Statistical analysis of credit risk

Topics in Default and Dependence modelling

Daniel Berg

Dissertation presented for the degree of
philosophiæ doctor

Department of mathematics
University of Oslo
2008

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
Nr. 707

ISSN 1501-7710

Cover: Inger Sandved Anfinsen.
Printed in Norway: AiT e-dit AS, Oslo, 2008.

Produced in co-operation with Unipub AS.
The thesis is produced by Unipub AS merely in connection with the
thesis defence. Kindly direct all inquiries regarding the thesis to the copyright
holder or the unit which grants the doctorate.

Unipub AS is owned by
The University Foundation for Student Life (SiO)

Statistical analysis of credit risk

Topics in Default and Dependence modelling

Ph.D. thesis

Daniel Berg

Department of Mathematics
Faculty of Mathematics and Natural Sciences
University of Oslo

Thesis advisors:	Fred Espen Benth, University of Oslo
	Xeni K. Dimakos, Norwegian Computing Center
	Kjersti Aas, Norwegian Computing Center
Thesis committee:	Christian Genest, Université Laval
	Mark Salmon, University of Warwick
	Nils Lid Hjort, University of Oslo

Acknowledgements

Several people have played a role in the completion of this work and it is my pleasure to express my gratitude to them here. Thanks are due to all my colleagues at the Norwegian Computing Center and at the Department of Mathematics of the University of Oslo who provided a stimulating environment and continuous support. In particular I would like to thank my supervisors at the Norwegian Computing Center, Kjersti Aas and Xeni Kristine Dimakos who demonstrated their expertise in the field through many discussions and enthusiastic guidance along the way. I would like to emphasize the pleasure I have had from close collaborations with Kjersti Aas. I am further indebted to all my other co-authors; Henrik Bakken, Claudia Czado, Rada Dakovic and Jean-François Quessy for our fruitful collaborations. Anders Løland deserves a special thank, always taking his time to help me with technical issues. I also want to express gratitude to my supervisor Fred Espen Benth at the University of Oslo for sharing his time with me. Finally, participants at several workshops and conferences as well as some anonymous referees provided useful comments. Needless to say, while I am grateful for all the support, I still claim full responsibility for any errors that may have occurred.

Parts of my work was carried out while visiting the Department of Mathematical Sciences at the Norwegian University of Science and Technology (NTNU) in Trondheim, Norway. I want to thank the administration, in particular Anne M. Kajander, and the statistics group at the Department for making my stay so pleasant. When deciding to start my PhD in 2004 my M.Sc. supervisor at NTNU, Håvard Rue played an important role. I am very grateful for his encouragements.

Finally, I would like to thank my wonderful wife Ingrid, our son Eivind Andreas, my parents, siblings, father- and sister in law for their encouragement and most importantly for reminding me that there are other, more important aspects of life besides mathematics.

Oslo, March 2008 Daniel Berg

Preface

This thesis was prepared in partial fulfillment of the requirements for the Ph.D. degree at the Department of Mathematics, Faculty of Mathematics and Natural Sciences at the University of Oslo. The work was carried out in the period from August 2004 to October 2007 under the supervision of Professor Fred Espen Benth, Chief Research Scientist Xeni Kristine Dimakos and Assistant Research Director Kjersti Aas. The work was carried out as part of the project *Statistical Analysis of Risk* at the Department of Mathematics of the University of Oslo, in collaboration with the Norwegian Computing Center. The project received financial support from the Norwegian Research Council, grant number 154079/420.

The thesis contains a brief introductory chapter and six papers. The introductory chapter was prepared for this thesis exclusively whereas the papers have all been submitted for publication. The introductory chapter and each paper is self-contained and can be read independently from the rest. The first page of each of the six papers contain the title, authors and their affiliation, an abstract, keywords and details of publication. Page numbers within the papers are given in parenthesis at the bottom of each page, underlining that the papers have been prepared and written separately. To emphasize the unity of the thesis, papers are also numbered consecutively at the top of each page and the list of references are collected in one bibliography placed at the end of the thesis.

Summary

This thesis consists of an introductory chapter and six papers. The papers can be grouped in two parts. The first part covers deafult prediction models and this is the topic of papers I and II. The second part covers dependency modelling using copulas, in particular copula goodness-of-fit testing. This is the topic of papers III - VI. In the following we summarize the abstracts, keywords and publication status of each paper. In cases of multiple authors the roles of each author is described.

Paper I: Bankruptcy prediction by generalized additive models.

Author(s): Daniel Berg

Abstract: We compare several accounting based models for bankruptcy prediction. The models are developed and tested on large data sets containing annual financial statements for Norwegian limited liability firms. Out-of-sample and out-of-time validation shows that generalized additive models significantly outperform popular models like linear discriminant analysis, generalized linear models and neural networks at all levels of risk. Further, important issues like default horizon and performance depreciation are examined. We clearly see a performance depreciation as the default horizon is increased and as time goes by. Finally a multi-year model, developed on all available data from three consecutive years, is compared with a one-year model, developed on data from the most recent year only. The multi-year model exhibit a desirable robustness to yearly fluctuations that is not present in the one-year model.

Keywords: Bankruptcy Prediction, Generalized Additive Models, Default Horizon, Performance Depreciation, Multi-year model.

Publication details: Applied Stochastic Models in Business and Industry, Vol. 23, No. 2 (2007), p. 129–143.

Paper II: Bankruptcy prediction in Norway: A comparison study.

Author(s): Rada Dakovic, Claudia Czado, Daniel Berg

Abstract: In this paper we develop statistical models for bankruptcy prediction of Norwegian firms in the limited liability sector using annual balance sheet information. We fit generalized linear-, generalized linear mixed- and generalized additive models in a discrete hazard setting. It is demonstrated that careful examination of the functional relationship between the explanatory variables and the probability of bankruptcy enhances the models' forecasting performance. Using information on the industry sector we model the unobserved heterogeneity between different sectors through an industry-specific random factor in the generalized linear mixed model. The models developed in this paper are shown to outperform the model with Altman's variables at all levels of risk. As a measure of models'

forecasting accuracy the area under the ROC curve is used.

Keywords: Bankruptcy Prediction, Industry Effects, Hazard Model, Generalized Linear Model, Generalized Linear Mixed Model, Generalized Additive Model.

Publication details: Submitted for publication, September 2007.

Comments: The major part of the analysis presented in this paper was run by Rada Dakovic. Claudia Czado suggested numerous improvements and corrected the manuscript recurrently. My contributions were primarily in project idea and definition phase, in preparation of the data set, some problem solving along the way and in the finalization of the manuscript.

Paper III: A copula goodness-of-fit test based on the conditional probability integral transform.

Author(s): Daniel Berg, Henrik Bakken

Abstract: We investigate a copula goodness-of-fit approach based on the conditional probability integral transform. The approach implicitly weights observations at corners and edges of the unit hypercube which makes it very powerful at detecting tail heaviness for large sample sizes. However, it is shown to perform rather poor for small sample sizes. We propose a generalization that allows for any weighting, making it more robust and more powerful for small sample sizes. Another weakness is that some deviations from the null hypothesis may be neglected. We show an example and propose an extension. Results from extensive Monte Carlo experiments show that our approach keeps prescribed levels well and that certain weighting schemes produce superior power for three alternative hypotheses. The margins are treated as unknown nuisance parameters and are replaced by their empirical distribution functions. A parametric bootstrap procedure is required to obtain reliable p-value estimates. Applied to daily log-returns of large cap stock portfolios the Gaussian- and one-parameter Clayton and Gumbel copulae are all strongly rejected, increasingly so for increasing dimension and sample size. The Student-t copula on the other hand, provides a good fit, indicating the presence of tail dependence in the daily log-returns of stock data.

Keywords: Copula, goodness-of-fit, conditional probability integral transformation, order statistic, parametric bootstrap, Anderson-Darling.

Publication details: Submitted for publication, April 2007.

Comments: Henrik Bakken contributed to the project idea and definition phase and to the early developments of the theoretical results. Daniel Berg did most of the programming, simulations and manuscript work.

Paper IV: Copula goodness-of-fit testing: an overview and power comparison.

Author(s): Daniel Berg

Abstract: Several copula goodness-of-fit approaches are examined, three of which are proposed in this paper. Results are presented from an extensive Monte Carlo study, where we examine the effect of dimension, sample size and strength of dependence on the nominal level and power of the different approaches. While no approach is always the best,

some stand out and conclusions and recommendations are made. A novel study of p-value variation due to permuation order, for approaches based on Rosenblatt's transformation is also carried out. Results show significant variation due to permutation order for some of the approaches based on this transform. However, when approaching rejection regions, the additional variation is negligible. Finally, motivated by the permutation study, new versions of some goodness-of-fit approaches are proposed and examined. The new versions consider all permutation orders of the variables and we see some power improvement over the approaches that consider one permutation order only.

Keywords: Copula, Cramér-von Mises statistic, empirical copula, goodness-of-fit, parametric bootstrap, pseudo-observations, Rosenblatt's transform.

Publication details: Submitted for publication, October 2007.

Paper V: Local sensitivity analyses of goodness-of-fit tests for copulas.

Author(s): Daniel Berg, Jean-François Quessy

Abstract: The asymptotic behavior of several goodness-of-fit statistics for copula families is obtained under contiguous alternatives. Many comparisons between a Cramér-von Mises functional of the empirical copula process and new moment-based goodness-of-fit statistics are made by considering their associated asymptotic local power curves. It is shown that the choice of the estimator for the unknown parameter can have a significant influence on the power of the Cramér-von Mises test, and that some of the moment-based statistics can provide simple and efficient goodness-of-fit methods. The paper ends with an extensive simulation study that aims to extend the conclusions to small and moderate sample sizes.

Keywords: Contiguous alternatives, copula, Cramér-von Mises statistic, empirical copula process, goodness-of-fit test, local power curves, rank-based estimators.

Publication details: Submitted for publication, October 2007.

Comments: The authors of this work are equal partners. Professor Jean-François Quessy derived most of the theoretical results and corresponding proofs and wrote the majority of the manuscript. Daniel Berg did most of the programming and simulations as well as some work on the manuscript. Both authors contributed equally to the project idea and definition phase.

Paper VI: Models for construction of multivariate dependence: a comparison study.

Author(s): Daniel Berg, Kjersti Aas

Abstract: We review models for construction of higher-dimensional dependence that have arisen recent years. A multivariate data set, which exhibit complex patterns of dependence, particularly in the tails, can be modelled using a cascade of lower-dimensional copulae. We examine two such models that differ in their construction of the dependency structure, namely the nested Archimedean constructions and the pair-copula constructions (also referred to as vines). The constructions are compared, and estimation- and simulation techniques are examined. The fit of the two constructions is tested on two different four-dimensional data sets; precipitation values and equity returns, using state of the art

copula goodness-of-fit procedures. The nested Archimedean construction is strongly rejected for both our data sets, while the pair-copula construction provides a much better fit. Through VaR calculations, we show that the latter does not overfit data, but works very well even out-of-sample.

Keywords: Nested Archimedean copulas, Pair-copula decompositions, Equity returns, Precipitation values, Goodness-of-fit, Out-of-sample validation.

Publication details: Submitted for publication, October 2007.

Comments: This work was almost equally split between Kjersti Aas and myself. Daniel Berg did the majority of the work concerning the hierarchical Archimedean copulas, while Kjersti Aas did the majority of the work concerning pair-copula constructions. Daniel Berg had the idea to the paper, and did most of the programming.

Table of Contents

1 Introduction

The last 30 years the international banking system has experienced many significant structural changes. Through mergers, acquisitions and globalization institutions have grown in size. Nevertheless, competition has substantially increased since regulations have been relaxed and has allowed banks to offer new products and to enter new markets and business activities. This expansion of the activities lead to new risks and thus an increasing demand for the quantification and pricing of risk (Crouhy et al., 2001).

Financial risks can be divided into market risk, credit risk, liquidity risk, operational risk, legal and regulatory risk and human factor risk (Crouhy et al., 2001). This thesis focuses on credit risk, the risk that the value of a portfolio changes due to unexpected changes in the credit quality of issuers or trading partners (McNeil et al., 2006). This includes both losses due to defaults and losses caused by changes in credit quality such as the downgrading of a counterparty in some rating system. A typical financial institution is exposed to credit risk in several areas of its business. Through lending and corporate bond portfolios, through any over-the-counter derivative transactions and through the credit derivative market. Hence, credit risk is very important as it relates to the core activity of most financial institutions (McNeil et al., 2006). Driven by the explosive markets for credit derivatives and the upcoming implementation of the regulatory requirements of Basel II, active credit risk management is of crucial importance for financial institutions. Consequently, there is a need for advanced mathematical and statistical methods.

Credit risk if often termed default risk. This term illustrates the fact that the probability of default is at the core of credit risk. Any model for the quantification of credit risk relies substantially on a good prediction of default probability. Other elements of credit risk is loss given default, exposure at default and dependencies between defaults. In this thesis we examine two such elements of credit risk. The first part of the thesis, constituting papers I and II, examines models for default prediction, more specifically accounting based models. Flexible nonparametric models are introduced to the problem of default prediction and compared to popular existing models. Important practical issues concerning default horizon and prediction performance are also examined. The second part, constituting papers III-VI, treats models for dependencies between defaults, more specifically copula models. While the evaluation of univariate distributions is well documented, the study of goodness-of-fit tests for copulas recently emerged as a challenging problem. Papers III-V are devoted to the topic of copula goodness-of-fit testing. New tests are proposed and existing tests are examined and, in some cases, generalized and extended. Most literature on copulas still focus on the bivariate case. However, in many practical applications the problem is higher dimensional. Paper VI deals with the construction of higher-dimensional copulas; proposed models are examined, compared and applied.

The purpose of the following introductory sections is to give a brief introduction to some of the issues encountered in credit risk management and some basic theory behind the material presented in the papers. First, a very brief introduction to the modelling of default probability is given in Section 1.1. Then an introduction to dependency modelling

through copulas is given in Section 1.2. Following the introductory chapter, the six papers and main contributions of the thesis are presented.

1.1 Default modelling

Two groups of models for the determination of default probability are commonly addressed in literature, accounting based models and market based models. Discriminant analysis and logistic regression models belong to the first group. The popular Z-score model (Altman, 1968) is based on discriminant analysis while Ohlsons O-score model (Ohlson, 1980) is based on generalized linear models with the logit link function. More recently several alternative models have been introduced to the problem of default prediction, e.g. neural networks (Wilson and Sharda, 1994), Bayesian methods (Posch et al., 2005), support vector machines (Chen et al., 2006) and generalized additive models (Berg, 2007a), among others.

The second group of models are market based models, also referred to as structural models. The popular Moody's KMV model (Crosbie, 1997) is a market based model based on the work of Merton (Merton, 1974). The market based models estimate the probability of default from the asset value of a firm. Stock prices are commonly used as proxies for the asset value. Several studies compare market based models with accounting based models, and most conclude that the market based models are superior, see e.g. Brockman and Turtle (2003) and Hillegeist et al. (2004). However, these studies use early accounting based models in their comparisons. More importantly, market based models require that firms are registered on a stock exchange. This is quite often not the case, in particular for small- and medium sized companies.

Papers I and II of this thesis both deal with topics in accounting based models and the reader is referred to these papers for further details.

1.2 Copulae

We now look closer at the modelling of dependence between several random variables using copulas. These random variables can for example be the returns of different assets. Traditionally, dependence modelling in the world of finance has been equivalent with the assumption of multivariate normality, with the correlation coefficient as a measure of dependence. However, the correlation coefficient is only sufficient for capturing the full dependence structure under the assumption of multivariate normality. The last few decades we have seen several important extensions. First, it is now clear that univariate distributions of financial returns are far from Gaussian. Second, the correlation coefficient has been shown to be grossly insufficient to provide an accurate description of the dependence structure of assets. Hence it is necessary to characterize the full joint multivariate distribution.

To characterize the dependence structure of several assets it is essential to realize that the joint multivariate distribution embodies two qualitatively different pieces of information on the assets. On one hand we have the marginal distributions while on the other hand we have the dependence structure of the assets, irrespective of their individual marginal distributions. Only the introduction of the copula allows a proper separation between these two pieces of information (Malevergne and Sornette, 2006). From an applied point of view, dependence is at the core of risk management. A proper modelling of dependence is crucial for i.a. diversification of risks, hedging strategies, securitizations and capital allocation.

The concept of copulas was first introduced independently by Hoeffding (1940, 1941) and Fréchet (1951) while Sklar (1959) first used the word copula in a statistical setting. He provided some general properties, established the copula function and showed that any joint distribution function can be considered a copula function. The literature on copulas has been growing rapidly, in particular in finance and insurance, starting with the work of Wang (1998), Frees and Valdez (1998) and Embrechts et al. (1999). Copulas are now successfully applied in finance and insurance, biostatistics (see e.g. Lambert and Vandenhende (2002)), hydrology (see e.g. Zhang and Singh (2006)) and environmental data (see e.g. Michele and Salvadori (2006)) among others. Mikosch (2006) (with corresponding discussion) discusses some critical remarks about the use of copulas, but despite his objections and scepticism copulas remain a flexible and popular tool for practitioners and an interesting topic for researchers. While the theory of copula models is now fairly well understood, inference for copula models is, to an extent, still under development (Genest and Favre, 2007).

This chapter continues as follows. First, Section 1.2.1 gives a brief survey of the most important properties of copulas. Then, some examples of parametric copula families are examined while Section 1.2.3 presents important and popular dependence measures.

For exhaustive and general introductions to copulas the reader is referred to Joe (1997) and Nelsen (1999) and for introductions oriented to financial applications Malevergne and Sornette (2006) and Cherubini et al. (2004)

1.2.1 Definition and main properties

Definition 1.1 (Copula) A d-dimensional copula is a multivariate distribution function C on $[0,1]^d$ with standard uniform marginal distributions.

C is a mapping of the form $C : [0,1]^d \to [0,1]$, i.e. a mapping of the unit hypercube into the unit interval. The following three properties must hold for $C(\mathbf{u}) = C(u_1, \ldots, u_d)$ to be a copula (McNeil et al., 2006):

1. $C(u_1, \ldots, u_d)$ is increasing in each component u_i.
2. $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ for all $i \in \{1, \ldots, d\},\ u_i \in [0,1]$.
3. For all $(a_1, \ldots, a_d),\ (b_1, \ldots, b_d) \in [0,1]^d$ with $a_i \leq b_i$ we have

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\ldots+i_d} C(u_{1i_1}, \ldots, u_{di_d}) \geq 0, \tag{1.1}$$

   where $u_{j1} = a_j$ and $u_{j2} = b_j$ for all $j \in \{1, \ldots, d\}$.

The first property is clearly required of any multivariate distribution function and the second property is the requirement of uniform marginal distributions. The third property is less obvious, but this so-called rectangle inequality ensures that if the random vector $(U_1, \ldots, U_d)'$ has distribution function C, then $P(a_1 \leq U_1 \leq b_1, \ldots, a_d \leq U_d \leq b_d)$ is non-negative.

The usefulness and importance of copulas for representing multivariate distributions with arbitrary marginals is summarized in the following theorem of Sklar (1959):

Theorem 1 (Sklar's theorem)
Let H be a joint distribution function with margins $F_1, \ldots, F_d$. Then there exists a copula $C : [0,1]^d \to [0,1]$ such that $F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$. Conversely, for any

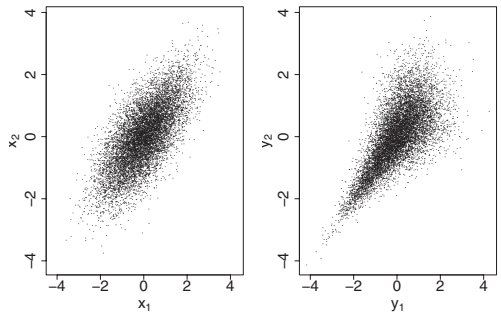


Figure 1.1: Two simulated data sets with standard normal marginal distributions, both with correlation coefficient 0.70.

distribution functions $F_1, \ldots, F_d$ and any copula C, the function H defined above is a d-dimensional distribution function with marginals $F_1, \ldots, F_d$. Furthermore, if $F_1, \ldots, F_d$ are continuous, C is unique.

A very powerful property shared by all copulas is their invariance under arbitrary, strictly increasing transformations of the random variables as stated in the following well-known theorem. For a proof in the bivariate case see Malevergne and Sornette (2006).

Theorem 2 (Invariance theorem)
Consider d continuous random variables $X_1, \ldots, X_d$ with copula C. Then, if $h_1(X_1), \ldots, h_d(X_d)$ are increasing transformations on the ranges of $X_1, \ldots, X_d$, the random variables $Y_1 = h_1(X_1), \ldots, Y_d = h_d(X_d)$ have exactly the same copula C.

This result demonstrates that the full dependence between the d random variables is completely captured by the copula, independent of the shape of the marginal distributions. In other words, Theorem 2 shows that the copula is an intrinsic measure of dependence between random variables. Under a monotonic change of variable from an old variable to a new variable, these two variables are comonotonic by definition. Intuitively it is natural that a measure of dependence between two random variables should be insensitive to the substitution of one of the variables by a comonotonic variable. This is precisely the content of Theorem 2 on copulas. In contrast, a measure of dependence such as the correlation coefficient is not invariant under a monotonic change of variable since it is a function of both the copula and the marginal distributions. Hence, it is not an intrinsic measure of dependence. We will return to this issue in Section 1.2.3.

To further illustrate the shortcomings of the linear correlation coefficient and the need for copulas and other measures of dependence, consider Figure 1.1. This figure shows two bivariate data sets, both having standard normal marginal distributions and correlation coefficient 0.70. Hence, in terms of the correlation coefficient the dependency is the same in the two data sets which is obviously not true.

1.2.2 Some copula examples

We now consider some examples of copulas, categorized as follows. First, we consider some *fundamental* copulas - important special dependency structures. We then proceed with *elliptical* copulas and finally *Archimedean* copulas. For simplicity we restrict ourselves to the bivariate case.

It is common to represent a bivariate copula by its distribution function

$$C(u_1, u_2) = P(U_1 \leq u_1, U_2 \leq u_2) = \int_0^1 \int_0^1 c(s,t)\mathrm{d}s\mathrm{d}t,$$

where the density $c(u_1, u_2)$ is given by

$$c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2}. \tag{1.2}$$

For *implicit* copulas the double integral on the right hand side of (1.2) does not have a closed form but is implied by a typically well-known bivariate distribution function. For *explicit* copulas the double integral has a closed form.

Fundamental copulas

First consider the simplest case where the continuous random variables X_1, X_2 are stochastically independent if and only if $C = \Pi$, where $\Pi(u_1, u_2) = u_1 u_2$. This is commonly referred to as the independence copula. At the other extreme, due to the rectangle inequality in (1.1), it can be shown that in order for X_1 to be a deterministic function of X_2, C must be either one of the two copulas

$$W(u_1, u_2) = \max(0, u_1 + u_2 - 1) \quad \text{or} \quad M(u_1, u_2) = \min(u_1, u_2)$$

which are usually referred to as the Fréchet-Hoeffding lower- and upper bounds, respectively. Any copula C represents a model of dependence that lies somewhere between these two extremes:

$$W(u_1, u_2) \leq C(u_1, u_2) \leq M(u_1, u_2), \quad u_1, u_2 \in [0, 1].$$

Figure 1.2 show perspective- and contour plots of the three fundamental copulas.

Elliptical copulas

Elliptical copulas do not have a simple closed form, but are derived from multivariate elliptical distribution functions. Hence, they are sometimes referred to as *implicit* copulas. In this survey we will consider two examples of elliptical copulas; the Gaussian and the Student copula. By construction, these two copulas are similar in their central parts and become more and more similar in the tails as the number of degrees of freedom of the Student copula increases. The two copulas can have drastically different behaviours with respect to the dependence between extremes as we will see in Section 1.2.3.

The Gaussian copula is derived from the multivariate Gaussian distribution and provides a natural setting for the generalization of multivariate Gaussian distributions to so-called meta-Gaussian distributions, applied in many areas. Meta-Gaussian distributions was introduced in Krzysztofowicz and Kelly (1996) and generalized to meta-ellipticity by Fang et al. (2002). The meta-Gaussian distribution has a Gaussian copula but differs

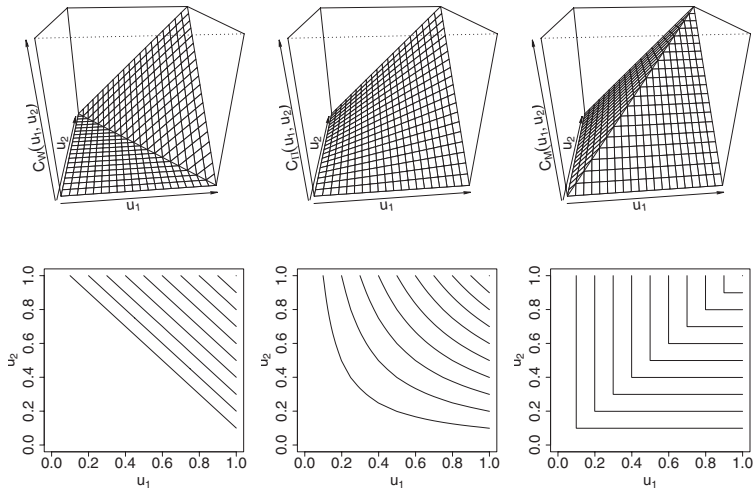


Figure 1.2: Surface (top) and contour (bottom) plots of three fundamental copulas, the countermonotonicity (left), independence (center) and comonotonicity (right) copulas.

from the multivariate Gaussian distribution in the marginal distributions which may be arbitrary.

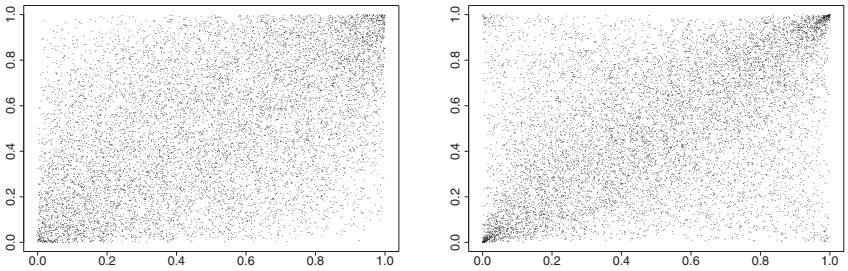
Elliptical copulas have the advantage of being rich in parameters, i.e. the number of parameters of a d-dimensional elliptical copula is $d(d-1)/2$ and $d(d-1)$ for the Gaussian and Student copulas, respectively. Elliptical copulas are also easily simulated, which makes them convenient for numerical simulations. A disadvantage is the restriction to radial linearity.

Example 1 (Gaussian copula)
Let Φ_ρ denote the standard Gaussian cumulative distribution with correlation coefficient ρ. Then the Gaussian copula with correlation coefficient ρ is given by

$$\begin{aligned} C_\rho(u_1, u_2) &= \Phi_\rho\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right) \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right\} dx\, dy, \end{aligned}$$

where ρ is the correlation coefficient and $\Phi^{-1}(\cdot)$ is the inverse of the standard univariate Gaussian distribution function. Figure 1.3(a) shows a scatter plot of the Gaussian copula.

Example 2 (Student copula)
The Student copula allows for joint fat tails and an increased probability of joint extreme events compared to the Gaussian copula. Let $T_{s,\nu}$ denote the Student cumulative distribution with shape coefficient s and degrees of freedom ν. Then the Student copula with shape



(a) Gaussian copula ($\rho = 0.5$). (b) Student copula ($s = 0.5,\ \nu = 2$).

Figure 1.3: Scatter plots of bivariate Gaussian (left) and Student (right) copulas.

coefficient s and degrees of freedom ν is given by

$$\begin{aligned} C_{s,\nu}(u_1, u_2) &= T_{s,\nu}\left(T_\nu^{-1}(u_1), T_\nu^{-1}(u_2)\right) \\ &= \int_{-\infty}^{T_\nu^{-1}(u_1)} \int_{-\infty}^{T_\nu^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-s^2}} \left\{1 + \frac{x^2 - 2sxy + y^2}{\nu(1-s^2)}\right\}^{-\frac{\nu+2}{2}} dx\, dy, \end{aligned}$$

where s is the shape parameter and $T_\nu^{-1}(\cdot)$ is the inverse of the standard univariate Student distribution function with ν degrees of freedom, expectation 0 and variance $\nu/(\nu-2)$. Figure 1.3(b) shows a scatter plot of the Student copula with $\nu = 2$ degrees of freedom.

Archimedean copulas

This extensively studied class of copulas has proved useful in a variety of applications. The class encompasses a large number of different copulas which arise very naturally in the context of survival analysis and frailty models. This explains their popularity in survival, insurance and credit risk literature. These copulas have closed form expressions and are thus sometimes referred to as *explicit* copulas. An Archimedean copula is defined as follows:

Definition 1.2 (Archimedean copula) Let φ be a continuouos, strictly decreasing and convex function from $[0, 1]$ to $[0, \infty)$ such that $\varphi(1) = 0$. Let $\varphi^{[-1]}$ denote the pseudo-inverse of φ:

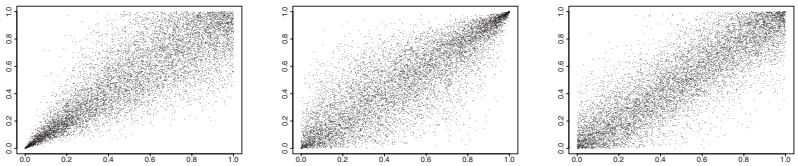
$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & 0 \le t \le \varphi(0), \\ 0, & t \ge \varphi(0). \end{cases}$$

Then the function

$$C(u, v) = \varphi^{[-1]}\{\varphi(u) + \varphi(v)\}$$

is an Archimedean copula with generator φ. When $\varphi(0) = \infty$, φ is said to be a strict generator and $\varphi^{[-1]} = \varphi^{-1}$.

For Archimedean copulas, the complexity of the dependence between two variables is reduced and embedded into the function of one single variable, the generator φ. This transforms the bivariate formulation into a simpler univariate one (Malevergne and Sornette, 2006).



(a) Clayton copula ($\theta = 4$). (b) Gumbel copula ($\theta = 3$). (c) Frank copula ($\theta = 10$).

Figure 1.4: Scatter plots of bivariate Clayton (left), Gumbel (middle) and Frank (right) copulas.

Note that the bivariate Fréchet-Hoeffding lower bound is an Archimedean copula, while the upper bound copula is not. Also, the independence copula is an Archimedean copula with generator $\varphi(t) = \ln t$. Among the large number of Archimedean copulas, the following three examples are particularly popular in the literature.

Example 3 (Clayton's copula)
The Clayton copula is sometimes also referred to as a Gamma frailty model or the Cook-Johnson copula. This copula has so-called left tail dependence, to which we will return in Section 1.2.3. It is given by

$$C_\theta^{Cl}(u_1, u_2) = \max\left\{\left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-\frac{1}{\theta}}, 0\right\}, \quad \theta \in [-1, \infty)$$

with generator $\varphi(t) = (t^{-\theta} - 1)/\theta$. Figure 1.4(a) shows a scatter plot of Clayton's copula.

Example 4 (Gumbel's copula)
The Gumbel copula is sometimes also referred to as the Gumbel-Hougaard or the Gumbel-Barnett copula. It is an extreme value copula and has upper tail dependence. It is given by

$$C_\theta^{Gu}(u_1, u_2) = \exp\left\{-\left[(-\ln u_1)^\theta + (-\ln u_2)^\theta\right]^{\frac{1}{\theta}}\right\}, \quad \theta \in [1, \infty),$$

with generator $\varphi(t) = (-\ln t)^\theta$. Figure 1.4(b) shows a scatter plot of Gumbel's copula.

Example 5 (Frank's copula)
The Frank copula has no tail dependence. It is given by

$$C_\theta^{Fr}(u_1, u_2) = -\frac{1}{\theta}\ln\left\{1 + \frac{(\exp(-\theta u_1) - 1)(\exp(-\theta u_2) - 1)}{\exp(-\theta) - 1}\right\}, \quad \theta \in \mathbb{R}$$

with generator $\varphi(t) = -\ln\{(\exp(-\theta t) - 1)/(\exp(-\theta) - 1)\}$. Figure 1.4(c) shows a scatter plot of Frank's copula.

Marshall and Olkin (1988) proved that, given a distribution function F defined on $\mathbb{R}^+$ such that $F(0) = 0$, the inverse generator $\varphi^{-1}(t)$ of an Archimedean copula, is the Laplace transform of F,

$$\varphi^{-1}(t) = \int_0^\infty \exp(-tx)\mathrm{d}F(x).$$

This suggests that frailty models (Shih, 1998) can provide a natural mechanism for generating random variables with Archimedean copulas (Malevergne and Sornette, 2006). Such

models are common in actuarial science because they offer a simple way of studying the joint mortality of a group of individuals sharing common risk factors (see e.g. Frees et al. (1996) among many others). In finance, they can also model the joint distribution of defaults of different obligators subjected to the same set of economic factors.

1.2.3 Measures of dependence

We have now seen how the general dependene structure of several random variables may be described through copulas. However, the use of copulas does not exclude more specific measures of dependence. In this section we describe some important dependence measures starting with the most basic concept, namely linear correlation. We then focus on on the more interesting family of concordance measures. Finally, we turn to tail dependence, a measure of extreme dependence. In each case we relate the measure to copulas. The two latter kinds of dependence measures, concordance measures and tail dependence coefficients, are both *copula based* dependence measures. In contrast to the ordinary linear correlation, these measures are functions of the copula only and can thus be used in the parametrization of copulas (McNeil et al., 2006).

Linear correlation

The linear correlation is probably still the most used measure of dependence in general. It plays an important role in financial theory, but it is important to realize that the concept is only really a natural one in the context of multivariate normal or, more generally, elliptical models.

Given two random variables X_1 and X_2, the linear correlation coefficient is given by

$$\rho(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)\,\mathrm{Var}(X_2)}},$$

provided that the variances $\mathrm{Var}(X_1)$ and $\mathrm{Var}(X_2)$ exist. $\mathrm{Cov}(X_1, X_2)$ is the covariance of X_1 and X_2. It is a measure of *linear* dependence and takes values in $[-1, 1]$. If X_1 and X_2 are independent then $\rho(X_1, X_2) = 0$, but the converse does not hold in general.

The linear correlation coefficient is invariant under strictly increasing linear transformations. However, it is not invariant under nonlinear strictly increasing transformations. Also, the linear correlation is only defined when the variances of X_1 and X_2 are finite. This restriction to finite-variance models is not ideal for a dependence measure and can cause problems when we work with heavy tailed distributions (McNeil et al., 2006).

Although the linear correlation takes values in $[-1, 1]$, these bounds can not always be reached. There are so-called *attainable* linear correlations that form a strict subset of $[-1, 1]$, governed by the Fréchet-Hoeffding bounds. As an illustration consider the following well known example from Embrechts et al. (1999). If we have two random variables with log-normal marginal distributions, $X_1 \sim \log\mathcal{N}(0, 1)$ and $X_2 \sim \log\mathcal{N}(0, \sigma^2)$, the lower- and upper bounds for $\rho(X_1, X_2)$ are given by $\rho_{\min} = \rho(\exp(Z), \exp(-\sigma Z))$ and $\rho_{\max} = \rho(\exp(Z), \exp(\sigma Z))$ respectively, where Z is a standard Gaussian random variable. One can easily show that

$$\rho_{\min} = \frac{\exp(-\sigma) - 1}{\sqrt{(\exp(1) - 1)(\exp(\sigma^2) - 1)}} \quad \text{and} \quad \rho_{\max} = \frac{\exp(\sigma) - 1}{\sqrt{(\exp(1) - 1)(\exp(\sigma^2) - 1)}}.$$

These attainable bounds for the linear correlation are shown in Figure 1.5 for different values of σ. Notice how the boundaries tend rapidly to zero as σ is increased. This

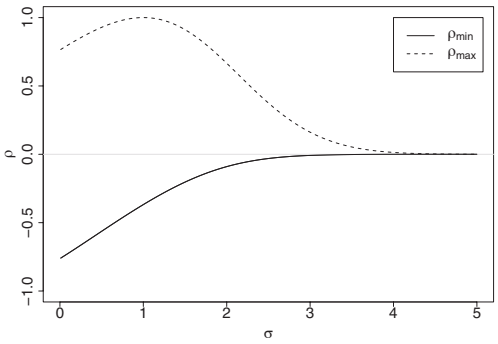


Figure 1.5: Maximum and minimum attainable linear correlations for log-normal random variables X_1 and X_2 where $X_1 \sim \log\mathcal{N}(0,1)$ and $X_2 \sim \log\mathcal{N}(0,\sigma^2)$.

shows, for example, that we can have situations where comonotonic random variables have very small linear correlation. Since comonotonicity is the strongest form of positive dependence, this provides a correction to the widely held view that small correlations imply weak dependence (McNeil et al., 2006).

Concordance measures

Kendall's tau. An important question for financial risk management is whether prices of two or more assets tend to rise or fall together. If they do, the diversification of risks will be difficult since diversification is based upon the fact that the fall of an asset is statistically balanced by the rise of another (Malevergne and Sornette, 2006). A natural way to quantify the tendency of assets to move together is to compare the probability that they rise (or fall) together with the probability that one of the two assets rises (falls) while the other one falls (rises). This can be translated mathematically as follows. Starting with two independent realizations (x_1, x_2) and $(\tilde{x}_1, \tilde{x}_2)$ of the same pair of random variables (X_1, X_2), let us consider the quantity

$$\rho_\tau = P\{(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) > 0\} - P\{(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) < 0\}. \tag{1.3}$$

The left-most term on the right hand side gives the probability of *concordance*, i.e. the probability that X_1 and X_2 move together upward or downward. In contrast, the right-most term on the right hand side represents the probability of *discordance*, i.e. the probability that the two random variables move in opposite directions. Equation (1.3) defines the population version of the so-called Kendall's tau. This quantity is invariant under increasing transformation of the marginal distributions. As a consequence, Kendall's tau depends only on the copula of (X_1, X_2). For continuous random variables, (1.3) can be transformed to $\rho_\tau = 2P\{(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) > 0\} - 1$, which yields the following expression in terms of a functional of the copula C of the two random variables (Malevergne and Sornette, 2006):

$$\rho_\tau(X_1, X_2) = 4 \int_0^1 \int_0^1 C(u_1, u_2) \mathrm{d}C(u_1, u_2) - 1. \tag{1.4}$$

From this equation, one may easily verify that Kendall's tau takes values in $[-1, 1]$, the lower bound being reached if and only if (X_1, X_2) are countermonotonic, while the upper

bound is reached if (X_1, X_2) are comonotonic. In addition, ρ_τ is zero for independent random variables. However, as for the linear correlation, ρ_τ may vanish even for non-independent random variables.

To evaluate (1.4), numerical integration is often called for. However, there are some nice expressions for particular families of copulas. Genest and MacKay (1986) have shown a relation between Kendall's tau and the generator of an Archimedean copula:

$$\rho_\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} \mathrm{d}t.$$

Table 1.1 provides closed form expressions for the relation between Kendall's tau and the dependence parameter of the Clayton, Gumbel and Frank copulas. For elliptical copulas, Lindskog et al. (2003) have shown that the relation

$$\rho_\tau = \frac{2}{\pi} \arcsin \rho$$

holds for any pair of random variables whose dependency structure is given by an elliptical copula.

Spearman's rho. This dependence measure can also be defined in terms of concordance and discordance for random pairs, but the most intuitive definition for our purposes involves copulas. For random variables X_1 and X_2 with marginal distributions F_1 and F_2, Spearman's rho is given by $\rho_s(X_1, X_2) = \rho(F_1(X_1), F_2(X_2))$. In other words, Spearman's rho is simply the linear correlation of the probability transformed random variables, which for continuous random variables is the linear correlation of their unique copula. As for Kendall's tau, Spearman's rho can be expressed in terms of the copula of the two random variables (McNeil et al., 2006):

$$\rho_s(X_1, X_2) = 12 \int_0^1 \int_0^1 \{C(u_1, u_2) - u_1 u_2\} \mathrm{d}u_1 \mathrm{d}u_2.$$

For higher dimensions it is common to define both concordance measures as matrices of pairwise Kendall's tau's and Spearman's rho's.

Tail dependence

Like the concordance measures, the coefficient of tail dependence is a measure of pairwise dependence which depends only on the copula of a pair of random variables X_1 and X_2 with continuous marginal distribution functions. The motivation for looking at tail dependence is that they provide measures of *extremal dependence*, or in other words, measurements of the strength of dependence in the tails of a bivariate distribution. The coefficients are defined in terms of limiting conditional probabilities of *quantile exceedances* (McNeil et al., 2006).

In the case of upper tail dependence, we look at the probability that X_2 exceeds its qth quantile, given that X_1 exceeds its qth quantile. Then we consider this limit as q goes to infinity. Obviously the roles of X_1 and X_2 are interchangeable. Formally, we have the following.

Table 1.1: Kendall's tau and coefficients of tail dependence for five popular copulas.

Copula	ρ_τ	λ_L	λ_U
Gaussian	$(2/\pi)\arcsin\rho$	0	0
Student(ν)	$(2/\pi)\arcsin\rho$	$2T_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right)$	$2T_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right)$
Clayton	$\theta/(\theta+2)$	$2^{-1/\theta}$ if $\theta > 0$, 0 otherwise	0
Gumbel	$1-1/\theta$	0	$2-2^{1/\theta}$
Frank	$1-4\theta^{-1}(1-D_1(\theta))$	0	0

Note: D_1 denotes the Debye function $D_1(\theta) = \theta^{-1}\int_0^\theta t/(\exp(t)-1)\mathrm{d}t$, and $T_{\nu+1}$ denotes the univariate Student distribution with $\nu+1$ degrees of freedom.

Definition 1.3 (Tail dependence) Let X_1 and X_2 be random variables with distribution functions F_1 and F_2. The coefficient of upper tail dependence of X_1 and X_2 is

$$\lambda_U = \lim_{q\to 1^-} P\{X_1 > F_1^{-1}(q) | X_2 > F_2^{-1}(q)\},$$

provided a limit $\lambda_U \in [0,1]$ exists. If $\lambda_U \in (0,1]$, then X_1 and X_2 are said to show upper tail dependence or extremal dependence in the upper tail. If $\lambda_U = 0$ they are said to be *asymptotically independent* in the upper tail. Analogously, the coefficient of lower tail dependence is

$$\lambda_L = \lim_{q\to 0^+} P\{X_1 < F_1^{-1}(q) | X_2 < F_2^{-1}(q)\},$$

provided a limit $\lambda_L \in [0,1]$ exists.

If F_1 and F_2 are continuous distribution functions, then we get simple expressions for λ_L and λ_U in terms of the unique copula C of the bivariate distribution. Indeed, one can show (McNeil et al., 2006) that

$$\lambda_L = \lim_{q\to 0^+} \frac{C(q,q)}{q},$$
$$\lambda_U = \lim_{u\to 1^-} \frac{1-2q+C(q,q)}{1-q}.$$

Using these expressions and L'Hopital's rule, one can easily show the tail dependence coefficients in Table 1.1. Note the positive tail dependence for the Student copula for all $\rho > -1$.

I

Bankruptcy prediction by generalized additive models

Daniel Berg
University of Oslo & Norwegian Computing Center

Abstract

We compare several accounting based models for bankruptcy prediction. The models are developed and tested on large data sets containing annual financial statements for Norwegian limited liability firms. Out-of-sample and out-of-time validation shows that generalized additive models significantly outperform popular models like linear discriminant analysis, generalized linear models and neural networks at all levels of risk. Further, important issues like default horizon and performance depreciation are examined. We clearly see a performance depreciation as the default horizon is increased and as time goes by. Finally a multi-year model, developed on all available data from three consecutive years, is compared with a one-year model, developed on data from the most recent year only. The multi-year model exhibit a desirable robustness to yearly fluctuations that is not present in the one-year model.

Key words

Bankruptcy Prediction, Generalized Additive Models, Default Horizon, Performance Depreciation, Multi-year model

Publication details

Applied Stochastic Models in Business and Industry
Vol. 23, No. 2 (2007), p. 129–143

I.1 Introduction

Since the work of Beaver (1966) and Altman (1968), bankruptcy prediction have been studied actively by academics and practitioners. This field of risk management continues to be very active, much due to the continuous development of new financial derivatives. For example, the pricing of credit derivatives relies on good estimates of counterparty risk. Two kinds of models are commonly addressed in the literature. First, there are accounting based models, for example discriminant analysis and logistic regression models. Second, there are market based models, also referred to as structural models, for example Merton models (e.g. the Moody's KMV model). The market models are based on the value of a firm set by the market. Stock prices are commonly used as proxies for the value. Several studies compare market based models with accounting based models, and most come to the conclusion that the market based models are superior, see e.g. Brockman and Turtle (2003) and Hillegeist et al. (2004). However, these studies use early accounting based models in their comparisons. And more importantly, market based models require that firms are registered on a stock exchange and this is quite often not the case. In Norway the majority of limited liability firms are not registered on any exchange. Hence, our focus is on accounting based models.

Linear discriminant analysis models have been widely used. Altmans popular Z-Score (Altman, 1968) is for example based on linear discriminant analysis. Generalized linear models, or multiple logistic regression models are also popular. Ohlsons O-Score (Ohlson, 1980) is based on generalized linear models with the logit link function, also referred to as logit analysis. Neural network models are powerful and popular alternatives, with the ability to incorporate a very large number of features in an adaptive nonlinear model, see for example Wilson and Sharda (1994).

Several authors have compared various models for bankruptcy prediction, see for example Tam (1991), Tam and Kiang (1992), Dimitras et al. (1996) and Altman and Narayanan (1997). Most of these compare NNs with discriminant- and/or logit analysis. See Atiya (2001) for a summary of these and several other studies.

Our main objective is to introduce Generalized Additive Models (GAM) as a flexible non-parametric alternative for bankruptcy prediction, and show that it performs significantly better than discriminant analysis, linear models and neural networks. The only application of GAM to bankruptcy prediction we have come across is De Giorgi and Burkhard (2006). To our knowledge there is no study comparing GAM with the methods mentioned above.

GAM is a generalization of the linear regression model. It replaces the usual linear function of a covariate with a sum of unspecified smooth functions, helping us discover potential non-linear shapes of covariate effects. The estimation of GAM and neural networks is more computationally demanding than for linear models, but with the rapidly increasing power of computers we expect an increasing application of such models in practice.

We develop several models using the same explanatory variables. To compare the models we use the validation methodology that is referred to as "out-of-sample" and "out-of-time" validation in Sobehart et al. (2000). The data set used is an extensive collection of annual financial statements of Norwegian limited liability firms in the period $1996-2000$ as well as the year of bankruptcy for all firms that filed for bankruptcy in the years $1996-2001$.

In addition to model comparison we examine the sensitivity of the GAM model to default horizon, and we test the depreciation of the prediction models, examining how the prediction power of a model depreciates as time goes by. This is very important to

consider when determining cut-off levels and also when considering model risk. Finally the performance of a multi-year model, developed on statements from three consecutive years, is compared with a one-year model, developed on statements from one year only.

The contributions of this paper is the comparison of GAM with other popular models for bankruptcy prediction as well as the illuminating results from the other important issues such as model validation methodology, default horizon, performance depreciation and model robustness with regards to the number of years of data included when building the model. The paper is aimed particularly at practitioners, but also applied researchers, who should consider all the above issues when building models for bankruptcy prediction based on financial statement data.

The paper proceeds as follows. Section I.2 describes the models we will examine. Section I.3 describes the data set and the explanatory variables, while Section I.4 discusses model development and validation methodologies. Section I.5 compares the prediction power of various models, out-of-sample and out-of-time. Section I.6 shows the sensitivity of a GAM model to default horizon, while Section I.7 shows the depreciation rate of a GAM model. Section I.8 compares the performance of a multi-year model and a one-year model and finally, Section I.9 presents a summary of our findings.

I.2 Prediction Models

When handling bankruptcy data it is natural to label one of the categories as success (healthy) and the other as failure (default) and to score these as 0 and 1 respectively. A typical data set will have a series of ones and zeros as the response variable Y. Associated with each Y there will often be observations on a set of explanatory variables $X_1, X_2, \ldots, X_p$. A bank will typically have information on the earnings and debt of each customer.

Since Altman (1968) proposed to use Linear Discriminant Analysis (LDA) to predict bankruptcy, several contributions have been made to improve the results, using different parametric, semiparametric and non-parametric models.

In contrast to normal-based regression models like the LDA, in which we wish to predict the value Y, given values for the explanatory variables, we will also be interested in predicting the probability π that $Y = 1$, given values for the explanatory variables (Krzanowski, 1998). Any probability is restricted to take values between 0 and 1, but a linear model can give rise to any value between $-\infty$ and ∞. It is thus necessary to transform π into a quantity that takes values in the interval $(-\infty, \infty)$ before a linear model can be sensibly applied. There are several such transformations, or link functions. We will only consider the logit link, $\varepsilon = \ln(\pi/(1-\pi))$, often denoted by $\varepsilon = \text{logit}(\pi)$.

I.2.1 Linear Discriminant Analysis

Linear discriminant analysis (LDA) is a multivariate statistical technique that leads to the development of a linear discriminant function maximizing the ratio of among-group to within-group variability, assuming that the variables follow a multivariate normal distribution and that the dispersion matrices of the groups are equal. Clearly, both of the assumptions pose a significant problem for the application of LDA in real-world situations, since they are difficult to meet (Doumpos and Zopounidis, 1999).

I.2.2 Generalized Linear Models

Generalized Linear Models (GLM) is a generalization of the multiple regression model

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon, \tag{I.1}$$

where ε has mean vector $\mathbf{0}$ and covariance matrix $\sigma^2\mathbf{I}$. The generalization makes use of the exponential family of distributions

$$f(y;\theta,\varphi) = \exp\{[y\theta - b(\theta)]/a(\varphi) + c(y,\varphi)\},$$

for some specific functions $a(\cdot), b(\cdot)$ and $c(\cdot)$, and parameters θ and φ. The GLM has the following features:

1. The Y_i's ($i = 1,\ldots,n$) are independent random variables sharing the same form of distribution from the exponential family.

2. The explanatory variables provide a set of linear predictors $\eta = \sum_j \beta_j X_j$ for $j = 1,\ldots,p$.

3. The link between 1 and 2 is that $g(\mu) = \eta$, where μ is the mean of Y. $g(\cdot)$ is called the link function of the model.

Two extensions of the multiple regression model (I.1) that characterize the GLM, are its applicability to any member of the exponential family of distributions, and the presence of a link function when connecting the linear predictor η to the mean μ of Y. This link is determined by the distribution of the random term ε in (I.1). If ε is logistically distributed, we use the logit link and the GLM is referred to as the logit model. We will consider the binomial distribution and the logit link function $g(\mu) = \ln(\mu/(n-\mu))$. Re-expressing this function in terms of π instead of μ, we obtain $g(\pi) = \ln(\pi/(1-\pi))$ or $\pi = \exp(\eta)/(1+\exp(\eta))$ (Krzanowski, 1998).

I.2.3 Generalized Additive Models

Hastie and Tibshirani (1986) proposed Generalized Additive Models (GAM). These models assume that the mean of the dependent variable depends on an additive predictor through a nonlinear link function. GAM extends the GLM by replacing the linear form $\sum_j X_j\beta_j$ with the additive form $\sum_j f_j(X_j)$. The linear regression step in GLM is replaced by a non-parametric additive regression step, where the data is used to determine the appropriate smooth function f. This is done through iterative smoothing operations and allows for various non-linear effects of the explanatory variables.

The logistic additive model, when applied to binary response data, takes the form $\ln(\pi/(1-\pi)) = \sum_j f_j(X_j)$ or $\pi = \exp(\sum_j f_j(X_j))/(1+\exp(\sum_j f_j(X_j)))$.

One of the main reasons for using GAM is that they do not involve strong assumptions about the relationship between two or more variables that is implicit in standard parametric regression. Such assumptions may force the fitted relationship away from its natural path at critical points. Also, since each of the individual additive terms are estimated using univariate smoothers, GAM avoids the problem of rapidly increasing variance for increasing dimensionality. This problem is referred to as the "curse of dimensionality" and is present in many nonparametric methods.

I.2.4 Feed-forward Neural Networks

We consider the supervised class of Neural Networks (NNs) called Multi-Layer Perceptrons. One hidden layer is used and no skip-layer connections are allowed. The probability of belonging to class k is then computed by

$$f_k(\mathbf{x}) = f_o\left(\alpha_k + \sum_{j=1}^{M} v_{jk} f_h(\beta_j + \sum_{i=1}^{N} w_{ij} x_i)\right), \tag{I.2}$$

from inputs to outputs. Here N, M and K are the number of input nodes (i.e. the number of explanatory variables), the number of nodes in the hidden layer and the number of output nodes (i.e. the number of possible classes), respectively. The activation function, $f_h(x)$, of the hidden layer is always taken to be the logistic function $f_h(x) = \exp(x)/(1 + \exp(x))$, while the output activation function, $f_o(x)$, may either be logistic or linear (Aas et al., 1999). We use the logistic output activation function only.

A NN with no hidden layers is identical to the GLM, while a NN with one hidden layer, where the hidden layer uses nonlinear activation functions such as the logistic function, is nonlinear in the parameters and corresponds to multivariate nonlinear logistic regression (Aas et al., 1999).

I.3 Data

Our data sets are extensive collections of annual financial statements for all limited liability firms registered at the Norwegian register for business enterprises over the years $1996 - 2000$. The 5 data sets all include a company identification number, explanatory variables examined and the year of bankruptcy. For firms that had not failed at the time of bankruptcy data extraction (2001), the year of bankruptcy was set to missing. When referring to, for example, a model developed from 1996 data with a 2 year default horizon, we mean a model developed from the 1996 financial statements, where a response variable Y is set to 1 if the year of bankruptcy was 1997 or 1998 and 0 otherwise.

One could argue that the study is distorted by the fact that firms that are in a jam often either put off submitting their statements or are unable to submit them at all. All firms in Norway are obliged by law to submit their annual financial statements by a certain date. When constructing the data set, firms delaying their submission have been allowed for and the bankruptcy date have been set in retrospect. However, there may be certain firms that are excluded from the study due to such circumstances, and this may distort the predicted default probabilities somewhat. However, it will not affect the primary objectives and contributions of this paper, stated in the Introduction.

A particular feature of the data is the very small number of defaults. Of approximately $100,000$ firms each year only about 1% defaulted the next year. This is representative of bankruptcy prediction. Bankruptcy is a rare and extreme event. However, since we have such a large data set, 1% of $100,000$ firms is still $1,000$ firms, we are able to develop and validate models in a proper manner.

I.3.1 Explanatory variables

The choice of, and investigation of explanatory variables is not one of the objectives in this paper. There are several studies of properties, relationships and empirical selection

Table I.1: Explanatory variables employed and their definition

Variable	*Definition*
REVANM	No. of auditor remarks
AGE	Age of firm
EKA*	Equity share of total assets (solidity)
TKR*	Return on capital employed (profitability)
UBE*	Outstanding public dues to total assets
LEV*	Trade credit to total assets
LIK*	Cash minus short term debt to revenue from operations (liquidity)
LDEB*	Consolidated long term liabilities to total assets
DIV*	Dividends to total assets
INDUSTRY	Which industry sector a firm belongs to
CurrentR*	Current assets to current liabilities (liquidity)
QuickR*	Current assets less inventory to current liabilities (liquidity)
RetAss*	Return on assets (profitability)

Note: For the variables marked with an asterisk the first differences are also investigated.

of explanatory variables, see for example Beaver (1966). The appropriate variables to use will vary with region and industry.

The explanatory variables considered here are found mainly in Bernhardsen (2001) and is a collection of financial ratios, an industry indicator, the number of auditor remarks and some first differences of the ratios. Through these first differences (the change since the previous year) we are able to utilize not only the most recent financial statement data of a firm, but also data from the previous year. In a preliminary analysis we removed variables that were not significant in any model. The remaining 13 variables and 10 first differences, i.e. 23 variables in total, are summarized in Table I.1. First differences are included for variables marked with an asterisk.

All variables, except for the industry indicator, the number of auditor remarks and the first differences, are defined as the deviance from their industry mean. These variables will then reflect a firms risk compared to other firms within the same industry.

I.4 Methodology

I.4.1 Model Development Framework

When developing models we include all the explanatory variables summarized in Table I.1. In practice a stepwise procedure should be applied to only include explanatory variables that add significant predictive power to the model. Since we develop and test so many models such a stepwise procedure is too time-consuming.

We do not exclude variables that are highly correlated. The inclusion of highly correlated explanatory variables may cause problems in practice, but only if interpretations of the individual effects of the explanatory variables are attempted. When including highly correlated variables such interpretations should be avoided, due to the phenomena multicollinearity. However, if a model is constructed solely for the purpose of prediction, then multicollinearity will not be of concern.

When developing models we generally use 60% of the data set, randomly selected from the full data set and referred to as the training set. The remaining 40% is used for validation and is referred to as the out-of-sample test set.

I.4.2 Validation Framework

The performance statistics of models can be highly sensitive to the data sample used for validation. To avoid embedding unwanted sample dependency, quantitative models should be validated on observations of firms that are not included in the sample used to build the model. This is referred to as out-of-sample validation in Sobehart et al. (2000).

If we develop a model from 1996 financial statements, using a two year default horizon, we are predicting probabilities that firms will fail during $1997 - 1998$. That means we can't build this model until 1999, when the 1998 data is available. The model can then be applied to 1998 financial statements, predicting default probabilities for $1999 - 2000$. But how good will the model perform on these 1998 data? Validating the model on 1998 data is referred to as out-of-time validation and is the measure most interesting for practitioners. We investigate both out-of-sample and out-of-time validation.

To compare models we consider so-called power curves, visually indicating the predictive performance of the various models. Power curves display the trade-off between Type I and Type II error for all possible values of the measure of interest. Type I and Type II errors are the errors of misclassifying a bankrupt firm as healthy, and misclassifying a healthy firm as bankrupt, respectively. In statistical terms, power curves represent the cumulative probability distribution of default events for different default probabilities (Sobehart et al., 2000).

While power curves is a convenient way of visualizing model performance, it is often desirable to have a single measure that summarizes the predictive accuracy of each risk measure for both Type I and Type II errors into a single statistic. We employ one of the metrics proposed in Sobehart et al. (2000), namely the Accuracy Ratio (AR). This metric is obtained by comparing the power curve of the model under investigation with that of the perfect model. The closer the power curve is to the perfect power curve, the better the model performs. To calculate the summary statistic we focus on the area A that lies above the power curve of a random model (the 45° line) and below the power curve of the model under investigation. The larger the area below the curve and above the 45° line, the better the model is doing overall. The maximum area B that can be enclosed above the 45° line is achieved by the perfect curve. This maximum area is equal to 0.5. The ratio, A/B is referred to as the AR. It summarizes the predictive power over the entire range of possible risk values and is a fraction between 0 and 1.

To compare models we employ a resampling scheme where several subsets are resampled, randomly, from the full test set. For each of these subsets the AR is calculated and a t-test is performed to determine if a model performs significantly better than another, at a certain confidence level. When validating models, we sample 100 subsets, each consisting of 5000 firms, hence we have 99 degrees of freedom for the Student-t distributed variable. We use a 99.5% confidence level.

I.5 Model Comparison

We now present the results from a comparison of two year default horizon models. Linear discriminant analysis (LDA), generalized linear models (GLM), generalized additive models

Table I.2: Accuracy ratio means and standard deviations for various default prediction models.

Model	*AR Mean*	*AR Std*
LDA	0.713	0.03
GLM-Logit	0.720	0.04
NN	0.723	0.05
GAM-Logit	0.773	0.04

Note: 1996 data, two year default horizon, **out-of-sample** validation.

(GAM) and single-hidden-layer neural networks (NN) are compared. For the NN models we use weight decay to help the optimization process and avoid overfitting. We use a weight decay of 0.01. We also use an accuracy ratio maximizing function to determine the optimal network size. The network size corresponds to the number of nodes in the hidden layer, M in Equation (I.2). The output function $f_o(x)$ is chosen to be logistic. We wish to compare the various models using the same training and test data sets, hence no cross-validation was performed.

I.5.1 Out-of-sample Validation

We first perform out-of-sample validation. We develop one model from the 1996 training set and test this model on the 1996 out-of-sample test set. We then develop one model from the 1997 training set and test this model on the 1997 out-of-sample test set, and correspondingly for the 1998 and 1999 data sets. The results for the 1996 models are displayed in the left graph of Figure I.1, showing the power curves of each model. The LDA, GLM and NN models seem to perform equally well, while the GAM model seems to outperform the others. To confirm this visual impression we look at the sampled AR statistics, displayed in Table I.2. We see that all models have approximately the same standard deviation and that the GAM model has a higher mean than the other models. Table I.3 shows whether or not a model performs significantly better than the models above it in the table, from left to right, the uppermost model to the model directly above. The table includes the results of the $1996, 1997, 1998$ and 1999 models. For the 1996 models our visual impression from the power curves is confirmed. There is no significant difference between LDA, GLM and NN while GAM significantly outperforms the others. For the 1997 models the GLM and NN models significantly outperform the LDA. For 1998 and 1999 the GLM does not perform significantly better than the LDA, but now the NN performs significantly better than the GLM. For all years the GAM model, with a confidence level of 99.5%, performs significantly better than all other models tested.

I.5.2 Out-of-time Validation

The results from the out-of-sample validation are interesting, but not exactly what we are interested in. We seek the performance on future data, hence we perform out-of-time validation on the 1998 data. The resulting power curves are displayed in the right graph of Figure I.1, and the corresponding AR statistics are displayed in Table I.4. The significance indicator tells us whether or not a model performs significantly better than the models above it in the table. We see that GAM still significantly outperforms all

Table I.3: Significance indicators stating whether or not a model performs significantly better than the models above.

Model	*1996*	*1997*	*1998*	*1999*
LDA	-	-	-	-
GLM-Logit	F	T	F	F
NN	FF	TF	TT	TT
GAM-Logit	TTT	TTT	TTT	TTT

Note: The combination 'TF' indicates that a model does and does not perform significantly better than the uppermost model and the model directly above it in the table, respectively. Two year default horizon, **out-of-sample** validation, 99.5% confidence level.

Table I.4: Accuracy ratio means and standard deviations for various models.

Model	*AR Mean*	*AR Std*	*Signif.*
GLM	0.676	0.04	-
LDA	0.678	0.04	F
NN	0.695	0.04	FF
GAM	0.726	0.04	TTT

Note: The significance indicator states whether or not a model is significantly better than the ones above. 1996 data, two year default horizon, **out-of-time** validation on 1998 data, 99.5% confidence level.

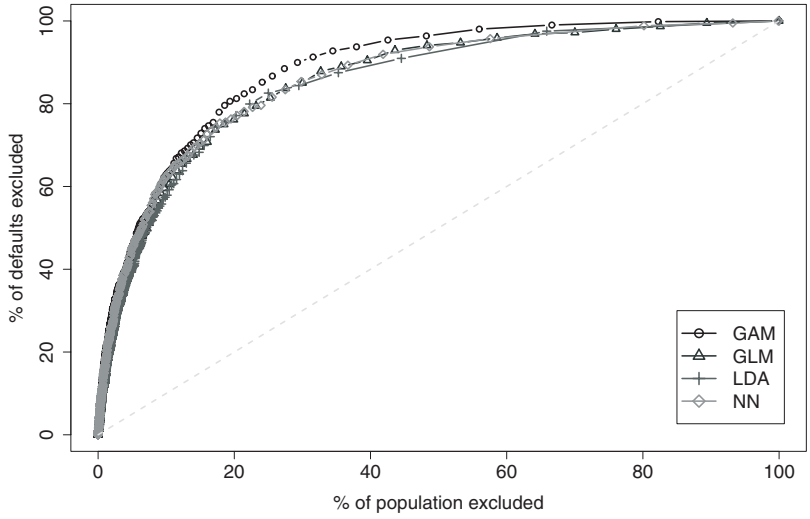
the other models. LDA, GLM and NN do not differ significantly. At high risk levels it seems like the NN model performs almost as good as the GAM, but as we move towards lower risk levels the GAM model outperforms all the other models, including the NN. This nicely demonstrates the importance of examining the model at the appropriate levels of risk. Note that while it varies which model performs second best, the GAM model seems to perform best at all levels of risk.

I.6 Default Horizon

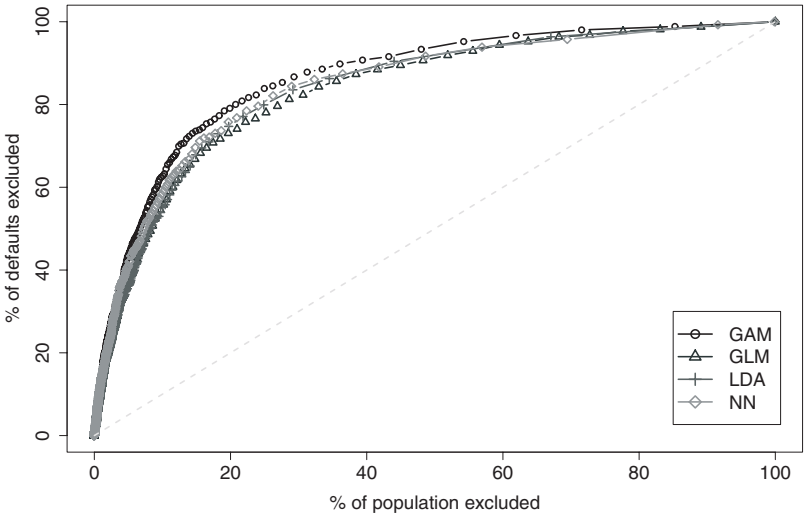
The term default horizon refers to the time horizon for which the model tries to predict. A one year default horizon model will define firms that fail during the first year after model development as default, while a two year default horizon model will define firms that fail during the first two years as bankrupt.

There are several reasons why the default horizon is important to consider. When entering a credit derivative contract for example. The time period of the contract will determine the default horizon. Or say the intended use is credit granting decision making. Then the highest prediction power is achieved by choosing a short default horizon, but this will 'hide' firms that are in distress but not as urgent and severe as those identified by the model. Alternatively, choosing a long default horizon will yield early warnings of distress, enabling preventive actions. Perhaps the best solution is to continuously use several models, each serving their own specific purpose.

We develop several GAM models on the same data set (1996), but with varying default horizons. In order to test a default horizon up to five years, we perform out-of-sample



(a)



(b)

Figure I.1: Prediction power of LDA, GLM, NN and GAM models. 1996 data, two year default horizon, **out-of-sample** and **out-of-time** validation: (a) Out-of-sample validation; (b) Out-of-time validation on 1998 data.

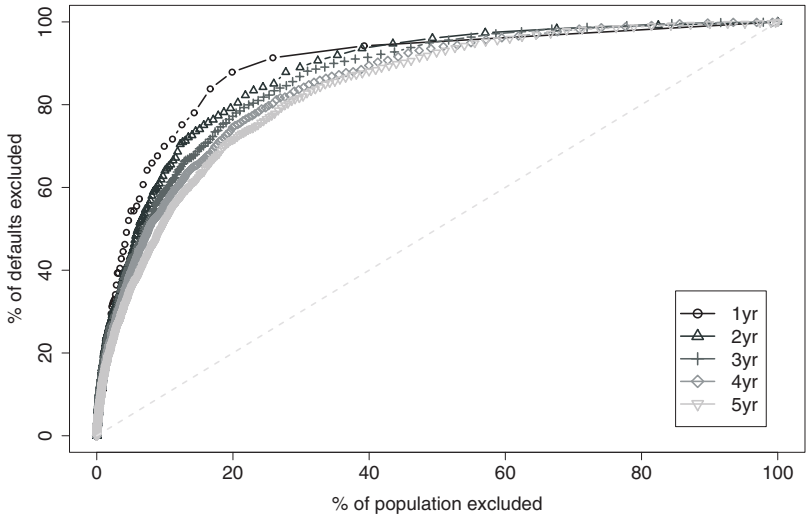


Figure I.2: Power graph indicating discriminating power as default horizon varies. GAM models, 1996 data, **out-of-sample** validation.

validation. Figure I.2 shows the results and we clearly see that the performance is reduced as the default horizon is increased. This is an expected, but nevertheless important result, and practitioners should keep this in mind when choosing a default horizon and assessing model risk. Table I.5 displays the results from the resampling procedure, and we see that the performance is reduced significantly for each year added to the default horizon.

By looking at which explanatory variables prove most significant we find that the longer the default horizon the more variables proved significant. This is especially evident if we compare 1 and 5 year default horizons. This indicates that signs of short term financial distress can be detected by looking at quite few variables. In general we can conclude that for longer default horizons the signs of distress are not so easily detected and much more complex interrelational structures are present. In such cases good statistical models are crucial for detecting important information and insight regarding the riskiness of firms. We also note that for all models the strongest variables are the ones we expected would be dominating: number of accountant remarks, age, industry, outstanding public dues and trade credit.

I.7 Performance Depreciation

When developing a model one might wish to keep this model for some time. In this case it is very important to be aware of the depreciation rate of the model. If for example a bank wishes to exclude 80% of the defaults at all times, the cut-off point needs to be adjusted as the model depreciates. This depreciation is also very important to consider if we are to

Table I.5: Accuracy ratio means and standard deviations for GAM models as default horizon varies.

Default Horizon	*AR Mean*	*AR Std*	*Signif.*
5 years	0.672	0.02	-
4 years	0.701	0.03	T
3 years	0.732	0.02	TT
2 years	0.760	0.04	TTT
1 year	0.784	0.07	TTTT

Note: The significance indicator states whether or not a model performs significantly better than the models above it in the table. 1996 data, varying default horizons, **out-of-sample** validation, 99.5% confidence level.

attempt to estimate the model risk. These are some reasons to examine the depreciation rate of bankruptcy prediction models as time goes by.

Let us look at how the performance of a one year default horizon model depreciates as time goes by. When performing out-of-time validation in Section I.5.2 this was basically what we did. We built a two year default horizon GAM model on the 1996 data and tested its performance on 1998 data, that is 2 years into the future. We now repeat this exercise and test the 1996 model on 1996, 1997, 1998, 1999 and 2000 data, that is $0-4$ years into the future. Table I.6 shows us the AR statistics from the resampling procedure. We see that there is a big decrease in mean performance from 0 to 1 year ahead. We also see that there is no significant difference in performance on data 1 and 2 years ahead. From 2 to 3 and 4 years ahead we again see a significant decrease in performance. Figure I.3 shows us the power curves for the models tested. This figure adds important information compared to the numbers in Table I.6. An interesting property seen is that the performance stays quite good, even 4 years into the future for high risk levels. However, the figure shows that to maintain an exclusion of for example 80% of the defaults, the cut-off point will have to be drastically increased as time goes by. We also note that the greatest depreciation happens the first year. The model performs much better on out-of-sample data than on out-of-time data. This is a natural effect of overfitting. The depreciation from 3 to 4 years ahead is relatively small in comparison. The point discussed in Section I.4.2, of models performing better than others at some risk levels but worse at other risk levels is also nicely demonstrated. Consider the performance 1 and 2 years ahead, we see that the model seems to perform better 1 year ahead for high risk values, that is when approximately $0-18\%$ of the population is excluded, while it performs better 2 years ahead if more than 18% of the population is excluded.

I.8 Multi-Year Model

We suspect that several actors in the market use only the most recent data when building bankruptcy prediction models. This is justified by the fact that the most recent data best reflect the characteristics of the data on which it will be used. But then the assumption is made that these characteristics change from year to year, and if this is true then the developed model will not be interesting anyway since it will only be applicable on contemporary data. So we must assume, unless we have good reason to believe otherwise, that the

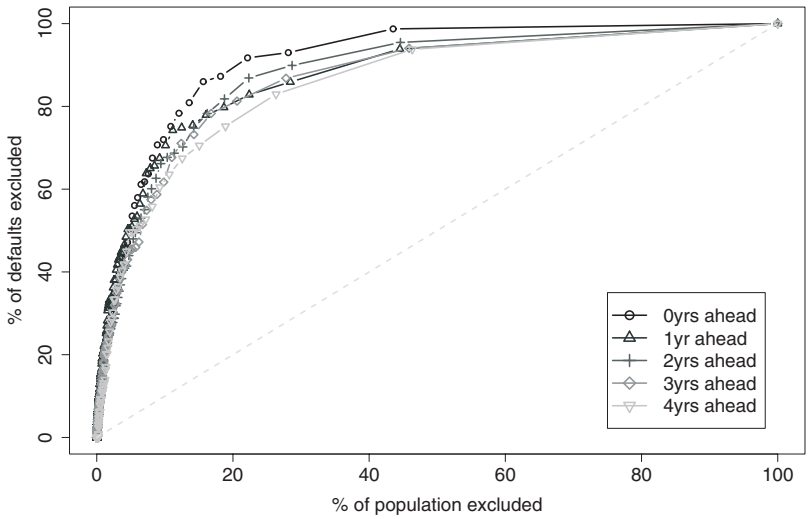


Figure I.3: Predictive power depreciation $0-4$ years into the future for a GAM model. 1996 data, one year default horizon.

Table I.6: Accuracy ratio means and standard deviations showing performance depreciation as time goes by.

No. yrs into future	*AR Mean*	*AR Std*	*Signif.*
4	0.699	0.09	-
3	0.735	0.08	T
1	0.756	0.08	FT
2	0.770	0.07	FTT
0	0.824	0.06	TTTT

Note: The significance indicator states whether or not a model performs significantly better than the models above. 1996 data, one year default horizon, 99.5% confidence level.

Table I.7: Accuracy ratio means and standard deviations for the M_{98} and M_{96-98} models.

Test Data	M_{96-98}	M_{98}	*Signif.*
1998	0.780 (0.06)	0.752 (0.07)	T
1999	0.755 (0.07)	0.707 (0.09)	T
2000	0.759 (0.08)	0.747 (0.09)	F

Note: Standard deviations in parentheses. The significance indicator states whether or not the M_{96-98} model performs significantly better than the M_{98} model, with a confidence level of 99.5%. One year default horizon, **out-of-sample** and **out-of-time** validation.

characteristics driving bankruptcy are constant. And if this is constant we should include as much data as possible when developing the model, since more data will give better estimates of default risk. Considering this and having seen the depreciation of models as time goes by, we compare a one-year model with a multi-year model. The one-year model is built on 1998 data while the multi-year model is built on data from three consecutive years, $1996 - 1998$. Both models are one year default horizon GAM models. For the multi-year model we utilize much more data than for the one-year model. Henceforth we will refer to the multi-year model and the one-year model as M_{96-98} and M_{98} to ease notation. The subscript denotes the years of data used to develop the model.

There are several arguments to consider multi-year models, in addition to those already mentioned. We are able to utilize more data, giving our models a better basis for detecting signs of distress. The significance of variables in a multi-year model is less dependent on the macroeconomic conditions specific to one year. A model, developed on one year of data only, will build signs of distress specific to that year into the model. A multi-year model on the other hand is expected to smooth out such year-specific effects. This way we would expect a multi-year model to be more robust than a one-year model, making it interesting for practitioners, especially those who know there might be some years until a new model is developed.

We developed several one year default horizon GAM models, one model for each year of data. Out-of-sample validation for each of these models shows that the out-of-sample performance varies quite much from year to year. This justifies considering a multi-year model. We never know if next year will be a good or bad year for model development. By using several years of data we better guard ourselves against such yearly fluctuations.

We perform out-of-sample validation, testing the models on the 1998 data, and out-of-time validation on 1999 and 2000 data. Unfortunately we do not have data that enables us to test the performance of the multi-year model more than two years into the future. However, Table I.7, still shows us interesting results. We see that M_{96-98} is more robust than M_{98}, as expected. The AR for M_{98} falls quite low for the 1999 test data while the multi-year model performs well for all test sets. The resampling procedure shows that M_{96-98} performs significantly better than M_{98} on the 1998 and 1999 test data, with a confidence level of 99.5%. On the 2000 data there is no significant difference in performance. The fact that the multi-year model outperforms the one-year model on the 1998 data is interesting. Apparently the $1996 - 1997$ data adds information about the 1998 out-of-sample test set, that the 1998 training set does not include.

I.9 Summary and discussion

We have shown, through out-of-sample and out-of-time validation, that generalized additive models significantly outperforms other models like linear discriminant analysis, generalized linear models and neural networks.

If the IT system prevents the implementation of GAM models or the method is deemed non-intuitive and hard to justify to managers, an approximation can be used. One can define dummy variables, a number of variables each representing an interval of the values of the original variable, often referred to as binning. For example $d_1 = \mathbf{1}_{\{DIV \leq 0\}}$, which means that d_1 will equal 1 if DIV is less than or equal to zero, and zero otherwise. Then a regression is performed with all the dummies. This will be an approximation since it allows for non-linear effects. The advantage is that it is very easy to explain the effect and meaning of each variable and that once the dummies are defined all we need to do is apply simple linear or ridge regression on the dummies. The disadvantage is the process of defining the intervals for each dummy. This process can be subjective and cumbersome if not automated. Also, the advantage of interpretation comes with a price, variables that are highly correlated must be excluded from the model to avoid multicollinearity problems.

We recommend further use of the out-of-time validation framework, employing re-sampling procedures. The ability to say whether a model is significantly better than another, given a certain confidence level, is of uttermost importance and is best achieved by re-sampling. Also, power curves adds valuable visual information about performance at different levels of risk. In practice one may only be interested in the performance for certain risk levels. In this case one can simply modify the AR-calculation to only consider the risk levels of interest.

Further we have shown how sensitive models are to the choice of default horizon. This is important to consider when for example negotiating a credit derivative contract and for banks monitoring and actively managing their portfolios.

We also examined the depreciation rate of models. This is very important to consider when deciding on desired levels of risk and cut-off points and also when estimating model risk. Figure I.3 nicely demonstrates the need for cut-off adjustment as time goes by.

Finally we compared a one-year model, estimated from one year of data, with a multi-year model, estimated from three consecutive years of data. The multi-year model performed significantly better on out-of-sample test data and also on some out-of-time test data. The multi-year model seemed to be more robust, performing stable across the test data sets while the one-year model performed rather poor on one of the test data sets. The main reason for the multi-year model outperforming the one-year model is believed to be the size of the training data set. Unless there are good reasons to believe that the characteristics driving bankruptcies have changed, we argue that data from several years should be utilized. An interesting issue for further research would be finding how far back it is optimal to go. More data is needed to do such a study.

The results and conclusions obtained in this paper are only indicative since they are conditioned on the data set at hand. Nonetheless, the discussions are fully applicable and the topics discussed should be considered by practitioners developing a model.

Acknowledgements: The author gratefully acknowledge the helpful advice and comments from an anonymous referee and editor. Thank is also due to colleagues at the Norwegian Computing Center, in particular Assistant Research Director Kjersti Aas and Chief Research Scientist Xeni Kristine Dimakos for their support and guidance. Finally,

thank is due to Associate Professor Sjur Westgaard at the Department of Industrial Economy and Technology Management, Norwegian University of Science and Technology, for providing the data set on which the entire work is based.

II

Bankruptcy prediction in Norway: a comparison study

Rada Dakovic
Technische Universität München

Claudia Czado
Technische Universität München

Daniel Berg
University of Oslo & Norwegian Computing Center

Abstract

In this paper we develop statistical models for bankruptcy prediction of Norwegian firms in the limited liability sector using annual balance sheet information. We fit generalized linear-, generalized linear mixed- and generalized additive models in a discrete hazard setting. It is demonstrated that careful examination of the functional relationship between the explanatory variables and the probability of bankruptcy enhances the models' forecasting performance. Using information on the industry sector we model the unobserved heterogeneity between different sectors through an industry-specific random factor in the generalized linear mixed model. The models developed in this paper are shown to outperform the model with Altman's variables at all levels of risk. As a measure of models' forecasting accuracy the area under the ROC curve is used.

Key words

Bankruptcy Prediction, Industry Effects, Hazard Model, Generalized Linear Model, Generalized Linear Mixed Model, Generalized Additive Model

Publication details

Submitted for publication - September 2007.

II.1 Introduction

Bankruptcy prediction is attracting the attention of both academics and practitioners since the seminal works of Beaver (1966) and Altman (1968) in the late 1960s. Researchers traditionally rely on linear combinations of financial ratios as predictors and use a single observation per company. Several recent studies emphasize that the relationship between explanatory variables and the logit (probit) of the default probability is often *non-linear* (see Sobehart and Stein (2000), Falkenstein et al. (2000), Berg (2007a)). Additionally, Shumway (2001), Chava and Jarrow (2004) and Hillegeist et al. (2004) emphasize that a single-period approach neglects important information when a company is at risk but remains solvent. To avoid these deficiencies of the traditional approaches, suggestions have been made to use neural networks or generalized additive models to model non-linearities and hazard models instead of single-period static models to incorporate information from the complete period at risk (see Shumway (2001), Chava and Jarrow (2004) and Hillegeist et al. (2004)). In this paper we follow the pattern of non-linear modeling and evaluate the forecast performance of both static- and hazard models. Our models are broad in scope in the sense that they apply to all industry sectors, including financial institutions. Additionally, we model the unobserved heterogeneity between different industry sectors by introducing an industry-specific intercept as a random factor in our non-linear logistic regression.

The purpose of this paper is to develop statistical models for bankruptcy prediction of firms in the limited liability sector of Norway. The $98,421$ firms in our database are observed on an annual basis and most of them are not registered on any exchange. Therefore, we have to rely on traditional accounting-based methods. We examine whether one can enhance bankruptcy prediction accuracy by a careful examination of the functional relationship between explanatory variables and the probability of bankruptcy. We utilize generalized additive models (GAM) in exploratory analysis to reveal non-linear relations to be used in the generalized linear model (GLM). Further, we show that when one carefully models, through linear and non-linear transformations of covariates in GLM, prediction accuracies of GLM and GAM are approximately the same. A slight improvement of model performance is further obtained by estimating an industry-specific intercept as a random factor. In the assessment of model accuracy we use ROC and CAP curve analysis, which became widely accepted in the bankruptcy prediction literature since they were introduced in Sobehart et al. (2000) and Sobehart and Keenan (2001). Forecasting ability of our models is stable over different subsets of the dataset and over time. The models are then compared to the celebrated Altman's Z-score model which uses a linear combination of 5 financial ratios as a proxy of the default probability. The Altman's model is reestimated and shown to capture less publicly-available information than the models we use in our analysis for this specific dataset. Improvements obtained by using a hazard instead of static setting are minimal, possibly because the maximal period at risk in our sample is low compared to previous studies of hazard models.

The paper is organized as follows. In Section II.2 we give a brief historical overview of bankruptcy prediction methods and outline statistical methods used in our analysis. Section II.3 describes our bankruptcy database. In Section II.4 we present the fitted models and evaluate their out-of-time prediction performance. Section II.5 concludes and discusses open problems for further research.

II.2 Statistical models for bankruptcy prediction

The bankruptcy prediction literature involves a number of statistical techniques used to obtain reliable estimates of default probability. The studies of Beaver (1966) and Altman (1968) that employ univariate and multivariate discriminant analysis respectively, are considered pioneering investigations of the relationship between the financial status of a company and its probability of failure. Subsequently, new statistical methods including gambler's ruin and option pricing theory, as well as linear regression, have been successfully applied in empirical analysis (see Wilcox (1971), Merton (1974), Martin (1977), Ohlson (1980), Zmijewski (1984)). Techniques used nowadays to construct bankruptcy prediction models involve neural and Bayesian networks (Tam and Kiang, 1992; Sun and Shenoy, 2007), theory of point processes (Das et al., 2007; Duffie et al., 2007), support vector machines (Härdle et al., 2005), and many others. For extensive reviews of related literature the reader is referred to Altman and Hotchkiss (2005), Altman and Narayanan (1997) or Falkenstein et al. (2000). We focus our attention on methods that emphasize the use of survival analysis and industry effects in failure prediction.

Recently several studies, including Shumway (2001), Chava and Jarrow (2004) and Hillegeist et al. (2004), have indicated that conventional models have a drawback of being based on the utilization of only a single observation per company. Traditionally the default probability of a company, irrespective of its bankruptcy status, has been dependent solely on its last available set of predictors. Such models, often called static, are shown to be outperformed by dynamical hazard models that incorporate the financial history of a company from the entire observation period. Applications of survival analysis techniques in bank failure prediction has a long history (see LeClere (2000) and Haling and Hayden (2006) for a review), while in bankruptcy prediction of non-financial institutions these methods have been disregarded since the work of Shumway (2001). Shumway argues that information neglected by static models can significantly improve model's forecasting accuracy, and highlights the simplicity of maximum likelihood estimation in the dynamical framework. Since the discrete hazard model plays one of the central roles in our investigation, we briefly outline its setting.

Assume that each firm i in the study has a failure time T_i and a censoring time C_i, both observed at discrete times, and that T_i, C_i are independent random variables with values in $\{1, \ldots, k\}$, where k denotes the end of the observation period. The observable lifetime of a firm i is then $S_i = \min(T_i, C_i)$. Let Δ_i denote the random censoring indicator given by

$$\Delta_i = \begin{cases} 1, & T_i \leq C_i, \quad \text{(non-censored)}, \\ 0, & T_i > C_i, \quad \text{(censored)}. \end{cases}$$

In addition to the observed lifetime s_i, we consider firm-specific time-varying covariates $\mathbf{x}_{it} \in \mathbb{R}^p$, that are assumed to have an influence on the lifetime. The data is given by

$$(s_i, \delta_i, \mathbf{x}_i(s_i)), \qquad i = 1, \ldots, n,$$

where $\mathbf{x}_i'(s_i) := (\mathbf{x}_{i1}', \ldots, \mathbf{x}_{is_i}')$ is the history of firm i until the observed lifetime s_i, and δ_i is the observed censoring indicator.

The basic quantity characterizing S_i is the *discrete hazard function*

$$\lambda(t|\, \mathbf{x}_i(t)) := \mathrm{P}(S_i = t|\, S_i > t-1, \mathbf{x}_i(t)), \qquad t \in \{1, \ldots, k\}, \tag{II.1}$$

which is assumed to be dependent on parameters or functions to be estimated. The exact form of the dependence of the hazard rate $\lambda(t|\,\mathbf{x}_i(t))$ on time-varying firm-specific covariates is given in Section II.4.2. Under certain conditions, parameters of dynamical hazard model can be estimated in the framework of ordinary binary regression by treating the annual bankruptcy indicators as independent binomials (see Fahrmeir and Tutz (2001) or Shumway (2001)). Precise assumptions under which the correspondence between the two models holds can be found in Arjas and Haara (1987) or Fahrmeir and Tutz (2001, p. 396). We note here that the hazard model built on only one year of data coincides with the static model.

We conclude this section with a short outline of research where the significance of industry effects in bankruptcy prediction modeling was discussed. Plat and Plat (1990, 1991) are among the first studies that illustrate the importance of industry-relative adjustments in failure prediction. Subsequently, a number of papers documented the impact of industry groupings on bankruptcy announcements. Lang and Stulz (1992) examine contagion and competitive intra-industry effects on default rate, while Alfo et al. (2005) use random industry effects to anticipate problematic firms. For a detailed review of reasons for presence of industry-specific information in bankruptcy prediction models, and an extensive list of references where these reasons are elaborated, the reader is referred to Chava and Jarrow (2004).

II.3 Data set

II.3.1 The Data Set

Financial statements and bankruptcy status for limited-liability firms in Norway are observed on an annual basis in the time period 1996-2000. Firms reporting non-positive total assets were eliminated. Balance sheets with book equity, short term debt or revenue from operations equal to 0 were excluded from further investigation in order to avoid null divisions when calculating financial ratios. Exploratory data analysis indicated a substantial lag between the date of the last reported financial statement and the bankruptcy date. This phenomenon is also described in Bernhardsen (2001). Among companies that were declared bankrupt in the time-period 1997-2001, only 25% report their financial statements in the last year of their existence, while for the remaining 75% we observe at least one year of missing data. For this reason, all companies (bankrupt as well as non-bankrupt) with missing financial statements for at least one year before bankruptcy or the end of the observation period were excluded from further analysis. Since salaried household work and internal organs and organizations were represented by only 3, respectively 1 firm in the resulting sample, these two industry sectors were not considered in our paper. For each of the continuous covariates used to estimate the model, the values below 0.2%-quantile and above 99.8%-quantile were calculated, and firms with these financial statements were also excluded from further consideration. Truncation of the data is often performed in order to remove outliers that frequently occur due to typos or recording errors (see Shumway (2001), Chava and Jarrow (2004)). Note that the discrete hazard model can be estimated in the framework of binary regression, since censoring can occur only at the end of the observation period due to the data requirements. Our final sample consists of $436,145$ firm-years corresponding to $98,421$ unique firms, and contains $2,270$ bankruptcies.

II.3.2 Explanatory variables

The set of covariates included in our model building process combines conventional accounting ratios used in bankruptcy prediction studies, and covariates traditionally employed in the credit risk analysis at Norges Bank, presented in Bernhardsen (2001). We take into account 5 frequently used default risk factors: profitability, solidity, liquidity, size and leverage. Additionally, we include industry indicator variables, information on the number of auditor remarks, age of a company, and an indicator of dividends paid current year as predictors of default probability. The list of time-varying explanatory variables considered in our analysis consists of

1. REVANM_{it} - the number of auditor remarks of firm i at time t,
2. AGE_{it} - age of a firm i at time t measured in years,
3. DIV_{it} - indicator for dividends paid by firm i at time t (dichotomous),
4. EKA_{it} - book value of equity to total assets of firm i at time t (solidity),
5. SIZE_{it} - logarithm of total assets of firm i at time t (size),
6. CashR_{it} - cash and marketable securities to current liabilities of firm i at time t (liquidity),
7. RetAss_{it} - return on assets to total assets of firm i at time t (profitability),
8. CLTA_{it} - current liabilities to total assets of firm i at time t (leverage).

Here $i = 1, \ldots, n, t = t_0(i), \ldots, S_0(i)$, where $t_0(i)$ and $S_0(i)$ denote the starting and survival time of firm i. In addition, the information about the sector a firm belongs to (fixed over time) is included into the model through an industry-specific intercept estimated as a fixed or random factor. The distribution of firms and bankruptcies with respect to the industry sector is given in Table II.1.

In the remainder of this section we shortly discuss properties of covariates described above placing the emphasis on the difference between bankrupt and solvent firms.

Summary statistics for REVANM can be found in Table II.2. The table indicates that distributions of the number of auditor remarks among solvent and bankrupt firms are distinct. Additionally, we notice the change in the distribution in 1999, when the percentage of companies with more then one remark becomes lower compared to the period 1996-1998. This change is possibly due to the fact that prior to 1998 Norwegian law was imposing only moderate sanctions for non-reporting financial information while more stringent regulations were introduced in 1998.

In Figure II.1 the histogram of AGE (bellow 50 years) and kernel density estimators of covariates EKA, SIZE, RetAss, CashR and CLTA [1] for bankrupt and non-bankrupt firms in the complete data-set are given. The difference between bankrupt and solvent companies is clearly visible. For all covariates except of CashR, the difference in modes of the respective distributions is evident. In agreement with econometric intuition, we observe that bankrupt companies are more likely to have low values of EKA, RetAss,

[1] Since continuous covariates that we consider have heavy-tailed distributions, the kernel density estimators are given for values of EKA above 2%-quantile, RetAss between 1%- and 99%-quantile, CashR and CLTA below 80%- and 98%-quantile , respectively.

Table II.1: Distribution of firms with respect to the industry sector

Industry sector	*Total*	*Bankrupt*	*Bankrupt (%)*
Forestry and agriculture	607	17	2.80
Fishing	1059	18	1.70
Mining and extraction	599	6	1.00
Industry	9810	315	3.21
Water and power supply	300	1	0.33
Building and construction	8503	242	2.85
Commodity trade, vehicle and domestic appliance repair	26340	863	3.28
Hotel and catering activity	3356	224	6.67
Transport and communication	5616	123	2.19
Finance and insurance	4225	22	0.52
Property operations, rental business and commercial services	32373	348	1.07
Public administration	46	2	4.35
Education	558	12	2.15
Health and social service	2086	20	0.96
Other social and personal services	2943	57	1.94
Total	98421	2,270	2.31

CashR and high values of CLTA. Notice that for all covariates except of SIZE, the shape of the density of bankrupt firms differs from respective shape estimated within the group of solvent companies. Further empirical analysis (not presented here) shows that distributions of all covariates except of REVANM are stable over time, and that indicator of dividends payed can be seen as a potentially powerful predictor of failure.

II.4 Results

The aim of this section is to describe the models fitted to our dataset and evaluate and compare their forecasting performance. We fit a GLM, transforming covariates according to the exploratory analysis in Section II.4.1. We compare it to the generalized linear mixed model (GLMM) with random, industry-specific intercept, and a GAM. All models are estimated using the same set of explanatory variables, presented in Section II.3.2. The three models (GLM, GLMM, GAM) are estimated and validated on different subsets of the complete sample. To illustrate the discriminative power of our models, we fit them using financial statements 1996-1999, predict default probabilities in 2000, and construct kernel density estimates of their logits for bankrupt and solvent firms, respectively. Finally, we estimate a GLM with Altman's variables (see Altman (1968)), using only linear transformations of covariates, and compare it to the models described above. All four models are estimated in a static and hazard setting, respectively. Their forecasting accuracy does not improve substantially if the hazard model is used instead of the static one.

(II.6)

Table II.2: Distribution of bankrupt (B) and non-bankrupt (NB) firms (percent) according to number of auditor remarks

		No. of auditor remarks				
Year	*Status*	*0*	*1*	*2*	*3*	≥ 4
1996	NB	75.42	22.38	0.57	1.46	1.63
	B	11.85	64.22	0.71	18.25	23.22
1997	NB	76.33	21.63	0.52	1.38	1.52
	B	15.33	59.80	1.26	18.59	23.62
1998	NB	76.03	21.93	0.55	1.35	1.49
	B	13.82	63.09	1.09	17.09	22.00
1999	NB	81.58	17.39	0.85	0.02	0.18
	B	23.17	67.20	6.60	0.18	3.03
2000	NB	81.82	17.22	0.81	0.00	0.15
	B	23.89	68.14	5.90	0.00	2.06

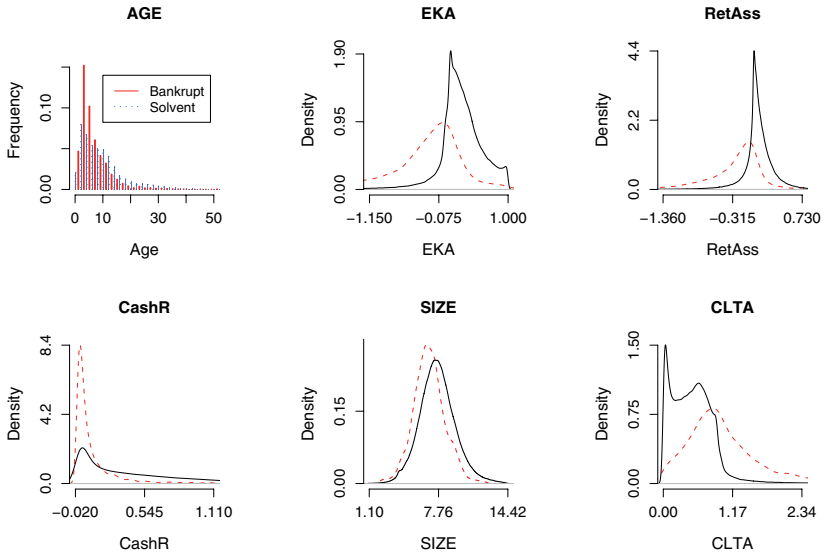


Figure II.1: Histogram of AGE. Kernel density estimators of EKA, RetAss, CashR, SIZE and CLTA for bankrupt (dashed line) and solvent (solid line) firms separately.

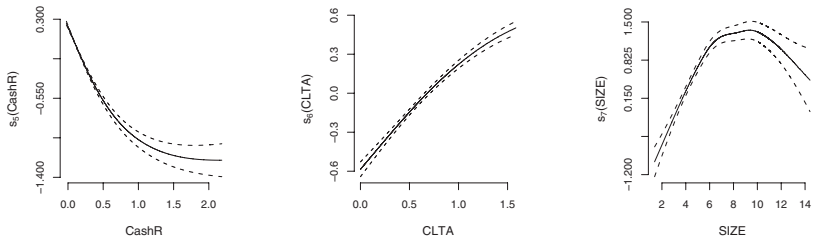


Figure II.2: Estimated transformations of CashR (left, values of CashR below 90%-quantile), CLTA (middle, values of CLTA below 98%-quantile), and SIZE (right).

II.4.1 Exploratory Data Analysis

In order to make inference about the form of the functional relationship between the logit of the hazard rate (II.1) and continuous explanatory variables, we fit the GAM

$$\begin{aligned}\operatorname{logit}\lambda(t|\,\mathbf{x}_{it}) = {} & \beta_1\mathrm{REVANM}_{it} + \beta_2\mathrm{AGE}_{it} + \sum_{j=1}^{15}\beta_{3j}D_{ji} + \beta_4\mathrm{DIV}_{it} \\ & + s_5(\mathrm{CashR}_{it}) + s_6(\mathrm{CLTA}_{it}) \\ & + s_7(\mathrm{SIZE}_{it}) + s_8(\mathrm{RetAss}_{it}) + s_9(\mathrm{EKA}_{it})\end{aligned} \tag{II.2}$$

to the complete dataset. Here $D_{ji},\ j = 1,\ldots,15$, are dummy variables being 1 if firm i belongs to industry j and 0 otherwise and $\hat{s}_5,\ldots,\hat{s}_9$ are smoothing splines to be estimated. We observe non-linear relations in the spline terms $\hat{s}_5, \hat{s}_7, \hat{s}_8$ and $\hat{s}_9$, i.e. for variables CashR, SIZE, RetAss and EKA. The $\hat{s}_6$ term, describing the effect of CLTA on default probability, can be considered as linear. The forms of the estimated functions are depicted in Figures II.2 and II.3. For more details regarding the theory of GAM, the reader is referred to Hastie and Tibshirani (1990).

In Figure II.2 functions $\hat{s}_5, \hat{s}_6$ and $\hat{s}_7$ are plotted. Function $\hat{s}_5$ is plotted for values of CashR lower than the 90%-quantile, since above that value the form of $\hat{s}_5$ becomes unstable, possibly due to outliers. The plot suggests to use the function $\exp(-\mathrm{CashR})$ in a corresponding GLM model. Similarly, $\hat{s}_6$ is depicted for values of CLTA below the 98%-quantile. In that range, the estimated function can be considered as linear, and therefore CLTA enters linearly in our final GLM model. Function $\hat{s}_7$ is plotted on the whole range of the covariate SIZE, and we decide to use a polynomial of degree 2 to model its influence on the default probability.

Visual examination of the left plots in Figure II.3, where $\hat{s}_8$ and $\hat{s}_9$ are plotted for the entire range of variables RetAss and EKA respectively, suggests that possibly separate functions should be fitted for negative and positive values for two of the covariates considered. In the middle and right hand side plots of Figure II.3, the estimated functions are depicted on the negative and positive half axes, respectively. We decide to use a polynomial of degree 3(2) for modeling the influence of negative (positive) values of RetAss, while the effect of EKA is modeled by 2 separate polynomials of degree 2, each fitted on the corresponding half axis.

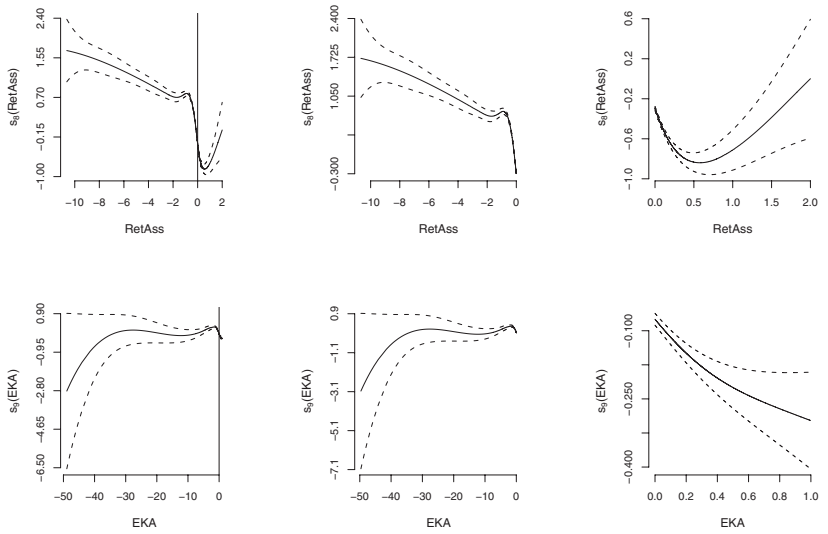


Figure II.3: Estimated transformation of RetAss (upper) and EKA (lower). Whole range of RetAss, EKA (left), negative values (middle), positive values (right).

We emphasize here that both the shape of functions we use in our analysis and the list of explanatory variables should be seen as suggestion. Our recommendation is to carefully investigate functional relationship of covariates to default probability, and use the results of exploratory data analysis to build the final model.

II.4.2 Fitted models

We assume that the discrete hazard rate (II.1) depends only on the last value of the covariates [2], namely

$$\lambda(t|\mathbf{x}_i(t)) := \lambda(t|\mathbf{x}_{it}),$$

and consider the following three models for $\lambda(t|\mathbf{x}_{it})$.

Model 1: GLM

[2]Other specifications of $\lambda(t|\mathbf{x}_i(t))$ that include time-lagged covariates are possible, but rarely used in practice. Results not reported here indicate that models with $\lambda(t|\mathbf{x}_i(t)) = \lambda(t|\mathbf{x}_{it}, \mathbf{x}_{it-1})$ when fitted to our dataset do not improve forecasting ability.

The GLM has the form indicated in Section II.4.1, namely

$$\begin{aligned}\text{logit}\lambda(r|\,\mathbf{x}_{it}) &= \beta_1 I(\text{REVANM}_{it} > 0) + \beta_2 \text{AGE}_{it} + \sum_{j=1}^{15} \beta_{3j} D_{ji} + \beta_4 \text{DIV}_{it} \\ &+ \beta_{51} I(\text{EKA}_{it} \geq 0)\text{EKA}_{it} + \beta_{52} I(\text{EKA}_{it} \geq 0)\text{EKA}_{it}^2 \\ &+ \beta_{53} I(\text{EKA}_{it} < 0)\text{EKA}_{it} + \beta_{54} I(\text{EKA}_{it} < 0)\text{EKA}_{it}^2 \\ &+ \beta_{61}\text{SIZE}_{it} + \beta_{62}\text{SIZE}_{it}^2 + \beta_{71} I(\exp(-\text{CashR}_{it})) \\ &+ \beta_{81} I(\text{RetAss}_{it} \geq 0)\text{RetAss}_{it} + \beta_{82} I(\text{RetAss}_{it} \geq 0)\text{RetAss}_{it}^2 \\ &+ \beta_{83} I(\text{RetAss}_{it} < 0)\text{RetAss}_{it} + \beta_{84} I(\text{RetAss}_{it} < 0)\text{RetAss}_{it}^2 \\ &+ \beta_{85} I(\text{RetAss}_{it} < 0)\text{RetAss}_{it}^3 + \beta_9 \text{CLTA}_{it}\end{aligned} \tag{II.3}$$

We remark here that all coefficients included in Model (II.3) were significant at the 5%-level when being estimated from the complete dataset. Analysis not reported here show that interactions of continuous ratios with industry sector indicators are not significant. Therefore industry-specific slopes for EKA, RetAss, CashR, SIZE and CLTA are not included into our model.

Model 2: GLMM

In order to incorporate homogeneity within industry sectors, while allowing for heterogeneity between different sectors, we fit the GLMM with random industry-specific intercept

$$\begin{aligned}\text{logit}\lambda(t|\,\mathbf{x}_{it}) &= \beta_1 I(\text{REVANM}_{it} > 0) + \beta_2 \text{AGE}_{it} + \sum_{j=1}^{15} b_j D_{ji} + \beta_4 \text{DIV}_{it} \\ &+ \beta_{51} I(\text{EKA}_{it} \geq 0)\text{EKA}_{it} + \beta_{52} I(\text{EKA}_{it} \geq 0)\text{EKA}_{it}^2 \\ &+ \beta_{53} I(\text{EKA}_{it} < 0)\text{EKA}_{it} + \beta_{54} I(\text{EKA}_{it} < 0)\text{EKA}_{it}^2 \\ &+ \beta_{61}\text{SIZE}_{it} + \beta_{62}\text{SIZE}_{it}^2 + \beta_{71} I(\exp(-\text{CashR}_{it})) \\ &+ \beta_{81} I(\text{RetAss}_{it} \geq 0)\text{RetAss}_{it} + \beta_{82} I(\text{RetAss}_{it} \geq 0)\text{RetAss}_{it}^2 \\ &+ \beta_{83} I(\text{RetAss}_{it} < 0)\text{RetAss}_{it} + \beta_{84} I(\text{RetAss}_{it} < 0)\text{RetAss}_{it}^2 \\ &+ \beta_{85} I(\text{RetAss}_{it} < 0)\text{RetAss}_{it}^3 + \beta_9 \text{CLTA}_{it}\end{aligned}$$

where $b_j \sim N(0, \sigma^2), j = 1, \ldots, 15$, are independent random variables representing the frailty effect. The model is estimated by the penalized quasi-maximum likelihood method described in Breslow and Clayton (1993).

Model 3: GAM

Additionally, we compare the previous two models to the GAM indicated in (II.2) which was used in the exploratory data analysis.

Model 4: Altman

Finally, our three models are compared to the static and hazard model with Altman's variables

$$\text{logit}\,\lambda(t|\,\mathbf{x}_{it}) = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \beta_4 X_{4it} + \beta_5 X_{5it},$$

(II.10)

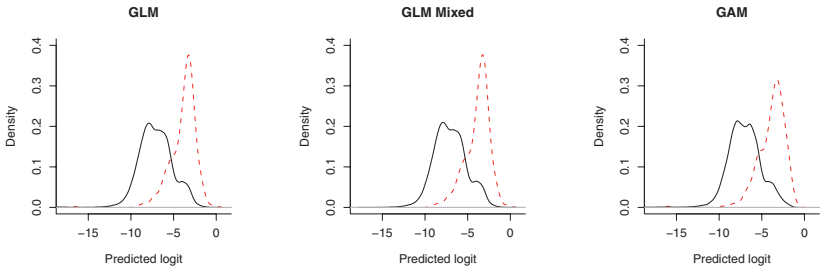


Figure II.4: Kernel density estimators of logits of predicted default probabilities for data from 2000 using GLM, GLM mixed and GAM model for bankrupt (dashed line) and solvent (solid line) firms separately (1996-1999 data used to fit the model).

where $X_1, \ldots, X_5$ are as defined in Altman (1968) [3].

II.4.3 Predictive performance

The three models, GLM, GLMM and GAM, are fitted to various subsets of our dataset, and their predictive performance is evaluated. In the assessment of model performance we use ROC and CAP curve analysis. More specifically, we use the measure of the area under the ROC curve, denoted AUC. The AUC is the area under the ROC curve and above the 45° line corresponding to the random model divided by 1/2 (the area between the ROC curve of the perfect and random model, respectively). It is a number between zero and one, one corresponding to the perfect model correctly classifying all firms and zero indicating the random model. The summary statistic of the CAP curve, the accuracy ratio AR, can be calculated directly from AUC (see Engelmann et al. (2003)).

Table II.3 shows the AUC for the models GLM, GLMM and GAM, evaluated at different fitting and prediction periods. We notice that GLM and GAM perform equally well, while GLMM has a slightly better forecasting accuracy. The lowest values of AUC are obtained when bankruptcies in 1999 were predicted. This is possibly due to the change in Norwegian law regarding sanctions for non-reporting of financial statements. Apart from evaluation of forecasts one year ahead, we have also computed AUC when bankruptcy prediction is done 2, 3 and 4 years into the future. The results are presented in the lower part of Table II.3. We observe that although the forecasting accuracy of models decline when we increase the number of forthcoming years for prediction, the power of the depreciation is not very pronounced, and the performance of the models can be considered as stable even when predicting bankruptcies several years into the future.

In order to illustrate the discriminatory power of our models we estimate them using the data 1996-1999, predict default probabilities for companies at risk in 2000, and plot kernel density estimates of their logits for bankrupt and solvent firms, respectively. Results are presented in Figure II.4. We note that all models have relatively high discriminatory power, and conclude that although plots obtained are suitable for illustration purposes, one needs more sensitive tools to decide which model has the highest forecasting accuracy.

[3] To construct the variables X_2 and X_4 we use the book value instead of market value of equity, since market information is unavailable.

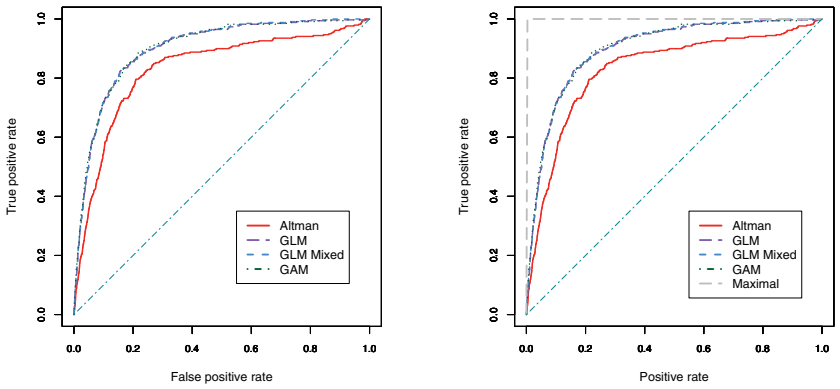


Figure II.5: ROC (left) and CAP (right) curve of GLM model, GLM with random effects, GAM model and model with Altman's variables for 1-year default-horizont out-of-time prediction (1996-1999 data used to fit the model, validation on 2000 data).

We then compare the GLM, GLMM and GAM to the model with Altman's variables. Both the static and hazard models were fitted using the data from 1996-1999, and validated on the 2000 data. The corresponding ROC and CAP curves are given in Figure II.5, and the AUC's are given in Table II.4. We observe that the forecasting accuracy of the model with Altman's variables, in both the static- and hazard setting, is lower than the corresponding accuracy of models that include non-linear transformations of covariates. Improvements obtained by utilizing the hazard instead of the static model are not pronounced, possibly due to the fact that firms in our sample are observed only for 5 years.

II.5 Conclusion and discussion

This paper presents an empirical investigation of bankruptcy prediction using the GLM, GLMM and GAM, in both the static- and hazard setting. Construction of a proper default prediction model is of crucial importance to practitioners. Potential applications include credit risk analysis, development of investment guidelines and rating methodologies, among others.

We develop empirical bankruptcy prediction models for the limited liability sector in Norway over the period 1996-2000 using annual balance sheet information. Application of non-linear modeling techniques allow us to depict complex relationships between the hazard rate of a firm at risk and its time-varying covariates. The structure of the relationship is estimated using a GAM. The final GLM (II.3) was constructed after a careful visual inspection of the plots obtained in the exploratory data analysis. Further, the unobserved heterogeneity was taken into account by including a random industry-specific intercept into the model. We utilized the AUC to compare models. While GLM and GAM perform equally well, the GLMM is shown to have slightly higher ability to anticipate problematic

Table II.3: Area under the ROC curve for GLM, GLM mixed and GAM models

Prediction 1 year ahead			
	Prediction in 2000		
Data used	*GLM*	*GLM Mixed*	*GAM*
96, 97, 98, 99	0.899	0.901	0.900
97, 98, 99	0.899	0.901	0.901
98, 99	0.894	0.900	0.897
99	0.889	0.896	0.893
	Prediction in 1999		
Data used	*GLM*	*GLM Mixed*	*GAM*
96, 97, 98	0.891	0.891	0.892
97, 98	0.890	0.891	0.892
98	0.891	0.891	0.893
	Prediction in 1998		
Data used	*GLM*	*GLM Mixed*	*GAM*
96, 97	0.897	0.905	0.898
97	0.894	0.902	0.894
	Prediction in 1997		
Data used	*GLM*	*GLM Mixed*	*GAM*
96	0.915	0.918	0.915
Prediction 2 years ahead			
	Prediction in 1999 and 2000		
Data used	*GLM*	*GLM Mixed*	*GAM*
96, 97, 98	0.894	0.895	0.894
97, 98	0.893	0.894	0.895
98	0.892	0.894	0.894
Prediction 3 years ahead			
	Prediction in 1998, 1999 and 2000		
Data used	*GLM*	*GLM Mixed*	*GAM*
96, 97	0.891	0.898	0.890
97	0.888	0.896	0.887
Prediction 4 years ahead			
	Prediction in 1997, 1998, 1999 and 2000		
Data used	*GLM*	*GLM Mixed*	*GAM*
96	0.896	0.903	0.895

Table II.4: Area under the ROC curve for model with Altman's variables, GLM, GLM mixed and GAM static and hazard model.

Model	*Static*	*Hazard*
Altman	0.816	0.830
GLM	0.897	0.899
GLM mixed	0.899	0.901
GAM	0.900	0.900

Note: All models are estimated using the data from 1996 until 1999, and validated on the data from 2000.

firms. Comparisons of models' forecasting accuracy were performed over different subsets of the complete sample. Utilization of the hazard instead of static setting does not improve the models' performance substantially, probably due to the fact that the maximal period at risk for firms in our dataset is only 5 years. Additionally, the model with Altman's variables was reestimated in both the hazard and traditional static setup. The AUC for the model with Altman's variables was substantially lower then the corresponding AUC of GLM, GLMM and GAM.

Future development of issues addressed in this paper may follow numerous directions. Primarily, more refined pattern in industry-effects modeling can be introduced by including information regarding the intra-industry groupings. The assumption of independence among firms may possibly be relaxed in the presence of empirical results presented in Lang and Stulz (1992), Das et al. (2007), Duffie et al. (2006) and references therein. Finally, the appropriate treatment of firms not reporting their balance sheet information, which were excluded from our analysis, should be established.

Acknowledgements: The authors thank Associate Professor Sjur Westgaard at the Department of Industrial Economy and Technology Management, Norwegian University of Science and Technology, for providing the original data set on which the entire work is based. Rada Dakovic and Claudia Czado acknowledge the support of Deutsche Forschungsgemeinschaft (CZ 86/1-1). Daniel Berg's research is supported by the Norwegian Research Council, grant number 154079/420.

III

A copula goodness-of-fit test based on the conditional probability integral transform

Daniel Berg
University of Oslo & Norwegian Computing Center

Henrik Bakken
Norwegian Unviversity of Science and Technology

Abstract

We investigate a copula goodness-of-fit approach based on the conditional probability integral transform. The approach implicitly weights observations at corners and edges of the unit hypercube which makes it very powerful at detecting tail heaviness for large sample sizes. However, it is shown to perform rather poor for small sample sizes. We propose a generalization that allows for any weighting, making it more robust and more powerful for small sample sizes. Another weakness is that some deviations from the null hypothesis may be neglected. We show an example and propose an extension. Results from extensive Monte Carlo experiments show that our approach keeps prescribed levels well and that certain weighting schemes produce superior power for three alternative hypotheses. The margins are treated as unknown nuisance parameters and are replaced by their empirical distribution functions. A parametric bootstrap procedure is required to obtain reliable p-value estimates. Applied to daily log-returns of large cap stock portfolios the Gaussian- and one-parameter Clayton and Gumbel copulae are all strongly rejected, increasingly so for increasing dimension and sample size. The Student-t copula on the other hand, provides a good fit, indicating the presence of tail dependence in the daily log-returns of stock data.

Key words

Copula, goodness-of-fit, conditional probability integral transformation, order statistic, parametric bootstrap, Anderson-Darling

Publication details

Submitted for publication - April 2007.

III.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 for 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 modeled by a chosen family of copulae. Information criterions, such as Aikaike's Information Criterion (AIC), are commonly employed for model selection. Such pure model selection criterions do not provide us with any understanding of the size of the decision rule employed, nor its power. This means that one can not say how well the selected family of copulae fits the data. Neither can we say whether one family of copulae fits the data significantly better than another. A goodness-of-fit (gof) approach on the other hand, will provide this information.

Copula gof testing recently emerged as a challenging inferential problem and some approaches have been proposed. Genest et al. (1995) assess the fit of bivariate Archimedean copulae. Shih (1998), Glidden (1999) and Cui and Sun (2004) test the Clayton model (also referred to as the gamma frailty model in survival analysis). Breymann et al. (2003), Chen et al. (2004) and Dobrić and Schmid (2007) apply the conditional probability integral transformation (cpit) and tests for independence. Malevergne and Sornette (2003) compare the empirical distribution of the data with a χ^2-distribution using a bootstrap method, testing the Gaussian copula hypothesis for financial asset dependencies. Fermanian (2005) approximates the underlying probability density function by kernel smoothing of the empirical density. Scaillet (2006) propose a test statistic based on the integrated squared difference between kernel estimators of the copula density and the parametric copula density. Dobrić and Schmid (2005) propose a chi-squared- and a likelihood ratio test, both based on partitioning the probability space. Panchenko (2005) focuses on positive definite bilinear forms, while Genest and Rémillard (2008) compare the empirical copula function to the parametric, null hypothesis copula function. Savu and Trede (2004) and Genest et al. (2006a) assess the cumulative distribution function (cdf) of the copula function. Finally, Genest et al. (2008) propose an approach based on the cpit and the copula function.

We will consider in more detail the approach proposed by Breymann et al. (2003), based on the cpit and henceforth denoted the cpit-approach. The cpit, also known as Rosenblatt's transformation, transforms a set of dependent variables into a set of independent variables, given the multivariate distribution. Breymann et al. (2003) perform a cpit under a parametric null hypothesis copula. Then they employ a dimension reduction technique to the d-variate cpit copula and compute a univariate test statistic on the resulting univariate vector. Their dimension reduction strongly weights data along the boundaries of the cpit copula, i.e. corners and edges of the d-dimensional unit hypercube. This makes it less robust for small sample sizes. We generalize the cpit-approach to allow for any weight function in the dimension reduction. In addition, the dimension reduction is not consistent in the sense that some deviations from the null hypothesis may be neglected. We show an example and propose an extension using an additional cpit, based on order statistics. Our generalized and extended approach is henceforth denoted the cpit2-approach.

The paper is organized as follows. Section III.2 presents preliminaries. In Section III.3 we introduce copula gof testing, the cpit-approach and our cpit2-approach. In Section III.4 we present the results from an extensive Monte Carlo study under several weighting

schemes. The study visualizes the size and power of the cpit2-approach in distinguishing the Gaussian copula from the Student-t-, Clayton- and Gumbel copulae under various dimensions and sample sizes. In Section III.5 we apply the best performing weighting scheme to analyze the dependence structure of the daily log-returns of some large cap stock portfolios. Finally, Section III.6 summarizes our results and concludes.

III.2 Preliminaries

Suppose we have n samples of the d-variate vector $\mathbf{X} = (X_1, \ldots, X_d)$. This vector comes from a population with unknown margins and linking copula C. We wish to test the hypotheses that the linking copula belongs to some parametric copula family C_θ:

$$\mathcal{H}_0 : C \in \{C_\theta; \theta \in \Theta\} \quad \text{vs.} \quad \mathcal{H}_a : C \notin \{C_\theta; \theta \in \Theta\}.$$

III.2.1 Empirical Marginals

To extract the copula we transform the vector $\mathbf{X}$ into a pseudo-vector $\mathbf{Z}$, through the empirical marginal distribution functions, $\mathbf{Z}_j = (Z_{j1}, \ldots, Z_{jd}) = (\widehat{F}_1(X_{j1}), \ldots, \widehat{F}_d(X_{jd}))$, $j = 1, \ldots, n$, where

$$\widehat{F}_i(x) = \frac{1}{n+1} \sum_{j=1}^{n} I\{X_{ji} \leq x\}. \tag{III.1}$$

Equivalently, the pseudo-vector can be expressed in terms of normalized ranks,

$$\mathbf{Z}_j = (Z_{j1}, \ldots, Z_{jd}) = \left(\frac{R_{j1}}{n+1}, \ldots, \frac{R_{jd}}{n+1} \right), \; j = 1, \ldots, n. \tag{III.2}$$

Here R_{ji} is the rank of X_{ji} in $X_{1i}, \ldots, X_{ni}$.

III.2.2 Anderson-Darling test statistic

Suppose we have a random vector $\mathbf{W} = (w_1, \ldots, w_n)$ which is iid $U(0,1)^n$ and that the cdf of $\mathbf{W}$ is $F(w) = w$. The AD statistic is then defined as

$$\mathcal{T} = n \int \frac{\left\{ \hat{F}(w) - w \right\}^2}{w(1-w)} \, \mathrm{d}w, \qquad w \in [0,1].$$

The AD statistic strongly weights deviations near $w = 0$ and $w = 1$. This is justified since the experimental deviations are small here due to the constraints $\{\hat{F}(w) - F(w)\} = 0$ at $w = (0,1)$ (Aslan and Zech, 2002).

The empirical version of the AD statistic for uniform variables can be shown to be (Marsaglia and Marsaglia, 2004):

$$\widehat{\mathcal{T}} = -n - \frac{1}{n} \sum_{j=1}^{n} (2j-1) \left\{ \ln \left[\widehat{F} \left(\frac{j}{n+1} \right) \right] + \ln \left[1 - \widehat{F} \left(\frac{n+1-j}{n+1} \right) \right] \right\}, \tag{III.3}$$

where the empirical cdf, $\widehat{F}$, is given by (III.1).

III.2.3 The conditional probability integral transform

There are several probability integral transformations, see e.g. D'Agostino and Stephens (1986) for a discussion. We will consider the transformation proposed by Rosenblatt (1952). This transformation was denoted the conditional probability integral transform (cpit) by D'Agostino and Stephens (1986) and it transforms a set of dependent variables into a new set of independent $U(0,1)$ variables, given their multivariate distribution. The cpit 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.

Definition III.1 (Conditional probability integral transform)
Let $\mathbf{Z} = (Z_1, \ldots, Z_d)$ denote a random vector with marginal distributions $F_i(z_i) = P(Z_i \leq z_i)$ and conditional distributions $F_{i|1\ldots i-1}(Z_i \leq z_i | Z_1 = z_1, \ldots, Z_{i-1} = z_{i-1})$ for $i = 1, \ldots, d$. The cpit of $\mathbf{Z}$ is defined as $T(\mathbf{Z}) = (T_1(Z_1), \ldots, T_d(Z_d))$ where

$$\begin{aligned} T_1(Z_1) &= P(Z_1 \leq z_1) = F_1(z_1), \\ T_2(Z_2) &= P(Z_2 \leq z_2 | X_1 = z_1) = F_{2|1}(z_2|z_1), \\ &\vdots \\ T_d(Z_d) &= P(Z_d \leq x_d | Z_1 = z_1, \ldots, Z_{d-1} = z_{d-1}) = F_{d|1\ldots d-1}(z_d|z_1, \ldots, z_{d-1}). \end{aligned}$$

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

A recent application of the cpit is to multivariate gof tests. A cpit is applied to a data set, assuming a multivariate null distribution, and then a test of multivariate independence is carried out on the resulting, transformed data set. The null hypothesis in our setting is a parametric copula family. The parameters of this copula family needs to be estimated before applying the cpit. We shortly present parameter estimation in Section III.2.4.

An advantage with the cpit in a gof setting is that the null- and alternative hypotheses are the same, regardless of the distribution before the cpit. The cpit also enables weighting in a simple way since the data, after the cpit, is i.i.d. $U(0,1)$ under the null hypothesis. Hong and Li (2002) report Monte Carlo evidence of multivariate tests using cpit 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 cpit is the invariance with respect to the permutation of the variables since there are $d!$ possible permutations. However, as long as the permutation is decided randomly, the results will not be influenced in any particular direction. D'Agostino and Stephens (1986) discuss this issue and propose solutions for some special cases, e.g. the cpit based on ordered variables, which does not suffer from permutation invariance. We will consider this in more detail when presenting our cpit2-approach in Section III.3.2.

III.2.4 Parameter estimation

There are two main ways of estimating the parameters of a copula, the fully parametric method or a semi-parametric method. The fully parametric method, termed the inference functions for margins (IFM) method (Joe, 1997), relies on the assumption of parametric, univariate margins. First, the parameters of the margins are estimated and then each parametric margin is plugged into the copula likelihood which is then maximized with

respect to the copula parameters. Since we treat the margins as nuisance parameters, we rather proceed with the pseudo-vector $\mathbf{Z}$ and the semi-parametric method. This method is denoted the pseudo-likelihood (Demarta and McNeil, 2005) or the canonical maximum likelihood (CML) (Romano, 2002) method and is described in Genest et al. (1995) and in Shih and Louis (1995) in the presence of censorship. Having obtained the pseudo-vector $\mathbf{Z}$, using (III.2), the copula parameters can be estimated using either maximum likelihood (ML) or using the well-known relations to Kendall's tau (for a survey of copulae and their relationship with measures of association, see Nelsen (1999)).

For the elliptical copulae in higher dimensions, we pairwise invert the sample Kendall's tau. This gives the correlation- and scale matrix for the Gaussian and Student-t copulae, respectively. For the Student-t copula we also need to estimate the degrees of freedom. Genest et al. (2008) estimate the scale matrix by inversion of Kendall's tau but assume the degrees of freedom to be known/fixed. We rather follow the approach used by Mashal and Zeevi (2002) and Demarta and McNeil (2005). This is a two-stage approach in which the scale matrix is first estimated by inversion of Kendall's tau, and then the pseudo-likelihood function is maximized with respect to the degrees of freedom, given the estimate of the scale matrix. For the Archimedean copulae, we consider the so-called exchangeable construction with one dependency parameter. We estimate this parameter by numerically maximizing the pseudo-likelihood.

III.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. Economic theory sheds little light on the dependence structure between financial assets, and multivariate normality is often assumed a priori. Evidence shows, however, that more appropriate dependence structures are available (Chen et al., 2004; Dobrić and Schmid, 2005).

Several approaches (e.g. Breymann et al. (2003); Genest et al. (2006a)) project the multivariate problem to a univariate problem applying some dimension reduction technique and then compute a univariate test statistic. This leads to numerically efficient algorithms even for problems of high dimension. Any univariate statistic may be used, e.g. Kolmogorov-Smirnov, Anderson-Darling, Cramér-von Mises or kernel smoothing based L2 statistics. For a thorough treatment of these and other statistics we refer to D'Agostino and Stephens (1986). In this paper we focus on the Anderson-Darling (AD) statistic.

For copula gof testing we are interested in the fit of the copula alone, hence the margins are commonly treated as nuisance parameters. I.e. we use empirical margins (or equivalently, normalized ranks). The use of empirical margins will alter the asymptotics of any test statistic. In addition, since we are testing a hypothesized, parametric, copula, parameter estimation error will influence the asymptotics. Breymann et al. (2003) fail to recognize these issues. They assume that the limiting distribution of their statistic is the same whether the margins and parameters are estimated or not. As a result, the p-values that they report are not correct. This erroneous assumption is pointed out by Genest and Rémillard (2008). It is also thoroughly investigated by Dobrić and Schmid (2007) who modify the test procedure by Breymann et al. (2003) such that the p-value estimates become reliable. Henceforth, when referring to the approach by Breymann et al. (2003), the cpit-approach, we mean the approach proposed by Breymann et al. (2003) but using

the test procedure of Dobrić and Schmid (2007).

III.3.1 The cpit-approach

The approach proposed by Breymann et al. (2003) is a generalization of the approach proposed by Malevergne and Sornette (2003) which corresponds to the special case of testing the Gaussian copula null hypothesis. It is a dimension reduction approach and we will denote the test observator by G.

The testing is based on the pseudo-vector $\mathbf{Z}$, see (III.2). A cpit is applied to $\mathbf{Z}$, assuming a null hypothesis copula C_θ. The d-variate vector $\mathbf{V} = (V_1, \ldots, V_d)$, resulting from the cpit, is i.i.d. $U(0,1)^d$ under the null hypothesis. Due to parameter- and margin estimation errors, this is only close to, but not exactly true. We will consider this issue in Section III.3.3. Until then we assume that this holds. The dimension reduction is now performed as

$$W_G = \sum_{i=1}^{d} \Phi^{-1}(V_i)^2. \tag{III.4}$$

The variable W_G should, under the null hypothesis, be χ^2_d distributed. The test observator G can now be defined.

Definition III.2 (Cpit test observator G)
Let W_G be defined by (III.4) and $F_{\chi^2_d}(\cdot)$ be the χ^2_d cdf. $G(w)$ is then defined as the cdf of $F_{\chi^2_d}(W_G)$:

$$G(w) = P[F_{\chi^2_d}(W_G) \leq w].$$

Under $\mathcal{H}_0$, all V_i are i.i.d. $U(0,1)$, hence $G(w) = w$ and the density of $G(w)$ is $g(w) = 1$.

Suppose we have n samples of $\mathbf{V}$, $\mathbf{v}_j = (v_{j1}, \ldots, v_{jd}),\ j = 1, \ldots, n$. After performing the dimension reduction in (III.4), we have n samples of W_G. The empirical version of the approach then becomes

$$\widehat{G}(w) = \frac{1}{n+1} \sum_{j=1}^{n} I\{F_{\chi^2_d}(W_{G,j}) \leq w\}, \qquad w = \frac{1}{n+1}, \ldots, \frac{n}{n+1}.$$

In the cpit-approach $\widehat{G}(w)$ is plugged in for $\widehat{F}(w)$ in the expression for the AD statistic (III.3).

The cpit-approach is computationally very efficient and conceptually simple. However, it has its weaknesses. First of all, the dimension reduction, through the use of $\Phi^{-1}(\cdot)^2$, strongly weights data along the boundaries of the d-dimensional unit hypercube. This may be appropriate when the sample size is large. However, for small sample sizes, this weighting makes the approach less robust and less powerful since there will be few observations in the boundary regions. We will see the effects of this in Section III.4. In addition, some deviations from the null hypothesis may be overlooked by the cpit-approach. Figure III.1 shows three constructed bivariate data sets, one that is independent in the left panel and two that are clearly dependent in the center- and right panels. Recall the null hypothesis of independence. We thus wish for the lack of independence in these panels to be detected. However, $G(w)$ will be exactly the same for all three data sets. The explanation is that a value of 0.2 and a value of 0.8 will both contribute with the exact same value to W_G, since $\Phi^{-1}(0.2)^2 = \Phi^{-1}(0.8)^2$. Hence, we suspect the approach to perform poor in cases where the cpit data set is radially asymmetric.

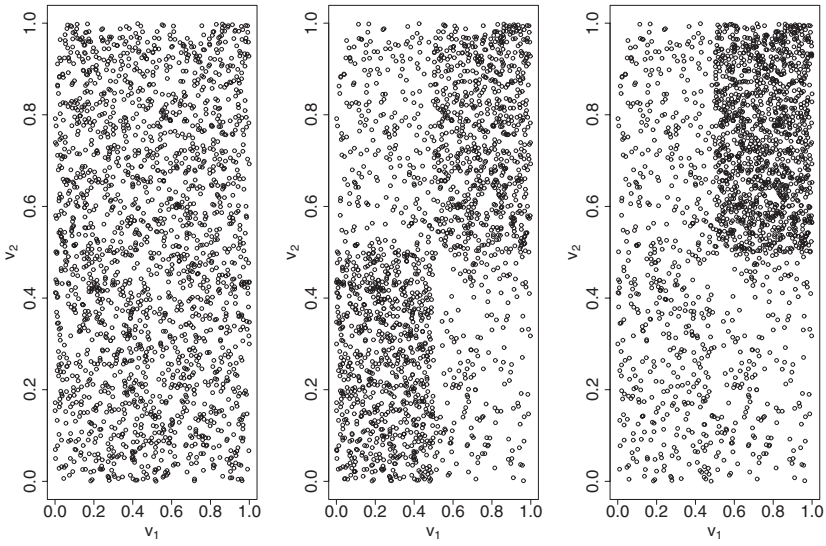


Figure III.1: Three $U(0,1)^2$ data sets, one that is independent (left panel) and two that are clearly dependent (center- and right panels). $\widehat{G}(w)$ is equivalent for all three.

(III.7)

III.3.2 The cpit2-approach: a generalization and extension

With the weaknesses of the cpit-approach in mind, we propose a new approach, denoted the cpit2-approach. This approach generalizes and extends the cpit-approach. First, any weight function can be employed in the dimension reduction (III.4). In addition, through the use of an additional cpit, based on order statistics, we are able to detect radial asymmetry in the cpit data as illustrated in Figure III.1.

We interpret the problem of multivariate gof testing as follows. To perform a gof assessment of a multivariate data set we can essentially perform two univariate gof tests. First, we test the fit in the d-space, e.g. through some dimension reduction technique such as (III.4). The result of this first, univariate, gof test is n values of a test statistic. If we know the distribution of these n samples, under the null hypothesis, we can perform another, univariate, gof test in the n-space. This will give us the desired test statistic for the multivariate problem. In what follows we are mainly concerned with the first gof test, in the d-space. For the second, in the n-space, we use the Anderson-Darling statistic, as in the cpit-approach.

We first perform the cpit on the original copula data set $\mathbf{Z}$. The resulting data $\mathbf{V}$ should be i.i.d. $U(0,1)^d$ under the null hypothesis. We now propose to test whether this is true, i.e. an additional test in the d-space, testing for independent uniformity of $\mathbf{V}$. This problem is well known and is discussed in great detail in D'Agostino and Stephens (1986). They suggest a cpit, based on ordered variables, that will be permutation invariant. Thus, we perform a regular cpit first, on $\mathbf{Z}$, and then a second cpit on $\mathbf{V}$ which is based on the order statistics of $\mathbf{V}$.

As before, let $\mathbf{V} = (V_1, \ldots, V_d)$ be the i.i.d. $U(0,1)^d$ random vector, obtained from applying the cpit to $\mathbf{Z}$. For $d = 1, 2, \ldots$, we denote the order statistics of $V_1, \ldots, V_d$ by

$$V_{(1)} \leq V_{(2)} \leq \cdots \leq V_{(d-1)} \leq V_{(d)}.$$

If $V_{(1)}, \ldots, V_{(d)}$ are the order statistics of a sample from a $U(0,1)$ parent distribution, then $V_{(i)}$ is a beta distributed variable with parameters $(i, d-(i-1))$ (D'Agostino and Stephens, 1986, ch. 8). To compute the expressions for the order statistic cpit, we resort to David (1981, Theorem 2.7) who shows the Markov nature of the order statistics. Using Deheuvels (1984, Theorem 1) and the fact that $\mathbf{V}$ is an i.i.d. $U(0,1)^d$ random vector under the null hypothesis, we obtain the following expression for the order statistic cpit of V:

$$H_i = F_{V_{(i)}|V_{(i-1)}}(v_{(i)}) = 1 - \left(\frac{1 - v_{(i)}}{1 - v_{(i-1)}}\right)^{d-(i-1)}, \; i = 1, \ldots, d, \; v_{(0)} = 0. \tag{III.5}$$

Intuitively, poor fit in the d-space is indicated by extreme values of H. Any H too low or too high can indicate a poor fit (Glen et al., 2001). We can now conduct the dimension reduction based on $\mathbf{V}$ and $\mathbf{H}$:

$$W_B = \sum_{i=1}^{d} \Gamma_V(V_{(i)}; \boldsymbol{\alpha}) \cdot \Gamma_H(H_i; \boldsymbol{\alpha}), \tag{III.6}$$

where Γ_V and Γ_H are weight functions used for weighting the information in V and H, respectively, and $\boldsymbol{\alpha}$ is the set of weight parameters. Any weight function may be used, depending on the use and the region of the copula one wishes to emphasize. Some obvious candidates for both $\Gamma_V(X; \boldsymbol{\alpha})$ and $\Gamma_H(X; \boldsymbol{\alpha})$ are:

(III.8)

(i) $\Phi^{-1}(X)^2$,

(ii) $|X - 0.5|$,

(iii) $(X - 0.5)^\alpha,\ \alpha = (2, 4, \ldots)$.

Consider for example the special case $\Gamma_V(X; \boldsymbol{\alpha}) = \Phi^{-1}(X)^2$ and $\Gamma_H(X; \boldsymbol{\alpha}) = 1$. We then obtain (III.4), the cpit-approach. Since both $\mathbf{V}$ and $\mathbf{H}$ are i.i.d. $U(0,1)^d$ under the null hypothesis we have the following result. By choosing $\Gamma_V(X; \boldsymbol{\alpha}) = 1$ and $\Gamma_H(X; \boldsymbol{\alpha}) = \Phi^{-1}(X)^2$, W_B in (III.6), as for (III.4), should follow a χ^2_d distribution under the null hypothesis. However, in general, the distribution of W_B is not known and we must turn to a double bootstrap to approximate the cdf. Suppose we have computed W_B, using some weight functions $\Gamma_V(\cdot; \boldsymbol{\alpha})$ and $\Gamma_H(\cdot; \boldsymbol{\alpha})$. Now we simply draw d i.i.d. $U(0,1)$ variables $\widetilde{V}$, compute $\widetilde{H}$ and $\widetilde{W}_B$ using the same weight functions as for W_B. By repeating this a large number of times (10000 times in this paper), we can approximate the cdf of W_B, F_B, under the null hypothesis. Again, as for W_G, W_B is only close to, but not exactly distributed according to F_B. This discussion is deferred to Section III.3.3. Our new test observator B can now be defined.

Definition III.3 (Cpit2 test observator B)
Let W_B be defined by (III.6) and $F_B(\cdot)$ be the cdf of W_B. $B(w)$ is then defined as the cdf of $F_B(W_B)$:

$$B(w) = P[F_B(W_B) \leq w].$$

Under $\mathcal{H}_0$, all V_i are i.i.d. $U(0,1)$, hence $B(w) = w$ and the density of $B(w)$, $b(w) = 1$.

Suppose we have n samples of $\mathbf{V}$, $\mathbf{v}_j = (v_{j1}, \ldots, v_{jd})$, $j = 1, \ldots, n$. The empirical version then becomes

$$\widehat{B}(w) = \frac{1}{n+1} \sum_{j=1}^{n} I\{F_B(W_{B,j}) \leq w\}, \qquad w = \frac{1}{n+1}, \ldots, \frac{n}{n+1}, \tag{III.7}$$

which can be plugged in for $\widehat{F}(w)$ in the expression for the AD statistic (III.3).

To summarize our cpit2-approach, we have performed two cpit's, the first to $\mathbf{Z}$ and the second to the order statistics of $\mathbf{V}$. By doing this additional order statistic cpit, our dimension reduction approach becomes more robust to phenomena like the one in Figure III.1. For the data sets in Figure III.1 we obtain $G(w) = (0.91, 0.91, 0.91)$ for the left-, center- and right panels, respectively, using the AD statistic. With the cpit2-approach, using $\Gamma_V(X; \boldsymbol{\alpha}) = 1$ and $\Gamma_H(X; \boldsymbol{\alpha}) = \Phi^{-1}(X)^2$, we obtain $B(w) = (0.36, 66.46, 111.62)$. We clearly see that our extension H detects the asymmetry. Figures III.2-III.4 show W_B as a surface, with respect to the cpit data set $\mathbf{V}$ and we see that the cpit-approach heavily emphasizes the boundaries. The use of the cpit2-approach, with weight combination (ii) still emphasizes these regions but less extremely. We also see that the Γ_H term adds weight to the diagonal as well. This is why this extension will help detect radial asymmetry in the cpit data. Finally, the generalization adds flexibility and robustness to small sample sizes. Both weight functions for the dimension reduction, Γ_V and Γ_H, can be decided freely, depending on the specific use.

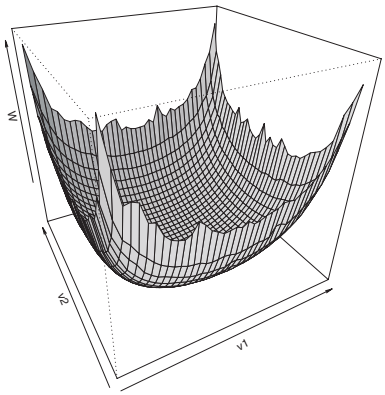


Figure III.2: Weight (W_B) surface with respect to cpit data, $\mathbf{V}$. Data generating process is the Gaussian copula ($\rho = 0.71$, corresponding to a Kendall's tau of 0.5) and we perform the cpit assuming the true null hypothesis of a Gaussian copula. Weight combination: $\Gamma_V(X;\boldsymbol{\alpha}) = \Phi^{-1}(X),\ \Gamma_H(X;\boldsymbol{\alpha}) = 1$.

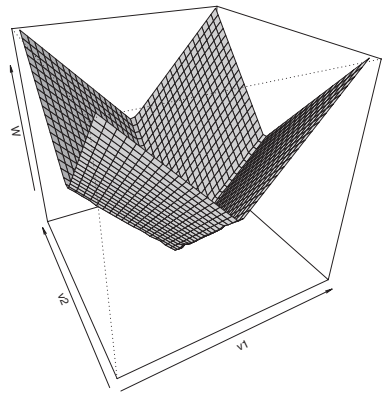


Figure III.3: Weight (W_B) surface with respect to cpit data, $\mathbf{V}$. Data generating process is the Gaussian copula ($\rho = 0.71$, corresponding to a Kendall's tau of 0.5) and we perform the cpit assuming the true null hypothesis of a Gaussian copula. Weight combination: $\Gamma_V(X;\boldsymbol{\alpha}) = |X - 0.5|,\ \Gamma_H(X;\boldsymbol{\alpha}) = 1$.

(III.10)

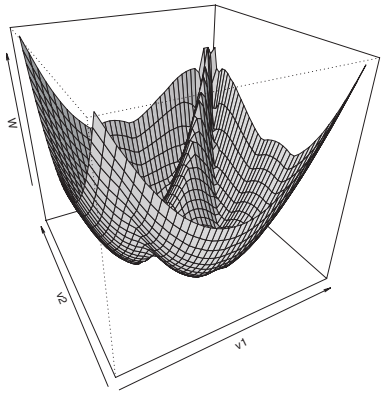


Figure III.4: Weight (W_B) surface with respect to cpit data, $\mathbf{V}$. Data generating process is the Gaussian copula ($\rho = 0.71$, corresponding to a Kendall's tau of 0.5) and we perform the cpit assuming the true null hypothesis of a Gaussian copula. Weight combination: $\Gamma_V(X;\boldsymbol{\alpha}) = |X - 0.5|,\ \Gamma_H(X;\boldsymbol{\alpha}) = |X - 0.5|$.

III.3.3 Testing Procedure

In Section III.3.1 we assume that W_G in (III.4) follow a χ^2_d distribution. Similarly, in Section III.3.2 we assume that the distribution of W_B in (III.6) can be approximated by a double bootstrap procedure. The estimation of the margins and the parameters of the null copula, introduces dependence in the cpit data. Hence, W_G is only close to, but not exactly χ^2_d distributed. Similarly, W_B is only close to, but not exactly distributed according to F_B.

To cope with this issue and obtain a proper estimate of the p-value of the observed statistic, one should perform a parametric bootstrap procedure, where both margin- and parameter estimation effects are accounted for. We adopt the parametric bootstrap procedure used in Genest et al. (2006a), the validity of which is established in Genest and Rémillard (2008). Dobrić and Schmid (2007) propose a very similar procedure in their modification of the original procedure used in Breymann et al. (2003). The asymptotic validity of the bootstrap procedure, applied to our test observator, has not yet been proved. However, our numerical results in Section III.4 strongly indicates that the procedure is valid.

Suppose we have a sample $\mathbf{x}$, n observations of the d-variate vector $\mathbf{X}$. The testing procedure for the cpit2-approach is then as given below. Remember that the cpit-approach is a special case, hence the same test procedure can be applied for this approach. Note the use of empirical margins (normalized ranks) in step (1) and (9a), the parametric bootstrap procedure in step (9) and the double bootstrap procedure in steps (6) and (9f).

(1) Extract the pseudo-observations $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ by converting the sample data $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ into normalized ranks according to (III.2).

(2) Estimate the parameters Θ of the null hypothesis copula, by a consistent estimator $\widehat{\Theta} = \widehat{\mathcal{L}}(\mathbf{z}_1, \ldots, \mathbf{z}_n)$.

(3) Compute the cpit sample data $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ by applying the cpit to $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ assuming the parametric null hypothesis copula $C_{\widehat{\Theta}}$.

(4) Compute the cpit2 sample data $(\mathbf{h}_1, \ldots, \mathbf{h}_n)$ by applying the cpit (III.5) to $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$.

(5) Compute W_B according to (III.6), using weight functions Γ_V and Γ_H.

(6) If W_B follows a known distribution under the null hypothesis, compute $F_B(W_B)$ accordingly. If not, approximate F_B as follows. For some large integer m, repeat the following steps for every $l \in \{1, \ldots, m\}$:

   (i) Generate a random sample $(v^*_{1,l}, \ldots, v^*_{d,l})$ from the null hypothesis copula, namely an i.i.d. $U(0,1)^d$ vector.

   (ii) Compute $(h^*_{1,l}, \ldots, h^*_{d,l})$ by applying the cpit (III.5) to $(v^*_{1,l}, \ldots, v^*_{d,l})$.

   (iii) Compute $\widehat{W}^*_{B,l}$ according to (III.6) using the same weight functions as in step (5).

   (iv) Compute $F_B(W) = \frac{1}{m+1} \sum_{l=1}^{m} I(\widehat{W}'_{B,l} > \widehat{W}_B)$.

(7) Compute $\widehat{B}(w)$ according to (III.7).

(8) The estimated AD statistic $\widehat{\mathcal{T}}$ is obtained by plugging $\widehat{B}(w)$ into (III.3).

(9) For some large integer N_b, repeat the following steps for every $k \in \{1, \ldots, N_b\}$:

   (a) Generate a random sample $(\mathbf{x}^*_{1,k}, \ldots, \mathbf{x}^*_{n,k})$ from the null hypothesis copula $C_{\widehat{\Theta}}$ and compute the associated pseudo-samples $(\mathbf{z}^*_{1,k}, \ldots, \mathbf{z}^*_{n,k})$ according to (III.2).

   (b) Estimate the parameters Θ, of the null hypothesis copula, with $\widehat{\Theta}^*_k = \widehat{\mathcal{L}}(\mathbf{z}^*_{1,k}, \ldots, \mathbf{z}^*_{n,k})$.

   (c) Compute the cpit sample data $(\mathbf{v}^*_{1,k}, \ldots, \mathbf{v}^*_{n,k})$ by applying the cpit to $(\mathbf{z}^*_{1,k}, \ldots, \mathbf{z}^*_{n,k})$, assuming the parametric null hypothesis copula $C_{\widehat{\Theta}^*_k}$.

   (d) Compute the cpit2 sample data $(\mathbf{h}^*_{1,k}, \ldots, \mathbf{h}^*_{n,k})$ by applying the cpit (III.5) to $(\mathbf{v}^*_{1,k}, \ldots, \mathbf{v}^*_{n,k})$.

   (e) Compute $W^*_{B,k}$ according to (III.6), using the same weight functions as in step (5).

   (f) If $W^*_{B,k}$ follows a known distribution, compute $F_B(W_B)$ accordingly. If not, approximate F_B as follows. For some large integer m, repeat the following steps for every $l \in \{1, \ldots, m\}$:

      (i) Generate a random sample $(v^{**}_{1,l,k}, \ldots, v^{**}_{d,l,k})$ from the null copula, an i.i.d. $U(0,1)^d$ vector.

      (ii) Compute $(h^{**}_{1,l,k}, \ldots, h^{**}_{d,l,k})$ by applying the cpit (III.5) to $(v^{**}_{1,l,k}, \ldots, v^{**}_{d,l,k})$

      (iii) Compute $\widehat{W}^{**}_{B,l,k}$ according to (III.6) using the same weight functions as in step (5).

      (iv) Compute $F_B(W^{**}_B) = \frac{1}{m+1} \sum_{k=1}^{m} I(\widehat{W}^{**}_{B,l,k} > \widehat{W}^*_{B,k})$.

   (g) Compute $\widehat{B}^*_k(w)$ according to (III.7).

(III.12)

(h) The estimated AD statistic $\widehat{\mathcal{T}}_k^*$ is obtained by plugging $\widehat{B}_k^*(w)$ into (III.3).

(10) An approximate p-value for the cpit2 test observator B is then given by

$$\widehat{p} = \frac{1}{N_b + 1} \sum_{k=1}^{N_b} I(\widehat{\mathcal{T}}_k^* > \widehat{\mathcal{T}}).$$

Steps $6 - 9$ may seem abundant and arbitrary. We could have used the W_B's directly and performed some test of its distribution. However, the distribution of W_B is usually not known and numerical- or simulation procedures are needed to approximate F_B.

III.4 Monte Carlo study

By performing so-called mixing tests we examine the ability of the cpit2-approach to keep nominal sizes and detect tail heaviness and skewness properties. The tests are performed by mixing a Gaussian copula 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. For $\beta = 0$, $C^{Mix} = C^{Ga}$ while for $\beta = 1$, $C^{Mix} = C^{Alt}$. For $0 < \beta < 1$ we sample from the Gaussian copula with probability $(1 - \beta)$ and from the alternative copula with probability β.

The alternative copulae considered in this paper are the Student-t-, Clayton- and Gumbel copulae. The ability to distinguish the Gaussian from the Student-t copula indicates the power at detecting lower and upper tail dependency, while the ability to distinguish the Gaussian from the Clayton- and Gumbel copulae indicates the power at detecting lower and upper tail dependency, respectively. For all copulae, the dependency parameter is set to correspond to a Kendall's tau of 0.2, i.e. a weak level of dependence. This should make the various copulae hard to distinguish. For the Student-t copula, the degree of freedom ν, is set to 4, i.e. very heavy tails. For the Gaussian copula, the upper and lower tail dependencies are both 0 while for the Student-t copula the lower and upper tail dependencies, for a Kendall's tau of 0.2, both equal 0.17. For the Clayton copula the lower- and upper tail dependencies equal 4 and 0, respectively. Finally, for the Gumbel copula, the lower- and upper tail dependencies equal 0 and 0.26, respectively. See Nelsen (1999) for the definition of tail dependency.

For the cpit-approach we examined all possible combinations of the weight functions $\Gamma_V(X; \boldsymbol{\alpha})$ and $\Gamma_H(X; \boldsymbol{\alpha})$, listed in Section III.3.2, namely $\Phi^{-1}(X)^2$, $|X - 0.5|$, $(X - 0.5)^\alpha$, $\alpha = (2, 8, 20)$. Note again that the cpit-approach is a special case of the cpit2-approach, with $\Gamma_V(X; \boldsymbol{\alpha}) = \Phi^{-1}(X)^2$ and $\Gamma_H(X; \boldsymbol{\alpha}) = 1$.

Our null hypothesis is that the mixed copula is a Gaussian copula. $\widehat{\mathcal{T}}$ and the corresponding estimate of the p-value is computed according to the test procedure in Section III.3.3, using $N_b = 500$ for the parametric bootstrap and $m = 10000$ for the double bootstrap. The entire procedure is repeated $N_{mix} = 2000$ times in order to obtain rejection rates and corresponding power curves. The resulting rejection rates for the best performing weight combinations (at $\beta = 1$), are given in Tables III.1-III.3. The weight combination corresponding to the cpit-approach is also included for comparison, although it did not perform very well compared to other combinations.

First, we examine the effect of dimension and sample size. For all combinations of Γ_V and Γ_H, the power increases with dimension and sample size, as expected and visualized in Figure III.5. We next examine the nominal levels, i.e. the rejection rates for $\beta = 0$, and they all roughly match the prescribed level of 5%. This indicates the validity of our bootstrap procedure. Finally, for all combinations of $d = (2, 5)$, $n = (125, 250, 500)$ and C^{Alt} =(Student-t, Clayton, Gumbel), we examine the power. The best combinations varies with dimension, sample size and whether we consider lower-, upper- or both lower and upper tail dependency. All over, the combinations

(i) $\Gamma_V(X; \boldsymbol{\alpha}) = |X - 0.5|,\ \Gamma_H(X; \boldsymbol{\alpha}) = 1$ and

(ii) $\Gamma_V(X; \boldsymbol{\alpha}) = (X - 0.5)^2,\ \Gamma_H(X; \boldsymbol{\alpha}) = 1$

stand out as superior. I.e. combinations where we only consider the cpit data $\mathbf{V}$, however with a different weight function than the one in the cpit-approach. When the alternative copula is the Student-t copula, the combinations

(iii) $\Gamma_V(X; \boldsymbol{\alpha}) = (X - 0.5)^8,\ \Gamma_H(X; \boldsymbol{\alpha}) = 1$ and

(iv) $\Gamma_V(X; \boldsymbol{\alpha}) = (X - 0.5)^8,\ \Gamma_H(X; \boldsymbol{\alpha}) = |X - 0.5|$

perform equally well as combinations (i) and (ii). For this case, the cpit-approach also perform quite well. If the alternative copula is either the Clayton- or the Gumbel copula, combinations (i) and (ii) are, by far, the best. In addition, the combinations

(v) $\Gamma_V(X; \boldsymbol{\alpha}) = 1,\ \Gamma_H(X; \boldsymbol{\alpha}) = \Phi^{-1}(X)^2$,

(vi) $\Gamma_V(X; \boldsymbol{\alpha}) = 1,\ \Gamma_H(X; \boldsymbol{\alpha}) = (X - 0.5)^\alpha,\ \alpha = (2, 8, 20)$ and

(vii) $\Gamma_V(X; \boldsymbol{\alpha}) = |X - 0.5|,\ \Gamma_H(X; \boldsymbol{\alpha}) = |X - 0.5|$

perform quite well.

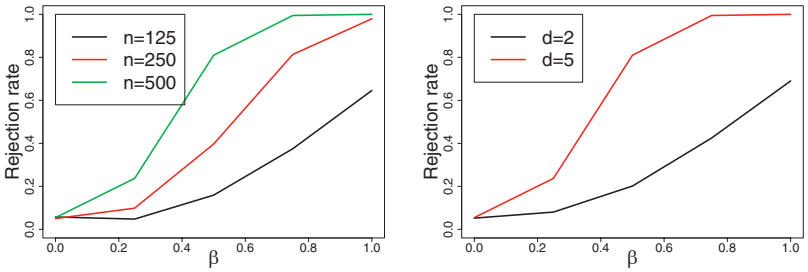
Consider the particular case where we have very few samples, i.e. $n = 125$, and high dimension, i.e. $d = 5$ and the alternative copula is the Clayton copula. In this case the Γ_H term adds power compared to combinations only includingthe Γ_V term. However, for large sample sizes (i) and (ii) are superior.

Figure III.6 illustrates the difference in power for some combinations. We see that the cpit-approach ($\Gamma_V(X; \boldsymbol{\alpha}) = \Phi^{-1}(X)^2,\ \Gamma_H(X; \boldsymbol{\alpha}) = 1$) has quite low power in some cases, while the cpit2-approach with weight combinations (i), (ii) and several other combinations perform better.

III.5 Application

The choice of dependency structure can have big impacts in several applications, e.g. capital allocation and the pricing of credit derivatives, such as basket default swaps. We analyze the dependency structure of stock portfolios by looking at their daily log-returns. The historical data consists of 1000 samples of 45 large cap stocks from the New York Stock Exchange, spanning the period from January 13th, 2003 to December 29th, 2006.

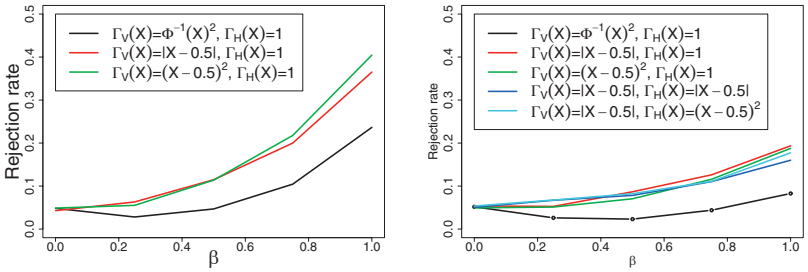
Asset collections of dimension 2 and 5 were randomly selected 2000 times from the full data set. As in Chen et al. (2004) and Panchenko (2005) we examine the raw returns and the GARCH(1, 1) filtered returns, i.e. each individual assets return is filtered through a



(a) Effect of n, the sample size. $C^{Alt} = C^{t_4}$, $\Gamma_V(x) = |x - 0.5|$, $\Gamma_H(x) = 1$, $d = 5$.

(b) Effect of d, the dimension. $C^{Alt} = C^{t_4}$, $\Gamma_V(x) = |x - 0.5|$, $\Gamma_H(x) = 1$, $n = 500$.

Figure III.5: Power curves for the approach B, for varying sample sizes and dimension. 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 null copula) is rejected. 5% significance level.



(a) $d = 2$, $n = 250$, C^{Alt} is Student-t.

(b) $d = 5$, $n = 500$, C^{Alt} is Clayton.

Figure III.6: Power curves for the approach B, comparing various weight combinations (Γ_V, Γ_H). 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 null copula) is rejected. 5% significance level.

Table III.1: Percentage rejection (5% level) of the Gaussian copula under various dimensions, sample sizes and weight combinations.

d	n	$\Gamma_V(X),\ \Gamma_H(X)$	β				
			0	0.25	0.5	0.75	1
2	125	$\Phi^{-1}(X)^2,\ 1$	4.93	2.71	2.22	3.00	6.06
		$\lvert X-0.5\rvert,\ 1$	5.67	4.93	7.24	8.72	16.45
		$(X-0.5)^2,\ 1$	5.67	4.24	6.95	9.51	16.16
		$(X-0.5)^8,\ 1$	4.63	2.86	2.51	3.3	7.34
		$(X-0.5)^8,\ \lvert X-0.5\rvert$	5.17	3.00	1.92	2.86	6.31
	250	$\Phi^{-1}(X)^2,\ 1$	4.83	2.81	4.68	10.44	23.65
		$\lvert X-0.5\rvert,\ 1$	4.29	6.31	11.48	20	36.50
		$(X-0.5)^2,\ 1$	4.88	5.52	11.38	21.77	40.44
		$(X-0.5)^8,\ 1$	4.53	3.30	4.88	12.81	28.62
		$(X-0.5)^8,\ \lvert X-0.5\rvert$	5.02	3.10	3.84	11.97	26.11
	500	$\Phi^{-1}(X)^2,\ 1$	5.42	4.73	15.12	38.33	67.88
		$\lvert X-0.5\rvert,\ 1$	5.22	8.03	20.15	42.46	69.01
		$(X-0.5)^2,\ 1$	5.12	7.93	21.72	48.92	77.00
		$(X-0.5)^8,\ 1$	5.22	4.83	15.07	38.33	66.6
		$(X-0.5)^8,\ \lvert X-0.5\rvert$	5.57	4.24	13.84	34.19	64.19
5	125	$\Phi^{-1}(X)^2,\ 1$	4.58	1.13	12.71	42.96	79.85
		$\lvert X-0.5\rvert,\ 1$	5.76	4.78	15.91	37.49	64.58
		$(X-0.5)^2,\ 1$	5.81	5.91	22.41	52.51	81.92
		$(X-0.5)^8,\ 1$	4.48	3.10	20.49	56.55	87.83
		$(X-0.5)^8,\ \lvert X-0.5\rvert$	4.53	1.87	14.53	45.96	82.07
	250	$\Phi^{-1}(X)^2,\ 1$	4.48	6.06	52.41	95.07	100.00
		$\lvert X-0.5\rvert,\ 1$	5.02	9.85	39.70	81.33	97.93
		$(X-0.5)^2,\ 1$	4.63	13.00	56.6	93.94	99.85
		$(X-0.5)^8,\ 1$	5.57	10.94	63.15	96.70	100.00
		$(X-0.5)^8,\ \lvert X-0.5\rvert$	4.93	8.67	53.55	93.69	99.9
	500	$\Phi^{-1}(X)^2,\ 1$	5.57	26.9	95.86	100.00	100.00
		$\lvert X-0.5\rvert,\ 1$	5.47	23.69	80.99	99.46	100.00
		$(X-0.5)^2,\ 1$	5.52	32.32	93.45	100.00	100.00
		$(X-0.5)^8,\ 1$	5.62	36.16	96.75	100.00	100.00
		$(X-0.5)^8,\ \lvert X-0.5\rvert$	4.88	29.01	94.19	100.00	100.00

Note: The alternative hypothesis is the Student-t copula with 4 degrees of freedom. The dependency parameter corresponds to a Kendall's tau of 0.2.

Table III.2: Percentage rejection (5% level) of the Gaussian copula under various dimensions, sample sizes and weight combinations.

d	n	$\Gamma_V(X),\ \Gamma_H(X)$	β				
			0	0.25	0.5	0.75	1
2	125	$\lvert X-0.5\rvert,\ 1$	5.32	4.93	4.68	4.78	6.40
		$(X-0.5)^2,\ 1$	5.62	5.32	4.58	4.58	5.12
		$\lvert X-0.5\rvert,\ \lvert X-0.5\rvert$	5.67	5.02	4.43	4.29	4.48
		$\lvert X-0.5\rvert,\ (X-0.5)^2$	5.32	5.22	4.98	4.38	5.12
		$1,\ \lvert X-0.5\rvert$	5.76	6.06	6.45	5.71	6.40
	250	$\lvert X-0.5\rvert,\ 1$	5.37	5.02	6.11	6.55	6.75
		$(X-0.5)^2,\ 1$	5.62	4.73	5.17	5.17	5.57
		$\lvert X-0.5\rvert,\ \lvert X-0.5\rvert$	4.48	6.06	5.27	4.63	5.32
		$\lvert X-0.5\rvert,\ (X-0.5)^2$	4.68	5.32	4.98	4.58	4.43
		$1,\ \lvert X-0.5\rvert$	5.42	6.01	5.27	6.55	7.29
	500	$\lvert X-0.5\rvert,\ 1$	4.04	4.53	6.16	8.03	9.61
		$(X-0.5)^2,\ 1$	4.38	4.53	6.11	6.90	7.83
		$\lvert X-0.5\rvert,\ \lvert X-0.5\rvert$	4.58	4.58	6.21	4.68	6.90
		$\lvert X-0.5\rvert,\ (X-0.5)^2$	4.43	4.78	5.12	5.62	5.81
		$1,\ \lvert X-0.5\rvert$	5.27	5.52	6.75	7.14	8.62
5	125	$\lvert X-0.5\rvert,\ 1$	4.29	5.22	5.02	6.45	5.62
		$(X-0.5)^2,\ 1$	5.42	4.24	3.74	4.73	4.48
		$\lvert X-0.5\rvert,\ \lvert X-0.5\rvert$	4.78	5.91	5.37	6.21	6.70
		$\lvert X-0.5\rvert,\ (X-0.5)^2$	4.53	6.06	5.57	7.00	6.85
		$1,\ \lvert X-0.5\rvert$	5.07	5.86	5.22	5.71	5.47
	250	$\lvert X-0.5\rvert,\ 1$	4.33	5.37	6.11	8.18	11.63
		$(X-0.5)^2,\ 1$	4.68	4.88	4.19	6.26	8.92
		$\lvert X-0.5\rvert,\ \lvert X-0.5\rvert$	4.83	5.12	6.50	7.24	10.20
		$\lvert X-0.5\rvert,\ (X-0.5)^2$	4.53	5.52	6.90	7.83	11.77
		$1,\ \lvert X-0.5\rvert$	5.57	4.93	5.91	6.35	7.54
	500	$\lvert X-0.5\rvert,\ 1$	5.37	5.27	8.67	12.61	19.36
		$(X-0.5)^2,\ 1$	4.93	5.12	7.04	11.58	18.77
		$\lvert X-0.5\rvert,\ \lvert X-0.5\rvert$	5.07	6.70	7.83	11.03	16.01
		$\lvert X-0.5\rvert,\ (X-0.5)^2$	5.37	6.75	8.23	11.08	17.73
		$1,\ \lvert X-0.5\rvert$	5.42	5.57	7.04	7.29	8.62

Note: The alternative hypothesis is the Clayton. The dependency parameter corresponds to a Kendall's tau of 0.2.

Table III.3: Percentage rejection (5% level) of the Gaussian copula under various dimensions, sample sizes and weight combinations.

d	n	$\Gamma_V(X),\ \Gamma_H(X)$	β				
			0	0.25	0.5	0.75	1
2	125	$\lvert X-0.5\rvert,\ 1$	5.91	4.88	5.71	6.06	6.55
		$(X-0.5)^2,\ 1$	6.16	4.88	5.76	5.02	5.37
		$1,\ \Phi^{-1}(X)^2$	5.96	5.22	4.88	4.58	4.73
		$1,\ (X-0.5)^8$	5.91	5.71	4.58	4.98	5.07
	250	$\lvert X-0.5\rvert,\ 1$	5.07	5.27	5.91	6.75	8.37
		$(X-0.5)^2,\ 1$	5.67	4.78	4.93	5.52	6.90
		$1,\ \Phi^{-1}(X)^2$	4.93	5.57	5.22	4.19	4.63
		$1,\ (X-0.5)^8$	5.22	5.17	5.91	3.99	4.73
	500	$\lvert X-0.5\rvert,\ 1$	5.52	6.40	6.75	9.85	12.51
		$(X-0.5)^2,\ 1$	5.27	5.22	6.75	9.06	12.32
		$1,\ \Phi^{-1}(X)^2$	5.32	5.12	5.07	5.22	6.11
		$1,\ (X-0.5)^8$	5.22	5.12	5.17	5.42	5.67
5	125	$\lvert X-0.5\rvert,\ 1$	4.68	4.09	4.29	4.14	3.94
		$(X-0.5)^2,\ 1$	4.88	4.09	3.30	3.15	3.60
		$1,\ \Phi^{-1}(X)^2$	5.32	6.06	4.98	5.22	5.17
		$1,\ (X-0.5)^8$	5.22	5.76	4.93	5.07	4.68
	250	$\lvert X-0.5\rvert,\ 1$	5.22	4.04	3.99	5.27	6.60
		$(X-0.5)^2,\ 1$	5.07	4.04	4.29	4.88	7.00
		$1,\ \Phi^{-1}(X)^2$	5.71	4.63	5.47	5.42	5.52
		$1,\ (X-0.5)^8$	5.32	5.37	5.76	5.71	5.47
	500	$\lvert X-0.5\rvert,\ 1$	4.63	4.24	4.98	8.33	12.81
		$(X-0.5)^2,\ 1$	4.53	3.65	5.57	9.41	17.09
		$1,\ \Phi^{-1}(X)^2$	4.98	5.57	5.32	6.06	7.44
		$1,\ (X-0.5)^8$	6.06	4.83	4.53	6.11	6.50

Note: The alternative hypothesis is the Gumbel copula. The dependency parameter corresponds to a Kendall's tau of 0.2.

standard GARCH$(1,1)$ process. This filtering is done to remove serial dependence in each individual time series. For details of GARCH processes, see e.g. Bollerslev (1986). We fit the Gaussian-, Student-t-, Clayton- and Gumbel copulae to the portfolios and apply the cpit2-approach, with $\Gamma_V(X;\boldsymbol{\alpha}) = |X - 0.5|$ and $\Gamma_H(X;\boldsymbol{\alpha}) = 1$, to investigate how often each copula is rejected.

Table III.4 shows the rejection rates for the raw and filtered returns. We see that for all but the Student-t copula, the rejection rate is increasing with sample size. For $d = 5$ the rejection rates for the Clayton- and Gumbel copulae are very high. This is not surprising since we are only considering the so-called exchangeable Clayton- and Gumbel copulae, having only one dependency parameter. Fitting a 5-dimensional distribution with only one parameter is usually not sufficient. The Gaussian copula is not that easily rejected for small sample sizes in the bivariate case. However, for higher dimensions and sample sizes, we see that the Gaussian copula is strongly rejected for both raw and filtered returns. The Student-t copula seems to provide a very good fit for all dimensions and sample sizes and for both raw and filtered returns. It is not surprising that the Student-t copula outperforms the other copulae since it has more parameters. Nevertheless, the low rejection rates for the Student-t copula are interesting. Also, note the reduced rejection rates in most cases for the filtered returns. This is also expected since serial dependence is removed.

III.6 Concluding remarks

We have generalized and extended the copula gof approach proposed by Breymann et al. (2003). The main contribution is the flexibility in the dimension reduction function. The generalization enables the user to apply any weight function combination to the cpit data sets $\mathbf{V}$ and $\mathbf{H}$, depending on the use. The additional cpit step, based on order statistics, should make the dimension reduction more robust to the inconsistency issue illustrated in Figure III.1. We have not been able to reconstruct the inconcistency issue in our Monte Carlo study, except when the alternative copula is the Clayton copula, for $d = 5$, $n = 125$ where the Γ_H term seems to add power. However, the added power in this case may also be due to the high dimension and very few samples. Neither have we found a real world data set where this issue manifests itself. However, the danger of this issue coming into play will always be there, justifying our extension.

Monte Carlo results show that our approach keeps the prescribed nominal level for all weight combinations examined. We also see that the cpit-approach has low power in some circumstances, particularly for low sample sizes. The reason is that the dimension reduction strongly weights the boundaries of the d-dimensional unit hypercube. If we have few samples there are few observations in the boundary regions and the cpit-approach becomes less robust and less powerful. An important result is the superior performance, in all our numerical tests, of the two weight combinations (i): $\Gamma_V(X;\boldsymbol{\alpha}) = |X - 0.5|$, $\Gamma_H(X;\boldsymbol{\alpha}) = 1$ and (ii): $\Gamma_V(X;\boldsymbol{\alpha}) = (X - 0.5)^2$, $\Gamma_H(X;\boldsymbol{\alpha}) = 1$. Hence, based on our numerical experiments, these combinations are recommended. For skewness properties, it may seem like the additional Γ_H term has some effect, in particular for higher dimensions and small sample sizes. With the inconsistency issue from Figure III.1 in mind, the additional use of $\Gamma_V(X;\boldsymbol{\alpha}) = |X - 0.5|$, $\Gamma_H(X;\boldsymbol{\alpha}) = |X - 0.5|$ is recommended.

Application of the cpit2-approach to a collection of large cap stock portfolios show that the Student-t copula provide a fairly good fit to the data while the Gaussian copula is strongly rejected for higher dimensions. A GARCH(1,1) filtering of the original data

Table III.4: Percentage rejection (5% level) of the Gaussian-, Student-t and one-parameter Clayton- and Gumbel copulae applied to daily log-returns of large cap stock portfolios.

Copula	*Dimension*	*Sample size*	*Raw returns*	*Filtered returns*
Gaussian	2	250	6.2	5.2
		500	11.4	5.3
		1000	26.2	11.7
	5	250	12.8	7.4
		500	16.5	12.1
		1000	49.7	25.1
Student-t	2	250	5.1	7.5
		500	6.2	5.4
		1000	6.4	3.5
	5	250	7.3	5.3
		500	9.8	6.4
		1000	10.7	10.3
Clayton	2	250	7.3	7.3
		500	9.9	5.6
		1000	21.5	8.4
	5	250	32.1	19.6
		500	41.3	35.4
		1000	77.8	60.4
Gumbel	2	250	5.9	3.1
		500	5.3	7.9
		1000	14.4	5.4
	5	250	44.8	31.6
		500	44.0	33.3
		1000	75.3	51.0

Note: Raw and GARCH(1,1) filtered log-returns. Weight combination $\Gamma_V(X) = |X - 0.5|$ and $\Gamma_H(X) = 1$.

only marginally reduced the rejection of the Gaussian copula. This is in accordance with the findings of Dobrić and Schmid (2005) and Chen et al. (2004) and indicates that the Student-t copula, in general provides a superior fit to daily log-returns for equity prices.

Further work involve comparison of the cpit2-approach with other approaches, e.g. Panchenko (2005), Genest and Rémillard (2008) and Genest et al. (2006a). Further tests of various weight combinations and their relative performance with respect to various degrees of dependence, dimensions, sample sizes and null- and alternative hypotheses, is also of interest. Finally work needs to be done to better understand how the weight combinations relate to the original data set $\mathbf{Z}$, not only the cpit data set $\mathbf{V}$.

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 gof testing and for her continuous interest and drive. Also, thank is due to an anonymous referee and participants at conferences and workshops, in particular Professor Christian Genest and Professor Dr. Friedrich Schmid, for valuable comments.

IV

Copula goodness-of-fit testing: An overview and power comparison

Daniel Berg
University of Oslo & Norwegian Computing Center

Abstract

Several copula goodness-of-fit approaches are examined, three of which are proposed in this paper. Results from an extensive Monte Carlo study are presented, where we examine the effect of dimension, sample size and strength of dependence on the nominal level and power of the different approaches. While no approach is always the best, some stand out and conclusions and recommendations are made. A novel study of p-value variation due to permuation order, for approaches based on Rosenblatt's transformation is also carried out. Results show significant variation due to permutation order for some of the approaches based on this transform. However, when approaching rejection regions, the additional variation is negligible. Finally, motivated by the permutation study, new versions of some goodness-of-fit approaches are proposed and examined. The new versions consider all permutation orders of the variables and we see some power improvement over the approaches that consider one permutation order only.

Key words

Copula, Cramér-von Mises statistic, empirical copula, goodness-of-fit, parametric bootstrap, pseudo-observations, Rosenblatt's transform

Publication details

Submitted for publication - October 2007.

IV.1 Introduction

A copula contains all the information about the dependency structure of a random vector. Due to the representation theorem of Sklar (1959), every distribution function H can be written as $H(x_1, \ldots, x_d) = C\{F_1(x_1), \ldots, F_d(x_d)\}$, where $F_1, \ldots, F_d$ are the marginal distributions and $C : [0,1]^d \to [0,1]$ is the copula. This enables the modelling of marginal distributions and the dependence structure in separate steps. This feature in particular has motivated successful applications in areas such as survival analysis, hydrology, actuarial science and finance. For exhaustive and general introductions to copulae, the reader is referred to Joe (1997) and Nelsen (1999), and for introductions oriented to financial applications, Malevergne and Sornette (2006) and Cherubini et al. (2004). While the evaluation of univariate distributions is well documented, the study of goodness-of-fit (GoF) tests for copulas emerged only recently as a challenging inferential problem.

Let C be the underlying d-variate copula of a population. Suppose one wants to test the composite GoF hypothesis

$$\mathcal{H}_0 : C \in \mathcal{C} = \{C_\theta; \theta \in \Theta\} \quad \text{vs.} \quad \mathcal{H}_1 : C \notin \mathcal{C} = \{C_\theta; \theta \in \Theta\}, \tag{IV.1}$$

where Θ is the parameter space. Lately, several contributions have been made to test this hypothesis, e.g. Genest and Rivest (1993), Shih (1998), Breymann et al. (2003), Malevergne and Sornette (2003), Scaillet (2006), Genest and Rémillard (2008), Fermanian (2005), Panchenko (2005), Genest et al. (2006a), Berg and Bakken (2007), Dobrić and Schmid (2007), Quessy et al. (2007), Genest et al. (2008), among others. However, the field is still in its infancy and general guidelines and recommendations are sparse.

For univariate distributions, the GoF assessment can be performed using e.g. the well-known Anderson-Darling statistic (Anderson and Darling, 1954), or less quantitatively using a QQ-plot. In the multivariate domain there are fewer alternatives. A simple way to build GoF approaches for multivariate random variables is to consider multi-dimensional chi-square approaches, as in for example Dobrić and Schmid (2005). The problem with this approach, as with all binned approaches based on gridding the probability space, is that they will not be feasible for high dimensional problems due to the curse of dimensionality. 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 a highly irregular empirical cumulative distribution function (cdf) due to the limited amount of data. For these reasons, multivariate binned approaches are not considered in this study. Multivariate kernel density estimation (KDE) approaches such as the ones proposed by Fermanian (2005) and Scaillet (2006) are also excluded from this study as they are simply too computationally exhaustive for high dimensional problems. The author believes GoF to be most useful for high-dimensional problems since copulae are then harder to conceptualize. Moreover, the consequences of poor model choice is often much greater in higher dimensional problems, e.g. risk assessments for high dimensional financial portfolios.

The class of dimension reduction approaches is a more promising alternative. Dimension reduction approaches reduce the multivariate problem to a univariate problem, and then apply some univariate test, leading to numerically efficient approaches even for high dimensional problems. These approaches primarily differ in the way the dimension reduction is carried out. For the univariate test it is common to apply standard univariate statistics such as Kolmogorov- or Cramér-von Mises type statistics. Examples include

Breymann et al. (2003), Malevergne and Sornette (2003), Genest et al. (2006a), Berg and Bakken (2007), Quessy et al. (2007) and Genest and Rémillard (2008) among others.

This paper is organized as follows. In Section IV.2 some preliminaries are presented. Section IV.3 gives an overview of the nine GoF approaches considered, including three new ones. In Section IV.4 results from an extensive Monte Carlo study are presented, where we examine the effect of dimension, sample size and strength of dependence on the nominal level and power of the approaches. Several null- and alternative hypothesis copulae are considered. Further, this section also presents results from a novel numerical study of the effect of permutation order for approaches based on Rosenblatt's transform. New versions of some of the approaches based on this transform are proposed and examined. These new versions utilize all permutation orders of the data in an attempt to extract more information, and hence increase the power. Finally, Section IV.5 discusses and recommends. In addition, detailed testing procedures, leading to proper p-value estimates for all approaches, are given in the appendix.

IV.2 Preliminaries

For copula GoF testing one is interested in the fit of the copula alone. Typically, one does not wish to introduce any distributional assumptions for the margins. Instead the testing is carried out using rank data. Suppose we have a random d-variate vector $\mathbf{X}$. The inference is then based on the so-called pseudo-vector $\mathbf{Z}$:

$$\mathbf{Z}_j = (Z_{j1}, \ldots, Z_{jd}) = \left(\frac{R_{j1}}{n+1}, \ldots, \frac{R_{jd}}{n+1} \right), \tag{IV.2}$$

where R_{ji} is the rank of X_{ji} amongst $(X_{1i}, \ldots, X_{ni})$. This transformation of each margin through their normalized ranks is often denoted the empirical marginal transformation. Given the independent samples $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$, the pseudo-observations $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ can be considered to be samples from the underlying copula C. However, due to the rank transformation, $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ are no longer independent samples. In addition, since we are testing a hypothesized parametric copula model, as summarized by (IV.1), parameter estimation error will influence the limiting distribution of any GoF approach. The practical consequence is the need for parametric bootstrap procedures to obtain reliable p-value estimates. This is treated in more detail in Secion IV.3.10.

IV.2.1 Rosenblatt's transformation

The Rosenblatt transformation, proposed by Rosenblatt (1952), transforms a set of dependent variables into a set of independent $U[0,1]$ variables, given the multivariate distribution. The transformation is a universally applicable way of creating a set of i.i.d. $U[0,1]$ variables from any set of dependent variables with known distribution. Given a test for multivariate, independent uniformity, the transformation can be used to test the fit of any assumed model.

Definition IV.1 (Rosenblatt's transformation)
Let $\mathbf{Z} = (Z_1, \ldots, Z_d)$ denote a random vector with marginal distributions $F_i(z_i) = P(Z_i \leq z_i)$ and conditional distributions $F_{i|1\ldots i-1}(Z_i \leq z_i | Z_1 = z_1, \ldots, Z_{i-1} = z_{i-1})$ for $i = 1, \ldots, d$. Rosenblatt's transformation of $\mathbf{Z}$ is defined as $\mathcal{R}(\mathbf{Z}) = (\mathcal{R}_1(Z_1), \ldots, \mathcal{R}_d(Z_d))$

where

$$\begin{aligned}
\mathcal{R}_1(Z_1) &= P(Z_1 \leq z_1) = F_1(z_1), \\
\mathcal{R}_2(Z_2) &= P(Z_2 \leq z_2 | X_1 = z_1) = F_{2|1}(z_2|z_1), \\
&\vdots \\
\mathcal{R}_d(Z_d) &= P(Z_d \leq x_d | Z_1 = z_1, \ldots, Z_{d-1} = z_{d-1}) = F_{d|1\ldots d-1}(z_d|z_1, \ldots, z_{d-1}).
\end{aligned}$$

The random vector $\mathbf{V} = (V_1, \ldots, V_d)$, where $V_i = \mathcal{R}_i(Z_i)$, is now i.i.d. $U[0,1]^d$.

A recent application of this transformation is multivariate GoF tests. The Rosenblatt transformation is applied to a data set, assuming a multivariate null hypothesis distribution, and then a test of multivariate independence is carried out on the resulting transformed data set. The null hypothesis is typically a parametric copula family. The parameters of this copula family needs to be estimated before performing the transformation.

One advantage with Rosenblatt's transformation in a GoF setting is that the null- and alternative hypotheses are the same, regardless of the distribution before the transformation. Hong and Li (2002) report Monte Carlo evidence of multivariate tests using transformed variables outperforming tests using the original random variables. Chen et al. (2004) believe that a similar conclusion also applies to GoF tests for copulae. Another advantage is that computationally intensive double bootstrap procedures can be avoided for some approaches.

A disadvantage with tests based on Rosenblatt's transformation is the lack of invariance with respect to the permutation of the variables since there are $d!$ possible permutations. However, as long as the permutation is decided randomly, the results will not be influenced in any particular direction. The practical implications of this disadvantage is studied in Section IV.4.2.

IV.2.2 Parameter estimation

Testing the hypothesis in (IV.1) involves the estimation of the copula parameters $\boldsymbol{\theta}$ by some consistent estimator $\widehat{\boldsymbol{\theta}}$. There are mainly two ways of estimating these parameters; the fully parametric method or a semi-parametric method. The fully parametric method, termed the inference functions for margins (IFM) method (Joe, 1997), relies on the assumption of parametric, univariate margins. First, the parameters of the margins are estimated and then each parametric margin is plugged into the copula likelihood which is then maximized with respect to the copula parameters. Since we treat the margins as nuisance parameters we choose to proceed with the pseudo-vector $\mathbf{Z}$ and the semi-parametric method. This method is denoted the pseudo-likelihood (Demarta and McNeil, 2005) or the canonical maximum likelihood (CML) (Romano, 2002) method, and is described in Genest et al. (1995) and in Shih and Louis (1995) in the presence of censorship. Having obtained the pseudo-vector $\mathbf{Z}$ as described in (IV.2), the copula parameters can be estimated using either maximum likelihood (ML) or using the well-known relation to Kendall's tau.

For the elliptical copulae in higher dimensions the pairwise sample Kendall's tau's are inverted. This gives the correlation- and scale matrix for the Gaussian and Student copulae, respectively. For the Student copula one must also estimate the degree-of-freedom. We follow Mashal and Zeevi (2002) and Demarta and McNeil (2005), who propose a two-stage approach in which the scale matrix is first estimated by inversion of Kendall's tau, and

then the pseudo-likelihood function is maximized with respect to the degree-of-freedom ν, using the estimate of the scale matrix. For the Archimedean copulae the parameter is estimated by inversion of Kendall's tau. For dimension $d > 2$ we estimate the parameter as the average of the $d(d-1)/2$ pairs of Kendall's tau's.

IV.3 Copula goodness-of-fit approaches

The following nine copula GoF approaches are examined:

$\mathcal{A}_1$: Based on Rosenblatt's transformation, proposed by Berg and Bakken (2007). This approach includes, as special cases, the approaches proposed by Malevergne and Sornette (2003), Breymann et al. (2003), and the second approach in Chen et al. (2004).

$\mathcal{A}_2$: Based on the the empirical copula and the copula distribution function, proposed by Genest and Rémillard (2008).

$\mathcal{A}_3$: Based on approach $\mathcal{A}_2$ and the Rosenblatt transformation, proposed by Genest et al. (2008).

$\mathcal{A}_4$: Based on the empirical copula and the cdf of the copula function, proposed by Savu and Trede (2004) and Genest et al. (2006a).

$\mathcal{A}_5$: Based on Spearman's dependence function, proposed by Quessy et al. (2007).

$\mathcal{A}_6$: A new approache that extends Shih's test (Shih, 1998) for the bivariate Clayton model to arbitrary dimension.

$\mathcal{A}_7$: Based on the inner product between two vectors as a measure of their distance, proposed by Panchenko (2005).

$\mathcal{A}_8$: A new approach based on approach $\mathcal{A}_7$ and the Rosenblatt transformation.

$\mathcal{A}_9$: A new approach based on averages of the approaches above.

Approaches $\mathcal{A}_1$-$\mathcal{A}_5$ are all dimension reduction approaches, while $\mathcal{A}_6$ is a moment-based approach and $\mathcal{A}_7$-$\mathcal{A}_8$ are full multivariate approaches. For all the dimension reduction approaches only the Cramér-von Mises statistic is considered for the unviariate test.

IV.3.1 Approach $\mathcal{A}_1$

Berg and Bakken (2007) propose a generalization of the approches proposed by Breymann et al. (2003) and Malevergne and Sornette (2003). The approach is based on Rosenblatt's transformation applied to the pseudo-vector $\mathbf{Z}$ from (IV.2), assuming a null hypothesis copula $C_{\widehat{\theta}}$. The d-variate vector $\mathbf{V}$, resulting from the transformation, is i.i.d. $U[0,1]^d$ under the null hypothesis.[1] Berg and Bakken (2007) also propose a second Rosenblatt transformation, applied to $\mathbf{V}$ but this term will not be considered here.

[1] Since we are working with rank data this is only close to, but not exactly true. This issue is discussed in Section IV.3.10. Until then it is assumed that this holds.

The dimension reduction of approach $\mathcal{A}_1$ is based on $\mathbf{V}$:

$$W_1 = \sum_{i=1}^{d} \Gamma\{V_i; \boldsymbol{\alpha}\}, \tag{IV.3}$$

where Γ is any weight function used to weight the information in $\mathbf{V}$ and $\boldsymbol{\alpha}$ is the set of weight parameters. Any weight function may be used, depending on the use and the region of $\mathbf{V}$ one wishes to emphasize. Consider for example the special case $\Gamma\{V_i; \boldsymbol{\alpha}\} = \Phi^{-1}(V_i)^2$ which corresponds to the approach proposed by Breymann et al. (2003). If the null hypothesis is the Gaussian copula this is also equivalent with the approach proposed by Malevergne and Sornette (2003). Both of the latter studies apply the Anderson-Darling (Anderson and Darling, 1954) statistic. Berg and Bakken (2007) show that the Anderson-Darling statistic with $\Gamma\{V_i; \boldsymbol{\alpha}\} = |V_i - 0.5|$ performs particularly well for testing the Gaussian null hypothesis. Hence, when performing the numerical studies in Section IV.4.1 the following two special cases of approach $\mathcal{A}_1$ are considered:

$$\mathcal{A}_1^{(i)} : \Gamma\{V_i; \boldsymbol{\alpha}\} = \Phi^{-1}(V_i)^2 \quad \text{and} \quad \mathcal{A}_1^{(ii)} : \Gamma\{V_i; \boldsymbol{\alpha}\} = |V_i - 0.5|.$$

For approach $\mathcal{A}_1^{(i)}$ it is easy to see that the distribution of W_1 is a χ^2_d distribution[1]. However, for approach $\mathcal{A}_1^{(ii)}$, and in general, the distribution of W_1 is not known and one must turn to a double bootstrap procedure to approximate the cdf F_1 under the null hypothesis. The test observator S_1 of approach $\mathcal{A}_1$ is defined as the cdf of $F_1(W_1)$:

$$S_1(w) = P\{F_1(W_1) \leq w\}, \qquad w \in [0, 1].$$

Under the null hypothesis, all V_i are i.i.d. $U[0, 1]$, hence $S_1(w) = w$. Suppose we have the random samples $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ from $\mathbf{V}$. Then the empirical version of the test observator can be computed as

$$\widehat{S}_1(w) = \frac{1}{n+1} \sum_{j=1}^{n} I\{F_1(W_1) \leq w\}. \tag{IV.4}$$

This paper only considers the Cramér-von Mises statistic. In Appendix IV.B we show that this becomes:

$$\begin{aligned} \widehat{T}_1 &= n \int_0^1 \{\widehat{S}_1(w) - S_1(w)\}^2 \, \mathrm{d}S_1(w) \\ &= \frac{n}{3} + \frac{n}{n+1} \sum_{j=1}^{n} \widehat{S}_1\left(\frac{j}{n+1}\right)^2 - \frac{n}{(n+1)^2} \sum_{j=1}^{n} (2j+1) \widehat{S}_1\left(\frac{j}{n+1}\right). \end{aligned} \tag{IV.5}$$

IV.3.2 Approach $\mathcal{A}_2$

Genest and Rémillard (2008) propose to use the copula distribution function for the dimension reduction. The approach is based on the empirical copula process, introduced by Deheuvels (1979):

$$\widehat{C}(\mathbf{u}) = \frac{1}{n+1} \sum_{j=1}^{n} I\left\{Z_{j1} \leq u_1, \ldots, Z_{jd} \leq u_d\right\}, \tag{IV.6}$$

where $\mathbf{Z}_j$ is given by (IV.2) and $\mathbf{u} = (u_1, \ldots, u_d) \in [0, 1]^d$. The empirical copula is the observed frequency of $P(Z_1 < u_1, \ldots, Z_d < u_d)$. Suppose we have the random samples

$(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ from $\mathbf{Z}$. The idea is then to compare $\widehat{C}(\mathbf{z})$ with an estimation $C_{\hat{\theta}}(\mathbf{z})$ of C_θ. This is a very natural approach for copula GoF testing considering that most univariate GoF tests are based on a distance between an empirical- and null hypothesis distribution function. Genest et al. (2008) state that, given that it is entirely non-parametric, $\widehat{C}$ is the most objective benchmark for testing the copula GoF. We expect this approach to be very powerful since there are so few transformations of the data. A Cramér-von Mises statistic for approach $\mathcal{A}_2$ becomes (Genest et al., 2008):

$$\widehat{T}_2 = n \int_{[0,1]^d} \left\{ \widehat{C}(\mathbf{z}) - C_{\hat{\theta}}(\mathbf{z}) \right\}^2 \, \mathrm{d}\widehat{C}(\mathbf{z}) = \sum_{j=1}^{n} \left\{ \widehat{C}(\mathbf{z}_j) - C_{\hat{\theta}}(\mathbf{z}_j) \right\}^2. \tag{IV.7}$$

IV.3.3 Approach $\mathcal{A}_3$

Genest et al. (2008) propose to apply approach $\mathcal{A}_2$ to the vector $\mathbf{V}$ resulting from applying the Rosenblatt transform to $\mathbf{Z}$. Suppose we have the random samples $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ from $\mathbf{V}$. The idea is then to compare $\widehat{C}$ with the independence copula $C_\perp$. A Cramér-von Mises statistic for approach $\mathcal{A}_3$ becomes (Genest et al., 2008):

$$\widehat{T}_3 = n \int_{[0,1]^d} \left\{ \widehat{C}(\mathbf{v}) - C_\perp(\mathbf{v}) \right\}^2 \, \mathrm{d}\widehat{C}(\mathbf{v}) = \sum_{j=1}^{n} \left\{ \widehat{C}(\mathbf{v}_j) - C_\perp(\mathbf{v}_j) \right\}^2. \tag{IV.8}$$

IV.3.4 Approach $\mathcal{A}_4$

Savu and Trede (2004) and Genest et al. (2006a) propose to use Kendall's dependence function $K(w) = P(C(\mathbf{Z}) \leq w)$ as a GoF approach. The test observator S_4 of approach $\mathcal{A}_4$ becomes

$$S_4(w) = P\{C(\mathbf{Z}\} \leq w\}, \qquad w \in [0,1],$$

where $\mathbf{Z}$ is the pseudo-vector from (IV.2). Under the null hypothesis, $S_4(w) = S_{4,\hat{\theta}}(w)$ which is copula specific. Suppose we have the random samples $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ from $\mathbf{Z}$. The empirical version of test observator S_4 then equals

$$\widehat{S}_4(w) = \frac{1}{n+1} \sum_{j=1}^{n} I\{\widehat{C}(\mathbf{z}_j) \leq w\}. \tag{IV.9}$$

A Cramér-von Mises statistic for approach $\mathcal{A}_4$ is given by:

$$\widehat{T}_4 = n \int_0^1 \{\widehat{S}_4(w) - S_{4,\hat{\theta}}(w)\}^2 \, \mathrm{d}\widehat{S}_4(w) = \sum_{j=1}^{n} \left\{ \widehat{S}_4\left(\frac{j}{n+1}\right) - S_{4,\hat{\theta}}\left(\frac{j}{n+1}\right) \right\}^2. \tag{IV.10}$$

IV.3.5 Approach $\mathcal{A}_5$

Quessy et al. (2007) propose a GoF approach for bivariate copulae based on Spearman's dependence function $L_2(w) = P(Z_1 Z_2 \leq w)$. Notice that $L_2(w) = P(C_\perp(Z_1, Z_2) \leq w)$. A natural extension to arbitrary dimension d is then $L_d(w) = P(C_\perp(\mathbf{Z}) \leq w)$ and the test observator S_5 of approach $\mathcal{A}_5$ becomes

$$S_5(w) = P\{C_\perp(\mathbf{Z}) \leq w\}, \quad w \in [0,1],$$

where $\mathbf{Z}$ is the pseudo-vector from (IV.2). Under the null hypothesis, $S_5(w) = S_{5,\widehat{\theta}}(w)$, which is copula specific. Suppose we have the random samples $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ from $\mathbf{Z}$. The empirical version of test observator S_5 then equals

$$\widehat{S}_5(w) = \frac{1}{n+1} \sum_{j=1}^{n} I\{C_\perp(\mathbf{z}_j) \leq w\}. \tag{IV.11}$$

A Cramér-von Mises statistic for approach A_5 is given by:

$$\widehat{T}_5 = n \int_0^1 \{\widehat{S}_5(w) - S_{5,\widehat{\theta}}(w)\}^2 \, \mathrm{d}\widehat{S}_5(w) = \sum_{j=1}^{n} \left\{ \widehat{S}_5\left(\frac{j}{n+1}\right) - S_{5,\widehat{\theta}}\left(\frac{j}{n+1}\right) \right\}^2. \tag{IV.12}$$

IV.3.6 Approach $\mathcal{A}_6$

Shih (1998) propose a moment-based GoF test for the bivariate gamma frailty model, also known as Clayton's copula. Shih (1998) considered unweighted and weighted estimators of the dependency parameter θ via Kendall's tau and a weighted rank-based estimator, namely

$$\widehat{\theta}_\tau = \frac{2\widehat{\tau}}{1-\widehat{\tau}} \quad \text{and} \quad \widehat{\theta}_W = \frac{\sum_{i<j} \Delta_{ij}/W_{ij}}{\sum_{i<j}(1-\Delta_{ij})/W_{ij}}, \tag{IV.13}$$

where $\widehat{\tau} = -1 + 4\sum_{i<j} \Delta_{ij}/\{n(n-1)\}$, $\Delta_{ij} = I\{(Z_{i1} - Z_{j1})(Z_{i2} - Z_{j2}) > 0\}$ and $W_{ij} = \sum_{k=1}^{n} I\{Z_{k1} \leq \max(Z_{i1}, Z_{j1}), Z_{k2} \leq \max(Z_{i2}, Z_{j2})\}$. Since $\widehat{\theta}_\tau$ and $\widehat{\theta}_W$ are both unbiased estimators of θ under the null hypothesis that $C = C_\theta$ for some $\theta \geq 0$, Shih (1998) propose the GoF statistic

$$\widehat{T}_{Shih} = \sqrt{n}\{\widehat{\theta}_\tau - \widehat{\theta}_W\}.$$

Shih (1998) shows that this statistic is asymptotically normal under the null hypothesis. Unfortunately, the variance provided by Shih (1998) was found to be wrong by Genest et al. (2006c), where a corrected formula is provided.

One way of extending this approach to arbitrary dimension d is to compare each pairwise element of $\widehat{\theta}_\tau$ and $\widehat{\theta}_W$. The resulting vector of $d(d-1)/2$ statistics will tend, asymptotically, to a $d(d-1)/2$ dimensional normal vector with a non-trivial covariance matrix. The normalized version of the vector, i.e. the inverted square root of the covariance matrix multiplied with the vector of statistics, will be asymptotically standard normal and hence the sum of squares will now be chi-squared with $d(d-1)/2$ degrees of freedom. The covariance matrix of the vector of statistics remains to be computed and is deferred to future research. For now we simply compute the non-normalized sum of squares and perform a parametric bootstrap to estimate the p-value, as for all other approaches.

The test statistic for approach $\mathcal{A}_6$ then becomes:

$$\widehat{T}_6 = \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} \left\{ \widehat{\theta}_{\tau,ij} - \widehat{\theta}_{W,ij} \right\}^2. \tag{IV.14}$$

$\widehat{\theta}_W$, and hence approach $\mathcal{A}_6$, is constructed specifically for testing the Clayton copula and will not be considered for testing any other copula model.

IV.3.7 Approach $\mathcal{A}_7$

Approaches $\mathcal{A}_1$-$\mathcal{A}_5$ are all two-stage dimension reduction approaches. First the problem is reduced to a univariate problem, second a univariate test statistic is applied. In contrast, the approach proposed by Panchenko (2005) tests the entire data set in one step. The approach is based on the inner product of $\mathbf{Z}$ and $\mathbf{Z}_{\widehat{\theta}}$, where $\mathbf{Z}$ is the pseudo-vector from (IV.2) and $\mathbf{Z}_{\widehat{\theta}}$ is the null hypothesis vector with $\widehat{\theta}$ being a consistent estimator of the copula parameter. The inner product can be used as a measure of the distance between two vectors. Now define the squared distance Q between the two vectors as

$$Q = \left\langle \mathbf{Z} - \mathbf{Z}_{\widehat{\theta}} \, |\kappa_d| \, \mathbf{Z} - \mathbf{Z}_{\widehat{\theta}} \right\rangle .$$

Here κ_d is a positive definite symmetric kernel such as the Gaussian kernel:

$$\kappa_d(\mathbf{Z}, \mathbf{Z}') = \exp\left\{-\|\mathbf{Z} - \mathbf{Z}'\|^2/(2dh^2)\right\},$$

with $\|\cdot\|$ denoting the Euclidean norm in $\mathbb{R}^d$ and $h > 0$ being a bandwidth. Q will be zero if and only if $\mathbf{Z} = \mathbf{Z}_{\widehat{\theta}}$. Suppose we have the random samples $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ from $\mathbf{Z}$. Now generate the random samples $(\mathbf{z}_1^*, \ldots, \mathbf{z}_n^*)$ from the null hypothesis vector $\mathbf{Z}_{\widehat{\theta}}$. Following the properties of an inner product, Q can be decomposed as $Q = Q_{11} - 2Q_{12} + Q_{22}$. Each term of this decomposition is estimated using V-statistics (see Denker and Keller (1983) for an introduction to U- and V-statistics) and the test statistic for approach $\mathcal{A}_7$ is given by:

$$\widehat{T}_7 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\kappa_d(\mathbf{z}_i, \mathbf{z}_j) - \frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\kappa_d(\mathbf{z}_i, \mathbf{z}_j^*) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\kappa_d(\mathbf{z}_i^*, \mathbf{z}_j^*). \tag{IV.15}$$

IV.3.8 Approach $\mathcal{A}_8$

Along the lines of approach $\mathcal{A}_3$ we propose a version of approach $\mathcal{A}_7$ based on $\mathbf{V}$, the vector resulting from the Rosenblatt transformation applied to $\mathbf{Z}$. Suppose we have the random samples $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ from $\mathbf{V}$. Now generate the random samples $(\mathbf{v}_1^*, \ldots, \mathbf{v}_n^*)$ from the independence copula. The statistic for approach $\mathcal{A}_8$ is simply

$$\widehat{T}_8 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\kappa_d(\mathbf{v}_i, \mathbf{v}_j) - \frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\kappa_d(\mathbf{v}_i, \mathbf{v}_j^*) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\kappa_d(\mathbf{v}_i^*, \mathbf{v}_j^*). \tag{IV.16}$$

IV.3.9 Approach $\mathcal{A}_9$

Finally, we propose to use averages of the approaches already introduced, as new approaches. Such averages will capture several aspects of the data and its potential deviation from the null hypothesis. Surely one can find optimal weights for a weighted average and the average should be taken over standardized variables, i.e. all approaches should be scaled appropriately. However, due to the computational load, this approach is included here in its most simple form as an interesting supplement and a hint of further research. Two averages are considered, first the average of all approaches and second the average of the three approaches based on the empirical copula, i.e. $\mathcal{A}_2$, $\mathcal{A}_3$ and $\mathcal{A}_4$. The corresponding statistics are defined as

$$\widehat{T}_9^{(i)} = \frac{1}{9}\left\{\widehat{T}_1^{(i)} + \widehat{T}_1^{(ii)} + \sum_{j=2}^{8}\widehat{T}_j\right\} \quad \text{and} \quad \widehat{T}_9^{(ii)} = \frac{1}{3}\left\{\widehat{T}_2 + \widehat{T}_3 + \widehat{T}_4\right\}. \tag{IV.17}$$

IV.3.10 Testing procedures

In Section IV.3.1 it was assumed that $\mathbf{V}$, resulting from applying Rosenblatt's transformation to $\mathbf{Z}$, is i.i.d. $U[0,1]^d$. The non-parametric margins introduce dependence in $\mathbf{V}$. Hence, it is only close to, but not exactly independent. This applies to all approaches considered here. In addition, we have small sample estimation error from the estimation of the null hypothesis copula parameter. To cope with these issues and obtain a proper estimate of the p-value of a statistic, one turns to parametric bootstrap procedures. The parametric bootstrap procedure used in Genest et al. (2006a) is adopted, the validity of which is established in Genest and Rémillard (2008). Dobrić and Schmid (2007) and Berg and Bakken (2007) propose a very similar procedure in their modification of the original procedure used in Breymann et al. (2003). The asymptotic validity of the bootstrap procedure has only been proved for the approaches $\mathcal{A}_2$ and $\mathcal{A}_4$. However, the results in Dobrić and Schmid (2007) and Berg and Bakken (2007) strongly indicates that the procedure is valid also for approach $\mathcal{A}_1$. This is further discussed in view of the results in Section IV.4.1 and in Section IV.5. The test procedure for approach $\mathcal{A}_7$, originally proposed in Panchenko (2005), gave us too low nominal levels (i.e. the rejection rate when the null hypothesis is true was lower than the prescribed size). However, a small fix, in line with the procedure of Genest and Rémillard (2008), solved this issue. Details of the test procedures for all approaches are given in Appendix IV.C. In many cases one must resort to a double parametric bootstrap to compute a statistic. This means that there are two bootstrap parameters that needs to be chosen, the sample size N_b for the double bootstrap step and the number of replications K for the estimation of p-values. In this paper the number of replications K is chosen to equal 1000, while the double bootstrap sample size N_b is chosen to equal 10000 for approach $\mathcal{A}_1$, and 2500 in dimension $d = \{2, 4\}$ and 5000 in dimension $d = 8$ for approaches $\mathcal{A}_2$, $\mathcal{A}_4$ and $\mathcal{A}_5$. See Appendix IV.C for details.

IV.4 Numerical experiments

IV.4.1 Size and power simulations

A large Monte Carlo study is performed to assess the properties of the approaches for various dimensions, sample sizes, levels of dependence and alternative dependence structures. The nominal levels and the power against fixed alternatives are of particular interest. The simulations are carried out according to the following factors:

- $\mathcal{H}_0$ copula (5 choices: Gaussian, Student, Clayton, Gumbel, Frank),
- $\mathcal{H}_1$ copula (5 choices: Gaussian, Student ($\nu = 6$), Clayton, Gumbel, Frank),
- Kendall's tau (2 choices: $\tau = \{0.2, 0.4\}$),
- Dimension (3 choices: $d = \{2, 4, 8\}$),
- Sample size (2 choices: $n = \{100, 500\}$).

Due to extreme computational load, the Student copula is only considered as null hypothesis in the bivariate case. In each of the remaining 260 cases, a sample of dimension d and size n is drawn from the $\mathcal{H}_1$ copula with dependence parameter corresponding to τ. The statistics of the various GoF approaches are then computed under the null hypothesis

$\mathcal{H}_0$ and p-values are estimated. This entire procedure is repeated $10,000$ times in order to estimate the nominal level and power for each approach under consideration.

Since we apply a parametric bootstrap procedure in the estimation of p-values, critical values are obtained by simulating from the null hypothesis, and hence all reported powers are so-called size-adjusted powers and approaches can be compared appropriately (see e.g. Hendry (2006) and Florax et al. (2006) for size-adjustment suggestions).

The critical values of each statistic under the true null hypothesis were tabulated for each dimension and sample size considered and for many levels of dependence. For the power simulations we used table look-up with linear interpolation to ensure comparison with the appropriate critical value. Despite the tabulation this computationally exhaustive experiment would not have been feasible without access to the *Titan* computer grid at the University of Oslo, a cluster of $1,750$ computing cores, 6.5 TB memory, 350 TB local disk and 12.5 Tflops.

Testing the Gaussian hypothesis

Let us first consider testing the Gaussian hypothesis under several fixed alternatives. Table IV.1 shows the results from our simulations.

Notice that the nominal levels of all approaches match the prescribed size of 5% well. Next, note that the power generally (but not always) increases with level of dependence, as expected since two copulae differs more and more as we move away from independence where all copulae are equivalent to the independence copula. Also note that the power increases with sample size, as it should for the approaches to be consistent. The power generally (but not always) also increases with dimension. This is as expected since it is natural to believe that the difference between two distributions increases with dimension, see for example Chen et al. (2004) who show that the Kullbach-Leibler Information Criterion (a measure of distance between two copulae) between the Gaussian- and Student copulae increases with dimension. Also, one can imagine that there is more for a GoF approach to work with the higher dimension is.

Next, we note that no approach is always the best. There are special cases where they perform well and cases where they perform poorly. For example, approaches $\mathcal{A}_1$ and $\mathcal{A}_3$ perform particularly well for testing against heavy tails, i.e. the Student copula alternative. $\mathcal{A}_1^{(i)}$ performs extremely well for high dimensions and large sample sizes while $\mathcal{A}_3$ performs very well for the bivariate case and for small sample sizes in higher dimensions. When Clayton and Gumbel are the alternatives, two of the approaches based on the empirical copula, $\mathcal{A}_2$ and $\mathcal{A}_4$, perform very well. In addition, in particular for Gumbel alternatives in higher dimensions, approach $\mathcal{A}_5$ performs very well. And finally, as expected, approach $\mathcal{A}_9^{(ii)}$, the average of $\mathcal{A}_2$, $\mathcal{A}_3$ and $\mathcal{A}_4$ perform very well for Clayton and Gumbel alternatives. For the Frank alternative, approach $\mathcal{A}_3$ performs particularly well for the bivariate case, but then, surprisingly, extremely poorly for higher dimensions while approaches $\mathcal{A}_4$ and $\mathcal{A}_5$ perform quite well for all dimensions. This shows us the danger of concluding for higher dimensions based on bivariate power results. We also note that approaches $\mathcal{A}_7$ and $\mathcal{A}_8$ are generally quite poor, they almost never perform among the best. However, at the same time they are usually not among the worst. Finally, we see that the average approaches perform quite well in most cases, sometimes being the most powerful ones.

One aspect of the power comparison that is lost when only looking at the best approach (bold in the tables), or when ranking the approaches, is that an approach can be almost as good as the best approach in all cases, but not necessarily the very best. For example

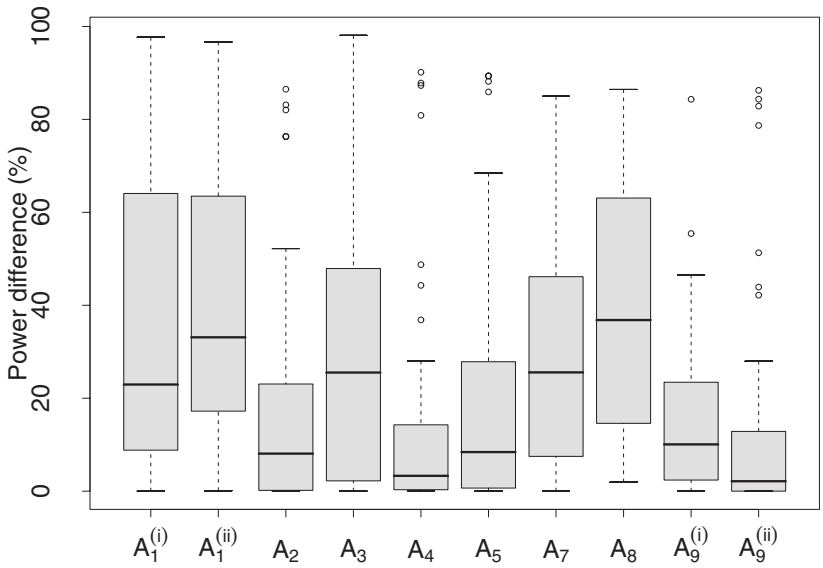


Figure IV.1: Distribution of power difference from the very best approach for testing the Gaussian copula.

when the alternative is the Gumbel copula for $d = 4$, $n = 500$ and $\tau = 0.40$, approach $\mathcal{A}_9^{(ii)}$ will be ranked 1 with a power of 99.8 while approach $\mathcal{A}_5$ will be ranked number 5 when its power is 98.1. This small difference in power may not even be statistically significant and purely due to Monte Carlo variation. Hence, in addition to the table we also examine a boxplot of the differences in power, from the best performing approach. This is depicted in Figure IV.1. From this figure we see that although approaches $\mathcal{A}_2$ and $\mathcal{A}_4$ are the best performing approaches in addition to the average approaches, the power in some very few cases is remarkably low compared to the best in those specific cases. All cases of poor performance of these approaches are for the Student alternative. Hence, for testing the Gaussian copula one should apply more than one approach, e.g $\mathcal{A}_2$ and $\mathcal{A}_3$ and in higher dimensions with large sample sizes also approach $\mathcal{A}_1^{(i)}$. The average approaches represent an attempt of applying several approaches simultaneously and we see that they have very stable and good performance. However, also for these approaches there are cases, although very few, of very poor performance compared to the very best approach.

For approach $\mathcal{A}_1$, Berg and Bakken (2007) report results where the weight function $\Gamma\{V_i; \boldsymbol{\alpha}\} = |V_i - 0.5|$ outperformed $\Gamma\{V_i; \boldsymbol{\alpha}\} = \Phi^{-1}(V_i)^2$, in particular for small sample sizes. These results are not confirmed in this paper where the conclusion is the opposite in almost all cases. However, in this paper the Cramér–von Mises statistic was applied while Berg and Bakken (2007) considered the Anderson–Darling statistic. Since the Anderson–Darling statistic emphasizes the tails of the distribution, when mixed with the extreme weight on the corners and edges of the unit hypercube from $\Phi^{-1}(V_i)$ it may be too extreme for small sample sizes. When using the Cramér–von Mises statistic this is apparently not the case.

Table IV.1: Percentage of rejections (at 5% significance level) of the Gaussian copula by approaches $\mathcal{A}_1$-$\mathcal{A}_9$.

d	n	τ	*True copula*	$\mathcal{A}_1^{(i)}$	$\mathcal{A}_1^{(ii)}$	$\mathcal{A}_2$	$\mathcal{A}_3$	$\mathcal{A}_4$	$\mathcal{A}_5$	$\mathcal{A}_6$	$\mathcal{A}_7$	$\mathcal{A}_8$	$\mathcal{A}_9^{(i)}$	$\mathcal{A}_9^{(ii)}$
2	100	0.2	Gaussian	*5.3*	*5.0*	*5.0*	*4.6*	*5.4*	*5.7*	–	*4.7*	*5.2*	*5.0*	*5.1*
			Student-t	0.9	4.2	7.0	**8.8**	6.1	5.3	–	5.6	6.0	3.3	6.4
			Clayton	2.6	5.0	19.7	19.6	19.9	15.6	–	7.1	6.9	10.6	**24.0**
			Gumbel	1.9	4.6	10.7	3.6	**11.6**	8.4	–	6.2	5.9	4.9	9.7
			Frank	3.4	3.2	6.0	**7.4**	6.0	6.2	–	5.4	5.5	3.4	6.1
		0.4	Gaussian	*5.2*	*5.0*	*4.7*	*5.4*	*4.8*	*4.7*	–	*5.0*	*4.7*	*5.0*	*4.9*
			Student-t	1.3	2.4	5.9	**11.6**	4.8	3.9	–	5.3	5.8	2.3	6.4
			Clayton	1.1	2.5	57.4	59.6	49.7	33.7	–	14.9	15.8	22.2	**63.9**
			Gumbel	1.3	2.6	**19.1**	5.0	18.5	8.2	–	7.0	7.9	4.1	16.2
			Frank	0.8	1.2	10.6	11.6	10.1	8.9	–	6.1	6.3	1.5	**11.8**
	500	0.2	Gaussian	*4.7*	*4.9*	*5.2*	*4.8*	*5.2*	*5.1*	–	*5.1*	*4.9*	*4.9*	*5.0*
			Student-t	19.5	16.9	10.0	16.9	8.4	8.5	–	10.3	9.8	**21.4**	10.0
			Clayton	2.0	5.8	72.5	71.3	71.9	57.2	–	23.8	20.3	56.5	**79.5**
			Gumbel	2.5	6.9	33.2	8.5	33.9	25.8	–	12.3	11.1	21.2	**34.3**
			Frank	2.2	2.9	11.4	**21.9**	11.1	9.9	–	7.6	8.1	5.8	14.5
		0.4	Gaussian	*5.0*	*5.0*	*4.6*	*5.4*	*4.9*	*4.8*	–	*4.9*	*5.5*	*5.1*	*4.8*
			Student-t	23.8	12.5	8.2	**30.5**	6.6	6.9	–	10.1	12.6	20.6	12.0
			Clayton	6.8	4.3	99.8	**100**	99.6	96.2	–	78.1	84.3	99.0	99.9
			Gumbel	8.8	6.0	**65.3**	18.9	62.9	39.8	–	26.4	32.4	42.3	**65.3**
			Frank	15.1	12.2	36.9	**60.7**	33.4	26.4	–	17.0	20.6	36.9	52.1
4	100	0.2	Gaussian	*4.8*	*5.0*	*4.6*	*4.8*	*4.8*	*5.3*	–	*5.6*	*5.0*	*5.0*	*4.9*
			Student-t	5.1	6.5	8.9	**15.4**	8.5	7.0	–	6.7	6.6	7.5	9.7
			Clayton	1.1	5.0	45.6	30.5	52.5	19.2	–	9.4	7.0	20.2	**55.9**
			Gumbel	1.2	3.1	12.8	0.7	42.5	**56.4**	–	13.9	8.8	13.2	34.9
			Frank	2.0	1.4	1.8	3.0	12.2	**19.6**	–	7.5	6.8	2.0	8.4
		0.4	Gaussian	*4.5*	*4.8*	*5.2*	*5.4*	*5.1*	*5.1*	–	*4.9*	*5.3*	*4.9*	*5.3*
			Student-t	9.2	3.7	8.6	**24.4**	6.1	5.3	–	6.9	7.1	7.5	8.1
			Clayton	1.1	1.8	**90.8**	80.4	84.0	45.6	–	27.9	18.3	48.8	90.1
			Gumbel	1.5	1.7	41.0	3.6	**52.0**	48.7	–	25.8	15.4	17.1	50.1
			Frank	1.6	2.2	10.1	7.3	**23.6**	20.6	–	12.6	8.3	5.6	21.2
	500	0.2	Gaussian	*5.8*	*5.3*	*5.3*	*5.0*	*4.8*	*4.9*	–	*5.0*	*5.5*	*4.9*	*4.7*
			Student-t	**98.5**	71.8	16.5	47.1	11.2	12.6	–	13.6	15.0	96.5	15.7
			Clayton	4.3	7.7	99.0	94.4	98.0	88.4	–	39.3	22.2	94.6	**99.2**
			Gumbel	8.0	5.9	84.2	48.0	97.7	**98.5**	–	70.3	34.7	92.3	98.0
			Frank	3.6	6.6	25.4	5.0	64.3	**66.2**	–	20.3	17.2	39.1	63.8
		0.4	Gaussian	*4.7*	*4.7*	*4.8*	*4.9*	*4.7*	*4.8*	–	*5.1*	*5.0*	*4.4*	*4.6*
			Student-t	**98.1**	67.5	11.6	72.1	8.0	8.8	–	16.4	18.7	94.0	13.8
			Clayton	44.3	13.2	**100**	**100**	**100**	99.9	–	97.2	91.2	**100**	**100**
			Gumbel	63.2	34.7	98.9	70.1	99.6	98.1	–	95.5	77.4	99.4	**99.8**
			Frank	79.3	74.2	73.2	19.5	88.6	74.5	–	61.2	40.7	**97.4**	90.6
8	100	0.2	Gaussian	*5.0*	*5.2*	*5.9*	*4.7*	*5.8*	*5.2*	–	*5.3*	*5.2*	*5.4*	*5.7*
			Student-t	**40.4**	16.4	9.8	15.0	12.3	7.7	–	7.9	6.9	35.9	12.4
			Clayton	0.7	4.1	48.7	24.3	**66.0**	1.2	–	11.8	6.6	19.5	65.5
			Gumbel	0.6	1.7	22.0	2.3	61.5	**98.3**	–	56.9	13.8	14.0	56.1
			Frank	0.4	0.6	3.8	1.3	7.3	**56.0**	–	14.4	7.2	0.6	4.7
		0.4	Gaussian	*5.1*	*5.2*	*5.0*	*4.6*	*5.3*	*5.7*	–	*5.5*	*5.1*	*5.3*	*5.1*
			Student-t	**51.7**	16.1	8.3	17.6	7.4	6.1	–	8.0	8.5	39.2	7.8
			Clayton	1.6	2.4	**96.6**	49.2	93.3	28.1	–	40.4	19.9	59.9	95.0
			Gumbel	16.2	10.1	70.5	2.7	78.4	**92.8**	–	67.9	28.1	52.7	78.6
			Frank	4.8	8.3	19.6	2.9	**28.7**	23.9	–	26.7	7.5	14.6	25.7
	500	0.2	Gaussian	*5.5*	*4.8*	*4.4*	*5.1*	*4.8*	*5.4*	–	*5.2*	*5.1*	*4.6*	*4.8*
			Student-t	**100**	99.9	23.7	56.4	19.1	11.8	–	21.7	20.9	**100**	21.3
			Clayton	11.8	12.9	**100**	74.3	99.7	84.8	–	50.5	13.6	97.2	99.9
			Gumbel	30.0	13.4	**100**	71.7	**100**	**100**	–	**100**	63.0	99.9	**100**
			Frank	22.9	38.3	99.8	10.5	98.4	**99.9**	–	69.6	19.4	90.7	99.8
		0.4	Gaussian	*4.9*	*5.4*	*4.9*	*5.2*	*5.4*	*5.1*	–	*4.7*	*5.9*	*5.1*	*5.2*
			Student-t	**100**	99.8	16.9	71.5	12.2	10.6	–	21.4	32.0	**100**	13.7
			Clayton	78.0	52.6	**100**	99.8	**100**	**100**	–	99.2	81.5	**100**	**100**
			Gumbel	**100**	98.7	**100**	33.9	**100**	**100**	–	**100**	94.7	**100**	**100**
			Frank	99.5	99.5	**100**	1.9	99.8	95.6	–	97.3	37.7	**100**	**100**

Note: The Student copula alternative hypothesis with degree-of-freedom $\nu = 6$. Numbers in *italic* are nominal levels and should correspond to the prescribed size of 5%. Numbers in **bold** indicates the best performing approach. All powers are size-adjusted.

Table IV.2: Percentage of rejections (at 5% significance level) of the bivariate Student copula by approaches $\mathcal{A}_1$-$\mathcal{A}_9$.

d	n	τ	*True copula*	$\mathcal{A}_1^{(i)}$	$\mathcal{A}_1^{(ii)}$	$\mathcal{A}_2$	$\mathcal{A}_3$	$\mathcal{A}_4$	$\mathcal{A}_5$	$\mathcal{A}_6$	$\mathcal{A}_7$	$\mathcal{A}_8$	$\mathcal{A}_9^{(i)}$	$\mathcal{A}_9^{(ii)}$
2	100	0.2	Gaussian	**5.7**	5.4	4.9	4.0	5.0	5.2	–	5.6	5.3	5.6	4.8
			Student-t	*4.4*	*4.6*	*4.8*	*4.1*	*5.1*	*4.8*	–	*5.1*	*5.0*	*4.6*	*4.8*
			Clayton	4.8	5.3	19.2	11.0	20.1	17.2	–	7.3	6.8	15.4	**21.3**
			Gumbel	4.7	5.1	9.2	4.9	**10.5**	7.0	–	5.9	5.8	7.6	10.1
			Frank	4.9	5.4	6.0	4.4	6.6	**7.1**	–	5.8	5.7	6.5	6.6
		0.4	Gaussian	4.7	5.4	4.9	4.0	5.2	5.4	–	**5.7**	4.9	5.2	4.8
			Student-t	*4.1*	*4.5*	*4.2*	*4.4*	*4.8*	*5.1*	–	*4.9*	*4.9*	*4.4*	*4.4*
			Clayton	4.2	4.9	55.0	31.7	53.3	41.1	–	15.4	14.8	39.9	**57.3**
			Gumbel	4.4	5.0	17.2	6.1	**18.7**	9.1	–	7.2	7.4	10.5	17.5
			Frank	2.9	3.4	11.8	5.3	**12.5**	10.5	–	7.5	6.3	6.9	11.6
	500	0.2	Gaussian	5.8	5.8	5.1	5.1	5.0	5.6	–	5.8	5.5	**6.0**	5.3
			Student-t	*5.1*	*5.1*	*4.5*	*4.5*	*4.5*	*5.3*	–	*5.1*	*5.2*	*4.8*	*4.6*
			Clayton	5.6	4.8	69.9	60.4	72.4	61.3	–	22.0	19.9	65.7	**77.5**
			Gumbel	5.2	5.3	28.6	18.6	30.0	19.7	–	11.0	10.0	23.5	**33.2**
			Frank	5.2	6.3	12.3	8.3	12.7	12.6	–	7.4	7.8	11.6	**13.4**
		0.4	Gaussian	**5.6**	5.2	4.5	5.3	5.0	5.5	–	5.2	4.9	5.4	5.0
			Student-t	*4.9*	*4.6*	*5.3*	*4.4*	*4.5*	*4.8*	–	*4.7*	*5.0*	*4.7*	*4.6*
			Clayton	6.4	7.0	99.8	99.6	99.6	97.7	–	74.6	78.4	99.5	**99.9**
			Gumbel	4.5	5.1	61.7	40.0	61.2	34.1	–	22.4	24.1	49.2	**68.3**
			Frank	11.6	5.9	41.2	15.4	40.4	31.7	–	17.2	14.2	36.0	**44.8**

Note: Student copula alternative hypothesis with degree-of-freedom $\nu = 6$. Numbers in *italic* are nominal levels and should correspond to the prescribed size of 5%. Numbers in **bold** indicates the best performing approach. All powers are size-adjusted.

Testing the Student hypothesis

Next, we consider testing the Student copula hypothesis, for the bivariate case only. Table IV.2 and Figure IV.2 show the results. Again we note that the nominal levels match the prescribed size well. The powers against the Gaussian copula are also very close to the nominal levels which makes sense since the Student copula approaches the Gaussian as the degrees of freedom increases. As for testing the Gaussian hypothesis, approaches $\mathcal{A}_2$, $\mathcal{A}_4$, and in particular $\mathcal{A}_9^{(ii)}$, perform very well. Approaches $\mathcal{A}_1$, $\mathcal{A}_7$ and $\mathcal{A}_8$ all perform rather poorly. While approach $\mathcal{A}_1$ performed very well for Student alternatives when testing the Gaussian copula, this is of course not the case when testing the Student copula since this is now the null hypothesis and nominal levels should, and do indeed, match the prescribed size of 5%.

Testing the Clayton hypothesis

Table IV.3 shows the results of testing the Clayton hypothesis and Figure IV.3 shows the power differences. The nominal levels match the prescribed size well. Again notice the very good performance of approaches $\mathcal{A}_2$ and $\mathcal{A}_4$. $\mathcal{A}_6$ does however outperform all other approaches. $\mathcal{A}_9^{(i)}$ also perform very well, but is highly dominated by $\mathcal{A}_6$ and does not provide additional knowledge in this case. Approach $\mathcal{A}_6$, the multivariate version of Shih's statistic, is constructed specifically for testing the Clayton copula. With this in mind, the performance of approaches $\mathcal{A}_2$ and $\mathcal{A}_4$ is quite impressive. While approach $\mathcal{A}_3$ performed very well for testing the Gaussian copula it performs very poorly for testing the Clayton copula, with terrible performance in some cases. Finally, note that the powers are in general higher than that for testing the Gaussian hypothesis, i.e. it is simpler to detect deviations from the Clayton copula than from the Gaussian copula.

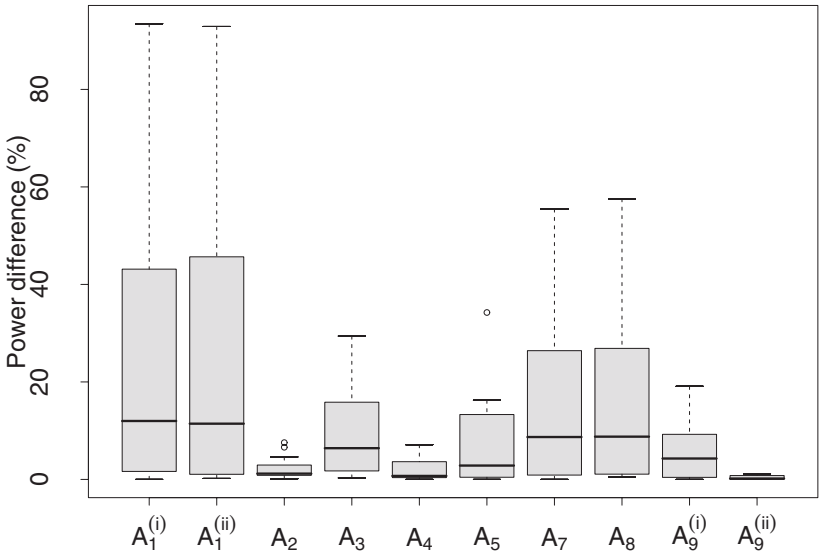


Figure IV.2: Distribution of power difference from the very best approach for testing the bivariate Student copula.

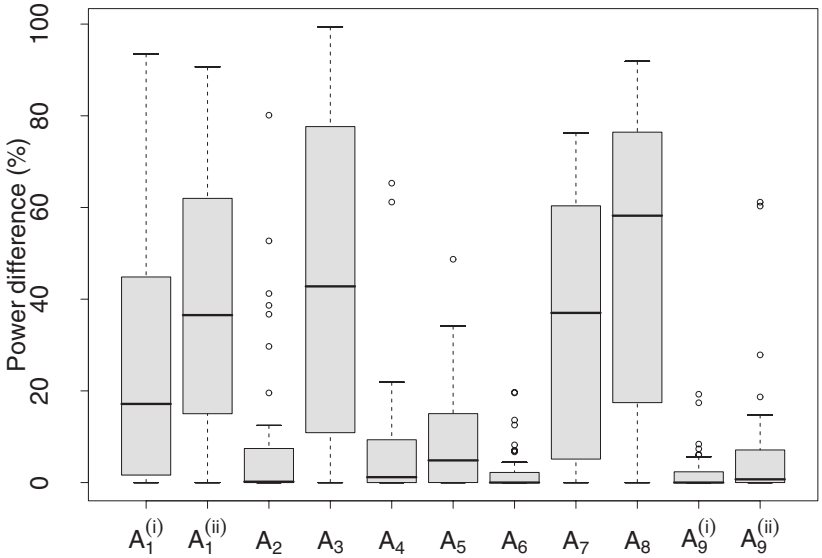


Figure IV.3: Distribution of power difference from the very best approach for testing the Clayton copula.

(IV.15)

Table IV.3: Percentage of rejections (at 5% significance level) of the Clayton copula by approaches $\mathcal{A}_1$-$\mathcal{A}_9$.

d	n	τ	*True copula*	$\mathcal{A}_1^{(i)}$	$\mathcal{A}_1^{(ii)}$	$\mathcal{A}_2$	$\mathcal{A}_3$	$\mathcal{A}_4$	$\mathcal{A}_5$	$\mathcal{A}_6$	$\mathcal{A}_7$	$\mathcal{A}_8$	$\mathcal{A}_9^{(i)}$	$\mathcal{A}_9^{(ii)}$
2	100	0.2	Gaussian	7.5	7.3	21.3	6.6	**23.2**	14.5	20.9	7.3	6.9	20.8	22.4
			Student-t	8.0	8.5	23.8	8.4	**24.1**	16.3	15.9	7.5	7.0	21.0	23.7
			Clayton	*4.9*	*5.1*	*5.0*	*5.2*	*5.0*	*5.2*	*4.5*	*5.2*	*5.2*	*5.0*	*5.1*
			Gumbel	6.2	9.4	46.7	13.0	**47.3**	32.3	40.4	12.4	11.1	41.2	47.1
			Frank	7.0	6.9	24.6	6.4	27.1	16.3	**30.3**	8.6	7.4	25.1	25.8
		0.4	Gaussian	24.0	26.7	58.9	26.4	58.2	33.7	62.1	16.6	15.3	**66.5**	60.6
			Student-t	13.4	19.0	**60.6**	16.0	58.4	35.1	53.6	15.4	13.7	58.2	57.3
			Clayton	*4.4*	*4.8*	*4.8*	*5.4*	*4.9*	*4.9*	*4.8*	*4.7*	*4.8*	*4.6*	*4.8*
			Gumbel	29.7	38.9	91.6	41.2	90.6	70.1	90.2	34.9	31.7	**92.0**	90.2
			Frank	24.1	19.2	64.8	24.2	66.2	35.6	**84.3**	19.3	16.5	77.0	65.6
	500	0.2	Gaussian	20.6	13.3	78.7	44.8	70.2	52.9	**85.9**	24.0	20.5	68.5	75.3
			Student-t	26.9	23.3	**82.1**	33.4	73.7	64.8	68.5	26.1	22.2	76.1	77.6
			Clayton	*5.2*	*5.1*	*5.0*	*4.8*	*5.1*	*5.4*	*5.1*	*5.3*	*4.5*	*4.8*	*5.2*
			Gumbel	12.6	23.2	**99.2**	84.9	97.9	94.0	99.0	60.1	52.0	97.2	98.6
			Frank	18.8	9.0	86.6	42.9	82.2	63.4	**97.6**	30.4	22.7	78.3	84.8
		0.4	Gaussian	94.8	85.6	**100**	99.5	99.7	95.5	**100**	77.7	82.3	99.9	99.9
			Student-t	65.3	71.4	**99.9**	89.7	99.6	97.3	99.8	74.7	74.9	99.8	99.8
			Clayton	*5.3*	*5.1*	*5.0*	*5.2*	*4.7*	*4.8*	*4.9*	*4.7*	*4.4*	*5.0*	*4.7*
			Gumbel	98.4	97.8	**100**	**100**	**100**	**100**	**100**	99.4	99.5	**100**	**100**
			Frank	97.8	69.9	**100**	99.4	99.9	96.7	**100**	84.6	86.8	**100**	**100**
4	100	0.2	Gaussian	10.8	10.6	37.4	3.2	38.5	39.1	**49.8**	10.6	6.5	49.2	37.9
			Student-t	27.1	21.3	48.4	17.8	37.7	42.2	37.7	10.1	7.3	**57.2**	42.5
			Clayton	*4.7*	*5.1*	*5.3*	*5.6*	*5.2*	*5.1*	*4.6*	*6.3*	*4.7*	*5.0*	*5.2*
			Gumbel	8.8	12.0	64.4	3.0	91.1	**94.1**	81.5	31.9	14.0	88.4	88.6
			Frank	7.7	6.5	36.0	1.4	**74.7**	68.9	73.0	15.1	7.2	72.8	68.8
		0.4	Gaussian	78.3	65.7	89.8	3.0	83.0	73.9	91.6	31.0	16.7	**95.2**	84.3
			Student-t	53.9	45.7	**92.9**	6.1	82.6	76.0	86.2	29.9	15.8	92.2	85.6
			Clayton	*5.2*	*4.7*	*5.6*	*5.5*	*5.2*	*5.1*	*4.5*	*5.3*	*4.9*	*5.1*	*5.3*
			Gumbel	79.1	62.1	99.3	4.9	99.8	99.8	99.8	80.8	40.1	**99.9**	99.8
			Frank	68.7	37.9	91.4	3.2	97.0	84.8	**99.6**	52.4	15.1	99.3	96.3
	500	0.2	Gaussian	89.6	38.1	99.4	18.1	97.0	91.2	**99.9**	38.8	23.0	99.4	98.0
			Student-t	93.7	76.9	99.9	89.7	95.8	94.5	97.9	44.1	30.8	**100**	98.7
			Clayton	*4.8*	*4.7*	*5.2*	*5.6*	*5.6*	*4.7*	*5.0*	*4.8*	*5.3*	*5.1*	*5.6*
			Gumbel	71.1	37.8	**100**	80.3	**100**	**100**	**100**	97.8	83.4	**100**	**100**
			Frank	82.6	11.8	99.8	14.5	**100**	99.9	**100**	67.9	24.8	**100**	**100**
		0.4	Gaussian	**100**	**100**	**100**	99.7	**100**	99.9	**100**	97.4	95.5	**100**	**100**
			Student-t	**100**	99.8	**100**	80.0	**100**	**100**	**100**	96.9	90.1	**100**	**100**
			Clayton	*4.9*	*5.2*	*5.3*	*5.7*	*5.6*	*5.2*	*5.6*	*4.8*	*5.5*	*5.1*	*5.4*
			Gumbel	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
			Frank	**100**	99.0	**100**	99.9	**100**	**100**	**100**	**100**	93.6	**100**	**100**
8	100	0.2	Gaussian	14.3	12.6	29.9	9.9	21.4	53.5	**82.6**	8.1	6.6	74.2	22.3
			Student-t	57.8	61.0	44.3	40.9	20.2	54.3	65.9	9.3	8.6	**85.5**	24.4
			Clayton	*5.5*	*5.0*	*5.2*	*5.5*	*5.6*	*5.4*	*4.3*	*4.7*	*5.2*	*5.1*	*5.5*
			Gumbel	7.6	10.5	63.2	52.6	91.9	**100**	98.0	68.7	26.5	97.0	90.8
			Frank	3.2	6.0	16.6	4.2	74.8	96.5	**96.7**	20.4	6.3	93.4	68.9
		0.4	Gaussian	97.5	91.7	96.9	2.5	87.1	89.0	98.2	34.8	10.9	**99.1**	90.2
			Student-t	86.3	80.5	**98.4**	29.5	86.1	89.4	96.0	32.4	10.7	97.7	91.4
			Clayton	*5.7*	*5.4*	*4.8*	*5.1*	*4.7*	*4.8*	*4.6*	*5.3*	*5.0*	*4.7*	*4.7*
			Gumbel	93.0	82.2	99.8	19.9	**100**	**100**	**100**	97.3	43.4	**100**	**100**
			Frank	85.2	62.8	93.7	0.6	99.6	97.7	**100**	76.5	8.1	**100**	99.6
	500	0.2	Gaussian	**100**	71.6	**100**	24.9	98.9	97.4	**100**	41.8	17.0	**100**	99.5
			Student-t	**100**	**100**	**100**	99.3	96.7	98.1	**100**	50.8	32.0	**100**	99.3
			Clayton	*5.3*	*4.8*	*5.0*	*4.8*	*4.9*	*5.3*	*4.6*	*5.3*	*5.4*	*5.4*	*4.7*
			Gumbel	98.3	40.7	**100**	96.6	**100**	**100**	**100**	**100**	96.8	**100**	**100**
			Frank	99.9	11.0	**100**	3.7	**100**	**100**	**100**	92.8	15.5	**100**	**100**
		0.4	Gaussian	**100**	**100**	**100**	96.1	**100**	**100**	**100**	98.7	84.4	**100**	**100**
			Student-t	**100**	**100**	**100**	93.2	**100**	**100**	**100**	98.7	78.1	**100**	**100**
			Clayton	*4.5*	*4.8*	*4.8*	*4.9*	*4.9*	*5.2*	*5.1*	*5.5*	*4.9*	*4.8*	*4.8*
			Gumbel	**100**	**100**	**100**	88.5	**100**	**100**	**100**	**100**	**100**	**100**	**100**
			Frank	**100**	**100**	**100**	69.5	**100**	**100**	**100**	**100**	76.0	**100**	**100**

Note: The Student copula alternative hypothesis with degree-of-freedom $\nu = 6$. Numbers in *italic* are nominal levels and should correspond to the prescribed size of 5%. Numbers in **bold** indicates the best performing approach. All powers are size-adjusted.

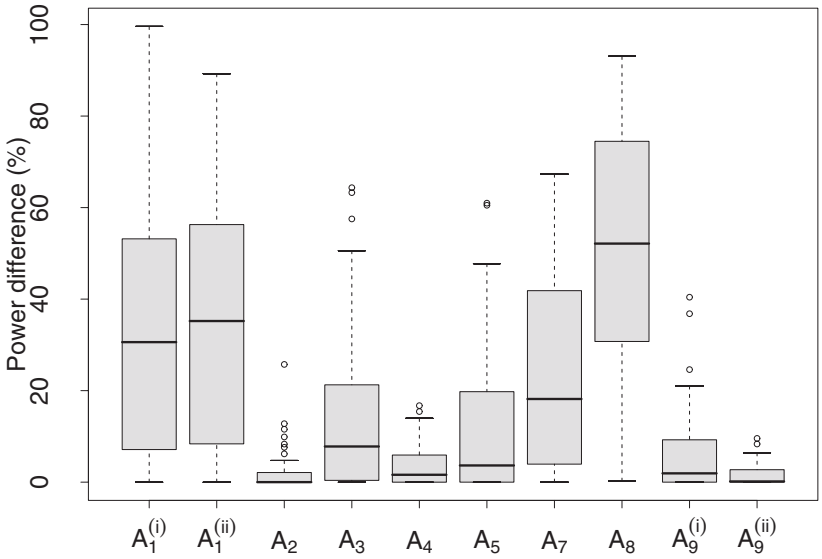


Figure IV.4: Distribution of power difference from the very best approach for testing the Gumbel copula.

Testing the Gumbel hypothesis

We now test the Gumbel hypothesis. The results are shown in Table IV.4 and the power differences in Figure IV.4. Notice that the nominal levels match the prescribed size well. Note also, again, the very good performance of approaches $\mathcal{A}_2$ and $\mathcal{A}_4$. Finally, approach $\mathcal{A}_9^{(ii)}$ perform very well. This is not surprising since it is the average of $\mathcal{A}_2$, $\mathcal{A}_3$ and $\mathcal{A}_4$.

Testing the Frank hypothesis

Finally, we test the Frank hypothesis. The results are shown in Table IV.5 and the power differences in Figure IV.5. The nominal levels match the prescribed size well. Note again the very good performance of approach $\mathcal{A}_2$. Approaches $\mathcal{A}_4$ and $\mathcal{A}_9^{(ii)}$ also perform very well.

IV.4.2 Effect of permutation order for Rosenblatt's transform

Approaches $\mathcal{A}_1$, $\mathcal{A}_3$ and $\mathcal{A}_8$ are all based on Rosenblatt's transform and a consecutive test of independence. The lack of invariance to the order of permutation may pose a problem to these approaches in the sense that the statistic for a given data set may prove very different depending on the permutation order. This is an undesirable feature of a statistical testing procedure. However, the practical consequence of this permutation invariance has not yet been investigated.

Table IV.6 shows the effect of permutation order on the estimated p-value for the three approaches based on Rosenblatt's transformation. The reported values are means and stan-

Table IV.4: Percentage of rejections (at 5% significance level) of the Gumbel copula by approaches $\mathcal{A}_1$-$\mathcal{A}_9$.

d	n	τ	*True copula*	$\mathcal{A}_1^{(i)}$	$\mathcal{A}_1^{(ii)}$	$\mathcal{A}_2$	$\mathcal{A}_3$	$\mathcal{A}_4$	$\mathcal{A}_5$	$\mathcal{A}_6$	$\mathcal{A}_7$	$\mathcal{A}_8$	$\mathcal{A}_9^{(i)}$	$\mathcal{A}_9^{(ii)}$
2	100	0.2	Gaussian	7.7	6.6	9.9	7.3	9.6	9.6	–	6.4	6.6	**10.2**	9.8
			Student-t	7.1	6.2	**11.2**	9.8	9.0	7.6	–	5.9	6.2	8.8	10.4
			Clayton	5.9	6.5	45.8	31.1	44.0	35.1	–	12.3	10.8	33.1	**47.5**
			Gumbel	*5.3*	*5.1*	*5.1*	*4.9*	*5.1*	*5.1*	–	*5.1*	*5.3*	*5.1*	*4.9*
			Frank	6.7	5.2	12.1	8.0	11.3	**13.3**	–	7.4	6.8	10.4	11.7
		0.4	Gaussian	11.4	11.2	17.5	8.9	16.4	13.7	–	8.1	7.2	**19.1**	17.6
			Student-t	5.8	6.2	**20.2**	15.2	16.1	11.3	–	7.5	6.7	13.9	19.7
			Clayton	8.1	14.0	**92.6**	75.4	89.8	75.3	–	34.7	31.4	83.4	**92.6**
			Gumbel	*4.8*	*4.6*	*4.8*	*5.1*	*4.9*	*4.7*	–	*4.7*	*5.2*	*4.8*	*5.0*
			Frank	8.1	7.1	**28.7**	9.4	24.8	24.3	–	10.3	9.0	20.9	25.7
	500	0.2	Gaussian	19.9	9.8	**37.0**	23.9	29.2	26.9	–	11.7	10.2	31.4	33.1
			Student-t	16.6	11.6	**39.1**	33.7	25.2	17.3	–	11.8	10.2	27.7	30.8
			Clayton	8.4	10.3	**99.6**	98.5	98.5	95.9	–	57.5	51.5	97.1	99.3
			Gumbel	*4.7*	*4.6*	*5.1*	*4.8*	*4.6*	*5.1*	–	*5.0*	*4.6*	*4.6*	*4.6*
			Frank	16.0	7.4	**53.9**	30.7	38.5	42.6	–	16.2	12.7	37.1	44.3
		0.4	Gaussian	49.9	32.4	**74.1**	38.4	61.6	46.8	–	25.4	28.9	73.8	67.7
			Student-t	9.0	10.8	**74.1**	56.7	57.3	36.0	–	20.9	21.1	53.0	68.4
			Clayton	43.6	57.8	**100**	**100**	**100**	**100**	–	99.3	99.6	**100**	**100**
			Gumbel	*5.4*	*4.9*	*5.2*	*5.5*	*5.0*	*5.0*	–	*4.8*	*5.2*	*5.0*	*4.9*
			Frank	45.3	13.8	**95.5**	47.8	85.1	82.2	–	44.4	42.1	86.2	89.2
4	100	0.2	Gaussian	6.8	13.0	54.7	43.4	51.1	24.0	–	14.9	7.5	41.6	**57.3**
			Student-t	24.9	24.8	56.8	55.7	52.8	21.1	–	13.0	8.8	58.7	**60.1**
			Clayton	3.4	15.1	89.6	85.4	97.1	82.2	–	29.9	10.1	90.6	**97.2**
			Gumbel	*5.0*	*4.9*	*5.0*	*4.5*	*5.0*	*5.3*	–	*5.0*	*5.6*	*4.8*	*5.0*
			Frank	4.6	5.4	22.2	13.1	29.2	**30.6**	–	12.6	5.5	18.6	30.0
		0.4	Gaussian	29.7	36.6	66.7	44.0	59.9	33.7	–	28.8	9.2	**70.5**	65.0
			Student-t	15.1	22.0	68.0	66.1	60.7	30.2	–	26.2	9.9	60.0	**68.9**
			Clayton	26.8	29.9	99.9	99.1	**100**	98.8	–	82.4	32.8	99.8	**100**
			Gumbel	*5.0*	*5.0*	*5.0*	*5.2*	*5.1*	*5.1*	–	*5.0*	*5.4*	*5.5*	*5.0*
			Frank	17.8	9.0	51.4	12.5	54.3	**56.1**	–	26.2	7.3	46.5	53.7
	500	0.2	Gaussian	75.9	59.1	**99.4**	98.5	98.3	96.0	–	68.4	19.5	**99.4**	99.2
			Student-t	92.0	88.5	99.1	99.7	97.7	94.5	–	67.4	27.3	**100**	99.2
			Clayton	34.2	64.9	**100**	**100**	**100**	**100**	–	98.1	53.3	**100**	**100**
			Gumbel	*4.7*	*4.8*	*4.8*	*4.6*	*4.7*	*5.0*	–	*4.7*	*4.2*	*4.6*	*4.7*
			Frank	47.7	10.0	86.6	47.5	92.7	**98.1**	–	58.0	9.8	93.2	94.0
		0.4	Gaussian	99.9	98.2	**100**	99.7	99.6	97.6	–	95.9	54.8	**100**	99.9
			Student-t	86.1	91.3	**100**	**100**	99.6	97.1	–	93.9	60.2	**100**	**100**
			Clayton	**100**	95.7	**100**	**100**	**100**	**100**	–	**100**	99.8	**100**	**100**
			Gumbel	*4.7*	*5.1*	*4.9*	*5.3*	*5.1*	*4.8*	–	*4.6*	*5.1*	*4.8*	*5.2*
			Frank	99.4	31.8	99.9	58.9	99.8	**100**	–	93.0	23.7	**100**	99.9
8	100	0.2	Gaussian	1.0	30.0	89.8	73.2	87.1	29.9	–	37.6	6.7	50.0	**90.4**
			Student-t	52.3	70.3	89.4	76.6	86.2	30.9	–	36.1	8.3	**91.9**	89.9
			Clayton	0.2	29.9	93.6	95.4	**99.8**	81.2	–	53.3	8.6	89.3	99.7
			Gumbel	*5.4*	*5.1*	*4.1*	*4.8*	*4.9*	*4.8*	–	*4.6*	*5.1*	*5.1*	*4.8*
			Frank	0.3	4.3	14.6	10.3	**40.4**	19.4	–	28.4	5.5	3.6	36.8
		0.4	Gaussian	36.8	68.2	**98.1**	72.3	90.2	50.3	–	70.1	6.8	93.7	93.7
			Student-t	45.3	65.7	**97.8**	83.8	90.8	51.8	–	65.0	11.7	94.1	94.6
			Clayton	38.5	45.9	**100**	99.6	**100**	99.9	–	98.2	42.0	**100**	**100**
			Gumbel	*5.2*	*5.1*	*5.3*	*5.1*	*5.3*	*5.4*	–	*5.0*	*5.5*	*5.2*	*5.4*
			Frank	16.0	8.7	54.3	9.6	**67.1**	63.5	–	53.4	4.9	42.5	66.2
	500	0.2	Gaussian	99.9	99.1	**100**	**100**	**100**	**100**	–	99.2	14.8	**100**	**100**
			Student-t	**100**	**100**	**100**	**100**	**100**	**100**	–	98.9	31.7	**100**	**100**
			Clayton	79.4	98.9	**100**	**100**	**100**	**100**	–	**100**	33.0	**100**	**100**
			Gumbel	*5.1*	*4.9*	*4.1*	*4.8*	*5.1*	*5.2*	–	*4.3*	*4.8*	*5.2*	*5.0*
			Frank	78.6	18.6	90.1	36.7	99.9	**100**	–	93.7	7.0	99.2	99.9
		0.4	Gaussian	**100**	**100**	**100**	**100**	**100**	**100**	–	**100**	37.5	**100**	**100**
			Student-t	**100**	**100**	**100**	**100**	**100**	**100**	–	**100**	67.5	**100**	**100**
			Clayton	**100**	99.9	**100**	**100**	**100**	**100**	–	**100**	99.7	**100**	**100**
			Gumbel	*5.3*	*4.9*	*5.1*	*5.3*	*5.2*	*5.4*	–	*4.9*	*5.0*	*5.2*	*5.1*
			Frank	**100**	48.8	**100**	35.6	**100**	**100**	–	99.8	9.5	**100**	**100**

Note: Student copula alternative hypothesis with degree-of-freedom $\nu = 6$. Numbers in *italic* are nominal levels and should correspond to the prescribed size of 5%. Numbers in **bold** indicates the best performing approach. All powers are size-adjusted.

(IV.18)

Table IV.5: Percentage of rejections (at 5% significance level) of the Frank copula by approaches $\mathcal{A}_1$-$\mathcal{A}_9$.

d	n	τ	*True copula*	$\mathcal{A}_1^{(i)}$	$\mathcal{A}_1^{(ii)}$	$\mathcal{A}_2$	$\mathcal{A}_3$	$\mathcal{A}_4$	$\mathcal{A}_5$	$\mathcal{A}_6$	$\mathcal{A}_7$	$\mathcal{A}_8$	$\mathcal{A}_9^{(i)}$	$\mathcal{A}_9^{(ii)}$
2	100	0.2	Gaussian	5.8	5.5	6.0	**7.5**	6.9	6.6	–	4.9	5.1	6.2	7.4
			Student-t	10.6	8.4	8.8	9.9	8.9	7.9	–	6.0	5.7	**11.9**	10.1
			Clayton	5.1	5.3	24.4	21.3	26.2	18.5	–	7.9	7.4	17.4	**29.4**
			Gumbel	5.2	6.0	13.5	8.8	14.2	11.4	–	6.3	6.3	10.0	**14.9**
			Frank	*5.8*	*5.6*	*5.5*	*7.3*	*5.6*	*5.4*	–	*5.4*	*4.8*	*5.7*	*5.9*
		0.4	Gaussian	12.2	9.1	9.4	9.2	9.5	6.8	–	5.6	6.5	**13.1**	10.7
			Student-t	8.2	6.4	13.7	10.4	13.3	9.4	–	6.2	7.1	12.0	**14.7**
			Clayton	6.8	5.2	65.4	47.5	62.4	34.6	–	15.9	16.9	46.6	**68.2**
			Gumbel	6.5	6.0	**29.1**	9.6	26.0	15.7	–	8.4	9.1	18.0	26.6
			Frank	*5.9*	*4.8*	*4.9*	*6.3*	*5.2*	*4.7*	–	*4.1*	*5.1*	*5.3*	*5.3*
	500	0.2	Gaussian	7.6	6.7	11.2	**15.3**	10.3	10.3	–	6.7	7.3	10.3	11.8
			Student-t	47.8	26.9	28.0	20.5	26.5	25.2	–	12.4	13.4	**48.0**	29.2
			Clayton	7.6	7.1	87.7	81.0	84.2	66.4	–	27.5	27.5	74.3	**87.8**
			Gumbel	11.4	10.3	**55.6**	31.9	44.5	41.8	–	15.1	15.9	41.1	49.2
			Frank	*5.5*	*4.9*	*4.5*	*7.2*	*5.4*	*5.1*	–	*4.6*	*5.4*	*4.9*	*5.5*
		0.4	Gaussian	30.3	23.1	42.5	35.1	32.7	23.2	–	14.0	14.9	**47.5**	42.2
			Student-t	20.9	14.5	**68.5**	28.6	57.1	46.2	–	22.3	21.5	58.9	63.8
			Clayton	11.9	9.5	**100**	99.9	**100**	97.6	–	83.9	85.2	99.9	**100**
			Gumbel	9.9	12.2	**95.2**	47.5	85.8	77.3	–	41.7	41.2	81.2	89.9
			Frank	*6.0*	*4.8*	*4.2*	*6.4*	*4.7*	*4.0*	–	*4.6*	*5.0*	*4.9*	*5.0*
4	100	0.2	Gaussian	4.8	9.3	27.6	27.0	24.8	10.3	–	6.9	6.9	18.2	**29.8**
			Student-t	44.0	25.9	40.0	41.1	36.8	20.3	–	8.2	7.7	**59.2**	44.5
			Clayton	6.5	8.5	68.0	75.0	87.1	41.9	–	13.2	8.5	71.9	**88.4**
			Gumbel	10.2	5.3	19.6	3.9	33.8	**50.5**	–	11.2	7.2	27.3	31.1
			Frank	*5.5*	*5.3*	*4.5*	*4.9*	*4.8*	*4.7*	–	*5.2*	*5.1*	*5.2*	*4.8*
		0.4	Gaussian	14.1	29.4	30.1	33.1	31.3	18.4	–	10.8	7.6	**43.9**	37.3
			Student-t	18.5	16.7	47.4	53.0	43.3	29.2	–	13.0	9.3	49.8	**53.6**
			Clayton	4.5	9.8	95.5	97.5	98.0	62.1	–	47.1	19.4	93.8	**98.8**
			Gumbel	9.7	5.1	58.0	7.2	54.7	**65.3**	–	21.3	9.1	44.0	56.6
			Frank	*5.6*	*4.8*	*5.4*	*5.4*	*5.3*	*5.7*	–	*5.2*	*4.6*	*5.4*	*5.5*
	500	0.2	Gaussian	13.4	38.1	**86.1**	79.1	66.0	57.7	–	19.8	15.9	77.3	76.2
			Student-t	99.0	90.2	97.4	95.7	88.3	88.7	–	34.3	27.9	**99.9**	95.2
			Clayton	11.2	31.1	**100**	**100**	**100**	99.7	–	66.7	37.3	**100**	**100**
			Gumbel	26.6	7.8	84.7	22.0	91.9	**97.5**	–	56.8	25.5	91.2	92.5
			Frank	*5.6*	*5.4*	*5.1*	*4.9*	*4.4*	*5.6*	–	*4.9*	*5.0*	*5.8*	*4.5*
		0.4	Gaussian	78.9	93.7	98.3	95.3	90.9	74.2	–	58.9	40.3	**99.9**	95.7
			Student-t	72.0	78.8	99.9	99.6	98.6	95.8	–	72.2	52.2	**100**	99.6
			Clayton	8.0	36.9	**100**	**100**	**100**	**100**	–	99.9	96.5	**100**	**100**
			Gumbel	35.0	6.9	**99.9**	51.9	99.7	**99.9**	–	91.5	54.4	99.7	99.8
			Frank	*4.9*	*5.1*	*5.3*	*6.0*	*5.0*	*5.1*	–	*5.7*	*4.8*	*5.0*	*5.3*
8	100	0.2	Gaussian	1.0	20.5	**81.2**	68.2	60.8	12.5	–	11.2	6.3	26.9	72.6
			Student-t	75.6	68.9	84.6	73.1	69.2	27.1	–	12.6	7.9	**94.3**	79.5
			Clayton	2.6	15.5	83.6	94.6	**97.7**	36.5	–	22.7	8.6	79.5	97.4
			Gumbel	20.3	5.0	35.7	22.2	63.2	**87.7**	–	39.8	7.8	43.7	60.4
			Frank	*4.5*	*5.1*	*4.7*	*5.2*	*4.8*	*4.8*	–	*5.5*	*5.1*	*4.9*	*4.8*
		0.4	Gaussian	11.7	62.0	**93.6**	81.4	60.1	24.2	–	25.7	8.2	78.1	73.4
			Student-t	47.8	55.9	**95.2**	91.3	74.1	38.4	–	28.3	10.8	90.9	86.2
			Clayton	1.3	18.1	98.7	99.8	**99.9**	69.4	–	81.0	39.4	98.5	**99.9**
			Gumbel	26.5	7.9	72.8	29.5	74.7	**93.7**	–	50.3	11.0	67.6	77.0
			Frank	*5.0*	*4.8*	*4.6*	*5.2*	*5.1*	*5.5*	–	*4.7*	*4.4*	*4.9*	*5.0*
	500	0.2	Gaussian	47.7	94.1	**100**	**100**	99.8	99.0	–	66.6	15.1	**100**	**100**
			Student-t	**100**	**100**	**100**	**100**	**100**	**100**	–	77.4	32.3	**100**	**100**
			Clayton	6.3	82.8	**100**	**100**	**100**	**100**	–	93.7	35.8	**100**	**100**
			Gumbel	71.4	6.0	95.9	74.3	**100**	**100**	–	98.5	34.1	98.9	**100**
			Frank	*4.5*	*4.8*	*4.3*	*5.1*	*5.2*	*5.3*	–	*5.6*	*5.3*	*5.5*	*5.1*
		0.4	Gaussian	**100**	**100**	**100**	**100**	99.9	93.1	–	97.6	37.9	**100**	**100**
			Student-t	**100**	**100**	**100**	**100**	**100**	99.7	–	98.6	61.5	**100**	**100**
			Clayton	8.3	83.7	**100**	**100**	**100**	**100**	–	**100**	99.6	**100**	**100**
			Gumbel	93.3	16.3	**100**	95.1	**100**	**100**	–	99.9	62.5	**100**	**100**
			Frank	*5.0*	*4.6*	*4.7*	*4.9*	*4.6*	*4.2*	–	*5.3*	*4.7*	*4.4*	*4.6*

Note: Student copula alternative hypothesis with degree-of-freedom $\nu = 6$. Numbers in *italic* are nominal levels and should correspond to the prescribed size of 5%. Numbers in **bold** indicates the best performing approach. All powers are size-adjusted.

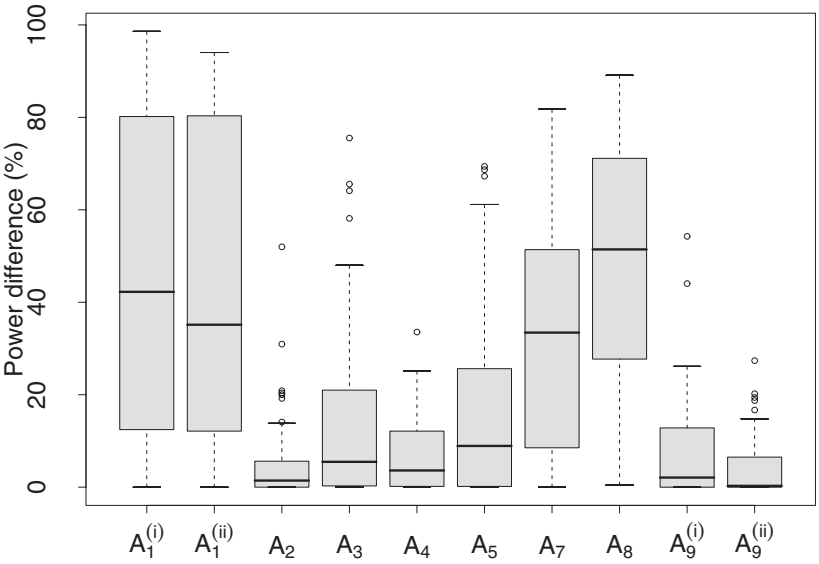


Figure IV.5: Distribution of power difference from the very best approach for testing the Frank copula.

dard deviations of the estimated p-values (over $d!$ permutations). The study is restricted to dimension $d = 5$ for which there are $d! = 120$ different permutations, sample size $n = 100$ and dependence $\tau = 0.5$. All reported values are averaged over 1000 independent simulations. For some of the approaches there are two sources of variation; permutation order and double bootstrap procedure. In order to see the effect of permutation order only, we report the same p-value variation results when the permutation is kept fixed, see Table IV.7.

From the two tables one can see that the permutation order adds no variance for approach $\mathcal{A}_1^{(i)}$ when the null hypothesis is the Gaussian copula. This permutation invariance of approach $\mathcal{A}_1^{(i)}$ under the Gaussian null hypothesis is proved in Appendix IV.A. However, when using a different weight function or when the null hypothesis is different from the Gaussian copula, variation is added due to the permutation order. Note that in- or close to rejection regions, the variation due to permutation order is as great as in other regions, relative to the mean. However, the practical effect will not be so important as the conclusion will most probably be the same, regardless of permutation order. We see the same for the other approaches. When the null- and alternative hypotheses are the same we see that the average of the mean p-values are approximately 0.5 as they should be. We also see that the variation in these cases are quite large, typically around 0.25 for approaches $\mathcal{A}_1^{(ii)}$, $\mathcal{A}_3$ and $\mathcal{A}_8$. For approach $\mathcal{A}_1^{(i)}$ we see that the variation is in general lower than for the other approaches. Also note that for approach $\mathcal{A}_8$ the permutation order adds almost no variation in any case as the estimated p-value will vary heavily even when keeping the permutation order fixed. This is due to the construction of the approach. Random samples

Table IV.6: Estimated mean p-value (mean of $d!$ permutations) for approaches based on Rosenblatt's transformation. In parentheses the standard deviation over all permutations is given. All quoted values are averaged over 100 simulations.

$\mathcal{H}_0$	$\mathcal{H}_1$	$\mathcal{A}_1^{(i)}$	$\mathcal{A}_1^{(ii)}$	$\mathcal{A}_3$	$\mathcal{A}_8$
Gaussian	Gaussian	0.514 (0.000)	0.520 (0.263)	0.513 (0.287)	0.510 (0.290)
	Clayton	0.501 (0.000)	0.480 (0.239)	0.021 (0.038)	0.205 (0.201)
	Gumbel	0.479 (0.000)	0.460 (0.237)	0.549 (0.294)	0.294 (0.247)
	Frank	0.415 (0.000)	0.419 (0.232)	0.535 (0.311)	0.428 (0.287)
Clayton	Gaussian	0.003 (0.002)	0.008 (0.015)	0.312 (0.187)	0.248 (0.237)
	Clayton	0.520 (0.159)	0.535 (0.263)	0.519 (0.269)	0.501 (0.283)
	Gumbel	0.002 (0.002)	0.016 (0.024)	0.370 (0.222)	0.103 (0.139)
	Frank	0.008 (0.004)	0.040 (0.051)	0.424 (0.226)	0.265 (0.242)
Gumbel	Gaussian	0.082 (0.027)	0.095 (0.118)	0.109 (0.100)	0.390 (0.279)
	Clayton	0.035 (0.012)	0.214 (0.181)	0.000 (0.001)	0.101 (0.129)
	Gumbel	0.533 (0.110)	0.533 (0.270)	0.528 (0.264)	0.506 (0.287)
	Frank	0.113 (0.034)	0.340 (0.239)	0.417 (0.246)	0.463 (0.286)
Frank	Gaussian	0.242 (0.102)	0.129 (0.152)	0.104 (0.086)	0.380 (0.274)
	Clayton	0.536 (0.153)	0.400 (0.248)	0.000 (0.001)	0.173 (0.184)
	Gumbel	0.396 (0.135)	0.492 (0.265)	0.325 (0.227)	0.365 (0.267)
	Frank	0.509 (0.151)	0.508 (0.272)	0.506 (0.245)	0.486 (0.281)

Note: Applied to $n = 100$ samples of $d = 5$ dimensional copulae with dependence parameter $\tau = 0.5$.

from the null hypothesis copula are drawn in every computation of the statistic, inducing large variation, at least when we are far from rejection regions.

To illustrate further, we look at so-called mixing tests. Two copulae are mixed in the following way:

$$C^{\text{mix}} = (1 - \beta)C_1 + \beta C_2,$$

where $\beta \in [0, 1]$ is the mixing parameter. We consider the case where C_1 is the Clayton copula while C_2 is the Gumbel copula. So when $\beta = 0$, the mixed copula is equivalent to the Clayton copula, while when $\beta = 1$ it is equivalent to the Gumbel copula. We draw $n = 500$ random samples from the $d = 5$ dimensional mixed copula with dependences $\tau_1 = \tau_2 = 0.4$. We then estimate the p-value under a Clayton null hypothesis for all values of β, using approaches $\mathcal{A}_1^{(i)}$, $\mathcal{A}_1^{(ii)}$, $\mathcal{A}_3$ and $\mathcal{A}_8$, i.e. all approaches based on Rosenblatt's transformation. The p-value is estimated for each of the $d!$ permutations of the variables and the 95% confidence interval over the $d!$ permutations is computed. This is repeated 1000 times and Figure IV.6 shows the resulting confidence intervals, averaged over the 1000 repetitions. Included in the figure are also the corresponding confidence intervals when the permutation order is kept fixed. This way we can see the additional p-value variation solely due to permutation order. We see that for approach $\mathcal{A}_1^{(i)}$ the additional variation is substantial when the null hypothesis is true. However, as we move towards rejection, the additional variation becomes negligible in the sense that the conclusion will be the same no matter which permutation order is chosen. Again we note that the additional variation due to permutation order is smaller for approach $\mathcal{A}_1^{(i)}$ than for the other approaches based on Rosenblatt's transformation. Note also, that for approach $\mathcal{A}_1^{(ii)}$ there is p-value variation even when the permutation is kept fixed. This is due to the double parametric bootstrap step concerned with the approximation of F_1 in (IV.3.1). This is also the case for approach $\mathcal{A}_8$ where we see only marginal additional variation due to permutation order.

Finally, we examine whether the utilization of all permutations may give us additional power. The idea is that by computing a statistic for each permutation of the data, more

Table IV.7: Estimated mean p-value (mean of $d!$ separate estimations based on the same data set) for approaches based on Rosenblatt's transformation. In parentheses the standard deviation over all permutations is given. All quoted values are averaged over 100 simulations.

$\mathcal{H}_0$	$\mathcal{H}_1$	$\mathcal{A}_1^{(i)}$	$\mathcal{A}_1^{(ii)}$	$\mathcal{A}_3$	$\mathcal{A}_8$
Gaussian	Gaussian	0.514 (0.000)	0.530 (0.057)	0.523 (0.000)	0.510 (0.284)
	Clayton	0.501 (0.000)	0.483 (0.056)	0.021 (0.000)	0.205 (0.194)
	Gumbel	0.479 (0.000)	0.458 (0.052)	0.559 (0.000)	0.294 (0.239)
	Frank	0.415 (0.000)	0.416 (0.048)	0.551 (0.000)	0.432 (0.282)
Clayton	Gaussian	0.002 (0.000)	0.008 (0.003)	0.318 (0.000)	0.250 (0.216)
	Clayton	0.517 (0.000)	0.535 (0.056)	0.524 (0.000)	0.501 (0.275)
	Gumbel	0.002 (0.000)	0.013 (0.003)	0.382 (0.000)	0.105 (0.125)
	Frank	0.008 (0.000)	0.038 (0.007)	0.436 (0.000)	0.262 (0.218)
Gumbel	Gaussian	0.080 (0.000)	0.089 (0.023)	0.104 (0.000)	0.390 (0.268)
	Clayton	0.036 (0.000)	0.205 (0.036)	0.000 (0.000)	0.100 (0.123)
	Gumbel	0.527 (0.000)	0.531 (0.061)	0.532 (0.000)	0.508 (0.281)
	Frank	0.112 (0.000)	0.342 (0.050)	0.421 (0.000)	0.461 (0.278)
Frank	Gaussian	0.240 (0.000)	0.129 (0.031)	0.109 (0.000)	0.381 (0.263)
	Clayton	0.541 (0.000)	0.395 (0.055)	0.000 (0.000)	0.170 (0.174)
	Gumbel	0.391 (0.000)	0.489 (0.059)	0.320 (0.000)	0.366 (0.257)
	Frank	0.502 (0.000)	0.510 (0.063)	0.501 (0.000)	0.485 (0.274)

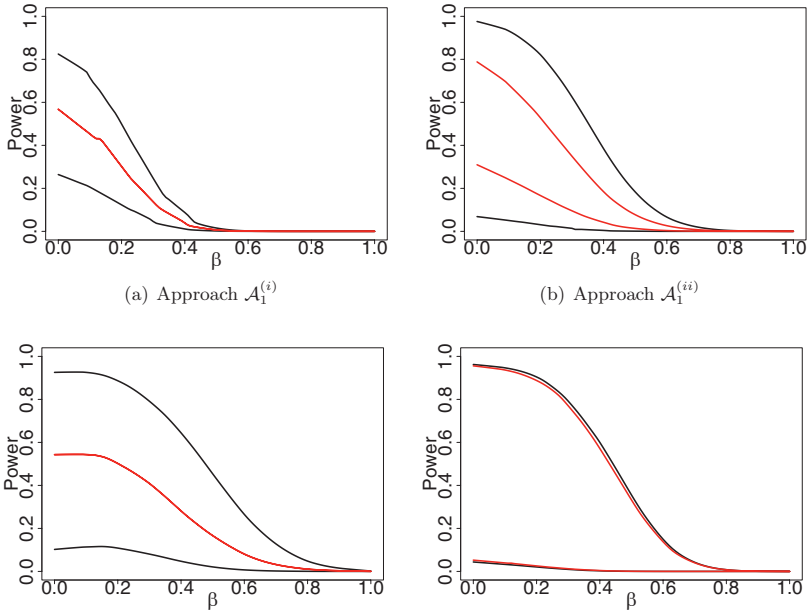
Note: Applied to $n = 100$ samples of $d = 5$ dimensional copulae with dependence parameter $\tau = 0.5$.

information is extracted from the data and we may achieve higher power. This is investigated for approaches $\mathcal{A}_1^{(i)}$ and $\mathcal{A}_3$ in the case $d = 4$, $n = 100$, $\tau = 0.4$ for $\mathcal{H}_0$ and $\mathcal{H}_1$ being one of the Gaussian-, Clayton-, Gumbel- or Frank copulae. We simply compute the average of the statistics over the $d!$ permutations. Table IV.8 shows the results, along with corresponding results (permutation fixed) from Tables IV.1, IV.3, IV.4 and IV.5. We see that averaging over all $d!$ permutations adds some power, e.g. for $\mathcal{A}_3$ for $\mathcal{H}_0$ =Gaussian, $\mathcal{H}_1$ =Clayton where the power increases from 81% to 95%. Hence, this might be a fruitful idea to pursue in future research. Perhaps one can find clever ways of averaging only over a few of the $d!$ permutations, and still gain most of the power increase.

IV.5 Discussion and recommendations

An overview of six copula GoF approaches was given, along with the proposal of three new approaches. A large Monte Carlo study was presented, examining the nominal levels and the power against some fixed alternatives under several combinations of problem dimension, sample size and dependence.

Results show, in general, increasing power with dimension, sample size and dependence, which is expected. Further, the results show that approach $\mathcal{A}_2$, the approach based on a distance between the empirical- and null hypothesis copula distribution functions, is in general the best approach, with approach $\mathcal{A}_4$ as a strong runner up. However, in some cases, e.g. when testing the Gaussian hypothesis against heavy tails, approach $\mathcal{A}_2$ does not perform so well. In this case, however, the otherwise poor approach $\mathcal{A}_1$ performs very well for high dimensions and large sample sizes. However, in general, approach $\mathcal{A}_2$ is recommended. One should consider supplmenting it with approaches $\mathcal{A}_4$ and $\mathcal{A}_1^{(i)}$, in particular if no strong a priori opinions exist as to which distribution we are testing for and what kind of deviations to detect. Average approaches merge the qualities of all the approaches included in the averaging and provides more stable power performance than the



(a) Approach $\mathcal{A}_1^{(i)}$ (b) Approach $\mathcal{A}_1^{(ii)}$

(c) Approach $\mathcal{A}_3$ (d) Approach $\mathcal{A}_8$

Figure IV.6: P-value variation due to permutation order for approaches based on Rosenblatt's transformation. Average 95% confidence intervals over 1000 separate mixed copula simulations. The null hypothesis is the Clayton copula while the alternative hypothesis is the Gumbel copula. The black lines represent the variation over $d!$ permutation orders while the red lines represent the variation when permutation is kept fixed.

individual approaches. However, the topic of averaging different approaches was included as a hint of further research and needs more work. Finally, to decide which approaches to consider, a preliminary test of ellipticity (see e.g. Huffera and Park (2007)) can be helpful.

When doing model evaluation, it is recommended to also examine various diagnostic tests such as GoF plots, e.g. plotting $S_4(w)$ with simulated null hypothesis confidence bands as done in Genest et al. (2006a). This may give valuable information on the fit of a copula. However, there is still a need for intuitive and informative diagnostic plots. Ideally such a plot should show, in some way and in case of rejection by the formal tests, which variable (i.e. which dimension) and/or which samples causes the rejection. Is it actually a deviation in the dependence structure between the variables or some extreme samples that cause the rejection? More research is needed on this topic.

Next, results were reported on the variation of the p-value estimates due to permutation order for approaches based on Rosenblatt's transformation. In general, one does not want a statistical testing procedure to give different values when running it several times on the same data set. However, for some of the approaches based on Rosenblatt's transformation, the estimated p-value will be different depending on which permuation order that

Table IV.8: Percentage of rejections (at 5% significance level) by approaches $\mathcal{A}_1^{(i)}$ and $\mathcal{A}_3$ when computing the average of $\widehat{T}_1^{(i)}$ and $\widehat{T}_3$ over all $d!$ permutations. These are denoted by $\mathcal{A}_{1,d!}^{(i)}$ and $\mathcal{A}_{3,d!}$ and they are compared to corresponding rejection rates for the original approaches $\mathcal{A}_1^{(i)}$ and $\mathcal{A}_3$, that only consider one, fixed permutation.

$\mathcal{H}_0$	$\mathcal{H}_1$	$\mathcal{A}_1^{(i)}$	$\mathcal{A}_{1,d!}^{(i)}$	$\mathcal{A}_3$	$\mathcal{A}_{3,d!}$
Gaussian	Gaussian	*4.7*	*5.2*	*5.1*	*4.9*
	Clayton	1.0	1.0	80.8	94.9
	Gumbel	1.5	1.7	3.6	2.2
	Frank	1.6	1.8	7.3	6.6
Clayton	Gaussian	78.3	83.4	3.0	4.7
	Clayton	*5.2*	*5.7*	*5.5*	*5.1*
	Gumbel	79.1	83.2	4.9	6.1
	Frank	68.7	74.5	3.2	3.4
Gumbel	Gaussian	29.7	30.0	44.0	62.9
	Clayton	26.8	26.0	99.1	99.9
	Gumbel	*5.0*	*5.3*	*5.2*	*5.2*
	Frank	17.8	18.4	12.5	18.3
Frank	Gaussian	14.1	14.6	33.1	48.4
	Clayton	4.5	4.0	97.5	99.5
	Gumbel	9.7	8.9	7.2	6.8
	Frank	*5.6*	*5.1*	*5.4*	*5.3*

Note: Dimension $d = 4$, sample size $n = 100$ and dependence $\tau = 0.4$. Numbers in *italic* are nominal levels and should correspond to the prescribed size of 5%. All powers are size-adjusted.

is chosen. This effect decreases as the p-value estimates approach critical levels. Hence, the author does not believe that the permutation effect is something to worry about. On the contrary, the permutational invariance may actually be useful, as seen when averaging over all permutations increases the power of some of the approaches. Also, as long as the permutation order is chosen randomly, the results are not influenced in any particular direction.

The results concerning the permutation of variables also point in direction of important future research. The variation of p-value estimates also depends on the bootstrap parameters M and N_b. These parameters are usually, in a rather arbitrary way, set to what is believed to be large values. This is also the case in this paper. However, there has been no study of the effect these choices may have on the power, and even more importantly the nominal levels of an approach. Originally, in the power studies of Section IV.4.1, a double bootstrap parameter $N_b = 2500$ was chosen for all combinations of dimension, sample size, dependence and alternative copula. However, for dimension $d = 8$ we observed some peculiar results, e.g. decreasing power as sample size increased. These peculiarities vanished when increasing N_b to 5000 for dimension $d = 8$. Choosing appropriately large values for these parameters and thus achieving proper nominal levels is crucial for any study and/or application of these GoF approaches. Hence, a study of the effects of these parameters and required minimum values would be highly interesting.

The computational aspect also deserves some attention. An important quality of approaches based on Rosenblatt's transform is computational efficiency. Approaches $\mathcal{A}_2$, $\mathcal{A}_4$ and $\mathcal{A}_5$ need computationally intensive double parametric bootstrap procedures to estimate p-values in some cases (e.g. for the elliptical copulae, in particular for higher dimensions and large sample sizes). Approaches based on Rosenblatt's transformation does not, in general, need this double bootstrap step, since after Rosenblatt's transformation, the null

hypothesis is always independence.

Finally, a word of warning. As emphasized in Genest et al. (2008), the asymptotics of several of the procedures presented here are not known. Hence, one cannot know for sure whether a bootstrap procedure will converge in every case. However, all the results so far on the performance of the proposed approaches and bootstrap procedures are comforting and strongly indicate the validity of the test procedures. Keep in mind though, the original approach and test procedure proposed by Breymann et al. (2003), which showed terrible performance in the study of Dobrić and Schmid (2007). This shows how wrong it can all go if our test procedure is not valid. Approaches $\mathcal{A}_2$ and $\mathcal{A}_4$, that turned out to be the best in our study, both have known asymptotics and the bootstrap procedures for these approaches are well established from Quessy (2005), Genest et al. (2006a) and Genest and Rémillard (2008). Hence, for the time being, these are the recommended for copula goodness-of-fit testing.

Acknowledgements: The author acknowledges 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. I would especially like to express my gratitude to Kjersti Aas for introducing me to the field of copulae and copula goodness-of-fit testing and for her continuous drive and interest. Credit is also due to my friend and former colleague Henrik Bakken for fruitful discussions and collaboration in the early phase of this and related work. Finally, I would like to thank colleagues and participants at various workshops and conferences for valuable comments, in particular Professor Nils Lid Hjort at the University of Oslo, Professor Jean-François Quessy at Université du Québec à Trois-Rivières, Professor Mark Salmon at Warwick Business School and Professor Alexander McNeil at Heriott-Watt University.

IV.A Proof of permutation invariance of $\mathcal{A}_1^{(i)}$ under Gaussian copula null hypothesis

To prove that approach $\mathcal{A}_1^{(i)}$ is permutatoin invariant under the Gaussian copula null hypothesis, let us first look at how Rosenblatt's transformation is carried out. For the Gaussian copula null hypothesis, this transformation is easily computed using the Cholesky decomposition of the covariance matrix. Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be a d-dimensional vector, where $\boldsymbol{\mu} = \mathbb{E}(\mathbf{X})$ and $\boldsymbol{\Sigma}$ is the $d \times d$ positive definite covariance matrix.

Since $\boldsymbol{\Sigma}$ is positive definite it can be written as $\boldsymbol{\Sigma} = \mathbf{A}^T\mathbf{A}$, where $\mathbf{A}$ is a lower triangular matrix and $\mathbf{A}^T$ denotes its transpose. Next, it is well known that $\mathbf{X}$ can be expressed as $\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}^T\mathbf{Y}$ where $\mathbf{Y} \sim \mathcal{N}(0, \boldsymbol{I})$ and $\boldsymbol{I}$ is the d-dimensional identity matrix. I.e. $\mathbf{Y}$ is a vector of d i.i.d. standard normally distributed variables. Solving for $\mathbf{Y}$ gives $\mathbf{Y} = (\mathbf{A}^T)^{-1}(\mathbf{X} - \boldsymbol{\mu})$. We now see that the vector $\mathbf{V} = \Phi(\mathbf{Y})$ is i.i.d. $U(0,1)^d$ under the Gaussian null hypothesis.

For approach $\mathcal{A}_1^{(i)}$ we now need to compute $W_1 = \sum_{i=1}^d \Phi^{-1}(V_i)^2 = \sum_{i=1}^d Y_i^2 = \mathbf{Y}^T\mathbf{Y}$. We now proceed with the bivariate setting for simplicity but the proof can easily be extended to arbitrary dimension d. Consider the Cholesky decomposition of the covariance matrix $\boldsymbol{\Sigma} = \mathbf{A}^T\mathbf{A}$ in detail:

$$\boldsymbol{\Sigma}^1 = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix} \begin{pmatrix} a_{11} & 0 \\ a_{12} & a_{22} \end{pmatrix} = \begin{pmatrix} a_{11}^2 + a_{12}^2 & a_{12}a_{22} \\ a_{12}a_{22} & a_{22}^2 \end{pmatrix},$$

where the superscript 1 in $\boldsymbol{\Sigma}^1$ denotes permutation order 1. We see now that $a_{11} = \sqrt{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}/\sigma_2$, $a_{12} = \sigma_{12}/\sigma_2$ and $a_{22} = \sigma_2$. Next, we see that

$$(\mathbf{A}^T)^{-1} = \begin{pmatrix} \frac{1}{a_{11}} & -\frac{a_{12}}{a_{11}a_{22}} \\ 0 & \frac{1}{a_{22}} \end{pmatrix}$$

and that

$$\mathbf{Y} = (\mathbf{A}^T)^{-1}(\mathbf{X} - \boldsymbol{\mu}) = \begin{pmatrix} \frac{1}{a_{11}}(X_1 - \mu_1) - \frac{a_{12}}{a_{11}a_{22}}(X_2 - \mu_2) \\ \frac{1}{a_{22}}(X_2 - \mu_2) \end{pmatrix}.$$

Now to compute $W_1^1 = \mathbf{Y}^T\mathbf{Y}$, superscript 1 denoting permutation order 1, we get

$$\begin{aligned} W_1^1 &= \frac{(X_1 - \mu_1)^2}{a_{11}^2} + \frac{a_{12}^2}{a_{11}^2 a_{22}^2}(X_2 - \mu_2)^2 - \frac{2a_{12}}{a_{11}^2 a_{22}}(X_1 - \mu_1)(X_2 - \mu_2) + \frac{(X_2 - \mu_2)^2}{a_{22}^2} \\ &= \frac{(X_1 - \mu_1)^2\sigma_2^2 + (X_2 - \mu_2)^2\sigma_1^2 - 2(X_1 - \mu_1)(X_2 - \mu_2)\sigma_{12}}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2} \end{aligned}$$

by inserting σ's for the a's.

By doing the same exercise with permutation order 2 we first get

$$\boldsymbol{\Sigma}^2 = \begin{pmatrix} \sigma_2^2 & \sigma_{12} \\ \sigma_{12} & \sigma_1^2 \end{pmatrix}$$

and $a_{11} = \sqrt{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}/\sigma_1$, $a_{12} = \sigma_{12}/\sigma_1$ and $a_{22} = \sigma_1$. Next, in the same manner as above, it is easily shown that

(IV.26)

$$W_1^2 = \frac{(X_2-\mu_2)^2\sigma_1^2 + (X_1-\mu_1)^2\sigma_2^2 - 2(X_1-\mu_1)(X_2-\mu_2)\sigma_{12}}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2} = W_1^1.$$

Hence we have shown that approach $\mathcal{A}_1^{(i)}$ is permutation invariant under the Gaussian copula null hypothesis. This is not so for other weight functions or other null hypothesis copulae. The invariance stems from the use of Φ^{-1} which cancels out with the Φ in $\mathbf{V} = \Phi(\mathbf{Y})$ and the squaring $\Phi(V_i)^2$.

IV.B Derivation of a Cramér-von Mises statistic

Consider the Cramér–von Mises (CvM) statistic

$$T = n\int_0^1 \{\widehat{F}(w) - F(w)\}^2 \mathrm{d}F(w),$$

where $\widehat{F}(w) = \frac{1}{n+1}\sum_{j=1}^n I(X_j \le t)$ is the empirical distribution function, equivalent to the normalized ranks used in the construction of the pseudo-vector $\mathbf{Z}$ in (IV.2). Given a random sample $(x_1, \ldots, x_n)$, the empirical version $\widehat{T}$ of the CvM statistic can be derived as follows.

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

Since $\widehat{F}(w)$ 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:

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

For approach $\mathcal{A}_1$ the test observator $S_1(w)$ is $U[0,1]$ under the null hypothesis. Hence $F(w) = w$ and we easily see that $\widehat{T}$ reduces to

$$\widehat{T}' = \frac{n}{3} + \frac{n}{n+1}\sum_{j=1}^n \widehat{F}\left(\frac{j}{n+1}\right)^2 - \frac{n}{(n+1)^2}\sum_{j=1}^n (2j+1)\widehat{F}\left(\frac{j}{n+1}\right).$$

IV.C Test procedures

Suppose we have observed the sample data $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$. The following parametric bootstrap procedures lead to proper p-value estimates for a parametric null hypothesis copula.

IV.C.1 Approach $\mathcal{A}_1$ (Berg and Bakken, 2007)

(1) Extract the pseudo-observations $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ by converting the sample data $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ into normalized ranks according to (IV.2).

(2) Estimate the parameters θ with a consistent estimator $\widehat{\theta} = \widehat{\mathcal{V}}(\mathbf{z}_1, \ldots, \mathbf{z}_n)$.

(3) Compute $(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \mathbf{R}(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ assuming the parametric null hypothesis copula $C_{\widehat{\theta}}$. Here $\mathbf{R}(\mathbf{z}_1, \ldots, \mathbf{z}_n) = (\mathcal{R}(z_{11}, \ldots, z_{1d}), \ldots, \mathcal{R}(z_{n1}, \ldots, z_{nd}))$ where $\mathcal{R}(z_{11}, \ldots, z_{1d}) = (\mathcal{R}_1(z_{11}), \ldots, \mathcal{R}_d(z_{1d}))$ denotes Rosenblatt's transformation as presented in Definition IV.1.

(4) Compute $(\mathbf{h}_1, \ldots, \mathbf{h}_n) = \mathbf{R}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$.

(5) Compute W_1 according to (IV.3), using weight functions Γ_V and Γ_H on $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ and $(\mathbf{h}_1, \ldots, \mathbf{h}_n)$ respectively.

(6) If W_1 follows a known distribution under the null hypothesis, compute $F_1(W_1)$ accordingly and jump to step (8).
If not, approximate F_1 as follows. For some large integer N_b, repeat the following steps for every $l \in \{1, \ldots, N_b\}$:

  (i) Generate a random sample $\mathbf{v}_l^* = (v_{1,l}^*, \ldots, v_{d,l}^*)$ from the null hypothesis copula, namely an i.i.d. $U[0,1]^d$ vector.

  (ii) Compute $\mathbf{h}_l^* = (h_{1,l}^*, \ldots, h_{d,l}^*) = \mathcal{R}(v_{1,l}^*, \ldots, v_{d,l}^*)$.

  (iii) Compute $W_{1,l}^*$ according to (IV.3) using the same weight functions Γ_V and Γ_H as in step (5) but now on $(v_{1,l}^*, \ldots, v_{d,l}^*)$ and $(h_{1,l}^*, \ldots, h_{d,l}^*)$ respectively.

(7) Compute $F_1(W_1) = \frac{1}{N_b+1} \sum_{l=1}^{N_b} I\{W_{1,l}^* > W_1\}$.

(8) Compute $\widehat{T}_1$ according to (IV.4) and (IV.5).

(9) For some large integer K, repeat the following steps for every $k \in \{1, \ldots, K\}$:

  (a) Generate a random sample $(\mathbf{x}_{1,k}^0, \ldots, \mathbf{x}_{n,k}^0)$ from the null hypothesis copula $C_{\widehat{\theta}}$ and compute the associated pseudo-samples $(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$ according to (IV.2).

  (b) Estimate the parameters θ^0 with $\widehat{\theta}_k^0 = \widehat{\mathcal{V}}(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$.

  (c) Compute $(\mathbf{v}_{1,k}^0, \ldots, \mathbf{v}_{n,k}^0) = \mathbf{R}(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$ assuming the parametric null hypothesis copula $C_{\widehat{\theta}_k^0}$.

  (d) Compute $(\mathbf{h}_{1,k}^0, \ldots, \mathbf{h}_{n,k}^0) = \mathbf{R}(\mathbf{v}_{1,k}^0, \ldots, \mathbf{v}_{n,k}^0)$.

  (d) Compute $W_{1,k}^0$ according to (IV.3), using the same weight functions Γ_V and Γ_H as in step (5), now on $(\mathbf{v}_{1,k}^0, \ldots, \mathbf{v}_{n,k}^0)$ and $(\mathbf{h}_{1,k}^0, \ldots, \mathbf{h}_{n,k}^0)$ respectively.

  (e) If $W_{1,k}^0$ follows a known distribution under the null hypothesis, compute $F_1(W_{1,k}^0)$ accordingly and jump to step (g).
  If not, approximate F_1 as follows. For some large integer N_b, repeat the following steps for every $l \in \{1, \ldots, N_b\}$:

    (i) Generate a random sample $\mathbf{v}_{l,k}^{0*} = (v_{1,l,k}^{0*}, \ldots, v_{d,l,k}^{0*})$ from the null copula, an i.i.d. $U[0,1]^d$ vector.

    (ii) Compute $\mathbf{h}_{l,k}^{0*} = (h_{1,l,k}^{0*}, \ldots, h_{d,l,k}^{0*}) = \mathcal{R}(v_{1,l,k}^{0*}, \ldots, v_{d,l,k}^{0*})$.

(iii) Compute $W_{1,l,k}^{0*}$ according to (IV.3) using the same weight functions Γ_V and Γ_H as in step (5) but now on $(v_{1,l,k}^{0*},\dots,v_{d,l,k}^{0*})$ and $(h_{1,l,k}^{0*},\dots,h_{d,l,k}^{0*})$ respectively.

(f) Compute $F_1(W_1^0)=\frac{1}{N_b+1}\sum_{l=1}^{N_b} I\{W_{1,l,k}^{0*}>W_{1,k}^0\}$.

(g) Compute $\widehat{T}_{1,k}^0$ according to (IV.4) and (IV.5).

(10) An approximate p-value for approach $\mathcal{A}_1$ is then given by $\widehat{p}=\frac{1}{K+1}\sum_{k=1}^{K} I\{\widehat{T}_{1,k}^0>\widehat{T}_1\}$.

IV.C.2 Approach $\mathcal{A}_2$ (Genest et al., 2008)

(1) Extract the pseudo-observations $(\mathbf{z}_1,\dots,\mathbf{z}_n)$ by converting the sample data $(\mathbf{x}_1,\dots,\mathbf{x}_n)$ into normalized ranks according to (IV.2).

(2) Estimate the parameters θ with a consistent estimator $\widehat{\theta}=\widehat{\mathcal{V}}(\mathbf{z}_1,\dots,\mathbf{z}_n)$.

(3) Compute $\widehat{C}(\mathbf{z})$ according to (IV.6).

(4) If there is an analytical expression for C_θ, compute the estimated statistic $\widehat{T}_2$ by plugging $\widehat{C}(\mathbf{z})$ and $C_{\widehat{\theta}}(\mathbf{z})$ into (IV.7). Jump to step (5).
If there is no analytical expression for C_θ then choose $N_b\geq n$ and carry out the following steps:

(i) Generate a random sample $(\mathbf{x}_1^*,\dots,\mathbf{x}_{N_b}^*)$ from the null hypothesis copula $C_{\widehat{\theta}}$ and compute the associated pseudo-samples $(\mathbf{z}_1^*,\dots,\mathbf{z}_{N_b}^*)$ according to (IV.2).

(ii) Approximate $C_{\widehat{\theta}}$ by $C_{\widehat{\theta}}^*(\mathbf{u})=\frac{1}{N_b+1}\sum_{l=1}^{N_b} I\{\mathbf{z}_l^*\leq\mathbf{u}\},\quad \mathbf{u}\in[0,1]^d$.

(iii) Approximate the CvM statistic in (IV.7) by $\widehat{T}_2=\sum_{j=1}^{n}\left\{\widehat{C}(\mathbf{z}_j)-C_{\widehat{\theta}}^*(\mathbf{z}_j)\right\}^2$.

(5) For some large integer K, repeat the following steps for every $k\in\{1,\dots,K\}$:

(a) Generate a random sample $(\mathbf{x}_{1,k}^0,\dots,\mathbf{x}_{n,k}^0)$ from the null hypothesis copula $C_{\widehat{\theta}}$ and compute the associated pseudo-samples $(\mathbf{z}_{1,k}^0,\dots,\mathbf{z}_{n,k}^0)$ according to (IV.2).

(b) Estimate the parameters θ^0 with a consistent estimator $\widehat{\theta}_k^0=\widehat{\mathcal{V}}(\mathbf{z}_{1,k}^0,\dots,\mathbf{z}_{n,k}^0)$.

(c) Let $\widehat{C}_k^0(\mathbf{u})=\frac{1}{n+1}\sum_{j=1}^{n} I\{\mathbf{z}_{j,k}^0\leq\mathbf{u}\},\quad \mathbf{u}\in[0,1]^d$.

(d) If there is an analytical expression for C_θ, let $\widehat{T}_{2,k}^0=\sum_{j=1}^{n}\left\{\widehat{C}_k^0(\mathbf{z}_{j,k}^0)-C_{\widehat{\theta}_k^0}(\mathbf{z}_{j,k}^0)\right\}^2$ and jump to step (6).
If there is no analytical expression for C_θ then choose $N_b\geq n$ and proceed as follows:

(i) Generate a random sample $(\mathbf{x}_{1,k}^{0*},\dots,\mathbf{x}_{N_b,k}^{0*})$ from the null hypothesis copula $C_{\widehat{\theta}_k^0}$ and compute the associated pseudo-samples $(\mathbf{z}_{1,k}^{0*},\dots,\mathbf{z}_{N_b,k}^{0*})$ according to (IV.2).

(ii) Approximate $C_{\widehat{\theta}_k^0}$ by $C_{\widehat{\theta}_k^0}^{0*}(\mathbf{u})=\frac{1}{N_b+1}\sum_{l=1}^{N_b} I\{\mathbf{z}_{l,k}^{0*}\leq\mathbf{u}\},\quad \mathbf{u}\in[0,1]^d$,

(iii) Approximate the CvM statistic in (IV.7) by $\widehat{T}_{2,k}^*=\sum_{j=1}^{n}\left\{\widehat{C}_k^0(\mathbf{z}_{j,k}^0)-C_{\widehat{\theta}_k^0}^{0*}(\mathbf{z}_{j,k}^0)\right\}^2$.

(6) An approximate p-value for approach $\mathcal{A}_2$ is then given by $\widehat{p}=\frac{1}{K+1}\sum_{k=1}^{K} I\{\widehat{T}_{2,k}^0>\widehat{T}_2\}$.

IV.C.3 Approach $\mathcal{A}_3$ (Genest et al., 2008)

(1) Extract the pseudo-observations $(\mathbf{z}_1,\dots,\mathbf{z}_n)$ by converting the sample data $(\mathbf{x}_1,\dots,\mathbf{x}_n)$ into normalized ranks according to (IV.2).

(2) Estimate the parameters θ with a consistent estimator $\widehat{\theta}=\widehat{\mathcal{V}}(\mathbf{z}_1,\dots,\mathbf{z}_n)$.

(3) Compute $(\mathbf{v}_1,\dots,\mathbf{v}_n)=\mathbf{R}(\mathbf{z}_1,\dots,\mathbf{z}_n)$ assuming the parametric null hypothesis copula $C_{\widehat{\theta}}$.

(3) Compute $\widehat{C}(\mathbf{v})$ according to (IV.6).

(4) Compute $\widehat{T}_3$ according to (IV.8).

(5) For some large integer K, repeat the following steps for every $k \in \{1, \ldots, K\}$:

(a) Generate a random sample $(\mathbf{x}_{1,k}, \ldots, \mathbf{x}_{n,k}^0)$ from the null hypothesis copula $C_{\widehat{\theta}}$ and compute the associated pseudo-samples $(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$ according to (IV.2).

(b) Estimate the parameters θ^0 with a consistent estimator $\widehat{\theta}_k^0 = \widehat{\mathcal{V}}(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$.

(c) Compute $(\mathbf{v}_{1,k}^0, \ldots, \mathbf{v}_{n,k}^0) = \mathbf{R}(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$.

(d) Let $\widehat{C}_k^0(\mathbf{u}) = \frac{1}{n+1} \sum_{j=1}^n I\{\mathbf{v}_{j,k}^0 \leq \mathbf{u}\}, \quad \mathbf{u} \in [0,1]^d$.

(e) Compute $\widehat{T}_{3,k}^0 = \sum_{j=1}^n \left\{ \widehat{C}_k^0(\mathbf{v}_{j,k}^0) - C_\perp(\mathbf{v}_{j,k}^0) \right\}^2$.

(6) An approximate p-value for approach $\mathcal{A}_3$ is then given by $\widehat{p} = \frac{1}{K+1} \sum_{k=1}^K I\{\widehat{T}_{3,k}^0 > \widehat{T}_3\}$.

IV.C.4 Approach $\mathcal{A}_4$ (Genest et al., 2008)

(1) Extract the pseudo-observations $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ by converting the sample data $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ into normalized ranks according to (IV.2).

(2) Estimate the parameters θ with a consistent estimator $\widehat{\theta} = \widehat{\mathcal{V}}(\mathbf{z}_1, \ldots, \mathbf{z}_n)$.

(3) Compute $\widehat{C}(\mathbf{z})$ according to (IV.6).

(4) If there is an analytical expression for $S_{4,\theta}$, compute the statistic $\widehat{T}_4$ according to (IV.9) and (IV.10). Jump to step (5).
If there is no analytical expression for $S_{4,\theta}$ then choose $N_b \geq n$ and proceed as follows:

(i) Generate a random sample $(\mathbf{x}_1^*, \ldots, \mathbf{x}_{N_b}^*)$ from the null hypothesis copula $C_{\widehat{\theta}}$ and compute the associated pseudo-samples $(\mathbf{z}_1^*, \ldots, \mathbf{z}_{N_b}^*)$ according to (IV.2).

(ii) Approximate $S_{4,\widehat{\theta}}$ by $\widehat{S}_4^*(w) = \frac{1}{N_b+1} \sum_{l=1}^{N_b} I\{\widehat{C}^*(\mathbf{z}_l^*) \leq w\}$, where $\widehat{C}^*(\mathbf{u}) = \frac{1}{N_b+1} \sum_{l=1}^{N_b} I\{\mathbf{z}_l^* \leq \mathbf{u}\}, \quad \mathbf{u} \in [0,1]^d$.

(iii) Approximate the CvM statistic in (IV.10) by $\widehat{T}_4 = \frac{n}{N_b} \sum_{l=1}^{N_b} \left\{ \widehat{S}_4\left(\widehat{C}^*(\mathbf{z}_l^*)\right) - \widehat{S}_4^*\left(\widehat{C}^*(\mathbf{z}_l^*)\right) \right\}$.

(5) For some large integer K, repeat the following steps for every $k \in \{1, \ldots, K\}$:

(a) Generate a random sample $(\mathbf{x}_{1,k}^0, \ldots, \mathbf{x}_{n,k}^0)$ from the null hypothesis copula $C_{\widehat{\theta}}$ and compute the associated pseudo-samples $(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$ according to (IV.2).

(b) Estimate the parameters θ^0 with a consistent estimator $\widehat{\theta}_k^0 = \widehat{\mathcal{V}}(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$.

(c) Let $\widehat{S}_{4,k}^0(w) = \frac{1}{n+1} \sum_{j=1}^n I\{\widehat{C}_k^0(\mathbf{z}_{j,k}^0) \leq w\}$, where $\widehat{C}_k^0(\mathbf{u}) = \frac{1}{n+1} \sum_{j=1}^n I\{\mathbf{z}_{j,k}^0 \leq \mathbf{u}\}$.

(d) If there is an analytical expression for $S_{4,\theta}$, compute the statistic $\widehat{T}_{4,k}^0$ by using $\widehat{S}_{4,k}^0$ and $S_{4,\widehat{\theta}_k^0}$ in (IV.10). Jump to step (6).
If there is no analytical expression for $S_{4,\theta}$ then choose $N_b \geq n$ and proceed as follows:

(i) Generate a random sample $(\mathbf{x}_{1,k}^{0*}, \ldots, \mathbf{x}_{N_b,k}^{0*})$ from the null hypothesis copula $C_{\widehat{\theta}_k^0}$ and compute the associated pseudo-samples $(\mathbf{z}_{1,k}^{0*}, \ldots, \mathbf{z}_{N_b,k}^{0*})$ according to (IV.2).

(ii) Approximate $S_{4,\widehat{\theta}_k^0}$ by $\widehat{S}_{4,k}^{0*}(w) = \frac{1}{N_b+1} \sum_{l=1}^{N_b} I\{\widehat{C}_k^{0*}(\mathbf{z}_{l,k}^{0*}) \leq w\}$, where $\widehat{C}_k^{0*}(\mathbf{u}) = \frac{1}{N_b+1} \sum_{l=1}^{N_b} I\{\mathbf{z}_{l,k}^{0*} \leq \mathbf{u}\}, \quad \mathbf{u} \in [0,1]^d$.

(iii) Approximate the CvM statistic in (IV.10) by $\widehat{T}_{4,k}^0 = \frac{n}{N_b} \sum_{l=1}^{N_b} \left\{ \widehat{S}_{4,k}^0\left(\widehat{C}_k^{0*}(\mathbf{z}_{l,k}^{0*})\right) - \widehat{S}_{4,k}^{0*}\left(\widehat{C}_k^{0*}(\mathbf{z}_{l,k}^{0*})\right) \right\}$.

(6) An approximate p-value for approach $\mathcal{A}_4$ is then given by $\widehat{p} = \frac{1}{K+1} \sum_{k=1}^K I\{\widehat{T}_{4,k}^0 > \widehat{T}_4\}$.

IV.C.5 Approach $\mathcal{A}_5$

(1) Extract the pseudo-observations $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ by converting the sample data $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ into normalized ranks according to (IV.2).

(2) Estimate the parameters θ with a consistent estimator $\widehat{\theta} = \widehat{\mathcal{V}}(\mathbf{z}_1, \ldots, \mathbf{z}_n)$.

(3) If there is an analytical expression for $S_{5,\theta}$, compute the statistic $\widehat{T}_5$ according to (IV.11) and (IV.12). Jump to step (4).
If there is no analytical expression for $S_{5,\theta}$ then choose $N_b \geq n$ and proceed as follows:

(i) Generate a random sample $(\mathbf{x}_1^*, \ldots, \mathbf{x}_{N_b}^*)$ from the null hypothesis copula $C_{\widehat{\theta}}$ and compute the associated pseudo-samples $(\mathbf{z}_1^*, \ldots, \mathbf{z}_{N_b}^*)$ according to (IV.2).

(ii) Approximate $S_{5,\widehat{\theta}}$ by $\widehat{S}_5^*(w) = \frac{1}{N_b+1} \sum_{l=1}^{N_b} I\{C_\perp(\mathbf{z}_l^*) \leq w\}$.

(iii) Approximate the CvM statistic in (IV.12) by
$\widehat{T}_5 = \frac{n}{N_b} \sum_{l=1}^{N_b} \left\{ \widehat{S}_5\left(C_\perp(\mathbf{z}_l^*)\right) - \widehat{S}_5^*\left(C_\perp(\mathbf{z}_l^*)\right) \right\}$.

(4) For some large integer K, repeat the following steps for every $k \in \{1, \ldots, K\}$:

(a) Generate a random sample $(\mathbf{x}_{1,k}^0, \ldots, \mathbf{x}_{n,k}^0)$ from the null hypothesis copula $C_{\widehat{\theta}}$ and compute the associated pseudo-samples $(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$ according to (IV.2).

(b) Estimate the parameters θ^0 with a consistent estimator $\widehat{\theta}_k^0 = \widehat{\mathcal{V}}(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$.

(c) Let $\widehat{S}_{5,k}^0(w) = \frac{1}{n+1} \sum_{j=1}^{n} I\{C_\perp(\mathbf{z}_{j,k}^0) \leq w\}$.

(d) If there is an analytical expression for $S_{5,\theta}$, compute the statistic $\widehat{T}_{5,k}^0$ by using $\widehat{S}_{5,k}^0$ and $S_{5,\widehat{\theta}_k^0}$ in (IV.12). Jump to step (5).
If there is no analytical expression for $S_{5,\theta}$ then choose $N_b \geq n$ and proceed as follows:

(i) Generate a random sample $(\mathbf{x}_{1,k}^{0*}, \ldots, \mathbf{x}_{N_b,k}^{0*})$ from the null hypothesis copula $C_{\widehat{\theta}_k^0}$ and compute the associated pseudo-samples $(\mathbf{z}_{1,k}^{0*}, \ldots, \mathbf{z}_{N_b,k}^{0*})$ according to (IV.2).

(ii) Approximate $S_{5,\widehat{\theta}_k^0}$ by $\widehat{S}_{5,k}^{0*}(w) = \frac{1}{N_b+1} \sum_{l=1}^{N_b} I\{C_\perp(\mathbf{z}_{l,k}^{0*}) \leq w\}$.

(iii) Approximate the CvM statistic in (IV.12) by
$\widehat{T}_{5,k}^0 = \frac{n}{N_b} \sum_{l=1}^{N_b} \left\{ \widehat{S}_{5,k}^0\left(C_\perp(\mathbf{z}_{l,k}^{0*})\right) - \widehat{S}_{5,k}^{0*}\left(C_\perp(\mathbf{z}_{l,k}^{0*})\right) \right\}$.

(5) An approximate p-value for approach $\mathcal{A}_5$ is then given by $\widehat{p} = \frac{1}{K+1} \sum_{k=1}^{K} I\{\widehat{T}_{5,k}^0 > \widehat{T}_5\}$.

IV.C.6 Approach $\mathcal{A}_6$

(1) Extract the pseudo-observations $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ by converting the sample data $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ into normalized ranks according to (IV.2).

(2) Estimate the parameters θ with a consistent estimator $\widehat{\theta} = \widehat{\mathcal{V}}(\mathbf{z}_1, \ldots, \mathbf{z}_n)$.

(3) Estimate the parameters $\widehat{\theta}_\tau$ and $\widehat{\theta}_W$ according to (IV.13).

(4) Compute $\widehat{T}_6$ according to (IV.14).

(5) For some large integer K, repeat the following steps for every $k \in \{1, \ldots, K\}$:

(a) Generate a random sample $(\mathbf{x}_{1,k}^0, \ldots, \mathbf{x}_{n,k}^0)$ from the null hypothesis copula $C_{\widehat{\theta}}$ and compute the associated pseudo-samples $(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$ according to (IV.2).

(b) Estimate the parameters $\widehat{\theta}_{\tau,k}^0$ and $\widehat{\theta}_{W,k}^0$ according to (IV.13).

(c) Compute $\widehat{T}_{6,k}^0$ according to (IV.14) using $\widehat{\theta}_{\tau,k}^0$ and $\widehat{\theta}_{W,k}^0$.

(6) An approximate p-value for approach $\mathcal{A}_6$ is then given by $\widehat{p} = \frac{1}{K+1} \sum_{k=1}^{K} I\{\widehat{T}_{6,k}^0 > \widehat{T}_6\}$.

IV.C.7 Approach $\mathcal{A}_7$ (Panchenko (2005) – corrected)

(1) Extract the pseudo-observations $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ by converting the sample data $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ into normalized ranks according to (IV.2).

(2) Estimate the parameters θ with a consistent estimator $\widehat{\theta} = \widehat{\mathcal{V}}(\mathbf{z}_1, \ldots, \mathbf{z}_n)$.

(3) Generate a random sample $(\mathbf{x}_1^*, \ldots, \mathbf{x}_n^*)$ from the null hypothesis copula $C_{\widehat{\theta}}$ and compute the associated pseudo-samples $(\mathbf{z}_1^*, \ldots, \mathbf{z}_n^*)$ according to (IV.2).

(4) Compute $\widehat{T}_7$ according to (IV.15) using $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ and $(\mathbf{z}_1^*, \ldots, \mathbf{z}_n^*)$.

(5) For some large integer K, repeat the following steps for each $k \in \{1, \ldots, K\}$:

   (a) Generate a random sample $(\mathbf{x}_{1,k}^0, \ldots, \mathbf{x}_{n,k}^0)$ from the null hypothesis copula $C_{\widehat{\theta}}$ and compute the associated pseudo-samples $(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$ according to (IV.2).

   (b) Estimate the parameters θ^0 with a consistent estimator $\widehat{\theta}_k^0 = \widehat{\mathcal{V}}(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$.

   (c) Generate a random sample $(\mathbf{x}_{1,k}^{0*}, \ldots, \mathbf{x}_{n,k}^{0*})$ from the null hypothesis copula $C_{\widehat{\theta}_k^0}$ and compute the associated pseudo-samples $(\mathbf{z}_{1,k}^{0*}, \ldots, \mathbf{z}_{n,k}^{0*})$ according to (IV.2).

   (d) Compute $\widehat{T}_{7,k}^0$ according to (IV.15) using $(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$ and $(\mathbf{z}_{1,k}^{0*}, \ldots, \mathbf{z}_{n,k}^{0*})$.

(6) An approximate p-value for approach $\mathcal{A}_7$ is then given by $\widehat{p} = \frac{1}{K+1} \sum_{k=1}^{K} I\{\widehat{T}_{7,k}^0 > \widehat{T}_7\}$.

IV.C.8 Approach $\mathcal{A}_8$

(1) Extract the pseudo-observations $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ by converting the sample data $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ into normalized ranks according to (IV.2).

(2) Estimate the parameters θ with a consistent estimator $\widehat{\theta} = \widehat{\mathcal{V}}(\mathbf{z}_1, \ldots, \mathbf{z}_n)$.

(3) Compute $(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \mathbf{R}(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ assuming the parametric null hypothesis copula $C_{\widehat{\theta}}$.

(4) Generate a random sample $(\mathbf{v}_1^*, \ldots, \mathbf{v}_n^*)$ from the independence copula.

(5) Compute $\widehat{T}_8$ according to (IV.16) using $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ and $(\mathbf{v}_1^*, \ldots, \mathbf{v}_n^*)$.

(6) For some large integer K, repeat the following steps for each $k \in \{1, \ldots, K\}$:

   (a) Generate a random sample $(\mathbf{x}_{1,k}^0, \ldots, \mathbf{x}_{n,k}^0)$ from the null hypothesis copula $C_{\widehat{\theta}}$ and compute the associated pseudo-samples $(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$ according to (IV.2).

   (b) Estimate the parameters θ^0 with a consistent estimator $\widehat{\theta}_k^0 = \widehat{\mathcal{V}}(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$.

   (c) Compute $(\mathbf{v}_{1,k}^0, \ldots, \mathbf{v}_{n,k}^0) = \mathbf{R}(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$ assuming the parametric null hypothesis copula $C_{\widehat{\theta}_k^0}$.

   (d) Generate a random sample $(\mathbf{v}_{1,k}^{0*}, \ldots, \mathbf{v}_{n,k}^{0*})$ from the independence copula.

   (e) Compute $\widehat{T}_{8,k}^0$ according to (IV.16) using $(\mathbf{v}_{1,k}^0, \ldots, \mathbf{v}_{n,k}^0)$ and $(\mathbf{v}_{1,k}^{0*}, \ldots, \mathbf{v}_{n,k}^{0*})$.

(7) An approximate p-value for approach $\mathcal{A}_8$ is then given by $\widehat{p} = \frac{1}{K+1} \sum_{k=1}^{K} I\{\widehat{T}_{8,k}^0 > \widehat{T}_8\}$.

IV.C.9 Approach $\mathcal{A}_9$

(1) Extract the pseudo-observations $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ by converting the sample data $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ into normalized ranks according to (IV.2).

(2) Estimate the parameters θ with a consistent estimator $\widehat{\theta} = \widehat{\mathcal{V}}(\mathbf{z}_1, \ldots, \mathbf{z}_n)$.

(3) Compute $\widehat{T}_1^{(i)}$, $\widehat{T}_1^{(ii)}$, $\widehat{T}_2$-$\widehat{T}_8$ by carrying out the appropriate steps of test procedures IV.C.1-IV.C.8 using $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ and $\widehat{\theta}$.

(4) Compute $\widehat{T}_9$ according to (IV.17).

(5) For some large integer K, repeat the following steps for every $k \in \{1, \ldots, K\}$:

  (a) Generate a random sample $(\mathbf{x}_{1,k}^0, \ldots, \mathbf{x}_{n,k}^0)$ from the null hypothesis copula $C_{\widehat{\theta}}$ and compute the associated pseudo-samples $(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$ according to (IV.2).

  (b) Estimate the parameters θ^0 with a consistent estimator $\widehat{\theta}_k^0 = \widehat{\mathcal{V}}(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$.

  (c) Compute $\widehat{T}_{1,k}^{0,(i)}$, $\widehat{T}_{1,k}^{0,(ii)}$, $\widehat{T}_{2,k}^0$-$\widehat{T}_{8,k}^0$ by carrying out the appropriate steps of test procedures IV.C.1-IV.C.8 using $(\mathbf{z}_{1,k}^0, \ldots, \mathbf{z}_{n,k}^0)$ and $\widehat{\theta}_k^0$.

  (d) Compute $\widehat{T}_{9,k}^0$ according to (IV.17) using $\widehat{T}_{1,k}^{0,(i)}$, $\widehat{T}_{1,k}^{0,(ii)}$, $\widehat{T}_{2,k}^0$-$\widehat{T}_{8,k}^0$.

(6) An approximate p-value for approach $\mathcal{A}_9$ is then given by $\widehat{p} = \frac{1}{K+1} \sum_{k=1}^{K} I\{\widehat{T}_{9,k}^0 > \widehat{T}_9\}$.

V

Local sensitivity analysis of goodness-of-fit tests for copulas

Daniel Berg
University of Oslo & Norwegian Computing Center

Jean-François Quessy
Université du Québec à Trois-Rivières

Abstract

The asymptotic behaviour of several goodness-of-fit statistics for copula families is obtained under contiguous alternatives. Many comparisons between a Cramér–von Mises functional of the empirical copula process and new moment-based goodness-of-fit statistics are made by considering their associated asymptotic local power curves. It is shown that the choice of the estimator for the unknown parameter can have a significant influence on the power of the Cramér-von Mises test, and that some of the moment-based statistics can provide simple and efficient goodness-of-fit methods. The paper ends with an extensive simulation study that aims to extend the conclusions to small and moderate sample sizes.

Key words

contiguous alternatives, copula, Cramér-von Mises statistic, empirical copula process, goodness-of-fit test, local power curves, rank-based estimators.

Publication details

Submitted for publication - October 2007.

V.1 Introduction

Copula functions contain all the information about the dependence structure of a random vector. Indeed, due to the representation theorem of Sklar (1959), every bivariate distribution function H can be written as $H(x,y) = C\{F(x), G(y)\}$, where F and G are the marginal distributions and $C : [0,1]^2 \to [0,1]$ is the copula. It turns out that C, which is unique when F and G are continuous, is a distribution function with uniform marginals on $[0,1]$. This representation enables practitioners to model the marginal behaviours and the dependence structure in separate steps. While the adjustment of univariate distributions is well documented, the study of goodness-of-fit tests for copulas emerged only recently as a challenging inferential problem.

Let C be the underlying copula of a bivariate population with continuous marginals and suppose one wants to test the goodness-of-fit hypotheses $\mathcal{H}_0 : C \in \mathcal{F} = \{C_\theta; \theta \in \mathcal{M}\}$ and $\mathcal{H}_1 : C \notin \mathcal{F} = \{C_\theta; \theta \in \mathcal{M}\}$, where $\mathcal{M}$ is the parameter space. Test statistics that help discriminate between $\mathcal{H}_0$ and $\mathcal{H}_1$ have been proposed by Fermanian (2005), Genest et al. (2006a), Scaillet (2006) and Chen and Fan (2005), among others. A bayesian selection procedure has also been investigated by Huard et al. (2006). In most cases, the efficiency of these methods, i.e. the power, is approximated by simulating repeatedly from a fixed alternative copula $D \notin \mathcal{F}$. This is done, in particular, in the works of Genest et al. (2008) and Berg (2007b), where extensive simulation results and recommendations are provided.

One of the most desirable properties of a statistical procedure is its ability to detect small departures from the null hypothesis. In the context of testing the fit to a particular copula family, such perturbations from $\mathcal{H}_0$ are given by the sequence of distributions

$$\mathcal{Q}_{\delta_n}(x,y) = \left(1 - \frac{\delta}{\sqrt{n}}\right) C(x,y) + \frac{\delta}{\sqrt{n}}\, D(x,y), \tag{V.1}$$

where $\delta_n = n^{-1/2}\delta$, $\delta > 0$ and C, D are bivariate copulas such that $C \in \mathcal{F}$. This mixture distribution is a copula for all $0 < \delta \leq n^{1/2}$. It is supposed throughout the paper that $\mathcal{Q}_{\delta_n}$ belongs to $\mathcal{F}$ only at the limit when $n \to \infty$. Moreover, in order to ensure that the departure from $\mathcal{H}_0$ increases as δ becomes larger (at least for large values of n), it is assumed throughout that the copula D stochastically dominates C, i.e. $D(x,y) \geq C(x,y)$ for all $(x,y) \in [0,1]^2$. The skill of a goodness-of-fit test to reject $\mathcal{H}_0$ under (V.1) can easily be motivated from applications in finance, where it is often advisable to detect changes in the dependence pattern over time, e.g. regime shifts for commodity markets.

In this paper, the asymptotic non-degenerate distribution of some goodness-of-fit statistics is investigated under the sequence $(\mathcal{Q}_{\delta_n})_{n\geq 1}$ of alternatives. The focus is put on a Cramér-von Mises type statistic computed from a version of the empirical copula process and on simple but efficient moment-based test statistics. The characterization of their limiting behaviour enables to compute asymptotic local power curves from which comparisons between the goodness-of-fit statistics under investigation can be made.

In Section V.2, the goodness-of-fit test statistics studied in this work are defined. In Section V.3, their asymptotic distribution under alternatives of the form (V.1) are obtained. These results enable to compute, in Section V.4, the local power curves of the statistics under study and hence to compare the latter under chosen scenarios of local distributions. In Section V.5, a new measure of asymptotic relative efficiency generalizing that of Pitman is described and computed for many cases. This index is particularly useful for the Cramér-von Mises goodness-of-fit statistic whose local power curve has no explicit expression. An extensive simulation study that aim to investigate the local behaviour of the testing procedures in small and moderate sample sizes and compare with the asymptotic results follows in Section V.6. The paper ends with a discussion of ideas for future investigations.

V.2 Some goodness-of-fit statistics for copula families

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from a bivariate population with continuous marginal distributions F, G and whose underlying copula is C. In Subsections V.2.1, V.2.2 and V.2.3, sta-

tistical procedures to determine if C belongs or not to a parametric family $\mathcal{F} = \{C_\theta; \theta \in \mathcal{M}\}$ are described. It is assumed throughout that $\mathcal{M}$ is a subset of the real line, so that θ can be estimated by an empirical version of a moment of C_θ. Since all statistics considered in this work are invariant under strictly increasing transformations of the variables, one can consider, for simplicity and without any loss of generality, that the marginal distributions are uniform on the interval $[0, 1]$.

V.2.1 The empirical copula goodness-of-fit process

A consistent estimation of a copula is possible via the empirical copula, which Deheuvels (1979) described as the distribution function of the sample of normalized ranks, i.e. $(\tilde{R}_{1,n}, \tilde{S}_{1,n})$, ..., $(\tilde{R}_{n,n}, \tilde{S}_{n,n})$, where $\tilde{R}_{i,n} = F_n(X_i)$ and $\tilde{S}_{i,n} = G_n(Y_i)$, with

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \leq x) \quad \text{and} \quad G_n(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(Y_i \leq y)$$

being the empirical marginal distributions. Explicitly, C is estimated by

$$C_n(x, y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(\tilde{R}_{i,n} \leq x, \tilde{S}_{i,n} \leq y\right). \tag{V.2}$$

The weak consistency of the empirical process $\mathcal{C}_{n,\theta} = \sqrt{n}(C_n - C_\theta)$ to a centered gaussian limit was obtained by Deheuvels (1979) under the hypothesis of independence, i.e. in the special case when $C_\theta(x, y) = xy$. This result was extended under general distributions by Gänssler and Stute (1987), Fermanian et al. (2004) and Tsukahara (2005). A suggestion made by Fermanian (2005) and exploited by Quessy (2005) and Genest et al. (2008) consists in basing a goodness-of-fit test on a modified version of $\mathcal{C}_{n,\theta}$, namely $\mathcal{C}_n = \sqrt{n}(C_n - C_{\hat{\theta}_n})$, where $\hat{\theta}_n$ consistently estimates θ. As shown by Quessy (2005), $\mathcal{C}_n$ is weakly consistent under $\mathcal{H}_0$, if the following assumptions are satisfied.

- $\mathcal{A}_1$. For all $\theta \in \mathcal{M}$, the first order partial derivatives of C_θ exist and are continuous;
- $\mathcal{A}_2$. $(\mathcal{C}_{n,\theta}, \Theta_n)$ converges jointly to a gaussian process $(\mathcal{C}_\theta, \Theta)$, where $\Theta_n = \sqrt{n}(\hat{\theta}_n - \theta)$. Moreover, for all $\theta \in \mathcal{M}$ and as $\varepsilon \downarrow 0$,

$$\sup_{\|\theta^\star - \theta\| < \varepsilon} \sup_{(x,y) \in [0,1]^2} \left| \dot{C}_{\theta^\star}(x, y) - \dot{C}_\theta(x, y) \right| \longrightarrow 0,$$

  where $\dot{C}_\theta = \partial C_\theta / \partial \theta$.

Under $\mathcal{A}_1$ and $\mathcal{A}_2$, the empirical goodness-of-fit process $\mathcal{C}_n$ converges weakly to a centered limit $\mathcal{C} = \mathcal{C}_\theta - \Theta \dot{C}_\theta$ having covariance function $\Gamma_{\mathcal{C}}(u, v, u', v') = \mathrm{Cov}\{\mathcal{C}(u, v), \mathcal{C}(u', v')\}$ whose expression is explicit but cumbersome. Thanks to this asymptotic result, it is then justified to base a goodness-of-fit test on some continuous functional computed from $\mathcal{C}_n$ in virtue of the continous mapping theorem (Billingsley, 1999). An omnibus statistic which has good power properties in general is the Cramer-von Mises distance function

$$\mathcal{V}_n = \int_0^1 \int_0^1 \{\mathcal{C}_n(x, y)\}^2 \, \mathrm{dxdy}. \tag{V.3}$$

Note that the use of other functional distances are possible, e.g. the Kolmogorov-Smirnov type statistics, but the latter have been found by Genest et al. (2006a) and by Genest et al. (2008) to be generally less powerful than the Cramér-von Mises statistic. Since statistic (V.3) has no explicit form in general, Genest and Rémillard (2008) propose to rely on the parametric bootstrap version

$$\mathcal{V}_{n,N} = \int_0^1 \int_0^1 \{\mathcal{C}_{n,N}(x, y)\}^2 \, \mathrm{dxdy},$$

where $\mathcal{C}_{n,N} = \sqrt{n}(C_n - C_N)$ and C_N is the empirical copula computed via Equation (V.2) from an artificial sample $(X^\star_{1,n}, Y^\star_{1,n}), \ldots, (X^\star_{N,n}, Y^\star_{N,n})$ from $C_{\hat{\theta}_n}$. These authors show that as $n, N \to \infty$, the process $\mathcal{C}_{n,N}$ converges to the same limit as $\mathcal{C}_n$ and consequently, $\mathcal{V}_{n,N}$ has the same asymptotic distribution as $\mathcal{V}_n$.

V.2.2 Moment-based goodness-of-fit statistics

Consider two real valued moments m_1 and m_2 of C_θ that are related to θ by one-to-one functions r_1, r_2 defined on $\mathcal{M}$ such that $m_1 = r_1(\theta)$ and $m_2 = r_2(\theta)$. Under the null hypothesis that the unknown copula of a population belongs to $\mathcal{F}$, one has $r_1^{-1}(m_1) = r_2^{-1}(m_2)$. If $\hat{m}_{1,n}$ and $\hat{m}_{2,n}$ are consistent for m_1 and m_2 respectively, then $\hat{\theta}_{1,n} = r_1^{-1}(\hat{m}_{1,n})$ and $\hat{\theta}_{2,n} = r_2^{-1}(\hat{m}_{2,n})$ provide consistent estimations of θ. In most cases of interest, $\sqrt{n}(\hat{\theta}_{j,n} - \theta)$ is asymptotically normal with mean zero and variance $\sigma_j^2(C_\theta)$ under $\mathcal{H}_0$. A simple, asymptotically normal goodness-of-fit statistic is then given by

$$\mathcal{S}_n = \sqrt{n}\left\{r_1^{-1}(\hat{m}_{1,n}) - r_2^{-1}(\hat{m}_{2,n})\right\}. \tag{V.4}$$

A goodness-of-fit test then consists in rejecting the null hypothesis whenever $|\mathcal{S}_n|/\sigma(C_\theta)$ exceeds $z_{\alpha/2}$, i.e. the $(1-\alpha/2)$-th percentile of a $\mathcal{N}(0,1)$ distribution, where $\sigma^2(C_\theta) = \lim_{n\to\infty} \mathrm{Var}(\mathcal{S}_n)$. Note that tests based on $\mathcal{S}_n$ may be inconsistent since it may happen that $r_1^{-1}(m_1) = r_2^{-1}(m_2)$ even if $\mathcal{H}_0$ is false.

The above method can be employed by considering two of the most popular measures of association, namely Spearman's rho and Kendall's tau. The latter are respectively defined, in terms of the underlying copula C_θ of the population, by

$$\rho_{C_\theta}(\theta) = 12\int_0^1\int_0^1 C_\theta(x,y)\mathrm{dxdy} - 3 \quad \text{and} \quad \tau_{C_\theta}(\theta) = 4\int_0^1\int_0^1 \mathrm{C}_\theta(\mathrm{x},\mathrm{y})\mathrm{dC}_\theta(\mathrm{x},\mathrm{y}) - 1. \tag{V.5}$$

Consistent estimators based on inversions of these rank statistics are $\hat{\theta}_{n,\rho} = \rho_{C_\theta}^{-1}(\rho_n)$ and $\hat{\theta}_{n,\tau} = \tau_{C_\theta}^{-1}(\tau_n)$, where

$$\rho_n = 1 - \frac{6n}{n^2-1}\sum_{i=1}^n\left(\tilde{R}_{i,n} - \tilde{S}_{i,n}\right)^2 \quad \text{and} \quad \tau_n = -1 + \frac{4}{n(n-1)}\sum_{i\neq j}\mathbf{1}(X_i \le X_j, Y_i \le Y_j)$$

are their sample versions. Another estimator arises from the so-called pseudo maximum-likelihood method, which is similar to the classical likelihood approach but where the normalized ranks are used instead of the observations. The resulting estimator $\hat{\theta}_{n,PL}$ has been studied by Genest et al. (1995), Shih and Louis (1995) and recently by Kim et al. (2006). Based on these three consistent estimators, one can build three goodness-of-fit statistics of the form (V.4), namely

$$S_{n1} = \sqrt{n}\left(\hat{\theta}_{n,\rho} - \hat{\theta}_{n,\tau}\right),\ S_{n2} = \sqrt{n}\left(\hat{\theta}_{n,\rho} - \hat{\theta}_{n,PL}\right) \text{ and } S_{n3} = \sqrt{n}\left(\hat{\theta}_{n,\tau} - \hat{\theta}_{n,PL}\right). \tag{V.6}$$

V.2.3 Shih's goodness-of-fit test for the gamma frailty model

The dependence function associated to the bivariate gamma frailty model, also referred to as Clayton's copula, is given in Equation (V.B.1) to be found in Appendix V.A. Shih (1998) considered unweighed and weighted estimations of the dependence parameter θ via Kendall's tau τ_n and the weighted rank-based statistic

$$\hat{\theta}_{n,W} = \frac{\sum_{i<j}\Delta_{ij}/W_{ij}}{\sum_{i<j}(1-\Delta_{ij})/W_{ij}},$$

where $\Delta_{ij} = \mathbf{1}\left\{(X_i - X_j)(Y_i - Y_j) > 0\right\}$ and

$$W_{ij} = \sum_{k=1}^n \mathbf{1}\{X_k \le \max(X_i, X_j), Y_k \le \max(Y_i, Y_j)\}.$$

Since $\hat{\theta}_{n,\tau} = 2\tau_n/(1-\tau_n)$ and $\hat{\theta}_{n,W}$ are both unbiased for θ under the null hypothesis that C belongs to Clayton's family of copulas, a version of a goodness-of-fit statistic proposed by Shih (1998) is $\mathcal{S}_{n4} = \sqrt{n}(\hat{\theta}_{n,\tau} - \hat{\theta}_{n,W})$. One deduces from arguments to be found in Shih (1998) that $\mathcal{S}_{n4}$ is asymptotically normal under the null hypothesis. Unfortunately, the variance provided by

Shih (1998) was found to be wrong by Genest et al. (2006c), where a corrected formula is provided. From the work of these authors, one may deduce the asymptotic representation

$$\mathcal{S}_{n4} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{K_\theta(X_i, Y_i) - L_\theta(X_i, Y_i)\} + o_{\mathrm{P}}(1), \tag{V.7}$$

where

$$K_\theta(x,y) = 2(\theta+2)^2 \left\{ 2\left(x^{-\theta} + y^{-\theta} - 1\right)^{-1/\theta} - x - y + \frac{1}{\theta+2} \right\}$$

and

$$L_\theta(x,y) = (\theta+1)(2\theta+1)\log\left(x^{-\theta} + y^{-\theta} - 1\right)^{-1/\theta} - (\theta+1)^2 \log(xy) + \theta.$$

Genest et al. (2006c) then used (V.7) to compute the asymptotic variance of $\mathcal{S}_{n4}$, whose complicated expression is given by

$$\begin{aligned} \sigma_4^2(C_\theta) &= \frac{136\theta^7 + 1352\theta^6 + 5171\theta^5 + 9449\theta^4 + 8281\theta^3 + 3001\theta^2 + 240\theta + 18}{3\theta^2(\theta+1)^2(\theta+3)^2} \\ &\quad + \frac{8(\theta+2)^4}{\theta^2(\theta+1)^2}\,\mathcal{I}_1(\theta) - \frac{4(\theta+1)^4}{\theta^4} \sum_{k=0}^{\infty} \frac{(-1)^k}{(k+1+1/\theta)^2} - \frac{8(\theta+1)(\theta+2)}{\theta^3}\,\mathcal{I}_2(\theta), \end{aligned}$$

where

$$\mathcal{I}_1(\theta) = \sum_{k=0}^{\infty} \frac{\Gamma^2(1/\theta)}{\Gamma(1/\theta)} \frac{k!\Gamma(k+1/\theta)}{\Gamma(k+1+2/\theta)} \quad \text{and} \quad \mathcal{I}_2(\theta) = \sum_{k=0}^{\infty} \frac{\Gamma(2/\theta)k!}{(k+1/\theta)\,\Gamma(k+1+2/\theta)}.$$

V.3 Asymptotic behaviour under local sequences

In order to derive non-degenerate limiting distributions for a given goodness-of-fit statistic under the sequence $(\mathcal{Q}_{\delta_n})_{n\geq 1}$ defined in Equation (V.1), one has to ensure that $\mathcal{Q}_{\delta_n}$ is *close* to $\mathcal{Q}_0 = C_\theta$ in a certain sense. One such criteria is given by van der Vaart and Wellner (1996), where it is supposed that

$$\lim_{n\to\infty} \int_0^1 \int_0^1 \left\{ \sqrt{n}\left(\sqrt{q_{\delta_n}(x,y)} - \sqrt{q_0(x,y)}\right) - \frac{\delta \dot{q}_0(x,y)}{2\sqrt{q_0(x,y)}} \right\}^2 \mathrm{dxdy} = 0, \tag{V.8}$$

for q_δ being the density associated to $\mathcal{Q}_\delta$ and $\dot{q}_\delta = \partial q_\delta / \partial\delta$. Note that condition (V.8) entails that the sequence $(\mathcal{Q}_{\delta_n})_{n\geq 1}$ is contiguous with respect to $\mathcal{Q}_0$. This is the key requirement that enables to derive the asymptotic local representation of the goodness-of-fit statistics $\mathcal{V}_{n,N}$ and $\mathcal{S}_{n1}, \ldots, \mathcal{S}_{n4}$. This is the subject of the remains of this section.

V.3.1 Local behaviour of some estimators of the dependence parameter

Many interesting estimators for the unknown parameter of a copula family admit the asymptotic representation

$$\Theta_{n,\Lambda} = \sqrt{n}\left(\hat{\theta}_{n,\Lambda} - \theta\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Lambda_{C_\theta}\left(\tilde{R}_{i,n}, \tilde{S}_{i,n}\right) + o_{\mathrm{P}}(1), \tag{V.9}$$

where $\Lambda_{C_\theta} : [0,1]^2 \to [0,1]$ is a twice differentiable score function such that for all $\theta \in \mathcal{M}$ and all $(x,y) \in [0,1]^2, \mathrm{E}_{C_\theta}\{\Lambda_{C_\theta}(\mathrm{X},\mathrm{Y})\} = 0$ and $\left|\Lambda''_{C_\theta}(x,y)\right| \leq g_\theta(x,y)$, where g_θ and $\Lambda^2_{C_\theta}$ are integrable with respect to $c_\theta(x,y) = \partial^2 C_\theta(x,y)/\partial x \partial y$. These conditions ensure that $\Theta_{n,\Lambda}$ converges in law to

$$\Theta_\Lambda = \Theta'_\Lambda + \int_{(0,1)^2} \Lambda_{C_\theta,10}(x,y)\beta_1(x)c_\theta(x,y)\mathrm{dxdy} + \int_{(0,1)^2} \Lambda_{C_\theta,01}(\mathrm{x},\mathrm{y})\beta_2(\mathrm{y})\mathrm{c}_\theta(\mathrm{x},\mathrm{y})\mathrm{dxdy},$$

where Θ'_Λ is the limit of $n^{-1/2}\sum_{i=1}^n \Lambda_{C_\theta}(X_i, Y_i)$ and β_1, β_2 are uniform brownian bridges, i.e. gaussian processes with covariance function $\mathrm{Cov}\{\beta_j(s), \beta_j(t)\} = \min(s,t) - st$, $j = 1, 2$, arising as the limits of $\sqrt{n}\{F_n(x) - x\}$ and $\sqrt{n}\{G_n(y) - y\}$ respectively. Here, $\Lambda_{C_\theta,10}(x,y) = \partial\Lambda_{C_\theta}(x,y)/\partial x$ and $\Lambda_{C_\theta,01}(x,y) = \partial\Lambda_{C_\theta}(x,y)/\partial y$.

Among the estimators that admit representation (V.9), one has the inversion of Spearman's rho and the pseudo maximum-likelihood estimator explored by Genest et al. (1995) and Shih and Louis (1995). More details will be given in Example 1 and Example 2. Another popular estimation strategy using a statistic that is not of the form (V.9) is based on $\hat{\theta}_{n,\tau}$, i.e. on the inversion of Kendall's measure of association.

The next proposition, whose proof is deferred to Appendix V.A.1, identifies the asymptotic distribution of $\Theta_{n,\Lambda}$ and $\Theta_{n,\tau} = \sqrt{n}(\hat{\theta}_{n,\tau} - \theta)$ under contiguous alternatives of the type (V.1). This result is a prerequisite in order to compute the local power of moment-based goodness-of-fit statistics described in Section V.2.2. Also, it will enable to characterize the asymptotic behaviour of the process $\mathcal{C}_n$, and consequently that of $\mathcal{V}_{n,N}$, under $(\mathcal{Q}_{\delta_n})_{n\geq 1}$ for several strategies that aim to estimate θ.

Proposition 1
Assume that condition (V.8) *holds for the sequence* $(\mathcal{Q}_{\delta_n})_{n\geq 1}$. *Then under* $(\mathcal{Q}_{\delta_n})_{n\geq 1}$,

(i) $\Theta_{n,\Lambda} \rightsquigarrow \Theta_\Lambda + \delta\mu_\Lambda(C_\theta, D)$, *where* $\mu_\Lambda(C_\theta, D) = \mathrm{E_D}\left\{\Lambda_{\mathrm{C}_\theta}(\mathrm{X},\mathrm{Y})\right\} - \mathrm{E_{C_\theta}}\left\{\Lambda_{\mathrm{C}_\theta}(\mathrm{X},\mathrm{Y})\right\}$ *and* Θ_Λ *is a normal random variable with mean* 0 *and variance*

$$\sigma^2_\Lambda = \mathrm{Var}\left\{\Lambda_{C_\theta}(X,Y) + \int_0^1\int_X^1 \Lambda_{C_\theta,10}(x,y)c_\theta(x,y) + \int_Y^1\int_0^1 \Lambda_{C_\theta,01}(x,y)c_\theta(x,y)\right\};$$

(ii) $\Theta_{n,\tau} \rightsquigarrow \Theta_\tau + \delta\mu_\tau(C_\theta, D)$, *where* $\mu_\tau(C_\theta, D) = 4\{\tau'(\theta)\}^{-1}\left\{\mathrm{E_D}(\mathrm{C}_\theta) - \mathrm{E_{C_\theta}}(\mathrm{C}_\theta)\right\}$ *and* Θ_τ *is a normal random variable with mean* 0 *and variance*

$$\sigma^2_\tau = \frac{16}{\{\tau'(C_\theta)\}^2}\,\mathrm{Var}\left\{2C_\theta(X,Y) - X - Y\right\}.$$

The next two examples are applications of part (i) of Proposition 1 when the estimator is based on an inversion of Spearman's rho and on the pseudo maximum-likelihood estimator.

Example 1
Let $\rho_{C_\theta}(\theta)$ *be the population value of Spearman's measure of association for a vector* (X, Y) *with underlying copula* C_θ. *Then* $\hat{\theta}_{n,\rho} = \rho^{-1}_{C_\theta}(\rho_n)$ *is a consistent estimator for* θ, *where* ρ_n *is Spearman's rank correlation coefficient. Using a Taylor expansion of order* 1, *one can show that this estimator can be written in the form* (V.9) *with* $\Lambda_{C_\theta}(x,y) = \{\rho'_{C_\theta}\}^{-1}\{12xy - 3 - \rho_{C_\theta}(\theta)\}$, *where* $\rho'_{C_\theta}(\theta) = \partial\rho_{C_\theta}(\theta)/\partial\theta$. *Thus, under the contiguous sequence* (V.1), $\Theta_{n,\rho} = \sqrt{n}(\hat{\theta}_{n,\rho} - \theta)$ *is asymptotically normal with drift parameter* $\mu_\rho(C_\theta, D) = \{\rho'_{C_\theta}(\theta)\}^{-1}\{\rho_D - \rho_{C_\theta}(\theta)\}$ *and variance*

$$\sigma^2_\rho(C_\theta) = \frac{144}{\{\rho'_{C_\theta}(\theta)\}^2}\ \mathrm{Var}\left\{XY + \int_0^1\int_X^1 yc_\theta(x,y)\mathrm{dxdy} + \int_Y^1\int_0^1 \mathrm{xc_\theta(x,y)dxdy}\right\}.$$

Example 2
Let $\hat{\theta}_{n,PL}$ *be the pseudo likelihood estimator. From the work of Genest et al. (1995), one has representation* (V.9) *with* $\Lambda_{C_\theta}(x,y) = \beta^{-1}_{C_\theta}\ell'_{C_\theta}(x,y)$, *where* $\ell_{C_\theta}(x,y) = \log c_\theta(x,y)$ *and* $\beta_{C_\theta} = \mathrm{E_{C_\theta}}[\{\ell'_{C_\theta}(\mathrm{X},\mathrm{Y})\}^2]$, *with* $\ell'_{C_\theta} = \partial\ell_{C_\theta}/\partial\theta$. *An application of Proposition 1 shows that* $\Theta_{n,PL} = \sqrt{n}(\hat{\theta}_{n,PL} - \theta)$ *converges in law to a normal distribution with variance* $\sigma^2_{PL}(C_\theta) = \beta^{-2}_{C_\theta}\,\mathrm{Var}\{\ell'_{C_\theta}(X,Y) - W_{C_\theta,1}(X) - W_{C_\theta,2}(Y)\}$, *where*

$$W_{C_\theta,1}(u) = \int_u^1\int_0^1 \ell'_{C_\theta}(x,y)\ell'_{C_\theta,1}(x,y)c_\theta(x,y)\mathrm{dxdy}$$

(V.6)

and

$$W_{C_\theta,2}(u) = \int_0^1 \int_u^1 \ell'_{C_\theta}(x,y)\ell'_{C_\theta,2}(x,y)c_\theta(x,y)\mathrm{dxdy},$$

with $\ell'_{C_\theta,1}(x,y) = \partial\ell_{C_\theta}(x,y)/\partial x$ *and* $\ell'_{C_\theta,2}(x,y) = \partial\ell_{C_\theta}(x,y)/\partial y$. *The asymptotic mean is*

$$\mu_{PL}(C_\theta, D) = \beta_{C_\theta}^{-1}\mathrm{E_D}\left\{\ell'_{C_\theta}(\mathrm{X},\mathrm{Y})\right\} - \beta_{C_\theta}^{-1}\mathrm{E}_{C_\theta}\left\{\ell'_{C_\theta}(\mathrm{X},\mathrm{Y})\right\} = \beta_{C_\theta}^{-1}\mathrm{E_D}\left\{\ell'_{C_\theta}(\mathrm{X},\mathrm{Y})\right\},$$

since by Lebesgue's dominated convergence theorem,

$$\mathrm{E}_{C_\theta}\left\{\ell'_{C_\theta}(\mathrm{X},\mathrm{Y})\right\} = \int_0^1\int_0^1 \dot{\mathrm{c}}_\theta(\mathrm{x},\mathrm{y})\mathrm{dxdy} = \frac{\partial}{\partial\theta}\int_0^1\int_0^1 \mathrm{c}_\theta(\mathrm{x},\mathrm{y})\mathrm{dxdy} = 0.$$

V.3.2 Local behaviour of the goodness-of-fit statistics

The first theoretical result of this section establishes the large-sample behaviour of $\mathcal{C}_n$ under the sequence $(\mathcal{Q}_{\delta_n})_{n\geq 1}$. It is assumed that the estimator of θ is either of the form (V.9) or based on the inversion of Kendall's tau.

Proposition 2
Suppose condition (V.8) *and Assumptions* $\mathcal{A}_1$*-*$\mathcal{A}_2$ *hold, and assume that* $\Theta_n = \sqrt{n}(\hat{\theta}_n - \theta)$ *converges in law to* $\tilde{\Theta} = \Theta + \delta\mu(C_\theta, D)$ *under the sequence* (V.1)*, where* Θ *is the limit in law of* Θ_n *under* $\mathcal{H}_0$. *Then under* $(\mathcal{Q}_{\delta_n})_{n\geq 1}$*, the empirical process* $\mathcal{C}_n = \sqrt{n}(C_n - C_{\hat{\theta}_n})$ *converges weakly to*

$$\tilde{\mathcal{C}} = \mathcal{C} + \delta\left\{D - C_\theta - \mu(C_\theta, D)\dot{C}_\theta\right\},$$

where $\mathcal{C}$ *is the weak limit of* $\mathcal{C}_n$ *under* $\mathcal{H}_0$ *and* $\dot{C}_\theta = \partial C_\theta/\partial\theta$.

Remark 1
As one may expect, a sequence of the form $\mathcal{Q}_{\delta_n} = C_{\theta+\delta_n}$ *yields absolutely no power for statistics based on* $\mathcal{C}_n$ *since* $\mathcal{Q}_{\delta_n} \in \mathcal{F}$ *in that case. Indeed, as one can deduce from computations made in the proof of Proposition 2, condition* (V.8) *enounced in van der Vaart and Wellner (1996) implies that* $\mathcal{C}_{n,\theta}$ *converges to* $\mathcal{C}_\theta + \delta\dot{C}_\theta$. *Moreover, since* Θ_n *converges to* $\Theta + \delta$ *in that case,* $\sqrt{n}(C_{\hat{\theta}_n} - C_\theta)$ *converges to* $(\Theta + \delta)\dot{C}_\theta$*, so that* $\mathcal{C}_n = \mathcal{C}_{n,\theta} - \sqrt{n}(C_{\hat{\theta}_n} - C_\theta)$ *converges to* $\mathcal{C}$*, i.e. to the same limit as under* $\mathcal{H}_0$.

The asymptotic local behaviour of the moment-based goodness-of-fit statistics (V.6) can easily be obtained as consequences of Proposition 1. This is the subject of Proposition 3, whose straightforward proof is omitted.

Proposition 3
Suppose condition (V.8) *holds. Then under* $(\mathcal{Q}_{\delta_n})_{n\geq 1}$,

(i) $\mathcal{S}_{n1} \rightsquigarrow \mathcal{S}_1 + \delta\{\mu_\rho(C_\theta, D) - \mu_\tau(C_\theta, D)\}$;

(ii) $\mathcal{S}_{n2} \rightsquigarrow \mathcal{S}_2 + \delta\{\mu_\rho(C_\theta, D) - \mu_{PL}(C_\theta, D)\}$;

(iii) $\mathcal{S}_{n3} \rightsquigarrow \mathcal{S}_3 + \delta\{\mu_\tau(C_\theta, D) - \mu_{PL}(C_\theta, D)\}$.

This result implies that the limiting distribution of S_{nj}, $j = 1, 2, 3$ under the contiguous sequence is normal with some mean $\delta\mu_j(C_\theta, D)$ and variance $\sigma_j^2(C_\theta)$. As long as $\mu(C_\theta, D) \neq 0$, a goodness-of-fit procedure based on S_{nj} will yield power locally.

V.3.3 Shih's statistic under contiguity

The asymptotic behaviour of $\mathcal{S}_{n4}$ under the contiguous sequence $(\mathcal{Q}_{\delta_n})_{n\geq 1}$ will follow from an application of Lecam's third lemma and the asymptotic representation (V.7). This result is summarized in Proposition 4.

Proposition 4
Under the contiguous sequence $(\mathcal{Q}_{\delta_n})_{n\geq 1}$, the goodness-of-fit statistic $\mathcal{S}_{n4}$ converges in law to a normal distribution with variance $\sigma_4^2(C_\theta)$ and mean $\delta\eta_1(C_\theta, D) - \delta\eta_2(C_\theta, D)$, where

$$\begin{aligned}
\eta_1(C_\theta, D) &= 4(\theta+2)^2 \left\{\mathrm{E_D}(\mathrm{C}_\theta) - \mathrm{E}_{\mathrm{C}_\theta}(\mathrm{C}_\theta)\right\} \\
\eta_2(C_\theta, D) &= (\theta+1)(\theta+2)\int_0^1\int_0^1 \{d(u,v) - c_\theta(u,v)\}\log C_\theta(u,v)\mathrm{d}u\mathrm{d}v \\
&\quad - (\theta+1)^2\int_0^1\int_0^1 \{d(u,v) - c_\theta(u,v)\}\log uv\mathrm{d}u\mathrm{d}v.
\end{aligned}$$

V.4 Local power comparisons

In this section, the asymptotic power of the goodness-of-fit tests based on $\mathcal{V}_{n,N}$ and $\mathcal{S}_{n1}, \ldots, \mathcal{S}_{n4}$ are investigated under alternatives of the form (V.1). Here, C and D are chosen to be in the same family with different levels of dependence. In other words, local alternatives of the form $\mathcal{Q}_{\delta_n} = (1-\delta_n)C_\theta(x,y) + \delta_n C_{\theta'}(x,y)$ are considered, where $\theta < \theta'$. It is assumed that θ is a dependence parameter for the family $\{C_\theta; \theta \in \mathcal{M}\}$, i.e. $C_\theta(x,y) \leq C_{\theta'}(x,y)$ for all $(x,y) \in [0,1]^2$. This requirement is fulfilled for most families of copulas. The above mixing distribution can represent a setting where the data generating process stays in the same family over time but the dependence strength suddenly changes, c.f. regime-shifting models. Structural changes of this kind can occur in mean-reverting processes such as those driving oil and other commodity prices, where the dependence pattern, i.e. the copula family, remains the same over time but the strength of this link becomes significantly stronger or weaker at some moment.

The following analyses will consider local distributions involving mixtures of Clayton, Frank, Gumbel-Barnett and Normal copulas whose analytical expressions are given in Equations (V.B.1)–(V.B.4) to be found in Appendix V.B.

V.4.1 Efficiency of the empirical copula process under various estimation strategies

Here, the influence of the estimation strategy on the power of the Cramér-von Mises statistic is investigated under local sequences. Here and in the sequel, $\mathcal{C}_{n,N,\rho}$, $\mathcal{C}_{n,N,\tau}$ and $\mathcal{C}_{n,N,PL}$ refer to the empirical copula goodness-of-fit process with the estimation of θ based respectively on Spearman's rho, Kendall's tau and the pseudo likelihood approach. Similarly, $\mathcal{V}_{n,N}^{\rho}$, $\mathcal{V}_{n,N}^{\tau}$ and $\mathcal{V}_{n,N}^{PL}$ are the associated Cramér-von Mises functionals.

According to Proposition 2, the weak limits of the empirical copula goodness-of-fit processes $\mathcal{C}_{n,N,\rho}$, $\mathcal{C}_{n,N,\tau}$ and $\mathcal{C}_{n,N,PL}$ under the contiguous sequence $(\mathcal{Q}_{\delta_n})_{n\geq 1}$ are

$$\tilde{\mathcal{C}}_\rho = \mathcal{C}_\rho + \delta(g - \mu_\rho \dot{C}_\theta), \quad \tilde{\mathcal{C}}_\tau = \mathcal{C}_\tau + \delta(g - \mu_\tau \dot{C}_\theta) \quad \text{and} \quad \tilde{\mathcal{C}}_{PL} = \mathcal{C}_{PL} + \delta(g - \mu_{PL}\dot{C}_\theta),$$

where $\mathcal{C}_\rho$, $\mathcal{C}_\tau$ and $\mathcal{C}_{PL}$ are the respective limits under $\mathcal{H}_0$ and $g(x,y) = D(x,y) - C_\theta(x,y)$. Computations of μ_ρ, μ_τ and μ_{PL} are detailed in Appendix V.B for mixtures of Clayton, Frank, Gumbel-Barnett and Normal copulas. The results are reported in Table V.1. Generally speaking, these drift terms are higher for $\Theta_{n,\rho}$ and $\Theta_{n,PL}$ than for $\Theta_{n,\tau}$. This indicates that the estimator based on Kendall's tau is more robust under perturbations of $\mathcal{H}_0$ of the type $\mathcal{Q}_{\delta_n}$, which is not necessarily a good property for goodness-of-fit testing where one wants to detect departures from $\mathcal{H}_0$.

There is no hope to obtain explicit representations for the asymptotic distributions of $\mathcal{V}_{n,N}^{\rho}$, $\mathcal{V}_{n,N}^{\tau}$ and $\mathcal{V}_{n,N}^{PL}$, and consequently for the associated power curves. A procedure to overcome this

Table V.1: Drift terms for the estimators based on Spearman's rho, the pseudo maximum-likelihood and Kendall's tau under mixtures of Clayton, Frank, Gumbel-Barnett and Normal copulas.

τ_{C_θ}	τ_D	*Model*	μ_ρ	μ_{PL}	μ_τ	*Model*	μ_ρ	μ_{PL}	μ_τ
0.1	0.2		0.244	0.250	0.030		0.901	0.926	0.111
0.1	0.3		0.475	0.487	0.059		1.789	1.815	0.231
0.1	0.4		0.692	0.697	0.086		2.615	2.704	0.333
0.1	0.5	Clayton	0.889	0.882	0.114	Frank	3.385	3.519	0.435
0.4	0.5		0.527	0.544	0.067		1.319	1.381	0.164
0.4	0.6		0.996	0.995	0.128		2.436	2.619	0.315
0.4	0.7		1.384	1.393	0.183		3.351	3.810	0.452
0.4	0.8		1.679	1.786	0.228		4.021	4.762	0.548
0.1	0.2		0.099	0.101	0.013		0.154	0.154	0.019
0.1	0.3		0.192	0.198	0.025		0.301	0.302	0.037
0.1	0.4		0.281	0.290	0.037		0.440	0.443	0.054
0.1	0.5	Gumbel–	0.485	0.379	0.049	Normal	0.565	0.572	0.071
0.4	0.5	Barnett	0.096	0.101	0.016		0.120	0.123	0.017
0.4	0.6		0.179	0.195	0.029		0.226	0.228	0.032
0.4	0.7		0.250	0.282	0.044		0.312	0.315	0.046
0.4	0.8		0.303	0.797	0.068		0.377	0.387	0.062

difficulty is explained next in order to compute the local power curve of the Cramér-von Mises tests. For simplicity, only the case involving $\mathcal{V}_{n,N}^{\rho}$ is detailed.

First note that under $(\mathcal{Q}_{\delta_n})_{n\geq 1}$,

$$\mathcal{V}_{n,N}^{\rho} \rightsquigarrow \tilde{\mathcal{V}}^{\rho} = \int_0^1\int_0^1 \left\{\tilde{\mathcal{C}}_\rho(x,y)\right\}^2 \mathrm{d}x\mathrm{d}y = \int_0^1\int_0^1 \{\mathcal{C}_\rho(x,y) + \delta h_\rho(x,y)\}^2 \mathrm{d}x\mathrm{d}y,$$

where $h_\rho(x,y) = D(x,y) - C_\theta(x,y) - \mu_\rho(C_\theta, D)\dot{C}_\theta(x,y)$. Hence, for large values of n and N, an approximation is given by

$$\tilde{\mathcal{V}}_{n,N}^{\rho} = \int_0^1\int_0^1 \{\mathcal{C}_{n,N,\rho}(x,y) + \delta h_\rho(x,y)\}^2 \mathrm{d}x\mathrm{d}y,$$

where $\mathcal{C}_{n,N,\rho}$ is the empirical copula goodness-of-fit process where θ is estimated through an inversion of Spearman's rho. One can see that $\tilde{\mathcal{V}}_{n,N}^{\rho} = \mathcal{V}_{n,N}^{\rho} + 2\delta V_1 + \delta^2 V_2$, where

$$\begin{aligned} V_1 &= \int_0^1\int_0^1 h_\rho(x,y)\mathcal{C}_{n,N,\rho}(x,y)\mathrm{d}x\mathrm{d}y \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{\tilde{R}_{i,n}}\int_{\tilde{S}_{i,n}} h_\rho(x,y)\mathrm{d}x\mathrm{d}y - \sqrt{n}\int_0^1\int_0^1 h_\rho(x,y)C_{\hat{\theta}_{n,\rho}}(x,y)\mathrm{d}x\mathrm{d}y \end{aligned}$$

and

$$V_2 = \int_0^1\int_0^1 \{h_\rho(x,y)\}^2 \mathrm{d}x\mathrm{d}y.$$

In Figure V.1 and Figure V.2, the local power curves of the Cramér-von Mises test statistic computed under the three considered estimation strategies using the above approximations with $n = 2500$ and $N = 2500$ are reported under mixtures of Clayton, Frank, Gumbel–Barnett and Normal copulas. The strength of the dependence of the null copula C and of the perturbation copula D, as measured by Kendall's tau, are $(\tau_C, \tau_D) = (0.1, 0.5)$ in Figure V.1 and $(\tau_C, \tau_D) = (0.4, 0.8)$ in Figure V.2.

It is first interesting to note that surprisingly, the choice of the estimator has a significant impact on the local power curves in almost all cases considered, except under Normal mixtures.

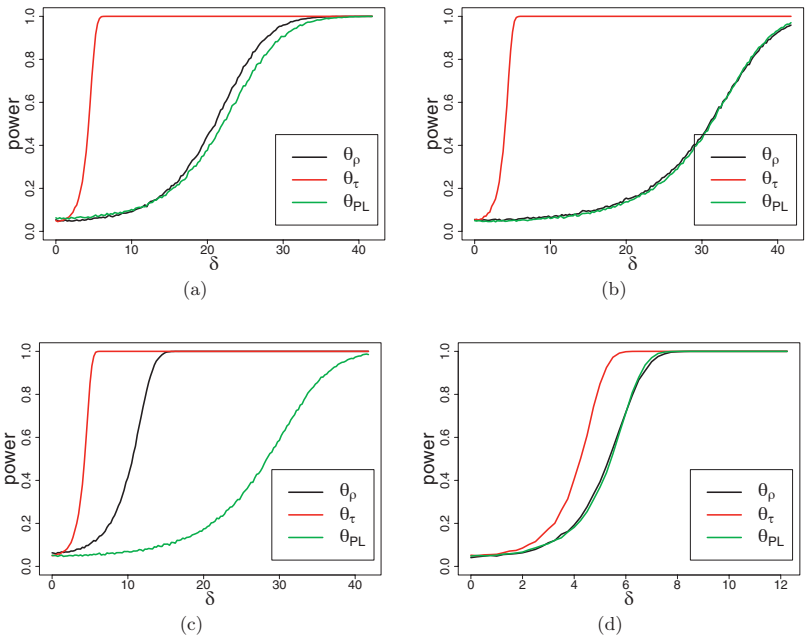


Figure V.1: Asymptotic local power curves of the tests based on $\mathcal{V}^{\rho}_{n,N}$, $\mathcal{V}^{\tau}_{n,N}$ and $\mathcal{V}^{PL}_{n,N}$ under mixtures of (a) Clayton, (b) Frank, (c) Gumbel–Barnett and (d) Normal copulas with $\tau_C = 0.1$ and $\tau_D = 0.5$.

Under Clayton alternatives, the conclusions are the same in Figure V.1 and Figure V.2, namely that $\mathcal{V}^{\tau}_{n,N}$ has a significantly much larger local power than its two competitors. Overall, $\mathcal{V}^{\rho}_{n,N}$ is the least powerful locally. Probably due to the fact that the drift terms μ_τ associated to the estimation by Kendall's tau are small (see Table V.1), $\mathcal{V}^{\tau}_{n,N}$ performs generally very well, especially in the case of small level of dependence, i.e. for $(\tau_C, \tau_D) = (0.1, 0.5)$. For higher degrees of dependence, $\mathcal{V}^{PL}_{n,N}$ is often better than $\mathcal{V}^{\tau}_{n,N}$ and constitutes a good choice under all scenarios, except for Clayton mixtures.

V.4.2 Comparison of the empirical copula process with the moment-based statistics

In view of Propositions 3 and 4, the asymptotic local power curves $\beta_1, \ldots, \beta_4$ of the goodness-of-fit tests based on $\mathcal{S}_{n1}, \ldots, \mathcal{S}_{n4}$ are of the form

$$\beta_j(\delta, C_\theta, D) = 1 - \Phi\left\{z_{\alpha/2} - \left|\frac{\delta\mu_j(C_\theta, D)}{\sigma_j(C_\theta)}\right|\right\} + \Phi\left\{-z_{\alpha/2} - \left|\frac{\delta\mu_j(C_\theta, D)}{\sigma_j(C_\theta)}\right|\right\}, \tag{V.10}$$

where $z_{\alpha/2}$ is the $(1-\alpha/2)$-th percentile of a $\mathcal{N}(0,1)$ distribution. Here, $\mu_1 = \mu_\rho - \mu_\tau$, $\mu_2 = \mu_\rho - \mu_{PL}$, $\mu_3 = \mu_\tau - \mu_{PL}$ and $\mu_4 = \eta_1 - \eta_2$. In view of equation (V.10), the local power of the test based on $\mathcal{S}_{nj}$ only depends on the absolute value of the ratio $\mu_j(C_\theta, D)/\sigma_j(C_\theta)$, i.e. the asymptotic local efficiency. Some values of μ_1, μ_2 and μ_3 are reported in Table V.2 under the four choices of

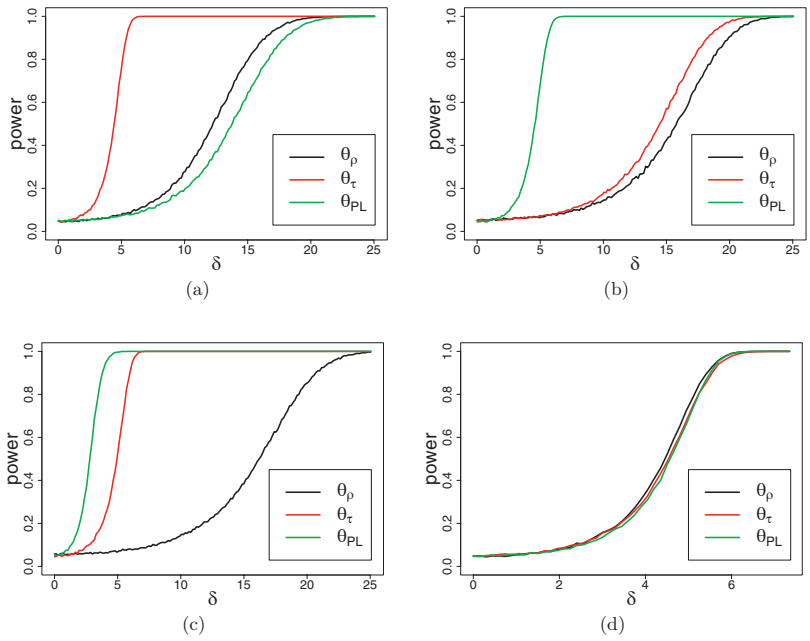


Figure V.2: Asymptotic local power curves of the tests based on $\mathcal{V}_{n,N}^{\rho}$, $\mathcal{V}_{n,N}^{\tau}$ and $\mathcal{V}_{n,N}^{PL}$ under mixtures of (a) Clayton, (b) Frank, (c) Gumbel–Barnett and (d) Normal copulas with $\tau_C = 0.4$ and $\tau_D = 0.8$.

mixture distributions. The highest local efficiencies, i.e. the one that yields the most power locally among the three, are identified in bold.

Table V.2 establishes a clear picture of which statistic is the best under a given scenario of mixture distributions: for Clayton, Gumbel–Barnett and Normal mixtures, $\mathcal{S}_{n1}$ is the most powerful locally, while $\mathcal{S}_{n3}$ is the best for local mixtures of Frank copulas. The test statistic $\mathcal{S}_{n2}$ is very poor in all cases, except when $(\tau_C, \tau_D) = (0.4, 0.8)$ under Gumbel-Barnett alternatives. It is also interesting to note that under Clayton mixtures, $\mathcal{S}_{n1}$ performs better than Shih's statistic $\mathcal{S}_{n4}$, even if the latter is specifically conceived for this particular case. To come to this conclusion, note that $|\mu_4|/\sigma_4 = 0.655$ when $(\tau_C, \tau_D) = (0.1, 0.5)$ and $|\mu_4|/\sigma_4 = 0.347$ when $(\tau_C, \tau_D) = (0.4, 0.8)$.

Figure V.3 compares the local power curves of $\mathcal{S}_{n1}$, $\mathcal{S}_{n2}$ and $\mathcal{S}_{n3}$ to the best statistic among $\mathcal{V}_{n,N,\rho}$, $\mathcal{V}_{n,N,\tau}$ and $\mathcal{V}_{n,N,PL}$ according to the results of subsection V.4.1. Only the case $(\tau_C, \tau_D) = (0.4, 0.8)$ is considered. For the mixture of Clayton copulas, the goodness-of-fit statistic of Shih, suitable only for this family, is also investigated.

The test statistic $\mathcal{S}_{n1}$ exhibit high power locally in all cases, while $\mathcal{S}_{n3}$ also performs very well. The most surprising discovery here is the rather poor performance of the Cramér-von Mises statistics compared to the very simple, asymptotically normal moment-based statistics. These conclusions must however be treated with care since the nature of the alternative distributions considered could have favored the moment-based statistics. Nevertheless, the latter deserve further investigations under other types of alternatives. Also, multivariate extensions of $\mathcal{S}_{n1}, \ldots, \mathcal{S}_{n4}$ could be considered as serious competitors to $\mathcal{V}_{n,N}^{\rho}$, $\mathcal{V}_{n,N}^{\tau}$ and $\mathcal{V}_{n,N}^{PL}$, the latter being very costly in terms of computing time.

Table V.2: Asymptotic local efficiency terms for the test statistics $\mathcal{S}_{n1}$, $\mathcal{S}_{n2}$ and $\mathcal{S}_{n3}$ under mixtures of Clayton, Frank, Gumbel-Barnett and Normal copulas

τ_{C_θ}	τ_D	*Mixture* model	$\mathcal{S}_{n1}$ $\lvert\mu_1\rvert/\sigma_1$	$\mathcal{S}_{n2}$ $\lvert\mu_2\rvert/\sigma_2$	$\mathcal{S}_{n3}$ $\lvert\mu_3\rvert/\sigma_3$	Mixture model	$\mathcal{S}_{n1}$ $\lvert\mu_1\rvert/\sigma_1$	$\mathcal{S}_{n2}$ $\lvert\mu_2\rvert/\sigma_2$	$\mathcal{S}_{n3}$ $\lvert\mu_3\rvert/\sigma_3$
0.1	0.2		**1.627**	0.006	0.227		3.329	0.065	**4.269**
0.1	0.3		**3.163**	0.013	0.442		6.566	0.067	**8.298**
0.1	0.4		**4.608**	0.009	0.632		9.617	0.230	**12.420**
0.1	0.5	Clayton	**5.894**	0.005	0.794	Frank	12.432	0.346	**16.155**
0.4	0.5		**0.762**	0.007	0.234		1.162	0.039	**1.873**
0.4	0.6		**1.438**	0.000	0.426		2.134	0.115	**3.547**
0.4	0.7		**1.989**	0.004	0.594		2.916	0.290	**5.169**
0.4	0.8		**2.403**	0.046	0.765		3.494	0.468	**6.487**
0.1	0.2		**1.920**	0.006	0.289		**3.971**	0.000	0.444
0.1	0.3		**3.728**	0.019	0.568		**7.765**	0.003	0.871
0.1	0.4		**5.446**	0.029	0.831		**11.353**	0.010	1.278
0.1	0.5	Gumbel–	**9.732**	0.338	1.084	Normal	**14.529**	0.023	1.646
0.4	0.5	Barnett	**0.795**	0.017	0.361		**1.459**	0.011	0.446
0.4	0.6		**1.491**	0.056	0.706		**2.748**	0.007	0.824
0.4	0.7		**2.048**	0.112	1.012		**3.768**	0.011	1.131
0.4	0.8		2.336	1.725	**3.099**		**4.462**	0.036	1.367

In bold, the most powerful statistic locally among $\mathcal{S}_{n1}$, $\mathcal{S}_{n2}$ and $\mathcal{S}_{n3}$

In some cases, e.g. in panel (b) of Figure V.3, it is difficult to decide whether $\mathcal{S}_{n2}$ performs better than $\mathcal{V}_{n,N}^{PL}$, locally. A way to circumvent this problem consists in computing some measure of asymptotic relative efficiency. This idea is developed in the next section.

V.5 Asymptotic relative efficiencies

V.5.1 A new ARE measure

For a goodness-of-fit statistic whose limiting distribution is normal with mean $\delta\mu(C_\theta, D)$ and variance $\sigma^2(C_\theta)$, the associated local power curve $\beta(\delta, C_\theta, D)$ is an increasing function of $\mu(C_\theta, D)/\sigma(C_\theta, D)$ for all fixed values of $\delta > 0$. It thus seems natural to compare the efficiency of two such statistics $\mathcal{S}_{nj}$ and $\mathcal{S}_{nk}$ via Pitman's measure of asymptotic relative efficiency (ARE), namely

$$\mathcal{ARE}_{\text{Pitman}}(\mathcal{S}_{nj}, \mathcal{S}_{nk}) = \left\{ \frac{\mu_j(C_\theta, D)/\sigma_j(C_\theta)}{\mu_k(C_\theta, D)/\sigma_k(C_\theta)} \right\}^2 .$$

However, it is not entirely clear how to extend this measure in the case when the limiting distribution of a test statistic is no longer normal, which is the case with many of the goodness-of-fit statistics. A generalization of Pitman's measure proposed by Genest et al. (2006b) and Genest et al. (2007) is

$$\widetilde{\mathcal{ARE}}(\mathcal{S}_{nj}, \mathcal{S}_{nk}) = \lim_{\delta \to 0} \frac{\beta_{\mathcal{S}_{nj}}(\delta) - \beta_{\mathcal{S}_{nj}}(0)}{\beta_{\mathcal{S}_{nk}}(\delta) - \beta_{\mathcal{S}_{nk}}(0)}$$

in terms of the local power functions $\beta_{\mathcal{S}_{nj}}$, $\beta_{\mathcal{S}_{nk}}$ of two tests $\mathcal{S}_{nj}$ and $\mathcal{S}_{nk}$. For most cases of interest, however, this measure requires the derivatives of the power curves in a neighborhood of $\delta = 0$. Since the asymptotic local power functions of the tests based on $\mathcal{V}_{n,N,\rho}$, $\mathcal{V}_{n,N,\tau}$ and $\mathcal{V}_{n,N,PL}$ admit no explicit representations, this causes a serious problem when trying to apply the latter definition.

(V.12)

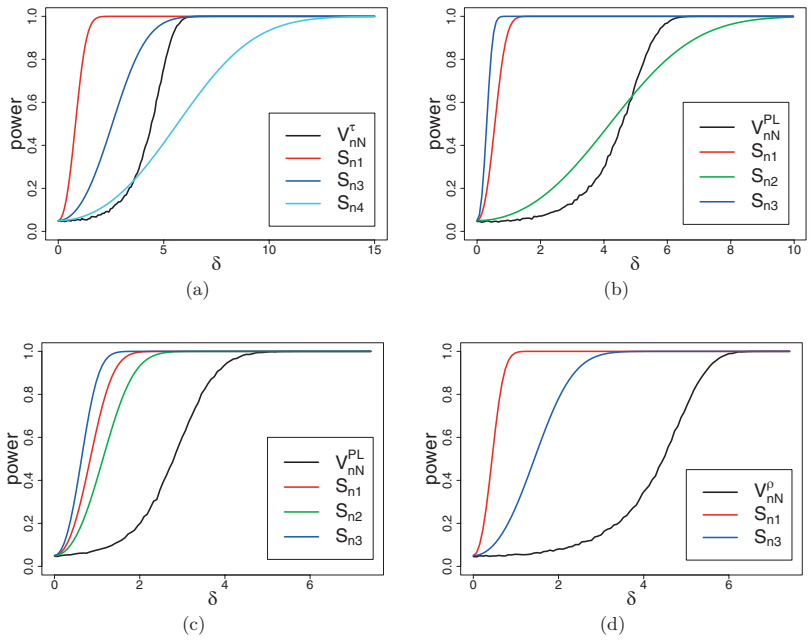


Figure V.3: Asymptotic local power of the Cramér-von Mises tests and of $\mathcal{S}_{n1}$, $\mathcal{S}_{n2}$, $\mathcal{S}_{n3}$ and $\mathcal{S}_{n4}$ under (a) Clayton, (b) Frank, (c) Gumbel-Barnett and (d) Normal mixtures with $\tau_C = 0.4$ and $\tau_D = 0.8$.

Here, another generalization of $\mathcal{ARE}_{\text{Pitman}}$ is proposed :

$$\mathcal{ARE}(\mathcal{S}_{nj}, \mathcal{S}_{nk}) = \left\{ \lim_{M\to\infty} \frac{\int_0^M \{1 - \beta_{\mathcal{S}_{nk}}(\delta)\}\,\mathrm{d}\delta}{\int_0^M \{1 - \beta_{\mathcal{S}_{nj}}(\delta)\}\,\mathrm{d}\delta} \right\}^2.$$

The first motivation for such a definition is the possibility to estimate $\int_0^M \{1 - \beta_{\mathcal{S}_{nj}}(\delta)\}\,\mathrm{d}\delta$ and $\int_0^M \{1 - \beta_{\mathcal{S}_{nk}}(\delta)\}\,\mathrm{d}\delta$ when accurate approximations $\hat{\beta}_{\mathcal{S}_{nj}}$ and $\hat{\beta}_{\mathcal{S}_{nk}}$ are available. This is the case for the power curves of the tests based on $\mathcal{V}^{\rho}_{n,N}$, $\mathcal{V}^{\tau}_{n,N}$ and $\mathcal{V}^{PL}_{n,N}$. To be specific, suppose $\hat{\beta}(\delta)$ is available at the points iM/N, $i = 1, \ldots, N$, for sufficiently large N in order to achieve some numerical accuracy. Upper and lower approximations of $\int_0^M \{1 - \beta_{\mathcal{S}_{nj}}(\delta)\}\,\mathrm{d}\delta$ are

$$I_1 = \frac{M}{N} \sum_{i=1}^{N} \left\{ 1 - \hat{\beta}\left(\frac{iM}{N}\right) \right\} \quad \text{and} \quad I_2 = \frac{M}{N} \sum_{i=0}^{N-1} \left\{ 1 - \hat{\beta}\left(\frac{iM}{N}\right) \right\},$$

and the chosen approximation, provided M is selected such that $\hat{\beta}(M) = 1$, is

$$\frac{I_1 + I_2}{2} = \frac{M}{N} \sum_{i=1}^{N-1} \left\{ 1 - \hat{\beta}\left(\frac{iM}{N}\right) \right\} + \frac{M}{N}\left(\frac{1-\alpha}{2}\right).$$

Another interesting feature of $\mathcal{ARE}(\mathcal{S}_{nj}, \mathcal{S}_{nk})$ is the fact that it generalizes Pitman's notion of asymptotic relative efficiency. To see this, let $\beta(\delta) = 1 - \Phi(z_{\alpha/2} - \delta\mu) + \Phi(-z_{\alpha/2} - \delta\mu)$ and

compute

$$\begin{aligned}\int_0^\infty \{1-\beta(\delta)\}\,\mathrm{d}\delta &= \int_0^\infty \Phi\left(z_{\alpha/2}-\delta\mu\right)\delta\delta - \int_0^\infty \Phi\left(-z_{\alpha/2}-\delta\mu\right)\mathrm{d}\delta \\ &= \frac{1}{\mu}\left\{\int_{-\infty}^{z_{\alpha/2}}\Phi(x)\mathrm{d}x - \int_{-\infty}^{-z_{\alpha/2}}\Phi(x)\mathrm{d}x\right\} = \frac{1}{\mu}\int_{-z_{\alpha/2}}^{z_{\alpha/2}}\Phi(x)\mathrm{d}x = \frac{z_{\alpha/2}}{\mu}.\end{aligned}$$

As a consequence, one has

$$\int_0^\infty \{1-\beta_j(\delta,C_\theta,D)\}\,\delta\delta = z_{\alpha/2}\left\{\frac{\mu_j(C_\theta,D)}{\sigma_j(C_\theta)}\right\}^{-1}$$

for local power functions of the form (V.10). Computations of $\mathcal{ARE}$ for some of the goodness-of-fit statistics encountered in this paper are provided in the next subsection.

V.5.2 Local efficiency comparisons

In all situations considered in subsection V.4.2, the best moment-based statistic locally outperform the best Cramér-von Mises statistic. Hence, it seems useless to compare the latter in terms of their asymptotic relative efficiency. However, since the power curves of $\mathcal{V}_{n,N}^{\rho}$, $\mathcal{V}_{n,N}^{\tau}$ and $\mathcal{V}_{n,N}^{PL}$ are often very close to each other, such computations could be very interesting. They are presented in Table V.3.

Table V.3: Estimated values of $\lim_{M\to\infty}\int_0^M\{1-\beta(\delta)\}\mathrm{d}\delta$ for the goodness-of-fit statistics $\mathcal{V}_{n,N}^{\rho}$, $\mathcal{V}_{n,N}^{\tau}$ and $\mathcal{V}_{n,N}^{PL}$ and asymptotic relative efficiencies under mixtures of Clayton, Frank, Gumbel-Barnett and Normal copulas.

Mixture model	τ_{C_θ}	τ_D	$\lim_{M\to\infty}\int_0^M\{1-\beta(\delta)\}\mathrm{d}\delta$ $\mathcal{V}_{n,N}^{\rho}$	$\mathcal{V}_{n,N}^{\tau}$	$\mathcal{V}_{n,N}^{PL}$	*Asymptotic relative efficiency* $(\mathcal{V}_{n,N}^{\rho},\mathcal{V}_{n,N}^{\tau})$	$(\mathcal{V}_{n,N}^{\rho},\mathcal{V}_{n,N}^{PL})$	$(\mathcal{V}_{n,N}^{\tau},\mathcal{V}_{n,N}^{PL})$
Clayton	0.1	0.5	12.018	2.540	12.618	0.211	1.050	4.968
	0.4	0.8	23.469	8.349	26.091	0.356	1.112	3.125
Frank	0.1	0.5	17.464	2.381	17.594	0.136	1.007	7.389
	0.4	0.8	29.483	27.079	8.670	0.918	0.294	0.320
Gumbel-Barnett	0.1	0.5	5.954	2.506	16.143	0.421	2.711	6.442
	0.4	0.8	30.369	9.282	5.475	0.306	0.180	0.590
Normal	0.1	0.5	3.142	2.491	3.150	0.793	1.003	1.265
	0.4	0.8	8.390	8.527	8.609	1.016	1.026	1.010

These computations show, among other things, that $\mathcal{V}_{n,N}^{\tau}$ is generally more powerful than $\mathcal{V}_{n,N}^{PL}$ for low dependence alternatives, i.e. close to independence. An opposite conclusion arises for mixture of high dependence copulas, namely when $(\tau_C,\tau_D)=(0.4,0.8)$. The performance of $V_{n,N}^{PL}$ and $V_{n,N}^{\rho}$ are quite similar for low dependence, except under Gumbel-Barnett mixtures. Overall, $V_{n,N}^{\tau}$ seems the best choice close to the independence copula, while $V_{n,N}^{PL}$ performs well under high levels of dependence.

Looking back at panel (b) of Figure V.3, it is difficult to decide whether $\mathcal{S}_{n2}$ performs better than $\mathcal{V}_{n,N}^{PL}$. Even though the local power curve of $\mathcal{V}_{n,N}^{PL}$ reaches 1 more quickly, the asymptotic relative efficiency is given by $\mathcal{ARE}(\mathcal{V}_{n,N}^{PL},\mathcal{S}_{n2})=0.950$, which supports the choice of $\mathcal{S}_{n2}$ over $\mathcal{V}_{n,N}^{PL}$ if a mixture of Frank distributions is suspected as a possible alternative.

V.6 Sensitivity in small samples

This section is devoted to the sensitivity in small samples and under fixed alternatives of the test statistics encountered in this paper, namely $\mathcal{V}_{n,N}^{\rho}$, $\mathcal{V}_{n,N,}^{\tau}$, $\mathcal{V}_{n,N}^{PL}$, $\mathcal{S}_{n1}$, $\mathcal{S}_{n2}$, $\mathcal{S}_{n3}$ and $\mathcal{S}_{n4}$. The main goal is to relate the asymptotic local efficiency results of Section V.4 and Section V.5 with empirical situations. In subsection V.6.1, the specific influence of the estimators on the power of the Cramér-von Mises statistics is investigated. In subsection V.6.2, comparisons with the moment-based statistics are made. These results will be paralleled with those presented in subsections V.4.1 and V.4.2 under contiguous sequences.

V.6.1 Influence of the estimators on the power of the Cramér-von Mises statistics

It was seen in subsection V.4.1 that the asymptotic local powers of the goodness-of-fit tests based on the empirical copula process are sensitive to the choice of the estimator of the dependence parameter, at least under the mixture distributions considered. In this section, the ability of $\mathcal{V}_{n,N}^{\rho}$, $\mathcal{V}_{n,N}^{\tau}$ and $\mathcal{V}_{n,N}^{PL}$ to reject the null hypothesis is first examined under fixed alternatives and many sample sizes. The results can be found in Tables V.4-V.7. First note that all 5% nominal levels are maintained, keeping in mind a margin of error of the magnitude of $\pm 1\%$ when estimating proportions from 10 000 replicates.

When Clayton's family of copulas is in the null hypothesis, one can see from Table V.4 that $\mathcal{V}_{n,N}^{\rho}$ performs very well against all alternatives, especially in small samples, while $\mathcal{V}_{n,N}^{\tau}$ is almost as powerful. The latter are significantly superior to $\mathcal{V}_{n,N}^{PL}$ under Gumbel-Barnett alternatives, especially in small samples. The performance of $\mathcal{V}_{n,N}^{PL}$ however surpasses that of $\mathcal{V}_{n,N}^{\rho}$ and $\mathcal{V}_{n,N}^{\tau}$ under Frank and Normal alternatives, and this advantage is particularly marked for higher degrees of dependence.

Things are much simpler in Table V.5 when testing the membership to Frank's family, where the three considered estimation strategies yield almost the same power for the Cramér-von Mises statistics. For the null hypothesis of belonging to Gumbel-Barnett's class, the statistic $\mathcal{V}_{n,N}^{PL}$ is remarkably better than its two competitors under Frank and Normal alternatives, especially for large sample sizes, as one can notice from the entries in Table V.6. An opposite conclusion must however be made under Clayton alternatives, where $\mathcal{V}_{n,N}^{\rho}$ and $\mathcal{V}_{n,N}^{\tau}$ are slightly better.

Finally, the most powerful statistics for testing the Normal hypothesis are $\mathcal{V}_{n,N}^{\rho}$ and $\mathcal{V}_{n,N}^{\tau}$ under Clayton alternatives, while $\mathcal{V}_{n,N}^{PL}$ is the best choice under observations that come from the Frank copula. Here again, the performance of the latter increases as the sample size becomes larger.

In a second series of analyses, the power of the Cramér-von Mises statistics under mixture distributions of the type $\mathcal{Q}_{\delta_n} = (1-\delta_n)C_\theta + \delta_n C_{\theta'}$ have been considered for samples of size $n = 500$. The corresponding empirical power curves are presented in Figure V.4. In this setting, $100\times\delta/\sqrt{500}$ % of the observations come from the distribution $C_{\theta'}$, so the power increases with δ. However, from a certain threshold, the observed powers suddenly decreases toward the nominal level. This occurs because $C_{\theta'}$ also belongs to the family of copulas under $\mathcal{H}_0$. One may have expected, however, that the powers would start to decrease at the middle point, i.e. when $\delta = \sqrt{500}/2 \approx 11.2$. The observed asymmetry in all four cases is probably an indication that the goodness-of-fit tests are better to detect discrepancies from $\mathcal{H}_0$ when the data come from a copula with a high level of dependence. The fact that $\theta' > \theta$ probably explained that the middle point is skewed to the right.

As expected, the differences in power between $\mathcal{V}_{n,N}^{\rho}$, $\mathcal{V}_{n,N}^{\tau}$ and $\mathcal{V}_{n,N}^{PL}$ are less apparent in small sample sizes than it was asymptotically (see Figure V.2 to compare). Nevertheless, the conclusions here are very similar to the asymptotic situation, except that the performance of $\mathcal{V}_{n,N}^{\rho}$ is not as bad as for $n \to \infty$ under Clayton and Gumbel-Barnett mixtures. Briefly, the choice of the estimator doesn't seem to have a significant influence under Gumbel-Barnett and Normal mixtures, while for Clayton mixtures, the pseudo-likelihood estimator is not recommended. The latter is however the best choice under Frank mixtures.

Table V.4: Estimated percentage of rejection of the null hypothesis of belonging to Clayton's family for the goodness-of-fit tests based on $\mathcal{V}_{n,N}^{\rho}$, $\mathcal{V}_{n,N}^{\tau}$, $\mathcal{V}_{n,N}^{PL}$, $\mathcal{S}_{n1}$, $\mathcal{S}_{n2}$ and $\mathcal{S}_{n3}$ under fixed copula alternatives.

		$\mathcal{H}_1$: Clayton						$\mathcal{H}_1$: Gumbel-Barnett					
n	τ	$\mathcal{V}_{n,N}^{\rho}$	$\mathcal{V}_{n,N}^{\tau}$	$\mathcal{V}_{n,N}^{PL}$	$\mathcal{S}_{n1}$	$\mathcal{S}_{n2}$	$\mathcal{S}_{n3}$	$\mathcal{V}_{n,N}^{\rho}$	$\mathcal{V}_{n,N}^{\tau}$	$\mathcal{V}_{n,N}^{PL}$	$\mathcal{S}_{n1}$	$\mathcal{S}_{n2}$	$\mathcal{S}_{n3}$
100	0.10	5.0	4.5	4.2	3.2	3.4	6.2	22.4	20.2	12.0	2.7	11.2	6.0
	0.15	5.7	5.0	5.2	3.1	4.2	7.2	38.7	36.6	24.0	2.3	18.6	11.0
	0.20	6.0	5.5	5.6	2.8	4.5	6.8	55.4	53.3	38.6	1.4	27.6	19.1
250	0.10	5.1	4.6	5.1	3.9	4.9	6.1	37.7	36.1	26.6	2.4	33.8	26.6
	0.15	5.3	5.1	5.0	4.0	4.9	5.9	65.4	64.6	53.2	1.7	58.2	49.9
	0.20	5.1	5.3	5.1	3.1	5.1	5.9	86.2	85.5	77.7	1.1	78.0	74.5
500	0.10	5.0	5.0	4.3	3.5	5.1	5.3	57.1	54.9	46.6	1.3	64.8	59.8
	0.15	5.6	5.1	4.8	4.8	4.9	5.7	86.6	86.2	79.9	1.4	90.1	88.2
	0.20	5.0	5.2	5.3	3.6	5.1	5.8	97.5	97.3	95.6	1.1	98.4	98.2
1000	0.10	5.1	5.1	4.7	3.0	5.2	5.4	73.3	73.8	69.7	0.6	90.5	89.9
	0.15	4.8	5.3	5.3	5.0	5.5	5.7	97.4	97.5	96.0	0.8	99.7	99.7
	0.20	5.1	5.3	5.3	4.8	4.9	5.2	99.9	100	99.9	2.2	100	100
2500	0.10	4.7	4.7	6.2	4.2	4.8	5.2	90.1	89.8	90.5	0.4	99.9	99.9
	0.15	4.8	4.8	5.5	4.7	4.3	4.6	99.9	99.9	99.9	0.3	100	100
	0.20	4.5	5.3	5.5	5.8	5.9	5.2	100	100	100	13.0	100	100
		$\mathcal{H}_1$: Frank						$\mathcal{H}_1$: Normal					
n	τ	$\mathcal{V}_{n,N}^{\rho}$	$\mathcal{V}_{n,N}^{\tau}$	$\mathcal{V}_{n,N}^{PL}$	$\mathcal{S}_{n1}$	$\mathcal{S}_{n2}$	$\mathcal{S}_{n3}$	$\mathcal{V}_{n,N}^{\rho}$	$\mathcal{V}_{n,N}^{\tau}$	$\mathcal{V}_{n,N}^{PL}$	$\mathcal{S}_{n1}$	$\mathcal{S}_{n2}$	$\mathcal{S}_{n3}$
100	0.10	12.8	11.6	8.1	3.9	7.5	3.8	12.3	11.2	6.7	3.4	5.6	3.1
	0.15	20.8	19.2	13.9	4.3	11.0	4.8	20.2	18.4	11.5	3.6	8.0	3.9
	0.20	31.3	29.5	23.9	5.1	17.6	9.3	29.4	26.9	17.9	3.2	11.5	6.1
250	0.10	18.8	18.3	17.8	4.7	24.0	16.4	18.6	17.3	12.7	4.0	17.4	11.4
	0.15	36.0	34.8	36.5	7.3	44.0	34.2	33.0	32.3	25.5	5.3	29.4	21.1
	0.20	55.7	54.6	58.3	9.3	64.7	56.5	49.7	47.8	41.2	5.0	44.3	36.1
500	0.10	28.7	27.1	32.0	5.3	49.9	42.8	25.8	24.6	20.8	3.9	35.8	29.1
	0.15	54.5	52.9	61.3	10.2	81.2	75.4	48.6	46.9	43.6	7.0	61.4	53.7
	0.20	77.1	76.1	84.7	14.1	95.3	93.6	69.7	68.4	66.7	7.9	81.8	78.0
1000	0.10	37.3	37.8	50.5	5.5	81.4	78.2	33.8	33.0	35.5	3.9	63.6	59.5
	0.15	72.3	72.3	83.5	15.7	98.7	98.1	66.7	65.8	65.8	10.2	92.5	89.4
	0.20	92.4	92.8	97.6	25.0	100	100	88.2	87.3	89.4	15.2	99.0	98.8
2500	0.10	50.2	48.5	73.3	9.0	99.6	99.6	43.4	42.4	52.4	5.8	96.1	95.9
	0.15	88.9	88.6	96.5	20.5	100	100	83.5	82.2	88.0	14.0	100	100
	0.20	99.1	99.2	99.9	40.4	100	100	97.5	97.8	98.6	28.7	100	100

Table V.5: Estimated percentage of rejection of the null hypothesis of belonging to Frank's family for the goodness-of-fit tests based on $\mathcal{V}^{\rho}_{n,N}$, $\mathcal{V}^{\tau}_{n,N}$, $\mathcal{V}^{PL}_{n,N}$, $\mathcal{S}_{n1}$, $\mathcal{S}_{n2}$ and $\mathcal{S}_{n3}$ under fixed copula alternatives.

		$\mathcal{H}_1$: Clayton						$\mathcal{H}_1$: Gumbel-Barnett					
n	τ	$\mathcal{V}^{\rho}_{n,N}$	$\mathcal{V}^{\tau}_{n,N}$	$\mathcal{V}^{PL}_{n,N}$	$\mathcal{S}_{n1}$	$\mathcal{S}_{n2}$	$\mathcal{S}_{n3}$	$\mathcal{V}^{\rho}_{n,N}$	$\mathcal{V}^{\tau}_{n,N}$	$\mathcal{V}^{PL}_{n,N}$	$\mathcal{S}_{n1}$	$\mathcal{S}_{n2}$	$\mathcal{S}_{n3}$
100	0.10	7.1	6.4	6.4	4.0	1.1	6.5	10.5	9.6	10.0	3.0	1.0	6.5
	0.15	11.3	10.2	10.4	3.6	1.3	8.3	15.2	14.2	14.6	2.4	1.0	8.5
	0.20	16.3	14.8	15.5	2.6	1.1	10.1	17.3	17.0	17.3	1.8	1.3	11.0
250	0.10	12.7	12.0	12.9	3.2	1.6	9.4	12.6	12.8	12.9	2.1	0.8	9.6
	0.15	24.8	24.8	26.0	2.3	1.4	15.1	19.7	19.6	20.3	0.9	1.6	15.4
	0.20	43.4	43.5	43.7	1.4	2.3	18.9	28.3	29.6	29.6	0.5	3.7	20.3
500	0.10	22.6	22.1	21.7	2.8	1.1	15.9	16.5	16.2	16.0	1.0	1.0	16.1
	0.15	47.3	47.1	47.0	1.8	2.4	25.9	28.3	28.6	28.8	0.5	5.1	27.1
	0.20	73.2	74.2	73.7	1.0	6.5	31.8	42.8	45.7	45.1	1.0	14.4	34.1
1000	0.10	36.4	39.3	38.5	2.2	1.5	26.8	21.4	22.9	22.7	0.4	3.6	29.0
	0.15	72.5	73.1	72.1	1.3	8.4	41.7	41.1	42.8	41.9	0.9	19.2	45.8
	0.20	92.8	92.9	93.2	1.3	18.0	47.9	60.6	62.1	63.9	8.0	37.1	54.7
2500	0.10	53.2	52.2	51.9	1.4	8.5	51.1	26.0	26.2	26.5	1.0	26.3	59.5
	0.15	90.6	91.1	91.5	1.1	32.3	74.8	53.6	56.0	56.9	19.8	66.0	82.4
	0.20	99.6	99.5	99.4	10.9	54.1	79.9	79.9	79.7	81.2	65.9	85.6	88.7

		$\mathcal{H}_1$: Frank						$\mathcal{H}_1$: Normal					
n	τ	$\mathcal{V}^{\rho}_{n,N}$	$\mathcal{V}^{\tau}_{n,N}$	$\mathcal{V}^{PL}_{n,N}$	$\mathcal{S}_{n1}$	$\mathcal{S}_{n2}$	$\mathcal{S}_{n3}$	$\mathcal{V}^{\rho}_{n,N}$	$\mathcal{V}^{\tau}_{n,N}$	$\mathcal{V}^{PL}_{n,N}$	$\mathcal{S}_{n1}$	$\mathcal{S}_{n2}$	$\mathcal{S}_{n3}$
100	0.10	6.0	5.6	5.5	4.6	2.3	4.3	6.0	5.1	5.4	3.8	2.1	4.7
	0.15	6.0	5.7	5.6	4.5	2.8	4.5	6.4	5.7	5.7	3.6	2.5	5.0
	0.20	5.4	5.4	5.2	4.5	3.3	4.7	6.7	6.4	6.7	3.3	2.8	5.9
250	0.10	4.8	4.9	4.9	4.2	3.1	4.5	6.2	5.6	6.0	3.3	3.2	5.3
	0.15	4.8	4.7	4.7	3.9	3.4	4.6	6.6	6.0	6.3	2.8	3.5	6.6
	0.20	4.5	5.1	4.8	4.2	3.6	4.7	8.3	7.7	7.9	2.0	3.0	7.8
500	0.10	4.6	4.6	4.5	3.9	4.1	4.6	6.2	5.6	5.4	2.7	3.7	6.6
	0.15	4.7	4.9	4.6	4.5	4.2	5.2	8.0	7.5	7.6	2.2	4.3	8.1
	0.20	5.0	5.3	5.1	4.7	4.4	5.1	10.8	11.4	10.1	1.6	4.1	8.6
1000	0.10	4.3	5.9	5.1	4.7	4.7	5.3	7.3	7.6	6.8	2.5	4.1	8.2
	0.15	5.4	5.7	5.1	5.0	5.0	5.1	9.8	10.3	9.2	1.6	3.9	8.9
	0.20	4.8	5.1	5.1	4.7	4.9	4.7	14.7	14.3	13.8	1.0	3.5	8.8
2500	0.10	5.3	5.1	4.3	4.6	4.6	4.4	7.6	7.4	6.8	1.7	4.5	9.0
	0.15	5.0	5.5	5.0	4.8	5.1	5.7	10.9	11.3	11.2	1.2	4.8	11.2
	0.20	5.4	5.0	5.3	4.2	4.7	4.5	17.2	16.2	17.3	0.6	4.0	9.5

Table V.6: Estimated percentage of rejection of the null hypothesis of belonging to Gumbel-Barnett's family for the goodness-of-fit tests based on $\mathcal{V}_{n,N}^{\rho}$, $\mathcal{V}_{n,N}^{\tau}$, $\mathcal{V}_{n,N}^{PL}$, $\mathcal{S}_{n1}$, $\mathcal{S}_{n2}$ and $\mathcal{S}_{n3}$ under fixed copula alternatives.

		$\mathcal{H}_1$: Clayton						$\mathcal{H}_1$: Gumbel–Barnett					
n	τ	$\mathcal{V}_{n,N}^{\rho}$	$\mathcal{V}_{n,N}^{\tau}$	$\mathcal{V}_{n,N}^{PL}$	$\mathcal{S}_{n1}$	$\mathcal{S}_{n2}$	$\mathcal{S}_{n3}$	$\mathcal{V}_{n,N}^{\rho}$	$\mathcal{V}_{n,N}^{\tau}$	$\mathcal{V}_{n,N}^{PL}$	$\mathcal{S}_{n1}$	$\mathcal{S}_{n2}$	$\mathcal{S}_{n3}$
100	0.10	9.3	7.9	5.4	6.2	16.5	5.8	3.6	3.7	4.9	5.7	5.9	7.0
	0.15	18.8	16.8	12.1	7.6	22.4	7.2	4.7	4.3	5.2	5.6	5.1	6.3
	0.20	31.5	29.9	22.2	7.8	26.7	9.6	4.9	4.7	5.3	5.3	4.4	4.8
250	0.10	26.8	25.1	20.7	8.4	36.3	24.2	4.6	4.5	5.4	5.5	5.8	5.9
	0.15	53.9	52.4	45.7	9.4	53.1	38.4	4.8	4.9	4.9	4.9	5.0	4.8
	0.20	78.9	77.5	70.6	10.4	67.9	55.1	5.2	5.0	5.0	5.1	4.7	4.6
500	0.10	48.8	48.0	41.7	9.9	59.8	51.8	4.7	5.1	4.8	4.8	4.9	4.8
	0.15	83.2	83.0	78.6	12.4	83.8	78.1	5.1	5.0	5.6	4.1	5.0	4.7
	0.20	96.9	96.8	95.6	13.0	94.2	92.3	4.9	4.4	4.7	4.4	5.1	4.9
1000	0.10	73.5	72.4	69.1	13.6	88.8	85.9	5.0	4.8	5.6	4.8	4.9	4.8
	0.15	97.5	97.0	96.7	19.3	98.9	98.7	5.1	4.9	5.3	4.5	4.8	5.1
	0.20	100	100	99.9	22.5	99.9	99.9	5.3	4.9	5.3	4.6	5.0	4.8
2500	0.10	92.6	91.6	90.4	16.3	99.9	99.9	5.8	4.9	5.0	3.5	5.6	5.4
	0.15	99.9	99.9	99.9	35.0	100	100	5.6	5.2	5.6	4.4	5.2	4.9
	0.20	100	100	100	41.7	100	100	5.5	5.0	5.3	5.2	5.1	5.5
		$\mathcal{H}_1$: Frank						$\mathcal{H}_1$: Normal					
n	τ	$\mathcal{V}_{n,N}^{\rho}$	$\mathcal{V}_{n,N}^{\tau}$	$\mathcal{V}_{n,N}^{PL}$	$\mathcal{S}_{n1}$	$\mathcal{S}_{n2}$	$\mathcal{S}_{n3}$	$\mathcal{V}_{n,N}^{\rho}$	$\mathcal{V}_{n,N}^{\tau}$	$\mathcal{V}_{n,N}^{PL}$	$\mathcal{S}_{n1}$	$\mathcal{S}_{n2}$	$\mathcal{S}_{n3}$
100	0.10	4.2	3.8	3.2	7.7	13.2	4.7	4.0	3.6	3.0	7.2	10.3	4.1
	0.15	5.4	5.2	5.1	9.7	16.4	4.6	4.9	4.3	3.8	8.5	11.9	3.7
	0.20	7.3	7.3	7.1	12.9	20.1	4.8	5.6	5.0	4.4	9.5	12.2	3.0
250	0.10	7.2	7.4	10.1	10.2	27.7	16.5	6.4	6.3	6.6	9.5	18.4	10.7
	0.15	12.8	13.0	18.1	15.3	42.0	26.0	10.4	9.8	9.8	11.9	25.3	13.4
	0.20	18.2	18.4	27.2	22.1	55.8	38.8	15.2	13.7	13.6	16.1	32.1	17.9
500	0.10	12.4	12.9	21.4	14.4	48.7	38.8	10.3	10.2	12.0	12.6	31.4	22.4
	0.15	22.2	23.1	37.5	24.1	73.3	63.6	18.3	16.8	19.4	18.5	47.3	35.2
	0.20	36.2	35.4	56.1	36.5	89.9	84.4	26.4	23.9	28.6	27.7	62.7	49.7
1000	0.10	18.7	19.2	36.8	19.5	79.6	74.5	14.9	14.2	20.7	16.9	57.4	50.1
	0.15	37.0	36.1	62.3	38.8	97.0	95.7	28.1	25.6	35.5	32.3	79.8	73.5
	0.20	56.9	57.1	81.9	58.3	99.7	99.6	41.0	40.7	51.6	50.3	91.1	86.3
2500	0.10	28.4	25.5	54.8	25.8	99.6	99.5	23.1	20.3	31.8	21.6	94.3	92.5
	0.15	54.7	55.1	81.4	65.9	100	100	39.8	40.6	54.8	62.1	99.4	99.2
	0.20	79.8	79.5	95.2	88.0	100	100	62.6	60.5	72.8	85.1	100	100

Table V.7: Estimated percentage of rejection of the null hypothesis of belonging to the Normal family for the goodness-of-fit tests based on $\mathcal{V}^{\rho}_{n,N}$, $\mathcal{V}^{\tau}_{n,N}$, $\mathcal{V}^{PL}_{n,N}$, $\mathcal{S}_{n1}$, $\mathcal{S}_{n2}$ and $\mathcal{S}_{n3}$ under fixed copula alternatives.

		$\mathcal{H}_1$: Clayton						$\mathcal{H}_1$: Gumbel–Barnett					
n	τ	$\mathcal{V}^{\rho}_{n,N}$	$\mathcal{V}^{\tau}_{n,N}$	$\mathcal{V}^{PL}_{n,N}$	$\mathcal{S}_{n1}$	$\mathcal{S}_{n2}$	$\mathcal{S}_{n3}$	$\mathcal{V}^{\rho}_{n,N}$	$\mathcal{V}^{\tau}_{n,N}$	$\mathcal{V}^{PL}_{n,N}$	$\mathcal{S}_{n1}$	$\mathcal{S}_{n2}$	$\mathcal{S}_{n3}$
100	0.10	4.8	4.4	4.5	4.9	5.0	6.2	7.1	6.8	7.0	4.1	5.4	6.7
	0.15	7.7	7.5	7.7	4.3	3.9	6.0	10.5	9.6	9.6	2.9	5.1	6.9
	0.20	12.3	12.2	11.4	4.3	4.2	5.9	13.9	13.5	13.5	2.4	5.1	6.9
250	0.10	10.5	9.6	9.5	5.0	4.2	6.8	10.3	9.1	9.7	2.8	5.4	8.3
	0.15	21.7	21.4	19.6	4.5	4.5	7.3	15.5	15.6	14.6	2.0	5.5	8.2
	0.20	36.6	37.4	32.9	3.2	3.9	6.5	21.5	21.4	19.7	1.2	5.3	8.0
500	0.10	19.7	20.5	16.9	4.8	5.5	8.3	14.1	13.7	12.8	2.1	6.6	9.6
	0.15	41.3	42.3	36.7	3.5	5.8	8.7	22.6	22.8	19.8	0.8	7.2	9.5
	0.20	65.1	65.4	58.9	2.6	5.3	7.2	32.9	32.8	28.8	1.6	6.6	8.3
1000	0.10	33.7	31.8	28.2	4.6	8.0	10.5	18.6	16.8	15.9	1.1	9.5	11.7
	0.15	64.3	64.2	59.1	2.7	10.4	12.7	31.3	30.2	27.7	1.7	11.5	12.1
	0.20	87.8	88.2	84.2	2.6	9.8	11.1	46.0	48.4	43.4	10.2	10.9	10.3
2500	0.10	43.2	45.1	41.9	3.0	15.3	17.8	18.1	19.7	18.7	1.6	18.4	18.4
	0.15	83.1	85.1	82.1	3.3	21.9	23.2	37.9	41.6	37.8	27.0	21.8	18.3
	0.20	98.6	98.5	98.0	11.2	25.8	24.9	61.9	62.2	59.0	67.0	26.7	18.4

		$\mathcal{H}_1$: Frank						$\mathcal{H}_1$: Normal					
n	τ	$\mathcal{V}^{\rho}_{n,N}$	$\mathcal{V}^{\tau}_{n,N}$	$\mathcal{V}^{PL}_{n,N}$	$\mathcal{S}_{n1}$	$\mathcal{S}_{n2}$	$\mathcal{S}_{n3}$	$\mathcal{V}^{\rho}_{n,N}$	$\mathcal{V}^{\tau}_{n,N}$	$\mathcal{V}^{PL}_{n,N}$	$\mathcal{S}_{n1}$	$\mathcal{S}_{n2}$	$\mathcal{S}_{n3}$
100	0.10	4.1	4.0	4.3	5.2	11.7	7.4	4.5	4.1	4.1	4.9	6.0	5.1
	0.15	4.5	4.4	5.1	5.9	15.6	10.0	4.8	4.7	4.4	4.9	5.6	4.7
	0.20	5.9	6.1	7.3	7.3	20.9	14.7	4.7	4.7	5.0	5.3	6.3	5.4
250	0.10	5.0	5.0	5.8	6.1	15.9	11.2	4.6	4.1	4.2	5.2	5.2	5.2
	0.15	6.3	6.7	9.7	7.2	25.6	19.6	4.8	4.9	5.1	5.1	5.3	5.3
	0.20	7.3	7.8	12.7	9.4	35.3	30.2	4.5	4.8	4.6	4.9	5.0	5.4
500	0.10	6.1	6.6	8.1	6.8	22.7	18.2	5.0	5.0	4.5	4.9	4.7	5.4
	0.15	7.1	8.2	13.2	8.8	36.9	31.5	5.2	5.4	4.7	5.1	5.2	5.6
	0.20	9.5	11.1	20.1	11.1	55.2	51.2	4.9	4.9	4.6	5.2	4.9	4.9
1000	0.10	7.5	7.1	11.2	8.4	34.9	30.0	6.2	5.2	5.0	5.5	5.3	5.1
	0.15	8.7	9.5	19.5	11.0	59.8	55.3	5.5	5.1	4.9	5.4	4.8	4.6
	0.20	13.2	14.7	31.5	13.8	82.2	80.4	5.0	5.1	5.4	4.5	5.1	5.1
2500	0.10	6.4	7.2	14.1	9.8	63.7	60.7	4.3	5.0	4.9	4.5	4.8	4.8
	0.15	10.0	11.7	26.7	14.5	91.7	90.5	4.4	5.2	5.1	5.4	4.8	4.7
	0.20	17.7	18.1	45.1	17.0	99.3	99.4	5.2	5.1	5.3	5.0	5.3	5.7

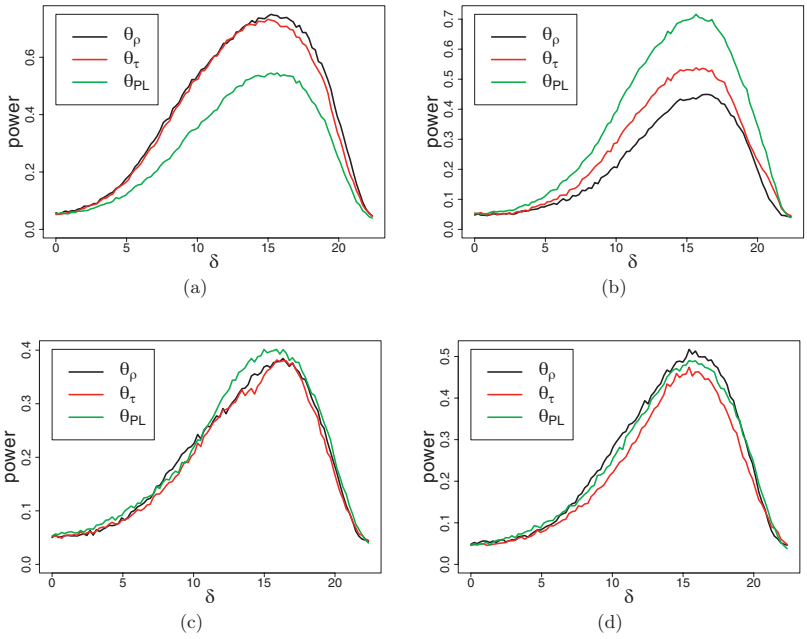


Figure V.4: Power curves for the tests based on $\mathcal{V}_{n,N}^{\rho}$, $\mathcal{V}_{n,N}^{\tau}$ and $\mathcal{V}_{n,N}^{PL}$ under (a) Clayton, (b) Frank, (c) Gumbel-Barnett and (d) Normal mixtures with $(\tau_C, \tau_d) = (0.4, 0.8)$, $n = 500$ and $N = 2500$.

V.6.2 Power of the Cramér-von Mises statistics compared to the moment-based statistics

It was seen in subsection V.6.1 that the test statistic $\mathcal{V}_{n,N}^{\rho}$ was a good choice for small sample sizes when testing the goodness-of-fit under the hypothesis of belonging to the Clayton family. The ability to reject $\mathcal{H}_0$ in that case is almost as good for tests based on $\mathcal{S}_{n2}$ and $\mathcal{S}_{n3}$, with a slight advantage to $\mathcal{S}_{n2}$. The power of the latter even becomes larger than that of $\mathcal{V}_{n,N}^{\rho}$ when $n \geq 500$ and is often better than the best Cramér-von Mises statistic in large samples, namely $\mathcal{V}_{n,N}^{PL}$. Note the poor performance of $\mathcal{S}_{n1}$ in all cases considered.

When testing the hypothesis of belonging to the Frank family, $\mathcal{S}_{n1}$ and $\mathcal{S}_{n2}$ are bad choices. However, $\mathcal{S}_{n3}$ is sometimes comparable with the Cramér-von Mises statistics when the sample size is large, especially under Gumbel-Barnett alternatives.

The null hypothesis of a Gumbel-Barnett family provides an example of a very powerful moment-based statistic. Here, $\mathcal{S}_{n2}$ is more powerful than the best Cramér-von Mises statistic, namely $\mathcal{V}_{n,N}^{\rho}$ under Clayton and $\mathcal{V}_{n,N}^{PL}$ under Frank and Normal copulas. Another example is given when testing the hypothesis of belonging to the Normal family against Frank alternatives, where $\mathcal{S}_{n2}$ and $\mathcal{S}_{n3}$ are clearly the most powerful. The latter are unfortunately inefficient to detect Clayton and Gumbel-Barnett dependence structures.

A final analysis have been made to compare the power of the tests under $\mathcal{Q}_{\delta_n} = (1 - \delta_n)C_\theta + \delta_n C_{\theta'}$. The results are to be found in Figure V.5. Here, the ordering in the power curves are often quite different to the ones encountered in Figure V.3 in the asymptotic situation. An explanation

probably lies in the fact that the moment-based statistics are especially good in very large samples, and the result is that the latter outclass the Cramér-von Mises statistics when $n \to \infty$. This domination is weaker in moderate sample sizes. This is particularly evident under Clayton mixtures where the best Cramér-von Mises statistic outperforms all moment-based statistics. Note here the very poor performance of $\mathcal{S}_{n1}$, in contrast to the extremely good performance of the same statistic when $n \to \infty$. Under Frank mixtures the moment-based statistics perform very well even for moderate sample sizes, where they outperform the best Cramér-von Mises statistic. Under Gumbel-Barnett mixtures, $\mathcal{S}_{n1}$ is clearly the best statistic while under Normal mixtures, $\mathcal{S}_{n3}$ is the best and $\mathcal{S}_{n1}$ provides a very poor performance.

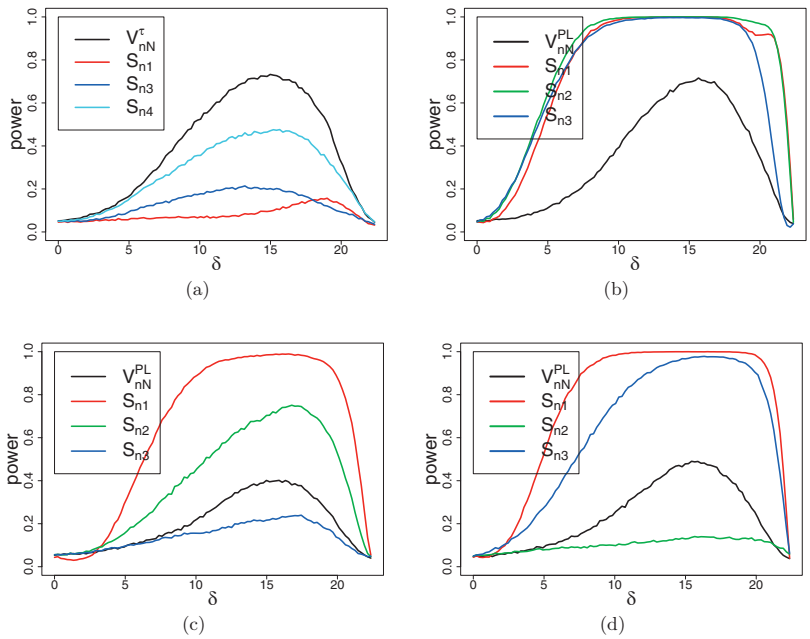


Figure V.5: Power of the tests based on $\mathcal{V}_{n,N}$, $\mathcal{S}_{n1}$, $\mathcal{S}_{n2}$, $\mathcal{S}_{n3}$ and $\mathcal{S}_{n4}$ when $n = 500$ under (a) Clayton, (b) Frank, (c) Gumbel-Barnett and (d) Normal mixtures with $\tau_C = 0.4$ and $\tau_D = 0.8$

V.7 Discussion

In this paper, the local power curves of tests based on Cramér-von Mises distances of the empirical copula goddness-of-fit process have been investigated and compared to that of moment-based statistics involving Spearman's rho, Kendall's tau and the pseudo-maximum likelihood estimator. Many discoveries have been made, in particular that the estimation strategy can have a significant impact on the power of the Cramér-von Mises statistics, and that some of the moment-based statistics provide very powerful tests under many distributional scenarios. Also, it seems that the ability of the Cramér-von Mises statistics to detect departures from $\mathcal{H}_0$ is better under fixed alternatives rather than under mixtures, while an opposite conclusion can be expressed for the moment-based statistics.

In future works, these kind of investigations could also be accomplished for other popular goodness-of-fit tests like those proposed by Scaillet (2006), Huard et al. (2006) and Genest et al. (2006a). The latter authors based their tests on Kendall's process $\mathcal{K}_n(t) = \sqrt{n}\{K_n(t) - K_{\hat{\theta}_n}(t)\}$, where $K_\theta(t) = \mathrm{P}\{C_\theta(X,Y) \leq t\}$, with $(X,Y) \sim C_\theta$, is the bivariate probability integral transformation of C_θ and K_n is a fully nonparametric estimator of K_θ. Suitable adaptations of the arguments to be found in Ghoudi and Rémillard (1998) should enable to establish that $\mathcal{K}_n \rightsquigarrow \mathcal{K} + \delta(\dot{L}_0 - \mu\dot{K}_\theta)$ under alternatives of the type $\mathcal{Q}_{\delta_n}$, where $\mathcal{K}$ is the weak limit of $\mathcal{K}_n$ under $\mathcal{H}_0$, L_δ is the probability integral transformation of $\mathcal{Q}_\delta$ and μ is the drift term associated to the limit of $\Theta_n = \sqrt{n}(\hat{\theta}_n - \theta)$ identified in Proposition 1.

It could also be interesting to exploit the idea of moment-based statistics to test the fit to families of multivariate copulas. For example, possible estimators of a univariate parameter θ are those based on inversions of the multivariate extensions of Spearman's rho described by Schmid and Schmidt (2007), namely

$$\rho_{n,\star} = \xi(d)\left\{2^d \int_{(0,1)^d} C_n(u)\mathrm{d}u - 1\right\} \quad \text{and} \quad \rho_{n,\star\star} = \xi(d)\left\{2^d \int_{(0,1)^d} \bar{C}_n(u)\mathrm{d}u - 1\right\},$$

where $\xi(d) = (d+1)(2^d - d - 1)^{-1}$, C_n is the multivariate empirical copula and $\bar{C}_n$ is the survival version of C_n. Then, the local behavior of the goodness-of-fit statistic

$$\mathcal{S}_n = \sqrt{n}\left\{\rho_\star^{-1}\left(\rho_{n,\star}\right) - \rho_{\star\star}^{-1}\left(\rho_{n,\star\star}\right)\right\},$$

where $\rho_\star$ and $\rho_{\star\star}$ are the population versions of $\rho_{n,\star}$ and $\rho_{n,\star\star}$, will be a consequence of that of $\mathcal{C}_{n,\theta}$ that can be deduced from the proof of Proposition 2.

It may be noted that the form of the alternative hypothesis (V.1) is not the only one under which asymptotic power curves could be derived. Another possibility is given by

$$\mathcal{Q}_\delta^\star(x,y) = \psi_\delta^{-1}\left[C\left\{\psi_\delta(x), \psi_\delta(y)\right\}\right],$$

where ψ_δ must satisfy some conditions to ensure that $\mathcal{Q}_\delta^\star$ is a copula and the perturbation function ψ_δ is chosen such that $\psi_0(t) = t$. Then, by arguments similar to that in the proof of Proposition 2, it would be possible to establish that $\mathcal{C}_{n,\theta} \rightsquigarrow \mathcal{C}_\theta + \delta\dot{\mathcal{Q}}_0^\star$, where

$$\dot{\mathcal{Q}}_0^\star(x,y) = C_{10}(x,y)\dot{\psi}_0(x) + C_{01}(x,y)\dot{\psi}_0(y) - \dot{\psi}_0\left\{C(x,y)\right\}.$$

Acknowledgements: Partial funding in support of this work was provided by the Norwegian Research Council (grant number 154079/420) and by the Natural Sciences and Engineering Research Council of Canada.

V.A Proofs

V.A.1 Proof of Proposition 1

Assumption (V.8) enables to deduce, from Lemma 3.10.11 of van der Vaart and Wellner (1996), that the log-likelihood ratio of $\mathcal{Q}_{\delta_n}$ with respect to $\mathcal{Q}_0$ has the asymptotic representation

$$L_n = \frac{\delta}{\sqrt{n}}\sum_{i=1}^n \left\{\frac{d(X_i,Y_i) - c_\theta(X_i,Y_i)}{c_\theta(X_i,Y_i)}\right\} - \frac{\delta^2}{2n}\sum_{i=1}^n \left\{\frac{d(X_i,Y_i) - c_\theta(X_i,Y_i)}{c_\theta(X_i,Y_i)}\right\}^2 + o_{\mathrm{P}}(1).$$

The proofs for (i) and (ii) are achieved in separate steps.

(i) From the asymptotic representation (V.9), it follows that

$$\Theta_{n,\Lambda} = \Theta'_{n,\Lambda} + \frac{1}{n}\sum_{i=1}^n \Lambda_{C_\theta,10}(X_i,Y_i)\beta_{n1}(X_i) + \frac{1}{n}\sum_{i=1}^n \Lambda_{C_\theta,01}(X_i,Y_i)\beta_{n2}(y_i) + o_{\mathrm{P}}(1),$$

(V.22)

where $\beta_{n1}(x) = \sqrt{n}\{F_n(x) - x\}$ and $\beta_{n2}(y) = \sqrt{n}\{G_n(y) - y\}$. From Slutsky's lemma, the bivariate central limit theorem and arguments that one can find in Ghoudi and Rémillard (1998), the vector $(\Theta_{n,\Lambda}, L_n)$ converges to a bivariate normal distribution with mean vector and covariance matrix

$$\mu = \left(0, \frac{-\delta^2\sigma^2(L)}{2}\right) \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_\Lambda^2(C_\theta) & \delta\mu_\Lambda(C_\theta, D) \\ \delta\mu_\Lambda(C_\theta, D) & \delta^2\sigma_\Lambda^2(\mathcal{Q}_\delta) \end{pmatrix},$$

where $\sigma^2(L) = \mathrm{Var}_{C_\theta}\{d(X,Y)/c_\theta(X,Y)\}$. One may then conclude, in view of Lecam's third lemma, that $\Theta_{n,\Lambda}$ is asymptotically normal with mean $\delta\mu_\Lambda(C_\theta, D)$ and variance $\sigma_\Lambda^2(C_\theta)$ under the contiguous sequence $(\mathcal{Q}_{\delta_n})_{n\geq 1}$.

(ii) From Hájek's projection method (Hájek and Sidák, 1967), one deduces the large-sample representation

$$\Theta_{n,\tau} = \frac{4}{\tau'_{C_\theta}(\theta)} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ 2C_\theta(X_i, Y_i) - X_i - Y_i + \frac{1 - \tau_{C_\theta}(\theta)}{2} \right\} + o_{\mathrm{P}}(1).$$

Hence, the vector (Θ_n, L_n) converges to a bivariate normal distribution with mean vector and covariance matrix

$$\mu = \left(0, \frac{-\delta^2\sigma^2(L)}{2}\right) \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_\tau^2 & \delta\mu_\tau(\theta) \\ \delta\mu_\tau(\theta) & \delta^2\sigma_\tau^2(\mathcal{Q}_\delta) \end{pmatrix},$$

from which it follows that $\Theta_{n,\tau}$ is asymptotically $\mathcal{N}(\delta\mu_\tau(\theta), \sigma_\tau^2)$ under $(\mathcal{Q}_{\delta_n})_{n\geq 1}$.

V.A.2 Proof of Proposition 2

Let $(X_1^{(n)}, Y_1^{(n)}), \ldots, (X_n^{(n)}, Y_n^{(n)})$ be a random sample from $\mathcal{Q}_{\delta_n}$. Write $\mathcal{C}_n^{(n)} = \mathcal{C}_{n,\theta}^{(n)} - \mathcal{B}_n^{(n)}$, where $\mathcal{C}_{n,\theta}^{(n)} = \sqrt{n}(C_n^{(n)} - C_\theta)$ and $\mathcal{B}_n^{(n)} = \sqrt{n}(C_{\hat{\theta}_n^{(n)}} - C_\theta)$. Here, $\hat{\theta}_n^{(n)}$ is the estimator based on the sample from $\mathcal{Q}_{\delta_n}$ and

$$C_n^{(n)}(x,y) = H_n^{(n)}\left\{ \left(F_n^{(n)}\right)^{-1}(x), \left(G_n^{(n)}\right)^{-1}(y) \right\},$$

where

$$H_n^{(n)}(s,t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(X_i^{(n)} \leq s, Y_i^{(n)} \leq t\right),$$

$F_n^{(n)}(s) = H_n^{(n)}(s,1)$ and $G_n^{(n)}(t) = H_n^{(n)}(1,t)$. From van der Vaart and Wellner (1996), condition (V.8) implies that $\mathbb{H}_n^{(n)}(s,t) = \sqrt{n}(H_n^{(n)} - C_\theta) \rightsquigarrow \mathbb{H} + \delta(D - C_\theta)$. In particular,

$$\beta_{1,n}^{(n)}(x) = \sqrt{n}\{F_n^{(n)}(x) - x\} = \mathbb{H}_n^{(n)}(x,1) \rightsquigarrow \mathbb{H}(x,1)$$

and

$$\beta_{2,n}^{(n)}(y) = \sqrt{n}\{G_n^{(n)}(y) - y\} = \mathbb{H}_n^{(n)}(1,y) \rightsquigarrow \mathbb{H}(1,y)$$

since $D(x,1) - C_\theta(x,1) = D(1,y) - C_\theta(1,y) = 0$. From Chapter 3 in Shorack and Wellner (1986), both

$$\sup_{0\leq x\leq 1}\left|F_n^{(n)}(x) - x\right| = \sup_{0\leq x\leq 1}\left|\left(F_n^{(n)}\right)^{-1}(x) - x\right|$$

and

$$\sup_{0\leq y\leq 1}\left|G_n^{(n)}(y) - y\right| = \sup_{0\leq y\leq 1}\left|\left(G_n^{(n)}\right)^{-1}(y) - y\right|$$

converge in probability to zero, so that

$$\sqrt{n}\left\{\left(F_n^{(n)}\right)^{-1} - I\right\} \rightsquigarrow -\mathbb{H}(\cdot,1) \quad \text{and} \quad \sqrt{n}\left\{\left(G_n^{(n)}\right)^{-1} - I\right\} \rightsquigarrow -\mathbb{H}(1,\cdot).$$

(V.23)

Hence, since one can write

$$\begin{aligned}\mathcal{C}_{n,\theta}^{(n)}(x,y) &= I\!H_n^{(n)}\left\{\left(F_n^{(n)}\right)^{-1}(x),\left(G_n^{(n)}\right)^{-1}(y)\right\} \\ &\quad+\sqrt{n}\left\{C_\theta\left(\left(F_n^{(n)}\right)^{-1}(x),\left(G_n^{(n)}\right)^{-1}(y)\right)-C_\theta(x,y)\right\} \\ &= I\!H_n^{(n)}\left\{\left(F_n^{(n)}\right)^{-1}(x),\left(G_n^{(n)}\right)^{-1}(y)\right\}+C_{\theta,10}(x,y)\sqrt{n}\{(F_n^{(n)})^{-1}(x)-x\} \\ &\quad+C_{\theta,01}(x,y)\sqrt{n}\{(G_n^{(n)})^{-1}(y)-y\}+o_{\mathrm{P}}(1),\end{aligned}$$

one deduces that $\mathcal{C}_{n,\theta}^{(n)}$ converges weakly to $\mathcal{C}_\theta+\delta(D-C_\theta)$, where $\mathcal{C}_\theta = I\!H - C_{\theta,10}I\!H(\cdot,1) - C_{\theta,01}I\!H(1,\cdot)$ is the limit identified, e.g. by Gänssler and Stute (1987) and Tsukahara (2005) under the null hypothesis. The second part of Assumption $\mathcal{A}_2$ and the mean-value theorem enable to establish that $\mathcal{B}_n^{(n)}$ converges to $\tilde{\Theta}\dot{C}_\theta = \Theta\dot{C}_\theta+\mu(C_\theta,D)\dot{C}_\theta$, while the joint consistency of $(\mathcal{C}_{n,\theta}^{(n)},\mathcal{B}_n^{(n)})$ to $(\mathcal{C}_\theta+\delta(D-C_\theta),\Theta\dot{C}_\theta+\mu(C_\theta,D)\dot{C}_\theta)$ rises from Assumption $\mathcal{A}_1$.

V.B Computation of the drift terms

In the case of Clayton, Frank and Gumbel-Barnett copulas, the value of Spearman's rho cannot be expressed explicitly in terms of the dependence parameter, and hence the population value of formula (V.5) must be estimated through numerical methods. Such is also the case for

$$\rho'_{C_\theta}(\theta)=12\int_0^1\int_0^1\dot{C}_\theta(x,y)\mathrm{d}x\mathrm{d}y,\quad \mathrm{E}_D\left\{C_\theta(X,Y)\right\}=\int_0^1\int_0^1 C_\theta(x,y)d(x,y)\mathrm{d}x\mathrm{d}y,$$

$$\beta_{C_\theta}=\int_0^1\int_0^1\frac{\{\dot{c}_\theta(x,y)\}^2}{c_\theta(x,y)}\,\mathrm{d}x\mathrm{d}y\quad\text{and}\quad \mathrm{E}_D\left\{\ell'_{C_\theta}(X,Y)\right\}=\int_0^1\int_0^1\frac{\dot{c}_\theta(x,y)}{c_\theta(x,y)}\,d(x,y)\mathrm{d}x\mathrm{d}y,$$

where $c_\theta(x,y)=\partial^2C_\theta(x,y)/\partial x\partial y$, $\dot{c}_\theta(x,y)=\partial c_\theta(x,y)/\partial\theta$ and $\dot{C}_\theta(x,y)=\partial C_\theta(x,y)/\partial\theta$. Note that for Archimedean copulas, i.e. dependence models of the form $C_\theta(x,y)=\varphi_\theta^{-1}\{\varphi_\theta(x)+\varphi_\theta(y)\}$, one can show that

$$\dot{C}_\theta(x,y)=\frac{\dot{\varphi}_\theta(x)+\dot{\varphi}_\theta(y)-\dot{\varphi}_\theta\left\{C(x,y)\right\}}{\varphi'_\theta\left\{C_\theta(x,y)\right\}},$$

where $\dot{\varphi}_\theta(x)=\partial\varphi_\theta(x)/\partial\theta$ and $\varphi'_\theta(x)=\partial\varphi_\theta(x)/\partial x$. The Clayton, Frank and Gumbel-Barnett copulas are member of this important class of models.

V.B.1 The Clayton family

The copulas in this class and their associated densities are

$$C_\theta^{\mathrm{CL}}(x,y)=\left(x^{-\theta}+y^{-\theta}-1\right)^{-1/\theta}\text{ and }c_\theta^{\mathrm{CL}}(x,y)=(\theta+1)\,(xy)^{-\theta-1}\left(x^{-\theta}+y^{-\theta}-1\right)^{-1/\theta-2},$$

where $\theta>-1$. The associated value of Kendall's tau is $\tau_{C_\theta^{\mathrm{CL}}}(\theta)=\theta/(\theta+2)$, from which one deduces easily that $\mathrm{E}_{C_\theta}(C_\theta)=(\theta+1)/2$ and $\tau'_{C_\theta^{\mathrm{CL}}}(\theta)=2/(\theta+2)^2$. Further,

$$\dot{C}_\theta(x,y)=\frac{C_\theta(x,y)}{\theta}\left\{\frac{x^{-\theta}\log x+y^{-\theta}\log y}{x^{-\theta}+y^{-\theta}-1}-\log C_\theta(x,y)\right\}.$$

V.B.2 The Frank family

Frank's copula is given by

$$C_\theta^{\mathrm{F}}(x,y)=-\frac{1}{\theta}\ln\left\{1-\frac{\left(1-e^{-\theta x}\right)\left(1-e^{-\theta y}\right)}{1-e^{-\theta}}\right\},$$

(V.24)

where $\theta \in \mathbb{R} \setminus \{0\}$. As reported in Frees and Valdez (1998), Spearman's rho and Kendall's tau in this family are expressed by

$$\rho_{C_\theta^{\mathrm{F}}}(\theta) = 1 + \frac{12}{\theta^2}\int_0^\theta \frac{t(2t-\theta)}{e^t-1}\mathrm{d}t \quad \text{and} \quad \tau_{C_\theta^{\mathrm{F}}}(\theta) = 1 - \frac{4}{\theta} + \frac{4}{\theta^2}\int_0^\theta \frac{t}{e^t-1}\mathrm{d}t.$$

Hence, one deduces

$$\rho'_{C_\theta^{\mathrm{F}}}(\theta) = \frac{12}{\theta\,(e^\theta-1)} - \frac{24}{\theta^4}\int_0^\theta \frac{t(3t-\theta)}{e^t-1}\mathrm{d}t$$

and

$$\tau'_{C_\theta^{\mathrm{F}}}(\theta) = \frac{4}{\theta^2} + \frac{4}{\theta\,(e^\theta-1)} - \frac{8}{\theta^3}\int_0^\theta \frac{t}{e^t-1}\mathrm{d}t.$$

The other necessary computations, however, must be accomplished numerically.

V.B.3 The Gumbel-Barnett family

The analytical form of this extreme-value copula (Ghoudi and Rémillard, 1998) is

$$C_\theta^{\mathrm{GB}}(x,y) = \exp\left\{-\left(|\log x|^{1/(1-\theta)} + |\log y|^{1/(1-\theta)}\right)^{1-\theta}\right\},$$

where $0 \le \theta \le 1$. Computations of the drift terms in this class of models are difficult and must be done numerically. The only explicit expressions are for Kendall's tau and its derivative, namely $\tau_{C_\theta^{\mathrm{GB}}}(\theta) = \theta$ and $\tau'_{C_\theta^{\mathrm{GB}}}(\theta) = 1$.

V.B.4 The Normal family

The Normal copula, which arises as the dependence function associated to the classical normal model, is defined by

$$C_\theta^{\mathrm{N}}(x,y) = \int_{-\infty}^{\Phi^{-1}(x)}\int_{-\infty}^{\Phi^{-1}(y)} h_\theta(s,t)\mathrm{d}s\mathrm{d}t,$$

where

$$h_\theta(s,t) = \frac{\left(1-\theta^2\right)^{-1/2}}{2\pi}\exp\left\{-\frac{1}{2(1-\theta^2)}\left(s^2+t^2-2\theta st\right)\right\}$$

is the standard bivariate normal density with correlation coefficient θ. Despite the implicit form of C_θ^{N} involving the percentile function of a standard univariate normal distribution, there exists explicit relationships between the dependence parameter θ and Kendall and Spearman measures of association. Explicitly,

$$\tau_{C_\theta}(\theta) = \frac{2}{\pi}\sin^{-1}(\theta) \quad \text{and} \quad \rho_{C_\theta}(\theta) = \frac{6}{\pi}\sin^{-1}\left(\frac{\theta}{2}\right),$$

from which it follows easily that

$$\mathrm{E}_{C_\theta}(C_\theta) = \frac{2\sin^{-1}(\theta)+\pi}{4\pi}, \quad \tau'_{C_\theta}(\theta) = \frac{2}{\pi\sqrt{1-\theta^2}} \quad \text{and} \quad \rho'_{C_\theta}(\theta) = \frac{6}{\pi\sqrt{4-\theta^2}}.$$

Hence, if $D \equiv C_{\theta_D}^{\mathrm{N}}$, i.e. if one considers a mixture of Normal copulas, then

$$\mu_\rho(C_\theta, D) = \frac{\sin^{-1}(\theta_D/2) - \sin^{-1}(\theta/2)}{\sqrt{4-\theta^2}}.$$

Also, the density associated to C_θ^{N} is

$$c_\theta^{\mathrm{N}}(x,y) = h_\theta\left\{\Phi^{-1}(x), \Phi^{-1}(y)\right\}\left(\Phi^{-1}\right)'(x)\left(\Phi^{-1}\right)'(y),$$

(V.25)

and it is possible to establish that

$$\ell'_{C^{\mathrm{N}}_\theta}(x,y) = \frac{\dot{c}^{\mathrm{N}}_\theta(x,y)}{c^{\mathrm{N}}_\theta(x,y)} = \left.\frac{\theta(1-\theta^2)-\theta(s^2+t^2)+(\theta^2+1)st}{(1-\theta^2)^2}\right|_{s=\Phi^{-1}(x),\, t=\Phi^{-1}(y)}.$$

This enables to compute

$$\begin{aligned}
\mathrm{E}_D\left\{\ell'_{C^{\mathrm{N}}_\theta}(X,Y)\right\} &= \int_0^1\int_0^1 \ell'_{C^{\mathrm{N}}_\theta}(x,y)c^{\mathrm{N}}_{\theta_D}(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{\frac{\theta(1-\theta^2)-\theta(s^2+t^2)+(\theta^2+1)st}{(1-\theta^2)^2}\right\}h_{\theta_D}(s,t)\mathrm{d}s\mathrm{d}t \\
&= \mathrm{E}_{\theta_D}\left\{\frac{\theta(1-\theta^2)-\theta(S^2+T^2)+(\theta^2+1)ST}{(1-\theta^2)^2}\right\},
\end{aligned}$$

where (S,T) follows a bivariate normal distribution with means 0, variances 1 and correlation coefficient θ_D. Here, E_{θ_D} denotes expectation with respect to h_{θ_D}. Thus, noting that $\mathrm{E}_{\theta_D}(S^2) = \mathrm{E}_{\theta_D}(T^2) = 1$ and $\mathrm{E}_{\theta_D}(ST) = \theta_D$, straightforward computations yield

$$\mathrm{E}_D\left\{\ell'_{C^{\mathrm{N}}_\theta}(X,Y)\right\} = \frac{(\theta^2+1)(\theta_D-\theta)}{(1-\theta^2)^2}.$$

Long but similar computations enable to obtain $\beta_{C_\theta} = \theta^2+1$ and hence

$$\mu_{PL}(C_\theta, D) = \frac{\theta_D-\theta}{(1-\theta^2)^2}.$$

VI

Models for construction of multivariate dependence: a comparison study

Daniel Berg
University of Oslo & Norwegian Computing Center

Kjersti Aas
Norwegian Computing Center

Abstract

We review models for construction of higher-dimensional dependence that have arisen recent years. A multivariate data set, which exhibit complex patterns of dependence, particularly in the tails, can be modelled using a cascade of lower-dimensional copulae. We examine two such models that differ in their construction of the dependency structure, namely the nested Archimedean constructions and the pair-copula constructions (also referred to as vines). The constructions are compared, and estimation- and simulation techniques are examined. The fit of the two constructions is tested on two different four-dimensional data sets; precipitation values and equity returns, using state of the art copula goodness-of-fit procedures. The nested Archimedean construction is strongly rejected for both our data sets, while the pair-copula construction provides a much better fit. Through VaR calculations, we show that the latter does not overfit data, but works very well even out-of-sample.

Key words

Nested Archimedean copulas, Pair-copula decompositions, Equity returns, Precipitation values, Goodness-of-fit, Out-of-sample validation.

Publication details

Submitted for publication - October 2007.

VI.1 Introduction

A copula is a multivariate distribution function with standard uniform marginal distributions. While the literature on copulae is substantial, most of the research is still limited to the bivariate case. Building higher-dimensional copulae is a natural next step, however, this is not an easy task. Apart from the multivariate Gaussian and Student copulae, the set of higher-dimensional copulae proposed in the literature is rather limited.

The Archimedean copula family (see e.g. Joe (1997) for a review) is a class that has attracted particular interest due to numerous properties which make them simple to analyse. The most common multivariate extension, the exchangeable multivariate Archimedean copula (EAC), is extremely restrictive, allowing the specification of only one generator, regardless of dimension. There have been some attempts at constructing more flexible multivariate Archimedean copula extensions, see e.g. Joe (1997); Nelsen (1999); Embrechts et al. (2003); Whelan (2004); Morillas (2005); Savu and Trede (2006). In this paper we discuss one group of such extensions, the nested Archimedean constructions (NACs). For the d-dimensional case, all NACs allow for the modelling of up to $d-1$ bivariate Archimedean copulae.

For a d-dimensional problem there are in general $d(d-1)/2$ parings of variables. Hence, while the NACs constitute a huge improvement compared to the EAC, they are still not rich enough to model all possible mutual dependencies amongst the d variates. An even more flexible structure, here denoted the pair-copula construction (PCC) allows for the free specification of $d(d-1)/2$ copulae. This structure was originally proposed by Joe (1996), and later discussed in detail by Bedford and Cooke (2001, 2002), Kurowicka and Cooke (2006) (simulation) and Aas et al. (2007) (inference). Similar to the NACs, the PCC is hierarchical in nature. The modelling scheme is based on a decomposition of a multivariate density into a cascade of bivariate copulae. In contrast to the NACs, the PCC is not restricted to Archimedean copulae. The bivariate copulae may be from any family and several families may well be mixed in one PCC.

This paper has several contributions. In Section VI.2 we compare the two ways of constructing higher dimensional dependency structures, the NACs and the PCCs. We examine properties and estimation- and simulation techniques, focusing on the relative strengths and weaknesses of the different constructions. In Section VI.3 we apply the NAC and the PCC to two four-dimensional data sets; precipitation values and equity returns. We examine the goodness-of-fit and validate the PCC out-of-sample with respect to one day value at risk (VaR) for the equity portfolio. Finally, Section VI.4 provides some summarizing comments and conclusions.

VI.2 Constructions of higher dimensional dependence

VI.2.1 The nested Archimedean constructions (NACs)

The most common multivariate Archimedean copula, the exchangeable Archimedean copula (EAC), is extremely restrictive, allowing the specification of only one generator, regardless of dimension. Hence, all k-dimensional marginal distributions ($k < d$) are identical. For several applications, one would like to have multivariate copulae which allows for more flexibility. In this section, we review one group of such extensions, the nested Archimedean constructions (NACs). We first review two simple special cases, the FNAC and the PNAC, in Sections VI.2.1 and VI.2.1, and then we turn to the general case in Section VI.2.1. However, before reviewing NACs, we give a short description of the EAC in Section VI.2.1, since this structure serves as a baseline.

The exchangeable multivariate Archimedean copula (EAC)

The most common way of defining a multivariate Archimedean copula is the EAC, defined as

$$C(u_1, u_2, \ldots, u_d) = \varphi^{-1}\left\{\varphi(u_1) + \ldots + \varphi(u_d)\right\}, \tag{VI.1}$$

where the function φ is a decreasing function known as the generator of the copula and φ^{-1} denotes its inverse (see e.g. Nelsen (1999)). Here, we assume that the generator has only one parameter,

θ. There are however cases in which the generator have more parameters, see e.g. Joe (1997). For $C(u_1, u_2, \ldots, u_d)$ to be a valid d-dimensional Archimedean copula, φ^{-1} should have an analytical property known as d-monotonicity. Se McNeil and Neslehova (2007) for details. One usually also assumes that $\varphi(0) = \infty$, i.e. that the Archimedean copula is strict. Consider for example the popular Gumbel (Gumbel, 1960) and Clayton (Clayton, 1978) copulae. The generator functions for these two copulae are given by $\varphi(t) = (-\log(t))^\theta$ and $\varphi(t) = (t^{-\theta} - 1)/\theta$, respectively.

The fully nested Archimedean construction (FNAC)

A simple generalization of (VI.1) can be found in Joe (1997) and is also discussed in Embrechts et al. (2003), Whelan (2004), Savu and Trede (2006) and McNeil (2007). The structure, which is shown in Figure VI.1 for the four-dimensional case, is quite simple, but notationally cumbersome. As seen from the figure, one simply adds a dimension step by step. The nodes u_1 and u_2 are coupled through copula C_{11}, node u_3 is coupled with $C_{11}(u_1, u_2)$ through copula C_{21}, and finally node u_4 is coupled with $C_{21}(u_3, C_{11}(u_1, u_2))$ through copula C_{31}. Hence, the copula for the 4-dimensional case requires three bivariate copulae C_{11}, C_{21}, and C_{31}, with corresponding generators $\varphi_{11}, \varphi_{21}$, and φ_{31}:

$$\begin{aligned} C(u_1, u_2, u_3, u_4) &= C_{31}(u_4, C_{21}(u_3, C_{11}(u_1, u_2))) \\ &= \varphi_{31}^{-1}\left\{\varphi_{31}(u_4) + \varphi_{31}(\varphi_{21}^{-1}\{\varphi_{21}(u_3) + \varphi_{21}(\varphi_{11}^{-1}\{\varphi_{11}(u_1) + \varphi_{11}(u_2)\})\})\right\}. \end{aligned}$$

For the d-dimensional case, the corresponding expression becomes

$$\begin{aligned} C(u_1, \ldots, u_d) \quad = \quad & \varphi_{d-1,1}^{-1}\{\varphi_{d-1,1}(u_d) + \varphi_{d-1,1} \circ \varphi_{d-2,1}^{-1}\{\varphi_{d-2,1}(u_{d-1}) + \varphi_{d-2,1} \\ \circ \ldots \circ \quad & \varphi_{11}^{-1}\{\varphi_{11}(u_1) + \varphi_{11}(u_2)\}\}\}. \end{aligned} \tag{VI.2}$$

In this structure, which Whelan (2004) refers to as fully nested, all bivariate margins are themselves Archimedean copulae. It allows for the free specification of $d-1$ copulae and corresponding distributional parameters, while the remaining $(d-1)(d-2)/2$ copulae and parameters are implicitly given through the construction. More specifically, in Figure VI.1, the two pairs (u_1, u_3) and (u_2, u_3) both have copula C_{21} with dependence parameter θ_{21}. Moreover, the three pairs (u_1, u_4), (u_2, u_4) and (u_3, u_4) all have copula C_{31} with dependence parameter θ_{31}. Hence, when adding variable k to the structure, we specify the relationships between k pairs of variables.

The FNAC is a construction of partial exchangeability and there are some technical conditions that need to be satisfied for (VI.2) to be a proper d-dimensional copula. The consequence of these conditions for the FNAC is that the degree of dependence, as expressed by the copula parameter, must decrease with the level of nesting, i.e. $\theta_{11} \geq \theta_{21} \geq \ldots \geq \theta_{d-1,1}$, in order for the resulting d-dimensional distribution to be a proper copula.

The partially nested Archimedean construction (PNAC)

An alternative multivariate extension is the PNAC. This structure was originally proposed by Joe (1997) and is also discussed in Whelan (2004), McNeil et al. (2006) (where it is denoted partially exchangeable) and McNeil (2007).

The lowest dimension for which there is a distinct structure of this class is four, when we have the following copula:

$$\begin{aligned} C(u_1, u_2, u_3, u_4) &= C_{21}(C_{11}(u_1, u_2), C_{21}(u_3, u_4)) \\ &= \varphi_{21}^{-1}\{\varphi_{21}(\varphi_{11}^{-1}\{\varphi_{11}(u_1) + \varphi_{11}(u_2)\}) + \varphi_{21}(\varphi_{12}^{-1}\{\varphi_{12}(u_3) + \varphi_{12}(u_4)\})\}. \end{aligned}$$

Figure VI.2 illustrates this structure graphically. Again the construction is notationally cumbersome although the logic is straightforward. We first couple the two pairs (u_1, u_2) and (u_3, u_4) with copulae C_{11} and C_{12}, having generator functions φ_{11} and φ_{12}, respectively. We then couple these two copulae using a third copula C_{21}. The resulting copula is exchangeable between u_1 and u_2 and also between u_3 and u_4. Hence, it can be understood as a composite of the EAC and the FNAC.

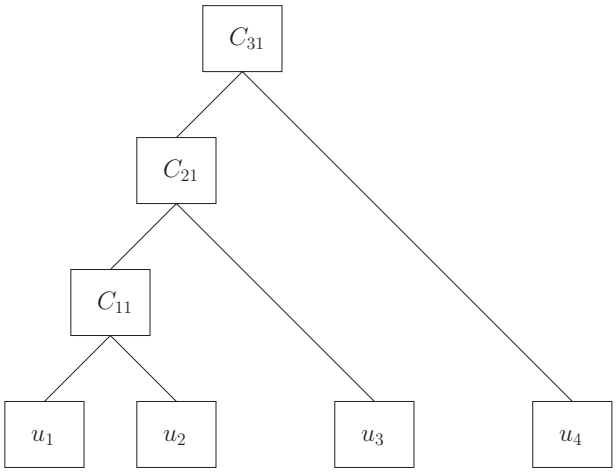


Figure VI.1: Fully nested Archimedean construction.

For the PNAC, as for the FNAC, $d-1$ copulae and corresponding distributional parameters are freely specified, while the remaining copulae and parameters are implicitly given through the construction. More specifically, in Figure VI.2, the four pairs (u_1, u_3), (u_1, u_4) (u_2, u_3) and (u_2, u_4) will all have copula C_{21}, with dependence parameter θ_{21}. Similar constraints on the parameters are required for the PNACs as for the FNACs.

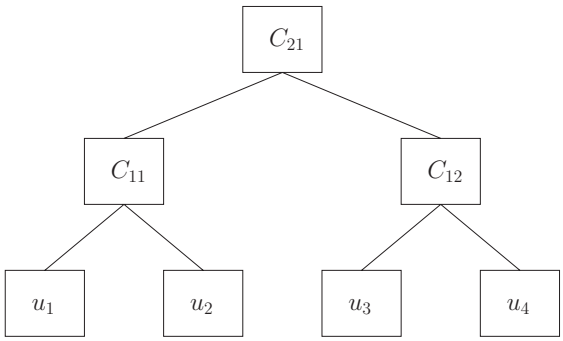


Figure VI.2: Partially nested Archimedean construction.

The general case

The generally nested Arcimedean construction was originally treated by Joe (1997, Chapter 4.2), and is also mentioned in Whelan (2004). However, Savu and Trede (2006) were the first to provide the notation for arbitrary nesting, and to show how to calculate the d-dimensional density in general.

Savu and Trede (2006) use the notation hierarchical Archimedean copula for the generally nested case. The idea is to build a hierarchy of Archimedean copulas. Assume that there are L

levels. At each level l, there are n_l distinct objects (an object is either a copula or a variable). At level $l = 1$ the variables $u_1, \ldots, u_d$ are grouped into n_1 exchangeable multivariate Archimedean copulae. These copulae are in turn coupled into n_2 copulae at level $l = 2$, and so on.

Figure VI.3 shows an example. The 9-dimensional copula in the figure is given by

$$C(u_1, \ldots u_9) = C_{41}(C_{31}(C_{21}(C_{11}(u_1, u_2), u_3, u_4), u_5, u_6), C_{22}(u_7, C_{12}(u_8, u_9))).$$

At level one, there are two copulae. Both are two-dimensional EACs. The first, C_{11} joins the variables u_1 and u_2, while the other, C_{12}, joins u_8 and u_9. At the second level, there are also two copulae. The first, C_{21}, joins the copula C_{11} with the two variables u_3 and u_4, while the other, C_{22} joins C_{12} and u_7. At the third level there is only one copula, C_{31}, joining C_{21}, u_5 and u_6. Finally, at level four, the copula C_{41} joins the two copulae C_{31} and C_{22}.

There are a number of conditions to ensure that the resulting structure is a valid Archimedean copula (Savu and Trede, 2006). The number of copulas must decrease at each level, the top level may only contain one copula and all the inverse generator functions must be completely monotone. Further, we must have that the degree of dependence must decrease with the level of nesting. For example in Figure VI.3 we must have that $\theta_{11} \geq \theta_{21} \geq \theta_{31} \geq \theta_{41}$ and $\theta_{12} \geq \theta_{22} \geq \theta_{41}$. If mixing copula generators belonging to different Archimedean families, even this requirement might not be sufficient. Two Archimedean copulas from families 1 and 2 can only be nested if the derivative of the product $\varphi_1 \circ \varphi_2^{-1}$ is completely monotonic (McNeil, 2007). The issue of which copula families that can be mixed has been considered in some detail in Joe (1997), but it is still not fully explored. Hence, here we only work with structures for which all the generators are from the same family.

Unfortunately, it is not possible to obtain a simple expression for the density of a hierarchical Archimedean copula. Due to the complex structure of this construction, one has to use a recursive approach. One differentiates the d-dimensional top level copula with respect to its arguments using the chain rule. See Savu and Trede (2006) for more details.

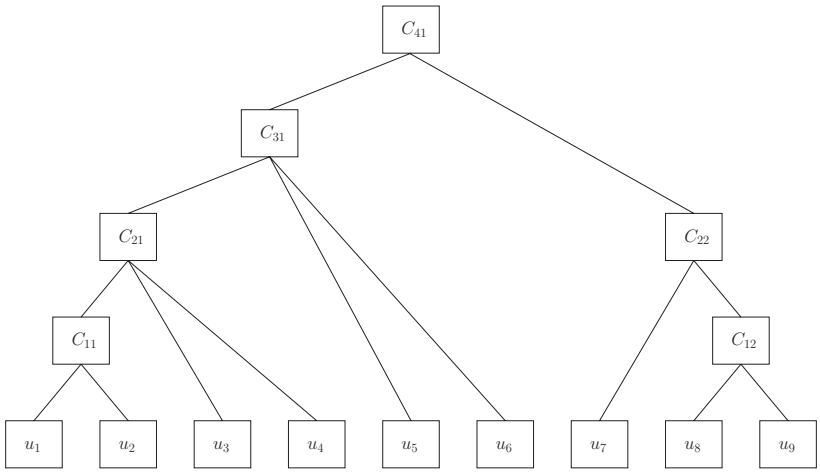


Figure VI.3: Hierarchically nested Archimedean construction.

Parameter estimation

For the NACs, as for the EAC, the parameters may be estimated by maximum likelihood. However, not even for the EAC it is straightforward to derive the density in general for all parametric

families. For instance, for the Gumbel family, one has to resort to a computer algebra system, such as *Mathematica*, or the function D in R, to derive the d-dimensional density.

Savu and Trede (2006) give the density expression for a general NAC. The density is obtained using a recursive approach. Hence, the number of computational steps needed to evaluate the density increases rapidly with the complexity of the copula, and parameter estimation becomes very time consuming in high dimensions.

Simulation

Simulating from the higher-dimensional constructions is a very important and central practical task. Simulating from an EAC is usually rather simple, and several algorithms exist. A popular algorithm utilizes the representation of the Archimedean copula generator using Laplace transforms (see e.g. Frees and Valdez (1998)). McNeil (2007) shows how to use the Laplace-transform method also for the NACs, in the case where all generators are taken from either the Gumbel- or the Clayton family. A problem with the Laplace transform method is that it is limited to copulae for which we can find a distribution that equals the Laplace transform of the inverse generator function, and from which we can easily sample. For some copulae, e.g. Frank, there is, as of now, no alternative to the conditional inversion method described in e.g. Embrechts et al. (2003). This procedure involves the $d-1$ first derivatives of the copula function and, in most cases, numerical inversion. The higher-order derivatives are usually extremely complex expressions (see e.g. Savu and Trede (2006)). Hence, simulation may become very inefficient for high dimensions.

VI.2.2 The pair-copula constructions (PCC)

While the NACs constitute a large improvement compared to the EAC, they still only allow for the modelling of up to $d-1$ copulae. An even more flexible structure, the PCC, allows for the free specification of $d(d-1)/2$ copulae. This structure was orginally proposed by Joe (1996), and it has later been discussed in detail by Bedford and Cooke (2001, 2002), Kurowicka and Cooke (2006) (simulation) and Aas et al. (2007) (inference). Similar to the NAC's, the PCC's are hierarchical in nature. The modelling scheme is based on a decomposition of a multivariate density into $d(d-1)/2$ bivariate copula densities, of which the first $d-1$ are unconditional, and the rest are conditional.

While the NACs are defined through their distribution functions, the PCCs are usually represented in terms of the density. Two main types of PCCs have been proposed in the literature; canonical vines and D-vines (Kurowicka and Cooke, 2004). Here, we concentrate on the D-vine representation, for which the density is (Aas et al., 2007):

$$f(x_1,\dots x_d) = \prod_{k=1}^{d} f(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,i}\left\{F(x_i|x_{i+1},\dots,x_{i+j-1}), F(x_{i+j}|x_{i+1},\dots,x_{i+j-1})\right\}. \quad \text{(VI.3)}$$

The conditional distribution functions are computed using (Joe, 1996)

$$F(x|\boldsymbol{v}) = \frac{\partial\, C_{x,v_j|\boldsymbol{v}_{-j}}\{F(x|\boldsymbol{v}_{-j}).F(v_j|\boldsymbol{v}_{-j})\}}{\partial F(v_j|\boldsymbol{v}_{-j})}, \quad \text{(VI.4)}$$

where $C_{ij|\boldsymbol{k}}$ is a bivariate copula distribution function. To use this construction to represent a dependency structure through copulas, we assume that the univariate margins are uniform in [0,1]. One 4-dimensional case of (VI.3) is

$$\begin{aligned} c(u_1,u_2,u_3,u_4) &= c_{11}(u_1,u_2)\cdot c_{12}(u_2,u_3)\cdot c_{13}(u_3,u_4) \\ &\cdot\ c_{21}(F(u_1|u_2),F(u_3|u_2))\cdot c_{22}(F(u_2|u_3),F(u_4|u_3)) \\ &\cdot\ c_{31}(F(u_1|u_2,u_3),F(u_4|u_2,u_3)), \end{aligned}$$

where $F(u_1|u_2) = \partial\, C_{11}(u_1,u_2)/\partial u_2$, $F(u_3|u_2) = \partial\, C_{12}(u_2,u_3)/\partial u_2$, $F(u_2|u_3) = \partial\, C_{12}(u_2,u_3)/\partial u_3$, $F(u_4|u_3) = \partial\, C_{13}(u_3,u_4)/\partial u_3$, $F(u_1|u_2,u_3) = \partial\, C_{21}(F(u_1|u_2),F(u_3|u_2))/\partial F(u_3|u_2)$ and $F(u_4|u_2,u_3) = \partial\, C_{22}(F(u_4|u_3),F(u_2|u_3))/\partial F(u_2|u_3)$. Figure VI.2.2 illustrates this structure.

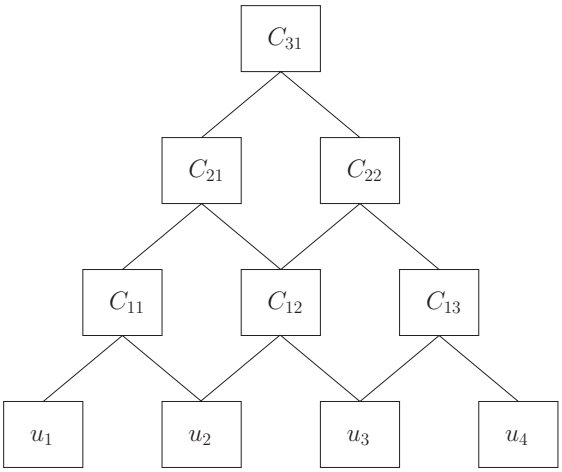


Figure VI.4: Pair-copula construction.

The copulae involved in (VI.3) do not have to belong to the same family. In contrast to the NACs they do not even have to belong to the same class. The resulting multivariate distribution will be valid even if we choose, for each pair of variables, the parametric copula that best fits the data. As seen from (VI.3) the PCC consists of $d(d-1)/2$ bivariate copulae of known parametric families, of which $d-1$ are copulae of pairs of the original variables, while the remaining $(d-1)(d-2)/2$ are copulae of pairs of variables constructed using (VI.4) recursively. This means that in contrast to the NACs, the unspecified bivariate margins will belong to a known parametric family in general. However, it can be shown, that e.g. upper (lower) tail dependence on the bivariate copulae at the lowest level is a sufficient condition for all bivariate margins to have upper (lower) tail dependence[1].

Parameter estimation

The parameters of the PCC may be estimated by maximum likelihood. In contrast to the NACs, the density is explicitly given. However, also for this construction, a recursive approach is used (see Aas et al. (2007, Algorithm 4)). Hence, the number of computational steps to evaluate the density increases rapidly with the complexity of the copula, and parameter estimation becomes time consuming in high dimensions.

Simulation

The simulation algorithm for a D-vine is straightforward and simple to implement, see Aas et al. (2007, Algorithm 2). Like for the NACs, the conditional inversion method is used. However, to determine each of the conditional distribution functions involved, only the first partial derivative of a bivariate copula needs to be computed (see Aas et al. (2007)). Hence, the simulation procedure for the PCC is in general much simpler and faster than for the NACs.

VI.2.3 Comparison

In this section we summarize the differences between the NACs and the PCCs with respect to ease of interpretation, applicability and computational complexity.

[1] Personal communication with Harry Joe.

Table VI.1: Summary of construction properties for the EAC, NAC and PCC constructions.

Constr.	*Max no. of copulae freely specified*	*Parameter constraints*	*Copula mixing*
EAC	1	None	Only one copula
NAC	$d-1$	Dependence must decrease with level of nesting	May combine different Arch. families but under complete monotonicity restrictions
PCC	$d(d-1)/2$	None	May combine any copula families from any class

First, the main advantage of the PCCs is the increased flexibility compared to the NACs. While the NACs only allow for the free specification of $d-1$ copulae, $d(d-1)/2$ copulae may be specified in a PCC. Next, for the NACs there are restrictions on which Archimedean copulae that can be mixed, while the PCCs can be built using copulae from different families and classes. Finally, the NACs have another even more important restriction in that the degree of dependence must decrease with the level of nesting. When looking for appropriate data sets for the applications in Section VI.3, it turned out to be quite difficult to find real-world data sets satisfying this restriction. Hence, this feature of the NACs might prevent them from being extensively used in real-world applications. For the PCCs, on the other hand, one is always guaranteed that all parameter combinations are valid. Table VI.1 summarizes these properties.

It is our opinion that another advantage of the PCCs is that they are represented in terms of the density and hence easier to handle than the NACs that are defined through their distribution functions. The PCCs are also in general more computationally efficient than the NACs. Table VI.2 shows computational times (s) in *R* [2] for likelihood evaluation, parameter estimation and simulation for different structures. The parameter estimation is done for the data set described in Section VI.3.3, and the simulation is performed using the parameters in Table VI.6 (based on 1000 samples). The values for NAC were computed using density expressions found in Savu and Trede (2006). However, general expressions may also easily be obtained symbolically using e.g. the function `D` in *R*. The estimation times in Table VI.2 are only indicative and included as examples since they are very dependent on size and structure of the data set. It is more appropriate to study the times needed to compute one evaluation of the likelihood given in the leftmost column. As can be seen from the table, the PCC is superior to the NAC for likelihood evaluation in both the Gumbel and the Frank case. Moreover, it is much faster for simulation in the Frank case, since one in this case must use the general conditional inversion algorithm with numerical inversion for the NAC. In the Gumbel case, however, one can perform much more efficient simulation from the NAC using the algorithms given in McNeil (2007). Hence, in this case, the NAC is superior to the PCC.

The multivariate distribution defined through a NAC will always by definition be an Archimedean copula (assuming that all requirements are satisfied), and all bivariate margins will belong to a known parametric family. This is not the case for the PCCs, for which neither the multivariate distribution nor the unspecified bivariate margins will belong to a known parametric family in general. However, we do not view this as a problem, since both might easily be obtained by simulation.

Finally, it should be noted that for both structures, an important part of the full estimation problem is how to select the ordering of the variables. For smaller dimensions (say 3 and 4), one may estimate the parameters of all possible orderings and compare the resulting log-likelihoods. This is in practice infeasible for higher dimensions, since the number of different orderings increases very rapidly with the dimension of the data set. One may instead determine which bivariate

[2] The experiments were run on a Intel(R) Pentium(R) 4 CPU 2.80GHz PC.

Table VI.2: Computational times in sec. for different constructions and copulae, fitted to the equity data in Section VI.3.3.

Method	*Likelihood evaluation*	*Estimation*	*Simulation*
	Gumbel		
NAC	0.32	34.39	0.02
PCC	0.04	5.09	7.56
	Frank		
NAC	0.12	5.34	64.83
PCC	0.02	1.22	5.82

relationships that are most important to model correctly and let this determine which ordering to choose. Very recently, there has been some attempts of formalising this procedure, both for the NACs (Okhrin et al., 2007) and for the PCCs (Min and Czado, 2007).

VI.3 Applications

The fit of the NAC and the PCC is assessed for two different four-dimensional data sets; precipitation values and equity returns. Appropriate modelling of precipitation is of great importance to insurance companies which are exposed to growth in damages to buildings caused by external water exposition. Modelling precipitation and valuing related derivative contracts is also indeed a frontier in the field of weather derivatives, see e.g. Musshoff et al. (2006). The dependencies within an equity portfolio can have enormous impacts on e.g. capital allocation and the pricing of collateralized debt obligations. Before these two applications are further treated, we describe the tests used for goodness-of-fit in our study.

VI.3.1 Goodness-of-fit

To evaluate whether a copula or copula construction appropriately fit the data at hand, goodness-of-fit testing is called upon. Lately, several procedures have been proposed, see e.g. Berg (2007b) for an overview and power comparison. These power comparisons show that no procedure is always the best. However, the procedure that showed to have the overall best performance in the study referred to above, was one based on the empirical copula C_n introduced by Deheuvels (1979),

$$C_n(\mathbf{u}) = \frac{1}{n+1}\sum_{j=1}^{n} \mathbf{1}(U_{j1} \leq u_1, \ldots, U_{jd} \leq u_d), \quad \mathbf{u} = (u_1, \ldots, u_d) \in (0,1)^d,$$

where $\mathbf{U}_j = (U_{j1}, \ldots, U_{jd})$ are the $U(0,1)^d$ pseudo-observations, defined as normalized ranks. This procedure is based on the process $\mathcal{C}_n = \sqrt{n}\{C_n - C_{\hat{\theta}_n}\}$ where $\hat{\theta}_n$ is some consistent estimator of θ. Basing a goodness-of-fit procedure on $\mathcal{C}_n$ was originally proposed by Fermanian (2005), but there dismissed due to poor statistical properties. However, it has later been shown that it has the necessary asymptotic properties to be a justified goodness-of-fit procedure (Quessy, 2005; Genest and Rémillard, 2008). Moreover, Genest et al. (2008) and Berg (2007b) have examined the power of $\mathcal{C}_n$ and concluded that it is a very powerful procedure in most cases.

We use the Cramér-von Mises statistic, defined by:

$$S_n = n\int_{[0,1]^d} \{C_n(\mathbf{u}) - C_{\theta_n}(\mathbf{u})\}^2 \, \mathrm{d}C_n(\mathbf{u}) = \sum_{j=1}^{n} \{C_n(\mathbf{U}_j) - C_{\theta_n}(\mathbf{U}_j)\}^2 .$$

Large values of S_n means a poor fit and leads to the rejection of the null hypothesis copula. In practice, the limiting distribution of S_n depends on θ. Hence, approximate p-values for the test

must be obtained through a parametric bootstrap procedure. We adopt the procedure in Appendix A in Genest et al. (2008), setting the bootstrap parameters m and N to 5000 and 1000, respectively. The validity of this bootstrap procedure was established in Genest and Rémillard (2008).

It will be shown in Section VI.3.2 that for the precipitation data set, S_n leads to the rejection of all the different NAC and PCC structures that are investigated. Hence, to be able to compare the two structures for this data set, we also use another goodness-of-fit procedure based on the process $\mathcal{K}_n = \sqrt{n}\{K_n - K_{\hat{\theta}_n}\}$, where

$$K_n(t) = \frac{1}{n+1}\sum_{j=1}^{n} \mathbf{1}(C_n(\mathbf{U}_j) \leq t),$$

is the empirical distribution function of $C_n(\mathbf{u})$. See Genest et al. (2006a) for details. Also for this procedure we use the Cramér-von Mises statistic, i.e.:

$$T_n = \int_{[0,1]^d} \left\{K_n(\mathbf{u}) - K_{\hat{\theta}}(\mathbf{u})\right\}^2 \mathrm{d}K_n(\mathbf{u}) = \sum_{j=1}^{n} \left\{K_n(\mathbf{U}_j) - K_{\hat{\theta}_n}(\mathbf{U}_j)\right\}^2,$$

and parametric bootstrap to obtain the p-values.

For both procedures, we use a 5% significance level for all experiments in this section.

VI.3.2 Application 1: Precipitation data

In this section we study daily precipitation data (mm) for the period 01.01.1990 to 31.12.2006 for 4 meteorological stations in Norway; Vestby, Ski, Nannestad and Hurdal, obtained from the Norwegian Meteorological Institute. According to Musshoff et al. (2006), the stochastic process of daily precipitation can be decomposed into a stochastic process of "rainfall"/"no rainfall", and a distribution for the amount of precipitation given that it rains. Here, we are only interested in the latter. Hence, before further processing, we remove days with non-zero precipitation values for at least one station, resulting in 2065 observations for each variable. Figures VI.5-VI.6 show the daily precipitation values and corresponding copulae for pairs of meteorological stations. Since we are mainly interested in estimating the dependence structure of the stations, the precipitation vectors are converted to uniform pseudo-observations before further modelling. In light of recent results due to Chen and Fan (2006), the method of maximum pseudo-likelihood is consistent even when time series models are fitted to the margins.

Based on visual inspection and preliminary goodness-of-fit tests for bivariate pairs (the copulae taken into consideration were the Student, Clayton, survival Clayton, Gumbel and Frank copulae), we decided to examine Gumbel and Frank NACs and Gumbel, Frank and Student PCCs for the precipitation data.

Hierarchically nested Archimedean construction

The most appropriate ordering of the variates in the decomposition is found by comparing Kendall's tau values for all bivariate pairs. These are shown in Table VI.3. They confirm the intuition that the degree of dependence between the variables corresponds to the distances between the stations. Ski and Vestby are closely located, and so is Hurdal and Nannestad, while the distance from Ski/Vestby to Hurdal/Nannestad is larger. Hence, we choose C_{11} and C_{12} to be the copulae of Vestby and Ski and Nannestad and Hurdal, respectively, while C_{21} is the copula of the remaining pairs.

The two leftmost columns of Table VI.4 show the estimated parameter values, resulting log-likelihoods, and estimated p-values for the Gumbel and Frank NACs, fitted to the precipitation data. We see that both goodness-of-fit procedures strongly reject the two NAC constructions. Hence we conclude that the NACs considered are not flexible enough to fit the precipitation data appropriately.

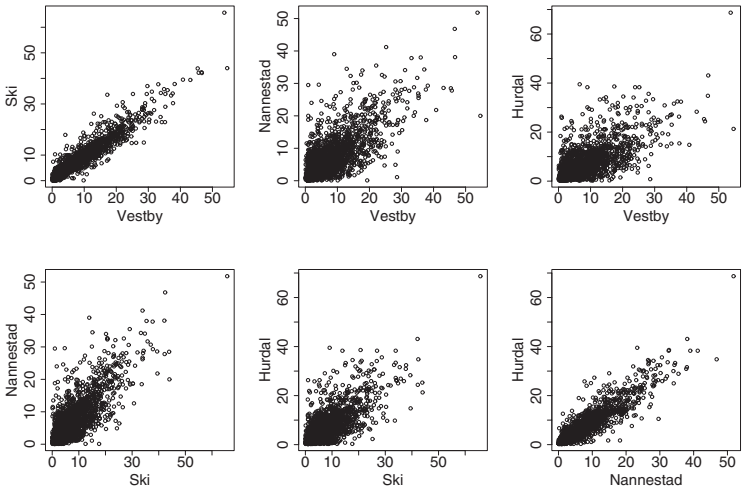


Figure VI.5: Daily precipitation (mm) for pairs of meteorological stations for the period 01.01.1990 to 31.12.2006, zeros removed.

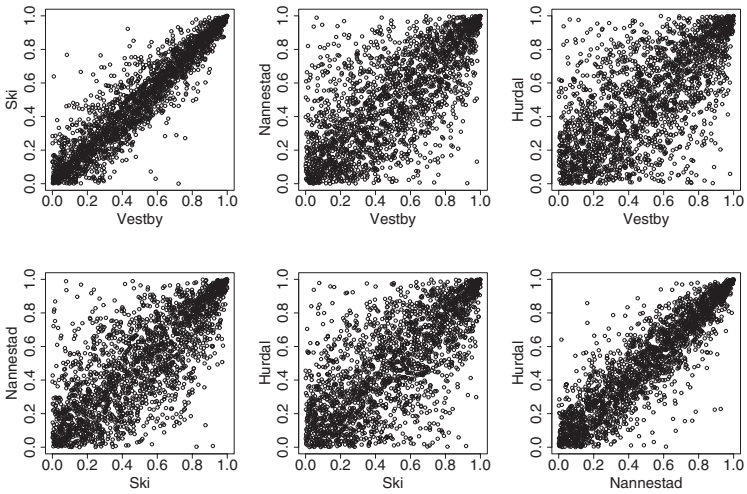


Figure VI.6: Pseudo-observations corresponding to Figure VI.5.

Table VI.3: Estimated Kendall's tau for pairs of variables.

Location	Ski	Nannestad	Hurdal
Vestby	0.79	0.49	0.47
Ski		0.56	0.53
Nannestad			0.71

Table VI.4: Estimated parameters, log-likelihood and estimated p-values for NACs and PCCs fitted to the precipitation data.

Parameter	*NAC*		*PCC*		
	Gumbel	*Frank*	*Gumbel*	*Frank*	*Student*
$\theta_{11} \setminus \nu_{11}$	4.32	16.69	4.34	16.78	$0.93 \setminus 3.6$
$\theta_{12} \setminus \nu_{12}$	3.45	13.01	2.24	7.10	$0.78 \setminus 6.7$
$\theta_{13} \setminus \nu_{13}$	-	-	3.45	12.98	$0.90 \setminus 5.5$
$\theta_{21} \setminus \nu_{21}$	1.97	5.96	1.01	0.08	$0.01 \setminus 9.6$
$\theta_{22} \setminus \nu_{22}$	-	-	1.02	0.61	$0.09 \setminus 14.5$
$\theta_{31} \setminus \nu_{31}$	-	-	1.03	0.27	$0.04 \setminus 17.3$
Log-likelihood	4741.05	4561.72	4842.25	4632.19	4643.38
p-value of S_n	0.000	0.000	0.000	0.000	0.000
p-value of T_n	0.002	0.000	0.089	0.013	0.070

Pair-copula construction

Also for the PCCs, the variables are ordered such that the copulae fitted at level 1 in the decomposition are those corresponding to the three largest Kendall's tau values. Hence, C_{11} is the copula of Vestby and Ski, C_{12} is the copula of Ski and Nannestad, and C_{13} is the copula of Ski and Hurdal. The parameters of the PCC are estimated using Algorithm 4 in Aas et al. (2007). The three rightmost columns of Table VI.4 show the estimated parameter values, resulting log-likelihoods and p-values for the Gumbel, Frank and Student PCCs. We see that, as for the NACs, all considered PCCs are strongly rejected by S_n. Hence, from the S_n-results, it is not possible to determine which of the two constructions that best fit the precipitation data and we therefore also used T_n. This procedure also rejects both NACs, but it fails to reject the Gumbel and Student PCCs, with the Gumbel PCC seemingly the best. Hence, we conclude that the Gumbel PCCs provides the best fit, but that there is a need for further research to find copula types that better captures the properties of the precipitation data.

VI.3.3 Application 2: Equity returns

In this section, we study an equity portfolio. The portfolio is comprised of four time series of daily log-return data from the period 14.08.2003 to 29.12.2006 (852 observations for each firm). The data set was downloaded from `http://finance.yahoo.com`. The firms are British Petroleum (BP), Exxon Mobile Corp (XOM), Deutsche Telekom AG (DT) and France Telecom (FTE). Financial log-returns are usually not independent over time. Hence, the original vectors of log-returns are processed by a GARCH filter before further modelling. We use the GARCH(1,1)-model (Bollerslev, 1986):

$$\begin{aligned} r_t &= c + \sigma_t \, z_t \\ \mathrm{E}[z_t] = 0 \text{ and } \mathrm{Var}[z_t] &= 1 \\ \sigma_t^2 &= a_0 + a \, \varepsilon_{t-1}^2 + b \, \sigma_{t-1}^2. \end{aligned} \tag{VI.5}$$

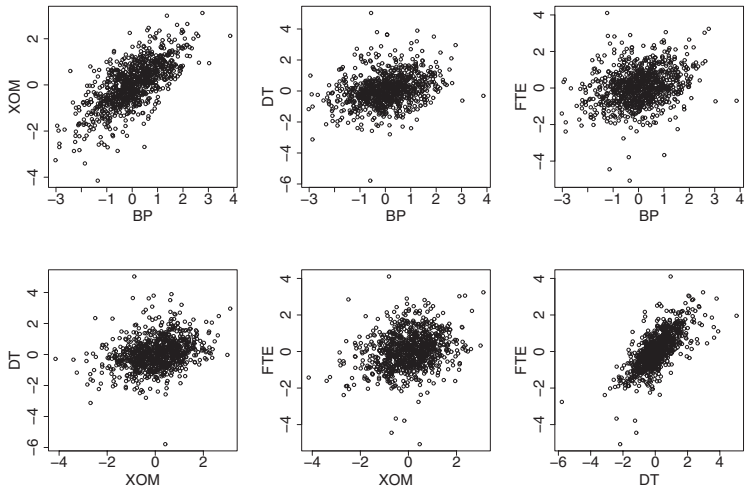


Figure VI.7: GARCH-filtered daily log-returns for our four stocks for the period from 14.08.2003 to 29.12.2006.

It is well recognised that GARCH models, coupled with the assumption of conditionally normally distributed errors are unable to fully account for the tails of the distributions of daily returns. Hence, we follow Venter and de Jongh (2002) and use the Normal Inverse Gaussian (NIG) distribution (Barndorff-Nielsen, 1997) as the conditional distribution. In a study performed by Venter and de Jongh (2004) the NIG distribution outperforms a skewed Student's t-distribution and a non-parametric kernel approximation as the conditional distribution of a one-dimensional GARCH process. After filtering the original returns with (VI.5) (estimated parameter values are shown in Appendix VI.A), the standardised residual vectors are converted to uniform pseudo-observations. Figures VI.6-VI.8 show the filtered daily log-returns and pseudo-observations for each pair of assets.

Based on visual inspection and preliminary goodness-of-fit tests for bivariate pairs (like for the precipitation data, the copulae taken into consideration were the Student, Clayton, survival Clayton, Gumbel and Frank copulae), we decided to examine a Frank NAC and Frank and Student PCC's for this data set.

Hierarchically nested Archimedean construction

Also for this data set, the most appropriate ordering of the variates in the decomposition is found by comparing Kendall's tau values for all bivariate pairs. The Kendall's tau values are shown in Table VI.5. As expected, stocks within one industrial sector are more dependent than stocks from different sectors. Hence, we choose C_{11} as the copula of BP and XOM, C_{12} as the copula of DT and FTE, and C_{21} as the copula of the remaining pairs. The leftmost column of Table VI.6 shows the estimated parameter values, resulting log-likelihood and p-value for the Frank HNAC. Even though this structure is not rejected by T_n, the strong rejection by S_n suggests that the fit is not very good. Hence, we conclude that the Frank NAC is not able to appropriately fit the equity data.

(VI.13)

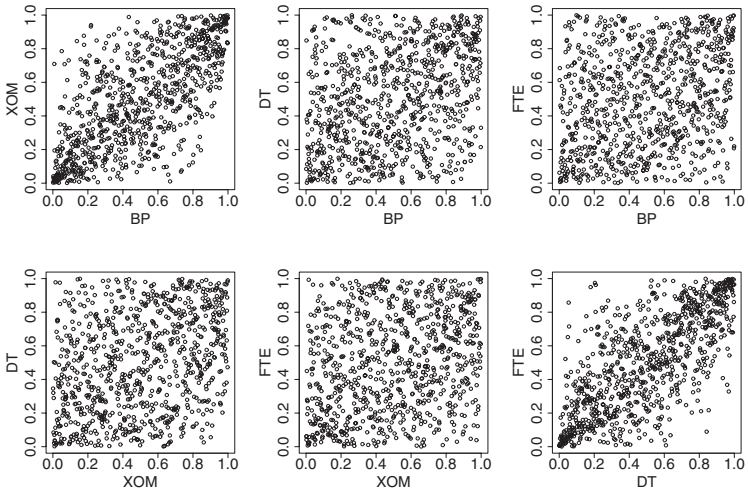


Figure VI.8: Pseudo-observations corresponding to Figure VI.7.

Table VI.5: Estimated Kendall's tau for pairs of variables for our four stocks.

Firm	XOM	DT	FTE
BP	0.45	0.19	0.20
XOM		0.23	0.17
DT			0.48

(VI.14)

Table VI.6: Estimated parameters, log-likelihood and estimated p-values for NAC and PCCs fitted to the filtered equity data.

Parameter	*NAC*	*PCC*	
	Frank	*Frank*	*Student*
$\theta_{11} \setminus \nu_{11}$	5.57	5.56	$0.70 \setminus 13.8$
$\theta_{12} \setminus \nu_{12}$	6.34	1.89	$0.32 \setminus 134.5$
$\theta_{13} \setminus \nu_{13}$	-	6.32	$0.73 \setminus 6.4$
$\theta_{21} \setminus \nu_{21}$	1.78	0.91	$0.14 \setminus 12.0$
$\theta_{22} \setminus \nu_{22}$	-	0.30	$0.06 \setminus 20.6$
$\theta_{31} \setminus \nu_{31}$	-	0.33	$0.07 \setminus 17.8$
Log-likelihood	616.45	618.63	668.49
p-value of S_n	0.006	0.008	0.410
p-value of T_n	0.385	0.385	0.697

Pair-copula construction

Again, the most appropriate ordering of the variates in the decomposition is determined by the size of the Kendall's tau values. Hence, we choose C_{11} as the copula of BP and XOM, C_{12} as the copula of XOM and DT, and C_{13} as the copula of DT and FTE. The parameters of the PCC are estimated by maximum likelihood, see Algorithm 4 in Aas et al. (2007). The two rightmost columns of Table VI.6 shows the estimated parameter values, resulting log-likelihood and estimated p-values for the Frank and Student PCCs. We see that the Frank PCC is rejected by S_n. Moreover, the p-value of T_n is equal to the one for the Frank NAC. The Student PCC, on the other hand, provides a very good fit and is not even rejected by S_n. Hence, we conclude that it fits the equity data very well.

VI.3.4 Validation

With the increasing complexity of models there is always the risk of overfitting the data. To examine whether this is the case for the PCC, we validate it out-of-sample for the equity portfolio. More specifically, we use the GARCH-NIG-Student PCC described in Section VI.3.3 to determine the risk of the return distribution for an equally weighted portfolio of BP, XOM, DT, and FTE over a one-day horizon. The equally-weighted portfolio is only meant as an example. In practice, the weights will fluctuate unless the portfolio is rebalanced every day.

The model estimated from the period 14.03.2003 to 29.12.2006 is used to forecast 1-day VaR at different significance levels for each day in the period from 30.12.2006 to 11.06.2007 (110 days). The test procedure is as follows: For each day t in the test set:

1. For each variable $j = 1, \ldots, 4$, compute the one-step ahead forecast of $\sigma_{j,t}$, given information up to time t.
2. For each simulation $n = 1, \ldots, 10,000$
   - Generate a sample $u_1, \ldots u_4$ from the estimated Student PCC.
   - Convert $u_1, \ldots u_4$ to NIG(0,1)-distributed samples $z_1, \ldots, z_4$ using the inverses of the corresponding NIG distribution functions.
   - For each variable $j = 1, \ldots, 4$, determine the log-return $r_{j,t} = c_{j,t} + \sigma_{j,t}\, z_j$. (Here $c_{j,t}$ is computed as the mean of the last 100 observed log-returns.)
   - Compute the return of the portfolio as $r_{p,t} = \sum_{j=1}^{4} \frac{1}{4} r_{j,t}$.
3. For significance levels $q \in \{0.005, 0.01, 0.05\}$
   - Compute the 1-day VaR_t^q as the qth-quantile of the distribution of $r_{p,t}$.

(VI.15)

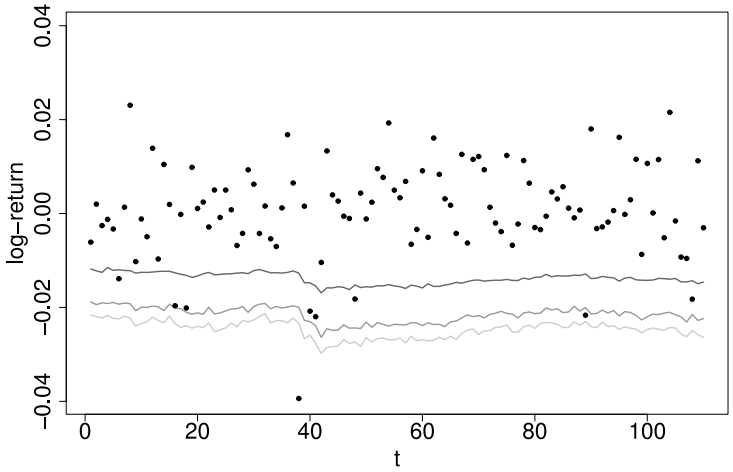


Figure VI.9: Log-returns for the equity portfolio for the period 30.12.2006 - 11.06.2007 along with 0.5%, 1%, 5% VaR simulated from the estimated GARCH-NIG-Student PCC.

Table VI.7: Number of violations of VaR, expected number of violations and p-values for the Kupiec test.

Significance level	0.005	0.01	0.05
Observed	1	2	9
Expected	0.55	1.1	5.5
P-value	0.13	0.44	0.16

- If VaR_t^q is greater than the observed value of $r_{p,t}$ this day, a violation is said to occur.

Figure VI.9 shows the actual log-returns for the portfolio in the period 30.12.2006 to 11.06.2007 and the corresponding VaR levels obtained from the procedure described above. Further, the two upper rows of Table VI.7 gives the number of violations x, of VaR for each significance level and with the expected values, respectively. To test the significance of the differences between the observed and the expected values, we use the likelihood ratio statistic by Kupiec (1995). The null hypothesis is that the expected proportion of violations is equal to α. Under the null hypothesis, the likelihood ratio statistic given by

$$2\ln\left(\left(\frac{x}{N}\right)^x\left(1-\frac{x}{N}\right)^{N-x}\right)-2\ln\left(\alpha^x(1-\alpha)^{N-x}\right),$$

where N is the length of the sample, is asymptotically distributed as $\chi^2(1)$. We have computed p-values of the null hypothesis for each quantile. The results are shown in the lower row of Table VI.7. If we use a 5% level for the Kupiec LR statistic, the null hypothesis is not rejected for any of the three quantiles. Hence, the GARCH-NIG-Student PCC seems to work very well out-of-sample.

VI.4 Summary and Conclusions

In this paper we have reviewed two classes of structures for construction of higher-dimensional dependence; the nested Archimedean constructions (NACs) and the pair-copula constructions

(PCCs). For both structures, a multivariate data set is modelled using a cascade of lower-dimensional copulae. They differ however in their construction of the dependence structure, the PCC being more flexible in that it allows for the free specification of $d(d-1)/2$ copulae, while the NAC's only allow for $d-1$.

Simulation and estimation techniques for the two structures have been examined, and we have shown that the PCCs in general are more computationally efficient than the NACs. The fit of the two constructions has been tested on two different four-dimensional data sets; precipitation values and equity returns, using state of the art copula goodness-of-fit procedures. The NACs considered are strongly rejected for both our data sets. For the precipitation data the Gumbel PCC provides a better fit. However, since even this structure is rejected by one of the goodness-of-fit tests used, one should look for other copula types that might capture the properties of the precipitation data even better than the Gumbel copula does. For the equity data, the Student PCC provides a good fit, and through VaR calculations we have shown that it does not overfit the training data, but works very well also out-of-sample.

Based on the properties presented and the results from the two applications we recommend in general the PCC over the NAC for the following reasons. First, the NAC has an important restriction in that the degree of dependence must decrease with the level of nesting. When looking for appropriate data sets for the applications in this paper, it turned out to be quite difficult to find real-world data sets satisfying this restriction. In addition, the NAC is restricted to the Archimedean class, and there are even restrictions on which Archimedean copulae that can be mixed. There might be real-world situations where there are natural hierarchy groupings of variables. In such cases the NAC's may come into consideration. However, the technical restrictions of the NAC might prevent extensive use.

The PCC, on the other hand, can be built using copulas from any class and there are no restrictions on the parameters of the structure. As far as we are concerned, the only potential disadvantage of the PCC compared to the NAC is that neither the unspecified bivariate margins nor the multivariate distribution in general will belong to a known parametric family. However, we do not view this as a significant problem since these distributions might easily be obtained through simulation.

Acknowledgements: Daniel Bergs work is supported by the Norwegian Research Council, grant number 154079/420 and Kjersti Aas' part is sponsored by the Norwegian fund Finansmarkedsfondet. We are very grateful to Cornelia Savu, Institute for Econometrics, University of Münster, Germany, for providing us with her code for the NAC's along with helpful comments. In addition we would like to express our deep gratitude for assistance on the GARCH-NIG filtration to Professor J.H. Venter, Centre of Business Mathematics and Informatics, North-West University, Potchefstroom, South Africa. Finally, we thank participants at the conference on copulae and multivariate probability distributions at Warwick Business School in September 2007 for their valuable comments, in particular Professor Alexander McNeil.

Table VI.8: Estimated GARCH and NIG parameters for our four stocks.

Parameter	*BP*	*XOM*	*DT*	*FTE*
a_0	1.598e-06	1.400e-06	1.801e-06	1.231e-06
a	0.010	0.023	0.025	0.028
b	0.978	0.968	0.963	0.966
β	-0.357	-0.577	0.105	0.037
ψ	3.686	2.293	1.173	1.670

VI.A Parameters for GARCH-NIG model

Table VI.8 shows the estimated parameters for the GARCH-NIG model used in Section VI.3.3. For further details of the estimation procedure see Venter and de Jongh (2002).

Bibliography

Aas, K., C. Czado, A. Frigessi, and H. Bakken (2007). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics 42*, In press.

Aas, K., R. Huseby, and M. Thune (1999, June). Data Mining: A Survey. Report, Norwegian Computing Centre. ISBN 82–539–0426–6.

Alfo, M., S. Caiazza, and G. Trovato (2005). Extending a Logistic Approach to Risk Modeling through Semiparametric Mixing. *Journal of Financial Services Research 28*, 163–176.

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance 23*(4), 589–609.

Altman, E. I. and E. Hotchkiss (2005). *Corporate Financial Distress and Bankruptcy* (Third ed.). John Willey & Sons, New York.

Altman, E. I. and P. Narayanan (1997). An International Survey of Business Failure Classification Models. *Financial Markets, Institutions and Instruments 6*(2), 1–57.

Anderson, T. W. and D. A. Darling (1954). A Test of Goodness of Fit. *Journal of the American Statistical Association 49*(268), 765–769.

Arjas, E. and P. Haara (1987). A logistic regression model for hazard: Asymptotic results. *Scandinavian journal of Statistics 14*(1), 1–18.

Aslan, B. and G. Zech (2002). Comparison of Different Goodness–of–Fit Tests. In *Advanced Statistical Techniques in PARTICLE Physics*. Grey College, Durham.

Atiya, A. F. (2001). Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results. *IEEE Transactions on Neural Networks 12*(4), 929–935.

Barndorff-Nielsen, O. E. (1997). Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling. *Scandinavian journal of Statistics 24*, 1–13.

Beaver, W. H. (1966). Financial Ratios as Predictors of Failure. *Journal of Accounting Research 4*, 71–111.

Bedford, T. and R. M. Cooke (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence 32*, 245–268.

Bedford, T. and R. M. Cooke (2002). Vines - a new graphical model for dependent random variables. *Annals of Statistics 30*, 1031–1068.

Berg, D. (2007a). Bankruptcy prediction by generalized additive models. *Applied Stochastic Models in Business and Industry 23*(2), 129–143.

Berg, D. (2007b). Copula Goodness–of–fit Tests: An overview and Power Comparison. Working paper, University of Oslo and The Norwegian Computing Center. Submitted for review.

Berg, D. and H. Bakken (2007). A copula goodness-of-fit approach based on the conditional probability integral transform. Working paper, University of Oslo and The Norwegian Computing Center. Submitted for review.

Bernhardsen, E. (2001). A Model of Bankruptcy Prediction. Working paper 10/2001, Norges Bank. ISBN 82–7553–186–1.

Billingsley, P. (1999). *Convergence of probability measures*. Wiley Series in Probability and Statistics. Probability and Statistics. New York: John Wiley & Sons.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics 31*, 307–327.

Breslow, N. E. and D. G. Clayton (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association 88*(421), 9–25.

Breymann, W., A. Dias, and P. Embrechts (2003). Dependence structures for multivariate high–frequency data in finance. *Quantitative Finance 3*(1), 1–16.

Brockman, P. and H. J. Turtle (2003). A barrier option framework for corporate security valuation. *Journal of Financial Economics 67*(3), 511–529.

Chava, S. and R. A. Jarrow (2004). Bankruptcy Prediction with Industry Effects. *Review of Finance 8*(4), 537–569.

Chen, S., W. Härdle, and R. Moro (2006, November). Estimation of default probabilities with support vector machines. SFB 649 Discussion Papers SFB649DP2006-077, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany. Available at http://ideas.repec.org/p/hum/wpaper/sfb649dp2006-077.html.

Chen, X. and Y. Fan (2005). Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection. *Canadian Journal of Statistics 33*, 389–414.

Chen, X. and Y. Fan (2006). Estimation and model selection of semi-parametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics 135*, 125–154.

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.

Cherubini, U., E. Luciano, and W. Vecchiato (2004). *Copula Methods in Finance*. Wiley.

Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika 65*, 141–151.

Crosbie, P. (1997). Modeling default risk. Technical Document, KMV Corporation.

Crouhy, M., D. Galai, and R. Mark (2001). *Risk Management*. McGraw-Hill.

Cui, S. and Y. Sun (2004). Checking for the gamma frailty distribution under the marginal proportional hazards frailty model. *Statistica Sinica 14*, 249–267.

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

Das, S. R., D. Duffie, N. Kapadia, and L. Saita (2007). Common Filings: How Corporate Defaults Are Correlated. *Journal of Finance 62*(1), 93–117.

David, H. A. (1981). *Order Statistics* (2 ed.). John Wiley and Sons.

De Giorgi, E. G. and J. Burkhard (2006). An Intensity Based Non–Parametric Default Model for Residential Mortgage Portfolios. *Journal of Risk 8*(4), 57–95.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés: Un test non paramétrique d'indépendence. *Acad. Royal Bel., Bull. Class. Sci., 5^e série 65*, 274–292.

Deheuvels, P. (1984). The Characterization of Distributions by Order Statistics and Record Values: A Unified Approach. *Annals of Eugenics 21*(2), 326–334.

Demarta, S. and A. J. McNeil (2005). The t Copula and Related Copulas. *International Statistical Review 73*(1), 111–129.

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.

Dimitras, A., S. Zanakis, and C. Zopounidis (1996). A survey of business failures with an emphasis on prediction methods and industrial applications. *European journal of Operational Research 90*, 487–513.

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*(4), 1053–1068.

Dobrić, J. and F. Schmid (2007). A Goodness of Fit Test for Copulas based on Rosenblatt's Transformation. *Computational Statistics & Data Analysis 51*(9), 4633–4642.

Doumpos, M. and C. Zopounidis (1999). A Multicriteria Discrimination Method for the Prediction of Financial Distress: The Case of Greece. *Multinational Finance journal 3*, 71–101.

Duffie, D., A. Eckner, G. Horel, and L. Saita (2006, May). Frailty correlated default. Working Paper, Graduate School of Business, Stanford University.

Duffie, D., L. Saita, and K. Wang (2007). Multi–period corporate default prediction with stochastic covariates. *Journal of Financial Economics 83*(3), 635–665.

Embrechts, P., F. Lindskog, and A. McNeil (2003). Modelling dependence with copulas and applications to risk management. In S.T.Rachev (Ed.), *Handbook of Heavy Tailed Distributions in Finance*. North-Holland: Elsevier.

Embrechts, P., A. McNeil, and D. Straumann (1999, May). Correlation and Dependence in Risk Management: Properties and Pitfalls. *RISK*, 69–71.

Engelmann, B., E. Hayden, and D. Tasche (2003). Testing Rating Accuracy. *Risk 16*(1), 82–86.

Fahrmeir, L. and G. Tutz (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models* (Second ed.). Springer.

Falkenstein, E., A. Borat, and L. Carty (2000). RiskCalc For Private Companies: Moody's Default Methodology. *Moody's Investors Service*.

Fang, H. B., K. T. Fang, and S. Kotz (2002). The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis 82*, 1–16.

Fermanian, J. (2005). Goodness of Fit Tests for Copulas. *Journal of Multivariate Analysis 95*, 119–152.

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

Florax, R. J. G. M., H. Folmer, and S. J. Rey (2006). A comment on specification searches in spatial econometrics: The relevance of hendry's methodology: A reply. *Regional Science and Urban Economics 36*, 300–308.

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges son données. *Ann. Univ. Lyon, Sect. A 9*, 53–77.

Frees, E. W., J. Carriere, and E. A. Valdez (1996). Annuity valuation with dependent mortality. *Journal of Risk & Insurance 63*, 229–261.

Frees, E. W. and E. A. Valdez (1998). Understanding relationships using copulas. *North American Actuarial journal 2*(1), 1–25.

Gänssler, P. and W. Stute (1987). Seminar on empirical processes. Volume 9 of *DMV Seminar*, Birkhäuser, Basel.

Genest, C. and A.-C. Favre (2007). Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of Hydrologic Engineering 12*, 347–368.

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

Genest, C. and R. MacKay (1986). The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician 40*, 280–283.

Genest, C., J.-F. Quessy, and B. Rémillard (2006a). Goodness-of-fit procedures for copula models based on the probability integral transform. *Scandinavian journal of Statistics 33*, 337–366.

Genest, C., J.-F. Quessy, and B. Rémillard (2006b). Local efficiency of a Cramér–von Mises test of independence. *Journal of Multivariate Analysis 97*, 274–294.

Genest, C., J.-F. Quessy, and B. Rémillard (2006c). On the joint asymptotic behavior of two rank-based estimators of the association parameter in the gamma frailty model. *Statistics & Probability Letters*, 10–18.

Genest, C., J.-F. Quessy, and B. Rémillard (2007). Asymptotic local efficiency of Cramér–von Mises tests for multivariate independence. *The Annals of Statistics 35*(1), 166–191.

Genest, C. and B. Rémillard (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Poincaré: Probabilités et statistiques 44*, In press.

Genest, C., B. Rémillard, and D. Beaudoin (2008). Omnibus goodness–of–fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics 42*, in press.

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

Ghoudi, K. and B. Rémillard (1998). Empirical processes based on pseudo-observations. In B. Szyskowicz (Ed.), *Asymptotic Methods in Probability and Statistics, A Volume in Honour of Miklos Csorgo*, pp. 171–197. Elsevier, Amsterdam.

Glen, A. G., L. M. Leemis, and D. R. Barr (2001). Order Statistics in Goodness–of–Fit Testing. *IEEE Transactions on Reliability 50*(2), 209–213.

Glidden, D. V. (1999). Checking the adequacy of the gamma frailty model for multivariate failure times. *Biometrika 86*, 381–393.

Gumbel, E. (1960). Distributions des valeurs extrémes en plusieurs dimensions. *Publications de l'Institute de Statistíque de l'Université de Paris 9*, 171–173.

Hájek, J. and Z. Sidák (1967). *Theory of Rank Tests*. Academic Press, New York.

Haling, M. and E. Hayden (2006). Bank Failure Prediction: A 2–Step Approach. Working Paper. Oesterreichische Nationalbank and University of Vienna.

Härdle, W., R. A. Moro, and D. Schäfer (2005). Predicting Bankruptcy with Support Vector Machines. SFB 649 Discussion Paper 2005–009.

Hastie, T. and R. Tibshirani (1986). Generalized Additive Models. *Statistical Science 1*, 297–318.

Hastie, T. J. and R. J. Tibshirani (1990). *Generalized Additive Models*. Chapman & Hall.

Hendry, D. F. (2006). Specification searches in spatial econometrics: The relevance of hendry's methodology. *Regional Science and Urban Economics 36*, 309–312.

Hillegeist, S. A., E. K. Keating, D. P. Cram, and K. G. Lundstedt (2004). Assessing the Probability of Bankruptcy. *Review of Accounting Studies 9*, 5–34.

Hoeffding, W. (1940). Masstabinvariante Korrelationstheorie. *Schriften des Matematischen Instituts und des Instituts für Angewandte Matematikde Universität Berlin 5*, 179–233. [Reprinted as: Scale-invariant correlation theory. In: Fisher, N. I. and Sen, P. K., editors, (1994). *The Collected Works of Wassily Hoeffding*, 57–107. Springer, New York.].

Hoeffding, W. (1941). Masstabinvariante Korrelationsmasse für diskontinuierliche Verteilungen. *Arkiv fr matematischen Wirtschaften und Sozialforschung 7*, 49–70. [Reprinted as: Scale-invariant correlation measures for discontinuous distributions. In: Fisher, N. I. and Sen, P. K., editors, (1994). *The Collected Works of Wassily Hoeffding*, 109–133. Springer, New York.].

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.

Huard, D., G. Évin, and A.-C. Favre (2006). Bayesian copula selection. *Computational Statistics and Data Analysis 51*, 809–822.

Huffera, F. W. and C. Park (2007). A test for elliptical symmetry. *Journal of Multivariate Analysis 98*(2), 256–281.

Joe, H. (1996). Families of *m*-variate distributions with given margins and *m(m-1)/2* bivariate dependence parameters. In L. Rüschendorf and B. Schweizer and M. D. Taylor (Ed.), *Distributions with Fixed Marginals and Related Topics*.

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

Kim, G., M. J. Silvapulle, and P. Silvapulle (2006). Comparison of semiparametric and parametric methods for estimating copulas. To be published in Computational Statistics and Data Analysis.

Krzanowski, W. J. (1998). *An Introduction to Statistical Modelling*. Arnold.

Krzysztofowicz, R. and K. S. Kelly (1996). A meta-Gaussian distribution with specified marginals. Technical report, University of Virginia.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of derivatives 2*, 173–184.

Kurowicka, D. and R. M. Cooke (2004). Distribution - free continuous bayesian belief nets. In *Fourth International Conference on Mathematical Methods in Reliability Methodology and Practice, Santa Fe, New Mexico*.

Kurowicka, D. and R. M. Cooke (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. New York: Wiley.

Lambert, P. and F. Vandenhende (2002). A copula-based model for multivariate non-normal longitudinal data: analysis of a dose titration safety study on a new antidepressant. *Statistics in Medicine 21*, 3197–3217.

Lang, L. and R. Stulz (1992). Contagion and competitive intra-industry effects of bankrutcy announcements. *Journal of Financial Economics 32*(1), 45–60.

LeClere, M. (2000). The occurence and timing of events: survival analysis applied to the study of financial distress. *Journal of Accounting Literature 19*, 158–189.

Lindskog, F., A. J. McNeil, and U. Schmock (2003). Kendall's tau for elliptical distributions. In G. Bol, G. Nakhaeizadeh, S. Rachev, T. Ridder, and K.-H. Vollmer (Eds.), *Credit Risk: Measurement, Evaluation and Management*, pp. 149–156. Heidelberg: Physica.

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

Malevergne, Y. and D. Sornette (2006). *Extreme Financial Risks: From Dependence to Risk Management*. Springer-Verlag Berlin Heidelberg.

Marsaglia, G. and J. Marsaglia (2004). Evaluating the Anderson-Darling Distribution. *Journal of Statistical Software 9*.

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

Martin, D. (1977). Early warning of bank failure : A logit regression approach. *Journal of Banking & Finance 1*(3), 249–276.

Mashal, R. and A. Zeevi (2002). Beyond correlation: Extreme co–movements between financial assets. Technical report, Columbia University.

McNeil, A., R. Frey, and P. Embrechts (2006). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

McNeil, A. J. (2007). Sampling nested archimedean copulas. *To appear in Journal of Statistical Computation and Simulation*.

McNeil, A. J. and J. Neslehova (2007). Multivariate archimedean copulas, d-monotone functions and l1-norm symmetric distributions. *To appear in Annals of Statistics*.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance 29*, 449–470.

Michele, C. D. and G. Salvadori (2006). Copulas and extreme storm structure. *Geophysical Research Abstracts 8*. 06124.

Mikosch, T. (2006). Copulas: tales and facts. *Extremes 9*, 3–20. Discussion, 21–53. Rejoinder 55–62.

Min, A. and C. Czado (2007). Bayesian estimation of multivariate copulas using pair-copula constructions. Working paper, Technische Universität München.

Morillas, P. M. (2005). A method to obtain new copulas from a given one. *Metrika 61*, 169–184.

Musshoff, O. M., M. Odening, and W. Xu (2006, July 23-26). Modeling and hedgning rain risk. In *Annual Meeting of the American Agricultural Economics Association (AAEA), Long Beach, USA*.

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

Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research 18*, 109–131.

Okhrin, O., Y. Okhrin, and W. Schmid (2007, May 30-31). 2007. In *Radon Workshop on Financial and Actuarial Mathematics for Young Researchers, Linz, Austria*.

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

Plat, H. D. and M. B. Plat (1990). Improving ex ante bankruptcy classification using stable predictive variables. *Journal of Business Finance and Accounting 17*(1), 31–51.

Plat, H. D. and M. B. Plat (1991). A note on the use of industry–relative ratios in bankruptcy prediction. *Journal of Banking and Finance 15*(6), 1183–1194.

Posch, P. N., L. Gunter, and S. Christiane (2005, May). Bayesian methods for improving credit scoring models. Finance 0505024, EconWPA. Available at http://ideas.repec.org/p/wpa/wuwpfi/0505024.html.

Quessy, J.-F. (2005). *Théorie et application des copules : tests d'adéquation, tests d'indépendance et bornes pour la valeur–à-risque.* Ph. D. thesis, Université Laval.

Quessy, J.-F., M. Mesfioui, and M.-H. Toupin (2007). A goodness-of-fit test based on Spearmans dependence function. Working paper, Laboratoire de probabilités et statistique de l'Université du Québec à Trois-Rivières.

Romano, C. (2002). Calibrating and simulating copula functions: An application to the Italian stock market. Working paper n. 12, CIDEM, Universit'a degli Studi di Roma "La Sapienza".

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

Savu, C. and M. Trede (2004). Goodness–of–fit tests for parametric families of Archimedean copulas. CAWM discussion paper, No. 6.

Savu, C. and M. Trede (2006). Hierarchical Archimedean Copulas. Technical report, Institute of Econometrics, University of Mnster.

Scaillet, O. (2006). Kernel Based Goodness–of–Fit Tests for Copulas with Fixed Smoothing Parameters. *Journal of Multivariate Analysis 98*, 533–543.

Schmid, F. and R. Schmidt (2007). Multivariate extensions of spearman's rho and related statistics. *Statistics and Probability Letters 77*, 407–416.

Shih, J. H. (1998). A goodness–of–fit test for association in a bivariate survival model. *Biometrika 85*, 189–200.

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

Shorack, G. R. and J. A. Wellner (1986). *Empirical processes with applications to statistics*. Wiley, New York.

Shumway, T. (2001). Forecasting Bankruptcy More Accuratelly: A Simple Hazard Model. *Journal of Business 74*(1), 101–124.

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

Sobehart, J. and S. Keenan (2001). Measuring Default Accuracy. *Risk 14*(3), 31–33.

Sobehart, J. and R. Stein (2000). Moody's Public Firm Risk Model: A Hybrid Approach To Modeling Short Term Default Risk. *Moody's Investors Service*.

Sobehart, J. R., S. Keenan, and R. Stein (2000). Rating Methodology – Benchmarking Quantitative Default Risk Models: A Validation Methodology. Technical report, Moody's Investors Service.

Sun, L. and P. P. Shenoy (2007). Using Bayesian Networks for Bankruptcy Prediction: Some Methodological Issues. Forthcoming in European journal of Operational Research.

Tam, K. (1991). Neural network models and the prediction of bank bankruptcy. *Omega 19*, 429–445.

Tam, K. and M. Kiang (1992). Managerial Applications of Neural Networks: The Case of Bank Failure Predictions. *Management Science 38*(7), 926–947.

Tsukahara, H. (2005). Semiparametric estimation in copula models. *Canadian Journal of Statistics 33*, 357–375.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes. With applications to statistics*. Springer Series in Statistics. New York: Springer–Verlag.

Venter, J. H. and P. J. de Jongh (2002). Risk estimation using the normal inverse Gaussian distribution. *Journal of Risk 4*(2), 1–24.

Venter, J. H. and P. J. de Jongh (2004). Selecting an innovation distribution for garch modelsto improve efficiency of risk and volatility estimation. *Journal of Risk 6*(3), 27–53.

Wang, D. (1998). Aggregation of correlated risk portfolios: models and algorithms. In *Proceedings of the Casualty Actuarial Society*, Volume 85, pp. 848–939.

Whelan, N. (2004). Sampling from Archimedean copulas. *Quantitative Finance 4*, 339–352.

Wilcox, J. W. (1971). A Simple Theory of Financial Ratios as Predictors of Failure. *Journal of Accounting Research 9*(2), 389–395.

Wilson, R. and R. Sharda (1994). Bankruptcy Prediction Using Neural Networks. *Decision Support Systems 11*, 545–557.

Zhang, L. and V. P. Singh (2006). Bivariate flood frequency analysis using the copula model. *Journal of Hydrologic Engineering 11*(2), 150–164.

Zmijewski, M. E. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research 22*, 59–82.