

## Tilburg University

### On the Optimality of Multivariate S-Estimators

Croux, C.; Dehon, C.; Yadine, A.

*Publication date:*  
2010

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*

Croux, C., Dehon, C., & Yadine, A. (2010). *On the Optimality of Multivariate S-Estimators*. (CentER Discussion Paper; Vol. 2010-39). *Econometrics*.

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

No. 2010–39

**ON THE OPTIMALITY OF MULTIVARIATE S-ESTIMATORS**

By C. Croux, C. Dehon, A. Yadine

March 2010

ISSN 0924-7815

# On the Optimality of Multivariate $S$ -estimators

C. Croux

K.U. Leuven & Tilburg University

C. Dehon

A. Yadine

Université Libre de Bruxelles

Université Libre de Bruxelles

## Abstract

In this paper we maximize the efficiency of a multivariate  $S$ -estimator under a constraint on the breakdown point. In the linear regression model, it is known that the highest possible efficiency of a maximum breakdown  $S$ -estimator is bounded above by 33% for Gaussian errors. We prove the surprising result that in dimensions larger than one, the efficiency of a maximum breakdown  $S$ -estimator of location and scatter can get arbitrarily close to 100%, by an appropriate selection of the loss function.

**Keywords:** Breakdown point, Multivariate Location and Scatter, Robustness,  $S$ -estimator.

# 1 Introduction

Multivariate S-estimators are estimating the center and the scatter matrix of a multivariate data cloud. They have excellent robustness properties. Their breakdown point, which is the maximal fraction of outliers that an estimator can resist, can go up to 50%. S-estimators are fast to compute (Salibián-Barrera and Yohai 2006), and they were shown to be useful in robust multivariate analysis, e.g. for principal components analysis (Croux and Haesbroeck 2000) and for discriminant analysis (Bashir and Carter 2005). The limiting distribution of these estimators is multivariate normal, and a formula for their asymptotic variances has been given by Lopuhaä (1989). In this paper we want to study the maximal efficiency an S-estimator can attain under a constraint on the breakdown point.

For a given  $p$ -dimensional sample  $x_1, x_2, \dots, x_n$ , multivariate S-estimators for location and scatter are defined as the vector  $t_n$  and the positive definite symmetric matrix  $C_n$  that minimize the determinant  $\det(C)$ , subject to

$$\frac{1}{n} \sum_{i=1}^n \rho \left( \sqrt{(x_i - t)^t C^{-1} (x_i - t)} \right) = b, \quad (1.1)$$

over all positive definite symmetric matrices  $C$  and  $p$  dimensional vectors  $t$ . The function  $\rho$  is a loss function, and  $b$  is constant selected to have consistency.

The standard choice for  $\rho$  is the Biweight loss function, given by

$$\rho_c(u) = \min(u^2/2 - u^4/(2c^2) + u^6/(6c^4), c^2/6). \quad (1.2)$$

By appropriately selecting  $c$ , the corresponding Biweight S-estimator attains a 50% breakdown point. For smaller values of the dimension  $p$ , the Gaussian efficiency of the location Biweight S-estimator is fairly low: 25%, for  $p = 1$ , 57%, for  $p = 2$ , and 72% for  $p = 3$ . The question we want to address is whether we can improve the efficiency of the Biweight S-estimator by using other loss function than the Biweight. More precisely, given a certain value of the desired breakdown point  $0 < \varepsilon \leq 0.5$ , find the  $\rho$ -function yielding the maximal efficiency, under the constraint that the breakdown point is at least  $\varepsilon$ .

This problem was solved by Hössjer (1992) for the regression case, which corresponds to  $p = 1$ . He obtained that the efficiency of S-estimators of regression is bounded above. The highest possible Gaussian efficiency a 50% breakdown point regression S-estimator can attain is about 33%. This finding motivated researchers in robust statistics to construct new types of regression estimators, like MM-estimators (Yohai and Zamar 1988), to combine robustness with high efficiency. Multivariate S-estimators for location and scatter were proposed by Davies (1987). Also for multivariate S-estimators, alternative estimators aiming at higher efficiencies were proposed, like multivariate MM-estimators (Tatsuoka and Tyler 2000), and several others. We refer to (Maronna et al. 2006) for a review on robust multivariate estimation of location and scatter.

The result we obtain is surprising: it turns out to be possible to construct loss functions yielding multivariate S-estimators with efficiency arbitrarily close to 100%, while still keeping the desired level for the breakdown point. In the multivariate case, for  $p \geq 2$ , combining a high breakdown point and arbitrarily high efficiency is possible for an S-estimator .

The paper is organized as follows. Section 2 sets the notations and contains the preliminary results. Section 3 presents the main theorem of the paper. Numerical illustrations are given in Section 4. The last section discusses the estimation of the scatter matrix by multivariate S-estimators, for which it also holds that the efficiency can get arbitrarily close to 1.

## 2 Notations and preliminary results

Let  $X$  be a  $p$ -variate random variable having an elliptically symmetric distribution  $H$  with density

$$f_H(x) = g((x - \mu)^t \Sigma^{-1} (x - \mu)), \quad (2.1)$$

with  $g$  a positive real valued function. Since S-estimators are affine equivariant, we assume in the remainder of this article that  $\mu = 0$  and  $\Sigma = I_p$ . Let  $G(t)$  be the

distribution function of the Mahalanobis distances , so

$$G(t) = P_H(\|X\| \leq t), \text{ for all } t \geq 0,$$

and define for any  $0 < \alpha < 1$

$$G_\alpha(t) = G(t) - (1 - \alpha), \text{ for all } t \geq 0. \quad (2.2)$$

We first restate the results on the breakdown point and the asymptotic variance of  $S$ -estimators, obtained in Davies (1987) and Lopuhaä (1989). We need the following standard condition on the loss function

(A)  $\rho(0) = 0$ ,  $\rho$  is non decreasing on  $(0, +\infty)$  and absolutely continuous with bounded derivative  $\psi$ . Furthermore, there exists a constant  $c > 0$  such that  $\rho(c) = \rho(\infty)$ .

To ensure consistency of the S-estimator defined in (1.1) it is required that  $E_G[\rho(Y)] = b$ . The breakdown point of an S-estimator is then given by

$$\varepsilon^*(\rho) = \min\left(\frac{E_G[\rho(Y)]}{\rho(\infty)}, 1 - \frac{E_G[\rho(Y)]}{\rho(\infty)}\right). \quad (2.3)$$

The asymptotic variance of the location S-estimator is characterized by the number

$$ASV(\psi) = \frac{p}{4c_p} \frac{A(\psi, H)}{(B(\psi, H))^2}, \quad (2.4)$$

where the constant  $c_p = (\int_0^\infty g(z^2)z^{p-1}dz)^{-1} = 2\pi^{\frac{p}{2}}/\Gamma(\frac{p}{2})$ , and

$$A(\psi, H) = \int_0^{+\infty} g(z^2)\psi^2(z)z^{p-1}dz \quad (2.5)$$

$$B(\psi, H) = \int_0^{+\infty} g'(z^2)\psi(z)z^p dz. \quad (2.6)$$

We want to select  $\rho$ , or equivalently  $\psi$ , such that the asymptotic variance is minimal, under the condition that the breakdown point is at least equal to  $\varepsilon$ , the desired level for the breakdown point, with  $0 < \varepsilon \leq 0.5$ . As was noticed by Hössjer (1992), it is possible to express the condition on the breakdown point in terms of the  $\psi$  function. Using partial integration, one can rewrite  $E_G[\rho(Y)] = \alpha\rho(\infty)$  as

$$\int_0^{+\infty} G_\alpha(t)\psi(t)dt = 0, \quad (2.7)$$

with  $G_\alpha$  defined in (2.2), for any  $0 < \alpha < 1$ . Since  $ASV(\psi)$  is independent of scalar multiplication of  $\psi$ , we may set  $B(\psi, H) = C$ , with  $C$  a fixed constant. The problem we want to solve is

**Problem  $\mathbf{P}_\varepsilon$**  . For a given breakdown point  $\varepsilon$ , minimize  $A(\psi, H)$  with respect to  $\psi$  under the constraints that  $B(\psi, H) = C$  and (2.7) is satisfied for some  $\varepsilon \leq \alpha \leq 1 - \varepsilon$ . The function  $\psi$  needs to satisfy the regularity condition (A).

This problem was solved by (Hössjer 1992) for  $p = 1$ , but it turns out that no solution of the above problem exists for the multivariate case, as will be discussed further in Section 2. An additional restriction is required; we ask that the loss function remains zero in a neighborhood of  $\rho(0) = 0$ . This means that observations with a very small Mahalanobis distance, the “inliers”, receive a zero weight. Let  $a$  be a fixed number. Problem  $\mathbf{P}_\varepsilon$  with the additional constraint that the  $\psi$ -function equals zero between 0 and  $a$  is abbreviated as  $\mathbf{P}_{\varepsilon,a}$  .

Solving problem  $\mathbf{P}_{\varepsilon,a}$  goes along the same lines as Hössjer (1992). The method of Lagrange multipliers suggests that the solution to problem  $\mathbf{P}_{\varepsilon,a}$  is given by

$$\psi_{\varepsilon,a}(x) = \begin{cases} \psi_\varepsilon(x) & \text{if } x \geq a \\ 0 & \text{if } 0 < x < a, \end{cases} \quad (2.8)$$

with

$$\psi_\varepsilon(x) = (\Lambda_g(x) - k \frac{G_\varepsilon(x)}{g(x^2)x^{p-1}})_+, \quad (2.9)$$

where  $k = k(\varepsilon, a)$  depends on  $\varepsilon$  and  $a$ ,  $(u)_+ = \max(0, u)$  takes the positive part of any number  $u$ . The function  $\Lambda_g$  is the score function of the maximum likelihood estimator, given by

$$\Lambda_g(x) = -\frac{2g'(x^2)x}{g(x^2)}. \quad (2.10)$$

In a similar way as in Hössjer (1992), a formal proof of the optimality of the score function (2.8) for problem  $\mathbf{P}_{\varepsilon,a}$  can be given. The proof builds on two lemmas and requires three conditions (B1)-(B3) on the model distribution  $H$ . They are listed in the Appendix, and are the direct multivariate extensions of the lemmas and conditions in Hössjer (1992).

**Theorem 1** *Assume that condition (A) on the loss function  $\rho$ , and conditions (B1), (B2) and (B3) on the model distribution hold. Take any  $0 < \varepsilon < 0.5$  and any  $0 < a < G^{-1}(1 - \varepsilon)$ . The score functions  $\{\delta\psi_{\varepsilon,a}\}$ , with  $\delta > 0$  arbitrary, solve the optimization problem  $\mathbf{P}_{\varepsilon,a}$  uniquely (almost surely).*

Full details of the proofs are given in the technical note (Croux et al. 2009), where it is also shown that the multivariate normal, the multivariate  $t$  and multivariate power-distributions verify conditions (B1)-(B3).

### 3 Optimal Multivariate S-estimators of Location

In this section we present the main result of the paper: the efficiency of a multivariate S-estimator can get arbitrarily close to 100%. We will first show that the score function  $\psi_{\varepsilon,a}$ , solving  $\mathbf{P}_{\varepsilon,a}$  tends to  $\Lambda_g$  for  $a$  tending to zero. The score function  $\Lambda_g$  corresponds to the Maximum Likelihood estimator, and provides the smallest possible asymptotic variance. For many models, including the Gaussian,  $\Lambda_g$  will not be bounded, and correspond to an estimator with a zero breakdown point.

Having a closer look at the expression for  $\psi_{\varepsilon,a}$  in (2.8), reveals that  $\psi_{\varepsilon,a}$  and  $\Lambda_g$  will coincide when the constant  $k = k(\varepsilon, a)$  vanishes. The following proposition, proven in the Appendix, shows that this indeed happens.

**Proposition 1** *For any  $0 < \varepsilon \leq 0.5$  we have that:*

- (i)  $\lim_{a \downarrow 0} k(\varepsilon, a) := k(\varepsilon)$  exists
- (ii) If  $p \geq 2$ , then  $k(\varepsilon) = 0$ .

We conclude from the previous proposition that there is pointwise convergence of  $\psi_{\varepsilon,a}(x)$  to  $\Lambda_g(x)$ , for every  $x > 0$ . If the convergence would have been uniform, then convergence of  $ASV(\psi_{\varepsilon,a})$  to  $ASV(\Lambda_g)$ , the smallest possible variance, is immediate. However, the function values of  $\psi_{\varepsilon,a}(x)$  for  $x$  close to 0, tend to infinity for  $a$  tending to zero. This is caused by the factor  $x^{p-1}$  in the denominator of (2.9), tending to



infinity for  $x$  converging to zero (if  $p \geq 2$ ). Before stating the final result, we need two additional conditions

$$(\mathbf{C}_1) \quad \lim_{x \rightarrow +\infty} x^p g'(x^2) \ln(g(x^2)) = 0$$

$$(\mathbf{C}_2) \quad E_H[\ln(g(\|X\|^2))] \text{ exists.}$$

It can be verified, see Croux et al. (2009) for details, that the multivariate normal, the multivariate  $t$  and multivariate power-distributions all verify conditions (C1)-(C2).

**Theorem 2** *For  $p \geq 2$ , and  $0 < \varepsilon \leq 0.5$ , we have*

$$ASV(\psi_{\varepsilon,a}) \longrightarrow ASV(\Lambda_g) \text{ for } a \rightarrow 0.$$

In the next section we will compute asymptotic efficiencies, defined as

$$\text{Eff}(\psi_{\varepsilon,a}) = \frac{ASV(\Lambda_g)}{ASV(\psi_{\varepsilon,a})}, \quad (3.1)$$

for several values of  $\varepsilon$  and  $a$ . As Theorem 2 shows, we have that  $\text{Eff}(\psi_{\varepsilon,a})$  converges to 100%, for  $a$  tending to zero, and this at any model distribution satisfying the conditions (B1)-(B3), and (C1)-(C2).

## 4 Numerical Illustrations

In this section we compute asymptotic efficiencies at the multivariate normal distribution, where  $g(y) = \exp(-y)/(2\pi)^{(p/2)}$ . Hence  $G(x) = F_{\chi_p^2}(x^2)$ , with  $F_{\chi_p^2}(\cdot)$  the cdf of a chi-square distribution with  $p$  degrees of freedom, and  $\Lambda_g(x) = x$ . The constant  $k = k(a, \varepsilon)$  is selected such that the breakdown point condition (2.7) is verified. The constant  $k$  is the root of a strictly decreasing real valued function, see equation (A.1), and computing it poses no numerical difficulties.

In Figure 4 we plot the loss function solving problem  $\mathbf{P}_{\varepsilon,a}$ , together with the Biweight loss and the quadratic loss (corresponding to the Maximum Likelihood estimator). We see that the optimal loss functions (solid lines; left panel) are continuous and bounded, as the Biweight. There are two differences (i) their curvature

Table 1: Gaussian efficiency of the optimal location S-estimator for different values of  $a$  and  $p$ . The breakdown point  $\varepsilon$  is 0.5 or 0.25. The first row corresponds to the Biweight S-estimator.

		$p = 2$	$p = 3$	$p = 4$	$p = 5$
		<hr/>			
Biweight		0.5796	0.7224	0.7998	0.8463
$\varepsilon = 0.5$	$a = 0.1$	0.6469	0.8463	0.9558	0.9925
	$a = 0.01$	0.7111	0.9504	0.9980	1.0000
	$a = 0.001$	0.7470	0.9887	1.0000	1.0000
		<hr/>			
Biweight		0.9118	0.9514	0.9676	0.9760
$\varepsilon = 0.25$	$a = 0.1$	0.9452	0.9855	0.9970	0.9996
	$a = 0.01$	0.9559	0.9952	0.9998	1.0000
	$a = 0.001$	0.9618	0.9989	1.0000	1.0000
		<hr/>			

is more quadratic (ii) they increases sharply in a neighborhood of zero. These two differences are reflected in the  $\psi_{\varepsilon,a}$  functions (solid lines; right panel); they (i) are close to  $\Lambda_g(x) = x$  (ii) attain high, but bounded, values if the argument is close to zero.

In Table 1, we give the efficiency of the location S-estimator with score function  $\psi_{\varepsilon,a}$  for different values of  $p$  and  $a$ . We see that it is indeed possible to improve substantially the efficiency of the Biweight S-estimators using other loss functions. Table 1 confirms that there is convergence towards 100% efficiency, for  $a$  tending to zero. This convergence, however, is rather slow, in particular for  $p = 2$ . For larger values of  $p$ , the efficiency of the Biweight S-estimator is higher, and convergence towards the maximal efficiency is faster.

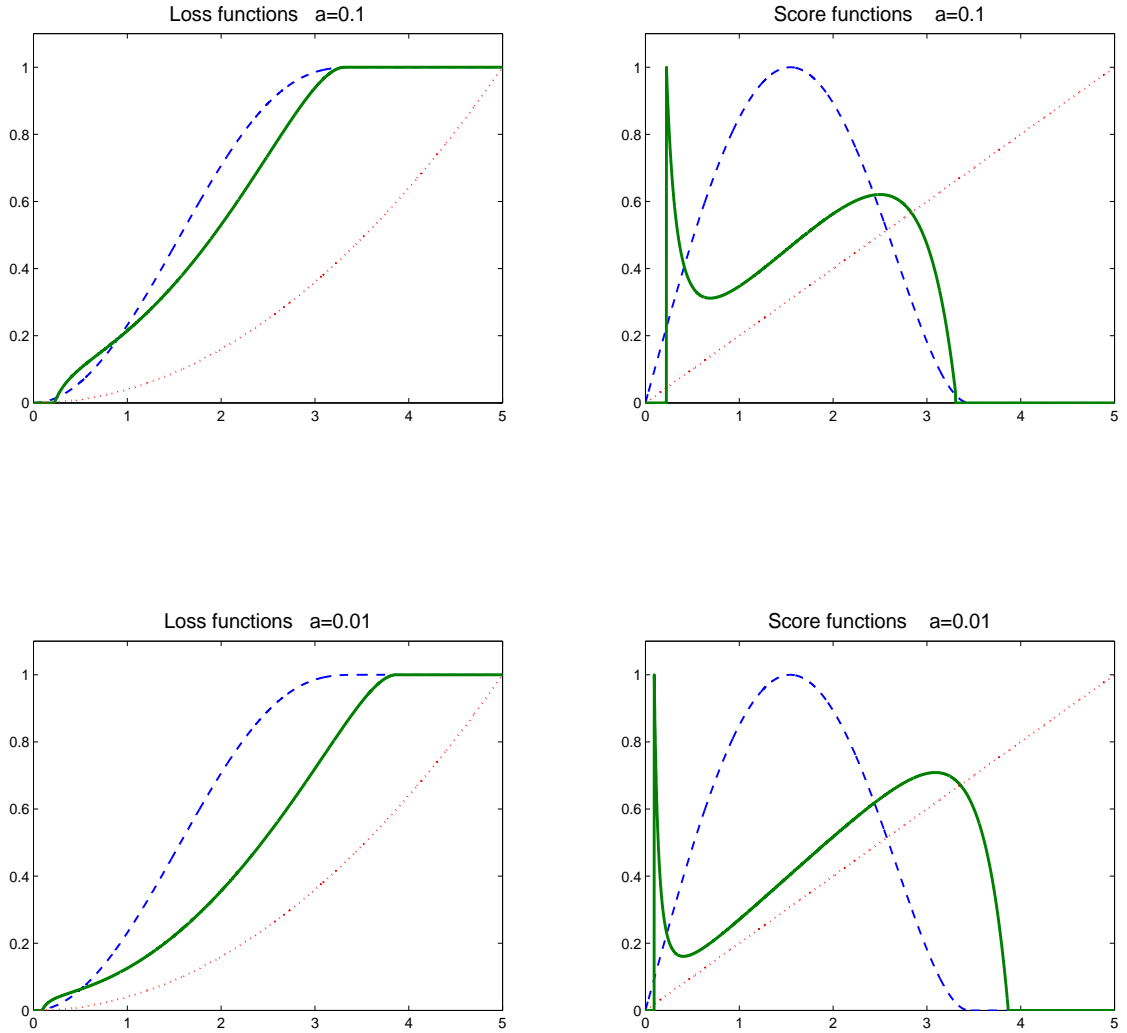


Figure 1: The loss functions solving problem  $\mathbf{P}_{\varepsilon,a}$ , for  $\varepsilon = 0.5$  and  $p = 3$ , together with the Biweight loss (dashed line) and the quadratic loss (dotted line). The  $\rho$  functions are given in the left panel, the score functions  $\psi$  in the right panel. Two values of  $a$  are considered:  $a = 0.1$  and  $a = 0.01$ .

## 5 Discussion

In section 3 we proved that a location S-estimator with a given breakdown point can attain an efficiency arbitrarily close to 100%. For the scatter matrix estimator one can follow exactly the same approach. We characterize the precision of a scatter matrix estimator by the asymptotic variance of an off-diagonal element, as it determines the limiting variance of the standardized scatter matrix (i.e. the shape matrix), e.g. Frahm (2009). Its formula resembles (2.4), but with

$$A(\psi, H) = \int_0^{+\infty} g(z^2)\psi^2(z)z^{p+1}dz \quad (5.1)$$

$$B(\psi, H) = \int_0^{+\infty} g'(z^2)\psi(z)z^{p+2}dz. \quad (5.2)$$

The difference with (2.5) and (2.6) is that the exponent  $p$  for the location problem needs to be replaced by the exponent  $p + 2$  for the scatter problem, but all proofs still carry through. Under the same conditions on the model distribution, it then follows that also S-estimators of scatter can attain an efficiency arbitrarily close to 100%. A difference with the location case is that we do not need the condition  $p \geq 2$  anymore; optimal efficiency can also be attained for univariate scale estimators, as was already shown in Croux (1994)

In Table 2 we give the efficiency of the optimal scatter matrix S-estimator for different values of the breakdown point,  $a$  and  $p$ . Convergence towards 100% efficiency for  $a$  tending to zero is faster than for the location case. Note that for  $p = 4$ , the biweight estimator already has a high efficiency; see Maronna et al. (2006) for a discussion on the behavior of S-estimators in high dimensions.

To conclude, let us restate the main result of this paper. While regression S-estimators have limited efficiency, this is not true for S-estimators of multivariate location/scatter. The optimal S-estimators we discussed combine high efficiency and high breakdown point. In this paper, robustness is only measured by means of the breakdown point. While the breakdown point is the most well-known measure of robustness, it is not the only one. For example, the local shift sensitivity of the

Table 2: Gaussian efficiency of the optimal scatter matrix S-estimator for different values of  $a$  and  $p$ . The breakdown point  $\varepsilon$  is 0.5 or 0.25. The first row corresponds to the Biweight S-estimator.

		$p = 2$	$p = 3$	$p = 4$	$p = 5$
Biweight		0.3765	0.5794	0.7025	0.7784
$\varepsilon = 0.5$	$a = 0.1$	0.9144	0.9884	0.9989	0.9999
	$a = 0.01$	0.9972	1.0000	1.0000	1.0000
Biweight		0.8498	0.9243	0.9528	0.9669
$\varepsilon = 0.25$	$a = 0.1$	0.9882	0.9988	0.9999	1.0000
	$a = 0.01$	0.9992	1.0000	1.0000	1.0000

S-estimators based on  $\psi_{\varepsilon,a}$  will be infinite. A topic for future research is to compare the maxbias of several highly efficient high breakdown estimators for multivariate location and scatter, including multivariate S-estimators.

## References

- Bashir, S. and Carter, E. M. (2005), “High breakdown mixture discriminant analysis,” *Journal of Multivariate Analysis*, 93, 102 – 111.
- Croux, C. (1994), “Efficient High-Breakdown M-Estimators of Scale,” *Statistics and Probability Letters*, 19, 371–379.
- Croux, C., Dehon, C., and Yadine, A. (2009), “Technical note: on the optimality of multivariate S-estimators,” Document available at [www.econ.kuleuven.be/christophe.croux/public/public.htm](http://www.econ.kuleuven.be/christophe.croux/public/public.htm).
- 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.

- Davies, P. L. (1987), “Asymptotic Behaviour of S-Estimates of Multivariate Location Parameters and Dispersion Matrices,” *Annals of Statistics*, 15, 1269–1292.
- Frahm, G. (2009), “Asymptotic distributions of robust shape matrices and scale,” *Journal of Multivariate Analysis*, 100, 1329–1337.
- Hössjer, O. (1992), “On the optimality of S-estimators,” *Statistics & Probability Letters*, 14, 413–419.
- Lopuhaä, H. (1989), “On the relation between S-estimators and M-estimators of multivariate location and covariance,” *Annals of Statistics*, 17, 1662–1683.
- Maronna, R. A., Martin, D. R., and Yohai, V. J. (2006), *Robust Statistics: Theory and Methods*, New York: Wiley.
- Salibian-Barrera, M. and Yohai, V. (2006), “A fast algorithm for S-regression estimates,” *Journal of Computational and Graphical Statistics*, 15, 414–427.
- Tatsuoka, K. and Tyler, D. (2000), “On the uniqueness of S-functionals and M-functionals under nonelliptical distributions,” *The Annals of Statistics*, 28, 1219–1243.
- Yohai, V. and Zamar, R. (1988), “High breakdown-point estimates of regression by means of the minimization of an efficient scale,” *Journal of the American statistical association*, 83, 406–413.

## A Appendix

The following conditions on the model distribution  $H$  are the multivariate versions of the conditions in Hössjer (1992).

(**B**<sub>1</sub>) The function  $g$  is strictly positive and bounded, with derivative  $g' < 0$ .

(**B**<sub>2</sub>) For every  $k > 0$  and  $0 < \alpha < 1$ , the function  $h_{\alpha,k}(x) = 2g'(x^2)x^p + kG_\alpha(x)$  is continuous, has a unique zero  $c(\alpha, k)$ , such that  $h_{\alpha,k}(x) > 0$  for  $x > c(\alpha, k)$ , and  $h_{\alpha,k}(x) < 0$  for  $x < c(\alpha, k)$ .

(**B**<sub>3</sub>) Let  $\Lambda_g(x) = -2g'(x^2)x/g(x^2)$  be the likelihood score function. We require that  $\int_0^{+\infty} \Lambda_g(x)dx = \infty$ .

Using these conditions, one can proof the following two lemmas

**Lemma 1** *Under assumptions ( $B_1$ ), ( $B_2$ ) and ( $B_3$ ), there exists for each  $k > 0$  and  $0 < \alpha < 1$  a unique  $c = c(\alpha, k)$  such that  $\Lambda_g(x) - k \frac{G_\alpha(x)}{g(x^2)x^{p-1}} \stackrel{\leq}{>} 0$ , whenever  $x > 0$  and  $x \stackrel{\geq}{<} c(\alpha, k)$ . Moreover, with  $\alpha$  fixed,  $c(\alpha, k)$  is a continuous and strictly decreasing function of  $k$ , with  $c(\alpha, k) \rightarrow +\infty$  as  $k \rightarrow 0^+$  and  $c(\alpha, k) \rightarrow c_\alpha^+$  as  $k \rightarrow +\infty$ .*

The next step is to find for each  $\alpha$  and  $a$  a corresponding  $k = k(\alpha, a)$  such that (2.7) is satisfied. For this purpose we introduce the following function

$$J_{\alpha,a}(k) = \int_a^{c(\alpha,k)} \psi_{\alpha,a,k}(x)G_\alpha(x)dx. \quad (\text{A.1})$$

**Lemma 2** *Let  $0 < \alpha < 1$ , and take  $0 < a < c_\alpha$ . Assume that ( $B_1$ ) – ( $B_3$ ) hold. Then  $J_{\alpha,a}(k)$  is a continuous, strictly decreasing function of  $k$  with a unique zero  $k(\alpha, a)$ .*

The proofs are almost identical to those in Hössjer (1992). Full details are given in the technical note (Croux et al. 2009).

## B Appendix

In this Appendix we proof the proposition and the Theorem of Section 3.

*Proof of Proposition 1.* To prove (i) and since  $k(\varepsilon, a)$  is positive, it is sufficient to show that  $k(\varepsilon, a)$  is increasing for  $0 < a < G^{-1}(1 - \varepsilon) = c_\varepsilon$ . For this, take  $0 < a < b < c_\varepsilon$ , and let us show that  $k(\varepsilon, a) < k(\varepsilon, b)$ . By definition,  $k(\varepsilon, a)$  solves  $J_{\varepsilon,a}(k) = 0$  and  $k(\varepsilon, b)$  solves  $J_{\varepsilon,b}(k) = 0$ , see Lemma 2 . From (A.1) it follows that

$$J_{\varepsilon,a}(k) = \int_a^b \psi_{\varepsilon,a,k}(x)G_\varepsilon(x)dx + J_{\varepsilon,b}(k), \quad (\text{B.1})$$

for every  $k$ . Since  $b < c_\varepsilon$ , the first term of (B.1) is strictly positive, implying that  $J_{\varepsilon,a}(k(\varepsilon, b)) < J_{\varepsilon,b}(k(\varepsilon, b)) = 0 = J_{\varepsilon,a}(k(\varepsilon, a))$ . Since  $J_{\varepsilon,b}(k)$  is strictly decreasing in  $k$ , see Lemma 2, we must have  $k(\varepsilon, a) < k(\varepsilon, b)$ .

Now we proof (ii) by contradiction. Assume that  $k(\varepsilon) > 0$ . Then we also have that  $c(\varepsilon, k(\varepsilon))$  is finite. By definition of  $k(\varepsilon, a)$  and  $c(\varepsilon, k(\varepsilon, a))$ , we have for every  $a > 0$

$$\int_a^{c(\varepsilon, k(\varepsilon, a))} \Lambda_g(x)G_\varepsilon(x)dx = k(\varepsilon, a) \int_a^{c(\varepsilon, k(\varepsilon, a))} \frac{G_\varepsilon^2(x)}{g(x^2)x^{p-1}}dx.$$

Taking the limit to  $a = 0$ , yields

$$\int_0^{c(\varepsilon, k(\varepsilon))} \Lambda_g(x)G_\varepsilon(x)dx = k(\varepsilon) \int_0^{c(\varepsilon, k(\varepsilon))} \frac{G_\varepsilon^2(x)}{g(x^2)x^{p-1}}(x)dx. \quad (\text{B.2})$$

The first term of the above equality is finite, using condition (B3) and the second is not, since  $p \geq 2$ . Hence we obtain a contradiction and must conclude that  $k(\varepsilon) = 0$ .

□

*Proof of Proposition 2.* Denote  $\psi_a = \psi_{\varepsilon,a,k(\varepsilon,a)}$ ,  $c = c(\varepsilon, k(\varepsilon, a))$ , and  $k = k(\varepsilon, a)$ . Recall that if  $a \downarrow 0$  then  $k \downarrow 0$ , and  $c \uparrow \infty$ . Using the definition of  $\psi_a$ , we get

$$\begin{aligned} A(\psi_a) &= \int_a^c g(r^2) \left[ \Lambda_g(r) - k \frac{G_\varepsilon(r)}{g(r^2)r^{p-1}} \right]^2 r^{p-1} dr \\ &= \int_a^c g(r^2) \Lambda_g^2(r) r^{p-1} dr - k \int_a^c \Lambda_g(r) G_\varepsilon(r) dr + k \int_a^c G_\varepsilon(r) \left( \frac{k G_\varepsilon(r)}{g(r^2)r^{p-1}} \right) - \Lambda_g(r) dr. \end{aligned}$$



The last integral is equal to 0, by definition of  $\psi_a$  and  $k$ , see Lemma 2. We can write

$$A(\psi_a) = \int_a^c \Lambda_g^2(r)g(r^2)r^{p-1}dr - k \int_a^c \Lambda_g(r)G_\varepsilon(r)rdr. \quad (\text{B.3})$$

Furthermore,

$$\begin{aligned} B(\psi_a) &= \int_a^c \psi_a(r)g'(r^2)r^p dr \\ &= -\frac{1}{2} \int_a^c (\Lambda_g(r) - k \frac{G_\varepsilon(r)}{g(r^2)r^{p-1}}) \Lambda_g(r)g(r^2)r^{p-1} dr \\ &= \frac{-A(\psi_a)}{2} \end{aligned}$$

which implies that  $ASV_p(\psi_a) = p/(c_p A(\psi_a))$ . To show that  $ASV(\psi_a) \longrightarrow ASV(\Lambda_g)$  for  $a \rightarrow 0^+$ , we thus need to show that

$$A(\psi_a) \longrightarrow A(\Lambda_g) = \int_0^{+\infty} \Lambda_g(r)g(r^2)r^{p-1}dr.$$

Since  $c(\varepsilon, k(\varepsilon, a))$  tends to infinity (see Lemma 1 and Proposition 1), it follows from (B.3) that it is sufficient to show

$$k \int_a^c \frac{g'(r^2)}{g(r^2)} G_\varepsilon(r)rdr \longrightarrow 0, \quad a \rightarrow 0.$$

We can develop this last expression as follows,

$$k \int_a^c \frac{g'(r^2)}{g(r^2)} G_\varepsilon(r)rdr \quad (\text{B.4})$$

$$= -(1-\varepsilon)k \int_a^c \frac{g'(r^2)}{g(r^2)} rdr + k \int_a^c \frac{g'(r^2)}{g(r^2)} G(r)rdr \quad (\text{B.5})$$

$$= -\frac{(1-\varepsilon)}{2}k(\ln g(c^2) - \ln g(a^2)) + k \int_a^c \frac{g'(r^2)}{g(r^2)} G(r)rdr \quad (\text{B.6})$$

For  $a \rightarrow 0$ , we have  $g(a^2) \rightarrow g(0) > 0$ , such that  $\ln(g(a^2)) \rightarrow 0, a \rightarrow 0$ . So it remains to show that

$$k \ln(g(c^2)) \longrightarrow 0, \quad a \rightarrow 0, \quad (\text{B.7})$$

and

$$k \int_a^c \frac{g'(r^2)}{g(r^2)} G(r)rdr \longrightarrow 0, \quad a \rightarrow 0. \quad (\text{B.8})$$

First note that  $c$  satisfies the equality  $\psi_a(c) = 0$ , or

$$-\frac{g'(c^2)}{g(c^2)}c = k \frac{G_\varepsilon(c)}{g(c^2)c^{p-1}},$$

then

$$\begin{aligned} k \ln g(c^2) &= -\frac{g'(c^2)}{g(c^2)} \frac{g(c^2)c^p \ln(g(c^2))}{G_\varepsilon(c)} \\ &= -\frac{g'(c^2)c^p \ln(g(c^2))}{G_\varepsilon(c)}. \end{aligned}$$

We have that  $G_\varepsilon(c) \rightarrow \varepsilon$ , since  $c \rightarrow \infty$ . Using condition (C<sub>1</sub>) results in (B.7).

Finally, using partial integration, we get

$$k \int_a^c \frac{g'(r^2)}{g(r^2)} G(r) r dr = \frac{k}{2} \ln g(c^2) G(c) - \frac{k}{2} \ln g(a^2) G(a) - \frac{k}{2} \int_a^c c_p r^{p-1} g(r^2) \ln g(r^2) dr. \quad (\text{B.9})$$

The first 2 terms in the above equation are tending again to zero; for the last term we have

$$k \int_a^c r^{p-1} g(r^2) \ln g(r^2) dr \rightarrow 0 \times E[\ln g(\|X\|^2)], \text{ for } a \rightarrow 0.$$

Using (C<sub>2</sub>), (B.8) follows.

□