MINIMUM HELLINGER DISTANCE ESTIMATION

FOR NORMAL MODELS

by

Paul W. Eslinger
Battelle Pacific Northwest Laboratory

Wayne A. Woodward
Southern Methodist University

Technical Report No. SMU/DS/TR-246

Department of Statistical Science

September 1990

# MINIMUM HELLINGER DISTANCE ESTIMATION FOR NORMAL MODELS

Paul W. Eslinger

and

Wayne A. Woodward

## AUTHOR'S FOOTNOTE

ABSTRACT

A robust estimator introduced by Beran (1977a, 1977b) which is based on the minimum Hellinger distance between a projection model density and a nonparametric sample density is studied empirically. An extensive simulation provides an estimate of the small sample distribution and supplies empirical evidence of the estimator performance for a normal location-scale model. Empirical robustness is also investigated, with performance competitive with that obtained from M-estimators and Cramér-von Mises minimum distance estimators. The minimum Hellinger distance estimator is shown to be an exception to the usual perception that a robust estimator cannot achieve full efficiency. Beran also introduced a goodness-of-fit statistic, $H^2$, based on the minimized Hellinger distance between a member of a parametric family of densities and a nonparametric density estimate. We investigate the statistic H (the square root of $H^2$) as a test for normality when both location and scale are unspecified. Empirically derived critical values are given which do not require extensive tables. The power of the statistic H is compared with four other widely used tests for normality.

Key Words:     Minimum distance; Robustness; Efficiency; Hellinger distance, Influence curve, Nonparametric; Goodness of Fit; Power.

# 1. INTRODUCTION

Robust estimators are those which are insensitive to small deviations from the assumptions, usually at the expense of not being optimal at the true model. Bickel (1978) describes robustness as "paying a price in terms of efficiency at the (true) model in terms of reasonably good maximum MSE over the neighborhood." Beran(1977a, 1977b) introduced a minimum distance estimator based on the Hellinger distance between a member of a parametric family of densities and a nonparametric density estimator. This estimator, called the minimum Hellinger distance estimator (MHDE), was shown by Beran and also by Stather (1981), under suitable conditions, to be consistent, asymptotically normal and asymptotically fully efficient. Tamura and Boos (1986) studied the MHDE in the multivariate setting. The theoretical results obtained by all of these authors indicate that the MHDE plays a special role in the issue of efficiency versus robustness in that it obtains robustness without sacrificing efficiency at the true model. However, the strength of the Hellinger metric, and the fact that the MHDE has an unbounded influence curve, causes some concern that the actual robustness of the MHDE is minimal. Although Beran, Stather, and Tamura and Boos provided some limited empirical evidence concerning the performance of the estimator, the available numerical results are limited. In this article we present an extensive numerical examination of the MHDE in the univariate setting in order to provide a better understanding of its potential usefulness.

Beran (1977b) suggested using the square of the minimized Hellinger distance, $H^2$, as a statistic for testing the goodness-of-fit of the parametric family. Beran concluded that the statistic $H^2$ is relatively insensitive to minor failures such as a few outliers. Bickel (1978) claims, apparently based on Beran's comment, that a goodness-of-fit test using the Hellinger distance does not have detecting power outside the Hellinger neighborhood. In this paper we reexamine the use of the Hellinger metric for purposes of obtaining a goodness-of-fit statistic in order to better understand the power attainable using this approach.

In Section 2 we provide background material concerning the MHDE. Section 3 is devoted to implementation issues for computation of the MHDE. Density estimation and numerical maximization are discussed and guidelines are given for calculating the MHDE. An extensive empirical study described in Section 4 investigates the robustness and small sample properties of the estimator and compares the MHDE with maximum likelihood, minimum distance and M-estimators. In Section 5 we propose the use of the statistic H, the square root of Beran's $H^2$, and we discuss motivation for its use and the derivation of critical values. In Section 6 we present the results of a simulation study comparing the power of H with four commonly used tests for normality for a wide range of alternative distributions.

## 2. BACKGROUND AND ESTIMATOR DEFINITION

Let $X_1$, $X_2$, . . ., $X_n$ denote a random sample from a population with distribution function G, and let $G_n$ denote the empirical distribution function, i.e.

$$G_n(t) = n^{-1}\left[\sum_{i=1}^{n} I(X_i \leq t)\right],$$

where I denotes the indicator function. Also, let $\left\{\mathcal{F}_\theta, \theta \varepsilon \Theta\right\}$ denote a family of distributions, called the projection model, involving the vector valued parameter $\theta$. A minimum distance estimate of $\theta$ is usually defined to be the value of $\theta$ for which the distance between $G_n$ and $F_\theta$ is minimized, where the distance is based on a measure of distance between distribution functions.

The Hellinger distance between two absolutely continuous distributions F and G is the distance between the square roots of the densities f=dF/d$\mu$ and g=dG/d$\mu$ defined by

$$H(f, g) = \left[\int \left(f^{\frac{1}{2}} - g^{\frac{1}{2}}\right)^2 d\mu\right]^{\frac{1}{2}}. \tag{2.1}$$

The Hellinger distance is independent of the choice of measure, so we shall use Lebesgue measure. It will be useful to note that minimizing H(f,g) in (2.1) is equivalent to maximizing $\int f^{\frac{1}{2}}(t)\, g^{\frac{1}{2}}(t)\, dt$. The MHDE is defined in terms of a functional T over the set of densities. Specifically, for a density g, we define T(g) as the value of the parameter $\theta$ which minimizes the distance between g and $\mathcal{F}_\theta$, i.e.

$$H(f_{T(g)}, g) = \min_{\theta \epsilon \Theta} H(f_\theta, g). \tag{2.2}$$

A MHDE of $\theta$ is the random variable $T(g_n)$ where $g_n$ is a suitable nonparametric density estimator based on the sample. If there is not a unique solution to (2.2) then $T(g_n)$ will denote any one of the minimizing values. We base our implementation of the MHDE on kernel density estimators of the form

$$g_n(y) = \left(nh_n\right)^{-1} \sum_{i=1}^{n} w\left[\left(y - X_i\right)/h_n\right] \tag{2.3}$$

where w is a density on the real line and $\{h_n\}$ is a sequence of constants which converge to 0 at an appropriate rate.

Beran gives conditions which guarantee the existence of MHD estimators for $\Theta$ compact and discusses the extension of the result for noncompact $\Theta$. Tamura and Boos (1986) prove consistency and asymptotic normality of $T(g_n)$ under conditions which include the two parameter normal family. Their asymptotic result requires that $g_n$ be a kernel density estimator with nonrandom $h_n$.

## 3. IMPLEMENTATION DETAILS

In this section we discuss the major steps of density estimation and numerical maximization in the evaluation of the MHDE. It should be noted that throughout the paper we will use the normal projection model with both location and scale unknown.

### Density Estimation

We follow Beran and choose the Epanechnikov kernel (Epanechnikov, 1967) as the kernel density estimator for the MHDE because of its optimal properties in density estimation. The Epanechnikov kernel has the form:

$$w(x) = .75(1-z^2), \qquad -1 \le z \le 1 \qquad (3.1)$$
$$= 0, \qquad \text{elsewhere.}$$

The sequence of constants, $\{h_n\}$, must be chosen to complete the definition of the kernel density estimator. The optimal sequence (see, for example, Schucany and Sommers, 1977) based on estimating the density $g(y)$ at the mean of the normal distribution is

$$h_n = \sigma \ (15\sqrt{2\pi}/n)^{0.2} = 2.161\sigma n^{-0.2}. \qquad (3.2)$$

The optimization criteria used is to minimize $E\left\{\int \left[g(x) - g_n(x)\right]^2 dx\right\}$. This sequence yielded both a bias and a large variance in scale estimation for the MHDE, apparently not emphasizing the tails of $g_n$ as much as desired for optimal MHDE performance.

It is not evident that a simple function involving g and $g_n$ can be minimized to yield optimal $h_n$ values for the MHDE, although the expression given above does suggest a functional form for the dependence of $h_n$ on the variance of g. Because analytic efforts to choose an appropriate form for $h_n$

have been unfruitful, an empirical study was conducted to determine a functional form for $h_n$ for a normal location-scale model. Based on the form for $h_n$ in (3.2), we chose to treat $h_n$ as the product of a scale estimator, $s_n$, and a sequence of constants, $c_n$. Throughout this article, the sequence $\{s_n\}$ of scale estimators is given by $s_n = $ SMAD where SMAD is the sample median absolute deviation given by SMAD = (median $|X_i - m|$)/0.6745, where m is the sample median. The $c_n$ sequence was chosen to yield an unbiased MHDE scale estimate. This approach did not cause any significant degradation in the performance of the location estimator because it is relatively insensitive to changes in $c_n$. Values for $c_n$ were obtained for 15 different sample sizes in the range 20 to 1000, using 5000 data sets for each sample size. The entire process of calculating the $c_n$ values was repeated using a different sequence of random numbers to allow examination of the variability in the procedure. A (log-linear) model of the form $c_n = kn^p$ was fit to the resulting values yielding the equation

$$c_n = 2.283 \; n^{-0.287} \; .$$ (3.4)

The $R^2$ for the fit was 0.9937. The replications at each sample size provided the opportunity to do a lack of fit test. The lack of fit test was not significant at the 0.05 alpha level, thus the functional form chosen is a reasonable approximation to the unknown true form.

<u>Numerical Maximization</u>

Calculation of the MHDE requires finding the maximum of

$$\int f^{\frac{1}{2}}(t) \; g_n^{\frac{1}{2}}(t) \; dt$$ (3.5)

with respect to two parameters, the mean and the standard deviation. The iterative, quadratically convergent, Gauss-Newton method described by Beran (1977b) was implemented for numerical investigations. This method finds the simultaneous zero's of the partial derivatives of (3.5) with respect to the parameters being estimated.

A composite Gauss quadrature integration rule was used to evaluate the integrals. The range of the integrals was divided into 75 subintervals of equal length and then each subinterval was integrated using a 4 point Gauss quadrature rule. Because the estimator maximizes (3.5) and the Epanechnikov kernel was used, the integrand is nonzero only over the support of $g_n(x)$, i.e., the interval $\left(X_{(1)} - c_n s_n, \; X_{(n)} + c_n s_n\right)$ where $X_{(1)}$ and $X_{(n)}$ are the smallest and largest sample values respectively.

Normal termination of the iterative solution to (3.5) occurred if both location and scale estimates differed by less than $10^{-4}$ from the estimate on the preceding step. Tests with a larger number of integration steps indicated that an approximate accuracy of $10^{-4}$ was being obtained in the solutions for location and scale.

## Sensitivity to Starting Values

The sensitivity of the MHDE to starting values was investigated by generating 1000 samples of size 40 from the standard normal distribution and counting the number of acceptable solutions obtained using different starting values. The starting location values ranged from $-1$ to 1 and the starting standard deviation values ranged from 0.5 to 2. Convergence percentages ranged from 6.2% at the starting location, standard deviation pair $(-1, 0.5)$ to 99.9% at the starting pair $(0,1)$. Although the procedure shows to be sensitive to starting values, convergence occurred 100% of the time at this sample size when the initial values used were the median for location and SMAD for scale. These two values, denoted by IV, were used as initial values by the iterative MHDE routines and two other iterative estimators described subsequently.

## Other Estimators

Three other estimators were evaluated for comparison with the MHDE. Two of the estimators are representative of the types of robust estimators currently in use. The Maximum Likelihood Estimator (MLE) for normal data, i.e., $\overline{X}$ and the sample standard deviation (with divisor n), S, was included because its distribution is known theoretically for all sample sizes. Note that S is a biased estimator.

The second estimator was obtained using the Cramér-von Mises minimum distance technique and is denoted by CVM. The paper by Parr and Schucany (1980) provides a reference on the Cramér-von Mises minimum distance estimation technique. The CVM estimator is obtained by choosing $\theta=(\mu,\sigma)$ to minimize

$$\sum_{i=1}^{n} \left[ F_\theta\left(X_{(i)}\right) - (i-0.5)/n \right]^2 \tag{3.7}$$

where the $X_{(i)}$ are sample order statistics and $F_\theta$ denotes the cumulative normal distribution function. The minimization to find the CVM estimate was accomplished by using the International Mathematics and Statistics Library (IMSL) subroutine ZXSSQ which implements an iterative nonlinear finite difference Levenberg-Marquardt least squares method.

The third estimator was an M-Estimator (MEST) based on work by Huber (1964). Defining the function

$$\Psi_H(t) = \begin{cases} t & \text{if } |t| < k_H \\ k_H \text{sgn}(t) & \text{if } |t| \geq k_H \end{cases}$$

for some constant $k_H > 0$, the M-estimator was obtained by first solving

$$\sum_{i=1}^{n} \Psi_H\left\{\left(X_{(i)} - \mu\right)/\text{SMAD}\right\} = 0 \tag{3.8}$$

for the location estimate $\mu$. Next, letting $Z_i = (X_i - \mu)/\sigma_H$ denote a standardized observation, the scale estimate, $\sigma_H$, was found by solving

$$(n-1)^{-1} \sum_{i=1}^{n} \Psi_H^2(Z_i) = B \tag{3.9}$$

where $B = E[\Psi_H^2(Y)]$ and Y comes from the standard normal distribution. A value of 1.4 was used for $k_H$ for location estimation which is in the range of values shown to perform well in the Princeton Robustness Study (see Andrews, et. al, 1972). To our knowledge, not much is known about the optimal selection of $k_H$ for scale estimation. Our experience indicates that $k_H = 1.4$ produced biased estimates, and thus we used $k_H = 2$ so that only a small amount of trimming is being done.

Both the CVM and MEST procedures are iterative in nature. The initial values used for both procedures were the same as the initial values used in the MHDE iterative solution procedure. In each case, the convergence criteria was set to obtain an accuracy rate of about $10^{-4}$, assuming that standard normal data was being used.

## Computation Time

Computation time for the MHDE, using an older version of the code which employed trapezoidal integration, was measured relative to the other estimators by calling a system clock before and after the call to each estimation subroutine. A new version of the code runs about 20% faster for the MHDE than the results reported here. The program was coded in FORTRAN on a CYBER 760 and the pseudo-random number generators used were from the IMSL software. The clock had an accuracy of approximately 0.01 seconds on each call. Cumulative computation times in seconds are

given in Table 1 for 5000 replications using standard normal data, where each data set was sorted before any estimation was done. The times for the subroutine which calculated the intital values (IV) reflect the use of an ineffecient sort routine which has since been replaced. The MHDE has comparable computation times at sample size 200 relative to the CVM and 800 for the MEST. Computation time for the MHDE increased by only about 25 percent as the sample size increased from 20 to 800. This appears to be due to three effects. First, to evaluate the kernel density at a point y, one has to include only those sample points, $X_i$, which satisfy $| y - X_i | \leq c_n s_n$ . Since $c_n$ decreases with increasing sample size, the number of points which satisfy this condition increases more slowly than the sample size. Second, as the quality of initial estimates (IV) improves with sample size, less iterations were required. At sample size 20, about 98% of the solutions were obtained by the 5th iteration. The same percentage of solutions were reached by 4 iterations at sample size 100, and 3 iterations at sample size 800. Third, the implementation of the MHDE only requires one pass through the data to calculate the kernel density estimate at the grid points for numerical integration. Both CVM and MEST require a pass through the entire data set for every iteration. For very large data sets the MHDE can be much cheaper to compute than the CVM and MEST.

## 4. EMPIRICAL RESULTS

The empirical study reported in this section was designed to investigate the small sample efficiency, small sample distribution, and empirical robustness of the MHDE. The performance of the MHDE is compared with the other estimators described in Section 3.

### Small Sample Efficiency

An empirical measure of the efficiency of an estimator relative to the Maximum Likelihood estimator is obtained from the ratio of MSE's i.e., E=MSE(MLE)/MSE(MHDE). A standard error estimate for the efficiencies was obtained from the approximate formula (Taylor Series with 2 terms) for the variance of a ratio of dependent random variables. The efficiencies of the four robust estimators under consideration are given in Table 2 for a range of sample sizes. The missing entries were not computed because of long processing times. Standard error estimates are given in parentheses after the efficiency values. The results show that the MHDE obtains high efficiency for small sample sizes and dominates the estimators IV, CVM, and MEST with respect to efficiency at the true model. The efficiency of the MHDE for location estimation is higher than that for scale estimation. An efficiency of 0.98 is attained for location estimation at sample size 40, while scale estimation requires sample size 700 to obtain the same efficiency. The corresponding asymptotic values (where known) are included on

the line headed by "∞".

## Small Sample Distribution

The simulation runs used to obtain the empirical efficiencies also yielded an empirical description, given in Table 3, of the small sample distribution of the MHDE. In the empirical comparison, the MHDE has a larger variance, but less bias (as expected from the choice for $c_n$), than the MLE for the scale component, while the location components do not appear to differ appreciably. Location and scale estimates are known to be independent for the MLE for all sample sizes, while the relationship is unknown for the MHDE. These results indicate that location and scale estimates for the MHDE have at most a low correlation for sample sizes as small as 20. The low, or possibly nonexistent, correlation between location and scale could be anticipated because $H(f, g_n)$ is invariant under location and scale changes. Correlations between the estimators for one simulation run using standard normal data sets of size 40 are provided in Table 4. The MLE and MHDE estimates are highly correlated.

## Empirical Robustness

One method of examining the robustness of an estimator is to calculate the Influence Curve (IC), (see Hampel, 1974) with the usual interpretation being that a robust estimator will have a bounded influence curve. A modification of Hampel's definition of the influence curve must be made for the MHDE (Beran, 1977b) because the minimum Hellinger distance function $T(g_n)$ has as its domain the space of densities rather than the space of distribution functions. Let

$$f(x; \theta, \alpha, z) = (1-\alpha) \, f(x;\theta) + \alpha\delta_z(x) \qquad (4.1)$$

for $\alpha\epsilon(0,1)$ and real $z$ where $\delta_z(x)$ is the uniform density on the interval $(z-\Delta, z+\Delta)$ where $\Delta > 0$ is very small. Define first the quotient ($\alpha$-IC),

$$\alpha\text{-IC}(z) = \{T[f(x; \theta, \alpha, z)] - \theta\} \, / \, \alpha \, ,$$

and then the influence curve is defined to be

$$IC(z) = \lim_{\alpha \to 0} \, \alpha\text{-IC}(z) \, . \qquad (4.2)$$

For the normal location-scale model (see Beran (1977b) for details) the MHDE has influence curve

$$IC(z) = [(z-\mu, \{(z-\mu)^2 - \sigma^2 + \Delta^3/3\}/2] . \qquad (4.3)$$

As $\Delta \to 0$ the influence curve of the MHDE becomes identical to the unbounded influence curve of the MLE. Beran (1977b) also shows that for $\alpha \epsilon (0,1)$

$$\lim_{z \to \infty} \alpha\text{-IC}(z) = 0 \qquad (4.4)$$

so the MHDE is robust at $f(x; \theta, \alpha, z)$ against $100\alpha\%$ contamination by gross errors at arbitrary real z, even though the influence curve is unbounded.

Hampel (1974) claims that the use of the $\alpha$-IC (before the limit) is preferable to the use of the influence curve to assess estimator robustness. The limiting form is often used because it is usually easier to evaluate, and it does not depend on $\alpha$. The MHDE is an example of an estimator for which the limiting form does not reliability provide information about the form of the $\alpha$-IC for $\alpha > 0$.

The $\alpha$-IC for the model in (4.1) using the standard normal density for f was obtained by numerical integration and is plotted in Figures 1 and 2 for several values of $\alpha$. The form of the $\alpha$-IC for both location and scale shows that the influence of an extreme value is diminshed to almost zero by the time it is removed by 5 standard deviations from the center of the data. The robustness indicated by the limit (4.4) should then be attainable for a typical data set; it is not just a mathematical anomaly.

An empirical estimate of the $\alpha$-IC for the MHDE and other robust estimators was generated by drawing 1000 replications at sample size 40 from the mixture distribution

$$f(x; \theta, \alpha) = (1-\alpha) f_1(x) + \alpha f_2(x) \qquad (4.5)$$

with $\alpha = 0.025$. The symbol $f_1$ denotes a standard normal distribution and $f_2$ denotes a normal distribution with mean d in the interval [0, 5] and standard deviation $\sigma_2 = 0.05$. This density differs from the density in (4.1) but maintains the concept of "near" point contamination. Because of the similarity in the $\alpha$-ICs for different values of $\alpha$ in both Figure 1 and Figure 2, a single value of $\alpha$ was used here. Figure 3 shows the estimate of the $\alpha$-IC for location and Figure 4 shows the estimate of the $\alpha$-IC for scale. The plots indicate that the contamination has the maximum influence on the MHDE at about d = 3 and then decreases. The $\alpha$-IC's for the IV, MEST and CVM estimators appear to reach

an asymptote around d = 2. The implication is that the MEST and CVM would be more robust than the MHDE against moderate contamination while the MHDE performs better near the true model, and also when there are a few extremely wild points.

The robustness of the MHDE is displayed empirically in Figures 3 and 4 but is guaranteed theoretically only when the sampled data is within a Hellinger neighborhood of the projection model (Staudte, 1980). Using the normal mixture model (4.5), the Hellinger neighborhood, within which one would expect the MHDE to be robust, is the region satisfying

$$\int f_1^{\frac{1}{2}}(x) \left[(1-\alpha)\, f_1(x) + \alpha f_2(x)\right]^{\frac{1}{2}} dx \leq \alpha^2/n \tag{4.6}$$

The radius of the Hellinger neighborhood (largest shift possible in the mean, d, of $f_2$ while still satisfying eq. 4.6) is given in Table 5 when $n = 40$ and $\alpha = 0.025$ for different values of $\sigma_2$. The data density used to generate the empirical $\alpha$-IC displayed in Figures 3 and 4 is far outside the Hellinger neighborhood. Thus, the MHDE exhibits robustness against alternatives which are outside a Hellinger neighborhood.

Table 5. Hellinger Neighborhood Radius for the Mixture of Normals Model

| $\sigma_2$ | 1.0 | 0.9 | 0.8 | 0.7642 | ≤0.7642 |
|---|---|---|---|---|---|
| radius | 0.3093 | 0.3022 | 0.1863 | 0.0000 | none |

In the examination of the robustness of the MHDE, we also considered three other simulation models: the Student's t distribution with 2 and 4 degrees of freedom and the Laplace (Double Exponential) distribution. Simulation comparisons for location efficiency with respect to the MLE for these models are given in Table 6, using the format of Table 2, for sample sizes 20, 40, 100 and 400 based on 1000 replications. For these three models the MHDE is seen to be robust relative to the MLE, but CVM, MEST (and often IV) obtain slightly higher efficiencies.

The maximum Hellinger distance between any two densities can be seen to be $\sqrt{2}$, and the Hellinger topology completely separates the sets of densities which are continuous from those which are discrete in the sense that the Hellinger distance between a continuous density and a discrete density is $\sqrt{2}$. This prompted an examination of the performance of the MHDE when the sample data is quantized. Tests were run on samples with sizes ranging from 20 to 800 where standard normal data

was rounded to the nearest .01, .05, .1 and .2. The performance of the MHDE appeared to be unaffected by the quantization. It appears that smoothing by the density estimator removes most of the grouping effects induced by the quantization.

## 5. DISTRIBUTION OF THE TEST STATISTIC

Beran (1977b) suggested using the square of the minimized Hellinger distance, $H^2$, as a statistic for testing the goodness-of-fit of the parametric family, in our case the two parameter normal. Beran (1977b) showed that the limiting null distribution of $H^2$, using a sample of independent observations of size n, is normal with mean $3R_n/20nc_n$ and variance $167R_n/(3080n^2c_n)$, where $R_n$ denotes the sample range. His result is essentially based on the following major assumptions (Theorem 8, Beran 1977b): 1) the parameter space is a compact subset of $\mathbf{R}^p$, 2) the support of the projection model is a closed interval on the real line, and 3) the Epanechnikov kernel density estimator is used. The method of proof for Beran's theorem does not extend to the situation where there is an infinite support for the projection model density.

The limiting null distribution of H, under the above conditions, is shown by Eslinger (1983) to be normal with mean $[3R_n/(20nc_n)]^{1/2}$ and variance $167/(1848n)$. This result follows from an application of Theorem A in Serfling (1980, p.118) to the distribution obtained by Beran. Beran suggested using critical values obtained from the limiting null distribution of $H^2$ for the small sample test for normality when location and scale were unspecified. Actually, even the large sample results do not apply in this situation since the support of the normal density is not a compact interval, and Beran noted that the accuracy of this approximate application was not known. The small sample distribution of H was studied by Eslinger (1983). Some small sample statistics from that study are reproduced in Table 7 where it can be seen that the small sample statistics for H approach the values for normality much faster than those of $H^2$. In the table the mean values are standardized by subtracting the asymptotic mean using the expected sample range under normality, $E(R_n)$, rather than $R_n$. Expected sample ranges for normal samples have been tabled extensively by Pearson and Hartley (1958). They also can be computed accurately using a FORTRAN algorithm such as the one given by Beasley and Springer (1977). The standard deviation values are standardized by dividing by the asymptotic standard deviation. The current results are slightly different from those shown in Table 7 because a different bandwidth sequence has been selected.

We derived critical values empirically for testing the null hypothesis of normality using the statistic H. The sequence $c_n$ given in (3.4) was used in this empirical study. The critical values reported in this paper are inappropriate for other definitions of $g_n$ and $f_\theta$. The critical values are

presented in a computationally compact form and do not require extensively tabled values. The method used to obtain the critical values was to generate 2000 sets of normal deviates for each of 20 distinct sample sizes in the range 20 to 1000, compute the statistic H, and then estimate the null distribution percentiles for H from the sample percentiles. Three replications of 2000 sets using different random sequences were made at each sample size, resulting in 60 percentile estimates for each $\alpha$. The sample percentiles were based on normal data computed using a linear congruential uniform random number generator (multiplier of 16807 and modulus of $2^{31}$) and a numerical inversion procedure (Griffiths and Hill, 1985) to transform from the uniform distribution to the normal distribution.

Values of the coefficients for a function that yields critical values for the statistic H as a function of alpha level and sample size, n, are given in Table 8. The functional form used to obtain the critical values is

$$H_\alpha = (a_1 + a_2 \, n^{a_3}) \, / \, (b_1 + b_2 n + b_3 n^{b_4}) \, .$$

A similar approach to obtaining critical values for goodness of fit statistics was reported by Stephens (1974). The functional form used here is a modification of the form used by Stephens (1974) which performed very well in our setting. For each $\alpha$, the $R^2$ value for fitting a curve of this type to the 60 percentile estimates was above .99. Also for each $\alpha$, the fitted curve fell within the 95% confidence interval of the corresponding percentile for all sample sizes. The $\alpha$ values are for the upper $100(1-\alpha)$ percentiles of the null distribution of H under normality. A one tailed test is appropriate for H since large values of the test statistic indicate a poor match between the projection model and the nonparametric density estimator. A test statistic value of 0 would indicate an exact fit by a member of the projection model. The accuracy of the functional form has not been verified for sample sizes over 1000.

Critical values could be obtained for larger sample sizes using the limiting null distribution of H. The method is to employ the form of the limiting null distribution for H under the assumption that the data come from a distribution which has a compact range, except that as in Table 7, the sample range, $R_n$, is replaced by the expected sample range under normality. This approach yields critical values which for sample sizes over 200 appear to be very similar to the empirically derived values. For smaller sample sizes investigated in the range 20 - 200, there was up to a 5% difference in the critical values given by the two methods. The true alpha level of the test using this approximation has not been examined.

## 6. POWER CONSIDERATIONS

### Detectable Alternatives

Bickel (1978) examined the neighborhoods within which goodness-of-fit statistics do not have detection capabilities. He showed that the Hellinger neighborhood was a subset of the neighborhood of the Kolmogrov-Smirnov (K-S) statistic, indicating that the H test should detect a broader class of model deviations than the K-S test. However, as mentioned in Section 1, Bickel also claims that a goodness-of-fit test using the Hellinger distance does not have detecting power outside the Hellinger neighborhood.

An indication of possible lower power of the MHDE compared to other statistics is that the convergence rate of the nonparametric density estimator to the true density under the Hellinger metric is $O(n^{-\frac{1}{2}+\Delta})$ where $\Delta > 0$ depends on the $c_n$ value used (Staudte, 1980). The convergence rate in the Kolmogorov metric of the empirical distribution function to the true distribution function is $O(n^{-\frac{1}{2}})$, so the K-S test converges slightly faster than the Hellinger metric test.

Staudte (1980) notes that the stronger the metric, the more sensitive the goodness-of-fit test based on the metric. The Hellinger metric is stronger than all the other metrics currently used for goodness-of-fit tests based on a minimum distance philosophy, indicating that the H statistic should provide a powerful test.

Theoretical arguments do not give clear indication of the performance of H in comparison with other goodness-of-fit statistics. The empirical studies which follow give an indication of the performance of H in relation to other statistics, and also indicates how the convergence rate of the kernel density estimator in the Hellinger metric effects the power of the test.

### Comparisons With Other Statistics

The power of the H statistic when testing for normality was evaluated by comparing its performance to the test statistics $A^2$ (Anderson and Darling, 1952), R (Filliben, 1975), Cramér-von Mises Minimum Distance $W^2$ (discussed by Stephens, 1974) and, W (Shapiro and Wilk, 1965) as extended to large samples by Royston (1982a, 1982b, 1983).

Eight alternative distributions were used for power comparisons. These distributions were a subset of those used by Stephens (1974) and Filliben (1975) in empirical power studies of tests for normality. The computational accuracy of the current study can be verified by comparing results with

those of Stephens and Filliben. The alternative distributions are listed in Table 9, along with skewness, $\sqrt{\beta_1}$, and tail length, $\beta_2$ and $\tau_2$, measures. The tail length measure, $\tau_2$, is given by $\tau_2 = (1-1/\tau_1)^{.57854}$ where

$$\tau_1 = \Big(G(.9975) - G(.0025)\Big) / \Big(G(.975) - G(.025)\Big)$$

and $G(p)$ is the percent point function of the distribution (see Filliben, 1975).

Table 10 gives a comparison of the power of the tests for normality for the distributions given in Table 9. The entry for each statistic and distribution is the number of rejections expressed as a fraction. The fractions are based on 2000 replications at each sample size. The W statistic generally had the highest power for most sample sizes for the alternatives which are shorter tailed than normal. For the uniform distribution W performed best at all sample sizes. The H statistic gave the second highest power in this case for n = 100, but its performance was somewhat below that associated with the other statistics for n $\leq$ 40. For the triangular distribution, the empirical power associated with all of the statistics was low for n $\leq$ 40. For n = 100, H had the second highest power, with the highest power again being associated with W. When the alternative distributions were symmetric with longer tails than the normal distribution, the R statistic generally had the highest power, with the power of the H statistic being quite competitive with all of the statistics considered. For skewed alternatives, the W statistic generally performed slightly better than the other statistics. For the Weibull(2) distribution the H statistic gave the lowest power for n $\leq$ 40 but had the second highest power for n = 100. For n $\leq$ 20 the H statistic had the smallest power for the exponential and chi-square(2) alternatives. For n $\geq$ 40, however, the power associated with H for these alternatives was at least .97. In general, the H statistic gives results which are competitive with the other four statistics.

Beran (1977b) generated a realization of length n=40 from a N(0,1) distribution. He examined the effect on the parameter estimates of varying one of the observations, $X_{22}$. He also investigated the sensitivity of his goodness-of-fit test based on $H^2$ to variations in the one data value. We employed the H statistic on Beran's data and for the original set of 40 observations, the H test had a value of 0.1217, (we use a different $c_n$ value than Beran did) which was smaller than the upper 5 percent critical value of 0.2106, indicating nonrejection of normality. None of the other four tests rejected the null hypothesis of normality at this alpha level. When the value of $X_{22}$ was changed from $-0.0192038$ to 10.0, the H test had a value of 0.1963, again indicating nonrejection of normality. However, all four of the other tests rejected the null hypothesis of normality. These results are consistent with Beran's results which led him to conclude that his test is insensitive to a few gross outliers.

A more general situation was devised to test the sensitivity of H to a small percentage of gross outliers. The test data used 1000 samples of size 100 from the normal mixture density with $100(1-\alpha)$ percent from the standard normal density and $100\alpha$ percent from the normal density with mean 5.0 and unit variance, with the randomized $\alpha = 0.01$. The empirical powers for the test statistics were 0.362 for $A^2$, 0.309 for $W^2$, 0.591 for R, 0.362 for H, and 0.458 for W. This case shows that the statistic H can detect the situation where there are a few gross outliers, though with lower power than two of the other statistics considered.

## 8. CONCLUSIONS

This article discusses a practical implementation of the minimum Hellinger distance estimator suggested by Beran (1977b). The choice of a kernel density estimator was discussed and a practical choice of bandwidth parameter was obtained. The computation time of the MHDE was shown to be high for small samples, but better than competing robust estimators for samples sizes on the order of several hundred. The MHDE was shown to be highly dependent on starting values, though the starting values suggested by Beran resulted in convergence of the iterative procedure a high percentage of the time. While calculation of the MHDE requires more computer time than the other robust estimators considerd for smaller sample sizes, it is shown to be computationally faster than other robust estimators for very large sample sizes primarily due to the fact that the MHDE requires only one pass through the data.

The small sample distribution of the MHDE from the normal location-scale model was studied empirically and compared to the small sample distribution of the maximum likelihood estimator. The distribution of the MHDE appears to have uncorrelated location and scale estimates. There were no models studied where the MHDE did not exhibit some robustness properties. When the sampled densities had extremely heavy or extremely light tails, MEST and CVM generally had slightly higher efficiency than the MHDE. Near the projection model the MHDE tended to dominate the other robust estimators. It also dominated in the situation of a few extreme wild shots. The MHDE demonstrated unexpected empirical robustness far outside Hellinger neighborhoods of the projection model.

Critical values have been obtained which allow use of the H statistic for testing a null hypothesis of normality, and a functional form for the critical values allows application for any sample size. Computation of the H statistic requires the computation of the minimum Hellinger distance estimates for location and scale. Once the estimation process is done, however, the value of H is relatively easy to obtain. This two step procedure suggests a "natural" use for H in the setting of minimum distance estimation. The H statistic is shown to provide a test for normality which is

competitive with the tests based on the other test statistics considered especially when the sample size is at least 40.

The H statistic was suggested by Beran (1977b) as a goodness-of-fit test which was insensitive to a few gross outliers, hence providing a reasonable test to determine if a robust model for normality was appropriate. The current study suggests that the H statistic is quite sensitive to model deviations and therefore does not provide an answer to the question of the appropriateness of a robust model.

# REFERENCES

Anderson, T. W. and Darling, D. A. (1952), "Asymptotic Theory of Certain 'Goodness of Fit' Criteria Based on Stochastic Processes", <u>Annals of Mathematical Statistics</u> <u>23</u>, 193-212.

Beasley, J. D. and Springer, S. G. (1977), Algorithm AS 111, "The Percentage Points of the Normal Distribution", <u>Applied Statistics</u> <u>26</u>, 118-121.

Beran, Rudolf (1977a), "Robust Location Estimates", <u>The Annals of Statistics</u> <u>5</u>, 431-444.

Beran, Rudolf (1977b), "Minimum Hellinger Distance Estimates for Parametric Models", <u>The Annals of Statistics</u> <u>5</u>, 445-463.

Bickel, P. J. (1978), "Some Recent Developments in Robust Statistics", Paper presented to the Fourth Australian Statistical Conference.

Boos, Dennis D. (1981), "Minimum Distance Estimators for Location and Goodness of Fit", <u>Journal of the American Statistical Association</u> <u>76</u>, 663-670.

Epanechnikov, V. A. (1969), "Non-parametric Estimation of a Multivariate Probability Density", <u>Theory of Probability and its Application</u> <u>XIV</u>, 153-158.

Eslinger, Paul (1983), <u>Minimum Hellinger Distance Estimation</u>, Ph.D. Dissertation, Statistics Department, Southern Methodist University, Dallas, Texas.

Filliben, J. J. (1975), "The Probability Plot Correlation Coefficient Test for Normality", <u>Technometrics</u> <u>17</u>, 111-117.

Griffiths, P. and Hill, I.D. (Editors). *Applied Statistics Algorithms.* Ellis and Horwood, Ltd.: Chichester, England.

Hampel, Frank R. (1974). "The Influence Curve and Its Role in Robust Estimation", Journal of the Americal Statistical Association 69, 383-393.

Huber, P. J. (1964), "Robust Estimation of a Location Parameter", Annals of Mathematical Statistics 35, 73-101.

Parr, W. C. and Schucany, W. R. (1980), "Minimum Distance and Robust Estimation", Journal of the American Statistical Association 75, 616-624.

Pearson, E. S. and Hartley, H. O. (1958), Biometrika Tables For Statisticians, Table 27, Cambridge University Press.

Royston, J. P. (1982), "An Extension of Shapiro and Wilk's W Test for Normality to Large Samples", Applied Statistics 31, 115-124.

Royston, J. P. (1982), Algorithm AS 181, "The W test for Normality", Applied Statistics 31, 176-180.

Royston, J. P. (1983), Correction to Algorithm AS 181, "The W test for Normality", Applied Statistics 32, 224.

Schucany, W. R. and Sommers, John P. (1977), "Improvements of Kernel Type Density Estimators", Journal of the American Statistical Association 72, No. 358, 428-433.

Serfling, Robert J. (1980), Approximation Theorems of Mathematical Statistics, John Wiley & Sons, New York.

Shapiro, S. S. and Wilk, M. B. (1965), "An Analysis of Variance Test for Normality (Complete Samples)", Biometrika 52, 591-611.

Stather, C. R. (1981), <u>Robust Statistical Inference Using Hellinger-Distance Methods</u>, Ph.D. Dissertation, Department of Mathematical Science, School of Physical Science, La Trobe University, Australia.

Staudte, Robert G. (1980), <u>Robust Estimation</u>, Queen's Papers in Pure and Applied Mathematics, No. 53, Queen's University, Kingston, Ontario, Canada

Stephens, M. A. (1974), "EDF Statistics for Goodness of Fit and Some Comparisons," <u>Journal of the American Statistical Association</u> <u>69</u>, No. 347, 730-737.

Tamura, Roy N. and Boos, Dennis D. (1986), "Minimum Hellinger Distance Estimation for Multivariate Location and Covariance," <u>Journal of the American Statistical Association</u> <u>81</u>, 223-229.

**Table 1.** Cumulative Computation Times (sec) for 5000 Replications Using a Sorted Data Set

| Estimator | Sample Size | | | | | |
|---|---|---|---|---|---|---|
| | 40 | 70 | 100 | 200 | 400 | 800 |
| IV | 6.7 | 17.6 | 34.2 | 130.6 | 521.9 | 2033.2 |
| MLE | .6 | .7 | .9 | 1.6 | 3.0 | 5.6 |
| MHDE | 449.0 | 440.2 | 444.7 | 461.1 | 504.9 | 557.3 |
| CVM | 83.5 | 127.8 | 172.6 | 319.0 | 608.1 | 1094.9 |
| MEST | 32.6 | 53.4 | 74.4 | 142.3 | 279.8 | 533.5 |

Table 2. Empirical Efficiency Under the Standard Normal Distribution
(Based on 5000 iterations)

| | Estimator | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | IV | | MHDE | | CVM | | MEST | |
| Sample Size | Location | | | | | | | |
| 20 | .692 | (.011) | .976 | (.004) | .915 | (.007) | .944 | (.006) |
| 40 | .668 | (.011) | .982 | (.004) | .914 | (.007) | .950 | (.006) |
| 60 | .648 | (.011) | .991 | (.003) | .923 | (.008) | .964 | (.006) |
| 100 | .650 | (.011) | .990 | (.002) | .915 | (.007) | .954 | (.006) |
| 200 | .640 | (.011) | .991 | (.003) | .913 | (.008) | .958 | (.006) |
| 400 | .638 | (.011) | .993 | (.002) | .913 | (.008) | .953 | (.005) |
| 700 | .638 | (.011) | .990 | (.002) | | | | |
| 1000 | .633 | (.011) | .990 | (.002) | | | | |
| $\infty$ | .637 | | 1.000 | | .914 | | .953 | |
| | Scale (Standard Deviation) | | | | | | | |
| 20 | .399 | (.008) | .823 | (.011) | .632 | (.011) | .881 | (.009) |
| 40 | .390 | (.008) | .902 | (.009) | .659 | (.011) | .891 | (.009) |
| 60 | .392 | (.009) | .929 | (.008) | .669 | (.012) | .891 | (.009) |
| 100 | .369 | (.008) | .933 | (.007) | .645 | (.011) | .879 | (.009) |
| 200 | .367 | (.008) | .957 | (.006) | .647 | (.011) | .878 | (.008) |
| 400 | .372 | (.008) | .970 | (.005) | .659 | (.011) | .898 | (.009) |
| 700 | .371 | (.008) | .979 | (.004) | | | | |
| 1000 | .371 | (.008) | .980 | (.004) | | | | |
| $\infty$ | *** | | 1.000 | | .651 | | *** | |

***   Unknown

**Table 3.** Empirical Comparison of MHDE and MLE Small Sample Distributions for the Standard Normal Distribution Using 5000 Replications

Statistic

| Sample Size | | | min | max | mean | n*var | skew | kurt | corr |
|---|---|---|---|---|---|---|---|---|---|
| 20 | MLE | l | -0.821 | 0.885 | -0.0035 | 0.990 | -0.017 | 3.059 | 0.011 |
| | | s | 0.439 | 1.540 | 0.9617 | 0.493 | 0.148 | 2.973 | |
| | MHDE | l | -0.807 | 0.881 | -0.0040 | 1.015 | -0.016 | 3.047 | 0.014 |
| | | s | 0.200 | 1.694 | 0.9937 | 0.633 | 0.057 | 3.141 | |
| 40 | MLE | l | -0.586 | 0.588 | 0.0022 | 0.996 | -0.027 | 2.897 | -0.014 |
| | | s | 0.606 | 1.494 | 0.9827 | 0.506 | 0.059 | 2.976 | |
| | MHDE | l | -0.593 | 0.615 | 0.0026 | 1.014 | -0.025 | 2.985 | -0.014 |
| | | s | 0.602 | 1.518 | 1.0018 | 0.574 | 0.037 | 2.981 | |
| 100 | MLE | l | -0.365 | 0.337 | -0.0006 | 1.021 | 0.010 | 2.899 | -0.017 |
| | | s | 0.751 | 1.258 | 0.9921 | 0.492 | 0.114 | 2.979 | |
| | MHDE | l | -0.367 | 0.336 | -0.0007 | 1.031 | 0.005 | 2.882 | -0.022 |
| | | s | 0.735 | 1.275 | 1.0006 | 0.534 | 0.099 | 3.020 | |
| 400 | MLE | l | -0.205 | 0.180 | -0.0001 | 1.036 | 0.054 | 3.024 | -0.005 |
| | | s | 0.875 | 1.142 | 0.9981 | 0.499 | 0.040 | 3.042 | |
| | MHDE | l | -0.203 | 0.181 | 0.0000 | 1.044 | 0.057 | 3.007 | -0.005 |
| | | s | 0.867 | 1.144 | 1.0001 | 0.516 | 0.031 | 3.058 | |

l denotes the location (mean) estimate
s denotes the scale (standard deviation) estimate

Table 4. Correlations Between Estimators for Sample Size 40.

| | | Location | | |
|---|---|---|---|---|
| IV | MLE | MHDE | CVM | MEST |
| 1.000 | 0.821 | 0.820 | 0.907 | 0.865 |
| | 1.000 | 0.994 | 0.961 | 0.980 |
| | | 1.000 | 0.962 | 0.980 |
| | | | 1.000 | 0.991 |
| | | | | 1.000 |

| | | Scale | | |
|---|---|---|---|---|
| IV | MLE | MHDE | CVM | MEST |
| 1.000 | 0.605 | 0.748 | 0.870 | 0.669 |
| | 1.000 | 0.962 | 0.827 | 0.960 |
| | | 1.000 | 0.905 | 0.975 |
| | | | 1.000 | 0.894 |
| | | | | 1.000 |

Table 6.  Location Efficiency for the Normal Projection Mode
          Data Distributions

| Sample Size | | Distribution | | |
| --- | --- | --- | --- | --- |
| | | t2 | t4 | D.E. |
| 20 | IV | 5.601 (1.135) | 1.069 (0.047) | 1.486 (0.064) |
| | MHDE | 4.778 (0.965) | 1.220 (0.041) | 1.122 (0.028) |
| | CVM | 5.681 (1.141) | 1.314 (0.042) | 1.410 (0.037) |
| | MEST | 5.268 (1.055) | 1.304 (0.038) | 1.333 (0.033) |
| 40 | IV | 4.818 (0.731) | 1.195 (0.060) | 1.645 (0.116) |
| | MHDE | 4.434 (0.673) | 1.241 (0.048) | 1.152 (0.044) |
| | CVM | 5.340 (0.807) | 1.416 (0.054) | 1.525 (0.066) |
| | MEST | 4.994 (0.753) | 1.378 (0.049) | 1.381 (0.052) |
| 100 | IV | 6.859 (1.124) | 1.147 (0.047) | 1.594 (0.109) |
| | MHDE | 5.645 (0.919) | 1.252 (0.037) | 1.130 (0.040) |
| | CVM | 7.322 (1.187) | 1.419 (0.039) | 1.535 (0.068) |
| | MEST | 6.685 (1.078) | 1.379 (0.035) | 1.421 (0.055) |
| 400 | IV | 5.363 (0.413) | 1.209 (0.050) | 1.971 (0.122) |
| | MHDE | 4.370 (0.316) | 1.226 (0.036) | 1.167 (0.038) |
| | CVM | 5.717 (0.411) | 1.468 (0.043) | 1.613 (0.059) |
| | MEST | 5.222 (0.371) | 1.418 (0.037) | 1.452 (0.046) |

Table 7. Empirical Small Sample Distribution Characteristics

| Statistic | n | Standardized Mean | Standardized Stand. Dev. | Skew | Kurt | E(Rn) |
|---|---|---|---|---|---|---|
| H | 20 | −.01023 | .79397 | 1.415 | 5.891 | 3.73495 |
| $H^2$ | 20 | −.00057 | .33158 | 2.803 | 16.410 | |
| H | 40 | −.00643 | .79043 | 1.043 | 4.838 | 4.32156 |
| $H^2$ | 40 | −.00040 | .30869 | 2.202 | 12.733 | |
| H | 100 | −.00122 | .81866 | .640 | 3.615 | 5.42909 |
| $H^2$ | 100 | .00033 | .30812 | 1.387 | 6.483 | |
| H | 400 | .00215 | .86490 | .342 | 3.086 | 5.93636 |
| $H^2$ | 400 | .00049 | .32398 | .840 | 3.992 | |
| H | ∞ | 0.0 | 1.0 | 0.0 | 3.0 | |
| $H^2$ | ∞ | 0.0 | 1.0 | 0.0 | 3.0 | |

Table 8. Coefficients for Critical Values of H

| α | a1 | a2 | a3 | b1 | b2 | b3 | b4 |
|---|------|------|------|------|------|------|------|
| 0.150 | 0.5309 | 0.2292 | 0.6392 | -0.7079 | 0.2602 | 1.3332 | 0.4614 |
| 0.100 | 0.1040 | 0.1807 | 0.6453 | -2.1490 | 0.2134 | 1.2031 | 0.3668 |
| 0.075 | 0.2607 | 0.1923 | 0.6425 | -1.9988 | 0.2177 | 1.2293 | 0.3900 |
| 0.050 | 0.8026 | 0.2309 | 0.5557 | -1.0948 | 0.1070 | 1.2914 | 0.5361 |
| 0.025 | 0.6244 | 0.2116 | 0.5513 | -1.8918 | 0.0954 | 1.2564 | 0.5018 |
| 0.020 | 0.6553 | 0.2121 | 0.5503 | -1.8935 | 0.0942 | 1.2438 | 0.5014 |
| 0.010 | 0.7574 | 0.2108 | 0.5458 | -1.9138 | 0.0868 | 1.2065 | 0.5048 |
| 0.005 | 0.7580 | 0.1406 | 0.4557 | -1.9551 | 0.0251 | 1.3658 | 0.4257 |
| 0.001 | 0.6323 | 0.2288 | 0.5958 | -2.5915 | 0.1396 | 0.9343 | 0.4913 |

## Table 9

## Alternative Distributions

### Symmetric, shorter tailed than normal

|  | $\sqrt{\beta_1}$ | $\beta_2$ | $\tau_2$ |
|---|---|---|---|
| Uniform | 0.0 | 1.800 | .167 |
| Triangular | 0.0 | 2.400 | .352 |

### Symmetric, longer tailed than normal

|  | $\sqrt{\beta_1}$ | $\beta_2$ | $\tau_2$ |
|---|---|---|---|
| Students t(4) | 0.0 | ---- | .673 |
| Students t(2) | 0.0 | ---- | .810 |
| Cauchy | 0.0 | ---- | .941 |

### Skewed, ordered by power of W

|  | $\sqrt{\beta_1}$ | $\beta_2$ | $\tau_2$ |
|---|---|---|---|
| Weibull (2) | .631 | 3.245 | .464 |
| Exponential | 2.000 | 9.000 | .579 |
| Chi-Square (1) | 2.828 | 15.000 | .631 |

## Table 10

### Empirical Power for Alternative Distributions

#### Sample Size = 10

|  | A$^2$ | W$^2$ | R | H | W |
|---|---|---|---|---|---|
| Uniform | .084 | .078 | .049 | .053 | .084 |
| Triangular | .045 | .043 | .036 | .056 | .041 |
| Student's t(4) | .131 | .126 | .155 | .116 | .140 |
| Student's t(2) | .315 | .306 | .349 | .220 | .318 |
| Cauchy | .627 | .628 | .642 | .517 | .613 |
| Weibull (2) | .088 | .078 | .086 | .063 | .089 |
| Exponential | .438 | .410 | .442 | .282 | .472 |
| Chi-Square (1) | .704 | .674 | .689 | .552 | .733 |

#### Sample Size = 20

|  | A$^2$ | W$^2$ | R | H | W |
|---|---|---|---|---|---|
| Uniform | .184 | .131 | .062 | .037 | .188 |
| Triangular | .045 | .038 | .018 | .027 | .031 |
| Student's t(4) | .258 | .224 | .285 | .220 | .249 |
| Student's t(2) | .543 | .502 | .578 | .465 | .518 |
| Cauchy | .897 | .882 | .900 | .857 | .876 |
| Weibull (2) | .153 | .121 | .137 | .101 | .153 |
| Exponential | .796 | .725 | .789 | .655 | .836 |
| Chi-Square (1) | .975 | .953 | .974 | .922 | .984 |

## Sample Size = 40

|  | A² | W² | R | H | W |
|---|---|---|---|---|---|
| Uniform | .463 | .336 | .294 | .147 | .706 |
| Triangular | .047 | .041 | .013 | .032 | .063 |
| Student's t(4) | .354 | .322 | .464 | .332 | .336 |
| Student's t(2) | .789 | .770 | .826 | .759 | .747 |
| Cauchy | .988 | .988 | .989 | .982 | .978 |
| Weibull (2) | .262 | .209 | .290 | .206 | .340 |
| Exponential | .988 | .966 | .990 | .970 | .998 |
| Chi-Square (1) | 1.000 | .999 | 1.000 | .999 | 1.000 |

## Sample Size = 100

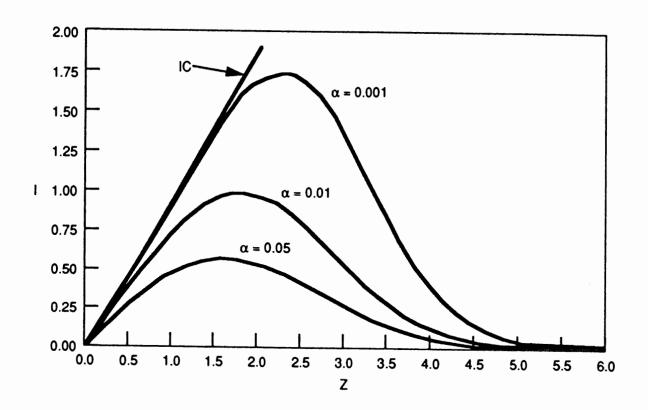|  | A² | W² | R | H | W |
|---|---|---|---|---|---|
| Uniform | .937 | .835 | .942 | .993 | 1.000 |
| Triangular | .071 | .062 | .026 | .105 | .303 |
| Student's t(4) | .643 | .601 | .785 | .627 | .502 |
| Student's t(2) | .982 | .976 | .991 | .980 | .947 |
| Cauchy | 1.000 | 1.000 | 1.000 | .998 | 1.000 |
| Weibull (2) | .591 | .510 | .687 | .713 | .825 |
| Exponential | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Chi-Square (1) | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

Figure 1.

Location IC and Location $\alpha$−IC's at Several Values of $\alpha$
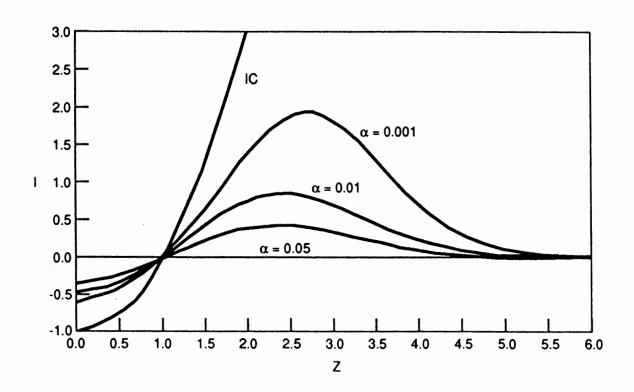
for the Normal Projection Model

Figure 2.

Scale IC and Scale $\alpha$−IC's at Several Values of $\alpha$
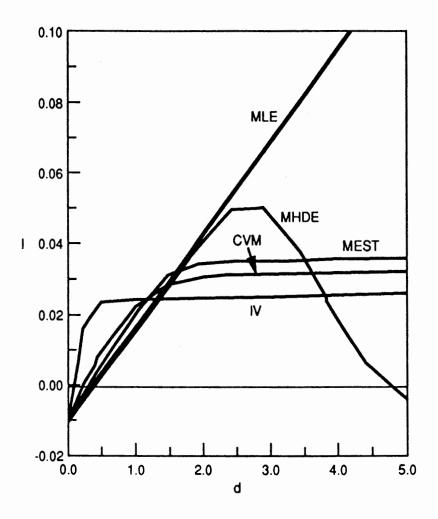
for the Normal Projection Model

Figure 3.

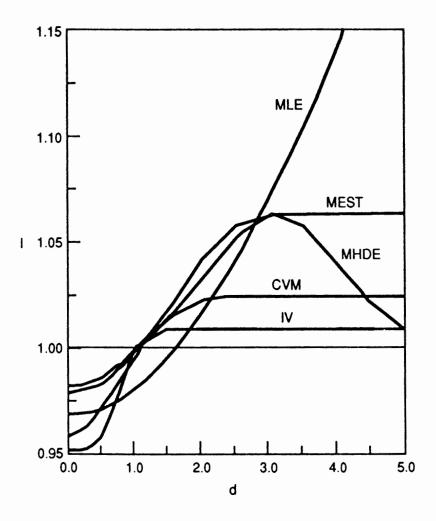Empirical Location $\alpha$−IC's for the Normal Projection Model

Figure 4.

Empirical Scale $\alpha-$IC's for the Normal Projection Model