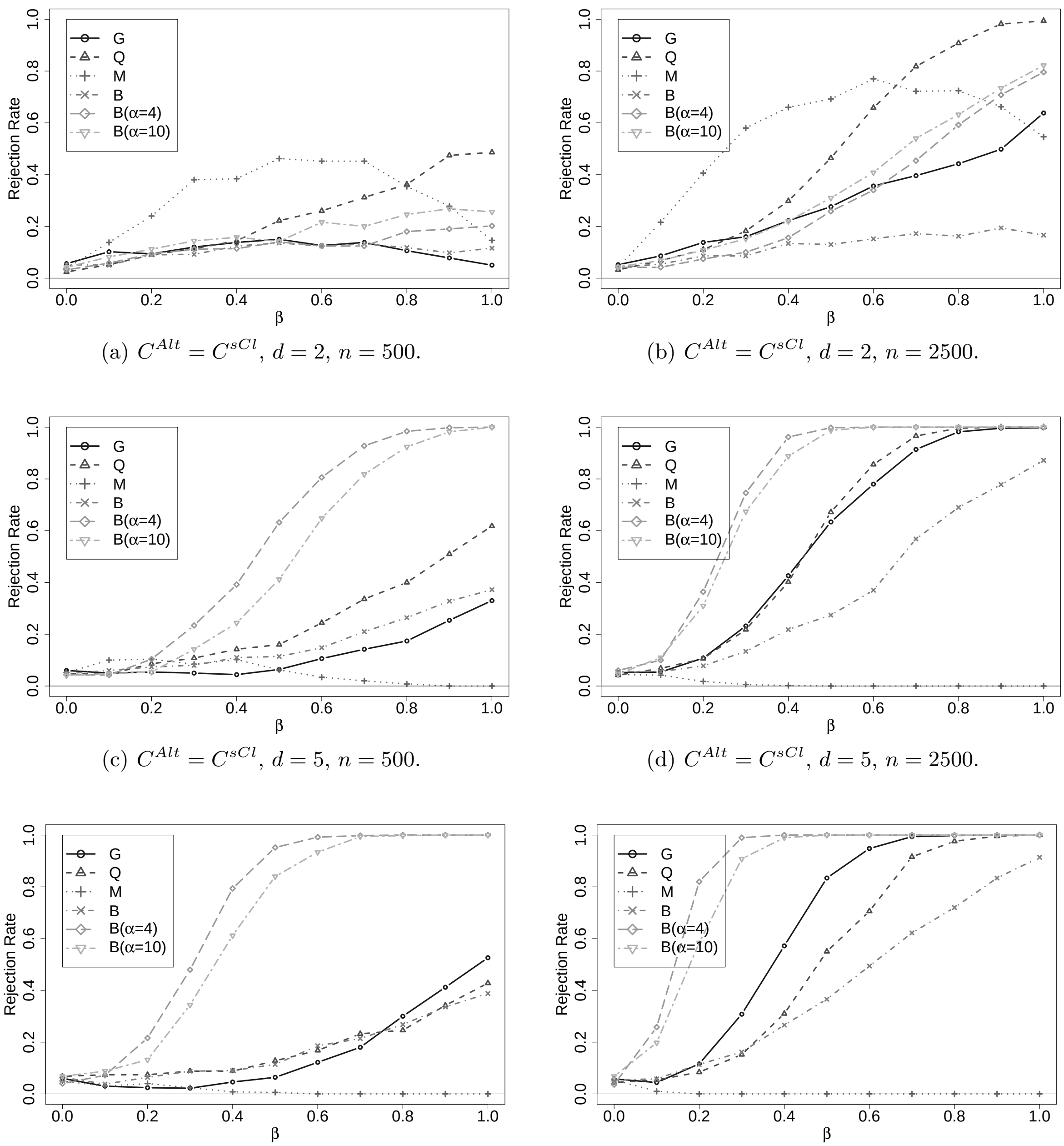
(c) $C^{Alt} = C^{Cl}$, $d = 5$, $n = 500$.

(d) $C^{Alt} = C^{Cl}$, $d = 5$, $n = 2500$.

(e) $C^{Alt} = C^{Cl}$, $d = 10$, $n = 500$.

(f) $C^{Alt} = C^{Cl}$, $d = 10$, $n = 2500$.

Figure 4.2. Power curves for all copula GOF approaches, $d = (2, 5, 10)$, $n = (500, 2500)$, $C^{Alt} = C^{Cl}$. For G, B and M the AD statistic is used. On the x-axis we see the mixing parameter β, while on the y-axis we see the portion of times the Gaussian copula (i.e. the $\mathcal{H}_0$ copula) is rejected, with a 5% significance level.



(a) $C^{Alt} = C^{sCl}$, $d = 2$, $n = 500$.

(b) $C^{Alt} = C^{sCl}$, $d = 2$, $n = 2500$.

(c) $C^{Alt} = C^{sCl}$, $d = 5$, $n = 500$.

(d) $C^{Alt} = C^{sCl}$, $d = 5$, $n = 2500$.

(e) $C^{Alt} = C^{sCl}$, $d = 10$, $n = 500$.

(f) $C^{Alt} = C^{sCl}$, $d = 10$, $n = 2500$.

Figure 4.3. Power curves for all copula GOF approaches, $d = (2, 5, 10)$, $n = (500, 2500)$, $C^{Alt} = C^{sCl}$. For G, B and M the AD statistic is used. On the x-axis we see the mixing parameter β, while on the y-axis we see the portion of times the Gaussian copula (i.e. the $\mathcal{H}_0$ copula) is rejected, with a 5% significance level.

- Q: No weight. Performs very well for skewness testing in the bivariate case. Not so strong for higher dimensions or for tail heaviness testing. A simple test in the sense that we do not have to bother with which univariate statistic to use. However, this can also be considered a lack of flexibility. No weighting flexibility, and the test is computationally demanding.

- B: Any weight. Adds valuable weighting flexibility, and with the appropriate weight, both tail heaviness and skewness performance is very good. Superior for higher dimensions. The approach is computationally more demanding than the approach G, but less demanding than the approaches M and Q.

Based on the above we give the following recommendations. Note that the results and recommendations are based on the particular mixing tests that we performed and should thus be interpreted as indications.

- If you have no prior opinion on which part of the copula that may deviate from $\mathcal{H}_0$, then use Q for bivariate problems and B with no weighting for higher dimensions.

- If you wish to emphasize the tails of the copula, use G or B with power tail weight.

- If you wish to emphasize other regions, i.e. one of the tails only, use Q in bivariate problems and B for higher dimensions.

- Tests based on binning or smoothing are also of interest in the bivariate case.

Further work should be done to better understand the approach M and its properties. Also, work remains in understanding the effect of weighting various regions of the copula. How does the fact that M weights the left tail influence its power for various alternative copulae? And plenty of work remains testing various weight functions for B and their influence. Finally, does the order in which we PIT the data significantly effect the p-values?

Acknowledgements

The authors acknowledge the support and guidance of colleagues at the Norwegian Computing Center, in particular Assistant Research Director Kjersti Aas and Chief Research Scientist Xeni Kristine Dimakos. We would especially like to express our gratitude to Kjersti Aas for introducing us to the field of copulae and copula goodness-of-fit testing. Most of this work was carried out while Henrik Bakken was a MSc. student at the Norwegian University of Science and Technology, working on his diploma thesis at the Norwegian Computing Center. Henrik is now a consultant at Boston Consulting Group, Oslo office, Norway. Daniel Berg is a PhD student at the Norwegian Computing Center and the University of Oslo.

References

Anderson, T. W. and D. A. Darling (1954). A test of goodness of fit. *Journal of the American Statistical Association 49*, 765–769.

Aslan, B. and G. Zech (2002). Comparison of different goodness-of-fit tests. In *Advanced Statistical Techniques in Particle Physics*. Grey College, Durham.

Azzalini, A. and A. Bowman (1997). *Applied Smoothing Techniques for Data Analysis*. New York: Oxford University Press.

Barbe, P., C. Genest, K. Ghoudi, and B. Rémillard (1996). On Kendall's process. *Journal of Multivariate Analysis 58*, 197–229.

Berg, D. and H. Bakken (2005, December). A goodness-of-fit test for copulae based on the probability integral transform. Note, Norwegian Computing Centre, SAMBA/41/05.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business and Economic Statistics 19*, 465–474.

Breymann, W., A. Dias, and P. Embrechts (2003). Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance 1*, 1–14.

Chen, X., Y. Fan, and A. Patton (2004). Simple tests for models of dependence between multiple financial time series, with applications to U.S. equity returns and exchange rates. Financial Markets Group, London School of Economics, Discussion Paper 483. Revised July 2004.

D'Agostino, R. B. and M. A. Stephens (1986). *Goodness-of-Fit Techniques*. New York: Marcel Dekker Inc.

Denker, M. and G. Keller (1983). On u-statistics and v. mises' statistics for weakly dependent processes. *Zeitschrift fuer Wahrscheinlichkeitstheorie und Verwandte Gebiete 64*, 505–522.

Diebold, F. X., T. Gunther, and A. S. Tay (1998). Evaluating density forecasts, with applications to financial risk management. *International Economic Review 39*, 863–883.

Diebold, F. X., J. Hahn, and A. S. Tay (1999). Multivariate density forecast and calibration in financial risk management: High-frequency returns on foreign exchange. *The Review of Economics and Statistics 81(4)*, 661–673.

Diks, C. and H. Tong (1999). A test for symmetry of multivariate probability distributions. *Biometrika 86 (3)*, 605–614.

Dobrić, J. and F. Schmid (2005). Testing goodness of fit for parametric families of copulas - application to financial data. *Communications in Statistics: Simulation and Computation 34*.

Embrechts, P., F. Lindskog, and A. McNeil (2003). *Modelling dependence with copulas and applications to risk management. In Handbook of heavy tailed distributions in finance*. Elsevier/North-Holland, Amsterdam.

Embrechts, P., A. McNeil, and D. Straumann (1999, May). Correlation and dependence in risk management: Properties and pitfalls. *RISK*, 69–71.

Farebrother, R. W. (1990). The distribution of a quadratic form in normal variables. *Applied Statistics 39*, 294–309.

Fermanian, J. (2003). Goodness of fit tests for copulas. Working Paper CREST 2003-34. To appear in Journal of Multivariate Analysis.

Genest, C., K. Ghoudi, and L. Rivest (1995). A semi-parametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika 82*, 543–552.

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

Genest, C. and L.-P. Rivest (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association*, 1034–1043.

Hong, Y. (2000). Evaluation of out-of-sample density forecasts with applications to s&p 500 stock prices. Working Paper, Department of Economics and Department of Statistical Science, Cornell University.

Hong, Y. and H. Li (2002). Nonparametric specification testing for continuous-time models with application to spot interest rates. Working Paper, Department of Economics, Cornell University.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Malevergne, Y. and D. Sornette (2003). Testing the gaussian copula hypothesis for financial assets dependence. *Quantitative Finance 3*, 231–250.

Marshall, A. and I. Olkin (1988). Families of multivariate distributions. *Journal of the American Statistical Association 83*, 834–841.

Nelsen, R. B. (1999). *An Introduction to Copulas.* New York: Springer Verlag.

Panchenko, V. (2005). Goodness-of-fit test for copulas. *Physica A 355(1)*, 176–182.

Peterson, L. (2004). Kernel density estimation. Lecture Notes, Baylor College of Medicine, Houston, Texas.

Pollard, D. (1979). General chi-square goodness-of-fit tests with data-dependent cells. *Z. Wahrsch. verw. Gebiete 50*, 317–331.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics 23*, 470–472.

Sklar, A. (1959). Fonctions dé repartition á n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris 8*, 299–231.

Snedecor, G. and W. Cochran (1986). *Statistical Methods*, Volume 8. Iowa State University Press.

Thompson, S. B. (2002). Evaluating the goodness of fit of conditional distributions, with an application to affine term structure models. Working paper, Department of Economics, Harvard University.

A. Gumbel K

In order to fit the approach by Genest et al. (2006) into the PIT framework, as done in Section 3.1.3, we need to impose a dependency structure on the PIT data, preferably one with upper tail dependence. The Gumbel copula is a natural choice for this purpose. In order to do this successfully we need an approach for sequential sampling from the Gumbel copula, e.g. the approach described in Embrechts et al. (2003, Section 6). When we have transformed the data into a data set with Gumbel dependency structure we need to compute $F_K(w)$ for this data. I.e. we need to derive an expression for $F_{K,Gumbel}(w)$.

A.1. Gumbel Transformation

To convert the PIT data set $\mathbf{Z}$, into a data set $\mathbf{U}$, with Gumbel dependency structure, the appropriate algorithm can be derived from Embrechts et al. (2003, Algorithm 6.1, Section 6.5), using $C(u_1, u_2, u_3) = C(C(u_1, u_2), u_3)$ and $C(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2))$, where $\phi(w) = (-\ln w)^\theta$ is the Gumbel copula generator function. We also utilize the function $F_{K,C}(w) = w - \phi(w)/\phi'(w)$, which is equal to Equation (3.15) for $d = 2$. The inverse, $F^{-1}_{K,C}$ must be found numerically, e.g. by bisection.

Algorithm A.1 Gumbel transformation

for $j \leftarrow 1,\ n$ **do**
    $t = F^{-1}_{K,C}(Z_{j,1})$
    for $i \leftarrow 2,\ d$ **do**
        $s = Z_{j,i}$
        $a = \varphi^{-1}(s\varphi(t))$
        $U_{j,i-1} = \varphi^{-1}((1-s)\varphi(t))$
        $t = F^{-1}_{K,C}(a)$
    end for
    $U_{j,d} = a$
end for

A.2. $F_{K,Gumbel}$

The expression for $F_{K,Gumbel}$ basically involves deriving an expression for $\frac{\mathrm{d}^i}{\mathrm{d}x^i}\varphi^{-1}(x)$. Remember that for the Gumbel copula $\varphi(w) = (-\log w)^\theta$ and $\varphi^{-1}(x) = \exp\left(-x^{1/\theta}\right)$.

We start with Equations (3.17) and (3.18):

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\varphi^{-1}(x) &= -\frac{1}{\theta}x^{1/\theta-1}\cdot\varphi^{-1}(x),\\
\frac{\mathrm{d}^2}{\mathrm{d}x^2}\varphi^{-1}(x) &= \left[\left(\frac{1}{\theta}-1\right)x^{-1} - \frac{1}{\theta}x^{1/\theta-1}\right]\cdot\frac{\mathrm{d}}{\mathrm{d}x}\varphi^{-1}(x) = \xi_2(x)\cdot\frac{\mathrm{d}}{\mathrm{d}x}\varphi^{-1}(x)\\
\frac{\mathrm{d}^3}{\mathrm{d}x^3}\varphi^{-1}(x) &= \left[\xi_2(x)\xi_2(x) + \xi_2^{(1)}(x)\right]\cdot\frac{\mathrm{d}}{\mathrm{d}x}\varphi^{-1}(x) = \xi_3(x)\cdot\frac{\mathrm{d}}{\mathrm{d}x}\varphi^{-1}(x)\\
&\;\;\vdots\\
\frac{\mathrm{d}^i}{\mathrm{d}x^i}\varphi^{-1}(x) &= \xi_i(x)\frac{\mathrm{d}}{\mathrm{d}x}\varphi^{-1}(x),
\end{aligned}
$$

where

$$
\xi_i(x) = \xi_2(x)\xi_{i-1}(x) + \xi_{i-1}^{(1)}(x).
$$

Let us continue by showing Equation (3.19):

$$\begin{aligned}
\xi_i^{(1)}(x) =& \xi_2^{(1)}(x)\xi_{i-1}(x) + \xi_2(x)\xi_{i-1}^{(1)}(x) + \xi_{i-1}^{(2)}(x) \\
\xi_i^{(2)}(x) =& \xi_2^{(2)}(x)\xi_{i-1}(x) + 2\xi_2^{(1)}(x)\xi_{i-1}^{(1)}(x) + \xi_2(x)\xi_{i-1}^{(2)}(x) + \xi_{i-1}^{(3)}(x) \\
\xi_i^{(3)}(x) =& \xi_2^{(3)}(x)\xi_{i-1}(x) + 3\xi_2^{(2)}(x)\xi_{i-1}^{(1)}(x) + 3\xi_2^{(1)}(x)\xi_{i-1}^{(2)}(x) + \xi_2(x)\xi_{i-1}^{(3)}(x) + \xi_{i-1}^{(4)}(x) \\
&\vdots \\
\xi_i^{(j)}(x) =& \xi_{i-1}^{(j+1)}(x) + \sum_{m=0}^{j} \tbinom{j}{m} \xi_2^{(m)}(x)\xi_{i-1}^{(j-m)}(x).
\end{aligned}$$

Next, we show Equation (3.20):

$$\begin{aligned}
\xi_2(x) =& x^{-1}\left[\frac{1}{\theta}\left(1 - x^{\frac{1}{\theta}}\right) - 1\right] \\
\xi_2^{(1)}(x) =& \frac{\mathrm{d}}{\mathrm{d}x}\xi_2(x) = (-1)^2\left(1 - \frac{1}{\theta}\right)x^{-1} + (-1)^2 \cdot \frac{1}{\theta} \cdot \left(1 - \frac{1}{\theta}\right) \cdot x^{\frac{1}{\theta}-2} \\
\xi_2^{(2)}(x) =& (-1)^3 \cdot 2 \cdot \left(1 - \frac{1}{\theta}\right) \cdot x^{-3} + (-1)^3 \cdot \frac{1}{\theta} \cdot \left(1 - \frac{1}{\theta}\right)\left(2 - \frac{1}{\theta}\right) \cdot x^{\frac{1}{\theta}-3} \\
&\vdots \\
\xi_2^{(j)}(x) =& \frac{(-1)^j}{x^{j+1}}\left[x^{\frac{1}{\theta}} \prod_{m=0}^{j}\left(m - \frac{1}{\theta}\right) - \left(1 - \frac{1}{\theta}\right) \cdot j!\right].
\end{aligned}$$

Now, since

$$F_K(w) = w + \sum_{i=1}^{d-1} \frac{(-1)^i}{i!}\varphi(w)^i \frac{\mathrm{d}^i}{\mathrm{d}x^i}\varphi^{-1}(x)\bigg|_{x=\varphi(w)},$$

we only need to compute the $d-1$ first derivatives of $\varphi^{-1}(x)$, which mainly consists of computing $\xi_{d-1}(x)$, according to Equation (3.17). Before we compute $\xi_{d-1}(x)$ we derive an alternative expression for $\xi_i(x)$:

$$\begin{aligned}
\xi_i(x) =& \xi_2(x)\xi_{i-1}(x) + \xi_{i-1}^{(1)}(x) \\
=& \xi_2(x)\left(\xi_2(x)\xi_{i-1}(x) + \xi_{i-2}^{(1)}\right) + \xi_{i-1}^{(1)} \\
=& \xi_2^3(x)\xi_{i-3}(x) + \xi_2^2(x)\xi_{i-3}^{(1)}(x) + \xi_2(x)\xi_{i-2}^{(1)}(x) + \xi_{i-1}^{(1)}(x) \\
&\vdots \\
=& \xi_2^{i-1}(x) + \sum_{m=1}^{i-2} \xi_2^{m-1}(x)\xi_{i-m}^{(1)}(x). \qquad \text{(A.1)}
\end{aligned}$$

Using Equation (A.1) we have that $\xi_{d-1}(x) = \xi_2^{d-2}(x) + \sum_{m=1}^{d-3} \xi_2^{m-1}(x)\xi_{d-1-m}^{(1)}(x)$. The only part involving the derivatives of $\xi_2(x)$ is $\xi_{d-1-m}^{(1)}(x)$. For $m = 1$ we have $\xi_{d-2}^{(1)}(x)$ which will be the part involving the most derivations of $\xi_2(x)$:

$$\begin{aligned}
\xi_{d-2}^{(1)}(x) =& \xi_{d-3}^{(2)}(x) + \ldots \\
=& \xi_{d-4}^{(3)}(x) + \ldots \\
&\vdots \\
=& \xi_{d-(d-2)}^{(d-3)}(x) = \xi_2^{(d-3)}(x) + \ldots
\end{aligned}$$

I.e. we only have to compute the first $d-3$ derivatives of $\xi_2(x)$ in order to compute all $\xi_i,\ i=2,\dots,d-1$.

We encounter numerical problems for large d. When x becomes very small, $\xi_2^{(j)}(x)$, and thus $\xi_i(x)$, becomes extremely large. This problem is countered by observing that $\xi_i^{(j)}(x)$ is proportional to x^{1-j-i}. This can be showed by induction. We start with $i=3$, showing that $\xi_3^{(j)}(x) \propto x^{-j-2}$:

$$\xi_3^{(j)}(x) = \xi_2^{(j+1)}(x) + \sum_{m=0}^{j} \binom{j}{m} \xi_2^{(m)}(x)\xi_2^{(j-m)}(x),$$

where

$$\xi_2^{(j+1)}(x) = \frac{(-1)^{j+1}}{x^{j+2}} \left[x^{\frac{1}{\theta}} \prod_{m=0}^{j+1} \left(m - \frac{1}{\theta} \right) - \left(1 - \frac{1}{\theta} \right)(j+1)! \right].$$

Hence $\xi_2^{(j+1)}(x) \propto x^{-j-2}$. Further, $\xi_2^{(m)}(x) \propto x^{-m-1}$ and $\xi_2^{(j-m)}(x) \propto x^{-j+m-1}$. Hence $\xi_2^{(m)}(x)\xi_2^{(j-m)}(x) \propto x^{-j-2}$, and we have showed that $\xi_3^{(j)} \propto x^{-j-2}$. Next we show that $\xi_i^{(j)}(x) \propto x^{1-j-i}$ implies that $\xi_{i+1}^{(j)}(x) \propto x^{-j-i}$:

$$\xi_{i+1}^{(j)}(x) = \xi_i^{(j+1)}(x) + \sum_{m=0}^{j} \binom{j}{m} \xi_2^{(m)}(x)\xi_i^{(j-m)}(x),$$

where $\xi_2^{(m)}(x) \propto x^{-m-1}$, and $\xi_i^{(j-m)}(x) \propto x^{1-j+m-i}$. Hence $\xi_2^{(m)}(x)\xi_i^{(j-m)}(x) \propto x^{-j-i}$, and since $\xi_i^{(j)}(x) \propto x^{1-j-i}$, we have showed that $\xi_{i+1}^{(j)}(x) \propto x^{-j-i}$.

We have now showed that $\xi_i^{(j)}(x)$ is proportional to x^{1-j-i}. This makes $\xi_i(x)$ proportional to x^{1-i}, which is shown using Equation (A.1):

$$\xi_i(x) = \xi_2^{i-1}(x) + \sum_{m=1}^{i-2} \xi_2^{m-1}(x)\xi_{i-m}^{(1)}(x),$$

where $\xi_2^{i-1}(x) \propto x^{1-i}$, $\xi_2^{m-1}(x) \propto x^{1-m}$ and $\xi_{i-m}^{(1)}(x) \propto x^{m-i}$. Hence $\xi_2^{m-1}(x)\xi_{i-m}^{(1)}(x) \propto x^{1-i}$ and $\xi_i(x) \propto x^{1-i}$. Now, since $\xi_i(x)$ is proportional to x^{1-i}, $\frac{\mathrm{d}}{\mathrm{d}x}\varphi^{-1}(x)$ is also proportional to x^{1-i}. In Equation (3.15), $x=\varphi(w)$ and the proportionality factor x^{1-i} cancels out with $\varphi(w)^i$ and becomes $\varphi(w)$.

B. Empirical Univariate GOF Test Statistics

We derive the expressions for the empirical statistics treated in Section 2.4.

The proof of the KS statistic in Equation (2.12) is simple. Since the cdf $\widehat{F}(z)$ is a discrete step function and $F(z)$ is a strictly increasing function, the optimal z is restricted to $F(j/(n+1))$ for $j=1,\dots,n$. Further, $F(j/(n+1)) = j/(n+1)$. The parameter i is needed due to the discrete form of the empirical distribution, $\widehat{F}(z)$ has two values at each step and both can be the maximum.

The proof of the KSAD statistic in Equation (2.15) is analogous to the proof of the KS statistic. However, for the KSAD statistic, to avoid nulldivision, i can not equal 1 when $j=n$.

The proof for the CvM (Equation (2.13)) statistic is slightly more complicated. Since the empirical cdf $\widehat{F}(z)$ is a step function and $F(z)=z$, $\mathcal{T}^{CvM}$ from Equation (2.9) becomes:

$$\begin{aligned}
\mathcal{T}^{CvM} =& n \int_0^1 \left(\widehat{F}(z) - F(z)\right)^2 \mathrm{d}F(z) \\
=& n \int_0^1 \left(\widehat{F}(z) - F(z)\right)^2 \mathrm{d}F(z) \\
=& n \int_0^1 \widehat{F}(z)^2 \mathrm{d}F(z) - 2n \int_0^1 \widehat{F}(z)F(z)\mathrm{d}F(z) + n\int_0^1 F(z)^2 \mathrm{d}F(z).
\end{aligned}$$

Since $\widehat{F}(z)$ is constant and equal to $\widehat{F}(j/(n+1))$ between $j/(n+1)$ and $(j+1)/(n+1)$ for $j = 1, \ldots, n$, the first two integrals can be split into n smaller integrals:

$$\begin{aligned}
\widehat{\mathcal{T}}^{CvM} =& n \sum_{j=1}^{n} \int_{j/(n+1)}^{(j+1)/(n+1)} \widehat{F}\left(\frac{j}{n+1}\right)^2 \mathrm{d}F(z) \\
&- 2n \sum_{j=1}^{n} \int_{j/(n+1)}^{(j+1)/(n+1)} \widehat{F}\left(\frac{j}{n+1}\right) F(z) \mathrm{d}F(z) + n\left[\frac{1}{3}F(z)^3\right]_0^1 \\
=& n \sum_{j=1}^{n} \widehat{F}\left(\frac{j}{n+1}\right)^2 \Big[F(z)\Big]_{j/(n+1)}^{(j+1)/(n+1)} \\
&- n \sum_{j=1}^{n} \widehat{F}\left(\frac{j}{n+1}\right) \Big[F(z)^2\Big]_{j/(n+1)}^{(j+1)/(n+1)} + \frac{n}{3} \\
=& \frac{n}{3} + n \sum_{j=1}^{n} \widehat{F}\left(\frac{j}{n+1}\right)^2 \left\{\frac{j+1}{n+1} - \frac{j}{n+1}\right\} \\
&- n \sum_{j=1}^{n} \widehat{F}\left(\frac{j}{n+1}\right) \left\{\left(\frac{j+1}{n+1}\right)^2 - \left(\frac{j}{n+1}\right)^2\right\} \\
=& \frac{n}{3} + \frac{n}{n+1} \sum_{j=1}^{n} \widehat{F}\left(\frac{j}{n+1}\right)^2 - \frac{n}{(n+1)^2} \sum_{j=1}^{n} (2j+1) \widehat{F}\left(\frac{j}{n+1}\right)
\end{aligned}$$

To prove the empirical AD statistic (Equation (2.14)) we proceed in the same manner as for the CvM statistic:

$$\begin{aligned}
\mathcal{T}^{AD} =& n \int_0^1 \frac{\left(\widehat{F}(z) - z\right)^2}{z(1-z)} \mathrm{d}F(z) \\
=& n \sum_{j=1}^{n} \int_{j/(n+1)}^{(j+1)/(n+1)} \frac{\left(\widehat{F}\left(\frac{j}{n+1}\right) - \frac{j}{n+1}\right)^2}{\frac{j}{n+1}\left(1 - \frac{j}{n+1}\right)} \mathrm{d}F(z) \\
=& n \sum_{j=1}^{n} \frac{\left(\widehat{F}\left(\frac{j}{n+1}\right) - \frac{j}{n+1}\right)^2}{\frac{j}{n+1}\left(1 - \frac{j}{n+1}\right)} \Big[F(z)\Big]_{j/(n+1)}^{(j+1)/(n+1)} \\
=& \frac{n}{n+1} \sum_{j=1}^{n} \frac{\left(\widehat{F}\left(\frac{j}{n+1}\right) - \frac{j}{n+1}\right)^2}{\frac{j}{n+1}\left(1 - \frac{j}{n+1}\right)}.
\end{aligned}$$