A Robust Analysis of the General Linear Model Based on One Step R- Estimates
Author(s): Joseph W. McKean and Thomas P. Hettmansperger
Source: *Biometrika*, Vol. 65, No. 3 (Dec., 1978), pp. 571-579
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: https://www.jstor.org/stable/2335909
Accessed: 14-06-2019 01:50 UTC

Biometrika (1978), **65**, 3, *pp.* 571–9
With 1 *text-figure*
Printed in Great Britain

A robust analysis of the general linear model based on one step *R*-estimates

BY JOSEPH W. McKEAN
Program of Mathematical Sciences, University of Texas, Dallas

AND THOMAS P. HETTMANSPERGER
Department of Statistics, Pennsylvania State University, University Park

SUMMARY

Classical analysis of variance with least squares fitting is often used to discern structure in a linear model. McKean & Hettmansperger (1976) proposed a robust analysis based on ranks using the *R*-estimates proposed by Jaeckel (1972). These rank procedures depend on minimizing a dispersion surface and as a result are computationally restricted to small to moderate sized sets. In this paper we propose one step iterations based on a second derivative approximation to the surface. These estimates can be obtained quickly from initial estimates. Further the analysis resulting from these estimates is asymptotically equivalent to the minimum dispersion analysis. Thus it can be recommended for large data sets.

Some key words: Gauss–Newton; General linear hypotheses; Linear model; Regression; Robust estimation.

1. INTRODUCTION AND SUMMARY

Classical analysis of variance is often used to discern structure in a linear model. Generally this involves fitting reduced and full models by the method of least squares. The least squares fit may be spoiled by small but reasonable deviations from normal error structure (Huber, 1972; Andrews, 1974). In such cases the analysis of variance would also seem to be impaired.

Jaeckel (1972) proposed a class of robust *R*-estimates for estimating the regression coefficients of a linear model. Using Jaeckel's estimating procedure to fit reduced and full models, McKean & Hettmansperger (1976) proposed an analysis of linear models. These procedures are a natural extension of the rank tests and estimates in the simple location model and inherit their efficiency and robustness properties. Hettmansperger & McKean (1977) discuss the implementation of these procedures including small sample corrections. We feel that this analysis is a reasonable, robust alternative to least squares for moderate sized data sets. In this paper we shall refer to this analysis as the robust analysis or the fully iterated analysis.

Jaeckel's estimates are obtained by minimizing a dispersion surface and require iterative techniques. Each iteration requires ranking of the residuals. Note, however, since the rankings will not change much from one iteration to the next, the ranking algorithm used should depend on the last set of rankings. Since the surface is convex, first derivative methods such as steepest descent can be used, but these are often slow, requiring many iterations, and hence restrict the application of these methods to moderately sized data sets. In this paper we consider faster, second derivative methods. Based on initial estimates, discussed in § 5, the first step is an approximate Gauss–Newton iteration towards the minimum of the dispersion surface. These estimates can be quickly obtained from starting values requiring only an additional estimate of scale. They are similar to the one step Huber *M*-estimates

proposed by Bickel (1975). Based on these estimates, an analysis of a linear model can be performed similar to the fully iterated analysis. This analysis is outlined in § 3 and we shall refer to it in this paper as the k step rank analysis.

As discussed in the last section, practical problems and preliminary Monte Carlo results indicate that generally one or two steps of this k step analysis yield an analysis which is quite close to the fully iterated analysis. Generally obtaining the fully iterated estimates necessitates leaving these Gauss–Newton approximations and proceeding with the slower first derivative methods such as steepest descent. Hence for a large data set closing the narrow gap between the k step analysis and the fully iterated analysis may be computationally expensive. Thus we feel the k step analysis is computationally more efficient. Some discussion of starting values and stopping values for k is also presented in the last section.

In § 4 we show that this k step analysis has the same asymptotic distribution and efficiency properties as the robust rank analysis. Hence we conclude that robust rank tests and estimates can now be computed for moderate to large linear models.

2. NOTATION AND DESCRIPTION OF THE ROBUST RANK ANALYSIS

2·1. *Estimation*

Let Y be an $n \times 1$ vector of observations which is expressed as the linear model

$$Y = 1b_0 + X_1 b + e, \tag{2·1}$$

where $X = [1, X_1]$ is a known, full column rank $n \times (r+1)$ matrix, b_0 is the intercept parameter, and b is an $r \times 1$ vector of parameters. Assume that the components of e are independent and identically distributed with an absolutely continuous density f which is symmetric about 0. Let $\bar{X}_1$ be the $n \times r$ matrix whose (i,j)th entry is the mean of the jth column of X_1 and let $X_{1c} = X_1 - \bar{X}_1$.

We shall consider rank scores of the form $a(i) = \phi\{i/(n+1)\}$, where ϕ is a nondecreasing function on $(0,1)$, and which further satisfy

$$\Sigma a(i) = 0 = \int \phi(u)\,du, \quad \phi(1-u) = -\phi(u), \quad \int \phi^2(u)\,du = 1. \tag{2·2}$$

The related one-sample scores are defined by $a^+(i) = \phi^+\{i/(n+1)\}$, where $\phi^+(u) = \phi\{\frac{1}{2}(u+1)\}$. Define the scale parameter τ by

$$\tau^{-1} = \int_0^1 \phi(u)\left[-\frac{f'\{F^{-1}(u)\}}{f\{F^{-1}(u)\}}\right] du, \tag{2·3}$$

where F is the distribution function of e_i. Wilcoxon scores can be generated in this way by taking $\phi(u) = \phi_w(u)$, where

$$\phi_w(u) = 12^{\frac{1}{2}}(u - \tfrac{1}{2}). \tag{2·4}$$

For these scores, $\tau^{-1} = 12^{\frac{1}{2}} \int f^2(u)\,du$.

Jaeckel's (1972) estimate of b is a value, $\hat{b}$, which minimizes the convex function

$$D(b) = \Sigma a\{R(Y_i - x_i' b)\}(Y_i - x_i' b), \tag{2·5}$$

where $R(Y_i - x_i' b)$ denotes the rank of the residual $Y_i - x_i' b$ among $Y_1 - x_1' b, \ldots, Y_n - x_n' b$ and x_i' is the ith row of X_1.

The one-sample residual process is defined by

$$S^+(t) = \Sigma a^+(R|\hat{e}_i - t|)\,\mathrm{sgn}\,(\hat{e}_i - t), \tag{2·6}$$

where $\hat{e}_i = Y_i - x_i'\hat{b}$ and $\operatorname{sgn}(u)$ is 1, 0 or -1 depending on whether u is positive, zero or negative. Let $\hat{b}_0$ be a value which solves $S^+(\hat{b}_0) \doteq 0$. Then $\hat{b}_0$ is a Hodges & Lehmann (1963) type of estimate of the intercept. If the Wilcoxon scores, (2·4), are used then

$$\hat{b}_0 = \operatorname{med}\{\tfrac{1}{2}(\hat{e}_i + \hat{e}_j)\} \quad (i \leqslant j).$$

Jaeckel (1972) showed that $\hat{b}$ has approximately a multivariate normal distribution with mean vector b and variance–covariance matrix $\tau^2(X_{1c}'X_{1c})^{-1}$. This is extended in Theorem 4·2 to the joint distribution of $(\hat{b}_0, \hat{b}')$.

The scale parameter τ must also be estimated. Let L and U be statistics which solve the equations $S^+(L) \doteq n^{\frac{1}{2}} z_\alpha$ and $S^+(U) \doteq -n^{\frac{1}{2}} z_\alpha$, where z_α is the upper α quantile of the standard normal distribution. Define

$$\hat{\tau} = n^{\frac{1}{2}}(U - L)/(2z_\alpha). \tag{2·7}$$

Under regularity conditions the authors (1976) show that $\hat{\tau} \to \tau$, in probability. As a small sample correction, Hettmansperger & McKean (1977) found in a Monte Carlo study that, when z_α is replaced by the $t_\alpha(n-r-1)$ critical point in the equations which determine L and U, the resulting estimates of τ had lower mean squared error. The same study indicated that values of α around 0·10 tended to give estimates of τ with lower mean squared error than did smaller, that is 0·05, or larger, 0·25, values.

2·2. *General linear hypotheses*

We shall consider general linear hypotheses about the vector b. As in most practical problems the intercept b_0 will be treated as a nuisance parameter. Following Searle (1971, p. 110) we shall consider hypotheses of the form

$$H_0\colon K'b = m \quad \text{versus} \quad H_A\colon K'b \neq m, \tag{2·8}$$

where K' is a known $q \times r$ matrix of full row rank and m is a known $q \times 1$ vector. Without loss of generality we can assume $m = 0$; see § 4. Define a new parameter $a_1 = K'b$. Then (Graybill, 1961, p. 242) we can reparameterize the model as

$$Y = 1b_0 + Za + e, \tag{2·9}$$

where a is partitioned as $a = (a_1', a_2')'$ and Z is an $n \times r$ matrix whose columns span the same space as the columns of X_1. In this formulation the hypotheses with a_2 unspecified, become

$$H_0\colon a_1 = 0 \quad \text{versus} \quad H_A\colon a_1 \neq 0. \tag{2·10}$$

We find it convenient at times to use this second formulation.

We call model (2·1) the full model. Let D_2 be the minimum value of the function D, (2·5), under the full model. We call model (2·1) subject to the hypothesis H_0, (2·8), the reduced model. Let D_1 be the minimum value of D under the reduced model. Let $\hat{\tau}$ be the full model estimate of scale given by (2·7). Define the statistic

$$F_R = (D_1 - D_2)/(q\tfrac{1}{2}\hat{\tau}). \tag{2·11}$$

As a test of the hypothesis H_0, McKean & Hettmansperger (1976) proposed rejecting H_0 for large values of F_R. Under H_0 and regularity conditions, they show that qF_R has an asymptotic chi-squared distribution with q degrees of freedom. The test statistic is consistent for H_A; see Theorem 4·3. In a Monte Carlo study Hettmansperger & McKean (1977) found it better as a small sample correction to compare F_R with a central $F(q, n-r-1)$ critical point. The empirical levels of the resulting test were quite close to the true levels; whereas, without this correction, the empirical levels were inflated.

3. The k-step analysis

Let $\hat{b}^{(0)}$ be an initial estimate of b; see § 5. The local quadratic approximation to D at $\hat{b}^{(0)}$, based on a Taylor expansion, is

$$Q^{(0)}(b) = D(\hat{b}^{(0)}) + \frac{\partial D(\hat{b}^{(0)})}{\partial b}(b-\hat{b}^{(0)}) + \tfrac{1}{2}(b-\hat{b}^{(0)})' H(\hat{b}^{(0)})(b-\hat{b}^{(0)}), \tag{3·1}$$

where H is the matrix of second partial derivatives of D. The first partial derivatives of D are $-S(b)$, where

$$S_j(b) = \sum_{i=1}^{n}(x_{ij}-\bar{x}_j)\,a\{R(Y_i-x_i'b)\} \quad (j = 1, \ldots, r). \tag{3·2}$$

The functions $S_j(b)$ are step functions in each b_i; hence, although the second derivatives are zero, the idea of slope persists. Numerical differentiation would not seem practical because of an increase in computation time due to further rankings. Using an asymptotic linearity result of Jureckova (1971a), S can be approximated by

$$S(b) \doteq -\hat{\tau}^{(0)-1}(X_{1c}'X_{1c})(b-\hat{b}^{(0)}) + S(\hat{b}^{(0)}), \tag{3·3}$$

where $\hat{\tau}^{(0)}$ is the estimator (2·7) based on the residuals $Y_i - x_i'\hat{b}^{(0)}$. With H replaced by $\hat{\tau}^{(0)-1}(X_{1c}'X_{1c})$ the resulting quadratic is similar to the one used by Jaeckel (1972) and the authors (1976) for obtaining asymptotic results.

The value which minimizes Q is given by (3·4) with $k = 1$. This estimate is similar to the one step estimate proposed by Kraft & van Eeden (1972), except the scale parameter here is simpler. It is analogous to Bickel's (1975) one step M estimate of type II.

Using $\hat{b}^{(1)}$ another estimate of τ can be formed which will lead to another quadratic (3·1) and, hence, a second step estimate. In general we define the k step estimate as

$$\hat{b}^{(k)} = \hat{b}^{(k-1)} + \hat{\tau}^{(k-1)}(X_{1c}'X_{1c})^{-1}S(\hat{b}^{(k-1)}). \tag{3·4}$$

Based on the residuals $Y_i - x_i'\hat{b}^{(k)}$ the parameter τ and the intercept b_0 can be estimated as in § 2. Denote these by $\hat{\tau}^{(k)}$ and $\hat{b}_0^{(k)}$.

The k step test statistic of H_0, (2·8), is defined as

$$F_R^{(k)} = \frac{(D_1^{(k)} - D_2^{(k)})/q}{\frac{1}{2}\hat{\tau}^{(k)}},$$

where $D_1^{(k)}$ and $D_2^{(k)}$ would be the values of D at the k step estimates in the reduced and full models.

4. Asymptotic distribution theory

The asymptotic theory for the rank analysis has been found by Jaeckel (1972) and McKean & Hettmansperger (1976). In this section we extend these results to the k step analysis. The consistency of the test statistic and the asymptotic joint distribution of $(\hat{b}_0, \hat{b}')'$ are new results for the rank analysis. This latter result, which is an extension of Adichie's (1967) work on the simple linear model, is useful for obtaining estimates and confidence intervals for predicted values and other linear combinations of the parameters.

We need some additional assumptions. Assume the density f has finite Fisher information. The design matrix should satisfy the assumptions of Jureckova (1971a). Among these are that $n^{-1}X_{1c}'X_{1c}$ converges to a positive-definite matrix. Assume that for the initial estimates, $n^{\frac{1}{2}}(\hat{b}^{(0)} - b)$ is bounded in probability. Finally, assume the following equivariance (Bickel, 1975) assumptions on the initial estimates:

$$\hat{b}^{(0)}(Y + b_0^* + X_1 b^*) = \hat{b}^{(0)}(Y) + b^*, \quad \hat{b}^{(0)}(cY) = c\hat{b}^{(0)}(Y). \tag{4·1}$$

It follows that $\hat{\tau}^{(0)}$ is scale equivariant, that is, $\hat{\tau}^{(0)}(cY + b_0^* + X_1 b) = |c|\hat{\tau}^{(0)}(Y)$; and, hence, the k step estimates also satisfy (4·1).

One consequence of these invariance properties and the symmetry of the error distribution is the analogue to unbiasedness found in Theorem 4·1. The asymptotic properties of the k step estimates are found in Theorem 4·2, which is proved in the Appendix.

THEOREM 4·1. *For $j = 1, \ldots, r$ the k step estimators $\hat{b}_0^{(k)}$ and $\hat{b}_j^{(k)}$ are symmetrically distributed about b_0 and b_j. Further, the scale estimate $\hat{\tau}^{(k)}$ is uncorrelated with each of $\hat{b}_0^{(k)}$ and $\hat{b}_j^{(k)}$.*

THEOREM 4·2.

(a) *For $k \geqslant 0$, $\hat{\tau}^{(k)} \to \tau$, in probability.*

(b) *For $k \geqslant 1$, $n^{\frac{1}{2}}(\hat{b}^{(k)} - \hat{b}) \to 0$, in probability.*

(c) *For $k \geqslant 1$, $(\hat{b}_0^{(k)}, \hat{b}^{(k)\prime})'$ is asymptotically distributed as the multivariate normal distribution* MVN $\{(b_0, b')', \tau^2(X'X)^{-1}\}$.

For asymptotic efficiency results of the test statistic, the second formulation of the linear model, (2·9), is more convenient. We need the sequence of alternative hypotheses: H_n: $a_n = (n^{-\frac{1}{2}}\Theta', a_2')'$, where $\Theta \neq 0$.

THEOREM 4·3. *For $k \geqslant 1$,*

(a) *under H_0, $qF_R^{(k)}$ has an asymptotic central chi-squared distribution with q degrees of freedom;*

(b) *$qF_R^{(k)}$ is a consistent test statistic for H_A, (2·10);*

(c) *under H_n, $qF_R^{(k)}$ has an asymptotic noncentral chi-squared distribution with q degrees of freedom and the same noncentrality parameter as qF_R (McKean & Hettmansperger, 1976).*

Part (b) follows essentially from the consistency of the full model estimates, see J. W. McKean's Pennsylvania State University thesis. The proofs of (a) and (c) are similar to the corresponding results in the fully iterated analysis (McKean & Hettmansperger, 1976).

If we assume var $(e_i) = \sigma^2 < \infty$, then by part (c) of Theorem 4·3 the asymptotic relative efficiency of $F_R^{(k)}$ compared to the classical least squares F test is σ^2/τ^2. If Wilcoxon scores (2·4) are used this results in the well-known value for the efficiency, and, in particular, if the underlying error distribution is normal this efficiency is 0·955. Hence there is little efficiency lost if the rank analysis is used rather than least squares. On the other hand, if the error distribution has tails heavier than the normal or is prone to outliers the rank analysis can be much more efficient than least squares.

5. STARTING VALUES AND A MONTE CARLO STUDY

For extremely heavy-tailed error structure the first few step estimates starting from least squares may be far from the fully iterated estimates. In the general regression problem further research is needed on resistant quick starting values. The work of Hinich & Talwar (1975) may prove useful in this context.

For many practical designs convenient starting values can be constructed. For example, in a two-way layout with replicates, linear combinations of cell medians can be used as starting values. These proved effective in the Monte Carlo study below when the errors followed a Cauchy distribution. If there are no replicates and we further assume no interaction then appropriate linear combinations of row and column medians, i.e. replacing the row and column means, would seem to make good starting values, although these are not equivariant as defined in § 4. In the following example we wanted to compare the effect starting values had on the k step estimates with data as extreme as Cauchy.

Example. We considered a 4×6 two-way layout with no replicates. The observations, given in Table 1, were obtained from the model $y_{ij} = \mu_{ij} + e_{ij}$, where the errors e_{ij} have the Cauchy distribution $f(x) = \{\pi(1+x^2)\}^{-1}$. The parameters were all zero, except that $\mu_{i2} = 6$ for all i. Thus there were column effects but no row effects.

Table 1. *The data for the example*

i	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$
1	1·46	6·33	−0·03	0·06	0·98	−0·27
2	−2·15	2·95	−0·46	0·88	10.53	7·25
3	−4·90	8·44	−1158·9	2·38	0·23	0·31
4	−1·54	5·89	−0·72	−1·89	0·20	−39·32

Wilcoxon scores were used. Both least squares and combinations of row and column medians were used as starting values. For the example and the Monte Carlo study the statistic $\hat{\tau}$, (2·7), was computed using the value of $\alpha = 0{\cdot}10$, see §2·1. For the full model, Fig. 1 depicts the values of the function D at each set of initial estimates, the resulting steps 1–5 and the minimum value of D. It is clear from the slopes of the graphs that the least squares starting values are poor while the median starting values are quite stable. A summary of test statistics is found in Table 2.

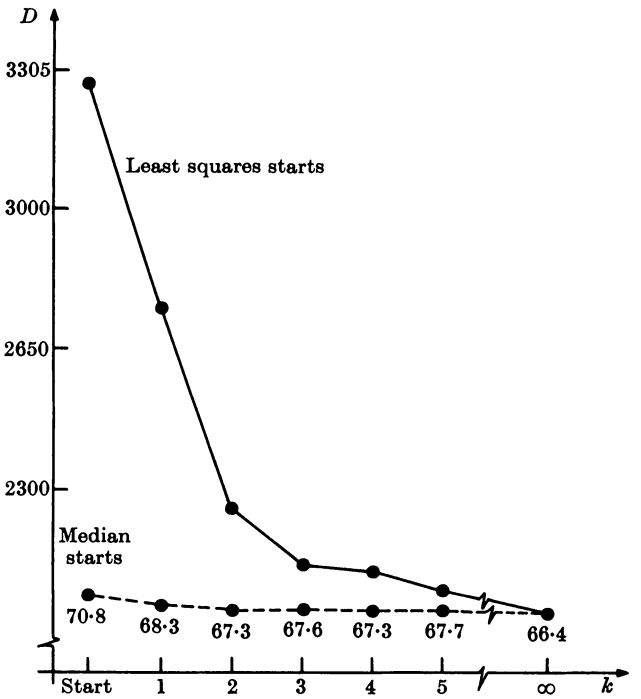

Fig. 1. Evaluation of the function D for the example. The solid line denotes the values of D at least squares, steps 1–5, and the minimum value of D. The broken line shows the corresponding values of D with median starting values. The numbers shown are for $k = 1, \ldots, 5$, the last 3 digits of the values of D for the median starts and for $k = \infty$, the last 3 digits of the minimum value of D.

Note from Fig. 1 that the graph based on median starting values begins to oscillate from the second to third steps. The analysis based on the second step is closer to the fully iterated analysis. Generally we suggest stopping when this oscillation occurs or if the fully iterated estimates are desired we suggest leaving this Gauss–Newton routine and proceeding with a

first derivative routine such as steepest descent. This example offers encouragement that with resistant robust starting values in a few steps the k step estimates are close to the fully iterated estimates in highly nonnormal situations.

Table 2. *Values of test statistics for hypotheses of the example*

	Hypotheses	
	No row effect; degrees of freedom (3, 15)	No column effect; degrees of freedom (5, 15)
F (least squares)	0·98	1·00
$F_R^{(1)}$ (median starts)	1·68	3·02
$F_R^{(2)}$ (median starts)	0·98	3·05
F_R (fully iterated)	0·96	3·18

A small Monte Carlo study was made to compare the analyses of a linear model based on a few iterations with that of full iteration. The model simulated was the 2×2 two-way layout

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk} \quad (i, j = 1, 2;\ k = 1, \ldots, n_{ij}).$$

The design was unbalanced with cell sizes $n_{11} = n_{22} = 8$ and $n_{12} = n_{21} = 5$. We performed 440 simulations of this model for each of three distributions: the logistic, $f(x) = e^{-x}(1+e^{-x})^{-2}$; the double exponential, $f(x) = \frac{1}{2}e^{-|x|}$; and the Cauchy $f(x) = \{\pi(1+x^2)\}^{-1}$. Only the null model, all parameters zero, was considered.

Table 3. *Empirical means, $\hat{\mu}$, empirical standard deviations, $\hat{\sigma}$, and asymptotic standard deviations, σ, for predicted mean of cell* (1, 1) *in Monte Carlo study*

	Logistic			Double exponential			Cauchy		
	$\hat{\mu}$	$\hat{\sigma}$	σ	$\hat{\mu}$	$\hat{\sigma}$	σ	$\hat{\mu}$	$\hat{\sigma}$	σ
Least squares	0·026	0·654	0·641	0·019	0·513	0·500	—	—	—
One step	0·018	0·628	0·612	0·010	0·447	0·408	0·021	0·794	0·641
Two step	0·022	0·627	0·612	0·013	0·445	0·408	0·026	0·792	0·641
Fully iterated	0·023	0·629	0·612	0·016	0·442	0·408	0·027	0·806	0·641

Least squares estimate not obtained for median starts; see text.

One step values based on least squares starts for logistic and double exponential, median starts for Cauchy.

Table 4. *Empirical means $\hat{\mu}$, empirical standard deviation, $\hat{\sigma}$, and empirical* 0·10 *and* 0·05 *levels of the test statistics for interaction*; 1 *and* 22 *degrees of freedom*

	Logistic				Double exponential				Cauchy			
	$\hat{\mu}$	$\hat{\sigma}$	0·10	0·05	$\hat{\mu}$	$\hat{\sigma}$	0·10	0·05	$\hat{\mu}$	$\hat{\sigma}$	0·10	0·05
LS	1·13	1·61	0·109	0·045	1·13	1·57	0·113	0·047	1·15	1·89	0·115	0·065
$F_R^{(1)}$	1·12	1·60	0·095	0·052	1·09	1·55	0·100	0·050	1·12	1·65	0·100	0·052
$F_R^{(2)}$	1·12	1·59	0·100	0·052	1·11	1·55	0·095	0·052	1·12	1·61	0·097	0·054
F_R	1·14	1·60	0·097	0·052	1·13	1·54	0·102	0·050	1·11	1·60	0·109	0·050

LS, the classical least squares F test for logistic and double exponential; F_R evaluated at median starting values for Cauchy.

$F_R^{(1)}$ and $F_R^{(2)}$ based on least squares starts for logistic and double exponential, median starts for Cauchy.

Wilcoxon scores were used for the rank analyses. Least squares starting values were used for the logistic and double exponential, while linear combinations of cell medians were used for the Cauchy distribution.

Table 3 displays the empirical means and standard deviations of the full model predicted cell mean, median in the Cauchy case, for cell (1, 1). The estimate of this parameter is a linear combination of the intercept a_0 and a for a full rank reparameterized model such as (2·9).

Since a_0 is not needed to obtain the k step estimates this estimate was not found for the median starts. We might note that for the parameters estimated the median starts had, for the Cauchy case, smaller standard deviations than the respective Wilcoxon estimates. This is expected for Cauchy errors. For comparison the asymptotic standard deviations from Theorem 4·2 are also given. Note that the ratio of sample size to number of parameters that are estimated is $26/4 = 6{\cdot}5$. For such low ratios and in the case of the Cauchy distribution, the discrepancy between the asymptotic and empirical standard deviations is similar to what Huber (1973, p. 819) found for his M-estimates. In general the two step procedure seemed quite satisfactory, especially for such small samples. Sign rather than Wilcoxon scores would be more natural to use for the double exponential case.

Evidence which supports the small sample correction of treating the test statistics $F_R^{(k)}$ and F_R as having a central F distribution under H_0 can be found in Table 4. The empirical means are close to 1·10 the mean of a central $F(1, 22)$ distribution. Also the 0·05 and 0·10 empirical levels seem to be close to the true levels.

For the fully iterated analysis the drop in dispersion, the numerator of F_R, is nonnegative. For the k step procedures this drop in dispersion was not constrained to be positive; hence in this study of the null model a small percentage of these drops were negative. This accounts for the slightly lower mean of the step test statistics. Note that the empirical levels are about the same.

These preliminary simulations along with more extensive examples run on real data indicate the k step procedures can provide a computationally feasible, robust analysis of large data sets.

This research was partially supported by National Science Foundation grants. The authors wish to thank the referees for their many helpful comments.

Appendix

Proof of Theorem 4·2

Because of the equivariant properties, we can assume that the true parameters (b_0, b) are $(0, 0)$. It suffices to prove the results for $k = 1$.

Since $\sqrt{n}$ times the initial estimates are bounded in probability it follows as in § 3 of McKean & Hettmansperger (1976) that $\hat{\tau}^{(0)} \to \tau$, in probability. As an approximation of $\hat{b}^{(1)}$, define $\dot{b} = \tau(X_{1c}' X_{1c})^{-1} S(0)$; then

$$n^{\frac{1}{2}}(\hat{b}^{(1)} - \dot{b}) = n(X_{1c}' X_{1c})^{-1}\{\hat{\tau}^{(0)} n^{-\frac{1}{2}} S(\hat{b}^{(0)}) - n^{-\frac{1}{2}} \tau S(0) + (n^{-1} X_{1c} X_{1c}) n^{\frac{1}{2}} \hat{b}^{(0)}\}.$$

Let $\Sigma = \lim n^{-1}(X_{1c}' X_{1c})$. Since $n^{\frac{1}{2}} \hat{b}^{(0)}$ is bounded in probability it follows from Jureckova's (1971a) asymptotic linearity result that $n^{\frac{1}{2}}(\hat{b}^{(1)} - \dot{b}) \to 0$ in probability. Further from Jureckova it follows that $n^{-\frac{1}{2}} S(0)$ is asymptotically distributed as MVN $(0, \Sigma)$. Thus $\hat{b}^{(1)}$ has the desired asymptotic distribution which in turn implies $\hat{\tau}^{(1)} \to \tau$ in probability; hence (a) holds. Jaeckel (1972) showed that $n^{\frac{1}{2}}(\hat{b} - \dot{b}) \to 0$ in probability; hence (b) holds.

We lack only the joint asymptotic distribution of $(\hat{b}_0^{(1)}, \hat{b}^{(1)})'$. This is an extension of Adichie's (1967) result. Define

$$\ddot{b} = \tau(n^{-1} X_{1c}' X_{1c})^{-1} n^{-1} \ddot{S}(0),$$

where the jth component of $\ddot{S}$ is

$$\ddot{S}_j(0) = \Sigma(x_{ij} - \bar{x}_j)\, \phi\{F(Y_i)\}.$$

Using results of Hájek & Šidák (1967, pp. 160, 164) we can show that $n^{\frac{1}{2}}\{S_j(0) - \ddot{S}_j(0)\} \to 0$, in probability. Hence, $n^{\frac{1}{2}}(\hat{b}^{(1)} - \ddot{b}) \to 0$ in probability.

Reparameterize the model as

$$Y = 1d + (X_1 - \bar{X}_1)b + e.$$

The one sample process corresponding to this model is

$$Z^*(d) = \sum_{i=1}^{n} a^+\{R|Y_i - d - (x_i - \bar{x})'\hat{b}^{(1)}|\}\operatorname{sgn}\{Y_i - d - (x_i - \bar{x})'\hat{b}^{(1)}\}.$$

Let $\hat{d}$ be the corresponding Hodges–Lehmann type estimate, that is, $Z^*(\hat{d}) \doteq 0$. Then $\hat{d} = \hat{a}^{(1)} + \bar{x}'\hat{b}^{(1)}$. From Jureckova (1971b) it follows that $n^{\frac{1}{2}}\hat{d}$ is bounded in probability. From the asymptotic linearity result of § 3 of McKean & Hettmansperger (1976), $n^{\frac{1}{2}}(\dot{d} - \hat{d}) \to 0$, in probability, where

$$\dot{d} = n^{-1}\tau\sum_{i=1}^{n} a^+(R|Y_i|)\operatorname{sgn}(Y_i).$$

Finally, let

$$\ddot{d} = n^{-1}\tau\sum_{i=1}^{n} \phi^+(U_i^+)\operatorname{sgn}(Y_i),$$

where $U_i^+ = F^+(|Y_i|)$, F^+ is the distribution function of $|Y|$. Once more referring to Hájek & Šidák (1967, p. 166) we have $n^{\frac{1}{2}}(\dot{d} - \ddot{d}) \to 0$ in probability. Thus $n^{\frac{1}{2}}(d - \ddot{d}) \to 0$, in probability. But by the symmetry of f and the relationship between the scores ϕ and ϕ^+ we get $\ddot{d} = n^{-1}\tau\Sigma\phi\{F(Y_i)\}$. From this and the conditions we imposed on the design we can show using the Lindeberg–Feller central limit theorem that

$$n^{\frac{1}{2}}\tau^{-1}(\ddot{d}, \Sigma b) \to \text{MVN}\left(0, \begin{bmatrix} 1 & 0 \\ 0 & \Sigma \end{bmatrix}\right)$$

in distribution, and from this obtain part (c).

References

Adichie, J. N. (1967). Estimates of regression parameters based on rank tests. *Ann. Math. Statist.* **38**, 894–904.

Andrews, D. F. (1974). A robust method for multiple linear regression. *Technometrics* **16**, 523–31.

Bickel, P. J. (1975). One-step Huber estimates in the linear model. *J. Am. Statist. Assoc.* **70**, 428–34.

Graybill, F. A. (1961). *An Introduction to Linear Statistical Models*, Vol. 1. New York: McGraw-Hill.

Hájek, J. & Šidák, Z. (1967). *Theory of Rank Statistics*. New York: Academic Press.

Hettmansperger, T. P. & McKean, J. W. (1977). A robust alternative based on ranks to least squares in analyzing linear models. *Technometrics* **19**, 275–84.

Hinich, M. J. & Talwar, P. (1975). A simple method for robust regression. *J. Am. Statist. Assoc.* **70**, 113–9.

Hodges, J. L. & Lehmann, E. L. (1963). Estimation of location based on rank tests. *Ann. Math. Statist.* **34**, 598–611.

Huber, P. J. (1972). Robust statistics: A review. *Ann. Math. Statist.* **43**, 1041–67.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1**, 799–821.

Jaeckel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.* **43**, 1449–58.

Jureckova, J. (1971a). Nonparametric estimate of regression coefficients. *Ann. Math. Statist.* **42**, 1328–38.

Jureckova, J. (1971b). Asymptotic independence of rank test statistic for testing symmetry on regression. *Sankhyā* A **33**, 1–18.

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

McKean, J. W. & Hettmansperger, T. P. (1976). Tests of hypothesis based on ranks in the general linear model. *Comm. Statist.* A **5**, 693–709.

Searle, S. R. (1971). *Linear Models*. New York: Wiley.

[*Received August* 1976. *Revised March* 1978]