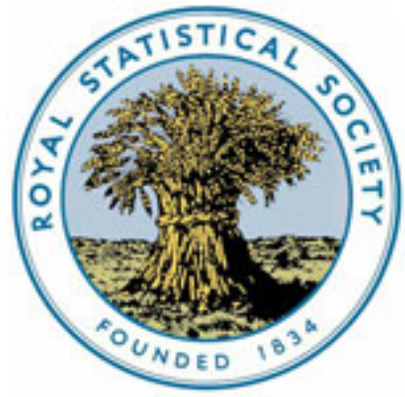



*J. R. Statist. Soc.* B (1994)
**56**, *No.* 1, *pp.* 209–220

# Bounded Influence Rank Regression

By J. D. NARANJO† and T. P. HETTMANSPERGER

*Western Michigan University, Kalamazoo, USA* *Penn State University, University Park, USA*

[Received March 1991. Revised December 1992]

SUMMARY

When $\epsilon_i = y_i - x_i'\beta$, it is known that minimizing $\Sigma\Sigma|\epsilon_i - \epsilon_j|$ yields an estimate of regression that attains a bounded *influence of the residual* with 95% efficiency for the normal distribution. We show that introducing weights $\Sigma\Sigma b_{ij}|\epsilon_i - \epsilon_j|$ achieves bounded total influence with positive breakdown. Mallows weights in particular are optimally efficient under a predefined bound on the gross error sensitivity. A generalization of Mallows weights allows additional local stability against high leverage points. Two numerical examples illustrate the behaviour of the estimate.

*Keywords*: BOUNDED INFLUENCE; BREAKDOWN; EFFICIENCY; MALLOWS WEIGHTS; RANK ESTIMATE; REGRESSION

## 1. INTRODUCTION

Consider the linear regression model $y_i = \alpha + x_i'\beta + \epsilon_i$, $i = 1, \ldots, n$, where $x_i$ is a $p \times 1$ vector of explanatory variables. In estimating $\beta$, the sensitivity of least squares estimates to the effect of outlying observations has led to the development of more robust fitting methods. A general family of rank estimates was proposed by Jaeckel (1972) which achieved some robustness against outliers, while allowing the user a choice of scores for efficiency considerations. The use of Wilcoxon scores in particular has been popular because it achieves good efficiency for a normal error distribution, and robustness against outlying $y$-values. However, Jaeckel's estimates remained sensitive to observations with outlying $x$-values, or high leverage points. Sievers (1983) proposed a weighted rank estimate that reduces to the Wilcoxon scores under constant weights. In this paper we show that the estimate proposed achieves robustness against both $x$- and $y$-outliers (in the sense of bounded influence). A breakdown point is derived which is a measure of the proportion of outliers that the estimator can handle. Following Krasker (1980), weights are derived which optimize efficiency given a required degree of robustness. Finally, a test is proposed for the hypothesis $H_0$: $\beta' = (\beta_1', \mathbf{0}')$.

Two examples in Section 9 illustrate the practical performance of the estimator. In example 1, a single outlying $x$-value makes the least squares and Wilcoxon fits totally unreliable, but the proposed estimator does well. Example 2 fits multiple-regression data with three independent variables and multiple outliers. Our estimator with generalized Mallows weights exposes the cluster of high leverage points. This example shows that varying the weight function to achieve additional local stability is useful in an exploratory framework. The weights allow the user to control and vary the degree of robustness in $x$-space only. Robustness against $y$-outliers

†*Address for correspondence*: Department of Mathematics and Statistics, Western Michigan University, Kalamazoo, MI 49008-5152, USA.

 0035–9246/94/56209

remains the same as for the Wilcoxon estimator.

This paper contributes to developments in rank-based regression by providing bounded influence fully iterated estimates. (Tableman (1990) proposed a one-step rank-based estimator.) Good efficiency is retained for the normal model (typically 90–95% in simple regression, depending on the distribution of $x$ and the choice of $c$). Finally, the estimate is iterated to convergence and is thus insensitive to estimates of nuisance scale parameters that determine step size.

## 2. THE ESTIMATE

Consider estimating $\beta$ by minimizing

$$D(\beta) = \sum\sum_{i<j} b_{ij}|z_i - z_j| \tag{2.1}$$

where $z_i = y_i - x_i'\beta$. The weights $b_{ij}$ may be a function of the $X$-matrix. The dispersion function (2.1) was originally proposed by Sievers (1983) in the context of the general linear model. Note that $D(\beta)$ is free of the intercept $\alpha$ which may be estimated as a second stage. When $b_{ij} \equiv 1$, equation (2.1) reduces to Jaeckel's (1972) rank dispersion function for Wilcoxon scores

$$D(\beta) = \sum\sum_{i<j} |z_i - z_j| = 2\sum_{i=1}^{n} \{R(z_i) - (n+1)/2\}z_i \tag{2.2}$$

where $R(z_i)$ is the rank of $z_i$ among $z_1, \ldots, z_n$.

The estimate is regression and scale equivariant, and is affine equivariant if $b_{ij} = b(x_i, x_j)$ is invariant with respect to non-singular transformations $Ax_i$.

We shall refer to the estimate as the weighted pairwise absolute deviation (WPAD).

## 3. ASYMPTOTIC NORMALITY

We shall treat $(x_i, y_i)$, $i = 1, \ldots, n$, as observations from a $(p+1)$-dimensional distribution with cumulative density function (CDF) $H$, where $X$ has marginal CDF $M(x)$ and the conditional distribution of $Y$ given $X$ is denoted $y|x \sim F(y - \alpha - x'\beta)$.

Suppose that the weights satisfy $b_{ij} = b_{ji}$. Create the symmetric $n \times n$ weight matrix $W_n = [w_{ij}]$ with off-diagonal elements $w_{ij} = -(1/n)b_{ij}$ and $i$th diagonal element $w_{ii} = (1/n)\Sigma_{j \neq i} b_{ij}$. Let $X_n^*$ be the $n \times p$ matrix with $i$th row $x_i$, $i = 1, \ldots, n$, and let $X_n$ denote the centred $X_n^*$-matrix. Suppose that there are $p \times p$ positive definite matrices $C$, $V$ and $\Sigma$ such that

$$n^{-1}X_n' W_n X_n \xrightarrow{p} C, \tag{3.1}$$

$$n^{-1}X_n' W_n^2 X_n \xrightarrow{p} V, \tag{3.2}$$

$$n^{-1}X_n' X_n \xrightarrow{p} \Sigma \tag{3.3}$$

where $A \xrightarrow{p} B$ means elementwise convergence in probability. Expressions (3.1) and (3.2) are weighted analogues of the more familiar assumption (3.3). Expanding the matrix products shows that

$$\Sigma = \tfrac{1}{2}\int\int (x_2 - x_1)(x_2 - x_1)' \, dM(x_2)\, dM(x_1),$$

$$C = \tfrac{1}{2}\int\int (x_2 - x_1)(x_2 - x_1)' \, b(x_1, x_2)\, dM(x_2)\, dM(x_1),$$

$$V = \int\left\{\int (x_2 - x_1)\, b(x_2, x_1)\, dM(x_2)\right\}\left\{\int (x_2 - x_1)\, b(x_2, x_1)\, dM(x_2)\right\}' dM(x_1)$$

which are all equal if $b_{ij} \equiv 1$.

The following theorem is a result from Sievers (1983), extended to where $x$ is random. The extension is proved by conditioning on $x$ and applying a multivariate version of Slutsky's theorem. In the assumptions from Sievers (1983) which are cited in the theorem, replace *convergence* with *convergence in probability*.

*Theorem 1.* Suppose that $\hat{\beta}_n$ minimizes $D(\beta)$ and let $\beta_0$ denote the true parameter value. Under assumptions A1–A3, A6 and A7 of Sievers (1983), plus expressions (3.1) and (3.2),

$$(\hat{\beta}_n - \beta_0)\sqrt{n} \xrightarrow{d} n\{0, (1/12\gamma^2)C^{-1}VC^{-1}\} \tag{3.4}$$

where $\gamma = \int f^2(y)\, dy$.

The scalar multiple $(12\gamma^2)^{-1}$ is a measure of scale that is the rank analogue of $\sigma_f^2 = \int (y - \mu_y)^2 \, dF(y)$ in the least squares procedure. It is the height of the density of $Y_1 - Y_2$ at the point of symmetry. In the unweighted case $b_{ij} = 1$, it can be shown that $C^{-1}VC^{-1} = \Sigma^{-1}$ so that expression (3.4) reduces to

$$(\hat{\beta}_n - \beta_0)\sqrt{n} \xrightarrow{d} n\{0, \ (1/12\gamma^2)\Sigma^{-1}\}, \tag{3.5}$$

which is similar to the least squares result except for the constant $(12\gamma^2)^{-1}$.

## 4. INFLUENCE FUNCTION

The influence function is a measure of the sensitivity of an estimator to small changes in the data and hence is a measure of robustness. More specifically, $\mathrm{IF}(x_0, y_0; H)$ measures approximately $n$ times the change in the estimator when the point $(x_0, y_0)$ is added to a very large sample of size $n - 1$ from the distribution $H$.

The gradient vector $\nabla D(\beta)$ exists almost everywhere and we define

$$\begin{aligned} S(\beta) &= -\nabla D(\beta) \\ &= \sum\sum_{i<j} b_{ij}(x_i - x_j)\,\mathrm{sgn}(z_i - z_j) \end{aligned} \tag{4.1}$$

where $x_i$ is a $p \times 1$ vector and $\mathrm{sgn}(a) = +1, 0, -1$ as $a >, =$ or $< 0$. The estimate $\hat{\beta}$ is a solution to $S(\beta) = 0$. Equation (4.1) has asymptotic functional form

$$S\{\beta(H)\} = \int\int b(x_1, x_2)(x_2 - x_1)\,\mathrm{sgn}\{(y_2 - x_2'\beta) - (y_1 - x_1'\beta)\}\, dH(x_2, y_2)\, dH(x_1, y_1)$$

and if $H_n(x, y)$ is the joint empirical CDF then

$$S\{\hat{\beta}(H_n)\} = 0$$

yields the rank estimate $\hat{\beta}$. The following proposition gives the influence function of the estimate $\hat{\beta}(H_n)$ as defined by Hampel (1974).

*Proposition 1.* Suppose that $\epsilon_i - \epsilon_j$ has density $g$ continuous at 0 with $g(0) > 0$, and expression (3.1) holds. The influence function of $\hat{\beta}$ at the model distribution $H$ in the direction of $(x_0, y_0)$ is

$$\mathrm{IF}(x_0, y_0) = \frac{F(y_0 - \alpha - x_0'\beta) - \frac{1}{2}}{\gamma} C^{-1} \int (x_0 - x)\, b(x_0, x)\, \mathrm{d}M(x). \tag{4.2}$$

The proof is given in Appendix A. Note that $E(\mathrm{IF})(\mathrm{IF})'$ gives the asymptotic covariance matrix (3.4), as expected. Equation (4.2) is the (asymptotic) average of the columns of the influence matrix in Sievers (1983).

The influence function factors into the *influence of residual* and *influence of position in factor space* (Hampel *et al.* (1986), p. 313), i.e.

$$\mathrm{IF}(x_0, y_0) = \mathrm{IR}(r_0)\, \mathrm{IP}(x_0)$$

where $\mathrm{IR} = \{F(r_0) - \frac{1}{2}\}/\gamma$ and

$$\mathrm{IP} = C^{-1} \int (x_0 - x)\, b(x_0, x)\, \mathrm{d}M(x).$$

IR is bounded naturally through the error CDF $F(\ )$. IP is unbounded in the unweighted case and can be bounded by an appropriate choice of weights.

By equivariance, we can assume without loss of generality that $Ex = 0$. If the weights are chosen so that

$$b(x_1, x_2) = h(x_1)\, h(x_2) \tag{4.3}$$

and assuming that

$$Ex\, h(x) = 0 \tag{4.4}$$

then $C = Exx'\, E\, h(x)$ and equation (4.2) reduces to

$$\mathrm{IF}(x_0, y_0) = \gamma^{-1}\{F(y_0 - \alpha - x_0'\beta) - \tfrac{1}{2}\}\{Exx'\, h(x)\}^{-1} x_0\, h(x_0). \tag{4.5}$$

Compare equation (4.5) with the least squares influence function

$$\mathrm{IF}(x_0, y_0; \mathrm{LS}) = (y_0 - \alpha - x_0'\beta)(Exx')^{-1} x_0, \tag{4.6}$$

which is unbounded in both residual and position in factor space.

## 5. BREAKDOWN POINT

The breakdown point, unlike the influence function, measures the effect on the estimator of changes in a large proportion of the sample. If we control a sufficiently large proportion of the sample, we should be able to shift the estimator an arbitrarily large distance from its original value. The breakdown defined below measures the maximum proportion that we can control and still not be able to cause an infinitely large shift.

Let $H_t = (1-t)H + t\delta_{(x_0, y_0)}$ where $\delta$ is the CDF of a point mass. Let $\mathrm{bias}(t) = \|\beta(H_t) - \beta(H)\|$. The breakdown point is defined as $\epsilon^* = \sup[t < \frac{1}{2}$:

$\max_{x_0, y_0}\{\text{bias}(t)\} < \infty]$. This is a special case of the more general definition by Hampel (1968).

*Theorem 2.* Let $x_1$ and $x_2$ denote independent random vectors from the marginal distribution $M$ of the explanatory variables. The estimate $\hat{\beta}$ has gross error breakdown

$$\epsilon^* = \inf_{\|\lambda\|=1}\left\{\frac{\frac{1}{2}E|\lambda'(x_1-x_2)|b(x_1,x_2)}{\frac{1}{2}E|\lambda'(x_1-x_2)|b(x_1,x_2)+\sup_{x_0\in\chi}E|\lambda'(x_0-x)|b(x_0,x)}\right\}. \quad (5.1)$$

The proof builds on the proof of Maronna *et al.* (1979) and is given in Appendix A.

*Corollary 1.* Let $c_p = E|z_1|$ where $(z_1, \ldots, z_p)$ is uniformly distributed on the sphere $|z| = 1$. If $M(x)$ is such that $x_1 - x_2$ is spherically symmetric and if $b(x_1, x_2)$ is independent of $z = (x_1-x_2)/\|x_1-x_2\|$, then

$$\epsilon^* \geqslant \frac{\frac{1}{2}c_p}{\frac{1}{2}c_p + R(p, b, M)} \quad (5.2)$$

where

$$R(p, b, M) = \frac{\sup_{x_0}\{E\|x_0-x\|\, b(x_0, x)\}}{E\|x_1-x_2\|b(x_1, x_2)}. \quad (5.3)$$

The proof follows from the fact that $E|\lambda'(x_1-x_2)|b(x_1,x_2) = c_pE\|(x_1-x_2)\|b(x_1, x_2)$. $c_p$ can be calculated recursively from the relation $c_p = 2\{\pi(p-1)c_{p-1}\}^{-1}$ with $c_1 = 1$ (Maronna *et al.* (1979), p. 97). $R$ depends on the dimension $p$ and weights $b$ and weakly on the distribution $M(x)$. Suppose that $b$ can be chosen so that $R \leqslant k$; then inequality (5.2) reduces to

$$\epsilon^* \geqslant \frac{\frac{1}{2}c_p}{\frac{1}{2}c_p + k}. \quad (5.4)$$

Table 1 gives values of the right-hand side of inequality (5.4) for several values of $p$ and $k$.

*Corollary 2.* $\epsilon^* \leqslant \frac{1}{3}$.

TABLE 1
*Lower bounds on* $\epsilon^*$

| $k$ | *Lower bound for the following values of p:* $p=1$ | $p=2$ | $p=3$ | $p=4$ |
|---|---|---|---|---|
| 2 | 0.20 | 0.14 | 0.11 | 0.10 |
| 3 | 0.14 | 0.10 | 0.08 | 0.07 |
| 4 | 0.11 | 0.07 | 0.06 | 0.05 |
| 5 | 0.09 | 0.06 | 0.05 | 0.04 |

The corollary follows from equation (5.1) and the inequality

$$E|\lambda'(x_1-x_2)|b(x_1,x_2) \leqslant \sup_{x_0}\{E|\lambda'(x_0-x)|b(x_0,x)\}$$

for all $\|\lambda\|=1$.

## 6. OPTIMAL ROBUST WEIGHTS

In this section we give weights $h(\ )$ that minimize the trace of the asymptotic covariance matrix subject to a bound on the norm of the influence function. By equation (4.5), the problem reduces to minimizing

$$\operatorname{tr}[\{Exx'h(x)\}^{-1}Exx'h^2(x)\{Exx'h(x)\}^{-1}]$$

given a bound on $\|xh(x)\|$. A parallel optimality problem for Mallows generalized M (GM) estimators minimizes

$$\operatorname{tr}[\{Exx'h(x)\psi'(r)\}^{-1}Exx'h^2(x)\psi^2(r)\{Exx'h(x)\psi'(r)\}^{-1}]$$

over $\psi$ and $h$ subject to a bound on $\|xh(x)\psi(r)\|$ (see Hampel *et al.* (1986), p. 321). Following the discussion of section 6.3 of Hampel *et al.* (1986), we have the following result.

*Theorem 3.* Suppose that $F$ has differentiable density $f$, $\int f^2 < \infty$, and expressions (3.1), (3.2), (4.3) and (4.4) hold. Let $h_c(x) = \min(1, c/\|Bx\|)$ where $B$ is defined by $E\{xx'h_c(Bx)\} = B^{-1}$. Then $h_c$ minimizes $\operatorname{tr}\{\operatorname{var}(\hat{\beta})\}$ over all $h(\ )$ that satisfy $\sup\|\mathrm{IF}(x,y)\| \leqslant d$ $(=(2\gamma)^{-1}\sup\|Bxh_c(x)\|)$.

For applications we suggest a slightly modified but considerably more convenient weight function, as follows. By affine equivariance we may (and should) standardize the $x$s initially, preferably by some robust location vector $\mu_x$ and scale matrix $S_x$. The $B$-matrix in theorem 3 is then close to the identity matrix, and the optimal weights are approximately

$$h(x) = \min\left[1, \frac{c}{\{(x-\mu)'S^{-1}(x-\mu)\}^{1/2}}\right]$$

which are the weights of the regular Mallows GM estimate.

## 7. GENERALIZED MALLOWS WEIGHTS

'If one point changes the estimate by many standard errors, . . . it is small consolation that the change is bounded by some large number' (Simpson *et al.*, 1992). In the known presence of extreme leverage points, we may choose to focus on achieving this local stability property, giving less consideration to efficiency. In these instances, sensitivity of the estimate to leverage points may be further reduced by using generalized Mallows weights

$$h(x) = \min\left[\left\{1, \frac{c}{(x-\mu)'S^{-1}(x-\mu)}\right\}^{r/2}\right] \tag{7.1}$$

considered by Simpson *et al.* (1992) for their one-step Mallows estimate. See section 1 of their paper for a discussion on the choice of $r$. In an example in Section 9,

we show that $r = 2$ is effective in uncovering outliers. We shall refer to the estimate using weights (7.1) as the WPAD($r$) estimate. WPAD(1) corresponds to the optimal efficient estimate of the previous section, and WPAD(0) is the regular Wilcoxon rank estimate.

## 8. TEST BASED ON GRADIENT $S(\beta)$

Let $\beta' = (\beta_1', \beta_2')$ and consider testing $H_0$: $\beta_2 = 0$ *versus* $H_a$: $\beta_2 \neq 0$ where $\beta_2$ is a $q \times 1$ vector ($q \leqslant p$). $\beta_1$ is a vector of unspecified nuisance parameters. Partition $S'(\beta) = (S_1'(\beta), S_2'(\beta))$ and let $\hat{\beta}_0' = (\hat{\beta}_1', \mathbf{0}')$ minimize $D(\beta)$ over all $\beta' = (\beta_1', \mathbf{0}')$. Then $\hat{\beta}_0$ is a solution to

$$S_1(\hat{\beta}_0) = \frac{\mathrm{d}}{\mathrm{d}\beta_1} D(\beta)\,|_{\beta=\hat{\beta}_0} = 0 \tag{8.1}$$

which is a set of $p - q$ equations in $p - q$ unknown. Under the null hypothesis, $S(\hat{\beta}_0)$ should be close to the zero vector, where $S(\ )$ is given by equation (4.1). Thus a test may be constructed that rejects the null hypothesis for values of $S(\hat{\beta}_0)$ away from 0. By equation (8.1), $S(\hat{\beta}_0)' = (S_1'(\hat{\beta}_0), S_2'(\hat{\beta}_0)) = (\mathbf{0}', S_2'(\hat{\beta}_0))$ so the test statistic need only depend on $S_2$.

Partition $C$ according to the partition $\beta' = (\beta_1', \beta_2')$.

*Theorem 4.* Under the assumptions of theorem 1, if $H_0$: $\beta_2 = 0$ is true, then

$$G^* = 3n^{-3} S_2'(\hat{\beta}_0) M^{-1} S_2(\hat{\beta}_0) \xrightarrow{\mathrm{d}} \chi^2(q)$$

where $M = (-C_{21}C_{11}^{-1} \quad I)V(-C_{21}C_{11}^{-1} \quad I)' = V_{22} - V_{21}C_{11}^{-1}C_{12} - C_{21}C_{11}^{-1}V_{12} + C_{21}C_{11}^{-1}V_{11}C_{11}^{-1}C_{12}$.

See Appendix A for the proof.

The hypothesis $H_0$: $\beta_2 = 0$ clearly includes tests for main effects, presence of interaction and significance of covariates, among others. With a proper transformation the hypothesis may include tests for any $q$ linearly independent constraints on $\beta$, i.e. $H_0$: $L\beta = 0$ where $L$ is a specified $q \times p$ matrix of constraints.

## 9. EXAMPLES

The following examples illustrate robustness of the WPAD estimate to outlying observations with high leverage. The second example shows the usefulness of increased local stability in flagging a high proportion of extreme leverage points.

The weights (7.1) are used in the calculations. For $(\mu, S)$, we chose the minimum volume ellipsoid proposed by Rousseeuw and van Zomeren (1987), i.e. the determinant of the scale matrix $S_x$ is minimized subject to

$$\#\{i\colon\ (x_i - \mu_x)' S_x^{-1} (x_i - \mu_x) \leqslant a^2\} \geqslant h$$

where $h = [(n+p+1)/2]$ ([ ] denotes 'the integer part of') and $a^2$ is a constant (taken to be $a^2 = \chi^2_{p,0.50}$). The constant $c$ in equation (7.1) is set at $c = \mathrm{med}\{d_i\} + 3\,\mathrm{MAD}\{d_i\}$ where $d_i = (x_i - \mu)' S^{-1} (x_i - \mu)$. This choice is equivalent to downweighting $x_i$ whenever $d_i$ is approximately two standard deviations larger than the mean.

In the simple linear regression case of example 1, standardization is done by using $(\mu, S) = [\mathrm{med}\{x_i\}, (1.483\,\mathrm{MAD}\{x_i\})^2]$.

### 9.1. *Pilot-plant Data*

Consider the pilot-plant data from Rousseeuw and Leroy (1987), originally from Daniel and Wood (1971). There are 20 observations and one independent variable. Consider the fitted equations for the least squares, Wilcoxon and the WPAD(1) estimate. Since the data have no outliers the three fits are quite similar, respectively

$$\hat{y} = 35.5 + 0.322x,$$

$$\hat{y} = 35.4 + 0.323x,$$

$$\hat{y} = 35.4 + 0.323x.$$

Suppose that we introduce an artificial outlier with high leverage (by changing $x_1 = 123$ to $x_1 = 1230$). The resulting fitted equations are then respectively

$$\hat{y} = 65.58 + 0.0191x,$$

$$\hat{y} = 65.16 + 0.0180x,$$

$$\hat{y} = 35.87 + 0.3150x.$$

Whereas the least squares and Wilcoxon estimates have dramatically changed, the WPAD estimate has remained relatively stable. Table 2 shows the effect on the test statistics for testing $H_0$: $\beta = 0$. Shown are the $F$-statistic for least squares and the gradient test statistics for the Wilcoxon and WPAD estimates. Only the WPAD estimate has unchanged inference results.

Similar behaviour was observed when $x_1$ was moved even further to $x_1 = 12300$. Pursuing another direction, we observed the behaviour of the WPAD estimate as more artificial outliers were introduced (by multiplying successive $x$-values by 10). The WPAD estimate remained stable up to three contaminated points. This exhibits an empirical breakdown of at least 15% for $p = 1$.

### 9.2. *Hawkins-Bradu-Kass Data*

We now look at multiple regression with multiple high leverage outliers. The artificial data generated by Hawkins *et al.* (1984) have 75 observations and three independent variables. Cases 1–10 are bad leverage points whereas cases 11–14 are good leverage points, i.e. they agree with the model for the bulk of the data. Table 3 gives selected standardized residuals for the least squares, Wilcoxon and the WPAD estimate for $r = 1$ and $r = 2$. None of the remaining residuals are larger than 1.4 in absolute value. The least squares residuals are externally standardized whereas the Wilcoxon and WPAD residuals have been standardized according to a first-

TABLE 2
*Test statistic values for $H_0$: $\beta = 0$ (pilot-plant data)*

| $x_1$ | *Least squares estimate* | *Wilcoxon estimate* | *WPAD estimate* |
|---|---|---|---|
| 123 | 3381.06† | 19.556† | 19.556† |
| 1230 | 1.770 | 1.692 | 16.484† |

† Significant at $\alpha = 0.01$.

TABLE 3
*Selected standardized residuals for the Hawkins data*

| *Case* | *Least squares estimate* | *Wilcoxon estimate* | *WPAD(1) estimate* | *WPAD(2) estimate* |
|---|---|---|---|---|
| 1 | 1.57 | 0.92 | 1.97 | 7.19 |
| 2 | 1.86 | 1.28 | 2.23 | 7.39 |
| 3 | 1.40 | 0.67 | 1.97 | 7.24 |
| 4 | 1.19 | −0.35 | 0.90 | 6.61 |
| 5 | 1.42 | 0.34 | 1.55 | 7.01 |
| 6 | 1.60 | 0.85 | 1.92 | 7.18 |
| 7 | 2.12 | 1.84 | 2.75 | 7.74 |
| 8 | 1.79 | 1.49 | 2.55 | 7.60 |
| 9 | 1.26 | 0.04 | 1.28 | 6.81 |
| 10 | 1.42 | 0.71 | 1.91 | 7.17 |
| 11 | −4.03 | −13.25 | −10.51 | 0.07 |
| 12 | −5.29 | −15.55 | −10.70 | 0.00 |
| 13 | −3.04 | −12.38 | −9.73 | 0.49 |
| 14 | −2.67 | −11.35 | −8.91 | −0.03 |

order approximation to $\sqrt{\text{var}(\hat{e}_i)}$ proposed by McKean *et al.* (1990). Note that the least squares, Wilcoxon and WPAD(1) methods flag cases 11–14 as outlying, their fits pulled by the 10 bad leverage points. At $r=2$, the WPAD estimate has gained additional robustness and flags cases 1–10 instead.

## ACKNOWLEDGEMENT

The work of T. P. Hettmansperger was partially supported by Office of Naval Research contract N00014-80-C0741.

## APPENDIX A

### A.1. *Proof of Proposition 1*

Let $\phi(a, b) = 1, \frac{1}{2}, 0$ as $a <, =$ or $> b$. Then the estimator has asymptotic functional form

$$\int\int x_1 b(x_1, x_2)\,[\phi\{y_2 - y_1 < (x_2 - x_1)'\beta(H)\} - \tfrac{1}{2}]\,\mathrm{d}H_2\,\mathrm{d}H_1 = 0.$$

Replace $H$ by $H_t = (1-t)H + t\delta_0$ where $\delta_0(x, y)$ is the CDF of a point mass at $(x_0, y_0)$. Taking derivatives of both sides with respect to $t$ and evaluating at $t=0$ gives

$$0 = (\mathrm{d}/\mathrm{d}t)|_{t=0} \int\int\int x_1 b(x_1, x_2)\,(F[y_1 - \alpha - x_1'\beta(H_t) + x_2'\{\beta(H_t) - \beta\}]$$

$$-\tfrac{1}{2})\,\mathrm{d}F(y_1 - \alpha - x_1'\beta)\,\mathrm{d}M_2\,\mathrm{d}M_1 + \int x_1 b(x_1, x_2)\,\{F(y_1 - \alpha - x_1'\beta) - \tfrac{1}{2}\}\,\mathrm{d}M_2\,\mathrm{d}\delta_0(x_1, y_1)$$

$$+ \int\int x_1 b(x_1, x_2)\,\{I(y_1 - x_1'\beta > y_2 - x_2'\beta) - \tfrac{1}{2}\}\,\mathrm{d}\delta_0(x_2, y_2)\,\mathrm{d}H(x_1, y_1)$$

$$= \int\int b(x_1, x_2) x_1 (x_2 - x_1)' \, dM(x_2) \, dM(x_1) \int f^2(y_1 - \alpha - x_1'\beta) \, dy_1 \, \dot{\beta}$$

$$+ F(y_0 - \alpha - x_0'\beta) \int x_0 \, b(x_0, x) \, dM(x) - \{F(y_0 - \alpha - x_0'\beta) - \tfrac{1}{2}\} \int x \, b(x, x_0) \, dM(x)$$

$$= -\gamma C\dot{\beta} + \{F(y_0 - \alpha - x_0'\beta) - \tfrac{1}{2}\} \int (x_0 - x) \, b(x_0, x) \, dM(x)$$

where $\dot{\beta} = (d/dt)\beta(H_t)|_{t=0}$ is the influence function.

## A.2. *Proof of Theorem 2*

$$0 = S\{\beta(H_t)\}$$

$$= \int\int (x_1 - x_2) \, b_{12}[I\{y_2 < y_1 + (x_2 - x_1)'\beta(H_t)\} - \tfrac{1}{2}] \, dH_t(x_2, y_2) \, dH_t(x_1, y_1)$$

$$= (1-t)^2 \int\int (x_1 - x_2) \, b_{12}[I\{y_2 < y_1 + (x_2 - x_1)'\beta_t\} - \tfrac{1}{2}] \, dH_2 \, dH_1$$

$$+ 2t(1-t) \int (x_0 - x) \, b(x_0, x) \, [I\{y < y_0 + (x - x_0)'\beta_t\} - \tfrac{1}{2}] \, dH(x, y).$$

Since the terms on the right-hand side sum to 0, the magnitude of the two terms in the sum have to be equal. Thus, for every vector $\lambda \in R^p$ such that $|\lambda| = 1$,

$$2t|E_H \lambda'(x_0 - x) \, b(x_0, x) \, [I\{y < y_0 + (x - x_0)'\beta_t\} - \tfrac{1}{2}]|$$

$$= (1-t)|E\lambda'(x_1 - x_2) \, b_{12}[I\{y_2 < y_1 + (x_2 - x_1)'\beta_t\} - \tfrac{1}{2}]|. \quad \text{(A.1)}$$

Now suppose that $\max\{\text{bias}(t)\} = \infty$. Then there is a sequence of point mass distributions $\{\delta_{0,k}\}$ such that $\|\beta(H_{t,k})\| \to \infty$ and $\lambda_k = \{1/\|\beta(H_{t,k})\|\}\beta(H_{t,k}) \to \lambda^*$ for some $\lambda^* \in R^p$. From equation (A.1)

$$(1-t)|E_{H_2, H_1} \lambda_k'(x_1 - x_2) \, b(x_1, x_2) \, [I\{y_2 < y_1 + (x_2 - x_1)'\beta_{t,k}\} - \tfrac{1}{2}]|$$

$$= 2t|E_H \lambda_k'(x_{0,k} - x) \, b(x_{0,k}, x) \, [I\{y < y_{0,k} + (x - x_{0,k})'\beta_{t,k}\} - \tfrac{1}{2}]$$

$$\leqslant 2tE_H|\lambda_k'(x_{0,k} - x) \, b(x_{0,k}, x)||\tfrac{1}{2}|.$$

Taking limits,

$$\tfrac{1}{2}(1-t)E_{H_2, H_1}|\lambda^{*\prime}(x_1 - x_2)|b(x_1, x_2) \leqslant t \sup_{x_0}\{E_H|\lambda^{*\prime}(x_0 - x)|b(x_0, x)\}$$

so that

$$t \geqslant \frac{\tfrac{1}{2}E_{H_2, H_1}|\lambda'(x_1 - x_2)|b(x_1, x_2)}{\tfrac{1}{2}E_{H_2, H_1}|\lambda'(x_1 - x_2)|b(x_1, x_2) + \sup_{x_0}\{E_H|\lambda'(x_0 - x)|b(x_0, x)\}}. \quad \text{(A.2)}$$

To prove the reverse inequality, suppose that $\max\{\text{bias}(t)\} < \infty$. Fix $x_0$ in equation (A.1). Take a sequence $y_{0,k}$ such that

$$I\{y < y_0 + (x - x_0)'\beta_t\} - \tfrac{1}{2} \xrightarrow{p} \tfrac{1}{2},$$

$$t|E_H \lambda'(x_0 - x) \, b(x_0, x)| \leqslant (1-t)|E_{H_2, H_1}\lambda'(x_1 - x_2) \, b(x_1, x_2)||\tfrac{1}{2}|.$$

Take a sequence $\{x_0\}$ such that the left-hand side tends to $t \sup_{x_0}|E_H \lambda'(x_0 - x) \, b(x_0, x)|$,

$$t \sup_{x_0}|E_H \lambda'(x_0 - x) \, b(x_0, x)| \leqslant \tfrac{1}{2}(1-t)E_{H_2, H_1}|\lambda'(x_1 - x_2)|b(x_1, x_2).$$

## A.3. *Proof of Theorem 4*

We need the following results from Sievers (1983) which we state here as two lemmas. The assumptions required are a subset of the assumptions of theorem 4. Let $\gamma$, $C$ and $V$ be as defined in Section 3, let $B$ be a constant and suppose that $\beta^*$ denotes the true parameter value.

*Lemma 1:* asymptotic linearity.

$$P\{\sup_{\sqrt{n}\|\beta-\beta^*\|\leqslant B}\|n^{-3/2}S(\beta)-n^{-3/2}S(\beta^*)+2\gamma C(\beta-\beta^*)\sqrt{n}\|\geqslant\epsilon\}\to 0.$$

*Lemma 2:* $n^{-3/2}S(\beta^*)\xrightarrow{d} n(0,\frac{1}{3}V)$. Now, under the null hypothesis $\beta^{*\prime}\equiv\beta_0'=(\beta_{10}',\mathbf{0}')$. By lemma 1, for $\|\beta-\beta^*\|\sqrt{n}\leqslant B$,

$$n^{-3/2}\begin{pmatrix}S_1(\beta)\\S_2(\beta)\end{pmatrix}=n^{-3/2}\begin{pmatrix}S_1(\beta_0)\\S_2(\beta_0)\end{pmatrix}-2\gamma\sqrt{n}\begin{pmatrix}C_{11}&C_{12}\\C_{21}&C_{22}\end{pmatrix}\begin{pmatrix}\beta_1-\beta_{10}\\\beta_2-0\end{pmatrix}+o_p(1).$$

Since $(\hat\beta_0-\beta_0)\sqrt{n}$ is bounded in probability under the null hypothesis, substituting $\hat\beta_0$ for $\beta$ above, we obtain

$$n^{-3/2}S_1(\hat\beta_0)\doteq n^{-3/2}S_1(\beta_0)-2\gamma C_{11}(\hat\beta_1-\beta_{10})\sqrt{n}, \tag{A.3}$$

$$n^{-3/2}S_2(\hat\beta_0)\doteq n^{-3/2}S_2(\beta_0)-2\gamma C_{21}(\hat\beta_1-\beta_{10})\sqrt{n}. \tag{A.4}$$

By equations (8.1) and (A.3), $2\gamma(\hat\beta_1-\beta_{10})\sqrt{n}\doteq n^{-3/2}C_{11}^{-1}S_1(\beta_0)$. Substituting in equation (A.4),

$$\begin{aligned}n^{-3/2}S_2(\hat\beta_0)&\doteq n^{-3/2}S_2(\beta_0)-n^{-3/2}C_{21}C_{11}^{-1}S_1(\beta_0)\\&=n^{-3/2}(-C_{21}C_{11}^{-1}\quad I)\,S(\beta_0).\end{aligned}$$

By lemma 2,

$$n^{-3/2}S_2(\hat\beta_0)\sqrt{3}\xrightarrow{d} n(0,M),$$

and the theorem immediately follows.

## REFERENCES


Daniel, C. and Wood, F. S. (1971) *Fitting Equations to Data.* New York: Wiley.

Hampel, F. R. (1968) Contributions to the theory of robust estimation. *PhD Thesis.* University of California, Berkeley.

——— (1974) The influence curve and its role in robust estimation. *J. Am. Statist. Ass.*, **69**, 383–393.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986) *Robust Statistics: the Approach Based on Influence Functions.* New York: Wiley.

Hawkins, D. M., Bradu, D. and Kass, G. V. (1984) Location of several outliers in multiple regression data using elemental sets. *Technometrics*, **26**, 197–208.

Jaeckel, L. A. (1972) Estimating regression coefficients by minimizing the dispersion of residuals. *Ann. Math. Statist.*, **43**, 1449–1458.

Krasker, W. S. (1980) Estimation in linear regression models with disparate data points. *Econometrica*, **48**, 1333–1346.

Maronna, R. A., Bustos, O. H. and Yohai, V. J. (1979) Bias and efficiency-robustness of general *M*-estimators for regression with random carriers. *Lect. Notes Math.*, **757**, 91–116.

McKean, J. W., Sheather, S. J. and Hettmansperger, T. P. (1990) Regression diagnostics for rank-based methods. *J. Am. Statist. Ass.*, **85**, 1018–1028.

Rousseeuw, P. J. and Leroy, A. M. (1987) *Robust Regression and Outlier Detection.* New York: Wiley.

Rousseeuw, P. J. and van Zomeren, B. C. (1990) Unmasking multivariate outliers and leverage points (with comments). *J. Am. Statist. Ass.*, **85**, 633–651.

Sievers, G. L. (1983) A weighted dispersion function for estimation in linear models. *Communs Statist. Theory Meth.*, **12**, 1161–1179.

Simpson, D. G., Ruppert, D. and Carroll, R. J. (1992) On one-step GM-estimates and stability of inferences in linear regression. *J. Am. Statist. Ass.*, **87**, 439–450.

Tableman, M. (1990) Bounded-influence rank regression: a one-step estimator based on Wilcoxon scores. *J. Am. Statist. Ass.*, **85**, 508–513.