DOE/ER/02310-167

PROGRESS REPORT ON
ROBUST ESTIMATES OF LOCATION
CALLED COMPROMISE ESTIMATES

by

Roberta Guarino
Princeton University

DISCLAIMER

This book was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

MASTER

Technical Report No. 167, Series 2
Department of Statistics
Princeton University
January 1980

Research on this report was supported in part by the Department of Energy under Contract No. EY-76-S-02-2310, awarded to the Department of Statistics, Princeton University, Princeton, New Jersey.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

A B S T R A C T

This is a progress report on new types of robust estimates of location called compromise estimates. Two types,

(i) psi-compromised M-estimates
and

(ii) gap-compromised estimates,
are discussed and compared via a simulation. The relationships between the two types of estimators are seen to depend on both the sample size and the underlying distribution of the data.

ACKNOWLEDGEMENTS

The author is grateful for the support of the Gurdien State Graduate Fellowship and would like to acknowledge many helpful discussions with P. Bloomfield and J. W. Tukey.

1. INTRODUCTION

This is a progress report on new types of robust estimates of location called *compromise* estimates. In particular, we will discuss and compare two types of compromise estimates. The first is a special kind of M-estimate called a psi-compromised M-estimate. The second type of estimate, a gap-compromised estimate, has as its underlying foundation the same working mechanism which has made many M-estimates so successful, namely, the downweighting of observations which are extreme with respect to the rest of the data.

A distinction between this work and most theoretical robustness studies is that the class of distributions over which we are trying to do well is of a different sort than those classes which are usually considered. For example, when Huber (1964) introduced M-estimates, he solved a minimax problem over the set of distributions

$$F_{\epsilon} = \{F = (1 - \epsilon)G + \epsilon H\}$$

where ϵ and G are fixed and G and H are symmetric distributions with the same center. The estimates discussed here are constructed to do well when the data are from one of two distributions. If, however, the two distributions are sufficiently different, we may find that we have constructed an estimator which performs well over a class of distributions that is much larger than the one over which we demanded good performance. Moreover, the procedures presented here can be generalized easily for optimization over any finite set of distributions. An obvious compromise of this sort would be one among the distributions, Gaussian, slash, and one-wild, that compose Tukey's three corners or among any other suitable set of corners. (Hampel has recently proposed some alternative corners.)

Section 2 includes some background information on M-estimates. The class of psi-compromised M-estimates is derived in Section 3 and the class of gap-compromised estimates is derived in Section 4. These two classes are compared in Sections 5, 6, and 7 via a Monte Carlo. Section 8 indicates the directions in which work on this topic is presently continuing.

2. BACKGROUND

The influence curve (also known as the influence function) of an estimator T at a distribution F was introduced by Hampel (1974) and is given by

$$IC_{T,F}(x) = \lim_{\epsilon \downarrow 0} \frac{T((1-\epsilon)F + \epsilon\alpha_x) - T(F)}{\epsilon}$$

where

$$\alpha_x(y) = \begin{cases} 0 & y < x \\ 1 & y \geq x \end{cases}$$

and $T(F)$ denotes the (population) value of the estimator at the distribution F . Thus $IC_{T,F}(x)$ can be thought of as describing the effect that a new observation at x has on the estimate $T(F)$. In his discussion, Hampel gave a heuristic demonstration that boundedness and continuity of the influence curve are two properties that are necessary to ensure the robustness of an estimator. By studying the influence curve, one is able to get a good feeling as to how an estimator will perform in a variety of sampling situations.

An estimate T_n is said to be an M-estimate of location if

$$\text{it satisfies } \sum_{i=1}^n \psi\left(\frac{x_i - T_n}{s_n}\right) = 0 \text{ for some specified function } \psi(\cdot)$$

and some ancillary s_n which is consistent under F for a scale

parameter, δ . Any function which, like ψ above, is reasonably thought of as defining an M-estimator will be called a *psi-function*. (Due to our concentration on symmetric random variables, all psi-functions used here will be assumed to be odd.) Let $T(F)$ satisfy

$$\int \psi\left(\frac{x-T(F)}{\delta}\right) dF = 0. \text{ Then it can be shown that } IC_{T,F}(x) \text{ is propor-}$$

tional to $\psi(x/\delta)$ and that under regularity conditions $\sqrt{n}(T_n - T(F))$ is asymptotically normally distributed with mean zero and variance

$$V(\psi, F) = \delta^2 \frac{\int \psi^2(x/\delta) dF(x)}{\left[\int \psi'(x/\delta) dF(x) \right]^2}.$$

This last fact will be used in the next section to derive the class of psi-compromised M-estimates.

3. PSI-COMPROMISED M-ESTIMATES

The following paragraph is intended to *motivate* the class of estimators which is derived below. It makes no optimality claims.

Suppose that we are willing to believe that our data come from one of two distributions, F_1 or F_2 . The plot of the asymptotic variance under F_2 against the asymptotic variance under F_1 for all M-estimates gives a region like the shaded one in Figure 3.1. We find that this region has a convex lower boundary with A and B being given by the asymptotic variances of the maximum likelihood estimates (MLE's) under F_1 and F_2 , respectively. Each point on the boundary corresponds to some M-estimate and therefore to some psi-function. If we are only considering F_1 and F_2 and we have a lot of data, we are wise to use one of these boundary psi-functions to construct an M-estimate. If we have only a

Figure 3.1 - Plot of the asymptotic variance under F_2 against the asymptotic variance under F_1 for all² M-estimates. Points outside of shaded region are unattainable by M-estimates.

moderate amount of data, we cannot claim any optimality for an M-estimate calculated with a boundary psi-function, but it is not unlikely that such an estimator would perform well.

It is easily seen that the set of boundary psi-functions is composed of the solutions to a set of minimization problems. In particular, the psi-function corresponding to a boundary tangent with slope $-\lambda/(1-\lambda)$, for λ between zero and one, is the solution of

$$H(\psi, \lambda) = \lambda V(\psi, F_1) + (1 - \lambda) V(\psi, F_2) = \min!$$

(F_1 and F_2 are scaled so that s_n asymptotically estimates the same value under either distribution.) A simple variational argument which is presented in Appendix A shows that $H(\psi, \lambda)$ is minimized by

$$\psi_\lambda(x) = - \frac{\lambda f'_1(x) + (((1-\lambda)f'_2(x))/u_1)}{\lambda f_1(x) + (((1-\lambda)f_2(x))/u_2)}$$

where

$$u_1 = (a_2/a_1)^3 (b_1/b_2), \quad u_2 = (a_2/a_1)^2 \quad (1)$$

and

$$a_i = \int \psi_\lambda(x) f_i'(x) dx$$

$$i = 1, 2.$$

$$b_i = \int \psi_\lambda^2(x) f_i(x) dx$$

Let $U(\phi, F) = \log V(\phi, F)$ where ϕ is a psi-function. An alternative to the above minimization is to consider the function

$$L(\phi, \gamma) = \gamma U(\phi, F_1) + (1 - \gamma) U(\phi, F_2)$$

seeking its minimum over ϕ . The solution to this problem is given by

$$\phi_\gamma(x) = - \frac{\gamma f_1'(x) + \left[((1-\gamma) f_2'(x)) / r_1 \right]}{\gamma f_1(x) + \left[((1-\gamma) f_2(x)) / r_2 \right]}$$

where

$$r_1 = (c_2/c_1), \quad r_2 = (d_2/d_1)$$

and

$$c_i = \int \phi_\gamma(x) f_i'(x) dx$$

$$i = 1, 2,$$

$$d_i = \int \phi_\gamma^2(x) f_i(x) dx$$

a form closely resembling (1).

It is reassuring to know that the sets $\Psi = \{\psi_\lambda(\cdot): 0 \leq \lambda \leq 1\}$ and $\Phi = \{\phi_\gamma(\cdot): 0 \leq \gamma \leq 1\}$ are in one to one correspondence. The proof of this fact and some discussion of the relationship of ψ_λ and ϕ_γ are presented in Appendix B in order not to disturb the continuity of the present discussion. The first set of simulations, those for the Gaussian-logistic compromise, was based on compromises of the log variances. All other compromises use variances. However, results of compromising variances for the Gaussian-logistic situation can be approximated

using the results of Appendix B.

The method of determining u_1 and u_2 (similarly r_1 and r_2) was to fix λ and initial values $u_1^{(0)}$ and $u_2^{(0)}$ and to iterate until convergence using the scheme suggested by (1). All numerical integrations were carried out using adaptive cautious Romberg extrapolation [2]. However, since the equations defined in (1) could have more than one fixed point, it is prudent to check that our iteration has indeed led us to a local minimum of our objective function. For example, the contour plot of $L(\phi_{.1}, .1)$, shown in Figure 3.2, agrees well with the result that for the Gaussian-logistic compromise, $L(\phi_{.1}, .1)$ is minimized at $(r_1, r_2) = (.98075, 1.04362)$.

The form of $\psi_\lambda(x)$ allows us to make some simple statements about the behavior of the influence curve, IC_λ , of the corresponding M-estimator for large $|x|$. We make the following

Definition: Let G and F be two distributions which are scaled so that for some specified $\alpha \in (0,1)$, $G^{-1}(\alpha) = F^{-1}(\alpha)$.

G will be said to be long-tailed relative to F ($G \gg F$) if

$$\lim_{|x| \rightarrow \infty} \{f(x)/g(x)\} = 0.$$

Then, IC_λ will behave for large $|x|$ as the influence curve of the more long-tailed of F_1 and F_2 . To see this, suppose $F_2 \gg F_1$. Write

$$\psi_\lambda(x) = - \frac{\lambda f_1'(x)}{\lambda f_1(x) + \left[((1-\lambda) f_2(x)) / u_2 \right]} - \frac{(1-\lambda) f_2'(x) / u_1}{f_1(x) + \left[((1-\lambda) f_2(x)) / u_2 \right]},$$

Gaussian-Logistic $\lambda = .1$

Figure 3.2 - Contour plot of $L(\phi_1, .1)$ as a function of r_1 and r_2 . The values of the vertices of these polygons have the indicated values. They have been connected by line segments rather than curves for convenience.

multiply by $f_2(x)/f_2'(x)$, and take limits as $|x| \rightarrow \infty$. (All necessary limits exist when, as would usually be the case, f_1 and f_2 approach zero monotonically as $|x| \rightarrow \infty$.) Thus the procedure protects us from the F_i with the longer tail by choosing a psi-function which, for large $|x|$, is optimal for that "worse distribution."

4. GAP-COMPROMISED ESTIMATES

4.1 The Gap Form of an M-estimate

Let f be a density with unknown location, θ , and scale,

δ . Then

$$\begin{aligned} \frac{\partial}{\partial \theta} \log \prod_{i=1}^n f((y_i - \theta)/\delta) &= \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f((y_i - \theta)/\delta) \\ &= \sum_{i=1}^n \text{sco}((y_i - \theta)/\delta) \\ &= S(y, \theta, \delta) . \end{aligned}$$

Define $T_n^*(y)$ as the principal zero of $S(y, \cdot, s_n)$ where under F , s_n is a consistent estimate of δ . Also, define a weight function

$$w((y-T)/s) = \frac{\text{sco}((y-T)/s)}{((y-T)/s)} .$$

Thus

$$0 = S(y, T_n^*, s_n) = \sum_i w((y_i - T_n^*)/s_n) (y_i - T_n^*) . \quad (2)$$

Rearrangement of (2) gives

$$T_n^* = \frac{\sum_i y_i w((y_i - T_n^*)/s_n)}{\sum_i w((y_i - T_n^*)/s_n)} , \quad (3)$$

the familiar w-iteration form of the MLE of θ .

(N.B. It is not necessary or even advisable to restrict ourselves to maximum likelihood estimation. Some of the best estimators that we have (e.g., bisquare) do not correspond to MLE's for any distribution. The discussion here is given in terms of MLE's rather than general M-estimates because we will later be concerned with the specific form of the weight function that corresponds to maximum likelihood estimation.)

Consider the case in which we are dealing with a sample of size n . Write $y(i)$ for the ordered values and denote the sample by:

$$y(-[n/2]) < \dots < y(-2) < y(-1) < y(0) < y(1) < y(2) < \dots < y([n/2]) .$$

If n is even, the sample under consideration is the original sample augmented by the sample median $y(0) = (y(-1) + y(1))/2$. Now define gaps by

$$\begin{aligned} g_j &= y(j) - y(j-1) & j &= 1, \dots, [n/2] \\ g_{-j} &= y(-(j-1)) - y(-j) & j &= 1, \dots, [n/2] \\ g_0 &= 0 . \end{aligned}$$

Thus, $g_j \geq 0$ for all j and

$$y(i) = \begin{cases} \sum_{j=1}^i g_j + y(0) & i = 1, \dots, [n/2] \\ -\left[\sum_{j=1}^{-i} g_j \right] + y(0) & i = -[n/2], \dots, -1 \end{cases} \quad (4)$$

Substituting (4) into (3) shows that

$$T_n^* = \sum_{j=-[n/2]}^{[n/2]} g_j (V_j/V_+) + y(0) \quad (5)$$

where

$$V_j = \begin{cases} w\left(\frac{y(j)-T_n^*}{s_n}\right) + \dots + w\left(\frac{y([n/2])-T_n^*}{s_n}\right), & j=1, \dots, [n/2] \\ -\left[w\left(\frac{y(j)-T_n^*}{s_n}\right) + \dots + w\left(\frac{y(-[n/2])-T_n^*}{s_n}\right)\right], & j=-[n/2], \dots, -1 \end{cases}$$

$$V_0 = 0$$

$$V_+ = \sum_i w\left(\frac{y(i)-T_n^*}{s_n}\right).$$

The estimate in (5) is the MLE from the w -iteration written in gap form.

4.2. Some Heuristics and the Compromise Procedure

In this section, we will go through some heuristics that will help to motivate the actual procedure that was used to form gap-compromised estimates.

Consider an estimate of the form

$$T(y) = \sum_i g_i v_i + y(0) \quad (6)$$

where the g_i are as in the preceding section and the v_i are some fixed set of suitable gap-weights. Assuming that n is large enough that $\text{cov}_F(g_i, g_j)$ is small for all $i \neq j$, we can write

$$\text{var}_F(T) \approx \sum_i v_i^2 \text{var}_F(g_i) + \text{var}_F(y(0)) + 2 \sum_i v_i \text{cov}_F(g_i, y(0)). \quad (7)$$

Suppose that (7) is minimized when $v_i = v_i^F$. Then it is easy to show that $\text{cov}_F(g_i, y(0)) = -v_i^F \text{var}_F(g_i)$ and thus the variance of the estimate T in (6) is

$$\text{var}_F(T) \cong \sum v_i^2 \text{var}_F(g_i) + \text{var}_F(y(0)) - 2 \sum v_i v_i^F \text{var}_F(g_i) .$$

Suppose now that there are two disparate specifications, F_1 and F_2 , to be compromised, the first with optimal gap-weights $v_i^{F_1}$ and $A_i^{F_1} = \text{var}_{F_1}(g_i)$ and the second with optimal gap-weights $v_i^{F_2}$ and $A_i^{F_2} = \text{var}_{F_2}(g_i)$. We would like to construct a set of gap-weights which is in some sense optimal as a compromise between those which are optimal at F_1 and those which are optimal at F_2 .

Consider the function

$$K(v, \lambda) = \lambda \text{var}_{F_1}(T) + (1-\lambda) \text{var}_{F_2}(T) .$$

Differentiation shows that K is minimized when

$$v_i = \left[\lambda A_i^{F_1} v_i^{F_1} + (1-\lambda) A_i^{F_2} v_i^{F_2} \right] / \left[\lambda A_i^{F_1} + (1-\lambda) A_i^{F_2} \right] . \quad (8)$$

The form of (8) tells us that if the variance of a gap is especially large under one distribution relative to its variance under the other, then that gap should be weighted as if the data were from the distribution for which the gap has high variance. Thus (8) offers some protection against the dangerous situation of mis-weighting a potentially highly variable gap.

Now, we are not going to actually use $v_i^F = \frac{-\text{cov}_F(g_i, y(0))}{\text{var}_F(g_i)}$

in forming our compromise gap-weights. Rather, we will use weights that have been tailored to the data sample that we have by using in place of the v_i^F the gap-weights that are derived from the w-iteration form of MLE. That is, we will replace the v_i^F in (8) by the $(v_i/v_+)^F$ shown in (5) and use the resulting v_i weights to form an estimate of the form given by (6). Such an estimate will be referred to as a gap-compromised estimate.

5. SIMULATIONS

The simulations presented in the following sections were carried out on the PDP 11/40 mini-computer in the Department of Statistics, Princeton University. All calculations were done in double precision (56-bit mantissa). Uniform random numbers were generated using Knuth's Algorithm M [6] with the two uniform sequences being congruentially generated. Gaussian samples were generated using an algorithm due to Forsythe and Ahrens-Dieter [3]. Finally, logistic samples were generated by using the logistic generating function and Cauchy variates were generated as the ratio of two Gaussians.

For each sample, the median was used as the initial estimate of location. Scale estimation was by the median absolute deviation (MAD)

$$s_n = \operatorname{median}_i |x_i - T_n|.$$

Consequently, all distributions were scaled to have quartiles at ± 1.0 . For both the psi-compromised M-estimates and the gap-compromised estimates, the convergence criterion in the definition of T_n was that the difference between two successive iterates be less than 5×10^{-4} or that 20 iterations had been performed.

6. GAUSSIAN-LOGISTIC COMPROMISE

Let F_1 denote the Gaussian distribution with density

$$f_1(x) = (.6745/\sqrt{2\pi}) \exp\{-(1/2)(.6745x)^2\} \quad (9)$$

and let F_2 denote the logistic distribution with density

$$f_2(x) = \frac{(\ln 3)\exp(-x \ln 3)}{(1 + \exp(-x \ln 3))^2}.$$

Table 6.1 shows the values of r_1 and r_2 that were used in the definitions of the ϕ_γ .

	<u>r_1</u>	<u>r_2</u>
.1	.98075	1.04362
.2	.97962	1.05000
.3	.97868	1.05728
.4	.97798	1.06569
$\gamma =$.5	.97759	1.07555
.6	.97761	1.08736
.7	.97823	1.10197
.8	.97983	1.12105
.9	.98338	1.14898

Table 6.1

Plots of $\phi_\gamma(x)$ for $x > 0$ are presented in Figure 6.1(a-k).

As mentioned in Section 2, the influence curve of an M-estimate is proportional to this psi-function and we need only rescale each plot by the appropriate constant in order to retrieve the influence curve. (The constant will depend on which distribution we are considering.)

Figure 6.1 - Plots of the Gaussian-logistic compromise psi-function, $\phi_{\gamma}(x)$, for various values of γ and for $x > 0$. Values of the function for $x < 0$ can be obtained with $\phi_{\gamma}(-x) = -\phi_{\gamma}(x)$.

6.1(a) $\gamma = 0.0$

6.1(b) $\gamma = 0.1$

6.1(c) $\gamma = 0.2$

Figure 6.1 (Continued)

Figure 6.1 (Continued)

6.1(g) $\gamma = 0.6$

6.1(h) $\gamma = 0.7$

6.1(i) $\gamma = 0.8$

Figure 6.1 (Continued)

6.1(j) $\gamma = 0.9$

6.1(k) $\gamma = 1.0$

Tükey [9] has suggested a finite sample approximation to the influence curve which has come to be known as the sensitivity curve. In particular, Figure 6.2(a-k) are the *stylized* sensitivity curves [1] for the Gaussian-logistic psi-compromised M-estimates based on a sample of size 10. These curves were generated by

- (i) starting with a pseudo-sample consisting of the 9 expected Gaussian order statistics from a sample of 9,
- (ii) augmenting this sample by a changing point x ,
- (iii) evaluating the estimate, denoted by $T(x)$, for this augmented sample, and
- (iv) plotting $nT(x)$ as a function of x .

The similarity of shapes between Figures 6.1 and 6.2 is striking.

Table 6.2 gives values of $A_i^{F_1}$ and $A_i^{F_2}$ for a sample of size 10 ($A_{-i}^F = A_i^F$).

<u>Gap Index</u> (median)	<u>Gaussian</u> <u>Variance</u>	<u>Logistic</u> <u>Variance</u>
1	.02803	.00283
2	.11869	.12295
3	.14353	.16190
4	.21167	.28319
5	.47522	.92456
(extreme)		

Table 6.2

At present, the influence curve for the gap-compromised estimator is still being calculated. Stylized sensitivity curves for the Gaussian-logistic gap-compromised estimates based on a sample of

Figure 6.2 - Stylized sensitivity curves for the Gaussian-logistic psi-compromised M-estimates for sample size 10. The background observations are the expected values of the order statistics from a Gaussian sample of size 9.

6.2(a) $\gamma = 0.0$

6.2(b) $\gamma = 0.1$

6.2(c) $\gamma = 0.2$

Figure 6.2 (Continued)

6.2(d) $\gamma = 0.3$

6.2(e) $\gamma = 0.4$

6.2(f) $\gamma = 0.5$

Figure 6.2 (Continued)

6.2(g) $\gamma = 0.6$

6.2(h) $\gamma = 0.7$

6.2(i) $\gamma = 0.8$

Figure 6.2 (Continued)

6.2(j) $\gamma = 0.9$

6.2(k) $\gamma = 1.0$

size 10 are shown in Figure 6.3(a-1).

The compromised estimates were computed for 1,000 samples of size 10 from the Gaussian distribution and for 1,000 samples of size 10 from the logistic distribution. Variances of the estimates under the Gaussian distribution were computed using the location-scale swindle described in [8] to improve the numerical accuracy of the results. A swindle was not possible with the logistic samples because as far as the author knows, the logistic probability law is not easily expressed as a ratio Z/Y with Z distributed symmetrically about zero and having a sufficient statistic for location and Y a positive random variable independent of Z (see [7]). Consequently, it will be necessary to use a larger number of samples in order to calculate the variances of the estimates under the logistic distribution to the desired precision. The current variance estimates and their estimated standard errors are given in Appendix C.

Define the efficiency of an estimator T under a distribution F as the ratio of the variance of the MLE_F under F to the variance of T under F and estimate it by:

$$eff_{F_n}(T) = \frac{\text{var}_F(MLE_F)}{\text{var}_{F_n}(T)}.$$

We shall refer to this quantity as the Gaussian efficiency if F is the Gaussian distribution, as the logistic efficiency if F is the logistic distribution, and so forth.

Figure 6.3 - Stylized sensitivity curves for the Gaussian-logistic gap-compromised estimates for sample size 10. The background observations are the expected values of the order statistics from a Gaussian sample of size 9.

6.3(a) $\lambda = 0.0$

6.3(b) $\lambda = 0.1$

6.3(c) $\lambda = 0.2$

Figure 6.3 (Continued)

6.3(d) $\lambda = 0.3$

6.3(e) $\lambda = 0.4$

6.3(f) $\lambda = 0.5$

Figure 6.3 (Continued)

6.3(g) $\lambda = 0.6$

6.3(h) $\lambda = 0.7$

6.3(i) $\lambda = 0.8$

Figure 6.3 (Continued)

6.3(j) $\lambda = 0.9$

6.3(k) $\lambda = 1.0$

6.3(l) Plots 6.3(a) - 6.3(k) superimposed.

Figure 6.4(a) is the plot of (logistic efficiency)⁻¹ against (Gaussian efficiency)⁻¹ for the psi-compromised M-estimates (plotting character P) and for the gap-compromised estimates (plotting character G) for samples of size 10. The G curve lies to the southwest of the P curve, indicating that for this compromise situation we have succeeded in constructing estimators which perform better than the "best" M-estimators. Figure 6.4(b) is the same plot for samples of size 20. The fact that the G curve no longer offers an improvement over the P curve, except at the extreme left, indicates the importance of investigating the dependence of these efficiency relationships on sample size. Thus, determination of the influence curves for gap-compromised estimates, which will allow us to construct this same plot in terms of asymptotic efficiencies, while useful for elucidating the limiting behavior of these relationships, is not sufficient for answering the practical questions concerning finite sample sizes.

7. CAUCHY-GAUSSIAN COMPROMISE

This section presents the results of compromising (9) and the Cauchy distribution with density

$$f_2(x) = 1/(\pi(1 + x^2)) .$$

Table 7.1 shows the values of u_1 and u_2 that were used in the definitions of the compromise psi-functions, ψ_λ .

Figure 6.4(a) - Plot of $(\text{logistic efficiency})^{-1}$ against $(\text{Gaussian efficiency})^{-1}$ for the psi-compromised M-estimates (plotting character P) and for the gap-compromised estimates (plotting character G) for samples of size 10.

Figure 6.4(b) - Plot of $(\text{logistic efficiency})^{-1}$ against $(\text{Gaussian efficiency})^{-1}$ for the psi-compromised M-estimates (plotting character P) and for the gap-compromised estimates (plotting character G) for samples of size 20.

	<u>u_1</u>	<u>u_2</u>
.1	1.68327	1.20178
.2	1.47064	1.09747
.3	1.31260	1.01921
.4	1.18221	.95448
$\lambda =$.5	1.06695	.89742
.6	.95927	.84457
.7	.85324	.79342
.8	.74215	.74161
.9	.61384	.68676

Table 7.1

Figure 7.1(a-k) are plots of the corresponding $\psi_\lambda(\cdot)$ functions. As was discussed in Section 3, the behavior of these functions for large $|x|$ is seen to be like x^{-1} , the optimal shape for Cauchy data. As in the Gaussian-logistic compromise, the resemblance of the stylized sensitivity curves, which are not shown, to these psi-functions is very close.

Table 7.2 gives values of the gap variances for samples of size 10 under the Cauchy distribution.

<u>Gap Index</u> (median)	<u>Cauchy Variance</u>
1	.0376
2	.2562
3	1.3206
4	∞
5	∞
(extreme)	

Table 7.2

Figure 7.1 - Plots of the Cauchy-Gaussian compromise psi-function, $\psi_{\lambda}(x)$, for various values of λ and for $x > 0$. Values of the function for $x < 0$ can be obtained with $\psi_{\lambda}(-x) = -\psi_{\lambda}(x)$.

7.1(a) $\lambda = 0.0$

7.1(b) $\lambda = 0.1$

7.1(c) $\lambda = 0.2$

Figure 7.1 (Continued)

7.1(d) $\lambda = 0.3$

7.1(e) $\lambda = 0.4$

7.1(f) $\lambda = 0.5$

Figure 7.1 (Continued)

7.1(g) $\lambda = 0.6$

7.1(h) $\lambda = 0.7$

7.1(i) $\lambda = 0.8$

Figure 7.1 (Continued)

7.1(j) $\lambda = 0.9$

7.1(k) $\lambda = 1.0$

Finally, Figure 7.2(a-k) shows the stylized sensitivity curves for the Cauchy-Gaussian gap-compromised estimates based on sample size 10.

As with the Gaussian-logistic compromise, 1,000 samples of size 10 were generated from (a) the Gaussian distribution (the same 1,000 samples that were used in the Gaussian-logistic simulations) and (b) the Cauchy distribution. All variances of the estimates were computed using the swindle and are presented, along with their standard errors, in Appendix C. A plot of $(\text{Cauchy efficiency})^{-1}$ against $(\text{Gaussian efficiency})^{-1}$ for the two types of estimates is given in Figure 7.3. We see that for values of λ that are close to zero (more weight on the Cauchy distribution), the two estimators seem to be equivalent. However, as λ increases toward one, the psi-compromised M-estimates outperform the gap-compromised estimates. This loss of Gaussian efficiency for the gap-compromised estimates is due to the fact that the infinite Cauchy variances of the last two gaps cause these gaps to receive very low weights, thus giving up too much information in (small) Gaussian samples. Attempts to rectify this problem by treating these last two Cauchy variances as large, but finite have failed in the sense that the Cauchy variances of the estimates for λ close to one become unacceptably large.

Figure 7.2 - Stylized sensitivity curves for the Cauchy-Gaussian gap-compromised estimates for sample size 10. The background observations are the expected values of the order statistics from a Gaussian sample of size 9.

7.2(a) $\lambda = 0.0$

7.2(b) $\lambda = 0.1$

7.2(c) $\lambda = 0.2$

Figure 7.2 (Continued)

7.2(d) $\lambda = 0.3$

7.2(e) $\lambda = 0.4$

7.2(f) $\lambda = 0.5$

Figure 7.2 (Continued)

7.2(g) $\lambda = 0.6$

7.2(h) $\lambda = 0.7$

7.2(i) $\lambda = 0.8$

Figure 7.2 (Continued)

7.2(j) $\lambda = 0.9$

7.2(k) $\lambda = 1.0$

CAUCHY-GAUSSIAN

Figure 7.3 - Plot of $(\text{Cauchy efficiency})^{-1}$ against $(\text{Gaussian efficiency})^{-1}$ for the psi-compromised M-estimates (plotting character P) and for the gap-compromised estimates (plotting character G) for samples of size 10.

8. ONGOING INVESTIGATIONS

Work that is currently in progress consists of the following:

- (a) Calculation of the influence curve for gap-compromised estimates:

As mentioned in Section 6, this will allow us to compare the estimates in terms of asymptotic efficiencies and will make the limiting dependence on sample size of our results more clear.

- (b) Computation of order statistic (gap) moments:

Unavailability of variances and covariances of order statistics is the one factor which prevents us from calculating gap-compromised estimates for a variety of compromise situations. Work by A. Bruce, D. Pregibon and J. W. Tukey on exactly this problem should help alleviate this difficulty.

- (c) Computing gap-compromised estimates based on a new form of the $v(i)$:

This point relates to (b) above in that the new form of $v(i)$ does not involve gap variances. If the estimators based on these new compromise gap weights perform well relative to the other two types of estimators, we may be able to circumvent the difficulty of (b).

- (d) Confidence bounds on $(\text{efficiency})^{-1}$:

In order to assess the strength of the indications in Figures 6.4 and 7.2 (particularly 6.4a), it is desirable to calculate confidence bounds for the G and P curves.

(e) The three corners:

Calculation of the boundary of the M-estimate region (see Figure 3.1) based on psi-compromises of the three curves is now underway. The three corners are:

- i) the standard Gaussian distribution;
- ii) the slash distribution;
- iii) the (.95, .05) mixture of $G(0,1)$ with $G(0,100)$.

APPENDIX A

We seek

$$\psi : \int_{-\infty}^{\infty} \frac{\lambda \psi^2 dF_1}{\left(\int \psi f_1' \right)^2} + \frac{(1-\lambda) \psi^2 dF_2}{\left(\int \psi f_2' \right)^2} = \min \quad (\text{A.1})$$

with $\psi f_i \rightarrow 0$ as $|x| \rightarrow \infty$ and ψ differentiable.

Consider a one-parameter family of admissible functions which includes $\psi(x)$, namely, the set of all functions of the form

$$\psi(x) + \epsilon \eta(x)$$

where $\eta(x)$ is any arbitrarily chosen differentiable function with $\eta f_i \rightarrow 0$ as $|x| \rightarrow \infty$.

Let

$$I(\epsilon) = \int_{-\infty}^{\infty} (\psi + \epsilon \eta)^2 \left\{ (\lambda f_1) / \left(\int (\psi + \epsilon \eta) f_1' \right)^2 + ((1-\lambda) f_2) / \left(\int (\psi + \epsilon \eta) f_2' \right)^2 \right\} dx.$$

It then follows that $I(\epsilon)$ takes on its minimum value when $\epsilon = 0$, that is, when the variation of ψ is zero. Hence, it must follow that

$$\left. \frac{dI(\epsilon)}{d\epsilon} \right|_{\epsilon=0} = 0$$

$$\frac{dI(\epsilon)}{d\epsilon} = \lim_{\theta \rightarrow 0} \frac{I(\epsilon + \theta) - I(\epsilon)}{\theta}$$

$$= \lim_{\theta \rightarrow 0} \left[\frac{\int ((\psi + \epsilon \eta) + \theta \eta)^2 \lambda f_1}{\left(\int ((\psi + \epsilon \eta) + \theta \eta) f_1' \right)^2} + \frac{\int ((\psi + \epsilon \eta) + \theta \eta)^2 (1-\lambda) f_2}{\left(\int ((\psi + \epsilon \eta) + \theta \eta) f_2' \right)^2} \right. \\ \left. - \frac{\int (\psi + \epsilon \eta)^2 \lambda f_1}{\left(\int (\psi + \epsilon \eta) f_1' \right)^2} - \frac{\int (\psi + \epsilon \eta)^2 (1-\lambda) f_2}{\left(\int (\psi + \epsilon \eta) f_2' \right)^2} \right] / \theta.$$

$$\text{Let } b_i = \int (\psi + \varepsilon \eta)^2 f_i \quad \text{and} \quad a_i = \int (\psi + \varepsilon \eta) f'_i ,$$

$$\Rightarrow \frac{dI(\varepsilon)}{d\varepsilon} = \lim_{\theta \rightarrow 0} \left[\frac{\lambda b_1 + 2\lambda\theta \int (\psi + \varepsilon \eta) \eta f_1 + \theta^2 \lambda \int \eta^2 f_1}{a_1^2 + 2\theta \left(\int (\psi + \varepsilon \eta) f'_1 \right) \left(\int \eta f'_1 \right) + \theta^2 \left(\int \eta f'_1 \right)^2} \right. \\ \left. + \frac{(1-\lambda)b_2 + 2(1-\lambda)\theta \int (\psi + \varepsilon \eta) \eta f_2 + \theta^2 (1-\lambda) \int \eta^2 f_2}{a_2^2 + 2\theta \left(\int (\psi + \varepsilon \eta) f'_2 \right) \left(\int \eta f'_2 \right) + \theta^2 \left(\int \eta f'_2 \right)^2} \right. \\ \left. - \frac{\lambda b_1}{a_1^2} - \frac{(1-\lambda)b_2}{a_2^2} \right] / \theta .$$

Collect all terms involving a_1 , b_1 , and f_1 over a common denominator and similarly collect all terms involving a_2 , b_2 , and f_2 .

$$\Rightarrow \frac{dI(\varepsilon)}{d\varepsilon} = \frac{2\lambda a_1^2 \int (\psi + \varepsilon \eta) \eta f_1 - 2\lambda b_1 a_1 \int \eta f'_1}{a_1^4} + \frac{2(1-\lambda) a_2^2 \int (\psi + \varepsilon \eta) \eta f_2 - 2(1-\lambda) b_2 a_2 \int \eta f'_2}{a_2^4}$$

$$\left. \frac{dI(\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0} = 0$$

$$\Rightarrow 0 = \int \left[\frac{\psi \lambda \tilde{a}_1^2 \eta f_1 - \lambda \tilde{b}_1 \tilde{a}_1 \eta f'_1}{\tilde{a}_1^4} + \frac{\psi (1-\lambda) \tilde{a}_2^2 \eta f_2 - (1-\lambda) \tilde{b}_2 \tilde{a}_2 \eta f'_2}{\tilde{a}_2^4} \right] \\ = \int \left[\frac{\lambda \psi f_1}{\tilde{a}_1^2} - \frac{\lambda \tilde{b}_1 f'_1}{\tilde{a}_1^3} + \frac{\psi (1-\lambda) f_2}{\tilde{a}_2^2} - \frac{(1-\lambda) \tilde{b}_2 f'_2}{\tilde{a}_2^3} \right] \eta ,$$

where

$$\tilde{a}_i = \int \psi f'_i \\ \tilde{b}_i = \int \psi^2 f_i .$$

Since the above equation is true for any function η which is differentiable and for which $\eta f_i \rightarrow 0$ as $|x| \rightarrow \infty$, the coefficient of $\eta(x)$ in the integrand must be zero everywhere in $(-\infty, \infty)$ so the condition

$$\frac{\lambda \psi f_1}{\tilde{a}_1^2} - \frac{\lambda \tilde{b}_1 f_1'}{\tilde{a}_1^3} + \frac{\psi(1-\lambda)f_2}{\tilde{a}_2^2} - \frac{(1-\lambda)\tilde{b}_2 f_2'}{\tilde{a}_2^3} = 0$$

must be satisfied

$$\Rightarrow \psi(x) = \left[\frac{\lambda f_1'(x) \tilde{b}_1}{\tilde{a}_1^3} + \frac{((1-\lambda)f_2'(x) \tilde{b}_2)}{\tilde{a}_2^3} \right] / \left[\frac{\lambda f_1(x)}{\tilde{a}_1^2} + \frac{((1-\lambda)f_2(x))}{\tilde{a}_2^2} \right]$$

Since (A.1) defines ψ only up to a multiplicative constant, take

$$\psi_\lambda(x) = (\tilde{a}_1/\tilde{b}_1)\psi(x) .$$

APPENDIX B

Let $\Psi = \{\psi_\lambda(\cdot): 0 \leq \lambda \leq 1\}$ and $\Phi = \{\phi_\gamma(\cdot): 0 \leq \gamma \leq 1\}$ where $\psi_\lambda(\cdot)$ and $\phi_\gamma(\cdot)$ are as in Section 3.

Fact: Let $\lambda \in [0,1]$ be given and let $\psi_\lambda(\cdot)$, u_1 , and u_2 be as in Section 3. Then taking

$$\gamma = \lambda / (\lambda(1-p) + p)$$

where $p = \sqrt{u_2}/u_1 = (V(\psi_\lambda, F_2)) / (V(\psi_\lambda, F_1))$ assures that $\gamma \in [0,1]$.

Furthermore,

$$\phi(x, \gamma) = \left[\gamma f'_1(x) + \frac{(1-\gamma)f'_2(x)}{\sqrt{u_2}} \right] / \left[\gamma f_1(x) + \frac{(1-\gamma)f_2(x)u_1}{u_2^{3/2}} \right]$$

is an element of Φ and is identified with $\psi_\lambda(\cdot)$.

Proof: The result is trivial for $\lambda = 0$ and $\lambda = 1$. That $\gamma \in (0,1)$ for $\lambda \in (0,1)$ is easily verified using the fact that $p > 0$.

Also:

$$\phi(x, \gamma) = \left[\lambda f'_1(x) + \frac{(1-\lambda)f'_2(x)}{u_1} \right] / \left[\lambda f_1(x) + \frac{(1-\lambda)f_2(x)}{u_2} \right] = \psi_\lambda(x).$$

Finally,

$$\begin{aligned} u_2^{1/2} &= \left[\int \psi_\lambda(x) f'_2(x) dx \right] / \left[\int \psi_\lambda(x) f'_1(x) dx \right] \\ &= \left[\int \phi(x, \gamma) f'_2(x) dx \right] / \left[\int \phi(x, \gamma) f'_1(x) dx \right]; \\ \frac{u_2^{3/2}}{u_1} &= \left[\int \psi_\lambda^2(x) f_2(x) dx \right] / \left[\int \psi_\lambda^2(x) f_1(x) dx \right] \\ &= \left[\int \phi^2(x, \gamma) f_2(x) dx \right] / \left[\int \phi^2(x, \gamma) f_1(x) dx \right], \end{aligned}$$

establishing that $\phi(\cdot, \gamma)$ is in Φ .

To study the relationship between γ and λ a bit more thoroughly, we note that if p were not a function of λ , then $\gamma(\lambda)$ would be the equation of a hyperbola passing through $(0,0)$ and $(1,1)$ and having asymptotes $x_0 = 1/(1-(1/p))$ and $y_0 = 1/(1-p)$. Also $\left. \frac{d\gamma}{d\lambda} \right|_{\lambda=1} = p$, so that $p > 1$ would imply that $\gamma(\lambda)$ was represented by the dotted curve in Figure B.1 while $p < 1$ would mean that $\gamma(\lambda)$ was given by the solid curve in the same figure.

Figure B.1 - Two possibilities for the plot of $\gamma(\lambda)$. The dotted curve corresponds to $p > 1$ while the solid curve corresponds to $p < 1$.

Thus, if p is approximately constant as λ ranges between zero and one, $\gamma(\lambda)$ will be well approximated by a hyperbola. Such is the case for the Gaussian-logistic compromise (Section 6) where p ranges from 2.86 at $\lambda = 0$ to 3.29 at $\lambda = 1$, a relative change of 13%. Figure B.2 is a plot showing the relationship between γ and λ for the Gaussian-logistic compromise.

Figure B.2 - Plot of $\gamma(\lambda)$ for the Gaussian-logistic compromise.
Also shown is the line, $y = x$.