Journal of Econometrics 193 (2016) 215–233

Contents lists available at ScienceDirect

Journal of Econometrics

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

# Goodness-of-fit test for specification of semiparametric copula dependence models

Shulin Zhang[a], Ostap Okhrin[b,*], Qian M. Zhou[c,d], Peter X.-K. Song[e]

[a] Center of Statistical Research, School of Statistics, Southwestern University of Finance and Economics, China
[b] Faculty of Transportation, Dresden University of Technology, Germany
[c] Department of Statistics and Actuarial Science, Simon Fraser University, Canada
[d] Department of Mathematics and Statistics, Mississippi State University, USA
[e] Department of Biostatistics, University of Michigan, USA

ARTICLE INFO

*Article history:*
Received 5 September 2013
Received in revised form
25 February 2016
Accepted 26 February 2016
Available online 17 March 2016

*JEL classification:*
C12
C22
C32
C52
G15

*Keywords:*
Bootstrap hybrid test
In-and-out-of sample likelihood
Power

ABSTRACT

This paper concerns goodness-of-fit tests for semiparametric copula models. Our contribution is two-fold: we first propose a new test constructed via the comparison between "in-sample" and "out-of-sample" pseudo-likelihoods. Under the null hypothesis that the copula model is correctly specified, we show that the proposed test statistic converges in probability to a constant equal to the dimension of the parameter space. We establish the asymptotic normality and investigate the local power of the test. We also extend the proposed test to the specification test of a class of multivariate time series models, and propose a new bootstrap procedure to establish the finite-sample null distribution, which is shown to have better control of type I error than the commonly used bootstrap. Secondly, we introduce a Bonferroni-based hybrid mechanism to combine several test statistics, which yields a useful test. This hybrid method is particularly appealing when there exists no single dominant optimal test. We conduct comprehensive simulation experiments to compare the proposed new test and hybrid approach with two of the best "blanket" tests in the literature. For illustration, we apply the proposed tests to analyze two real datasets.


## 1. Introduction

Assessing dependency among multiple variables is a primary task in business economics or financial applications. Copula is becoming increasingly popular in such fields due to its flexibility in seamlessly integrating sophisticated dependence structures and varying marginal distributions of multivariate random variables. For example, in finance, copulas are widely applied to study dependency in asset pricing, asset allocation and risk management; see Klugman and Parsa (1999) and Cherubini et al. (2004, 2011), among others. More examples in other fields can be found in Frees and Valdez (1998), Wang and Wells (2000), Song (2007) and Danaher and Smith (2011), just to name a few.

Essentially, a parametric copula is a cumulative distribution function (CDF) specified by a certain known functional form up to some unknown dependence parameters. When a parametric copula is used in applications, misspecification on any of its parametric structure may cause erroneous statistical estimation and inference. To check for the adequacy of a copula model, specification tests have been extensively investigated in the literature. Wang and Wells (2000) proposed a rank based test for bivariate copulas. Malevergne and Sornette (2003) developed a test for the specification of Gaussian copulas. Fermanian (2005) and Scaillet (2007) established goodness-of-fit tests through kernel techniques. Other types of specification tests include Panchenko's (2005) $V$-statistic type test, Prokhorov and Schmidt's (2009) conditional moment based test, Mesfioui et al.'s (2009) Spearman dependence based test, and Genest et al.'s (2011) Pickands dependence based test. Very recently, Huang and Prokhorov (2014) adopted White's test based on the information matrix test (White, 1982) to derive a test for copula model specification. With the utility of either Kendall's or Rosenblat's probability integral transformations, several other versions of specification tests have been proposed in the literature, including those proposed by Breymann et al. (2003), Dobrić and Schmid (2007) and Genest and Favre (2007), among others.

* Corresponding author.
*E-mail addresses:* slzhang@swufe.edu.cn (S. Zhang), ostap.okhrin@tu-dresden.de (O. Okhrin), qmzhou@sfu.ca (Q.M. Zhou), pxsong@umich.edu (P.X.-K. Song).

http://dx.doi.org/10.1016/j.jeconom.2016.02.017
0304-4076/

In a recent paper, Genest et al. (2009) made a thorough comparison for most of the existing "blanket tests". A blanket test refers to a test whose implementation does not require either an arbitrary categorization of data or any strategic choice of smoothing parameter, weight function, kernel or bandwidth. It is demonstrated by Genest et al. (2009) that none of these blanket tests perform uniformly the best. It is interesting to note that almost all of them had illustrated nearly no power in differentiating Gaussian copulas from Student's $t$ copulas, both of which are very important symmetric copulas with different tail dependence properties. Another challenge in the use of the blanket tests considered in Genest et al. (2009) is that they rely on certain probability integral transformations, which may be difficult to derive analytically in many popular copula dependence models, e.g. Student's $t$ copulas and vine copulas (e.g. Kurowicka and Joe, 2011).

To overcome the difficulties above, we propose an alternative specification test for semiparametric copulas in this paper. The proposed test statistic takes a form of ratio constructed via two types of pseudo-likelihoods: one is "in-sample" pseudo-likelihood and the other is "out-of-sample" pseudo-likelihood. The idea behind the construction of the new test is rooted in the fact that, heuristically, a goodness-of-fit test is to examine how a model fits the data. Thus, we vary data by the means of jackknife and quantify how sensitive the pseudo likelihood is to the varying data. Naturally, a comparison of pseudo likelihoods over different datasets are utilized to characterize how well the model fits the data. Inspired by Presnell and Boos's (2004) likelihood based in-and-out-of-sample test, we term our proposed test as the pseudo in-and-out-of-sample (PIOS) test. In comparison to the blanket tests in Genest et al. (2009), which are all indeed rank-based tests, our PIOS test is a pseudo likelihood based test, which does not require any probability integral transformation. Thus, as demonstrated later in the paper, the PIOS test is computationally simple and numerically stable.

Under the null hypothesis of the assumed copula model being correctly specified, we show that under some mild regularity conditions, the PIOS test statistic converges in probability to a constant equal to the dimension of the parameter space of the null copula model. Also, we establish both consistency and asymptotic normality for the PIOS test statistic. Compared to the fully parametric in-and-out-of-sample test proposed by Presnell and Boos (2004), our work makes the following new contributions. First, the PIOS test is applicable to a semiparametric copula model in which the marginal CDFs may be fully unspecified. Secondly, Presnell and Boos's (2004) test is based on a single point data in-and-out-of-sample procedure. As a useful extension, the PIOS test is based on a data block in-sample and out-of-sample procedure, where the size of block is allowed to increase with the sample size. Such flexibility is useful to extend the original method to serially dependent time series data. Thirdly, the development of asymptotic properties of the PIOS test involves the use of the theory of empirical processes with varying block size, and therefore such theoretical work is new and fundamentally different from that established in Presnell and Boos (2004). Fourthly, we develop the asymptotic local power theory in the Pitman sense. Finally, the PIOS test is extended to the case of semi-parametric copula based multivariate dynamic (SCOMDY) model. However, the commonly used bootstrap procedure (Chen and Fan, 2006), based on resampling from estimated innovation processes, may fail to attain the nominal test sizes. We propose a new bootstrap procedure, which involves resampling from the time series data and re-estimating the dynamic parameters of the SCOMDY model in each bootstrap path. The simulation studies have shown that our proposed bootstrap would better control type I error due to accounting for uncertainty in estimating the dynamic parameters.

Another primary focus of the paper is the adoption of Bonferroni correction in combining several test statistics and the resulting test is termed as the hybrid test in this paper. As demonstrated in Genest et al. (2009), there exists no single dominant asymptotically optimal test against general alternatives; see also Freedman (2009). The hybrid test offers a compromise of several different tests, which is particularly appealing when there is no *a priori* knowledge about the top performer in the hypothesis test. We show that the hybrid test can control type I error, as long as each of them does, and that it will be a consistent test as long as there exists one consistent test among the involved tests, regardless of the performance of the remaining tests. The basic setup for the hybrid test is different from that for multiple testing. The difference between these two settings is rooted in the number of null hypotheses involved in the analysis. In our case of hybrid test, there is only one null hypothesis versus one alternative hypothesis, to which several different test statistics (e.g. $S_n, J_n, R_n, T_n$ defined in the following sections) are applied on the same data, so that the test statistics are intrinsically correlated and thus Bonferroni procedure is deemed to control the size of hybrid test. On contrary, in the case of multiple testing, many different null hypotheses are considered and tested simultaneously for whether or not all these null hypotheses hold together, in which only one test statistic is used repeatedly in each hypothesis; see an example of goodness-of-fit test proposed by Hofert and Mächler (2013). Although our setting appears to be different from the multiple testing, the method of Bonferroni procedure is applicable to the hybrid test for the type I error control. Our simulation studies clearly illustrate that, in general, the proposed hybrid test enjoys the desirable finite sample performance.

This paper is organized as follows. Section 2 is devoted to the details for the construction of the PIOS test. Section 3 discusses the hybrid test. Section 4 presents the large sample properties of the proposed PIOS test statistic. Section 5 presents an extension of the PIOS test to multivariate time series data. Section 6 concerns Monte Carlo simulation studies to evaluate finite sample performances of the proposed PIOS test and hybrid test. In Section 7, the proposed tests are applied to two real datasets. The final section provides some concluding remarks. All technical details are included in the appendices.

## 2. Pseudo in-and-out-of-sample (PIOS) test

Suppose that $X_1 = (X_{11}, \dots, X_{1d})^T, \dots, X_n = (X_{n1}, \dots, X_{nd})^T$ is a random sample of size $n$ drawn from a multivariate distribution $H(x) = H(x_1, x_2, \dots, x_d)$ with continuous marginal CDF $F(x) \overset{\Delta}{=} \{F_1(x_1), \dots, F_d(x_d)\}$. According to Sklar's theorem (Sklar, 1959), we suppose that the joint distribution $H(\cdot)$ can be expressed by the following representation:

$$H(x_1, x_2, \dots, x_d) \overset{\Delta}{=} C_0\{F(x)\} = C_0\{F_1(x_1), \dots, F_d(x_d)\},$$

where $C_0(\cdot)$ is the true copula function. The corresponding joint density function of $H(\cdot)$, denoted by $h(\cdot)$, takes the form of

$$h(x_1, x_2, \dots, x_d) = c_0\{F_1(x_1), \dots, F_d(x_d)\} \prod_{k=1}^{d} f_k(x_k),$$

where, $c_0(u)$, $u = (u_1, \dots, u_d) \in (0, 1)^d$ is the resulting copula density function of copula $C_0(\cdot)$ and $f_k(\cdot)$ are the corresponding marginal density functions of $F_k(\cdot)$, $k = 1, \dots, d$. Throughout this paper, the marginal CDF $F(\cdot)$ is not specified by any parametric forms.

In practice, we often assume that the underlying true copula $C_0$ belongs to a parametric class, *say*,

$$\mathcal{C} \overset{\Delta}{=} \{C(\cdot; \theta), \theta \in \Theta\},$$

where $\Theta \subset \mathcal{R}^p$ is a $p$-dimensional parameter space. It is well known that misspecification on any of its parametric structure