AUSTRIAN JOURNAL OF STATISTICS
Volume 32 (2003), Number 1&2, 163–177

Goodness-of-Fit Test of Shapiro-Wilk Type with Nuisance Regression and Scale

Pranab Kumar Sen, Jana Jurečková, and Jan Picek
University of North Carolina at Chapel Hill
Charles University in Prague
Technical University in Liberec

Abstract: Shapiro and Wilk (1965) proposed a highly intuitive goodness-of-fit test of normality with nuisance location and scale parameters. The test has received a considerable attention in the literature; its asymptotic null distribution is covered by the results of de Wet and Wenter (1973), and was recently studied by Sen (2002).

We extend the Shapiro-Wilk test to the situation with nuisance regression and scale, and construct a test based on the pair of the maximum likelihood estimator and of a pseudo-L-estimator of the standard deviation in the linear regression model. The asymptotic equivalence of these estimators is a characteristic property of the normal distribution of the errors. We shall show that the asymptotic null distribution of the test criterion under the hypothesis of normality is similar to that of the Shapiro-Wilk test in the location-scale model. The main tool of the proof is the second-order asymptotics of L-estimators (see Jurečková and Sen, 1996) extended to pseudo-L-estimators in regression model. The proposed test is numerically compared with a test earlier proposed by the authors (see Jurečková et al., 2003), based on robust estimators of scale.

Keywords: BLUE of Scale Parameter, Goodness-of-fit Test, MLE of Scale Parameter, Shapiro-Wilk Test of Normality.

1 Introduction

Let $Y_1, \ldots, Y_n$ be independent observations following the linear model

$$Y_i = \theta + \mathbf{x}_i'\boldsymbol{\beta} + \sigma e_i,\ i = 1, \ldots, n, \tag{1}$$

where $\mathbf{x}_i \in \mathbb{R}^p,\ i = 1, \ldots, n$ are given regressors, not all equal, $\theta \in \mathbb{R}^1,\ \boldsymbol{\beta} \in \mathbb{R}^p$ and $\sigma > 0$ are unknown intercept, regression and scale parameters, and the errors e_i are independent and identically distributed according to a continuous distribution function F with location 0 and scale parameter 1.

We want to test the hypothesis

$$\mathbf{H}_0 :\ F \equiv \Phi, \quad \text{against} \quad \mathbf{H}_1 :\ F \equiv F_1 \neq \Phi \tag{2}$$

where Φ is the standard normal d.f., F_1 is a general nonnormal d.f., and $\theta,\ \boldsymbol{\beta}$, and σ are treated as nuisance parameters.

For the location-scale model (*i.e.* when $\boldsymbol{\beta} = \mathbf{0}$), Shapiro and Wilk (1965) considered a goodness-of-fit test based on two estimators of σ: L_n, the BLUE (best linear unbiased estimator) under $\mathbf{H}_0$, and $\hat{\sigma}_n$, the maximum likelihood estimator (MLE) under $\mathbf{H}_0$.

If $Y_1, \ldots, Y_n$ are *i.i.d.* observations with the distribution $\mathcal{N}(\mu, \sigma^2)$, then the MLE of σ is $\hat{\sigma}_n$, where

$$\hat{\sigma}_n^2 = n^{-1} \sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2. \tag{3}$$

Let $(Y_1^*, \ldots, Y_n^*)'$ be the vector of the corresponding order statistics. Then the best linear unbiased estimate (BLUE) L_n of σ has the form

$$L_n = \sum_{i=1}^{n} a_{ni} Y_i^* \tag{4}$$

where

$$\mathbf{a}' = (a_1, \ldots, a_n) = (\mathbf{M}_n' \mathbf{V}_n^{-1} \mathbf{M}_n)^{-1} (\mathbf{M}_n' \mathbf{V}_n^{-1}), \quad \mathbf{a}_n' \mathbf{1}_n = 0, \tag{5}$$

where $\mathbf{M}_n = \mathbf{M}$ denotes the vector of expected values of order statistics and $\mathbf{V}_n = \mathbf{V}$ is the corresponding variance matrix. Shapiro and Wilk (1965) modified the BLUE of σ to $L_{n0} = \sum_{i=1}^{n} a_{ni,0} Y_i^*$ where

$$\mathbf{a}_{n0}' = \frac{\mathbf{M}'\mathbf{V}^{-1}}{(\mathbf{M}'\mathbf{V}^{-1}\mathbf{V}^{-1}\mathbf{M})^{1/2}}; \tag{6}$$

then $\mathbf{a}_{n0}' \mathbf{1}_n = 0$ and $\mathbf{a}_{n0}' \mathbf{a}_{n0} = 1$. L_{n0} is asymptotically equivalent to

$$T_n = \frac{1}{n} \sum_{i=1}^{n} \Phi^{-1} \left(\frac{i}{n+1} \right) Y_i^* \tag{7}$$

(see, e.g. Serfling, 1980, p. 267). Let us write the Shapiro-Wilk criterion in the form

$$W_n = n \left(1 - \frac{L_{n0}^2}{\hat{\sigma}_n^2} \right). \tag{8}$$

Two scale estimators L_{n0} and $\hat{\sigma}_n$ are asymptotically (first-order) equivalent if and only if $F \equiv \Phi$, i.e. if the hypothesis of normality is true, while under non-normal alternative F_1 with the finite second moment, the sequence $\sqrt{n}\left(1 - \frac{L_{n0}^2}{\hat{\sigma}_n^2}\right)$ has a nondegenerate asymptotic (normal) distribution. A similar idea was used by Jurečková and Sen (2001) for testing a goodness-of-fit to a general symmetric F_0; the test criterion was a difference of two estimators of location which were asymptotically equivalent (of the second order) if and only if $F \equiv F_0$.

2 Extension of the Shapiro-Wilk Test

In order to find a form of the statistic (8), easily extendable to the linear regression model (1), consider the transformed errors of model (1)

$$\mathbf{e}_n^0 = \sqrt{\tfrac{n}{n-1}}(\mathbf{e}_n - \bar{\mathbf{e}}_n) = \sqrt{\tfrac{n}{n-1}}[\mathbf{I}_n - \mathbf{H}_{n0}]\mathbf{e}_n \tag{1}$$

where

$$\mathbf{H}_{n0} = n^{-1}\mathbf{1}_n'\mathbf{1}_n \tag{2}$$

and write

$$W_n = n\left\{1 - \left(\frac{L_n^0}{S_n^0}\right)^2\right\}, \tag{3}$$

where

$$S_n^0 = \left(n^{-1}(\mathbf{e}_n^0)'(\mathbf{e}_n^0)\right)^{\frac{1}{2}}, \qquad L_n^0 = \sum_{i=1}^{n} a_{ni,0}e_{n:i}^0 \tag{4}$$

are the corresponding pseudoestimators of σ based on $\mathbf{e}_n^0$, where $e_{n:1}^0 \leq \ldots \leq e_{n:n}^0$ are ordered $e_{n1}^0, \ldots, e_{nn}^0$. The asymptotic distribution of general quadratic statistics was studied by de Wet and Wenter (1973), and specifically the asymptotic distribution of the Shapiro-Wilk statistic under $\mathbf{H}_0$ was studied by Sen (2002); it turns out that, as $n \to \infty$,

$$W_n \xrightarrow{\mathcal{D}} \sum_{k\geq 1} \lambda_k Z_k^2 \tag{5}$$

where λ_k are real numbers connected with the coefficients of the quadratic form, and Z_k are *i.i.d.* $\mathcal{N}(0,1)$ variables, $k = 1, 2, \ldots$.

We propose an extension of the Shapiro-Wilk test to the linear regression model (1), treating $\theta,\ \boldsymbol{\beta}$ and σ as nuisance. The test is based on the transformed residuals of the Y_i with respect to the maximum likelihood estimators of the nuisance parameters. The MLE of parameters $\theta,\ \boldsymbol{\beta},\ \sigma$ under normal Φ have the form

$$\begin{aligned}
&\hat{\theta}_n = \bar{Y}_n = n^{-1}\mathbf{1}_n'\mathbf{Y}_n = \theta + \bar{\mathbf{e}}_n,\ \bar{\mathbf{e}}_n = n^{-1}\mathbf{1}_n'\mathbf{e}_n;\\
&\widehat{\boldsymbol{\beta}}_n = (\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n'\mathbf{Y}_n = \boldsymbol{\beta} + \sigma(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{e}_n;\\
&\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^{n}(Y_i - \hat{\theta}_n - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}_n)^2 = \sigma^2 n^{-1}\mathbf{e}_n'[\mathbf{I}_n - \mathbf{H}_{n0} - \mathbf{H}_n]\mathbf{e}_n,
\end{aligned} \tag{6}$$

where

$$\mathbf{H}_n = \left[h_{n,ij}\right]_{i,j=1}^{n} = \mathbf{X}_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n', \tag{7}$$

while $\mathbf{H}_{n0}$ was defined in (2), and $\mathbf{X}_n$ is the $n \times p$ matrix with the rows $\mathbf{x}_1', \ldots, \mathbf{x}_n'$. We shall assume that $\mathbf{X}$ satisfies

$$\begin{aligned}
&\mathbf{X}'\mathbf{1}_n = \mathbf{0}, \quad \mathrm{Rank}(\mathbf{X}_n) = p < n-1\\
&\text{and}\\
&\max_{1\leq i\leq n} h_{n,ii} = \mathcal{O}(n^{-1}) \ \text{ as } \ n \to \infty \ \text{ (the balanced design).}
\end{aligned} \tag{8}$$

The matrices $\mathbf{H}_{n0}$ and $\mathbf{H}_n$ are both of order $n \times n$ and idempotent; moreover, by (8), they satisfy

$$\mathbf{H}_{n0}\mathbf{H}_n = \mathbf{0} = \mathbf{H}_n\mathbf{H}_{n0}; \tag{9}$$

and

$$\mathrm{Tr}(\mathbf{H}_{n0}) = 1 \ \text{ and } \ \mathrm{Tr}(\mathbf{H}_n) = \sum_{i=1}^{n} h_{n,ii} = p. \tag{10}$$

Let $\tilde{\mathbf{r}}_n$ denote the vector of residuals of $\mathbf{Y}$ with respect to the MLE of θ and $\boldsymbol{\beta}$:

$$\tilde{\mathbf{r}}_n = \mathbf{Y}_n - \hat{\theta}\mathbf{1}_n - \mathbf{X}_n\widehat{\boldsymbol{\beta}}_n = \sigma[\mathbf{I}_n - \mathbf{H}_{n0} - \mathbf{H}_n]\mathbf{e}_n = \sigma\tilde{\mathbf{e}}_n \ \text{ (say).} \tag{11}$$

Our test statistic will be based on the standardized residuals

$$\mathbf{r}_n = \mathbf{D}_n^{-\frac{1}{2}}\tilde{\mathbf{r}}_n = \sigma\mathbf{D}_n^{-\frac{1}{2}}[\mathbf{I}_n - \mathbf{H}_{n0} - \mathbf{H}_n]\mathbf{e}_n = \sigma\hat{\mathbf{e}}_n \ \text{ (say)} \tag{12}$$

where

$$\mathbf{D}_n = \mathrm{diag}\left(1 - \tfrac{1}{n} - h_{n,11}, \ldots, 1 - \tfrac{1}{n} - h_{n,nn}\right). \tag{13}$$

Under $\mathbf{H}_0$, the vector $\tilde{\mathbf{r}}_n$ has the normal distribution

$$\mathcal{N}_n(\mathbf{0}, \sigma^2[\mathbf{I}_n - \mathbf{H}_{n0} - \mathbf{H}_n]) \tag{14}$$

while the transformed r_{ni} have the normal $\mathcal{N}(0, \sigma^2)$ distribution, $i = 1, \ldots, n$. Moreover, the correlation between the r_{ni} is only of order $\mathcal{O}(n^{-1})$. It follows from (8), (12) and (14) that $\mathbb{E}(\tilde{r}_{ni} - r_{ni})^2 = \mathcal{O}(n^{-2})$; hence, the difference of the linear estimates of σ based on $\mathbf{r}_n$ and on $\tilde{\mathbf{r}}_n$, respectively, will be only $\mathcal{O}_p(n^{-1})$.

We propose the goodness-of-fit test of the hypothesis (2) of the normality, based on the observations $Y_1, \ldots, Y_n$, following the linear regression model (1) with unknown $\theta,\ \boldsymbol{\beta}$ and σ. The test criterion is

$$\widehat{W}_n = n\left\{1 - \left(\frac{\hat{L}_n}{\hat{s}_n}\right)^2\right\}. \tag{15}$$

where

$$\hat{s}_n^2 = \frac{1}{n}\sum_{i=1}^{n} r_{ni}^2$$
$$\hat{L}_n = \sum_{i=1}^{n} a_{ni}^0 r_{n:i}, \tag{16}$$

are the residual variance and the linear estimator of σ with $a_{ni,0},\ i = 1, \ldots, n$ defined in (6) and $r_{n:i}$ being the order statistics corresponding to the residuals (12). We shall show that the asymptotic null distribution of $\widehat{W}_n$ coincides with that of W_n; hence, the test rejects the hypothesis of the normality on the asymptotic significance level α provided

$$\widehat{W}_n \geq \tau_\alpha \tag{17}$$

where τ_α is the asymptotic critical value of the Shapiro-Wilk test of normality with nuisance location and scale. The coefficients $a_{ni,0},\ i = 1, \dots, n$ and the critical values of the original Shapiro-Wilk test for $n \leq 50$ are tabulated in Shapiro and Wilk (1965). For $n > 50$, the critical values should be calculated, e.g., by a Monte Carlo procedure (see Section 4 devoted to a numerical illustration). To justify this, we must prove the following theorem:

Theorem 2.1 *Let $Y_1, \dots, Y_n$ follow the linear regression model (1) with the standard normal distribution of the errors $e_1, \dots, e_n$. Then,*

$$\widehat{W}_n - W_n = o_p(n^{-1}) \text{ as } n \to \infty. \tag{18}$$

The proof is postponed to the Section 3.

Remark 2.1 *Once we have established the asymptotic equivalence of $\widehat{W}_n$ and W_n, the question of the form of the asymptotic distribution of $\widehat{W}_n$ becomes simpler, because we can refer to the rich literature on the location/scale model (see, e.g. Shapiro, 1998). From the point of view of applications, we are mainly concerned with the case of moderate or small sample sizes. The relation (18) does not provide a good approximation for finite sample sizes, when we can obtain the critical values by means of a simulation. For moderately large sample sizes, we can adopt some resampling technique, to get a better approximation. A methodological justification for this approach was given by Hušková and Janssen (1993), who validated the bootstrapping technique for degenerate U-statistics and related functionals. Incidentally, jackknifing may not be applicable, as the anticipated laws are not normal. More details can be found in Section 4 on numerical illustrations.*

3 Proof of the Asymptotic Null Distribution of $\widehat{W}_n$

Because $\hat{L}_n$ and $\hat{s}_n$ are both scale-equivariant, we may put $\sigma = 1$, without loss of generality. Let $\mathbf{e}_n^0$ and $\mathbf{r}_n$ be the residuals given in (1) and (12), respectively. Then under $\mathbf{H}_0$,

$$\mathbf{e}_n^0 \overset{\mathcal{D}}{=} \mathcal{N}_n(\mathbf{0}, \mathbf{\Lambda}_n^0), \quad \mathbf{r}_n \overset{\mathcal{D}}{=} \mathcal{N}_n(\mathbf{0}, \mathbf{\Lambda}_n) \tag{1}$$

where

$$\mathbf{\Lambda}_n^0 = \frac{n}{n-1}[\mathbf{I}_n - \mathbf{H}_{n0}] \text{ and } \mathbf{\Lambda}_n = \mathbf{D}_n^{-\frac{1}{2}}[\mathbf{I}_n - \mathbf{H}_{n0} - \mathbf{H}_n]\mathbf{D}_n^{-\frac{1}{2}}. \tag{2}$$

Notice that the diagonal elements of $\mathbf{\Lambda}_n^0$ and of $\mathbf{\Lambda}_n$ are all equal to 1. (1) and (12) imply that

$$\mathbf{r}_n - \mathbf{e}_n^0 \sim \mathcal{N}_n(\mathbf{0}, \mathbf{\Gamma}_n) \tag{3}$$

where

$$\mathbf{\Gamma}_n = \Big[\gamma_{n,ij}\Big]_{i,j=1}^{n} \text{ and } \max_{1 \leq i \leq n} |\gamma_{n,ij}| = \mathcal{O}(n^{-1}) \tag{4}$$

by (8). This together with the Bernstein inequality implies that

$$\max_{1\le i\le n} |r_{ni} - e^0_{ni}| = \mathcal{O}\left(\sqrt{\frac{\log\, n}{n}}\right) \text{ a.s., as } n \to \infty \tag{5}$$

and hence also

$$\max_{1\le i\le n} |r_{n:i} - e^0_{n:i}| = \mathcal{O}\left(\sqrt{\frac{\log\, n}{n}}\right) \text{ a.s., as } n \to \infty \tag{6}$$

where $r_{n:1} < \ldots < r_{n:n}$ and $e^0_{n:1}, \ldots < e^0_{n:n}$ are the sets of order statistics for $\mathbf{r}_n$ and $\mathbf{e}^0_n$, respectively. Denote

$$Z_n(x) = n^{\frac{3}{4}}[\hat{F}_n(x) - F^0_n(x)],\ x \in I\!R \tag{7}$$

where

$$F^0_n(x) = \tfrac{1}{n}\sum_{i=1}^{n} I[e^0_{ni} \le x], \quad \hat{F}_n(x) = \tfrac{1}{n}\sum_{i=1}^{n} I[r_{ni} \le x],\ x \in I\!R \tag{8}$$

are the respective empirical distribution functions of the residuals. Then $I\!EZ_n(x) = 0$ and

$$\begin{aligned} I\!EZ^2_n(x) &= n^{-\frac{1}{2}}\sum_{i=1}^{n}[2\Phi(x) - 2I\!P(r_{ni} \le x, e^0_{ni} \le x)] \\ &+ n^{-\frac{1}{2}}\sum_{i\ne j}^{n}\Big[I\!P(r_{ni} \le x, r_{nj} \le x)) + I\!P(e^0_{ni} \le x, e^0_{nj} \le x) \\ &- I\!P(r_{ni} \le x, e^0_{nj} \le x) - I\!P(e^0_{ni} \le x, r_{nj} \le x)\Big],\ x \in I\!R \end{aligned} \tag{9}$$

and, in view of (1), (12) and (1)-(4), this further implies that

$$I\!EZ^2_n(x) \le K\phi(x)\{1 + \mathcal{O}(n^{-\frac{1}{2}})\},\ x \in I\!R,\ K < \infty,\ \phi(x) = \frac{d\Phi(x)}{dx}. \tag{10}$$

Similarly we can find the upper bound for the $2k$-th moment of $Z_n(x)$:

$$I\!EZ^{2k}_n(x) \le K^*(\phi(x))^k\{1 + \mathcal{O}(n^{-\frac{1}{2}})\},\ x \in I\!R \tag{11}$$

as $n \to \infty$, where K^* is a finite constant, independent of n (but it may depend on k). Hence, by (7) and (11),

$$I\!P\left\{n^{\frac{5}{8}}|\hat{F}_n(x) - F^0_n(x)| > \varepsilon\right\} \le n^{-\frac{k}{4}}C^*(\varepsilon) \tag{12}$$

for every $\varepsilon > 0$ and some $C^*(\varepsilon) < \infty$. Specifically, using (12) with $k > 8$, we obtain

$$\max_{1\le i\le n}\left\{n^{\frac{5}{8}}\left|\hat{F}_n\left(\tfrac{i}{n+1}\right) - F^0_n\left(\tfrac{i}{n+1}\right)\right|\right\} = o(1) \text{ a.s., as } n \to \infty \tag{13}$$

and because both $\hat{F}_n$ and F^0_n are monotone, we conclude from (13) that

$$\sup_{x\in R}\left\{n^{\frac{5}{8}}|\hat{F}_n(x) - F^0_n(x)|\right\} = o(1) \text{ a.s., as } n \to \infty, \tag{14}$$

hence

$$\max_{1\le j\le n}|F_n^0(r_{n:j})-\tfrac{j}{n}|=o\left(n^{-\frac{5}{8}}\right)\ \text{a.s.} \tag{15}$$

Because of the identity

$$F_n^0(x)=\tfrac{1}{n}\sum_{i=1}^{n}I\left[e_i\le\sqrt{\tfrac{n-1}{n}}x+\bar e_n\right]=F_n\left(\sqrt{\tfrac{n-1}{n}}x+\bar e_n\right),$$

the Bahadur representation (see Bahadur, 1966) yields

$$\begin{aligned}\sup\Big\{&\left|F_n^0(x)-F_n^0(y)-\Phi\left(\sqrt{\tfrac{n-1}{n}}x+\bar e_n\right)+\Phi\left(\sqrt{\tfrac{n-1}{n}}y+\bar e_n\right)\right|\\ &:|x-y|\le n^{-\frac{1}{2}}(\log\ n)^{\frac{1}{2}}\Big\}=\mathcal{O}\left(n^{-\frac{3}{4}}\log\ n\right)\ \text{a.s. as } n\to\infty.\end{aligned} \tag{16}$$

Inserting $x=r_{n:j},\ y=e^0_{n:j}$ in (16), we obtain with the aid of (6) and (13) that

$$\begin{aligned}&\max_{1\le j\le n}\left\{\left|\Phi\left(\sqrt{\tfrac{n-1}{n}}r_{n:j}+\bar e_n\right)-\Phi\left(\sqrt{\tfrac{n-1}{n}}e^0_{n:j}+\bar e_n\right)\right|\right\}\\ &=\max_{1\le j\le n}\left\{\left|\Phi\left(\sqrt{\tfrac{n-1}{n}}r_{n:j}+\bar e_n\right)-\Phi(e_{n:j})\right|\right\}\\ &=\max_{1\le j\le n}\left\{\left|\Phi\left(\sqrt{\tfrac{n-1}{n}}r_{n:j}+\bar e_n\right)-U_{n:j}\right|\right\}=o\left(n^{-\frac{5}{8}}\right)\ \text{a.s. as } n\to\infty\end{aligned} \tag{17}$$

where $U_{n:j}$ is the order statistic corresponding to the sample of size n from the uniform $R[0,1]$ distribution. Then (17) implies

$$\max_{n^{\frac{15}{16}}\le j\le n-n^{\frac{15}{16}}}|r_{n:j}-e^0_{n:j}|=o\left(n^{-\frac{1}{2}-\frac{1}{15}}\right)\ \text{a.s. as } n\to\infty. \tag{18}$$

For $j<n^{\frac{15}{16}}$ or $j>n-n^{\frac{15}{16}}$ we shall use the approximation (6).

It follows from (3) and (15) that

$$\begin{aligned}&\frac{L_n^0}{S_n^0}=\left(1-\frac{W_n}{n}\right)^{\frac{1}{2}}=1-\frac{W_n}{2n}+\mathcal{O}_p(n^{-2})\\ \text{and}&\\ &\frac{\hat L_n}{\hat s_n}=1-\frac{\widehat{W}_n}{2n}+\mathcal{O}_p(n^{-2})\end{aligned} \tag{19}$$

and it further follows from (4), (6) and (12) that

$$\hat s_n^2=S_n^{02}-\frac{1}{\sigma^2 n}\left\{\left(\widehat{\boldsymbol\beta}-\boldsymbol\beta\right)'(\mathbf{X}_n'\mathbf{X}_n)\left(\widehat{\boldsymbol\beta}-\boldsymbol\beta\right)\right\}+o_p(n^{-1}) \tag{20}$$

where $\frac{1}{\sigma^2}\left(\widehat{\boldsymbol\beta}-\boldsymbol\beta\right)'(\mathbf{X}_n'\mathbf{X}_n)\left(\widehat{\boldsymbol\beta}-\boldsymbol\beta\right)=\mathbf{e}_n'\mathbf{H}_n\mathbf{e}_n=\mathcal{O}_p(1)$. Therefore,

$$S_n^{02}=\hat s_n^2+\frac{1}{n}(\mathbf{e}_n'\mathbf{H}_n\mathbf{e}_n)+o_p(n^{-1}), \tag{21}$$

and hence,

$$S_n^0 = \hat{s}_n \left\{ 1 + \frac{1}{2} \frac{\mathbf{e}_n' \mathbf{H}_n \mathbf{e}_n}{\mathbf{e}_n' (\mathbf{I} - \mathbf{H}_{n0}) \mathbf{e}_n} + o_p(n^{-1}) \right\} \tag{22}$$

By (19),

$$\widehat{W}_n - W_n^0 = 2n \left\{ \frac{L_n^0}{S_n^0} - \frac{\hat{L}_n}{\hat{s}_n} \right\} + o_p(1) = \frac{2n}{S_n^0} \left\{ L_n^0 - \frac{S_n^0}{\hat{s}_n} \hat{L}_n \right\} + o_p(1) \tag{23}$$

$$= \frac{2n}{S_n^0} \left\{ L_n^0 - \left[1 + \frac{1}{2} \frac{\mathbf{e}_n' \mathbf{H}_n \mathbf{e}_n}{\mathbf{e}_n' (\mathbf{I} - \mathbf{H}_{n0}) \mathbf{e}_n} + o_p(n^{-1}) \right] \hat{L}_n \right\} + o_p(1)$$

$$= \frac{2n}{S_n^0} (L_n^0 - \hat{L}_n) - \frac{\mathbf{e}_n' \mathbf{H}_n \mathbf{e}_n}{s_n^{03}} \hat{L}_n + o_p(1).$$

Since $S_n^0 \xrightarrow{p} 1$ and $L_n^0 \xrightarrow{p} 1$ as $n \to \infty$ and $\hat{L}_n = L_n^0 + (\hat{L}_n - L_n^0)$, it suffices to show that $\hat{L}_n - L_n^0 \xrightarrow{p} 0$ as $n \to \infty$. Denote $\xi_{ni} = \Phi^{-1}\left(\frac{i}{n+1}\right),\ i = 1, \ldots, n$. Using the approximation (7), we have

$$\hat{L}_n - L_n^0 = \frac{1}{n} \sum_{i=1}^{n} (r_{ni} - e_{ni}^0) \xi_{ni}$$

$$= \frac{1}{n} \sum_{i=1}^{n} (r_{ni} - e_{ni}^0) \left\{ \frac{1}{2}(r_{ni} + e_{ni}^0) - \left[\frac{1}{2}(r_{ni} + e_{ni}^0) - \xi_{ni} \right] \right\} \tag{24}$$

$$= \frac{1}{2n} \left(\sum_{i=1}^{n} r_{ni}^2 - \sum_{i=1}^{n} e_{ni}^{02} \right) - \frac{1}{n} \sum_{i=1}^{n} (r_{ni} - e_{ni}^0) \left[\frac{1}{2}(r_{ni} + e_{ni}^0) - \xi_{ni} \right]$$

$$= \frac{1}{2}(\hat{s}_n^2 - S_n^{02}) - B_n \quad \text{(say).}$$

Using (22), the first term on the right-hand side of (24) is

$$-\frac{1}{2n} (\mathbf{e}_n' \mathbf{H}_n \mathbf{e}_n) + o_p(n^{-1}). \tag{25}$$

Applying the Cauchy-Schwarz inequality to B_n, we obtain that

$$B_n^2 \le \left\{ \frac{1}{n} \sum_{i=1}^{n} (r_{ni} - e_{ni}^0)^2 \right\} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{2}(r_{ni} + e_{ni}^0) - \xi_{ni} \right]^2 \right\}. \tag{26}$$

Using (5) for $i < n^{\frac{15}{16}}$ and $i > n - n^{\frac{15}{16}}$, and (18) for $n^{\frac{15}{16}} \le i < n - n^{\frac{15}{16}}$, we obtain

$$\frac{1}{n} \sum_{i=1}^{n} (r_{ni} - e_{ni}^0)^2 = \mathcal{O}\left(n^{-1-\frac{1}{16}} \log n \right). \tag{27}$$

On the other hand, it follows from (5) and from the convergence properties of the $e_{n:i}^0$ proved in Hoeffding (1953) that

$$\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{1}{2}(r_{n:i} + e_{n:i}^0) - \xi_{ni} \right\}^2 = \mathcal{O}_p\left(\frac{\log n}{n} \right), \quad \text{as } n \to \infty. \tag{28}$$

Therefore, it follows from (26)–(28)

$$B_n^2 = \mathcal{O}_p\left(\left(\frac{\log n}{n^2}\right)^2 \frac{1}{n^{\frac{1}{16}}}\right) = o_p\left(\frac{1}{n^2}\right), \text{ as } n \to \infty. \tag{29}$$

As a result,

$$\hat{L}_n - L_n^0 = -\frac{1}{2n}(\mathbf{e}_n'\mathbf{H}_n\mathbf{e}_n) - o_p\left(\frac{1}{n}\right), \tag{30}$$

so that by (23) and (30),

$$\widehat{W}_n - W_n = o_p\left(\frac{1}{n}\right), \tag{31}$$

and that completes the proof of the theorem. □

4 Numerical Illustration

We shall illustrate the performance of the proposed test on simulated regression models. As the regression matrix is used the Mayer matrix (based on real historical physical measurements) of order (27×3), or its multiple, with the first column $\mathbf{1}_n$; the value of the nuisance regression parameter vector is $\boldsymbol{\beta} = (14.5,\ 1.542417,\ 0.05)'$. For more details on this model we refer to Jurečková et al. (2003).

In our numerical experiment, the errors were generated by sampling from the following alternative densities:

normal $\mathcal{N}(0,1)$:	$f(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$	normal $\mathcal{N}(0,16)$:	$f(x) = \frac{1}{4\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{32}}$
logistic $(0,1)$:	$f(x) = \frac{\mathrm{e}^{-x}}{(1+\mathrm{e}^{-x})^2}$	logistic $(0,4)$:	$f(x) = \frac{\mathrm{e}^{-x/4}}{(1+\mathrm{e}^{-x/4})^2}$
Laplace $(0,1)$:	$f(x) = \frac{1}{2}\mathrm{e}^{-\lvert x\rvert}$	Laplace $(0,4)$:	$f(x) = \frac{4}{2}\mathrm{e}^{-4\lvert x\rvert}$
Cauchy:	$f(x) = \frac{1}{\pi(1+x^2)}$.		

In order to gain an insight into larger sample size behavior of our proposed test, we also replicate the Mayer design matrix 4, 8, 16 and 32 times, resulting in the design matrix of 108, 216, 432 and 864 rows, respectively; in each case the errors e_i are generated to insure independence.

1000 replications were simulated for each case. Based on these data, we calculated the test statistics

$$\widehat{W}_n = n\left(1 - \frac{\hat{L}_n^2}{\hat{s}_n^2}\right) = n\left(1 - \frac{\left(\sum_{i=1}^n a_{ni,0}r_{n:i}\right)^2}{\sum_{i=1}^n r_{ni}^2}\right), \tag{1}$$

where $a_{ni,0}$ and r_{ni} are defined in (6) and (12), respectively, $i = 1, \ldots, n$.

Because the asymptotic null distributions of the test statistics $\widehat{W}_n$ and coefficients $(a_1, \ldots, a_n)$ are not known for $n > 50$, they were approximated in the following way:

The coefficients $a_i,\ i = 1, \ldots, n,\ n > 50$ were approximated as suggested in Shapiro and Wilk (1965); namely, we calculated

$$\hat{a}_i^* = 2M_i \quad (i = 2, 3, \ldots, n-1)$$

and

$$\hat{a}_1^2 = \hat{a}_n^2 = \frac{\Gamma\left(\frac{1}{2}(n+1)\right)}{\sqrt{2}\Gamma\left(\frac{1}{2}n+1\right)}.$$

The coefficients $\hat{a}_1, \hat{a}_n$ are directly usable but $\hat{a}_i^*\ i = 2, \ldots, n-1,$ must be normalized in the following way:

$$\hat{a}_i = \hat{a}_i^* \sqrt{\frac{1 - 2a_1^2}{\sum_{i=2}^{n-1} \hat{a}_i^{*2}}} \quad (i = 2, \ldots, n-1)$$

The expected values M_i of order statistics, $i = 2, \ldots, n-1$, were computed by the numerical integration. A similar approach was used by Royston (1982b), Royston (1982a), Royston (1995), and his algorithm is used in many software packages offering the Shapiro-Wilk test.

The distribution of the test statistic was then approximated by the following Monte Carlo procedure: For a fixed $n,$ a random sample of size n from the normal distribution was generated, and $\widehat{W}_n$ was computed with the coefficients $\hat{a}_i,\ i = 1, \ldots, n.$ This random experiment was repeated 100 000 times. For the sake of comparison, the nonparametric test, proposed by the authors in Jurečková et al. (2003) was performed for testing the normality on the same data. This test is based on the criterion

$$T_n^* = n^{\frac{1}{2}} \left\{ \log \frac{S_{n0}}{S_{n1}} - \log\ \xi(F_0) \right\}$$

(see Jurečková et al., 2003,for the details of the notation). Tables 1–6 give the numbers of rejections of $\mathbf{H}_0$ (among 1000 tests) for both test criteria described above. Tables 5 and 6 are based on the same vectors of the simulated errors, for the sake of comparison; while Table 5 is based on the sixteenfold Mayer matrix, the slope columns of design matrix used for Table 6 are simulated from the uniform distribution and standardized to have an average zero.

4.1 Conclusion

The numerical results illustrate that the Shapiro-Wilk type test distinguishes well the normal distribution of the errors in the linear regression models from the above distribution shapes. With its easy calculation and a high efficiency, the proposed test of the Shapiro-Wilk type is practically appealing.

Notice that for moderate samples the empirical power of the Shapiro-Wilk test is higher than the power of the test proposed in Jurečková et al. (2003) (cf. Tables 1 and 2), while for larger sample it seems to be an opposite case (cf. Tables 3–6). A comparison of Tables 5 and 6 shows that this is not due to the choice of the matrix. This fact can be explained by the way how were both tests performed: While the Shapiro-Wilks critical values were calculated for each fixed n in the above manner, the critical values of the test of Jurečková et al. (2003) are purely asymptotic; and the asymptotics naturally works better for larger n.

Table 1: Numbers of rejections of $\mathbf{H}_0$ among 1000 cases on level α for Mayer matrix (27×3)

	α=0.01		α=0.05		α=0.1	
Distribution of errors	$\widehat{W}_n$	T_n^*	$\widehat{W}_n$	T_n^*	$\widehat{W}_n$	T_n^*
Normal $N(0,1)$	19	12	43	60	102	143
Normal $N(0,16)$	9	12	52	57	91	119
Logistic $(0,1)$	42	22	126	76	179	137
Logistic $(0,4)$	42	18	111	76	180	131
Laplace $(0,1)$	131	53	275	127	361	210
Laplace $(0,4)$	138	50	255	120	324	214
Cauchy $(0,1)$	828	624	900	720	922	763

Table 2: Numbers of rejections of $\mathbf{H}_0$ among 1000 cases on level α for quadruple Mayer matrix (108×3)

	α=0.01		α=0.05		α=0.1	
Distribution of errors	$\widehat{W}_n$	T_n^*	$\widehat{W}_n$	T_n^*	$\widehat{W}_n$	T_n^*
Normal $N(0,1)$	8	12	44	60	104	143
Normal $N(0,16)$	8	12	47	57	95	119
Logistic $(0,1)$	51	22	119	76	174	137
Logistic $(0,4)$	44	18	112	76	158	131
Laplace $(0,1)$	333	53	499	127	581	210
Laplace $(0,4)$	354	50	529	120	606	214
Cauchy $(0,1)$	1000	624	1000	720	1000	763

Table 3: Numbers of rejections of $\mathbf{H}_0$ among 1000 cases on level α for eightfold Mayer matrix (216×3)

	α=0.01		α=0.05		α=0.1	
Distribution of errors	$\widehat{W}_n$	T_n^*	$\widehat{W}_n$	T_n^*	$\widehat{W}_n$	T_n^*
Normal $N(0,1)$	10	9	48	56	95	108
Normal $N(0,16)$	10	14	50	58	100	114
Logistic $(0,1)$	52	88	95	205	146	305
Logistic $(0,4)$	44	86	99	196	143	288
Laplace $(0,1)$	543	664	727	837	783	891
Laplace $(0,4)$	560	685	718	831	797	890
Cauchy $(0,1)$	1000	1000	1000	1000	1000	1000

Table 4: Numbers of rejections of $\mathbf{H}_0$ among 1000 cases on level α for sixteenfold Mayer matrix (432×3)

	α=0.01		α=0.05		α=0.1	
Distribution of errors	$\widehat{W}_n$	T_n^*	$\widehat{W}_n$	T_n^*	$\widehat{W}_n$	T_n^*
Normal $N(0,1)$	10	16	57	54	103	109
Normal $N(0,16)$	8	9	55	51	100	100
Logistic $(0,1)$	55	198	96	388	146	493
Logistic $(0,4)$	47	183	95	367	133	487
Laplace $(0,1)$	866	966	939	991	961	995
Laplace $(0,4)$	854	962	906	990	937	995
Cauchy $(0,1)$	1000	1000	1000	1000	1000	1000

Table 5: Numbers of rejections of $\mathbf{H}_0$ among 1000 cases on level α for $32\times$ Mayer matrix (864×3)

	α=0.01		α=0.05		α=0.1	
Distribution of errors	W_n	T_n	W_n	T_n	W_n	T_n
Normal $N(0,1)$	18	15	59	49	105	109
Normal $N(0,16)$	14	12	53	49	107	100
Logistic $(0,1)$	76	459	134	688	205	770
Logistic $(0,4)$	72	427	132	652	172	756
Laplace $(0,1)$	991	1000	999	1000	1000	1000
Laplace $(0,4)$	991	1000	998	1000	998	1000
Cauchy $(0,1)$	1000	1000	1000	1000	1000	1000

Table 6: Numbers of rejections of $\mathbf{H}_0$ among 1000 cases on level α for simulated matrix (864×2)

	α=0.01		α=0.05		α=0.1	
Distribution of errors	W_n	T_n	W_n	T_n	W_n	T_n
Normal $N(0,1)$	16	11	61	50	105	86
Normal $N(0,16)$	9	7	52	47	100	96
Logistic $(0,1)$	78	452	141	678	200	763
Logistic $(0,4)$	71	403	131	630	181	735
Laplace $(0,1)$	993	999	999	1000	1000	1000
Laplace $(0,4)$	990	1000	996	1000	998	1000
Cauchy $(0,1)$	1000	1000	1000	1000	1000	1000

Acknowledgements

Research of the first author was supported by the Boshamer Research support, a component of the Boshamer professorship endowed to him. Research of the second author was supported by the grants GAČR 201/02/0621 and MSM 113200008. Research of the third author was supported by the grants GAČR 201/02/0049 and MSM 245100303.

References

R.R. Bahadur. A note of quantiles in large samples. *Ann. Math. Statist.*, 37:557–580, 1966.

T. de Wet and J.H. Wenter. Asymptotic distributions of quadratic forms with application to test of fit. *Ann. Statist.*, 31:276–295, 1973.

W. Hoeffding. On the distribution of the expected values of the order statistics. *Ann. Math. Statist.*, 24:93–100, 1953.

M. Hušková and P. Janssen. Consistency of generalized bootstrap for degenerate U-statistics. *Ann. Statist.*, 21:1811–1823, 1993.

J. Jurečková, J. Picek, and P.K. Sen. Goodness-of-fit tests with nuisance regression and scale. *Metrika (to appear)*, 2003.

J. Jurečková and P.K. Sen. *Robust Statistical Procedures: Asymptotics and Interrelations*. J. Wiley, New York, 1st edition, 1996.

J. Jurečková and P.K. Sen. Goodness-of-fit tests and second order asymptotic relations. *Journ. Statist. Planning and Inference*, 91:377–397, 2001.

P. Royston. Algorithm AS 181: The W test for normality. *Applied Statistics*, 31:176–180, 1982a.

P. Royston. An extension of Shapiro and Wilk W test for normality to large samples. *Applied Statistics*, 31:115–124, 1982b.

P. Royston. A remark to algorithm AS 181: The W test for normality. *Applied Statistics*, 44:547–551, 1995.

P.K. Sen. Shapiro-Wilk type goodness-of-fit tests for normality: Asymptotics revisited. In C. Huber-Carol, N. Balakrishnan, M.S. Nikulin, and M. Mesbah, editors, *Goodness-of-Fit Tests and Model Validity*, pages 73–88. Birkhäuser, Boston, 2002.

R.J. Serfling. *Approximation Theorems of Mathematical Statistics*. J. Wiley, New York, 1st edition, 1980.

S.S. Shapiro. Distribution assessment. In N. Balakrishnan and C.R. Rao, editors, *Handbook of Statistics, Vol 17: Order Statistics: Applications*, pages 475–494. Elsevier, Amsterdam, 1998.

S.S. Shapiro and M.B. Wilk. An analysis of variance for normality (complete samples). *Biometrika*, 52:591–611, 1965.

Authors' addresses:

Pranab Kumar Sen
Departments of Biostatistics and Statistics
The University of North Carolina
3105E McGavran - Greenberg Hall, CB #7420
Chapel Hill, NC 27599-7420
USA

Tel. +919 9667274
Fax +919 9663804
E-mail: pksen@bios.unc.edu
http://www.bios.unc.edu/

Jana Jurečková
Charles University in Prague
Department of Statistics
Sokolovská 83
CZ-186 75 Prague 8
Czech Republic

Tel. +420 221913285
Fax +420 222323316
E-mail jurecko@karlin.mff.cuni.cz
http://www.karlin.mff.cuni.cz/∼jurecko

Jan Picek
Technical University in Liberec
Department of Applied Mathematics
Hálkova 6
CZ-461 17 Liberec
Czech Republic

Tel. +420 485354263
Fax +420 485354301
E-mail: jan.picek@vslib.cz
http://www.fp.vslib.cz/picek