Department of Applied Mathematics
Faculty of EEMCS

University of Twente
The Netherlands

P.O. Box 217
7500 AE Enschede
The Netherlands
Phone: +31-53-4893400
Fax: +31-53-4893114
Email: memo@math.utwente.nl
www.math.utwente.nl/publications

Memorandum No. 1689

**Detecting positive quadrant dependence
and positive function dependence**

A. Janic-Wróblewska,[1] W.C.M. Kallenberg
and T. Ledwina[2]

September, 2003

[1]Institute of Mathematics, Wrocław University of technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
[2]Institute of Mathematics, Polish Academy of Science, ul. Kopernika 18, 51-617 Wrocław, Poland

Detecting positive quadrant dependence and positive function dependence

A. Janic-Wróblewska, W.C.M. Kallenberg and T. Ledwina

Institute of Mathematics
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27
50-370 Wrocław
Poland

Faculty of Electrical Engineering, Mathematics and Computer Science
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

Institute of Mathematics
Polish Academy of Science
ul. Kopernika 18
51-617 Wrocław
Poland

Abstract There is a lot of interest in positive dependence going beyond linear correlation. In this paper three new rank tests for testing independence against positive dependence are introduced. The first one is directed on positive quadrant dependence, the second and third one concentrate on positive function dependence. The new testing procedures are not only sensitive for positive grade linear correlation, but also for positive grade correlations of higher order. They are based on the principle of data driven tests, which consists of three steps. Firstly, parametric families are introduced spanning up the space of null hypothesis and alternatives; secondly, within the families good tests are used; thirdly, a selection rule determines the appropriate model. The new tests improve standard tests for linear correlation as Spearman's rank correlation test substantially in case some proper higher order correlations are exhibited by the data, while the loss in power under alternatives with dominating linear correlation is not very high. Monte Carlo results clearly show this behavior.

Keyword and phrases: positive quadrant dependence, positive function dependence, rank test, model selection, Monte Carlo study, projected Legendre polynomials.

2000 Mathematics Subject Classification: 62 G 10, 62 H 20, 65 C 05.

1   Introduction

Recently there is much attention on the effects of positive dependence among risks. Positive dependence may lead to substantial deviations, for instance in the stop-loss premiums, compared to the independence case; see e.g. Albers (1999), Denuit et al. (2001) and references therein. Another area, where positive (or negative) dependence is important is mathematical finance. For instance, one wish to know whether certain stocks are negatively dependent in order to build a well balanced portfolio. Also in many other fields there is a lot of interest in positive (or negative) dependence. The following citation from Mari and Kotz (2001) makes this clear. "Examples of interdependent meteorological phenomena in nature and interdependence in the medical, social, and political aspects of our existence, not to mention the economic structures, are too numerous to be cited individually."

The interest in positive dependence is not restricted to linear correlation. On the contrary, there is a still growing interest in going beyond linear correlation, see e.g. Embrechts et al. (2002). In this paper we concentrate on positive dependence, not restricting ourselves to linear correlation. Similar results can be formulated and proved for negative dependence, but see Remark 1.2.

The idea behind positive dependence of two random variables X and Y is the property that large (respectively small) values of X (or functions of X) go together with large (respectively small) values of Y (or functions of Y). This can be expressed in various ways resulting in many notions of positive dependence. Metry and Sampson (1991) introduce a family of 64 partial orders for positive dependence on bivariate distributions with fixed marginals. The fact that there are so many notions of "positive dependence" confirms the importance of and the high interest in the concept. For a review see e.g. Shaked (1979), Scarsini and Shaked (1996) or Joe (1997).

In view of the tremendous consequences of wrongly ignoring positive dependence in favour of independence, it is of great importance to test independence against the restricted alternative of positive dependence. Under bivariate normality dependence is completely described by linear correlation and therefore testing independence under bivariate normality is equivalent to testing linear correlation between X and Y. In general however, X and Y are independent if and only if $cov(f_1(X), f_2(Y)) = 0$ for all f_1 and f_2 ranging over a separating class of functions (see e.g. Breiman (1968) page 165 ff). Hence, linear correlation is, although being an important aspect of dependence, not the only one, as was already clear from the fact that there are so many notions of positive dependence.

In this paper we concentrate on testing independence against two types of positive dependence, called positive quadrant dependence (PQD) and positive function dependence (PFD). In fact, we like to develop tests which have high power at (strictly) PQD- and PFD-alternatives, respectively and low power at distributions which are not strictly PQD and not strictly PFD, respectively. We look for tests which arrange their rejection region as much as possible on detecting distributions with positive dependence and have "low power" at distributions with negative dependence (or at distributions which are neither positive nor negative dependent). Although we are not formally testing "not positive dependent" (in the sense of PQD or PFD) against strictly positive dependent, from our point of view tests having high power at distributions which are not positive dependent give the wrong signal in such a problem.

New tests are introduced, which start with investigating linear correlation and consider step by step the correlation of larger and larger classes of functions. Since the marginal distributions of X and Y are unknown, we take the grade representation of the joint distribution of X and Y, that is we deal with $F(X)$ and $G(Y)$, where F and G are the unknown marginal distribution functions of X and Y, respectively. The nonnegative (linear) correlation between $F(X)$ and

$G(Y)$ and the two types of positive dependence, PQD and PFD, can be described as follows: (tacitly it assumed that we restrict to existing covariances)

1. $cov(F(X), G(Y)) \geq 0$, that concerns linear correlation;

2. $cov(f_1(F(X)), f_2(G(Y))) \geq 0$ for all nondecreasing f_1 and f_2; this is equivalent with PQD (see Theorem 4.4 in Esary et al. (1967)); we speak of strictly PQD when inequality holds for at least some pair of nondecreasing (f_1, f_2);

3. $cov(f(F(X)), f(G(Y))) \geq 0$ for all f; this is PFD (Joe (1997), page 25), or, equivalently, positive definite dependence, see Proposition 2.2 in Shaked (1979); we speak of strictly PFD when inequality holds for at least some function f.

Remark 1.1 For all three forms the covariances are 0 in case of independence. Another still stronger notion of positive dependence is defined by $cov(f_1(F(X), G(Y)), f_2(F(X), G(Y))) \geq 0$ for all f_1 and f_2 nondecreasing in both arguments. The random variables X and Y are called associated when this holds (Esary et al. (1967)). However, taking $f_1(u, v) = f_2(u, v)$ it is seen that independence of X and Y does not imply $cov(f_1(F(X), G(Y)), f_2(F(X), G(Y))) = 0$, although independence of X and Y implies that X and Y are associated, see Theorem 2.1 in Esary et al. (1967). Therefore, we do not use this notion of positive dependence.

Remark 1.2 The concept of PFD deals with positive dependence of functions of X and Y. Its counterpart, negative function dependence, could be defined as $cov(f(F(X)), f(G(Y))) \leq 0$ for all f. Consider random variables X, Y both with as marginal distribution the uniform distribution on $(0, 1)$. When $Y = 1 - X$ one might intuitively call this the strongest form of negative dependence. However, taking $f(x) = (x - \frac{1}{2})^2$, we get $cov(f(F(X)), f(G(Y))) = var\left\{(X - \frac{1}{2})^2\right\} > 0$ and hence, the pair $(X, 1 - X)$ is not negative function dependent.

It is clear that PQD (notion 2) and PFD (notion 3) imply positive linear correlation (notion 1). The idea is when investigating positive dependence, to start with linear correlation, as given in form 1 (or $f_1(u) = u, f_2(v) = v$ in terms of form 2 and $f(u) = u$ in terms of form 3). Subsequently, other pairs f_1, f_2 of nondecreasing functions (notion 2) or functions f (notion 3) are considered, thus describing further important aspects of positive dependence by the corresponding positive covariances. Successively, more complicated functions are involved. This is along the same line as with the simple goodness-of-fit testing problem, where we test the null hypothesis that a random variable has a given distribution. The first interest in that case will not be to examine whether the 289th moment agrees with that of the given distribution, but first attention will be focused on the location, then on the variance etc.

The (pairs of) functions should be chosen in such a way, that a new pair really offers new aspects. For instance, if our second pair of functions (notion 2) should be $f_1(u) = u$, $f_2(v) = v + 0.00001v^2$ we are repeating basically the linear correlation.

The idea is to build with k selected (pairs of) functions a k-dimensional parametric model, to test independence against strictly PQD. The (nonparametric) testing problem of testing independence against strictly PQD is relaxed in this k-dimensional parametric model to a parametric testing problem. Within this (easier) framework a suitable test statistic is derived. A similar approach has been carried out in Kallenberg and Ledwina (1999). The main difference with testing independence against dependence as treated in that paper is that the alternative of positive dependence leads to a *restricted* testing problem, which has implications both in building the model and in the testing problem within the k-dimensional parametric model.

In principle, the parametric models are based on the orthonormal Legendre polynomials. This orthonormal system has been successfully applied in the full independence testing problem

(see Kallenberg and Ledwina (1999)) and in many goodness-of-fit and other testing problems, where a large (nonparametric) testing problem is tackled by a sequence of parametric simplifications. However, PQD concerns *nondecreasing* functions and therefore we project the Legendre polynomials into the space of nondecreasing functions and take these projections as our building blocks.

In case of testing independence against strictly PFD all functions are involved and hence we do not need to project the Legendre polynomials in that situation.

Building up in this way more and more aspects of positive dependence, the question arises how much complexity should be taken into account. Since taking more (pairs of) functions in consideration will correspond with a higher dimensional model, the problem of complexity boils down to a model selection problem and hence an answer can be given from that area of statistics by applying Schwarz's (1978) rule, or a modification of it. This method of test construction has been introduced by Ledwina (1994) and has proven to give a nice solution in a broad spectrum of similar testing problems; see e.g. Ledwina (1994), Kallenberg and Ledwina (1999), Albers et al. (2001) and references therein. Having selected the appropriate dimension k, the test statistic in the k-dimensional parametric model is chosen. In deriving the critical value for the whole procedure we should of course take into account that the k-dimensional parametric model is chosen by the data.

Summarizing the data driven test construction has the following three steps:

step 1 introduce nice nested families spanning up the space of null hypothesis and alternatives (strictly PQD or strictly PFD);

step 2 propose related to the families in step 1 good tests to test within the family under consideration;

step 3 propose a good selection rule.

The paper is organized as follows. We start with testing independence against strictly PQD. In Section 2 the k-dimensional model with the projected Legendre polynomials is introduced and our score type test for testing independence against strictly PQD is derived within this k-dimensional model. In particular we take into account that we should get low power when the distribution is not strictly PQD. These test statistics in combination with the selection rule lead to our data driven test statistic V^+. Although the data driven test can be based on k-dimensional models with $k = 1, .., d(n)$, where $d(n)$ is a sequence of numbers tending to infinity as $n \to \infty$, we concentrate on the one- and two-dimensional models, similarly as in Kallenberg and Ledwina (1999). In Section 3 the same program is executed for testing independence against strictly PFD resulting in the test statistic T^+. A more liberal test statistic T^o is also introduced. An extensive simulation study is performed and results are presented in Section 4. The new tests are compared with several tests for positive dependence. It turns out that all three new tests improve substantially for instance Spearman's rank correlation test when higher order correlations play an important role in the distribution at hand. Moreover, when rejecting with a data driven test, one can look at the several components of the procedure to figure out what the reason is for rejecting and model the underlying distribution using these components. The three tests V^+, T^+ and T^o are expected to have higher power than the corresponding unrestricted tests V and $TS2$ in Kallenberg and Ledwina (1999), since V^+, T^+ and T^o are tuned to distributions with positive dependence. Indeed, in our simulation study this clearly comes up.

To help to imagine which deviations can be detected by V^+ we present in the Appendix the first five projected Legendre polynomials. Also we give there details about the distributions

4

used in the simulation study and further we present there some theoretical properties of T^+ and T^o, their asymptotic null distribution and consistency. The proofs are get patterning those of Kallenberg and Ledwina (1999). Finally, we give an easy, but accurate approximation for the critical values and/or p-values of T^+ and T^o. Similar derivations for V^+ would require much more involved technical work and are not developed here.

2   Testing PQD

Let $(X_1, Y_1), ..., (X_n, Y_n)$ be i.i.d. random variables with distribution function D and marginal distribution functions F and G for X_1 and Y_1, respectively. It will be assumed that F and G are continuous, but further they are unknown.

It is the aim to test the null hypothesis of independence against the alternative hypothesis of strictly PQD. The null hypothesis can be written as

$$H_0 : D(x,y) = F(x)G(y), \text{ for all } x, y \in \mathbb{R}$$

or, equivalently,

$$H_0 : cov(f_1(F(X)), f_2(G(Y))) = 0 \text{ for all nondecreasing } f_1 \text{ and } f_2,$$

see Theorem 4.4 in Esary et al. (1967) and Lemma 3 in Lehmann (1966). The alternative hypothesis is

$$H_1 : D(x,y) \geq F(x)G(y), \text{ for all } x, y \in \mathbb{R} \text{ with strict inequality for some pair } (x, y),$$

or, equivalently,

$$H_1 : cov(f_1(F(X)), f_2(G(Y))) \geq 0 \text{ for all nondecreasing } f_1 \text{ and } f_2$$
$$\text{with strict inequality for at least some pair of nondecreasing } (f_1, f_2).$$

The class of all nondecreasing functions is very large, which makes the testing problem complicated. Our strategy is to take some well chosen functions and apply the testing problem in the (much) simpler framework of these functions. For instance, for given nondecreasing functions f_1 and f_2 (with $\int f_j(u)du = 0, \int f_j^2(u)du = 1, j = 1, 2$) we consider densities on the unit square given by

$$1 + \theta_{11}f_1(u)f_1(v) + \theta_{12}f_1(u)f_2(v)$$

as one of the parametric models.

To propose a handsome selection of nondecreasing functions we start with the normalized Legendre polynomials on $[0, 1]$. The first five of them are given by

$$
\begin{aligned}
b_1(x) &= \sqrt{3}(2x - 1) \\
b_2(x) &= \sqrt{5}(6x^2 - 6x + 1) \\
b_3(x) &= \sqrt{7}(20x^3 - 30x^2 + 12x - 1) \\
b_4(x) &= 3(70x^4 - 140x^3 + 90x^2 - 20x + 1) \\
b_5(x) &= \sqrt{11}\left(252x^5 - 630x^4 + 560x^3 - 210x^2 + 30x - 1\right).
\end{aligned}
$$

Using the Moriguti projection algorithm nicely described and generalized in Rychlik (2001), page 14, we have derived projections of the b_j's onto the cone of nondecreasing functions. After

normalization we got the functions b_j^* explicitly given in the Appendix. The functions obey the properties

$$\int_0^1 b_j^*(u)du = 0, \quad \int_0^1 \{b_j^*(u)\}^2 du = 1.$$

Obviously, b_1 is increasing on $(0,1)$ and thus $b_1^* = b_1$. It will serve as our first function to be investigated. Hence, the one-dimensional model is given by the densities

$$1 + \theta_{11}b_1^*(u)b_1^*(v), \quad -\tfrac{1}{3} \leq \theta_{11} \leq \tfrac{1}{3}.$$

(The restriction on θ_{11} deserves to have nonnegative functions.) It is easily seen that $\theta_{11} = 0$ corresponds to independence, while strictly PQD coincides with $\theta_{11} \in (0,1/3)$. Hence our testing problem here boils down to $H_0 : \theta_{11} = 0$ against $H_1 : \{\theta_{11} > 0\} \cap (0,1/3)$, while we want to have low power when $\theta_{11} < 0$.

If we would know the marginal distributions F and G, we would base the test on

$$\sum_{i=1}^n b_1^* \left(F(X_i)\right) b_1^* \left(G(Y_i)\right).$$

Since F and G are unknown, we replace them by the corresponding empirical distribution functions. Writing R_i for the rank of X_i among $X_1, ..., X_n$ and S_i for the rank of Y_i among $Y_1, ..., Y_n$, and applying the familiar correction of continuity we arrive at the statistic (after some rescaling)

$$V(1,1) = \frac{1}{\sqrt{n}} \sum_{i=1}^n b_1^* \left(\frac{R_i - \frac{1}{2}}{n}\right) b_1^* \left(\frac{S_i - \frac{1}{2}}{n}\right)$$

and we reject for large values of this test statistic. To match our notation with the two-dimensional case, we shall consider below, we write the test statistic as

$$V^+(1,1) = V(1,1)^2 I \{V(1,1) > 0\},$$

where $I\{A\}$ denotes the indicator function of the set A.

Remark 2.1 By putting $1 + \theta_{11}b_1^*(u)b_1^*(v)$ as one-dimensional model, the linear correlation of $F(X)$ and $G(Y)$ immediately comes in and hence is considered as the most important one, being investigated at the first place.

Remark 2.2 The test statistic $V(1,1)$ is in fact nothing else than *Spearman's rank correlation rho*.

The function b_2 is not nondecreasing and therefore we project this function onto the space of nondecreasing functions. The result, after normalization, is

$$
\begin{aligned}
b_2^*(u) &= -\frac{\sqrt{10}}{\sqrt{47}} I \left\{0 \leq u \leq \frac{3}{4}\right\} + \frac{8\sqrt{2}}{\sqrt{47}} b_2(u) I \left\{\frac{3}{4} < u \leq 1\right\} \\
&= -0.4613 \, I \{0 \leq u \leq 0.75\} + 1.6503 \, b_2(u) \, I \{0.75 < u \leq 1\}.
\end{aligned}
$$

We consider the two-dimensional models $\{(1,1), (r,s)\}$ with $r,s = 1,2$ and $(r,s) \neq (1,1)$, given by the densities

$$1 + \theta_{11}b_1^*(u)b_1^*(v) + \theta_{rs}b_r^*(u)b_s^*(v), \text{ where } (\theta_{11}, \theta_{rs}) \in \Theta_{rs}$$

with Θ_{rs} such that the functions are nonnegative and hence densities. Obviously, when $\theta_{11} = \theta_{rs} = 0$ we have independence and when $(\theta_{11}, \theta_{rs}) \in \Theta_{rs} \cap \{\theta_{11} \geq 0, \theta_{rs} \geq 0\} \setminus \{\theta_{11} = 0, \theta_{rs} = 0\}$ we get alternatives which are strictly PQD. Define

$$V(r,s) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} b_r^* \left(\frac{R_i - \frac{1}{2}}{n} \right) b_s^* \left(\frac{S_i - \frac{1}{2}}{n} \right)$$

and

$$\widetilde{V}(r,s) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} b_r^* \left(F(X_i) \right) b_s^* \left(G(Y_i) \right).$$

Remark 2.3 The statistic $V(r,s)$ is a so called linear rank statistic of the form

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} J \left(\frac{R_i - \frac{1}{2}}{n}, \frac{S_i - \frac{1}{2}}{n} \right)$$

with score function

$$J(u,v) = b_r^*(u) b_s^*(v),$$

see e.g. Schriever (1987). The corresponding functional equals

$$\int \int J \left(F(x), G(y) \right) dD(x,y),$$

giving

$$\int \int J \left(F(x), G(y) \right) dD(x,y) = cov(b_r^*(F(X)), b_s^*(G(Y)))$$

and hence $V(1,1)$ is related to our notion 1 of positive dependence, while for all r,s the statistic $V(r,s)$ is related to our notion 2 of positive dependence, since b_r^* and b_s^* are nondecreasing.

Define Γ_{rs} as the covariance matrix under independence of the vector

$$\left(\widetilde{V}(1,1), \widetilde{V}(r,s) \right),$$

that is, (with U uniformly distributed on $(0,1)$)

$$\Gamma_{12} = \Gamma_{21} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \text{ and } \Gamma_{22} = \begin{pmatrix} 1 & \rho^2 \\ \rho^2 & 1 \end{pmatrix}$$

$$\text{with } \rho = cov \left(b_1^*(U), b_2^*(U) \right) = \frac{27\sqrt{30}}{32\sqrt{47}} = 0.674101.$$

The score test for testing $(\theta_{11}, \theta_{rs}) = (0,0)$ against $(\theta_{11}, \theta_{rs}) \neq (0,0)$ when F and G are known is given by

$$\left(\widetilde{V}(1,1), \widetilde{V}(r,s) \right) \Gamma_{rs}^{-1} \left(\widetilde{V}(1,1), \widetilde{V}(r,s) \right)^{\top}, \tag{1}$$

where $^{\top}$ denotes the transpose of a vector. We adapt this statistic by taking the ranks (since F and G are unknown) and by adding indicator functions since we have a restricted testing problem, leading to our test statistic

$$V^+(r,s) = (V(1,1), V(r,s)) \Gamma_{rs}^{-1} (V(1,1), V(r,s))^{\top} I \{V(1,1) \geq 0, V(r,s) \geq 0\}$$

in the model $\{(1,1),(r,s)\}$. The indicator functions are putted there to get high power at strictly PQD distributions and low power otherwise, thus checking if *simultaneously* the covariances

$$cov(b_1^*(F(X)), b_1^*(G(Y))), cov(b_r^*(F(X)), b_s^*(G(Y)))$$

are positive.

The last step is the choice of the selection rule. The idea is that a more complex model is only applied if it gives a sufficiently large improvement. This is reflected by penalizing for higher k. Mimicking the Schwarz (1978) solution and replacing twice the log-likelihood by our score type statistic we choose dimension two if

$$\max\left\{V^+(1,2), V^+(2,1), V^+(2,2)\right\} - V^+(1,1) \geq \log n$$

and otherwise dimension one is chosen. This leads to the following data driven test statistic

$$V^+ = \begin{cases} V^+(1,1) & \text{if } \max\{V^+(1,2), V^+(2,1), V^+(2,2)\} - V^+(1,1) < \log n \\ \max\{V^+(1,2), V^+(2,1), V^+(2,2)\} - V^+(1,1) & \text{otherwise.} \end{cases} \quad (2)$$

Remark 2.4 It should be emphasized that one cannot have a strictly PQD distribution with linear correlation equal to 0. As soon as we have PQD and linear correlation equal to 0, we automatically have independence, see Lemma 3 in Lehmann (1966). As a consequence, one might say, that we can simply test for linear correlation, using for instance Spearman's rank correlation coefficient as test statistic. However, on the one hand not every distribution with positive linear correlation is PQD and hence such a test wastes its power to distributions which are not PQD. On the other hand, when for a strictly PQD-alternative higher order correlations are more prominent than linear correlation, a test directed only to linear correlation will have lower power for such alternatives. As mentioned in the introduction, recently there is a growing interest in other forms of positive dependence than only linear correlation and such forms should be detected as well.

In contrast, there exist strictly PFD distributions with linear correlation equal to 0. An example is the Sub-Gaussian (α) distribution, see d9 in Section 4. The fact that the linear correlation for strictly PQD distributions is always positive indicates that testing independence against strictly PQD is more difficult than testing independence against strictly PFD.

Remark 2.5 In principle, the test V^+ can be extended straightforwardly to general $d(n)$. For $d(n) = 3$ the three-dimensional models $[(1,1),(r,s),(k,l)]$ with $r, s, k, l = 1, 2, 3$, all pairs being different, come in (see also Remark 4.1). To obtain the test statistic within the model, we start with the score test for testing independence against dependence as in (1). Then we replace the unknown $F(X_i)$ and $G(Y_i)$ by their ranks (corrected for continuity) and add the appropriate indicator functions, resulting in the test statistic V_Γ^+ in the model Γ. The test statistic V^+ is now defined as follows. Let $\Gamma(k)$ be the set of k-dimensional models and let

$$V_k^+ = \max_{\Gamma \in \Gamma(k)} V_\Gamma^+, \quad (3)$$

then

$$V^+ = V_{S2}^+ \text{ with } S2 = \min\left\{k : 1 \leq k \leq d(n) : V_k^+ - k \log n \geq V_r^+ - r \log n, 1 \leq r \leq d(n)\right\}. \quad (4)$$

Unfortunately, the concrete implementation is rather laborious due to the large number of models,

$$\sum_{k=1}^{d(n)} \binom{d(n)^2 - 1}{k - 1},$$

8

yielding for instance 1918314818520 models for $d(n) = 10$. To overcome this problem, we may select a suitable subset $\check{\Gamma}(k)$ of $\Gamma(k)$ and define as test statistic $\check{V}^+ = \check{V}_{S2}^+$ with $\check{V}_k^+ = \max_{\Gamma \in \check{\Gamma}(k)} V_\Gamma^+$ and $S2$ given as in (4) with V_k^+ replaced by $\check{V}_k^+$. Note that formally the test statistic given in (2) can be viewed as $\check{V}^+$ with $\check{\Gamma}(k) = \Gamma(2)$ for $k \geq 2$.

3   Testing PFD

Here we follow the same program as in the previous section. The alternative hypothesis is

$$H_1 : cov(f(F(X)), f(G(Y))) \geq 0 \text{ for all } f \text{ with strict inequality for at least some } f.$$

Our family of models on the unit square for this testing problem is based on the Legendre polynomials and is given by

$$1 + \sum_{j=1}^{k} \theta_j b_j(u) b_j(v), \qquad (\theta_1, ..., \theta_k) \in \Theta_k,$$

where Θ_k is such that the above functions are nonnegative and hence probability densities. Obviously, when $\theta_1 = ... = \theta_k = 0$ we have independence. Restricting attention to parameter values in Θ_k satisfying $\theta_1 \geq 0, ..., \theta_k \geq 0$ with at least one strict inequality, we get alternatives which are strictly PFD. The score test for testing $(\theta_1, ..., \theta_k) = (0, ..., 0)$ against $(\theta_1, ..., \theta_k) \neq (0, ..., 0)$ when F and G are known is given by

$$\sum_{j=1}^{k} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} b_j(F(X_i)) b_j(G(Y_i)) \right\}^2. \tag{5}$$

Define

$$C_j = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} b_j\left(\frac{R_i - \frac{1}{2}}{n}\right) b_j\left(\frac{S_i - \frac{1}{2}}{n}\right). \tag{6}$$

Remark 3.1 The statistic C_j is a so called linear rank statistic of the form

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} J\left(\frac{R_i - \frac{1}{2}}{n}, \frac{S_i - \frac{1}{2}}{n}\right)$$

with score function

$$J(u, v) = b_j(u) b_j(v),$$

see e.g. Schriever (1987). The corresponding functional equals

$$\int \int J(F(x), G(y)) \, dD(x, y),$$

giving

$$\int \int J(F(x), G(y)) \, dD(x, y) = cov(b_j(F(X)), b_j(G(Y)))$$

and hence C_1 is related to our notion 1 of positive dependence, while for all j the statistic C_j is related to our notion 3 of positive dependence.

We adapt the test statistic (5) by taking the ranks (since F and G are unknown) and by adding indicator functions, since we have a restricted testing problem, similarly as in V^+, thus

checking *simultaneously* whether all the covariances $cov(b_j(F(X)), b_j(G(Y)))$ are positive. This leads to the test statistic

$$T_k^+ = \left[\sum_{j=1}^k C_j^2 \right] \left[\prod_{j=1}^k I\{C_j \geq 0\} \right] \tag{7}$$

in the k-dimensional model. The indicator function serves to get high power at strictly PFD alternatives and low power at distributions which are not PFD, similarly as in V^+. The selection rule is given by

$$S2^+ = \min\left\{k : 1 \leq k \leq d(n) : T_k^+ - k\log n \geq T_r^+ - r\log n, 1 \leq r \leq d(n)\right\}, \tag{8}$$

resulting in the data driven test statistic

$$T^+ = T_{S2^+}^+ \tag{9}$$

and we reject for large values of this test statistic. An accurate approximation of the distribution of T^+ under the null hypothesis facilitates its application by giving easy approximations for critical values and/or p-values. This approximation is given in the Appendix.

In contrast to strictly PQD distributions, there exist distributions which are strictly PFD with linear correlation equal to 0, see for instance the distributions denoted by d7 with $\rho = 0$, d8 and d9 in Section 4.2. On the set $\{C_1 < 0\}$ we have that each T_k^+ takes it smallest value (0) and hence we do not reject with the test statistic T^+ on the set $\{C_1 < 0\}$. In case of a linear correlation equal to 0, this set has asymptotic probability $\frac{1}{2}$ and hence for these strictly PFD distributions we should have (for large n) at most power $\frac{1}{2}$. Therefore, we also consider a more liberal version, where we check not simultaneously whether $cov(b_j(F(X)), b_j(G(Y)))$ is positive, but for each j separately. As soon as one of the C_j's is positive, the test statistic is positive. Its definition is as follows. Let

$$T_k^o = \sum_{j=1}^k [C_j I\{C_j \geq 0\}]^2$$

and

$$S2^o = \min\left\{k : 1 \leq k \leq d(n) : T_k^o - k\log n \geq T_r^o - r\log n, 1 \leq r \leq d(n)\right\},$$

then this data driven test statistic is given by

$$T^o = T_{S2^o}^o \tag{10}$$

and we reject for large values of this test statistic. Also for T^o an accurate and simple approximation of its distribution under the null hypothesis is given in the Appendix.

4   Simulation study

To see how well the new tests behave an extensive simulation study has been performed. Here we present the main results of this study. We do not restrict ourselves to alternatives, but consider occasionally also distributions with negative dependence (or distributions which are neither positive nor negative dependent) to see how the several tests react on those distributions. As mentioned in the introduction we look for tests that are sensitive for PQD or PFD and do not have very high power at distributions which are not PQD or PFD, respectively.

Note that all tests are based on the ranks $(R_1, S_1), ..., (R_n, S_n)$ and hence monotone transformations of X_i's and Y_i's have no influence on the power. For the distributions in the simulation

study we therefore may take, for instance, expectations equal to 0 and variances equal to 1 if desired.

The simulated "power" comparison presented here concerns $n = 50$ and level 0.05. Every Monte Carlo experiment reported below has been repeated $10\,000$ times. Hence, the standard deviation of simulated powers does not exceed $(40\,000)^{-\frac{1}{2}} = 0.005$.

Apart from presenting the powers, we also give (simulated) correlations, thus characterizing the distributions we deal with. They are defined as follows. When testing independence against strictly PQD, we consider the correlation c_{rs}^*, defined as

$$c_{rs}^* = E_{\mathbb{P}} b_r^* \left(F(X) \right) b_s^* \left(G(Y) \right),$$

where $\mathbb{P}$ is the distribution under consideration and F and G are the marginal distribution functions of X and Y, respectively, under $\mathbb{P}$. Denote the simulations of a distribution by $(X_{ij}, Y_{ij}), 1 \leq i \leq 50, 1 \leq j \leq 10\,000$. If an explicit formula for the grade presentation is available (as e.g. in the case of the Plackett distributions, see d1), c_{rs}^* is estimated by

$$\widehat{c}_{rs}^* = \frac{1}{10\,000} \sum_{j=1}^{10\,000} \frac{1}{50} \sum_{i=1}^{50} b_r^* \left(F(X_{ij}) \right) b_s^* \left(G(Y_{ij}) \right).$$

Otherwise, c_{rs}^* is estimated by

$$\widehat{c}_{rs}^* = \frac{1}{10\,000} \sum_{j=1}^{10\,000} \frac{1}{50} \sum_{i=1}^{50} b_r^* \left(\frac{R_{ij} - \frac{1}{2}}{50} \right) b_s^* \left(\frac{S_{ij} - \frac{1}{2}}{50} \right).$$

Here R_{ij} is the rank of X_{ij} among $X_{1j}, ..., X_{50j}$ and S_{ij} is the rank of Y_{ij} among $Y_{1j}, ..., Y_{50j}$ for each $j = 1, ..., 10\,000$. The (estimated) correlation $\widehat{c}_{rs}^*$ is presented as (r^*, s^*) in Figures 1 and 2.

In case of testing independence against strictly PFD, the correlations are in fact Fourier coefficients. For uniformity in presentation and since we like to detect higher order correlations we speak of the correlation c_{rs}, defined as

$$c_{rs} = E_{\mathbb{P}} b_r \left(F(X) \right) b_s \left(G(Y) \right).$$

They are estimated by

$$\widehat{c}_{rs} = \frac{1}{10\,000} \sum_{j=1}^{10\,000} \frac{1}{50} \sum_{i=1}^{50} b_r \left(F(X_{ij}) \right) b_s \left(G(Y_{ij}) \right)$$

or

$$\widehat{c}_{rs} = \frac{1}{10\,000} \sum_{j=1}^{10\,000} \frac{1}{50} \sum_{i=1}^{50} b_r \left(\frac{R_{ij} - \frac{1}{2}}{50} \right) b_s \left(\frac{S_{ij} - \frac{1}{2}}{50} \right).$$

The (estimated) correlation $\widehat{c}_{rs}$ is presented as (r, s) in Figure 3. In both cases we considered up to five b_j^*'s and b_j's, respectively.

4.1 Testing PQD

When testing independence against strictly PQD, we consider the following four test statistics.

(a) Spearman's rank correlation coefficient

$$RS = \sqrt{n-1} \left\{ \frac{12}{n \left(n^2 - 1 \right)} \sum_{i=1}^{n} R_i S_i - \frac{3 \left(n + 1 \right)}{n - 1} \right\}.$$

(b) The Behnen-Neuhaus statistic $S_n(V)$ given by (6.3.10) with $r = 4$ and score functions (6.3.11) on page 319 of Behnen and Neuhaus (1989). Here we denote this test statistic as BN^+. We have also considered the one-sided test S^0 defined on page 90 of Behnen and Neuhaus (1989). This is a very complicated "quadratic" statistic. Since its power behavior is rather similar to that of BN^+, we prefer to present BN^+ only.

(c) The statistic L^+ introduced in Ledwina (1986)

$$L^+ = \sup_{0<p<1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{R_i}{n} \right) (p - I\{S_i \leq np\})$$

in relation to the monotonic dependence function proposed by Kowalczyk and Pleszczyńska (1977).

(d) The new test statistic V^+, see (2).

For all four test statistics we reject for large values of it. The simulations yield the following critical values for $n = 50$ and level 0.05

$$RS : 1.66, BN^+ : 5.80, L^+ : 0.40, V^+ : 4.45.$$

The following distributions are presented. The PQD-properties of these distributions are derived in the Appendix. We borrowed the idea of modeling alternatives through some regression models from the simulation study contained in the PhD dissertation of Thas (2001).

d1 **Plackett** (ψ). This family of distributions is nicely described on pages $191 - 197$ of Johnson (1987). For $\psi = 1$ we have independence, $\psi > 1$ gives strictly PQD, while $\psi < 1$ corresponds to strictly negative quadrant dependence (NQD). Plackett noted the considerable similarity of his distribution with normal marginals to the usual bivariate normal distribution. Therefore, the usual bivariate normal distribution is not included for this testing problem. For the Plackett distributions the first correlation c_{11}^* is dominating, as is seen by $(1^*, 1^*)$ in Figure 1.

d2 **Cubic regression model** (β). Let X, Z be independent random variables each with a standard normal distribution and let

$$Y = \beta X^3 + Z.$$

For $\beta = 0$ we have independence, $\beta > 0$ gives strictly PQD, while $\beta < 0$ corresponds to strictly negative quadrant dependence. The correlation c_{33}^* is dominating, see $(3^*, 3^*)$ in Figure 1.

d3 **Bivariate stable** (a_3). In general, (X, Y) has a bivariate stable distribution with parameters $(a_1, a_2, a_3, b_1, b_2, b_3, h)$ if $X = Z_1 + hZ_3$, $Y = Z_2 + hZ_3$, $h \in \mathbb{R}$, where Z_1, Z_2, Z_3 are independent stable random variables, $Z_i \sim S(a_i, b_i)$, $i = 1, 2, 3$. In our application we used the parametrization of Chambers et al. (1976), page 341, and fixed $a_1 = 2, a_2 = 0.2, b_1 = 1, b_2 = -1, b_3 = 0, h = 1$. Then for any a_3 we have a strictly PQD distribution. For these distributions the correlation c_{22}^* is dominating, see $(2^*, 2^*)$ in Figure 1.
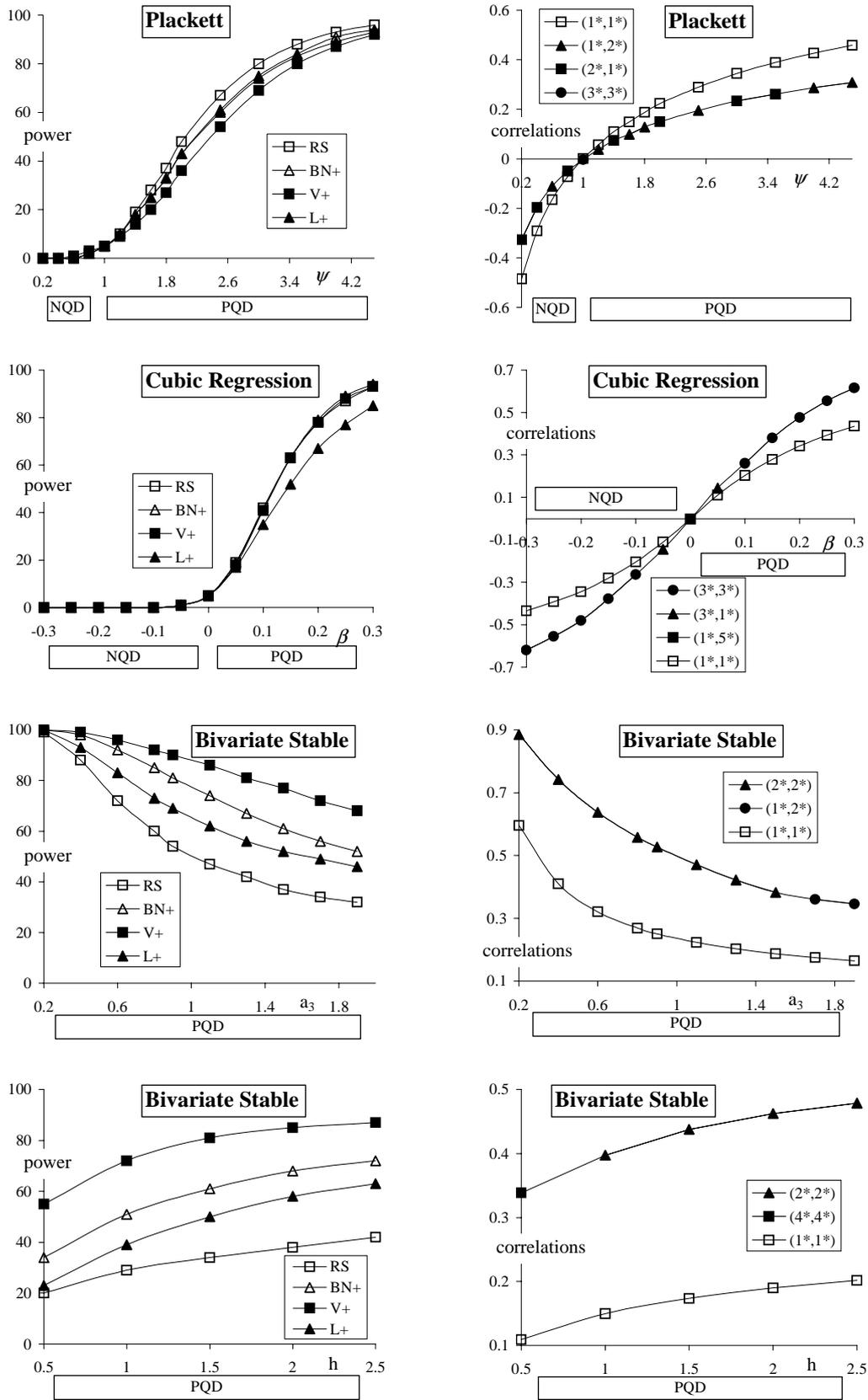
12

Figure 1. The simulated powers of 3 tests and the simulated correlation $\widehat{c}^*_{11}$ together with the (next) most important simulated correlation $\widehat{c}^*_{rs}$ at distributions $d1 - d4$.

d4 **Bivariate stable** (h) with $a_1 = 1.5, a_2 = 0.1, a_3 = 0.9, b_1 = 1, b_2 = -1, b_3 = -0.5, h$ varying. When $h = 0$ we have independence, otherwise we get strictly PQD. Here the correlation c_{22}^* is dominating, see $(2^*, 2^*)$ in Figure 1.

d5 **Heteroscedastic linear regression** (p). The distribution is obtained as follows. Let X be uniformly distributed at $(0, 1)$. Let Z be independent of X with a standard normal distribution. Then $Y = X + 4(1 - X^p) Z$, where $p > 0$ varies. These distributions are not PQD. The dominating correlation c_{12}^* is negative, see $(1^*, 2^*)$ in Figure 2 (at $p = 0.25$ we have $(1^*, 2^*) = -0.236$ and $(2^*, 1^*) = 0.287$).

The (simulated) powers of the three tests at the distributions d1 − d4 are shown in Figure 1. Moreover, the simulated correlation $\widehat{c}_{11}^*$ and the (next) most important simulated correlation $\widehat{c}_{rs}^*$ (among $\widehat{c}_{rs}^*$ with $r, s = 1, ..., 5$) of these distributions are also shown in Figure 1. For instance, in case of the Plackett distribution the most important (simulated) correlation (that is the one with the highest absolute value) among $\widehat{c}_{rs}^*$ with $r, s = 1, ..., 5$ is $\widehat{c}_{11}^* = (1^*, 1^*)$. The next most important one depends on the specific parameter value and is for example $(1^*, 2^*)$ at $\psi = 1.8$, $(2^*, 1^*)$ at $\psi = 2$, etc.

Since c_{11}^* is dominating for the Plackett alternatives, it is no surprise that RS has the highest power for those alternatives. However, the differences are not that large. In the cubic regression model the three test have almost exactly the same power. It is clearly seen that for the bivariate stable alternatives with dominating c_{22}^* great differences in power may occur. Although strictly PQD implies positive linear correlation, it is seen by the bivariate stable alternatives that V^+ improves RS substantially when higher order correlations are exhibited by the data.
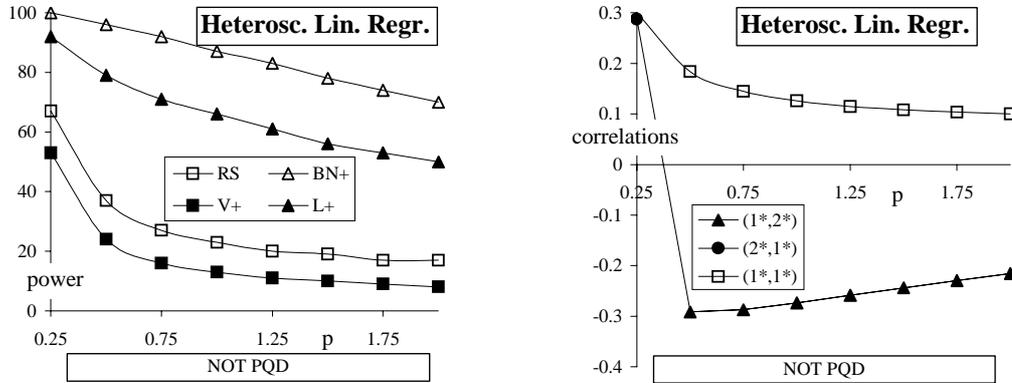

Figure 2. The simulated powers of 3 tests and the simulated correlation $\widehat{c}_{11}^*$ together with the (next) most important simulated correlation $\widehat{c}_{rs}^*$ at distribution d5.

Figure 2 shows that BN^+ and L^+ have very high powers at the d5 distributions, although they are not PQD: at all presented parameter values p, except for $p = 0.25$ the most important correlation, $(1^*, 2^*)$ is negative (at $p = 0.25$ we have $(1^*, 2^*) = -0.236$ and $(2^*, 1^*) = 0.287$). The other tests do not have such high power here, which we consider as an advantage, since the tests should concentrate as much as possible on positive dependence and not waste their power to distributions which are not PQD. Another advantage of V^+ is that it is based on easily interpretable estimated grade correlations and additional inspection of these correlations can be helpful in situations when we are not sure that the data at hand obey the PQD pattern.

Remark 4.1 We have also performed simulations for V^+ when $d(n) = 3$, see Remark 2.5. The simulated powers are as expected: some power loss when linear correlation is dominating and some gain otherwise.

In view of the results presented here and the other simulations that we have done we conclude that for testing independence against strictly PQD the new test based on V^+ performs very well for a wide variety of strictly PQD alternatives, going also behind the linear correlation.

4.2 Testing PFD

When testing independence against strictly PFD, we consider the following three test statistics.

(a) Spearman's rank correlation coefficient RS, see 4.1 (a).

(b) The new test statistic T^+, see (9).

(c) The new test statistic T^o, see (10).

For all three test statistics we reject for large values of it. The simulations yield as critical values of T^+ for $n = 50$ and level 0.05 : 2.81 if $d(50) = 1$, 3.12 if $d(50) = 2$, 3.20 if $d(50) = 3$, 3.24 if $d(50) \geq 4$. The critical values of T^o for $n = 50$ and level 0.05 are: 2.81 if $d(50) = 1$, 3.97 if $d(50) = 2$, 4.15 if $d(50) = 3$, 4.21 if $d(50) \geq 4$. This shows that for $d(n) \geq 2$ there is almost no change in the critical values of T^+ and T^o. This is a nice feature of these test statistics. Moreover, for data driven tests the power behavior is also very stable w.r.t. $d(n)$, see e.g. Kallenberg and Ledwina (1997). Therefore, the problem of choosing k in T_k^+ and T_k^o is definitely not replaced by the problem of choosing $d(n)$. In the simulation study the tests T^+ and T^o are applied with $d(50) = 10$.

Since BN^+ is designed for testing independence against strictly PQD we do not present this test here. Anyway, we simulated the powers and they are high for most of these distributions as well, although for the alternatives d8 and d9 (see below) serious power loss w.r.t T^o arises.

The following distributions are presented. The PFD-properties of these distributions are derived in the Appendix.

d6 **Bivariate normal** (ρ) with expectations equal to 0, variances equal to 1 and correlation coefficient ρ. The bivariate normal distribution is strictly PFD for $\rho > 0$, independence occurs for $\rho = 0$ and, remarkably (but see Remark 1.2), for $\rho < 0$ we get neither PFD nor strictly negative function dependence. For the bivariate normal distributions the first order correlation is dominating, as is seen by $(1, 1)$ in Figure 3, being positive for $\rho > 0$ and negative for $\rho < 0$, while the second order correlation, see $(2, 2)$, is positive for all $\rho \neq 0$ (for $\rho = -0.2, -0.1$ we get $(2, 2) = 0.029, 0.007$, respectively). Note that $b_2(x) = 6\sqrt{5} \left(x - \frac{1}{2}\right)^2 - \frac{1}{2}\sqrt{5}$, see also Remark 1.2.

d7 **Khintchine** (ρ), defined as $X = ZU, Y = ZV$, where (U, V) has a bivariate normal distribution with expectations equal to 0, variances equal to 1 and correlation coefficient ρ, and Z being independent of (U, V), is distributed as $\sqrt{\chi_3^2}$, where χ_3^2 is a random variable with a chi-square distribution with 3 degrees of freedom. (This is not literally the Khintchine distribution, see e.g. Bryson and Johnson (1982), but a slight modification of it. Nevertheless we simply call it the Khintchine distribution.) The Khintchine (ρ) distributions with $\rho \geq 0$ are strictly PFD, although for $\rho = 0$ the linear correlation equals 0, as is easily seen by symmetry arguments and independence of Z, U and V. For the Khintchine distributions with not too large ρ, the correlation c_{22} is dominating, while for $|\rho| \geq 0.3$ the linear correlation dominates, see Figure 3.
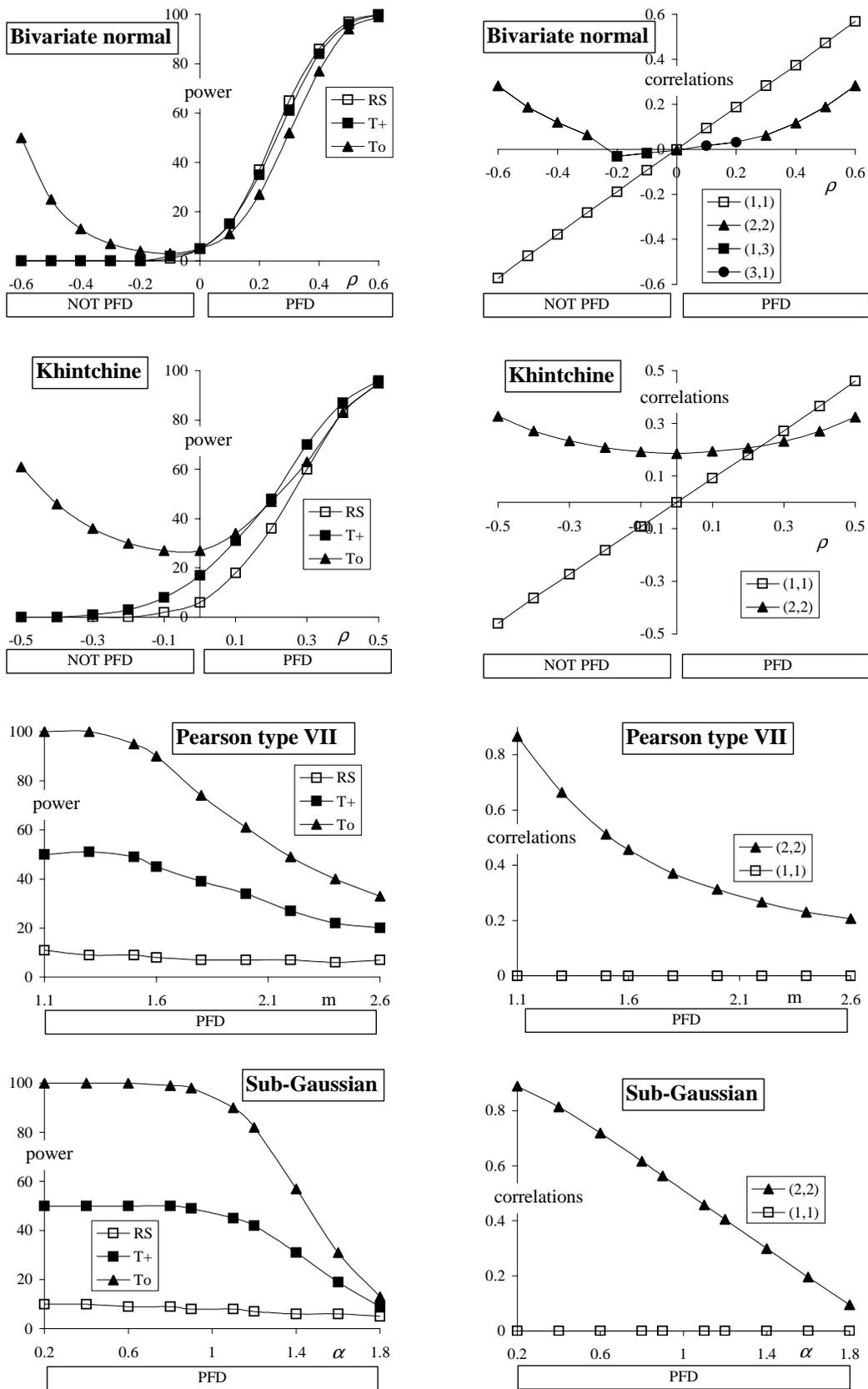
15

Figure 3. The simulated powers of 3 tests and the simulated correlation $\widehat{c}_{11}$ together with the (next) most important simulated correlation $\widehat{c}_{rs}$ at distributions $d6 - d9$.

d8 **Pearson type VII** (m) with $m > 1$, defined as $X = U/\sqrt{Z}, Y = V/\sqrt{Z}$, where U, V and Z are independent random variables, U and V with a standard normal distribution and Z having a Gamma distribution with parameters $(m-1, \frac{1}{2})$. The Pearson type VII (m) distributions with $m > 1$ are strictly PFD, although they have linear correlation equal to 0, as is easily seen by symmetry arguments. The correlation c_{22} is dominating, see Figure 3.

d9 **Sub-Gaussian** (α). This subfamily of the sub-Gaussian distributions is defined as $X = U\sqrt{Z}, Y = V\sqrt{Z}$, where U, V and Z are independent, U and V have a standard normal distribution and Z has a stable distribution, totally skewed to the right, with index of stability $\alpha/2$, see Samorodnitsky and Taqqu (1994). The Sub-Gaussian (α) distributions are strictly PFD, although they have linear correlation equal to 0, as is easily seen by symmetry arguments. The correlation c_{22} is dominating, see Figure 3.

The (simulated) powers of the three tests at the distributions d6 − d9 are shown in Figure 3. Moreover, the simulated correlation $\widehat{c}_{11}$ and the (next) most important simulated correlation $\widehat{c}_{rs}$ (among $\widehat{c}_{rs}$ with $r, s = 1, ..., 10$) of these distributions are also shown in Figure 3.

It is seen that T^+ and T^o perform far much better than RS, unless the first order correlation is dominating, as in the bivariate normal distribution (d6). A gain in power up to 90% for T^o is seen in d8 and d9! On the other hand, the loss in power at alternatives with dominating first order correlation is not very large. As explained in Section 3 when introducing T^+, in case of a linear correlation equal to 0 its asymptotic power is at most $\frac{1}{2}$. This is clearly shown at the alternatives d8 and d9: while for small parameter values the power of T^o comes close to 100%, the power of T^+ reaches 50%.

The asymptotic null distribution of T^+ and T^o and the consistency of the tests based on T^+ and T^o for a very large class of positive dependent alternatives are shown in the Appendix. Although formally T^+ is not consistent when the linear correlation equals 0, the alternatives d7 − d9 show that a substantial power gain w.r.t. Spearman's rank correlation test can be obtained when the linear correlation is (close to) 0. To facilitate their application a simple, but accurate approximation for the critical values and/or p-values of T^+ and T^o is given as well in the Appendix.

It is seen in the bivariate normal distribution (d6) that T^o may have substantial power at distributions which are not PFD. In that case the power comes from the fact that the correlation c_{22} is positive. For instance, when $\rho = -0.6$ we get $\widehat{c}_{22} = 0.28$, see Figure 3. This agrees with the consistency result (Theorem 2 in the Appendix), which states that the test based on T^o is consistent against any alternative with positive c_{jj} for some j. In this sense the test statistic T^o is liberal, which is no problem as long as we consider distributions that are either independent or strictly PFD.

In view of the results presented here and the other simulations that we have done we conclude that for testing independence against strictly PFD the new tests based on T^+ and T^o are useful and flexible tools to detect PFD alternatives, going also behind linear correlation.

Remark 4.2 All three tests V^+, T^+ and T^o are expected to have higher power than the corresponding unrestricted tests V and $TS2$ in Kallenberg and Ledwina (1999), since V^+, T^+ and T^o are tuned to distributions with positive dependence. Indeed, in our simulation study this clearly comes up. For instance, at each of the alternatives d1 − d4 the power of V^+ is about 0.20 higher than the power of V for parameter values such that the power is not too high

or too low. So, the gain of concentrating on PQD alternatives (V^+) or PFD alternatives (T^+ and T^o) is pretty much.

Appendix

In this appendix we give details about the projected Legendre polynomials, the distributions used in the simulation study and further we present here some theoretical properties of T^+ and T^o, their asymptotic null distribution and their consistency. Moreover, we give an easy, but accurate approximation for the critical values and/or p-values of T^+ and T^o.

Projected Legendre polynomials

The first five projected Legendre polynomials on the space of nondecreasing functions, normalized such that

$$\int\limits_0^1 \{b_i^*(u)\}^2 \, du = 1, \quad i = 1, 2, ...,$$

are given by

$$
\begin{aligned}
b_1^*(u) &= b_1(u) = \sqrt{3}(2u - 1) = 3.4641u - 1.7321; \\
b_2^*(u) &= -\frac{\sqrt{10}}{\sqrt{47}}I\left\{0 \leq u \leq \frac{3}{4}\right\} + \frac{8\sqrt{2}}{\sqrt{47}}b_2(u)I\left\{\frac{3}{4} < u \leq 1\right\} \\
&= -0.4613\, I\left\{0 \leq u \leq 0.75\right\} + 1.6503\, b_2(u)\, I\left\{0.75 < u \leq 1\right\}; \\
b_3^*(u) &= \frac{b_3(u)}{\sqrt{1 - \frac{18\sqrt{15}}{10}}}\left[I\left\{0 \leq u \leq \frac{1}{2} - \frac{\sqrt{15}}{10}\right\} + I\left\{\frac{1}{2} + \frac{\sqrt{15}}{10} \leq u \leq 1\right\}\right] \\
&= 1.2262\, b_3(u)\left[I\left\{0 \leq u \leq 0.1127\right\} + I\left\{0.8873 \leq u \leq 1\right\}\right]; \\
b_4^*(u) &= -0.8387\, I\left\{0 \leq u \leq 0.2950\right\} + 2.1219\, b_4(u)I\left\{0.2950 < u \leq 0.3348\right\} \\
&\quad + 0.1137\, I\left\{0.3348 < u \leq 0.9324\right\} + 2.1219\, b_4(u)I\left\{0.9324 < u \leq 1\right\}; \\
b_5^*(u) &= 1.8625\, b_5(u)I\left\{0 \leq u \leq 0.0452\right\} - 0.1476\, I\left\{0.0452 < u \leq 0.4936\right\} \\
&\quad + 1.8625\, b_5(u)I\left\{0.4936 < u \leq 0.5064\right\} + 0.1476\, I\left\{0.5064 < u \leq 0.9548\right\} \\
&\quad + 1.8625\, b_5(u)I\left\{0.9548 < u \leq 1\right\}.
\end{aligned}
$$

Verification PQD

d1 **Plackett** (ψ). By direct calculation from the distribution function as given on page 192 in Johnson (1987) it is easily verified that we have strictly PQD when $\psi > 1$, independence when $\psi = 1$ and strictly negative quadrant dependence when $\psi < 1$.

d2 **Cubic regression model** (β). It follows immediately from Example 1 (iv) in Lehmann (1966) that for $\beta > 0$ we get PQD and, similarly, for $\beta < 0$ negative quadrant dependence. Moreover, $cov(X, Y) \neq 0$ when $\beta \neq 0$ and hence we have strictly PQD for $\beta > 0$ and strictly negative quadrant dependence for $\beta < 0$. By definition of Y, we have independence when $\beta = 0$.

d3 **Bivariate stable** (a_3) with further the fixed values $a_1 = 2, a_2 = 0.2, b_1 = 1, b_2 = -1, b_3 = 0, h = 1$. This distribution is of the form $X = Z_1 + hZ_3, Y = Z_2 + hZ_3$ with Z_1, Z_2, Z_3

independent random variables. By Example 1 (ii) in Lehmann (1966) it follows that the bivariate stable distribution is PQD. When $h = 0$, obviously we have independence. Since $cov(X, Y) > 0$ when $h \neq 0$, we conclude that the bivariate stable distribution is strictly PQD for $h \neq 0$.

d4 **Bivariate stable** (h) with $a_1 = 1.5, a_2 = 0.1, a_3 = 0.9, b_1 = 1, b_2 = -1, b_3 = -0.5$. See d3.

d5 **Heteroscedastic linear regression** (p). As remarked before these distributions are not PQD.

Verification PFD

d6 **Bivariate normal** (ρ) with expectations equal to 0, variances equal to 1 and correlation coefficient ρ. For $\rho \geq 0$ the bivariate normal is a prominent example of distributions with a diagonal expansion, called PDE in Shaked (1979), page 69. On page 70 Shaked shows that each PDE is positive definite dependent (PDD), which in turn is equivalent to PFD, see Proposition 2.2 in Shaked (1979). Obviously, this concerns for $\rho > 0$ strictly PFD, since $cov(X, Y) > 0$. For $\rho = 0$ we have independence. For $\rho < 0$ we do not get strictly negative function dependence, but neither PFD (since $c_{11} < 0$) nor negative function dependence (since $c_{22} > 0$).

d7 **Khintchine** (ρ) is defined as $X = ZU, Y = ZV$, where (U, V) has a bivariate normal with expectations equal to 0, variances equal to 1 and correlation coefficient ρ and Z being independent of (U, V) is distributed as $\sqrt{\chi_3^2}$. It follows from Theorem 3.1 in Shaked (1979) with $(X_1, Y_1) = (U, V), (X_2, Y_2) = (Z, Z)$ and $h(x, y) = xy$ that Khintchine (ρ) with $\rho \geq 0$ is PFD. Furthermore, using $E\left(U^2 V^2\right) = 1 + 2\rho^2 \geq 1$, we get $cov(X^2, Y^2) = EZ^4 E\left(U^2 V^2\right) - \left(EZ^2\right)^2 \geq var(Z^2) > 0$ and hence we have strictly PFD.

d8 **Pearson type VII** (m) is defined as $X = U/\sqrt{Z}, Y = V/\sqrt{Z}$, where U, V and Z are independent random variables, U and V with a standard normal distribution and Z having a Gamma distribution with parameters $(m - 1, \frac{1}{2})$. Hence, by Proposition 2.1 in Shaked (1979) it is seen that this distribution is positively dependent by mixture, which implies PFD, see page 70 in Shaked (1979). Let $t > 0$ be such that $E\left(\frac{1}{\sqrt{Z}}\right)^{2t} < \infty$. Then we get $cov(|X|^t, |Y|^t) = var\left(\left(\frac{1}{\sqrt{Z}}\right)^t\right) E|U|^t E|V|^t > 0$ and hence we have strictly PFD.

d9 **Sub-Gaussian** (α) is defined as $X = U\sqrt{Z}, Y = V\sqrt{Z}$, where U, V and Z are independent, U and V have a standard normal distribution and Z has a stable distribution, totally skewed to the right, with index of stability $\alpha/2$. By Proposition 2.1 in Shaked (1979) it is seen that this distribution is positively dependent by mixture, which implies PFD, see page 70 in Shaked (1979). Let $t > 0$ be such that $E\left(\sqrt{Z}\right)^{2t} < \infty$. Then we get $cov(|X|^t, |Y|^t) = var\left(\left(\sqrt{Z}\right)^t\right) E|U|^t E|V|^t > 0$ and hence we have strictly PFD.

Asymptotic null distribution T^+ and T^o

As defined in (6), we have

$$(C_1, ..., C_k) =$$
$$\frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} b_1 \left(\frac{R_i - \frac{1}{2}}{n} \right) b_1 \left(\frac{S_i - \frac{1}{2}}{n} \right), ..., \sum_{i=1}^{n} b_k \left(\frac{R_i - \frac{1}{2}}{n} \right) b_k \left(\frac{S_i - \frac{1}{2}}{n} \right) \right).$$

The derivations of the asymptotic null distributions of T^+ and T^o go along the same lines of argument. As might be expected the selection rule concentrates on dimension 1 under H_0. Since T^+ and T^o are the same when restricted to dimension 1, we get the same asymptotic result for T^+ and T^o. Furthermore, to prove that the selection rule concentrates on dimension 1, it is enough to know that they are dominated by $\sum_{j=1}^{k} C_j^2$. To show that this is indeed essential and doing so making the proof more transparent, we therefore derive at once the asymptotic null distribution for the whole class of tests statistics with these features, defined as

$$\check{T}_{\check{S}} \text{ with } 0 \leq \check{T}_k \leq \sum_{j=1}^{k} C_j^2 \text{ and } \check{T}_1 = C_1^2 I\left(C_1 > 0\right),$$

$$\text{where } \check{S} = \min\left\{ k : 1 \leq k \leq d(n) : \check{T}_k - k \log n \geq \check{T}_r - r \log n, 1 \leq r \leq d(n) \right\}.$$

Note that T^+ and T^o belong to this class.

Theorem 1 *Assume that H_0 is true and denote the probability measure under H_0 by P_0. If $d(n) = o\left(\{n/\log n\}^{\frac{1}{10}}\right)$, then for every test statistic*

$$\check{T}_{\check{S}} \text{ with } 0 \leq \check{T}_k \leq \sum_{j=1}^{k} C_j^2 \text{ and } \check{T}_1 = C_1^2 I\left(C_1 > 0\right),$$

$$\text{where } \check{S} = \min\left\{ k : 1 \leq k \leq d(n) : \check{T}_k - k \log n \geq \check{T}_r - r \log n, 1 \leq r \leq d(n) \right\},$$

we have

$$\lim_{n \to \infty} P_0(\check{S} = 1) = 1 \text{ and } \lim_{n \to \infty} P_0\left(\check{T}_{\check{S}} \leq x\right) = \begin{cases} \Phi\left(\sqrt{x}\right) & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}, \tag{11}$$

where Φ denotes the standard normal distribution function. Note that $\lim_{n \to \infty} P_0\left(\check{T}_{\check{S}} = 0\right) = \frac{1}{2}$.

Proof. The proof of Theorem 1 follows from the proof of Theorem 1 in Kallenberg and Ledwina (1999). By definition of $\check{S}$ and the assumption $0 \leq \check{T}_k \leq \sum_{j=1}^{k} C_j^2$ we get

$$P_0(\check{S} \geq 2) = \sum_{k=2}^{d(n)} P_0(\check{S} = k) \leq \sum_{k=2}^{d(n)} P_0\left(\check{T}_k \geq (k-1)\log n\right)$$
$$\leq \sum_{k=2}^{d(n)} P_0\left(\sum_{j=1}^{k} C_j^2 \geq (k-1)\log n\right).$$

Application of (A.1) and (A.2) on page 297 in Kallenberg and Ledwina (1999) gives

$$\lim_{n \to \infty} \sum_{k=2}^{d(n)} P_0\left(\sum_{j=1}^{k} C_j^2 \geq (k-1)\log n\right) = 0,$$

implying

$$\lim_{n\to\infty} P_0(\check{S} = 1) = 1. \tag{12}$$

Using

$$
\begin{aligned}
P_0\left(\check{T}_{\check{S}} \leq x\right) &= P_0\left(\check{T}_1 \leq x, \check{S} = 1\right) + P_0\left(\check{T}_{\check{S}} \leq x, \check{S} \geq 2\right) \\
&= P_0\left(\check{T}_1 \leq x\right) - P_0\left(\check{T}_1 \leq x, \check{S} \geq 2\right) + P_0\left(\check{T}_{\check{S}} \leq x, \check{S} \geq 2\right),
\end{aligned}
$$

the limiting distribution of $\check{T}_{\check{S}}$ under H_0 immediately follows from (12), the assumption $\check{T}_1 = C_1^2 I\left(C_1 > 0\right)$ and the asymptotic standard normality of C_1 under H_0. ∎

Consistency of T^o

The consistency of T^o is presented in the following theorem.

Theorem 2 *Let $\mathbb{P}$ be an alternative and let F and G be the marginal distribution functions of X and Y, respectively under $\mathbb{P}$. Let*

$$c_{jj} = E_{\mathbb{P}} b_j\left(F\left(X\right)\right) b_j\left(G\left(Y\right)\right) > 0 \tag{13}$$

for some j. If $d(n)$ tends to infinity and $d(n) = o\left(\{n/\log n\}^{\frac{1}{10}}\right)$, then T^o is consistent against $\mathbb{P}$.

Proof. By (A.11) and (A.12) on page 299 in Kallenberg and Ledwina (1999) we have for any r, s,

$$\frac{1}{n}\sum_{i=1}^{n} b_r\left(\frac{R_i - \frac{1}{2}}{n}\right) b_s\left(\frac{S_i - \frac{1}{2}}{n}\right) \xrightarrow{\mathbb{P}} E_{\mathbb{P}} b_r\left(F\left(X\right)\right) b_s\left(G\left(Y\right)\right) \tag{14}$$

as $n \to \infty$. Let J be the smallest $j \geq 1$ for which (13) holds. Let C_j be defined as in (6). In view of (14) we get for any $j \in \{1, ..., J-1\}$ that $I\left(C_j > 0\right) \xrightarrow{\mathbb{P}} 0$ if $E_{\mathbb{P}} b_j\left(F\left(X\right)\right) b_j\left(G\left(Y\right)\right) < 0$ and $n^{-1} C_j^2 \xrightarrow{\mathbb{P}} 0$ if $E_{\mathbb{P}} b_j\left(F\left(X\right)\right) b_j\left(G\left(Y\right)\right) = 0$. Hence, $n^{-1} T_j^o \xrightarrow{\mathbb{P}} 0$ for any $j \in \{1, ..., J-1\}$. Moreover, $I\left(C_J > 0\right) \xrightarrow{\mathbb{P}} 1, n^{-1} C_J^2 \xrightarrow{\mathbb{P}} \{E_{\mathbb{P}} b_J\left(F\left(X\right)\right) b_J\left(G\left(Y\right)\right)\}^2 > 0$, implying $n^{-1} T_J^o \xrightarrow{\mathbb{P}} \{E_{\mathbb{P}} b_J\left(F\left(X\right)\right) b_J\left(G\left(Y\right)\right)\}^2 > 0$. Because $d(n) \to \infty$, we have $d(n) \geq J$ for all sufficiently large n. For any $j \in \{1, ..., J-1\}$, the foregoing implies

$$n^{-1} T_j^o - jn^{-1}\log n < n^{-1} T_J^o - Jn^{-1}\log n$$

for all sufficiently large n and hence, $\mathbb{P}\left(S2^o = j\right) \xrightarrow{\mathbb{P}} 0$ for any $j \in \{1, ..., J-1\}$ and therefore, $\mathbb{P}\left(S2^o \geq J\right) \xrightarrow{\mathbb{P}} 1$.

For any $j \geq J$, it follows from (7) that $T_j^o \geq T_J^o$. Noting that $T_J^o \xrightarrow{\mathbb{P}} \infty$ and $\mathbb{P}\left(S2^o \geq J\right) \xrightarrow{\mathbb{P}} 1$ we also get $T^o \xrightarrow{\mathbb{P}} \infty$. By Theorem 1 the proof of Theorem 2 is complete. ∎

The class of positive dependent alternatives for which consistency of T^o holds is rather large. In particular, for any strictly PQD alternative $\mathbb{P}$ the linear correlation is strictly positive, cf. Remark 2.4, and hence, by Theorem 2, T^o is consistent against any strictly PQD alternative. For any strictly PFD alternative $\mathbb{P}$ we have $cov_{\mathbb{P}}(f(F(X)), f(G(Y))) \geq 0$ for all f with inequality for at least some f, implying $E_{\mathbb{P}} b_j\left(F\left(X\right)\right) b_j\left(G\left(Y\right)\right) \geq 0$ for all j. As a rule, for a strictly PFD alternative we will have inequality for some j and hence consistency of T^o.

Consistency of T^+

The consistency of T^+ is presented in the following theorem. Noting that $T^+ \geq T_1^+$ allows to pattern the proof of Theorem 2. We omit the details.

Theorem 3 *Let $\mathbb{P}$ be an alternative and let F and G be the marginal distribution functions of X and Y, respectively under $\mathbb{P}$. Let $c_{11} = E_{\mathbb{P}} b_1 (F(X)) b_1 (G(Y)) > 0$. If $d(n)$ tends to infinity and $d(n) = o\left(\{n/\log n\}^{\frac{1}{10}}\right)$, then T^+ is consistent against $\mathbb{P}$.*

The test based on T^+ is consistent against any alternative $\mathbb{P}$ for which the linear correlation $E_{\mathbb{P}} b_1 (F(X)) b_1 (G(Y)) > 0$ and hence in particular for any strictly PQD alternative, cf. Remark 2.4. That T^+ is not consistent at strictly PFD alternatives with linear correlation equal to 0 was already discussed when introducing T^+ in Section 3.

Approximation for critical values and p-values

The first order approximations of T^o and T^+ given in (11) is not very accurate. For instance the simulated critical values for $n = 50$, $d(n) \geq 4$ and $\alpha = 0.05$ equal 3.24 and 4.21, respectively, while the approximation based on Theorem 1 yields 2.71. The same phenomenon occurs in data driven tests for other testing problems like testing goodness of fit. The remedy is a second order approximation following the line of argument given by Kallenberg and Ledwina (1995, Section 4).

The idea is that $T^o \geq T_1^o = C_1^2 1 (C_1 > 0)$ and $T^+ \geq T_1^+ = C_1^2 1 (C_1 > 0)$, implying that $P(T^o \leq x)$ and $P(T^+ \leq x)$ are underestimated by the first order approximation $\Phi(\sqrt{x})$. We apply the following approximations (we denote by $\doteq$ approximately equals to)

$$
\begin{aligned}
P(T^o \leq x) &\doteq P(T_1^o \leq x, S2 = 1) + P(T_2^o \leq x, S2 = 2) \\
&\doteq P(T_1^o \leq x) - P(T_1^o \leq x \leq T_2^o, S2 = 2) \\
&\doteq P(C_1 \leq \sqrt{x}) \\
&\quad - P\left((C_1 I\{C_1 > 0\})^2 \leq x \leq (C_1 I\{C_1 > 0\})^2 + (C_2 I\{C_2 > 0\})^2, (C_2 I\{C_2 > 0\})^2 > \log n\right).
\end{aligned}
$$

Replacing C_1 and C_2 by (independent) $N(0,1)$-distributed random variables, a further approximation leads to the final proposal

$$
P(T^o \leq x) \doteq \begin{cases} \Phi(\sqrt{x}) \Phi(\sqrt{\log n}) & \text{if } x \leq \log n \\ \Phi(\sqrt{x}) - \frac{1}{2} \Phi(-\sqrt{x}) - \dfrac{e^{-x/2}}{4} & \text{if } x \geq 2 \log n \\ \text{linearize} & \text{if } \log n < x < 2 \log n. \end{cases} \tag{15}
$$

Inserting the simulated critical value 4.21 for $n = 50$ and $\alpha = 0.05$ in (15) gives 0.956, which is rather close to 0.95 and a big improvement compared to the approximation based on Theorem 1, yielding $\Phi(\sqrt{4.21}) = 0.980$.

Similarly we get

$$
P(T^+ \leq x) \doteq \begin{cases} \Phi(\sqrt{x}) \Phi(\sqrt{\log n}) + \frac{1}{2} \Phi(-\sqrt{\log n}) & \text{if } x \leq \log n \\ \Phi(\sqrt{x}) - \dfrac{e^{-x/2}}{4} & \text{if } x \geq 2 \log n \\ \text{linearize} & \text{if } \log n < x < 2 \log n. \end{cases} \tag{16}
$$

Inserting the simulated critical value 3.24 for $n = 50$ and $\alpha = 0.05$ in (16) gives 0.953, which is rather close to 0.95 and an improvement compared to the approximation based on Theorem 1, yielding $\Phi(\sqrt{3.24}) = 0.964$.

References

Albers, W. (1999). Stop-loss premiums under dependence. *Insurance Math. Econom.* **24** 173-185.

Albers, W., Kallenberg, W.C.M. and Martini, F. (2001). Data driven rank tests for classes of tail alternatives. *J. Amer. Statist. Assoc.* **96** 685-696.

Behnen, K. and Neuhaus, G. (1989). *Rank Tests with Estimated Scores and Their Application.* Teubner, Stuttgart.

Breiman (1968). *Probability*, Addison-Wesley, Reading, Massachusetts.

Bryson, M.C. and Johnson, M.E. (1982). Constructing and simulating multivariate distributions using Khintchine's theorem. *J. Statist. Comput. Simul.* **16** 129-137.

Chambers, J.M., Mallows, C.L. and Stuck, B.W. (1976). A method for simulating stable random variables. *J. Amer. Statist. Assoc.* **71** 340-344.

Denuit, M. , Dhaene, J. and Ribas, C. (2001). Does positive dependence between individual risks increase stop-loss premiums? *Insurance Math. Econom.* **28** 305-308.

Embrechts, P., McNeil, A. and Straumann, D. (2002). Correlation and dependency in risk management: properties and pitfalls. In: *Risk Management: Value at Risk and Beyond* (Eds. Dempster, M. and Moffatt, H.). Cambridge University Press, Cambridge, pp. 176-223.

Esary, J.D., Proschan, F. and Walkup, D.W. (1967). Association of random variables, with applications. *Ann. Math. Statist.* **38** 1466-1474.

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

Johnson, M.E. (1987) *Multivariate Statistical Simulation.* Wiley, New York.

Kallenberg, W.C.M. and Ledwina, T. (1995). On data driven Neyman's tests. *Probab. Math. Statist.* **15** 409-426.

Kallenberg, W.C.M. and Ledwina, T. (1997). Data-driven smooth tests when the hypothesis is composite. *J. Amer. Statist. Assoc.* **92** 1094-1104.

Kallenberg, W.C.M. and Ledwina, T. (1999). Data-driven rank tests for independence. *J. Amer. Statist. Assoc.* **94** 285-301.

Kowalczyk,T. and Pleszczyńska, E. (1977). Monotonic dependence functions of bivariate distributions. *Ann. Statist.* **5** 1221-1227.

Ledwina, T. (1986). Large deviations and Bahadur slopes of some rank tests of independence. *Sankhyā* A **48** 188-207.

Ledwina, T. (1994). Data driven version of Neyman's smooth test of fit. *J. Amer. Statist. Assoc.* **89** 1000-1005.

Lehmann, E. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37** 1137-1153.

Mari, D.D. and Kotz, S. (2001). *Correlation and Dependence.* Imperial College Press, London.

Metry, M.H. and Sampson, A.R. (1991). A family of partial orderings for positive dependence among fixed marginal bivariate distributions. In: *Advances in Probability Distributions with Given Marginals : Beyond the Copulas* (Eds. G. Dall' Aglio, S. Kotz and G. Salinetti). Kluwer, Dordrecht, Netherlands, pp. 129-138.

Rychlik, T. (2001) *Projecting Statistical Functionals.* Lecture Notes in Statistics 160, Springer, New York.

Samorodnitsky, G. and Taqqu, M.S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models With Infinite Variance*, New York, Chapman and Hall.

Scarsini, M. and Shaked, M. (1996). Positive dependence orders: a survey. In: *Athens Conference on Applied Probability and Time Series I: Applied Probability* (Eds. C.C. Heyde, Yu V. Prohorov, R. Pyke and S.T. Rachev). Springer, New York, pp. 70-91.

Schriever, B.F. (1987). An ordering of positive dependence. *Ann. Statist.* **15** 1208-1214.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461-464.

Shaked, M. (1979). Some concepts of positive dependence for bivariate interchangeable distributions. *Ann. Inst. Statist. Math.* **31** 67-84.

Thas, O. (2001). Nonparametrical tests based on sample space partitions. PhD Dissertation, University of Gent.