ELSEVIER

Statistics & Probability Letters 46 (2000) 287–299

STATISTICS & PROBABILITY LETTERS

www.elsevier.nl/locate/stapro

Jackknifing, weighting, diagnostics and variance estimation in generalized M-estimation

Zhiyi Du[a], Douglas P. Wiens[b,*,1]

[a] *Department of Mathematics and Statistics, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario, Canada K1S 5B6*
[b] *Statistics Centre, Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1*

Received February 1999; received in revised form May 1999

Abstract

We study and compare methods of covariance matrix estimation, and some diagnostic procedures, to accompany generalized ("Bounded Influence") M-estimation of regression in the linear model. The methods derive from one-step approximations to the delete-one estimates of the regression parameters. Two weighting schemes are also compared. The comparisons are made through a simulation study and a case study. The jackknife-based covariance estimates are successful at improving the coverages of associated confidence intervals. One of the weighting schemes is found to be quite generally superior to the other, with respect to mean-squared error and to confidence interval coverage, on data containing a realistic proportion of outliers and high leverage points.

MSC: primary 62F35; 62J05; secondary 62J20

Keywords: Bounded influence M-estimation; Finite sample correction; Iteratively reweighted least squares; Least median of squares; Least trimmed squares; Minimum volume ellipsoid; Regression; Robustness

1. Introduction

In this article we address several problems concerning generalized M-estimation (GM-estimation) of regression in the linear model. The first is the derivation of accurate estimates of the covariance matrix of the regression estimates. In Section 2 below we propose weighted jackknife methods of covariance estimation, based on approximate delete-one estimates of the regression parameters. The second problem, studied in Section 3, is the choice of weights used to bound the influence of outliers in the factor space. Two measures of influence for GM-regression are proposed in Section 4. These are based on the same delete-one estimates

* Corresponding author. Tel.: +1-780-492-4406; fax: +1-780-492-6823.
E-mail addresses: zdu@math.carleton.ca (Z. Du), doug.wiens@ualberta.ca (D.P. Wiens)

[1] This research is supported by the Natural Sciences and Engineering Research Council of Canada.

PII: S0167-7152(99)00117-0

as are our covariance estimates. All of these proposals are evaluated, by simulation and in a case study, in Section 5.

We assume a model of the form $y_i = \boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\theta} + \sigma\varepsilon_i$, where $\boldsymbol{\theta}$ is a p-dimensional parameter vector, $\sigma > 0$ is a scale parameter and the ε_i are zero-mean, uncorrelated random errors. When the model contains an intercept term we write $\boldsymbol{z}_i = (1, \boldsymbol{x}_i^{\mathrm{T}})^{\mathrm{T}}$ and $q = p - 1$; otherwise we define $\boldsymbol{x}_i = \boldsymbol{z}_i$ and $q = p$. The q-dimensional vectors $\boldsymbol{x}_i$ may themselves be random, in which case we assume that they are distributed independently of the errors. We adopt a representation of the estimates as in Simpson and Chang (1997):

$$\hat{\boldsymbol{\theta}}_{\mathrm{GM}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \rho\left(\frac{y_i - \boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\theta}}{S w_i^{\alpha}}\right) w_i^{1+\alpha}, \tag{1}$$

where ρ is a specified function (even, absolutely continuous, non-decreasing on $[0, \infty)$), S is an estimate of scale and $w_i = w(\boldsymbol{x}_i)$ are weights chosen by the statistician. We normalize the weights by $n^{-1}\sum w_i = 1$.

Common choices of α are $\alpha = 0, 1, -1$, yielding estimates of the Mallows, Schweppe and Hill–Ryan types, respectively — see the discussion in Hampel et al. (1986, pp. 315–316). To estimate scale we use the modified median absolute deviation:

$$S = \frac{\mathrm{med}_{1 \leqslant i \leqslant n}\{|y_i - \boldsymbol{z}_i^{\mathrm{T}}\hat{\boldsymbol{\theta}}|\}}{\Phi^{-1}(0.75)}.$$

The standard normal quartile $\Phi^{-1}(0.75)$ is included for consistency under normal errors.

With $\psi = \rho'$, (1) yields the equation

$$\sum_{i=1}^{n} \psi\left(\frac{y_i - \boldsymbol{z}_i^{\mathrm{T}}\hat{\boldsymbol{\theta}}}{S w_i^{\alpha}}\right) \boldsymbol{z}_i w_i = \boldsymbol{0}. \tag{2}$$

Simpson et al. (1992) proposed a k-step (Newton–Raphson) solution to (2) using $\alpha = 0$ and recommended $k = 3$. As a starting value they took the least median of squares (LMS) estimator (Rousseeuw, 1984). Coakley and Hettmansperger (1993) proposed a k-step solution using $\alpha = 1$ and the initial least trimmed squares (LTS) estimator (Rousseeuw, 1984). In both cases, the high breakdown point of the initial estimator, and the asymptotic properties of the fully iterated solution to (2), are inherited by the k-step estimate if k is fixed in advance.

The usual variance/covariance estimates are obtained as follows. Define residuals $e_i = y_i - \boldsymbol{z}_i^{\mathrm{T}}\hat{\boldsymbol{\theta}}$, standardized residuals $r_i = e_i/S$ and matrices

$$\begin{aligned} \boldsymbol{P} &= \sum_{i=1}^{n} \psi'\left(\frac{r_i}{w_i^{\alpha}}\right) w_i^{1-\alpha} \boldsymbol{z}_i \boldsymbol{z}_i^{\mathrm{T}}, \\ \boldsymbol{Q} &= \sum_{i=1}^{n} \psi^2\left(\frac{r_i}{w_i^{\alpha}}\right) w_i^{2} \boldsymbol{z}_i \boldsymbol{z}_i^{\mathrm{T}}. \end{aligned} \tag{3}$$

Under appropriate and commonly assumed regularity conditions — see Maronna and Yohai (1981) and Wiens (1996) — the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ is consistently estimated by

$$\boldsymbol{C} = S^2 \cdot \boldsymbol{P}^{-1}\boldsymbol{Q}\boldsymbol{P}^{-1}. \tag{4}$$

When $\alpha = 0$ a commonly used alternative is the 'exchangeable' estimate

$$\boldsymbol{C}_{\mathrm{exch}} = S^2 \cdot \boldsymbol{P}_{\mathrm{exch}}^{-1}\boldsymbol{Q}_{\mathrm{exch}}\boldsymbol{P}_{\mathrm{exch}}^{-1}, \tag{5}$$

where

$$\boldsymbol{P}_{\text{exch}} = \frac{1}{n}\sum_{i=1}^{n}\psi'(r_i)\sum_{i=1}^{n} w_i \boldsymbol{z}_i \boldsymbol{z}_i^{\mathrm{T}},$$

$$\boldsymbol{Q}_{\text{exch}} = \frac{1}{n-p}\sum_{i=1}^{n}\psi^2(r_i)\sum_{i=1}^{n} w_i^2 \boldsymbol{z}_i \boldsymbol{z}_i^{\mathrm{T}}.$$

Our experience has been that both (4) and (5) can be improved by finite-sample corrections as proposed in Section 2.

Weights $w_i \equiv 1$ give an ordinary M-estimate, which has a breakdown point of 0 in the presence of high leverage points. The weights in GM-estimation are thus chosen so as to downweight highly influential points in the factor space. The hat-matrix weights $w_i \propto (1-h_{ii})^{1/2}$ (Handshin et al., 1975) or $w_i \propto (1-h_{ii})/\sqrt{h_{ii}}$ (Welsch, 1980) clearly have this property. However, the leverages h_{ii} are well known to be highly susceptible to the masking effect of clusters of outliers in the factor space. The linear relationship between the leverages and the classical Mahalanobis distance defined via the sample mean and covariance matrix of the regressors suggests weighting schemes based on the Robust Mahalanobis distance (Rousseeuw and Leroy, 1987)

$$\mathrm{RM}_i^2 = (\boldsymbol{x}_i - \boldsymbol{m}_X)^{\mathrm{T}} \boldsymbol{C}_X^{-1} (\boldsymbol{x}_i - \boldsymbol{m}_X),$$

where $\boldsymbol{m}_X$ and $\boldsymbol{C}_X$ are the minimum volume ellipsoid (MVE) estimates of location and scatter.

The weights w_i used by Simpson et al. (1992) and by Coakley and Hettmansperger (1993) are

$$w^{(0)}(\boldsymbol{x}_i;\beta) \propto \min\left[1, \frac{\chi_q^2(\beta)}{\mathrm{RM}_i^2}\right], \tag{6}$$

where $\chi_q^2(\beta)$ is a suitably chosen β-quantile of the χ_q^2 distribution. A choice motivated in part by the considerations in the preceding paragraph, and in part by a study of optimal weights for fixed design regression models (Wiens, 1999) is that of weights w_i given by

$$w^{(1)}(\boldsymbol{x}_i;\gamma) \propto (1+\gamma^2 \mathrm{RM}_i^2)^{-1/2} \tag{7}$$

for a suitably chosen parameter γ. Simpson and Chang (1997) studied the behaviour of k-step GM-estimates, with the kth Newton–Raphson step followed by one step of iteratively reweighted least-squares (IRLS). In this study the GM-weights used were $w^{(1)}(\boldsymbol{x}_i; 1/\sqrt{2})$; this choice was essentially ad hoc but somewhat motivated by the smoothness properties of these weights (Simpson, personal communication). Our simulation studies indicate that weights $w^{(1)}(\cdot;\gamma)$ can result in significantly reduced mean-squared errors of estimation, relative to $w^{(0)}(\cdot;\beta)$.

A wealth of diagnostic procedures, and in particular influence measures, has long existed for least-squares regression analyses; recently some of these methods have been extended to rank-based robust estimates of regression (McKean et al., 1990) and to M- and GM-estimates of regression (Cook and Weisberg, 1982; McKean et al., 1993; Simpson and Chang, 1997). Two influence measures based on the delete-one regression estimates are proposed in Section 4 below.

2. Covariance estimation

To obtain approximate (one-step) delete-one estimates we approximate the solutions to the equations

$$\sum_{j\neq i}\psi\left(\frac{y_j - \boldsymbol{z}_j^{\mathrm{T}}\boldsymbol{\theta}}{S w_j^{\alpha}}\right)\boldsymbol{z}_j w_j = \boldsymbol{0}, \quad i = 1,\ldots,n$$

by performing one step of Newton's method, starting with $\hat{\boldsymbol{\theta}}_{\mathrm{GM}}$. These approximate solutions are

$$\hat{\boldsymbol{\theta}}_{-i} = \hat{\boldsymbol{\theta}}_{\mathrm{GM}} - S \cdot \boldsymbol{P}^{-1}\boldsymbol{z}_i \frac{\psi(r_i/w_i^{\alpha})w_i}{1-p_i}, \tag{8}$$

where

$$p_i = \psi'\left(\frac{r_i}{w_i^{\alpha}}\right) w_i^{1-\alpha}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{P}^{-1}\boldsymbol{z}_i\,.$$

Note that $\sum_{i=1}^{n} p_i = p$; if ψ is monotone then $0 < p_i < 1$. In these and other respects the p_i play a role similar to the hat-matrix leverages h_{ii}. For ordinary M-estimates, (8) is given in Cook and Weisberg (1982, p. 202).

One can proceed to obtain a weighted jackknife estimate similar to that of Hinkley (1977). First compute weighted pseudovalues $\boldsymbol{\theta}^i = \hat{\boldsymbol{\theta}}_{\mathrm{GM}} + n(1-p_i)(\hat{\boldsymbol{\theta}}_{\mathrm{GM}} - \hat{\boldsymbol{\theta}}_{-i})$; the weighted jackknife estimate is then $\hat{\boldsymbol{\theta}}_{\mathrm{WJ}} = (1/n)\sum \boldsymbol{\theta}^i \approx \hat{\boldsymbol{\theta}}_{\mathrm{GM}}$. This approximation is exact if $\hat{\boldsymbol{\theta}}_{\mathrm{GM}}$ is an exact solution to (2). In the spirit of Wu (1986, exhibit 5.1) we propose the weighted jackknife covariance estimate

$$\begin{aligned}\boldsymbol{C}_{\mathrm{J}} &= \sum_{i=1}^{n}(1-p_i)(\hat{\boldsymbol{\theta}}_{-i} - \hat{\boldsymbol{\theta}}_{\mathrm{GM}})(\hat{\boldsymbol{\theta}}_{-i} - \hat{\boldsymbol{\theta}}_{\mathrm{GM}})^{\mathrm{T}} \\ &= S^2 \cdot \boldsymbol{P}^{-1}\boldsymbol{Q}_{\mathrm{J}}\boldsymbol{P}^{-1},\end{aligned} \tag{9}$$

where

$$\boldsymbol{Q}_{\mathrm{J}} = \sum_{i=1}^{n} \frac{\psi^2(r_i/w_i^{\alpha})w_i^2}{1-p_i}\boldsymbol{z}_i\boldsymbol{z}_i^{\mathrm{T}}.$$

Wu (1986) chose the OLS-based version of this estimate for its favourable properties with respect to bias, when the errors are heteroscedastic.

For Mallows estimates ($\alpha = 0$) an option is to adjust (9) as

$$\boldsymbol{C}_{\mathrm{J,adj}} = S^2\left(\frac{n_{\mathrm{adj}}}{n}\right)^2 \cdot \boldsymbol{P}_{\mathrm{adj}}^{-1}\boldsymbol{Q}_{\mathrm{J,adj}}\boldsymbol{P}_{\mathrm{adj}}^{-1}, \tag{10}$$

where

$$n_{\mathrm{adj}} = \sum_{i=1}^{n} I(\psi'(r_i) > 0),$$

$$\boldsymbol{P}_{\mathrm{adj}} = \frac{1}{n}\sum_{i=1}^{n}\psi'(r_i)\sum_{i=1}^{n} I(\psi'(r_i) > 0)w_i\boldsymbol{z}_i\boldsymbol{z}_i^{\mathrm{T}},$$

$$\boldsymbol{Q}_{\mathrm{J,adj}} = \frac{1}{n-p}\sum_{i=1}^{n}\psi^2(r_i)\sum_{i=1}^{n}\frac{w_i^2\boldsymbol{z}_i\boldsymbol{z}_i^{\mathrm{T}}}{1-p_{\mathrm{adj},i}},$$

$$p_{\mathrm{adj},i} = \psi'(r_i)I(\psi'(r_i) > 0)w_i \cdot \boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{P}_{\mathrm{adj}}^{-1}\boldsymbol{z}_i\,.$$

The adjustment of $\boldsymbol{P}$ is an attempt to account for finite-sample bias introduced by assuming that $E[\psi'(\varepsilon)w(\boldsymbol{x})\boldsymbol{z}(\boldsymbol{x})\boldsymbol{z}(\boldsymbol{x})^{\mathrm{T}}] = E[\psi'(\varepsilon)] \cdot E[w(\boldsymbol{x})\boldsymbol{z}(\boldsymbol{x})\boldsymbol{z}(\boldsymbol{x})^{\mathrm{T}}]$ and then estimating the latter (by $\boldsymbol{P}_{\mathrm{exch}}/n$), rather than the former. If ψ is strictly increasing, then $\boldsymbol{P}_{\mathrm{adj}} = \boldsymbol{P}_{\mathrm{exch}}$ and $p_{\mathrm{adj},i} = p_i$. The option of replacing the indicator $I(\psi'(r_i) > 0)$ by the sign of $\psi'(r_i)$ was investigated, but all too often led to breakdown through the near singularity of the resulting matrix, when used with a redescending ψ. The choice $(\boldsymbol{P}_{\mathrm{exch}}, \boldsymbol{Q}_{\mathrm{J}})$ rather then

$(\boldsymbol{P}_{\text{adj}}, \boldsymbol{Q}_{\text{J,adj}})$ was also considered; it gave acceptable results when used with weights $w^{(0)}(\cdot;\beta)$ but resulted in low-confidence interval coverages when used with the weights $w^{(1)}(\cdot;\gamma)$.

3. Weights

For purposes of comparison we shall choose γ^2 in (7) and β in (6) in order to attain a specified asymptotic efficiency, relative to unit weights, in the linear regression model with regressors $\boldsymbol{z}(\boldsymbol{x}) = (1, \boldsymbol{x}_{q\times 1}^{\mathrm{T}})^{\mathrm{T}}$, $\boldsymbol{x}_i \sim$ i.i.d. $\mathrm{N}(\boldsymbol{0}, \boldsymbol{I}_q)$. We shall do this for $\alpha = 0$ only, the choice then being independent of the choice of ψ. For weights $w(\boldsymbol{x})$, make the definition

$$i_{a,b} := E_{\Phi}[\|\boldsymbol{x}\|^{2a} w^{2b}(\boldsymbol{x})] = \begin{cases} E\left[Z^a \min\left(1, \dfrac{\chi_q^2(\beta)}{Z}\right)^{2b}\right], & w = w^{(0)}, \\ E\left[\dfrac{Z^a}{(1+\gamma^2 Z)^b}\right], & w = w^{(1)}, \end{cases}$$

where $Z \sim \chi_q^2$. Replace $n^{-1}\boldsymbol{P}_{\text{exch}}$ and $n^{-1}\boldsymbol{Q}_{\text{exch}}$ by their asymptotic values $E[\psi'(\varepsilon)] \cdot E[w(\boldsymbol{x})\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}}]$ and $E[\psi^2(\varepsilon)] \cdot E[w^2(\boldsymbol{x})\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}}]$, respectively. Then $\boldsymbol{P}_{\text{exch}}^{-1}\boldsymbol{Q}_{\text{exch}}\boldsymbol{P}_{\text{exch}}^{-1}$ becomes proportional to

$$\boldsymbol{C}_\gamma := \frac{i_{0,1}}{i_{0,1/2}^2} \oplus q\frac{i_{1,1}}{i_{1,1/2}^2}\boldsymbol{I}_q.$$

For the loss functions $l_A(\boldsymbol{C}_\gamma) = \operatorname{trace} \boldsymbol{C}_\gamma$ and $l_D(\boldsymbol{C}_\gamma) = |\boldsymbol{C}_\gamma|^{1/(q+1)}$ the efficiencies of $w^{(0)}(\cdot;\beta)$ and $w^{(1)}(\cdot;\gamma)$ relative to constant weights ($\beta = 1, \gamma = 0$) are

$$e_A = \frac{l_A(\boldsymbol{C}_0)}{l_A(\boldsymbol{C}_\gamma)} = (q+1)\left(\frac{i_{0,1}}{i_{0,1/2}^2} + q^2\frac{i_{1,1}}{i_{1,1/2}^2}\right)^{-1}$$

and

$$e_D = \frac{l_D(\boldsymbol{C}_0)}{l_D(\boldsymbol{C}_\gamma)} = \left[\frac{i_{0,1}}{i_{0,1/2}^2} \cdot \left(q\frac{i_{1,1}}{i_{1,1/2}^2}\right)^q\right]^{-1/(q+1)}.$$

Table 1 gives some particular values of γ^2 and β, for relative efficiencies of 0.90 and 0.95. Note that the A- and D-relative efficiencies are attained at almost identical values, in each case. Table 2 gives the limiting efficiencies as $\gamma^2 \to \infty$ or as $\beta \to 0$. We use the equivalent weights $\gamma w^{(1)}(\boldsymbol{x}_i;\gamma)$ and $w^{(0)}(\boldsymbol{x}_i;\beta)/\chi_q^2(\beta)$, which tend to RM_i^{-1} and RM_i^{-2}, respectively, to evaluate the asymptotic values. For $w^{(1)}(\cdot;\gamma)$ and hence for $w_i = \sqrt{w^{(0)}(\boldsymbol{x}_i;\beta)}$ (a suggestion of Simpson et al. (1992)) and moderately large q these cannot fall much below unity.

Table 1
Values of γ^2 and β for specified asymptotic relative efficiencies of $w^{(1)}(\cdot;\gamma)$ and $w^{(0)}(\cdot;\beta)$

	γ^2				β			
	e_D		e_A		e_D		e_A	
q	0.9	0.95	0.9	0.95	0.9	0.95	0.9	0.95
1	1.798	0.643	1.798	0.643	0.804	0.890	0.807	0.891
2	1.817	0.620	1.816	0.620	0.718	0.838	0.720	0.839
3	2.251	0.600	2.247	0.600	0.643	0.793	0.644	0.793
4	3.832	0.629	3.811	0.629	0.577	0.751	0.577	0.751
5	24.921	0.698	23.736	0.698	0.516	0.711	0.517	0.712

Table 2
Limiting efficiencies of $w^{(1)}(\cdot;\gamma)$ as $\gamma^2 \to \infty$ and (in parentheses) of $w^{(0)}(\cdot;\beta)$ as $\beta \to 0$

	$q \leqslant 2$	$q = 3$	$q = 4$	$q = 5$	$q = 6$	$q = 10$	$q = 20$
e_D	0.000 (0.000)	0.790 (0.000)	0.863 (0.000)	0.896 (0.529)	0.915 (0.636)	0.950 (0.795)	0.975 (0.900)
e_A	0.000 (0.000)	0.784 (0.000)	0.862 (0.000)	0.895 (0.544)	0.915 (0.640)	0.950 (0.795)	0.975 (0.900)

4. Diagnostic measures

In this section we report a large sample approximation to the variance of the residual, leading to a form of studentized residual. The result is as in McKean et al. (1990,1993); we modify it somewhat by incorporating the same flexibility in the choices of $\boldsymbol{P}$ and $\boldsymbol{Q}$ as in the estimator being evaluated. We then propose two measures of influence for GM-estimates based on the approximate delete-one estimates.

4.1. Studentized residuals

Using the methods of McKean et al. (1990, 1993) we find that a large-sample estimate of VAR$[e_i]$, ignoring terms which are $o_p(n^{-1})$, is

$$S_i^2 = S^2 \left[1 - 2w_i \boldsymbol{z}_i^{\mathrm{T}} \boldsymbol{P}^{-1} \boldsymbol{z}_i \frac{1}{n-p} \sum_{j=1}^{n} \psi\left(\frac{r_j}{w_j^{\alpha}}\right) r_j + \boldsymbol{z}_i^{\mathrm{T}} \boldsymbol{P}^{-1} \boldsymbol{Q} \boldsymbol{P}^{-1} \boldsymbol{z}_i \right]. \tag{11}$$

Unfortunately, (11) can be negative in small samples. When this occurs we replace S_i^2 by $S^2(1-p_i)$. Although we have never encountered such an event in practice, when ψ is redescending there is a remote possibility that $p_i > 1$. Should this occur one could resort to the classical expression $S^2(1-h_{ii})$, as done by McKean et al. (1990).

We plot the studentized residuals $r_i' = e_i/S_i$. Here and in the influence measures below, we use the same versions of $\boldsymbol{P}$ and $\boldsymbol{Q}$ as are used in the corresponding computation of COV$[\hat{\boldsymbol{\theta}}]$. In particular, for the special case of least squares we use $\boldsymbol{P}_{\mathrm{exch}} = \boldsymbol{Q}_{\mathrm{exch}} = \sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{z}_i^{\mathrm{T}}$ and both forms of S_i reduce to the familiar $S\sqrt{1-h_{ii}}$.

4.2. Robust change in fit

Similar to the Welsch–Kuh distance (see Chatterjee and Hadi, 1986) we define the change in fit $\hat{Y}_i - \hat{Y}_{i,-i} = \boldsymbol{z}_i^{\mathrm{T}}(\hat{\boldsymbol{\theta}}_{\mathrm{GM}} - \hat{\boldsymbol{\theta}}_{-i})$ due to the ith data point, scaled with respect to the standard error $s(\hat{Y}_i) = \sqrt{\boldsymbol{z}_i^{\mathrm{T}} \boldsymbol{C} \boldsymbol{z}_i}$ of this fitted value:

$$\mathrm{RCF}_i = \frac{\hat{Y}_i - \hat{Y}_{i,-i}}{s(\hat{Y}_i)} = \frac{\boldsymbol{z}_i^{\mathrm{T}} \boldsymbol{P}^{-1} \boldsymbol{z}_i \psi(r_i/w_i^{\alpha}) w_i}{(1-p_i)\sqrt{\boldsymbol{z}_i^{\mathrm{T}} \boldsymbol{P}^{-1} \boldsymbol{Q} \boldsymbol{P}^{-1} \boldsymbol{z}_i}}.$$

For least squares RCF_i reduces to $\sqrt{h_{ii}/(1-h_{ii})}(e_i/S\sqrt{1-h_{ii}})$. Apart from the use of S rather than the delete-one scale S_{-i}, this is the classical Welsch–Kuh distance. In line with a suggestion of Cook and Weisberg (1982) we compare RCF_i with the benchmark values $\pm\sqrt{p}$.

4.3. Robust Cook's statistics

The sample influence curve (Cook and Weisberg, 1982, pp. 109ff.) is

$$\mathrm{SIC}_i = (n-1)(\hat{\boldsymbol{\theta}}_{\mathrm{GM}} - \hat{\boldsymbol{\theta}}_{-i}) = (n-1) \cdot S \cdot \boldsymbol{P}^{-1} \boldsymbol{z}_i \cdot \frac{\psi(r_i/w_i^{\alpha}) w_i}{1-p_i}.$$

The classical version of Cook's statistic is obtained as the scale and affine invariant measure $D_i(\boldsymbol{M}, c) = \mathrm{SIC}_i^{\mathrm{T}} \cdot \boldsymbol{M} \cdot \mathrm{SIC}_i/c$ with $\boldsymbol{M} = \sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{z}_i^{\mathrm{T}}$ and $c = p(n-1)^2 S^2$. Three choices which are then suggested as robust replacements are $(\boldsymbol{M}, c) = (\boldsymbol{P}, p(n-1)^2 S^2)$, $(\boldsymbol{C}^{-1}, p(n-1)^2)$ and $(\boldsymbol{Q}, p(n-1)^2 S^2)$. Although all three seem to tell much the same story in the examples at which we have looked, we prefer the last since $\boldsymbol{Q}$ is closely related to the covariance matrix of $\boldsymbol{z}_i \cdot \psi(r_i/w_i^{\alpha})w_i$. Thus we take

$$\mathrm{RC}_i = D_i(\boldsymbol{Q}, p(n-1)^2 S^2) = \frac{1}{p}\left(\frac{\psi(r_i/w_i^{\alpha})w_i}{1-p_i}\right)^2 \boldsymbol{z}_i^{\mathrm{T}} \boldsymbol{P}^{-1} \boldsymbol{Q} \boldsymbol{P}^{-1} \boldsymbol{z}_i.$$

Following classical usage, we compare RC_i to the median of the F_{n-p}^{p} distribution. We note the relationship

$$\mathrm{RC}_i = \frac{1}{p}\mathrm{RCF}_i^2 \left(\frac{\boldsymbol{z}_i^{\mathrm{T}} \boldsymbol{P}^{-1} \boldsymbol{Q} \boldsymbol{P}^{-1} \boldsymbol{z}_i}{\boldsymbol{z}_i^{\mathrm{T}} \boldsymbol{P}^{-1} \boldsymbol{z}_i}\right)^2. \tag{12}$$

Our proposed influence measures and benchmarks are then in agreement to the extent that $F_{n-p}^{p}(0.5) \approx 1$ and the two quadratic forms in (12) agree with each other. These quadratic forms are equal for least-squares estimates but can be substantially different otherwise.

There is a revealing decomposition of RC_i in terms of the regression weights v_i (see (13) below) which are employed if the estimates are computed via IRLS:

$$\mathrm{RC}_i = \frac{1}{p}\left(\frac{e_i}{S\sqrt{1-p_i}}\right)^2 \frac{v_i^2}{1-p_i} \frac{s^2(\hat{Y}_i)}{S^2}.$$

Thus large values of RC_i derive from large absolute values of the alternate form of the studentized residuals r_i' proposed above when (11) is negative, from large regression weights v_i, from leverages p_i near 1, or from large standardized values of $s^2(\hat{Y}_i)$.

5. Simulations and case study

For the simulations and examples below we computed eight types of estimates: OLS; a three-step Huber estimate (H); a three-step Mallows estimate ($\mathrm{M}^{(0)}$) using the weights $w^{(0)}(\cdot;\beta)$, a three-step Mallows estimate ($\mathrm{M}^{(1)}$) using weights $w^{(1)}(\cdot;\gamma)$, three-step Schweppe estimates $\mathrm{S}^{(0)}$ and $\mathrm{S}^{(1)}$ using weights $w^{(0)}(\cdot;\beta)$ and $w^{(1)}(\cdot;\gamma)$, respectively, and three-step Hill–Ryan estimates $\mathrm{HR}^{(0)}$ and $\mathrm{HR}^{(1)}$ using weights $w^{(0)}(\cdot;\beta)$ and $w^{(1)}(\cdot;\gamma)$, respectively. The initial estimator $\boldsymbol{\theta}_{(0)}$ for all except OLS was LTS. The constants γ and β were chosen for an asymptotic efficiency of 95% relative to unit weights. We used Hampel's piecewise linear ψ function with tuning constants (1.5,3,8). Very similar results were obtained using Huber's ψ function, or by using LMS as initial estimator.

Tables 3 and 4 give summary results of the simulation study comparing the various estimation and weighting methods. For each of two situations we simulated 5000 samples of size 30 each. In the first of these ("clean" data) the two non-constant columns of the design matrix were independently distributed, each consisting of 30 independent N(0,4) values. The random errors were i.i.d. N(0,1) and we took $\boldsymbol{\theta} = (2, -2, 1)^{\mathrm{T}}$. The second situation ("corrupted" data) was like the first, except that the first two rows of the design matrices had (0,10,10) and $(0, 10, -10)$, respectively, added to them. The first three y-values y_1, y_2, y_3 had -5, -30, 5, respectively, added to them. Thus, apart from random error, y_i exceeded $E[Y_i|\boldsymbol{x}_i]$ by -5, 0, 5 for $i = 1, 2, 3$ so that the first point was outlying with respect to both its $\boldsymbol{x}$-value and its y-value (a "bad" leverage point), the second with respect only to its $\boldsymbol{x}$-value (a "good" leverage point); the third was an outlier but not a high leverage point.

The same initial estimate was used for each of the robust procedures, and the same MVE estimates were used for each of $w^{(0)}(\cdot;\beta)$ and $w^{(1)}(\cdot;\gamma)$. This eliminated differences between the estimates due to subsampling

Table 3
Point estimates, biases, standard errors and root-mean-squared errors obtained from simulations (standard errors, in third decimal place, of bias and scale estimates are in parentheses)

	OLS	H	$M^{(0)}$	$M^{(1)}$	$S^{(0)}$	$S^{(1)}$	$HR^{(0)}$	$HR^{(1)}$
Clean data								
bias($\hat{\theta}_2$)	0.002 (1)	0.002 (1)	0.003 (2)	0.002 (1)	0.003 (2)	0.002 (2)	0.003 (2)	0.002 (2)
bias($\hat{\tau}$)	0.002 (5)	0.002 (5)	0.000 (5)	0.002 (5)	0.002 (6)	0.000 (5)	0.001 (6)	0.003 (5)
$\hat{\sigma}$	0.991 (2)	0.940 (3)	0.949 (3)	0.946 (3)	0.942 (3)	0.931 (3)	0.960 (3)	0.955 (3)
s.e.($\hat{\theta}_2$)	0.097	0.100	0.111	0.105	0.118	0.108	0.127	0.112
s.e.($\hat{\tau}$)	0.329	0.338	0.372	0.355	0.392	0.362	0.417	0.379
rmse($\hat{\theta}_2$)	0.097	0.100	0.111	0.105	0.118	0.108	0.127	0.112
rmse($\hat{\tau}$)	0.329	0.338	0.372	0.355	0.392	0.362	0.417	0.379
Corrupted data								
bias($\hat{\theta}_2$)	0.163 (1)	0.087 (2)	0.011 (2)	0.033 (2)	0.004 (2)	0.009 (2)	0.005 (2)	0.035 (2)
bias($\hat{\tau}$)	0.860 (4)	0.415 (7)	0.089 (6)	0.185 (6)	0.083 (6)	0.105 (6)	0.057 (6)	0.187 (5)
$\hat{\sigma}$	1.462 (3)	1.027 (3)	1.051 (3)	1.038 (3)	1.058 (3)	1.029 (3)	1.062 (3)	1.061 (3)
s.e.($\hat{\theta}_2$)	0.072	0.116	0.119	0.109	0.134	0.124	0.138	0.110
s.e.($\hat{\tau}$)	0.263	0.486	0.410	0.421	0.439	0.435	0.446	0.376
rmse($\hat{\theta}_2$)	0.178	0.145	0.120	0.114	0.134	0.124	0.138	0.115
rmse($\hat{\tau}$)	0.900	0.639	0.420	0.460	0.447	0.448	0.449	0.420

Table 4
Estimated coverage probabilities of 95% confidence intervals[a] and ratios of self-estimated standard errors to standard errors obtained from simulations (standard errors, in third decimal place, of coverage probabilities are in parentheses)

	Clean data				Corrupted data			
	Coverages		S.E. ratios		Coverages		S.E. ratios	
	$\hat{\theta}_2$	$\hat{\tau}$	$\hat{\theta}_2$	$\hat{\tau}$	$\hat{\theta}_2$	$\hat{\tau}$	$\hat{\theta}_2$	$\hat{\tau}$
OLS	0.953 (3)	0.948 (3)	0.985	1.002	0.539 (7)	0.277 (6)	1.176	1.291
H	0.948 (3)	0.947 (3)	0.979	0.996	0.586 (7)	0.521 (7)	0.574	0.551
$M^{(0)}_{exch}$	0.949 (3)	0.949 (3)	0.975	0.986	0.944 (3)	0.936 (3)	0.949	0.950
$M^{(0)}_{J,adj}$	0.964 (3)	0.961 (3)	1.056	1.059	0.968 (2)	0.959 (3)	1.070	1.077
$M^{(1)}_{exch}$	0.948 (3)	0.951 (3)	0.975	0.989	0.859 (5)	0.821 (5)	0.759	0.739
$M^{(1)}_{J,adj}$	0.963 (3)	0.959 (3)	1.053	1.058	0.940 (3)	0.916 (4)	1.029	1.009
$S^{(0)}_C$	0.917 (4)	0.925 (4)	0.882	0.896	0.924 (4)	0.914 (4)	0.851	0.904
$S^{(0)}_J$	0.932 (4)	0.937 (3)	0.924	0.935	0.937 (3)	0.927 (4)	0.895	0.948
$S^{(1)}_C$	0.926 (4)	0.932 (4)	0.903	0.930	0.918 (4)	0.896 (4)	0.947	1.015
$S^{(1)}_J$	0.935 (3)	0.941 (3)	0.941	0.966	0.927 (4)	0.904 (4)	0.981	1.049
$HR^{(0)}_C$	0.907 (4)	0.910 (4)	0.830	0.848	0.912 (4)	0.916 (4)	0.829	0.890
$HR^{(0)}_J$	0.921 (4)	0.922 (4)	0.869	0.886	0.931 (4)	0.928 (4)	0.872	0.933
$HR^{(1)}_C$	0.924 (4)	0.923 (4)	0.907	0.910	0.949 (3)	0.936 (3)	1.143	1.250
$HR^{(1)}_J$	0.938 (3)	0.936 (3)	0.957	0.947	0.957 (4)	0.941 (3)	1.185	1.291

[a]Based on t_{n-p} approximations to the distributions of the estimates divided by their standard errors.

variation in the LTS and MVE algorithms; we have sometimes found variation in the latter algorithm to have a particularly significant effect on the final estimates.

Each of the three steps employed a modified Newton's method. We initially determined iterates by

$$\boldsymbol{\theta}_{(k)} = \boldsymbol{\theta}_{(k-1)} + \kappa S \boldsymbol{P}^{-1} \sum_{i=1}^{n} \psi\left(\frac{y_i - \boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\theta}_{(k-1)}}{S w_i^{\alpha}}\right) \boldsymbol{z}_i w_i, \quad k = 1,2,3$$

with $\kappa = 1$ and $\boldsymbol{P}$ given by (3) and evaluated at $\boldsymbol{\theta}_{(k-1)}$. For the Mallows estimates we replaced $\boldsymbol{P}$ by $\boldsymbol{P}_{\mathrm{exch}}$. The step factor κ was then repeatedly decreased by a factor of $\frac{1}{2}$ until the right-hand side of (1) was found to have decreased. If this failed to occur after nine attempts, iterations ceased. Despite this safeguard the Schweppe estimates occasionally gave huge biases on the corrupted data, presumably due to the decreased effect of the weights on $\boldsymbol{P}$ when $\alpha = 1$. The Hill–Ryan estimates were occasionally highly biased on clean data. This aberrant behaviour was ameliorated in both cases by following the three Newton–Raphson steps by one step of IRLS:

$$\hat{\boldsymbol{\theta}} = \left(\sum_{i=1}^{n} v_i \boldsymbol{z}_i \boldsymbol{z}_i^{\mathrm{T}}\right)^{-1} \sum_{i=1}^{n} v_i \boldsymbol{z}_i y_i,$$

$$v_i = \begin{cases} \psi\left(\dfrac{r_i}{w_i^{\alpha}}\right)\dfrac{w_i}{r_i}, & r_i \neq 0, \\ \psi'(0) w_i^{1-\alpha}, & r_i = 0, \end{cases} \tag{13}$$

$$r_i = \frac{y_i - \boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\theta}_{(3)}}{S}.$$

Simpson and Chang (1997) showed that $\hat{\boldsymbol{\theta}}$ is asymptotically equivalent to $\hat{\boldsymbol{\theta}}_{\mathrm{GM}}$ of (1), and proposed some related diagnostic procedures.

For each estimate $\hat{\theta}_j$ we computed biases, simulated standard errors and sample estimates rmse of the root-mean-squared error: $\mathrm{rmse} = (N^{-1}\sum_{i=1}^{N}(\hat{\theta}_{j,i} - \theta_j)^2)^{1/2}$. These are presented in Table 3 and Fig. 1 for the slope estimate $\hat{\theta}_2$ and for the linear combination $\hat{\tau} = \hat{\theta}_0 + 2\hat{\theta}_1 + 2\hat{\theta}_2$. For the Mallows estimates we computed covariance estimates from $\mathrm{M}^{(0)}$ and $\mathrm{M}^{(1)}$ using each of (5) and (10), obtaining $\mathrm{M}^{(k)}_{\mathrm{exch}}$ and $\mathrm{M}^{(k)}_{\mathrm{J,adj}}$, respectively ($k = 0, 1$). For the Schweppe and Hill–Ryan estimates we computed these estimates from (4) and (9), obtaining $\mathrm{S}^{(k)}_{\mathrm{C}}$, $\mathrm{HR}^{(k)}_{\mathrm{C}}$ and $\mathrm{S}^{(k)}_{\mathrm{J}}$, $\mathrm{HR}^{(k)}_{\mathrm{J}}$. For H we use (5) together with a correction factor due to Huber (1981, p. 173, Eq. 6.5). The accuracy of these methods was measured by the observed coverage of nominal 95% confidence intervals, and by the ratios of the self-estimated to simulated standard errors. These are presented for $\hat{\theta}_2$ and for $\hat{\tau}$ in Table 4 and Fig. 1.

On the clean data all estimates performed well. The confidence intervals based on the Mallows estimates tended to become somewhat conservative when computed from the jackknife-based covariance estimates. The coverages based on $\mathrm{S}^{(k)}$ or $\mathrm{HR}^{(k)}$ were all closer to the nominal level for $k = 1$ than for $k = 0$, and were closer for J than for C. Similarly, the rmse figures for $k = 1$ were slightly lower than those for $k = 0$, for all three GM estimates.

The effect of corrupting the data is evident in the abysmal coverage properties of the OLS- and H-based confidence intervals. The intervals based on the GM estimates did not lose coverage appreciably (with the exception of $\mathrm{M}^{(1)}_{\mathrm{exch}}$, but this was rectified by $\mathrm{M}^{(1)}_{\mathrm{J,adj}}$). All benefited from the use of the jackknife-based covariance estimates. In most cases the rmse figures were lower using weights $w^{(1)}$ than using weights $w^{(0)}$.

Example 4.1. *Hawkins–Bradu–Kass data*. Fig. 2 shows a plot of the LMS standardized residuals r_i against the robust Mahalanobis distances RM_i for the artificial data set of Hawkins et al. (1984). These data are

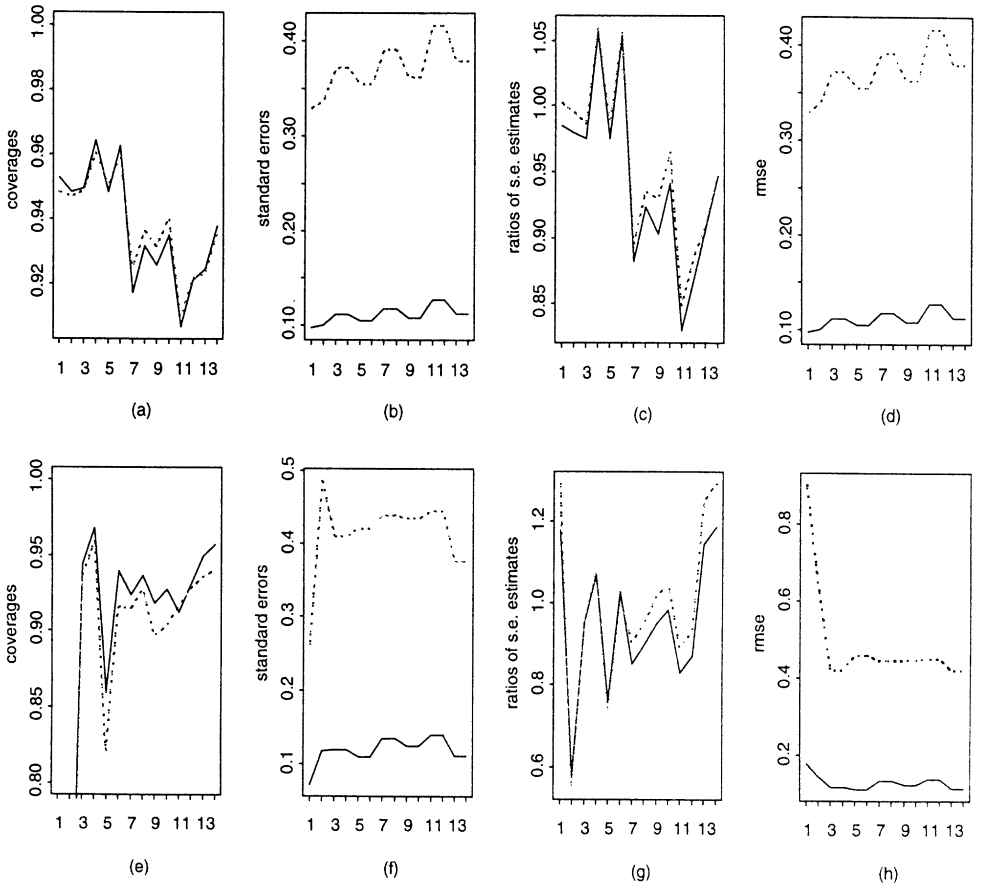

Fig. 1. Comparative measures from simulation study. Plots (a)–(d) use "clean" data; (e)–(h) use "corrupted" data. Solid lines refer to slope estimate $\hat{\theta}_2$, broken lines to $\hat{\tau} = \hat{\theta}_0 + 2\hat{\theta}_1 + 2\hat{\theta}_2$. Plots (a) and (e) give coverages of nominal 95% confidence intervals; (b) and (f) give standard errors computed from the simulations; (c) and (g) give ratios of self-estimated standard errors to those derived from the simulations; (d) and (h) give root mean-squared errors. Horizontal axes give estimation methods: 1. LS 2. H 3. $M^{(0)}_{exch}$ 4. $M^{(0)}_{J,adj}$ 5. $M^{(1)}_{exch}$ 6. $M^{(1)}_{J,adj}$ 7. $S^{(0)}_{C}$ 8. $S^{(0)}_{J}$ 9. $S^{(1)}_{C}$ 10. $S^{(1)}_{J}$ 11. $HR^{(0)}_{C}$ 12. $HR^{(0)}_{J}$ 13. $HR^{(1)}_{C}$ 14. $HR^{(1)}_{J}$.

discussed in Rousseeuw and Leroy (1987) and in Rousseeuw and van Zomeren (1990), where plots as in Fig. 2 are proposed. There are 75 observations on three variables. Observations 1–10 are 'bad' leverage points in that their y-values do not conform with the bulk of the data; these are the points with large standardized residuals ($|r_i| > 2.5$) and large RM_i ($> \sqrt{\chi^2_3(0.975)}$) in Fig. 2. Points 11–14 on the other hand are 'good' leverage points, with small LMS residuals in spite of their large RM values.

The results of several regressions both with and without points 1–10 are shown in Table 5. The estimates were calculated in the same way as in the simulation study, with the exception that LMS was used as initial estimate. Fig. 3 gives diagnostic plots for the full dataset. In the studentized residual plot the horizontal lines are at 0 and at ± 2.5. None of the estimation methods were fooled by the pattern of outliers; all assigned large studentized residuals to points 1–10 and accommodated points 11–14. With respect to RCF and RC the estimate $HR^{(1)}_{J}$ declared points 1–10 more influential — although still not near the benchmark values — than did the other estimates. An examination of Table 5 reveals that points 1–10 do indeed have more influence on $HR^{(1)}_{J}$ than on the other estimates.

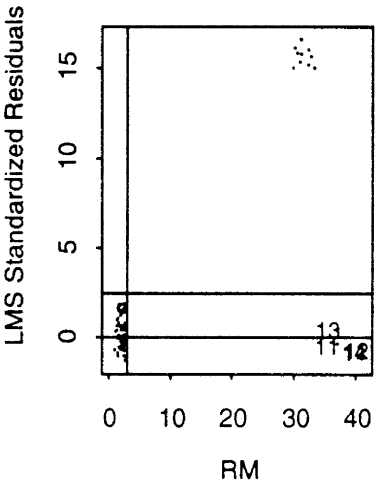

Fig. 2. Standardized LMS residuals vs. robust Mahalanobis distances; Hawkins–Bradu–Kass data.

Table 5
Hawkins–Bradu–Kass data — parameter estimates with standard errors in parentheses

	H	$M_{J,adj}^{(0)}$	$M_{J,adj}^{(1)}$	$S_J^{(0)}$	$S_J^{(1)}$	$HR_J^{(0)}$	$HR_J^{(1)}$
Full data set							
θ_0	−0.180 (0.110)	−0.040 (0.217)	−0.166 (0.109)	−0.010 (0.200)	−0.010 (0.200)	−0.024 (0.197)	−0.396 (0.160)
θ_1	0.081 (0.070)	0.073 (0.076)	0.093 (0.066)	0.062 (0.071)	0.062 (0.071)	0.080 (0.068)	0.128 (0.073)
θ_2	0.040 (0.041)	0.020 (0.076)	0.031 (0.055)	0.012 (0.070)	0.012 (0.070)	0.018 (0.067)	0.054 (0.065)
θ_3	−0.052 (0.034)	−0.110 (0.079)	−0.060 (0.045)	−0.107 (0.072)	−0.107 (0.072)	−0.124 (0.067)	0.021 (0.063)
Cleaned data set, excluding points 1–10							
θ_0	−0.180 (0.104)	−0.049 (0.226)	−0.169 (0.115)	−0.023 (0.201)	−0.017 (0.205)	−0.034 (0.197)	−0.078 (0.169)
θ_1	0.081 (0.067)	0.074 (0.079)	0.093 (0.071)	0.069 (0.072)	0.069 (0.074)	0.081 (0.069)	0.109 (0.075)
θ_2	0.040 (0.040)	0.024 (0.079)	0.032 (0.059)	0.019 (0.070)	0.016 (0.073)	0.024 (0.067)	−0.004 (0.065)
θ_3	−0.052 (0.035)	−0.111 (0.082)	−0.061 (0.048)	−0.117 (0.072)	−0.116 (0.073)	−0.128 (0.067)	−0.092 (0.067)

6. Summary

We have proposed estimates of the covariance matrix of a GM-estimator which are based on approximate delete-one estimates of the regression coefficients. These, and the proposed weights $w^{(1)}(\cdot;\gamma)$, have been shown in the simulation studies to result quite generally in significant gains over the standard competitors, with respect to the mean-squared errors of the regression estimates and the coverages of confidence intervals. The influence measures RCF and RC proposed in Section 4 have shown themselves, in the case study, to be useful complements to other robust diagnostic measures.

We note that McKean et al. (1993) questioned and investigated the usefulness of plots of GM residuals against predicted values given that, in contrast to least squares, these vectors are not orthogonal. Simpson and Chang (1997) addressed this question through their use of weighted residual plots following an iterative estimation method culminating in one step of IRLS. Although we have not presented them in this paper, such plots also provide useful information on possible curvature and bias.

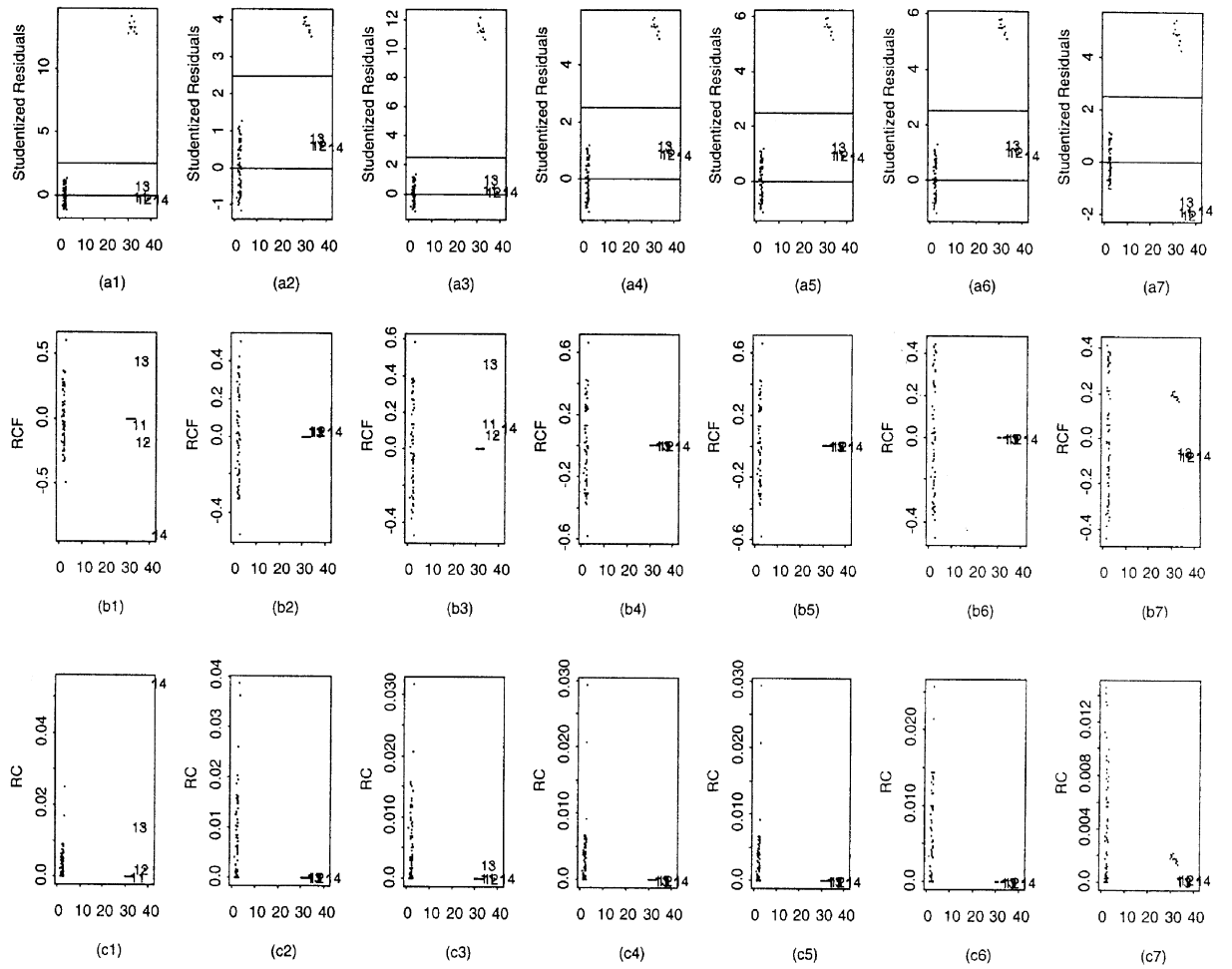

Fig. 3. Diagnostic plots; Hawkins–Bradu–Kass data. (a) Studentized residuals, (b) robust changes in fit, (c) robust Cook's distances; all against robust Mahalanobis distances. Estimation methods are: 1. H 2. $M^{(0)}_{J,adj}$ 3. $M^{(1)}_{J,adj}$ 4. $S^{(0)}_{J}$ 5. $S^{(1)}_{J}$ 6. $HR^{(0)}_{J}$ 7. $HR^{(1)}_{J}$.

References

Chatterjee, S., Hadi, A., 1986. Influential observations, high leverage points, and outliers in linear regression (with discussion). Statist. Sci. 1, 379–416.

Coakley, C.W., Hettmansperger, T.P., 1993. A bounded influence, high breakdown, efficient regression estimator. J. Amer. Statist. Assoc. 88, 872–880.

Cook, R.D., Weisberg, S., 1982. Residuals and Influence in Regression. Chapman & Hall, New York.

Hampel, F.R., Ronchetti, E., Rousseeuw, R.J., Stahel, W., 1986. Robust Statistics: The Approach Based on Influence Functions. Wiley, Toronto.

Handshin, E., Schweppe, F.C., Kohlas, J., Fiechter, A., 1975. Bad data analysis for power system state estimation (with discussion). IEEE Trans. Power Apparatus Systems 94, 329–337.

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

Hinkley, D.V., 1977. Jackknifing in unbalanced situations. Technometrics 19, 285–292.

Huber, P.J., 1981. Robust Statistics. Wiley, New York.

Maronna, R.A., Yohai, V.J., 1981. Asymptotic behaviour of general M-estimators for regression and scale with random carriers. Z. Wahrscheinlichkeitstheorie Verwandte Gebiete 58, 7–20.

McKean, J.W., Sheather, S.J., Hettmansperger, T.P., 1990. The use and interpretation of residuals based on robust estimation. J. Amer. Statist. Assoc. 88, 1254–1263.

McKean, J.W., Sheather, S.J., Hettmansperger, T.P., 1993. Regression diagnostics for rank-based methods. J. Amer. Statist. Assoc. 85, 1018–1028.

Rousseeuw, P.J., 1984. Least median of squares regression. J. Amer. Statist. Assoc. 79, 871–880.

Rousseeuw, P.J., Leroy, A.M., 1987. Robust Regression and Outlier Detection. Wiley, Toronto.

Rousseeuw, P.J., van Zomeren, B.C., 1990. Unmasking multivariate outliers and leverage points (with discussion). J. Amer. Statist. Assoc. 85, 633–639.

Simpson, D.G., Chang, Y.-C.I., 1997. Reweighting approximate GM estimators: asymptotics and residual-based graphics. J. Statist. Plann. Inference 57, 273–293.

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

Welsch, R.E., 1980. Regression sensitivity analysis and bounded influence estimation. In: Kmenta, J., Ramsey, J.B. (Eds.), Evaluation of Econometric Models. Academic Press, New York, pp. 153–167.

Wiens, D.P., 1996. Asymptotics of generalized M-estimation of regression and scale with fixed carriers, in an approximately linear model. Statist. Probab. Lett. 30, 271–285.

Wiens, D.P., 1999. Robust weights and designs for biased regression models: least squares and generalized M estimation. J. Statist. Plann. Inference, in press.

Wu, C.F.J., 1986. Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). Ann. Statist. 14, 1261–1295.