The Annals of Statistics
1986, Vol. 14, No. 3, 1246–1251

BAHADUR REPRESENTATIONS FOR ROBUST SCALE ESTIMATORS BASED ON REGRESSION RESIDUALS

BY A. H. WELSH

University of Chicago

We investigate the asymptotic behaviour of the median deviation and the semi-interquartile range based on the residuals from a linear regression model by deriving weak asymptotic representations for the estimators. These representations may be used to obtain a variety of central limit theorems and yield conditions under which the median deviation and the semi-interquartile range are asymptotically equivalent. The results justify the use of the estimators as concommitant scale estimators in the general scale equivariant M-estimation of a regression parameter problem. Finally, the results contain as a special case those obtained by Hall and Welsh (1985) for independent and identically distributed random variables.

1. Introduction. In this paper, we investigate the asymptotic properties of two popular robust scale estimators, the median absolute deviation from the median (sometimes called MAD or, at least since Hampel (1974), the median deviation) and the semi-interquartile range, applied to the residuals from a linear regression model. An important (but not the only) motivation is the problem of concommitant scale estimation in M-estimation.

Suppose that we observe $Y_1, \dots, Y_n$ where

$$Y_j = x_j'\theta_0 + e_j, \qquad 1 \le j \le n, \tag{1.1}$$

with $\{x_j = (x_{j1}, \dots, x_{jp})'\}$ a sequence of known p-vectors $(p \ge 1)$, θ_0 a unique unknown regression parameter to be estimated, and $\{e_j\}$ a sequence of independent and identically distributed random variables with unknown distribution function F. Relles (1968) and Huber (1973) investigated the class of M estimators of the regression parameter θ_0 as solutions of equations of the form

$$\sum_{j=1}^{n} x_j \psi(Y_j - x_j'\theta) = 0,$$

where ψ: $\mathbf{R} \to \mathbf{R}$. In general, scale equivariant M-estimators may be obtained by calculating a location invariant and scale equivariant scale estimator σ_n from the data and then solving the system of equations

$$\sum_{j=1}^{n} x_j \psi\big((Y_j - x_j'\theta)/\sigma_n\big) = 0.$$

Huber (1964) made three proposals for obtaining a suitable scale estimator σ_n.

Received April 1985; revised October 1985.

AMS 1980 *subject classifications*. Primary 62F35, 60F05; secondary 62G30.

Key words and phrases. Linear regression, median deviation, quantiles, robust estimation, scale estimation, semi-interquartile range.

The asymptotic theory of the estimators resulting from proposals 1 and 2 may be derived from the results of Huber (1967); proposal 3 has proved efficacious for a particular M-estimator in the regression problem (see Welsh (1985) for references) and has been investigated for M-estimators in the location subproblem by Bell (1980). Another conceptually and computationally simple approach is to apply an explicit robust scale estimator to the residuals. This procedure is frequently adopted in the location subproblem for which the median is a natural initial estimator. The small sample performance of the resulting regression M-estimators has been investigated by Holland and Welsch (1977), Hill and Holland (1977), and Denby and Mallows (1977). The asymptotic results of Bickel (1975) and Yohai and Maronna (1979), which implicitly assume that F is symmetric, are applicable provided there exists a scale estimator σ_n such that $n^{1/2}(\sigma_n - \sigma_0)$ is bounded in probability for some $\sigma_0 > 0$. However, the results are incomplete in that no robust scale estimator satisfying the above requirement has been exhibited. We will show in this paper that under very mild conditions, the median deviation and the semi-interquartile range are appropriately bounded in probability.

Our main result (Theorem 2) yields a central limit theorem for the median deviation in the general asymmetric case and establishes that (Corollary 2.1) under mild conditions implied by symmetry, the semi-interquartile range is asymptotically equivalent to the median deviation.

2. Results. We assume throughout that the model (1.1) holds so that, for any $\theta \in \mathbf{R}^p$, the residuals from θ are

$$e_j(\theta) = Y_j - x_j'\theta = e_j - x_j'(\theta - \theta_0), \qquad 1 \le j \le n,$$

almost surely, and $e_j(\theta_0) = e_j$, $1 \le j \le n$, almost surely. Under mild smoothness conditions on F, the results below hold for any initial regression parameter estimator θ_n such that

(Ci) $\qquad n^{1/2}(\theta_n - \theta_0)$ is bounded in probability,

provided that

(Cii) $x_{j1} = 1$, $1 \le j \le n$ and there exists a positive definite matrix Δ such that $\lim_{n\to\infty} \mathrm{n}^{-1} \sum_{j=1}^{n} x_j x_j' = \Delta$

holds. Condition (Ci) holds for a large class of estimators including M-estimators. It is convenient but not necessary to assume that θ_n estimates an intercept; with $z_j' = (1, x_j')$, $1 \le j \le n$, instead of x_j, $1 \le j \le n$ in condition (Cii), the results below still hold. For examples of estimators which do not estimate an additive main effect, see Jaeckel (1972) and Welsh (1985).

We derive weak Bahadur representations for the median deviation and the semi-interquartile range through the sample quantiles. For $\theta \in \mathbf{R}^p$ in the neighborhood of θ_0, let

$$F_n(x,\theta) = n^{-1}\sum_{j=1}^{n} I\{e_j(\theta) \le x\}, \qquad x \in \mathbf{R},$$

where I is the indicator function. For simplicity, set $F_n(x, \theta_0) = F_n(x)$. The qth sample quantile $\xi_{nq}(\theta)$ is defined by

$$F_n(\xi_{nq}(\theta), \theta) = q,$$

and the population quantile ξ_q is defined by

$$F(\xi_q) = q, \qquad 0 < q < 1.$$

The first result which is of independent interest, slightly generalises Lemma 1 of Ruppert and Carroll (1980) by weakening their first condition. The result may be proved by an extension of the argument of Ghosh (1971) with Lemma 4.1 of Bickel (1975) or using results of Pierce and Kopecky (1979) or Loynes (1980) or by modifying the argument of Ruppert and Carroll (1980). The proof is omitted.

THEOREM 1. *Suppose that conditions* C *hold and the derivative of F exists in a neighborhood of ξ_q, is continuous at ξ_q, and $F'(\xi_q) > 0$. Then for fixed q,* $0 < q < 1$,

$$n^{1/2}\{\xi_{nq}(\theta_n) - \xi_q\} + n^{1/2}\frac{\{F_n(\xi_q) - q\}}{F'(\xi_q)} + n^{1/2}\bar{x}'(\theta_n - \theta_0) \to_P 0.$$

In the special case that θ_0 is known, conditions C are redundant and the smoothness condition can be weakened to yield the result of Ghosh (1971). The representation in Theorem 1 is useful in determining the joint asymptotic distribution of any finite number of fixed quantiles, possibly in conjunction with other statistics. In particular, we immediately obtain a Bahadur representation for the semi-interquartile range.

COROLLARY 1.1. *Suppose that conditions* C *hold and that the smoothness condition of Theorem* 1 *holds for $q = 3/4$ and $q = 1/4$. Let $Q_n(\theta_n) = \{\xi_{n,3/4}(\theta_n) - \xi_{n,1/4}(\theta_n)\}/2$* and $Q_0 = \{\xi_{3/4} - \xi_{1/4}\}/2$. *Then*

$$n^{1/2}\{Q_n(\theta_n) - Q_0\} + n^{1/2}\frac{F_n(\xi_{3/4}) - 3/4}{2F'(\xi_{3/4})} - \frac{F_n(\xi_{1/4}) - 1/4}{2F'(\xi_{1/4})} \to_P 0.$$

The next result, the Bahadur representation for the median deviation, is the main result of this paper. The sample median deviation $S_n(\theta_n, \xi_{n,1/2}(\theta_n))$ is the median of $|e_j(\theta_n) - \xi_{n,1/2}(\theta_n)|, 1 \le j \le n$, while the population median deviation $s_0 > 0$ is defined by $F(\xi_{1/2} + s_0) - F(\xi_{1/2} - s_0 -) = \frac{1}{2}$.

THEOREM 2. *Suppose that conditions* C *hold and*

(i) *$F'(\xi_{1/2} + x)$ exists for x in a neighborhood of the origin and is continuous and positive at $x = 0$;*

(ii) *$F'(\xi_{1/2} \pm s_0 + x)$ exists for x in a neighborhood of the origin and is continuous at $x = 0$;*

(iii) *$F'(m_0 + x) + F'(m_0 - x) > 0$ for x in a neighborhood of s_0.*

Then

$$\begin{aligned} &n^{1/2}\{S_n(\theta_n, \xi_{n,1/2}(\theta_n)) - s_0\} \\ &\quad + n^{1/2}\{F_n(\xi_{1/2} + s_0) - F_n(\xi_{1/2} - s_0 -) - \tfrac{1}{2}\}/g_1 \\ &\quad - n^{1/2}\{F_n(\xi_{1/2}) - \tfrac{1}{2}\}\{g_2/g_1F'(\xi_{1/2})\} \to_P 0, \end{aligned}$$

where$g_1 = F'(\xi_{1/2} + s_0) + F'(\xi_{1/2} - s_0)$*and*$g_2 = F'(\xi_{1/2} + s_0) - F'(\xi_{1/2} - s_0)$.

OUTLINE OF PROOF. Let $\tau_n' = (\theta_{n1} + \xi_{n,1/2}(\theta_n), \theta_{n2}, \dots, \theta_{np})$ and $\tau_0' = (\theta_{01} + \xi_{1/2}, \theta_{02}, \dots, \theta_{0p})$ so the median deviation is the median of $|Y_j - x_j'\tau_n|$, $1 \le j \le n$, and $S_n(\theta_n, \xi_{n,1/2}(\theta)) = S_n(\tau_n)$. For $\tau \in \mathbf{R}^p$ in a neighborhood of τ_0, put

$$\overline{G}_n(z, \tau) = n^{-1} \sum_{j=1}^{n} \{F(x_j'\tau + z) - F(x_j'\tau - z -)\}, \qquad z \ge 0.$$

By condition (iii), for τ in a neighborhood of τ_0, $s_n(\tau)$ defined by $\overline{G}_n(s_n(\tau), \tau) = \frac{1}{2}$ is unique and $s_n(\tau_0) = s_0$. Now

$$n^{1/2}\{S_n(\tau_n) - s_0\} = n^{1/2}\{S_n(\tau_n) - s_n(\tau_n)\} + n^{1/2}\{s_n(\tau_n) - s_0\}.$$

The first term may be handled by a modified version of the proof of Theorem 3 of Hall and Welsh (1985) with Lemma 4.1 of Bickel (1975), while the second term may be handled by a one-term Taylor series expansion since $s_n(\tau)$ is differentiable at τ_0 and $s_n'(\tau_0) = -(g_2/g_1)\bar{x}$ at τ_0. □

If (1.1) does not include an intercept, the above argument applies with $\tau_n' = (\xi_{n,1/2}(\theta_n), \theta_n') \in \mathbf{R}^{p+1}$, $\tau_0' = (\xi_{1/2}, \theta_0') \in \mathbf{R}^{p+1}$, and x_j replaced by z_j, $1 \le j \le n$.

If the regression parameter θ_0 is known, the result generalises Corollary 3.1 of Hall and Welsh (1985) by discarding their fourth condition. The resulting Bahadur representation provides an alternative derivation of the central limit theorem (Theorem 2) of Hall and Welsh. For the present problem, the conditioning argument used by Hall and Welsh to prove their central limit theorem is of limited utility because conditioning on the estimate of the regression parameter does not simplify the structure of the problem.

Combining Corollary 3.1 and Theorem 2 we obtain the following analogue of Corollary 3.1.1 of Hall and Welsh (1985).

COROLLARY 2.1. *Suppose that in addition to the conditions of Theorem 2, $1 - F(\xi_{1/2} + s_0) = F(\xi_{1/2} - s_0 -)$ and $F'(\xi_{1/2} + s_0) = F'(\xi_{1/2} - s_0)$. Then*

$$n^{1/2}\{S_n(\theta_n, \xi_{n,1/2}(\theta_n)) - Q_n(\theta_n)\} \to_P 0.$$

The above conditions hold if F is symmetric, has connected support, and a positive, continuous density on its support.

If we include an intercept in (1.1) and assume that $\xi_{1/2} = 0$ so the underlying error distribution F is centered about the origin, we may consider the alternative scale estimator $R_n(\theta_n)$, the median of $|e_j(\theta_n)|, 1 \le j \le n$. Note that the estimator $R_n(\theta_n)$ is only location invariant if θ_n includes an intercept estimator but $Q_n(\theta_n)$ and $S_n(\theta_n, \xi_{n,1/2}(\theta_n))$ are location invariant whether or not θ_n includes an intercept estimator. Specifically, if $\theta_n = (\alpha_n, \beta_n')'$, $\alpha_n \in \mathbf{R}$ an intercept estimator, and $\beta_n \in \mathbf{R}^{p-1}$, then $Q_n(\theta_n) = Q_n(\beta_n)$ and $S_n(\theta_n, \xi_{n,1/2}(\theta_n)) = S_n(\beta_n, \xi_{n,1/2}(\beta_n))$; if α_n is the median of $Y_j - x_j'\beta_n, 1 \le j \le n$, then $R_n(\theta_n) = S_n(\theta_n, \xi_{n,1/2}(\theta_n))$. If the conditions of Theorem 2 hold, then, by the same argument as that used to derive Theorem 2, it follows that

$$n^{1/2}\{R_n(\theta_n) - r_0\} + n^{1/2}\frac{\{F_n(r_0) - F_n(-r_0) - 1/2\}}{F'(r_0) + F'(-r_0)}$$

$$+ \frac{\{F'(r_0) - F'(-r_0)\}}{F'(r_0) + F'(-r_0)} n^{1/2}\bar{x}'\theta_n \to_P 0,$$

where $r_0 > 0$ is defined by $F(r_0) - F(-r_0) = \frac{1}{2}$. Moreover, it then follows that if in addition F is symmetric about the origin, $S_n(\theta_n, \xi_{n,1/2}(\theta_n))$, $R_n(\theta_n)$ and $Q_n(\theta_n)$ are all asymptotically equivalent.

Finally, many regression parameter estimators (including least squares, least absolute deviations, and other M estimators) admit a representation of the form

$$n^{1/2}(\theta_n - \theta_0) - n^{-1/2}\sum_{j=1}^{n} \Delta^{-1}x_j\gamma(e_j) \to_P 0,$$

for some real function γ. It is straightforward to use such a representation and the results of the present paper to obtain central limit theorems for $\xi_{nq}(\theta_n), 0 < q < 1, (\theta_n', S_n(\theta_n, \xi_{n,1/2}(\theta_n)))', (\xi_{n,1/2}(\theta_n), S_n(\theta_n, \xi_{n,1/2}(\theta_n)))'$, etc.

Acknowledgments. I am grateful to David Pollard and Stephen Stigler for helpful conversations and to a referee and Associate Editor for comments which improved the presentation of this paper.

REFERENCES

BAHADUR, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–580.

BELL, R. M. (1980). An adaptive choice of the scale parameter for M-estimators. Technical Report, Stanford Univ.

BICKEL, P. J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70** 428–434.

DENBY, L. and MALLOWS, C. L. (1977). Two diagnostic displays for robust regression analysis. *Technometrics* **19** 1–13.

GHOSH, J. K. (1971). A new proof of the Bahadur representation of quantiles and an application. *Ann. Math. Statist.* **42** 1957–1961.

HALL, P. and WELSH, A. H. (1985). Limit theorems for the median deviation. *Ann. Inst. Statist. Math.* **37** 27–36.

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–397.

HILL, R. W. and HOLLAND, P. W. (1977). Two robust alternatives to least squares regression. *J. Amer. Statist. Assoc.* **72** 828–833.

HOLLAND, P. W. and WELSCH, R. E. (1977). Robust regression using iteratively reweighted least squares. *Commun. Statist. A* **6** 813–827.

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

HUBER, P. J. (1967). The behaviour of maximum likelihood estimates under non-standard conditions. In *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 221–233. Univ. California Press.

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** 1449–1458.

LOYNES, R. M. (1980). The empirical distribution function of residuals from generalized regression. *Ann. Statist.* **8** 285–298.

PIERCE, D. A. and KOPECKY, K. J. (1979). Testing goodness of fit for the distribution of errors in regression models. *Biometrika* **66** 1–5.

RELLES, D. (1968). Robust regression by modified least squares. Ph.D. thesis, Yale Univ.

RUPPERT, D. and CARROLL, R. J. (1980). Trimmed least squares estimation in the linear model. *J. Amer. Statist. Assoc.* **75** 828–838.

WELSH, A. H. (1985). An angular approach for linear data. *Biometrika* **72** 441–450.

YOHAI, V. J. and MARONNA, R. A. (1979). Asymptotic behaviour of M-estimators for the linear model. *Ann. Statist.* **7** 258–268.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
5734 SOUTH UNIVERSITY AVENUE
CHICAGO, ILLINOIS 60637