Detection of non-Gaussianity

Teresa Ledwina and Grzegorz Wyłupek*

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

(Received 20 September 2013; accepted 28 October 2014)

We develop two tests sensitive to various departures from composite goodness-of-fit hypothesis of normality. The tests are based on the sums of squares of some components naturally arising in decomposition of the Shapiro–Wilk-type statistic. Each component itself has diagnostic properties. The numbers of squared components in sums are determined via some novel selection rules based on the data. The new solutions prove to be effective tools in detecting a broad spectrum of sources of non-Gaussianity. We also discuss two variants of the new tests adjusted to verification of simple goodness-of-fit hypothesis of normality. These variants also compare well to popular competitors.

Keywords: Akaike selection rule; correlation test; goodness-of-fit; higher criticism; L-statistic; model selection; sample quantile function; Shapiro–Wilk test; test of normality; Wasserstein distance

2010 Mathematics Subject Classification: 62F03; 62G10

1. Introduction

Detection of non-Gaussianity, though being a classical problem, is still a subject of intensive research. This is motivated by wide applications and the existing needs to detect non-standard and complex deviations from Gaussianity. For discussion and some evidence, see, for example, Graham et al.,[1] Jin et al.,[2] Güner et al.,[3] Romão et al.,[4] Tarongi and Camps [5] and references therein.

One of the best existing procedures for verifying Gaussianity is the test of Shapiro and Wilk.[6] Since its application to large data sets was not easy while the asymptotic theory related to this statistic was very difficult, many simplified variants of this test have been elaborated. For an extensive discussion see Brown and Hettmansperger [7] and LaRiccia.[8] We would like to supplement the list of solutions presented in these references by two more recent and important ones by Chen and Shapiro [9] and del Barrio et al.[10] In extensive study of Romão et al.,[4] the three just mentioned statistics have been evaluated as the best omnibus tests of normality.

Several variants and modifications of the Shapiro–Wilk test can be, in principle, classified into three groups corresponding to: optimization of plotting positions,[7,9,11] measuring a discrepancy between empirical and estimated normal quantile function [10,12–14] and procedures attempting to summarize more carefully the appearance of nonlinearity in normal probability plots.[8,15–17] Obviously, these classes are not disjoint and other grouping is also possible. The

*Corresponding author. Email: wylupek@impan.pan.wroc.pl

© 2014 Taylor & Francis
goal of the present paper is to modify, combine, and refine two selected solutions from the first and third group, respectively.

Some details are as follows. We consider the classical set-up. The observed variables $X_1, \ldots, X_n$ are independent and identically distributed with finite second moment, the common distribution function $F$, and the quantile function $Q = F^{-1}$. In terms of the quantile function, the null hypothesis that $X_i$’s have Gaussian distribution can be written as

$$H: Q = \mu + \sigma \Phi^{-1},$$

where $\Phi^{-1}$ is the quantile function of the standard normal distribution function $\Phi$ while $\mu \in (-\infty, +\infty)$ and $\sigma \in (0, +\infty)$ are unknown.

Let $X_{1:n} \leq \cdots \leq X_{n:n}$ denote order statistics of the sample $X_1, \ldots, X_n$ and let $Q_n$ be the sample quantile function, i.e. $Q_n(t) = X_{i:n}$ for $i - 1 < m \leq i$, $i = 1, \ldots, n$. Additionally, set $Q_n(0) = \lim_{t\downarrow 0} Q_n(t) = X_{1:n}$. The non-normality of $F$ can be measured through the departure of $Q_n$ from the linearity specified by $H$. As a measure of the discrepancy one can consider the Wasserstein distance between $Q_n$ and the best fitted function $\mu + \sigma \Phi^{-1}$, i.e. the quantity

$$W_n = \left\{ \int_0^1 [Q_n(t) - \hat{\mu}_n - \hat{\sigma}_n \Phi^{-1}(t)]^2 \, dt \right\}^{1/2},$$

where $\hat{\mu}_n$ and $\hat{\sigma}_n$ solve the problem

$$\minimize_{\mu, \sigma} \int_0^1 [Q_n(t) - \mu - \sigma \Phi^{-1}(t)]^2 \, dt.$$

For relation of the minimization procedure to the problem of selection of the plotting positions in construction of normality plot see Brown and Hettmansperger.[7] On the other hand, by LaRiccia,[8]

$$W_n^2 = \sum_{j=2}^{\infty} \left\{ \int_0^1 Q_n(t) h_j(t) \, dt \right\}^2,$$

where $\{h_j\}_{j \geq 0}$ is transformed orthonormal Hermite polynomials system with $h_j(t) = H_j(\Phi^{-1}(t))/\sqrt{j!}$ and the equality is understood in the $L_2$-sense. Therefore, the quadratic measure of fit of the residuals

$$R_n(t) = Q_n(t) - \hat{\mu}_n - \hat{\sigma}_n \Phi^{-1}(t)$$

(2)

can be decomposed into the sum of squares of components which can be interpreted as empirical Fourier coefficients of unknown $Q$ in the system $\{h_j\}_{j \geq 0}$.

Our approach to construction of the test of $H$ is via analysis of components of $W_n^2$. In this respect, we follow the ideas contained in [7] as well as LaRiccia [8] where tests based on two first components of $W_n^2$ were considered. Their approach belongs to a stream of popular solutions based on empirical skewness and kurtosis such as D’Agostino–Pearson, Bowman–Shenton or Jarque–Bera tests, for example. Such solutions are effective in some situations but fail to detect many non-normal distributions. For some evidence, see Romão et al.[4] Therefore, in our approach the number of components of $W_n^2$, that are taken into account, is not restricted to one or two first ones, but is specified by some data-driven Akaike-type selection rule. We define and investigate omnibus test of this type. In this way, we supplement and refine the existing ideas by providing a final touch which makes the resulting solution to be much more flexible and sensitive to a larger spectrum of deviations from $H$. Moreover, motivated by the scope of the papers by Jin et al. [2] and Güner et al.,[3] among others, we also propose a variant of the solution which
is adjusted to detection of symmetrical alternatives. Extensive simulations show that empirical
detection rule of new solutions is appealing.

The rest of this article is organized as follows. Section 2 introduces notations, some
heuristics and basic elements of our constructions. In Section 3, we state basic asymptotic
properties of the new tests of \( H \) saying that the test statistics have non-degenerate null distributions and the
tests are consistent for a large set of alternatives. We also present useful result on the behaviour
of the selection rule in large samples. Moreover, some precautions on accuracy of finite sam-
ple results are provided. Section 4 introduces the first- and second-moment corrections which
improve finite sample behaviour of the components of \( \mathcal{W}_n^2 \) under \( H \) and, in consequence, which
make them useful auxiliary diagnostic tools. Section 4 also contains a selection of simulated
powers under small and moderate sample sizes. We included here six cases of symmetrical alter-
natives only, while other outcomes we postponed to the supplementary material. The results are
couraging. Therefore, we also considered less traditional setting and some specialized alter-
natives. Namely, following Jin et al.,[2] we studied large sample case and symmetrical alternatives
being sparse mixtures and linear combinations of normal and power-tail distributed observations.
In Section 4, we included two such cases for an illustration. Section 5 contains some real data
examples. In Section 6, we additionally discuss variants of our solutions adjusted to testing sim-
ple null hypothesis \( H^* \) : \( Q = \Phi^{-1} \) and compare them to the higher criticism statistic popularized
by Donoho and Jin.[18]

The conclusions are that in both cases (testing \( H \) and \( H^* \)), the new solutions work well, outper-
form simple constructions based on the sums of squares of the first few components and compare
to the best existing solutions.

2. Some methodology of testing \( H \)

2.1. Preliminaries

Let \( \varphi \) denote the density of the standard normal distribution. Recall that the \( j \)th Hermite polynomial \( H_j \) is defined by \( H_j(x) = (-1)^j \big( d/dx \big)^j \varphi(x) \big) \big( \varphi(x) \big) \big. j \geq 1, \ H_0(x) \equiv 1, \ x \in \mathbb{R}. \) The
system is complete in \( \mathbb{L}_2(\mathbb{R}, \varphi(x) \, dx) \) and \( H_j \)'s satisfy:
\[
\int_{\mathbb{R}} H_i(x) H_j(x) \varphi(x) \, dx = 0, \quad i \neq j, \text{ and } \int_{\mathbb{R}} H_j^2(x)\varphi(x) \, dx = j!.
\]
Therefore, the related system of functions on \((0,1)\) given by
\[
h_j(t) = \frac{1}{\sqrt{j!}} H_j(\Phi^{-1}(t)), \quad j \geq 1, \quad h_0(t) \equiv 1, \quad t \in (0,1),
\]
is orthonormal and complete in \( \mathbb{L}_2((0,1), dt) \). For illustration, we plot in the supplementary
material the functions \( h_1, \ldots, h_6 \). The formula (3) in [7] relates this system to the inverted Edgeworth
expansion of \( F \).

Basing on pp.1670–1671 of the last mentioned paper and Section 5 of LaRiccia,[8] we list
below the useful results on the statistics \( \hat{\mu}_n, \hat{\sigma}_n, \) and \( R_n(t), \) defined in Section 1.

\[
\hat{\mu}_n = \int_0^1 Q_n(t) h_0(t) \, dt = n^{-1} \sum_{i=1}^n X_{i:n}, \tag{3}
\]

\[
\hat{\sigma}_n = \int_0^1 Q_n(t) h_1(t) \, dt = \sum_{i=1}^n b_i X_{i:n} = \sum_{i=1}^{n-1} \varphi_i (X_{i+1:n} - X_{i:n}), \tag{4}
\]

where
\[
b_i = \int_{(i-1)/n}^{i/n} \Phi^{-1}(t) \, dt = \varphi \left( \Phi^{-1} \left( \frac{i-1}{n} \right) \right) - \varphi \left( \Phi^{-1} \left( \frac{i}{n} \right) \right) = \varphi_{i-1} - \varphi_i,
\]
while \( \varphi_l = \varphi(\Phi^{-1}(l/n)) \), \( l = 1, \ldots, n - 1 \), and for \( l = 0 \) and \( l = n \) we set \( \varphi_0 = \varphi_n = 0 \), i.e. we defined the quantities as appropriate limits.

For \( j = 0, 1, 2, \ldots \) set

\[
\hat{\delta}_{j,n} = \int_0^1 R_n(t)h_j(t) \, dt. \tag{5}
\]

Then, by the properties of \( h_j \)'s, \( \hat{\delta}_{0,n} = \hat{\delta}_{1,n} = 0 \), while for \( j \geq 2 \), \( \hat{\delta}_{j,n} = \int_0^1 Q_n(t)h_j(t) \, dt \) and

\[
\hat{\delta}_{j,n} = \sum_{i=1}^n w_{ji}X_{i,n}, \quad \text{where} \quad w_{ji} = \int_{(i-1)/n}^{i/n} h_j(t) \, dt = \frac{1}{\sqrt{j}} \left[ \varphi_{i-1}h_{j-1} \left( \frac{i-1}{n} \right) - \varphi_i h_{j-1} \left( \frac{i}{n} \right) \right].
\]

The above shows that both the estimates \( \hat{\mu}_n \), \( \hat{\sigma}_n \) and Fourier coefficients \( \hat{\delta}_{j,n} \) of \( R_n \), \( j \geq 2 \), are \( L \)-statistics and the successive Fourier coefficients of \( Q_n \) in the system \( \{h_{j}\}_{j \geq 0} \). An important observation is that, for each \( j \geq 1 \), it holds \( \sum_{i=1}^n w_{ji} = 0 \), provided that, similarly as above, we set \( \varphi_0 h_1(0) = \varphi_0 h_1(1) = 0 \), \( l \geq 0 \). This observation implies, in particular, that \( \hat{\delta}_{j,n} \), \( j \geq 2 \), is invariant on change of shift. The formula for \( \hat{\sigma}_n \) is obviously a special case of Equation (6).

### 2.2. Some heuristics and basic components

If \( \int_{\mathbb{R}} x^2 \, dF(x) \) exists and is finite then the corresponding \( Q(t) \) belongs to \( L_2((0, 1), \, dt) \). Since the null hypothesis \( H \) asserts that \( Q(t) - \mu - \sigma \Phi^{-1}(t) = 0 \), having in mind the properties of \( \{h_{j}\}_{j \geq 0} \), for some fixed \( k \), consider an alternative of the form

\[
Q(t) - \mu - \sigma \Phi^{-1}(t) = \sum_{j=2}^{k+1} \delta_j h_j(t), \quad k \geq 1, \ \delta_j \in \mathbb{R}, \ j \geq 2.
\]

In such circumstances, we can replace \( H \) by the relation \( \sum_{j=2}^{k+1} \delta_j h_j(t) = 0 \) or, equivalently, by an auxiliary hypothesis

\[
H(k) : \delta_2 = \cdots = \delta_{k+1} = 0.
\]

To propose a statistic for testing \( H(k) \), introduce the components

\[
C_j = \frac{\sqrt{n(j+1)} \delta_{j,n}}{\hat{\sigma}_n}, \quad j \geq 2,
\]

where \( \delta_{j,n} \) is given by (5) and (6). Note that \( C_j \) is location and scale invariant. Moreover, under \( H \), \( C_j \) is asymptotically \( N(0, 1) \) while \( C_i \) and \( C_j \), \( i \neq j \), \( i, j \geq 2 \), are asymptotically uncorrelated, see the supplementary material for a formal proof of these facts. By the above, it seems to be natural to reject \( H(k) \) when

\[
N_k = \sum_{j=2}^{k+1} C_j^2
\]

is too large. This heuristics can be further supported by a formal argument; see LaRiccia and Mason [19] for general theory and LaRiccia [8] for its exemplification in the special case of testing normality under \( k = 2 \) and local alternatives of the form \( \sum_{j=2}^{3} (d_j/\sqrt{n})h_j(t) \), for some \( d_2 \) and \( d_3 \). The optimal weight \( w_0 \), introduced in [19], is then \( (\sqrt{3}d_2, \sqrt{4}d_3) \) and leads to the defined above statistic \( N_k \) with \( k = 2 \).

However, in practice, \( k \) is not known while the decision on \( k \) is critical to power behaviour of such a test. Therefore, in the next section, we propose and discuss some rule to decide about \( k \).
2.3. Selection of $k$ in $N_k$ and basic forms of the new test statistics

The statistic $N_k$ resembles Neyman’s smooth statistic for uniformity, based on $k$ empirical Fourier coefficients of $F$ in some orthonormal system on $(0,1)$. Ledwina [20] proposed to select $k$ in Neyman’s statistic via Schwarz selection rule while later on some simplified variant of the rule was introduced and investigated; see Inglot and Ledwina.[21] Simulations show that Schwarz rule and its simplified variant are very useful to select the number of components when only a few first Fourier coefficients are relatively large whereas the number of observations is small or moderate. Otherwise, the Schwarz penalty is too heavy. Since, in the present paper, we are aiming few first Fourier coefficients are relatively large whereas the number of observations is small or moderate. Therefore, we propose a more complex rule. Recall that small values of $R$ indicates $H$, $R_n$ tends slowly to $+\infty$.

The selection rules related to $R_n$ and the corresponding test statistics that we propose below, and adjust to practical use in Section 4, are designed to satisfy the following aims:

(i) reject null hypothesis in most cases when $R_n$ does and retain good overall power of $R_n$;
(ii) take more careful control of the components $\delta_{j,n}$’s when $R_n$ accepts $H$ and possibly improve $R_n$ in some situations;
(iii) in the case when the alternatives are symmetrical only exploit parsimonious set of components of $W_n^2$.

To describe our solutions step by step, let us introduce a sequence $\{d(n)\}$ of nondecreasing natural numbers, interpreted as the maximal number of components in $N_k$ which we shall consider given size of sample $n$. In other words, we can imagine that we model the function $Q - \mu - \sigma \Phi^{-1}$ considering nested family of models

$$M : \{\delta_2 h_2; \delta_2 \in \mathbb{R}\}, \{\delta_2 h_2 + \delta_3 h_3; \delta_2, \delta_3 \in \mathbb{R}\}, \ldots,$$

$$\{\delta_2 h_2 + \cdots + \delta_{d(n)} h_{d(n)+1}; \delta_2, \ldots, \delta_{d(n)+1} \in \mathbb{R}\}$$

of the successive dimensions 1, 2, \ldots, $d(n)$. Next, we define an auxiliary selection rule

$$A_c = \min\{k : N_k - ck \geq N_j - cj, \ k, j = 1, \ldots, d(n)\}.$$

The choice $c = 2$ leads to the Akaike’s penalty. Such a penalty is relatively light and a use of $N_A$ would result in relatively large critical values. This is not very profitable to power of the corresponding test. Therefore, we propose a more complex rule. Recall that small values of $R_n$ indicate $H$ and denote by $r_\alpha$ the critical value of such a test on the significance level $\alpha$. The new
selection rule $A$ depends, in particular, on $\mathcal{R}_n$ and the level $\alpha$, and is defined as follows:

$$A = \begin{cases} 
A_{c(n,\alpha)} & \text{if } \mathcal{R}_n \leq r_\alpha, \\
A_2 & \text{otherwise,}
\end{cases}$$

(8)

where $c(n, \alpha)$ is the smallest value of $c$ such that, under $H$, it holds that $P(A_c = 1) \geq 1 - \alpha$. In this place note that, in practice, $c(n, \alpha)$ is determined by simulations, basing on artificial samples of sizes $n$ from the $N(0, 1)$ distribution, as $N_k$ and $\mathcal{R}_n$ are location and scale invariant. So, in particular, the choice of $c(n, \alpha)$ is independent from the data at hand. Given $n$, $r_\alpha$ is also found by Monte Carlo experiment. Our simulations show that the penalty $c(n, \alpha)$ and analogous quantities defined below are much greater than 2. Therefore, the rule $A$ is restrictive when the ‘switch’ $\{\mathcal{R}_n \leq r_\alpha\}$ indicates $H$ and is liberal otherwise. This results in relatively small critical values of the corresponding tests and simultaneously allows for flexibility when the ‘switch’ indicates an alternative.

As mentioned in Section 1, in some applications, it is natural to restrict attention to symmetrical distributions, only. For such distributions $Q$ is asymmetrical and hence $\int_0^1 Q(t)h_2(t) \, dt = 0$, $l = 1, 2, \ldots$. Therefore, then it is enough to concentrate on odd functions and the related list of models

$$M^o : \{\delta h_3; \delta \in \mathbb{R}\}, \{\delta h_3 + \delta h_5; \delta, \delta_5 \in \mathbb{R}\}, \ldots,$$

$$\{\delta_3 h_3 + \cdots + \delta_2 d^o(n)+1 h_2 d^o(n)+1; \delta_3, \ldots, \delta_2 d^o(n)+1 \in \mathbb{R}\}$$

to approximate $Q - \mu - \sigma \Phi^{-1}$. In consequence, as above, introduce

$$N_k^o = \sum_{l=1}^k C_{2l+1}^2$$

and the related quantities $A^o_c$, $d^o(n, \alpha)$, and $A^o$. In particular,

$$A^o_c = \min\{k : N_k^o - ck \geq N_j^o - cj, \; k, j = 1, \ldots, d^o(n)\}.$$  

(9)

Our first solution rejects $H$ for large values of

$$N_A = \sum_{j=2}^{A+1} C_j^2$$

whereas the variant specialized to symmetrical cases rejects $H$ for large values of

$$N_{A'}^o = \sum_{l=1}^{A'} C_{2l+1}^2.$$  

3. Theoretical properties of the new tests of $H$ and preliminary comments

In this section, we assume that both $d(n)$ and $d^o(n)$, appearing in the definitions of $A_c$ and $A^o_c$, are independent from $n$ and are equal to fixed numbers $D$ and $D^o$, say. Such an assumption is sufficient in practical applications, as $D$ and $D^o$ can be arbitrary large. On the other hand, it allows us to skip difficult and laborious mathematical arguments, needed otherwise. Basic asymptotic
properties of the rules $A$ and $A^o$, and the corresponding test statistics $N_A$ and $N_{A^o}^o$ are collected in Propositions 1 and 2.

Proposition 1 says that asymptotic null distributions of the new test statistics are non-degenerate. Proposition 2 shows flexibility of the new selection rules under the alternatives and demonstrates that the new solutions are able to detect a large spectrum of alternatives, when sample size grows to infinity. More precisely, recall that by the argument of Section 2.2, a distribution function $F$ with finite second moment is an alternative to the null $N(\mu, \sigma)$ distribution, if $c_l = \int F^{-1}(\ell) h_l(t) \, dt \neq 0$ for some $l \geq 2$. Obviously, $l$ depends on $F$. Assume that $2 \leq l \leq D (3 \leq l \leq D^o$ and odd). This gives rise to introduce $\ell = \ell(F)$ ($\ell^o = \ell^o(F)$) being the smallest index $l \in \{2, \ldots, D\}$ ($l \in \{3, \ldots, D^o\}$ and odd) such that $c_l \neq 0$. The basic property of the selection rule $A (A^o)$ is that, for $n$ large enough, it chooses the dimension greater than or equal to $\ell - 1$ ($\ell^o - 2$) and includes the $\ell$th ($\ell^o$th) corresponding component into the test statistic. On the other hand, we shall show that, the components $C_l$ with indices $l$ such that $c_l \neq 0$ tend to infinity in probability and, in consequence of the aforementioned properties of $A (A^o)$, the test based on $N_A(N_{A^o}^o)$ is consistent, provided that $\ell \leq D (\ell^o \leq D^o)$.

Proofs of Propositions 1 and 2 appear in the supplementary material. A similar argument works and similar results are valid in the case of other data-driven statistics considered below.

**Proposition 1** Suppose that $d(n)$ and $d^o(n)$ do not depend on $n$ and are equal to $D$ and $D^o$, respectively. Then, under $H$ and $n \to \infty$, both $N_A$ and $N_{A^o}^o$ are bounded in probability.

**Proposition 2** Assume that $d(n)$ and $d^o(n)$ do not depend on $n$ and are equal to $D$ and $D^o$, respectively. Let $X$ has a distribution function $F$. Suppose that the second moment of $X$ exists and is finite. Then, for all alternatives obeying the corresponding $Q = F^{-1}$ such that $\int Q(t) h_l(t) \, dt \neq 0$ for some $2 \leq l \leq D (3 \leq l \leq D^o$ and odd), and $\ell$ and $\ell^o$ defined above, it holds that

(i) $\lim_{n \to \infty} P(A \geq \ell - 1) = \lim_{n \to \infty} P(A^o \geq \ell^o - 2) = 1$;

(ii) the test rejecting $H$ for large values of $N_A (N_{A^o}^o)$ is consistent.

Asymptotic behaviour of $N_A$ and $N_{A^o}^o$ under $H$ is a consequence of asymptotic normality of the components $C_j$, $j = 1, \ldots, D$, their asymptotic independence and Theorem 3 in [10]. Note that, however, the convergence in these asymptotic results is extremely slow. For some evidence, see Brown and Hettmansperger,[7] Krauczki.[22] and Tables S1 and S3 of the supplementary material. Therefore, in the next section, we introduce some finite sample corrections. Similar corrections are introduced in Section 6 in the case of testing simple null hypothesis $H^o$. The corrections improve substantially finite sample behaviour of the components $C_j$'s; cf. Tables S2 and S5 of the supplementary material. Moreover, since the selection rules $A$ and $A^o$ have built into the penalties which are adjusted to given $\alpha$ and $n$, relying on the slow asymptotic is further eliminated in this way, as well. In Section 4.1, below, and in Section 5 of the supplementary material we illustrate how qualitative asymptotic results on $A$ and $A^o$, presented in Proposition 2, manifest quantitatively in some finite sample situations.

### 4. Finite sample corrections, test statistics and simulation results

Table 2 of Brown and Hettmansperger [7] provides some evidence how slowly, under $H$, estimated mean and standard deviation of $C_3$, approach the limiting values, when sample size ranges from 5 to 100. The drawback does not vanish even for very large sample sizes; see Table S1 in the supplementary material for this article. Therefore, some finite sample corrections are welcome. For this purpose, given sample size $n$, we generate $n \tau$ independent realizations of the $N(0, 1)$
random variables and calculate the resulting values $C_{ij,0}$, $i = 1, \ldots, nr$, say, of $C_j$, $j \geq 2$. The notation $nr$ shall be used throughout for the number of Monte Carlo runs. However, note that the number $nr$ is not the same in all places.

We set $e_{j,0} = (nr)^{-1} \sum_{i=1}^{nr} C_{ij,0}$, $\frac{q^2}{v_{j,0}} = (nr)^{-1} \sum_{i=1}^{nr} (C_{ij,0} - e_{j,0})^2$ and

$$C_j = \frac{(C_j - e_{j,0})}{\frac{q}{v_{j,0}}}, \quad j \geq 2.$$  \hfill (10)

Since, under $H$, the distribution of $X_{1:n}, \ldots, X_{n:n}$ is symmetrical, therefore, $EC_{2l} = 0$ and one could skip the calculation of the related $e_{2l,0}$’s. However, to do not complicate the notation and presentation, we estimate all the quantities. Table S2 in the supplementary material shows that, under moderately large sample size, finite sample distribution of $C_j$’s, under $H$, is reasonably close to the limiting one $N(0, 1)$. Therefore, significance of single component can be easily noticed and verified.

With usage of the corrected variants $C_j$’s define

$$N_k = \sum_{j=2}^{k+1} C_j^2 \quad \text{and} \quad N_k^\alpha = \sum_{l=1}^{k} C_{2l+1}^2,$$

$$A_s = \min\{k : N_k - sk \geq N_j - sj, \ k, j = 1, \ldots, d(n)\}$$

and the counterpart $A^{\alpha}_s$ of $A_s$. Set $s(n, \alpha)$ and $s'(n, \alpha)$ to be the smallest $s$’s satisfying, under $H$, $P(A_s = 1) \geq 1 - \alpha$ and $P(A_s^{\alpha} = 1) \geq 1 - \alpha$, respectively. Finally, define the selection rule $A$ similarly as in (8). Its counterpart $A^{\alpha}$ is constructed analogously.

The resulting data-driven statistics $N_A$ and $N^{\alpha}_A$ shall be investigated in our simulation study presented below and in the supplementary material.

In this section, we consider moderately large sample sizes ranging from $n = 50$ to $n = 150$ and large $n$ equal to 2442. We compare four tests based on $R_n$, $N_A$, $N^{\alpha}_A$, and $QH$. The statistic $QH$ is the already mentioned solution by Chen and Shapiro [9] and is given by

$$QH = \frac{1}{\sqrt{n(n-1)}S_n} \sum_{i=1}^{n-1} \frac{X_{i+1:n} - X_{i:n}}{q_{i+1} - q_i},$$

where $q_i = \Phi^{-1}([i - 3/8]/[n + 1/4])$. This test rejects normality for small values of $QH$. Simulated critical values of these test statistics as well as the related values of $s(n, \alpha)$ and $s'(n, \alpha)$ are collected in Table 1. The table also contains Monte Carlo estimates, under $H$, of probabilities of the events $\{A_s(n, \alpha) = 1\}$ and $\{A^{\alpha}_{s'}(n, \alpha) = 1\}$ as well as $\{A = 1\}$ and $\{A^{\alpha} = 1\}$. Throughout the paper for each $n$ we took $d(n) = 20$ and $d'(n) = 10$.

Table 1 shows that the concentration of the selection rules $A$ and $A^{\alpha}$ on the first models of the respective lists $M$ and $M^{\alpha}$ is high under $H$. This results in relatively small critical values; cf. the comment in Section 2.3. To illustrate this comment, let us note that for instance, for $\alpha = 0.05$ and $n = 50$ the simulated critical value of the test rejecting $H$ for large values of $N_A$, would be 41.38. Observe also that, surprisingly enough, for moderate sample sizes, empirical critical values of $N^{\alpha}_A$ are much greater than those of $N_A$. We discuss this question in the supplementary material. Note also that for large sample sizes and in the case of testing simple null hypothesis $H^\alpha$, discussed in Section 6, the problem disappears.

4.1. Empirical powers under small and moderate sample sizes

We considered several alternatives both symmetrical and asymmetrical. Below, in Figures 1 and 2 we present six cases of symmetrical alternatives. The alternatives are described as follows. Let
Table 1. Simulated $(1 - \alpha)$-quantiles of $\mathcal{N}_A$, $\mathcal{N}_A^0$, $\mathcal{R}_n$, $\mathcal{QH}$ and the related quantities under $\mathbf{H}$, different $n$, $n \tau = 100,000$. Frequencies multiplied by 100.

| Significance level | $0.05$ | $0.01$ |
|--------------------|--------|--------|
| $n$                | 50     | 75     | 100    | 125    | 150    | 244$^2$ | 244$^2$ |
| Statistic          |        |        |        |        |        |         |         |
| $\mathcal{N}_A$    | Quantile | 8.456  | 8.305  | 8.296  | 7.718  | 7.780  | 7.960   | 11.542  |
| $s(n, \alpha)$     | 4.4    | 4.4    | 4.4    | 4.4    | 4.4    | 4.3     | 6.9     |
| $\hat{P}(A_{n, \alpha}) = 1$ | 93.02  | 93.12  | 93.04  | 93.35  | 93.29  | 93.12   | 98.47   |
| $\mathcal{N}_A^0$  | Quantile | 11.609 | 10.976 | 10.679 | 10.494 | 10.162 | 7.092   | 10.163  |
| $s^0(n, \alpha)$   | 4.8    | 4.8    | 4.8    | 4.7    | 4.7    | 4.4     | 7.0     |
| $\hat{P}(A_{n, \alpha}^0) = 1$ | 93.38  | 93.36  | 93.29  | 93.20  | 93.20  | 93.43   | 98.59   |
| $\mathcal{R}_n$    | 2.530  | 2.6357 | 2.708  | 2.786  | 2.832  | 3.721   | 4.711   |
| $\mathcal{QH}$     | 0.99589| 0.99762| 0.99839| 0.99887| 0.99916| 1.00001 | 1.000039|

Figure 1. Symmetrical alternatives. First row: medians of $C^j_{2j+1}$’s, $j = 1, \ldots, 10$, estimated under $n = 150$. Second row: empirical powers (in percentages) against $n$; $\alpha = 0.05$. $n \tau = 10,000$.

$U$ denote a random variable with the uniform distribution on $(0, 1)$ while $Z$ stands for a standard normal $N(0, 1)$ random variable. Then, $TU(\lambda_1, \lambda_2)$, $\lambda_1 > 0$, $\lambda_2 > 0$ denotes the distribution of a random variable $(U^{\lambda_1} - 1)/\lambda_1 - [(1 - U)^{\lambda_2} - 1]/\lambda_2$, i.e. $TU(\cdot, \cdot)$ is well-known Tukey’s family. $SU(g, d)$, $g \in \mathbb{R}$, $d > 0$, stands for the unbounded Johnson’s distribution of a random variable sinh$((Z - g)/d)$. $RT(k)$, $k \in \mathbb{N}$, is the distribution of a random variable $Z/U^{1/(k+1)}$, see Rogers and Tukey.[23] $M(\varepsilon, m)$, $\varepsilon \in (0, 1)$, $m \in \mathbb{R}$, denotes the distribution of the mixture $(1 - \varepsilon)N(0, 1) + (\varepsilon/2)N(-m, 1) + (\varepsilon/2)N(m, 1)$. $Cos(j, d)$, $j \in \mathbb{N}$, $d \in [0, 1]$, has the density $[1 + d \cos(\pi j \Phi(x))]\varphi(x)$, $x \in \mathbb{R}$, where $\varphi$ and $\Phi$ are the density and the distribution function of the standard normal law, cf. Inglot et al.[24]

The first three cases, displayed in Figure 1, concern distributions considered by Krauczi.[22] She, in turn, followed Shapiro et al. [25] and Gan and Koehler.[26] In course of our simulations, for each alternative we estimated medians of the components $C_j$’s. Note that, given $n$ and $j$, each
The median is simply calculated from values of $C_j$ obtained in the successive 10,000 Monte Carlo runs. We present medians, not the averages, as we investigated and included into our study some long-tailed alternatives. Next, immediate observation is that for many different alternatives the components $C_j$’s behave similarly. Therefore, in particular, the components are a basis to select really different cases. In Figure 2, we present three relatively complex (in terms of the components) and interesting symmetrical cases. Results for three other symmetrical cases from Krauczi [22] as well as for six asymmetrical alternatives are shown in Figures S2–S4 in the supplementary material. All these figures have the same structure. We summarize the data by presenting the estimated medians of $C_j$’s and below we display empirical powers for $n = 50, \ldots, 150$. In this study, $\alpha = 0.05$.

The results show that, in general, empirical behaviour of $QH$, $R_n$, and $N_A$ is similar. In the cases $A_3$, $A_5$, and $A_6$, $N_A$ is slightly better than $R_n$. $N^{\alpha}_{A^o}$ provides considerable gain in power in case of symmetrical alternatives that are picked ($A_3$) or multi-modal ($A_5, A_6$). The alternative $A_5$ represents here normal heterogeneous mixtures while $A_6$ describes some sinusoidal deviation from $H$. The importance of detecting mixtures is extensively discussed in [27], for example, while several sinusoidal signal models are considered in the literature on microwave radiometry; see Tarongi and Camps [5] for an overview. Moreover, in some applications the symmetry is naturally built-into the underlying model. For some illustration, see De Roo et al. [28] and the models discussed in [3]. In such cases, $N^{\alpha}_{A^o}$ can be safely applied. In case of asymmetrical alternatives, use of $N^{\alpha}_{A^o}$ is obviously not sensible.

Now, let us discuss briefly the behaviour of the selection rule $A^o$ under three representative alternatives and an impact of the rule into empirical power of $N^{\alpha}_{A^o}$.

Table 2 contains the illustrative frequencies ($n = 150$) of attaining the successive dimensions, among $1, \ldots, 10$, under consideration, by the rule $A^o$ in the case of alternatives $A_1, A_5$, and $A_6$.

The outcomes reflect changes of the consecutive cumulative sums of squared components influenced mainly by the magnitude and the index of significantly large components. To have more detailed insight into the role of the rule $A^o$ in possible improvement of the power of the test based on $R_n$ let us introduce three ‘counters’ of three important events in our Monte Carlo repetitions of simulation experiments: $K_1$ giving the percentage of cases when $\{R_n > r_\alpha\}$ takes
place; $K_2$ summarizing the percentage of cases when the test based on $\mathcal{N}_{A^o}$ rejects $H$ in the situation when the one based on $\mathcal{R}_n$ also does it; i.e. when $\{\mathcal{R}_n > r_\alpha\}$ holds; $K_3$ counting the cases when $\mathcal{N}_{A^o}$ rejects $H$, in the situation when $\mathcal{R}_n$ accepts it. Obviously, $K_1$ coincides with the empirical power of the test rejecting $H$ for large values of $\mathcal{R}_n$ while $K_2 + K_3$ equals the empirical power of the new test. Table 3 collects some evidence in the case of alternatives $A_1$, $A_5$, and $A_6$. Observe that $K_2$ is close to $K_1$, as the penalty in $A^o$ is small when $\{\mathcal{R}_n > r_\alpha\}$. The quantity $K_3$ exhibits the gain in empirical power due to more careful inspection of components in the cases when the test based on the standardized Wasserstein distance $\mathcal{R}_n$ accepts $H$.

Table 3 shows that if only few first odd components are relatively large, then the selection rule only provides little extra impact to the power of $\mathcal{N}_{A^o}$, as in such cases almost all information was caught by $\mathcal{R}_n$. However, if components of higher orders play the main role the rule $A^o$ recovers to some extent the information which was partially lost by $\mathcal{R}_n$, due to the inexplicit weighting built-into it. Obviously, the magnitude of components essentially influences the counts $K_3$. Therefore, if the signal is weaker then the ability of $A^o$ to repair some weakness of $\mathcal{R}_n$ shall be smaller. The above illustrates the way in which we fulfilled the postulates stated in Section 2.3.

We also studied empirical powers of tests based on two first components. Since the results were as it could be expected, we are not presenting them here.

We close with discussing useful relation of tests based on $\mathcal{R}_n$ and $Q\mathcal{H}$. Namely, note that the tests are very similar in structure. Indeed, by Equations (4) and (7) it holds that $0 \leq \hat{\sigma}_n \leq S_n$ and $\mathcal{R}_n$ rejects the null hypothesis when

$$\mathcal{R}_n^o \doteq \frac{\sum_{i=1}^{n-1} \psi_i (X_{i+1:n} - X_{i:n})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2}} = \frac{\sum_{i=1}^{n} b_i X_{i:n}}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2}}$$

is too small. The test based on $Q\mathcal{H}$ rejects $H$ for small values of

$$Q\mathcal{H}^o = \frac{\sum_{i=1}^{n-1} \psi_i (X_{i+1:n} - X_{i:n})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2}} = \frac{\sum_{i=1}^{n} a_i X_{i:n}}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2}},$$

where $\psi_i = 1/[q_{i+1} - q_i]$, $i = 1, \ldots, n - 1$ while $a_1 = -\psi_1$, $a_i = \psi_{i-1} - \psi_i$, $i = 2, \ldots, n - 1$, $a_n = \psi_{n-1}$. Therefore, possible differences in the behaviour of $\mathcal{R}_n$ and $Q\mathcal{H}$ are consequences
of the underlying coefficients of the successive order statistics. Question of that kind, in the case of comparison of $\mathcal{QH}$ with some celebrated competitors, was already discussed in [9] and further studied in [29,30]. Following these important contributions, we consider the normalized coefficients $a_i^\circ = (a_i/\sqrt{\sum_{j=1}^{n} a_j^2})$ and $b_i^\circ = (b_i/\sqrt{\sum_{j=1}^{n} b_j^2})$ and call them the weights. Bai and Chen,[29, p.487] concluded that:

comparing the weights ... one finds that tests with large weights for extreme order statistics are more powerful to detect short tailed alternatives and tests with small weights for the extreme order statistics are more powerful to detect heavy-tailed alternatives.

Therefore, to have some insight into the present situation, we calculated the ratios $r_i(a, b) = a_i^\circ/b_i^\circ$ of the weights. For some comparison to earlier results discussed in [30]; in Table 4, we present the ratios for $n = 10, 20, 30$ and $i = \lfloor n/2 \rfloor + 1, \ldots, n$. Moreover, for $n = 50, 150$, and $n = 244^2$, i.e. for the selection of sample sizes considered above and in Section 4.2, we display 14 ratios for $i = \lfloor n/2 \rfloor + 1, \ldots, \lfloor n/2 \rfloor + 7$ and for $i = n - 6, \ldots, n$, only. In all the considered cases, we observed that, given $n$, $r_i(a, b)$ are slowly growing with $i$, for $i$ large enough, $r_{[n/2]+1}(a, b) < 1$ and $r_n(a, b) > 1$.

On the other hand, the distributions in $\mathcal{A}_1$, $\mathcal{A}_5$, and $\mathcal{A}_6$ have kurtosis less than 3. For these alternatives, $\mathcal{QH}$ outperforms $\mathcal{R}_n$. In the case of long-tailed alternatives, $\mathcal{A}_2$ and $\mathcal{A}_4$ the situation is reversed. For $\mathcal{A}_3$ the kurtosis is close to three while empirical powers of $\mathcal{QH}$ and $\mathcal{R}_n$ are almost identical. Therefore, the above along with the evidence in Table 4 further support usefulness of the quoted guideline of Bai and Chen.[29]

### 4.2. Empirical powers when detecting symmetrical mixtures and sums, large $n$

In this section, we impose the set-up similar to that described in [2]. In particular, now $n = 244^2$ and alternatives are sparse heterogeneous mixture $M(\cdot,\cdot)$, defined above, and linear combination of $N(0, 1)$ variable and Pareto one; denoted here by $P(\cdot,\cdot)$. More precisely, $P(\lambda, a)$, $\lambda > 0$, $a > 0$, stands for the distribution of $\sqrt{1-\lambda}Z + \sqrt{\lambda}W$ where $Z \sim N(0, 1)$, $W$ has the symmetrical Pareto distribution with the tail index $a$, while $Z$ and $W$ are independent. The density of $W$ has the form $p(w) = (a/2)|w|^{-1-a} \times 1(|w| > 1)$, where $1(\cdot)$ is the indicator of the set $\cdot$; cf. Grabchak and Samorodnitsky.[31]
In this section, we consider two significance levels: .05 and .01. As before, \( d(n) = 20 \) and \( d'(n) = 10 \). In contrast to Jin et al.,[2] here the null hypothesis is a composite one.

Figure 3 shows that the test based on the statistic \( N_A^0 \), specialized to detect symmetrical alternatives to \( H \), has stable empirical power that compete to that of omnibus tests considered in our study. However, in contrast to some cases shown in Section 4.1, the gain of power of this new test is rather moderate. The reason is that, in the cases studied in this section, one of the two ‘early’ Fourier coefficients on the underlying list, i.e. \( \hat{\delta}_3, n \) or \( \hat{\delta}_5, n \), is dominating. Moreover, the magnitude of \( \hat{\delta}_5, n \) is rather moderate. In such circumstances, substantial improvement over the test based on \( R_n \) is hard to get. We shall see that situation changes in the case of these and similar alternatives when testing simple null hypothesis. Simple null hypothesis was considered in [2] and is discussed in Section 6.

5. Real data examples

First, we considered \( n = 20 \) measurements of percentage of silt from the successive plots in the depth 3, presented in [32, p.121]. It is of interest to test for normality.

We summarize the data in Figure 4 via normal Q–Q plot and empirical components \( C_2, \ldots, C_{21} \). The Q–Q plot provides a warning that the data distribution and theoretical normal one do not match very well and suggests that empirical distribution has shorter tails. It also exhibits some symmetry in data distribution. The components supplement these findings also suggesting symmetry of the empirical distribution. Moreover, the magnitude of some of components provides quite strong evidence of non-normality. Table 5 summarizes briefly the outcomes of our formal tests. It can be seen that \( N_A^0 \) detects non-normality on the level \( \alpha = 0.03 \) while \( N_A \) on
Figure 4. Normal Q–Q plot and components $C_j, j = 2, \ldots, 21$ for the silt data.

Table 5. The behaviour of the statistics $N_A$ and $N_{A^0}$ for the silt data.

| Statistic          | $\alpha$ | 0.14 | 0.13 | 0.12 | 0.10 | 0.05 | 0.04 | 0.03 | 0.02 |
|--------------------|----------|------|------|------|------|------|------|------|------|
| $R_n$              | $r_n$    | 1.728| 1.763| 1.799| 1.881| 2.199| 2.303| 2.435| 2.619|
| $N_A$              | Quantile | 6.381| 6.637| 7.032| 8.477| 10.712| 11.465| 12.101| 13.788|
| $x(n, \alpha)$    |          | 3.0  | 3.1  | 3.2  | 3.4  | 4.4  | 4.7  | 5.2  | 5.9  |
| $N_{A^0}$          | Quantile | 8.403| 9.082| 9.693| 10.623| 14.850| 15.780| 16.947| 19.039|
| $x'(n, \alpha)$   |          | 3.2  | 3.3  | 3.4  | 3.7  | 4.8  | 5.2  | 5.7  | 6.3  |
| $N_A$              | Value of statistic 22.025 | 17.861 | 5.395 | 5.395 | 5.395 | 5.395 | 5.395 | 5.395 | 0.079 |
| $N_{A^0}$          | Value of statistic 21.827 | 21.827 | 21.827 | 21.827 | 17.748 | 17.748 | 17.748 | 5.316 |
| $N_A$              | Value of the rule $A$ 8 | 6 | 2 | 2 | 2 | 2 | 2 | 1 |
| $N_{A^0}$          | Value of the rule $A^0$ 4 | 4 | 4 | 4 | 4 | 3 | 3 | 3 | 1 |

Note: Quantiles and parameters calculated on $10^5$ MC runs.

Figure 5. Normal Q–Q plot and components $C_j, j = 2, \ldots, 21$ for the tephra data.

the level $\alpha = 0.13$. For comparison, we also calculated $p$-values of $R_n$, $QH$ as well as the Shapiro–Wilk (SW), Anderson–Darling (AD), and Cramér–von Mises (CvM) tests obtaining 0.13, 0.09, 0.11, 0.20, and 0.29, respectively. So, $N_A$ and these tests do not detect significant non-normality, when applying standard significance levels.

Next, we considered the tephra data described and analysed by Bowman and Azzalini,[33, pp.27, 38–40]. As there, we applied logistic transformation to the original data. The $n = 59$ outcomes are summarized in Figure 5. Normal Q–Q plot indicates some skewness. The observed magnitude of some of components $C_j, j$-even, supports existence of such feature of the data.
The tests $R_n$, $QH$, $SW$, $AD$, and $CvM$ when applied to these data lead to $p$-values: .12, .14, .13, .03, and .02, respectively. So, $CvM$ and $AD$ indicate significant non-normality on relatively small levels. In contrast, the first three tests lead to non-significant $p$-values. Similar outcomes result from usage of $N_A$ and $N_A^*$, cf. Table 6, though $N_A$ appears to be slightly more sensitive.

6. Some remarks on testing simple null hypothesis

Since recent years testing simple null hypothesis

$$H^*: Q = \Phi^{-1}$$

received considerable attention as well, see Donoho and Jin [18] and the consecutive papers; in this section, we introduce and discuss some counterparts of $N_A$, $N_A^*$, and $R_n$ adjusted to $H^*$, and compare them to the higher criticism statistic introduced in the aforementioned paper.

Natural counterparts of $R_n$, $W_n$, and $R_n$ have the form

$$R_n^*(t) = Q_n(t) - \Phi^{-1}(t), \quad W_n^* = \left\{ \int_0^1 (R_n^*(t))^2 \, dt \right\}^{1/2} \quad \text{and} \quad R_n^* = n(W_n^*)^2.$$ 

Let $\hat{\delta}_{0,n}^*, \ldots, \hat{\delta}_{k,n}^*$ minimize over $\delta_0, \ldots, \delta_k$, $\delta_l, l = 0, \ldots, k$, the expression $\int_0^1 (R_n^*(t) - \sum_{j=0}^k \delta_j h_j(t))^2 \, dt$. Then,

$$\hat{\delta}_{j,n}^* = \int_0^1 R_n^*(t) h_j(t) \, dt.$$ 

Hence, due to the properties of $h_j$’s and Equations (3), (4), (6),

$$\hat{\delta}_{0,n}^* = \hat{\mu}_n, \quad \hat{\delta}_{1,n}^* = \hat{\sigma}_n - 1, \quad \hat{\delta}_{j,n}^* = \hat{\delta}_{j,n}, \quad j \geq 2,$$

and $\{W_n^*\} = (\hat{\delta}_{0,n}^*)^2 + (\hat{\delta}_{1,n}^*)^2 + W_n^2$. (11)

Set $C_j^* = \sqrt{n(j + 1)} \hat{\delta}_{j,n}^*$ and denote finite sample corrections of the mean and the standard deviation of $C_j^*$, under $H^*$, by $e_{j,0}^*$ and $v_{j,0}^*$, respectively. Now, define

$$C_j^* = \frac{C_j^* - e_{j,0}^*}{v_{j,0}^*}$$

as well as

$$N_k^* = \sum_{j=0}^{k-1} (C_j^*)^2 \quad \text{and} \quad N_k^{\alpha*} = \sum_{l=0}^{k-1} (C_{2l+1}^*)^2.$$ 

Analogously as before also define $A^*$ and $A^{\alpha*}$ and the related data-driven statistics $N_A^*$ and $N_A^{\alpha*}$. For completeness, we give some formulas, the related corrections and some critical values in Section 4 of the supplementary material.
Table 7. Simulated powers (in percentages) of the tests for verifying $H^*$ against symmetrical and asymmetrical alternatives; $\alpha = 0.05, n = 75, n_\epsilon = 10,000$.

| Alternative   | $N^*_A$ | $N^*_A^{\epsilon o}$ | $R^*_n$ | $HC$ | $K^*$ |
|---------------|---------|----------------------|---------|------|------|
| $TU(0.5, 0.5)$ | 11      | 38                   | 7       | 0    | 52   |
| $SU(0, 1.5)$   | 70      | 73                   | 53      | 16   | 63   |
| $Cos(4, 0.6)$  | 39      | 57                   | 23      | 12   | 3    |
| $t(5)$         | 77      | 76                   | 71      | 73   | 59   |
| $LC(0.05, 3)$  | 71      | 63                   | 66      | 56   | –    |
| $TruncN(-2, 1)$| 100     | 100                  | 100     | 1    | –    |
| $Cos(3, 0.6)$  | 50      | 23                   | 34      | 11   | –    |

Following Jin et al. [2] we consider the higher criticism statistic of the form

$$HC = \sqrt{n} \sup_{x \in \mathbb{R}} \frac{|F_n(x) - \Phi(x)|}{\sqrt{\Phi(x)[1 - \Phi(x)]}},$$  \hspace{1cm} (12)

where $F_n$ stands for the empirical distribution function, and the two-sided test based on the empirical kurtosis given by

$$K^* = \sqrt{\frac{n}{24}} \left( \frac{n^{-1} \sum_{i=1}^{n} X_i^4}{n^{-1} \sum_{i=1}^{n} X_i^2} - 3 \right).$$

The formula (12) can be recognized as the well-known Jaeschke [34] statistic.

Tables 7 and 8 contain some simulated powers under $n = 75$ (symmetrical and asymmetrical cases) and $n = 244^2$ (symmetrical alternatives). The alternatives mostly comes from the set considered and discussed in Section 4. Three additional ones are defined as follows. $t(5)$ is the Student distribution with 5 degrees of freedom. $LC(\epsilon, m)$ has the density $\epsilon \phi(x - m) + (1 - \epsilon) \phi(x), \epsilon \in (0, 1), x \in \mathbb{R}$, where $\phi$ is the standard normal density. TruncN$(a, b), a \in \mathbb{R}, b \in \mathbb{R}, a < b$, stands for the standard normal distribution truncated at $a$ and $b$.

The results show that the new data-driven tests perform well in the study. Moreover, in contrast to the case of $H$, when testing $H^*$, under $n = 244^2, N^*_A$ outperforms $R^*_n$ considerably. This is mainly caused by two facts. Due to (11), critical values of the test based on $R^*_n$ are much greater than those ones based on $R_n$ while, in case of many alternatives, the first two empirical Fourier coefficients $\hat{\delta}^*_0, n$ and $\hat{\delta}^*_1, n$ are small and their impact into the empirical power does not compensate for the increase in critical values. For the data-driven solutions $N^*_A$, and $N^*_A^{\epsilon o}$, critical values under $n = 75$ are much smaller while under $n = 244^2$ are similar to or even slightly less than respective values under $H$, while the selection rules provide considerable amount of additional information. We present some selected simulation results and some further discussions on this point in Section 5 of the supplementary material.

The test based on $K^*$ is one of the leading solutions in the cases when the component $C^*_3$ is the largest one. However, in the situations when $C^*_3$ is close to 0 the kurtosis $K^*$ completely breaks down. For values of $C^*_j$, related to alternatives in Table 8, see Figure S5 in the supplementary material.

Empirical power of $HC$ is unstable and, therefore, the test is not a highly reliable omnibus goodness-of-fit procedure. Jin et al. [2, p.2482] mention that ‘when the actual evidence lies in the middle of the distribution’, then $HC$ will be very weak. This corresponds with earlier theoretical developments of Eicker,[35, p.117] who infers on sensitivity of (12) only in moderate tails and Révész,[36] who also suggests sensitivity of (12) for some changes in the tails only. The above findings are good motivation to propose new solutions behaving stably under a large
Table 8. Simulated powers (in percentages) of the tests for verifying $H^*$ against symmetrical alternatives; $\alpha = 0.01$, $n = 2442$, $n_r = 10,000$.

| Alternative | $N_{A*}^*$ | $N_{A*}^{\text{cov}}$ | $R_n^*$ | $\mathcal{HC}$ | $K^*$ |
|-------------|------------|-----------------|--------|--------------|------|
| $M(0.0015, 2.5)$ | 31 | 61 | 26 | 23 | 58 |
| $M(0.00025, 3.5)$ | 17 | 40 | 15 | 41 | 27 |
| $M(0.000075, 4.5)$ | 23 | 41 | 22 | 54 | 22 |
| $P(0.14, 11.0)$ | 99 | 99 | 97 | 1 | 54 |
| $P(0.03, 5.0)$ | 67 | 65 | 42 | 2 | 1 |
| $P(0.02, 4.1)$ | 62 | 60 | 38 | 3 | 2 |

Spectrum of alternatives. It seems that the constructions proposed in this paper meet such a goal to a satisfactory extent.

Similar solutions can be constructed for testing of fit to $N(0, \sigma)$ and $N(\mu, 1)$, $\mu, \sigma$-unknown, distributions.

7. Conclusion

We have proposed some data-driven tests to detect departures from Gaussianity. The main focus is on composite null hypothesis. There are many kinds of non-Gaussianity and many kinds of tests to detect them. Several tests focus on specific deviations and are not sensitive to others. Our goal was to propose solutions which are sensitive to commonly studied deviations from Gaussianity and which, simultaneously, are able to detect some complex alternatives. Our constructions have simple interpretation in terms of normalized Fourier coefficients of the sample quantile function. These normalized coefficients are called components and obey approximately the standard normal distribution under the null hypothesis. The number of components, which is taken into consideration, depends on the data at hand and is defined via Akaike-type selection rule. The resulting solutions prove to behave stably and efficiently under a large spectrum of alternatives and compare well to the best existing competitors. Similar approach can be in principle applied to some other location/scale families. This, however, requires some further careful investigation. Note that De Wet [37] provides insightful discussion on tests for location and scale families based on a weighted Wasserstein distance. The paper by Csörgő [38] further extends these considerations.

Supplemental data

Supplemental data for this article can be accessed http://dx.doi.org/10.1080/00949655.2014.983110.

Acknowledgments

This work was supported by Grant N N201 608440 from the National Science Centre, Poland. Calculations have been carried out in Wroclaw Centre for Networking and Supercomputing (http://www.wcss.wroc.pl), grant No. 199. We are deeply grateful for kind cooperation of the Centre. We also thank the referee for useful remarks.

References

[1] Graham P, Turok N, Lubin PM, Schuster JA. A simple test for non-Gaussianity in cosmic microwave background radiation measurements. Astrophys J. 1995;449:404–412.
[2] Jin J, Starck J-L, Donoho DL, Aghanim N, Forni O. Cosmological non-Gaussian signature detection: comparing performance of different statistical tests. EURASIP J Appl Signal Process. 2005;15:2470–2485.
[3] Güner B, Frankford MT, Johnson JT. A study of the Shapiro-Wilk test for the detection of pulsed sinusoidal radio frequency interference. IEEE Trans Geosc Rem Sens. 2009;47:1745–1751.
[4] Romão X, Delgado R, Costa A. An empirical power comparison of univariate goodness-of-fit tests for normality. J Statist Comput Simul. 2010;80:545–591.
[5] Tarongi JM, Camps A. Normality analysis for RFI detection in microwave radiometry. Remote Sens. 2010;2: 191–210.
[6] Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965;52: 591–611.
[7] Brown BM, Hettmansperger TP. Normal scores, normal plots, and tests for normality. J Am Statist Assoc. 1996;91:1668–1675.
[8] LaRiccia VN. Optimal goodness-of-fit tests for normality against skewness and kurtosis alternatives. J Statist Plan Inference. 1986;13:67–79.
[9] Chen L, Shapiro SS. An alternative test for normality based on normalized spacings. J Statist Comput Simul. 1995;53:269–288.
[10] Del Barrio E, Cuesta-Albertos JA, Matrán C, Rodríguez-Rodríguez JM. Tests of goodness of fit based on the $L_2$-Wasserstein distance. Ann Statist. 1999;27:1230–1239.
[11] Shapiro SS, Francia RS. An approximate analysis of variance test for normality. J Am Statist Assoc. 1972;67: 215–216.
[12] De Wet T, Venter JH. Asymptotic distributions of certain test criteria of normality. S Afr Statist J. 1972;6: 139–149.
[13] Aly E-EAA, M Csörgő. Quadratic nuisance-parameter-free goodness-of-fit tests in the presence of location and scale parameters. Canad J Statist. 1985;13:53–70.
[14] LaRiccia VN, Mason DM. Cramér-von Mises statistics based on the sample quantile function and estimated parameters. J Multivar Anal. 1986;18:93–106.
[15] Puri ML, Rao CR. Augmenting Shapiro–Wilk test for normality. In: Walter Joh. Ziegler, editor. Contributions to applied statistics, a festschrift to A. Linder. New York: Birkhäuser; 1976. p. 129–139.
[16] LaBrecque J. Goodness-of-fit tests based on nonlinearity in probability plots. Technometrics. 1977;19:293–306.
[17] LaRiccia VN. Smooth goodness-of-fit tests: A quantile function approach. J Am Statist Assoc. 1991;86:427–431.
[18] Donoho DL, Jin J. Higher criticism for detecting sparse heterogeneous mixtures. Ann Statist. 2004;32: 962–994.
[19] LaRiccia VN, Mason DM. Optimal goodness-of-fit tests for location/scale families of distributions based on the sum of squares of $L$-statistics. Ann Statist. 1985;13:315–330.
[20] Ledwina T. Data-driven version of Neyman’s smooth test of fit. J Amer Statist Assoc. 1994;89:1000–1005.
[21] Inglot T, Ledwina T. Intermediate approach to comparison of some goodness-of-fit-tests. Ann Inst Statist Math. 2001;53:810–834.
[22] Krauzci É. A study of the quantile correlation test for normality. TEST. 2009;18:156–165.
[23] Rogers WH, Tukey JW. Understanding some long-tailed symmetrical distributions. Statist Neerlandica. 1972;26:211–226.
[24] Inglot T, Jurlęwicz T, Ledwina T. On Neyman-type smooth tests of fit. Statistics. 1990;21:549–568.
[25] Shapiro SS, Wilk MB, Chen HJ. A comparative study of various tests for normality. J Am Statist Assoc. 1968;63:1343–1372.
[26] Gan FF, Koehler KJ. Goodness-of-fit tests based on P–P probability plots. Technometrics. 1990;32:289–303.
[27] Cai TT, Jeng XJ, Jin J. Optimal detection of heterogeneous and heteroscedastic mixtures. J R Statist Soc B. 2011;73:629–662.
[28] De Rood RD, Misra S, Ruf CS. Sensitivity of the kurtosis statistic as a detector of pulsed sinusoidal RFI. IEEE Trans Geosc Remote Sens. 2007;45:1938–1946.
[29] Bai ZD, Chen L. Weighted $W$ test for normality and asymptotics: a revisit of Chen-Shapiro test for normality. J Statist Plan Inference. 2003;113:485–503.
[30] Chen L. C470. Normality tests based on weighted order statistics. J Statist Comput Simul. 2003;73:603–606.
[31] Grabchak M, Samorodnitsky G. Do financial returns have finite or infinite variance? A paradox and an explanation. Quant Financ. 2010;10:883–893.
[32] Andrews DF, Herzberg AM. Data: a collection of problems from many fields for student and research worker. New York: Springer; 1985.
[33] Bowman A, Azzalini A. Applied smoothing techniques for data analysis: the kernel approach with S-plus illustrations. Oxford: Clarendon Press; 1997.
[34] Jaeschke D. The asymptotic distribution of the supremum of the standardized empirical distribution function on subintervals. Ann Statist. 1979;7:108–115.
[35] Eicker F. The asymptotic distribution of the suprema of the standardized empirical processes. Ann Statist. 1979;7:116–138.
[36] Révész P. A joint study of the Kolmogorov–Smirnov and the Eicker–Jaeschke statistics. Statist Decis. 1982;1: 57–65.
[37] De Wet T. Goodness-of-fit tests for location and scale families based on a weighted $L_2$-Wasserstein distance measure. TEST. 2002;11:89–107.
[38] S Csörgő. Weighted correlation tests for location-scale families. Math Comput Model. 2003;38:753–762.