Available online at www.sciencedirect.com
ScienceDirect
Insurance: Mathematics and Economics 44 (2009) 199–213
INSURANCE MATHEMATICS & ECONOMICS
www.elsevier.com/locate/ime

Goodness-of-fit tests for copulas: A review and a power study

Christian Genest[a,*], Bruno Rémillard[b], David Beaudoin[c]

[a] Département de mathématiques et de statistique, Université Laval, 1045, avenue de la Médecine, Québec (Québec), Canada G1V 0A6
[b] Service d'enseignement des méthodes quantitatives de gestion, HEC Montréal, 3000, chemin de la Côte-Sainte-Catherine, Montréal (Québec), Canada H3T 2A7
[c] Département d'opérations et systèmes de décision, Université Laval, 2325, rue de la Terrasse, Québec (Québec), Canada G1V 0A6

Received October 2007; accepted 11 October 2007

Abstract

Many proposals have been made recently for goodness-of-fit testing of copula models. After reviewing them briefly, the authors concentrate on "blanket tests", i.e., those whose implementation requires neither an arbitrary categorization of the data nor any strategic choice of smoothing parameter, weight function, kernel, window, etc. The authors present a critical review of these procedures and suggest new ones. They describe and interpret the results of a large Monte Carlo experiment designed to assess the effect of the sample size and the strength of dependence on the level and power of the blanket tests for various combinations of copula models under the null hypothesis and the alternative. To circumvent problems in the determination of the limiting distribution of the test statistics under composite null hypotheses, they recommend the use of a double parametric bootstrap procedure, whose implementation is detailed. They conclude with a number of practical recommendations.

Keywords: Anderson–Darling statistic; Copula; Cramér–von Mises statistic; Gaussian process; Goodness-of-fit; Kendall's tau; Kolmogorov–Smirnov statistic; Monte Carlo simulation; Parametric bootstrap; Power study; Pseudo-observations; P-values

1. Introduction

Consider a continuous random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with joint cumulative distribution function H and margins $F_1, \ldots, F_d$. The copula representation of H is given by $H(x_1, \ldots, x_d) = C\{F_1(x_1), \ldots, F_d(x_d)\}$, where C is a unique cumulative distribution function having uniform margins on $(0, 1)$. A copula model for $\mathbf{X}$ arises when C is unknown but assumed to belong to a class

$$\mathcal{C}_0 = \{C_\theta : \theta \in \mathcal{O}\},$$

where $\mathcal{O}$ is an open subset of $\mathbb{R}^p$ for some integer $p \geq 1$. The books of Joe (1997) and Nelsen (2006) provide handy compendiums of the most common parametric families of copulas.

Copula modeling has found many successful applications of late, notably in actuarial science, survival analysis and hydrology; see, e.g., Frees and Valdez (1998), Cui and Sun (2004) and Genest and Favre (2007) and references therein. However, nowhere has the methodology been adopted and used with greater intensity than in finance. Ample illustrations are provided in the books of Cherubini et al. (2004) and McNeil et al. (2005), notably in the context of asset pricing and credit risk management.

Given independent copies $\mathbf{X}_1 = (X_{11}, \ldots, X_{1d}), \ldots, \mathbf{X}_n = (X_{n1}, \ldots, X_{nd})$ of $\mathbf{X}$, the problem of estimating θ under the assumption

$$H_0 : C \in \mathcal{C}_0$$

has already been the object of much work; see, e.g., Genest et al. (1995), Shih and Louis (1995), Joe (1997, 2005), Tsukahara (2005) or Chen et al. (2006). However, the complementary issue of testing H_0 is only beginning to draw attention.

The situation is evolving rapidly but at this point in time, the literature on the subject can be divided broadly into three groups:

(1) Procedures developed for testing specific dependence structures such as the Normal copula (Malevergne and Sornette, 2003) or the equally popular Clayton family, also referred to as the gamma frailty model in survival analysis (Shih, 1998; Glidden, 1999; Cui and Sun, 2004).

* Corresponding author.
E-mail address: genest@mat.ulaval.ca (C. Genest).

0167-6687/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.insmatheco.2007.10.005

(2) Statistics that can be used to test the goodness-of-fit of any class of copulas but whose implementation involves:
  (a) an arbitrary parameter, as in the rank-based statistic due to Wang and Wells (2000);
  (b) kernels, weight functions and associated smoothing parameters, as in Berg and Bakken (2005), Fermanian (2005), Panchenko (2005) and Scaillet (2007);
  (c) ad hoc categorization of the data into a multiway contingency table in order to apply an analogue of the standard chi-squared test, along the lines of Genest and Rivest (1993), Klugman and Parsa (1999), Andersen et al. (2005), Dobrić and Schmid (2005) or Junker and May (2005).

(3) "Blanket tests", i.e., those applicable to all copula structures and requiring no strategic choice for their use. Included in this category are variants of the Wang–Wells approach due to Genest et al. (2006), but also the procedures investigated or used by Breymann et al. (2003), Genest and Rémillard (in press) and Dobrić and Schmid (2007).

And then there are authors who, in applied work, use standard goodness-of-fit statistics as a tool for choosing between several copulas, but without attempting to formally test whether the selected model is appropriate, in the light of a P-value. See, e.g., the analysis of stock index returns by Ané and Kharoubi (2003).

The purpose of this paper is to present a critical review of the *blanket* goodness-of-fit tests proposed to date, to suggest variants or improvements, and to compare the relative power of these procedures through a Monte Carlo study involving a large number of copula alternatives and dependence conditions. After some general considerations given in Section 2, existing tests are described in Section 3 and new statistics are proposed in Section 4. Listed in Section 5 are the factors considered in the study designed to assess the level and compare the power of the selected tests. Results are reported and discussed in Section 6. Finally, various observations and methodological recommendations are made in the Conclusion.

2. General considerations

There is a fundamental difference between the problem of estimating the dependence parameter of a copula model $\mathcal{C}_0 = \{C_\theta : \theta \in \mathcal{O}\}$ and the complementary issue of testing the validity of the null hypothesis $H_0 : C \in \mathcal{C}_0$ for some class $\mathcal{C}_0$ of copulas. The distinction is spelled out below, as it helps to understand the technical challenges associated with goodness-of-fit testing in this context.

2.1. Estimation

Two broad approaches to the estimation of the dependence parameter θ have been developed. They differ mainly through the user's willingness to make parametric assumptions or not about the unknown margins.

Given specific choices of parametric families $\mathcal{F}_j = \{F_{\gamma_j} : \gamma_j \in \Gamma_j\}$ of univariate distributions, estimation can proceed via the full standard maximum likelihood method under H_0 and the additional assumption that

$$H_0' : F_1 \in \mathcal{F}_1, \dots, F_d \in \mathcal{F}_d.$$

An alternative technique that is computationally more convenient has been advocated by Joe (1997). His "Inference Functions for Margins" or IFM approach proceeds in two steps: parametric estimates $F_{\hat\gamma_1}, \dots, F_{\hat\gamma_d}$ of the margins are first obtained under H_0'; they are then plugged into the log-likelihood, viz.

$$L(\theta) = \sum_{i=1}^{n} \log[c_\theta\{F_{\hat\gamma_1}(X_{i1}), \dots, F_{\hat\gamma_d}(X_{id})\}],$$

in which c_θ denotes the density of the copula C_θ (assuming that it exists). The function $L(\theta)$ is then maximized. As illustrated by Joe (2005), however, the gain in computational convenience often comes at the expense of efficiency. Kim et al. (2007) further show that an inappropriate choice of models for the margins may have detrimental effects on the estimation of the dependence parameter per se.

If one is unwilling to assume H_0', nonparametric estimation of the margins must be used. The most natural choice consists in replacing F_j by its empirical counterpart

$$\hat F_j(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(X_{ij} \le t),$$

and then estimating θ by the value $\hat\theta$ that maximizes the log pseudo-likelihood

$$\ell(\theta) = \sum_{i=1}^{n} \log[c_\theta\{\hat F_1(X_{i1}), \dots, \hat F_d(X_{id})\}].$$

This amounts to working with the ranks of the observations, because for all $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, d\}$, $R_{ij} = n\hat F_j(X_{ij})$ is the rank of X_{ij} among $X_{1j}, \dots, X_{nj}$.

The asymptotic normality of $\hat\theta$ was established by Genest et al. (1995), and by Shih and Louis (1995) in the presence of censorship. As shown by Genest and Werker (2002), however, this method is not asymptotically semi-parametrically efficient in general. See Klaassen and Wellner (1997) for a notable exception and Tsukahara (2005) or Chen et al. (2006) for other rank-based estimators.

2.2. Goodness-of-fit testing

When testing the hypothesis $H_0 : C \in \mathcal{C}_0$ that the dependence structure of a multivariate distribution is well-represented by a specific parametric family $\mathcal{C}_0$ of copulas, the option of modeling the margins by parametric families is no longer viable. For, it would be tantamount to testing the much narrower null hypothesis $H_0 \cap H_0'$ corresponding to a full parametric model. In this context, the marginal distributions $F_1, \dots, F_d$ are (infinite-dimensional) nuisance parameters.

Given that the underlying copula C of a random vector is invariant by continuous, strictly increasing transformations of its components, it appears that the only reasonable option for testing H_0 consists of basing the inference on the maximally invariant statistics with respect to this set of transformations, i.e., the ranks. Indeed, all formal goodness-of-fit tests mentioned in the introduction are rank-based. Alternatively, they can be viewed as functions of the collection $\mathbf{U}_1 = (U_{11}, \ldots, U_{1d})$, …, $\mathbf{U}_n = (U_{n1}, \ldots, U_{nd})$ of pseudo-observations deduced from the ranks, viz. $U_{ij} = R_{ij}/(n+1) = n\hat{F}_j(X_{ij})/(n+1)$, where the scaling factor $n/(n+1)$ is only introduced to avoid potential problems with c_θ blowing up at the boundary of $[0, 1]^d$.

The pseudo-observations $\mathbf{U}_1, \ldots, \mathbf{U}_n$ can be interpreted as a sample from the underlying copula C. It is plain, however, that they are *not* mutually independent and that their components are only *approximately* uniform on $(0, 1)$. Accordingly, any inference procedure based on these constructs should take these features into account. As will be seen, testing procedures that mistakenly ignore these considerations not only lack power but fail to hold their nominal level.

3. "Blanket tests" currently available

This section describes five rank-based procedures that have been recently proposed for testing the goodness-of-fit of any class of d-variate copulas. Of all the tests listed in Section 1, these are the only ones that qualify as "blanket", in the sense that they involve no parameter tuning or other strategic choices.

3.1. Two tests based on the empirical copula

As mentioned in Section 2, the pseudo-observations $\mathbf{U}_1, \ldots, \mathbf{U}_n$ constitute the maximally invariant statistics on which to test $H_0 : C \in \mathcal{C}_0$. The information they contain is conveniently summarized by the associated empirical distribution, viz.

$$C_n(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(U_{i1} \le u_1, \ldots, U_{id} \le u_d),$$
$$\mathbf{u} = (u_1, \ldots, u_d) \in [0, 1]^d. \tag{1}$$

It is usually called the "empirical copula", though it is neither a copula nor exactly the same (except asymptotically) as originally defined by Deheuvels (1979).

Gänßler and Stute (1987), Fermanian et al. (2004) and Tsukahara (2005) give various conditions under which C_n (or slight variants thereof) is a consistent estimator of the true underlying copula C, i.e., whether H_0 is true or not. Given that it is entirely nonparametric, C_n is arguably the most objective benchmark for testing $H_0 : C \in \mathcal{C}_0$. Therefore, natural goodness-of-fit tests consist in comparing a "distance" between C_n and an estimation C_{θ_n} of C obtained under H_0. Here and in the sequel, $\theta_n = \mathcal{T}_n(\mathbf{U}_1, \ldots, \mathbf{U}_n)$ stands for an estimate of θ derived from the pseudo-observations.

Goodness-of-fit tests based on the empirical process

$$\mathbb{C}_n = \sqrt{n}(C_n - C_{\theta_n})$$

are briefly considered by Fermanian (2005), who comments that they "seem to be unpractical, except by bootstrapping". Genest and Rémillard (in press) examine the implementation issues in detail. In particular, they consider rank-based versions of the familiar Cramér–von Mises and Kolmogorov–Smirnov statistics, viz.

$$S_n = \int_{[0,1]^d} \mathbb{C}_n(\mathbf{u})^2 \mathrm{d}C_n(\mathbf{u}) \quad \text{and} \quad T_n = \sup_{\mathbf{u}\in[0,1]^d} |\mathbb{C}_n(\mathbf{u})|. \tag{2}$$

Large values of these statistics lead to the rejection of H_0. Approximate P-values can be deduced from their limiting distributions, which depend on the asymptotic behavior of the process $\mathbb{C}_n$. Genest and Rémillard (in press) establish the convergence of the latter under appropriate regularity conditions on the parametric family $\mathcal{C}_0$ and the sequence (θ_n) of estimators. They also show that the tests based on S_n and T_n are consistent; i.e., if $C \notin \mathcal{C}_0$, then H_0 is rejected with probability 1 as $n \to \infty$.

In practice, the limiting distributions of S_n and T_n depend on the family of copulas under the composite null hypothesis, and on the unknown parameter value θ in particular. As a result, the asymptotic distribution of the test statistics cannot be tabulated and approximate P-values can only be obtained via specially adapted Monte Carlo methods. A specific parametric bootstrap procedure is described in Appendix A. Its validity is established by Genest and Rémillard (in press).

3.2. Two tests based on Kendall's transform

Another avenue successively explored by Genest and Rivest (1993), Wang and Wells (2000) and Genest et al. (2006) consists in basing a test of H_0 on a probability integral transformation of the data. The specific mapping they consider is

$$\mathbf{X} \mapsto V = H(\mathbf{X}) = C(\mathcal{U}_1, \ldots, \mathcal{U}_d),$$

where $\mathcal{U}_i = F_i(X_i)$ for $i \in \{1, \ldots, d\}$ and the joint distribution of $\boldsymbol{\mathcal{U}} = (\mathcal{U}_1, \ldots, \mathcal{U}_d)$ is C. This has come to be called Kendall's transform, because the expectation of V is an affine transformation of the multivariate version of Kendall's coefficient of concordance; see Barbe et al. (1996) or Jouini and Clemen (1996).

Let K denote the (univariate) distribution function of V. Genest and Rivest (1993) show that K can be estimated nonparametrically by the empirical distribution function of a rescaled version of the pseudo-observations $V_1 = C_n(\mathbf{U}_1), \ldots, V_n = C_n(\mathbf{U}_n)$. Barbe et al. (1996) give weak regularity conditions under which a central limit theorem can be proved for the slight variant

$$K_n(v) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(V_i \le v), \quad v \in [0, 1]. \tag{3}$$

In particular, the latter is a consistent estimator of the underlying distribution K.

Now under H_0, the vector $\boldsymbol{\mathcal{U}} = (\mathcal{U}_1, \ldots, \mathcal{U}_d)$ is distributed as C_θ for some $\theta \in \mathcal{O}$, and hence the Kendall transform $C_\theta(\boldsymbol{\mathcal{U}})$ has distribution K_θ. Through a measure of distance between K_n and a parametric estimation K_{θ_n} of K, one can test

$$H_0'' : K \in \mathcal{K}_0 = \{K_\theta : \theta \in \mathcal{O}\}.$$

Because $H_0 \subset H_0''$, of course, the nonrejection of H_0'' does not entail the acceptance of H_0. Consequently, tests based on the empirical process

$$\mathbb{K}_n = \sqrt{n}(K_n - K_{\theta_n})$$

are not generally consistent. Although they point out this limitation, Genest et al. (2006) investigate tests of H_0 based on this process. The idea had been put forward earlier (but not carried through) by Wang and Wells (2000) in the case of bivariate Archimedean copulas, for which H_0'' and H_0 are equivalent.

The specific statistics considered by Genest et al. (2006) are rank-based analogues of the Cramér–von Mises and Kolmogorov–Smirnov statistics, viz.

$$S_n^{(K)} = \int_0^1 \mathbb{K}_n(v)^2 \mathrm{d}K_{\theta_n}(v) \quad \text{and} \quad T_n^{(K)} = \sup_{v \in [0,1]} |\mathbb{K}_n(v)|. \tag{4}$$

Large values of either one of these statistics lead to the rejection of H_0''. Approximate P-values can be deduced from their limiting distributions, which depend on the asymptotic behavior of $\mathbb{K}_n$. The convergence of the latter is established by Genest et al. (2006) under appropriate regularity conditions on the parametric families $\mathcal{C}_0$, $\mathcal{K}_0$, and the sequence (θ_n) of estimators.

As the asymptotic distributions of $S_n^{(K)}$ and $T_n^{(K)}$ depend both on the unknown copula C_θ and on θ, approximate P-values for these statistics must again be found via simulation. See Appendix B for a parametric bootstrap procedure.

3.3. A test based on Rosenblatt's transform

Another well-known probability integral transformation on which goodness-of-fit tests could be based is due to Rosenblatt (1952). This mapping, which is commonly used for simulation, provides a simple way of decomposing a random vector with a given distribution into mutually independent components that are uniformly distributed on the unit interval. Its standard definition is recalled below for convenience.

Definition. Rosenblatt's probability integral transform of a copula C is the mapping $\mathcal{R} : (0,1)^d \to (0,1)^d$ which to every $\mathbf{u} = (u_1, \ldots, u_d) \in (0,1)^d$ assigns another vector $\mathcal{R}(\mathbf{u}) = (e_1, \ldots, e_d)$ with $e_1 = u_1$ and for each $i \in \{2, \ldots, d\}$,

$$e_i = \frac{\partial^{i-1} C(u_1, \ldots, u_i, 1, \ldots, 1)}{\partial u_1 \cdots \partial u_{i-1}} \Bigg/ \frac{\partial^{i-1} C(u_1, \ldots, u_{i-1}, 1, \ldots, 1)}{\partial u_1 \cdots \partial u_{i-1}}. \tag{5}$$

A critical property of Rosenblatt's transform is that $\boldsymbol{\mathcal{U}}$ is distributed as C, denoted $\boldsymbol{\mathcal{U}} \sim C$, if and only if the distribution of $\mathcal{R}(\boldsymbol{\mathcal{U}})$ is the d-variate independence copula

$$C_\perp(e_1, \ldots, e_d) = e_1 \times \cdots \times e_d, \quad e_1, \ldots, e_d \in [0,1].$$

Thus $H_0 : \boldsymbol{\mathcal{U}} \sim C \in \mathcal{C}_0$ is equivalent to $H_0^* : \mathcal{R}_\theta(\boldsymbol{\mathcal{U}}) \sim C_\perp$ for some $\theta \in \mathcal{O}$.

To test this hypothesis, therefore, one can use the fact that under H_0, the pseudo-observations $\mathbf{E}_1 = \mathcal{R}_{\theta_n}(\mathbf{U}_1), \ldots, \mathbf{E}_n = \mathcal{R}_{\theta_n}(\mathbf{U}_n)$ can be interpreted as a sample from the independence copula $C_\perp$. Of course, these pseudos are *not* mutually independent and only *approximately* uniform on $(0,1)^d$. Any inference procedure involving these constructs should thus take these features into account. This point is raised though eventually ignored by Breymann et al. (2003).

To describe the procedure of Breymann et al. (2003), let Φ denote the cumulative distribution function of a standard $\mathcal{N}(0,1)$ random variable and define

$$\chi_i = \sum_{j=1}^d \{\Phi^{-1}(E_{ij})\}^2, \quad i \in \{1, \ldots, n\}.$$

Exploiting the fact that $\mathbf{E}_1, \ldots, \mathbf{E}_n$ are "approximately" uniformly distributed over $(0,1)^d$, these authors argue that $\chi_1, \ldots, \chi_n$ can be interpreted as a sample from G, the distribution function of a chi-square random variable with d degrees of freedom. Now a natural estimate of G is the empirical distribution of the set $\chi_1, \ldots, \chi_n$, viz.

$$G_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\chi_i \le t), \quad t \ge 0. \tag{6}$$

For convenience, Breymann et al. (2003) assume that the empirical process $\mathbb{G}_n = \sqrt{n}(G_n - G)$ behaves asymptotically as if $\mathbf{E}_1, \ldots, \mathbf{E}_n$ were exactly uniform. They further suppose that the asymptotic distribution is independent of θ, and hence that it can be represented as $\beta \circ G$, where β is the standard Brownian bridge.

Should these assumptions hold true, Breymann et al. (2003) argue that it would then be possible to test H_0 with the Anderson–Darling statistic

$$A_n = -n - \frac{1}{n} \sum_{i=1}^n (2\mathrm{i} - 1)[\log\{G(\chi_{(i)})\} + \log\{1 - G(\chi_{(n+1-i)})\}], \tag{7}$$

where $\chi_{(1)} \le \cdots \le \chi_{(n)}$ are the order statistics corresponding to $\chi_1, \ldots, \chi_n$. The P-value would be simply given by reference to the limiting distribution of the original Anderson–Darling statistic; see e.g., Shorack and Wellner (1986).

As mentioned by Dobrić and Schmid (2007), however, the conclusions of Breymann et al. (2003) are too optimistic. Simulations show clearly that if the tabulated values of the Anderson–Darling statistic are used to perform their test, the resulting procedure has essentially no power and does not even maintain its nominal level.

To fix this problem, Dobrić and Schmid (2007) explain how the results of Genest and Rémillard (in press) could be exploited to compute reliable P-values for test statistics based on $\mathbb{G}_n$. In their paper, the Anderson–Darling test statistic A_n is used, together with the parametric bootstrap procedure described in Appendix C. Note, however, that the validity of the parametric bootstrap depends critically on the existence of a limiting distribution for A_n. The conditions (if any) under which this happens remain to be determined. Nevertheless, this test was included in the simulation study.

4. New procedures based on Rosenblatt's transform

One avenue not covered by Breymann et al. (2003) or Dobrić and Schmid (2007) consists in working directly with the process, using the full power of Rosenblatt's transform. The idea is not new, as it appeared in Klugman and Parsa (1999) for bivariate censored data. These authors propose a Pearson chi-square statistic computed from $\mathbf{E}_1, \ldots, \mathbf{E}_n$. However, their P-value calculation is incorrect, because it assumes wrongly that the limiting distribution is chi-square. The fact that the margins were estimated using parametric families is not taken into account in their work.

Under the null hypothesis H_0, the empirical distribution function

$$D_n(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(\mathbf{E}_i \le \mathbf{u}\right), \quad \mathbf{u} \in [0,1]^d \tag{8}$$

associated with the pseudo-observations $\mathbf{E}_1, \ldots, \mathbf{E}_n$ should be "close" to $C_\perp$. Thus, any reasonable notion of distance between D_n and $C_\perp$ is a good candidate for testing goodness-of-fit. Here, two Cramér–von Mises statistics are considered, namely

$$\begin{aligned} S_n^{(C)} &= n\int_{[0,1]^d} \{D_n(\mathbf{u}) - C_\perp(\mathbf{u})\}^2 \, \mathrm{d}D_n(\mathbf{u}) \\ &= \sum_{i=1}^{n} \{D_n(\mathbf{E}_i) - C_\perp(\mathbf{E}_i)\}^2 \end{aligned} \tag{9}$$

and

$$\begin{aligned} S_n^{(B)} &= n\int_{[0,1]^d} \{D_n(\mathbf{u}) - C_\perp(\mathbf{u})\}^2 \, \mathrm{d}\mathbf{u} \\ &= \frac{n}{3^d} - \frac{1}{2^{d-1}}\sum_{i=1}^{n}\prod_{k=1}^{d}\left(1 - E_{ik}^2\right) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\prod_{k=1}^{d}\left(1 - E_{ik} \vee E_{jk}\right), \end{aligned}$$

where $a \vee b = \max(a, b)$. These statistics only differ in their integration measure.

Using the tools described in the paper of Ghoudi and Rémillard (2004), one can easily determine the asymptotic null behavior of $\sqrt{n}\,(D_n - C_\perp)$ and, in turn, the convergence of $S_n^{(B)}$ and $S_n^{(C)}$. The limiting null distributions of these statistics are both unwieldy and, as in previous cases, they are functions both of the underlying copula and of its unknown parameter value θ. Nevertheless, goodness-of-fit testing is possible through the parametric bootstrap procedure described in Appendix D.

5. Experimental design

A large-scale Monte Carlo experiment was conducted to assess the finite-sample properties of the proposed goodness-of-fit tests for various choices of dependence structures and degrees of association. Two characteristics of the tests were of interest: their ability to maintain their nominal level, arbitrarily fixed at 5% throughout the study, and their power under a variety of alternatives.

To curtail the computational effort, comparisons were limited to the bivariate case and to three degrees of dependence, viz. $\tau = 0.25, 0.50, 0.75$. Seven one-parameter families of copulas were also considered, both under the null hypothesis and under the alternative. They fall into three categories:

(1) Three meta-elliptical copula families uniquely determined from the following classical bivariate distributions with correlation coefficient $\rho = \sin(\pi\tau/2)$:
   (a) the Gaussian distribution;
   (b) the Student distribution with $\nu = 4$ degrees of freedom;
   (c) the Pearson type II distribution with $\nu = 4$ degrees of freedom.
(2) Three of the most common Archimedean copula models, namely
   (a) the Clayton family, also known in the survival analysis literature as the gamma frailty model (Clayton, 1978; Cook and Johnson, 1981);
   (b) the Frank family (Nelsen, 1986; Genest, 1987);
   (c) the Gumbel–Hougaard family originally considered by Gumbel (1960) in the context of extreme-value theory.
(3) The Plackett family of copulas (Plackett, 1965).

The class of meta-elliptical copulas was introduced by Fang et al. (2002, 2005); its properties were examined by Frahm et al. (2003) and Abdous et al. (2005). These dependence structures are popular in actuarial science and in finance, where data often (but not always) exhibit heavy-tail dependence; see Malevergne and Sornette (2003), Cherubini et al. (2004) and McNeil et al. (2005) and references therein.

The Archimedean models are also commonly used in practice, particularly in survival analysis, because of their interpretation as mixture models and the natural extension they provide for Cox's proportional hazards model; see, e.g., Oakes (1989), Faraggi and Korn (1996) or Wang and Wells (2000). Refer also to Frees and Valdez (1998) and Klugman and Parsa (1999) for actuarial applications.

Finally, the Plackett system of distributions, which is neither Archimedean nor meta-elliptical, has found applications in biostatistics because of its constant cross-ratio property; see, e.g., Burzykowski et al. (2004). Dobrić and Schmid (2005), among others, investigated the relevance of this specific copula model in a financial context.

For every possible choice of copula and fixed value of τ, 10,000 random samples of size $n = 50$ were generated. An equal number of samples of size $n = 150$ was also obtained. Each of these samples was then used to test the goodness-of-fit of the seven families of distributions. Each of the following eight tests was applied in turn:

(1) The two tests derived by Genest and Rémillard (in press) from the empirical copula process, i.e., those based on the statistics S_n and T_n.
(2) The two tests developed by Genest et al. (2006) using Kendall's transform, i.e., those involving statistics $S_n^{(K)}$ and $T_n^{(K)}$.
(3) The test of Breymann et al. (2003) based on the statistic A_n and its corrected version developed by Dobrić and Schmid (2007), which both rely on Rosenblatt's transform.

(4) The two new procedures suggested in Section 4, i.e., those based on the statistics $S_n^{(B)}$ and $S_n^{(C)}$.

In all cases, the number of (primary-level) bootstrap samples was fixed at $N = 1000$. Whenever necessary, $m = 2500$ samples were drawn for the second-level bootstrap. This occurred when a closed-form expression was unavailable for the copula C_θ or the associated Kendall distribution K_θ. Two of the meta-elliptical copula models fall into this category on both accounts; for the Normal and the Plackett distributions, only K_θ needed to be estimated via a two-level parametric bootstrap.

Finally, whenever the parameter of a copula model had to be estimated, this was done by inversion of Kendall's tau. Given the sample version τ_n of τ, this involved solving for θ in the equation

$$4\int_{[0,1]^2} C_\theta(u_1, u_2)\mathrm{d}C_\theta(u_1, u_2) - 1 = \tau_n.$$

In all families considered, the solution is unique. See Nelsen (2006) for appropriate formulas and Genest and Rémillard (in press) for arguments showing that this method meets all the conditions required for the validity of the parametric bootstrap.

To sum up, the simulations were run according to a balanced experimental design involving the following factors:

$\mathcal{C}_0$: hypothesized copula model under H_0 (7 choices);
$\mathcal{C}$: copula model from which the data were generated (7 choices);
τ: level of dependence in $\mathcal{C}$, as measured by Kendall's tau (3 choices);
n: size of each sample drawn from $\mathcal{C}$ (2 choices).

In each of these $7 \times 7 \times 3 \times 2 = 294$ cases, 10,000 repetitions were performed in order to estimate the level or power of each of the eight tests under consideration.

This is by far the largest Monte Carlo experiment carried out to date in this area. The results presented below required the nearly exclusive use of 140 CPUs over a one-month period. By comparison, for example, the simulation study of Fermanian (2005) is limited to testing the goodness-of-fit of Frank's family; his alternative hypotheses are mixtures of the Frank and independence copula. The level and power of his tests are assessed from samples of size $n = 200$ for 30 combinations of the dependence and mixture parameters. With his choice of kernel and window, the tests turn out to be conservative.

In another study, Dobrić and Schmid (2007) look at the level and power of the statistic A_n using samples of size $n = 2500$ when the null hypothesis is either Normal, or Student with 3 degrees of freedom. The alternatives are either Clayton, Normal or Student, and three scenarios are considered for parameter estimation: (i) margins and dependence parameter ρ known; (ii) margins unknown but ρ known; (iii) margins and ρ both unknown. The study shows the effectiveness of the bootstrap algorithm, but no comparisons with alternative tests are included.

Recently, Scaillet (2007) also studied the performance of Fermanian's tests when the smoothing parameters are held fixed. His Monte Carlo simulations reveal that a bootstrap-based version of the tests performs well in samples of size $n = 50$ and 200. Mixtures of Frank, Normal and Student copulas were used. Comparisons with the tests based on $S_n^{(K)}$ and $T_n^{(K)}$ are also included but deemed inconclusive, as per the author's self-admission.

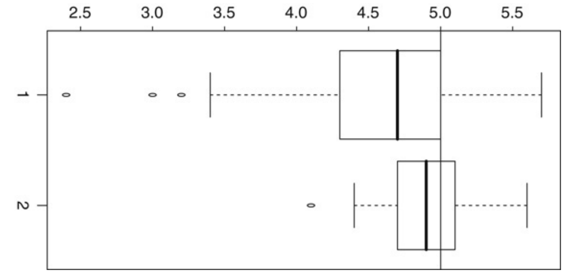

Fig. 1. Level of seven goodness-of-fit tests, as observed across $21 = 7 \times 3$ choices of $\mathcal{C}_0$ and τ. Top panel: $n = 50$; bottom panel: $n = 150$.

6. Results

Tables 1–3 report the level and power of the blanket tests from Sections 3 and 4. Each table corresponds to a specific combination of $\tau \in \{0.25, 0.50, 0.75\}$ and $n = 150$. Each line of a table shows the percentage of rejection of $H_0 : C \in \mathcal{C}_0$ associated with the different tests, given a choice of $\mathcal{C}_0$ and a true underlying copula family $\mathcal{C}$.

As an example, Table 1 shows that when testing for the Frank copula from a random sample of size $n = 150$, there are approximately 33.4% of chances that the test based on the Cramér–von Mises statistic S_n will reject the null hypothesis when the data are from the Gumbel–Hougaard copula with $\tau = 0.25$.

Due to space limitations, the corresponding tables for sample size $n = 50$ are not presented; they can be found in the doctoral dissertation of Beaudoin (2007). Note that the results for the test of Breymann et al. (2003) are omitted from all tables, given that the percentage of rejection of H_0 observed in the simulations was never higher than 1.5%. This portrays vividly the difficulties associated with an improper identification of the limiting distribution of a test statistic, as independently reported by Dobrić and Schmid (2007).

Because of the sheer amount of information in Tables 1–3, it is difficult to get a quick grasp of the relative performance of the tests in terms of level and power. To assist with the interpretation of the results, various aspects of the question are examined and illustrated with the help of box plots in the following subsections.

6.1. Level of the tests

Given that their finite-sample distribution is approximated by a parametric bootstrap procedure, the tests based on statistics

$$1 : S_n, \quad 3 : S_n^{(K)}, \quad 5 : S_n^{(B)} \quad 7 : A_n$$
$$2 : T_n, \quad 4 : T_n^{(K)}, \quad 6 : S_n^{(C)}$$

are expected to hold their nominal level. A cursory look at the figures highlighted in Tables 1–3 confirms that this happens in the vast majority of cases.

The same message is conveyed graphically by Fig. 1, where box plots show the dispersion in the levels observed across

Table 1
Percentage of rejection of H_0 by various tests for data sets of size $n = 150$ arising from different copula models with $\tau = 0.25$

Copula under H_0	True copula	Test based on						
		S_n	T_n	$S_n^{(K)}$	$T_n^{(K)}$	$S_n^{(B)}$	$S_n^{(C)}$	A_n
Clayton	Clayton	**4.6**	**4.8**	**4.1**	**4.5**	**4.7**	**4.8**	**4.9**
	Gumbel–Hougaard	86.1	62.4	57.9	42.7	80.9	76.7	22.4
	Frank	56.3	32.7	37.4	26.4	42.8	36.2	6.2
	Plackett	56.0	31.2	33.7	23.4	43.9	39.0	8.1
	Normal	50.2	27.5	24.5	16.8	41.8	34.6	6.4
	Student 4 dl	56.5	32.3	23.2	15.5	51.0	52.7	32.9
	Pearson 4 dl	49.9	28.7	26.1	17.3	43.3	32.9	6.4
Gumbel–Hougaard	Clayton	72.1	62.6	92.3	82.1	65.1	60.5	8.3
	Gumbel–Hougaard	**5.0**	**5.0**	**4.7**	**5.1**	**5.1**	**5.0**	**5.0**
	Frank	15.4	15.4	19.9	15.1	12.9	10.0	5.9
	Plackett	14.3	14.7	18.9	14.7	12.5	10.6	4.9
	Normal	10.1	11.7	24.4	18.9	10.2	7.5	5.9
	Student 4 dl	14.1	12.9	29.8	26.2	14.3	18.2	17.4
	Pearson 4 dl	10.2	12.6	23.6	18.5	12.8	7.8	11.3
Frank	Clayton	40.0	36.8	77.3	70.6	36.2	36.1	9.6
	Gumbel–Hougaard	33.4	18.5	9.1	6.1	27.8	29.5	12.4
	Frank	**5.3**	**5.1**	**5.1**	**5.0**	**4.9**	**4.9**	**5.1**
	Plackett	5.7	5.2	5.4	5.1	5.2	6.1	6.6
	Normal	7.8	7.3	10.5	9.9	6.2	6.3	5.3
	Student 4 dl	18.5	11.4	22.0	19.7	14.6	23.0	40.7
	Pearson 4 dl	6.5	7.3	7.7	7.6	6.5	5.2	7.0
Plackett	Clayton	37.6	34.2	69.8	60.5	33.3	31.9	6.2
	Gumbel–Hougaard	30.4	16.6	7.2	5.4	24.6	24.8	6.8
	Frank	5.0	5.2	4.8	5.1	5.0	4.2	6.5
	Plackett	**5.2**	**5.0**	**4.8**	**4.8**	**4.5**	**4.7**	**5.0**
	Normal	6.8	6.8	8.2	7.6	6.1	5.4	5.7
	Student 4 dl	14.1	9.8	15.6	14.4	10.1	15.6	26.2
	Pearson 4 dl	6.2	7.3	6.6	6.4	7.5	5.4	12.0
Normal	Clayton	31.6	26.6	56.9	45.8	33.3	33.0	7.2
	Gumbel–Hougaard	23.8	11.9	7.1	5.5	24.7	27.0	8.9
	Frank	7.9	7.2	5.6	5.3	7.2	7.0	5.5
	Plackett	7.9	6.8	4.4	4.4	8.2	9.4	6.0
	Normal	**5.1**	**5.0**	**4.7**	**5.2**	**4.7**	**5.0**	**4.8**
	Student 4 dl	10.5	6.8	7.4	7.4	16.6	27.8	29.9
	Pearson 4 dl	4.8	5.3	4.9	4.7	4.7	3.4	8.2
Student 4 dl	Clayton	27.7	26.2	52.1	39.0	25.1	17.4	11.2
	Gumbel–Hougaard	19.1	11.4	7.4	6.0	17.3	11.5	9.5
	Frank	9.1	8.2	9.5	7.6	8.9	4.5	23.3
	Plackett	7.7	7.7	7.3	6.2	6.6	3.6	13.9
	Normal	4.9	5.9	5.4	5.0	7.9	3.1	23.0
	Student 4 dl	**4.8**	**5.3**	**4.6**	**4.7**	**4.5**	**4.8**	**5.4**
	Pearson 4 dl	6.2	7.1	6.5	5.9	15.7	5.9	42.9
Pearson 4 dl	Clayton	35.7	29.3	60.0	50.9	40.5	44.8	16.9
	Gumbel–Hougaard	28.2	13.7	7.8	5.7	32.1	40.1	21.2
	Frank	9.0	7.0	4.9	4.7	10.4	12.4	7.6
	Plackett	9.2	7.2	4.2	4.1	13.2	17.8	12.4
	Normal	6.2	5.2	5.5	5.5	6.5	9.0	8.1
	Student 4 dl	16.4	8.6	10.2	9.3	28.0	47.3	52.6
	Pearson 4 dl	**5.2**	**5.1**	**5.0**	**4.8**	**4.9**	**4.7**	**5.3**

the seven tests and the $21 = 7 \times 3$ combinations of null hypothesis $\mathcal{C}_0$ and level of dependence τ. The data for $n = 50$ (from Beaudoin (2007)) and $n = 150$ (from Tables 1–3) are in the top and bottom panel, respectively.

In Fig. 1, the dimensions of each box are defined by the three quartiles of the empirical distribution of levels; outliers are indicated by open dots. The graphs show that overall, the parametric bootstrap algorithm does a very good job of approximating the null distribution of the various statistics. Except in a few cases, the performance is quite acceptable when $n = 50$. It is almost irreproachable when $n = 150$.

6.2. Effect of sample size

It is a classical fact of statistics that the power of a test increases with sample size. As Fig. 2 clearly shows, the present

Table 2
Percentage of rejection of H_0 by various tests for data sets of size $n = 150$ arising from different copula models with $\tau = 0.50$

Copula under H_0	True copula	Test based on						
		S_n	T_n	$S_n^{(K)}$	$T_n^{(K)}$	$S_n^{(B)}$	$S_n^{(C)}$	A_n
Clayton	Clayton	**5.3**	**5.0**	**4.5**	**4.5**	**5.1**	**5.0**	**5.0**
	Gumbel–Hougaard	99.9	98.3	98.5	91.4	99.7	99.5	78.3
	Frank	95.7	81.2	89.5	74.9	94.4	90.3	37.2
	Plackett	95.8	77.7	83.5	63.5	92.9	90.4	62.0
	Normal	93.7	74.1	75.1	53.7	89.0	85.5	35.2
	Student 4 dl	94.8	78.0	75.0	54.4	87.9	87.6	50.4
	Pearson 4 dl	94.0	74.3	75.8	55.0	91.9	88.0	31.9
Gumbel–Hougaard	Clayton	99.6	98.4	99.9	99.0	99.7	99.5	33.4
	Gumbel–Hougaard	**4.6**	**5.0**	**4.6**	**4.9**	**4.5**	**4.9**	**5.0**
	Frank	39.8	37.5	42.4	28.4	52.1	37.0	9.3
	Plackett	29.8	27.2	32.0	23.1	43.2	37.0	21.6
	Normal	18.3	21.1	37.7	27.4	33.7	25.2	4.9
	Student 4 dl	21.8	21.1	40.6	31.7	29.7	31.9	10.0
	Pearson 4 dl	18.1	21.7	36.6	26.2	41.2	28.9	4.0
Frank	Clayton	89.1	84.9	98.6	96.3	86.9	90.4	13.3
	Gumbel–Hougaard	63.0	39.6	28.3	15.8	44.1	57.6	9.2
	Frank	**4.8**	**5.1**	**4.8**	**5.2**	**4.8**	**4.8**	**5.1**
	Plackett	8.4	6.3	7.5	6.8	10.5	19.9	12.5
	Normal	19.9	15.0	22.6	17.3	8.9	14.4	4.8
	Student 4 dl	35.1	19.6	37.2	27.2	22.9	44.3	19.1
	Pearson 4 dl	15.0	13.0	17.3	13.1	7.1	8.9	5.8
Plackett	Clayton	83.9	78.4	95.5	86.4	79.6	78.0	12.5
	Gumbel–Hougaard	48.8	28.1	16.4	10.1	29.1	30.4	8.1
	Frank	6.8	7.8	8.2	8.0	10.2	3.9	10.5
	Plackett	**5.0**	**5.3**	**5.0**	**5.1**	**4.9**	**5.2**	**4.7**
	Normal	9.8	11.2	9.4	7.9	6.9	5.1	12.3
	Student 4 dl	15.1	11.4	15.1	10.6	7.4	11.7	7.4
	Pearson 4 dl	8.2	11.0	7.7	6.5	9.4	5.5	21.3
Normal	Clayton	80.0	68.8	90.3	75.2	90.8	88.2	7.8
	Gumbel–Hougaard	38.3	17.8	16.1	10.8	42.0	44.4	5.7
	Frank	20.2	14.3	17.4	14.1	13.4	8.5	8.7
	Plackett	13.2	9.7	6.8	6.6	18.0	22.7	18.1
	Normal	**4.9**	**5.0**	**4.9**	**5.2**	**5.0**	**5.3**	**4.8**
	Student 4 dl	8.2	5.3	5.9	5.2	20.4	32.1	8.8
	Pearson 4 dl	4.6	4.9	5.0	5.0	4.8	3.0	5.5
Student 4 dl	Clayton	77.3	70.5	90.6	73.2	84.9	74.9	6.0
	Gumbel–Hougaard	33.9	18.2	17.3	11.8	30.3	20.9	4.9
	Frank	26.9	18.9	29.3	20.7	24.2	8.1	6.0
	Plackett	13.8	11.0	11.6	9.5	10.2	6.9	10.4
	Normal	5.2	6.4	5.9	6.1	9.9	2.9	6.7
	Student 4 dl	**5.0**	**4.9**	**4.9**	**5.0**	**5.1**	**5.2**	**4.9**
	Pearson 4 dl	5.6	6.8	6.8	6.7	18.3	5.2	9.4
Pearson 4 dl	Clayton	81.8	69.6	91.2	76.3	92.9	92.8	11.2
	Gumbel–Hougaard	41.9	19.3	16.2	10.7	51.7	59.8	8.2
	Frank	18.9	13.5	13.9	11.7	14.5	12.8	11.4
	Plackett	13.9	9.7	5.6	5.7	28.8	37.5	23.7
	Normal	5.5	5.0	5.0	4.8	7.4	11.0	5.0
	Student 4 dl	10.9	6.2	6.7	5.7	34.3	52.4	14.1
	Pearson 4 dl	**4.7**	**4.7**	**4.7**	**4.8**	**4.7**	**4.9**	**4.8**

case is no exception. The box plots displayed there portray the variation in the ratio power ($n = 150$)/power ($n = 50$) for each of the seven tests, as observed across $126 = 7 \times 6 \times 3$ combinations of factors $\mathcal{C}_0$, $\mathcal{C}$ and τ, when the first two factors are different.

One can readily see from Fig. 2 that on average, the tests double their power as sample size goes from $n = 50$ to 150. In many instances, the improvement is more than four-fold but needless to say, it would quickly level off (to 1) as n keeps growing. However, there are also a few cases where no gain in power occurs. It is instructive to examine more carefully what happens in those extreme cases.

(1) What are the outliers identified in Fig. 2 and why is the increase in power so large in those cases?
   (a) Most outliers occur either at $\tau = 0.25$ or 0.75.
   (b) The statistics S_n, T_n and $S_n^{(B)}$ have very few outliers, if any.

Table 3
Percentage of rejection of H_0 by various tests for data sets of size $n = 150$ arising from different copula models with $\tau = 0.75$

Copula under H_0	True copula	Test based on						
		S_n	T_n	$S_n^{(K)}$	$T_n^{(K)}$	$S_n^{(B)}$	$S_n^{(C)}$	A_n
Clayton	Clayton	**5.4**	**5.0**	**4.9**	**5.1**	**5.1**	**5.2**	**5.0**
	Gumbel–Hougaard	99.9	99.9	99.9	98.7	99.9	99.9	49.1
	Frank	99.1	86.2	97.0	81.2	99.9	99.7	76.7
	Plackett	99.5	89.1	93.6	73.6	99.6	99.5	64.1
	Normal	99.8	91.7	94.9	77.7	99.5	99.6	23.8
	Student 4 dl	99.8	95.1	94.3	79.4	99.0	99.1	18.2
	Pearson 4 dl	99.7	90.7	95.1	77.1	99.7	99.7	29.8
Gumbel–Hougaard	Clayton	99.9	99.5	99.9	99.2	99.9	99.9	29.0
	Gumbel–Hougaard	**4.5**	**4.7**	**4.4**	**4.6**	**5.2**	**4.8**	**4.9**
	Frank	51.7	45.4	61.6	38.0	83.8	72.4	75.0
	Plackett	25.8	20.3	29.8	17.9	67.8	62.8	39.6
	Normal	12.3	17.0	29.4	18.6	60.7	53.6	5.9
	Student 4 dl	16.1	17.4	32.9	19.8	54.8	52.0	3.9
	Pearson 4 dl	11.8	18.6	30.1	19.6	66.9	58.7	6.5
Frank	Clayton	96.6	91.7	99.6	95.5	99.7	99.7	26.8
	Gumbel–Hougaard	81.9	43.6	53.2	27.1	59.9	74.2	40.0
	Frank	**4.7**	**4.7**	**4.5**	**4.7**	**5.0**	**5.1**	**5.2**
	Plackett	20.6	8.0	15.4	8.8	18.6	36.0	7.9
	Normal	40.9	21.2	40.2	20.5	18.4	30.1	49.8
	Student 4 dl	59.4	26.0	56.0	27.9	34.4	58.2	42.3
	Pearson 4 dl	34.2	21.0	34.5	18.0	15.0	22.3	54.1
Plackett	Clayton	89.8	86.8	97.7	78.6	99.5	99.1	18.8
	Gumbel–Hougaard	45.8	23.4	19.1	11.4	35.5	29.4	37.4
	Frank	14.9	15.4	18.5	15.3	9.7	3.6	10.9
	Plackett	**4.7**	**5.0**	**4.9**	**5.1**	**4.9**	**5.2**	**5.2**
	Normal	7.7	12.9	7.7	6.0	2.5	1.2	44.3
	Student 4 dl	11.0	12.3	11.4	6.7	4.3	3.6	45.2
	Pearson 4 dl	7.4	13.8	6.6	5.5	2.9	1.5	44.2
Normal	Clayton	91.8	82.4	97.3	75.4	99.9	99.9	8.2
	Gumbel–Hougaard	38.5	13.2	17.9	10.6	55.5	54.0	4.7
	Frank	42.2	22.9	41.4	24.6	32.8	20.1	70.2
	Plackett	16.5	7.6	7.0	7.0	23.0	30.6	30.0
	Normal	**4.9**	**4.4**	**4.4**	**4.8**	**4.9**	**4.6**	**5.1**
	Student 4 dl	6.6	4.3	4.9	4.5	12.3	18.3	4.9
	Pearson 4 dl	4.4	5.3	4.6	4.8	4.8	3.7	5.1
Student 4 dl	Clayton	90.6	86.6	97.7	78.6	99.9	99.7	10.9
	Gumbel–Hougaard	33.9	15.1	19.2	11.5	48.4	39.3	4.6
	Frank	48.2	30.5	53.9	32.4	39.3	20.3	81.8
	Plackett	15.7	8.9	11.0	9.7	16.4	17.2	43.5
	Normal	4.1	5.7	5.1	6.0	5.0	2.1	5.9
	Student 4 dl	**4.9**	**4.7**	**4.8**	**4.9**	**5.6**	**5.3**	**4.5**
	Pearson 4 dl	4.3	6.4	5.4	6.0	6.2	2.7	7.1
Pearson 4 dl	Clayton	93.0	81.4	97.4	75.1	99.9	99.9	7.3
	Gumbel–Hougaard	42.0	13.3	18.4	10.5	58.3	60.3	4.5
	Frank	41.2	21.0	37.4	22.8	28.4	20.2	63.4
	Plackett	17.5	7.3	6.5	6.2	26.7	37.3	25.8
	Normal	5.3	4.3	5.0	4.8	5.6	7.3	5.1
	Student 4 dl	8.3	4.3	5.4	4.3	18.3	28.9	5.2
	Pearson 4 dl	**4.5**	**4.8**	**4.6**	**4.7**	**4.8**	**5.0**	**4.7**

(c) The outliers at $\tau = 0.25$ are for $S_n^{(K)}$ and $T_n^{(K)}$, which prove particularly apt at detecting that data are not of the Clayton type as n increases.

(d) Most of the outliers at $\tau = 0.75$ are for $T_n^{(K)}$ and A_n; when $n = 150$, the first is much better as a goodness-of-fit test for the Frank copula, while the second can discriminate a Clayton dependence structure more easily.

(2) In what cases does one observe an increase in power of 10% or less (as identified by the vertical line crossing the box plots), and why?

(a) This phenomenon occurs mostly when $\tau = 0.25$ or 0.75, and twice as often in the former case than in the latter.

(b) This problem spares $S_n^{(K)}$ and $T_n^{(K)}$ and affects all others equally.

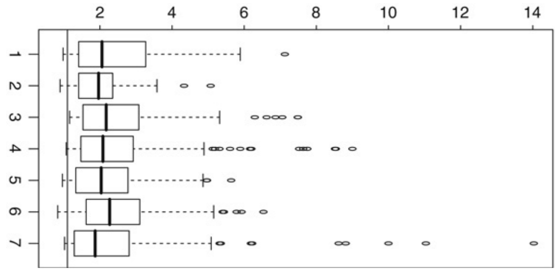

Fig. 2. Ratio power ($n = 150$)/power ($n = 50$) for seven goodness-of-fit tests, as observed across $126 = 7 \times 6 \times 3$ combinations of factors $\mathcal{C}_0$, $\mathcal{C}$ and τ for which $\mathcal{C}_0 \neq \mathcal{C}$. The vertical line is at 1.1, to help identify the cases where the improvement in power is less than 10% when n goes from 50 to 150.

(c) In half of the cases, the problem occurs because of a failure to distinguish between the Normal and the Pearson copula; most of the other instances of low increase in power occur when the null and the alternative are the Frank and Plackett copulas, or vice versa.

To illustrate the difficulties associated with the proper identification of a dependence structure from as small a sample as $n = 50$, Fig. 3 portrays typical scatter plots for the seven copula models considered in the study. For comparative purposes, $\tau = 0.5$ in all cases. The distinctive features of the models are hardly distinguishable and would be even fuzzier if one were to set $\tau = 0.25$.

In contrast, Fig. 4 displays scatter plots of random samples of size $n = 1000$ from the same copulas, again with $\tau = 0.5$. The characteristics of the different models are then much easier to pick out. The Clayton and the Gumbel–Hougaard are particularly easy to spot: their lower- and upper-tail dependences translate into greater densities of points in the lower-left and upper-right corners of the unit square, respectively.

While a trained eye could perhaps distinguish consistently between other pairs of copulas at $n = 1000$, some differences remain tenuous, e.g., between the Frank and the Plackett, or between the Normal and the Pearson copulas. Thus the fact that many of the power figures in Tables 1–3 are low does not come as a total surprise.

As shown by Genest et al. (2006), goodness-of-fit testing using $S_n^{(K)}$ and $T_n^{(K)}$ performs quite well when $n = 250$. As they point out, however, these tests are not generally consistent. In particular, they fail to discriminate bivariate extreme-value copulas having the same level of dependence, because for any such model, the theoretical Kendall distribution is $K(w) = w - (1 - \tau)w \log(w)$ for all $w \in (0, 1]$.

6.3. Which test performs best?

As might have been expected, no single test is preferable to all others, irrespective of the circumstances. It is clear from Tables 1–3 that the choice of the most powerful test depends

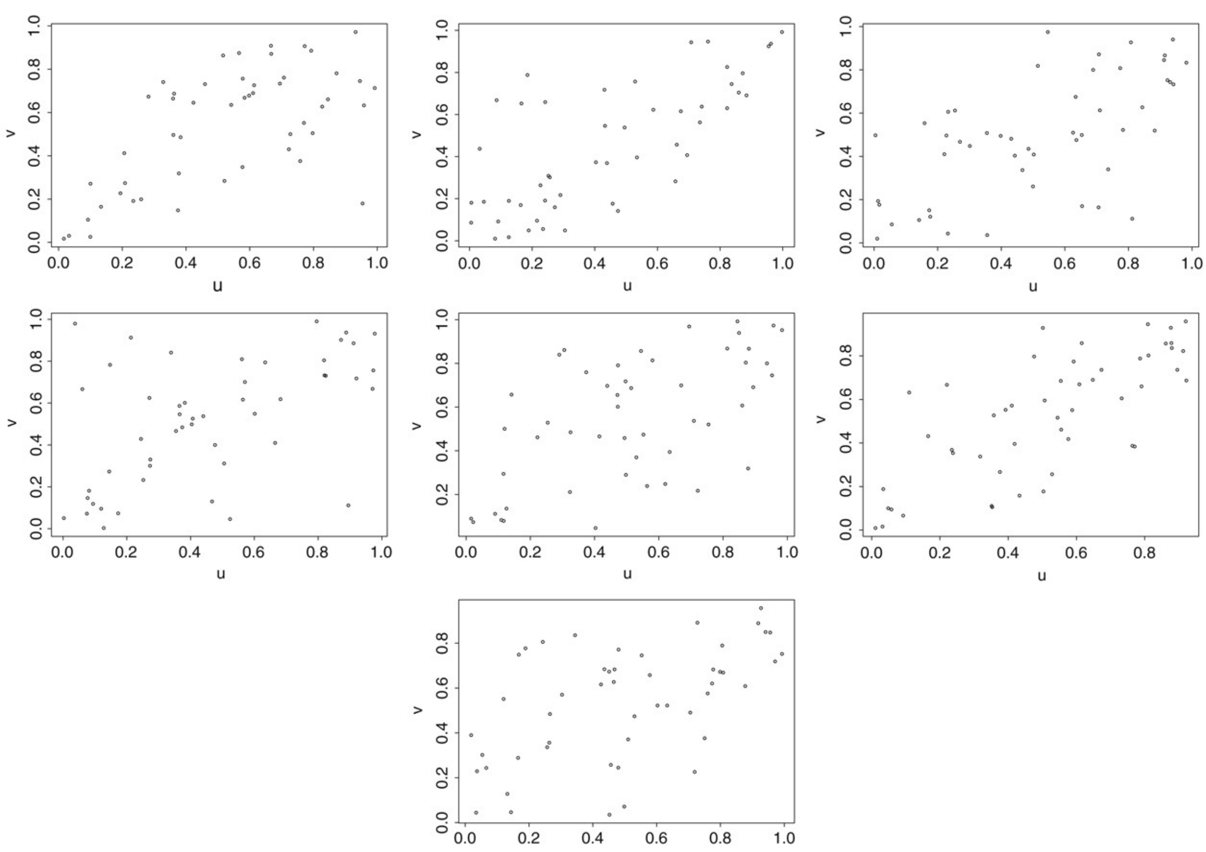

Fig. 3. Samples of size $n = 50$ from seven different copulas with parameter $\tau = 0.50$. From left to right, and top to bottom: Clayton, Gumbel–Hougaard, Frank, Plackett, Normal, Student and Pearson with 4 degrees of freedom.

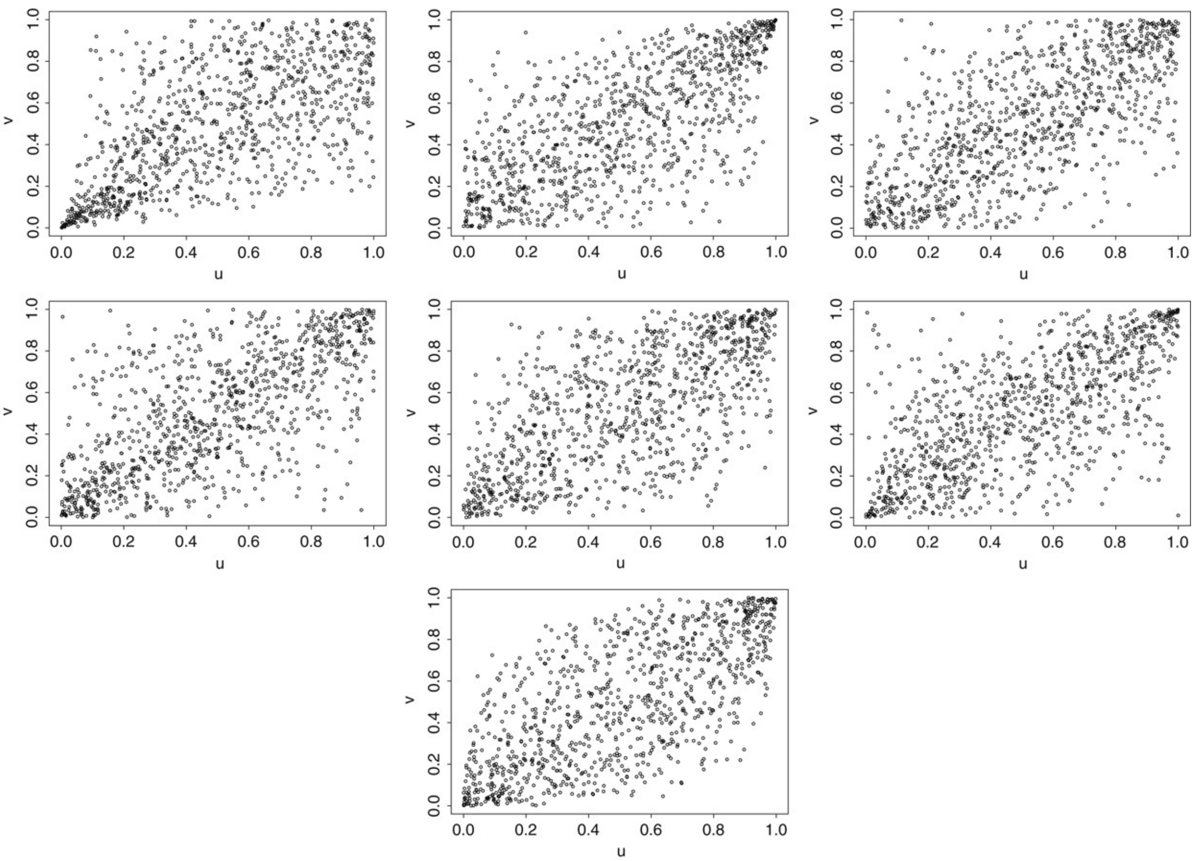

Fig. 4. Samples of size $n = 1000$ from seven different copulas with parameter $\tau = 0.50$. From left to right, and top to bottom: Clayton, Gumbel–Hougaard, Frank, Plackett, Normal, Student and Pearson with 4 degrees of freedom.

on the combination of factors τ, $\mathcal{C}$, and $\mathcal{C}_0$. From the additional tables reported in Beaudoin (2007), a dependence on n is also evident.

In practice, of course, only the last two factors are known for sure, i.e., the null hypothesis under investigation and the sample size. At the expense of mild "data snooping", one can get also a fairly good idea of the level of dependence in the data, as measured by Kendall's tau. Prior knowledge of the exact nature of dependence in the data, however, would defeat the purpose of goodness-of-fit testing.

In order to extract methodological recommendations from the mass of data contained in Tables 1–3, it is convenient to rank the tests from 1 to 7 in each of the $126 = 7 \times 6 \times 3$ experimental conditions corresponding to the seven possible choices of $\mathcal{C}_0$, the six alternatives $\mathcal{C}$, and the three values of tau. The sample size was fixed at $n = 150$ throughout, as those results are less subject to random variation and possibly more representative of situations one would encounter in practice.

Table 4 displays average ranks computed over the alternatives, for given $\mathcal{C}_0$ and τ. In the table, the best test is highlighted in each of the $21 = 7 \times 3$ scenarios considered. In Table 4, the tests are ranked from 1 to 7 in *increasing* order of power. Based on the number of times each test had the highest rank, it appears that:

(1) The best procedures overall are those based on $S_n^{(B)}$, S_n and $S_n^{(K)}$ with 6.5, 5 and 5 "wins", respectively.
(2) The tests based on $S_n^{(C)}$ and A_n are average with 2.5 and 2 wins, respectively.
(3) The performance of the tests involving T_n and $T_n^{(K)}$ is much less impressive, as they had no victory.

These observations are consistent with the common wisdom of the goodness-of-fit literature, to the effect that test statistics based on the Cramér–von Mises functional of a process tend to be more powerful than those based on the Kolmogorov–Smirnov distance taken on the same process.

A similar message is conveyed by the average ranks reported at the bottom of Table 4, which yield the following preference ranking:

$$S_n^{(B)} \succ S_n \succ S_n^{(K)} \succ S_n^{(C)} \succ T_n \succ A_n \succ T_n^{(K)}.$$

Although their differences may not be statistically significant, these means suggest that the tests based on T_n, A_n and $T_n^{(K)}$ are much less powerful than the others.

Other salient features of Table 4 are as follows:

(1) Among the tests based on a Cramér–von Mises statistic, there seems to be little to choose between a construction involving C_n, K_n or Rosenblatt's transform. Their averages are comparable, as are their respective number of wins (although $S_n^{(C)}$ had only 2.5 wins).
(2) The statistic S_n unequivocally yields the most powerful test of the Clayton hypothesis; it also does quite well for goodness-of-fit testing of Frank's model.

Table 4
Average ranking over factor $\mathcal{C}$ of the seven goodness-of-fit tests in $21 = 7 \times 3$ combinations of factors $\mathcal{C}_0$ and τ

H_0	τ	Test based on						
		S_n	T_n	$S_n^{(K)}$	$T_n^{(K)}$	$S_n^{(B)}$	$S_n^{(C)}$	A_n
Clayton	0.25	**7.0**	3.5	3.3	1.8	5.8	5.0	1.5
	0.50	**7.0**	3.2	3.8	2.0	6.0	5.0	1.0
	0.75	**5.9**	3.5	4.0	2.0	5.8	5.8	1.0
Gumbel–Hougaard	0.25	3.6	4.0	**7.0**	5.6	3.7	2.3	1.8
	0.50	3.5	2.8	**6.3**	3.5	6.2	4.7	1.0
	0.75	2.9	2.7	4.8	2.7	**6.8**	5.8	2.5
Frank	0.25	4.6	3.4	**5.3**	4.0	2.8	4.0	3.8
	0.50	5.3	2.8	**5.5**	4.0	3.3	5.2	1.8
	0.75	**5.8**	2.3	4.8	2.2	3.6	5.6	3.7
Plackett	0.25	4.2	4.1	**4.8**	4.2	3.8	2.8	4.3
	0.50	**4.9**	4.3	4.8	3.3	3.9	2.7	4.1
	0.75	4.8	5.0	4.6	2.8	3.3	2.3	**5.2**
Normal	0.25	4.7	3.8	3.7	2.4	**5.0**	4.8	3.8
	0.50	4.3	3.3	4.3	2.8	**5.0**	4.7	3.7
	0.75	4.3	3.2	3.7	2.5	**5.5**	4.8	4.1
Student 4 dl	0.25	4.6	4.4	4.5	2.6	4.7	1.9	**5.3**
	0.50	4.8	3.9	5.1	3.0	**5.7**	2.3	3.2
	0.75	3.7	3.3	4.2	3.3	**5.2**	3.5	4.8
Pearson 4 dl	0.25	4.2	2.2	3.1	2.3	5.3	**6.5**	4.5
	0.50	4.8	3.0	3.3	1.8	**6.2**	**6.2**	2.7
	0.75	4.8	2.3	3.8	1.9	5.8	**5.9**	3.5
Average		4.75	3.38	4.50	2.89	4.92	4.36	3.21
Standard error		1.04	0.75	0.99	0.96	1.15	1.45	1.39

(3) The test based on $S_n^{(B)}$ seems particularly good at detecting the lack of Normal or Student types of dependence, while $S_n^{(C)}$ is most powerful for the Pearson hypothesis; it would be interesting to see whether this conclusion extends to other meta-elliptical copula structures.
(4) Among the tests constructed using the Kendall transform, the procedure based on $S_n^{(K)}$ was far superior and offered the best performance when testing the goodness-of-fit of Gumbel–Hougaard and Frank copula structures.
(5) No clear recommendation emerges for goodness-of-fit testing of the Plackett.

7. Observations and recommendations

Based on the experience gained from carrying out this comparative power study of the existing blanket goodness-of-fit tests for copula models, the following general observations and specific recommendations can be made.

I. General observations:
   (a) In goodness-of-fit testing as in any other inferential context, the greater the sample size, the better. Large data sets not only help to distinguish between copula models but play a role in the reliability of the parametric bootstrap procedures used to approximate the statistics' null distribution.
   (b) In order for the double bootstrap to be efficient, the number m of repetitions must be substantially larger than the sample size n. In the present study, $m = 2500$ was found to be an acceptable minimum. While this is not a problem when using a test once, it quickly becomes computationally demanding in the context of a simulation study. In the present case, the recourse to a double bootstrap whenever C_θ or K_θ was not available in closed form made it totally impractical to run the experiment at a sample size of $n = 250$, for lack of sufficient computing resources.
   (c) In this regard, the tests based on A_n, $S_n^{(B)}$ and $S_n^{(C)}$ are at an advantage: because they rely on Rosenblatt's transform, a single bootstrap is enough to approximate their null distribution and extract P-values. However, the value of these statistics depends on the order in which the variables are successively conditioned. While it is traditional to take $U_2|U_1, U_3|(U_1, U_2), \ldots$ as in (5), any other sequence could be used. Different decisions could possibly ensue. (This point will need to be the object of future research.)
   (d) When statistics based on Cramér–von Mises and Kolmogorov–Smirnov functionals of the same empirical process are compared, the former are almost invariably more powerful. The present simulations and those reported earlier by Genest et al. (2006) both point strongly in that direction.

II. Specific recommendations, based on the present state of knowledge:
   (a) Overall, statistics S_n and $S_n^{(B)}$ yield the best blanket goodness-of-fit test procedures for copula models.

While $S_n^{(B)}$ is slightly more consistent than S_n in its performance across models and can be implemented without ever calling upon a double bootstrap, it relies on a nonunique (and therefore somewhat arbitrary) Rosenblatt transform.

(b) Statistics $S_n^{(C)}$ and $S_n^{(K)}$ are also recommendable, and the latter is especially convenient when the null hypothesis is Archimedean, since the Kendall distribution K is then available in closed form.

(c) The jury is still out on the merits of the test based on A_n. Anderson–Darling type statistics have proved useful in many other contexts, particularly in circumstances where differences in the tail of a distribution were deemed to be important. While it seems plausible that the same would hold in a copula context, the simulation results are not convincing in this regard. The asymptotic behavior of this statistic also remains to be studied.

(d) There are no strong arguments in favor of using the tests based on T_n or $T_n^{(K)}$. As for the uncorrected version of the test proposed by Breymann et al. (2003), it should never be used.

In future work, it would be interesting to investigate the sensitivity of tests based on the Rosenblatt transform to the order in which conditioning is done. It would also be useful to expand the present study to include comparisons with general goodness-of-fit tests involving tuning parameters, as well as with procedures developed to test for specific dependence structures such as the Clayton or the Normal copula.

On the theoretical front, several of the procedures that have been proposed recently for goodness-of-fit testing of copula models remain on shaky grounds. As illustrated by the appalling performance of the test proposed by Breymann et al. (2003), the dependence between pseudo-observations must imperatively be taken into account.

Nontrivial mathematics are required before one can conclude (or not) that the limiting distribution of a rank-based statistic is the same as in the classical multivariate context in which it was originally developed. Furthermore, conditions are required for the convergence of bootstrap algorithms, and failure to check them may lead to disaster. No sleight of hand will change that fact.

Acknowledgments

Partial funding in support of this work was granted by the Natural Sciences and Engineering Research Council of Canada, by the Fonds québécois de la recherche sur la nature et les technologies, and by the Institut de finance mathématique de Montréal. The authors gratefully acknowledge the GERAD (Montréal) and the Salle des marchés at Université Laval for extensive use of their computing facilities.

Appendix A. A parametric bootstrap for S_n and T_n

The following procedure leads to an approximate P-value for the test based on S_n. The adaptations required for T_n or any other rank-based statistic are obvious.

(1) Compute C_n as per formula (1) and estimate θ with $\theta_n = \mathcal{T}_n(\mathbf{U}_1, \ldots, \mathbf{U}_n)$.

(2) If there is an analytical expression for C_θ, compute the value of S_n, as defined in (2). Otherwise, proceed by Monte Carlo approximation. Specifically, choose $m \geq n$ and carry out the following extra steps:

(a) Generate a random sample $\mathbf{U}_1^*, \ldots, \mathbf{U}_m^*$ from distribution C_{θ_n}.

(b) Approximate C_{θ_n} by

$$B_m^*(\mathbf{u}) = \frac{1}{m}\sum_{i=1}^{m} \mathbf{1}\left(\mathbf{U}_i^* \leq \mathbf{u}\right), \quad \mathbf{u} \in [0,1]^d.$$

(c) Approximate S_n by

$$S_n = \sum_{i=1}^{n} \left\{C_n(\mathbf{U}_i) - B_m^*(\mathbf{U}_i)\right\}^2.$$

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

(a) Generate a random sample $\mathbf{Y}_{1,k}^*, \ldots, \mathbf{Y}_{n,k}^*$ from distribution C_{θ_n} and compute their associated rank vectors $\mathbf{R}_{1,k}^*, \ldots, \mathbf{R}_{n,k}^*$.

(b) Compute $\mathbf{U}_{i,k}^* = \mathbf{R}_{i,k}^*/(n+1)$ for $i \in \{1, \ldots, n\}$ and let

$$C_{n,k}^*(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(\mathbf{U}_{i,k}^* \leq \mathbf{u}\right), \quad \mathbf{u} \in [0,1]^d$$

and estimate θ by $\theta_{n,k}^* = \mathcal{T}_n(\mathbf{U}_{1,k}^*, \ldots, \mathbf{U}_{n,k}^*)$.

(c) If there is an analytical expression for C_θ, let

$$S_{n,k}^* = \sum_{i=1}^{n}\{C_{n,k}^*\left(\mathbf{U}_{i,k}^*\right) - C_{\theta_{n,k}^*}\left(\mathbf{U}_{i,k}^*\right)\}^2.$$

Otherwise, proceed as follows:

(i) Generate a random sample $\mathbf{Y}_{1,k}^{**}, \ldots, \mathbf{Y}_{m,k}^{**}$ from distribution $C_{\theta_{n,k}^*}$.

(ii) Approximate $C_{\theta_{n,k}^*}$ by

$$B_{m,k}^{**}(\mathbf{u}) = \frac{1}{m}\sum_{i=1}^{m} \mathbf{1}\left(\mathbf{Y}_{i,k}^{**} \leq \mathbf{u}\right), \quad \mathbf{u} \in [0,1]^d$$

and let

$$S_{n,k}^* = \sum_{i=1}^{n} \left\{C_{n,k}^*\left(\mathbf{U}_{i,k}^*\right) - B_{m,k}^{**}\left(\mathbf{U}_{i,k}^*\right)\right\}^2.$$

An approximate P-value for the test is then given by $\sum_{k=1}^{N} \mathbf{1}(S_{n,k}^* > S_n)/N$.

Appendix B. A parametric bootstrap for $S_n^{(K)}$ and $T_n^{(K)}$

For the sake of simplicity, the following algorithm is described in terms of statistic $S_n^{(K)}$. However, it is also valid *mutatis mutandis* for $T_n^{(K)}$ or any other rank-based statistic.

(1) Compute K_n as per formula (3) and estimate θ with $\theta_n = \mathcal{T}_n(\mathbf{U}_1, \ldots, \mathbf{U}_n)$.

(2) If there is an analytical expression for K_θ, compute the value of $S_n^{(K)}$, as defined in (4). Otherwise, proceed by Monte Carlo approximation. Specifically, choose $m \geq n$ and carry out the following extra steps:

(a) Generate a random sample $\mathbf{U}_1^*, \ldots, \mathbf{U}_m^*$ from distribution C_{θ_n}.

(b) Approximate K_{θ_n} by

$$B_m^*(t) = \frac{1}{m}\sum_{i=1}^{m} \mathbf{1}\left(V_i^* \le t\right), \quad t \in [0,1],$$

where

$$V_i^* = \frac{1}{m}\sum_{j=1}^{m} \mathbf{1}\left(\mathbf{U}_j^* \le \mathbf{U}_i^*\right), \quad i \in \{1,\ldots,m\}.$$

(c) Approximate $S_n^{(K)}$ by

$$S_n^{(K)} = \frac{n}{m}\sum_{i=1}^{m} \left\{K_n\left(V_i^*\right) - B_m^*\left(V_i^*\right)\right\}^2.$$

Note in passing that $m \times B_m^*(V_i^*)$ is the rank of V_i^* among $V_1^*, \ldots, V_m^*$.

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

(a) Generate a random sample $\mathbf{Y}_{1,k}^*, \ldots, \mathbf{Y}_{n,k}^*$ from distribution C_{θ_n} and compute their associated rank vectors $\mathbf{R}_{1,k}^*, \ldots, \mathbf{R}_{n,k}^*$.

(b) Compute

$$V_{i,k}^* = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}\left(\mathbf{Y}_{j,k}^* \le \mathbf{Y}_{i,k}^*\right), \quad i \in \{1,\ldots,n\}$$

$$K_{n,k}^*(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(V_{i,k}^* \le t\right), \quad t \in [0,1]$$

and estimate θ by $\theta_{n,k}^* = \mathcal{T}_n\{\mathbf{R}_{1,k}^*/(n+1), \ldots, \mathbf{R}_{n,k}^*/(n+1)\}$.

(c) If there is an analytical expression for K_θ, let

$$S_{n,k}^{(K)*} = \int_0^1 \{C_{n,k}^*(t) - K_{\theta_{n,k}^*}(t)\}^2 \mathrm{d}K_{\theta_{n,k}^*}(t),$$

for which an explicit expression can easily be deduced from (4). Otherwise, proceed as follows:

(i) Generate a random sample $\mathbf{Y}_{1,k}^{**}, \ldots, \mathbf{Y}_{m,k}^{**}$ from distribution $C_{\theta_{n,k}^*}$.

(ii) Approximate $K_{\theta_{n,k}}^*$ by

$$B_{m,k}^{**}(t) = \frac{1}{m}\sum_{i=1}^{m} \mathbf{1}\left(V_{i,k}^{**} \le t\right), \quad t \in [0,1],$$

where

$$V_{i,k}^{**} = \frac{1}{m}\sum_{j=1}^{m} \mathbf{1}\left(\mathbf{Y}_{j,k}^* \le \mathbf{Y}_{i,k}^*\right), \quad i \in \{1,\ldots,m\}.$$

Then set

$$S_{n,k}^{(K)*} = \frac{n}{m}\sum_{i=1}^{m} \left\{K_{n,k}^*\left(V_{i,k}^*\right) - B_{m,k}^{**}\left(V_{i,k}^*\right)\right\}^2,$$

where $m \times B_{m,k}^{**}(V_{i,k}^*)$ is the rank of $V_{i,k}^*$ among $V_{1,k}^*, \ldots, V_{m,k}^*$.

An approximate P-value for the test is then given by $\sum_{k=1}^{N} \mathbf{1}(S_{n,k}^{(K)*} > S_n^{(K)})/N$.

Appendix C. A parametric bootstrap for A_n

Although the following algorithm is described in terms of statistic A_n, it is also valid *mutatis mutandis* for any other rank-based statistic based on $\chi_1, \ldots, \chi_n$.

(1) Compute G_n as per formula (6) and estimate θ with $\theta_n = \mathcal{T}_n(\mathbf{U}_1, \ldots, \mathbf{U}_n)$.

(2) Compute the value of A_n as per formula (7).

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

(a) Generate a random sample $\mathbf{Y}_{1,k}^*, \ldots, \mathbf{Y}_{n,k}^*$ from distribution C_{θ_n} and compute their associated rank vectors $\mathbf{R}_{1,k}^*, \ldots, \mathbf{R}_{n,k}^*$.

(b) Compute $\mathbf{U}_{i,k}^* = \mathbf{R}_{i,k}^*/(n+1)$ for $i \in \{1,\ldots,n\}$.

(c) Estimate θ with $\theta_{n,k}^* = \mathcal{T}_n(\mathbf{U}_{1,k}^*, \ldots, \mathbf{U}_{n,k}^*)$, and compute $\chi_{1,k}^*, \ldots, \chi_{n,k}^*$, where

$$\chi_{i,k}^* = \sum_{j=1}^{d}\left\{\Phi^{-1}(E_{ij,k}^*)\right\}^2 \quad \text{and} \quad \mathbf{E}_{i,k}^* = \mathcal{R}_{\theta_{n,k}^*}\left(\mathbf{U}_{i,k}^*\right),$$

$i \in \{1,\ldots,n\}$.

(d) Let

$$G_{n,k}^*(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(\chi_{i,k}^* \le t\right), \quad t \ge 0$$

and define

$$A_{n,k}^* = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\cdot[\log\{G(\chi_{(i),k}^*)\} + \log\{1 - G(\chi_{(n+1-i),k}^*)\}].$$

An approximate P-value for the test is then given by $\sum_{k=1}^{N} \mathbf{1}(A_{n,k}^* > A_n)/N$.

Appendix D. A parametric bootstrap for $S_n^{(C)}$ and $S_n^{(B)}$

The following algorithm is described in terms of statistic $S_n^{(C)}$. However, it is also valid *mutatis mutandis* for $S_n^{(B)}$ or any other rank-based statistic.

(1) Compute D_n as per formula (8) and estimate θ by $\theta_n = \mathcal{T}_n(\mathbf{U}_1, \ldots, \mathbf{U}_n)$.

(2) Compute the value of $S_n^{(C)}$, as defined in (9).

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

(a) Generate a random sample $\mathbf{Y}_{1,k}^*, \ldots, \mathbf{Y}_{n,k}^*$ from distribution C_{θ_n} and compute their associated rank vectors $\mathbf{R}_{1,k}^*, \ldots, \mathbf{R}_{n,k}^*$.

(b) Compute $\mathbf{U}_{i,k}^* = \mathbf{R}_{i,k}^*/(n+1)$ for $i \in \{1,\ldots,n\}$.

(c) Estimate θ by $\theta_{n,k}^* = \mathcal{T}_n(\mathbf{U}_{1,k}^*, \ldots, \mathbf{U}_{n,k}^*)$ and compute $\mathbf{E}_{1,k}^*, \ldots, \mathbf{E}_{n,k}^*$, where

$$\mathbf{E}_{i,k}^* = \mathcal{R}_{\theta_{n,k}^*}\left(\mathbf{U}_{i,k}^*\right), \quad i \in \{i,\ldots,n\}.$$

(d) Let

$$D_{n,k}^*(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(\mathbf{E}_{i,k}^* \le \mathbf{u}\right), \quad \mathbf{u} \in [0,1]^d$$

and set

$$S_{n,k}^{(C)*} = \sum_{i=1}^{n} \left\{D_{n,k}^*\left(\mathbf{E}_{i,k}^*\right) - C_\perp\left(\mathbf{E}_{i,k}^*\right)\right\}^2.$$

An approximate P-value for the test is then given by $\sum_{k=1}^{N} \mathbf{1}(S_{n,k}^{(C)*} > S_n^{(C)})/N$.

References

Abdous, B., Genest, C., Rémillard, B., 2005. Dependence properties of meta-elliptical distributions. In: Statistical Modeling and Analysis for Complex Data Problems. In: GERAD 25th Anniv. Ser., vol. 1. Springer, New York, pp. 1–15.

Andersen, P.K., Ekstrøm, C.T., Klein, J.P., Shu, Y., Zhang, M.-J., 2005. A class of goodness of fit tests for a copula based on bivariate right-censored data. Biometrical Journal 47, 815–824.

Ané, T., Kharoubi, C., 2003. Dependence structure and risk measure. Journal of Business 76, 411–438.

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

Beaudoin, D., 2007. Estimation de la dépendance et choix de modèles pour des données bivariées sujettes à censure et à troncation. PhD thesis, Université Laval, Québec, Canada.

Berg, D., Bakken, H., 2005. A goodness-of-fit test for copulae based on the probability integral transform. Technical Report SAMBA/41/05, Norsk Regnesentral, Oslo, Norway.

Breymann, W., Dias, A., Embrechts, P., 2003. Dependence structures for multivariate high-frequency data in finance. Quantitative Finance 3, 1–14.

Burzykowski, T., Molenberghs, G., Buyse, M., 2004. The validation of surrogate end points by using data from randomized clinical trials: A case-study in advanced colorectal cancer. Journal of the Royal Statistical Society Series A 167, 103–124.

Chen, X., Fan, Y., Tsyrennikov, V., 2006. Efficient estimation of semiparametric multivariate copula models. Journal of the American Statistical Association 101, 1228–1240.

Cherubini, U., Luciano, E., Vecchiato, W., 2004. Copula Methods in Finance. Wiley, New York.

Clayton, D.G., 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.

Cook, R.D., Johnson, M.E., 1981. A family of distributions for modelling nonelliptically symmetric multivariate data. Journal of the Royal Statistical Society Series B 43, 210–218.

Cui, S., Sun, Y., 2004. Checking for the gamma frailty distribution under the marginal proportional hazards frailty model. Statistica Sinica 14, 249–267.

Deheuvels, P., 1979. La fonction de dépendance empirique et ses propriétés: Un test non paramétrique d'indépendance. Académie Royale de Belgique. Bulletin de la Classe des Sciences, 5e Série 65, 274–292.

Dobrić, J., Schmid, F., 2005. Testing goodness of fit for parametric families of copulas: Application to financial data. Communications in Statistics. Simulation and Computation 34, 1053–1068.

Dobrić, J., Schmid, F., 2007. A goodness of fit test for copulas based on Rosenblatt's transformation. Computational Statistics & Data Analysis 51, 4633–4642.

Fang, H.-B., Fang, K.-T., Kotz, S., 2002. The meta-elliptical distributions with given marginals. Journal of Multivariate Analysis 82, 1–16.

Fang, H.-B., Fang, K.-T., Kotz, S., 2005. Corrigendum to: "The meta-elliptical distributions with given marginals" [J. Multivariate Anal. 82: 1–16 (2002)]. Journal of Multivariate Analysis 94, 222–223.

Faraggi, D., Korn, E.L., 1996. Competing risks with frailty models when treatment affects only one failure type. Biometrika 83, 467–471.

Fermanian, J.-D., 2005. Goodness-of-fit tests for copulas. Journal of Multivariate Analysis 95, 119–152.

Fermanian, J.-D., Radulović, D., Wegkamp, M.H., 2004. Weak convergence of empirical copula processes. Bernoulli 10, 847–860.

Frahm, G., Junker, M., Szimayer, A., 2003. Elliptical copulas: Applicability and limitations. Statistics & Probability Letters 63, 275–286.

Frees, E.W., Valdez, E.A., 1998. Understanding relationships using copulas. North American Actuarial Journal 2, 1–25.

Gänßler, P., Stute, W., 1987. Seminar on Empirical Processes. Birkhäuser Verlag, Basel, Switzerland.

Genest, C., 1987. Frank's family of bivariate distributions. Biometrika 74, 549–555.

Genest, C., Favre, A.-C., 2007. Everything you always wanted to know about copula modeling but were afraid to ask. Journal of Hydrologic Engineering 12, 347–368.

Genest, C., Ghoudi, K., Rivest, L.-P., 1995. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika 82, 543–552.

Genest, C., Quessy, J.-F., Rémillard, B., 2006. Goodness-of-fit procedures for copula models based on the integral probability transformation. Scandinavian Journal of Statistics 33, 337–366.

Genest, C., Rémillard, B., 2008. Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models, Annales de I'Institut Henri Poincaré. Probabilités et Statistiques 44 (in press).

Genest, C., Rivest, L.-P., 1993. Statistical inference procedures for bivariate Archimedean copulas. Journal of the American Statistical Association 88, 1034–1043.

Genest, C., Werker, B.J.M., 2002. Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models. In: Distributions with Given Marginals and Statistical Modelling. Kluwer, Dordrecht, The Netherlands, pp. 103–112.

Ghoudi, K., Rémillard, B., 2004. Empirical processes based on pseudo-observations. II. The multivariate case. In: Asymptotic Methods in Stochastics. In: Fields Inst. Commun., vol. 44. Amer. Math. Soc., Providence, RI, pp. 381–406.

Glidden, D.V., 1999. Checking the adequacy of the gamma frailty model for multivariate failure times. Biometrika 86, 381–393.

Gumbel, E.J., 1960. Distributions des valeurs extrêmes en plusieurs dimensions. Publications de l'Institut de Statistique de l'Université de Paris 9, 171–173.

Joe, H., 1997. Multivariate Models and Dependence Concepts. Chapman & Hall, London.

Joe, H., 2005. Asymptotic efficiency of the two-stage estimation method for copula-based models. Journal of Multivariate Analysis 94, 401–419.

Jouini, M.N., Clemen, R.T., 1996. Copula models for aggregating expert opinions. Operations Research 44, 444–457.

Junker, M., May, A., 2005. Measurement of aggregate risk with copulas. The Econometrics Journal 8, 428–454.

Kim, G., Silvapulle, M.J., Silvapulle, P., 2007. Comparison of semiparametric and parametric methods for estimating copulas. Communications in Statistics. Simulation and Computation 51, 2836–2850.

Klaassen, C.A.J., Wellner, J.A., 1997. Efficient estimation in the bivariate normal copula model: Normal margins are least favourable. Bernoulli 3, 55–77.

Klugman, S., Parsa, R., 1999. Fitting bivariate loss distributions with copulas. Insurance: Mathematics & Economics 24, 139–148.

Malevergne, Y., Sornette, D., 2003. Testing the Gaussian copula hypothesis for financial assets dependences. Quantitative Finance 3, 231–250.

McNeil, A.J., Frey, R., Embrechts, P., 2005. Quantitative Risk Management. Princeton University Press, Princeton, NJ.

Nelsen, R.B., 1986. Properties of a one-parameter family of bivariate distributions with specified marginals. Communications in Statistics A. Theory and Methods 15, 3277–3285.

Nelsen, R.B., 2006. An Introduction to Copulas, second ed. Springer, New York.

Oakes, D., 1989. Bivariate survival models induced by frailties. Journal of the American Statistical Association 84, 487–493.

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

Plackett, R.L., 1965. A class of bivariate distributions. Journal of the American Statistical Association 60, 516–522.

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

Scaillet, O., 2007. Kernel based goodness-of-fit tests for copulas with fixed smoothing parameters. Journal of Multivariate Analysis 98, 533–543.

Shih, J.H., 1998. A goodness-of-fit test for association in a bivariate survival model. Biometrika 85, 189–200.

Shih, J.H., Louis, T.A., 1995. Inferences on the association parameter in copula models for bivariate survival data. Biometrics 51, 1384–1399.

Shorack, G.R., Wellner, J.A., 1986. Empirical Processes with Applications to Statistics. Wiley, New York.

Tsukahara, H., 2005. Semiparametric estimation in copula models. The Canadian Journal of Statistics 33, 357–375.

Wang, W., Wells, M.T., 2000. Model selection and semiparametric inference for bivariate failure-time data. Journal of the American Statistical Association 95, 62–72.