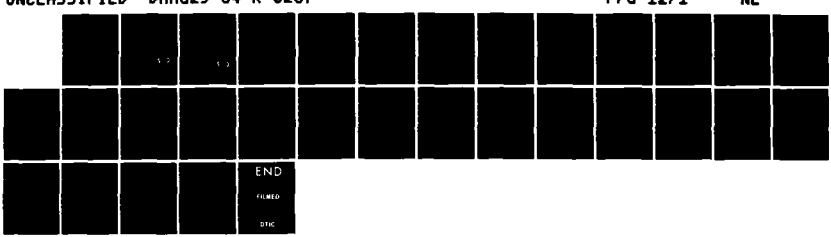
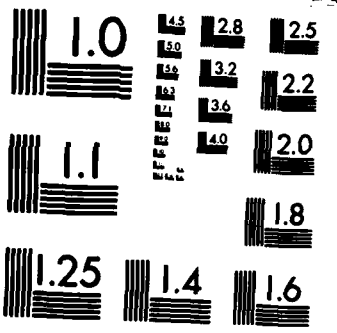
AD-A153 301 ROBUST REGRESSION THROUGH ROBUST COVARIANCES(U) YALE 1/1
UNIV NEW HAVEN CONN DEPT OF STATISTICS
S MORGENTHALER ET AL. JAN 85 TR-2-ARO ARO-21779.2-MA
UNCLASSIFIED DAAG29-84-K-0207 F/G 12/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

STATISTICS CENTER

Massachusetts Institute of Technology

77 Massachusetts Avenue, Rm. E40-111, Cambridge, Massachusetts 02139 (617) 253-8722

ROBUST REGRESSION THROUGH ROBUST COVARIANCES

BY

STEPHAN MORGENTHALER
DEPARTMENT OF STATISTICS, YALE UNIVERSITY

AND

RICARDO MARONNA
UNIVERSIDAD NACIONAL DE LA PLATA

TECHNICAL REPORT NO. ARO 2

JANUARY 1985

PREPARED UNDER CONTRACT

DAAG29-84-K-0207

FOR THE U.S. ARMY RESEARCH OFFICE

ROBUST REGRESSION THROUGH ROBUST COVARIANCES

By

Stephan Morgenthaler

Department of Statistics, Yale University

And

Ricardo Maronna

Universidad Nacional de La Plata

ABSTRACT

This paper describes a regression estimator which builds on the investigation of the covariance structure in the full space of explanatory variables and response variable. It is robust since it down weights outlying observations. A comparison to other robust estimators is included.

1. INTRODUCTION

In recent years a variety of estimators for regression parameters have been proposed. The driving force behind this kind of research is the non-robustness of the ordinary least squares (OLS) estimator. Just as in the case of location estimation the OLS solution is unduly influenced by departures from the "usual" model assumptions. Two different approaches to robustification have been proposed. The first one is due to Huber (1973) (see also Mosteller and Tukey (1977)) and works by "Huberizing" residuals. The second one is based on Hampel's work (1974) and due to Krasker and Welsch (1982). These methods bound the "influence" on the estimated parameter vector.

In this paper we will describe an approach to the regression problem which is based on covariance estimation. Instead of attacking the problem of estimating the regression parameters directly as solutions to some "normal equation", we will view the problem in a larger context and try to summarize what we learn from the data in an estimated covariance matrix which contains all the information we need. This approach has the advantage that the response variable and the explanatory variables are treated symmetrically. We can therefore compute any of the possible regressions -- i.e. any choice of response -- in a single run without having to recalculate weights.

In Section 2 we will give the formulas and discuss one particular estimator based on an affine equivariant covariance estimator. Section 3 contains a (limited) comparison with other robust regression estimators through asymptotics and experimental sampling.

2. ROBUST REGRESSION THROUGH ROBUST COVARIANCES

2.1 DEFINITION OF THE ESTIMATOR

Let $z_i' = (x_i', y_i)$, $i=1,\ldots,n$ be n vectors in $\mathbb{R}^{p+1}$ which are created by n independent observations from the stochastic model

$$Y = \beta' X + E \qquad (2.1)$$

with fixed, unknown regression parameter $\beta \in \mathbb{R}^p$, random carrier $X \in \mathbb{R}^p$ and random error E independent of the carrier.

In order to describe the probability structure in (2.1) two distributions are needed. First the error distribution $H(e/\sigma)$ and second the carrier distribution $G()$. Together with the independence assumption in (2.1) these two determine the distribution of $Z' = (X', Y)$. As we have indicated there is at least one more parameter of primary importance, namely the "scale" σ in our error distribution. All of the bounded influence research is taking place in a model like (2.1).

The classical solution to the inference problems posed by (2.1) is based on the principle of least squares. In recent years several alternative robust estimators b of β have been proposed.

They all can be conveniently described by "normal equations":

$$[1] \quad \sum_{i=1}^{n} x_i r_i = 0, \; x_i \in \mathbb{R}^p, \; r_i = y_i - b'x_i$$

$$[2] \quad \sum_{i=1}^{n} x_i \psi\left(\frac{r_i}{s}\right) = 0 \quad \text{Huber (1973).}$$

Here s is an auxillary or simultaneous estimate of the error "scale" and $\psi()$ is a somewhat arbitrary function which is usually chosen in such a way that the efficiency at a central model like $H() = \Phi()$ is high.

$$[3] \quad \sum_{i=1}^{n} w_i x_i \psi\left(\frac{r_i}{s}\right) = 0 \quad \text{Mallows (1975).}$$

Here w_i is a weight which depends on x_i (and possibly all the other x_j's).

$$[4] \quad \sum_{i=1}^{n} w_i x_i \psi\left(\frac{r_i}{sw_i}\right) = 0 \quad \text{Schweppe (1975).}$$

This form is optimal with respect to the heuristic notion of bounded influence (Hampel (1974) and Krasker and Welsch (1982)). A review of these estimators can be found in Maronna, Bustos and Yohai (1979).

[1] gives the ordinary least squares estimator (OLS), [2]

corresponds to the classical approach to robust regression and there are two popular choices for the ψ-function:

(i) $\psi(x) = \psi_k(x) = \max(-k, \min(k, x))$

(ii) $\psi(x) = bsq_k(x) = x/k\ (1 - (x/k)^2)^2$ if $-k \leq |x| \leq k$

$= 0$ otherwise

The second one is preferable since it gives excellent protection against heavy-tailed error distributions. [3] and [4] are bounded influence estimators (see Hampel (1974)). Asymptotic optimality theory tells us that estimators of the form [4] are preferable. Dividing and multiplying each term in the sum by $r_i/(sw_i)$ we get

$$[4'] \quad \sum_{i=1}^{n} x_i r_i \frac{\psi(\frac{r_i}{sw_i})}{\frac{r_i}{sw_i}} = 0$$

which we immediately recognize as the normal equation of a weighted least squares problem. The optimal weights are of the form

$$\frac{\psi_k(\frac{r_i}{sw_i})}{\frac{r_i}{sw_i}}$$

where $\psi_k()$ is as in (i) above and $w_i = (x_i' S^{-1} x_i)^{-1/2}$ (S an estimated covariance matrix of the carriers).

The theory behind this statement can be found in Krasker and Welsch (1982).

We have started this section by introducing the vectors $z_i \in \mathbb{R}^{p+1}$ which we get by joining the carriers x_i to the observations y_i. Since regression deals with the linear dependence between these two elements it is quite natural to summarize our data $z_1,\ldots,z_n$ at a first stage by a covariance matrix and then extract the specific information we are primarely interested in -- about β and σ -- at the second stage.

Remark: To implement the above program it seems advantageous to take out the constant term in our regression model. This term corresponds to a carrier identically equal to 1 and if we allow for it, we will be in the rather special position that all vectors z_i lie in a p-dimensional subspace of $\mathbb{R}^{p+1}$. The natural way to proceed is to estimate the constant term via simultaneous location estimation. This way our target will be the value of the regression function at a central carrier value. The formulas we will present are all written for the case where we do not estimate location. These formulas can, however, easily be modified.

Let us define a robust variance-covariance matrix W^* through the implicit equation

$$\frac{1}{n}\sum_{i=1}^{n} u(z_i' W^{*-1} z_i)\, z_i z_i' = W^* \qquad (2.2)$$

where u() is a weight function such that W^* is Fisher consistent at a multivariate Gaussian model. Such M-estimators of covariance

matrices are discussed and examined in Maronna (1976) (see also Huber (1981)). Now we partition W^*

$$W^* = \begin{bmatrix} V^* & c^* \\ c^{*\prime} & e^* \end{bmatrix},$$

where V^* is the pxp covariance matrix of the carriers, c^* is the px1 vector of carrier-response cross terms and e^* is the total variance of the response. In analogy to the OLS-case we finally put

$$\text{estimated parameter } b = V^{*-1} c^*$$
$$\text{and estimated error variance } s^2 = e^* - b'V^*b \qquad (2.3)$$

If we now look back we realize that the system of equations (2.2) is equivalent to

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} w_i x_i r_i = 0 \\ &\frac{1}{n}\sum_{i=1}^{n} w_i r_i^2 = s^2 \qquad (2.4) \\ &\frac{1}{n}\sum_{i=1}^{n} w_i x_i' x_i = V^*, \text{ where} \end{aligned}$$

the weights are $w_i = u(z_i'W^{*-1}z_i) = u(x_i'V^{*-1}x_i + r_i^2/s^2)$ and $r_i = y_i - x_i'b$.

proof: In order to get (2.4) from (2.2) we have to partition according to the blocks of W^*. E.g.

$$\frac{1}{n}\sum_{i=1}^{n} w_i\, y_i\, x_i = c^* = V^* b = \left(\frac{1}{n}\sum_{i=1}^{n} w_i\, x_i\, x_i'\right) b$$

$$= \frac{1}{n}\sum_{i=1}^{n} w_i\, x_i\, (x_i' b),$$ which yields the first equation in (2.4).

Remarks:

(1) From (2.4) we immediately infer that in the special case where $u() = 1$ we are almost back at the usual OLS estimates for β and σ^2.

(2) It would be of interest for the sake of high breakdown (see Huber (1981)) to estimate the covariance matrix via a high breakdown estimator (Donoho and Huber (1983) and Stahel(1981)). At the moment this does not seem to be practical, but we nevertheless want to point out this important possibility. We plan to report our results on an M-estimator approach as in (2.2), hoping that they will be helpful.

The solution W^* in (2.2), i.e. the estimated covariance matrix, satisfies a very broad equivariance relation. If we transform the data linearly $\tilde{z}_i = Az_i$ $i=1,\ldots,n$ the covariance matrix will be transformed as

$$W^*(\tilde{z}_1,\ldots,\tilde{z}_n) = A\, W^*(z_1,\ldots,z_n)\, A' \qquad (2.5)$$

(affine equivariance).

The parameters of interest in (2.1) are β and σ. The equivariance relations we might wish to impose on an estimator of these parameters are:

(i) $\tilde{y}_i = \eta(y_i + \gamma' x_i)$; $\tilde{x}_i = x_i$ $i=1,\ldots,n \longrightarrow$

$b(\tilde{y}_1,\ldots,\tilde{y}_n) = \eta(b(y_1,\ldots,y_n) + \gamma)$ and

$s(\tilde{y}_1,\ldots,\tilde{y}_n) = \eta\, s(y_1,\ldots,y_n)$

(<u>regression equivariance</u>)

and (2.6)

(ii) $\tilde{y}_i = y_i$; $\tilde{x}_i = M x_i$ $i=1,\ldots,n \longrightarrow$

$b(\tilde{y}_1,\ldots,\tilde{y}_n) = M'^{-1} b(y_1,\ldots,y_n)$ and

$s(\tilde{y}_1,\ldots,\tilde{y}_n) = s(y_1,\ldots,y_n)$

(<u>carrier equivariance</u>).

Regression equivariance means that an estimator behaves reasonably if we add an exactly linear function of the carriers to the response. The carriers are viewed as fixed and our answer b is always defined in relation to them. Carrier equivariance is quite a different concept. If the parametrization of our model (2.1) is changed in a linear fashion, it then makes sense to use the inverse transformation on the parameters.

The importance of (ii) lies in the fact that by transforming our parametrization we can sometimes get parameters which we can estimate better, i.e. with decreased variability. Carrier equivariance lets us move between these different parametrizations without inconsistencies.

It is easily seen that affine equivariance (2.5) of our covariance estimator is a sufficient condition to ensure (2.6).

Proof:

Put $A = \begin{bmatrix} I & 0 \\ \eta\gamma' & \eta \end{bmatrix}$ or $A = \begin{bmatrix} M & 0 \\ 0 & 1 \end{bmatrix}$ in (2.5).

Remark: The robust estimators described at the beginning are most easily described as solutions to a weighted least squares scheme where the weights have to be found iteratively. It is easily seen that regression equivariance and carrier equivariance are obtained in the case where the final weights are invariant under the transformations given in (2.6). Quantities which fulfill this invariance requirement are standardized residuals r_i/s and squared standardized distances in the carrier space $(x_i'V^{*-1}x_i)$, where s is a scale equivariant function of the residuals and V^* is an affine equivariant function of the carriers. In the case of the classical Huber estimator the final weights only depends on r_i/s. This estimator -- fully iterated -- is therefore equivariant. If we use a few-step version, i.e. we do not fully iterate but rather use a pre-fixed number of steps in some iteration scheme, the resulting estimator will not be equivariant but hopefully nearly so.

The Kraser-Welsch estimator is equivariant as well. Its weights depend on the product of standardized residuals times standardized carrier distance. The Mallows estimator finally will be equivariant if we are careful in the choice of carrier weights -- they should

only depend on the standardized carrier distances.

2.2. ASYMPTOTIC PROPERTIES OF THE REGRESSION ESTIMATOR (2.3)

In this subsection we want to discuss the asymptotic properties of a regression estimator defined through a covariance M-estimator (2.2). We saw in (2.4) that it corresponds to a weighted least squares estimator whose weights depend on the sum of the squared standardized residual and the squared standardized carrier distance. Maronna (1976) develops the asymptotic theory for covariance M-estimators. We plan to use these results and to look what happens if we apply (2.3). But first let us examine the influence function (see Hampel (1974)).

In order to simplify the formulas we will first consider the case where the carrier distribution G() (see beginning of Section 2) is symmetric with respect to each coordinate, i.e. the 2^p vectors $(\pm x_1,\ldots,\pm x_p)$ for all possible choices of signs are identically distributed with distribution G(). If in addition the error distribution $H(x/\sigma)$ is symmetric the influence on the jth component b_j $(j = 1,\ldots,p)$ in (2.3) is

$$IF_{b_j,G,H}(x_0,y_0) = u(d_0^2)\, r_0 \frac{x_{0,j}}{a_j} \qquad (2.7)$$

where $d_0^2 = (x_0',y_0)\, W^{-1}(G,H)\, (x_0',y_0)'$ and $r_0 = y_0 - \beta'(G,H)x_0$, and $a_j = V_{jj} + E_{G,H}[u'(d^2)x_j^2 r^2\, 2/\sigma^2]$ (note that x_j refers to the jth component of x).

In order to understand this formula we have to remember that the influence function is an asymptotic "tool" and that therefore the population values of our estimators appear in the formula. W(G,H) is defined through

$$W(G,H) = E_{G,H}\,[u(d^2)\,(x',y)'(x',y)]\quad (d^2 = (x',y)\,W^{-1}(x',y)')$$

and can be partitioned just as W^* (see (2.3)). V_{jj} denotes the jth diagonal element of W(G,H).

Remarks:

(1) If we do not impose the symmetry requirements on the carrier distribution G() the influence on the vector b in (2.3) is

$$IF_{b,G,H}\,(x_o,y_o) = u(d_o^2)\,r_o\,A^{-1}x_o,$$

where $A_{jk} = V_{jk} + E_{G,H}\,[u'(d^2)\,2r^2/\sigma^2\;\; x_j x_k]$.

Proof: Our estimator can be put into the usual M-estimator form, i.e. we estimate the parameter $\xi = (\beta,\ \sigma,\ V)$ based on the data $z_1,\ldots,z_n$ via $\sum \varphi(z_i,\xi) = 0$. Now we can apply the standard formulas to get influence function (see Huber (1981)). Our φ-function splits into three parts according to

$$\varphi_1(z,\xi) = u(d^2)\;x\;r$$

$$\varphi_2(z,\xi) = u(d^2)\;r^2 - \sigma^2$$

$$\text{and } \varphi_3(z,\xi) = u(d^2)\;xx' - V.$$

The matrix $D = E_{G,H}\,[\delta\varphi/\delta\xi]$ has the block form

$$-D = \begin{bmatrix} A & 0 & 0 \\ 0 & & \\ 0 & & * \end{bmatrix},$$

where A is the matrix which turns up in the influence for b. The above block form is ensured if only the error distribution $H(x/\sigma)$ is symmetric around 0.

(2) The influence function for a general class of robust regression estimators can be found in Maronna, Bustos and Yohai (1979).

(3) The influence function gives a description of the bias introduced by infinitesimal perturbations of an ideal model. It turns out that *a bounded influence is desirable* (see Hampel (1971)). Under the strong symmetry conditions we have

$$\sup_{z_0} \{|IF_{b,G,H}(z_0)|\} = \frac{\sigma}{2} \text{ const}(G,H) \sup_{c} \{cu(c)\}.$$

Proof: see (2.7) and note that $W(G,H) = \text{diag}(w_1,\ldots,w_p,\sigma)$.

Our estimator ((2.2) and (2.3)) has therefore a bounded influence function provided the weight function u(c) goes to zero fast enough as c gets large. This is not surprising since the weight depends on both the size of the residual as well as the distance form the center in carrier space.

The asymptotic normality of our estimates can be proved using a standard result due to Huber (1967). The proof follows the same

steps as Maronna (1976). Unfortunately asymptotic normality is only proved under the extremely stringent condition that the joint distribution of x and y is spherically symmetric up to an affine transformation. This is an assumption which can possibly be relaxed, but another proof is required. The other regularity conditions needed are boundedness and monotonicity conditions on u(c) and cu(c), respectively. Under all of these assumptions it is true that the estimate b_j (see (2.3)) is asymptotically Gaussian with mean β_j and asymptotic variance

$$\frac{E_{G,H}\,[u^2(d^2)\; r^2\; x_j^{\;2}]}{a_j^2} \qquad \text{(see (2.7))},$$

the components of b furthermore are independent.

Maronna (1976) showed that the breakdown properties of covariance estimates of the form (2.2) are not encouraging. As the dimensionality of the z-space increases, i.e. if we add more carriers, the breakdown point of W^* (see (2.2)) is necessarily decreasing like 1/(1+dimensionality). The regression estimator (2.3) which is based on W^* has the same breakdown point. If we add a contamination along an arbitray regression line and let the contaminating carrier value go to infinity the estimator breaks down.

We have discussed above two types of properties. First the behavior at the presumed central model which is described by the asymptotic parameters. And second the breakdown point which is a

simple indicator for the behavior under severe deviations from the central model. The asymptotic influence curve stands somewhat inbetween. On one hand it provides a decompositon of the asymptotic variance at the central model, on the other hand it is useful as an indicator for the behavior under perturbations of the linearity assumption.

In the next section we will use the formulas derived above to give us numerical values. But these asymptotic findings are of course of secondary importance. What we really need to know is the small sample behavior. To see how well asymptotics predicts the corresponding small sample values some simulation results for samples of size 20 are included.

3. COMPARISON OF REGRESSION ESTIMATORS

In order to compare the various estimators discussed in Section 2 we plan to do a small Monte Carlo study. To further simplify the questions involved we will restrict attention to the simple case p=1, i.e. to the model

$$y_i = \beta x_i + \varepsilon_i \quad i=1,\ldots,n \tag{3.1}$$

with random carriers (independent of the errors). For model (3.1) the questions of how to handle a constant term does not appear and (2.2) together with (2.3) defines a reasonable estimator of β. To fully specify this estimator we need to define a weight function $u(c)$. A somewhat typical choice is

$$u(c) = \frac{\psi_k(c)}{c(1-\exp(-k/2))} \qquad (3.2)$$

where $\psi_k(c) = \max(-k,\min(c,k))$ denotes Huber's ψ-function. The justification for this paricular form of a weight function lies in the fact that near the Gaussian error model a good behavior of the asymptotic efficiency can be expected. The weight function is chosen in such a way that the resulting estimator is Fisher consistent at the Gaussian model.

We still have the constant k at our disposal. It determines the trade-off between resistance and efficiency. For k=5.77 the asymptotic variance of our estimator of β at the standard Gaussian model (i.e. both the error distribution and the carrier distribution are standard Gaussian) is 1.05 -- or the efficiency is 95%. This appears to be a reasonable normalization.

3.1 ASYMPTOTIC NUMBERS

The sampling situation we take into consideration are polyGaussians of the form

$$(1-\varepsilon)\ N(0,1) + \varepsilon\ N(0,\tau^2) =: (\varepsilon/\tau^2) \qquad (3.3)$$

and will be used both as a model for errors as well as carriers.

Table 3.1 contains the values of the asymptotic variance for the regression estimator under consideration.

Table I: asymptotic variance of b (see (2.3)), (ε_1/τ_1^2) refers to the carrier distribution, (ε_2/τ_2^2) denotes the error distribution (see (3.3))

ε_1/τ_1^2 \ ε_2/τ_2^2	N(0,1)	0.1/9	0.1/25	0.05/100
N(0,1)	1.05*	1.33	1.39	1.16
0.1/9	0.71	0.90	0.96	0.79
0.1/25	0.55	0.69	0.74	0.61
0.05/100	0.77	0.97	1.03	0.86

(These values were computed numerically with a 24-point Gaussian quadrature procedure)

*: normalization as discussed above

Since the covariance estimator (2.3) has bounded influence if used with the weight function (3.2) its asymptotic behavior depends on the distribution of the carriers. Table I shows that this dependence is quite remarkable. The first column combines standard Gaussian errors with increasingly heavy tailed carrier distributions. The least squares estimators are of course optimal in these situations with asymptotic variances of 1.00, 0.56, 0.29, and 0.17 . The asymptotic efficiency of our estimator in these situations is therefore 95%, 79%, 53%, and 22%. It appears that the

normalization of bounded influence regression estimators is a non-trivial matter if we want to achieve a fair comparison with estimators which do not treat outliers in carrier space any differently. Most of the information about the slope parameter β (see (3.1)) in the first column of Table 3.1 lies exactly in the points far out in carrier space. Disregarding this information results in a big loss.

Maronna, Bustos and Yohai (1979) give tables similar to Table I for a Mallows-type estimator and the Hampel-Krasker estimator. Table II shows the numbers for the Mallows estimator -- the Hampel-Krasker estimator behaves similarly, but a bit worse overall because of our normalization (the Hampel-Krasker estimator is the "most robust" if we normalize this way!).

Table II: asymptotic variance of Mallows estimator with Huber weight function, (ε_1/τ_1^2) refers to the carrier distribution, (ε_2/τ_2^2) denotes the error distribution (see (3.3))

	ε_2/τ_2^2			
ε_1/τ_1^2	N(0,1)	0.1/9	0.1/25	0.05/100
N(0,1)	1.05	1.42	1.61	1.42
0.1/9	0.66	0.86	0.98	0.86
0.1/25	0.46	0.59	0.66	0.59

0.05/100	0.36	0.45	0.51	0.46

Table III contains the efficiencies of the Mallows estimator with respect to the covariance estimator (2.3).

Table III: (asymptotic variance of covariance estimator)/(asymptotic variance of Mallows estimator), (ε_1/τ_1^2) refers to the carrier distribution, (ε_2/τ_2^2) denotes the error distribution (see (3.3))

	ε_2/τ_2^2			
ε_1/τ_1^2	N(0,1)	0.1/9	0.1/25	0.05/100
N(0,1)	1.00	1.00	0.97	0.92
0.1/9	1.06	1.07	1.05	0.99
0.1/25	1.17	1.15	1.15	1.07
0.05/100	2.03	2.02	1.98	1.89

As the carrier distribution gets heavier tails the covariance estimator is outperformed by the Mallows estimator. But in most cases the relative loss compared to the Mallows estimator is not serious.

It should be noted that the robust estimator based on covariance is not intended as a replacement for other bounded

influence fitters. Its weight function (see (2.4)) has the appealing feature that the response variable and the carriers are interchangable. All possible choices of "the response" can be easily fitted. In some situations this can be helpful.

Table IV finally shows the gross-error sensitivities of the different estimators together with their asymptotic breakdown points. The gross-error sensitivity is defined as the maximum of the (Euclidian norm) of the influence curve (see Hampel (1974)) and therefore serves as a description of the behavior under deviations from the central model. In a sense we could call these our tools of measuring the resistance of the estimators.

Table IV: gross-error sensitivity and breakdown point under standard Gaussian error distribution, $(\varepsilon_1, \tau_1^2)$ refers to the carrier distribution (see (3.3))

	estimators		
	Covariance	Mallows	Hampel-Krasker
	breakdown points		
	0.16	0.31	0.50
ε_1/τ_1^2	gross error sensitivities		
N(0,1)	3.34	3.40	2.94
0.1/9	2.86	2.66	2.22
0.1/25	2.63	2.15	1.73

0.05/100	2.99	2.07	1.64

(for the last two columns see Maronna, Bustos and Yohai (1979), Table 4)

The asymptotic optimality property of the Hampel-Krasker estimator is clearly visible in Table IV -- that particular estimator minimizes the gross-error sensitivity under a fixed asymptotic variance (all of this at the central model). This is exactly the reason why we are interested in using it.

3.2 RESULTS OF A SIMULATION EXPERIMENT

Table V contains empirical variances of the regression estimators (2.3) for samples of size 20. The weight function u() is described at the beginning of Section 3 and the model is described in (3.3).

Table V: empirical variances of $n^{1/2}$ b (see (2.3)) for samples of size n=20, (ε_1/τ_1^2) refers to the carrier distribution, (ε_2/τ_2^2) denotes the error distribution (see (3.3))

ε_2/τ_2^2

ε_1/τ_1^2	N(0,1)	0.1/25	0.05/100
N(0,1)	1.17	1.88	1.61
0.1/25	0.69	1.16	0.94
0.05/100	1.14	1.34	1.14
0.1/100	0.57	0.94	--

(The numbers in this table are based on 500 replicas and the estimated relative errors are between 9% and 15%)

A comparison with Table I shows that the asymptotic variance is in all situations lower than the actual small sample variance. But otherwise the pattern in these estimated variances follows the behavior of Table I. We do not have the corresponding numbers for the Mallows estimator. A look at Table 5 in Maronna, Bustos and Yohai (1979) shows, however, that the step from asymptotic variances to small sample values very closely follows the behavior we have seen in our numbers. It therefore appears that the asymptotic theory is a good enough approximation to the small sample (n=20) behavior.

As pointed out in Maronna, Bustos and Yohai (1979) the distribution of the estimated regression coefficient is apparently non-Gaussian in a lot of sampling situations. This makes the comparison between different estimators more difficult since the variance -- or mean-square-error -- might not be the criterion to

use. From our simulation studies it appears that the distribution of the slope estimator is heavier tailed that a Gaussian. This is true in the case of the covariance estimator for either a "contaminated" carrier distribution or a heavier tailed error distribution.

We would finally like to point to a future research topic. Donoho and Huber have recently pointed out the intuitive appeal of the small sample breakdown point (see Donoho and Huber (1983) and Huber (1984)) and it would be useful to study this aspect of our estimators. In the case of regression these studies will not be simple, however, since the breakdown point will presumably depend on the actual carrier values.

4. CONCLUSIONS AND AN EXAMPLE

We have discussed an approach to robust regression estimation based on robust covariance estimation. We have seen in Section 2 that if we adopt an affine equivariant M-estimator for the covariance side, we do nothing else but a weighted least squares approach on the regression side. The weights depend on two quantities:

(1) standardized residual

(2) standardized norm in the carrier space.

Both of these are estimated robustly and the sum of their squares is the quantity that matters. This particular form of the weight

function is rather unique and symmetric with respect to the notions "response" and "explanatory". This symmetry makes the choice attractive. The covariance estimator fits a linear model to the bulk of the data. Points far out in the space of explanatory variables are identified as not belonging to the majority of the data just as are points which do not fit the linear structure well. This is the primary difference to other bounded influence methods which try to identify only the second class of outliers.

Our particular estimator seems to downweights points far out in the carrier space too much. This can possibly be corrected by using a different weight function.

Let us finally look at an example. We choose the stack loss data given in Daniel and Wood (1982, p. 61) involving three independent variables

x_1 = air flow

x_2 = cooling water temperature

and x_3 = acid concentration.

The response variable is the stack loss. The sample size n is 21. For our covariance estimator we use the weight function $u(x) = \varphi_k(x)/x$ $(x \geq 0)$, where $\varphi_k(x) = \min(k, \max(-k,x))$ is Huber's φ-function. The tuning constant k is chosen as 5.0.

On that particular data set the weights determined by the covariance estimator after 20 iterations turn out to be small for

the first four and the last point. Figure I shows a plot of the residuals of this fit against the first variable x_1. The weight is indicated by the area of the point. We can identify the 5 outlying points and we can also appreciate the comment made by Daniel & Wood that one of these points fits the plane determined by the rest quite well.

The final conclusion of Daniel and Wood is that X_1^2 (or $x_1 x_2$) should be included and that x_3 can be dropped. Our analysis supports those findings and in fact recovers the final fit given by Daniel & Wood.

BIBLIOGRAPHY:

Daniel C. and Wood F.S. (1980): Fitting Equations to Data, Wiley, New York.

Donoho D. and Huber P.J. (1983): "The notion of breakdown point", in a Festschrift for Erich L. Lehmann, P. Bickel et al, edts., Wadsworth, Belmont CA

Hampel, F.R. (1971): "A general qualitative definition of robustness". Ann. Math. Statist., 42, 1887-1896

Hampel, F.R. (1974): "The influence curve and its role in robust estimation". J. Amer. Statist. Assoc., 62, 1179-1186

Handschin, E., Kohlas, J., Fiechter, A. and Schweppe, F. (1975): "Bad data analysis for power system state estimation". IEEE Transactions on Power Apparatus and Systems, PAS-94, 2, 329-337

Huber, P.J. (1967): "The behavior of maximum likelihood estimates under nonstandard conditions". Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, Univ. of California Press, Berkeley

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

Huber, P.J. (1981): Robust Statistics. Wiley, New York

Huber, P.J. (1984): "Finite sample breakdown of M- and P-estimators". Annals Statist., 12, 119-126

Krasker, W.S. and Welsch, R.E. (1982): "Efficient bounded influence regression estimation". J. Amer. Statist. Assoc., 77, 595-605

Mallows, C.L. (1975): "On some topics in robustness". unpublished memorandum, Bell Telephone Laboratories, Murray Hill, N.J.

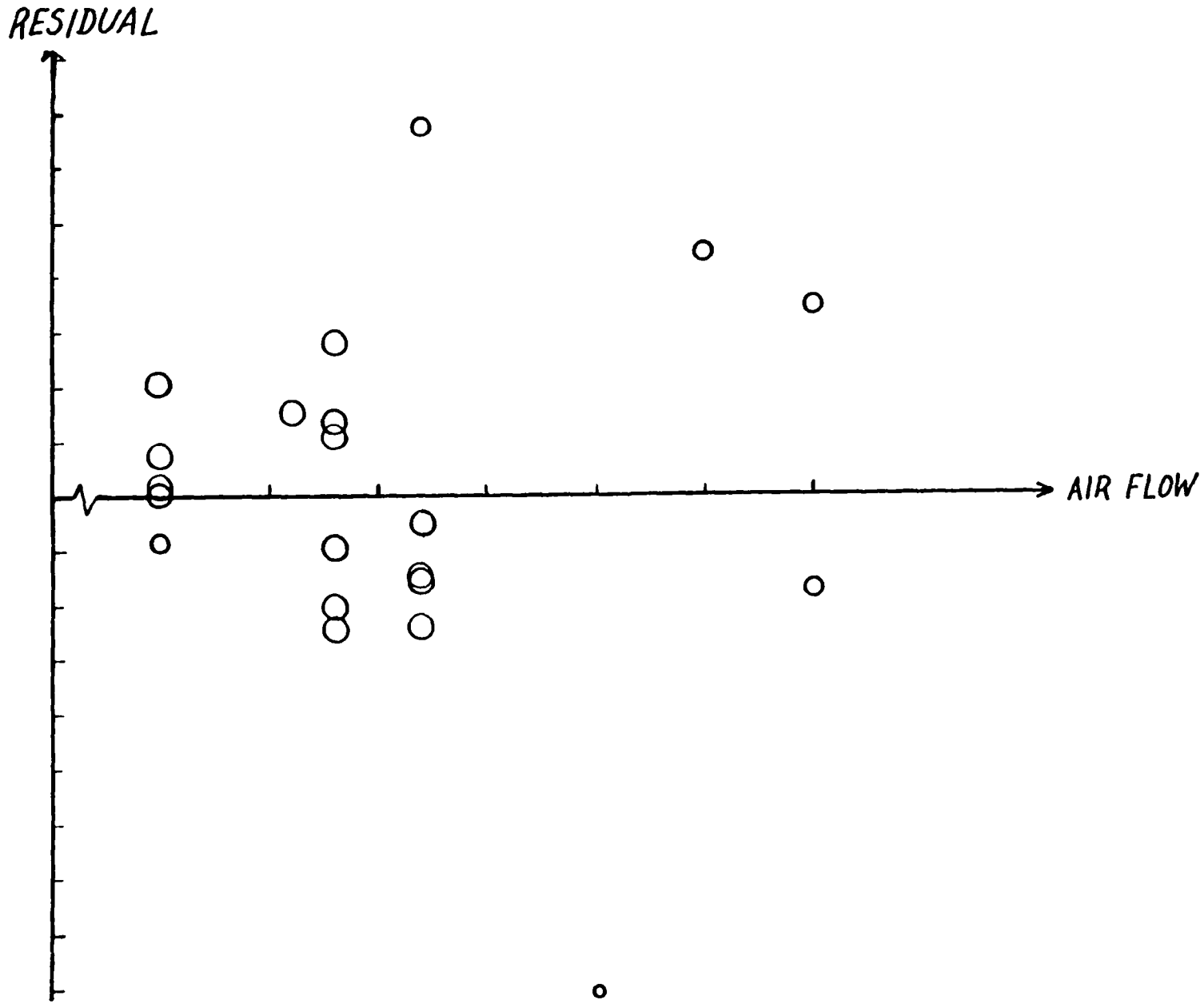
Maronna, R.A. (1976): "Robust M-estimators of multivariate location and scatter". Ann. Statist., 4, 51-67

Maronna, R.A. , Bustos O., and Yohai V. (1979): "Smoothing Techniques for Curve Estimation". in Lecture Notes in Mathematics, No. 757, Springer, Heidelberg

Mosteller, F. and Tukey, J.W. (1977): Data Analysis and Regression: A second course in Statistics. Addison-Wesley, Reading, Massachusetts

Stahel, W. (1981): "Robust estimation: infinitesimal optimality and covariance matrix estimators". unpublished PhD thesis, Swiss Federal Inst. of Technology, Zurich, Switzerland

FIGURE I: Residual against x_1 for the stack loss data after fitting y on x_1, x_2 and x_3 with the covariance estimator. The weights are proportional to the area of the dots.



UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

MASTER COPY - FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARO 21779.2-MA	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (and Subtitle) Robust Regression Through Robust Covariances		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Stephan Morgenthaler and Ricardo Maronna		8. CONTRACT OR GRANT NUMBER(s) DAAG29-84-K-0207
9. PERFORMING ORGANIZATION NAME AND ADDRESS Statistics Center Massachusetts Institute of Technology Cambridge, Massachusetts 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS N/A
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE January 1985 13. NUMBER OF PAGES 27
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Bounded Influence; asymptotics; and simulation.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper describes a regression estimator which builds on the investigation of the covariance structure in the full space of explanatory variables and response variable. It is robust since it down weights outlying observations. A comparison to other robust estimators is included. → Send hyperlink to include: --

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

FILMED

5-85

DTIC