A robust and efficient adaptive reweighted estimator of multivariate location and scatter

Daniel Gervini [1]
Department of Biostatistics, University of Zurich
Sumatrastrasse 30
Zurich, CH-8006, Switzerland

July 30, 2002

[1]The research of this paper was done while the author was visiting the University of Illinois at Urbana-Champaign, USA

Abstract

This article proposes a reweighted estimator of multivariate location and scatter, with weights adaptively computed from the data. Its breakdown point and asymptotic behavior under elliptical distributions are established. This adaptive estimator is able to attain simultaneously the maximum possible breakdown point for affine equivariant estimators and full asymptotic efficiency at the multivariate normal distribution. For the special case of hard-rejection weights and the MCD as initial estimator, it is shown to be more efficient than its non-adaptive counterpart for a broad range of heavy-tailed elliptical distributions. A Monte Carlo study shows that the adaptive estimator is as robust as its non-adaptive relative for several types of bias-inducing contaminations, while it is remarkably more efficient under normality for sample sizes as small as 200.
AMS subject classification: 62H12, 62F35.
Key words and phrases: efficient estimation, high breakdown point, minimum covariance determinant (MCD), outlier detection, robust estimation.

1 Introduction

The properties of most multivariate statistical methods, such as principal component analysis, factor analysis, discrimination and classification, depend strongly on the estimators of location and scatter they are based on. The estimators most commonly used, the sample mean and covariance, are efficient when the data are normally distributed but are extremely sensitive to outliers and elliptical distributions with heavy tails. Since visual detection of outliers in high dimensions is very difficult, the use of good robust multivariate estimators should constitute an essential part of the statistical analysis of this kind of problems.

Several robust estimators of multivariate location and scatter have been proposed since Maronna's pioneering paper on multivariate M-estimation [14]. Among others, we can mention the minimum volume ellipsoid (MVE) and the minimum covariance determinant (MCD) estimators by Rousseeuw [20], multivariate S-estimators by Davies [3] and Lopuhaä [11], and the Stahel-Donoho (S-D) estimator proposed in [26, 5] and revisited in [15]. A thorough overview of robust multivariate estimation can be found in the article by Maronna and Yohai [16]. All these methods offer more resistance to outliers than the classical sample mean and covariance, but they also have some shortcomings. Maronna's M-estimators are computationally the simplest but posses a low breakdown point, which is bounded above by $1/(p+1)$ as shown in [5]. The MVE has a slow $n^{-1/3}$ rate of convergence (see [4]) and it is hard to compute in high dimensions. The MCD, although $n^{1/2}$-consistent (see [1, 2]), has a very low asymptotic efficiency under normality if one requires high breakdown point. S-estimators defined with a smooth ρ function are also $n^{1/2}$-consistent and can be very efficient under the normal model in high dimensions, but if the function ρ is not properly chosen they can be extremely sensitive to outliers, as pointed out by Rocke [17]. S-D estimators require a formidable computational effort and, although $n^{1/2}$-consistent, their asymptotic distribution is as yet unknown and appears to be non-normal.

Besides high outlier resistance, if robust multivariate estimators are to be of practical use in statistical inference they should offer a reasonable efficiency under the normal model and a manageable asymptotic distribution. Certainly MVE, MCD and S-D are not in this category. Overall, if one wants to take care of both robustness and efficiency considerations, the best choice seems to be a two-stage procedure. First a highly robust but perhaps inefficient estimator is computed, which is used as staring point to find a local solution of an S-estimating equation as in Woodruff and Rocke [29] or for detecting outliers and computing the sample mean and covariance of the "cleaned" data set as in Rousseeuw and van Zomeren [24]. Specifically, the latter proposal consists of discarding those observations whose Mahalanobis distances exceed a certain fix threshold value. In the past, the MVE was commonly used as initial estimator for these procedures. But recently Rousseeuw and van Driessen [23] have proposed an algorithm for the MCD that, although it does not guarantee that the exact estimator is found, it is faster and more accurate than previously existing algorithms, even for very large data sets. This fact, added to its $n^{-1/2}$ rate of convergence, seems to point to the MCD as the current best choice for initial estimator of a two-step procedure.

In the context of linear regression, many estimators have been proposed that aim to reconcile high efficiency and robustness. Typically, these methods are also two-stage procedures. The best known proposals consist of one-step estimators computed by reweighting [19] or Newton-Raphson steps [10, 25], and estimators that minimize an efficient objective function on the second stage (MM-estimators [30] and τ-estimators [31]). A different approach is the cross-checking method proposed by He [9]. All these estimators can attain the maximum breakdown point and arbitrarily high efficiency. However, the gains in efficiency come at the price of a larger bias, as Rousseeuw well pointed out in [21]. The reason is that all these methods are non-adaptive, and higher efficiency can only be obtained by increasing tunning parameters, which in turn affects the bias.

What we propose in this paper is essentially an improvement over Rousseeuw and van Zomeren's proposal [24]. It consists of a reweighted one-step estimator that uses adaptive threshold values. This adaptive reweighting scheme is able to maintain the outlier resistance of the initial estimator in breakdown and bias and, at the same time, attain (100%) efficiency at the normal distribution. This kind of adaptive reweighting was first proposed in [7] for the linear regression model. What follows is an extension of that idea to the multivariate location-scatter estimation problem.

2 The estimator

Given a sample $x_1, \ldots, x_n$ in $\mathcal{R}^p$ and initial robust estimators of location and scatter (t_{0n}, V_{0n}), consider the Mahalanobis distances

$$d_i := d\left(x_i, t_{0n}, V_{0n}\right) = \left\{\left(x_i - t_{0n}\right)' V_{0n}^{-1} \left(x_i - t_{0n}\right)\right\}^{1/2}.$$

An outlier will typically have a larger Mahalanobis distance than a "good" observation. If one assumes a normal distribution, d_i^2 is approximately χ_p^2 distributed and it is reasonable to suspect of those observations with, for instance, $d_i^2 \geq \chi_{p,.975}^2$. What Rousseeuw and van Zomeren propose in [24] is to skip those outlying observations and compute the sample mean and covariance matrix of the rest of the data, obtaining in this way new estimators (t_{1n}, V_{1n}). This reweighting step is known to improve the efficiency of the initial estimator while retaining (most of) its robustness. However, the threshold value $\chi_{p,.975}^2$ is an arbitrary number. For large data sets, that are becoming more and more frequent nowadays, a considerable number of observations will be discarded even if they do follow the normal model. One way to avoid this problem is to increase the threshold value to another arbitrary fix number, but this will affect the bias of the reweighted estimator. A better alternative is to use an adaptive threshold value that increases with n if the data is "clean" but remains bounded if there are outliers in the sample.

We propose one method of constructing such adaptive threshold values. Let

$$G_n\left(u\right) = \frac{1}{n} \sum_{i=1}^{n} I\left(d^2\left(x_i, t_{0n}, V_{0n}\right) \leq u\right)$$

be the empirical distribution of the squared Mahalanobis distances. Let $G_p\left(u\right)$ be the χ_p^2 distribution function. For a normally distributed sample we expect G_n to converge to

G_p. Therefore a way to detect outliers is to compare the tails of G_n with the tails of G_p. If $\eta = \chi^2_{p,1-\alpha}$ for a certain small α, say $\alpha = .025$, define

$$\alpha_n = \sup_{u \geq \eta} \{G_p(u) - G_n(u)\}^+ \tag{1}$$

where $\{\cdot\}^+$ indicates the positive part. This α_n can be regarded as a measure of outliers in the sample. Note that we only take into account positive differences in (1) because a negative difference would not indicate presence of outliers. If $d^2_{(i)}$ denotes the i-th order statistic of the squared Mahalanobis distances and $i_0 = \max\left\{i : d^2_{(i)} < \eta\right\}$, then (1) comes down to

$$\alpha_n = \max_{i > i_0} \left\{ G_p\left(d^2_{(i)}\right) - \frac{i-1}{n} \right\}^+ .$$

Those observations corresponding to the largest $\lfloor \alpha_n n \rfloor$ distances are considered outliers and eliminated in the reweighting step (here $\lfloor a \rfloor$ is the largest integer that is less than or equal to a). The cut-off value is then defined as

$$c_n = G_n^{-1}(1 - \alpha_n), \tag{2}$$

where as usual $G_n^{-1}(u) = \min\{s : G_n(s) \geq u\}$. Note that $c_n = d^2_{(i_n)}$ with $i_n = n - \lfloor \alpha_n n \rfloor$ and that $i_n > i_0$ as a consequence of the definition of α_n. Hence $c_n > \eta$.

To define the reweighted estimator we will use weights of the form

$$w_{in} = w\left(\frac{d^2(x_i, t_{0n}, V_{0n})}{c_n}\right) \tag{3}$$

with a weight function that satisfies

(W) $w : [0, \infty) \to [0, 1]$ is non-increasing, $w(0) = 1$, $w(u) > 0$ for $u \in [0, 1)$ and $w(u) = 0$ for $u \in [1, \infty)$.

The simplest choice among those functions satisfying (W) is the hard-rejection function $w(u) = I(u < 1)$, which is the one most commonly used in practice. Once the weights (3) are computed, the one-step reweighted estimators are defined as

$$t_{1n} = \sum_{i=1}^n w_{in} x_i \Big/ \sum_{i=1}^n w_{in} \tag{4}$$

$$V_{1n} = \sum_{i=1}^n w_{in} (x_i - t_{1n})(x_i - t_{1n})' \Big/ \sum_{i=1}^n w_{in}. \tag{5}$$

It is clear that under appropriate conditions, the threshold values (2) will tend to infinity under the multivariate normal model and then (4) and (5) will be asymptotically equivalent to the common sample mean and covariance, and thus will attain full asymptotic efficiency. This result is formally established in Section 4. What is less obvious, though, is that this adaptive reweighting scheme is able to maintain the breakdown point of the initial estimator. This issue is addressed in the next section.

3 Global robustness

3.1 Finite-sample breakdown point

A global measure of robustness of an estimator is given by the finite-sample replacement breakdown point (BDP) introduced in [6]. Roughly speaking, it is the smallest fraction of outliers that can spoil the estimator completely. Formally, it is defined as follows. Let $\mathbf{X} = (x_1, \ldots, x_n)$ be a sample in $\mathcal{R}^p$ and $t_n(\mathbf{X})$ and $V_n(\mathbf{X})$ the corresponding location and scatter estimators based on $\mathbf{X}$. Given $m \leq n$, let $\mathcal{X}_m$ be the set of all corrupted $n \times p$ matrices $\mathbf{X}^*$ that are obtained after replacing m data points (columns) of $\mathbf{X}$ by arbitrary vectors. Then the BDP of the location estimator t_n at the sample $\mathbf{X}$ is defined as

$$\varepsilon^*(t_n, \mathbf{X}) = \min\left\{m \in \{1, \ldots, n\} : \sup_{\mathbf{X}^* \in \mathcal{X}_m} \|t_n(\mathbf{X}^*)\| = \infty\right\}/n.$$

For a scatter estimator we also want to avoid implosion. Then if $\operatorname{cond}(V_n(\mathbf{X}))$ denotes the condition number of $V_n(\mathbf{X})$, which is the ratio between the largest and the smallest eigenvalue, the BDP of V_n at $\mathbf{X}$ is given by

$$\varepsilon^*(V_n, \mathbf{X}) = \min\left\{m \in \{1, \ldots, n\} : \sup_{\mathbf{X}^* \in \mathcal{X}_m} \operatorname{cond}(V_n(\mathbf{X}^*)) = \infty\right\}/n,$$

where we set $0/0 = \infty$ for completeness of the definition.

Under the following conditions Theorem 1 shows that the BDPs of the adaptive reweighted estimators are not less than those of the initial estimators:

(GP) $n \geq p + 1$ and $\mathbf{X}$ is in general position (that is, a hyperplane of dimension less than p cannot contain more than p points of the sample).

(E) There exist $c < \sqrt{\eta}$ and $k \geq p + 1$ such that

$$\#\left\{i : d^2(x_i, t_{0n}(\mathbf{X}), V_{0n}(\mathbf{X})) \leq c^2\right\} \geq k$$

for every sample $\mathbf{X}$.

Theorem 1 *Let t_{1n} and V_{1n} be the one-step estimators defined by (4) and (5). Under conditions (W), (GP) and (E), if*

$$\varepsilon_0^*(\mathbf{X}) = \min\left\{\varepsilon^*(t_{0n}, \mathbf{X}), \varepsilon^*(V_{0n}, \mathbf{X}), (k-p)/n\right\}$$

then $\min\left\{\varepsilon^*(t_{1n}, \mathbf{X}), \varepsilon^*(V_{1n}, \mathbf{X})\right\} \geq \varepsilon_0^*(\mathbf{X})$.

Throughout this paper we will consider only affine equivariant estimators. That is, for any $b \in \mathcal{R}^p$ and any non-singular $p \times p$ matrix A,

$$\begin{aligned} t_n(A\mathbf{X} + b) &= At_n(\mathbf{X}) + b \\ V_n(A\mathbf{X} + b) &= AV_n(\mathbf{X})A'. \end{aligned}$$

Under condition (GP), Theorem 6 in [3] shows that the BDP of affine equivariant location and scatter estimators can be at most $\lfloor (n-p+1)/2 \rfloor$. When $k = \lfloor (n+p+1)/2 \rfloor$ we have that $(k-p) \geq \lfloor (n-p+1)/2 \rfloor$ and then $\varepsilon_0^*(\mathbf{X})$ is just the minimum between $\varepsilon^*(t_{0n}, \mathbf{X})$ and $\varepsilon^*(V_{0n}, \mathbf{X})$.

Condition (E) seems rather artificial, so it is important to examine cases where it might hold. Theorem 3.1 in [22] shows that the BDP of the MVE equals the upper bound $\lfloor (n-p+1)/2 \rfloor$ when it is defined as the minimizer of $\det(V)$ for $(t, V) \in \mathcal{R}^p \times \mathrm{SPD}(p)$ subject to

$$\#\left\{i : d^2(x_i, t, V) \leq c^2\right\} \geq \lfloor (n+p+1)/2 \rfloor, \tag{6}$$

where $\mathrm{SPD}(p)$ denotes the space of symmetric positive definite $p \times p$ matrices. The tuning constant c^2 only determines the magnitude of V_{0n} and does not affect t_{0n} or the shape of V_{0n} (we call $H(V_{0n})$ a shape function of V_{0n} if $H(aV_{0n}) = H(V_{0n})$ for every $a > 0$). Taking $c^2 = \chi^2_{p,.5}$ makes V_{0n} consistent under normality. Note that the MVE defined this way satisfies (E) with $k = \lfloor (n+p+1)/2 \rfloor$ for any $\eta > \chi^2_{p,.5}$ (and recall that we take $\eta = \chi^2_{p,1-\alpha}$ with $\alpha = .025$ or a similar small value).

More generally, consider an S-estimator defined as the minimizer of $\det(V)$ subject to

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(d(x_i, t, V)\right) \leq r\rho(\infty) \tag{7}$$

where $0 < r < 1$, $\rho(0) = 0$, ρ is continuous and strictly increasing on $[0, c]$ and constant on $[c, \infty)$. When $r \leq (n-p)/2n$, Theorem 3.2 in [22] shows that the breakdown point of (t_{0n}, V_{0n}) is $\lceil nr \rceil / n$ (where $\lceil a \rceil$ is the smallest integer that is grater than or equal to a) and then the choice $r = (n-p)/2n$ yields exactly the upper bound $\lfloor (n-p+1)/2 \rfloor$, although in practice one takes $r = 1/2$ for simplicity. For elliptical distributions (as considered in Section 4) the location estimator t_{0n} and the shape of V_{0n} are consistent regardless of the value of c. In order to make V_{0n} consistent for the covariance matrix of a normal distribution, ρ must satisfy $E\{\rho(Q)\} = r\rho(\infty)$ where $Q^2 \sim \chi^2_p$. When ρ is in the family of Tukey's bisquared functions,

$$\rho_c(u) = \min\left\{\frac{u^2}{2} - \frac{u^4}{2c^2} + \frac{u^6}{6c^4}, \frac{c^2}{6}\right\},$$

this consistency condition uniquely determines the tuning constant c as a function of r. Even though $\#\{i : d_i^2 \leq c^2\} \geq nr$ for this type of S-estimators, c^2 goes to infinity with p (see Theorem 1 in [17] and comments thereafter) and then (E) cannot be guaranteed to hold for any fix k. This seems to be a problem inherent to all smooth ρ functions (it does not affect the MVE, which is an S-estimator defined with a jump function).

As for the MCD, recall that its objective is to find a subset of h observations whose covariance matrix has the lowest determinant. If $x_{i_1}, \ldots, x_{i_h}$ are those observations, then

$$\begin{aligned} t_{0n} &= \frac{1}{h}\sum_{j=1}^{h} x_{i_j} \\ S_{0n} &= \frac{1}{h-1}\sum_{j=1}^{h} \left(x_{i_j} - t_{0n}\right)\left(x_{i_j} - t_{0n}\right)'. \end{aligned}$$

S_{0n} is not consistent to the covariance matrix of a normal distribution unless multiplied by a consistency factor. One possible consistency correction is as follows. Let $d^2_{(i)}(\cdot, t_{0n}, S_{0n})$ be the i-th ordered statistic of the squared Mahalanobis distances with respect to (t_{0n}, S_{0n}), and define

$$V_{0n} = \frac{d^2_{(h)}(\cdot, t_{0n}, S_{0n})}{\chi^2_{p,h/n}} S_{0n}. \tag{8}$$

Then $\#\left\{i : d_i^2 \leq \chi^2_{p,h/n}\right\} \geq h$. Since h is typically chosen to be at most $.75n$, (E) will be satisfied with $k = h$ and any $\eta > \chi^2_{p,.75}$. The choice $h = \lfloor (n+p+1)/2 \rfloor$ makes both the MCD and the reweighted estimators attain the maximum BDP.

3.2 Asymptotic breakdown point

The concept of finite-sample BDP is easily interpretable and thus a useful tool for assessing the robustness of an estimator. However, it does not give any information about the behavior of the estimator when $n \to \infty$ and one systematically samples from contaminated distributions. Or to be more precise, the results obtained for finite samples cannot be immediately extrapolated to the asymptotic scenario. However, we show in this section that Theorem 1 does hold asymptotically under certain conditions, for the asymptotic BDP as defined below.

Let (t, V) be a pair of location and scatter functionals. That is, t and V are such that if F_n is the empirical distribution function of the sample $\mathbf{X}$ then $(t_n(\mathbf{X}), V_n(\mathbf{X})) = (t(F_n), V(F_n))$. It is clear that the functional expressions corresponding to (1) and (2) are

$$\alpha(F) = \sup_{u \geq \eta} \{G_p(u) - G_F(u)\}^+ \tag{9}$$

$$c(F) = G_F^{-1}(1 - \alpha(F)) \tag{10}$$

where $G_F(u) = P(d^2(X, t_0(F), V_0(F)) \leq u)$ with $X \sim F$. The functional expressions of the reweighted estimators are

$$\begin{array}{rcl} t_1(F) & = & E_F\{w_F(X)\, X\} / E_F\{w_F(X)\} \\ V_1(F) & = & E_F\left\{w_F(X)\,(X - t_1(F))\,(X - t_1(F))'\right\} / E_F\{w_F(X)\} \end{array} \tag{11}$$

where $w_F(X) = w(d^2(X, t_0(F), V_0(F))/c(F))$. In the proof of Theorem 2 it is shown that the expectations in (11) are well defined for any distribution F for which $(t_0(F), V_0(F))$ is defined. Note that the initial estimator might not have a well defined functional expression for a completely arbitrary distribution F. For the MCDE, [2] constructs a functional expression that is valid for any F. For the MVEE, Theorem 3 in [4] proves the existence of the functional under certain conditions on F that we mention below. Although these functionals might be not unique, that will not be a problem for our purposes.

Consider now a target distribution F_0 and the gross-error contamination neighborhood

$$\mathcal{V}_\varepsilon = \{(1-\varepsilon) F_0 + \varepsilon H : H \text{ is an arbitrary distribution on } \mathcal{R}^{\mathrm{p}}\}. \tag{12}$$

Even though $\mathcal{V}_\varepsilon$ is not a neighborhood in the topological sense, it is appropriate for studying asymptotic robustness properties of the estimators, because a distribution in $\mathcal{V}_\varepsilon$ will produce a proportion ε of arbitrary outliers in the sample. Thus $\mathcal{V}_\varepsilon$ is (roughly) the asymptotic counterpart of $\mathcal{X}_m$ with $\varepsilon = m/n$. The asymptotic breakdown point (ABDP) of a location estimator at F_0 is defined as

$$\varepsilon^*(t, F_0) = \sup \left\{ \begin{array}{c} \varepsilon > 0 : \text{there is a compact set } K(\varepsilon) \subset \mathcal{R}^p \\ \text{such that } t(F) \in K(\varepsilon) \text{ for all } F \in \mathcal{V}_\varepsilon \end{array} \right\}$$

and for an scatter estimator it is

$$\varepsilon^*(V, F_0) = \sup \left\{ \begin{array}{c} \varepsilon > 0 : \text{there is a compact set } K(\varepsilon) \subset (0, \infty) \\ \text{such that } \operatorname{cond}(V(F)) \in K(\varepsilon) \text{ for all } F \in \mathcal{V}_\varepsilon \end{array} \right\}.$$

Theorem 2 below gives lower bounds for the ABDPs of the reweighted estimators under two different sets of conditions.

(C) F_0 is continuous.

(B) $P_{F_0}(B) > 0$ for any open ball $B \subset \mathcal{R}^p$.

(E') There exist $c < \sqrt{\eta}$ and $\kappa > 0$ such that

$$P_F\left(d^2(X, t_0(F), V_0(F)) \leq c^2\right) \geq \kappa$$

for every $F \in \mathcal{V}_\varepsilon$ with $\varepsilon < \min\{\varepsilon^*(t_0, F_0), \varepsilon^*(V_0, F_0)\}$.

Theorem 2 *Let* $\varepsilon_0^*(F_0) = \min\{\varepsilon^*(t_0, F_0), \varepsilon^*(V_0, F_0)\}$. *Then:*

1. *If (W), (C) and (B) hold,* $\min\{\varepsilon^*(t_1, F_0), \varepsilon^*(V_1, F_0)\} \geq \varepsilon_0^*(F_0)$.
2. *If (W), (C) and (E') hold,* $\min\{\varepsilon^*(t_1, F_0), \varepsilon^*(V_1, F_0)\} \geq \min\{\varepsilon_0^*(F_0), \kappa\}$.

Part 2 of Theorem 2 is consistent with Theorem 1. Note that condition (E') is just the asymptotic version of (E) and will hold with $\kappa = 1/2$ for suitably tuned MCDE and MVEE. Specifically, if the MVEE is defined with $c^2 = \chi^2_{p,.5}$ and coverage κ, then (E') holds for any $\eta > \chi^2_{p,.5}$. If the MCDE has coverage κ and is scaled for consistency as explained in Section 3.1, then (E') holds for any $\eta > \chi^2_{p,\kappa}$. Recall that $\eta = \chi^2_{p,1-\alpha}$ and we typically take $\alpha = .025$, while the coverage κ is usually less than .75. Moreover, for the MVEE we deduce from Theorem 5 in [4] that $\varepsilon_0^*(F_0) \geq \min\{\kappa, 1-\kappa\}$ if F_0 is an elliptical distribution as in (13) with h non-increasing and strictly decreasing at c^2 (note that Davies defines ABDP using neigborhoods that are actually broader than $\mathcal{V}_\varepsilon$).

Part 1 of Theorem 2 makes the difference here, because (B) is a condition on the target distribution and not on the initial estimators. This way the result may be applied to S-estimators, which were precluded by assumptions (E) and (E').

Another tool for assessing the robustness of an estimator is the influence function. Since it is closely related to the asymptotic distribution of the estimator, we defer its treatment to the next section.

4 Asymptotics under elliptical models

Let us now turn to the asymptotic behavior of the adaptive reweighted estimators when the sample follows an elliptical distribution F with density

$$f(x) = |\Sigma|^{-1/2} h\left((x-\mu)' \Sigma^{-1} (x-\mu)\right), \tag{13}$$

with $\mu \in \mathcal{R}^p$, $\Sigma \in \mathrm{SPD}(p)$, and $h : [0,\infty) \to [0,\infty)$. The multivariate normal distribution, denoted by $N_p(\mu,\Sigma)$, corresponds to $h(u) = (2\pi)^{-p/2} e^{-u/2}$. Model (13) also accommodates heavy-tailed distributions such as the multivariate t with ν degrees of freedom, that will be denoted by $t_p(\nu)$ and corresponds to

$$h(u) = \frac{\Gamma\left(\frac{p+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\pi\nu)^{\frac{p}{2}}} \left(1+\frac{u}{\nu}\right)^{-\frac{p+\nu}{2}}.$$

The $t_p(1)$ distribution is the multivariate Cauchy.

For the $N_p(\mu,\Sigma)$ distribution the sample mean and sample covariance matrix are the uniform minimum variance unbiased estimators of μ and Σ. Also, they are the maximum likelihood estimators (up to an asymptotically negligible factor $(n-1)/n$ for the covariance) and hence attain the minimum asymptotic variance. For these reasons, besides computational simplicity, the sample mean and sample covariance matrix are the optimal estimators if the normal model holds. Robust estimators, on the other hand, may be very inefficient under normality, as mentioned in the introduction. We prove in this section that, under general conditions, the adaptive reweighted estimators (4) and (5) are asymptotically equivalent to the sample mean and covariance and hence fully efficient under normality.

Let us analyze first the asymptotic behavior of the cut-off values c_n given by (2). The distribution function of $d^2(X,\mu,\Sigma)$ under (13) is

$$\begin{aligned} G(u) &= P\left(d^2(X,\mu,\Sigma) \le u\right) \\ &= \int I\left(\|x\|^2 \le u\right) h\left(\|x\|^2\right) dx \\ &= \frac{2\pi^{p/2}}{\Gamma(p/2)} \int_0^{\sqrt{u}} h\left(r^2\right) r^{p-1} dr. \end{aligned}$$

(The last equality in the display and similar expressions in this section are obtained by applying Lemma 2.1 of [12].) If F is the $N_p(\mu,\Sigma)$ distribution, G is the χ^2_p distribution. For F the $t_p(\nu)$ distribution, G is p times the $\mathcal{F}(p,\nu)$ distribution.

Lemma 1 *Let $t_{0n} \to \mu$ and $V_{0n} \to \gamma\Sigma$ in probability, for some $\gamma > 0$. If α_n is as in (1) then $\alpha_n \to \alpha_0$ in probability with*

$$\alpha_0 = \sup_{u \ge \eta} \left\{G_p(u) - G(\gamma u)\right\}^+. \tag{14}$$

If in addition G is strictly increasing in its support and c_n is as in (2) then $c_n \to c_0$ in probability with

$$c_0 = G^{-1}(1-\alpha_0)/\gamma. \tag{15}$$

Remark. Note that in the notation of Section 3.2 it is $G(\gamma u) = G_F(u)$, so that Lemma 1 is just saying that $\alpha(F_n) \to \alpha(F)$ and $c(F_n) \to c(F)$ in probability, as expected.

In the preceding Lemma we have not required that $V_{0n} \to \Sigma$ in probability because this will not hold for *every* elliptical distribution. Scatter estimators are typically calibrated to be consistent for the normal distribution, but for other elliptical distributions V_{0n} will converge to $\gamma\Sigma$ for some $\gamma \neq 1$. For example, if V_{0n} is the MVE or the MCD with coverage $h \approx \beta n$ and calibrated for consistency under the normal distribution, then $\gamma = G^{-1}(\beta)/G_p^{-1}(\beta)$. Note that if G is stochastically less than or equal to G_p, then $\gamma \geq 1$, $\alpha_0 = 0$ and $c_0 = G^{-1}(1)/\gamma$. If in addition G is of unbounded support then $c_0 = \infty$, which means that no observations are discarded in the limit. This eventually makes the estimator asymptotically efficient for the multivariate normal model. Precise conditions are given in Theorem 3. For that theorem, which gives asymptotic expansions for the reweighted estimators, we need the following assumption:

(H) The function h in (13) is continuously differentiable.

Theorem 3 *Let t_{1n} and V_{1n} be the adaptive reweighted estimators given by (4) and (5). Suppose that (W), (H) and all the conditions of Lemma 1 are satisfied. Let $\Sigma = B^2$ and $\gamma^{-1}V_{0n} = B_{0n}^2$, with B and B_{0n} in $\mathrm{SPD}(p)$. Then*

$$\begin{aligned} t_{1n} &= \mu + \frac{a_2(\gamma c_0)}{a_1(\gamma c_0)}(t_{0n} - \mu) + \frac{1}{a_1(\gamma c_0)}\frac{1}{n}\sum_{i=1}^{n} w\left(\frac{d^2(X_i, \mu, \Sigma)}{\gamma c_0}\right)(X_i - \mu) \\ &\quad + o_P\left(n^{-1/2}\right) + o_P\left(\|(t_{0n} - \mu, B_{0n} - B)\|\right) \end{aligned}$$

and

$$\begin{aligned} V_{1n} &= \frac{a_3(\gamma c_0)}{a_1(\gamma c_0)}\Sigma + \frac{a_4(\gamma c_0)}{a_1(\gamma c_0)}\left\{\mathrm{tr}\left(B^{-1}(B_{0n} - B)\right)\Sigma + 2B^{-1}(B_{0n} - B)\Sigma\right\} \\ &\quad + \frac{1}{a_1(\gamma c_0)}\frac{1}{n}\sum_{i=1}^{n}\left\{w\left(\frac{d^2(X_i, \mu, \Sigma)}{\gamma c_0}\right)(X_i - \mu)(X_i - \mu)' - a_3(\gamma c_0)\Sigma\right\} \\ &\quad + o_P\left(n^{-1/2}\right) + o_P\left(\|(t_{0n} - \mu, B_{0n} - B, t_{1n} - \mu)\|\right), \end{aligned}$$

where

$$\begin{aligned} a_1(u) &= E_F\left\{w\left(\frac{\|X\|^2}{u}\right)\right\} = \frac{2\pi^{p/2}}{\Gamma(p/2)}\int_0^\infty w\left(\frac{r^2}{u}\right)h\left(r^2\right)r^{p-1}dr, \\ a_2(u) &= \frac{2\pi^{p/2}}{\Gamma(p/2)}\int_0^\infty w\left(\frac{r^2}{u}\right)\left\{h\left(r^2\right) + \frac{2}{p}h'\left(r^2\right)r^2\right\}r^{p-1}dr, \\ a_3(u) &= \frac{1}{p}E_F\left\{w\left(\frac{\|X\|^2}{u}\right)\|X\|^2\right\} = \frac{2\pi^{p/2}}{\Gamma(p/2)}\int_0^\infty \frac{1}{p}w\left(\frac{r^2}{u}\right)h\left(r^2\right)r^{p+1}dr, \\ a_4(u) &= \frac{2\pi^{p/2}}{\Gamma(p/2)}\int_0^\infty w\left(\frac{r^2}{u}\right)\left\{\frac{r^2}{p}h\left(r^2\right) + \frac{2r^4}{p(p+2)}h'\left(r^2\right)\right\}r^{p-1}dr. \end{aligned}$$

Both expansions hold even if $c_0 = \infty$.

Remark: When $w(u) = I(u < 1)$ and F is the $N_p(\mu, \Sigma)$ distribution we have

$$\begin{aligned} a_1(u) &= G_p(u), \\ a_2(u) &= \frac{\pi^{\frac{p}{2}}}{\Gamma\left(\frac{p}{2}+1\right)} h(u)\, u^{\frac{p}{2}} \\ &= \frac{1}{\Gamma\left(\frac{p}{2}+1\right)} \left(\frac{u}{2}\right)^{\frac{p}{2}} e^{-u/2}, \\ a_3(u) &= G_{p+2}(u), \\ a_4(u) &= \frac{\pi^{\frac{p}{2}}}{2\Gamma\left(\frac{p}{2}+2\right)} h(u)\, u^{\frac{p}{2}+1} \\ &= \frac{1}{\Gamma\left(\frac{p}{2}+2\right)} \left(\frac{u}{2}\right)^{\frac{p}{2}+1} e^{-u/2}. \end{aligned}$$

Analogous expressions for the $t_p(\nu)$ distribution are given in the Appendix.

Remark: Note that in the preceding Theorem we do not need to impose a finite fourth moment assumption like H2 in [12], because weights that satisfy (W) are redescending when c_0 is finite, and $c_0 = \infty$ only if $G(\gamma \cdot)$ is stochastically less than or equal to G_p in the tails, in which case fourth moments must be finite.

When (13) is the multivariate normal, and V_{0n} is calibrated for consistency under normality, we have $c_0 = \infty$, $\gamma = 1$, $a_1(c_0) = a_3(c_0) = 1$, and $a_2(c_0) = a_4(c_0) = 0$. If in addition t_{0n} and V_{0n} are $n^{1/2}$-consistent, Theorem 3 implies that the adaptive reweighted estimators are asymptotically equivalent to the sample mean and variance and hence asymptotically efficient under the multivariate normal model.

When the initial estimators are n^{τ}-consistent with $\tau < 1/2$, such as the MVE, if $c_0 = \infty$ we obtain that $t_{1n} - \mu$ and $V_{1n} - \left\{E_F\left(\|X\|^2\right)/p\right\}\Sigma$ are $o_P\left(n^{-\tau}\right)$. Unfortunately we cannot say whether or not the rate of convergence is improved. Lemma 3 in [7], which is the key to prove that the adaptive estimator improves the rate of convergence in the linear regression model, relies heavily on the assumption of symmetric error distribution and hence does not carry over to the present setting. On the other hand, if $c_0 < \infty$ the one-step estimators remain n^{τ}-consistent and the coefficients $a_2(\gamma c_0)/a_1(\gamma c_0)$ and $a_4(\gamma c_0)/a_3(\gamma c_0)$ are the relative efficiencies of the (unbiased) reweighted estimators with respect to the initial estimators. As Lopuhaä [12] shows, for hard-rejection weights and a density with non-increasing function h these coefficients are less than one, so that the reweighted estimators are still more efficient than the initial estimators.

4.1 Influence function

While the asymptotic breakdown point is a measure of global robustness of an estimator, the influence function is useful to study the so-called local robustness. It measures the sensitivity of an estimator to small amounts of contamination (see [8] for a thorough discussion of the influence function and its uses in robust statistics). Formally, if Δ_x is the distribution that puts mass 1 on $x \in \mathcal{R}^p$, the influence function of an estimator T at

the distribution F is defined as

$$\mathrm{IF}\left(x, T, F\right) = \lim_{\varepsilon \downarrow 0} \frac{T\left(\left(1-\varepsilon\right)F + \varepsilon \Delta_x\right) - T\left(F\right)}{\varepsilon}$$

when the limit exists. The following theorem gives the influence functions of the reweighted estimators t_1 and V_1 for elliptical models.

Theorem 4 *Let F be an elliptical distribution as in (13). Assume that (W) and (H) are satisfied and that $G(u)$ is strictly increasing in its support. If* $\mathrm{IF}(x, t_0, F)$ *and* $\mathrm{IF}(x, V_0, F)$ *exist and $d^2\left(x, \mu, \Sigma\right)/c_0\gamma$ is a continuity point of w, then* $\mathrm{IF}(x, t_1, F)$ *and* $\mathrm{IF}(x, V_1, F)$ *exist and are given by*

$$\begin{aligned}
\mathrm{IF}\left(x, t_1, F\right) &= \frac{a_2\left(\gamma c_0\right)}{a_1\left(\gamma c_0\right)} \mathrm{IF}\left(x, t_0, F\right) + \frac{1}{a_1\left(\gamma c_0\right)} w\left(\frac{d^2\left(x, \mu, \Sigma\right)}{\gamma c_0}\right)\left(x - \mu\right) \\
\mathrm{IF}\left(x, V_1, F\right) &= \frac{a_4\left(\gamma c_0\right)}{a_1\left(\gamma c_0\right)} \left\{\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\mathrm{IF}\left(x, V_0, F\right)\Sigma^{-1}\right)\Sigma + B^{-1}\mathrm{IF}\left(x, V_0, F\right)B^{-1}\right\} \\
&\quad + \frac{1}{a_1\left(\gamma c_0\right)} \left\{w\left(\frac{d^2\left(x, \mu, \Sigma\right)}{\gamma c_0}\right)\left(x-\mu\right)\left(x-\mu\right)' - a_3\left(\gamma c_0\right)\Sigma\right\}
\end{aligned}$$

where $B \in \mathrm{SPD}(p)$ is such that $B^2 = \Sigma$.

Note how the characteristic adaptivity of these estimators manifests itself in the influence functions, which depend on the model distribution F through c_0. If the initial estimators are bounded-influence, so will be the reweighted estimators when $c_0 < \infty$, that is, for heavy-tailed elliptical models. For the multivariate normal, though, we have an infinite c_0 and consequently unbounded influence. Hence at the normal model we have to give up local robustness if we want to attain maximal efficiency. Whether this is acceptable or if it is better to sacrifice some efficiency in order to keep the influence functions bounded for *all* elliptical distributions is largely a matter of preference. In general, high breakdown point and reasonably low bias seem to be regarded as the fundamental indicators of robustness. In linear regression, for instance, several robust estimators (such as S- and τ-estimators) have unbounded influence functions yet high breakdown points and good bias properties, hence they are extensively used in practice. The situation for the multivariate location-scatter model is different in that all robust estimators used in practice (with the exception of the MVE) are bounded-influence. But the simulations reported in Section 5 show that the estimators we propose have good bias properties, so we feel confident to recommend their use in practice.

4.2 The MCD as initial estimator

Theorem 3 provides a way to compute the asymptotic variances of the adaptive estimators for elliptical distributions other than normal, and we use it in this subsection to analyze the effect of heavy-tailed elliptical distributions on the asymptotic variances. We use the MCD as initial estimator and the hard-rejection weight function.

Let (t_{0n}, V_{0n}) be the MCD with coverage $h \approx \beta n$ and (t_{1n}, V_{1n}) the adaptive one-step estimator. Let $K_{p,p}$ be the $p^2 \times p^2$ commutation matrix, $q = G^{-1}(\beta)$, $\gamma = q/G_p^{-1}(\beta)$ and

$$\begin{aligned} a_5(u) &= \frac{1}{p(p+2)} E_F \left\{ w\left(\frac{\|X\|^2}{u} \right) \|X\|^4 \right\} \\ &= \frac{2\pi^{p/2}}{\Gamma(p/2)\, p(p+2)} \int_0^{\sqrt{u}} h\left(r^2\right) r^{p+3} dr. \end{aligned}$$

Then

$$\begin{aligned} \mathrm{AV}\left\{ \sqrt{n}(t_{0n} - \mu) \right\} &= \kappa_0 \Sigma, \\ \mathrm{AV}\left\{ \sqrt{n}\mathrm{vec}(V_{0n} - \gamma\Sigma) \right\} &= \sigma_0 (\mathrm{I} + K_{p,p}) \Sigma \otimes \Sigma + \lambda_0 \mathrm{vec}(\Sigma) \mathrm{vec}(\Sigma)', \end{aligned}$$

where AV stands for asymptotic variance and

$$\begin{aligned} \kappa_0 &= \frac{a_3(q)}{\{a_1(q) - a_2(q)\}^2}, \\ \sigma_0 &= \frac{a_5(q)}{\{a_3(q) - a_4(q)\}^2}. \end{aligned}$$

Similarly

$$\begin{aligned} \mathrm{AV}\left\{ \sqrt{n}(t_{1n} - \mu) \right\} &= \kappa_1 \Sigma, \\ \mathrm{AV}\left\{ \sqrt{n}\mathrm{vec}\left(V_{1n} - \frac{a_3(\gamma c_0)}{a_1(\gamma c_0)} \Sigma \right) \right\} &= \sigma_1 (\mathrm{I} + K_{p,p}) \Sigma \otimes \Sigma + \lambda_1 \mathrm{vec}(\Sigma) \mathrm{vec}(\Sigma)', \end{aligned}$$

with

$$\begin{aligned} \kappa_1 &= \frac{1}{a_1^2(\gamma c_0)} \left[\frac{a_2^2(\gamma c_0)\, a_3(q)}{\{a_1(q) - a_2(q)\}^2} + a_3(\gamma c_0) \right. \\ &\quad \left. + \frac{2 a_2(\gamma c_0)\, a_3(\gamma c_0 \wedge q)}{\{a_1(q) - a_2(q)\}} \right], \end{aligned} \tag{16}$$

$$\begin{aligned} \sigma_1 &= \frac{1}{a_1^2(\gamma c_0)} \left[\frac{a_4^2(\gamma c_0)\, a_5(q)}{\{a_3(q) - a_4(q)\}^2} + a_5(\gamma c_0) \right. \\ &\quad \left. + \frac{2 a_4(\gamma c_0)\, a_5(\gamma c_0 \wedge q)}{\{a_3(q) - a_4(q)\}} \right]. \end{aligned} \tag{17}$$

Obviously the same expressions hold for asymptotic variances of fix-threshold one-step estimators if c_0 is the corresponding threshold value. The derivation of these expressions is explained in the Appendix, where computable expressions for λ_0 and λ_1 are given. Here we focus only on the σ-parameters because asymptotic distributions of shape functions of scatter estimators do not involve λ-parameters, as established by Theorem 1 in [28].

The first type of heavy-tailed elliptical distributions we will consider is the $t_p(\nu)$ family. Figure 1 displays the asymptotic relative efficiencies κ_0/κ_1 and σ_0/σ_1 as functions

of the degrees of freedom ν for the adaptive and the non-adaptive estimators, for $p = 3$ and $p = 10$. We have taken $\beta = .5$ for the MCD and $\eta = \chi^2_{p,.9}$. For the non-adaptive estimator we took $\chi^2_{p,.9}$ as threshold value. This value, smaller than the usual $\chi^2_{p,.975}$, was chosen to make the difference between the adaptive and the non-adaptive estimator more clear-cut. We can see in Figure 1 that the one-step estimators are always more efficient than the initial estimators except for the Cauchy distribution. The adaptive location estimator already outperforms its non-adaptive counterpart for ν as small as 3. The adaptive scatter estimator, in contrast, needs larger degrees of freedom to outperform the non-adaptive one, yet its efficiency for small ν is still acceptable.

Another family of elliptical distributions is given by the contaminated normal model,

$$F_{\varepsilon,k} = (1-\varepsilon)\, N_p\,(0, \mathrm{I}) + \varepsilon N_p\,(0, k\mathrm{I})$$

with $0 \leq \varepsilon < .5$ and $k > 0$. The asymptotic relative efficiencies κ_0/κ_1 and σ_0/σ_1 were computed for $\varepsilon = .025, .05, \ldots, .25$ and several values of k. For the location estimators, the smallest ARE for each ε is attained at $k = 0$. For the scatter estimators, the least-favorable $k(\varepsilon)$ was found by grid search. Figure 2 displays the asymptotic relative efficiencies as functions of ε for the least favorable $F_{\varepsilon,k(\varepsilon)}$. The AREs of the estimators are decreasing in ε but remain greater than 1. In particular, the adaptive estimators remain more efficient than their non-adaptive counterparts over fairly large neighborhoods of the normal distribution.

5 Monte Carlo study

In order to assess the finite sample efficiency and robustness of the proposed estimator we carried out some simulations. Rather than considering the broadest possible combination of estimators and sampling situations, we focused on the most commonly used multivariate estimators and compared the effect of a fix threshold reweighting step with that of the adaptive reweighting step proposed in this paper, using the non-reweighted estimator as a benchmark.

As initial estimators we considered the following:

1. Minimum Volume Ellipsoid (MVE) as defined in (6). It is included to illustrate the effect of the adaptive weighting scheme on a $n^{1/3}$-consistent estimator. Although there exist exact algorithms to compute the MVE, they are impractical for the large samples considered in the present study, so we opted for the approximate subsampling algorithm described in [22], based on 1000 subsamples.

2. Minimum Covariance Determinant (MCD) as defined in (8), with coverage $h = \lfloor (n+p+1)/2 \rfloor$ that yields maximum breakdown point. We used the FAST-MCD algorithm proposed in Section 5 of [23] with 500 random starts.

3. Multivariate S-estimator (S) as defined in (7), with Tukey's bisquared function and $r = 1/2$. We also tried the Winsorized squares ρ function, which is supposed to

alleviate the problem of lack of robustness in high dimensions according to [17], but that was not the case in our simulations and hence the results are not reported. A subsampling algorithm based on 1000 subsamples was used to find a good starting point and then iterated to a local minimum.

Starting from the three estimators mentioned above, we computed one-step reweighted estimators with fix threshold value $\chi^2_{p,.975}$ and with adaptive threshold values (2) with $\eta = \chi^2_{p,.975}$. These estimators are respectively denoted by the suffixes 1F and 1A in Tables 1 to 4.

The sampling situations considered were the multivariate normal, the multivariate Cauchy and the "shifted" normal (explained below). As a measure of error for the location estimator we considered the standardized squared error $\mathrm{SE} = (t_n - \mu)' \Sigma^{-1} (t_n - \mu)$. For the covariance matrix we again concentrated on the estimation of the shape of Σ and thus considered $\mathrm{LCN} = \log \mathrm{cond}\left(\Sigma^{-1/2} V_n \Sigma^{-1/2}\right)$ as error measure. By equivariance of the estimators and invariance of the error measures, without loss of generality we took $\mu = 0$ and $\Sigma = \mathrm{I}$.

For the multivariate normal model we generated 1000 data sets of sizes $n = 50, 100, 200, 500$ and dimensions $p = 3, 10$. Tables 1 and 2 summarize the results. For location estimators Table 1 reports their relative MSEs with respect to the sample mean. For scatter estimators Table 2 reports their relative mean LCNs with respect to the sample covariance. Figure 3 displays the relative errors of the adaptive one-step estimator with respect to the non-adaptive one when the initial estimator is the MCD.

To study the robustness of the estimators with respect to a heavy-tailed elliptical distribution we considered the multivariate Cauchy. Tables 3 and 4 report under CAU the relative median errors with respect to the maximum likelihood estimators of μ and Σ. As an example of outlier contamination we chose what has been called the "shifted" normal distribution, $(1-\varepsilon) N_p(0, \mathrm{I}) + \varepsilon N_p(\mu^*, \mathrm{I})$. This type of contamination where the covariance of the "bad" data coincides with that of the "good" data is particularly troublesome, even more than point-mass contaminations (see comment after Theorem 1 in [18]). In order to reduce sampling variability we did not randomize outliers. Rather, we just added μ^* to the first $n\varepsilon$ points of each simulated sample. By affine equivariance we took without loss of generality $\mu^* = ke_1$ for $k = 1, \ldots, 20$. The proportions of outliers considered were $\varepsilon = .10, .20$ and the sample size was $n = 50$. All situations were replicated 1000 times. Tables 3 and 4 report the maximum (with respect to k) of the median errors. Note that for this model we have $\mu = \varepsilon k e_1$ and $\Sigma = \mathrm{I} + \varepsilon(1-\varepsilon) k^2 e_1 e_1'$, thus the median SE of the sample mean converges to $\varepsilon^2 k^2$ and the median LCN of the sample covariance converges to $\log(1 + \varepsilon(1-\varepsilon) k^2)$, the maximums being attained at $k = 20$. The entries on Tables 3 and 4 are consistent with this.

Overall, we can say that the gains yielded by the weighting step are important. Under the normal model the gains in efficiency are considerable, especially when the initial estimators are the MVE and the MCD. For the MVE the theory predicts that the relative efficiency of the fix-threshold reweighted estimators eventually decreases to zero because the rate of consistency remains $n^{1/3}$, but for a sample size as large as 500 the gains in efficiency are still considerable. In all the situations considered in the simulations, the

adaptive estimator is noticeably more efficient than the fix-threshold estimator for samples of size 200 and the superiority is evident for 500. The S-estimator is more efficient than the other two under the normal model, especially for large dimensions (this is consistent with the asymptotic efficiencies displayed in Figure 3 in [17]), but it performs very poorly for contaminated normal distributions. Note that for SN(.20) it behaves as if it had broken down. Rather than being a problem inherent to the estimator, this is a problem of Tukey's bisquared function, that requires for consistency a cut-off constant that grows quicly with the dimension (for $p = 10$ it is already $c = 6.75$). Rocke [17] proposes a ρ function which is a translated bisquared function that depends on two tuning constants. However, I think that an adaptive reweighted one step estimator starting from the MCD does the job as well and is a much simpler procedure.

A Proofs

Proof of Theorem 1 Replace at most $m = n\varepsilon_0^*(\mathbf{X}) - 1$ points of $\mathbf{X}$ and denote the corrupted sample by $\mathbf{X}^*$. To simplify the notation, let $t_{0n}^* = t_{0n}(\mathbf{X}^*)$ and analogously define $t_{1n}^*, V_{0n}^*, V_{1n}^*$ and the Mahalanobis distances d_i^*. If $i_0 = \max\left\{i : d_{(i)}^{*2} < \eta\right\}$ then

$$\alpha_n = \max_{i>i_0}\left\{G_p\left(d_{(i)}^{*2}\right) - \frac{i-1}{n}\right\}^+$$

and $c_n = d_{(i_n)}^{*2}$ with $i_n = n - \lfloor n\alpha_n \rfloor$. Remember that $i_n > i_0$ and $c_n > \eta$. Let $w_{in}^* = w\left(d_i^{*2}/c_n\right)$ and, abusing the notation, $w_{(i)}^* = w\left(d_{(i)}^{*2}/c_n\right)$. Then

$$\begin{aligned}\left\|t_{1n}^* - t_{0n}^*\right\|^2 &\leq \frac{1}{\left(\sum_{i=1}^n w_{in}^*\right)^2}\sum_{i=1}^n w_{in}^{*2}\left\|x_i^* - t_{0n}^*\right\|^2 \\ &\leq \frac{\lambda_1\left(V_{0n}^*\right)}{\left(\sum_{i=1}^n w_{in}^*\right)^2}\sum_{i=1}^n w_{(i)}^{*2}d_{(i)}^{*2}\end{aligned}$$

where $\lambda_1\left(V_{0n}^*\right)$ is the largest eigenvalue of V_{0n}^*. Since $w_{(i)} = 0$ for $i \geq i_n$ and

$$d_{(i)}^{*2} \leq G_p^{-1}\left(\frac{i + \lfloor n\alpha_n \rfloor}{n}\right) \; for \; i \geq i_0 + 1$$

we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n w_{(i)}^{*2}d_{(i)}^{*2} &\leq \eta + \frac{1}{n}\sum_{i=i_0+1}^{n-\lfloor n\alpha_n \rfloor - 1} d_{(i)}^{*2} \\ &\leq \eta + \frac{1}{n}\sum_{i=i_0+1}^{n-\lfloor n\alpha_n \rfloor - 1} G_p^{-1}\left(\frac{i + \lfloor n\alpha_n \rfloor}{n}\right)\end{aligned}$$

$$
\begin{aligned}
&\leq \eta + \frac{1}{n}\sum_{i=1}^{n} G_p^{-1}\left(\frac{i-1}{n}\right) \\
&\leq \eta + \int_0^1 G_p^{-1}(u)\, du \\
&= \eta + p.
\end{aligned}
$$

On the other hand, by assumption (E) we have that $d_i^{*2} \leq c^2$ for at least $k-m$ of the original points. Since $\varepsilon_0^*(\mathbf{X}) \leq (k-p)/n$ we have $k-m \geq p+1$. So at least $p+1$ points of the original sample, that without loss of generality we can assume to be $x_1, \ldots, x_{p+1}$, are such that $w\left(d^2\left(x_i, t_{0n}^*, V_{0n}^*\right)/c_n\right) \geq w\left(c^2/\eta\right) > 0$. Then

$$
\sum_{i=1}^{n} w_{in}^* \geq w\left(\frac{c^2}{\eta}\right)(p+1)
$$

Therefore

$$
\left\|t_{1n}^* - t_{0n}^*\right\|^2 \leq \frac{\lambda_1\left(V_{0n}^*\right) n\,(\eta+p)}{w^2\left(c^2/\eta\right)(p+1)^2}
$$

Since $m \leq n\varepsilon_0^*(\mathbf{X}) - 1$, $\lambda_1\left(V_{0n}^*\right)$ and t_{0n}^* stay bounded for every $\mathbf{X}^* \in \mathcal{X}_m$ and then it must be $\varepsilon^*\left(t_{1n}, \mathbf{X}\right) \geq \varepsilon_0^*(\mathbf{X})$.

Let us turn now to V_{1n}^*. Since

$$
\begin{aligned}
\operatorname{tr}\left(V_{1n}^*\right) &= \sum_{i=1}^{n} w_{in}^* \left\|x_i^* - t_{1n}^*\right\|^2 \Big/ \sum_{i=1}^{n} w_{in}^* \\
&\leq \lambda_1\left(V_{0n}^*\right)(\eta+p) \Big/ \left\{(p+1)\, w\left(c^2/\eta\right)\right\},
\end{aligned}
$$

$\lambda_1\left(V_{1n}^*\right)$ must remain bounded for every $\mathbf{X}^* \in \mathcal{X}_m$ with $m \leq n\varepsilon_0^*(\mathbf{X}) - 1$. On the other hand, if $\lambda_p\left(V_{1n}^*\right)$ is the smallest eigenvalue of V_{1n}^*, by a well known property of eigenvalues of symmetric positive semi-definite matrices (see for instance Lemma 5.1 in [13]) we have

$$
\begin{aligned}
\lambda_p\left(V_{1n}^*\right) &= \lambda_p\left(\sum_{i=1}^{n} w_{in}^*\left(x_i^* - t_{1n}^*\right)\left(x_i^* - t_{1n}^*\right)' \Big/ \sum_{i=1}^{n} w_{in}^*\right) \\
&\geq \frac{w\left(c^2/\eta\right)}{n}\lambda_p\left(\sum_{i=1}^{p+1}\left(x_i - t_{1n}^*\right)\left(x_i - t_{1n}^*\right)'\right) \\
&\geq \frac{w\left(c^2/\eta\right)}{n}\lambda_p\left(\sum_{i=1}^{p+1}\tilde{x}_i\tilde{x}_i'\right) \\
&\geq \frac{w\left(c^2/\eta\right)}{n}\lambda_0(\mathbf{X})
\end{aligned}
$$

where

$$
\lambda_0(\mathbf{X}) = \min\left\{\lambda_p\left(\sum_{j=1}^{p+1}\tilde{x}_{i_j}\tilde{x}_{i_j}'\right) : 1 \leq i_1 < \ldots < i_{p+1} \leq n\right\}
$$

and $\tilde{x}_{i_j}$ denotes observation x_{i_j} centered by the mean of the corresponding subsample. Since $\lambda_0(\mathbf{X}) > 0$ by (GP), $\lambda_p(V_{1n}^*)$ stays bounded away from 0 for every $\mathbf{X}^* \in \mathcal{X}_m$ with $m \leq n\varepsilon_0^*(\mathbf{X}) - 1$. Then $\varepsilon^*(V_{1n}, \mathbf{X}) \geq \varepsilon_0^*(\mathbf{X})$ and the proof is complete. □

Proof of Theorem 2 This proof mimics that of Theorem 1. We prove both parts of the theorem simultaneously because they are very similar. The key is the inequality

$$E_F\left\{w_F(X)\, d^2(X, t_0(F), V_0(F))\right\} \leq \eta + p$$

for all $F \in \mathcal{V}_\varepsilon$ with $\varepsilon < \varepsilon_0^*(F_0)$. This inequality follows from the fact that $G_p(u) - G_F(u) \leq \alpha(F)$ for all $u \geq \eta$ and $G_F(u) < 1 - \alpha(F)$ for $\eta \leq u < c(F)$. If $\eta \leq c < c(F)$ we have

$$\begin{aligned}
\int_\eta^c u dG_F(u) &= cG_F(c) - \eta G_F(\eta) - \int_\eta^c G_F(u)\, du \\
&\leq c(1 - \alpha(F)) - \eta G_p(\eta) + \eta\alpha(F) \\
&\quad - \int_\eta^c \{G_p(u) - \alpha(F)\}\, du \\
&= c(1 - G_p(c)) + \int_\eta^c u dG_p(u) \\
&\leq \int_\eta^\infty u dG_p(u)
\end{aligned}$$

and then

$$\begin{aligned}
E_F\left\{w_F(X)\, d^2(X, t_0(F), V_0(F))\right\} &= \int_0^\infty w\left(\frac{u}{c(F)}\right) u dG_F(u) \\
&\leq \int_0^{c(F)} u dG_F(u) \\
&\leq \eta + \int_\eta^\infty u dG_p(u) \\
&\leq \eta + p.
\end{aligned}$$

Since

$$E_F\left\{w_F(X)\, \|X - t_0(F)\|^2\right\} \leq \lambda_1(V_0(F))\, E_F\left\{w_F(X)\, d^2(X, t_0(F), V_0(F))\right\}$$

it follows that

$$t_1(F) = \arg\min_{t \in \mathcal{R}^p} E_F\left\{w_F(X)\, \|X - t\|^2\right\}$$

is well defined for all $F \in \mathcal{V}_\varepsilon$ with $\varepsilon < \varepsilon_0^*(F_0)$ and the expression in (11) is valid. By taking the trace, it is clear that the expression for $V_1(F)$ in (11) is valid as well. Moreover, we have that

$$\|t_1(F) - t_0(F)\| \leq \frac{\sqrt{\lambda_1(V_0(F))}\sqrt{\eta + p}}{E_F\{w_F(X)\}}.$$

If condition (E') is assumed,

$$E_F\{w_F(X)\} \geq w\left(\frac{c^2}{\eta}\right)\kappa$$

for all $F \in \mathcal{V}_\varepsilon$ with $\varepsilon < \varepsilon_0^*(F_0)$ and then $\varepsilon^*(t_1, F_0) \geq \varepsilon_0^*(F_0)$. Now suppose that (B) holds instead of (E'). If $\varepsilon^*(t_1, F_0) < \varepsilon_0^*(F_0)$, there would be an $\varepsilon < \varepsilon_0^*(F_0)$ and a sequence of distributions $\{F_n\} \subset \mathcal{V}_\varepsilon$ such that $\|t_1(F_n)\| \to \infty$, which would imply that $E_{F_n}\{w_{F_n}(X)\} \to 0$ and then

$$\lim_{n\to\infty} E_{F_0}\left\{w\left(\frac{d^2(X, t_0(F_n), V_0(F_n))}{\eta}\right)\right\} = 0.$$

Now, since the initial estimators do not breakdown for such ε, there would be a subsequence $\{F_{n_k}\}$ such that $t_0(F_{n_k}) \to t_0^*$ for some vector t_0^* and $V_0(F_{n_k}) \to V_0^*$ for some non-singular matrix V_0^*. Then by dominated convergence we would have that $E_{F_0}\{w(d^2(X, t_0^*, V_0^*)/\eta)\} = 0$, which is impossible under assumption (B). Therefore it must be $\varepsilon^*(t_1, F_0) \geq \varepsilon_0^*(F_0)$.

As for V_1, we have that

$$\begin{aligned}\operatorname{tr}(V_1(F)) &= E_F\left\{w_F(X)\|X - t_1(F)\|^2\right\}/E_F\{w_F(X)\} \\ &\leq \lambda_1(V_0(F))(\eta + p)/E_F\{w_F(X)\}.\end{aligned}$$

As in the preceding paragraph, we deduce that $\lambda_1(V_1(F))$ remains bounded on $\mathcal{V}_\varepsilon$ if $\varepsilon < \varepsilon_0^*(F_0)$. To prove that $\lambda_p(V_1(F))$ remains bounded away from zero, consider $e(F)$ the eigenvector with norm 1 associated with $\lambda_p(V_1(F))$, so that

$$\lambda_p(V_1(F)) = \frac{E_F\left\{w_F(X)\left|(X - t_1(F))'e(F)\right|^2\right\}}{E_F\{w_F(X)\}}.$$

Under condition (E') we have

$$E_F\left\{w_F(X)\left|(X - t_1(F))'e(F)\right|^2\right\}$$

$$\begin{aligned}&\geq w\left(\frac{c^2}{\eta}\right)\xi^2 P_F\left(d^2(X, t_0(F), V_0(F)) \leq c^2 \text{ and } \left|(X - t_1(F))'e(F)\right| \geq \xi\right) \\ &\geq w\left(\frac{c^2}{\eta}\right)\xi^2\left(\kappa - P_F\left(\left|(X - t_1(F))'e(F)\right| < \xi\right)\right) \\ &\geq w\left(\frac{c^2}{\eta}\right)\xi^2\left(\kappa - \varepsilon - (1-\varepsilon)P_{F_0}\left(\left|(X - t_1(F))'e(F)\right| < \xi\right)\right)\end{aligned}$$

where ξ is any positive number. If there were a sequence $\{F_n\} \subset \mathcal{V}_\varepsilon$ with $\varepsilon < \min\{\varepsilon_0^*(F_0), \kappa\}$ such that $\lambda_p(V_1(F_n)) \to 0$, there would be a subsequence $\{F_{n_k}\}$ such that $e(F_{n_k}) \to e^*$ and $t_1(F_{n_k}) \to t_1^*$ for some vectors e^* and t_1^* for which we would then have that

$$P_{F_0}\left(\left|(X - t_1^*)'e^*\right| < \xi\right) \geq \frac{\kappa - \varepsilon}{1 - \varepsilon}.$$

But $P_{F_0}\left(\left|(X-t_1^*)'\,e^*\right|=0\right)=0$ by (C), so we can always choose a $\xi(\varepsilon)>0$ that violates the preceding inequality. Therefore it must be $\varepsilon^*(V_1,F_0)\geq\min\left\{\varepsilon_0^*(F_0),\kappa\right\}$. Now suppose that (B) holds instead of (E'). If there were a sequence $\{F_n\}\subset\mathcal{V}_\varepsilon$ with $\varepsilon<\varepsilon_0^*(F_0)$ such that $\lambda_p(V_1(F_n))\to 0$, there would be a subsequence $\{F_{n_k}\}$ such that $e(F_{n_k})\to e^*$, $t_0(F_{n_k})\to t_0^*$, $t_1(F_{n_k})\to t_1^*$ and $V_0(F_{n_k})\to V_0^*$ for some vectors e^*, t_0^* and t_1^* and a non-singular matrix V_0^*, for which

$$E_{F_0}\left\{w\left(\frac{d^2(X,t_0^*,V_0^*)}{\eta}\right)\left|(X-t_1^*)'\,e^*\right|^2\right\}=0.$$

But this contradicts assumption (B) again. Therefore $\varepsilon^*(V_1,F_0)\geq\varepsilon_0^*(F_0)$ and the proof is complete. □

Proof of Lemma 1 Since

$$|\alpha_n-\alpha_0|\leq\sup_{u\geq 0}|G_n(u)-G(\gamma u)|$$

we will show that the right-hand side is $o_P(1)$. Consider the family of functions

$$\mathcal{F}=\left\{f(x,t,V,u)=I\left(d^2(x,t,V)\leq u\right):t\in\mathcal{R}^p,V\in\mathrm{SPD}(p),u\geq 0\right\}.$$

We can write

$$G_n(u)-G(\gamma u)=E_nf(\cdot,t_{0n},V_{0n},u)-Ef(\cdot,\mu,\gamma\Sigma,u)$$

where E_n denotes expectation with respect to the empirical measure (in the variable x only) and E denotes expectation with respect to the model distribution. Since $\mathcal{F}$ is a VC-subgraph class of functions (this follows easily from Lemma 2.6.15 in [27]) then $\mathcal{F}$ is a Glivenko-Cantelli class by Theorem 2.4.3 in [27] and we have that

$$\sup_{f\in\mathcal{F}}|E_nf-Ef|\to 0\text{ almost surely.}$$

Also, by dominated convergence and continuity of G we have

$$\sup_{u\geq 0}|E\{f(\cdot,t_{0n},V_{0n},u)-f(\cdot,\mu,\gamma\Sigma,u)\}|\to 0\text{ in probability.}$$

These two convergencies together imply that $\alpha_n\to\alpha_0$ in probability. That $c_n\to c_0$ in probability will follow from the fact that G^{-1} is continuous when G is strictly increasing. □

Proof of Theorem 3 Let us write

$$\begin{aligned}t_{1n}&=\frac{E_n\Psi_2(\cdot,t_{0n},V_{0n},c_n)}{E_n\Psi_1(\cdot,t_{0n},V_{0n},c_n)}\\V_{1n}&=\frac{E_n\Psi_3(\cdot,t_{0n},V_{0n},c_n,t_{1n})}{E_n\Psi_1(\cdot,t_{0n},V_{0n},c_n)}\end{aligned}$$

with

$$\begin{aligned}
\Psi_1(x,t,V,c) &= w\left(d^2(x,t,V)/c\right) \\
\Psi_2(x,t,V,c) &= w\left(d^2(x,t,V)/c\right)x \\
\Psi_3(x,t,V,c,m) &= w\left(d^2(x,t,V)/c\right)(x-m)(x-m)'.
\end{aligned}$$

The families of functions

$$\left\{\Psi_j(x,t,V,c) : t \in \mathcal{R}^p, V \in \mathrm{SPD}(p),\, c \geq 0\right\}$$

for $j = 1, 2$ and

$$\left\{\Psi_3(x,t,V,c,m) : t, m \in \mathcal{R}^p, V \in \mathrm{SPD}(p),\, c \geq 0\right\}$$

are VC-subgraph with square-integrable envelopes and hence Donsker, according to Theorem 2.5.2 in [27]. Observe that

$$\begin{aligned}
E_n\Psi_2(\cdot,t_{0n},V_{0n},c_n) &= E_n\Psi_2\left(\cdot,t_{0n},\gamma^{-1}V_{0n},\gamma c_n\right) \\
&= E\Psi_2\left(\cdot,t_{0n},\gamma^{-1}V_{0n},\gamma c_n\right) + (E_n - E)\Psi_2(\cdot,\mu,\Sigma,\gamma c_0) \\
&\quad + (E_n - E)\left\{\Psi_2\left(\cdot,t_{0n},\gamma^{-1}V_{0n},\gamma c_n\right) - \Psi_2(\cdot,\mu,\Sigma,\gamma c_0)\right\}.
\end{aligned}$$

Since $\gamma^{-1}V_{0n} \to \Sigma$ in probability we can use the results of Lopuhaä [12] in a straightforward manner. From his Lemma 3.1 we obtain

$$E\Psi_2\left(\cdot,t_{0n},\gamma^{-1}V_{0n},\gamma c_n\right) = a_2(\gamma c_n)(t_{0n} - \mu) + o_P\left(\|(t_{0n} - \mu, B_{0n} - B)\|\right)$$

with $a_2(\gamma c_n) \to a_2(\gamma c_0)$ in probability by Lemma 1. The second term is just

$$(E_n - E)\Psi_2(\cdot,\mu,\Sigma,\gamma c_0) = \frac{1}{n}\sum_{i=1}^{n} w\left(\frac{d^2(X_i,\mu,\Sigma)}{\gamma c_0}\right)(X_i - \mu).$$

For the third term we use the Donsker property: since

$$E\left|\Psi_2\left(\cdot,t_{0n},\gamma^{-1}V_{0n},\gamma c_n\right) - \Psi_2(\cdot,\mu,\Sigma,\gamma c_0)\right|^2 \to 0 \text{ in probability}$$

we have that

$$\sqrt{n}(E_n - E)\left\{\Psi_2\left(\cdot,t_{0n},\gamma^{-1}V_{0n},\gamma c_n\right) - \Psi_2(\cdot,\mu,\Sigma,\gamma c_0)\right\} = o_P(1).$$

In a similar way it is shown that

$$E_n\Psi_1\left(\cdot,t_{0n},\gamma^{-1}V_{0n},\gamma c_n\right) = a_1(\gamma c_0) + o_P(1)$$

and then the expansion for t_{1n} follows. For V_{1n} we proceed similarly, now using that $t_{1n} \to \mu$ in probability and the expansion

$$E\Psi_3\left(\cdot,t_{0n},\gamma^{-1}V_{0n},\gamma c_n,t_{1n}\right) =$$

$$a_3(\gamma c_n)\Sigma + a_4(\gamma c_n)\left\{\mathrm{tr}\left(B^{-1}(B_{0n}-B)\right)\Sigma + 2B^{-1}(B_{0n}-B)\Sigma\right\}$$
$$+o_P\left(\|(t_{0n}-\mu, B_{0n}-B, t_{1n}-\mu)\|\right).$$

(Note that $B^{-1}(B_{0n}-B) = (B_{0n}-B)B^{-1} = A_n$ in Lopuhaä's notation.) □

Proof of Theorem 4 We will use the notation of the proof of Theorem 3 again, as well as the results of Lemma 3.1 in [12]. First note that if $F_{\mu,\Sigma}$ denotes an elliptical distribution with parameters μ and Σ, we have

$$\begin{aligned}\mathrm{IF}(x, t, F_{\mu,\Sigma}) &= B\,\mathrm{IF}\left(B^{-1}(x-\mu), t, F_{0,\mathrm{I}}\right)\\ \mathrm{IF}(x, V, F_{\mu,\Sigma}) &= B\,\mathrm{IF}\left(B^{-1}(x-\mu), V, F_{0,\mathrm{I}}\right)B\end{aligned}$$

for all equivariant location and scatter estimators. Therefore the proof will be done for $\mu = 0$ and $\Sigma = \mathrm{I}$. Let $F_{\varepsilon,x} = (1-\varepsilon)F_{0,\mathrm{I}} + \varepsilon\Delta_x$. Since

$$t_1(F_{\varepsilon,x}) = \frac{(1-\varepsilon)E\Psi_2\left(\cdot, t_0(F_{\varepsilon,x}), \gamma^{-1}V_0(F_{\varepsilon,x}), \gamma c(F_{\varepsilon,x})\right) + \varepsilon w_{F_{\varepsilon,x}}(x)\,x}{(1-\varepsilon)E\Psi_1\left(\cdot, t_0(F_{\varepsilon,x}), \gamma^{-1}V_0(F_{\varepsilon,x}), \gamma c(F_{\varepsilon,x})\right) + \varepsilon w_{F_{\varepsilon,x}}(x)}$$

(with expectations taken with respect to $F_{0,\mathrm{I}}$) and

$$E\Psi_2\left(\cdot, t_0(F_{\varepsilon,x}), \gamma^{-1}V_0(F_{\varepsilon,x}), \gamma c(F_{\varepsilon,x})\right) = a_2(\gamma c(F_{\varepsilon,x}))\,t_0(F_{\varepsilon,x}) + o(\varepsilon),$$

then $\lim_{\varepsilon\downarrow 0} t_1(F_{\varepsilon,x})/\varepsilon$ exists for those x such that $\|x\|^2/\gamma c_0$ is a continuity point of w, and equals $\mathrm{IF}(x, t_1, F_{0,\mathrm{I}})$ as given in the statement of the theorem, provided

$$\lim_{\varepsilon\downarrow 0} c(F_{\varepsilon,x}) = c_0. \tag{18}$$

We prove this at the end. Now, to obtain $\mathrm{IF}(x, V_1, F_{0,\mathrm{I}})$ we proceed analogously, using the expansion

$$E\Psi_3\left(\cdot, t_0(F_{\varepsilon,x}), \gamma^{-1}V_0(F_{\varepsilon,x}), \gamma c(F_{\varepsilon,x}), t_1(F_{\varepsilon,x})\right) =$$
$$a_3(\gamma c(F_{\varepsilon,x}))\,\mathrm{I} + a_4(\gamma c(F_{\varepsilon,x}))\left\{\mathrm{tr}\left(A_0(F_{\varepsilon,x})\right)\mathrm{I} + 2A_0(F_{\varepsilon,x})\right\} + o(\varepsilon),$$

where $A_0(F_{\varepsilon,x}) = B^{-1}\left(B_0(F_{\varepsilon,x}) - B\right) = \left(B_0(F_{\varepsilon,x}) - B\right)B^{-1}$ and $B_0^2(F_{\varepsilon,x}) = \gamma^{-1}V_0(F_{\varepsilon,x})$, and consequently the fact that

$$\begin{aligned}\mathrm{IF}\left(x, \gamma^{-1}V_0, F_{0,\mathrm{I}}\right) &= B\,\mathrm{IF}(x, B_0, F_{0,\mathrm{I}}) + \mathrm{IF}(x, B_0, F_{0,\mathrm{I}})\,B\\ &= 2B\,\mathrm{IF}(x, B_0, F_{0,\mathrm{I}})\\ &= 2B\,\mathrm{IF}(x, A_0, F_{0,\mathrm{I}})\,B.\end{aligned}$$

To complete the proof let us show (18). Actually we prove that

$$\lim_{\varepsilon\downarrow 0}\sup_{u\geq 0}\left|G_{F_{\varepsilon,x}}(u) - G_F(u)\right| = 0 \tag{19}$$

which implies that $\lim_{\varepsilon\downarrow 0}\alpha(F_{\varepsilon,x}) = \alpha_0$ and then (18) follows by continuity of G^{-1}. To prove (19) simply note that

$$\left|G_{F_{\varepsilon,x}}(u) - G_F(u)\right|$$

$$\leq (1-\varepsilon)\left|P_F\left(d^2\left(X, t_0(F_{\varepsilon,x}), V_0(F_{\varepsilon,x})\right) \leq u\right) - G_F(u)\right| + \varepsilon$$

which goes to zero when $\varepsilon \downarrow 0$, uniformly in u by continuity of G_F. $\square$

Derivation of expressions (16) and (17) Assume again that $\mu = 0$ and $\Sigma = \mathrm{I}$. Let t_0 and V_0 denote the location and scatter MCD functionals respectively, so that $V_0(F) = \gamma \mathrm{I}$. Croux and Haesbroeck [2] give the following expressions for their influence functions at F:

$$\begin{aligned} IF(x, t_0, F) &= \frac{1}{a_1(q) - a_2(q)} I\left(\|x\|^2 \leq q\right) x \\ IF\left(x, \gamma^{-1} V_0, F\right) &= \frac{1}{a_3(q) - a_4(q)} I\left(\|x\|^2 \leq q\right) xx' + v(\|x\|)\,\mathrm{I} \end{aligned}$$

for a certain real function v whose explicit expression is irrelevant for us. It follows that

$$IF(x, t_1, F) = \left[\frac{a_2(\gamma c_0) I\left(\|x\|^2 \leq q\right)}{a_1(\gamma c_0)\{a_1(q) - a_2(q)\}} + \frac{w\left(\|x\|^2/\gamma c_0\right)}{a_1(\gamma c_0)}\right] x.$$

The asymptotic variance of t_1 is $E_F\left\{IF(X, t_1, F)\, IF(X, t_1, F)'\right\}$ and when $w(u) = I(u < 1)$ it is

$$E_F\left\{IF(X, t_1, F)\, IF(X, t_1, F)'\right\} = \kappa_1 \mathrm{I}$$

with κ_1 as in (16). For the scatter estimator we have that

$$IF(x, V_1, F) = l(\|x\|)\frac{xx'}{\|x\|^2} + m(\|x\|)\,\mathrm{I}$$

with

$$l(\|x\|) = \frac{a_4(\gamma c_0) I\left(\|x\|^2 \leq q\right)\|x\|^2}{a_1(\gamma c_0)\{a_3(q) - a_4(q)\}} + \frac{w\left(\|x\|^2/\gamma c_0\right)\|x\|^2}{a_1(\gamma c_0)}$$

and

$$m(\|x\|) = \frac{a_4(\gamma c_0)}{a_1(\gamma c_0)}\left(\frac{p}{2} + 1\right) v(\|x\|) + \frac{a_4(\gamma c_0) I\left(\|x\|^2 \leq q\right)\|x\|^2}{a_1(\gamma c_0) 2\{a_3(q) - a_4(q)\}} - \frac{a_3(\gamma c_0)}{a_1(\gamma c_0)}.$$

Using Lemma 5.1 in [11] we obtain that

$$E_F\left[\mathrm{vec}\{IF(X, V_1, F)\}\,\mathrm{vec}\{IF(X, V_1, F)\}'\right] = \sigma_1\left(\mathrm{I}_{p^2} + K_{p,p}\right) + \lambda_1 \mathrm{vec}(\mathrm{I})\,\mathrm{vec}(\mathrm{I})'$$

with

$$\sigma_1 = \frac{E_F\{l^2(\|x\|)\}}{p(p+2)}$$

and

$$\lambda_1 = \frac{E_F\{l^2(\|x\|)\}}{p(p+2)} + 2E_F\{l(\|x\|)\, m(\|x\|)\} + E_F\left\{m^2(\|x\|)\right\}.$$

This is the asymptotic variance of V_1. For the hard-rejection weight $w(u) = I(u < 1)$, σ_1 comes down to (17).

When w is the hard-rejection weight function and F is the $t_p(\nu)$ distribution with ν other than 2 or 4, we have the following expressions for the $a_i(u)$:

$$\begin{aligned}
a_1(u) &= G\left(\frac{u}{p}; p, \nu\right), \\
a_2(u) &= \frac{\Gamma\left(\frac{p+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{p}{2}+1\right)} \frac{(u/\nu)^{p/2}}{(1+u/\nu)^{(p+\nu)/2}}, \\
a_3(u) &= \frac{\nu}{\nu-2} G\left(\frac{\nu-2}{\nu}\frac{u}{p+2}; p+2, \nu-2\right), \\
a_4(u) &= \frac{\nu}{2}\frac{\Gamma\left(\frac{p+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{p}{2}+2\right)} \frac{(u/\nu)^{p/2+1}}{(1+u/\nu)^{(p+\nu)/2}}, \\
a_5(u) &= \frac{\nu^2}{(\nu-2)(\nu-4)} G\left(\frac{\nu-4}{\nu}\frac{u}{p+4}; p+4, \nu-4\right),
\end{aligned}$$

where $G(u; p, \nu)$ is the distribution function of an $\mathcal{F}(p, \nu)$ distribution, which can be formally defined even for negative ν. For F the $N_p(\mu, \Sigma)$ distribution, $a_5(u) = G_{p+4}(u)$. □

Acknowledgments

The author thanks Professor Steve Portnoy for carefully reading the preliminar draft of this work and making many suggestions for improvement. Comments by the referees were also very helpful

References

[1] R. W. Butler, P. L. Davies, M. Jhun, Asymptotics for the minimum covariance determinant estimator, *Ann. Statist.* **21** (1993), 1385-1400.

[2] C. Croux, G. Haesbroeck, Influence function and efficiency of the minimum covariance determinant scatter matrix estimator, *J. Multivariate Anal.* **71** (1999), 161-190 doi: 10.1006/jmva.1999.1839.

[3] P. L. Davies, Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices, *Ann. Statist.* **15** (1987), 1269-1292.

[4] P. L. Davies, The asymptotics of Rousseeuw's minimum volume ellipsoid estimator, *Ann. Statist.* **20** (1992), 1828-1843.

[5] D. L. Donoho, Breakdown properties of multivariate location estimators, Ph.D. qualifying paper, Harvard University, 1982.

[6] D. L. Donoho, P. J. Huber, The notion of breakdown point, *in* "A Festschrift for Erich L. Lehmann" (P. J. Bickel, K. A. Docksum, J. L. Hodges, Jr., Eds.), pp. 157-184, Wadsworth, Belmont, CA, 1983.

[7] D. Gervini, V. J. Yohai, A class of robust and fully efficient regression estimators, to appear in *Ann. Statist.* **30** (2002).

[8] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel, "Robust Statistics: The Approach Based on Influence Functions," Wiley, New York, 1986.

[9] X. He, A local breakdown property of robust tests in linear regression, *J. Multivariate Anal.* **38** (1991), 294-305.

[10] J. Jurecková, S. Portnoy, Asymptotics for one-step M-estimators in regression with application to combining efficiency and high breakdown point, *Comm. Statist. Theory Methods* **16** (1987), 2187-2199.

[11] H. P. Lopuhaä, On the relation between S-estimators and M-estimators of multivariate location and covariance, *Ann. Statist.* **17** (1989), 1662-1683.

[12] H. P. Lopuhaä, Asymptotics of reweighted estimators of multivariate location and scatter, *Ann. Statist.* **27** (1999), 1638-1665.

[13] H. P. Lopuhaä, P. J. Rousseeuw, Breakdown points of affine equivariant estimators of multivariate location and covariance matrices, *Ann. Statist.* **19** (1991), 229-248.

[14] R. A. Maronna, Robust M-estimators of multivariate location and scatter, *Ann. Statist.* **4** (1976), 51-67.

[15] R. A. Maronna, V. J. Yohai, The behavior of the Stahel-Donoho robust multivariate estimator, *J. Amer. Statist. Assoc.* **90** (1995), 330-341.

[16] R. A. Maronna, V. J. Yohai, Robust estimation of multivariate location and scatter, *in* "Encyclopedia of Statistical Sciences, Update Volume 2" (S. Kotz, C. Read, D. Banks, Eds.), pp.589-596, Wiley, New York, 1998.

[17] D. M. Rocke, Robustness properties of S-estimators of multivariate location and shape in high dimension, *Ann. Statist.* **24** (1996), 1327-1345.

[18] D. M. Rocke, D. L. Woodruff, Identification of outliers in multivariate data, *J. Amer. Statist. Assoc.* **91** (1996), 1047-1061.

[19] P. J. Rousseeuw, Least median of squares regression, *J. Amer. Statist. Assoc.* **79** (1984), 871-880.

[20] P. J. Rousseeuw, Multivariate estimation with high breakdown point, *in* "Mathematical Statistics and Applications, Vol B" (W. Grossman, G. Pflug, I. Vincze, W. Wertz, Eds.), pp. 283-297, I. Reidel, Dordrecht, 1985.

[21] P. J. Rousseeuw, Unconventional features of positive-breakdown estimators, *Statist. Probab. Lett.* **19** (1994), 417-431.

[22] P. J. Rousseeuw, A. M. Leroy, "Robust Regression and Outlier Detection", Wiley, New York, 1987.

[23] P. J. Rousseeuw, K. van Driessen, A fast algorithm for the minimum covariance determinant estimator, *Technometrics* **41** (1999), 212-224.

[24] P. J. Rousseeuw, B. C. van Zomeren, Unmasking multivariate outliers and leverage points, *J. Amer. Statist. Assoc.* **85** (1990), 633-639.

[25] D. G. Simpson, D. Ruppert, R. J. Carroll, On one-step GM-estimates and stability of inferences in linear regression, *J. Amer. Statist. Assoc.* **87** (1992), 439-450.

[26] W. A. Stahel, Breakdown of covariance estimators, Research Report 31, Fachgruppe für Statistik, E.T.H. Zürich, 1981.

[27] A. W. van der Vaart, J. A. Wellner, "Weak Convergence and Empirical Processes", Springer, New York, 1996.

[28] D. E. Tyler, Robustness and efficiency properties of scatter matrices, *Biometrika* **70** (1983), 411-420.

[29] D. L. Woodruff, D. M. Rocke, Computable robust estimation of multivariate location and shape in high dimension using compound estimators, *J. Amer. Statist. Assoc.* **89** (1994), 888-896.

[30] V. J. Yohai, High breakdown point and high efficiency robust estimates for regression, *Ann. Statist.* **15**, 642-656.

[31] V. J. Yohai, R. H. Zamar, High breakdown point estimates of regression by means of the minimization of an efficient scale, *J. Amer. Statist. Assoc.* **83**, 406-413.

Table 1: Relative MSE of location estimators for the multivariate normal.

	$p = 3$*				$p = 10$			
Estimator	50	100	200	500	50	100	200	500
MVE	.23	.16	.11	.06	.26	.14	.07	.03
MVE-1F	.61	.70	.72	.75	.66	.73	.72	.65
MVE-1A	.63	.74	.79	.86	.67	.75	.75	.71
MCD	.32	.27	.22	.17	.50	.42	.37	.34
MCD-1F	.55	.66	.73	.81	.52	.61	.78	.89
MCD-1A	.57	.69	.79	.90	.52	.63	.81	.94
S	.71	.71	.70	.74	.93	.93	.93	.95
S-1F	.86	.87	.89	.92	.95	.94	.95	.96
S-1A	.88	.92	.95	.97	.95	.96	.98	.99

* Each sub-column represents a different sample size.

Table 2: Relative mean LCN of scatter estimators for the multivariate normal.

	$p = 3$*				$p = 10$			
Estimator	50	100	200	500	50	100	200	500
MVE	.32	.28	.23	.16	.40	.29	.21	.13
MVE-1F	.58	.65	.67	.65	.65	.72	.73	.66
MVE-1A	.59	.69	.74	.77	.66	.74	.76	.71
MCD	.30	.27	.24	.22	.43	.42	.42	.40
MCD-1F	.51	.59	.66	.70	.44	.58	.78	.88
MCD-1A	.52	.63	.72	.83	.44	.60	.81	.92
S	.72	.74	.76	.77	.94	.95	.96	.96
S-1F	.81	.85	.87	.88	.92	.94	.95	.96
S-1A	.83	.88	.92	.95	.93	.96	.98	.99

* Each sub-column represents a different sample size.

Table 3: Errors of location estimators for multivariate Cauchy and shifted normal.

	$p = 3$			$p = 10$		
Estimator	CAU*	SN(.10)**	SN(.20)**	CAU	SN(.10)	SN(.20)
MEAN	.01	4.08	16.14	.01	4.17	16.16
MVE	.17	.24	.31	.06	.99	5.91
MVE-1F	.56	.09	.19	.29	.37	4.78
MVE-1A	.55	.09	.19	.29	.37	4.82
MCD	.67	.17	.21	.55	.40	.63
MCD-1F	.67	.10	.17	.55	.39	.62
MCD-1A	.64	.10	.17	.55	.39	.62
S	.81	.08	.17	.63	.30	11.19
S-1F	.59	.08	.29	.47	.35	14.67
S-1A	.59	.08	.29	.47	.36	14.82

* Column reports relative median SE with respect to the Cauchy MLE.
** Column reports maximum median SE over contamination grid as explained in text.

Table 4: Errors of scatter estimators for multivariate Cauchy and shifted normal.

	$p = 3$			$p = 10$		
Estimator	CAU*	SN(.10)**	SN(.20)**	CAU	SN(.10)	SN(.20)
COV	.25	3.87	4.43	.30	4.60	5.16
MVE	.37	1.83	1.94	.30	4.55	6.14
MVE-1F	.57	1.04	1.24	.44	2.92	5.05
MVE-1A	.58	1.02	1.23	.44	2.89	5.06
MCD	.51	2.08	2.09	.58	3.79	4.05
MCD-1F	.64	1.18	1.26	.58	3.69	3.95
MCD-1A	.64	1.14	1.24	.58	3.67	3.94
S	.73	.87	1.16	.70	2.20	5.14
S-1F	.67	.84	1.18	.64	2.30	5.19
S-1A	.67	.84	1.18	.64	2.30	5.18

* Column reports relative median LCN with respect to the Cauchy MLE.
** Column reports maximum median LCN over contamination grid as explained in text.

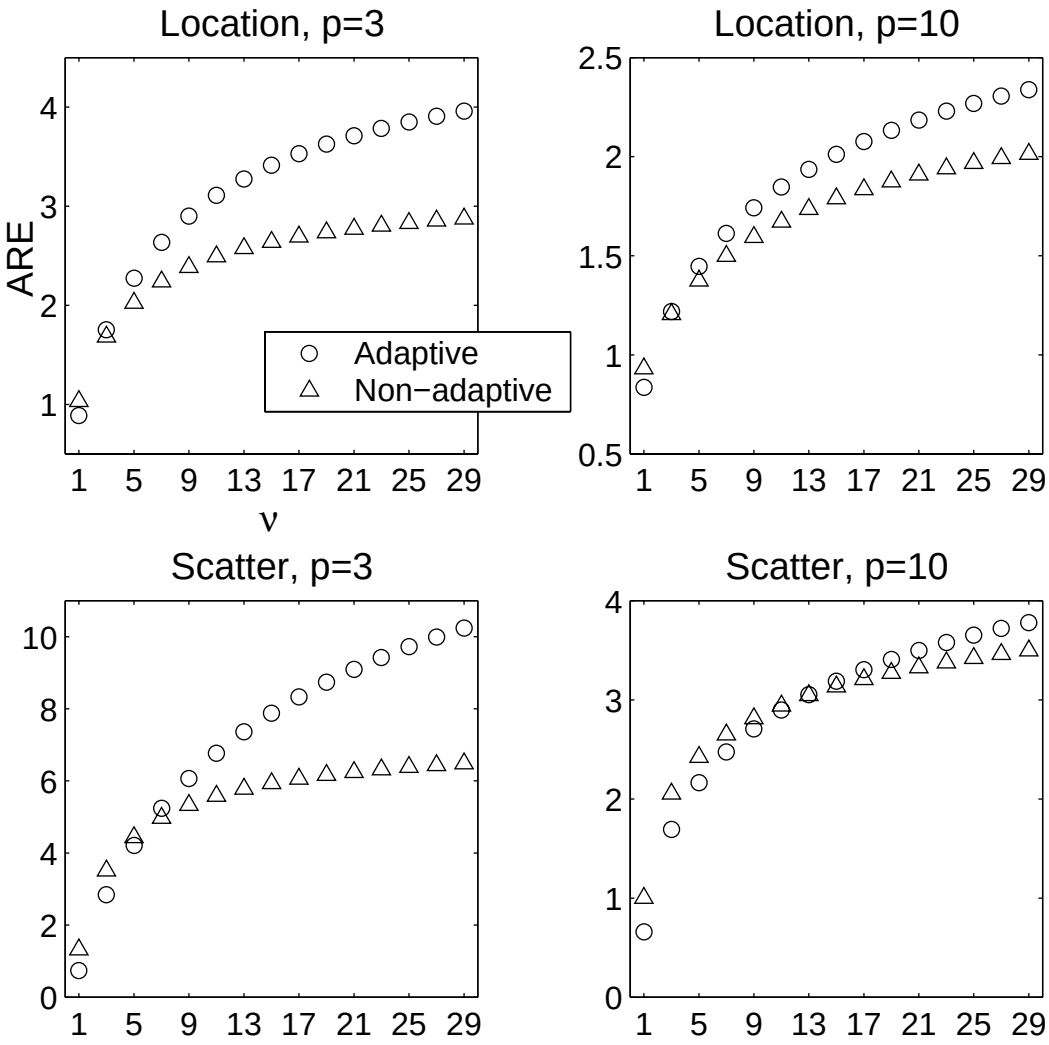

Figure 1: AREs of one-step estimators for $t_p(\nu)$ distributions.

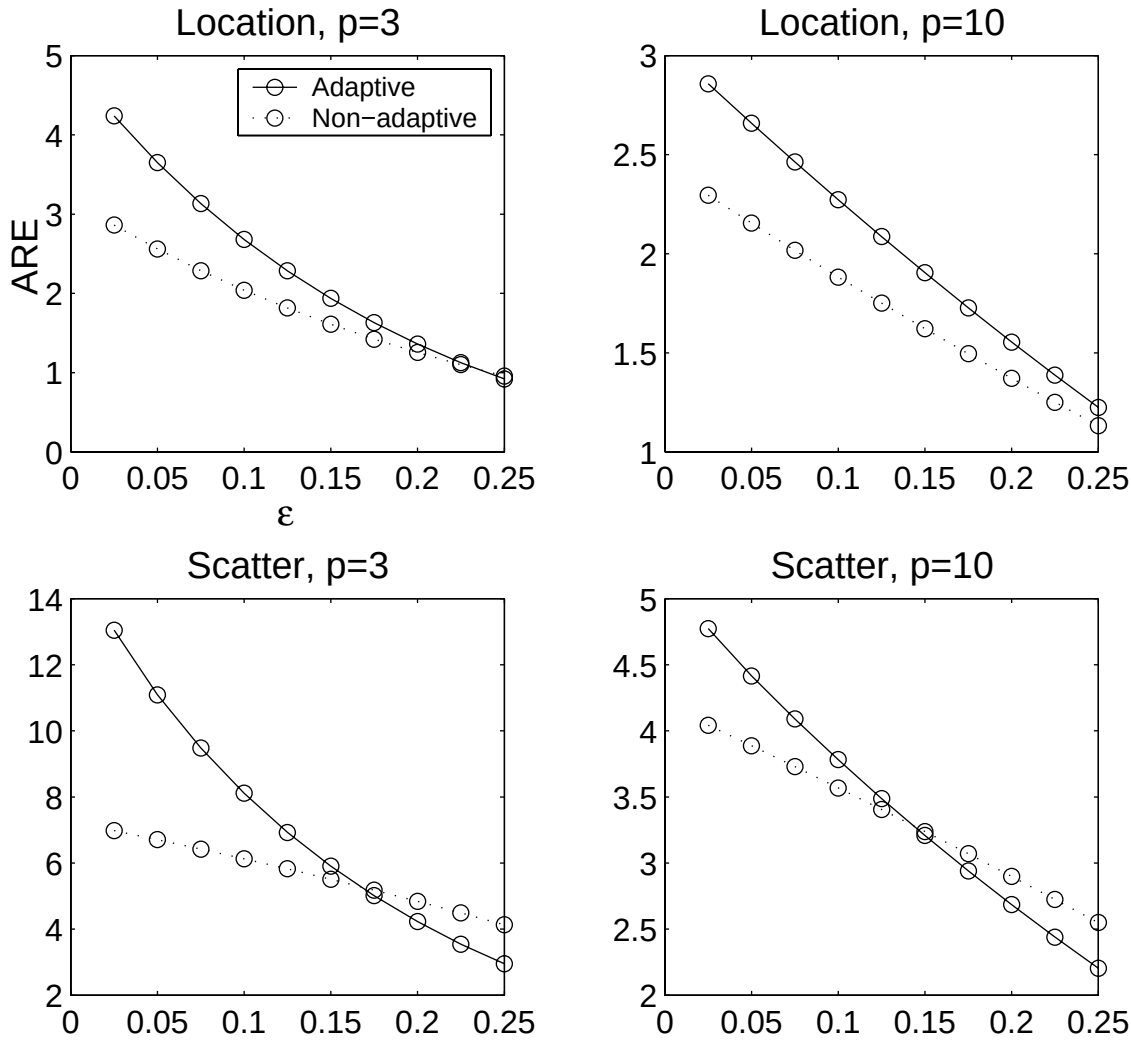

Figure 2: AREs of one-step estimators for contaminated normal distributions.

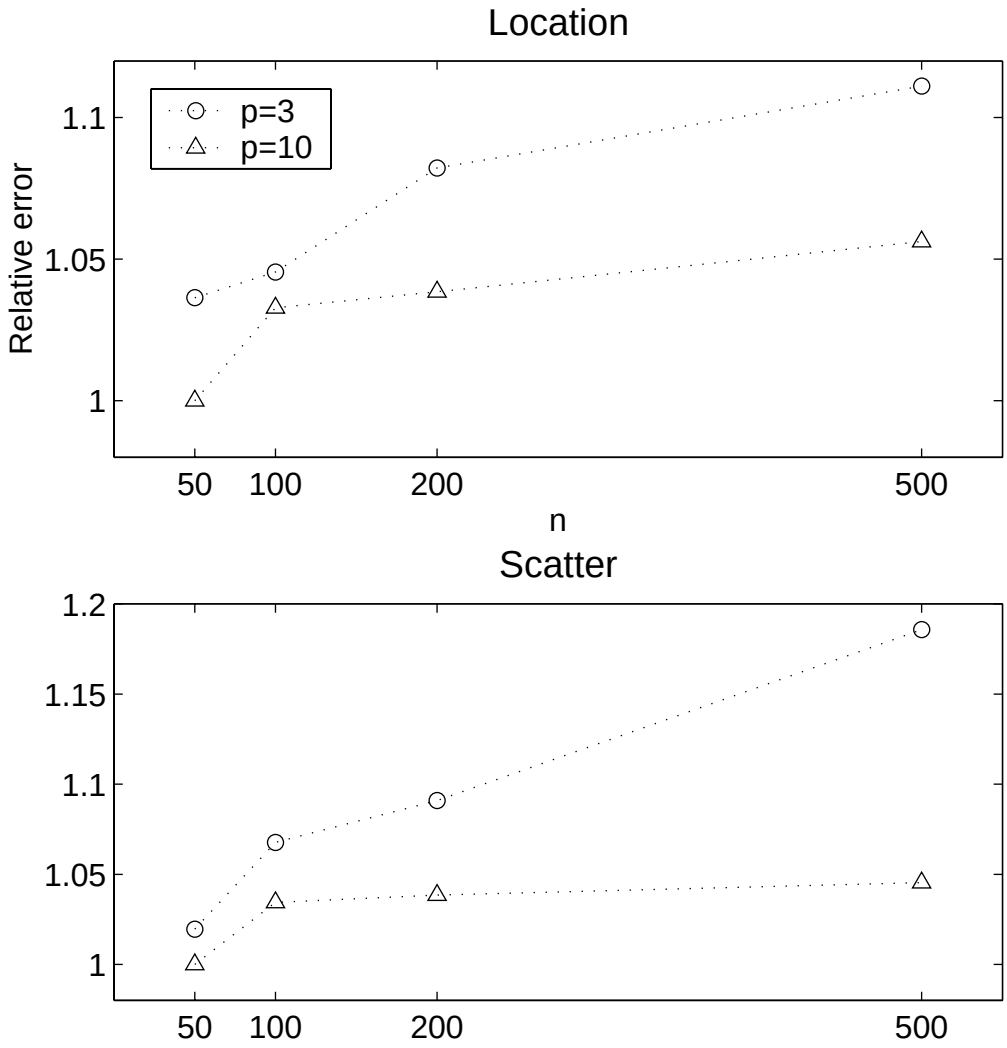

Figure 3: Relative errors of adaptive one-step estimator with respect to the non-adaptive one when MCD is the initial estimator and the distribution is normal.