Journal of Multivariate Analysis 138 (2015) 34–52

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Efficient information based goodness-of-fit tests for vine copula models with fixed margins: A comprehensive review

Ulf Schepsmeier

Center for Mathematical Sciences, Technische Universität München, Boltzmannstraße 3, 85748 Garching b. München, Germany

HIGHLIGHTS

- We present a new goodness-of-fit test for regular vine copula models.
- The test arises from the information matrix ratio.
- The test's power is investigated and compared to 14 other goodness-of-fit tests.
- A simulation study shows the excellent performance with respect to size and power.
- An application to stock indices and their related volatility indices is performed.

ARTICLE INFO

Article history:
Received 25 April 2014
Available online 14 January 2015

AMS subject classifications:
62H15
62F03
62F05
62P05

Keywords:
Copula
Goodness-of-fit tests
Information matrix ratio test
Power comparison
R-vine

ABSTRACT

We introduce a new goodness-of-fit test for regular vine (R-vine) copula models, a flexible class of multivariate copulas based on a pair-copula construction (PCC). The test arises from the information matrix ratio and assumes fixed margins. The corresponding test statistic is derived and its asymptotic normality is shown. The test's power is investigated and compared to 14 other goodness-of-fit tests, adapted from the bivariate copula case, in a high-dimensional setting. The extensive simulation study on the copula level shows the excellent performance with respect to size and power as well as the superiority of the information matrix ratio based test against most other goodness-of-fit tests. The best performing tests are applied to a portfolio of stock indices and their related volatility indices validating different R-vine specifications.

1. Introduction

Analyzing complex correlated data has received considerable attention in the current statistical literature. Among many approaches to modeling correlation structures, copula based models offer a powerful and flexible toolbox to characterize dependence profiles among variables, which have been studied extensively. Unfortunately, there is little progress known in the theory and method concerning a goodness-of-fit (GOF) test, an important aspect of statistical model diagnostics. In fact, most of the published work has been only focused on bivariate copula models (see for example the comparison study of [20]).

Copulas join marginal distributions $F_1, \ldots, F_d$ of a (continuous) random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ with their dependency structure by a joint cumulative distribution function (cdf) $H(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$. Here C is the unique cdf with uniform margins on the unit hypercube [38]. Classical copula classes such as the elliptical or Archimedean copulas are

E-mail address: ulf.schepsmeier@tum.de.

http://dx.doi.org/10.1016/j.jmva.2015.01.001
0047-259X/

very limited with respect to flexibility in higher dimensions. But they are very powerful and well understood in the bivariate case. Thus, Joe [26] and later Bedford and Cooke [3,4] independently constructed multivariate densities using $d(d-1)/2$ bivariate copulas. They permit flexibility and feasibility of constructing and computing a relatively large dimensional copula model. In Aas et al. [1] this process is termed a pair-copula construction (PCC) and the statistical inference is developed for it. Since then the theory of vine copulas arising from the PCC was studied in the literature (see for example [12,39,13,14]).

Along with the breakthrough of vine copula constructions, model diagnosis becomes ever so imperative in the application of multi-dimensional vine copulas. Developing efficient GOF tests is now a timely task as already noted in Fermanian [16], and an important addition to the current literature of vine copulas. In addition, comprehensive comparisons for many of the classical GOF tests are lacking in terms of their relative merits when they are applied to multi-dimensional copulas. So far model verification methods for vine copulas are usually based on the likelihood, or on the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) as classical comparison measures, which take the model complexity into account.

In our goodness-of-fit (GOF) tests we would like to test

$$H_0 : C \in \mathcal{C}_0 = \{C_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\} \quad \text{against} \quad H_1 : C \notin \mathcal{C}_0 = \{C_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}, \tag{1}$$

where C denotes the vine copula cdf with known margins and $\mathcal{C}_0$ is a class of parametric vine copulas with $\Theta \subseteq \mathbb{R}^p$ being the parameter space of dimension p.

Many GOF tests were studied in the literature (see for example [18,20,5,25]). To our knowledge they however were only applied to elliptical and Archimedean copulas. However, a GOF test for vine copula models verifying the chosen pair-copula families has, to our knowledge, only be treated in Schepsmeier [35]. Although, already Aas et al. [1] suggested a GOF test for vine copulas based on the multivariate probability integral transformation (PIT) of Rosenblatt [33] given in the Appendix A, but never investigated its small sample performance. We will show that this test and many other copula GOF tests have little to no power in the high-dimensional setting of a vine and thus are not appropriate to be utilized there.

The main contribution of this paper is a new GOF test to perform model verification of vine copula models using hypothesis tests. As in Schepsmeier [35] it is based on the Bartlett identity ($-\mathbb{H}(\boldsymbol{\theta}) = \mathbb{C}(\boldsymbol{\theta})$) as generally suggested by White [41] and assumes fixed margins. Here $\mathbb{H}(\boldsymbol{\theta})$ is the expected Hessian or variability matrix, and $\mathbb{C}(\boldsymbol{\theta})$ is the expected outer product of the gradient or sensitivity matrix. In contrast to the White test, which relies on the difference between $-\mathbb{H}(\boldsymbol{\theta})$ and $\mathbb{C}(\boldsymbol{\theta})$, our new test is based on the information matrix ratio (IMR), $\Psi(\boldsymbol{\theta}) := -\mathbb{C}(\boldsymbol{\theta})^{-1}\mathbb{H}(\boldsymbol{\theta})$ [43].

First, the IMR based test statistic for vine models will be derived and its asymptotic normality under the Bartlett identity will be shown. Secondly, the small sample performance for size and power will be investigated and compared to 14 other GOF tests for vines in a high-dimensional setting ($d = 5$ and $d = 8$). The general assumption for all these tests as well as for the IMR based test is that the margins are known. In particular, we will compare to GOF tests based on the

- difference of Bartlett identity or
- difference between empirical and estimated copula, i.e. $\hat{C}_n(\boldsymbol{v}) - C_{\hat{\boldsymbol{\theta}}_n}(\boldsymbol{v})$, with $\boldsymbol{v} = (v_1, \ldots, v_d) \in [0, 1]^d$

$$\hat{C}_n(\boldsymbol{v}) = \frac{1}{n+1} \sum_{t=1}^{n} \mathbf{1}_{\{u_{t1} \le v_1, \ldots, u_{td} \le v_d\}}, \tag{2}$$

  and $C_{\boldsymbol{\theta}}(\boldsymbol{u})$ being the copula with estimated parameter(s) $\hat{\boldsymbol{\theta}}_n$. Here, $(u_{t1}, \ldots, u_{td})$ are the so-called copula data, i.e. data transformed to the hypercube after fitting univariate margins.
- Further, we will compare to GOF tests based on the multivariate probability integral transform (PIT).

The approach based on difference between empirical and estimated copula is closely related to the empirical copula process approach by Genest and Rémillard [19]. Since we work with fixed margins the empirical copula is based on copula data rather than ranks as original defined in e.g. Genest and Rémillard [19]. The tests based on the multivariate PIT aggregation from multivariate transformed data to univariate test data are facilitated using different aggregation functions. For the resulting univariate test data then standard univariate GOF test statistics such as Anderson–Darling (AD), Cramér–von Mises (CvM) and Kolmogorov–Smirnov (KS) are used to test for uniformity. In contrast, the empirical copula process (ECP) based test use the multivariate approaches of the Cramér–von Mises (mCvM) and Kolmogorov–Smirnov (mKS) test statistics. The different GOF tests are given in the Appendix B for the convenience of the reader.

The power study will expose that the information based GOF tests such as the information matrix difference approach of Schepsmeier [35] and in particular our new IMR based test outperform the other GOF tests in terms of size and power. The PIT based GOF tests reveal little to no power against the considered alternatives. But applying the PIT transformed data to the empirical copula process, as first suggested by Genest et al. [20], is more promising. Here $C_{\hat{\boldsymbol{\theta}}_n}(\boldsymbol{u})$ is replaced by the independence copula $C_\perp$ in the ECP.

The remainder of this paper is structured as follows: Section 2 gives an introduction on vine copula models. The new proposed Information ratio (IR) test is introduced and its test statistics derived in Section 3. Additionally the asymptotic normality of the test statistic is shown. Further GOF tests extended from known copula GOF tests are given in Section 4 for the extensive power comparison study in Section 5 investigating their size and power. An application of an 8-dimensional portfolio of stock indices and their related volatility indices is performed in Section 6 comparing different vine specifications and proposed GOF tests. The final Section 7 summarizes and shows areas of further research.