IWQW

Institut für Wirtschaftspolitik und Quantitative
Wirtschaftsforschung

Diskussionspapier
Discussion Papers

No. 10/2015

A multivariate rank test of independence based on a multiparametric polynomial copula

Benedikt Mangold
University of Erlangen-Nürnberg

ISSN 1867-6707

Friedrich-Alexander-Universität
IWQW
Institut für Wirtschaftspolitik und Quantitative Wirtschaftsforschung

A multivariate rank test of independence based on a multiparametric polynomial copula

Benedikt Mangold[1,2]

University of Erlangen-Nürnberg, Department of Statistics and Econometrics, Lange Gasse 20, 90403 Nürnberg, Germany

Sunday 19th February, 2017

Abstract

This paper introduces a copula based multivariate rank test for independence extending existing approaches from literature to p dimensions. Then, a multiparametric p-dimensional generalization of the FGM copula is provided that can model the behavior in each vertex of the p-dimensional unit cube using exactly one parameter per vertex – the family of polynomial copulas. The independence copula is nested in this family if and only if every parameter is zero. In this case, a popular way to test for independence is comparing an estimate of the vector of parameters to a vector containing zeros only. Unfortunately, due to the mere quantity of parameters, no established estimation procedure can be used in higher dimensions. Instead, the developed multivariate rank test is applied sequentially to every parameter to test for joint squared deviation from independence. Applying this new test to the polynomial copula results in the new vertex test which is a test for independence with focus on the high dimensional tail regions. It is compared to similar nonparametric rank tests of independence by means of calculation time and power under several alternatives and sample sizes.

Keywords: Rank-based inference, multiparametric, copula, independence, dependogramm, partial dependence, multivariate tail

Email address: benedikt.mangold@fau.de (Benedikt Mangold)

[1]This paper has formerly been published as *A multivariate linear rank test of independence based on a multiparametric copula with cubic sections.*

[2]The author has benefited from many helpful discussions with Ingo Klein.

Testing for independence is an area of huge interest in statistics and finds application in almost every scientific discipline. Probably the most common case of appliance is testing for independence when there are two samples that are each independent and identically distributed (i.i.d.). The parameter space of a bivariate parametric probability distribution often includes a value that is equivalent to the independence of the marginals, say 0. If so, a natural test for independence requires an estimate of the parameter which is compared to 0, i.e. independence, if the distribution of the estimator is known. The estimation is usually done using either maximum likelihood – if a certain distribution of the sample is assumed – or kernel density estimation. If no assumptions on the family of the population's distribution can be made or in presence of outliers, nonparametric methods of inference concerning the parameter are often available.

This procedure of testing for independence can be applied easily if the assumed distribution of the sample satisfies standard regularity conditions, if the amount of parameters q of the distribution is low, and if the dimension p of the sample is not too large. However, it is debatable if having only one or few dependence parameters is sufficient for modeling complex dependency structures, especially in higher dimensions. More flexible statistical models with more parameters as the vine copulas however suffer from the curse of dimensionality as the amount of parameters increases exponentially, see e.g. Czado (2010).

Shirahata (1974) introduces a nonparametric rank-based test of independence for a distribution function of dimension p, allowing distributions with $q > 1$ parameters. The test statistic is based on the aggregation of ranks of the samples, which is a natural, nonparametric and consistent way of estimating the marginal probability functions. This rank test is locally most powerful against the alternative that the sample is from the aforementioned distribution. However, for a multiparameteric probability distribution the proposed test statistic of Shirahata (1974) has one disadvantage: Summing up the components of the related vector of test statistics can cancel out deviations from zero, corresponding to deviations from independence. This can result in test decisions that are far too conservative, since the hypothesis of independence will erroneously not be rejected in the latter case. Without loss of generality, this mechanism of finding locally most powerful rank tests can be converted to copulas, which will be done throughout the paper.

This disadvantage and the potential lack of unique characterization of independence in a multiparametric setting vanishes if distributions with only one parameter are discussed. In particular, Garralda-Guillén (1998), later Genest and Verret (2005), transferred the procedure of Shirahata (1974) exclusively into the world of bivariate copulas with one single parameter of dependence. They identify several well known rank test statistics to be locally most powerful under specific alternatives and compare the power of these tests to other established tests in various scenarios.

This article is concerned with various generalizations. First, a principle of construction for locally most powerful rank tests for multivariate copulas with dimension $p \geq 2$ and one parameter is provided, extending the work of Garralda-Guillén (1998) and Genest and Verret (2005) which has not been addressed in literature yet. Second, this procedure is applied to copulas with $q > 1$ parameters consecutively for each parameter, resulting in a vector of rank test statistics $\boldsymbol{T}$. If this copula allows a unique characterization of independence by a vector of parameters, this enables jointly testing for independence using all entries of $\boldsymbol{T}$ simultaneously. The difference to the approach of Shirahata (1974) however is, that knowing the distribution of $\boldsymbol{T}$ allows a squared standardized aggregation which avoids the canceling-out-effect described earlier. Third, a new copula family $C_{\boldsymbol{\theta}}^{\alpha}$ with parameter vector $\boldsymbol{\theta} = (\theta_1, ..., \theta_q)'$, $q > 1$, and nuisance parameter α is introduced. $C_{\boldsymbol{\theta}}^{\alpha}$ is called polynomial copula and is equal to the independence copula if and only if $\boldsymbol{\theta} = (0, ..., 0)'$. The polynomial copula generalizes the copula with cubic sections introduced by Nelsen et al. (1997) by extension to higher dimensions and by the additional shape parameter α. Every component of $\boldsymbol{\theta}$ is uniquely linked to one vertex of the p dimensional unit cube $[0, 1]^p$. $\theta_j = 0$ indicates that the density function is equal to 1 in the vertex $\boldsymbol{v}_j$ that is associated to θ_j for $j = 1, ..., q$ – the value of the density function of the independence copula. $\theta_j \neq 0$ states that the probability of an occurrence located in regions close to $\boldsymbol{v}_j$ is lower or higher than if the marginals were independent. This focus on the vertex regions of the unit cube, which represent high dimensional tails of a probability distribution, is of interested in many applications (see e.g. Demarta and McNeil (2005)), namely hydrology, banking and finance, insurance, and geology. Thus, $C_{\boldsymbol{\theta}}^{\alpha}$ incorporates a way for flexible modeling of multivariate tail constellations.

The downside of using multiparametric copulas can be that the amount of parameters q

increases exponentially as p is getting larger. As an example, estimating $\boldsymbol{\theta}$ of the polynomial copula from a sample to test for independence is troublesome since $q = 2^p$. For maximum likelihood estimation, there is often no closed form expression for the estimator, and numerical maximization is not feasible for high dimensions due to the lack of a decent ratio of observations to parameters for large p and possible dependencies within $\boldsymbol{\theta}$. Additionally, the admissibility restrictions for $\boldsymbol{\theta}$ are complex and complicate the optimization problem. A possible alternative, the kernel density estimation (kde), suffers from similar problems, but provides bad approximations at the tails of distributions additionally, especially in higher dimensions. For copulas with parameters modeling the behavior in the tails, it is especially not recommended to use kde in a multiparametric, multidimensional setting.

For the polynomial copula, instead of using an estimate of $\boldsymbol{\theta}$ to test for independence, $\boldsymbol{T}$ is derived for $C^{\alpha}_{\boldsymbol{\theta}}$, denoted $\boldsymbol{T}_{\alpha}$. It is shown that there is a closed form expression for every entry $T_{\alpha,j},\ j = 1, ..., q$, which enables fast computation. The components of the realization $\boldsymbol{t}_{\alpha}$ can be computed separately, since they have no interpretation as parameter of $C^{\alpha}_{\boldsymbol{\theta}}$ and are hence not subject of admissibility restrictions any longer. Every component $t_{\alpha,j},\ j = 1, ..., q$, provides information on deviations from the density function value of 1 in the associated vertex. The larger the absolute value of a component of $\boldsymbol{t}_{\alpha}$, the larger the deviation from the density function of the independence copula. The sum of squared standardized deviations will be used testing for independence in the following, since it will be shown that $\boldsymbol{T}_{\alpha}$ is asymptotically multivariate normally distributed, by constructing a χ^2-type test of independence (vertex test).

In a simulation study, the power of the vertex test is compared to similar nonparametric tests, which are a multivariate version of Spearman's ρ introduced by Schmid and Schmidt (2007) and a test based on the average squared distance between independence copula and the empirical copula according to Deheuvels (1979) discussed in Genest and Rémillard (2004). The distributions of both alternative tests have no closed form expression, which means that the critical values need to be simulated at high computational costs for each sample size n and dimension p. Key advantage of the vertex test is that the asymptotic distribution of the test statistic is available in explicit form and therefore the critical values can be immediately determined. This distinguishes the vertex test from other nonparametric tests of independence and can have a crucial impact on real-time systems where calculation time is of importance.

This paper is organized as follows: Section 1 initially introduces the general class of polynomial copulas $C^{\alpha}_{\boldsymbol{\theta}}$ as foundation of the vertex test along with the k-reduced version, where every parameter in $\boldsymbol{\theta}$ is set to zero except the k-th. Further, admissibility restrictions and important properties of $C^{\alpha}_{\boldsymbol{\theta}}$ are discussed. Section 2 extends the concept of locally most powerful rank tests to p dimensions and applies it sequentially to multiparametric copulas in general, and to the polynomial copula of the previous section in particular. It is shown that the resulting vector of test statistics $\boldsymbol{T}$ containing the univariate rank test statistics is asymptotically multivariate normal distributed and a χ^2-type test of global independence is provided. Section 3 provides possible applications of the resulting vertex test and compares its performance to other similar nonparametric rank tests of independence with respect to power and calculation time. This article concludes with a short summary and an outlook in section 4.

1. Polynomial copula

This section introduces the flexible class of polynomial copulas, together with some preliminary definitions and terms that are used throughout the paper:

Definition 1 (Nelsen (2006))**.** *A p-dimensional copula C with parameter vector $\boldsymbol{\theta} \in \mathbb{R}^p$ is mapping from $[0,1]^p \mapsto [0,1]$ with the following properties:*

- *For all $u_i \in [0,1]$, $i = 1, ..., p$, holds*

$$C_{\boldsymbol{\theta}}(u_1, ..., u_{i-1}, 0, u_{i+1}, ..., u_p) = 0$$
$$C_{\boldsymbol{\theta}}(1, ..., 1, u_i, 1, ..., 1) = u_i.$$

- *For every $H = \prod_{i=1}^{p}[x_i, y_i] \subseteq [0,1]^p$ the integral*

$$\int_H \mathrm{d}C(\boldsymbol{u}) \geq 0$$

  with respect to C, $\boldsymbol{u} \in [0,1]^p$.

If $C_{\boldsymbol{\theta}}$ is a copula, $\boldsymbol{\theta}$ is called admissible. The copula density function is denoted as $c_{\boldsymbol{\theta}}$.

1.1. Bivariate polynomial copula

For the sake of simplicity we will initially introduce and discuss the bivariate case of the new class of polynomial copulas. The generalization to higher dimension is given in section 1.2.

Definition 2 (Bivariate polynomial copula)**.** *For dimension $p = 2$, the new family of polynomial copulas with dependence parameters $(\theta_1, \theta_2, \theta_3, \theta_4)'$ and shape parameter $\alpha > 0$ is given by*

$$
\begin{aligned}
C^{\alpha}_{\boldsymbol{\theta}}(u_1, u_2) = C^{\alpha}_{(\theta_1,\theta_2,\theta_3,\theta_4)'}(u_1, u_2) = u_1 u_2 \left(1 + (1-u_1)(1-u_2) \times \right. & \\
\left. \left(\theta_1 \left((1-u_1)(1-u_2)\right)^{\alpha} + \theta_2 \left((1-u_1) u_2\right)^{\alpha} + \theta_3 \left(u_1 (1-u_2)\right)^{\alpha} + \theta_4 \left(u_1 u_2\right)^{\alpha}\right)\right), & \qquad (1)
\end{aligned}
$$

$u_1, u_2 \in [0, 1]$. *The density function is denoted as $c^{\alpha}_{\boldsymbol{\theta}}$.*

The polynomial copula includes several copulas as special case, such as the iterated FGM copula (Kotz and Johnson, 1977), the Lin copula (Lin, 1987), the copula of Kimeldorf and Sampson (1975), the copula family of Sarmanov (1974), and the copula with cubic sections of Nelsen et al. (1997) is provided as a special case (for $\alpha = 1$). Figure 1 gives an impression of the density function of a polynomial copula. The polynomial degree of the copula is regulated by the shape parameter α. The higher α, the higher the overall degree of the polynomial section of the copula. For $\alpha \in \mathbb{N}^+$, the density function of the copula has a simple form which enables fast calculations and an intuitive understanding of its properties. The density function $c^{\alpha}_{\boldsymbol{\theta}}$ corresponding to equation (1) has the following properties:

$$
c^{\alpha}_{\boldsymbol{\theta}}(0,0) = 1 + \theta_1, \quad c^{\alpha}_{\boldsymbol{\theta}}(0,1) = 1 - \theta_2, \quad c^{\alpha}_{\boldsymbol{\theta}}(1,0) = 1 - \theta_3, \quad c^{\alpha}_{\boldsymbol{\theta}}(1,1) = 1 + \theta_4, \qquad (2)
$$

i.e. the parameters describe the magnitude and the direction of deviations from the density function of the independence copula in its $2^p = 4$ vertices, since independence is characterized by

$$
C^{\alpha}_{\boldsymbol{\theta}}(u_1, u_2) = u_1 u_2 \leftrightarrow \boldsymbol{\theta} = (0,0,0,0)' = \boldsymbol{\theta}_0.
$$

However, not every $\boldsymbol{\theta} \in \mathbb{R}^4$ results in proper copula function satisfying the requirements of definition 2. The next lemma provides admissibility restrictions on $\boldsymbol{\theta}$ for $\alpha = 1$, which can be evaluated in a closed form expression:

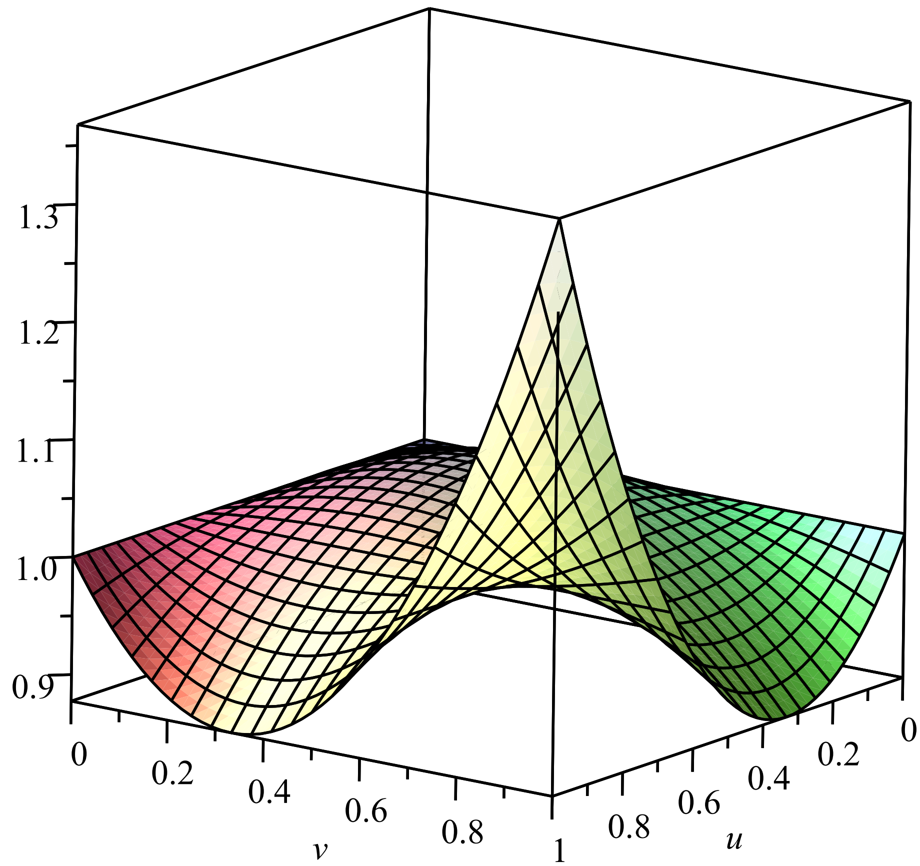

Figure 1: Density function of the polynomial copula from equation (1) with parameter $\boldsymbol{\theta} = (0, 0, 0, 0.37)'$ and $\alpha = 1$

Lemma 1 (Bivariate admissibility for $\alpha = 1$)**.** *The polynomial copula from equation (1) is admissible for $\alpha = 1$, if $(\theta_1, \theta_3)'$, $(\theta_4, \theta_2)'$, $(\theta_4, \theta_3)'$ and $(\theta_1, \theta_2)'$ are element of*

$$S^1 := \{[-1, 2] \times [-2, 1]\} \cap \left\{x, y \in \mathbb{R} | x^2 - xy + y^2 - 3x - 3y \leq 0\right\}$$

Proof. See Nelsen et al. (1997). □

Figure 2 gives a graphical intuition of the set S^1 of lemma 1. One has to distinguish between three cases:

- Triangle with vertices $(-1, 1)'$, $(1, 1)'$, $(-1, -1)'$: The minimum of the density function

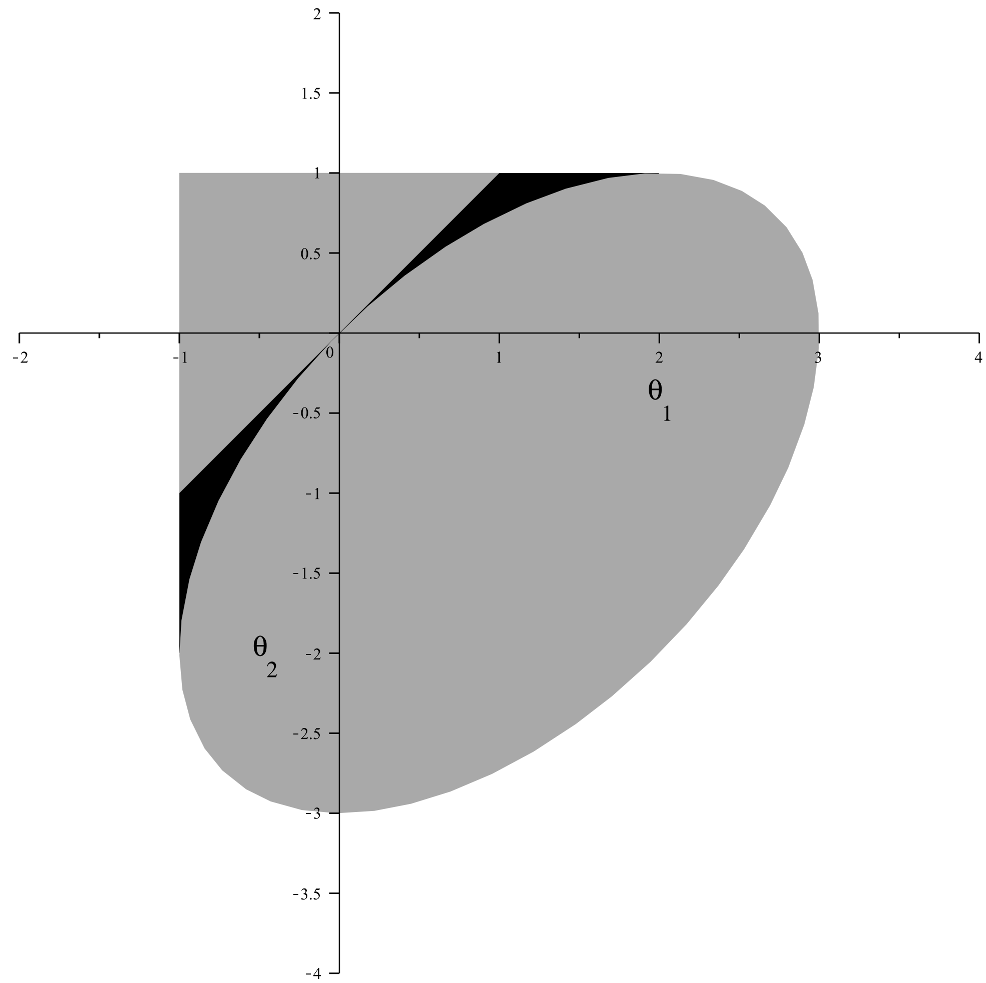

Figure 2: Area S^1 which ensures the admissibility of the bivariate polynomial copula for $\alpha = 1$ using the example of $(\theta_1, \theta_2)'$.

is nonnegative and is attained in a vertex.

- Inner area of the ellipse with equation $x^2 - xy + y^2 - 3x - 3y = 0$: The minimum is attained in between two vertices and is nonnegative.
- Black area: The minimum attains a negative value outside the interval $[0, 1]$. However, the density function is nonnegative on $[0, 1]^2$.

For $\alpha \neq 1$, determining the range of admissibility of $\boldsymbol{\theta}$ is not possible in a closed form expression. Sets S^α for different values of α that are simulated based on 10,000 randomly drawn parameter vectors $\boldsymbol{\theta}$ such that they result in positive density values of $c^\alpha_{\boldsymbol{\theta}}$ are presented in figure 6 in the appendix. However, since the aim is testing for independence rather than parameter estimation, for this purpose it is sufficient to show that there exists a region around $(0,0,0,0)'$ containing admissible $\boldsymbol{\theta}$ for every value of $\alpha > 0$:

Lemma 2 (Bivariate admissibility $\alpha \neq 1$)**.** *For every $0 < \alpha < \infty$, there exists an $\epsilon > 0$, such that $\boldsymbol{\theta} \in [-\epsilon, \epsilon]^4$ is admissible and $C^\alpha_{\boldsymbol{\theta}}$ is a copula function.*

Proof. Since $C^\alpha_{\boldsymbol{\theta}}$ is a copula for $\boldsymbol{\theta} = (0,0,0,0)$, the stated follows directly from the uniform continuity of polynomials on an closed interval $[0,1]$. □

$C^\alpha_{\boldsymbol{\theta}}$ has a strong and weak tail index coefficient of 0 due to its polynomial structure. Kendall's τ takes on moderate values – large α result in small ranges of τ whereas values of $\alpha \in [0.25, 1]$ result in τ roughly between -0.4 and 0.4. The effect of large α on τ results from the high polynomial degree which pushes the density function $c^\alpha_{\boldsymbol{\theta}}$ close to 1 in non tail regions – the value of the density function of the independence copula. Only in the vertex regions the density drifts to the values of equation (2). This implies that generally rather weak dependencies can be modeled with the polynomial copula, see table 1 for explicit formulas of τ and summary statistics of simulated values based on 10,000 repetitions.

1.2. Multivariate polynomial copula

The polynomial copula from definition 2 shall now be extended to p dimensions – at first for a vector of parameters $\boldsymbol{\theta} \in \mathbb{R}^q$; later only a reduced special case $q = 1$ is discussed.

We use the following notation in order to uniquely identify vertices, even in higher dimensions:

Definition 3. *The labeling of a vertex $\boldsymbol{v}$ of the p-dimensional cuboid $[0,1]^p$ is given by the coordinates as a binary number plus 1, which results in vertices $\boldsymbol{v}_i$, $i = 1, ..., 2^p$. A vertex $\boldsymbol{v} = (u_1, ..., u_p)'$, $u_i \in \{0,1\}$, $i = 1, ..., p$, is called even, if the amount of $u_i = 1$ is even or zero, odd otherwise.*

Example 1. *Let $p = 4$. The vertex $(1,0,1,0)'$ is the eleventh vertex $\boldsymbol{v}_{10+1=11}$, since $1010_2 = 10_{10}$. $\boldsymbol{v}_{11}$ is even, since the value 1 appears two times.*

In order to keep the notation simple, we require the following definition:

Table 1: Descriptive statistics of 10,000 simulated Kendall's τ values for several shape parameters α. Each value is calculated from a random vector of parameters $\boldsymbol{\theta}$ sampled with respect to the admissibility constraints from figure 6 in the appendix. The second column contains explicit formulas for τ, if available.

		Summary statistics					
α	Formula	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.1	-	-0.336	-0.098	0.004	0.003	0.104	0.333
0.25	-	-0.383	-0.098	0.000	-0.001	0.097	0.382
0.33	-	-0.378	-0.098	0.000	0.001	0.100	0.393
0.5	-	-0.388	-0.096	0.001	0.001	0.099	0.382
1	$\frac{\frac{\theta_1\theta_4-\theta_2\theta_3}{25}+\theta_1\theta_2\theta_3\theta_4}{18}$	-0.379	-0.094	-0.001	0.000	0.095	0.386
2	$\frac{\frac{\theta_1\theta_4-\theta_2\theta_3}{49}+\theta_1\theta_2\theta_3\theta_4}{50}$	-0.229	-0.050	-0.001	-0.001	0.048	0.223
3	$\frac{\frac{\theta_1\theta_4-\theta_2\theta_3}{98}+\theta_1\theta_2\theta_3\theta_4}{225}$	-0.098	-0.023	0.000	0.000	0.023	0.101
4	$\frac{\frac{\theta_1\theta_4-\theta_2\theta_3}{1089}+\theta_1\theta_2\theta_3\theta_4}{441}$	-0.053	-0.011	0.000	0.000	0.011	0.051

Definition 4. *Let $\boldsymbol{u} = (u_1, ..., u_p)'$. The function*

$$perm(\boldsymbol{u}) = \left(\begin{Bmatrix} 1-u_1 \\ u_1 \end{Bmatrix} \cdots \begin{Bmatrix} 1-u_p \\ u_p \end{Bmatrix} \right)$$

maps to a vector of dimension 2^p, whose components are the products of all possible combinations of the elements in braces.

Example 2. *For $\boldsymbol{u} \in [0,1]^2$ we have*

$$\text{perm}(\boldsymbol{u}) = \left(\begin{Bmatrix} 1-u_1 \\ u_1 \end{Bmatrix} \begin{Bmatrix} 1-u_2 \\ u_2 \end{Bmatrix} \right) = \begin{pmatrix} (1-u_1)(1-u_2) \\ (1-u_1)u_2 \\ u_1(1-u_2) \\ u_1u_2 \end{pmatrix}$$

A generalization of the bivariate polynomial copula to p dimensions is given by the following definition:

Definition 5 (p-dimensional polynomial copula). *For a vector $\boldsymbol{u} = (u_1, ..., u_p)' \in [0,1]^p$, $\boldsymbol{\theta} = (\theta_1, ..., \theta_q)'$, $\alpha > 0$ and $q = 2^p$, the copula function*

$$C^{\alpha}_{\boldsymbol{\theta}}(\boldsymbol{u}) = \prod_{i=1}^{p} u_i \left(1 + \langle \boldsymbol{\theta}, perm(\boldsymbol{u})^{\alpha} \rangle \prod_{i=1}^{p} (1 - u_i)\right) \tag{3}$$

with $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^{p} x_i y_i$ for $\boldsymbol{x} = (x_1, ..., x_p)'$ respective $\boldsymbol{y} = (y_1, ..., y_p)'$ is an extension of the polynomial copula from definition 2 to p dimensions.

The important vertex property of the polynomial copula from equation (2) is summarized by the following lemma:

Lemma 3. *For the p-dimensional density function $c^{\alpha}_{\boldsymbol{\theta}}(\boldsymbol{u})$, $\boldsymbol{u} = (u_1, ..., u_p)'$, $\boldsymbol{\theta} = (\theta_1, ..., \theta_q)'$, $q = 2^p$, we have for $\alpha > 0$ that*

$$c^{\alpha}_{\boldsymbol{\theta}}(\boldsymbol{v}_i) = 1 + a(\boldsymbol{v}_i)\theta_i, \quad i = 1, ..., 2^p,$$

with

$$a(\boldsymbol{v}_i) = \begin{cases} 1, & \text{if } \boldsymbol{v}_i \text{ is an even vertex} \\ -1, & \text{if } \boldsymbol{v}_i \text{ is an odd vertex} \end{cases}.$$

The parameter θ_i is called associated with the vertex $\boldsymbol{v}_i$.

The following lemma provides a closed form expression for the space of admissible parameter vectors $\boldsymbol{\theta}$ for the p dimensional polynomial copula if $\alpha = 1$:

Corollary 1 (Multivariate admissibility for $\alpha = 1$). *The p-dimensional copula $C^{1}_{\boldsymbol{\theta}}(\boldsymbol{u})$ is well-defined for $p \geq 2$ if the pair of parameters $(\theta', \theta'')'$ associated to the vertices of every two dimensional edge of the cuboid $[0,1]^p$ are element of the set*

$$S^1 := [-1; 2] \times [-2, 1] \cup \left\{ \theta''^2 - \theta' \theta'' + 3\,\theta'' + \theta'^2 - 3\,\theta' \leq 0 \right\} \tag{4}$$

Proof. The requirements of definition 1 are verified. $C^1(u_1, ..., u_{i-1}, 0, u_{i+1}, ..., u_p) = 0$ is true since if any $u_i = 0$ the first product in (3) and hence the whole expression gets 0. $C^1(1, ..., 1, u_i, 1, ..., 1) = u_i$ is true since the second product is zero for at least one occurrence of 1 and therefore the whole parenthesis in (3). Now it is shown that the density function does only attain nonnegative values which is sufficient for the second requirement of definition 2 due to the polynomial structure of $C^1_{\boldsymbol{\theta}}(\boldsymbol{u})$. The minimum of the density function is attained at a vertex or an edge. If the restriction

$$1 + a(\boldsymbol{v}_i)\theta_i \geq 0. \tag{5}$$

holds for all parameters θ_i, $i = 1, ..., 2^p$, all vertices always have nonnegative function values. Let $\boldsymbol{e}_i$ be an arbitrary, twodimensional edge of $[0,1]^p$ with the respective vertices $\boldsymbol{v}'$ and $\boldsymbol{v}''$, w.l.o.g. $\boldsymbol{e}_i = (0, ..., 0, u, 1, ..., 1)'$ with $\boldsymbol{v}' = (0, ..., 0, 0, 1, ..., 1)'$ and $\boldsymbol{v}'' = (0, ..., 0, 1, 1, ..., 1)'$ resp. the associated parameters θ' and θ''. Then we have

$$c^1_{\boldsymbol{\theta}}(\boldsymbol{e}_i) = (3\theta' - 3\theta'')^2 u^2 + (2\theta'' - 4\theta')u + \theta' + 1. \tag{6}$$

For $\theta' > \theta''$ the minimum of (6) is greater than 0 if

$$-\frac{1}{3}\frac{\theta'^2 - \theta'\,\theta'' + 3\,\theta'' + \theta''^2 - 3\,\theta'}{\theta' - \theta''} \geq 0.$$

The latter expression implies, that the pair (θ', θ'') needs to satisfy the ellipsoid inequality

$$\theta'^2 - \theta'\,\theta'' + 3\,\theta'' + \theta''^2 - 3\,\theta' \leq 0.$$

For $0 < \theta' \leq \theta''$ and $\theta' \geq -1$ resp. $\theta'' \leq 1$ the density function $c^1_{\boldsymbol{\theta}}(\boldsymbol{e}_i)$ is concave and the minimum is either attained at $\boldsymbol{v}'$ or at $\boldsymbol{v}''$ and is in either case nonnegative. In the remaining black area in figure 2 one has $\theta' > \theta''$ and therefore $c^1_{\boldsymbol{\theta}}(\boldsymbol{e}_i)$ is a convex parabola. Although the minimum has in this case a negative value, it is attained outside of the interval $[0,1]$, and the nonnegativity of (6) at both vertices implies the nonnegativity on $[0,1]$. □

As in the bivariate case, there is no closed form expression for the parameter space $\Theta \subset \mathbb{R}^{2^p}$ of the p-dimensional polynomial copula for $\alpha \neq 1$, containing all admissible parameter vectors $\boldsymbol{\theta}$. However, the following lemma ensures that for every α there exists a neighborhood of the p dimensional vector $(0, ..., 0)'$ that contains admissible vectors of parameters:

Lemma 4 (Multivariate admissibility for $\alpha \neq 1$)**.** *For every $0 < \alpha < \infty$, there exists an $\epsilon > 0$, such that $\boldsymbol{\theta} \in [-\epsilon, \epsilon]^q$, $q = 2^p$, is admissible and $C^\alpha_{\boldsymbol{\theta}}$ is a copula function.*

Proof. Since $C^\alpha_{\boldsymbol{\theta}}$ is a copula for $\boldsymbol{\theta} = (0, ..., 0)'$, the stated follows directly from the uniform continuity of polynomials on a closed interval. □

Even though the polynomial copula has several dependence parameters, the characterization of independence is unique. The latter is shown by the following proposition:

Proposition 1 (Nested independence copula)**.** *The polynomial copula is equal to the independence copula if and only if* $\boldsymbol{\theta} = \mathbf{0}$.

Proof. Inserting $\boldsymbol{\theta} = \mathbf{0}$ directly in the expression of definition 5 results in

$$C_{\mathbf{0}}^{\alpha}(\boldsymbol{u}) = \prod_{i=1}^{p} u_i = \Pi,$$

where Π denotes the independence copula. Let w.l.o.g. be θ_1 from $\boldsymbol{\theta} = (\theta_1, ..., \theta_q)'$ unequal to 0. Due to lemma 3 the value that the density function attains in vertex $\boldsymbol{v}_1$ is unequal to 1. Hence, $C_{\boldsymbol{\theta}}^{\alpha}(\boldsymbol{u})$ is not the independence copula. □

The locally most powerful rank tests (LMPRT) for copulas from Garralda-Guillén (1998) and Genest and Verret (2005) have been developed for $\theta \in \mathbb{R}$ and $p = 2$. Conversely, the polynomial copula is equipped with 2^p parameters. For applying a generalized LMPRT to the q-parametric copulas (see section 2.2), an auxiliary copula is required:

Definition 6 (k-reduced form)**.** *Let* $C_{\boldsymbol{\theta}}$ *be a* p*-dimensional,* q*-parametric copula function where the independence copula is uniquely determined by* $\boldsymbol{\theta} = \mathbf{0}$*. The copula resulting from the parameter vector*

$$\boldsymbol{\theta}_{[k]} = (\underbrace{0, ..., 0}_{k-1}, \theta_k, \underbrace{0, ..., 0}_{q-k})'$$

is called k*-reduced copula,* $k = 1, ..., q$*, and has only one remaining parameter* θ_k*.*

For $\alpha = 1$, the rules of admissibility for the k-reduced polynomial copula are even simpler than in the general case:

Lemma 5 (Admissibility of k-reduced polynomial copula)**.** *For* $\alpha = 1$*, the* k*-reduced polynomial copula is well-defined for* $\theta_k \in [-1, 3]$ *if* $\boldsymbol{v}_k$ *is an even vertex and* $\theta_k \in [-3, 1]$ *if* $\boldsymbol{v}_k$ *is an odd vertex,* $k = 1, ..., q$*. For* $\alpha \neq 1$ *there exists an interval* $[-\epsilon, \epsilon]$*,* $\epsilon > 0$*, which contains admissible values for* $C_{\boldsymbol{\theta}_{[k]}}^{\alpha}$*.*

Proof. First part of the lemma follows from corollary 1 when one of the parameters is taking the value 0. The second part from lemma 4, since $0 \in [-\epsilon, \epsilon]$, $\epsilon > 0$. □

Clearly, the k-reduced polynomial copula itself will not find application in practical work since it can only consider dependencies in the single vertex $\boldsymbol{v}_k$. However, a joint analysis of all vertices can be used to determine a deviation from independence. As pointed out in the introduction, using an estimate of $\boldsymbol{\theta}$ for the purpose of testing for independence is not an option due to problematic complications in the estimation process. As a way out, a new concept of locally most powerful rank tests for independence designed for p dimensional copulas with one parameter of dependence is introduced in the next section. This concept will be sequentially applied to every k-reduced form of a multiparametric copula to test simultaneously for independence.

2. The vertex test for independence

2.1. Multivariate locally most powerful rank tests

Sklar's theorem (Sklar, 1959) states that a copula describes the functional relation between continuous marginal distributions and the joint distribution of several random variables in a unique manner. In the following, the marginal distributions are assumed to be known, hence w.l.o.g. they can be considered as uniformly distributed. For $C(u_1, ..., u_p) = \Pi(u_1, ..., u_p) = \prod_{i=1}^{p} u_i$, the joint distribution is equal to the product of the margin distributions – the respective random variables are stochastically independent. Hence, a simple idea for a test of independence is to examine whether the empirical dependency structure of a sample of size n and dimension p corresponds to an independence copula of dimension p. In this section, the tests of independence for bivariate copulas, developed by Garralda-Guillén (1998) and Genest and Verret (2005), are generalized to p dimensions. For this purpose, an important requirement is the generalization of the concept of positive quadrant dependency to higher dimensions:

Definition 7 (Positive orthant dependency of Joe (1997))**.** *A distribution function $H(\boldsymbol{x})$, $\boldsymbol{x} = (x_1, ..., x_p)' \in \mathbb{R}^p$, with margins $X_i \sim F_i(x_i)$ is called positive lower orthant dependend*

(PLOD) if

$$H(\boldsymbol{x}) - \prod_{i=1}^{p} F_i(x_i) \geq 0 \quad \textit{for all } x_i \in \mathbb{R}, \quad i = 1, ..., p.$$

A copula is called PLOD if $C \geq \Pi$, where Π denotes the independence copula. If C is parametrized by $\theta \in \mathbb{R}$, $\theta' > \theta \Rightarrow C_{\theta'}(\boldsymbol{u}) \geq C_{\theta}(\boldsymbol{u})$ for $u \in [0,1]^p$ can be derived by PLOD.

Definition 7 generalizes the bivariate positive quadrant dependency concept used by Garralda-Guillén (1998) and Genest and Verret (2005). Note, that until now independence is still characterized by one single parameter.

Definition 8 (Locally most powerful rank test, Genest and Verret (2005))**.** *Let $\mathcal{M}$ be the set of all rank tests with a fixed level of significance. Let $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ be i.i.d. continuous random variables of dimension p. A test T_{opt} for the hypotheses*

$$H_0 : \theta = \theta_0 = 0 \ \textit{vs.} \ H_1 : \theta > \theta_0 \tag{7}$$

is called locally most powerful rank test (LMPRT) if

$$\forall T \in \mathcal{M} \, \exists \, \epsilon > 0 \, \forall \, 0 < \theta < \epsilon : 1 - \beta_{T_{\text{opt}}}(\theta) > 1 - \beta_T(\theta),$$

where $\beta(\cdot)$ denotes the power function of a test.

The requirements for the following proposition 2, that provides a method of constructing a LMPRT for p-variate copulas with $q = 1$ dependence parameter, are:

- A1 The parameter space $\Theta \subset \mathbb{R}$ is a closed interval and there exists a $\theta_0 \in \Theta$ such that $C_{\theta_0}(\boldsymbol{u}) = \Pi(\boldsymbol{u}) = \prod_{i=1}^{p} u_i$.
- A2 The family C_θ is PLOD.
- A3 For all $\theta \in \Theta$, $C_\theta(\boldsymbol{u})$ and the respective density function $c_\theta(\boldsymbol{u})$ are absolutely continuous in θ for all $\boldsymbol{u} \in (0,1)^p$.
- A4 $\dot{c}_\theta(u_1, ..., u_p) := \frac{\partial c_\theta(u_1,...,u_p)}{\partial \theta}$ is continuous in an environment around θ_0 with respect to θ and it holds that

  $$\lim_{\theta \to \theta_0} \int_{(0,1)^p} |\dot{c}_\theta(u_1, ..., u_p)| \mathrm{du_{i_1}} \cdots \mathrm{du_{i_p}} < \infty,$$

  for all $\{i_1, ..., i_p\} \in \mathcal{S}_p$, where $\mathcal{S}_p$ denotes the set of all permutations of $\{1, ..., p\}$.

Proposition 2 (Generalization of proposition 1, Genest and Verret (2005))**.** *Let* $\boldsymbol{R}_i = (R_{1i}, ..., R_{ji}, ..., R_{pi})'$ *be the ranks associated with a p-dimensional sample* $\boldsymbol{X}_i = (X_{1i}, ..., X_{ji}, ..., X_{pi})'$, $i = 1, ..., n,\ j = 1, ..., p$. *Let* $\boldsymbol{X}_i$ *be from a population that follows a copula from the class* C_θ *satisfying the requirements A1 to A4. Then the following statistic* T_n^* *is associated to the LMPRT for a fixed level of significance:*

$$T_n^* = \frac{1}{n}\sum_{i=1}^{n} T(R_{1i}, ..., R_{pi}), \tag{8}$$

with

$$T(r_1, ..., r_p) = \mathbb{E}\left[\frac{\partial}{\partial\theta}\log c_\theta(B_{r_1}, ..., B_{r_p})\bigg|_{\theta=\theta_0}\right],$$

where B_{r_j}, $j = 1, ..., p$, *are independent random variables with* $B_{r_j} \sim \beta(r_j, n - r_j + 1)$.

Proof. Straightforward generalization of the proof in Garralda-Guillén (1998), Genest and Verret (2005) in the sense of Shirahata (1974) for $q = 1$ to p dimensions. $\square$

In order to show the asymptotic normality of the test statistic of equation (8), the following lemma is needed:

Lemma 6. *Let* $\varphi(u_1, ..., u_p)$ *be a continuously differentiable element of* $\mathcal{L}^2([0,1]^p)$. *Then*

$$\lim_{n\to\infty} \mathbb{E}\left[(a_n^\varphi(R_{n11}, ..., R_{n1p}) - \varphi(U_{11}, ..., U_{1p}))^2\right] = 0$$

where

$$a_n^\varphi(i_1, ..., i_p) = \mathbb{E}\left[\varphi(U_{11}, ..., U_{1p})|R_{n11} = i_1, ..., R_{n1p} = i_p\right].$$

Proof. Following Hájek et al. (1999), p. 189, with the addition, that continuously differentiable functions of measurable functions are measurable again. $\square$

Further requirements for proposition 3, which ensures asymptotic normality of the test statistic from proposition 2, are:

A5 Let $\dot{c}_{\theta_0}(\boldsymbol{u})$ be such that for all $\boldsymbol{u} = (u_1, ..., u_p)' \in (0,1)^p$ we have:

$$\int_0^1 \dot{c}_{\theta_0}(u_1, ..., u_p)\mathrm{du_i} = 0,\ i = 1, ..., p, \text{ and } \int_{[0,1]^p} \dot{c}_{\theta_0}(\boldsymbol{u})\mathrm{du} \geq 0.$$

A6 $\dot{c}_{\theta_0}$ can be expressed by a finite sum of squared integrable functions that are monotone in every argument and it holds

$$\mathbb{E}\left[\dot{c}_{\theta_0}\left(\frac{R_{1i}}{n+1},...,\frac{R_{pi}}{n+1}\right)\dot{c}_{\theta_0}\left(\frac{R_{1j}}{n+1},...,\frac{R_{pj}}{n+1}\right)\right] = \mathrm{o}\left(\frac{1}{\mathrm{n}}\right), \quad \mathrm{i} \neq \mathrm{j}.$$

Proposition 3 (Generalization of proposition 2, Genest and Verret (2005))**.** *If $\dot{c}_{\theta_0}$ satisfies the requirements A5 and A6, $\sqrt{n}T_n^*$ converges to a normal distribution with expectation 0 and variance $\sigma^2(\dot{c}_{\theta_0})$ if the H_0 hypothesis is true, where*

$$\sigma^2(\dot{c}_{\theta_0}) = \int_{(0,1)^p} \left|\dot{c}_{\theta_0}(\boldsymbol{u})\right|^2 d\boldsymbol{u}.$$

Further T_n^* and the statistic

$$T_n = \frac{1}{n}\sum_{i=1}^{n} \dot{c}_{\theta_0}\left(\frac{R_{1i}}{n+1},...,\frac{R_{pi}}{n+1}\right)$$

are asymptotically equivalent.

Proof. Using theorem 1 of Behnen (1971) together with lemma 6. □

Proposition 3 provides a construction mechanism for a LMPRT for a p-dimensional copula with one parameter. As argued before, it is disputable whether or not one single parameter can reflect a complex, high dimensional association between the marginal distributions. If the independence copula is uniquely identified by a parameter vector, the following proposition allows the application of the developed multivariate LMPRT to more flexible, q-parametric copula families such as the polynomial copula from definition 5 by simultaneously testing the hypotheses $H_0: \ \theta_k = \theta_{0,k}$ for every k-reduced form, $k = 1, ..., q$, short $H_0: \ \boldsymbol{\theta} = \boldsymbol{\theta}_0$, $\boldsymbol{\theta}, \boldsymbol{\theta}_0 \in \Theta \subset \mathbb{R}^q$. In the following, $\boldsymbol{T}$ denotes a vector containing the q LMPRT statistics $T_{n,k}$ from proposition 2 of each k-reduced form from definition 6.

Proposition 4 (Asymtotic normality of $\boldsymbol{T}$)**.** *Let $C_{\boldsymbol{\theta}}$ be a p-dimensional, q-parametric copula family with k-reduced copulas satisfying the assumptions A1 to A6, $k = 1, ..., q$. Let $\boldsymbol{T} = (T_{n,1}, ..., T_{n,q})'$, short $(T_1, ..., T_q)'$, be a vector containing all LMPRT statistics $T_{n,k}$ from proposition 2 of each k-reduced form $C_{\boldsymbol{\theta}_{[k]}}$, $k = 1, ..., q$. Then, $\sqrt{n}\boldsymbol{T}$ is asymptotically multivariate normally distributed with $\boldsymbol{\mu} = \boldsymbol{\theta_0}$ and covariance matrix*

$$\Sigma(\dot{c}_{\boldsymbol{\theta}_0})_{i,j} = \int_{[0,1]^p} \left(\left.\frac{\partial c_{\boldsymbol{\theta}}(\boldsymbol{u})}{\partial \theta_i}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right) \times \left(\left.\frac{\partial c_{\boldsymbol{\theta}}(\boldsymbol{u})}{\partial \theta_j}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right) d\boldsymbol{u}$$

for $i, j = 1, ..., q$. Further, we have that

$$T = n\boldsymbol{T}'\Sigma(\dot{c}_{\boldsymbol{\theta}_0})^{-1}\boldsymbol{T} \overset{a}{\sim} \chi^2(q). \tag{9}$$

Proof. Proposition 3 states that every component of $\boldsymbol{T}$ is asymptotically normal distributed. The joint normal distribution of $\boldsymbol{T}$ with expectation $\boldsymbol{\mu} = \boldsymbol{\theta_0}$ and covariance matrix $\Sigma(\dot{c}_{\mathbf{0}})$ follows directly from theorem 4.1 in Sen and Puri (1967) together with lemma 6. Hence, it can be concluded that the quadratic form of T is asymptotically χ^2 distributed with q degrees of freedom. $\square$

In the following, w.l.o.g. the vectors $\boldsymbol{\theta}_0$ and $\mathbf{0} := (0, ..., 0)$ are used interchangeably. As initially pointed out, a classical way of testing for independence would be the estimation of $\boldsymbol{\theta}$ from data and testing if deviations between the estimate and $\boldsymbol{\theta}_0$ cannot be explained by the sampling error alone. However, for many multiparametric copulas in general and for high dimensional polynomial copulas in particular, the estimation of $\boldsymbol{\theta}$ is problematic for several reasons. First, the amount of parameters q that need to be estimated can grow exponentially in dimension p. Second, there often is no closed form expression for the classical maximum likelihood estimator, thus numerical optimization is required which is cumbersome. Third, alternative estimation procedures such as the kernel density estimation (kde) are not practicable since the domain of the copula is bounded and thus the estimation in tail regions, which are often of interest, imprecise. Fourth, admissibility restrictions on $\boldsymbol{\theta}$ often imply dependence within the components of $\boldsymbol{\theta}$ and needs to be considered in the estimation process.

This difficulties can partially be avoided by discussing deviations from $\boldsymbol{\theta}_0 = \mathbf{0}$ in $\boldsymbol{t}$, the realization of $\boldsymbol{T}$, instead of the estimate of $\boldsymbol{\theta}$. While the curse of dimensionality keeps persisting, the components of $\boldsymbol{t}$ are easily calculated due to its rank-based form. Importantly, they can not be interpreted as parameters of a copula any longer. This solves the admissibility problem for higher dimensions and allows computing the components of $\boldsymbol{t}$ sequentially. Proposition 4 simultaneously tests the hypotheses that each component of $\boldsymbol{\theta}$ deviates from $\mathbf{0}$. Thus, a large value of the sum of squared scaled deviations of $\boldsymbol{t}$ from $\boldsymbol{\theta}_0 = \mathbf{0}$ can be used as statistic of a test of independence for copula families with multiple parameters that allow a unique characterization of the independence copula, w.l.o.g with $\boldsymbol{\theta}_0 = \mathbf{0}$.

Corollary 2 (Testing for multivariate independence)**.** *Let $C_{\boldsymbol{\theta}}$ be a q-parametric, p-dimensional copula family, $\boldsymbol{\theta} \in \Theta^q \subseteq \mathbb{R}^q$, where $C_{\boldsymbol{\theta}}$ is the independence copula w.l.o.g. if and only if $\boldsymbol{\theta} = \mathbf{0}$. If $C_{\boldsymbol{\theta}}$ satisfies the assumptions A1–A6, the test statistic T from proposition 4 can be used to test for independence.*

One class of copulas whose k-reduced forms satisfy conditions A1–A6 is the polynomial copula from section 1. Next, an explicit test statistics T_α which results from the application of corollary 2 to polynomial copulas is provided and the covariance matrix for the asymptotic distribution of T_α is derived.

2.2. Locally most powerful rank test for the polynomial copula and the vertex test

In the following, proposition 3 shall be applied sequentially to the new k-reduced polynomial copula. To keep the notation simple, another definition is required:

Definition 9. *Let $\boldsymbol{u} = (u_1, ..., u_p)'$ and $\boldsymbol{c} = (c_1, ..., c_p)'$. Then*

$$\mathrm{pow}(\boldsymbol{u}, \boldsymbol{c}) := (u_1^{c_1}, ..., u_p^{c_p})' \quad \textit{and} \quad \mathrm{pow}_\pi(\boldsymbol{u}, \boldsymbol{c}) := \prod_{i=1}^{p} u_i^{c_i}.$$

The next lemma ensures that the k-reduced polynomial copula satisfies the assumption A2:

Lemma 7. *The k-reduced polynomial copula from Definition 6 is PLOD.*

Proof. Let $\theta_k' > \theta_k$. Let $\boldsymbol{v}_k = (v_{k,1}, ..., v_{k,p})'$, $v_i \in \{0, 1\}$, $i = 1, ..., p$, be the vertex associated to θ_k' resp. θ_k. Then we have

$$\begin{aligned} &C^\alpha_{\boldsymbol{\theta}'_{[k]}}(u_1, ..., u_p) - C^\alpha_{\boldsymbol{\theta}_{[k]}}(u_1, ..., u_p) = \\ &\qquad \underbrace{(\theta_k' - \theta_k)}_{>0} \times \underbrace{\mathrm{pow}_\pi(\boldsymbol{u}, \boldsymbol{v}_k)^\alpha}_{\geq 0} \times \underbrace{\mathrm{pow}_\pi(1 - \boldsymbol{u}, 1 - \boldsymbol{v}_k)^\alpha}_{\geq 0} \times \underbrace{\prod_{i=1}^{p} u_i(1 - u_i)}_{\geq 0} \geq 0, \end{aligned} \tag{10}$$

where $1 - \boldsymbol{x} = (1 - x_1, ..., 1 - x_p)'$. □

Example 3. *The bivariate 4-reduced polynomial copula from definition 2 (see figure 1) for $\alpha = 1$ is given by*

$$C^1_{\theta_4}(u_1, u_2) = u_1 u_2 + \theta_4 u_1^2 u_2^2 (1 - u_1)(1 - u_2) \tag{11}$$

for $\theta_4 \in [-3, 1]$ and $u_1, u_2 \in [0, 1]$. $C^1_{\theta_4}$ satisfies the properties A1–A6 due to its simple polynomial structure and lemma 7. Therefore the test based on the statistic

$$T_n = \frac{1}{n} \sum_{i=1}^{n} \left(3\frac{R_{2i}}{n+1} - 2\right) \frac{R_{1i}}{n+1} \left(3\frac{R_{1i}}{n+1} - 2\right) \frac{R_{2i}}{n+1} \tag{12}$$

is LMPRT for the 4-reduced polynomial copula (11). According to proposition 3, $\sqrt{n}T_n$ is asymptotically normal distributed with expectation 0 and variance $\frac{4}{225}$. The asymptotic p-value can be calculated using $1 - \Phi(t_n)$, where t_n denotes the realization of the test statistic T_n and Φ is the cumulative distribution function (cdf) of the standard normal distribution. Note that according to equation (2), the sign and absolute value of t_n give information on the direction and magnitude of a deviation from independence: a negative (positive) sign implies deviation from independence due to a higher (lower) density of observations in the vertex $\boldsymbol{v}_4$ than in the case of independence. The larger the absolute value of t_n, the larger the deviation from independence.

Applying corollary 2 to the family of polynomial copulas for general α results in the test statistic T_α which is based on $\boldsymbol{T}_\alpha$ from proposition 4. $\boldsymbol{t}_\alpha$ and t_α denote their respective realizations. The k-th entry of $\boldsymbol{T}_\alpha$, $T_{\alpha,k}$, is the LMPRT only against the alternative hypothesis of the k-reduced polynomial copula.

Example 4 (Vertex test for a bivariate sample)**.** *For $p = 2$ and $\alpha = 1$ we have under the null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 = (0, 0, 0, 0)'$ that the vector of LMPRT statistics is*

$$\boldsymbol{T}_1 = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} \left(\frac{R_{2i}}{n+1} - 1\right)\left(3\frac{R_{1i}}{n+1} - 1\right)\left(\frac{R_{1i}}{n+1} - 1\right)\left(3\frac{R_{2i}}{n+1} - 1\right) \\ \frac{1}{n}\sum_{i=1}^{n} -\left(3\frac{R_{2i}}{n+1} - 2\right)\left(3\frac{R_{1i}}{n+1} - 1\right)\left(\frac{R_{1i}}{n+1} - 1\right)\frac{R_{2i}}{n+1} \\ \frac{1}{n}\sum_{i=1}^{n} -\left(\frac{R_{2i}}{n+1} - 1\right)\left(3\frac{R_{2i}}{n+1} - 1\right)\frac{R_{1i}}{n+1}\left(3\frac{R_{1i}}{n+1} - 2\right) \\ \frac{1}{n}\sum_{i=1}^{n} \left(3\frac{R_{2i}}{n+1} - 2\right)\frac{R_{1i}}{n+1}\left(3\frac{R_{1i}}{n+1} - 2\right)\frac{R_{2i}}{n+1} \end{pmatrix}.$$

Further, $\sqrt{n}\boldsymbol{T}_1$ is asymptotically normal distributed with expectation $\boldsymbol{\mu}' = (0, 0, 0, 0)'$ and

$$\Sigma(\dot{c}_{\mathbf{0}})_{i=1,\ldots,4,j=1,\ldots,4} = \begin{pmatrix} 64 & -16 & -16 & 4 \\ -16 & 64 & 4 & -16 \\ -16 & 4 & 64 & -16 \\ 4 & -16 & -16 & 64 \end{pmatrix}^{-1}.$$

Therefore, we have

$$T_1 = n\boldsymbol{T}_1{}'\Sigma(\dot{c}_{\mathbf{0}})^{-1}\boldsymbol{T}_1 \overset{a}{\sim} \chi^2(4)$$

and the asymptotic p-value is obtained using $1 - F_{\chi^2(4)}(t_1)$.

Tables of test statistics $\boldsymbol{T}_\alpha$ and covariance matrices $\Sigma(\dot{c}_{\mathbf{0}})$ associated to bivariate vertex tests with other values of α can be found in the appendix (table 5 and 6), together with test statistic and covariance matrix of the vertex test for $p = 3$ and $\alpha = 1$. Note that a generalization to dimension $p > 3$ is straightforward but omitted in this paper for the purposes of simplicity.

Rejecting H_0 for the k-th component of $\boldsymbol{t}_\alpha$ implies a substantial deviation of $t_{\alpha,k}$ from $\theta_{0,k} = 0$, $k = 1, ..., q$. Thus, $t_{\alpha,k}$ gives insights into the frequency of occurrences in a region close to the vertex $\boldsymbol{v}_k$ with respect to the frequency that would have been expected if the marginal distributions were independent.

This concrete interpretation of the components of $\boldsymbol{t}_\alpha$ as deviations from the density function of the independence copula, evaluated at the vertices of the unit cube, encourages the usage of T_α as general test for independence. In this case however, the single components of $\boldsymbol{T}_\alpha$ are no longer a LMPRT in general but inherit information on the deviation from independence in regions located close to the vertices only.

Conclusively, the effect of the nuisance parameter α on the sensitive regions of the vertex test is investigated. For this purpose, the influence function (Hampel, 1986) is simulated, which is a Gâteaux derivative for measuring the impact of small contamination in a sample on the statistical functional $\boldsymbol{T}_\alpha$. Figure 3 shows as an example influence functions for $\alpha = 1$ and $\alpha = 3$ based on 1,000 iterations.

As pointed out earlier, the vertex test focuses on deviations from independence in the vertices of the unit cube which correspond to high dimensional tails of a copula. One can deduce from figure 3 that the higher α, the more sensitive T_α reacts to small changes in the tail regions. This finding is supported by the simulation study in section 3.2 where large values of α are especially suitable if the tails of the sample distribution are extremely heavy.

Note that the vector $\boldsymbol{t}_\alpha$ itself could also find a possible application to financial stock market data, where a certain pattern of joint extreme directions of the returns is desirable

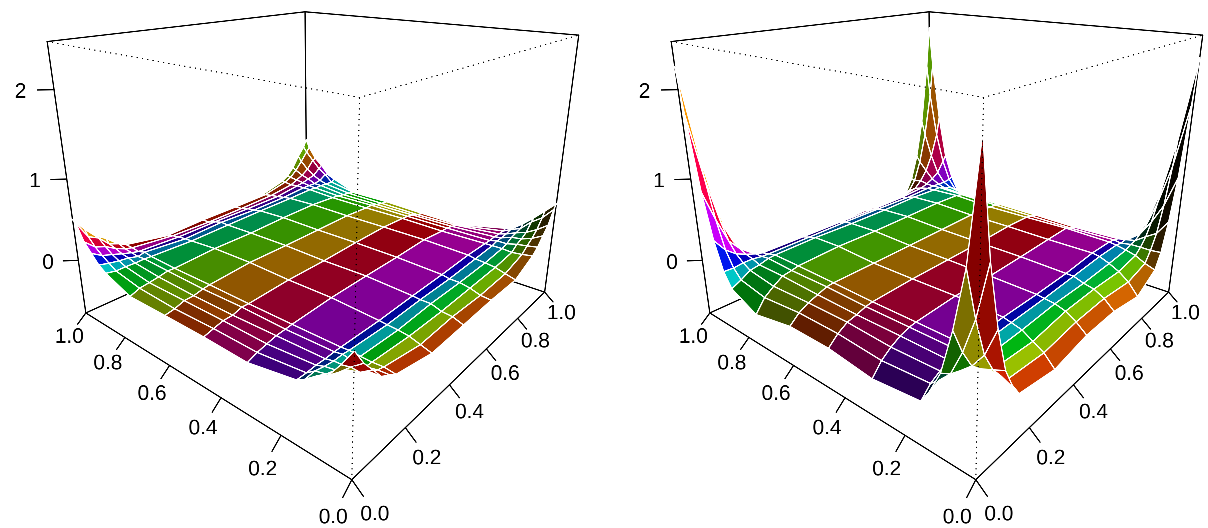

Figure 3: Simulated influence function (1,000 iterations) of T_α for $\alpha = 1$ (left) and $\alpha = 3$ (right) for a bivariate sample of size $n = 50$ with independent, uniformly distributed margins.

i.e. to identify a portfolio that has an opposing movement of its partial returns out of a pool of many assets. Another application of $\boldsymbol{T}_\alpha$ as the basis of new concepts of symmetry similarly to radial symmetry including a testing procedure can be found in Mangold (2017).

The following section illustrates the theoretical findings of T_α in a short simulation study and compares its power to other nonparametric tests.

3. Application

In this section an implementation of the vertex test in the language `R` (R Core Team, 2016) is provided and compared to other nonparametric rank tests of independence by means of power and calculation time. First, an implementation of an dependogram based on the polynomial vertex test mimicking Genest and Rémillard (2004) testing simultaneously for every possible partial dependency is introduced.

3.1. Dependogram

In this section, only the vertex test for $\alpha = 1$ is discussed as an example. A generalization for $\alpha \neq 1$ is straightforward.

Definition 10 (Dependogram). *Let* $\boldsymbol{X}_i = (X_{1i}, ..., X_{ji}, ..., X_{pi})'$, $i = 1, ..., n$, $j = 1, ..., p$, *be from sample of size* n *of a* p*-variate distribution. The dependogram provides (graphical) information if every partial sample, indexed by*

$$\mathcal{P}_{>1}(\{1, ..., p\}) = \{U \subseteq \{1, ..., p\} : |U| > 1\}, \tag{13}$$

is stochastically independent by calculating and plotting all test statistics T_1 *and critical values of the vertex test applied to every partial sample. Thereby, the level of significance is corrected via Bonferroni correction – all in all one has to perform* $2^p - p - 1$ *tests.*

Initially, the dependogram has been introduced by Genest and Rémillard (2004) using a test statistic that is based on a Cramér-von Mises distance between the empirical copula by Deheuvels (1979) and the independence copula. An implementation to create this dependogram is available by the command `dependogram` provided by the `R` package `copula` (Yan, 2007; Kojadinovic and Yan, 2010; Hofert and Mächler, 2011; Hofert et al., 2015).

Note that the distribution of the used test statistic has no closed form expression under the null hypotheses which implies that the critical values have to be obtained via Monte-Carlo simulation for every sample size n and dimension p. Since the effort in the sense of calculation time is huge, the application of this method can be cumbersome, especially for large sample sizes and/or higher dimensions. Information about the required calculation time is provided in section 3.4.

The implementation of the dependogram based on the vertex test is now introduced and its usage presented exemplarily by a dependency structure used in Genest and Rémillard (2004).

Example 5 (Genest and Rémillard (2004)). *Let* $\boldsymbol{x}$ *be a random sample of a* 5*-variate Gaussian distributed random variable with* $\boldsymbol{\mu} = (0, 0, 0, 0, 0)'$ *and with an identity matrix as covariance matrix. The dependency structure is obtained by:*

```
x <- matrix(rnorm(500),100,5)
x[,1] <- abs(x[,1]) * sign(x[,2] * x[,3])
x[,5] <- x[,4]/2 + sqrt(3) * x[,5]/2
```

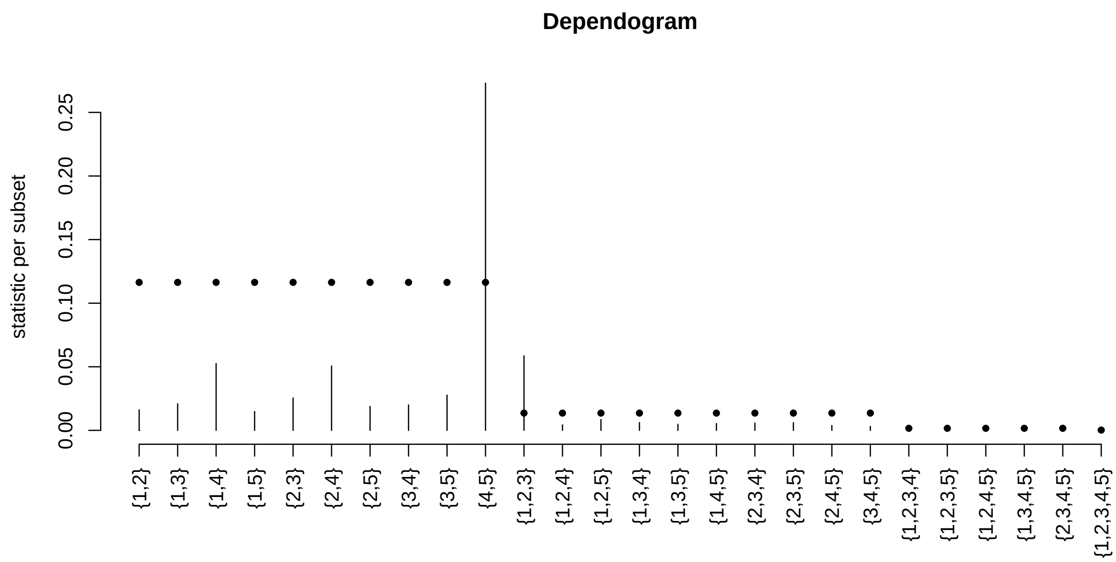

Figure 4: Dependogram based on the test statistic of Genest and Rémillard (2004) from the R package `copula`. The critical values are presented as points, the values of the test statistics as bars.

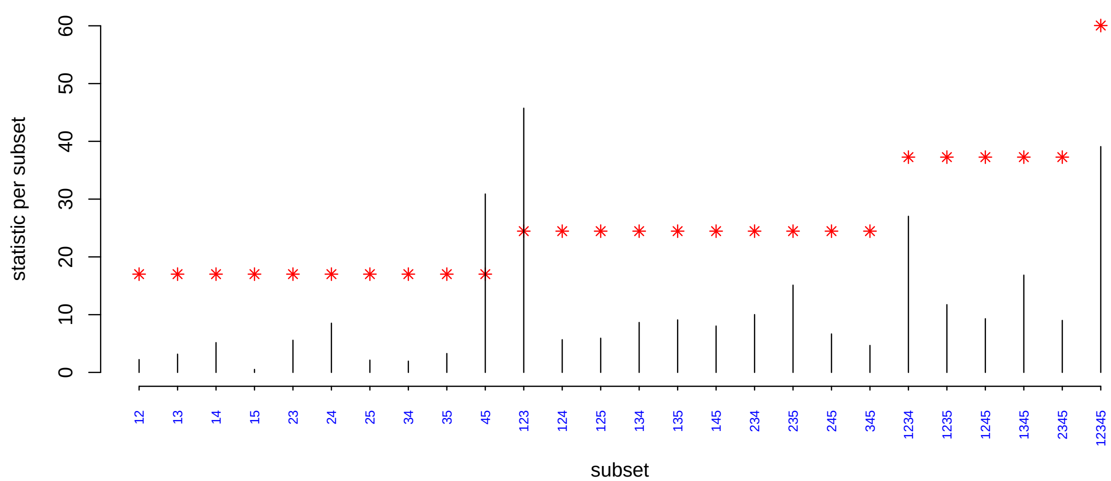

Figure 5: New implementation the dependogram based on the polynomial vertex with $\alpha = 1$. The critical values are presented as stars, the values of the test statistics as bars.

The figures 4 and 5 show the dependogram of the R *functions and the implementation of the polynomial vertex test applied to the example: Both reject the null hypothesis for the pairwise* *test on* $\{4,5\}$ *and the triple test of* $\{1,2,3\}$ *whereas the global hypotheses of all 5 samples would not have been rejected. Even if the global hypothesis would have been rejected, the finding of the partial dependencies allows a deeper insight into the relationship between the variables.*

As further discussed in section 3.4, the calculation time for creating the dependograms of figures 4 and 5 differs rigorously: The determination of the critical values of the test statistic of Genest and Rémillard (2004) for the example with $p = 5$ implies a Monte Carlo simulation for the dimensions $p = 2, 3, 4, 5$. The vertex test however compares T_1 with a quantile of a univariate χ^2 distribution which is available right away.

3.2. Comparison of power

In a simulation study it shall be investigated how the vertex test performs in comparison to the test of Genest and Rémillard (2004) (ff. Genest test) from section 3.1 and a third test, based on a multivariate version of Spearman's ρ by Schmid and Schmidt (2007) (ff. Schmid test):

Definition 11 (Multivariate version of Spearman's ρ by Schmid and Schmidt (2007))**.** *Let* $X_{1i}, ..., X_{ni}$ *be an i.i.d. sample of a* p *variate population,* $i = 1, ..., p$*. One generalization of the correlation coefficient by Spearman is given by*

$$\rho_{1,p} = h(p)\left(2^p \int_{[0,1]^p} \hat{C}_n(\boldsymbol{u})\mathrm{du} - 1\right) = h(p)\left(\frac{2^p}{n}\left(\sum_{j=1}^{n}\prod_{i=1}^{p}\left(1 - \frac{R_{ji}}{n}\right)\right) - 1\right)$$

whereas R_{ji} *denotes the rank of* X_{ji} *in* $X_{1i}, ..., X_{ni}$*,* $i = 1, ..., p$*. The distribution of* $\rho_{1,p}$ *itself depends on the dimension* p *and the sample size* n *under the null hypothesis of independence and needs to be determined via simulation.*

Table 2: Power of the vertex test of independence ($\alpha = 1$), the Genest test and the Schmid test if the sample is from a Student t distribution with sample size n, degrees of freedom ν and Kendall's τ. The power (in percent) is based on $10,000$ repetitions at a significance level of 5%.

	$n = 25$						$n = 50$					
	$p = 2$			$p = 5$			$p = 2$			$p = 5$		
	Vertex	Genest	Schmid	Vertex	Genest	Schmid	Vertex	Genest	Schmid	Vertex	Genest	Schmid
$\tau = 0.05$												
$\nu = 1$	36.98	9.56	10.01	94.92	97.86	15.13	88.07	12.96	12.36	99.96	99.91	21.56
$\nu = 2$	11.84	7.30	8.27	70.59	79.91	13.88	39.69	8.92	10.10	96.17	92.45	19.49
$\nu = 3$	6.29	6.77	7.40	48.75	59.43	12.72	20.41	7.79	9.17	84.48	73.96	18.31
$\nu = 4$	4.25	6.82	7.14	34.71	44.84	12.06	12.86	7.09	8.51	69.55	56.56	18.56
$\nu = 5$	3.58	6.35	6.77	25.84	35.96	11.14	9.91	6.74	8.28	58.46	45.31	17.14
$\tau = 0.1$												
$\nu = 1$	38.68	11.37	11.77	94.88	98.09	24.48	87.93	16.11	15.09	99.98	99.96	36.81
$\nu = 2$	12.47	9.08	9.93	71.34	80.36	23.92	41.26	11.44	13.13	96.05	92.74	38.96
$\nu = 3$	6.92	8.38	9.16	49.50	60.96	23.32	22.83	10.80	12.90	85.21	75.56	39.97
$\nu = 4$	5.04	7.73	8.73	35.22	46.53	23.46	15.52	10.14	12.05	71.33	58.31	40.24
$\nu = 5$	4.24	8.00	8.70	26.76	36.80	22.82	12.44	10.62	12.86	59.87	46.45	41.10
$\tau = 0.25$												
$\nu = 1$	46.02	22.42	22.62	95.20	98.27	60.97	93.12	40.00	38.00	99.96	99.93	86.34
$\nu = 2$	20.71	20.89	22.57	74.24	83.93	65.03	59.35	37.21	39.91	97.29	94.67	90.94
$\nu = 3$	13.47	19.99	22.24	55.92	66.42	68.87	42.82	36.69	41.11	88.97	81.41	93.35
$\nu = 4$	11.28	20.26	22.27	43.53	54.85	69.04	34.44	36.52	41.31	78.74	67.79	93.66
$\nu = 5$	9.83	20.73	22.43	34.74	45.97	70.93	30.39	36.19	41.74	69.26	57.60	94.23

The power is compared in a Monte Carlo simulation with $N = 10{,}000$ iterations for dimensions $p \in \{2, 5\}$ at a significance level of 5% and various sample sizes $n \in \{25, 50\}$. Since the vertex test has the focus on deviations from independence in the tails of the a distribution, the generated data is from a multivariate Student t distribution with low degrees of freedom, $\nu = 1, ..., 5$. To determine the critical value a Monte-Carlo simulation with $N = 1{,}000$ iterations has been carried out under the null hypothesis. This is the default in the package `copula`. The parameter of dependence is set such that it results in $\tau = 0.05, 0.1, 0.25$. Table 2 gives an overview on the simulated values.

To give a résumé, the vertex test performs best in situations with small τ, fairly large sample size $n = 50$ and a low degree of freedom. This is expected, since a certain sample size is required for observing values in the tails where the vertex test is sensitive. The Schmid test has the highest power if the degrees of freedom and the dependency τ are relatively high. It seems that the test statistic $\rho_{1,p}$ tends to be a sound choice for strong linear deviation from independence towards a normal distribution. For smaller sample sizes from a high dimensional sample, the test of Genest and Rémillard (2004) achieves the highest power.

Interestingly, there is a difference in the tendencies of the behavior of power: With increasing degrees of freedom, the power of the vertex and the Genest test diminishes where the opposite is true for the Schmid test. In a situation where one needs to decide which test should be used, an analysis of the existence of the first moments seems to be advisable.

There is strong evidence that all of the three tests are consistent, since the power is approximately 100% in all cases for very large sample sizes (not listed in table). This finding emboldens the usage of the vertex test based on the polynomial copula as a test for deviations from independence with a focus on the high dimensional tail regions.

Throughout the section, the nuisance parameter α has been set to 1. However, different values of α lead to different simulated power values. The next section provides a possible selection procedure for α using a semiparametric estimation in the bivariate case.

3.3. Setting the nuisance parameter α

This section presents an adaptive selection procedure for α that can be used to estimate α_{auto} from data prior to the actual test for independence. Amblard and Girard (2005) introduced a semiparametric method of estimating the generating function ϕ of bivariate copulas with one dependence parameter from data, if the copula function can be written as

$$C_{\theta,\phi}(u_1, u_2) = uv + \theta\phi(u_1)\phi(u_2). \tag{14}$$

The function ϕ must satisfy the assumptions $\phi(0) = \phi(1) = 0$ and $|\phi(x) - \phi(y)| \leq |x - y|$ for all $x, y \in [0, 1]^2$.

Example 6. *The* 4*-reduced form of the bivariate polynomial copula can be written in the sense of equation* (14) *with* $\phi_\alpha(x) = x^{\alpha+1}(1 - x)$*.* ϕ_α *is a proper generating function since* $\phi_\alpha(0) = \phi_\alpha(1) = 0$ *and* $|\phi_\alpha(x) - \phi_\alpha(y)| \leq |x - y|$ *for all* $x, y \in [0, 1]^2$*. The latter is true, since* ϕ'_α *has an upper bound of* $L = 1$ *and is therefore Lipschitz continuous.*

The 1-reduced form can also be expressed by a generating function as in example 6. However, this is not true for the remaining two reduced forms. Thus, using the method of Amblard and Girard (2005) for calibrating α can only involve the 1- and the 4-reduced form of the polynomial copula.

Once the semiparametric estimate of $\hat{\phi}$ is obtained from a bivariate sample, the best parameter α_{auto} is the value of α that minimizes the functional quadratic distance between $\hat{\phi}$ and ϕ_α. Table 3 provides simulated power values from an identical simulation setting as in section 3.2 for fix values $\alpha = 1, 2, 3$ and α_{auto}, as well as several sample sizes n.

For small sample sizes, a small value of α leads to the highest power. This is expected, since the tail regions are sparse for $n = 25$ and therefore the sensitive regions should not be focused too much on the tails alone. With augmenting sample size however, the best results are obtained with a higher value of α.

Table 3: Power of the vertex test of independence if the sample is from a bivariate Student t distribution with sample size n, degrees of freedom ν and Kendall's τ. The power (in percent) is based on $10,000$ repetitions at a significance level of 5%. The tuning parameter α is set to the fix values 1, 2 and 3 in contrast to an adaptive tuning procedure (auto).

	$n = 25$				$n = 50$				$n = 100$				$n = 200$			
	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$	auto	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$	auto	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$	auto	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$	auto
$\tau = 0.05$																
$\nu = 1$	37.80	36.54	29.41	22.68	87.12	88.31	90.12	87.72	99.74	99.79	99.87	99.82	100.00	100.00	100.00	99.82
$\nu = 2$	12.11	10.92	8.76	5.70	40.11	43.07	47.27	40.24	78.00	80.11	84.41	80.73	98.42	98.51	99.22	80.73
$\nu = 3$	6.72	6.11	4.28	3.07	21.34	24.12	26.28	20.88	46.88	50.03	55.98	49.94	81.18	82.40	87.06	49.94
$\nu = 4$	4.62	3.68	2.44	1.86	14.25	15.35	16.66	13.09	31.57	34.56	38.57	31.98	59.90	62.17	67.45	31.98
$\nu = 5$	3.65	2.69	1.67	1.42	11.10	11.89	12.46	9.17	22.76	24.70	27.45	23.32	44.31	46.89	51.17	23.32
$\tau = 0.1$																
$\nu = 1$	40.88	38.30	30.65	22.87	89.18	90.20	91.46	88.15	99.81	99.84	99.93	99.87	100.00	100.00	100.00	99.87
$\nu = 2$	14.31	12.55	9.00	6.32	46.51	48.41	50.25	43.17	84.10	84.76	87.30	85.90	99.08	98.96	99.38	85.90
$\nu = 3$	8.26	7.15	4.96	2.55	26.88	28.43	29.86	24.11	58.54	60.14	62.58	60.32	90.30	89.82	91.24	60.32
$\nu = 4$	6.48	4.96	3.35	1.91	19.64	20.25	20.51	15.47	44.49	45.49	45.86	44.02	77.11	76.81	76.55	44.02
$\nu = 5$	4.91	3.83	2.66	1.63	15.67	15.46	15.61	12.03	34.44	36.00	35.24	35.10	66.28	66.52	64.14	35.10
$\tau = 0.25$																
$\nu = 1$	59.17	52.73	41.54	28.79	96.75	95.81	95.45	93.80	100.00	100.00	99.99	100.00	100.00	100.00	100.00	100.00
$\nu = 2$	33.49	27.38	17.71	11.82	78.13	74.63	69.50	64.67	98.70	97.76	96.79	98.45	100.00	99.98	99.96	98.45
$\nu = 3$	25.78	19.13	12.45	7.43	67.31	61.79	52.97	45.97	95.49	93.72	88.79	95.07	99.97	99.86	99.52	95.07
$\nu = 4$	21.82	15.69	9.15	5.50	61.54	55.11	44.61	36.68	93.43	90.41	81.18	91.99	99.92	99.66	98.37	91.99
$\nu = 5$	19.88	13.61	7.63	4.31	56.76	49.63	37.45	31.32	91.85	88.02	75.60	89.03	99.83	99.55	97.04	89.03

The results for applying the semiparametric routine of estimating α_{auto} from the data prior to the actual test are promising: using α_{auto} instead of a fixed α results in simulated power values that are similar to the best power obtained with fixed α if the sample size is large. The obtained power is often not the overall highest, but there are only few fixed α with a slightly higher performance together with some α with considerably lower power. However, small samples sizes $n < 100$ are insufficient for the semiparametric method of Amblard and Girard (2005), who use a minimum of $n = 100$ in their paper. The bad quality of the estimate $\hat{\phi}$ illuminates why the fitted α_{auto} results in a lower power than for any of the fixed α. Using α_{auto} is therefore recommended in situations where the vertex test has not been applied yet, so no further information on how to choose α properly is available, and if the sample size is sufficiently large.

A generalization of this semiparametric routine to higher dimensions is possible, but even more cumbersome. For the p-dimensional polynomial copula however, yet the 1- and the 2^p-reduced form have a representation with a generating function ϕ – the remaining 2^{p-1} reduced forms cannot be considered. Thus, the natural choice of $\alpha = 1$ is recommended for higher dimensions in remembrance of Occam's razor.

3.4. Calculation time

This section discusses the required calculation time with respect to dimension p and sample size n. To avoid redundancy, the vertex test is exclusively discussed for $\alpha = 1$ and only compared to the Genest test which has similar calculation requirements as the Schmid test.

Since the variances of the distributions of the test statistics of the Genest and the Schmid test contain a Brownian bridge, their quantiles cannot be calculated directly – they have to be determined via a preceded simulation for every constellation of dimension p and sample size n. Both, quality of the derived critical values and the required calculation time are augmenting with increasing amount of iterations (the default setting in the package `copula` is 1,000 Monte Carlo iterations). In addition to that, it is obvious that the required calculation time increases for larger sample sizes and higher dimensions.

Although only the asymptotic distribution of the vertex test statistic T_α is known, it is provided in a closed form expression. For small sample sizes the error of approximation of the critical value can be large. However, with increasing sample size this error diminishes as table 2 suggests. Since the dimension p only affects the degrees of freedom needed to determine the critical value of an univariate distribution, the dimension has virtually no influence on the required calculation time of the critical value compared to the other nonprametric tests. Naturally, p influences the time that is needed to make a test decision because the vector of LMPRT statistics of length 2^p needs to be calculated from data. Similarly to any of the three proposed tests, this calculation can get very cumbersome especially in high dimensions despite the formula of the LMPRT statistics is relatively simple.

Table 4: Calculation time in seconds (except as noted otherwise) that is needed to make a test decision. Genest test: Time to simulate a critical value. Vertex test: Time to generate the calculation routine.

	Genest test					**Vertex test**
	$n = 25$	$n = 50$	$n = 100$	$n = 200$	$n = 500$	$n \leq 1000$
p						
2	0.06	0.07	0.15	0.58	3.80	0.00
3	0.04	0.11	0.33	1.29	17.74	0.00
4	0.09	0.20	0.67	2.78	48.48	0.00
5	0.17	0.45	1.62	6.53	132.56	0.00
6	0.33	1.00	3.82	14.83	321.15	0.04
7	0.70	2.23	8.91	34.31	767.48	0.14
...	...	...	...	...	...	...
13	15[1]	$2m$[1]	$20m$[1]	$1h$[1]	$5d$[1]	$23m$

[1]Extrapolated

Table 4 gives an overview on the calculation time that is required for the entire testing procedure, including the simulation of the critical values and the calculation of the test statistic. Obviously, the time that is needed to simulate a critical value is increasing exponentially for both, the Genest and the vertex test as the sample sizes or the dimension rises. However, the degree of exponentiality differs massively. Especially samples of large size illuminate that the time needed to simulate the critical values for the Genest test plays an important role for actual applications. Similarly, the calculation time augments for the vertex test with higher dimension p. Hereby the bottleneck is rather the calculation of the 2^p dimensional vector of test statistics and the required covariance matrix for scaling rather than the determination of the critical value. The actual calculation time is barely influenced by the sample size n – quite the contrary, the vertex test rather profits of increasing sample sizes since the error of approximation is shrinking as pointed out by section 3.2. The time provided in table 4 is the run-time needed in `R` in order to generate the routine that can be used to calculate the vector of test statistics $\boldsymbol{t}_1$ from data.

Thus, the vertex test can be a sound alternative in situations where the time it lasts to perform a test of independence is relevant. This finding is especially important if the sample size is varying, which would require several cumbersome simulations of the critical values of the other nonparametric tests, or if the sample size is very large.

4. Summary and Outlook

In this article, a nonparametric multivariate rank-based test for independence has been developed. The test consists of two generalizations: First, it is based on an extension of bivariate locally most powerful rank tests for copulas with one dependence parameter to p dimensions. Second, a generalization of the famous FGM-copula with more than one dependence parameter and one nuisance parameter is introduced, the polynomial copula.

The multivariate test of independence for one parameter, sequentially applied to a copula family with several dependence parameters, can be used as a joint test for independence provided that there is a unique parameter constellation that results in the independence copula. The squared scaled form of the vector of test statistics leads to a simple χ^2-type test of independence. Its key feature is an asymptotic distribution that has a closed form expression.

Deriving this test for the polynomial copula leads to an expression that can be interpreted as sum of squared scaled deviations from the density function of the independence copula evaluated in regions close to the vertices of the unit cube – the vertex test. Thus, if one is interested in independent behavior of marginal distributions with focus on multivariate tails, this vertex test can be applied to samples from any distribution.

In a Monte Carlo simulation study, this vertex test based on the polynomial copula has been compared to other nonparametric rank tests. For samples from a Student t distribution with non-existing lower moments, three tests have been compared by evaluating the power with a focus on dimensions $p > 2$. The influence of the choice of the nuisance parameter on the power has been discussed and an adaptive procedure for determining this parameter from data has been introduced. However, a broader simulation study considering further alternative hypotheses could illuminate the differences of the examined tests and the choice of the nuisance parameter in a more detailed manner.

One key advantage of the vertex test is the good performance for large sample sizes n, since the critical values used for test decisions are from an asymptotic distribution and are therefore not affected by sample size. Other state-of-the-art rank-based tests for independence need a cumbersome simulation of the critical values preceding the test. This time-consuming procedure can be omitted using the vertex test which makes it attractive in areas where the computation time is crucial. The loss in power which arises by the use of an asymptotic instead of a finite sample distribution was shown to be minor in the simulation.

Decoding the violation of independence into deviations from independence in every single vertex (e.g. the more frequent joint occurrences of high ranks as in the case of independence) gives insight into the nature of dependence. This finding could be used e.g. in the area of financial market data – if one wants to build a portfolio of p assets from a pool of N assets, one could identify the one out of $\binom{N}{p}$ combinations that provides favorable joint movement of the returns. In this way, one could search for a combination of assets, whose returns move in the most diametrical of parallel manner. The former is of interest for diversification, the latter for an investment strategy and has already found application in Stübinger et al. (2016). The focus on the high dimensional tail regions, or vertices of the unit cube, enables a new concept of symmetry for copulas that can easily be generalized to higher dimensions, see Mangold (2017) for further details.

Bibliography

Amblard, C. and Girard, S. (2005). Estimation procedures for a semiparametric family of bivariate copulas. *Journal of Computational and Graphical Statistics*, 14(2):363–377.

Behnen, K. (1971). Asymptotic optimality and ARE of certain rank-order tests under contiguity. *The Annals of Mathematical Statistics*, 42(1):325–329.

Czado, C. (2010). Pair-copula constructions of multivariate copulas. In Jaworski, P., Durante, F., Härdle, W. K., and Rychlik, T., editors, *Copula Theory and its Applications: Proceedings of the Workshop held in Warsaw, 25-26 September 2009*, pages 93–109. Springer, Berlin/Heidelberg, Germany.

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

Demarta, S. and McNeil, A. J. (2005). The t copula and related copulas. *International Statistical Review / Revue Internationale de Statistique*, 73(1):111–129.

Garralda-Guillén, A. I. (1998). Dependencia y tests de rangos para leyes bidimensionales. *Doctoral thesis, Universidad de Granada, Spain.*

Genest, C. and Rémillard, B. (2004). Test of independence and randomness based on the empirical copula process. *Test*, 13(2):335–369.

Genest, C. and Verret, F. (2005). Locally most powerful rank tests of independence for copula models. *Journal of Nonparametric Statistics*, 17(5):521–539.

Hájek, J., Šidák, Z., and Sen, P. K. (1999). *Theory of Rank Tests*. Probability and mathematical statistics. Academic Press, San Diego, 2nd edition.

Hampel, F. R. (1986). *Robust Statistics: The Approach based on Influence Functions*. Wiley series in probability and mathematical statistics. John Wiley & Sons, New York, USA.

Hofert, M., Kojadinovic, I., Mächler, M., and Yan, J. (2015). *copula: Multivariate Dependence with Copulas*. R package version 0.999-14.

Hofert, M. and Mächler, M. (2011). Nested archimedean copulas meet R: The nacopula package. *Journal of Statistical Software*, 39(9):1–20.

Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*, volume 73 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, USA.

Kimeldorf, G. and Sampson, A. (1975). Uniform representations of bivariate distributions. *Communications in Statistics – Theory and Methods*, 4(7):617–627.

Kojadinovic, I. and Yan, J. (2010). Modeling multivariate distributions with continuous margins using the copula R package. *Journal of Statistical Software*, 34(9):1–20.

Kotz, S. and Johnson, N. L. (1977). Propriétés de dépendance des distributions itérées, généralisées à deux variables Farlie-Gumbel-Morgenstern. *Comptes Rendus de l'Académie des Sciences. Série I. Mathématique*, 285:277–280.

Lin, G. D. (1987). Relationships between two extensions of Farlie-Gumbel-Morgenstern distribution. *Annals of the Institute of Statistical Mathematics*, 39(1):129–140.

Mangold, B. (2017). New concepts of symmetry for copulas. FAU Discussion Papers in Economics 6, Friedrich-Alexander University Erlangen-Nürnberg, Germany.

Nelsen, R. B. (2006). *An Introduction to Copulas.* Springer series in statistics. Springer, New York, USA, 2nd edition.

Nelsen, R. B., Quesada-Molina, J. J., and Rodriguez-Lallena, J. A. (1997). Bivariate copulas with cubic sections. *Journal of Nonparametric Statistics*, 7(3):205–220.

R Core Team (2016). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna.

Sarmanov, I. O. (1974). New forms of correlation relationships between positive quantities applied in hydrology. *Mathematical Models in Hydrology*, 100:104–109.

Schmid, F. and Schmidt, R. (2007). Multivariate extensions of Spearman's rho and related statistics. *Statistics & Probability Letters*, 77(4):407–416.

Sen, P. K. and Puri, M. L. (1967). On the theory of rank order tests for location in the multivariate one sample problem. *The Annals of Mathematical Statistics*, 38(4):1216–1228.

Shirahata, S. (1974). Locally most powerful rank tests for independence. *Bulletin of Mathematical Statistics*, 16(1):11–21.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231.

Stübinger, J., Mangold, B., and Krauß, C. (2016). Statistical arbitrage with vine copulas. FAU Discussion Papers in Economics 11, Friedrich-Alexander University Erlangen-Nürnberg, Germany.

Yan, J. (2007). Enjoy the joy of copulas: With a package copula. *Journal of Statistical Software*, 21(4):1–21.

Appendix

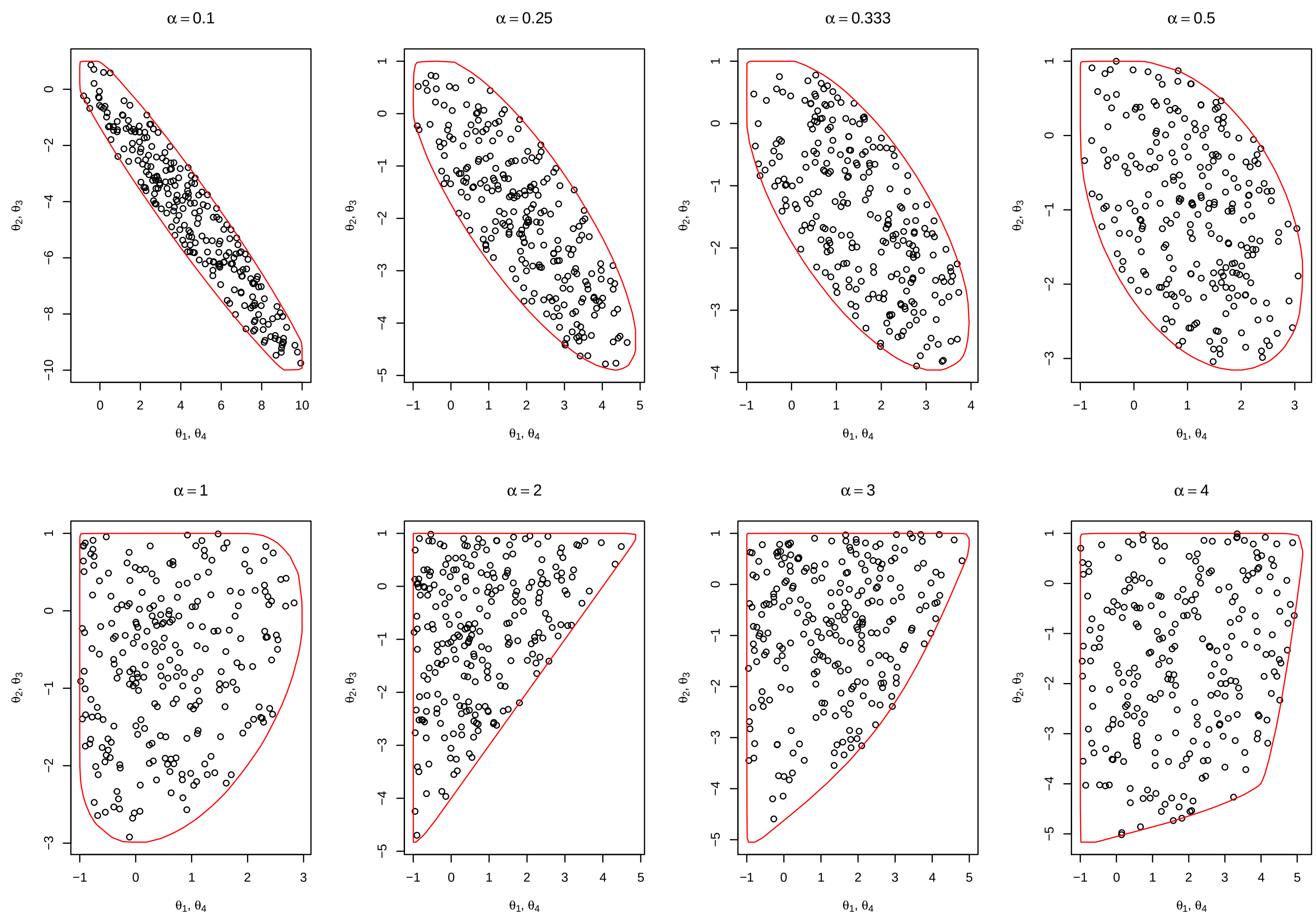

Figure 6: 250 simulated $\boldsymbol{\theta}$ for several values of α, for which the bivariate polynomial copula is admissible. Convex hull approximates $\mathcal{S}^{\alpha}$ and is based on 10,000 simulated values.

Table 5: Test statistics $\boldsymbol{T}_\alpha$ for $p = 2$ and several values of α.

$$\boldsymbol{T}_{0.5} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} \sqrt{\left(1-\frac{R_{1i}}{n+1}\right)\left(1-\frac{R_{2i}}{n+1}\right)} \left(5\frac{R_{1i}}{n+1}-2\right)\left(5\frac{R_{2i}}{n+1}-2\right) \\ \frac{1}{n}\sum_{i=1}^{n} \left(5\frac{R_{2i}}{n+1}-3\right)\sqrt{\frac{R_{2i}}{n+1}\left(1-\frac{R_{1i}}{n+1}\right)}\left(5\frac{R_{1i}}{n+1}-2\right) \\ \frac{1}{n}\sum_{i=1}^{n} \left(5\frac{R_{1i}}{n+1}-3\right)\sqrt{\left(1-\frac{R_{2i}}{n+1}\right)\frac{R_{1i}}{n+1}}\left(5\frac{R_{2i}}{n+1}-2\right) \\ \frac{1}{n}\sum_{i=1}^{n} \left(5\frac{R_{2i}}{n+1}-3\right)\left(5\frac{R_{1i}}{n+1}-3\right)\sqrt{\frac{R_{1i}}{n+1}\frac{R_{2i}}{n+1}} \end{pmatrix} \quad \boldsymbol{T}_1 = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} \left(\frac{R_{2i}}{n+1}-1\right)\left(3\frac{R_{1i}}{n+1}-1\right)\left(\frac{R_{1i}}{n+1}-1\right)\left(3\frac{R_{2i}}{n+1}-1\right) \\ \frac{1}{n}\sum_{i=1}^{n} -\left(3\frac{R_{2i}}{n+1}-2\right)\left(3\frac{R_{1i}}{n+1}-1\right)\left(\frac{R_{1i}}{n+1}-1\right)\frac{R_{2i}}{n+1} \\ \frac{1}{n}\sum_{i=1}^{n} -\left(\frac{R_{2i}}{n+1}-1\right)\left(3\frac{R_{2i}}{n+1}-1\right)\frac{R_{1i}}{n+1}\left(3\frac{R_{1i}}{n+1}-2\right) \\ \frac{1}{n}\sum_{i=1}^{n} \left(3\frac{R_{2i}}{n+1}-2\right)\frac{R_{1i}}{n+1}\left(3\frac{R_{1i}}{n+1}-2\right)\frac{R_{2i}}{n+1} \end{pmatrix}$$

$$\boldsymbol{T}_2 = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} \left(4\frac{R_{2i}}{n+1}-1\right)\left(1-\frac{R_{2i}}{n+1}\right)^2\left(4\frac{R_{1i}}{n+1}-1\right)\left(1-\frac{R_{1i}}{n+1}\right)^2 \\ \frac{1}{n}\sum_{i=1}^{n} \left(\frac{R_{2i}}{n+1}\right)^2\left(4\frac{R_{2i}}{n+1}-3\right)\left(4\frac{R_{1i}}{n+1}-1\right)\left(1-\frac{R_{1i}}{n+1}\right)^2 \\ \frac{1}{n}\sum_{i=1}^{n} \left(4\frac{R_{2i}}{n+1}-1\right)\left(1-\frac{R_{2i}}{n+1}\right)^2\left(\frac{R_{1i}}{n+1}\right)^2\left(4\frac{R_{1i}}{n+1}-3\right) \\ \frac{1}{n}\sum_{i=1}^{n} \left(\frac{R_{2i}}{n+1}\right)^2\left(4\frac{R_{2i}}{n+1}-3\right)\left(\frac{R_{1i}}{n+1}\right)^2\left(4\frac{R_{1i}}{n+1}-3\right) \end{pmatrix} \quad \boldsymbol{T}_3 = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} \left(5\frac{R_{2i}}{n+1}-1\right)\left(1-\frac{R_{2i}}{n+1}\right)^3\left(1-\frac{R_{1i}}{n+1}\right)^3\left(5\frac{R_{1i}}{n+1}-1\right) \\ \frac{1}{n}\sum_{i=1}^{n} \left(\frac{R_{2i}}{n+1}\right)^3\left(1-\frac{R_{1i}}{n+1}\right)^3\left(5\frac{R_{2i}}{n+1}-4\right)\left(5\frac{R_{1i}}{n+1}-1\right) \\ \frac{1}{n}\sum_{i=1}^{n} \left(5\frac{R_{2i}}{n+1}-1\right)\left(5\frac{R_{1i}}{n+1}-4\right)\left(1-\frac{R_{2i}}{n+1}\right)^3\left(\frac{R_{1i}}{n+1}\right)^3 \\ \frac{1}{n}\sum_{i=1}^{n} \left(\frac{R_{2i}}{n+1}\right)^3\left(5\frac{R_{1i}}{n+1}-4\right)\left(\frac{R_{1i}}{n+1}\right)^3\left(5\frac{R_{2i}}{n+1}-4\right) \end{pmatrix}$$

$$\boldsymbol{T}_4 = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} \left(1-\frac{R_{1i}}{n+1}\right)^4\left(6\frac{R_{2i}}{n+1}-1\right)\left(6\frac{R_{1i}}{n+1}-1\right)\left(1-\frac{R_{2i}}{n+1}\right)^4 \\ \frac{1}{n}\sum_{i=1}^{n} \left(1-\frac{R_{1i}}{n+1}\right)^4\left(6\frac{R_{2i}}{n+1}-5\right)\frac{R_{2i}}{n+1}^4\left(6\frac{R_{1i}}{n+1}-1\right) \\ \frac{1}{n}\sum_{i=1}^{n} \left(\frac{R_{1i}}{n+1}\right)^4\left(6\frac{R_{1i}}{n+1}-5\right)\left(6\frac{R_{2i}}{n+1}-1\right)\left(1-\frac{R_{2i}}{n+1}\right)^4 \\ \frac{1}{n}\sum_{i=1}^{n} \left(6\frac{R_{2i}}{n+1}-5\right)\left(\frac{R_{1i}}{n+1}\right)^4\left(6\frac{R_{1i}}{n+1}-5\right)\left(\frac{R_{2i}}{n+1}\right)^4 \end{pmatrix}$$

Table 6: Covariance matrices $\Sigma(\dot{c}_{\mathbf{0}})$ of $\boldsymbol{T}_\alpha$ for $p = 2$ for several values of α.

$$\alpha = 0.5$$

$$\begin{pmatrix} \frac{9}{256} & \frac{63\pi}{8192} & \frac{63\pi}{8192} & \frac{441\pi^2}{26214} \\ \frac{63\pi}{8192} & \frac{9}{256} & \frac{441\pi^2}{26214} & \frac{63\pi}{8192} \\ \frac{63\pi}{8192} & \frac{441\pi^2}{26214} & \frac{9}{256} & \frac{63\pi}{8192} \\ \frac{441\pi^2}{26214} & \frac{63\pi}{8192} & \frac{63\pi}{8192} & \frac{9}{256} \end{pmatrix}$$

$$\alpha = 1$$

$$\begin{pmatrix} \frac{4}{225} & \frac{1}{225} & \frac{1}{225} & \frac{1}{900} \\ \frac{1}{225} & \frac{4}{225} & \frac{1}{900} & \frac{1}{225} \\ \frac{1}{225} & \frac{1}{900} & \frac{4}{225} & \frac{1}{225} \\ \frac{1}{900} & \frac{1}{225} & \frac{1}{225} & \frac{4}{225} \end{pmatrix}$$

$$\alpha = 2$$

$$\begin{pmatrix} \frac{9}{1225} & \frac{-3}{2450} & \frac{-3}{2450} & \frac{1}{4900} \\ \frac{-3}{2450} & \frac{9}{1225} & \frac{1}{4900} & \frac{-3}{2450} \\ \frac{-3}{2450} & \frac{1}{4900} & \frac{9}{1225} & \frac{-3}{2450} \\ \frac{1}{4900} & \frac{-3}{2450} & \frac{-3}{2450} & \frac{9}{1225} \end{pmatrix}$$

$$\alpha = 3$$

$$\begin{pmatrix} \frac{16}{3969} & \frac{-2}{2835} & \frac{-2}{2835} & \frac{1}{8100} \\ \frac{-2}{2835} & \frac{16}{3969} & \frac{1}{8100} & \frac{-2}{2835} \\ \frac{-2}{2835} & \frac{1}{8100} & \frac{16}{3969} & \frac{-2}{2835} \\ \frac{1}{8100} & \frac{-2}{2835} & \frac{-2}{2835} & \frac{16}{3969} \end{pmatrix}$$

$$\alpha = 4$$

$$\begin{pmatrix} \frac{25}{9801} & \frac{-5}{19602} & \frac{-5}{19602} & \frac{1}{39204} \\ \frac{-5}{19602} & \frac{25}{9801} & \frac{1}{39204} & \frac{-5}{19602} \\ \frac{-5}{19602} & \frac{1}{39204} & \frac{25}{9801} & \frac{-5}{19602} \\ \frac{1}{39204} & \frac{-5}{19602} & \frac{-5}{19602} & \frac{25}{9801} \end{pmatrix}$$

Example 7 (Vertex test for a trivariate sample). *For $p = 3$ and $\alpha = 1$ we have under the null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 = \mathbf{0}$ that that the vector of LMPRT statistics is*

$$\boldsymbol{T}_1 = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n}\left(\frac{R_{1i}}{n+1}-1\right)\left(\frac{R_{3i}}{n+1}-1\right)\left(3\frac{R_{3i}}{n+1}-1\right)\left(\frac{R_{2i}}{n+1}-1\right)\left(3\frac{R_{2i}}{n+1}-1\right)\left(3\frac{R_{1i}}{n+1}-1\right) \\ \frac{1}{n}\sum_{i=1}^{n}-\left(\frac{R_{1i}}{n+1}-1\right)\left(\frac{R_{2i}}{n+1}-1\right)\left(3\frac{R_{2i}}{n+1}-1\right)\frac{R_{3i}}{n+1}\left(3\frac{R_{1i}}{n+1}-1\right)\left(3\frac{R_{3i}}{n+1}-2\right) \\ \frac{1}{n}\sum_{i=1}^{n}-\left(\frac{R_{1i}}{n+1}-1\right)\left(\frac{R_{3i}}{n+1}-1\right)\left(3\frac{R_{3i}}{n+1}-1\right)\left(3\frac{R_{2i}}{n+1}-2\right)\left(3\frac{R_{1i}}{n+1}-1\right)\frac{R_{2i}}{n+1} \\ \frac{1}{n}\sum_{i=1}^{n}\left(\frac{R_{1i}}{n+1}-1\right)\left(3\frac{R_{2i}}{n+1}-2\right)\frac{R_{3i}}{n+1}\left(3\frac{R_{1i}}{n+1}-1\right)\left(3\frac{R_{3i}}{n+1}-2\right)\frac{R_{2i}}{n+1} \\ \frac{1}{n}\sum_{i=1}^{n}-\left(\frac{R_{3i}}{n+1}-1\right)\left(3\frac{R_{3i}}{n+1}-1\right)\left(\frac{R_{2i}}{n+1}-1\right)\left(3\frac{R_{2i}}{n+1}-1\right)\left(3\frac{R_{1i}}{n+1}-2\right)\frac{R_{1i}}{n+1} \\ \frac{1}{n}\sum_{i=1}^{n}\left(\frac{R_{2i}}{n+1}-1\right)\left(3\frac{R_{2i}}{n+1}-1\right)\frac{R_{3i}}{n+1}\left(3\frac{R_{1i}}{n+1}-2\right)\frac{R_{1i}}{n+1}\left(3\frac{R_{3i}}{n+1}-2\right) \\ \frac{1}{n}\sum_{i=1}^{n}\left(\frac{R_{3i}}{n+1}-1\right)\left(3\frac{R_{3i}}{n+1}-1\right)\left(3\frac{R_{2i}}{n+1}-2\right)\left(3\frac{R_{1i}}{n+1}-2\right)\frac{R_{1i}}{n+1}\frac{R_{2i}}{n+1} \\ \frac{1}{n}\sum_{i=1}^{n}-\left(3\frac{R_{2i}}{n+1}-2\right)\frac{R_{3i}}{n+1}\left(3\frac{R_{1i}}{n+1}-2\right)\frac{R_{1i}}{n+1}\left(3\frac{R_{3i}}{n+1}-2\right)\frac{R_{2i}}{n+1} \end{pmatrix}.$$

Further, $\sqrt{n}\boldsymbol{T}_1$ is normal distributed with expectation $\boldsymbol{\mu}' = (0,0,0,0,0,0,0,0)'$ and

$$\Sigma(\dot{c}_{\mathbf{0}})_{i=1\ldots8,j=1\ldots8} = \begin{pmatrix} 512 & -128 & -128 & 32 & -128 & 32 & 32 & -8 \\ -128 & 512 & 32 & -128 & 32 & -128 & -8 & 32 \\ -128 & 32 & 512 & -128 & 32 & -8 & -128 & 32 \\ 32 & -128 & -128 & 512 & -8 & 32 & 32 & -128 \\ -128 & 32 & 32 & -8 & 512 & -128 & -128 & 32 \\ 32 & -128 & -8 & 32 & -128 & 512 & 32 & -128 \\ 32 & -8 & -128 & 32 & -128 & 32 & 512 & -128 \\ -8 & 32 & 32 & -128 & 32 & -128 & -128 & 512 \end{pmatrix}^{-1}.$$

Therefore, we have

$$T_1 = n\boldsymbol{T}_1{}'\Sigma(\dot{c}_{\mathbf{0}})^{-1}\boldsymbol{T}_1 \overset{a}{\sim} \chi^2(8)$$

and the asymptotic p-value is obtained using $1 - F_{\chi^2(8)}(t_1)$.