Robust Multivariate Regression

Peter J. Rousseeuw, Stefan Van Aelst, Katrien Van Driessen
and Jose Agulló

September 17, 2003

Abstract

We introduce a robust method for multivariate regression, based on robust estimation of the joint location and scatter matrix of the explanatory and response variables. As robust estimator of location and scatter we use the minimum covariance determinant (MCD) estimator of Rousseeuw (1984). Based on simulations we investigate the finite-sample performance and robustness of the estimator. To increase the efficiency we propose a reweighted estimator, which was selected from several possible reweighting schemes. The resulting multivariate regression does not need much computation time and is applied to real datasets. We shown that the multivariate regression estimator has the appropriate equivariance properties, has a bounded influence function, and inherits the breakdown value of the MCD estimator. These theoretical robustness properties confirm the good finite-sample results obtained from the simulations.

Key words: Breakdown value; Diagnostic plot; Influence function; Minimum covariance determinant; Reweighting.

Peter J. Rousseeuw is Professor, Dept. of Mathematics and Computer Science, University of Antwerp, Midelheimlaan 1, B-2020 Antwerp, Belgium. Stefan Van Aelst is Assistant Professor, Dept. of Applied Mathematics and Computer Science, Ghent University, Krijgslaan 281 S9, B-9000 Ghent, Belgium. Katrien Van Driessen is Assistant, Dept. of Applied Economics, University of Antwerp, Prinsstraat 13, B-2000 Antwerp, Belgium. Jose Agulló is Professor at Departamento de Fundamentos del Análisis Económico, University of Alicante, E-03080 Alicante, Spain. This research started while Stefan Van Aelst was Postdoctoral Researcher with the FWO-Flanders (Belgium).

1 Introduction

Suppose we have a p-variate predictor $x = (x_1, \ldots, x_p)^t$ and a q-variate response $y = (y_1, \ldots, y_q)^t$. The multivariate regression model is given by $y = \mathcal{B}^t x + \alpha + \varepsilon$ where $\mathcal{B}$ is the $(p \times q)$ slope matrix, α is the q-dimensional intercept vector, and the errors $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_q)^t$ are i.i.d. with zero mean and with $\mathrm{Cov}(\varepsilon) = \Sigma_\varepsilon$ a positive definite matrix of size q. Let us denote the location of the joint (x, y) variables by μ and their scatter matrix by Σ. Partitioning μ and Σ yields the notation

$$\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \qquad \text{and} \qquad \Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}.$$

Traditionally, μ is often estimated by the empirical mean $\hat{\mu}$ and Σ by the empirical covariance matrix $\hat{\Sigma}$. It turns out that the least squares estimator of $\mathcal{B}$, α and Σ_ε can be written as a function of the components of $\hat{\mu}$ and $\hat{\Sigma}$, namely

$$\hat{\mathcal{B}} = \hat{\Sigma}_{xx}^{-1} \hat{\Sigma}_{xy} \tag{1}$$

$$\hat{\alpha} = \hat{\mu}_y - \hat{\mathcal{B}}^t \hat{\mu}_x \tag{2}$$

$$\hat{\Sigma}_\varepsilon = \hat{\Sigma}_{yy} - \hat{\mathcal{B}}^t \hat{\Sigma}_{xx} \hat{\mathcal{B}} \tag{3}$$

(see e.g. Johnson and Wichern 1998, page 440). Multivariate regression has applications in chemometrics, engineering, econometrics, psychometrics, and other fields. Recent work on multivariate regression includes Barret and Ling (1992), Breiman and Friedman (1997), Cook and Setodji (2003), Davis and McKean (1993), Gleser (1992), Koenker and Portnoy (1990) and Ollila et al. (2002, 2003).

It is well known that classical multiple regression is extremely sensitive to outliers in the data. This problem also holds in the case of multivariate regression as can be seen from the following example.

Example 1. We consider a dataset (Lee 1992) which contains measurements of properties of pulp fibres and the paper made from them. The aim is to investigate relations between pulp fibre properties and the resulting paper properties. The dataset contains $n = 62$ measurements of the following four pulp fibre characteristics: arithmetic fibre length, long fibre fraction, fines fibre fraction, and zero span tensile. The four paper properties that have been measured are breaking length, elastic modulus, stress at failure, and burst strength. The dataset is available at http://win-www.uia.ac.be/u/statis/datasets/pulpfibre.html.

Our goal is to predict the $q = 4$ paper properties from the $p = 4$ fibre characteristics. For this purpose, we first applied classical multivariate regression to the data.

Figure 1 represents the result of the classical analysis. It plots the Mahalanobis distances of the residuals $r_i = y_i - \hat{\mathcal{B}}^t x_i - \hat{\alpha}$ as given by

$$d(r_i) := \sqrt{r_i^t (\hat{\Sigma}_\varepsilon)^{-1} r_i}$$

versus the Mahalanobis distances of the carriers:

$$d(x_i) := \sqrt{(x_i - \mu_x)^t (\hat{\Sigma}_{xx})^{-1} (x_i - \mu_x)}.$$

This diagnostic plot combines the information on regression outliers and leverage points, and is much more useful than either distance separately. The horizontal and vertical lines are the usual cutoff values $\sqrt{\chi^2_{p,0.975}}$ and $\sqrt{\chi^2_{q,0.975}}$ which both equal 3.34 since $p = q = 4$ in this example. From this plot we see that observations $51, 52$, and 56 are detected as vertical outliers. On the other hand, some observations are identified as leverage points (observations 60 and 61 are the largest) but they are not considered to be regression outliers because they have small residual distance.

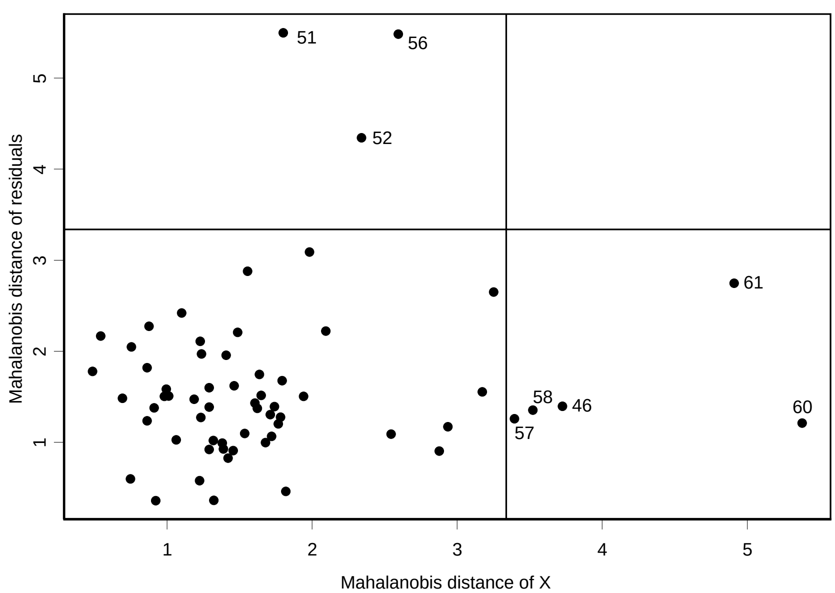

Figure 1: Plot of Mahalanobis distances of LS residuals versus Mahalanobis distances of the carriers for the pulp fibre data.

To check the result obtained by classical multivariate regression we start by applying univariate robust regression with the same regressors but for each of the responses separately. Here we use the least trimmed squares (LTS) estimator of Rousseeuw (1984) which can be computed quickly with the FAST-LTS algorithm of Rousseeuw and Van Driessen (2000). To obtain reliable outlier identification we use the reweighted LTS with finite-sample correction factor as proposed by Pison et al. (2002).

Figure 2 shows the standardized residuals resulting from LTS regression with the first response (breaking length). From this plot we immediately see that observations $51, 52, 56$ and 61 are detected as outliers. Similarly, outliers can be identified from standardized LTS residuals corresponding to the other three responses. Table 1 summarizes the outliers detected by applying LTS for each of the four responses. From Table 1 we see that the

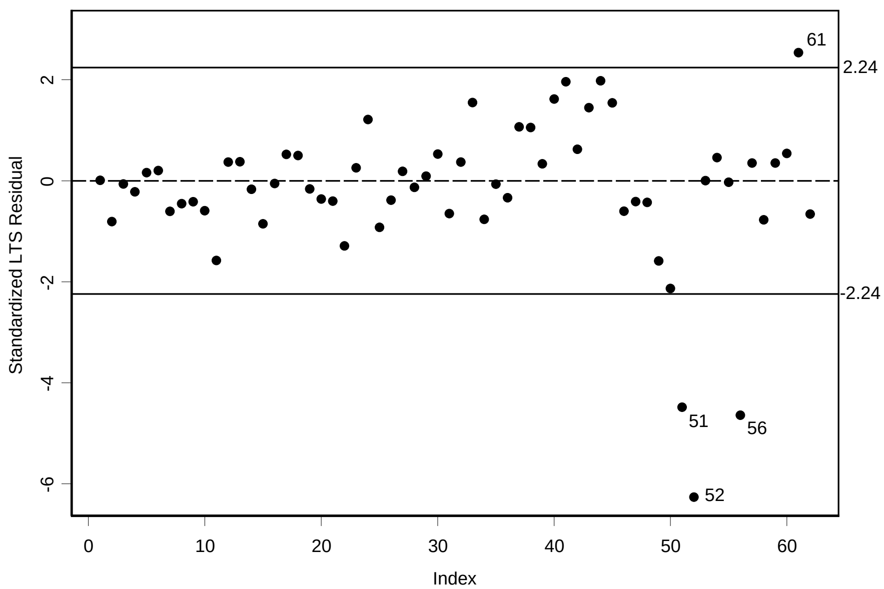

Figure 2: Plot of the standardized LTS residuals corresponding to the first response (breaking length) versus the case number.

univariate LTS regressions identify observations $51, 52, 56$ and 61 as outliers. This already shows that the classical multivariate regression based on least squares in Figure 1 has been influenced by outliers since it did not detect observation 61 as a regression outlier. Hence, clearly the least squares multivariate regression has been influenced by this leverage point. This analysis shows that we need robust estimators to investigate these data. However,

response	outliers
y_1	51,52,56,61
y_2	61
y_3	52,56,61
y_4	51,52,56

Table 1: List of observations in the pulp fibre data that are detected as outliers by applying LTS regression to each of the four responses separately.

applying univariate LTS regressions to each of the response variables separately does not yield a solution which is equivariant under affine transformations of the response variables. Moreover, this approach only allows us to detect outliers in the coordinate directions of the responses but does not detect outliers that are masked in these directions. Therefore, we aim to construct a robust method for multivariate regression which allows us to detect all the outliers and is also reasonably efficient in both the statistical and the computational sense. After developing such a robust method we will analyze these data further in Section 5.

In the next section we introduce a robust method for multivariate regression based on the minimum covariance determinant estimate of the joint (x, y) variables. We study the performance of the estimator by simulations. In section 3 we investigate several reweighted versions of the estimator which improve the performance of the initial estimator, and select the reweighting scheme that works best. The finite-sample robustness of the optimal estimator is studied in Section 4. In Section 5 we continue the analysis of the previous example, and describe an application to chemical engineering. Section 6 shows that the robust estimator has the equivariance properties that we expect from a multivariate regression method. In Section 7 we discuss the robustness properties of the estimator and derive studentized residual distances. Section 8 summarizes our conclusions. All proofs are given in the Appendix.

2 MCD regression

We propose to use robust estimators for the center μ and scatter matrix Σ in expressions (1) to (3) to construct a robust multivariate regression method. This robust method will have the equivariance properties required for a multivariate regression estimator. Many robust estimators of multivariate location and scatter have been investigated in the literature, such as

M-estimators (Maronna 1976), the minimum volume ellipsoid and minimum covariance determinant (MCD) estimator (Rousseeuw 1984, 1985), S-estimators (Davies 1987, Rousseeuw and Leroy 1987, Lopuhaä 1989), CM-estimators (Kent and Tyler 1996), and τ-estimators (Lopuhaä 1991). More recently, also depth based location and scatter estimators have been introduced (Zuo et al. 2001, Zuo and Cui 2002). Robust estimators of location and scatter in high dimensions have been investigated by Woodruff and Rocke (1994), Rocke (1996), and Rocke and Woodruff (1996). In the multiple regression case Maronna and Morgenthaler (1986) used multivariate M-estimators in (1)–(3), but their method inherits the low breakdown value of M-estimators. A multivariate regression method of M-type was proposed by Koenker and Portnoy (1990), who noted that their method lacks affine equivariance.

We will use the MCD to estimate the center and scatter matrix of the joint (x, y) variables because the MCD is a robust estimator with high breakdown value and bounded influence function (Croux and Haesbroeck 1999). Moreover, the MCD estimator is asymptotically normal (Butler, Davies, and Jhun 1993). The resulting robust multivariate regression method will be called *MCD regression*.

Consider a dataset $Z_n = \{z_i; i = 1, \ldots, n\} \in I\!R^{p+q}$. The MCD looks for the subset $\{z_{i_1}, \ldots, z_{i_h}\}$ of size h whose covariance matrix has the smallest determinant, where $\lceil n/2 \rceil \leq h \leq n$. We will denote $\gamma = (n-h)/n$ so $0 \leq \gamma \leq 0.5$. The estimate for the center is then defined as the mean $t_n = \frac{1}{h}\sum_{j=1}^{h} z_{i_j}$ and the covariance estimate is given by $C_n = c_n\, c_\gamma \frac{1}{h}\sum_{j=1}^{h}(z_{i_j} - t_n)(z_{i_j} - t_n)^t$ where c_γ is a consistency factor and c_n is a small sample correction factor (see Pison et al. 2002). The MCD estimator has breakdown value approximately equal to γ. Two common choices for h are $h = [(n+p+q+1)/2] \approx n/2$ so $\gamma \approx 0.5$ which yields the highest possible breakdown value, and $h \approx 3n/4$ (i.e. $\gamma \approx 0.25$) which gives a better compromise between efficiency and breakdown. Recently, Rousseeuw and Van Driessen (1999) constructed a fast algorithm to compute the MCD. This FAST-MCD algorithm made the MCD very useful for analyzing large datasets, e.g. with n in the hundred thousands. Other robust methods to analyze large datasets have been developed and used by Knorr et al. (2001), Alqallaf et al. (2002), and Maronna and Zamar (2002).

Since computation of the MCD regression estimates consists of computation of the MCD of the joint (x, y) variables followed by standard matrix operations, we obtain a computationally efficient method. Moreover, from (1) to (3) we immediately see that regressions of all possible splits in x and y variables can be carried out once the MCD of the joint (x, y) variables has been computed. It has been shown that observations which lie far from the center can have only a small effect on the MCD estimates. Therefore, both leverage points (which have a large x-distance) and regression outliers (which are deviating in y-space) can have only a small effect on the MCD regression estimates. However, it has been noted that the MCD can have a low efficiency (Croux and Haesbroeck 1999).

To investigate the efficiency of the MCD regression we performed the following simulation study. For various sample sizes n, and for different choices of p and q, we generated m datasets of size n from the multivariate standard Gaussian distribution $N(0, I_{p+q})$, which corresponds to putting $\mathcal{B} = 0$ and $\alpha = 0$. For each dataset $Z^{(l)}, l = 1, \ldots, m$ we carried out MCD regression yielding the $(p \times q)$ slope matrix estimate $\hat{\mathcal{B}}^{(l)}$, the intercept vector $\hat{\alpha}^{(l)} \in I\!R^q$, and the $(q \times q)$ covariance matrix estimate $\hat{\Sigma}_\varepsilon^{(l)}$ of the errors.

The Monte Carlo variance of a slope coefficient $\hat{\mathcal{B}}_{jk}$ is measured as

$$\mathrm{Var}(\hat{\mathcal{B}}_{jk}) = n \operatorname*{var}_l(\hat{\mathcal{B}}_{jk}^{(l)}) \quad \text{for } j = 1, \ldots, p \text{ and } k = 1, \ldots, q. \tag{4}$$

The overall variance of the estimated matrix $\hat{\mathcal{B}}$ is defined as $\mathrm{Var}(\hat{\mathcal{B}}) = \mathrm{ave}_{j,k}(\mathrm{Var}\hat{\mathcal{B}}_{jk})$. The corresponding finite-sample efficiency of the slope is then given by $1/\mathrm{Var}(\hat{\mathcal{B}})$. Analogously we compute the finite-sample efficiency of the intercept vector. To measure the accuracy of the error scatter matrix, we use the standardized variance (Bickel and Lehmann 1976) of the elements of the error covariance matrix, defined as

$$\mathrm{Stvar}((\hat{\Sigma}_\varepsilon)_{jk}) = \frac{n \operatorname*{var}_l((\hat{\Sigma}_\varepsilon^{(l)})_{jk})}{[\operatorname*{ave}_l \operatorname*{ave}_j((\hat{\Sigma}_\varepsilon^{(l)})_{jj})]^2} \quad \text{for } j = 1, \ldots, q \text{ and } k = 1, \ldots, q. \tag{5}$$

The overall finite-sample efficiency of the off-diagonal elements is then given by $1/\operatorname*{ave}_{j \neq k}(\mathrm{Stvar}((\hat{\Sigma}_\varepsilon)_{jk}))$. For the diagonal elements the finite-sample efficiency is given by $2/\mathrm{ave}_j(\mathrm{Stvar}((\hat{\Sigma}_\varepsilon)_{jj}))$ since the Fisher information equals 2 in this case.

The top panel of Table 2 shows the simulation results for $p = 4$ and $q = 4$, but the results were similar for many other choices of p and q. The table contains sample sizes between 50 and 500. All simulations were done with $m = 1000$ replications. The cells contain the finite-sample efficiencies of $\hat{\mathcal{B}}$, $\hat{\alpha}$, the diagonal elements of $\hat{\Sigma}_\varepsilon$ and the off-diagonal elements of $\hat{\Sigma}_\varepsilon$. We see that the finite-sample efficiencies are very low for $\gamma = 0.5$ and are somewhat better for $\gamma = 0.25$. In the next section we propose the use of reweighted estimators to improve these efficiencies.

		$n = 50$	$n = 100$	$n = 300$	$n = 500$	$n = \infty$
MCD regression: $\hat{\mathcal{B}}$, $\hat{\alpha}$, $\hat{\Sigma}_\varepsilon$						
$\gamma = 0.50$	slope	0.176	0.167	0.166	0.169	0.166
	intercept	0.268	0.290	0.300	0.298	0.307
	Σ_{diag}	0.211	0.205	0.196	0.190	0.182
	Σ_{offdiag}	0.194	0.183	0.166	0.172	0.166
$\gamma = 0.25$	slope	0.371	0.387	0.401	0.410	0.403
	intercept	0.506	0.545	0.568	0.543	0.578
	Σ_{diag}	0.401	0.431	0.439	0.432	0.430
	Σ_{offdiag}	0.387	0.415	0.393	0.401	0.403
MCD regression with reweighted location: $\hat{\mathcal{B}}^L$, $\hat{\alpha}^L$, $\hat{\Sigma}_\varepsilon^L$						
$\gamma = 0.50$	slope	0.200	0.354	0.662	0.762	0.851
	intercept	0.303	0.525	0.811	0.838	0.934
	Σ_{diag}	0.245	0.391	0.677	0.727	0.794
	Σ_{offdiag}	0.222	0.384	0.664	0.735	0.851
$\gamma = 0.25$	slope	0.403	0.598	0.772	0.830	0.864
	intercept	0.545	0.747	0.883	0.877	0.936
	Σ_{diag}	0.434	0.613	0.793	0.798	0.812
	Σ_{offdiag}	0.427	0.629	0.782	0.813	0.864
MCD regression with reweighted regression: $\hat{\mathcal{B}}^R$, $\hat{\alpha}^R$, $\hat{\Sigma}_\varepsilon^R$						
$\gamma = 0.50$	slope	0.245	0.465	0.812	0.902	0.957
	intercept	0.338	0.582	0.862	0.875	0.959
	Σ_{diag}	0.251	0.387	0.685	0.735	0.858
	Σ_{offdiag}	0.232	0.399	0.684	0.763	0.880
$\gamma = 0.25$	slope	0.538	0.758	0.895	0.948	0.960
	intercept	0.622	0.804	0.927	0.906	0.961
	Σ_{diag}	0.463	0.627	0.820	0.815	0.874
	Σ_{offdiag}	0.462	0.665	0.812	0.841	0.892
MCD regression with reweighted location and regression: $\hat{\mathcal{B}}^{LR}$, $\hat{\alpha}^{LR}$, $\hat{\Sigma}_\varepsilon^{LR}$						
$\gamma = 0.50$	slope	0.233	0.628	0.906	0.955	0.961
	intercept	0.332	0.721	0.928	0.920	0.962
	Σ_{diag}	0.252	0.501	0.826	0.829	0.881
	Σ_{offdiag}	0.233	0.542	0.824	0.860	0.900
$\gamma = 0.25$	slope	0.508	0.801	0.913	0.959	0.961
	intercept	0.614	0.849	0.942	0.924	0.962
	Σ_{diag}	0.458	0.680	0.864	0.839	0.881
	Σ_{offdiag}	0.459	0.728	0.854	0.872	0.900

Table 2: Finite-sample efficiencies of the slope matrix, intercept vector and error covariance matrix of the four types of MCD regression, for $p = 4$ and $q = 4$. The number of replications was $m = 1000$.

3 Reweighted multivariate regression

To increase the efficiencies obtained in the previous section, we now consider reweighted versions of our estimator. These reweighted estimators inherit the robustness properties of the initial estimator, while attaining a higher efficiency. We will consider three versions, based on reweighting the location estimator, reweighting the regression estimator, and both.

3.1 Reweighting the location estimator

To increase the efficiency of the location and scatter estimators it is customary to compute one-step reweighted versions (Rousseeuw and Leroy 1987, Lopuhaä 1999, Zuo et al. 2001, Zuo and Cui 2002). The one-step reweighted MCD estimates with nominal trimming portion δ_l are defined as

$$t_n^1 = \frac{\sum_{i=1}^n w(d^2(z_i))z_i}{\sum_{i=1}^n w(d^2(z_i))} \quad \text{and} \quad C_n^1 = d_{\delta_l}\frac{\sum_{i=1}^n w(d^2(z_i))(z_i - t_n^1)(z_i - t_n^1)^t}{\sum_{i=1}^n w(d^2(z_i))} \tag{6}$$

where d_{δ_l} is a consistency factor. The weights are computed as $w(d^2(z_i)) = I(d^2(z_i) \leq q_{\delta_l})$ where $q_{\delta_l} = \chi^2_{p+q,1-\delta_l}$ and $d(z_i) = ((z_i - t_n)^t C_n^{-1}(z_i - t_n))^{1/2}$ is the robust distance of observation z_i based on the initial MCD estimates (t_n, C_n). It is customary to take $\delta_l = 0.025$ (Rousseeuw and Van Driessen 1999). The robustness properties of the one-step reweighted MCD estimators are similar to those of the initial MCD (Lopuhaä and Rousseeuw 1991, Lopuhaä 1999). Other methods to increase the efficiency of the MCD location and scatter include one-step M-estimators and cross-checking (He and Wang 1996).

We can now compute the multivariate regression estimates (1), (2), and (3) based on the reweighted location and scatter (t_n^1, C_n^1). We denote the resulting regression by $\hat{\mathcal{B}}^L$, $\hat{\alpha}^L$, and $\hat{\Sigma}_\varepsilon^L$ where the "L" indicates that the reweighting was done in the Location stage. The simulation results for the reweighted location estimators are shown in the second panel of Table 2. We see that multivariate regression estimates based on the reweighted MCD have a much higher efficiency than the estimates based on the initial unweighted MCD.

3.2 Reweighting the regression

In a regression analysis it is natural to use weights based on the residuals corresponding to the initial fit (Rousseeuw and Leroy 1987). Denote the residual of the observation z_i by

$r_i = y_i - \hat{\mathcal{B}}^t x_i - \hat{\alpha}$. We now define the reweighted regression estimators

$$T_n^R = (\sum_{i=1}^n w(d^2(r_i))u_i u_i^t)^{-1} \sum_{i=1}^n w(d^2(r_i))y_i u_i \quad (7)$$

$$\hat{\Sigma}_\varepsilon^R = d_{\delta_r} \frac{\sum_{i=1}^n w(d^2(r_i))(r^R)_i (r^R)_i^t}{\sum_{i=1}^n w(d^2(r_i))} \quad (8)$$

where $T_n^R = ((\hat{\mathcal{B}}^R)^t, \hat{\alpha}^R)^t$, $u_i = (x_i^t, 1)^t$, $(r^R)_i = y_i - (\hat{\mathcal{B}}^R)^t x_i - \hat{\alpha}^R$, δ_r is the trimming portion, and d_{δ_r} is a consistency factor. Following Rousseeuw and Leroy (1987) we take $\delta_r = 0.01$ as our default. The superscript R says that the weights were based on the initial Regression. In particular, the weights are computed as $w(d^2(r_i)) = I(d^2(r_i) \leq q_{\delta_r})$ where $q_{\delta_r} = \chi^2_{q,1-\delta_r}$ and $d(r_i) = (r_i^t(\hat{\Sigma}_\varepsilon)^{-1} r_i)^{1/2}$ is the robust distance of the ith residual. The robustness properties of these reweighted regression estimators follow from the properties of the initial regression estimators. Note that the weights now depend only on the size of the residual distance $w(d^2(r_i))$ so in contrast with the initial estimates, good leverage points (which have large distance in x-space but small residual distance and thus are not outliers for the regression model) are not downweighted anymore.

The third panel of Table 2 shows the simulation results for the reweighted regression estimators. We see that the reweighted multivariate regression estimates have a much higher efficiency than the initial estimates based on MCD. Moreover, the efficiency of the reweighted regression estimates is also higher than the efficiency of the estimates based on the reweighted MCD.

3.3 Reweighting both location and regression

A further possibility is to use the robust distances $d(r_i^L) = ((r_i^L)^t(\hat{\Sigma}_\varepsilon^L)^{-1} r_i^L)^{1/2}$ in (7) and (8), where $r_i^L = y_i - (\hat{\mathcal{B}}^L)^t x_i - \hat{\alpha}^L$. This yields a weighted regression estimator with weights based on the residuals of the method in 3.1. We denote the resulting estimators by $T_n^{LR} = ((\hat{\mathcal{B}}^{LR})^t, \hat{\alpha}^{LR})^t$ and $\hat{\Sigma}_\varepsilon^{LR}$. Also here, good leverage points are not downweighted anymore. The simulation results for the reweighted location estimators are shown in the last panel of Table 2. From this table we see that the efficiency of LR-weighting is comparable for small samples ($n = 50$) and clearly better for larger samples than the efficiency of the other reweighting schemes. Overall, we also see that $\gamma = 0.25$ consistently outperformed $\gamma = 0.50$, the difference being larger at small samples. Hence, from the efficiency viewpoint, LR-weighted MCD regression with $\gamma = 0.25$ comes out best. It will be shown in Section 6 that the breakdown value of MCD regression is approximately equal to γ, so Table 2 shows that there is a trade-off between efficiency and breakdown. In practice, data with more than 20% of outliers rarely occur, so we recommend using the LR-weighted MCD regression with $\gamma = 0.25$ as the default to obtain a better efficiency. Only if the data are of very low quality such that a higher level of outliers can be expected, using LR-weighted MCD regression with $\gamma = 0.50$ is more appropriate.

4 Finite-sample robustness

To study the finite-sample robustness, we carried out simulations with datasets contaminated by different types of outliers: A point (x_i, y_i) which does not follow the linear pattern of the majority of the data but whose x_i is not outlying is called a vertical outlier. A point (x_i, y_i) whose x_i is outlying is called a leverage point. We say that such an (x_i, y_i) is a bad leverage point when it does not follow the pattern of the majority, otherwise it is called a good leverage point (which does not harm the fit).

Because regression estimators often break down in the presence of vertical outliers or bad leverage points, we generated datasets with both types of outliers. For sample sizes between $n = 50$ and $n = 500$, and with $p = 4$ and $q = 4$ we generated $m = 1000$ datasets from the multivariate standard Gaussian distribution $N(0, I_{p+q})$. (This is the same situation as described in Section 2.) We then replaced 10% of the data as follows. The x_i are kept, but the q response variables are distributed as $N(2\sqrt{\chi^2_{p+q,.99}}, 0.1)$. This yields vertical outliers, because only their responses are outlying. We also replaced 10% of the data with bad leverage points for which the p independent variables are generated according to $N(2\sqrt{\chi^2_{p,.99}}, 0.1)$ and the q dependent variables are generated from $N(2\sqrt{\chi^2_{q,.99}}, 0.1)$.

As in the previous simulations, for each dataset $Z^{(l)}, l = 1, \ldots, m$ we computed the $(p \times q)$ slope matrix $\hat{\mathcal{B}}^{(l)}$, the intercept vector $\hat{\alpha}^{(l)} \in I\!R^q$ and the $(q \times q)$ covariance matrix $\hat{\Sigma}_\varepsilon^{(l)}$ of the errors. To measure robustness, we use the bias and the MSE. As commonly defined, the bias and MSE of a univariate component T are given by

$$\begin{aligned} \text{bias}(T) &= \underset{l}{\text{ave}}(T^{(l)} - \theta) \\ \text{MSE}(T) &= n\,\underset{l}{\text{ave}}(T^{(l)} - \theta)^2 \end{aligned}$$

with θ the true value of the parameter. The bias and MSE of the slope are defined as

$$\begin{aligned} \text{bias}(\hat{\mathcal{B}}) &= \sqrt{\operatorname*{ave}_{j,k}(\text{bias}(\hat{\mathcal{B}}_{jk})^2)} \\ \text{MSE}(\hat{\mathcal{B}}) &= \operatorname*{ave}_{j,k}(\text{MSE}(\hat{\mathcal{B}}_{jk})) \end{aligned}$$

and similarly for the intercept $\hat{\alpha}$ and for the diagonal and off-diagonal elements of $\hat{\Sigma}_\varepsilon$.

Table 3 shows the simulation results when the estimates of the slope matrix, intercept vector and error covariance matrix were obtained from the LR-weighted method with $\gamma = 0.25$, and from the classical multivariate least squares regression. Simulations for other sample sizes n and different dimensions p and q gave similar results. In Table 3 we see that in the presence of vertical outliers and bad leverage points, both the bias and MSE obtained from the LR-weighted MCD regression are much lower than those obtained from least squares regression. The low bias and MSE values of the LR-weighted method are in line with the asymptotic robustness properties in Section 6.

	n=50		n=100		n=500	
	bias	MSE	bias	MSE	bias	MSE
LR-weighted MCD regression ($\gamma = 0.25$)						
slope	0.0066	1.637	0.0038	1.462	0.0013	1.307
intercept	0.0104	1.501	0.0036	1.415	0.0021	1.336
Σ_{diag}	0.1349	3.326	0.0704	3.240	0.0104	2.845
Σ_{offdiag}	0.0050	1.245	0.0049	1.319	0.0021	1.369
LS regression						
slope	0.2068	10.883	0.2071	12.233	0.2064	28.806
intercept	1.0225	54.275	1.0214	105.876	1.0243	525.924
Σ_{diag}	6.5387	2156.323	6.8122	4655.881	7.0394	24788.392
Σ_{offdiag}	6.8076	2332.350	7.0464	4977.298	7.2440	26246.846

Table 3: Bias and MSE of the slope matrix, intercept vector and error covariance matrix obtained by the LR-weighted MCD regression with $\gamma = 0.25$ and multivariate least squares regression. The data contained 20% of outliers. The number of replications was $m = 1000$.

To compare the MCD regression with the univariate robust regressions approach used in Section 1, we used the simulation setup above but we now generated correlated multivariate Gaussian responses with correlation $r_{jk} = 0.5,\ (j \neq k)$. Thus, we obtain a regression model with correlated errors. We generated 10% of vertical outliers and 10% of bad leverage points as before.

The results for the LR-weighted MCD regression ($\gamma = 0.25$) in Table 4 are comparable to the results in Table 3, as expected from the equivariance of the estimator. Table 4 shows that the LR-weighted MCD regression in general outperforms the coordinatewise LTS regressions both in bias and MSE. The differences are largest for the slope estimates. Note that Table 4 does not contain results for the off-diagonal elements of the error covariance matrix because these elements are not estimated in the univariate LTS approach. Hence, another advantage of the multivariate MCD regression method is that it gives a robust estimate of the full error covariance matrix.

	n=50		n=100		n=500	
	bias	MSE	bias	MSE	bias	MSE
LR-weighted MCD regression ($\gamma = 0.25$)						
slope	0.0044	1.644	0.0043	1.483	0.0017	1.319
intercept	0.0069	1.512	0.0017	1.364	0.0021	1.313
Σ_{diag}	0.1316	3.357	0.0676	3.200	0.0097	2.900
Combination of univariate LTS regressions						
slope	0.2360	4.630	0.2403	7.270	0.2406	30.214
intercept	0.0264	2.144	0.0342	1.880	0.0331	2.272
Σ_{diag}	0.0976	4.166	0.0868	6.158	0.1004	9.044

Table 4: Bias and MSE of the slope matrix, intercept vector and error variances obtained by the LR-weighted MCD regression with $\gamma = 0.25$ and univariate LTS regressions. The data contained 20% of outliers. The number of replications was $m = 1000$.

Based on the performance results in the previous section and the robustness results here, we recommend the LR-weighted method with $\gamma = 0.25$ in practice to identify all outliers and robustly estimate the full error covariance matrix.

5 Examples

Example 1 (continued). We now continue the analysis of example 1 in Section 1 by applying the LR weighted robust multivariate regression method with $\gamma = 0.25$ to these data. Figure 3 shows the diagnostic plot corresponding to the robust analysis. This plot is a generalization of the diagnostic plot for multiple regression due to Rousseeuw and van Zomeren (1990). In this display the robust distances of the q-dimensional residuals $d(r_i^{LR}) = ((r_i^{LR})^t(\hat{\Sigma}_\varepsilon^{LR})^{-1} r_i^{LR})^{1/2}$ are plotted versus the robust distances of the p-dimensional x_i given by $d(x_i) = ((x_i -$

$(t_n^1)_x)^t((C_n^1)_{xx})^{-1}(x_i - (t_n^1)_x))^{1/2}$. The plot enables us to classify the data points into regular observations, vertical outliers, good leverage points, and bad leverage points. Moreover, one can see whether a point is an extreme outlier or merely a borderline case. Being a graphical tool, this plot also allows us to discover unexpected structure in the data. Note that all the estimates on which the plot is based are byproducts of the robust multivariate regression algorithm, so the plot requires very little computation time.

From Figure 3 we see that thirteen observations have residuals with robust distance above the horizontal cutoff line at $\sqrt{\chi^2_{4,0.975}} = 3.34$ and thus are detected as regression outliers. Eight of these points also have a large x-distance and therefore are bad leverage points. Note that classical multivariate regression only detected three of these outliers (51, 52, and 56) and considered four of the outliers (46, 58, 60 and 61) to be good leverage points. Moreover, by applying LTS for each of the responses separately we only detected one additional outlier (61) but nine other outliers remained hidden, among which the bad leverage points 59, 60 and 62 are the most severe.

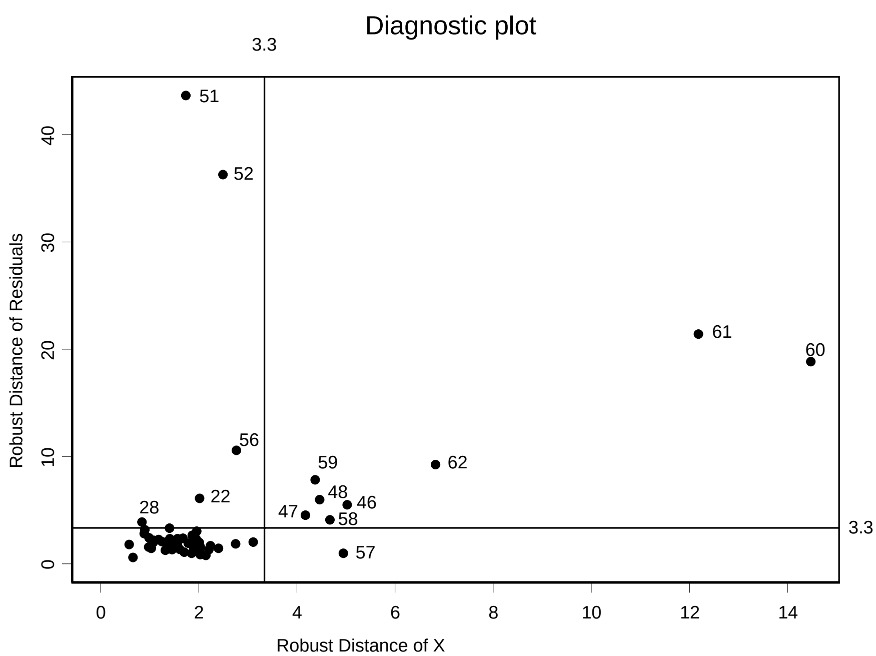

Figure 3: Plot of robust distances of residuals versus robust distances of the carriers for the pulp fibre data.

By exploring the origin of the collected data we found out that all but the last four pulp samples (observations 59-62) were produced from fir wood. Moreover, most of the outlying samples were obtained using different pulping processes. For example, observation 62 is the only sample from a chemi-thermomechanical pulping process, observations 60 and 61 are the only samples from a solvent pulping process, and observations 51, 52 and 56 are obtained from a kraft pulping process. Finally, the smaller outliers (22, 46-48 and 58) all were Douglas fir samples.

Example 2. This example describes an actual dataset obtained from Shell's polymer laboratory in Ottignies, Belgium by courtesy of dr. Christian Ritter. For reasons of confidentiality, all variables have been standardized and their exact meanings are not given. The dataset consists of $n = 217$ observations with $p = 4$ predictor variables and $q = 3$ response variables. The predictor variables describe the chemical characteristics of a piece of foam, whereas the response variables measure its physical properties such as tensile strength. Foam product specifications are expressed in terms of its physical properties. Production units around the world have to meet the prescribed physical requirements. The physical properties are determined by the chemical composition used in the production process. However, different chemical compositions will lead to foams meeting all required specifications. Moreover, depending on the location of the production unit there is a strong variation in the price of the necessary chemicals. Therefore, the goal is to establish a relationship between the chemical inputs and the resulting physical properties which can then be used to determine the cheapest chemical composition resulting in foams meeting all physical requirements. We use multivariate regression to determine the relationship between the chemical inputs and the physical properties. A few cases with missing values had been omitted in advance. After an initial exploratory study of the seven variables, including their Q-Q plots, we have applied Box-Cox transformations to them. We then ran a robust multivariate regression using the LR-weighted method with $\gamma = 0.25$. This computation took only 43 seconds on a Sun SparcStation 20/514.

Figure 4 shows the diagnostic plot of the Shell foam data (robust distances of the residuals r_i^{LR} versus the robust distances of the x_i). Observations 215 and 110 lie far from both the horizontal cutoff line at $\sqrt{\chi^2_{3,0.975}} = 3.06$ and the vertical cutoff line at $\sqrt{\chi^2_{4,0.975}} = 3.34$. These two observations can thus be classified as bad leverage points. Several observations lie substantially above the horizontal cutoff but not to the right of the vertical cutoff, which means that they are vertical outliers (their residuals are outlying but their x-values are not).

When this list of special points was presented to the scientists who had made the measurements, we learned that 8 observations in Figure 4 were produced with a different production

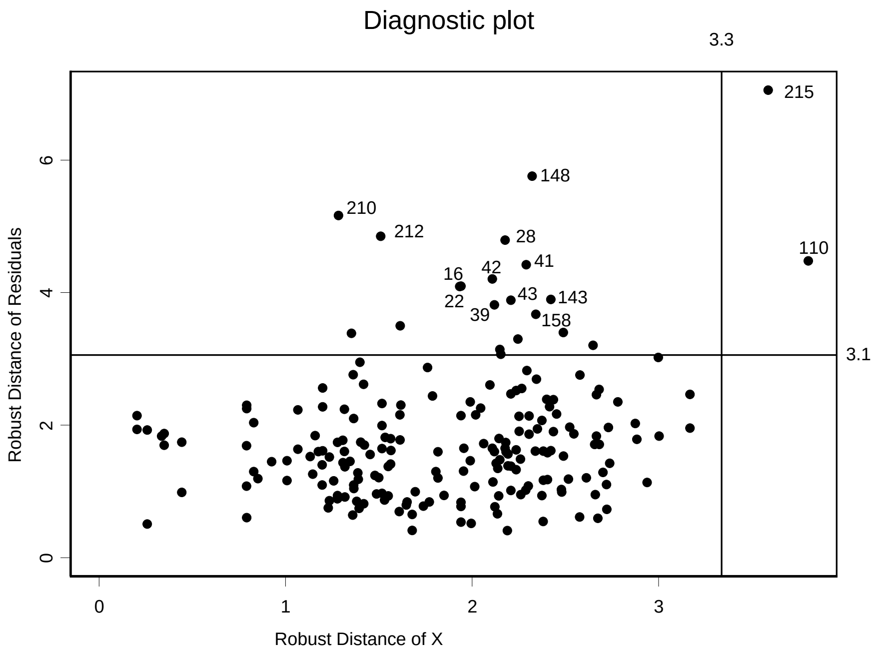

Figure 4: Diagnostic plot of robust residuals versus robust distances of the carriers for the foam data.

technique and hence belong to a different population with other characteristics. These include the observations 210, 212 and 215. We therefore remove these 8 observations from the data, and retain only observations from the intended population.

Running the method again yields the diagnostic plot in Figure 5. Observation 110 is still a bad leverage point, and also several of the vertical outliers remained. No chemical/physical mechanism was found to explain why these points are outliers, leaving open the possibility of some large measurement errors. But the detection of these substantial outliers at least provides us with the option to choose whether or not to allow them to affect the final result.

6 Equivariance and robustness properties

The theorems in this section show that the proposed LR weighted method based on MCD (1) has the natural equivariance properties of multivariate regression estimators, and (2) is robust. These generalize the regression, scale, and affine equivariance (see Rousseeuw and Leroy 1987, page 116) and robustness of multiple regression estimators. All proofs are given in the Appendix.

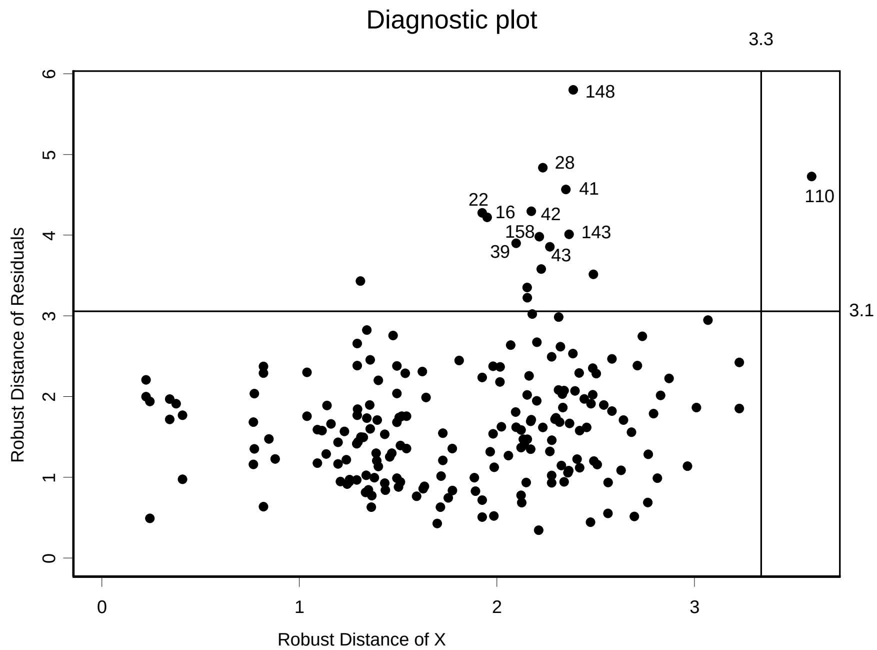

Figure 5: Diagnostic plot of robust residuals versus robust distances for the corrected foam data.

Denote $T_n(X,Y) = (\hat{\mathcal{B}}^t, \hat{\alpha})^t$ where the matrix X is $(n \times p)$ and Y is $(n \times q)$. The estimator T_n is called *regression equivariant* if

$$T_n(X, Y + XD + 1_n w^t) = T_n(X,Y) + (D^t, w)^t \tag{9}$$

where D is any $(p \times q)$ matrix, w is any $(q \times 1)$ vector, and $1_n = (1,1,\ldots,1)^t \in I\!R^n$. Regression equivariance means that if we add a linear function of the explanatory variables to the responses, then the coefficients of this linear function are also added to the estimator.

The estimator T_n is said to be *y-affine equivariant* if

$$T_n(X, YC + 1_n d^t) = T_n(X,Y)\,C + (O_{pq}^t, d)^t \tag{10}$$

where C is any nonsingular $(q \times q)$ matrix, d is any $(q \times 1)$ vector and O_{pq} is the $(p \times q)$ matrix consisting of zeroes. If the response variables are transformed linearly then y-affine equivariance implies that the estimator T transforms accordingly.

We say that the estimator T_n is *x-affine equivariant* if

$$T_n(XA^t + 1_n v^t, Y) = (\hat{\mathcal{B}}^t A^{-1}, \hat{\alpha} - \hat{\mathcal{B}}^t A^{-1} v)^t \tag{11}$$

for any nonsingular $(p \times p)$ matrix A and any column vector $v \in I\!R^{p\times 1}$. If the explanatory variables are transformed linearly then x-affine equivariance says that the estimator T_n transforms correctly.

Theorem 1. *The LR weighted multivariate MCD-regression estimator* $T_n = ((\hat{\mathcal{B}}^{LR})^t, \hat{\alpha}^{LR})^t$ *is regression, y-affine, and x-affine equivariant.*

We also study the theoretical robustness properties of the estimator in terms of its breakdown value and its influence function which also yields its asymptotic variance. These theoretical properties will confirm the finite-sample results obtained in Sections 3.3 and 4.

The finite-sample breakdown value (Donoho and Huber 1983) of a regression estimator T_n at a dataset $Z_n = (X, Y) \in I\!R^{n\times(p+q)}$ is defined as the smallest fraction of observations of Z_n that need to be replaced to carry T_n beyond all bounds. Formally,

$$\varepsilon_n^*(T_n, Z_n) = \min_{1\leq m\leq n}\{\frac{m}{n} : \sup_{Z_n'}\|T_n(Z_n) - T_n(Z_n')\| = \infty\}, \tag{12}$$

where the supremum is over all possible collections Z_n' that differ from Z_n in at most m points. The following theorem shows that the LR weighted MCD regression estimator $T_n = ((\hat{\mathcal{B}}_n^{LR})^t, \hat{\alpha}_n^{LR})^t$ inherits the breakdown value of the initial MCD location and scatter estimators applied to the $(p + q)$-dimensional dataset Z_n. Note that the breakdown value of a covariance estimator is the smallest fraction of outliers that can make the largest eigenvalue arbitrarily large or the smallest eigenvalue arbitrarily small.

Theorem 2. *Let Z_n be a set of $n \geq p + q + 1$ observations and t_n^1, C_n^1 the reweighted MCD estimators of location and scatter with* $\min\{\varepsilon_n^*(t_n^1, Z_n), \varepsilon_n^*(C_n^1, Z_n)\} = \lceil n\gamma \rceil/n$ *where* $\gamma = (n-h)/n \leq (n-(p+q))/(2n)$*. Then the multivariate regression estimator* $T_n = ((\hat{\mathcal{B}}_n^{LR})^t, \hat{\alpha}_n^{LR})^t$ *also satisfies* $\varepsilon_n^*(T_n, Z_n) = \lceil n\gamma \rceil/n$.

The influence function of an estimator T at a distribution H measures the effect on T of an infinitesimal contamination at a single point (Hampel et al. 1986). If we denote the point mass at $z = (x^t, y^t)^t$ by Δ_z and write $H_\varepsilon = (1-\varepsilon)H + \varepsilon\Delta_z$ then the influence function is given by

$$IF(z, T, H) = \lim_{\varepsilon\downarrow 0}\frac{T(H_\varepsilon) - T(H)}{\varepsilon} = \frac{\partial}{\partial\varepsilon}T(H_\varepsilon)|_{\varepsilon=0}. \tag{13}$$

The following theorem gives the influence functions of the LR weighted MCD regression estimators at the standard Gaussian distribution. The influence function at general Gaussian distributions then follows from the equivariance properties in Section 6.

Theorem 3. *The influence functions of $\hat{\mathcal{B}}^{LR}$, $\hat{\alpha}^{LR}$ and $\hat{\Sigma}_\varepsilon^{LR}$ at the standard Gaussian distribution $H = N(0, I_{p+q})$ are given by*

$$\begin{aligned}
IF(z, \hat{\mathcal{B}}^{LR}_{jk}, H) &= [c_1\, I(\|z\|^2 \leq q_\gamma) + c_2\, I(\|z\|^2 \leq q_{\delta_l}) + c_3\, I(\|y\|^2 \leq q_{\delta_r})]x_j y_k \\
IF(z, \hat{\alpha}^{LR}_{j}, H) &= [c_4\, I(\|z\|^2 \leq q_\gamma) + c_5\, I(\|z\|^2 \leq q_{\delta_l}) + c_3\, I(\|y\|^2 \leq q_{\delta_r})]y_j \\
IF(z, (\hat{\Sigma}^{LR}_{\varepsilon})_{jk}, H) &= [c_6\, I(\|z\|^2 \leq q_\gamma) + c_7\, I(\|z\|^2 \leq q_{\delta_l}) + c_3\, I(\|y\|^2 \leq q_{\delta_r})]y_j y_k \\
IF(z, (\hat{\Sigma}^{LR}_{\varepsilon})_{jj}, H) &= [c_8\, I(\|z\|^2 \leq q_\gamma) + c_9\, I(\|z\|^2 \leq q_{\delta_l}) + c_3\, I(\|y\|^2 \leq q_{\delta_r})]y_j^2 + \\
&\quad g_1(\|z\|, \|y\|) I(\|z\|^2 \leq q_\gamma) + g_2(\|y\|) I(\|z\|^2 \leq q_{\delta_l}) + c_{10} I(\|y\|^2 \leq q_{\delta_r}) + c_{11}
\end{aligned}$$

Here $q_\gamma = \chi^2_{p+q,1-\gamma}$. The constants $c_1, \ldots, c_{11}$ and the functions g_1 and g_2 can be found in the appendix. Note that the influence functions of the slope and intercept become zero as soon as $\|y\|$ becomes large, so vertical outliers as well as bad leverage points have no effect on the regression estimates. For the covariance of the errors, the influence on the off-diagonal elements becomes zero, and the influence on the diagonal elements becomes constant for observations with large $\|y\|$ so the effect of outliers and leverage points is bounded. On the other hand, good leverage points (which have large $\|x\|$ but small $\|y\|$ and thus are not outliers for the regression model) are not downweighted.

Figure 6 shows the influence functions of the LR weighted MCD regression estimators with $\gamma = 0.25$ at the bivariate Gaussian distribution $H = N_2(0, I)$ $(p = q = 1)$. The influence functions of the slope $\hat{\beta}^{LR} = \hat{\mathcal{B}}^{LR}$ and the intercept $\hat{\alpha}^{LR}$ are shown in Figures 6a and 6b. The influence function of the error scale $(\hat{\sigma}^{LR})^2 = \hat{\Sigma}^{LR}_\varepsilon$ is shown in Figure 6c.

From the influence function we can compute the asymptotic variance of the elements of the slope matrix $\hat{\mathcal{B}}^{LR}$ at the standard Gaussian distribution as

$$ASV(\hat{\mathcal{B}}^{LR}_{ij}, H) = E_H[IF(z, \hat{\mathcal{B}}^{LR}_{jk}, H)^2] \tag{14}$$

(see Hampel et al. 1986), and similarly for $\hat{\alpha}$ and $\hat{\Sigma}_\varepsilon$. It can easily be shown that the asymptotic variances of the slope and intercept elements of the least squares estimator equal 1. For the least squares estimator of the error covariance it holds that the asymptotic variance equals 1 for the off-diagonal elements and equals 2 for the diagonal elements. Therefore, the asymptotic relative efficiency (ARE) of the slope $\hat{\mathcal{B}}^{LR}$ relative to the least squares slope $\hat{\mathcal{B}}_{LS}$ is given by

$$\text{ARE}(\hat{\mathcal{B}}^{LR}, H) = 1/ASV(\hat{\mathcal{B}}^{LR}_{ij}, H) \tag{15}$$

and similarly for the intercept $\hat{\alpha}^{LR}$ and off-diagonal elements of $\hat{\Sigma}^{LR}_\varepsilon$. The ARE of the

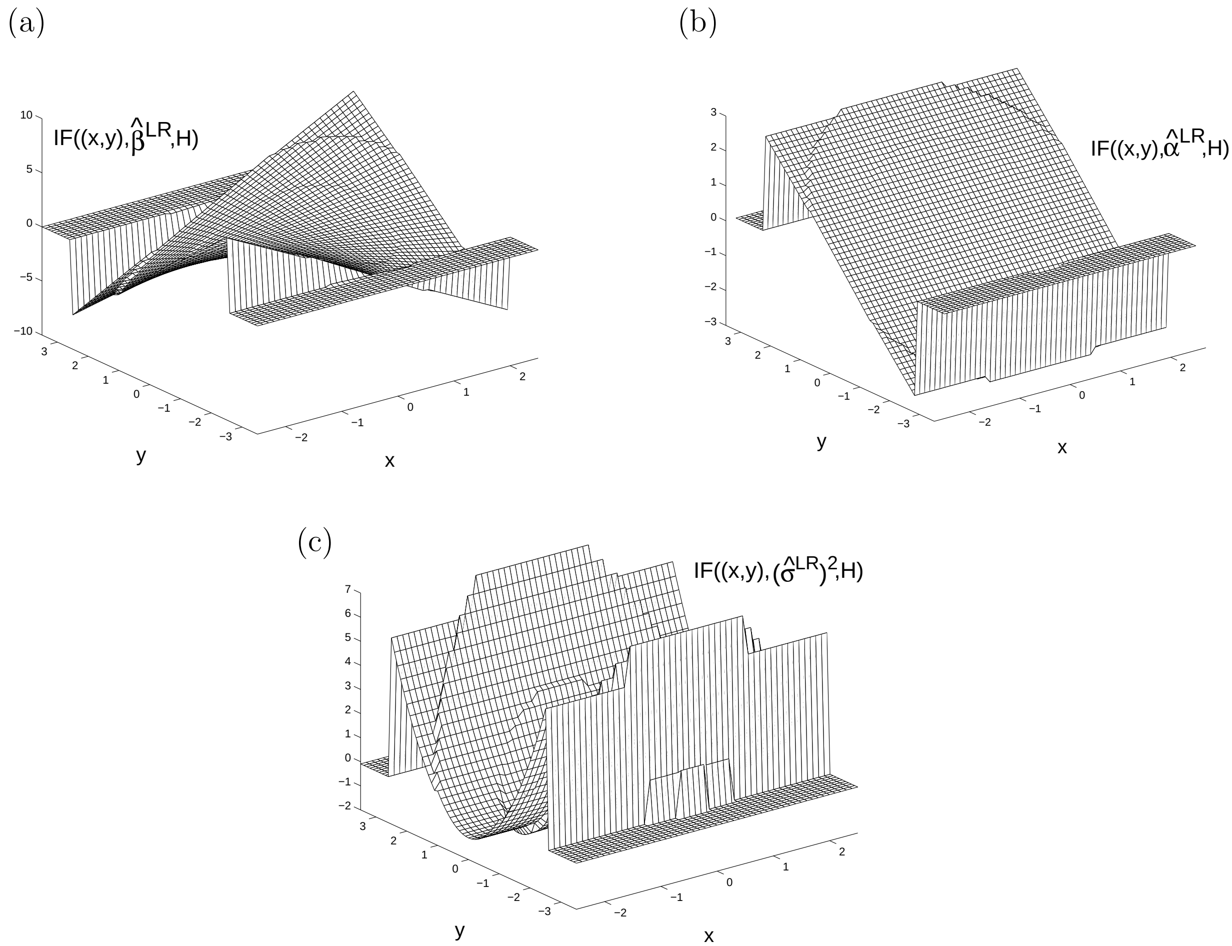

Figure 6: Influence functions at the bivariate Gaussian distribution of (a) slope, (b) intercept, and (c) error scale of LR weighted MCD regression.

diagonal elements of $\hat{\Sigma}_{\varepsilon}^{LR}$ equals

$$\text{ARE}((\hat{\Sigma}_{\varepsilon})^{LR})_{jj}, H) = 2/ASV((\hat{\Sigma}_{\varepsilon}^{LR})_{jj}). \quad (16)$$

For $p = 4$, $q = 4$, the ARE of slope, intercept and diagonal and off-diagonal elements of the error covariance are given in Table 2 under $n = \infty$. For the initial MCD regression and the L and R weighted methods the efficiencies can be obtained from additional results in the Appendix. It is reassuring to note that the finite-sample efficiencies correspond quite well to the asymptotic efficiencies. The difference is often negligible already for $n = 500$.

To obtain outlier diagnostics that take into account the residual error and the location of the observation in x-space, we now introduce studentized robust residual distances. These studentized residual distances generalize the studentized residuals for univariate robust regression (McKean et al. 1990, 1993) to multivariate regression. They also extend the studentized residual distances for multivariate least squares regression (Caroni 1987) to robust multivariate regression.

Consider the asymptotic representation of the estimator given by the influence function:

$$T_n = \theta + \frac{1}{n}\sum_{j=1}^{n} IF(z_j, T_n, G) + o(n^{-1/2})$$

where $\theta = (\mathcal{B}^t, \alpha)^t$ and G is the joint distribution of $z = (x^t, y^t)^t$. We then obtain the following first order approximation for the residuals

$$r_i \doteq \varepsilon_i - \frac{1}{n}\sum_{j=1}^{n}[IF(z_j, \hat{\mathcal{B}}^{LR}, G)^t x_i + IF(z_j, \hat{\alpha}^{LR}, G)] \tag{17}$$

from which the covariance matrix $\text{cov}(r_i)$ can be derived as outlined in the appendix. Studentized residual distances are now defined as

$$sd_i = \sqrt{r_i^t(\widehat{\text{cov}(r_i)})^{-1} r_i}.$$

Here, $\widehat{\text{cov}(r_i)}$ is the estimated covariance matrix for residual r_i obtained by replacing the unknown error covariance matrix Σ_ε with an estimate $\hat{\Sigma}_\varepsilon$. If the estimate $\hat{\Sigma}_\varepsilon$ is derived from the fitted model based on all data points, then we obtain *internally* studentized residual distances. If $\hat{\Sigma}_\varepsilon$ comes from the model using all data points except z_i when computing sd_i, then we obtain *externally* studentized residual distances. For large outlying points, there will be little difference between internally and externally studentized residual distances because large outliers have only small influence on the LR-weighted MCD regression estimates, but for intermediate points externally studentized residuals will be larger than internally studentized residuals. To identify outliers we compare the squared studentized residuals with quantiles of the χ_q^2 distribution. Figure 7 shows the externally studentized residuals for the pulp fibre and foam datasets analyzed before. The horizontal line in both plots is the square root of the 97.5% quantile of the corresponding Chisquare distribution. The labeled points in Figure 7 even lie above the 99.5% quantile of the Chisquare distribution. These outliers have also been labeled in the diagnostic plots (Figures 3 and 4) in Section 5.

7 Conclusions

Least squares multivariate regression is sensitive to outliers in the dataset. Therefore, alternative methods that can detect and resist outliers are needed so that reliable results can be

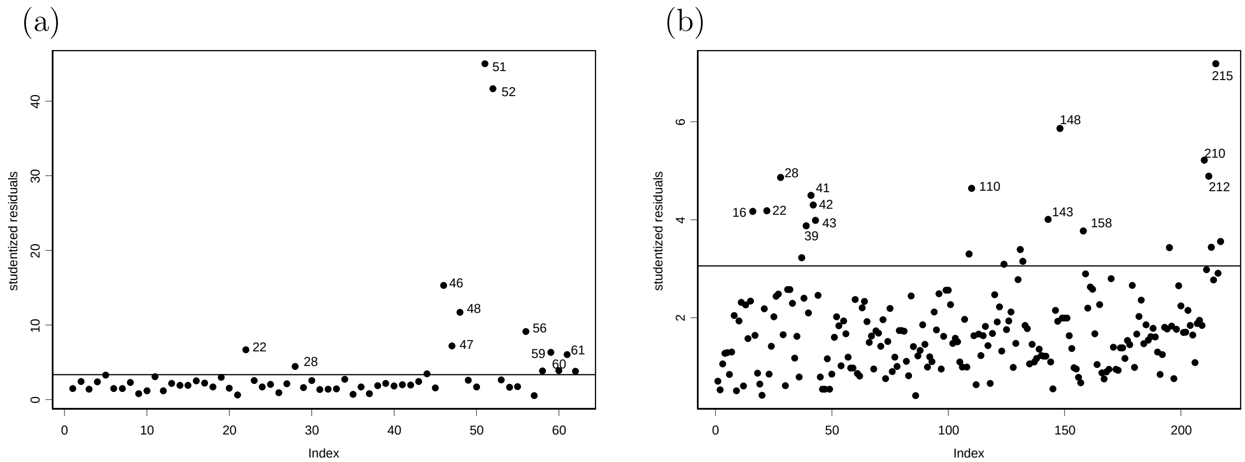

Figure 7: Externally studentized robust residuals for (a) the pulp fibre data and (b) the foam data. The horizontal line is the 97.5% quantile of the χ^2_q distribution.

obtained also in the presence of outliers. Substantial work has been done to develop influence measures for multivariate regression (Hossain and Naik 1989, Barrett and Ling 1992, Hadi et al. 1995, Kim 1995, Seaver et al. 1998). Much less has been done so far to develop robust estimators with bounded influence and/or high breakdown value. Singer and Sen (1985) and Koenker and Portnoy (1990) proposed robust methods based on M-estimators. Methods based on affine equivariant sign and ranks have been recently proposed by Ollila et al. (2002, 2003). However these methods still have zero breakdown value.

We have shown that substituting robust estimates of location and scatter in the classical expressions for the slope, intercept and error scale yields a robust multivariate regression method. By inserting the MCD estimator of location and scatter we obtain a positive-breakdown and bounded-influence method, albeit with a rather low efficiency. To improve the efficiency we have studied several types of reweighting schemes. We found that the best result is obtained by using the MCD-based robust distances to form a reweighted estimator of location and scatter, which then yields the initial regression. The robust residuals from this initial regression then give us the weights for the final regression. We call this the LR-weighted MCD regression. This approach gave the best finite-sample performance in our simulations and also yielded the highest asymptotic efficiency. Moreover, simulations with contaminated datasets indicated that its robustness properties also hold at finite samples. These simulations also showed that the LR-weighted MCD regression clearly outperforms classical least squares regression as well as univariate LTS regression applied to each of the responses separately. The proposed method was illustrated on two real data applications, where a new diagnostic plot turned out to be a very useful graphical tool to detect special points. Formal outlier diagnostics have been constructed based on studentized robust residual distances. MCD regression also is an essential part of robust principal component regression (Hubert and Verboven 2003) and robust partial least squares regression (Hubert and Vanden Branden 2003) procedures that are used to analyze high-dimensional data from spectra with several responses.

Appendix

To prove Theorem 1 we first show the following lemma.

Lemma 1. *From the affine equivariance of the reweighted MCD location and scatter estimators* (t_n^1, C_n^1) *it follows that the L weighted MCD regression estimator* $T_n^L = ((\hat{\mathcal{B}}^L)^t, \hat{\alpha}^L)^t$ *is regression, y-affine, and x-affine equivariant.*

Proof of Lemma 1: Affine equivariance of (t_n^1, C_n^1) means that for any nonsingular $(p+q)\times(p+q)$ matrix M and any vector $a \in I\!R^{p+q}$ it holds that $t_n^1(ZM^t + 1_n a^t) = Mt_n^1(Z) + a$ and $C_n^1(ZM^t + 1_n a^t) = MC_n^1 M^t$. To prove regression equivariance we take

$$M = \begin{pmatrix} I_p & 0 \\ D^t & I_q \end{pmatrix} \quad \text{and} \quad a = \begin{pmatrix} 0 \\ w \end{pmatrix}.$$

Then $ZM^t + 1_n a^t = (X, Y)M^t + 1_n a^t = (X, Y + XD + 1_n w^t)$ and

$$\begin{aligned}
(t_n^1)_x(ZM^t + 1_n a^t) &= (t_n^1)_x(Z) \\
(t_n^1)_y(ZM^t + 1_n a^t) &= (t_n^1)_y(Z) + D^t(t_n^1)_x(Z) + w \\
(C_n^1)_{xx}(ZM^t + 1_n a^t) &= (C_n^1)_{xx}(Z) \\
(C_n^1)_{xy}(ZM^t + 1_n a^t) &= (C_n^1)_{xx}(Z)D + (C_n^1)_{xy}(Z).
\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
\hat{\mathcal{B}}^L(ZM^t + 1_n a^t) &= (C_n^1)_{xx}^{-1}(ZM^t + 1_n a^t)(C_n^1)_{xy}(ZM^t + 1_n a^t) \\
&= D + (C_n^1)_{xx}^{-1}(Z)(C_n^1)_{xy}(Z) \\
&= D + \hat{\mathcal{B}}^L(Z)
\end{aligned}$$

and

$$
\begin{aligned}
\hat{\alpha}^L(ZM^t + 1_n a^t) &= (t_n^1)_y(ZM^t + 1_n a^t) - \hat{\mathcal{B}}^L(ZM^t + 1_n a^t)(t_n^1)_x(ZM^t + 1_n a^t) \\
&= (t_n^1)_y(Z) + D^t(t_n^1)_x(Z) + w - (D + \hat{\mathcal{B}}^L(Z)^t(t_n^1)_x(Z) \\
&= (t_n^1)_y(Z) - \hat{\mathcal{B}}^L(Z)^t(t_n^1)_x(Z) + w \\
&= \hat{\alpha}^L(Z) + w
\end{aligned}
$$

which is the desired result. To prove y- affine equivariance, we put

$$
M = \begin{pmatrix} I_p & 0 \\ 0 & C^t \end{pmatrix} \quad \text{and} \quad a = \begin{pmatrix} 0 \\ d \end{pmatrix}.
$$

Finally, to prove x-affine equivariance we put

$$
M = \begin{pmatrix} A & 0 \\ 0 & I_q \end{pmatrix} \quad \text{and} \quad a = \begin{pmatrix} v \\ 0 \end{pmatrix}.
$$

Proof of Theorem 1. From Lemma 1 we already have that the equivariance properties hold for the regression estimator based on reweighted MCD. It immediately follows that the reweighted regression estimator is also regression, x-affine, and y- affine equivariant because the weights $d(r_i^L)$ are invariant under these transformations which can be proved similarly as in Lemma 1.

Lemma 2. *Let Z_n be a set of $n \geq p+q+1$ observations and t_n^1, C_n^1 the reweighted MCD estimators of location and scatter with $\min\{\varepsilon_n^*(t_n^1, Z_n), \varepsilon_n^*(C_n^1, Z_n)\} = \lceil n\gamma \rceil / n$ where $\gamma = (n-h)/n \leq (n-(p+q))/(2n)$. Then the estimator $T_n^L = ((\hat{\mathcal{B}}_n^L)^t, \hat{\alpha}_n^L)^t$ also satisfies $\varepsilon_n^*(T_n^L, Z_n) = \lceil nr \rceil / n$.*

Proof of Lemma 2: Since the estimator T_n^L is regression, y-affine, and x-affine equivariant (Lemma 1), we may assume without loss of generality that $t_n^1(Z_n) = 0$. Let $\tilde{Z}_n$ be a dataset obtained by replacing $m < \lceil n\gamma \rceil / n$ points from the original dataset Z_n by arbitrary values. We first show that the slope $\hat{\mathcal{B}}^L(\tilde{Z}_n)$ remains bounded. Denote the eigenvalues of $(C_n^1)_{xx}(\tilde{Z}_n)$ by $\lambda_1((C_n^1)_{xx}(\tilde{Z}_n)) \leq \cdots \leq \lambda_p((C_n^1)_{xx}(\tilde{Z}_n))$. Note that $\|\hat{\mathcal{B}}^L(\tilde{Z}_n)\| = \|(C_n^1)_{xx}^{-1}(\tilde{Z}_n)(C_n^1)_{xy}(\tilde{Z}_n)\| \leq \|(C_n^1)_{xx}^{-1}(\tilde{Z}_n)\| \, \|(C_n^1)_{xy}(\tilde{Z}_n)\|$. Now we have that

$$
\|(C_n^1)_{xx}^{-1}(\tilde{Z}_n)\| = \sup_{\|x\| \neq 0} \frac{\|(C_n^1)_{xx}^{-1}(\tilde{Z}_n)x\|}{\|x\|} = \left(\inf_{\|x\| \neq 0} \frac{\|(C_n^1)_{xx}(\tilde{Z}_n)x\|}{\|x\|} \right)^{-1} = \frac{1}{\lambda_1((C_n^1)_{xx}(\tilde{Z}_n))}
$$

which is bounded because the covariance matrix $C_n^1(\tilde{Z}_n)$ does not break down for $m < \lceil n\gamma \rceil / n$. Denote $\lambda_1(\tilde{Z}_n) \leq \cdots \leq \lambda_{p+q}(\tilde{Z}_n)$ the eigenvalues of $C_n^1(\tilde{Z}_n)$, then we have that $\|(C_n^1)_{xy}(\tilde{Z}_n)\| \leq \|C_n^1(\tilde{Z}_n)\| \leq \lambda_{p+q}(\tilde{Z}_n)$ which is also bounded for $m < \lceil n\gamma \rceil / n$. For the intercept it clearly holds that $\|\hat{\alpha}^L(\tilde{Z}_n)\| = \|(t_n)_y(\tilde{Z}_n) - (\hat{\mathcal{B}}^L)^t(\tilde{Z}_n)(t_n)_x(\tilde{Z}_n)\| \leq \|(t_n)_y(\tilde{Z}_n)\| + \|(\hat{\mathcal{B}}^L)^t(\tilde{Z}_n)\| \|(t_n)_x(\tilde{Z}_n)\|$ is bounded for $m < \lceil n\gamma \rceil / n$ since $\|(\hat{\mathcal{B}})^L(\tilde{Z}_n)\|$ and $\|t_n(\tilde{Z}_n)\|$ are bounded.

Proof of Theorem 2. Lemma 2 shows that the L weighted MCD regression estimator T_n^L inherits the breakdown value of the reweighted MCD estimators. It now easily follows that (under certain regularity conditions of the design matrix) the reweighted regression estimator T_n^{LR} inherits the breakdown value of the initial regression estimator T_n^L.

Lemma 3. *Denote t, C the functionals corresponding to the reweighted MCD location and scatter estimators, then the influence functions of $\hat{\mathcal{B}}^L$, $\hat{\alpha}^L$, and $\hat{\Sigma}_\varepsilon^L$ satisfy*

$$IF(z, \hat{\mathcal{B}}^L, H) = IF(z, C_{xy}^1, H) \tag{18}$$

$$IF(z, \hat{\alpha}^L, H) = IF(z, t_y^1, H). \tag{19}$$

$$IF(z, \hat{\Sigma}_\varepsilon^L, H) = IF(z, C_{yy}^1, H)). \tag{20}$$

Proof of Lemma 3: First we derive the influence function of the slope $\hat{\mathcal{B}}^L$. Since $\hat{\mathcal{B}}^L(H_\varepsilon) = (C_{xx}^1)^{-1}(H_\varepsilon)C_{xy}^1(H_\varepsilon)$ we obtain that

$$\begin{aligned} IF(z, \hat{\mathcal{B}}^L, H) &= \frac{\partial}{\partial\varepsilon}((C_{xx}^1)^{-1}(H_\varepsilon)C_{xy}^1(H_\varepsilon))|_{\varepsilon=0} \\ &= IF(z, (C_{xx}^1)^{-1}, H)C_{xy}^1(H) + (C_{xx}^1)^{-1}(H)IF(z, C_{xy}^1, H) \\ &= IF(z, C_{xy}^1, H) \end{aligned}$$

since consistency of C^1 yields $C^1(H) = I_{p+q}$. Similarly, with $\hat{\alpha}^L(H_\varepsilon) = t_y^1(H_\varepsilon) - (\hat{\mathcal{B}}^L)^t(H_\varepsilon)t_x^1(H_\varepsilon)$ we have that

$$\begin{aligned} IF(z, \hat{\alpha}^L, H) &= \frac{\partial}{\partial\varepsilon}(t_y^1(H_\varepsilon) - (\hat{\mathcal{B}}^L)^t(H_\varepsilon)t_x^1(H_\varepsilon))|_{\varepsilon=0} \\ &= IF(z, t_y^1, H) - IF(z, \hat{\mathcal{B}}^L, H)^t t_x^1(H) - (\hat{\mathcal{B}}^L)^t(H)IF(z, t_x^1, H) \\ &= IF(z, t_y^1, H) \end{aligned}$$

since $t^1(H) = 0$ and $\hat{\mathcal{B}}^L(H) = 0$. Finally, $\hat{\Sigma}_\varepsilon^L(H_\varepsilon) = C_{yy}^1(H_\varepsilon) - (\hat{\mathcal{B}}^L)^t(H_\varepsilon)C_{xx}^1(H_\varepsilon)(\hat{\mathcal{B}}^L)(H_\varepsilon)$

yields

$$\begin{aligned} IF(z,\hat{\Sigma}_{\varepsilon}^{L},H) &= \frac{\partial}{\partial\varepsilon}(C_{yy}^{1}(H_{\varepsilon})-(\hat{\mathcal{B}}^{L})^{t}(H_{\varepsilon})C_{xx}^{1}(H_{\varepsilon})\hat{\mathcal{B}}^{L}(H_{\varepsilon}))|_{\varepsilon=0} \\ &= IF(z,C_{yy}^{1},H)-IF(z,\hat{\mathcal{B}}^{L},H)^{t}C_{xx}^{1}(H)\hat{\mathcal{B}}^{L}(H)-(\hat{\mathcal{B}}^{L})^{t}(H)IF(z,C_{xx}^{1},H)\hat{\mathcal{B}}^{L}(H) \\ &\quad -(\hat{\mathcal{B}}^{L})^{t}(H)C_{xx}^{1}(H)IF(z,\hat{\mathcal{B}}^{L},H) \\ &= IF(z,C_{yy}^{1},H) \end{aligned}$$

since $\hat{\mathcal{B}}^{L}(H)=0$.

Proof of Theorem 3. Combining Lemma 3 with the results of Croux and Haesbroeck (1999) we obtain that the influence functions of $\hat{\mathcal{B}}^{L}$, $\hat{\alpha}^{L}$ and $\hat{\Sigma}_{\varepsilon}^{L}$ equal

$$\begin{aligned} IF(z,\hat{\mathcal{B}}_{jk}^{L},H) &= [\frac{a_2}{c_2}\,I(\|z\|^2\leq q_{\gamma})+\frac{1}{d_1}\,I(\|z\|^2\leq q_{\delta_l})]x_jy_k \\ IF(z,(\hat{\alpha}^{L})_j,H) &= [(1-\frac{d_1}{1-\delta_l})\frac{1}{c_1}\,I(\|z\|^2\leq q_{\gamma})+\frac{1}{1-\delta_l}I(\|z\|^2\leq q_{\delta_l})]y_j \\ IF(z,(\hat{\Sigma}_{\varepsilon}^{L})_{jk},H) &= [\frac{a_2}{c_2}\,I(\|z\|^2\leq q_{\gamma})+\frac{1}{d_1}\,I(\|z\|^2\leq q_{\delta_l})]y_jy_k \\ IF(z,(\hat{\Sigma}_{\varepsilon}^{L})_{jj},H) &= [\frac{a_2}{c_2}\,I(\|z\|^2\leq q_{\gamma})+\frac{1}{d_1}\,I(\|z\|^2\leq q_{\delta_l})]y_j^2+\frac{a_2}{2c_2}\,\|y\|^2I(\|z\|^2\leq q_{\gamma})-1+ \\ &\quad \frac{p+q+2}{2}\frac{a_2}{c_2[a_3-(p+q)a_4]}\,[a_4\|z\|^2\,I(\|z\|^2\leq q_{\gamma})+\frac{a_3}{p+q}q_{\gamma}(1-\gamma-I(\|z\|^2\leq q_{\gamma}))-1]. \end{aligned}$$

Denote the incomplete gamma function by $\Gamma(u;v)=\Gamma(u)^{-1}\int_0^v t^{u-1}e^{-t}\,dt$. Then the constants a_1,a_2,a_3,a_4 are given by $a_1=1/d_1$, $a_2=(d_1-d_2)/d_1$, $a_3=c_2/c_1$ and $a_4=\frac{1}{2}-\frac{1}{2c_1}[c_2-\frac{q_{\gamma}}{p+q}(c_1+\gamma-1)]$ where $c_1=\Gamma(\frac{p+q}{2}+1;\frac{q_{\gamma}}{2})$, $c_2=\Gamma(\frac{p+q}{2}+2;\frac{q_{\gamma}}{2})$, $d_1=\Gamma(\frac{p+q}{2}+1;\frac{q_{\delta_l}}{2})$ and $d_2=\Gamma(\frac{p+q}{2}+2;\frac{q_{\delta_l}}{2})$.

It can easily be shown that the influence functions of reweighted regression estimators defined by (7) and (8) are connected to the influence functions of the initial regression estimators $\hat{\mathcal{B}}^{L}$, $\hat{\alpha}^{L}$, and $\hat{\Sigma}_{\varepsilon}^{L}$ through

$$\begin{aligned} IF(z,\hat{\mathcal{B}}^{LR},H) &= (1-\frac{d_1^R}{1-\delta_r})IF(z,\hat{\mathcal{B}}^{L},H)+\frac{I(\|y\|^2\leq q_{\delta_r})}{1-\delta_r}xy^t \\ IF(z,\hat{\alpha}^{LR},H) &= (1-\frac{d_1^R}{1-\delta_r})IF(z,\hat{\alpha}^{L},H)+\frac{I(\|y\|^2\leq q_{\delta_r})}{1-\delta_r}y \\ IF(z,\hat{\Sigma}_{\varepsilon}^{LR},H) &= \frac{d_1^R-d_2^R}{d_1^R}(IF(z,(\hat{\Sigma}_{\varepsilon}^{L}),H)+\frac{1}{2}tr(IF(z,\hat{\Sigma}_{\varepsilon}^{L},H))I_q)+\frac{I(\|y\|^2\leq q_{\delta_r})}{d_1^R}yy^t-I_q. \end{aligned}$$

The constants d_1^R and d_2^R are given by $d_1^R=\Gamma(\frac{q}{2}+1;\frac{q_{\delta_r}}{2})$ and $d_2^R=\Gamma(\frac{q}{2}+2;\frac{q_{\delta_r}}{2})$. Note that the above results extend the expressions for the influence functions of reweighted multivariate location and scatter functionals given by (Lopuhaä 1999).

Studentized residual distances. First note that for any $(x^t, y^t)t$ such that

$$\begin{aligned} x &= \Sigma_{xx}^{1/2} u + \mu_x \\ y &= \mathcal{B}^t x + \alpha + \Sigma_\varepsilon^{1/2}\varepsilon \end{aligned}$$

with $(u^t, \varepsilon^t)^t \sim H$, i.e. the standard Gaussian distribution. Then it follows from Theorem 3 and the equivariance properties in Theorem 1 that the influence function of $\hat{\mathcal{B}}^{LR}$ and $\hat{\alpha}^{LR}$ at the joint distribution G of $(x^t, y^t)^t$ can be written as

$$\begin{aligned} IF(z, \hat{\mathcal{B}}^{LR}, G) &= \left((1 - \frac{d_1^R}{1-\delta_r})[\frac{a_2}{c_2}\, I(d^2(z) \le q_\gamma) + \frac{1}{d_1}\, I(d^2(z) \le q_{\delta_l})] + \frac{I(d^2(r) \le q_{\delta_r})}{1-\delta_r}\right) \Sigma_{xx}^{-1}(x - \mu_x) r^t \\ IF(z, \hat{\alpha}^{LR}, G) &= \left((1 - \frac{d_1^R}{1-\delta_r})[(1 - \frac{d_1}{1-\delta_l})\frac{1}{c_1}\, I(d^2(z) \le q_\gamma) + \frac{1}{1-\delta_l} I(d^2(z) \le q_{\delta_l})] + \frac{I(d^2(r) \le q_{\delta_r})}{1-\delta_r}\right) r \\ &\quad - IF(z, \hat{\mathcal{B}}^{LR}, G)^t \mu_x \end{aligned}$$

where $d^2(z)$ and $d^2(r)$ are the squared robust distances of the point z and its corresponding residual. By substituting the above expressions for the influence functions in the right hand side of (17) the following approximation for the covariance matrix of residual r_i can be obtained

$$\begin{aligned} \operatorname{cov}(r_i) &\doteq \Bigg(1 - \frac{2}{n}[f_i(z_i) + \frac{d_1^R}{1-\delta_r}(d^2(x_i) + 1)] \\ &\quad + \frac{1}{n^2}\sum_{j=1}^{n}[f_i^2(z_j)\frac{d_1^R}{(1-\delta_r)^2}(d_{ji} + 1)^2 + \frac{2d_1^R}{1-\delta_r} f_i(z_j)(d_{ji} + 1)]\Bigg) \Sigma_\varepsilon \end{aligned}$$

where $d_{ji} = (x_j - \mu_x)^t \Sigma_{xx}^{-1}(x_i - \mu_x)$ and $f_i(z_j) = (1 - \frac{d_1^R}{1-\delta_r})[\frac{a_2}{c_2}\, I(d^2(z) \le q_\gamma) + \frac{1}{d_1}\, I(d^2(z) \le q_{\delta_l})] d_{ji} + (1 - \frac{d_1^R}{1-\delta_r})[(1 - \frac{d_1}{1-\delta_l})\frac{1}{c_1}\, I(d^2(z) \le q_\gamma) + \frac{1}{1-\delta_l} I(d^2(z) \le q_{\delta_l})]$.

References

Alqallaf, F.A., Konis, K.P., Martin, R.D. and Zamar, R.H. (2002), "Scalable robust covariance and correlation estimates for data mining," *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* Edmonton, Alberta 14-23.

Barrett, B.E., and Ling, R.F. (1992), "General Classes of Influence Measures for Multivariate Regression," *Journal of the American Statistical Association*, 87, 184-191.

Bickel, P.J., and Lehmann, E.L. (1976), "Descriptive Statistics for Nonparametric Models III: Dispersion," *The Annals of Statistics*, 4, 1139–1159.

Breiman, L., and Friedman, J. H. (1997), “Predicting Multivariate Responses in Multiple Linear Regression,” *Journal of the Royal Statistical Society B,* 59, 3-54.

Butler, R.W., Davies, P.L., and Jhun, M. (1993), “Asymptotics for the Minimum Covariance Determinant Estimator,” *The Annals of Statistics*, 21, 1385–1400.

Caroni, C. (1987), “Residuals and Influence in the Multivariate Linear Model,” *The statistician,* 36, 365-370.

Cook, D. R., and Setodji, M. C. (2003), “A Model-Free Test for Reduced Rank in Multivariate Regression,” *Journal of the American Statistical Association*, 98, 340–351.

Croux, C., and Haesbroeck, G. (1999), “Influence Function and Efficiency of the Minimum Covariance Determinant Scatter Matrix Estimator,” *Journal of Multivariate Analysis*, 71, 161–190.

Davies, L. (1987), “Asymptotic Behavior of S-Estimators of Multivariate Location Parameters and Dispersion Matrices,” *The Annals of Statistics*, 15, 1269–1292.

Davis, J. B., and McKean, J. W. (1993), “Rank-Based Methods for Multivariate Linear Models,” *Journal of the American Statistical Association*, 88, 245–251.

Dollinger, M.B., and Staudte, R.G. (1991), “Influence Functions of Iteratively Reweighted Least Squares Estimators,” *Journal of the American Statistical Association*, 86, 709–716.

Donoho, D.L., and Huber, P.J. (1983), “The Notion of Breakdown Point,” in *A Festschrift for Erich Lehmann*, edited by P.J. Bickel, K.A. Doksum and J.L. Hodges. Belmont: Wadsworth, 157–184.

Gleser, L. J. (1992) “The Importance of Assessing Measurement Reliability in Multivariate Regression,” *Journal of the American Statistical Association*, 87, 696–707.

Hadi, A.S., Jones, W.D., Ling, R.F. (1995), “A Unifying Representation of some Case-Deletion Influence Measures in Univariate and Multivariate Linear Regression,” *Journal of Statistical Planning and Inference,* 46, 123-135.

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

He, X., and Wang, G. (1996), "Cross-checking Using the Minimum Volume Ellipsoid Estimator," *Statistica Sinica*, 6, 367–374.

Hossain, A., and Naik, D.N. (1989), "Detection of Influential Observations in Multivariate Regression," *Journal of Applied Statistics,* 16, 25-37.

Hubert, M., and Vanden Branden, K. (2003), "Robust Methods for Partial Least Squares Regression," *Journal of Chemometrics,* to appear.

Hubert, M., and Verboven, S. (2003), "A Robust PCR method for High-dimensional Regressors," *Journal of Chemometrics,* to appear.

Johnson, R.A., and Wichern, D.W. (1998), *Applied Multivariate Statistical Analysis*, 4th edition, Upper Saddle River: Prentice Hall.

Kent, J.T., and Tyler, D.E. (1996), "Constrained M-Estimation for Multivariate Location and Scatter," *The Annals of Statistics,* 24, 1346-1370.

Kim, M.G. (1995), "Local Influence in Multivariate Regression," *Communications in Statistics-Theory and Methods,* 24, 1271-1278.

Knorr, E.M., Ng, R.T. and Zamar, R.H. (2001), "Robust space transformation for distance based operations," *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* San Francisco.

Koenker, R., and Portnoy, S. (1990), "M Estimation of Multivariate Regressions," *Journal of the American Statistical Association*, 85, 1060–1068.

Lee, J. (1992), "Relationships between Properties of Pulp-Fibre and Paper," Phd thesis, University of Toronto.

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.

Lopuhaä, H.P. (1991), "Multivariate τ-Estimators for Location and Scatter," *Canadian Journal of Statistics,* 19, 307-321.

Lopuhaä, H.P. (1999), "Asymptotics of Reweighted Estimators of Multivariate Location and Scatter," *The Annals of Statistics*, 27, 1638–1665.

Lopuhaä, H.P. and Rousseeuw, P.J. (1991), "Breakdown points of affine equivariant estimators of multivariate location and covariance matrices," *The Annals of Statistics*, 19, 229–248.

Maronna, R.A. (1976), "Robust M-Estimates of Multivariate Location and Scatter," *The Annals of Statistics*, 4, 51–67.

Maronna, R.A., and Morgenthaler, S. (1986), "Robust Regression through Robust Covariances," *Communications in Statistics-Theory and Methods*, 15, 1347–1365.

Maronna, R. and Zamar, R.H (2002), "Robust multivariate estimates for high dimensional data sets," *Technometrics,* 44, 307-317

McKean, J.W., Sheather, S.J., Hettmansperger, T.P. (1990), "Regression Diagnostics for Rank-Based Methods," *Journal of the American Statistical Association,* 85, 1018-1028.

McKean, J.W., Sheather, S.J., Hettmansperger, T.P. (1990), "The Use and Interpretation of Residuals Based on Robust Estimation," *Journal of the American Statistical Association,* 88, 1254-1263.

Ollila, E., Hettmansperger, T.P., and Oja, H. (2002), "Estimates of regression coefficients based on sign covariance matrix," *Journal of the Royal Statistical Society B,* 64, 447-466.

Ollila, E., Oja, H., and Koivunen, V., (2003), "Estimates of regression coefficients based on rank covariance matrix," *Journal of the American Statistical Association,* 98, 90-98.

Pison, G., Van Aelst, S., and Willems, G. (2002), "Small Sample Corrections for LTS and MCD," *Metrika*, 111-123.

Rocke D.M. (1996), "Robustness Properties of S-Estimators of Multivariate Location and Shape in High Dimension," *The Annals of Statistics*, 24, 1327–1345.

Rocke, D.M., and Woodruff, D.L. (1996), "Identification of Outliers in Multivariate Data," *Journal of the American Statistical Association*, 91, 1047–1061.

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 *Mathematical Statistics and Applications, Vol. B*, edited by W. Grossmann, G. Pflug, I. Vincze and W. Wertz. Dordrecht: Reidel Publishing Company, 283–297.

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

Rousseeuw, P.J., and Van Driessen, K. (1999), "A Fast Algorithm for the Minimum Covariance Determinant Estimator," *Technometrics*, 41, 212–223.

Rousseeuw, P.J., and Van Driessen, K. (2000), "An Algorithm for Positive-Breakdown Regression Based on Concentration Steps," in *Data Analysis: Scientific Modeling and Practical Application*, edited by W. Gaul, O. Opitz and M. Schader. New York: Springer-Verlag, 335-346.

Rousseeuw, P.J., and van Zomeren, B.C. (1990), "Unmasking Multivariate Outliers and Leverage Points," *Journal of the American Statistical Association*, 85, 633–651.

Seaver, B., Blankenship, A., and Triantis, K.P. (1998), "Identifying Potentially Influential Subsets in Multivariate Regression," Technical Report, University of Tennessee

Singer, J.M., and Sen, P.K. (1985) "M-Methods in Multivariate Linear Models," *Journal of multivariate Analysis,* 17, 168-184.

Woodruff, D.L., and Rocke, D.M. (1994), "Computable Robust Estimation of Multivariate Location and Shape in High Dimension Using Compound Estimators," *Journal of the American Statistical Association*, 89, 888–896.

Zuo, Y., and Cui, H. (2002), "Depth Weighted Scatter Estimators," Technical Report, Michigan State University.

Zuo, Y., Cui, H., and He, X. (2001), "On the Stahel-Donoho Estimator and Depth-Weighted Means of Multivariate Data," *The Annals of Statistics,* to appear