Bounded Influence Regression using High Breakdown Scatter Matrices

Christophe Croux,[1] Stefan Van Aelst,[2] and Catherine Dehon[1]

Abstract

In this paper we estimate the parameters of a regression model using S-estimators of multivariate location and scatter. The approach is proven to be Fisher-consistent, and the influence functions are derived. The corresponding asymptotic variances are obtained and it is shown how they can be estimated in practice.

Keywords: Fisher-Consistency, Influence Function, Robust Regression, S-Estimators.

1 Introduction

Consider the classical regression model

$$y_i = \alpha + \beta^t u_i + \varepsilon_i,$$

$i = 1, \ldots, n$ where the error terms $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. and independent of the p-dimensional carriers $u_1, \ldots, u_n$. The least-squares (LS) estimators $\hat{\alpha}_{LS}$ and $\hat{\beta}_{LS}$ are defined as the minimizers of the sum of squared residuals

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \alpha - \beta^t u_i)^2. \tag{1.1}$$

Since the least squares estimator is very sensitive to the presence of outliers, robust alternatives need to be looked for. Many of these robust regression methods consist of minimizing a robust loss function of the residuals, instead of a quadratic loss function. Main examples here are the Least Median of Squares and Least Trimmed Squares estimator (Rousseeuw 1984), who can attain the maximum breakdown value. The breakdown value is the smallest fraction of data points that needs to be replaced to carry the estimator arbitrarily far away

[1]ECARES, Université Libre de Bruxelles, CP-139, Av. F.D. Roosevelt 50, B-1050 Brussels, Belgium.

[2]Research Assistant with the FWO, Belgium, University of Antwerp, Dept. of Mathematics and Computer Science, Universiteitsplein 1, B-2610 Wilrijk, Belgium. E-mail: Stefan.VanAelst@uia.ua.ac.be

(for a formal definition, see Rousseeuw and Leroy 1987, page 117). Generalized S-estimators (Croux, Rousseeuw and Hössjer 1994) and τ-estimators (Yohai and Zamar 1988) combine this high breakdown value with a high efficiency. However, their unbounded influence function is sometimes seen as a drawback.

Another way of robustifying LS consists of robustifying the first order conditions associated to the minimization of (1.1):

$$\frac{1}{n}\sum_{i=1}^{n}(y_i-\alpha-\beta^t u_i)u_i=0 \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n}(y_i-\alpha-\beta^t u_i)=0. \tag{1.2}$$

This lead to the construction of M and GM-estimators which are defined as solutions of robustified versions of the first order equations (1.2). Unfortunately, they have no high breakdown point (see e.g. Hampel et al. 1986). To remediate this, MM- (Yohai 1987) and one step GM-estimators (Simpson, Ruppert, and Carroll 1992, Coakley and Hettmansperger 1993) were proposed.

In case of the LS-estimator, the solution of the normal equations (1.2) is explicit:

$$\hat\beta=\left(\frac{1}{n}\sum_{i=1}^{n}(u_i-\hat\mu_u)(u_i-\hat\mu_u)^t\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat\mu_y)(u_i-\hat\mu_u)\right) \text{ and } \hat\alpha=\hat\mu_y-\hat\mu_u^t\hat\beta,$$

with $\hat\mu_u=\frac{1}{n}\sum_{i=1}^n u_i$ and $\hat\mu_y=\frac{1}{n}\sum_{i=1}^n y_i$. With the use of the empirical covariance matrices $\hat S_{uy}$ and $\hat S_{uu}$, we may rewrite the above equation as

$$\hat\beta=\hat S_{uu}^{-1}\hat S_{uy} \quad \text{and} \quad \hat\alpha=\hat\mu_y-\hat\mu_u^t\hat\beta. \tag{1.3}$$

The idea now is not to robustify the normal equations, but its solutions. Therefore, we will replace the empirical mean and covariance in (1.3) by robust equivalents. Many proposals for robust location and covariance matrices have been made, such as M-estimators (Maronna 1976), the Stahel-Donoho estimator (Stahel 1981), the Minimum Volume Ellipsoid and Minimum Covariance Determinant estimator (Rousseeuw 1984,1985) and S-estimators (Davies 1987, Rousseeuw and Leroy 1987).

Maronna and Morgenthaler (1986) used multivariate M-estimators to insert into (1.3) and showed that the resulting estimators have all the desired equivariance properties. They also gave an expression for the influence function of this approach based on M-estimators, but only for a regression without intercept. Visuri et al. (2000) used rank based covariance matrices and derived results at elliptically symmetric models. In this paper, S-estimators of location and scatter will be used. For a finite sample $\{z_1,\ldots,z_n\}\subset I\!R^{p+1}$ the S-estimates are defined as the couple $(\hat{\mu}, \hat{S})$ which minimizes $\det(S)$ under the constraint

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\sqrt{(z_i-\mu)^t S^{-1}(z_i-\mu)}\,\right) \leq b, \tag{1.4}$$

over all $\mu \in I\!R^{p+1}$ and $S \in \mathrm{PDS}(p+1)$, where $\mathrm{PDS}(p+1)$ is the set of all positive definite symmetric matrices of size $p+1$. The function ρ is chosen by the statistician and b is a selected constant.

At first sight, this approach based on robust covariance matrix estimators seems to be restricted to regression models with elliptically symmetric carrier distribution. Indeed, consistency of robust covariance matrices is always proven under this symmetry assumption. In practice, this restriction cannot be retained. Even an ordinary quadratic regression would then not be covered by the hypothesis of the model. The main contribution of this paper is that we prove the approach to be valid for arbitrary carrier distributions.

In Section 2 we define the regression functionals based on robust S-estimators of location and scatter. The corresponding influence function is computed in Section 3 and shown to be bounded for ρ functions with bounded derivative. An estimator for the covariance matrix of the estimator is presented as well. Section 4 gives a real data example, while Section 5 concludes. The Appendix contains all the proofs.

2 The Functional

The functional form of S-estimators of multivariate location and scatter is defined as follows. Let K be an arbitrary $(p+1)$-dimensional distribution. For our purposes, K represents the joint distribution of the carriers and response variable. Define now the S-estimator $(M(K), S(K))$ as the couple (M, S) which minimizes $\det(S)$ under the constraint

$$\int \rho\left(\sqrt{(z-M)^t S^{-1}(z-M)}\,\right) dK(z) \leq b, \tag{2.1}$$

over all $M \in I\!R^{p+1}$ and $S \in \mathrm{PDS}(p+1)$. The function ρ satisfies

(R) ρ is even, continuous, non decreasing on $[0, +\infty[$ with $\rho(0) = 0$, and almost everywhere twice differentiable with derivative $\rho' = \psi$.

The constant b satisfies $0 < b < \rho(\infty)$ and determines the breakdown point of the estimator which equals $\min(\frac{b}{\rho(\infty)}, 1 - \frac{b}{\rho(\infty)})$ (see Lopuhaä 1989). The vector $M(K)$ corresponds with the location S-estimator, and $S(K)$ with the scatter S-estimator.

Let u contain the first p components of the variable $z \sim K$ and y the last component, so $z = (u^t, y)^t$. The variable y will be the dependent variable of the regression equation while u contains the explanatory variables. Split up the vector $M(K)$ and matrix $S(K)$ accordingly, that is

$$M(K) = \begin{pmatrix} M_u(K) \\ M_y(K) \end{pmatrix} \quad \text{and} \quad S(K) = \begin{pmatrix} S_{uu}(K) & S_{uy}(K) \\ S_{yu}(K) & S_{yy}(K) \end{pmatrix}.$$

The functional of interest is now defined as $T(K) = (a(K), b(K)^t)^t$ where

$$b(K) = S_{uu}^{-1}(K) S_{uy}(K) \tag{2.2}$$

is called the regression slope functional and

$$a(K) = M_y(K) - b(K)^t M_u(K) \tag{2.3}$$

the intercept functional. One has that $T = (a, b^t)^t$ is regression, scale, and carrier equivariant (Maronna and Morgenthaler, 1986). This means that, using the notation $a(K) = a(u, y)$ and $b(K) = b(u, y)$ for $(u^t, y)^t \sim K$,

$$\begin{aligned} a(Au, cy + l^t u + d) &= c\, a(u, y) + d \\ b(Au, cy + l^t u + d) &= (A^{-1})^t (c\, b(u, y) + l) \end{aligned}$$

for every $l \in I\!R^p$, $c, d \in I\!R$ and nonsingular $(p \times p)$ matrix A.

Consider now the regression model

$$y = \alpha + u^t \beta + \varepsilon$$

where u is the vector of random explicative variables and ε the error term. We suppose that ε is independent of u and that $F(t) = P(\varepsilon \leq t)$ satisfies

(F) The distribution F has a strictly positive, symmetric and unimodal density f.

We denote by H the distribution of $z = (u^t, y)^t$, and call it the model distribution. A regularity condition (to avoid degenerate situations) on the distribution G of the carriers u is that

(G) $P_G(u^t \gamma = \delta) < 1 - \frac{b}{\rho(\infty)}$ for all $\gamma \in I\!R^p \setminus \{0\}$ and $\delta \in I\!R$.

When using a 50% breakdown estimator, this means that not more than half of the mass of the distribution of G is lying on the same hyperplane. For unbounded ρ functions it implies that the distribution of G is not completely concentrated on a hyperplane. A first result is that the functionals a and b defined in (2.3) and (2.2) are Fisher-consistent for the intercept and slope parameters α and β.

Theorem 1. *The functional T is Fisher-consistent for the parameter $\theta = (\alpha, \beta^t)^t$ at the model distribution H, that is*

$$T(H) = \begin{pmatrix} a(H) \\ b(H) \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \theta.$$

Note that no symmetry conditions for the distribution of the carriers have been required.

3 Influence function

Before deriving the influence function we recall that S-estimators satisfy the following first-order conditions (Lopuhaä 1989):

$$\int w_1(d_K^2(z))(z - M(K))\, dK(z) \;=\; 0 \tag{3.1}$$

$$\int w_1(d_K^2(z))(z - M(K))(z - M(K))^t\, dK(z) \;=\; \int w_2(d_K^2(z))\, dK(z)\, S(K), \tag{3.2}$$

where the weight functions equal $w_1(t) = \psi(\sqrt{t})/\sqrt{t}$ and $w_2(t) = \frac{\psi(\sqrt{t})\sqrt{t} - \rho(\sqrt{t}) + b}{p+1}$, and $d_H^2(z) = (z - M(H))^t S(H)^{-1}(z - M(H))$ is a squared Mahalanobis distance. It will be shown that w_1 determines the form of the influence function.

The influence function of the functional T at the distribution H measures the effect on T of adding a small mass at $z = (u^t, y)^t$. If we denote the point mass at z by Δ_z and consider the contaminated distribution $H_{\varepsilon,z} = (1-\varepsilon)H + \varepsilon\Delta_z$ then the influence function is given by

$$IF(z; T, H) = \lim_{\varepsilon \downarrow 0} \frac{T(H_{\varepsilon,z}) - T(H)}{\varepsilon} = \frac{\partial}{\partial\varepsilon} T(H_{\varepsilon,z})\big|_{\varepsilon=0}.$$

(See Hampel et al. 1986.) The next theorem gives an expression for the influence functions of the regression functional b and intercept functional a at a model distribution.

Theorem 2. *Let $y = \alpha + u^t\beta + \varepsilon$, where ε is independent of u, and $\varepsilon \sim F$ satisfying condition (F). Let H be the distribution of $z = (u^t, y)^t$, H_0 the distribution of $(u^t, \varepsilon)^t$ and* *denote $x = (1, u^t)^t$ and $\theta = (\alpha, \beta^t)^t$. Then the influence function of the functional T at the distribution H is given by*

$$\mathrm{IF}(z; T, H) = C(H_0)^{-1} w_1(d_H^2(z)) x(y - x^t\theta) \tag{3.3}$$

where

$$C(H_0) = \int w_1(d_{H_0}^2(z)) xx^t \, dH_0(z) + \frac{2}{S_{yy}(H_0)} \int w_1'(d_{H_0}^2(z)) y^2 xx^t \, dH_0(z). \tag{3.4}$$

Moreover, if the score function $\Lambda_f(t) = -f'(t)/f(t)$ associated to the density f exists, then

$$C(H_0) = \int xx^t w_1(d_{H_0}^2(z)) y \Lambda_f(y) \, dH_0(z). \tag{3.5}$$

From the above Theorem, it is seen that the influence function is bounded as soon as w_1 is bounded.

Remark 1: Let $z \sim H$ the model distribution and denote

$$A = \begin{pmatrix} I_p & 0 \\ -\beta^t & 1 \end{pmatrix} \qquad \text{and} \qquad c = \begin{pmatrix} 0 \\ -\alpha \end{pmatrix}.$$

Then $Az + c \sim H_0$ and by affine equivariance of the S-estimator

$$S(H_0) = AS(H)A^t = \begin{pmatrix} S_{uu}(H) & 0 \\ 0 & S_{yy}(H) - \beta^t S_{uy}(H) \end{pmatrix}.$$

The scale functional $\sigma_\varepsilon(H) := S_{yy}(H_0)$ equals therefore

$$\sigma_\varepsilon(H) = \sqrt{S_{yy}(H) - \beta^t S_{uu}(H)\beta}\,. \tag{3.6}$$

Since $\det(A) = 1$ we can rewrite (3.4) as

$$C(H_0) = \int w_1(d_H^2(z)) xx^t \, dH(z) + \frac{2}{\sigma_\varepsilon^2(H)} \int w_1'(d_H^2(z))(y - x^t\theta)^2 xx^t \, dH(z), \tag{3.7}$$

which is an expression in terms of the observed distribution H. Equivalently,

$$C(H_0) = \int xx^t w_1(d_H^2(z))(y - x^t\theta)\Lambda_f(y - x^t\theta) \, dH(z). \tag{3.8}$$

Remark 2: Note that for $\rho(t) = t^2$, we have $w_1(t) = 1$ and from (3.7) $C(H_0) = E_H[xx^t]$. Therefore, the influence function becomes $\mathrm{IF}(z; T, H) = E_H[xx^t]^{-1} x(y - x^t\theta)$ which is, as expected, the influence function of the least squares estimator.

4 Estimating the asymptotic variance

Asymptotic variances can be obtained in a heuristic way from the influence function by means of

$$\mathrm{ASV}(T,H)=\int \mathrm{IF}(z;T,H)\,\mathrm{IF}(z;T,H)^t\,dH(z)$$

(cfr. Hampel et al. 1986, page 226). Together with expression (3.3) this yields

$$\mathrm{ASV}(T,H)=C(H_0)^{-1}D(H_0)\,C(H_0)^{-1}$$

with

$$D(H_0)=\int w_1^2(d^2_{H_0}(z))y^2xx^t\,dH_0(z)=\int w_1^2(d^2_H(z))(y-x^t\theta)^2xx^t\,dH(z). \tag{4.1}$$

At the sample level, we estimate the parameters α,β by $\hat\alpha=a(H_n)$ and $\hat\beta=b(H_n)$, where H_n is the empirical distribution function of the data $z_i=(x_i^t,y_i)^t$ $(1\le i\le n)$. With $\hat\mu=M(H_n)$ and $\hat S=S(H_n)$ we retrieve the estimators defined in the introduction (equations (1.3) and (1.4)). The covariance matrix of $\hat\theta=(\hat\alpha,\hat\beta^t)^t$ is now estimated in a natural way by replacing H by H_n in the right hand side of expressions (3.7) and (4.1):

$$\widehat{\mathrm{Cov}(\hat\theta)}=\frac{1}{n}\widehat{\mathrm{ASV}(T,H)}=\frac{1}{n}\widehat{C(H_0)}^{-1}\widehat{D(H_0)}\,\widehat{C(H_0)}^{-1} \tag{4.2}$$

with

$$\begin{aligned}\widehat{D(H_0)}&=\frac{1}{n}\sum_{i=1}^n w_1^2(d_i^2)r_i^2x_ix_i^t\\ \widehat{C(H_0)}&=\frac{1}{n}\sum_{i=1}^n\{w_1(d_i^2)+\frac{2}{\hat\sigma^2_{\varepsilon,n}}\,w_1'(d_i^2)r_i^2\}x_ix_i^t,\end{aligned}$$

where $x_i=(1,u_i^t)^t$, $r_i=y_i-u_i^t\hat\beta-\hat\alpha$, $d_i=\sqrt{(z_i-\hat\mu)^t\hat S^{-1}(z_i-\hat\mu)}$ is the robust Mahalanobis distance of z_i (Rousseeuw and van Zomeren 1990), and

$$\hat\sigma_{\varepsilon,n}=\sqrt{\hat S_{yy}-\hat\beta^t\hat S_{uu}^{-1}\hat\beta}\,.$$

Alternatively, $C(H_0)$ can be estimated, by using (3.8), as

$$\widehat{C(H_0)}=\frac{1}{n}\sum_{i=1}^n x_ix_i^tw_1(d_i^2)r_i\Lambda_{\hat f_n}(r_i), \tag{4.3}$$

which requires however a nonparametric estimate $\hat{f}_n$ of the density f. If f is specified to be $N(0, \sigma^2)$ then we have $\Lambda_f(t) = -\frac{d}{dt} \log f(t) = t/\sigma^2$. The parameter σ can be estimated from the residuals by a consistent scale estimator $\hat{\sigma}_n(r_1, \ldots, r_n)$. For Gaussian errors (4.3) then results in

$$\widehat{C(H_0)} = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^t w_1(d_i^2) \frac{r_i^2}{\hat{\sigma}_n^2}$$

yielding

$$\widehat{\mathrm{Cov}}(\hat{\theta}) = \hat{\sigma}_n^4 \left(\sum_{i=1}^{n} w_1(d_i^2) r_i^2 x_i x_i^t \right)^{-1} \left(\sum_{i=1}^{n} w_1^2(d_i^2) r_i^2 x_i x_i^t \right) \left(\sum_{i=1}^{n} w_1(d_i^2) r_i^2 x_i x_i^t \right)^{-1} . \tag{4.4}$$

If the function ρ becomes constant for values larger than a certain c^*, then the function w_1 is redescending to zero. It follows that in this case the estimators for $\mathrm{Cov}(\hat{\theta})$ are robust since outliers are downweighted to zero in expressions (4.2) and (4.4).

5 Example

As an example we consider the famous Hawkins-Bradu-Kass data (Hawkins et al. 1984), which is a constructed data set with $n = 75$ and $p = 3$. The first 14 observations are known to be outliers. As ρ function the Tukey Biweight function was taken

$$\rho_c(t) = \min(\frac{t^2}{2} - \frac{t^4}{2c^2} + \frac{t^6}{6c^4}, \frac{c^2}{6}) \tag{5.1}$$

and the constant b was set to $\frac{\rho(\infty)}{2}$ to ensure that the estimator will have a 50% breakdown point. The choice of the tuning constant c is arbitrary in this regression setup, but it is customary to select it such that $E_H[\rho(d_H(z))] = b$ for $H = N(0, I_{p+1})$. The function ρ_c is bounded and sufficiently smooth, with an associated weight function w_1 being redescending. For computing the S-estimator of location and scatter, the fast and accurate algorithm of Ruppert (1992) has been used. The estimate for the covariance matrix of the coefficients has been computed from formula (4.2). In Table 1 (a), we report the estimates obtained with the classical estimator, the proposed robust covariance based estimator and the MM-estimator (Yohai and Zamar 1988) with 50% breakdown point. We have chosen to make a comparison with robust MM-estimators, since this is an established robust regression method with high breakdown point and good efficiency properties. It is standard implemented in S-plus and also reports standard errors and correlations between the estimates (as described in Yohai, Stahel and Zamar 1991).

Table 1: Estimates of the intercept and regression parameters for (a) the Hawkins-Bradu-Kass data and (b) the clean Hawkins-Bradu-Kass data. Standards errors are reported between parenthesis, correlations between estimated coefficients are in the right panel of the table. The estimators considered are the Least Squares (LS) estimator, the estimator based on the robust S-estimator of location/scatter, and an MM-estimator.

(a)

$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$		$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$
LS-estimator				$\hat{\beta}_1$	-0.637		
-0.388	0.239	-0.335	0.383	$\hat{\beta}_2$	-0.180	-0.084	
(0.405)	(0.255)	(0.015)	(0.125)	$\hat{\beta}_3$	0.470	-0.540	-0.775
Robust Covariance Based				$\hat{\beta}_1$	-0.360		
-0.018	0.097	0.004	-0.130	$\hat{\beta}_2$	-0.635	-0.009	
(0.226)	(0.079)	(0.078)	(0.077)	$\hat{\beta}_3$	-0.386	-0.316	-0.086
MM-estimator				$\hat{\beta}_1$	-0.648		
-0.181	0.081	0.040	-0.052	$\hat{\beta}_2$	-0.164	-0.084	
(0.114)	(0.073)	(0.044)	(0.039)	$\hat{\beta}_3$	0.426	-0.487	-0.795

(b)

$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$		$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$
LS-estimator				$\hat{\beta}_1$	-0.456		
-0.010	0.062	0.012	-0.107	$\hat{\beta}_2$	-0.527	-0.031	
(0.190)	(0.067)	(0.066)	(0.069)	$\hat{\beta}_3$	-0.481	-0.102	-0.123
Robust Covariance Based				$\hat{\beta}_1$	-0.328		
-0.021	0.123	-0.001	-0.147	$\hat{\beta}_2$	-0.672	-0.020	
(0.253)	(0.087)	(0.088)	(0.082)	$\hat{\beta}_3$	-0.371	-0.354	-0.058
MM-estimator				$\hat{\beta}_1$	-0.463		
-0.011	0.062	0.012	-0.107	$\hat{\beta}_2$	-0.533	-0.008	
(0.245)	(0.086)	(0.086)	(0.090)	$\hat{\beta}_3$	-0.451	-0.127	-0.149

We see that the two robust methods give quite similar results, while the classical estimates are very different since they are highly influenced by the outliers. Note that none of the variables is declared as significant by the robust approach, while the second and third slope parameter are significantly different from zero according to the LS method. It is instructive to compare these results with those based on the clean data-set with the 14 artificial outliers deleted. From Table 1 (b), we notice that the results for the method based on the robust covariance matrix hardly change, neither for the estimates, neither for the covariance matrix of the estimates. The MM-estimator appears to be less stable for the correlations between the coefficients. Note that, on the basis of the clean data, LS finds none of the variables to be significant.

Several diagnostic plots can be produced. We will illustrate them for the estimator based on the robust S-covariance matrix defined above. In Figure 1a, the standardized residuals $r_i/\hat{\sigma}_n$ are represented versus their index. We immediately observe that the first 14 observations are not following the linear relation imposed by the majority of the data.

The robust distance d_i of an observation $z_i = (x_i, y_i)$ indicates how far the data point is from the bulk of the data cloud. In Figure 1b the robust distances d_i are compared with the constant c of (5.1). If d_i is bigger than this critical value then $w_1(d_i^2)$ will vanish, resulting in a zero influence on the estimator according to (3.3). We see that the 14 outliers were all above this critical value, and therefore are completely downweighted.

To verify whether condition (F) on the residuals is reasonable, a diagnostic plot will be used (cfr. Figure 1c). The solid line is the kernel density estimate $\hat{f}_h(t)$ of the distribution of the residuals. The Gaussian kernel has been used and the bandwidth h was selected using maximum-likelihood cross-validation (see e.g. Härdle 1991, p. 93). Afterwards, a symmetric unimodal version of this density has been added to this plot. It has been constructed as follows: first we computed $\hat{f}_h^s(t_j) = (\hat{f}_h(t_j) + \hat{f}_h(-t_j))/2$ for a grid of equidistant positive points starting from zero. Then a classical monotonic regression algorithm (see e.g. Cox and Cox 1994, page 51) has been applied on the $\hat{f}_h^s(t_j)$ to obtain $\hat{f}_h^{sm}(t_j)$. Putting $\hat{f}_h^{sm}(-t_j) = \hat{f}_h^{sm}(t_j)$ and connecting all the obtained values results in the dashed line of Figure 1c, which is a symmetric and unimodal function. Note that the initial density estimate is reasonably close to the unimodal symmetric version, so we assume that condition (F) is satisfied. Of course, more formal tests for unimodality and symmetry could be applied. To put emphasis on the central part of the data, the density estimate has been restricted to the interval

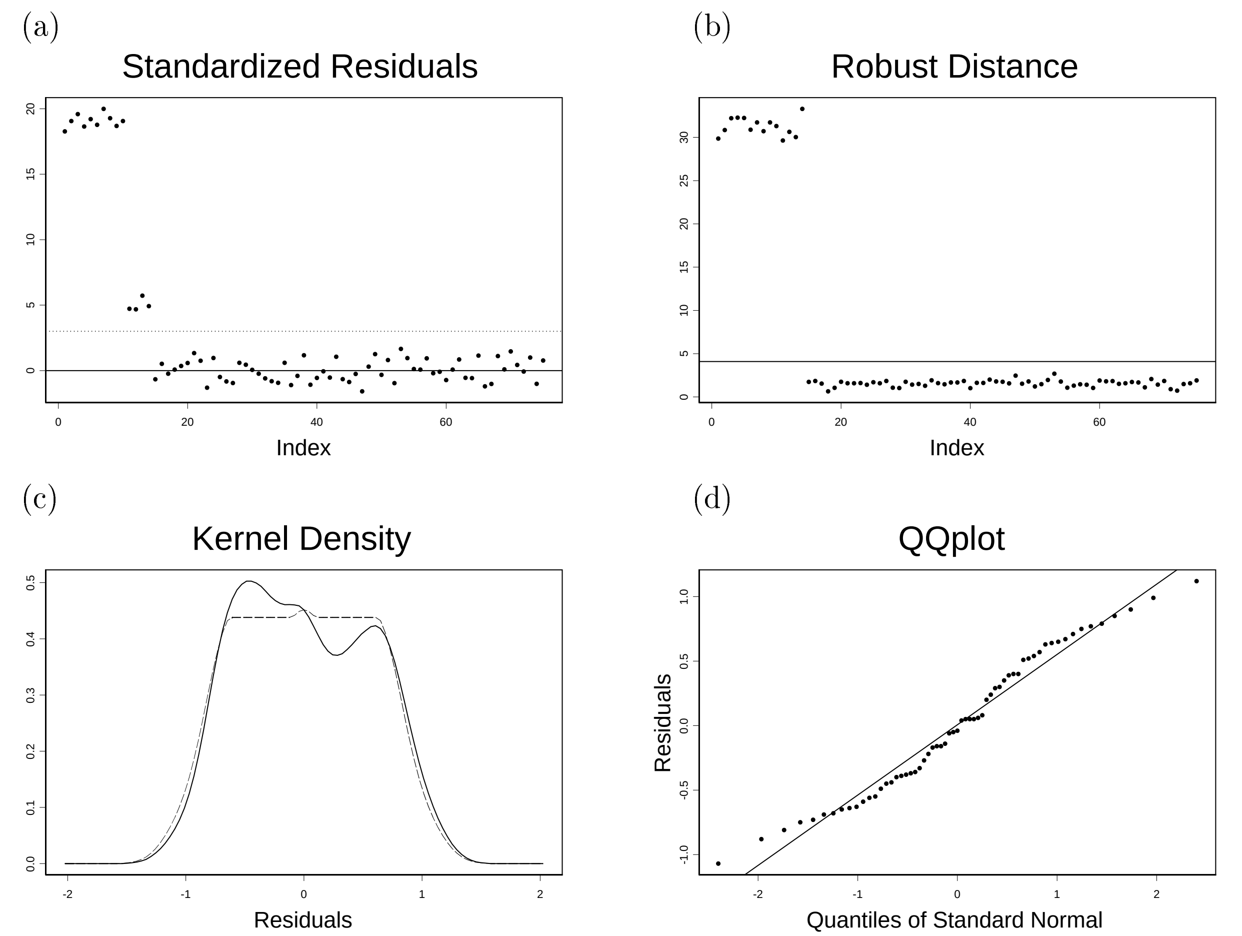

Figure 1: Diagnostic plots for the residuals of the regression estimator based on an S-estimator of multivariate location/scatter for the Hawkins-Bradu-Kass data: (a) standardized residuals (b) robust distances versus their index (c) kernel-based density estimate (solid line) and its symmetric, unimodal version (dashed line) (d) QQ-plot of the residuals.

Table 2: Estimates of the intercept and regression parameters for the Hawkins-Bradu-Kass data by the method based on a robust covariance S-estimator, as in Table 1. Now, for computing standards errors and correlations between estimated coefficients the hypothesis of normality was used.

$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$		$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$
Complete data set				$\hat{\beta}_1$	-0.538		
-0.018	0.097	0.004	-0.130	$\hat{\beta}_2$	-0.457	-0.028	
(0.297)	(0.102)	(0.101)	(0.111)	$\hat{\beta}_3$	-0.572	0.063	-0.114
Clean data set				$\hat{\beta}_1$	-0.530		
-0.021	0.123	-0.001	-0.147	$\hat{\beta}_2$	-0.461	-0.002	
(0.231)	(0.080)	(0.079)	(0.089)	$\hat{\beta}_3$	-0.534	0.003	-0.150

$[-3\hat{\sigma}_n, 3\hat{\sigma}_n]$. Finally, a classical QQ-plot is presented in Figure 1d. Once again, since we do not want the outliers to dominate this picture and make the interpretation harder, the plot is based on all residuals with absolute value smaller than $3\hat{\sigma}_n$. Figure 1d suggest that normality will not be rejected. This is confirmed by a Kolmogorov-Smirnoff test (P-value>0.2).

Supposing normality of the error terms allows us to use formula (4.4) to estimate the covariance matrix of the estimator. The scale of the Gaussian error distribution ε was estimated by an A-estimator of scale (see Iglewicz 1982, page 417), which has the maximal breakdown point, a redescending influence function (like the regression and intercept estimators) and is standard implemented in S-plus. Results are reported in Table 2. We see that estimates, standard errors, and correlations between the coefficients remain robust: the outcomes based on the whole data set and just on the clean data are not too different and close to the LS result computed from the data with the outliers deleted.

To conclude this Section, we note that an unbounded (in the x-direction) influence estimator, like the MM-estimator or the Least Trimmed Squares estimator, will yield small standardized values for residuals 11 upto 14. Using these estimators, observations 11-14 are not declared as regression outliers, but are called "good leverage points": they are outlying in the design space, but still follow the regression model. Our analysis yields a different result: we declare 11-14 as regression outliers, although far less outlying than the first 10 observations. We claim that our interpretation is the correct one: the unbounded influence function approach has the risk that leverage points, not too far located from the "true" regression plane, can still have a huge influence on the estimator and attract the regression line towards them. Extensive diagnostic testing confirmed our conclusion. For example, computing standardized residuals for the suspected observations 11 to 14 from the LS-fit based on the clean data set (observations 15 upto 75) yielded values higher than 4.5.

6 Conclusions

In this paper we discussed properties of regression estimators based on high breakdown S-estimators of location and scatter. We proved Fisher consistency of the method, without making the hypothesis of elliptical symmetry on the distribution of the explanatory variables. We derived the influence function, which appears to be bounded for the usual choices of ρ functions in robust statistics. Moreover, it can easily be shown that the resulting regression estimator inherits the breakdown point of the location/scatter S-estimator.

S-estimators of location and scatter have very attractive properties. It was shown that they are asymptotically normal (Davies 1987) with a quite high efficiency also in higher dimensions (Lopuhaä 1989, Croux and Haesbroeck 1999). At the same time they are extremely robust and have a smooth influence function. Moreover, there exist very fast algorithms to compute them (Ruppert 1992, Woodruff and Rocke 1994), even in high dimensions. They seem to be an excellent choice as robust covariance matrix estimators in multivariate analysis. In the context of principal components analysis, they have been successfully applied by Croux and Haesbroeck (2000).

Although many robust regression approaches have already been proposed in the literature, we think that the approach based on robust covariance matrices merits to be added to the list of available robust regression estimators. A thorough comparison with other methods would lead us too far, but let us mention some important advantages.

First of all, a robust estimate for the covariance matrix of the estimator is available. Practitioners ask for standard errors around their robust estimates, but robust standard errors are not so often available in the literature (among the exceptions are the bounded influence regression estimators proposed by Chang, Mckean, Naranjo and Sheather 1999 and Ferretti et al. 1999). Secondly, a similar approach can be applied to more general regression models, like multivariate regression (Rousseeuw, Van Aelst en Van Driessen 2000) and calibration models (Cheng and Van Ness 1997). Finally, the method is simple and easy to explain.

7 Appendix

To prove Theorem 1 we will use the following lemma.

Lemma 1. *If the function ρ satisfies condition (R) and the distribution F satisfies condition (F), then the function*

$$\lambda_{\sigma,c}(t) = \int \rho(\sqrt{(y-t)^2\sigma + c}\,)\, dF(y)$$

is symmetric and increasing on $[0, +\infty[$ for every $\sigma > 0$, $c \geq 0$. Moreover, for $c < c^* = \inf\{t > 0 \,|\, \rho(t) = \rho(\infty)\}$, $\lambda_{\sigma,c}(t)$ is strictly increasing on $[0, +\infty[$.

Proof of Lemma 1: The symmetry of $\lambda_{\sigma,c}$ follows from the symmetry of F:

$$\lambda_{\sigma,c}(-t) = \int \rho\left(((y+t)^2\sigma + c)^{1/2}\right) dF(y) = \int \rho\left(((-y+t)^2\sigma + c)^{1/2}\right) dF(y) = \lambda_{\sigma,c}(t).$$

Now $\lambda_{\sigma,c}$ has a positive derivative $\lambda'_{\sigma,c}(t)$ on $]0, +\infty[$ which can be seen as follows:

$$\begin{aligned}
\lambda'_{\sigma,c}(t) &= \frac{\partial}{\partial t} \int \rho\left(((y-t)^2\sigma + c)^{1/2}\right) dF(y) \\
&= -\sigma \int \frac{(y-t)}{((y-t)^2\sigma + c)^{1/2}}\, \psi\left(((y-t)^2\sigma + c)^{1/2}\right) f(y)\, dy \\
&= -\sigma \left\{ \int_{-\infty}^{t} \frac{(y-t)}{((y-t)^2\sigma + c)^{1/2}}\, \psi\left(((y-t)^2\sigma + c)^{1/2}\right) f(y)\, dy \right. \\
&\quad \left. + \int_{t}^{\infty} \frac{(y-t)}{((y-t)^2\sigma + c)^{1/2}}\, \psi\left(((y-t)^2\sigma + c)^{1/2}\right) f(y)\, dy \right\}.
\end{aligned}$$

By transforming the integration variables in these last two integrals, we obtain

$$\begin{aligned}
\lambda'_{\sigma,c}(t) &= -\sigma \left\{ -\int_0^{\infty} \frac{s}{s^2\sigma + c}\, \psi\left((s^2\sigma + c)^{1/2}\right) f(t-s)\, ds \right. \\
&\quad \left. + \int_0^{\infty} \frac{s}{s^2\sigma + c}\, \psi\left((s^2\sigma + c)^{1/2}\right) f(t+s)\, ds \right\} \\
&= \sigma \int_0^{\infty} \frac{s}{s^2\sigma + c}\, \psi\left((s^2\sigma + c)^{1/2}\right) [f(t-s) - f(t+s)]\, ds.
\end{aligned}$$

For every $s, t > 0$ we have $f(t-s) - f(t+s) > 0$ from the unimodality of F. Condition (R) ensures that $\psi\left((s^2\sigma + c)^{1/2}\right) \geq 0$ implying that $\lambda'_{\sigma,c}(t) \geq 0$. Moreover, if $c < c^*$, then $\{s > 0 \,|\, \psi\left((s^2\sigma + c)^{1/2}\right) > 0\}$ has a non-zero Lebesgue measure, so that in this case $\lambda'_{\sigma,c}(t) > 0$ for $t > 0$. □

Proof of Theorem 1: First of all, due to equivariance, we may suppose that $\alpha = 0$ and $\beta = 0$, so $y = \varepsilon$. Lopuhaä (1989) has shown that a solution $(M(H), S(H))$ of problem (2.1) always exists. It is now sufficient to prove that $M_y(H) = 0$ and $S_{uy}(H) = 0$, which will imply immediately that $T(H) = 0 = (\alpha, \beta^t)^t$. Denote $M \equiv M(H)$, $S \equiv S(H)$ and

$$S^{-1} = \begin{pmatrix} S^{uu} & S^{uy} \\ S^{yu} & S^{yy} \end{pmatrix}$$

where $0 < S^{yy} < \infty$ since S is a positive definite matrix. Suppose that (i) $S^{uy} \neq 0$ or (ii) ($S^{uy} = 0$ and $M_y \neq 0$). With $\tilde{S}^{uu} = S^{uu} - (S^{uy}S^{yu})/S^{yy}$, define $\tilde{S}$ by

$$\tilde{S}^{-1} = \begin{pmatrix} \tilde{S}^{uu} & 0 \\ 0 & S^{yy} \end{pmatrix} \text{ and put } \tilde{M} = \begin{pmatrix} M_u \\ 0 \end{pmatrix}.$$

Now by definition of $T(H) = (M(H), S(H))$, and using independence of y and u, we may write

$$b \geq \iint \rho(\alpha_u(y)^{1/2})\, dF(y)\, dG(u)$$

with

$$\alpha_u(y) = (u - M_u)^t S^{uu}(u - M_u) + (y - M_y)^2 S^{yy} + 2(u - M_u)^t S^{uy}(y - M_y)$$

With $t(u) = M_y - \frac{(u-M_u)^t S^{uy}}{S^{yy}} \in I\!R$, we have that $\alpha_u(y) = (y - t(u))^2 S^{yy} + (u - M_u)^t \tilde{S}^{uu}(u - M_u)$. From Lemma 1 it follows that the function

$$t \to \int \rho\left(((y-t)^2 S^{yy} + (u - M_u)^t \tilde{S}^{uu}(u - M_u))^{1/2} \right) dF(y)$$

is symmetric and increasing on $[0, +\infty[$. Therefore, it holds for every u that

$$\int \rho(\alpha_u(y)^{1/2})\, dF(y) \geq \int \rho\left((y^2 S^{yy} + (u - M_u)^t \tilde{S}^{uu}(u - M_u))^{1/2} \right) dF(y),$$

with strict inequality if $t(u) \neq 0$ and $c_u < c^*$, where $c_u = \sqrt{(u - M_u)^t \tilde{S}^{uu}(u - M_u)}$ and c^* defined as in Lemma 1. Denote $A = \{u \,|\, t(u) = 0\}$ and $B = \{u \,|\, c_u \geq c^*\}$. Since for all $u \in B$, $\alpha_u(y)^{1/2} \geq c^*$ for every y, we have that

$$b \geq E_H[\rho(\alpha_u(y)^{1/2})] \geq E_H[\rho(\alpha_u(y)^{1/2}) I(u \in B)] = \rho(\infty) P(B).$$

If $P_G(A \cup B) = 1$, then we would have that $P(A) \geq 1 - P(B) \geq 1 - \frac{b}{\rho(\infty)}$ contradicting hypothesis (G), since A forms a hyperplane in $I\!R^p$. Therefore, we have $P_G(A \cup B) < 1$ and

$$\begin{aligned} b \geq \iint \rho(\alpha_u(y)^{1/2})\, dF(y)\, dG(u) &> \iint \rho\left((y^2 S^{yy} + (u - M_u)^t \tilde{S}^{uu}(u - M_u))^{1/2} \right) dF(y)\, dG(u) \\ &= \int \rho\left(((z - \tilde{M})^t \tilde{S}^{-1}(z - \tilde{M}))^{1/2} \right) dH(z), \end{aligned}$$

while at the same time $\det(S) = \det(\tilde{S})$. Therefore there exists a constant $c < 1$ such that $E_H\left[\rho\left(((z-\tilde{M})^t(c\tilde{S})^{-1}(z-\tilde{M}))^{1/2}\right)\right] \leq b$ while $\det(c\tilde{S}) = c^{p+1}\det(\tilde{S}) < \det(S)$, hereby contradicting the definition of $T(H) = (M(H), S(H))$. We conclude that case (i) and (ii) are excluded, and therefore $M_y = 0$ and $S^{uy} = 0$ (which implies $S_{uy} = 0$). $\square$

To prove Theorem 2 we need the following two lemmas.

Lemma 2. *From the first order condition (3.2) for the scatter matrix functional S, it follows that*

$$\left[\frac{2}{S_{yy}(H_0)}\int w_1'(d^2_{H_0}(z))\, y^2\, \tilde{u}\, dH_0(z)\right] \mathrm{IF}(z; M_y, H_0) +$$
$$\left[\int w_1(d^2_{H_0}(z))\, \tilde{u}\tilde{u}^t\, dH_0(z) + \frac{2}{S_{yy}(H_0)}\int w_1'(d^2_{H_0}(z))\, y^2\, \tilde{u}\tilde{u}^t\, dH_0(z)\right] S_{uu}^{-1}(H_0)\mathrm{IF}(z; S_{uy}, H_0) =$$
$$w_1(d^2_{H_0}(z)) y\tilde{u}. \tag{7.1}$$

where $\tilde{u} = u - M_u(H_0)$.

Proof of Lemma 2: Consider the contaminated distribution $H_\varepsilon = (1-\varepsilon)H_0 + \varepsilon\Delta_z$. Lopuhaä (1989) has shown that a solution of problem (2.1) exists for contaminated distributions of this type when ε is sufficiently small. From the (u, y) component of equation (3.2) we obtain

$$(1-\varepsilon)\int w_1(d^2_{H_\varepsilon}(z))(u - M_u(H_\varepsilon))(y - M_y(H_\varepsilon))\, dH_0(z) + \varepsilon\, w_1(d^2_{H_\varepsilon}(z))(u - M_u(H_\varepsilon))(y - M_y(H_\varepsilon)) =$$
$$(1-\varepsilon)\int w_2(d^2_{H_\varepsilon}(z))\, dH_0(z)\, S_{uy}(H_\varepsilon) + \varepsilon\, w_2(d^2_{H_\varepsilon}(z))\, S_{uy}(H_\varepsilon).$$

Differentiating both sides of the above equation w.r.t. ε and evaluating at 0 yields

$$-\int w_1(d^2_{H_0}(z))\tilde{u}(y - M_y(H_0))\, dH_0(z) + \int w_1'(d^2_{H_0}(z))\frac{\partial}{\partial\varepsilon} d^2_{H_\varepsilon}(z)\Big|_{\varepsilon=0}\, \tilde{u}(y - M_y(H_0))\, dH_0(z) -$$
$$\int w_1(d^2_{H_0}(z))\mathrm{IF}(z; M_u, H_0)(y - M_y(H_0))\, dH_0(z) - \int w_1(d^2_{H_0}(z))\tilde{u}\mathrm{IF}(z; M_y, H_0)\, dH_0(z) +$$
$$w_1(d^2_{H_0}(z))\tilde{u}(y - M_y(H_0)) = -\int w_2(d^2_{H_0}(z))\, dH_0(z)\, S_{uy}(H_0) + w_2(d^2_{H_0}(z))\, S_{uy}(H_0) +$$
$$\int w_2'(d^2_{H_0}(z))\frac{\partial}{\partial\varepsilon} d^2_{H_\varepsilon}(z)\Big|_{\varepsilon=0}\, dH_0(z)\, S_{uy}(H_0) + \int w_2(d^2_{H_0}(z))\, dH_0(z)\, \mathrm{IF}(z; S_{uy}, H_0).$$

where $\tilde{u} = u - M_u(H_0)$. From (3.2) it follows that the first term on the left hand side equals the first term on the right hand side in the above equation. Since $M_y(H_0) = 0$ and $S_{uy}(H_0) = 0$ and using that

$$\frac{\partial}{\partial\varepsilon} d^2_{H_\varepsilon}(z)\Big|_{\varepsilon=0} = (z - M(H_0))^t\mathrm{IF}(z; S^{-1}, H_0)(z - M(H_0)) - 2(z - M(H_0))^t S^{-1}(H_0)\mathrm{IF}(z; M, H_0)$$

the previous equation becomes

$$
\begin{aligned}
&\int w_1'(d^2_{H_0}(z))(z-M(H_0))^t\mathrm{IF}(z;S^{-1},H_0)(z-M(H_0))\tilde{u}y\,dH_0(z) -\\
&2\int w_1'(d^2_{H_0}(z))(z-M(H_0))^tS^{-1}(H_0)\mathrm{IF}(z;M,H_0)\tilde{u}y\,dH_0(z) -\\
&\int w_1(d^2_{H_0}(z))y\,dH_0(z)\ \mathrm{IF}(z;M_u,H_0) - \int w_1(d^2_{H_0}(z))\tilde{u}\,dH_0(z)\ \mathrm{IF}(z;M_y,H_0) +\\
&w_1(d^2_{H_0}(z))\tilde{u}y = \int w_2(d^2_{H_0}(z))\,dH_0(z)\,\mathrm{IF}(z;S_{uy},H_0). \qquad (7.2)
\end{aligned}
$$

Now $\mathrm{IF}(z;S^{-1},H_0) = -S(H_0)^{-1}\mathrm{IF}(z;S,H_0)S(H_0)^{-1}$ (this inequality follows immediately from matrix derivation rules see e.g. (Pullman 1976, page 120), so

$$
\begin{aligned}
&(z-M(H_0))^t\mathrm{IF}(z;S^{-1},H_0)(z-M(H_0)) = (z-M(H_0))^tS(H_0)^{-1}\mathrm{IF}(z;S,H_0)S(H_0)^{-1}(z-M(H_0)) =\\
&\tilde{u}^t\mathrm{IF}(z;S_{uu}^{-1},H_0)\tilde{u} + \mathrm{IF}(z;S_{yy},H_0)(y/S_{yy}(H_0))^2 + 2\tilde{u}^tS_{uu}^{-1}(H_0)\mathrm{IF}(z;S_{uy},H_0)y/S_{yy}(H_0) \qquad (7.3)
\end{aligned}
$$

since $M_y(H_0)=0$ and $S_{uy}(H_0)=0$. Also $d^2_{H_0}(z) = y^2/S_{yy}(H_0) + \tilde{u}^tS_{uu}^{-1}(H_0)\tilde{u}$. Therefore, using (7.3), the first integral in expression (7.2) can be split up into three parts. The first part equals

$$
\int \tilde{u}^t\mathrm{IF}(z;S_{uu}^{-1},H_0)\tilde{u}\left\{\int w_1'(y^2/S_{yy}(H_0)+\tilde{u}^tS_{uu}^{-1}(H_0)\tilde{u})y\,dF(y)\right\}\tilde{u}\,dG(u) = 0,
$$

since the inner integral is zero thanks to symmetry of F. For the same reason we have for the second part

$$
\int\left\{\int w_1'(y^2/S_{yy}(H_0)+\tilde{u}^tS_{uu}^{-1}(H_0)\tilde{u})(y/S_{yy}(H_0))^2y\,dF(y)\right\}dG(u)\mathrm{IF}(z;S_{yy},H_0) = 0.
$$

Therefore, the first integral of equation (7.2) becomes

$$
\begin{aligned}
&\int w_1'(d^2_{H_0}(z))(z-M(H_0))^t\mathrm{IF}(z;S^{-1},H_0)(z-M(H_0))\tilde{u}y\,dH_0(z) =\\
&\qquad\frac{2}{S_{yy}(H_0)}\int w_1'(d^2_{H_0}(z))\,y^2\,\tilde{u}\tilde{u}^t\,dH_0(z)\ S_{uu}^{-1}(H_0)\mathrm{IF}(z;S_{uy},H_0). \qquad (7.4)
\end{aligned}
$$

The second integral of equation (7.2) can be split up into two parts by using

$$
(z-M(H_0))^tS^{-1}(H_0)\mathrm{IF}(z;M,H_0) = \tilde{u}^tS_{uu}^{-1}(H_0)\mathrm{IF}(z;M_u,H_0) + \mathrm{IF}(z;M_y,H_0)y/S_{yy}(H_0).
$$

The first part equals

$$
\int \tilde{u}^tS_{uu}^{-1}(H_0)\mathrm{IF}(z;M_u,H_0)\left\{\int w_1'(y^2/S_{yy}(H_0)+\tilde{u}^tS_{uu}^{-1}(H_0)\tilde{u})y\,dF(y)\right\}\tilde{u}\,dG(u) = 0,
$$

by symmetry of F. Therefore, the second integral of equation (7.2) reduces to

$$\int w_1'(d^2_{H_0}(z))(z - M(H_0))^t S^{-1}(H_0)\mathrm{IF}(z; M, H_0)\tilde{u}y\, dH_0(z) =$$
$$\frac{1}{S_{yy}(H_0)} \int w_1'(d^2_{H_0}(z))\, y^2 \tilde{u}\, dH_0(z)\ \mathrm{IF}(z; M_y, H_0). \tag{7.5}$$

The integral in the third term of expression (7.2) also equals zero by symmetry of F. From the u component of (3.1) it follows that also the fourth term of (7.2) equals zero. Using the (u, u) component of (3.2) the right hand term of (7.2) can be rewritten as

$$\int w_2(d^2_{H_0}(z))\, dH_0(z)\, \mathrm{IF}(z; S_{uy}, H_0) = \int w_1(d^2_{H_0}(z))\tilde{u}\tilde{u}^t\, dH_0(z)\ S^{-1}_{uu}(H_0) \tag{7.6}$$

Substituting (7.4), (7.5), and (7.6) in expression (7.2) yields the desired result. □

Starting from the y component of (3.1), with similar computations as in Lemma 2, the next lemma can be proven.

Lemma 3. *From the first order condition* (3.1) *for the location functional* M*, it follows that*

$$\left[\int w_1(d^2_{H_0}(z))\, dH_0(z) + \frac{2}{S_{yy}(H_0)} \int w_1'(d^2_{H_0}(z))\, y^2\, dH_0(z)\right]\ \mathrm{IF}(z; M_y, H_0) +$$
$$\frac{2}{S_{yy}(H_0)} \int w_1'(d^2_{H_0}(z))\, y^2\, \tilde{u}^t\, dH_0(z)\ S^{-1}_{uu}(H_0)\mathrm{IF}(z; S_{uy}, H_0) = w_1(d^2_{H_0}(z))y \tag{7.7}$$

where $\tilde{u} = u - M_u(H_0)$.

Using Lemma 2 and lemma 3 given above we can prove Theorem 2.

Proof of Theorem 2: We first derive the influence function at H_0. We write

$$T(H_0) = \begin{pmatrix} a(H_0) \\ b(H_0) \end{pmatrix} = Q(H_0)^{-1} \begin{pmatrix} M_y(H_0) \\ S^{-1}_{uu}(H_0)S_{uy}(H_0) \end{pmatrix},$$

with

$$Q(H_0) = \begin{pmatrix} 1 & M_u^t(H_0) \\ 0 & I_p \end{pmatrix},$$

and I_p the identity matrix. Since $M_y(H_0) = 0$ and $S_{uy}(H_0) = 0$ it follows that

$$\mathrm{IF}(z; T, H_0) = Q(H_0)^{-1} \begin{pmatrix} \mathrm{IF}(z; M_y, H_0) \\ S^{-1}_{uu}(H_0)\mathrm{IF}(z; S_{uy}, H_0) \end{pmatrix}. \tag{7.8}$$

We can combine Lemma 2 and Lemma 3 in one single equation:

$$\left[\int w_1(d^2_{H_0}(z))\,\tilde{x}\tilde{x}^t\,dH_0(z) + \frac{2}{S_{yy}(H_0)}\int w_1'(d^2_{H_0}(z))\,y^2\,\tilde{x}\tilde{x}^t\,dH_0(z)\right]\begin{pmatrix}\mathrm{IF}(z;M_y,H_0)\\ S_{uu}^{-1}(H_0)\mathrm{IF}(z;S_{uy},H_0)\end{pmatrix} = w_1(d^2_{H_0}(z))y\tilde{x} \tag{7.9}$$

with $\tilde{x}=(1,\tilde{u}^t)^t$ and where we used $\int w_1(d^2_{H_0}(z))\,\tilde{u}\,dH_0(z)=0$ (which follows from (3.1)). Together with expression (7.8) this yields

$$\mathrm{IF}(z;T,H_0) = Q(H_0)^{-1}A(H_0)^{-1}w_1(d^2_{H_0}(z))\tilde{x}y \tag{7.10}$$

where

$$A(H_0) = \int w_1(d^2_{H_0}(z))\,\tilde{x}\tilde{x}^t\,dH_0(z) + \frac{2}{S_{yy}(H_0)}\int w_1'(d^2_{H_0}(z))\,y^2\,\tilde{x}\tilde{x}^t\,dH_0(z) \tag{7.11}$$

Now we can easily check that $Q(H_0)^t\tilde{x}=(1,u^t)^t=x$. It follows from (7.10) and $Q(H_0)^{-1}=-Q(H_0)$ that

$$\begin{aligned}\mathrm{IF}(z;T,H_0) &= -Q(H_0)^{-1}A(H_0)^{-1}Q(H_0)^t w_1(d^2_{H_0}(z))xy\\ &= (Q(H_0)^tA(H_0)Q(H_0))^{-1}w_1(d^2_{H_0}(z))xy\\ &= C(H_0)^{-1}w_1(d^2_{H_0}(z))xy \end{aligned}\tag{7.12}$$

where $C(H_0)=Q(H_0)^tA(H_0)Q(H_0)$ equals, using (7.11),

$$C(H_0) = \int w_1(d^2_{H_0}(z))\,xx^t\,dH_0(z) + \frac{2}{S_{yy}(H_0)}\int w_1'(d^2_{H_0}(z))\,y^2\,xx^t\,dH_0(z). \tag{7.13}$$

Using integration by parts, expression (7.13) can be rewritten as (3.5).

Finally, note that if $z=(u^t,y)^t\sim H$, then $Az+c\sim H_0$ with $A=\begin{pmatrix} I_p & 0\\ -\beta^t & 1\end{pmatrix}$, and $c=\begin{pmatrix}0\\-\alpha\end{pmatrix}$. By equivariance of the functional T we have $T(H)=T(H_0)+\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$ and therefore

$$\mathrm{IF}(z;T,H) = \mathrm{IF}(Az+c;T,H_0) = \mathrm{IF}((u,y-x^t\theta);T,H_0) \tag{7.14}$$

Due to the affine equivariance of the S-estimator, we have

$$\begin{aligned} d^2_H(z) &= (z-M(H))^tS(H)^{-1}(z-M(H))\\ &= (Az-(M(H_0)-c))^tS(H_0)^{-1}(Az-(M(H_0)-c))\\ &= d^2_{H_0}(Az+c)\end{aligned}$$

for all $z\in I\!R^{p+1}$. Therefore, it follows from (7.14) and (7.12) that

$$\mathrm{IF}(z;T,H) = C(H_0)^{-1}w_1(d^2_H(z))x(y-x^t\theta). \qquad \square$$

References

Chang, W. H., J. W. Mckean, J. D. Naranjo, and S. J. Sheather (1999), High-breakdown rank regression, *Journal of the American Statistical Association* **94**, 205-219.

Cheng, C. and J. W. Van Ness (1997), Robust calibration, *Technometrics* **39**, 401-411.

Coakley, C. W. and T. P. Hettmansperger (1993), A bounded influence, high breakdown, efficient regression estimator, *Journal of the American Statistical Association* **88**, 872-880.

Cox, T. F. and M. A. A. Cox (1994), *Multidimensional scaling,* Chapman and Hall, London.

Croux, C. and G. Haesbroeck (1999), Influence function and efficiency of the minimum covariance determinant scatter matrix estimator, *Journal of Multivariate Analysis* **71**, 161–190.

Croux, C. and G. Haesbroeck (2000), Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies, *Biometrika,* to appear.

Croux, C., P. J. Rousseeuw, and O. Hössjer (1994), Generalized S-estimators, *Journal of the American Statistical Association* **89**, 1271-1281.

Davies, L. (1987), Asymptotic behavior of S-estimators of multivariate location parameters and dispersion matrices, *The Annals of Statistics* **15**, 1269-1292.

Ferretti, N., D. Kelmansky, V. J. Yohai, and R. H. Zamar (1999), A class of locally and globally robust regression estimates, *Journal of the American Statistical Association* **94**, 174-188.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (1986), *Robust statistics: the approach based on influence functions,* John Wiley and Sons, New York.

Härdle, W. (1991), *Smoothing techniques with implementation in S,* Springer-Verlag, New York.

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

Iglewicz, B. (1982), Robust Scale Estimates, in: D. C. Hoaglin, F. Mosteller, and J. W. Tukey (eds.), *Understanding robust and exploratory data Analysis*, John Wiley and Sons, New York.

Lopuhaä H. P. (1989), On the relation between S-estimators and M-estimators of multivariate location and covariance, *The Annals of Statistics* **17**, 1662-1683.

Maronna, R. (1976), Robust M-estimates of multivariate location and scatter, *The Annals of Statistics* **4**, 51-67.

Maronna, R. and S. Morgenthaler (1986), Robust regression through robust covariances, *Communications in Statistics: Theory and Methods* **15**, 1347-1365.

Pullman, N. J. (1976), *Matrix theory and its applications: selected topics,* Marcel Dekker, New York.

Rousseeuw, P. J. (1984), Least median of squares regression, *Journal of the American Statistical Association* **79**, 871–880.

Rousseeuw, P. J. (1985), Multivariate estimation with high breakdown point, in: W. Grossmann, G. Pflug, I. Vincze and W. Wertz (eds.), *Mathematical statistics and applications, Vol. B*, Reidel, Dordrecht, 283-297.

Rousseeuw, P. J. and A. M. Leroy (1987), *Robust regression and outlier detection,* Wiley-Interscience, New York.

Rousseeuw, P. J., S. Van Aelst, and K. Van Driessen (2000), Robust multivariate regression, Technical Report, University of Antwerp.

Rousseeuw, P. J. and B. C. van Zomeren (1990), Unmasking multivariate outliers and leverage points, *Journal of the American Statistical Association* **85**, 633-651.

Ruppert, D. (1992), Computing S-estimators for regression and multivariate location/dispersion, *Journal of Computational and Graphical Statistics* **1**, 253-270.

Simpson, D. G., D. Ruppert, and R. J. Carroll (1992), On one-step GM estimates and stability of inferences in linear regression, *Journal of the American Statistical Association* **87**, 439-450.

Stahel, W.A. (1981), Robust Estimation: Infinitesimal optimality and covariance matrix estimators, PhD. Thesis (in German), ETH, Zurich.

Visuri, S., V. Koivunen, J. Möttönen, E. Ollila, and H. Oja (2000), Affine equivariant multivariate rank methods, submitted.

Woodruff, D. L. and D. M. Rocke (1994), Computable robust estimation of multivariate location and shape in high dimension using compound estimators, *Journal of the American Statistical Association* **89**, 888-896.

Yohai, V. J. (1987), High breakdown point and high efficiency robust estimates for regression, *The Annals of Statistics* **15**, 642-656.

Yohai, V. J., W. A. Stahel, and R. H. Zamar (1991), A procedure for robust estimation and inference in linear regression, in: W. Stahel and S. Weisberg (eds.), *Directions in robust statistics and diagnostics,* Springer-Verlag, New York, 365-374.

Yohai, V. J. and R. H. Zamar (1988), High breakdown-point estimates of regression by means of the minimization of an efficient scale, *Journal of the American Statistical Association* **83**, 406-413.