TILBURG ◆ 🏛 ◆ UNIVERSITY

Tilburg University

The K-Step Spatial Sign Covariance Matrix

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). *The K-Step Spatial Sign Covariance Matrix*. (CentER Discussion Paper; Vol. 2010-41). Econometrics.

Download date: 12. May. 2021

No. 2010–41

THE K-STEP SPATIAL SIGN COVARIANCE MATRIX

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

March 2010

The k-step Spatial Sign Covariance Matrix

C. Croux · C. Dehon · A. Yadine

Abstract The Sign Covariance Matrix is an orthogonal equivariant estimator of multivariate scale. It is often used as an easy-to-compute and highly robust estimator. In this paper we propose a k-step version of the Sign Covariance Matrix, which improves its efficiency while keeping the maximal breakdown point. If k tends to infinity, Tyler's M-estimator is obtained. It turns out that even for very low values of k, one gets almost the same efficiency as Tyler's M-estimator.

Keywords Breakdown point · Multivariate analysis · Principal components · Robust estimation · Spatial signs ·

————————————

Christophe Croux
Faculty of Business and Economics, K.U.Leuven, Leuven, Belgium
Faculty of Economics and Business Administration, Tilburg University, The Netherlands
Tel.: +32-(0)16-326958
Fax: +32-(0)16-326732
E-mail: Christophe.Croux@econ.kuleuven.be

Catherine Dehon
Université libre de Bruxelles, ECARES, and Institut de Recherche en Statistique, CP-114,
Av. F.D. Roosevelt 50, B-1050 Brussels, Belgium
Tel.: +32-(0)2-650-3858
Fax: +32-(0)2-650-4012
E-mail: cdehon@ulb.ac.be

Abdelilah Yadine Université libre de Bruxelles, Institut de Recherche en Statistique, Av. F.D.
Roosevelt 50, B-1050 Brussels, Belgium

1 Introduction

Let $X = \{X_1, X_2, ..., X_n\}$ be a set of n multivariate observations, with each X_i a vectors of dimension p. These observations are a random sample from a distribution F. We assume that the distribution F is elliptically symmetric with center μ and scatter matrix Σ. The density of F can then be written as

$$f_{\mu,\Sigma}(x) = \frac{1}{\det(\Sigma)} g((x - \mu)^t \Sigma^{-1}(x - \mu)), \tag{1}$$

with $\det(\Sigma)$ the determinant of Σ and g a function taking positive values and scaled such that the density f integrates to one. If the second moment of F exists, then Σ is a multiple of the covariance matrix. In this paper we focus on $F = N(0, \Sigma)$, but the obtained results extend to any elliptically symmetric distribution.

It is well known that the sample covariance matrix, while being the most efficient at normal distributions, is very vulnerable in presence of outliers. A very simple robust estimator of Σ is the *Sign Covariance Matrix* (SCM), defined as

$$\hat{\Sigma}_S(X) = \frac{1}{n} \sum_{i=1}^n \frac{(X_i - \hat{\mu}_n)(X_i - \hat{\mu}_n)^t}{\|X_i - \hat{\mu}_n\|^2}. \tag{2}$$

This estimator is nothing else but the usual covariance matrix computed from the spatial signs of the observations, defined as

$$U(X_i) = \frac{(X_i - \hat{\mu}_n)}{\|X_i - \hat{\mu}_n\|}.$$

Since the spatial signs are bounded vectors, with unit norm, the SCM always remains bounded, indicating its robustness. The SCM and its usefulness in different applications are discussed in Locantore et al (1999), Visuri et al (2000), Sirkia et al (2009), Oja (2010), among others. For the location estimator $\hat{\mu}_n$ in (2) we take the L_1-*median*, also called the spatial median, defined as

$$\hat{\mu} = \operatorname*{argmin}_{\mu} \frac{1}{n} \sum_{i=1}^n \|X_i - \mu\|. \tag{3}$$

The L_1-estimator is highly robust, and as solution of a convex optimization problem very fast to compute.

A first contribution of this paper is that we formally show that the breakdown point of the SCM is the highest possible, namely 50%. The breakdown point of an estimator is a standard measure of robustness, and gives the highest fraction of outliers the estimator can withstand. A formal definition is given in Section 2. At first sight one may think that the breakdown point of the SCM should be 100%, since its norm is always bounded by one. However, breakdown may also occur if the estimator implodes, meaning that the smallest eigenvector of the SCM tends to zero. Implosion breakdown is important, since one often inverts scatter matrices, and a full rank of the scatter matrix estimator is desirable.

A major drawback of the SCM is that it is only *orthogonally equivariant*. This means that $\hat{\Sigma}_S(AX) = A\hat{\Sigma}_S(X)A^t$ for any orthogonal matrix A, but not for any non-singular matrix A, which would imply affine equivariance. The lack of affine equivariance also results in a severe loss of statistical power when the true distribution deviates strongly

from sphericity, as was shown in Croux et al (2002). To increase the efficiency of the SCM, we propose to perform a one-step update of the SCM,

$$\hat{\Sigma}_1(X) = \frac{1}{n} \sum_{i=1}^{n} \frac{(X_i - \hat{\mu}_n)(X_i - \hat{\mu}_n)^t}{(X_i - \hat{\mu}_n)^t \hat{\Sigma}_S^{-1}(X_i - \hat{\mu}_n)}. \tag{4}$$

More generally, the k-step SCM estimator is defined by

$$\hat{\Sigma}_k(X) = \frac{1}{n} \sum_{i=1}^{n} \frac{(X_i - \hat{\mu}_n)(X_i - \hat{\mu}_n)^t}{(X_i - \hat{\mu}_n)^t \hat{\Sigma}_{k-1}^{-1}(X_i - \hat{\mu}_n)}, \tag{5}$$

for $k > 1$. We will show that the k-step SCM estimator keeps the breakdown point of the initial SCM, but achieves a higher efficiency at non-spherical distributions.

If k tends to infinity, then the k-step SCM estimator converges to Tyler's M, an affine equivariant estimator. In fact, Tyler (1987) proposed an iterative algorithm to compute his estimator and proved its convergence. One has that $\hat{\Sigma}_k$ corresponds to the kth step approximation of Tyler's M estimator. In this paper, we consider $\hat{\Sigma}_k$ as an estimator in its own right, being orthogonal equivariant and having good robustness properties. While Tyler's M-estimator has a breakdown point decreasing with the dimension, this does not hold for $\hat{\Sigma}_k$. So keeping k fixed results in a high breakdown point, a property one looses by running the iterative algorithm to infinity.

The paper is organized as follows. Section 2 shows that the SCM and its k-step version have the maximal breakdown point property. Section 3 contains results on statistical efficiency. We derive an analytical expression for the asymptotic efficiency in the bivariate Gaussian case. Simulation results are presented in Section 4, and Section 5 contains the conclusions and limitations of this paper.

2 Breakdown point

In this section we compute the breakdown point of the SCM and the k-step version. We prove that they attain the highest possible breakdown point of 50%. The location estimator used in definitions (2) and (5) is the L_1-median. The breakdown point of a multivariate location estimator $\hat{\mu}_n$ at the sample X is defined as

$$\varepsilon^*(\hat{\mu}_n, X) = \min_{1 \le m \le n} \{ \frac{m}{n} : \sup_{X'} \|\hat{\mu}_n(X) - \hat{\mu}_n(X')\| = +\infty \}.$$

The supremum is taken over all possible corrupted collections X' that can be obtained by replacing any m points $X_1, ..., X_m$ of X by arbitrary values $X'_1, ..., X'_m$. Lopuhaä and Rousseeuw (1991) showed that the L_1-median has the largest possible breakdown point of any translation equivariant estimator:

$$\varepsilon^*(\hat{\mu}_n, X) = \frac{[\frac{n+1}{2}]}{n}. \tag{6}$$

The breakdown point of a multivariate scale estimator $\hat{\Sigma}$ at a data set X is defined as the smallest fraction of outliers that can either take the largest eigenvalue over all bounds, or take the smallest eigenvalue arbitrarily close to 0. Denote the ordered

eigenvalues of any matrix Σ by $\lambda_1(\Sigma) \geq \ldots \geq \lambda_p(\Sigma)$. This formal definition of the breakdown point is then

$$\varepsilon^*(\hat{\Sigma}, X) = \min_{1 \leq m \leq n} \{\frac{m}{n} : \sup_{X'} \max\{\lambda_1(\hat{\Sigma}(X')), \ \lambda_p(\hat{\Sigma}(X'))^{-1}\} = +\infty\},$$

where the supremum is taken over the same collections X' as before. We first show that the SCM estimator has the same breakdown point as the L_1-median. A proof of this proposition is given in the Appendix.

Proposition 1 *Let X be a sample of size n such that no $\lfloor n/2 \rfloor + 1$ points are contained in the same hyperplane. Then*

$$\varepsilon^*(\Sigma_S, X) = [\frac{n+1}{2}]/n.$$

The $k-$step estimator keeps the breakdown point of the initial SCM estimator. The proof relies on the following identity

$$\hat{\Sigma}_k(X) = \hat{\Sigma}_{k-1}(X)^{1/2} \hat{\Sigma}_S \left(\hat{\Sigma}_{k-1}^{-1/2}(X)X \right) \hat{\Sigma}_{k-1}(X)^{1/2}, \tag{7}$$

for any $k \geq 1$, and with $\hat{\Sigma}_0 := \hat{\Sigma}_S$. It is not difficult to show, combining (7) and Proposition 1, that $\hat{\Sigma}_k$ will inherit the breakdown point of $\hat{\Sigma}_{k-1}$. By induction we get that

Proposition 2 *Let X be a sample of size n such that no $\lfloor n/2 \rfloor + 1$ points are contained in the same hyperplane. Then*

$$\varepsilon^*(\Sigma_k, X) = [\frac{n+1}{2}]/n, \tag{8}$$

for every $k \leq 1$.

The above result holds for every fixed value of k. When we let k tend to infinity, by iterating formula (5) up to convergence, the breakdown point is not maximal anymore and depends on the dimension. Dumbgen and Tyler (2005) showed that an upper bound for Tyler's M-estimator is given by $1/p$, with p the dimension. This finding provides some theoretical support for the results of Hettmansperger and Randles (2002), who said that Tyler's M-estimator has a 'practical' breakdown point of 50%.

3 Asymptotic efficiencies

In this section we study some aspects of the limiting distribution of the k-step SCM estimators. In particular we compute asymptotic efficiencies at the bivariate normal distribution, but the obtained results can be easily extended to other elliptically symmetric distribution. Let us first define the population quantities that are estimated. For a given distribution F, let

$$T(F) = \underset{\mu}{\arg\min} \, E_F[\|X - \mu\|]$$

be the population version of the spatial median. The population counterpart of the SCM is

$$\Sigma_S(F) := \Sigma_0(F) = E_F \left[\frac{(X - T(F))(X - T(F))^t}{(X - T(F))^t(X - T(F))} \right]$$

and the functional k-step SCM is defined recursively as

$$\Sigma_k(F) = E_F \left[\frac{(X - T(F))(X - T(F))^t}{(X - T(F))^t \Sigma_{k-1}^{-1}(F)(X - T(F))} \right], \tag{9}$$

for $k \geq 1$. It is worth mentioning that also the SCM verifies (9), with $\Sigma_{-1}(F) = I_p$, the identity matrix. In the limit, for $k \to \infty$, we get the population version of Tyler's estimator, as solution of the equation

$$\Sigma_\infty(F) = E_F \left[\frac{(X - T(F))(X - T(F))^t}{(X - T(F))^t \Sigma_\infty^{-1}(F)(X - T(F))} \right], \tag{10}$$

where the constraint $\text{Trace}(\Sigma_\infty(F)) = p$ is imposed to ensure a unique solution. Alternative, one could use the constraint $\det(\Sigma_\infty(F)) = 1$, see Paindaveine (2008).

A scatter matrix Σ can always be decomposed as

$$\Sigma = \lambda U D U^t, \tag{11}$$

where U contains the eigenvectors of Σ, D is a diagonal matrix containing the scaled eigenvalues of Σ, and $\lambda = \det(\Sigma)$. The *size* of the scatter matrix is then determined by λ, the matrix UDU^t is the *shape matrix* while U determines the *orientation* of the scatter matrix (Bensmail and Celeux, 1996). Since all estimators we consider are at least orthogonal equivariant, we may assume without lost of generality that the scatter matrix Σ of the distribution F is diagonal. Using symmetry arguments, it is then immediate to check that $\Sigma_k(F)$ is also diagonal, for every $k \geq 0$. We thus have that the SCM and its k-step versions have the same orientation as the scatter matrix Σ, but the shape and the size will be different.

To compare the precision of the different k-step estimators in a meaningful way, we compare the asymptotic variances of their eigenvectors, who are all estimating the same quantity at elliptical model distributions. We only present results for the bivariate case ($p = 2$), to facilitate the exposition, and because it allows for an explicit expression of the asymptotic variances. The model distribution is then denoted by F_γ, having scatter matrix

$$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & \gamma \end{pmatrix}, \quad 0 < \gamma < 1. \tag{12}$$

Since eigenvectors have norm one, and are orthogonal, it suffices to study the distribution of the first component of the first eigenvector of $\hat{\Sigma}_k$ (for $p = 2$).

3.1 Asymptotic Variance

For $p = 2$, the asymptotic variance of the first component of the first eigenvector of $\hat{\Sigma}_k$ at F_γ equals

$$ASV_k(F_\gamma) = \frac{1}{(\lambda_{k,2} - \lambda_{k,1})^2} E[IF(X, \Sigma_{k,12}; F_\gamma)^2], \tag{13}$$

see Croux and Haesbroeck (2000). The computation of the influence function of the off-diagonal element of $\hat{\Sigma}_k$ is not difficult. Using standard influence function techniques, we

get that the influence function of an off-diagonal element of the k-step SCM estimator at $x = (x_1, x_2)^t$ equals

$$IF(x, \Sigma_{k,12}; F_\gamma) = \frac{x_1 x_2}{x^t \Sigma_{k-1}^{-1}(F_\gamma)x} + c_{k-1} IF(x, \Sigma_{k-1,12}; F_\gamma). \qquad (14)$$

We have that

$$c_s = 2 E_{F_\gamma} \left[\frac{X_1^2 X_2^2}{(X^t \Sigma_s(F_\gamma)^{-1} X)^2} \right] \frac{1}{\lambda_{s,1} \lambda_{s,2}},$$

and

$$\Sigma_s(F_\gamma) = \begin{pmatrix} \lambda_{s,1} & 0 \\ 0 & \lambda_{s,1} \end{pmatrix}, \qquad (15)$$

for every $s \geq 0$. Hence the influence function can be computed recursively from (14). The IF for $k = 0$, corresponding to the SCM, is given by

$$IF(x, \Sigma_{0,12}; F_\gamma) = \frac{x_1 x_2}{\|x\|^2},$$

and still verifies (14) if we set $c_{-1} = 0$, and $\lambda_{-1,1} = \lambda_{-1,2} = 1$, such that $\Sigma_{-1}(F_\gamma) = I$.

Working out the recursive relation (14), combined with (13), results in

$$ASV_k(F_\gamma) = \frac{1}{(\lambda_{k,2} - \lambda_{k,1})^2} a_k^t B_k a_k \qquad (16)$$

with $a_k = (1, c_{k-1}, c_{k-1} c_{k-2}, \ldots, c_{k-1} c_{k-2} \cdots c_{-1})^t$ a vector of length $k+1$, and B_k a square matrix of size $k+1$, with elements

$$(B_k)_{j_1, j_2} = E_{F_\gamma} \left[\frac{X_1^2 X_2^2}{(X^t \Sigma_{k-j_1}(F_\gamma)^{-1} X)(X^t \Sigma_{k-j_2}(F_\gamma)^{-1} X)} \right]$$

for $1 \leq j_1, j_2 \leq k+1$, and for every $k \geq 0$. The vector a_k, the matrix B_k and the scalar c_k all depend on γ.

For the SCM, so $k = 0$, the asymptotic variance was computed by Croux et al (2002). At the normal distribution they obtain

$$ASV_0(F_\gamma) = \frac{\sqrt{\gamma}}{2(1 - \sqrt{\gamma})^2}.$$

Below we show how analytic expressions for the asymptotic variance for $k \geq 1$ can be obtained.

3.2 Calculus

In this section we provide some lemmas allowing us to compute the ASV_k in (16). For computing the constant vector a_k and the matrix B_k in (16), we need to evaluate quantities of the form

$$\Psi(a, b) = E_0 \left[\frac{X_1^2 X_2^2}{(X_1^2 + a X_2^2)(X_1^2 + b X_2^2)} \right],$$

with $a, b > 0$. Furthermore, we also need the eigenvalues of $\Sigma_k(F_\gamma)$. The first lemma gives an expression for these eigenvalues. The second lemma gives two formulas allowing to compute analytically the asymptotic variances at the normal distribution. The results can be obtained by straightforward calculus, and details can be found in the Ph.D. manuscript of Yadine (2006).

Lemma 1 *Let $\lambda_{k,1}$ and $\lambda_{k,2}$ be the eigenvalues of $\Sigma_k(F_\gamma)$, wiht $0 < \gamma < 1$. Then the following recursion relations hold:*

$$\lambda_{k,1} = \lambda_{k-1,1}\left(1 - \gamma\frac{\lambda_{k-1,1}}{\lambda_{k-1,2}}\varphi(\gamma\frac{\lambda_{k-1,1}}{\lambda_{k-1,2}})\right)$$

and

$$\lambda_{k,2} = \gamma\lambda_{k-1,1}\varphi(\gamma\frac{\lambda_{k-1,1}}{\lambda_{k-1,2}}),$$

with

$$\varphi(a) = E_{F_1}\left[\frac{X_1^2}{X_1^2 + aX_2^2}\right],$$

for any $a > 0$, and $k \geq 0$. Recall that $\lambda_{-1,1} = \lambda_{-1,2} = 1$.

Lemma 2 *For F_γ a normal distribution, we have*

$$\varphi(a) = \frac{1}{\sqrt{a}(1 + \sqrt{a})}$$

and

$$\Psi(a,b) = \frac{1}{(\sqrt{a} + \sqrt{b})(1 + \sqrt{a})(1 + \sqrt{b})},$$

for all $a, b > 0$.

3.3 Efficiency of k-step estimators for bivariate Gaussian distributions

At the normal distribution, the most efficient estimator for Σ is the sample covariance matrix, being the maximum likelihood estimator. It is not difficult to verify that

$$ASV_{ML}(F_\gamma) = \frac{\gamma}{(1-\gamma)^2}.$$

We then define the efficiency of the k-step SCM as

$$\text{Eff}_k(F_\gamma) = \frac{ASV_{ML}(F_\gamma)}{ASV_k(F_\gamma)} \tag{17}$$

for $k \geq 0$. Using the results of subsections 3.1 and 3.2 the efficiency can be computed analytically at Gaussian distributions. For example, for the one-step SCM we have

$$\text{Eff}_1(F_\gamma) = \frac{2\gamma^{1/4}(1 - \sqrt{\gamma}\gamma^{1/4})^2(1 + \gamma^{1/4})^2}{(1-\gamma)^2\left[(1 + \gamma^{1/4})^2 + 5\gamma^{1/4}\right]}.$$

In Figure 1 we plot the efficiencies of the k-step SCM estimator for different values of k, as a function of γ, where γ is the ratio between the smallest and the largest eigenvalue of the model covariance matrix. We see that the efficiencies converges very quickly to the value 0.5, when k tends to infinity. This limiting value corresponds to the efficiency of Tyler's M-estimator, being $p/(p+2) = 0.5$ (Tyler 1987, Frahm 2009). This values does not depend on the value of γ since Tyler's M shape is an affine equivariant estimator (see also Ollila et al 2002). For the k-step SCM we observe a loss of efficiency if we deviate strongly from spherical distribution, i.e. when γ is close to zero and where Σ_γ is close to non-singular The surprising finding is that already for small values of k, say $k = 3$, there is almost no difference anymore between the efficiency of Tyler's M and the k-step SCM, over almost the complete range of possible values for γ. For instance, for $\gamma = 0.001$, $\text{Eff}_3(F_\gamma)$ still equals 0.44. Hence the loss in efficiency of using the $k = 3$ version instead of the fully iterated Tyler's M-estimator is only important for extremely small values of γ.
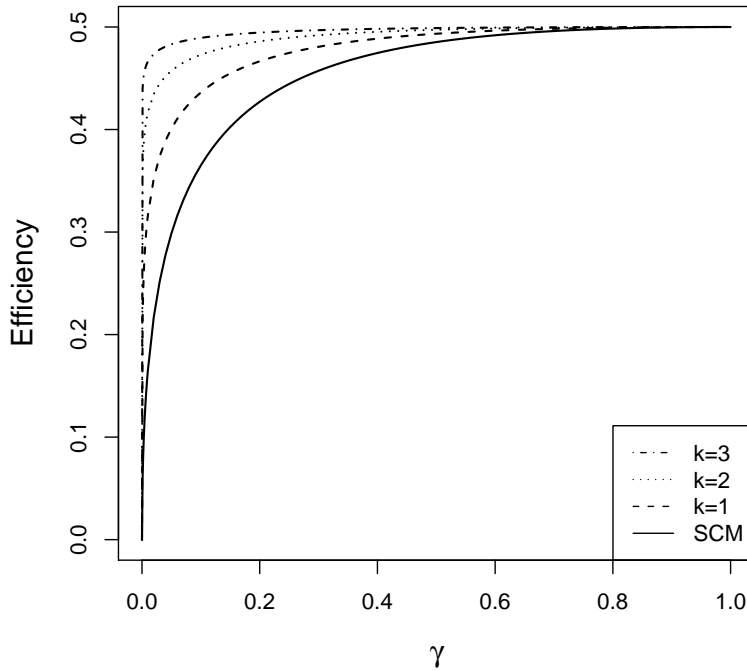
Fig. 1 Efficiencies of the SCM and its k-step version at the bivariate normal distribution, as a function of γ, the ratio between the smallest and the largest eigenvalue of the model covariance matrix. The efficiency of Tyler's M-estimator is constant and equal to 0.5. We see that, for increasing k, there is fast convergence to this value.

4 Simulations

In this section we perform a modest simulation study to confirm the asymptotic efficiencies obtained in the previous Section. We generate $m = 10000$ samples from a bivariate normal distribution with covariance matrix Σ_γ, as in (12).

For every generated sample, we compute the first eigenvector $\hat{v}_1^j$ of the multivariate scale estimator, and summarize the outcomes by the

$$MSE(\hat{v_1}) = \frac{1}{m} \sum_{j=1}^{m} \left(\arccos\{|v_1^t \hat{v}_1^j|\} \right)^2 ,$$

with $v_1 = (1,0)^t$ the true eigenvector. This MSE converges to the asymptotic variance (16) introduced in Section 3, see Croux et al (2002). Dividing the MSE obtained using the sample covariance matrix by the MSE resulting from the k-step SCM yields the finite sample counterpart of (17). Finite sample efficiencies are obtained for the SCM, its k-step version, with $k = 1, 2, 3, 4$, and for the fully iterated version ($k = \infty$), i.e. Tyler's M-estimator. We consider sample sizes $n = 20, 50, 100, 200$. Furthermore, we

Table 1 Finite sample efficiencies, over 10000 simulation runs, of the first eigenvector estimates, for samples of size $n = 20, 50, 100, 200$, generated from a bivariate normal distribution with mean 0 and covariance matrix $\Sigma = diag(1, \gamma)$, for $\gamma = 0.1$ and $\gamma = 0.7$. We considered the SCM, and its $k - step$ version. The column with $k = \infty$ corresponds to Tyler's M-estimator.

	n	SCM	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = \infty$
	20	0.338	0.382	0.404	0.411	0.414	0.415
	50	0.359	0.421	0.453	0.467	0.473	0.476
$\gamma = 0.1$	100	0.365	0.433	0.467	0.482	0.489	0.492
	200	0.363	0.432	0.467	0.483	0.489	0.493
	∞	0.365	0.436	0.473	0.489	0.496	0.500
	n	SCM	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = \infty$
	20	0.791	0.790	0.789	0.788	0.788	0.786
	50	0.655	0.655	0.655	0.654	0.654	0.653
$\gamma = 0.7$	100	0.523	0.523	0.522	0.522	0.521	0.521
	200	0.427	0.427	0.427	0.427	0.427	0.427
	∞	0.496	0.498	0.499	0.500	0.500	0.500

take $\gamma = 0.1$, where the deviation of sphericity is very strong, and $\gamma = 0.7$, where we are closer to a spherical model distribution. The standard errors around the results reported in Table 1 are at most 0.02.

We can see from Table 1 that the finite sample efficiencies converges to their asymptotic counterparts ($n = \infty$). Note that for $\gamma = 0.7$ and $n = 20, 50$ the finite sample efficiencies are considerably higher than expected. When we are closer to the spherical model distribution ($\gamma = 0.7$), there appears to be no significant difference between the efficiency of the different k-step estimators and Tyler's M.

5 Conclusion

The use of k-step estimators is widespread in the statistical literature (see Hallin et al, 2006 for a recent contribution). Starting from an initial estimate, one makes sequential updates of the estimator, resulting in a sequence of k-step estimators. The initial estimator is consistent, and the k-step versions increase the efficiency. Most often, the initial estimator is easy to compute, and the k-step updates come with almost no additional computational effort. As such, Taskinen et al (2009) use Tyler's M-estimator, or its symmetrized version, as a starting estimator, and increase its efficiency by carrying out k-steps. However, since this starting estimator has a low breakdown point, the same holds for the k-step improvement. Our approach is different, we start from a high breakdown estimator and all further k-step SCM inherit this high breakdown point.

As we showed in Section 3, the efficiency of the k-step SCM estimator increases quickly to $p/(p + 2)$, the efficiency of Tyler's M-estimator. To increase further the efficiency, additional weights can be added, resulting in

$$\hat{\Sigma}_k(X) = \frac{1}{n} \sum_{i=1}^{n} \frac{(X_i - \hat{\mu}_n)(X_i - \hat{\mu}_n)^t}{(X_i - \hat{\mu}_n)^t \hat{\Sigma}_{k-1}^{-1}(X_i - \hat{\mu}_n)} w((X_i - \hat{\mu}_n)^t \hat{\Sigma}_{k-1}^{-1}(X_i - \hat{\mu}_n))$$

for $k \geq 1$. Kent and Tyler (1991) show that this sequence of estimators converges to an M-estimator of shape if the weight function w is strictly increasing. Using a bounded weight function will retain the breakdown point of the initial SCM. Assume that we take a weight function w that is bounded, strictly increasing, and such that $w(s) = s$

for $s \leq c$. For c and k large enough, this results in a high breakdown point estimator having a Gaussian efficiency arbitrarily close to 100%.

Our paper has several limitations. In this paper we keep the location estimator fixed. One could consider to update the location estimator as well, as in Hettmansperger and Randles (2002). We claim that the k-step spatial median retains the maximal breakdown point of the L_1-median, and that updating the location estimator will not change the efficiency of the k-step SCM estimator, neither its breakdown point.

Another limitation is that the k-step SCM estimator is only consistently estimating the orientation of the scatter matrix. However, this is the most crucial part, since the eigenvalues of Σ can be estimated afterwards by applying an efficient and robust scale estimator to the data projected on the respective eigenvectors. Finally, we only measure robustness by means of the breakdown point. It would be of interest to consider also the maxbias curve, as was done by Croux and Rousseeuw (1994) for univariate k-step M-estimators.

Appendix

Proof of proposition 1. Denote $\hat{\Sigma} = \hat{\Sigma}_S$. We first show that $\varepsilon^*(\hat{\Sigma}, X) \geq \lfloor \frac{n+1}{2} \rfloor / n$. Let $m < \lfloor \frac{n+1}{2} \rfloor$ and replace m observations of X to get $X' = \{x_1, \ldots, x_{n-m}, x'_{n-m+1}, \ldots, x'_n\}$. Without loss of generality, we assume that the first $n-m$ values of X remain unchanged. Let $\hat{\mu}'$ be the L_1-median computed from X'. We need to show that there are existing constants $\delta_X, N_X > 0$ such that $\lambda_{\min}(\hat{\Sigma}(X')) > \delta_X$ and $\lambda_{\max}(\hat{\Sigma}(X') < N_X$. We have

$$
\begin{aligned}
\lambda_{\max}(\hat{\Sigma}(X')) &= \sup_{\|u\|=1} u^t \hat{\Sigma}(X') u \\
&= \sup_{\|u\|=1} \frac{1}{n} \sum_{i=1}^n \frac{u^t (x'_i - \hat{\mu}')(x'_i - \hat{\mu}')^t u}{\|x'_i - \hat{\mu}'\|^2} \\
&= \sup_{\|u\|=1} \frac{1}{n} \sum_{i=1}^n \frac{(u^t(x'_i - \hat{\mu}'))^2}{\|x'_i - \hat{\mu}'\|^2} \\
&\leq \sup_{\|u\|=1} \frac{1}{n} \sum_{i=1}^n \frac{\|u\|^2 \|x'_i - \hat{\mu}'\|^2}{\|x'_i - \hat{\mu}'\|^2} = 1 = N_X.
\end{aligned}
$$

For every subset J of size $n-m$ from $\{1, \ldots, n\}$, define $\eta_J = \max_{i \in J} \sum_{i \in J} d^2(x_i, H_J)$, where H_J is the hyperplane minimizing $\sum_{i \in J} d^2(x_i, H)$ over all possible hyperplanes H, and $d(x, H)$ is the Euclidean distance between an observation x and a hyperplane H. Define then $\eta_X = \min_J \eta_J$. Since $n - m \geq \lfloor \frac{n+1}{2} \rfloor$, and no $\lfloor \frac{n+1}{2} \rfloor$ original observations are on the same hyperplane, we have that $\eta_X > 0$. Furthermore, since there is not yet breakdown of the L_1-median, there exists a constant $\bar{M}$ such that $\|\hat{\mu}'\| \leq \bar{M}_X$. Finally, let $M_X = \max_{1 \leq i \leq n} \|x_i\|$, and $\delta_X = 0.5 \eta_X / (n(M_X + \bar{M}_X)^2)$. Using all these

notations, we obtain

$$\lambda_{\min}(\hat{\Sigma}, X') = \min_{\|u\|=1} u^t \hat{\Sigma}(X') u$$

$$= \min_{\|u\|=1} \frac{1}{n} \sum_{i=1}^{n} \frac{u^t(x_i' - \hat{\mu}')(x_i' - \hat{\mu}')^t u}{\|x_i' - \hat{\mu}'\|^2}$$

$$= \min_{\|u\|=1} \frac{1}{n} \sum_{i=1}^{n} \left[\frac{u^t(x_i' - \hat{\mu}')}{\|x_i' - \hat{\mu}'\|^2} \right]^2$$

$$\geq \min_{\|u\|=1} \frac{1}{n} \sum_{i=1}^{n-m} \left[\frac{u^t(x_i - \hat{\mu}')}{\|x_i' - \hat{\mu}'\|^2} \right]^2$$

$$\geq \min_{\|u\|=1} \sum_{i=1}^{n-m} \left[u^t(x_i' - \hat{\mu}') \right]^2 \frac{1}{n(M_X + \bar{M}_X)^2}$$

$$\geq \sum_{i=1}^{n-m} d^2(x_i, H_{\{1,\dots,n-m\}}) \frac{1}{n(M_X + \bar{M}_X)^2}$$

$$\geq \eta_X \frac{1}{n(M_X + \bar{M}_X)^2} = \delta_X.$$

Now we will show that $\varepsilon^*(\Sigma_S, X) \leq [\frac{n+1}{2}]/n$. We replace $m = [\frac{n+1}{2}]$ observations of X to a constant vector $\tilde{x}$. Since the L_1-median has the multivariate exact fit property (Martin, Maronna, Yohai, 2006), we have $\hat{\mu}' = \tilde{x}$. We take $\tilde{x} > M_X$, such that $\|x_i - \tilde{x}\| \geq \|\tilde{x}\| - \|x_i\| \geq \|\tilde{x}\| - M_X$, for every index i. Let $\tilde{u}$ a vector of norm one, orthogonal to $\tilde{x}$, then

$$\lambda_{\inf}(\hat{\Sigma}(X')) = \inf_{\|u\|=1} \frac{1}{n} \sum_{i=1}^{n-m-1} \left[\frac{u^t(x_i - \tilde{x})}{\|x_i - \tilde{x}\|} \right]^2$$

$$\leq \frac{1}{n} \sum_{i=1}^{n-m} \frac{(\tilde{u}^t x_i)^2}{\|x_i - \tilde{x}\|^2}$$

$$\leq \frac{1}{n} \sum_{i=1}^{n-m} \frac{\|x_i\|^2}{\|x_i - \tilde{x}\|^2}$$

$$\leq M_X^2 \frac{1}{n} \sum_{i=1}^{n-m} \frac{1}{\|x_i - \tilde{x}\|^2}$$

$$\leq \frac{n-m}{n} \frac{M_X^2}{(\|\tilde{x}\| - M_X)^2},$$

which tends to 0 when $\|\tilde{x}\|$ is tending to infinity. $\qquad\square$

References

1. Bensmail, H. and Celeux, G., Regularized Gaussian Discriminant Analysis through Eigenvalue Decomposition, Journal of the American Statistical Association, 91, 1743-1749 (1996)

2. Croux, C., and Haesbroeck, G., Principal Component Analysis based on Robust Estimators of the Covariance or Correlation Matrix: Influence Functions and Efficiencies, Biometrika, 87, 603-618 (2000)
3. Croux, C., Ollila, E., and Oja, H., Sign and Rank Covariance Matrices: Statistical Properties and Application to Principal Components Analysis, Statistical Data Analysis Based on the L1-Norm and Related Methods, pp 257-271. Ed. Y. Dodge, Basel: Birkhauser (2002)
4. Dümbgen, L. and Tyler, D.E., On the Breakdown Properties of some Multivariate M-Functionals, Scandinavian Journal of Statistics, 32, 247-264 (2005)
5. Frahm, G., Asymptotic distiributions of robust shape matrices and scale, Journal of Multivariate Analysis, 100, 1329-1337 (2009)
6. Hallin, M., Oja, H., and Paindaveine, D., Semiparametrically efficient rank-based inference for shape. II. Optimal R-estimation of shape, Annals of Statisitcs, 34, 2757-2789 (2006)
7. Kent, J.T., and Tyler, D.E., Redescending M-estimates of multivariate location and scatter, The Annals of Statistics, 19, 2102-2119 (1991)
8. Locantore N., Marron J. S., Simpson D. G., Tripoli N., Zhang J. T., Cohen K. L., Robust principal components for functional data, Test, 8, 1-28 (1999)
9. Marden, J.I. , Some Robust Estimates of Principal Components, Statistics and Probability Letters, 43, 349-359 (1999)
10. Maronna, R., Martin, D. and Yohai, V., Robust Statistics. Wiley, New York, NY. (2006)
11. Oja, H., Multivariate Nonparametric Methods with R. An approach based on spatial signs and ranks, Springer, in print (2010)
12. Ollila, E., Heetmansperger, T.P., and Oja, H. Affine Equivariant Multivariate Sign Methods, unpublished manuscript (2002)
13. Rousseeuw, P.J., and Croux, C., The Bias of k-step M-estimators, Statistics and Probability Letters, 20, 411-420 (1994)
14. Paindaveine, D., A canonical definition of shape, Statistics and Probability Letters, 78, 2240-2247 (2008)
15. Sirkia, S., Taskinen, S., Oja, H. and Tyler, D., Tests and estimates of shape based on spatial signs and ranks, Journal of Nonparametric Statistics, 21, 15 -176, (2009)
16. Taskinen, S., Sirkiä, S., and Oja, H., k-step Shape Estimators based on Spatial Signs and Ranks, manuscipt (2009)
17. Tyler, D.E. , A distribution-free M-estimator of multivariate scatter, Annals of Statistics, 15, 234-251 (1987)
18. Visuri, S., Ollila, E., Koivunen, V., Möttönen, J., and Oja, H., Affine equivariant multivariate rank methods, Journal of Statistical Planning and Inference, 114, 161-185 (2002)
19. Yadine, A., Robustness and Efficiency of Multivariate Scatter Estimators", Ph.D. dissertation, Université Libre de Bruxelles, Bruxelles (2006)