Faculty of Mathematical Sciences

University of Twente

University for Technical and Social Sciences

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

Memorandum No. 1504

Data driven rank tests for classes of tail alternatives

W. Albers, W.C.M. Kallenberg and F. Martini

October 1999

ISSN 0169-2690

Data Driven Rank Tests for Classes of Tail Alternatives

Willem Albers, Wilbert C.M. Kallenberg and Felix Martini
Faculty of Mathematical Sciences
University of Twente
P.O. Box 217, 7500 AE Enschede
The Netherlands

Abstract Tail alternatives describe the frequent occurrence of a non-constant shift in the two-sample problem with a shift function increasing in the tail. The classes of shift functions can be built up using Legendre polynomials. It is important to rightly choose the number of involved polynomials. Here this choice is based on the data, using a modification of Schwarz' selection rule. Given the data driven choice of the model, appropriate rank tests are applied. Simulations show that the new data driven rank tests work very well. While other tests for detecting shift alternatives as Wilcoxon's test may completely break down for important classes of tail alternatives, the new tests have high and stable power. The new tests have also higher power than data driven rank tests for the unconstrained two-sample problem. Theoretical support is obtained by proving consistency of the new tests against very large classes of alternatives, including all common tail alternatives. A simple but accurate approximation of the null distribution makes application of the new tests easy.

Keywords and phrases: shift function, model selection, Monte Carlo study, consistency, Legendre polynomials.

1991 Mathematics Subject Classification: 62 G 10, 62 E 25, 62 G 20

1 Introduction

In the standard two-sample problem a (possible) constant shift in location is the point of interest. For many applications, however, a constant shift is far from realistic. Medical examples can be found for instance in Fleming et al. (1980), but a non-constant shift appears also in many other areas as economics, finance etc. Often the shift increases when we come further in the tail of the distribution. We speak in that case of *tail alternatives*. Common tests for a constant shift like Wilcoxon's two-sample test do not detect adequately such tail alternatives. On the other hand, the tail alternatives have a special feature and hence application of tests in the two-sample problem which are developed to detect *all* kind of alternatives, may be inappropriate either.

The notion of generalized shift alternatives was introduced by Neuhaus (1987, 1988; cf. also Behnen and Neuhaus 1989). He distinguishes between shifting only the "upper part" of the distribution, only the "central part" and only the "lower part". Recently, Hájek, Šidák and Sen (1999, page 352) refer to the pioneering approach to the two-sample problem by Behnen and Neuhaus (1989) by stating: "e.g. it is plausible that the extreme parts of the population react in quite another way than the central part". In the one sample case, when testing goodness of fit with a simple hypothesis, Mason and Schuenemeyer (1983, 1992) consider deviations from the hypothesized distribution occurring in the tails and speak of "heavy tail alternatives" and "light tail alternatives". The one-sided form of Mason and Schuenemeyer's heavy tail alternatives belongs to the class of so called *late alternatives* in Albers and Schut (1996). These alternatives are completely concentrated in the right-tail of the distribution. Albers and Schut also introduce *early tail alternatives*, based on survival functions studied by Harrington and Fleming (1982) and related to score functions for censored rank tests (see Albers and Akritas 1987). Early tail alternatives contain a constant part throughout and have a slow increase going into the tail. Between these two extremes they consider the class of *increasing tail alternatives*.

It is the aim of this paper to present new tests for the challenging two-sample problem of detecting shift alternatives with an increasing shift when we go into the tail of the distribution. More precisely, let $X_1, \dots, X_m, Y_1, \dots, Y_n$ be independent random variables (rv's), the $X_i, i = 1, \dots, m$, having distribution function (df) F and the $Y_j, j = 1, \dots, n$ having df G. The standard two-sample problem concerns a constant shift, $Y = X + \theta$, where we test $H_0 : \theta = 0$ against e.g. $H_1 : \theta > 0$. To admit tail alternatives, we consider the more general model, where $Y = X + \theta B(F(X))$ with $B \geq 0$ and nondecreasing on (0,1). The larger X, the larger $F(X)$, and, since B is nondecreasing, the larger is the shift $\theta B(F(X))$, thus getting a technical definition of what we call *tail alternatives* in this paper.

In finding an appropriate test for the testing problem in the setting of tail alternatives, there are two problems: B is unknown and F is unknown. The latter leads to rank tests. To get rid of the unknown B, we take the orthonormal system of Legendre polynomials. The contribution of the first component of this system to the test statistic, turns out to correspond to Wilcoxon's test, thus taking into account especially the *constant* term in the shift. The second component is essentially Mood's test statistic, being a rank test of *scale*, etc, cf. Eubank et al. (1987, page 821). In the one sample problem, Hušková and Sen (1985, 1986) have used the orthonormal system of Legendre polynomials to estimate the unknown score generating function, see also Hájek, Šidák and Sen (1999, page 348).

A major problem in this context is the number of components which one should take into account. Recent research (cf. Bickel and Ritov 1992; Bogdan 1995; Bogdan and Ledwina 1996; Eubank 1997; Eubank and LaRiccia 1992; Fan 1996; Inglot et al. 1997, 1998; Inglot and Ledwina 1996; Janic-Wróblewska and Ledwina 1999; Kallenberg and Ledwina 1995a,b, 1997a,b, 1999; Kallenberg et al. 1997; Ledwina 1994) strongly indicates that a deterministic choice, even when it is sequential as in Hušková and Sen (1985, 1986), does not give a satisfactory solution. We follow the line of argument as described in Janic-Wróblewska and Ledwina (1999). Firstly, the testing problem is reparametrized as e.g. in Neuhaus (1987). Then we model alternatives in terms of exponential families with growing dimension, thus covering more and more the whole space of alternatives. Here the Legendre polynomials come in. The suitable dimension is determined by the data, using a modification of Schwarz' (1978) selection rule.

A new element, compared to the two-sample problem treated in Janic-Wróblewska and Ledwina (1999), is of course, that we deal with *tail alternatives.* Hence, the testing problems in the exponential families are restricted ones. Likelihood ratio tests for such testing problems can be defined, but they are difficult to implement in practice, even in the limit, where we have multivariate normality, cf. e.g. Follmann (1996, page 854). One of the attractive points of most of the data-driven tests is their simplicity in application. Therefore, we do not take the complicated likelihood ratio tests, but adjust Follmann's test to our situation. We repair also a slight ambiguity in Follmann's test. These lines of argument lead to two new tests in the two-sample problem with tail alternatives. The new tests are introduced and motivated in more detail in Section 2.

For several types of tail alternatives a simulation study has been performed. It turns out that we benefit from pointing our new tests on tail alternatives: substantial power gain is achieved compared to the data driven test for the unrestricted two-sample problem, proposed by Janic-Wróblewska and Ledwina (1999). When the constant shift part in a tail alternative is not dominant, Wilcoxon's test may break down completely. The new tests have high and stable power, resulting in comparison to Wilcoxon's test in a relatively small power loss when the constant shift part is dominant and a large power gain otherwise. The simulation results are presented in Section 3.

The asymptotic null distribution is derived in Section 4. As often with this kind of data driven test statistics, the asymptotic null distribution itself does not give accurate approximations. Fortunately, a slightly more delicate approach leads to a very simple, but also accurate approximation for critical values and/or p-values. This makes the new tests easily applicable in practice. Furthermore, the observed nice empirical performance of the new tests is supported and explained by proving consistency against very broad classes of alternatives, including all alternatives with Y stochastically larger than X and hence in particular all tail alternatives.

2 Test statistics

We start with a reformulation of the testing problem along the lines of Neuhaus (1987), cf. also Behnen and Neuhaus (1989, sec. 1.3). Recall that $X_1, \ldots, X_m$, $Y_1, \ldots, Y_n$ are independent rv's, the $X_i, i = 1, \ldots, m$, having df F and the $Y_j, j = 1, \ldots, n$, having df G. Let $N = m + n$ and $H(x) = \frac{m}{N}F(x) + \frac{n}{N}G(x)$. Then we may write

$$\begin{aligned} F(x) &= H(x) + \frac{n}{N}\{F(x) - G(x)\} \\ G(x) &= H(x) - \frac{m}{N}\{F(x) - G(x)\}. \end{aligned} \tag{1}$$

The function H may be seen as the nuisance parameter and $F - G$ as the parameter of interest.

Assume that F is differentiable with derivative f and H is differentiable with derivative h. Writing for short $D(x) = B(F(x))$, we get in our case for small θ (we denote by $\doteq$ approximately equals to)

$$G(x) = P\left(X + \theta D(X) \le x\right) \doteq F(x - \theta D(x)) \doteq F(x) - \theta D(x) f(x). \tag{2}$$

Inserting (2) in (1), we arrive at

$$\begin{aligned} F(x) &\doteq H(x) + \frac{n}{N}\theta D(x) f(x) \\ G(x) &\doteq H(x) - \frac{m}{N}\theta D(x) f(x). \end{aligned}$$

Therefore, writing $\Psi = -f(F^{-1})$, the df of $H(X)$ at u when X has df F is up to first order equal to

$$\begin{aligned} &u + \frac{n}{N}\theta D\left(H^{-1}(u)\right) f\left(H^{-1}(u)\right) \doteq u + \frac{n}{N}\theta D\left(F^{-1}(u)\right) f\left(F^{-1}(u)\right) \\ &= u - \frac{n}{N}\theta B(u)\Psi(u). \end{aligned}$$

The corresponding density equals

$$1 - \frac{n}{N}\theta J(u), \text{ where } J = (B\Psi)'.$$

Similarly, the density of $H(Y)$ at v when Y has df G is up to first order equal to

$$1 + \frac{m}{N}\theta J(v).$$

A slightly more convenient way is to write these densities in the form of an (first order equivalent) exponential family, yielding, up to normalizing constants,

$$\exp\left\{-\frac{n}{N}\theta J(u)\right\} \text{ for } H(X) \text{ and } \exp\left\{\frac{m}{N}\theta J(v)\right\} \text{ for } H(Y).$$

2.1 The k-dimensional model

To cover broad classes of alternatives we do not take one fixed direction J, but k directions $J_1, \ldots, J_k$ and extend the exponential family accordingly to, again up to normalizing constants,

$$\exp\left\{-\frac{n}{N}\sum_{r=1}^{k}\theta_r J_r(u)\right\} \text{ for } H(X) \text{ and } \exp\left\{\frac{m}{N}\sum_{r=1}^{k}\theta_r J_r(v)\right\} \text{ for } H(Y).\quad(3)$$

The larger k, the broader the class of alternatives under consideration. The role of θJ is now replaced by $\sum_{r=1}^{k}\theta_r J_r$ and, since normalizing is not important, we simply define the function B_k in dimension k by

$$(B_k\Psi)' = \sum_{r=1}^{k}\theta_r J_r. \quad (4)$$

So, instead of $Y = X + \theta B(F(X))$, we now have $Y = X + B_k(F(X))$ with the shift function B_k given by (4), where the θ's now are absorbed in the function B_k.

To get a convenient class of alternatives, we take for the J_r's the orthonormal system of Legendre polynomials on (0,1). The first four Legendre polynomials are given by

$$J_1(x) = \sqrt{3}(2x-1),$$

$$J_2(x) = \sqrt{5}(6x^2 - 6x + 1),$$

$$J_3(x) = \sqrt{7}(20x^3 - 30x^2 + 12x - 1),$$

$$J_4(x) = 3(70x^4 - 140x^3 + 90x^2 - 20x + 1).$$

As in Albers and Schut (1996) we take for Ψ the one related to the Wilcoxon case:

$$\Psi(t) = -\sqrt{3}t(1-t),$$

according to the logistic distribution with variance $\pi^2/9 = 1.097$, having df $F(x) = 1/(1+e^{-x\sqrt{3}})$. Noting that $\psi = \Psi' = J_1$, (4) implies

$$B_k(t) = \frac{\sum_{r=1}^{k} \theta_r \int_0^t J_r(u)du}{\int_0^t J_1(u)du}.$$

Since the numerator is 0 at $t = 0$ and $t = 1$, it contains a factor $t(1-t)$ and therefore, J_k is a polynomial of degree $k-1$. More precisely, using (10') on page 176 of Sansone (1959) we get

$$\frac{r(r+1)\int_0^t J_r(u)du}{t^2 - t} = J_r'(t)$$

and hence

$$B_k(t) = \sum_{r=1}^{k} \theta_r \frac{J_r'(t)}{\sqrt{3}r(r+1)}. \qquad (5)$$

In terms of the model given by (3) the null hypothesis $H_0 : F = G$ reads as $\theta_1 = 0, \ldots, \theta_r = 0$. The alternatives are characterized by the fact that $B_k \geq 0$ and that B_k is nondecreasing on (0,1). These requirements are not easily expressed in simple statements about $\theta_1, \ldots, \theta_k$. Therefore, we replace these characterization by the relaxation

$$B_k(0) \geq 0, B_k'(0) \geq 0 \text{ and } B_k'(1) \geq 0. \qquad (6)$$

As B_k is a polynomial of degree $k-1$, the conditions given by (6) are for the special cases $k = 1, 2, 3$ equivalent to "$B_k \geq 0$ and B_k nondecreasing on (0,1)".

By (11) on page 251 and (6_3) on page 173 of Sansone (1959) we get

$$J_r'(0) = (-1)^{r+1}\sqrt{2r+1}r(r+1), \; J_r''(1) = \tfrac{1}{2}\sqrt{2r+1}(r+2)(r+1)r(r-1)$$

and

$$J_r''(0) = (-1)^r J_r''(1).$$

Hence, the requirements $B_k(0) \geq 0, B_k'(0) \geq 0$ and $B_k'(1) \geq 0$ read as

$$\sum_{r=1}^{k} \theta_r (-1)^{r+1} \frac{\sqrt{2r+1}}{\sqrt{3}} \geq 0, \ \sum_{r=2}^{k} \theta_r (-1)^{r} \frac{\sqrt{2r+1}(r+2)(r-1)}{2\sqrt{3}} \geq 0, \ \text{and}$$
$$\sum_{r=2}^{k} \theta_r \frac{\sqrt{2r+1}(r+2)(r-1)}{2\sqrt{3}} \geq 0. \tag{7}$$

So, in the k-dimensional model we want to test $H_0 : \theta_1 = 0, \ldots, \theta_k = 0$ against H_1 given by (7).

2.2 Test statistic in the k-dimensional model

As a consequence of (3), the simultaneous density of $H(X_1), \ldots, H(X_m)$, $H(Y_1), \ldots, H(Y_n)$ at $u_1, \ldots, u_m, \ v_1, \ldots, v_n$ is, up to normalizing constants,

$$\exp \left[\sum_{r=1}^{k} \theta_r \left\{ \frac{m}{N} \sum_{j=1}^{n} J_r(v_j) - \frac{n}{N} \sum_{i=1}^{m} J_r(u_i) \right\} \right].$$

Suppose for the moment that the nuisance parameter H is known. Then we should base the test statistic on the (sufficient) statistic

$$\left(\frac{m}{N} \sum_{j=1}^{n} J_1\left(H(Y_j)\right) - \frac{n}{N} \sum_{i=1}^{m} J_1(H(X_i)), \ldots, \frac{m}{N} \sum_{j=1}^{n} J_k(H(Y_j)) - \frac{n}{N} \sum_{i=1}^{m} J_k(H(X_i)) \right). \tag{8}$$

Since

$$H = \frac{m}{N} F + \frac{n}{N} G$$

is unknown, it is estimated by

$$\hat{H} = \frac{m}{N} \hat{F} + \frac{n}{N} \hat{G},$$

where $\hat{F}$ and $\hat{G}$ are the empirical df's, based on $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$, respectively. Hence,

$$\hat{H}(X_i) = \frac{R_i}{N} \text{ and } \hat{H}(Y_j) = \frac{R_{m+j}}{N}$$

with R_i the rank of X_i and R_{m+j} the rank of Y_j in the pooled sample. Inserting the estimators $\hat{H}(X_i)$ and $\hat{H}(Y_j)$ in (8) and applying the familiar correction for continuity we arrive at the statistics (after some rescaling)

$$
\begin{aligned}
&(Z_1, \ldots, Z_k) \\
&= \sqrt{\frac{N}{nm}} \left(\frac{m}{N} \sum_{i=m+1}^{N} J_1\left(\frac{R_i - \frac{1}{2}}{N}\right) - \frac{n}{N} \sum_{i=1}^{m} J_1\left(\frac{R_i - \frac{1}{2}}{N}\right), \right. \qquad (9) \\
&\qquad \left. \ldots, \frac{m}{N} \sum_{i=m+1}^{N} J_k\left(\frac{R_i - \frac{1}{2}}{N}\right) - \frac{n}{N} \sum_{i=1}^{m} J_k\left(\frac{R_i - \frac{1}{2}}{N}\right) \right).
\end{aligned}
$$

It is not hard to show that $(Z_1, \ldots, Z_k)$ is asymptotically normal with under H_0 expectation $(0, \ldots, 0)$ and under local alternatives of the form (3), with the θ's of order $O(N^{-1/2})$, expectation $\sqrt{\frac{nm}{N}}(\theta_1, \ldots, \theta_k)$, while the covariance matrix is the identity. According to Fisher's tactic to reduce a complicated inference problem to a simple form (cf. Efron 1998, page 97), but still preserving the main feature of the problem, we consider the asymptotic situation. That is, for a k-variate normal distribution with expectation $(\theta_1, \ldots, \theta_k)$ we have the testing problem that the multivariate normal mean is 0 against the alternative given by (7). Define

$$
W_1 = \sum_{r=1}^{k} Z_r(-1)^{r+1}\frac{\sqrt{2r+1}}{\sqrt{3}}, \; W_2 = \sum_{r=2}^{k} Z_r(-1)^{r}\frac{\sqrt{2r+1}(r+2)(r-1)}{2\sqrt{3}},
$$

$$
W_3 = \sum_{r=2}^{k} Z_r\frac{\sqrt{2r+1}(r+2)(r-1)}{2\sqrt{3}} \text{ and } W_r = Z_r \text{ for } r \geq 4.
$$

In terms of the W_r's, writing ϑ_r for the expectation of W_r, the testing problems reads as $H_0 : (\vartheta_1, \ldots, \vartheta_k) = (0, \ldots, 0)$ against $H_1 : \vartheta_1 \geq 0, \vartheta_2 \geq 0$, $\vartheta_3 \geq 0, (\vartheta_1, \ldots, \vartheta_k) \neq (0, \ldots, 0)$. Since the covariance matrix of the W_r's is no longer the identity, the likelihood ratio test for this testing problem is difficult to implement. In the similar situation that the expectation under the alternative hypothesis has all its components nonnegative, Follmann (1996) presents a simple test with good power properties.

Before applying Follmann's approach to our testing problem, a comment should be given on Follmann's test. If we replace W_r by $c_r W_r$ for some constants $c_r > 0$, the testing problem remains the same. However, Follmann's test statistic is not invariant under a scale transformation with different scale factors c_r. To avoid the ambiguity, some standardizing should be made. Therefore, we replace W_r by $W_r/\sqrt{\text{var}W_r}$, thus considering the correlation matrix instead of the covariance matrix. (Note that in the numerical results of Follmann (1996) indeed also correlation matrices are involved!) This leads to the Y_r's given by

$$Y_1 = \frac{\sum_{r=1}^{k} Z_r (-1)^{r+1} \frac{\sqrt{2r+1}}{\sqrt{3}}}{\sqrt{\frac{1}{3}k(k+2)}},$$

$$Y_2 = \frac{\sum_{r=2}^{k} Z_r (-1)^{r} \frac{\sqrt{2r+1}(r+2)(r-1)}{2\sqrt{3}}}{\sqrt{\frac{1}{720}k(k-1)(24k^3+189k^2+499k+454)}}, \tag{10}$$

$$Y_3 = \frac{\sum_{r=2}^{k} Z_r \frac{\sqrt{2r+1}(r+2)(r-1)}{2\sqrt{3}}}{\sqrt{\frac{1}{720}k(k-1)(24k^3+189k^2+499k+454)}} \text{ and } Y_r = Z_r \text{ for } r \geq 4.$$

Let $1(A)$ be the indicator function of the set A. Write ^{T} for the transpose. Denoting by $\mathbf{\Sigma}$ the covariance matrix of $\mathbf{Y} = (Y_1, \ldots, Y_k)^T$ (which is the same as the correlation matrix of $\mathbf{Y}$), the test statistic equals

$$T_k = \mathbf{Y}^T \mathbf{\Sigma}^{-1} \mathbf{Y}\; 1(Y_1 + Y_2 + Y_3 > 0). \tag{11}$$

We reject for large values of T_k. For $k = 1$, the indicator in T_1 should be read as $1(Y_1 > 0)$ and in case of T_2 we get $1(Y_1 + Y_2 > 0)$. (Note that for $k = 2$ the last two conditions of (7) coincide.)

Since $\mathbf{Y}^T \mathbf{\Sigma}^{-1} \mathbf{Y} = \mathbf{Z}^T \mathbf{Z}$, with $\mathbf{Z} = (Z_1, \ldots, Z_k)^T$, the test statistic in (11) can be simply written as

$$T_k = \mathbf{Z}^T \mathbf{Z}\, 1(Y_1 + Y_2 + Y_3 > 0) = \mathbf{Z}^T \mathbf{Z}\, 1\left(\sum_{r=1}^{k} c_r Z_r > 0 \right) \text{ with for } k \neq 2$$

$$c_r = \begin{cases} \sqrt{\dfrac{2r+1}{k(k+2)}} & \text{if } r \text{ is odd} \\ -\sqrt{\dfrac{2r+1}{k(k+2)}} + \dfrac{\sqrt{2r+1}(r+2)(r-1)}{\sqrt{\frac{1}{240}k(k-1)(24k^3+189k^2+499k+454)}} & \text{if } r \text{ is even} \end{cases} \tag{12}$$

and $c_1 = \sqrt{\frac{3}{8}}$, $c_2 = -\sqrt{\frac{5}{8}} + 1$ for $k = 2$.

2.3 Selection rule and data driven test statistics

The approach so far is along the same line as applying Neyman's smooth tests in goodness of fit testing problems, cf. e.g. Rayner and Best (1989). Recent research in this area has shown that the smooth tests behave very well, but that the right choice of the number k of components is extremely important. Since the right choice depends on the type of alternative, which is of course unknown, a deterministic good choice of k is only possible if the main interest is in a very particular type of alternatives. A solution to this problem is to make a choice depending on the data. In a series of papers (see Bogdan 1995; Bogdan and Ledwina 1996; Inglot et al. 1997, 1998; Inglot and Ledwina 1996; Janic-Wróblewska and Ledwina 1999; Kallenberg and Ledwina 1995a,b, 1997a,b, 1999; Kallenberg et al. 1997; Ledwina 1994) the data driven procedure based on (modifications of) Schwarz' selection rule has been shown to be very successful. The idea is that a higher dimensional and hence more complex model should be penalized. This idea is applied here as well.

The selection rule in the present situation reads as

$$S = \min\{k : 1 \le k \le d(N) : T_k - k \log N \ge T_r - r \log N, 1 \le r \le d(N)\} \,. \tag{13}$$

The number $d(N)$ is the largest dimension of the exponential family models of type (3), which are under consideration when we have N observations. It should be noted that the data driven test procedure is stable for large $d(N)$ and that the problem of choosing the number of components k is certainly not replaced by the choice of $d(N)$, cf. e.g. Kallenberg and Ledwina (1997b, sec. 6.2).

The data driven test statistic for our testing problem is given by T_S and the null hypothesis is rejected for large values of T_S.

Apart from T_k and T_S we consider also more simple test statistics $T1_k$ and $T1_{S1}$, given by

$$\begin{aligned} T1_k &= \mathbf{Z}^T\mathbf{Z}\, 1\left(\sum_{r=1}^{k} Z_r > 0\right) \quad \text{and} \\ S1 &= \min\{k : 1 \le k \le d(N) : T1_k - k \log N \ge T1_r - r \log N, 1 \le r \le d(N)\} \,, \end{aligned} \tag{14}$$

where we reject the null hypothesis for large values of $T1_{S1}$. Since $\mathbf{Z}^T\mathbf{Z}$ is the squared Wilcoxon statistic for $k = 1$, the squared Mood statistic for $k = 2$, $T1_k$ starts for $k = 1$ with investigating whether there is a positive shift, then the scale is coming in, etc. It is seen from (12) and (14) that T_k and $T1_k$ only differ in the weights of the Z_r's in the indicator function. For instance, in case of T_2 the weights are $\sqrt{\frac{3}{8}} = 0.61$ for Z_1 and $-\sqrt{\frac{5}{8}} + 1 = 0.21$ for Z_2. The higher weight for Z_1 indicates that T_S is slightly closer to Wilcoxon's test than $T1_{S1}$. This is seen in the simulation, when considering the lognormal alternative, cf. Figure 5 in Section 3.

Simple and accurate approximations of the critical values of T_S and $T1_{S1}$ are discussed in Section 4.

3 Simulation

To see how well the new tests perform, a simulation study has been done. The Monte Carlo experiments for getting critical values are repeated 50,000 times and the Monte Carlo experiments for the simulated powers 10,000 times. Hence the standard deviation of the simulated powers does not exceed $(40,000)^{-1/2} = 0.005$. The random number generator is taken from Matsumoto and Nishimura (1998) and has a period of $2^{19937} - 1$.

3.1 Alternatives

The alternatives considered in the simulation study are the tail alternatives discussed in Albers and Schut (1996), tail alternatives presented in Neuhaus (1987) and tail alternatives considered in Janic-Wróblewska and Ledwina (1999).

The tail alternatives from Albers and Schut (1996) are of three types: late tail alternatives, early tail alternatives and increasing tail alternatives. Late tail alternatives are completely concentrated in the right-tail of the distribution and are investigated also in Mason and Schuenemeyer (1983, 1992). Early tail alternatives contain a constant part throughout and have a slow increase going into the tail. Between these two extremes there is the class of increasing tail alternatives. A more precise description is given in Table 1.

Table 1. Alternatives in the simulation study

Name	df
	if not specified, F denotes the logistic df with variance $\pi^2/9$, given by $F(x) = \dfrac{1}{1+e^{-x\sqrt{3}}}$; the standard logistic df refers to $F(x) = \dfrac{1}{1+e^{-x}}$
Late tail alternatives	$F(x - \theta B_a(F(x)))$ with $B_a(t) = \begin{cases} 0 & , 0 < t \leq 1-a, \\ a^{-3/2}\left(1 - \dfrac{1-a}{t}\right) & , 1-a < t < 1 \end{cases}$
Early tail alternatives	$F\left(x - \theta B_\tau(F(x))\right)$ with $B_\tau(t) = \left(\dfrac{1+2\tau}{3}\right)^{1/2} \left\{\dfrac{1-(1-t)^\tau}{\tau t}\right\}$
Increasing tail alternatives	$F\left(x - \theta B_r(F(x))\right)$ with $B_r(t) = \left(\dfrac{4r^2-1}{3r}\right)^{1/2} t^{r-1}$
Upper shift (Neuhaus)	$G(x) = F(x - F(4x)/2)$ with F the standard normal(N) or standard logistic(L) df
Pure Shift (Neuhaus)	$G(x) = F(x - 1/2)$ with F the standard normal(N) or standard logistic(L) df
Lognormal	$G(x) = F\left(\dfrac{\log x}{\sigma}\right)$ with F the standard normal df

For the late and early tail alternatives we get a constant shift by taking $a = 1, \tau = 1$, respectively; the closer we come with a or τ to 1, the closer we come to the constant shift. For the increasing tail alternatives the constant shift corresponds to $r = 1$.

3.2 Test statistics

For power comparison we consider the following tests.

- W: the one-sided Wilcoxon test
- $T0$: the optimal rank test for the given alternative
- $T2$: the data driven smooth test for the two-sample problem of Janic-Wróblewska and Ledwina
- NH: Neuhaus' test with bandwidth $a = 0.4$
- T_S: the new test given in Section 2.3
- $T1_{S1}$: the other new test given in Section 2.3.

Comparison with Wilcoxon's test is of interest for seeing what we would get if we did not worry about the dependence on x in the shift and simply would act as if there was a constant shift. Although the optimal rank test $T0$ changes from alternative to alternative and hence cannot be seen as competitor, its power is of interest as a kind of upper bound for the power at the given alternative. The data driven smooth test for the two-sample problem of Janic-Wróblewska and Ledwina has been developed for the general two-sample problem. It is of interest

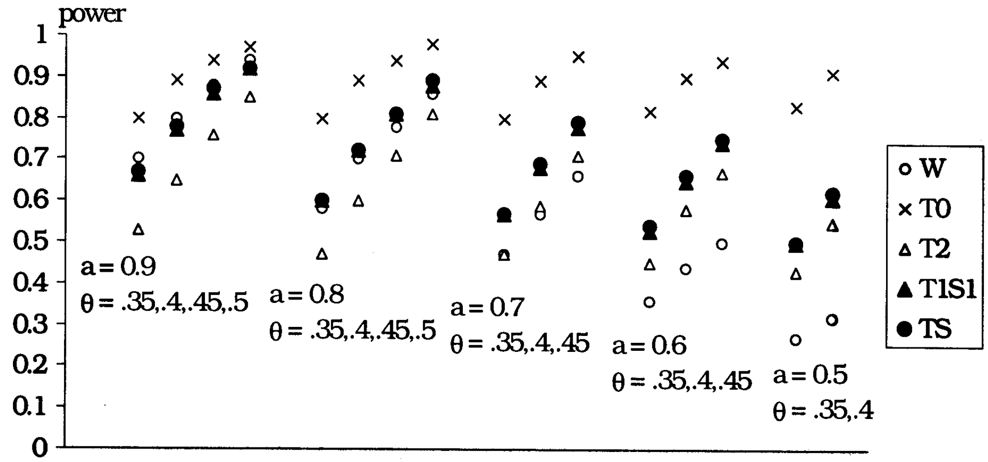

Figure 1: Simulated Powers of five tests at late tail alternatives; $m = n = 100; \alpha = 0.05$

to see whether power gain can be obtained when one focuses attention on tail alternatives. Finally, Neuhaus' test is "recommended in situations where there are doubts that all parts of the underlying distribution are shifted at the same rate" (Neuhaus, 1987 page 512) and hence this test is a natural competitor in our testing problem. For the lognormal and logistic distribution (see the pure shift alternative) further comparison can be made with tests of Fan (1996) and Schmid and Trede (1995) by combining our results with those of Janic-Wróblewska and Ledwina (1999).

3.3 Simulation results

The results of the simulation study for alternatives, mentioned in Table 1, and test statistics, given in Section 3.2 are presented in Figures 1–5.

Comparing the new tests with Wilcoxon's test, we see that the new tests clearly outperform Wilcoxon's test when the shift is not constant but becomes larger going into the tail of the distribution. This is seen obviously in Figures 1, 3 and 4 passing from the left to the right, cf. also the remarks about the parameters a and r shortly after Table 1. As early tail alternatives contain a constant part throughout and have a slow increase going into the tail, it is no surprise that Wilcoxon's test performs better for such type of alternatives. However, the power differences between Wilcoxon's test and the new tests in Figure 2 are not very large. In the lognormal case the power of Wilcoxon's test completely breaks down, leading to a very large gain of power by the new tests in Figure 5. The lack of power for Wilcoxon's test when dealing with lognormal alternatives is of

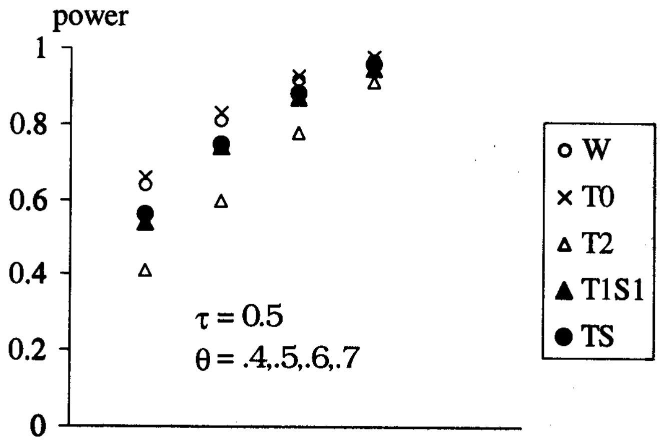

Figure 2: Simulated Powers of five tests at early tail alternatives;
$m = n = 50; \alpha = 0.05$

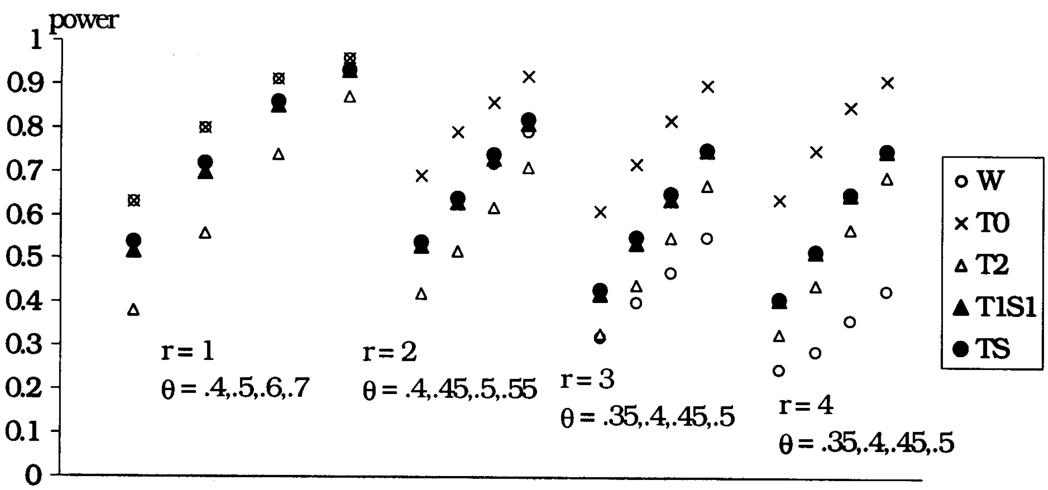

Figure 3: Simulated Powers of five tests at increasing tail alternatives;
$m = n = 50; \alpha = 0.05$

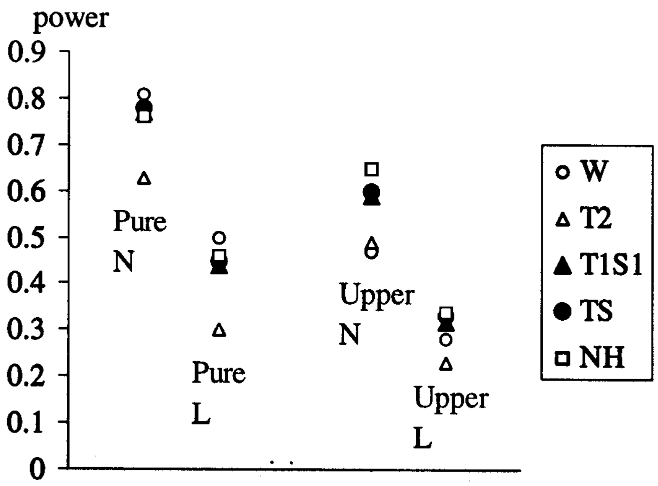

Figure 4: Simulated Powers of five tests at Neuhaus alternatives; $m = n = 40; \alpha = 0.1$

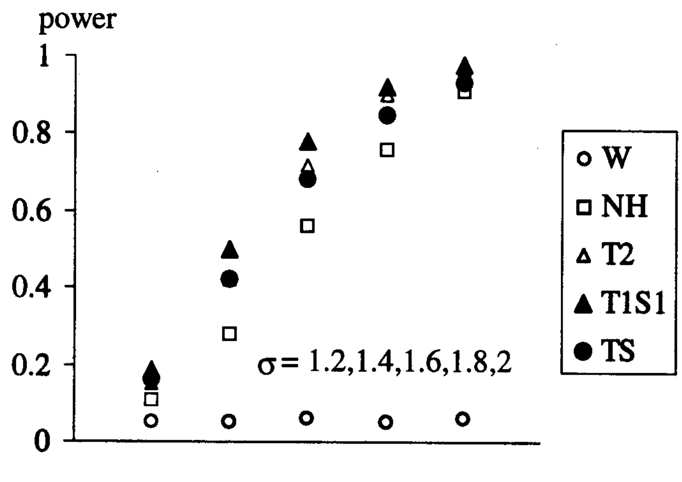

Figure 5: Simulated Powers of five tests at lognormal alternatives; $m = n = 100; \alpha = 0.05$

course no surprise, because here we have more or less a scale alternative and no shift, cf. Section 4.

The data driven smooth test for the two-sample problem of Janic-Wróblewska and Ledwina has been developed for the general two-sample problem. Figures 1–5 show that indeed substantial power can be gained by exploiting the tail alternatives.

In Figures 4 and 5 comparison with Neuhaus' test can be made. The lognormal case shows that the new test can lead to substantial power gain for the new tests compared to Neuhaus' test. In the case of the alternatives of Neuhaus (1987) the powers of the new tests are close to those of Neuhaus' test.

Combining our results with the simulation study of Janic-Wróblewska and Ledwina shows that in the logistic case (see Figure 4 Pure L) the new tests have comparable power to that of the test of Schmid and Trede, while the test of Fan has very poor power. For the lognormal alternatives (see Figure 5) the situation is the other way around: the power of the test of Schmid and Trede is very poor, while the test of Fan has comparable power to the powers of the new tests.

In the Figures 1–4 the two new tests have almost the same power with T_S slightly better than $T1_{S1}$. Since T_S is closer to Wilcoxon's test than $T1_{S1}$, and the lognormal case has no pure shift but is more a scale alternative, it is not surprising that $T1_{S1}$ has somewhat more power than T_S for this alternative. Nevertheless, T_S behaves rather well also for this alternative, being far better than Wilcoxon's test and substantially better than Neuhaus' test. For some more quantitative details concerning this alternative see Section 4.

The conclusions from the simulation results presented in Figures 1–5 and from other simulation results that we have performed are as follows.

- The new tests have high and stable power at the several types of tail alternatives.
- The new tests are far better than Wilcoxon's test when the shift function really increases in the tail, and have only slightly less power in case of (almost) constant shift.
- The new tests benefit from pointing on tail alternatives: substantial power gain is achieved compared to the data driven test for the unrestricted two-sample problem, proposed by Janic-Wróblewska and Ledwina.
- The new tests can obtain substantial higher power than Neuhaus' test, the test of Fan and the test of Schmid and Trede.
- Test T_S is often slightly better than $T1_{S1}$, but sometimes $T1_{S1}$ performs a little bit better than T_S.

4 Asymptotic null distribution and consistency

In Section 3 it was seen that the new tests T_S and $T1_{S1}$ behave very well in the simulation study. Here we first discuss the asymptotic null distribution of a class of test statistics containing the new tests T_S and $T1_{S1}$. For application of the new tests we need critical values and/or p-values. In the simulation study critical values are obtained by Monte Carlo methods. For practical use of the new tests simulation is rather inconvenient and hence accurate and simple approximations of critical values and p-values are needed. Such approximations, based on a second order analysis, are presented in Section 4.2. Furthermore, the nice power behavior of the new tests, which was shown in the simulations, is supported by proving consistency of the new tests for broad classes of alternatives, including all alternatives with Y stochastically larger than X and hence in particular all tail alternatives.

4.1 Asymptotic null distribution

We derive the asymptotic null distributions of test statistics

$$\tilde{T}_{\tilde{S}} \text{ with } 0 \le \tilde{T}_k \le \mathbf{Z}^T\mathbf{Z} \text{ and } \tilde{T}_1 = Z_1^2 1(Z_1 > 0),$$
$$\text{where } \tilde{S} = \min\left\{k : 1 \le k \le d(N) : \tilde{T}_k - k\log N \ge \tilde{T}_r - r\log N, 1 \le r \le d(N)\right\}.$$

Note that both test statistics T_S and $T1_{S1}$ are of this form.

We always assume that the sequences $m = m(N)$ and $n = n(N)$ tend to infinity as $N \to \infty$ in such a way that

$$\delta < m/N < 1 - \delta$$

for some $\delta > 0$.

The first result concerns the limiting behavior of the selection rule under H_0. As may be expected it concentrates on the lowest dimension, dimension 1.

Theorem 4.1 *Assume that H_0 is true and denote the probability measure under H_0 by P_0. If $d(N) = o(\{N/\log N\}^{1/9})$, then for every test statistic*

$$\tilde{T}_{\tilde{S}} \textit{ with } 0 \le \tilde{T}_k \le \mathbf{Z}^T\mathbf{Z} \textit{ and } \tilde{T}_1 = Z_1^2 1(Z_1 > 0),$$
$$\textit{where } \tilde{S} = \min\left\{k : 1 \le k \le d(N) : \tilde{T}_k - k\log N \ge \tilde{T}_r - r\log N, 1 \le r \le d(N)\right\}$$

we have

$$\lim_{N\to\infty} P_0(\tilde{S} = 1) = 1 \textit{ and } \lim_{N\to\infty} P_0(\tilde{T}_{\tilde{S}} \le x) = \begin{cases} \Phi\left(\sqrt{x}\right) & \textit{if } x \ge 0 \\ 0 & \textit{if } x < 0 \end{cases},$$

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

Proof The proof of Theorem 4.1 is a generalization of the method of proof, used in Kallenberg and Ledwina (1999) to prove Theorem 1 of that paper and is similar to the proof of Theorem 4.1 in Janic-Wróblewska and Ledwina (1999).

Because $\tilde{S} = k$ implies that dimension k "beats" dimension 1, we get by definition of $\tilde{S}$ and by the assumption $0 \le \tilde{T}_k \le \mathbf{Z}^T\mathbf{Z}$

$$P_0(\tilde{S} \ge 2) = \sum_{k=2}^{d(N)} P_0(\tilde{S} = k) \le \sum_{k=2}^{d(N)} P_0\left(\tilde{T}_k \ge (k-1)\log N\right)$$
$$\le \sum_{k=2}^{d(N)} P_0\left(\mathbf{Z}^T\mathbf{Z} \ge (k-1)\log N\right).$$

Application of (A.2) in Janic-Wróblewska and Ledwina (1999) yields

$$\lim_{N\to\infty} \sum_{k=2}^{d(N)} P_0\left(\mathbf{Z}^T\mathbf{Z} \ge (k-1)\log N\right) = 0$$

and hence

$$\lim_{N\to\infty} P_0(\tilde{S} = 1) = 1. \tag{15}$$

Using

$$P_0(\tilde{T}_{\tilde{S}} \le x) = P_0(\tilde{T}_1 \le x, \tilde{S} = 1) + P_0(\tilde{T}_{\tilde{S}} \le x, \tilde{S} \ge 2)$$
$$= P_0(\tilde{T}_1 \le x) - P_0(\tilde{T}_1 \le x, \tilde{S} \ge 2) + P_0(\tilde{T}_{\tilde{S}} \le x, \tilde{S} \ge 2),$$

the limiting distribution of $\tilde{T}_{\tilde{S}}$ under H_0 immediately follows from (15), the assumption that $\tilde{T}_1 = Z_1^2 1(Z_1 > 0)$ and the asymptotic standard normality of Z_1 under H_0.

4.2 Approximation for critical values and *p*-values

The first order approximation of T_S and $T1_{S1}$ given by Theorem 4.1 is not accurate. For instance, the simulated critical value for $m = n = 50$ and $\alpha = 0.05$ equals 3.710 for T_S and 3.872 for $T1_{S1}$, whereas the approximation based on Theorem 4.1 yields 2.706 for both. The same phenomenon occurs in data driven goodness of fit tests and in the data driven two sample test of Janic-Wróblewska and Ledwina. The remedy is a second order approximation, following the line of argument given by Kallenberg and Ledwina (1995a, sec. 4), where more details can be found.

The idea is that $T_S \ge T_1 = Z_1^2 1(Z_1 > 0)$ and $T1_{S1} \ge T1_1 = Z_1^2 1(Z_1 > 0)$, implying that $P(T_S \le x)$ and $P(T1_{SI} \le x)$ are overestimated by the first order approximation $\Phi(\sqrt{x})$. We apply the following approximations

$$
\begin{aligned}
P(T_S \le x) &\doteq P(T_1 \le x, S=1) + P(T_2 \le x, S=2) \\
&\doteq P(T_1 \le x) - P(T_1 \le x < T_2, S=2) \\
&\doteq P(T_1 \le x) - P(T_1 \le x < T_2, T_2 > T_1 + \log N).
\end{aligned}
$$

Replacing Z_1 and Z_2 by (independent) $N(0,1)$-distributed rv's, a further approximation leads to the final proposal

$$
\begin{array}{ll}
\Phi(\sqrt{x}) - \dfrac{0.05245}{\sqrt{N}} & \\
-\left\{2\Phi\left(\sqrt{x}\right) - \frac{1}{2} - \Phi\left(0.342\sqrt{\log N}\right)\right\}\Phi\left(-\sqrt{\log N}\right) & \text{if } x \le \log N \\
\Phi\left(\sqrt{x}\right) - \frac{1}{2}e^{-x/2} & \text{if } x \ge 2\log N \\
\text{linearize} & \text{if } \log N < x < 2\log N.
\end{array}
\tag{16}
$$

Similarly, a simple approximation of $P(T1_{S1} \le x)$ is given by

$$
\begin{array}{ll}
\Phi(\sqrt{x}) - \dfrac{1}{8\sqrt{N}} - \left\{\Phi(\sqrt{x}) - \frac{1}{2}\right\}\Phi\left(-\sqrt{\log N}\right) & \text{if } x \le \log N \\
\Phi(\sqrt{x}) - \frac{1}{2}e^{-x/2} & \text{if } x \ge 2\log N \\
\text{linearize} & \text{if } \log N < x < 2\log N.
\end{array}
\tag{17}
$$

To illustrate the accuracy of the approximations (16) and (17), we calculate the approximations for the simulated critical values. The results are presented in Tables 2 and 3.

Table 2. Accuracy of the approcimation (16) *at the simulated critical values of* T_S *for* $\alpha = 0.05$

$P(T_S \le x)$				
(m,n)	(25,25)	(50,50)	(100,100)	(40,60)
x	4.6043	3.7096	3.1799	3.6721
appr. (16)	0.9581	0.9569	0.9522	0.9563

Table 3. Accuracy of the approximation (17) *at the simulated critical values of* $T1_{S1}$ *for* $\alpha = 0.05$

$P(T1_{S1} \le x)$				
(m,n)	(25,25)	(50,50)	(100,100)	(40,60)
x	4.4450	3.8715	3.3123	3.8088
appr. (17)	0.9525	0.9554	0.9518	0.9544

It is seen that the approximations work very well. Therefore, p-values and critical values can be obtained from (16) and (17) with sufficient precision. This makes the new tests easily applicable in practice.

4.3 Consistency

The consistency of the tests based on T_S and $T1_{S1}$ is given in the following theorem.

Theorem 4.2 *Let P be any alternative and let F and G be the marginal distribution functions of X and Y, respectively, under P. Assume that $\lim_{N\to\infty} m(N)/N = \gamma$ for some $0 < \gamma < 1$. Let $H_\gamma(x) = \gamma F(x) + (1-\gamma)G(x)$.*

(i) Suppose that for some k,

$$\sum_{r=1}^{k} c_r \left\{E_P J_r(H_\gamma(Y)) - E_P J_r(H_\gamma(X))\right\} > 0 \tag{18}$$

with c_r given by (11). If $d(N)$ tends to infinity and $d(N) = o(\{N/\log N\}^{\frac{1}{9}})$, then T_S is consistent at P. If $\lim_{N\to\infty} d(N) = d < \infty$, say, then T_S is consistent at P, provided that (18) holds for some $k \le d$.

(ii) Suppose that for some k,

$$\sum_{r=1}^{k} \left\{E_P J_r(H_\gamma(Y)) - E_P J_r(H_\gamma(X))\right\} > 0. \tag{19}$$

If $d(N)$ tends to infinity and $d(N) = o(\{N/\log N\}^{\frac{1}{9}})$, then $T1_{S1}$ is consistent at P. If $\lim_{N\to\infty} d(N) = d < \infty$, say, then $T1_{S1}$ is consistent at P, provided that (19) holds for some $k \le d$.

Proof We present a proof of (i). The consistency of $T1_{S1}$ is proved in exactly the same way. Note that

$$\frac{R_i}{N} = \frac{m}{n}\hat{F}(X_i) + \frac{n}{N}\hat{G}(X_i) \text{ for } 1 \le i \le m$$

and

$$\frac{R_i}{N} = \frac{m}{n}\hat{F}(Y_i) + \frac{n}{N}\hat{G}(Y_i) \text{ for } m+1 \le i \le N$$

and hence, using $\max\{|J_r'(x)| : 0 \le x \le 1\} = r(r+1)\sqrt{2r+1}$,

$$\left|\sqrt{\frac{N}{mn}} Z_r - \frac{N}{mn}\left\{\frac{m}{N}\sum_{i=m+1}^{N} J_r(H_\gamma(Y_i)) - \frac{n}{N}\sum_{i=1}^{m} J_r(H_\gamma(X_i))\right\}\right|$$

$$\leq r(r+1)\sqrt{2r+1}\left[\begin{array}{l} 2\left\{\frac{m}{N}\left\|\hat{F}-F\right\|_\infty+\frac{n}{N}\left\|\hat{G}-G\right\|_\infty+\frac{1}{2N}\right\} \\ +\frac{1}{n}\sum_{i=m+1}^{N}\left\{\left|\frac{m}{N}-\gamma\right|F(Y_i)+\left|\frac{n}{N}-(1-\gamma)\right|G(Y_i)\right\} \\ +\frac{1}{m}\sum_{i=1}^{m}\left\{\left|\frac{m}{N}-\gamma\right|F(X_i)+\left|\frac{n}{N}-(1-\gamma)\right|G(X_i)\right\} \end{array}\right]. \quad (20)$$

Define $\mathbf{E} = (E_1, \ldots, E_k)^T$ with $E_r = E_P J_r(H_\gamma(Y)) - E_P J_r(H_\gamma(X))$. Since the right-hand side of (20) tends to 0 in probability and since by the law of large numbers

$$\frac{N}{mn}\left\{\frac{m}{n}\sum_{i=m+1}^{N} J_r(H_\gamma(Y_i)) - \frac{n}{N}\sum_{i=1}^{m} J_r(H_\gamma(X_i))\right\} \xrightarrow{P} E_r,$$

we get

$$\sqrt{\frac{N}{mn}} Z_r \xrightarrow{P} E_r. \quad (21)$$

Let k^* be the smallest $k \geq 1$ for which (18) holds. In view of (21) we have for any $k \in \{1, \ldots, k^* - 1\}$ that

$$\frac{N}{mn} T_k \xrightarrow{P} 0 \text{ and furthermore, } \frac{N}{mn} T_{k^*} \xrightarrow{P} \mathbf{E}^T\mathbf{E}\, 1\left(\sum_{r=1}^{k^*} c_r E_r > 0\right) > 0. \quad (22)$$

Hence $T_{k^*} \xrightarrow{P} \infty$. Assume $d(N) \geq k^*$, which holds for sufficiently large N, because $d(N)$ tends to infinity or $k^* \leq d = \lim_{N\to\infty} d(N)$ (the latter implies $d(N) = d$ for sufficiently large N). Since for any $k \in \{1, \ldots, k^* - 1\}, S = k$ implies

$$\frac{N}{mn} T_k - \frac{N}{mn} k \log N \geq \frac{N}{mn} T_{k^*} - \frac{N}{mn} k^* \log N,$$

(22) yields $P(S = k) \to 0$ for any $k \in \{1, \ldots, k^* - 1\}$ and thus $P(S \geq k^*) \to 1$ as $N \to \infty$. If $S = s$ with $s \geq k^*$, then $T_s \geq T_{k^*}$ by definition of the selection rule. Because $P(S \geq k^*) \to 1$ and $T_{k^*} \xrightarrow{P} \infty$, we also get $T_s \xrightarrow{P} \infty$. Application of Theorem 4.1 completes the proof of the consistency.

Corollary 4.3 *Assume that* $\lim_{N\to\infty} m(N)/N = \gamma$ *for some* $0 < \gamma < 1$ *and* $d(N) = o(\{N/\log N\}^{\frac{1}{9}})$. *If* Y *is stochastically larger than* X, *we have consistency for* T_S *and* $T1_{S1}$. □

Proof By considering $k = 1$ in (18) and (19) we see that consistency of T_S and $T1_{S1}$ is obtained if

$$E_P H_\gamma(Y) > E_P H_\gamma(X)$$

and this is satisfied if Y is *stochastically larger* than X.

For all alternatives considered in the simulation study, except for the lognormal case, it is immediately seen that Y is *stochastically larger* than X and hence by Corollary 4.3 consistency of T_S and $T1_{S1}$ is obtained for these alternatives.

For the lognormal alternative we get $E_1 = E_P J_1 H_\gamma(Y)) - E_P J_1(H_\gamma(X)) = 0$ and hence we can not expect power for $k = 1$. This is clearly seen in Figure 5, where the power of Wilcoxon's test completely breaks down. By direct calculation we get, writing U for a rv with a standard normal distribution,

$$\begin{aligned} E_2 = E_P J_2(H_\gamma(Y)) - E_P J_2(H_\gamma(X)) = 6\sqrt{5}\,[\gamma E\{\Phi(\sigma U)\}^2 \\ + (1-2\gamma)E\{\Phi(U)\}^2 - (1-\gamma)E\{\Phi(U/\sigma)\}^2]\,. \end{aligned} \tag{23}$$

As (23) is linear inγ, it is seen that $E_2 > 0$ for $\sigma > 1$ by considering $\gamma = 1$ and $\gamma = 0$ and noting that $E\{\Phi(tU)\}^2$ is increasing in t. Hence, by (18) and (19) with $k = 2$, both T_S and $T1_{S1}$ are consistent also for the lognormal alternatives. For $\gamma = \frac{1}{2}$, used in the simulation, we get $E_2 = E_P J_2(H_{1/2}(Y)) - E_P J_2(H_{1/2}(X)) =$ 0.223, 0.405, 0.553, 0.674, 0.775 for $\sigma = 1.2,\ 1.4,\ 1.6,\ 1.8,\ 2.0$, respectively. The facts that $E_1 = 0$ and $E_2 > 0$, and that T_S gives higher weight to Z_1 than $T1_{S1}$, explain the somewhat higher power for $T1_{S1}$ at the lognormal alternatives.

Remark Many modifications are possible for defining data driven smooth tests, such as other orthonormal systems and other penalties in the selection rule. These issues are extensively discussed e.g. in Kallenberg and Ledwina (1997b) and Inglot et al. (1997). Here we mention the replacement of Legendre polynomials by certain Jacobi polynomials associated with inner products giving, in particular, other weights at the tails of the interval (0,1). In our simulations these tests performed not substantially better and hence we have not presented the results here. Another modification is not to use orthonormal systems, but simply to start with the functions

$$B_r(t) = \sqrt{\frac{4r^2-1}{3r}}\,t^{r-1},$$

given by the increasing tail alternatives (see Albers and Schut, 1996). This leads to

$$J_r(t) = \sqrt{\frac{4r^2-1}{r}}\left[(r+1)t^r - rt^{r-1}\right].$$

Although such a direct approach sounds attractive, serious problems occur, because the functions J_r are too close to each other and hence do not span efficiently the space of alternatives of interest.

References

Albers,W., and Akritas, M. G. (1987), "Combined Rank Tests for the Two-Sample Problem With Randomly Censored Data," *Journal of the American Statistical Association*, 82, 648-655.

Albers, W., and Schut, C. (1996), "A Survey of Rank Tests for Classes of Tail Alternatives," In *Research Developments in Probability and Statistics*, Festschrift in Honor of Madan L. Puri on the Occasion of his 65th Birthday, Edited by E. Brunner and M. Denker, 289-302, Utrecht: VSP.

Bickel, P. J., and Ritov Y. (1992), "Testing for Goodness of Fit: A New Approach," In *Nonparametric Statistics and Related Topics*, Edited by A. K. Md. E. Saleh, 51-57, Amsterdam: North-Holland.

Behnen, K., and Neuhaus, G. (1989), *Rank Tests With Estimated Scores and Their Applications*, Stuttgart: Teubner.

Bogdan, M. (1995), "Data Driven Versions of Pearson's Chi-Square Test for Uniformity," *Journal of Statistical Computation and Simulation*, 52, 217-237.

Bogdan, M., and Ledwina, T. (1996). "Testing Uniformity via Log-Spline Modeling," *Statistics*, 28, 131-157.

Efron, B. (1998), "R. A. Fisher in the 21st Century," *Statistical Science*, 13, 95-122.

Eubank, R. L. (1997), "Testing Goodness of Fit With Multinomial Data," *Journal of the American Statistical Association*, 92, 1084-1093.

Eubank, R. L., and LaRiccia, V. N. (1992), "Asymptotic Comparison of Cramér-von Mises and Nonparametric Function Estimation Techniques for Testing Goodness-of-Fit," *The Annals of Statistics*, 20, 2071-2086.

Eubank, R. L., LaRiccia, V. N., and Rosenstein, R. B. (1987), "Test Statistics Derived as Components of Pearson's Phi-Squared Distance Measure," *Journal of the American Statistical Association*, 82, 816-825.

Fan, J. (1996), "Test of Significance Based on Wavelet Thresholding and Neyman's Truncation," *Journal of the American Statistical Association*, 91, 647-688.

Fleming, T. R., O'Fallon, J. R., O'Brien, P. C., and Harrington, D. P. (1980), "Modified Kolmogorov-Smirnov Test Procedures With Applications to Arbitrarily Right Censored Data," *Biometrics*, 69, 553-566.

Follmann, D. (1996), "A Simple Multivariate Test for One-Sided Alternatives," *Journal of the American Statistical Association*, 91, 854-861.

Hájek, J., Šidák, Z., and Sen, P. K. (1999), *Theory of Rank Tests*, Second Edition, San Diego: Academic Press.

Harrington, D. P., and Fleming, T. R. (1982), "A Class of Rank Test Procedures for Censored Survival Data," *Biometrics*, 69, 553-566.

Hušková, M., and Sen, P. K. (1985), "On Sequentially Adaptive Asymptotically Efficient Rank Statistics," *Sequential Analysis*, 4, 225-251.

—— (1986), "On Sequentially Adaptive Signed Rank Statistics," *Sequential Analysis*, 5, 237-251.

Inglot, T., Kallenberg, W. C. M., and Ledwina, T. (1997) "Data Driven Smooth Tests for Composite Hypotheses," *The Annals of Statistics*, 25, 1222-1250.

—— (1998), "Vanishing Shortcoming of Data Driven Neyman's tests," In *Asymptotic Methods in Probability and Statistics*, A Volume in Honour of Miklós Csörgő, Edited by B. Szyszkowicz, 811-829, Amsterdam: North-Holland.

Inglot, T., and Ledwina, T. (1996), "Asymptotic Optimality of Data-Driven Neyman's Tests for Uniformity," *The Annals of Statistics*, 24, 1982-2019.

Janic-Wróblewska, A., and Ledwina, T. (1999), "Data Driven Rank Test for Two-Sample Problem," *Scandinavian Journal of Statistics*, to appear.

Kallenberg, W. C. M., and Ledwina, T. (1995a) "On Data Driven Neyman's Tests," *Probability and Mathematical Statistics*, 15, 409-426.

—— (1995b), "Consistency and Monte Carlo Simulation of a Data Driven Version of Smooth Goodness-of-Fit Tests," *The Annals of Statistics*, 23, 1594-1608.

—— (1997a), "Data-Driven Smooth Tests for Composite Hypotheses: Comparison of Powers," *Journal of Statistical Computation and Simulation*, 59, 101-121.

—— (1997b), "Data-Driven Smooth Tests When the Hypothesis is Composite," *Journal of the American Statistical Association*, 92, 1094-1104.

—— (1999), "Data-Driven Rank Tests for Independence," *Journal of the American Statistical Association*, 94, 285-301.

Kallenberg, W. C. M., Ledwina, T., and Rafajłowicz, E. (1997), "Testing Bivariate Independence and Normality," *Sankhyā* Ser. A, 59, 42-59.

Ledwina, T. (1994), "Data Driven Version of Neyman's Smooth Test of Fit," *Journal of the American Statistical Association*, 89, 1000-1005.

Mason, D. M., and Schuenemeyer, J. H. (1983), "A Modified Kolmogorov-Smirnov Test Sensitive to Tail Alternatives," *The Annals of Statistics*, 11, 933-946.

—— (1992), "Correction A Modified Kolmogorov-Smirnov Test Sensitive to Tail Alternatives," *The Annals of Statistics*, 20, 620-621.

Matsumoto, M., and Nishimura, T. (1998), "Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator," *ACM Transactions on Modeling and Computer Simulation*, 8, 3-30.

Neuhaus, G. (1987), "Local Asymptotics for Linear Rank Statistics With Estimated Score Functions," *The Annals of Statistics*, 15, 491-512.

—— (1988), "Addendum to: Local Asymptotics for Linear Rank Statistics With Estimated Score Functions," *The Annals of Statistics*, 16, 1342-1343.

Rayner, J. C. W., and Best, D. J. (1989), *Smooth Tests of Goodness of Fit*, New York: Oxford University Press.

Sansone, G. (1959), Ortogonal Functions, Interscience, New York.

Schmid,F., and Trede, M. (1995), "A Distribution Free Test for the Two Sample Problem for General Alternatives," *Computational Statistics & Data Analysis*, 20, 409-419.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461-464.