THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

A Comparison of Seven Imputation Procedures for ISDP

No. 14

Vicki J. Huggins

U.S. Census Bureau

November 1985

U.S. Department of Commerce U.S. CENSUS BUREAU

Survey of Income and Program Participation

A COMPARISON OF SEVEN
IMPUTATION PROCEDURES FOR ISDP

No. 8606 14

Vicki J. Huggins

SIPP Working Paper #14

November 1985

ACKNOWLEDGEMENT

This paper was prepared by Vicki J. Huggins of the Statistical Research Division, Bureau of the Census. The author would like to thank Brian Greenberg for suggesting this research and providing a number of helpful recommendations along the way.

SUGGESTED CITATION

Huggins, Vicki J. "A Comparison of Seven Imputation Procedures for ISDP," SIPP Working Paper Series No. 8606. Washington, DC: U.S. Bureau of the Census, 1986.

Table of Contents

I. INTRODUCTION

Missing data for longitudinal surveys occur in a variety of patterns which can be sorted and categorized into different classes of missingness depending on the survey unit. For this study, the survey unit is a person. Therefore the missingness that occurs in the data can be person nonresponse, whereby no data are available for a person at any given time period in the survey, record-type nonresponse where an entire module of related data for a person is unavailable, and item nonresponse in which data are missing sporadically throughout the person record. For this study we focused on record-type nonresponse for a single continuous variable. It is important that this type of nonresponse be addressed as it occurs generously throughout a longitudinal survey. Also, simulation of record-type nonresponse provides reasonably sized data files to study and manipulate. It is important to note that the techniques investigated can be employed to compensate for both item and record-type nonresponse.

The objective of this study is to evaluate seven different methods of imputation for continuous data in a longitudinal survey. The methods compared are described below as are the procedures to compare them. In our comparisons, we employed a variety of summary statistics and graphic techniques. The particular findings are detailed in the body of the text and a number of graphs and tables are included in the Appendix to support these findings. No information was observed to support any assumptions of normality in the data studied, and the analysis proceeds using a variety of nonparametric techniques.

In Section II we describe the data used in this study and discuss how it was used. In Section III we discuss each of the alternative imputation strategies that are compared against one another. In Section IV the methods used to compare the different procedures are described and the results of our analyses are presented. Findings are summarized in Sections V and VI, and an Appendix contains the tables, graphs, and summary statistics used in our analyses.

I. SIMULATING MISSING DATA PATTERNS

Twelve-month longitudinal data extracted from the 1979 ISDP (Income Survey Development Program) survey were used in this study. These data were entered into a SIR (Scientific Information Retrieval) database, from which fixed-format simulation data files were extracted. Subsequent manipulation and evaluation were performed using special purpose FORTRAN programs and the SPSS-X statistical package on a Univac 1100 and IBM-XT at the Bureau of the Census.

For this study, missing data were simulated using records on which the variable of interest was completely reported, and for technical reasons records with zero responses for the variable of interest were excluded. We then had the original values for the missingness simulated in the file which we use later in analyzing the properties of imputations obtained by the selected imputation methods. The continuous variable used in the study is wages and salary. The following indicates the simulation procedure used to induce missing data onto records.

(1) Define a longitudinal record for wages and salary to be a person record of responses to the question: What were your wages and salaries for month j, j=1, 12 in 1979? Denote the response by wage (j), j=1,12.

	1	2	3	4	5	6	7	8	9	10	11	12
(j):												
Ex: Wage(j):	100	100	150	145	120	200	150	200	100	100	150	175

Longitudinal record

(2) Randomly select 500 person records for persons, age ≥ 16, with at least one missing response, i.e., wage (j) = -1 for some j, and at least one complete response, i.e., wage (j) > 0 for some j. (The value "-1" is a place holder for a missing response.)

	1	2	3	4	5	6	7	8	9	10	11	12
(j):												
Ex: Wage(j):	100	100	-1	-1	-1	100	100	100	150	150	150	150

(3) Select approximately 2,000 person records with complete responses for every

month (j), i.e., wage (j) > 0 for all j=1, 12.

(4) Induce the missing pattern from a record in the set (2) onto a record for a

complete respondent in set (3) by a nearest match procedure. That is, let $X_{n,j}$=wage (j) for some case n from data set (2) and let $Y_{i,j}$ = wage (j) for some case i from data set (3), and find the record Y_i in the set (3) to minimize:

$$\sum_{j=1}^{12} (X(n,j) - Y(i,j))^2.$$

We set X(n,j)-Y(i,j) = 0 for X(n,j) missing.

One then induces the n^{th} missing data pattern from (2) onto the i^{th} full respondent in (3) to obtain 500 simulated person records with missing wave responses. In all, 410 unique complete respondents were used in simulating the 500 records with induced missing responses.

III. SEVEN IMPUTATION PROCEDURES

The seven imputation procedures examined in this study are described below. The first three employ regression-type techniques which utilize the entire data set to (1) model the missingness that occurs in the entire set of data and (2) derive model-based imputes for the missing values. The last four procedures implement averaging techniques in which only data for the current case are used in determining an impute for a missing month's value. The regression-based imputation procedures: Iterated Buck, Logarithmic Iterated Buck, and Cube Iterated Buck; and the four averaging techniques: Arithmetic Smoothing (1) and (2) and Multiplicative Smoothing (1) and (2) were tested and evaluated on the simulated data set described above.

(a) Iterated Buck Techniques

The Iterated Buck procedure is a sequential regression technique that estimates regression parameters, derives imputes based on these parameters, and repeats this

process until the sequence of estimated parameters converge. For a detailed description and derivation of the Iterated Buck method the reader is referred to papers by S. F. Buck [2] and Beale and Little [1] pertaining to missing values in multivariate analysis. The important thing to note here is that Iterated Buck is an EM-Algorithm that gives maximum likelihood estimates of the population parameters when there is the assumption that the data have a multivariate normal distribution.

However, no distributional assumption of normality of the data is justified here, as indicated in Figures 1-4. Histograms of the residuals, observed value - imputed value, for Iterated Buck, Logarithmic Iterated Buck and Cube Iterated Buck are presented with a normal overlay represented by the dotted line on the histograms. Comparing the two distributions in each of the histograms suggests that a normality assumption for the data is unjustified. Even in the absence of normality the Iterated Buck method can be used to derive imputations. Of course, since the data are not normal, our analysis will proceed along nonparametric lines, and considerations especially appropriate to normal data will not be addressed.

We now describe the steps involved in the Iterated Buck procedures. Assume for a set of N observations and n variables that $x_{i,j}$ represents the value of the j^{th} variable in the i^{th} observation for $j=1,\ldots,n$ and $i=1,\ldots,N$. Let m_j denote the sample mean value of the j^{th} variable over all complete observations and $u_{j,k}$ denote the sample covariance between variables m_j and m_k over all complete observations. The Iterated Buck method uses m_j and $u_{j,k}$ to compute:

(1) $x_{i,j}$ = $x_{i,j}$, if $x_{i,j}$ is observed
a linear combination of the set of variables observed in the i^{th} observation, otherwise

(2) $c_{i,j,k}$ = partial covariance of m_j and m_k if $x_{i,j}$ and $x_{i,k}$ are both unknown
0, otherwise

(3) $\bar{x}_j = \sum_{i=1}^{N} x_{i,j}/N$,

(4) $a_{j,k} = \sum_{i=1}^{N} (x_{i,j} - \bar{x}_j)(x_{i,k} - \bar{x}_k) + c_{i,j,k}$.

Set $m_j = \bar{x}_j$ and $u_{j,k} = a_{j,k}/(N-1)$ and repeat (1) thru (4) until there are no further changes in m_j and $u_{j,k}$. The term $c_{i,j,k}$ is a correction term for the bias that would normally occur in the formation of $a_{j,k}$. The procedure is applied to a longitudinal record for the variable wages and salaries by setting $x_{i,j} = AMT(i,j)$ for person record i, i=1,N and month j, j=1,12. If instead $x_{i,j} = \log(AMT(i,j))$ for the i^{th} person record and j^{th} month, this is the reason we omitted records containing zero responses, the resulting algorithm is called the Logarithmic Iterated Buck. After the algorithm is satisfied, $x_{i,j}$ is transformed back to original amounts and corresponding imputes. By using the logarithm of amounts of wages and salaries one reduces the impact of skewness in the data and avoids the problem of generating negative imputes. Similarly, Cube Interated Buck operates on $x_{ij} = (AMT(i,j))^{1/3}$ until closeness criteria are met. The $x_{i,j}$ are transformed back to original values and corresponding imputes.

(b) Smoothing Procedures

The two averaging techniques examined here are termed Arithmetic Smoothing and Multiplicative Smoothing because the imputes are based on the arithmetic mean and geometric mean respectively.

Arithmetic Smoothing essentially allocates an equal additive subdivision to each missing value which depends on the length of the interval of missing values in the data record and the reported values on either side of the missing data. For example, suppose March, April, and May values for variable X, denoted $x_{m,j}$, j = 3,5, were missing for a particular record. Then the record looks like the following:

J	F	M	A	M	J	J	A	S	O	N	D
x_1	x_2	$x_{m,3}$	$x_{m,4}$	$x_{m,5}$	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
		missing interval									

We determine the difference in the bounding reported values of the missing interval and divide by the number of subintervals to arrive at

$$d = \frac{x_6 - x_2}{4} .$$

We then add d to x_2 consecutively to obtain imputes for $x_{m,3}$, $x_{m,4}$ and $x_{m,5}$. Explicitly,

$$x_{m,3} = x_2 + d$$
$$x_{m,4} = x_2 + 2d$$
$$x_{m,5} = x_2 + 3d .$$

For the general case, let $\underline{r} = (x_1,\ldots,x_{12})$ be a longitudinal record of amounts. Suppose x_m is a missing response bound below by x_i and above by x_j.

(1) Compute $k = j-i$

(2) Compute $d = (x_j - x_i)/k$

then (3) $x_m = x_i + (m-i)d$.

Note that $x_j = x_i + k \cdot d$.

One difficulty with this method is that bounds may not exist around missing responses, specifically, when endpoints (month (1) and/or month (12)) of the record are missing. Two solutions to this problem are examined. The first solution is to substitute the arithmetic mean of the record's complete responses, $(\sum_{i=1}^{p} x_i)/p$, where p is the number of reported responses, into the endpoints whenever one or both endpoints of the record is missing. The second solution is to substitute the arithmetic mean of the two nearest values for missing endpoints. Numerical comparisons of both methods are included with all other results at the end of this report.

Multiplicative Smoothing basically conforms to the same principles as Arithmetic Smoothing with the difference that the geometric mean of a missing interval's bounding responses is employed, and equal multiplicative subdivisions are allocated to missing values in an interval of missing responses. That is, for Multiplicative Smoothing we determine the quotient of the bounding reported values of the missing interval and base

our imputation on that value. For the general case let $\underline{r} = (x_1,\ldots,x_{12})$ be a longitudinal record of amounts and let x_m denote a missing response bound below by x_i and above by x_j.

(1) Compute $k = j-i$

(2) Compute $q = (x_j / x_i)^{1/k}$

Then (3) $x_m = x_i \cdot q^{(m-i)}$.

Note that $x_j = x_i \cdot q^k$.

The two methods used to correct for missing endpoints on a record corresponding to the situation for Arithmetic Smoothing were, (1) use the geometric mean of the record's complete responses, $\left(\prod_{i=1}^{p} x_i\right)^{1/p}$, and (2) use the geometric mean of the nearest two values for any missing endpoints.

It should be noted that Multiplicative Smoothing of amounts of wages and salaries and Arithmetic Smoothing of the logarithm of amounts of wages and salaries give identical results. The following shows the relationship between the two procedures.

The basis for Multiplicative Smoothing is that for some missing interval of length k bounded below by x_a and above by x_b, and with x_m missing in that interal, $(a \leq m \leq b)$,

(1) $x_m = x_a \cdot q^{(m-a)}$ where $q = (x_b / x_a)^{1/k}$.

Taking the logarithm of (1) gives

(2) $\log x_m = \log x_a + (m-a) \log q$.

and by setting $y_a = \log x_a$ and $y_m = \log x_m$ we get

(3) $$\log q = \frac{\log x_m - \log x_a}{(m-a)}$$

$$= \frac{y_m - y_a}{(m-a)}.$$

Letting log q equal d and substituting into (2) we obtain

(4) $y_m = y_a + (m-a)d$

which is the basis for Arithmetic Smoothing as discussed above.

IV. COMPARING THE PROCEDURES

There are several questions to be addressed when analzying the effectiveness and efficiency of an imputation procedure, and by focusing on these questions particular imputation procedures can be identified that maximize the desired end results. The final decision as to which imputation strategy is best to use for particular survey items must rest with subject-matter specialists who are familiar with the subject-matter of the survey, the questionnaire form, and the underlying target population. In this report, we present a number of descriptive statistics for each of the procedures described above. These can be compared against one another and serve as a basis for an informed decision as to which procedure is to be preferred. In general, the questions that must be addressed are:

(1) What does a completely reported data record look like? Is it typically reported consistently, erratically, in particular patterns, or does it follow some distribution?

(2) What are the imputations expected to accomplish? Should the derived imputation resemble the reported data, implement a presumed relationship, or smooth over the missingness?

(3) What criteria should be used to evaluate and compare methods?

The data for wages and salary are at times reported consistently across a 12-month period, reported erratically other times, and may or may not follow a particular pattern of responses based on ISDP waves (the 3-month interval to which a questionnaire refers). Ideally, the optimal imputation procedure would adhere to patterns of consistency or erraticism of the reported data for each individual person record.

As discussed in Section II, we start with completely reported longitudinal records and then blank out responses conforming to missing patterns from a set of longitudinal records having nonresponse. We then impute for the induced nonresponse and compare the imputes with the original values that were blanked out. These comparisons form the basis of our analysis. As noted earlier, normality assumptions are not supported by the data, and accordingly, the analysis is nonparametric.

We let

$$\mathbf{x}_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,12})$$

be the i^{th} completely reported record from a set of N records, and we assume that for some month j, j = 1,12, $x_{i,j}$ was blanked out. We then have the following:

$x_{i,j}$ = The amount of wages and salaries for the j^{th} month of the i^{th} person record,

$\hat{x}_{i,j}$ = Imputed value of $x_{i,j}$ from some imputation procedure,

$r_{i,j} = x_{i,j}/x_{i,j+1}$, and

$\hat{r}_{i,j} = x_{i,j}/x_{i,j+1}$ where at least one of $x_{i,j}$ or $x_{i,j+1}$ was imputed.

The analytical variables computed and evaluated for each imputation method are

(1) $c_{i,j} = x_{i,j} - \hat{x}_{i,j}$

(2) $c_{i,j} = (x_{i,j} - \hat{x}_{i,j})/x_{i,j}$

(3) $c_{i,j} = r_{i,j} - \hat{r}_{i,j}$

(4) $c_{i,j} = (r_{i,j} - \hat{r}_{i,j})/r_{i,j}$.

Note that:

(a) $x_{i,j} - \hat{x}_{i,j}$ represents the difference between original value and imputed value,

(b) $(x_{i,j} - \hat{x}_{i,j})/x_{i,j}$ represents the relative difference,

(c) $r_{i,j} - \hat{r}_{i,j}$ represents the difference between the ratio of adjacent months when one was imputed, and

(d) $(r_{i,j} - \hat{r}_{i,j})/r_{i,j}$ measures the relative difference of these ratios.

The statistics we will use to examine these analytic variables are

(i) $$S_1 = \sum_{i=1}^{N} \sum_{j=1}^{12} c_{i,j} ,$$

(ii) $$S_2 = \sum_{i=1}^{N} \sum_{j=1}^{12} c_{i,j}^2 ,$$

(iii) $$S_3 = \left(\sum_{i=1}^{N} \sum_{j=1}^{12} c_{i,j} \right)/n ,$$

(iv) $$S_4 = \sum_{i=1}^{N} \sum_{j=1}^{12} (c_{i,j} - \bar{c})^2/n ,$$

where n is the total number of cases for which $c_i \neq 0$ and

$$\bar{c} = \left(\sum_{i=1}^{N} \sum_{j=1}^{12} c_{i,j} \right)/n .$$

V. OBSERVED RESULTS OF THE COMPARISONS

Table 1 contains numerical comparisons for analytical variable $c_{i,j} = x_{i,j} - \hat{x}_{i,j}$. The seven imputation procedures are listed horizontally and the four derived statistics used for evaluation are listed vertically. If one of the smoothing imputation methods has a (1) appended to its name, the method substitutes the mean of all reported months for missing endpoints of a record; if a (2) is appended to the name of the procedure, the mean of the two nearest reported values are substituted for missing endpoints. Table 2 presents the numerical results for the analytical variable $c_{i,j} = (x_{i,j} - \hat{x}_{i,j})/x_{i,j}$ and is set up identical to Table 1. In both Table 1 and Table 2, n = 3183 in calculations of S_1 through S_4. Tables 3 and 4 contain, respectively, the numerical results for the two analytical variables $c_{i,j} = r_{i,j} - \hat{r}_{i,j}$ and $c_{i,j} = (r_{i,j} - \hat{r}_{i,j})/r_{i,j}$. In imputing S_1 and S_4, n = 2820 and the table formats, for Table 3 and 4 follow that of Tables 1 and 2.

(a) Tables 1-4

In examining numerical results presented in table 1, it appears that the Arithmetic Smoothing (1) method and the Logarithmic Iterated Buck method both perform well when using $c_{ij} = x_{i,j} - \hat{x}_{i,j}$ as comparison criteria. The sum and mean of the $c_{i,j}$'s are closer to zero for the Arithmetic Smoothing method than all other methods, and the sum of squares and variance of the $c_{i,j}$'s are closer to zero for the Logarithmic Iterated Buck method.

Results from table 2 indicate that the Multiplication Smoothing (1) method performs well in comparison to all other methods when using the statistic $c_{i,j} = \frac{x_{i,j} - \hat{x}_{i,j}}{x_{i,j}}$ as criteria. The sum and mean of the $c_{i,j}$'s are closest to zero for the Multiplicative Smoothing (1) method. The Cube Iterated Buck method has the smallest sum of squares of the $c_{i,j}$'s and the Logarithmic Iterated Buck method has the smallest variance.

Results in both Tables 3 and 4 using

$$c_{i,j} = r_{i,j} - \hat{r}_{i,j} \text{ and } c_{i,j} = \frac{r_{i,j} - \hat{r}_{i,j}}{r_{i,j}}$$

respectively, as comparison criteria, indicate that the Cube Iterated Buck method performs better than all other methods for the sum, sum of squares, mean and variance of the $c_{i,j}$'s.

For each of the analytic variables, the better a procedure simulates an aspect of missing data, the closer the relevant derived statistic (either S_1, S_2, S_3, or S_4) will approach zero. One initial reason for carrying out this study was to determine whether the Iterated Buck method is a better imputation procedure than its counterparts, the Logarithmic Iterated Buck and the Cube Interated Buck.

The most decisive finding in this study is that for every derived statistic, the Logarithmic Iterated Buck method outperformed the Iterated Buck method. Using the logarithm of wages and salary rather than actual amounts provides a two-fold improvement over the Iterated Buck procedure by eliminating negative imputes and increasing the accuracy of the imputes. Moreover, in every statistic except the first and third on Table 1, Cube Iterated Buck outperformed Iterated Buck. From these

observations, it is clear that better results are obtained using the Buck Procedure with transformed data.

Results comparing the Logarithmic Iterated Buck method with the Cube Iterated Buck method are mixed. In Tables 3 and 4 Cube Iterated procedure performs better. Most often in Tables 1 and 2, the Logarithmic Iterated Buck procedure does better. All in all, the results are close. One interesting observation is that for the statistic

$$\sum_{i=1}^{N} \sum_{j=1}^{12} ((x_{i,j} - \hat{x}_{i,j})/x_{i,j})^2 ,$$

the Cube Iterated Buck procedure far out performs all other procedures; that is, it seems to do well for scaled residuals. On the other hand, for the statistic

$$\sum_{i=1}^{N} \sum_{j=1}^{12} (x_{i,j} - \hat{x}_{i,j})^2$$

the Logarithmic Iterated Buck procedure does best of all. For the last two analytical statistics presented in Tables 3 and 4 the Cube Iterated Buck procedure outperformed all other imputation procedures for each statistic calculated, with the Logarithmic Iterated Buck procedure a fairly close second best.

Arithmetic Smoothing (1) and Multiplicative Smoothing (1) are virtually identical on comparison and outperform their counterparts Arithmetic Smoothing (2) and Multiplicative Smoothing (2) the majority of the time. Logarithmic Iterated Buck and Cube Iterated Buck do a little better, all in all, than the smoothing techniques. However, the ease in implementing either of the two smoothing techniques may strongly argue in their favor.

(b) Figures 4-11

In Figure 4 we present a histogram of the amounts of reported wages and salaries that fall into the range \$0. to \$5,000. Histograms of values produced by each of the seven imputation procedures appear in Figures 5 through 11.

Histograms of the data completed by the Logarithmic Iterated Buck method in Figure 6, Cube Iterated Buck method in Figure 7, Arithmetic Smoothing (1) in Figure 8, and Multiplicative Smoothing (1) in Figure 9 look very much alike and also appear to

reasonably resemble the true data in Figure 1. Although histograms of Arithmetic and Multiplicative Smoothing (2) in Figures 10 and 11 look somewhat similar to the true data, there appears to be slightly more grouping of the data than in the reported data. Overall, the imputed data sets for all procedures except the Iterated Buck procedure appear to emulate the distribution of the reported data very well.

The data for this study were not edited. However one extremely large value for monthly wage and salary amount was deleted as an obvious edit failure as it caused some problems in obtaining informative graphs of the data. Unbounded histograms were produced but offered very little extra information so were not included here.

(c) Figures 12-18

Figures 12 thru 18 present scatterplots of the reported amounts of wages and salaries versus imputes obtained from each of the imputation procedures in the same order as the histograms are listed. The more linear the relationship the better the imputation procedure is in simulating the reported data on a month-to-month basis. Ideally, we would like the standard error of the estimate

$$\left(\sum_{i=1}^{n} (x_i - \hat{x}_i)^2/(n-1) \right)^{1/2}$$

to be small, the intercept near zero, and the slope close to one. Here, the index represents any month for any record when a wage and salary report x_i received an impute. The correlation and R-square values which measure the relationship between the values and the goodness of fit of the linear model, respectively, should approach one for the best method. The standard error of the estimate, intercept, and slope of the linear relationship listed at the bottom of each scatterplot all appear best overall for the Logarithmic Iterated Buck procedure, Figure 13. The Iterated Buck procedure gives a negative intercept as a result of negative imputes and the standard error of the estimate is the largest of all the methods. Statistics for Logarithmic and Cube Iterated Buck procedures compare favorably to each other. Scatterplots of the Arithmetic and Multiplicative Smoothing (1) procedures basically have the same statistics and both have smaller standard errors and intercepts closer to zero than Iterated Buck. Arithmetic and Multiplicative Smoothing (2) have the worst slope and intercept but the best fit based on the R-squared value.

(d) Figures 19-25

Histograms of scaled residuals, that is, $(x_{i,j} - \hat{x}_{i,j})/x_{i,j}$, are presented in Figures 19 thru 25. The imputation procedure used to obtain the imputations is listed at the top left of each histogram. The Iterated Buck procedure and the Logarithmic Iterated Buck procedure most often overestimate true values and all of the smoothing techniques underestimate true values. However, the Cube Iterated Buck procedures provide underestimates more often than any of the other techniques. This is determined by counting the number of negative scaled residuals in each of Figures 19 thru 25 and comparing them to the number of positive scaled residuals. The smoothing techniques tend to spike around zero, indicating that these procedures exactly estimate reported amounts more often that the Iterated Buck techniques.

(e) Brief Summary of Observations:

Based on the statistics generated as part of this analysis, the four procedures that appear best are Logarithmic Iterated Buck, Cube Iterated Buck, Arithmetic Smoothing (1) and Multiplicative Smoothing (1). The residual sum of squares presented in Table 1, Row 2, is a traditionally used comparison criterion, and based on this statistic, the Logarithmic Iterated Buck is the best procedure. When examining histograms of data completed using each of the imputation procedures to the true data, Cube Iterated Buck, Arithmetic and Multiplicative Smoothing (1) appear almost as good as Logarithmic Iterated Buck. Other statistics provided in Tables 1 thru 4 indicate that each of the four methods are favored by different criteria. The issue is to choose comparison criteria that address specific needs of the data problem at hand. Survey-specific needs should be brought to bear in assessing the merit of each of the procedures discussed.

VI. CONCLUDING REMARKS

Of the imputation procedures examined in this report, the Logarithmic Iterated Buck and Cube Iterated Buck outperformed the Iterated Buck procedure. Of the smoothing techniques, Arithmetic Smoothing (1) and Multiplicative Smoothing (1) outperformed Arithmetic Smoothing (2) and Multiplicative Smoothing (2), respectively. All Iterated Buck procedures must consider a sample of cases with missing values to derive parameters for imputing for nonresponse. Both smoothing techniques need only consider one record at a time and bounding values when deriving an imputation for nonresponse.

A variety of summary statistics are presented. One conclusion to be drawn from this simulation study is that the "simple" imputation procedures, i.e., the Arithmetic and Multiplicative Smoothing procedures, work quite well and the additional cost and complexity of using the more complicated procedures, i.e., the Iterated Buck procedures, may not be worth the effort.

In this report we did not add variability to the imputes in the form of a residual. To the extent that this is a comparative study, we felt adding residuals could be omitted at this stage. Of course, in implementing any one of these procedures, one may add some random components. Variability can be computed from the entire data set and added into each imputation or computed on a record-by-record basis where the variability added to the imputes for each record is based on the record under consideration. An alternate form to adding variability on a record-by-record basis is to split the data file into two or more groups of records. One group might contain cases that report consistently over time and the other group might contain erratic data reporters. The variability added to each record will be determined by the group in which the record lies.

Acknowledgement

I would like to thank Brian Greenberg for suggesting this research and providing a number of helpful recommendations along the way.

REFERENCES

1 Beale, E.M.L. and Little, R.J.A. (1975). Missing values in multivariate analysis. J.R. Statistical Society, B. 37, 129-146.

2 Buck, S.F. (1960). A method of estimation of missing values in multivariate data suitable for use with an electronic computer. J.R. Statistical Society. B. 22, 302-306.

FIGURE 1

HISTOGRAM OF RESIDUALS

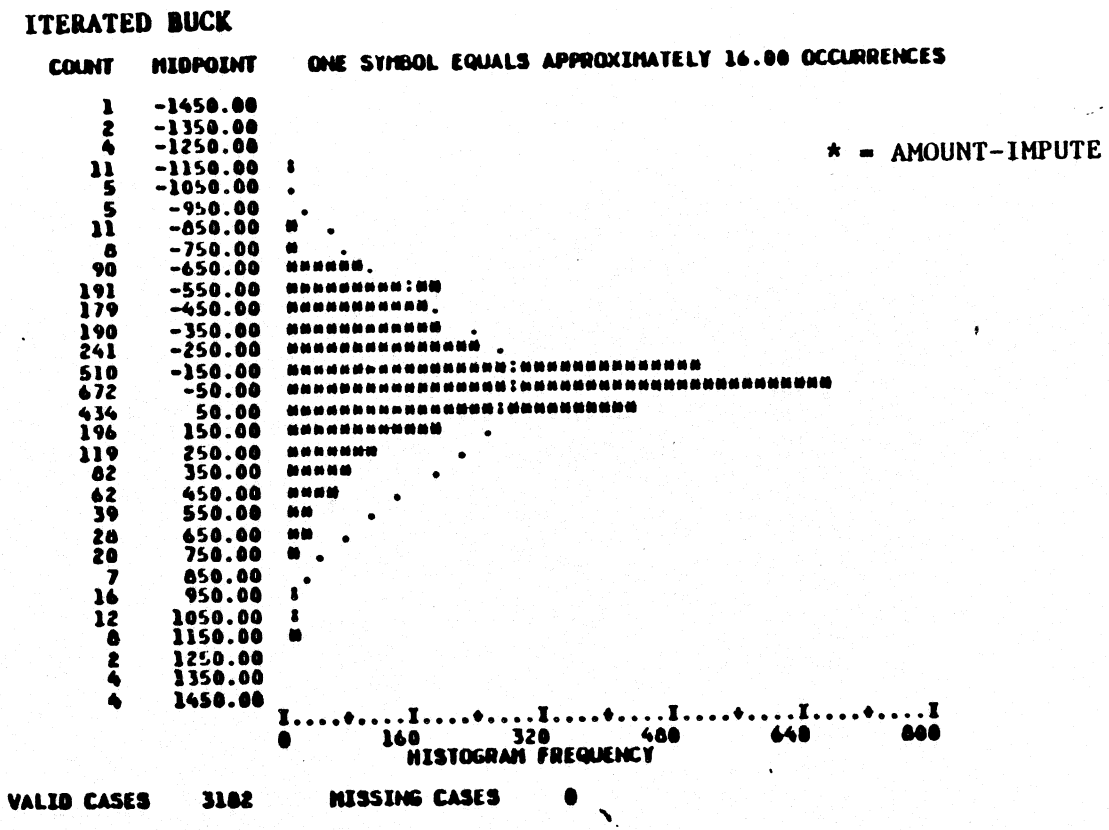
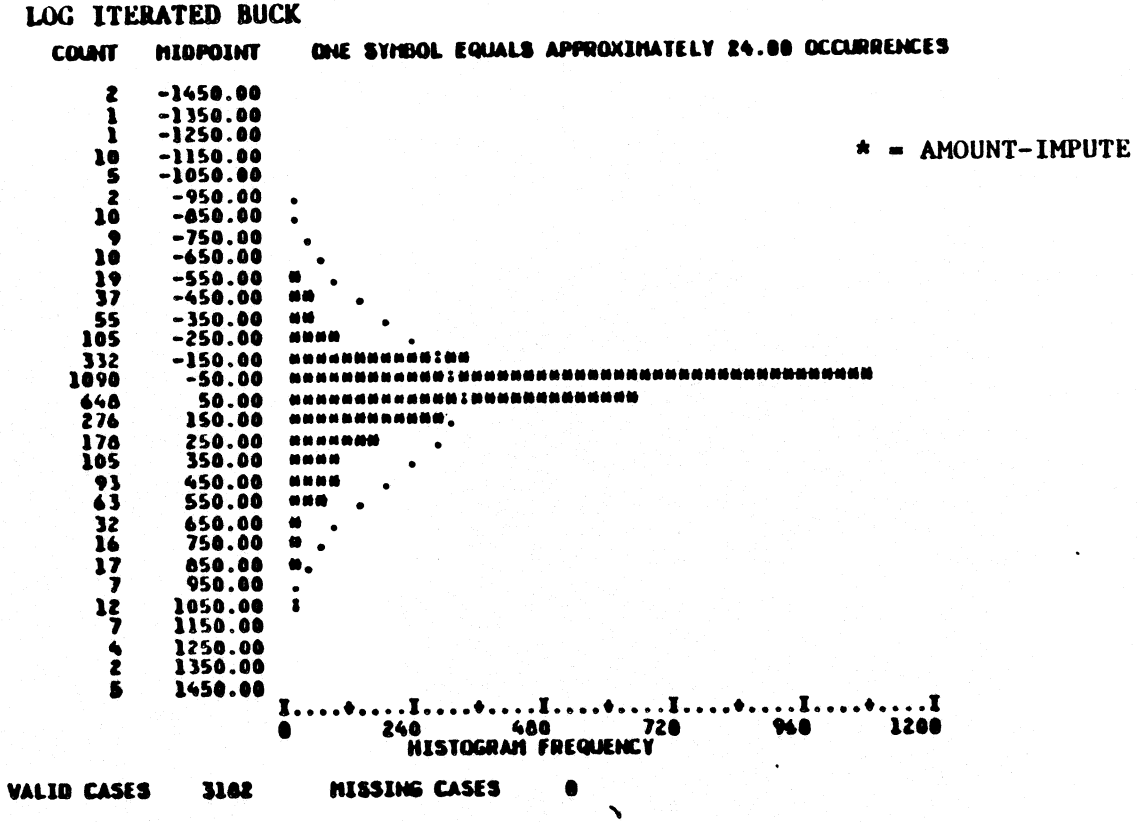


FIGURE 2

HISTOGRAM OF RESIDUALS



VALID CASES 3182 MISSING CASES 0

FIGURE 3

HISTOGRAM OF RESIDUALS

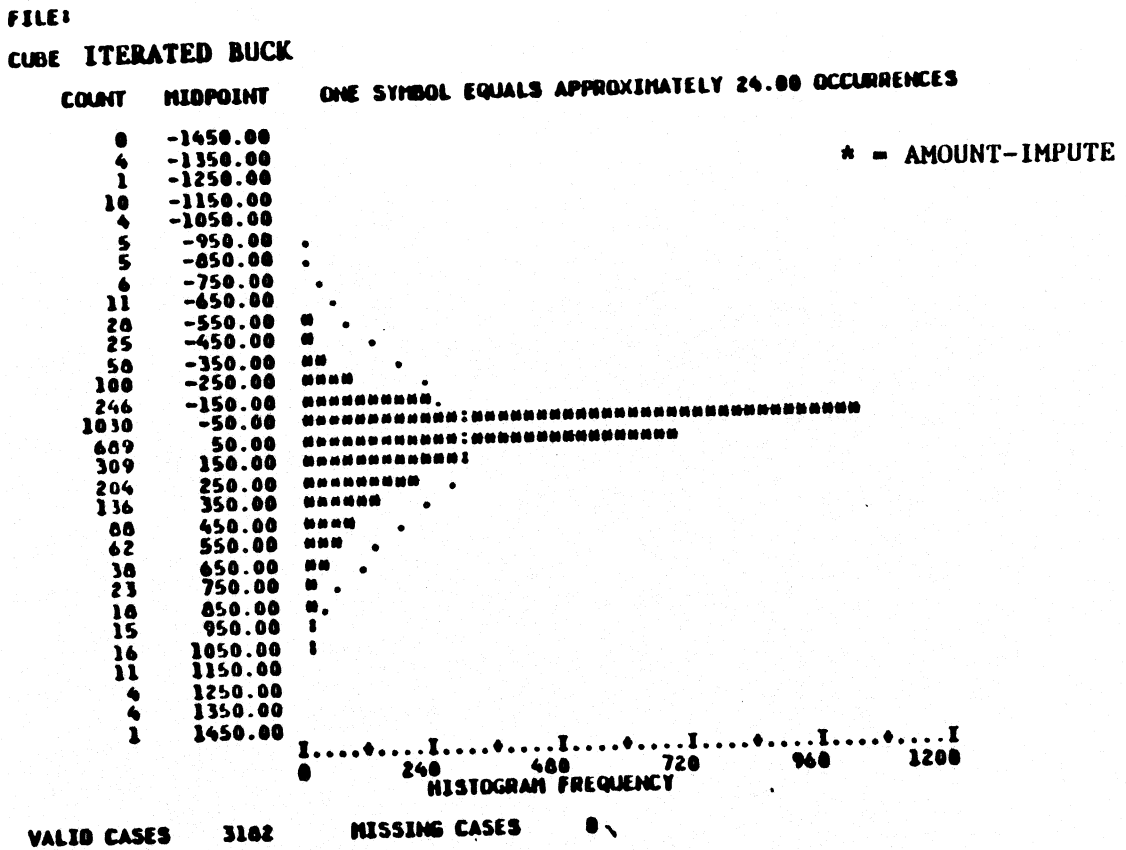


FIGURE 4

HISTOGRAM OF REPORTED AMOUNTS

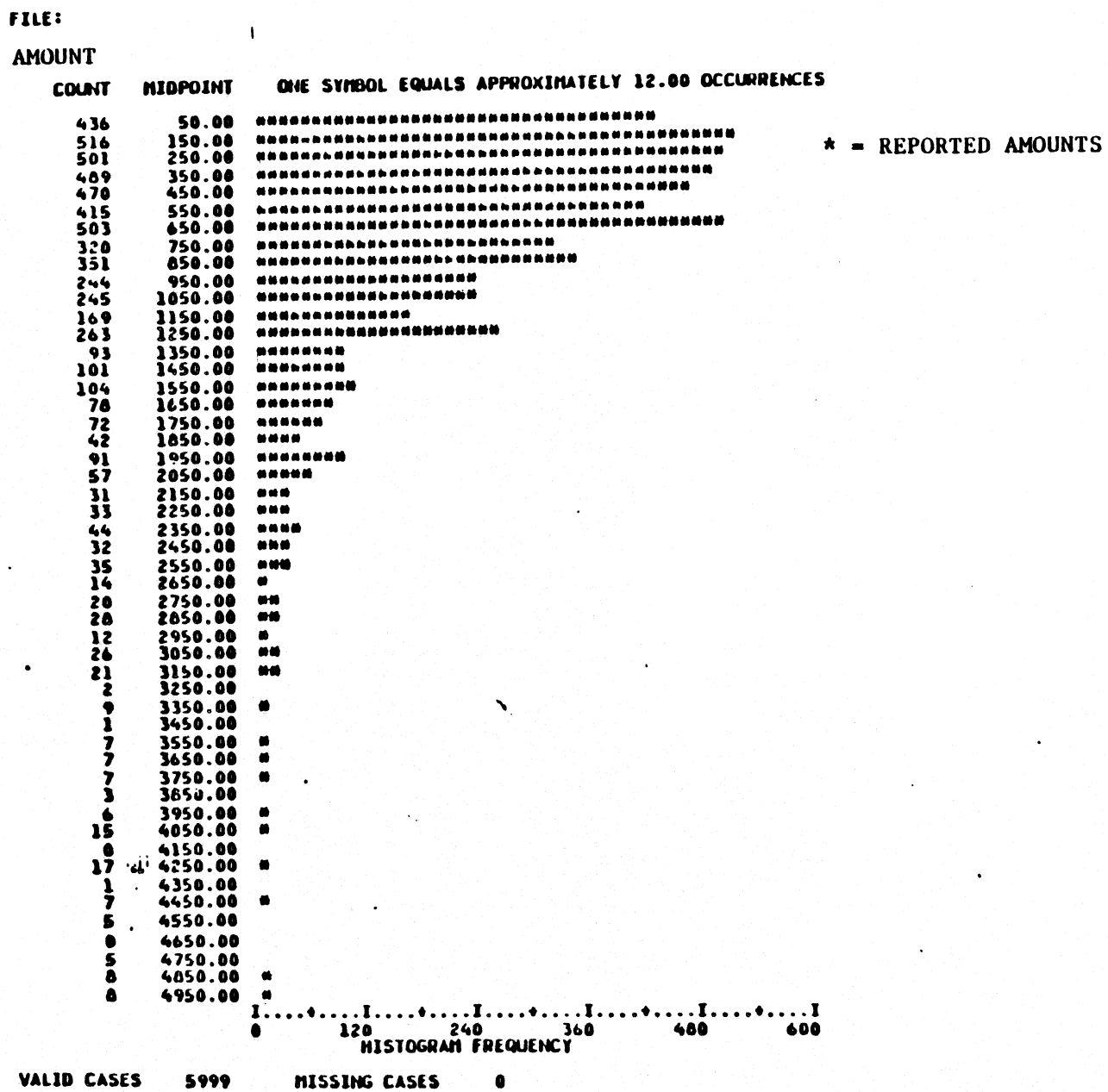


FIGURE 5

HISTOGRAM OF DATA COMPLETED BY IMPUTATION

FILE:

ITER\TED BUCK

COUNT	MIDPOINT	ONE SYMBOL EQUALS APPROXIMATELY 12.00 OCCURRENCES
198	50.00	##################
392	150.00	################################## * = IMPUTED AMOUNTS
444	250.00	######################################
480	350.00	##
390	450.00	##################################
415	550.00	###################################
583	650.00	##
493	750.00	###
457	850.00	##
311	950.00	##########################
303	1050.00	#########################
239	1150.00	####################
232	1250.00	###################
128	1350.00	##########
137	1450.00	##########
125	1550.00	##########
83	1650.00	#######
76	1750.00	######
45	1850.00	####
62	1950.00	#####
31	2050.00	###
28	2150.00	##
26	2250.00	##
40	2350.00	####
21	2450.00	##
20	2550.00	##
19	2650.00	##
22	2750.00	##
16	2850.00	#
7	2950.00	#
18	3050.00	##
20	3150.00	##
9	3250.00	#
5	3350.00	
2	3450.00	
9	3550.00	#
8	3650.00	#
8	3750.00	#
3	3850.00	
6	3950.00	#
7	4050.00	#
2	4150.00	
9	4250.00	#
1	4350.00	
6	4450.00	#
6	4550.00	#
0	4650.00	
7	4750.00	#
7	4850.00	#
1	4950.00	

I....+....I....+....I....+....I....+....I....+....I....+....I

0 120 240 360 480 600

HISTOGRAM FREQUENCY

VALID CASES 5999 MISSING CASES 0

FIGURE 6

HISTOGRAM OF DATA COMPLETED BY IMPUTATION

FILE:

LOG ITERATED BUCK

COUNT	MIDPOINT	ONE SYMBOL EQUALS APPROXIMATELY 12.00 OCCURRENCES
422	50.00	**********************************
555	150.00	**
516	250.00	**
499	350.00	***
473	450.00	***
436	550.00	************************************
476	650.00	***
343	750.00	*****************************
334	850.00	****************************
228	950.00	*******************
264	1050.00	**********************
173	1150.00	**************
227	1250.00	*******************
112	1350.00	*********
109	1450.00	*********
111	1550.00	*********
96	1650.00	********
80	1750.00	*******
48	1850.00	****
88	1950.00	*******
40	2050.00	***
23	2150.00	**
22	2250.00	**
35	2350.00	***
23	2450.00	**
24	2550.00	**
17	2650.00	*
16	2750.00	*
16	2850.00	*
16	2950.00	*
20	3050.00	**
22	3150.00	**
6	3250.00	*
9	3350.00	*
7	3450.00	*
10	3550.00	*
12	3650.00	*
6	3750.00	*
6	3850.00	*
8	3950.00	*
5	4050.00	
2	4150.00	
8	4250.00	*
2	4350.00	
2	4450.00	
4	4550.00	
4	4650.00	
4	4750.00	
7	4850.00	*
1	4950.00	

I....+....I....+....I....+....I....+....I....+....I

0 120 240 360 480 600

HISTOGRAM FREQUENCY

* = IMPUTED AMOUNTS

VALID CASES 5999 MISSING CASES 0

FIGURE 7

HISTOGRAM OF DATA COMPLETED BY IMPUTATION

FILE:

CUBE ITERATED BUCK

COUNT	MIDPOINT	ONE SYMBOL EQUALS APPROXIMATELY 12.00 OCCURRENCES
397	50.00	**********************************
558	150.00	**
546	250.00	***
537	350.00	***
500	450.00	**
424	550.00	************************************
455	650.00	**************************************
331	750.00	****************************
354	850.00	******************************
246	950.00	*********************
263	1050.00	**********************
168	1150.00	**************
197	1250.00	*****************
105	1350.00	*********
113	1450.00	*********
115	1550.00	*********
98	1650.00	********
72	1750.00	******
37	1850.00	***
79	1950.00	*******
36	2050.00	***
24	2150.00	**
22	2250.00	**
35	2350.00	***
25	2450.00	**
22	2550.00	**
14	2650.00	*
18	2750.00	**
16	2850.00	*
15	2950.00	*
23	3050.00	**
21	3150.00	**
5	3250.00	
14	3350.00	*
4	3450.00	
9	3550.00	*
8	3650.00	*
6	3750.00	*
7	3850.00	*
3	3950.00	
9	4050.00	*
0	4150.00	
9	4250.00	*
3	4350.00	
2	4450.00	
5	4550.00	
3	4650.00	
4	4750.00	
8	4850.00	*
2	4950.00	

* = IMPUTED AMOUNTS

I....+....I....+....I....+....I....+....I....+....I....+....I
0 120 240 360 480 600

HISTOGRAM FREQUENCY

VALID CASES 5999 MISSING CASES 0

FIGURE 8

HISTOGRAM OF DATA COMPLETED BY IMPUTATION

FILE:

ARITHMETIC SMOOTHING (1)

COUNT	MIDPOINT	ONE SYMBOL EQUALS APPROXIMATELY 12.00 OCCURRENCES
508	50.00	**
565	150.00	**
496	250.00	***
529	350.00	**
503	450.00	***
394	550.00	**********************************
446	650.00	**
294	750.00	**************************
366	850.00	********************************
249	950.00	*********************
238	1050.00	********************
174	1150.00	***************
198	1250.00	*****************
113	1350.00	**********
89	1450.00	*******
99	1550.00	********
82	1650.00	*******
89	1750.00	*******
38	1850.00	***
109	1950.00	*********
51	2050.00	****
19	2150.00	**
16	2250.00	*
35	2350.00	***
25	2450.00	**
29	2550.00	**
7	2650.00	*
26	2750.00	**
21	2850.00	**
18	2950.00	**
20	3050.00	**
17	3150.00	*
4	3250.00	
6	3350.00	*
3	3450.00	
8	3550.00	*
4	3650.00	
9	3750.00	*
7	3850.00	*
3	3950.00	
11	4050.00	*
11	4150.00	*
11	4250.00	*
3	4350.00	
5	4450.00	
4	4550.00	
0	4650.00	
7	4750.00	*
6	4850.00	*
2	4950.00	

* = IMPUTED AMOUNTS

I....+....I....+....I....+....I....+....I....+....I

0 120 240 360 480 600

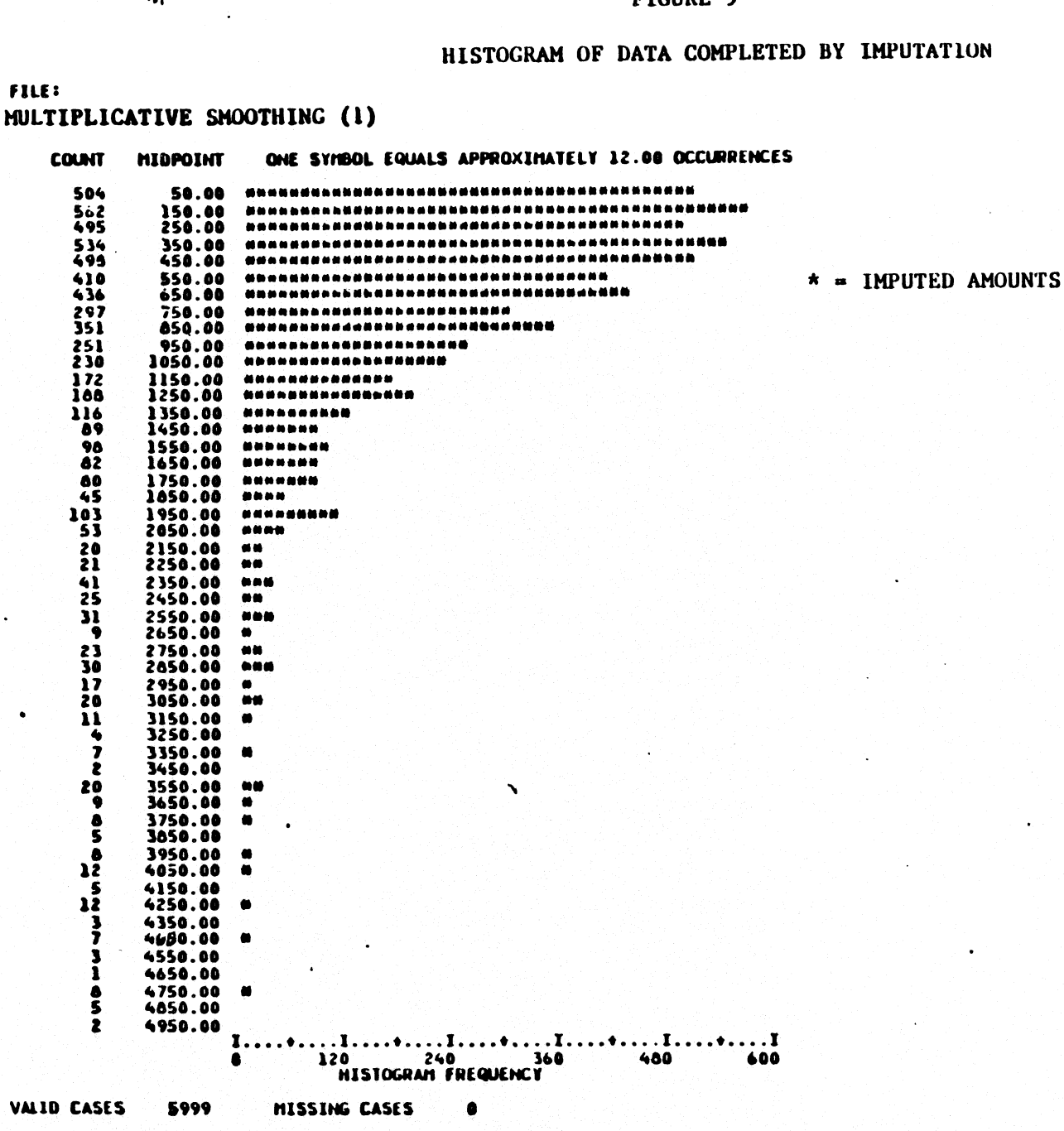
HISTOGRAM FREQUENCY

VALID CASES 5999 MISSING CASES 0

FIGURE 9

HISTOGRAM OF DATA COMPLETED BY IMPUTATION

FILE:
MULTIPLICATIVE SMOOTHING (1)



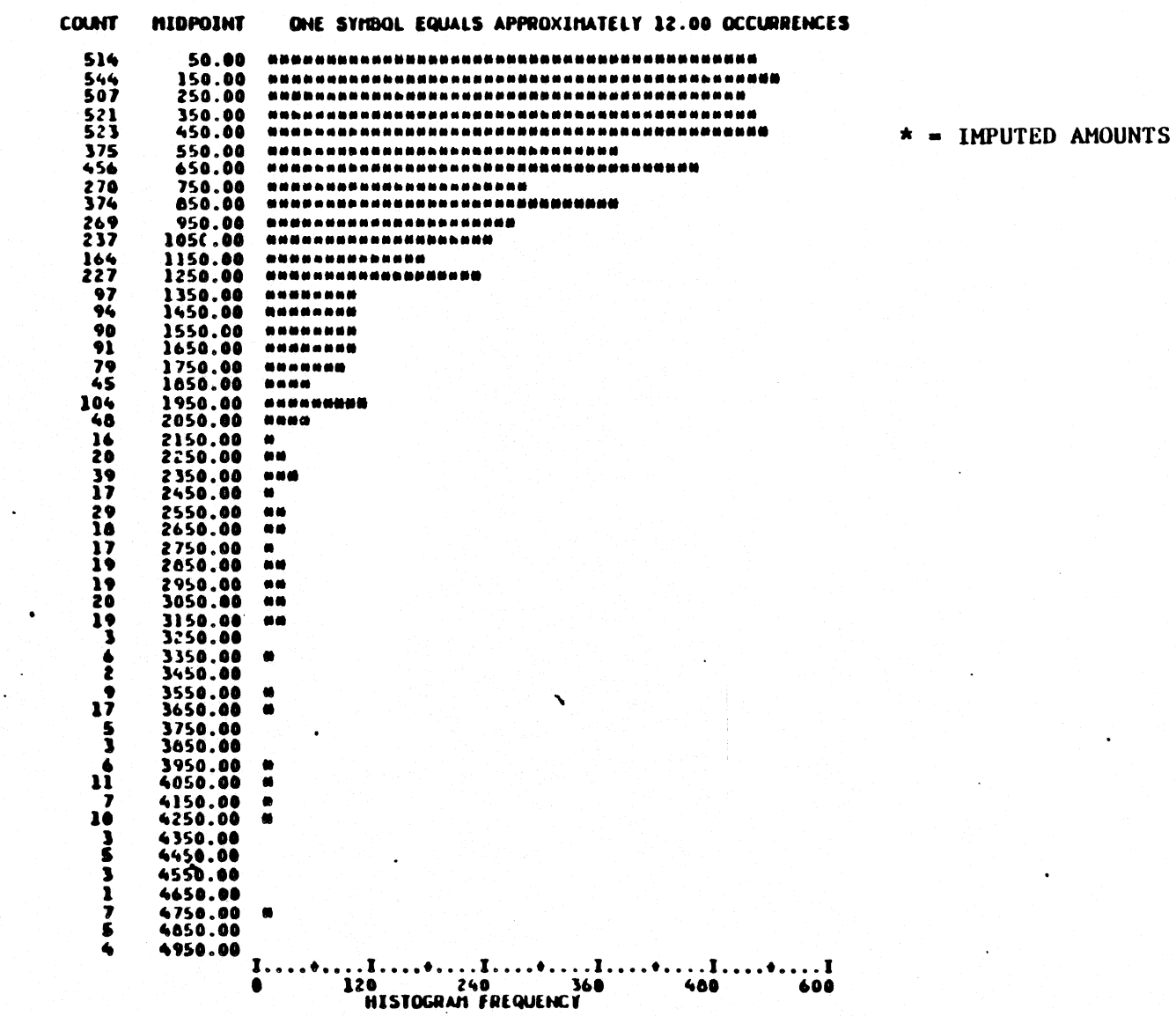
VALID CASES 5999 MISSING CASES 0

FIGURE 10

HISTOGRAM OF DATA COMPLETED BY IMPUTATION

FILE:

ARITHMETIC SMOOTHING (2)



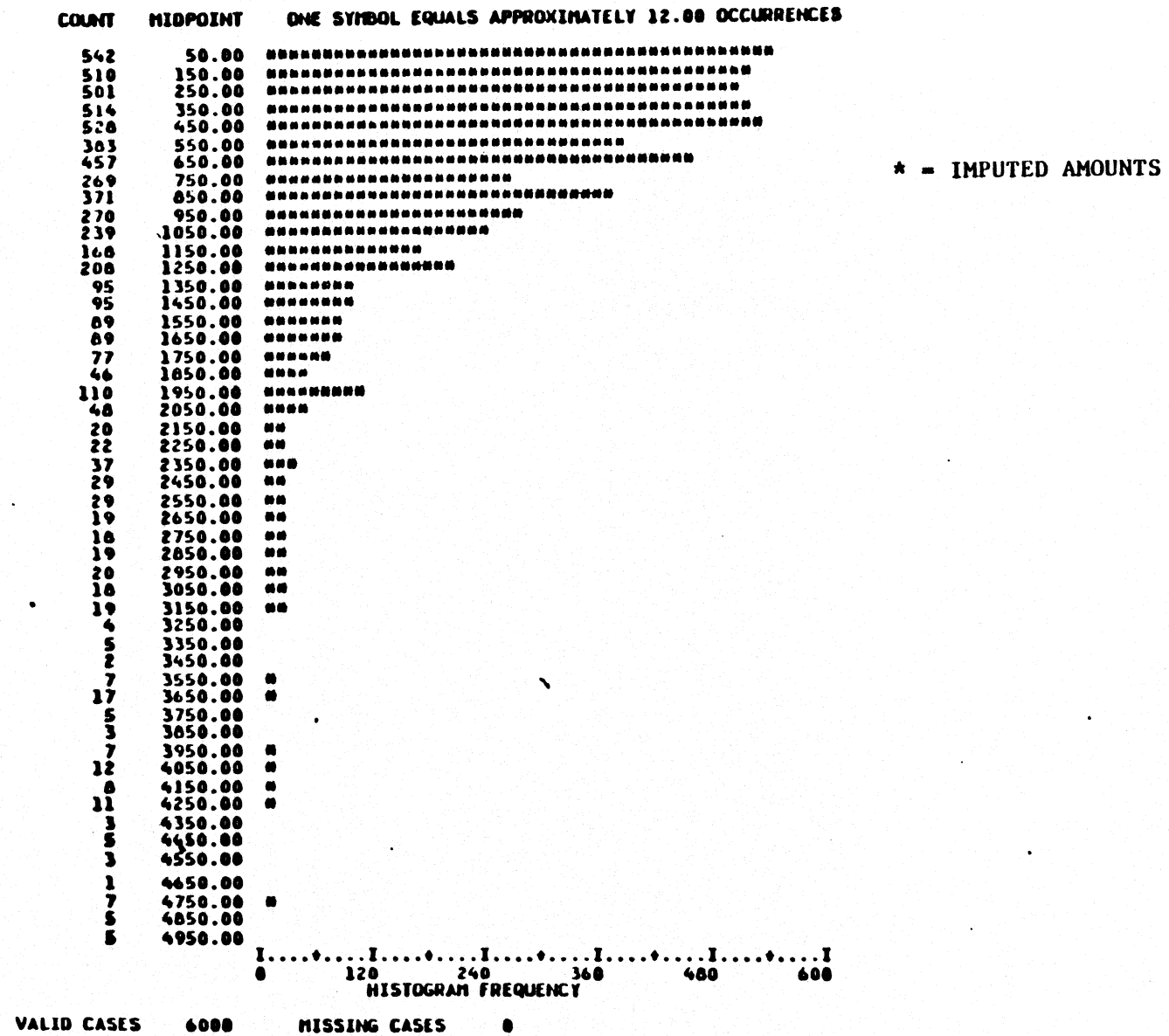
VALID CASES 6000 MISSING CASES 0

FIGURE 11

HISTOGRAM OF DATA COMPLETED BY IMPUTATION

FILE:

MULTIPLICATIVE SMOOTHING (2)

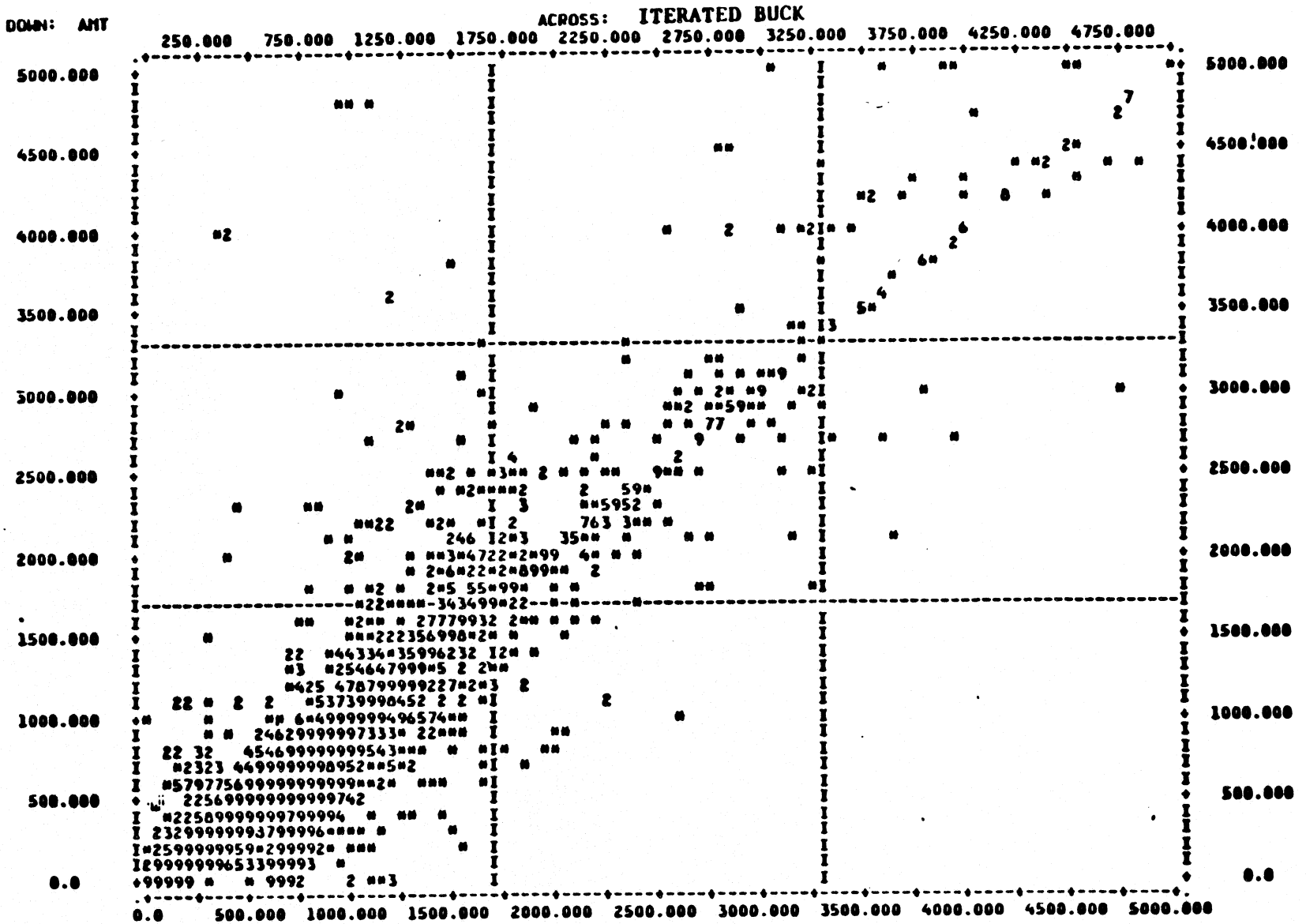


* = IMPUTED AMOUNTS

VALID CASES 6000 MISSING CASES 0

FIGURE 12

REPORTED AMOUNTS BY IMPUTED AMOUNTS



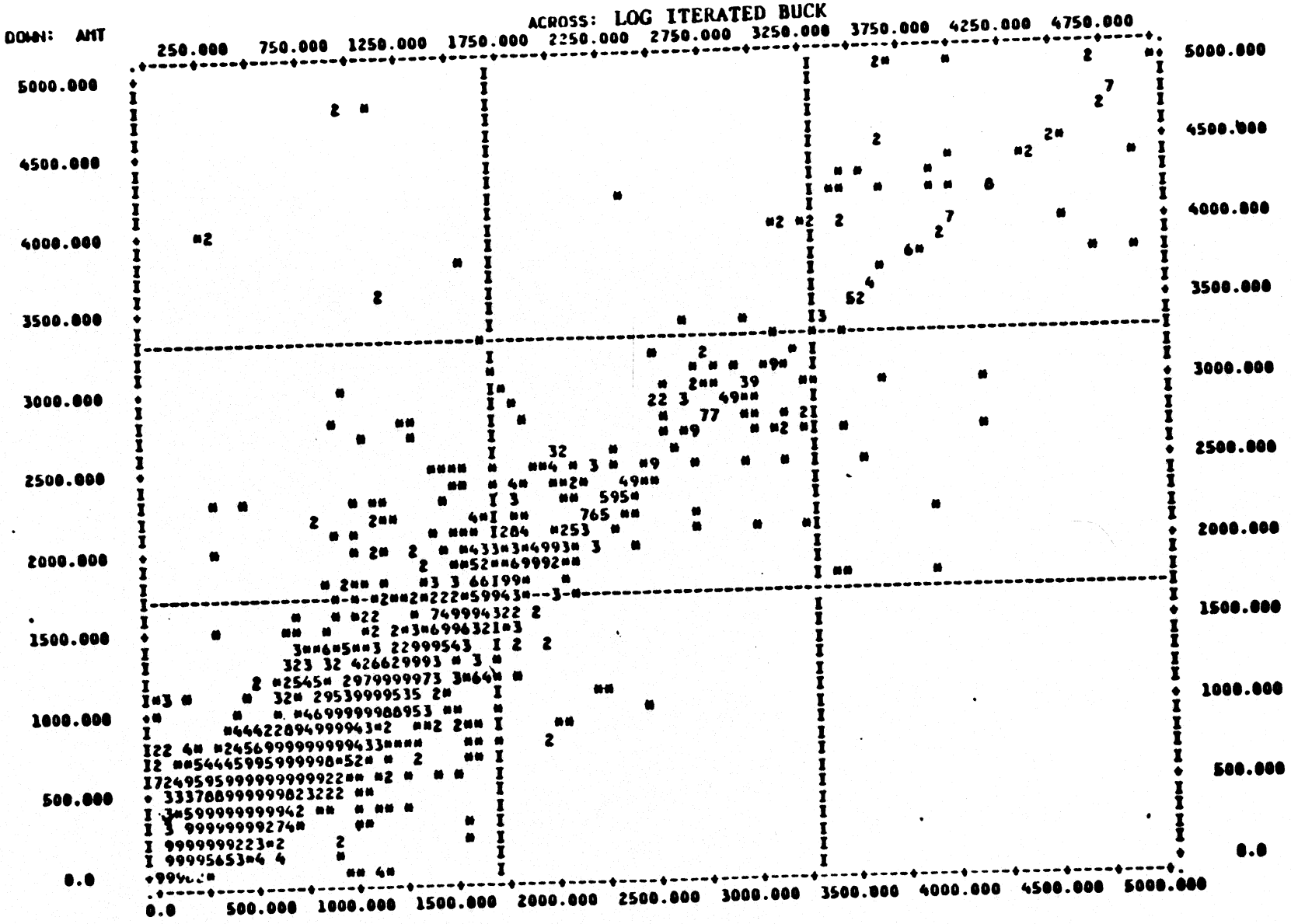
STATISTICS..

CORRELATION (R)-	.92884	R SQUARED -	.86274	SIGNIFICANCE -	.00000
STD ERR OF EST -	284.54653	INTERCEPT (A) -	-58.87632	SLOPE (B) -	1.02211
PLOTTED VALUES -	5949	EXCLUDED VALUES-	50	MISSING VALUES -	0

[redacted] COEFFICIENT CANNOT BE COMPUTED.

FIGURE 13

REPORTED AMOUNTS BY IMPUTED AMOUNTS



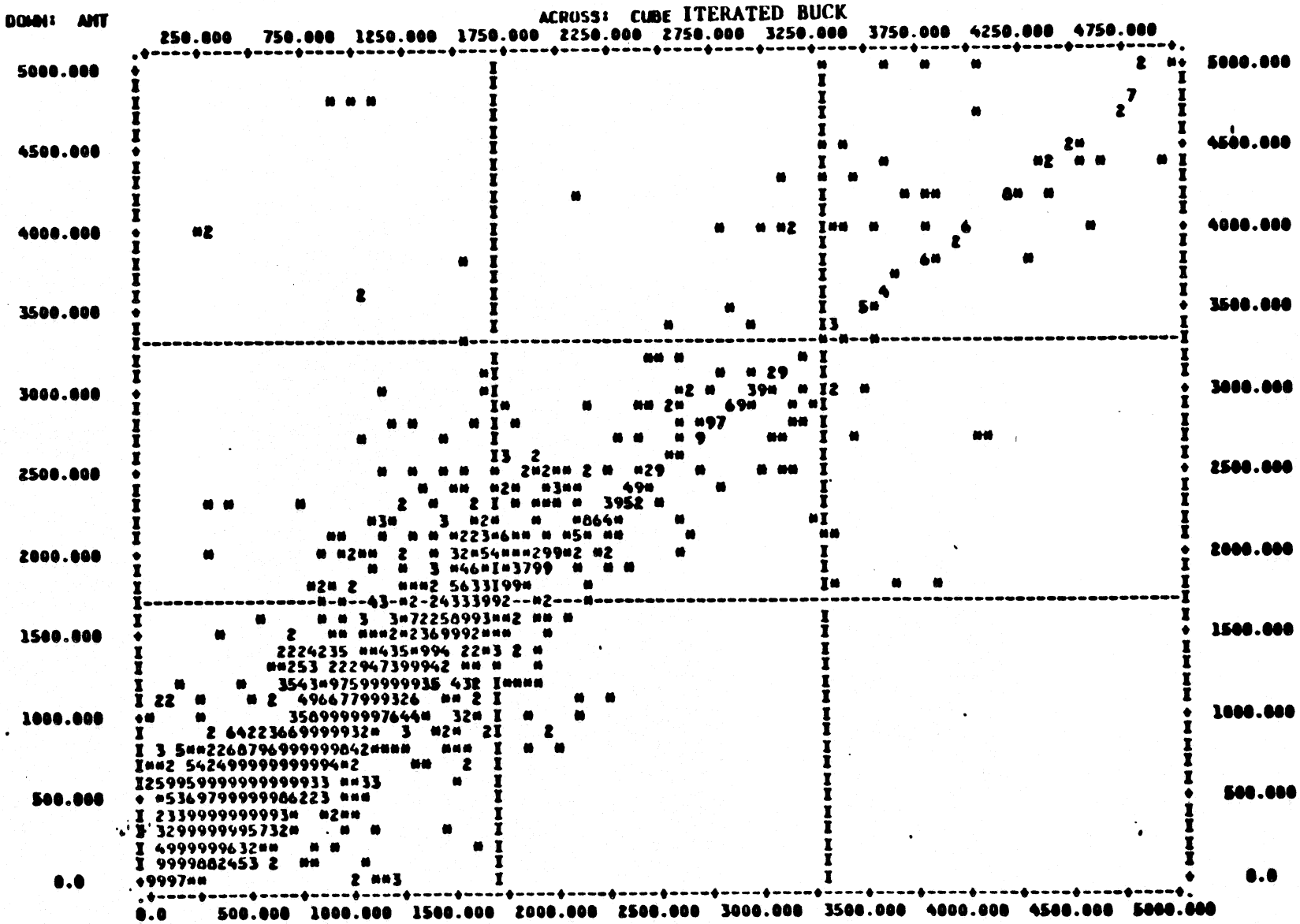
STATISTICS..

CORRELATION (R)-	.94601	R SQUARED -	.89493	SIGNIFICANCE -	.00000
STD ERR OF EST -	240.21565	INTERCEPT (A) -	29.57185	SLOPE (B) -	.99223
PLOTTED VALUES -	5958	EXCLUDED VALUES-	41	MISSING VALUES -	0

'********' IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

FIGURE 14

REPORTED AMOUNTS BY IMPUTED AMOUNTS



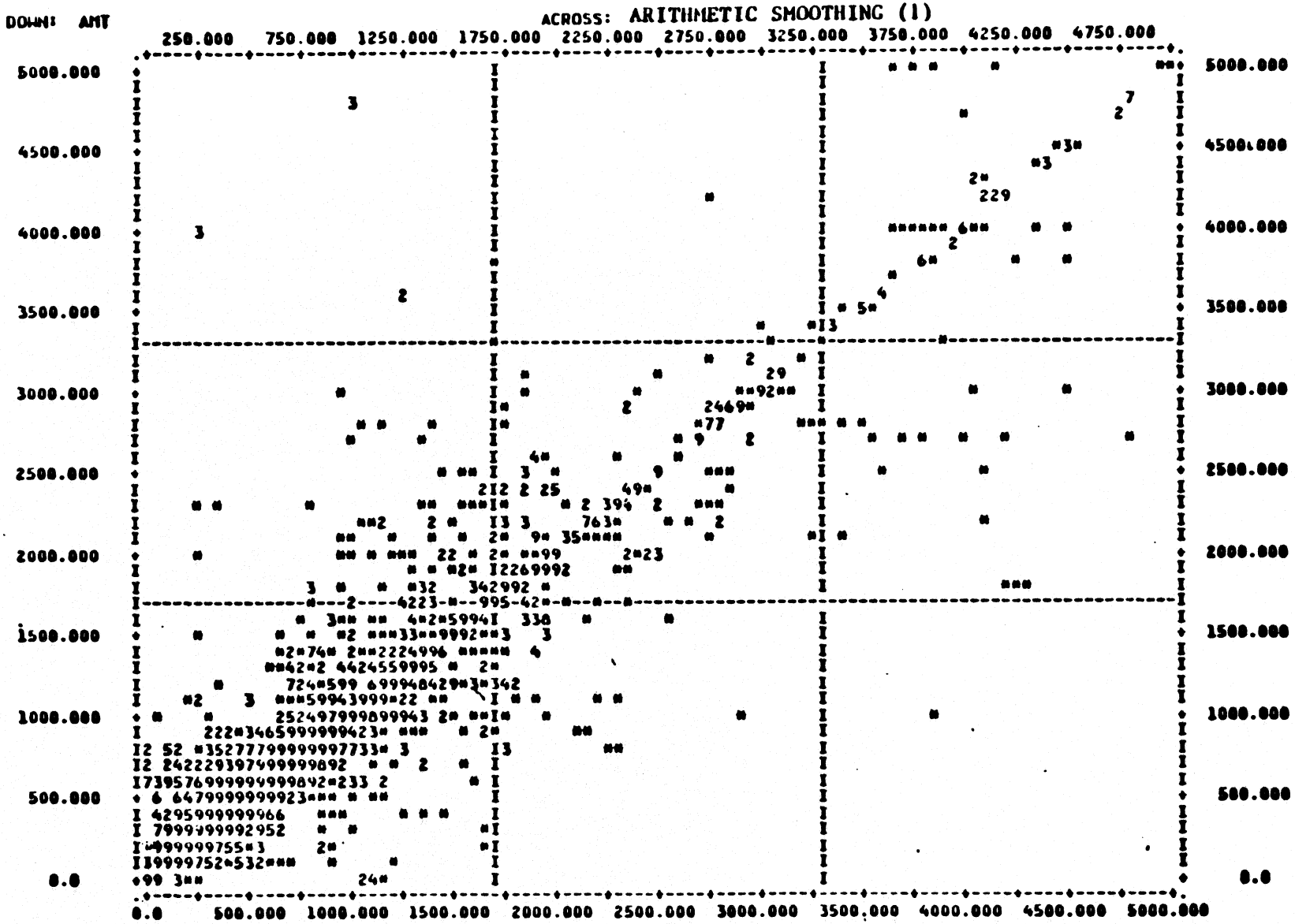
STATISTICS..

CORRELATION (R)-	.94384	R SQUARED -	.89083	SIGNIFICANCE -	.00000
STD ERR OF EST -	254.02342	INTERCEPT (A) -	34.55960	SLOPE (B) -	.99963
PLOTTED VALUES -	5959	EXCLUDED VALUES-	40	MISSING VALUES -	0

'*********' IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

FIGURE 15

REPORTED AMOUNTS BY IMPUTED AMOUNTS



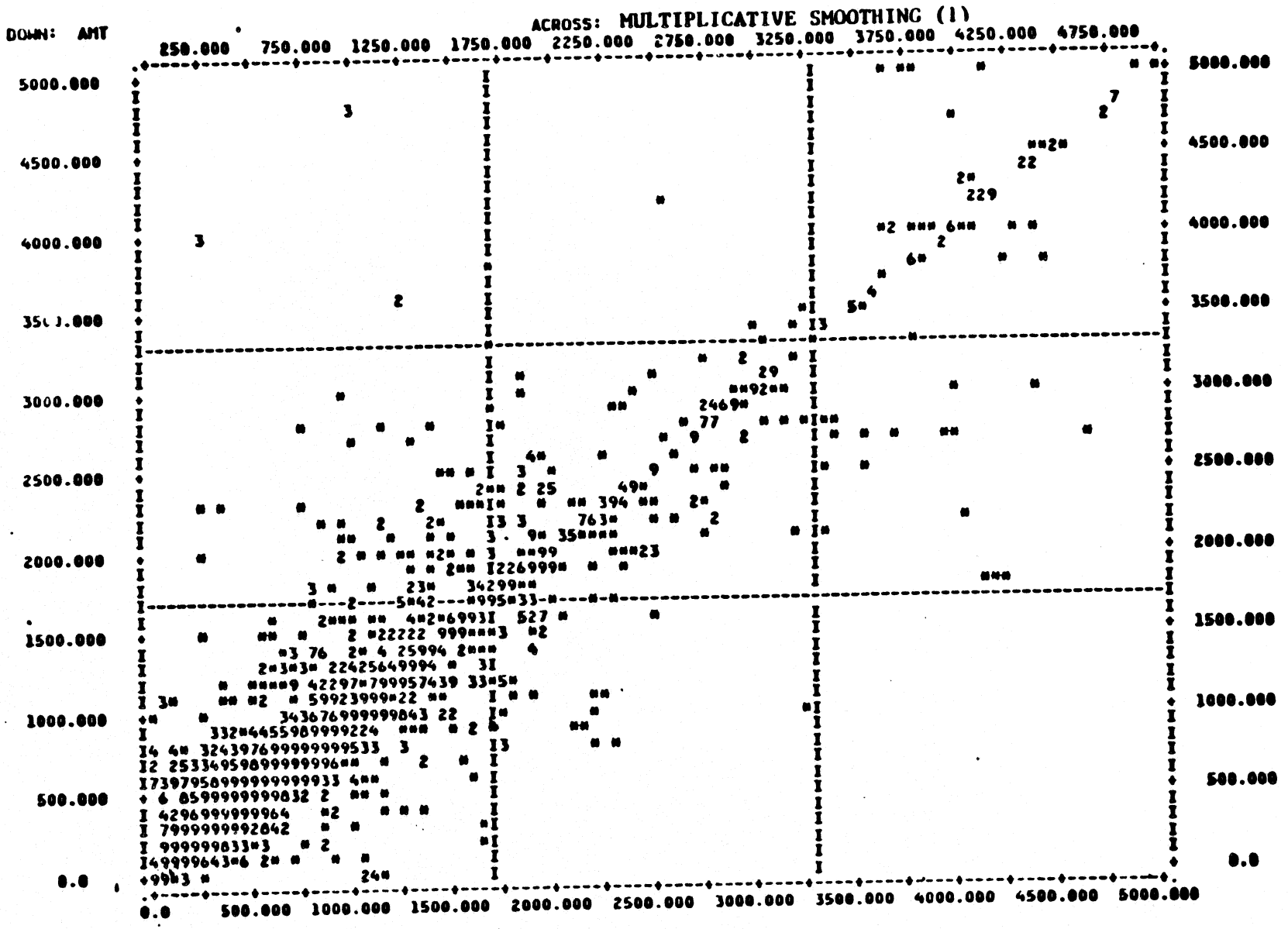
STATISTICS..

CORRELATION (R)-	.94223	R SQUARED -	.88780	SIGNIFICANCE -	.00000
STD ERR OF EST -	255.96089	INTERCEPT (A) -	61.27008	SLOPE (B) -	.95075
PLOTTED VALUES -	5957	EXCLUDED VALUES-	42	MISSING VALUES -	0

'**********' IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

FIGURE 16

REPORTED AMOUNTS BY IMPUTED AMOUNTS



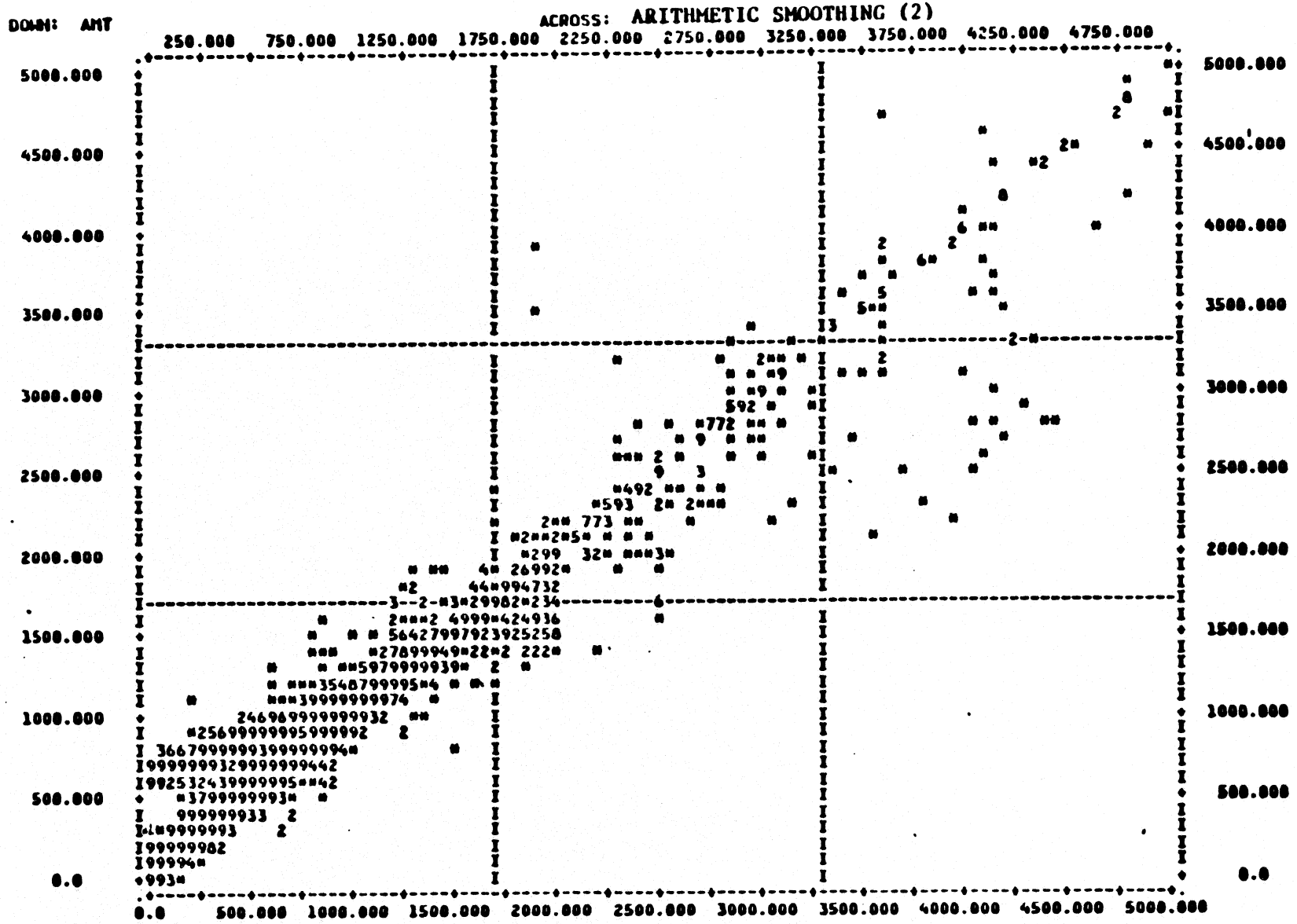
STATISTICS..

CORRELATION (R)-	.94273	R SQUARED -	.88874	SIGNIFICANCE -	.00000
STD ERR OF EST -	254.88906	INTERCEPT (A) -	63.88467	SLOPE (B) -	.95477
PLOTTED VALUES -	5957	EXCLUDED VALUES-	42	MISSING VALUES -	0

'*********' IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

FIGURE 17

REPORTED AMOUNTS BY IMPUTED AMOUNTS



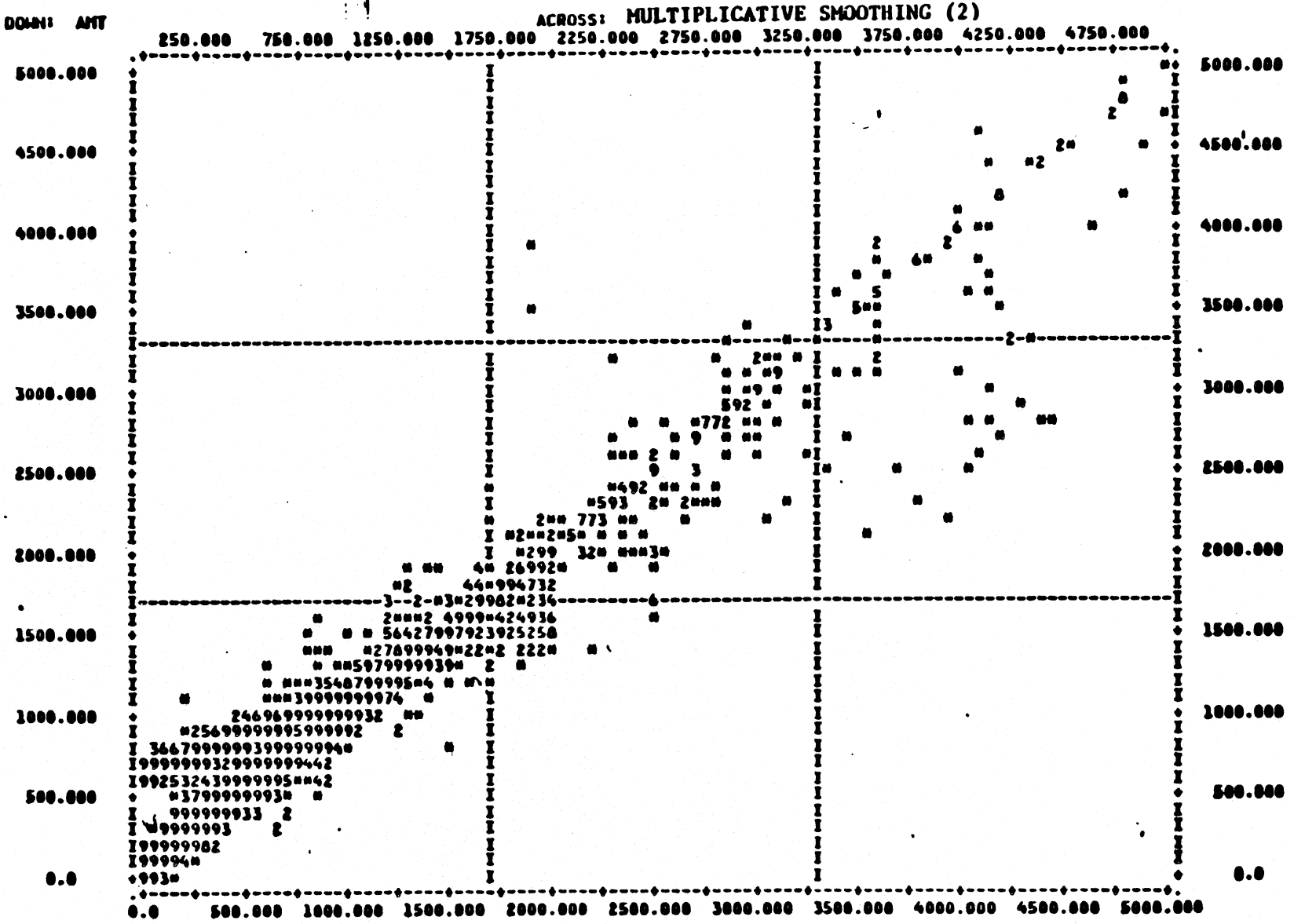
STATISTICS..

CORRELATION (R)-	.96460	R SQUARED -	.93046	SIGNIFICANCE -	.00000
STD ERR OF EST -	184.66221	INTERCEPT (A) -	155.05095	SLOPE (B) -	.89032
PLOTTED VALUES -	5955	EXCLUDED VALUES-	45	MISSING VALUES -	0

'**********' IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

FIGURE 18

REPORTED AMOUNTS BY IMPUTED AMOUNTS



STATISTICS..

CORRELATION (R)-	.96460	R SQUARED -	.93046	SIGNIFICANCE -	.00000
STD ERR OF EST -	184.66222	INTERCEPT (A) -	155.85096	SLOPE (B) -	.89032
PLOTTED VALUES -	5955	EXCLUDED VALUES-	45	MISSING VALUES -	0

'**********' IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

FIGURE 19

HISTOGRAM OF SCALED DIFFERENCES

ITERATED BUCK

Count	Midpoint	
10	-1.725	\\\
12	-1.675	\\\
13	-1.625	\\\
10	-1.575	\\\
21	-1.525	\\\\\
12	-1.475	\\\
10	-1.425	\\\
19	-1.375	\\\\\
18	-1.325	\\\\\
12	-1.275	\\\
14	-1.225	\\\\
14	-1.175	\\\\
21	-1.125	\\\\\
14	-1.075	\\\\
29	-1.025	\\\\\\\
33	-.975	\\\\\\\\
15	-.925	\\\\
32	-.875	\\\\\\\\
28	-.825	\\\\\\\
26	-.775	\\\\\\\
43	-.725	\\\\\\\\\\\
45	-.675	\\\\\\\\\\\
46	-.625	\\\\\\\\\\\\
49	-.575	\\\\\\\\\\\\
66	-.525	\\\\\\\\\\\\\\\\\
55	-.475	\\\\\\\\\\\\\
61	-.425	\\\\\\\\\\\\\\
62	-.375	\\\\\\\\\\\\\\
84	-.325	\\\\\\\\\\\\\\\\\\\\\
81	-.275	\\\\\\\\\\\\\\\\\\\
103	-.225	\\\\\\\\\\\\\\\\\\\\\\\\\
125	-.175	\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\
170	-.125	\\\
192	-.075	\\
182	-.025	\\\
194	.025	\\
174	.075	\\\
141	.125	\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\
112	.175	\\\\\\\\\\\\\\\\\\\\\\\\\\\
103	.225	\\\\\\\\\\\\\\\\\\\\\\\\\
55	.275	\\\\\\\\\\\\\
59	.325	\\\\\\\\\\\\\\
48	.375	\\\\\\\\\\\\
26	.425	\\\\\\\
19	.475	\\\\\
22	.525	\\\\\\
23	.575	\\\\\\
13	.625	\\\
15	.675	\\\\
11	.725	\\\
6	.775	\\
10	.825	\\\
7	.875	\\
5	.925	\
1	.975	

\ = AMOUNT-IMPUTE / AMOUNT

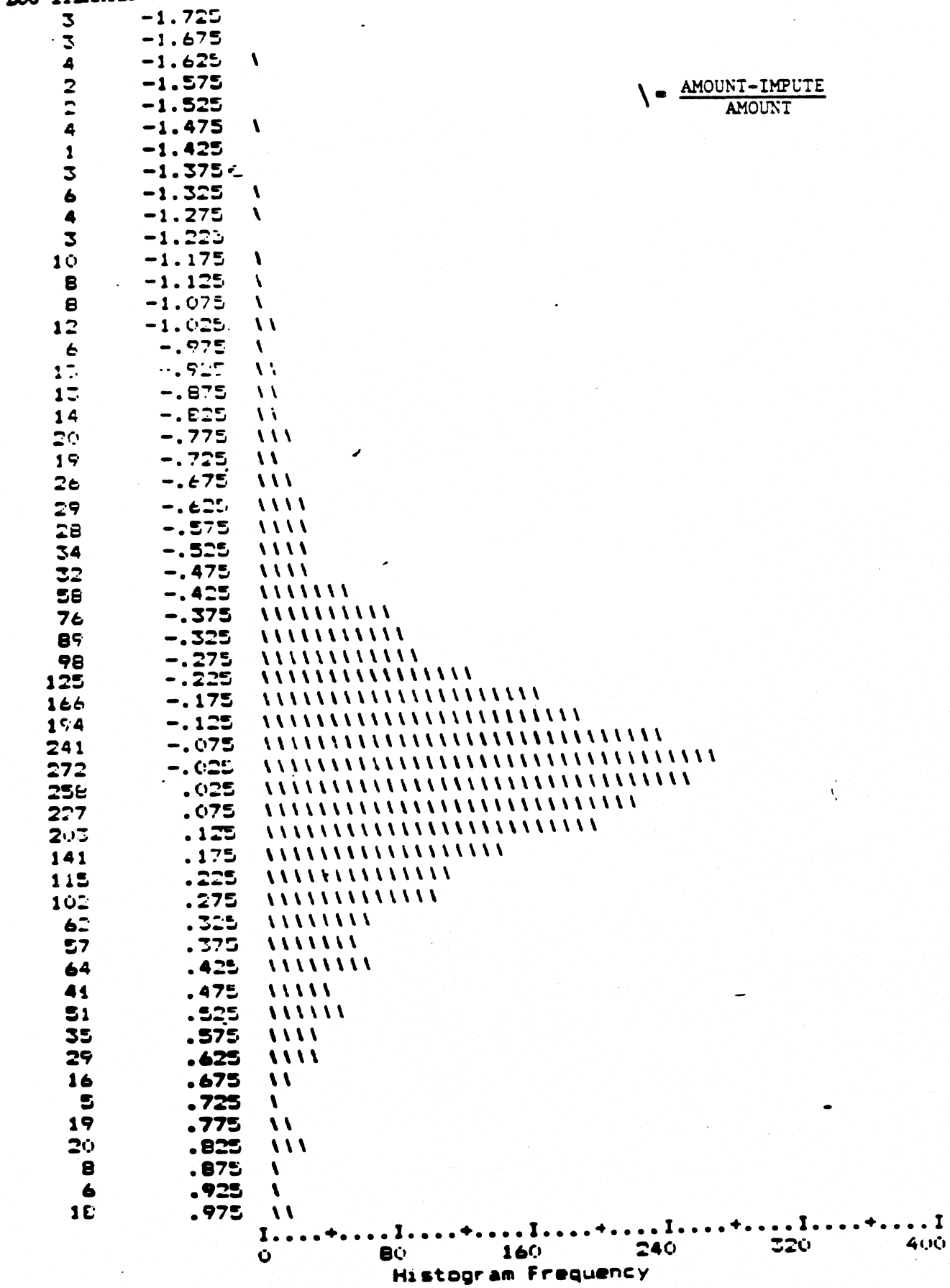
I....+....I....+....I....+....I....+....I....+....I

0 40 80 120 160 200

Histogram Frequency

FIGURE 20

HISTOGRAM OF SCALED DIFFERENCES



Missing Cases 0

FIGURE 21

CUBE ITERATED BUCK — HISTOGRAM OF SCALED DIFFERENCES

Count	Midpoint	
0	-1.7750	
5	-1.7250	\
0	-1.6750	
6	-1.6250	\
1	-1.5750	
0	-1.5250	
5	-1.4750	\
8	-1.4250	\\
2	-1.3750	
7	-1.3250	\
7	-1.2750	\
9	-1.2250	\\
6	-1.1750	\
9	-1.1250	\\
15	-1.0750	\\\
13	-1.0250	\\\
9	-.9750	\\
11	-.9250	\\
15	-.8750	\\\
20	-.8250	\\\\
12	-.7750	\\
11	-.7250	\\
22	-.6750	\\\\
21	-.6250	\\\\
32	-.5750	\\\\\\
35	-.5250	\\\\\\\
48	-.4750	\\\\\\\\\\
51	-.4250	\\\\\\\\\\
49	-.3750	\\\\\\\\\\
73	-.3250	\\\\\\\\\\\\\\\
91	-.2750	\\\\\\\\\\\\\\\\\\
116	-.2250	\\\\\\\\\\\\\\\\\\\\\\\
174	-.1750	\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\
168	-.1250	\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\
187	-.0750	\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\
234	-.0250	\\
249	.0250	\\
250	.0750	\\
219	.1250	\\
182	.1750	\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\
126	.2250	\\\\\\\\\\\\\\\\\\\\\\\\\
121	.2750	\\\\\\\\\\\\\\\\\\\\\\\\
95	.3250	\\\\\\\\\\\\\\\\\\\
60	.3750	\\\\\\\\\\\\
60	.4250	\\\\\\\\\\\\
66	.4750	\\\\\\\\\\\\\\
48	.5250	\\\\\\\\\\
24	.5750	\\\\\
25	.6250	\\\\\
12	.6750	\\
16	.7250	\\\
25	.7750	\\\\\
20	.8250	\\\\
4	.8750	\
14	.9250	\\\
8	.9750	\\

\= AMOUNT-IMPUTE / AMOUNT

I....+....I....+....I....+....I....+....I....+....I

0 50 100 150 200 250

Histogram Frequency

alid Cases 3183 Missing Cases 0

FIGURE 22

HISTOGRAM OF SCALED DIFFERENCES

ARITHMETIC SMOOTHING (1)

\ = (AMOUNT-IMPUTE) / AMOUNT

Count	Midpoint
0	-1.725
0	-1.675
0	-1.625
1	-1.575
4	-1.525
4	-1.475
2	-1.425
7	-1.375
3	-1.325
4	-1.275
1	-1.225
2	-1.175
5	-1.125
8	-1.075
11	-1.025
15	-.975
10	-.925
10	-.875
9	-.825
12	-.775
9	-.725
14	-.675
13	-.625
25	-.575
25	-.525
35	-.475
32	-.425
47	-.375
61	-.325
111	-.275
102	-.225
115	-.175
154	-.125
181	-.075
427	-.025
234	.025
223	.075
159	.125
227	.175
150	.225
170	.275
56	.325
83	.375
70	.425
56	.475
46	.525
32	.575
25	.625
26	.675
19	.725
16	.775
23	.825
16	.875
14	.925
14	.975

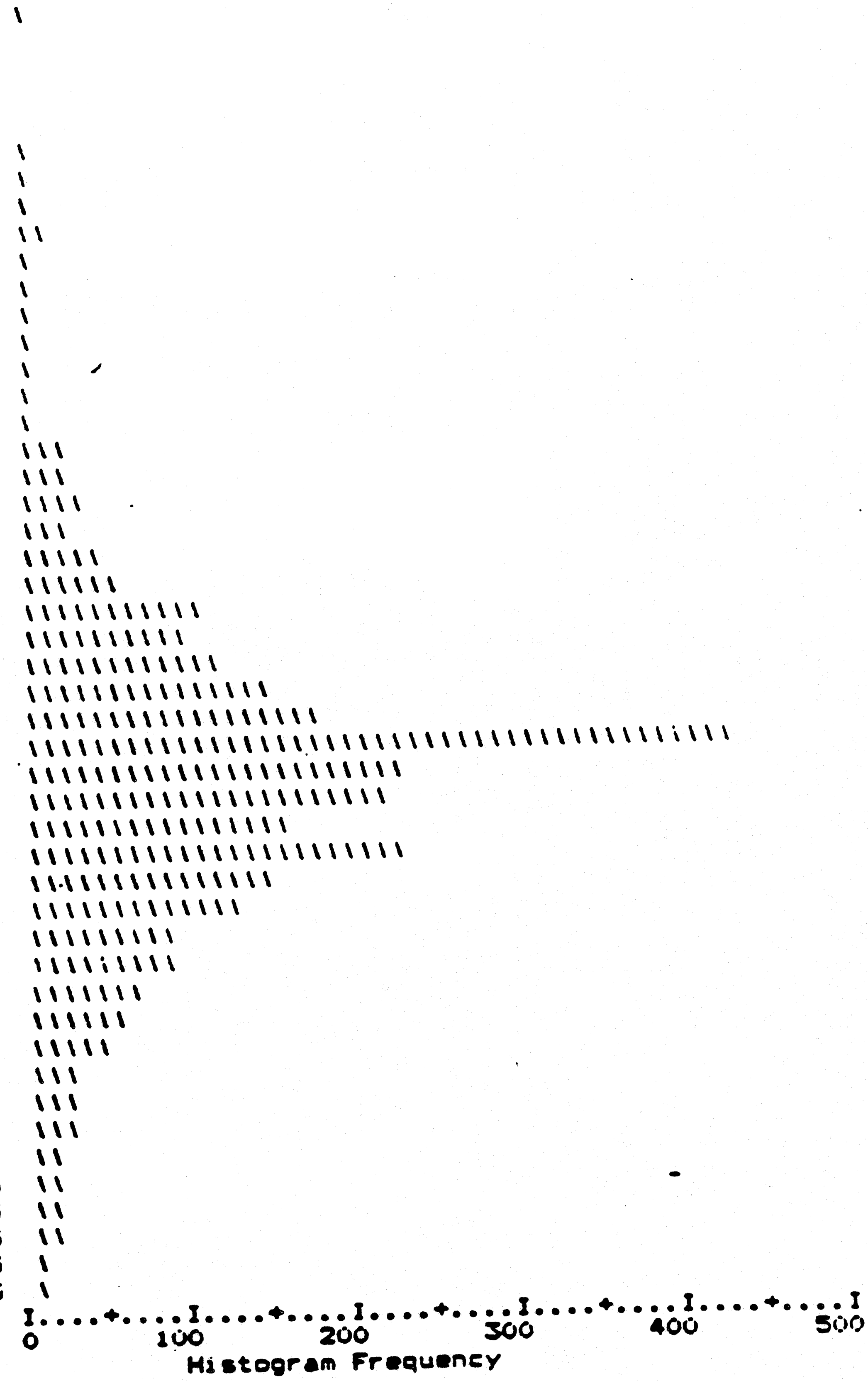


FIGURE 23

HISTOGRAM OF SCALED DIFFERENCES

MULTIPLICATIVE SMOOTHING (1)

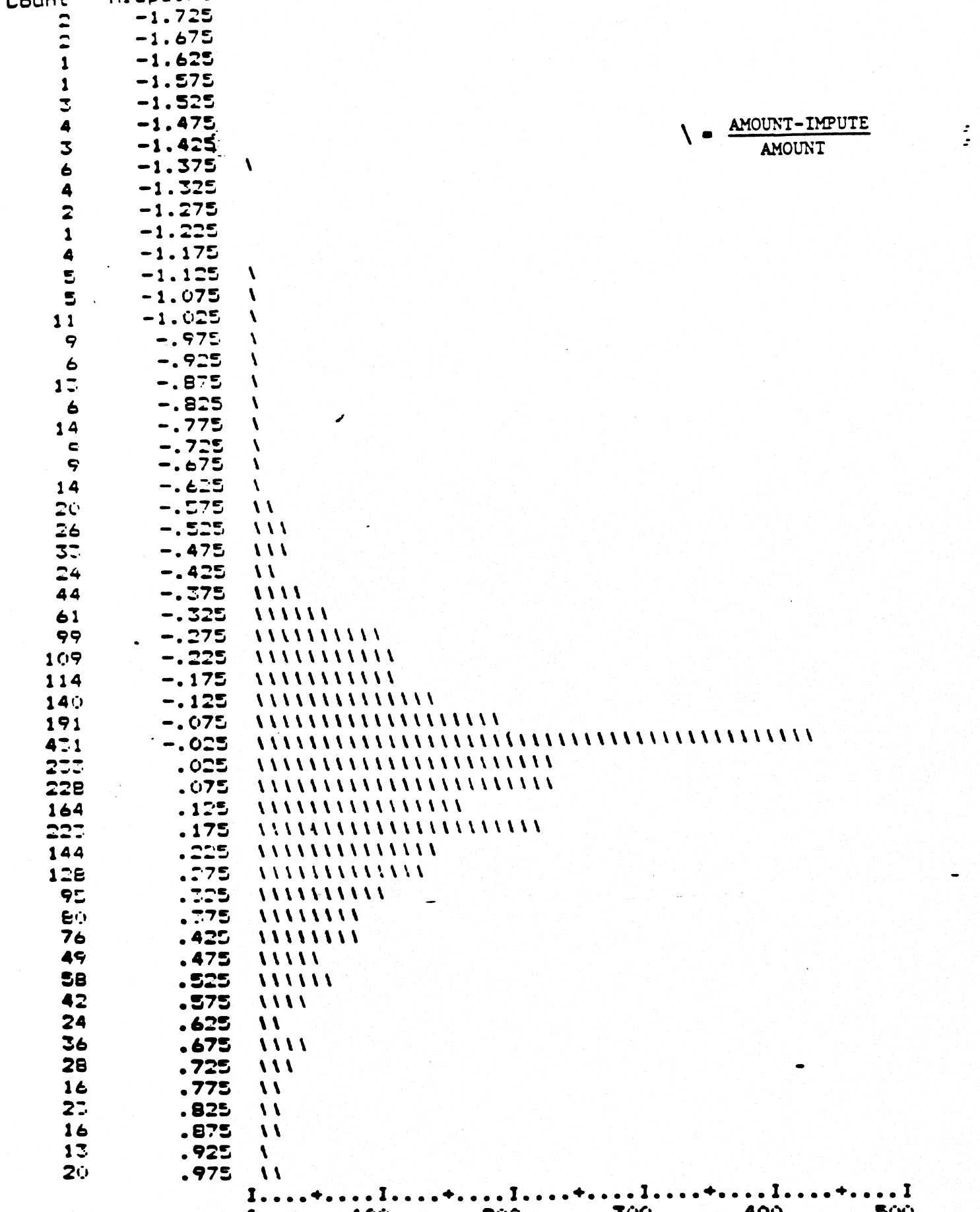


FIGURE 24

HISTOGRAM OF SCALED DIFFERENCES

ARITHMETIC SMOOTHING (2)

\ = AMOUNT-IMPUTE / AMOUNT

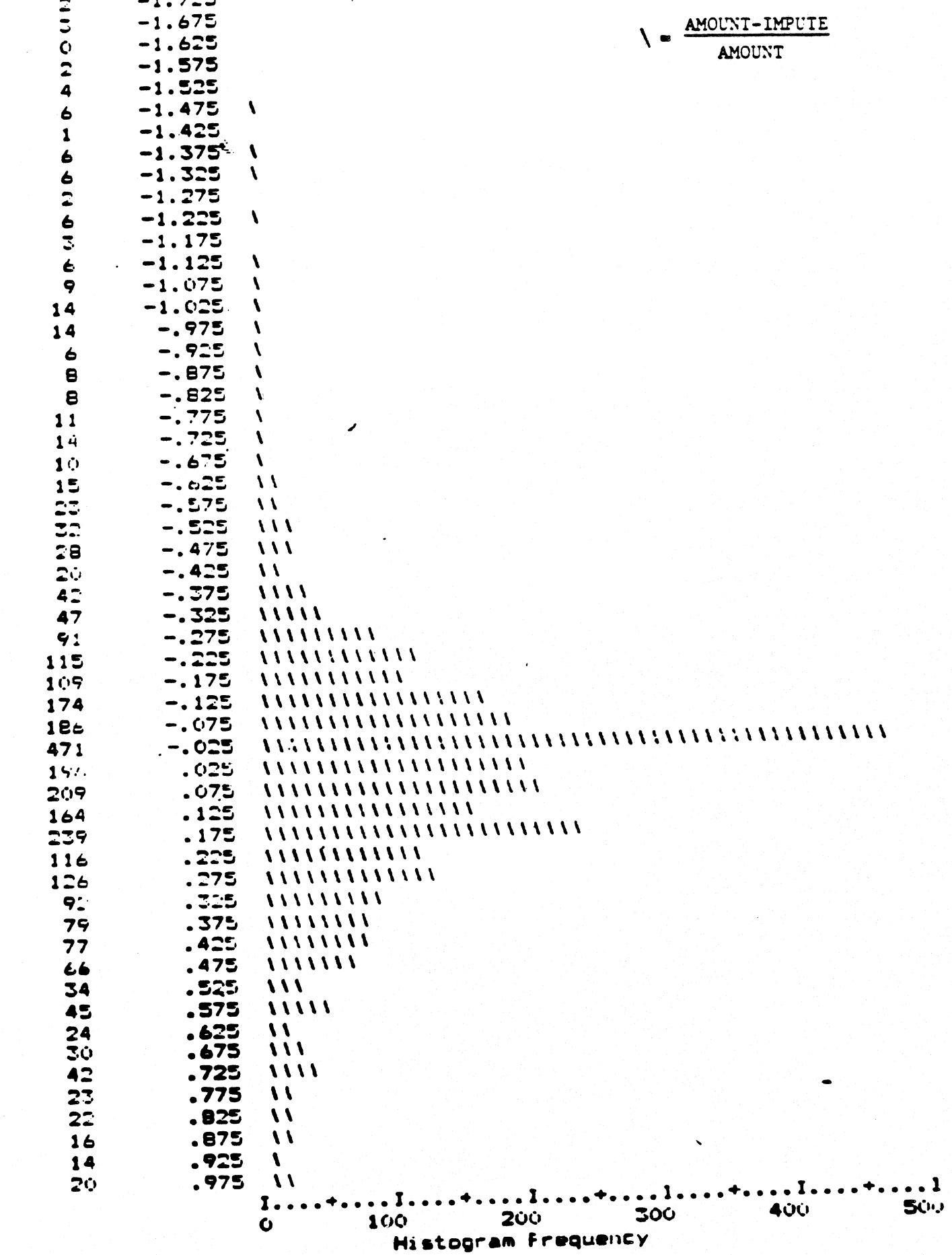


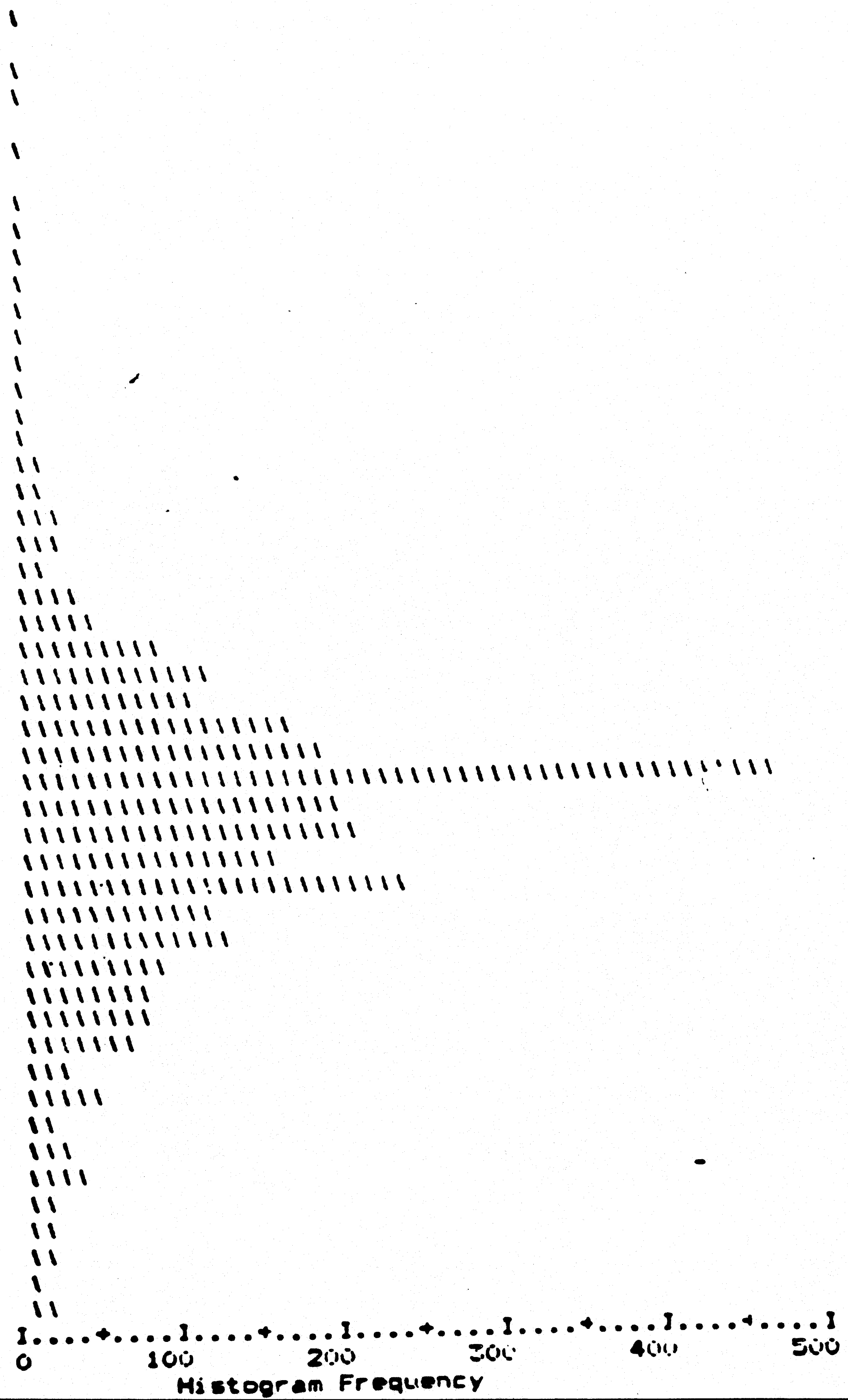
FIGURE 25

HISTOGRAM OF SCALED DIFFERENCES

MULTIPLICATIVE SMOOTHING (2)

\ = AMOUNT-IMPUTE / AMOUNT

Count	Midpoint
2	-1.725
3	-1.675
0	-1.625
2	-1.575
4	-1.525
6	-1.475
1	-1.425
6	-1.375
6	-1.325
2	-1.275
6	-1.225
3	-1.175
6	-1.125
9	-1.075
14	-1.025
14	-.975
6	-.925
8	-.875
8	-.825
11	-.775
14	-.725
10	-.675
15	-.625
23	-.575
32	-.525
28	-.475
20	-.425
42	-.375
47	-.325
91	-.275
115	-.225
109	-.175
174	-.125
186	-.075
471	-.025
196	.025
209	.075
164	.125
239	.175
116	.225
126	.275
92	.325
79	.375
77	.425
66	.475
34	.525
45	.575
24	.625
30	.675
42	.725
23	.775
22	.825
16	.875
14	.925
20	.975



Sum, Sum of Squares, Mean, and Variance of Residuals

$c_{ij} = x_{ij} - \hat{x}_{ij}$	Iterated Buck	Logarithmic Iterated Buck	Cube Iterated Buck	Arithmetic Smoothing (1)	Multiplicative Smoothing (1)	Arithmetic Smoothing (2)	Multiplicative Smoothing (2)
$\sum_{i=1}^{N} \sum_{j}^{12} c_{i,j}$	-198,198.8	182,713.8	244,565.5	150,246.2	191,657.9	177,842.6	198,243.3
$\sum_{i=1}^{N} \sum_{j=1}^{12} c_{i,j}^2$	881,607,100	791,591,000	835,393,408	836,990,100	823,631,700	880,732,300	874,453,700
$\frac{\sum_{i=1}^{N} \sum_{j=1}^{12} c_{i,j}}{n}$	-62.268	57.403	76.835	47.203	60.213	55.873	62.282
$\frac{\sum_{1}^{N} \sum_{j=1}^{12} (c_{i,j} - \bar{c})^2}{n}$	273,096.5	245,398.1	258,550.8	260,729.8	255,135	273,577.6	270,848.4

n = 3183 cases

$c_{ij} \neq 0$

TABLE 1

Sum, Sum of Squares, Mean, and Variance of Scaled Residuals

$c_{ij} = \frac{x_{i,j} - \hat{x}_{i,j}}{x}$	Iterated Buck	Logarithmic Iterated Buck	Cube Iterated Buck	Arithmetic Smoothing (1)	Multiplicative Smoothing (1)	Arithmetic Smoothing (2)	Multiplicative Smoothing (2)
$\sum_{i=1}^{N} \sum_{j=1}^{12} c_{i,j}$	-5792.08	-1001.723	-1120.027	-734.073	-661.089	-699.351	-661.261
$\sum_{i=1}^{N} \sum_{j=1}^{12} c_{i,j}^2$	3,594,429	952,807.2	58973.7	1,024,158	1,007,097	1,022,134	1,013,160
$\frac{\sum_{i=1}^{N} \sum_{j=1}^{12} c_{i,j}}{n}$	-1.820	-.315	-.352	-.231	-.208	-.220	-.208
$\frac{\sum_{i}^{N} \sum_{j=1}^{12} (c_{i,j} - \bar{c})^2}{n}$	148.434	16.454	18.404	16.904	16.827	17.016	16.985

n = 3183 cases

$c_{ij} \neq 0$

TABLE 2

Sum, Sum of Squares, Mean, and Variance of Difference in Adjacent Ratios

$c_{i,j} = r_{ij} - \hat{r}_{ij}$	Iterated Buck	Logarithmic Iterated Buck	Cube Iterated Buck	Arithmetic Smoothing (1)	Multiplicative Smoothing (1)	Arithmetic Smoothing (2)	Multiplicative Smoothing (2)
$\sum_{i}^{N} \sum_{j=1}^{12} c_{i,j}$	-2,252.849	-1040.282	-957.756	-1482.371	-1514.899	-1512.631	-1522.117
$\sum_{i}^{N} \sum_{j=1}^{12} c_{i,j}^2$	4,071,898	122,167.8	119,478.695	202,261.6	221,854.2	213,856.469	226,826.703
$\frac{\sum_{i}^{N} \sum_{j=1}^{12} c_{i,j}}{n}$	-.799	-.369	-.340	-.526	-.537	-.536	-.540
$\frac{\sum_{i}^{N} \sum_{j=1}^{12} (c_{i,j} - \bar{c})^2}{n}$	1,443.303	43.186	42.253	71.448	78.383	75.547	80.144

$n = 2820$

$c_{ij} \neq 0$

TABLE 3

Sum, Sum of Squares, Mean, and Variance of Difference in Scaled Adjacent Ratios

$c_{i,j} = \frac{r_{i,j} - \hat{r}_{i,j}}{r_{i,j}}$	Iterated Buck	Logarithmic Iterated Buck	Cube Iterated Buck	Arithmetic Smoothing (1)	Multiplicative Smoothing (1)	Arithmetic Smoothing (2)	Multiplicative Smoothing (2)
$\sum_{i=1}^{N}\sum_{j=1}^{11} c_{i,j}$	-2960.78	-1756.161	-1679.103	-2258.004	-2,305.211	-2295.991	-2320.885
$\sum_{i=1}^{N}\sum_{j=1}^{11} c_{i,j}^2$	4,074,304	123,128.9	121,931.687	215,218.4	232,074.2	227,100.922	237,474.719
$\frac{\sum_{i=1}^{N}\sum_{j=1}^{11} c_{i,j}}{n}$	-1.050	-.623	-.595	-.801	-.817	-.814	-.823
$\frac{\sum_{i=1}^{N}\sum_{j=1}^{11} (c_{i,j}-\bar{c})^2}{n}$	1443.693	43.275	42.884	75.677	81.628	79.868	83.533

$n = 2820$

$c_{ij} \neq 0$

TABLE 4