Multivariate Nonparametric Tests of Independence

Sara Taskinen [*], Hannu Oja[+] and Ronald H. Randles[#]

[*]*Dept. of Mathematics and Statistics, University of Jyväskylä*
[+]*Tampere School of Public Health, University of Tampere*
[#]*Dept. of Statistics, University of Florida*

April 13, 2005

Abstract

New test statistics are proposed for testing whether two random vectors are independent. Gieser and Randles as well as Taskinen, Kankainen and Oja introduced and discussed multivariate extensions of the quadrant test by Blomqvist. This paper serves as a sequel to this work and presents new multivariate extensions of Kendall's tau and Spearman's rho statistics. Two different approaches are discussed: First, interdirection proportions are used to estimate the cosines of angles between centered observation vectors and between differences of observation vectors. Second, covariances between affine equivariant multivariate signs and ranks are used. The test statistics arising from these two approaches appear to be asymptotically equivalent if each vector is elliptically symmetric. The spatial sign versions are easy to compute for data in common dimensions. They provide practical, robust alternatives to normal theory methods. Asymptotic theory is developed to approximate the finite-sample null distributions as well as to calculate limiting Pitman efficiencies. Small sample null permutation distributions are also described. A simple simulation study is used to compare the proposed tests to the classical Wilks' test. Finally, the theory is illustrated by an example.

KEY WORDS: Affine invariance, Kendall's tau, Pitman efficiency, Quadrant test, Robustness, Spearman's rho

1 INTRODUCTION

In many problem settings multiple measurements are made on each experimental unit, resulting in high dimension multivariate data. It is often of interest to explore potential relationships among subsets of these measurements. For example, some measurements may represent attributes of psychological characteristics while others represent attributes of physical characteristics. It may be of interest to determine whether there is a relationship between the psychological and the physical characteristics. This requires a test of independence between pairs of vectors, where the vectors potentially have different measurement scales and dimensions.

Let $\boldsymbol{x}_i^T = (\boldsymbol{x}_i^{(1)^T}, \boldsymbol{x}_i^{(2)^T})$ for $i = 1, \ldots, n$ denote a random sample of vector pairs, where $\boldsymbol{x}_i^{(1)^T}$ is a continuous vector of dimension p and $\boldsymbol{x}_i^{(2)^T}$ is a continuous vector of dimension q. We seek to test:

$$H_0: \quad \boldsymbol{x}_i^{(1)} \text{ and } \boldsymbol{x}_i^{(2)} \text{ are independent} \quad \text{vs} \quad H_a: \quad \text{they are dependent.}$$

The classical parametric test due to Wilks (1935) is based on

$$W = \frac{|A|}{|A_{11}||A_{22}|},$$

where $A = \sum_{i=1}^n (\boldsymbol{x}_i - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_i)^T$ is partitioned into $A_{st} = \sum_{i=1}^n (\boldsymbol{x}_i^{(s)} - \bar{\boldsymbol{x}}_i^{(s)})(\boldsymbol{x}_i^{(t)} - \bar{\boldsymbol{x}}_i^{(t)})^T$ for $s, t = 1, 2$. Under the assumption of multivariate normality, Wilks' test is optimal, that is, the most efficient test. Under H_0 with finite fourth moments, $-n \log W \to_d \chi^2_{pq}$.

A nonparametric analogue to Wilks' test was given by Puri and Sen (1971). They developed a class of tests based on componentwise ranking which use a test statistic of the form

$$S^J = \frac{|T|}{|T_{11}||T_{22}|}.$$

Here the elements of $(p + q) \times (p + q)$ matrix T are

$$T_{st} = \frac{1}{n} \sum_{i=1}^n J_s\left(\frac{C_{si}}{n+1}\right) J_t\left(\frac{C_{ti}}{n+1}\right),$$

where C_{si} denotes the rank of the sth component of $\boldsymbol{x}_i$ among the sth components of all n vectors, J_s and J_t are arbitrary (standardized) score functions and T is partitioned in the same manner as in the Wilks' test. Under H_0, $-n \log S^J \to_d \chi^2_{pq}$.

Muirhead (1982) examined the effect of the group of transformations $\{\boldsymbol{x} \to A\boldsymbol{x} + \boldsymbol{b}\}$ on this problem. Here $\boldsymbol{b}$ is a $p + q$ vector and

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$$

is an arbitrary nonsingular matrix with $p \times p$ matrix A_1 and $q \times q$ matrix A_2. The Wilks' test is invariant under this group of transformations. Thus its performance does not depend on the variance-covariance structure of either $\boldsymbol{x}_i^{(1)}$ or $\boldsymbol{x}_i^{(2)}$. The test given by Puri and Sen is not invariant under this group of transformations. Gieser and Randles (1997) proposed a nonparametric test based on interdirection counts that generalized the univariate ($p = q = 1$) quadrant test of Blomqvist (1950) and which is invariant under this transformation group. Taskinen, Kankainen and Oja (2003a) proposed a more practical invariant extension of the quadrant test based on spatial signs. It is easy to compute for data in dimensions (say, less than 15) commonly encountered in practice. These two invariant quadrant test extensions share certain asymptotic properties, for example, their asymptotic efficiencies are the same.

This paper develops tests which generalize the popular univariate tests due to Kendall (1938) and Spearman (1904) to any dimensions (arbitrary p and q). They have advantages over the quadrant test extensions in that they do not require centering (subtracting a location estimator) and they generally have better power properties than the quadrant test extensions. Moreover, the spatial sign versions are easy to compute for data in common dimensions, thus providing intuitive, practical, robust alternatives to multivariate normal theory methods.

The univariate tests based on Kendall's tau, Spearman's rho and Blomqvist's quadrant statistic are described in Section 2. Their multivariate analogues based on interdirections are also sketched. In Section 3, the more practical versions of these tests based on spatial signs and ranks are described. Section 4 shows their large sample properties under the null hypothesis as well as under a sequence of alternatives under mild assumptions; e.g. even the assumption of the existence of first moments can be avoided. Asymptotic efficiencies are given and simulations are used to compare the finite sample powers of these tests in section 5. In Section 6, the theory is illustrated by an example and the paper is concluded with some comments in Section 7.

2 ANALOGUES TO KENDALL'S TAU AND SPEARMAN'S RHO

Consider the problem of measuring dependence within the components of bivariate vectors. Let $(x_i^{(1)}, x_i^{(2)})^T$, $i = 1, \ldots, n$, be a random sample from a bivariate, continuous population. Using the univariate sign funtion $S(x) = \text{sign}(x) = -1, 0, 1$ as $x <, =, > 0$, the **sign** of the univariate $x_i^{(1)}$ is $S_i^{(1)} = S(x_i^{(1)})$ and the sign of the difference $x_i^{(1)} - x_j^{(1)}$ is $S_{ij}^{(1)} = S(x_i^{(1)} - x_j^{(1)})$. The **centered rank** of $x_i^{(1)}$ is described as $R_i^{(1)} = \text{ave}_j S(x_i^{(1)} - x_j^{(1)})$, and the univariate median $\widehat{\mu}^{(1)}$ of the $x_i^{(1)}$'s satisfies $\text{ave}_j S(\widehat{\mu}^{(1)} - x_j^{(1)}) = 0$. The **centered sign** of $x_i^{(1)}$ is then $\widehat{S}_i^{(1)} = S(x_i^{(1)} - \widehat{\mu}^{(1)})$. Using $x_1^{(2)}, \ldots, x_n^{(2)}$ to define $S_i^{(2)}$, $S_{ij}^{(2)}$, $R_i^{(2)}$, $\widehat{\mu}^{(2)}$ and $\widehat{S}_i^{(2)}$ analogously, the popular nonparametric measures of dependence are now conveniently expressed. They are the Blomqvist's quadrant statistic

$$Q = \text{ave}_i \{\widehat{S}_i^{(1)} \widehat{S}_i^{(2)}\},$$

Kendall's tau

$$\tau = \text{ave}_{i,j} \{S_{ij}^{(1)} S_{ij}^{(2)}\}$$

and Spearman's rho

$$\rho = \text{ave}_i \{R_i^{(1)} R_i^{(2)}\} = \text{ave}_{i,j,k} \{S_{ij}^{(1)} S_{ik}^{(2)}\}.$$

The test statistics are thus covariances (or correlations) between centered signs, signs of the pairwise differences, and centered ranks, respectively.

When correlating univariate pairs, the most interpretable feature is the magnitude of the correlation, that is, the square of the correlation coefficient. Multivariate measures of correlation provide analogues to

$$Q^2 = \text{ave}\{(\widehat{S}_i^{(1)} \widehat{S}_{i'}^{(1)})(\widehat{S}_i^{(2)} \widehat{S}_{i'}^{(2)})\},$$

$$\tau^2 = \text{ave}\{(S_{ij}^{(1)} S_{i'j'}^{(1)})(S_{ij}^{(2)} S_{i'j'}^{(2)})\}$$

and

$$\rho^2 = \text{ave}\{(S_{ij}^{(1)} S_{i'j'}^{(1)})(S_{ik}^{(2)} S_{i'k'}^{(2)})\},$$

where the averages are computed over all possible indices. We can now interpret $\widehat{S}_i^{(1)} \widehat{S}_{i'}^{(1)}$ and $S_{ij}^{(1)} S_{i'j'}^{(1)}$, for example, as the cosines of the angles between $x_i^{(1)} - \widehat{\mu}^{(1)}$ and $x_{i'}^{(1)} - \widehat{\mu}^{(1)}$ and between $x_i^{(1)} - x_j^{(1)}$ and $x_{i'}^{(1)} - x_{j'}^{(1)}$, respectively. In the univariate cases, the cosine values are $+1$ or -1.

In (1989) Randles used so called interdirection proportions to estimate the cosines of angular distance between two vectors relative to the positions of the other vectors. In the current context, for example, the cosine of the angle

between $\boldsymbol{x}_i^{(1)} - \boldsymbol{x}_j^{(1)}$ and $\boldsymbol{x}_{i'}^{(1)} - \boldsymbol{x}_{j'}^{(1)}$ could be measured with $\cos(\pi p^{(1)}(i,j;i',j'))$ where $p^{(1)}(i,j;i',j')$ represents the fraction of hyperplanes formed by differences $\boldsymbol{x}_{i*}^{(1)} - \boldsymbol{x}_{j*}^{(1)}$ and $\boldsymbol{0}$ such that $\boldsymbol{x}_i^{(1)} - \boldsymbol{x}_j^{(1)}$ and $\boldsymbol{x}_{i'}^{(1)} - \boldsymbol{x}_{j'}^{(1)}$ are on opposite sides of the hyperplane. (Here $i*$ and $j*$ would not use any of $\{i,j,i'$ or $j'\}$). Defining similar proportions among $\boldsymbol{x}^{(2)}$ vector differences, yields

$$\tau_1^2 = \mathrm{ave}\{\cos(\pi p^{(1)}(i,j;i',j'))\cos(\pi p^{(2)}(i,j;i',j'))\}$$

and

$$\rho_1^2 = \mathrm{ave}\{\cos(\pi p^{(1)}(i,j;i',j'))\cos(\pi p^{(2)}(i,k;i',k'))\}$$

as natural multivariate analogues to Kendall's τ^2 and Spearman's ρ^2, respectively. Also here the averages are computed over all possible indices. Kendall's tau and Spearman's rho analogues provide natural counterparts to the quadrant test analogue

$$Q_1^2 = \mathrm{ave}\{\cos(\pi\widehat{p}^{(1)}(i;i'))\cos(\pi\widehat{p}^{(2)}(i;i'))\},$$

which was studied by Gieser and Randles (1997), where, for example, $\widehat{p}^{(1)}(i;i')$ is the fraction of hyperplanes formed by $p-1$ vectors $\boldsymbol{x}_{i*}^{(1)} - \widehat{\boldsymbol{\mu}}^{(1)}$ and $\boldsymbol{0}$ such that $\boldsymbol{x}_i^{(1)} - \widehat{\boldsymbol{\mu}}^{(1)}$ and $\boldsymbol{x}_{i'}^{(1)} - \widehat{\boldsymbol{\mu}}^{(1)}$ are on opposite sides of the hyperplane, where $i*$ is not equal to either i or i'. Here the observations are centered on some affine equivariant location estimator $\widehat{\boldsymbol{\mu}}^{(1)}$ based on $\boldsymbol{x}_1^{(1)}, \ldots, \boldsymbol{x}_n^{(1)}$, for example, the Oja (1983) median or the transformation retransformation spatial median (Hettmansperger and Randles, 2002). The Kendall and Spearman analogues do not require centering on a location estimator.

While Q_1^2, τ_1^2 and ρ_1^2 are meaningful extensions of the quadrant test statistic, Kendall's tau and Spearman's rho, respectively, they are difficult to compute if either p or q exceed 2. More practical extensions are described in the next section.

3 TESTS BASED ON SPATIAL SIGNS AND RANKS

Intuitive, computationally convenient extensions of Kendall's and Spearman's statistics can be created using notions of spatial signs and spatial centered ranks as developed in Möttönen and Oja (1995). To make the resulting correlation measures affine invariant, the data points are transformed before spatial signs and ranks are formed. Consider a sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ where each vector is of dimension k. The data points are transformed into $\boldsymbol{z}_i = \widehat{V}^{-1/2}(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})$ where $\widehat{\boldsymbol{\mu}}$ is an affine equivariant location estimator and $\widehat{V}$ is a shape matrix estimate.

Definition 1. $\widehat{V}$ *is called a* **shape matrix estimate**, *if it is a positive definite* $k \times k$ *matrix with* $Tr(\widehat{V}) = k$ *and affine equivariant in the sense that if* $\widehat{V}^*$ *is calculated from* $A\boldsymbol{x}_i + \boldsymbol{b}$, $i = 1, \ldots, n$, *then*

$$\widehat{V}^* = \frac{k}{Tr(A\widehat{V}A^T)} A\widehat{V}A^T,$$

for every $k \times k$ *nonsingular matrix* A *and* k *vector* $\boldsymbol{b}$. *Here* $Tr(\cdot)$ *denotes the trace.*

The spatial sign function is defined as

$$\boldsymbol{S}(\boldsymbol{x}) = \left\{ \begin{array}{ll} \|\boldsymbol{x}\|^{-1}\boldsymbol{x}, & \boldsymbol{x} \neq \boldsymbol{0} \\ \boldsymbol{0}, & \boldsymbol{x} = \boldsymbol{0}, \end{array} \right.$$

where $\|\boldsymbol{x}\| = (\boldsymbol{x}^T\boldsymbol{x})^{1/2}$ is the (Euclidean) length of the vector $\boldsymbol{x}$. Applying the spatial sign function to the transformed data points produces **standardized sign vectors** $\widehat{\boldsymbol{S}}_i = \boldsymbol{S}(\boldsymbol{z}_i)$ and $\widehat{\boldsymbol{S}}_{ij} = \boldsymbol{S}(\boldsymbol{z}_i - \boldsymbol{z}_j)$. The **standardized rank vector** is then $\widehat{\boldsymbol{R}}_i = \text{ave}_j \widehat{\boldsymbol{S}}_{ij}$. Because the $\widehat{\boldsymbol{S}}_{ij}$'s and $\widehat{\boldsymbol{R}}_i$'s are based on differences, they do not depend on $\widehat{\boldsymbol{\mu}}$ at all, unless it plays a role in the computation of $\widehat{V}$. The following is proved in the Appendix A.

Lemma 1. *The standardized sign vectors* $\widehat{\boldsymbol{S}}_i$ *and* $\widehat{\boldsymbol{S}}_{ij}$ *and standardized rank vector* $\widehat{\boldsymbol{R}}_i$ *are affine equivariant (and location invariant) in the sense that if* $\widehat{\boldsymbol{S}}_i^*$, $\widehat{\boldsymbol{S}}_{ij}^*$ *and* $\widehat{\boldsymbol{R}}_i^*$ *are calculated from* $A\boldsymbol{x}_i + \boldsymbol{b}$, $i = 1, \ldots, n$, *then* $\widehat{\boldsymbol{S}}_i^* = P\widehat{\boldsymbol{S}}_i$ *and* $\widehat{\boldsymbol{S}}_{ij}^* = P\widehat{\boldsymbol{S}}_{ij}$ *and* $\widehat{\boldsymbol{R}}_i^* = P\widehat{\boldsymbol{R}}_i$, *where* $P = (A\widehat{V}A^T)^{-1/2}A\widehat{V}^{1/2}$ *is orthogonal.*

The multivariate extension of Kendall's tau is created by forming sign vectors $\widehat{\boldsymbol{S}}_{ij}^{(1)}$ based on the first components $\boldsymbol{x}_1^{(1)}, \ldots, \boldsymbol{x}_n^{(1)}$, transformed by a shape matrix estimate $\widehat{V}^{(1)}$ chosen so that

$$p \, \text{ave}_{i,j} \{\widehat{\boldsymbol{S}}_{ij}^{(1)} \widehat{\boldsymbol{S}}_{ij}^{(1)T}\} = I_p,$$

where I_p is the $p \times p$ identity matrix. This is the transformation studied by Tyler (1987) only computed on differences $\boldsymbol{x}_i^{(1)} - \boldsymbol{x}_j^{(1)}$. See also Dümbgen (1998). With this choice, the $\widehat{\boldsymbol{S}}_{ij}^{(1)}$'s do not depend on $\widehat{\boldsymbol{\mu}}^{(1)}$ at all. Similarly, q-dimensional sign vectors $\widehat{\boldsymbol{S}}_{ij}^{(2)}$ are formed based on $\boldsymbol{x}_1^{(2)}, \ldots, \boldsymbol{x}_n^{(2)}$ transformed by a $\widehat{V}^{(2)}$ chosen to satisfy

$$q \, \text{ave}_{i,j} \{\widehat{\boldsymbol{S}}_{ij}^{(2)} \widehat{\boldsymbol{S}}_{ij}^{(2)T}\} = I_q.$$

7

The multivariate Kendall's tau squared is then

$$\tau_2^2 = ||\text{ave}_{i,j}\{\widehat{\boldsymbol{S}}_{ij}^{(1)}\widehat{\boldsymbol{S}}_{ij}^{(2)^T}\}||^2,$$

where $|| \cdot ||^2 = Tr(\cdot^T \cdot)$ is the squared matrix norm. τ_2^2 is not the same as τ_1^2, but like τ_1^2 it can be described as a correlation of cosines. Unlike τ_1^2, the statistic τ_2^2 is easily computed for data in common dimensions.

The multivariate extension of Spearman's rho uses $\widehat{\boldsymbol{R}}_i^{(1)}$ based on the first components $\boldsymbol{x}_1^{(1)}, \ldots, \boldsymbol{x}_n^{(1)}$ transformed by a shape matrix $\widehat{V}^{(1)}$ chosen so that

$$p\,\text{ave}_i\{\widehat{\boldsymbol{R}}_i^{(1)}\widehat{\boldsymbol{R}}_i^{(1)^T}\} = \text{ave}_i\{\widehat{\boldsymbol{R}}_i^{(1)^T}\widehat{\boldsymbol{R}}_i^{(1)}\}I_p.$$

See Visuri, Oja and Koivunen (2000) for the development of these rank covariance matrices. With this choice, the rank vectors $\widehat{\boldsymbol{R}}_i^{(1)}$ do not depend on $\widehat{\boldsymbol{\mu}}^{(1)}$ in any way.

With analogous descriptions of the q-dimensional rank vectors $\widehat{\boldsymbol{R}}_i^{(2)}$, the multivariate Spearman's rho squared is then

$$\rho_2^2 = ||\text{ave}_i\{\widehat{\boldsymbol{R}}_i^{(1)}\widehat{\boldsymbol{R}}_i^{(2)^T}\}||^2.$$

It is not the same as ρ_1^2, but it has a similar interpretation as a correlation of cosines. The statistic ρ_2^2 is easy to compute for data in common dimensions. This property makes it much more practical than ρ_1^2.

The quadrant statistic of Taskinen et al. (2003a) uses p-dimensional standardized sign vectors $\widehat{\boldsymbol{S}}_i^{(1)}$ based on $\boldsymbol{x}_1^{(1)}, \ldots, \boldsymbol{x}_n^{(1)}$ with $\widehat{\boldsymbol{\mu}}^{(1)}$ and $\widehat{V}^{(1)}$ chosen to satisfy

$$\text{ave}_i\{\widehat{\boldsymbol{S}}_i^{(1)}\} = \boldsymbol{0} \quad \text{and} \quad p\,\text{ave}_i\{\widehat{\boldsymbol{S}}_i^{(1)}\widehat{\boldsymbol{S}}_i^{(1)^T}\} = I_p.$$

The estimates $\widehat{\boldsymbol{\mu}}^{(1)}$ and $\widehat{V}^{(1)}$ are the transformation retransformation spatial median and Tyler's shape estimate as proposed in Hettmansperger and Randles (2002). With analogous descriptions of the q-dimensional standardized signs $\widehat{\boldsymbol{S}}_i^{(2)}$, the Taskinen et al. test is based on the statistic

$$Q_2^2 = ||\text{ave}_i\{\widehat{\boldsymbol{S}}_i^{(1)}\widehat{\boldsymbol{S}}_i^{(2)^T}\}||^2.$$

Again, Q_2^2 is much easier to compute than Q_1^2 and is therefore more practical.

4   LIMITING DISTRIBUTIONS

To establish a large sample distribution for the statistics described in Section 3, we first examine their behaviour under the null hypothesis condition

that $\boldsymbol{x}^{(1)}$ is independent from $\boldsymbol{x}^{(2)}$. Assume now that $\boldsymbol{x}^{(1)}$ has a continuous elliptically symmetric distribution with density function of the form

$$f(\boldsymbol{x}^{(1)}) = |\Sigma_1|^{-1/2} \exp\left\{-\Psi_1\left(||\Sigma_1^{-1/2}(\boldsymbol{x}^{(1)} - \boldsymbol{\mu}_1)||\right)\right\},$$

where $\boldsymbol{\mu}_1$ denotes the location of the distribution and Σ_1 defines the shape of its contours. Similarly assume that $\boldsymbol{x}^{(2)}$ has an elliptically symmetric distribution with density

$$g(\boldsymbol{x}^{(2)}) = |\Sigma_2|^{-1/2} \exp\left\{-\Psi_2\left(||\Sigma_2^{-1/2}(\boldsymbol{x}^{(2)} - \boldsymbol{\mu}_2)||\right)\right\}.$$

In the following we also need to assume that the location and shape estimates used in the standardization are $\sqrt{n}$-consistent. Note that, if for example the standardization for τ_2^2 and ρ_2^2 is based on Dümbgen's (1998) shape estimate, no assumptions on the existence of first moments of f and g are needed.

Under above assumptions, and under contiguous alternatives, the statistics τ_1^2 and τ_2^2, ρ_1^2 and ρ_2^2 and Q_1^2 and Q_2^2 are asymptotically equivalent within each pair, respectively:

Theorem 1. *Under the null hypothesis with elliptically symmetric marginal distributions, $n(\tau_1^2 - \tau_2^2) \xrightarrow{P} 0$, $n(\rho_1^2 - \rho_2^2) \xrightarrow{P} 0$ and $n(Q_1^2 - Q_2^2) \xrightarrow{P} 0$.*

Also small sample simulations showed only minor differences in the performance of the statistics within each pair. Thus, in what follows, we examine exclusively the more practical versions τ_2^2, ρ_2^2 and Q_2^2.

Theorem 2. *Under the null hypothesis assumptions described above, the limiting distributions of*

$$\frac{npq}{4c_F^2 c_G^2}\, \tau_2^2, \quad \frac{npq}{c_F^2 c_G^2}\, \rho_2^2 \quad and \quad npq\, Q_2^2$$

are chi-squared distributions with pq degrees of freedom. The constants c_F^2 and c_G^2 depend on the marginal distributions F and G, respectively.

The constants c_F^2 and c_G^2 depend on the underlying distributions Ψ_1 and Ψ_2, that is, the distributions of the respective standardized component radius. When $\Sigma_1 = I$, the centered rank function is $\boldsymbol{R}_F(\boldsymbol{x}) = E_F[\boldsymbol{S}(\boldsymbol{x} - \boldsymbol{x}_i^{(1)})]$. Letting $\boldsymbol{x} = r\boldsymbol{u}$, where $r = ||\boldsymbol{x}||$ and $\boldsymbol{u} = \boldsymbol{S}(\boldsymbol{x})$, $\boldsymbol{R}_F(r\boldsymbol{u}) = q_F(r)\boldsymbol{u}$, for some bounded increasing function $q_F(r)$. The constant $c_F^2 = E_F[q_F^2(r)]$ and c_G^2 is analogously defined. See Appendix B for expressions of c_F^2 at selected distributions. To use τ_2^2 and ρ_2^2 as test statistics for detecting correlation, obviously the constants c_F^2 and c_G^2 must be consistently estimated since they

depend on the respective marginal distributions. Fortunately, this is not difficult since convergent estimates

$$\widehat{c}_F^2 = \text{ave}_i\{\widehat{\boldsymbol{R}}_i^{(1)^T}\widehat{\boldsymbol{R}}_i^{(1)}\} \quad \text{and} \quad \widehat{c}_G^2 = \text{ave}_i\{\widehat{\boldsymbol{R}}_i^{(2)^T}\widehat{\boldsymbol{R}}_i^{(2)}\}$$

are readily available.

To consider and compare the efficiencies of different test statistics, we next derive the limiting distributions of τ_2^2, ρ_2^2 and Q_2^2 under certain contiguous alternative sequences. We use similar sequences of alternatives as Gieser and Randles (1997) and Taskinen et al. (2003a): Let $\boldsymbol{x}_i^{(1)}$ and $\boldsymbol{x}_i^{(2)}$ be independent with spherical marginal densities $\exp\{-\Psi_1(||\boldsymbol{x}^{(1)}||)\}$ and $\exp\{-\Psi_2(||\boldsymbol{x}^{(2)}||)\}$ and write, for some choices of M_1 and M_2,

$$\begin{pmatrix} \boldsymbol{y}_i^{(1)} \\ \boldsymbol{y}_i^{(2)} \end{pmatrix} = \begin{pmatrix} (1-\Delta)I_p & \Delta M_1 \\ \Delta M_2 & (1-\Delta)I_q \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_i^{(1)} \\ \boldsymbol{x}_i^{(2)} \end{pmatrix}, \tag{1}$$

with $\Delta = \delta/\sqrt{n}$. Let f_Δ be the density of $(\boldsymbol{y}_i^{(1)^T}, \boldsymbol{y}_i^{(2)^T})^T$. The optimal likelihood ratio test statistic for testing H_0 against H_Δ is then

$$L = \sum_{i=1}^n \{\log f_\Delta(\boldsymbol{y}_i^{(1)}, \boldsymbol{y}_i^{(2)}) - \log f_0(\boldsymbol{y}_i^{(1)}, \boldsymbol{y}_i^{(2)})\}.$$

We need the general assumption that, under the null hypothesis,

$$L = \sqrt{n}\delta K - \frac{1}{2}\delta^2\sigma^2 + o_P(1) \tag{2}$$

where $K = \text{ave}_i\{k_i\} = \text{ave}_i\{k(\boldsymbol{y}_i^{(1)}, \boldsymbol{y}_i^{(2)})\}$ with $E(k_i) = 0$ and $Var(k_i) = \sigma^2$ is bounded. This should be checked separately in each case. See the Appendix A for the formula for k_i. Under this assumption, the sequence of alternatives is then contiguous to the null hypothesis (LeCam's first lemma).

If now τ_2^{2*}, ρ_2^{2*} and Q_2^{2*} represent τ_2^2 and ρ_2^2 and Q_2^2 calculated from these transformed observations, then we get the following result.

Theorem 3. *For $max(p,q) > 1$, The limiting distributions of $[npq/4c_F^2 c_G^2]\,\tau_2^{2*}$ and $[npq/c_F^2 c_G^2]\,\rho_2^{2*}$ are noncentral chi-squared distributions with pq degrees of freedom and noncentrality parameter*

$$\frac{\delta^2}{4\,pq\,c_F^2 c_G^2}\,||d_1 M_1 + d_2 M_2^T||^2,$$

where

$$d_1 = (p-1)E(||\boldsymbol{x}_i^{(2)} - \boldsymbol{x}_j^{(2)}||)E(||\boldsymbol{x}_i^{(1)} - \boldsymbol{x}_j^{(1)}||^{-1})$$

and

$$d_2 = (q-1)E(||\boldsymbol{x}_i^{(1)} - \boldsymbol{x}_j^{(1)}||)E(||\boldsymbol{x}_i^{(2)} - \boldsymbol{x}_j^{(2)}||^{-1}).$$

The limiting distribution of $npq\,Q_2^{2*}$ is a noncentral chi-squared distribution with pq degrees of freedom and noncentrality parameter

$$\frac{\delta^2}{pq}||c_1 M_1 + c_2 M_2^T||^2,$$

where

$$c_1 = (p-1)E(||\boldsymbol{x}_i^{(2)}||)E(||\boldsymbol{x}_i^{(1)}||^{-1})$$

and

$$c_2 = (q-1)E(||\boldsymbol{x}_i^{(1)}||)E(||\boldsymbol{x}_i^{(2)}||^{-1}).$$

For the case $p = q = 1$, see Taskinen et al. (2003b). The use of these noncentrality parameters in asymptotic efficiencies requires some clarifying comments. We are considering efficiencies for the dependent alternatives in the 'directions' determined by fixed M_1 and M_2. The alternative distributions are not elliptical (except in the normal case) but the sequence is contiguous to an elliptical null distribution. Next note that the efficiencies are of the same type $||h_1 M_1 + h_2 M_2^T||^2$ where h_1 and h_2 depend on the marginal spherical distributions and the test used. If the marginal distributions are the same (which implies $h_1 = h_2$), then the efficiencies do not depend on M_1 and M_2 at all. We say that the test is "asymptotically unbiased" if the noncentrality parameter is positive. Our tests are clearly not asymptotically unbiased for choices M_1 and M_2 such that $h_1 M_1 = -h_2 M_2^T$. The only exception is again the case where both marginal distributions are multivariate normal; in that case $M_1 = -M_2^T$ implies uncorrelateness as well as independence. Finally note that, if the marginal distributions are elliptical, then due to affine invariance, the efficiencies are given by $||h_1 \Sigma_1^{-1/2} M_1 \Sigma_2^{1/2} + h_2 \Sigma_1^{1/2} M_2^T \Sigma_2^{-1/2}||^2$. In the next section, our efficiency comparisons use $M_1 = M_2^T$.

As test statistics for testing the null hypothesis of independence, none of the τ_2^2, ρ_2^2 or Q_2^2 have an unconditional distribution-free property for small n. However, their permutation null distribution and hence permutation p-values are easy to generate. The permutation principle dictates that all possible $n!$ pairings of the given set of n first component $\boldsymbol{x}^{(1)}$ vectors with the set of n second component $\boldsymbol{x}^{(2)}$ vectors are equally likely. Since the computation of marginal standardized signs and ranks do not depend on the pairing, the null permutation distributions can be generated by finding

$$\tau_2^2(\boldsymbol{\alpha}) = ||\text{ave}_{i,j}\{\widehat{\boldsymbol{S}}_{ij}^{(1)} \widehat{\boldsymbol{S}}_{\alpha_i \alpha_j}^{(2)T}\}||^2,$$

$$\rho_2^2(\boldsymbol{\alpha}) = ||\text{ave}_i\{\widehat{\boldsymbol{R}}_i^{(1)} \widehat{\boldsymbol{R}}_{\alpha_i}^{(2)T}\}||^2$$

11

and

$$Q_2^2(\boldsymbol{\alpha}) = ||\text{ave}_i\{\widehat{\boldsymbol{S}}_i^{(1)}\widehat{\boldsymbol{S}}_{\alpha_i}^{(2)^T}\}||^2,$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T$ is drawn at random from the $n!$ permutations of the integers $(1, \ldots, n)$.

5 LIMITING AND FINITE-SAMPLE EFFICIENCIES

We next compare τ_2^2 and ρ_2^2 to Wilks' test W through limiting efficiencies and simple simulation studies. For comparisons between quadrant statistics and W, see Gieser and Randles (1997) and Taskinen et al. (2003a). As the limiting distributions are of the same type, χ_{pq}^2, the efficiency comparisons may be based on noncentrality parameters only. The efficiency comparisons are now made in the multivariate normal distribution, t distribution and contaminated normal distribution cases. For simplicity, we assume that $M_1 = M_2^T$.

Since W has a limiting noncentral chi-squared distribution with pq degrees of freedom and noncentrality parameter $\delta^2||M_1 + M_2^T||^2$, the asymptotic efficiencies are simply given by

$$ARE(\tau_2^2, W) = ARE(\rho_2^2, W) = \frac{(d_1 + d_2)^2}{16\,pq\,c_F^2 c_G^2},$$

with d_1 and d_2 given in Theorem 3. For a k-dimensional multivariate normal distribution with cdf Φ,

$$E_\Phi(||\boldsymbol{x}_i - \boldsymbol{x}_j||) = \frac{2\,\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \quad \text{and} \quad E_\Phi(||\boldsymbol{x}_i - \boldsymbol{x}_j||^{-1}) = \frac{\Gamma\left(\frac{k-1}{2}\right)}{2\,\Gamma\left(\frac{k}{2}\right)}.$$

And for a k-dimensional standardized t-distribution with ν degrees of freedom

$$E_{t_\nu}(||\boldsymbol{x}_i - \boldsymbol{x}_j||) = \frac{\sqrt{\nu-2}\,\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{2\nu-1}{2}\right)\Gamma^2\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma^2\left(\frac{\nu}{2}\right)\Gamma(\nu-1)}$$

and

$$E_{t_\nu}(||\boldsymbol{x}_i - \boldsymbol{x}_j||^{-1}) = \frac{\Gamma\left(\frac{k-1}{2}\right)\Gamma\left(\frac{2\nu+1}{2}\right)\Gamma^2\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu-2}\,\Gamma\left(\frac{k}{2}\right)\Gamma^2\left(\frac{\nu}{2}\right)\Gamma(\nu+1)}.$$

The resulting limiting efficiencies for selected degrees of freedom and dimensions are listed in Table 1. Note that since limiting multinormality of the regular covariance matrix holds if the fourth moments of the underlying

distribution are finite, W has a limiting distribution only when $\nu > 4$. When the underlying distribution is multivariate normal ($\nu = \infty$), the Wilks' test is the best, but τ_2^2 and ρ_2^2 are very competitive with it. As the underlying population becomes heavy-tailed (ν gets smaller), then τ_2^2 and ρ_2^2 are better than Wilks' test.

Table 1: $ARE(\tau_2^2, W)$ and $ARE(\rho_2^2, W)$ at different p- and q-variate t distributions for selected $\nu = \nu_1 = \nu_2$.

				p		
	q	2	3	5	8	10
	2	1.12	1.13	1.14	1.15	1.16
	3		1.14	1.16	1.17	1.17
$\nu = 5$	5			1.17	1.18	1.19
	8				1.19	1.20
	10					1.20
	2	1.00	1.01	1.02	1.03	1.03
	3		1.02	1.03	1.04	1.04
$\nu = 10$	5			1.04	1.05	1.05
	8				1.06	1.06
	10					1.07
	2	0.93	0.94	0.95	0.95	0.96
	3		0.95	0.96	0.96	0.96
$\nu = \infty$	5			0.96	0.97	0.97
	8				0.97	0.98
	10					0.98

Consider next the contaminated normal distribution with cdf $\Phi_{\epsilon,c}(\boldsymbol{x}) = (1 - \epsilon)\, \Phi(\boldsymbol{x}) + \epsilon\, \Phi(\boldsymbol{x}/c)$, where $c > 0$ and Φ is the cdf of $N_k(\mathbf{0}, I_k)$. Now

$$E_{\Phi_{\epsilon,c}}(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|) = \left[(1-\epsilon)^2 + \epsilon(1-\epsilon)\sqrt{2(1+c^2)} + c\epsilon^2\right] \frac{2\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}$$

and

$$E_{\Phi_{\epsilon,c}}(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^{-1}) = \left[(1-\epsilon)^2 + 2\epsilon(1-\epsilon)\sqrt{\frac{2}{1+c^2}} + \frac{\epsilon^2}{c}\right] \frac{\Gamma\left(\frac{k-1}{2}\right)}{2\Gamma\left(\frac{k}{2}\right)}$$

13

and limiting efficiencies for $\epsilon = 0.1$ and for selected values of c are listed in Table 2. Here the superior performance of the nonparametric tests is obvious.

Table 2: $ARE(\tau_2^2, W)$ and $ARE(\rho_2^2, W)$ at different p- and q-variate contaminated normal distributions for $\epsilon = 0.1$ and for selected values of c.

				p		
	q	2	3	5	8	10
	2	1.17	1.19	1.20	1.21	1.22
	3		1.20	1.22	1.23	1.23
$c = 3$	5			1.23	1.24	1.25
	8				1.25	1.26
	10					1.26
	2	1.92	1.95	1.98	2.00	2.01
	3		1.98	2.01	2.04	2.04
$c = 6$	5			2.05	2.07	2.08
	8				2.09	2.10
	10					2.11

A simulation study was used to compare the finite sample powers of τ_2^2, ρ_2^2 and W. For comparisons between quadrant statistics, see Taskinen et al. (2003a). Independent $\boldsymbol{x}^{(1)}$- and $\boldsymbol{x}^{(2)}$-samples of sizes $n = 30$ and 60 were generated from a multivariate standard normal distribution, from a standardized t distribution with $\nu_1 = \nu_2 = 5$ and from a contaminated normal distribution with $\epsilon = 0.1$ and $c = 6$. The transformation in (1) with $M_1 = M_2^T = I$ was used for chosen values of δ to introduce dependence into the model. The tests were applied and the process was replicated 1500 times. The critical values used in test constructions were based on the chi-square approximations to the null distributions as given in Theorem 1 (with estimated $\widehat{c}_F^2$ and $\widehat{c}_G^2$).

The empirical powers for $p = q = 3$ are given in Figure 1 and for $p = q = 5$ in Figure 2. Consider first the simulation results in the case $p = q = 3$. In the multivariate normal case W is slightly better than the other tests. In the considered t distribution case tests are almost equally powerful and in the contaminated normal case τ_2^2 and ρ_2^2 perform much better than W. As $n = 30$, ρ_2^2 seems to be slightly more powerful than τ_2^2, but as n increases, no significant differences can be seen between tests. As $p = q = 5$ and $n = 30$, ρ_2^2 is clearly more powerful than τ_2^2, but again for large n, the tests are almost

14

equally powerful. The size of ρ_2^2 is very close to the designated size 0.05 in all cases. For small sample sizes, the size of τ_2^2 is often slightly below 0.05 and for heavy-tailed distributions W varies widely above 0.05.

As interdirection based τ_1^2 and ρ_1^2 are computationally very intensive, we compared the empirical powers of τ_1^2 and ρ_1^2 to those of τ_2^2 and ρ_2^2 for $n = 15$ and $p = q = 2$ only. As seen in Figure 3, both Spearman's rho statistics are clearly better than the two Kendall's tau statistics, as was expected. However, no significant differences can be seen between τ_1^2 and τ_2^2 and between ρ_1^2 and ρ_2^2, respectively.

6   A REAL DATA EXAMPLE

McNaughton and Davies (1987) studied the effect of an aerobic conditioning program on cholesterol levels. Several physical and blood-related variables of 12 subjects were measured before and after an aerobic conditioning program. Table 3 lists changes in respiratory variables: vital capacity (VC), forced expiratory volume (FEV) and maximum oxygen uptake (VO_2) and changes in cholesterol variables: total cholesterol (TC), triglycerides (TG) and HDL cholesterol (HDL). We wish to test whether the respiratory changes are independent of cholesterol changes.

Table 3: Respiratory and cholesterol changes of 12 subjects.

Subject	Respiratory changes			Cholesterol changes		
	VC	FEV	VO_2	TC	TG	HDL
1	0.30	0.45	24.00	-0.50	-0.20	0.00
2	0.40	0.50	10.00	0.60	0.10	0.20
3	0.20	0.45	4.00	0.40	-0.10	0.00
4	0.20	0.30	2.00	0.10	0.10	-0.10
5	0.60	0.50	13.00	-0.50	-0.20	0.00
6	0.30	0.10	10.00	0.10	0.60	0.00
7	0.70	1.00	28.00	-0.90	-1.40	0.40
8	0.10	0.10	8.00	-0.50	0.10	-0.30
9	0.20	0.30	3.00	0.40	0.00	-0.10
10	0.20	0.35	0.00	0.40	0.10	0.00
11	0.30	0.59	-5.00	0.30	0.00	0.10
12	0.00	20.00	0.00	0.50	0.10	0.00

The p-values given by different tests were calculated using the chi-square approximations to the null distributions. For the original data set, the p-values for W, τ_2^2, ρ_2^2 and Q_2^2 are 0.022, 0.028, 0.018 and 0.025. To illustrate the robustness of τ_2^2, ρ_2^2 and Q_2^2, the first VO_2 measurement was moved from -25 to 50 and the p-values were recalculated. The results are shown in Figure 4. When the first VO_2 measurement decreases, the p-value given by W increases substantially, while only small changes can be seen in the p-values of other tests.

Finally, the permutation p-values were obtained for the original data set using 1500 permutation samples. The p-values for W, τ_2^2, ρ_2^2 and Q_2^2 are 0.069, 0.011, 0.015 and 0.013. Note that even for this very small sample size ($n = 12$), the permutation p-values of τ_2^2, ρ_2^2 and Q_2^2 are quite similar to those given by the chi-square approximation, and the approximation for ρ_2^2 is extremely good. More variation is seen in the p-value of W.

7   FINAL REMARKS

The paper develops new nonparametric intuitively appealing tests for testing whether two random vectors are independent. The tests generalize the popular univariate tests due to Kendall (1938) and Spearman (1904). In the first approach, Randles' (1989) interdirections were used in the test construction. It is remarkable that, in this approach, no explicit location vector and scatter matrix estimates are needed. In the second approach, the extensions based on the spatial signs and ranks of the standardized observations are introduced. Natural shape matrix estimates for the standardization are proposed. The test statistics arising from these two approaches appear to be asymptotically equivalent if each vector is elliptically symmetric.

The new tests have an appealing invariance property and a convenient limiting chi-square null distribution. The tests have excellent asymptotic and finite-sample efficiencies. The spatial sign and rank versions are easy to compute for data in any dimension. As demonstrated in the example, the tests are resistant to outliers, and the p-values for the conditionally distribution-free permutation test are also easily generated. To conclude, the spatial sign and rank versions provide intuitively appealing, practical, robust alternatives to multivariate normal theory methods.

APPENDIX A: PROOFS OF THE RESULTS

Proof of Lemma 1   Using $\widehat{V}^* = k\,[Tr(A\widehat{V}A^T)]^{-1}A\widehat{V}A^T$ and $\widehat{\boldsymbol{\mu}}^* = A\widehat{\boldsymbol{\mu}}+\boldsymbol{b}$, it is straightforward to show that

$$\widehat{\boldsymbol{S}}_i^* = (A\widehat{V}A^T)^{-1/2}A\widehat{V}^{1/2}\widehat{\boldsymbol{S}}_i =: P\widehat{\boldsymbol{S}}_i.$$

Transformation matrix P is clearly orthogonal. Similarly $\widehat{\boldsymbol{S}}_{ij}^* = P\widehat{\boldsymbol{S}}_{ij}$ and $\widehat{\boldsymbol{R}}_i^* = P\widehat{\boldsymbol{R}}_i$.

Proof of Theorem 1   To shorten the notations in what follows, we write $\boldsymbol{x}_{ij}^{(1)} = \boldsymbol{x}_i^{(1)} - \boldsymbol{x}_j^{(1)}$ and $\boldsymbol{x}_{ij}^{(2)} = \boldsymbol{x}_i^{(2)} - \boldsymbol{x}_j^{(2)}$. Due to the affine invariance of the tests we consider only the case with spherical marginal densities. First, Gieser and Randles (1997) showed that Q_1^2 and

$$\left\| \frac{1}{n}\sum_i \frac{\boldsymbol{x}_i^{(1)}\boldsymbol{x}_i^{(2)^T}}{||\boldsymbol{x}_i^{(1)}||||\boldsymbol{x}_i^{(2)}||} \right\|^2$$

are asymptotically equivalent. Same asymptotic representation is found for Q_2^2 in the proof of Theorem 1 in Taskinen et al. (2003a). Second, under the null model, the interdirection proportions $p^{(1)}(i,j;i',j')$ are U-statistics with bounded kernels and expected values

$$E(p^{(1)}(i,j;i',j')) = \frac{1}{\pi}\cos^{-1}\left(\frac{\boldsymbol{x}_{ij}^{(1)^T}\boldsymbol{x}_{i'j'}^{(1)}}{||\boldsymbol{x}_{ij}^{(1)}||||\boldsymbol{x}_{i'j'}^{(1)}||} \right),$$

and similarly for $p^{(2)}(i,j;i',j')$. Thus

$$p^{(1)}(i,j;i',j') = E(p^{(1)}(i,j;i',j')) + o_p(1).$$

As in Gieser and Randles (1997), one can then show that the multivariate Kendall's τ_1^2 is asymptotically equivalent to

$$\frac{1}{n^4}\sum_i\sum_j\sum_{i'}\sum_{j'}\cos(\pi E(p^{(1)}(i,j;i',j')))\cos(\pi E(p^{(2)}(i,j;i',j')))$$

$$= \left\| \frac{1}{n^2}\sum_i\sum_j \frac{\boldsymbol{x}_{ij}^{(1)}\boldsymbol{x}_{ij}^{(2)^T}}{||\boldsymbol{x}_{ij}^{(1)}||||\boldsymbol{x}_{ij}^{(2)}||} \right\|^2,$$

and similarly for Spearman's ρ_1^2. The proof is then completed by finding the same asymptotic representations for τ_2^2 and ρ_2^2; see the proof of Theorem 2.

Proof of Theorem 2

The limiting distribution of τ_2^2: As in the proof of Theorem 1 in Taskinen et al. (2003a), it can be shown that under H_0,

$$\sqrt{n}\left(\frac{1}{n^2}\sum_i\sum_j \widehat{\boldsymbol{S}}_{ij}^{(1)}\widehat{\boldsymbol{S}}_{ij}^{(2)^T} - \frac{1}{n^2}\sum_i\sum_j \frac{\boldsymbol{x}_{ij}^{(1)}\boldsymbol{x}_{ij}^{(2)^T}}{||\boldsymbol{x}_{ij}^{(1)}||\,||\boldsymbol{x}_{ij}^{(2)}||}\right) \xrightarrow{P} 0.$$

Hence it is enough to derive the limiting distribution of

$$W_K = \frac{1}{n^2}\sum_i\sum_j \frac{\boldsymbol{x}_{ij}^{(1)}\boldsymbol{x}_{ij}^{(2)^T}}{||\boldsymbol{x}_{ij}^{(1)}||\,||\boldsymbol{x}_{ij}^{(2)}||}.$$

Since $vec(W_K)$ is asymptotically equivalent with a U-statistic with kernel

$$h\left(\begin{pmatrix}\boldsymbol{x}_1^{(1)}\\\boldsymbol{x}_1^{(2)}\end{pmatrix},\begin{pmatrix}\boldsymbol{x}_2^{(1)}\\\boldsymbol{x}_2^{(2)}\end{pmatrix}\right) = vec\left(\frac{\boldsymbol{x}_{12}^{(1)}\boldsymbol{x}_{12}^{(2)^T}}{||\boldsymbol{x}_{12}^{(1)}||\,||\boldsymbol{x}_{12}^{(2)}||}\right),$$

the limiting distribution of $\sqrt{n}vec(W_K)$ is asymptotically normal with expected value

$$E\left[vec\left(\frac{\boldsymbol{x}_{12}^{(1)}\boldsymbol{x}_{12}^{(2)^T}}{||\boldsymbol{x}_{12}^{(1)}||\,||\boldsymbol{x}_{12}^{(2)}||}\right)\right] = \boldsymbol{0}$$

and variance $4\Sigma_K$, where

$$\Sigma_K = E\left[vec\left(\frac{\boldsymbol{x}_{12}^{(1)}\boldsymbol{x}_{12}^{(2)^T}}{||\boldsymbol{x}_{12}^{(1)}||\,||\boldsymbol{x}_{12}^{(2)}||}\right)vec^T\left(\frac{\boldsymbol{x}_{13}^{(1)}\boldsymbol{x}_{13}^{(2)^T}}{||\boldsymbol{x}_{13}^{(1)}||\,||\boldsymbol{x}_{13}^{(2)}||}\right)\right].$$

Now for spherically distributed $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$,

$$\begin{aligned}
\Sigma_K &= E\left[E\left[vec\left(\frac{\boldsymbol{x}_{12}^{(1)}\boldsymbol{x}_{12}^{(2)^T}}{||\boldsymbol{x}_{12}^{(1)}||\,||\boldsymbol{x}_{12}^{(2)}||}\right)vec^T\left(\frac{\boldsymbol{x}_{13}^{(1)}\boldsymbol{x}_{13}^{(2)^T}}{||\boldsymbol{x}_{13}^{(1)}||\,||\boldsymbol{x}_{13}^{(2)}||}\right)\Bigg|\begin{pmatrix}\boldsymbol{x}_1^{(1)}\\\boldsymbol{x}_1^{(2)}\end{pmatrix}\right]\right]\\
&= E\left[vec\left(\boldsymbol{R}_F(\boldsymbol{x}^{(1)})\boldsymbol{R}_G^T(\boldsymbol{x}^{(2)})\right)vec^T\left(\boldsymbol{R}_F(\boldsymbol{x}^{(1)})\boldsymbol{R}_G^T(\boldsymbol{x}^{(2)})\right)\right]\\
&= E\left[vec\left(q_F(r)q_G(r)\boldsymbol{u}^{(1)}\boldsymbol{u}^{(2)^T}\right)vec^T\left(q_F(r)q_G(r)\boldsymbol{u}^{(1)}\boldsymbol{u}^{(2)^T}\right)\right]\\
&= E\left[q_F^2(r)q_G^2(r)\right]E\left[vec\left(\boldsymbol{u}^{(1)}\boldsymbol{u}^{(2)^T}\right)vec^T\left(\boldsymbol{u}^{(1)}\boldsymbol{u}^{(2)^T}\right)\right]\\
&= c_F^2 c_G^2 I_{pq}/pq,
\end{aligned}$$

and the limiting distribution of $\sqrt{n}vec(W_K)$ is $N_{pq}\left(\boldsymbol{0}, 4\,c_F^2 c_G^2 I_{pq}/pq\right)$. Consequently,

$$\frac{npq}{4c_F^2 c_G^2}\tau_2^2 = \frac{npq}{4c_F^2 c_G^2}vec(W_K)^T vec(W_K) \xrightarrow{d} \chi_{pq}^2.$$

The limiting distribution of ρ_2^2. First note that under H_0,

$$\sqrt{n}\left(\frac{1}{n^3}\sum_i\sum_j\sum_k \widehat{\boldsymbol{S}}_{ij}^{(1)}\widehat{\boldsymbol{S}}_{ik}^{(2)^T} - \frac{1}{n^3}\sum_i\sum_j\sum_k \frac{\boldsymbol{x}_{ij}^{(1)}\boldsymbol{x}_{ik}^{(2)^T}}{||\boldsymbol{x}_{ij}^{(1)}||||\boldsymbol{x}_{ik}^{(2)}||}\right) \xrightarrow{P} 0.$$

Then write

$$W_S = \frac{1}{n^3}\sum_i\sum_j\sum_k \frac{\boldsymbol{x}_{ij}^{(1)}\boldsymbol{x}_{ik}^{(2)^T}}{||\boldsymbol{x}_{ij}^{(1)}||||\boldsymbol{x}_{ik}^{(2)}||}.$$

Since $vec(W_S)$ is asymptotically equivalent with a U-statistic with kernel

$$h\left(\begin{pmatrix}\boldsymbol{x}_1^{(1)}\\\boldsymbol{x}_1^{(2)}\end{pmatrix}, \begin{pmatrix}\boldsymbol{x}_2^{(1)}\\\boldsymbol{x}_2^{(2)}\end{pmatrix}, \begin{pmatrix}\boldsymbol{x}_3^{(1)}\\\boldsymbol{x}_3^{(2)}\end{pmatrix}\right)$$

$$= \frac{1}{3}\, vec\left(\frac{\boldsymbol{x}_{12}^{(1)}\boldsymbol{x}_{13}^{(2)^T}}{||\boldsymbol{x}_{12}^{(1)}||||\boldsymbol{x}_{13}^{(2)}||} + \frac{\boldsymbol{x}_{21}^{(1)}\boldsymbol{x}_{23}^{(2)^T}}{||\boldsymbol{x}_{21}^{(1)}||||\boldsymbol{x}_{23}^{(2)}||} + \frac{\boldsymbol{x}_{31}^{(1)}\boldsymbol{x}_{32}^{(2)^T}}{||\boldsymbol{x}_{31}^{(1)}||||\boldsymbol{x}_{32}^{(2)}||}\right),$$

the limiting distribution of $\sqrt{n}vec(W_S)$ is asymptotically normal with expected value $\boldsymbol{0}$ and variance $9\Sigma_S$, where

$$\Sigma_S = E\left[h\left(\begin{pmatrix}\boldsymbol{x}_1^{(1)}\\\boldsymbol{x}_1^{(2)}\end{pmatrix}, \begin{pmatrix}\boldsymbol{x}_2^{(1)}\\\boldsymbol{x}_2^{(2)}\end{pmatrix}, \begin{pmatrix}\boldsymbol{x}_3^{(1)}\\\boldsymbol{x}_3^{(2)}\end{pmatrix}\right) \cdot h^T\left(\begin{pmatrix}\boldsymbol{x}_1^{(1)}\\\boldsymbol{x}_1^{(2)}\end{pmatrix}, \begin{pmatrix}\boldsymbol{x}_4^{(1)}\\\boldsymbol{x}_4^{(2)}\end{pmatrix}, \begin{pmatrix}\boldsymbol{x}_5^{(1)}\\\boldsymbol{x}_5^{(2)}\end{pmatrix}\right)\right].$$

As in the case of τ_2^2, it can be shown that $\Sigma_S = c_F^2 c_G^2 I_{pq}/9pq$. Hence the limiting distribution of $\sqrt{n}vec(W_S)$ is $N_{pq}\left(\boldsymbol{0}, c_F^2 c_G^2 I_{pq}/pq\right)$ and $[npq/c_F^2 c_G^2]\rho_2^2 \xrightarrow{d} \chi_{pq}^2$.

The limiting distribution of Q_2^2 is derived in Taskinen et al. (2003a).

Proof of Theorem 3

Gieser (1993) considered the asymptotic representation $L = \sqrt{n}\delta K - \frac{1}{2}\delta^2\sigma^2 + o_P(1)$, where

$$K = \frac{1}{n}\sum_i k_i = \frac{1}{n}\sum_i\left[p - \psi_1(r_i^{(1)})r_i^{(1)} + \psi_1(r_i^{(1)})r_i^{(2)}\boldsymbol{u}_i^{(1)^T}M_1\boldsymbol{u}_i^{(2)}\right.$$

$$\left. + q - \psi_2(r_i^{(2)})r_i^{(2)} + \psi_2(r_i^{(2)})r_i^{(1)}\boldsymbol{u}_i^{(2)^T}M_2\boldsymbol{u}_i^{(1)}\right],$$

with $r_i^{(1)} = ||\boldsymbol{x}_i^{(1)}||$, $\boldsymbol{u}_i^{(1)} = ||\boldsymbol{x}_i^{(1)}||^{-1}\boldsymbol{x}_i^{(1)}$ and $\psi_1(r) = \Psi_1'(r)$, and similarly for $r_i^{(2)}$, $\boldsymbol{u}_i^{(2)}$ and $\psi_2(r)$. If in the above representation $E(k_i) = 0$ and $E(k_i^2) = \sigma^2$,

the sequence of alternatives is contiguous to the null hypothesis. See Gieser (1993) for mild conditions.

As our test statistics are just squared norms of the U-statistics with bounded kernels, LeCam's third lemma could be easily applied to find the limiting distributions of the linearized versions of the U-statistics under the alternative sequences. We however give an alternative, somewhat heuristic proof to keep the formulas as simple and easily interpretable as possible.

The limiting distribution of τ_2^{2*}. The sequence of alternatives for $\Delta = \delta/\sqrt{n}$ is thus contiguous to the null hypothesis $\Delta = 0$. Therefore (see the proof of Theorem 2) also

$$\sqrt{n}\left(\frac{1}{n^2}\sum_i\sum_j \tilde{\boldsymbol{S}}_{ij}^{(1)}\tilde{\boldsymbol{S}}_{ij}^{(2)^T} - \frac{1}{n^2}\sum_i\sum_j \frac{\boldsymbol{y}_{ij}^{(1)}\boldsymbol{y}_{ij}^{(2)^T}}{||\boldsymbol{y}_{ij}^{(1)}||||\boldsymbol{y}_{ij}^{(2)}||}\right) \xrightarrow{P} 0,$$

where $\tilde{\boldsymbol{S}}_{ij}^{(1)}$ and $\tilde{\boldsymbol{S}}_{ij}^{(2)}$ are the standardized sign vectors of $\boldsymbol{y}_{ij}^{(1)}$ and $\boldsymbol{y}_{ij}^{(2)}$. As

$$\boldsymbol{y}_{ij}^{(1)} = \boldsymbol{x}_{ij}^{(1)} - \frac{\delta}{\sqrt{n}}\boldsymbol{x}_{ij}^{(1)} + \frac{\delta}{\sqrt{n}}M_1\boldsymbol{x}_{ij}^{(2)}$$

and

$$\boldsymbol{y}_{ij}^{(2)} = \boldsymbol{x}_{ij}^{(2)} - \frac{\delta}{\sqrt{n}}\boldsymbol{x}_{ij}^{(2)} + \frac{\delta}{\sqrt{n}}M_2\boldsymbol{x}_{ij}^{(1)},$$

we get

$$\frac{\boldsymbol{y}_{ij}^{(1)}}{||\boldsymbol{y}_{ij}^{(1)}||} = \frac{\boldsymbol{x}_{ij}^{(1)}}{||\boldsymbol{x}_{ij}^{(1)}||} - \frac{\delta}{\sqrt{n}}\frac{\boldsymbol{x}_{ij}^{(1)}\boldsymbol{x}_{ij}^{(1)^T}M_1\boldsymbol{x}_{ij}^{(2)}}{||\boldsymbol{x}_{ij}^{(1)}||^3} + \frac{\delta}{\sqrt{n}}\frac{M_1\boldsymbol{x}_{ij}^{(2)}}{||\boldsymbol{x}_{ij}^{(1)}||} + o_P(n^{-1/2})$$

and

$$\frac{\boldsymbol{y}_{ij}^{(2)}}{||\boldsymbol{y}_{ij}^{(2)}||} = \frac{\boldsymbol{x}_{ij}^{(2)}}{||\boldsymbol{x}_{ij}^{(2)}||} - \frac{\delta}{\sqrt{n}}\frac{\boldsymbol{x}_{ij}^{(2)}\boldsymbol{x}_{ij}^{(2)^T}M_2\boldsymbol{x}_{ij}^{(1)}}{||\boldsymbol{x}_{ij}^{(2)}||^3} + \frac{\delta}{\sqrt{n}}\frac{M_2\boldsymbol{x}_{ij}^{(1)}}{||\boldsymbol{x}_{ij}^{(2)}||} + o_P(n^{-1/2}).$$

But then in the spherical case,

$$\frac{\sqrt{n}}{n^2}\sum_i\sum_j \tilde{\boldsymbol{S}}_{ij}^{(1)}\tilde{\boldsymbol{S}}_{ij}^{(2)^T} = \frac{\sqrt{n}}{n^2}\sum_i\sum_j \frac{\boldsymbol{x}_{ij}^{(1)}\boldsymbol{x}_{ij}^{(2)^T}}{||\boldsymbol{x}_{ij}^{(1)}||||\boldsymbol{x}_{ij}^{(2)}||}$$

$$-\frac{\delta}{n^2}\sum_i\sum_j \frac{\boldsymbol{x}_{ij}^{(1)}\boldsymbol{x}_{ij}^{(1)^T}M_2^T\boldsymbol{x}_{ij}^{(2)}\boldsymbol{x}_{ij}^{(2)^T}}{||\boldsymbol{x}_{ij}^{(1)}||||\boldsymbol{x}_{ij}^{(2)}||^3} + \frac{\delta}{n^2}\sum_i\sum_j \frac{\boldsymbol{x}_{ij}^{(1)}\boldsymbol{x}_{ij}^{(1)^T}M_2^T}{||\boldsymbol{x}_{ij}^{(1)}||||\boldsymbol{x}_{ij}^{(2)}||}$$

$$-\frac{\delta}{n^2}\sum_i\sum_j \frac{\boldsymbol{x}_{ij}^{(1)}\boldsymbol{x}_{ij}^{(1)^T}M_1\boldsymbol{x}_{ij}^{(2)}\boldsymbol{x}_{ij}^{(2)^T}}{||\boldsymbol{x}_{ij}^{(2)}||||\boldsymbol{x}_{ij}^{(1)}||^3} + \frac{\delta}{n^2}\sum_i\sum_j \frac{M_1\boldsymbol{x}_{ij}^{(2)}\boldsymbol{x}_{ij}^{(2)^T}}{||\boldsymbol{x}_{ij}^{(2)}||||\boldsymbol{x}_{ij}^{(1)}||} + o_P(1).$$

Hence

$$\sqrt{n}\left(\frac{1}{n^2}\sum_i\sum_j \tilde{\boldsymbol{S}}_{ij}^{(1)}\tilde{\boldsymbol{S}}_{ij}^{(2)^T} - \frac{1}{n^2}\sum_i\sum_j \frac{\boldsymbol{x}_{ij}^{(1)}\boldsymbol{x}_{ij}^{(2)^T}}{||\boldsymbol{x}_{ij}^{(1)}||\,||\boldsymbol{x}_{ij}^{(2)}||}\right) \xrightarrow{P}$$

$$\frac{\delta}{pq}\Bigg[(p-1)E(||\boldsymbol{x}_{ij}^{(2)}||)E(||\boldsymbol{x}_{ij}^{(1)}||^{-1})M_1$$

$$+ (q-1)E(||\boldsymbol{x}_{ij}^{(1)}||)E(||\boldsymbol{x}_{ij}^{(2)}||^{-1})M_2^T\Bigg],$$

which completes the proof.

The limiting distribution of ρ_2^{2*}. If we proceed as above, we get

$$\sqrt{n}\left(\frac{1}{n^3}\sum_i\sum_j\sum_k \tilde{\boldsymbol{S}}_{ij}^{(1)}\tilde{\boldsymbol{S}}_{ik}^{(2)^T} - \frac{1}{n^3}\sum_i\sum_j\sum_k \frac{\boldsymbol{y}_{ij}^{(1)}\boldsymbol{y}_{ik}^{(2)^T}}{||\boldsymbol{y}_{ij}^{(1)}||\,||\boldsymbol{y}_{ik}^{(2)}||}\right) \xrightarrow{P}$$

$$\frac{\delta}{pq}\Bigg[(p-1)E_G(q_G(r)r)E(||\boldsymbol{x}_{ij}^{(1)}||^{-1})M_1$$

$$+ (q-1)E_F(q_F(r)r)E(||\boldsymbol{x}_{ij}^{(2)}||^{-1})M_2^T\Bigg].$$

Finally note that $2E_G(q_G(r)r) = E(||\boldsymbol{x}_{ij}^{(2)}||)$ and $2E_F(q_F(r)r) = E(||\boldsymbol{x}_{ij}^{(1)}||)$.

The limiting distribution of Q_2^{2*} is derived in Taskinen et al. (2003a).

APPENDIX B: EXPRESSIONS FOR c_F^2 AT CERTAIN MULTIVARIATE DISTRIBUTIONS

The constants c_F^2 for standard multivariate normal distribution $F = \Phi$, multivariate t distribution $F = t_\nu$ and multivariate contaminated normal distribution $F = \Phi_{\epsilon,c}$ are

$$c_\Phi^2 = \frac{1}{2\pi}\frac{\Gamma^2\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}\sum_{i=0}^{\infty}\frac{\Gamma^2\left(\frac{1}{2}+i\right)}{\Gamma\left(\frac{k+2}{2}+i\right)i!}\left(\frac{1}{4}\right)^i,$$

$$c_{t_\nu}^2 = \frac{\Gamma\left(\frac{\nu+k}{2}\right)\Gamma\left(\frac{3\nu}{2}\right)}{\Gamma^3\left(\frac{\nu}{2}\right)\Gamma\left(\frac{k}{2}\right)}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\Bigg\{\frac{\Gamma\left(\frac{k+1}{2}+i\right)\Gamma\left(\frac{\nu+1}{2}+i\right)\Gamma\left(\frac{k+1}{2}+j\right)}{\Gamma\left(\frac{k+2}{2}+i\right)\Gamma\left(\frac{k+2}{2}+j\right)}$$

$$\frac{\Gamma\left(\frac{\nu+1}{2}+j\right)\Gamma\left(\frac{k+2}{2}+i+j\right)}{\Gamma\left(\frac{3\nu+k+2}{2}+i+j\right)i!j!}\Bigg\}$$

and

$$c^2_{\Phi_{\epsilon,c}} = \frac{1}{2\pi} \frac{\Gamma^2\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \sum_{i=0}^{\infty} \frac{\Gamma^2\left(\frac{1}{2}+i\right)}{\Gamma\left(\frac{k+2}{2}+i\right)i!} c_{i,\epsilon,c},$$

where

$$c_{i,\epsilon,c} = \frac{\epsilon^3 + (1-\epsilon)^3}{4^i} + \frac{2(1-\epsilon)\epsilon^2 + 2\epsilon(1-\epsilon)^2 c^{4i+2}}{(c^2+1)^{2i+1}}$$
$$+ \frac{\epsilon(1-\epsilon)^2 + (1-\epsilon)\epsilon^2 c^{2i+1}}{2^{i-\frac{3}{2}}(c^2+1)^{i+\frac{1}{2}}}.$$

Expressions for the multivariate normal and t distribution cases are derived in Möttönen, Oja and Tienari (1997). The expression for the contaminated normal case is found in a similar way. Constants for selected distributions and dimensions are listed in Table 4.

Table 4: Constants c^2_F for selected distributions and dimensions.

k	c^2_Φ	$c^2_{t_{10}}$	$c^2_{t_5}$	$c^2_{\Phi_{0.1,3}}$	$c^2_{\Phi_{0.1,6}}$
2	0.4063	0.4046	0.4028	0.4025	0.3980
3	0.4360	0.4333	0.4302	0.4299	0.4231
5	0.4614	0.4574	0.4531	0.4529	0.4441
8	0.4760	0.4711	0.4660	0.4659	0.4560
10	0.4808	0.4757	0.4702	0.4702	0.4600

REFERENCES

Blomqvist, N. (1950), "On a Measure of Dependence Between Two Random Variables", *Annals of Mathematical Statistics*, 21, 593–600.

Dümbgen, L. (1998), "On Tyler's M-Functional of Scatter in High Dimension", *Annals of the Institute of Statistical Mathematics*, 50, 471–491.

Gieser, P.W. (1993), *A New Nonparametric Test of Independence Between Two Sets of Variates*, unpublished Ph.D. dissertation, University of Florida, Gainesville. (http://www.cyberpete.com/thesis.do)

Gieser, P.W. and Randles, R.H. (1997), "A Nonparametric Test of Independence Between Two Vectors", *Journal of the American Statistical Association*, 92, 561–567.

Hettmansperger, T.P. and Randles, R. (2002), "A Practical Affine Equivariant Multivariate Median", *Biometrika*, 89, 851–860.

Kendall, M.G. (1938), "A New Measurement of Rank Correlation", *Biometrika*, 30, 81–93.

McNaughton L. and Davies, P. (1987), "The Effects of a 16 Week Aerobic Conditioning Program on Serum Lipids, Lipoproteins and Coronary Risk Factors", *J. Sports Med.*, 27, 296–302.

Muirhead, R.J. (1982), *Aspects of Multivariate Statistical Theory*, New York: Wiley.

Möttönen J. and Oja, H. (1995), "Multivariate Spatial Sign and Rank Methods", *Journal of nonparametric statistics*, 5, 201–213.

Möttönen J., Oja, H. and Tienari J. (1997), "On the Efficiency of Multivariate Spatial Sign and Rank Tests", *The Annals of Statistics*, 25, 542–552.

Oja, H. (1983), "Descriptive Statistics for Multivariate Distributions", *Statistics & Probability Letters*, 1, 327–332.

Puri, M.L. and Sen, P.K. (1971), *Nonparametric Methods in Multivariate Analysis*, New York: Wiley.

Randles, R.H. (1989), "A Distribution-Free Multivariate Sign Test Based on Interdirections", *Journal of the American Statistical Association*, 84, 1045–1050.

Spearman, C. (1904), "The Proof and Measurement of Association Between Two Things", *American Journal of Psychology,*, 15, 72–101.

Taskinen, S., Kankainen, A. and Oja, H. (2003a), "Sign Test of Independence Between Two Random Vectors", *Statistics & Probability Letters*, 62, 9–21.

—— (2003b), "Tests of Independence Based on Sign and Rank Covariances", in *Developments in Robust Statistics. International Conference on Robust Statistics 2001*, eds. R. Dutter, P. Filzmoser, U. Gather and P.J. Rousseeuw, Heidelberg: Springer-Verlag, pp. 387–403.

Tyler, D.E. (1987), "A Distribution-Free M-Estimator of Multivariate Scatter", *The Annals of Statistics*, 15, 234–251.

Visuri, S., Oja, H. and Koivunen, V. (2000), "Sign and Rank Covariance Matrices", *Journal of Statistical planning and inference*, 91, 557–575.

Wilks, S.S. (1935), "On the Independence of k Sets of Normally Distributed Statistical Variables", *Econometrica*, 3, 309–326.
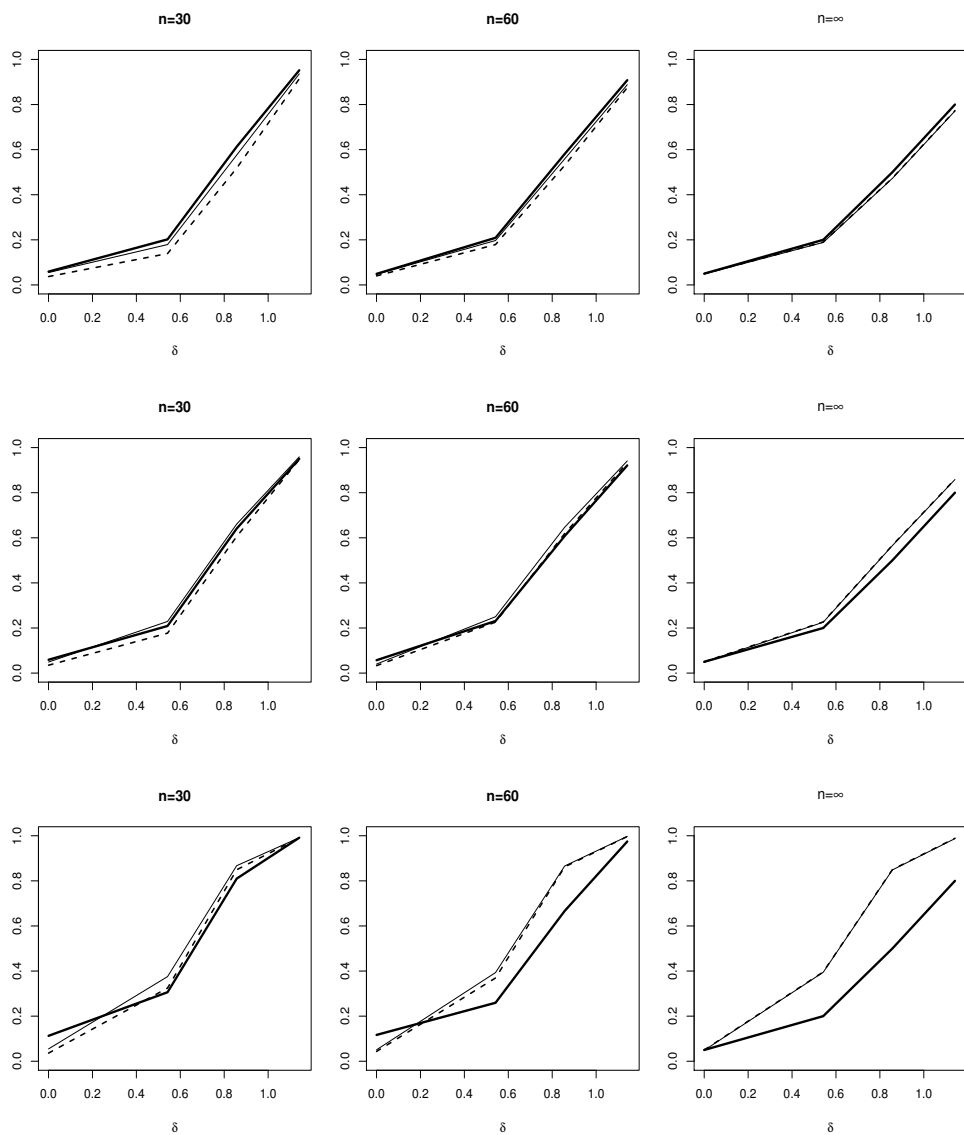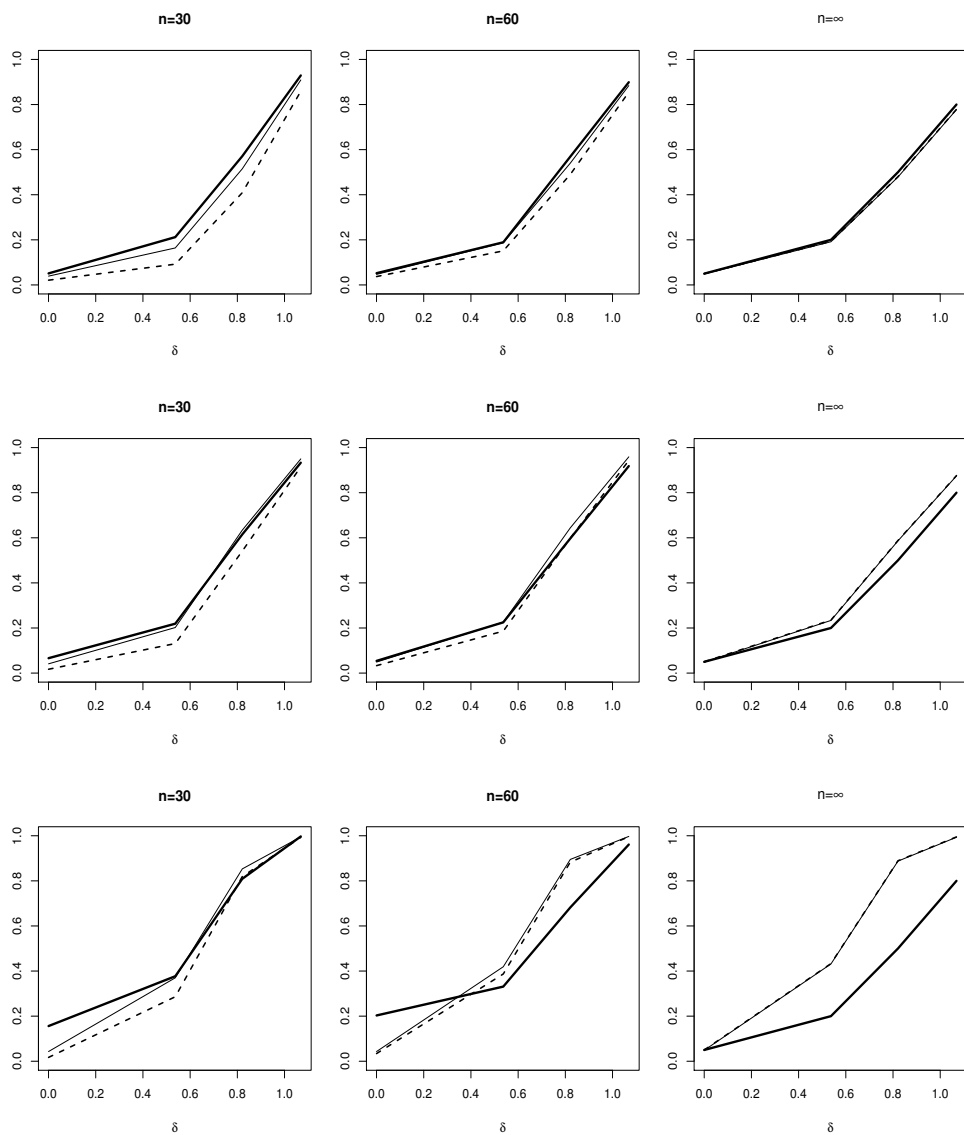
Figure 1: Empirical powers for $p = q = 3$ using the multivariate normal distribution (first row), multivariate t distribution with $\nu = 5$ (second row) and contaminated normal distribution with $\epsilon = 0.1$ and $c = 6$ (third row). The thick solid line denotes W, the thin solid line ρ_2^2 and the dotted line τ_2^2.

24

Figure 2: Empirical powers for $p = q = 5$ using the multivariate normal distribution (first row), multivariate t distribution with $\nu = 5$ (second row) and contaminated normal distribution with $\epsilon = 0.1$ and $c = 6$ (third row). The thick solid line denotes W, the thin solid line ρ_2^2 and the dotted line τ_2^2.
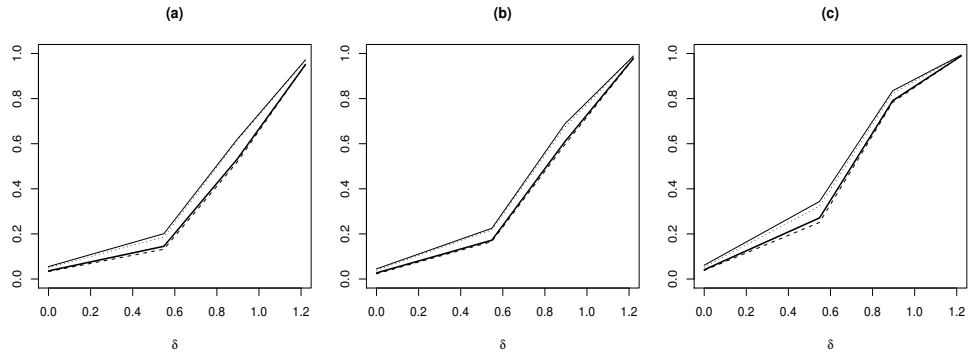
Figure 3: Empirical powers for $n = 15$ and $p = q = 2$ using the multivariate normal distribution (a), multivariate t distribution with $\nu = 5$ (b) and contaminated normal distribution with $\epsilon = 0.1$ and $c = 6$ (c). The thick solid line denotes τ_2^2, the thin solid line ρ_2^2, the thick dotted line τ_1^2 and the thin dotted line ρ_2^2.
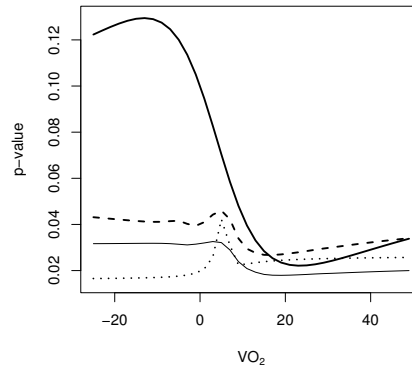

Figure 4: p-values for aerobic conditioning program data when the first VO_2 measurement is moved from -25 to 50. The thick solid line denotes W, the thin solid line ρ_2^2, the thick dotted line τ_2^2 and the thin dotted line Q_2^2.