Ann. Inst. Statist. Math.
Vol. 52, No. 2, 351–366 (2000)

ROBUSTNESS COMPARISONS OF SOME CLASSES OF LOCATION PARAMETER ESTIMATORS

JOHN R. COLLINS*

Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada T2N 1N4

(Received May 13, 1997; revised August 28, 1998)

Abstract. Asymptotic biases and variances of M-, L- and R-estimators of a location parameter are compared under ε-contamination of the known error distribution F_0 by an unknown (and possibly asymmetric) distribution. For each ε-contamination neighborhood of F_0, the corresponding M-, L- and R-estimators which are asymptotically efficient at the least informative distribution are compared under asymmetric ε-contamination. Three scale-invariant versions of the M-estimator are studied: (i) one using the interquartile range as a preliminary estimator of scale: (ii) another using the median absolute deviation as a preliminary estimator of scale; and (iii) simultaneous M-estimation of location and scale by Huber's Proposal 2. A question considered for each case is: when are the maximal asymptotic biases and variances under asymmetric ε-contamination attained by unit point mass contamination at ∞? Numerical results for the case of the ε-contaminated normal distribution show that the L-estimators have generally better performance (for small to moderate values of ε) than all three of the scale-invariant M-estimators studied.

Key words and phrases: Robust estimation, M-, L- and R-estimators, asymptotic biases, asymptotic variances, asymmetric contamination.

1. Introduction and summary

This paper carries out a comparative study of the robust M-, L- and R-estimators of location that are derived from the asymptotic minimax theory for robust estimators (Huber (1981)). Comparisons of maximal asymptotic biases and variances are carried out when the error distribution model is the asymetrically ε-contaminated normal distribution. In all the asymptotics, the proportion of contamination, ε, remains fixed as the sample size approaches ∞.

In Section 2, the comparisons are described. The key idea is to parameterize the comparisons on the value of ε. For each $\varepsilon > 0$, consider the M-, L- and R-estimators of location (represented by the functionals $T_{M,\varepsilon}$, $T_{L,\varepsilon}$ and $T_{R,\varepsilon}$, respectively) which are asymptotically efficient at the least informative ε-contaminated normal distribution. Under *symmetric* ε-contamination each of the three estimators is unbiased and each has the same maximal asymptotic variance (the reciprocal of the Fisher information of the least informative distribution). The program is then to compare the maximal asymptotic biases and variances of the same three estimators under asymmetric contamination. Although the estimators $\{T_{M,\varepsilon}, T_{L,\varepsilon}, T_{R,\varepsilon}\}$ have equivalent behavior (i.e., unbiased with tied asymptotic variances) under symmetric contamination, both their influence functions and their maximal asymptotic biases and variances differ under asymmetric

* Supported by Natural Sciences and Engineering Research Council of Canada Grant A-4499.

contamination. So restriction of the comparisons to triples $\{T_{M,\varepsilon}, T_{R,\varepsilon}, T_{L,\varepsilon}\}$ provides a reasonable and concise way of quantifying the differences in behavior of the three classes of estimators under asymmetric contamination.

Section 3 presents computations of the asymptotic biases and variances of the three estimators at F_∞, the stochastically largest distribution in the ε-contamination neighborhood. For each estimator, the question of whether the suprema of the asymptotic biases and variances are attained at F_∞ is then considered. Theorem 1 shows that the maximal asymptotic variance of M-estimators with monotone score functions is attained at F_∞.

Although the L- and R-estimators studied in Section 3 are scale-invariant, the M-estimator is not. Section 4 presents the influence functions and asymptotic variance functionals for three scale-invariant versions of the M-estimator: (i) the M-estimator with the interquantile range (IQR) as preliminary scale estimator; (ii) the M-estimator with the median absolute deviation (MAD) as preliminary scale estimator; and (iii) the M-estimator of location obtained from Huber's Proposal 2.

Section 5 presents computations of asymptotic biases and variances of the scale-invariant M-estimators at F_∞. The question of whether the suprema of the asymptotic biases and variances are attained at F_∞ is then considered. Theorem 2 shows that the supremum of the asymptotic bias of the M-estimator with IQR as preliminary scale estimator is attained at F_∞. Also calculations in Section 5 show, for M-estimators with either IQR or MAD as preliminary scale estimators, that the supremum of the asymptotic variance is *not* attained at F_∞. Larger asymptotic variances can be achieved by moving some of the contaminating mass away from ∞ to a neighborhood of a discontinuity of the influence function of the estimator.

The conclusions of the study are given in Section 6. The main conclusions are:

(1) When the scale parameter is known, M-estimators have smaller maximal asymptotic biases and variances than both R- and L-estimators under asymmetric contamination.

(2) When the scale parameter is unknown, the L-estimators have smaller maximal asymptotic biases and variances than all three scale-invariant versions of the M-estimators for values of ε ranging from 0.10 to 0.30.

Remark 1. The particular M- and L-estimators studied here (i.e., the Huber M-estimator, especially its scale-invariant versions, and the α-trimmed mean) are often used in practice. The corresponding R-estimator, although not used in practice because of its computational complexity, is included here for its theoretical interest. The asymptotic theory is relevant to practice when the tuning constants for the estimators correspond to the proportion of contamination ε. In applications one never really knows ε exactly; however, there are situations in which one has an approximate idea of the value of ε based on, say, previous data sets of a similar type. In these cases, the asymptotic theory provides a good guide to the choice of estimator for reasonably large sample sizes.

Remark 2. With ε fixed, the asymptotic comparisons of estimators depend on the value of ε. Fortunately our results yield qualitatively uniform comparisons over reasonably wide ranges of values of ε. An alternative to fixed-ε asymptotics is the local asymptotic approach, in which $\varepsilon \to 0$ at rate $1/\sqrt{n}$ as the sample size $n \to \infty$. In this framework Huber M-estimators and minimum distance estimators are known to be asymptotically optimal. A good guide to the local asymptotic approach literature is Rieder (1994) — see Chapter 5 in particular. An interesting open question is whether the fixed-ε results here have any connection to the local asymptotic optimality theory.

2. The estimators and their breakdown points

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution $F((x-\theta)/\sigma)$, where θ is an unknown location parameter to be estimated, and σ is a known scale parameter. All estimators of θ considered here will be location-invariant, so without loss of generality we assume that $\theta = 0$; also we assume that the known value of σ is 1. Assume that F is an unknown member of an ε-contaminated neighborhood, defined by

$$\mathcal{P}_{F_0,\varepsilon} = \{F : F = (1-\varepsilon)F_0 + \varepsilon G \\ \text{for some unknown and possibly asymmetric distribution } G\},$$

where the contamination proportion $\varepsilon(0 < \varepsilon < 1)$ is assumed to be known, and the fixed distribution F_0 is assumed to satisfy:

ASSUMPTION 1. F_0 is absolutely continuous; and its density $f_0 = F_0'$ is symmetric about 0, with $f_0(x)$ strictly decreasing in $x > 0$.

In this study, F_0 will sometimes also be assumed to satisfy:

ASSUMPTION 2. The density f_0 is absolutely continuous, and the function $\xi_0(x) = -f_0'(x)/f_0(x)$ is monotone nondecreasing in x.

All calculations will be done for the special case of the standard normal distribution $F_0(x) = \Phi(x) = \int_{-\infty}^{x} \phi(t)dt$, where $\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$.

We now summarize the asymptotic minimax theory for three classical types of estimators of the location parameter θ; see Huber (1981) for details. All three types of estimators (M, L and R) are generated from functionals $T(F)$ by taking the estimator to be $T(F_n)$, where F_n is the empirical distribution function of the random sample X_1, $X_2, \ldots, X_n$. The M-, L- and R-estimators arise as solutions of $\int \psi(x - T(F))dF(x) = 0$, $T(F) = \int xm(F(x))dF(x)$ and $\int J\{\frac{1}{2}[s + 1 - F(2T(F) - F^{-1}(s))]\}ds = 0$, respectively. When F_0 satisfies Assumptions 1 and 2, there is a unique $F_\varepsilon^* \in \mathcal{P}_{F_0,\varepsilon}$ which has minimal Fisher information. Futhermore one can choose the score functions ψ, m and J so that the resulting M-, L- and R-estimators are asymptotically efficient at F_ε^*. These estimators will be denoted throughout by $T_{M,\varepsilon}$, $T_{L,\varepsilon}$ and $T_{R,\varepsilon}$. In the important special case $F_0 = \Phi$, they are given as follows:

$T_{M,\varepsilon}$ is the solution of $\int \psi_c(x - T(F))dF(x) = 0$ where

$$\psi_c(x) = \max[-c, \min(c, x)], \tag{2.1}$$

with c determined from ε by

$$\frac{2\phi(c)}{c} - 2\Phi(-c) = \frac{\varepsilon}{1-\varepsilon}; \tag{2.2}$$

$T_{L,\varepsilon}$ is the α-trimmed mean (L-estimator with $m(t) = 1/(1-2\alpha)$ for $\alpha < t < 1-\alpha, = 0$ otherwise) with

$$\alpha = (1-\varepsilon)\Phi(-c) + \varepsilon/2; \tag{2.3}$$

and $T_{R,\varepsilon}$ is the R-estimator defined through

$$J(t) = \begin{cases} -c, & t \le \alpha \\ \Phi^{-1}\left(\frac{t-\varepsilon/2}{1-\varepsilon}\right), & \alpha \le t \le 1-\alpha \\ c, & t \ge 1-\alpha, \end{cases} \tag{2.4}$$

where c and α are given by (2.2) and (2.3).

Under mild regularity conditions on F, each of the estimators $T_{M,\varepsilon}$, $T_{L,\varepsilon}$ and $T_{R,\varepsilon}$ (represented generically by T_ε below) has as asymptotic distribution of $n^{1/2}[T_\varepsilon(F_n) - T_\varepsilon(F)]$ the normal distribution with mean 0 and variance $V(T_\varepsilon, F)$, where

$$V(T_\varepsilon, F) = \int \mathrm{IC}^2(x; F, T_\varepsilon)dF(x) \tag{2.5}$$

and where $\mathrm{IC}(x; F, T_\varepsilon)$ is the influence function of T_ε at F. Let $\mathcal{P}^s_{F_0,\varepsilon}$ denote the subset of $\mathcal{P}_{F_0,\varepsilon}$ consisting of distributions symmetric about 0. Then, under Assumptions 1 and 2, as F varies over $\mathcal{P}^s_{F_0,\varepsilon}$, each of the three estimators are unbiased and have the same maximal asymptotic variance, namely $1/I(F^*_\varepsilon)$, where $I(\cdot)$ denotes Fisher information for location. This minimax result was obtained by Huber (1964) for M-estimators; it was later extended by Jaeckel (1971) for L- and R-estimators.

To study the effects of asymmetric contamination, our aim is to compare, for each $\varepsilon > 0$, the maximal asymptotic biases

$$\sup\{|T_\varepsilon(F)| : F \in \mathcal{P}_{F_0,\varepsilon}\} \tag{2.6}$$

and the maximal asymptotic variances

$$\sup\{V(T_\varepsilon, F) : F \in \mathcal{P}_{F_0,\varepsilon}\} \tag{2.7}$$

as T_ε ranges over $\{T_{M,\varepsilon}, T_{L,\varepsilon}, T_{R,\varepsilon}\}$. What we will actually compute in the next section, for $F_0 = \Phi$, is $T_\varepsilon(F_\infty)$ and $V(T_\varepsilon, F_\infty)$ where $F_\infty = (1-\varepsilon)\Phi + \varepsilon\delta_\infty$, the stochastically largest member of $\mathcal{P}_{\Phi,\varepsilon}$. In general $T_\varepsilon(F_\infty)$ and $V(T_\varepsilon, F_\infty)$ are lower bounds for (2.6) and (2.7), respectively. For each estimator, we will consider the question of whether the maximal asymptotic biases and variances over $\mathcal{P}_{F_0,\varepsilon}$ are attained at F_∞. This question will also be considered for three scale-invariant versions of M-estimators studied in Section 5.

Because the estimators $T_{M,\varepsilon}$, $T_{L,\varepsilon}$ and $T_{R,\varepsilon}$ depend on ε (rather than staying fixed as $\varepsilon \to 0$), a slight variant of the usual notion of breakdown point is appropriate. For a class of estimators $\{T_\varepsilon\}$ indexed by ε, define the breakdown point as the supremum of the values of ε for which $\sup\{T_\varepsilon(F) : F \in \mathcal{P}_{F_0}\} < \infty$. We now calculate the breakdown points for $\{T_{M,\varepsilon}\}$, $\{T_{L,\varepsilon}\}$ and $\{T_{R,\varepsilon}\}$ in the special case when $F_0 = \Phi$.

Since the Huber M-estimator, with ψ_c given by (2.1), has the (usual) breakdown point $\frac{1}{2}$ for each fixed value of $c > 0$, it is clear that $\varepsilon^*_{T_M} = \frac{1}{2}$.

Since $T_{L,\varepsilon}$ is an α-trimmed mean, no breakdown occurs as long as ε and α in equations (2.2) and (2.3) satisfy $\varepsilon < \alpha$. Breakdown (i.e., $\varepsilon = \alpha$) occurs when the value of c determined by ε in (2.2) satisfies $\phi(c) - 2c\Phi(-c) = 0$, from which calculation yields $\varepsilon^*_{T_L} = 0.3509$.

For R-estimators of location, there is no breakdown as long as the constants ε, c and α and the function J determined by equations (2.2), (2.3) and (2.4) satisfy

$$\int_{1-\varepsilon/2}^{1} J(s)ds - \int_{1/2}^{1-\varepsilon/2} J(s)ds < 0,$$

which yields (after further calculation):

$$c(\varepsilon-\alpha)-\int_{1/2}^{1-\alpha}\Phi^{-1}\left(\frac{t-\varepsilon/2}{1-\varepsilon}\right)dt<0. \tag{2.8}$$

Note that (2.8) holds when $\varepsilon = \alpha$ (corresponding to breakdown for T_L), showing that $\varepsilon^*_{T_R} > \varepsilon^*_{T_L}$. Further calculation shows that breakdown occurs when $c(\varepsilon-\alpha)-(1-\varepsilon)[\phi(0)-\phi(c)]=0$, yielding $\varepsilon^*_{T_R}=0.4465$.

3. Asymptotic bias and variance comparisons when σ is known

Columns (3)–(5) of Table 1 present the asymptotic biases of $T_{M,\varepsilon}$, $T_{R,\varepsilon}$ and $T_{L,\varepsilon}$ at $F_\infty=(1-\varepsilon)\Phi+\varepsilon\delta_\infty$ for various values of ε. The values of c in Column (2) are determined from ε by formula (2.2), and the asymptotic biases are given in the normalized form $T(F_\infty)/\varepsilon$. More explicitly, $T_{M,\varepsilon}(F_\infty)$ is the solution t_∞ of

$$(1-\varepsilon)\int\psi_c(x-t_\infty)\phi(x)dx+\varepsilon c=0, \tag{3.1}$$

where ψ_c is given by (2.1) and (2.2); $T_{L,\varepsilon}(F_\infty)$ simplifies to

$$\frac{1-\varepsilon}{1-2\alpha}\left[\phi\left(\Phi^{-1}\left(\frac{\alpha}{1-\varepsilon}\right)\right)-\phi\left(\Phi^{-1}\left(\frac{1-\alpha}{1-\varepsilon}\right)\right)\right];$$

and $T_{R,\varepsilon}(F_\infty)$ is the solution t_∞ of

$$\begin{aligned}c(1-s_2-s_1)+\int_{s_1}^{s_2}\Phi^{-1}\Bigg[\frac{1}{2(1-\varepsilon)}\Bigg[s+1-(1-\varepsilon)\\ \cdot\Phi\left(2t_\infty-\Phi^{-1}\left(\frac{s}{1-\varepsilon}\right)\right)\Bigg]-\varepsilon\Bigg]ds=0,\end{aligned} \tag{3.2}$$

where s_1 and s_2 are the solutions of

$$\frac{1}{2}\left[s_1+1-(1-\varepsilon)\Phi\left(2t_\infty-\Phi^{-1}\left(\frac{s_1}{1-\varepsilon}\right)\right)\right]=\alpha \quad \text{and}$$
$$\frac{1}{2}\left[s_2+1-(1-\varepsilon)\Phi\left(2t_\infty-\Phi^{-1}\left(\frac{s_2}{1-\varepsilon}\right)\right)\right]=1-\alpha.$$

Columns (3)–(5) of Table 2 present the asymptotic variances of $T_{M,\varepsilon}$, $T_{R,\varepsilon}$ and $T_{L,\varepsilon}$ at $F_\infty=(1-\varepsilon)\Phi+\varepsilon\delta_\infty$. For comparison, the corresponding values of $1/I(F^*_\varepsilon)$, the minimax asymptotic variance under symmetric ε-contamination, are given in Column (2). For the M-estimator, one obtains

$$V(T_{M,\varepsilon},F_\infty)=\frac{\int\psi_c^2(x-t_\infty)\phi(x)dx+\varepsilon c^2}{(1-\varepsilon)^2[\Phi(t_\infty+c)-\Phi(t_\infty-c)]^2},$$

where t_∞ is given by (3.1). For the L-estimator, one obtains

$$\begin{aligned}V(T_{L,\varepsilon},F_\infty)=(1-2\alpha)^{-2}\Bigg\{(1-\varepsilon)\int_a^b x^2\phi(x)dx+\alpha(a^2+b^2)\\ -[(1-\varepsilon)(\phi(a)-\phi(b))+\alpha(a+b)]^2\Bigg\},\end{aligned}$$

Table 1. Asymptotic biases under $F_\infty = (1-\varepsilon)\Phi + \varepsilon\delta_\infty$. Tabled values are $|T_\varepsilon(F_\infty)|/\varepsilon$.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		scale known					
			scale unknown				
ε	c	$T_{M,\varepsilon}$	$T_{R,\varepsilon}$	$T_{L,\varepsilon}$	$T_{M,\varepsilon}$ MAD	$T_{M,\varepsilon}$ IQR	$T_{M,\varepsilon}$ Prop2
0.01	1.945	2.072	2.073	2.075	2.091	2.091	2.105
0.02	1.717	1.917	1.919	1.921	1.948	1.949	1.965
0.05	1.398	1.758	1.761	1.766	1.816	1.818	1.835
0.10	1.140	1.704	1.713	1.720	1.800	1.806	1.818
0.15	0.980	1.727	1.744	1.754	1.862	1.883	1.880
0.20	0.862	1.792	1.820	1.832	1.974	2.041	1.993
0.25	0.766	1.892	1.936	1.954	2.140	∞	2.166
0.30	0.685	2.034	2.104	2.135	2.383	∞	2.437
0.3509	0.612	2.240	2.362	∞	2.762	∞	2.968
0.40	0.549	2.540	2.802	∞	3.365	∞	5.114
0.4465	0.495	3.031	∞	∞	4.411	∞	∞
0.50	0.436	∞	∞	∞	∞	∞	∞
Is sup $\|T(F)\|$ attained at F_∞?		yes	yes	yes	not known	yes (Thm 2)	yes

where $a = \Phi^{-1}(\alpha/(1-\varepsilon))$, $b = \Phi^{-1}((1-\alpha)/(1-\varepsilon))$ and α is given by (2.2) and (2.3). For the R-estimator, one obtains

$$V(T_{R,\varepsilon}, F_\infty) = \frac{\int U^2(x)dF_\infty(x) - \left[\int U(x)dF_\infty(x)\right]^2}{\left[\int U'(x)dF_\infty(x)\right]^2},$$

where $U(x)$ is an indefinite integral of

$$U'(x) = J'\left\{\frac{1}{2}\left[s + 1 - (1-\varepsilon)\Phi\left(2t_\infty - \Phi^{-1}\left(\frac{s}{1-\varepsilon}\right)\right)\right]\right\}(1-\varepsilon)\phi(2t_\infty - x)$$

where J is given as before, and t_∞ is the solution of (3.2). The calculation of $V(T_{R,\varepsilon}, F_\infty)$ was carried out by numerical integration.

We now consider the question of whether the maximal asymptotic biases and variances over $\mathcal{P}_{\Phi,\varepsilon}$ are attained at F_∞.

For the asymptotic biases, we have that $\sup\{|T(F)| : F \in \mathcal{P}_{\Phi,\varepsilon}\} = T(F_\infty)$ for each of $T_{M,\varepsilon}$, $T_{L,\varepsilon}$ and $T_{R,\varepsilon}$ by monotonicity; see Chapter 3 of Huber (1981).

For the asymptotic variances, the identity

$$\sup\{V(T,F) : F \in \mathcal{P}_{F_0,\varepsilon}\} = V(T, F_\infty) \tag{3.3}$$

is proved for $F_0 = \Phi$ and $T = T_{L,\varepsilon}$ in Collins (1986). For $T = T_{R,\varepsilon}$, it is not known whether (3.3) is true. For $T = T_{M,\varepsilon}$, (3.3) follows as a special case of Theorem 1 below.

Define Ψ to be the class of functions ψ which satisfy: (i) $\psi : \mathcal{R} \to \mathcal{R}$ is continuous, odd, monotone nondecreasing (but not $\equiv 0$) with $\psi(x) \to \psi(\infty) < \infty$ as $x \to \infty$; and (ii) ψ' (at points where it exists) is monotone nondecreasing on $[0,\infty]$ (so in particular that $\psi'(x) \downarrow 0$ as $x \to \infty$). Define the M-estimator $T_\psi(F)$ to be the solution to

$$\int \psi(x - T_\psi(F))dF = 0. \tag{3.4}$$

Table 2. Asymptotic variances under $F_\infty = (1-\varepsilon)\Phi + \varepsilon\delta_\infty$. (See Section 5 for an explanation of the additional values in parentheses in Columns (6) and (7).)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		scale known					
			scale unknown				
ε	$\frac{1}{I(F_\varepsilon^*)}$	$T_{M,\varepsilon}$	$T_{R,\varepsilon}$	$T_{L,\varepsilon}$	$T_{M,\varepsilon}$ MAD	$T_{M,\varepsilon}$ IQR	$T_{M,\varepsilon}$ Prop2
0.01	1.065	1.066	1.067	1.069	1.067 (1.102)	1.087 (1.091)	1.067
0.02	1.116	1.118	1.122	1.125	1.120 (1.181)	1.153 (1.163)	1.123
0.05	1.256	1.266	1.282	1.289	1.279 (1.398)	1.335 (1.365)	1.289
0.10	1.490	1.534	1.582	1.598	1.586 (1.787)	1.679 (1.752)	1.650
0.15	1.748	1.869	1.974	2.005	2.012 (2.290)	2.196 (2.324)	2.098
0.20	2.046	2.323	2.531	2.600	2.679 (3.015)	3.341 (3.556)	2.875
0.25	2.397	2.991	3.410	3.586	3.838 (4.174)	∞ (∞)	4.378
0.30	2.822	4.073	5.009	5.645	6.331 (6.229)	∞ (∞)	8.284
0.3509	3.353	6.150	8.866	∞	12.50 (10.93)	∞ (∞)	28.90
0.40	3.996	10.96	24.79	∞	30.58 (23.26)	∞ (∞)	1232.
0.4465	4.765	28.27	∞	∞	104.7 (71.94)	∞ (∞)	∞
0.50	5.928	∞	∞	∞	∞ (∞)	∞ (∞)	∞
Is sup $V(T,F)$ attained at F_∞?		yes (Thm 1)	not known	yes	no (Sec 5)	no (Sec 5)	not known

and define $V(T_\psi, F)$ by

$$V(T_\psi, F) = \frac{\int \psi^2(x - T_\psi(F))dF}{[\int \psi'(x - T_\psi(F))dF]^2}. \tag{3.5}$$

We note that when $\psi \in \Psi$ and F_0 satisfies Assumption 1, there is a dense subset of $\mathcal{P}_{F_0,\varepsilon}$ (containing F_∞) for which the estimator $T_\psi(F)$ is uniquely determined by (3.4) and its asymptotic variance is given by (3.5).

THEOREM 1. *Let $\psi \in \Psi$, $0 < \varepsilon < \frac{1}{2}$ and let F_0 be a distribution satisfying Assumption* 1. *Then*

$$\sup\{V(T_\psi, F) : F \in \mathcal{P}_{F_0,\varepsilon}\} = V(T_\psi, F_\infty)$$

where $F_\infty = (1-\varepsilon)F_0 + \varepsilon\delta_\infty$.

A proof of Theorem 1 is given in the Appendix.

Conclusions following from the calculations and results in this section are given in Section 6.

4. Scale invariant M-estimators

Consider the model $F((x-\theta)/\sigma)$, with $F \in \mathcal{P}_{F_0,\varepsilon}$ with both θ and σ unknown. Since the L- and R-estimators of θ are scale-invariant, the asymptotic bias and variance comparisons of $T_{L,\varepsilon}$ and $T_{R,\varepsilon}$ in the previous section go through in the scale unknown case. However the M-estimators $T_{M,\varepsilon}$ are not scale-invariant, so they will now be replaced in the comparisons by some scale-invariant versions. Without loss of generality, we will assume that $\theta = 0$ and $\sigma = 1$.

One well-known scale-invariant version is the M-estimator of location with a preliminary estimator of scale, defined by

$$\int \psi\left(\frac{x-T(F)}{S(F)}\right) dF = 0, \tag{4.1}$$

where $S(F)$ is a scale functional. We consider two different scale functionals: the α-interquantile range

$$S_\alpha(F) = \frac{F^{-1}(1-\alpha) - F^{-1}(\alpha)}{k} \tag{4.2}$$

and the symmetrized α-interquantile range

$$\tilde{S}_\alpha(F) = S_\alpha(\tilde{F}) = \frac{\tilde{F}^{-1}(1-\alpha) - \tilde{F}^{-1}(\alpha)}{k}, \tag{4.3}$$

where

$$\tilde{F}(x) = \frac{1}{2}\left\{F(x) + 1 - F\left[2F^{-1}\left(\frac{1}{2}\right) - x - 0\right]\right\}. \tag{4.4}$$

Here k is a positive constant; in the calculations of the next section, we set $k = F_0^{-1}(1-\alpha) - F_0^{-1}(\alpha)$ to ensure Fisher-consistency of both S_α and $\tilde{S}_\alpha$ at the uncontaminated symmetric F_0. For the special case α=.25, $S_{.25}$ is the IQR and $\tilde{S}_{.25}$ is the MAD.

Another well-known scale-invariant method is by simultaneous estimation of location and scale: that is, to solve the system of equations

$$\int \psi\left(\frac{x-T(F)}{S(F)}\right) dF = 0 \tag{4.5}$$

and

$$\int \chi\left(\frac{x-T(F)}{S(F)}\right) dF = 0 \tag{4.6}$$

for $(T(F), S(F))$. This is considered here for the particular choice of score functions known as Huber's Proposal 2 (Huber (1964, 1981)): set $\psi = \psi_c$ (as in Section 2) and $\chi(x) = \psi_c^2(x) - \beta(c)$, with $\beta(c) = \int \psi_c^2(x) dF_0(x)$ to ensure Fisher-consistency at F_0.

Note that the breakdown points for the M-estimators of location with preliminary estimators of scale are the breakdown points of the latter: 0.25 when $S = S_{.25}$ and 0.5 when $S = \tilde{S}_{.25}$. For Huber's Proposal 2, combining the breakdown point for fixed c, $\beta(c)/[\beta(c)+c^2]$ (Huber (1981), p. 143), with the relation (2.2) easily yields $\varepsilon^* = 0.4255$.

The influence function of an M-estimator of location with preliminary estimator of scale $S(F)$ is given by:

$$(4.7)\quad \mathrm{IC}(x;F,T) = \frac{\psi\left(\frac{x-T(F)}{S(F)}\right)S(F) - \mathrm{IC}(x;F,S)\int \frac{x-T(F)}{S(F)}\psi'\left(\frac{x-T(F)}{S(F)}\right)F(dx)}{\int \psi'\left(\frac{x-T(F)}{S(F)}\right)F(dx)},$$

where $\mathrm{IC}(x;F,S)$ is the influence function of $S(F)$. Hence the corresponding asymptotic variance functional (see formula (2.5)) is given by:

$$(4.8)\quad \begin{aligned} V(T,F) = \Bigg\{ &\int \psi^2\left(\frac{x-T(F)}{S(F)}\right)F(dx) \\ &- \frac{2}{S(F)}\int \frac{x-T(F)}{S(F)}\psi'\left(\frac{x-T(F)}{S(F)}\right)F(dx) \\ &\cdot \int \psi\left(\frac{x-T(F)}{S(F)}\right)\mathrm{IC}(x;F,S)F(dx) \\ &+ \left[\frac{1}{S(F)}\int \frac{x-T(F)}{S(F)}\psi'\left(\frac{x-T(F)}{S(F)}\right)F(dx)\right]^2 \\ &\cdot \left[\int \mathrm{IC}^2(x;F,S)F(dx)\right]^2 \Bigg\} \\ \div &\left[\frac{1}{S(F)}\int \psi'\left(\frac{x-T(F)}{S(F)}\right)F(dx)\right]^2. \end{aligned}$$

The influence function of the α-interquantile range $S_\alpha(F) = [F^{-1}(1-\alpha) - F^{-1}(\alpha)]/k$ is given by (Hampel *et al.* (1986), p. 110)

$$(4.9)\quad k\cdot \mathrm{IC}(x;F,S_\alpha) = \begin{cases} \dfrac{1-\alpha}{f(a)} - \dfrac{\alpha}{f(b)}, & \text{for } x < a, \\ -\alpha\left[\dfrac{1}{f(a)} + \dfrac{1}{f(b)}\right], & \text{for } a < x < b, \\ \dfrac{1-\alpha}{f(b)} - \dfrac{\alpha}{f(a)}, & \text{for } x > b, \end{cases}$$

where $a = F^{-1}(\alpha)$ and $b = F^{-1}(1-\alpha)$. For the symmetrized version $\tilde{S}_\alpha(F)$, the influence function is given by (Collins (1991), formula (3.6)):

$$(4.10)\quad k\cdot[(f(\tilde{a}) + f(\tilde{b}))]\cdot \mathrm{IC}(x;F,\tilde{S}_\alpha) = \begin{cases} 1-2\alpha-(f(\tilde{a})-f(\tilde{b}))/[2f(m)], & \text{for } x<\tilde{a}, \\ -2\alpha-(f(\tilde{a})-f(\tilde{b}))/[2f(m)], & \text{for } \tilde{a}<x<m, \\ -2\alpha+(f(\tilde{a})-f(\tilde{b}))/[2f(m)], & \text{for } m<x<\tilde{b}, \\ 1-2\alpha+(f(\tilde{a})-f(\tilde{b}))/[2f(m)], & \text{for } x>\tilde{b}; \end{cases}$$

where $m = F^{-1}(1/2)$, $\tilde{a} = \tilde{F}^{-1}(\alpha)$ and $\tilde{b} = \tilde{F}^{-1}(1-\alpha) = 2m - \tilde{a}$. In order for formula (4.9) ((4.10)) to make sense, we assume that F has a density f in some neighborhood of each of the points a and b ($\tilde{a}, m$ and $\tilde{b}$).

For simultaneous estimation of location and scale, the influence function of $T(F)$ is

$$\text{(4.11)} \qquad \mathrm{IC}(x;F,T)=\frac{\psi(y)S(F)\int\chi'(y)yF(dx)-\chi(y)S(F)\int\psi'(y)yF(dx)}{\int\psi'(y)F(dx)\int\chi'(y)yF(dx)-\int\chi'(y)F(dx)\int\psi'(y)yF(dx)}$$

where $y=(x-T(F))/S(F)$ and where $(T(F),S(F))$ is the solution to (4.5) and (4.6). The corresponding asymptotic variance functional is

$$\text{(4.12)} \quad V(T,F)=\Bigg\{\int\psi^2(y)S^2(F)F(dx)\left[\int\chi'(y)yF(dx)\right]^2 - 2\int\psi(y)\chi(y)S^2(F)F(dx)\left[\int\chi'(y)yF(dx)\right]\cdot\left[\int\psi'(y)yF(dx)\right] + \int\chi^2(y)S^2(F)F(dx)\left[\int\psi'(y)yF(dx)\right]^2\Bigg\} \div\left\{\int\psi'(y)F(dx)\int\chi'(y)yF(dx)-\int\chi'(y)F(dx)\int\psi'(y)yF(dx)\right\}^2,$$

where $y=(x-T(F))/S(F)$.

5. Asymptotic bias and variance comparisons for scale-invariant M-estimators

In this section we consider, for each of the three scale-invariant M-estimators: (i) computation of the asymptotic biases and variances under F_∞; and (ii) the question of whether the suprema of the asymptotic biases and variances over $\mathcal{P}_{F_0,\varepsilon}$ are attained at F_∞.

Inserting $F_\infty=(1-\varepsilon)F_0+\varepsilon\delta_\infty$ into (4.1) and (4.2) yields that $T(F_\infty)$ is determined by

$$\text{(5.1)} \qquad (1-\varepsilon)\int\psi\left(\frac{x-T(F_\infty)}{S(F_\infty)}\right)f_0(x)dx+\varepsilon\psi(\infty)=0,$$

with $S(F_\infty)$ given by

$$\text{(5.2)} \qquad S(F_\infty)=S_\alpha(F_\infty)=\frac{\left[F_0^{-1}\left(\frac{1-\alpha}{\varepsilon}\right)-F_0^{-1}\left(\frac{\alpha}{1-\varepsilon}\right)\right]}{k}$$

for the M-estimator with α-interquantile range as auxiliary scale estimator. For the symmetrized version, substitution of F_∞ into (4.3) and (4.4) yields

$$\text{(5.3)} \qquad S(F_\infty)=\tilde{S}_\alpha(F_\infty)=\frac{2(m_0-\tilde{a}_0)}{k}$$

where $m_0=\Phi^{-1}(.5/(1-\varepsilon))$ and $\tilde{a}_0$ is the solution of

$$\text{(5.4)} \qquad (1-\varepsilon)[\Phi(2m_0-\tilde{a}_0)-\Phi(\tilde{a}_0)]=\frac{1}{2}.$$

The values of $T(F_\infty)/\varepsilon$ in Column (7) ((6)) of Table 1 were calculated from (5.1) and (5.2) ((5.1), (5.3) and (5.4)) with $F_0=\Phi$, $\alpha=.25$, $k=2\Phi^{-1}(.75)$, and $\psi=\psi_c$ with c

given by (2.2). The values in Column (8) were obtained by inserting $F_\infty = (1-\varepsilon)\Phi+\varepsilon\delta_\infty$, $\psi = \psi_c$ and $\chi = \psi_c^2 - \int \psi_c^2 d\Phi$ into (4.5) and (4.6) and solving for $(T(F_\infty), S(F_\infty))$.

The values of the asymptotic variances $V(T, F_\infty)$ for the three scale-invariant M-estimators, given in Columns (6)–(8) of Table 2, were computed from formulas (4.7)–(4.12) by inserting $F = F_\infty = (1-\varepsilon)\Phi + \varepsilon\delta_\infty$, $\alpha = .25$, $k = 2\Phi^{-1}(.75)$, $\psi = \psi_c$ with c given by (2.2), and $\chi = \psi_c^2 - \int \psi_c^2 d\Phi$.

Consider now the question of whether

$$\sup\{|T(F)| : F \in \mathcal{P}_{\Phi,\varepsilon}\} = T(F_\infty) \tag{5.5}$$

holds for the cases tabulated in Columns (6)–(8) of Table 1. For Huber's Proposal 2 (Column (8)), (5.5) holds as a special case of a result proved in Section 4 of Martin and Zamar (1993). For the M-estimator with MAD as preliminary scale estimator (Column (6)), it is not known whether (5.5) holds. For the M-estimator with IQR as preliminary scale estimator (Column (7)), (5.5) holds as a special case of the following theorem, which gives some general conditions on F_0 and ψ under which the maximal asymptotic bias is attained at F_∞.

THEOREM 2. *Let $\psi : \mathcal{R} \to \mathcal{R}$ be continuous, odd and monotone nondecreasing with $0 < \psi(\infty) < \infty$. Let $0 < \varepsilon < \alpha < \frac{1}{3}$, and suppose that F_0 satisfies Assumption* 1. *Let $\mathcal{P}'_{F_0,\varepsilon}$ be the subset of F's in $\mathcal{P}_{F_0,\varepsilon}$ for which F has a density f in some neighborhood of each of the points $F^{-1}(\alpha/(1-\varepsilon))$ and $F^{-1}((1-\alpha)/(1-\varepsilon))$. For $F \in \mathcal{P}'_{F_0,\varepsilon}$, let $T(F)$ be the (necessarily unique) solution of*

$$\int \psi\left(\frac{x - T(F)}{S_\alpha(F)}\right) dF = 0,$$

where $S_\alpha(F) = [F^{-1}(1-\alpha) - F^{-1}(\alpha)]/k$ for some fixed $k > 0$. Then $\sup\{T(F) : F \in \mathcal{P}'_{F_0,\varepsilon}\}$ *is attained at $F_\infty = (1-\varepsilon)F_0 + \varepsilon\delta_\infty$.*

A proof of Theorem 2 appears in the Appendix.

Consider now the question of whether $\sup\{V(T,F) : F \in \mathcal{P}_{\Phi,\varepsilon}\}$ is attained at F_∞. First consider the M-estimator with IQR as preliminary scale estimator. Since the influence function of S_α (formula (4.9)) is discontinuous at both $a = F^{-1}(\alpha)$ and $b = F^{-1}(1-\alpha)$, the influence function of the corresponding M-estimator (formula (4.7)) is also discontinuous. For F in $\mathcal{P}'_{F_0,\varepsilon}$, the term $\int \mathrm{IC}^2(x; F, S_\alpha) F(dx)$ in formula (4.8) is equal to

$$\frac{1}{k^2}\left[\frac{\alpha(1-\alpha)}{f^2(a)} + \frac{\alpha(1-\alpha)}{f^2(b)} - \frac{2\alpha^2}{f(a)f(b)}\right]. \tag{5.6}$$

Collins (1991) showed that the supremum of (5.6) is (depending on the values of α and ε) either attained at F_∞ or is equal to $\alpha(1-\alpha)/\{k^2(1-\varepsilon)^2 f_0^2[F_\infty^{-1}(\alpha)]\}$. The latter is achieved as a limit of a sequence of F's in $\mathcal{P}'_{F_0,\varepsilon}$ with most of the contaminating mass near ∞ but with a small (approaching 0) proportion of the mass in a neighborhood of $a = F^{-1}(\alpha)$ with $f(a)$ approaching ∞. This suggests that the asymptotic variance functional (4.8) can likewise sometimes be inflated by moving some of the mass in F_∞ to a neighborhood of the discontinuity at a.

The values in parentheses in Column (7) of Table 2 were obtained from (4.8) and (4.9) by formally substituting $f(a) = \infty$, $f(b) = (1-\varepsilon)f_0[F_\infty^{-1}(1-\alpha)]$ and $F(dx) =$

$F_\infty(dx)$. These values, clearly attainable as a limit for a sequence of F's in $\mathcal{P}_{\Phi,\varepsilon}$, exceed $V(T, F_\infty)$ for the tabulated values of ε up through 0.20, so that we have

$$V(T, F_\infty) < \sup\{V(T, F) : F \in \mathcal{P}'_{\Phi,\varepsilon}\}. \tag{5.7}$$

Anologous calculations for the case of the M-estimator with MAD as preliminary scale estimator yield the values shown in parentheses in Column (6) of Table 2. In this case (5.7) holds for the tabulated values of ε up through 0.25. In both the cases of the IQR and the MAD as preliminary scale estimates, it is not known whether the values in parentheses — now established as lower bounds for $\sup V(T, F)$ — are equal to $\sup V(T, F)$.

In the case of Huber's Proposal 2, the influence function of the estimator is continuous, but it is not known whether the values of $V(T, F_\infty)$ in Column (8) of Table 2 are equal to $\sup V(T, F)$.

6. Conclusions

The bottom row of Table 1 (Table 2) summarizes answers to the question of whether $\sup\{|T(F)| : F \in \mathcal{P}_{\Phi,\varepsilon}\}$ $[\sup\{(V, F) : F \in \mathcal{P}_{\Phi,\varepsilon}\}]$ is attained over $\mathcal{P}_{\Phi,\varepsilon}$ at $F_\infty = (1 - \varepsilon)\Phi + \varepsilon\delta_\infty$, with citations to new results in this paper given in parentheses.

The following conclusions are drawn from Tables 1 and 2.

1. In the scale-known case (Columns (3)–(5)), using maximal asymptotic bias as the criterion, the M-estimator outperforms the R-estimator, which in turn outperforms the L-estimator, uniformly over the tabulated values of ε.

2. In the scale-known case with maximal asymptotic variance as criterion, the M-estimator outperforms both the R- and L-estimators for all the tabulated values of ε.

3. In the scale-unknown case (Columns (4)–(8)) with maximal asymptotic bias as criterion, the R-estimator outperforms the L-estimator, which in turn outperforms each of the three scale-invariant M-estimators for the tabulated values of ε up to 0.30.

4. In the scale-unknown case with the maximal asymptotic variance criterion, the L-estimator outperforms both of the M-estimators with preliminary scale estimates for the tabulated values of ε up to 0.30. The L-estimator also outperforms the Huber Proposal 2 M-estimator for values of ε ranging from 0.10 to 0.30, but comparison for ε ranging from 0.01 to 0.05 is not possible because the values in Column (8) are known only to be a lower bound on the supremum of the asymptotic variance over $\mathcal{P}_{\Phi,\varepsilon}$. Also although the tabulated values of $V(T_{R,\varepsilon}, F_\infty)$, for ε ranging from 0.02 to 0.40, are less than the established lower bounds on the maximal asymptotic variances of the four other scale-invariant estimators, no comparison can be made since the question of whether $V(T_{R,\varepsilon}, F_\infty)$ is equal to $\sup\{V(T_{R,\varepsilon}, F) : F \in \mathcal{P}_{\Phi,\varepsilon}\}$ is open.

Acknowledgements

I would like to thank the referees for valuable suggestions which greatly improved the presentation of the results.

Appendix

PROOF OF THEOREM 1. First note that

$$V(T_\psi, F_\infty) = \frac{(1-\varepsilon)\int \psi^2(x - t_\infty)^2 f_0(x)dx + \varepsilon\psi^2(\infty)}{\left[(1-\varepsilon)\int \psi'(x - t_\infty) f_0(x)dx\right]^2},$$

where t_∞ is the solution to

$$(1-\varepsilon)\int \psi(x-t_\infty)f_0(x)dx + \varepsilon\psi(\infty) = 0.$$

It suffices to show (see formula (3.5)), for all $F = (1-\varepsilon)F_0 + \varepsilon G$, that:

(a) $\int \psi^2(x - T_\psi(F))f_0(x)dx \le \int \psi^2(x-t_\infty)f_0(x)dx$;
(b) $\int \psi^2(x - T_\psi(F))dG(x) \le \psi^2(\infty)$;
(c) $\int \psi'(x - T_\psi(F))f_0(x)dx \ge \int \psi'(x-t_\infty)f_0(x)dx \ge 0$; and
(d) $\int \psi'(x - T_\psi(F))dG(x) \ge 0$.

Both (b) and (d) are immediate consequences of the definition of Ψ. To show (a), first note that t_∞, attained at F_∞, is the maximum possible value of $T_\psi(F)$ as F ranges over $\mathcal{P}_{F_0,\varepsilon}$. By symmetry, we need only consider the F's in $\mathcal{P}_{F_0,\varepsilon}$ for which $T_\psi(F) \in [0, t_\infty]$. Since ψ is odd and both ψ' and f_0 are even, differentiation with respect to t under the integral sign (justified by dominated convergence) yields:

$$\begin{aligned}\frac{d}{dt}\int \psi^2(x-t)f_0(x)dx &= -2\int \psi(x-t)\psi'(x-t)f_0(x)dx\\ &= 2\int_0^\infty \psi(y)\psi'(y)[f_0(t-y) - f_0(t+y)]dy \ge 0\end{aligned}$$

for all $t \in [0, t_\infty]$, since ψ and ψ' are ≥ 0 on $[0,\infty)$ and since $f_0(t-y) - f_0(t+y) \ge 0$ for all $t > 0$ and $y > 0$ by Assumption 1. Thus $\int \psi^2(x-t)f_0(x)dx$ takes its maximum value over $[0, t_\infty]$ at $t = t_\infty$, proving (a).

To show that $\int \psi'(x-t)f_0(x)dx$ attains its minimum over $[0, t_\infty]$ at $t = t_\infty$, first note that it follows from Assumption 1 that f_0' exists a.e. x, is odd, and satisfies $f_0'(x) < 0$ a.e. $x > 0$. Then we calculate that

$$\begin{aligned}\frac{d}{dt}\int \psi'(x-t)f_0(x)dx &= \frac{d}{dt}\int \psi'(x)f_0(x+t)dx\\ &= \int \psi'(x)f_0'(x)(x+t)dx\\ &= \int_0^\infty f_0'(x)[\psi'(x-t) - \psi'(x+t)]dx \le 0\end{aligned}$$

for all $t \in [0, t_\infty]$, since $f_0'(x) < 0$ a.e. $x > 0$ and $\psi'(x-t) - \psi'(x+t) \ge 0$ for all $x > 0$ and $t > 0$, since ψ' is even and monotone nonincreasing on $[0, \infty]$. This proves (c) and completes the proof of the theorem. □

PROOF OF THEOREM 2. First note that the range of values of $S_\alpha(F)$ as F varies over $\mathcal{P}_{F_0,\varepsilon}$ is the range of

$$s(\gamma) = \left[F_0^{-1}\left(\frac{1-\alpha-\varepsilon\gamma}{1-\varepsilon}\right) - F_0^{-1}\left(\frac{\alpha-\varepsilon\gamma}{1-\varepsilon}\right)\right]\Big/ k$$

as γ varies over $[0, \frac{1}{2}]$. Furthermore, $s(\gamma)$ is strictly decreasing in γ and $s(0) = \sup\{S_\alpha(F) : F \in \mathcal{P}_{F_0,\varepsilon}\} = S_\alpha(F_\infty)$ (see the proof of Theorem 1 of Collins (1991)).

By symmetry, we need only consider F's in $\mathcal{P}'_{F_0,\varepsilon}$ for which $T(F) \ge 0$. Also we will use the notation $a = F^{-1}(\alpha)$, $b = F^{-1}(1-\alpha)$ and $s = (b-a)/k$, where it is assumed throughout that the only triples (a, b, s) considered are those which are attainable by some $F \in \mathcal{P}'_{F_0,\varepsilon}$. With this notation, $T(F)$ is the solution t of

$$-\frac{1-\varepsilon}{\varepsilon}\int \psi\left(\frac{x-t}{s}\right) f_0(x)dx = \int \psi\left(\frac{x-t}{s}\right) G(dx), \tag{A.1}$$

where G necessarily satisfies the following two conditions:

$$G(a) = [\alpha - (1-\varepsilon)F_0(a)]/\varepsilon \tag{A.2}$$

and

$$G(a+sk) = [1 - \alpha - (1-\varepsilon)F_0(a+sk)]/\varepsilon. \tag{A.3}$$

By the conditions on ψ, it follows that: (i) the left side of (A.1) is an increasing function of t; (ii) the right side of (A.1) is decreasing in t for fixed G; and (iii) $\int \psi[(x-t)/s]G^*(dx) \geq \int \psi[(x-t)/s]G(dx)$ for all values of t whenever G^* is stochastically larger than G. Using the notation $t(G)$ for the solution t of (A.1) when s is fixed, an immediate consequence of (i), (ii) and (iii) is that $t(G^*) \geq t(G)$ whenever G^* is stochastically larger than G.

The stochastically largest distribution G satisfying both (A.2) and (A.3) clearly places mass $G(a)$ at a, mass $G(a+sk) - G(a)$ at $a+sk$ and mass $1 - G(a+sk)$ at ∞. The supremum of $T(F)$ over the subclass of $\mathcal{P}'_{F_0,\varepsilon}$ on which $F^{-1}(\alpha) = a$ and $S(F) = s$ is therefore the solution t of

$$\begin{aligned} -\frac{1-\varepsilon}{\varepsilon}&\int \psi\left(\frac{x-t}{s}\right) f_0(x)dx \\ &= m_1(a)\psi\left(\frac{a-t}{s}\right) + m_2(a)\psi\left(\frac{a+ks-t}{s}\right) + m_3(a)\psi(\infty), \end{aligned} \tag{A.4}$$

where

$$\begin{aligned} m_1(a) &= \frac{\alpha - (1-\varepsilon)F_0(a)}{\varepsilon}, \\ m_2(a) &= \frac{(1-2\alpha) - (1-\varepsilon)[F_0(a+ks) - F_0(a)]}{\varepsilon}, \quad \text{and} \\ m_3(a) &= 1 - \frac{(1-\alpha) - (1-\varepsilon)[F_0(a+ks)]}{\varepsilon}. \end{aligned}$$

Now consider finding the supremum of $T(F)$ over $\mathcal{P}'_{F_0,\varepsilon}$ subject only to the side condition $S_\alpha(F) = s$. The derivative of the right side of (A.4) with respect to a is

$$\begin{aligned} m_1(a)\frac{1}{s}\psi'\left(\frac{a-t}{s}\right) &+ m_2(a)\frac{1}{s}\psi'\left(\frac{a+ks-t}{s}\right) \\ &+ \frac{1-\varepsilon}{\varepsilon}\left\{ f_0(a+ks)\left[\psi(\infty) - \psi\left(\frac{a+ks-t}{s}\right)\right]\right. \\ &\left. + f_0(a)\left[\psi\left(\frac{a+ks-t}{s}\right) - \psi\left(\frac{a-t}{s}\right)\right]\right\}, \end{aligned}$$

which is nonnegative a.e. x. Thus the maximal t satisfying (A.4) occurs at the largest possible value of a corresponding to $S_\alpha(F) = s$. To find the latter, write the identity $s = (b-a)/k$ as

$$F_0^{-1}\left(\frac{\alpha - \varepsilon G(a)}{1-\varepsilon}\right) + ks = F_0^{-1}\left(\frac{1-\alpha-\varepsilon G(b)}{1-\varepsilon}\right), \tag{A.5}$$

and note that finding the largest possible value of a subject to this constraint is equivalent to finding the smallest value of $G(a)$, $0 \le G(a) \le \frac{1}{2}$, for which (A.5) is satisfied for some $G(b)$ with $G(a) \le G(b)$. As $G(a) \to 0$ the choice of $G(b)$ satisfying (A.5) is also decreasing, until a is large enough that $G(a) = G(b)$. Thus $a(s)$, the maximal possible value of a corresponding to $S_\alpha = s$, is given by

$$\text{(A.6)} \qquad a(s) = F_0^{-1}\left(\frac{\alpha - \varepsilon\gamma(s)}{1-\varepsilon}\right),$$

where $\gamma(s)$ is the unique number in $[0, \frac{1}{2}]$ determined from s by

$$\text{(A.7)} \qquad F_0^{-1}\left(\frac{1-\varepsilon-\alpha\gamma(s)}{1-\varepsilon}\right) - F_0^{-1}\left(\frac{\alpha-\varepsilon\gamma(s)}{1-\varepsilon}\right) = ks;$$

and the values of m_1, m_2 and m_3 at $a(s)$ are clearly $\gamma(s)$, 0 and $1-\gamma(s)$, respectively.

The final step is to show that $t(s)$, defined by

$$\text{(A.8)} \quad -(1-\varepsilon)\int \psi\left(\frac{x-t(s)}{s}\right) f_0(x)dx = \varepsilon\gamma(s)\psi\left(\frac{a(s)-t(s)}{s}\right) + [1-\gamma(s)]\psi(\infty),$$

is nondecreasing in s. First note that $\psi'(s) < 0$ at all possible values of s, and use (A.6) and (A.7) to calculate

$$\text{(A.9)} \qquad a'(s) = \frac{f_0(b(s))}{k[f_0(a(s)) - f_0(b(s))]} > 0,$$

where $b(s) = F_0^{-1}\left(\frac{1-\alpha-\varepsilon\gamma(s)}{1-\varepsilon}\right)$. The inequality in (A.9) follows from symmetric unimodality of F_0 (Assumption 1) and the fact that $b(s) \ge |a(s)|$ for all $\gamma(s) \in [0, \frac{1}{2}]$, with strict inequality when $\gamma \neq \frac{1}{2}$.

Differentiating (A.8) with respect to s yields the following identity after a rearrangement of terms:

$$\text{(A.10)} \quad \left[(1-\varepsilon)\int \frac{1}{s}\psi'\left(\frac{x-t(s)}{s}\right) f_0(x)dx + \varepsilon\gamma(s)\frac{1}{s}\psi'\left(\frac{a(s)-t(s)}{s}\right)\right] t'(s)$$
$$= -(1-\varepsilon)\int \frac{1}{s}\left(\frac{x-t(s)}{s}\right)\psi'\left(\frac{x-t(s)}{s}\right) f_0(x)dx + \varepsilon\gamma(s)\psi'\left(\frac{a(s)-t(s)}{s}\right)$$
$$\cdot\left(\frac{sa'(s)-a(s)+t(s)}{s^2}\right) - \gamma'(s)\left(\psi(\infty) - \psi\left(\frac{a(s)-t(s)}{s}\right)\right).$$

The coefficient of $t'(s)$ on the left side of (A.10) is clearly positive for all possible values of s. The first term on the right side of (A.10) can be written as

$$\frac{1-\varepsilon}{s^2}\int_0^\infty y\psi'(y)\,[f_0(t(s)-sy) - f_0(t(s)+sy)]\,dy,$$

which is nonnegative for all s since $t(s) \ge 0$ and F_0 satisfies Assumption 1. Since $\gamma(s)$, $a(s)$, $t(s)$ and ψ' are nonnegative, the second term on the right side will be nonnegative if it can be shown that $a(s) < 0$ for all possible values of s. But $a(s) \le F_0^{-1}(\alpha/(1-\varepsilon)) < F_0^{-1}(\alpha/(1-\alpha)) < F_0^{-1}\left(\frac{1}{2}\right) = 0$, since $0 < \varepsilon < \alpha < \frac{1}{3}$ by hypothesis. Finally the third term on the right side is nonnegative since $\gamma'(s) \le 0$ and ψ is monotone nondecreasing.

It now follows that $t'(s) \geq 0$ for all possible values of s. Since $t(s)$ is nondecreasing in s, $T(F)$ must attain its maximum over $\mathcal{P}'_{F_0,\varepsilon}$ at the same F where $S_\alpha(F)$ attains its maximum over $\mathcal{P}'_{F_0,\varepsilon}$, namely at $F = F_\infty$. $\square$

References

Collins, J. R. (1986). Maximum asymptotic variances of trimmed means under asymmetric contamination, *Ann. Statist.*, **14**, 348–354.

Collins, J. R. (1991). Maximal asymptotic biases and variances of symmetrized interquantile ranges under asymmetric contamination, *Statistics*, **22**, 379–402.

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

Huber, P. J. (1964). Robust estimation of a location parameter, *Ann. Math. Statist.*, **35**, 73–101.

Huber, P. J. (1981). *Robust Statistics*, Wiley, New York.

Jaeckel, L. A. (1971). Robust estimates of location: Symmetry and asymmetric contamination, *Ann. Math. Statist*, **42**, 1020–1034.

Martin, R. D. and Zamar, R. H. (1993). Bias robust estimation of scale, *Ann. Statist.*, **21**, 991–1017.

Rieder, H. (1994). *Robust Asymptotic Statistics*, Springer, New York.