Netherlands school of economics

ECONOMETRIC INSTITUTE

ESTIMATING QUARTERLY MODELS WITH PARTLY MISSING QUARTERLY OBSERVATIONS

G.M.M. GELAUFF and R. HARKEMA

Erasmus

REPORT 7724/E

ERASMUS UNIVERSITY ROTTERDAM, P.O. BOX 1738, ROTTERDAM, THE NETHERLANDS

ESTIMATING QUARTERLY MODELS WITH PARTLY MISSING QUARTERLY OBSERVATIONS

by G.M.M. Gelauff and R. Harkema*

ABSTRACT

In this paper Monte Carlo simulation is used in order to compare the performance of five different methods to estimate quarterly models with partly missing quarterly observations. The methods are compared on the basis of the parameter estimates they produce. The first three methods solve the estimation problem in two steps: first the yearly series is disaggregated into a quarterly one and then the quarterly model is estimated. The fourth method considers disaggregation of the yearly series within the context of the model to be estimated and arrives simultaneously at estimates of the missing quarterly observations and of the parameters of the model. The last method simply consists of maximum likelihood estimation of the yearly model. The conclusions from this simulation study are twofold: (i) none of the methods that are developed for the purpose of estimating quarterly models with partly missing observations performs significantly better than maximum likelihood estimation of the yearly model; (ii) the standard errors that result from application of the first three methods are deceptive.

Contents

* A preliminary version of this paper has been presented at the European Meeting of the Econometric Society at Vienna, September 6-9, 1977.

1. INTRODUCTION

After being specified an econometric model requires data about the included variables to allow of estimation and decision making or prediction. It may happen that for some variables data are incomplete. Statistical literature has provided various ways of dealing with the then arising missing data problems. A subset of these problems consists of cases in which data for one of the included variables are only available on a yearly but not on a quarterly basis.

Roughly speaking, the methods that have been developed to treat this estimation problem can be split up into two categories, viz., formal and causal methods. The former category solves the estimation problem in two steps: first the yearly series is disaggregated into a quarterly one and then the quarterly model is estimated. The latter category, which is relatively ill-represented, considers disaggregation of the yearly series within the context of the model to be estimated and arrives simultaneously at estimates of the missing quarterly observations and of the parameters of the model.

As econometric estimation is often employed with the aim to arrive at decision rules or predictions it is interesting to compare the various methods on the basis of the parameter estimates they produce. Unfortunately such a comparison cannot easily be performed by means of analytical devices. Therefore, the main purpose of the present paper is to investigate the sampling characteristics of the various estimators by using Monte Carlo methods.

The plan of the paper is as follows. In Section 2 a review is presented of both categories of methods. In Section 3 we estimate a quarterly consumption function with missing quarterly data for personal income and present the results obtained by applying five different methods. In Section 4 we perform a Monte Carlo experiment in order to examine the small sample characteristics of these methods. Finally, Section 5 contains some concluding comments.

2. REVIEW OF SOME DISAGGREGATION TECHNIQUES

A simple way of deriving quarterly values from annual totals satisfying some reasonable criteria has been brought forward by Lisman and Sandee (1964), but has the disadvantage that the first and last yearly totals cannot be disaggregated and that the criteria cannot escape a sense of arbitrariness. Boot, Feibes and Lisman (1967) therefore suggest to use a minimization criterion, which consists of the sum of squares of the first or second differences between the successive quarterly values. During the minimization the restriction that the sum of the quarterly figures in a year equals the yearly total is taken into account. The method of Boot, Feibes, and Lisman has been generalised to disaggregating a set of values for n periods into a set of np subperiod values by Cohen, Müller, and Padberg (1971). These authors infer from matrix inversion considerations that when the total number of period values exceeds $n = 13$ it would be desirable to use overlapping sets of periods, each of length $n \leq 13$ for obtaining subperiod values. This should also be borne in mind when the technique of Boot, Feibes, and Lisman is employed.

The approach of Van de Aker and Van Reeken (1971) consists of fitting a cubic spline through the cumulative series of yearly totals. Then the quarterly figures are computed by interpolating the spline and subtracting adjacent interpolation results. Van de Aker and Van Reeken show that their technique and the first difference method of Boot, Feibes, and Lisman are closely related. Their numerical experiments exhibit no extreme deviations between the spline interpolation and the method of Boot, Feibes, and Lisman when the computed quarterly figures are compared.

The only thing the above mentioned methods require is a set of annual totals. Other, more promising possibilities arise when related series are available. In its most simple form this can be illustrated as follows. Suppose two series x^* and y^* of annual totals are given. The corresponding quarterly series x is also known, but the quarterly series y is missing. Then ordinary least squares estimation of a_1 and a_0 in the linear model

$$y_i^* = a_0 + a_1 x_i^* + v_i \qquad (i = 1, \ldots, n) \tag{2.1}$$

yields estimates $\hat{a}_0$ and $\hat{a}_1$ which may be used to compute a series of quarterly figures for y by means of

$$y_j = \hat{a}_0 + \hat{a}_1 x_j \qquad (j = 1, \ldots, 4n) \tag{2.2}$$

A disadvantage of this approach is that generally the quarterly figures resulting from this procedure will not add up to the yearly totals. Ginsburgh (1973) suggests to combine the Boot, Feibes, and Lisman approach with the related series approach. He first disaggregates x^* and y^* by the Boot, Feibes, and Lisman method giving $\hat{y}$ and $\hat{x}$ and thereafter computes the final series of quarterly figures y according to

$$(2.3) \qquad y_j = \hat{y}_j + \hat{a}_1(x_j - \hat{x}_j) \qquad (j = 1, \ldots, 4n)$$

Ginsburgh shows this procedure to be tantamount to minimizing

$$(2.4) \qquad \sum_{j=2}^{4n} (\Delta y_j - \hat{a}_1 \Delta x_j)^2$$

subject to the constraint that the quarterly figures add up to the yearly totals, the so called consistency requirement. According to the numerical experiments of Ginsburgh this modified related series approach produces quarterly figures that are closer to the actual ones than those resulting from the Boot, Feibes, Lisman approach.

Chow and Lin (1971) start from q related series $x_1, \ldots, x_q$ and suppose the quarterly relation between y and x_i $(i = 1, \ldots, q)$ to follow a linear model

$$(2.5) \qquad y = X\beta + u$$

where y denotes a vector of length 4n, X a matrix of order $4n \times q$, and u a vector of disturbances of length 4n with mean zero and variance-covariance matrix V. To estimate the series of quarterly figures for y they use the best linear unbiased estimator Ay^* where A is of order $4n \times n$ and y^* denotes the series of yearly totals. Recently, Chow and Lin (1976) applied their method to a problem where quarterly values for y are known for the last part of the series but only yearly figures for the first part.[1] They prove that in this case too the covariance matrix of their estimator is not larger than that of any other unbiased estimator which is a linear combination of the known quarterly and yearly observations on y.

All so far treated methods belong to the first category characterized in the Introduction. They give techniques for constructing quarterly figures irrespective of the appearance of the variable concerned in an econometric model and may be called formal methods. The other category

[1] In fact Chow and Lin discuss a problem with quarterly data given and monthly data missing but this makes no fundamental difference. To maintain a uniform description of the methods there will be spoken of missing quarterly data.

estimates quarterly figures within the context of an econometric model using the causal relationships between the variables in that model. Therefore they may be called causal methods. The first of these has been developed by Sargan and Drettakis and is described in Drettakis (1973) and Sargan and Drettakis (1974). The authors consider a linear system of simultaneous equations and compute maximum likelihood estimates of the missing data as well as the parameters of the model. It is assumed that quarterly data are missing for a subset of the endogenous variables and in a part of the series. An advantage of this method is its allowance for simultaneity.

Another approach in this category developed by Somermeyer, Jansen and Louter (1976), considers the linear model

$$(2.6)\qquad y_{jk} = \sum_{h=1}^{H} \beta_h x_{hjk} + \beta_I x_{Ijk} + u_{jk} \qquad (j = 1, \ldots, J;\ k = 1, 2, 3, 4)$$

where y_{jk} denotes the value of the variable to be explained, x_{hjk} denotes the value of the h-th explanatory variable, x_{Ijk} denotes the (missing) value of the I-th explanatory variable and u_{jk} denotes a disturbance term, all in quarter k of year j. Somermeyer _et al_. assume the unknown quantities x_{Ijk} to be a weighted average of the yearly observations on x_I in the same year j, the preceding year j-1 and the succeeding year j+1, with weights depending on both the quarter k and the time shift τ ($\tau = -1, 0, 1$). Hence

$$(2.7)\qquad x_{Ijk} = \mu_j \sum_{\tau=-1}^{+1} \alpha_{k\tau} x_{Ij+\tau} \qquad (j = 1, \ldots, J;\ k = 1, 2, 3, 4)$$

with $\alpha_{k\tau}$ representing fixed, but _a priori_ unknown weights satisfying

$$(2.8)\qquad \sum_{k=1}^{4} \sum_{\tau=-1}^{1} \alpha_{k\tau} = 1$$

and

$$\mu_j = \frac{x_{Ij}}{\sum_{k'=1}^{4} \sum_{\tau=-1}^{1} \alpha_{k'\tau} x_{Ij+\tau}} \qquad (j = 1, \ldots, J)$$

The correction factors μ_j ensure the consistency restraint to be satisfied. After substituting (2.7) into (2.6) and specifying the stochastic properties of the disturbance terms u_{jk}, Somermeyer _et al_. estimate the parameters, weights, and missing quarterly figures by means of an iterative generalized least squares procedure. They conclude from numerical experiments that their method performs somewhat better than the Boot, Feibes, and Lisman method. In addition they conjecture that their method is also able to handle the case where no quarterly data are available for the variable to be explained.

A third approach that can be thought of is maximum likelihood estimation of the quarterly model treating the missing observations as normally distributed random variables with unknown means. This approach ia similar to the one proposed by Drettakis (1973) apart from the fact that now all quarterly observations on some explanatory variable are supposed to be missing and not only a subset. Attempts to proceed along this track did not prove to be very succesful and therefore this approach will not be discussed here. More details can be found in Gelauff and Harkema (1977).

Before ending up this Section it should be pointed out that most of the formal methods suffer from the drawback that it is difficult, if not impossible, to assess analytically the statistical properties of the parameter estimates they produce. The problems are similar to those encountered in the statistical analysis of models with errors in variables on the understanding that in the present case the measurement errors are linear combinations of the true missing data themselves. Therefore the traditional assumption saying that the measurement errors are asymptotically uncorrelated with the true values will in general not be tenable in the present case. Nevertheless, from the wellknown results of statistical analysis of models with errors in variables it may be conjectured that the formal methods yield parameter estimates that are inconsistent. As a matter of fact there is only one formal method which we can prove to yield consistent parameter estimates, i.e., the very simple one that assigns to each quarterly figure one quarter of the corresponding yearly total. Therefore we have also taken along this method in the simulation experiments that have been performed in Section 4.

3. ESTIMATION OF A QUARTERLY CONSUMPTION FUNCTION

In this section we present some empirical results obtained by application of five different methods. As an illustrative example we estimate a quarterly specified consumption function which is taken from Somermeyer, Jansen and Louter (1976) and reads as follows

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t \qquad (3.1)$$

with y_t = consumption;

x_{1t} = personal wealth, approximated by deposits;

x_{2t} = product of the rate of interest in deviation of its mean over the period of observation, and income;

x_{3t} = income itself;

u_t = disturbance.

Somermeyer et al. estimate (3.1) using United States data, assuming the actually known quarterly income figures to be missing, and Dutch ones for which no quarterly income figures exist. A problem arises because of the presence of the variable x_2 for it depends on the unknown quarterly income series. In order to cope with this problem Somermeyer et al. employ an iterative procedure. For reasons of simplicity and because at least for U.S. data the influence of the variable x_2 is extremely small, this variable has been deleted from the computations carried out with U.S. data. Using a somewhat different notation which will prove to be useful in the next section the following model arises

$$y_{it} = \alpha + \gamma z_{it} + \beta x_{it} + v_{it} \tag{3.2}$$

where y_{it} denotes consumption, z_{it} denotes personal wealth and x_{it} denotes income, all in quarter i of year t.

The methods which have been used to estimate (3.2) are: (i) the very simple method that assigns to each quarterly figure one quarter of the corresponding yearly total (I); (ii) the Boot-Feibes-Lisman approach (II); (iii) the Van de Aker-Van Reeken approach (III); (iv) the Somermeyer-Jansen-Louter approach (IV); (v) maximum likelihood estimation of the yearly model (V).

As regards methods I, II, and III maximum likelihood estimation has been carried out assuming the quarterly disturbances to be normally distributed with zero means and variance-covariance matrix $P \otimes \sigma^2 I_4$, where the symbol $\otimes$ denotes the operation of forming the Kronecker product, P denotes the wellknown matrix

$$P = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \vdots & & & & \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \dots & 1 \end{bmatrix}$$

and I_4 denotes the identity matrix of order 4. The results of method IV have been taken from Somermeyer et al. and method V has been carried out assuming the yearly disturbances to be normally distributed with zero means and variance-covariance matrix $4\sigma^2 P$. Numerical optimizations were performed using the Modified Fibonacci approach as described in e.g. Beveridge and Schechter (1970).

The resulting estimates are presented in Table 1 together with the associated asymptotic standard errors which were computed by means of the Hessian of the loglikelihood function evaluated at its optimum.[2]

Table 1. Estimates of the Parameters of a Macro Economic Consumption Function with Partly Missing Observations Using Five Alternative Estimation Methods*

Method	Parameter				
	α	β	γ	ρ	σ
I	29.20 (20.87)	0.865 (0.043)	-0.302 (1.588)	0.811	8.889
II	30.32 (5.13)	0.874 (0.010)	-0.528 (0.455)	-0.091	3.605
III	30.36 (5.13)	0.874 (0.010)	-0.532 (0.455)	-0.085	3.597
IV**	28.0 (6.8)	0.856 (0.011)	0.00 (0.5)	-0.257	-
V	114.07 (31.98)	0.872 (0.016)	-0.391 (0.717)	-0.176	11.046

* The figures between parentheses denote the asymptotic standard errors.

** These results are adjusted in comparison with Table 1 of Somermeyer et al. In Somermeyer et al. (1974) the basic data (running from 1959-I up to and including 1971-IV) are presented to which the methods have been applied. Consulting the authors it appeared that for computational reasons Somermeyer et al. had divided the quarterly consumption and income series by hundred and the deposits series by ten but had not readjusted their parameter estimates. Therefore $\hat{\beta}_0$ is multiplied by one hundred and $\hat{\beta}_1$ by ten in Table 1.

A comparison of the estimates of α and β in Table 1 shows that they are fairly similar. The estimates of α are almost identical when it is borne in mind that aggregation from quarterly to annual data results in

[2] See e.g. Goldfeld and Quandt (1972).

a value that is four times as large so that the estimate obtained by means of method V has to be divided by four in order to be comparable with the results of the other methods. Method I yields estimates of ρ and σ that are much more different from zero than those obtained by means of other methods. The large value of $\hat{\sigma}$ is also responsible for the large asymptotic standard errors of method I because these standard errors depend multiplicatively on σ. The approaches of Boot _et_ _al_. (II) and of Van de Aker and Van Reeken (III) yield almost identical estimates and asymptotic standard errors for all parameters. Moreover the standard errors computed by means of these methods are smaller than the ones obtained by the other approaches. The estimate of β obtained by Somermeyer _et_ _al_. deviates slightly from the estimates obtained by means of the other methods which may well be a result of not using the x_2 series in the other approaches. The estimate of γ differs somewhat more although not significantly for it can be seen from the standard errors that in all other cases too $\hat{\gamma}$ is not significantly different from zero. Remembering that aggregation from quarterly to annual data results in values of α and σ^2 that are four times as large, it appears that method V yields parameter estimates that are not too different from those obtained by the other approaches but asymptotic standard errors which are considerably larger.

Judging on the basis of the precision of the estimates methods II and III seem to perform best. From the discussion at the end of Section 2, however, it will be clear that the results obtained by using formal methods like methods II and III do not have a very clearcut statistical interpretation. Moreover, the preceding comparison is only based on asymptotic arguments. In practice, samples may be too small to validate this type of argument. Therefore it is desirable to compare the small sample properties of the methods treated so far. For this purpose Monte Carlo simulation is in order.

4. MONTE CARLO RESULTS

In the preceding section we have asserted that it may be useful to apply Monte Carlo methods in order to get some insight into the small sample properties of the various estimators. To perform such a simulation experiment we proceed as follows. We employ a simple quarterly consumption function, viz., (4.1) $$y_{it} = \alpha + \beta x_{it} + v_{it}$$

where y_{it} denotes quarterly consumption expenditures, x_{it} quarterly income, and v_{it} a disturbance term. First of all the data of the model have to be generated. Evidently, this can be done in various ways. The approach that we will use consists of generating annual income figures according to a stochastic first-order difference equation and disaggregating the annual figures into quarterly ones according to a moving average process. More specifically, annual income figures are generated according to

(4.2) $$j_t = (1 + g)j_{t-1} + \xi_t \qquad t = 2, \ldots, T$$

with

(4.3) $$\xi_t = \rho_1 \xi_{t-1} + \pi_t$$

where g denotes the growth rate of the income series and ρ_1 is an autocorrelation coefficient, both of which are chosen in advance. To start with we draw π_t ($t = 2, \ldots, T$) at random from a normal distribution with zero mean and variance $\sigma_1^2(1+g)^{2t-4}$, σ_1^2 being known in advance. Next, we draw ξ_1 at random from a normal distribution with zero mean and variance $\sigma_1^2[(1+g)^2 - \rho_1^2]^{-1}$. The quantities ξ_t ($t = 2, \ldots, T$) can then be generated through successive application of formula (4.3). As a result of this procedure the standard deviations of the distributions of the ξ_t will grow by the same rate as the expectations of the annual income figures j_t, viz., (1+g). The annual income figures themselves are generated through successive application of formula (4.2) starting from some _a priori_ chosen value for j_1.

In order to disaggregate the annual income figures into quarterly ones we use the following moving average process

(4.4) $$x_{it} = \mu_t(\alpha_{i-1}j_{t-1} + \alpha_{i0}j_t + \alpha_{i\tau}j_{t+1})$$

where $$\mu_t = j_t/(\sum_{i=1}^{4}\sum_{\tau=-1}^{1} \alpha_{i\tau}j_{t+\tau})$$

and $$\sum_{i=1}^{4}\sum_{\tau=-1}^{1} \alpha_{i\tau} = 1$$

The weights $\alpha_{i\tau}$ ($i = 1, \ldots, 4$; $\tau = -1, 0, 1$) are fixed in advance and, for example, taken equal to the values estimated by Somermeyer _et al_. From (4.4) it is easily seen that in order to generate quarterly figures for the periods 2 through T we need two additional annual income figures, viz., for the periods 1 and T+1.

Finally, we impose a yearly autocorrelation scheme on the disturbances v_{it} by postulating

$$v_{it} = \rho_2 v_{i,t-1} + \varepsilon_{it} \qquad (i = 1, \ldots, 4;\ t = 2, \ldots, T) \tag{4.5}$$

As before we draw ε_{it} $(i = 1, \ldots, 4)$ at random from a normal distribution with zero mean and variance $\sigma_2^2(1 + g)^{2t-4}$, σ_2^2 being known in advance. Next, we draw v_{i1} $(i = 1, \ldots, 4)$ at random from a normal distribution with zero mean and variance $\sigma_2^2[(1+g)^2 - \rho_2^2]^{-1}$. The quantities v_{it} $(i = 1, \ldots, 4;\ t = 2, \ldots, T)$ can then be generated through successive application of formula (4.5).

Once the quarterly income figures and disturbances have been generated and α and β have been selected the quarterly consumption figures y_{it} can be computed from (4.1). With the annual income figures at hand and assuming the quarterly income figures missing we are in a position to apply the methods mentioned above. By using the Somermeyer-Jansen-Louter approach we obtain estimates of α and β in a straightforward way; the formal methods may be brought into the experiment by first disaggregating the series of annual income figures into a quarterly one and then estimating α and β from the quarterly model (4.1) by means of a maximum likelihood estimation procedure. The likelihood function that has been used is similar to the one used in Section 3, i.e., based on the assumption that the quarterly disturbances v_{it} are normally distributed with means zero and variance-covariance matrix $P \otimes \sigma^2 I_4$. Therefore numerical optimization is required with respect to the autocorrelation parameter. Finally the yearly specified model can be estimated by aggregating the quarterly consumption figures to annual ones. Once again, as in Section 3, the likelihood function that has been used is based on the assumption that the yearly disturbances are normally distributed with zero means and variance-covariance matrix $4\sigma^2 P$. The small sample distribution of the various estimators can be exhibited by repeating the above mentioned procedure a number of times.

In order to perform the simulation experiment we first have to attach numerical values to the parameters on the basis of which the data are generated. When doing this two things have to be kept in mind, viz., (i) we have to obtain data possessing so much variation that they may be expected to be reasonably different from one set of drawings to the other;

(ii) the data series should as much as possible resemble a real world situation. As regards the yearly income figures we estimated $\hat{\rho}_1$ and $\hat{g}$ from a regression of j_t on j_{t-1}, using the U.S. income series and imposing a first order Markov scheme on the disturbances. Maximum likelihood estimation resulted in a growth rate of 7.2 % and a value for ρ_1 of 0.51. To meet the first point mentioned above we have chosen σ_1 in such a way that $2\sigma_1[(1+g)^2 - \rho_1^2]^{-\frac{1}{2}}$ is equal to 4 % of the first annual income figure giving $\sigma_1 = 23.7$. With a growth rate of 7.2 % this will result in a very small probability of drawing an annual income figure in year t which is smaller than the one in year t-1, which would be incompatible with our second point made above. We have fixed j_1 upon the first figure of the U.S. annual income series ($j_1 = 1271.6$). The weights $\alpha_{i\tau}$ which are necessary to disaggregate the annual income figures according to (4.4) are taken equal to the values estimated in Somermeyer _et al_. (1974) for the U.S. data. As regards generating the disturbances v_{it}, ρ_2 has been fixed upon 0.8 to investigate the effect of the presence of serious autocorrelation and σ_2 has been taken equal to 27 in order to generate fairly fluctuating disturbances. The coefficients α and β have been fixed upon the values corresponding to the estimates of the consumption function using U.S. data. This leads to $\alpha = 28$ and $\beta = 0.85$.

The simulation experiments have been performed for a subset of the methods available because of the large amount of computing time required. From the set of formal methods the related series approaches have been deleted because the results of these approaches depend heavily on the related series used. Generating some kind of related series is too arbitrary and therefore we have selected only those methods[3] which were also applied in Section 3.

For each of these methods we have computed the means of the estimates of the parameters α and β over the 500 replications performed and the sums of squared differences between the true parameter values and the individual estimates $\hat{\alpha}$ and $\hat{\beta}$. The mean estimates and the sums of squared differences are presented in Table 2. The small sample distributions of the estimators for β are exhibited by means of the histograms at the end of this paper.

From Table 2 it appears that the mean estimates are fairly similar

[3] We wish to thank Messrs. Somermeyer, Jansen, and Louter for providing us with the computer programs necessary to apply their method.

Table 2. Mean Estimates and Sums of Squared Differences

Method	Mean Estimates		Sums of Squared Differences	
	α	β	α	β
I	26.68	0.8547	3534	0.0161
II	27.72	0.8527	3516	0.0160
III	27.50	0.8532	3521	0.0160
IV	20.12	0.8643	7779	0.0290
V	101.96	0.8569	57377	0.0161

and almost unbiased for all methods except method IV. As regards the sums of squared differences all methods yield almost identical results apart from the approach of Somermeyer et al.the dispersion of which is about two times larger. From the histograms it can be seen that the differences in the performance of the three formal methods can be neglected. The histogram corresponding to the approach of Somermeyer et al. illustrates the larger dispersion that was already apparent from the computed sum of squared differences. In addition it shows that this larger dispersion is mainly due to a relatively large number of outliers.

The most important conclusion from the results of the simulation procedure is that the methods I, II, III, and V yield almost identical and better estimates than those provided by method IV. As a consequence it does not seem to be very rewarding to apply the methods that have been developed to estimate quarterly models with partly missing quarterly observations. Simply estimating the yearly specified model seems to perform just as well. In addition, it should be concluded that the estimation results shown in Table 1 are obviously misleading since they wrongly indicate methods II and III to be considerably more efficient than estimating the yearly specified model. According to the above experiments and taking into account its simplicity, estimating the yearly model seems most appropriate.

5. CONCLUSION

The methods that have been developed to estimate quarterly models with partly missing quarterly observations can be split up into two categories, viz., formal and causal methods. The former category solves the estimation problem in two steps: first the yearly series is disaggregated into a quarterly one and then the quarterly model is estimated. The latter category, which is relatively ill-represented, considers disaggregation of the yearly series within the context of the model to be estimated and arrives simultaneously at estimates of the missing quarterly observations and of the parameters of the model. In the present paper Monte Carlo methods are used to compare the performance of the various methods.

Because econometric estimation is often employed with the aim to arrive at decision rules or predictions it seems most useful to compare the various methods on the basis of the parameter estimates they produce. Computations with respect to a quarterly consumption function for the U.S.A. lead to the conclusion that two of the formal methods, viz., the approach of Boot et al. and of Van de Aker and Van Reeken yield estimates that are asymptotically more efficient than those produced by other methods. As, however, the results obtained by using formal methods like those of Boot et al. and Van de Aker and Van Reeken do not have a very clearcut statistical interpretation while, moreover, the sample may be too small to validate asymptotic arguments it seems desirable to investigate the small sample characteristics of the various estimators using a Monte Carlo simulation procedure.

The most important conclusion from the results of the simulation procedure is that none of the methods that are developed for the purpose of estimating quarterly models with partly missing quarterly observations performs significantly better than maximum likelihood estimation of the yearly model. The formal methods developed by Boot et al. and Van de Aker and Van Reeken yield standard errors that are deceptive in the sense that they strongly underestimate the true standard errors, while the approach due to Somermeyer et al. results in a dispersion that is about two times as large as the dispersion resulting from maximum likelihood estimation of the yearly model. For the time being it should therefore be concluded that according to the results of the simulation procedure and taking into account its simplicity maximum likelihood estimation of the yearly model seems most appropriate.

It goes without saying that the conclusions stated above cannot straightforwardly be generalized to other models than the one considered in this paper. If, for example, the number of explanatory variables which are measured on a quarterly basis increases (e.g. by including variables that account for seasonal effects), it may be presumed that the quarterly estimation methods will do increasingly better and actually, after passing some critical bound, may perform better than maximum likelihood estimation of the yearly model. In addition, it should be pointed out that the model which has been considered in this paper does not include any lagged values of the explanatory variable whose quarterly figures were supposed to be missing. Clearly, in such cases estimation of the yearly model does not make sense and it would be desirable to know how the quarterly estimation methods perform in this kind of models. These questions, however, are beyond the scope of the present paper and are subject of further research.

REFERENCES

Aker, A.W. van de, and A.J. van Reeken (1971), A Method for k-fold Interpolation of Time-Series, R.C. notitie nr. 11 of the Tilburg School of Economics.

Beveridge, G.S.G., and R.S. Schechter (1970), Optimization; theory and practice, Mc Graw Hill, New York.

Boot, J.C.G., W. Feibes and J.H.C. Lisman (1967), "Further Methods of Derivation of Quarterly Figures from Annual Data", Applied Statistics, Vol. 16, pp. 65-75.

Chow, G.C., and A. Lin (1971), "Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series", The Review of Economics and Statistics, Vol. 53, pp. 372-375.

Chow, G.C., and A. Lin (1976), "Best Linear Unbiased Estimation of Missing Observations in an Economic Time Series", Journal of the American Statistical Association, Vol. 71, pp. 719-721.

Cohen, K.J., W. Müller and M.W. Padberg (1971), "Autoregressive Approaches to Disaggregation of Time Series Data", Applied Statistics, Vol. 20, pp. 119-129.

Drettakis, E.G. (1973), "Missing Data in Econometric Estimation", Review of Economic Studies, Vol. 40, pp. 537-551.

Gelauff, G.M.M., and R. Harkema (1977), Estimating Quarterly Models with Partly Missing Quarterly Observations, unpublished working paper of the Econometric Institute of the Erasmus University Rotterdam.

Ginsburgh, V.A. (1973), "A Further Note on the Derivation of Quarterly Figures Consistent with Annual Data", Applied Statistics, Vol. 22, pp. 368-374.

Goldfeld, S.M., and R.E. Quandt (1972), Nonlinear Methods in Econometrics, North-Holland, Amsterdam.

Lisman, J.H.C., and J. Sandee (1964), "Derivation of Quarterly Figures from Annual Data", Applied Statistics, Vol. 13, pp. 87-90.

Sargan, J.D., and E.G. Drettakis (1974), "Missing Data in an Autoregressive Model", International Economic Review, Vol. 15, pp. 39-58.

Somermeyer, W.H., R. Jansen and A.S. Louter (1974), "Estimating Quarterly Values of Annually Known Variables in Quarterly Relationships", Report 7403 (revised) of the Econometric Institute of the Erasmus University Rotterdam.

Somermeyer, W.H., R. Jansen and A.S. Louter (1976), "Estimating Quarterly Values of Annually Known Variables in Quarterly Relationships", Journal of the American Statistical Association, Vol. 71, pp. 588-595.

FIGURE 1. FREQUENCY DISTRIBUTION OF THE ESTIMATES OF BETA
OBTAINED ACCORDING TO METHOD I.

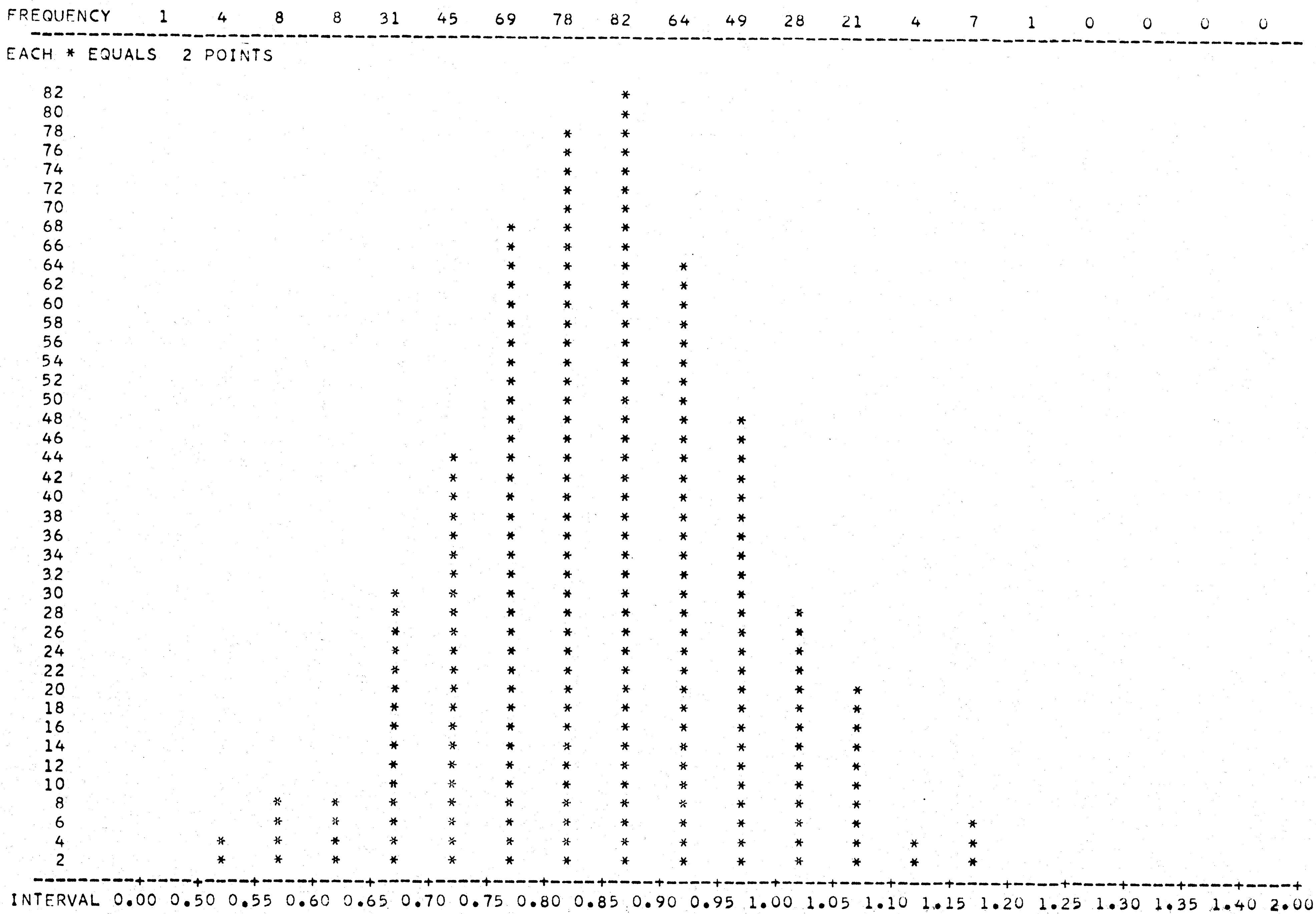


FIGURE 2. FREQUENCY DISTRIBUTION OF THE ESTIMATES OF BETA OBTAINED ACCORDING TO METHOD II.

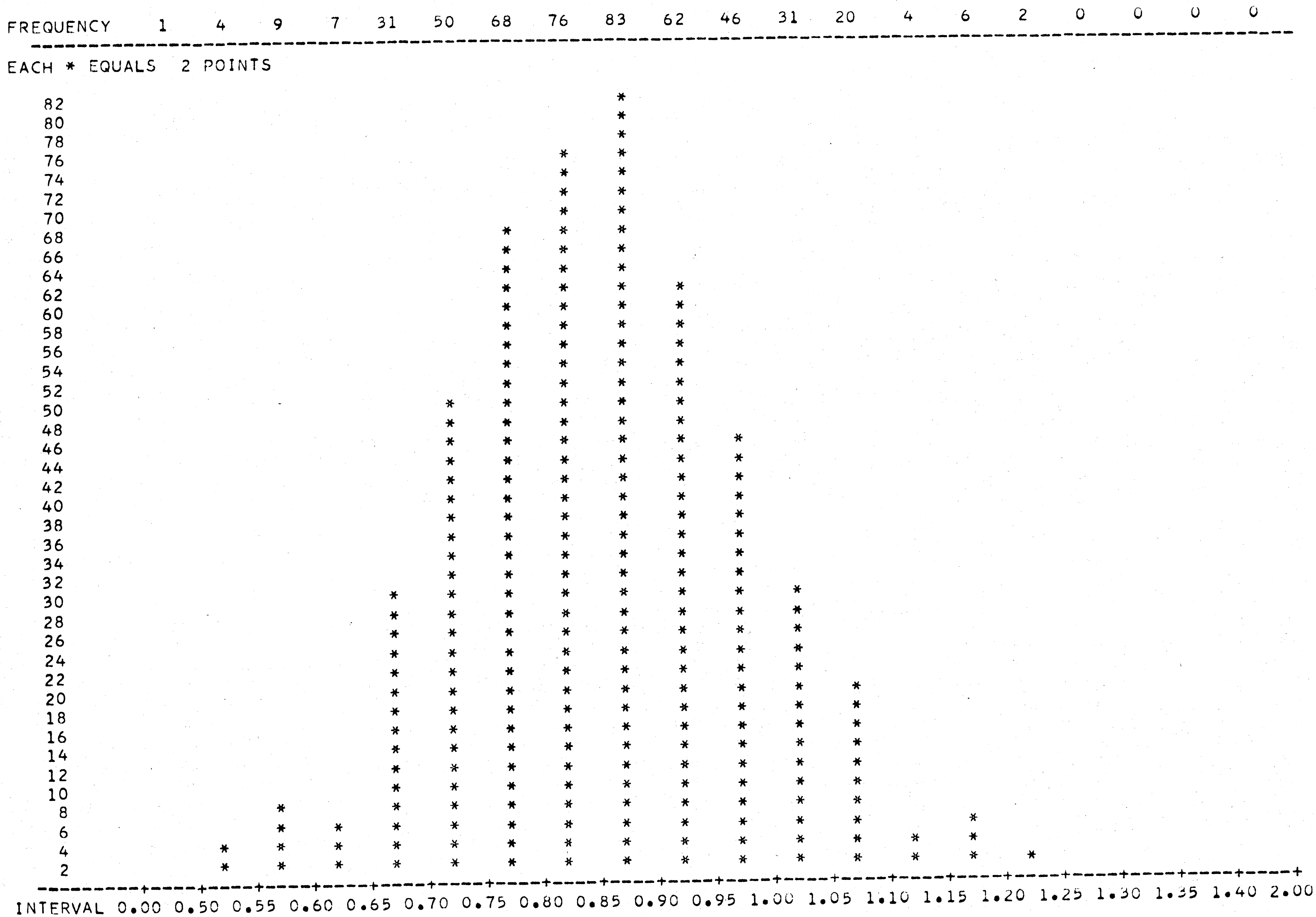


FIGURE 3. FREQUENCY DISTRIBUTION OF THE ESTIMATES OF BETA
OBTAINED ACCORDING TO METHOD III.

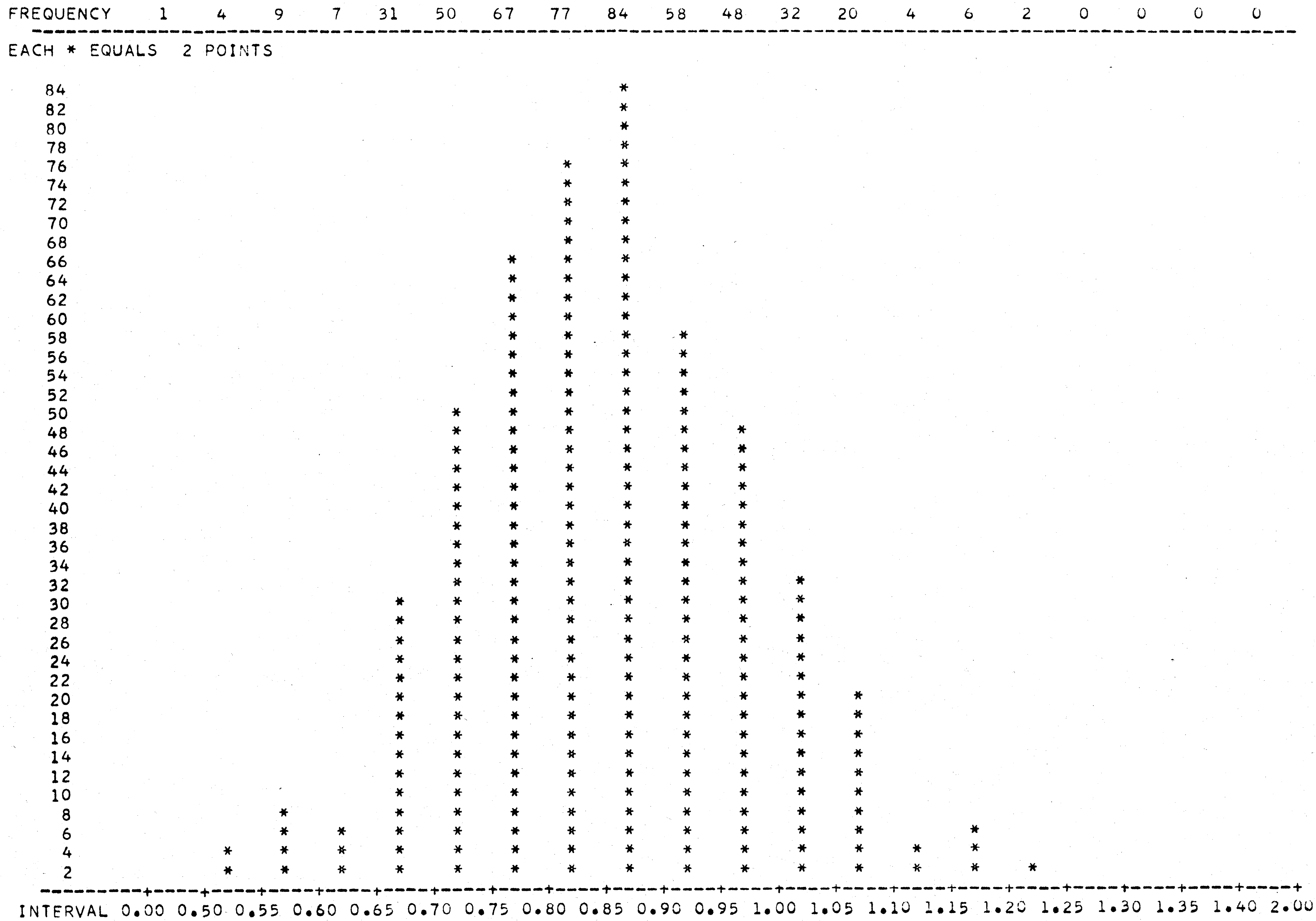


FIGURE 4. FREQUENCY DISTRIBUTION OF THE ESTIMATES OF BETA
OBTAINED ACCORDING TO METHOD IV.

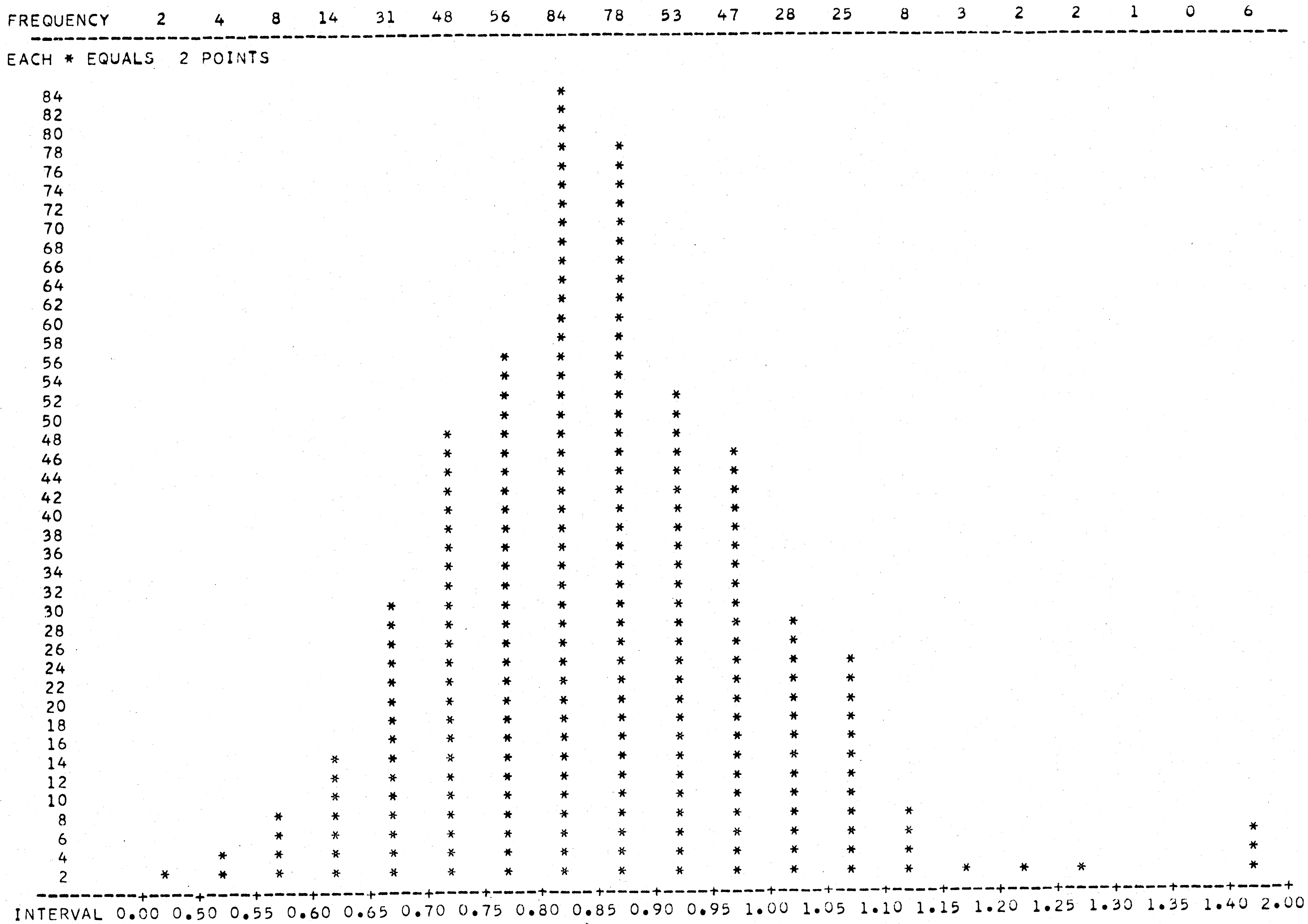
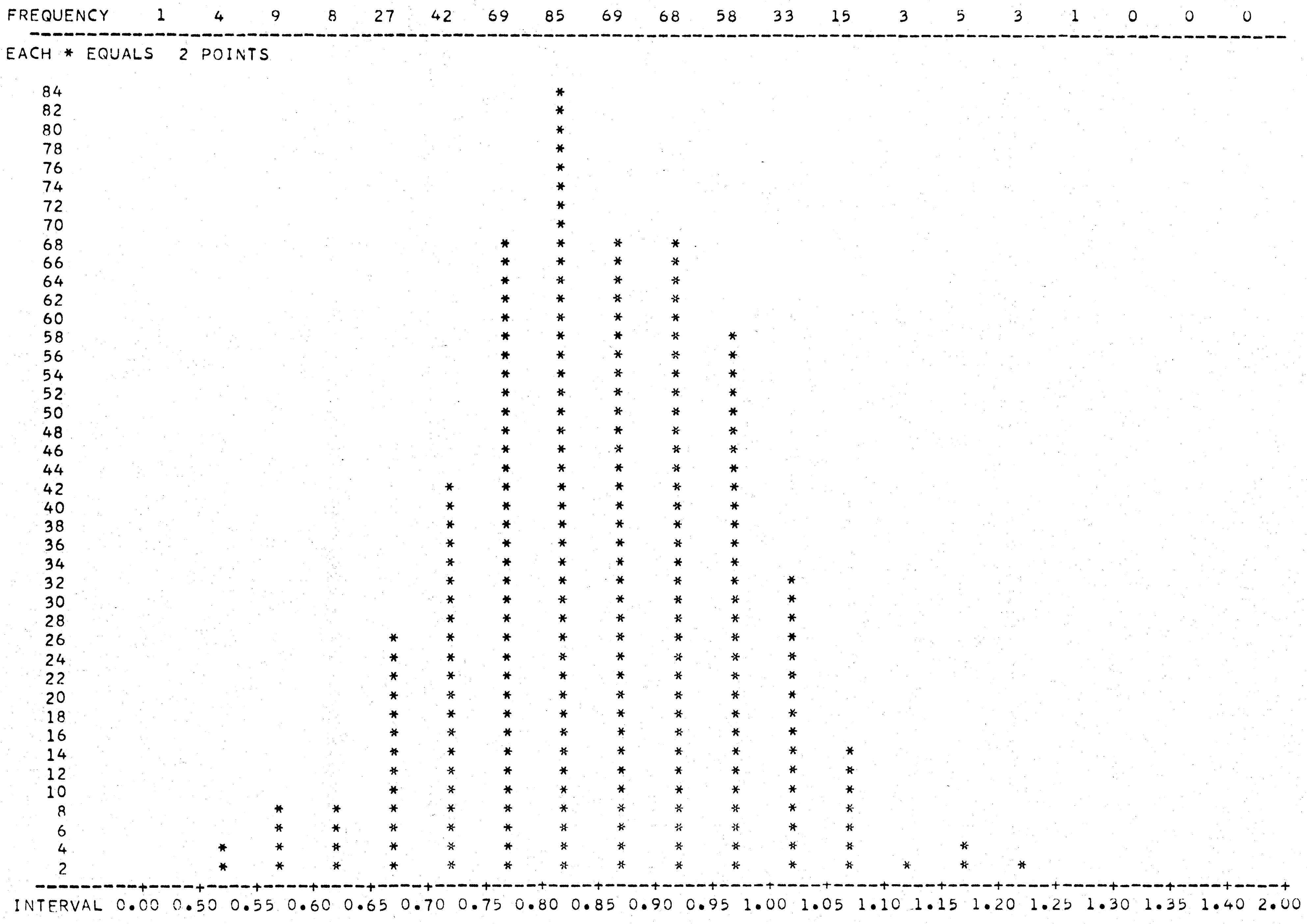


FIGURE 5. FREQUENCY DISTRIBUTION OF THE ESTIMATES OF BETA
OBTAINED ACCORDING TO METHOD V.



REPORTS 1977

7700 List of Reprints, nos. 179-194; List of Reports, 1976.

7701/M "Triangular - Square - Pentagonal Numbers", by R.J. Stroeker.

7702/ES "The Exact MSE-Efficiency of the General Ridge Estimator relative to OLS", by R. Teekens and P.M.C. de Boer.

7703/ES "A Note on the Estimation of the Parameters of a Multiplicative Allocation Model", by R. Teekens and R. Jansen.

7704/ES "On the Notion of Probability: A Survey", by E. de Leede and J. Koerts.

7705/ES "A Mathematical Theory of Store Operation", by B. Nooteboom.

7706/ES "An Analysis of Efficiency in Retailing, by B. Nooteboom.

7707/S "A Note on Theil's Device for choosing a Blus Base", by C. Dubbelman

7708/E "A General Market Model of Labour Income Distribution: An Outline", by W.H. Somermeyer.

7709/E "Further Results on Efficient Estimation of Income Distribution Parameters", by T. Kloek and H.K. van Dijk.

7710 "List of Reprints, nos. 195-199; Abstracts of Reports First Half 1977".

7711/M "Degenerating Families of Linear Dynamical Systems I", by M. Hazewinkel.

7712/M "Twisted Lubin-Tate Formal Group Laws, Ramified Witt Vectors and (Ramified) Artin-Hasse Exponential Mappings", by M. Hazewinkel.

7713/EM "An Efficient Way in Programming 'Eaves' Fixed Point Algorithm" by R. Jansen and A.S. Louter.

7714/E "An Alternative Derivation and a Generalization of Nataf's Theorem", by W.H. Somermeyer and J. van Daal.

7715/M "Application of Non-Linear Programming to Plane Geometry", by R. Stroeker.

7716/S "Stochastic Compactness of Sample Extremes" by L. de Haan and G. Ridder.

7717/ES "Cross-Section Studies of Efficiency in Retailing, Part I: Individual Grocers and Butchers", by B. Nooteboom.

7718/O "On Redundancy in Systems of Linear Inequalities", by J. Telgen.

7719/O "On R.W. Llewellyn's Rules to Identify Redundant Constraints; A Detailed Critique and Some Generalizations", by J. Telgen.

7720/O "Redundant and Non-Binding Constraints in Linear Programming Problems", by J. Telgen.

7721/MO "The Complexity of the Network Design Problem", by D.S. Johnson, J.K. Lenstra and A.H.G. Rinnooy Kan.

7722/O "Branching from the Largest Upper Bound: Folklore and Facts", by B.L. Fox, J.K. Lenstra, A.H.G. Rinnooy Kand and L.E. Schrage.

7723/E "A New Diskz and a New Coldstart for the IBM 1130 Computing System", by F. van Doeland.

7724/E "Estimating Quarterly Models with Partly Missing Quarterly Observations", by G.M.M. Gelauff and R. Harkema.