ORGDP

OAK
RIDGE
GASEOUS
DIFFUSION
PLANT

UNION
CARBIDE

MASTER

FITTING A THREE-PARAMETER LOGNORMAL DISTRIBUTION
WITH APPLICATIONS TO HYDROGEOCHEMICAL
DATA FROM THE NATIONAL URANIUM
RESOURCE EVALUATION PROGRAM

V. E. Kane

October 1979

OPERATED BY
UNION CARBIDE CORPORATION
FOR THE UNITED STATES
DEPARTMENT OF ENERGY

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

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

DISCLAIMER

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

Reference to a company or product name does not imply approval or recommendation of the product by Union Carbide Corporation or the U. S. Department of Energy to the exclusion of others that may meet specifications.

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use or the results of such use of any information, apparatus, product or process disclosed in this report, or represents that its use by such third party would not infringe privately owned rights.

PAGES 1 to 2
WERE INTENTIONALLY
LEFT BLANK

CONTENTS

	<u>Page</u>
ABSTRACT.	7
INTRODUCTION.	9
1.0 PERSPECTIVE	10
2.0 STANDARD ESTIMATION METHODS	12
2.1 Maximum Likelihood.	13
2.2 Moments	15
2.3 Percentile.	16
3.0 ESTIMATION FROM GOODNESS-OF-FIT CRITERIA.	18
3.1 Goodness-of-Fit Estimation.	18
3.2 WOS Estimation for the Three-Parameter Lognormal Distribution.	21
4.0 MONTE CARLO STUDY	25
4.1 Computational Considerations.	25
4.2 Results of Monte Carlo Studies.	28
5.0 ESTIMATION CONSIDERATIONS	30
5.1 Bias.	31
5.2 Robustness.	34
6.0 EXAMPLES.	35
7.0 CONCLUSIONS	37
ACKNOWLEDGMENTS	39
REFERENCES.	41
APPENDIX.	45

DISCLAIMER

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

THIS PAGE
WAS INTENTIONALLY
LEFT BLANK

LIST OF TABLES

<u>No.</u>	<u>Title</u>	<u>Page</u>
1	Comparison of Lognormal Parameter Estimation Methods for $\gamma=0$, $\mu=0$, $\sigma=1$	31
2	Comparison of Lognormal Parameter Estimation Methods for $\gamma=0$, $\mu=0$, $\sigma=0.5$	32
3	Comparison of Lognormal Parameter Estimation Methods for $\gamma=0$, $\mu=0$, $\sigma=0.1$	33
4	Robustness Evaluation for the Three-Parameter Lognormal Estimation Methods.	36
5	Analysis of Example Data Sets	38

LIST OF FIGURES

<u>No.</u>	<u>Title</u>	<u>Page</u>
1	Typical Plot of Shapiro-Francia and Likelihood Functions	26

**THIS PAGE
WAS INTENTIONALLY
LEFT BLANK**

ABSTRACT

The standard maximum likelihood and moment estimation procedures are shown to have some undesirable characteristics for estimating the parameters in a three-parameter lognormal distribution. A class of goodness-of-fit estimators is found which provides a useful alternative to the standard methods. The class of goodness-of-fit tests considered include the Shapiro-Wilk and Shapiro-Francia tests which reduce to a weighted linear combination of the order statistics that can be maximized in estimation problems. The weighted-order statistic estimators are compared to the standard procedures in Monte Carlo simulations. Bias and robustness of the procedures are examined and example data sets analyzed including geochemical data from the National Uranium Resource Evaluation Program.

Keywords: lognormal distribution, three-parameter lognormal distribution, weighted-order statistic estimation, Shapiro-Wilk test, Shapiro-Francia test, local maximum likelihood estimation.

THIS PAGE
WAS INTENTIONALLY
LEFT BLANK

INTRODUCTION

Many analyses of geologic data use the lognormal distribution to explain the observed variation in some measured quantity. In particular, analyses of hydrogeochemical data from the National Uranium Resource Evaluation (NURE) Program often assume lognormality of the data in application of many standard statistical procedures. These standard applications implicitly assume a two-parameter lognormal model; this report addresses consideration of a three-parameter lognormal model. The extension of the lognormal model by the addition of a shift parameter gives the model greater flexibility, but unfortunately increases the difficulty of estimating the parameters.

This work addresses characteristics of local maximum likelihood [ml], estimates, moment, percentile, Cohen (1951), and modified Cohen estimators that can be used to fit the three-parameter lognormal model. A class of weighted-order statistic estimators, having desirable limiting properties, is compared to the above estimators by Monte Carlo methods. All procedures yield biased estimators of the shift parameter for certain values of the mean and variance. The robustness of the estimation methods is addressed, and several general example data sets are analyzed. The appendix presents applications of the suggested methodology to hydrogeochemical data from the NURE Program.

1.0 PERSPECTIVE

The random variable Y is said to have a three-parameter lognormal distribution if $\log(Y-\gamma) \sim N(\mu, \sigma^2)$ with $\gamma < Y < \infty$ and $\sigma^2 > 0$. The two-parameter lognormal distribution assumes that γ is a known constant. A detailed development of both distributions is given by Aitchison and Brown (1969), and Johnson and Kotz (1970, ch. 14). Both parameterizations have been used for many years to describe a variety of measurement data in areas such as geology (Ahrens, 1954), biology (Kleczkowski, 1949), economics (Gibrat, 1931), and agriculture (Cochran, 1938). More specific application areas include analysis of failure rates (Mann, et al, 1974), probits (Bliss, 1934), and particle sizes (Smith and Jordan, 1964). Additionally, many standard statistical analyses, such as analysis of variance, must often consider a logarithmic transformation to obtain the required normality.

The two-parameter lognormal model leads to standard normal estimation methods, where the three-parameter lognormal model results in estimation difficulties that often lead to adopting the two-parameter model. It is often reasoned that the range of the random variable is actually zero to infinity and not $\gamma (>0)$ to infinity. However, this range justification typically ignores practical measurement limitations.

The three-parameter model may provide a marked improvement over the two-parameter model in some practical situations (e.g., Kleczkowski, 1949). Theoretically, it should be noted that in the Pearson system (Johnson and Kotz, 1970, ch. 12) the lognormal distribution is characterized by a single curve in the β_1 (squared skewness) and β_2 (kurtosis)

plane. However, any point on the lognormal β_1, β_2 curve represents a lognormal distribution with γ unspecified. Thus, from both the practical and theoretical viewpoint, it is often important to entertain the three-parameter model. In practice, this model may often result in γ estimated to be close to zero.

Aitchison and Brown (1969) discuss early attempts at estimation in the three-parameter model, which were primarily restricted to graphical methods due to the lack of computers. Cohen (1951) gave the maximum likelihood (ml) equations and an approximate estimation procedure that he later pursued more fully (Cohen and Whitten, 1978). Hill (1963) noted that traditional likelihood estimation methods result in only lml estimates. Box and Cox (1964) graphically estimated general power-shift transformation parameters which were lml estimates for the three-parameter lognormal. Lambert (1964) again obtained lml estimates, but by using the method of scoring. Likelihood, percentile, and moment methods were compared by Calitz (1973) using Monte Carlo methods. Weighted least square methods were formulated by Munro and Wixley (1970). Penalty and barrier functions were introduced by Wingo (1975, 1976) to ensure convergence to lml estimate. Giesbrecht and Kempthorn (1976) obtain ml estimates from a discrete model by introducing a grouping interval for lognormal data. Numerous other publications consider estimation when the data are censored, a problem not addressed here.

2.0 STANDARD ESTIMATION METHODS

Let the three-parameter lognormal distribution of the random variable Y , with probability density function (pdf)

$$f(y; \gamma, \mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} (y-\gamma)^{-1} \exp \{-[\log(y-\gamma) - \mu]^2 / 2\sigma^2\},$$

$$\sigma^2 > 0, \gamma < y < \infty$$

$$= 0, \text{ otherwise} \quad (2.1)$$

be denoted by $\Lambda(\gamma, \mu, \sigma^2)$. It is useful for estimation to examine the limiting behavior of $f(y; \gamma, \mu, \sigma^2)$ as $\gamma \rightarrow -\infty$ or $\sigma \rightarrow 0$. First, note that $f(y; \gamma, \mu, \sigma^2)$ can be reparameterized in terms of the moments of Y as $f(y; \gamma, \mu_Y, \sigma_Y^2)$ with

$$\mu = \log(\mu_Y - \gamma) - 1/2 \log(1 + \sigma_Y^2 / (\mu_Y - \gamma)^2),$$

$$(2.2)$$

$$\sigma^2 = \log(1 + \sigma_Y^2 / (\mu_Y - \gamma)^2).$$

Lemma 1: $\lim_{\gamma \rightarrow -\infty} f(y; \gamma, \mu, \sigma^2) = \sigma_Y^{-1} \phi\left(\frac{y - \mu_Y}{\sigma_Y}\right)$

where ϕ is the standard normal pdf.

Proof: Using the reparameterization (2.2)

$$f(y; \gamma, \mu_Y, \sigma_Y^2) = (2\pi)^{-1/2} (\mu_Y - \gamma)^{-1} [\log(1 + \sigma_Y^2/(\mu_Y - \gamma)^2)]^{-1/2} \\ \exp \left\{ -\frac{1}{2} [\log((y - \gamma)/(\mu_Y - \gamma))]^2 / \log(1 + \sigma_Y^2/(\mu_Y - \gamma)^2) \right. \\ \left. - \frac{1}{8} \log(1 + \sigma_Y^2/(\mu_Y - \gamma)^2) \right\}.$$

For γ sufficiently small, $\log(1 + \sigma_Y^2/(\mu_Y - \gamma)^2) = \sigma_Y^2/(\mu_Y - \gamma)^2 + o(\gamma^3)$ thus, the first term approaches $(2\pi\sigma_Y^2)^{-1/2}$. Since $\log[(y - \gamma)/(\mu_Y - \gamma)] = (\mu_Y - y)/\gamma + o(\gamma)$ the exponent approaches $-(y - \mu_Y)^2/2\sigma_Y^2$.

Johnson and Kotz (1970, p. 117) give an alternate two-parameter version of Lemma 1 that as $\sigma \rightarrow 0$ the standardized lognormal pdf ($\mu_Y = 0$, $\sigma_Y = 1$) approaches the normal pdf. Since the lognormal pdf approaches the normal pdf as $\gamma \rightarrow -\infty$, it may be difficult to distinguish between the two density functions for large negative values of γ , creating estimation difficulties.

2.1 Maximum Likelihood

Consider a three-parameter lognormal sample $y_1, y_2, \dots, y_n$ with ordered values $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}$ from which estimates of γ, μ , and σ^2 are desired. Cohen (1951) gave the likelihood equations which reduce to determining the maximum of

$$L(\gamma) = \sup_{\mu, \sigma^2} (2\pi)^{n/2} \prod_{i=1}^n f(y_i; \gamma, \mu, \sigma^2) = [\hat{\sigma}(\gamma)]^{-n} \exp[-n\hat{\mu}(\gamma)]$$

where

$$\hat{\mu}(\gamma) = n^{-1} \sum_1^n \log(y_i - \gamma), \quad (2.3)$$

$$\hat{\sigma}^2(\gamma) = n^{-1} \sum_1^n [\log(y_i - \gamma) - \hat{\mu}(\gamma)]^2.$$

The corresponding log likelihood equation $\ell(\gamma) = 0$ is given by Cohen (1951). Two results given by Hill (1963) reflect on finding the maximum of $L(\gamma)$.

Corollary 2 (Hill, 1963):

$$\lim_{\gamma \rightarrow -\infty} L(\gamma) = [n^{-1} \sum_1^n (y_i - \bar{y})^2]^{-n/2} \quad (2.4)$$

Proof: Computing the limit as $\gamma \rightarrow -\infty$ of the likelihood equations using Lemma 1

$$\begin{aligned} \lim_{\gamma \rightarrow -\infty} L(\gamma) &= \lim_{\gamma \rightarrow -\infty} \sup_{\mu, \sigma^2} (2\pi)^{n/2} \prod_1^n f(y_i; \gamma, \mu, \sigma^2) \\ &= \sup_{\mu, \sigma^2} (2\pi)^{n/2} \prod_1^n \sigma_Y^{-1} \phi\left(\frac{y_i - \mu_Y}{\sigma_Y}\right) \\ &= [n^{-1} \sum_1^n (y_i - \bar{y})^2]^{-n/2}. \end{aligned}$$

Hill notes that (2.4) implies that the likelihood does not approach zero in remote regions of the parameter space. Alternatively, (2.4) is implied from Lemma 1 by the limiting relationship between the lognormal and normal pdf's. Large negative estimates of γ may suggest that the

normal model should be preferred to the three-parameter lognormal model. This data-dependent model specification is similar to the goodness-of-fit ideas suggested by Kempthorne (1966) and Easterling (1976).

Lemma 3 (Hill, 1963): $\lim_{\gamma \rightarrow y_{(1)}} L(\gamma) = +\infty$.

The ml estimate of γ, μ , and σ^2 is $(y_{(1)}, -\infty, +\infty)$ from Lemma 3. Because of this undesirable property, Wingo (1975, 1976) computed lml estimates by restricting the parameter space such that $\hat{\gamma} < y_1$. Harter and Moore (1966) suggest that these lml estimates may have many of the optimal properties of traditional ml estimators.

2.2 Moments

Johnson and Kotz (1970, p 124) give a standard moment estimation method. Equating the first three central sample moments s_1, s_2 , and s_3 to the corresponding population moments results in

$$\hat{\sigma}^2 = \log(d) ,$$

$$\hat{\mu} = \frac{1}{2} \log[s_2/(d(d-1))] , \quad (2.5)$$

$$\hat{\gamma} = s_1 - [s_2/(d-1)]^{1/2} ,$$

where

$$d = [e + (e^2 - 1)^{1/2}]^{1/3} + [e - (e^2 - 1)^{1/2}]^{1/3} - 1 \quad (2.6)$$

with $e = 1 + 1/2 s_3^2/s_2^3$. From (2.6), $d > 1$ (and $\hat{\sigma}^2 > 0$) for any $s_2 \neq 0$ and $s_3 \neq 0$. However, for samples with $s_3 \leq 0$, moment estimates are not well defined since s_3 or $-s_3$ would yield the same d . Since $\beta_1^2 = [\exp(\sigma^2) - 1] [\exp(\sigma^2) + 2]^2$, sampling variation due to either small σ^2 or small sample sizes could cause $s_3 < 0$ (Cohen and Whitten, 1979).

2.3 Percentile

Let $y(p)$ denote the p -th sample percentile, then percentile estimates from Johnson and Kotz (1970) are ($0 < p < .5$)

$$\hat{\sigma} = \{\log [y(.5) - y(p)] - \log [y(1-p) - y(.5)]\} / v(p)$$

$$\hat{\mu} = \log [y(.5) - y(p)] / [1 - \exp(v(p) \hat{\sigma})] \quad (2.7)$$

$$\hat{\gamma} = y(.5) - \exp(\hat{\mu})$$

where Φ^{-1} is the inverse standard normal cumulative distribution function (cdf) and $v(p) = \Phi^{-1}(p)$. A slight modification of (2.7) was found to be useful. Let Z_i denote the i -th standard normal order statistic from a sample of size n . Harter (1961) suggested the approximation

$$E(Z_i) \doteq \Phi^{-1} \left(\frac{i - \alpha}{n - 2\alpha + 1} \right) \quad (2.8)$$

where

$$\alpha = .315065 + .057974 \lambda_n - .009776 \lambda_n^2 \quad (i = 1, i = n)$$

$$\alpha = .327511 + .058212 \lambda_n - .007909 \lambda_n^2 \quad (2 \leq i \leq n - 1)$$

with $\lambda_n = \log_{10}(n)$. Harter discourages use of the α approximations for $n > 400$. Substituting in (2.7) $y_{(i)}$, $y_{(n-i+1)}$ and (2.8) for $y(p)$, $y(1-p)$ and $v(p)$ produced improved percentile estimates. These estimates will be said to not exist when $(y_{(n-i+1)} - y(.5))/(y(.5) - y_{(i)}) < 1$, that is, the order statistics are "skewed" to the right.

An alternative approach was suggested by Cohen (1951), and developed further by Cohen and Whitten (1979). First, note that

$$Y_i = \gamma + e^{\mu + \sigma Z_i}, \quad (2.9)$$

thus

$$\Phi^{-1}\left(\frac{i}{n+1}\right) \doteq E(Z_{(i)}) = E[(\log(Y_{(i)} - \gamma) - \mu)/\sigma]. \quad (2.10)$$

Then substituting the observed $y_{(i)}$ for the expected $Y_{(i)}$ in (2.10) yields

$$\log(y_{(i)} - \gamma) \doteq \mu + \sigma \Phi^{-1}\left(\frac{i}{n+1}\right). \quad (2.11)$$

Substituting (2.3) into (2.11) for $i=1$ gives

$$h(\gamma) = \hat{\mu}(\gamma) + \hat{\sigma}(\gamma) \Phi^{-1}\left(\frac{1}{n+1}\right) - \log(y_{(1)} - \gamma) = 0 \quad (2.12)$$

which can be solved iteratively. Cohen and Whitten (1979) consider estimates for $i = 1, 2, 3$ in (2.11), but conclude $i = 1$ gives the best results. As with the standard percentile estimate, modifying Cohen's procedure by substituting (2.8), with α determined for $i = 1$, into (2.12) for $\Phi^{-1}\left(\frac{1}{n+1}\right)$ was found to be useful. Both estimates are compared in Section 4.

3.0 ESTIMATION FROM GOODNESS-OF-FIT CRITERIA

3.1 Goodness-of-fit Estimation

Interest in goodness-of-fit estimation probably originates with minimum- χ^2 methods that were first developed primarily for discrete and categorical data. Ferguson (1958) discusses several modifications to minimum- χ^2 estimation, developed to overcome the computational difficulty of the original method. The resulting estimates are BAN. Easterling (1976) discusses minimum- χ^2 estimation along with general goodness-of-fit estimation. He suggests that parameter estimates obtained by optimizing a goodness-of-fit test statistic reflect on the statistical model being considered (through the obtained significance level of the test statistic) and require no assumptions about seemingly nonexistent "super sampling populations".

Another type of goodness-of-fit estimation involves the use of distance measures between the empirical and hypothesized distribution

functions. Typical examples would include use of the Kolmogorov-Smirnov or Cramer-von Mises statistics (e.g., Sahler, 1970; Rao, et al, 1975). Minimum-distance estimators also have desirable asymptotic properties, and are robust in some situations (Parr and Schucany, 1978).

Two considerations motivate the use of another type of goodness-of-fit estimator for the three-parameter lognormal distribution. First, the class of weighted-order statistic (wos) estimators suggested here are based on test statistics, such as the Shapiro-Wilk test, that have markedly greater power (Shapiro and Wilk, 1965) than the traditional χ^2 -test or most minimum-distance tests. This increased power suggests that the wos estimators may be sensitive to the assumption of lognormality, which is desirable since the lognormal model is assumed to be correct. The second consideration involves the need for estimation methods that are alternatives to the methods discussed in Section 2. Obtaining lml estimates can be cumbersome and in many situations moment and percentile estimates can be improved upon.

Consider sample $x_1, \dots, x_n$ with $g(X_i; \underline{\theta}) \sim F$ where g is a monotonic function, and $\underline{\theta} = (\theta_1, \dots, \theta_k) \in \Theta$ are unknown parameters. Define the wos estimate of $\underline{\theta}$ by $\hat{\underline{\theta}}$ where

$$W(\hat{\underline{\theta}}) = \sup_{\underline{\theta}} W(\underline{\theta}) \quad (3.1)$$

with

$$W(\underline{\theta}) = \frac{[\sum_{i=1}^n a_i g(x_{(i)}; \underline{\theta})]^2}{\sum_{i=1}^n [g(x_i; \underline{\theta}) - n^{-1} \sum_{j=1}^n g(x_j; \underline{\theta})]^2}, \quad (3.2)$$

the constants $\underline{a} = (a_1, \dots, a_n)$ are chosen such that $\underline{1}'\underline{a} = 0$, and $\underline{a}'\underline{a} = 1$. Additionally, $\underline{a}$ is defined such that $W(\underline{\theta})$ for $\underline{\theta}$ known is a goodness-of-fit test statistic that rejects an assumed distribution for small values. When F is symmetric, then by the symmetry of the order statistics $-a_i = a_{n-i+1}$. For estimation, $W(\underline{\theta})$ is a squared correlation measure in which the ordered a_i 's are expected to agree closely with the ordered $g(x_{(i)}; \hat{\underline{\theta}})$'s if $\hat{\underline{\theta}}$ is close to the unknown population $\underline{\theta}$. The range of $W(\underline{\theta})$ can be obtained from arguments presented by Shapiro and Wilk (1965).

$$\text{Lemma 4: } a_1^2/(1-n^{-1}) \leq W(\underline{\theta}) \leq 1.$$

It should be noted that since $W(\underline{\theta})$ is location and scale invariant, it is not possible to estimate simple changes of location and scale in X .

Applications of wos estimation include transformations where F is a standard distribution and g is a transformation involving $\underline{\theta}$ of the observed random variable. Probably the simplest example is considered here, the three-parameter lognormal distribution where F is the normal distribution and $g(x; \underline{\theta}) = \log(x - \underline{\theta})$. Examples where wos estimates are not defined include the normal and two-parameter lognormal distributions.

3.2 WOS Estimation for the Three-Parameter Lognormal Distribution

Let the wos estimate of γ in the three-parameter lognormal distribution be defined as $\hat{\gamma}$ where

$$W(\hat{\gamma}) = \sup_{\gamma} W(\gamma) \quad (3.3)$$

with

$$W(\gamma) = \frac{[\sum_{i=1}^n a_i \log(y_{(i)} - \gamma)]^2}{\sum_{i=1}^n [\log(y_{(i)} - \gamma) - n^{-1} \sum_{j=1}^n \log(y_{(j)} - \gamma)]^2} \quad (3.4)$$

the constants $\underline{a} = (a_1, \dots, a_n)'$ are chosen such that $a_i = -a_{n-i+1}$, $\underline{1}'\underline{a} = 0$, and, $\underline{a}'\underline{a} = 1$. Four test statistics in the class (3.4) are considered, each corresponding to a different choice of $\underline{a}$. Define

$$\begin{aligned} E(Z_{(i)}) &= m_i, \\ \text{Median}(Z_{(i)}) &= \tilde{m}_i \\ \text{Cov}(Z_{(i)}, Z_{(j)}) &= v_{ij}, \end{aligned} \quad (3.5)$$

with $\underline{m} = (m_1, \dots, m_n)'$ and $V = (v_{ij})$. The four test estimators are defined by the following choices of $\underline{a}$:

1. Shapiro and Wilk (1965)

$$\underline{a}' = \underline{m}' V^{-1} / (\underline{m}' V^{-1} V^{-1} \underline{m})^{1/2}, \quad (3.6)$$

2. Shapiro and Francia (1972)

$$\underline{a}' = \underline{m}' / (\underline{m}' \underline{m})^{1/2}, \quad (3.7)$$

3. Filliben (1975)

$$\underline{a}' = \underline{b}' / (\underline{b}' \underline{b})^{1/2}, \quad (3.8)$$

where $\underline{b} = (\phi^{-1}(\tilde{m}_1), \dots, \phi^{-1}(\tilde{m}_n))'$ and

4. Uniform

$$a_i = (i - \frac{n+1}{2}) \quad (3.9)$$

Four test statistics are considered because of the varying degree of computational difficulty involved in obtaining $\underline{a}$. The most difficult computations involved the Shapiro-Wilk $\underline{a}$ in (3.6) in which tabular values are given for $3 \leq n \leq 50$ by Shapiro and Wilk (1965). Next in difficulty are the Shapiro-Francia and Filliben computations. The approximation given in (2.8) was used for the m_i 's and the approximations given by Filliben (1975)

$$\tilde{m}_i = \begin{cases} 1-\tilde{m}_n & i=1 \\ (i-.3175)/(n+.365) & 2 \leq i \leq n-1 \\ .5(1/n) & i=n \end{cases} \quad (3.10)$$

was used for the $\tilde{m}_i$'s. The simplest coefficients are those used for the uniform weights (3.9).

The maximization of (3.4) for varying $\underline{a}$ is analogous to finding γ that yields $\log(y_{(i)} - \gamma)$ that are most closely correlated with the a_i . The $\underline{a}$ in (3.6), (3.7), and (3.8) are related to the order statistics of a normal distribution. The uniform $\underline{a}$ (3.9) are related only to the symmetry of $\log(y_{(i)} - \gamma)$ as computed in the D'Agostino (1971) test statistic. Comparison of estimates obtained using the uniform weights to the other choices of $\underline{a}$ allows assessment of the normality assumption and the value of increased computational difficulty for (3.6), (3.7), and (3.8).

Two limiting results for $W(\gamma)$ correspond to Corollary 2 and Lemma 3 for likelihood estimation

$$\text{Lemma 5: } \lim_{\gamma \rightarrow -\infty} W(\gamma) = \frac{[\sum_1^n a_i y_{(i)}]^2}{\sum_1^n (y_i - \bar{y})^2}$$

Proof: Noting that $\log(y - \gamma) = \log(-\gamma) - y/\gamma + o(\gamma^{-1})$ which upon substitution in $W(\gamma)$ and taking the limit as $\gamma \rightarrow -\infty$ yields the desired result.

$$\text{Lemma 6: } \lim_{\gamma \rightarrow y_{(1)}} W(\gamma) = a_1^2 / (1 - 1/n) = \min W(\gamma)$$

Proof: Let $\epsilon = 1/\log(y_{(1)} - \gamma)$, by rewriting $W(\gamma)$ to isolate $\log(y_{(1)} - \gamma)$

$$\lim_{\gamma \rightarrow y_{(1)}} W(\gamma) = \lim_{\epsilon \rightarrow 0} (a_1 + \epsilon C_1)^2 / \{ [(1-n^{-1}) + \epsilon C_2]^2 + \sum_{i \neq 1} [n^{-1} + \epsilon C_{3i}]^2 \}$$

where the C 's are constants not dependent on ϵ . The limit is equal to $a_1^2/(1-n^{-1})$, the minimum value of $W(\gamma)$ from Lemma 4.

Corresponding to Corollary 2, Lemma 5 implies that as $\gamma \rightarrow -\infty$, $W(\gamma)$ converges to the test statistic defined from the untransformed or normally distributed data. In contrast to Lemma 3, Lemma 6 implies that the $\hat{\gamma}$ which maximizes $W(\gamma)$ will not approach $y_{(1)}$, since as $\gamma \rightarrow y_{(1)}$, $W(\gamma)$ approaches a minimum value.

A solution of (3.3) can be obtained by finding γ such that

$$w(\gamma) = \frac{\partial}{\partial \gamma} \log W(\gamma) = 0 \quad (3.11)$$

with

$$w(\gamma) = \sum (u_i - \bar{u})(v_i - \bar{v}) \sum a_i u_i - \sum (u_i - \bar{u})^2 \sum a_i v_i$$

where $u_i = \log(y_{(i)} - \gamma)$ and $v_i = (y_{(i)} - \gamma)^{-1}$. Only one solution to (3.11) was found in the analyzed data. If $\hat{\gamma}$ is the first solution such that $w(\hat{\gamma}) = 0$, then from Lemma 5 and 6, $w(\gamma) < 0$ for $\hat{\gamma}' < \gamma < y_1$ and $w(\gamma) > 0$ for $\hat{\gamma}' < \gamma < \hat{\gamma}$ where $\hat{\gamma}' < \hat{\gamma}$ is a second solution to (3.11).

4.0 MONTE CARLO STUDY

4.1 Computational Considerations

A Monte Carlo sampling experiment is used to compare the likelihood, moment, percentile, Cohen, modified Cohen, and four was estimation methods for the three-parameter lognormal distribution. Sample sizes of $n=15, 25, 50,$ and 100 are used to indicate small, medium, moderate, and large sample behavior of the estimation procedures. Since Aitchison and Brown (1969) recommend that $p = .05$ in (2.7) for the percentile estimate, $i = 1, 2, 3, 5$ was chosen in (2.8) for the above sample sizes. Also, note that the Shapiro-Wilk a in (3.6) are not tabulated for $n > 51$.

The likelihood, Cohen, modified Cohen, and four was procedures all require iterative solutions. Figure 1 for the Shapiro-Francia $W(\gamma)$ (with y 's of 0.256, 0.344, 0.358, 0.589, 0.602, 0.611, 0.613, 0.616, 0.799, 0.946, 1.26, 1.64, 2.12, 2.33, 5.52) shows a typical shape of the functions to be maximized. However, $W(\gamma)$ does not change appreciably for γ appreciably less than $\hat{\gamma}$, hence derivative based optimization methods could be expected to perform poorly when an initial estimate is far from the solution. (Perhaps other functions such as $(1.01 - W^{-1}(\gamma))^{-1}$ would have better optimization properties.) The boundary problem at $\gamma = y_{(1)}$ for $L(\gamma)$ is also depicted in Figure 1.

The was estimation algorithm involved finding a solution to $w(\gamma) = 0$ in two stages. The first stage is a false position method where $\hat{\gamma}_{\text{high}} = y_{(1)} - n^{-k}$ for the smallest $k = 1, 2, \dots, k_0$ such that $w(\hat{\gamma}_{\text{high}}) < 0$

Figure 1

TYPICAL PLOT OF SHAPIRO-FRANCIA AND LIKELIHOOD FUNCTIONS

with k_0 defined so that $n_0^{-k} > \epsilon$ (say, 10^{-8}). Also, $\hat{\gamma}_{low} = \hat{\gamma}_{high} - (y_{([n/2])} - y_{(1)})^{k'}$ for the smallest $k' = 1, \dots, 10$ such that $w(\hat{\gamma}_{low}) > 0$. Note that if $k' > 1$ then the final $\hat{\gamma}_{high}$ is set equal to the last $\hat{\gamma}_{low}$ where $w(\hat{\gamma}_{low}) < 0$. Five "halving" iterations are then performed by defining $\hat{\gamma}_{new} = (\hat{\gamma}_{high} + \hat{\gamma}_{low})/2$ and replacing $\hat{\gamma}_{high}$ by $\hat{\gamma}_{new}$ if $w(\hat{\gamma}_{new}) < 0$ or $\hat{\gamma}_{low}$ by $\hat{\gamma}_{new}$ if $w(\hat{\gamma}_{new}) > 0$. The second stage consists of using the final $\hat{\gamma}_{new}$ as a starting value in Newton's method which typically converged in less than 5 iterations. An analogous procedure was used to solve for the likelihood, Cohen and modified Cohen estimates. It should be noted that multiple solutions for $\ell(\gamma) = 0$, $h(\gamma) = 0$ or $w(\gamma) = 0$ may be possible, but the above iterative scheme will find the $\hat{\gamma}$ closest to $y_{(1)}$. Cohen and Whitten (1979), in fact, observed multiple solutions for the likelihood equation.

Without loss of generality, it is possible to consider lognormal distributions of the form $\Lambda(0,0,\sigma^2)$ in the Monte Carlo simulations. To show this, let $Z_i \sim N(0,1)$ for $i=1, \dots, n$ and consider the corresponding $Y_i \sim \Lambda(\gamma, \mu, \sigma^2)$ from (2.9), and $Y_i^* \sim \Lambda(\gamma^*, \mu^*, \sigma^2)$ where

$$Y_i^* = \gamma^* + e^{\mu^* - \mu(Y_i - \gamma)}. \quad (4.1)$$

Denote by $\hat{\gamma}$ and $\hat{\gamma}^*$ estimates of γ and γ^* obtained from Y_i and Y_i^* respectively. From (4.1)

$$\log (Y_i - \hat{\gamma}) = \log \{ Y_i^* - [\gamma^* - \gamma e^{\mu^* - \mu} + \hat{\gamma} e^{\mu^* - \mu}] \} - (\mu^* - \mu) \quad (4.2)$$

$$= \log (Y_i^* - \hat{\gamma}^*) - (\mu^* - \mu)$$

with

$$\hat{\gamma}^* = \gamma^* + e^{\mu^* - \mu} (\hat{\gamma} - \gamma) \quad (4.3)$$

Using (4.1), it follows that the moment and percentile estimates from Y_i corresponding to the estimates from Y_i^* are $\hat{\gamma}^*$, $\hat{\mu}^* = \hat{\mu} + (\mu^* - \mu)$ and $\hat{\sigma}^* = \hat{\sigma}$. Similarly, using (4.2) it follows that the same relationship holds for the likelihood, wos and Cohen estimators. It should be noted that from (4.3) if $\mu^* - \mu > 0$ the bias of $\hat{\gamma}^*$ is exponentially inflated from the bias of $\hat{\gamma}$ and if $\mu^* - \mu < 0$ the bias is exponentially reduced.

The Monte Carlo study involved the generation of n uniform psuedo-random numbers from a 3^{19} congruential generator, using the KR algorithm given by Kinderman and Ramage (1976) to obtain Z_i and finally transforming the Z_i by (2.9) to obtain $Y_i \sim \Lambda(0, 0, \sigma^2)$ for $\sigma=1, .5$, and $.1$. The expected mean, standard deviation, skewness, and kurtosis of Y are respectively $\sigma=1$: 1.65, 2.16, 6.18, 114; $\sigma=0.5$: 1.28, 0.60, 1.75, 8.9; $\sigma=0.1$: 1.05, 0.10, 0.30, 3.2.

4.2 Results of Monte Carlo Studies

The estimation procedures can be compared by examining the squared deviations suggested by Aitchison and Brown (1969)

$$\Delta_{\theta} = [m^{-1} \sum_1^m (\hat{\theta}_i - \theta)^2]^{1/2} \quad (4.4)$$

where θ is the parameter being estimated, and $\hat{\theta}_1, \dots, \hat{\theta}_m$ the estimates of θ in m simulations. In the study, if m_0 simulated data sets did not produce an estimate of γ for the reasons described in Section 2, m was replaced by $m-m_0$ in (4.4).

Simulated data sets where the sample skewness was negative were excluded. Other than the moment estimator, no technical difficulties exist for negatively skewed samples. However, the lognormal distribution has strictly positive skewness and estimates of γ from negatively skewed samples resulted in large negatively biased values of $\hat{\gamma}$. Additionally, estimation for small n or σ resulted in a few very small estimates of γ that would be considered inadmissible in applied problems. For example, the likelihood method produced a minimum $\hat{\gamma}$ of -392 ($\sigma=.1$) and -207 ($\sigma=.5$) for $n=15$ where $Y_{(1)}$ was around 0.4 . The other methods were considerably worse. From Lemma 1, the large negative $\hat{\gamma}$'s appear to imply the normal model would be more appropriate than the lognormal model. To make meaningful comparisons of the methods, 0.5% of the largest and smallest $\hat{\gamma}$ values were deleted from the analyses.

Results of the Monte Carlo simulations appear in Tables 1, 2, and 3. Evaluation of the estimation methods must consider both n and σ . For all methods the accuracy of the estimates increases with increasing n or σ . Estimates of μ and σ change much less than the estimates of γ for all the methods. It is thus appropriate to assess the usefulness of an estimator by considering $\hat{\sigma}$ in an applied problem. For $\sigma=1$, methods other than the moment were acceptable with the exception of the Cohen ($n=15$) and percentile ($n=15, 25$) methods. For $\sigma=.5$, only the local ml estimator is appropriate for all n considered. The was estimators are useful for $n \geq 25$; the modified Cohen for $n \geq 50$ and the Cohen for $n \geq 100$. The moment and percentile estimates seem generally

inferior for all n with $\sigma=.5$. For $\sigma=.1$, estimation results in much larger Δ 's and only the local ml estimator seems appropriate for $n < 50$. It should be noted that the Δ_γ 's would improve appreciably for all estimators if seemingly inadmissible large negative values of $\hat{\gamma}$ were excluded.

The local ml estimator was uniformly superior, but convergence and estimation were slightly more difficult. Nonconvergence increased with increasing σ but was only encountered for $n=15$. Several of the w estimators are competitors of the likelihood estimator for $n \geq 25$ ($\sigma=1$, .5) and $n \geq 100$ ($\sigma=.1$). The Shapiro-Wilk estimator seems inferior to the Shapiro-Francia estimator which does not differ appreciably from the Filliben estimator. Interestingly, the uniform w yields Δ 's comparable to the Filliben Δ 's for $\sigma \geq .5$. The modification (2.8) in the Cohen estimator has made a substantial improvement in the Δ 's. Relative to the other estimators the moment and percentile estimator do not seem generally desirable.

5.0 ESTIMATION CONSIDERATIONS

5.1 Bias

All estimation procedures produce distributions of $\hat{\gamma}$ that are skewed to the left. From (4.3), the relatively small bias in the mean $\hat{\gamma}$ in Table 1 will increase exponentially with increasing μ^* . However, as is shown in Tables 2 and 3, a more serious problem is the marked increase in the negative bias which occurs for the smaller values of σ . It is apparent even for sample sizes of 100 ($\sigma=.1$) that large negatively biased estimates of γ are possible. However, for most estimation

Table 1

COMPARISON OF LOGNORMAL PARAMETER ESTIMATION METHODS FOR $\gamma=0$, $\mu=0$, $\sigma=1$

Estimation Method ^(a)	Mean $\hat{\gamma}$	Median $\hat{\gamma}$	Δ_{γ}	Δ_{γ}	Δ_{σ}	Estimation Method	Mean $\hat{\gamma}$	Median $\hat{\gamma}$	Δ_{γ}	Δ_{γ}	Δ_{σ}
n=15 (29) ^(b)						n=50					
1 Likelihood (107)	.026	.084	.324	.487	.431	1	.024	.028	.084	.205	.173
2 Shapiro-Wilk	-.123	.028	.847	.496	.357	2	.011	.021	.093	.206	.169
3 Shapiro-Francia	-.121	.011	.603	.476	.344	3	-.013	-.003	.097	.200	.162
4 Filliben	-.122	.009	.592	.475	.343	4	-.016	-.005	.098	.200	.162
5 Uniform WOS	-.163	-.034	.551	.469	.334	5	-.073	-.058	.134	.218	.176
6 Modified Cohen	-.135	-.032	.479	.461	.343	6	-.001	.003	.089	.202	.167
7 Cohen	-.219	-.240	3,631	.992	.414	7	-.061	-.051	.129	.218	.180
8 Moment	-2.25	-1.78	2.88	1.23	.577	8	-1.13	-1.06	1.25	.832	.405
9 Percentile (125)	-.406	.016	1.72	.711	.507	9	-.079	-.021	.256	.318	.255
n=25 (1)						n=100					
1 Likelihood	.033	.051	.158	.328	.286	1	.014	.015	.051	.130	.109
2 Shapiro-Wilk	-.024	.017	.219	.320	.254	2	-.006	-.004	.054	.128	.105
3 Shapiro-Francia	-.043	-.001	.221	.316	.247	3	-.008	-.005	.055	.129	.105
4 Filliben	-.045	-.003	.222	.316	.247	4	-.007	-.006	.094	.159	.133
5 Uniform WOS	-.099	-.050	.253	.325	.250	5	-.004	.005	.055	.132	.111
6 Modified Cohen	-.034	-.011	.178	.307	.246	6	-.030	-.027	.070	.140	.117
7 Cohen	-.177	-.116	.336	.372	.283	7	-.841	-.812	.920	.678	.327
8 Moment	-1.61	-1.43	1.84	1.04	.500	8	-.028	-.009	.119	.198	.171
9 Percentile (37)	-.213	.008	.902	.537	.403	9					

(a) For each sample size (n), 5,000 data sets were generated from $\Lambda(0, 0, 1)$ and the estimation procedures applied if the sample skewness was positive. The number of simulations where an estimation method did not converge or give $\hat{\sigma} > 0$ is given parenthetically. Note that 0.5% of the largest and smallest $\hat{\gamma}$ values were deleted.

(b) The number of simulations where the sample skewness was negative is given parenthetically.

Table 2

COMPARISON OF LOGNORMAL PARAMETER ESTIMATION METHODS FOR $\gamma=0$, $\mu=0$, $\sigma=0.5$

Estimation Method ^(a)	Mean $\hat{\gamma}$	Median $\hat{\gamma}$	Δ_{γ}	Δ_{γ}	Δ_{σ}	Estimation Method	Mean $\hat{\gamma}$	Median $\hat{\gamma}$	Δ_{γ}	Δ_{μ}	Δ_{σ}
n = 15 (259) ^(b)						n = 50 (3)					
1 Likelihood (13)	-.257	.145	1.66	.707	.515	1	-.028	.028	.291	.288	.505
2 Shapiro-Wilk	-20.6	.069	250	.982	.545	2	-.049	.021	.333	.300	.512
3 Shapiro-Francia	-3.44	.043	69.9	.786	.554	3	-.085	-.014	.339	.296	.530
4 Filliben	-2.98	.040	65.0	.773	.554	4	-.088	-.017	.340	.296	.531
5 Uniform WOS	-.548	-.015	2.21	.694	.574	5	-.167	-.090	.392	.315	.563
6 Modified Cohen	-1,552	-.057	10,034	2.11	.599	6	-.078	-.001	.362	.329	.520
7 Cohen	-10,677	-.789	33,142	4.31	.739	7	-.364	-.186	.800	.469	.599
8 Moment	-1.89	-.783	4.76	1.01	.725	8	-.513	-.360	.829	.527	.630
9 Percentile (480)	-.732	.130	3.69	.884	.524	9 (53)	-.382	-.061	1.35	.548	.560
n = 25 (82)						n = 100					
1 Likelihood	-.138	.082	.903	.502	.510	1	.003	.324	.138	.179	.496
2 Shapiro-Wilk	-.361	.030	2.94	.548	.537						
3 Shapiro-Francia	-.283	.004	1.33	.519	.548	3	-.028	-.001	.170	.180	.512
4 Filliben	-.283	.007	1.30	.518	.549	4	-.030	-.004	.171	.180	.514
5 Uniform WOS	-.339	-.064	1.13	.520	.575	5	-.112	-.083	.220	.207	.551
6 Modified Cohen	-19.7	-.010	620	.687	.554	6	-.015	.018	.204	.221	.499
7 Cohen	-4,112	-.352	24,825	2.40	.669	7	-.146	-.094	.314	.274	.555
8 Moment	-1.13	-.582	2.48	.800	.687	8	-.274	-.224	.454	.368	.583
9 Percentile (421)	-.689	.048	2.93	.789	.540	9	-.096	-.015	.352	.300	.525

(a) For each sample size (n), 5,000 data sets were generated from $\Lambda(0, 0, 0.5)$ and the estimation procedures applied if the sample skewness was positive. The number of simulations where an estimation method did not converge or give $\hat{\sigma} > 0$ is given parenthetically. Note that 0.5% of the largest and smallest $\hat{\gamma}$ values were deleted.

(b) The number of simulations where the sample skewness was negative is given parenthetically.

Table 3

COMPARISON OF LOGNORMAL PARAMETER ESTIMATION METHODS FOR $\gamma=0$, $\mu=0$, $\sigma=0.1$

Estimation Method (a)	Mean $\hat{\gamma}$	Median $\hat{\gamma}$	Δ_{γ}	Δ_{μ}	Δ_{σ}	Estimation Method	Mean $\hat{\gamma}$	Median $\hat{\gamma}$	Δ_{γ}	Δ_{μ}	Δ_{σ}
n = 15 (1832) ^(b)						n = 50 (1077)					
1 Likelihood (3)	.024	.613	2.46	1.35	.737	1	-.681	.281	3.64	.947	.854
2 Shapiro-Wilk	-39.7	.578	230	1.80	.762	2	-9.10	.272	56.0	1.28	.856
3 Shapiro-Francia	-17.2	.557	135	1.54	.769	3	-3.21	.245	27.3	1.05	.862
4 Filliben	-16.2	.555	128	1.53	.770	4	-3.08	.245	26.5	1.05	.862
5 Uniform WOS	-13.3	.527	118	1.44	.782	5	-7.87	.184	53.8	1.21	.872
6 Modified Cohen	-4,344	.418	11,020	4.75	.806	6	-9,551	.311	27,675	4.49	.846
7 Cohen	-13,899	-4,201	22,642	7.31	.911	7	24,039	-1.01	47,928	6.62	.912
8 Moment	-.849	.309	4.87	1.03	.852	8	-1.19	.091	4.94	.968	.881
n = 25 (1555)						n = 100 (653)					
1 Likelihood	-.313	.482	3.33	1.12	.797	1	-.591	.137	2.66	.792	.880
2 Shapiro-Wilk	-40.4	.450	276	1.57	.810						
3 Shapiro-Francia	-20.4	.437	181	1.36	.816	3	-.804	.110	3.87	.823	.884
4 Filliben	-20.8	.435	187	1.35	.816	4	-.802	.106	3.85	.822	.884
5 Uniform WOS	-28.3	.399	220	1.42	.828	5	-3.42	.024	24.6	.990	.894
6 Modified Cohen	-10,037	.421	27,361	4.77	.817	6	-1,081	.226	3,289	3.48	.867
7 Cohen	-27,596	-1.97	49,579	7.19	.905	7	-2,665	-.639	5,629	5.01	.918
8 Moment	-1.05	.220	5.26	1.01	.863	8	-.830	.026	3.10	.885	.893

(a) For each sample size (n), 5,000 data sets were generated from $\Lambda(0, 0, 0.1)$ and the estimation procedures applied if the sample skewness was positive. The number of simulations where an estimation method did not converge or give $\hat{\sigma} > 0$ is given parenthetically. Note that 0.5% of the largest and smallest $\hat{\gamma}$ values were deleted.

(b) The number of simulations where the samples skewness was negative is given parenthetically.

methods, the median $\hat{\gamma}$ remains relatively constant and close to the true $\gamma = 0$. From (2.3), the negative bias of $\hat{\gamma}$ causes a positive bias in $\hat{\mu}$.

An average bias correction should be possible by fixing σ and n and using the estimated biases $\hat{\gamma}_b$ and $\hat{\mu}_b$ when $\gamma = 0$ and $\mu = 0$. Consider $\hat{\gamma}^*$, $\hat{\mu}^*$, and $\hat{\sigma}$ to be estimated from $\Lambda(\gamma^*, \mu^*, \sigma^2)$. Then using (4.3), the corrected estimates are

$$\hat{\mu}_c = \hat{\mu}^* - \hat{\mu}_b, \quad (5.1)$$

$$\hat{\gamma}_c = \hat{\gamma}^* - \hat{\gamma}_b \exp(\hat{\mu}_c)$$

Values of $\hat{\gamma}_b$ and $\hat{\mu}_b$ can be obtained from Table 2. However, use of (5.1) is not desirable since average corrections are probably inappropriate because of the large variations in $\hat{\gamma}$.

The bias problem for $\hat{\gamma}$ can cause estimation difficulties for small n or σ . It is desirable to compute the sample skewness, if it is negative the lognormal approximation is probably inappropriate. In any case, the negative sample skewness will cause the moment estimate to be undefined and may cause the wos or likelihood estimators to produce extreme negative estimates of γ . Comparison of several estimators may be desirable.

5.2 Robustness

A possible concern for any of the estimation procedures is the sensitivity to departures from lognormality and the effect of outliers. Relatively extreme departures from lognormality were considered by generating lognormal data from (2.9) with $\gamma=0$, $\mu=0$ and Z_i having a t -distribution with three degrees of freedom. Table 4 indicates that Δ_γ was reduced for most of the procedures, possibly due to an increased number of smaller values of Z_i defining the boundary at γ more precisely. The

Table 4

ROBUSTNESS EVALUATION FOR THE THREE-PARAMETER
LOGNORMAL ESTIMATION METHODS

Estimation Method (a)	Mean $\hat{\gamma}$	Δ_{γ}	$\Delta_{\mu}^{(b)}$	$\Delta_{\sigma}^{(b)}$
t-distribution				
Likelihood (2)	- .006	.059	.200	.461
Shapiro-Wilk (1)	- .009	.067	.198	.474
Shapiro-Francia (1)	- .019	.073	.203	.490
Filliben (1)	- .020	.073	.204	.491
D'Agostino (1)	- .046	.096	.224	.528
Modified Cohen	- .027	.069	.217	.502
Cohen	- .059	.100	.242	.546
Moment (62)	-28.2	83.5	2.33	.187
Percentile	- .047	.206	.275	.579
.94 N(0,1) + .06 N(3,1)				
Likelihood	.087	.105	.181	.424
Shapiro-Wilk	.089	.105	.178	.426
Shapiro-Francia	.077	.097	.183	.391
Filliben	.076	.097	.183	.389
D'Agostino	.049	.083	.203	.335
Modified Cohen	.058	.089	.201	.361
Cohen	.027	.082	.229	.312
Moment	5.20	7.18	1.69	.168
Percentile	.068	.191	.276	.474
.94 N(0,1) + .06 N(10,1)				
Likelihood	.118	.129	.335	1.91
Shapiro-Wilk	.118	.129	.338	1.89
Shapiro-Francia	.118	.129	.338	1.89
Filliben	.118	.129	.338	1.89
D'Agostino	.117	.128	.358	1.84
Modified Cohen	.116	.128	.367	1.82
Cohen	.114	.126	.382	1.79
Percentile	.108	.175	.268	.641

(a) Each estimation method was applied to the same samples of size $n = 50$ from 1000 simulations. The number of simulations where an estimation method did not converge or give $\hat{\sigma} > 0$ is given parenthetically; the moment method was not performed if $s_3 > 10^{10}$ due to computational considerations.

(b) In (4.15), $\mu = 0$ was assumed; in the t-distribution simulations $\sigma = 1.73$, and in the normal distribution simulations $\sigma = 1$.

precision of $\hat{\mu}$ was roughly comparable with the simulations in Table 1. Direct comparison of Δ_{σ} between the two tables is not possible since σ increased with the t-distribution.

Outliers in the left tail of the lognormal distribution can be expected to alter all estimates appreciably since $\gamma < y_{(1)}$. It would be desirable if outliers in the right tail would not affect $\hat{\gamma}$ substantially. To evaluate the influence of large outliers, 47 Z_i 's in (2.9) were generated from a $N(0,1)$ distribution and 3 Z_i 's from $N(3,1)$ in the first group of simulations; 3 Z_i 's from $N(10,1)$ in a second group of simulations. From Table 4, the small number of outliers causes positive bias in $\hat{\gamma}$ and increased variation of $\hat{\gamma}$, $\hat{\mu}$, and $\hat{\sigma}$. The moment and Cohen estimators appear less affected by the outliers, but this is probably due to the outliers compensating for the negative bias exhibited in Tables 1, 2, and 3. The percentile estimates seem less affected by the outliers since extreme points are truncated in the estimation procedure.

6.0 EXAMPLES

The first example is a simulated data set given by Cohen (1951) for which $\gamma=100$, $\mu=3.91$, and $\sigma=.4$. The minimum value is $y_{(1)} = 127$, the sample skewness is 0.75, the Shapiro-Wilk statistic of the untransformed data is .937 ($P=.206$), the Shapiro-Wilk statistic assuming $\gamma=0$ is .954 ($P=.424$), and the final Shapiro-Wilk statistic is .969 ($P=.713$). This data set provides an interesting example since most analysts would not transform the original data. Results of several analyses of this data

appear in Table 5. Additional analyses of this data set are given by Wingo (1975, 1976) using penalized likelihood and moving barrier optimization of the likelihood. The same results (to three significant figures) as given by Cohen (1951) were obtained by Wingo.

The second example given by Wilson and Worcester (1945) is the percentage of net state and municipal debt relative to wealth for the forty-eight contiguous states. The minimum value is $y_{(1)} = 1.5$, the sample skewness is 1.1, the Shapiro-Wilk statistic of the untransformed data is .800 ($P < .001$), the Shapiro-Wilk statistic assuming $\gamma = 0$ is .978 ($P = .663$), and the final Shapiro-Wilk statistic is .992 ($P = .99$). Various analyses appear in Table 5.

The third example given in Lambert (1964) consists of observations from the annual 24-hour maximum rainfalls at Sydney, Australia from 1859 to 1945. The minimum value is $y_{(1)} = 154$, the sample skewness is 2.3, the Shapiro-Francia statistic of the untransformed data is .928 ($P < .01$), the Shapiro-Francia statistic assuming $\hat{\gamma} = 0$ is .993 ($P = .85$), and the final Shapiro-Francia statistic is .994 ($P = .90$). Various analyses appear in Table 5; the estimates of γ vary appreciably.

7.0 CONCLUSIONS

The standard ml estimates for the three-parameter lognormal distribution have been shown to have some undesirable characteristics. It is possible to compute local ml estimates but specialized methods must be used to insure the optimization procedure does not let $\hat{\gamma} = y_{(1)}$. The wos estimators provide a useful alternative for moderate to large n and σ to

Table 5

ANALYSIS OF EXAMPLE DATA SETS

Source of Data	Source of Analysis	Type of Estimator	$\hat{\gamma}$	$\hat{\mu}$	$\hat{\sigma}$
Cohen (1951) n = 20	Cohen (1951)	moment	72.8	4.35	.241
	Cohen (1951)	lmle	117.7	3.37	.604
	Cohen (1951)	Cohen	104.7	3.78	.407
		wos	107.6	3.71	.448
Wilson and Worchester (1945) n = 48	W & W (1945)	moment	.009	1.43	.602
	W & W (1945)	lmle	1.02	1.11	.763
	Wingo (1976)	mmle	1.02	1.01	.764
		wos	1.02	1.11	.772
Lambert (1964) n = 87	Lambert (1964)	percentile	-113.5	6.23	.314
	Lambert (1964)	lmle	- .77	5.97	.403
	Wingo (1975)	mmle(a)	- 4.24	5.99	.399
	Wingo (1976)	mmle	- 3.94	5.98	.400
		moment	-72.3	6.15	.344
		wos	- 18.39	6.02	.387

(a) Modified maximum likelihood estimates.

the local mle estimators and other estimation procedures since they yield comparable point estimates of the parameters and can be easily computed by standard iterative procedures.

Interval estimates have not been addressed, but direct computation of consonance intervals as described by Easterling (1976) is possible using tabulated Shapiro-Francia percentiles. However, the original Shapiro-Francia (1972) percentiles may be inaccurate (Pearson, et al, 1977). The Shapiro-Francia estimators seem appropriate for wos estimation. For all estimators, the sample skewness should be examined and estimates computed only for positively skewed distributions. Finally, for small σ (say $\sigma \leq .5$) and even moderate n (say, $n \leq 50$), any estimator of γ may be negatively biased and the estimate of μ positively biased. Comparison of wos and moment estimators may be desirable.

ACKNOWLEDGMENTS

The author wishes to thank D. S. Robson and A. C. Cohen for their suggestions concerning sections of the manuscript, and C. L. Begovich for her extensive computational assistance.

THIS PAGE
WAS INTENTIONALLY
LEFT BLANK

REFERENCES

1. Ahrens, L. H. (1954), "The Lognormal Distribution of Elements", *Geochimica et Cosmochimica Acta*, 5, 49-73.
2. Aitchison, J., and Brown, J. A. C. (1969), *The Lognormal Distribution*, 4th ed., London: Cambridge University Press.
3. Bliss, C. I. (1934), "The Method of Probits", *Science*, 79, 38-39.
4. Box, G. E. P., and Cox, D. R. (1964), "An Analysis of Transformations", *Journal of the Royal Statistical Society, Series B*, 26, 211-251.
5. Calitz, F. (1973), "Maximum Likelihood Estimation of the Parameters of the Three-Parameter Lognormal Distribution - a Reconsideration", *Australian Journal of Statistics*, 3, 185-190.
6. Cochran, W. G. (1938), "Some difficulties in the Statistical Analysis of Replicated Experiments", *Empire Journal of Experimental Agriculture*, 6, 157-163.
7. Cohen, A. C. (1951), "Estimating Parameters of Logarithmic-Normal Distributions by Maximum Likelihood", *Journal of the American Statistical Association*, 46, 206-212.
8. Cohen, A. C., and Whitten, B. J. (1979), "Estimation in the Three Parameter Log-normal Distribution", *Journal of the American Statistical Association*, to appear.
9. D'Agostino, R. B. (1971), "An Omnibus Test of Normality for Moderate and Large Size Samples", *Biometrika*, 58, 341-348.
10. Easterling, R. G. (1976), "Goodness of Fit and Parameter Estimation", *Technometrics*, 18, 1-9.
11. Ferguson, T. S. (1958), "A Method of Generating Best Asymptotically Normal Estimates with Application to the Estimation of Bacterial Densities", *Annals of Mathematical Statistics*, 29, 1046-1062.
12. Filliben, J. J. (1975), "The Probability Plot Correlation Coefficient Test for Normality", *Technometrics*, 17, 111-117.
13. Gibrat, R. (1931), *Les Inegalites Economiques*, Paris: Libraire du Recueil Sirey.
14. Giesbrecht, F., and Kempthorne, O. (1976), "Maximum Likelihood Estimation in the Three-Parameter Lognormal Distribution", *Journal of the Royal Statistical Society, Series B*, 38, 257-264.

15. Harter, H. L. (1961), "Expected Values of Normal Order Statistics", *Biometrika*, 48, 151-165.
16. Harter, H. L., and Moore, A. H. (1966), "Local Maximum Likelihood Estimation of the Parameters of Three-Parameter Lognormal Populations from Complete and Censored Samples", *Journal of the American Statistical Association*, 61, 842-851.
17. Hill, B. M. (1963), "The Three-Parameter Lognormal Distribution and Bayesian Analysis of a Point-Source Epidemic", *Journal of the American Statistical Association*, 58, 72-84.
18. Johnson, N. L., and Kotz, S. (1970), *Continuous Univariate Distributions* 1, New York: John Wiley & Sons.
19. Kempthorne, O. (1966), "The Classical Problem of Inference-Goodness of Fit", *Proceedings of the Fifth Berkley Symposium*, 1, 235-249.
20. Kinderman, A. J., and Ramage, J. G. (1976), "Computer Generation of Normal Random Variables", *Journal of the American Statistical Association*, 71, 893-896.
21. Kleczkowski, A. (1949), "The Transformation of Local Lesion Counts for Statistical Analysis", *Annals of Applied Biology*, 36, 139-152.
22. Lambert, J. A. (1964), "Estimation of Parameters in the Three-Parameter Lognormal Distribution", *The Australian Journal of Statistics*, 6, 29-32.
23. Mann, N. R., Schafer, R. E., and Singpurwalla, N. D. (1974), *Methods for Statistical Analysis of Reliability and Life Data*, New York: John Wiley & Sons.
24. Munro, A. H., and Wixley, R. A. J. (1970), "Estimators Based on Order Statistics of Small Samples from a Three-Parameter Lognormal Distribution", *Journal of the American Statistical Association*, 65, 212-225.
25. Parr, W. C., and Schucany, W. R. (1978), "Minimum Distance and Robust Estimation", unpublished manuscript.
26. Pearson, E. S., D'Agostino, B., and Bowman, K. O. (1977), "Tests for Departure from Normality: Comparison of powers", *Biometrika*, 64, 231-246.
27. Rao, P. V., Schuster, E. F., and Littell, R. C. (1975), "Estimation of Shift and Center of Symmetry Based on Kolmogorov-Smirnov Statistics", *Annals of Statistics*, 3, 862-873.
28. Sahler, W. (1970), "Estimation by Minimum-Discrepancy Methods", *Metrika*, 15, 85-106.

29. Shapiro, S. S., and Francia, R. S. (1972), "An Approximate Analysis of Variance Test for Normality", *Journal of the American Statistical Association*, 67, 215-216.
30. Shapiro, S. S., and Wilk, M. B. (1965), "An Analysis of Variance Test for Normality (Complete Samples)", *Biometrika*, 52, 591-611.
31. Sinclair, A. J. (1976), *Applications of Probability Graphs in Mineral Exploration*, Special Vol. 4, The Association of Exploration Geochemists, Richmond Printers, Richmond, B. C., Canada.
32. Smith, J. E., and Jordan, M. L. (1964), "Mathematical and Graphical Interpretation of the Log-Normal Law for Partical Size Distribution Analysis", *Journal of Colloid Science*, 19, 549-559.
33. Uraninum Resource Evaluation Project, *Hydrogeochemical and Stream Sediment Reconnaissance Basic Data for Plainview NTMS Quadrangle, Texas*, Union Carbide Corporation, Nuclear Division, Oak Ridge Gaseous Diffusion Plant, Oak Ridge, Tennessee, K/UR-124 (May 11, 1979). United States Department of Energy, Grand Junction, Colorado [GJBX-97(79)].
34. Wilson, E. B., and Worcester, J. (1945), "The Normal Logarithmic Transform", *Review of Economic Statistics*, 27, 17-22.
35. Wingo, D. R. (1975), "The Use of Interior Penalty Functions to Overcome Lognormal Distribution Parameter Estimation Anomalies", *Journal of Statistical Computation and Simulation*, 4, 49-61.
36. Wingo, D. R. (1976), "Moving Truncations Barrier-Function Methods for Estimation in Three-Parameter Lognormal Models", *Communication in Statistics Simulation and Computation*, B5, 65-80.

APPENDIX
ANALYSIS OF NURE PROGRAM HYDROGEOCHEMICAL DATA

THIS PAGE
WAS INTENTIONALLY
LEFT BLANK

APPENDIX A

LIST OF TABLES

<u>No.</u>	<u>Title</u>	<u>Page</u>
A-1	Sample Statistics for Selected Variables From the Plainview Quadrangle.	A-6

LIST OF FIGURES

<u>No.</u>	<u>Title</u>	<u>Page</u>
A-1	Standard and Shifted Lognormal Probability Plots for Uranium	A-8
A-2	Standard and Shifted Lognormal Probability Plots for Uranium and Arsenic	A-10
A-3	Standard and Shifted Lognormal Probability Plots for Conductivity and Total Alkalinity	A-11

THIS PAGE
WAS INTENTIONALLY
LEFT BLANK

APPENDIX

Groundwater data from the Plainview Quadrangle³³, collected as part of the NURE reconnaissance program, was analyzed to evaluate the use of the three-parameter lognormal model as well as the wos estimation method. Reconnaissance geochemical data have two problems that influence the three-parameter lognormal estimation problem. First, mutiple populations are often expected because of the varying geologic populations that are sampled. Second, laboratory analysis often result in data below the detection limit, which creates a censored sample. The analysis of the Plainview data will divide the samples into two major geologic regions, the Ogallala Formation (code TPO) and the Permian Units, consisting of the Blaine Formation (PGEB), the Whitehorse and Cloud Chief Group (PGWC), and the Quartermaster Group (POQ). The censoring problem is ignored by selecting variables that have few, if any, censored values. If several values are censored, they are omitted from the analysis.

Sample statistics and parameter estimates appear in Table A-1 for eight geochemical variables. These results allow assessments of the appropriateness of the following: (1) no transformation, (2) the two-parameter lognormal versus the three-parameter lognormal distribution, and (3) separating the data by the geologic unit. With the exception of arsenic in PGEB and PGWC where there were numerous censored values, it is possible to assess the preceding three transformation options.

Table A-1

SAMPLE STATISTICS FOR SELECTED VARIABLES FROM THE PLAINVIEW QUADRANGLE

Untransformed Data						Log (Y) Transformed Data						Log (Y - y) Transformed Data					
Geologic Unit ^(a)	Sample Size	Minimum	Maximum	Skewness ^(b)	Kurtosis ^(b)	Shapiro-Francia Statistic ^(b)	Skewness ^(b)	Kurtosis ^(b)	Shapiro-Francia Statistic ^(b)	Mean $\bar{\mu}$	Standard Deviation $\bar{\sigma}$	Skewness ^(b)	Kurtosis	Shapiro-Francia Statistic ^(b)	Shift $\bar{y}$	Mean $\bar{\mu}$	Standard Deviation $\bar{\sigma}$
Uranium																	
Ogallala Formation	356	0.19	37.9	2.63	16.9	0.810	-1.7	7.79	0.851	1.74	0.734	0.08	4.65	0.962	-4.32	2.36	0.337
Permian Units	473	0.10	106.9	1.10	25.5	0.619	-0.74	5.50	0.956	2.00	0.947	0.12	3.70	0.989	-1.00	2.19	0.757
PGEB	121	0.10	68.4	2.87	15.6	0.725	-1.03	5.03	0.924	1.79	1.07	0.08	3.11	0.983	-1.48	2.11	0.729
PGWC	275	0.10	106.9	4.32	29.1	0.615	-0.60	5.48	0.966	2.06	0.874	0.10	3.65	0.991	-0.87	2.21	0.732
POQ	77	0.42	103.2	3.32	15.2	0.586	-0.17	4.36	0.965	2.14	0.958	0.24	3.84	0.974	-0.53	2.23	0.865
Arsenic																	
Ogallala Formation	355	0.25	15.0	1.10	5.32	0.941	-0.88	4.72	0.960	1.35	0.560	0.00	2.99	0.997	-2.37	1.87	0.329
Permian Units	472	0.25	17.8	3.11	23.9	0.757	-0.17	2.01	0.930								
PGEB	121	0.25	3.8	2.31	8.76	0.654	0.91	2.66	0.779								
PGWC	274	0.25	8.0	1.33	4.96	0.887	-0.52	2.50	0.942								
POQ	77	0.25	17.8	3.92	24.1	0.633	-0.17	3.52	0.973	0.628	0.769	0.07	3.48	0.976	-0.15	0.73	0.692
Boron																	
Ogallala Formation	354	16	2,383	7.42	89.2	0.522	-0.59	5.69	0.911	5.16	0.558	0.17	6.05	0.929	-39.7	5.39	0.457
Permian Units	473	7	10,670	4.10	26.8	0.592	-0.02	3.81	0.990	6.26	0.955	0.17	3.34	0.993	-12.4	6.29	0.955
PGEB	121	41	10,670	3.28	18.5	0.677	-0.30	3.26	0.990	6.67	1.01	0.02	2.90	0.996	-60.6	6.79	0.898
PGWC	275	7	10,090	4.66	34.4	0.543	0.04	4.61	0.976	6.16	0.921	0.31	3.85	0.982	-14.2	6.20	0.875
POQ	77	14	6,875	4.15	24.8	0.542	0.03	3.80	0.981	5.95	1.03	0.24	3.44	0.984	-11.5	6.00	0.981
Calcium																	
Ogallala Formation	354	1.7	509.8	11.1	170.7	0.401	-2.06	20.9	0.751	3.74	0.452	0.74	14.6	0.848	-13.7	4.04	0.296
Permian Units	473	70.0	1,049.0	-0.827	3.3	0.882	-2.8	15.6	0.668	5.99	0.751						
PGEB ^(c)	121	70.1	1,049.0	-1.05	5.44	0.855	-2.55	9.67	0.661	6.23	0.441						
PGWC ^(c)	275	70.0	1,027.0	-0.85	3.84	0.886	-4.09	30.9	0.604	6.04	0.658						
POQ ^(c)	77	12.4	720.0	-0.06	1.38	0.882	-0.78	2.51	0.866	5.41	1.03						
Sodium																	
Ogallala Formation	354	3.8	281.2	3.93	22.2	0.607	-0.05	6.05	0.939	3.64	0.579	0.33	5.71	0.943	-2.96	3.72	0.525
Permian Units	473	3.3	2,882	4.70	33.4	0.517	0.28	3.10	0.993	4.56	1.10	0.19	3.20	0.994	1.45	4.54	1.13
PGEB	121	3.8	2,882	4.78	29.8	0.475	0.22	3.10	0.991	4.81	1.08	0.06	3.15	0.993	3.98	4.75	1.14
PGWC	275	3.3	1,416	3.13	13.5	0.582	0.38	3.07	0.983	4.39	1.09	0.27	3.24	0.985	1.50	4.36	1.12
POQ	77	5.1	5,686	4.48	27.9	0.495	0.11	3.53	0.982	4.79	1.10	0.19	3.44	0.982	-1.46	4.81	1.07
Conductivity																	
Ogallala Formation	356	290	2,815	1.23	29.8	0.795	0.095	5.49	0.949	6.52	0.272	0.03	5.37	0.949	25.4	6.48	0.283
Permian Units	473	240	12,200	1.20	6.70	0.923	-1.30	7.00	0.909	8.10	0.508	0.05	4.01	0.976	-2,937	8.77	0.238
PGEB	121	800	5,900	0.91	4.86	0.936	-0.883	5.08	0.934	8.27	0.409	0.52	4.16	0.960	-12,983	9.75	0.090
PGWC	275	520	12,200	1.72	10.2	0.877	-0.847	5.69	0.927	8.10	0.427	0.08	4.67	0.958	-1,865	8.57	0.256
POQ	77	140	5,760	1.15	4.48	0.916	-0.979	4.23	0.937	7.84	0.748	0.31	2.84	0.990	-1,591	8.39	0.404
Sulfate																	
Ogallala Formation	356	2.5	1,557	15.4	166.	0.17	0.46	5.29	0.972	3.05	0.814	0.34	5.19	0.972	.49	3.01	0.841
Permian Units ^(c)	473	6	4,322	-0.08	3.7	0.952	-2.3	8.6	0.711	6.99	0.936						
PGEB ^(c)	121	25	3,340	-0.32	4.14	0.938	-3.13	16.3	0.667	7.21	0.640						
PGWC ^(c)	275	20	3,760	-0.22	3.65	0.963	-2.67	10.9	0.683	7.09	0.83						
POQ	77	6	4,322	0.93	4.12	0.896	-0.95	3.11	0.896	6.33	1.41	0.04	1.82	0.932	.770	7.36	0.490
Total Alkalinity																	
Ogallala Formation	356	42	498	0.05	7.44	0.947	-2.75	24.2	0.821	5.53	0.137	-0.03	7.39	0.947	-5,137	8.59	0.008
Permian Units	473	25	468	0.89	3.98	0.946	-0.39	2.59	0.982	4.89	0.572	-0.02	2.37	0.989	-54.9	5.27	0.387
PGEB	121	46	468	1.02	5.92	0.935	-0.79	5.34	0.941	5.18	0.356	0.07	4.25	0.972	-144.5	5.79	0.186
PGWC	275	26	348	0.94	3.31	0.923	-0.08	2.29	0.990	4.69	0.566	0.09	2.29	0.990	1.1	4.68	0.562
POQ	77	25	468	0.63	2.94	0.961	-0.72	3.26	0.960	5.11	0.626	0.00	2.38	0.990	-153	5.81	0.299

(a) Permian Units: PGEB - Blaine Formation, PGWC - Whitehorse and Cloud Creek Group, POQ - Quartermaster Group.

(b) Normal distribution values are: Skewness = 0, Kurtosis = 3, Shapiro-Francia statistic = 1.0.

(c) Negative skewness implies lognormal model is not appropriate.

The calcium and sulfate data in the Permian units have mostly negative skewness values, which suggests that a lognormal distribution is not appropriate. Alternately, the uranium in the Ogallala Formation provides a good example where the addition of the shift parameter provides an increased fit. This is best shown by the increase in the Shapiro-Francia statistic from 0.851 to 0.962, as well as the change in the sample skewness and kurtosis to more nearly normal values.

The possible importance in the shift can be illustrated by considering the determination of an anomalous percentile (e.g., 95) for uranium. For the two-parameter lognormal model:

$$\begin{aligned} Y &\geq \exp (1.64 \hat{\sigma}_2 + \hat{\mu}_2) \\ &\geq 19.0 \end{aligned}$$

are the anomalous values, where $\hat{\mu}_i$, $\hat{\sigma}_i$ ($i = 2,3$) are the respective moment estimates from the two- and three-parameter models. For the three-parameter lognormal model,

$$\begin{aligned} Y &\geq \exp (1.64 \hat{\sigma}_3 + \hat{\mu}_3) + \hat{\gamma}_3 \\ &\geq 14.1 \end{aligned}$$

The three-parameter percentile results in 12 instead of 8 anomalous uranium values in the Permian Units. Note that the estimates of the mean and standard deviation change appreciable for the two models.

The Ogallala Formation uranium example also illustrates where it may not be desirable to use a single lognormal distribution. In Figures A-1(a), the S-curvature usually indicates more than one population (Sinclair, 1976). Use of the three-parameter model often eliminates the curvature, but Figures A-1(b) still is indicative of multiple populations.

(a) Ogallala Formation

(b) Ogallala Formation

(c) Permian Units

(d) Permian Units

(e) Blaine Formation

(f) Blaine Formation

Figure A-1

STANDARD AND SHIFTED LOGNORMAL PROBABILITY PLOTS FOR URANIUM

The remaining probability plots in Figures A-1 to A-3 show varying degrees of linearization for the shifted data. For example, arsenic in the Ogallala Formation [Figure A-2(e)] or conductivity in the Quaternary Group [Figure A-3(c)] show some S-curvature in the standard probability plots, but the shifted data seem adequately linear.

It should be noted that the shift for many of the elements is negative, which seems to imply that negative values are possible concentrations. For example, in the Ogallala Formation uranium data $Y > -4.32$ instead of $Y > 0$. This is an unfortunate possible consequence of fitting the three-parameter lognormal model. An explanation is that the fit to the data is over the whole range of Y rather than requiring $Y > 0$.

More importantly, it is possible to estimate the probability that Y is in the interval $(-4.32, 0)$ assuming the three-parameter model,

$$\begin{aligned} \Pr\{-4.32 > Y > 0\} &= \Pr\{[\text{Log}(Y+4.32) - \hat{\mu}]/\hat{\sigma} < [\text{Log}(4.32) - \hat{\mu}]/\hat{\sigma}\} \\ &= 0.0039 \end{aligned}$$

Thus, there is a very small probability of negative values under the fitted three-parameter model.

The sodium values exhibit a positive shift with $Y > 3.98$ in the Blaine Formation. In this case, the $\hat{\gamma} = 3.98$ is often interpreted as a lower limit threshold value. However, for the Plainview data few positive threshold values were obtained, and these were probably of little consequence.

In summary, the methodology described in this report provides a useful tool to assess the distributional form of noncensored geochemical data. It is possible to determine whether a two- or three-parameter

(a) Whitehorse and Cloud Chief Group

(b) Whitehorse and Cloud Chief Group

(c) Quartermaster Group

(d) Quartermaster Group

(e) Ogallala Formation

(f) Ogallala Formation

Figure A-2

STANDARD AND SHIFTED LOGNORMAL PROBABILITY PLOTS FOR URANIUM AND ARSENIC

(a) Blaine Formation

(b) Blaine Formation

(c) Quartermaster Group

(d) Quartermaster Group

(e) Quartermaster Group

(f) Quartermaster Group

Figure A-3

STANDARD AND SHIFTED LOGNORMAL PROBABILITY PLOTS
FOR CONDUCTIVITY AND TOTAL ALKALINITY

lognormal distribution fits the data adequately and to estimate the appropriate parameters. Examples have shown that even small shifts can effectively make the lognormal probability plots more linear. It is important to note that curvature in the lower region of a probability plot may indicate the need for a shift rather than assuming the existence of multiple populations.

DISTRIBUTION

INTERNAL

1 -34. <u>Computer Sciences Division</u> Bayne, C. K. Beauchamp, J. J. Begovich, C. L. Carter, H. P. Gardiner, D. A. Gosslee, D. G. Green, R. O. Hammerling, F. D. Kane, V. E. (25) Westley, G. W.	40. <u>Operations Analysis and Planning Division</u> Bradbury, J. T.
	41. <u>Technical Director</u> Wilcox, W. J., Jr.
	42-62. <u>Uranium Resource Evaluation Project</u> Arendt, J. W.
35. <u>Ecological Sciences Information Center</u> Daniel, E. W.	63. <u>ORGDP Records Department (RC)</u>
36. <u>Enrichment Technology Div.</u> Merriman, J. R.	64-65. <u>Oak Ridge National Laboratory</u> Row, T. H. Shults, W. D.
37-38. <u>K-25 Technical Services Div.</u> Levin, R. W. Napolitan, D. S.	66-69. <u>Y-12 Plant</u> Bernander, N. K. Hibbs, R. F. Vanstrum, P. R. White, J. C.
39. <u>Library</u>	

EXTERNAL

70. M. A. Boyd
Development Division
Oak Ridge Operations Office
Department of Energy
P.O. Box E
Oak Ridge, Tennessee 37830
71. Al Clebsch, Jr.
Water Resources Division
U. S. Geological Survey
Denver Federal Center
Mail Stop 406
Denver, Colorado 80225

72. David H. Dahlem, Project Officer
NURE Project Office - Geochemistry
Grand Junction Office
Department of Energy
P.O. Box 2567
Grand Junction, Colorado 81501
73. A. P. d'Silva
Ames Laboratory
Iowa State University
of Science and Technology
Ames, Iowa 50011
- 74-75. Donald L. Everhart, Manager
Grand Junction Office
Department of Energy
P.O. Box 2567
Grand Junction, Colorado 81501
76. R. G. Garrett
Geological Survey of Canada
601 Booth Street
Ottawa, Canada K1A0E8
77. M. L. Hyder
NURE Program
Savannah River Laboratory
Aiken, South Carolina 29801
78. Victor J. Janzer
Water Resources Division
U. S. Geological Survey
Federal Center
Box 25046, Mail Stop 407
Lakewood, Colorado 80225
79. C. B. Kiser, Assistant Manager
for Development and Planning
Oak Ridge Operations Office
Department of Energy
P.O. Box E
Oak Ridge, Tennessee 37830
80. A. T. Miesch
U. S. Geological Survey
Federal Center
Box 25046
Denver, Colorado 80225

81. R. B. McCammon
U. S. Geological Survey
National Center 920
Reston, Virginia 22092
82. W. Morris
Alternate Group Leader, J-5
Los Alamos Scientific Laboratory
P.O. Box 1663, Mail Stop 586
Los Alamos, New Mexico 87545
- 83-84. Donald E. Livingston
Bendix Field Engineering Corporation
P.O. Box 1569
Grand Junction, Colorado 81501
- 85-87. T. W. Offield, Chief
Branch of Uranium-Thorium Resources
U. S. Geological Survey
Box 25046, Mail Stop 916
Denver Federal Center
Denver, Colorado 80225
88. Arthur W. Rose
Pennsylvania State University
119 Mineral Sciences Building
University Park, Pennsylvania 16802
89. Nathaniel Stetson, Manager
Savannah River Operations Office
P.O. Box A
Aiken, South Carolina 29801
90. Robert J. Wright, Jr.
Chief Geologist, URE
Department of Energy - Headquarters
Washington, D. C. 20545
91. D. S. Robson
Biometrics Unit
Department of Plant Breeding
Cornell University
Ithaca, New York 14850
- 92-93. Technical Information Center
Department of Energy
P.O. Box 62
Oak Ridge, Tennessee 37830
- 94-193. Technical Library
Bendix Field Engineering Corporation
P. O. Box 1569
Grand Junction, Colorado 81501