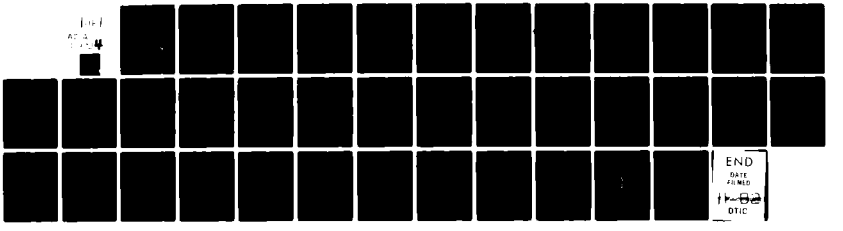
AD-A119 724 RENSSELAER POLYTECHNIC INST TROY NY F/G 12/1
SELF-CRITICAL AND ROBUST PROCEDURES FOR THE ANALYSIS OF UNIVARI--ETC(U)
JUN 82 A S PAULSON, M A PRESSER, E H NICKLIN DAAG29-81-K-0110
UNCLASSIFIED A-5 ARO-18072.5-MA NL



SELF-CRITICAL AND ROBUST PROCEDURES FOR THE ANALYSIS OF UNIVARIATE COMPLETE DATA

by

A. S. Paulson *
Rensselaer Polytechnic Institute

M. A. Presser
General Foods Corporation

and

E. H. Nicklin
Alex Brown and Sons

*Research supported in part by U.S. Army Research Office under contract DAA G29-81-K-0110.

Summary

A statistical sensitivity analysis may be defined and performed in terms of the response of a vector of parameter estimates to variation in the way sample information is processed vis-a-vis to tentative underlying model. The mode of information processing is a generalization of likelihood and is indexed on a non-statistical parameter c. The case c=0 corresponds to maximum likelihood. If the vector of parameter estimates is stable under moderate increase of the index c from 0, the tentative model and the data are internally consistent. A general procedure for the conduct of such sensitivity analyses is given along with several illustrations. For fixed, positive values of the index, one obtains a general robust estimation procedure.

Key Words: sensitivity analysis, robust procedure, Gaussian, logistic, Weibull, extreme value, Poisson, generalized likelihood

1. Introduction

Sensitivity analyses are routinely and usefully performed in the engineering disciplines, in operational research, to name just a few. The objective of a sensitivity analysis is to determine the response of a solution to changes in the assumptions, to changes in the data, or to changes in the information that are used in modeling representations of reality. Also within the purview of a sensitivity analysis is the determination of responses of a solution to changes in the way in which information is processed. If small or moderate perturbations in the assumptions, data, processing, etc. produce large changes in a solution, then valuable information has been provided for the analyst or the modeler since certain facets of the model or of the information are critical or require further attention.

Residual analysis, jackknife procedures, and robust procedures provide several ways in which a sensitivity analysis may be performed in a statistical setting. In this setting the data and a tentatively assumed model should be considered as a single entity. The objective of a sensitivity analysis is to determine whether the data and the tentatively assumed model are internally consistent and this may be effected by controlling the way information provided by the data is incorporated in the evaluation of model parameters. A sensitivity analysis can also be most useful in model evolution. We propose herein a general procedure for performing a sensitivity analysis of a data-model unit and, secondarily, a general procedure for producing robust estimators of location, scale, shape, etc. parameters.

There is by now an extensive literature on robust methods for location problems. Barnett and Lewis (1978), Huber (1981), and Rey (1977) provide useful summaries of currently available robust methods. There are few papers on construction of robust estimators for non-location parameters and asymmetric distributions. This paper provides such a construction. It is based on what is believed to be a new generalization of the likelihood or Shannon's information. The procedures we propose are easily implemented and are readily applied in a modeling or structured data framework. Several examples are provided. The Gaussian, Weibull, gamma, logistic, and Poisson distributions are explicitly considered.

2. Construction of a Family of Estimators

Let $x_1, x_2, \ldots, x_n$ be a random sample from the normal density

$$n(x;\mu,\sigma^2) = (2\pi\sigma^2)^{\frac{1}{2}} \exp(-\tfrac{1}{2}(\tfrac{x-\mu}{\sigma})^2). \tag{2.1}$$

The log likelihood for the location parameter μ, σ assumed momentarily to be known, is

$$L(\mu) = \sum_{j=1}^{n} \log f(x_j;\mu,\sigma^2) = -\frac{n}{2}\log 2\pi = -\frac{n}{2}\log\sigma^2 - \tfrac{1}{2}\sum_{j=1}^{n}\left(\frac{x_j-\mu}{\sigma}\right)^2. \tag{2.2}$$

The maximum likelihood estimator of μ is determined by minimizing the quadratic form

$$\tfrac{1}{2}\sum_{j=1}^{n}(x_j-\mu)^2 \tag{2.3}$$

with respect to μ. In a seminal paper Huber (1964) suggested that μ be estimated by replacing the quadratic in (2.3) by

$$\tfrac{1}{2} \sum_{j=1}^{n} \rho(x_j-\mu), \tag{2.4}$$

where $\rho(\cdot)$ is a convex function which increases less rapidly than a quadratic; in particular the choice

$$\rho(u) = \begin{cases} \tfrac{1}{2} u^2 & , \quad |u| \le \kappa \\ \kappa|u| - \tfrac{1}{2}\kappa^2, & |u| > \kappa \end{cases} \tag{2.5}$$

is optimal in a minimax sense. Thus (2.4) represents a change (from (2.3)) in the way the information provided by the x_j for the model $n(x;\mu,\sigma^2)$ of (2.1) is processed.

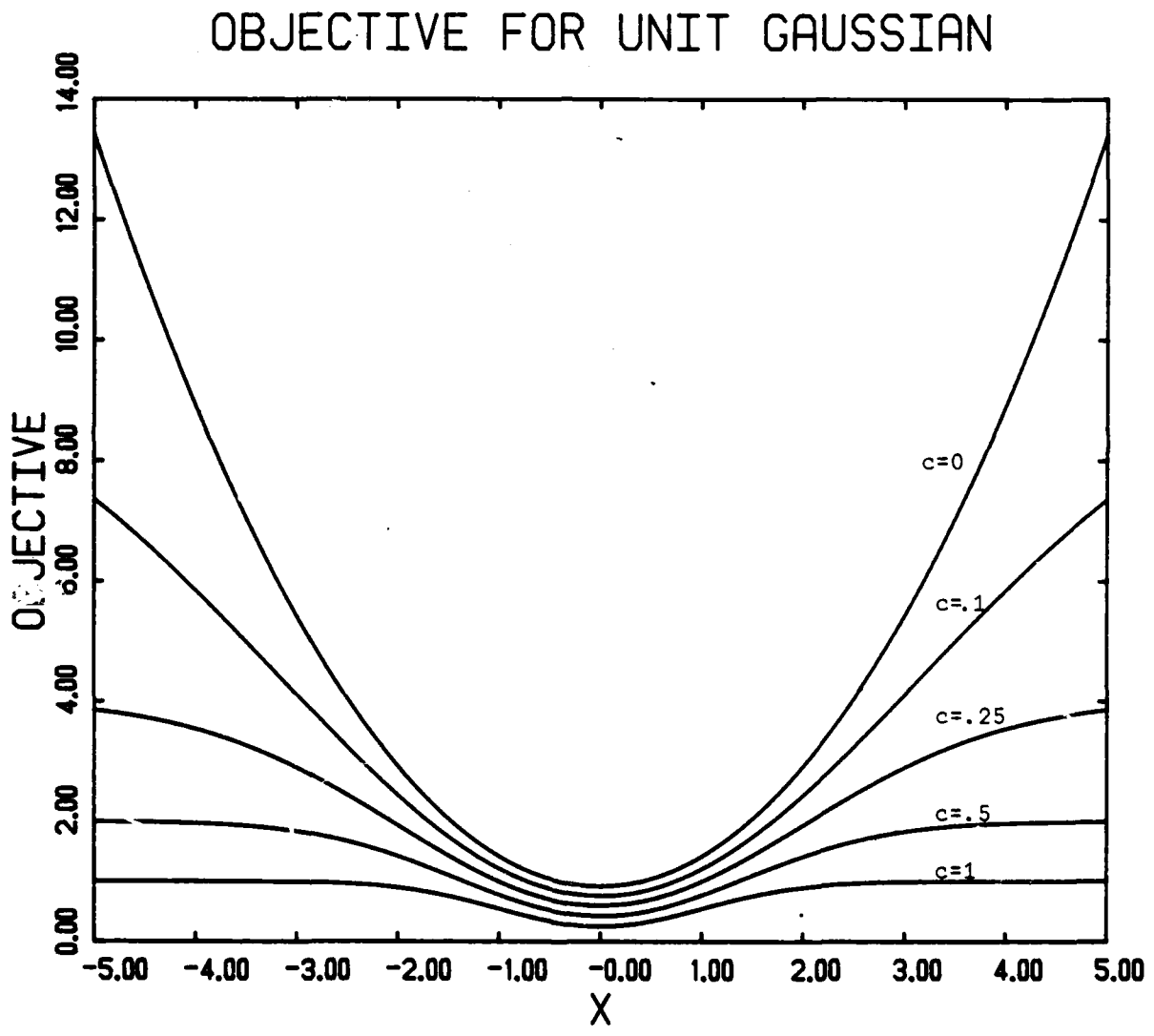
Whereas Huber recommended perturbing the quadratic form of (2.3), we shall perturb the negative of the likelihood or the Shannon information

$$I = - \sum_{j=1}^{n} \log n(x_j;\mu,\sigma^2). \tag{2.6}$$

The more surprising an item of information, that is an x_j which must be extreme, the larger the information - log $n(x_j;\mu,\sigma^2)$. See Barnett and Lewis (1978, Chapter 9). The information (2.6) is unbounded. If difficulties with the data-model entity are expected, it is natural to curb the information since it is unbounded. Figure 1 gives a plot of - log $n(x;0,1)$ labeled c=0, a single term of I in (2.6). We do not advocate discarding the large information items but rather wish to partition the information so that we can find and isolate the surprising items and call attention to them for further study. The isolation of surprising items is not especially difficult when structure is not involved; graphical techniques, for example, are especially informative. However, when

Figure 1

The information measure - ℓ_{cx} for the Gaussian distribution and several values of c



structure is involved, the matter is no longer simple and powerful tools may be needed.

How should the likelihood be perturbed in order to produce procedures for sensitivity analysis and robust estimation?

We first address the problem of a general density and return to the Gaussian case in the next section. Let $x_1, x_2, \ldots, x_n$ be now a random sample from an absolutely continuous density $f(x|\theta)$ whose interval of support is non-trivially $-\infty<x<\infty$ and where, for concreteness of discussion ant without loss of generality, θ is taken to be a scalar taking on values in some open set Θ. Suppose further that $f(x|\theta)$ possesses sufficient regularity to permit the standard maximum likelihood operations (Kendall and Stuart, 1966, Vol. II, Ch. 18). For each i, i=1,2,...,n

$$\int_{-\infty}^{\infty} f(x_i|\theta)\, dx_i = 1 \tag{2.7}$$

On taking partial derivatives of both sides of (2.7) with respect to θ we find

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} f(x_i|\theta)\, dx_i = \int_{-\infty}^{\infty} \left\{ \frac{\partial \log f(x_i|\theta)}{\partial\theta} \right\} f(x_i|\theta)\, dx_i = 0 \tag{2.8}$$

The equation (2.8) implies that $E\left(\frac{\partial \log f(x_i|\theta)}{\partial\theta}\right) = 0$. Further an estimator $\hat{\theta}$, say, of θ may be obtained from the zeros of

$$\sum_{i=1}^{n} \frac{\partial \log f(x_i|\theta)}{\partial\theta} = 0 \tag{2.9}$$

which is arrived at on summing the quantity in brackets $\{\cdot\}$ in (2.8)

over i and setting the resulting expression to zero. This expression coincides with the estimating equation derived from differentiating the log likelihood

$$\ell_0(\theta) = \sum_{i=1}^{n} \log f(x_i|\theta) = \lim_{c\to 0} \sum_{i=1}^{n} \frac{f^c(x_i|\theta)-1}{c} \tag{2.10}$$

with respect to θ.

Now the likelihood is essentially a geometric mean which is in turn a special case of the generalized mean

$$M(\theta,c) = \left(\frac{1}{n} \sum_{i=1}^{n} f^c(x_i|\theta)\right)^{1/c}, \tag{2.11}$$

This suggests that it may be useful to determine estimators of θ which make use of the density raised to the cth power. There is some evidence for such an approach in the work of Paulson and Nicklin (1981) involving estimators derived from distances between characteristic functions. Under very mild regularity conditions on the density $f(x_i|\theta)$ there exists a function $Q(\theta;c)$ such that

$$\int_{-\infty}^{\infty} f^{1+c}(x_i|\theta)\, dx_i = Q(\theta;c) \tag{2.12}$$

for some $-k_1<c<k_2$, $k_1,k_2>0$. Then

$$\int_{-\infty}^{\infty} \frac{f^{1+c}(x_i|\theta)}{Q(\theta;c)}\, dx_i = 1, \tag{2.13}$$

and

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}\frac{f^{1+c}(x_i|\theta)}{Q(\theta;c)}\,dx_i=\int_{-\infty}^{\infty}\left[(1+c)\,\frac{f_i^{1+c}}{Q}\,\frac{\partial\log f_i}{\partial\theta}-\frac{f_i^{1+c}}{Q}\,\frac{\partial\log Q}{\partial\theta}\right]dx_i=0, \tag{2.14}$$

on interchange of differentiation and integration. The arguments of $f(x_i|\theta)$ and $Q(\theta;c)$ have been deleted in (2.14) for notational convenience and this will often be done where there is no danger of misinterpretation. From (2.14) we find

$$\int_{\infty}^{\infty}f_i\left\{\frac{f_i^c}{Q}\left[(1+c)\,\frac{\partial\log f_i}{\partial\theta}-\frac{\partial\log Q}{\partial\theta}\right]\right\}dx_i=0. \tag{2.15}$$

We thus choose as our estimating equation for θ,

$$\sum_{i=1}^{n}f^c(x_i|\theta)\left[(1+c)\,\frac{\partial\log f(x_i|\theta)}{\partial\theta}-\frac{\partial\log Q(\theta;c)}{\partial\theta}\right]=0, \tag{2.16}$$

the quantity in $\{\cdot\}$ in (2.15), indexed on i and summed over i. This is exactly analogous to the way in which (2.9) was determined to be the maximum likelihood estimator of θ. Observe that $Q(\theta;0)=1$ and that (2.16) reduces to the usual likelihood equation. The maximum likelihood estimating equation (2.9) was developed by means of score function arguments. Equation (2.16) is also developed by means of score function arguments. A bona fide objective function which gives rise to (2.16) would be of considerable practical and theoretical importance. If we regard (2.16) as a differential equation we find that the corresponding objective function is given by

$$\ell_c(\theta)=\frac{1}{c}\sum_{i=1}^{n}\left|\frac{f^c(x_i|\theta)}{(Q(\theta;c))^{c/(1+c)}}-1\right| \tag{2.17}$$

as may be verified by differentiation with respect to θ. When $c \to 0$ in (2.17), $\ell_c(\theta) \to \ell_0(\theta)$, the log likelihood. It is interesting to note that $\ell_c(\theta)$ does not coincide with the right-most side of (2.10).

Perhaps a comment on what we are not doing is merited here. We are not introducing a new density which is now a function of the original parameter θ and a new statistical parameter c. It is helpful to view c as an index of how information is to be processed and which is entirely at the data analyst's disposal. The objective is still the estimation of the parameter θ of the assumed (and tentative) density $f(x|\theta)$ based on the sample data $x_1, x_2, \ldots, x_n$ but now we are interested in separating out surprising items by utilizing in the estimation process different measures of information, namely those indexed on c and given in (2.17). For example, Figure 1 provides plots of

$$\ell_{cx} = \frac{1}{c}\left|\frac{n^c(x|0,1)}{[Q(0,1;c)]^{c/(1+c)}} - 1\right|$$

with c=0, .1, .25, .5, 1. This measure of information is bounded for $c>0$ but is unbounded for $c \leq 0$. Increasingly greater weight is given to tail observations as c decreases from zero.

Apart from the case c=0, an interesting special case is provided by c=1 for which θ is estimated by maximizing $\sum_{i=1}^{n} f(x_i|\theta)/Q^{\frac{1}{2}}(\theta;1)$. No other special case seems to be of articular interest. An estimator $\hat{\theta}_c$ may be determined as that value of θ which maximizes $\ell_c(\theta)$. Equation (2.17) holds under quite general conditions. For example, θ and the x_i need not be restricted to scalar values, and the density f need not be absolutely continuous.

The above construction provides a means by which we can determine the response of a solution to changes in the way the information is processed. In particular, the quantity

$$\frac{\hat{\theta}_c - \hat{\theta}_{c'}}{c-c'}$$

will provide a qualitative measure of the sensitivity of response of the estimator $\hat{\theta}_c$ to changes in the user-specified index c. The values c>0, c'=0 are especially interesting. Surprising items are not permitted to exert a large influence on the information measure under c>0 and thus are not permitted to exert a large influence on the parameter estimates. When the interval of support of the density $f(x|\theta)$ is infinite on both the left and the right then surprising items will be identified by a low weight $f^c(x|\theta)/Q(\theta;c)$ in (2.16) (see also (2.15)). Indeed, the pattern of the weights $f^c(x_i|\theta)/Q(\theta;c)$ provides very useful diagnostic information. We now examine some special cases.

3. The Gaussian Distribution

The most important error model in statistics is provided by the Gaussian distribution given in (2.1). Suppose c>0 is fixed. We shall derive robust estimators for μ and σ^2 from (2.17). We find by straightforward integration that

$$\int_{-\infty}^{\infty} n^{1+c}\,(x|\mu,\sigma^2)dx = [(1+c)(2\pi\sigma^2)^c]^{-\frac{1}{2}} = Q(\mu,\sigma^2;c). \qquad (3.1)$$

Let $\ell_c(\mu,\sigma^2)$ be the two-parameter analogue of (2.17). Differentiation

of $\ell_c(\mu,\sigma^2)$ with respect to μ and σ^2 or substitution in (2.16) of $Q(\mu,\sigma^2,c)$ leads to the simultaneous estimating equations

$$\sum_{i=1}^{n} (x_i-\mu) \exp\left(-\frac{c}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right) = 0, \tag{3.2}$$

$$\sum_{i=1}^{n} \{(1+c)(x_i-\mu)^2 - \sigma^2\} \exp\left(-\frac{c}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right) = 0. \tag{3.3}$$

The estimators $\hat{\mu}_c$ and $\hat{\sigma}_c^2$ jointly satisfy the implicit equations

$$\mu = \frac{\sum x_i v_{ic}}{\sum v_{ic}}, \tag{3.4}$$

$$\sigma^2 = (1+c)\frac{\sum (x_i-\mu)^2 v_{ic}}{\sum v_{ic}}, \tag{3.5}$$

where

$$v_{ic} = \exp\left(-\frac{c}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right). \tag{3.6}$$

The estimators $\hat{\mu}_c$ and $\hat{\sigma}_c^2$ have a single solution for c in a neighborhood of zero. Just how large this neighborhood is as a function of n and c is not known. However, when c becomes too large multiple solutions may arise. The estimators $\hat{\mu}_c$ and $\hat{\sigma}_c^2$ are M-estimators and are consistent for μ and σ^2 when the $x_1,x_2,\ldots,x_n$ are a random sample from a Gaussian distribution with mean μ and variance σ^2, provided the consistent zeros of (3.2) and (3.3) are chosen. In this case $n^{\frac{1}{2}}(\hat{\mu}_c-\mu, \hat{\sigma}_c^2-\sigma^2)$ is asymptotically bivariate normal with mean vector $\underset{\sim}{0}$ and covariance matrix

$$\underset{\sim}{V} = \sigma^2 \left(\frac{(1+c)^2}{1+2c}\right)^{3/2} \begin{pmatrix} 1 & 0 \\ 0 & \frac{2+4c+3c^2}{1+2c} \end{pmatrix}. \tag{3.7}$$

(Asymptotic expressions are given as we proceed since they will usually be more naturally presented in this manner. Details are given in section 7 .) Thus $\hat{\mu}_c$ and $\hat{\sigma}_c^2$ are asymptotically independently distributed. The value $c = -\frac{1}{2}$ is a singularity of the covariance matrix $\underset{\sim}{V}$ and only values $-\frac{1}{2}<c<\infty$ are permissable. The asymptotic efficiencies for $\hat{\mu}_c$ and $\hat{\sigma}_c^2$ relative to $\hat{\mu}_0$ and $\hat{\sigma}_0^2$ are readily computed from $\underset{\sim}{V}$ and are given in Table 1.

Table 1

Efficiencies of the Estimators $\hat{\mu}_c$, $\hat{\sigma}_c^2$ for Selected Values of c

Estimators	-.3	-.2	-.1	0	.1	.2	.3	.4	.5	1.0
$\hat{\mu}_c$	.738	.908	.982	1	.988	.959	.921	.880	.838	.650
$\hat{\sigma}_c^2$	.631	.825	.964	1	.975	.919	.850	.777	.706	.433

The asymptotic efficiencies remain high over a broad range of values of c. The influence function for $\hat{\mu}_c$ and $\hat{\sigma}_c^2$ at the Gaussian distribution with mean μ and variance σ^2 are proportional to the score functions (Huber, 1981, p. 45) determined from (3.2) and (3.3) respectively; that is

$$IC(\hat{\mu}_c;y,N) \propto (y-\mu)\, f^c(y|\mu,\sigma^2), \tag{3.8}$$

$$IC(\hat{\sigma}_c^2;y,N) \propto \{(1+c)(y-\mu)^2 - \sigma^2\}\, f^c(y|\mu,\sigma^2). \tag{3.9}$$

Both influence functions are bounded and re-descendent to zero for all $c>0$. Both estimators $\hat{\mu}_c$ and $\hat{\sigma}_c^2$ have high breakdown bounds for sample sizes $n\geq 10$.

The data presented in Table 2 are taken from David and Quesenberry (1961) and are tentatively from a Gaussian population. Also presented

in Table 2 are the final weights $\hat{v}_{ic}$ for c=0, .3, .5. As c increases the weights $\hat{v}_{ic}$, i=14,15,16 decrease dramatically indicating that these observations and the single Gaussian parent assumption may not be mutually consistent. Of course tail observations will be weighted lower as c increases even if the data were Gaussian but not to the extent seen here. The rate of change $(\hat{v}_{14,.3} - \hat{v}_{14,0})/(.3-0) = -1.02$ is substantial but it is not known if this is satatistically significant. Resolution of the distribution of the rate of change statistic seems to be a difficult problem but one worth some attention and is perhaps best addressed by simulation. The parameter estimates $\hat{\mu}_c$ and $\hat{\sigma}_c$ are presented in Table 3. As c increases from 0 the rate of change of $\hat{\mu}_c$ and $\hat{\sigma}_c$ is large. However, when c is approximately .85, $\hat{\sigma}_c$ no longer decreases with c but begins to increase. This behavior is typical of the behavior of $\hat{\sigma}_c$ when the data are indeed Gaussian. The estimators $\hat{\mu}_c$ and $\hat{\sigma}_c$ should be approximately independent if the data were from a Gaussian population. When c increases from 0 to .2, the estimated asymptotic correlation $\hat{\rho}$ increases from 0 to .88. However, when $c \doteq .85$ this estimated correlation is approximately zero while for c>.85 the correlation becomes negative.

This example captures several aspects of the procedure which hold generally for Gaussian error models with or without structure. First, the weights $\hat{v}_{ic}$ can be advantageously used to determine the extent to which data and model are internally consistent. Observations which receive low weights are prime candidates for further study. Second, one might expect an estimate of σ^2 computed from (3.5) to decrease with increasing c. This is not the case. When the data are Gaussian the estimate of σ^2

Table 2

Data and final observational weights $\hat{v}_{ic}$ (×1000) for several values of c

	Observation	c=0	c=.3	c=.5
1	.32	62.5	61	55
2	.35	62.5	66	67
3	.37	62.5	69	74
4	.38	62.5	71	77
5	.39	62.5	72	80
6	.44	62.5	76	88
7	.45	62.5	76	88
8	.46	62.5	76	87
9	.47	62.5	76	86
10	.48	62.5	76	85
11	.52	62.5	74	74
12	.53	62.5	73	71
13	.57	62.5	67	55
14	.74	62.5	32	7
15	.74	62.5	32	7
16	1.09	62.5	.09	.34(-3)

will remain roughly constant or increase slightly as c is increased. The estimate of μ also remains roughly constant as c increases. Numerical and sample size considerations determine the magnitude which c may take. We typically find $0 \leq c \leq 1$ most useful for the Gaussian distribution although we have made use of values of $c > 1$ in practical settings. Apart from numerical difficulties $\hat{\mu}_c \to \max\{f(x_i)\}$, i.e. the mode, as c becomes large. Third, the estimated asymptotic correlation provides a useful diagnostic in a data analysis. These three comments are based on extensive practical and simulation experience.

4. Other Distributions

We now show that (2.17) may be used to construct sensitivity analyses and robust estimators for c variety of distributions other than the Gaussian. It is worth emphasizing at this point that sensitivity analyses which are concerned primarily with error models are not of primary importance because problems associated with error models without intervening structure are relatively easy to deal with. The main interest is in having a procedure which is capable of dealing with the combination of error and structural models since problems with the model or with the data or with both may be very difficult to detect since, quite often, efficient estimation procedures hide more than they illuminate. At the heart of these problems is the modeling process itself. The ultimate objective is to explore and uncover an appropriate, hopefully parsimonious, model

Table 3

Parameter Estimates $\hat{\mu}_c$ and $\hat{\sigma}_c$ and Estimated Asymptotic Correlation $\hat{\rho}$

c	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}$
0.0	.5189	.189	0
0.2	.4786	.141	.88
0.3	.4628	.116	.83
0.5	.4456	.091	.56
0.6	.4427	.087	.30
0.7	.4415	.085	.15
0.8	.4410	.085	.05
0.85	.4409	.086	.01
0.9	.4408	.086	-.03
1.0	.4407	.087	-.09
1.5	.4411	.091	-.31
2.0	.4421	.094	-.51
-0.1	.5344	.198	.78
-0.2	.5497	.203	.78
-0.3	.5651	.204	.77
-0.5	.5989	.195	.77

to describe a body of data. However, once we have in hand a procedure for various types of error models, the extension to error-structural models is not nearly as difficult. Few papers have dealt with error models other than the Gaussian. Thall (1979) has proposed a procedure for the exponential distribution based on a modification of Huber's procedure. Hampel (1968) has proposed a general procedure but it does not seem to be numerically viable. The following error distributions are explicitly considered because they are used in structured situations.

a. The Logistic Distribution. The logistic distribution is a location and scale family which is not a member of the exponential class and which is sometimes used in place of the Gaussian distribution. It has density

$$l(x|\mu,\tau) = \frac{1}{\tau} \frac{\exp((x-\mu)/\tau)}{\{1 + \exp((x-\mu)/\tau)\}^2} \tag{4.1}$$

for $\tau>0$, $-\infty<\mu<\infty$. The corresponding distribution function is

$$L(x|\mu,\tau) = \frac{\exp((x-\mu)/\tau)}{1 + \exp((x-\mu)/\tau)} . \tag{4.2}$$

The integral

$$\int_{-\infty}^{\infty} l^{1+c}(x|\mu,\tau)\, dx = \frac{1}{\tau^c} \int_{-\infty}^{\infty} \{L(1-L)\}^c \ell\, dx$$

$$= \frac{1}{\tau^c} \int_0^1 u^c(1-u)^c\, du$$

$$= \frac{B(1+c,1+c)}{\tau^c} = Q(\mu,\tau;c) \tag{4.3}$$

on making the transform $u = L(x|\mu,\tau)$. $B(1+c,1+c)$ denotes the complete

beta function with arguments 1+c and 1+c. Estimators $\hat{\mu}_c$ and $\hat{\tau}_c$ are determined from maximization of the analogue $\ell_c(\mu,\tau)$ of (2.17). We find with a little effort that the estimators $\hat{\mu}_c$, $\hat{\tau}_c$ jointly satisfy

$$\sum_{i=1}^{n} 1^c(x_i|\mu,\tau) \tanh\left(\frac{x_i-\mu}{2\tau}\right) = 0, \tag{4.4}$$

and

$$\sum_{i=1}^{n} 1^c(x_i|\mu,\tau) \left\{1+(1+c)\frac{x_i-\mu}{\tau} \tanh\left(\frac{x_i-\mu}{2\tau}\right)\right\} = 0. \tag{4.5}$$

The equations (4.4) and (4.5) reduce to those of maximum likelihood when c=0. The score functions at the logistic distribution are proportional to the influence functions and satisfy

$$S(\hat{\mu}_c;y,1) = 1^c(y|\mu,\tau) \tanh\left(\frac{y-\mu}{2\tau}\right), \tag{4.6}$$

$$S(\hat{\tau}_i;y,1) = 1^c(y|\mu,\tau) \{1+(1+c)\frac{y-\mu}{\tau} \tanh\left(\frac{y-\mu}{2\tau}\right)\}, \tag{4.7}$$

both of which are bounded and redescend to zero. The right hand side of (4.6) is bounded for c=0. The right hand side of (4.7) is, however, not bounded for c=0 and increases linearly in $|y|$.

b. <u>The Gamma Distribution</u>. The gamma distribution with scale parameter β and shape parameter α has density

$$g(x|\alpha,\beta) = \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}}, \quad x > 0 \tag{4.8}$$

for $\alpha, \beta>0$. This distribution, especially when $\alpha=1$, is widely used as a failure distribution. Using (2.17) we find that the estimators $\hat{\alpha}_c$, $\hat{\beta}_c$ jointly satisfy

$$\sum_{i=1}^{n} g^c(x_i|\alpha,\beta)\ \{\log x_i - \log\left(\frac{\beta}{1+c}\right) - \psi(\alpha(1+c)\} = 0 \qquad (4.9)$$

$$\sum_{i=1}^{n} g^c(x_i|\alpha,\beta)\ \{(1+c)\ \frac{x_i}{\beta} - (\alpha(1+c)-c)\} = 0, \qquad (4.10)$$

when $x_1, x_2, \ldots, x_n$ is a random sample putatively from the distribution (4.11). Here $\psi(z)$ is the digamma function with argument z. As a special case take $\alpha=1$. Then the estimator $\hat{\beta}_c$ for the mean of an exponential distribution satisfies the implicit equation

$$\beta = (1+c)\ \frac{\sum_{i=1}^{n} x_i\ g^c(x_i|1,\beta)}{\sum_{i=1}^{n} g^c(x_i|1,\beta)}$$

$$= (1+c)\ \frac{\sum x_i\ e^{-cx_i/\beta}}{\sum e^{-cx_i/\beta}}\ . \qquad (4.11)$$

Thus observations which are far removed from β will receive a low weight $e^{-cx_i/\beta}$ when $c>0$.

Even though the estimators $\hat{\alpha}_c$ and $\hat{\beta}_c$ are in a reasonably attractive form it is numerically and statistically more appealing to make the transformation $y = \log x$. The resulting density is

$$g_*(y|\alpha,\phi) = \frac{1}{\Gamma(\alpha)} \exp\{\alpha(y-\phi) - \exp(y-\phi)\},\ -\infty<y<\infty, \qquad (4.12)$$

where $\phi = \log \beta$ plays the role of a location parameter and now α is similar to a scale parameter. The log-gamma density is unimodal with well-behaved tails. On evaluating $Q_*(\alpha,\phi;c)$ corresponding to (4.12) and

substituting in (2.17) we find directly that the estimators $\hat{\phi}_c$ and $\hat{\alpha}_c$ satisfy

$$\sum_{i=1}^{n} g_*^c(y_i|\alpha,\phi)\ \{\exp(y_i-\phi) - \alpha\} = 0, \tag{4.13}$$

$$\sum_{i=1}^{n} g_*^c\ (y_i|\alpha,\phi)\ \{y_i - \phi + \log(1+c) - \psi(\alpha)\} = 0. \tag{4.14}$$

In this case $g_*^c(y|\alpha,\phi)/Q_*^{c/(1+c)}$ is bounded in y for all ϕ and $\alpha>0$ while $g^c(x|\alpha,\beta)/Q^{c/(1+c)}$ is not bounded for $\alpha<1$ and all $c>0$ as $x\to 0$. When the support of the random variable is nontrivially on $(-\infty,\infty)$ the information quantity is always bounded when $c>0$.

In the special case when $\alpha=1$ and assumed known and only β is being estimated the estimator $\hat{\beta}_c$ satisfies the implicit relationship

$$\beta = \frac{\sum_{i=1}^{n} x_i^{1+c}\ e^{-cx_i/\beta}}{\sum_{i=1}^{n} x_i^{c}\ e^{-cx_i/\beta}} \tag{4.15}$$

as may be verified from (4.13) and letting $x = e^y$. The structural differences in the estimators for β provided by (4.11) and (4.15) are interesting in their own right.

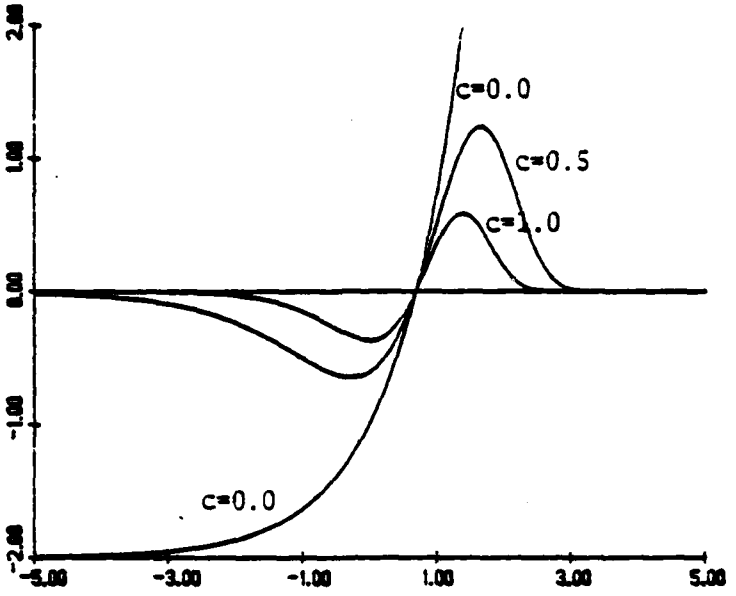
The score functions for $\hat{\alpha}_c$ and $\hat{\beta}_c$ at the log-gamma distribution with $\alpha=2$, $\phi=0$ are plotted in Figure 2 for several values of c. Observe that the score functions for $\hat{\alpha}_c$ and $\hat{\phi}_c$ are both unbounded when $c=0$.

c. <u>The Weibull Distribution</u>. The Weibull distribution with shape parameter k and scale parameter θ is given by

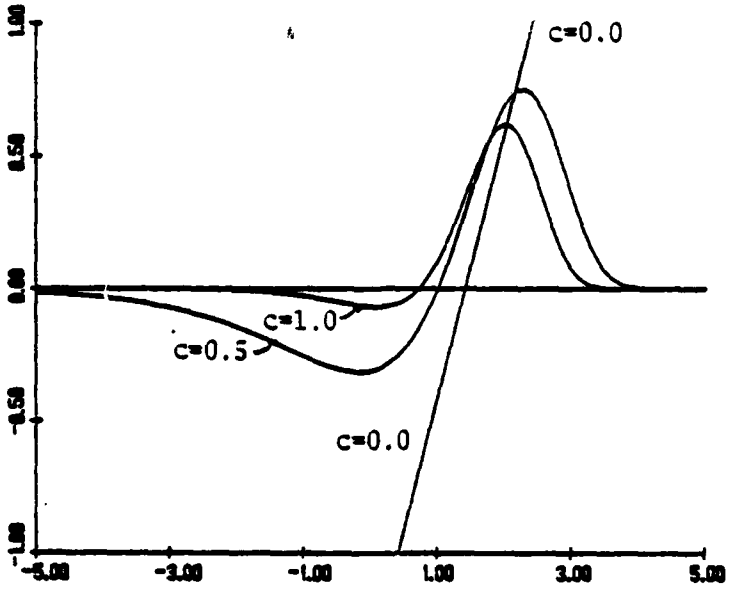
Figure 2

Score functions for the estimators $\hat{\phi}_c$ and $\hat{\alpha}_c$ at the log-gamma density, $\phi=0$, $\alpha=2$.

(a)



(b)



$$w(x|k,\theta) = \frac{k}{\theta}\left(\frac{x}{\theta}\right)^{k-1} \exp\{-\left(\frac{x}{\theta}\right)^{k}\},\ \theta>0,\ k>0. \tag{4.16}$$

It is more convenient to deal with the distribution of $y = \log x$ than with that of x itself. The log-Weibull (or Type I extreme value) density is given by

$$w_*(y|k,\phi) = k \exp[k(y-\phi) - \exp\{k(y-\phi)\}],\ -\infty<y<\infty, \tag{4.17}$$

where $\phi = \log \theta$. The function

$$Q_*(k,\phi,c) = \int_{-\infty}^{\infty} w_*^{1+c}(y|k,\phi)\, dy = \frac{k^c\, \Gamma(1+c)}{(1+c)^{1+c}}. \tag{4.18}$$

We thus obtain from (2.16) or (2.17) the joint estimating equations for the parameters ϕ and k, respectively,

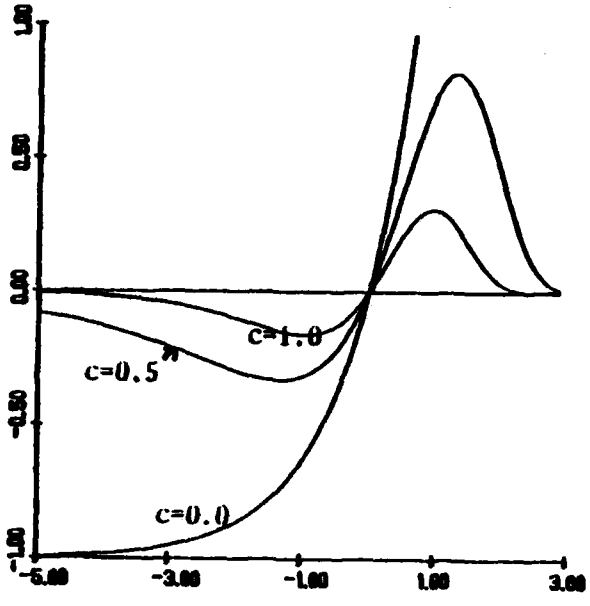
$$\sum_{i=1}^{n} [-k + k \exp\{k(y_i-\phi)\}]\, k^c \exp\{ck(y_i-\phi) - c \exp\{k(y_i-\phi)\}\} = 0, \tag{4.19}$$

$$\sum_{i=1}^{n} [(1+c)\{\frac{1}{k} + (y_i-\phi) - (y_i-\phi)\exp\{k(y_i-\phi)\}\} - \frac{c}{k}]\, k^c \exp[ck(y_i-\phi) - c \exp\{k(y_i-\phi)\}\}] = 0. \tag{4.20}$$

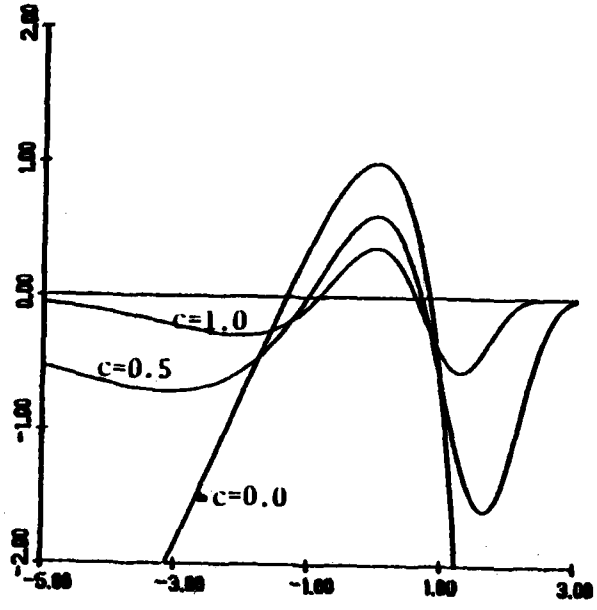
Figure 3 depicts the score functions $S(\hat{k},y,w_*)$ and $S(\hat{\phi},y,w_*)$ at the log-Weibull distribution with $\phi=0$ and $k=1$. These functions are bounded and redescending to zero when $c>0$. The estimators are thus qualitatively robust for fixed $c>0$. Estimators for k and θ can also be developed without the transformation $y = \log x$ but we prefer those of (4.19) and (4.20). The efficiencies of the estimators $\hat{\phi}_c$ and $\hat{k}_c$ and the asymptotic correlation $\rho(\hat{\phi}_c,\hat{k}_c)$ between the estimators are given in Table 4. It is interesting to observe that for a given index c the efficiencies for the

Figure 3

Some functions for the estimators $\hat{\phi}_c$ and $\hat{k}_c$ at the extreme value distribution with location $\phi=0$ and shape $k=1$



3(a)



3(b)

Table 4

Joint Asymptotic Relative Efficiencies of the
Log-Weibull (or extreme value) Estimators

c	eff. ($\hat{\phi}$)	eff. ($\hat{k}$)	asym. corr. ($\hat{\phi}$,$\hat{k}$)
0.0	1.0	1.0	.313
0.1	.987	.974	.313
0.2	.957	.917	.312
0.3	.919	.847	.311
0.4	.877	.775	.308
0.5	.835	.705	.306
0.6	.793	.639	.303
0.7	.754	.580	.300
0.8	.716	.527	.296
0.9	.680	.479	.293
1.0	.647	.436	.289
1.5	.511	.283	.272
2.0	.414	.193	.256
3.0	.290	.103	.230
0.0	1.0	1.0	.313
-0.1	.980	.961	.314
-0.2	.898	.814	.320
-0.3	.703	.522	.349
-0.4	.325	.155	.475

estimators of the location parameter ϕ and of the scale parameter k are very close to those given in Table 1 for the location and scale estimators of the Gaussian distribution for the same c. For example, the efficiency of $\hat{\phi}_{.5}$ is .835 while the efficiency of $\hat{\mu}_{.5}$ is .838. This characteristic seems to hold more generally.

d. The Poisson Distribution. The Poisson distribution with mean μ has frequency function

$$P(x|\mu) = \frac{\mu^x e^{-\mu}}{x!}, \qquad x = 0,1,2,\ldots \quad . \tag{4.21}$$

The development of the self-critical estimators is exactly as above. However, in this case the function

$$Q(\mu;c) = \sum_{x=0}^{\infty} P^{1+c}(x|\mu) \tag{4.22}$$

cannot be evaluated in closed form. The estimator $\hat{\mu}$ of μ must involve an iterative evaluation of $Q(\mu;c)$ in the implicit equation (2.16). If μ is large, a normal approximation to the Poisson may be used, say through (3.2), to reduce the computational effort. The exact implicit equation for μ is

$$\sum_{j=1}^{n} P^c(x_j|\mu) \left\{ \frac{x_i}{\mu} - \frac{\sum_{x=0}^{\infty} \mu^{(1+c)x-1}/(x!)^{1+c}}{\sum_{x=0}^{\infty} \mu^{(1+c)x}/(x!)^{1+c}} \right\} = 0. \tag{4.23}$$

e. Clearly, equations (2.16) and (2.17) may be applied to a wide variety of distributions. For example, the self-critical procedure has been successfully used for the negative binomial, the binomial, the Pareto and the multivariate normal distributions in addition to those discussed here.

5. An Example

The data in Table 5 represents the time to death of 64 infants. These infants were ostensibly the victims of sudden infant death syndrome. Apart from the seven longest survival times, the Weibull model provides a reasonable model for these data. However, since the hazard function of the extreme value distribution is strictly increasing, the Weibull or extreme value model would not be appropriate. Biological considerations suggest that the hazard function for sudden infant death syndrome should be first increasing and then decreasing. The lognormal distribution may provide an appropriate statistical model for these data. Our main purpose, however, is to illustrate the procedure we propose. Table 6 provides the estimates of extreme value and lognormal parameters for various values of c.

For both the log-Weibull and log-normal models, the variation in c produces a dramatic change in the response surface generated by the parameter estimates. For example, the estimate of k changes from 1.36 to 2.33 as c goes from zero to unity. The estimate of σ^2 changes from .45 to .20 as c moves from zero to unity. The estimate of θ diminishes from 118.9 to 85.4 as c moves from zero to unity. This sensitivity of parameter estimates to change in the value of c which reflects the way in which the information is processed indicates the inconsistency of both models with the data or that some of the data are not consistent with the model. Of course, a simple probability plot will be equally effective in generating the same conclusion in the case of unstructured data. However, when we are dealing with combined structure and error models, inconsistencies

Table 5

Times to Death (in days) in an Epidemiological Study

17	[illegible]	77	113
22	63	80	123
25	63	82	128
34	65	87	135
34	65	87	146
34	65	88	148
35	66	92	149
39	66	93	158
42	67	96	160
43	68	99	218
44	73	100	234
54	74	101	267
55	75	102	329
56	76	106	372
57	77	108	455
57	77	110	492

Table 6

Paremeter Estimates for the Sudden Infant Death Syndrome Data

	log-Weibull(n=64)		log-normal(n=64)		log-gamma(n=64)		log-Weibull(n=57)	
c	$\hat{\theta}=e^{\hat{\phi}}$	$\hat{k}$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\theta}$	$\hat{k}$
0	118.9	1.36	4.43	.45	49.7	2.16	89.4	2.43
.3	97.2	1.67	4.39	.38	30.3	2.97	88.3	2.35
.5	89.1	2.14	4.37	.31	21.3	3.94	87.5	2.32
1	85.4	2.33	4.36	.20	15.6	5.19	85.3	2.35

are often hidden by the c=0 analyses. Many of these inconsistencies may be uncovered by changing the way in which the data is processed, namely, by studying the response of parameter estimates and observational weights to variation in c.

The sensitivity analysis identifies the largest seven and the smallest three times as being most inconsistent with the log-Weibull or log-normal model. Early deaths are generally consistent with an alternate mode of failure rather then the sudden infinite death syndrome. Several of the late deaths, including the latest, were of a suspicious nature. When the largest seven observations are removed and the analyses repeated, the parameter estimates and the observational weights $\hat{v}_{ic}$ remain much more stable as do the parameter estimates for both the log-Weibull and log-normal models. This is illustrated for the log-Weibull model in Table 6.

6. Some Regression Models

A great deal of attention has been devoted to robust regression in the case in which the underlying error distribution is assumed to be approximately Gaussian. The self-critical procedure we introduced in section 2 can be expected to be very useful when there are inconsistencies between model and data in the y-direction. Difficulties in the x-direction, that is in the factor space, will require augmentation of the procedures given in section 2. We shall briefly examine the cases in which the error distribution is Gaussian, extreme value (or log-Weibull), and Poisson.

First consider the regression model

$$y_{ij} = h(\underset{\sim}{x}_i, \underset{\sim}{\theta}) + z_{ij} \tag{6.1}$$

where $\underset{\sim}{x}_i = (x_{i1}, x_{i2}, \ldots, x_{im})$ is the ith set of values of the m independent variables, n_i is the number of replicates of the ith experimental condition, $\underset{\sim}{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^T$ is a p×1 vector of unknown parameters, y_{ij} is a particular realization of the experiment, and z_{ij} are the error terms. The regression function $h(\underset{\sim}{x}, \underset{\sim}{\theta})$ relates the expected value of the dependent variable to the independent variables and the parameters, and given the $\underset{\sim}{x}_i$ and the y_{ij}, we wish to estimate $\underset{\sim}{\theta}$. Assume first that the z_{ij} are approximately Gaussian with mean 0 and variance σ^2. Then given the $\underset{\sim}{x}_i$, we choose as parameter estimates those values of $\underset{\sim}{\theta}$ and σ^2 which maximize

$$\frac{1}{c} \sum_i \sum_j \left\{ \frac{n^c(z_{ij} | \underset{\sim}{\theta}, \underset{\sim}{x}, \sigma^2)}{[Q(\underset{\sim}{\theta}, \sigma^2; c)]^{c/(1+c)}} - 1 \right\} , \tag{6.2}$$

where

$$n(z_{ij} | \underset{\sim}{\theta}, x, \sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\{- \frac{1}{2\sigma^2} (y_{ij} - h(\underset{\sim}{x}_i, \underset{\sim}{\theta}))^2\}, \tag{6.3}$$

and $Q(\underset{\sim}{\theta}, \sigma^2; c) = [(1+c)(2\pi\sigma^2)^c]^{-\frac{1}{2}}$ as in (3.1). The joint estimators of $\underset{\sim}{\theta}$ and σ^2 for various values of c will allow us to perform a sensitivity analysis.

Type I extreme value regression arises as a natural consequence of the parametric proportional hazards model of Cox (1972). In this case the z_{ij} of (6.1) will follow the distribution (4.15). Estimators for $\underset{\sim}{\theta}$ and k will be determined from maximization of

$$\frac{1}{c} \sum_i \sum_j \left\{ \frac{w_*^c(z_{ij}|\underset{\sim}{\theta},\underset{\sim}{x},k)}{[Q(\underset{\sim}{\theta},k;c]^{c/(1+c)}} - 1 \right\} \tag{6.4}$$

where

$$w_*(z_{ij}|\underset{\sim}{\theta},\underset{\sim}{x},k) = k \exp[k(y_{ij}-f(\underset{\sim}{x}_i,\underset{\sim}{\theta})) - \exp(k(y_{ij}-f(\underset{\sim}{x}_i,\underset{\sim}{\theta}))], \tag{6.5}$$

and $Q(\underset{\sim}{\theta},k;c)$ is given by (4.15). We have found extreme value regression to be a useful alternative to the usual Gaussian regression in several engineering applications where an analysis of the hazard structure dictated a model other than Gaussian. The ability to perform a sensitivity analysis or a robust analysis in this case has been of considerable use in model evolution as well.

If we constrain $f(\underset{\sim}{x}_i,\underset{\sim}{\theta})$ to be positive we may also develop in a straightforward fashion a sensitivity analysis or, for fixed $c>0$, a robust analysis for Poisson regression. Finally, the procedure of section 2 produces attractive and easy to use procedures for experimental design. We shall consider the regression and design topics elsewhere.

7. Asymptotic Covariances and Efficiencies

The self-critical estimators are M-estimators and as such are consistent and asymptotically normal under regularity conditions similar to those of regular maximum likelihood estimators. Just as in the case of maximum likelihood, $c=0$, we cannot always be sure that there will be a unique solution for the estimating equations. From among the local optima, the consistent solution is assumed to be the one taken. Asymptotic variance-covariance matrices of the estimators are readily

determined from (2.16) or (2.17) by standard expansion arguments. Let $x_1, x_2, \ldots, x_n$ be a random sample from the density or frequency function $f(x|\underset{\sim}{\theta})$ where $\underset{\sim}{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^T$ is a p×1 vector of parameters. Let

$$\ell_{cX} = c^{-1} \left\{ \frac{f^c(X|\underset{\sim}{\theta})}{[Q(\underset{\sim}{\theta};c)]^{c/(c+1)}} - 1 \right\} \tag{7.1}$$

where the random variable X has the same distribution as the x_j. Then the asymptotic variance-covariance matrix for the estimator $\hat{\underset{\sim}{\theta}}_c$ is determined from the matrices $\underset{\sim}{H}$ and $\underset{\sim}{\sum}$ whose elements are given by

$$V_{\theta\theta'} = E\left(\frac{\partial \ell_{cX}}{\partial \theta} \frac{\partial \ell_{cX}}{\partial \theta'} \right), \tag{7.2}$$

$$H_{\theta\theta'} = E\left(\frac{\partial^2 \ell_{cX}}{\partial \theta \partial \theta'} \right), \tag{7.3}$$

where $\theta, \theta' = \theta_1, \theta_2, \ldots, \theta_p$. The asymptotic covariance matrix of the estimator $\hat{\underset{\sim}{\theta}}_c$ is

$$\underset{\sim}{\sum} = n^{-1} \underset{\sim}{H}^{-1} \underset{\sim}{V} \underset{\sim}{H}^{-1}. \tag{7.4}$$

From (2.15) we have

$$\int_{-\infty}^{\infty} f^{1+c} \left[(1+c) \frac{\partial \log f}{\partial \theta} - \frac{\partial \log Q}{\partial \theta} \right] dx = 0;$$

differentiating this expression with respect to θ' we find

$$\int_{-\infty}^{\infty} \left[(1+c) f^{1+c} \frac{\partial \log f}{\partial \theta'} \left\{ (1+c) \frac{\partial \log f}{\partial \theta} - \frac{\partial \log Q}{\partial \theta} \right\} + f^{1+c} \left\{ (1+c) \frac{\partial^2 \log f}{\partial \theta \partial \theta'} - \frac{\partial^2 \log Q}{\partial \theta \partial \theta'} \right\} \right] dx = 0$$

which implies

$$E\left[f^c \frac{\partial \log f}{\partial \theta'} \left| \frac{\partial \log f}{\partial \theta} - \frac{1}{1+c} \frac{\partial \log Q}{\partial \theta} \right.\right] = - \frac{1}{1+c} E\left[f^c \left| \frac{\partial^2 \log f}{\partial \theta \partial \theta'} \right.\right.$$

$$\left.\left. - \frac{1}{(1+c)} \frac{\partial^2 \log Q}{\partial \theta \partial \theta'} \right|\right], \qquad (7.5)$$

a generalization of Fisher's information identity. We thus find that

$$H_{\theta\theta'} = E\left[\frac{\partial^2 \ell_{cX}}{\partial \theta \partial \theta'}\right] = \frac{1}{1+c} E\left[f^c \left| \frac{\partial^2 \log f}{\partial \theta \partial \theta'} - \frac{1}{1+c} \frac{\partial^2 \log Q}{\partial \theta \partial \theta'} \right|\right], \qquad (7.6)$$

while

$$V_{\theta\theta'} = E\left[f^{2c} \left(\frac{\partial \log f}{\partial \theta} - \frac{1}{1+c} \frac{\partial \log Q}{\partial \theta}\right)\left(\frac{\partial \log f}{\partial \theta'} - \frac{1}{1+c} \frac{\partial \log Q}{\partial \theta'}\right)\right]. \qquad (7.7)$$

The expressions (7.6) and (7.7) are often easy to evaluate. Equation (7.4) has been used to produce the efficiencies quoted in section 3 and 4. Both (7.6) and (7.7) are easy to approximate in the practical setting where the true value of $\underset{\sim}{\theta}$ is not known, for example

$$\hat{H}_{\theta\theta'} = n^{-1} \sum_{j=1}^{n} \frac{\partial^2 \ell_{cx_j}}{\partial \theta \partial \theta'} \Bigg|_{\underset{\sim}{\theta}=\hat{\underset{\sim}{\theta}}_c}, \qquad (7.8)$$

or the estimators $\hat{\underset{\sim}{\theta}}$ may be substituted directly in the expression for expectations.

Based on simulation experience, the estimators $\hat{\underset{\sim}{\theta}}_c$ approach their asymptotic distributions very rapidly, often for n as small as 10 and 20. This is due to the smoothing induced by the term f^c in (2.16).

8. Additional Families of Estimators and Discussion

Let $f(x|\theta)$ be continuous density defined over the nontrivial interval of support $(-\infty,\infty)$ and let θ take on values in an open set Θ. Let $g(x|\theta)$ be a function of x over $-\infty<x<\infty$, $\theta\epsilon\Theta$. Suppose further that there exits a function $k_g(\theta)$ such that

$$\int_{-\infty}^{\infty} f(x|\theta)\, g(x|\theta)\, dx = k_g(\theta). \tag{8.1}$$

We have then that the inner product of f with g is normalized by $k_g(\theta)$, i.e.

$$\int_{-\infty}^{\infty} \frac{f(x|\theta)\, g(x|\theta)}{k_g(\theta)}\, dx = 1. \tag{8.2}$$

Under mild regularity conditions

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} \frac{fg}{k_g}\, dx = \int_{-\infty}^{\infty} \frac{fg}{k_g}\left\{\frac{\partial \log f}{\partial\theta} + \frac{\partial \log g}{\partial\theta} - \frac{\partial \log k_g}{\partial\theta}\right\} dx. \tag{8.3}$$

If $x_1,x_2,\ldots,x_n$ is a random sample from $f(x|\theta)$, then an estimator for θ may be determined from the consistent zero of (see (2.15))

$$\sum_{i=1}^{n} g(x_i|\theta)\left\{\frac{\partial \log f(x_i|\theta)}{\partial\theta} + \frac{\partial \log g(x_i|\theta)}{\partial\theta} - \frac{\partial \log k_g}{\partial\theta}\right\} = 0. \tag{8.4}$$

Because we have taken $g(x_i|\theta)$ to be $f^c(x_i|\theta)$, in sections 2-6, we have termed the estimators of this paper self-critical. The choice $f^c(x_i|\theta)$ seems to be most useful in practice although the estimators determined from (8.4) may prove to be useful in some contexts.

The procedure proposed in this paper is envisioned to be of greatest use in sensitivity analysis and model evolution settings. It also provides a tool for the identification of outliers in the structural data case since the model plays a direct role (through the f^c term in (2.16) or (2.17)) in the processing of information. Finally, we have provided a general, easily used, computationally attractive procedure which stems from a single primitive concept for the construction of simultaneous robust estimators.

References

1. Barnett, V. and Lewis, T. (1978). Outliers in Statistical Data. New York: Wiley.

2. Cox, D.R. (1972). Regression models and life tables. Journal of the Royal Statistical Society, B, 34, 187-220.

3. David, H.A. and Quesenberry, C.P. (1961). Some tests for outliers. Biometrika, 48, 379-390.

4. Hampel, F. (1968). Contributions to the Theory of Robust Estimation. Unpublished Ph.D. dissertation, University of California-Berkeley. Ann Arbor: University Microfilms.

5. Huber, P.J. (1964). Robust estimation of a location parameter. Annals of Mathematical Statistics, 35, 73-101.

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

7. Kendall, M.G. and Stuart, A. (1961). The Advanced Theory of Statistics, Vol. II, New York: Hafner.

8. Paulson, A.S. and Nicklin, E.H. (1982). The form, and some robustness properties, of integrated distance estimators for linear models, applied to some published data sets. To appear in Applied Statistics.

9. Rey, W.J.J. (1977). Robust Statistical Methods. New York: Springer-Verlag.

10. Thall, P.F. (1979). Huber-sense robust M-estimations of a scale parameter, with application to the exponential distribution. Journal of the American Statistical Association, 74, 147-152.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: A-5	2. GOVT ACCESSION NO.: AD-A119724
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): SELF-CRITICAL AND ROBUST PROCEDURES FOR THE ANALYSIS OF UNIVARIATE COMPLETE DATA	5. TYPE OF REPORT & PERIOD COVERED: Interim Technical Report
	6. PERFORMING ORG. REPORT NUMBER: A-5
7. AUTHOR(s): A.S. Paulson, M.A. Presser, E.H. Nicklin	8. CONTRACT OR GRANT NUMBER(s): DAA G29-81-K-0110
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Rensselaer Polytechnic Institute, Troy, New York 12181	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
11. CONTROLLING OFFICE NAME AND ADDRESS: Approved for public release; distribution unlimited.	12. REPORT DATE: 1 June 1982
	13. NUMBER OF PAGES: 35
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Department of the Navy, Office of Naval Research, 715 Broadway (5th Floor), New York, New York 10003	15. SECURITY CLASS. (of this report):
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

U. S. Army Research Office
Post Office Box 12211
Research Triangle Park, NC 27709

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
sensitivity analysis, robust procedure, Gaussian, logistic, Wiebull, extreme value, Poisson, generalized likelihood

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
A statistical sensitivity analysis may be defined and performed in terms of the response of a vector of parameter estimates to variation in the way sample information is processed vis-a-vis the tentative underlying model. The mode of information processing is a generalization of likelihood and is indexed on a non-statistical parameter c. The case c=0 corresponds to maximum likelihood. If the vector of parameter estimates is stable under moderate increase of the index c from 0, the tentative model and the data are internally consistent. A general procedure for the conduct of such sensitivity analyses is

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract (cont'd)

given along with several illustrations. For fixed, positive values of the index, one obtains a general robust estimation procedure.