TESTING SYMMETRY OF THE ERRORS OF A LINEAR MODEL

THOMAS P. HETTMANSPERGER
Penn State University

JOSEPH W. MCKEAN
Western Michigan University

SIMON J. SHEATHER
University of New South Wales

A test for symmetry of the distribution of the errors in a linear model is proposed. It is a goodness-of-fit type test based on the discrepancy between two robust fits. The first fit is appropriate under symmetric errors while the second is appropriate for skewed as well as symmetric distributions. The proposed test is robust and is asymptotically distribution free. Besides deriving the test statistic's null asymptotic distribution, its efficiency under a general class of local alternatives is obtained which allows for the determination of the test's asymptotic relative efficiency with its competitors.

1. Introduction

Constance van Eeden has made many important contributions to the development of signed rank procedures. Symmetry is a crucial assumption necessary for the validity of such procedures. In many situations encountered in practice, a test of symmetry is quite useful. For example, consider a randomized paired design. Under the null hypothesis of no treatment effect, the paired differences are symmetrically distributed; however, under alternatives that involve a change in scale as well as one in location, this is not true.

In this paper, we propose a test for the hypothesis that the errors in a linear model are symmetrically distributed. It is a goodness-of-fit type test based on the discrepancy between two robust rank-based fits. The first fit uses a robust signed-rank (SR) fitting criterion that is appropriate under the assumption of symmetric errors. It yields the distance, the minimum of the objective function, $D_{\text{SR}}(\widehat{\mathbf{Y}}_{\text{SR}})$, between the vector of responses, $\mathbf{Y}$, and the vector of fitted values, $\widehat{\mathbf{Y}}_{\text{SR}}$. The second fitted vector, $\widehat{\mathbf{Y}}_{\text{R}}$, is based on a robust rank (R) fitting criterion that is appropriate for either symmetric or asymmetric error distributions. For this second fit we obtain $D_{\text{SR}}(\widehat{\mathbf{Y}}_{\text{R}})$ the distance between $\mathbf{Y}$ and $\widehat{\mathbf{Y}}_{\text{R}}$ using the symmetric "yardstick," i.e., distance based on the SR norm. The test statistic is the standardized difference in these distances, $H_{\varphi} = RD/\widehat{\delta}$ where $RD = D_{\text{SR}}(\widehat{\mathbf{Y}}_{\text{R}}) - D_{\text{SR}}(\widehat{\mathbf{Y}}_{\text{SR}})$; see (2.13). Under symmetry, the fits, and hence the distances, should be similar. Thus the null hypothesis of symmetry is rejected for large values of RD.

Keywords and phrases: asymptotic distribution-free; asymptotic relative efficiency; linear rank scores; rank-based regression; robust; signed-rank regression; Wilcoxon scores.

In Section 2, we describe the new test of symmetry and derive its asymptotic null distribution. We discuss what score functions to use for the test of symmetry in Section 3. Finally we consider the behavior of the newly proposed test under contiguous alternatives in Section 5.

2. The test statistic H_φ and its null asymptotic distribution

Consider the linear model,

$$\mathbf{Y} = \alpha\mathbf{1} + \mathbf{X}_c\boldsymbol{\beta} + \mathbf{e}, \tag{2.1}$$

where $\mathbf{1}$ is a vector of n ones, $\mathbf{X}_c$ is a $n \times p$ design matrix of full column rank, and $\mathbf{e}$ is a vector of iid random errors with common density f and distribution function F. Since the model includes an intercept there is no loss in generality in assuming that the $\mathbf{X}_c$ is centered and that med $e_i = F^{-1}(\frac{1}{2}) = 0$. Denote the augmented matrix $[\mathbf{1} : \mathbf{X}_c]$ by $\mathbf{X}$ and let $\mathbf{b} = (\alpha, \boldsymbol{\beta}')'$. Let Ω denote the column space of $\mathbf{X}$. We are interested in the hypotheses of symmetry,

$$H_0 : \; f(-x) = f(x) \quad \text{versus} \quad H_A : \; f(-x) \neq f(x). \tag{2.2}$$

2.1. SR and R fits

As mentioned earlier, our test of symmetry is based on two fits, a signed-rank (SR) fit and a rank (R) fit. Under the assumption of symmetry, the SR estimates are consistent estimators while the R estimates are consistent estimators under both symmetry and asymmetry. First we briefly present the R estimates.

R estimates are based on rank regression scores. These are generated as $a_\varphi(i) = \varphi\big(i/(n+1)\big)$ where $\varphi(u)$ is a square-integrable, nondecreasing score function defined on the interval $(0,1)$, which, without loss of generality, is standardized as follows: $\int_0^1 \varphi(u)\,du = 0$ and $\int_0^1 \varphi^2(u)\,du = 1$. We will further assume that the scores are odd about $\frac{1}{2}$; i.e., $\varphi(1-u) = -\varphi(u)$. The most widely used such score is the Wilcoxon given by $\varphi(u) = \sqrt{12}(u - \frac{1}{2})$.

The R estimate based on the score function $\varphi(u)$ is given by

$$\widehat{\boldsymbol{\beta}}_{\mathrm{R}} = \operatorname{Argmin} \|\mathbf{Y} - \mathbf{X}_c\boldsymbol{\beta}\|_{\mathrm{R}}, \tag{2.3}$$

where $\|\cdot\|_{\mathrm{R}}$ is defined by

$$\|\mathbf{v}\|_{\mathrm{R}} = \sum_{i=1}^{n} a_\varphi(R(v_i))v_i, \quad \text{for } \mathbf{v} \in R^n, \tag{2.4}$$

and the $R(v_i)$ denotes the rank of v_i among $v_1, \ldots, v_n$. This is a pseudo-norm on R^n. It has all the properties of a norm except that the property

$\|\mathbf{v}\| = 0$ iff $\mathbf{v} = \mathbf{0}$ is replaced by $\|\mathbf{v}\|_{\mathrm{R}} = 0$ iff $\mathbf{v} = a\mathbf{1}$, for any scalar a. Hence, the intercept α cannot be estimated using a pseudo-norm. Instead, we will estimate it by

$$\widehat{\alpha} = \mathrm{med}\{Y_i - \mathbf{x}'_{ci}\widehat{\boldsymbol{\beta}}_{\mathrm{R}}\}. \tag{2.5}$$

The R estimate of $\mathbf{b}$ is $\widehat{\mathbf{b}}_{\mathrm{R}} = (\widehat{\alpha}, \widehat{\boldsymbol{\beta}}'_{\mathrm{R}})'$. Under the regularity conditions found in the Appendix, $\widehat{\mathbf{b}}_{\mathrm{R}}$ has an approximate

$$N_{p+1}\left(\begin{pmatrix}\alpha\\ \boldsymbol{\beta}\end{pmatrix}, \begin{bmatrix} n^{-1}\tau_{\mathrm{S}}^2 & \mathbf{0}' \\ \mathbf{0} & \tau_{\varphi}^2(\mathbf{X}_{\mathrm{c}}'\mathbf{X}_{\mathrm{c}})^{-1}\end{bmatrix}\right) \tag{2.6}$$

distribution; where the scale parameter τ_φ is given by,

$$\tau_\varphi = \left[\int_0^1 \varphi(u)\left(-\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}\right) du\right]^{-1}, \tag{2.7}$$

and the scale parameter τ_{S} is given by,

$$\tau_{\mathrm{S}} = 1/\big(2f(0)\big). \tag{2.8}$$

Chapter 3 of Hettmansperger and McKean (1998), (HM) gives a discussion of the fitting and the scale parameters. We will use the estimate of τ_φ proposed by Koul, Sievers and McKean (1987) because it is consistent under both symmetrical and asymmetrical errors and we will denote this estimate by $\widehat{\tau}_\varphi$. Consistent estimates of τ_{S} are discussed on page 26 of HM.

Given the score function $\varphi(u)$, the SR estimates are based on the associated signed-rank scores that are generated as $a_\varphi^+(i) = \varphi^+(i/(n+1))$, where $\varphi^+(u) = \varphi((u+1)/2)$, $0 < u < 1$. For example, the signed-rank Wilcoxon scores are generated by $\varphi^+(u) = \sqrt{3}u$. For a given score function $\varphi(u)$, the corresponding signed-rank norm on R^n is given by,

$$\|\mathbf{v}\|_{\mathrm{SR}} = \sum_{i=1}^{n} a_\varphi^+(R|v_i|)|v_i|, \quad \text{for } \mathbf{v} \in R^n, \tag{2.9}$$

and the $R|v_i|$ denotes the rank of $|v_i|$ among $|v_1|, \ldots, |v_n|$; see page 42 of HM. The signed-rank estimate is given by

$$\widehat{\mathbf{b}}_{\mathrm{SR}} = \mathrm{Argmin}\,\|\mathbf{Y} - \mathbf{X}\mathbf{b}\|_{\mathrm{SR}}. \tag{2.10}$$

Under H_0 and mild regularity conditions (given in the Appendix), $\widehat{\mathbf{b}}_{\mathrm{SR}}$ is asymptotically

$$N_{p+1}(\mathbf{b}, \tau_\varphi^2(\mathbf{X}'\mathbf{X})^{-1}), \tag{2.11}$$

where the scale parameter τ_φ is defined in (2.7).

The estimate $\widehat{\mathbf{b}}_{\mathrm{SR}}$ is asymptotically equivalent to the the signed-rank estimate proposed by Kraft and van Eeden (1972); see, also, van Eeden (1972). To see this, differentiate the norm expression found in display (2.9) with respect to $\mathbf{b}$. The solution to the resulting normal equations was the estimate proposed by Kraft and van Eeden; see Hettmansperger and McKean (1983) for discussion.

2.2. Test statistic

To test the hypothesis of symmetric errors, we will consider the SR and R fits based on a selected score function $\varphi(u)$. The SR minimizes the normed distance between $\mathbf{Y}$ and Ω, using the norm $\|\cdot\|_{\mathrm{SR}}$, (2.9). Under symmetry, it produces a consistent estimate of $\mathbf{b}$ which we called $\widehat{\mathbf{b}}_{\mathrm{SR}}$. The predicted value of $\mathbf{Y}$ is $\widehat{\mathbf{Y}}_{\mathrm{SR}} = \mathbf{X}\widehat{\mathbf{b}}_{\mathrm{SR}}$. Then the distance between $\mathbf{Y}$ and the space Ω is given by $D_{\mathrm{SR}}(\widehat{\mathbf{Y}}_{\mathrm{SR}}) = \|\mathbf{Y} - \widehat{\mathbf{Y}}_{\mathrm{SR}}\|_{\mathrm{SR}}$. This is the distance between $\mathbf{Y}$ and the space Ω, assuming symmetry; i.e, the yardstick under symmetry.

The R fit produces a consistent estimate $\widehat{\mathbf{b}}_{\mathrm{R}}$ of $\mathbf{b}$ under both H_0 and H_A. The associated predicted value of $\mathbf{Y}$ is $\widehat{\mathbf{Y}}_{\mathrm{R}} = \mathbf{X}\widehat{\mathbf{b}}_{\mathrm{R}}$. Under symmetry $\widehat{\mathbf{Y}}_{\mathrm{R}}$ should be close to $\widehat{\mathbf{Y}}_{\mathrm{SR}}$. To measure the disparity, consider the $\|\cdot\|_{\mathrm{SR}}$-distance between $\mathbf{Y}$ and the space Ω based on this fit which we define as $D_{\mathrm{SR}}(\widehat{\mathbf{Y}}_{\mathrm{R}}) = \|\mathbf{Y} - \widehat{\mathbf{Y}}_{\mathrm{R}}\|_{\mathrm{SR}}$. Our test statistic is a standardization (see below) of the difference in distances,

$$RD = D_{\mathrm{SR}}(\widehat{\mathbf{Y}}_{\mathrm{R}}) - D_{\mathrm{SR}}(\widehat{\mathbf{Y}}_{\mathrm{SR}}). \tag{2.12}$$

Note that $RD \geq 0$ because $D_{\mathrm{SR}}(\widehat{\mathbf{Y}}_{\mathrm{SR}})$ is the minimized distance. Small values of RD indicate H_0 while large values of RD indicate H_A. Our proposed test is a goodness of fit type test. $D_{\mathrm{SR}}(\widehat{\mathbf{Y}}_{\mathrm{SR}})$ is what we expect the distance between $\mathbf{Y}$ and Ω to be under H_0, while $D_{\mathrm{SR}}(\widehat{\mathbf{Y}}_{\mathrm{R}})$ is the distance, using the norm under symmetry, between $\mathbf{Y}$ and Ω based on a fit that does not assume symmetry.

Our proposed test statistic is

$$H_{\varphi} = \frac{RD}{\widehat{\delta}}, \tag{2.13}$$

where the scale parameter δ is given by

$$\delta = \frac{\tau_{\mathrm{S}}^2 - 2\kappa\tau_{\varphi}\tau_{\mathrm{S}} + \tau_{\varphi}^2}{2\tau_{\varphi}}, \tag{2.14}$$

κ is a known constant defined by

$$\kappa = 2\int_{1/2}^{1} \varphi(u)\, du, \tag{2.15}$$

and $\widehat{\delta}$ is the estimate of δ based on the estimators $\widehat{\tau}_\varphi$ and $\widehat{\tau}_S$ defined above. It is clear from the proof of the next theorem that $0 \leq \kappa < 1$; hence, $\tau_S^2 - 2\kappa\tau_\varphi\tau_S + \tau_\varphi^2 \geq (\tau_S - \tau_\varphi)^2$. Thus both δ and its estimate are always nonnegative.

The asymptotic decision rule, based on the theorem below, is to reject H_0 if $H_\varphi \geq \chi_1^2(\alpha)$ where $\chi_1^2(\alpha)$ is the upper α-critical value of a χ^2 random variable with one degree of freedom.

The proof of the following theorem can be found in the Appendix.

Theorem 2.1. *Under H_0 and the regularity conditions listed in the Appendix, $H_\varphi \xrightarrow{\mathcal{D}} \chi_1^2$ random variable.* ·

3. Score selection

In this section, we discuss appropriate score functions to use for H_φ. First, consider a univariate sample, $w_1, w_2, \ldots, w_n$. The classical test for symmetry is based on the statistic,

$$b_1 = \frac{\frac{1}{n}\sum_{i=1}^n (w_i - \overline{w})^3}{\left[\frac{1}{n}\sum_{i=1}^n (w_i - \overline{w})^2\right]^{3/2}}; \tag{3.1}$$

see Chapter 7 of D'Agostino and Stephens (1986). This test is certainly not robust, but it does suggest using a cubic rank score function. Further, if we assume that w_i follows a normal distribution, then the asymptotically most powerful rank test for a sequence of local skewness alternatives is based on the third Hermite polynomial which is a cubic in $\Phi^{-1}(u)$, the inverse of the standard normal distribution function; see Eubank, LaRiccia and Rosenstein (1987). In another paper, Eubank, LaRiccia and Rosenstein (1992) also considered the third Legendre polynomial which is a cubic in u to form a test statistic based on residuals from a location estimate. The Hermite and Legendre third degree polynomials, though, are not monotone functions, a condition required for $\varphi(u)$ as discussed in Section 2.1. Hence, the corresponding function defined as in (2.9) for these scores functions will not be a norm and the dispersion function will not be convex. This jeopardizes the computation of the regression coefficients as well as the asymptotic theory for the rank and signed-rank procedures.

Instead, we will consider the following simple cubic, which overcomes the problems described in the previous paragraph,

$$\varphi_C(u) = 8\sqrt{7}(u - \tfrac{1}{2})^3, \tag{3.2}$$

for rank regression scores and $\varphi_C^+(u) = \varphi_C((u+1)/2)$ for the associated signed rank scores. Let τ_C denote the scale parameter (2.7) for these cubic scores. The score function $\varphi_C(u)$ is monotone and bounded; hence, the theory outlined in Section 2 holds for these score functions. Further, the regression fits are easily computed as well as the Koul et al. (1987) estimate of τ_C; see Chapter 3 of HM. The test can be computed at the web site `http://www.stat.wmich.edu/mckean/Symm/testofsymm.html`. For $\varphi_C(u)$, $\kappa = \sqrt{7}/4$. Let δ_C denote the corresponding parameter given by (2.14). Our test statistic is given by

$$H_C = \frac{RD}{\widehat{\delta}_C}. \tag{3.3}$$

An asymptotic level α test is to reject H_0 if $H_C \geq \chi_1^2(\alpha)$, where $\chi_1^2(\alpha)$ denotes the upper α critical point of a χ^2 random variable with one degree of freedom.

4. A comparison with other tests for symmetry, in particular, the mean minus the median

In a companion paper, Hettmansperger, McKean and Sheather (2002) compared the Monte Carlo performance of the test described in the previous section with a number of procedures including the univariate procedures of Boos (1982), Gastwirth (1971), Eubank, LaRiccia and Rosenstein (1992) and a test based on the difference between the mean and the median (which we shall describe below). All of these procedures are tests for the univariate symmetry problem. For our setting, the linear model problem, we are interested in their behavior on residuals. Thus, these procedures depend on the LS and the Wilcoxon residuals.

We will denote the LS residuals by

$$r_{\mathrm{LS}i} = Y_i - \mathbf{x}_{ci}'\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}, \tag{4.1}$$

where $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}} = (\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_c'\mathbf{Y}$ and $\mathbf{x}_{ci}'$ is the ith row of $\mathbf{X}_c$. We will denote the Wilcoxon residuals by

$$r_{\mathrm{RW}i} = Y_i - \mathbf{x}_{ci}'\widehat{\boldsymbol{\beta}}_{\mathrm{RW}}, \tag{4.2}$$

where $\widehat{\boldsymbol{\beta}}_{\mathrm{RW}}$ is the R estimate, (2.3), when the Wilcoxon score function is used. Note that these are *not* the signed-rank Wilcoxon residuals. Also, for Wilcoxon scores we will denote the scale parameter τ, (2.7), by τ_W. Note that for Wilcoxon scores, it simplifies to

$$\tau_W = 1\Big/\left(\sqrt{12}\int f^2(x)\,dx\right). \tag{4.3}$$

Below we describe the two procedures which had the best Monte Carlo performance in terms of both level and power.

1. **H_φ Procedures**. Besides the test statistic H_{C}, (3.3), based on the cubic score function, (3.2), Hettmansperger et al. (2002) investigated the behavior of the test statistic H_φ for two other score functions: the Wilcoxon, $\varphi(u) = \sqrt{12}(u - \frac{1}{2})$ and the normal score function $\varphi(u) = \Phi^{-1}(u)$, where $\Phi(u)$ is the cdf for a standard normal random variable. In their Monte Carlo study, two choices of critical values were investigated: the $\chi^2(1)$, (χ^2-distribution with 1 degree of freedom), critical values as suggested by the asymptotic theory and the $F(1, n-p)$, (F-distribution with 1 and $n-p$ degrees of freedom), critical values, a standard degree of freedom correction. This leads to 6 different test procedures which we label as: the Wilcoxons, $H_{\mathrm{W}\chi^2}$ and $H_{\mathrm{W}F}$; the normals, $H_{\mathrm{N}\chi^2}$ and $H_{\mathrm{N}F}$; and the cubics, $H_{\mathrm{C}\chi^2}$ and $H_{\mathrm{C}F}$. The best performing procedure was found to be $H_{\mathrm{C}\chi^2}$.

2. **MM: The Mean minus the Median**. The mean minus the median is one of the simplest measures of symmetry that is discussed in practically every course in statistics. Its analogue for the regression problem is to compare estimates of the intercept parameter. We chose estimates based on fits which do not assume symmetry. Our analogues of the median and mean are the median of the Wilcoxon residuals, (4.2), ($\widehat{\alpha}_{\mathrm{W}} = \mathrm{med}\{r_{\mathrm{RW}i}\}$), and the mean of the least squares residuals, (4.1), ($\widehat{\alpha}_{\mathrm{LS}} = \bar{r}_{\mathrm{LS}}$). Under symmetrical errors e_i (see Hettmansperger and McKean, 1998, p. 166), we have the following asymptotic representations of these estimates:

$$
\begin{aligned}
\widehat{\alpha}_{\mathrm{W}} &= \frac{1}{n}\tau_{\mathrm{S}} \sum_{i=1}^{n} \mathrm{sgn}(e_i) + o_p(n^{-1/2}), \\
\widehat{\alpha}_{\mathrm{LS}} &= \frac{1}{n} \sum_{i=1}^{n} e_i + o_p(n^{-1/2}),
\end{aligned}
\tag{4.4}
$$

where τ_{S} is given by expression (2.8) and σ^2 is the variance of the error distribution. Based on these representations, it can be shown that the asymptotic variance of $\widehat{\alpha}_{\mathrm{LS}} - \widehat{\alpha}_{\mathrm{W}}$ is

$$
V_{\mathrm{MM}} = \mathrm{Var}(\widehat{\alpha}_{\mathrm{LS}} - \widehat{\alpha}_{\mathrm{W}}) \doteq \frac{\tau_{\mathrm{S}}^2}{n} + \frac{\sigma^2}{n} - 2\frac{\tau_{\mathrm{S}}}{n}\,\mathrm{E}(|e_1|). \tag{4.5}
$$

Hettmansperger et al. (2002) estimated τ_{S} based on the Wilcoxon residuals as discussed in Section 2, σ^2 by MSE of the least squares residuals; and $\mathrm{E}(|e_1|)$ by $n^{-1}\sum_{i=1}^{n} |\widehat{e}_{\mathrm{RW}i}|$). Let $\widehat{V}_{\mathrm{MM}}$ denote the resulting estimate of $\mathrm{Var}(\widehat{\alpha}_{\mathrm{LS}} - \widehat{\alpha}_{\mathrm{W}})$. Given a level α, the MM-testing procedure is: reject the hypothesis of symmetry if $|z_{\mathrm{MM}}| > z_{\alpha/2}$ where

$$
z_{\mathrm{MM}} = (\widehat{\alpha}_{\mathrm{LS}} - \widehat{\alpha}_{\mathrm{W}})/\sqrt{\widehat{V}_{\mathrm{MM}}}, \tag{4.6}
$$

and $z_{\alpha/2}$ is the upper $\alpha/2$ standard normal quantile.

5. Behavior under contiguous alternatives

We consider the asymptotic distribution of the test statistic H_φ, (2.13), under sequences of local, asymmetric alternatives. A general class of such local alternatives was considered by Eubank et al. (1992) for the univariate case. In our notation, consider a sequence of linear models, (2.1) indexed by the sample size n. Assume that the errors $e_{n1}, \dots, e_{nn}$ are iid with common cdf,

$$F_n(t) = F_0(t) + \frac{\lambda}{\sqrt{n}} G(t) + O(1/n), \tag{5.1}$$

where $F_0(t)$ is a symmetric distribution function, $G(t) = H(t) - F_0(t)$, $H(t)$ is a distribution function, and the $O(1/n)$ term is uniform in t. As indicated by Eubank et al. (1992), many of the local alternatives that have appeared in the literature can be formulated as in (5.1). This sequence of alternatives is contiguous to the symmetric cdf $F_0(t)$. We shall obtain the noncentrality parameter of our test statistic H_φ under this sequence of alternatives.

Theorem 5.1. *Under F_n given in* (5.1) *and the regularity conditions in the appendix,*

$$H_\varphi \xrightarrow{\mathcal{D}_{F_n}} \chi^2(1, \lambda^2 \mu_\varphi^2), \tag{5.2}$$

where

$$\mu_\varphi = \sqrt{\frac{\tau_S^2}{\tau_S^2 - 2\kappa\tau_S\tau_\varphi + \tau_\varphi^2}} \left\{ 1 - 2H(0) - \frac{\tau_\varphi}{\tau_S} \int \varphi(F_0(e))\, dH(e) \right\}. \tag{5.3}$$

Proof. Replacing $\widehat{\delta}$ with δ, we can use (A.14) of the Appendix to obtain the following asymptotic representation of the test statistic H_φ:

$$\begin{aligned} H_\varphi &\doteq \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\operatorname{sgn}(e_i) - (\tau_\varphi/\tau_S)\varphi(F(e_i))}{\sqrt{(\tau_S^2 - 2\kappa\tau_S\tau_\varphi + \tau_\varphi^2)/\tau_S^2}} \right\}^2 \\ &\stackrel{\text{Def}}{=} S_n^2, \end{aligned} \tag{5.4}$$

where κ is given in (2.15) and

$$\frac{\tau_S^2 - 2\kappa\tau_S\tau_\varphi + \tau_\varphi^2}{\tau_S^2} = \operatorname{Var}\left\{ \operatorname{sgn}(e_i) - \frac{\tau_\varphi}{\tau_S}\varphi(F(e_i)) \right\}.$$

Under symmetry, $\lambda = 0$, it follows from Theorem 2.1 that $S_n \xrightarrow{\mathcal{D}} \mathrm{N}(0,1)$.

The density of $F_n(x)$ is $f_n(x) = f_0(x) + \frac{\lambda}{\sqrt{n}} g(x)$. Hence the log of the ratio of the likelihood functions under H_n and H_0 is given by,

$$l_n = \sum_{i=1}^{n} \log \frac{f_n(x_i)}{f_0(x_i)} = \frac{\lambda}{\sqrt{n}} \sum_{i=1}^{n} \frac{g(x_i)}{f_0(x_i)} + o_p(1). \tag{5.5}$$

Under $F_n(t)$, $S_n \xrightarrow{\mathcal{D}} \mathrm{N}(\lambda\mu_\varphi, 1)$, where by LeCam's Third Lemma,

$$\begin{aligned}\mu_\varphi &= \lim \mathrm{Cov}_{F_0}(S_n, l_n) \\ &= \sqrt{\frac{\tau_S^2}{\tau_S^2 - 2\kappa\tau_S\tau_\varphi + \tau_\varphi^2}} \,\mathrm{E}_{F_0}\left\{ \frac{g(e)}{f_0(e)} \left[\mathrm{sgn}(e) - \frac{\tau_\varphi}{\tau_S}\varphi(F(e)) \right] \right\}.\end{aligned}$$

Since $g(e) = h(e) - f_0(e)$ and $\int \mathrm{sgn}(x)\, dF_0(x) = 0 = \int \varphi(F_0(e))\, dF_0(e)$, μ_φ simplifies to (5.3). □

5.1. Comparison with the MM procedure

Because of their comparable Monte Carlo performances in the study of Hettmansperger et al. (2002), it is of interest to determine the asymptotic relative efficiency between the $H_{c\chi^2}$ test statistic and the mean minus median (MM) test statistic, for the sequence of alternatives given by (5.1).

The noncentrality parameter for the test statistic $H_{c\chi^2}$ is given by $\lambda^2\mu_\varphi^2$, where μ_φ is given by expression (5.3) and $\varphi(u)$ is the score function given by (3.2). For this cubic score, denote μ_φ by μ_{C}.

To obtain the noncentrality parameter for the MM test, recall that we can represent the difference of the mean and median under the null hypothesis by the expression,

$$\sqrt{n}(\bar{e} - \mathrm{med}\, e_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(e_i - \tau_S\, \mathrm{sgn}(e_i)\big) + o_p(1). \tag{5.6}$$

Using the log of the ratio of the likelihoods functions under H_n and H_0, (5.5), we can apply LeCam's Third Lemma to show that z_{MM}^2 converges in distribution under F_n to a $\chi^2(1, \lambda^2\mu_{\mathrm{MM}}^2)$ distribution, where

$$\mu_{\mathrm{MM}} = \frac{E_H(e) - \tau_s\big(1 - 2H(0)\big)}{\sqrt{\mathrm{Var}_H(e) - 2\tau_s\, \mathrm{E}_H\, |e| + \tau_s^2}}. \tag{5.7}$$

The ARE between $H_{c\chi^2}$ and Z_{MM}^2 is the ratio $\mu_{\mathrm{C}}^2/\mu_{\mathrm{MM}}^2$. Large values favor $H_{c\chi^2}$.

Because the noncentrality parameter μ_{MM} contains the mean $\mathrm{E}_H\,|e|$ and variance $\mathrm{Var}_H(e)$ of the errors, it is not robust. As the following example shows, it is easy to obtain a family of asymmetric distributions where μ_{MM} is essentially 0.

Table 1. AREs between $H_{C\chi^2}$ and z^2_{MM} Test Statistics at the Distribution (5.8) with $\epsilon = .2$, $\eta = .95$, and $\mu_R = 1$.

μ_L	Mean(H)	Median(H)	μ^2_{MM}	$\mu^2_{C\chi^2}$	ARE
−2	.17	.1635	6.58E−4	5.92E−4	.90
−3	.16	.16395	1.70E−4	3.66E−4	2.12
−4	.15	.1629	2.30E−9	3.00E−4	1.30E+5

Example. Consider the contaminated normal distribution function with contamination in both directions:

$$H(t) = (1-\epsilon)\Phi(t) + \epsilon\eta\Phi(t-\mu_R) + \epsilon(1-\eta)\Phi(t-\mu_L); \tag{5.8}$$

where $\Phi(t)$ is the cdf of a standard normal random variable. If η is close to 1 then the majority of the contamination is on the right side. For various values of the parameters, the noncentrality parameter μ_{MM} of the MM procedure is 0. Table 1 displays such a situation for $\epsilon = .2$, $\eta = .95$, and $\mu_R = 1$. Here, μ_{MM} is essentially 0, if $\mu_L = -4$. Note that the contamination in this case drove the mean of H to the left of its median.

6. Conclusion

The proposed test statistic H_φ offers the user a robust and asymptotically distribution free test for symmetry of the error distribution in a linear model. It is a goodness-of-fit type test and is easily interpretable. We recommend the statistic $H_{C\chi^2}$ based on the simple cubic score function (3.2). Part of the theory behind the test is based on the fundamental work of Constance van Eeden on signed-rank procedures for the linear model.

Acknowledgements. We would like to thank the referees and the editor for their remarks on the paper which improved its exposition.

APPENDIX

Regularity conditions

Assume the following conditions, (see Chapter 3 of HM for discussion). On the density f, assume that

$$f \text{ is absolutely continuous, } 0 < I(f) < \infty. \tag{R.1}$$

On the design matrix $\mathbf{X}$, assume that

(R.2) $$\lim_{n\to\infty} \max_{1\le i\le n} h_{iin} = 0,$$

(R.3) $$\lim_{n\to\infty} n^{-1}\mathbf{X}'\mathbf{X} = \mathbf{\Sigma},$$

where h_{iin} denotes the ith diagonal entry of the projection matrix onto the column space of $\mathbf{X}$ and $\mathbf{\Sigma}$ is a $(p+1)\times(p+1)$ positive definite matrix.

Proof of Theorem 2.1. Without loss of generality assume that the true vector of parameters $\mathbf{b} = \mathbf{0}$. Write $D_{\mathrm{R}} = D_{\mathrm{SR}}(\widehat{\mathbf{Y}}_{\mathrm{R}})$, $D_{\mathrm{SR}} = D_{\mathrm{SR}}(\widehat{\mathbf{Y}}_{\mathrm{SR}})$. In this notation, RD is given by $RD = D(\widehat{\mathbf{b}}_{\mathrm{R}}) - D(\widehat{\mathbf{b}}_{\mathrm{SR}})$. The function D is a convex function and can be approximated, asymptotically by a quadratic function; see Hettmansperger and McKean (1983). Let a^* be

(A.1) $$a^*(R|e_i|) = a^+(R|e_i|)\,\mathrm{sgn}(e_i),$$

and denote the $n\times 1$ vector $(a^*(R|e_1|),\ldots,a^*(R|e_n|))'$ by $\mathbf{a}^*(R|\mathbf{e}|)$. We will use similar vector notations for the ranks, signs, and score functions, i.e., $\mathbf{a}(R(\mathbf{e}))$, $\mathbf{sgn}(\mathbf{e})$, and $\boldsymbol{\varphi}(F(\mathbf{e}))$. Consider the quadratic function,

(A.2) $$Q(\mathbf{b}) = (2\tau_\varphi)^{-1}\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} - \mathbf{b}'\mathbf{X}'\mathbf{a}^*(R|\mathbf{e}|) + D(\mathbf{e}).$$

This quadratic function approximates D in that $D(\mathbf{b}) - Q(\mathbf{b}) = o_p(1)$ for $\sqrt{n}\mathbf{b} = O_p(1)$.

Then write,

(A.3) $$RD = \left[D(\widehat{\mathbf{b}}_{\mathrm{R}}) - Q(\widehat{\mathbf{b}}_{\mathrm{R}})\right] + \left[Q(\widehat{\mathbf{b}}_{\mathrm{R}}) - Q(\widehat{\mathbf{b}}_{\mathrm{SR}})\right] + \left[Q(\widehat{\mathbf{b}}_{\mathrm{SR}}) - D(\widehat{\mathbf{b}}_{\mathrm{SR}})\right].$$

Based on the quadratic approximation, (see Hettmansperger and McKean, 1983), the first and third bracketed terms of (A.3) go to 0 in probability.

Before evaluating the middle term, we will state some useful asymptotic representations. Basically we will have two sets of representations: a general representation and a representation when the errors have a symmetric distribution. For our proof we will use the more convenient second set but the first set will be convenient for our discussion on local alternatives. First consider the quadratic function. Its definition above, (A.2), is the general result. Next, using the connection between $\varphi(u)$ and $\varphi^+(u)$, ($\varphi^+(u) = \varphi((u+1)/2)$), and substituting the distribution function for the empirical distribution function, results in

(A.4) $$\frac{1}{\sqrt{n}}\mathbf{X}'\mathbf{a}^*(R|\mathbf{e}|) = \frac{1}{\sqrt{n}}\mathbf{X}'\boldsymbol{\varphi}(F(\mathbf{e})) + o_p(1).$$

We can then write the quadratic function, (A.2) as

(A.5) $$Q(\mathbf{b}) = (2\tau_\varphi)^{-1}\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} - \mathbf{b}'\mathbf{X}'\boldsymbol{\varphi}(F(\mathbf{e})) + D(\mathbf{e}).$$

The quadratic approximation holds for this quadratic function also.

As discussed in HM (Theorem 3.5.11), asymptotic representations for $\widehat{\mathbf{b}}_{\mathrm{R}}$ are given by:

$$\widehat{\mathbf{b}}_{\mathrm{R}} = \begin{bmatrix} \tau_{\mathrm{S}}\frac{1}{n}\mathbf{1}'\,\mathbf{sgn}(\mathbf{e}) \\ \tau_{\varphi}(\mathbf{X}_{\mathrm{c}}'\mathbf{X}_{\mathrm{c}})^{-1}\mathbf{X}_{\mathrm{c}}'\mathbf{a}(R(\mathbf{e}))\end{bmatrix} + O_p\left(\frac{1}{\sqrt{n}}\right) \tag{A.6}$$

$$= \begin{bmatrix} \tau_{\mathrm{S}}\frac{1}{n}\mathbf{1}'\,\mathbf{sgn}(\mathbf{e}) \\ \tau_{\varphi}(\mathbf{X}_{\mathrm{c}}'\mathbf{X}_{\mathrm{c}})^{-1}\mathbf{X}_{\mathrm{c}}'\varphi(F(\mathbf{e}))\end{bmatrix} + O_p\left(\frac{1}{\sqrt{n}}\right). \tag{A.7}$$

Again, the second representation is useful in the proof while the first will be useful in the contiguous section. Algebra shows that if $\sqrt{n}(\mathbf{b}_1 - \mathbf{b}_2) = O_p(1)$, then $Q(\mathbf{b}_1) - Q(\mathbf{b}_2) = o_p(1)$. Hence, we can use the vector on the right in (A.7) for the evaluation in (A.3).

Next consider the signed-rank estimate $\widehat{\mathbf{b}}_{\mathrm{SR}}$. Similar to the asymptotic representations of the rank estimate,

$$\widehat{\mathbf{b}}_{\mathrm{SR}} = \tau_{\varphi}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{a}^*(R|\mathbf{e}|) + O_p\left(\frac{1}{\sqrt{n}}\right) \tag{A.8}$$

$$= \tau_{\varphi}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varphi(F(\mathbf{e})) + O_p\left(\frac{1}{\sqrt{n}}\right); \tag{A.9}$$

see Hettmansperger and McKean (1983).

Using these approximations, we will now evaluate the middle difference. We will first evaluate it using the first quadratic (A.2) and the set of first representations for the estimates. This results in

$$Q(\widehat{\mathbf{b}}_{\mathrm{SR}}) = -\frac{\tau_{\varphi}}{2}\mathbf{a}^*(R|\mathbf{e}|)'\mathbf{H}\mathbf{a}^*(R|\mathbf{e}|) + D(\mathbf{e}), \tag{A.10}$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Using the second set of asymptotic representations which hold under symmetric errors we have,

$$Q(\widehat{\mathbf{b}}_{\mathrm{SR}}) = -\frac{\tau_{\varphi}}{2}\varphi(F(\mathbf{e}))'\mathbf{H}_1\varphi(F(\mathbf{e})) - \frac{\tau_{\varphi}}{2}\varphi(F(\mathbf{e}))'\mathbf{H}_{\mathrm{c}}\varphi(F(\mathbf{e})) + D(\mathbf{e}), \tag{A.11}$$

where the projection matrices are $\mathbf{H}_1 = n^{-1}\mathbf{1}\mathbf{1}'$ and $\mathbf{H}_{\mathrm{c}} = \mathbf{X}_{\mathrm{c}}(\mathbf{X}_{\mathrm{c}}'\mathbf{X})^{-1}\mathbf{X}_{\mathrm{c}}'$.

Likewise under the first set of representations,

$$Q(\widehat{\mathbf{b}}_{\mathrm{R}}) = \frac{1}{2\tau_{\varphi}}\left[\tau_{\mathrm{S}}^2\,\mathbf{sgn}(\mathbf{e})'\mathbf{H}_1\,\mathbf{sgn}(\mathbf{e}) + \tau_{\varphi}^2\mathbf{a}(R(\mathbf{e}))'\mathbf{H}_{\mathrm{c}}\mathbf{a}(R(\mathbf{e}))\right] - \tau_{\mathrm{S}}\,\mathbf{sgn}(\mathbf{e})'\mathbf{H}_1\mathbf{a}^*(R(|\mathbf{e}|) - \tau_{\varphi}\mathbf{a}(R(\mathbf{e}))'\mathbf{H}_{\mathrm{c}}\mathbf{a}^*\left(R(|\mathbf{e}|)\right) + D(\mathbf{e}). \tag{A.12}$$

While under the second set of representations we get

$$Q(\widehat{\mathbf{b}}_{\mathrm{R}}) = \frac{\tau_{\mathrm{S}}^2}{2\tau_{\varphi}}\,\mathbf{sgn}(\mathbf{e})'\mathbf{H}_1\,\mathbf{sgn}(\mathbf{e}) + \frac{\tau_{\varphi}}{2}\varphi(F(\mathbf{e}))'\mathbf{H}_{\mathrm{c}}\varphi(F(\mathbf{e})) - \tau_{\mathrm{S}}\,\mathbf{sgn}(\mathbf{e})'\mathbf{H}_1\varphi(F(\mathbf{e})) - \tau_{\varphi}\varphi(F(\mathbf{e}))'\mathbf{H}_{\mathrm{c}}\varphi(F(\mathbf{e})) + D(\mathbf{e}). \tag{A.13}$$

Using (A.11) and (A.13), after simplification, we get

$$RD = \frac{\tau_S^2}{2\tau_\varphi}\left[\mathbf{sgn}(\mathbf{e}) - \frac{\tau_\varphi}{\tau_S}\boldsymbol{\varphi}\big(F(\mathbf{e})\big)\right]' \mathbf{H}_1 \left[\mathbf{sgn}(\mathbf{e}) - \frac{\tau_\varphi}{\tau_S}\boldsymbol{\varphi}\big(F(\mathbf{e})\big)\right]. \tag{A.14}$$

Let $\mathcal{D}$ denote the variance-covariance operator. Under symmetry,

$$E\left[\mathbf{sgn}(\mathbf{e}) - \frac{\tau_\varphi}{\tau_S}\boldsymbol{\varphi}\big(F(\mathbf{e})\big)\right] = \mathbf{0}, \tag{A.15}$$

and $\mathcal{D}\big(\mathbf{sgn}(\mathbf{e})\big) = \mathbf{I}$. Because the scores are standardized, $\mathcal{D}\left(\boldsymbol{\varphi}\big(F(\mathbf{e})\big)\right) = \mathbf{I}$. Finally, the matrix $E\big[\,\mathbf{sgn}(\mathbf{e})\boldsymbol{\varphi}\big(F(\mathbf{e})\big)'\big]$ has 0 entries off the main diagonal while on the main diagonal it has entries

$$\begin{aligned} \kappa &= E[\operatorname{sgn}(e_1)\varphi(F(e_1))] \\ &= E[\operatorname{sgn}(2F(e_1) - 1)\varphi(F(e_1))] \\ &= \int_0^1 \operatorname{sgn}(2u - 1)\varphi(u)\,du \\ &= 2\int_{\frac{1}{2}}^1 \varphi(u)\,du. \end{aligned} \tag{A.16}$$

Hence,

$$\mathcal{D}\left[\mathbf{sgn}(\mathbf{e}) - \frac{\tau_\varphi}{\tau_S}\boldsymbol{\varphi}\big(F(\mathbf{e})\big)\right] = \left[1 + \frac{\tau_\varphi^2}{\tau_S^2} - 2\frac{\tau_\varphi}{\tau_S}\kappa\right]\mathbf{I}. \tag{A.17}$$

The asymptotic normality of $\mathbf{sgn}(\mathbf{e}) - (\tau_\varphi/\tau_S)\boldsymbol{\varphi}\big(F(\mathbf{e})\big)$ can be established by standard Lindeberg Central Limit Theorem arguments; see p. 167 of HM. Because the rank of $\mathbf{H}_1$ is 1, the result follows. □

REFERENCES

Boos, D.D. (1982). A test for asymmetry associated with the Hodges-Lehmann estimator. *J. Amer. Statist. Assoc.* 77, 647–651.

D'Agostino, R.B. and Stephens, M.A. (eds.) (1986). *Goodness-of-Fit Techniques.* Dekker, New York.

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

Eubank, R.L., LaRiccia, V. and Rosenstein, R.B. (1992). Testing symmetry about an unknown median, via linear rank procedures. *J. Nonparametr. Statist.* 1, 301–311.

Gastwirth, J.L. (1971). On the sign test for symmetry. *J. Amer. Statist. Assoc.* 66, 821–823.

Hettmansperger, T.P. and McKean, J.W. (1983). A geometric interpretation of inferences based on ranks in the linear model. *J. Amer. Statist. Assoc.* 78, 885–893.

Hettmansperger, T.P. and McKean, J.W. (1998). *Robust Nonparametric Statistical Methods.* Arnold, London.

Hettmansperger, T.P., McKean, J.W., and Sheather, S.J. (2002). Finite sample performance of tests for symmetry of the errors in a linear model. *J. Stat. Comput. Simul.* 72, 863–879.

Koul, H.L., Sievers, G.L. and McKean, J.W. (1987). An estimator of the scale parameter for the rank analysis of linear models under general score functions. *Scand. J. Statist.* 14, 131–141.

Kraft, C.H. and van Eeden, C. (1972). Linearized rank estimates and signed-rank estimates for the general linear hypothesis. *Ann. Math. Statist.* 43, 42–57.

van Eeden, C. (1972). An analogue, for signed rank statistics, for Jurečková's asymptotic linearity theorem for rank statistics. *Ann. Math. Statist.* 43, 791–802.

Thomas P. Hettmansperger
Department of Statistics
Penn State University
326 Thomas Building
University Park, PA, 16802-2111
USA
tph@stat.psu.edu

Joseph W. McKean
Department of Mathematics and Statistics
Western Michigan University
Kalamazoo, MI 49008
USA
joe@stat.wmich.edu

Simon J. Sheather
Australian Graduate School of Management
The University of New South Wales
Sydney NSW 2052
Australia
simonsh@agsm.edu.au