F. Schmid and R. Schmidt, On the Asymptotic Behavior of Spearman's Rho and Related Multivariate Extensions, *Statistics and Probability Letters* (to appear), 2006.

Multivariate Extensions of Spearman's Rho and Related Statistics

Friedrich Schmid[1] and Rafael Schmidt[12]

Abstract

Multivariate measures of association are considered which, in the bivariate case, coincide with the population version of Spearman's rho. For these measures, nonparametric estimators are introduced via the empirical copula. Their asymptotic normality is established under rather weak assumptions concerning the copula. The asymptotic variances are explicitly calculated for some copulas of simple structure. For general copulas, a nonparametric bootstrap is established.

AMS 2000 subject class: Primary 62H20, 62G05, 62G20, Secondary 60G15, 62G30.

Key words: Spearman's rho, Multivariate measure of association, Copula, Nonparametric estimation, Empirical copula, Weak convergence, Asymptotic variance, Nonparametric bootstrap.

1 Introduction

Spearman's rho is a widely used measure for the strength of association between two random variables X and Y. It is invariant with respect to the marginal distributions of X and Y, and can be expressed as a function of their copula. This property is also known as 'scale-invariance'. There are various ways to extend Spearman's rho to a (multivariate) measure of association between d random variables $X_1, \ldots, X_d$. This is of interest in many fields of application, e.g. in the multivariate analysis of financial asset returns where one wants to express the amount of dependence in a portfolio by a single number.

The focus of this paper is on the nonparametric estimation of multivariate population versions of Spearman's rho via the empirical copula. Using empirical process theory it can be shown that the estimators are asymptotically normally distributed under rather weak assumptions concerning the copula of $X_1, \ldots, X_d$. We obtain compact expressions for the asymptotic variances of these estimators, which are determined by the copula and its partial derivatives. If the copula possesses a simple structure, these obtained formulas are suitable for explicit computations. Otherwise, we provide a bootstrap algorithm.

The structure of the paper is as follows. The next section introduces the notation and some definitions used in the following. Section 3 reviews three multivariate extensions of the population version of Spearman's rho and derives some analytical properties. Section 4 introduces the related nonparametric estimators which are based on the empirical copula and establishes their asymptotic normality. Thereafter, formulas for their asymptotic variances are given and a bootstrap procedure is presented. Finally, in Section 5 we explicitly calculate the asymptotic variance for various copulas.

[1]Department of Economic and Social Statistics, University of Cologne, Germany, rafael.schmidt@uni-koeln.de

[2]Department of Statistics, London School of Economics, UK. The author gratefully acknowledges financial support by the Deutsche Forschungsgemeinschaft (DFG).

2 Notation and Definitions

Throughout the paper we write bold letters for vectors, e.g., $\mathbf{x} := (x_1, \ldots, x_d)$ is a d-dimensional vector. Inequalities $\mathbf{x} \leq \mathbf{y}$ are understood componentwise, i.e, $x_i \leq y_i$ for all $i = 1, \ldots, d$. The indicator function on a set A is denoted by $\mathbf{1}_A$. The set $[a,b]^d$, $a < b$, refers to the d-dimensional cartesian product $[a,b] \times \cdots \times [a,b] \subset \mathbb{R}^d$. The space $\ell^\infty(T)$ comprises all uniformly bounded real-valued functions on some set T. We equip the space with the uniform metric $m(f_1, f_2) = \sup_{t \in T} |f_1(t) - f_2(t)|$. Let $X_1, X_2, \ldots, X_d$ be $d \geq 2$ random variables with joint distribution function

$$F(\mathbf{x}) = P(X_1 \leq x_1, \ldots, X_d \leq x_d), \quad \mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d,$$

and marginal distribution functions $F_i(x) = P(X_i \leq x)$ for $x \in \mathbb{R}$ and $i = 1, \ldots, d$. If not stated otherwise, we will always assume that the F_i are continuous functions. Thus, according to Sklar's theorem (Sklar, 1959), there exists a unique *copula* $C : [0,1]^d \to [0,1]$ such that

$$F(\mathbf{x}) = C(F_1(x_1), \ldots, F_d(x_d)) \quad \text{for all } \mathbf{x} \in \mathbb{R}^d.$$

The copula C is the joint distribution function of the random variables $U_i = F_i(X_i)$, $i = 1, \ldots, d$. Moreover, $C(\mathbf{u}) = F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))$ for all $\mathbf{u} \in [0,1]^d$. The generalized inverse function G^{-1} is defined via $G^{-1}(u) := \inf\{x \in \mathbb{R} \cup \{\infty\} \mid G(x) \geq u\}$ for all $u \in (0,1]$ and $G^{-1}(0) := \sup\{x \in \mathbb{R} \cup \{-\infty\} \mid G(x) = 0\}$. A detailed treatment of copulas is given in Nelsen (1999) and Joe (1997).

A copula is said to be *radially symmetric* if, and only if,

$$C(\mathbf{u}) = P(\mathbf{U} \leq \mathbf{u}) = P(\mathbf{U} > \mathbf{1} - \mathbf{u}) =: \bar{C}(\mathbf{1} - \mathbf{u}) \quad \text{for all } \mathbf{u} \in [0,1]^d.$$

It is well known that every copula C is bounded in the following sense:

$$\begin{aligned} W(\mathbf{u}) &:= \max\{u_1 + \ldots + u_d - (d-1), 0\} \\ &\leq C(\mathbf{u}) \leq \min\{u_1, \ldots, u_d\} =: M(\mathbf{u}) \quad \text{for all } \mathbf{u} \in [0,1]^d, \end{aligned}$$

where M and W are called the upper and lower *Fréchet-Hoeffding bounds*, respectively. The upper bound M is a copula itself and is also known as the comonotonic copula. It represents the copula of $X_1, \ldots, X_d$ if $F_1(X_1) = \cdots = F_d(X_d)$ with probability one, i.e., where there is (with probability one) a strictly increasing functional relationship between X_i and X_j $(i \neq j)$. By contrast, the lower bound W is a copula only for dimension $d = 2$. Another important copula is the independence copula

$$\Pi(\mathbf{u}) := \prod_{i=1}^d u_i, \quad \mathbf{u} \in [0,1]^d,$$

describing the dependence structure of stochastically independent random variables $X_1, \ldots, X_d$.

3 Multivariate Extensions of Spearman's rho

The following three multivariate population versions of Spearman's rho are considered $(d \geq 2)$:

$$\begin{aligned} \rho_1 &= h(d) \cdot \left\{ 2^d \int_{[0,1]^d} C(\mathbf{u})\, d\mathbf{u} - 1 \right\} \quad \text{with } h(d) = \frac{d+1}{2^d - (d+1)}, \\ \rho_2 &= h(d) \cdot \left\{ 2^d \int_{[0,1]^d} \Pi(\mathbf{u})\, dC(\mathbf{u}) - 1 \right\} \quad \text{and} \\ \rho_3 &= h(2) \cdot \left\{ 2^2 \sum_{k<l} \binom{d}{2}^{-1} \int_{[0,1]^2} C_{kl}(u,v)\, du dv - 1 \right\}, \end{aligned}$$

where $C_{kl}(u,v)$ refers to the bivariate marginal copula of C which corresponds to the k-th and l-th margin. The estimation of ρ_1 has been originally considered in Ruymgaart and van Zuijlen

(1978) and was later discussed by Wolff (1980), Joe (1990), and Nelsen (1996). A related class of multivariate measures of tail dependence is developed in Schmid and Schmidt (2006). The version ρ_2 appears first in Joe (1990) and later in Nelsen (1996). Further, ρ_3 is known as the population version of the average pair-wise Spearman's rho given, for example, in Kendall (1970), Chapter 6.

The motivation of ρ_3 as the weighted average of all pair-wise Spearman's rhos is obvious. By contrast, the motivation of ρ_1 and ρ_2 becomes clearer by the following. Spearman's rho of a two-dimensional random vector $\mathbf{X}$ with copula C can be written as

$$\rho = \frac{Cov\,(U_1, U_2)}{\sqrt{Var\,(U_1)}\sqrt{Var\,(U_2)}} = \frac{\int_0^1\int_0^1 uv\; dC\,(u,v) - \left(\frac{1}{2}\right)^2}{\sqrt{\frac{1}{12}}\sqrt{\frac{1}{12}}} = 12\int_0^1\int_0^1 C(u,v)\; dudv - 3,$$

where (U_1, U_2) have joint distribution function C. This expression is equal to

$$\rho = \frac{\int_{[0,1]^2} C(u,v)\; dudv - \int_{[0,1]^2} \Pi(u,v)\; dudv}{\int_{[0,1]^2} M(u,v)\; dudv - \int_{[0,1]^2} \Pi(u,v)\; dudv} = \frac{\int_{[0,1]^2} uv\; dC(u,v) - \int_{[0,1]^2} uv\; d\Pi(u,v)}{\int_{[0,1]^2} uv\; dM(u,v) - \int_{[0,1]^2} uv\; d\Pi(u,v)}$$

because of $\int_{[0,1]^2} M(u,v)\; dudv = 1/3$ and $\int_{[0,1]^2} \Pi(u,v)\; dudv = 1/4$. Thus, ρ can be interpreted as the normalized average distance between the copula C and the independent copula $\Pi(u,v) = uv$. The following d-dimensional extension of ρ to ρ_1 (and analogously to ρ_2) is now straightforward

$$\frac{\int_{[0,1]^d} C(\mathbf{u})\; d\mathbf{u} - \int_{[0,1]^d} \Pi(\mathbf{u})\; d\mathbf{u}}{\int_{[0,1]^d} M(\mathbf{u})\; d\mathbf{u} - \int_{[0,1]^d} \Pi(\mathbf{u})\; d\mathbf{u}} = \frac{d+1}{2^d - (d+1)}\left\{2^d \int_{[0,1]^d} C(\mathbf{u})\; d\mathbf{u} - 1\right\} = \rho_1.$$

For $d = 2$, Spearman's ρ coincides with $\rho_1 = \rho_2 = \rho_3$, though, for $d > 2$ the values of $\rho_i,\ i = 1, 2, 3$, are different in general. There exists, however, an interesting relationship between ρ_1 and ρ_2. Consider a random vector $\mathbf{X} = (X_1, \ldots, X_d)$ and an index set $I \subset \{1, \ldots, d\}$ with cardinality $2 \leq |I| \leq d$. We denote by $\rho_{1,I}$ the $|I|$-dimensional version of ρ_1 corresponding to those variables X_i where $i \in I$. The following relationship holds:

$$\rho_{2,\{1,\ldots,d\}} = \sum_{k=2}^{d} (-1)^k \frac{h(d)}{h(k)} \frac{2^d}{2^k} \sum_{\substack{I \subset \{1,\ldots,d\} \\ |I| = k}} \rho_{1,I}.$$

This can be derived via partial integration combined with the inclusion-exclusion principle. There is an immediate consequence of this relationship if C is radially symmetric. In this case $\rho_{1,\{1,\ldots,d\}}$ and $\rho_{2,\{1,\ldots,d\}}$ coincide. Moreover, if d is *odd*, then both measures of association can be expressed by the lower dimensional $\rho_{1,I}$:

$$\rho_{1,\{1,\ldots,d\}} = \rho_{2,\{1,\ldots,d\}} = \sum_{k=2}^{d-1} (-1)^k \frac{h(d)}{h(k)} \frac{2^{d-1}}{2^k} \sum_{\substack{I \subset \{1,\ldots,d\} \\ |I| = k}} \rho_{1,I}.$$

If d is *even*, we have

$$0 = \sum_{k=2}^{d-1} (-1)^k \frac{h(d)}{h(k)} \frac{2^d}{2^k} \sum_{\substack{I \subset \{1,\ldots,d\} \\ |I| = k}} \rho_{1,I}.$$

Further, ρ_1 and ρ_3 are closely related to each other for the following copula C_0, which is a convex combination of copulas:

$$C_0(\mathbf{u}) = \sum_{k<l} \binom{d}{2}^{-1} C_{kl}(u_k, u_l) \cdot \prod_{j \neq l,k} u_j, \tag{1}$$

where C_{kl} refers to the marginal copula which corresponds to the k-th and l-th margin of (any) copula C. In this particular case, $\rho_3(C) = h(2)\{2^d \int_{[0,1]^d} C_0(\mathbf{u})\ d\mathbf{u} - 1\}$, which coincides with $\rho_1(C_0)$ except for a normalizing factor. Note that $\rho_1(C_0) \leq h(d)/3$. We remark that ρ_3 has certain disadvantages as a multidimensional measure of association, as it is determined by the bivariate copulas only. Consider, for example, the 3-dimensional copula $C(u,v,w) = uvw + \lambda u(1-u)v(1-v)w(1-w)$, $|\lambda| \leq 1$, which is of FGM type discussed in Section 5. This copula possesses independent bivariate margins and, therefore, $\rho_3 = 0$. By contrast, $\rho_1 = \lambda \cdot 6^{-3} \neq 0$ if $\lambda \neq 0$.

4 Nonparametric estimation via the empirical copula

Consider a random sample $(\mathbf{X}_j)_{j=1,\dots,n}$ from a d-dimensional random vector $\mathbf{X}$ with joint distribution function F and copula C which are completely unknown. The marginal distribution functions F_i are estimated by their empirical counterparts

$$\hat{F}_{i,n}(x) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{\{X_{ij} \leq x\}}, \quad \text{for } i = 1,\dots,d \text{ and } x \in \mathbb{R}.$$

Further, set $\hat{U}_{ij,n} := \hat{F}_{i,n}(X_{ij})$ for $i = 1,\dots,d,\ j = 1,\dots,n$, and $\hat{\mathbf{U}}_{j,n} = (\hat{U}_{1j,n},\dots,\hat{U}_{dj,n})$. Note that

$$\hat{U}_{ij,n} = \frac{1}{n}(\text{rank of } \ X_{ij} \ \text{ in } \ X_{i1},\dots,X_{in}).$$

The estimation of $\rho_i,\ i = 1,2,3$, will therefore be based on ranks (and not on the original observations). In other words, we consider rank order statistics. The copula C is estimated by the empirical copula which is defined as

$$\hat{C}_n(\mathbf{u}) = \frac{1}{n}\sum_{j=1}^{n}\prod_{i=1}^{d} \mathbf{1}_{\{\hat{U}_{ij,n} \leq u_i\}} \quad \text{for} \quad \mathbf{u} = (u_1,\dots,u_d) \in [0,1]^d.$$

Empirical copulas were introduced and studied by Deheuvels (1979) under the name of 'empirical dependence functions'. The estimators of $\rho_i,\ i = 1,2,3$, are given by

$$\begin{aligned}
\hat{\rho}_{1,n} &= h(d)\cdot\Big\{2^d \int_{[0,1]^d} \hat{C}_n(\mathbf{u})\ d\mathbf{u} - 1\Big\} = h(d)\cdot\Big\{\frac{2^d}{n}\sum_{j=1}^{n}\prod_{i=1}^{d}(1-\hat{U}_{ij,n}) - 1\Big\} \\
\hat{\rho}_{2,n} &= h(d)\cdot\Big\{2^d \int_{[0,1]^d} \Pi(\mathbf{u})\ d\hat{C}_n(\mathbf{u}) - 1\Big\} = h(d)\cdot\Big\{\frac{2^d}{n}\sum_{j=1}^{n}\prod_{i=1}^{d}\hat{U}_{ij,n} - 1\Big\}, \\
\hat{\rho}_{3,n} + 3 &= 12\sum_{k<l}\binom{d}{2}^{-1}\int_{[0,1]^2}\hat{C}_{kl,n}(u,v)\ dudv = \frac{12}{n}\binom{d}{2}^{-1}\sum_{k<l}\sum_{j=1}^{n}(1-\hat{U}_{kj,n})(1-\hat{U}_{lj,n})
\end{aligned}$$

with $h(d) = (d+1)/(2^d - d - 1)$ and $\hat{C}_{kl,n}(u,v)$ being the bivariate marginal empirical copula of $\hat{C}_n$ which corresponds to the k-th and l-th margin. The right formula of $\hat{\rho}_{2,n}$ is developed in Equation (8) later. The estimator $\hat{\rho}_{1,n}$ for $d = 2$ differs slightly from the traditional sample version of Spearman's rho

$$\hat{\rho}_{S,n} = 1 - \frac{6n}{n^2 - 1}\sum_{j=1}^{n}(\hat{U}_{1j,n} - \hat{U}_{2j,n})^2, \tag{2}$$

which is used if no ties are present in the sample. It can be shown that $\hat{\rho}_{1,n} \leq \hat{\rho}_{S,n}$ for $n \in \mathbb{N}$ and $\lim_{n\to\infty}\sqrt{n}\{\hat{\rho}_{1,n} - \hat{\rho}_{S,n}\} = 0$ almost surely. Therefore, in the bivariate case, $\hat{\rho}_{1,n}$ and $\hat{\rho}_{S,n}$ have the same asymptotic distribution. An equivalent result for $\hat{\rho}_{3,n}$ and the traditional sample version of ρ_3, as given, e.g., in Kendall (1970), Chapter 6, holds. A method for approximating the asymptotic variance of $\hat{\rho}_{S,n}$ is developed in Borkowf (1999) and Borkowf (2002).

The derivation of the limiting laws for $\sqrt{n}(\hat{\rho}_{i,n} - \rho_i)$ involves the following theorem concerning the asymptotic behavior of the copula process $\sqrt{n}\{\hat{C}_n(\mathbf{u}) - C(\mathbf{u})\}$ which has been investigated in various setting, e.g., by Rüschendorf (1976), Stute (1984), Gänßler and Stute (1987), Fermanian, Radulović, and Wegkamp (2004), and Tsukahara (2005).

Theorem 1 *Let F be a continuous d-dimensional distribution function with copula C. Under the additional assumption that the i-th partial derivatives $D_iC(\mathbf{u})$ exist and are continuous for $i = 1, \ldots, d$, we have*

$$\sqrt{n}\{\hat{C}_n(\mathbf{u}) - C(\mathbf{u})\} \xrightarrow{w} \mathbb{G}_C(\mathbf{u}).$$

Weak convergence takes place in $\ell^\infty([0,1]^d)$ and

$$\mathbb{G}_C(\mathbf{u}) = \mathbb{B}_C(\mathbf{u}) - \sum_{i=1}^{d} D_iC(\mathbf{u})\mathbb{B}_C(\mathbf{u}^{(i)}).$$

The vector $\mathbf{u}^{(i)}$ denotes the vector where all coordinates, except the i-th coordinate of $\mathbf{u}$, are replaced by 1. The process $\mathbb{B}_C$ is a tight centered Gaussian process on $[0,1]^d$ with covariance function

$$E\{\mathbb{B}_C(\mathbf{u})\mathbb{B}_C(\mathbf{v})\} = C(\mathbf{u} \wedge \mathbf{v}) - C(\mathbf{u})C(\mathbf{v}),$$

i.e., $\mathbb{B}_C$ is a d-dimensional Brownian Bridge.

Theorem 1 is closely related to the next theorem which we need in order to establish the limiting law of $\sqrt{n}(\hat{\rho}_{2,n} - \rho_2)$.

Theorem 2 *Let F be a continuous d-dimensional distribution function with copula C. Suppose $\mathbf{U}$ is distributed with copula C and define $\bar{C}(\mathbf{u}) = P(\mathbf{U} > \mathbf{u})$. Consider the estimator*

$$\hat{\bar{C}}_n(\mathbf{u}) = \frac{1}{n}\sum_{j=1}^{n}\prod_{i=1}^{d}\mathbf{1}_{\{\hat{U}_{ij,n} > u_i\}} \quad \textit{for} \quad \mathbf{u} = (u_1, \ldots, u_d) \in [0,1]^d.$$

Under the assumptions and using the notation of Theorem 1 we have

$$\sqrt{n}\{\hat{\bar{C}}_n(\mathbf{u}) - \bar{C}(\mathbf{u})\} \xrightarrow{w} \mathbb{G}_{\bar{C}}(\mathbf{u}). \tag{3}$$

Weak convergence takes place in $\ell^\infty([0,1]^d)$ and $\mathbb{G}_{\bar{C}}(\mathbf{u}) = \mathbb{B}_{\bar{C}}(\mathbf{u}) + \sum_{i=1}^{d} D_i\bar{C}(\mathbf{u})\mathbb{B}_{\bar{C}}(\mathbf{u}_{(i)})$ with $\mathbf{u}_{(i)}$ denoting the vector where all coordinates, except the i-th coordinate of $\mathbf{u}$, are replaced by 0. The process $\mathbb{B}_{\bar{C}}$ is a tight centered Gaussian process on $[0,1]^d$ with covariance function $E\{\mathbb{B}_{\bar{C}}(\mathbf{u})\mathbb{B}_{\bar{C}}(\mathbf{v})\} = \bar{C}(\mathbf{u} \vee \mathbf{v}) - \bar{C}(\mathbf{u})\bar{C}(\mathbf{v})$.

Proof. Consider the estimator

$$\hat{\bar{C}}^\star_n(\mathbf{u}) = \frac{1}{n}\sum_{j=1}^{n}\prod_{i=1}^{d}\mathbf{1}_{\{U_{ij} > u_i\}} \quad \text{for} \quad \mathbf{u} \in [0,1]^d \quad \text{with } U_{ij} = F_i(X_{ij})$$

and $F_i,\ i = 1, \ldots, d$, denoting the marginal distribution functions of F. The corresponding empirical process converges weakly in $\ell^\infty([0,1]^d)$ to a d-dimensional Brownian bridge $\mathbb{B}_{\bar{C}}$ with covariance structure $E\{\mathbb{B}_{\bar{C}}(\mathbf{u})\mathbb{B}_{\bar{C}}(\mathbf{v})\} = \bar{C}(\mathbf{u} \vee \mathbf{v}) - \bar{C}(\mathbf{u})\bar{C}(\mathbf{v})$. The verification is standard; for example, marginal convergence is proven via a multivariate version of the Lindeberg-Feller theorem for triangular arrays, see Araujo and Giné (1980), p.41.

This empirical process and the empirical process in (3) are related to each other as follows:

$$\sqrt{n}\{\hat{\bar{C}}_n(\mathbf{u}) - \bar{C}(\mathbf{u})\} + O(1/\sqrt{n}) = \sqrt{n}\big[\hat{\bar{F}}_n\{\hat{F}_{1,n}^{-1}(u_1), \ldots, \hat{F}_{d,n}^{-1}(u_d)\} - \bar{C}(\mathbf{u})\big] = (*),$$

where we write

$$\hat{\bar{F}}_n(\mathbf{x}) = \frac{1}{n}\sum_{j=1}^{n}\prod_{i=1}^{d}\mathbf{1}_{\{X_{ij}>x_i\}}.$$

With $\bar{F}(\mathbf{x}) = P(\mathbf{X} > \mathbf{x})$ we have

$$\begin{aligned}(*) \quad = \quad & \sqrt{n}\{\hat{\bar{C}}^{\star}_n(\mathbf{u}) - \bar{C}(\mathbf{u})\} + \sqrt{n}\big[\bar{F}\{\hat{F}^{-1}_{1,n}(u_1), \ldots, \hat{F}^{-1}_{d,n}(u_d)\} - \bar{C}(\mathbf{u})\big] \\ & + \sum_{i=1}^{d-1}\Big[H_n\{F_1^{-1}(u_1), \ldots, F_{i-1}^{-1}(u_{i-1}), \hat{F}^{-1}_{i,n}(u_i), \ldots, \hat{F}^{-1}_{d,n}(u_d)\} \\ & \qquad - H_n\{F_1^{-1}(u_1), \ldots, F_i^{-1}(u_i), \hat{F}^{-1}_{i+1,n}(u_{i+1}), \ldots, \hat{F}^{-1}_{d,n}(u_d)\}\Big],\end{aligned} \tag{4}$$

where $H_n = \sqrt{n}(\hat{\bar{F}}_n - \bar{F})$. We mentioned that $\sqrt{n}\{\hat{\bar{C}}^{\star}_n(\mathbf{u}) - \bar{C}(\mathbf{u})\} \xrightarrow{w} \mathbb{B}_{\bar{C}}$ in $\ell^\infty([0,1]^d)$. Further,

$$\sqrt{n}[\bar{F}\{\hat{F}^{-1}_{1,n}(u_1), \ldots, \hat{F}^{-1}_{d,n}(u_d)\} - \bar{C}(\mathbf{u})] \xrightarrow{w} -\sum_{i=1}^{d} D_i\bar{C}(\mathbf{u})\mathbb{B}_C(\mathbf{u}^{(i)})$$

due to the Bahadur representation (Bahadur, 1966) of the empirical process for uniformly distributed random variables and an application of the functional Delta method (Van der Vaart and Wellner, 1996, p.374). Weak convergence takes place in $\ell^\infty([0,1]^d)$, cf. Fermanian, Radulović, and Wegkamp (2004). The last sum of Formula (4) converges to zero in probability due to the weak convergence of $\sqrt{n}(\hat{\bar{F}}_n - \bar{F})$ in $\ell^\infty([-\infty,\infty]^d)$. An application of the continuous mapping theorem yields the asserted weak convergence (utilize almost surely convergent versions of H_n.) Finally, the fact $\mathbb{B}_C(\mathbf{u}^{(i)}) = -\mathbb{B}_{\bar{C}}(\mathbf{u}_{(i)})$ a.s. for all $\mathbf{u}^{(i)},\ \mathbf{u}_{(i)} \in [0,1]^d,\ i = 1, \ldots, d$, completes the proof. ☐

Theorem 3 *Let $\hat{\rho}_{i,n},\ i = 1,2,3,$ be the estimators as defined above. Under the assumptions and using the notation of Theorem 1 and Theorem 2,*

$$\sqrt{n}(\hat{\rho}_{i,n} - \rho_i) \xrightarrow{d} Z_i \sim \mathcal{N}(0, \sigma_i^2).$$

The variances are

$$\sigma_1^2 = 2^{2d}h(d)^2 \int_{[0,1]^d}\int_{[0,1]^d} E\Big\{\mathbb{G}_C(\mathbf{u})\mathbb{G}_C(\mathbf{v})\Big\}\, d\mathbf{u}d\mathbf{v}, \tag{5}$$

$$\sigma_2^2 = 2^{2d}h(d)^2 \int_{[0,1]^d}\int_{[0,1]^d} E\Big\{\mathbb{G}_{\bar{C}}(\mathbf{u})\mathbb{G}_{\bar{C}}(\mathbf{v})\Big\}\, d\mathbf{u}d\mathbf{v}, \tag{6}$$

$$\sigma_3^2 = 144 \sum_{\substack{k<l \\ s<t}} \binom{d}{2}^{-2} \int_{[0,1]^d}\int_{[0,1]^d} E\Big\{\mathbb{G}_C\big(\mathbf{u}^{(k,l)}\big)\mathbb{G}_C\big(\mathbf{v}^{(s,t)}\big)\Big\}\, d\mathbf{u}d\mathbf{v} \tag{7}$$

with $\mathbf{u}^{(k,l)} := (1, ..., 1, u_k, 1, ..., 1, u_l, 1, ..., 1).$

For related results on bivariate linear rank order statistics of similar type we refer to Rüschendorf (1976), Gänßler and Stute (1987), and Genest and Rémillard (2004). Alternative derivations of the asymptotic behavior of rank order statistics such as $\hat{\rho}_{1,n}$ are given in Stepanova (2003).

Proof. The distributional convergence of $\sqrt{n}(\hat{\rho}_{1,n} - \rho_1)$ follows with Theorem 1 and the continuous mapping theorem (see Van der Vaart and Wellner (1996), Theorem 1.3.6), since the integral operator is a continuous linear map on $\ell^\infty([0,1]^d)$ into $\mathbb{R}$ and $\mathbb{G}_C$ is a tight Gaussian process. The formula for the variance σ_1^2 follows then by an application of Fubini's theorem.

For $\hat{\rho}_{2,n}$ we first establish a useful relationship. Let $\mathbf{U}_\omega$ be distributed according to the (empirical) copula $\hat{C}_n(\cdot)(\omega)$ for fixed ω. Note that the last expression is indeed a multivariate distribution function. Further, consider i.i.d. random variables $Z_1, \ldots, Z_d$ which are uniformly distributed on the interval $[0,1]$. Then for any fixed ω

$$\int_{[0,1]^d} \Pi(\mathbf{u})\, d\hat{C}_n(\mathbf{u})(\omega) = \int_{[0,1]^d} P(\mathbf{Z} < \mathbf{u})\, d\hat{C}_n(\mathbf{u})(\omega) = \int_{[0,1]^d} P(\mathbf{U}_\omega > \mathbf{u})\, d\mathbf{u} = \int_{[0,1]^d} \hat{\bar{C}}_n(\mathbf{u})(\omega)\, d\mathbf{u}.$$

Hence, we may rewrite the estimator $\hat{\rho}_{2,n}$ as follows

$$\hat{\rho}_{2,n}/h(d) + 1 = 2^d \int_{[0,1]^d} \Pi(\mathbf{u})\, d\hat{C}_n(\mathbf{u}) = 2^d \int_{[0,1]^d} \hat{\bar{C}}_n(\mathbf{u})\, d\mathbf{u} = \frac{2^d}{n} \sum_{j=1}^{n} \prod_{i=1}^{d} \hat{U}_{ij,n}. \tag{8}$$

Weak convergence of $\sqrt{n}(\hat{\rho}_{2,n} - \rho_2)$ and the form of σ_2^2 follows now by Theorem 2, along the same argumentation as above.

For the weak convergence of $\sqrt{n}(\hat{\rho}_{3,n} - \rho_3)$, observe that

$$\sqrt{n}(\hat{\rho}_{3,n} - \rho_3) \quad = \quad 12 \sum_{k<l} \binom{d}{2}^{-1} \int\limits_{[0,1]^d} \sqrt{n}\{\hat{C}_n(\mathbf{u}^{(k,l)}) - C(\mathbf{u}^{(k,l)})\}\, d\mathbf{u}$$

is a continuous linear map on $\ell^\infty([0,1]^d)$ of the empirical copula process $\sqrt{n}\{\hat{C}_n(\mathbf{u}) - C(\mathbf{u})\}$ and, thus, converges to $12 \sum_{k<l} \binom{d}{2}^{-1} \int_{[0,1]^d} \mathbb{G}_C(\mathbf{u}^{(k,l)})\, d\mathbf{u}$ according to Theorem 1. □

Remark. The process $\mathbb{G}_C(\mathbf{u}^{(k,l)})$ in Formula (7) takes the following form (using the notation of Theorem 1):

$$\mathbb{G}_C(\mathbf{u}^{(k,l)}) = \mathbb{B}_C(\mathbf{u}^{(k,l)}) - D_k C(\mathbf{u}^{(k,l)}) \mathbb{B}_C(\mathbf{u}^{(k)}) - D_l C(\mathbf{u}^{(k,l)}) \mathbb{B}_C(\mathbf{u}^{(l)})$$

because $\mathbb{B}_C(\mathbf{1}) = 0$ almost surely. Moreover, Formula (7) can be rewritten as

$$\begin{aligned}
\sigma_3^2 \quad = \quad & 144 \Big[\sum_{k<l} \binom{d}{2}^{-2} \int\limits_{[0,1]^2} \int\limits_{[0,1]^2} E\big\{\mathbb{G}_C(\mathbf{u}^{(k,l)}) \mathbb{G}_C(\mathbf{v}^{(k,l)})\big\}\, d(u_k, u_l) d(v_k, v_l) \\
& + \sum_{\substack{k<l, r<s \\ \{k,l\} \neq \{r,s\}}} \binom{d}{2}^{-2} \int\limits_{[0,1]^2} \int\limits_{[0,1]^2} E\big\{\mathbb{G}_C(\mathbf{u}^{(k,l)}) \mathbb{G}_C(\mathbf{v}^{(r,s)})\big\}\, d(u_k, u_l) d(u_r, u_s) \Big].
\end{aligned}$$

Proposition 4 *Let C be a radially symmetric copula. Under the additional assumptions and using the notation of Theorem 1 and Theorem 2, the processes $\mathbb{G}_C(\mathbf{u})$ and $\mathbb{G}_{\bar{C}}(\mathbf{1} - \mathbf{u})$ are equally distributed. Moreover, the asymptotic variances σ_1^2 and σ_2^2 in Formulas (5) and (6) coincide.*

Proof. Regarding the first assertion, note that $D_i \bar{C}(\mathbf{t})|_{\mathbf{t}=\mathbf{1}-\mathbf{u}} = -D_i \bar{C}(\mathbf{1} - \mathbf{u}) = -D_i C(\mathbf{u})$ and the covariance structure of $\mathbb{B}_C(\mathbf{u})$ equals

$$C(\mathbf{u} \wedge \mathbf{v}) - C(\mathbf{u})C(\mathbf{v}) = \bar{C}\{(\mathbf{1} - \mathbf{u}) \vee (\mathbf{1} - \mathbf{v})\} - \bar{C}(\mathbf{1} - \mathbf{u})\bar{C}(\mathbf{1} - \mathbf{v})$$

which corresponds to the covariance function of $\mathbb{B}_{\bar{C}}(\mathbf{1} - \mathbf{u})$. The second assertion follows by an appropriate substitution of the integrals in Formulas (5) and (6). □

Obviously, the integral expressions in Theorem 3 cannot be explicitly evaluated for the majority of known copulas, even in the case $d = 2$. However, the following theorem shows that the nonparametric bootstrap works. In this context, let $(\mathbf{X}_j^B)_{j=1,\ldots,n}$ denote the bootstrap sample which is obtained by sampling from $(\mathbf{X}_j)_{j=1,\ldots,n}$ with replacement. An empirical analysis of this bootstrap procedure is presented in Schmid and Schmidt (2006) and its applicability in the context of measuring price comovements in financial markets is discussed in Penzer et al. (2006).

Theorem 5 (The bootstrap) *Let $\hat{\rho}_{i,n}, i = 1, 2, 3$, be the estimators as defined in the beginning of the present section and $\hat{\rho}^B_{i,n}$ denote the corresponding estimators for the bootstrap sample $(\mathbf{X}^B_j)_{j=1,...,n}$. Then, under the assumptions of Theorems 1 and 2, the sequences $\sqrt{n}\{\hat{\rho}^B_{i,n} - \hat{\rho}_{i,n}\}, i = 1, 2, 3$, respectively, converge weakly to the same Gaussian limit as $\sqrt{n}\{\hat{\rho}_{i,n} - \rho_i\}, i = 1, 2, 3$, with probability one.*

Proof. Denote the empirical copula of $(\mathbf{X}^B_j)_{j=1,...,n}$ by $\hat{C}^B_n$. For dimension $d = 2$, Fermanian, Radulović, and Wegkamp (2004) show that the process $\sqrt{n}\{\hat{C}^B_n - \hat{C}_n\}$ converges weakly to the same Gaussian process as $\sqrt{n}\{\hat{C}_n - C\}$ with probability one. Weak convergence takes place in $\ell^\infty([0, 1]^2)$. The multidimensional generalization of this result is proven along the same lines as Theorem 2. The conclusion follows now by the continuous mapping theorem, see Van der Vaart and Wellner (1996), Theorem 1.3.6. □

5 Examples

The independence copula. Consider a random vector $\mathbf{X} = (X_1, ..., X_n)$ where $X_1, ..., X_n$ are stochastically independent (but not necessarily identically distributed). The related copula is the independence copula Π.

Proposition 6 *Let C be the independence copula $\Pi(\mathbf{u}) = \prod_{i=1}^d u_i$. Then, the asymptotic variances in Theorem 3 are given by*

$$\sigma_1^2 = \sigma_2^2 = -\frac{(d+1)^2\{3 + d - 3(4/3)^d\}}{3(1 + d - 2^d)^2} \quad \textit{and} \quad \sigma_3^2 = \binom{d}{2}^{-1} \tag{9}$$

Proof. Note that $C \equiv \Pi$ is radially symmetric and, hence, $\sigma_1^2 = \sigma_2^2$ according to Proposition 4. Obviously, $D_iC(\mathbf{u}) = \prod_{k \neq i} u_k$. Further, for $i \neq j$ and $\mathbf{u}^{(i)} := (1, ..., 1, u_i, 1, ..., 1)$ we have

$$\int_{[0,1]^d} \int_{[0,1]^d} E\left\{D_iC(\mathbf{u})B_C(\mathbf{u}^{(i)}) \cdot D_jC(\mathbf{v})B_C(\mathbf{v}^{(j)})\right\} d\mathbf{u}d\mathbf{v} = 0.$$

Moreover, $\int_{[0,1]^d} \int_{[0,1]^d} E\{B_C(\mathbf{u})B_C(\mathbf{v})\} \, d\mathbf{u}d\mathbf{v} = 3^{-d} - 2^{-2d}$ and for $i = 1, ..., d$ we derive

$$\begin{aligned} &\int_{[0,1]^d} \int_{[0,1]^d} E\left\{B_C(\mathbf{u}) \cdot D_iC(\mathbf{v})B_C(\mathbf{v}^{(i)})\right\} d\mathbf{u}d\mathbf{v} \\ = &\int_{[0,1]^d} \int_{[0,1]^d} E\left\{D_iC(\mathbf{u})B_C(\mathbf{u}^{(i)}) \cdot D_iC(\mathbf{v})B_C(\mathbf{v}^{(i)})\right\} d\mathbf{u}d\mathbf{v} = \frac{2^{2-2d}}{3} - 2^{-2d}. \end{aligned}$$

Collecting terms, we obtain $\int_{[0,1]^d} \int_{[0,1]^d} E\{\mathbb{G}_C(\mathbf{u})\mathbb{G}_C(\mathbf{v})\} \, d\mathbf{u}d\mathbf{v} = 3^{-d} - 2^{-2d} - d(2^{2-2d}3^{-1} - 2^{-2d})$. Finally, the respective normalization yields the left-hand formula in Equation (9). For σ_3^2 observe that

$$\int_{[0,1]^d} \int_{[0,1]^d} E\left\{\mathbb{G}_C(\mathbf{u}^{(k,l)})\mathbb{G}_C(\mathbf{v}^{(k,l)})\right\} d\mathbf{u}d\mathbf{v} = \frac{1}{144}$$

and for $k, l, r, s \in \{1, ..., d\}$ such that $k < l$ and $r < s$ and $\{k, l\} \cap \{r, s\} = \emptyset$ we have

$$\int_{[0,1]^d} \int_{[0,1]^d} E\left\{\mathbb{G}_C(\mathbf{u}^{(k,l)})\mathbb{G}_C(\mathbf{v}^{(r,s)})\right\} d\mathbf{u}d\mathbf{v} = 0.$$

The latter expression is also 0 if we consider the case where $\{k,l\} \cap \{r,s\}$ have exactly one element in common. Insertion of the above findings into Equation (5) results in the right-hand formula of Equation (9). □

A variance stabilizing transformation for the FGM copula. The family of Farlie-Gumbel-Morgenstern copulas (in short: FGM copulas) is defined by

$$C(u,v) = uv + \lambda uv(1-u)(1-v) \quad \text{for all} \quad |\lambda| \leq 1.$$

Because of their simple analytical form, FGM copulas have been widely used in statistics, for example, in order to obtain efficiency results on nonparametric tests for stochastic independence. A list of applications and references is given in Nelsen (1999), p.68. Recall that for $d = 2$ all the asymptotic variances σ_i^2, as given in Theorem 3, coincide. Hence, the subscripts i may be dropped. Direct calculation shows that

$$\sigma^2 = 1 - \frac{11}{45}\lambda^2.$$

Further, Spearman's rho takes the form $\rho = \lambda/3$ which implies that $|\rho| \leq 1/3$ and σ^2 can be expressed as a function of ρ, i.e, $\sigma^2 = 1 - \rho^2 \cdot 11/5$. A variance stabilizing transformation h, which satisfies

$$\sqrt{n}\,(h(\hat{\rho}_{i,n}) - h(\rho)) \xrightarrow{d} N(0,1), \quad i = 1,2,3$$

is then derived via the Delta method. In particular, we obtain

$$h(\rho) = \sqrt{\frac{5}{11}}\arcsin\left(\sqrt{\frac{11}{5}}\rho\right) \quad \text{for } |\rho| \leq 5/11.$$

This transformation may appropriately be extended to the domain $[-1,1]$ such that $h(\hat{\rho})$ is well defined. For $|\rho| \leq 1/3$, transformation h is close to Fisher's z-transformation $\ln\{(1+\rho)/(1-\rho)\}/2$.

The Kotz-Johnson copula. The following copula is called Kotz-Johnson copula (or Kotz and Johnson's iterated FGM copula), see Nelsen (1999), p.72.:

$$C(u,v) = uv + \lambda uv(1-u)(1-v)(1+\theta uv) \quad \text{for all} \quad -1 \leq \lambda, \theta \leq 1.$$

It forms a generalization of the FGM copula. Spearman's rho for this copula is $\rho = \lambda/3 + \lambda\theta/12$ and a direct calculation yields

$$\sigma^2 = 1 + \frac{1}{25}\lambda\theta - \frac{11}{45}\lambda^2 - \frac{11}{90}\lambda^2\theta - \frac{53}{5040}\lambda^2\theta^2 + \frac{1}{450}\lambda^3\theta + \frac{1}{1800}\lambda^3\theta^2.$$

A generic family of copulas. The next family of copulas, introduced in Rodríguez-Lallena and Úbeda-Flores (2004), can also be seen as a generalization of FGM copulas:

$$C(u,v) = uv + \lambda f(u)g(v),$$

where f and g are nonzero and absolutely continuous functions on $[0,1]$ such that $f(0) = f(1) = g(0) = g(1) = 0$. The range of the parameter λ, which obviously depends on the choice of f and g, is specified in the last reference. Spearman's ρ takes the form

$$\rho = 12 \cdot \lambda \int_0^1 f(u)du \int_0^1 g(v)dv.$$

The asymptotic variance σ^2 involves the following functionals:

$$T_1(f) \;=\; \int_0^1 f(u)\,du, \quad T_2(f) = \int_0^1 f^2(u)\,du, \text{ and } T_3(f) = \int_0^1 F(u)\,du = \int_0^1 \int_0^u f(x)\,dxdu$$

and $T_i(g)$, $i = 1,2,3$, which are analogously defined. For this family of copulas, we obtain

$$\begin{aligned}\sigma^2 \;=\; & 1 + \lambda \cdot 144\big\{2T_3(f) - T_1(f)\big\}\big\{2T_3(g) - T_1(g)\big\} \\ & + \lambda^2 \cdot 144\big\{2T_2(f)T_1^2(g) + 2T_2(g)T_1^2(f) - 7T_1^2(f)T_1^2(g)\big\}.\end{aligned}$$

References

Araujo, A., and E. Giné, 1980, *The Central Limit Theorem for Real and Banach Valued Random Variables.* (John Wiley & Sons, New York).

Bahadur, R.R., 1966, A note on quantiles in large samples, *Ann. Math. Statist.* 37, 577–580.

Borkowf, C.B., 1999, A new method for approximating the asymptotic variance of Spearman's rank correlation, *Statistica Sinica* 9, 535–558.

Borkowf, C.B., 2002, Computing the nonnull asymptotic variance and the asymptotic relative efficiency of Spearman's rank correlation, *Computational Statistics and Data Analysis* 39, 271–286.

Deheuvels, P., 1979, La fonction de dépendance empirique et ses propriétés, *Acad. Roy. Belg., Bull. Cl. Sci.* 65(5), 274–292.

Fermanian, J.D., D. Radulović, and M. Wegkamp, 2004, Weak convergence of empirical copula processes, *Bernoulli* 10(5), 847–860.

Gänßler, P., and W. Stute, 1987, *Seminar on Empirical Processes, DMV-Seminar Vol. 9.* (Birkhäuser Verlag, Basel).

Genest, C., and B. Rémillard, 2004, Tests of Independence and Randomness Based on the Empirical Copula Process, *Test* 13(2), 335–369.

Joe, H., 1990, Multivariate Concordance, *Journal of Multivariate Analysis* 35, 12–30.

Joe, H., 1997, *Multivariate Models and Dependence Concepts.* (Chapman and Hall, London).

Kendall, M.G., 1970, *Rank Correlation Methods.* (Griffin, London).

Nelsen, R.B., 1996, Nonparametric measures of multivariate association, *In: Distribution with Fixed Marginals and Related Topics, IMS Lecture Notes - Monograph Series* 28, 223–232.

Nelsen, R.B., 1999, *An Introduction to Copulas.* (Springer Verlag, New York).

Penzer, J., F. Schmid, and R. Schmidt, 2006, Measuring Large Comovements in Financial Markets, University of Cologne, *preprint.*

Rodríguez-Lallena, J.A., and M. Úbeda-Flores, 2004, A new class of bivariate copulas, *Statistics and Probability Letters* 66(3), 315–325.

Rüschendorf, L., 1976, Asymptotic normality of multivariate rank order statistics, *Annals of Statistics* 4, 912–923.

Ruymgaart, F.H., and M.C.A. van Zuijlen, 1978, Asymptotic normality of multivariate linear rank statistics in the non-i.i.d. case, *Annals of Statistics* 6(3), 588–602.

Schmid, F., and R. Schmidt, 2006, Multivariate conditional versions of Spearman's rho and related measures of tail dependence, Forthcoming in the *Journal of Multivariate Analysis.*

Schmid, F., and R. Schmidt, 2006, Bootstrapping Spearman's Multivariate Rho, Forthcoming in *Proc. Compstat 2006.*

Sklar, A., 1959, Fonctions de répartition à n dimensions et leurs marges, *Publ. Inst. Statist. Univ. Paris* 8, 229–231.

Stepanova, N.A., 2003, Multivariate rank tests for independence and their asymptotic efficiency, *Math. Methods Statist.* 12(2), 197–217.

Stute, W., 1984, The oscillation behavior of empirical processes: The multivariate case, *Annals of Probability* 12, 361–379.

Tsukahara, H., 2005, Semiparametric estimation in copula models, *Canadian Journal of Statistics* 33(3), 357–375.

Van der Vaart, A.W., and J.A. Wellner, 1996, *Weak Convergence and Empirical Processes.* (Springer Verlag, New York).

Wolff, E.F., 1980, N-dimensional measures of dependence, *Stochastica* 4(3), 175–188.