Statistical Methods and Applications manuscript No.
(will be inserted by the editor)

An L1-type estimator of multivariate location and shape

Ella Roelant*, **Stefan Van Aelst**

Ghent University - UGent, Department of Applied Mathematics and Computer Science, Gent, Belgium.

e-mail: `Ella.Roelant@UGent.be, Stefan.VanAelst@UGent.be`

Received: date

Abstract In this note we study a multivariate extension of the median obtained by considering the median as the L1 location estimator. Contrary to other multivariate extensions, this multivariate estimator yields simultaneously a location estimate and shape/scatter estimate. We investigate properties of the estimator such as the influence function and asymptotic variances and compare it with other estimators of location and shape.

Key words multivariate location and shape, influence function, efficiency

Send offprint requests to:

* *Present address:* Ghent University - UGent, Department of Applied Mathematics and Computer Science, Krijgslaan 281 S9, B-9000 Gent, Belgium, Tel. +32 264 4756, Fax +32 9 264 4995.

1 Introduction

Our aim is to estimate the location and shape of a multivariate data set. Desirable properties for such estimators are affine equivariance, a bounded influence function, a positive breakdown value, high efficiency and easy computability. In this note we consider a simple extension of the median to the multivariate setting.

Several multivariate generalizations of the median have already been proposed. An overview has been given by Small (1990). The spatial median (Brown 1983) or mediancentre (Gower 1974) and the orthomedian (Grübel 1996) have a 50% breakdown value (Lopuhaä and Rousseeuw 1991) but a drawback is that these estimators are only orthogonally equivariant. Other median type multivariate location estimators are Tukey's halfspace median (Tukey 1975), the Oja median (Oja 1983), the simplicial depth median (Liu 1990), the transformation-retransformation median (Chakraborty and Chaudury 1996) and the multivariate median of Hettmansperger and Randles (2002). A straightforward way to define an affine equivariant multivariate median is by minimizing the sum of the Mahalanobis distances. Contrary to the medians cited above, this estimator comes naturally with an accompanying shape/scatter estimator whose properties are given as well.

The outline of the paper is as follows. In Section 2 we formalize the estimator. Section 3 gives the influence function of the estimator at elliptical distributions and Section 4 studies the efficiency. Section 5 compares the estimator with other estimators. Section 6 contains an example and we conclude in Section 7.

2 The estimator

Let $x_1, \ldots, x_n$ be a sample of p-variate observations. The estimator is then defined as the solution $(\hat{\mu}_1, \hat{V}_1)$ that minimizes the sum of the distances $d_i(m, V) = \sqrt{(x_i - m)^T V^{-1} (x_i - m)}$, that is

$$(\hat{\mu}_1, \hat{V}_1) = \underset{m,V,\det V=1}{\operatorname{argmin}} \sum_{i=1}^{n} d_i(m, V) \tag{1}$$

among all $(m, V) \in \mathbb{R}^p \times \mathrm{PDS}(p)$ where $\mathrm{PDS}(p)$ is the class of positive definite symmetric matrices of size p. For $p = 1$ it is clear that the estimator is the (L1) median. The estimator is an extension of the univariate median that is obtained by considering it as the L1 location estimator. Note that by constraining the determinant of V to 1, we get an estimate for the shape of the data cloud.

On the other hand let us consider the estimator $(\hat{\mu}_2, \hat{\Sigma}_2)$ which is obtained as the solution to the problem of minimizing $\det C$ subject to the constraint

$$\frac{1}{n} \sum_{i=1}^{n} d_i(m, C) = b$$

among all $(m, C) \in \mathbb{R}^p \times \mathrm{PDS}(p)$ and for some $b > 0$. If the observations come from an elliptical distribution $F_{\mu,\Sigma}$ with density $f_{\mu,\Sigma}(x) = (\det \Sigma)^{-1/2} f[(x - \mu)^T \Sigma^{-1} (x - \mu)]$ then the constant b can be chosen equal to $\mathrm{E}_{F_{0,\mathrm{I}}}[\sqrt{X^T X}]$ to obtain a consistent scatter estimator $\hat{\Sigma}_2$ of Σ. Put

$\hat{V}_2 = \hat{\Sigma}_2/(\det \hat{\Sigma}_2)^{1/p}$, the corresponding shape estimator. For discriminant analysis, this estimator was already used in Croux and Dehon (2001) who called it the S-median.

As in Lopuhaä (1989) it can be shown that both $(\hat{\mu}_1, \hat{V}_1)$ and $(\hat{\mu}_2, \hat{V}_2)$ satisfy the first order conditions:

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\hat{V}_j^{-1}(x_i - \hat{\mu}_j)}{d_i(\hat{\mu}_j, \hat{V}_j)} = 0 \tag{2}$$

$$\frac{1}{n}\sum_{i=1}^{n} d_i(\hat{\mu}_j, \hat{V}_j)\hat{V}_j - \frac{p}{n}\sum_{i=1}^{n} \frac{(x_i - \hat{\mu}_j)(x_i - \hat{\mu}_j)^T}{d_i(\hat{\mu}_j, \hat{V}_j)} = 0 \tag{3}$$

with $j = 1, 2$. Hence, since the objective function is convex, both estimators are equivalent. In the remainder we will denote the estimator by $(\hat{\mu}, \hat{V})$ which due to (2)-(3) belongs to the class of multivariate M-estimators. The estimating equations (2)-(3) can be solved easily by using Iteratively Reweighted Least Squares with weights $w_i = 1/d_i(\mu, V)$. To obtain initial weights, we calculate the empirical mean and covariance. However, the convexity of the objective function in (1) implies that the algorithm always converges to the unique global minimum, independent of the choice of the initial weights.

3 Influence function

The influence function of a functional 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 by Δ_x the point mass at $x \in \mathbb{R}^p$ and consider the contaminated distribution $H_{\epsilon,x} = (1-\epsilon)H + \epsilon\Delta_x$ then the influence function is given by

$$\mathrm{IF}(x;T,H) = \lim_{\epsilon\to 0}\frac{T(H_{\epsilon,x}) - T(H)}{\epsilon} = \frac{\partial}{\partial\epsilon}T(H_{\epsilon,x})|_{\epsilon=0}.$$

For any distribution H on $\mathbb{R}^p$, the functional corresponding to the L1-type estimator is the solution $T(H) = (\hat{\mu}(H), \hat{V}(H))$ that minimizes

$$\int [(x-m)^T V^{-1}(x-m)]^{1/2} dH(x)$$

among all $(m,V) \in \mathbb{R}^p \times \mathrm{PDS}(p)$ with $\det V = 1$. We consider the influence function at elliptical distributions $F_{\mu,\Sigma}$. Due to affine equivariance of $T(H)$ it suffices to look at spherical distributions $F_{0,\mathrm{I}}$. Similarly as in Van Aelst and Willems (2005) the influence function of $T(H)$ can be derived as

$$\mathrm{IF}(x;\hat{\mu},F_{0,\mathrm{I}}) = \frac{x/||x||}{\mathrm{E}_{F_{0,\mathrm{I}}}\left[\left(1-\frac{1}{p}\right)\frac{1}{||X||}\right]} \tag{4}$$

$$\mathrm{IF}(x;\hat{V},F_{0,\mathrm{I}}) = \frac{p(p+2)||x||}{\mathrm{E}_{F_{0,\mathrm{I}}}[(p+1)||X||]}\left(\frac{xx^T}{||x||^2} - \frac{1}{p}I\right) \tag{5}$$

with $X \sim F_{0,\mathrm{I}}$. For the corresponding scatter functional $\hat{\Sigma}(H)$ we obtain that

$$\mathrm{IF}(x;\hat{\Sigma},F_{0,\mathrm{I}}) = \frac{2(||x||-b)}{\mathrm{E}_{F_{0,\mathrm{I}}}[||X||]}I + \frac{p(p+2)||x||}{\mathrm{E}_{F_{0,\mathrm{I}}}[(p+1)||X||]}\left(\frac{xx^T}{||x||^2} - \frac{1}{p}I\right). \tag{6}$$

An outline of the derivation can be found in the Appendix.

Note that for elliptical distributions the influence function of the location functional is the same as for the Oja median (see Niinimaa and Oja, 1995). Hence, similarly as for the Oja median, the influence function coincides with that of the spatial median for the special case of spherical distributions.

4 Efficiency

General results of M-estimators guarantee consistency and asymptotic normality of the L1-type estimator (see e.g. van der Vaart (1998) Theorems 5.7 and 5.23). The asymptotic variance-covariance matrix for $\hat{\mu}$ at the model distribution $F_{0,\mathrm{I}}$ can now be obtained as

$$\mathrm{ASV}(\hat{\mu}, F_{0,\mathrm{I}}) = \mathrm{E}_{F_{0,\mathrm{I}}}[\mathrm{IF}(x; \hat{\mu}, F_{0,\mathrm{I}}) \times \mathrm{IF}(x; \hat{\mu}, F_{0,\mathrm{I}})^T]$$

(see e.g. Hampel et al., 1986). The influence functions of the L1-functional (4)-(6) can be written in the same form as the influence function of the S-functional (see Lopuhaä 1989, formulas 5.7-5.8). Consequently, we can use the quantities α and β defined in Lopuhaä (1989, expression 5.3) with $\psi(x) = 1$, which yields $\alpha = \frac{1}{p}$ and $\beta = \mathrm{E}_{F_{0,\mathrm{I}}}\left[\left(1 - \frac{1}{p}\right)\frac{1}{||X||}\right]$. It follows that the asymptotic variance of the location estimator can be expressed as $\mathrm{ASV}(\hat{\mu}, F_{0,\mathrm{I}}) = \frac{\alpha}{\beta^2} I$. Analogously, the asymptotic variances of the diagonal and the off-diagonal elements of $\hat{V}$ are

$$\mathrm{ASV}(\hat{V}_{ii}, F_{0,\mathrm{I}}) = \left(2 - \frac{2}{p}\right)\sigma_1 \quad \text{and} \quad \mathrm{ASV}(\hat{V}_{ij}, F_{0,\mathrm{I}}) = \sigma_1$$

with

$$\sigma_1 = \frac{p(p+2)\mathrm{E}_{F_{0,\mathrm{I}}}[||X||^2]}{\{\mathrm{E}_{F_{0,\mathrm{I}}}[(p+1)||X||]\}^2}.$$

For the scatter matrix $\hat{\Sigma}$ the asymptotic variances become:

$$\mathrm{ASV}(\hat{\Sigma}_{ii}, F_{0,\mathrm{I}}) = 2\sigma_1 + \sigma_2 \text{ and } \mathrm{ASV}(\hat{\Sigma}_{ij}, F_{0,\mathrm{I}}) = \sigma_1$$

with

$$\sigma_2 = -\frac{2}{p}\sigma_1 + \frac{4\mathrm{E}_{F_{0,\mathrm{I}}}[(||X|| - b)^2]}{\mathrm{E}_{F_{0,\mathrm{I}}}[||X||]^2}.$$

The asymptotic relative efficiencies (ARE) of the location estimator with respect to the empirical mean (given by $\mathrm{ARE}(\hat{\mu}, \overline{X}) = \mathrm{ASV}(\overline{X})/\mathrm{ASV}(\hat{\mu})$) at elliptical distributions equal those of the Oja median (Hettmansperger, Nyblom and Oja 1994) and the multivariate median of Hettmansperger and Randles (2002). Figure 1 shows the ARE at the multivariate normal $N_p(0, \mathrm{I})$ and t_5- and t_8-distributions for increasing dimensions. For t-distributions the

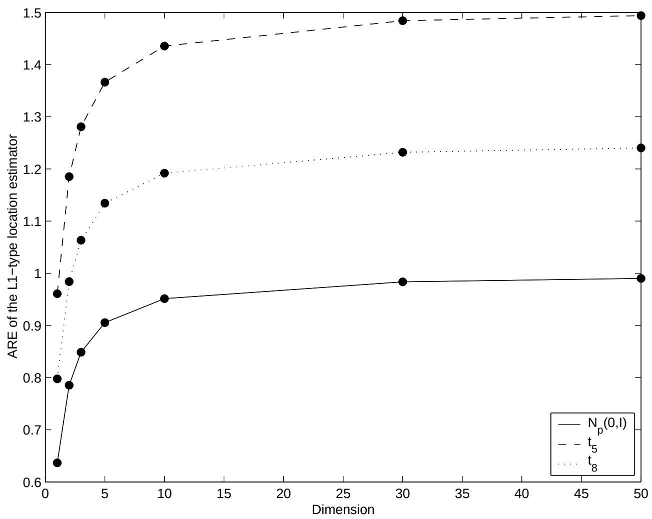

Fig. 1 ARE of the L1-type location estimator at $N_p(0, \mathrm{I})$, t_5 and t_8

location estimator is more efficient than the empirical mean. As for many estimators, the ARE increases with the dimension.

Figure 2 shows the ARE of the shape estimator. Note that for the empirical shape matrix estimator C it holds that $\mathrm{ASV}(C_{ii}, N_p(0, \mathrm{I})) = 2 - 2/p$ and $\mathrm{ASV}(C_{ii}, t_\nu) = (2 - 2/p)\frac{\nu-2}{\nu-4}$. The conclusions are the same as for the location estimator and the efficiency difference is even larger.

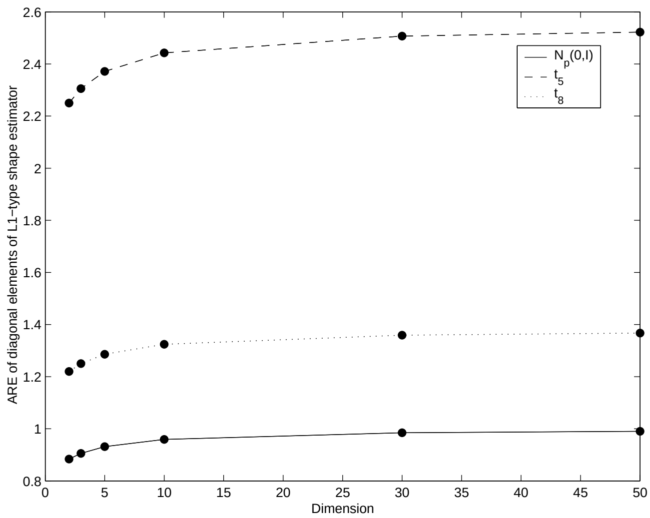

Fig. 2 ARE of the L1-type estimator of shape at $N_p(0,\mathrm{I})$, t_5 and t_8

For the off-diagonal elements of the scatter estimator the ARE is the same as for the shape matrix (Figure 2). Figure 3 shows the ARE for the diagonal elements which show a similar behavior.

We now compare the asymptotic results obtained above with the finite-sample behavior of the estimator. Based on $m = 5000$ data sets with n observations and p variables generated from an elliptical distribution we estimate the finite-sample variances of the location estimator $\hat{\mu}$ as $n \operatorname*{ave}_{1\leq j\leq p} \operatorname*{var}_{k}(\hat{\mu}_j^k)$ where $\hat{\mu}_j^k$ is the jth component of the location estimate for the kth sample ($1 \leq k \leq m$). Let $\hat{V}_{ij}^k$ denote the element (i,j) of the shape estimate obtained from the kth sample. The accuracy of a diagonal element is then measured by the standardized variance

$$\mathrm{StVar}(\hat{V}_{ii}) = \frac{n\mathrm{var}_k(\hat{V}_{ii}^k)}{[\mathrm{ave}_k(\hat{V}_{ii}^k)]^2}.$$

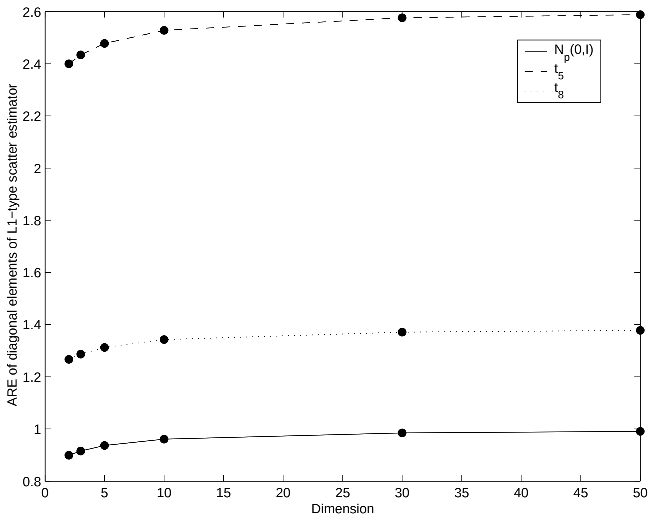

Fig. 3 ARE of the diagonal elements of the L1-type estimator of scatter at $N_p(0, \mathrm{I})$, t_5 and t_8

The accuracy of an off-diagonal element is measured by the mean squared error (MSE):

$$\mathrm{MSE}(\hat{V}_{ij}) = \frac{n}{m}\sum_{k=1}^{m}(\hat{V}_{ij}^{k})^2.$$

Finite-sample efficiencies for data generated from the multivariate t_8-distribution are given for the location estimator $\hat{\mu}$ (Table 1), off-diagonal elements of the shape estimator $\hat{V}$ (Table 2) and diagonal elements of the scatter estimator $\hat{\Sigma}$ (Table 3). Table 1 shows that both finite-sample and asymptotically the efficiencies of the L1-type location estimator are better than those of $\overline{X}$ for $p \geq 3$. The finite-sample efficiencies show the same behavior as the asymptotic relative efficiencies but are a bit lower. Table 2 and 3 show that the efficiencies of the L1-type estimators of shape and scatter are always better than those of the empirical estimators.

	$n = 30$	$n = 50$	$n = 100$	$n = 200$	$n = 500$	$n = \infty$
$p = 1$	0.7666	0.7886	0.7871	0.7986	0.7950	0.7975
$p = 2$	0.9531	0.9684	0.9783	0.9965	0.9827	0.9839
$p = 3$	1.0352	1.0551	1.0464	1.0559	1.0633	1.0634
$p = 5$	1.1085	1.1281	1.1243	1.1256	1.1300	1.1343
$p = 10$	1.1471	1.1736	1.1796	1.1826	1.1869	1.1918
$p = 30$		1.1397	1.1982	1.2149	1.2235	1.2320
$p = 50$			1.1697	1.2108	1.2288	1.2403

Table 1 Relative efficiencies of the L1-type estimate of location at t_8

	$n = 30$	$n = 50$	$n = 100$	$n = 200$	$n = 500$	$n = \infty$
$p = 2$	1.0923	1.1351	1.1724	1.1631	1.2063	1.2198
$p = 3$	1.1430	1.1728	1.2166	1.2034	1.2484	1.2500
$p = 5$	1.2046	1.2403	1.2627	1.2560	1.2778	1.2857
$p = 10$	1.2716	1.3101	1.3199	1.3138	1.3252	1.3243
$p = 30$		1.2728	1.3753	1.3763	1.3749	1.3592
$p = 50$			1.3414	1.3922	1.3946	1.3674

Table 2 Relative efficiencies of the off-diagonal elements of the L1-type shape estimator

	$n=30$	$n=50$	$n=100$	$n=200$	$n=500$	$n=\infty$
$p=2$	1.2034	1.2594	1.2662	1.2464	1.2521	1.2667
$p=3$	1.2333	1.2683	1.2788	1.2744	1.2706	1.2868
$p=5$	1.2405	1.2829	1.2994	1.3023	1.3079	1.3125
$p=10$	1.2394	1.2909	1.3180	1.3190	1.3420	1.3424
$p=30$		1.1904	1.2998	1.3381	1.3604	1.3711
$p=50$			1.2401	1.3229	1.3626	1.3781

Table 3 Relative efficiencies of the diagonal elements of the L1-type scatter estimator

5 Comparison to other estimators

For any affine equivariant location functional M that possesses an influence function, there exists a real-valued function γ_M such that $\mathrm{IF}(x; M, F_{\mu,\Sigma}) = \gamma_M(d(x))(x-\mu)/d(x)$ where $d^2(x) = (x-\mu)^T\Sigma^{-1}(x-\mu)$ (Ollila, Oja and Hettmansperger 2002). Furthermore, for any affine equivariant scatter matrix functional C that possesses an influence function, there exist two real-valued functions α_C and β_C such that

$$\mathrm{IF}(x; C, F_{\mu,\Sigma}) = \alpha_C(d(x))(x-\mu)(x-\mu)^T - \beta_C(d(x))\Sigma$$

(Croux and Haesbroeck 2000, lemma 1).

We compare the γ_M and α_C functions of several estimators: the L1-type estimator, the empirical mean ($\overline{X}$) and covariance (Cov), the Minimum Covariance Determinant (MCD) estimator (Rousseeuw 1984) and S-estimators (Davies 1987, Lopuhaä 1989) with Tukey's biweight ρ-function.

L1-type	$\gamma_{\mathrm{L1}}(t) = \frac{1}{\mathrm{E}_{F_{0,\mathrm{I}}} \quad 1-\frac{1}{p} \quad \frac{1}{\|\|X\|\|}}$
Mean	$\gamma_{\overline{X}}(t) = t$
MCD	$\gamma_{\mathrm{MCD}}(t) = \frac{I(t \leq \sqrt{q_\alpha})}{F_{\chi^2_{p+2}}(q_\alpha)}$ where $q_\alpha = \chi^2_{p,1-\alpha}$
S	$\gamma_{\mathrm{S}}(t) = \frac{\rho'(t)}{\mathrm{E}_{F_{0,\mathrm{I}}} \quad 1-\frac{1}{p} \quad \frac{\rho'(\|\|X\|\|)}{\|\|X\|\|} + \frac{1}{p}\rho''(\|\|X\|\|)}$

Table 4 Functions γ_M for the L1-type estimator, empirical mean, MCD- and S-estimator.

The γ_M functions are given in Table 4 and Figure 4 shows them for $p = 2$ at a bivariate normal distribution. We used the MCD- and S-estimator with 25% breakdown point. For robustness, γ_M should be continuous and

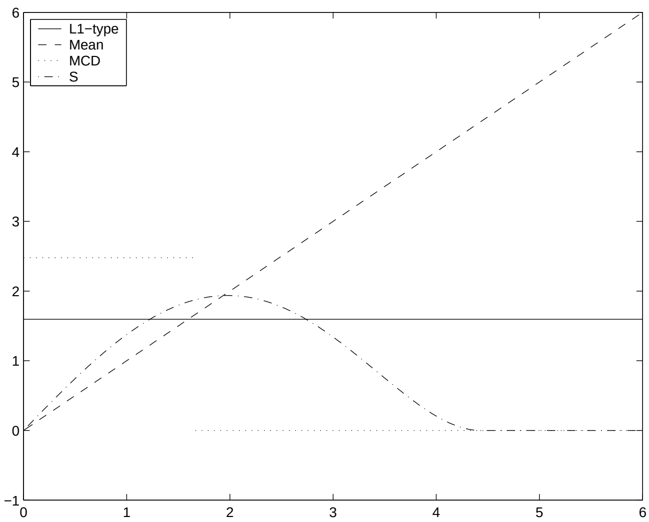

Fig. 4 Examples of the function γ_M for some estimators

bounded. The L1-type estimator satisfies these conditions, but because it is a constant function, outliers do not have a smaller effect than the rest of the data. Outliers have a larger effect on the empirical mean than 'regular' points which shows its lack of robustness. The S- and MCD-estimators are more robust because outliers have no effect after a certain point. On the other hand, this figure also illustrates the local robustness optimality (for small amounts of contamination) of the L1-type location estimator. It is most B-robust which means that it minimizes the maximal possible effect of a small amount of contamination (see Croux and Dehon 2001).

We now compare the α_C functions which are given in Table 5 and plotted in Figure 5. For the L1-type estimator, α_{L1} is decreasing meaning that the

L1-type	$\alpha_{\mathrm{L1}}(t) = \frac{p}{\gamma_1 t}$ where $\gamma_1 = \frac{\mathrm{E}_{F_{0,\mathrm{I}}}[(p+1)\|\|X\|\|]}{p+2}$
Cov	$\alpha_{\mathrm{Cov}}(t) = 1$
MCD	$\alpha_{\mathrm{MCD}}(t) = \frac{I(t \leq \sqrt{q_\alpha})}{F_{\chi^2_{p+4}}(q_\alpha)}$ where $q_\alpha = \chi^2_{p,1-\alpha}$
S	$\alpha_{\mathrm{S}}(t) = \frac{p\rho'(t)}{\gamma_1 t}$ where $\gamma_1 = \frac{\mathrm{E}_{F_{0,\mathrm{I}}}[\rho''(\|\|X\|\|)\|\|X\|\|^2 + (p+1)\rho'(\|\|X\|\|)\|\|X\|\|]}{p+2}$

Table 5 Functions α_C for the L1-type estimator, empirical covariance, MCD- and S-estimator.

effect of a point on the estimator decreases when its distance from the center increases. On the other hand, for the function α_{Cov} outliers do not have a smaller effect resulting in bad robustness properties. The functions for the S- and the MCD-estimator again become 0 after a certain point.

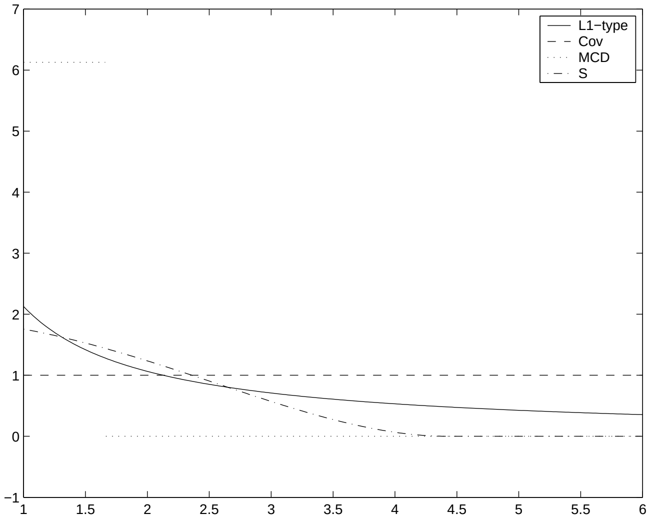

Fig. 5 Examples of the function α_C for some estimators

6 Example

We consider a data set taken from the 'The Data and Story Library' (http://lib.stat.cmu.edu/DASL/Stories/Forbes500CompaniesSales.html), which contains several facts about 79 companies selected from the Forbes 500 list of 1986. We look at the following six variables: Assets (amount of assets in millions), Sales (amount of sales in millions), Market-value (market-value of the company in millions), Profits (profits in millions), Cash-flow (cash-flow in millions) and Employees (number of employees in thousands). We use the L1-type estimator to find an estimate of location and shape/scatter. Table 6 compares the location estimate with the empirical mean. Clearly, there are large differences between both location estimates. The empirical means are much higher than the L1-type estimates. A comparison of the

Empirical mean	5940.53	4178.29	3269.75	209.84	400.93	37.60
L1-type mean	2679.33	1757.50	1099.14	89.90	164.10	15.63

Table 6 Empirical mean and L1-type mean estimate for the location of the Forbes data set

scatter matrix estimates (not shown) revealed similar effects. The difference between both estimates is caused by the presence of outliers in the data set as shown in Figure 6. This plot compares the Mahalanobis distances computed with empirical mean and covariance matrix with those based on the L1-type estimates of location and scatter using a distance-distance plot as proposed by Rousseeuw and Van Driessen (1999). The axes are rescaled to have the same length. If we draw horizontal and vertical lines at the usual

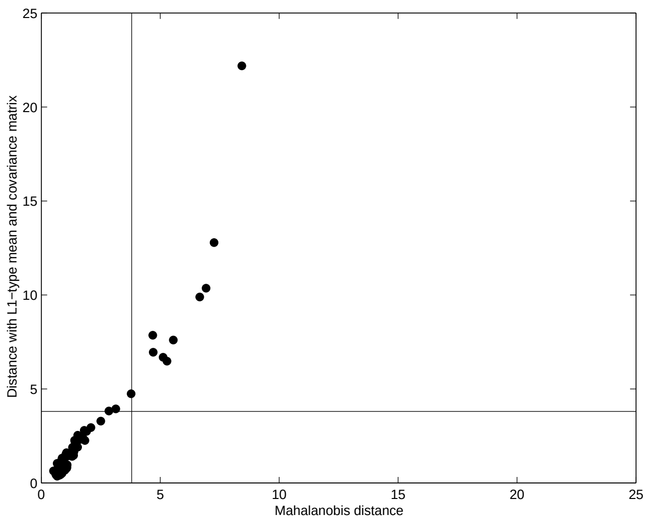

Fig. 6 Distances computed with the L1-type mean and covariance matrix versus those computed with the empirical mean and covariance matrix

cut off $\sqrt{\chi^2_{6,0.975}} = 3.0812$ we notice that outliers are more pronounced (further from the center) in the L1-type distances. Hence the L1-type estimates are less affected by the outliers (especially the scatter matrix is less inflated) than the empirical estimates which makes them more suitable in the presence of mild contamination.

7 Conclusion

We considered a multivariate estimator which is defined as a natural generalization of the univariate median. This results in an affine equivariant estimator which is highly efficient at long-tailed elliptical distributions such as multivariate t-distributions. Since the estimator has a zero breakdown point, it is less robust than high breakdown point estimators such as the MCD- and S-estimators. However, with small amounts of contamination or for heavily tailed distributions it offers an easy to compute, reliable alternative for estimating the location and shape of a data set. To illustrate the computational efficiency we give computation times for a data set of size $n = 100$ and dimension $p = 10$ the algorithm takes 0.005 seconds and still only 13.4 seconds for a data set with $n = 100\,000$ and $p = 50$.

Appendix

Derivation of (4)-(6)**:** Analogously to the estimating equations (2) and (3) it can be shown that the L1-functional $T(H) = (\hat{\mu}(H), \hat{\Sigma}(H))$ can be represented by the following equations:

$$\int \frac{1}{d_H(r)}(x - \hat{\mu}(H))^T dH = 0 \tag{7}$$

$$\int \frac{p}{d_H(r)}(x - \hat{\mu}(H))(x - \hat{\mu}(H))^T dH = \int v(d_H(r)) dH \hat{\Sigma}(H) \tag{8}$$

where $r = x - \hat{\mu}(H)$, $d_H(r) = (x - \hat{\mu}(H))^T \hat{\Sigma}(H)^{-1}(x - \hat{\mu}(H))$ and $v(d_H(r)) = b$. We assume now that $\hat{\mu}(H) = 0$ and $\hat{\Sigma}(H) = I$. Define $H_\epsilon := (1-\epsilon)H + \epsilon \Delta_{x_0}$. We only derive the influence function for the location $\hat{\mu}$. The influence function for $\hat{\Sigma}$ is derived in a similar way, by differentiating equation (8). From (7) it follows that

$$\frac{\partial}{\partial \epsilon} \int \frac{1}{d_{H_\epsilon}(r)}(x - \hat{\mu}(H_\epsilon))^T dH_\epsilon|_{\epsilon=0} = 0$$

and splitting dH_ϵ we get

$$\frac{\partial}{\partial \epsilon}\left[(1-\epsilon)\int \frac{1}{d_{H_\epsilon}(r)}(x - \hat{\mu}(H_\epsilon))^T dH + \epsilon \int \frac{1}{d_{H_\epsilon}(r)}(x - \hat{\mu}(H_\epsilon))^T d\Delta_{x_0}\right]|_{\epsilon=0} = 0.$$

Differentiating and accounting for equation (7) yields

$$-\int \frac{-1}{d_{H_\epsilon}(r)^2}\frac{\partial}{\partial \epsilon} d_{H_\epsilon}(r)|_{\epsilon=0}(x - \hat{\mu}(H))^T dH - \int \frac{1}{d_H(r)}(-\mathrm{IF}(x_0; \hat{\mu}, H))^T dH = \frac{1}{d_H(r_0)}(x_0 - \hat{\mu}(H))^T$$

Since $\hat{\mu}(H) = 0$ and $\hat{\Sigma}(H) = I$ we have $r = x$ and $d_H(r) = \sqrt{x^T x} = \|x\|$. So we can rewrite

$$\int \frac{1}{\|x\|^2}\frac{\partial}{\partial \epsilon} d_{H_\epsilon}(r)|_{\epsilon=0} x^T dH + \int \frac{1}{\|x\|} dH \mathrm{IF}(x_0; \hat{\mu}, H)^T = \frac{1}{\|x_0\|} x_0^T \tag{9}$$

It holds that

$$\frac{\partial}{\partial \epsilon} d_{H_\epsilon}(r)|_{\epsilon=0} = \frac{1}{2} d_H(r)^{-1}\left[-2\mathrm{IF}(x_0; \hat{\mu}, H)^T \hat{\Sigma}^{-1}(H)x + x^T \mathrm{IF}(x_0; \hat{\Sigma}^{-1}, H)x\right]$$

Combining this with (9) and using symmetry of H the latter term of this derivative vanishes, hence we obtain:

$$\mathrm{IF}(x_0; \hat{\mu}, H)^T \int \left(-\frac{1}{p} + 1\right) \frac{1}{\|x\|} dH = \frac{x_0^T}{\|x_0\|}$$

thus

$$\mathrm{IF}(x; \hat{\mu}, F_{0,\mathrm{I}}) = \frac{x/||x||}{\mathrm{E}_{F_{0,\mathrm{I}}}\left[\left(1 - \frac{1}{p}\right)\frac{1}{||X||}\right]}$$

References

1. Brown BM (1983) Statistical uses of the spatial median. Journal of the Royal Statistical Society Series B 45: 25-30
2. Chakraborty B and Chaudhuri P (1996) On a transformation and re-transformation technique for constructing an affine equivariant multivariate median. Proceedings of the American Mathematical Society 124: 2539-2547
3. Croux C and Dehon C (2001) Robust linear discriminant analysis using S-estimators. The Canadian Journal of Statistics 29: 473-493
4. Croux C and Haesbroeck G (2000) Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. Biometrika 87: 603-618
5. Davies PL (1987) Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. The Annals of Statistics 15: 1269-1292
6. Gower JC (1974) The mediancentre. Journal of the Royal Statistical Society Series C - Applied Statistics 23: 466-470
7. Grübel R (1996) Orthogonalization of multivariate location estimators: the orthomedian. The Annals of Statistics 24: 1457-1473

8. Hampel FR, Ronchetti EM, Rousseeuw PJ and Stahel WA (1986) Robust Statistics: The Approach Based on Influence Functions. Wiley, New York

9. Hettmansperger TP, Nyblom J and Oja H (1994) Affine invariant multivariate one-sample sign tests. Journal of the Royal Statistical Society Series B 56: 221-234

10. Hettmansperger TP and Randles RH (2002) A practical affine equivariant multivariate median. Biometrika 89: 851-860

11. Liu RY (1990) On a notion of data depth based on random simplices. The Annals of Statistics 18: 405-414

12. Lopuhaä HP (1989) On the relation between S-estimators and M-estimators of multivariate location and covariance. The Annals of Statistics 17: 1662-1683

13. Lopuhaä HP and Rousseeuw PJ (1991) Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. The Annals of Statistics 19: 229-248

14. Niinimaa A and Oja H (1995) On the influence functions of certain bivariate medians. Journal of the Royal Statistical Society Series B 57: 565-574

15. Oja H (1983) Descriptive statistics for multivariate distributions. Statistics and Probability Letters 1: 327-332

16. Ollila E, Oja H and Hettmansperger TP (2002) Estimates of regression coefficients based on the sign covariance matrix. Journal of the Royal Statistical Society Series B 64: 447-466

17. Rousseeuw PJ (1984) Least median of squares regression. Journal of the American Statistical Association 79: 871-880

18. Rousseeuw PJ and Van Driessen K (1999) A fast algorithm for the minimum covariance determinant estimator. Technometrics 41: 212-223

19. Small CG (1990) A survey of multidimensional medians. International Statistical Review 58: 263-277

20. Tukey JW (1975) Mathematics and the picturing of data. Proceedings of the International Congress of Mathematicians 2: 523-531

21. Van Aelst S and Willems G (2005) Multivariate regression S-estimators for robust estimation and inference. Statistica Sinica 15: 981-1001

22. van der Vaart AW (1998) Asymptotic Statistics. Cambridge University Press