AD-A045 132 NAVAL POSTGRADUATE SCHOOL MONTEREY CALIF F/G 12/2
ROBUST REGRESSION USING MAXIMUM-LIKELIHOOD WEIGHTING AND ASSUMI--ETC(U)
JUN 77 H R MOORE

UNCLASSIFIED NL

| OF |

AD
A045132

END
DATE
FILMED
11 -77
DDC

NAVAL POSTGRADUATE SCHOOL

Monterey, California

(9) Master's THESIS,

(6) ROBUST REGRESSION USING MAXIMUM-LIKELIHOOD WEIGHTING AND ASSUMING CAUCHY-DISTRIBUTED RANDOM ERROR.

by

(10) Harry Richard Moore, II

(11) June 1977

Thesis Advisor: D. P. Gaver

(12) 59p.

Approved for public release; distribution unlimited.

251 450

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Robust Regression Using Maximum Likelihood Weighting and Assuming Cauchy-Distributed Random Error		5. TYPE OF REPORT & PERIOD COVERED Master's thesis; June 1977
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) LT Harry Richard Moore II, USN		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE June 1977
		13. NUMBER OF PAGES 60
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

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

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Robust regression
Weighted regression
Data analysis
Multivariate analysis
Multiple regression

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Least-squares estimates of regression coefficients are extremely sensitive to large errors in even a single data point. Frequently, an ad-hoc procedure is used to weight the data in a manner to alleviate the effects of extreme observations.

This thesis is a study of the effectiveness of an iterative regression method using weights derived through maximum-likelihood arguments. Actual

Next Page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
(Page 1) S/N 0102-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. weights are calculated on the assumption of Cauchy-distributed error as a worst-case situation in which the errors have long, fat tails and no finite moments.

ROBUST REGRESSION USING MAXIMUM-LIKELIHOOD WEIGHTING AND ASSUMING CAUCHY-DISTRIBUTED RANDOM ERROR

by

H. Richard Moore II
Lieutenant, United States Navy
B.S., United States Naval Academy, 1969

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL

June 1977

Author: H. Richard Moore II

Approved by: Donald P. Gaver, Thesis Advisor

Donald K Bouchaud, Second Reader

M. Sovereign (James K Hartman), Chairman, Department of Operations Research

DA Schrady, Academic Dean

ABSTRACT

Least-squares estimates of regression coefficients are extremely sensitive to large errors in even a single data point. Frequently, an ad-hoc procedure is used to weight the data in a manner to alleviate the effects of extreme observations.

This thesis is a study of the effectiveness of an iterative regression method using weights derived through maximum-likelihood arguments. Actual weights are calculated on the assumption of Cauchy-distributed error as a worst-case situation in which the errors have long, fat tails and no finite moments.

TABLE OF CONTENTS

LIST OF FIGURES

I. INTRODUCTION

A. LEAST-SQUARES LINEAR REGRESSION

It is often desirable to model the behavior of a response variable as a function of another variable, sometimes referred to as a "carrier", since it carries information about the dependent variable. In the simplest case, the equation

$$y_i = b_0 + b_1 x_i$$

is fitted to a set of data points (x_i, y_i). Usually this is done using the "least-squares" procedure which selects the coefficients b_0 and b_1 that minimize the sum of squared residuals, r_i, defined as

$$r_i = y_i - \hat{b}_0 - \hat{b}_1 x_i \quad .$$

The procedure is based on the linear model

$$y_i = b_0 + b_1 x_i + \epsilon_i$$

where the ϵ_i are independent and identically distributed random variables with mean zero and constant (but unknown) variance. Then, by the Gauss-Markov Theorem, the estimates

$\hat{b}_0$ and $\hat{b}_1$ found by solving the "normal equations"

$$\Sigma y_i = n\hat{b}_0 + \hat{b}_1 \Sigma x_i$$

$$\Sigma x_i y_i = \hat{b}_0 \Sigma x_i + \hat{b}_1 \Sigma x_i^2$$

are the best (minimum variance) linear unbiased estimates of b_0 and b_1.

B. DEFICIENCIES OF LEAST-SQUARES

the least-squares procedure works very well when the ϵ_i are short-tailed and the other assumptions about the error distribution hold. If, however, the error distribution has very long tails, implying that extreme observations may well occur, least-squares quickly demonstrates its sensitivity to large random error. In real data, the analyst very rarely has a hint as to the nature of the true distribution of the ϵ_i. Heuristic arguments appealing to the central limit theorem are frequently made along the line that there are several scurces of variability in the data, which, in the aggregate, will be "normally" distributed and thus suitable for least-squares. Unfortunately, if any of the errors are long-tailed (such as may be described by the Cauchy distribution), then their aggregate effect will not be normal.

Figure 10 and Figure 11 in the Appendix are histograms of least-squares estimates for b_0 and b_1 in the linear model

$$y_i = 22 + 2(x_i - \bar{x}) .$$

The ϵ_i for these estimates came from a Normal, or Gaussian, distribution, for which the least-squares procedure is optimal. Figure 12 and figure 13 show the effects of Cauchy-distributed error on the estimates of these coefficients. The end cells contain points which would otherwise be off the scale of the histogram, and emphasize that large errors in estimating the coefficients are quite possible when using least-squares. A uniform distribution's adverse effect upon the coefficient estimates is shown in Figure 14 and Figure 15.

A function of Cauchy variates,

$$V = Ce^{C/100}$$

is the error density associated with the widely-varying coefficient estimates histogrammed in Figure 16 and figure 17. This distribution is virtually symmetric between ±12, but has a long tail extending toward $+\infty$. Another distribution of error, a function of normal variates,

$$Z = Ne^{N+0.01N^2}$$

has high positive skewness, a little bias, and an adverse effect upon the least-squares estimates for b_0 and b_1 as shown in Figure 18 and Figure 19.

All of these cases demonstrate that the variances of the coefficient estimates may be drastically increased by the presence of non-gaussian, and especially long-tailed, distributions of error. While the bulk of the estimates do indeed fall near the actual values, there is clearly an unacceptable probability of obtaining an extreme estimate when using the least-squares procedure.

C. USE OF THE CAUCHY DISTRIBUTION

Data disturbed by Cauchy-distributed error, with long, thick tails and lack of finite moments, may be considered an extremely difficult case for regression techniques to treat reliably. A procedure that works well for data subjected to such extremely straggly-tailed errors can reasonably be expected to work well, though not necessarily optimally, in many curve-fiting situations. This thesis uses maximum-liklihood estimates for regression coefficients to develop a robust regression procedure, then further assumes a Cauchy-distributed error to apply a specific technique to a series of controlled regression problems.

II. SINGLE-CARRIER ROBUST REGRESSION

A. MAXIMUM-LIKELIHOOD ESTIMATORS

The procedure to be presented is based upon the linear model

$$y_i = b_o + b_1(x_i - \bar{x}) + \epsilon_i$$

The $_i$ are assumed to be independent, identically distributed random variables centered at zero with spread (a scale parameter) and having a density of the form

$$\frac{1}{\xi} f\left(\frac{x}{\xi}\right)$$

The probability for any single observation y_i may be expressed as

$$P(Y = y_i) = f\left(\frac{y_i - \hat{b}_o - \hat{b}_1(x_i - \bar{x})}{\xi}\right)\frac{1}{\xi}\, dy$$

The likelihood function for n observations is the product of n of the above probilities. Taking logarithms, the log-likelihood function is then

$$L(\hat{b}_0, \hat{b}_1, \xi) = \Sigma \ln f\left(\frac{y_i - \hat{b}_0 - \hat{b}_1 (x_i - \bar{x})}{\xi}\right) - n \ln \xi$$

Partial derivatives are taken with respect to $\hat{b}_0$, $\hat{b}_1$ and ξ, and all set equal to zero to find the $\hat{b}_0$, $\hat{b}_1$ and ξ which maximize L. Using r_i above, $\psi(x)$ is defined as

$$\psi(x) = \frac{f'}{f} = \frac{\delta}{\delta f} \ln f(x) ,$$

the three equations obtained from setting the partial derivatives of L to zero may be written as

$$\Sigma \psi\left(\frac{r_i}{\xi}\right)\left(\frac{y_i - \hat{b}_0 - \hat{b}_1 (x_i - \bar{x})}{r_i}\right) = 0$$

$$\Sigma \psi\left(\frac{r_i}{\xi}\right)\left(\frac{y_i - \hat{b}_0 - \hat{b}_1 (x_i - \bar{x})}{r_i}\right)(x_i - \bar{x}) = 0$$

$$\Sigma \psi\left(\frac{r_i}{\xi}\right)\left(\frac{y_i - \hat{b}_0 - \hat{b}_1 (x_i - \bar{x})}{r_i}\right)(y_i - \hat{b}_0 - \hat{b}_1 (x_i - \bar{x})) \frac{1}{\xi^2} + \frac{n}{\xi} = 0$$

This system of non-linear equations may be solved iteratively. Defining $w_i^{(j)}$ as

$$w_i^{(j)} = -\psi\left(\frac{r_i^{(j-1)}}{\hat{\xi}_{j-1}}\right)\frac{1}{r_i^{(j-1)}}$$

where the superscript (j) refers to the number of iterations, The equations at the j^{th} iteration are

$$\sum w_i^{(j)}(y_i - \hat{b}_0 - \hat{b}_1(x_i - \bar{x})) = 0$$

$$\sum w_i^{(j)}(x_i - \bar{x})(y_i - \hat{b}_0 - \hat{b}_1(x_i - \bar{x})) = 0$$

$$\hat{\xi}_j^2 = \hat{\xi}_{j-1}\frac{1}{n}\sum w_i^{(j)}[r_i^{(j-1)}]^2$$

The first two equations are simply weighted least-squares normal equations which may be solved by standard iterative weighted least-squares algorithms in which the weights for each subsequent iteration are calculated from the above expression for $w_i^{(j)}$.

Assuming the error to be Cauchy-distributed, the weighting formula becomes

$$w_i^{(j)} = \frac{2}{\hat{\xi}_{j-1}\left(1 + \left(\frac{r_i^{(j-1)}}{\hat{\xi}_{j-1}}\right)^2\right)}$$

B. INITIAL ESTIMATES

It is necessary to begin the iterative process with an initial estimate, or guess, of the values of the coefficients. A robust estimate suggested by D. F. Andrews [1] using the median provides an estimate which is insensitive to arbitrarily large disturbances in up to 25% of the data

The first coefficient, b_o (which corresponds to the mean of the y_i in least-squares estimation) is estimated by the median of the y_i:

$$\hat{b}_o = y_i' \quad .$$

Next, the carriers, $(x_i - \bar{x})$, are ordered and then broken up into three groups of approximately equal size. Of interest are the upper group of carriers, x_U, the lower carriers, x_L, and the y_i corresponding to the $(x_i - \bar{x})$ in each group (y_u and y_l).
The estimate for b_1 is a rough slope computed from the medians of the four groups:

$$\hat{b}_1 = \frac{y_u' - y_l'}{x_u' - x_l'}$$

The median of the absolute values of the residuals from these estimates is the initial guess for ξ:

$$\hat{\xi}_0 = \text{median } |y_i - b_0 - b_1(x_i - \bar{x})|$$

Weights $w_j^{(1)}$ are calculated from the residuals and $\hat{\xi}_0$. The algorithm then proceeds until the values of the coefficients stabilize.

C. SUMMARY OF PROCEDURE

1. Overall Effect

Figure 1 is a typical scatterplot of data which includes extreme observations, or "outliers", and sketches of representative least-squares and robust fits. The effect of the weighting procedure is to pull the extreme observations in closer to the bulk of the data, reducing their tendency to distort the fit (note least-squares line). It should be noted that both the response variable and the carriers are weighted in this technique.

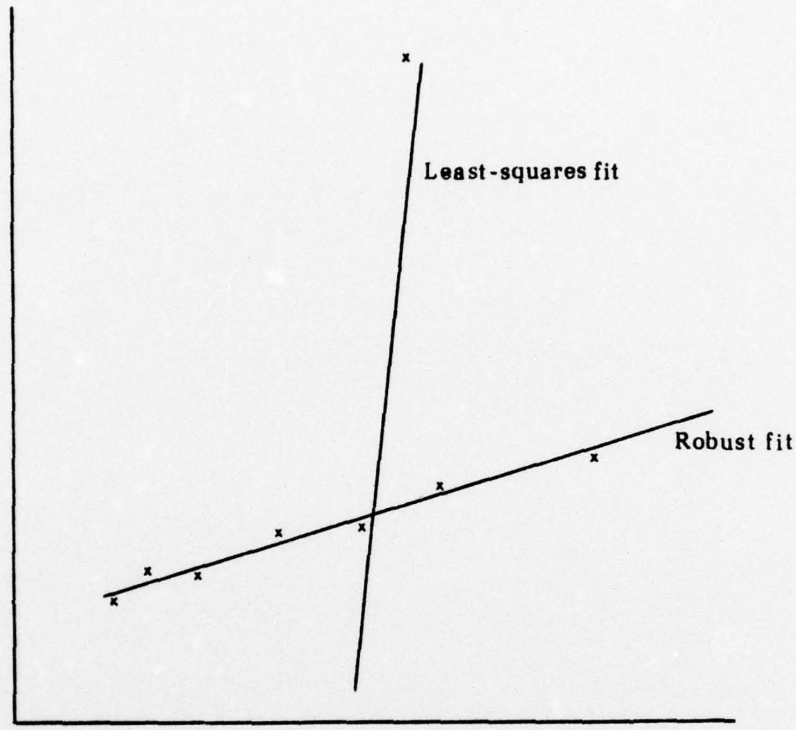


Figure 1 - FITS WITH OUTLIERS

2. Solution Not Unique

There are cases in which the robust procedure may not converge to a single global solution. Since the solution to the weighted normal equations is actually the solution to the three non-linear equations obtained by setting the partial derivatives of equal to zero, there exists the possibility of converging to a local solution not optimizing b_0 and b_1. Figure 2 is an example of a local solution. The scatterplot represents data which actually has two seperate means (the data might be drive-in movie attendence, where observations were made only on Wednesdays and Saturdays). A least-squares fit approximately splits the two groups of points as indicated. A robust fit may also split the data, but could converge to one of the two clusters if either is sparse enough to cause the weighting process to treat its points as outliers.

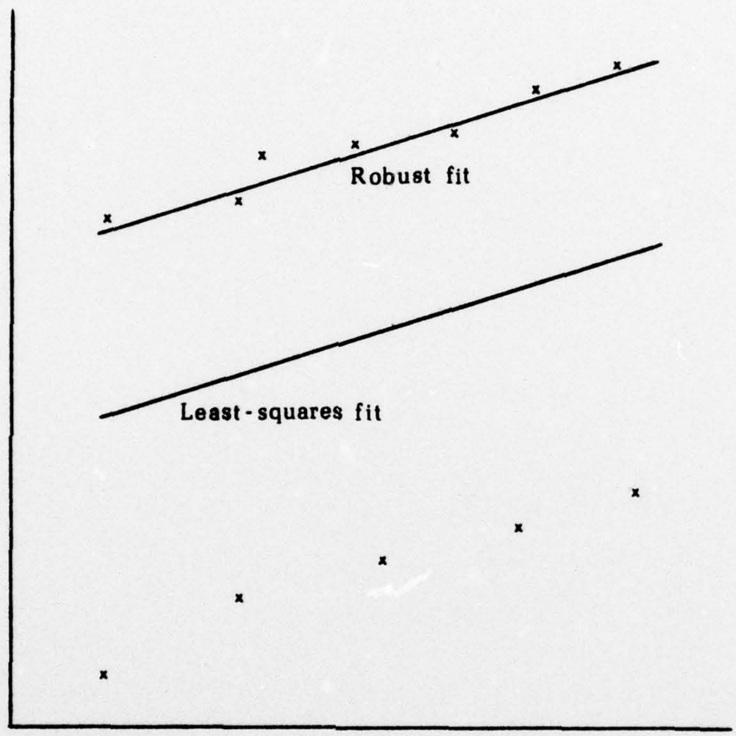


Figure 2 - LOCAL SOLUTION

3. Algorithm

The following flow chart depicts the algorithm for the Cauchy-weighting regression method. The criteria for convergence (change in both coefficients of less than 0.01% from one iteration to the next) was somewhat arbitrary, but was set to meet practical expectations inanalysis problems and not consume excessive amounts of computer time.

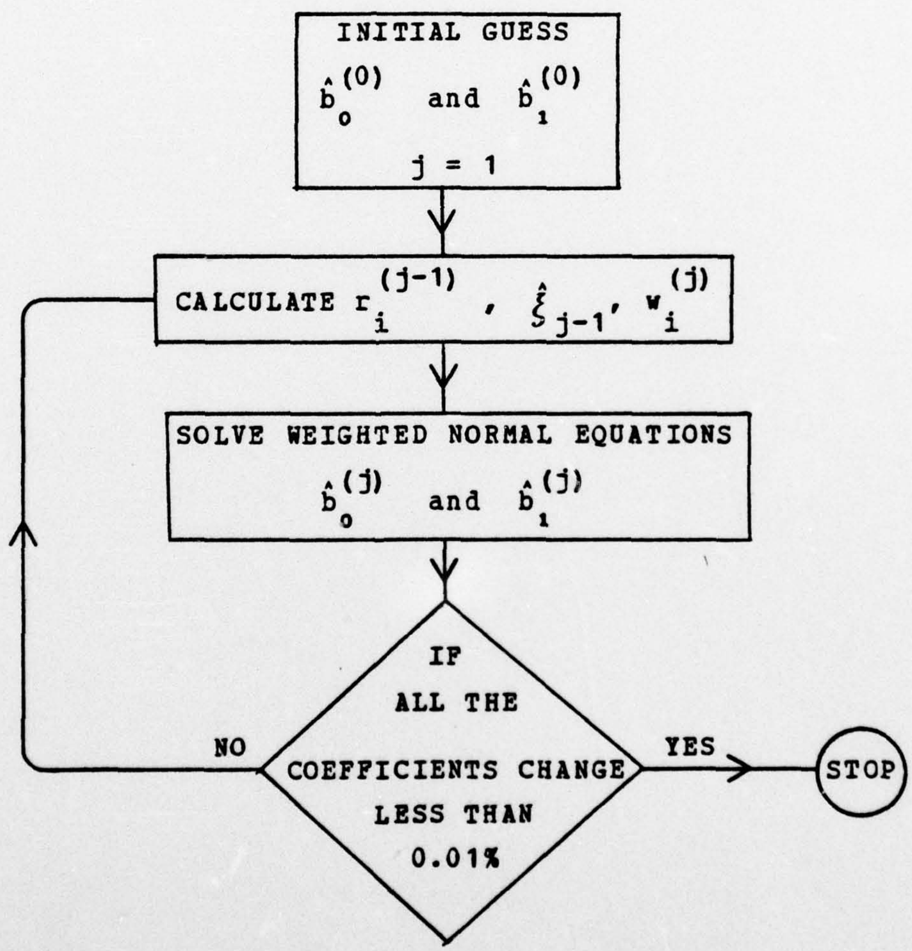


Figure 3 - ALGORTIHM FLOWCHART

D. INADEQUACY OF R^2

One of the measures of adequacy of fit for least-squares regression is R^2, the amount of variance explained by the regression. It is the ratio

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (\hat{y}_i - \bar{y})^2}$$

where

$$\hat{y}_i = \hat{b}_0 + \hat{b}_1 (x_i - \bar{x}) \ .$$

For least-squares, R^2 is a fraction between 0 and 1, but for a robust procedure, the above ratio may exceed 1. This occurs when the robust fit is "farther" from the mean of the data than the least-squares fit.

Consider the following set of observations.

y	3.75	6.00	7.00	8.00	10.25
x	1.00	2.00	4.00	6.00	7.00

The mean of the y_i is 7.00 and a least-squares fit of the model $y = b_0 + b_1 x$ to the data yields $\hat{b}_0 = 3.385$, $\hat{b}_1 = 0.094$ and $R^2 = 0.919$. A robust fit would reduce the effects of observations (2.00,6.00) and (6.00,8.00) since they lie somewhat off the line through the other three points. A robust procedure, bringing these "extreme" observations in closer to the rest of the data, might well yield coefficients of $\hat{b}_0 = 3.000$ and $\hat{b}_1 = 1.000$. From these coefficients and the data, $R^2 = 1.124$. Figure 4 is a scatterplot of the observations with drawings of the actual least-squares fit and the postulated robust fit. Note that the two fits are very close, but more importantly, that the robust fit is rotated so that $(\hat{y}_i - \bar{y})^2$ for the robust fit is everywhere greater than or equal to the same measure for the least-squares fit.

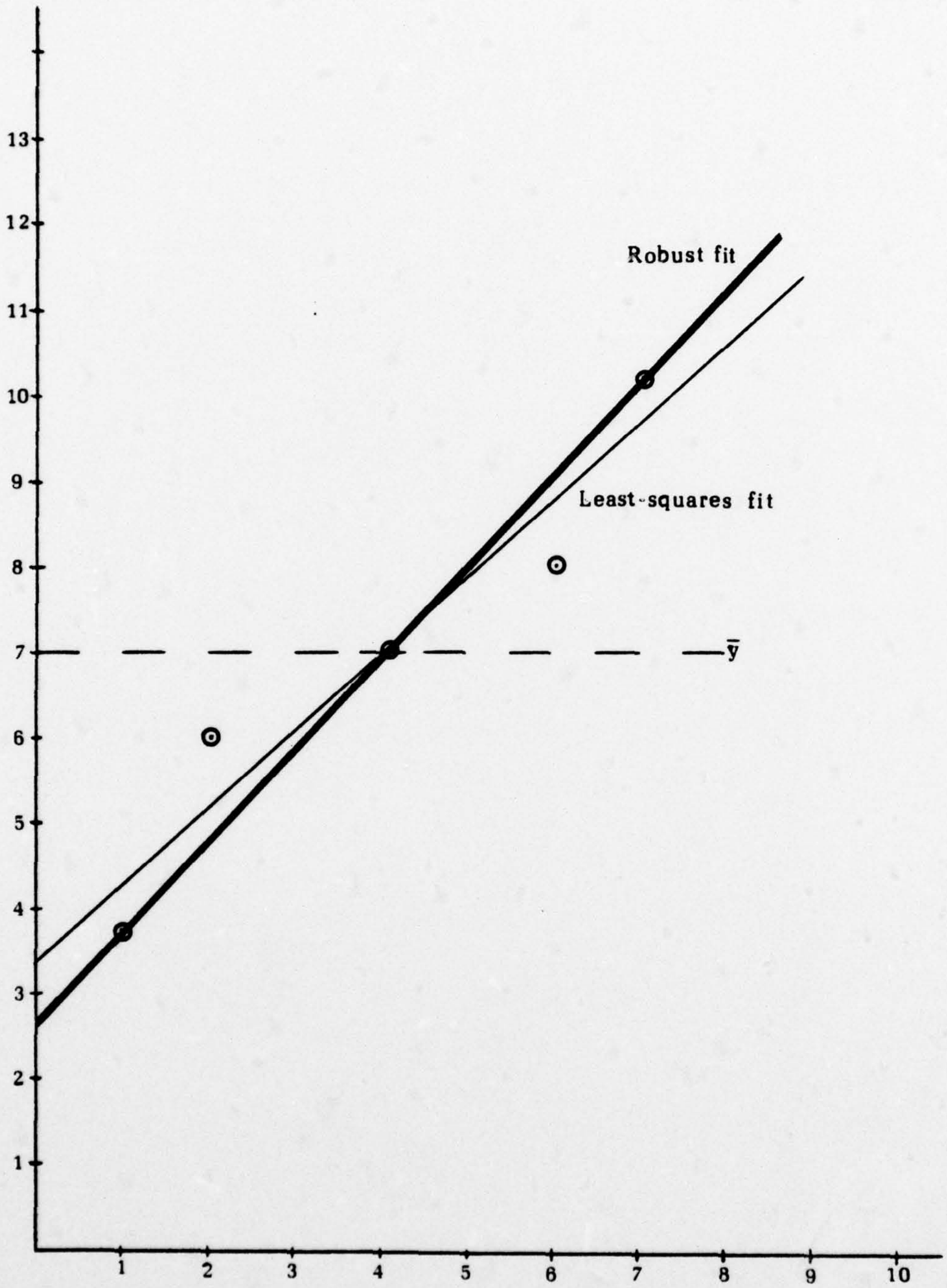


Figure 4 - SITUATION IN WHICH R^2 EXCEEDS 1

More generally, as in Figure 1, R^2 as calculated above is small due to the large deviations from the mean caused by outliers. When a response variable has only a single carrier, a plot of the data and the fitted line provide a visual evaluation of the fit. In multivariable cases, it is usually impossible to plot the data meaningfully, and the good fit of the robust line to the bulk of the data could be belied by inappropriately using R^2 as a measure of the fit.

III. EXPERIMENTAL PROCEDURE FOR A SINGLE CARRIER

A. GENERAL DESCRIPTION

A "true" model

$$y_i = 22 + 2(x_i - \bar{x}) + \epsilon_i$$

was established to enable comparisons of the Cauchy weighting technique and least-squares. The x_i were the integers 1 through 20, and random variates were selected from one of five controlled error distributions to produce 20 observations of the y_i. The y_i were then regressed on the $(x_i - \bar{x})$ to obtain estimates for b_0 and b_1 which could be compared to the actual values.

One thousand replications were made for each distribution and each method. Histograms were constructed for both $\hat{b}_0$ and $\hat{b}_1$ to reveal their distributions. Preliminary runs indicated that most problems converged (both coefficients changed less than 0.01% from one iteration to the next) within 10 iterations. To reduce the amount of time to perform the experiment, the Cauchy-weighting iterations were terminated no later than the seventh iteration. Values of the coefficients were recorded at the fourth iteration to see if there were significant changes between that and the final iteration. If the problem converged early, data normally collected at a later point were assigned the stabilized values.

B. ERROR DISTRIBUTIONS

Five controlled error distributions were used to disturb the observations. The first, the Gaussian or "Normal" distribution with mean zero was matched to the second, a Cauchy distribution. This was done by integrating the Cuachy density centered at zero to find the 75^{th} quantile, giving 1 as a measure of the spread of the distribution. Since the corresponding Normal(0,1) 75^{th} quantile is 0.6745, a Normal distribution with standard deviation 1.4826 will have the same interquartile range as a Cauchy distribution with spread parameter 1. The third source of error was a uniform distribution with mean zero and variance matched to the above Normal, giving it a range of -13.1886 to 13.1886.

The "V" density is a function of Cauchy variates C:

$$V = Ce^{C/100}$$

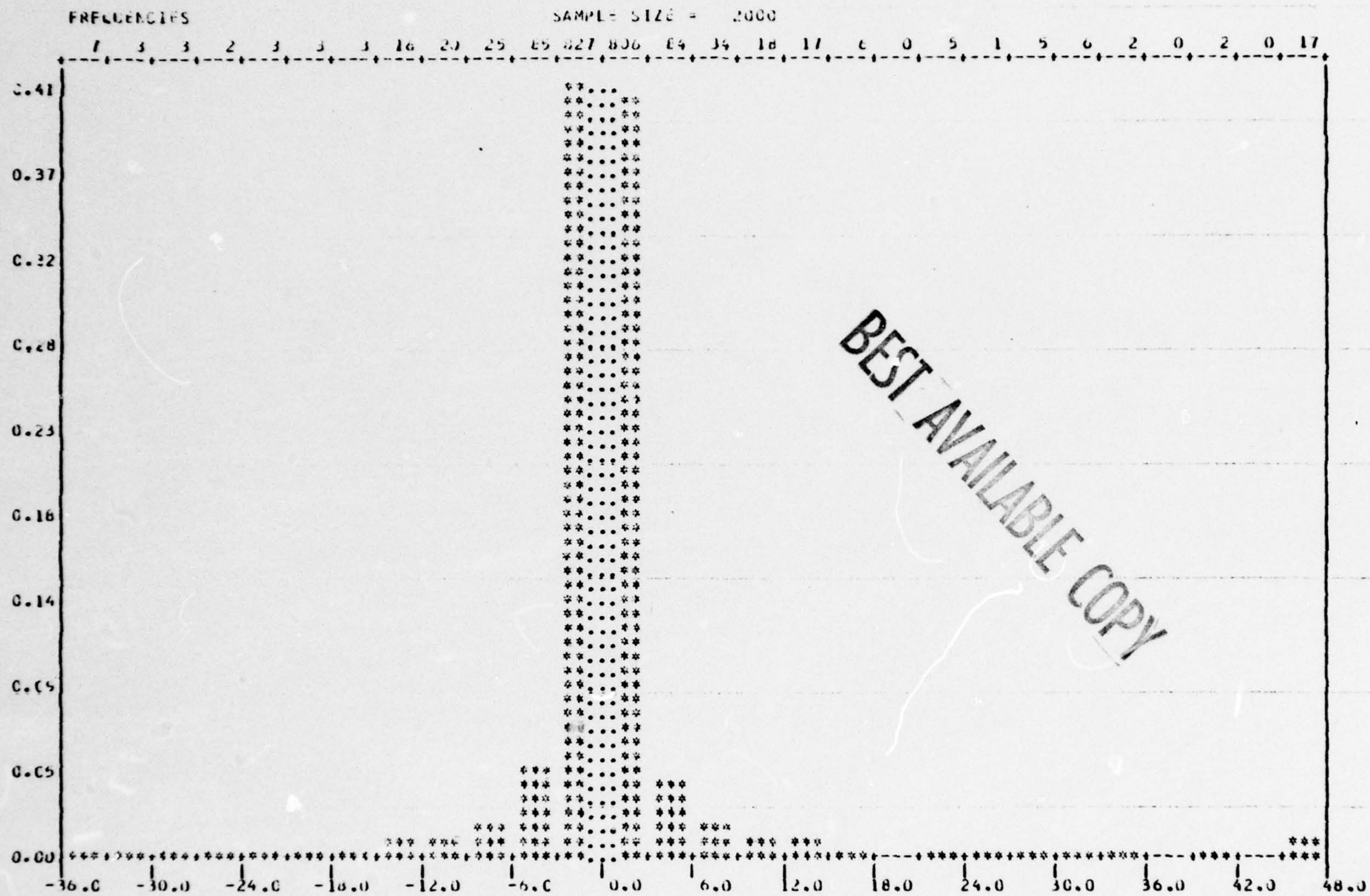
It is positively skewed, but virtually symmetric between -18 and +18 with a very pronounced central spike. Figure 5 is a histogram of 2000 "V" variates.

The final test density is a function of Normal variates N with mean zero and variance 1.

$$Z = Ne^{N+0.01N^2}$$

It is positively skewed and slightly biased. A histogram of 2000 "Z" variates is shown in Figure 6.

Figure 5 - "V" HISTOGRAM



BEST AVAILABLE COPY

Figure 6 - "Z" HISTOGRAM

IV. RESULTS OF SINGLE-CARRIER EXPERIMENT

A. LEAST-SQUARES ADVANTAGES

Summary statistics for the distributions of $\hat{b}_0$ and $\hat{b}_1$ for the single-carrier experiment are shown in Figures 7 and 8. Looking at means and standard deviations, least-squares estimates (maximum-liklihood estimates for normal-error data) are better for normally-distributed cases than the Cauchy method. Interestingly, least-squares is also noticeably better when the error comes from a uniform distribution. This result could be explained by the relatively broad area in which the data points may fall for the uniform error with respect to the range of the $(x_i-\bar{x})$ used, and the susceptibility of the Cauchy method to convergence to local solutions.

True value = 22

	Normal	Cauchy	Uniform	"V"	"Z"
Mean					
Initial	22.03	22.03	21.97	22.19	23.54
Fourth	22.01	22.00	22.03	21.99	22.31
Final	22.01	22.00	22.05	21.99	22.10
Least-squares	22.00	30.68	21.99	1.4E9	23.80
Std. Dev.					
Initial	.9518	2.00	2.75	1.96	1.54
Fourth	.4075	.394	2.68	.413	.373
Final	.4127	.382	2.84	.374	.298
Least-squares	.33	9240.2	1.69	1.8E10	2.78
Minimum					
Initial	18.90	15.36	12.67	15.46	21.63
Fourth	20.81	19.38	13.98	20.45	21.71
Final	20.66	19.38	13.61	20.54	21.67
Least-squares	21.02	-4054	17.24	17.83	21.86
.10 Quantile					
Initial	20.82	19.61	18.49	19.75	21.77
Fourth	21.47	21.54	18.66	21.52	21.93
Final	21.48	21.57	18.46	21.56	21.81
Least-squares	21.59	18.85	19.86	20.57	22.47
.25 Quantile					
Initial	21.39	20.79	20.06	21.05	22.26
Fourth	21.75	21.77	20.15	21.73	22.01
Final	21.73	21.78	20.01	21.76	21.88
Least-squares	21.77	20.88	20.78	21.34	22.86
.50 Quantile					
Initial	22.01	21.99	21.98	22.13	23.41
Fourth	22.01	22.00	22.06	21.98	22.24
Final	22.01	22.00	22.06	21.99	22.03
Least-squares	22.00	22.02	21.97	22.24	23.40
.75 Quantile					
Initial	22.63	23.34	23.86	23.46	24.43
Fourth	22.28	22.24	23.85	22.26	22.50
Final	22.29	22.23	24.02	22.21	22.23
Least-squares	22.23	23.05	23.18	23.58	24.16
.90 Quantile					
Initial	23.24	24.37	25.56	24.51	25.64
Fourth	22.53	22.47	25.50	22.47	22.77
Final	22.56	22.46	25.78	22.45	22.46
Least-squares	22.42	24.15	24.21	27.54	25.18
Maximum					
Initial	25.80	30.71	30.56	30.93	30.35
Fourth	23.30	23.64	29.34	24.32	24.96
Final	23.29	23.57	29.33	23.58	24.44
Least-squares	23.04	29216	27.30	2.3E11	98.14

Figure 7 - SUMMARY OF SINGLE-CARRIER $\hat{b}_0$ DISTRIBUTION

True value = 2

	Normal	Cauchy	Uniform	"Y"	"Z"
Mean					
Initial	2.00	2.00	1.99	2.00	2.00
Fourth	2.00	2.00	1.98	2.00	2.00
Final	2.00	2.00	1.99	2.00	2.00
Least-squares	2.00	8.54	1.99	2.8E7	1.99
Std. Dev.					
Initial	.092	.166	.410	.164	.132
Fourth	.067	.069	.450	.072	.052
Final	.070	.067	.470	.066	.040
Least-squares	.06	144.3	.29	2.8E9	.26
Minimum					
Initial	1.67	1.35	.71	1.41	1.51
Fourth	1.75	1.63	.66	1.69	1.76
Final	1.74	1.67	.66	1.76	1.77
Least-squares	1.79	-75.4	.99	-6E10	-.58
.10 Quantile					
Initial	1.88	1.80	1.48	1.80	1.83
Fourth	1.91	1.92	1.41	1.92	1.94
Final	1.91	1.92	1.40	1.92	1.96
Least-squares	1.93	1.48	1.63	1.53	1.77
.25 Quantile					
Initial	1.94	1.89	1.71	1.89	1.92
Fourth	1.96	1.96	1.67	1.96	1.97
Final	1.95	1.96	1.65	1.96	1.98
Least-squares	1.96	1.84	1.81	1.85	1.91
.50 Quantile					
Initial	2.00	2.01	2.00	2.01	2.00
Fourth	2.00	2.00	1.98	2.00	2.00
Final	2.00	2.00	1.99	2.00	2.00
Least-squares	2.00	2.00	1.99	2.00	2.00
.75 Quantile					
Initial	2.06	2.10	2.27	2.10	2.08
Fourth	2.04	2.04	2.29	2.04	2.02
Final	2.04	2.04	2.30	2.04	2.02
Least-squares	2.04	2.15	21.8	2.13	2.08
.90 Quantile					
Initial	2.11	2.20	2.52	2.20	2.16
Fourth	2.08	2.08	2.60	2.09	2.05
Final	2.09	2.08	2.62	2.08	2.04
Least-squares	2.07	2.37	2.38	2.36	2.20
Maximum					
Initial	2.39	2.56	3.11	2.53	2.52
Fourth	2.52	2.35	3.19	2.39	2.44
Final	2.42	2.34	3.19	2.35	2.36
Least-squares	2.17	4395	2.86	4.5E10	5.73

Figure 8 - SUMMARY OF SINGLE-CARRIER $\hat{b}_1$ DISTRIBUTION

Figure 9 is a diagram of the region in which data points may fall when the error is uniform between ±13.1886 . Since the observations may lie anywhere in the region with equal probability, the weighting process may not be able to clearly discriminate which points are outliers. Chance alignments of a series of points could determine a local optimum upon which the Cauchy-likelihood method would converge. While other distributions have longer tails, the bulk of their variates fall within a relatively small distance of their center, better defining a mean trend and clearly differentiating outliers from the rest of the observations.

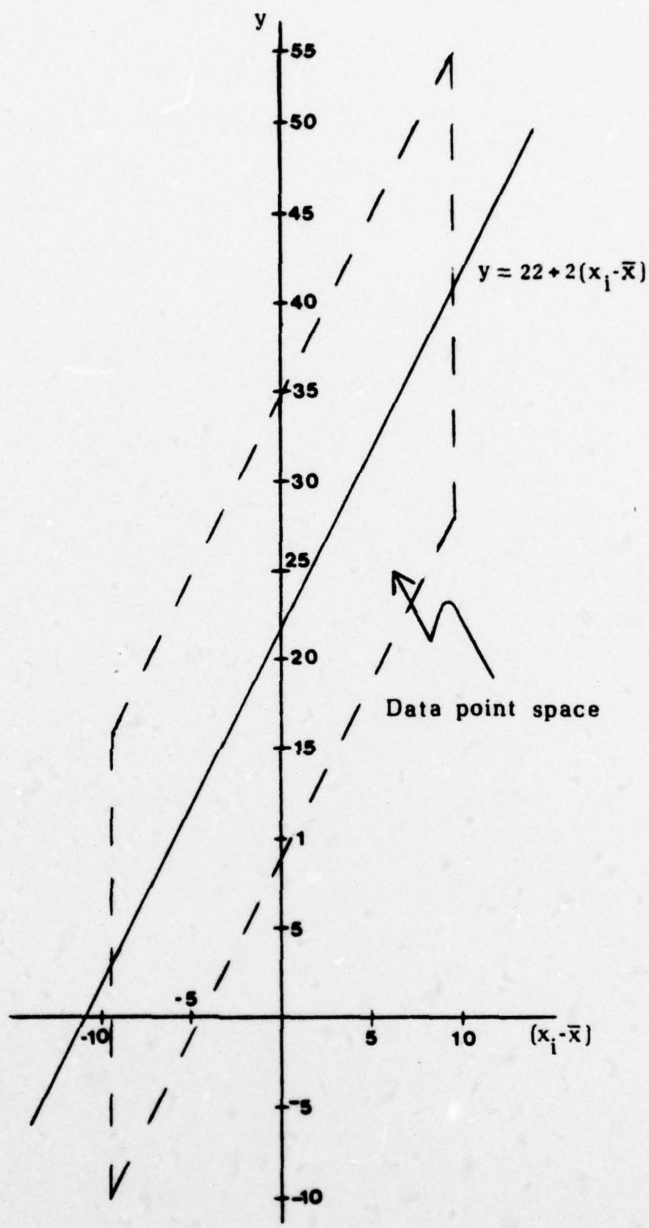


Figure 9 - EQUALLY-LIKELY REGION FOR UNIFORM DATA

B. CAUCHY METHOD ADVANTAGES

Comparing means and standard deviations for the two methods, the Cauchy-likelihood procedure is clearly the more reliable technique when the errors have long tails. Maximum and minimum estimates of the coefficients are closer to their true values when the Cauchy technique is used, and it never produces extreme estimates. There is little difference in the estimates from the fourth to the seventh iterations.

C. SIMILARITIES

The means and medians for both methods are not significantly different. The coefficient estimates between the 25^{th} and 75^{th} quantiles are virtually the same over all distributions, the Cauchy-based method doing better for the long-tailed distributions and the normal having some advantage principally when the error is uniform. Even between the 10^{th} and 90^{th} quantiles, the two procedures yield very similar results.

V. REGRESSION ON TWO INDEPENDENT CARRIERS

A. MULTIVARIATE MODEL

For two independent carriers, the linear model is assumed to be of the form

$$y_i = b_0 + b_1(x_{i1} - \bar{x}_1) + b_2(x_{i2} - \bar{x}_2) + \epsilon_i$$

The regression can be expressed in matrix terms as

$$Y = XB + E.$$

Y is an nx1 matrix of n response variable observations, X is an nxp matrix having all 1's in the first column, the n observations of the first carrier in the second column, and the n observations of each of the remaining p-2 carriers in each of the remaining columns. B is a px1 matrix of coefficients, $b_0, b_1, b_2, \ldots b_{p-1}$, and E is an nxn matrix of unknown random errors, independent and identically distributed, centered at zero and having constant spread.

Using a prime (') to designate the transpose of a matrix, the least-squares normal equations are

$$X'Y = X'X\hat{B} \ .$$

The weighted normal equations are

$$(WX)'Y = (WX)'X\hat{B}$$

where W is an nxn matrix having as its diagonal elements w_{ii} the k^{th} iteration weights, $w_i^{(k)}$, and zeros elsewhere. The weighted normal equations are equivalent to a system of p equations in p unknowns (the b_i, i = 0,1,...,p-1) which can be written as

$$U = V\hat{B}$$

and is easily solved for $\hat{B}$.

B. MODIFICATION TO INITIAL ESTIMATION PRODEDURE

Multiple-carrier regression problems require a modification to the initial estimate procedure to ensure that any interdependence among the carriers is removed prior to estimating the effects of the carriers on the response variable. D. F. Andrews [1] has suggested a rather time-consumming method applying a robust sweep operator to the columns of the X matrix in an iterative process. An alternate method inspired by Mosteller and Tukey [6] sequentially regresses the carriers on each of their predecessors in the X-matrix to eliminate unwanted effects.

Multiple regression may be viewed as a sequence of single-carrier regressions in which the dependent variable, y, is regressed on the first carrier alone according to the model

$$y = \gamma_1 x_1 + \text{residual}$$

Let $y_{;1}$ ("y adjusted for x_1) be the residual after the effects of x_1 are removed:

$$y_{;1} = y - \hat{\gamma}_1 x_1 \quad .$$

This residual is set aside while the effect of x_1 on x_2 is removed using the model

$$x_2 = d_{2;1} x_1 + x_{2;1} \quad ,$$

$x_{2;1}$ being "x_2 adjusted for x_1" . The residual of y ($y_{;1}$) is then regressed on $x_{2;1}$ to find $\hat{b}_2$ in the model

$$y_{;1} = b_2 x_{2;1} + \text{residual} \quad .$$

Substituting for $y_{;1}$ and $x_{2;1}$, $b_1 = (\hat{\gamma}_1 - \hat{d}_{2;1} b_2)$ so that

$$y = b_1 x_1 + b_2 x_2 + \text{residual} \quad .$$

For a model having a mean effect

$$y_i = b_0 + b_1 (x_{i1} - \bar{x}_1) + b_2 (x_{i2} - \bar{x}_2) + \text{residual}$$

In practice, $\hat{b}_0$ is found immediately, (from the median of the y_i, as before) and its effects not removed for computational considerations. It is not important to remove the mean effect since it is independent of the carrier effects that must be removed.

Estimates for γ_i are found using the median estimate described above:

$$\hat{\gamma}_i = \frac{{}_a y'_u - {}_a y'_l}{{}_a x'_{ui} - {}_a x'_{li}}$$

where the $_a$ subscript indicates the quantities have been adjusted for all preceeding carriers. For example, the estimate for $\hat{\gamma}_3$ would require that y_i and x_{i3} both be adjusted for the effects of carrier x_1 and carrier $x_{2;1}$. A similar procedure finds the $d_{j;a}$ coefficients for the j^{th} carrier regressed on its predecessors. The $\hat{\gamma}_i$ and $\hat{d}_{j;a}$ may then be arranged in a system of equations in the $\hat{b}_i$ and subsequently solved to yield the coefficients for the desired model.

C. TWO-CARRIER EXPERIMENT

The twc-carrier experiment was analagous to the one-carrier tests. Coefficients b_0, b_1 and b_2 were fixed to establish a known model

$$y_i = 13 + 3(x_{i1} - \bar{x}_1) - 0.5(x_{i2} - \bar{x}_2) \quad .$$

The x_{i1} were the integers 1 through 20 in ascending crder; the x_{i2} were the same integers shuffled to establish independence in the X-matrix columns. The "true" y_i were then calculated and subsequently disturbed by the same additive error ϵ_i as in the one-carrier case.

Since the single-carrier experiment showed little change in the values of the coefficients from the fourth iteration to the seventh, the two-carrier iterations were terminated after four cycles (or convergence) for each of 1000 replications for each of the five distributions. Only final values were recorded since the initial guesses tended to be somewhat unstable in the first experiment.

D. RESULTS OF THE TWO-CARRIER EXPERIMENT

The results of the second experiment are summarized in Figures 20, 21 and 22 in the Appendix. The estimates of the coefficients parallel the single-carrier cases exactly, with the excepticn of the Cauchy method applied to uniform disturbances. The standard deviations of the Cauchy estimates for the coefficients of the uniform-disturbed data are slightly lower than in the single-carrier case, in contrast to the general trend for the standard deviations to be higher for the two-carrier problems. While there seems to be some interaction between the carriers which raises the standard deviations in general, the use of two carriers may be reducing the tendency of the Cauchy method to converge to a local solution.

VI. CONCLUSIONS

The robust method developed and tested in this paper demonstrates extremely stable behavior over a variety of distributions of random error. Traditional least-squares estimation, on the other hand, is subject to potentially large errors in its estimates of regression coefficients. The method based on Cauchy-likelihood weighting has consistently smaller error when outliers are present and only slightly larger errors (though never any extreme errors) when the error distribution is closer to the Normal distribution.

The Cauchy-likelihood estimates appear to be very slightly biased. The centers of the y_i tend to be estimated too high, while the slopes of the carriers are consistently low. Possibly, if the experiment were run again with the signs of the error terms reversed, the apparent biasing would also reverse to imply that the procedure is robustly unbiased.

There are two drawbacks to the Cauchy method. The first is its requirement for more calculations and intermediate storage. The initial estimates of the coefficients alone require more computer assets than least-squares needs for a complete, though possibly erroneous, solution. As a general rule, the robust Cauchy-likelihood procedure requires twice the data storage capacity and five to six times as long to run as a basic least-squares routine. Clearly, the large reduction in risk for obtaining seriously inaccurate estimates of regression coefficients warrants the use of the

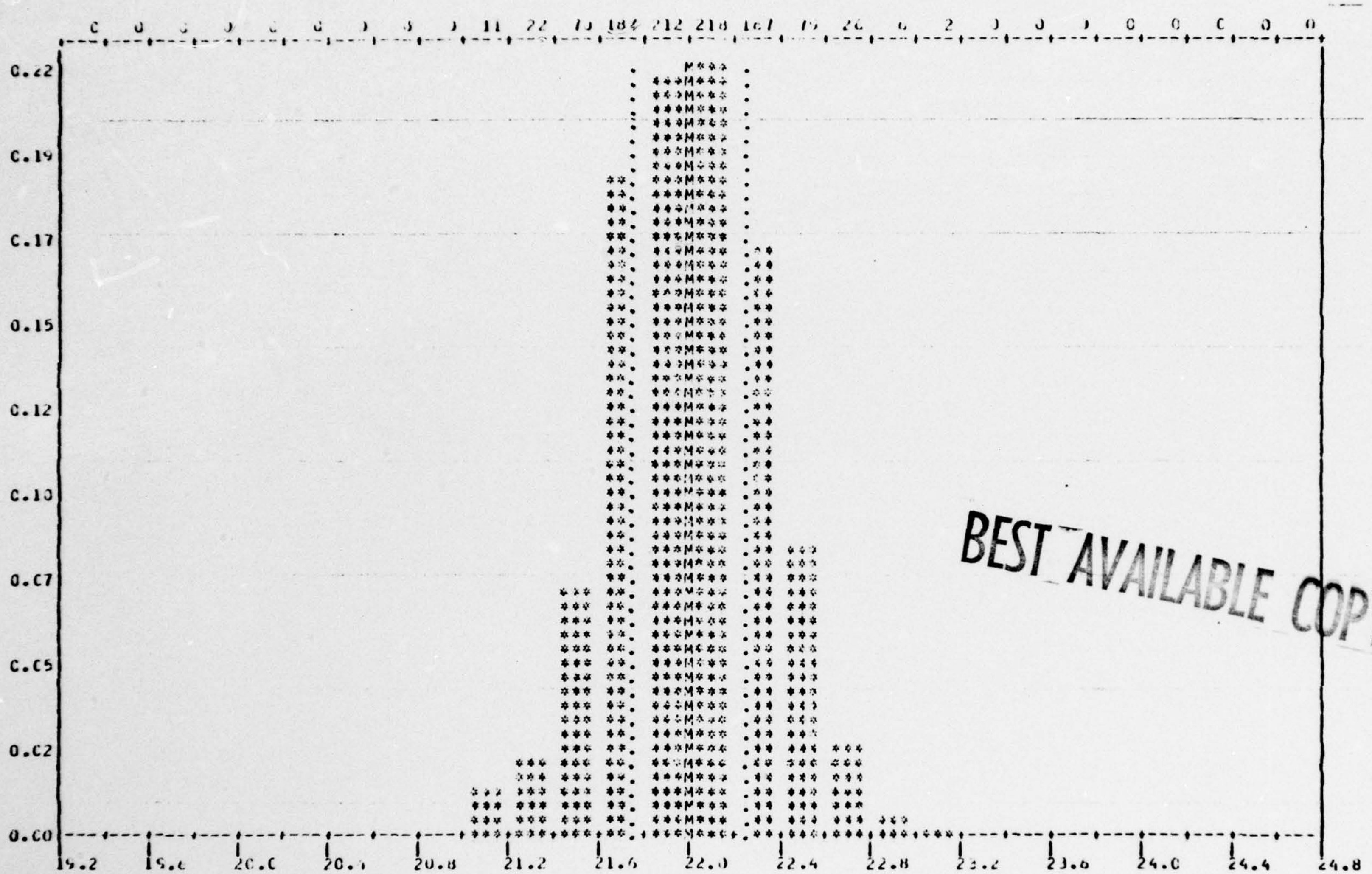
Cauchy, or some other robust procedure in every-day data analysis, even with the increase in computer requirements.

The other problem with using the Cauchy-likelihood method is the possible convergence upon a local solution. It should be noted, however, that traditional least-squares will also produce erroneous results when used under the same conditions which would cause the Cauchy-based technique to stabilize at a local solution.

APPENDIX A

HISTOGRAMS OF LEAST-SQUARES ESTIMATES

Figure 10 - LEAST-SQUARES ESTIMATE OF b_o WITH NORMAL ERROR



BEST AVAILABLE COPY

Figure 11 - LEAST-SQUARES ESTIMATE OF b_1 WITH NORMAL ERROR

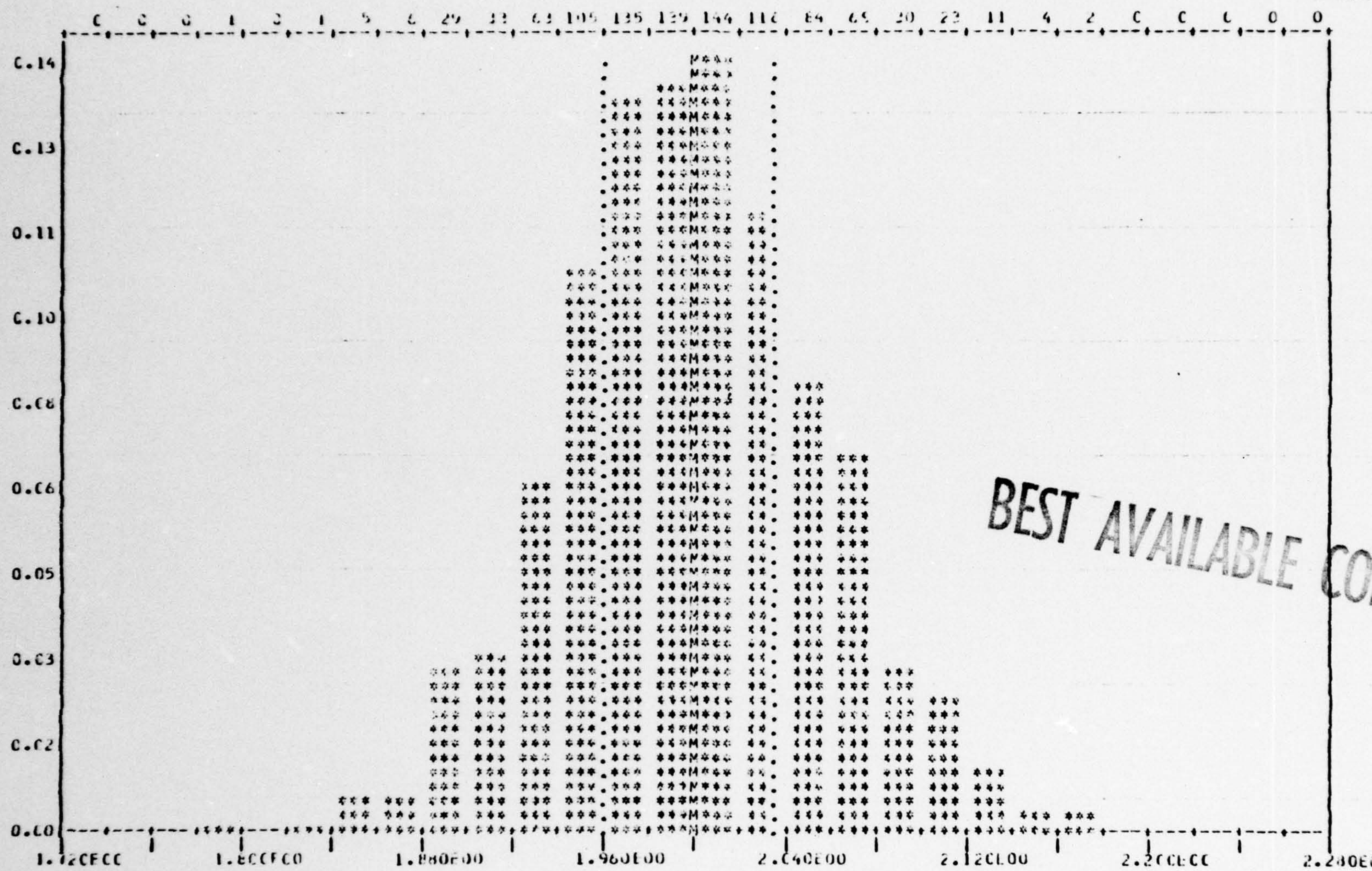


Figure 12 - LEAST-SQUARES ESTIMATE OF b_0 WITH CAUCHY ERROR

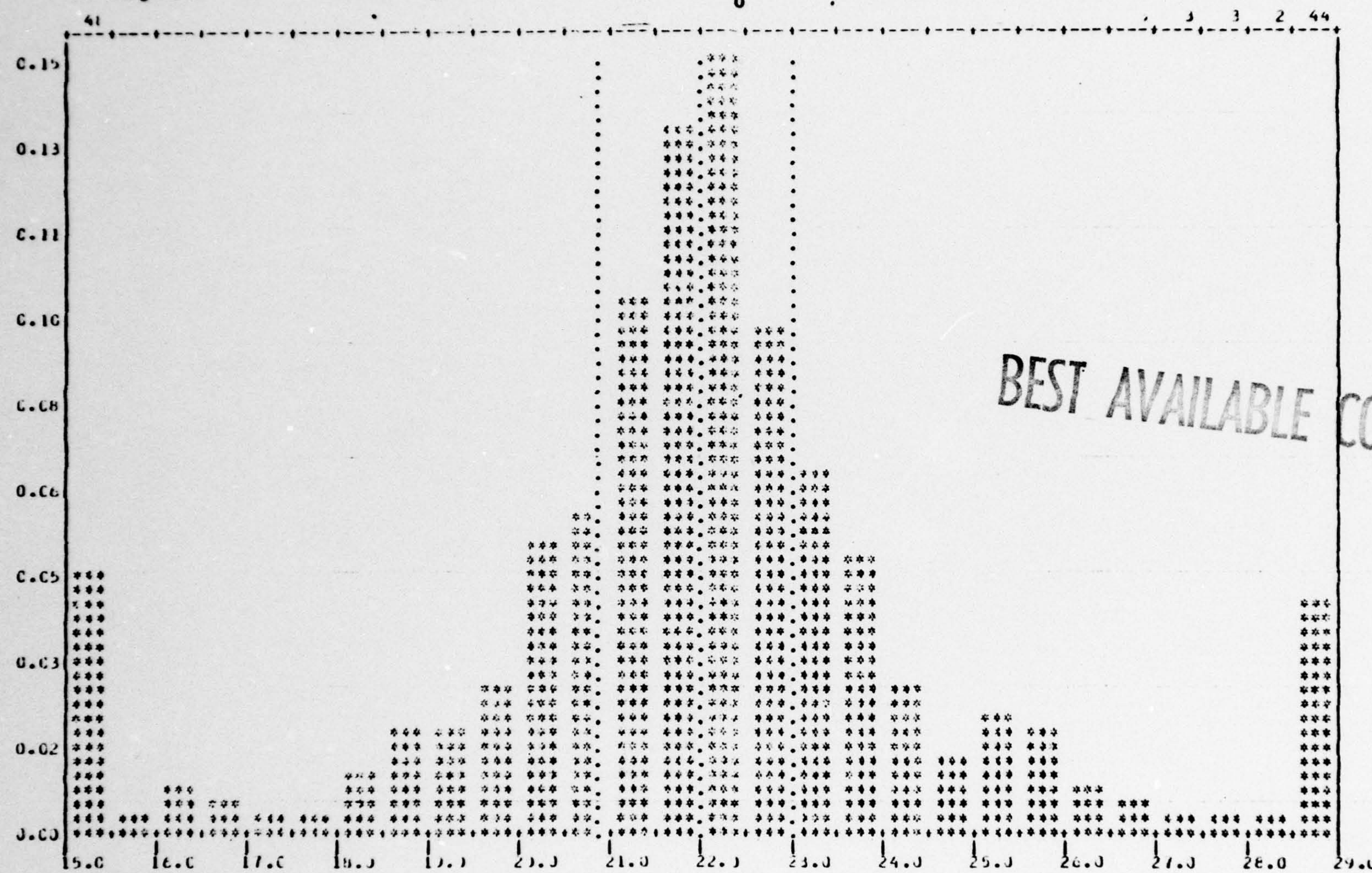


Figure 13 - LEAST-SQUARES ESTIMATE OF b_1 WITH CAUCHY ERROR

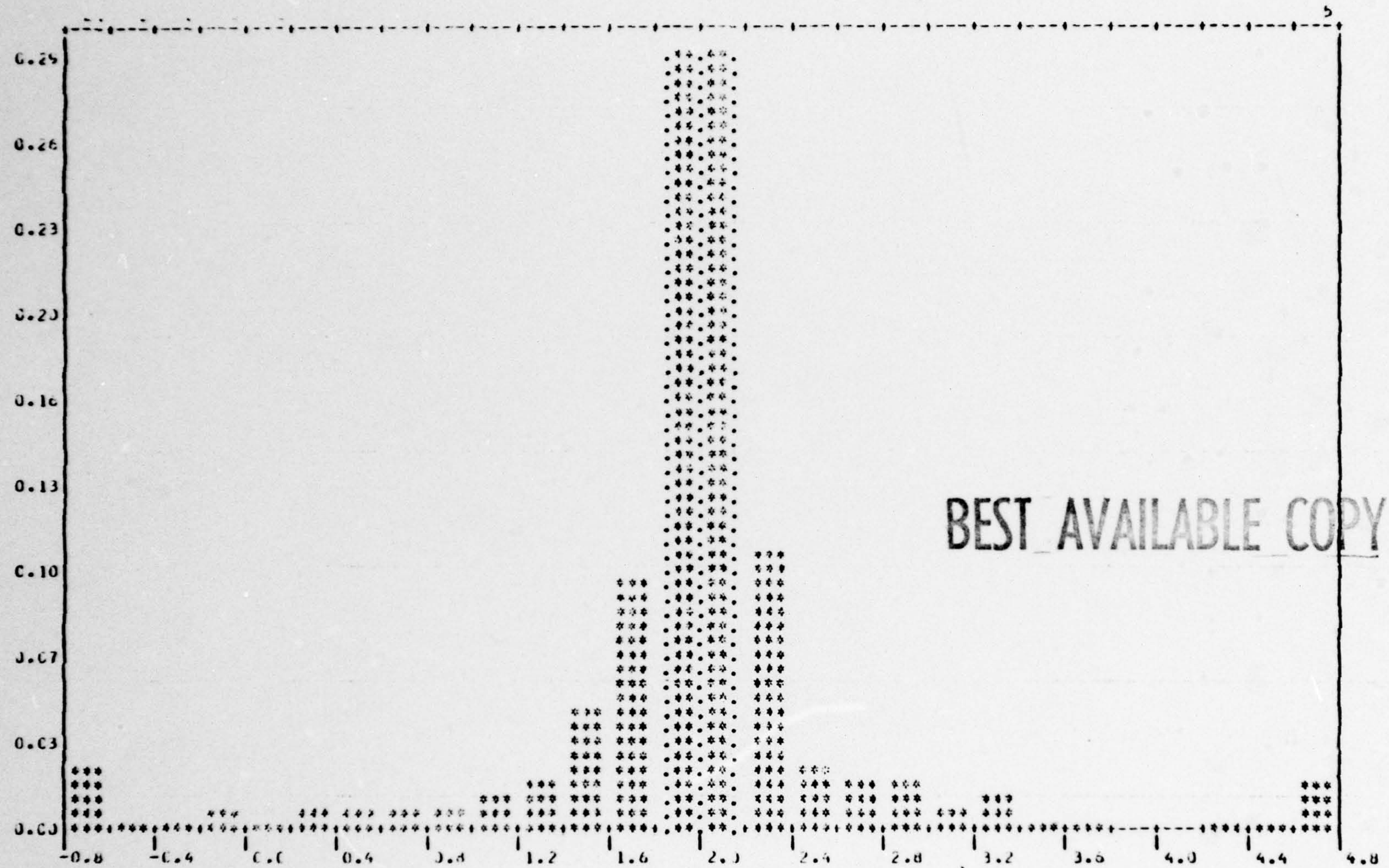


Figure 14 - LEAST-SQUARES ESTIMATE OF b_0 WITH UNIFORM ERROR

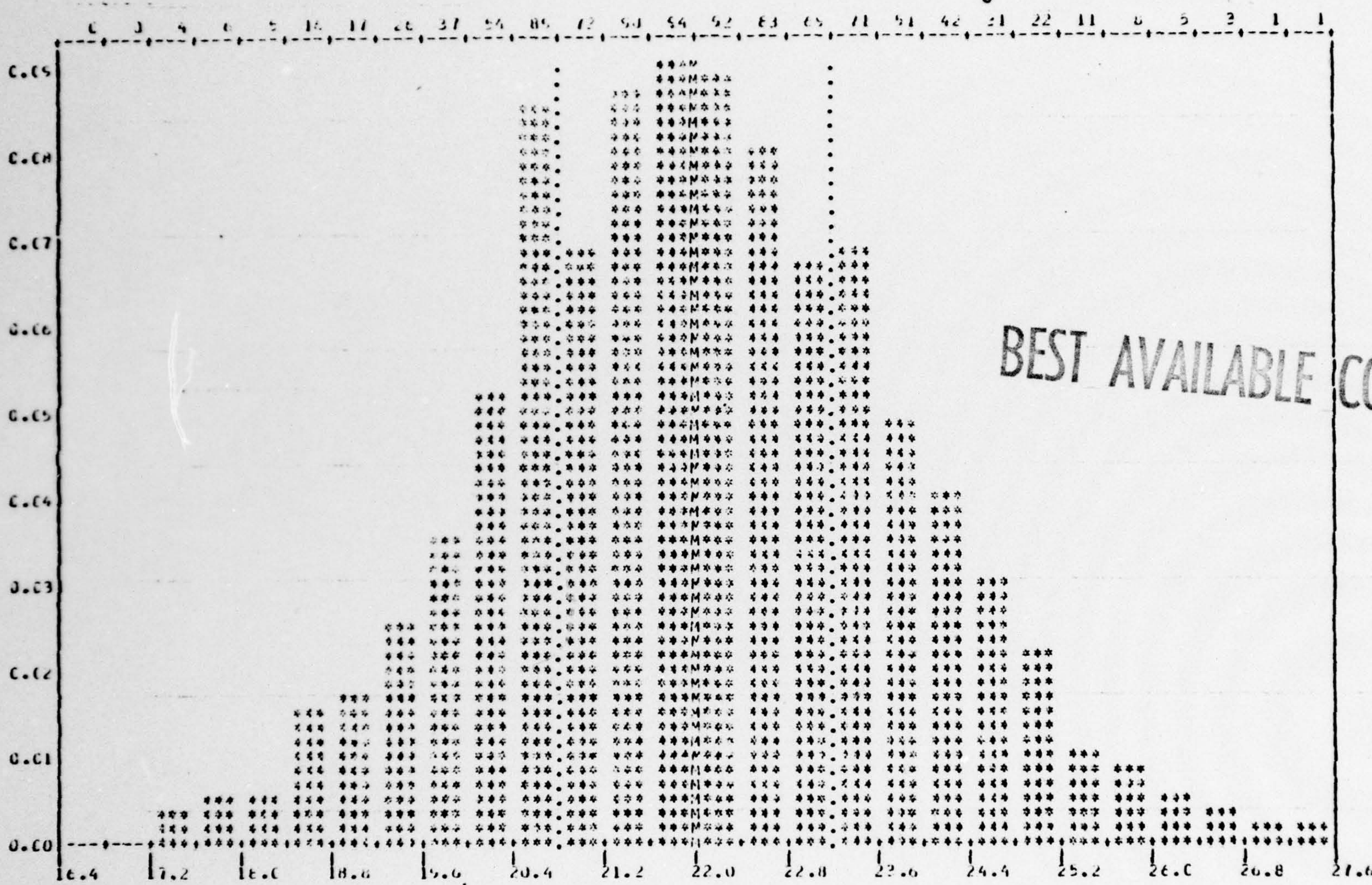


Figure 15 - LEAST-SQUARES ESTIMATE OF b_o WITH UNIFORM ERROR

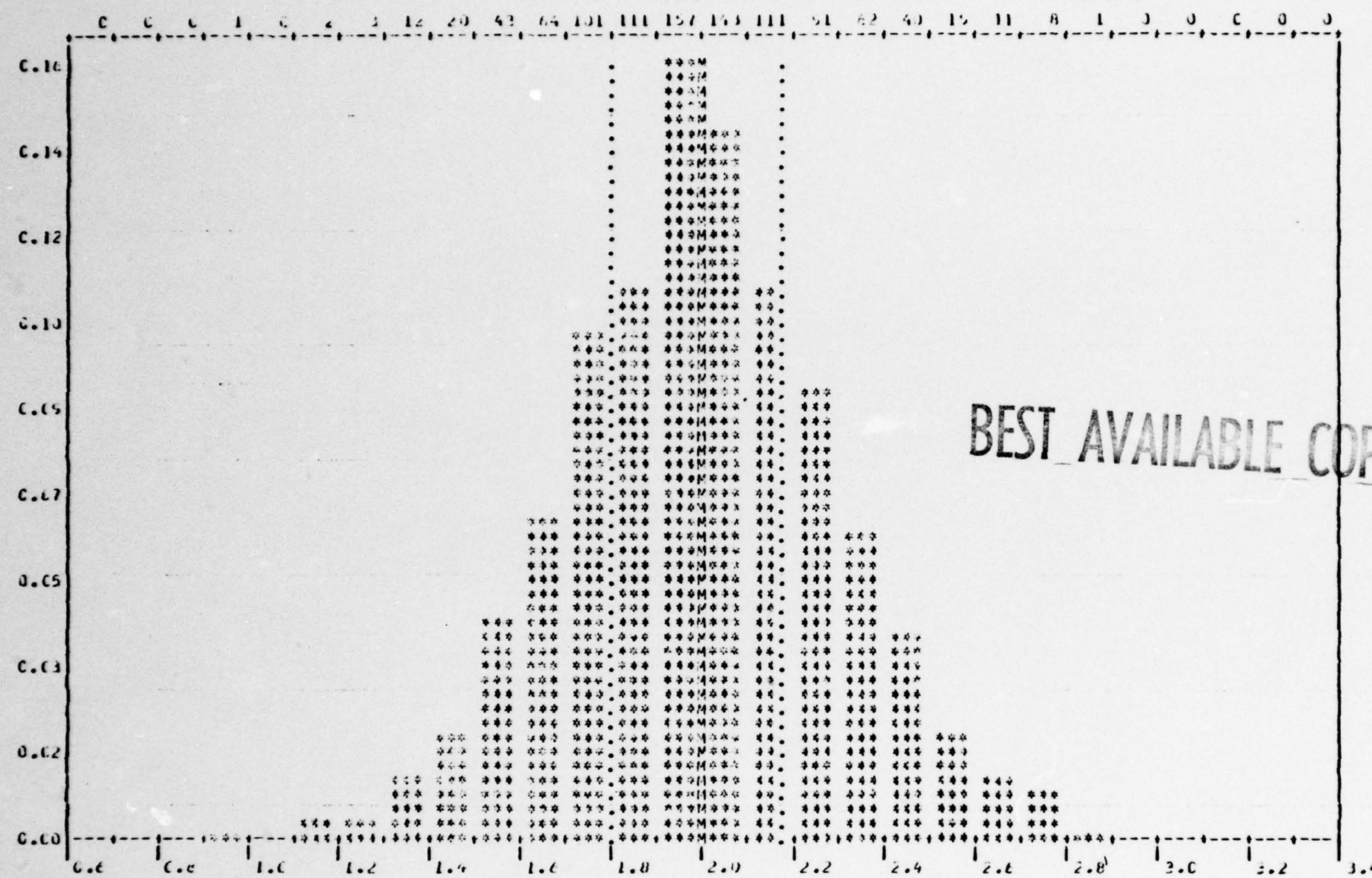


Figure 16 - LEAST-SQUARES ESTIMATE OF b_0 WITH "V" ERROR

Figure 17 - LEAST-SQUARES ESTIMATE OF b_1 WITH "V" ERROR

Figure 18 - LEAST-SQUARES ESTIMATE OF b_o WITH "Z" ERROR

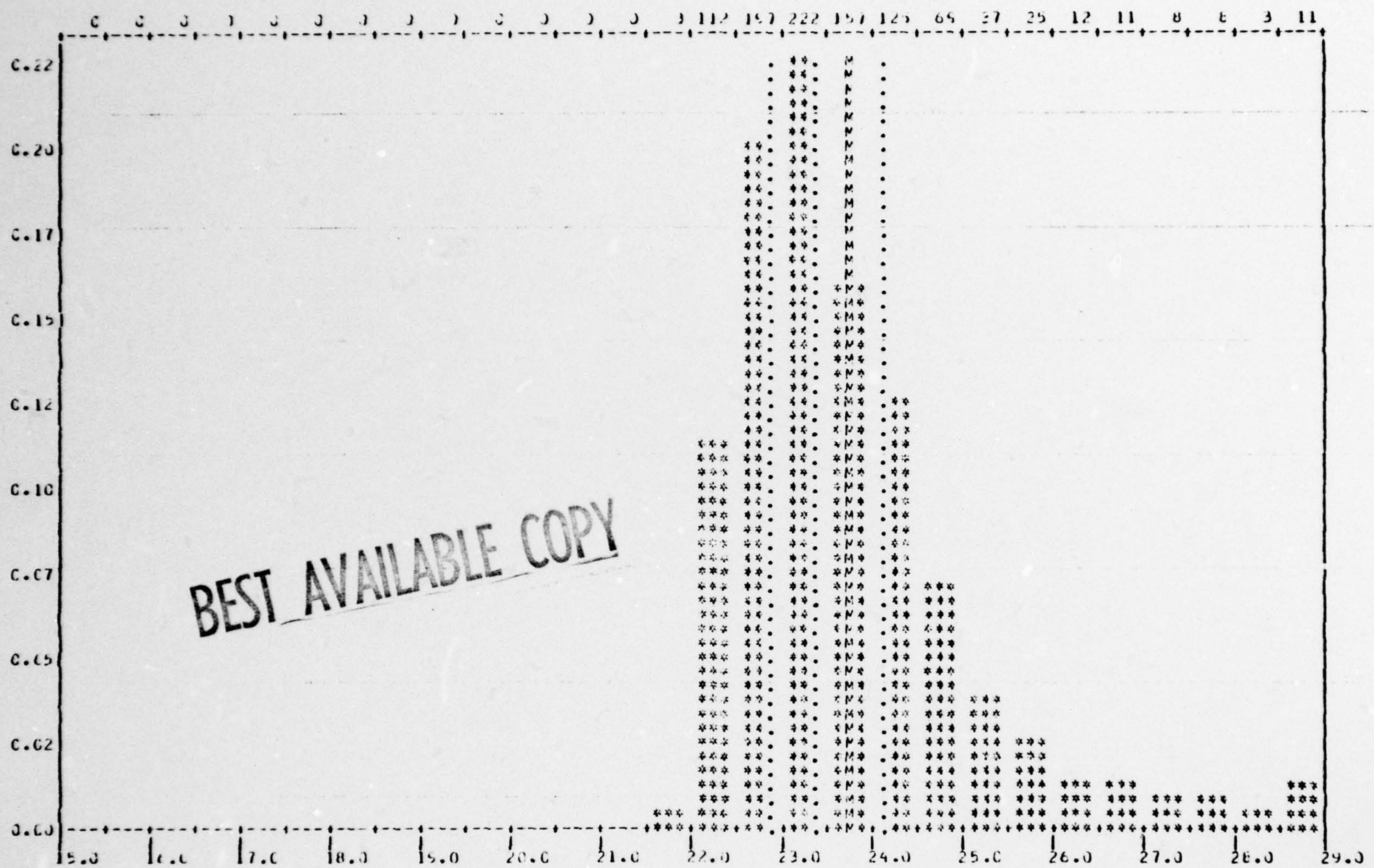
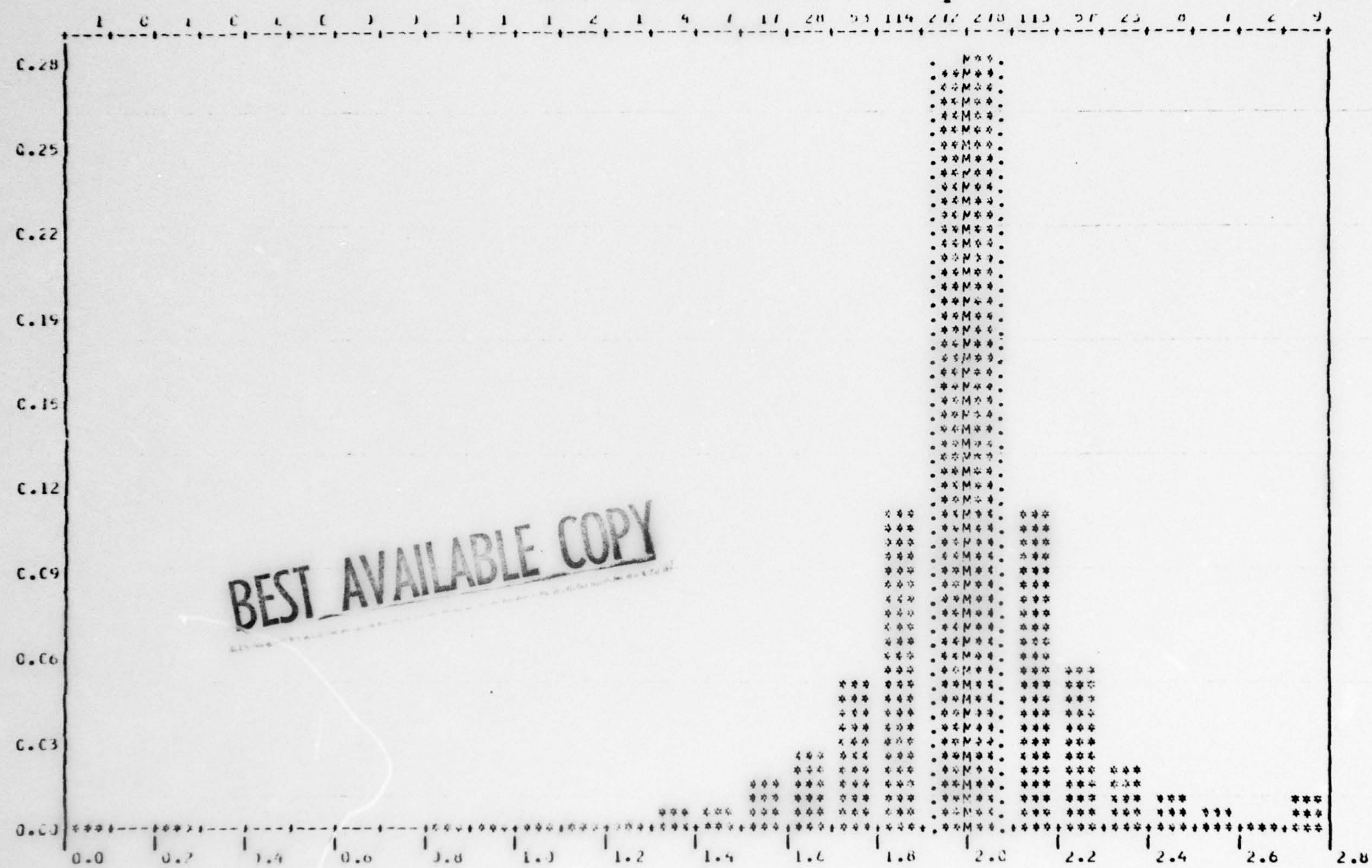


Figure 19 - LEAST-SQUARES ESTIMATE OF b_1 WITH "Z" ERROR



True value = 13

	Normal	Cauchy	Uniform	"V"	"Z"
Mean					
Cauchy	13.02	13.01	13.03	13.01	13.18
Least-squares	13.00	298	12.99	1.4E9	14.8
Std. Dev.					
Cauchy	.417	.406	2.71	.405	.358
Least-squares	.331	9240	1.69	1.8E10	2.78
Minimum					
Cauchy	11.66	11.52	4.68	11.49	12.69
Least-squares	12.02	-4063	8.24	8.83	12.86
.10 Quantile					
Cauchy	12.51	12.54	9.54	12.55	12.83
Least-squares	12.59	9.85	10.87	11.57	13.47
.25 Quantile					
Cauchy	12.74	12.76	11.21	12.75	12.93
Least-squares	12.77	11.88	11.78	12.34	13.86
.50 Quantile					
Cauchy	13.02	13.00	13.02	12.99	13.08
Least-squares	13.00	13.02	12.97	13.25	14.40
.75 Quantile					
Cauchy	13.31	13.27	14.87	13.26	13.32
Least-squares	13.23	14.04	14.18	14.58	15.16
.90 Quantile					
Cauchy	13.56	13.52	16.58	13.51	13.64
Least-squares	13.42	16.15	15.21	18.54	16.18
Maximum					
Cauchy	14.35	15.01	20.18	14.63	15.35
Least-squares	14.04	29.22	18.30	2.3E11	89.14

Figure 20 - SUMMARY OF TWO-CARRIER $\hat{b}_0$ DISTRIBUTION

True value = 3

	Normal	Cauchy	Uniform	"V"	"Z"
Mean					
Cauchy	2.99	2.99	2.96	2.99	3.00
Least-squares	3.00	16.28	2.99	1.2E7	2.99
Std. Dev.					
Cauchy	.071	.071	.446	.072	.044
Least-squares	.057	354	.287	2.7E9	.271
Minimum					
Cauchy	2.73	2.69	1.67	2.64	2.77
Least-squares	2.79	-65.17	2.00	-7E10	.144
.10 Quantile					
Cauchy	2.89	2.90	2.41	2.90	2.95
Least-squares	2.93	2.47	2.63	2.52	2.77
.25 Quantile					
Cauchy	2.94	2.95	2.66	2.95	2.98
Least-squares	2.96	2.84	2.80	2.86	2.91
.50 Quantile					
Cauchy	2.99	2.99	2.96	2.99	3.00
Least-squares	3.00	3.00	2.99	3.00	3.00
.75 Quantile					
Cauchy	3.03	3.03	3.25	3.03	3.02
Least-squares	3.04	3.15	3.18	3.12	3.08
.90 Quantile					
Cauchy	3.08	3.08	3.54	3.07	3.04
Least-squares	3.07	3.37	3.37	3.37	3.20
Maximum					
Cauchy	3.24	3.35	4.19	3.36	3.35
Least-squares	3.18	11140	3.85	3.97	6.29

Figure 21 - SUMMARY OF TWO-CARRIER $\hat{b}_1$ DISTRIBUTION

True value = -0.5

	Normal	Cauchy	Uniform	"V"	"Z"
Mean					
Cauchy	-.501	-.501	-.509	-.501	-.501
Least-squares	-.502	49.29	-.503	-1.3E8	-.502
Std. Dev.					
Cauchy	.070	.069	.435	.069	.044
Least-squares	.055	1576	.283	3.3E9	.443
Minimum					
Cauchy	-.767	-.873	-1.76	-.788	-.802
Least-squares	-.665	-388	-1.29	-5E10	-3.77
.10 Quantile					
Cauchy	-.591	-.585	-1.08	-.587	-.547
Least-squares	-.575	-.991	-.866	-.897	-.721
.25 Quantile					
Cauchy	-.550	-.542	-.824	-.543	-.518
Least-squares	-.536	-.663	-.682	-.650	-.589
.50 Quantile					
Cauchy	-.499	-.499	-.499	-.499	-.501
Least-squares	-.501	-.503	-.499	-.502	-.496
.75 Quantile					
Cauchy	-.455	-.460	-.214	-.458	-.482
Least-squares	-.468	-.368	-.317	-.367	-.414
.90 Quantile					
Cauchy	-.415	-.418	-.062	-.418	-.457
Least-squares	-.429	-.109	-.129	-.126	-.295
Maximum					
Cauchy	-.295	-.212	.619	-.234	-.294
Least-squares	-.311	49823	.365	3.9E10	1.07

Figure 22 - SUMMARY OF TWO-CARRIER $\hat{b}_2$ DISTRIBUTION

LIST OF REFERENCES

1. Andrews, D. F., "A Robust Method for Multiple Linear Regression," Technometrics, v. 16, p.523-531, November 1974.
2. Draper, N. R. and Smith, H., Applied Regression Analysis, Wiley, 1966.
3. Fishman, G. S., Concepts and Methods in Discrete Event Digital Simulation, Wiley, 1974.
4. Gillman, L. and Rose, A. J., APL An Interactive Approach, 2d ed., Wiley, 1974.
5. Hutchinson, D. R., APL Histogram, Density Estimation and Probability Plotting Routines, M.S. Thesis, Naval Postgraduate School, Monterey, California, December 1976.
6. Mosteller, F. and Tukey, J. W., Data Analysis and Regression: A Second Course in Statistics, Addison-Wesley, 1977.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93940	2
3.	Prof. M. G. Sovereign, Code 55 Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
4.	Prof. D. P. Gaver, Code 55Gv Department of Operations Research Naval Postgraduate School Monterey, California 93940	2
5.	Lt. D. R. Bouchoux, Code 55Bh Department of Operations Research Naval Postgraduate School Monterey, California 93940	2
6.	Mr. Dean Gonzalez ACEVAL/AIMVAL Joint Test Force Nellis AFB, Nevada 89191	2
7.	Lt. H. Richard Moore II 5368 Albright Drive Virginia Beach, Virginia 23462	3