Kernel Based Goodness-of-Fit Tests for Copulas with Fixed Smoothing Parameters

Olivier Scaillet
HEC Genève and Swiss Finance Institute
Université de Genève
Bd Carl Vogt, 102
CH - 1211 Genève 4, Suisse
olivier.scaillet@hec.unige.ch

First version: May 2005
Revised version: March 2006

Abstract

We study a test statistic based on the integrated squared difference between a kernel estimator of the copula density and a kernel smoothed estimator of the parametric copula density. We show for fixed smoothing parameters that the test is consistent and that the asymptotic properties are driven by a U-statistic of order 4 with degeneracy of order 1. For practical implementation we suggest to compute the critical values through a semiparametric bootstrap. Monte Carlo results show that the bootstrap procedure performs well in small samples. In particular size and power are less sensitive to smoothing parameter choice than they are under the asymptotic approximation obtained for a vanishing bandwidth.

Key words and phrases: Nonparametric, Copula density, Goodness-of-fit test, U-statistic.

JEL Classification: C12, D81, G10, G21, G22.

MSC 2000: 62G10, 62P05.

1 Introduction

Goodness-of-fit (gof) tests have a long history in statistics (D'Agostino and Stephens [4]) while copula models find new interesting applications in finance and insurance (see, e.g., McNeil, Frey and Embrechts [19], Denuit, Dhaene, Goovaerts and Kaas [6]) beside their long acclaimed applications in reliability and survival analysis. Nevertheless relatively little is known about properties of gof tests for copulas despite an obvious need for such tools in applied work; see however Genest, Quessy and Rémillard [15], and Chen, Fan and Patton [3] for study of gof tests based on the integral probability transformation. The main reason is the technical difficulty induced by the probabilistic behaviour of the empirical copula process (see, e.g., van der Vaart and Wellner [23], Fermanian, Radulovic and Wegkamp [11]). In order to circumvent this difficulty Fermanian [10] suggests to use a gof test based on the integrated weighted squared difference between a kernel estimator of the copula density and a kernel smoothed estimator of the parametric copula density. In particular he shows that the test statistic is asymptoticallly normally distributed when the bandwidth shrinks to zero. Fan [7] has previously established similar results for bias-corrected gof tests of standard pdf via kernel methods.

In this paper we study the asymptotic properties of the gof test introduced by Fermanian [10], but holding the smoothing parameters fixed and the weight function equal to one as in Fan [9]. We start with recalling the form of the test statistic in Section 2. We derive its interpretation in terms of a weighted integrated squared difference between the characteristic function of the empirical copula process and the characteristic function of the estimated parametric copula of the null hypothesis. A direct consequence of such an interpretation is test consistency for fixed smoothing parameters. We show that the asymptotic properties are driven by a U-statistic of order 4 with degeneracy of order 1. This is to be contrasted with the result for standard pdf obtained by Fan [9], namely asymptotic properties driven by a U-statistic of order 2 with degeneracy of order 1. To work with copulas carries here a price in terms of analytical tractability of the asymptotic distribution. Therefore for practical implementation we recommend to compute the critical values through a semiparametric bootstrap. Monte Carlo results of Section 3 reveal that the bootstrap performs well in small samples. In particular size and power are less sensitive to smoothing parameter choice than the same characteristics under the asymptotic approximation obtained for a vanishing bandwidth.

2 Test statistic and asymptotic properties

We consider a setting made of i.i.d. observations $\{\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})'; i = 1, \ldots, n\}$ of a random vector taking values in $\mathbb{R}^d$. The distribution of $\boldsymbol{X}$ is denoted by F, and the margins are denoted by F_j, $j = 1, ..., d$. The copula function is denoted by C, and its density by c.

Following Deheuvels [5], let us define the empirical copula function by

$$\hat{C}(\boldsymbol{u}) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{\hat{F}_1(X_{i1}) \leq u_1, \ldots, \hat{F}_d(X_{id}) \leq u_d\}, \quad \boldsymbol{u} = (u_1, \ldots, u_d)' \in [0,1]^d,$$

where $\hat{F}_j$ is the empirical cumulative distribution functions computed from $\{X_{ij}; i = 1, \ldots, n\}$,

$j = 1, ..., d$. Observe that $\hat{C}$ is actually a function of the ranks of the observations since $n\hat{F}_j(X_{ij})$ is the rank of X_{ij} among $X_{1j}, \ldots, X_{nj}$.

Nonparametric estimation procedures for the density of a copula function have been proposed by Behnen, Huskova, and Neuhaus [2], and Gijbels and Mielniczuk [17]. They rely on a kernel smoothing directly applied to the transformed sample $\left\{\hat{\mathbf{Y}}_i = (\hat{F}_1(X_{i1}), \ldots, \hat{F}_d(X_{id}))'; i = 1, \ldots, n\right\}$. The kernel estimator of the copula density at point $\boldsymbol{u}$ is simply

$$\hat{c}(\boldsymbol{u}) := \frac{1}{n}\sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{u} - \hat{\mathbf{Y}}_i), \tag{2.1}$$

where $K_{\boldsymbol{H}}(\boldsymbol{y}) = K(\boldsymbol{H}^{-1}\boldsymbol{y})/\det \boldsymbol{H}$, K is a d-dimensional kernel, and $\boldsymbol{H}$ is a nonsingular, symmetric matrix of smoothing parameters.

Fermanian [10] suggests to use the following gof test statistic for the parametric family $C(\boldsymbol{u}; \theta)$ with density $c(\boldsymbol{u}; \theta)$ and $\theta \in \Theta \subset \mathbb{R}^p$:

$$\hat{J}(w) := \int \left[\hat{c}(\boldsymbol{u}) - K_{\boldsymbol{H}} * c(\boldsymbol{u}; \hat{\theta})\right]^2 w(\boldsymbol{u}) d\boldsymbol{u},$$

where $*$ denotes convolution, and w is a weight function.

The estimator $\hat{\theta}$ can be computed by a semiparametric maximum likelihood method as in Genest, Ghoudi and Rivest [14], and Shih and Louis [22]. Let θ_0 be the true value of the parameter under the null hypothesis of well specification. Then the semiparametric estimator $\hat{\theta}$ satisfies under the null hypothesis:

$$\hat{\theta} - \theta_0 = n^{-1}\boldsymbol{A}(\theta_0)\sum_{i=1}^{n} \boldsymbol{B}(\boldsymbol{X}_i; \theta_0) + o_p(n^{-1/2}),$$

where $\boldsymbol{A}(\theta_0)$ is a $p \times p$ positive definite matrix and $\boldsymbol{B}(\boldsymbol{X}_i; \theta_0)$ is a $p \times 1$ random vector, such that $E[\boldsymbol{B}(\boldsymbol{X}_i; \theta_0)] = 0$ and $E[\boldsymbol{B}(\boldsymbol{X}_i; \theta_0)'\boldsymbol{B}(\boldsymbol{X}_i; \theta_0)] < \infty$.

The following lemma justifies the use of $\hat{J}$ even if the bandwidth is not assumed to shrink to zero when the sample size grows to infinity. The kernel K is chosen so that it is symmetric about zero, square integrable, and admits a Fourier transform which vanishes on a set of Lebesgue measure zero.

Lemma 2.1. *For $w = 1$,*

$$\hat{J} := \hat{J}(1) = \int |\bar{\hat{C}}(\boldsymbol{t}) - \bar{C}(\boldsymbol{t}; \hat{\theta})|^2 \bar{K}^2(\boldsymbol{H}\boldsymbol{t}) d\boldsymbol{t},$$

where $\bar{\hat{C}}(\boldsymbol{t}) := \int \exp(i\boldsymbol{t}'\boldsymbol{u})\hat{C}(d\boldsymbol{u})$, $\bar{C}(\boldsymbol{t}; \hat{\theta}) := \int \exp(i\boldsymbol{t}'\boldsymbol{u})C(d\boldsymbol{u}; \hat{\theta})$, *and* $\bar{K}(\boldsymbol{t}) := (2\pi)^{-d/2}\int \exp(i\boldsymbol{t}'\boldsymbol{u})K(\boldsymbol{u})d\boldsymbol{u}$.

The lemma states that $\hat{J}$ reduces to the comparison of two empirical characteristic functions when the weight function is equal to one (see Fan [8] for use of empirical characteristic functions in gof tests). The first term $\bar{\hat{C}}(\boldsymbol{t})$ is the empirical characteristic function of the transformed sample, while the second term is the estimated characteristic function under the parametric assumption. Decomposing the characteristic function $\bar{C}$ of the copula C in its real part $\operatorname{Re}\bar{C}(\boldsymbol{t}) = \int \cos(\boldsymbol{t}'\boldsymbol{u})C(d\boldsymbol{u})$ and its imaginary part $\operatorname{Im}\bar{C}(\boldsymbol{t}) = \int \sin(\boldsymbol{t}'\boldsymbol{u})C(d\boldsymbol{u})$, and recognizing that the maps which assign to each copula function C the real and imaginary parts are linear maps, we get from the delta method (van der Vaart and Wellner [23] Section 3.9) that $\operatorname{Re}\bar{\hat{C}}(\boldsymbol{t})$, $\operatorname{Im}\bar{\hat{C}}(\boldsymbol{t})$, $\operatorname{Re}\bar{C}(\boldsymbol{t};\hat{\theta})$, and $\operatorname{Im}\bar{C}(\boldsymbol{t};\hat{\theta})$, are consistent and jointly asymptotically normally distributed. Hence we deduce that $\bar{\hat{C}}(\boldsymbol{t})$ and $\bar{C}(\boldsymbol{t};\hat{\theta})$ are consistent and asymptotically normally distributed in the complex plane (see Feueverger and Mureika [12] for the results on the standard empirical characteristic function). Now recall that a copula function is a cdf on the unit cube, and that there is a one-to-one correspondence between distribution functions and characteristic functions. This gives that no matter the choice of the bandwidth matrix $\boldsymbol{H}$, the limit J of $\hat{J}$ when $w = 1$ is such that $J \geq 0$ and $J = 0$ iff the copula function is well specified, provided that the set $\{\boldsymbol{t} \in \mathbb{R}^d : \bar{K}(\boldsymbol{t}) = 0\}$ has Lebesgue measure zero. Hence we may conclude as in Fan [9] that a test based on $\hat{J}$ is consistent when holding the smoothing parameters fixed and the weight function equal to one.

The following proposition describes the asymptotic behaviour of $\hat{J}$ when the bandwidth matrix does not vanish.

Proposition 2.2. *Under the null hypothesis of well specification the asymptotic behaviour of $\hat{J}$ is that of a U-statistic of order 4 with degeneracy of order 1.*

Contrary to the case of a non-vanishing bandwidth for standard pdf (Fan [9]) the asymptotic behaviour of the gof test statistic is not that of a U-statistic of order 2 but of order 4. From the degeneracy of order 1 the rate of convergence is still n, and the exact form of the asymptotic distribution is an infinite sum of weighted chi-square random variables. The weights can be computed in theory from the eigenvalues of an integral equation (see Lee [18] p. 80). However the kernel of the U-statistic of order 4 involves 24 terms (see the proof of the proposition), and the dimension of the integral is 3. This casts doubt on the numerical accuracy of such a method in practice.

Hereafter we rely on a semiparametric bootstrap to compute the critical values of the test. First we draw from the estimated copula $C(\boldsymbol{u};\hat{\theta})$ in order to impose the dependence structure of the null hypothesis, and then we use the ranks of these draws to build the bootstrap transformed sample. This semiparametric bootstrap is already exploited in Genest, Quessy and Rémillard [15] since the distributions of their test statistics depend on the unknown parameter value, even in the limit. Its validity for a broad class of gof testing problems is shown in Genest and Rémillard [16]. In particular in the context of gof tests for copulas they show that the sequence associated with the empirical copula is regular for the parametric copula family of the null hypothesis (Genest and Rémillard [16] Proposition 4.2). They also show the regularity of the parametric estimators we use in this paper (Genest and Rémillard [16] Example 4.4). Since we work with linear maps the consistency of the semiparametric bootstrap in our setting is a straightforward consequence of their results.

3 Monte Carlo results

In this section we study the performance of kernel based gof tests for copulas in small samples when the weight function is kept equal to one. We compare the performance of rejection rules based on 1) asymptotic sets of the chi-square test statistic derived in Fermanian [10] Corollary 4 for a vanishing bandwidth, 2) sets computed with a bootstrap procedure for the same test statistic, 3) sets computed with a bootstrap procedure for the test statictic $n\hat{J}$.

First we examine gof tests for the Frank copula. Table I gathers results concerning the size of the different testing procedures, i.e., when the true copula is a Frank copula. This parametric family is often used in actuarial and financial applications, and is easy to draw from; see, e.g., Genest [13] or Nelsen [20]. The chosen values of the parameter are $\theta \in \{1, 2, 3\}$. They match low to moderate positive dependences as exhibited by the corresponding true values of the Kendall tau, $\tau \in \{.11, .21, .31\}$.

The sample sizes are fixed at $n = 50$ and $n = 200$. The first sample size can be thought as rather small since we face a bivariate inference problem. For the sake of interpretation samples are generated with both margins corresponding to an exponential distribution with a unit parameter. This can be seen as mimicking the behaviour of claim or duration data. Note that the numerical results below remain exactly the same if we use uniform margins or other margins with strictly monotonic continuously differentiable cdf (such as Gaussian or Student margins to mimick financial returns) and keep the same seeds in the pseudo-random generators. The reason is that the testing procedures rely intrinsically on ranks.

The kernel estimator of the copula density is based on a bivariate quartic product kernel. Then the Scott's rule of thumb (Scott [21]) to select the smoothing parameters gives $\boldsymbol{H} = 2.6073 n^{-1/6} \hat{\Sigma}^{1/2}$, where $\hat{\Sigma}$ is the estimated covariance matrix of the transformed data. To gauge the impact of the choice of the smoothing parameters we report sizes for multiples of this bandwith matrix, namely $\delta \boldsymbol{H}$ with $\delta \in \{.1, .25, .5, 1, 1.5\}$. In our simulations the diagonal terms of the selected $\boldsymbol{H}$ are close to one third. We use a bivariate Gauss-Legendre quadrature with 12×12 knots to compute the test statistic, and the optmum routine of the Gauss statistical software with user-supplied analytical gradient and Hessian to optimize the semiparametric loglikelihood. Since we use a Gauss-Legendre quadrature with knots belonging to $(0, 1)^2$ the restriction that the support of w should be strictly inside the unit square (Fermanian [10] Assumption T) for the asymptotic distribution to be theoretically valid has no practical impact here. The number of bootstrap samples to approximate the p-value is set equal to 500. For each case 200 Monte Carlo simulations are performed, and the rejection rates are computed for each method w.r.t. the conventional significance level of $\alpha = 0.05$. Programs are available on request.

The results in Table I show that the asymptotic testing procedure is highly sensitive to the bandwidth choice, and that the size distortion may be large. Both bootstrap methods do not suffer from these inconvenient features when $n = 200$. However they tend to slightly underreject when $n = 50$. From bootstrap theory on higher-order improvements, we know that the bootstrap is expected to yield better results when applied to asymptotic pivots. However the difference between both simulation based methods is not striking in our finite sample experiments. We might thus prefer to use the second bootstrap method, namely the one relying simply on $n^2\hat{J}$, in light of its computational ease and speed. This second method does not require computing complicated asymptotic bias and variance terms.

In Table II, we gather some results about the power of the testing procedures. We consider a case where 50% of the observed sample is substituted for data drawn from a Student copula with 4 degrees of freedom and a .95 correlation parameter. Again the results indicate that the asymptotic testing procedure is much more affected by the bandwidth choice than both bootstrap procedures. Note further that the smoothing parameters should not be chosen too large (oversmoothing) or too small (undersmoothing) in order to improve on power. We may then conclude that, since the power may be weak in some cases, it is even more crucial to get a well controlled size via the bootstrap.

To get further insight on the behavior of the testing procedure we have also considered a non-Archimedean copula. Tables III and IV correspond to Tables I and II but with the Gaussian copula replacing the Frank copula. The semiparametric and nonparametric estimation methods are kept the same. The chosen parameter values of the Gaussian copula are $\theta \in \{.17, .32, .47\}$ so that $\tau \in \{.11, .21, .31\}$. Results are akin to those got for the Frank copula, and conclusions remain unchanged.

Finally Table V allows for comparison with the Cramer-von Mises (CVM) and Kolmogorov-Smirnov (KS) testing procedures of Genest, Quessy and Rémillard (GQR) [15]. Their procedures are very easy to implement for parametric copula models admitting a distribution function of the probability integral transformation in closed form. This is the case for Archimedean copulas, such as the Frank copula. We can see that results on size are similar to those reported in Table I, but results on power seem to be better in Table II when $\delta \in \{.25, .5, 1\}$. Of course these results are not extensive enough to have a definitive answer (if possible) about which gof test for copulas to favour overall. Even if such an extensive Monte Carlo study is certainly of interest, this is beyond the scope of this note.

TABLE I: Impact of bandwidth choice on size

Rejection rates at 5% level with 200 replications

Pseudo copula: Frank, True copula: Frank

Size	$n = 50$					$n = 200$				
F: $\theta = 1$	$\delta = .1$	.25	.5	1	1.5	$\delta = .1$	.25	.5	1	1.5
Asym.	.44	.00	.00	.00	.00	.51	.06	.00	.00	.00
As. Boot.	.05	.03	.04	.03	.00	.06	.05	.06	.05	.05
Boot.	.05	.03	.05	.04	.04	.06	.05	.07	.05	.06
F: $\theta = 2$	$\delta = .1$	.25	.5	1	1.5	$\delta = .1$	.25	.5	1	1.5
Asym.	.50	.00	.00	.00	.00	.51	.06	.01	.00	.00
As. Boot.	.05	.06	.05	.02	.00	.06	.06	.06	.06	.04
Boot.	.05	.07	.05	.03	.03	.06	.06	.06	.06	.06
F: $\theta = 3$	$\delta = .1$	.25	.5	1	1.5	$\delta = .1$	.25	.5	1	1.5
Asym.	.51	.01	.00	.00	.00	.48	.04	.01	.00	.00
As. Boot.	.05	.06	.05	.03	.03	.03	.03	.06	.05	.05
Boot.	.05	.06	.06	.04	.03	.03	.03	.06	.05	.05

TABLE II: Impact of bandwidth choice on power

Rejection rates at 5% level with 200 replications

Pseudo copula: Frank, True copula: mixture of Frank and Student

Power	$n = 50$					$n = 200$				
F: $\theta = 1$	$\delta = .1$	.25	.5	1	1.5	$\delta = .1$	.25	.5	1	1.5
Asym.	.51	.08	.02	.00	.00	.65	.77	.77	.04	.00
As. Boot.	.06	.20	.26	.09	.00	.13	.71	.96	.72	.18
Boot.	.06	.19	.26	.18	.17	.14	.71	.96	.69	.26
F: $\theta = 2$	$\delta = .1$	.25	.5	1	1.5	$\delta = .1$	.25	.5	1	1.5
Asym.	.46	.04	.00	.00	.00	.58	.67	.52	.01	.00
As. Boot.	.07	.11	.22	.05	.00	.12	.58	.86	.50	.11
Boot.	.06	.12	.21	.12	.13	.12	.57	.83	.47	.16
F: $\theta = 3$	$\delta = .1$	.25	.5	1	1.5	$\delta = .1$	.25	.5	1	1.5
Asym.	.49	.00	.00	.00	.00	.64	.54	.26	.00	.00
As. Boot.	.04	.10	.15	.04	.00	.09	.39	.61	.25	.07
Boot.	.04	.10	.13	.10	.08	.10	.39	.60	.21	.12

TABLE III: Impact of bandwidth choice on size
Rejection rates at 5% level with 200 replications
Pseudo copula: Gaussian, True copula: Gaussian

Size	$n = 50$					$n = 200$				
F: $\theta = .17$	$\delta = .1$	.25	.5	1	1.5	$\delta = .1$	.25	.5	1	1.5
Asym.	.46	.00	.00	.00	.00	.48	.04	.00	.00	.00
As. Boot.	.03	.02	.06	.04	.00	.03	.08	.06	.06	.04
Boot.	.06	.04	.06	.05	.06	.05	.08	.07	.06	.05
F: $\theta = .32$	$\delta = .1$	.25	.5	1	1.5	$\delta = .1$	.25	.5	1	1.5
Asym.	.51	.00	.00	.00	.00	.62	.03	.01	.00	.00
As. Boot.	.04	.02	.04	.02	.02	.07	.05	.05	.05	.04
Boot.	.04	.04	.04	.04	.06	.09	.07	.08	.05	.05
F: $\theta = .47$	$\delta = .1$	.25	.5	1	1.5	$\delta = .1$	.25	.5	1	1.5
Asym.	.69	.00	.00	.00	.00	.73	.04	.01	.00	.00
As. Boot.	.02	.03	.06	.01	.01	.04	.04	.06	.05	.03
Boot.	.06	.05	.06	.06	.06	.06	.05	.07	.06	.06

TABLE IV: Impact of bandwidth choice on power
Rejection rates at 5% level with 200 replications
Pseudo copula: Gaussian, True copula: mixture of Gaussian and Student

Power	$n = 50$					$n = 200$				
F: $\theta = .17$	$\delta = .1$	.25	.5	1	1.5	$\delta = .1$	.25	.5	1	1.5
Asym.	.72	.01	.02	.01	.00	.64	.49	.85	.35	.00
As. Boot.	.06	.01	.38	.07	.01	.02	.39	1	1	.72
Boot.	.07	.22	.39	.23	.12	.16	.92	1	1	.88
F: $\theta = .32$	$\delta = .1$	.25	.5	1	1.5	$\delta = .1$	.25	.5	1	1.5
Asym.	.76	.02	.01	.00	.00	.75	.25	.52	.13	.00
As. Boot.	.06	.01	.29	.03	.00	.01	.08	.96	.96	.45
Boot.	.07	.20	.34	.18	.10	.12	.80	.99	.97	.69
F: $\theta = .47$	$\delta = .1$	.25	.5	1	1.5	$\delta = .1$	.25	.5	1	1.5
Asym.	.81	.09	.00	.00	.00	.83	.05	.21	.03	.00
As. Boot.	.03	.03	.17	.02	.00	.03	.01	.72	.77	.21
Boot.	.09	.12	.21	.14	.09	.10	.65	.96	.85	.48

TABLE V: Comparison with CVM and KS tests of GQR

Rejection rates at 5% level with 200 replications

Size/Power: Pseudo copula: Frank, True copula: Frank/mixture

Size	$n = 50$		$n = 200$		Power	$n = 50$		$n = 200$	
	CVM	KS	CVM	KS		CVM	KS	CVM	KS
F: $\theta = 1$	.05	.03	.06	.05	F: $\theta = 1$	.06	.04	.28	.23
F: $\theta = 2$	.04	.03	.06	.05	F: $\theta = 2$	.04	.04	.24	.19
F: $\theta = 3$	.04	.03	.06	.05	F: $\theta = 3$	.05	.04	.22	.16

APPENDIX

A Proof of Lemma 2.1

The Fourier transform of $\hat{c}(\boldsymbol{u}) - K_{\boldsymbol{H}} * c(\boldsymbol{u};\hat{\theta})$ is given by

$$\begin{aligned}
&(2\pi)^{-d/2} \int \exp(i\boldsymbol{t}'\boldsymbol{u}) \left[\hat{c}(\boldsymbol{u}) - K_{\boldsymbol{H}} * c(\boldsymbol{u};\hat{\theta})\right] d\boldsymbol{u} \\
&\quad = (2\pi)^{-d/2} \int \exp(i\boldsymbol{t}'\boldsymbol{u}) \left[\int K_{\boldsymbol{H}}(\boldsymbol{u}-\boldsymbol{v}) \left\{\hat{C}(d\boldsymbol{v}) - C(d\boldsymbol{v};\hat{\theta})\right\}\right] d\boldsymbol{u} \\
&\quad = \int \exp(i\boldsymbol{t}'\boldsymbol{v}) \bar{K}(\boldsymbol{H}\boldsymbol{t}) \left\{\hat{C}(d\boldsymbol{v}) - C(d\boldsymbol{v};\hat{\theta})\right\}.
\end{aligned}$$

The stated result is then deduced from an application of Parseval's identity to $\hat{J}$ and the last equality.

B Proof of Proposition 2.2

Let us introduce $\{\mathbf{Y}_i = (F_1(X_{i1}), \ldots, F_d(X_{id}))'; i = 1, \ldots, n\}$. Obviously C is the cdf of $\mathbf{Y}$. We consider

$$\hat{I} = \int \left[\frac{1}{n}\sum_{i=1}^{n} a(\boldsymbol{u}, \boldsymbol{X}_i)\right]^2 d\boldsymbol{u}, \tag{B.1}$$

with

$$\begin{aligned}
a(\boldsymbol{u}, \boldsymbol{X}_i) &= K_{\boldsymbol{H}}(\boldsymbol{u} - \mathbf{Y}_i) + K_{\boldsymbol{H}}^{(1)}(\boldsymbol{u} - \mathbf{Y}_i)'(\hat{\mathbf{Y}}_i - \mathbf{Y}_i) \\
&\quad - \int K_{\boldsymbol{H}}(\boldsymbol{u} - \boldsymbol{v}) C(d\boldsymbol{v};\theta_0) - \mu(\boldsymbol{u})\boldsymbol{A}(\theta_0)\boldsymbol{B}(\boldsymbol{X}_i;\theta_0),
\end{aligned}$$

where $K_{\boldsymbol{H}}^{(1)}$ denotes the first derivative of $K_{\boldsymbol{H}}$, and $\mu(\boldsymbol{u}) = \int K_{\boldsymbol{H}}(\boldsymbol{u}-\boldsymbol{v})\partial C(d\boldsymbol{v};\theta_0)/\partial\theta'$.

As in Fan [9] Lemma 3.1, via simple second-order Taylor expansions of $\hat{c}(\boldsymbol{u})$ around $\mathbf{Y}_i$, and $c(\boldsymbol{u};\hat{\theta})$ around θ_0, we may check that $n(\hat{J} - \hat{I}) = o_p(1)$ under the null hypothesis and our assumptions since $\sqrt{n}(\hat{\theta} - \theta_0) = O_p(1)$ and $\sqrt{n}(\hat{\mathbf{Y}}_i - \mathbf{Y}_i) = O_p(1)$. Indeed substituting the expansions into $\hat{J}$ and collecting terms yield $n(\hat{J} - \hat{I}) = n\sum_{i=1}^{7} \hat{I}_i$ with $\hat{I}_i = \int \hat{I}_i(\boldsymbol{u})d\boldsymbol{u}$ and

$$\begin{aligned}
\hat{I}_1(\boldsymbol{u}) &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \left[K_{\boldsymbol{H}}(\boldsymbol{u} - \mathbf{Y}_i) + K_{\boldsymbol{H}}^{(1)}(\boldsymbol{u} - \mathbf{Y}_i)'(\hat{\mathbf{Y}}_i - \mathbf{Y}_i)\right] \\
&\qquad \left[(\hat{\mathbf{Y}}_j - \mathbf{Y}_j)' K_{\boldsymbol{H}}^{(2)}(\boldsymbol{u} - \mathbf{Y}_j)(\hat{\mathbf{Y}}_j - \mathbf{Y}_j)\right],
\end{aligned}$$

$$\begin{aligned}
\hat{I}_2(\boldsymbol{u}) &= \frac{1}{4n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \left[(\hat{\mathbf{Y}}_i - \mathbf{Y}_i)' K_{\boldsymbol{H}}^{(2)}(\boldsymbol{u} - \mathbf{Y}_i)(\hat{\mathbf{Y}}_i - \mathbf{Y}_i)\right] \\
&\qquad \left[(\hat{\mathbf{Y}}_j - \mathbf{Y}_j)' K_{\boldsymbol{H}}^{(2)}(\boldsymbol{u} - \mathbf{Y}_j)(\hat{\mathbf{Y}}_j - \mathbf{Y}_j)\right],
\end{aligned}$$

$$\hat{I}_3(\boldsymbol{u}) = \int\int \Big[K_{\boldsymbol{H}}(\boldsymbol{u}-\boldsymbol{v})\Big(C(d\boldsymbol{v};\theta_0)+\partial C(d\boldsymbol{v};\theta_0)/\partial\theta'(\hat{\theta}-\theta_0)\Big)\Big] \Big[K_{\boldsymbol{H}}(\boldsymbol{u}-\boldsymbol{s})(\hat{\theta}-\theta_0)'\partial^2 C(d\boldsymbol{s};\theta_0)/\partial\theta\partial\theta'(\hat{\theta}-\theta_0)\Big],$$

$$\hat{I}_4(\boldsymbol{u}) = \frac{1}{4}\int\int \Big[K_{\boldsymbol{H}}(\boldsymbol{u}-\boldsymbol{v})(\hat{\theta}-\theta_0)'\partial^2 C(d\boldsymbol{v};\theta_0)/\partial\theta\partial\theta'(\hat{\theta}-\theta_0)\Big] \Big[K_{\boldsymbol{H}}(\boldsymbol{u}-\boldsymbol{s})(\hat{\theta}-\theta_0)'\partial^2 C(d\boldsymbol{s};\theta_0)/\partial\theta\partial\theta'(\hat{\theta}-\theta_0)\Big],$$

$$\hat{I}_5(\boldsymbol{u}) = -\frac{1}{n}\sum_{i=1}^{n}\Big[K_{\boldsymbol{H}}(\boldsymbol{u}-\mathbf{Y}_i)+K^{(1)}_{\boldsymbol{H}}(\boldsymbol{u}-\mathbf{Y}_i)'(\hat{\mathbf{Y}}_i-\mathbf{Y}_i)\Big] \Big[\int K_{\boldsymbol{H}}(\boldsymbol{u}-\boldsymbol{v})(\hat{\theta}-\theta_0)'\partial^2 C(d\boldsymbol{v};\theta_0)/\partial\theta\partial\theta'(\hat{\theta}-\theta_0)\Big],$$

$$\hat{I}_6(\boldsymbol{u}) = -\frac{1}{n}\sum_{i=1}^{n}\Big[(\hat{\mathbf{Y}}_i-\mathbf{Y}_i)'K^{(2)}_{\boldsymbol{H}}(\boldsymbol{u}-\mathbf{Y}_i)(\hat{\mathbf{Y}}_i-\mathbf{Y}_i)\Big] \Big[\int K_{\boldsymbol{H}}(\boldsymbol{u}-\boldsymbol{v})\Big(C(d\boldsymbol{v};\theta_0)+\partial C(d\boldsymbol{v};\theta_0)/\partial\theta'(\hat{\theta}-\theta_0)\Big)\Big],$$

$$\hat{I}_7(\boldsymbol{u}) = -\frac{1}{2n}\sum_{i=1}^{n}\Big[(\hat{\mathbf{Y}}_i-\mathbf{Y}_i)'K^{(2)}_{\boldsymbol{H}}(\boldsymbol{u}-\mathbf{Y}_i)(\hat{\mathbf{Y}}_i-\mathbf{Y}_i)\Big] \Big[\int K_{\boldsymbol{H}}(\boldsymbol{u}-\boldsymbol{v})(\hat{\theta}-\theta_0)'\partial^2 C(d\boldsymbol{v};\theta_0)/\partial\theta\partial\theta'(\hat{\theta}-\theta_0)\Big].$$

Then since $\frac{1}{n}\sum_{i=1}^{n}K_{\boldsymbol{H}}(\boldsymbol{u}-\mathbf{Y}_i)-\int K_{\boldsymbol{H}}(\boldsymbol{u}-\boldsymbol{v})C(d\boldsymbol{v};\theta_0)$ converges to zero and the absolute value of each of its elements is bounded for any $\boldsymbol{u}$, we can deduce that $n(\hat{I}_1+\hat{I}_6)=o_p(1)$ and $n(\hat{I}_3+\hat{I}_5)=o_p(1)$, while $n\hat{I}_2=O_p(n^{-1})$, $n\hat{I}_4=O_p(n^{-1})$, and $n\hat{I}_7=O_p(n^{-1})$.

Now we can rewrite Equation (B.1) as

$$\hat{I} = \frac{1}{n^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} g(\boldsymbol{X}_i,\boldsymbol{X}_j,\boldsymbol{X}_k,\boldsymbol{X}_l), \tag{B.2}$$

with

$$\begin{aligned} g(\boldsymbol{X}_i,\boldsymbol{X}_j,\boldsymbol{X}_k,\boldsymbol{X}_l) = & \int\{K_{\boldsymbol{H}}(\boldsymbol{u}-\mathbf{Y}_i)+K^{(1)}_{\boldsymbol{H}}(\boldsymbol{u}-\mathbf{Y}_i)'(\boldsymbol{I}[\boldsymbol{X}_k\le\boldsymbol{X}_i]-\mathbf{Y}_i) \\ & -\int K_{\boldsymbol{H}}(\boldsymbol{u}-\boldsymbol{v})C(d\boldsymbol{v};\theta_0)-\mu(\boldsymbol{u})\boldsymbol{A}(\theta_0)\boldsymbol{B}(\boldsymbol{X}_i;\theta_0)\} \\ & \{K_{\boldsymbol{H}}(\boldsymbol{u}-\mathbf{Y}_j)+K^{(1)}_{\boldsymbol{H}}(\boldsymbol{u}-\mathbf{Y}_j)'(\boldsymbol{I}[\boldsymbol{X}_l\le\boldsymbol{X}_j]-\mathbf{Y}_j) \\ & -\int K_{\boldsymbol{H}}(\boldsymbol{u}-\boldsymbol{v})C(d\boldsymbol{v};\theta_0)-\mu(\boldsymbol{u})\boldsymbol{A}(\theta_0)\boldsymbol{B}(\boldsymbol{X}_j;\theta_0)\}d\boldsymbol{u}, \end{aligned}$$

with $\boldsymbol{I}[\boldsymbol{X}_k \leq \boldsymbol{X}_i] = (I[X_{k1} \leq X_{i1}], ..., I[X_{kd} \leq X_{id}])'$.

By symmetrization of the kernel g (Lee [18] p. 7) we know that $\hat{I}$ shares the same asymptotic behaviour as the V-statistic:

$$\hat{V} = \frac{1}{n^4} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} \psi(\boldsymbol{X}_i, \boldsymbol{X}_j, \boldsymbol{X}_k, \boldsymbol{X}_l),$$

where ψ is a symmetric kernel given by:

$$\begin{aligned}
24\psi(\boldsymbol{X}_i, \boldsymbol{X}_j, \boldsymbol{X}_k, \boldsymbol{X}_l) \;=\; & g(\boldsymbol{X}_i, \boldsymbol{X}_j, \boldsymbol{X}_k, \boldsymbol{X}_l) + g(\boldsymbol{X}_i, \boldsymbol{X}_j, \boldsymbol{X}_l, \boldsymbol{X}_k) + g(\boldsymbol{X}_i, \boldsymbol{X}_k, \boldsymbol{X}_j, \boldsymbol{X}_l) \\
& + g(\boldsymbol{X}_i, \boldsymbol{X}_k, \boldsymbol{X}_l, \boldsymbol{X}_j) + g(\boldsymbol{X}_i, \boldsymbol{X}_l, \boldsymbol{X}_j, \boldsymbol{X}_k) + g(\boldsymbol{X}_i, \boldsymbol{X}_l, \boldsymbol{X}_k, \boldsymbol{X}_j) \\
& + g(\boldsymbol{X}_j, \boldsymbol{X}_i, \boldsymbol{X}_k, \boldsymbol{X}_l) + g(\boldsymbol{X}_j, \boldsymbol{X}_i, \boldsymbol{X}_l, \boldsymbol{X}_k) + g(\boldsymbol{X}_j, \boldsymbol{X}_k, \boldsymbol{X}_i, \boldsymbol{X}_l) \\
& + g(\boldsymbol{X}_j, \boldsymbol{X}_k, \boldsymbol{X}_l, \boldsymbol{X}_i) + g(\boldsymbol{X}_j, \boldsymbol{X}_l, \boldsymbol{X}_i, \boldsymbol{X}_k) + g(\boldsymbol{X}_j, \boldsymbol{X}_l, \boldsymbol{X}_k, \boldsymbol{X}_i) \\
& + g(\boldsymbol{X}_k, \boldsymbol{X}_i, \boldsymbol{X}_j, \boldsymbol{X}_l) + g(\boldsymbol{X}_k, \boldsymbol{X}_i, \boldsymbol{X}_l, \boldsymbol{X}_j) + g(\boldsymbol{X}_k, \boldsymbol{X}_j, \boldsymbol{X}_i, \boldsymbol{X}_l) \\
& + g(\boldsymbol{X}_k, \boldsymbol{X}_j, \boldsymbol{X}_l, \boldsymbol{X}_i) + g(\boldsymbol{X}_k, \boldsymbol{X}_l, \boldsymbol{X}_i, \boldsymbol{X}_j) + g(\boldsymbol{X}_k, \boldsymbol{X}_l, \boldsymbol{X}_j, \boldsymbol{X}_i) \\
& + g(\boldsymbol{X}_l, \boldsymbol{X}_i, \boldsymbol{X}_j, \boldsymbol{X}_k) + g(\boldsymbol{X}_l, \boldsymbol{X}_i, \boldsymbol{X}_k, \boldsymbol{X}_j) + g(\boldsymbol{X}_l, \boldsymbol{X}_j, \boldsymbol{X}_i, \boldsymbol{X}_k) \\
& + g(\boldsymbol{X}_l, \boldsymbol{X}_j, \boldsymbol{X}_k, \boldsymbol{X}_i) + g(\boldsymbol{X}_l, \boldsymbol{X}_k, \boldsymbol{X}_i, \boldsymbol{X}_j) + g(\boldsymbol{X}_l, \boldsymbol{X}_k, \boldsymbol{X}_j, \boldsymbol{X}_i).
\end{aligned}$$

Through the connection between a V-statistic $\hat{V}$ of order 4 and its associated U-statistics $\hat{U}_j$ of order $j = 1, ..., 4$ (Lee [18] p. 183) we can write

$$\hat{V} = \mathcal{S}_4^{(4)} \frac{(n-1)(n-2)(n-3)}{n^3} \hat{U}_4 + \mathcal{S}_4^{(3)} \frac{(n-1)(n-2)}{n^3} \hat{U}_3 + \mathcal{S}_4^{(2)} \frac{(n-1)}{n^3} \hat{U}_2 + \mathcal{S}_4^{(1)} \frac{1}{n^3} \hat{U}_1,$$

where the Stirling numbers of the second kind are given by $\mathcal{S}_4^{(4)} = 1$, $\mathcal{S}_4^{(3)} = 6$, $\mathcal{S}_4^{(2)} = 7$, $\mathcal{S}_4^{(1)} = 1$ (Abramowitz and Stegun [1] p. 835). Hence the asymptotic behaviour of $\hat{I}$ is that of the leading U-statistic $\hat{U}_4$ of order 4, whose kernel is equal to $\psi(\boldsymbol{X}_i, \boldsymbol{X}_j, \boldsymbol{X}_k, \boldsymbol{X}_l)$.

To determine the degree of degeneracy when the null hypothesis holds true, we need to investigate the nullity of appropriate conditional expectations of ψ, namely

$$\begin{aligned}
\sigma_1^2 &= \mathrm{Var}[\psi_1(\boldsymbol{X}_i)], \\
\sigma_2^2 &= \mathrm{Var}[\psi_2(\boldsymbol{X}_i, \boldsymbol{X}_j)], \\
\sigma_3^2 &= \mathrm{Var}[\psi_3(\boldsymbol{X}_i, \boldsymbol{X}_j, \boldsymbol{X}_k)],
\end{aligned}$$

where

$$\begin{aligned}
\psi_1(\boldsymbol{x}_i) &= E[\psi(\boldsymbol{X}_i, \boldsymbol{X}_j, \boldsymbol{X}_k, \boldsymbol{X}_l) | \boldsymbol{X}_i = \boldsymbol{x}_i], \\
\psi_2(\boldsymbol{x}_i, \boldsymbol{x}_j) &= E[\psi(\boldsymbol{X}_i, \boldsymbol{X}_j, \boldsymbol{X}_k, \boldsymbol{X}_l) | \boldsymbol{X}_i = \boldsymbol{x}_i, \boldsymbol{X}_j = \boldsymbol{x}_j], \\
\psi_3(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k) &= E[\psi(\boldsymbol{X}_i, \boldsymbol{X}_j, \boldsymbol{X}_k, \boldsymbol{X}_l) | \boldsymbol{X}_i = \boldsymbol{x}_i, \boldsymbol{X}_j = \boldsymbol{x}_j, \boldsymbol{X}_k = \boldsymbol{x}_k].
\end{aligned}$$

We can see that $\psi_1(\boldsymbol{x}_i) = 0$ under the null hypothesis, so that $\sigma_1^2 = 0$. On the contrary $\psi_2(\boldsymbol{x}_i, , \boldsymbol{x}_j) \neq 0$ since $E[g(\boldsymbol{X}_i, \boldsymbol{X}_j, \boldsymbol{X}_k, \boldsymbol{X}_l) | \boldsymbol{X}_i = \boldsymbol{x}_i, \boldsymbol{X}_j = \boldsymbol{x}_j] \neq 0$, for example, so that $\sigma_2^2 > 0$. Hence the degree of degeneracy is 1, and $n\hat{U}_4$ has a nondegenerate limit distribution (Lee [18] p. 90). This yields the stated result.

Acknowledgements

Olivier Scaillet thanks the editor and the three reviewers for constructive criticism and careful reading which lead to improving and correcting earlier versions of the paper. He also thanks Marcelo Fernandes and Patrick Gagliardini for helpful discussions. He has received support by the Swiss National Science Foundation through the National Center of Competence: Financial Valuation and Risk Management. Part of this research was done when he was visiting THEMA and ECARES.

References

[1] M. Abramowitz, I. Stegun, Handbook of Mathematical Functions, Dover, New York, 1972.

[2] K. Behnen, M. Huskova, G. Neuhaus, Rank estimators of scores for testing independence, Statistics and Decision 3 (1985) 239-262.

[3] X. Chen, Y. Fan, A. Patton, Simple tests for models of dependence between multiple financial time series, with applications to US equity returns and exchange rates, Working paper LSE (2003).

[4] R. D'Agostino, M. Stephens, Goodness-of-Fit Techniques, Marcel Dekker, New York, 1986.

[5] P. Deheuvels, La fonction de dépendance empirique et ses propriétés: Un test non paramétrique d'indépendance, Bulletin de l'Académie royale de Belgique: Classe des sciences 65 (1979) 274-292.

[6] M. Denuit, J. Dhaene, M. Goovaerts, R. Kaas, Actuarial Theory for Dependent Risks, Wiley, New York, 2005.

[7] Y. Fan, Testing the goodness-of-fit of a parametric density function by kernel methods, Econometric Theory 10 (1994) 316-356.

[8] Y. Fan, Goodness-of-fit tests for a multivariate distribution by the empirical characteristic function, Journal of Multivariate Analysis 62 (1997) 36-63.

[9] Y. Fan, Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters, Econometric Theory 14 (1998) 604-621.

[10] J.-D. Fermanian, Goodness of fit tests for copulas, Journal of Multivariate Analysis 95 (2005) 119-152.

[11] J.-D. Fermanian, D. Radulovic, M. Wegkamp, Weak convergence of empirical copula processes, Bernoulli 10 (2004) 847-860.

[12] A. Feueverger, R. Mureika, The empirical characteristic function and its applications, Annals of Statistics 5 (1977) 88-97.

[13] C. Genest, Frank's family of bivariate distributions, Biometrika 74 (1987) 549-555.

[14] C. Genest, K. Ghoudi, L.-P. Rivest, A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, Biometrika 82 (1998) 543-552.

[15] C. Genest, J.-F. Quessy, B. Rémillard, Goodness-of-fit procedures for copula models based on the probability integral transformation, Scandinavian Journal of Statistics forthcoming (2006).

[16] C. Genest, B. Rémillard, Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models, Working paper Université Laval (2005).

[17] I. Gijbels, J. Mielniczuk, Estimating the density of a copula function, Communications in Statistics: Theory and Methods 19 (1990) 445-464.

[18] A. Lee, U-Statistics: Theory and Practice, Marcel Dekker, New York, 1990.

[19] A. McNeil, R. Frey, P. Embrechts, Quantitative Risk Management: Concepts, Techniques and Tools, Princeton University Press, Princeton, 2005.

[20] R. Nelsen, An Introduction to Copulas, Lecture Notes in Statistics, Springer-Verlag, New York, 1999.

[21] D. Scott, Multivariate Density Estimation. Theory, Practice and Visualization, Wiley, New York, 1992.

[22] J. Shih, T. Louis, Inferences on the association parameter in copula models for bivariate survival data, Biometrics 51 (1995) 1384-1399.

[23] A. van der Vaart, J. Wellner, Weak Convergence and Empirical Processes, Springer-Verlag, New York, 1996.