PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 32, Fasc. 2 (2012), pp. 323–358

DATA DRIVEN TESTS FOR UNIVARIATE SYMMETRY

BY

TADEUSZ INGLOT (WROCŁAW), ALICJA JANIC (WROCŁAW),
AND JADWIGA JÓZEFCZYK (WROCŁAW)

Abstract. We propose new data driven score rank tests for univariate symmetry around a known center. We apply both Schwarz-type and recently introduced data driven penalty selection rules. Some key asymptotic results regarding the test statistics are given and some asymptotic optimality properties proved. In an extensive simulation study, we compare the empirical behaviour of these tests to tests found in the recent literature to be powerful. We show that, for a broad range of asymmetric distributions, data driven tests have stable power, which is comparable to their competitors for typical alternatives and much greater for some atypical alternatives.

2000 AMS Mathematics Subject Classification: Primary: 62G10; Secondary: 62G30, 65C05, 65C60.

Key words and phrases: Testing symmetry, data driven score test, selection rule, rank test, vanishing shortcoming, Kallenberg efficiency, Hungarian construction, modified sign test, hybrid test, optimal Bayes test, Monte Carlo study.

1. INTRODUCTION

It has been strongly argued that the symmetry or asymmetry of a distribution is essential for the validity of many statistical models. Simultaneously, testing for symmetry (about a known center) is one of the oldest classical nonparametric problems. Beginning with Fisher's sign test through the tests assembled in the monograph of Hájek and Šidák [13] and followed by a large variety of recent constructions (e.g., McWilliams [36], Modarres and Gastwirth [37], [38], Janic-Wróblewska [26], Cheng and Balakrishnan [4], Thas et al. [45], and Bakshaev [1] to mention only some examples), this problem has received constant attention in the literature. Hájek and Šidák [13] described a broad class of linear rank tests, each being optimal for a particular type of asymmetry. On the other hand, omnibus tests, based on Cramér–von Mises and Kolmogorov–Smirnov type statistics, are widely recommended in the literature (see, e.g., Orlov [40], Rothman and Woodroofe [41], Doksum et al. [6], Koziol [33]), but attain high power only for narrow classes of asymmetric distributions. The different approach adopted by

Behnen and Neuhaus [2] has turned out to be successful in improving the range of sensitivity of such tests to a wider class of asymmetric distributions.

In the last two decades, some new concepts have been introduced. One of them, proposed by Cheng and Balakrishnan [4], modifies the sign test by trimming the sum of signs only to that corresponding to a few of the largest and smallest observations. The resulting test can detect asymmetry in the tails of a distribution with very high power. An interesting construction was presented by Modarres and Gastwirth [38]. Their two-stage testing procedure, called the hybrid test, combines the sign test with a modification of the conditional Wilcoxon test introduced by Tajuddin [44] by considering only the tails of the observed distribution. The hybrid test performs very well and provides superior power in comparison with many classical or recently proposed procedures for alternatives with a typical form of asymmetry. Another construction based on a decomposition of the chi-square statistic for independence in a particular contingency table was recently proposed by Thas et al. [45].

Further progress was achieved by Janic-Wróblewska [26], who adopted the methodology of data driven tests developed by Ledwina [34], Kallenberg and Ledwina [29]–[31], and Inglot et al. [19] to the problem of testing symmetry. She proved some asymptotic results for a new test statistic and showed the good empirical behaviour of the test based on it by a small simulation study. Later on, such an approach proved to be successful in the construction of new powerful nonparametric procedures as, e.g., testing for independence (Kallenberg and Ledwina [32], or Janic-Wróblewska et al. [27]) or testing in two- and k-sample problems (Janic-Wróblewska and Ledwina [28], Wyłupek [46], Ledwina and Wyłupek [35]).

In the present paper we continue, extend and generalize the work of Janic-Wróblewska [26]. In Section 2 we recall the construction of the test statistic introduced in Janic-Wróblewska [26] and provide some new data driven rank test statistics by applying some new selection rules proposed in Inglot and Janic [17] for testing uniformity. In Section 3 we give an extensive study of the asymptotic behaviour of these test statistics. We also state some asymptotic optimality properties of the tests based on them. Roughly speaking, they say that our tests are as good as the most powerful test in the sense of asymptotic powers for converging alternatives. In Section 4 results of the simulation study are presented. We compare the empirical behaviour of our data driven tests with selected tests, which are all recommended in the recent literature as being powerful. We make these comparisons for a broad variety of alternatives representing different types of asymmetry. We show that data driven tests have stable power, which is comparable to their competitors for typical alternatives and much greater for some atypical alternatives. We show empirically that the proposed data driven tests are significantly more powerful (on average) than any of their competitors. These conclusions are in accordance with the theoretical optimality properties discussed in Section 3. Section 5 contains proofs of all the theorems, which rely heavily on the results given in the Appendix. In the Appendix we present some general asymptotic results for linear rank statistics using a new approach based on the Hungarian construction.

2. DATA DRIVEN TESTS

Let $X_1, \dots, X_n$ be i.i.d. real random variables with a continuous distribution function $F(x)$. We are interested in testing

$$H_0 : F(\mu + x) = 1 - F(\mu - x), \quad x \in \mathbb{R},$$

i.e. testing the symmetry of F about a known median μ, which throughout will be assumed to be equal to zero.

Furthermore, denote by $F_s(x) = \frac{1}{2}\big(F(x) + 1 - F(-x)\big)$ the distribution function of the symmetric part of F and put $F_a = F - F_s$. Then testing H_0 is equivalent to testing whether $F = F_s$. We transform the data into the unit interval using F_s to obtain $U_1, \dots, U_n$ with $U_i = F_s(X_i),\ i = 1, \dots, n$. Since F is absolutely continuous with respect to its symmetric part F_s, the transformed data U_i have a distribution function

$$F \circ F_s^{-1}(t) = t + A(t), \quad t \in [0, 1],$$

where $A(t)$ is an absolutely continuous function symmetric with respect to $t = 1/2$. Equivalently they have a density on $[0, 1]$ of the form

$$p(t) = 1 + a(t),$$

where $a(t)$ is an antisymmetric, with respect to $t = 1/2$, derivative of $A(t)$. So, testing H_0 is equivalent to testing that $a = 0$. Observe that $|a(t)| \leqslant 1$ a.s. due to the above definition of A.

The above analysis allows us to follow the idea of Neyman [39] and embed the null distribution (the uniform distribution over [0,1]) into the nested sequence of exponential families of densities

$$f_k(t, \vartheta) = c_k(\vartheta) \exp\Big\{ \sum_{j=1}^{k} \vartheta_j b_{2j-1}(t) \Big\}, \quad k = 1, 2, \dots, \tag{2.1}$$

where $b_1, b_3, \dots, b_{2k-1}$ denote the odd Legendre polynomials (cf. Sansone [42]), $\vartheta = (\vartheta_1, \vartheta_2, \dots, \vartheta_k)^T \in \mathbb{R}^k$, v^T stands for the transposition of a vector v, and $c_k(\vartheta)$ is the normalizing constant.

Suppose that $p(t) = 1 + a(t)$ can be treated approximately as a member of the family (2.1). Then H_0 reduces to testing $\vartheta = 0$. By the orthonormality of the Legendre polynomials, the score statistic for such a parametric problem takes the form

$$\sum_{j=1}^{k} \Big\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} b_{2j-1}\big(F_s(X_i)\big) \Big\}^2. \tag{2.2}$$

Since F_s is unknown (and therefore can be regarded as a nuisance parameter in our initial testing problem), we shall apply its natural estimator of the form

$$\mathcal{F}_{ns}(x) = \frac{1}{2}\big(\mathcal{F}_n(x) + 1 - \mathcal{F}_n(-x)\big),$$

where $\mathcal{F}_n(x)$ is the empirical distribution function of the original sample $X_1, \ldots, X_n$. Consequently, the test statistic for testing H_0 within the family f_k is of the form

$$N_k = \sum_{j=1}^{k} \hat{b}_{2j-1}^2, \tag{2.3}$$

where

$$\hat{b}_{2j-1} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} b_{2j-1}\left(\mathcal{F}_{ns}(X_i) - \frac{1}{4n}\right) \tag{2.4}$$

shall be called *empirical Fourier coefficients*. In (2.4) we have inserted a usual continuity correction. It is easily seen that (2.3) can be written in the following two equivalent forms:

$$\begin{aligned} N_k &= \sum_{j=1}^{k}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\operatorname{sign}(X_i)b_{2j-1}\left(\frac{n+R_i^+-1/2}{2n}\right)\right)^2 \\ &= \sum_{j=1}^{k}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}b_{2j-1}\left(\frac{R_i-1/2}{2n}\right)\right)^2, \end{aligned} \tag{2.5}$$

where R_i^+ is the rank of the absolute value $|X_i|$ among $|X_1|, \ldots, |X_n|$, $\operatorname{sign}(X_i)$ is the sign of X_i, while R_i is the rank of X_i in the pooled sample $X_1, \ldots, X_n, -X_1, \ldots, -X_n$.

To apply N_k for testing H_0, one needs to choose a number k of components from the set $\{1, \ldots, d(n)\}$, where $d(n) \geqslant 1$ can unboundedly increase for $n \to \infty$. To ensure that this choice is well suited to the observations, we base it on the data. Namely, following Kallenberg and Ledwina [31], Janic-Wróblewska [26] and Janic-Wróblewska and Ledwina [28], we take a score based selection rule S which is a modification of the Schwarz BIC criterion (cf. also Schwarz [43]), defined by

$$S = \min\{1 \leqslant k \leqslant d(n) : N_k - k\log n \geqslant N_j - j\log n,\ 1 \leqslant j \leqslant d(n)\}, \tag{2.6}$$

and the corresponding data driven test statistic is set to be N_S. This selection rule, with a relatively large penalty, is not able to detect highly oscillating alternatives with good power. So, one can adopt data driven choices of the penalty, which for the case of goodness-of-fit testing problem were proposed in Inglot and Ledwina [25] and Inglot and Janic [17]. The idea for such choices is to apply Schwarz penalty when empirical Fourier coefficients are of small magnitude (and therefore indicate for H_0) or to apply much lighter Akaike penalty, otherwise. Here, we shall define the selection rule L, which will be defined analogously to the one introduced in Inglot and Janic [17], where it was shown to have good empirical properties.

For this purpose take a natural number $D_n \geqslant 1$, $D_n \leqslant d(n)$, as well as a small positive number δ_n, and consider thresholds c_{jn} given by (cf. Inglot and Janic [17] for some explanation)

$$1 - \mathbf{\Phi}(c_{jn}) = \frac{1}{2}\left(\delta_n D_n^{-1}\binom{d(n)}{j}^{-1}\right)^{1/j},$$

where $\mathbf{\Phi}$ denotes the standard normal distribution function. Next, order $\hat{b}_1^2, \hat{b}_3^2, \ldots, \hat{b}_{2d(n)-1}^2$ from the smallest to the largest, obtaining $\hat{b}_{(1)}^2, \ldots, \hat{b}_{(d(n))}^2$, and consider the event

$$W_n = \{\hat{b}_{(d(n))}^2 \geqslant c_{1n}^2\} \cup \{\hat{b}_{(d(n)-1)}^2 \geqslant c_{2n}^2\} \cup \{\hat{b}_{(d(n)-D_n+1)}^2 \geqslant c_{D_n n}^2\}.$$

Denote by $\mathbf{1}_A$ the indicator of the set A. Then define the penalty

$$\pi(j,n) = j \log n \cdot \mathbf{1}_{W_n^c} + 2j \cdot \mathbf{1}_{W_n},$$

where W_n^c denotes the complement of W_n, the corresponding selection rule L,

$$L = \min\{1 \leqslant k \leqslant d(n) : \; N_k - \pi(k,n) \geqslant N_j - \pi(j,n), \; 1 \leqslant j \leqslant d(n)\}, \tag{2.7}$$

and the data driven test statistics N_L also denoted by $N_L(D_n, \delta_n)$. These statistics define upper-tailed tests for H_0. After some numerical comparisons, conducted for the usual significance level $\alpha = 0.05$, we restrict our attention mainly to the cases $\delta_n = 0.05, \; D_n = 1$ and $\delta_n = 0.05, \; D_n = 3$. In the sequel, we denote the corresponding selection rules succinctly by $L1$ and $L3$, respectively, and the corresponding statistics by N_{L1} and N_{L3}. For $\delta_n = 0$ we get the statistic N_S. For other significance levels α, we suggest to take δ_n comparable with α and again D_n between 1 and 3.

In Janic-Wróblewska [26], it was shown that, under H_0, the selection rule S is consistent (i.e. $S \to 1$ in probability) provided $d(n) = o(n/\log n)^{1/9}$ and, consequently, that asymptotically N_S has the central chi-square distribution with one degree of freedom. In the next section, we slightly strengthen this result and give a more complete study of asymptotic properties of the test statistics N_S and N_L. In Section 4 we report results of an extensive study of empirical behaviour of the three above described tests N_S, N_{L1}, and N_{L3} in comparison to other tests which are recommended in the literature as being good tests for symmetry.

Finally, note that although we restrict our attention to exponential families based on the odd Legendre polynomials, all our considerations remain true if we take any orthonormal system of absolutely continuous functions on $[0,1]$ which are antisymmetric with respect to 1/2.

3. ASYMPTOTIC BEHAVIOUR AND OPTIMALITY OF DATA DRIVEN TESTS

In this section, we provide asymptotic results for data driven tests based on N_S and N_L described in the previous section both under the null hypothesis (symmetry) and under alternatives. To this end let us put

$$\hat{b} = \hat{b}(n) = (\hat{b}_1, \hat{b}_3, \ldots) \tag{3.1}$$

to denote an infinite sequence of empirical Fourier coefficients given by (2.4) with respect to the odd Legendre polynomials. Then (cf. (2.3)) obviously

$$N_k = |\hat{b}|_k^2, \quad k \geqslant 1,$$

where $|v|_k = (v_1^2 + \ldots + v_k^2)^{1/2}$ stands for the k-dimensional Euclidean norm of a vector $v = (v_1, v_2, \ldots)^T$.

Recall $d(n) \geqslant 1$ is a number of components we allow for our statistics (a maximal dimension of the model we consider), and S and L are the selection rules defined in (2.6) and (2.7). Of course, we are mostly interested in the case $d(n) \to \infty$.

3.1. Null hypothesis. Suppose H_0 is true, i.e. the distribution of the sample satisfies $F = F_s$. The first statement of the following theorem slightly strengthens Theorem 3.1 in Janic-Wróblewska [26]. We present its proof in Section 5. It goes along another line of the argument than applied in Janic-Wróblewska [26]. We give it because a similar methodology has been used to prove all our next results.

THEOREM 3.1. *Assume $d(n) = O(n^\tau)$ for some $\tau < 1/6$.*

(1) *We have $S \to 1$ in probability and, consequently, $N_S \xrightarrow{\mathcal{D}} \chi_1^2$, where χ_k^2 denotes a random variable with the central chi-square distribution with k degrees of freedom.*

(2) *If, in addition, $D_n = D \geqslant 1$ is a fixed natural number, $D < d(n)$, and $\delta_n > 0$ is such that $\delta_n \to 0$, $\log(1/\delta_n) = o(n)$, and $\big(\log\log d(n)\big)/\log(1/\delta_n) \to 0$, then $P(L = S) \to 1$ and, consequently, $N_L \xrightarrow{\mathcal{D}} \chi_1^2$.*

It is worth noting that the convergence to the limiting distribution for both statistics under H_0 is rather slow, and exact critical values for upper-tailed tests based on N_S and N_L are, for moderate sample sizes, essentially larger than the asymptotic ones (cf., for example, Ledwina [34]). To overcome this problem, one can use simulated critical values or apply some approximation formula for the exact null distributions of these statistics. Both solutions are discussed in Subsection 4.1.

The moderate deviation theorem for N_S and N_L, we state below and prove in Section 5, will be useful to derive some optimality properties of our tests. Recall that $1 \leqslant S \leqslant L \leqslant d(n)$ a.s.

THEOREM 3.2. *Assume $d(n) = O(n^\tau \log n)$ for some $0 \leqslant \tau < 1/12$. Then for any sequence x_n of positive numbers satisfying the conditions*

$$nx_n^\vartheta \to 0, \qquad \frac{n^{1-2\tau}x_n^2}{\log^4 n} \to \infty$$

for some $\vartheta < 3/(1+3\tau)$ we have

$$\begin{aligned}(3.2)\qquad P(N_S \geqslant nx_n^2) &\leqslant P(N_L \geqslant nx_n^2) \leqslant P(N_{d(n)} \geqslant nx_n^2)\\ &= \exp\Big\{-\frac{1}{2}nx_n^2 + O(nx_n^{1+\vartheta/2}) + O\left(d(n)\log nx_n^2\right)\Big\}\\ &= \exp\Big\{-\frac{1}{2}nx_n^2 + o(\sqrt{n}x_n)\Big\}.\end{aligned}$$

3.2. Alternative hypothesis. Suppose that $F \neq F_s$. More precisely, let F_s be some (unknown) distribution function of a symmetric distribution on $\mathbb{R}$ and

$$F \circ F_s^{-1}(t) = t + A(t),$$

where $A(t)$ is an absolutely continuous function symmetric with respect to $1/2$. This means that the distribution of the original sample is $F(x) = F_s(x) + A\big(F_s(x)\big)$ with some unknown F_s.

The first result deals with fixed alternatives as described above.

THEOREM 3.3. *Assume $d(n) \to \infty$ and $d(n) = O(n^\tau)$ for some $\tau < 1/6$, and F is a fixed asymmetric distribution. Then $N_S \to \infty$ in probability and, consequently, $N_L \to \infty$ in probability. This means that data driven symmetry tests based on N_S and N_L are consistent against any fixed alternative (asymmetric distribution).*

The proof of Theorem 3.3 is given in Section 5.

Now, let us consider alternatives with distribution functions F_n converging to the null hypothesis at an intermediate rate. Namely, consider sequences of alternatives of the form

$$F_n(x) = F_s(x) + \rho_n A\big(F_s(x)\big), \tag{3.3}$$

where F_s is again some (unknown) distribution function of a symmetric distribution on $\mathbb{R}$,

$$\rho_n \to 0 \quad \text{and} \quad \frac{n\rho_n^2}{\log^2 n} \to \infty \tag{3.4}$$

and $A(t)$ has a normalized derivative $a(t)$ antisymmetric with respect to $t = 1/2$, i.e. $\int_0^1 a^2(t)dt = 1$ (recall that a is bounded a.s.).

Set

$$s_n = \rho_n \sqrt{n} \int_0^1 b(t) a(t) dt \tag{3.5}$$

being an asymptotic shift, where $b = (b_1, b_3, \ldots)$ denotes the sequence of the odd Legendre polynomials. The following limit theorem for N_S and N_L is proved in Section 5.

THEOREM 3.4. *Assume $d(n) \to \infty$ and $d(n) = O(n^\tau \log n)$ for some $0 \leqslant \tau < 1/12$, and the sequence (P_n) of alternatives is defined as in* (3.3) *with arbitrary normalized $a(t)$ antisymmetric with respect to $1/2$. Then for any ρ_n satisfying $n^{1-2\tau}\rho_n^2/\log^4 n \to \infty$ and $n\rho_n^4 = o(1)$ and for every $x \in \mathbb{R}$ we have*

$$P_n\biggl(\frac{N_S - |s_n|^2_{d(n)}}{2|s_n|_{d(n)}} \leqslant x\biggr) \to \mathbf{\Phi}(x),\ P_n\biggl(\frac{N_L - |s_n|^2_{d(n)}}{2|s_n|_{d(n)}} \leqslant x\biggr) \to \mathbf{\Phi}(x) \tag{3.6}$$

as $n \to \infty$.

3.3. Efficiency of data driven tests. The considerations presented in this subsection correspond to those of Inglot et al. [20] (cf. also Ducharme and Ledwina [7] or Inglot [15]).

Let w be an arbitrary real number and for the alternative given by (3.3) define the critical value of the upper-tailed test based on N_S by

$$t_n = |s_n|^2_{d(n)} + 2w|s_n|_{d(n)}. \tag{3.7}$$

Denote by α_n the size of this test and by β_n its power under F_n. Now, for each n consider the Neyman–Pearson test for testing $p_0(t) \equiv 1$ against $p_n(t) = 1 + \rho_n a(t)$, defined by the statistic

$$V_n = \frac{1}{\sqrt{n}\sigma_{0n}} \sum_{i=1}^n \bigl[\log\bigl(1 + \rho_n a(U_i)\bigr) - \mu_{0n}\bigr],$$

where $\mu_{0n} = \int_0^1 \log p_n(t) dt$ and $\sigma_{0n}^2 = \int_0^1 \log^2 p_n(t) dt - \mu_{0n}^2$ while $U_1, U_2, \ldots, U_n$ is the transformed sample as defined in Section 2. Choose the critical value for V_n which ensures exactly the size α_n. Then denote by β_n^+ the power of such a test and by $\mathcal{R}_n = \beta_n^+ - \beta_n$ the shortcoming of the test N_S with respect to V_n under the alternative F_n. The next theorem gives conditions under which $\mathcal{R}_n$ vanishes asymptotically.

THEOREM 3.5. *Assume $d(n) \to \infty$ and $d(n) = O(n^\tau \log n)$ for some $0 \leqslant \tau < 1/12$, and the sequence (P_n) of alternatives is defined as in* (3.3). *Suppose ρ_n satisfies $n^{1-2\tau}\rho_n^2/\log^4 n \to \infty$ and $n\rho_n^\vartheta \to 0$ for some $\vartheta < 3/(1+3\tau)$. Additionally, assume a, defined by* (3.3), *satisfies*

$$\frac{|s_n|^2_{d(n)} - n\rho_n^2}{\sqrt{n}\rho_n} \to 0. \tag{3.8}$$

Then for any $w \in \mathbb{R}$ we have

$$\mathcal{R}_n = \beta_n^+ - \beta_n \to 0 \quad \textit{as } n \to \infty.$$

An analogous statement also holds for the tests based on N_L.

Theorem 3.5 is proved in Section 5. The condition (3.8) means that a is "sufficiently smooth" and is obviously satisfied when a has finite expansion with respect to b (cf. (3.5)).

The optimality property obtained in the previous theorem can be stated in terms of asymptotic intermediate efficiency, i.e. Kallenberg efficiency (see Inglot [14] for the definition, conditions for existence, and the explicit formula). We omit here the precise definition and other details and refer the reader to, e.g., Inglot [14]. Applying Theorem 2.7 from Inglot [14] and Theorems 3.2 and 3.4, we get the following result which is proved in Section 5.

THEOREM 3.6. *Assume $d(n) \to \infty$ and $d(n) = O(n^\tau \log n)$ for some* $0 \leqslant \tau < 1/12$, *and the sequence (P_n) of alternatives is defined as in* (3.3) *with ρ_n satisfying the assumptions of Theorem* 3.5 *and with an arbitrary $a(t)$. Then the Kallenberg efficiency of the test based on N_S with respect to the Neyman–Pearson test for such a sequence of alternatives is equal to one. The same assertion holds for the test based on N_L.*

The statement of Theorem 3.6 reads as follows. Given a sequence of alternatives as in (3.3) satisfying the assumptions of Theorem 3.6. Consider the test N_S for the sample of size n with critical value defined by (3.7) with any $w \in \mathbb{R}$, the corresponding significance level α_n, and asymptotic power $1 - \mathbf{\Phi}(w)$. Next, consider the minimal sample size M_n ($M_n \leqslant n$) for the Neyman–Pearson test being on the same significance level α_n, which attains asymptotic power at least $1 - \mathbf{\Phi}(w)$. Then

$$\frac{M_n}{n} \to 1 \quad \text{as } n \to \infty.$$

Note that here the direction of alternative represented by $a(t)$ is arbitrary, while conditions for ρ_n give some restrictions on the convergence rate of sequences of alternatives to the null hypothesis (i.e. to some symmetric distribution F_s).

4. SIMULATION STUDY

To see how well new tests behave for finite samples, we present in this section an extensive simulation study in which we compare power behaviour of our tests with a broad variety of existing tests for different types of asymmetric distributions. For convenience, we shall write D and δ rather than D_n and δ_n omitting the subscript n. All programming work and computations were performed using `R` and `C++`. Every Monte Carlo experiment was repeated 10,000 times.

4.1. Critical values. As was said above, asymptotic critical values of our data driven tests determined in Theorem 3.1 are far from being close to the exact ones for moderate sample sizes. The same situation has occurred in the case of data driven goodness-of-fit tests. Since the distributions of our signed rank test statistics do not depend on the underlying distribution F_s, the simplest way is to find empirical critical values by MC experiment. In Table 1 we provide such critical values for several choices of D and δ.

TABLE 1. Simulated critical values of N_S and $N_L = N_L(D, \delta)$ for different sample sizes and selected values of D and $\delta,\ \alpha = 0.05$; 10,000 MC runs

δ	D		$n = 25$ d(25)=9	$n = 50$ d(50)=10	$n = 100$ d(100)=12	$n = 200$ d(200)=14	$n = 400$ d(400)=15	$n = 800$ d(800)=15
0		N_S	6.138	5.659	5.355	4.929	4.428	4.224
0.01	1	N_L	6.178	5.779	5.548	5.224	4.706	4.476
	3	N_L	6.157	5.717	5.474	5.164	4.663	4.442
0.03	1	N_L	6.343	6.085	5.997	5.814	5.443	5.113
	3	N_L	6.240	5.922	5.788	5.607	5.186	4.951
0.05	1	N_{L1}	6.571	6.461	6.514	6.441	6.312	5.986
	3	N_{L3}	6.333	6.153	6.121	6.042	5.803	5.573

From Table 1 it can be easily seen that critical values slowly approach the asymptotic ones and for a relatively large sample size $n = 800$ they are (in some cases) even 50% greater than the asymptotic one (i.e. 3.841). Such tendency is more distinct for larger δ. In the simulations below we shall use the critical values from Table 1.

Another possible way to get critical values for finite sample sizes is to find an approximation formula to the null distribution of the test statistic. Kallenberg and Ledwina [30] proposed an approximation formula to the null distribution of their data driven test statistic and proved its accuracy. A similar solution was adopted by Janic-Wróblewska and Ledwina [28] and also Janic-Wróblewska [26]. The formula in Janic-Wróblewska [26] concerns the statistic N_S and works well providing slightly smaller values in comparison with the exact ones. The reason is that the approximation formula takes into account events $\{S = 1\}$ and $\{S = 2\}$ omitting cases when $S \geqslant 3$ which occur with small but positive probability. The differences between values in Table 1 and calculated from that approximation formula do not exceed 0.3 for $n = 25$ and 0.05 for $n = 400$. However, that construction is hard to repeat for the statistic N_L since the selection rule L depends on two parameters D and δ and a resulting formula is rather complicated and much less accurate. So, we resign from presenting it here.

4.2. Tests for comparisons. As competitors of the data driven tests, described in Section 2, the following tests have been considered.

• The Behnen and Neuhaus [2] omnibus test, denoted here by BN ($S_n(a, K)$ in their notation). The authors propose the bandwidth 0.40 and the

Parzen-2 kernel K. For detailed description of the test statistic, we refer to (5.2.4), (5.2.14) and (5.2.15), p. 266, in Behnen and Neuhaus [2] (see also (10) and (11), p. 78, of their book).

• The Doksum et al. test, denoted here by DFA, introduced in Doksum et al. [6] (A_n in their notation, the weight function $\Psi_1(u, v)$, and ends of the interval $a = 0$, $b = 1$). For detailed description of the test statistic, we refer to Doksum et al. [6].

• The hybrid test, denoted here by MG and introduced by Modarres and Gastwirth [38] (Z_w in their notation). This is a two-stage testing procedure and has good power behaviour as it was shown in Modarres and Gastwirth [38]. Below, we take $p = 0.8$ and $\alpha_1 = 0.01$, $\alpha_2 = 0.0404$ as was suggested by the authors. For detailed description, we refer to Modarres and Gastwirth [38].

• The modified sign test, denoted here by CB, introduced by Cheng and Balakrishnan [4] (C_k in their notation). For detailed description of C_k we refer to Cheng and Balakrishnan [4].

Making the above choice we were aiming at taking the most representative tests from a large variety of constructions which have proved to be powerful. For example, an interesting recent construction by Thas et al. [45] leads to the test whose power behaviour is quite comparable with the hybrid test (cf. Table 3 in Thas et al. [45]). So, we have decided to include to our study only the hybrid test.

For many families of alternatives there are known locally most powerful rank tests. To see how far their powers are better than those of our new tests we considered the four simplest location families for logistic, normal, Laplace and Cauchy distributions. These families were studied in Hájek and Šidák [13]. In Figure 1 we show powers of the corresponding tests denoting them by OP (S^+ in the Hájek and Šidák [13] notation).

4.3. Alternatives. For power comparisons, we have considered a broad spectrum of alternatives. Although we have assumed that the median is equal to zero, we also want to check the ability of the considered tests to detect alternatives with non-zero median. To have a better insight into the connection between the magnitude of Fourier coefficients and sensitivity of the tests under consideration (and N_S, N_{L1}, N_{L3}, in particular) we classify alternatives into some groups using estimated Fourier coefficients with respect to the odd Legendre polynomials (cf. (2.4)):

$$\text{(4.1)} \qquad \langle 2j-1 \rangle = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{n} \sum_{i=1}^{n} b_{2j-1} \left(\frac{R_{i(k)} - 1/2}{2n} \right) = \frac{1}{N\sqrt{n}} \sum_{k=1}^{N} \widehat{b}_{2j-1(k)},$$

where $R_{i(k)}$ is the rank of the observation x_i in the kth simulated sample $x_1, \ldots, x_n$, $-x_1, \ldots, -x_n$, n is the sample size, and N is the number of Monte Carlo runs. These coefficients can be interpreted as estimated odd components (under the Legendre polynomials) of Pearson's Φ^2 measure of disparity between two distributions studied by Eubank et al. [8].

4.3.1. *Alternatives with non-zero median.* We take here some alternatives which are shifts of some typical symmetric distributions and mixtures of some typical distributions. We divide them into three groups according to a magnitude of absolute values of the components $\langle 2j-1\rangle$. Let $\beta_{(p,q)}(x)$, $p,q>0$, denote the beta density function and θ be a real parameter.

- **Alternatives with dominating component** $\langle 1\rangle$, denoted by $ALT1$:

Notation	**Density**
Logistic(θ)	$\exp(x-\theta)/\big(1+\exp(x-\theta)\big)^2,\quad x\in\mathbb{R};$
Normal (θ)	$\phi(x-\theta),\quad x\in\mathbb{R};$
Laplace (θ)	$0.5\exp(-\lvert x-\theta\rvert),\quad x\in\mathbb{R};$
Cauchy(θ)	$1/\big[\pi\big(1+(x-\theta)^2\big)\big],\quad x\in\mathbb{R},$

where ϕ denotes the standard normal density function.

- **Alternatives with dominating component** $\langle 3\rangle$, denoted by $ALT2$:

Notation	**Density**
EV(θ)	$\exp\big((x-\theta)-\exp(x-\theta)\big),\quad x\in\mathbb{R};$
C(θ)	$\theta\phi\big(\ln(x+1.2)\big)/(x+1.2)+(1-\theta)\phi(x),\ x>-1.2, \theta\in[0,1];$
LC(θ)	$0.7\phi(x-\theta/0.7)+0.3\phi(x+\theta/0.3),\quad x\in\mathbb{R}.$

- **Alternatives with 'mixed' structure of components,** denoted by $ALT3$:

Notation	**Density**
Beta(θ)	$0.3\big(\beta_{(1,2)}(x-1)+\beta_{(2,1)}(x)\big)+0.4\beta_{(1,\theta)}(x-0.5),$ $x\in[0,2], \theta>0.$

Alternatives $ALT1$ come from Eubank et al. [8], while LC(θ) was used in Fan [9].

4.3.2. *Alternatives with zero median.* We consider alternatives with mixed structure of components and we also include some families of alternatives which frequently appear in the literature.

- **Alternatives with 'mixed' structure of components,** denoted by $ALT4$:

Notation	**Density**
F(θ)	$0.5+2x\theta^{-2}(\theta-\lvert x\rvert)\mathbf{1}_{(\lvert x\rvert<\theta)},\quad x\in[-1,1], \theta\in[0,1];$
MixBeta(θ)	$0.1\big(\beta_{(1,2)}(x-1)+\beta_{(2,1)}(x)\big)+0.8\beta_{(1,\theta)}(x+2^{-1/\theta}-1),$ $x\in[0,2], \theta>0;$
Sin(θ, j)	$0.5+\theta\sin(\pi jx),\quad x\in[-1,1], \theta\in[-0.5,0.5].$

Note that F(θ) and Sin(θ, j) were also considered by Fan [9].

- **The Generalized Lambda Family** denoted by $ALT5$:

Following, e.g., Modarres and Gastwirth [38] the nine specific distributions are selected from the Generalized Lambda Family. The corresponding random variable is defined as

$$X=\frac{U^{\lambda_3}-(1-U)^{\lambda_4}}{\lambda_2}+\lambda_1, \tag{4.2}$$

where U is a uniform random variable on $[0,1]$, $\lambda_3, \lambda_4 \in \mathbb{R}, \lambda_2 \neq 0$, and in each case λ_1 is chosen in such a way that X has median zero. In each case $\langle 3 \rangle$ is the dominating component.

- **The Generalized Tukey-Lambda Family** denoted by $ALT6$:

The Generalized Tukey-Lambda Family was studied in detail by Freimer et al. [10]. The corresponding random variable is defined as

$$X = \frac{1}{\lambda_2}\left(\frac{U^{\lambda_3}-1}{\lambda_3} - \frac{(1-U)^{\lambda_4}-1}{\lambda_4}\right) + \lambda_1, \tag{4.3}$$

where U is a uniform random variable on $[0,1]$, λ_3, $\lambda_4 \in \mathbb{R}$, and in each case λ_1 is chosen in such a way that X has median zero. In all our simulations we took the scale parameter λ_2 equal to one. Thus we consider the two-parameter family denoted here by Tukey(λ_3, λ_4). The members of the family are usually classified into five categories. In simulation study we consider the following cases:

Notation	**Description**
Tukey($0.1, \lambda$)	$\lambda < 1$, class I;
Tukey($\lambda, 0.9$)	$\lambda > 1$, class II;
Tukey($7, \lambda$)	$1 < \lambda < 2$, class IV;
Tukey($4, \lambda$)	$\lambda > 2$, class V.

For this family of distributions, the third component $\langle 3 \rangle$ is usually dominating. For the cases Tukey$(0.1, \lambda)$, Tukey$(\lambda, 0.9)$, and Tukey$(4, \lambda)$, the component $\langle 5 \rangle$ is also significant. We omit class III of distributions corresponding to $\lambda_3,\ \lambda_4 \in (1,2)$ for which all the compared tests behave similarly to those in class I.

- **The Fechner Family** denoted by $ALT7$:

A simple family of asymmetric densities, proposed more than a century ago by Fechner, was recently revived by Arellano-Valle, Gómez, and Quintana (see Cassart et al. [3]). Denote by f a symmetric density. The f-Fechner family is the three-parameter collection of densities of the form

$$\frac{1}{\sigma}\left[f\left(\frac{x-\theta}{(1+\xi)\sigma}\right)\mathbf{1}_{(-\infty,\theta]}(x) + f\left(\frac{x-\theta}{(1-\xi)\sigma}\right)\mathbf{1}_{(\theta,\infty)}(x)\right], \quad x \in \mathbb{R},$$

with the location parameter (median) $\theta \in \mathbb{R}$, the scale parameter $\sigma \in (0,+\infty)$, and the skewness parameter $\xi \in (-1,1)$. So, we take in our simulations $\theta = 0$ and $\sigma = 1$. We consider two families: the Normal-Fechner Family and the Cauchy–Fechner Family. For the Normal-Fechner Family $\langle 3 \rangle$ is the dominating coefficient, while for the Cauchy–Fechner Family also $\langle 1 \rangle$ and $\langle 5 \rangle$ are significant.

Notation	**Description**
Normal-Fechner(ξ)	$f(x) = (\sqrt{2\pi})^{-1}\exp(-x^2/2)$;
Cauchy-Fechner(ξ)	$f(x) = \pi^{-1}(x^2+1)^{-1}$.

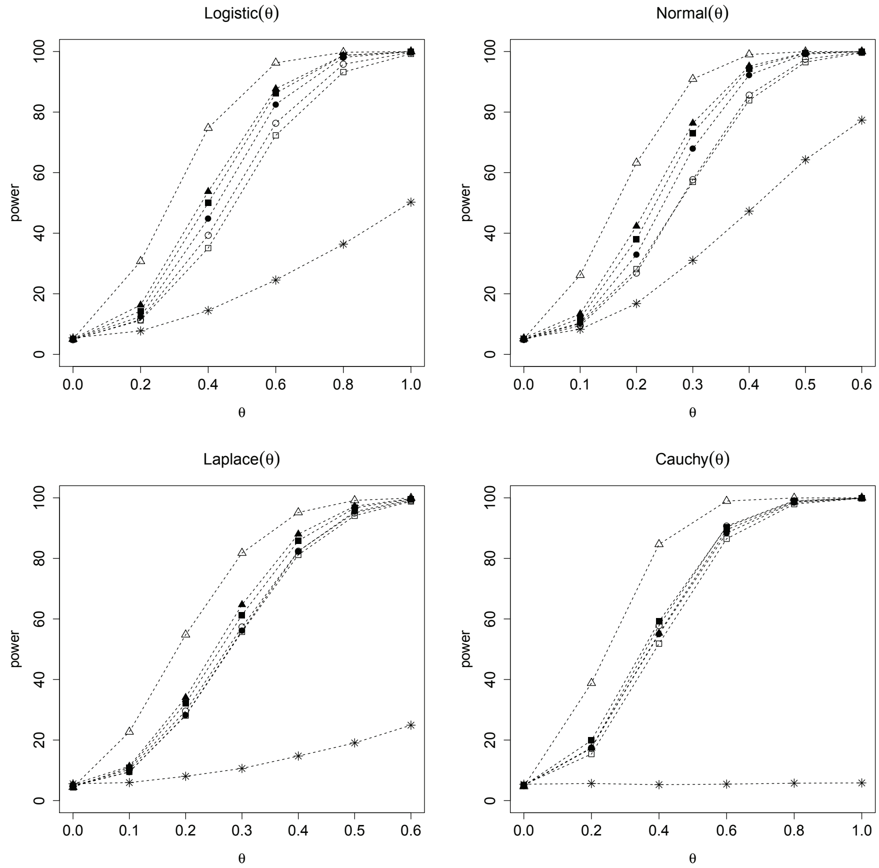

FIGURE 1. Alternatives $ALT1$ with the dominating component $\langle 1\rangle$. Comparison of empirical powers (in %) of NS (-■-), $NL3$ (-●-), BN (-○-), DFA (-▲-), MG (-□-), CB (-∗-), and OP (-△-). $n = 100$, $\alpha = 0.05$, $d(100) = 12$; 10,000 MC runs

4.3.3. *Orthogonal alternatives.* We consider the set of 24 alternatives defined by the exponential family $f_k(x, \vartheta)$ given by (2.1) by taking $k = 12$ and $\vartheta = \pm 0.25e_j$, $j = 1, \ldots, 12$, where $e_1, \ldots, e_{12}$ is the standard basis in the Euclidean space $\mathbb{R}^{12}$. Namely, we put

$$g^{+}_{12,j}(x) = f_{12}\left(\frac{x+1}{2}, 0.25e_j\right), \quad g^{-}_{12,j}(x) = f_{12}\left(\frac{x+1}{2}, -0.25e_j\right), \qquad x \in [-1, 1],$$

$j = 1, \ldots, 12$, respectively. So, we disturb the (symmetric) uniform distribution on $[-1, 1]$ nearly on the one (antisymmetric) axis.

4.4. Power comparisons. In the consecutive points, we present the power comparisons for alternatives described in Subsection 4.3. For brevity, we focus on the case $n = 100$ (except Table 2 in which we took $n = 50$ to get reasonable powers

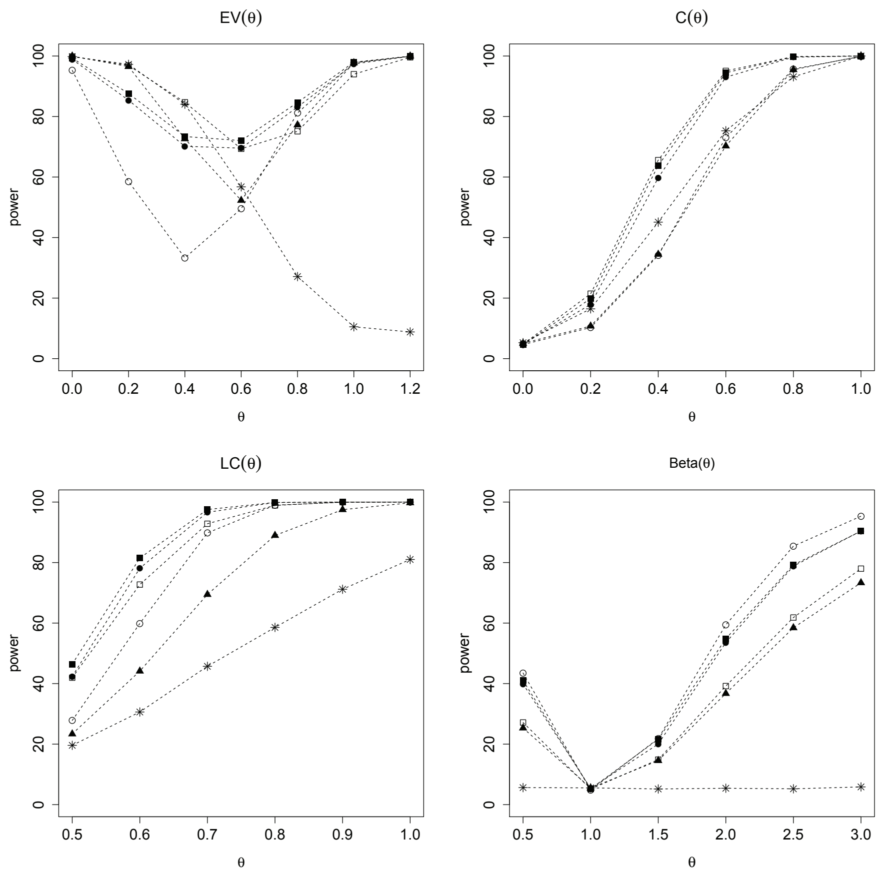

FIGURE 2. Alternatives $ALT2$ with the dominating component $\langle 3 \rangle$ and alternatives $ALT3$ with 'mixed' structure of components. Comparison of empirical powers (in %) of NS (-■-), $NL3$ (-●-), BN (-○-), DFA (-▲-), MG (-□-), and CB (-∗-). $n = 100$, $\alpha = 0.05$, $d(100) = 12$; 10,000 MC runs

of the compared tests). In our figures, we show two data driven tests N_S and N_{L3}. Only in Table 3 N_{L1} is also present. For notational convenience, we shall denote them by NS, $NL3$, and $NL1$, respectively.

4.4.1. *Power comparisons for alternatives with non-zero median.* The power behaviour of the considered tests for the alternatives described in Subsection 4.3.1 is reported in Figures 1 and 2.

We can observe that for alternatives from group $ALT1$ (see Figure 1) almost all the considered tests work very well. Only the test CB performs poor behaviour. In particular, for the Cauchy distribution (Cauchy(θ)), the test CB attains power at the significance level. A similar situation occurs for the Laplace distribution (Laplace(θ)). As could be expected, the test OP is optimal for detecting location

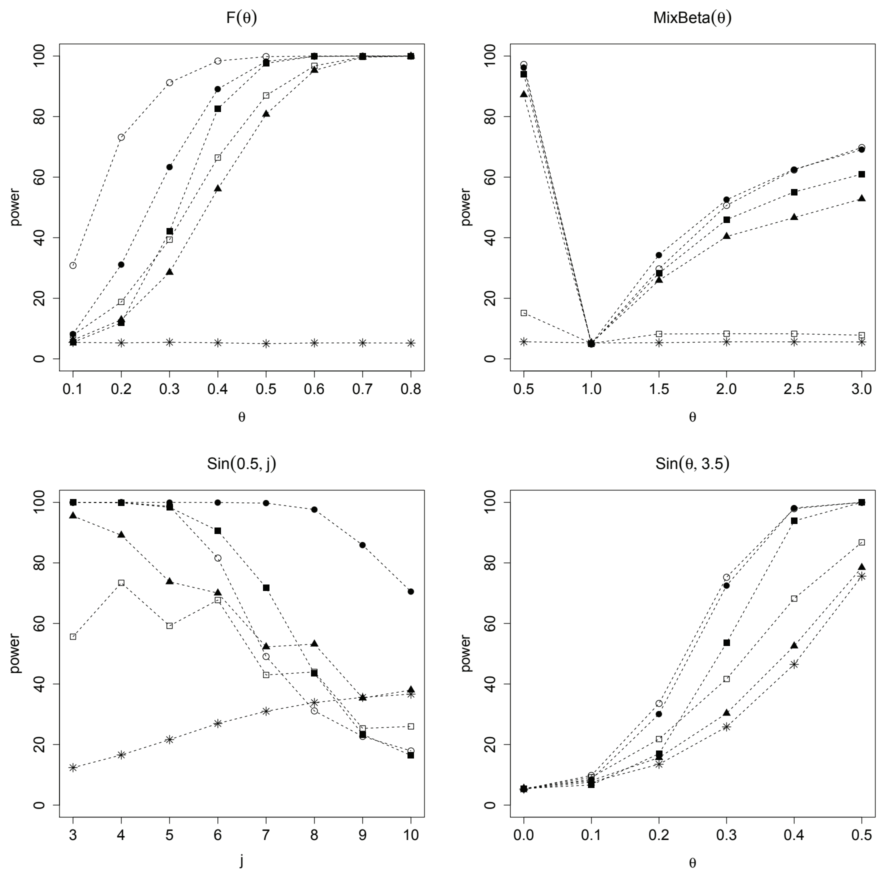

FIGURE 3. Alternatives $ALT4$ with 'mixed' structure of components. Comparison of empirical powers (in %) of NS (-■-), $NL3$ (-●-), BN (-○-), DFA (-▲-), MG (-□-), and CB (-∗-). $n = 100, \alpha = 0.05, d(100) = 12$; 10,000 MC runs

shifts in the known distribution. In particular, the Wilcoxon test is the best for logistic family (Logistic(θ)), the sign test for Laplace Family, etc. However, the loss in power for the data driven tests (NS and $NL3$) with respect to OP is at most ca. 30%. On the other hand, these linear rank tests are not able to detect alternatives when the component $\langle 1 \rangle$ is negligible. For alternatives $ALT2$–$ALT7$ these tests completely break down, and therefore the results of power simulations are not reported here. Moreover, note that the hybrid test MG is distinctly worse than NS and $NL3$ for alternatives from the group $ALT1$.

It can be seen from Figure 2 that for alternatives $ALT2$ and $ALT3$ the data driven tests provide superior power. Contrary to BN and CB, the tests NS and $NL3$ do not have weak points, and DFA and MG are often worse than NS and $NL3$. Moreover, NS is a little bit better than $NL3$. Observe that CB is not able to detect $ALT3$ at all. The reason is that the asymmetry of these alternatives has nothing to do with the tails of a distribution.

TABLE 2. Alternatives $ALT5$ from the Generalized Lambda Family.
Comparison of empirical powers (in %) of NS, $NL3$, BN, DFA, MG, and CB.
$n = 50$, $\alpha = 0.05$, $d(50) = 10$; 10,000 MC runs

λ_2	λ_3	λ_4	NS	$NL3$	BN	DFA	MG	CB
Case 1: 0.197454	0.134915	0.134915	5	5	5	5	4	5
Case 2: 1.0	1.4	0.25	83	80	49	81	94	95
Case 3: 1.0	0.00007	0.1	96	95	77	95	99	99
Case 4: 0.04306	0.025213	0.094029	39	35	16	24	49	43
Case 5: -1.0	-0.0075	-0.03	56	52	25	37	67	59
Case 6: -0.351663	-0.13	-0.16	7	7	6	5	6	7
Case 7: -1.0	-0.1	-0.18	19	17	10	11	22	19
Case 8: -1.0	-0.001	-0.13	99	98	88	98	100	100
Case 9: -1.0	-0.00001	-0.17	99	99	91	98	100	100

4.4.2. *Power comparisons for alternatives with zero median.* The power behaviour of the considered tests for the alternatives described in Subsection 4.3.2 is reported in Figures 3, 4, and 5 and Table 2.

In Figure 3, one can observe that the tests CB and MG provide no protection against alternatives with asymmetry in the "middle" part of a distribution when simultaneously the median equals zero (see, for example, MixBeta(θ)). The tests $NL3$, NS, and BN perform the best behaviour for alternatives $ALT4$. It turns out that for detecting "high-frequency" departures (see $\text{Sin}(0.5, j)$) $NL3$ is the best test. In this case, the loss in power for NS and for other tests is even equal to 50%.

Table 2 presents power behaviour of all the compared tests for nine cases from the Generalized Lambda Family, which are frequently studied in the literature (see, e.g., Modarres and Gastwirth [38] and also Cheng and Balakrishnan [4]). Exceptionally, in this case we took $n = 50$ to ensure reasonable powers. It is known that the alternatives $ALT5$, considered in Table 2, have asymmetry in tails of the distribution (the component $\langle 3 \rangle$ is dominating). This explains why CB and MG detect $ALT5$ with very high powers. But the data driven tests lose at most ca. 15% with respect to MG.

In Figure 4, we report powers of the six compared tests for alternatives $ALT6$ taken from the Generalized Tukey-Lambda Family. In this case the results are quite similar to those for alternatives $ALT5$ but here the data driven tests lose at most 40% with respect to (the most powerful) CB.

Finally, Figure 5 concerns the Fechner Family. Observe that for heavy tailed distribution (Cauchy-Fechner(ξ)) the data driven tests are much better than CB and slightly dominate MG although the asymmetry in tails is essential (but it seems the "middle" part is more significant).

4.4.3. *Average powers.* The aim of this section is to repeat an analogous experiment to that performed in Inglot and Janic [17]. We consider a finite set of nearly orthogonal alternatives $g^{+}_{12,j}(x)$, $g^{-}_{12,j}(x)$, $j = 1, \ldots, 12$, defined in Subsection 4.3.3, and look at average powers of all the considered tests in comparison to the

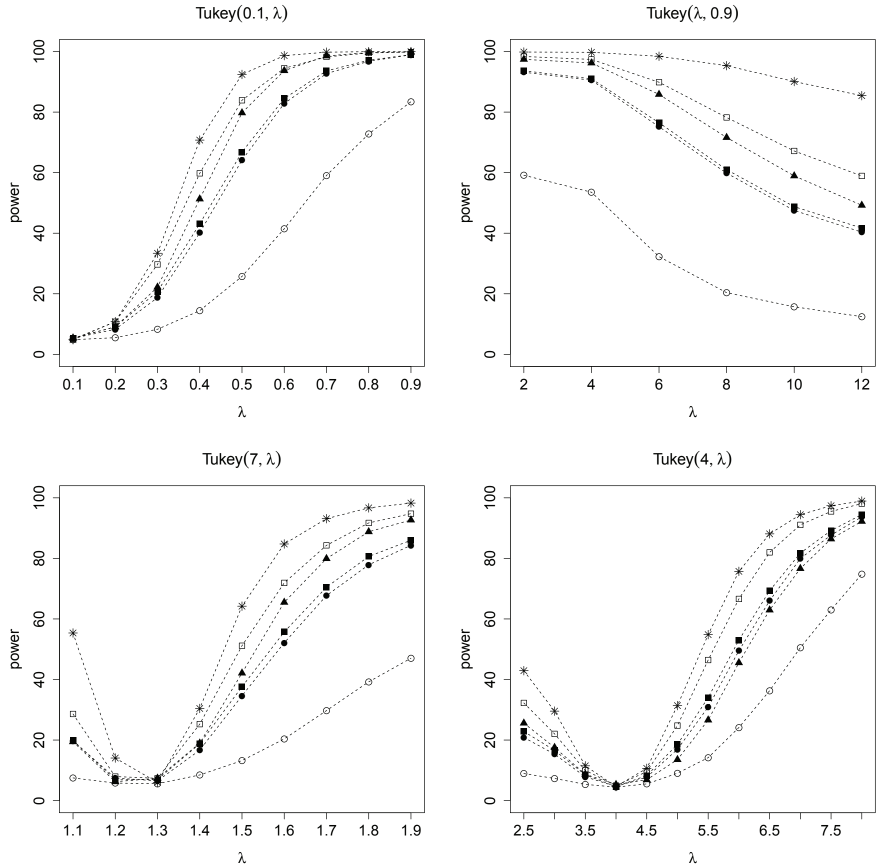

FIGURE 4. Alternatives $ALT6$ from the Generalized Tukey-Lambda Family. Comparison of empirical powers (in %) of NS (-■-), $NL3$ (-●-), BN (-○-), DFA (-▲-), MG (-□-), and CB (-∗-). $n = 100$, $\alpha = 0.05$, $d(100) = 12$; 10,000 MC runs

Neyman–Pearson test and the optimal Bayes test. Inglot and Janic [17] showed that, in the middle range of the power function, the loss in average power of the optimal Bayes test with respect to the Neyman–Pearson test can be measured by the Shannon entropy of a prior distribution.

Let T^* be the two-sided optimal Bayes test given as in Inglot and Janic [17] (cf. (A.2) in their Appendix), defined by 24 densities $g^+_{12,j}(x)$, $g^-_{12,j}(x)$, $j = 1, \ldots, 12$, under the uniform prior distribution on them. In Table 3 we present powers of all the six compared tests, $NL1$ and T^* for alternatives $g^+_{12,j}(x)$, $j = 1, \ldots, 12$. The powers for alternatives $g^-_{12,j}(x)$ are practically the same as that for $g^+_{12,j}(x)$, so we omit them in Table 3. To show the whole picture, we also include the one-sided Neyman–Pearson test denoted by NP (constructed for each alternative separately) and the score test based on $N_{12} = |\hat{b}|^2_{12}$. In the last row of Table 3

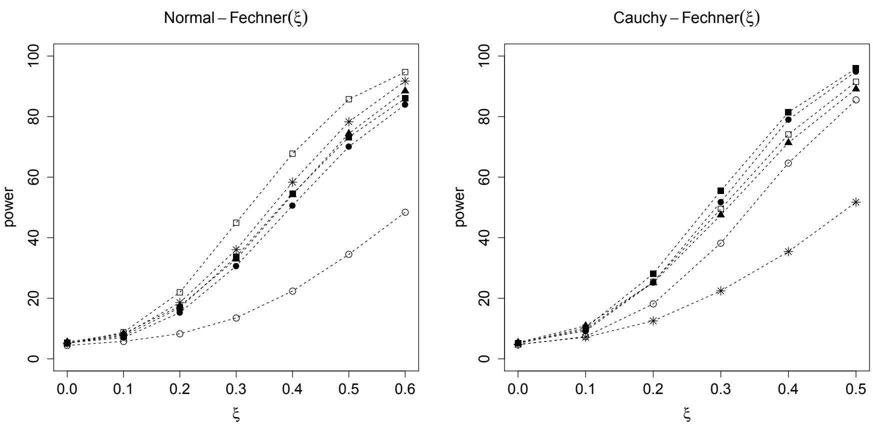

FIGURE 5. Alternatives $ALT7$ from the Fechner Family: the Normal-Fechner Family and the Cauchy–Fechner Family. Comparison of empirical powers NS (-■-), $NL3$ (-●-), BN (-○-), DFA (-▲-), MG (-□-), and CB (-∗-). $n = 100$, $\alpha = 0.05$, $d(100) = 12$; 10,000 MC runs

TABLE 3. Comparison of powers and average powers (in %) of NP, T^*, NS, $NL1$, $NL3$, BN, DFA, MG, CB, and N_{12}. $n = 100$, $\alpha = 0.05$, $d(n) = 12$; 10,000 MC runs, alternatives $g^+_{12,j}(x)$, uniform prior

ϑ	NP	T^*	NS	$NL1$	$NL3$	BN	DFA	MG	CB	N_{12}	$\langle 2j-1 \rangle$
$0.25e_1$	80	40	58	48	52	53	60	41	15	28	0.24
$0.25e_2$	80	40	60	53	56	45	25	54	22	27	0.24
$0.25e_3$	81	41	33	35	34	30	13	12	20	27	0.23
$0.25e_4$	80	40	16	30	24	21	9	8	13	27	0.22
$0.25e_5$	81	41	9	28	22	15	8	7	9	26	0.21
$0.25e_6$	79	39	6	23	20	11	6	5	7	26	0.19
$0.25e_7$	81	41	6	20	18	9	6	6	6	26	0.17
$0.25e_8$	80	39	5	16	16	7	6	6	6	25	0.15
$0.25e_9$	81	41	5	14	14	7	6	5	6	24	0.13
$0.25e_{10}$	81	40	5	12	12	6	6	6	6	22	0.11
$0.25e_{11}$	81	42	5	9	10	6	5	5	6	20	0.09
$0.25e_{12}$	82	43	6	8	9	6	6	6	7	15	0.07
average power	80.6	40.6	17.8	24.7	23.9	18.0	13.0	13.4	10.3	24.4	

we give average powers over 12 displayed alternatives. Note that average powers over all 24 alternatives differ from the presented ones at most 0.1%. Although N_{12} attains a similar average power to $NL3$, this last test keeps much higher power for "smooth" alternatives, and therefore better competes with other tests. Note that the power of NP and T^* tests is almost constant for each such artificially selected alternative and the loss in average power for T^* with respect to the case when full information about alternative is available equals ca. 40%. This agrees quite well with the approximation derived in Theorem A.2 of Inglot and Janic [17]. On the other hand, the loss in average power for $NL3$ and $NL1$ with respect to T^* in the middle range of the power function is about 16%. Moreover, as could be expected, the test $NL1$ attains a little bit better average power than $NL3$ (ca. 0.8%) and much better than NS (ca. 7%). DFA, MG, and CB behave like typical directional tests and break down beginning from the sixth alternative. BN looks much better and can compete with NS although being inferior to NS for the first three ("smooth") alternatives.

4.5. Conclusions. The results presented in Subsection 4.4 show that the newly introduced data driven tests perform well for a wide range of alternatives and are able to compete with other commonly used tests as well as recent constructions. Simulations show that the new test $NL3$ has high and stable power in situations when the other tests break down and simultaneously has comparable power under alternatives which can be detected by those tests. This observation is also confirmed by the results shown in Table 3. Therefore, we recommend $NL3$ as an omnibus test having the widest spectrum of sensitivity for detecting asymmetry. To make this conclusion more transparent we selected one specific alternative from each figure, took powers for every of the six compared tests for these alternatives (for which powers attained are in the middle range) and calculated average powers over all 18 cases. We obtained (in %) 57.5 for $NL3$, 56.2 for NS, 51.9 for MG, 48.5 for DFA, 44.9 for BN, and 36.9 for CB. This calculation additionally illustrates the above conclusion on the ability of data driven tests (and $NL3$ in particular) to cover the widest spectrum of asymmetric distributions preserving high power in each case.

5. PROOFS

In this section we provide proofs of all theorems stated in Section 3. They are based on auxiliary results presented in the Appendix. In the sequel we shall use letters C, c to denote positive constants possibly different in each case.

In our case we specify a system of functions Φ taken arbitrarily in the Appendix to be the vector b of the odd Legendre polynomials (cf. (2.4) and (A.3)). Hence a Gaussian vector γ defined in (A.6) is a standard Gaussian sequence and, consequently, $\Gamma = I$ is the identity matrix, and λ_n defined in Proposition A.3 are equal to one for all n. Using well-known properties of the Legendre polynomials we easily see that the coefficient $\psi(k)$ defined in (A.8) can be estimated by $\psi^2(k) \leqslant Ck^3 \log^2 k$ for every k.

Observe that the correction $1/(4n)$ we have inserted in (2.4) has no influence on asymptotic behaviour of our statistics. Indeed, if we set $\tilde{b}_{2j-1} = (1/\sqrt{n}) \sum_{i=1}^{n} b_{2j-1}\big(\mathcal{F}_{ns}(X_i)\big)$ and $\tilde{b} = (\tilde{b}_1, \tilde{b}_3, \ldots)$, then applying again properties of the Legendre polynomials we obtain, by a standard calculation,

$$\big||\hat{b}|_k - |\tilde{b}|_k\big| \leqslant C(k^{3/2} \log n)/\sqrt{n}.$$

This estimate allows us to omit the correction in proofs of all theorems. Therefore, throughout this section we shall use the notation $\hat{b}_{2j-1}$ and $\hat{b}$ for empirical Fourier coefficients without the correction $1/(4n)$. Then the vector of empirical Fourier coefficients $\hat{b}$ coincides with $\hat{\Phi}$ as in (A.2).

P r o o f o f T h e o r e m 3.1. Taking in (A.12) $x_n = (\log^2 n)/\sqrt{n}$ and $k(n) = 1$ for all n and applying the triangle inequality, we get

$$N_1 = |\hat{b}|_1^2 \xrightarrow{\mathcal{D}} |\gamma|_1^2 \stackrel{\mathcal{D}}{=} \chi_1^2.$$

Now, let $d(n) = O(n^\tau)$ for some $\tau < 1/6$ be as assumed in Theorem 3.1. Then

$$\begin{aligned} P(S \geqslant 2) &= \sum_{k=2}^{d(n)} P(S = k) \leqslant \sum_{k=2}^{d(n)} P\big(N_k \geqslant (k-1)\log n\big) \\ &= \sum_{k=2}^{d(n)} P\big(|\hat{b}|_k^2 \geqslant (k-1)\log n\big). \end{aligned} \tag{5.1}$$

For each fixed $k \geqslant 2$ put in Proposition A.3 $x_n^2 = [(k-1)\log n]/n$, $k(n) = k$, and take some $\nu < (1-4\tau)/(2-2\tau)$. Then (A.13) is satisfied and, consequently, by Proposition A.3 we have for n sufficiently large

$$\begin{aligned} &P\big(|\hat{b}|_k^2 \geqslant (k-1)\log n\big) \\ &\quad \leqslant \exp\left\{-\frac{1}{2}(k-1)(\log n)\big(1+o(1)\big)\right\} + C\exp\{-c\log^2 n\} \leqslant n^{-1/3}. \end{aligned}$$

Combining the above estimation with (5.1) we obtain $P(S \geqslant 2) \leqslant d(n)/n^{1/3}$, which tends to zero. This proves the assertion (1). Note that here we have needed only that $\tau < 1/4$.

To prove the assertion (2) let $\mathcal{E}_j$ be the family of all j-element subsets of $\{1, 2, \ldots, d(n)\}$. Then

$$P(W_n) \leqslant \sum_{j=1}^{D} P(\hat{b}^2_{(d(n)-j+1)} \geqslant c_{jn}^2) \leqslant \sum_{j=1}^{D} \sum_{E \in \mathcal{E}_j} P(|\hat{b}|_E^2 \geqslant j c_{jn}^2), \tag{5.2}$$

where $|v|_E^2 = \sum_{i \in E} v_i^2$ for a finite subset E of $\{1, 2, \ldots\}$. Arguing as in the proof of Proposition 1 in Inglot and Janic [17] and using the assumption $\big(\log\log d(n)\big)/\log(1/\delta_n) \to 0$ we get

$$j c_{jn}^2 \leqslant 2D \log\big(2Dd(n)/\delta_n\big) \quad \text{and} \quad j c_{jn}^2 \geqslant \log(1/\delta_n) + 2\log\binom{d(n)}{j}$$

for n sufficiently large. Since $\log(1/\delta_n) = o(n)$, we see that for every j and every $E \in \mathcal{E}_j$ we have $jc_{jn}^2 = o(n)$ and $c_{jn}^2 \to \infty$. In Proposition A.3 set $x_n^2 = jc_{jn}^2/n$ and $\nu = (1-6\tau)/3$. Then (A.13) is fulfilled. Hence, for n sufficiently large

$$P(|\hat{b}|_E^2 \geqslant jc_{jn}^2) \leqslant \exp\left\{-\frac{1}{2}\log\frac{1}{\delta_n} - \log\binom{d(n)}{j} + r_n\right\} + C\exp\{-cn^{\varepsilon}\}$$

with any positive $\varepsilon < \nu$, where the remainder r_n, due to the assumptions on $d(n)$ and δ_n, can be estimated by $C\log\log\big(d(n)/\delta_n\big) + Cn^{-\nu/2}\log^{1+\nu/2}(1/\delta_n)$ not depending on j and a particular set E. Again, by the assumptions on δ_n we have $r_n = o\big(\log(1/\delta_n)\big)$. Inserting this estimate into (5.2) we infer that $P(W_n) \to 0$, which completes the proof of the assertion (2). ■

P r o o f o f T h e o r e m 3.2. Given $d(n)$ with corresponding $\tau \in [0, 1/12)$ let x_n be any sequence such that $nx_n^{\vartheta} \to 0$ for some $\vartheta < 3/(1+3\tau)$ and $n^{1-2\tau}x_n^2/\log^4 n \to \infty$. To check the assumptions of Theorem A.2 let us set $\nu = \vartheta/2 - 1$ and $k(n) = d(n)$. Then

$$\frac{nx_n^2}{d(n)} = \frac{n^{1-2\tau}x_n^2}{\log^4 n}\,\frac{n^{\tau}\log n}{d(n)}\,(n^{\tau}\log^3 n) \to \infty.$$

Since $\eta = 3 - 6\tau - \vartheta(1-2\tau) > 0$ and $\psi^2\big(d(n)\big) = O\big(d^3(n)\log^2 n\big)$, we have

$$\frac{n^3x_n^{4+4\nu}}{\psi^4\big(d(n)\big)\log^6 n} \geqslant C\left(\frac{n^{1-2\tau}x_n^2}{\log^4 n}\right)^{\vartheta}\left(\frac{n^{\tau}\log n}{d(n)}\right)^6 \frac{n^{\eta}}{\log^{16-4\vartheta} n} \to \infty$$

and similarly, observing that $\zeta = (6-2\vartheta)/\vartheta - 6\tau > 0$, we get

$$x_n^{2-4\nu}\psi^4\big(d(n)\big) \leqslant C(nx_n^{\vartheta})^{(6-2\vartheta)/\vartheta}\left(\frac{d(n)}{n^{\tau}\log n}\right)^6 \frac{\log^{10} n}{n^{\zeta}} \to 0.$$

Hence (A.14) of Theorem A.2 can be applied, thus proving the first expansion in (3.2). Due to the assumptions on x_n the remainder terms in this expansion are of order $o(\sqrt{n}x_n)$. This proves our theorem. ■

P r o o f o f T h e o r e m 3.3. Let P denote the distribution of the sample $X_1, \ldots, X_n$ with distribution function $F(x)$. Since F is asymmetric, we have $F \neq F_s$ or, equivalently, $A \neq 0$. This implies $\int_0^1 b(t)dA(t) \neq 0$. Write $k_0 = \min\big\{k \geqslant 1 : \int_0^1 b_{2k-1}(t)dA(t) \neq 0\big\}$. Further let us put

$$\mathcal{S}_{nA} = \int_0^1 b\big(\mathcal{H}_{ns}(t)\big)dA(t)$$

and (cf. (3.5))

$$s_A = \int_0^1 b(t)dA(t).$$

We apply Proposition A.5 to $x_n = (\log^2 n)/\sqrt{n}$ and fixed $k(n) = 1, \ldots, k_0$. Then from (A.21) and (A.22) we get

$$|\hat{b} - \sqrt{n}\mathcal{S}_{nA} - \hat{b}_0|_j \xrightarrow{P} 0, \quad |\mathcal{S}_{nA} - s_A|_j \xrightarrow{P} 0,$$

where (cf. (A.5))

$$\hat{b}_0 = \int_0^1 b(t)d\varepsilon_n(t) = -\int_0^1 b'(t)\varepsilon_n(t)dt \xrightarrow{\mathcal{D}} -\int_0^1 b'(t)B\big(t + A(t)\big)dt,$$

while $\varepsilon_n(t)$ denotes the transformed empirical process as defined in the Appendix. In consequence, $|\sqrt{n}\mathcal{S}_{nA}|_{k_0} = C\sqrt{n}\big(1 + o_P(1)\big)$ and

$$|\hat{b}|_{k_0} \geqslant |\sqrt{n}\mathcal{S}_{nA}|_{k_0} - |\hat{b} - \sqrt{n}\mathcal{S}_{nA} - \hat{b}_0|_{k_0} - |\hat{b}_0|_{k_0} = C\sqrt{n}\big(1 + o_P(1)\big).$$

On the other hand, we have $|s_A|_j = 0$ for $j < k_0$, which results in $|\hat{b}|_j = o_P(\sqrt{n})$. Now, for the selection rule S we have

$$P(S < k_0) \leqslant \sum_{j=1}^{k_0-1} P\big(|\hat{b}|^2_{k_0} \leqslant |\hat{b}|^2_j + (k_0 - j)\log n\big) \to 0.$$

Hence, $P(S \geqslant k_0) \to 1$, $P(N_S \geqslant |\hat{b}|^2_{k_0}) \to 1$ and $N_S \xrightarrow{P} \infty$. An application of Theorem 3.1 (boundedness of a critical value) completes the proof. ■

P r o o f o f T h e o r e m 3.4. Given $d(n)$ with the corresponding τ. Let ρ_n be any sequence satisfying the assumptions of Theorem 3.4. We apply Theorem A.3 to $k(n) = d(n)$, $A_n = A$ for all n and $\delta_n = (\log n)/\sqrt{n\rho_n^\kappa}$ with $\kappa = (1 + \tau) \wedge (3/2 - 6\tau)$. Then $\zeta = (1 - 2\tau)(1 + \kappa) \leqslant 2$ and, consequently,

$$\frac{\rho_n}{\delta_n^2} = n^{1-\zeta/2} \log^{2\kappa} n \left(\frac{n^{1-2\tau}\rho_n^2}{\log^4 n}\right)^{(1+\kappa)/2} \to \infty.$$

Obviously, $n\rho_n\delta_n^2/(\log n) = (\log n)/\rho_n^{\kappa-1} \to \infty$ and (A.15) is fulfilled. So, by (A.20) we have for versions of $\hat{b}$ and γ on a common probability space $(\Omega, \mathcal{B}, \mathbf{P})$, defined by the KMT inequality (cf. (A.7) in the Appendix),

$$\mathbf{P}\big(|\hat{b}^{(1)} - s_n - \gamma_n^{(1)}|_{d(n)} \geqslant C\rho_n^{1-\kappa/2} d^{3/2}(n) \log^2 n\big) \leqslant C_1 \exp\{-c \log n\} \to 0.$$

Since $12\tau - 2 + \kappa < 0$, the assumption on $d(n)$ implies

$$\rho_n^{2-\kappa} d^3(n) \log^4 n \leqslant C\rho_n^{2-\kappa} n^{3\tau} \log^7 n = C(n\rho_n^4)^{(2-\kappa)/4} n^{3\tau-(2-\kappa)/4} \log^7 n \to 0.$$

Hence we have proved

$$|\hat{b}^{(1)} - s_n - \gamma_n^{(1)}|_{d(n)} \xrightarrow{\mathbf{P}} 0. \tag{5.3}$$

Since a is normalized, the assumption $d(n) \to \infty$ and the Parceval identity give

$$|s_n|^2_{d(n)}/n\rho_n^2 \to 1, \tag{5.4}$$

which immediately implies $|s_n|_{d(n)} \to \infty$. Thus

$$\begin{aligned} \frac{|\hat{b}|^2_{d(n)} - |s_n|^2_{d(n)}}{2|s_n|_{d(n)}} &\overset{D}{=} \frac{|\hat{b}^{(1)}|^2_{d(n)} - |s_n|^2_{d(n)}}{2|s_n|_{d(n)}} \\ = \frac{|\hat{b}^{(1)} - s_n|^2_{d(n)}}{2|s_n|_{d(n)}} &+ \frac{s_n}{|s_n|_{d(n)}} \circ_{d(n)} (\hat{b}^{(1)} - s_n - \gamma_n^{(1)}) + \frac{s_n}{|s_n|_{d(n)}} \circ_{d(n)} \gamma_n^{(1)}, \end{aligned} \tag{5.5}$$

where $\circ_{d(n)}$ denotes the usual Euclidean scalar multiplication in $\mathbb{R}^{d(n)}$. From (5.3) it follows that the first two terms on the right-hand side of (5.5) converge in probability to zero while the third one has the standard normal distribution for every n (we use the fact that for every n, $d(n)$-dimensional truncation of $s_n/|s_n|_{d(n)}$ is a point on the unit sphere in $\mathbb{R}^{d(n)}$ and $\gamma_n^{(1)}$ is a standard Gaussian sequence). Thus we have proved

$$\frac{|\hat{b}|^2_{d(n)} - |s_n|^2_{d(n)}}{2|s_n|_{d(n)}} \xrightarrow{D} N(0,1). \tag{5.6}$$

Now, let us consider a deterministic counterpart of the Schwarz type rule S defined by

$$l(n) = \min\{1 \leqslant k \leqslant d(n) : |s_n|^2_k - |s_n|^2_j \geqslant \mu(k-j)\log n, 1 \leqslant j \leqslant d(n)\}, \tag{5.7}$$

where $\mu > 1$ is chosen arbitrarily. We shall show that $P\big(S < l(n)\big) \to 0$ in a similar way to that of Inglot and Ledwina [23]. To this end let us write $|v|^2_{kl} = |v|^2_l - |v|^2_k$ for a vector v and $k < l$. For $k < l(n) = l$, from the triangle inequality we have

$$\begin{aligned} P\big(|\hat{b}|^2_{kl} \leqslant (l-k)\log n\big) &\leqslant P\big(|\hat{b} - s_n|_{kl} \geqslant |s_n|_{kl} - \sqrt{(l-k)\log n}\big) \\ &\leqslant P\big(|\hat{b} - s_n|_{kl} \geqslant (\sqrt{\mu} - 1)\sqrt{(l-k)\log n}\big) \\ &= \mathbf{P}\big(|\hat{b^{(1)}} - s_n|_{kl} \geqslant (\sqrt{\mu} - 1)\sqrt{(l-k)\log n}\big) \\ &\leqslant \mathbf{P}\big(|\hat{b^{(1)}} - s_n - \gamma_n^{(1)}|_{kl} \geqslant (1/2)(\sqrt{\mu} - 1)\sqrt{(l-k)\log n}\big) \\ &\quad + P\big(|\gamma|_{kl} \geqslant (1/2)(\sqrt{\mu} - 1)\sqrt{(l-k)\log n}\big). \end{aligned} \tag{5.8}$$

Set $\delta_n = (\log n)/\sqrt{n\rho_n}$. Then $\delta_n \to 0$ and

$$\begin{aligned}\psi^2(l-k)n\rho_n^2\delta_n^2 &\leqslant Cd^3(n)\rho_n \log^4 n \leqslant Cn^{3\tau}\rho_n \log^7 n\\ &= (n\rho_n^4)^{1/4} n^{(12\tau-1)/4}\log^7 n \to 0.\end{aligned}$$

Moreover, δ_n satisfies (A.15) and $\delta_n = O(\sqrt{\rho_n})$. So, applying again Theorem A.3 we estimate the first term in (5.8) by $C\exp\{-c\log^2 n\}$. By the classical bound for tails of a norm of a Gaussian vector, the second term in (5.8) is estimated by $\exp\{-c(l-k)\log n\}$. Using these estimates and (5.7) we get

$$\begin{aligned}P\big(S<l(n)\big) &\leqslant \sum_{k=1}^{l-1} P(|\hat{b}|_k^2 - k\log n \geqslant |\hat{b}|_l^2 - l\log n)\\ &= \sum_{k=1}^{l-1} P\big(|\hat{b}|_{kl}^2 \leqslant (l-k)\log n\big)\\ &\leqslant Cd(n)\exp\{-c\log^2 n\} + \sum_{k=1}^{l-1} n^{-c(l-k)} \to 0\end{aligned} \tag{5.9}$$

as the last term on the right-hand side is smaller than $1/(n^c-1)$.

It is easily seen that the proof of (5.3) can be repeated with $d(n)$ replaced by $l(n)$. Moreover, by (5.4), (5.7), and the assumptions on $d(n)$ and ρ_n, we have for sufficiently large n (cf. (3.5))

$$\frac{|s_n|_{l(n)d(n)}^2}{|s_n|_{d(n)}^2} \leqslant \frac{|s_n|_{d(n)}^2 - |s_n|_{l(n)}^2}{|s_n|_{d(n)}} \leqslant \frac{\mu\big(d(n)-l(n)\big)\log n}{|s_n|_{d(n)}} \leqslant C\frac{n^\tau \log^2 n}{\sqrt{n}\rho_n} \to 0.$$

Hence $|s_n|_{l(n)}/|s_n|_{d(n)} \to 1$ and $|s_n|_{l(n)} \to \infty$. This implies that (5.6) holds for $d(n)$ replaced by $l(n)$ and, consequently,

$$\frac{|\hat{b}|_{l(n)}^2 - |s_n|_{d(n)}^2}{2|s_n|_{d(n)}} \xrightarrow{D} N(0,1). \tag{5.10}$$

The relation $P(|\hat{b}|_{l(n)}^2 \leqslant N_S) \to 1$ and the obvious inequality $N_S \leqslant N_L \leqslant |\hat{b}|_{d(n)}^2$ combined with (5.6) and (5.10) complete the proof. ∎

REMARK 5.1. *An inspection of the proof of* (5.3) *and* (5.9) *shows that if* $\tau = 0$, *then* (3.6) *remains valid if*

$$\rho_n \log^7 n \to 0 \quad \textit{and} \quad \frac{n\rho_n^2}{\log^4 n} \to \infty,$$

thus covering almost the whole range of convergent alternatives between contiguous and fixed alternatives.

P r o o f o f T h e o r e m 3.5. The proof of Theorem 3.5 goes along the same line of argument as in Inglot et al. [20], Inglot and Ledwina [24] or Ducharme and Ledwina [7]. We provide it here to make the paper self-contained.

To make the notation precise, throughout this proof and the proof of Theorem 3.6, we shall write P_0 for the null distribution corresponding to some $F = F_s$ and P_n for the alternative distribution corresponding to the density $1 + \rho_n a(t)$ of the data transformed by F_s (cf. (3.3)).

Let $w \in \mathbb{R}$ be an arbitrary number and assume that the critical value of the upper tailed test based on N_S is given by $t_n = |s_n|^2_{d(n)} + 2|s_n|_{d(n)} w$. From Theorem 3.4 and (5.4) we obtain

$$\beta_n = P_n\Big(\frac{N_S - |s_n|^2_{d(n)}}{2|s_n|_{d(n)}} \geqslant w\Big) \to 1 - \mathbf{\Phi}(w) \tag{5.11}$$

as $n \to \infty$, where $\mathbf{\Phi}$ is the standard normal distribution function. By Theorem 3.2 and the assumption (3.8), the size of this test satisfies

$$\alpha_n = P_0(N_S \geqslant t_n) \leqslant \exp\big\{-(1/2)n\rho_n^2 - \sqrt{n}\rho_n w + o(\sqrt{n}\rho_n)\big\}. \tag{5.12}$$

Consider the Neyman–Pearson test statistic V_n for testing uniformity, given in Section 3, against the distribution given by the density $1 + \rho_n a(t)$ on the unit interval. Then from Theorems 3.1 and 3.2 in Inglot and Ledwina [24] (see also Proposition 4.1 and the formula (4.5) in Inglot [15]) it follows that

$$P_n(V_n \geqslant \sqrt{n} m_n + y) \to 1 - \mathbf{\Phi}(y) \tag{5.13}$$

as $n \to \infty$ for every $y \in \mathbb{R}$, where $m_n = \rho_n + O(\rho_n^2)$ and

$$P_0(V_n \geqslant \sqrt{n} x_n) = \exp\big\{-(1/2)n x_n^2 + O(n x_n^3) + O(\log n x_n^2)\big\} \tag{5.14}$$

for every $x_n \to 0$ and $n x_n^2 \to \infty$. Choose a real number y_n such that

$$P_0(V_n \geqslant \sqrt{n} m_n + y_n) = \alpha_n, \tag{5.15}$$

i.e. such that $\sqrt{n} m_n + y_n$ is an exact critical value for the Neyman–Pearson test, which ensures exactly the same significance level α_n as defined by (5.12) for the test N_S.

We have $y_n \leqslant w + o(1)$ since otherwise (5.11) and (5.13) would imply that the Neyman–Pearson test attains lower power than the test N_S, both being on the level α_n, which is impossible.

From (5.12), (5.14), and the assumptions on ρ_n it follows that for n sufficiently large $P_0(V_n \geqslant \sqrt{n}\rho_n/2) > \exp\{-n\rho_n^2/4\} > \alpha_n$, which means $y_n \geqslant -\sqrt{n}\rho_n/2$

for n sufficiently large. So, (5.14) can be applied for $x_n = m_n + y_n/\sqrt{n}$. Comparing the exponents from this expansion and those given by (5.12) we see that y_n satisfies the inequality

$$\frac{1}{2\sqrt{n}\rho_n} y_n^2 + y_n - w + o(1) \geqslant 0$$

and, consequently, the inequality $y_n \geqslant w + o(1)$. This proves that $y_n = w + o(1)$ and, by (5.11) and (5.13),

$$\lim_n \beta_n^+ = \lim_n P_n(V_n \geqslant \sqrt{n} m_n + y_n) = 1 - \mathbf{\Phi}(w) = \lim_n \beta_n.$$

Thus the assertion of the theorem follows. ■

P r o o f o f T h e o r e m 3.6. We shall apply Theorem 2.7 of Inglot [14]. Under the assumptions of our theorem it follows immediately from (3.2) that

$$\lim_{n\to\infty} \frac{1}{nx_n^2} \log P_0(N_S \geqslant nx_n^2) = -\frac{1}{2}$$

for every sequence x_n as in Theorem 3.2. This and Theorem 3.4 imply that the intermediate slope of the test N_S for the considered sequence (P_n) of alternatives is equal to $|s_n|^2_{d(n)}/2$, while the intermediate slope for the Neyman–Pearson test V_n for this sequence is equal to $n\rho_n^2/2$. This means that the ratio of slopes equals $\left|\int_0^1 b(t)a(t)dt\right|^2_{d(n)}$ and tends to one by the Parceval identity (cf. (5.4)). Since the choice of the critical value t_n given in (3.7) and the form of significance level α_n in (5.12) ensure nondegenerate power for the test N_S and $(\log \alpha_n)/n \to 0$, Theorem 2.7 of Inglot [14] can be applied, thus proving our theorem. ■

6. APPENDIX

We collect here some auxiliary, general results on asymptotic behaviour of linear rank statistics related to the one-sample problem. We present them to assure an easy and precise reference in Section 5 and for the convenience of the reader. Analogous results for linear rank statistics in the two-sample problem were investigated in Inglot [16].

Limit theorems for rank statistics have been investigated intensively by many authors, among them Hájek [12] and Govindarajulu [11]. Hájek's approach is quite general and his main results are often referred to also in recent papers. However, his proofs are technically involved and do not allow for explicit probability inequalities. To get asymptotic results more flexible for various applications we propose here to use Hungarian construction and, in particular, the celebrated Komlós, Major and Tusnády inequality (KMT inequality, for short). Such methodology was successfully exploited to establish limit behaviour of some classes of statistics like

Lipschitz functionals of the empirical process (Inglot and Ledwina [21], [22]) or bilinear forms of empirical process (Inglot et al. [18]) for the problem of testing uniformity.

Before we state our results we need some additional notation to that we have used in Section 2.

Let $X_1,\dots,X_n$ be a sample from a continuous distribution F on the real line and let $F_s(x) = \frac{1}{2}[F(x)+1-F(-x)]$ be its symmetric part. Transform our sample into the unit interval by F_s to get $U_i = F_s(X_i),\ i=1,\dots,n$. Then U_i have the distribution function $F\circ F_s^{-1}(t) = t + A(t),\ t\in[0,1]$, where A is an absolutely continuous function on $[0,1]$, symmetric with respect to $t=1/2$. The case $F=F_s$ corresponds to $A\equiv 0$.

As previously, let $\mathcal{F}_n(x)$ denote the empirical distribution function of the sample $X_1,\dots,X_n$ and $\mathcal{F}_{ns}$ its symmetric part which we have used as an estimator of the unknown distribution F_s. Moreover, denote by $\mathcal{H}_n = \mathcal{F}_n\circ F_s^{-1}$ the empirical distribution function of the transformed sample $U_1,\dots,U_n$. Next, let $\varepsilon_n = \sqrt{n}(\mathcal{H}_n - F\circ F_s^{-1}) \stackrel{\mathcal{D}}{=} e_n\circ F\circ F_s^{-1}$ be the transformed empirical process on $[0,1]$, where e_n denotes the uniform empirical process. It is easy to verify that

$$\text{(A.1)}\qquad \mathcal{H}_{ns}(t) = \frac{1}{2}[\mathcal{H}_n(t)+1-\mathcal{H}_n(1-t)] = \frac{1}{2\sqrt{n}}[\varepsilon_n(t)-\varepsilon_n(1-t)].$$

Now, let $\varphi_1,\varphi_2,\dots$ be a system of linearly independent, absolutely continuous functions on $[0,1]$, antisymmetric with respect to $t=1/2$ and such that $\int_0^1\varphi_j^2(t)dt = 1$ for all j. Let $\Phi = (\varphi_1,\varphi_2,\dots)$ stand for the vector of these functions. We shall use φ_j's as score functions to define the following linear rank statistic which is our main object in this section:

$$\text{(A.2)}\quad \hat{\Phi} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \Phi\big(\mathcal{F}_{ns}(X_i)\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \Phi\big(\mathcal{H}_{ns}(U_i)\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \Phi\left(\frac{R_i}{2n}\right),$$

where R_i is the rank of X_i in the pooled sample $X_1,\dots,X_n - X_1,\dots,-X_n$ (cf. Section 2). Components $\hat{\varphi}_j$ of $\hat{\Phi}$ may be considered as empirical Fourier coefficients of the antisymmetric part of F, i.e., $F_a = F - F_s$ with respect to the system Φ. A simple calculation shows that $\hat{\Phi}$ can be written in the form

$$\text{(A.3)}\qquad \hat{\Phi} = \int_0^1 \Phi\big(\mathcal{H}_{ns}(t)\big)d\varepsilon_n(t) + \sqrt{n}\int_0^1 \Phi\big(\mathcal{H}_{ns}(t)\big)dA(t),$$

where we have used the fact that $\Phi\big(\mathcal{H}_{ns}(t)\big)$ is antisymmetric with respect to $t=1/2$.

Observe that, under $F=F_s$, $\mathcal{H}_{ns}$ may be considered as an estimator of the identity function $\mathrm{id}(t)=t$. The following simple fact will be used in the sequel to replace $\mathcal{H}_{ns}(t)$ by t in (A.3) when studying an asymptotics of $\hat{\Phi}$.

PROPOSITION A.1. *Under the above notation we have*

$$\int_0^1 \left[\Phi\big(\mathcal{H}_{ns}(t)\big) - \Phi(t)\right] d\varepsilon_n(t) = \sum_{i=1}^{2n} \int_{(i-1)/2n}^{i/2n} \Phi'(s)[\varepsilon_n(s) - \varepsilon_n(U_{(i)})]ds, \tag{A.4}$$

where $U_{(i)}$ are the order statistics in the pooled transformed sample $U_1, \ldots, U_n$, $1 - U_1, \ldots, 1 - U_n$.

Proof of Proposition A.1. By the absolute continuity of Φ we have

$$\int_0^1 \left[\Phi\big(\mathcal{H}_{ns}(t)\big) - \Phi(t)\right] d\varepsilon_n(t) = \int_0^1 \int_t^{\mathcal{H}_{ns}(t)} \Phi'(s) ds d\varepsilon_n(t).$$

The last integral can be written as a double integral over the subset D of the unit square which lies between graphs of $\mathcal{H}_{ns}(t)$ and the identity function $\mathrm{id}(t)$ (cf. Lemma 2.1 in Inglot [16]). Using the Fubini theorem, dividing D into $2n$ parts by horizontal lines $s = i/(2n), i = 1, \ldots, 2n-1$, and integrating with respect to t over each part, we get (A.4). ■

The formula (A.4) suggests to define an auxiliary statistic of the form

$$\hat{\Phi}_0 = \int_0^1 \Phi(t) d\varepsilon_n(t) = -\int_0^1 \Phi'(t)\varepsilon_n(t)dt, \tag{A.5}$$

which corresponds to $\hat{\Phi}$ with $\mathcal{H}_{ns}$ replaced by the identity function.

Finally, let $B(t)$ denote the Brownian bridge and

$$\gamma = -\int_0^1 \Phi'(t)B(t)dt \tag{A.6}$$

be a Gaussian vector with mean zero and covariance matrix $\Gamma = \int_0^1 \Phi(t)\Phi^T(t)dt$.

Null hypothesis. Now, we consider the case $F = F_s$, i.e. the case where X_i have a symmetric distribution. Then U_i are uniformly distributed over $[0, 1]$ and $\varepsilon_n \overset{\mathcal{D}}{=} e_n$ is the uniform empirical process.

Now, let $k(n)$ be any sequence of natural numbers (including the constant sequence) and $|v|_k = (v_1^2 + \ldots + v_k^2)^{1/2}$ denote the k-dimensional Euclidean norm of a vector v (cf. Section 3). By a straightforward application of KMT inequality we get

$$\mathbf{P}\Big(|\hat{\Phi}_0^{(1)} - \gamma_n^{(1)}|_{k(n)} \geqslant \psi\big(k(n)\big)x_n\Big) \tag{A.7}$$

$$= \mathbf{P}\Big(\Big|\int_0^1 \Phi'(t)[\varepsilon_n^{(1)}(t) - B_n^{(1)}(t)]dt\Big|_{k(n)} \geqslant \psi\big(k(n)\big)x_n\Big) \leqslant C\exp\{-cx_n\sqrt{n}\}$$

provided positive numbers x_n satisfy the condition $nx_n^2/\log^2 n \to \infty$. Here and in the sequel $\varepsilon_n^{(1)}$ and $B_n^{(1)}$ are versions of ε_n and B defined on a common probability space $(\Omega, \mathcal{B}, \mathbf{P})$ constructed in the KMT inequality, $\hat{\Phi}_0^{(1)} = -\int_0^1 \Phi'(t)\varepsilon_n^{(1)}(t)dt$, $\gamma_n^{(1)} = -\int_0^1 \Phi'(t)B_n^{(1)}(t)dt$, and

$$\psi^2(k) = \sum_{j=1}^{k} \Big(\int_0^1 |\varphi_j'(t)|dt\Big)^2 \tag{A.8}$$

is the constant depending on the system Φ which is finite due to the absolute continuity of Φ.

PROPOSITION A.2. *If $F = F_s$, then for any sequence $k(n)$ of natural numbers and every sequence x_n of positive numbers satisfying the conditions $x_n \to 0$ and $nx_n^2/\log n \to \infty$ we have for n sufficiently large*

$$\mathbf{P}\left(|\hat{\Phi}^{(1)} - \hat{\Phi}_0^{(1)}|_{k(n)} \geqslant \psi\big(k(n)\big)\sqrt{nx_n^3}\right) \leqslant C\exp\{-cnx_n^2\}, \tag{A.9}$$

where $\psi(k)$ is given by (A.8).

Proof of Proposition A.2. Consider an event

$$E_n = \{\sup_t |\varepsilon_n^{(1)}(t)| \geqslant \sqrt{n}x_n\} \subset \Omega. \tag{A.10}$$

Then by the KMT inequality and the well-known inequality

$$\mathbf{P}\big(\sup_t |B_n^{(1)}(t)| \geqslant x\big) \leqslant 2\exp\{-2x^2\}, \quad x > 0,$$

we get for n sufficiently large

$$\begin{aligned}\mathbf{P}(E_n) &\leqslant \mathbf{P}\big(\sup_t |\varepsilon_n^{(1)}(t) - B_n^{(1)}(t)| \geqslant (1/2)\sqrt{n}x_n\big)\\ &\quad + \mathbf{P}\big(\sup_t |B_n^{(1)}(t)| \geqslant (1/2)\sqrt{n}x_n\big)\\ &\leqslant C\exp\{-(1/2)nx_n^2\}.\end{aligned} \tag{A.11}$$

On the set E_n^c, for each $i = 1, \ldots, 2n$ and $u \in [(i-1)/(2n), i/(2n)]$ we have by the formula (A.1)

$$|u - U_{(i)}^{(1)}| \leqslant \left|\frac{i}{2n} - U_{(i)}^{(1)}\right| + \frac{1}{2n} = |\mathcal{H}_{ns}^{(1)}(U_{(i)}^{(1)}) - U_{(i)}^{(1)}| + \frac{1}{2n} \leqslant x_n + \frac{1}{2n} = r_n$$

and, consequently, on the set E_n^c we have

$$\begin{aligned}&\max_{1\leqslant i\leqslant 2n} \sup_{u\in[(i-1)/2n, i/2n]} |\varepsilon_n^{(1)}(u) - \varepsilon_n^{(1)}(U_{(i)}^{(1)})|\\ &\quad \leqslant 2\sup_t |\varepsilon_n^{(1)}(t) - B_n^{(1)}(t)| + \sup_{0\leqslant t\leqslant 1-r_n} \sup_{0\leqslant u\leqslant r_n} |B_n^{(1)}(t+u) - B_n^{(1)}(t)|.\end{aligned}$$

Now, an application of Proposition A.1 and an analogous argument to that in the proof of Theorem 3.2 in Inglot [16] gives (A.9). ■

Combining (A.7) and (A.9) we obtain the following theorem.

THEOREM A.1. *If $F = F_s$, then for any sequence $k(n)$ of natural numbers and every sequence x_n of positive numbers satisfying $x_n \to 0$ and $nx_n^2/\log^2 n \to \infty$ we have for n sufficiently large*

$$\mathbf{P}\Big(|\hat{\Phi}^{(1)} - \gamma_n^{(1)}|_{k(n)} \geqslant \psi\big(k(n)\big)\sqrt{nx_n^3}\Big) \leqslant C\exp\{-cnx_n^2\}. \tag{A.12}$$

Observe that from (A.12) one can easily derive asymptotic normality of $\hat{\Phi}$.

Proposition A.2 allows us also to get a moderate deviation theorem for $\hat{\Phi}$. Since for any $\nu > 0$

$$\begin{aligned}&\mathbf{P}(|\hat{\Phi}^{(1)}|^2_{k(n)} \geqslant nx_n^2) \leqslant \mathbf{P}\big(|\gamma_n^{(1)}|_{k(n)} \geqslant (1 - x_n^\nu)x_n\sqrt{n}\big)\\ &+\mathbf{P}\big(|\hat{\Phi}^{(1)} - \hat{\Phi}_0^{(1)}|_{k(n)} \geqslant (1/2)x_n^{1+\nu}\sqrt{n}\big) + \mathbf{P}\big(|\hat{\Phi}_0^{(1)} - \gamma_n^{(1)}|_{k(n)} \geqslant (1/2)x_n^{1+\nu}\sqrt{n}\big)\end{aligned}$$

and

$$\begin{aligned}&\mathbf{P}(|\hat{\Phi}^{(1)}|^2_{k(n)} \geqslant nx_n^2) \geqslant \mathbf{P}\big(|\gamma_n^{(1)}|_{k(n)} \geqslant (1 + x_n^\nu)x_n\sqrt{n}\big)\\ &-\mathbf{P}\big(|\hat{\Phi}^{(1)} - \hat{\Phi}_0^{(1)}|_{k(n)} \geqslant (1/2)x_n^{1+\nu}\sqrt{n}\big) - \mathbf{P}\big(|\hat{\Phi}_0^{(1)} - \gamma_n^{(1)}|_{k(n)} \geqslant (1/2)x_n^{1+\nu}\sqrt{n}\big),\end{aligned}$$

using (A.7), (A.9), and the expansion for the tails of the distribution of the Euclidean norm of a Gaussian vector we get the following proposition (cf. the proof of Theorem 3.4 in Inglot [16]).

PROPOSITION A.3. *If $F = F_s$, then for any sequence $k(n)$ of natural numbers, any $\nu > 0$, and every sequence x_n of positive numbers satisfying*

$$x_n \to 0, \quad \frac{nx_n^2}{k(n)\lambda_n} \to \infty, \quad \frac{n^3 x_n^{4(1+\nu)}}{\psi^4(k(n))\log^3 n} \to \infty \tag{A.13}$$

we have for n sufficiently large

$$\begin{aligned}&\left|P(|\hat{\Phi}|^2_{k(n)} \geqslant nx_n^2) - \exp\left\{-\frac{nx_n^2}{2\lambda_n} + O\left(\frac{nx_n^{2+\nu}}{\lambda_n}\right) + O\left(k(n)\log nx_n^2\right)\right\}\right|\\ &\qquad\qquad \leqslant C\exp\left\{-c\frac{nx_n^{4(1+\nu)/3}}{\psi^{4/3}\big(k(n)\big)}\right\},\end{aligned}$$

where λ_n is the largest eigenvalue of the covariance matrix Γ_n of the first $k(n)$ components of γ.

Proposition A.3 immediately implies

THEOREM A.2. *If $F = F_s$, then for any sequence $k(n)$ of natural numbers, any $\nu \in (0, 1/2)$, and every sequence x_n of positive numbers such that $x_n \to 0$, $nx_n^2/\big(k(n)\lambda_n\big) \to \infty$, and $x_n^{2-4\nu}\psi^4\big(k(n)\big)/\lambda_n^3 \to 0$ we have*

$$P(|\hat{\Phi}|^2_{k(n)} \geqslant nx_n^2) = \exp\left\{-\frac{nx_n^2}{2\lambda_n} + O\left(\frac{nx_n^{2+\nu}}{\lambda_n}\right) + O\left(k(n)\log nx_n^2\right)\right\}, \tag{A.14}$$

where λ_n is the largest eigenvalue of the covariance matrix Γ_n of the first $k(n)$ components of γ and $\psi(k)$ is given by (A.8). *In particular, when Φ is an orthonormal system, then $\lambda_n = 1$ for all n.*

Convergent alternatives. Now, suppose alternatives $A(t) = A_n(t)$ converge to zero. Namely, denote by $\rho_n a_n,\ \rho_n > 0$, the derivative of A_n and suppose $\rho_n \to 0$, while a_n are uniformly bounded, i.e. $\sup_n \sup_t |a_n(t)| \leqslant a_\infty < \infty$, and normalized, i.e. $\int_0^1 a_n^2(t)dt = 1$. We have the following proposition (cf. Theorems 4.1 and 4.2 in Inglot [16]).

PROPOSITION A.4. *Suppose $X_1, \ldots, X_n$ have the distribution function of the form $F_n(x) = F_s(x) + A_n\big(F_s(x)\big)$, where F_s is a distribution function of a fixed symmetric distribution and A_n are as described above with $\rho_n \to 0$ such that $n\rho_n^2/\log n \to \infty$. Then for any sequence $k(n)$ of natural numbers and every sequence $\delta_n \to 0$ of positive numbers such that*

$$\frac{n\rho_n\delta_n^2}{\log n} \to \infty \tag{A.15}$$

we have for n sufficiently large

(i) $$\mathbf{P}\Big(|\hat{\Phi}_0^{(1)} - \gamma_n^{(1)}|_{k(n)} \geqslant \psi\big(k(n)\big)\sqrt{n}\rho_n\delta_n\Big) \leqslant C\exp\{-cn\rho_n\delta_n^2\}$$

and

(ii) $$\mathbf{P}\Big(\Big|\int_0^1 \Phi\big(\mathcal{H}_{ns}^{(1)}(t)\big)d\varepsilon_n^{(1)}(t) - \hat{\Phi}_0^{(1)}(t)\Big|_{k(n)} \geqslant \psi\big(k(n)\big)\sqrt{n}\rho_n\delta_n\Big) \leqslant C\exp\{-cn\rho_n^{3/2}\delta_n\},$$

where $\psi(k)$ is given by (A.8).

P r o o f o f P r o p o s i t i o n A.4. Since $\varepsilon_n^{(1)}(t) = e_n^{(1)}\big(F_n\big(F_s^{-1}(t)\big)\big)$ and, by the Schwarz inequality, $\sup_t \big|F_n\big(F_s^{-1}(t)\big) - t\big| = \sup_t |A_n(t)| \leqslant \rho_n/\sqrt{2}$, it follows that

$$\begin{aligned}\mathbf{P}\big(\sup_t |\varepsilon_n^{(1)}(t) - B_n^{(1)}(t)| \geqslant \sqrt{n}\rho_n\delta_n\big) &\leqslant \mathbf{P}\big(\sup_t |e_n^{(1)}(t) - B_n^{(1)}(t)| \geqslant \sqrt{n}\rho_n\delta_n^2\big)\\ &+ \mathbf{P}\big(\sup_{0\leqslant t\leqslant 1-\rho_n}\ \sup_{0\leqslant u\leqslant\rho_n} |B_n^{(1)}(t+u) - B_n^{(1)}(t)| \geqslant (1-\delta_n)\sqrt{n}\rho_n\delta_n\big),\end{aligned}$$

which can be estimated by $C\exp\{-cn\rho_n\delta_n^2\}$ due to the KMT inequality, a version of Lemma 1.1.1 of Csörgő and Révész [5] stated for the Brownian bridge and (A.15). Now, repeating an analogous calculation to that in (A.7) we prove (i).

Consider an event (cf. (A.10))

$$E_n = \{\sup_t |\varepsilon_n^{(1)}(t)| \geqslant \sqrt{n\rho_n}\delta_n\} \subset \Omega. \tag{A.16}$$

A similar argument to that in the proof of Proposition A.2 gives an estimation $\mathbf{P}(E_n) \leqslant C\exp\{-cn\rho_n\delta_n^2\}$. On the set E_n^c, for each $i = 1,\ldots,2n$ and $u \in [(i-1)/(2n), i/(2n)]$ we have by (A.1)

$$|u - U_{(i)}^{(1)}| \leqslant |\mathcal{H}_{ns}^{(1)}(U_{(i)}^{(1)}) - U_{(i)}^{(1)}| + \frac{1}{2n} \leqslant \sqrt{\rho_n}\delta_n + \frac{1}{2n} = r_n. \tag{A.17}$$

Again, arguing as in the proof of Proposition A.2 we get (ii). ∎

Set (cf. (A.3))

$$\mathcal{S}_n = \sqrt{n}\int_0^1 \Phi\big(\mathcal{H}_{ns}(t)\big)dA_n(t) = \sqrt{n}\rho_n\int_0^1 \Phi\big(\mathcal{H}_{ns}(t)\big)a_n(t)dt$$

and its deterministic counterpart (cf. (3.5))

$$s_n = \sqrt{n}\rho_n\int_0^1 \Phi(t)a_n(t)dt. \tag{A.18}$$

An analogue of Proposition A.1 applied to $\mathcal{S}_n$ and s_n gives

$$|\mathcal{S}_n - s_n|_{k(n)} \leqslant a_\infty\psi\big(k(n)\big)\sqrt{n}\rho_n \max_{1\leqslant i\leqslant 2n}\ \sup_{u\in[(i-1)/2n, i/2n]} |u - U_{(i)}|.$$

Using this estimate, considering the event E_n given by (A.16), the estimate (A.17) and

$$|\hat{\Phi}^{(1)} - s_n - \hat{\Phi}_0^{(1)}|_{k(n)} \leqslant \Big|\int_0^1 \Phi\big(\mathcal{H}_{ns}^{(1)}(t)\big)d\varepsilon_n^{(1)}(t) - \hat{\Phi}_0^{(1)}(t)\Big|_{k(n)} + |\mathcal{S}_n^{(1)} - s_n|_{k(n)}$$

we obtain the following theorem.

THEOREM A.3. *Suppose $X_1,\ldots,X_n$ have the distribution function of the form $F_n(x) = F_s(x) + A_n\big(F_s(x)\big)$, where F_s is a distribution function of a fixed symmetric distribution and A_n are as in Proposition* A.4 *with $\rho_n \to 0$ such that $n\rho_n^2/\log n \to \infty$. Then for any sequence $k(n)$ of natural numbers and every se-*

quence $\delta_n \to 0$ *of positive numbers satisfying* (A.15) *and such that* $\delta_n = O(\sqrt{\rho_n})$ *we have for* n *sufficiently large*

$$\text{(A.19)} \qquad \mathbf{P}\Big(|\mathcal{S}_n^{(1)} - s_n|_{k(n)} \geqslant a_\infty \psi\big(k(n)\big)\sqrt{n\rho_n^3}\delta_n\Big) \leqslant C\exp\{-cn\rho_n\delta_n^2\}$$

and

$$\text{(A.20)} \qquad \mathbf{P}\Big(|\hat{\Phi}^{(1)} - s_n - \gamma_n^{(1)}|_{k(n)} \geqslant \psi\big(k(n)\big)\sqrt{n}\rho_n\delta_n\Big) \leqslant C\exp\{-cn\rho_n\delta_n^2\},$$

where $\psi(k)$ *is given by* (A.8).

Fixed alternatives. Let $A(t) \neq 0$ be fixed. Observe that $\varepsilon_n(t) \stackrel{\mathcal{D}}{=} e_n\big(t + A(t)\big)$. Repeating an argument as in the proof of Proposition A.2 and (A.19) we obtain the following result.

PROPOSITION A.5. *Suppose* $X_1, \ldots, X_n$ *have the distribution function of the form* $F(x) = F_s(x) + A\big(F_s(x)\big)$, *where* F_s *is a distribution function of a fixed symmetric distribution and* A *is a fixed absolutely continuous function on* $[0,1]$, *symmetric with respect to* $t = 1/2$. *Then for any sequence* $k(n)$ *of natural numbers*, *any* $0 \leqslant \sigma < 1/2$, *and every sequence* x_n *of positive numbers with* $x_n \to 0$ *and* $nx_n^2/\log^2 n \to \infty$ *we have for* n *sufficiently large*

$$\text{(A.21)} \qquad P\Big(\Big|\int_0^1 \Phi\big(\mathcal{H}_{ns}(t)\big)d\varepsilon_n(t) - \hat{\Phi}_0\Big|_{k(n)} \geqslant \psi\big(k(n)\big)x_n^\sigma\Big) \leqslant C\exp\{-cnx_n^2\} + C\exp\{-cx_n^{2\sigma-1}\}$$

and

$$\text{(A.22)} \qquad P\Big(\Big|\int_0^1 \Phi\big(\mathcal{H}_{ns}(t)\big)dA(t) - \int_0^1 \Phi(t)dA(t)\Big|_{k(n)} \geqslant 3\psi\big(k(n)\big)x_n\Big) \leqslant C\exp\{-cnx_n^2\}.$$

Proof of Proposition A.5. Let us put $E_n = \{\sup_t|\varepsilon_n^{(1)}(t)| \geqslant \sqrt{n}x_n\}$ (cf. (A.10)). Then by a similar argument to that in the proof of Proposition A.2 we show that $\mathbf{P}(E_n) \leqslant C\exp\{-cnx_n^2\}$. Next, we pattern the proof of (A.9) with the only difference that $\max_{1\leqslant i\leqslant 2n}\sup_{u\in[(i-1)/2n, i/2n]}|\varepsilon_n^{(1)}(u) - \varepsilon_n^{(1)}(U_{(i)}^{(1)})|$ will be estimated by

$$2\sup_t|\varepsilon_n^{(1)}(t) - B_n^{(1)}(t)| + \sup_{0\leqslant t\leqslant 1-2r_n}\ \sup_{0\leqslant u\leqslant 2r_n}|B_n^{(1)}(t+u) - B_n^{(1)}(t)|,$$

where $r_n = x_n + 1/(2n)$. This is a consequence of the property $|a(t)| \leqslant 1$ (cf. Section 2) and the estimate $|s + A(s) - U_{(i)}^{(1)} - A(U_{(i)}^{(1)})| \leqslant 2r_n$ which holds on the set E_n^c for all $s \in [(i-1)/(2n), i/(2n)]$ and $i = 1, \ldots, 2n$. The proof is complete. ■

Acknowledgements. The authors thank the referee for careful reading of the manuscript and many useful remarks.

REFERENCES

[1] A. Bakshaev, *Nonparametric tests based on N-distances*, Lituanian Math. J. 48 (2008), pp. 368–379.

[2] K. Behnen and G. Neuhaus, *Rank Tests with Estimated Scores and Their Application*, Teubner, Stuttgart 1989.

[3] D. Cassart, M. Hallin, and D. Paindaveine, *Optimal detection on Fechner-asymmetry*, J. Statist. Plann. Inference 138 (2008), pp. 2499–2525.

[4] W.-H. Cheng and N. Balakrishnan, *A modified sign test for symmetry*, Comm. Statist. Simulation Comput. 33 (2004), pp. 703–709.

[5] S. Csörgő and P. Révész, *Strong Approximations in Probability and Statistics*, Akadémiai Kiadó, Budapest – Academic Press, New York 1981.

[6] K. A. Doksum, G. Fenstad, and R. Aaberge, *Plots and tests for symmetry*, Biometrika 64 (1977), pp. 473–487.

[7] G. R. Ducharme and T. Ledwina, *Efficient and adaptive nonparametric test for the two-sample problem*, Ann. Statist. 31 (2003), pp. 2036–2058.

[8] R. L. Eubank, V. N. LaRiccia, and R. B. Rosenstein, *Test statistics derived as components of Pearson's phi–squared distance measure*, J. Amer. Statist. Assoc. 82 (1987), pp. 816–825.

[9] J. Fan, *Test of significance based on wavelet thresholding and Neyman's truncation*, J. Amer. Statist. Assoc. 91 (1996), pp. 674–688.

[10] M. Freimer, G. Kollia, G. S. Mudholkar, and C. T. Lin, *A study of the generalized Tukey lambda family*, Comm. Statist. Theory Methods 17 (1988), pp. 3547–3567.

[11] Z. Govindarajulu, *Two sided confidence limits for $P(Y < X)$ based on normal samples of X and Y*, Sankhyā Ser. B 29 (1967), pp. 35–40.

[12] J. Hájek, *Asymptotic normality of simple linear rank statistics under alternatives,* Ann. Math. Statist. 39 (1968), pp. 325–346.

[13] J. Hájek and Z. Šidák, *Theory of Rank Tests*, Academia, Prague 1967.

[14] T. Inglot, *Generalized intermediate efficiency of goodness of fit tests*, Math. Methods Statist. 8 (1999), pp. 487–509.

[15] T. Inglot, *Intermediate efficiency by shifting alternatives and evaluation of power*, J. Statist. Plann. Inference 140 (2010), pp. 3263–3281.

[16] T. Inglot, *Asymptotic behaviour of linear rank statistics for the two-sample problem*, Probab. Math. Statist. 32 (2012), pp. 93–116.

[17] T. Inglot and A. Janic, *How powerful are data driven score tests for uniformity*, Appl. Math. (Warsaw) 36 (2009), pp. 375–395.

[18] T. Inglot, W. C. M. Kallenberg, and T. Ledwina, *Asymptotic behavior of some bilinear functionals of the empirical process*, Math. Methods Statist. 2 (1993), pp. 316–336.

[19] T. Inglot, W. C. M. Kallenberg, and T. Ledwina, *Data driven smooth tests for composite hypotheses*, Ann. Statist. 25 (1997), pp. 1222–1250.

[20] T. Inglot, W. C. M. Kallenberg, and T. Ledwina, *Vanishing shortcoming of data driven Neyman's tests*, in: *Asymptotic Methods in Probability and Statistics: A Volume in Honour of Miklós Csörgő*, North-Holland, 1998, pp. 811–829.

[21] T. Inglot and T. Ledwina, *On probabilities of excessive deviations for Kolmogorov–Smirnov, Cramér–von Mises and chi-square statistics*, Ann. Statist. 18 (1990), pp. 1491–1495.

[22] T. Inglot and T. Ledwina, *Moderately large deviations and expansions of large deviations for some functionals of weighted empirical processes*, Ann. Probab. 21 (1993), pp. 1691–1705.

[23] T. Inglot and T. Ledwina, *Asymptotic optimality of data-driven Neyman's tests for uniformity*, Ann. Statist. 24 (1996), pp. 1982–2019.

[24] T. Inglot and T. Ledwina, *Intermediate approach to comparison of some goodness-of-fit tests*, Ann. Inst. Statist. Math. 53 (2001), pp. 810–834.

[25] T. Inglot and T. Ledwina, *Towards data driven selection of a penalty function for data driven Neyman tests*, Linear Algebra Appl. 417 (2006), pp. 124–133.

[26] A. Janic-Wróblewska, *Data driven rank test for univariate symmetry*, Technical Report No. I18/98/P-020, Institute of Mathematics, Wrocław University of Technology.

[27] A. Janic-Wróblewska, W. C. M. Kallenberg, and T. Ledwina, *Detecting positive quadrant dependence and positive function dependence*, Insurance Math. Econom. 34 (2004), pp. 467–487.

[28] A. Janic-Wróblewska and T. Ledwina, *Data driven rank test for two-sample problem*, Scand. J. Statist. 27 (2000), pp. 281–297.

[29] W. C. M. Kallenberg and T. Ledwina, *Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests*, Ann. Statist. 23 (1995), pp. 1594–1608.

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

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

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

[33] J. A. Koziol, *On a Cramér-von Mises type statistic for testing symmetry*, J. Amer. Statist. Assoc. 75 (1980), pp. 161–167.

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

[35] T. Ledwina and G. Wyłupek, *Two-sample test against one-sided alternatives*, Scand. J. Statist. 39 (2012), pp. 358–381.

[36] T. P. McWilliams, *A distribution-free test for symmetry based on a runs statistic*, J. Amer. Statist. Assoc. 85 (1990), pp. 1130–1133.

[37] R. Modarres and J. L. Gastwirth, *A modified runs test for symmetry*, Statist. Probab. Lett. 31 (1996), pp. 107–112.

[38] R. Modarres and J. L. Gastwirth, *Hybrid test for the hypothesis of symmetry*, J. Appl. Stat. 25 (1998), pp. 777–783.

[39] J. Neyman, *'Smooth test' for goodness of fit*, Skand. Aktuarietidskr. 20 (1937), pp. 149–199.

[40] A. I. Orlov, *On testing the symmetry of distributions*, Theory Probab. Appl. 7 (1972), pp. 357–361.

[41] E. D. Rothman and M. Woodroofe, *A Cramér–von Mises type statistic for testing symmetry*, Ann. Math. Statist. 43 (1972), pp. 2035–2038.

[42] G. Sansone, *Orthogonal Functions*, Interscience, New York 1959.

[43] G. Schwarz, *Estimating the dimension of a model*, Ann. Statist. 6 (1978), pp. 461–464.

[44] I. H. Tajuddin, *Distribution-free test for symmetry based on Wilcoxon two-sample test*, J. Appl. Stat. 21 (1994), pp. 409–415.

[45] O. Thas, J. C. W. Rayner, and D. J. Best, *Tests for symmetry based on the one-sample Wilcoxon signed rank statistic*, Comm. Statist. Simulation Comput. 34 (2005), pp. 957–973.

[46] G. Wyłupek, *Data driven k-sample tests*, Technometrics 52 (2010), pp. 107–123.

Institute of Mathematics and Computer Science
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland
E-mail: Tadeusz.Inglot@pwr.wroc.pl
E-mail: Alicja.Janic@pwr.wroc.pl
E-mail: jadwiga.jozefczyk@pwr.wroc.pl

Received on 3.3.2012;
revised version on 30.11.2012