Efficient and Robust Fitting of Lognormal Distributions

Robert Serfling[1]

University of Texas at Dallas

May 2002

[1]Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688, USA. Supported by a grant from the Society of Actuaries, with administrative support from the Actuarial Education and Research Fund, and by NSF Grants DMS-9705209 and DMS-0103698.
Email: `serfling@utdallas.edu`. Website: `www.utdallas.edu/~serfling`.

Abstract

In parametric modeling of loss distributions in actuarial science, a versatile choice with intermediate tail weight is the lognormal distribution. Surprisingly, however, the fitting of this model using estimators which are at once efficient and robust has not been seriously addressed in the extensive literature. Consequently, for example, typical estimators of the lognormal mean and variance fail to be both efficient and robust. In particular, the highly efficient maximum likelihood estimators lack robustness. By robustness is meant limited sensitivity to outliers in the sample. For the two-parameter lognormal estimation problem, we consider equivalently the problem of efficient and robust joint estimation of the mean and variance of a normal model and introduce generalized median type estimators which are robust while also possessing very high efficiency compared to competitors already in the literature. These yield efficient and robust estimators of various parameters of interest in the lognormal model, and in this regard we provide detailed treatment of the lognormal mean. Extension of the approach to the much more complicated problem of estimation for the three-parameter lognormal model is also discussed.

AMS 1991 Subject Classification: Primary 62F35, Secondary 62F12.

Key words and phrases: robust estimation; efficient estimation; lognormal model; normal model; generalized median statistics.

1 Introduction and Preliminaries

In parametric modeling of loss distributions in actuarial science, a versatile choice with tail weight intermediate between that of the gamma and Pareto distributions is the *lognormal* distribution, which in its three-parameter form $L(\mu, \sigma, \tau)$ is the distribution of

$$Y = \tau + e^X,$$

where τ represents a threshold value and X is a normal random variable with mean μ and standard deviation σ. See Daykin, Pentikäinen and Pesonen (1994) for discussion and useful graphical illustrations and Klugman, Panjer and Willmot (1998) for detailed treatment including methods of fitting. Other applications arising in business and economics include modeling of firm sizes, incomes, stock prices, and lengths of service in labor turnover contexts, and the model serves many other kinds of applications as well. Complete books (Aitchison and Brown, 1957, and Crow and Shimizu, 1988) as well as Chapter 14 of Johnson, Kotz and Balakrishnan (1994) are dedicated to the theory and the diverse applications of the lognormal model.

In fitting a statistical model by estimation of parameters, two very important properties are desired of the estimators: *efficiency*, in the sense of small mean square error, and *robustness*, in the sense of low sensitivity to outliers in the data. By the term "outlier" is meant an observation sufficiently far afield from the bulk of the data that its representativeness of the underlying population becomes in question. Surprisingly, however, the goal of finding estimators which are not only efficient but also robust has not been seriously addressed in the extensive literature on the lognormal model. Here we focus on estimation of the *mean* of the lognormal distribution,

$$\eta = E\{Y\} = \tau + e^{\mu + \sigma^2/2},$$

and develop estimators of η meeting both of the above criteria.

Following Klugman, Panjer and Willmot (1998) and others, we confine attention in the present paper primarily to the *two-parameter* lognormal model corresponding to the case that the threshold parameter τ is known, as for example when τ represents a known deductible for claim amounts. (In Section 3.3, however, we briefly discuss extension of our results to the much more complicated problem of estimation for the three-parameter lognormal model.) Thus, setting $\tau = 0$ without loss of generality, we consider the lognormal model $L(\mu, \sigma)$ defined by the cdf

$$F(y) = \Phi\left(\frac{\log y - \mu}{\sigma}\right),\ 0 < y < \infty,$$

where $-\infty < \mu < \infty$, $0 < \sigma < \infty$, and Φ denotes the standard normal cdf. A random variable Y thus has the distribution $L(\mu, \sigma)$ if $X = \log Y$ has the normal distribution $N(\mu, \sigma^2)$ with mean μ and standard deviation σ.

While in principle, therefore, the fitting of a (two-parameter) lognormal model reduces simply to the fitting of a normal distribution, we will see, however, that our estimation goal central to the lognormal model corresponds in the associated normal model to a problem that has not received sufficient development. In particular, the problem of efficient and robust estimation of $\eta = e^{\mu+\sigma^2/2}$ clearly rests upon that of *simultaneously* efficient and robust *joint* estimation of μ and σ in the context of the corresponding model $N(\mu, \sigma^2)$. The latter problem has received but limited attention (see some general development in Hampel *et al.*, 1986) that does not meet present needs. Rather, treatments of the model $N(\mu, \sigma^2)$ have developed excellent efficient and robust estimators of μ but have left σ to be estimated merely consistently as a nuisance parameter. This paper extends the methodology for the normal model in a way that serves such applications as efficient and robust estimation of the lognormal mean.

Here let us clarify that, although the normal model comes into play, our focus remains on the *lognormal* model, in order to serve applications in which it is indeed the model of choice. Thus, for present purposes, the only relevant transformation is the logarithmic transformation. The Box-Cox power transformations and various other transformations, that arise in connection with the goal of exploring what kind of transformed normal model might fit a data set, are not relevant in the present context.

The efficiency criterion that we will employ is based on the performance of the maximum likelihood (ML) estimator, whose asymptotic optimality in terms of variance provides a quantitative benchmark. Thus, for a competing estimator, the asymptotic relative efficiency (ARE) is defined as the limiting ratio of sample sizes at which that estimator and the MLE perform "equivalently". Precise formulation appears in Section 1.1 below.

For robustness, two interrelated measures are used. The breakdown point (BP) of an estimator is the greatest fraction of data values that may be corrupted without the estimator becoming uninformative about the target parameter. The gross error sensitivity (GES) measures, approximately, the maximum contribution to the estimation error that can be produced by a single outlying observation, when the given estimator is used. From the discussion of the BP and GES measures in Section 1.2 below, it can be seen that as the anticipated proportion of outliers increases, suggesting the use of an estimator with high BP, it becomes of increased importance that the estimator have low GES.

Since higher BP comes at a higher price in terms of reduced ARE, however, one should choose estimators with BP no higher than actually needed. In typical situations, the range 0.05 to 0.30 for BP provides very adequate protection. An effective general approach is to set a minimum acceptable BP and a maximum acceptable GES and then maximize ARE subject to these constraints.

In this spirit, we develop estimators for μ, σ, and η which offer very high ARE along with adequately high BP and adequately low GES. Let us first examine the ML estimators as candidates. For a data set $Y_1, \ldots, Y_n$ from the model $L(\mu, \sigma)$, transformation to the equivalent model $N(\mu, \sigma^2)$ yields the well-known ML estimators of the location parameter μ and the scale parameter σ:

$$\hat{\mu}_{\mathrm{ML}} = n^{-1} \sum_{i=1}^{n} \log Y_i$$

and

$$\hat{\sigma}_{\mathrm{ML}} = \left(n^{-1} \sum_{i=1}^{n} (\log Y_i - \hat{\mu}_{\mathrm{ML}})^2 \right)^{1/2}.$$

These yield the MLE of $\eta = e^{\mu+\sigma^2/2}$: $\hat{\eta}_{\mathrm{ML}} = e^{\hat{\mu}_{\mathrm{ML}}+\hat{\sigma}^2_{\mathrm{ML}}/2}$. While the estimators $\hat{\mu}_{\mathrm{ML}}$ and $\hat{\sigma}_{\mathrm{ML}}$ each possess the favorable properties of converging to their respective parameters and having minimal asymptotic variance, they fail to be robust, each having BP = 0 and GES = ∞ (the worst cases). Such sensitivity to outliers is undesirable, and alternative estimators are desired, therefore, which give up some efficiency in return for a suitable degree of robustness. (This nonrobustness of lognormal model-based estimators is seen also, from a different perspective, in a study of Myers and Pepin, 1990, in the context of estimation of population abundance using a lognormal distribution for the nonzero observations.)

For the parameter μ in $N(\mu, \sigma^2)$, there already exist a number of robust competitors to $\hat{\mu}_{\mathrm{ML}}$ which pay relatively small prices in terms of reduced efficiency. Trimmed means, the Hodges-Lehmann estimator, M-estimators, and others are discussed in Hampel *et al.* (1986). In the present paper we use estimators of *generalized median* (GM) type (Serfling, 1984), which offer excellent trade-offs between efficiency and robustness and have other attractive properties.

For the parameter σ in $N(\mu, \sigma^2)$, there are classical robust competitors to $\hat{\sigma}_{\mathrm{ML}}$ based on the interquartile range and the median absolute deviation, but while offering very high BP (0.50) these sacrifice too much efficiency. These and some trimmed standard deviation type competitors to $\hat{\sigma}_{\mathrm{ML}}$ are improved upon, however, by estimators of Rousseeuw and Croux (1993) which also attain BP = 0.50 but give up a much smaller, though still substantial, amount of efficiency. In the present paper, giving greater emphasis to ARE while relaxing somewhat the very stringent BP = 0.50 requirement, we develop for σ new estimators of GM type which attain very high ARE with still high enough BP and low enough GES.

Formulation of the GM estimators for joint estimation of μ and σ in $N(\mu, \sigma^2)$ is carried out in Section 2, along with study of their BP, GES, and ARE performance measures. As the problem of efficient and robust fitting of normal models is very basic to statistical practice, these results are of general interest and have broad potential application. Particular application to the lognormal model is treated in Section 3, where GM competitors to $\hat{\eta}_{\mathrm{ML}}$ are obtained which have very favorable ARE as well as attractive BP and GES. The Appendix provides miscellaneous details and proofs. Following some remarks below, the remainder of the present section is devoted to formulation of the ARE, BP, and GES measures.

Remarks (i) A more comprehensive study of robustness would examine the efficiencies of the estimators within a neighborhood of the target model. Of course, then one must define what is meant by "nearby", in the sense of a suitable metric (for which there are a number of standard choices). Such an extended treatment, however, entails technical development beyond the scope of the present paper.

(ii) The BP and GES correspond to particular features of the *maxbias curve*, a more sophisticated robustness measure introduced by Martin, Yohai and Zamar (1989) (see also Ferretti, Kelmansky, Yohai and Zamar, 1999, for recent discussion and further references).

While this curve does provide somewhat more information on the robustness of the estimators than BP and GES alone, its use in general requires technical development beyond the scope of the present paper. In any case, the GM estimators presented in this paper are competitive with any estimators obtained via alternative approaches and points of view.

(iii) The abbreviation "GM" used here (and elsewhere in the literature) for "generalized median estimators" is used alternatively in some other parts of the literature to denote "generalized M-estimators". We assume that this will cause no difficulty to readers. □

1.1 Efficiency Criterion: ARE

We start with the fact (see, for example, Serfling, 1980, §2.2, Problem 2.P.8, and §3.3) that the bivariate ML estimator $(\hat{\mu}_{\mathrm{ML}}, \hat{\sigma}_{\mathrm{ML}})$ for (μ, σ) in $N(\mu, \sigma^2)$ is asymptotically bivariate normal with mean (μ, σ) and covariance matrix $n^{-1}\Sigma_0$, where

$$\Sigma_0 = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2/2 \end{bmatrix}.$$

That is,

$$n^{1/2}(\hat{\mu}_{\mathrm{ML}} - \mu,\, \hat{\sigma}_{\mathrm{ML}} - \sigma) \xrightarrow{\mathrm{d}} N((0,0), \Sigma_0),$$

as the sample size $n \to \infty$, where "$\xrightarrow{\mathrm{d}}$" denotes "converges in distribution" and $N((0,0), \Sigma_0)$ denotes bivariate normal with mean $(0,0)$ and covariance matrix Σ_0.

As discussed in Serfling (1980, §4.1), for estimation of a d-variate parameter ξ by a d-variate estimator $\hat{\xi}$ which is asymptotically d-variate normal with mean ξ and covariance matrix $n^{-1}\Sigma$, we have: confidence ellipsoids for ξ based on the estimator $\hat{\xi}$ have volume proportional to $|\Sigma|^{1/d}$. Thus the determinant $|\Sigma|$ plays in higher dimensions the role played by the variance in one dimension and is called the *generalized variance*. For two competing asymptotically d-variate normal estimators A and B with the same mean vector ξ and respective covariance matrices Σ_A and Σ_B, it follows that the ratio of respective sample sizes n_A and n_B at which the estimators perform "equivalently" (that is, have confidence ellipsoids of equal volume) approaches a limit value,

$$\frac{n_A}{n_B} \longrightarrow \left(\frac{|\Sigma_A|}{|\Sigma_B|}\right)^{1/d}, \tag{1.1}$$

as n_A and $n_B \to \infty$. This limit is then interpreted as the *asymptotic relative efficiency* (ARE) of estimator "B" with respect to estimator "A". Of course, in the case of a one-dimensional estimator we have $d = 1$ and the quantity in equation (1.1) is just the ratio of asymptotic variance parameters.

Now consider an estimator $(\hat{\mu}, \hat{\sigma})$ which like the MLE is asymptotically bivariate normal with mean (μ, σ) but with some other covariance matrix Σ_1:

$$n^{1/2}(\hat{\mu} - \mu,\, \hat{\sigma} - \sigma) \xrightarrow{\mathrm{d}} N((0,0), \Sigma_1).$$

Applying equation (1.1) with $d = 2$, we then obtain for the ARE of $(\hat{\mu}, \hat{\sigma})$ with respect to $(\hat{\mu}_{\mathrm{ML}}, \hat{\sigma}_{\mathrm{ML}})$:

$$\mathrm{ARE}\big((\hat{\mu}, \hat{\sigma}), (\hat{\mu}_{\mathrm{ML}}, \hat{\sigma}_{\mathrm{ML}})\big) = \left(\frac{|\Sigma_0|}{|\Sigma_1|}\right)^{1/2}. \tag{1.2}$$

In the special case that Σ_1, like Σ_0, is of form

$$\Sigma_1 = \begin{bmatrix} v_{11} & 0 \\ 0 & v_{22} \end{bmatrix}, \tag{1.3}$$

equation (1.2) becomes simply

$$\left(\frac{\sigma^2}{v_{11}} \times \frac{\sigma^2/2}{v_{22}}\right)^{1/2}, \tag{1.4}$$

that is,

$$\mathrm{ARE}\big((\hat{\mu}, \hat{\sigma}), (\hat{\mu}_{\mathrm{ML}}, \hat{\sigma}_{\mathrm{ML}})\big) = \big(\mathrm{ARE}(\hat{\mu}, \hat{\mu}_{\mathrm{ML}}) \times \mathrm{ARE}(\hat{\sigma}, \hat{\sigma}_{\mathrm{ML}})\big)^{1/2}. \tag{1.5}$$

Remarks (i) All choices of estimators $(\hat{\mu}, \hat{\sigma})$ that we consider here for estimation of (μ, σ) in $N(\mu, \sigma^2)$ will have asymptotic covariance matrices of the form in equation (1.3) and hence will satisfy equation (1.5). This is because each estimator $\hat{\mu}$ considered will be both *odd*,

$$\hat{\mu}(-X_1, \ldots, -X_n) = -\hat{\mu}(X_1, \ldots, X_n),$$

and *translation equivariant*,

$$\hat{\mu}(X_1 + c, \ldots, X_n + c) = \hat{\mu}(X_1, \ldots, X_n) + c, \quad \text{for any } c,$$

while each estimator $\hat{\sigma}$ considered will be both *even*,

$$\hat{\sigma}(-X_1, \ldots, -X_n) = \hat{\sigma}(X_1, \ldots, X_n),$$

and *translation invariant*,

$$\hat{\sigma}(X_1 + c, \ldots, X_n + c) = \hat{\sigma}(X_1, \ldots, X_n), \quad \text{for any } c.$$

As seen in Randles and Wolfe (1979, Corollary 1.3.33), in the case of data from a symmetric distribution, any odd translation equivariant statistic and any even translation invariant statistic are *uncorrelated*. Thus, throughout, we will have $\mathrm{Cov}\{\hat{\mu}, \hat{\sigma}\} = 0$.

(ii) Further, each estimator $\hat{\mu}$ considered will be *scale equivariant*,

$$\hat{\mu}(X_1/c, \ldots, X_n/c) = \hat{\mu}(X_1, \ldots, X_n)/c, \quad \text{for any } c > 0.$$

This property, together with translation equivariance, yields that v_{11} in Σ_1 must be of form $c_{11}\sigma^2$, where c_{11} is the value of v_{11} obtained in the case of standard normal data. In this case the ARE does not depend on μ or σ:

$$\mathrm{ARE}(\hat{\mu}, \hat{\mu}_{\mathrm{ML}}) = \frac{1}{c_{11}}.$$

Likewise, each estimator $\hat{\sigma}$ considered also will be *scale equivariant*, which together with translation invariance yields that v_{22} in Σ_1 must be of form $c_{22}\sigma^2$, where c_{22} is the value of v_{22} obtained in the case of standard normal data. Consequently, again the ARE does not depend on μ or σ:

$$\mathrm{ARE}(\hat{\sigma}, \hat{\sigma}_{\mathrm{ML}}) = \frac{1}{2\, c_{22}}.$$

With these simplifications, equations (1.4) and (1.5) reduce to

$$\mathrm{ARE}\big((\hat{\mu}, \hat{\sigma}), (\hat{\mu}_{\mathrm{ML}}, \hat{\sigma}_{\mathrm{ML}})\big) = \left(\frac{1}{c_{11}} \times \frac{1}{2\, c_{22}}\right)^{1/2}. \tag{1.6}$$

(iii) Such simplicity fails, however, to hold for estimators of the parameter $\eta = e^{\mu+\sigma^2/2}$. In particular, for $\hat{\eta} = e^{\hat{\mu}+\hat{\sigma}^2/2}$, we have that $\log \hat{\eta}$ is translation equivariant but not scale equivariant. Consequently, as we will find in Section 3, the quantity $\mathrm{ARE}(\hat{\eta}, \hat{\eta}_{\mathrm{ML}})$ turns out to be a function of σ. In this case the different choices of estimator $\hat{\eta}$ are compared with respect to the ARE criterion by comparing respective tables or plots of their ARE versus σ over a range of σ values.
$\square$

1.2 Robustness Criteria: BP and GES

1.2.1 Breakdown Point (BP)

A popular and effective criterion for robustness of an estimator is its *breakdown point* (BP), loosely characterized as the largest proportion of sample observations which themselves may be corrupted without the estimator itself becoming corrupted beyond use. When the BP is well-defined as a quantity not depending on the particular sample values but only on the sample size n, then we typically take as our criterion its limit value as $n \to \infty$. The BP of an estimator measures the degree to which the estimator remains uninfluenced by the presence of outliers. We thus define:

> *Breakdown Point*: the largest proportion of sample observations which may be given arbitrary values without taking the estimator to a limit uninformative about the parameter being estimated.

In particular, for the *location* parameter μ in $N(\mu, \sigma^2)$, we define $\mathrm{BP}(\hat{\mu})$ to be the largest proportion of observations which may be given arbitrary values without taking $\hat{\mu}$ to $\pm\infty$. For the *scale* parameter σ, we define $\mathrm{BP}(\hat{\sigma})$ to be the largest proportion of observations which may be given arbitrary values without taking $\hat{\sigma}$ to either 0 or ∞.

It is readily seen that the estimators $\hat{\mu}_{\mathrm{ML}}$ and $\hat{\sigma}_{\mathrm{ML}}$ each have $\mathrm{BP} = 0$ and thus are *nonrobust* in this sense. Clearly, estimators are desired which have *nonzero* breakdown points while possessing relatively high efficiency.

1.2.2 Gross Error Sensitivity (GES)

Associated with any estimator $\hat{\xi}$ of a parameter $\xi(G)$ associated with a distribution G is its *influence function* (IF), defined by

$$\mathrm{IF}(x) = \lim_{\lambda \downarrow 0} \frac{\xi((1-\lambda)G + \lambda\delta_x) - \xi(G)}{\lambda},$$

where δ_x denotes the distribution placing all mass at the point x. As the directional derivative of ξ at G in the direction of δ_x, $\mathrm{IF}(x)$ approximates the contribution to the total estimation error that is made by an observation located at x. It follows that for the estimator $\hat{\xi}$ based on a sample $X_1, \ldots, X_n$ from G, a first order approximation to the estimation error is given in terms of the IF:

$$\hat{\xi} - \xi \;\doteq\; n^{-1} \sum_{i=1}^{n} IF(X_i).$$

See Hampel *et al.* (1986) for a full treatment of the IF and its general role in statistical inference, and see Marceau and Rioux (2001) for a nice exposition of the IF and its application in the context of estimators of excess of loss premiums and other quantities arising in risk theory.

The quantity $sup_x|\mathrm{IF}(x)|$ is called the *Gross Error Sensitivity* (GES). Thus the maximum possible impact that any single outlier can produce on the estimation error for a sample of size n is measured, approximately, by

$$n^{-1}\,\mathrm{GES}.$$

Clearly, estimators with relatively low GES are desired. Indeed, as a minimal requirement for robustness, the IF should be *bounded.* It is also clearly desirable that the IF be *smooth.*

In particular, for estimation of μ and σ in $N(\mu, \sigma^2)$, the estimator $\hat{\mu}_{\mathrm{ML}}$ has IF

$$x - \mu,$$

and the estimator $\hat{\sigma}_{\mathrm{ML}}$ has IF

$$\frac{(x-\mu)^2 - \sigma^2}{2\sigma},$$

yielding in each case GES $= \infty$ (*nonrobustness*).

2 Generalized Median Estimators for μ and σ in $N(\mu, \sigma^2)$

2.1 Basic Formulation of Generalized Median Estimators

Generalized median (GM) estimators fall within the class of "generalized L-estimators" which were introduced and investigated in Serfling (1984). In general, for estimation of a parameter ξ on the basis of a sample $Y_1, \ldots, Y_n$, GM estimators are defined as follows. For a given choice of integer J, a "kernel" $h(y_1, \ldots, y_J)$ is selected such that $h(Y_1, \ldots, Y_J)$ is *median unbiased* for ξ, that is, the median of the distribution of $h(Y_1, \ldots, Y_J)$ is ξ. In the present paper we confine attention to kernels which are invariant under permutations of their arguments. In this case the corresponding GM estimator is given by taking the median of the evaluations $h(Y_{i_1}, \ldots, Y_{i_J})$ of the kernel h over all subsets of observations taken J at a time, that is, corresponding to all $\binom{n}{J}$ subsets $\{i_1, \ldots, i_J\}$ of distinct indices from $\{1, \ldots, n\}$. This yields for ξ the estimator

$$\hat{\xi}_{\text{GM}} = \text{Median}\{h(Y_{i_1}, \ldots, Y_{i_J})\}.$$

Different choices of J and kernel h lead to different GM estimators for ξ.

Letting H denote the cdf of a kernel evaluation $h(Y_1, \ldots, Y_J)$ and assuming differentiability of H at ξ, it follows readiy from Serfling (1984) that the estimator $\hat{\xi}_{\text{GM}}$ has IF

$$\frac{J}{H'(\xi)} \left(\frac{1}{2} - w(y) \right), \tag{2.1}$$

where

$$w(y) = P\{h(y, Y_1, \ldots, Y_{J-1}) \leq \xi\}. \tag{2.2}$$

Thus the IF of a GM estimator is bounded. For many typical h, the function $w(\cdot)$ is smooth, thus giving the IF as well this additional favorable property.

Since $0 \leq w(y) \leq 1$ must hold, an upper bound for the GES is given by

$$\text{GES} \leq \frac{J}{2\, H'(\xi)}.$$

Typically, either $\inf_y w(y) = 0$ or $\sup_y w(y) = 1$ (or both), in which case the GES actually equals this upper bound.

Further, the function $w(\cdot)$ is instrumental in obtaining the asymptotic distribution of the GM estimator. We have from Serfling (1984) that $\hat{\xi}_{\text{GM}}$ *is asymptotically normal with mean* ξ *and variance*

$$\frac{J^2\, \text{Var}\{w(Y)\}}{[H'(\xi)]^2}\, n^{-1},$$

as $n \to \infty$.

2.2 GM Estimators for μ in $N(\mu, \sigma^2)$

We now specialize the GM approach to estimate μ in $N(\mu, \sigma^2)$, on the basis of a random sample $X_1, \ldots, X_n$. For any fixed integer $k \geq 1$ not depending on the sample size n, we introduce the kernel

$$h_1(x_1, \ldots, x_k) = k^{-1} \sum_{i=1}^{k} x_i.$$

Clearly, $h_1(X_1, \ldots, X_k)$ is median unbiased for μ. The particular choice of kernel h_1 is motivated by the fact that each evaluation $h_1(X_{i_1}, \ldots, X_{i_k})$ is the MLE of μ based on just the observations $X_{i_1}, \ldots, X_{i_k}$. Denote the corresponding GM estimator by $\hat{\mu}_{(k)}$. Then $\hat{\mu}_{(1)}$ is just the ordinary median of the data, and $\hat{\mu}_{(2)}$ is the well-known Hodges-Lehmann estimator. For general choice of k, the estimator $\hat{\mu}_{(k)}$ was introduced in Serfling (1984) as a particular example of generalized L-statistic and has been further studied in Choudhury and Serfling (1988), Choudhury (1990), Chaudhuri (1992), Ambühl (2000), and Serfling (2000).

It is not difficult to obtain (see Appendix A.1) that $\hat{\mu}_{(k)}$ has asymptotic breakdown point

$$\mathrm{BP}(\hat{\mu}_{(k)}) = 1 - (1/2)^{1/k}$$

as $n \to \infty$. The random variable $h_1(X_1, \ldots, X_k)$ is found to have distribution $H_1 = N(\mu, \sigma^2/k)$, yielding

$$H_1'(\mu) = \sqrt{\frac{k}{2\pi}}\,\sigma^{-1}.$$

Also, it is readily derived that for this kernel the function in (2.2) is given by

$$w(x) = \Phi\left(\frac{\mu - x}{\sqrt{k-1}\,\sigma}\right).$$

Thus $\hat{\mu}_{(k)}$ has a smooth and bounded IF.

Noting that $w(x) \to 1$ or 0 as $x \to -\infty$ or $+\infty$, respectively, we obtain from equation (2.1) the gross error sensitivity,

$$\mathrm{GES}(\hat{\mu}_{(k)}) = \sqrt{\frac{\pi}{2}}\,\sqrt{k}\,\sigma = 1.2533\,\sqrt{k}\,\sigma.$$

In order to eliminate dependence on σ, we use a *standardized* version, $\mathrm{GES}^* = \mathrm{GES}/\sigma$.

Also (Chaudhuri, 1992), we have $\mathrm{Var}\{w(X)\} = (2\pi)^{-1}\sin^{-1}(1/k)$, and it follows that $\hat{\mu}_{(k)}$ is asymptotically normal with mean μ and variance $c_{11k}\,\sigma^2\,n^{-1}$, as $n \to \infty$ with k fixed, where

$$c_{11k} = k \sin^{-1}(1/k).$$

We thus arrive at

$$\mathrm{ARE}(\hat{\mu}_{(k)}, \hat{\mu}_{\mathrm{ML}}) = \frac{1}{c_{11k}}.$$

Evaluations of BP, GES*, ARE, and c_{11} for selected k are provided in Table 1.

Table 1

$\mathrm{BP}(\hat{\mu}_{(k)})$, $\mathrm{GES}^*(\hat{\mu}_{(k)})$, $\mathrm{ARE}(\hat{\mu}_{(k)}, \hat{\mu}_{\mathrm{ML}})$, and c_{11k}, for $k = 1:9$

	k								
	1	2	3	4	5	6	7	8	9
BP	0.500	0.293	0.206	0.159	0.129	0.109	0.094	0.083	0.074
GES*	1.253	1.772	2.171	2.507	2.802	3.070	3.316	3.545	3.760
ARE	0.637	0.955	0.981	0.989	0.993	0.995	0.997	0.997	0.998
c_{11k}	1.571	1.047	1.020	1.011	1.007	1.005	1.003	1.003	1.002

Remarks (i) While $\hat{\mu}_{(1)}$ (the median) has very high BP and very low GES, its ARE of 0.64 is unacceptably low. For $k = 2, \ldots, 9$, however, the estimators $\hat{\mu}_{(k)}$ provide a spectrum of favorable choices, trading off BP and GES by degrees in return for improved ARE.

(ii) Somewhat competitive estimators are provided by trimmed means, which tend to favor GES more strongly at a greater sacrifice of BP or ARE. For example (see Hampel, 1974), the 10% trimmed mean has BP = 0.10, GES = 1.60, and ARE = 0.943. For comparison, $\hat{\mu}_{(2)}$ (the Hodges-Lehmann estimator) has slightly worse GES but much better BP and slightly higher ARE of 0.955, and $\hat{\mu}_{(6)}$ has much worse GES but comparable BP and much better ARE.

(iii) As another type of competitor, one might also consider a particular M-estimator such as the Huber Proposal 2, H(1.3), for which (Huber, 1981, p. 144) the BP is 0.29 in agreement with $\hat{\mu}_{(2)}$, but whose ARE corresponds to that of the 10% trimmed mean at 0.943.

Thus GM estimators tend to offer somewhat more favorable trade-offs between ARE and BP than competing estimators, although it should be recognized that these are not the only criteria that we might choose for efficiency and robustness. For the purpose of estimating μ in the framework of the lognormal target problem considered in this paper, one could reasonably take a trimmed mean or an M-estimator instead of $\hat{\mu}_{(k)}$, without drastic change in the results.

□

2.3 GM Estimators for σ in $N(\mu, \sigma^2)$

For GM type estimation of σ in $N(\mu, \sigma^2)$, it is convenient first to develop estimators for σ^2 and then to take square roots.

2.3.1 GM Estimators for σ^2

We again utilize a kernel based on the method of maximum likelihood. For fixed integer $m \geq 2$ not depending on n, we use

$$\tilde{h}_2(x_1, \ldots, x_m) = m^{-1} \sum_{i=1}^{m} \left(x_i - m^{-1} \sum_{j=1}^{m} x_j \right)^2,$$

the "maximum likelihood kernel" for estimation of σ^2 on the basis of just m observations. It is readily seen that $m\,\tilde{h}_2(X_1, \ldots, X_m)/\sigma^2$ has cdf G_{m-1}, where G_ν denotes the chi-square distribution with ν degrees of freedom. With M_ν denoting the median of G_ν, we define

$$h_2(x_1, \ldots, x_m) = \frac{m}{M_{m-1}} \tilde{h}_2(x_1, \ldots, x_m)$$

and thus have that $h_2(X_1, \ldots, X_m)$ is median unbiased for σ^2. An alternative expression for h_2 is

$$h_2(x_1, \ldots, x_m) = \frac{1}{m\, M_{m-1}} \sum_{1 \leq i < j \leq m} (x_i - x_j)^2.$$

Denote the corresponding GM estimator by $\hat{\sigma}^2_{(m)}$. For $m = 2$ this reduces to

$$\frac{1}{2M_1}\,\mathrm{Median}_{1\le i<j\le n}\{(X_i - X_j)^2\},$$

an estimator formulated by Shamos (1976) and Bickel and Lehmann (1979) and studied in detail by Rousseeuw and Croux (1993). For $m \ge 3$, however, $\hat{\sigma}^2_{(m)}$ has not been investigated previously in the literature.

The asymptotic breakdown point for fixed m as $n \to \infty$ is found (Appendix A.1) to be

$$\mathrm{BP}(\hat{\sigma}^2_{(m)}) = 1 - (1/2)^{1/m}.$$

The random variable $h_2(X_1, \ldots, X_m)$ has cdf $H_2(z) = G_{m-1}(M_{m-1}z/\sigma^2)$, $z > 0$, yielding

$$H_2'(\sigma^2) = C_m\,\sigma^{-2},$$

where

$$C_m = \left(\frac{M_{m-1}}{2}\right)^{\frac{m-1}{2}}\ \frac{e^{-\frac{M_{m-1}}{2}}}{\Gamma(\frac{m-1}{2})},$$

with $\Gamma(\cdot)$ denoting the gamma function. Values of M_m and C_m for selected m are provided in Appendix A.2.

It is not difficult to see that the function in (2.2) is given by $w(x) = w_0((x-\mu)/\sigma)$, where

$$\begin{aligned} w_0(z) &= P\{h_2(z, Z_1, \ldots, Z_{m-1}) \le 1\} \\ &= P\left\{\sum_{i=1}^{m-1}(z - Z_i)^2 + \sum_{1\le i<j\le m-1}(Z_i - Z_j)^2 \le m\,M_{m-1}\right\}, \end{aligned}$$

with $Z_1, \ldots, Z_{m-1}$ independent standard normal random variables. Thus $\hat{\sigma}^2_{(m)}$ has a smooth and bounded IF.

Note that $w_0(z) \to 0$ as $z \to \pm\infty$. We then obtain from equation (2.1) the gross error sensitivity,

$$\mathrm{GES}(\hat{\sigma}^2_{(m)}) = \frac{m}{2\,C_m}\ \sigma^2.$$

In order to eliminate the dependence on σ^2, we use $\mathrm{GES}^* = \mathrm{GES}/\sigma^2$.

Also, $\mathrm{Var}\{w(X)\} = \mathrm{Var}\{w_0(Z)\}$ with Z standard normal. With this quantity denoted by ζ_m, we have that $\hat{\sigma}^2_{(m)}$ is asymptotically normal with mean σ^2 and variance $\tilde{c}_{22m}\,\sigma^4\,n^{-1}$, where

$$\tilde{c}_{22m} = \frac{m^2\zeta_m}{C_m^2}.$$

Values of ζ_m are provided in Appendix A.2. Also (Appendix A.3) the maximum likelihood estimator $\hat{\sigma}^2_{\mathrm{ML}}$ is asymptotically normal with mean σ^2 and variance $2\,\sigma^4\,n^{-1}$. We thus arrive at

$$\mathrm{ARE}(\hat{\sigma}^2_{(m)}, \hat{\sigma}^2_{\mathrm{ML}}) = \frac{2}{\tilde{c}_{22m}}.$$

Evaluations of BP, GES*, ARE, and $\tilde{c}_{22}$ for selected m are provided in Table 2.

Table 2

BP($\hat{\sigma}^2_{(m)}$), GES*($\hat{\sigma}^2_{(m)}$), ARE($\hat{\sigma}^2_{(m)}, \hat{\sigma}^2_{\mathrm{ML}}$), and $\tilde{c}_{22m}$, for $m = 2, 3, 5, 7$, and 9

	m				
	2	3	5	7	9
BP	0.293	0.206	0.129	0.094	0.074
GES*	4.666	4.328	4.754	5.308	5.841
ARE	0.864	0.862	0.910	0.940	0.956
$\tilde{c}_{22m}$	2.314	2.320	2.198	2.128	2.091

2.3.2 GM Estimators for σ

For the parameter σ, the GM and ML estimators are obtained by simply taking square roots of corresponding estimators for σ^2. Breakdown points clearly remain unchanged, but (Appendix A.4) the GES values change:

$$\mathrm{GES}(\hat{\sigma}_{(m)}) = \frac{1}{2\,\sigma}\,\mathrm{GES}(\hat{\sigma}^2_{(m)}) = \frac{1}{2}\,\mathrm{GES}^*(\hat{\sigma}^2_{(m)})\,\sigma.$$

Standardizing to $\mathrm{GES}^*(\hat{\sigma}_{(m)}) = \mathrm{GES}(\hat{\sigma}_{(m)})/\sigma$, we thus have

$$\mathrm{GES}^*(\hat{\sigma}_{(m)}) = \frac{1}{2}\,\mathrm{GES}^*(\hat{\sigma}^2_{(m)}).$$

Also (Appendix A.3), the asymptotic variance of $\hat{\sigma}_{(m)}$ is that of $\hat{\sigma}^2_{(m)}$ times the factor $1/4\sigma^2$. Thus $\hat{\sigma}_{\mathrm{ML}}$ and $\hat{\sigma}_{(m)}$ are each asymptotically normal with mean σ and respective asymptotic variances $0.5\,\sigma^2\,n^{-1}$ and $c_{22m}\,\sigma^2\,n^{-1}$, where

$$c_{22m} = \frac{\tilde{c}_{22m}}{4}.$$

Consequently, the ARE remains the same:

$$\mathrm{ARE}(\hat{\sigma}_{(m)}, \hat{\sigma}_{\mathrm{ML}}) = \mathrm{ARE}(\hat{\sigma}^2_{(m)}, \hat{\sigma}^2_{\mathrm{ML}}).$$

The corresponding analogue of Table 2 is provided in Table 3.

Table 3

BP($\hat{\sigma}_{(m)}$), GES*($\hat{\sigma}_{(m)}$), ARE($\hat{\sigma}_{(m)}, \hat{\sigma}_{\mathrm{ML}}$), and c_{22m}, for $m = 2, 3, 5, 7$, and 9

	m				
	2	3	5	7	9
BP	0.293	0.206	0.129	0.094	0.074
GES*	2.333	2.164	2.377	2.654	2.920
ARE	0.864	0.862	0.910	0.940	0.956
c_{22m}	0.579	0.580	0.549	0.532	0.523

Remarks (i) Well-known robust competitors to $\hat{\sigma}_{\text{ML}}$ given by suitably normalized versions of the interquartile range and the median absolute deviation have very favorable BP's of 0.25 and 0.50, respectively, and in common a very favorable GES of 1.27. Unfortunately, however, these estimators sacrifice too much efficiency, having in common ARE only 0.37 with respect to $\hat{\sigma}_{\text{ML}}$. Considerably higher ARE is achieved by trimmed versions of $\hat{\sigma}_{\text{ML}}$, at the cost of somewhat lower BP. For example, the 10% trimmed standard deviation has BP = 0.10 and ARE = .78 (see Bickel and Lehmann, 1976, and Janssen, Serfling and Veraverbeke, 1987). In turn, these estimators are substantially improved with respect to BP, while at the same time slightly improving ARE, by estimators introduced by Rousseeuw and Croux (1993), of which one is discussed in (ii) below. Alternatively, the GM estimators $\hat{\sigma}_{(m)}$ for m = 2, 3, 5, 7, and 9 provide a spectrum of favorable choices attaining much higher ARE by trading off BP and GES to some extent.

(ii) With respect to the estimator $\hat{\sigma}_{(2)}$, Rousseeuw and Croux (1993) consider modifications which achieve high BP = 0.50 with only a moderately small sacrifice of ARE. Specifically, they replace the median of the $\binom{n}{2}$ pairwise interpoint differences $|X_i - X_j|$ by the j_nth order statistic, where $j_n = \binom{n}{2}/4$, and they alter the constant factor to 2.222. In comparison with $\hat{\sigma}_{(2)}$, the resulting estimator has optimal BP and relatively low GES*, but at the costs of reduction in ARE and increase in small sample bias. We nevertheless include this estimator in our later discussions and denote it by $\hat{\sigma}_{(1)}$. Augmenting the information in Table 3, Table 4 provides similar information for $\hat{\sigma}_{(1)}$:

Table 4

BP($\hat{\sigma}_{(1)}$), GES*($\hat{\sigma}_{(1)}$), ARE($\hat{\sigma}_{(1)}, \hat{\sigma}_{\text{ML}}$), and c_{221}

BP	GES*	ARE	c_{221}
0.500	2.069	0.823	0.610

(iii) If desired, the estimators $\hat{\sigma}_{(m)}$ for $m \geq 3$ could also be modified in the above vein to improve BP by replacing the median by another of the ordered kernel evaluations and changing the multiplicative constant. We prefer, however, to retain the median because of its simplicity and intuitive appeal, and because it yields higher ARE and lower small sample bias.
□

2.4 Joint Estimation of μ and σ

Let us now consider *joint* estimation of (μ, σ) and compare the estimator $(\hat{\mu}_{(k)}, \hat{\sigma}_{(m)})$ with the estimator $(\hat{\mu}_{\text{ML}}, \hat{\sigma}_{\text{ML}})$, for various choices of the pair (k, m). As discussed in Section 1.1, we may use equation (1.6) for the ARE. In conjunction with the values in Tables 1, 3, and 4, this yields the ARE values in Table 5 for selected pairs (k, m).

Table 5

ARE$\big((\hat{\mu}_{(k)}, \hat{\sigma}_{(m)}), (\hat{\mu}_{\rm ML}, \hat{\sigma}_{\rm ML})\big)$ for k = 1, 2, 3, 5, 7, and 9, and m = 1, 2, 5, 7, and 9

	m				
k	1	2	5	7	9
1	0.724	0.742	0.761	0.774	0.780
2	0.887	0.908	0.932	0.947	0.955
3	0.899	0.921	0.945	0.960	0.968
5	0.904	0.926	0.951	0.966	0.974
7	0.906	0.928	0.953	0.968	0.976
9	0.906	0.929	0.953	0.969	0.977

While the most favorable choice in Table 5 is $(k, m) = (9, 9)$, with ARE = 0.98, we must also take into account the corresponding BP value, 0.07, from Tables 1 and 3. If somewhat higher BP values are needed, at the cost of a small reduction in ARE, the choice $(k, m) = (5, 5)$ offers ARE = 0.95 with BP = 0.13. If considerably higher BP values are needed, the choice $(k, m) = (2, 2)$ offers BP = 0.29, but the ARE slips to 0.91. The choice $(k, m) = (1, 1)$ offers the best BP, 0.50, but the ARE drops sharply to 0.72.

We note also that GES* values worsen as ARE improves. With this in mind, the choice $(k, m) = (5, 5)$ offers very good overall balance with respect to the factors BP, GES and ARE. There are also considerations of computational burden, however, and we discuss these in Section 3.1.3.

3 Application to Lognormal Models

For the two-parameter lognormal model, estimation of the mean is treated in Section 3.1 and of other target parameters in Section 3.2. Extension to the three-parameter lognormal model is treated briefly in Section 3.3.

3.1 Estimation of the Mean

We now address the problem of efficient and robust estimation of the mean

$$\eta = e^{\mu + \sigma^2/2}$$

of the lognormal distribution $L(\mu, \sigma)$, on the basis of a sample $Y_1, \ldots, Y_n$. The ML estimator is given by

$$\hat{\eta}_{\rm ML} = e^{\hat{\mu}_{\rm ML} + \hat{\sigma}^2_{\rm ML}/2},$$

with $\hat{\mu}_{\rm ML}$ and $\hat{\sigma}_{\rm ML}$ as in Section 1 based on the transformed observations $X_i = \log Y_i$, $1 \le i \le n$, which have the $N(\mu, \sigma^2)$ distribution. Although efficient, this estimator of η inherits the nonrobustness of $\hat{\mu}_{\text{ML}}$ and $\hat{\sigma}_{\text{ML}}$ and their BP values of 0 and GES values of ∞. Therefore, utilizing the development in Section 2, we consider the competing estimators

$$\hat{\eta}_{(k,\,m)} = e^{\hat{\mu}_{(k)}+\hat{\sigma}^2_{(m)}/2}$$

and examine their BP, GES (Section 3.1.1) and ARE (Section 3.1.2). Summary discussion is provided in Section 3.1.3.

3.1.1 BP and GES

For the BP we have in general

$$\text{BP}(\hat{\eta}_{(k,\,m)}) = \min\{\text{BP}(\hat{\mu}_{(k)}),\ \text{BP}(\hat{\sigma}_{(m)})\}.$$

In particular, the estimators $\hat{\eta}_{(1,\,1)}$, $\hat{\eta}_{(2,\,2)}$, $\hat{\eta}_{(5,\,5)}$, and $\hat{\eta}_{(9,\,9)}$ have BP's of 0.50, 0.29, 0.13, and 0.07, respectively.

For the GES we have (Appendix A.4)

$$\text{GES}(\hat{\eta}_{(k,\,m)}) = \eta\,(\text{GES}(\hat{\mu}_{(k)}) + \text{GES}(\hat{\sigma}_{(m)})\,\sigma).$$

In this case, standardizing does not completely eliminate the dependence on parameters. A partially standardized version, however, $\text{GES}^*(\hat{\eta}_{(k,\,m)}) = \text{GES}(\hat{\eta}_{(k,\,m)})/\eta\sigma(1+\sigma)$, satisfies

$$\text{GES}^*(\hat{\eta}_{(k,\,m)}) = \frac{1}{1+\sigma}\text{GES}^*(\hat{\mu}_{(k)}) + \frac{\sigma}{1+\sigma}\text{GES}^*(\hat{\sigma}_{(m)}).$$

This quantity has limit $\text{GES}^*(\hat{\mu}_{(k)})$ as $\sigma \to 0$ and limit $\text{GES}^*(\hat{\sigma}_{(m)})$ as $\sigma \to \infty$. For the estimators $\hat{\eta}_{(1,\,1)}$, $\hat{\eta}_{(2,\,2)}$, $\hat{\eta}_{(5,\,5)}$, and $\hat{\eta}_{(9,\,9)}$, in particular, these pairs of limits are $(1.253, 2.069)$, $(1.772, 2.333)$, $(2.802, 2.377)$, and $(3.760, 2.920)$, respectively.

3.1.2 ARE

By standard results on transformations of asymptotically normal random variables, it follows (Appendix A.3) that if joint estimators $(\hat{\mu}, \hat{\sigma})$ of (μ, σ) satisfy

$$n^{1/2}(\hat{\mu} - \mu,\, \hat{\sigma} - \sigma) \xrightarrow{\text{d}} N((0,0), [\sigma_{ij}]_{2\times 2})$$

as $n \to \infty$, then the corresponding estimator $\hat{\eta} = e^{\hat{\mu}+\hat{\sigma}^2/2}$ satisfies

$$n^{1/2}(\hat{\eta} - \eta) \xrightarrow{\text{d}} N(0, \eta^2(\sigma_{11} + 2\sigma\sigma_{12} + \sigma^2\sigma_{22}))$$

as $n \to \infty$. For all estimators $(\hat{\mu}, \hat{\sigma})$ under consideration, we have

$$\sigma_{11} = c_{11}\sigma^2, \quad \sigma_{12} = 0, \quad \sigma_{22} = c_{22}\sigma^2$$

for numerical constants c_{11} and c_{22}, yielding

$$n^{1/2}(\hat{\eta} - \eta) \xrightarrow{\text{d}} N(0, \eta^2\sigma^2(c_{11} + c_{22}\sigma^2))$$

as $n \to \infty$. For the ML estimators of μ and σ we have $c_{11} = 1$ and $c_{22} = 0.5$. Thus, for any estimator $\hat{\eta}$ satisfying the above conditions, the ARE is given by

$$\text{ARE}(\hat{\eta}, \hat{\eta}_{\text{ML}}) = \frac{1 + 0.5\sigma^2}{c_{11} + c_{22}\sigma^2}.$$

Note that this ARE converges to $\text{ARE}(\hat{\mu}, \hat{\mu}_{\text{ML}})$ as $\sigma \to 0$ and to $\text{ARE}(\hat{\sigma}, \hat{\sigma}_{\text{ML}})$ as $\sigma \to \infty$. In particular, for $\hat{\eta}_{(k,\,m)}$ we have

$$\text{ARE}(\hat{\eta}_{(k,\,m)}, \hat{\eta}_{\text{ML}}) = \frac{1 + 0.5\sigma^2}{c_{11k} + c_{22m}\sigma^2}.$$

Thus, utilizing Tables 1, 3 and 4, we obtain

$$\text{ARE}(\hat{\eta}_{(1,\,1)}, \hat{\eta}_{\text{ML}}) = \frac{1 + 0.5\sigma^2}{1.57 + 0.61\sigma^2},$$

which increases from 0.637 at $\sigma = 0$ to 0.820 as $\sigma \to \infty$,

$$\text{ARE}(\hat{\eta}_{(2,\,2)}, \hat{\eta}_{\text{ML}}) = \frac{1 + 0.5\sigma^2}{1.047 + 0.579\sigma^2}$$

which decreases from 0.955 to 0.864,

$$\text{ARE}(\hat{\eta}_{(5,\,5)}, \hat{\eta}_{\text{ML}}) = \frac{1 + 0.5\sigma^2}{1.007 + 0.549\sigma^2},$$

which decreases from 0.993 to 0.911, and

$$\text{ARE}(\hat{\eta}_{(9,\,9)}, \hat{\eta}_{\text{ML}}) = \frac{1 + 0.5\sigma^2}{1.002 + 0.523\sigma^2}$$

which decreases from 0.998 to 0.956. It is seen that the best ARE is attained as $\sigma \to 0$ for $k = m = 1$ but as $\sigma \to \infty$ in the other cases. A comparison of the above formulas shows that this is due to the considerable inefficiency of the ordinary median $\hat{\mu}_{(1)}$ in comparison with the estimators $\hat{\mu}_{(k)}$ for $k \geq 2$.

Table 6 exhibits for selected values of σ the ARE values for these four estimators.

Table 6

$\text{ARE}(\hat{\eta}_{(j,\,j)}, \hat{\eta}_{\text{ML}})$ for $j = 1$, 2, 5, and 9, and for selected σ

	σ						
	0	2.5	5.0	7.5	10.0	20.0	∞
$(1,1)$	0.637	0.766	0.803	0.812	0.815	0.819	0.820
$(2,2)$	0.955	0.884	0.870	0.866	0.865	0.864	0.864
$(5,5)$	0.993	0.929	0.916	0.913	0.912	0.911	0.911
$(9,9)$	0.998	0.966	0.959	0.957	0.957	0.956	0.956

3.1.3 Summary Discussion

A good overall estimator appears to be $\hat{\eta}_{(5,5)}$, which offers quite high ARE above 0.91 uniformly over σ, combined with favorable BP of 0.13 and acceptable standardized GES* within the range 2.4 to 2.8. The estimator $\hat{\eta}_{(9,9)}$, however, which offers ARE above 0.96 uniformly over σ, is more attractive if lower BP of 0.07 and higher GES* in the range 2.9 to 3.8 can be tolerated.

On the other hand, it should be noted that the estimators $\hat{\eta}_{(k,m)}$ become increasingly computationally intensive as k or m increase. If computational burden is a consideration, then the estimator $\hat{\eta}_{(2,2)}$ becomes attractive. It provides excellent BP of 0.29, GES* in the range 1.8 to 2.3, and ARE above 0.86 uniformly over σ. Alternatively, the estimators $\hat{\eta}_{(k,m)}$ may be modified to eliminate the computational intensiveness, as discussed in Appendix A.5.

A quick illustration of the robustness of the GM estimators over the MLE's is provided by the following simple experiment. A random sample of size 100 from $N(5,1)$ was taken, and the largest observation, already an "outlier", was increased in value to become a more extreme outlier, from 8.58 to 11.0. Standard boxplots of the original and modified samples are displayed in Figure 1, and the corresponding ML and GM$_{(2,2)}$ estimates of μ, σ, and η are listed in Table 7.

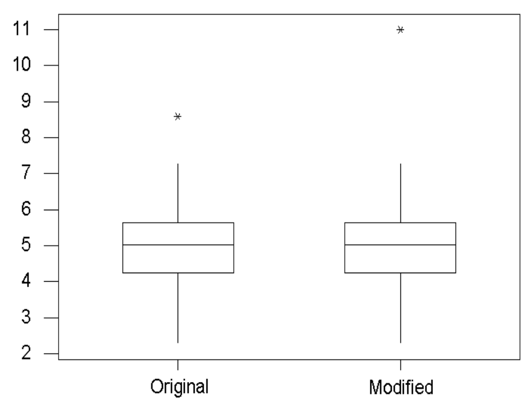

Figure 1

Table 7

ML and $GM_{(2,\,2)}$ estimates for original and modified samples

Estimator	Original Sample	Modified Sample
$\hat{\mu}_{\text{ML}}$	4.98	5.00
$\hat{\sigma}_{\text{ML}}$	1.04	1.14
$\hat{\eta}_{\text{ML}}$	249.6	282.9
$\hat{\mu}_{(2,\,2)}$	5.00	5.00
$\hat{\sigma}_{(2,\,2)}$	1.08	1.08
$\hat{\eta}_{(2,\,2)}$	266.0	266.0

For $L(5,1)$ we have $\eta = 244.7$. We see that the presence of the outlier in the original sample results in the MLE slightly overestimating η, with value 249.6. The less efficient estimator $\hat{\eta}_{(2,\,2)}$ overestimates by a greater amount, with value 266.0. A rather moderate modification of the value of the single outlier in the original sample, however, influences a dramatic change in the value of $\hat{\eta}_{\text{ML}}$, from 249.6 to 282.9. On the other hand, the robust estimator $\hat{\eta}_{(2,\,2)}$ remains quite stable with value unchanged (although for some data sets its value would change somewhat, but not dramatically). We would expect similar results with the more efficient competitor $\hat{\eta}_{(5,\,5)}$, whose computation, however, requires for each of μ and σ taking the median of $\binom{100}{5} = 75{,}287{,}520$ kernel evaluations instead of just $\binom{100}{2} = 4950$ as for $\hat{\eta}_{(2,\,2)}$. This computation may be carried out via an efficient algorithm or by the modified method described in Appendix A.5.

It is of interest to investigate the small sample performance of the estimator $\hat{\eta}_{(2,\,2)}$, and a suitable study will be carried out elsewhere. One can somewhat anticipate the results from those of a simulation study by Brazauskas and Serfling (2001) for some other GM estimators in a different context. There it was found that the superiority of the GM estimators over various competitors remained valid even for small sample sizes $n = 10$ and 25, and that the specific ARE values are valid for sample size $n \geq 100$.

3.2 Other Target Parameters

Besides the mean, the *variance*

$$\theta = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$$

is of interest and insertion of efficient and robust estimates for μ and σ then yields estimates for θ which likewise are efficient and robust. In particular, the properties of the estimators $\hat{\theta}_{(k,\,m)}$ may be developed along the lines of Section 3.1.

Also, parameters such as the "limited expected value", $E\{X \wedge x\}$, and the "limited second moment", $E\{(X \wedge x)^2\}$, play important roles in actuarial practice. For extensive discussion and treatment, see Daykin, Pentikäinen and Pesonen (1994) and Klugman, Panjer and Willmot (1998). The latter authors remark (p. 73) on the greater flexibility offered by parametric modeling over empirical modeling. In particular, for the lognormal model, explicit formulas are readily derived:

$$E\{X \wedge x\} = e^{\mu+\sigma^2/2}\,\Phi\left(\frac{\log x - \mu - \sigma^2}{\sigma}\right) \;+\; x\left[1 - \Phi\left(\frac{\log x - \mu}{\sigma}\right)\right],$$

and

$$E\{(X \wedge x)^2\} = e^{2\mu+2\sigma^2}\,\Phi\left(\frac{\log x - \mu - 2\sigma^2}{\sigma}\right) \;+\; x^2\left[1 - \Phi\left(\frac{\log x - \mu}{\sigma}\right)\right].$$

Again, insertion of efficient and robust estimates for μ and σ yields efficient and robust estimates for the limited expected value and limited second moment.

3.3 Extension to the Three-Parameter Lognormal Model

In the case of the three-parameter lognormal model $L(\mu, \sigma, \tau)$, even maximum likelihood estimation becomes highly problematic and a variety of competing modified maximum likelihood approaches have been developed, along with other types of estimation methods (see Johnson, Kotz and Balakrishnan, 1994, for general discussion). Here we propose the following estimators extending the foregoing methodology for the two-parameter case. First, let us denote the ordered Y_i's by

$$Y_{n1} \le Y_{n2} \le \cdots \le Y_{nn}$$

and estimate τ by

$$\hat{\tau}_n = Y_{n1} = \min\{Y_1, \ldots, Y_n\},$$

a natural estimator already considered in the literature. Next define transformed observations Z_i, $1 \le i \le n-1$ having ordered values given by

$$Z_{n-1,i} = \log(Y_{n,i+1} - \hat{\tau}_n) = \log(Y_{n,i+1} - Y_{n1}),\; 1 \le i \le n-1,$$

and estimate μ and σ by the estimators $\hat{\mu}_{(k)}$ and $\hat{\sigma}_{(m)}$ based on the Z_i's as surrogates of the strictly normal variates $X_i = \log(Y_i - \tau)$, $1 \le i \le n$ used when τ is known. These estimators should retain the favorable combination of efficiency and robustness (in the case of $\hat{\tau}_n$, against *upper* outliers) established in the two-parameter case, but precise quantification of these properties is highly technical and deferred to a future investigation.

Acknowledgments

Helpful comments provided by Dr. G. L. Thompson and Dr. V. Brazauskas, and insightful constructive remarks by several anonymous referees, are greatly appreciated and have led to substantial improvements in the paper. Also, support by a grant from the Society of Actuaries, with administrative support from the Actuarial Education and Research Fund, and support by NSF Grants DMS-9705209 and DMS-0103698, is gratefully acknowledged.

Appendix: Proofs and Further Details

A.1 Breakdown Points

For efficient and robust estimation of the tail index of a Pareto distribution, generalized median estimators using appropriate kernels have been developed and studied in Brazauskas and Serfling (2000a,b). Arguments given there regarding breakdown points apply in similar fashion here and are given only briefly.

For the estimator $\hat{\mu}_{(k)}$, the relevant kernel $h(x_1, \ldots, x_k) = k^{-1} \sum_{i=1}^{k} x_i \to \infty$ if one or more arguments x_i are taken to $+\infty$. Thus the GM estimator $\hat{\mu}_{(k)}$ can break down due to upper contamination unless the number M of upper contaminating observations is such that no more than half of the kernel evaluations contain a contaminating observation, that is,

$$\frac{\binom{n}{k} - \binom{n-M}{k}}{\binom{n}{k}} \leq 0.5.$$

A similar argument applies in the case of lower contamination. Thus we have

$$\mathrm{BP}(\hat{\mu}_{(k)}) = n^{-1} \max_{1 \leq M \leq n} \left\{ M : \frac{\binom{n-M}{k}}{\binom{n}{k}} \geq \frac{1}{2} \right\} \longrightarrow 1 - (\tfrac{1}{2})^{1/k}, \quad n \to \infty.$$

For the estimator $\hat{\sigma}_{(m)}$, similar arguments apply to the relevant kernel, and the same BP is obtained.

A.2 Selected Constants

Table A1 lists values of M_{m-1}, C_m and ζ_m for selected m.

Table A1

M_{m-1}, C_m and ζ_m for selected m

	m				
	2	3	5	7	9
M_{m-1}	0.45494	1.38629	3.35669	5.34812	7.34412
C_m	0.21434	0.34657	0.52586	0.65941	0.77043
ζ_m	0.02658	0.03096	0.02432	0.01890	0.01532

The values of ζ_m have been obtained by numerical integration using MAPLE and some "tweaking", after first setting up the computations via certain technical reexpressions of the problem. Specifically, it can be shown that for $m = 2$

$$w_0(z) = P\{(Z_0 - z)^2 \leq 2M_1\}$$

and for $m \geq 3$

$$w_0(z) = P\left\{\left(Z_0 - \frac{z+U}{m-1}\right)^2 \leq \frac{m}{(m-1)^2}\left[(z+U)^2 + (m-1)(M_{m-1} - z^2 - V)\right]\right\},$$

where $U = \sum_{i=1}^{m-2} Z_i$, $V = \sum_{i=1}^{m-2} Z_i^2$, and $Z_0, Z_1, \ldots, Z_{m-2}$ are independent $N(0,1)$ random variables.

A.3 Asymptotic Normality of Transformed Random Variables

As noted in Section 1.1, $\hat{\sigma}_{\text{ML}}$ is asymptotically normal with mean σ and variance $(\sigma^2/2)\, n^{-1}$. We apply the well-known "delta method", as follows. Given $\hat{\theta}$ asymptotically normal with mean θ and variance $\Delta\, n^{-1}$, if a function $g(\cdot)$ has nonzero derivative at θ, then $g(\hat{\theta})$ is asymptotically normal with mean $g(\theta)$ and variance $[g'(\theta)]^2\, \Delta\, n^{-1}$. In particular, applying this to $\hat{\sigma}_{\text{ML}}$ with $g(x) = x^2$, we obtain that $\hat{\sigma}^2_{\text{ML}}$ is asymptotically normal with mean σ^2 and variance $2\,\sigma^4\, n^{-1}$. Likewise, using $g(x) = \sqrt{x}$, we obtain that the asymptotic variance of $\hat{\sigma}_{(m)}$ is that of $\hat{\sigma}^2_{(m)}$ times the factor $1/(4\sigma^2)$. Similar considerations using transformations of asymptotically bivariate normal random variables yield the asymptotic results for $\hat{\eta}$ given in Section 3.1.2. See Serfling (1980, §3.3) for general treatment.

A.4 GES Under Transformation

By standard theory on influence functions (Serfling, 1980, and Hampel *et al.*, 1986), it is readily seen that the influence functions of $\hat{\theta}$ and $g(\hat{\theta})$ are related by

$$\text{IF}_{g(\hat{\theta})}(x) = g'(\theta)\,\text{IF}_{\hat{\theta}}(x).$$

(This is also given as Proposition 4 of Marcel and Rioux, 2001.) It follows that the corresponding GES values are related in the same fashion. This leads to the results on GES stated in Sections 2.3.2 and 3.1.1.

A.5 Computational Issues

For situations when the number $\binom{n}{k}$ of kernel evaluations needed for computation of $\hat{\mu}_{(k)}$, or number $\binom{n}{m}$ of kernel evaluations needed for computation of $\hat{\sigma}^2_{(m)}$, is extremely large (in excess of 10^7), we reduce the computational burden by randomly choosing 10^7 kernel evaluations. Such an approach maintains a high degree of numerical accuracy (up to 3 decimal places) and renders the computational burden negligible.

References

[1] Aitchison, J. and Brown, J. A. (1957). *The Lognormal Distribution.* Cambridge:Cambridge University Press.

[2] Ambühl, M. (2000). *Statistische Inferenz in Lokationsmodellen basierend auf verallgemeinerten Walsh-Mitteln*. Ph.D. Dissertation, Universität Bern.

[3] Bickel, P. J. and Lehmann, E. L. (1976). Descriptive statistics for nonparametric models. III. Dispersion. *Annals of Statistics* **4** 1139–1158.

[4] Bickel, P. J. and Lehmann, E. L. (1979). Descriptive statistics for nonparametric models. IV. Spread. In *Contributions to Statistics: Hájek Memorial Volume*, ed. J. Jureĉková. Prague: Academia, pp. 33–40.

[5] Brazauskas, V. and Serfling, R. (2000a). Robust and efficient estimation of the tail index of a single-parameter Pareto distribution. *North American Actuarial Journal* **4** 12–27.

[6] Brazauskas, V. and Serfling, R. (2000b). Robust estimation of tail parameters for two-parameter Pareto and exponential models via generalized quantile statistics. *Extremes* **3** 231–249.

[7] Brazauskas, V. and Serfling, R. (2001). Small sample performance of robust estimators of tail parameters for Pareto and exponential models. *Journal of Statistical Computation and Simulation*, **70** 1–19.

[8] Chaudhuri, P. (1992). Multivariate location estimation using extension of R-estimates through U-statistics type approach. *Annals of Statistics* **20** 897–916.

[9] Choudhury, J. (1990). Sequential confidence intervals based on generalized Hodges-Lehmann location estimators and related statistics. *Communications in Statistics – Simulation* **19** 287–303.

[10] Choudhury, J. and Serfling, R. (1988). Generalized order statistics, Bahadur representations and sequential nonparametric fixed-width confidence intervals. *Journal of Statistical Planning and Inference* **19** 269–282.

[11] Crow, E. L. and Shimizu, K. (eds.) (1988). *Lognormal Distributions: Theory and Applications*. New York: Dekker.

[12] Daykin, C. D., Pentikäinen, T., and Pesonen, M. (1994). *Practical Risk Theory for Actuaries*. Chapman & Hall, London.

[13] Derrig, R. A., Ostaszewski, K. M. and Rempala, G. A. (2000). Applications of resampling methods in actuarial science. *Proceedings of the Casualty Actuarial Society* **LXXXVII** 322–364.

[14] Ferretti, N., Kelmansky, D., Yohia, V. J. and Zamar, R. H. (1999). A class of locally and globally robust regression estimates. *Journal of the American Statistical Society* **94** 174–188.

[15] Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association* **69** 383–393.

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

[17] Janssen, P., Serfling, R., and Veraverbeke, N. (1987). Asymptotic normality of U-statistics based on trimmed samples. *Journal of Statistical Planning and Inference* **16** 63–74.

[18] Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Volume 1, 2nd edition. New York: Wiley.

[19] Klugman, S. A., Panjer, H. H., and Willmot, G. E. (1998). *Loss Models: From Data to Decisions*. New York: Wiley.

[20] Marceau, E. and Rioux, J. (2001). On robustness in risk theory. *Insurance: Mathematics and Economics* **29** 167–185.

[21] Martin, R. D., Yohai, V. J. and Zamar, R. H. (1989). Min-max bias robust regression. *Annals of Statistics* **17** 1608–1630.

[22] Myers, R. A. and Pepin, P. (1990). The robustness of lognormal-based estimators of abundance. *Biometrics* **46** 1185–1192.

[23] Randles, R. and Wolfe, D. A. (1979). *Introduction to the Theory of Nonparametric Statistics*. New York: Wiley.

[24] Rousseeuw, P. J. and Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association* **88** 1273–1283.

[25] Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

[26] Serfling, R. (1984). Generalized L-, M- and R-statistics. *Annals of Statistics* **12** 76–86.

[27] Serfling, R. (2000) Robust and nonparametric estimation via generalized L-statistics: theory, applications, and perspectives. In *Advances in Methodological and Applied Aspects of Probability and Statistics*, ed. N. Balakrishnan. Gordon & Breach, pp. 197–217.

[28] Shamos, M. I. (1976). Geometry and statistics: problems at the interface. In *New Directions and Recent Results in Algorithms and Complexity*, ed. J. F. Traub. New York: Academic Press, pp. 251–280.