# The influence function of the Stahel-Donoho estimator of multivariate location and scatter

Daniel Gervini *

University of Zurich

**Abstract**

This article derives the influence function of the Stahel-Donoho estimator of multivariate location and scatter for elliptical distributions. Local robustness and asymptotic relative efficiency are studied. The expressions obtained for the influence functions coincide with those of one-step reweighted estimators.

*Key words and phrases:* asymptotic relative efficiency, gross-error sensitivity, local robustness, reweighted estimators.

*Department of Biostatistics, Sumatrastrasse 30, 8006 Zurich, Switzerland

# 1 Introduction

The statistical analysis of multidimensional data typically relies, explicitly or implicitly, on estimators of location and scatter. For example, principal component analysis, canonical correlation, and discrimination, are all techniques based on the scatter estimator. The most common estimators are the sample mean and the sample covariance matrix, which are optimal if the data follows a normal distribution but are otherwise extremely sensitive to outliers. Since visual detection of outliers is unworkable in high dimensions, robust estimators are indispensable. An overview of robust estimation in the multivariate setting is given in Maronna and Yohai (1998). Among the estimators considered in that article, those proposed independently by Stahel (1981) and Donoho (1982) (which will be called S-D estimators hereafter) possess very good robust properties. They were the first equivariant estimators with maximum asymptotic breakdown point 1/2 (see Hampel, Rousseeuw, Ronchetti and Stahel, 1986, Theorem 5.5.3), and were in fact introduced for that reason. More recently, Maronna and Yohai (1995) computed the maximum bias function for contaminated normal models and carried out extensive simulations, showing that properly tuned S-D estimators outperform other robust estimators in many situations. Nevertheless, these estimators have not received much attention in comparison with other robust proposals. This is probably due to the practical difficulties involved in their computation. Another reason may have been the lack of asymptotic theory, which has been solved in part in a as yet unpublished manuscript by Zuo, Cui and He (2002).

In this article we derive the influence function of the S-D estimator for elliptical models. The influence function allows us to study the local robustness and the asymptotic efficiency of these estimators, providing a rationale for choosing appropriate weight functions and tuning parameters. It turns out that properly tuned S-D estimators sometimes outperform, and are always at least comparable to, the more popular S-estimators (Davies, 1987).

This article is organized as follows. Section 2 formally introduces the S-D estimator and reviews some basics about elliptical distributions and influence functions. Section 3 gives some non-standard results about univariate location-scale estimators of projections that are needed in Section 4, where the influence function of the S-D estimator is derived and the resulting properties are discussed. Proofs are left to the Appendix.

## 2 Definitions

The Stahel-Donoho estimator of location and scatter is defined as follows. Given a sample $\mathbf{X} = (x_1, \ldots, x_n)$ in $\mathbb{R}^p$, consider all univariate projections $a^\top \mathbf{X}$ with $a \in \mathbb{R}^p$, $a \neq 0$. Let $m(\cdot)$ and $s(\cdot)$ be univariate estimators of location and scale. Then, a measure of outlyingness of each $a^\top x_i$ is the standardized distance $\left|a^\top x_i - m(a^\top \mathbf{X})\right| / s(a^\top \mathbf{X})$. The overall outlyingness of the point $x_i$ is then defined as

$$r(x_i; \mathbf{X}) = \sup_{a \in \mathbb{R}^p} \left| \frac{a^\top x_i - m(a^\top \mathbf{X})}{s(a^\top \mathbf{X})} \right|.$$

The S-D estimator downweights each observation according to its overall outlyingness. The location estimator $T_n$ is then a weighted mean and the scatter estimator $V_n$ is a weighted covariance matrix:

$$\begin{aligned} T_n &= \frac{\text{ave}\{w(r^2(x_i; \mathbf{X}))x_i\}}{\text{ave}\{w(r^2(x_i; \mathbf{X}))\}} \\ V_n &= \frac{\text{ave}\{w(r^2(x_i; \mathbf{X}))(x_i - T(\mathbf{X}))(x_i - T(\mathbf{X}))^\top\}}{\text{ave}\{w(r^2(x_i; \mathbf{X}))\}}, \end{aligned}$$

where $w$ is a non-negative and usually non-increasing function. In what follows, we will express estimators as functionals on the space of distribution functions. Thus $T_n = T(F_n)$ and $V_n = V(F_n)$, where $F_n$ denotes the empirical distribution function of $\mathbf{X}$.

To analyze the asymptotic properties of these functionals, we will restrict ourselves to the class of elliptical distributions as target models. These models are flexible enough to accommodate heavy-tailed distributions without finite moments, while possessing well defined location and scatter parameters. A random vector $Y \in \mathbb{R}^p$ has an elliptical distribution if its density is of the form

$$f(y) = |\Sigma|^{-\frac{1}{2}} h((y-\mu)^\top \Sigma^{-1} (y-\mu)), \tag{1}$$

where $h$ is a non-negative and integrable function, $\mu \in \mathbb{R}^p$ is the location parameter, and $\Sigma \in \mathbb{SP}(p)$ is the scatter parameter ($\mathbb{SP}(p)$ denotes the family of symmetric positive definite matrices). The vector $X = \Sigma^{-\frac{1}{2}}(Y - \mu)$, where $\Sigma^{\frac{1}{2}} \in \mathbb{SP}(p)$ is the unique symmetric square root of $\Sigma$, has density $f(x) = h(\|x\|^2)$ and its distribution is called spherical. A more general definition of sphericity, which does not need existence of a density function, requires that the distribution of $X$ be rotationally invariant. That is, $\mathcal{L}(\Gamma X) = \mathcal{L}(X)$ for every $\Gamma \in \mathbb{O}(p)$, where $\mathbb{O}(p)$ is the family of orthogonal matrices. Several properties of spherical distributions are given in Bilodeau and Brenner (1999, chapters 4 and 13) and Hampel et al. (1986, chapter 5). Two of these, which will be used many times in this paper, are that $\mathcal{L}(a^\top X) = \mathcal{L}(\|a\| X_1)$ for any $a \in \mathbb{R}^p$, with $\mathcal{L}(X_1)$ symmetric about zero, and that $X$ can be factorized as $X = RU$, where $R = \|X\|$ is stochastically independent of $U = X/\|X\|$ and $U$ has the uniform distribution on $\mathcal{S}^{p-1}$ (the unit sphere in $\mathbb{R}^p$). Then $E(U) = 0$ and $E(UU^\top) = p^{-1}I$. Moreover, if $W = (U_1, \ldots, U_k)^\top$ with $1 \le k < p$ then $\|W\|^2 \sim \mathcal{B}(k/2, (p-k)/2)$ and $W$ has density

$$f(w) = \frac{\Gamma(\frac{p}{2})}{\pi^{\frac{k}{2}}\Gamma(\frac{p-k}{2})}\,(1 - \|w\|^2)^{\frac{p-k}{2}-1}, \quad 0 < \|w\|^2 < 1.$$

$R^2$ has density

$$f_{R^2}(t) = \frac{\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})}\, t^{\frac{p}{2}-1} h(t).$$

Two elliptical models of importance are the multivariate normal $\mathcal{N}_p(\mu, \Sigma)$, whose density is determined by

$$h(t) = \frac{e^{-\frac{t}{2}}}{(2\pi)^{\frac{p}{2}}},$$

and the multivariate $\mathcal{T}$ on $\nu$ degrees of freedom, $\mathcal{T}_{p,\nu}(\mu, \Sigma)$, whose density is determined by

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

If $X \sim \mathcal{N}_p(0, I)$ then $R^2 \sim \chi^2_p$ and the marginals of $X$ are independent $\mathcal{N}_1(0, 1)$. The $\mathcal{T}_{p,\nu}(0, I)$ distribution can be characterized as the distribution of $X = Z/(Q/\nu)^{\frac{1}{2}}$ with $Z \sim \mathcal{N}_p(0, I)$, $Q \sim \chi^2_\nu$, $Z$ and $Q$ independent; then $R^2/p \sim \mathcal{F}_{p,\nu}$, or equivalently $(1 + \nu/R^2)^{-1} \sim \mathcal{B}(p/2, \nu/2)$, and the marginals of $X$ have $\mathcal{T}_{1,\nu}$ distributions (but are not independent).

Let us turn now to the influence function. Let $T$ and $V$ be location and scatter functionals which are well defined on elliptical distributions and their point-mass contaminations $F_{\varepsilon,z} = (1-\varepsilon)F + \varepsilon\Delta_z$, for small enough $\varepsilon$. The influence function of $T$ at $F$ is

$$IF(z;T,F) = \lim_{\varepsilon\downarrow 0} \frac{T(F_{\varepsilon,z}) - T(F)}{\varepsilon},$$

whenever the limit exists. The definition of $IF(z;V,F)$ is analogous. We will restrict ourselves to affine equivariant estimators. Therefore, if $F_{\mu,\Sigma}$ is an elliptical distribution as in (1), we have that $T(F_{\mu,\Sigma}) = \Sigma^{\frac{1}{2}}T(F_{0,I}) + \mu$ and $V(F_{\mu,\Sigma}) = \Sigma^{\frac{1}{2}}V(F_{0,I})\Sigma^{\frac{1}{2}}$, which implies that

$$\begin{aligned} IF(z;T,F_{\mu,\Sigma}) &= \Sigma^{\frac{1}{2}} IF(\Sigma^{-\frac{1}{2}}(z-\mu);T,F_{0,I}) \\ IF(z;V,F_{\mu,\Sigma}) &= \Sigma^{\frac{1}{2}} IF(\Sigma^{-\frac{1}{2}}(z-\mu);V,F_{0,I})\Sigma^{\frac{1}{2}}, \end{aligned}$$

where $F_{0,I}$ is spherical. Thus, without loss of generality we can and will work exclusively with spherical distributions from now on. For arbitrary equivariant estimators and spherical distributions Hampel et al. (1986, chapter 5.3) show that $T(F) = 0$ and $V(F) = \beta_{V,F}I$ for some $\beta_{V,F} > 0$, and that the influence functions are of the form

$$IF(z;T,F) = z w_\mu(\|z\|^2) \tag{2}$$

$$IF(z;V,F) = \Big(zz^\top - \frac{\|z\|^2}{p} I\Big) w_\eta(\|z\|^2) + \frac{1}{p} I w_\tau(\|z\|^2), \tag{3}$$

for non-negative functions $w_\mu$ and $w_\eta$, and a function $w_\tau$ such that $E\{w_\tau(R^2)\} = 0$. It is known that, under appropriate conditions, $T(F_n)$ is asymptotically normal with covariance matrix given by

$$\mathcal{C}_T = E_F\{IF(X;T,F)IF(X;T,F)^\top\} = d_\mu I, \tag{4}$$

with $d_\mu = E\{R^2 w_\mu^2(R^2)\}/p$. A similar result holds for $\operatorname{vecs} V(F_n)$, where vecs is the operation that stacks the $p+p(p-1)/2$ non-redundant elements of $V$ into a vector, as follows: $\operatorname{vecs} V = (v_{11}/\sqrt{2}, \ldots, v_{pp}/\sqrt{2}, v_{21}, v_{31}, v_{32}, \ldots, v_{p,p-1})^\top$. The asymptotic covariance of $\operatorname{vecs} V(F_n)$ is

$$\mathcal{C}_V = E_F\{\operatorname{vecs} IF(X;V,F) \operatorname{vecs} IF(X;V,F)^\top\} = d_\eta\Big(I - \frac{1}{p}ww^\top\Big) + d_\tau \frac{1}{p} ww^\top, \tag{5}$$

where $w^\top = (\mathbf{1}_p^\top, \mathbf{0}_{p(p-1)/2}^\top)$, $d_\eta = E\{R^4 w_\eta^2(R^2)\}/(p(p+2))$ and $d_\tau = E\{w_\tau^2(R^2)\}/(2p)$. Note that $\mathcal{C}_T^{-1}$ and $\mathcal{C}_V^{-1}$ have the same forms as in (4) and (5) with $d_\mu$, $d_\eta$ and $d_\tau$ replaced by their respective inverses.

Useful summaries of local robustness are the information-standardized gross-error sensitivities,

$$\begin{aligned}\gamma_T^*(F) &= \sup_{z\in\mathbb{R}^p}\{IF(z;T,F)^\top \mathcal{I}_1(F) IF(z;T,F)\}^{1/2}\\ \gamma_V^*(F) &= \sup_{z\in\mathbb{R}^p}\{\mathrm{vecs}(IF(z;V,F))^\top \mathcal{I}_2(F)\mathrm{vecs}(IF(z;V,F))\}^{1/2},\end{aligned}$$

where $\mathcal{I}_1$ and $\mathcal{I}_2$ are the Fisher information matrices of $F$ for location and scatter, namely $\mathcal{I}_1(F) = E\{\ell^2(\|X\|^2)XX^\top\}$ and $\mathcal{I}_2(F) = E(MM^\top)$, with $M = \mathrm{vecs}(-\frac{1}{2}I + \frac{1}{2}\ell(\|X\|^2)XX^\top)$ and $\ell(t) = -2h'(t)/h(t)$. These matrices have the forms (4) and (5) with $d_\mu^* = E\{R^2\ell^2(R^2)\}/p$, $d_\eta^* = E\{R^4\ell^2(R^2)\}/(p(p+2))$ and $d_\tau^* = E\{(R^2\ell(R^2)-p)^2\}/(2p) = ((p+2)d_\eta^* - p)/2$. Then

$$\begin{aligned}\gamma_T^*(F) &= \sup_{t\geq 0}|tw_\mu(t^2)|d_\mu^{*\frac{1}{2}}\\ \gamma_V^*(F) &= \sup_{t\geq 0}\Big\{\frac{1}{2}\Big(1-\frac{1}{p}\Big)t^4 w_\eta^2(t^2)d_\eta^* + \frac{1}{2p}w_\tau^2(t^2)d_\tau^*\Big\}^{\frac{1}{2}}.\end{aligned}$$

For the $\mathcal{N}_p(0,I)$ distribution is $d_\mu^* = d_\eta^* = d_\tau^* = 1$, while for the $\mathcal{T}_{p,\nu}(0,I)$ distribution is $d_\mu^* = d_\eta^* = (p+\nu)/(p+\nu+2)$ and $d_\tau^* = \nu/(p+\nu+2)$.

What is of primary concern in many applications is not the estimation of $\Sigma$ itself but of a shape component of $\Sigma$. A shape component is a function $S$ such that $S(\lambda\Sigma) = S(\Sigma)$ for all $\lambda > 0$. Two common examples are the eigenvectors and the ratios of eigenvalues. The influence function of a shape component does not depend on $w_\tau$ (Kent and Tyler, 1997). In particular, the influence function of $S(V) = V/(\mathrm{tr}\, V/p)$ is

$$IF(z;S,F) = \frac{1}{\beta_{V,F}}\Big(zz^\top - \frac{\|z\|^2}{p}I\Big)w_\eta(\|z\|^2)$$

and then $\mathcal{C}_S$ is like (5) with $d_{\tau,S} = 0$ and $d_{\eta,S} = d_{\eta,V}/\beta_{V,F}^2$. Then

$$\gamma_S^*(F) = \sup_{t\geq 0}|t^2 w_\eta(t^2)|\Big\{\frac{1}{2}\Big(1-\frac{1}{p}\Big)\frac{d_\eta^*}{\beta_{V,F}^2}\Big\}^{\frac{1}{2}}.$$

In this article we will use $\gamma_S^*$ instead of $\gamma_V^*$ to evaluate the robustness of the scatter estimator. An additional advantage of working with $S(V)$ instead of $V$ is that the asymptotic relative efficiency with respect to the maximum likelihood estimator is determined by a single scalar, because the asymptotic covariance is just a multiple of $I - ww^\top$. Thus, we have

$$\begin{aligned} ARE(T;F) &= \frac{1/d_\mu^*}{d_\mu} \\ ARE(S;F) &= \frac{1/d_\eta^*}{d_\eta/\beta_{V,F}^2}. \end{aligned}$$

To summarize, in this article we will judge the performance of the estimators on the basis of $\gamma_T^*$, $\gamma_V^*$, $ARE(T)$ and $ARE(S)$.

Going back to the S-D estimators, we have that their functional forms are given by

$$\begin{aligned} T(F) &= \frac{E_F\{w(r^2(X;F))X\}}{E_F\{w(r^2(X;F))\}} \\ V(F) &= \frac{E_F\{w(r^2(X;F))(X-T(F))(X-T(F))^\top\}}{E_F\{w(r^2(X;F))\}}, \end{aligned}$$

where

$$r(x;F) = \sup_{a\in\mathbb{R}^p}\left|\frac{a^\top x - m(F^a)}{s(F^a)}\right|.$$

Here $F^a = \mathcal{L}(a^\top X)$ and $m(\cdot)$ and $s(\cdot)$ are univariate location and scale functionals. The S-D estimators are well defined for any $F$ if the weight function satisfies:

**W1.** $w : [0,\infty) \to [0,\infty)$ is bounded and $w(u^2)u^2$ is bounded.

In order that $T$ and $V$ be affine equivariant we also have to assume that the univariate estimators $m$ and $s$ are equivariant, that is, $m(\mathcal{L}(\alpha Z+\beta)) = \alpha m(\mathcal{L}(Z))+\beta$ and $s(\mathcal{L}(\alpha Z+\beta)) = |\alpha| s(\mathcal{L}(Z))$ for any $\alpha,\beta\in\mathbb{R}$. This implies that $m(\mathcal{L}(Z)) = 0$ if $\mathcal{L}(Z)$ is symmetric about zero, and that

$$r(x;F) = \sup_{a\in\mathcal{S}_+^{p-1}}\left|\frac{a^\top x - m(F^a)}{s(F^a)}\right|$$

where $\mathcal{S}_+^{p-1} = \{a \in \mathcal{S}^{p-1} : a_1 \geq 0\}$. For a spherical distribution $F$ we have $m(F^a) = 0$ and $s(F^a) = s_0$ for all $a \in \mathcal{S}^{p-1}$, so that $r^2(x;F) = \|x\|^2/s_0^2$. Then $T(F) = 0$ and $V(F) = E\{w(R^2)R^2\}/(pE\{w(R^2)\})I$.

To obtain the influence functions of $T$ and $V$, note that

$$
\begin{aligned}
T(F_{\varepsilon,z}) &= \frac{(1-\varepsilon)h_1(\varepsilon)+\varepsilon w(r^2(z;F_{\varepsilon,z}))z}{h_0(\varepsilon)} \\
V(F_{\varepsilon,z}) &= \frac{(1-\varepsilon)h_2(\varepsilon)+\varepsilon w(r^2(z;F_{\varepsilon,z}))(z-T(F_{\varepsilon,z}))(z-T(F_{\varepsilon,z}))^\top}{h_0(\varepsilon)}
\end{aligned}
$$

where

$$
\begin{aligned}
h_0(\varepsilon) &= E_F\{w(r^2(X;F_{\varepsilon,z}))\} \\
h_1(\varepsilon) &= E_F\{w(r^2(X;F_{\varepsilon,z}))X\} \\
h_2(\varepsilon) &= E_F\{w(r^2(X;F_{\varepsilon,z}))(X-T(F_{\varepsilon,z}))(X-T(F_{\varepsilon,z}))^\top\}.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
IF(z;T,F) &= \frac{h_1'(0)+w(r^2(z;F))z}{h_0(0)} \\
IF(z;V,F) &= \frac{h_2'(0)+w(r^2(z;F))zz^\top-h_2(0)}{h_0(0)}.
\end{aligned}
$$

Clearly, the problem is to obtain $h_1'(0)$ and $h_2'(0)$. This is not so straightforward because it involves computing the derivative of $r(x;F_{\varepsilon,z})$ with respect to $\varepsilon$, which in turn requires some non-standard results about the derivatives of $m(F^a_{\varepsilon,z})$ and $s(F^a_{\varepsilon,z})$. These results are given in the next section.

# 3 Derivatives of location-scale estimators of projections and of the outlyingness measure

Since it is necessary to find the derivatives of $m(F^a_{\varepsilon,z})$ and $s(F^a_{\varepsilon,z})$ with respect to $a$ and $\varepsilon$, we will restrict ourselves to the class of univariate location-scale estimators that are solutions of simultaneous M-estimating equations. This class is rich enough to include all estimators used in practice, and allows us to obtain nice expressions for the derivatives under fairly general conditions.

Let us consider then, a pair of location-scale functionals $(m,s)$ that solve

$$E\Big\{\psi\Big(\frac{Z-m}{s}\Big)\Big\} = 0 \quad (6)$$

$$E\Big\{\rho\Big(\frac{Z-m}{s}\Big)\Big\} = K. \quad (7)$$

Assume that

**A1.** $\psi$ is odd and $\rho$ is even.

**A2.** $\psi$ and $\rho$ are twice continuously differentiable and $\psi(u)$, $\psi'(u)u$, $\psi''(u)u^2$, $\rho(u)$, $\rho'(u)u$, and $\rho''(u)u^2$ are bounded.

Two important cases, for which the resulting S-D estimators attain the maximum asymptotic breakdown point 50%, are:

- The median and the standardized median absolute deviation, corresponding to $\psi(u) = \text{sign}(u)$, $\rho(u) = I(|u| \geq \Phi^{-1}(3/4))$ and $K = 1/2$.
- The maximum likelihood estimator of location and scale of a univariate Cauchy distribution, corresponding to $\psi(u) = u/(1+u^2)$, $\rho(u) = u^2/(1+u^2)$ and $K = 1/2$.

Note that A2 is satisfied in the Cauchy case but not in the med/mad case. Nevertheless, the final formulas of Theorem 3 can be applied in this case as well, if the central distribution has a density.

If $X$ is a random vector with spherical distribution $F$, and $F_1$ denotes the marginal distribution of $X_1$, then A1 implies that $(m(F^a), s(F^a)) = (0, \|a\|s_0)$ for every $a \in \mathbb{R}^p$, where $s_0 = s_0(F)$ is implicitly defined by

$$E_{F_1}\Big\{\rho\Big(\frac{X_1}{s_0}\Big)\Big\} = K. \quad (8)$$

Typically $\rho$ is calibrated so that $s_0 = 1$ when $F = \mathcal{N}_p(0, I)$, but this needs not be the case. All we assume is the following:

**A3.** Equation (8) admits a solution $s_0 > 0$ for which $E_{F_1}\{\psi'(X_1/s_0)\} \neq 0$ and $E_{F_1}\{\rho'(X_1/s_0)X_1/s_0\} \neq 0$.

The next proposition gives sufficient conditions for A3 to hold. These conditions are satisfied by most estimators and distributions that appear in practice (for the med/mad case, an ad-hoc proof is possible). Theorem 1 below provides expressions for the partial derivatives of $m(F^a_{\varepsilon,z})$ and $s(F^a_{\varepsilon,z})$ that are needed to compute the derivative of $r(x;F_{\varepsilon,z})$ at $\varepsilon=0$, which is given in Theorem 2.

**Proposition 1** *Equation (8) admits a solution $s_0 \in (0,\infty)$ if $\rho$ is bounded, even, non-decreasing in $[0,\infty)$, $\rho(0)<K<\rho(\infty)$, and $X_1$ has a symmetric distribution about zero. In addition, A3 holds if $\psi$ is bounded, odd, non-negative in $[0,\infty)$, both $\psi(|u|)$ and $\rho'(|u|)$ are strictly positive in a neighborhood of zero (except at zero), and $P(X_1=0)<1$.*

**Theorem 1** *Let $F_{\varepsilon,z}=(1-\varepsilon)F+\varepsilon\Delta_z$ with $F$ an spherical distribution. Suppose that A1, A2 and A3 hold. For each $z\in\mathbb{R}^p$, let $G(\varepsilon,a)=(m(F^a_{\varepsilon,z}),s(F^a_{\varepsilon,z}))$. Then $G$ is twice continuously differentiable at $(0,a)$ for every $a\neq 0$, and if $\tilde a = a/\|a\|$, the partial derivatives are:*

$$\begin{aligned}
\frac{\partial G_1}{\partial \varepsilon}(0,a) &= \frac{\|a\|s_0\psi(\tilde a^\top z/s_0)}{E_F\{\psi'(X_1/s_0)\}}\\
\frac{\partial G_1}{\partial a}(0,a) &= 0\\
\frac{\partial G_2}{\partial \varepsilon}(0,a) &= \frac{\|a\|s_0\{\rho(\tilde a^\top z/s_0)-K\}}{E_F\{\rho'(X_1/s_0)X_1/s_0\}}\\
\frac{\partial G_2}{\partial a}(0,a) &= \tilde a s_0\\
\frac{\partial^2 G_1}{\partial a\partial \varepsilon}(0,a) &= \frac{\tilde a\psi(\tilde a^\top z/s_0)s_0+\psi'(\tilde a^\top z/s_0)(I-\tilde a\tilde a^\top)z}{E_F\{\psi'(X_1/s_0)\}}\\
\frac{\partial^2 G_1}{\partial a\partial a^\top}(0,a) &= 0\\
\frac{\partial^2 G_2}{\partial a\partial \varepsilon}(0,a) &= \frac{\tilde a\{\rho(\tilde a^\top z/s_0)-K\}s_0+\rho'(\tilde a^\top z/s_0)(I-\tilde a\tilde a^\top)z}{E_F\{\rho'(X_1/s_0)X_1/s_0\}}\\
\frac{\partial^2 G_2}{\partial a\partial a^\top}(0,a) &= \frac{1}{\|a\|}(I-\tilde a\tilde a^\top)s_0.
\end{aligned}$$

**Theorem 2** *Under the conditions of Theorem 1, $r^2(x;F_{\varepsilon,z})$ is differentiable at $\varepsilon=0$ for each $x,z\in\mathbb{R}^p$, $x\neq 0$, and*

$$\frac{\partial}{\partial\varepsilon}r^2(x;F_{\varepsilon,z})\bigg|_{\varepsilon=0} = -2\,\frac{\|x\|}{s_0}\frac{\psi(\tilde x^\top z/s_0)}{E_F\{\psi'(X_1/s_0)\}} - 2\,\frac{\|x\|^2}{s_0^2}\frac{\{\rho(\tilde x^\top z/s_0)-K\}}{E_F\{\rho'(X_1/s_0)X_1/s_0\}}, \tag{9}$$

*where $\tilde x = x/\|x\|$.*

## 4 Influence function of S-D estimators

Now we can resume the derivation of the influence function of S-D estimators. Theorem 3 below gives the expressions of $IF(z;T,F)$ and $IF(z;V,F)$. For this theorem we need an additional assumption:

**W2.** $w$ is differentiable almost everywhere and $\sup |w'(u^2)u^4| < \infty$.

**Theorem 3** *Let $F$ be a spherical distribution such that $P_F(X=0)=0$ and assume that A1, A2, A3, W1 and W2 hold. Then*

$$IF(z;T,F) = \Big\{ \frac{c_1(F)}{c_0(F)} g_1(\|z\|) + \frac{w(\|z\|^2/s_0^2)\|z\|}{c_0(F)} \Big\} \frac{z}{\|z\|}$$

*and*

$$\begin{aligned} IF(z;V,F) &= \Big\{ \frac{c_2(F)}{c_0(F)} g_2(\|z\|) + \frac{w(\|z\|^2/s_0^2)\|z\|^2}{c_0(F)} \Big\} \Big\{ \frac{zz^\top}{\|z\|^2} - \frac{I}{p} \Big\} \\ &\quad + \Big\{ \frac{c_2(F)}{c_0(F)} g_3(\|z\|) + \frac{w(\|z\|^2/s_0^2)\|z\|^2 - c_3(F)}{c_0(F)} \Big\} \frac{I}{p} \end{aligned}$$

*if $z \neq 0$, and $IF(0;T,F)=0$, $IF(0;V,F)=0$. Here*

$$\begin{aligned} g_1(t) &= \frac{E\{\psi(U_1 t/s_0)U_1\}}{E\{\psi'(X_1/s_0)\}} \\ g_2(t) &= \frac{E\{\rho(U_1 t/s_0)U_1^2\}p - E\{\rho(U_1 t/s_0)\}}{E\{\rho'(X_1/s_0)X_1/s_0\}(p-1)} \\ g_3(t) &= \frac{E\{\rho(U_1 t/s_0)\} - K}{E\{\rho'(X_1/s_0)X_1/s_0\}} \\ c_0(F) &= E\left\{ w(R^2/s_0^2) \right\} \\ c_1(F) &= -2E\left\{ w'(R^2/s_0^2)R^2/s_0 \right\} \\ c_2(F) &= -2E\left\{ w'(R^2/s_0^2)R^4/s_0^2 \right\} \\ c_3(F) &= E\left\{ w(R^2/s_0^2)R^2 \right\} \end{aligned}$$

*where $X \sim F$, $R = \|X\|$ and $U = X/\|X\|$.*

Note that $IF(z;T,F)$ does not depend on the scale estimator $s$ (except through $s_0$) and $IF(z;V,F)$ does not depend on the location estimator $m$. More remarkable, although not completely surprising, is that $IF(z;T,F)$ and $IF(z;V,F)$ have the forms of influence functions of one-step reweighted estimators (Lopuhaä, 1999, Remark 4.1; in Lopuhaä's notation, $c_0(F) = c_1$, $c_1(F) = pc_2$, $c_2(F) = p(p+2)c_4$ and $c_3(F) = pc_3$). Specifically, suppose that $T_0$ and $V_0$ are estimators with influence functions of the form (2) and (3) with $w_\mu(\|z\|^2) = pg_1(\|z\|)/\|z\|$, $w_\eta(\|z\|^2) = p(p+2)g_2(\|z\|)/\|z\|^2$ and $w_\tau(\|z\|^2) = 2pg_3(\|z\|)$. Then the influence functions of $T$ and $V$ coincide with those of the one-step reweighted estimators

$$
\begin{aligned}
T_1(F) &= E_F\{w(d^2(X;T_0,V_0))X\}/E_F\{w(d^2(X;T_0,V_0))\} \\
V_1(F) &= E_F\{w(d^2(X;T_0,V_0))(X-T_1)(X-T_1)^\top\}/E_F\{w(d^2(X;T_0,V_0))\},
\end{aligned}
$$

where $d^2(X;T_0,V_0) = (X-T_0)^\top V_0^{-1}(X-T_0)$. We say that this result is not completely surprising because if one takes $m(F^a) = a^\top T_0(F)$ and $s(F^a) = \{a^\top V_0(F)a\}^{\frac{1}{2}}$ then $d(X;T_0,V_0) = r(X;F)$. Of course, $T_0$ and $V_0$ are only "virtual" estimators; while it is possible to construct M-estimating equations from the influence functions of $T_0$ and $V_0$, they do not satisfy the conditions for existence of solution required in Hampel et al. (1986, p.287, Theorem 1). Nevertheless, this heuristic argument sheds some light on certain facts that have been observed in simulations; for example, that a given S-D estimator can attain high relative efficiency under both Normal and Cauchy models (Yohai and Maronna, 1995, Section 4.4), a behavior that is not uncommon for reweighted estimators (see Lupohaä, 1999, and Gervini, 2002). It also allows us to conjecture that S-D estimators are asymptotically normal, which was doubted in the past. The asymptotic normality of $T$, with asymptotic variance given by $E_F\{IF(X;T,F)IF(X;T,F)^\top\}$, has been proved by Zuo et al. (2001, Theorem 3.1). The conditions assumed by these authors are more general than ours in principle, but when it comes to obtaining working formulas they also assume that the model distribution is elliptical and that $m$ is an M-estimator defined through a differentiable $\psi$ function, although they are allowed to use the MAD as scale estimator (see their Lemma 3.2). However, for the asymptotic normality of $V$ (which has not yet been rigorously established) it is presumably necessary to impose on $s$ smoothness conditions similar to ours.

The gross-error sensitivities and the asymptotic efficiencies that result from Theorem 3 provide a rational way to choose the weight function $w$. We need a weight function that offers a reasonable trade-off between local robustness and efficiency. Maronna and Yohai (1995) recommend a Huber-type weight function $w_H(t^2;c) = \min\{t^2, c\}/t^2$. Zuo et al. (2001) propose a weight function that gives weight one to observations with small outlyingness and then decreases exponentially to zero; following this idea, we consider a truncated exponential weight function $w_E(t^2;c) = \min\{1, \exp(5(1-t^2/c))\}$. This function decreases so rapidly to zero that it is essentially a smoothed down indicator function. Finally, we introduce a Gaussian weight function $w_G(t^2;c) = \varphi(t^2/c)/\varphi(1)$, which decreases exponentially for $t^2 > c$ but, unlike $w_E$ and $w_H$ which are constant on $[0, c)$, it is strictly decreasing on $[0,\infty)$. We take cut-off values of the form $c = \chi^2_{p,1-\alpha}$. Figures 1 and 2 plot gross-error sensitivities and asymptotic relative efficiencies as functions of $\alpha$, for $\mathcal{N}_p(0, I)$ models with $p = 2$ and $p = 10$, and the median and MAD as $m$ and $s$. We can see very clearly that the exponential weight is uniformly worse than the others. The reason is that $w_E$ decreases to zero too quickly after the cut-off value, and then $|w'_E|$ becomes too large, resulting in large coefficients $c_1(F)/c_0(F)$ and $c_2(F)/c_0(F)$. Figure 3 illustrates this, showing $c_1^2(F)/(pc_0^2(F))$ and $c_2^2(F)/(p(p+2)c_0^2(F))$ for a $\mathcal{N}_5(0, I)$ and a $\mathcal{T}_{5,3}(0, I)$ model. Note how these ratios are so close to zero for $\alpha = .05$ (the value suggested by Maronna and Yohai, 1995), that the contribution of $m$ and $s$ is virtually negligible and the asymptotic behavior is then dominated by the weight function.

Regarding the other two weights, neither of them stands out as uniformly better. Huber weights are the most efficient but the value $\alpha = .05$ yields poor local robustness. Overall, the best choices seem to be either a Gaussian weight with $\alpha = .10$ or a Huber weight with $\alpha = .50$. For further insight, we have compared these two estimators with the multivariate S-estimator corresponding to Tukey's $\rho$-function, with asymptotic breakdown point 1/2 and calibrated for consistency under the Normal model. Figures 4 and 5 plot gross error sensitivities and the asymptotic relative efficiencies as functions of $p$, for $\mathcal{N}_p(0, I)$ and $\mathcal{T}_{p,3}(0, I)$ models. Again, there are no clear winners between Huber and Gaussian weights, but we can confidently say that the S-D estimators perform fairly well compared to the S-estimator, and are almost always more efficient.

# A Proofs

**Proof of Proposition 1.** The first part is immediate, since $E(\rho(X_1/s))$ goes to $\rho(0)$ when $s \to \infty$, and goes to $\rho(\infty)$ when $s \to 0$. For the second part, there is by assumption a $\delta_1 > 0$ such that $1 - F_1(\delta_1) > 0$. Given $s > 0$, there exists a $\delta > 0$ small enough so that $\delta \le \delta_1/s$ and $\psi(\delta) > 0$. Then

$$\begin{aligned} E\Big\{\psi'\Big(\frac{X_1}{s}\Big)\Big\} &= 2s\int_0^\infty (1 - F_1(sx))\, d\psi(x) \\ &\ge 2s(1 - F_1(\delta_1))\psi(\delta), \end{aligned}$$

which is strictly positive. To show that $E\{\rho'(X_1/s)X_1/s\} > 0$, use the same argument replacing $\psi(u)$ by $\int_0^u \rho'(x)x\, dx$. □

**Proof of Theorem 1.** For a fixed $z \in \mathbb{R}^p$, let $H : [0,1] \times \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}^2$ be given by

$$\begin{aligned} H_1(\varepsilon, a, m, s) &= E_{F_{\varepsilon,z}}\Big\{\psi\Big(\frac{a^\top X - m}{s}\Big)\Big\} \\ H_2(\varepsilon, a, m, s) &= E_{F_{\varepsilon,z}}\Big\{\rho\Big(\frac{a^\top X - m}{s}\Big)\Big\} - K. \end{aligned}$$

Then $G(\varepsilon, a)$ is implicitly defined by the equation $H(\varepsilon, a, G(\varepsilon, a)) = 0$. To find the derivatives we will use the Implicit Function Theorem (Dieudonné, 1969, Theorems 10.2.2 and 10.2.3). By A1, A2 and a straightforward application of dominated convergence, we have

$$\begin{aligned} \frac{\partial H_1}{\partial \varepsilon} &= -E_F\Big\{\psi\Big(\frac{\|a\|X_1 - m}{s}\Big)\Big\} + \psi\Big(\frac{a^\top z - m}{s}\Big) \\ \frac{\partial H_1}{\partial a} &= (1-\varepsilon)E_F\Big\{\psi'\Big(\frac{\|a\|X_1 - m}{s}\Big)\frac{X_1}{s}\Big\}\frac{a}{\|a\|} + \varepsilon\psi'\Big(\frac{a^\top z - m}{s}\Big)\frac{z}{s} \\ \frac{\partial H_1}{\partial m} &= (1-\varepsilon)E_F\Big\{\psi'\Big(\frac{\|a\|X_1 - m}{s}\Big)\Big\}\Big(-\frac{1}{s}\Big) + \varepsilon\psi'\Big(\frac{a^\top z - m}{s}\Big)\Big(-\frac{1}{s}\Big) \\ \frac{\partial H_1}{\partial s} &= (1-\varepsilon)E_F\Big\{\psi'\Big(\frac{\|a\|X_1 - m}{s}\Big)\Big(\frac{\|a\|X_1 - m}{s}\Big)\Big\}\Big(-\frac{1}{s}\Big) \\ &\quad +\varepsilon\psi'\Big(\frac{a^\top z - m}{s}\Big)\Big(\frac{a^\top z - m}{s}\Big)\Big(-\frac{1}{s}\Big), \end{aligned}$$

with analogous expressions for the partial derivatives of $H_2$ replacing $\psi$ by $\rho - K$. All these partial derivatives possess themselves continuous partial derivatives, whence $H$ is twice continuously differentiable. Since $G(0,a) = (0, \|a\| s_0)$, we need to evaluate the partial derivatives of $H$ at $(0, a, 0, \|a\| s_0)$. The differential is

$$\mathcal{D}_{m,s} H(0,a,0,\|a\|s_0) = \begin{pmatrix} -\dfrac{E_{F_1}\{\psi'(X_1/s_0)\}}{\|a\|s_0} & 0 \\ 0 & -\dfrac{E_{F_1}\{\rho'(X_1/s_0)X_1/s_0\}}{\|a\|s_0} \end{pmatrix}$$

which is non-singular by A3. Hence G is (twice) continuously differentiable at $(0, a)$ for any $a \neq 0$, and $\mathcal{D}_{\varepsilon,a} G(0,a) = (\mathcal{D}_{m,s} H(0,a,0,\|a\|s_0))^{-1} \mathcal{D}_{\varepsilon,a} H(0,a,0,\|a\|s_0)$. Since

$$\mathcal{D}_{\varepsilon,a} H(0,a,0,\|a\|s_0) = \begin{pmatrix} \psi\Big(\dfrac{a^\top z}{\|a\|s_0}\Big) & 0 \\ \rho\Big(\dfrac{a^\top z}{\|a\|s_0}\Big) - K & E_{F_1}\Big\{\rho'\Big(\dfrac{X_1}{s_0}\Big)\dfrac{X_1}{\|a\|s_0}\Big\}\dfrac{a}{\|a\|} \end{pmatrix}$$

the expressions for the first partial derivatives given in the theorem follow. Second partial derivatives are obtained from them in a straightforward manner. □

**Proof of Theorem 2.** For a fixed $x \in \mathbb{R}^p$, $x \neq 0$, consider the Lagrangian

$$\mathcal{G}(\varepsilon, a, \lambda) = \Big(\frac{a^\top x - G_1(\varepsilon,a)}{G_2(\varepsilon,a)}\Big)^2 + \lambda(1 - \|a\|^2).$$

By Theorem 1 we have that $\mathcal{G}$ is differentiable at $(0, a, \lambda)$ for any $a \neq 0$. Let $H : [0,1] \times \mathbb{R}^p \times \mathbb{R}_+ \to \mathbb{R}^{p+1}$ be the vector of partial derivatives of $g$ with respect to $a$ and $\lambda$. That is,

$$\begin{aligned} H_1(\varepsilon,a,\lambda) &= 2\Big(\frac{a^\top x - G_1(\varepsilon,a)}{G_2(\varepsilon,a)}\Big)\Big(\frac{x - \frac{\partial G_1}{\partial a}(\varepsilon,a)}{G_2(\varepsilon,a)} - \frac{a^\top x - G_1(\varepsilon,a)}{G_2^2(\varepsilon,a)}\frac{\partial G_2}{\partial a}(\varepsilon,a)\Big) - 2\lambda a \\ H_2(\varepsilon,a,\lambda) &= 1 - \|a\|^2. \end{aligned}$$

Therefore the vector $a = a(\varepsilon)$ that realizes the supremum $r^2(x; F_{\varepsilon,z})$ satisfies the implicit equation $H(\varepsilon, a(\varepsilon), \lambda(\varepsilon)) = 0$. We will use the Implicit Function Theorem to show that $a(\varepsilon)$ is differentiable at $\varepsilon = 0$. The condition $H_2(0, a(0), \lambda(0)) = 0$ is equivalent to $\|a(0)\| = 1$. Since

$$H_1(0,a,\lambda) = 2\frac{a^\top x x}{\|a\|^2 s_0^2} - 2\frac{(a^\top x)^2 a}{\|a\|^4 s_0^2} - 2\lambda a,$$

by multiplying each term by $a$ we deduce that $H_1(0, a(0), \lambda(0)) = 0$ if and only if $\lambda(0) = 0$ and either $a(0)^\top x = 0$ or $a(0) = \pm x/\|x\|$. Clearly the maximizer of $r^2(x; F)$ cannot be orthogonal to $x$ because that would imply $r^2(x, F) = 0$; then it must be $a(0) = \pm x/\|x\|$, whichever belongs to $\mathcal{S}_+^{p-1}$. Let $a_0 = a(0)$. Using the formulas in Theorem 1 we have that

$$\begin{aligned}
\frac{\partial H_1}{\partial a}(0, a, \lambda) &= \frac{2}{\|a\|^2 s_0^2}\Big\{\Big(I - 2\frac{aa^\top}{\|a\|^2}\Big)xx^\top\Big(I - 2\frac{aa^\top}{\|a\|^2}\Big) - \frac{(a^\top x)^2}{\|a\|^2}I\Big\} - 2\lambda I \\
\frac{\partial H_1}{\partial \lambda}(0, a, \lambda) &= -2a \\
\frac{\partial H_2}{\partial a}(0, a, \lambda) &= -2a \\
\frac{\partial H_2}{\partial \lambda}(0, a, \lambda) &= 0
\end{aligned}$$

and then

$$\mathcal{D}_{a,\lambda} H(0, a_0, 0) = \begin{pmatrix} \frac{2}{s_0^2}\Big(xx^\top - \|x\|^2 I\Big) & -2a_0 \\ -2a_0^\top & 0 \end{pmatrix}$$

which is non-singular. Then, by the Implicit Function Theorem we have that $a(\varepsilon)$ and $\lambda(\varepsilon)$ are differentiable at $\varepsilon = 0$. Therefore

$$r^2(x, F_{\varepsilon,z}) = \Big(\frac{a(\varepsilon)^\top x - G_1(\varepsilon, a(\varepsilon))}{G_2(\varepsilon, a(\varepsilon))}\Big)^2$$

is also differentiable at $\varepsilon = 0$ and

$$\frac{\partial}{\partial \varepsilon} r^2(x; F_{\varepsilon,z})\Big|_{\varepsilon=0}$$

$$= 2\Big(\frac{a_0^\top x - G_1(0, a_0)}{G_2(0, a_0)}\Big)\Big(\frac{a'(0)^\top x - \frac{\partial G_1(\varepsilon, a(\varepsilon))}{\partial \varepsilon}\Big|_{\varepsilon=0}}{G_2(0, a_0)} - \frac{a_0^\top x - G_1(0, a_0)}{G_2^2(0, a_0)}\frac{\partial G_2(\varepsilon, a(\varepsilon))}{\partial \varepsilon}\Big|_{\varepsilon=0}\Big).$$

The explicit value of $a'(0)$ is irrelevant; we only use that $a'(0)^\top a_0 = 0$, a consequence of $\|a(\varepsilon)\|^2 = 1$. Since $G_1(0, a_0) = 0$ and $G_2(0, a_0) = s_0$, we have

$$\begin{aligned}
\frac{\partial G_1(\varepsilon, a(\varepsilon))}{\partial \varepsilon}\Big|_{\varepsilon=0} &= \frac{\partial G_1}{\partial \varepsilon}(0, a_0) + \frac{\partial G_1}{\partial a}(0, a_0)^\top a'(0) \\
&= \frac{s_0 \psi(a_0^\top z/s_0)}{E_F\{\psi'(X_1/s_0)\}} \\
\frac{\partial G_2(\varepsilon, a(\varepsilon))}{\partial \varepsilon}\Big|_{\varepsilon=0} &= \frac{\partial G_2}{\partial \varepsilon}(0, a_0) + \frac{\partial G_2}{\partial a}(0, a_0)^\top a'(0) \\
&= \frac{s_0\{\rho(a_0^\top z/s_0) - K\}}{E_F\{\rho'(X_1/s_0)X_1/s_0\}}
\end{aligned}$$

and (9) follows. □

**Proof of Theorem 3.** By dominated convergence we have

$$\begin{aligned} h_1'(0) &= E_F\Big\{w'(r^2(X,F))\frac{\partial}{\partial\varepsilon}r^2(x,F)\Big|_{\varepsilon=0}X\Big\} \\ h_2'(0) &= E_F\Big\{w'(r^2(X,F))\frac{\partial}{\partial\varepsilon}r^2(x,F)\Big|_{\varepsilon=0}XX^\top\Big\}. \end{aligned}$$

The theorem follows upon replacing the derivative of $r^2(X, F_{\varepsilon,z})$ by the expression given in Theorem 2 and using the factorization $X = RU$. We have to evaluate expectations of the form $E\{\vartheta(z^\top U/s_0)U\}$ and $E\{\vartheta(z^\top U/s_0)UU^\top\}$, where $\vartheta$ is either $\psi$ or $\rho - K$. Take a $\Gamma \in \mathbb{O}(p)$ with $z/\|z\|$ as its first column. Using that $\mathcal{L}(U) = \mathcal{L}(\Gamma U)$ we obtain that

$$E\Big\{\vartheta\Big(\frac{z^\top U}{s_0}\Big)U\Big\} = E\Big\{\vartheta\Big(\frac{\|z\|U_1}{s_0}\Big)U_1\Big\}\frac{z}{\|z\|} \tag{10}$$

$$E\Big\{\vartheta\Big(\frac{z^\top U}{s_0}\Big)UU^\top\Big\} = E\Big\{\vartheta\Big(\frac{\|z\|U_1}{s_0}\Big)U_2^2\Big\}I + E\Big\{\vartheta\Big(\frac{\|z\|U_1}{s_0}\Big)(U_1^2 - U_2^2)\Big\}\frac{zz^\top}{\|z\|^2} \tag{11}$$

and since

$$\begin{aligned} E\Big\{\vartheta\Big(\frac{\|z\|U_1}{s_0}\Big)\Big\} &= \operatorname{tr} E\Big\{\vartheta\Big(\frac{z^\top U}{s_0}\Big)UU^\top\Big\} \\ &= E\Big\{\vartheta\Big(\frac{\|z\|U_1}{s_0}\Big)U_1^2\Big\} + (p-1)E\Big\{\vartheta\Big(\frac{\|z\|U_1}{s_0}\Big)U_2^2\Big\} \end{aligned}$$

equation (11) can be rewritten entirely in terms of $U_1$ as

$$\begin{aligned} E\Big\{\vartheta\Big(\frac{z^\top U}{s_0}\Big)UU^\top\Big\} &= E\Big\{\vartheta\Big(\frac{\|z\|U_1}{s_0}\Big)U_1^2\Big\}\frac{zz^\top}{\|z\|^2} \\ &\quad + \frac{1}{p-1}E\Big\{\vartheta\Big(\frac{\|z\|U_1}{s_0}\Big)(1-U_1^2)\Big\}\Big(I - \frac{zz^\top}{\|z\|^2}\Big). \end{aligned}$$

The rest is straightforward, using that (10) is zero for $\vartheta = \rho - K$ and (11) is zero for $\vartheta = \psi$. □

**Some formulas for the med/mad case.** To compute the gross-error sensitivities of the S-D estimator, observe that

$$\begin{aligned} tw_\mu(t^2) &= \frac{c_1(F)g_1(t) + w(t^2/s_0^2)t}{c_0(F)} \\ t^2 w_\eta(t^2) &= \frac{c_2(F)g_2(t) + w(t^2/s_0^2)t^2}{c_0(F)} \\ w_\tau(t^2) &= \frac{c_2(F)g_3(t) + \{w(t^2/s_0^2)t^2 - c_3(F)\}}{c_0(F)} \\ \beta_F &= \frac{c_3(F)}{c_0(F)p}. \end{aligned}$$

When $m$ is the median and $s$ is the standardized MAD, we have

$$\begin{aligned} E\Big\{\psi\Big(\frac{U_1 t}{s_0}\Big)U_1\Big\} &= \frac{\Gamma(\frac{p}{2})}{\sqrt{\pi}\Gamma(\frac{p+1}{2})} \\ E\Big\{\rho\Big(\frac{U_1 t}{s_0}\Big)\Big\} &= 1 - F_{\mathcal{B}(\frac{1}{2},\frac{p-1}{2})}\Big(\frac{\Phi^{-1}(.75)^2 s_0^2}{t^2}\Big) \\ E\Big\{\rho\Big(\frac{U_1 t}{s_0}\Big)U_1^2\Big\} &= \frac{1}{p} - \frac{1}{p}F_{\mathcal{B}(\frac{3}{2},\frac{p-1}{2})}\Big(\frac{\Phi^{-1}(.75)^2 s_0^2}{t^2}\Big) \end{aligned}$$

and then, changing the order of integration to avoid $\psi'$ and $\rho'$, we have

$$\begin{aligned} g_1(t) &= \frac{\Gamma(\frac{p}{2})}{\sqrt{\pi}\Gamma(\frac{p+1}{2})}\frac{1}{2s_0 f_1(0)} \\ g_2(t) &= \frac{F_{\mathcal{B}(1/2,(p-1)/2)}(\Phi^{-1}(.75)^2 s_0^2/t^2) - F_{\mathcal{B}(3/2,(p-1)/2)}(\Phi^{-1}(.75)^2 s_0^2/t^2)}{(p-1)2s_0\Phi^{-1}(.75)f_1(s_0\Phi^{-1}(.75))} \\ g_3(t) &= \frac{1/2 - F_{\mathcal{B}(1/2,(p-1)/2)}(\Phi^{-1}(.75)^2 s_0^2/t^2)}{2s_0\Phi^{-1}(.75)f_1(s_0\Phi^{-1}(.75))} \end{aligned}$$

where $f_1$ is the marginal density function of $X_1$. For the $\mathcal{N}_p(0,I)$ model we have $s_0 = 1$, while for the $\mathcal{T}_{p,\nu}(0,I)$ model we have $s_0 = F^{-1}_{\mathcal{T}_{1,\nu}}(.75)/\Phi^{-1}(.75)$.

For $\gamma^*_T(F)$ we can obtain a explicit expression:

$$\gamma^*_T(F) = \Big\{\frac{c_1(F)}{c_0(F)}\frac{\Gamma(\frac{p}{2})}{\sqrt{\pi}\Gamma(\frac{p+1}{2})}\frac{1}{2s_0 f_1(0)} + \frac{\sup_{t\geq 0} w(t^2/s_0^2)t}{c_0(F)}\Big\} d_\mu^{*\frac{1}{2}}.$$

The maximum of $w(t^2/s_0)t$ is attained at $t = s_0\sqrt{c}$ for Huber and Exponential weights, and then $\sup_{t\geq 0} w(t^2/s_0^2)t = s_0\sqrt{c}$; for the Gaussian weight the maximum is attained at $t = s_0\sqrt{c}/2^{1/4}$ and then $\sup_{t\geq 0} w_G(t^2/s_0^2)t = \varphi(1/\sqrt{2})s_0\sqrt{c}/(2^{1/4}\varphi(1))$.

# References


Bilodeau, M. and Brenner, D. (1999), *Theory of Multivariate Statistics* (Springer, NY).

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

Dieudonné, J. (1969), *Foundations of Modern Analysis* (Academic Press, NY).

Donoho, D.L. (1982), Breakdown properties of multivariate location estimators, Ph.D. qualifying paper.(Department of Statistics, Harvard University).

Gervini, D. (2002), A robust and efficient adaptive reweighted estimator of multivariate location and scatter. To appear in *J. Multivar. Anal.*

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. (1986), *Robust Statistics: The Approach Based on Influence Functions* (Wiley, NY).

Kent, J.T. and Tyler, D.E. (1997), Constrained M-estimation for multivariate location and scatter, *Ann. Statist.* **24**, 1346–1370.

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

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

Maronna, R.A. and Yohai, V.J. (1998), Robust estimation of multivariate location and scatter, in: S. Kotz and N.L. Johnson, eds. *Encyclopedia of Statistical Sciences, Update Volume 2* (Wiley, NY) pp. 589–596.

Rousseeuw, P.J. and Yohai, V.J. (1984), Robust regression by means of S-estimators, in: J. Franke, W. Härdle and R.D. Martin, eds. *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statistics Vol. 26* (Springer, NY) pp. 256–272.

Stahel, W.A. (1981), Robust estimation: Infinitesimal optimality and covariance matrix estimators, Ph.D. thesis (ETH, Zurich).

Zuo, Y., Cui, H., and He, X. (2002), On the Stahel-Donoho estimator and depth-weighted means of multivariate data, unpublished manuscript.

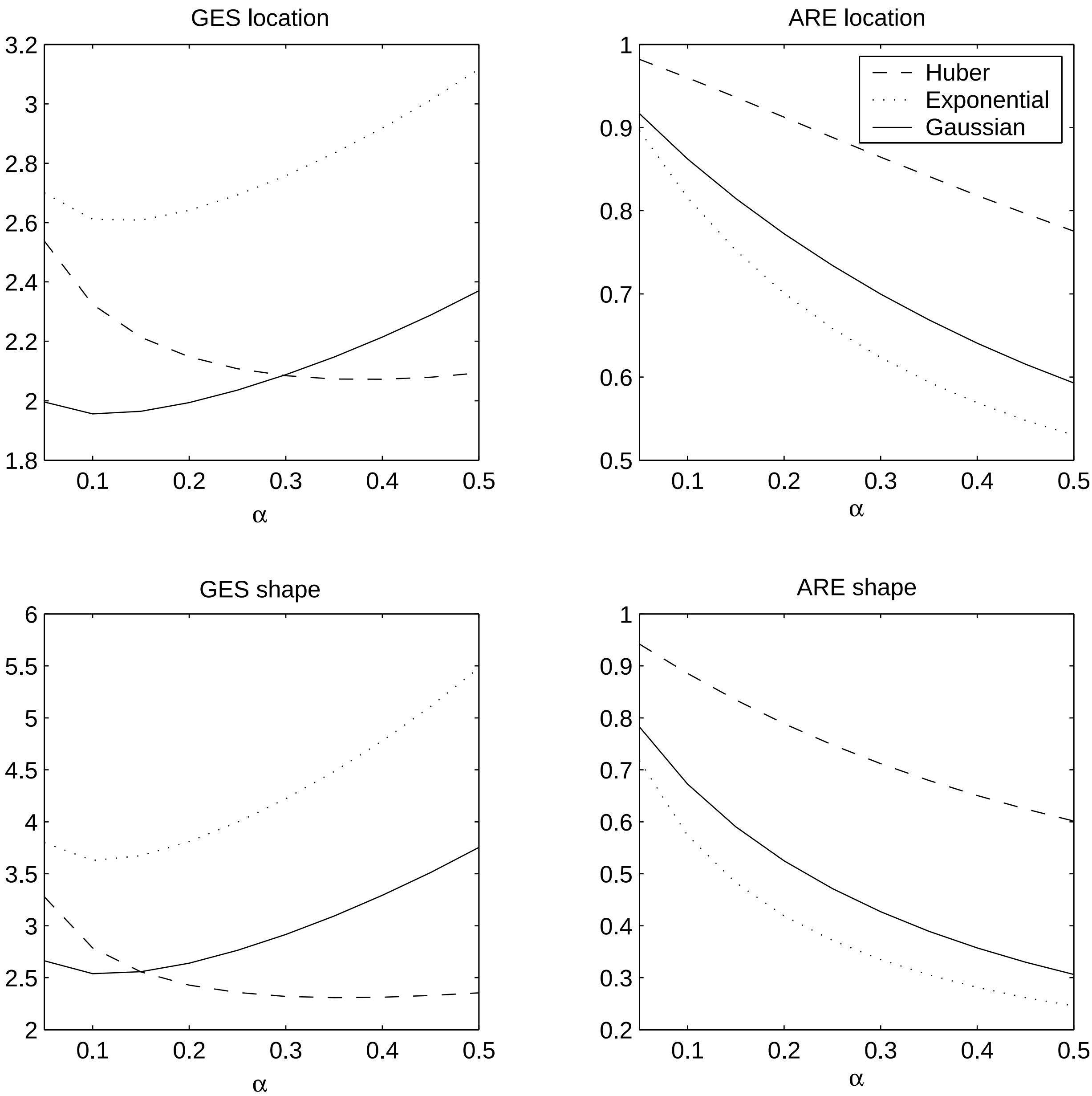


**Figure 1:** Gross-error sensitivities and asymptotic relative efficiencies of S-D estimators for three different weight functions and cut-off values $\chi^2_{p,1-\alpha}$. Model $\mathcal{N}_p(0, I)$ with $p = 2$.

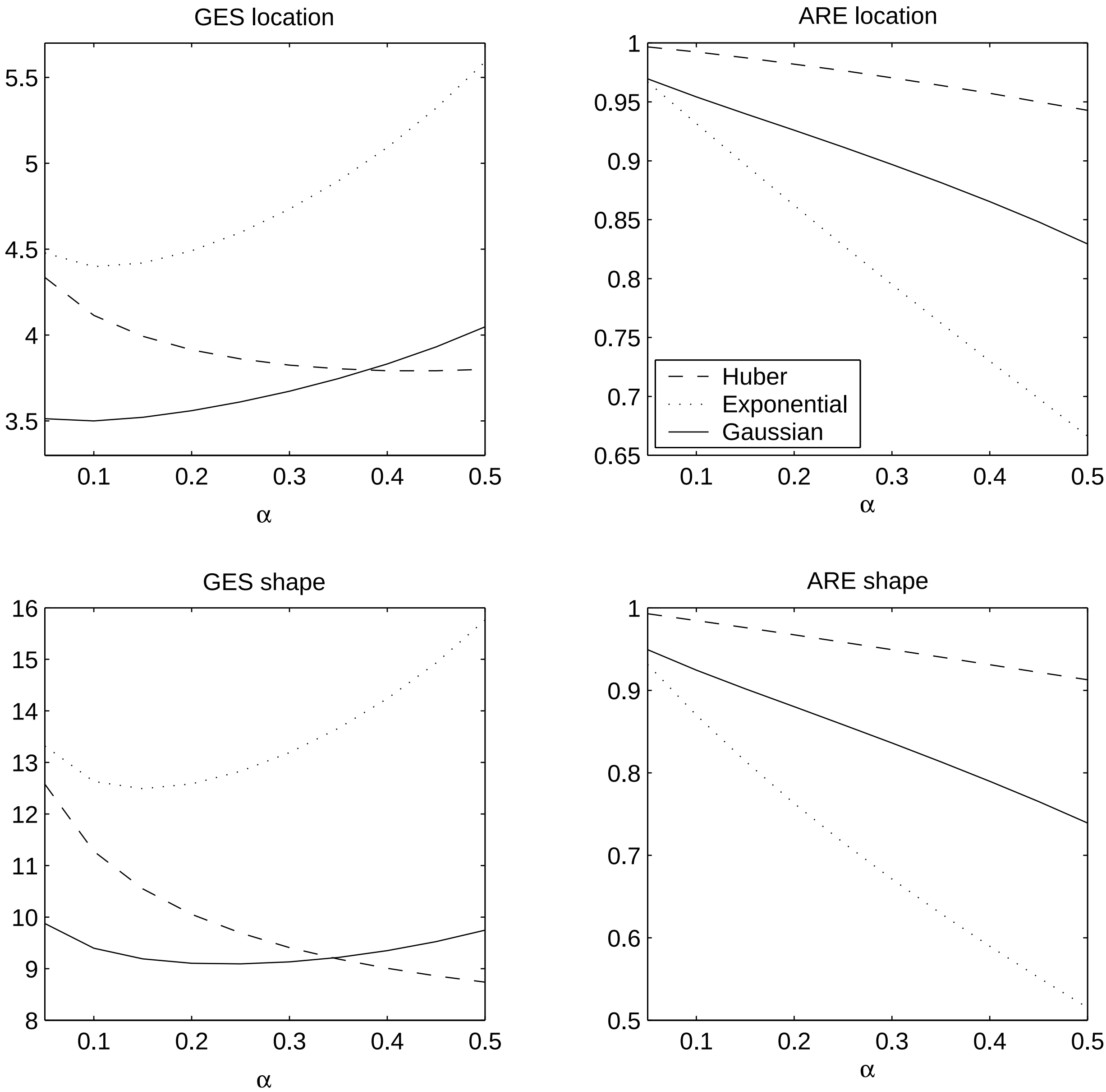


**Figure 2:** Gross-error sensitivities and asymptotic relative efficiencies of S-D estimators for three different weight functions and cut-off values $\chi^2_{p,1-\alpha}$. Model $\mathcal{N}_p(0, I)$ with $p = 10$.

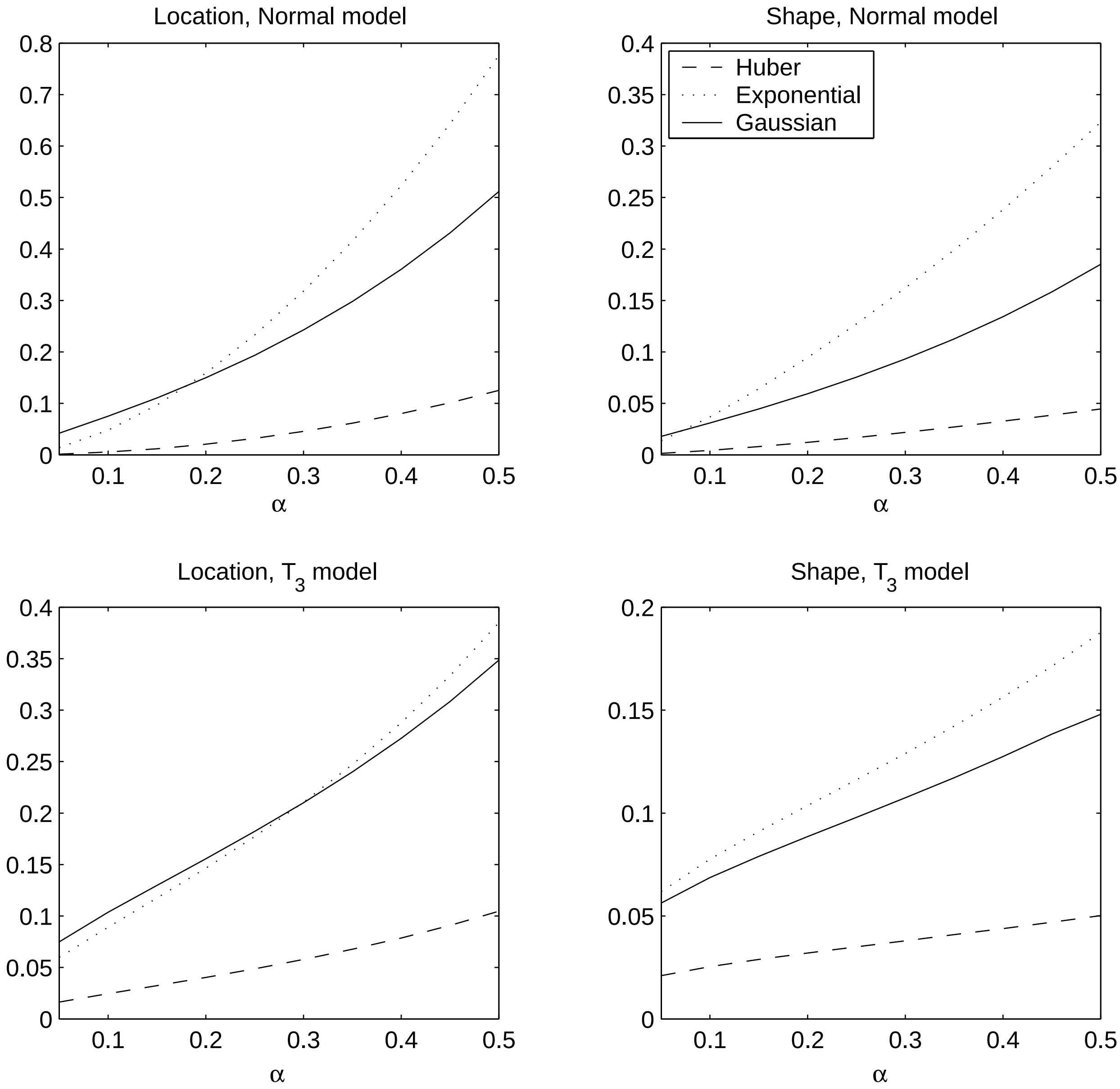


**Figure 3:** Coefficients $c_1^2/(pc_0^2)$ (for location) and $c_2^2/(p(p+2)c_3^2)$ (for scatter) under $\mathcal{N}_5(0, I)$ and $\mathcal{T}_{5,3}(0, I)$ models.

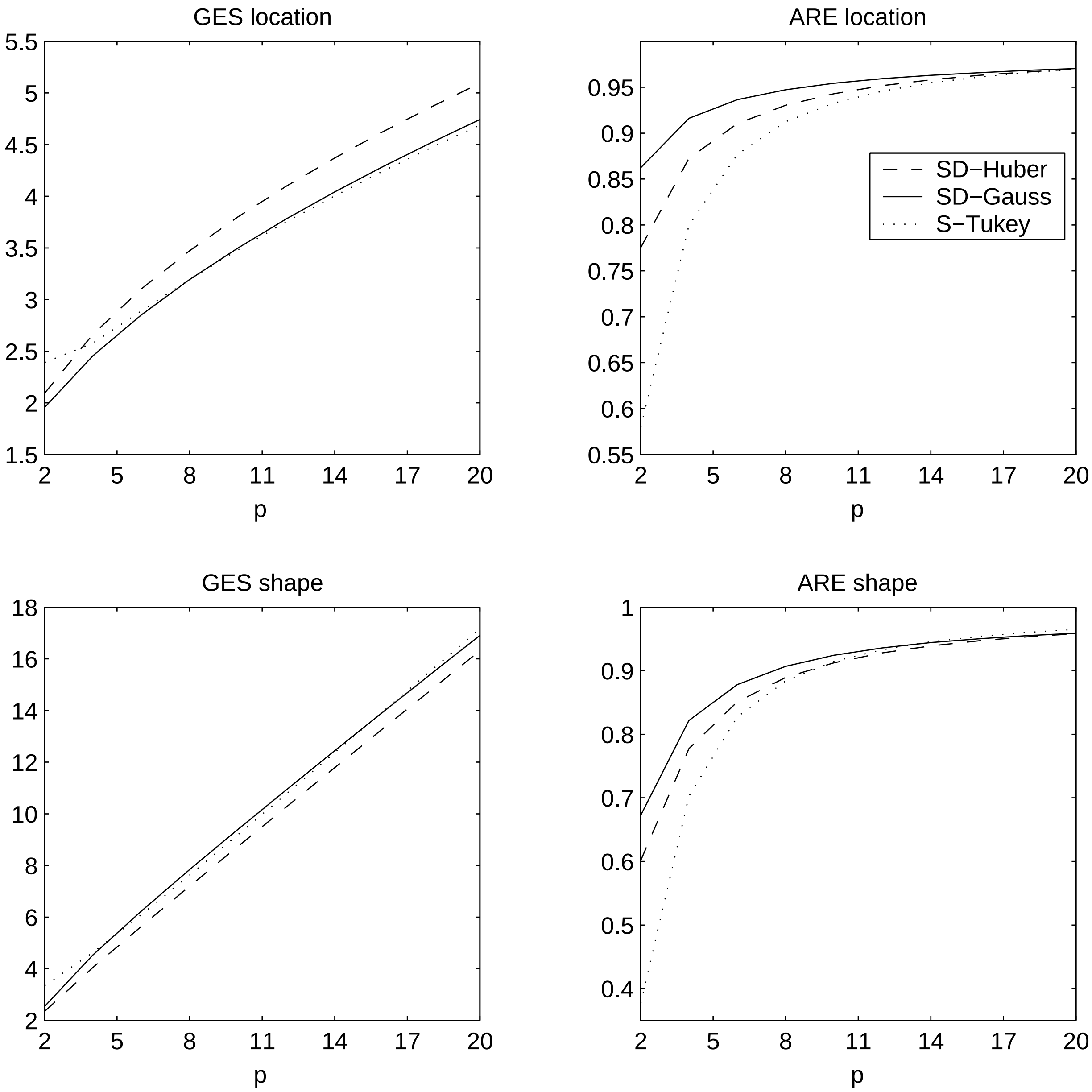


**Figure 4:** Gross-error sensitivities and asymptotic relative efficiencies of S-D estimators compared with S-estimators, for $\mathcal{N}_p(0, I)$ models.

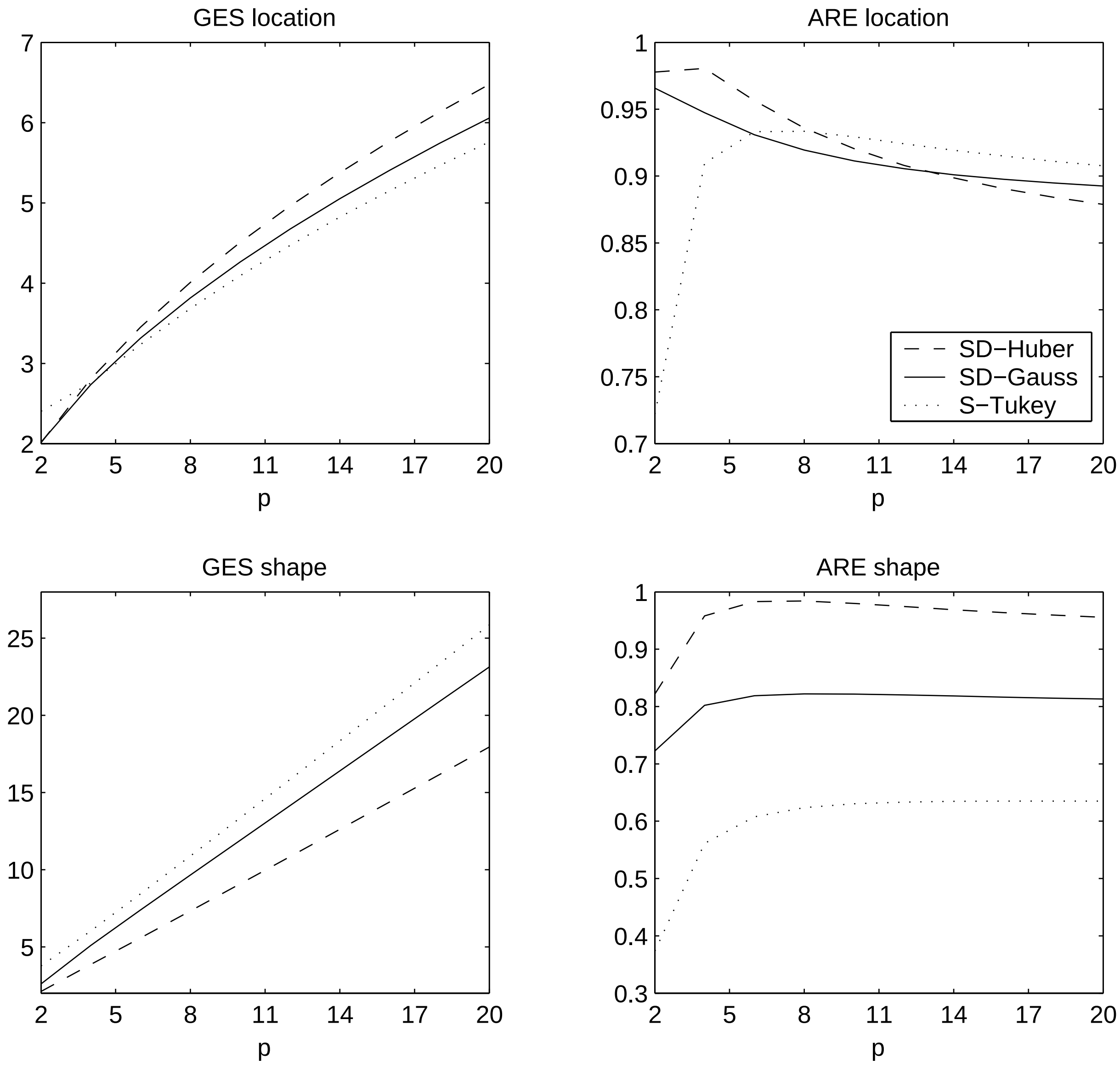


**Figure 5:** Gross-error sensitivities and asymptotic relative efficiencies of S-D estimators compared with S-estimators, for $\mathcal{T}_{p,3}(0, I)$ models.