

REVIEW

Open Access

Efficient and adaptive rank-based fits for linear models with skew-normal errors

Joseph W McKean^{1*} and John D Kloke²

*Correspondence:

mckean@wmich.edu

¹Department of Statistics, Western Michigan University, Kalamazoo, USA

Full list of author information is available at the end of the article

Abstract

The rank-based fit of a linear model is based on minimizing a norm. A score function needs to be selected for the fit and the proper choice leads to asymptotically efficient regression estimators, i.e., fits equivalent to the maximum likelihood estimators (mle). In this paper, we present the family of optimal scores functions for the skew-normal family of distributions. We show the easy computation of this rank-based fit using the R package Rfit. We present the results of a small simulation study comparing the rank-based estimators and the mles in terms of efficiency and validity over skew-normal and contaminated normal distributions. We also develop and present empirical results for a Hogg-type adaptive procedure for selecting among a family of these scores based on a robust initial fit.

Keywords: Linear models; Monte Carlo; Nonparametrics; Regression rank scores; Robust; Wilcoxon procedures

1 Introduction

Rank-based fitting of linear models offers an attractive alternative to least squares (LS) and maximum likelihood (mle) fitting. The geometry of the rank-based fit is similar to that of LS. Simply replace the Euclidean squared norm used in the LS fit with another norm which, unlike LS, results in a robust fit. The accompanying robust analysis is the analogue of the LS's ANOVA and ANCOVA. The rank-based analysis offers a complete analysis including robust diagnostics to check quality of fit. This rank based analysis has recently been extended to mixed and nonlinear models; see Kloke et al. (2009) and Abebe and McKean (2013), respectively. A full development of the rank-based analysis can be found in Chapters 3-5 of the monograph by Hettmansperger and McKean (2011). The rank-based analysis is, generally, highly efficient. It can easily be optimized depending on the information available concerning the distribution of the random errors. For example, if the form of the error distribution is known, then an appropriate rank-based procedure can be selected to attain full efficiency.

In this paper, we discuss rank-based analyses which are appropriate for the skew-normal (SN) family of distributions. This is a rich family of skewed distributions developed by Azzalini (1985). The skewness of a distribution in this family is controlled by a shape parameter α , $-\infty < \alpha < \infty$. Distributions are left or right skewed, depending on whether $\alpha < 0$ or $\alpha > 0$, respectively. If $\alpha = 0$ then the distribution is normal. As we discuss in Section 3, all of these distributions are light-tailed. Such families of distributions occur

frequently in accelerated failure time (AFT) models. The response of interest in such models is survival time and its log is frequently modeled in terms of a log linear model. The random errors for these log linear models are, generally, skewed. For instance, the family of distributions for log linear models when survival time follows an F -distribution contains a wide variety of skewed distributions with tail weights that range from moderate to heavy; see McKean and Sievers (1989). Hence, the skew-normal family adds a rich class of light-tailed skewed distributions which also includes the normal distribution.

In Section 2, we outline the rank-based analysis for a general linear model. The computation of these analyses is easily handled by the R package `Rfit`, developed by Kloke and McKean (2013), which can be freely downloaded at the CRAN <http://cran.us.r-project.org/>. We discuss this software for an example in this section and we continue such discussion in the remainder of the article. The data for the example and the R code, supplemental to `Rfit`, used in this article are available to the reader at the <http://www.stat.wmich.edu/mckean/SN/>.

In Section 3, we develop rank-based analyses for the skew-normal family. These analyses are efficient for this family and the appropriate analysis is fully efficient. As we discuss, these analyses are technically robust similar to the optimal rank-based analysis for normally distributed errors. In contrast, in a sensitivity analysis, we show that the maximum likelihood fit (mle) is not robust. In Section 4, we present the results of a Monte Carlo study which verify the robustness and validity of the rank-based analysis over a family of SN distributions and contaminated SN distributions. These studies confirm the nonrobustness of the mle analysis.

The rank-based analysis depends on the shape parameter α . One outcome of the Monte Carlo study is that rank-based analyses based on shape parameters in a neighborhood of the correct α had very similar behavior to that using the correct α . This suggests that a simple Hogg-type adaptive procedure would entertain excellent properties in this situation. In Section 5, we develop such an adaptive scheme for the family of SN distributions. In a simulation study, we verify the efficiency and validity of this scheme over SN situations and, further, over two contaminated situations.

2 Notation and rank-based analysis

Let Y be a $n \times 1$ vector of responses which follows the linear model given by

$$Y = \mathbf{1}_n \beta_0 + X\boldsymbol{\beta} + \mathbf{e}. \quad (1)$$

where $\mathbf{1}_n$ is a vector of n ones; X is a $n \times p$ design matrix which may contain predictors (covariates) as well as indicator (dummy) variables; β_0 is an intercept parameter; $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression parameters; and \mathbf{e} is a $n \times 1$ vector of random errors. Because we have an intercept parameter in the model, we assume without loss of generality that the design matrix X is centered, (all columns of X have mean 0). For the theory discussed below, assume that the components of \mathbf{e} are iid with pdf $f(x)$ and cdf $F(x)$, where $F(x)$ is unknown.

The least squares (LS) estimator of $\boldsymbol{\beta}$ is the vector $\widehat{\boldsymbol{\beta}}_{LS}$ which minimizes the Euclidean distance between Y and the column space of X ; that is, it satisfies

$$\widehat{\boldsymbol{\beta}}_{LS} = \text{Argmin} \|Y - X\boldsymbol{\beta}\|_2^2, \quad (2)$$

where $\|\mathbf{v}\|_2^2 = \sum_{i=1}^n v_i^2$ is the Euclidean norm.

For the rank-based estimates, simply replace the Euclidean norm $\|\cdot\|_2$ by the pseudo-norm

$$\|\mathbf{v}\|_\varphi = \sum_{i=1}^n a[R(v_i)] v_i, \quad \mathbf{v} \in R^n, \quad (3)$$

where $R(v_i)$ is the rank of v_i among v_1, \dots, v_n and the scores $a(i)$ are generated as $a[i] = \varphi[i/(n+1)]$, for a nondecreasing bounded square-integrable function $\varphi(u)$, satisfying, without loss of generality, the standardizing conditions $\int \varphi(u) du = 0$ and $\int \varphi^2(u) du = 1$. Then the rank-based estimator minimizes the $\|\cdot\|_\varphi$ -distance between \mathbf{Y} and the column space of \mathbf{X} ; i.e.,

$$\hat{\boldsymbol{\beta}}_\varphi = \text{Argmin} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_\varphi. \quad (4)$$

These estimators were proposed by Jaeckel (1972) and Jurečková (1971). An associated rank-based analysis, including diagnostics procedures, is discussed in Chapters 3-5 of the monograph by Hettmansperger and McKean (2011). A score function needs to be selected. Often the Wilcoxon (linear) score function is used, $\varphi[u] = \sqrt{12}[u - (1/2)]$. When Wilcoxon scores are used, we refer to the subsequent fit and analysis as the Wilcoxon analysis. Another frequent choice is the sign scores function, $\varphi[u] = \text{sgn}[u - (1/2)]$, which yields the l_1 -fit. Score functions are discussed in terms of optimality in Section 2.2.

As with LS, the rank-based estimator of the intercept is a location estimate based on the residuals. For LS, the arithmetic mean is used while for the rank-based estimates, generally, the median is used; i.e.,

$$\hat{\beta}_0 = \text{medi} \{ Y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_\varphi \}. \quad (5)$$

2.1 Theory

As shown in Hettmansperger and McKean (2011), the influence function of the rank-based estimator $\hat{\boldsymbol{\beta}}_\varphi$ is given by

$$\Omega(\mathbf{x}_0, y_0; \hat{\boldsymbol{\beta}}_\varphi) = \tau_\varphi (\mathbf{X}'\mathbf{X}/n)^{-1} \varphi[F(y_0)] \mathbf{x}_0, \quad (6)$$

where the point (y_0, \mathbf{x}_0^T) represents an outlier. The parameter τ_φ is the scale parameter given by

$$\tau_\varphi^{-1} = \int \varphi(u) \varphi_f(u) du, \quad (7)$$

where

$$\varphi_f(u) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}. \quad (8)$$

Based on this influence function it is clear that $\hat{\boldsymbol{\beta}}_\varphi$ is robust in Y -space if the scores function $\varphi(u)$ is bounded. Note, though, that $\hat{\boldsymbol{\beta}}_\varphi$ is not robust in X -space. A weighted version of the Wilcoxon estimator called the HBR (high breakdown rank-based) achieves 50% breakdown in both the X -space and the Y -space; see Chang et al. (1999).

As can be seen from the influence function, the asymptotic distribution of the rank-based estimator is given by

$$\hat{\boldsymbol{\beta}}_\varphi \sim N_p(\boldsymbol{\beta}, \tau_\varphi^2 (\mathbf{X}'\mathbf{X})^{-1}). \quad (9)$$

Note that the only difference between the theory for the LS and rank-based estimators is that σ^2 is replaced by τ_φ^2 . Hence, the asymptotic relative efficiency (ARE) between the

LS and rank-based estimator is σ^2/τ_φ^2 . For Wilcoxon scores, assuming that the random errors have a normal distribution, this ARE is the familiar 0.955; that is, for normal errors using the Wilcoxon analysis instead of the LS analysis results in only a 5% loss of efficiency. Koul et al. (1987) developed an estimator of τ_φ , $\hat{\tau}_\varphi$, which is computed by `Rfit`.

Chapters 3-5 of Hettmansperger and McKean (2011) discuss rank-based analyses of linear, mixed, and nonlinear models. The examples and simulation studies of this paper make use of these confidence intervals and a robust R^2 , which we briefly present. An asymptotic $(1 - \alpha)100\%$ confidence interval for a linear combination of regression parameters, say, $\mathbf{h}^T \boldsymbol{\beta}$ is given by

$$\mathbf{h}^T \hat{\boldsymbol{\beta}}_\varphi \pm t_{\alpha/2, n-p-1} \hat{\tau}_\varphi \sqrt{\mathbf{h}^T (\mathbf{X}'\mathbf{X})^{-1} \mathbf{h}}, \tag{10}$$

where $t_{\alpha/2, n-p-1}$ denotes the upper $\alpha/2$, t -critical value with $n - p - 1$ degrees of freedom.

Next, consider a general linear hypothesis of the form

$$H_0 : \mathbf{H}\boldsymbol{\beta} = \mathbf{0} \text{ versus } H_A : \mathbf{H}\boldsymbol{\beta} \neq \mathbf{0} \tag{11}$$

where \mathbf{H} is a specified $q \times p$ matrix. Let V_{Full} and V_{Red} denote the respective full model column space of \mathbf{X} and the reduced model subspace of V_{Full} constrained by H_0 . Denote the distance between \mathbf{Y} and each of these subspaces respectively by $\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_\varphi\|_\varphi$ and $\|\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\theta}}_\varphi\|_\varphi$, where \mathbf{W} denotes a reduced model $n \times (p - q)$ design matrix and $\hat{\boldsymbol{\theta}}_\varphi$ denotes the corresponding reduced model rank-based estimator. Then $RD = \|\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\theta}}_\varphi\|_\varphi - \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_\varphi\|_\varphi$ denotes the reduction in distance when passing from the reduced model to the full model. This is analogous to the LS reduction in Euclidean squared-distance (reduction in sums-of-squares). The corresponding rank-based F -test is given by

$$F_\varphi = \frac{RD/q}{\hat{\tau}_\varphi/2}. \tag{12}$$

An asymptotic level α test is to reject H_0 in favor of H_A , if $F_\varphi \geq F_\alpha(q, n - p - 1)$, where $F_\alpha(q, n - p - 1)$ denotes the upper α -critical value of an F -distribution with q and $n - p - 1$ degrees of freedom.

As an example of this test, consider the hypothesis that all regression coefficients except for the intercept parameter are 0; i.e.,

$$H_0 : \boldsymbol{\beta} = \mathbf{0} \text{ versus } H_A : \boldsymbol{\beta} \neq \mathbf{0}, \tag{13}$$

In this case the reduced model dispersion is $\|\mathbf{Y}\|_\varphi$. Thus the reduction in dispersion is $RD = \|\mathbf{Y}\|_\varphi - \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_\varphi\|_\varphi$. Using this reduction, the robust F -test statistic is given by expression (12). A nominal level α test of the hypotheses (13) is to reject H_0 if $F_\varphi \geq F_\alpha(p, n - p - 1)$. This test is a robust analogue of the least squares F test statistic that all regression coefficients are 0. Recall that the traditional coefficient of determination R^2 can be expressed as a one-to-one function of the LS F -test. In the same way, a robust coefficient of determination R^2 can be formulated as

$$R_2 = \frac{RD}{RD + (n - p - 1)(\hat{\tau}_\varphi/2)}; \tag{14}$$

see page 243 of Hettmansperger and McKean (2011). We refer to R_2 as a robust coefficient of determination in subsequent examples.

2.2 Optimal scores

The rank-based analysis outlined above requires the selection of a score function $\varphi(u)$. If the form of the underlying error distribution is known, we can obtain an optimal score function which minimizes the variance of the estimator. Using expressions (7) and (8), we can rewrite $1/\tau_\varphi$ as

$$\begin{aligned} (\tau_\varphi)^{-1} &= \int_0^1 \varphi(u)\varphi_f(u) du \\ &= \frac{\int_0^1 \varphi(u)\varphi_f(u) du}{\left(\int_0^1 \varphi_f^2(u) du\right)^{1/2}} \left(\int_0^1 \varphi_f^2(u) du\right)^{1/2} \\ &= \rho \left(\int_0^1 \varphi_f^2(u) du\right)^{1/2} = \rho\sqrt{I(f)}, \end{aligned} \tag{15}$$

where ρ is a correlation coefficient and $\sqrt{I(f)}$ is Fisher Information. Therefore, minimizing τ_φ is equivalent to maximizing the above identity. By the last equality, this is accomplished by making $\rho = 1$; i.e., by taking $\varphi(u)$ to be $\varphi_f(u)$. So expression (8) is the score function which optimizes the rank-based analysis. Since $\hat{\beta}_\varphi$ is location and scale equivariant, only the form of $f(x)$ is needed. Furthermore, since in this case $\tau_\varphi = 1/\sqrt{I(f)}$, the rank-based estimator $\hat{\beta}_\varphi$ is asymptotically fully efficient, i.e., $\hat{\beta}_\varphi$ has the same asymptotic distribution as the maximum likelihood estimator (mle).

For example, if the error distribution is normal, then the optimal score function simplifies to $\varphi(u) = \Phi^{-1}(u)$, the normal scores. If the error distribution is logistic, then the linear Wilcoxon scores are obtained, while double exponential (Laplace) distributed errors produces the sign scores.

2.3 Computation of the rank-based analysis

The computation of a rank-based analysis can be obtained by using the R package `Rfit` developed by Kloke and McKean (2012), which can be downloaded at CRAN. Like R, `Rfit` is freeware and can run on all platforms (windows, linux, and mac). As we discuss in Section 3, it is easy to install new scores in `Rfit` based on a general scores function. For now, we illustrate the computation of `Rfit` for a Wilcoxon analysis in the following example.

Example 2.1 (Linear Model with Skew-Normal Errors). We use a simulated data set based on the model $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + e$, where $x_1 = 1, \dots, 50$; x_2 and x_3 are variates from a standard normal distribution; and the random errors are generated from a standard skew-normal distribution with shape parameter $\alpha = -8$, as discussed in Section 3. We set $\beta_1 = 0.01$, $\beta_2 = 0.15$, and $\beta_3 = 0.0$. The sample size is $n = 50$. The data set can be downloaded at the url cited in Section 1. The code segments below assume that the R vectors `y`, `x1`, `x2`, `x3`, contain respectively the responses and values for x_1 , x_2 and x_3 . For this example, the following R code using the package `Rfit` computes the Wilcoxon fit of the model, prints out the table of coefficients, and saves the Studentized residuals in the vector `studw`.

```
> fitw <- rfit(y ~ x1+x2+x3)
> summary(fitw)
Call:
rfit.default(formula = y ~ x1 + x2 + x3)

Coefficients:
                Estimate Std. Error t.value  p.value
(Intercept) -0.4672019   0.1811862  -2.5786 0.013186 *
x1            0.0044180   0.0056997   0.7751 0.442228
x2            0.2435079   0.0803388   3.0310 0.003993 **
x3           -0.0012315   0.0688965  -0.0179 0.985817
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared (Robust): 0.1811501
Reduction in Dispersion Test: 3.39212 p-value: 0.02567

> studw <- rstudent(fitw)
```

Based on the summary table, the 95% confidence intervals for β_j , $j = 1, 2, 3$, (10), trap the true parameters. The overall F_φ test that all the regression coefficients are 0 except for the intercept is significant, $p = 0.0257$. The value of the robust coefficient of determination R_2 , expression (14), is 18%.

The last command stores the Studentized Wilcoxon residuals in the vector `studw`. These residuals are adjusted for both variance of the errors and location in the X -space; see McKean and Sheather (2009) for a review of robust diagnostic procedures. Figure 1 displays Studentized residual and normal $q-q$ plots based on the Wilcoxon fit.

As noted in Section 3, the skew-normal distribution chosen to generate the random variates in this example is left skewed. Hence, as expected, the residuals show longer left than right tails. These plots show that scores for left-skewed error distributions are more appropriate for this data than the Wilcoxon scores. There appears to be one large outlier in the left tail, also. ■

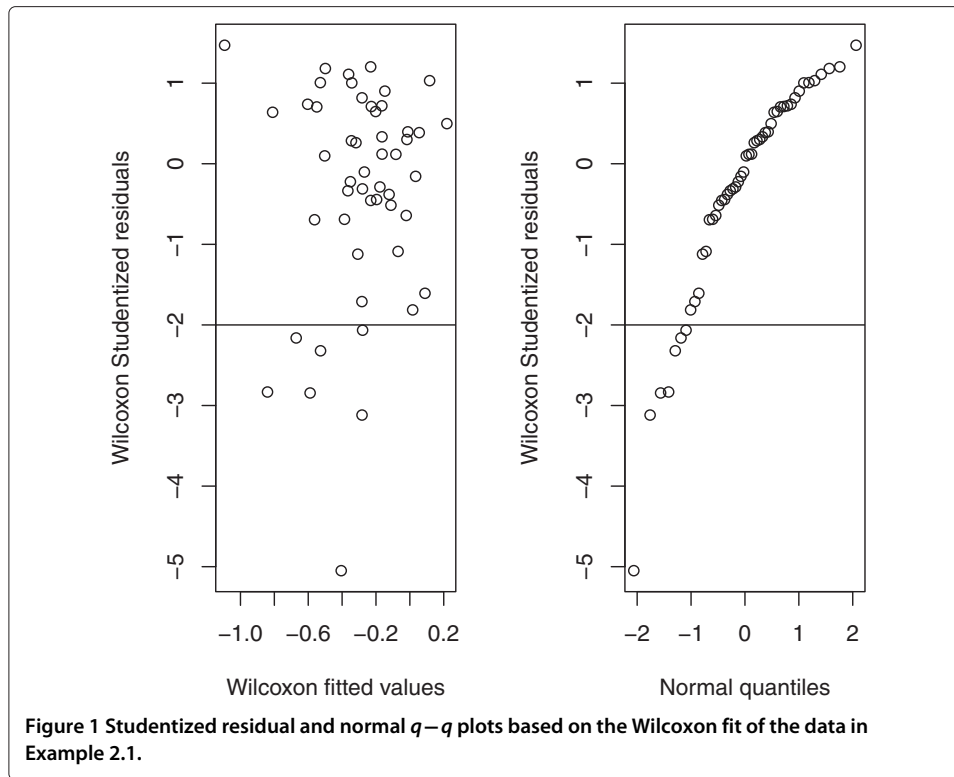
3 Skew-normal error distributions

The family of skew-normal distributions consists of left and right skewed distributions along with the normal distribution. The pdfs in this family are of the form

$$f(x; \alpha) = 2\phi(x)\Phi(\alpha x), \tag{16}$$

where the parameter α satisfies $-\infty < \alpha < \infty$ and $\phi(x)$ and $\Phi(x)$ are the pdf and cdf of a standard normal distribution, respectively. For this paper, if a random variable X has this pdf, we say that X has a standard skew-normal distribution with parameter α and write $X \sim SN(\alpha)$. If $\alpha = 0$, then X has a standard normal distribution. Further X is distributed left skewed if $\alpha < 0$ and right skewed if $\alpha > 0$. This family of distributions was introduced by Azzalini (1985), who discussed many of its properties.

In this paper, we are interested in linear models, (1), where the random errors may have skew-normal errors. In this case, the random error can be written as $e_i = b\varepsilon_i$, where ε_i has a standard skew-normal distribution and b is a scale parameter. The rank-based



estimator $\hat{\beta}_\varphi$ and corresponding analysis are regression and scale equivariant, so there is no need to estimate the scale parameter b . The only scale parameter requiring estimation for standard errors is τ_φ . Likewise, for inference on the vector of parameters β there is no need to estimate the shape parameter α .

What rank scores would be best for such error distributions? To get an idea, we next discuss the optimal scores for a specified α . To obtain the optimal rank-based scores, because of equivariance, we need only the form (down to scale and location) of the pdf. So for the derivation of the scores, assume that the random variable $X \sim SN(\alpha)$ with pdf (16). It easily follows that

$$\frac{f'(x; \alpha)}{f(x; \alpha)} = x - \frac{\alpha\phi(\alpha x)}{\Phi(\alpha x)}. \quad (17)$$

Denote the inverse of the cdf of X by $F^{-1}(u; \alpha)$. Then it follows from expression (15) that the optimal score function for X is

$$\varphi_\alpha(u) = F^{-1}(u; \alpha) - \frac{\alpha\phi(\alpha F^{-1}(u; \alpha))}{\Phi(\alpha F^{-1}(u; \alpha))}. \quad (18)$$

For all values of α , this score function is strictly increasing over the interval $(0, 1)$; see Azzalini (1985). As expected, for $\alpha = 0$, expression (17) simplifies to the normal scores. Due to the first term on the right-side of expression (18), all the score functions in this family are unbounded, indicating that the skew-normal family of distributions is light-tailed. Thus the influence functions of the rank-based estimators based on scores in this

family are unbounded in the Y -space and, hence, are not robust. This includes the normal scores, but Huber (1981) pointed out that normal scores are technically robust and, as our simulation studies show, the family of skew-normal scores seems also to be technically robust.

Figure 2 displays the pdf's and corresponding optimal scores for three values of α : -7 , 1 , 5 . Note that the pdf for $\alpha = -7$ is left skewed while those for positive α values are right skewed. Unsurprisingly, the pdf for $\alpha = 1$ is closer to being symmetric than the other pdfs. The score function for the left-skewed pdf emphasizes relatively the right tails over the left tails, while the reverse is true for the right-skewed pdfs.

3.1 Computation of the rank-based analysis using skew-normal scores

The computation of the rank-based analysis can be obtained by using the R package `Rfit`. It is easy to install the family of skew-normal scores. Briefly, rank-based scores form a class in `Rfit` consisting of three parts: the score function, its derivative, and a vector of parameters used in the definition of the function. For the skew-normal scores, details are given in the appendix, but for the readers convenience the necessary R code is contained in the R function `skewns`, which we have placed at the web site cited in the introduction.

Example 3.1 (Example 2.1, Continued). We now return to Example 2.1 and show the computation of the rank-based analysis of it based on the skew-normal scores with shape parameter $\alpha = -8$. The first two lines of code define the skew-normal scores as `salp` and the third line sets the shape parameter. Details of this definition can be found in the appendix.

```
> source("skewns.R")
> salp <- skewns
> salp@param <- c(-8)
> fits <- rfit(y ~ x1+x2+x3,scores=salp)
> summary(fits)
```

Call:

```
rfit.default(formula = y ~ x1 + x2 + x3, scores = salp)
```

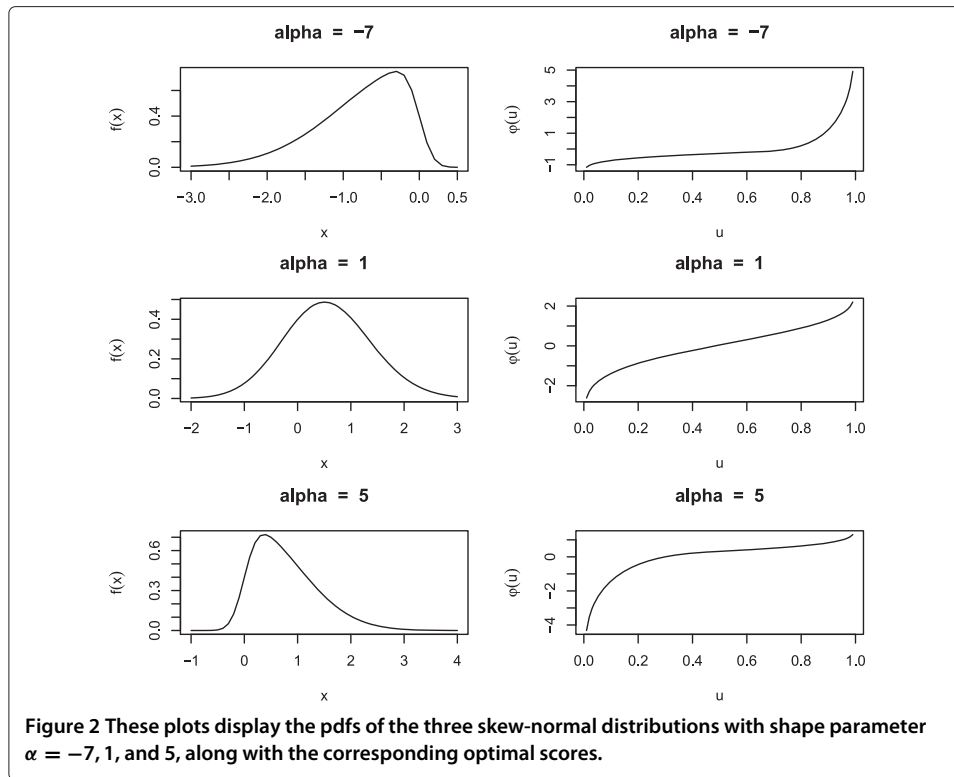
```
Coefficients: Estimate Std. Error t.value p.value
(Intercept) -0.4176052  0.1443377 -2.8933 0.0058088 **
x1 0.0042469  0.0034173  1.2428 0.2202441
x2 0.1830920  0.0481674  3.8012 0.0004219 ***
x3 0.0543903  0.0413072  1.3167 0.1944525
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Multiple R-squared (Robust): 0.2864934 Reduction in Dispersion Test:
6.15677 p-value: 0.00131
```

```
> studs <- rstudent(fits)
```

Note that the skew-normal analysis is much more precise than the Wilcoxon analysis of the last section. The empirical ARE is $(\tau_W/\tau_{\alpha=-8})^2 = 2.78$; i.e., for this data set, the



skew-normal analysis is 2.8 times more efficient than the Wilcoxon analysis. Note, also, that the robust coefficient of determination, R_2 , has increased from 18% to 28%.

3.2 Sensitivity analysis

For a verification of the technical robustness of the rank-based skew-normal analysis, we conducted a small sensitivity analysis. We generated $n = 50$ observations from a linear model of the form $y_i = x_i + e_i$, where x_i has a $N(0, 1)$ distribution and e_i has a $N(0, 10^2)$ distribution. The x_i 's and e_i 's are all independent. We added outliers of the form

$$y_{50} \leftarrow y_{50} + \Delta, \tag{19}$$

where Δ is in the set $\{0, 20, 40, 60, 80, 100, 1000, 2000\}$. The sensitivity curve for an estimator $\hat{\beta}$ is given by the function

$$S(\Delta; \hat{\beta}) = \hat{\beta} - \hat{\beta}(\Delta), \tag{20}$$

where $\hat{\beta}$ and $\hat{\beta}(\Delta)$ denote the estimates of β on the original and modified data (19), respectively. We obtained sensitivity curves for the estimators: Wilcoxon, normal scores, skew-normal ($\alpha = 3$), skew-normal ($\alpha = 5$), skew-normal ($\alpha = 7$), and maximum likelihood estimates (mle). The mles were computed by the package `sn`. For all values of Δ , the changes in all of the the rank-based estimates were less than 0.004. Thus the rank-based skew-normal estimators, including the normal scores estimator, exhibited technical robustness for this study. On the other hand, the mle was sensitive to the values of Δ . We show these changes in Table 1; hence, for this study, the mle was not robust.

Table 1 Values of the sensitivity function for the mle at the given values of Δ

Δ	0	20	40	60	80	100	1000	2000
mle	0.00	-0.07	-0.07	-0.00	0.12	0.30	-5.80	-6.32

3.3 Range of practical α parameters

In Sections 4 and 5, the results of simulation studies are presented. The error distributions involve families of skew-normal distributions. So a practical range of α values is needed. Pourahmadi (2007) derived properties of the skew-normal distribution including its moment generating function. In particular, he showed that if X has the $SN(\alpha)$ distribution, then X converges in distribution to $|Z|$ as $\alpha \rightarrow \infty$ where Z is $N(0, 1)$; i.e., the distribution of X converges to a half-normal distribution. Likewise, X converges in distribution to $-|Z|$ as $\alpha \rightarrow -\infty$. In terms of α , the convergence is fairly fast. Table 2 serves as an illustration of this as it displays the mean (μ), median ($\tilde{\mu}$), variance (σ^2), and coefficient of skewness (ξ) for various values of α . The last column of the table shows the value of these parameters for the half-normal distribution. Positive values of α suffice because if $X \sim SN(\alpha)$ then $-X \sim SN(-\alpha)$. There is little difference between the standard skew-normal distribution and the half-normal distribution for values of α near ± 8 . Based on these facts, we use skew-normal distributions with values of α between -12 and 12 for our Monte Carlo investigations.

Because we are interested in linear models, there is another practical reason for this range of α values. Note that the support of a skew-normal distribution is $(-\infty, \infty)$ making it ideal for error distributions for regression models. On the other hand, the support of a half-normal distribution is $(0, \infty)$, which is generally the support of a survival distribution. Often, the log's of such variables are modeled as accelerated failure time (AFT) models, as briefly discussed in Section 1.

4 Monte Carlo study

This section contains the results of a small simulation study concerning rank-based procedures based on skew-normal scores. The model simulated is

$$y_i = \beta_0 + \beta_1 x_i + \theta c_i + e_i, \tag{21}$$

where x_i is distributed $N(0, 1)$; e_i is distributed from a selected error distribution; $i = 1, \dots, 100$; the x_i 's and e_i 's are all independent; and the variable c_i is a treatment indicator with values of either 0 or 1. We selected two error distributions for the study. One is a skew-normal distribution with shape parameter $\alpha = 5$ while the other is a contaminated version of a skew-normal. The contaminated errors are of the form

$$e_i = (1 - I_{\epsilon,i})W_i + I_{\epsilon,i}V_i, \tag{22}$$

Table 2 Parameters (mean μ , median $\tilde{\mu}$, variance σ^2 , and coefficient of skewness ξ) for $SN(\alpha)$ distributions for the given values of α

α	0.00	1.00	2.00	4.00	6.00	7.00	8.00	10.00	15.00	20.00	∞
μ	0.00	0.56	0.71	0.77	0.79	0.79	0.79	0.79	0.80	0.80	0.80
$\tilde{\mu}$	0.00	0.55	0.66	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.67
σ^2	1.00	0.68	0.49	0.40	0.38	0.38	0.37	0.37	0.37	0.36	0.36
ξ	0.00	0.14	0.45	0.78	0.89	0.92	0.93	0.96	0.98	0.98	0.99

The values in the last column (∞) are those for a half-normal distribution.

where W_i has a skew-normal distribution with shape parameter $\alpha = 5$, V_i has a $N(\mu_c = 10, \sigma_c^2 = 36)$ distribution, $I_{\epsilon,i}$ has a binomial $(1, \epsilon = 0.15)$ distribution, and W_i, V_i , and $I_{\epsilon,i}$ are all independent. Hence, this contaminated distribution is skewed with heavy right tails. The design is slightly unbalanced with $n_1 = 45$ and $n_2 = 55$. Without loss of generality β, θ , and β_0 were set to 0.

For the rank based procedures, we selected the rank-based procedure based on the score function $\varphi_5(u)$, (18), which is optimal for a skew normal distribution with $\alpha = 5$ and then three on each side of the optimal, i.e., procedures based on the score functions $\varphi_\alpha(u)$ with $\alpha = 2, 3, 4, 6, 7$, and 8. With the discussion in Section 3.3 in mind, we also selected the rank-based procedure with $\alpha = 10$. The rank-based Wilcoxon, least squares (LS) procedure, and mle procedures complete the methods investigated. The empirical results presented are the empirical AREs, which for each estimator is the ratio of the empirical mean-square error (MSE) of the mle to the empirical MSE of the estimator; hence, values of this ratio less than 1 are favorable to the mle while values greater than 1 are favorable to the estimator. Secondly, we present the empirical confidence intervals with nominal confidence 0.95. For all the procedures, we chose asymptotic confidence intervals of the form $\hat{\beta} \pm 1.96SE(\hat{\beta})$. We used a simulation size of 10,000.

The results are presented in Table 3. For the skew-normal errors, for both parameters β and θ , all the rank-based estimators except the Wilcoxon estimator are more efficient than the mle estimator. Note that the most efficient estimator for both β and θ is the rank-based estimator with $\alpha = 5$; although, empirical efficiencies are not significantly different from the empirical efficiencies for a few of the nearby (α close to 5) rank-based estimators. In terms of validity, the empirical confidences of all the procedures are close to the nominal confidence of 0.95. Not surprisingly, LS performed the worst overall.

For the contaminated error distribution, the rank-based estimators are much more efficient than the mle procedure. Further, the estimator with scores based on $\alpha = 5$ is still the most empirically powerful in the study. It has empirical efficiency of 785% relative to the mle for β and 1310% for θ . Even the Wilcoxon procedure is over 756% more efficient

Table 3 Summary of results of simulation study of rank-based procedures and the mle procedure for the skew-normal with shape $\alpha = 5$ distribution and a skew-normal contaminated distribution

Proced.	Skew normal errors				Contaminated errors			
	β		θ		β		θ	
	ARE	Conf.	ARE	Conf.	ARE	Conf.	ARE	Conf.
rb $\alpha = 2$	1.02	0.96	1.04	0.96	6.61	0.98	10.84	0.98
rb $\alpha = 3$	1.09	0.96	1.11	0.96	7.43	0.97	12.24	0.98
rb $\alpha = 4$	1.13	0.96	1.15	0.96	7.79	0.97	12.91	0.98
rb $\alpha = 5$	1.14	0.96	1.16	0.96	7.85	0.96	13.10	0.97
rb $\alpha = 6$	1.13	0.95	1.16	0.96	7.73	0.96	13.02	0.97
rb $\alpha = 7$	1.11	0.95	1.14	0.95	7.49	0.95	12.72	0.97
rb $\alpha = 8$	1.09	0.95	1.12	0.95	7.17	0.95	12.30	0.96
rb $\alpha = 10$	1.04	0.94	1.07	0.94	6.46	0.94	11.22	0.95
rb Wil.	0.78	0.95	0.79	0.95	4.70	0.96	7.56	0.97
LS	0.70	0.95	0.71	0.95	0.20	0.95	0.31	0.95
mle	1.00	0.93	1.00	0.93	1.00	0.96	1.00	0.99

than the mle for θ . On the basis of the empirical confidences, for both parameters, all procedures appear to be from slightly to moderately conservative. Least squares performed extremely poor in the contaminated part of the study. All the rank-based procedures based on skew-normal scores display technical robustness in this study.

5 Hogg-Type adaptive procedure

In Section 3, we discussed the rank-based method based on the optimal score function for a specified shape parameter α . Asymptotically, it is as efficient as the mle and, at least for the situations covered in the simulation study, the rank-based estimator appears to be more efficient than the mle for finite samples. In practice, though, the true shape parameter is not known. One could obtain the mle of α and use that score function. The mle, however, is not robust. Reconsidering the empirical study, note from Table 3 that the rank-based estimates close to the optimal rank-based estimator were more efficient than the mle and most had efficiencies that were quite close to that of the optimal. That is, in selecting a score function, perhaps close would suffice. In this section, we consider a Hogg-type adaptive scheme which has this as its goal.

Hogg et al. (1975) proposed an adaptive procedure for tests of the difference in locations for the two sample problem. The null hypothesis is that the two population distributions are the same. The selection of the test is based on a pair of selector statistics that measure respectively skewness and tail weight of the underlying error distribution. These selector statistics are functions of the order statistics of the combined samples. Several distribution-free rank tests of significance level δ comprise the tests. Under the null hypothesis, it follows from the sufficiency and completeness of the combined order statistics and the distribution-freeness of the rank test statistics that the selected test maintains the level δ . See, also, the discussion in Chapter 10 of Hogg et al. (2013).

This is fine for simple location tests where we have distribution-free rank tests, but in our case we are fitting a linear model and, hence, the adaption must be based on the residuals from an initial fit. Thus the above mentioned sufficiency result is not true for our fitting case. Shomrani (2003) developed a Hogg-type adaptive scheme for fitting a linear model based on an initial fit. In Shomrani's scheme, the selector statistics are functions of the residuals from the initial fit. While the significance level is no longer maintained, based on the results of a large simulation study, the scheme's empirical levels were generally close to the nominal value. In Chapter 6 of Kloke and McKean (2014), R software is developed for this scheme.

The adaptive scheme of Shomrani (2003) was formed for a wide range of error distributions: from left to right skewed and from light to heavy tailed distributions. We refine this scheme for the skew-normal family of distributions. As discussed above, there are two selector statistics involved. One, Q_1 , selects based on skewness while the other, Q_2 selects based on tail thickness. In a preliminary study over the skew-normal family, tail thickness did not seem to be a paramount issue, so we focus on Q_1 alone.

Let $\mathbf{V} = (V_1, V_2, \dots, V_n)^T$ be a random vector and define

$$Q_1(\mathbf{V}) = \frac{\bar{U}_{.05} - \bar{M}_{.5}}{\bar{M}_{.5} - \bar{L}_{.05}}, \quad (23)$$

where $\bar{U}_{.05}$, $\bar{M}_{.5}$, and $\bar{L}_{.05}$ are the averages of the largest 5% of the V_i 's, the middle 50% of the V_i 's, and the smallest 5% of the V_i 's, respectively. Large values of Q_1 indicate that

the right tails of the sample are longer than the left tails; i.e., indicating an underlying right skewed distribution. Likewise, small values of Q_1 indicate left-skewness. Note that as left (right) skew increases, Q_1 is likely to decrease (increase). The statistic Q_1 is not robust. One scheme under current investigation is to replace the means by medians. Keep in mind, though, that a robust diagnostic analysis is available for the initial robust fit. Hence in practice, outliers are easily flagged and modifications to the adaptive scheme can be made.

5.1 Adaptive scheme for skew normals

Our adaptive scheme consists of the 7 optimal score functions for skew-normal distributions with $\alpha = -12, -8, -4, 0, 4, 8,$ and 12 . So there are three scores each for left and right skewed distributions along with the normal scores. The scheme utilizes residuals from an initial Wilcoxon fit. We chose the Wilcoxon because it is robust. Also, it is optimal for a symmetric distribution (logistic) and, hence, less likely to bias selection for left or right skewness.

We decided to set the benchmarks for the selector statistic Q_1 based on the medians of the distribution of Q_1 for the family of skew-normal distributions with $\alpha = -10, -6, -2, 2, 6,$ and 10 . Then the selection part of the adaptive scheme is given by:

$$\begin{aligned}
 & Q_1 < \text{med}(Q_1|\alpha = -10) \Rightarrow \text{Select score using } \alpha = -12 \\
 & \text{med}(Q_1|\alpha = -10) < Q_1 < \text{med}(Q_1|\alpha = -6) \Rightarrow \text{Select score using } \alpha = -8 \\
 & \text{med}(Q_1|\alpha = -6) < Q_1 < \text{med}(Q_1|\alpha = -2) \Rightarrow \text{Select score using } \alpha = -4 \\
 & \text{med}(Q_1|\alpha = -2) < Q_1 < \text{med}(Q_1|\alpha = 2) \Rightarrow \text{Select score using } \alpha = 0 \quad (24) \\
 & \text{med}(Q_1|\alpha = 2) < Q_1 < \text{med}(Q_1|\alpha = 6) \Rightarrow \text{Select score using } \alpha = 4 \\
 & \text{med}(Q_1|\alpha = 6) < Q_1 < \text{med}(Q_1|\alpha = 10) \Rightarrow \text{Select score using } \alpha = 8 \\
 & \text{med}(Q_1|\alpha = 10) < Q_1 \Rightarrow \text{Select score using } \alpha = 12
 \end{aligned}$$

We estimated the medians of the sampling distributions of the statistic Q_1 based on simulations of size 10,000 drawn from the appropriate skew-normal distributions. Because Q_1 is location and scale equivariant, simulation using standard skew-normal distributions suffices. The estimated (simulated) medians used for the scheme are given in Table 4.

In summary, the algorithm for our adaptive scheme is:

1. Fit using Wilcoxon scores \Rightarrow Obtain residuals \hat{e}_W .
2. Compute $Q_1(\hat{e}_W)$ and then select φ_α using expression (24), using the estimated medians of Q_1 .
3. Fit with selected score φ_α .
4. Inference is based on the fit of Step (3).

We next try the scheme on the data of Example 2.1.

Table 4 Simulated sample median of the distribution of Q_1 drawn from the skew-normal distribution with shape parameter α

α	-10	-6	-2	2	6	10
Median Q_1	0.44	0.49	0.73	1.37	2.05	2.26

The simulation size is 10,000. These estimated medians are used in the adaptive scheme (24).

Example 5.1 (Example 2.1, continued). For the data in Example 2.1, the selector Q_1 has the value 0.344; hence the adaptive scheme (24) selects the score function with $\alpha = -12$. Table 5 summarizes the results of the fits based on skew-normal scores with α 's in a neighborhood of -8 .

The values of the regression coefficients are given along with the robust coefficient of determinations R_2 and estimates of τ_φ . The fits are quite similar. Notice that in terms of precision, $\hat{\tau}_\varphi$'s, that the fit with $\alpha = -8$ is the most precise. ■

5.2 Simulation study

We investigated the validity and efficiency of the adaptive scheme in a Monte Carlo study using situations similar to those in Section 4. In particular, the sample size is set at $n = 100$ and the linear model contains one predictor coefficient, β , and one indicator coefficient for the treatment, θ . We considered three sampling situations. In the first situation (I), for each simulation, α is randomly selected from the set $\{-12, -11, \dots, 11, 12\}$ where the selections are made equilikely. Then the random errors for the model are generated from this $SN(\alpha)$ distribution. For Situation II, again α is randomly selected from the same set of values but now the random errors are random variables of the form

$$e_i = (1 - I_{.15})S_i + I_{.15}C_i,$$

where $S_i \sim SN(\alpha)$, $C_i \sim N(10, 6^2)$, $I_{.15}$ is Bernoulli with proportion of success 0.15, and $S_i, C_i, I_{.15}$ are independent. Thus, Situation II is the same as the second situation of Section 3, i.e., right-skewed contamination. Situation III is the same as situation II except that $C_i \sim N(0, 6^2)$, i.e., symmetric contamination. 10,000 simulations were used for each situation.

The methods considered are: our adaptive scheme (AdSch), least squares (LS), Wilcoxon (Wil), and maximum likelihood (mle). We also considered the procedure based on the correct α ; i.e., the α which is selected for the distribution of the random errors. Note that this is not a statistical method and we label it as Optrv, "rv" for random variable. Even for Situation I, the distribution of its rank-based estimate depends on the multinomial random variable involved in the selection of the simulated distribution. We only include it to serve as an yardstick for the four statistical methods.

As in the simulation study of Section 3, we considered empirical efficiency (relative to the mle) and validity of 95% confidence intervals for the parameters β and θ . The results for Situation I are summarized in Table 6. The adaptive scheme was more efficient than

Table 5 For the data of Example 2.1, the adaptive scheme chose the score function with $\alpha = -12$

α	β_1	β_2	β_3	$\hat{\tau}_\varphi$	Robust R_2
$\alpha = -12$ (Adaptive Choice)	0.0035	0.2190	0.0368	0.3675	0.2743
$\alpha = -10$	0.0030	0.2095	0.0480	0.3699	0.2732
$\alpha = -9$	0.0037	0.1923	0.0514	0.3531	0.2828
$\alpha = -8$	0.0042	0.1831	0.0544	0.3484	0.2865
$\alpha = -7$	0.0043	0.1822	0.0537	0.3533	0.2849
$\alpha = -6$	0.0044	0.1808	0.0540	0.3490	0.2884
$\alpha = -5$	0.0044	0.1805	0.0531	0.3580	0.2838
$\alpha = -4$	0.0043	0.1854	0.0449	0.3886	0.2668

This table shows the variation in the rank-based fits for score functions in a neighborhood of $\alpha = -8$.

Table 6 Empirical efficiencies and confidence coefficients for Situation I (error distributions are skew-normals)

	AdSch	Optrv	LS	Wil	mle
β , ARE	1.06	1.12	0.75	0.80	1.00
β , Conf	0.95	0.95	0.95	0.95	0.93
θ , ARE	1.05	1.11	0.73	0.79	1.00
θ , Conf	0.95	0.95	0.95	0.95	0.94

Empirical efficiencies are MSE's relative to the MSE of the mles.

all other statistical methods for both parameters. In particular, it was more efficient than the mle by 6 and 5% respectively. In terms of validity all of the methods are valid. The adaptive scheme was less efficient (by 6%) than the optimal non-statistical procedure.

For Situation I, the random errors have a skew-normal distribution with shape parameter α drawn from the set $\{-12, -11, \dots, 12\}$, while the scheme selects scores from the set $\{-12, -8, -4, 0, 4, 8, 12\}$. These sets are different; hence, it does not make sense to consider when the scheme made the "correct" selection. We did keep track of how often the selection was within two units of the distribution simulated. For the 10,000 simulations of Situation I, the estimate of this proportion is 0.584. Note that for Situations II and III, the random errors have a contaminated skew-normal distribution. In particular, it is not a skew-normal distribution. So for Situations II and III, this proportion is irrelevant.

The results for Situation II are summarized in Table 7. For this right-skewed contaminated situation, the adaptive scheme is much more efficient than the mle, 358% and 356% respectively for β and θ . The adaptive scheme is adapting to this heavy tailed situation and the mle is not robust. As expected, the Wilcoxon performs best. All the methods are valid. We included the optimal random variable procedure since it was in Situation I. Note, though, that in Situations II and III the distribution of the random errors is not a skew-normal distribution. Besides not being a statistical method it is no longer optimal in any sense.

The results for the situation, III, with symmetric contamination can be found in Table 8. The adaptive scheme is much more efficient than the mle, 367% and 477% respectively for β and θ . As expected, the Wilcoxon performs best. All the rank-based methods are valid. The mle is slightly conservative for θ . The same comments hold for Optrv as in Situation II.

6 Conclusion

Rank-based analyses of linear models depend on the selection of a score function. In practice, often the Wilcoxon (linear) score function is chosen. These scores require no tuning constants and, further, the Wilcoxon rank-based analysis attains 95.5% efficiency relative

Table 7 Empirical efficiencies and confidence coefficients for Situation II (error distributions are skewed contaminated skew-normals)

	AdSch	Optrv	LS	Wil	mle
β , ARE	3.58	1.86	0.92	7.27	1.00
β , Conf	0.97	0.97	0.95	0.95	0.94
θ , ARE	3.56	1.56	0.94	7.36	1.00
θ , Conf	0.97	0.96	0.95	0.95	0.95

Empirical efficiencies are MSE's relative to the MSE of the mles.

Table 8 Empirical efficiencies and confidence coefficients for Situation III (error distributions are symmetrically contaminated skew-normals)

	AdSch	Optrv	LS	Wil	mle
β , ARE	3.67	1.01	0.26	6.16	1.00
β , Conf	0.94	0.97	0.95	0.97	0.95
θ , ARE	4.77	0.85	0.35	8.18	1.00
θ , Conf	0.94	0.97	0.95	0.97	0.98

Empirical efficiencies are MSE's relative to the MSE of the mles.

to the traditional least squares (LS) analysis when the random error distribution is normal. However, rank-based analyses are easily optimized if there is knowledge of the distribution of the random errors of the linear model. For example, if the random errors are normally distributed then selecting the normal scores for the rank-based analysis results in the efficiency of 100% (fully efficient) relative to the LS analysis.

In this paper, we have presented the rank-based analyses based on appropriate score functions for random errors having a distribution from the family of skew-normal distributions. In this case, the score function depends on the shape parameter α , $-\infty < \alpha < \infty$. Of course, the rank-based analysis is fully efficient if the correct α is known. The rank-based analysis is a complete analysis, including fitting, inference (rank-based ANOVA), and robust diagnostic procedures. Based on the results of our Monte Carlo, these rank-based analyses appear to be more efficient than the maximum likelihood (mle) analysis for the skew-normal distributions considered. The most efficient rank-based analysis is based on the optimal score function, but even those rank-based analyses with shape parameters within three units of the correct α were more efficient than the mle in these situations. They were much more efficient than the mle's over situations where the error distribution had a contaminated skew-normal distribution. Based on empirical confidence levels, all the methods in the study were valid.

The good efficiency results for the rank-based analyses in a neighborhood of the true α suggest that a Hogg-type adaptive scheme would have high efficiency. In Section 5, we developed such a scheme for the skew-normal family of distributions based on an initial robust Wilcoxon fit. In the Monte Carlo studies we performed, this scheme was more efficient than the mle over the family of skew-normal distributions and was much more efficient than the mle over the contaminated skew-normal situations. Furthermore, for the situations covered, this adaptive scheme appears to be valid.

Kloke and McKean (2012) developed an R package `Rfit` for these rank-based analyses, which can be freely downloaded at CRAN. The default scores are the Wilcoxon scores, but, as we discuss in Section 3 it is easy to add classes of scores including the optimal scores for skew-normal distributions. The adaptive scheme of Section 5 is also easily coded using `Rfit`. The necessary code for the scores and the adaptive scheme can be found at the web site cited in Section 1. Hence, computation of these rank-based analyses is not a problem.

The rank-based analyses using skew-normal scores are robust in Y (response) space, but not in X (factor) space. The weighted Wilcoxon fit proposed by Chang et al. (1999) yield a robust rank-based analysis which possesses 50% breakdown in X (factor) space. We are now developing such an analysis for the skew-normal scores; see Abebe et al. (2014) for discussion. This analysis could also be part of an adaptive scheme.

A Appendix: R code for the class of skew normal scores

Rank-based scores form a class in `Rfit` which consists of three principal parts. The first part is the scores function itself and the third part consists of any parameters in the function expression. Thus parts (1) and (3) for the skew-normal scores are respectively given by expression (18) and the shape parameter α . The second part is the derivative of the score function which is used in the `Rfit` function which estimates τ_φ . Let $l(x)$ denote the function defined in expression (18). The derivative of the optimal score function is given by

$$\varphi'_\alpha(u) = l'[F^{-1}(u; \alpha)] \frac{1}{2\phi[F^{-1}(u; \alpha)] \Phi[\alpha F^{-1}(u; \alpha)]}, \quad (25)$$

where

$$l'(x) = 1 + \frac{\alpha^2 \phi(\alpha x) [\alpha x \Phi[\alpha F^{-1}(u; \alpha)] + \phi(\alpha x)]}{\Phi^2(\alpha x)}.$$

To complete the class statement for the skew-normal scores we need only compute the quantiles $F^{-1}(u; \alpha)$. Azzalini (2013) developed the R package `sn` (available at CRAN) which computes the quantile function $F^{-1}(u; \alpha)$ and, also, the corresponding pdf and cdf. The command `qsn(u, shape=alpha)` returns $F^{-1}(u; \alpha)$, for $0 < u < 1$. The package `sn` requires the package `mnormt`.

The following R code defines the class of skew-normal scores:

```
> library(mnormt)
> library(sn)
> source("skewns.R")
> skewns

An object of class "scores"
Slot "phi":
function (u, param)
{
  alp = param[1]
  f3 = qsn(u, shape = alp)
  top = f3 - ((alp * dnorm(alp * f3))/pnorm(alp * f3))
  top
}

Slot "Dphi":
function (u, param)
{
  alp = param[1]
  f3 = qsn(u, shape = alp)
  part1 = alp * f3 * pnorm(alp * f3) + dnorm(alp * f3)
  part2 = (alp^2 * dnorm(alp * f3) * part1)/(pnorm(alp * f3)^2)
  part3 = (1 + part2)/(2 * dnorm(f3) * pnorm(alp * f3))
  part3
}

Slot "param":
[1] 1
```

The next code segment obtains the data for a plot of the scores with shape parameter $\alpha = -7$.

```
> alp <- -7
> x <- seq(-3, .5, .1)
> y <- dsn(x, shape=alp)
> n <- 100
> ps <- (1:n)/(n+1)
> salp <- skewns
> salp@param <- c(alp)
> phialp <- getScores(salp, ps)
```

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

The authors contributed equally to the manuscript. All authors read and approved the final manuscript.

Acknowledgement

We acknowledge the helpful comments of an associate editor and a referee on the original manuscript.

Received: 10 February 2014 Accepted: 21 July 2014

Published online: 22 October 2014

References

- Abebe, A, McKean, JW: Weighted Wilcoxon estimators in Nonlinear Regression. *Aust N Z J. Stat.* **55**, 401–420 (2013)
- Abebe, A, McKean, JW, Kloke, JD, Bilgic, Y: Iterated Reweighted Rank-Based Estimates for GEE Models. Technical Report (2014)
- Azzalini, A: A class of distributions which includes the normal ones. *Scand. J. Stat.* **12**, 171–178 (1985)
- Azzalini, A: R package *sn*: The skew-normal and skew-*t* distributions (version 0.4-18). Università di Padova, Italia (2013)
- Chang, W, McKean, J, Naranjo, J, Sheather, S: High-breakdown rank regression. *J. Am. Stat. Assoc.* **94**, 205–219 (1999)
- Hettmansperger, TP, McKean, JW: Robust Nonparametric Statistical Methods. 2nd Edition. Chapman-Hall, Boca Raton, FL (2011)
- Hogg, RV, Fisher, DM, Randles, RH: A two-sample adaptive distribution-free test. *J. Am. Stat. Assoc.* **70**, 656–661 (1975)
- Hogg, RV, McKean, JW, Craig, AT: Introduction to Mathematical Statistics. 7th Ed. Pearson, Boston (2013)
- Huber, PJ: Robust Statistics. John Wiley & Son (1981)
- Jaeckel, LA: Estimating regression coefficients by minimizing the dispersion of residuals. *Ann. Math. Stat.* **43**, 1449–1458 (1972)
- Jurečková, J: Nonparametric estimate of regression coefficients. *Ann. Math. Stat.* **42**, 1328–1338 (1971)
- Kloke, JD, McKean, JW: Rfit: Rank-based estimation for linear models. *R J.* **4**, 57–64 (2012)
- Kloke, JD, McKean, JW: Small sample properties of JR estimators. In: *JSM Proceedings. American Statistical Association, Alexandria, VA* (2013)
- Kloke, JD, McKean, JW: Nonparametric statistical methods using R, Chapman-Hall, Boca Raton, FL (2014)
- Kloke, JD, McKean, JW, Rashid, M: Rank-based estimation and associated inferences for linear models with cluster correlated errors. *J. Am. Stat. Assoc.* **104**, 384–390 (2009)
- Koul, HL, Sievers, GL, McKean, JW: An estimator of the scale parameter for the rank analysis of linear models under general score functions. *Scand. J. Stat.* **14**, 131–141 (1987)
- McKean, J, Sheather, S: Diagnostic procedures. *Wiley Interdiscip. S. Rev.: Comput. Stat.* **1**(2), 221–233 (2009)
- McKean, J, Sievers, G: Rank scores suitable for analysis of linear models under asymmetric error distributions. *Technometrics.* **31**, 207–218 (1989)
- Pourahmadi, M: Construction of skew-normal random variables: Are they linear combinations of normals and half-normals. *J. Stat. Theory Appl.* **3**, 314–328 (2007)
- Shomrani, A: A comparison of different schemes for selecting and estimating score functions based on residuals. Ph.D. thesis, Western Michigan University, Department of Statistics (2003)

doi:10.1186/s40488-014-0018-0

Cite this article as: McKean and Kloke: Efficient and adaptive rank-based fits for linear models with skew-normal errors. *Journal of Statistical Distributions and Applications* 2014 **1**:18.