

August, 1979

LIDS-P-909
(revised)

SOFTWARE FOR EXPLORING DISTRIBUTION SHAPE

David C. Hoaglin
Abt Associates Inc. and Harvard University

and

Stephen C. Peters
Massachusetts Institute of Technology

28 May 1979

*This work was supported in part by Grants SOC75-15702, MCS77-26902 and MCS78-17697 from the National Science Foundation.

ABSTRACT

It is often important to study the shape of continuous distributions that arise in real data or in simulation studies. Some techniques introduced recently by J.W. Tukey offer considerable flexibility in describing and summarizing distribution shape, and they permit direct, resistant fitting to data quantiles. We illustrate their use in two examples, and we present a modularized collection of FORTRAN subroutines implementing these techniques.

keywords

exploratory data analysis
g-and-h-distributions
household income
Housing Allowance Demand Experiment
kurtosis
lognormal distribution
resistance
robustness
simulation
skewness
stable distributions
statistical computing

1. INTRODUCTION

For many statistical models, discussions of robustness are concerned primarily with variations in distribution shape, usually in the neighborhood of the Gaussian (or normal) distribution. The most extensively studied and the best understood of these is the problem of estimating the center of a symmetric unimodal distribution. A high degree of robustness (of efficiency) is possible in this problem, especially in the face of longer-than-Gaussian tails [1]; but Stigler [4], among others, has questioned the connection between real data and the theoretical distributions on which most of the simulation studies have been based. It is clear from the discussion surrounding this issue that all too little empirical information is available on the shapes of distributions encountered in practice in various fields. Some of the discussants of Stigler's paper suggest the routine calculation and collection of sample skewness and kurtosis as measures of distribution shape. We would like to suggest that such moment-based measures are too unreliable to provide a realistic picture of the shape of data. If, for example, a moment-based measure of spread such as the variance is already too sensitive to the presence of isolated outlying observations, then we must recognize that higher-order moments can only be more adversely affected. The more extreme observations in a sample do provide more information about spread (and higher-order properties) than about location in a symmetric population, but it is preferable to use measures other than sample moments in extracting this information.

In this paper we review a class of distributions which is well-suited to the task of exploring distribution shape in real data, and we describe a collection of Fortran routines that we have developed to facilitate its use. Section 2 discusses the distributions and the way in which they parametrize distribution shape. Section 3 presents two examples, and Section 4 describes the software. Section 5 provides some brief concluding discussion. An appendix gives the program listings.

2. THE "g-AND-h-DISTRIBUTIONS"

A distribution shape corresponds to a location-scale family of distributions. That is, two distributions have the same shape if the corresponding random variables, X_1 and X_2 , are related by $X_2 = a + bX_1$, for some constants a and b with $-\infty < a < \infty$ and $b > 0$. A shape may be a subfamily of a much richer family of distributions such as the gamma distributions, the symmetric stable laws, or the Pearson

curves.

In what follows, Y is the "standard" representative of a shape, and other random variables with the same shape are related to Y through $X = A + BY$. The way in which the data determine A and B will be clear when we have presented the particular family of shapes.

Several attributes are desirable in a procedure for probing distribution shape. It should be possible to fit a shape directly to a set of data (without, for example, having to determine parameters by maximum likelihood). The procedure should be resistant; that is, an arbitrary change in a small fraction of the data should produce only a small change in the fitted shape. And, because one generally begins with a Gaussian distribution and measures departures from it, a family of shapes could conveniently be matched to the Gaussian shape in the middle (specifically, at the median). Also, it is advantageous to be able to calculate quantiles for a fitted shape with only modest effort.

With a number of these attributes in mind, Tukey [6] has suggested a two-parameter family of shapes, the "g-and-h-distributions." These are defined in terms of their quantile function relative to the standard Gaussian distribution,

$$Q_{g,h}(z) = \frac{e^{gz} - 1}{g} e^{hz^2/2} \tag{2.1}$$

so that if z_p is the p-th quantile of the standard Gaussian distribution (i.e., $P\{Z \leq z_p\} = p$), $0 < p < 1$, then $Q_{g,h}(z_p)$ is the p-th quantile of a standard g-and-h-distribution (with the specified values of g and h). The parameter g controls asymmetry or skewness, while h controls elongation or the extent to which the tails are stretched (relative to the Gaussian). It is straightforward to show that, when $g=0$, equation (2.1) reduces to

$$Q_{0,h}(z) = z e^{hz^2/2} \tag{2.2}$$

The random variable $Y = Q_{0,h}(Z)$ has a symmetric distribution, and when $h = 0$, it is simply Z. Thus in equation (2.2), $e^{hz^2/2}$ serves as a tail-stretching operator; for $h > 0$, the further a quantile is into the tail, the more it is stretched from its standard Gaussian value. (Negative values of h are possible, but then we must be aware that $Q_{0,h}$ is no longer monotonic when $z^2 > -1/h$.) With g set to zero,

this subfamily is known as the "h-distributions." These distributions have Paretian tails, and the factor of 1/2 in the exponent of equation (2.2) yields a close approximation to the Cauchy distribution at $h \approx 1$.

The other obvious subfamily, the "g-distributions," comes about by setting $h=0$ in equation (2.1). We then have

$$Q_{g,0}(z) = \frac{e^{gz} - 1}{g} . \quad (2.3)$$

If we rewrite this as $Q_{g,0}(z) = [(e^{gz} - 1)/(gz)]z$, we can think of the expression in brackets as a skewing operator. It produces skewness to the right when $g > 0$, skewness to the left when $g < 0$, and no skewness when $g = 0$. From equation (2.3) it is easy to see how these distributions are matched to the standard Gaussian distribution at the median: $Q_{g,0}(0) = 0$ and $Q_{g,0}(z) \approx z$ for z near 0. Also, from the form of $Q_{g,0}$ it is straightforward to see that the shapes in the lognormal family of distributions correspond to g-distributions with positive values of g . Thus the g-distributions are essentially lognormal, but the g-and-h-distributions provide a much greater range of skewness and elongation.

Calculating g and h

To see how one fits these distributions to data (or, as approximations, to other theoretical distributions), it is easiest to begin with the g-distributions. From the data, we require the median, $x_{.5}$, and a set of additional quantiles symmetrically placed about the median. That is, these quantiles come in pairs, x_p and x_{1-p} , for suitable values of p , $0 < p < .5$. In small samples one can work with the full set of order statistics; but in moderate and large samples, we generally use the "letter values" [5], which correspond to taking the integer powers of 1/2 as the values of p .

Recalling that the data quantiles, x_p , are related to the "standard" quantiles, y_p , through $x_p = A + By_p$, it is easy to see that $A = x_{.5}$. Then straightforward algebra allows us to calculate a value of g which exactly fits the spacing of x_p and x_{1-p} about the median. We denote this value by g_p :

$$g_p = \frac{-1}{z_p} \ln \frac{x_{1-p} - x_{.5}}{x_{.5} - x_p} \quad (2.4)$$

where z_p is the corresponding standard Gaussian quantile. Of course, the value

of g_p determined in this way may vary from one value of p to another, and later we discuss a generalization of the g -distributions which allows for systematic variation in g_p . At this stage we concentrate on the simplest form, in which g_p does not depend on p . Equation (2.4) allows us to calculate g_p directly from the data, and working with several such values provides a basis for resistance. We might have as many as ten values of p , each yielding a value of g_p , and one or two unusual values should stand out. As a very simple resistant procedure, we take for our constant value of g the median of the available values of g_p . (More refined summaries are certainly possible.)

Remembering that we are still restricting our attention to g -distributions, we now determine a suitable value for the scale constant, B . A reasonable way to approach this is by means of a quantile-quantile plot (Q-Q plot) [8]. We plot the data quantiles, x_p , against the quantiles of the standard g -distribution, $Q_{g,0}(z_p)$, for the fitted value of g . The slope of this plot is B . This method of finding B provides a check on the appropriateness of using a constant value of g to describe the skewness of the data. Systematic curvature suggests that such a simple model may not be adequate.

We may look more broadly at the problem of fitting a value of g by allowing h to be nonzero in equation (2.1). It is then a simple matter to verify that the elongation factor cancels in calculating g_p , so that equation (2.4) is valid regardless of the value of h . This leaves the task of finding a value of h that is suitable for the data, first when $g=0$ and then when $g \neq 0$.

When $g=0$, the standard quantile function is $Q_{0,h}$ as in equation (2.2). The difference between the upper and lower p -th quantiles of the data is thus

$$x_{1-p} - x_p = -2B \times Q_{0,h}(z_p), \quad (2.5)$$

and rearranging this and taking logarithms yields

$$\ln\left(\frac{x_{1-p} - x_p}{-2z_p}\right) = \ln B + h\left(\frac{z_p^2}{2}\right). \quad (2.6)$$

(The square of the quantity $(x_{1-p} - x_p)/(-2z_p)$ is known as the 100 p % pseudo-variance [1].) Because both h and the scale constant B are unknown, we use a sequence of values of p (such as those yielding the letter values) and plot the left-hand side of equation (2.6) against z_p^2 . This makes reasonably direct use

of the data, provides a basis for resistance, and enables us to judge whether a constant value of h is adequate to describe the data. Because we have assumed symmetry ($g=0$), we could use equations analogous to (2.5) and (2.6) but involving the median and either the upper p -th quantile or the lower p -th quantile. We prefer the first approach stated because it tends to have a slight symmetrizing effect in the face of fluctuations in data.

Now when $g \neq 0$, it is necessary first to fit a value of g (as described earlier) and then to adjust the data toward symmetry before we can make a plot to find h . To illustrate the adjustment step, we work with the median and the upper p -th quantile of the data:

$$x_{1-p} - x_{.5} = \frac{B}{g} (e^{-gz_p} - 1) e^{hz_p^2/2} \quad (2.7)$$

Dividing through by $(e^{-gz_p} - 1)/g$ and taking logarithms puts us back in the situation of equation (2.6), and again a plot should lead us to a suitable value of h .

More General Forms

In discussing the process of choosing values for g and h , we have hinted that constant values may not always be adequate. What are we to do when this appears to be the case? The natural generalization [6] is to regard both g and h (as appropriate) as functions of z^2 , so that the constant values of g and h that we have so far used are only the constant terms of polynomials such as

$$g(z) = g_0 + g_1 z^2$$

and

$$h(z) = h_0 + h_1 z^2 .$$

To explore these possibilities, we need only to plot g_p from equation (2.4) against z_p^2 and to look for curvature in the plot based on equation (2.6).

The greater flexibility that comes from allowing non-constant g and h seems adequate to describe quite a wide range of distributions, both empirical and theoretical. The next section presents one of each to illustrate these methods.

3. TWO EXAMPLES

The first example keeps some aspects of the fitting process simple by

concentrating on h-distributions to provide an approximation for the symmetric stable distribution of index (or characteristic exponent) 1.5. For a particular standardization of the symmetric stable distributions, Fama and Roll [3] provide a table of percentage points. Selected values of these appear as x_{1-p} in Exhibit 1, which also shows the steps of calculation along the lines of equation (2.6). Exhibit 2 plots $\ln(x/z)$ against z^2 . Resistant fitting of linear and quadratic terms in z^2 does a good job of describing the behavior of $\ln(x/z)$, except for the point at $p = .0005$, which falls well below the fitted curve. Out at least as far as $p = .005$, however, the constant $h_0 = 0.111$, $h_1 = 0.038$ and $\ln B = 0.333$ work quite well, yielding

$$1.396ze^{h(z)z^2/2}$$

with

$$h(z) = 0.111 + 0.038z^2$$

as the approximation to the quantile function of this symmetric stable distribution (of course, the location parameter, A , is zero by symmetry). We must, however, return to the point at $p = .0005$ and reconcile the form of $h(z)$ with the known tail behavior of this stable distribution. The h-distributions have Paretian tails, and straightforward calculations indicate that a constant value of h corresponds to a characteristic exponent equal to $1/h$. Thus the stable distribution of this example should yield $h = 2/3$, at least once p is small enough. Using additional quantile values in the range $.01 \geq p \geq .0001$ (made available by W.H. DuMouchel), we have been able to look closer at the relationship between $\ln(x/z)$ and z^2 . In the interval between $p = .005$ and $p = .002$, the curve makes a smooth transition from the quadratic behavior of $h(z)$ to very nearly a straight line corresponding to $h = 2/3$. To describe this combination of initially quadratic and asymptotically linear behavior, we can try a rational function of the form

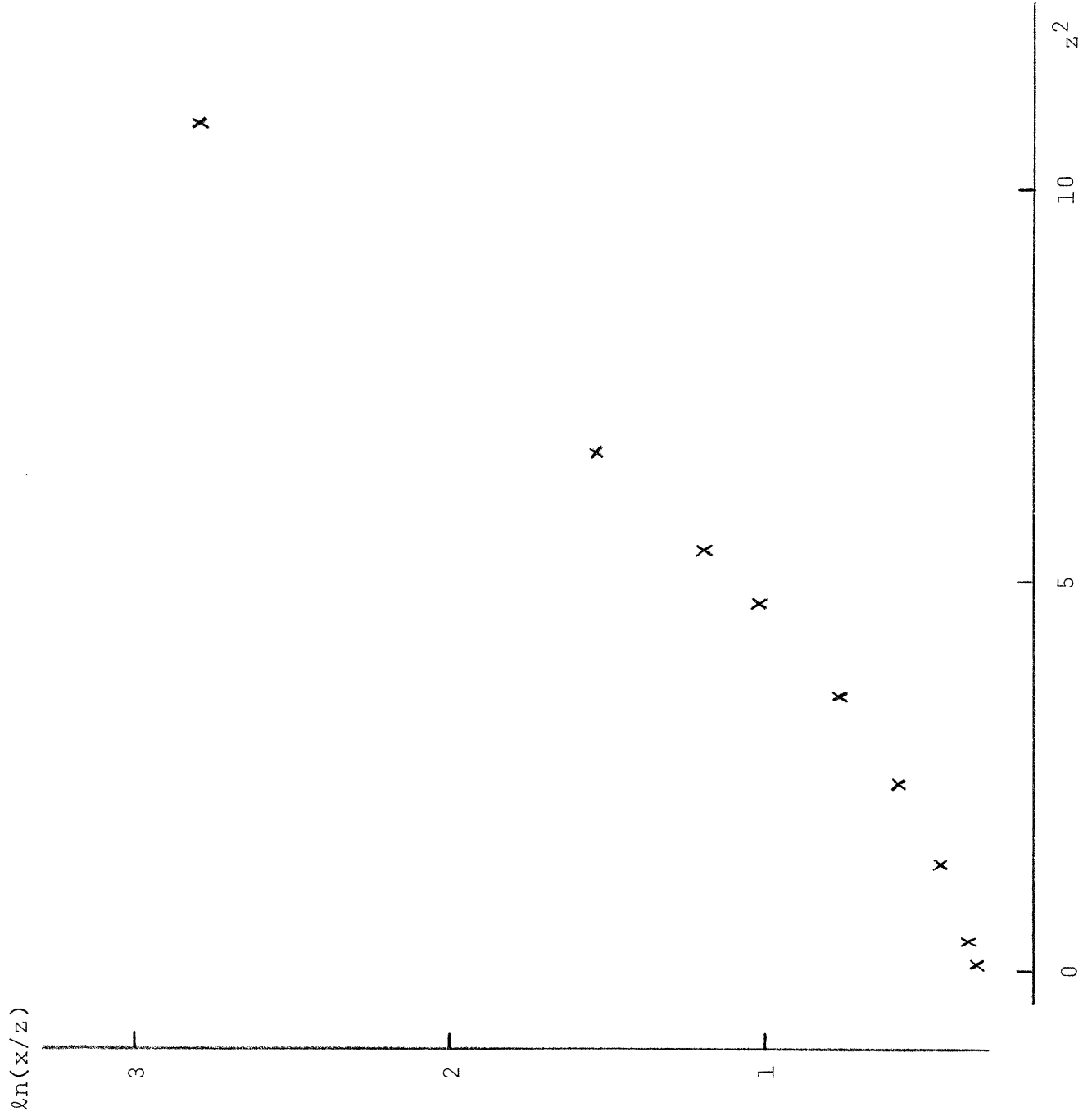
$$\ln\left(\frac{x}{z}\right) = c_0 + \frac{c_1 z^2 + c_2 z^4}{1 + 3c_2 z^2} .$$

Fitting this by nonlinear least squares yields $c_0 = 0.3771$, $c_1 = -0.09279$, and $c_2 = 0.08827$. While the systematic pattern of the residuals suggests that we could gain by increasing the degree (in z^2) of both the numerator and the denominator by 1, the fit of this rational function is excellent,

Exhibit 1. Quantiles of the Symmetric Stable Distribution of Index 1.5 and Calculations for Fitting an h-Distribution

<u>p</u>	<u>x_{1-p}</u>	<u>z_{1-p}</u>	<u>ln(x/z)</u>	<u>z²</u>
.38	0.427	0.3055	0.335	0.093
.26	0.921	0.6434	0.359	0.414
.12	1.837	1.1750	0.447	1.381
.06	2.763	1.5548	0.575	2.417
.03	4.049	1.8808	0.767	3.537
.015	6.043	2.1701	1.024	4.709
.01	7.737	2.3264	1.202	5.412
.005	11.983	2.5758	1.537	6.635
.0005	54.337	3.2905	2.804	10.827

Exhibit 2. Plot of $\ln(x/z)$ against z^2 for the symmetric stable distribution of index 1.5



leaving residuals of magnitude no larger than 0.04. It may be desirable to check the results between $p = .38$ and $p = .5$, but on the whole this rational function provides a good approximation to $\ln(x/z)$ and hence a reasonably accurate approximation to the quantile function of this symmetric stable distribution.

The second example involves both g and h and is based on data from the Housing Allowance Demand Experiment [2], part of the Experimental Housing Allowance Program established by the U.S. Department of Housing and Urban Development. Exhibit 3 shows a letter-value display [5] for the annual incomes (in dollars) reported at enrollment by 994 low-income households in Allegheny County, Pennsylvania. As equation (2.4) indicates, g_p depends on the differences between the median (here 3480) and the corresponding upper and lower quantiles, and these differences form the first two numerical columns of Exhibit 4. The remaining columns of this exhibit complete the calculation of g_p and include the value of z_p^2 . Because the values of g_p decrease steadily as p decreases, we will want to look further than a constant value of g by plotting g_p against z_p^2 . From eye-fitting a straight line to this plot (Exhibit 5), we find that $g_p \approx 0.493 - 0.025z_p^2$. The second point, based on the eighths, departs noticeably from this line, but so far we can offer no simple explanation. This summarizes the skewness, but the possibility of elongation remains.

To pursue elongation in these data, we first adjust for the fitted pattern of skewness, as indicated by equation (2.7) and the subsequent discussion. This leads us to plot the log of the adjusted upper semi-spread,

$$\ln \text{USS}_p^* = \ln \frac{(x_{1-p} - x_{.5})g(z_p)}{e^{-g(z_p)}z_p - 1},$$

against z_p^2 . Exhibit 6 shows the calculations for this, Exhibit 7 has the plot, and the result of fitting a resistant straight line is $7.52 - 0.0168z_p^2$, so that $B = 1845$ and $h = -0.0336$. Thus these income data exhibit slight negative elongation (i.e., the tails are squeezed), once we allow for the skewness. The overall fitted quantile function is

Exhibit 3. Letter-value display for a sample of reported household incomes

#	household income (dollars)
994	
M 497.5	3480
F 249	2412 4944
E 125	1788 6443
D 63	1517 7284
C 32	1248 8350
B 16.5	963.5 8994
A 8.5	727.5 9754.5
Z 4.5	579 10210
Y 2.5	345 10675.5
1	114 10874

Exhibit 4. Calculations for describing skewness in the household-income data

<u>tag</u>	<u>semi-spreads*</u>			$\frac{g_p}{p}$	$\frac{z_p^2}{p}$
	<u>lower</u>	<u>upper</u>	<u>(upper)/lower</u>		
F	1068	1464	1.371	0.468	0.455
E	1692	2963	1.751	0.487	1.323
D	1963	3804	1.938	0.431	2.353
C	2232	4870	2.182	0.419	3.470
B	2516.5	5514	2.191	0.364	4.639
A	2752.5	6274.5	2.280	0.341	5.845
Z	2901	6730	2.320	0.316	7.076
Y	3135	7195.5	2.295	0.288	8.327
X	3366	7394	2.197	0.254	9.593

*median = 3480, lower semi-spread equals median minus lower letter value, and upper semi-spread equals upper letter value minus median.

Exhibit 5. Plot of g_p against z_p^2 for the household
income data

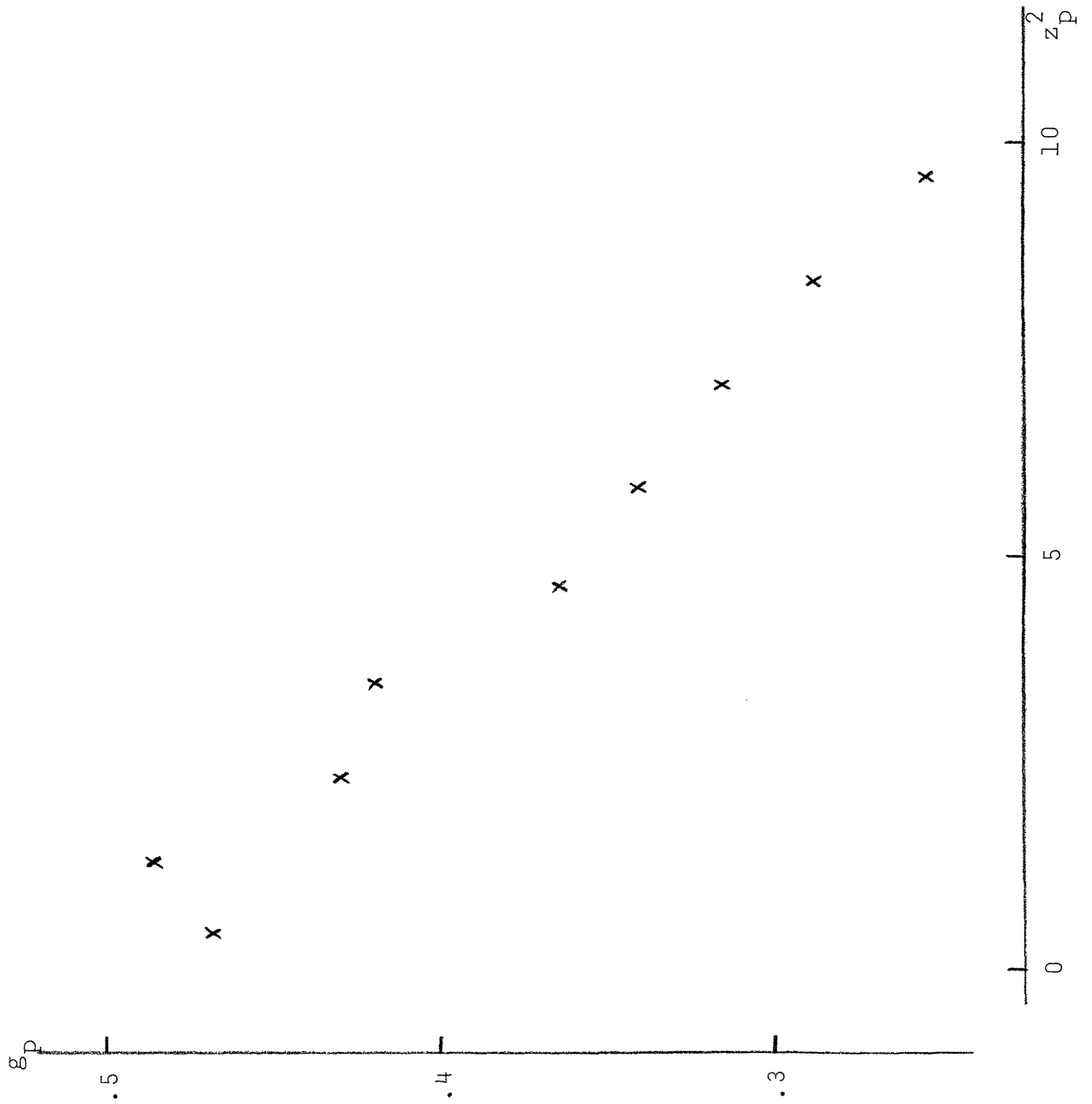
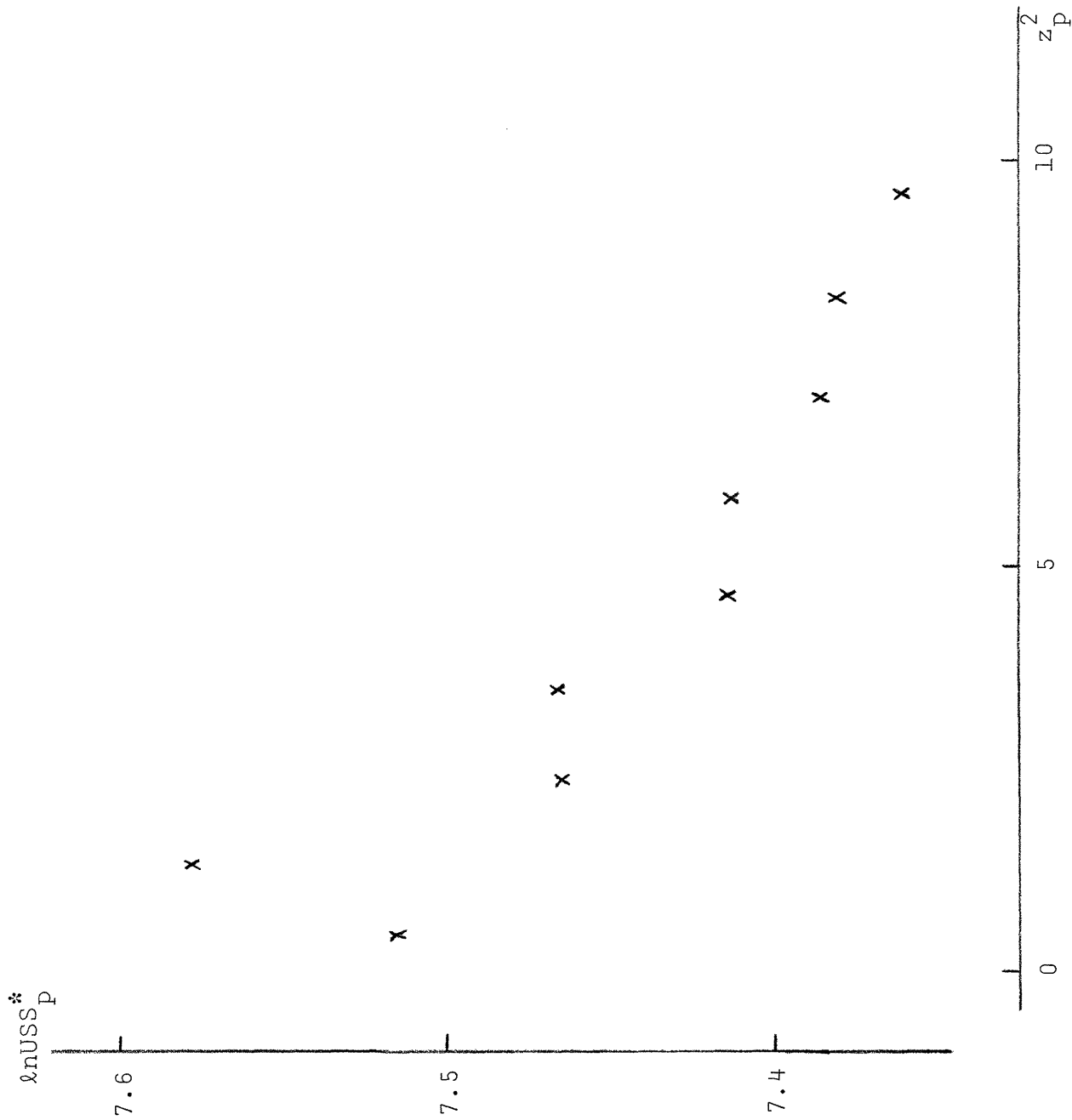


Exhibit 6. Adjusting the upper semi-spreads of the household income data for the fitted pattern of skewness, $g(z) = 0.493 - 0.025z^2$

<u>tag</u>	<u>$x_{1-p} - x_{.5}$</u>	<u>$-z_p$</u>	<u>$g(z_p)$</u>	<u>$G^*(-z_p)^a$</u>	<u>$\ln \text{USS}_p^*$</u>
F	1464	0.6745	0.482	0.797	7.516
E	2963	1.1503	0.460	1.516	7.578
D	3804	1.5341	0.434	2.180	7.465
C	4870	1.8627	0.406	2.784	7.467
B	5514	2.1539	0.377	3.322	7.414
A	6274.5	2.4176	0.347	3.786	7.413
Z	6730	2.6601	0.316	4.170	7.386
Y	7195.5	2.8856	0.285	4.477	7.382
X	7394	3.0973	0.253	4.701	7.361

^a $G^*(z) = Q_{g(z),0}(z) = (e^{g(z)z} - 1)/g(z)$

7
Exhibit 7. Plot of the logarithm of the adjusted upper semi-spread against z_p^2 for the household income data



$$3480 + 1845 \frac{e^{g(z)z} - 1}{g(z)} e^{-0.0168z^2}$$

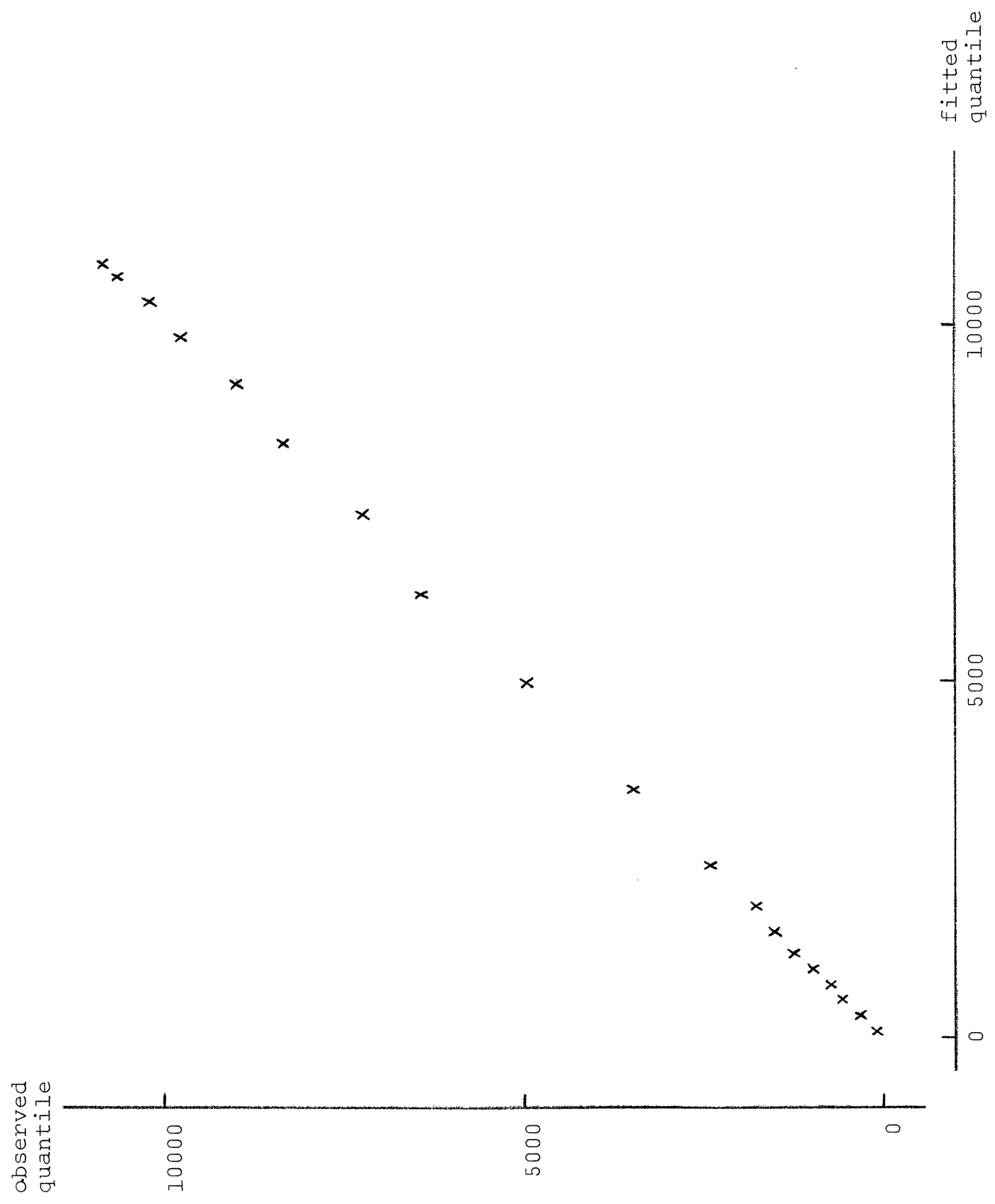
with $g(z) = 0.493 - 0.025z^2$. Using the letter values, we can make a Q-Q plot of the observed and fitted quantiles. This plot, Exhibit 8, shows that a g-and-h-distribution does quite well in describing the shape of this sample of 994 incomes by means of five constants. Perhaps the most interesting feature is the non-constant g : these data are less skewed in the tails than a lognormal distribution with the same skewness in the shoulders. Enrollment conditions for the Housing Allowance Demand Experiment involved income limits, but these varied according to the number of members in the household and the treatment group to which the household had been assigned, so that substantially more detailed analysis would be necessary to determine whether those limits could have affected the overall pattern of skewness. The g-and-h-distributions provide a flexible way of approaching such questions.

4. DESIGN OF THE SOFTWARE

An important objective in developing the present software for exploring distribution shape was to provide capabilities for dealing with a wide range of situations. For example, the data quantiles, x_p , may come from a sample or from a table for a theoretical distribution. Or, as in the first example in Section 3, it may be unnecessary to adjust for fitted skewness before trying to characterize elongation. For this reason we have written the component operations as a collection of Fortran subroutines.

Two basic data structures are appropriate in these algorithms. First, because calculations of g_p rely on paired quantiles, the letter-value data structure preserves these pairings in a two-dimensional array organized in the same way as the central portion of the letter-value display in Exhibit 3. (The only modification is that both columns contain the median as the first element.) Parallel to this array is a vector of depths, which tell how far into the ordered sample one must count to locate the corresponding letter value. (It is sometimes useful to think of the letter values as dividing a population or an ordered sample into segments such that each successive letter value is the median of the segment outside the preceding letter value. Thus the sequence of tail areas in each tail is $1/2, 1/4, 1/8, \dots$.) Second, the quantile data structure, used for Q-Q plots, is simply a vector

Exhibit 8. Q-Q plot for the household income data against the fitted g-and-h distribution



of quantiles in order from smallest to largest.

The subroutines for working with g-and-h-distributions fall into four groups: data manipulation, calculations related to g, calculations related to h, and calculation of fitted quantiles. In describing the subroutines, we do not present detailed argument lists because full listings of the routines are given in the appendix, but we should point out that several routines return intermediate results (which can be displayed or otherwise used as the programmer chooses).

Data Manipulation

LVALS, taken from the book by Velleman and Hoaglin [7], accepts a sorted batch of data and returns an array of letter values and a vector of depths.

LVLDSP, a modified version of LVDSPY [7], prints a letter-value display (and returns a vector of midsummaries and a vector of spreads as by-products).

DTOP (written more informatively as "D TO P") accepts a vector of depths and the corresponding sample size and returns a vector of fractions or p-values, calculated according to

$$p = (d - 1/3)/(n + 1/3),$$

where $d \leq (n + 1)/2$ is a depth and n is the sample size (if $d=1$, the formula used is $p = 0.695/(n + 0.390)$). For integer values of d, these formulas closely approximate the median value of the d-th order statistic in a sample of n from the uniform distribution on [0,1].

PTOZ accepts a vector of p-values and returns the corresponding standard Gaussian quantiles.

LVTQQ accepts an array of pairs of letter values and returns the corresponding vector of quantiles.

Calculations Related to g

LVQ2GP accepts an array of pairs of letter values (or, more generally, pairs of quantiles, centered at and including the median) and a vector of the corresponding Gaussian quantiles and returns the vector of g_p values (calculated according to equation (2.4)), as well as three vectors of

quantities calculated at intermediate steps in the process. These are the "upper semi-spread," $x_{1-p} - x_{.5}$, the "lower semi-spread," $x_{.5} - x_p$, and the logarithm of their ratio.

GPDSPY accepts the vectors of g_p and z_p , along with the intermediates returned by LVQ2GP, and displays the calculations of g_p (in a format similar to Exhibit 4).

Calculations Related to h

QTOH accepts a vector of upper quantiles (i.e., quantiles above the median) for a distribution or sample whose median is zero, also accepts the vector of corresponding Gaussian quantiles, and returns a vector containing the logarithm of their ratio, analogous to equation (2.6). As that equation indicates, the upper quantiles of a sample can be estimated as half the spread: $(x_{1-p} - x_p)/2$.

QTOHAG accepts a vector of either upper semi-spreads or lower semi-spreads (as defined under LVQ2GP), according to the setting of a parameter, also accepts vectors of the corresponding fitted g_p values and the corresponding Gaussian quantiles, adjusts the semi-spreads for the fitted pattern of skewness, and then returns the same log of ratio as calculated by QTOH. Note that the adjustment of g_p can accommodate any fitted pattern of skewness: constant g (including zero), or values found from a polynomial in z^2 , or any vector of values.

Calculation of Fitted Quantiles

Z2QGHC accepts a vector of standard Gaussian quantiles, a constant value of g , and a constant value of h ; and it returns a vector of corresponding quantiles of the g -and- h -distribution, according to equation (2.1).

Z2QGHP accepts a vector of standard Gaussian quantiles and corresponding vectors of (fitted) values of g_p and h_p , and it returns a vector of corresponding quantiles of the (generalized) g -and- h -distribution.

QG evaluates the g -distribution quantile function, $Q_{g,0}(z)$.

QH evaluates the h -distribution quantile function, $Q_{0,h}(z)$.

QGH evaluates the g -and- h -distribution quantile function, $Q_{g,h}(z)$.

5. DISCUSSION

Supported by the software that we have developed, the family of g-and-h-distributions offers a flexible way of exploring and describing distribution shape. To do this fairly routinely for data in a variety of fields would require little additional effort, and the results could add a great deal to what is known about the kinds of distributions that occur in practice.

As a single-number measure of skewness in data, the median of the values of g_p calculated from the letter values is likely to be much more informative than the classical moment-based measure. For a further step, to capture non-constant skewness, we would fit a resistant line (such as the one implemented by Velleman and Hoaglin [7]) to the relationship between g_p and z_p^2 . This should be a minimum requirement if the process is being automated and applied routinely to many substantial bodies of data.

Correspondingly, it will be more valuable to measure elongation in terms of h than to measure kurtosis. We would go at least as far as a constant value of h, and we suggest the right-hand side of equations like (2.6) should include the term $(h_1/2)z_p^4$ whenever a high degree of automation is planned. In any event, such measures of elongation in data should be determined resistantly.

Statisticians often produce large amounts of data in simulation experiments, but analyses of such data rarely go beyond reporting average values and their standard errors. These studies offer an excellent opportunity to learn about distribution shape, and the detailed data values are already available in a computer. Some of the uses of pseudovariances in the Princeton Robustness Study [1] represent an early step in this direction, and one of us (S.C.P.) is now including shape studies in a substantial series of simulation experiments.

6. REFERENCES

- [1] ANDREWS, D.F., P.J. BICKEL, F.R. HAMPEL, P.J. HUBER, W.H. ROGERS, and J.W. TUKEY (1972). Robust Estimates of Location: Survey and Advances. Princeton University Press.

- [2] BAKEMAN, H.E., S.D. KENNEDY, and J. WALLACE (1977). Fourth Annual Report on the Housing Allowance Demand Experiment. Abt Associates Inc., Cambridge, Mass.
- [3] FAMA, E.F. and R. ROLL (1968). Some properties of symmetric stable distributions. J. Amer. Statist. Assoc., 63, 817-836.
- [4] STIGLER, S.M. (1977), Do robust estimators work with real data? Ann. Statist., 5, 1055-1098.
- [5] TUKEY, J.W. (1977). Exploratory Data Analysis. Addison-Wesley Publishing Company.
- [6] TUKEY, J.W. (1977). Modern techniques in data analysis. NSF-sponsored regional research conference at Southeastern Massachusetts University, North Dartmouth, Mass.
- [7] VELLEMAN, P.F. and D.C. HOAGLIN (1979). Computing for Exploratory Data Analysis. Duxbury Press, forthcoming.
- [8] WILK, M.B. and R. GNANADESIKAN (1968). Probability plotting methods for the analysis of data. Biometrika, 55, 1-17.

7. ACKNOWLEDGEMENTS

The authors are grateful to John W. Tukey for comments on earlier related work and to William H. DuMouchel for providing additional quantile values of the symmetric stable distribution used in the example. The analysis of the data from the Housing Allowance Demand Experiment was carried out primarily under Contract H-2040R between the U.S. Department of Housing and Urban Development and Abt Associates Inc., but the opinions expressed are solely those of the authors.

APPENDIX

Listings of Fortran Subroutines

In addition to the routines described in Section 4, this appendix contains a sample driver routine, SHPSMP, which illustrates their use, and the function IGAU, which takes a value of p strictly between 0 and 1 and calculates the corresponding standard Gaussian quantile z_p by evaluating a rational approximation to the standard Gaussian inverse c.d.f.

We have not taken up space by listing the sorting routine, SORT, which simply puts the elements of its argument array, X , into order from smallest to largest.

The arrays connected with the letter-values data structure have fixed sizes and are adequate for up to 15 pairs of letter values (counting the median as one pair). This is unlikely to be a practical constraint because this size can accommodate sample sizes up to 24576.

```

SUBROUTINE SHPSMP(X,N, IOUNIT, A, B, GO, HO)
C
C   INTEGER N, IOUNIT
C   REAL X(N), A, B, GO, HO
C
C   A SAMPLE SHAPE ANALYSIS SEQUENCE.
C   FOR THE BATCH X CONTAINING N VALUES, SUMMARIZE ITS SHAPE
C   WITH A CONSTANT SKEW ADJUSTMENT GO (THE MEDIAN OF THE GP'S) AND
C   WITH THE SIMPLEST ELONGATION ADJUSTMENT, HO.
C   THE PLOTTING ROUTINES PINIT, PLOT1, ABLINZ, AND FINISZ FROM THE
C   PLOTTING PACKAGE GR-Z ARE CALLED TO DISPLAY SEVERAL IMPORTANT
C   RELATIONSHIPS.
C   (SEE 'STRUCTURED GRAPHICAL ALGORITHMS: GR-Z', BELL LABORATORIES
C   MURRAY HILL, NEW JERSEY ).
C   THE SHAPE SUMMARY PARAMETERS ARE RETURNED IN A, B, GO, AND HO.
C
C   INTEGER NLV, ITEXT(4), TNLVM1, NLVP1
C   LOGICAL LERROR
C   REAL DEPTHS(15), VLPAIR(15,2), MID(15), SPREAD(15), PVAL(15), ZP(15),
1  LSS(15), USS(15), LNRAT(15), GP(15), ZPSQ(15), HP(15),
2  FITLVS(15,2), Q(29), QFIT(29), SCRATCH(29), HODIV2
C
C   DATA ITEXT/4HFEDC, 4HBAZY, 4HXWVU, 4HTS /
C
C   SORT THE BATCH. THEN PRODUCE A LETTER-VALUE DISPLAY FROM THE
C   ORDERED DATA.
C
C   CALL SORT(X,N)
C   CALL LVALS(X,N,DEPTHS, VLPAIR, NLV)
C   CALL LVLDSP(DEPTHS, VLPAIR, NLV, MID, SPREAD, IOUNIT)
C
C   A, THE LOCATION PARAMETER, IS JUST THE MEDIAN OF THE BATCH.
C
C   A = VLPAIR(1,1)
C
C   OBTAIN GAUSSIAN QUANTILES CORRESPONDING TO LETTER VALUES, THEN
C   CALCULATE THE VALUES OF GP.
C
C   CALL D TO P (DEPTHS, NLV, N, PVAL)
C   CALL P TO Z (PVAL, NLV, ZP)
C   CALL LVQ2GP(VLPAIR, NLV, ZP, LSS, USS, LNRAT, GP)
C   CALL GPDSPY(NLV, LSS, USS, LNRAT, ZP, GP, IOUNIT)
C
C   PLOT THE VALUES OF GP AGAINST THE SQUARED GAUSSIAN QUANTILES.
C
C   DO 10 I=1, NLV
10      ZPSQ(I) = ZP(I)*ZP(I)
C
C   CALL PINIT

```


A3

```
CALL PLOT1(ZPSQ(2),GP(2),NLV-1,'GP VS. SQUARED NORMAL QUANTILES"',  
1 '','','ZP-SQUARED','','GP','','T=1//',ITEXT)
```

C
C
C
C
C
C

SUMMARIZE THE SKEWNESS BY THE MEDIAN OF THE GP VALUES.
(A MORE SOPHISTICATED ANALYSIS WOULD PROBABLY CONSIDER MORE
COMPLICATED FITS TO GP, BUT FOR THE DATA WE HAVE IN MIND, A
CONSTANT VALUE FOR GP IS ADEQUATE.)

```
CALL SORT(GP(2),NLV-1)  
GO = (GP(1 + NLV/2) + GP(1 + NLV - NLV/2))/2.0E0
```

C
C
C
C

SET GP TO THE FITTED VALUE G(P) (HERE SIMPLY GO)
FOR USE BY THE ELONGATION ANALYSIS.

C
20

```
DO 20 I=1,NLV  
    GP(I) = GO
```

C
C
C
C

CALCULATE THE VALUES OF HP, WHICH WILL INDICATE THE ELONGATION
ADJUSTMENT REQUIRED.

```
CALL Q TO HAG (NLV,USS,ZP,GP,HP,.TRUE.)
```

C
C
C
C
C

PLOT THE VALUES OF HP AGAINST SQUARED GAUSSIAN QUANTILES.
THE PLOT SHOULD BE ROUGHLY STRAIGHT, THE INTERCEPT
IS THE LOG OF THE SCALE PARAMETER B, THE SLOPE IS H0.

```
CALL PLOT1(ZPSQ(2),HP(2),NLV-1,  
1 'LN(USS/GSTAR) VS SQUARED NORMAL QUANTILES','','',  
2 'ZP-SQUARED','LN(USS/GSTAR)','','T=1//',ITEXT)
```

C
C
C

FIT A RESISTANT LINE TO OBTAIN THE SLOPE AND INTERCEPT.

```
CALL RLINE(ZPSQ(2),HP(2),NLV-1,4,.FALSE.,B,HODIV2,LERROR,SCRATCH)
```

C
C
C

DRAW THE RESISTANT LINE ON THE LAST PLOT.

```
CALL ABLINZ(B,HODIV2)
```

C
C
C

TRANSFORM B FROM THE LOG SCALE. COMPUTE H0.

```
B = EXP(B)  
H0 = HODIV2 + HODIV2
```

C
C
C

CONSTRUCT THE FITTED QUANTILES FROM THE SUMMARY PARAMETERS.

C
30

```
DO 30 I=1,NLV  
    HP(I) = H0
```

C
C
C

```
CALL Z2QGHP(ZP,NLV,GP,HP,FITLVS(1,1))
```

C
40
C

```
DO 40 I=1,NLV  
    ZP(I) = -ZP(I)
```

```
CALL Z2QGHP(ZP,NLV,GP,HP,FITLVS(1,2))
```

```
C
DO 50 I=1,NLV
    FITLVS(I,1) = A + B*FITLVS(I,1)
    FITLVS(I,2) = A + B*FITLVS(I,2)
50 CONTINUE
C
C LV TO Q REARRANGES THE COLUMN FORMAT OF THE TABLES TO
C VECTOR FORM, ORDERING THE QUANTILES SMALLEST TO LARGEST.
C
CALL LV TO Q (VLPAIR,NLV,Q)
CALL LV TO Q (FITLVS,NLV,QFIT)
C
C DRAW A Q-Q PLOT SHOWING THE LETTER VALUES AGAINST THE QUANTILES
C OF THE FITTED G-AND-H DISTRIBUTION.
C
TNLVM1 = NLV + NLV - 1
CALL PLOT1(Q,QFIT,TNLVM1,'"', 'FINAL ADJUSTED Q-Q PLOT"',
1 'FITTED QUANTILES"', 'LETTER VALUES"',0,0)
C
C DRAW REFERENCE LINE OF UNIT SLOPE.
C
CALL ABLINZ(0.0E0,1.0E0)
CALL FINISZ
C
RETURN
END
```

```

LVALS   FORTRAN 06/16/79 13:38
        SUBROUTINE LVALS (X,N,D,XLV,NLV)
C
C       INTEGER N, NLV
C
C       REAL X(N), D(15), XLV(15,2)
C
C       INTEGER I,J,K
C
C
C       FOR THE BATCH OF VALUES IN X, WHICH MUST ALREADY HAVE BEEN SORTED
C       INTO INCREASING ORDER FROM X(1) TO X(N), FIND THE SELECTED QUAN-
C       TILES KNOWN AS THE LETTER VALUES. UPON EXIT, XLV CONTAINS THE
C       LETTER VALUES, D CONTAINS THE CORRESPONDING DEPTHS AND
C       NLV IS THE NUMBER OF PAIRS OF LETTER VALUES.
C       SPECIFICALLY, XLV(1,1) AND XLV(1,2) ARE BOTH SET EQUAL TO THE
C       MEDIAN, WHOSE DEPTH, D(1), IS (N+1)/2. THE REST OF THE LETTER
C       VALUES COME IN PAIRS AND ARE STORED IN XLV IN ORDER FROM THE
C       HINGES OUT TO THE EXTREMES.
C       THUS XLV(2,1) AND XLV(2,2) ARE THE LOWER HINGE
C       AND THE UPPER HINGE, RESPECTIVELY, AND XLV(NLV,1) AND XLV(NLV,2)
C       ARE THE LOWER EXTREME (MINIMUM) AND UPPER EXTREME(MAXIMUM),
C       RESPECTIVELY.
C
C       FROM COMPUTING FOR EXPLORATORY DATA ANALYSIS BY
C       PAUL F. VELLEMAN AND DAVID C. HOAGLIN, 1979.
C
C
C       IF ((N .GT. 3) .AND. (N .LE. 24576)) GO TO 10
C           NLV=0
C           RETURN
C
C       HANDLE MEDIAN SEPARATELY BECAUSE IT IS NOT A PAIR OF LETTER VALUES
C
10      D(1)=FLOAT(N+1)/2.0
C           J=(N/2)+1
C           XLV(1,1)=(X(J)+X(N+1-J))/2.0
C           XLV(1,2)=XLV(1,1)
C
C           K=N
C           I=2
C
C
20      K=(K+1)/2
C           J=(K/2)+1
C           D(I)=FLOAT(K+1)/2.0
C
C           XLV(I,1)=(X(J)+X(K+1-J))/2.0
C           XLV(I,2)=(X(N-K+J)+X(N+1-J))/2.0
C
C           I=I+1
C           IF(D(I-1) .GT. 2.0) GO TO 20

```

C

```
NLV=I
D(I)=1.0
XLV(I,1)=X(1)
XLV(I,2)=X(N)
```

C

```
RETURN
END
```

```

LVL DSP   FORTRAN   05/07/79 12:31
      SUBROUTINE LVL DSP (DEPTHS, VLPAIR, NLV, MID, SPREAD, IOUNIT)
C
      INTEGER NLV, IOUNIT
      REAL DEPTHS(15), VLPAIR(15, 2), MID(15), SPREAD(15)
C
      INTEGER ITAG(15)
      DATA ITAG/1HM, 1HF, 1HE, 1HD, 1HC, 1HB, 1HA, 1HZ, 1HY, 1HX, 1HW, 1HV, 1HU,
1      1HT, 1HS/
C
C
C      FROM THE INFORMATION STORED IN VLPAIR AND DEPTHS, PRODUCE
C      A LETTER-VALUE DISPLAY ON LOGICAL UNIT IOUNIT.
C      NLV IS THE NUMBER OF LETTER VALUES (SELECTED QUANTILES) ACTUALLY
C      CALCULATED AND STORED IN VLPAIR.
C      RETURN MIDSUMMARIES AND (LETTER) SPREADS, WHICH WILL BE USEFUL
C      FOR A MID-VS-SPREAD PLOT.
C      THE (LETTER) SPREAD IS JUST THE DIFFERENCE OF CORRESPONDING UPPER
C      AND LOWER QUANTILES ( 0. FOR THE MEDIAN), WHILE THE MIDSUMMARY
C      IS THE POINT HALFWAY BETWEEN THESE QUANTILES.
C
      IF (NLV .LT. 2 .OR. NLV .GT. 15) RETURN
C
      DO 10 I=1, NLV
          MID(I) = (VLPAIR(I, 1) + VLPAIR(I, 2))/2.0E0
          SPREAD(I) = VLPAIR(I, 2) - VLPAIR(I, 1)
10     CONTINUE
C
      WRITE(IOUNIT, 9001) ITAG(1), DEPTHS(1), VLPAIR(1, 1), MID(1),
1     SPREAD(1)
9001  FORMAT(1H0, 20X, 20HLETTER-VALUE DISPLAY //
1     1H , 2X, 5HDEPTH, 2X, 10H  LOW  , 3X, 10H  HIGH  , 5X,
2     10H  MID  , 3X, 10H  SPREAD //
3     1H , A1, 1X, F5.1, 8X, F10.5, 12X, F10.5, 3X, F10.5)
C
      WRITE(IOUNIT, 9002) (ITAG(I), DEPTHS(I), VLPAIR(I, 1), VLPAIR(I, 2),
1     MID(I), SPREAD(I), I=2, NLV)
9002  FORMAT(1H , A1, 1X, F5.1, 2X, F10.5, 3X, F10.5, 5X, F10.5, 3X, F10.5)
C
      WRITE(IOUNIT, 9003)
9003  FORMAT(1H0)
C
      RETURN
      END

```

```
DTOP      FORTRAN 06/16/79 13:39
          SUBROUTINE D TO P (DEPTHS,NLV,N,PVAL)
C
          INTEGER NLV,N
          REAL DEPTHS(NLV),PVAL(NLV)
C
          REAL FLOATN,T
C
C          CONVERT THE VALUES IN DEPTHS TO FRACTIONS
C          N IS THE SIZE OF THE SAMPLE FROM WHICH DEPTHS WAS CALCULATED.
C
          IF (NLV .LT. 1) RETURN
C
          FLOATN = FLOAT(N)
          T = 3.0E0*FLOATN + 1.0E0
          DO 20 I=1,NLV
             IF (DEPTHS(I) .EQ. 1.0E0 .OR. DEPTHS(I) .EQ. FLOATN) GO TO 10
             PVAL(I) = (3.0E0*DEPTHS(I) - 1.0E0)/T
             GO TO 20
C
C          PVAL(I) = (DEPTHS(I) - .305E0)/(FLOATN + .390E0)
10        CONTINUE
20        RETURN
          END
```

```
PTOZ      FORTRAN 06/16/79 13:39
          SUBROUTINE P TO Z (PVAL,NLV,ZP)
C
          INTEGER NLV
          REAL PVAL(NLV),ZP(NLV)
C
          REAL IGAU
C
C
C      GIVEN P-VALUES FIND THE CORRESPONDING GAUSSIAN QUANTILE BY CALLING
C      ON IGAU WHICH COMPUTES THE GAUSSIAN INVERSE C.D.F.
C
C
C      IF (NLV .LT. 2) RETURN
C
          DO 10 I=1,NLV
10      ZP(I) = IGAU(PVAL(I))
C
          RETURN
          END
```

A10

```
LVTOQ   FORTRAN 06/16/79 13:40
        SUBROUTINE LV TO Q (VLPAIR,NLV,Q)
C
        INTEGER NLV
        REAL VLPAIR(15,2),Q(29)
C
C
C REARRANGE THE ARRAY OF LETTER-VALUES IN LVPAIR INTO A VECTOR
C OF QUANTILES Q. THE VALUE OF NLV SELECTS THE NUMBER
C OF VALUES SET INTO Q (THIS GIVES THE OPTION OF STOPPING SHORT
C OF THE EXTREMES).
C THE ELEMENTS OF Q WILL APPEAR IN INCREASING ORDER, FROM
C SMALLEST LETTER-VALUE TO LARGEST.
C
        Q(NLV) = VLPAIR(1,1)
C
        IF (NLV .LT. 2 .OR. NLV .GT. 15) RETURN
        DO 10 I=2,NLV
            Q(NLV-I+1) = VLPAIR(I,1)
            Q(NLV+I-1) = VLPAIR(I,2)
10      CONTINUE
C
        RETURN
        END
```


All

```
LVQ2GP  FORTRAN  05/07/79 12:04
      SUBROUTINE LVQ2GP(VLPAIR,NLV,ZP,LSS,USS,LNRAT,GP)
C
      INTEGER NLV
      REAL VLPAIR(15,2),ZP(NLV),LSS(NLV),USS(NLV),LNRAT(NLV),GP(NLV)
C
C
C      FROM THE LETTER-VALUES STORED IN VLPAIR, CALCULATE THE VALUES
C      OF GP: MINUS THE LOG OF THE RATIO OF THE UPPER TO LOWER
C      SEMI-SPREAD DIVIDED BY THE CORRESPONDING GAUSSIAN QUANTILE.
C      (A SEMI-SPREAD IS THE DIFFERENCE BETWEEN A LETTER VALUE AND
C      THE MEDIAN.)
C      ONLY ELEMENTS 2 THROUGH NLV OF THE VECTORS PASSED TO THIS ROUTINE
C      ARE REFERENCED, THE CALCULATIONS BEING DONE FOR ALL LETTER-VALUES
C      OTHER THAN THE MEDIAN.
C
C
C      IF (NLV .LT. 2 .OR. NLV .GT. 15) RETURN
C
      DO 10 I=2,NLV
          LSS(I) = VLPAIR(1,1) - VLPAIR(I,1)
          USS(I) = VLPAIR(I,2) - VLPAIR(1,1)
          LNRAT(I) = ALOG(USS(I)/LSS(I))
          GP(I) = -LNRAT(I)/ZP(I)
10  CONTINUE
C
      RETURN
      END
```

```

GPDSPY  FORTRAN  05/07/79 12:06
      SUBROUTINE GPDSPY(NLV,LSS,USS,LNRAT,ZP,GP,IOUNIT)
C
      INTEGER NLV,IOUNIT
      REAL LSS(NLV),USS(NLV),LNRAT(NLV),ZP(NLV),GP(NLV)
C
      INTEGER ITAG(15)
      DATA ITAG/1HM,1HF,1HE,1HD,1HC,1HB,1HA,1HZ,1HY,1HX,1HW,1HV,1HU,
1     1HT,1HS/
C
C
C     DISPLAY QUANTITIES USED TO DETERMINE GP ON LOGICAL UNIT IOUNIT.
C
C
      IF (NLV .LT. 2) RETURN
C
      WRITE(IOUNIT,9001)(ITAG(I),LSS(I),USS(I),LNRAT(I),ZP(I),GP(I),
1     I=2,NLV)
9001  FORMAT(1H0,27X,17HCALCULATION OF GP //
1     1H ,8X,12HSEMI-SPREADS,10X,10H LN RATIO ,5X,10H     ZP     ,
2     5X,10H     GP     //
3     (1H ,A1,1X,F10.5,3X,F10.5,5X,F10.5,5X,F10.5,5X,F10.5))
C
      WRITE(IOUNIT,9002)
9002  FORMAT(1H0)
C
      RETURN
      END

```

```
QTOH      FORTRAN  05/07/79 17:53
          SUBROUTINE QTOH(NLV,UPRQNT,ZP,LPSGMA)
C
          INTEGER NLV
          REAL UPRQNT(NLV),ZP(NLV),LPSGMA(NLV)
C
C      CALCULATE LPSGMA = LN(PSEUDO-SIGMA) FROM UPPER QUANTILES
C      AND CORRESPONDING GAUSSIAN QUANTILES.
C      THIS ROUTINE WILL NOT TAKE INTO ACCOUNT ANY SKEWNESS ADJUSTMENTS,
C      FOR THAT CAPABILITY SEE THE SUBROUTINE QTOHAG.
C      NOTE THAT THE UPPER QUANTILE OF A SYMMETRIC DISTRIBUTION
C      CAN BE ESTIMATED AS HALF THE (LETTER) SPREAD.
C
C      IF (NLV .LT. 2) RETURN
C
          DO 10 I=2,NLV
10         LPSGMA(I) = ALOG(UPRQNT(I)/(-ZP(I)))
C
          RETURN
          END
```

```

QTOHAG  FORTRAN  05/07/79 12:09
      SUBROUTINE QTOHAG(NLV,SEMISP,ZP,FITGP,HP,UPSPRD)
C
      INTEGER NLV
      REAL SEMISP(NLV),ZP(NLV),FITGP(NLV),HP(NLV)
      LOGICAL UPSPRD
C
C
C      CALCULATE HP FROM THE (LOWER OR UPPER) SEMISPREAD,
C      CORRESPONDING GAUSSIAN QUANTILE, AND THE SKEWNESS ADJUSTMENT
C      FITGP.  THE LOGICAL PARAMETER UPSPRD IF TRUE IMPLIES THE
C      UPPER SEMI-SPREAD WAS PASSED IN SEMISP, OTHERWISE THE
C      LOWER SEMI-SPREAD HAS BEEN PASSED.
C
C
C      IF (NLV .LT. 2) RETURN
C
C      DO 10 I=2,NLV
          IF (FITGP(I) .EQ. 0.0E0) HP(I) = ALOG(SEMISP(I)/(-ZP(I)))
C
C          IF (FITGP(I) .NE. 0.0E0 .AND. UPSPRD)
1          HP(I)=ALOG(SEMISP(I)/((EXP(FITGP(I))*(-ZP(I)))-1.0E0)/FITGP(I)))
C
C          IF (FITGP(I) .NE. 0.0E0 .AND. .NOT. UPSPRD)
1          HP(I)=ALOG(SEMISP(I)/((1.0E0-EXP(FITGP(I))*ZP(I)))/FITGP(I)))
10      CONTINUE
C
      RETURN
      END

```

```
Z2QGHC  FORTRAN  05/07/79 17:54
        SUBROUTINE Z2QGHC(Z,N,G,H,STDGHQ)
C
        INTEGER N
        REAL Z(N),G,H,STDGHQ(N)
C
C      THIS SUBROUTINE ACCEPTS A VECTOR OF STANDARD GAUSSIAN QUANTILES,
C      Z, A CONSTANT VALUE OF G AND A CONSTANT VALUE OF H AND IT RETURNS
C      A VECTOR OF CORRESPONDING QUANTILES OF THE G-AND-H DISTRIBUTION.
C
C
        IF (N .LT. 1) RETURN
C
        IF (G .NE. 0.0E0) GO TO 20
        DO 10 I=1,N
10          STDGHQ(I) = EXP(H/2.0E0 * Z(I) * Z(I)) * Z(I)
        RETURN
C
20        DO 30 I=1,N
30          STDGHQ(I) = (EXP(G*Z(I)) - 1.0E0)/G * EXP(H/2.0E0*Z(I)*Z(I))
        RETURN
        END
```

```
Z2QGHP  FORTRAN  05/07/79 11:38
      SUBROUTINE Z2QGHP(ZP,N,GP,HP,STDGHQ)
C
      INTEGER N
      REAL ZP(N),GP(N),HP(N),STDGHQ(N)
C
C      THIS SUBROUTINE ACCEPTS A VECTOR OF STANDARD GAUSSIAN QUANTILES,
C      ZP, AND CORRESPONDING VECTORS OF (FITTED) VALUES OF GP AND HP,
C      AND IT RETURNS A VECTOR OF CORRESPONDING QUANTILES OF THE
C      (GENERALIZED) G-AND-H DISTRIBUTION.
C
C      IF (N .LT. 1) RETURN
C
      DO 10 I=1,N
          IF (GP(I) .NE. 0.0E0) STDGHQ(I) =
1      (EXP(GP(I)*ZP(I)) - 1.0E0)/GP(I) * EXP(HP(I)/2.0E0*ZP(I)*ZP(I))
C
          IF (GP(I) .EQ. 0.0E0) STDGHQ(I) =
1      EXP(HP(I)/2.0E0*ZP(I)*ZP(I)) * ZP(I)
10  CONTINUE
      RETURN
      END
```

```
IGAU      FORTRAN 06/16/79 14:27
REAL FUNCTION IGAU(U)
REAL U

C
REAL ANSWER, ARG, TEMP

C
C   CALCULATE THE GAUSSIAN INVERSE CDF AT U BY RATIONAL APPROXIMATION
C   FORMULA 26.2.23 FROM HANDBOOK OF MATHEMATICAL FUNCTIONS,
C   M. ABRAMOWITZ AND I.A. STEGUN EDITORS, NATIONAL BUREAU OF
C   STANDARDS
C   U MUST LIE BETWEEN 0 AND 1 (EXCLUSIVE).
ARG=U
IF (U .GT. .5E0) ARG = 1.0E0 - U

C
TEMP=SQRT(-2.0E0*ALOG(ARG))

C
ANSWER=TEMP-(2.515517E0 + (.802853E0 + .010328E0 * TEMP) * TEMP) /
1 (1.0E0 + (1.432788E0 + (.189629E0 + .001308E0 * TEMP) * TEMP) *
2 TEMP)

C
IGAU=ANSWER
IF (U .GT. .5E0) RETURN
IGAU=-ANSWER
RETURN
END
```

```
QG      FORTRAN 06/16/79 13:57
REAL FUNCTION QG(Z,G0)
C
REAL Z,G0
C
C      THIS FUNCTION CALCULATES A QUANTILE OF THE G-DISTRIBUTION
C      CORRESPONDING TO THE GAUSSIAN QUANTILE Z.
C
IF (G0 .NE. 0.0E0) GO TO 10
    QG = Z
    RETURN
C
10     QG = (EXP(G0*Z) - 1.0E0)/G0
    RETURN
END
```



```
QH      FORTRAN 06/16/79 13:58
REAL FUNCTION QH(Z,H0)
C
C      REAL Z,H0
C
C      THIS FUNCTION CALCULATES A QUANTILE OF THE H-DISTRIBUTION
C      FROM THE CORRESPONDING GAUSSIAN QUANTILE Z.
C
C      QH = EXP(H0*Z*Z) * Z
RETURN
END
```

```
QGH      FORTRAN 06/16/79 14:00
REAL FUNCTION QGH(Z,G0,H0)
C
REAL Z,G0,H0
C
C
C      THIS FUNCTION CALCULATES A QUANTILE OF THE COMBINED G AND H
C      DISTRIBUTION FROM THE CORRESPONDING GAUSSIAN QUANTILE Z.
C
C
      IF (G0 .NE. 0.0E0)GO TO 10
          QGH = EXP(H0*Z*Z) * Z
          RETURN
C
10      QGH = (EXP(G0*Z) - 1.0E0)/G0 * EXP(H0*Z*Z)
      RETURN
      END
```