Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: http://www.tandfonline.com/loi/cjas20

# A powerful and interpretable alternative to the Jarque–Bera test of normality based on 2nd-power skewness and kurtosis, using the Rao's score test on the APD family

A. Desgagné & P. Lafaye de Micheaux

**To cite this article:** A. Desgagné & P. Lafaye de Micheaux (2018) A powerful and interpretable alternative to the Jarque–Bera test of normality based on 2nd-power skewness and kurtosis, using the Rao's score test on the APD family, Journal of Applied Statistics, 45:13, 2307-2327, DOI: 10.1080/02664763.2017.1415311

**To link to this article:** https://doi.org/10.1080/02664763.2017.1415311

JOURNAL OF APPLIED STATISTICS, 2018
VOL. 45, NO. 13, 2307–2327
https://doi.org/10.1080/02664763.2017.1415311

Taylor & Francis
Taylor & Francis Group



# A powerful and interpretable alternative to the Jarque–Bera test of normality based on 2nd-power skewness and kurtosis, using the Rao's score test on the APD family

A. Desgagné[a] and P. Lafaye de Micheaux[b]

[a]Département de mathématiques, Université du Québec à Montréal, Montréal, Québec, Canada; [b]School of Mathematics and Statistics, UNSW Sydney, Australia

**ABSTRACT**
We introduce the 2nd-power skewness and kurtosis, which are interesting alternatives to the classical Pearson's skewness and kurtosis, called 3rd-power skewness and 4th-power kurtosis in our terminology. We use the sample 2nd-power skewness and kurtosis to build a powerful test of normality. This test can also be derived as Rao's score test on the asymmetric power distribution, which combines the large range of exponential tail behavior provided by the exponential power distribution family with various levels of asymmetry. We find that our test statistic is asymptotically chi-squared distributed. We also propose a modified test statistic, for which we show numerically that the distribution can be approximated for finite sample sizes with very high precision by a chi-square. Similarly, we propose a directional test based on sample 2nd-power kurtosis only, for the situations where the true distribution is known to be symmetric. Our tests are very similar in spirit to the famous Jarque–Bera test, and as such are also locally optimal. They offer the same nice interpretation, with in addition the gold standard power of the regression and correlation tests. An extensive empirical power analysis is performed, which shows that our tests are among the most powerful normality tests. Our test is implemented in an R package called PoweR.

**ARTICLE HISTORY**
Received 1 December 2016
Accepted 5 December 2017

**KEYWORDS**
Normality test; asymmetric power distribution; Lagrange multiplier test; Rao's score test; skewness; kurtosis

## 1. Introduction

An important issue in statistics is the validity of the normality assumptions that are often required for the use of many popular methods of statistical analysis. Consequently, the problem of testing that a sample has been drawn from some normal distribution with unknown mean and variance is one of the most common problems of goodness of fit in statistical practice. For this reason, many test procedures have been proposed in the literature.

**CONTACT** P. Lafaye de Micheaux lafaye@unsw.edu.au School of Mathematics and Statistics, UNSW Sydney, NSW 2052, Australia

Supplemental data for this article can be accessed here. https://doi.org/10.1080/02664763.2017.1415311

A comprehensive power comparison study of 33 existing tests for normality can be found in [19], with a brief review of each test (see also [23]).

It is generally accepted that the regression and correlation tests are the most powerful, in particular the test of Shapiro and Wilk [21] (or its extensions, see for instance [20]) that is widely used in practice, or the tests of Chen and Shapiro [4] and Del Barrio *et al.* [7]. For the situations where it is known that the true distribution is symmetric, the directional test of Coin [5] emerges as the most powerful, according to the empirical study of Romão et al. [19]. Some tests based on the empirical distribution function are also powerful, in particular, the $Z_A$ and $Z_C$ tests of Zhang and Wu [24] and the test of Anderson and Darling [1]. However, apart from the Shapiro–Wilk and Anderson–Darling tests, these tests rely on simulated quantiles, which may limit their implementation.

Despite the remarkable qualities of the Shapiro–Wilk test, another test is widely used, especially in the econometric fields. This is the well-known Jarque–Bera (JB) test [12], first proposed by Bowman and Shenton [3] and based on sample Pearson's skewness and kurtosis, respectively, estimates of third and fourth standardized moments. Practitioners often see statistical procedures as decision aid tools and therefore require transparent methods that are easily interpretable. When normality is rejected using the JB test, one also obtains information on the process: the distribution may be skewed to the right (or to the left) and/or exhibit long (or short) tails. This knowledge is often valuable to users, and this feature may explain the popularity of the JB test, even if it has some power issues. The tests of D'Agostino and Pearson [6] and Doornik and Hansen [8], which combine different normalizing transformations of skewness and kurtosis, generally seem to be slightly more powerful than the JB test.

In this paper, we first propose a quasi omnibus test that presents the advantages of the JB test, with in addition the gold standard power of the regression and correlation tests. We also derive a directional test with the same benefits, for the situations where the true distribution is known to be symmetric. Our starting point was based on the idea that kurtosis can be measured in more than one way. Geary [9] proposed to use the first standardized sample moment $n^{-1}\sum_{i=1}^{n}|Z_i|$ as an alternative to the classical sample Pearson's kurtosis (defined as the fourth standardized sample moment $n^{-1}\sum_{i=1}^{n}|Z_i|^4$), where $Z_i$ represents the standardized observations. Note that Bonett and Seier [2] revisited the measure of Geary [9] with the G-kurtosis and their powerful associated directional normality test. They also discuss the benefits of both types of kurtosis to detect the non-normality of a sample. Our intuition was that the sweet spot lies in-between, and the second standardized sample moment emerged as the natural choice given the quadratic term in the normal density. However, by construction, $n^{-1}\sum_{i=1}^{n}|Z_i|^2 = 1$. Therefore, we considered instead the limit $n^{-1}\sum_{i=1}^{n}(|Z_i|^{2+\epsilon}-1)/\epsilon$ when $\epsilon \to 0$ and obtained $K_2 := n^{-1}\sum_{i=1}^{n} Z_i^2 \log|Z_i|$, which we will define formally later as the sample 2nd-power kurtosis. We also extended this idea to skewness. While Pearson's skewness is defined as the third standardized sample moment $n^{-1}\sum_{i=1}^{n} Z_i^3 = n^{-1}\sum_{i=1}^{n}|Z_i|^3\,\mathrm{sign}(Z_i)$, we consider instead $B_2 := n^{-1}\sum_{i=1}^{n}|Z_i|^2\,\mathrm{sign}(Z_i)$, which we will define formally later as sample 2nd-power skewness. In our terminology, the JB test uses the 3rd-power skewness and 4th-power kurtosis, while we propose instead to base our test of normality on a combination of the sample 2nd-power skewness and kurtosis. It happens that this approach permits to preserve the structure and benefits of the JB test, namely simple measures that are easily interpretable, with the promise of maximum performance.

However, for our test to be a serious alternative to the JB test, we believe that a formal justification is needed, along with a theoretical framework that will let us obtain the asymptotic distribution of our test statistic. To achieve this, we follow the same strategy adopted by JB, who used Rao's score test (also known as the Lagrange multiplier test, see [18]) on the Pearson family of distributions. It turns out that if one takes instead the family of the asymmetric power distribution (APD), introduced by Komunjer [14], the resulting test statistic is a combination of our new measures of 2nd-power skewness and kurtosis. The APD, described in Section 2, combines the vast range of exponential tail behavior provided by the exponential power distribution (EPD) family with various levels of asymmetry. The large size of this family makes us classify our test as *quasi omnibus*.

In Section 3, we develop Rao's score test on the APD family and easily find, in a first step, the test statistic and its asymptotic distribution, given fixed location and scale parameters. In a second step, we substitute these unknown parameters for their maximum-likelihood estimators under the null hypothesis of normality to test the composite hypotheses. The section is then devoted to finding the asymptotic distribution of the modified statistic. The result is very similar to that of JB: same local optimality; under the null, $B_2$ and $K_2$ are asymptotically independent and normally distributed; and the test statistic, given by the sum of the squares of the standardized 2nd-power skewness and kurtosis, is asymptotically $\chi_2^2$ distributed. (Proving this last result was quite challenging, in particular, the proof of Lemma 2 which is provided in Appendix A.)

As is often the case with asymptotic results, the approximation for small sample sizes is not good enough; for instance, Mantalos [17] shows that the JB test has rather poor small sample properties. In our situation, this is explained in part by the well-known fact that skewed distributions are often associated with heavy tails for small samples. In Section 4, we address this issue by considering $K_2 - B_2^2$ instead of $K_2$, which we will define formally later as the sample 2nd-power net kurtosis. It turns out that the dependency of this measure with $B_2$ is negligible even for small samples. We therefore create a modified statistic, based on standardized 2nd-power skewness and net kurtosis, for which we show numerically that the distribution can be approximated, with very high precision, by a $\chi_2^2$ for all sample sizes as small as 10. We believe that accurate $p$-values and thus reliable conclusions, without the need to rely on simulated quantiles or tables, is a desirable characteristic that can ease the acceptance and implementation of a test. This is rarely found in the (recent) literature for small sample sizes.

In Section 5, we derive a *directional* test of normality based on the sample 2nd-power kurtosis and apply the same strategy as above using Rao's score on the symmetric EPD family. We also provide a transformed version for which we show numerically that the distribution can be approximated with very high precision by a standard normal for sample sizes as small as 10. We obtain a test as powerful as the Coin test [5], a regression and correlation test considered the best. Furthermore, rejection of normality comes with a justification: the tails are too heavy if the statistic is positive, and the tails are too short otherwise.

Finally, an example is given in Section 6, using the computer code for the R software available in the supplementary material at the publisher's website (Appendix D). An extensive empirical power analysis is done in Section 7 (tables with numerical results are postponed to Appendix C). The conclusion follows in Section 8. Note that all proofs are provided in Appendix A.

## 2. Asymmetric power distribution

The APD, proposed by Komunjer [14], can be viewed as a generalization of the symmetrical EPD – also known as the generalized power distribution or the generalized error distribution – to a broader family that includes asymmetric densities. Thus, the APD family combines the large range of exponential tail behaviors provided by the EPD family with various levels of asymmetry. In particular, the normal distribution is included in this family, and therefore, in Section 3, we use the APD as an embedding family of alternatives to develop a new test of normality.

The probability density function $f(u)$ of the standard APD is defined in Section 2 of [14]. In order to obtain the standard normal density as a special case, we modify its scale with the change of variable $u = 2^{-1/\lambda}x$ and obtain

$$f(x \mid \alpha, \lambda) = \frac{\delta_{\alpha,\lambda}^{1/\lambda}}{2^{1/\lambda}\Gamma(1+1/\lambda)} \exp\left[-\frac{1}{2}\frac{\delta_{\alpha,\lambda}}{A_{\alpha,\lambda}(x)}|x|^{\lambda}\right], \quad \text{for all } x \in \mathbb{R}, \tag{1}$$

where $0 < \alpha < 1$, $\lambda > 0$ and $0 < \delta_{\alpha,\lambda} < 1$ with

$$\delta_{\alpha,\lambda} := \frac{2\alpha^{\lambda}(1-\alpha)^{\lambda}}{\alpha^{\lambda}+(1-\alpha)^{\lambda}} \quad \text{and} \quad A_{\alpha,\lambda}(x) := \left[1/2 + \operatorname{sign}(x)(1/2-\alpha)\right]^{\lambda}.$$

We observe that $A_{\alpha,\lambda}(x) = \alpha^{\lambda}$ if $x < 0$ and $A_{\alpha,\lambda}(x) = (1-\alpha)^{\lambda}$ if $x > 0$, which generates the asymmetry of the density with respect to the mode, given by the origin. Therefore, for a given value of $\lambda$, the degree of asymmetry is controlled by the parameter $\alpha$. Indeed, one can verify that $\alpha = \Pr[X < 0]$, which means that the density is skewed to the right if $0 < \alpha < 1/2$, symmetric if $\alpha = 1/2$ and skewed to the left if $1/2 < \alpha < 1$. The tails' behavior of the density is controlled by the parameter $\lambda$; heavier tails are associated with smaller values of $\lambda$ and shorter tails with larger values of $\lambda$. Note that location and scale parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ will be added in Section 3 to test for the composite hypothesis.

The APD family includes some known distributions. Naturally, if we let $\alpha = 1/2$ (it follows that $A_{\alpha,\lambda}(x) = \delta_{\alpha,\lambda} = 2^{-\lambda}$), we obtain the symmetric EPD distribution, which includes the Laplace distribution (also known as the double exponential) if $\lambda = 1$ and the standard normal distribution if $\lambda = 2$. For other values of $\alpha$, if we let $\lambda = 1$, we obtain the asymmetric Laplace distribution, also known as the two-piece double exponential (see [15]), while if we let $\lambda = 2$, we obtain the two-piece normal distribution, also known as the split normal (see [13]).

## 3. The score test on the APD family

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with density

$$g(x \mid \theta_1, \theta_2, \mu, \sigma) := \sigma^{-1} f\left(\sigma^{-1}(x-\mu) \mid \theta_1, \theta_2\right), \quad \text{for all } x \in \mathbb{R},$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$ are the unknown location and scale parameters and $f(x \mid \theta_1, \theta_2)$ is defined in Equation (1). Note that, for convenience, we changed the parameters $\alpha$ and $\lambda$ to $\theta_1$ and $\theta_2$. We will write $X \sim \text{APD}(\theta_1, \theta_2, \mu, \sigma)$ when the density of $X$ is given by $g(x \mid \theta_1, \theta_2, \mu, \sigma)$.

We wish to test the goodness-of-fit hypotheses

$$H_0 : X \sim \mathrm{N}(\mu,\sigma) \quad \text{vs.} \quad H_1 : X \not\sim \mathrm{N}(\mu,\sigma).$$

But given that we assume that $X_1,\ldots,X_n$ are i.i.d. APD$(\theta_1,\theta_2,\mu,\sigma)$, the non-parametric formulation in the above hypotheses can be transformed into the testing of the null hypothesis that the measurements of $X$ come from some $N(\mu,\sigma^2)$ (with $\mu$ and $\sigma$ unspecified) against the family of alternatives APD$(\theta_1,\theta_2,\mu,\sigma)$ (with $\theta_1 \in (0,1), \theta_2 > 0, \mu \in \mathbb{R}, \sigma > 0$). In other words, we wish to test

$$\begin{aligned} &H_0 : X \sim \mathrm{APD}(1/2,2,\mu,\sigma), \\ \text{against} \quad &H_1 : X \sim \mathrm{APD}(\theta_1,\theta_2,\mu,\sigma); \quad (\theta_1,\theta_2) \neq (1/2,2), \end{aligned}$$

which can be achieved using the parametric Rao's score test (also known as the Lagrange multiplier test) of

$$H_0 : (\theta_1,\theta_2) = (1/2,2) \quad \text{vs.} \quad H_1 : (\theta_1,\theta_2) \neq (1/2,2).$$

We consider, in a first step, that $\mu$ and $\sigma$ are some known nuisance parameters. Thus, if we define

$$\boldsymbol{\theta} := (\theta_1,\theta_2)^{\mathrm{T}},$$

the statistic to test the simple null hypothesis $H_0 : (\theta_1,\theta_2) = (1/2,2)$ is based on the vector $n^{-1}\sum_{i=1}^{n}(\partial/\partial\boldsymbol{\theta})\log g(X_i \mid \boldsymbol{\theta}^{\mathrm{T}},\mu,\sigma)|_{\boldsymbol{\theta}=(1/2,2)^{\mathrm{T}}}$. In the second step, we substitute $\mu$ and $\sigma$ for their maximum-likelihood estimators under the null, denoted by $\hat{\mu}_n$ and $\hat{\sigma}_n$, in order to test the composite hypotheses, and we study the asymptotic distribution of the modified statistic. This is the general idea; let us now take a closer look at the situation.

We first define three primary functions. Let $\boldsymbol{d_\theta}(y)$, $d_\mu(y)$ and $d_\sigma(y)$ be defined as

$$\begin{aligned} \boldsymbol{d_\theta}(y) &:= \frac{\partial}{\partial\boldsymbol{\theta}}\log g(x \mid \boldsymbol{\theta}^{\mathrm{T}},\mu,\sigma)\Big|_{\boldsymbol{\theta}=(1/2,2)^{\mathrm{T}},\ x=\mu+\sigma y} = \frac{\partial}{\partial\boldsymbol{\theta}}\log f(y \mid \boldsymbol{\theta}^{\mathrm{T}})\Big|_{\boldsymbol{\theta}=(1/2,2)^{\mathrm{T}}}, \\ d_\mu(y) &:= \sigma\frac{\partial}{\partial\mu}\log g(x \mid 1/2,2,\mu,\sigma)\Big|_{x=\mu+\sigma y} = -\frac{\partial}{\partial y}\log f(y \mid 1/2,2), \end{aligned} \tag{2}$$

$$\begin{aligned} d_\sigma(y) &:= \sigma\frac{\partial}{\partial\sigma}\log g(x \mid 1/2,2,\mu,\sigma)\Big|_{x=\mu+\sigma y} \\ &= -1 - y\frac{\partial}{\partial y}\log f(y \mid 1/2,2) = y d_\mu(y) - 1. \end{aligned} \tag{3}$$

We can verify that

$$\begin{aligned} \boldsymbol{d_\theta}(y) &= \begin{pmatrix} -2y^2\,\mathrm{sign}(y) \\ -2^{-1}[y^2\log|y| - (2-\log 2-\gamma)/2] \end{pmatrix}, \\ d_\mu(y) &= y \quad \text{and} \quad d_\sigma(y) = y^2 - 1, \end{aligned} \tag{4}$$

where

$$\gamma := -\psi(1) = 0.577215665\ldots \tag{5}$$

is the Euler–Mascheroni constant, $\psi(x) := (\mathrm{d}/\mathrm{d}x)\log\Gamma(x) = \Gamma'(x)/\Gamma(x)$ is the digamma function and $\Gamma(x)$ is the gamma function. Note that the result $\psi(3/2) = 2 - 2\log 2 - \gamma$ has been used in the derivations.

We observe that the term $y^2\,\mathrm{sign}(y)$ can also be written as $y|y|$. Furthermore, the function $y^2 \log|y|$ is not defined at $y=0$. Hence, we define $(y^2\log|y|)|_{y=0} := 0$, and as a result, this function is now continuous everywhere.

Using Rao's score test as described above, we consider in the first step that $\mu$ and $\sigma$ are known. Hence, the statistic to test the simple null hypothesis $H_0 : (\theta_1, \theta_2) = (1/2, 2)$ against $H_1 : (\theta_1, \theta_2) \neq (1/2, 2)$ is denoted by $\boldsymbol{r}_n(\mu, \sigma)$ and given by

$$\boldsymbol{r}_n(\mu,\sigma) := \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{d}_{\boldsymbol{\theta}}(Y_i) = \begin{pmatrix} -2\left[n^{-1}\displaystyle\sum_{i=1}^{n} Y_i^2\,\mathrm{sign}(Y_i)\right] \\ -2^{-1}\left[n^{-1}\displaystyle\sum_{i=1}^{n} Y_i^2\log|Y_i| - (2-\log 2-\gamma)/2\right] \end{pmatrix},$$

where

$$Y_i = \sigma^{-1}(X_i - \mu).$$

However, this statistic cannot be used directly to test composite hypotheses when $\mu$ and $\sigma$ are considered unknown. Therefore, the second step consists in substituting $\mu$ and $\sigma$ for their maximum-likelihood estimators $\hat{\mu}_n$ and $\hat{\sigma}_n$, under the null hypothesis given by $X_i \sim N(\mu, \sigma^2)$. Thus, we search for the values $\mu$ and $\sigma$ that jointly satisfy the equations $\sum_{i=1}^{n} d_\mu(Y_i) = 0$ and $\sum_{i=1}^{n} d_\sigma(Y_i) = 0$, and we obtain the well-known estimators

$$\hat{\mu}_n = \bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \hat{\sigma}_n = S_n := \left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right]^{1/2}. \tag{6}$$

Hence, we propose to base the composite test of normality on the statistic $\boldsymbol{r}_n(\hat{\mu}_n, \hat{\sigma}_n)$. The remainder of this section is devoted to establishing the asymptotic distribution of $n^{1/2}\boldsymbol{r}_n(\hat{\mu}_n, \hat{\sigma}_n)$ under the null hypothesis.

The strategy consists first in determining, using the central limit theorem, the asymptotic distribution of the vector $n^{1/2} \cdot n^{-1}\sum_{i=1}^{n}(\boldsymbol{d}_{\boldsymbol{\theta}}(Y_i)^{\mathrm{T}}, d_\mu(Y_i), d_\sigma(Y_i))^{\mathrm{T}}$ under the null hypothesis of normality. The second step consists in writing $n^{1/2}\boldsymbol{r}_n(\hat{\mu}_n, \hat{\sigma}_n)$ as a linear combination of this vector plus a negligible term $o_P(1)\mathbf{1}_2$, in order to obtain the asymptotic distribution of $n^{1/2}\boldsymbol{r}_n(\hat{\mu}_n, \hat{\sigma}_n)$ under the null hypothesis. Thus, for the rest of the section, we assume that $X_i$, and in general $X$, are normally distributed. Or equivalently, for $i = 1, \ldots, n$, we assume that

$$Y_i = \sigma^{-1}(X_i - \mu) \sim N(0,1) \quad \text{and} \quad Y = \sigma^{-1}(X - \mu) \sim N(0,1).$$

**Proposition 3.1:**

$$n^{1/2} \cdot \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} \boldsymbol{d}_{\boldsymbol{\theta}}(Y_i) \\ d_\mu(Y_i) \\ d_\sigma(Y_i)\end{pmatrix} \xrightarrow{\mathcal{D}} N_4\left(\mathbf{0}, \begin{pmatrix} J_{\boldsymbol{\theta}} & J_{\boldsymbol{\theta}\mu} & J_{\boldsymbol{\theta}\sigma} \\ J_{\boldsymbol{\theta}\mu}^{\mathrm{T}} & J_\mu & J_{\mu\sigma} \\ J_{\boldsymbol{\theta}\sigma}^{\mathrm{T}} & J_{\mu\sigma} & J_\sigma \end{pmatrix}\right),$$

*where* $\boldsymbol{d_\theta}(\cdot)$, $d_\mu(\cdot)$, $d_\sigma(\cdot)$ *are defined in Equation* (4), *and*

$$\begin{aligned}
J_{\boldsymbol{\theta}} &:= \mathbb{E}\left[\boldsymbol{d_\theta}(Y)\boldsymbol{d_\theta}(Y)^{\mathrm{T}}\right] = \begin{pmatrix} 12 & 0 \\ 0 & 32^{-1}\left[4(3-\log 2-\gamma)^2+3\pi^2-28\right]\end{pmatrix},\\
J_{\boldsymbol{\theta}\mu} &:= \mathbb{E}\left[\boldsymbol{d_\theta}(Y)d_\mu(Y)\right] = \left(-8(2\pi)^{-1/2};\ 0\right)^{\mathrm{T}},\\
J_{\boldsymbol{\theta}\sigma} &:= \mathbb{E}\left[\boldsymbol{d_\theta}(Y)d_\sigma(Y)\right] = \left(0;\ -(3-\log 2-\gamma)/2\right)^{\mathrm{T}},\\
J_\mu &:= \mathbb{E}\left[d_\mu^2(Y)\right] = 1, \quad J_\sigma := \mathbb{E}\left[d_\sigma^2(Y)\right] = 2, \quad J_{\mu\sigma} := \mathbb{E}\left[d_\mu(Y)d_\sigma(Y)\right] = 0.
\end{aligned}$$

***Proof:*** The proposition is a direct application of the central limit theorem. See Appendix A for the details of the calculations. ∎

In the next four propositions, we study $n^{1/2}\boldsymbol{r}_n(\hat{\mu}_n, \hat{\sigma}_n)$ with the aim of writing this statistic as a linear combination of the vector given in Proposition 3.1, plus an asymptotically negligible term.

**Proposition 3.2:**

$$\begin{aligned}
n^{1/2}\boldsymbol{r}_n(\hat{\mu}_n, \hat{\sigma}_n) = n^{1/2}\boldsymbol{r}_n(\mu,\sigma) &+ n^{1/2}(\hat{\mu}_n-\mu)\frac{\partial}{\partial\mu}\boldsymbol{r}_n(\mu,\sigma)\\
&+ n^{1/2}(\hat{\sigma}_n-\sigma)\frac{\partial}{\partial\sigma}\boldsymbol{r}_n(\mu,\sigma) + n^{1/2}R,
\end{aligned}$$

*where* $n^{1/2}R = o_P(1)\mathbf{1}_2$ *is a negligible term and* $\mathbf{1}_2 := (1,1)^{\mathrm{T}}$.

***Proof:*** We use the Taylor expansion of $\boldsymbol{r}_n(\hat{\mu}_n, \hat{\sigma}_n)$ around $(\mu,\sigma)$, where $R$ is the remainder term. Furthermore, we know from Proposition 3.1 that $n^{1/2}\boldsymbol{r}_n(\mu,\sigma) = n^{1/2}\cdot n^{-1}\sum_{i=1}^n \boldsymbol{d_\theta}(Y_i)$ is $O_P(1)\mathbf{1}_2$, and it is shown in Appendix A that $n^{1/2}R = o_P(1)\mathbf{1}_2$, thus a negligible term. ∎

We now study the terms $n^{1/2}(\hat{\mu}_n-\mu)$, $n^{1/2}(\hat{\sigma}_n-\sigma)$ in Proposition 3.3 and the derivatives $(\partial/\partial\mu)\boldsymbol{r}_n(\mu,\sigma)$, $(\partial/\partial\sigma)\boldsymbol{r}_n(\mu,\sigma)$ in Proposition 3.4.

**Proposition 3.3:**

$$n^{1/2}(\hat{\mu}_n-\mu) = n^{1/2}\sigma J_\mu^{-1}\cdot\frac{1}{n}\sum_{i=1}^n d_\mu(Y_i) + o_P(1)$$

*and*

$$n^{1/2}(\hat{\sigma}_n-\sigma) = n^{1/2}\sigma J_\sigma^{-1}\cdot\frac{1}{n}\sum_{i=1}^n d_\sigma(Y_i) + o_P(1),$$

*where* $J_\mu$ *and* $J_\sigma$ *are defined in Proposition* 3.1.

***Proof:*** See Appendix A. ∎

**Proposition 3.4:**

$$\frac{\partial}{\partial \mu} \boldsymbol{r}_n(\mu, \sigma) = -\sigma^{-1} J_{\boldsymbol{\theta}\mu} + o_P(1)\mathbf{1}_2 \quad \textit{and} \quad \frac{\partial}{\partial \sigma} \boldsymbol{r}_n(\mu, \sigma) = -\sigma^{-1} J_{\boldsymbol{\theta}\sigma} + o_P(1)\mathbf{1}_2.$$

***Proof:*** See Appendix A. ■

The next proposition is directly obtained by combining Proposition 3.2 with Propositions 3.3 and 3.4.

**Proposition 3.5:**

$$n^{1/2} \boldsymbol{r}_n(\hat{\mu}_n, \hat{\sigma}_n) = n^{1/2} \cdot \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{d}_{\boldsymbol{\theta}}(Y_i) - n^{1/2} J_\mu^{-1} J_{\boldsymbol{\theta}\mu} \cdot \frac{1}{n} \sum_{i=1}^{n} d_\mu(Y_i) - n^{1/2} J_\sigma^{-1} J_{\boldsymbol{\theta}\sigma} \cdot \frac{1}{n} \sum_{i=1}^{n} d_\sigma(Y_i) + o_P(1)\mathbf{1}_2.$$

*Equivalently, in matrix form, we have*

$$n^{1/2} \boldsymbol{r}_n(\hat{\mu}_n, \hat{\sigma}_n) = n^{1/2} \left( I_2; -J_\mu^{-1} J_{\boldsymbol{\theta}\mu};\ -J_\sigma^{-1} J_{\boldsymbol{\theta}\sigma} \right) \cdot \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} \boldsymbol{d}_{\boldsymbol{\theta}}(Y_i) \\ d_\mu(Y_i) \\ d_\sigma(Y_i) \end{pmatrix} + o_P(1)\mathbf{1}_2,$$

*where $I_2$ is the identity matrix of size 2.*

We observe, using the central limit theorem, that each term of the linear combination is $O_P(1)\mathbf{1}_2$, except obviously the term $o_P(1)\mathbf{1}_2$. Finally, combining Propositions 3.1 and 3.5, we obtain the asymptotic distribution of $\boldsymbol{r}_n(\hat{\mu}_n, \hat{\sigma}_n)$ under the null hypothesis of normality.

**Proposition 3.6:**

$$n^{1/2} \boldsymbol{r}_n(\hat{\mu}_n, \hat{\sigma}_n) \xrightarrow{\mathcal{D}} N_2 \left( \mathbf{0}, J_{\boldsymbol{\theta}} - J_\mu^{-1} J_{\boldsymbol{\theta}\mu} J_{\boldsymbol{\theta}\mu}^{\mathrm{T}} - J_\sigma^{-1} J_{\boldsymbol{\theta}\sigma} J_{\boldsymbol{\theta}\sigma}^{\mathrm{T}} \right),$$

*with*

$$J_{\boldsymbol{\theta}} - J_\mu^{-1} J_{\boldsymbol{\theta}\mu} J_{\boldsymbol{\theta}\mu}^{\mathrm{T}} - J_\sigma^{-1} J_{\boldsymbol{\theta}\sigma} J_{\boldsymbol{\theta}\sigma}^{\mathrm{T}} = \begin{pmatrix} 4(3 - 8/\pi) & 0 \\ 0 & (3\pi^2 - 28)/32 \end{pmatrix}.$$

***Proof:*** See Appendix A. ■

Before we introduce the statistic for the asymptotic test of normality, the following definition is given.

**Definition 3.7:** For a sample $X_1, \ldots, X_n$, '2nd-power skewness' and '2nd-power kurtosis' are, respectively, denoted by $B_2$ and $K_2$, and defined as

$$B_2 := \frac{1}{n}\sum_{i=1}^{n} Z_i^2 \operatorname{sign}(Z_i) \quad \text{and} \quad K_2 := \frac{1}{n}\sum_{i=1}^{n} Z_i^2 \log|Z_i|,$$

where $Z_i = S_n^{-1}(X_i - \bar{X}_n)$ and $\bar{X}_n$, $S_n$ are defined in Equation (6).

Analogously, 2nd-power skewness and kurtosis for a random variable $X$ are defined, respectively, as $E(Z^2 \operatorname{sign}(Z))$ and $E(Z^2 \log(Z))$, where $Z = \sigma^{-1}(X - \mu)$. Note that $B_2$ can also be written as $B_2 := n^{-1}\sum_{i=1}^{n} Z_i|Z_i|$. As mentioned in Section 1, $B_2$ is an alternative to Pearson's sample skewness given by $n^{-1}\sum_{i=1}^{n} Z_i^3$, which can also be written as $n^{-1}\sum_{i=1}^{n} |Z_i|^3 \operatorname{sign}(Z_i)$. In our proposed terminology, this would be 3rd-power skewness, while $B_2$ is 2nd-power skewness. Similarly, $K_2$ is an alternative to Pearson's sample kurtosis given by $n^{-1}\sum_{i=1}^{n} |Z_i|^4$, which can be called 4th-power kurtosis in our terminology, or to Geary's measure of kurtosis given by $n^{-1}\sum_{i=1}^{n} |Z_i|$, which can be called 1st-power kurtosis. However, for 2nd-power kurtosis, we must take the limiting case because, by construction, $n^{-1}\sum_{i=1}^{n} |Z_i|^2 = 1$ for any sample. Therefore, we consider $\lim_{\epsilon\to 0} n^{-1}\sum_{i=1}^{n}(|Z_i|^{2+\epsilon} - 1)/\epsilon$, which happens to be equal to $K_2$.

A significative positive (negative) value of $B_2$ suggests that the distribution is right-skewed (left-skewed), while a small value of $|B_2|$ suggests that the distribution is symmetric. Furthermore, we can show that $K_2$ is a positive random variable (for any sample size), with large (small) values of $K_2$ corresponding to long-tailed (short-tailed) distribution.

Using $B_2$ and $K_2$, Proposition 3.6 can be rewritten explicitly as follows:

$$n^{1/2}\begin{pmatrix} -2B_2 \\ -2^{-1}[K_2 - (2 - \log 2 - \gamma)/2] \end{pmatrix} \xrightarrow{\mathcal{D}} N_2\left(\mathbf{0}, \begin{pmatrix} 4(3 - 8/\pi) & 0 \\ 0 & (3\pi^2 - 28)/32 \end{pmatrix}\right). \tag{7}$$

We can now present our main theoretical result in Theorem 3.8, which follows directly from Equation (7).

**Theorem 3.8:** *The statistic for the asymptotic test of normality, for the composite hypothesis and based on Rao's score test with the APD family of alternatives, is denoted by* $X^a_{\text{APD}}$ *and given by*

$$X^a_{\text{APD}} := \frac{nB_2^2}{3 - 8/\pi} + \frac{n\left(K_2 - (2 - \log 2 - \gamma)/2\right)^2}{(3\pi^2 - 28)/8},$$

*where* $B_2, K_2$ *are given in Definition* 3.7 *and* $\gamma$ *is defined in Equation* (5). *Furthermore, under the null hypothesis,*

$$X^a_{\text{APD}} \xrightarrow{\mathcal{D}} \chi^2_2, \quad \text{as } n \to \infty.$$

*The null hypothesis is rejected if* $X^a_{\text{APD}}$ *is larger than the chi-squared quantile* $\chi^2_{2,\alpha}$, *at a significance level of* $\alpha$. *P-value can be computed as* $\Pr(X > X^a_{\text{APD}})$, *where* $X$ *is a* $\chi^2_2$*-distributed random variable.*

The statistic $X^a_{\mathrm{APD}}$ is remarkably simple, as a result of the asymptotic null covariance (and independence) between $B_2$ and $K_2$, as shown by Proposition 3.6. The form of $X^a_{\mathrm{APD}}$ is very similar to the JB statistic, which involves Pearson's skewness and kurtosis, or using our terminology, 3rd-power skewness and 4th-power kurtosis. Note that the superscript $a$ in the notation $X^a_{\mathrm{APD}}$ stands for 'asymptotic', to mark that the chi-squared distribution is valid for $n \to \infty$ or in practice for large $n$. Indeed, as is often the case with Rao's score tests, the chi-squared approximation for small sample sizes is not good enough. In the next section, we go one step further by proposing a modified version of the statistic $X^a_{\mathrm{APD}}$, denoted by $X_{\mathrm{APD}}$, for which we show numerically that the distribution can be approximated very precisely, under the null, by a chi-squared for sample sizes as small as 10.

## 4. The $X_{\mathrm{APD}}$ test for finite sample sizes

The first issue to address in the modification of $X^a_{\mathrm{APD}}$ is the dependency between $B_2$ and $K_2$ in small samples, because the $\chi^2_2$ distribution results from a sum of squares of two independent standard normals. Note that we assume in this section that $n \geq 10$. Indeed, small and large values of $B_2$ are associated with large values of $K_2$. The same issue can be observed with Pearson's skewness and kurtosis. A skewed distribution, to the left or to the right, often exhibits heavy tails when the sample size is small. To resolve this problem, we consider, instead of $K_2$, the following quantity.

**Definition 4.1:** The sample '2nd-power net kurtosis' is defined as $K_2 - B_2^2$, where $B_2$ and $K_2$ are given in Definition 3.7.

We can show that $K_2 - B_2^2 \geq 0$ for all samples of any size. The key part of the proof is the analysis of the case of an odd sample of size $2n+1$ that contains only two distinct values with $n$ replications of the largest. In this case, one can verify that $B_2 = (2n+1)^{-1}$ and $K_2 = 2^{-1} B_2 \log((n+1)/n)$. Finally, it suffices to observe that $K_2 - B_2^2 = B_2^2(K_2/B_2^2 - 1)$ and that $K_2/B_2^2 = 2^{-1}(2n+1)\log((n+1)/n)$ is decreasing and converges to 1 as $n \to \infty$.

The strategy consists in basing the modified statistic $X_{\mathrm{APD}}$ on transformed measures of $B_2$ and $K_2 - B_2^2$ that are approximately distributed as $N(0, 1)$. We find numerically that the dependency between $K_2 - B_2^2$ and $B_2$ is negligible. Furthermore, we find numerically that $(K_2 - B_2^2)^{1/3}$ and $B_2$ are closely distributed as a normal for all samples of $n \geq 10$, under the null hypothesis. Note that the power of 1/3 comes from a Wilson–Hilferty cubed root transformation that leads to normality because $K_2 - B_2^2$ can be approached by a gamma.

The next steps consist in the standardization of $B_2$ and $(K_2 - B_2^2)^{1/3}$, always under the null hypothesis, for all $n \geq 10$. We have $E(B_2) = 0$ for all sample sizes because $Z_i^2$ is independent of $\mathrm{sign}(Z_i)$ and $E(\mathrm{sign}(Z_i)) = 0$. Furthermore, we know from Theorem 3.8 that the asymptotic variance of $n^{1/2}B_2$ is given by $3 - 8/\pi$. The variance for finite samples is then estimated using a linear regression through the origin, where the variable $\mathrm{Var}\,(n^{1/2}B_2)/(3 - 8/\pi) - 1$ (simulated for various $n \geq 10$) is explained by $1/n^{\alpha}$, with $\alpha$ chosen to maximize the $R^2$ of the regression. We find that $\alpha = 1$, with a regression coefficient equal to $-1.9$. It leads us to the next definition.

**Definition 4.2:** The 'transformed 2nd-power skewness', denoted by $Z(B_2)$, is defined as

$$Z(B_2) := \frac{n^{1/2}B_2}{[(3-8/\pi)(1-1.9/n)]^{1/2}},$$

where $B_2$ is given in Definition 3.7.

Now, using Theorem 3.8, the delta method and $n^{1/2}B_2^2 = o_P(1)$, we find that the asymptotic expectation of $(K_2 - B_2^2)^{1/3}$ is given by $E^a := ((2-\log 2-\gamma)/2)^{1/3}$ and that the asymptotic variance of $n^{1/2}(K_2 - B_2^2)^{1/3}$ is given by $V^a := 9^{-1}((2-\log 2-\gamma)/2)^{-4/3}(3\pi^2-28)/8$. The expectation and variance for finite samples are then estimated using linear regressions through the origin, where the variables $E[(K_2 - B_2^2)^{1/3}]/E^a - 1$ and Var $[n^{1/2}(K_2 - B_2^2)^{1/3}]/V^a - 1$ (simulated for various $n \geq 10$) are explained by $1/n^{\alpha}$, with $\alpha$ chosen to maximize the $R^2$ of each regression. It leads us to the next definition.

**Definition 4.3:** The 'transformed 2nd-power net kurtosis', denoted by $Z(K_2 - B_2^2)$, is defined as

$$Z(K_2 - B_2^2) := \frac{n^{1/2}\left[(K_2 - B_2^2)^{1/3} - ((2-\log 2-\gamma)/2)^{1/3}(1-1.026/n)\right]}{\left[72^{-1}((2-\log 2-\gamma)/2)^{-4/3}(3\pi^2-28)(1-2.25/n^{0.8})\right]^{1/2}},$$

where $B_2$ and $K_2$ are given in Definition 3.7.

Considering that we have found numerically that $Z(B_2)$ and $Z(K_2 - B_2^2)$ are approximately distributed, under the null, as standard normal for all $n \geq 10$, with a negligible dependence between them, we present our statistic for finite sample sizes in the next proposition.

**Proposition 4.4:** *The proposed statistic to test the composite hypothesis of normality, for finite sample sizes* $n \geq 10$, *is denoted by* $X_{\text{APD}}$ *and given by*

$$X_{\text{APD}} := Z^2(B_2) + Z^2(K_2 - B_2^2),$$

*where the transformed 2nd-power skewness* $Z(B_2)$ *and the transformed 2nd-power net kurtosis* $Z(K_2 - B_2^2)$ *are, respectively, given in Definitions* 4.2 *and* 4.3. *Or written explicitly,*

$$X_{\text{APD}} := \frac{nB_2^2}{(3-8/\pi)(1-1.9/n)} + \frac{n\left[(K_2 - B_2^2)^{1/3} - ((2-\log 2-\gamma)/2)^{1/3}(1-1.026/n)\right]^2}{72^{-1}((2-\log 2-\gamma)/2)^{-4/3}(3\pi^2-28)(1-2.25/n^{0.8})},$$

*where* $B_2$, $K_2$ *are given in Definition* 3.7 *and* $\gamma$ *is defined in Equation* (5). *Furthermore, under the null hypothesis ('*$\overset{\text{app}}{\sim}$*' denotes 'approximately distributed as' with high numerical precision),*

$$X_{\text{APD}} \overset{\text{app}}{\sim} \chi_2^2, \quad \textit{for all } n \geq 10.$$

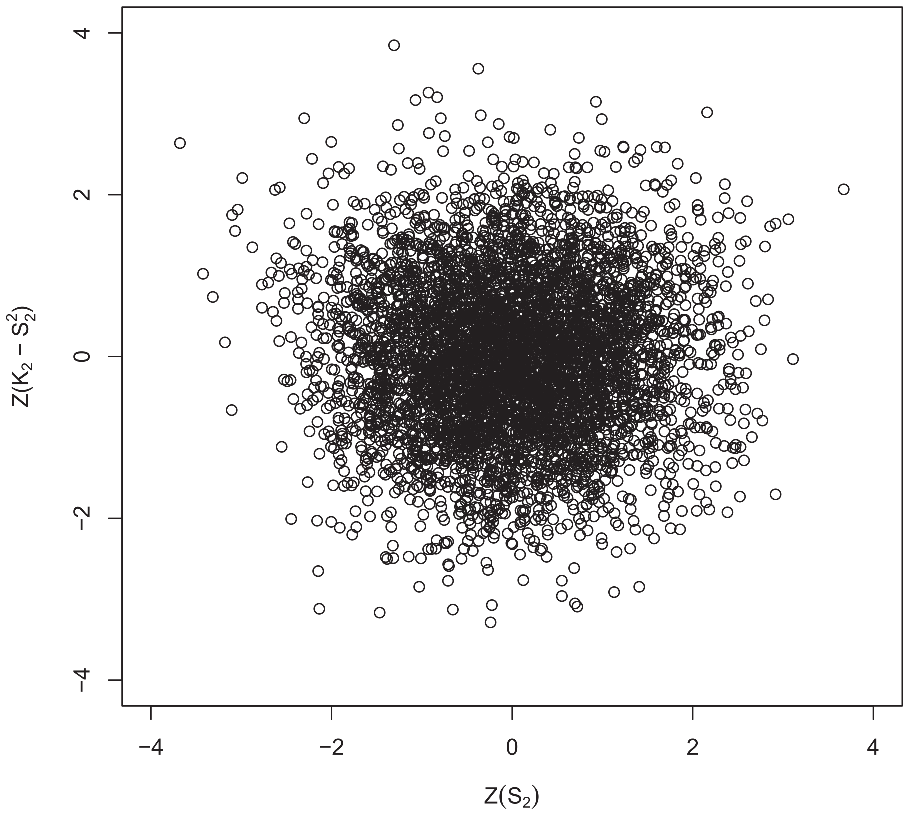


**Figure 1.** Transformed 2nd-power skewness $Z(B_2)$ and transformed 2nd-power net kurtosis $Z(K_2 - B_2^2)$, for 5,000 normal samples of size 20.

*The null hypothesis is rejected if* $X_{\text{APD}}$ *is larger than the chi-squared quantile* $\chi^2_{2,\alpha}$, *at a significance level of* $\alpha$. *P-value can be computed as* $\Pr(X > X_{\text{APD}})$, *where X is a* $\chi^2_2$*-distributed random variable.*

Note that $X_{\text{APD}}/X^a_{\text{APD}} \to 1$ as $n \to \infty$, as a result of the delta method and therefore $X_{\text{APD}} \xrightarrow{\mathcal{D}} \chi^2_2$ as $n \to \infty$. We observe that the modified statistic $X_{\text{APD}}$ remains relatively simple, although it is now adjusted for the sample size.

The question is now how good is the approximation of the distribution of $X_{\text{APD}}$ by a chi-squared distribution with two degrees of freedom. A preliminary answer is given visually in Figure 1, where $Z(B_2)$ and $Z(K_2 - B_2^2)$ are plotted for 5000 normal samples of size 20. It looks exactly as a plot of two independent $N(0, 1)$ variables. We also assessed the quality of the fit by performing a study of the empirical level (empirical power under the null hypothesis) of the statistic $X_{\text{APD}}$ based on 1,000,000 simulations and on $\chi^2_2$ quantiles, for different sample sizes. The results, given in Table 1, show that the significance level of the test is very accurate. To better appreciate the level of precision, we present in Appendix B (available in the supplementary material at the publisher's website) the same table for different competitor tests for which a formula for the computation of $p$-values is available. None of them reaches the accuracy provided by our test $X_{\text{APD}}$.

**Table 1.** Empirical power (in %) of the tests $X_{\text{APD}}$ and $Z_{\text{EPD}}$, under the null hypothesis, based on 1,000,000 simulations and on $\chi^2_2$ and $N(0,1)$ quantiles, for different sample sizes.

| | Empirical power, under the null, of $X_{\text{APD}}$ | | | | | | | | | Empirical power, under the null, of $Z_{\text{EPD}}$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha\backslash n$ | 10 | 12 | 14 | 16 | 18 | 20 | 50 | 100 | 200 | 10 | 12 | 14 | 16 | 18 | 20 | 50 | 100 | 200 |
| 1 | 0.9 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.1 | 1.1 | 1.1 | 1.0 | 1.1 | 1.1 | 1.1 | 1.1 | 1.1 | 1.1 | 1.1 | 1.0 |
| 2 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.1 | 2.1 | 2.1 | 2.0 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.0 |
| 3 | 3.1 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.1 | 3.1 | 3.1 | 3.1 | 3.1 | 3.1 | 3.1 | 3.1 | 3.1 | 3.1 | 3.0 |
| 4 | 4.2 | 4.1 | 4.0 | 4.0 | 4.0 | 4.0 | 4.0 | 4.1 | 4.1 | 4.1 | 4.1 | 4.1 | 4.1 | 4.1 | 4.1 | 4.1 | 4.0 | 4.0 |
| 5 | 5.3 | 5.1 | 5.0 | 5.0 | 5.0 | 5.0 | 5.0 | 5.1 | 5.1 | 5.1 | 5.1 | 5.1 | 5.1 | 5.1 | 5.1 | 5.1 | 5.0 | 5.0 |
| 6 | 6.3 | 6.1 | 6.0 | 6.0 | 6.0 | 6.0 | 6.0 | 6.1 | 6.1 | 6.2 | 6.1 | 6.1 | 6.1 | 6.1 | 6.1 | 6.0 | 6.0 | 6.0 |
| 7 | 7.4 | 7.1 | 7.1 | 7.0 | 6.9 | 6.9 | 7.0 | 7.0 | 7.0 | 7.2 | 7.1 | 7.1 | 7.1 | 7.1 | 7.0 | 7.0 | 7.0 | 7.0 |
| 8 | 8.5 | 8.2 | 8.1 | 8.0 | 7.9 | 7.9 | 7.9 | 8.0 | 8.0 | 8.2 | 8.0 | 8.0 | 8.1 | 8.0 | 8.0 | 8.0 | 8.0 | 8.0 |
| 9 | 9.5 | 9.2 | 9.1 | 9.0 | 8.9 | 8.9 | 8.9 | 9.0 | 9.1 | 9.2 | 9.0 | 9.0 | 9.0 | 9.0 | 9.0 | 9.0 | 9.0 | 9.0 |
| 10 | 10.6 | 10.3 | 10.1 | 10.0 | 9.9 | 9.9 | 9.9 | 10.0 | 10.0 | 10.2 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 |
| 12 | 12.7 | 12.3 | 12.1 | 12.0 | 11.8 | 11.9 | 11.8 | 12.0 | 12.0 | 12.1 | 11.9 | 12.0 | 12.0 | 11.9 | 11.9 | 11.9 | 12.0 | 12.0 |
| 14 | 14.8 | 14.3 | 14.1 | 14.0 | 13.8 | 13.8 | 13.8 | 14.0 | 14.0 | 14.1 | 13.9 | 13.9 | 13.9 | 13.8 | 13.9 | 13.9 | 13.9 | 14.0 |
| 16 | 16.9 | 16.4 | 16.1 | 16.0 | 15.8 | 15.8 | 15.8 | 15.9 | 16.0 | 16.0 | 15.9 | 15.9 | 15.9 | 15.8 | 15.8 | 15.9 | 15.9 | 16.0 |
| 18 | 18.9 | 18.4 | 18.1 | 18.0 | 17.8 | 17.8 | 17.7 | 18.0 | 18.0 | 18.0 | 17.8 | 17.8 | 17.9 | 17.8 | 17.8 | 17.8 | 17.9 | 18.0 |
| 20 | 21.0 | 20.4 | 20.2 | 19.9 | 19.8 | 19.7 | 19.7 | 19.9 | 20.0 | 19.9 | 19.8 | 19.8 | 19.8 | 19.7 | 19.8 | 19.8 | 19.9 | 19.9 |

## 5. The directional test for finite sample sizes

When it is known that the distribution of the random variable is symmetric, we can take advantage of this information by using a directional test and thus increasing the power. In this section, we consider a directional test based on sample 2nd-power kurtosis. As mentioned in Section 2, the symmetrical EPD is a particular case of the APD when the parameter of asymmetry is set to $\theta_1 = 1/2$; therefore, we will write

$$\text{EPD}(\theta_2, \mu, \sigma) := \text{APD}(1/2, \theta_2, \mu, \sigma)$$

and use the EPD as a family of alternatives.

We wish to test the null hypothesis that the measurements of $X$ come from some $N(\mu, \sigma^2)$ (with $\mu$ and $\sigma$ unspecified) against the family of $\text{EPD}(\theta_2, \mu, \sigma)$. In other words, considering that $\theta_2 > 0, \mu \in \mathbb{R}, \sigma > 0$, we wish to test

$$H_0 : X \sim \text{EPD}(2, \mu, \sigma) \quad \text{against} \quad H_1 : X \sim \text{EPD}(\theta_2, \mu, \sigma); \quad \theta_2 \neq 2.$$

To achieve this, we perform Rao's score test following the same strategy adopted in Section 3, and it is easy to verify that the resulting test statistic is based on $K_2$. Furthermore, we obtain the following result directly from Theorem 3.8:

**Corollary 5.1:** *Under the null hypothesis, we have*

$$Z^a_{\text{EPD}} := \frac{n^{1/2}\left(K_2 - (2 - \log 2 - \gamma)/2\right)}{[(3\pi^2 - 28)/8]^{1/2}} \xrightarrow{\mathcal{D}} N(0,1), \quad \text{as } n \to \infty.$$

As was the case for the $X_{\text{APD}}$ test statistic, the normal approximation for small sample sizes is not good enough. Therefore, we go one step further by proposing a simple transformation of $K_2$, for which we show numerically that the distribution can be closely approximated, under the null, by a normal for sample sizes as small as 10.

We first find, for each $n \geq 10$, the value $\alpha_n$ that maximizes the normality of the Box-Cox transformation $T_n := ((2K_2)^{\alpha_n} - 1)/\alpha_n$, using 10,000 values of $K_2$ simulated under the null. A simple linear regression is then used to explain $\alpha_n$ by $1/n^b$, where $b$ is chosen to maximize the $R^2$, and we obtain the following equation:

$$\alpha_n = -0.06 + 2.1/n^{0.67}. \tag{8}$$

Now, using the delta method with the results of Theorem 3.8, we find that the asymptotic expectation of $T_n$ is given by $E^a := -((2-\log 2-\gamma)^{-0.06}-1)/0.06$ and that the asymptotic variance of $n^{1/2}T_n$ is given by $V^a := 2^2(2-\log 2-\gamma)^{2(-0.06-1)}(3\pi^2-28)/8$. The expectation and variance for finite samples are then estimated using linear regressions through the origin where the variables $E(T_n) - E^a$ and Var $(n^{1/2}T_n) - V^a$ (simulated for various $n \geq 10$) are explained by $1/n^b$, with $b$ chosen to maximize the $R^2$ of each regression. It leads us to the next definition.

**Definition 5.2:** The 'transformed 2nd-power kurtosis', denoted by $Z(K_2)$, is defined as

$$Z(K_2) := \frac{n^{1/2}\left[((2K_2)^{\alpha_n}-1)/\alpha_n + ((2-\log 2-\gamma)^{-0.06}-1)/0.06 + 1.32/n^{0.95}\right]}{\left[(2-\log 2-\gamma)^{-2.12}(3\pi^2-28)/2 - 3.78/n^{0.733}\right]^{1/2}},$$

where $K_2$ is given in Definition 3.7, $\alpha_n$ is given in Equation (8) and $\gamma$ is defined in Equation (5).

The directional test follows directly in the next proposition.

**Proposition 5.3:** *The proposed directional statistic to test the composite hypothesis of normality, for finite sample sizes $n \geq 10$, is denoted by $Z_{\mathrm{EPD}}$ and given by*

$$Z_{\mathrm{EPD}} := Z(K_2),$$

*where the transformed 2nd-power kurtosis $Z(K_2)$ is given in Definition 5.2. Furthermore, under the null hypothesis,*

$$Z_{\mathrm{EPD}} \overset{\mathrm{app}}{\sim} N(0,1), \quad \text{for all } n \geq 10.$$

*The null hypothesis is rejected if $|Z_{\mathrm{EPD}}|$ is larger than the normal quantile $z_{\alpha/2}$, at a significance level of $\alpha$. P-value can be computed as $2\Pr(Z > |Z_{\mathrm{EPD}}|)$, where $Z$ is a $N(0,1)$-distributed random variable.*

Note that $Z_{\mathrm{EPD}}/Z^a_{\mathrm{APD}} \to 1$ as $n \to \infty$, as a result of the delta method and therefore $Z_{\mathrm{EPD}} \xrightarrow{\mathcal{D}} N(0,1)$ as $n \to \infty$. We also assessed the quality of the fit by performing a study of the empirical power of the statistic $Z_{\mathrm{EPD}}$, under the null hypothesis, based on 1,000,000 simulations and on the normal quantiles, for different sample sizes. The results, given in Table 1, show that the significance level of the test is again very accurate.

## 6. Example

In this section, we present an example using both tests $X_{\mathrm{APD}}$ and $Z_{\mathrm{EPD}}$ and interpret the results. To ease the calculations, we provide the computer code for the R software in

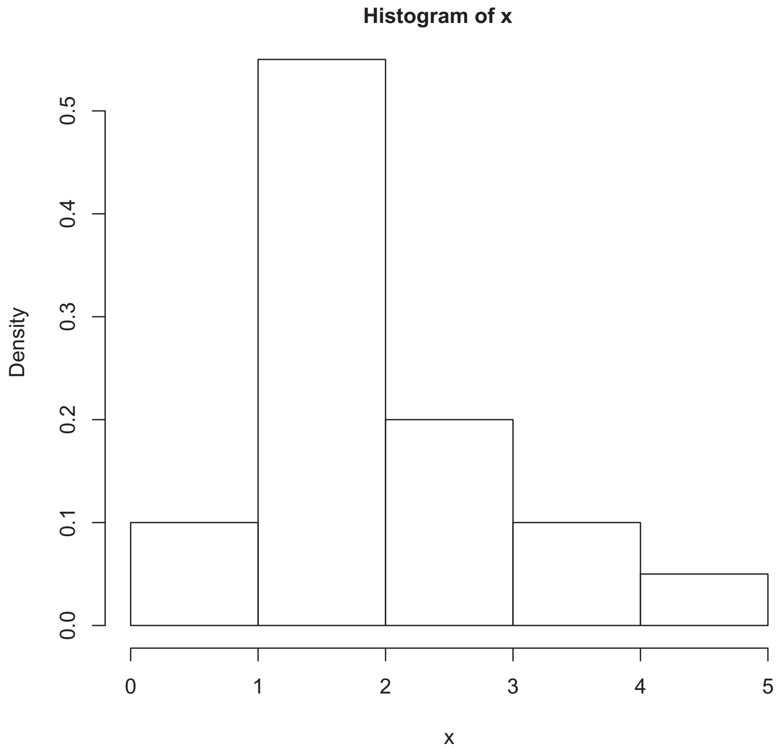


**Figure 2.** Histogram for the sample of the example, $n = 20$.

Appendix D (available in the supplementary material at the publisher's website). The tests can also be computed using the `PoweR` package (version 1.06) [16].

Consider the following sample $X_1, \ldots, X_{20}$, coded in R as

```
x <- c (0.2 , 0.5 , 1.1 , 1.4 , 1.6 , 1.6 , 1.7 , 1.7 ,
          1.7 , 1.8 , 1.9 , 2.0 , 2.0 , 2.1 , 2.1 , 2.1 ,
          2.7 , 3.2 , 4.0 , 4.6)
```

The histogram, given in Figure 2, shows a distribution that is skewed to the right. However, it is more difficult to visually evaluate the tails' thickness. We first find that 2nd-power skewness is $B_2 = 0.27073$ and 2nd-power kurtosis is $K_2 = 0.55356$. Second, we compute transformed 2nd-power skewness, transformed 2nd-power net kurtosis and transformed 2nd-power kurtosis. We find that $Z(B_2) = 1.88985$, $Z(K_2 - B_2^2) = 1.80266$ and $Z(K_2) = 2.07717$. Note that these transformed values can be interpreted as $Z$-scores, which means for instance that values smaller than $-1.96$ or larger than 1.96 can be considered as the most extreme 5%.

The statistics of the $X_{\mathrm{APD}}$ test and of the $Z_{\mathrm{EPD}}$ directional test are then given, respectively, by

$$X_{\mathrm{APD}} = Z^2(B_2) + Z^2(K_2 - B_2^2) = 6.82111 \quad \text{and} \quad Z_{\mathrm{EPD}} = Z(K_2) = 2.07717.$$

$P$-values for the $X_{\text{APD}}$ and $Z_{\text{EPD}}$ tests are, respectively, 0.0330 and 0.0378. The null hypothesis of normality is thus rejected if the significance level is 5%. Note that the Shapiro–Wilk test, using the instruction `shapiro.test(x)` in R, gives a $p$-value of .0405, which is consistent with our tests. However, the JB test gives a $p$-value of .1885, which does not allow us to reject the normality.

An interesting feature of our tests is the possibility of interpreting the results, beyond the rejection of the null. If it is known that the true distribution is symmetric, then the directional test is appropriate. Under the null hypothesis of normality, there is a 5% chance that the test statistic (in absolute value) $|Z_{\text{EPD}}|$ will be larger than the quantile $z_{0.025} = 1.96$, or equivalently that 2nd-power kurtosis $K_2$ will be smaller than 0.19143 or larger than 0.53852, given that the sample size is 20. For our sample, we observed $K_2 = 0.55356$ and $Z_{\text{EPD}} = 2.07717$, which means that the normality is rejected at a significance level of 5% because the tails of the observed distribution are heavier than what we could expect under the null for a sample size of 20.

If the symmetry is not assumed, as is generally the case, then the $X_{\text{APD}}$ test is more appropriate. Under the null hypothesis of normality, there is a 5% chance that the test statistic $X_{\text{APD}} = Z^2(B_2) + Z^2(K_2 - B_2^2)$ will be larger than the quantile $\chi^2_{2;0.5} = 5.99146$, or equivalently that the point $(Z(B_2), Z(K_2 - B_2^2))$ will lie outside a circle of radius $(\chi^2_{2;0.5})^{1/2} = 2.44775$ and centered at the origin (see Figure 1). For example, any point such that $|Z(B_2)| > 1.73082$ and $|Z(K_2 - B_2^2)| > 1.73082$ or $|Z(K_2 - B_2^2)| > 2.44775$ or $|Z(B_2)| > 2.44775$ is among the most extreme 5%. For our sample, we observed $(Z(B_2), Z(K_2 - B_2^2)) = (1.88985, 1.80266)$ and $X_{\text{APD}} = 6.82111$, which means that the normality is rejected at a significance level of 5%. The positive and relatively large values of both $Z$-scores indicate that this is partly because the observed right-skewness is important and partly because the tails of the distribution are long, given this level of asymmetry and the sample size of 20, considering what we could expect under the null. Note that in this case, each indication is not strong enough to singly lead to the rejection; instead, it is the combination of both of them that allows us to conclude with confidence that the sample is not normally distributed.

## 7. Empirical power analysis

In this section, we compare the empirical power of our tests, the quasi omnibus $X_{\text{APD}}$ and the directional $Z_{\text{EPD}}$, with the most powerful normality tests available in the literature. A preliminary analysis of the 33 tests analyzed (and described in detail) in [19] has been done to make an informed choice, and we have selected the three best omnibus tests in each of the three following categories: regression and correlation tests, tests based on the empirical distribution function and tests based on measures of skewness and kurtosis. We also selected the two best directional tests against symmetric alternatives. The present empirical power analysis is thus performed for the 13 normality tests given in Table 2.

For our study, we chose a total of 85 alternatives: 33 symmetric long-tailed, 26 symmetric short-tailed and 26 asymmetric. Naturally, we considered the APD and EPD alternatives, as well as usual distributions such as the Student's $t$, logistic, beta, $\chi^2$, gamma, Gumbel, log-normal and Weibull.

**Table 2.** Selected tests for the empirical power analysis.

| Abbreviations | Regression and correlation tests |
|---|---|
| $W$ | Shapiro–Wilk test |
| $CS$ | Chen–Shapiro test |
| $BCMR$ | del Barrio–Cuesta–Albertos–Matrán–Rodríguez-Rodríguez test |
| $\beta_3^2$ | (Directional) Coin test |
| **Abbreviations** | **Tests based on the empirical distribution function** |
| $AD*$ | Anderson–Darling test |
| $Z_A$ | Zhang–Wu $Z_A$ test |
| $Z_C$ | Zhang–Wu $Z_C$ test |
| **Abbreviations** | **Tests based on measures of skewness and kurtosis** |
| $K^2$ | D'Agostino–Pearson test |
| $DH$ | Doornik–Hansen test |
| $JB$ | Jarque–Bera test |
| $X_{\mathrm{APD}}$ | 2nd-power skewness and kurtosis-based test |
| $T_\omega$ | (Directional) Bonett–Seier test |
| $Z_{\mathrm{EPD}}$ | (Directional) 2nd-power kurtosis-based test |

The empirical power is computed for sample sizes of $n = 10, 20, 50, 100$ and 200, at a significance level of 5%. We use simulated critical values (based on 1,000,000 simulations) for each test to ensure that the true level is 5%. Note that for our tests $X_{\mathrm{APD}}$ and $Z_{\mathrm{EPD}}$, using either the simulated values or the chi-squared (or normal) quantiles has no impact given the high accuracy of the latter. For a given sample size, the empirical power of a test is measured by the proportion of samples (simulated from the alternative distribution) for which the composite hypothesis of normality is rejected. Each calculation of the power is based on 100,000 simulations using the R software with the `PoweR` package (version 1.06) [16].

The results of the study are presented in Appendix C (available in the supplementary material at the publisher's website). The detailed results are given in Tables C1–C3 and the results aggregated by groups of alternatives are given in Tables C4–C8 as follows: 33 symmetric long-tailed (C4), 26 symmetric short-tailed (C5), 26 asymmetric (C6), 59 symmetric (C7) and all 85 alternatives (C8). We focus our analysis on the aggregate results. For each of these tables, the average empirical power is given for each test and for each sample size. The best power among all tests is also given for each sample size in the column labeled 'Best'; this measure is used as a benchmark for the calculation of a total score that we define as follows. For each test and for each of the five sample sizes, we compute the deviation to the best, defined as the difference between the best power (the benchmark) and the power of the test. The total score, for each test, is the average deviation to the best (ADB). This score is reported in the second-to-last line of each table. The smaller the ADB, the better the performance of a test. The tests are ranked in the last line of the tables, based on this score.

Let us take a closer look at Table C7, which gives the average empirical power for the 59 symmetric alternatives, to determine the most powerful tests when it is known that the true distribution is symmetric. The best are the directional tests $\beta_3^2$ and $Z_{\mathrm{EPD}}$, with ADBs of 0.2 and 0.6. They are followed by the directional test $T_\omega$ and the quasi omnibus tests $X_{\mathrm{APD}}$, with ADBs of 3.1 and 3.2. If we take the analysis a step further, we see in Table C4 that a few tests perform well against symmetric long-tailed alternatives. These tests are $DH$, $X_{\mathrm{APD}}$, $JB$, $\beta_3^2$ and $Z_{\mathrm{EPD}}$, with ADBs of 0.4, 0.5, 0.6, 0.8 and 1.6. However, if we look at Table C5, the tests $Z_{\mathrm{EPD}}$ and $\beta_3^2$ clearly emerge as the best against symmetric short-tailed alternatives, with ADBs of 0.7 each, followed by far by $T_\omega$ with an ADB of 5.1.

Consider now Table C8, which gives the average empirical power for all 85 alternatives, to determine the most powerful omnibus tests. The best tests are *CS*, $X_{\mathrm{APD}}$ and *W* with ADBs of 0.4, 0.6 and 0.7. They are followed by the tests $Z_C$, *BCMR* and $Z_A$, with ADBs of 1.0, 1.1 and 1.3, which is also excellent performance. If we take the analysis a step further, we see in Table C6 that the tests $Z_A$, *CS*, *W*, *BCMR* and $Z_C$ clearly appear as the best against asymmetric alternatives, with ADBs of 0.1, 0.7, 0.9, 1.3 and 1.3, while as mentioned above, $X_{\mathrm{APD}}$ is the best omnibus test against symmetric alternatives (after the three directional tests).

It is interesting to compare our results with those given in Table 11 of [19]. We observe that the test $\beta_3^2$ dominates against symmetric alternatives for each of their considered sample sizes ($n = 25,50,100$), which is consistent with our conclusion that $\beta_3^2$ and our test $Z_{\mathrm{EPD}}$ are the most powerful when it is known that the true distribution is symmetric. Furthermore, if we compute the ADB for the powers given in their Table 11, we obtain that *CS*, *W*, $Z_A$, $Z_C$ and *BCMR* are the best omnibus tests, in this order. Again, this is consistent with our conclusion that *CS*, our test $X_{\mathrm{APD}}$, *W*, $Z_C$, *BCMR* and $Z_A$ are the most powerful omnibus tests.

## 8. Conclusion

This paper introduced a new test of normality based on sample 2nd-power skewness $B_2$ and kurtosis $K_2$ (see Definition 3.7), which are alternative measures to the classical sample Pearson's skewness (corresponding to 3rd-power skewness) and kurtosis (corresponding to 4th-power kurtosis). More precisely, the test statistic $X_{\mathrm{APD}}$ is the sum of the squares of what we defined as transformed 2nd-power skewness $Z(B_2)$ and transformed 2nd-power net kurtosis $Z(K_2 - B_2^2)$, two quantities that are virtually independent and closely distributed as a standard normal. Consequently, the distribution of the test statistic $X_{\mathrm{APD}}$ can be approximated, with a very high numerical accuracy, by a $\chi_2^2$ for any sample sizes of $n \geq 10$ (see Proposition 4.4). The test has been derived from Rao's score on the APD family, a generalization of the symmetric EPD to take into account the asymmetry. We thus obtain that the exact asymptotical distribution of $X_{\mathrm{APD}}$ is $\chi_2^2$.

Similarly, we introduced a directional test of normality based on sample 2nd-power kurtosis $K_2$, when the true distribution is known to be symmetric. More precisely, the test statistic $Z_{\mathrm{EPD}}$ is the transformed 2nd-power kurtosis $Z(K_2)$. Consequently, the distribution of the test statistic $Z_{\mathrm{EPD}}$ can thus be approximated very accurately by a $N(0,1)$ for any sample sizes of $n \geq 10$ (see Proposition 5.3). The test has been derived from Rao's score on the symmetric EPD family. We thus obtain that the exact asymptotical distribution of $Z_{\mathrm{EPD}}$ is $N(0,1)$.

We compared our tests, in terms of power, with those generally recognized as the best, in an extensive empirical power analysis against 85 alternatives, divided into symmetric long-tailed, symmetric short-tailed and asymmetric distributions. First, we found that the most powerful tests, when it is known that the true distribution is symmetric, are unequivocally the directional Coin test $\beta_3^2$ and our directional test $Z_{\mathrm{EPD}}$. While a few tests perform well against symmetric long-tailed alternatives, these two tests clearly emerge as the best against symmetric short-tailed alternatives. Note that the Bonett–Seier directional test $T_\omega$ and our quasi omnibus test $X_{\mathrm{APD}}$ follow as the next best tests against symmetric alternatives.

Second, our analysis showed that the most powerful omnibus tests are the Chen–Shapiro test *CS*, our test $X_{\mathrm{APD}}$ and the Shapiro–Wilk test $W$. They are followed closely by the *BCMR* test [7] and the Zhang–Wu tests $Z_C$ and $Z_A$ [24]. Furthermore, our results are consistent with those found in the extensive power analysis of [19].

Finally, we would like to comment on the 'omnibus' property of our test by making a link with its 'robustness'. Note that robustness can be apprehended from three different perspectives. First, a common definition of robustness for a statistical method is its ability to perform correctly outside of its assumed range of validity. In our case, the $X_{\mathrm{APD}}$ and $Z_{\mathrm{EPD}}$ tests can be derived from the Lagrange multiplier method and as such, they are known to have optimal large sample power properties for, respectively, APD and EPD distributions. Note that the large range of tail behavior and asymmetry of the APD makes the $X_{\mathrm{APD}}$ test quasi omnibus. However, in practice, the $X_{\mathrm{APD}}$ and $Z_{\mathrm{EPD}}$ tests are used, respectively, as omnibus test and directional test against all symmetric alternatives. Therefore, the tests need to be robust in the sense that they must exhibit very good power even for distributions not belonging to the APD and EPD families. This issue is addressed in our empirical power study where we showed that the $X_{\mathrm{APD}}$ and $Z_{\mathrm{EPD}}$ tests possess excellent power against alternatives such as the Student's $t$, logistic, beta, $\chi^2$, gamma, Gumbel, log-normal and Weibull distributions.

Second, certain authors (see for instance [10,11]) propose robust normality tests where they replace the non-robust sample mean and sample standard deviation in existing tests by robust (to outliers) estimators of location and scale, such as the median, the median absolute deviation from the median (MAD) or the average absolute deviation from the median (MAAD). In particular, Gel and Gastwirth [10] propose a robust modification of the JB test of normality where they replace the sample standard deviation by the MAAD in the calculation of (non-robust) sample skewness and kurtosis. Note that robustness to outliers here concerns only location and scale estimators, in the sense that the influence of outliers on these estimators is limited. However, the resulting normality test is not robust to outliers. On the contrary, this eventually leads to more powerful directed test against distributions with heavy-tailed alternatives and/or outliers. Further research to study this kind of robustification on the $X_{\mathrm{APD}}$ and $Z_{\mathrm{EPD}}$ tests can be of interest.

Third, we can consider robustness to outliers, in the sense that their influence on a test decision is limited. Suppose that a data set shows strong evidence of normality, except for one or a few extreme observations. A test of normality that is robust to outliers will not reject normality in this situation. As Stehlík *et al.* [22] explain, virtually all common tests for normality lack this kind of robustness. The reason lies simply in the question being asked. Usually, the null hypothesis to be rejected is the normality of data, and it is therefore desirable that the presence of outliers leads to its rejection. If the question is rather about the approximative normality of data, where the null hypothesis to be rejected is the normality of data with possibly a small percentage of contamination by outliers, robust tests are thus desirable. It is important to ask first the good question and then use the appropriate class of tests. In this paper, the $X_{\mathrm{APD}}$ and $Z_{\mathrm{EPD}}$ tests do not search this kind of robustness and are therefore suited if one wants to reject normality for distributions with outliers. However, our tests provide insights on the cause of rejection, e.g. asymmetry, short tails or heavy tails, and thus further analysis on outliers is possible to make an informed decision.

Alternatively, an easy way to modify the $X_{\mathrm{APD}}$ and $Z_{\mathrm{EPD}}$ tests to make them robust to outliers is to follow the adaptive procedures described in Section 3.4 of [22]. It consists simply in manually removing outliers using our preferred method and then using the standard test for normality. For instance, we can remove the smallest 5% of the observations along with the largest 5% (called trimmed method by the authors). Note that the level of the test will be affected and therefore new critical values should be numerically computed by simulations.

In summary, we propose a quasi omnibus test $X_{\mathrm{APD}}$ that offers at least the same benefits as the JB test: a simple test statistic based on measures of skewness and kurtosis that give information on the shape of the distribution. When normality is rejected, practitioners also obtain information on the process, namely if the distribution is asymmetric and/or long-tailed (or short-tailed). This knowledge is often valuable to users and is not available with the $W$, $\beta_3^2$, $CS$, $BCMR$, $Z_C$ or $Z_A$ tests. In addition to those features, the power of our test $X_{\mathrm{APD}}$ is clearly higher than that offered by the tests based on skewness and kurtosis, such as the Jarque–Bera $JB$, D'Agostino–Pearson $K^2$ or Doornik–Hansen $DH$ tests. In fact, in terms of power, it is comparable to the tests of Shapiro–Wilk and Chen–Shapiro, generally accepted as the most powerful. Finally, a key factor for the implementation in software is that the distribution of the test statistic $X_{\mathrm{APD}}$ is approximated, with an unequalled accuracy, by a $\chi_2^2$ for any sample sizes of $n \geq 10$. No tables or simulated quantiles are needed; $p$-values are computed with high precision using the $\chi_2^2$ distribution. We also propose the directional test $Z_{\mathrm{EPD}}$, when it is known that the true distribution is symmetric, which presents essentially the same benefits as its omnibus counterpart.

For all those reasons, we believe that the $X_{\mathrm{APD}}$ test, based on 2nd-power skewness and kurtosis, should be considered as a serious alternative to the JB test, especially in the econometric fields, where the latter is widely used. An implementation of the $X_{\mathrm{APD}}$ test in software, jointly with the directional test $Z_{\mathrm{EPD}}$, is easy and represents a valuable decision aid for practitioners.

## Acknowledgments

We would like to thank the two referees and the Associate Editor for their careful reading of our manuscript.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

The research of Pierre Lafaye de Micheaux was supported by the Natural Sciences and Engineering Research Council of Canada under grant RGPIN-2014-06348.

## References

[1] T.W. Anderson and D.A. Darling , *A test of goodness of fit*, J. Amer. Statist. Assoc. 49 (1954), pp. 765–769.

[2] D.G. Bonett and E. Seier, *A test of normality with high uniform power*, Comput. Stat. Data Anal. 40 (2002), pp. 435–445.

[3] K. Bowman and L. Shenton, *Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and $b_2$*, Biometrika 62 (1975), pp. 243–250.
[4] L. Chen and S.S. Shapiro, *An alernative test for normality based on normalized spacings*, J. Stat. Comput. Simul. 53 (1995), pp. 269–287.
[5] D. Coin, *A goodness-of-fit test for normality based on polynomial regression*, Comput. Stat. Data Anal. 52 (2008), pp. 2185–2198.
[6] R. D'Agostino and E. Pearson, *Tests for departure from normality. Empirical results for the distributions of $b_2$ and $\sqrt{b_1}$*, Biometrika 60 (1973), pp. 613–622.
[7] E. Del Barrio, J. Cuesta-Albertos, C. Matrán, and J. Rodríguez-Rodríguez, *Tests of goodness of fit based on the $l_2$-Wasserstein distance*, Ann. Statist. 27 (1999), pp. 1230–1239.
[8] J.A. Doornik and H. Hansen, *An omnibus test for univariate and multivariate normality*, Oxf. Bull. Econ. Stat. 70 (2008), pp. 927–939.
[9] R.C. Geary, *The ratio of the mean deviation to the standard deviation as a test of normality*, Biometrika 27 (1935), pp. 310–332.
[10] Y.R. Gel and J.L. Gastwirth, *A robust modification of the Jarque–Bera test of normality*, Econ. Lett. 99 (2008), pp. 30–32. Available at http://www.sciencedirect.com/science/article/pii/S0165176507001838.
[11] Y.R. Gel, W. Miao, and J.L. Gastwirth, *Robust directed tests of normality against heavy-tailed alternatives*, Comput. Stat. Data Anal. 51 (2007), pp. 2734–2746. Available at http://www.sciencedirect.com/science/article/pii/S0167947306002805.
[12] C.M. Jarque and A.K. Bera, *A test for normality of observations and regression residuals*, Int. Statist. Rev. 55 (1987), pp. 163–172.
[13] N. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, Vol. 1, 2nd ed., Wiley, New York, 1994.
[14] I. Komunjer, *Asymmetric power distribution: Theory and applications to risk measurement*, J. Appl. Econ. 22 (2007), pp. 891–921.
[15] S. Kotz, T.J. Kozubowski, and K. Podgorski, *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*, Birkhauser, Boston, MA, 2001.
[16] P. Lafaye de Micheaux and V. Tran, *PoweR: A reproducible research tool to ease Monte Carlo power simulation studies for goodness-of-fit tests in R*, J. Stat. Softw. 69 (2016), pp. 1–4.
[17] P. Mantalos, *Three different measures of sample skewness and kurtosis and their effects on the Jarque–Bera test for normality*, Int. J. Comput. Econ. Econom. 2 (2011), pp. 47–62.
[18] C. Rao, *Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation*, Math. Proc. Cambridge Philos. Soc. 44 (1948), pp. 50–57.
[19] X. Romão, R. Delgado and A. Costa, *An empirical power comparison of univariate goodness-of-fit tests for normality*, J. Statist. Comput. Simul. 80 (2010), pp. 545–591.
[20] P. Royston, *Remark as r94: A remark on algorithm as 181: The w-test for normality*, J. R. Statist. Soc. Ser. C 44 (1995), pp. 547–551.
[21] S.S. Shapiro and M.B. Wilk, *An analysis of variance test for normality (complete samples)*, Biometrika 52 (1965), pp. 591–611.
[22] M. Stehlík, L. Střelec, and M. Thulin, *On robust testing for normality in chemometrics*, Chemometr. Intell. Lab. Syst. 130 (2014), pp. 98–108. Available at http://www.sciencedirect.com/science/article/pii/S0169743913001913.
[23] T. Thadewald and H. Büning, *Jarque–Bera test and its competitors for testing normality - a power comparison*, J. Appl. Stat. 34 (2007), pp. 87–105.
[24] J. Zhang and Y. Wu, *Likelihood-ratio tests for normality*, Comput. Stat. Data Anal. 49 (2005), pp. 709–721.