DEPARTMENT OF MATHEMATICS

Final Report to the Ministry of State for Science and Technology

on

THE QUEEN'S MATHEMATICS DEPARTMENT POPULATION MODEL

by

J.H. Davis and J.H. Verner
Queen's University, Kingston, Ontario

Queen's Mathematical Preprints No. 1975-33

VOLUME I

Queen's University

AT KINGSTON
ONTARIO, CANADA

Final Report to the Ministry of State for
Science and Technology
on
THE QUEEN'S MATHEMATICS DEPARTMENT POPULATION MODEL

by
J.H. Davis and J.H. Verner
Queen's University, Kingston, Ontario

Queen's Mathematical Preprints No. 1975-33

VOLUME I

Contents

Letter of Transmittal

Dr. A.R. Demirdache,
Director,
Technological Forecasting
  and Technology Assessment Division,
Ministry of State for Science
  and Technology,
270 Albert Street,
Ottawa, Canada,
K1A 1A1.

Dear Dr. Demirdache:

I have pleasure in forwarding to you the final report prepared by Dr. Jon Davis and Dr. James Verner of the Department of Mathematics, Queen's University of the Population Model which was commissioned by the Ministry.

Since the general philosophy of the Model was set forth in our Interim Report of April 15, 1974, it was felt unnecessary to repeat that in the introduction to the Final Report, but rather to append the Interim Report as Appendix D .

I believe that you will agree with my judgement that the present work opens up a new and potentially extremely important approach to modelling the growth of populations. This task is an indispensable step in any serious attempt to model national, socio-economic problems.

I believe that you will wish to bring this Report to the attention of the statistical services of the Government, with a view to studying its implications for the form in which population statistics should be collected and encoded, in order that there be available the data necessary for realistic modelling of the population of Canada as a whole or of any of its regions.

I would remind you that according to the contract between us, the right of publication of the details of the procedures outlined in this Report in scientific journals has been reserved to Professors Davis and Verner. This, of course, does not preclude the Ministry from distributing the whole Report in any way it wishes. In particular, you may consider that it would be of interest to the International Institute for Applied Systems Analysis and to various organs of the United Nations concerned with economic development as related to population growth.

Yours faithfully,

AJC:mh

A.J. Coleman, Head,
Department of Mathematics,
Queen's University.

Abstract

This report discusses final development of the Queen's Mathematics Department Population Model.

The basic models and techniques presented in the authors' interim report on this work have been extended, and the algorithms have been tested on the available real data.

Also reported here is the development of methodology and software for the estimation of linear input-output models which includes a provision for the modelling of an unknown exogenous component in the observed output records.

Acknowledgements

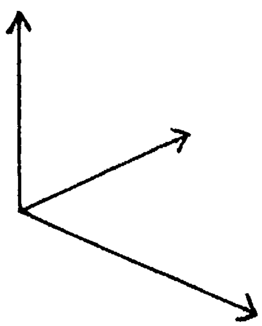
Unpublished data used here for modelling the Canadian population has been provided by two divisions of Statistics Canada: the Population Estimates and Projections Division, Census Field, and the Vital Statistics Section, Heath Division. The authors acknowledge with appreciation the provision of this data, and in particular, thank Mr. Mr. J. Kelly and Mr. D. Nagnur for their co-operation in transmitting the data.

Much of the computer programming contained herein results from the efforts of two graduates of Queen's program in Computer Science and Mathematics. For this assistance the authors express their gratitude to Cornelia Berghout and John L. Martin.

The authors would also like to thank Marsha Hartley and Gloria Thompson for their efforts in the typing and preparation of this report.

Surface Plots

Certain results in this report are displayed in the form of computer-drawn plots of surfaces in three-dimensions. The variables involved <u>increase</u> in the directions schematically indicated below:



The ranges of the "base variables" involved are mentioned in the body of the text; the range of the "height variable" appears beside each drawing.

I. Introduction

This report considers further aspects of the general problems of the construction of dynamical population models using methods described in the authors' Interim Report [3].

The basic structure of the models considered is described at length in [3], and the various reasons for adopting that structure are explained in that document as well. Our subsequent experience has not led us to modify our basic models, and the general numerical methods for estimation proposed in [3] have required only minor modification for use in practice. For these reasons, a lengthy review of the material in [3] has been omitted in this report.

The main purpose in undertaking this project was to attempt to produce a basic framework, and a set of adaptable, useable numerical algorithms for use in simulation studies involving population. A basic framework is presented in the Interim Report; this report is largely devoted to the task of making the work carried out more useful.

The models which have been considered in this work are of a relatively ambitious nature. Simulations with these models would allow study of detailed population distribution problems. These potential benefits are not without cost, however, as the methods used in this project may not be familiar to all possible users of these results. Further, as much as one might hope that the algorithms and associated computer software developed in this project would be entirely foolproof, this is not the case. From our experience using the algorithms reported here in connection with real data, it appears that a certain amount of judgement is required on the part of the user of these methods. The user must be aware of the underlying assumptions of the model, and of the limitations of the numerical methods employed in the software in order to successfully interpret the results of computations. It is our hope that this report will provide some insight in this area, as well as serve as a "user's manual" for the software included in the Appendix.

Chapter II of this report is devoted to the problem of estimating the coefficients $a_i(t)$, $d_i(x)$ and $b(t)$ in the model

$$\frac{\partial f}{\partial t}(x,t) = -\frac{\partial}{\partial x}\left(\sum_{i=1}^{q} a_i(t)d_i(x)f(x,t)\right) - b(t)\, f(x,t)$$

adopted (See [3] and Chapter II) to describe the evolution of the fertility curve over time. Data supplied by Statistics Canada was utilized, and the numerical results obtained support the utility of the model and associated estimation method.

In Chapter III we discuss problems arising in the estimation of the coefficients in the partial differential equation

$$\frac{\partial p}{\partial t} = -\frac{\partial p}{\partial x} - \frac{\partial}{\partial s}\left(\sum_{i=1}^{N} a_i(t) d_i(x,s) p\right) - r(x,t)\, p + i$$

governing the evolution of the age and income dependent population distribution.

In the case of this estimation problem, the data which we were able to obtain is much less extensive than would be desirable. Problems arise both because of the limited number of years for which data is available, and because of the level of aggregation in the published data.

The results of numerical experiments on the effect of aggregation on the accuracy of the estimation algorithm are displayed, together with the results of computations using data obtained from Revenue Canada publications.

As pointed out in [3], the next step _after_ the coefficients of the partial differential equation have been estimated is to attempt to develop a model for the dynamical behaviour of these coefficients. Chapter IV consists of one effort in this direction; we have attempted to model the dynamical relationship between the fertility coefficients and various economic time series.

It was anticipated in [3] that it would be possible to apply more or less conventional time series identification techniques (or at worst, perhaps non-linear regression methods) to this problem. The available fertility data, unfortunately, is such that it has been possible to compute estimates for the fertility parameters only for a relatively short period of time. The amount of available data simply appears too small to allow conventional statistical analysis. The situation is further complicated by the fact that one suspects at least the possibility of a large exogenous component (not "driven" by the economic factors) in the fertility parameters. This might well make use of standard approaches difficult even if a longer data run were available.

The approach that we have taken is to make an adaptation and extension of certain techniques developed for input-output identification of control systems to the present problem. While the results of numerical experiments on the available data may be described as inconclusive, it is hoped that the method developed will prove useful in modelling studies.

Finally, copies of the computer programs developed are included in the Appendix, together with some comments relevant to use of the programs. Considerable effort has been made to provide clear documentation in the programs.

II. Fertility Estimation

A General Observations

As mentioned above, for the purposes of this final report, we assume that the reader is familiar with the authors' interim report [3].

We recall the suggestion made in [3] that the dynamical behaviour of the fertility curve could be adequately modelled by the relatively simple partial differential equation

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial}{\partial x}(a(t)d(x)f(x,t)) - b(t)\,f(x,t) \quad .$$

It is more useful to hypothesize a slightly more general model of the form

$$\frac{\partial f}{\partial t}(x,t) = -\frac{\partial}{\partial x}\left(\sum_{i=1}^{q} a_i(t)d_i(x)f(x,t)\right) - b(t)\,f(x,t) \quad ,$$

in order to include the possibility of more than two time functions affecting the fertility curve behaviour. (Calculations on the actual data, however, suggest that the original simplified representation is adequate.)

Introduction of an integrating factor and relatively simple manipulations lead to an expression of the form

$$\sum_{i=1}^{q} a_i(t)\, d_i(x) = -\frac{1}{\bar{f}(x,t)} \int_{\underline{x}}^{x} \frac{\partial \bar{f}}{\partial t}(\zeta,t)\, d\zeta$$

from which $a_i(t)$, $d_i(x)$ are determined. (See the attached Appendix for more details.)

With real data, the expression on the right hand side of the above must be evaluated by numerical means. The spline routines utilized are described in the Appendix and the interim report [3].

In connection with the use of the methods proposed, two major issues arise. The first, and perhaps more philosophical problem, is that of validation of the model proposed above. As pointed out in [3], the function $b(t)$ is <u>defined</u> in such a way that the presence of the $b(t)$ term in the governing equation is essentially valid by the definition of $b(t)$. The question of the term involving $a_i(t)$ and $d_i(x)$ is quite a different matter, however, since the argument for this is essentially that the observed behaviour suggests a governing partial differential equation of the above form [12].

It is, of course, impossible to supply "proof" of the correctness of any hypothesized model. Essentially the only criterion which may be applied is that of consistency with the observed data.

One form of consistency, perhaps that which comes first to mind, is the requirement that a simulation of the fertility curve should reproduce the historical data to an acceptable degree of accuracy. While this is a necessary condition, one must demand more from the results of an estimation in order to have confidence in the results.

A second form of consistency is the requirement that the estimates be consistent not just "in the large" (as a trial simulation shows), but that they be consistent with subsets of the available data. Consistent results returned from repeated calculations of this sort strongly suggest that the results obtained "actually are present in the data". Needless to say, reinforcement of this sort is not available from a one-shot estimation procedure.

In the present situation, it is most natural to make a sequence of estimates based essentially on data years j to $j + L$, for varying starting date j (L here is the data length required in the computational procedure). For the problem of estimating

$$\sum_{i=1}^{q} a_i(t)\ d_i(x)$$

one ideally hopes to obtain "overlapping sections" of the smooth time functions $a_i(t)$ from this procedure, and consistent smooth age functions $d_i(x)$. The extent to which this occurs in the actual data runs may be taken as an indication of the validity of the model.

The second major issue that arises in connection with use of the estimation procedure suggested in [3] is that of use with real data. The algorithm has a firm theoretical basis, and the numerical experiments with artifically generated data reported in [3] show that the numerical procedures have reasonable behaviour. The real data, however, is subject to various (unknown) errors; this compounds the problems arising from the likelihood that the model will not provide an exact fit for the physical situation.

Looking at the estimation formula above, one may anticipate two major sources of trouble. The first is in the differentiation with respect to time, a procedure bound to accentuate errors. In fact, the integration serves in practice to smooth these accentuated errors considerably. The major errors seem to occur at the ends of the time interval; this is expected from the behaviour of the spline routine. The second large source of error occurs at the ends of the age interval. This is caused by the fact that $\bar{f}(x,t)$ approaches zero at the ends of the interval of interest, so that errors in

$$\int_{\underline{x}}^{x} \frac{\partial \bar{f}}{\partial t} (\zeta,t)\, d\zeta$$

are magnified considerably.

If we let $M(x,t)$ represent the numerically calculated values of

$$- \frac{1}{\bar{f}(x,t)} \int_{\underline{x}}^{x} \frac{\partial \bar{f}}{\partial t} (\zeta,t)\, d\zeta \quad .$$

Then the situation may be summaried by writing

$$M(x,t) = \sum_{i=1}^{q} a_i(t)\, d_i(x) + B(x,t) + N(x,t) \quad .$$

In the above, $B(x,t)$ represents the error due to the two effects mentioned above; this implies that $B(x,t) = 0$, except in the immediate neighbourhood of the edges of the region of interest. $N(x,t)$ represents a (hopefully small) modelling and numerical error, and is not restricted in location and extent. In order to remove the error term $B(x,t)$, we may simply multiply the above equation by a function

$$\chi_1(t)\, \chi_2(x)$$

(a product of characteristic functions) which is identically equal to unity on the interior of the region of interest, but equal to zero near the boundary of the region. This results in

$$M(x,t) \cdot \chi_1(t)\ \chi_2(x) = \sum_{i=1}^{q} (\chi_1(t) a_i(t)\ (\chi_2(x) d_i(x) + \chi_1(t)\ \chi_2(x)\ N(x,t) .$$

This procedure is easily implemented numerically simply by setting the "borders" of the matrix representing the function $M(x,t)$ to zero. The estimation procedure described in [3] (based on eigenvalue procedures) may then be applied to the <u>bordered array</u> to produce estimates of $(\chi_1(t)a_i(t))$ and $(\chi_2(x)d_i(x))$.

The need for this technique is easily seen by an attempt to run the estimation algorithm without the bordering procedure. The algorithm is essentially based on calculation of the eigenvalues and eigenvectors of the matrix

$$M^T M$$

where M is the array of (sample values of) the function $M(x,t)$ above. The eigenvalue package employed returns results ordered in magnitude; it is also easily seen that, if the representation

$$M(x,t) = \sum_{i=1}^{q} a_i(t)\, d_i(x)$$

is exact, then the eigenvalue associated with the eigenvector $a_i(t)$ is simply $d_i^T d_i$. (See the Appendix below.)

This means that running the algorithm on the "sliding window" of data should automatically return the $a_i(t)$, $d_i(x)$ functions in consistent order, namely in order of decreasing norm of $d_i(x)$. This makes consistency of repeated estimations very easy to see in plots of the calculated results.

The presence of large errors in the calculated value of $M(x,t)$, namely the error $B(x,t)$, has a disastrous effect on the above situation. $B(x,t)$ may (and in practice certainly does) contribute fairly large eigenvalues to the symmetric matrix MM^T . This destroys the consistent ordering of the returned eigenvectors, as well as producing considerable (and variable) distortion in the calculated values of the "real" $a_i(t)$, $d_i(x)$. In fact, it is the presence of calculated eigenvectors supported at the interval borders in preliminary calculations which has suggested the bordering procedure described above.

Use of the bordering procedure essentially introduces enough regularity to make computations possible. The price to be paid for this is that the calculations return estimates only for the "bordered" functions $\chi_1(t)\ a_i(t)$ and $\chi_2(x)\ d_i(x)$. This leaves the coefficients unknown on the border areas of the region. The lack of knowledge at the borders of the age interval causes little difficulty; since the fertility here is essentially zero, these values have little effect on simulations.

Loss of the borders in the time direction is somewhat more serious from the point of view of subsequent use of the estimates. With the limited amount of data available and the bordering losses, we obtain only a twelve year run of data for the $a_i(t)$. This small number of samples has made substantial difficulty in the problem of estimating a dynamical model for the evolution of these coefficients. (See Chapter IV below.)

There are other numerical problems (of smaller importance) associated with these calculations. These are described in the Appendix and on comment cards in the attached programs.

B Numerical Results

The sources and treatment of the data used for these computations are described in the Appendix. The complete computer programs and samples of the resulting output are also included in the appendix. We reproduce below the graphical summary of the computational results.

With use of the "sliding data window" and bordering procedure described above, the available data is sufficient to allow eight computations of estimated values of $a_i(t)$ and $d_i(x)$. Computations were made based on a 10×10 matrix. Greater or lesser consistency was obtained in the first four eigenvectors from each calculation, so that results from lower eigenvalues are suppressed in the output. The contribution from these components is small, and the eigenvectors show wide scatter. We attribute these to numerical and data error, and model inaccuracy.

In the plots below, only the highest run number is printed in case of coincident values. It is readily seen that the estimated values for the function $a_1(t)$ show a high degree of consistency, and that the variation in results appears relatively small. (In these plots, "9" represents the numerical average of the available runs. This is taken as the final estimate of the algorithm.)

The similar plots for $a_2(t)$, $a_3(t)$, and $a_4(t)$ show increasing scatter in the computed results. The plots do seem to indicate, however, the systematic presence of these higher order terms.

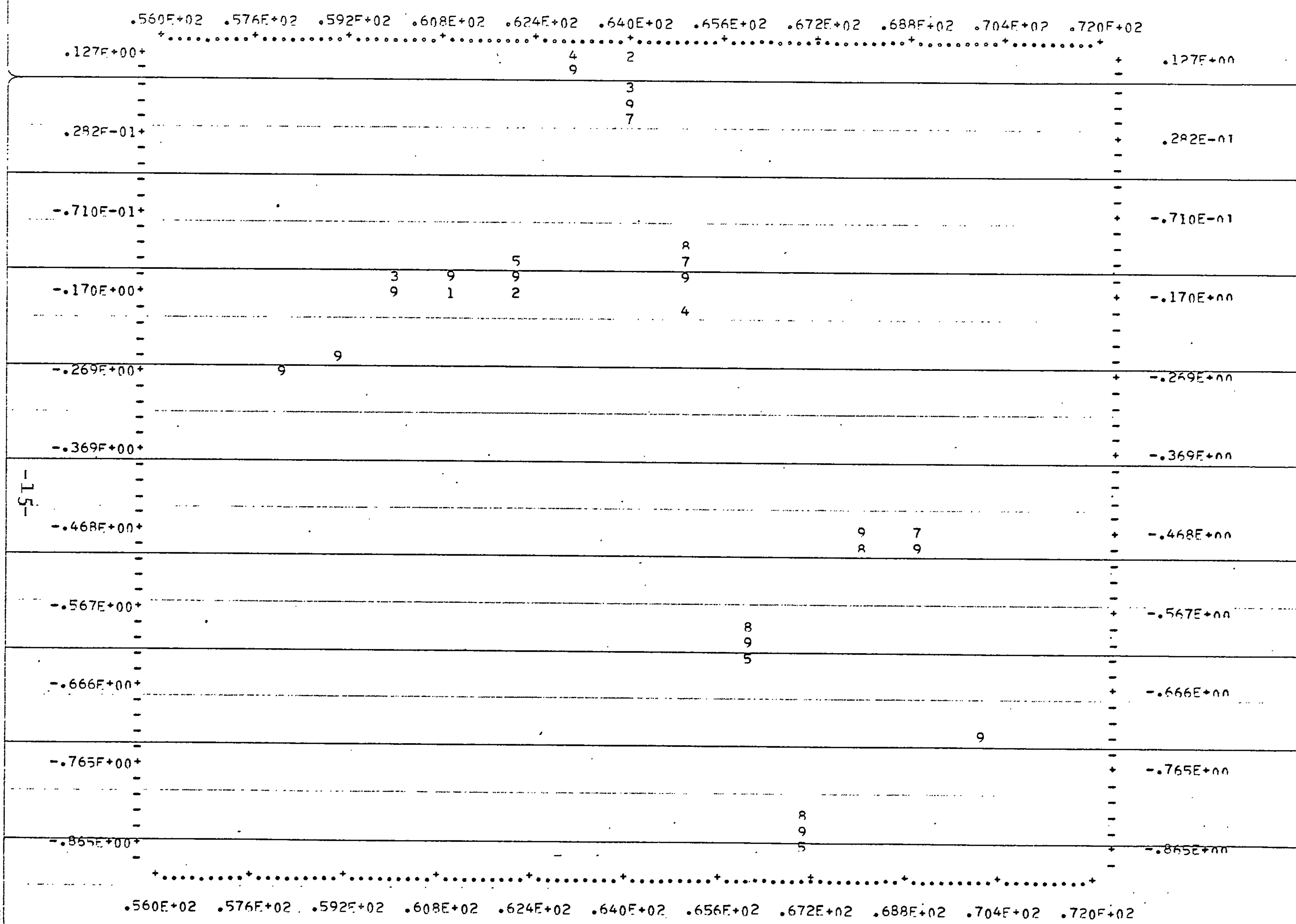
The computed results for the corresponding $d_i(x)$ show a similar degree of consistency, although a higher degree of scatter seems to be present. This is probably attributable to the fact that the estimation algorithm "back calculates" $d_i(x)$ according to

$$d_i = M a_i \quad .$$

While the eigenvector method used for calculating a_i apparently filters out errors in the calculated matrix M to some extent, no such effect is operating in the above computation. One may therefore expect greater error in the calculation of d_i .

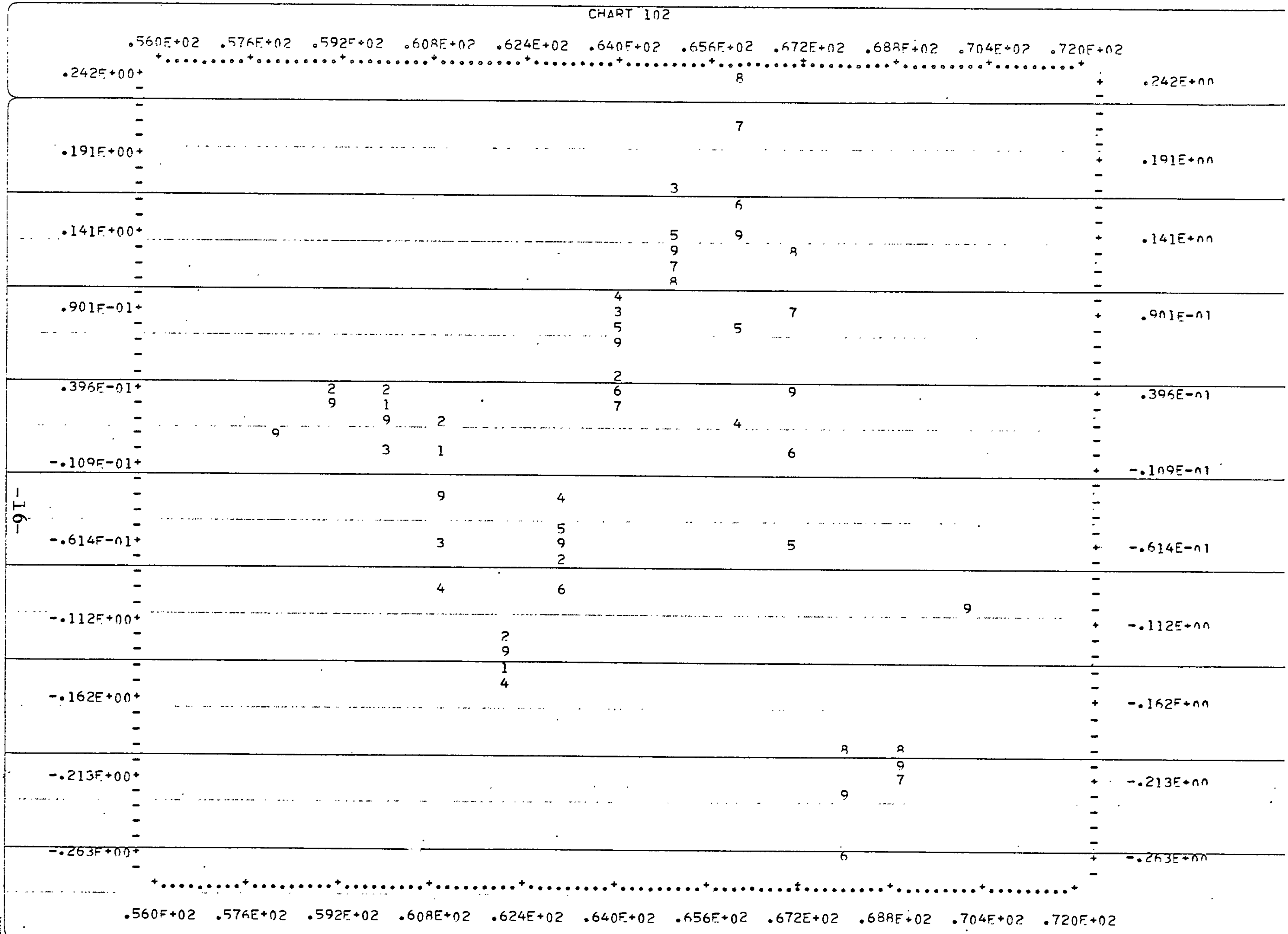
One might attempt to work with the adjoint of the matrix M, in which case a parallel method should return estimates of the d_i as eigenvectors. The resulting symmetric matrix is of the same dimension as the age interval in this case. As a result, the eigenvalue calculation is much more difficult computationally, and more subject to ill-conditioning problems.

CHART 101



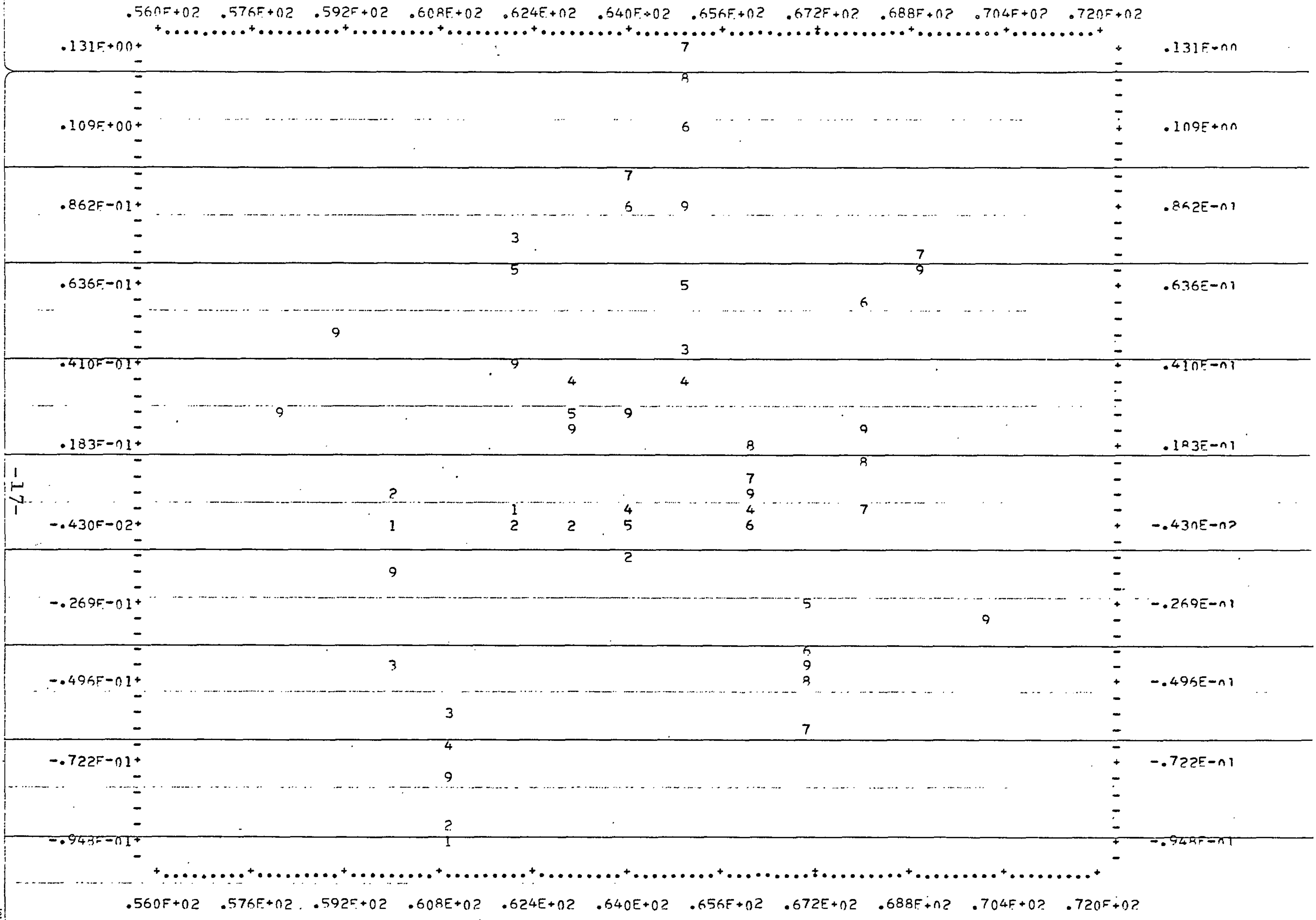
$a_1(t)$ - First Time Component

-15-

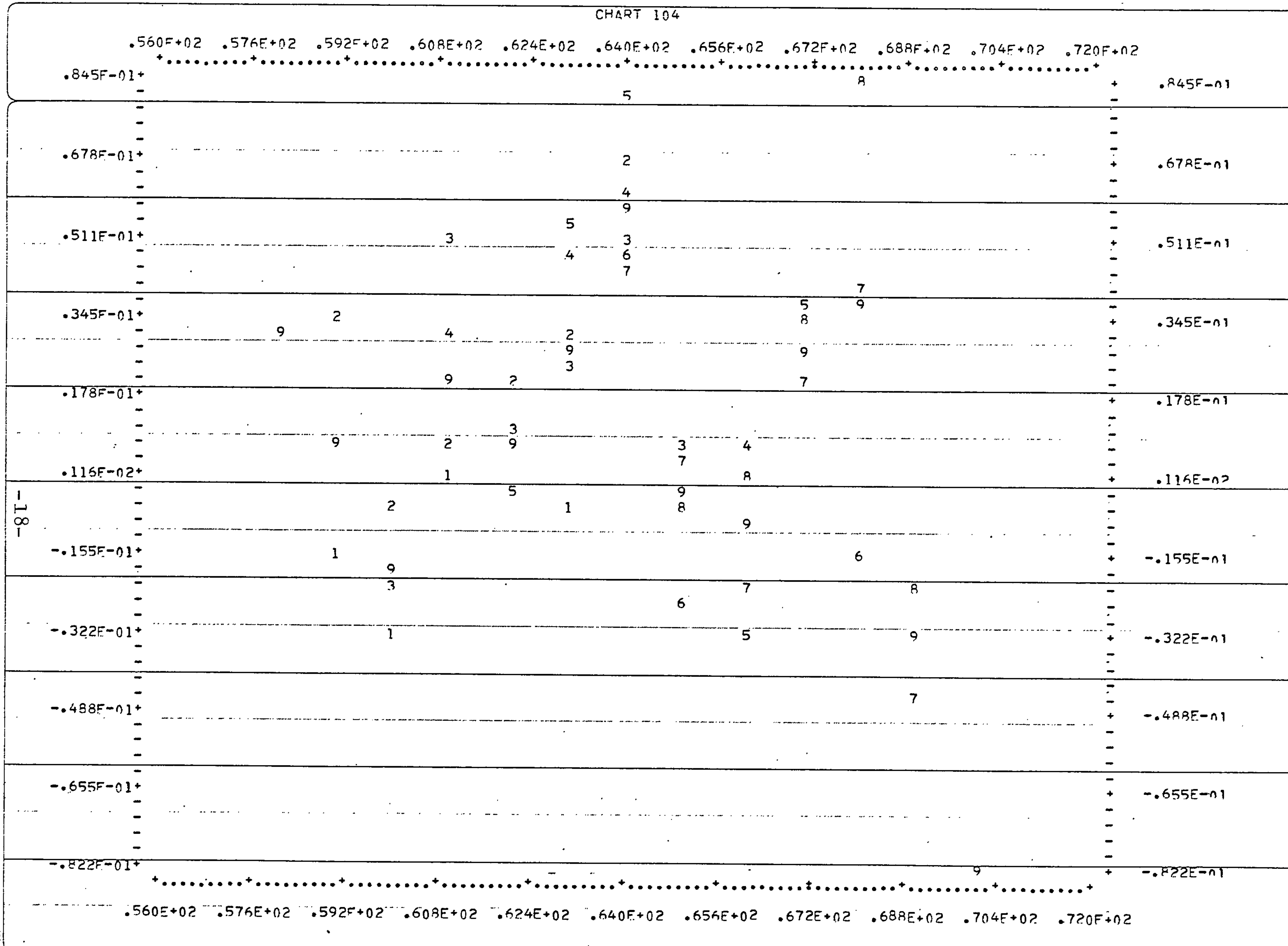


$a_2(t)$ - Second Time Component

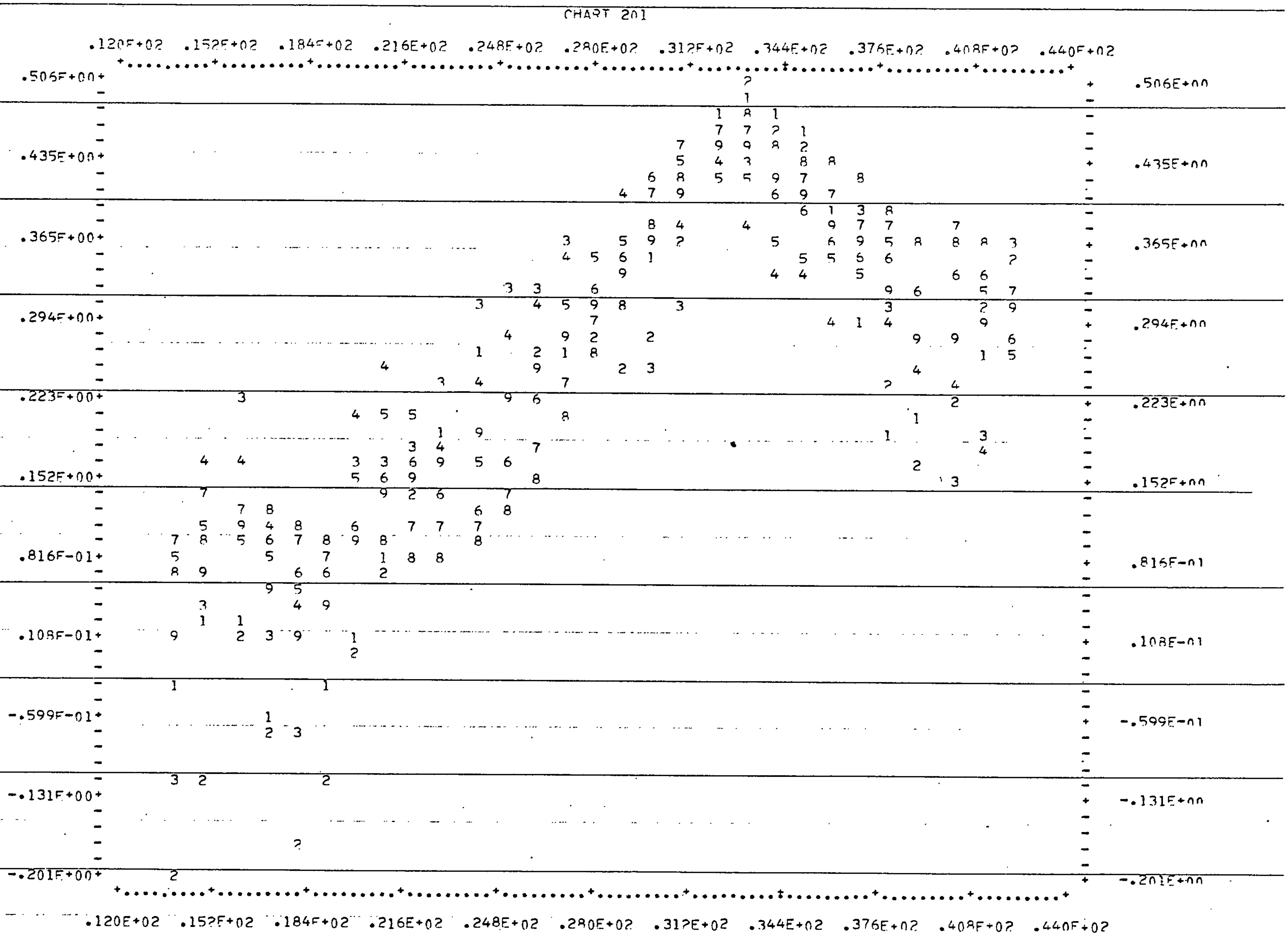
CHART 103



$a_3(t)$ - Third Time Component

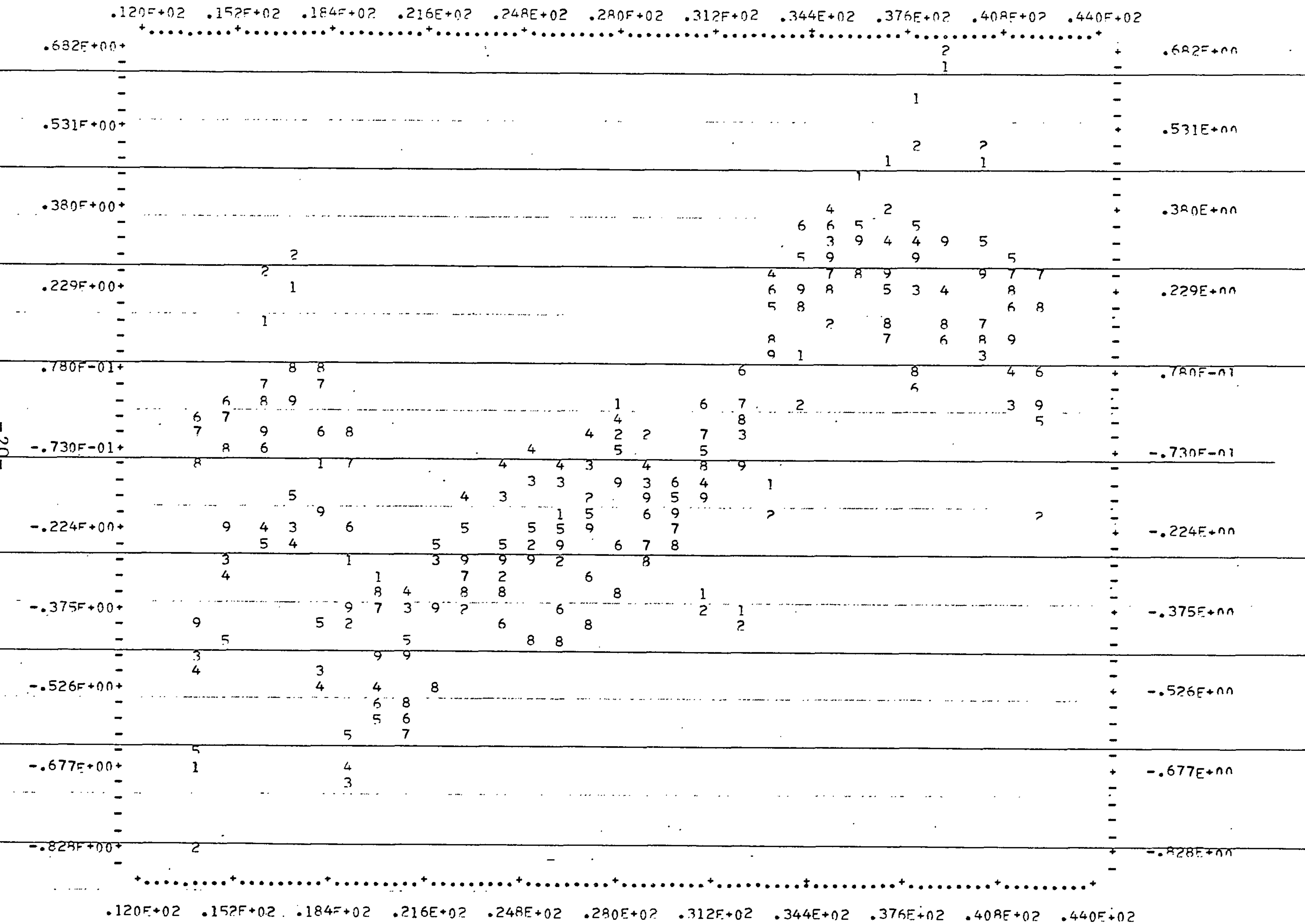


$a_4(t)$ - Fourth Time Component

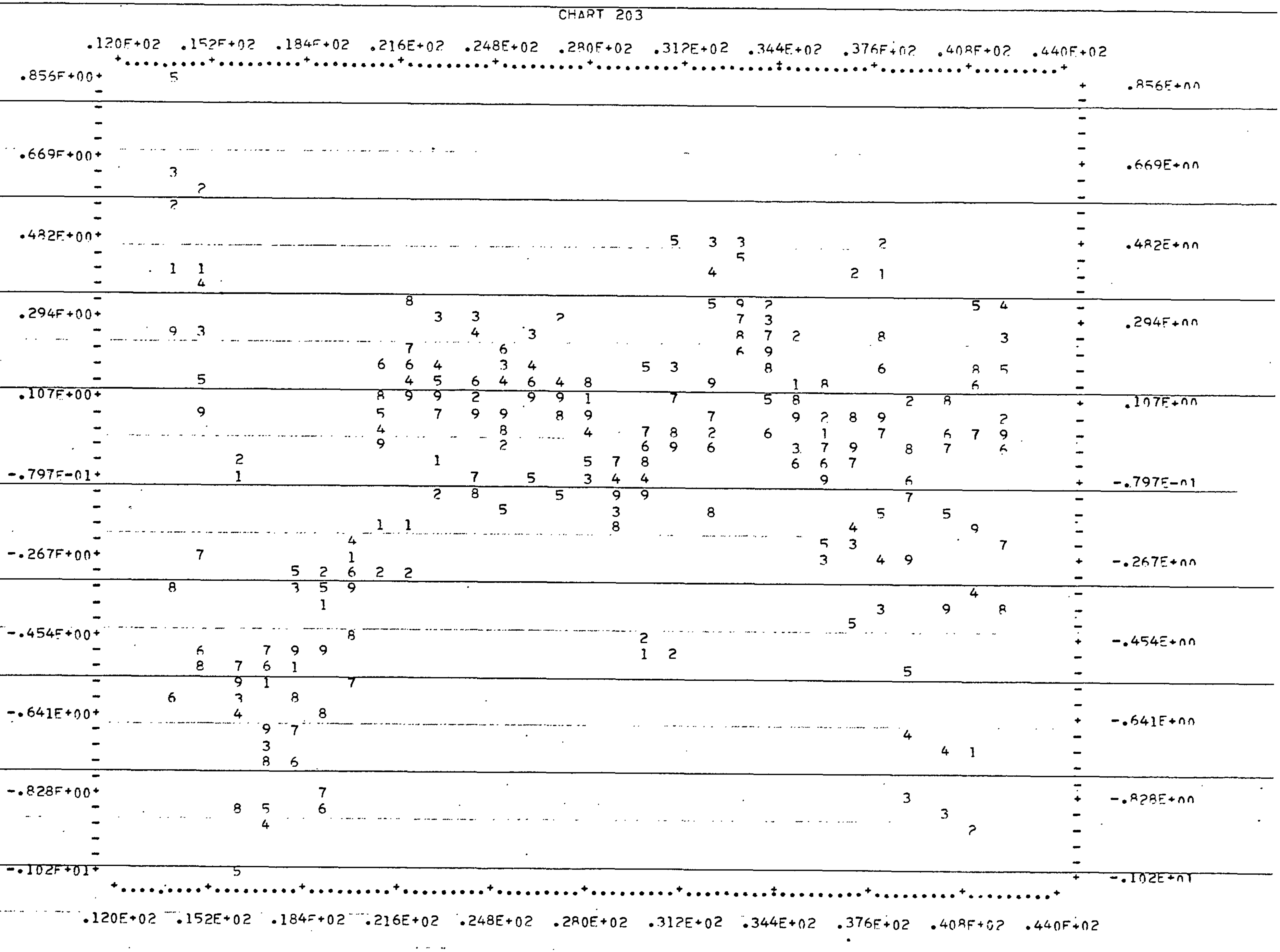


$d_1(x)$ - First Age Component

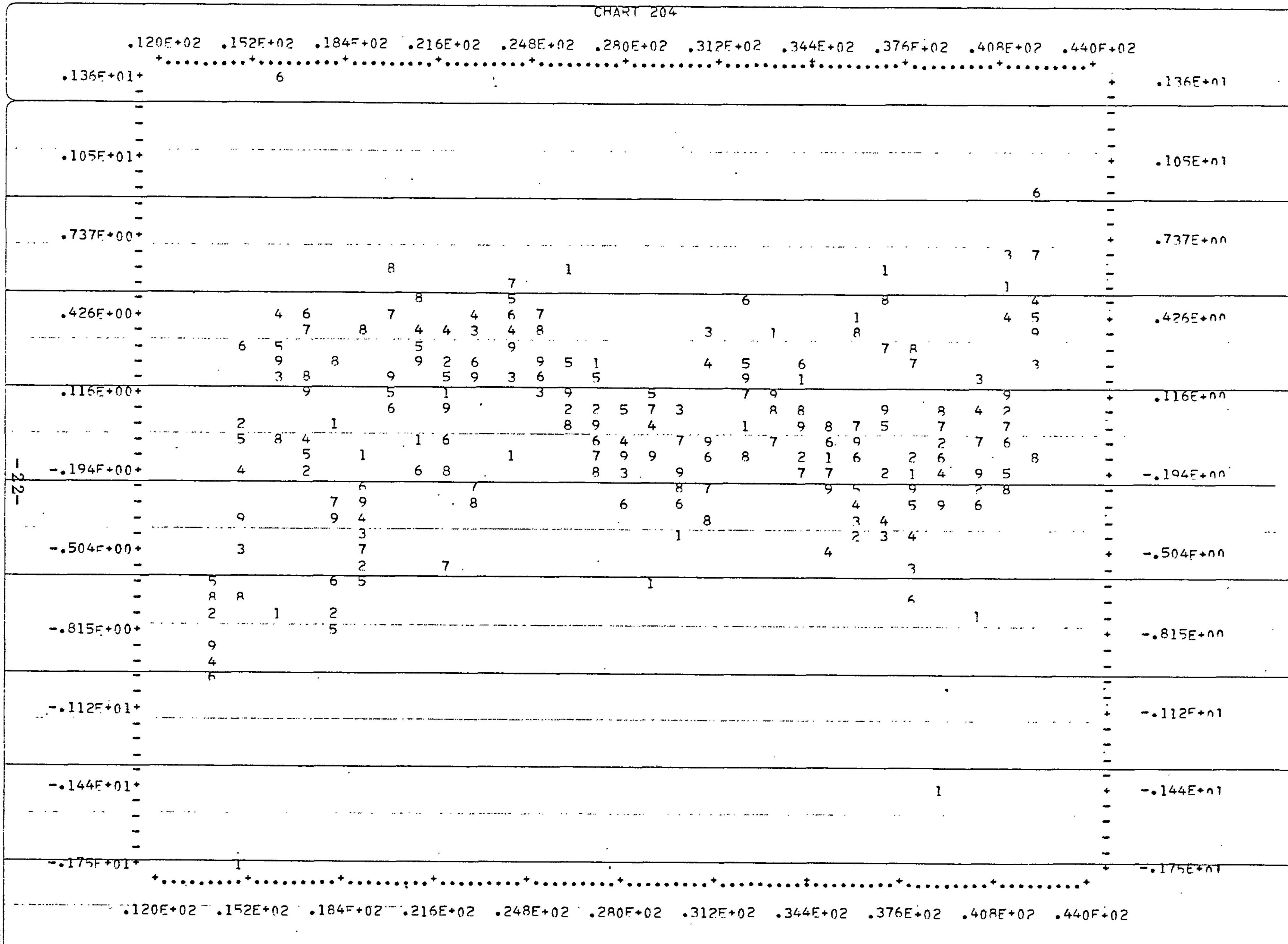
CHART 202



$d_2(x)$ - Second Age Component



$d_3(x)$ - Third Age Component



$d_4(x)$ - Fourth Age Component

The results of the above computations were checked by simulation of the evolution of the fertility curve over the years for which parameter esimates were calculated. Simulations were run using

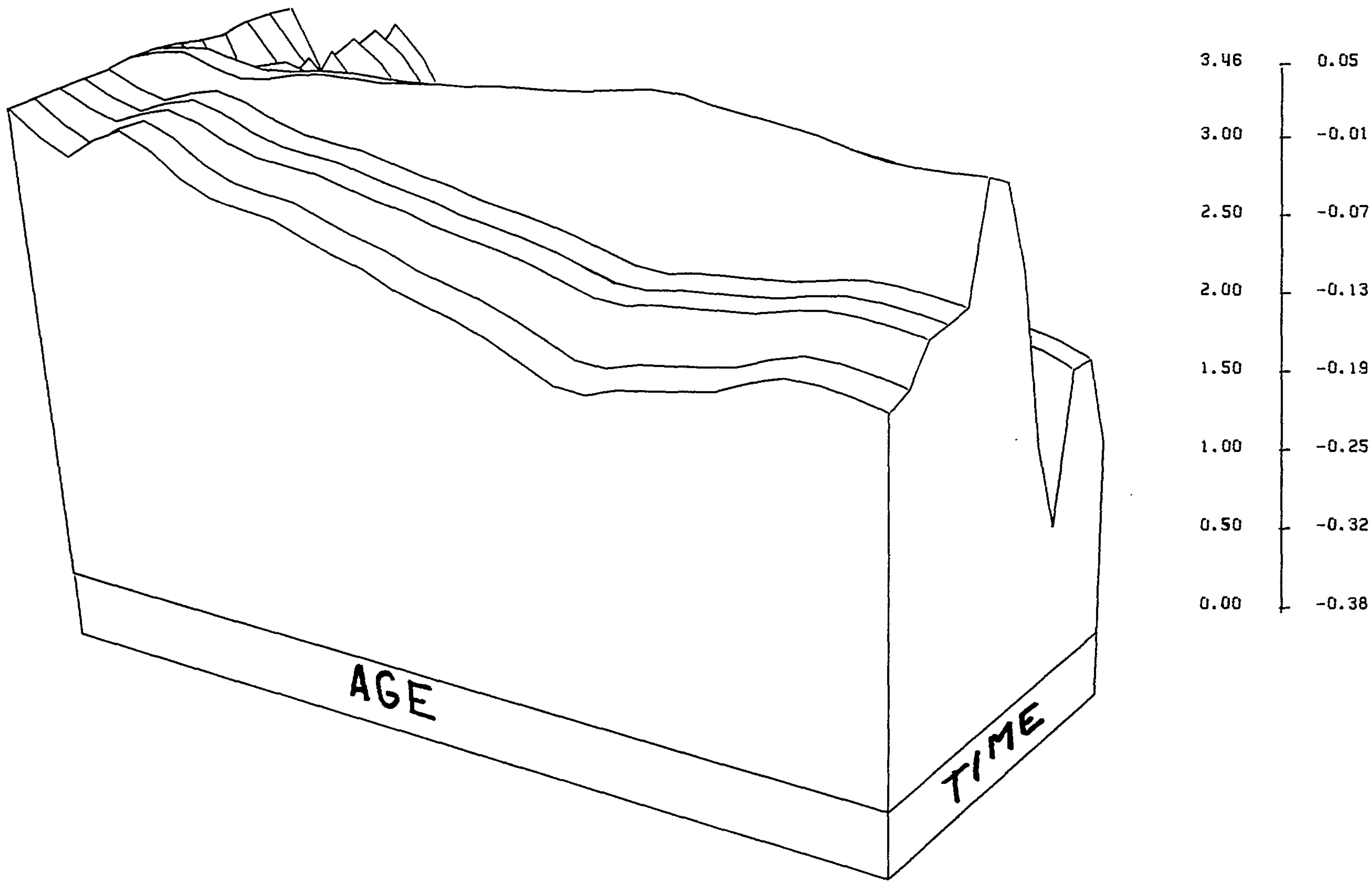
(i) $a_1(t)\, d_1(x)$

(ii) $a_1(t)d_1(x) + a_2(t)d_2(x)$

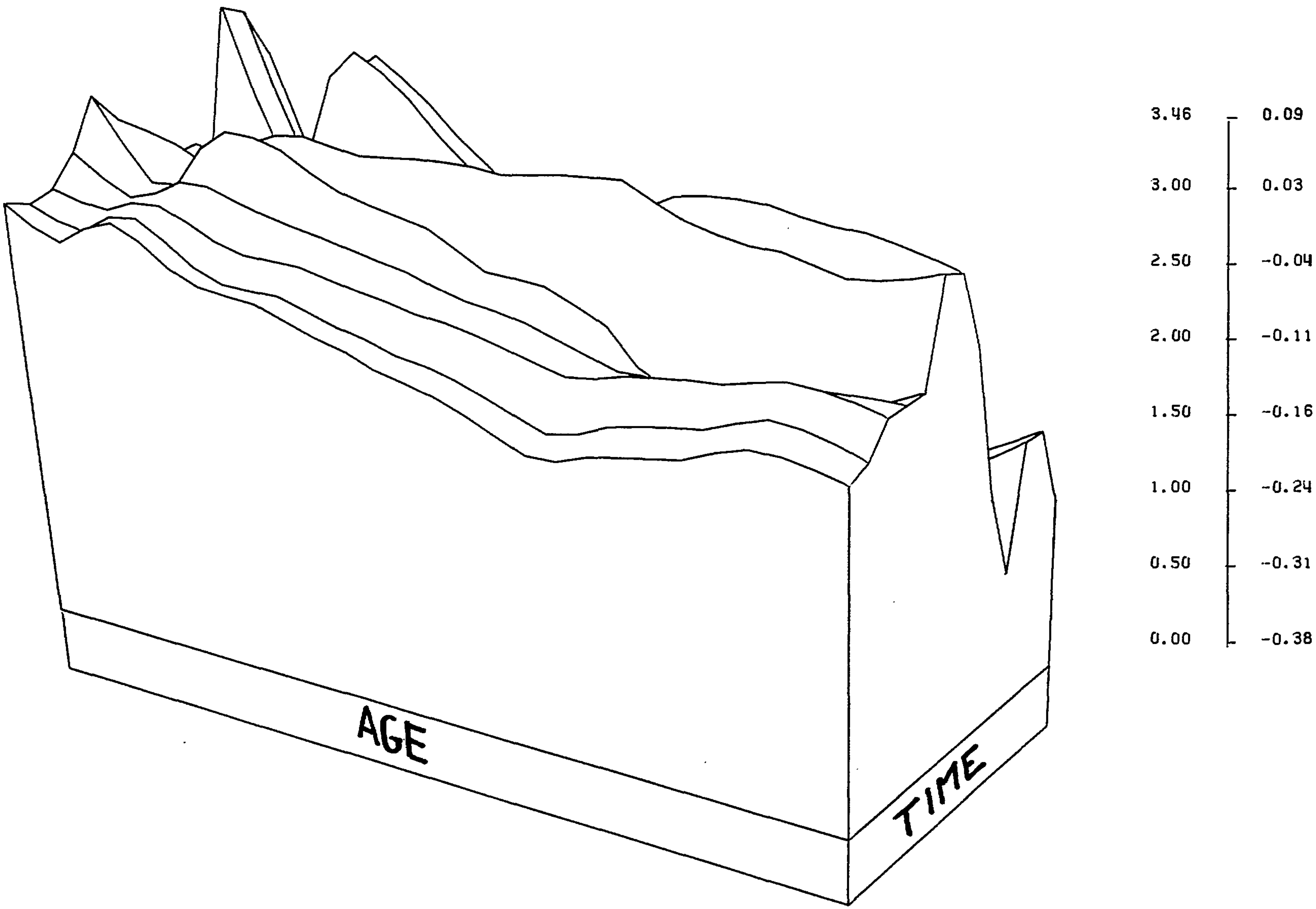
(iii) $a_1(t)d_1(x) + a_2(t)d_2(x) + a_3(t)d_3(x)$

(iv) $a_1(t)d_1(x) + a_2(t)d_2(x) + a_3(t)d_3(x) + a_4(t)d_4(x)$

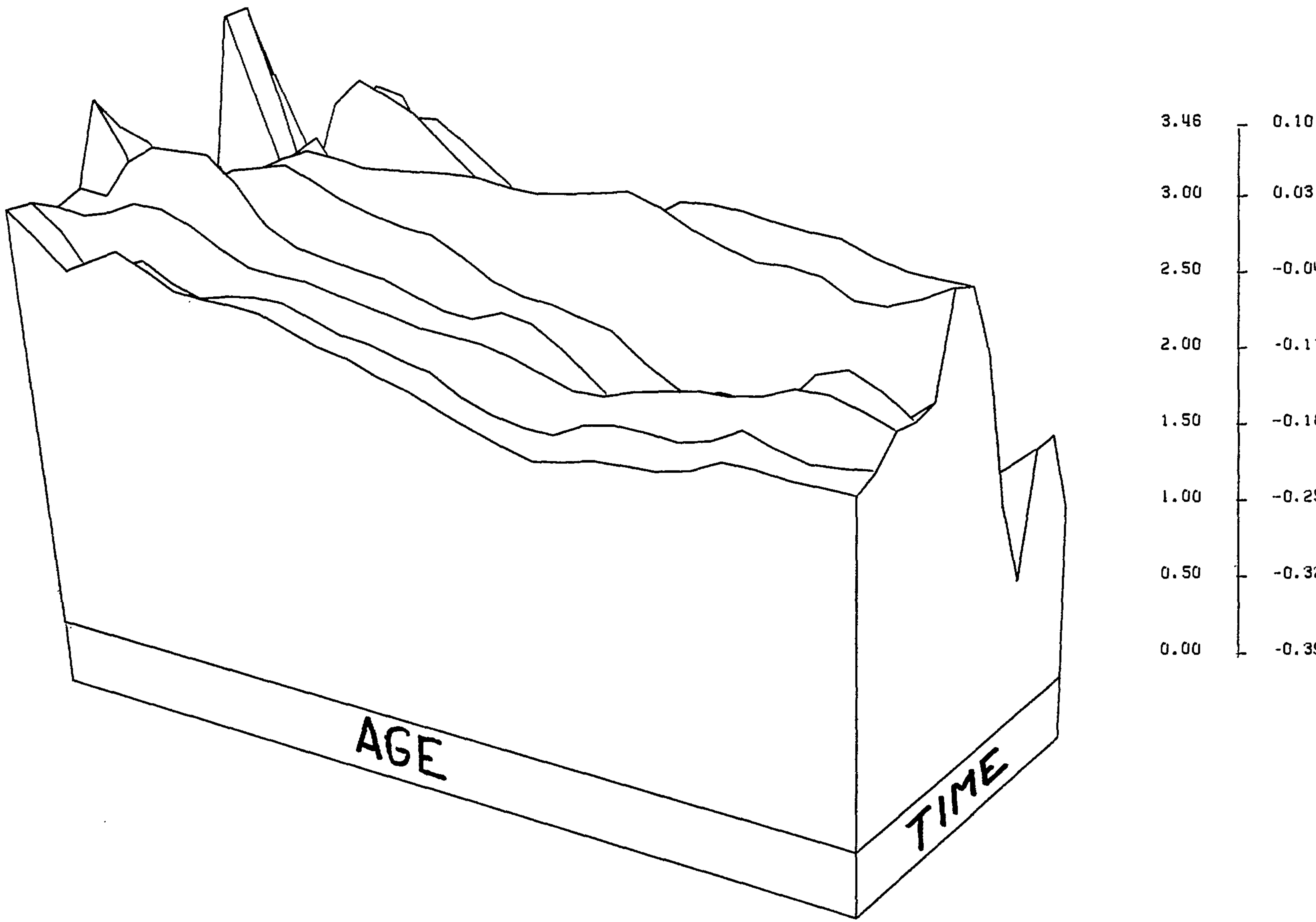
as coefficients in the model. Complete numerical results are presented in the Appendix, but a graphical presentation of the result is given below. Also included are plots of (i), (ii), (iii) and (iv) (labelled "fertility parameters" in the following plots).



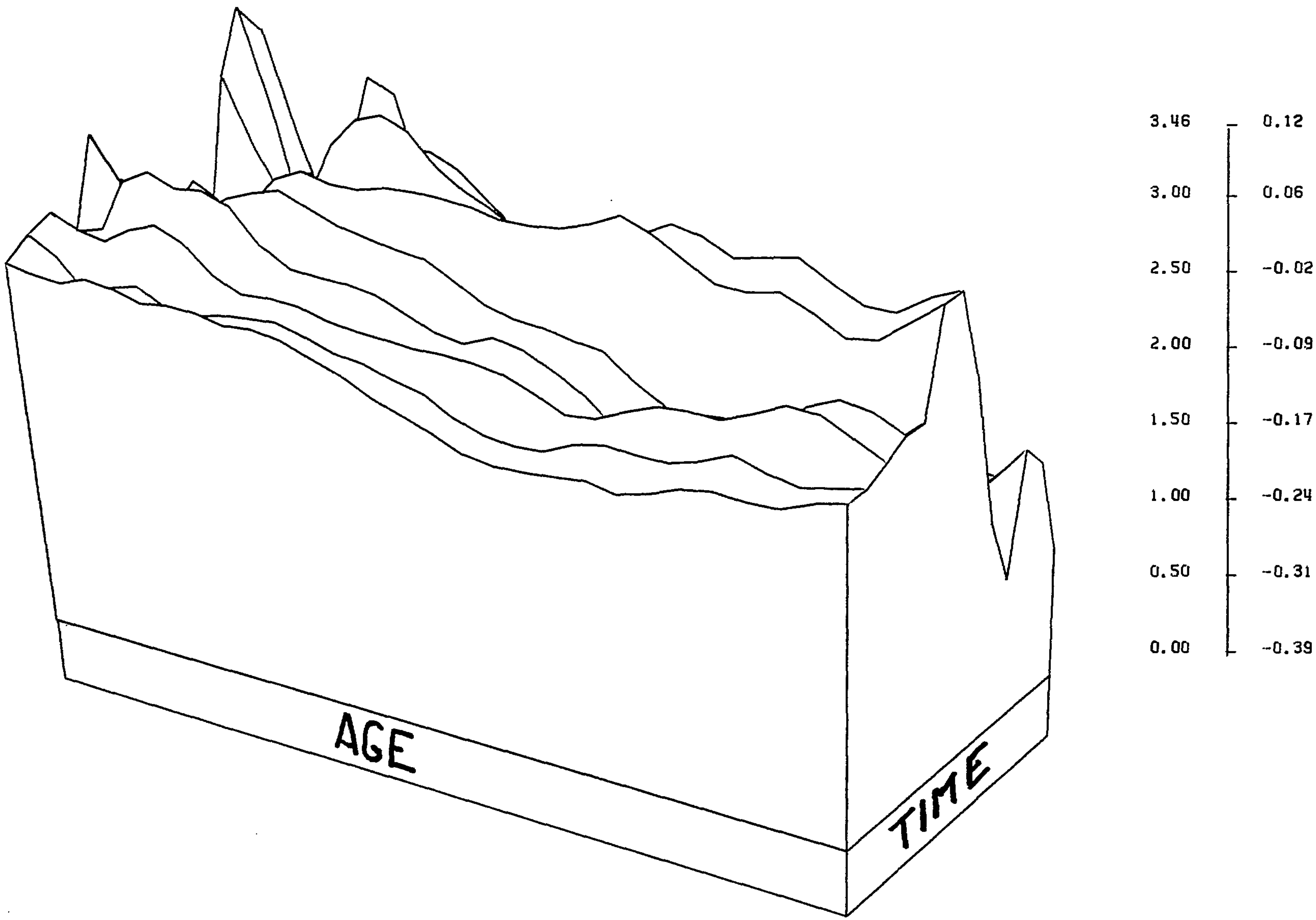
FERTILITY PARAMETERS - AGES 14 - 42 - YEARS 1958 - 1970 - 1 COMPONENT



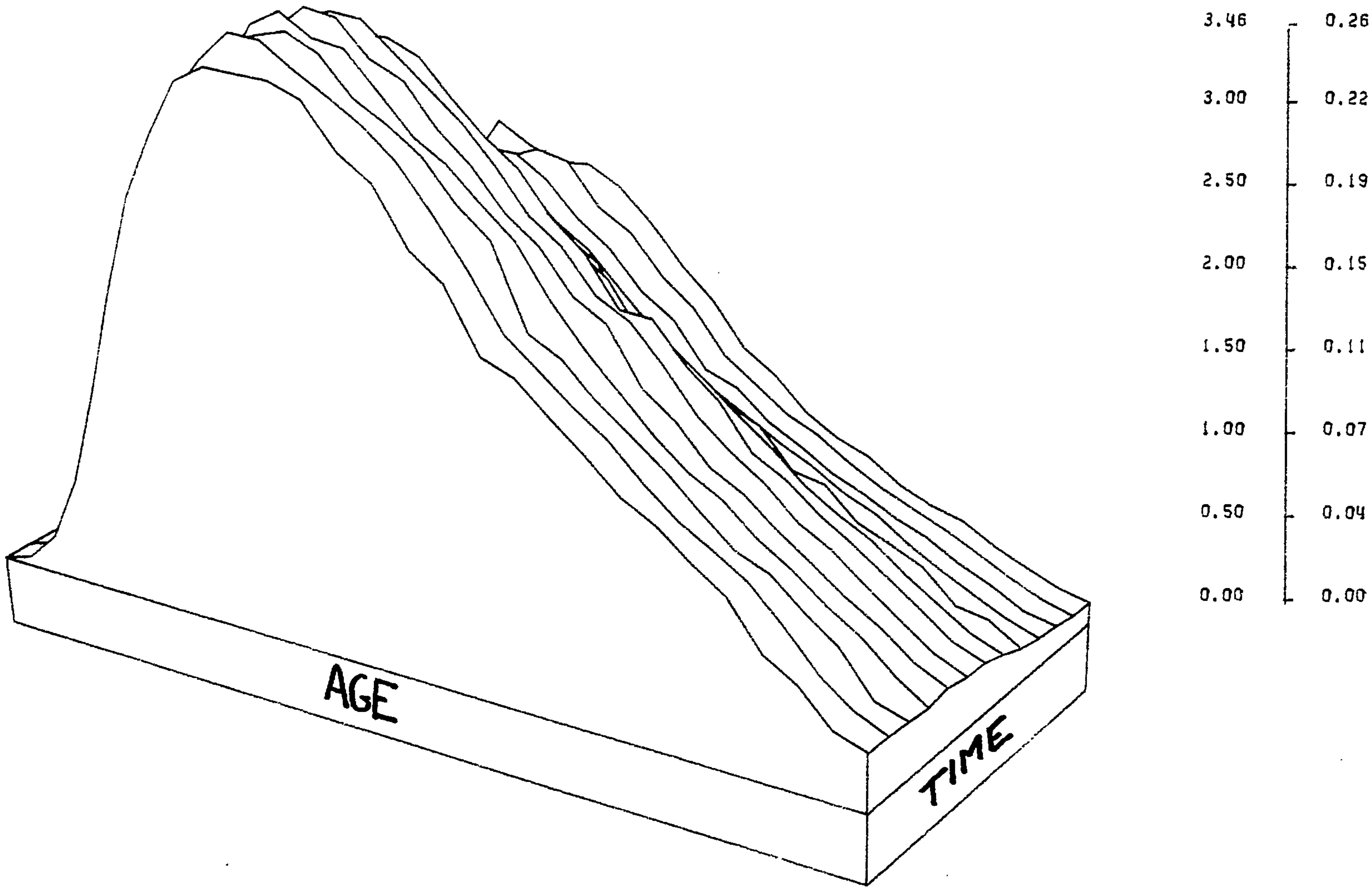
FERTILITY PARAMETERS - AGES 14 - 42 - YEARS 1958 - 1970 - 2 COMPONENTS



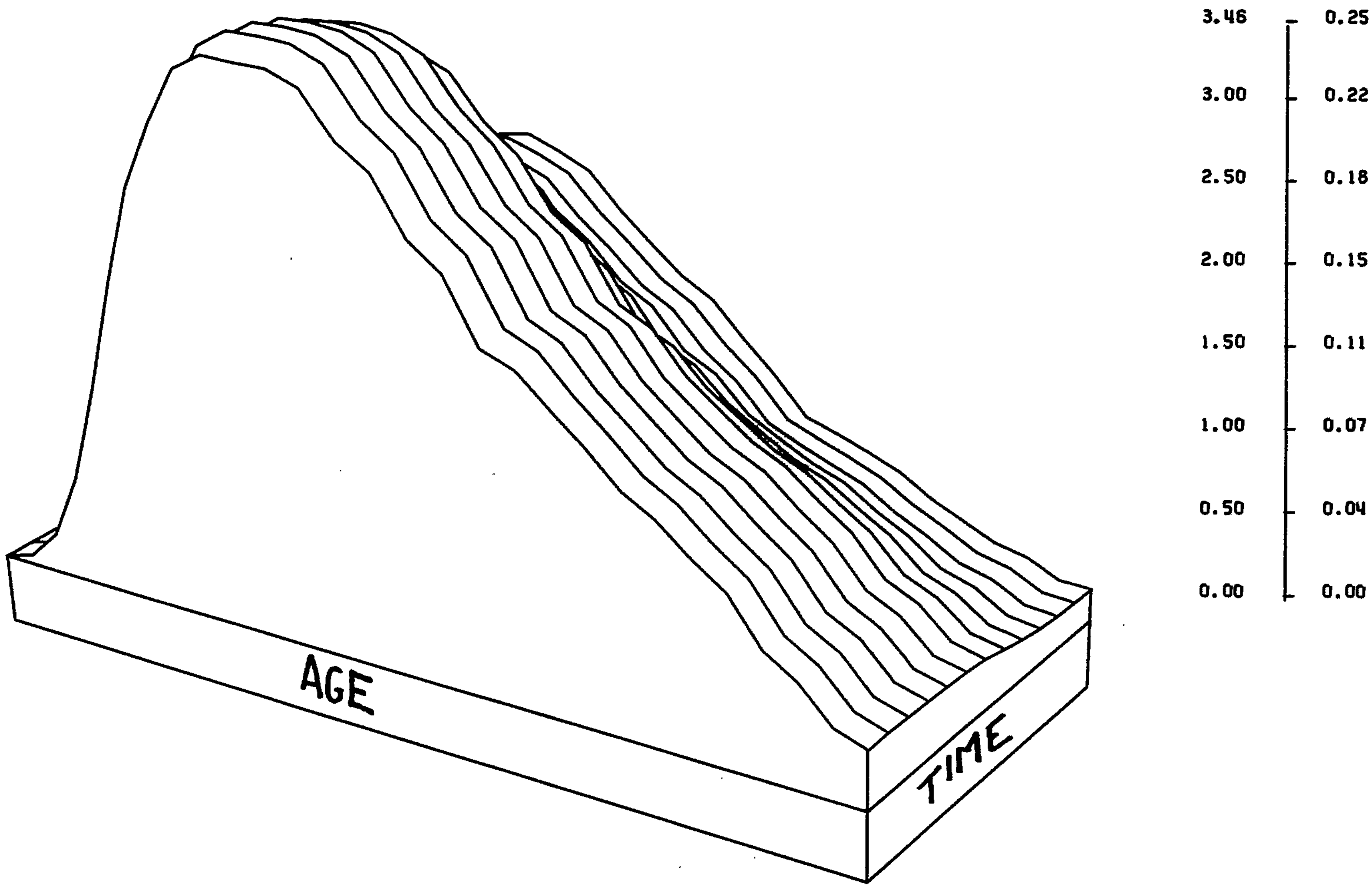
FERTILITY PARAMETERS - AGES 14 - 42 - YEARS 1958 - 1970 - 3 COMPONENTS



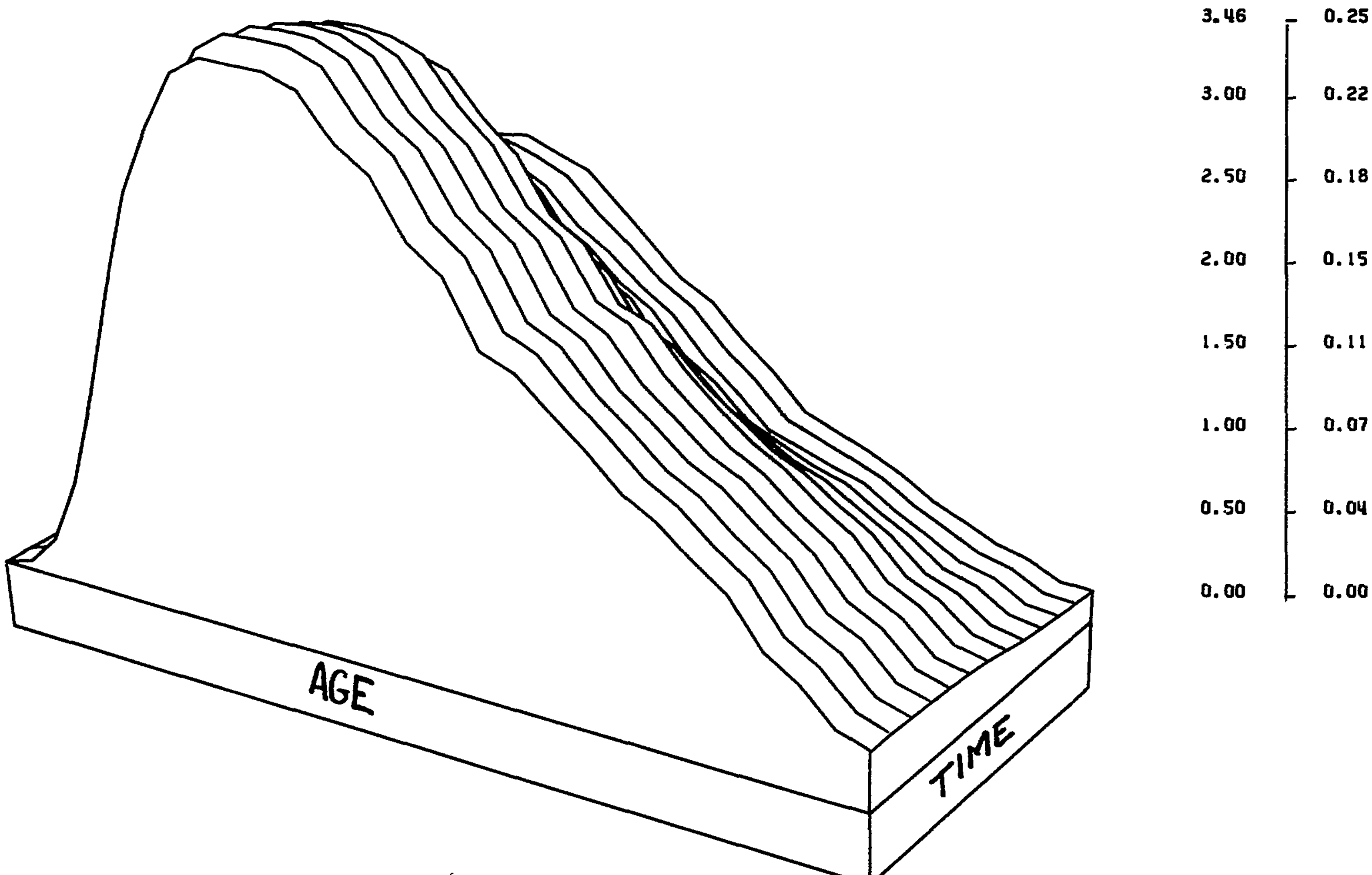
FERTILITY PARAMETERS - AGES 14 - 42 - YEARS 1958 - 1970 - 4 COMPONENTS



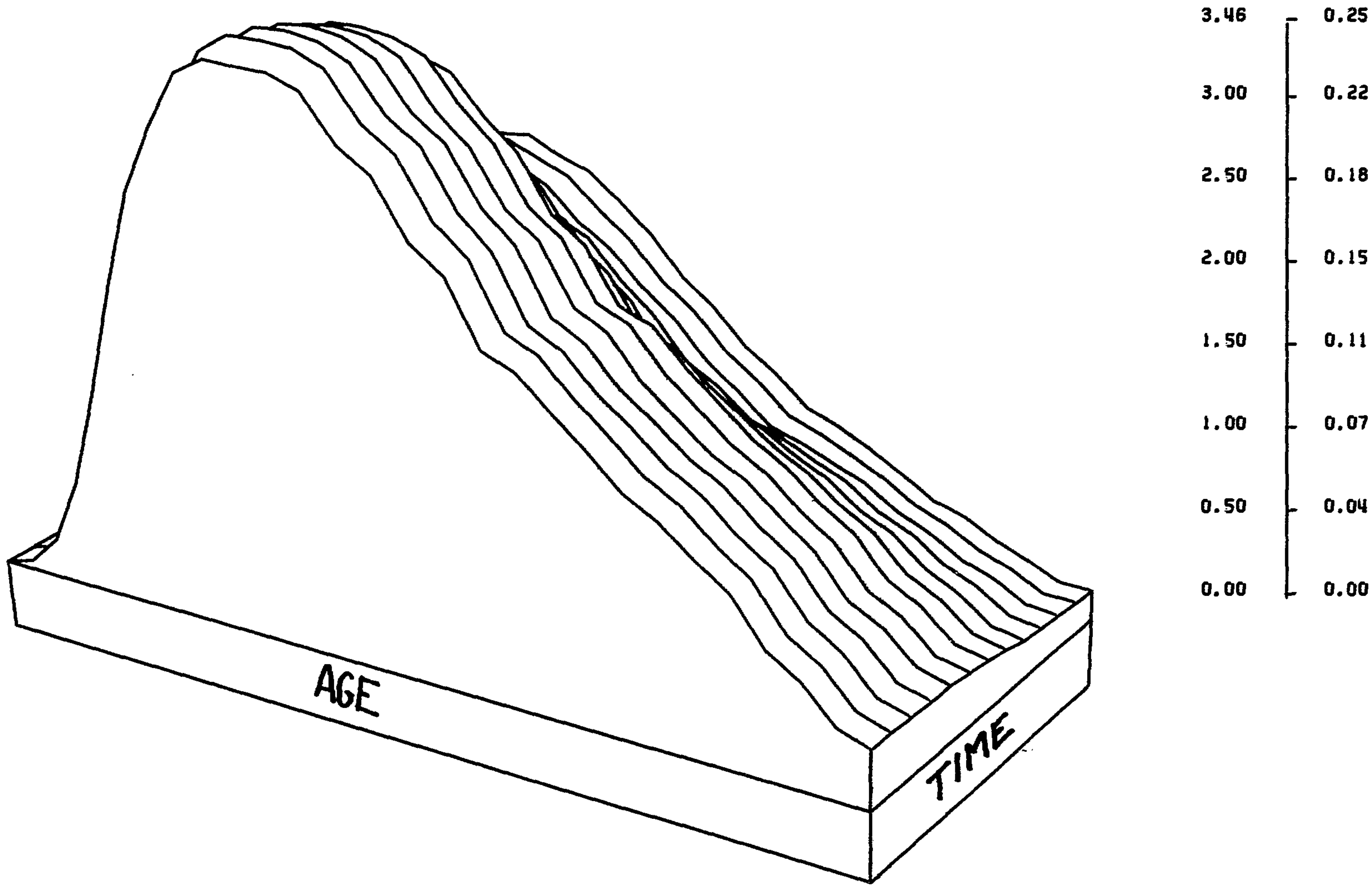
FERTILITY FROM VITAL STATISTICS DATA - 1958 TO 1970



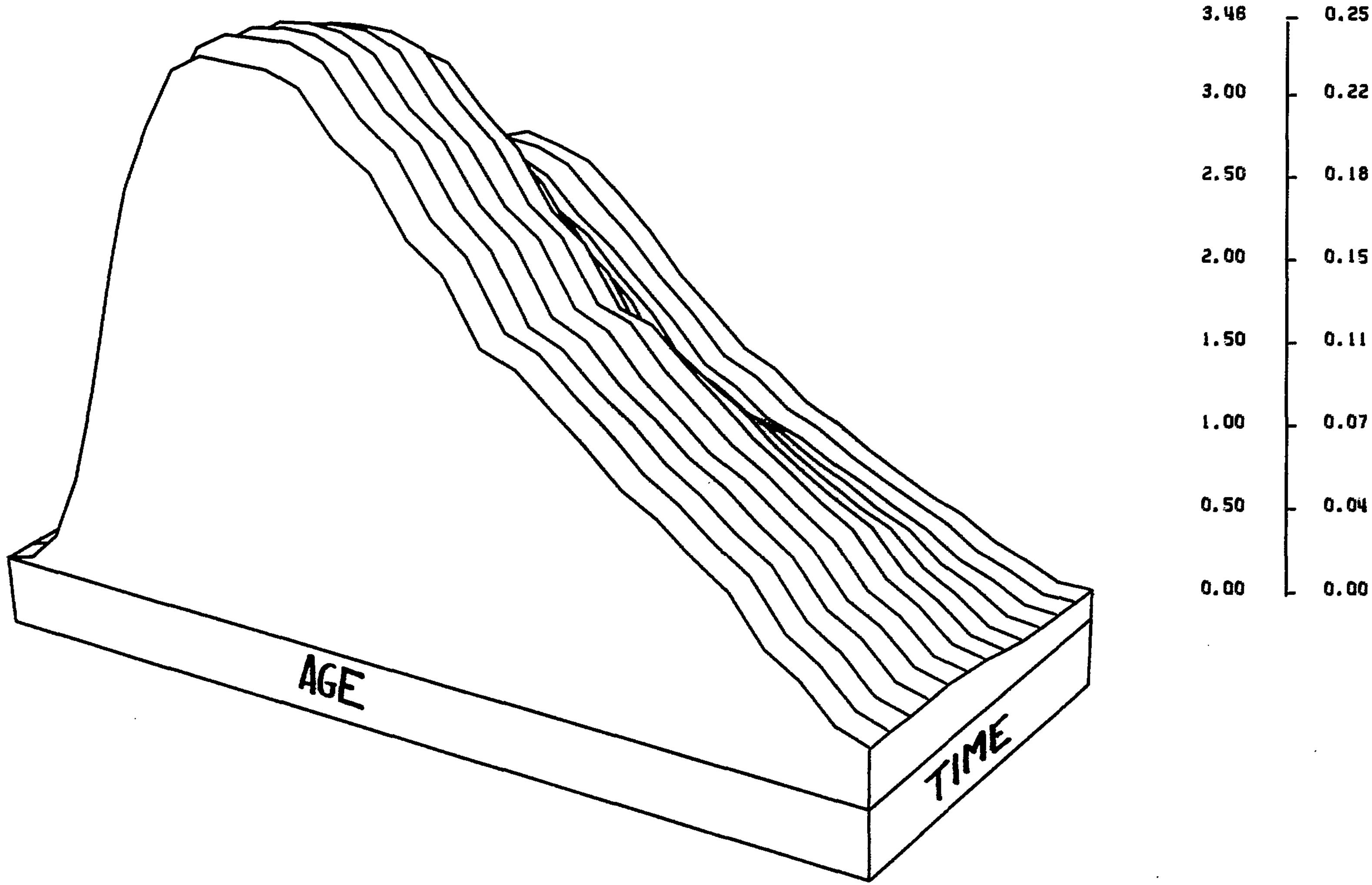
SIMULATED FERTILITY - AGES 14 - 42 - YEARS 1958 - 1970 - 1 COMPONENT



SIMULATED FERTILITY - AGES 14 - 42 - YEARS 1958 - 1970 - 2 COMPONENTS



SIMULATED FERTILITY - AGES 14 - 42 - YEARS 1958 - 1970 - 3 COMPONENTS



SIMULATED FERTILITY - AGES 14 - 42 - YEARS 1958 - 1970 - 4 COMPONENTS

It appears from the simulations that the model provides a good fit for the observed data. It also appears that no significant improvement in the results is gained by using more than one component. This suggests that in fact

$$\frac{\partial f}{\partial t}(x,t) = -\frac{\partial}{\partial x}\left(a_1(t)d_1(x)f(x,t)\right) - b(t)\, f(x,t)$$

provides an adequate model for the behaviour of the fertility curve.

III. Estimation of Economic Mobility

A General Remarks

It seems likely that results relating to the estimation of economic mobility would be of more interest than the results on fertility modelling reported in the previous chapter. For this reason, it is unfortunate that the numerical results we have obtained using available data are in a sense less satisfactory than those of the previous case.

Since the modelling of the income distribution requires use of a distribution over both age and income variables, it is to be expected that the numerical problems associated with the estimation procedure would be somewhat more delicate than in the previous case involving distribution only over age. One may expect the resulting algorithms to be somewhat more sensitive to inaccurate data, and so to require a relatively better data base for a comparable degree of estimation accuracy.

Fairly detailed age-income distribution data seems to be available for recent census years. However, in order to estimate and model the evolution of income distribution over time, data on an annual basis is required. The only annual data we have been able to obtain is that of reported income, as available from the annual reports on taxation provided by Revenue Canada.

Aside from difficulties arising from changes in tax law and reporting procedures, there are other problems associated with this data.

The first problem is that this information has apparently only been collected since 1963. This short length of available data has two effects. One is that the procedure mentioned above involving making sequential estimates on the basis of subsets of the available data is severely hampered, as the length of the available data is sufficient to support only one, or at most two estimates. (Sufficient data was available for eight estimates of the fertility parameters.) The length and uncertainty of the resulting estimates in turn make the prospect of estimating a dynamical model for the mobility parameters very dim indeed.

A second problem bearing on the use of the available data is that of the level of aggregation in the reported data. The data is essentially in histogram form, grouped into five year age brackets, and income brackets of various lengths. This data must be disaggregated on some basis in order to obtain a smooth distribution to which the estimation algorithm may be applied.

Initial runs using the real data produced evidence of distortions arising as a result of the aggregation of the data. This evidence includes an apparent oscillation of period ten years present in some intermediate computed results. This may well be due to the five year age aggregations in the data.

In an attempt to gauge the effect of the aggregation level on the computed results, experiments were run using differing levels of aggregation on simulated distribution data. As reported below, the experimental results indicate that the level of aggregation in the available data is likely to cause severe distortion in the computed estimates; the level of aggregation may well be so high that it is impossible to extract useful information regarding economic mobility from the published data records.

B Aggregation Experiments

In order to study the effects of data aggregation on the estimation procedure, a numerical integration of the model equation

$$\frac{\partial p}{\partial t} = -\frac{\partial p}{\partial x} - \frac{\partial}{\partial s}\left(\sum_{i=1}^{3} a_i(t)di(x_1 s)p\right) - r(x,t)p$$

was carried out (See Appendix E).

It is easy to show using the representation of economic mobility in the form

$$\mu(x,s,t) = \sum_{i=1}^{q} a_i(t)\ d_i(x,s)$$

that if both the time functions $\{a_i(t)\}$ and age-income functions $\{d_i(x,s)\}$ are orthogonal sets over their respective domains then (assuming $\{a_i(t)\}$ orthonormal) the eigenvalues of the symmetric matrix $M^T M$ occurring in the estimation algorithm are just $\{\|d_i(x,s)\|^2\}$, with $\{a_i(t)\}$ being the corresponding eigenvector set. Conversely, if one starts with an arbitrary representation

$$\mu(x,s,t) \quad \sum_{i=1}^{q} \tilde{a}_i(t)\ \tilde{d}_i(x,s)$$

(with no orthogonality conditions), then the result of the estimation algorithm is to produce the orthonormal basis $\{a_i(t)\}$ for the subspace consisting of the span of $\{\tilde{a}_i(t)\}$ which has the property that the resulting $\{d_i(x,s)\}$ in the above representation are <u>also</u> orthogonal. These remarks follow from Appendix C of the interim report [3], together with a bit of elementary linear algebra.

The above facts are mentioned for the reason that they must be kept in mind in the construction of simulation examples. In effect, if one wishes to be able to recognize the components $\{a_i(t)\}$ and $\{d_i(x,s)\}$ in the computed estimates generated from simulation data, these components must be chosen a-priori to be orthogonal. If this is not done, then the estimation algorithm will return the "orthogonalized versions" of the functions involved. (As a matter of computational tactics, one may construct

$$\mu(x,s,t) = \sum_{i=1}^{q} \tilde{a}_i(t)\, \tilde{d}_i(x,s)$$

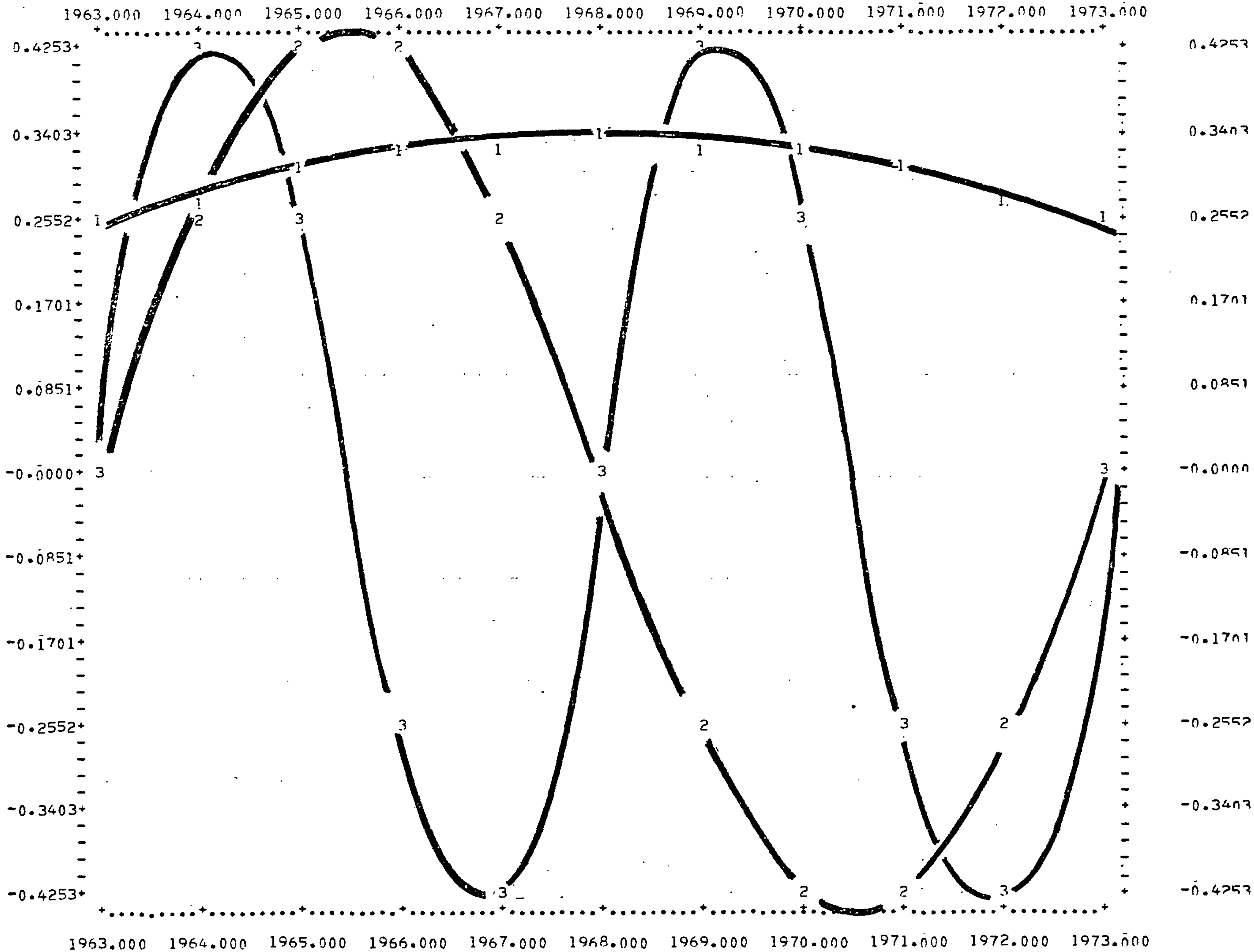
and use program 3.4 of Appendix E as a means of numerically generating the orthogonal components.)

For the simulation used here, $a_1(t)$ was chosen as a parabolic curve, $a_2(t)$ as a sine function of period ten years, and $a_3(t)$ as a sine function of period five years. The $\{d_i(x,s)\}$ were chosen orthogonal, and such that a "reasonable" economic mobility function was produced. It should be noted that the mobility function selected for these experiments is considerably more realistic than the example included as an algorithm test in the interim report [3]. Magnitudes of the mobility in this experiment are considerably larger, and were selected on the basis of preliminary computations utilizing the available real data.

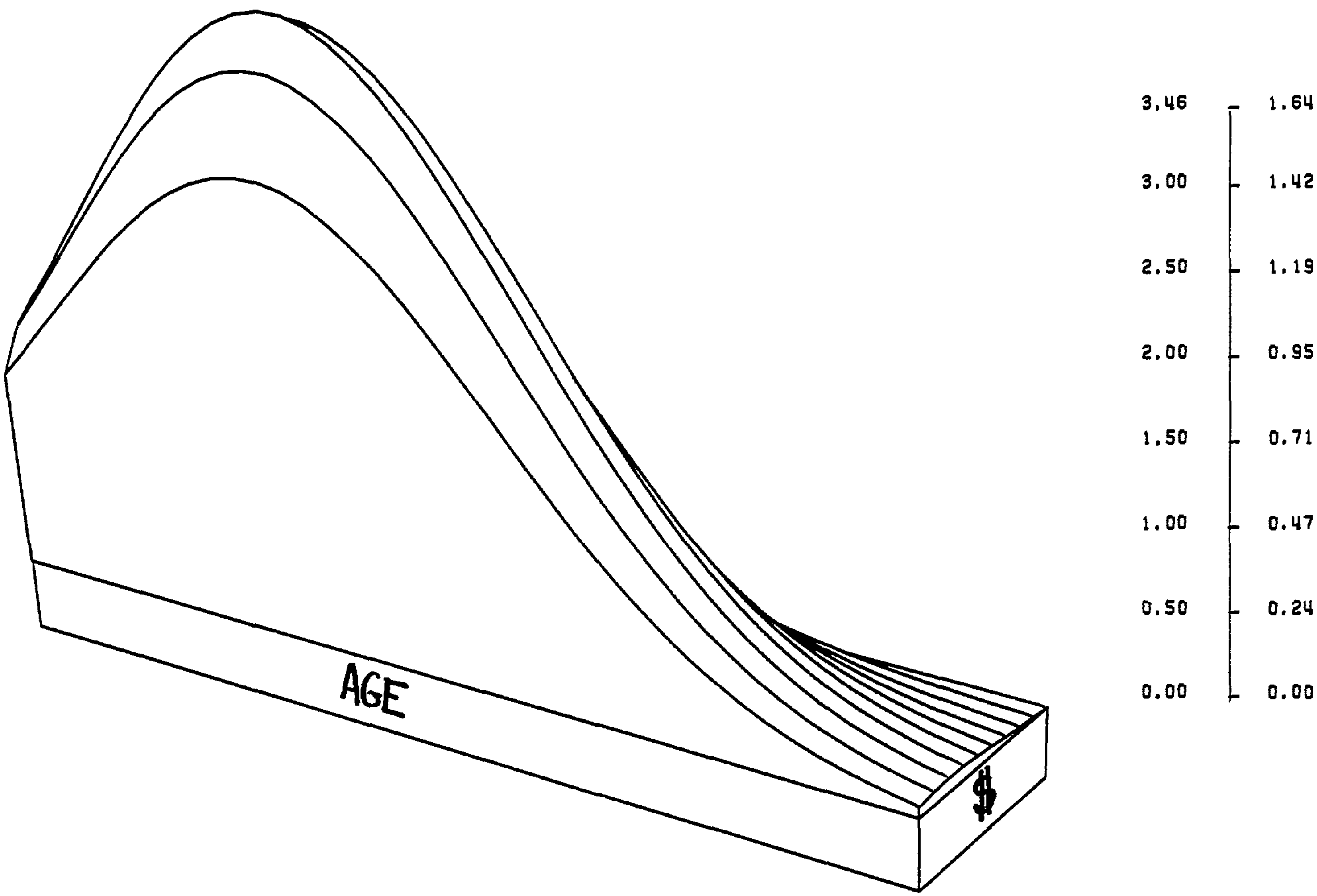
Below are reproduced plots of the time components and age-income components as reproduced from program 3.4

Also shown is a copy of the program output page giving numerical values of the eigenvalues of $M^T M$, and numerical values of the $\{a_i(t)\}$. It is easily seen from this output that the "parabolic" component is by far the dominant effect in the simulated economic mobility.

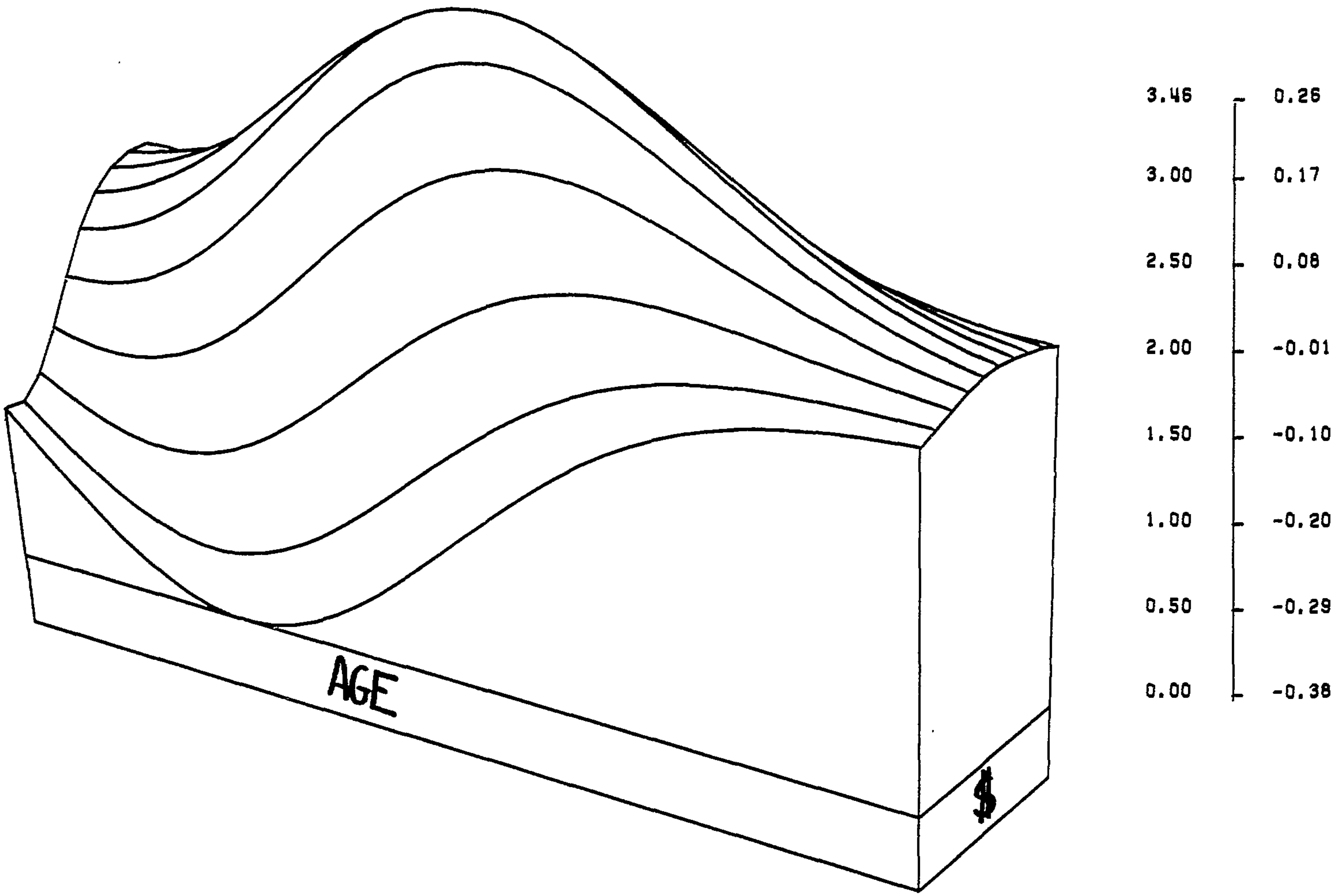
CHART 107



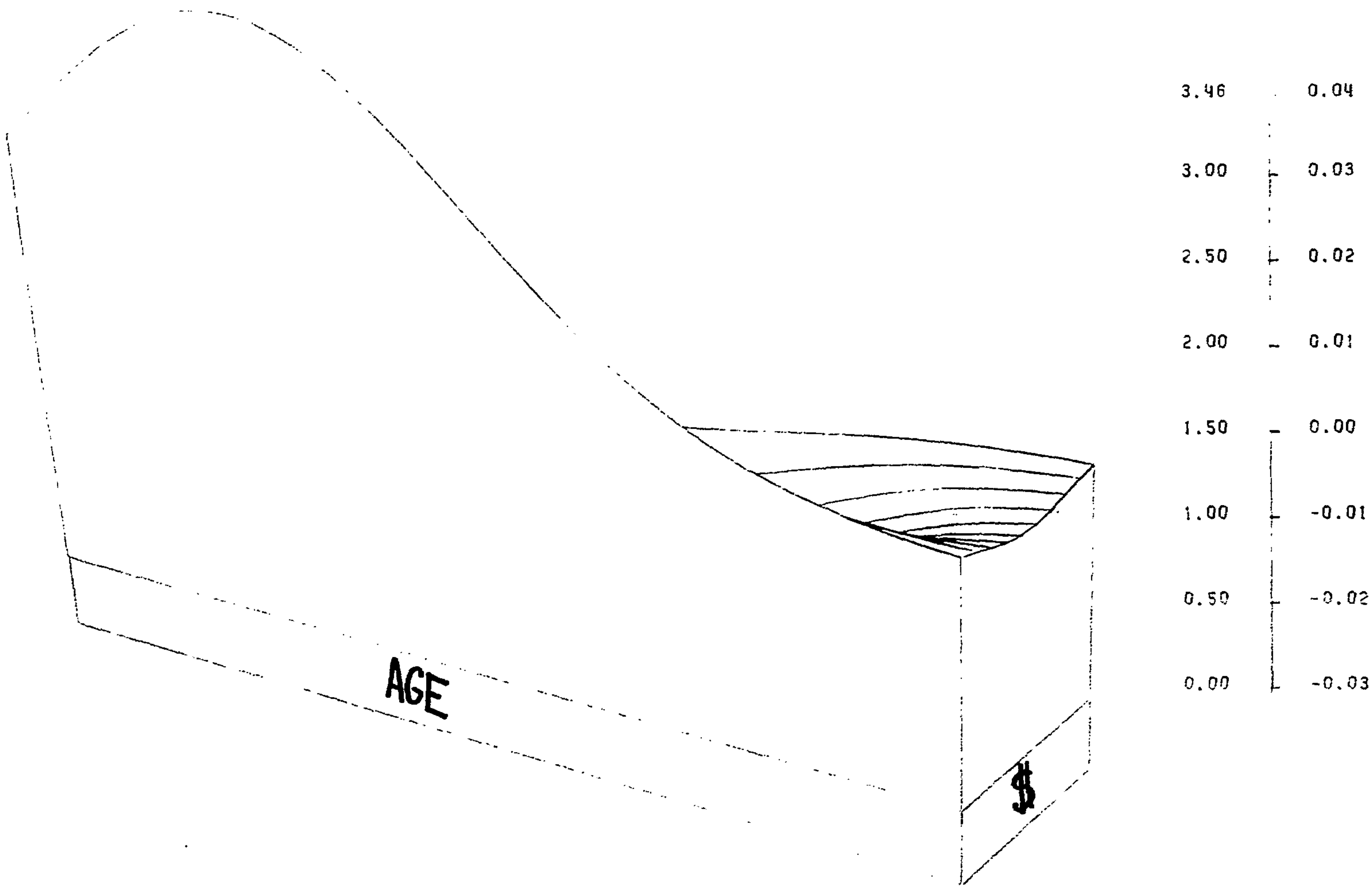
Three Time Components of Simulated Mobility



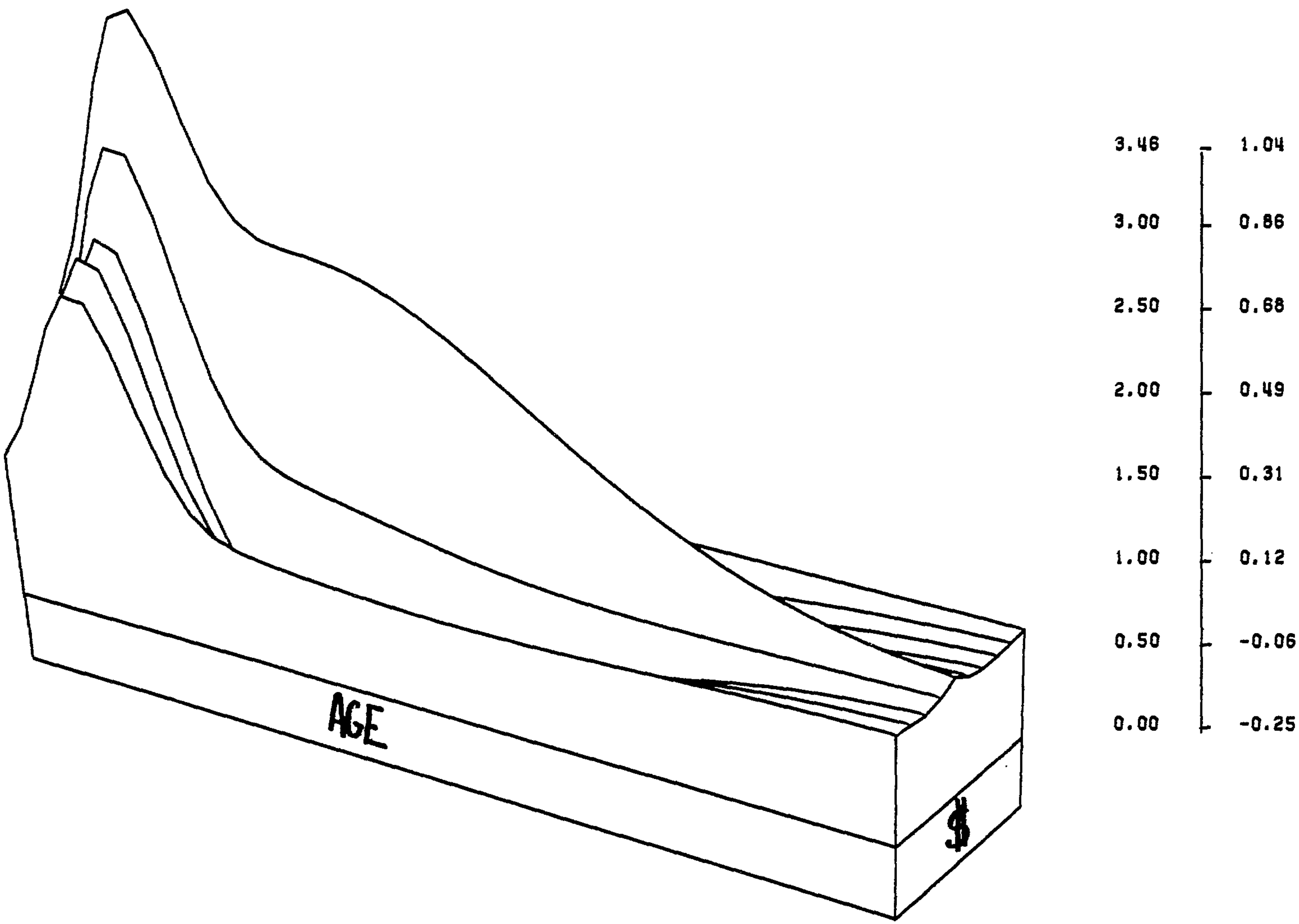
COMPONENTS OF SIMULATED MU : AGES 17 - 56 INCOMES $2,000 - 11,000 COMPONENT 1



COMPONENTS OF SIMULATED MU : AGES 17 - 56 INCOMES $2,000 - 11,000 COMPONENT 2



ORTHOGONAL COMPONENTS OF MU - AGES 17-56 INCOMES $2,000-11,000 COMPONENT 2 × (-1)



COMPONENTS OF SIMULATED MU : AGES 17 - 56 INCOMES $2,000 - 11,000 COMPONENT 3

```
EIGENPAIRS OF MTM(11,11)    BORDERING  0 ROWS AND COLUMNS

            EIGENVALUE =        169.19884
EVT(I, 1)
     0.250058     0.280041     0.303361     0.320029     0.330047
     0.333411     0.330107     0.320121     0.303447     0.280090
     0.250058
            EIGENVALUE =          0.06203
EVT(I, 2)
    -0.000034    -0.263045    -0.425531    -0.425219    -0.262699
    -0.000033     0.262634     0.425169     0.425414     0.263033
    -0.000034
            EIGENVALUE =          0.00271
EVT(I, 3)
    -0.000009     0.425118     0.262600    -0.263042    -0.425332
    -0.000081     0.425686     0.262993    -0.262781    -0.425194
    -0.000009
            EIGENVALUE =          0.00000
EVT(I, 4)
     0.000007    -0.630246     0.738538    -0.229651    -0.031331
    -0.000121     0.022794     0.036240     0.036026     0.022370
     0.000007
            EIGENVALUE =          0.00000
EVT(I, 5)
    -0.000006    -0.175358    -0.110943     0.100749     0.166936
    -0.303490     0.768610    -0.485839    -0.003315    -0.002060
    -0.000006
```

Using the economic mobility function, the equation was integrated numerically. The grid used was fairly coarse (see below), and the result of the computation was values of the simulated population density at one year time intervals, one year age increments, and one thousand dollar income increments.

This simulated data was aggregated to three different levels, referred to as low, intermediate and high. Low aggregation is essentially at the level of the integration grid, while high aggregation corresponds to the age and income brackets used in the real data on reported income obtained from the Revenue Canada publications. The brackets are presented in the chart below. Note that while the density and low aggregation appear to have essentially the same age and income brackets, the data is treated differently. Low aggregation data is generated by integrating the splined simulated density, and then disaggregating using fourth order splines as described in the Appendix to generate a density estimate.

Aggregation Age Income Brackets

Level	Ages	Incomes($1000's)
Density (unaggregated)	17, 18, 19,...	0, 1, 2, 3,...19
Low Aggregation	17, 18, 19,...	0, 1, 2, 3,...18, 20
Intermediate Aggregation	17, 20, 23,...65, 69, 73, 78	0, 1, 2,...18, 20
High Aggregation	17, 25, 30, 35, ...70, 78	0, 2, 3, 4,.... 10, 15, 20

The estimation procedure was run on the simulated data at the four levels of aggregation described in the chart above.

Since the algorithm used is basically similar to that utilized for the estimation of the fertility curve dynamics (Chapter III above), entirely similar considerations arise. In particular, problems arising from errors and inaccuracies on the <u>borders</u> of the regions involved require the use of a "bordering technique" parallel to the one adopted above. In all runs, the data is bordered in the age and income directions so that only data from ages 26 to 48, and incomes \$3000 to \$11,000 are utilized in the estimation. The data in the time direction was bordered by 0, 1, and 2 years (on each end), producing three estimates of the $\{a_i(t)\}$ for each level of aggregation. In addition, surface plots of the age-income components $\{d_i(x,s)\}$ corresponding to the zero bordering $\{a_i(t)\}$ were produced for each level of aggregation.

The estimation algorithm returns (estimates of) the time components $\{a_i(t)\}$ in order of decreasing eigenvalue magnitude. This should correspond to decreasing $\|d_i(x,s)\|^2$ in the simulated mobility. The following summary charts give the "position of appearance" of the parabolic and single frequency sine functions as a function of aggregation level and time-bordering.

Position of Parabolic Mode

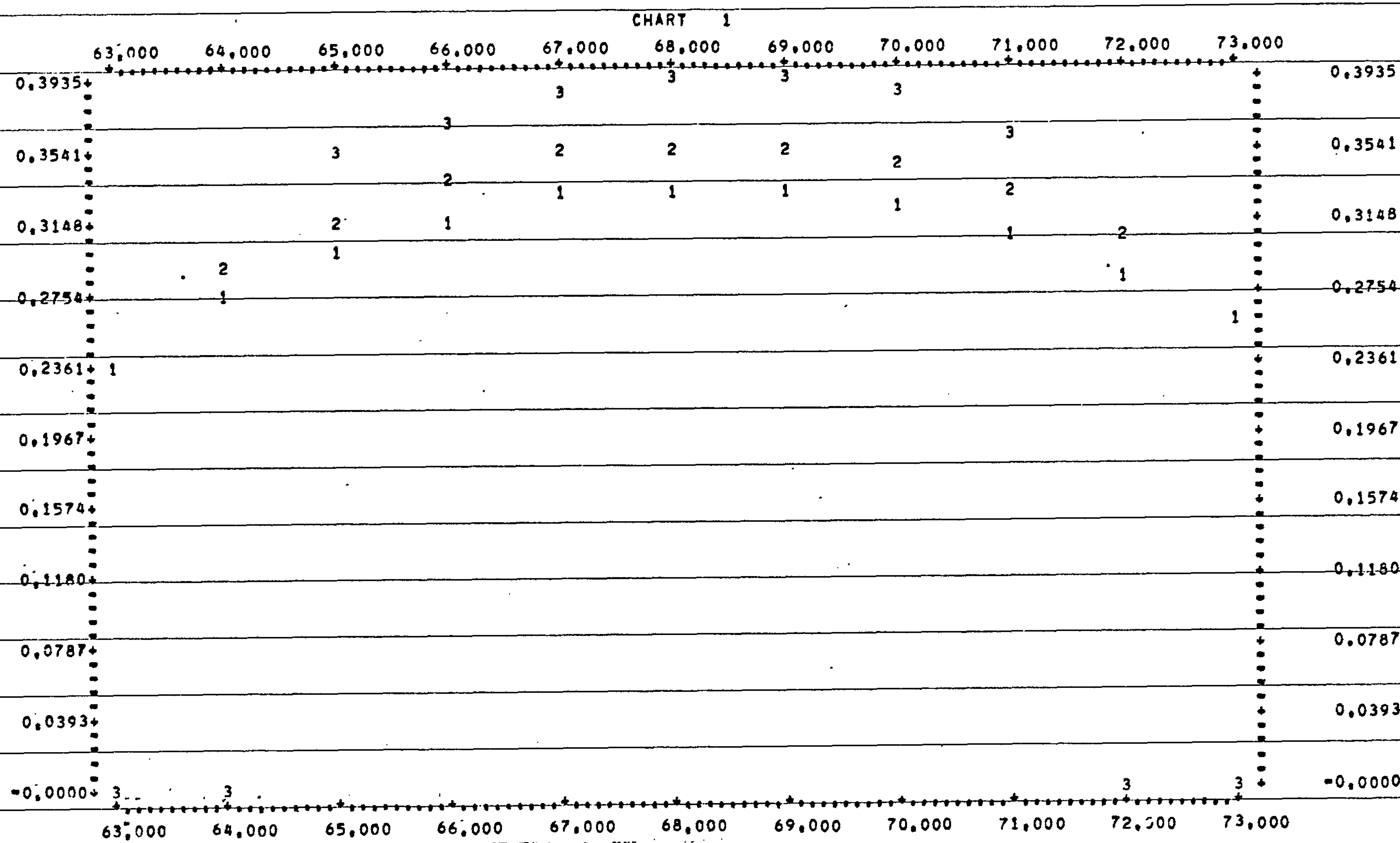
Aggregation	Bordering Years		
	0	1	2
density	1	1	1
low	1	1	1
intermediate	1	1	1
high	2*	?	?

* distorted with no bordering, deteriorates with bordering.

Position of Sinusoidal Mode

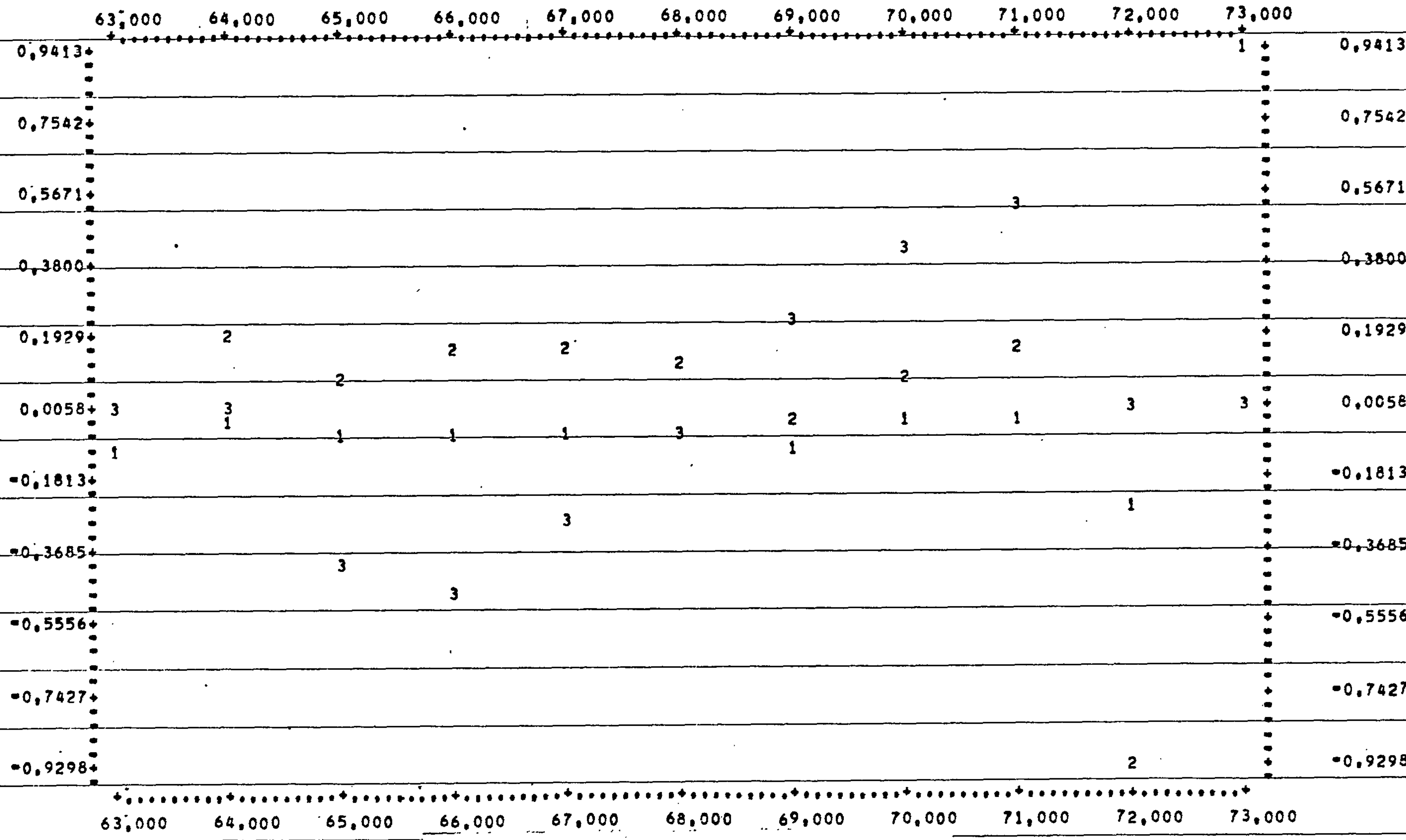
Aggregation	Bordering Years		
	0	1	2
density	3*	3*	2
low	3*	3*	2
intermediate	4	3	2
high	3?	?	?

The following plots consist of the estimated $\{a_i(t)\}$ for each aggregation level, and the corresponding age-income components $\{d_i(x,s)\}$ as computed from the estimation algorithm. These plots should be compared with the original components of the economic mobility used in the simulation.



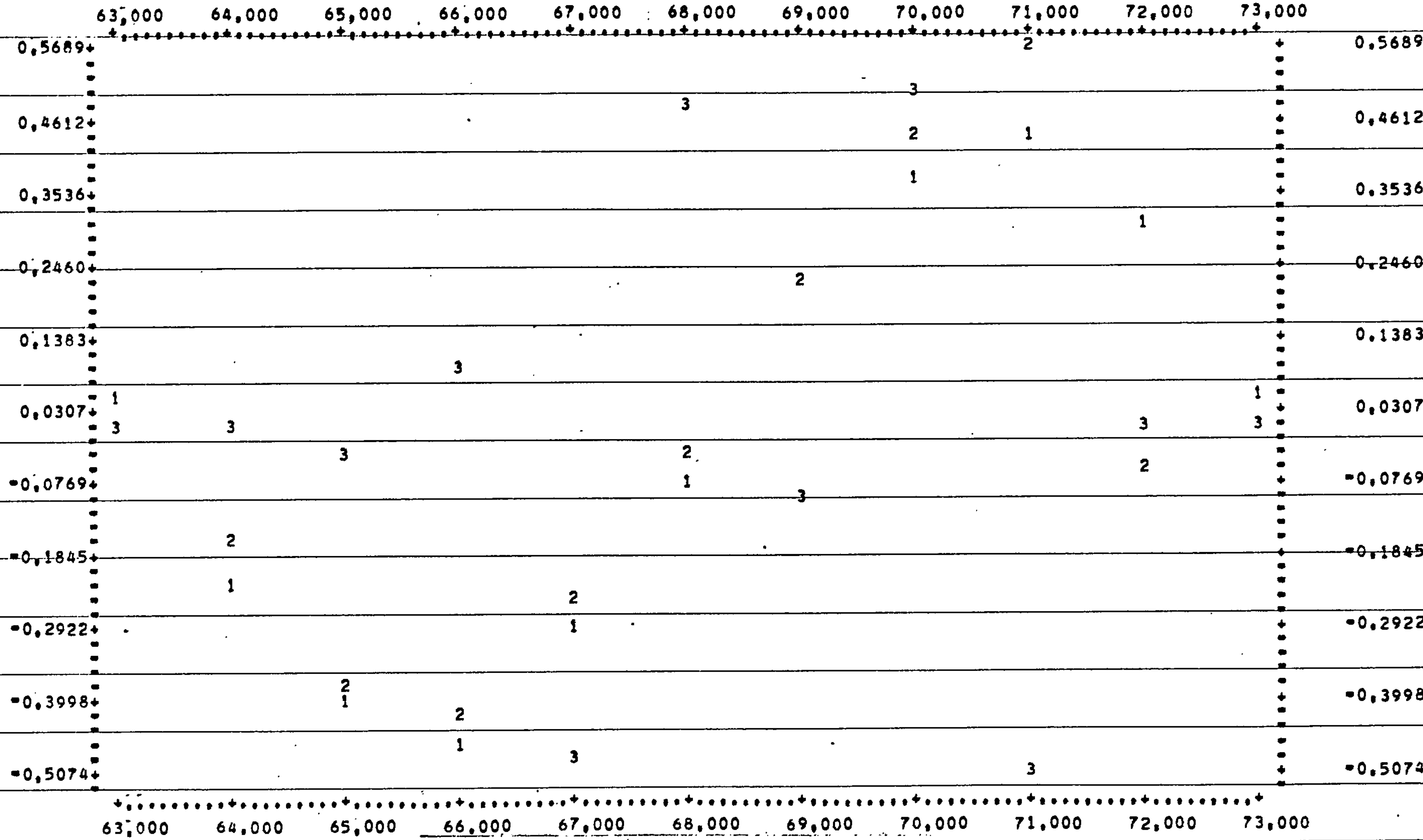
$a_1(t)$ - First Component from Density Data

CHART 2



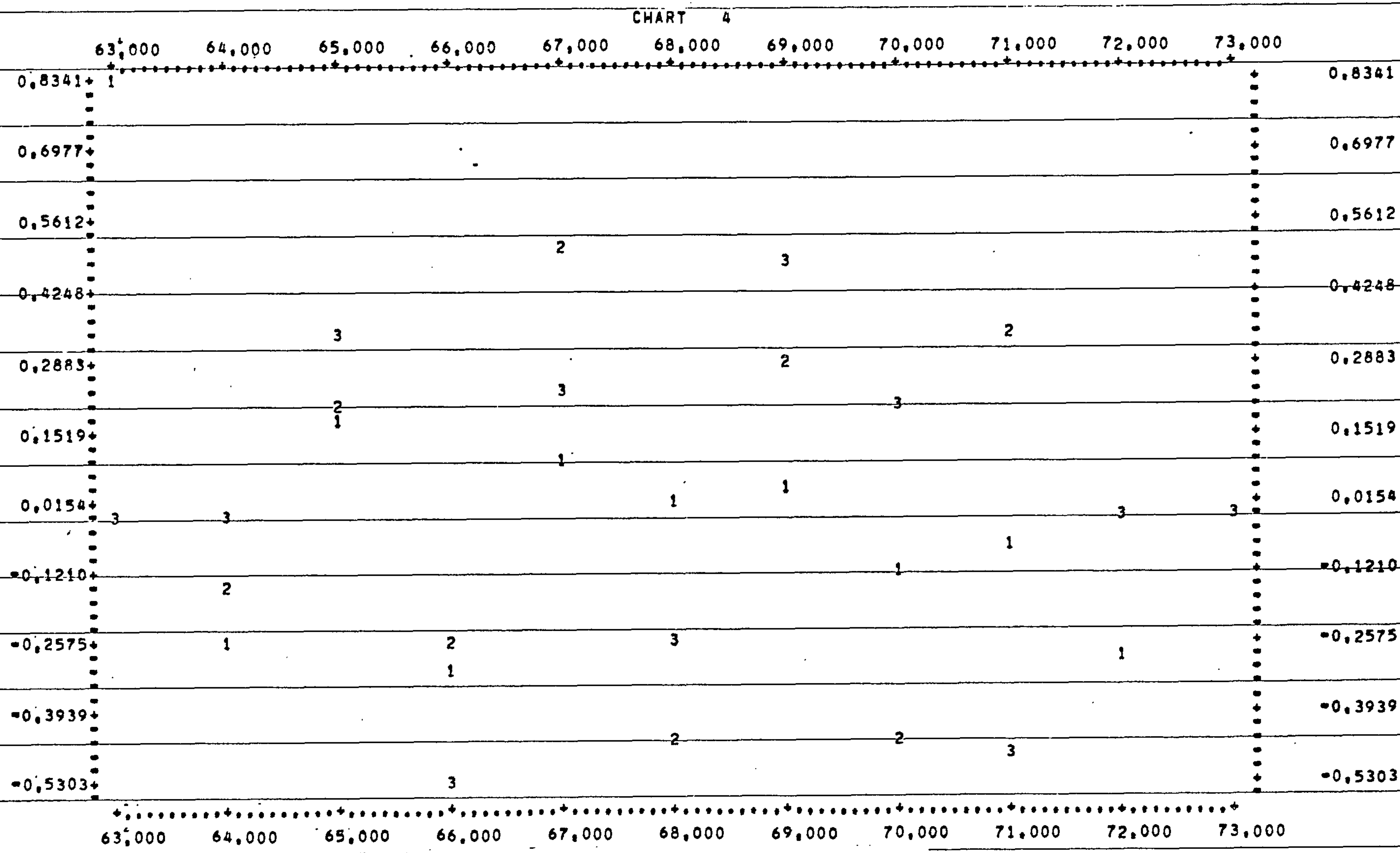
$a_2(t)$ - Second Component from Density Data

CHART 3



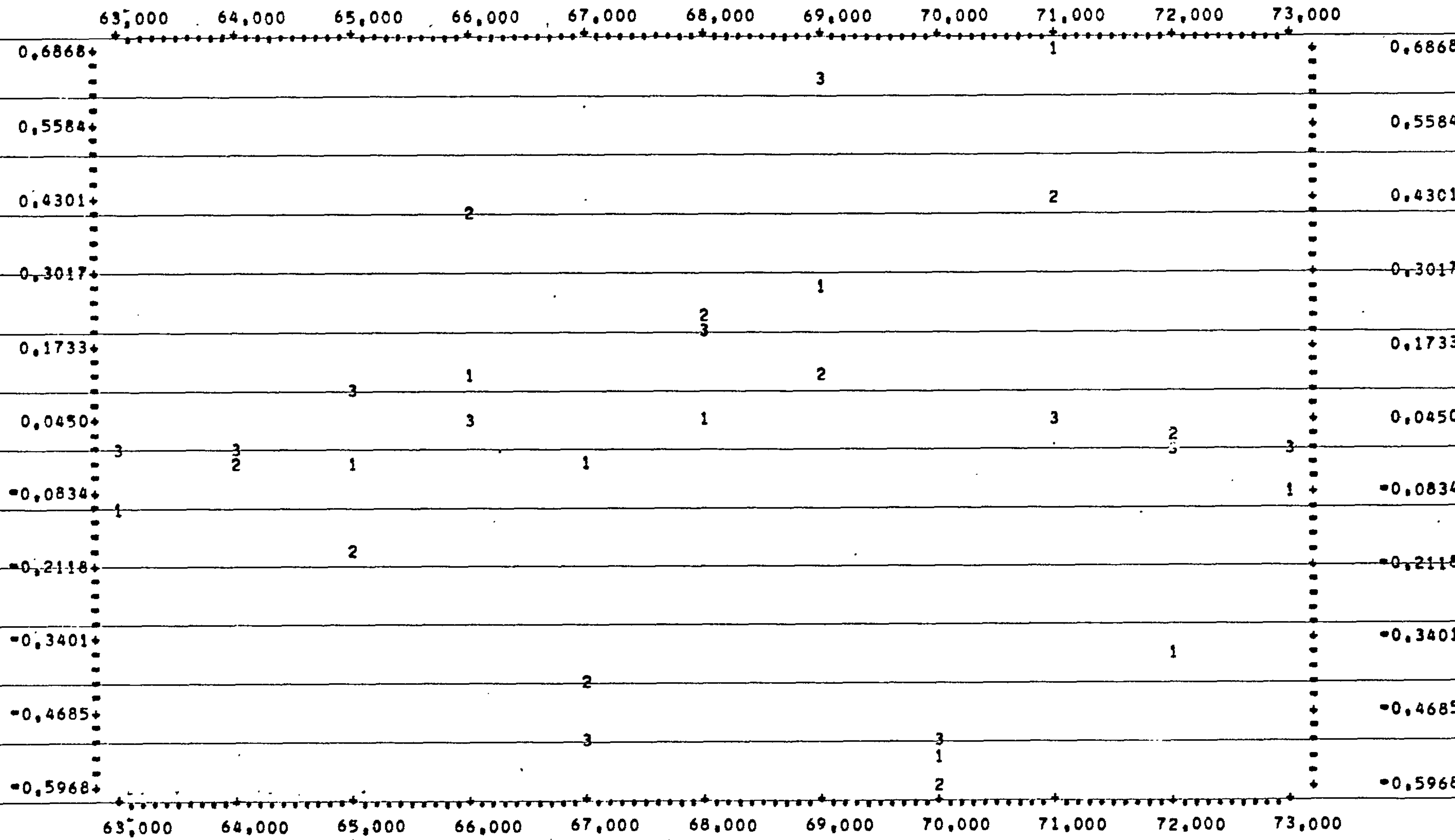
$a_3(t)$ - Third Component from Density Data

-55-

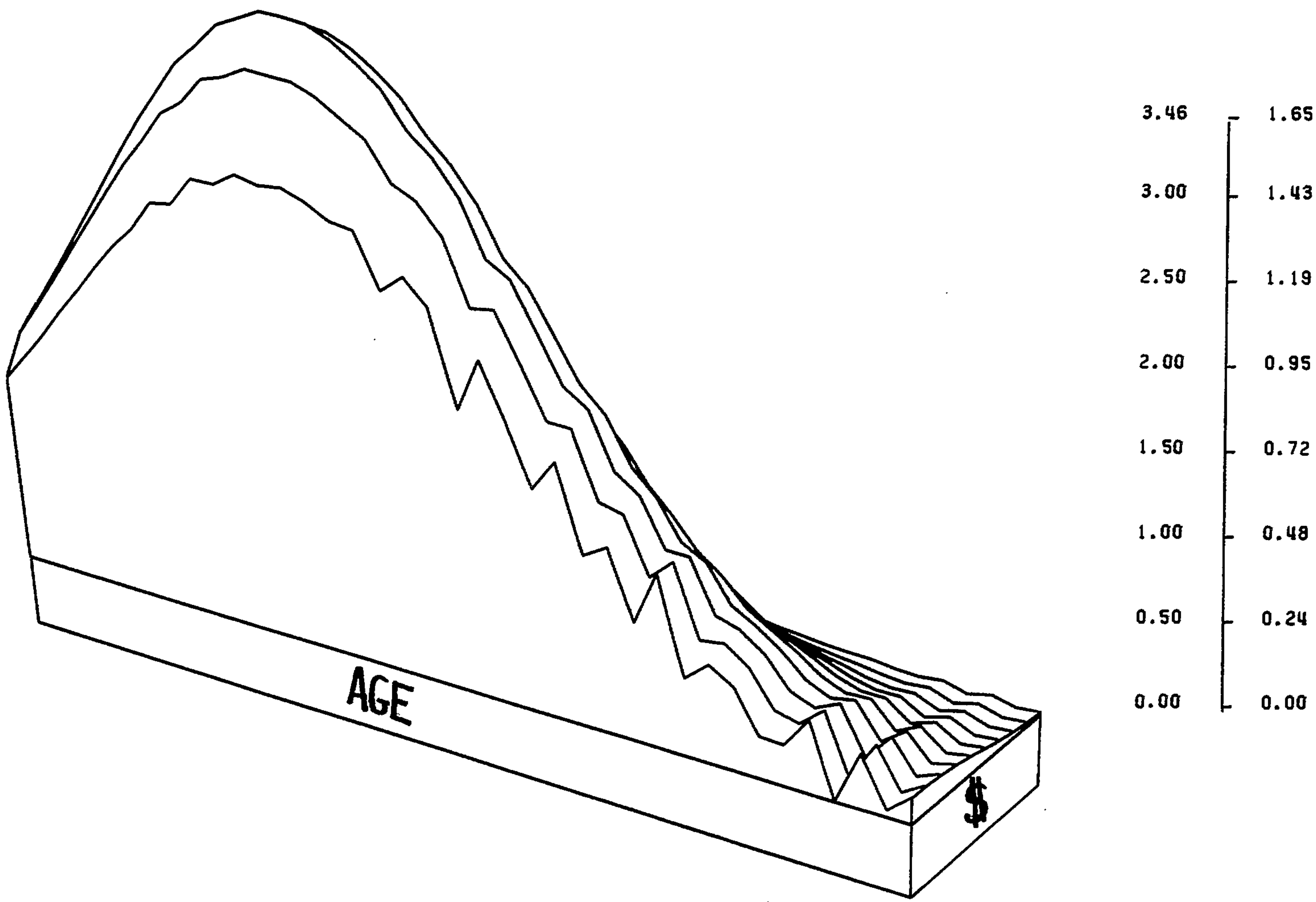


$a_4(t)$ - Fourth Component from Density Data

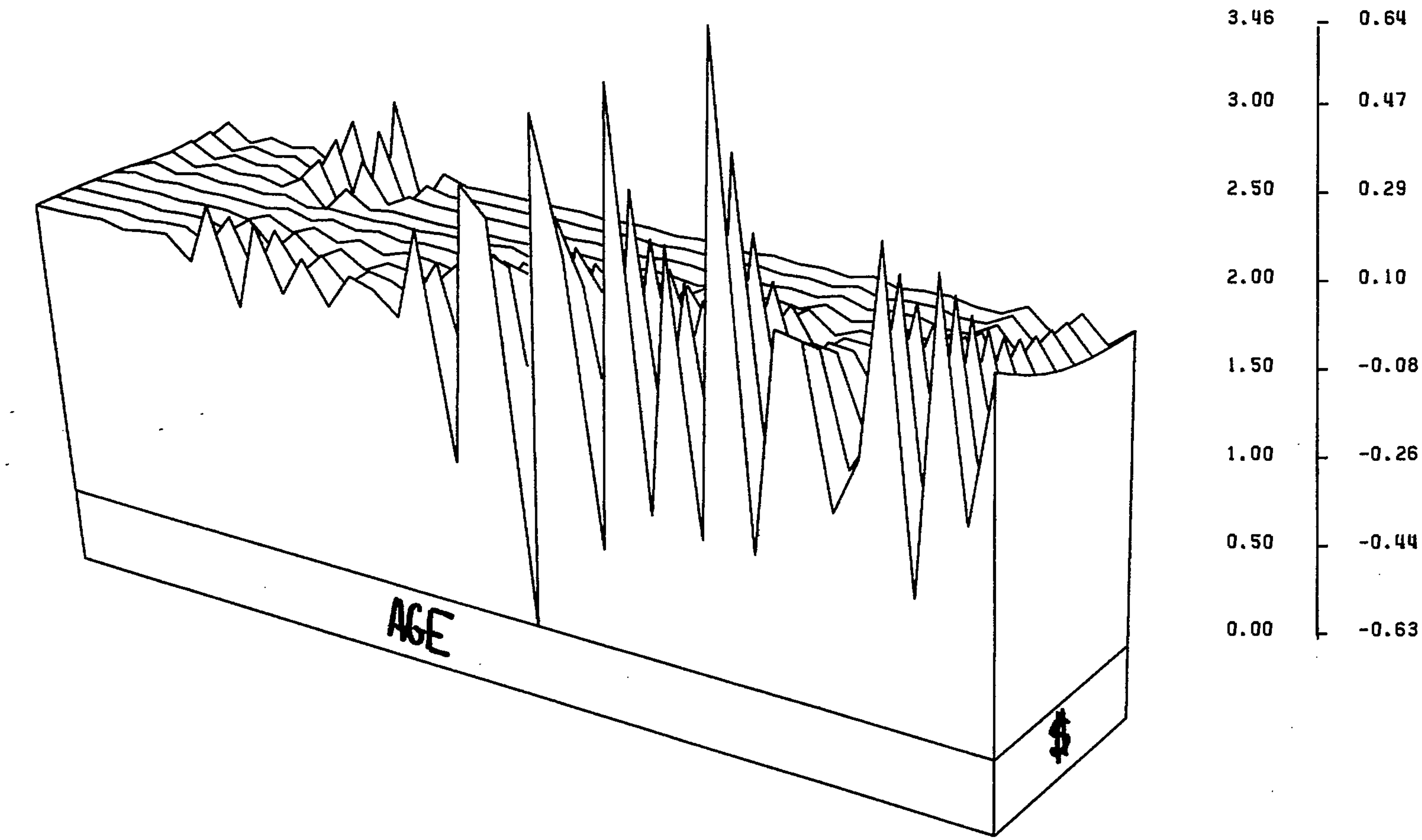
CHART 5



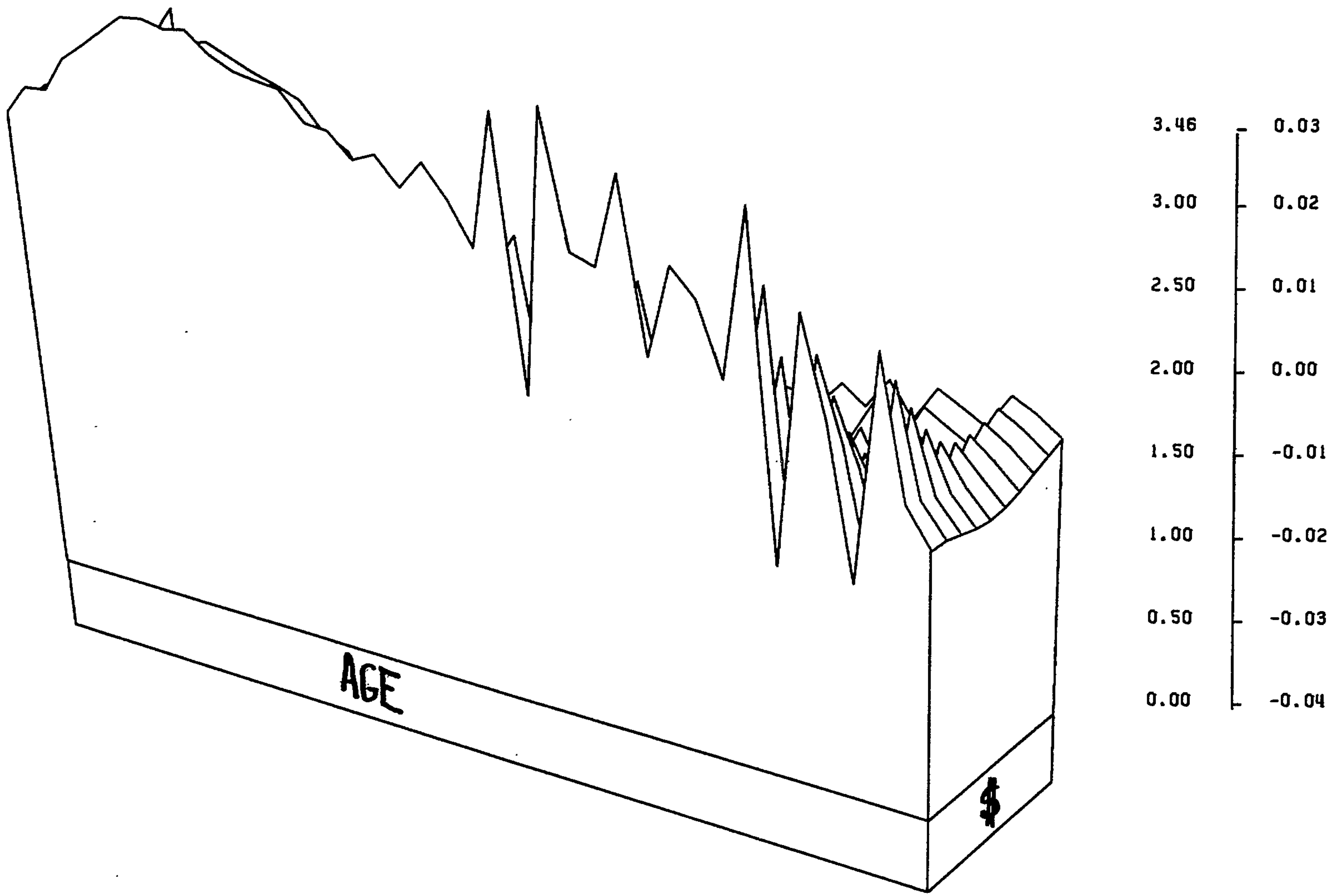
$a_5(t)$ - Fifth Component from Density



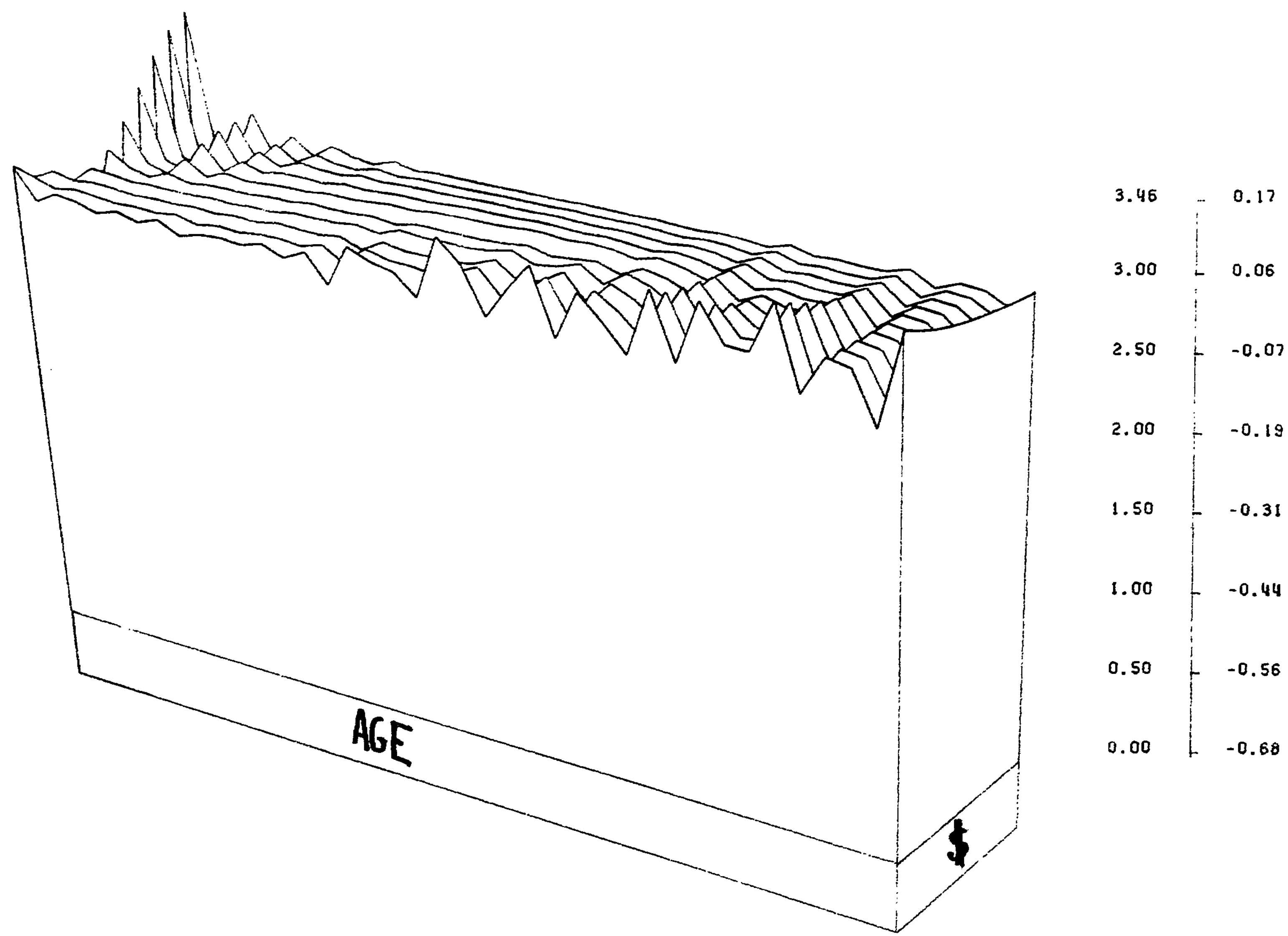
COMPONENTS OF MU ESTIMATED FROM DENSITY COMPONENT 1



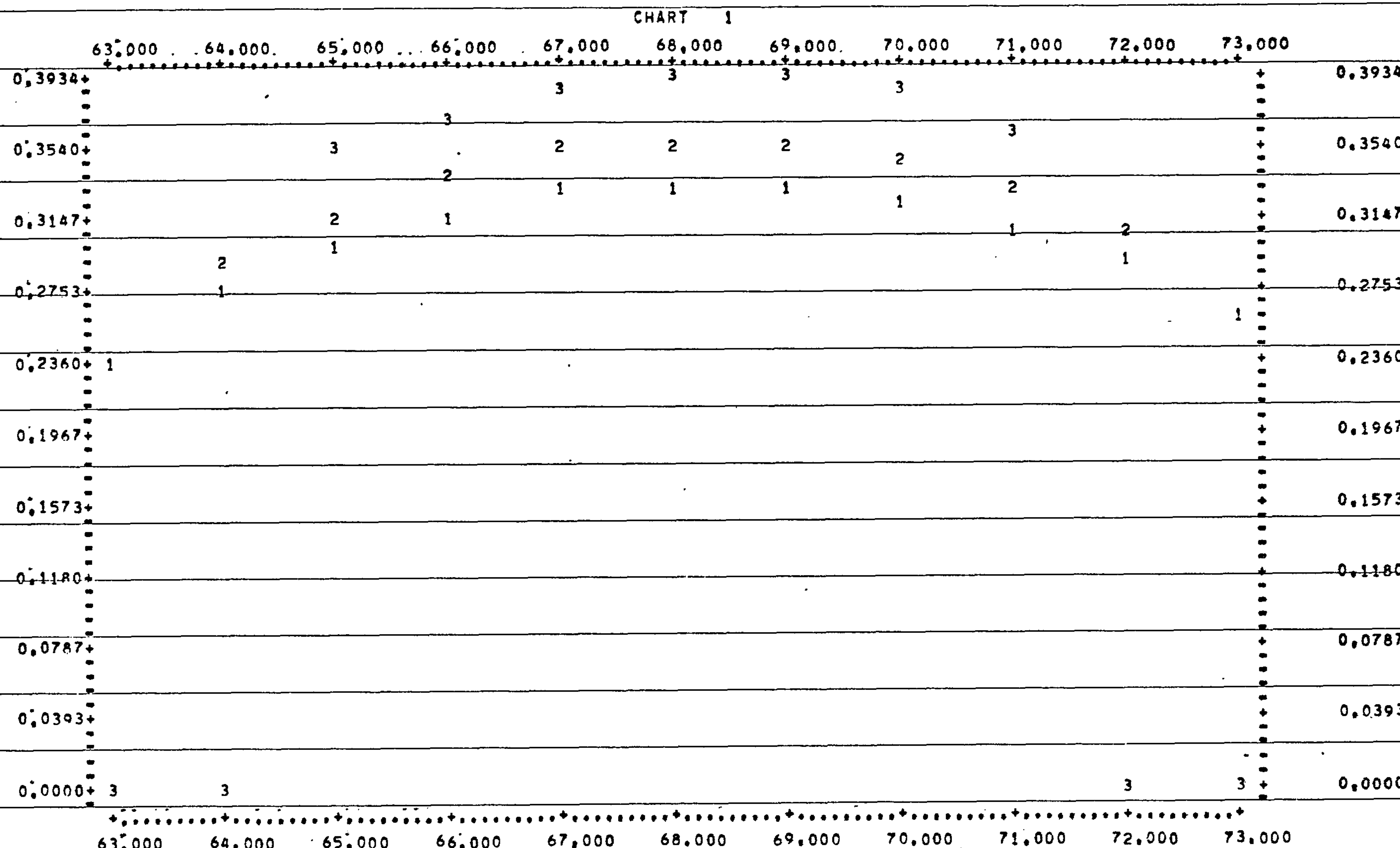
COMPONENTS OF MU ESTIMATED FROM DENSITY COMPONENT 2



COMPONENTS OF MU ESTIMATED FROM DENSITY COMPONENT 3

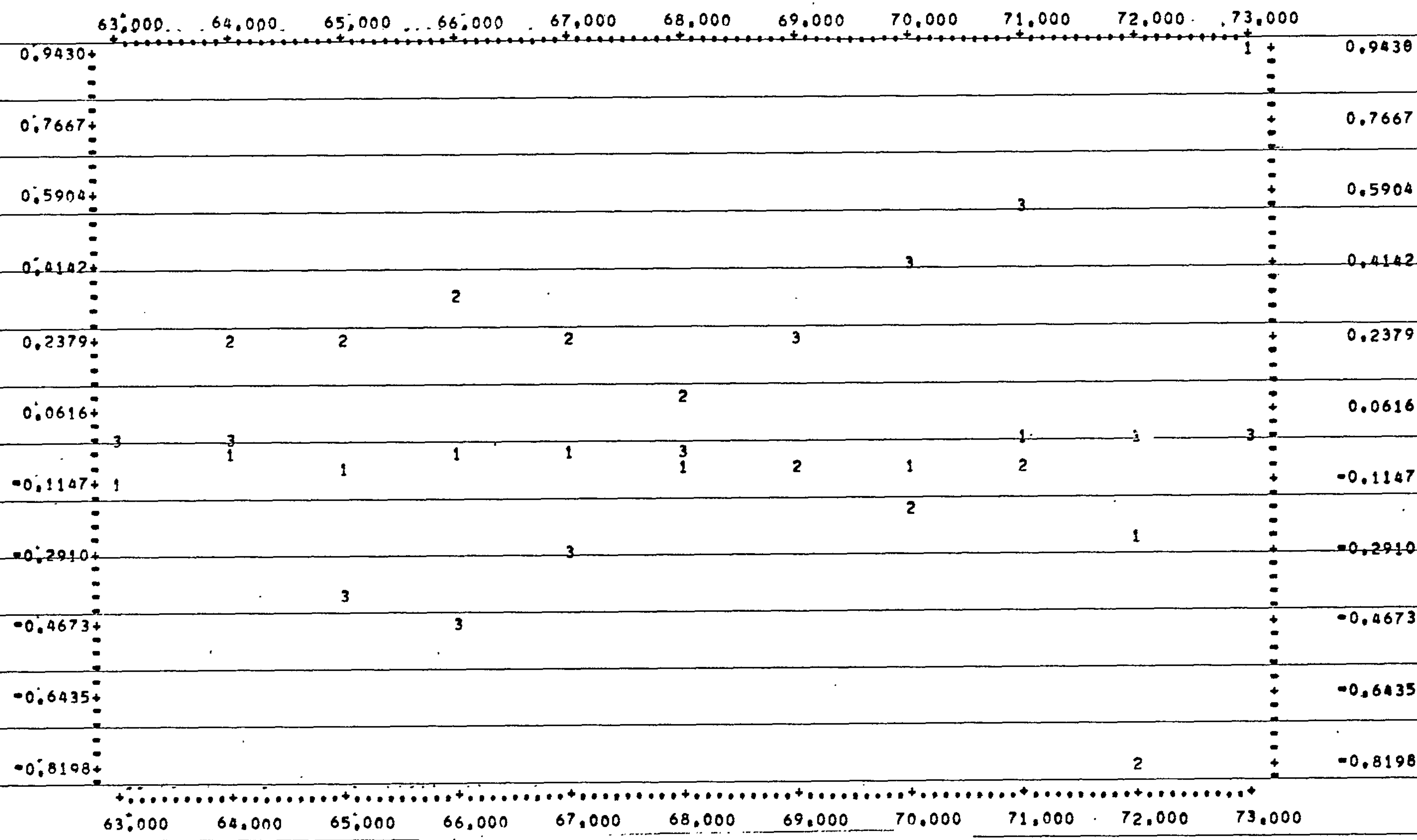


COMPONENTS OF MU FROM DENSITY WITH NO AGGREGATION - COMPONENT 4

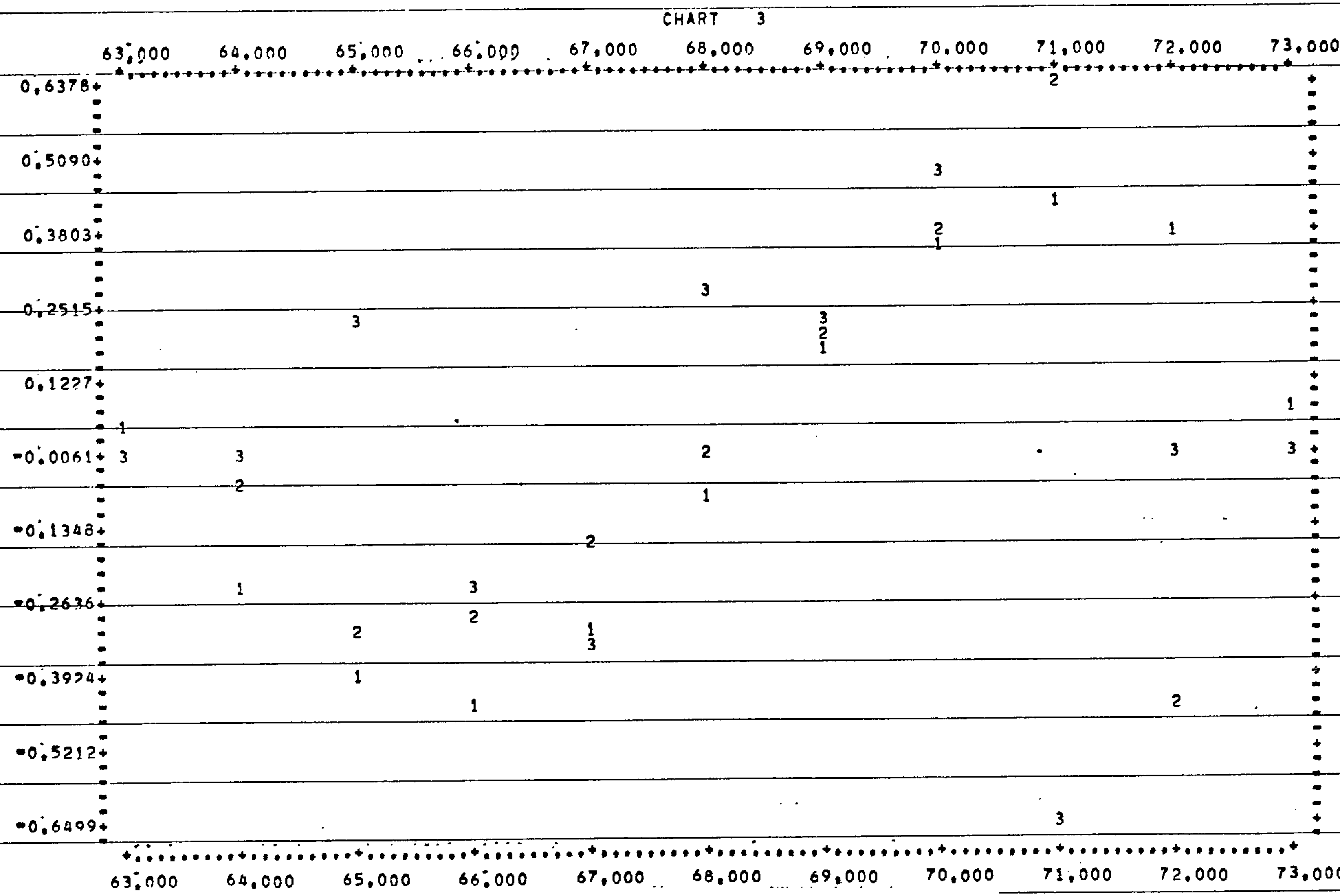


$a_1(t)$ - First Component from Low Aggregation

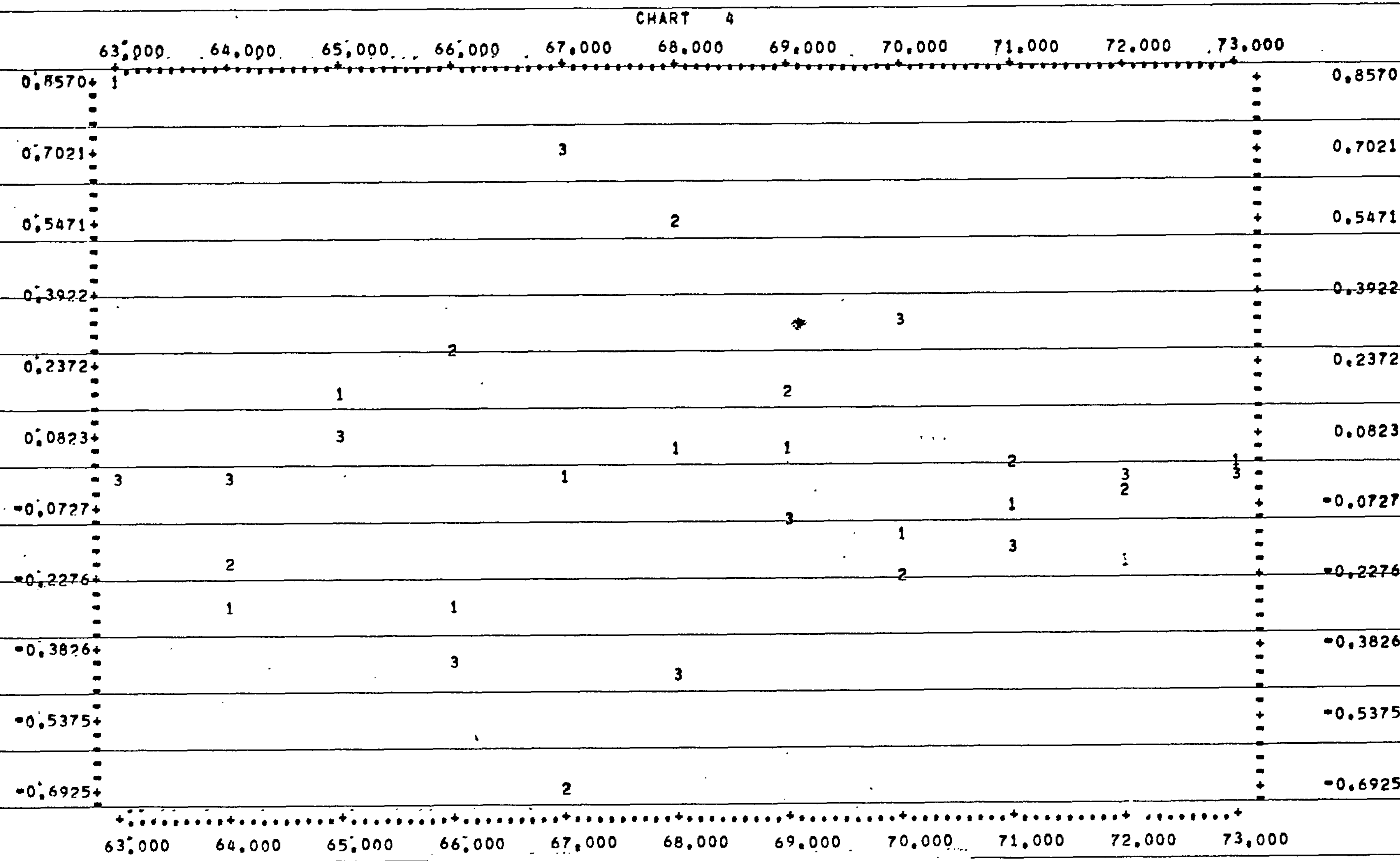
CHART 2



$a_2(t)$ - Second Component from Low Aggregation

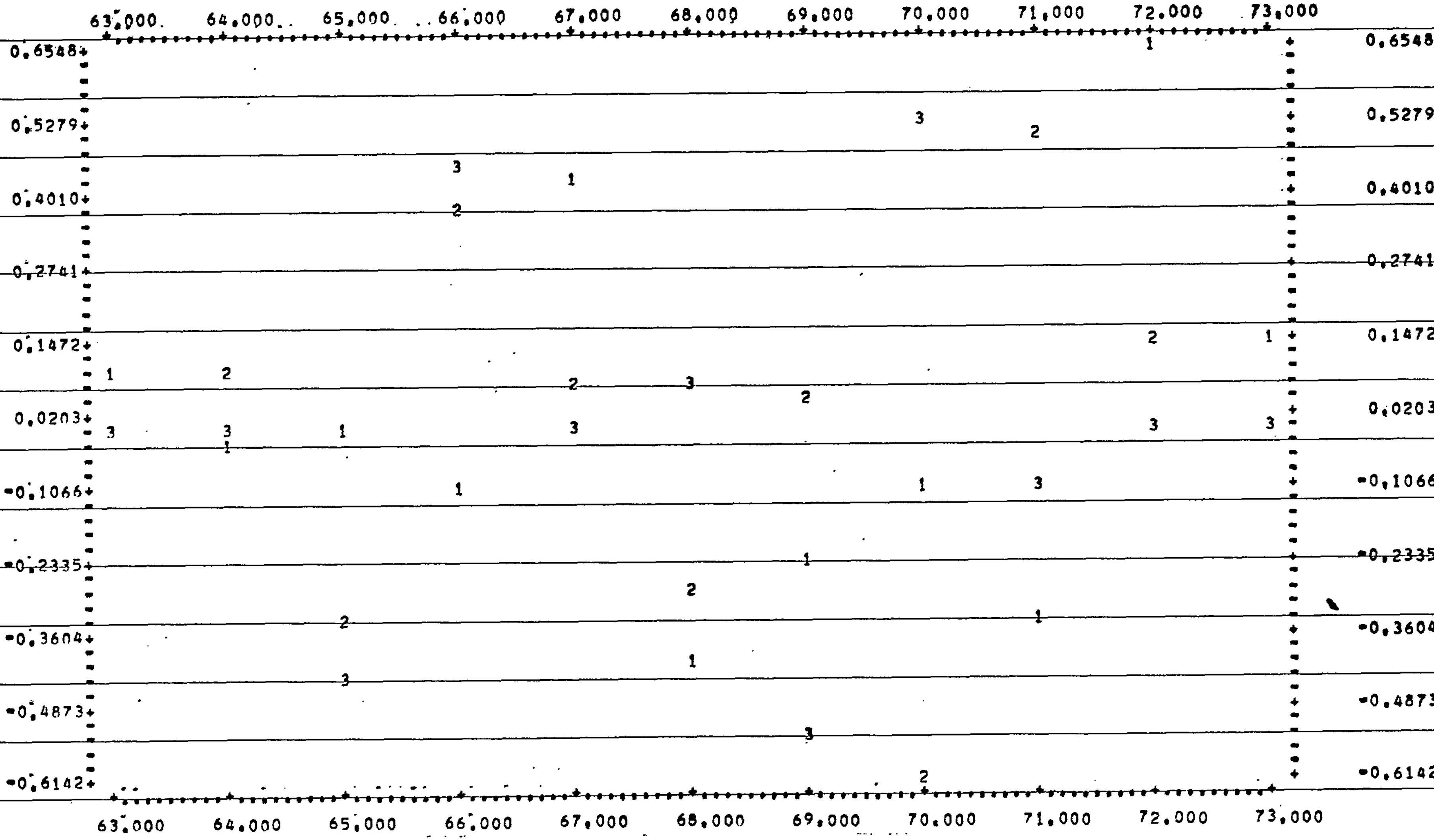


$a_3(t)$ - Third Component from Low Aggregation

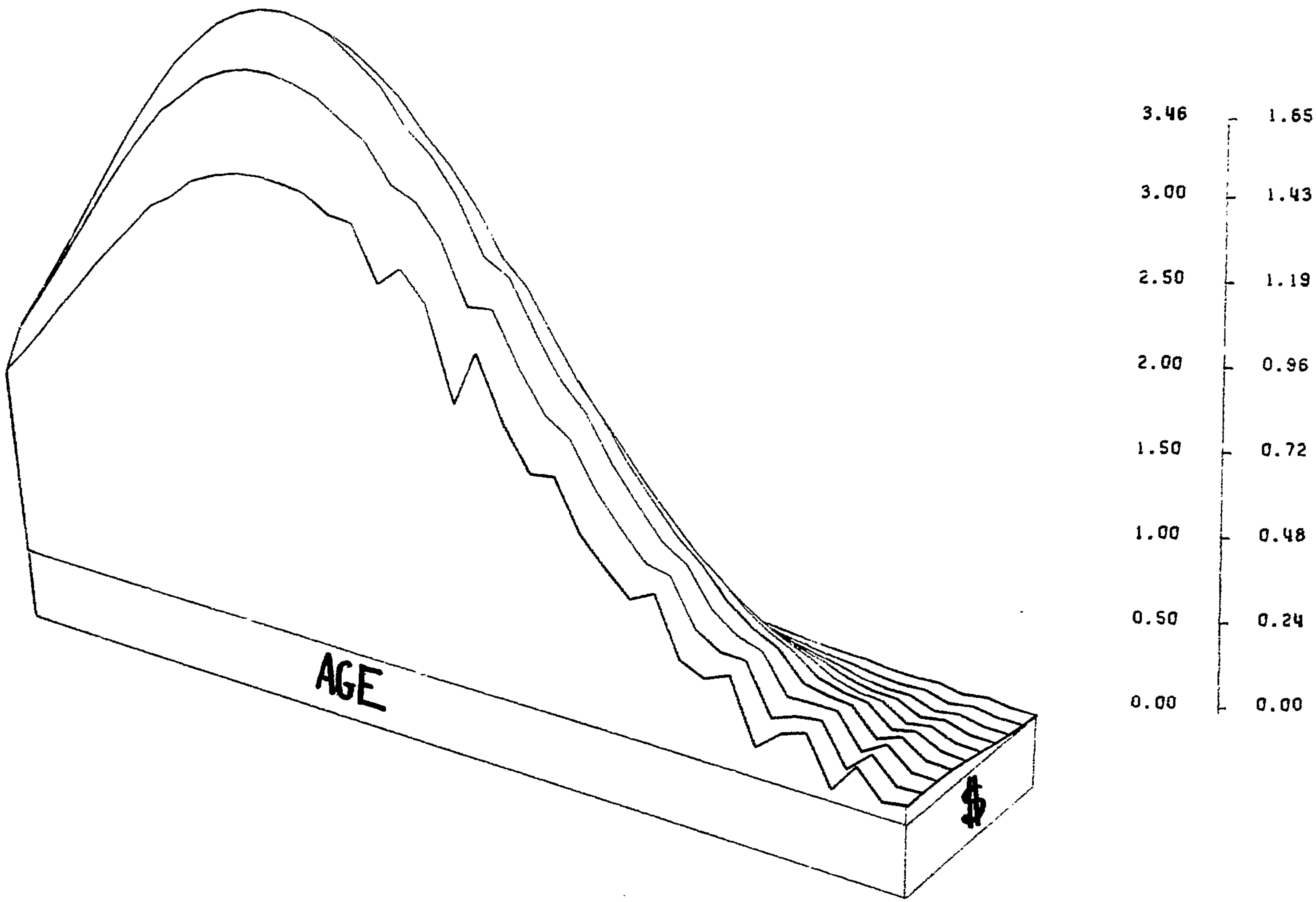


$a_4(t)$ - Fourth Component from Low Aggregation

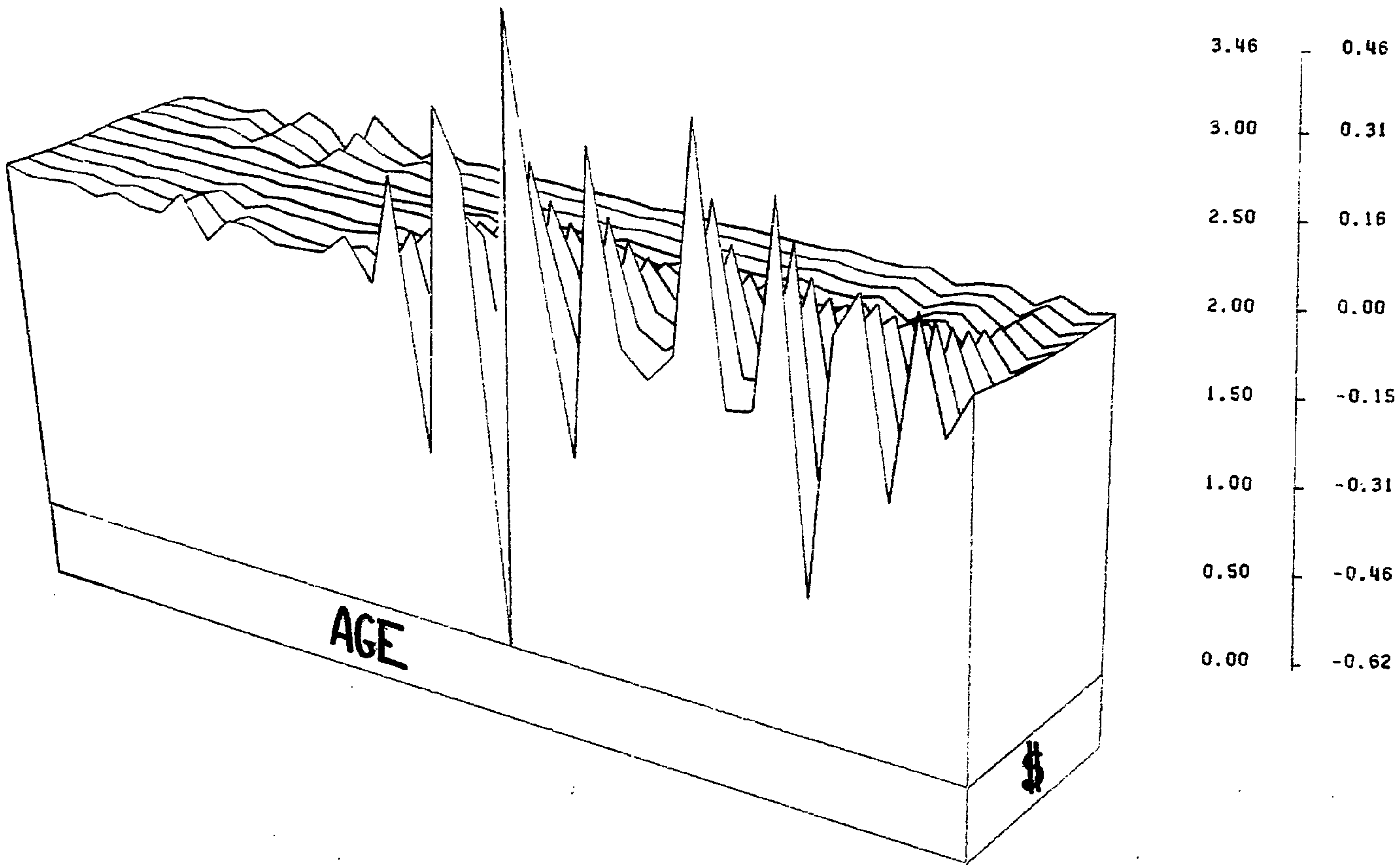
CHART 5



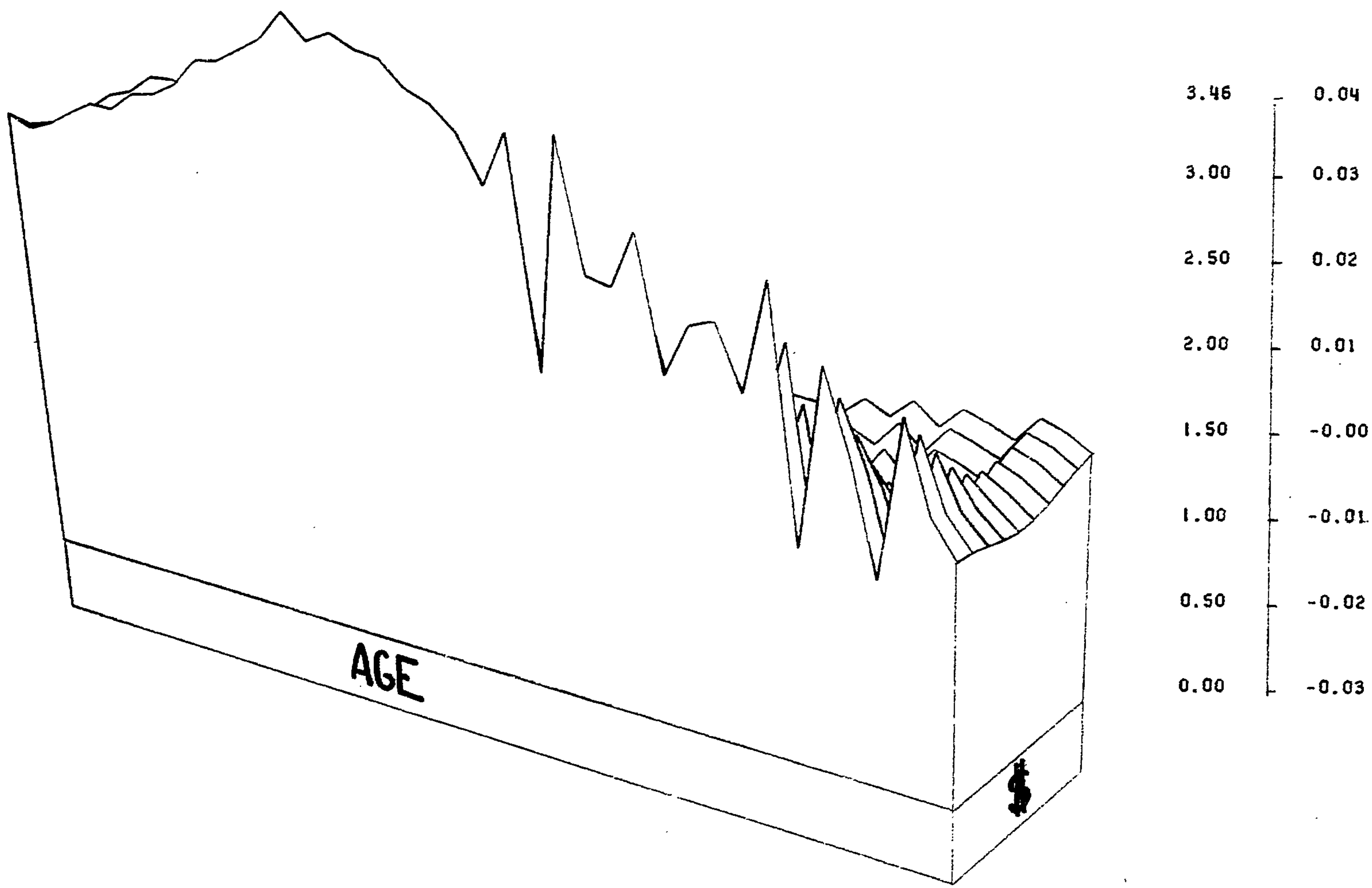
$a_5(t)$ - Fifth Component from Low Aggregation



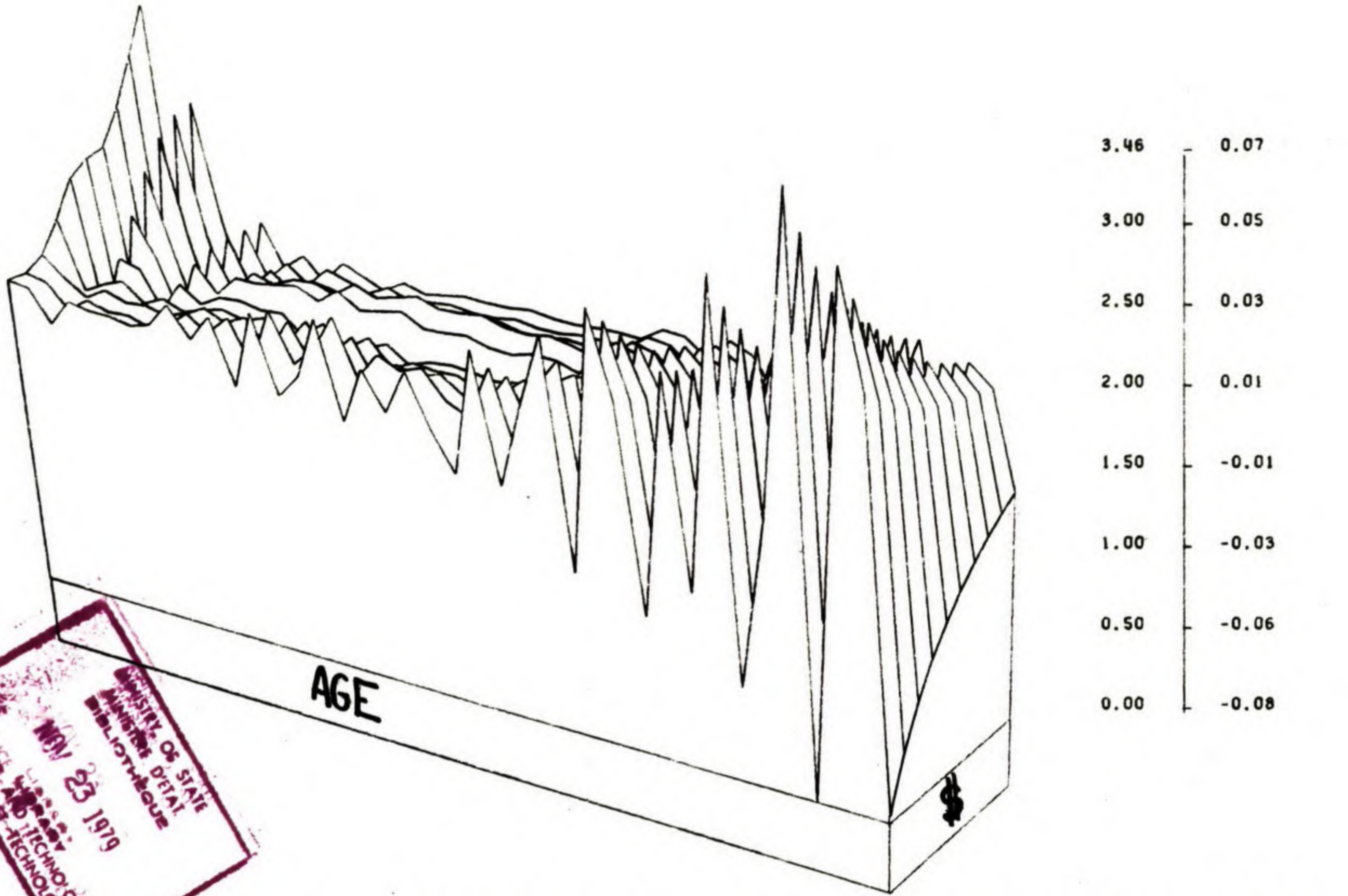
COMPONENTS OF MU FROM LOW AGGREGATION - AGES 17 - 56 COMPONENT 1



COMPONENTS OF MU FROM LOW AGGREGATION - AGES 17 - 56 COMPONENT 2

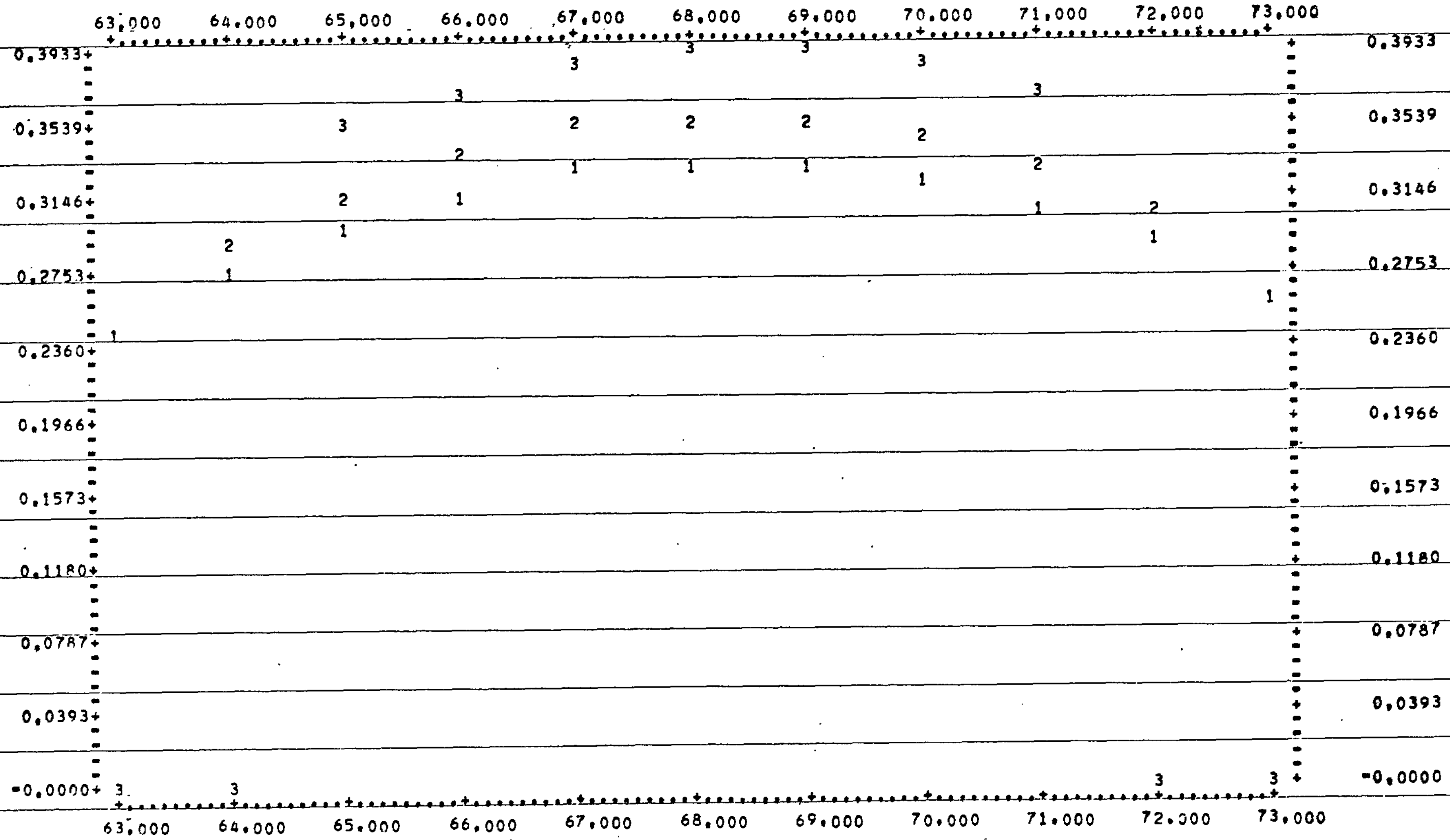


COMPONENTS OF MU FROM LOW AGGREGATION - AGES 17 - 56 COMPONENT 3



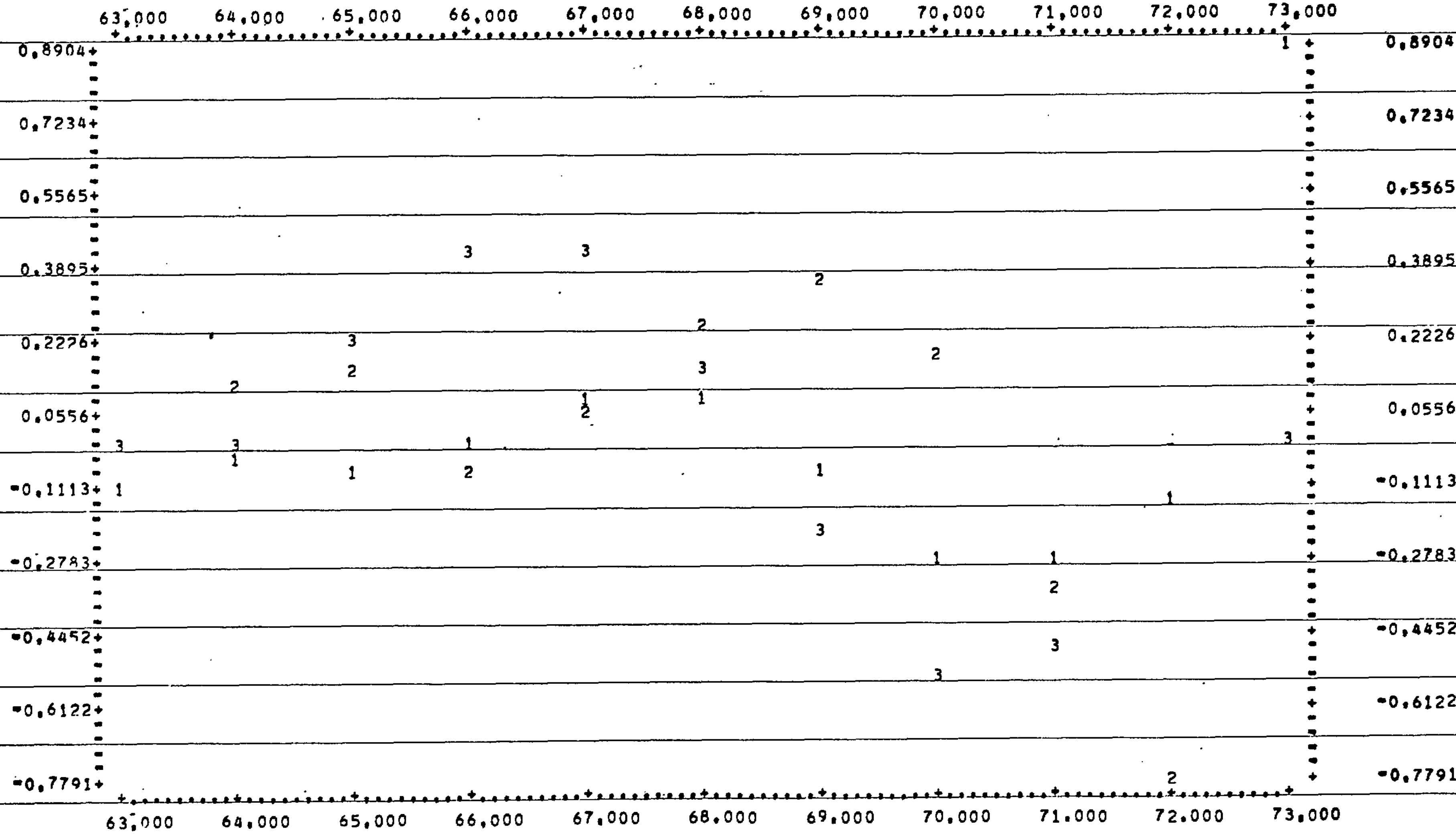
COMPONENTS OF MU FROM LOW AGGREGATION - AGES 17 - 56 COMPONENT 4

CHART 1



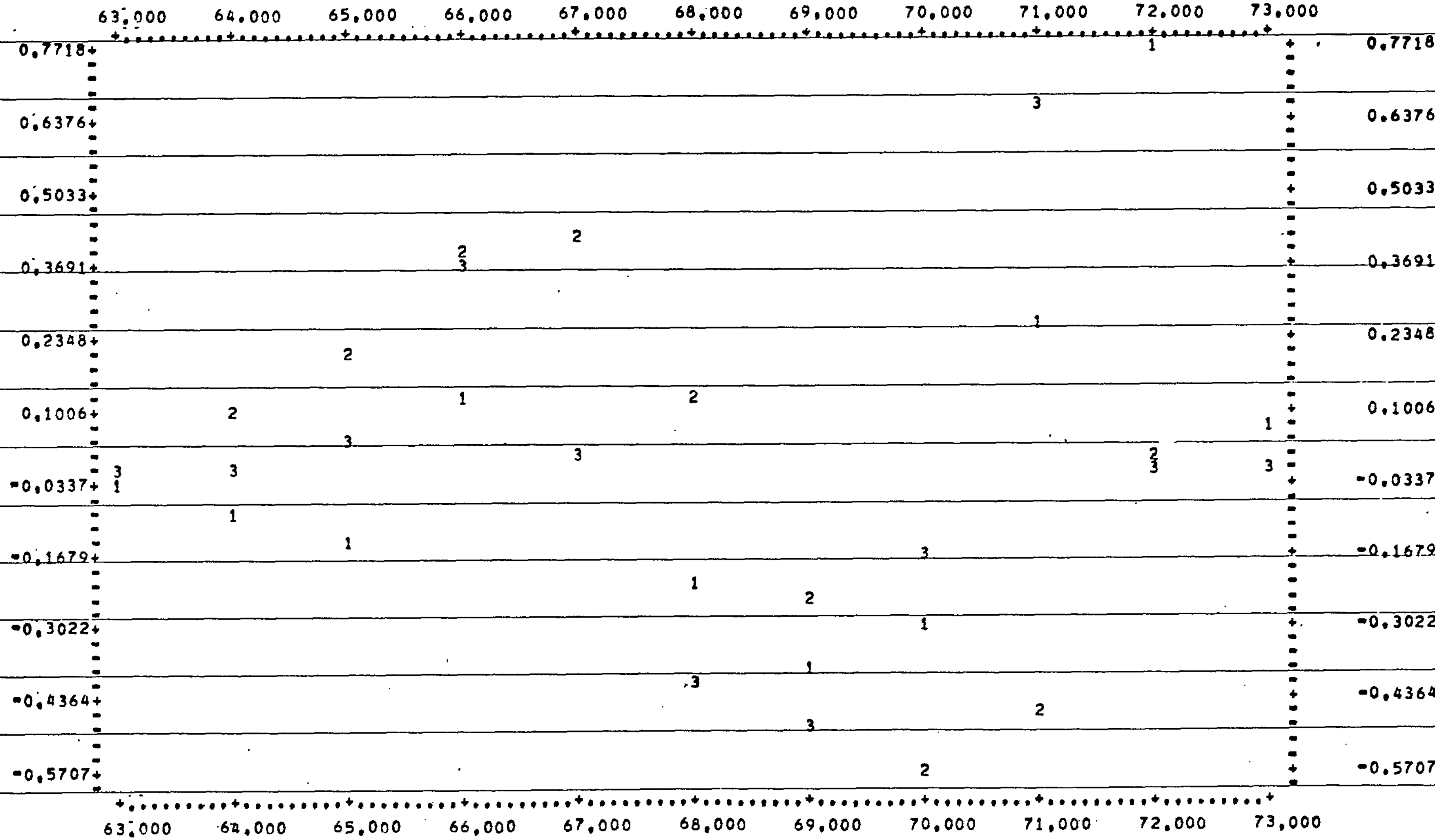
$a_1(t)$ - First Component from Intermediate Aggregation

CHART 2

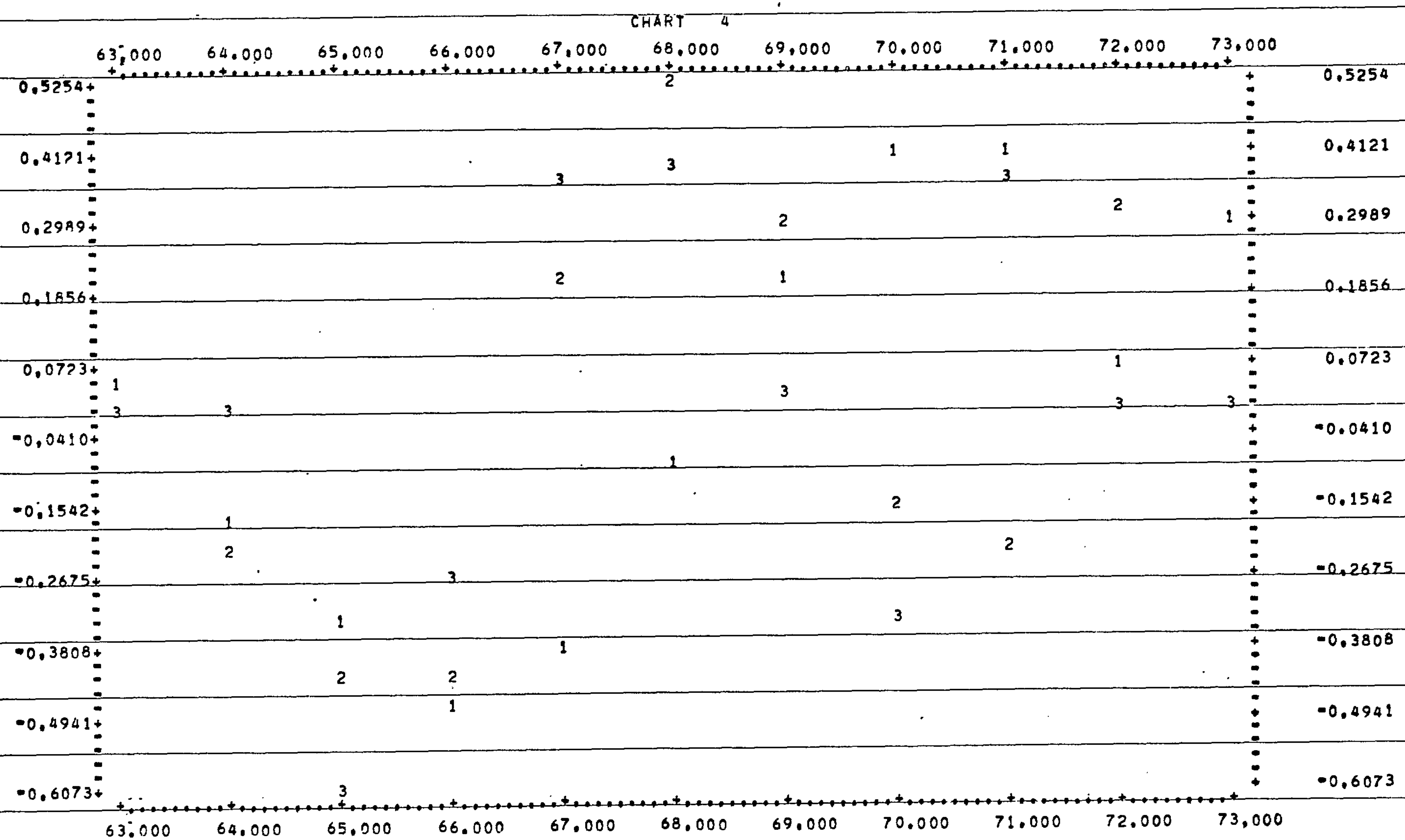


$a_2(t)$ - Second Component from Intermediate Aggregation

CHART 3

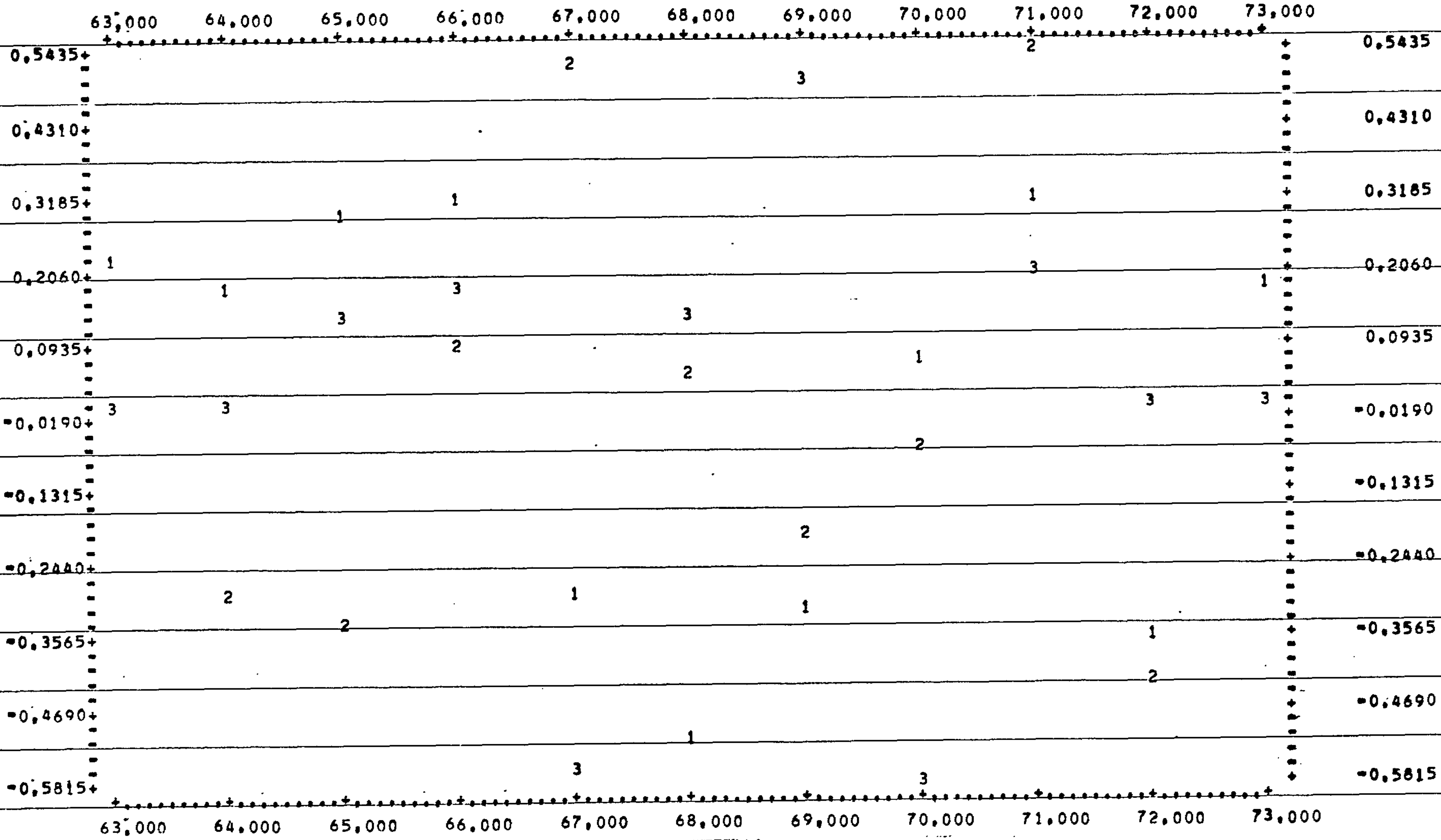


$a_3(t)$ - Third Component from Intermediate Aggregation

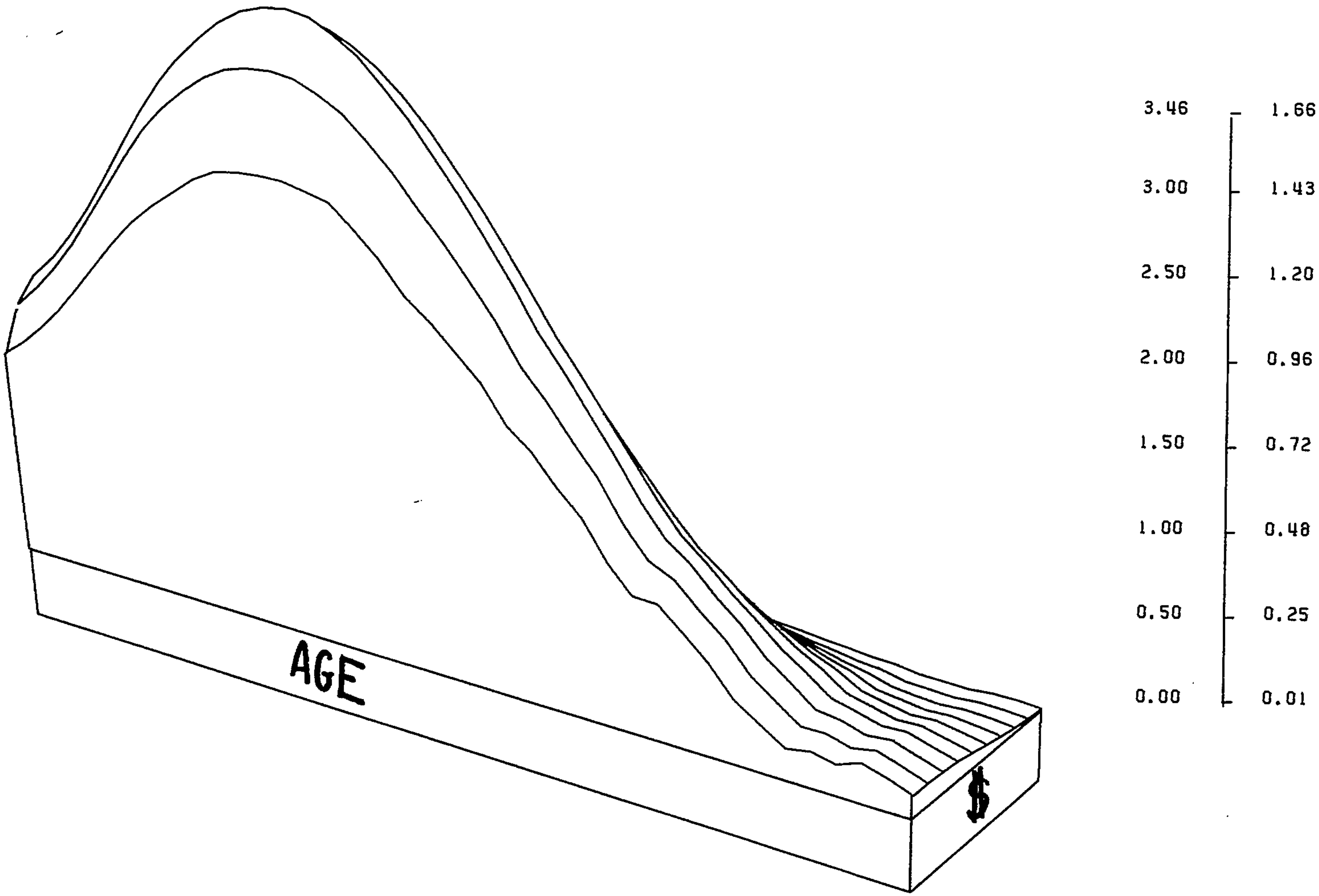


$a_4(t)$ - Fourth Component from Intermediate Aggregation

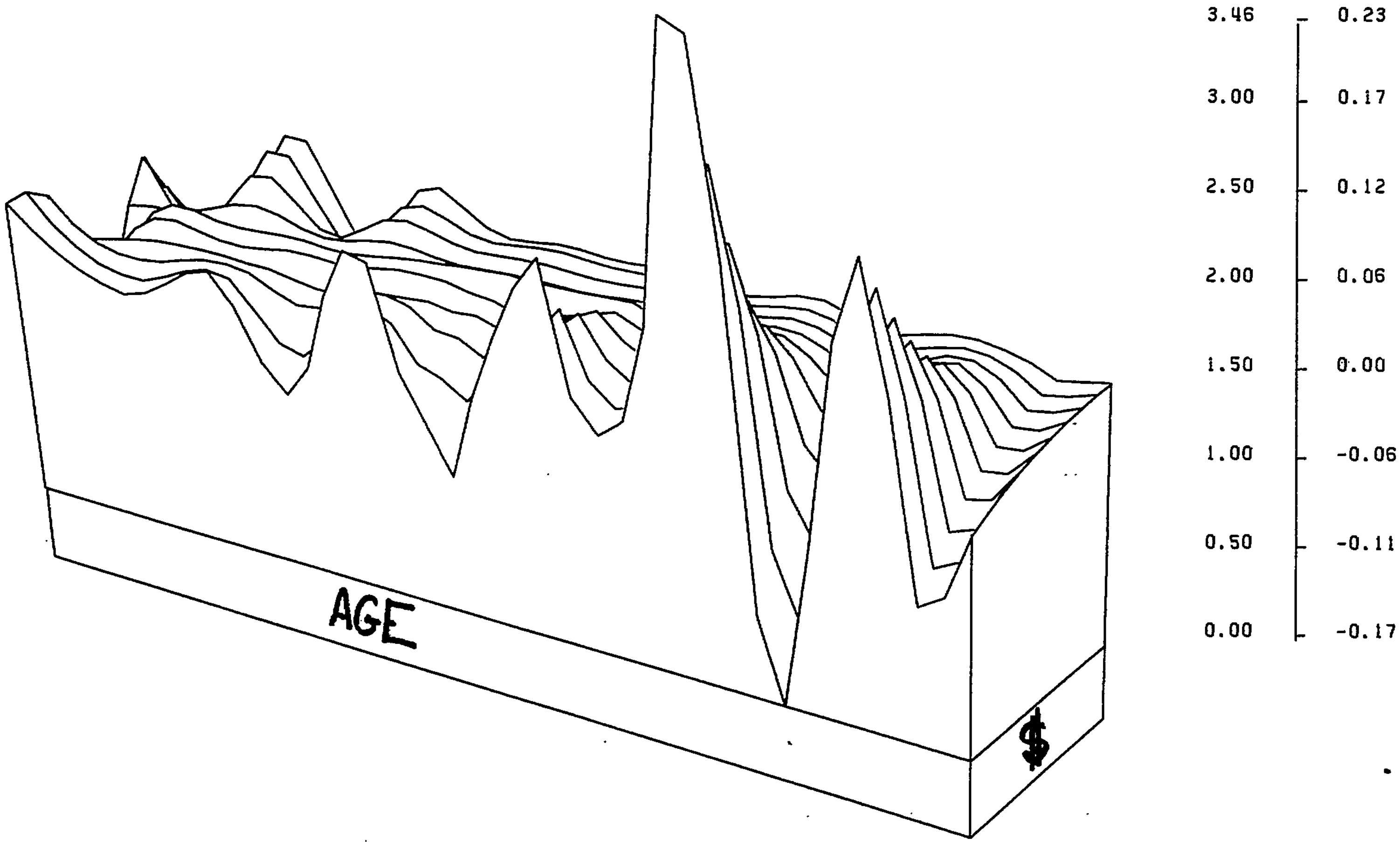
CHART 5



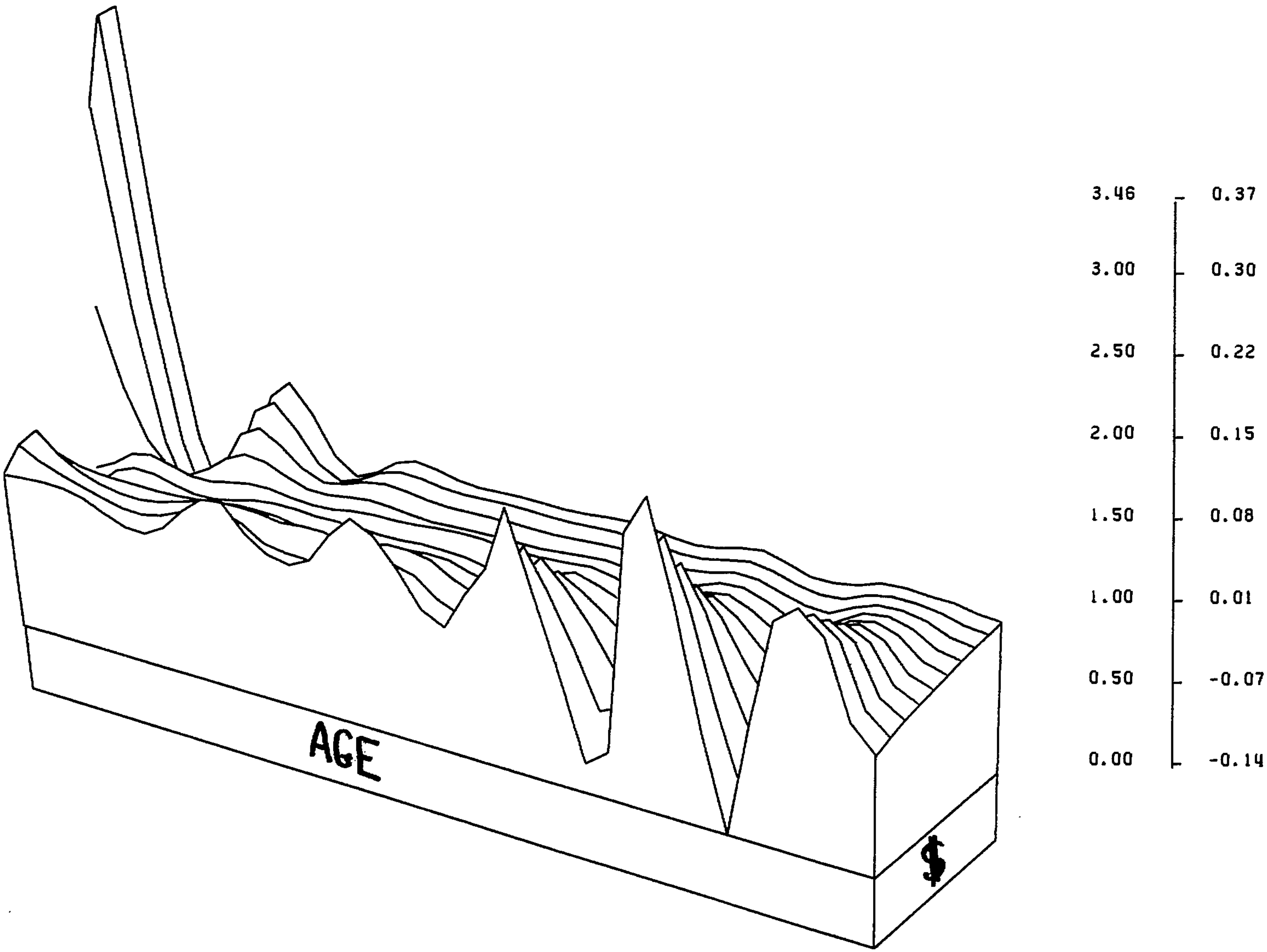
$a_5(t)$ - Fifth Component from Intermediate Aggregation



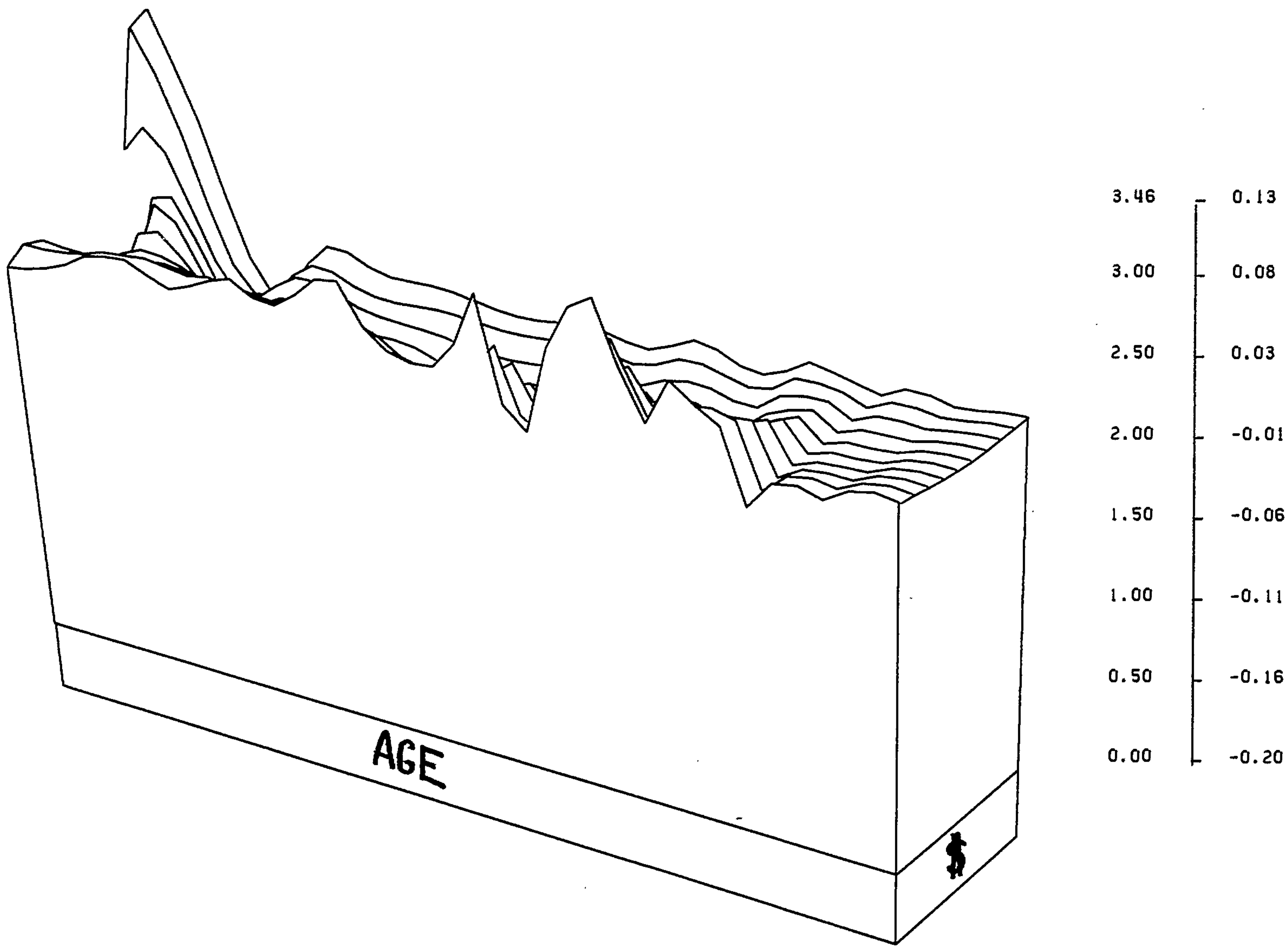
COMPONENTS OF MU FROM INTERMEDIATE AGGREGATION - COMPONENT 1 = FIRST MODE



COMPONENTS OF MU FROM INTERMEDIATE AGGREGATION - COMPONENT 2 = SPURIOUS MODE

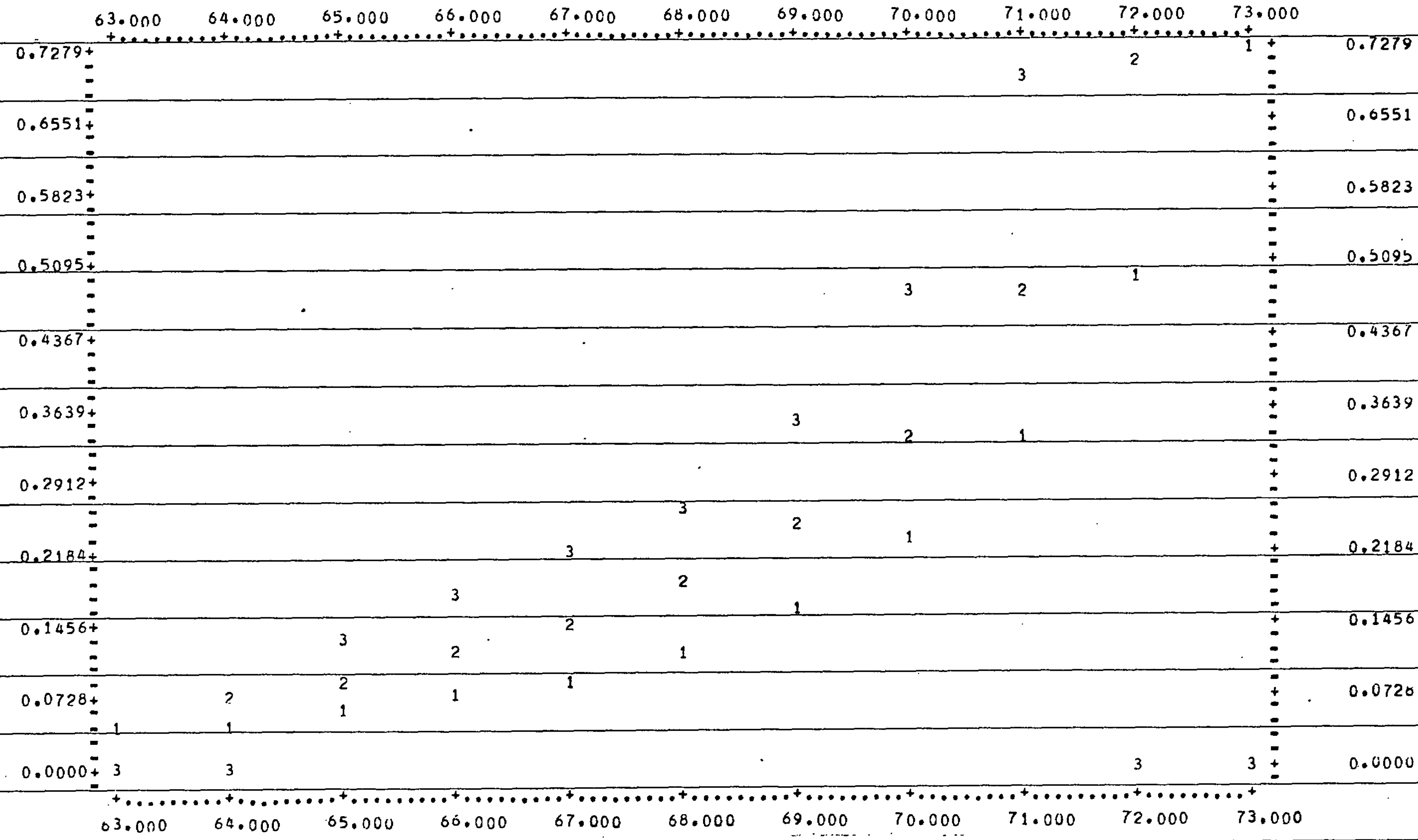


COMPONENTS OF MU FROM INTERMEDIATE AGGREGATION - COMPONENT 3 = SPURIOUS MODE



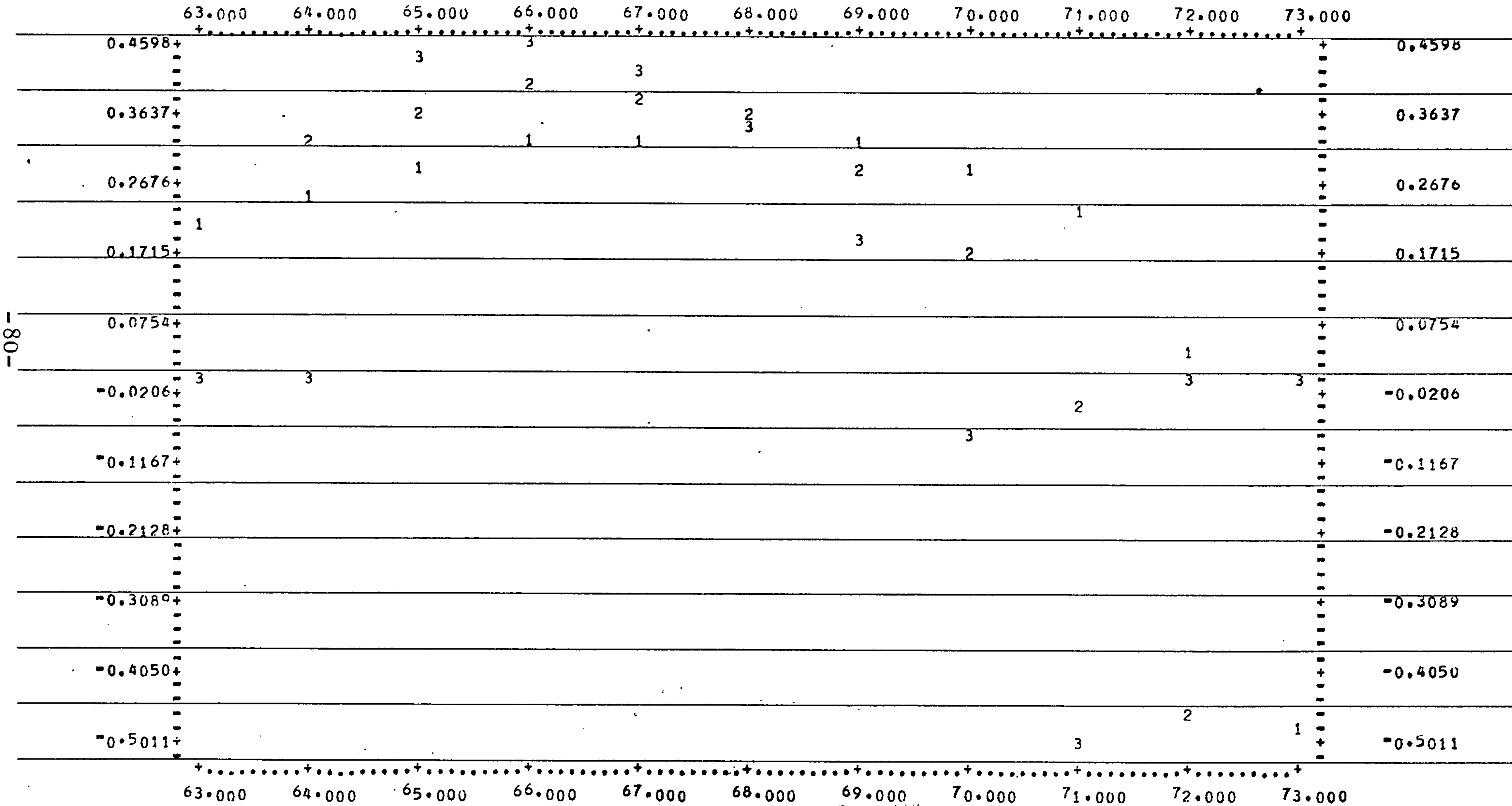
COMPONENTS OF MU FROM INTERMEDIATE AGGREGATION - COMPONENT 4 = SECOND MODE

CHART 1



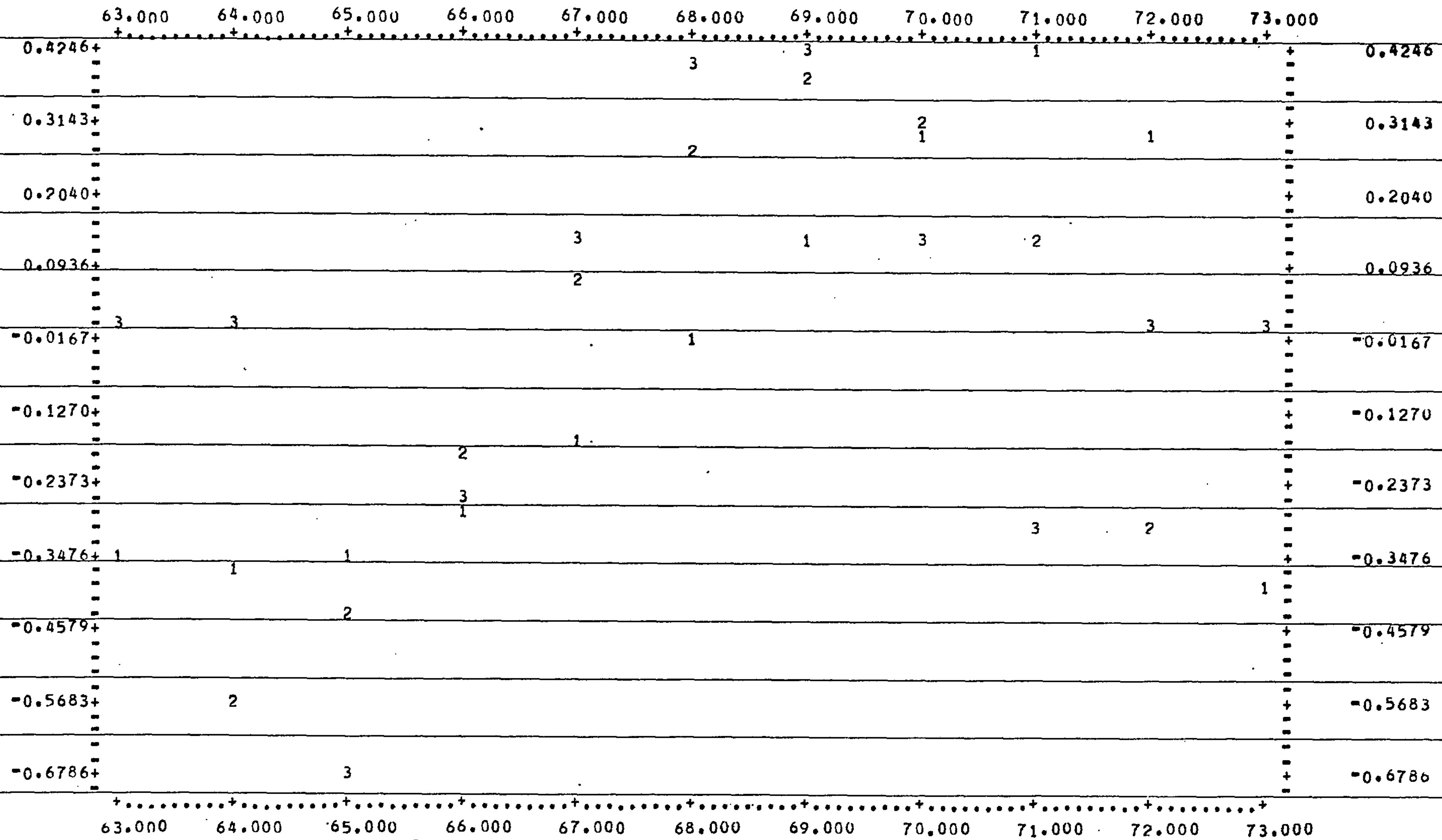
$a_1(t)$ - First Component from High Aggregation

CHART 2

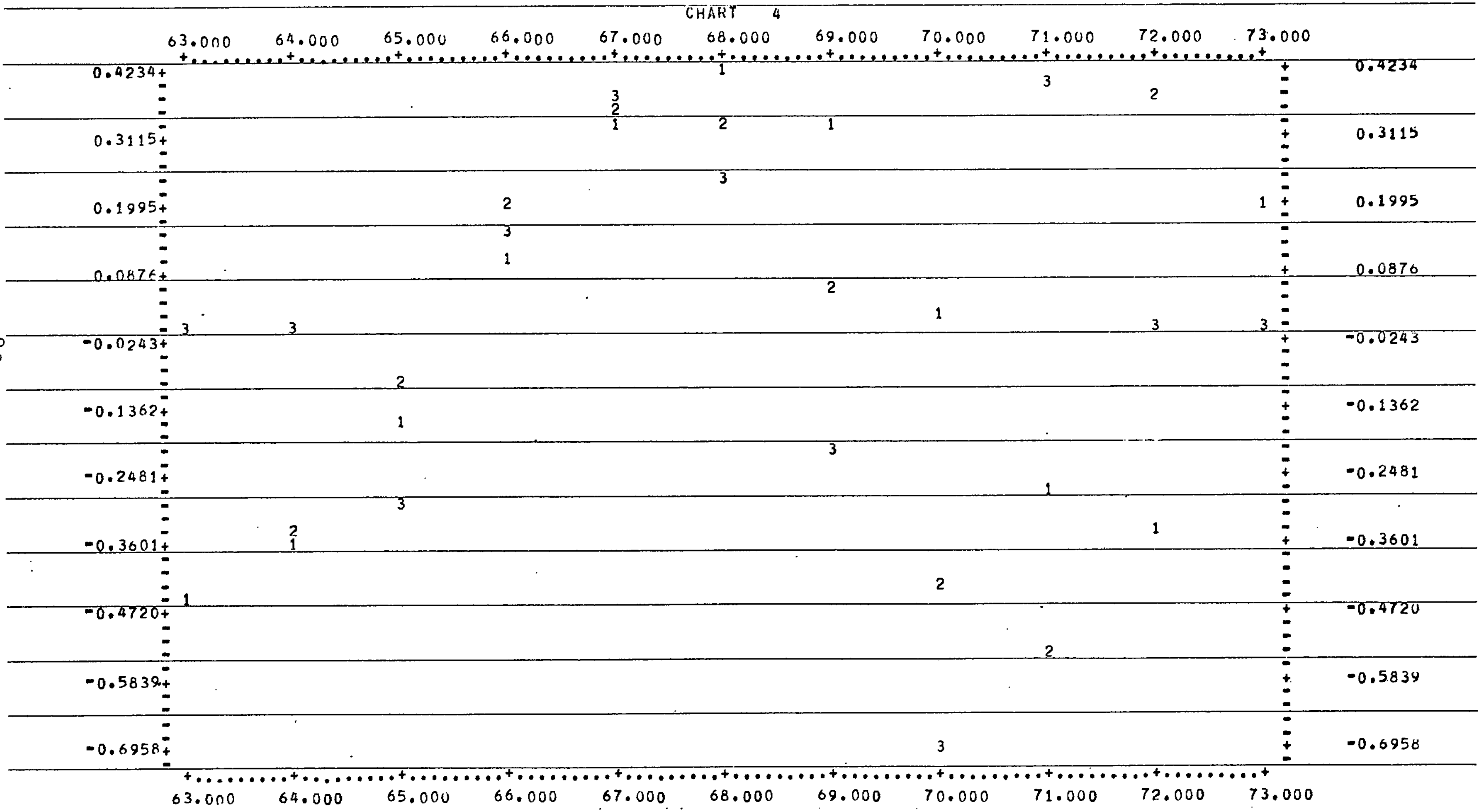


$a_2(t)$ - Second Component from High Aggregation

CHART 3

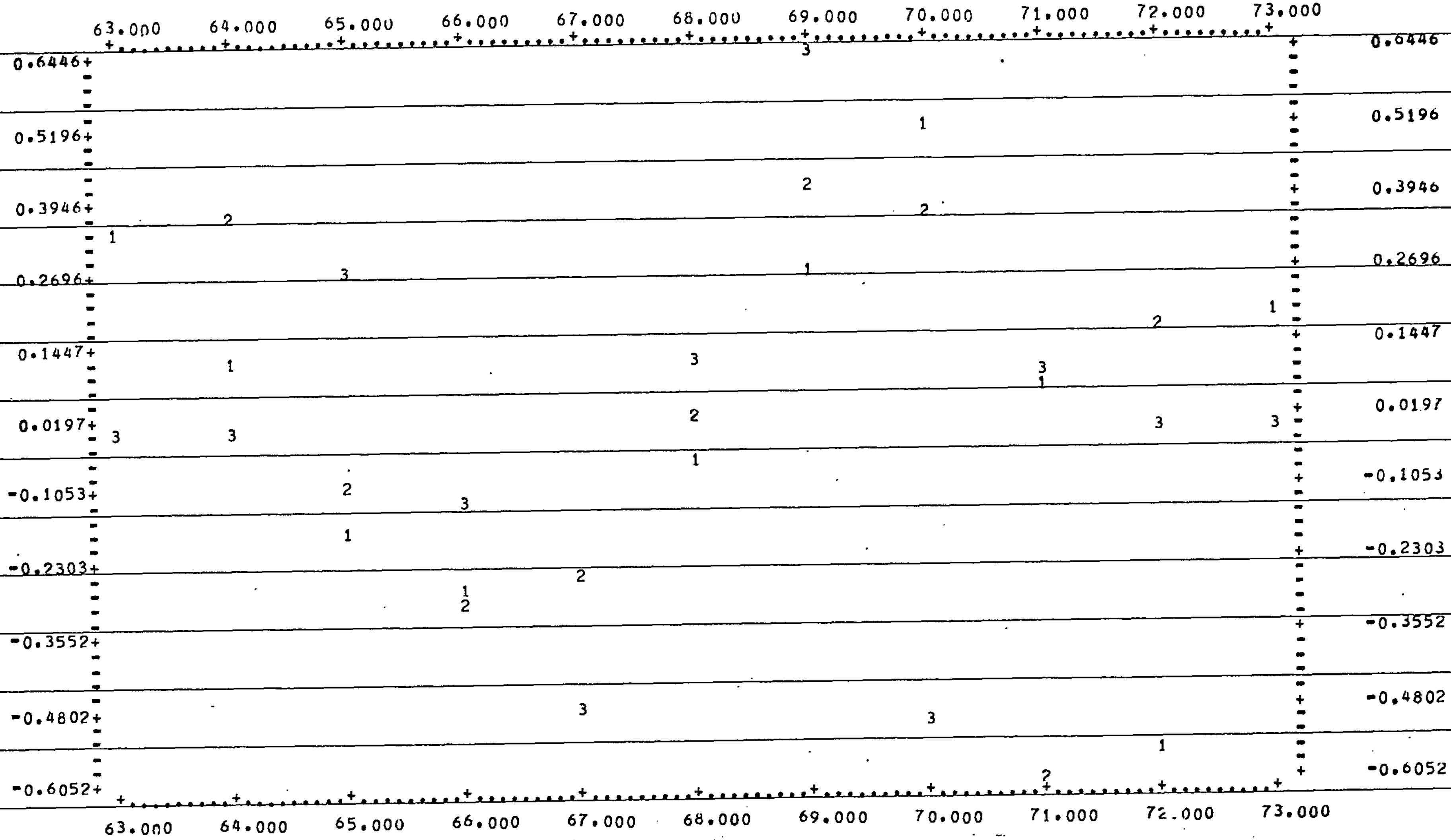


$a_3(t)$ - Third Component from High Aggregation

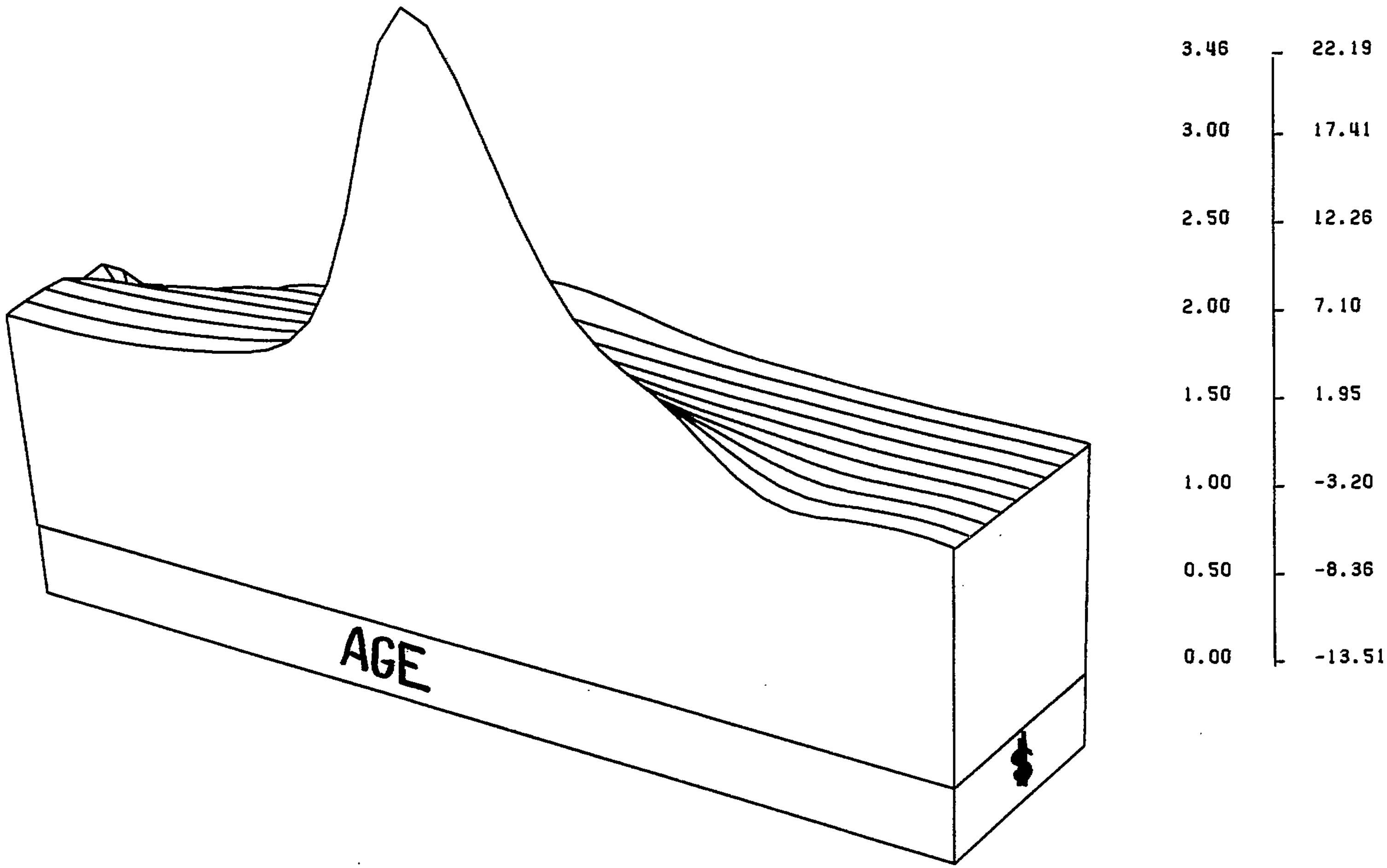


$a_4(t)$ - Fourth Component from High Aggregation

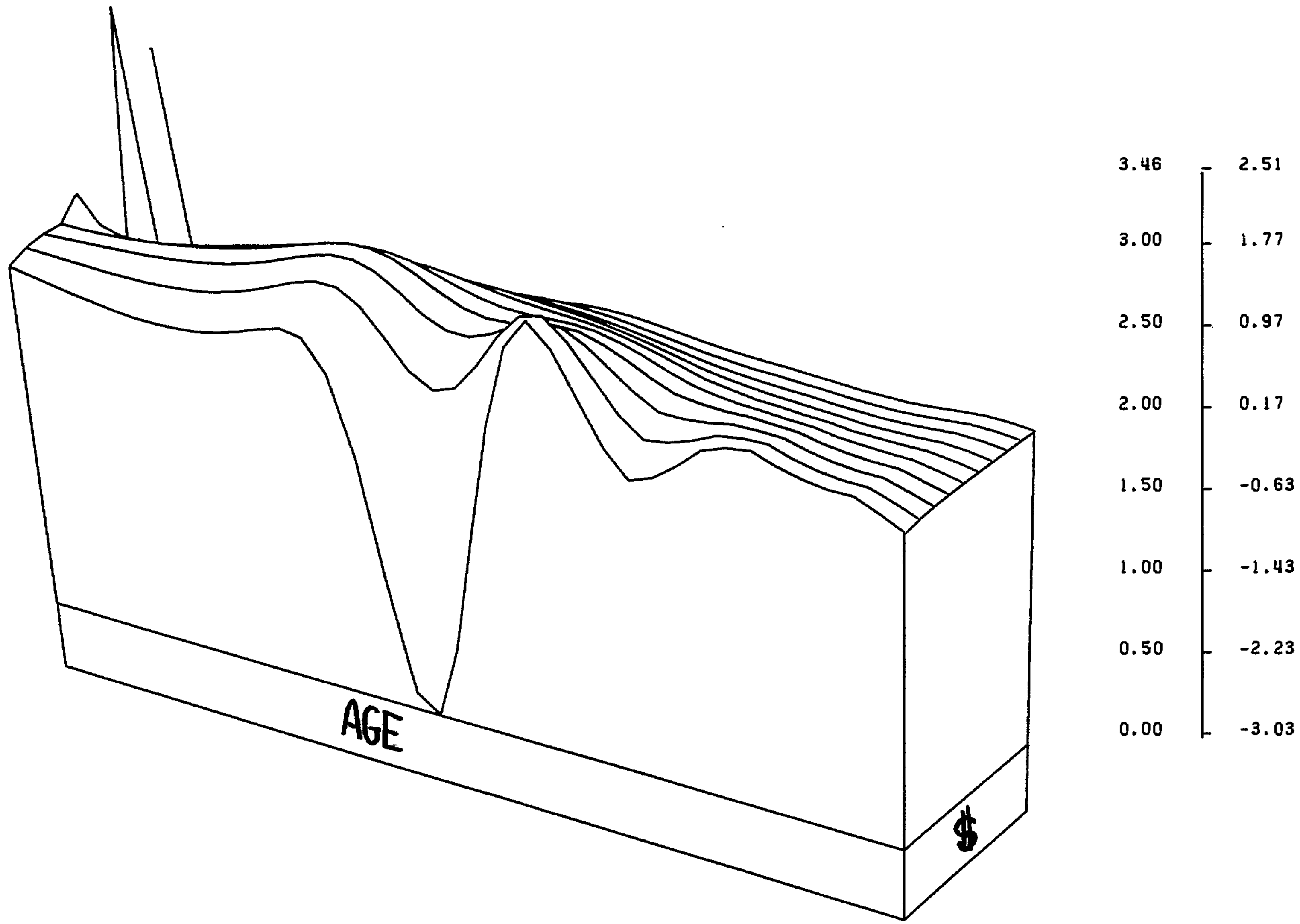
CHART 5



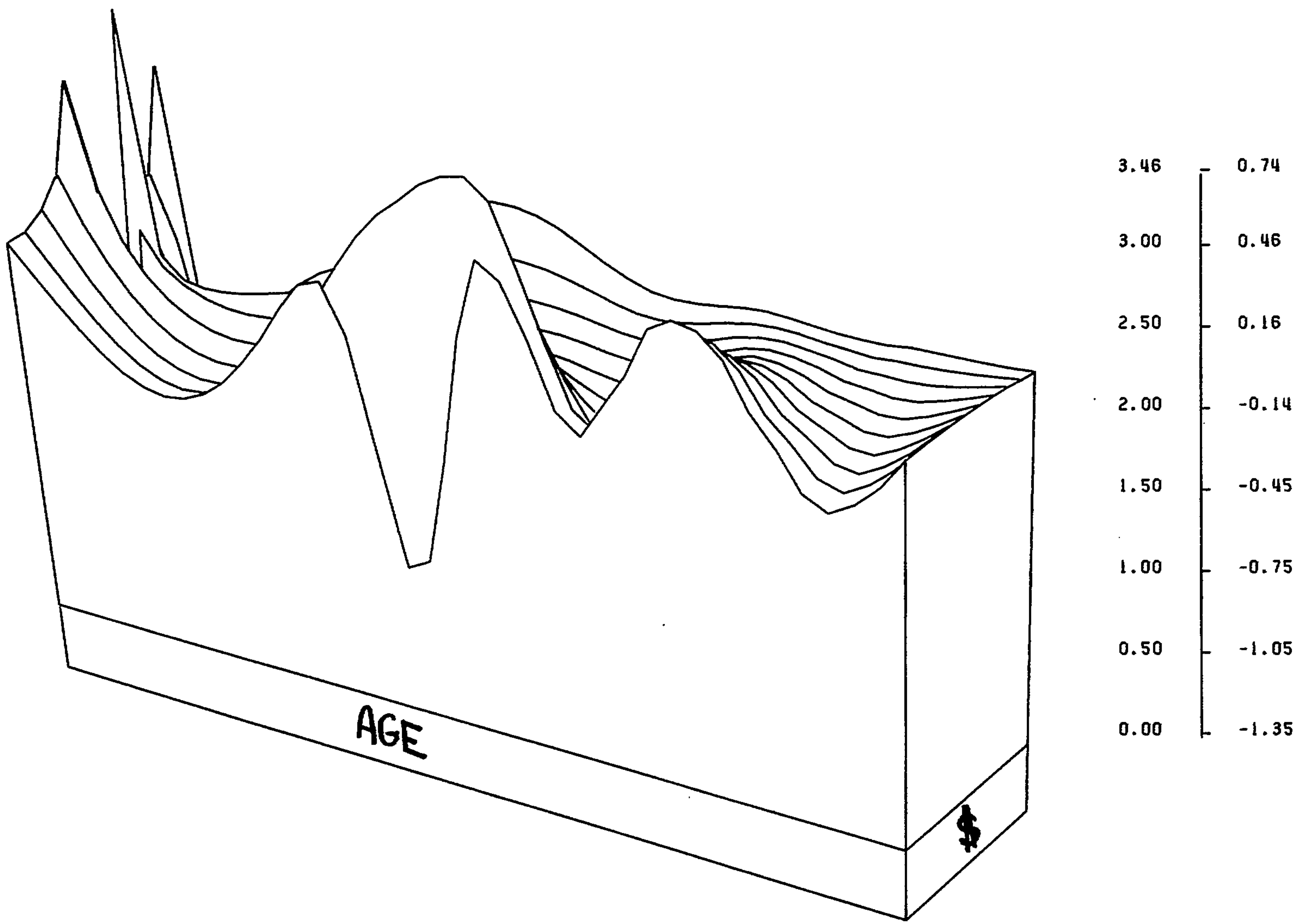
$a_5(t)$ - Fifth Component from High Aggregation



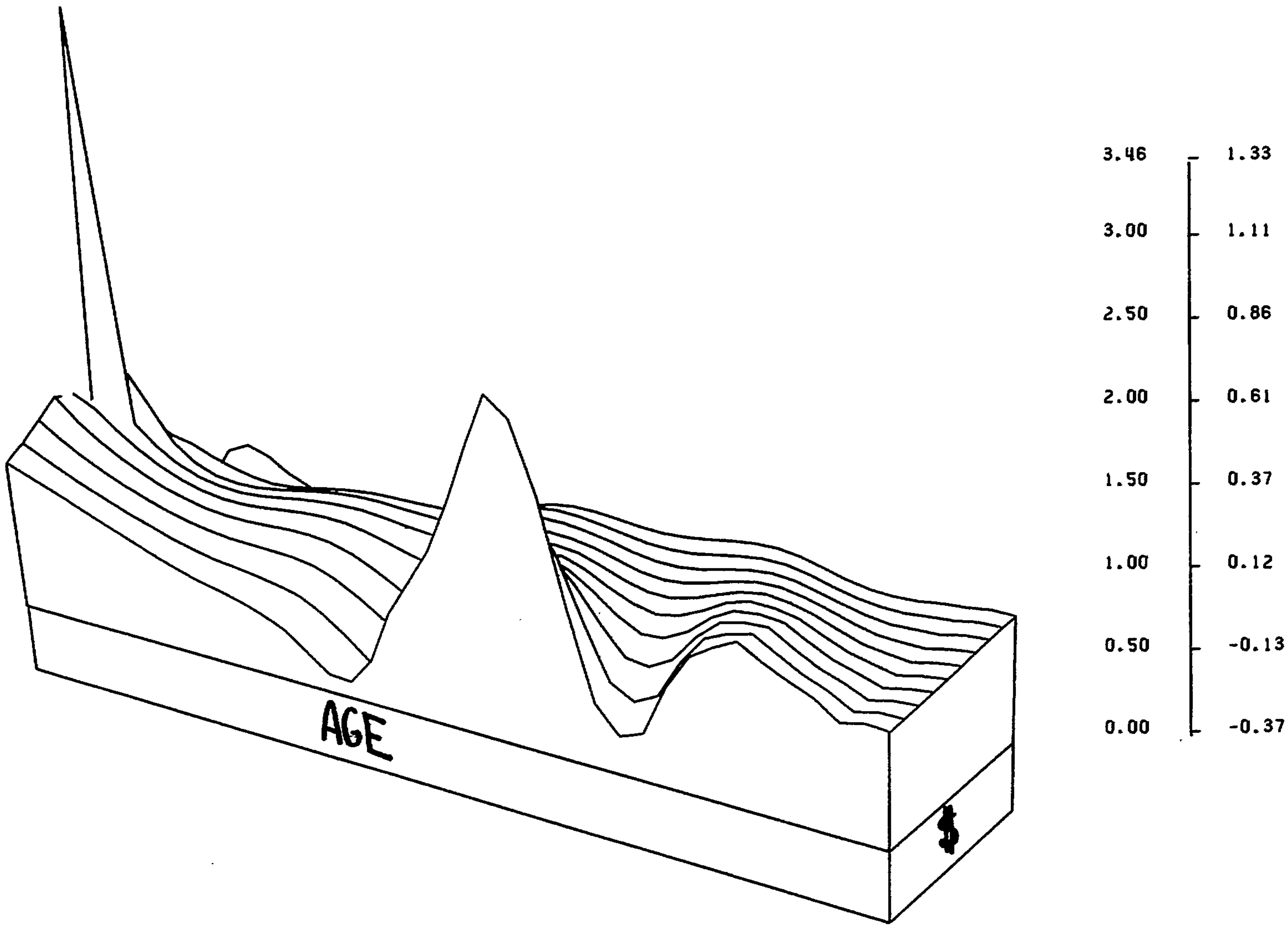
COMPONENTS OF MU FROM HIGH AGGREGATION - COMPONENT 1 = SPURIOUS MODE



COMPONENTS OF MU FROM HIGH AGGREGATION - COMPONENT 2 = FIRST MODE



COMPONENTS OF MU FROM HIGH AGGREGATION - COMPONENT 3 = SECOND MODE?



COMPONENTS OF MU FROM HIGH AGGREGATION - COMPONENT 4

The above plots contain a large amount of information; we confine our remarks to the most significant aspects of the computed results.

One effect clearly visible in the plots is the presence of a "saw tooth" component in the calculated age-income components. This effect is essentially due to the coarseness of the grid used for integrating the simulated density function. This could be cut down by use of smaller grid increments, with a corresponding increase in computing time and cost. The problem is not helped, and is probably aggravated by the fact that the initial distribution used for the simulation is not a "natural" initial condition for the mobility function moved. This could be overcome by running the simulation for a long enough period to ensure that the initial distribution disappears.

The presence of these errors is not entirely harmful, as it can simulate the inevitable data errors and modelling inaccuracies of the real situation. Also, it is evident that the integration process in the intermediate aggregation has had a smoothing effect on these errors.

Given the relatively small component of sinusoidal nature which is present in the mobility chosen for the simulation (See the eigenvalue calculations above), what is perhaps somewhat surprising is that the algorithm is able to extract a sinusoid at all from the simulated data. Note that with two year borderings, the sine function appears as second component in the density, low and intermediate aggregation cases. The density calculation seems to be the only one which returns a reasonable $d(x,s)$ for the sinusoid (but with noticeable sawtooth component).

Of much more practical interest is the fate of the dominant, or parabolic component in the simulated mobility at various levels of aggregation. As the results clearly show, this parabolic component is present with all degrees of bordering for the cases of the density, and low and intermediate aggregation. In all three cases, a reasonable representation of the age-income component is returned (essentially marred only by the integration error sawtooth disturbance).

When the aggregation is increased to the level of that present in the available data, however, the results are drastically altered. There is a dominant time component which persists through all levels of bordering, and which seems entirely spurious. We attribute its presence entirely to the level of aggregation. The "true" parabolic component seems to occur as the second component, only at the zero-bordering level, and there in a form distorted by the spurious component. With some imagination, it is possible to see the corresponding age-income component as a distorted form of the "true" age-income component.

We regard the above computational results as strongly suggesting that the aggregation level of the age-income data published in the annual reports of Revenue Canada is too high to allow determination of economic mobility by use of the estimation methods developed in this report.

C Real Data Computations

As mentioned above, the only annual income distribution data which we were able to obtain consists of the reported income figures from the annual reports of Revenue Canada.

A potential problem which arises from the use of this data is the fact that these reported income figures include the effects of inflation on wage and salary rates. While it was shown in the interim report that the form of the governing equations for the age-income distribution was invariant under an arbitrary monotone change of income scale, the possible effect of inflation on an income measure was not explicitly considered. In fact, derivation of the governing equation in [3] was originally carried out on the basis of an implicit assumption of an income measure constant over time.

However, one may consider the effect of using an inflating income scale in a manner analogous to the derivation in [3] regarding changes of income measure.

The simplest such change is given by defining

$$\sigma = c(t)\ s$$

where s represents the original income measure (constant dollars), $c(t)$ represents the current consumer price index (cumulative inflation factor), and so that σ represents income in "current" dollars.

One objection to the use of the above in connection with the reported income statistics is that reported income is made up of various components, all of which are unevenly affected by inflation; moreover, the proportion of these components in reported income varies across income brackets. To take account of this possibility, we consider the effect of a time-varying income scale change of the general form

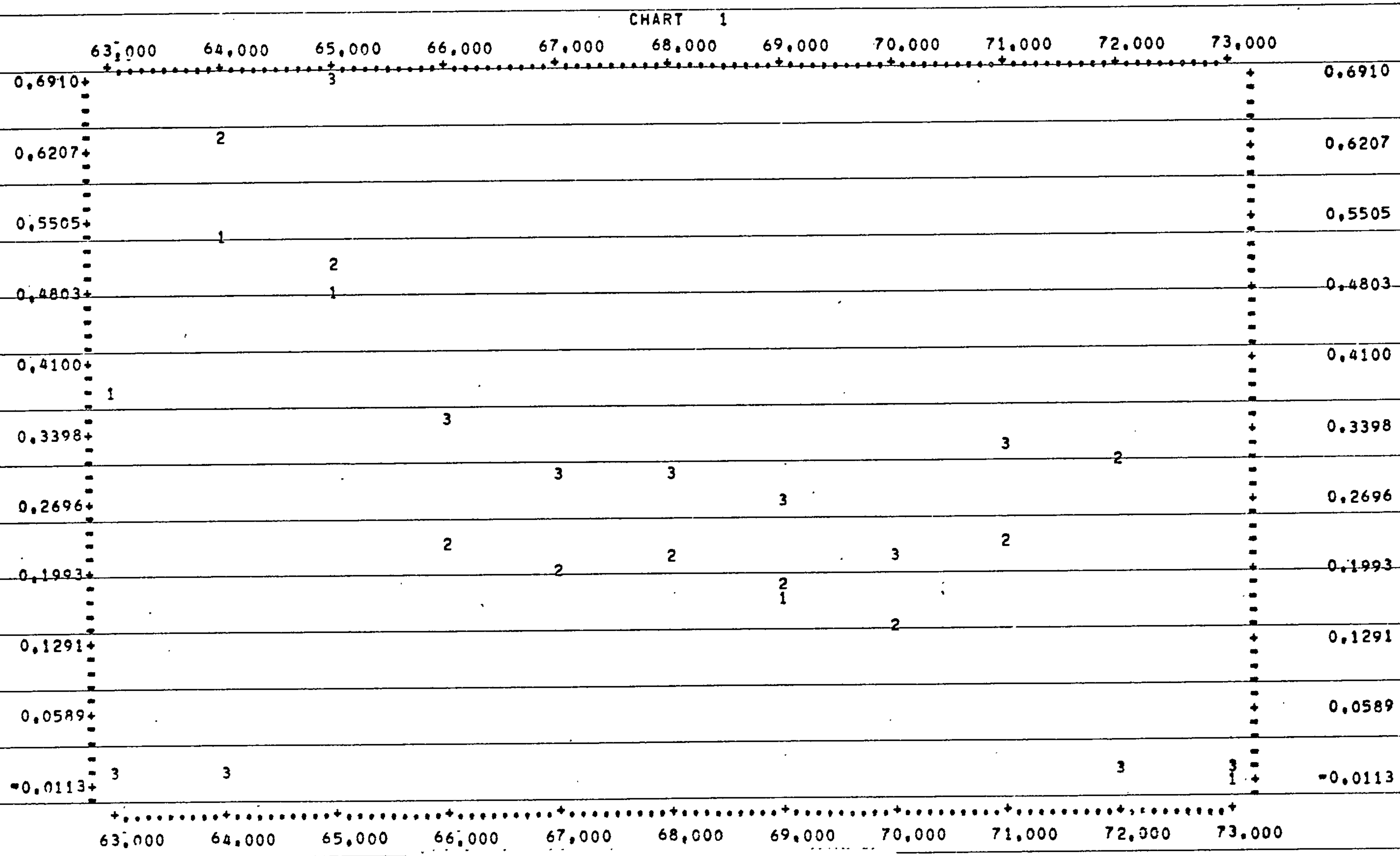
$$\sigma = \Sigma\ (s,t) \quad .$$

This formulation allows the possibility of uneven inflationary effects across income levels, and should be sufficient to allow a transition from reported income to virtually any fixed income measure. It is shown in Appendix A that the governing equation of the distribution is not changed in form by the introduction of such a change in income measure, that is, that there exists a well-defined economic mobility function defined in terms of the income level σ . This exercise shows that reported income data may be used with no change in the model formulation; estimates of economic mobility computed from this data simply refer to the income measure σ .

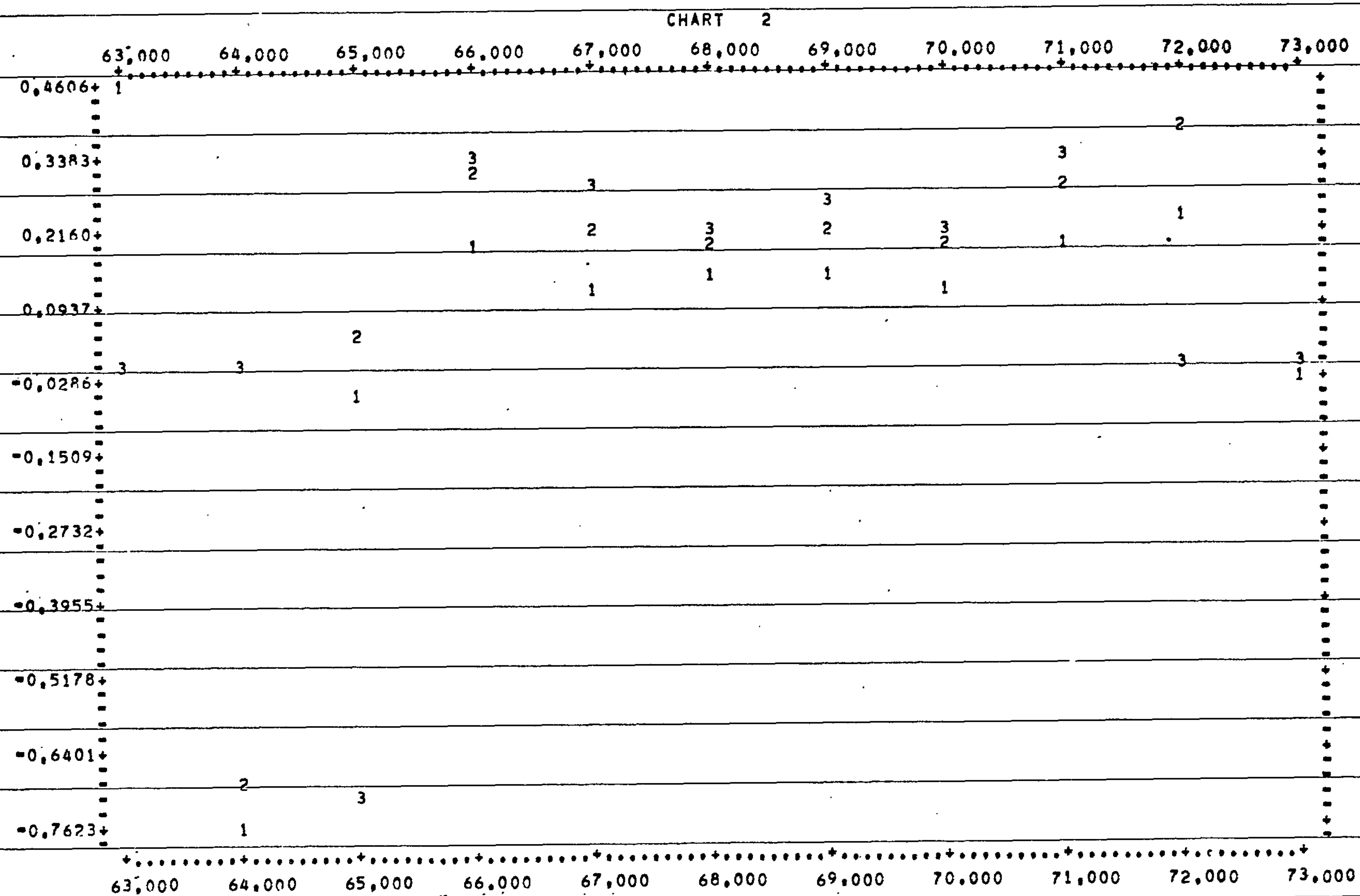
The above considerations effectively settle the question of the usability of the type of data available for the purposes of use in models of the sort considered; the next issue is whether or not the form of the available data is such that one may expect to extract useful model estimates.

The work reported in Section B above was undertaken to test the effects of data aggregation procedures on the validity of the computed estimates. The experiments run with simulated data indicates, for that example, that fairly reliable estimates may be extracted from data which is aggregated to about one half of the extent to which the available data has been aggregated. The results of the experimental run at high aggregation level do nothing to encourage confidence in results from data at this level of aggregation. In fact, the results indicate that spurious results are possible in such a situation.

In spite of the above situation, the fact remains that the highly aggregated real data is the <u>only</u> data available to us; natural curiosity has forced us to run the aggregated real data through the estimation algorithm. The results are reproduced below in a format parallel to that used for the aggregation experiments.

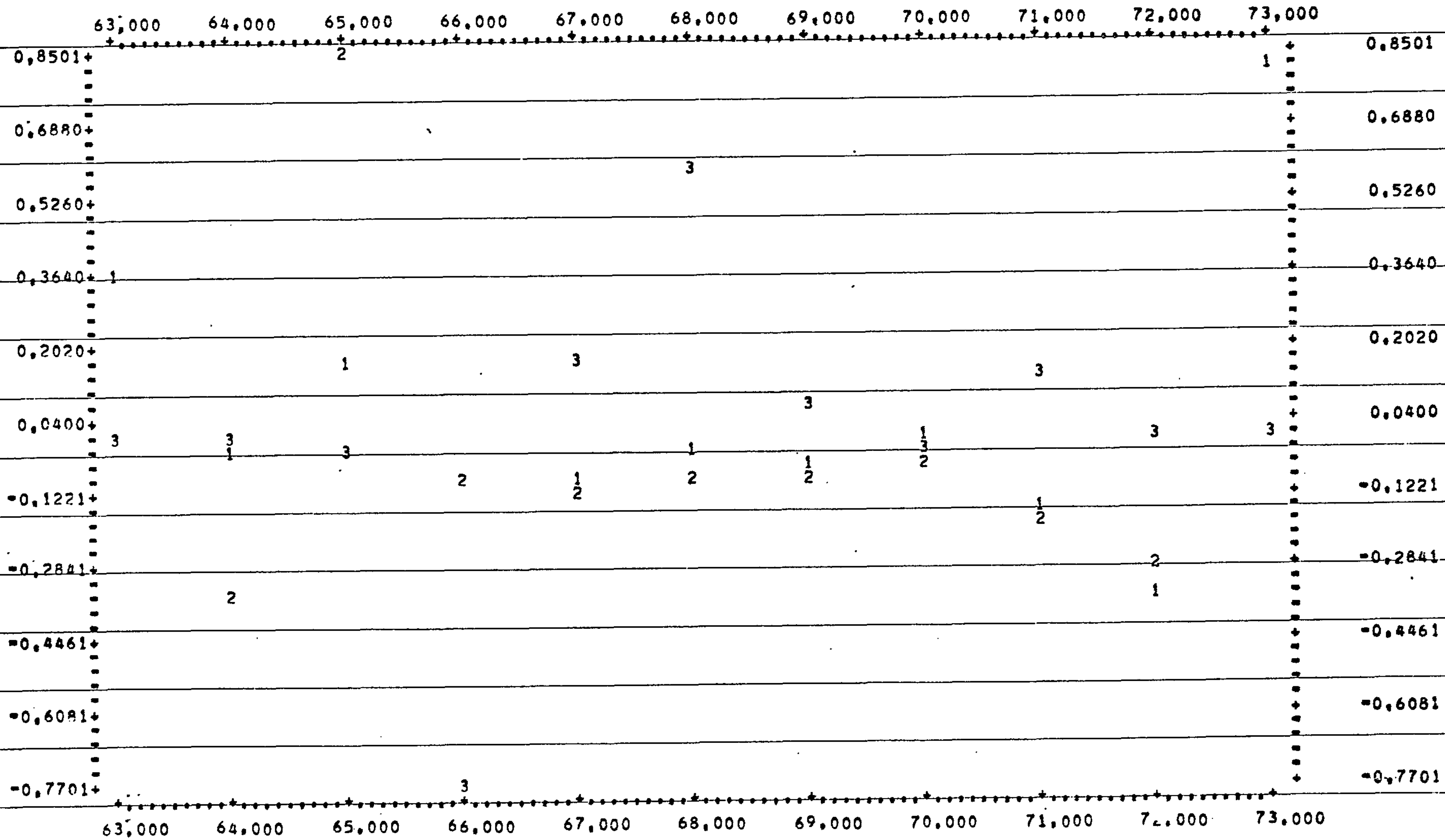


$a_1(t)$ - First Component from Real Data

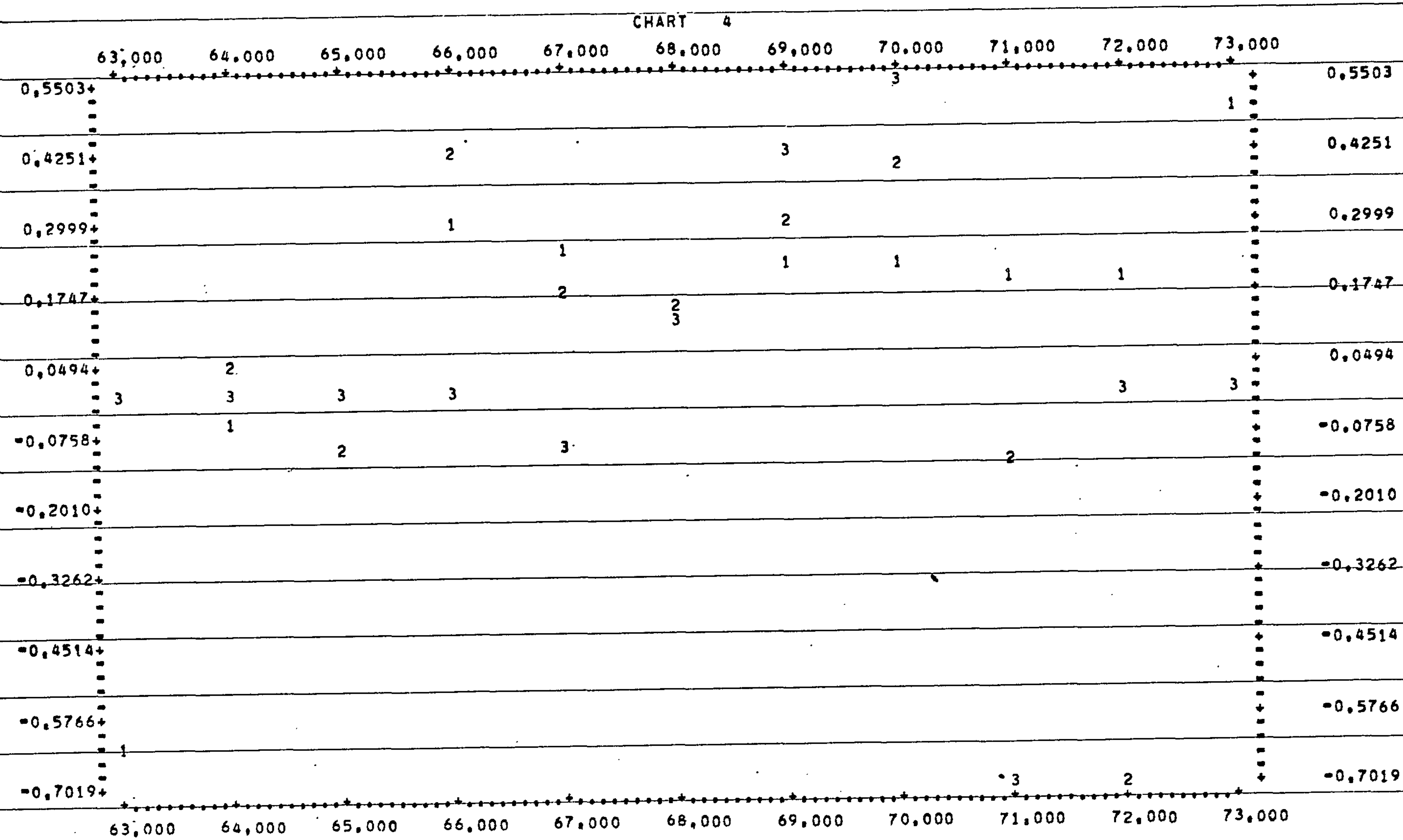


$a_2(t)$ - Second Component from Real Data

CHART 3

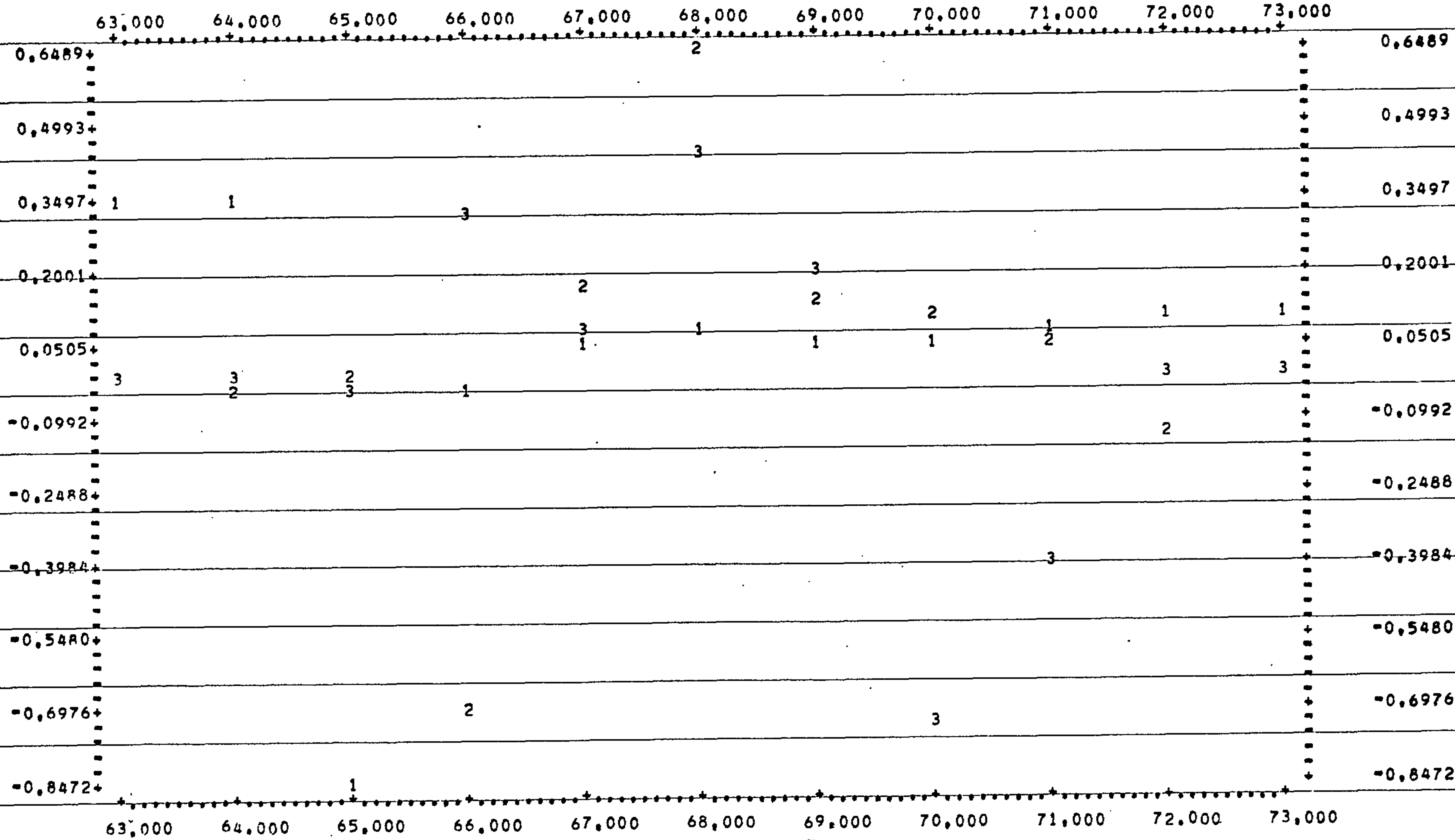


$a_3(t)$ - Third Component from Real Data

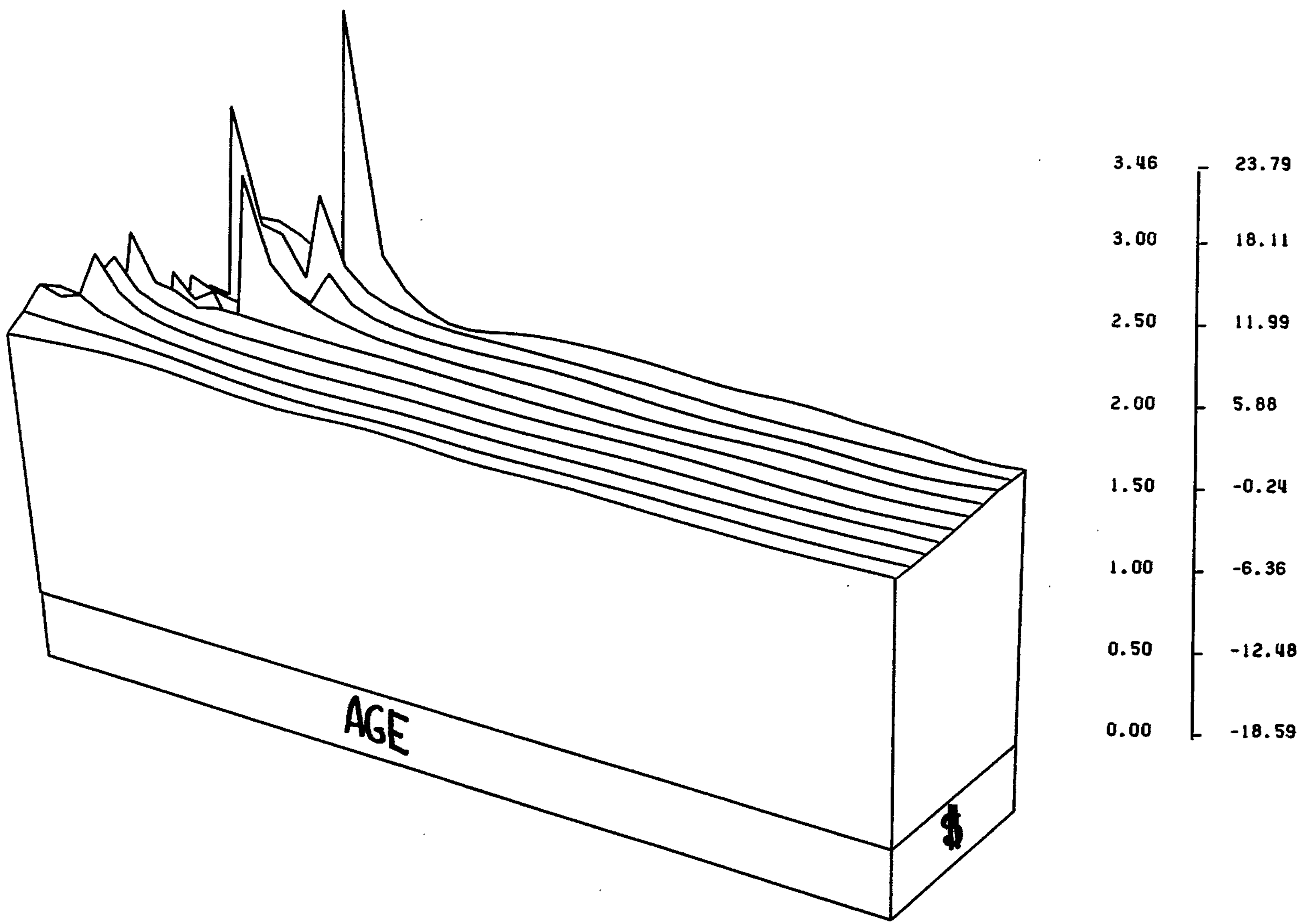


$a_4(t)$ - Fourth Component from Real Data

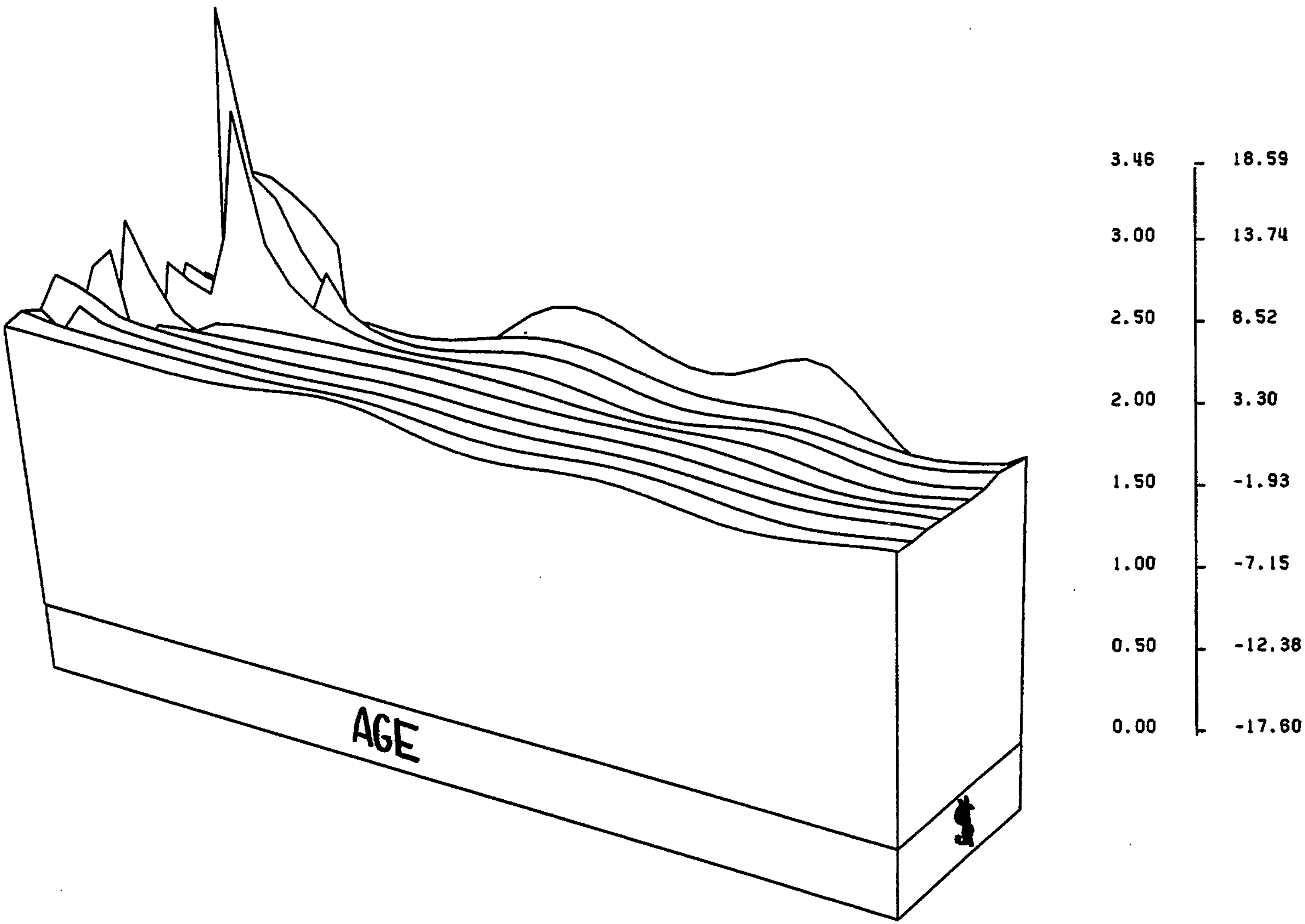
CHART 5



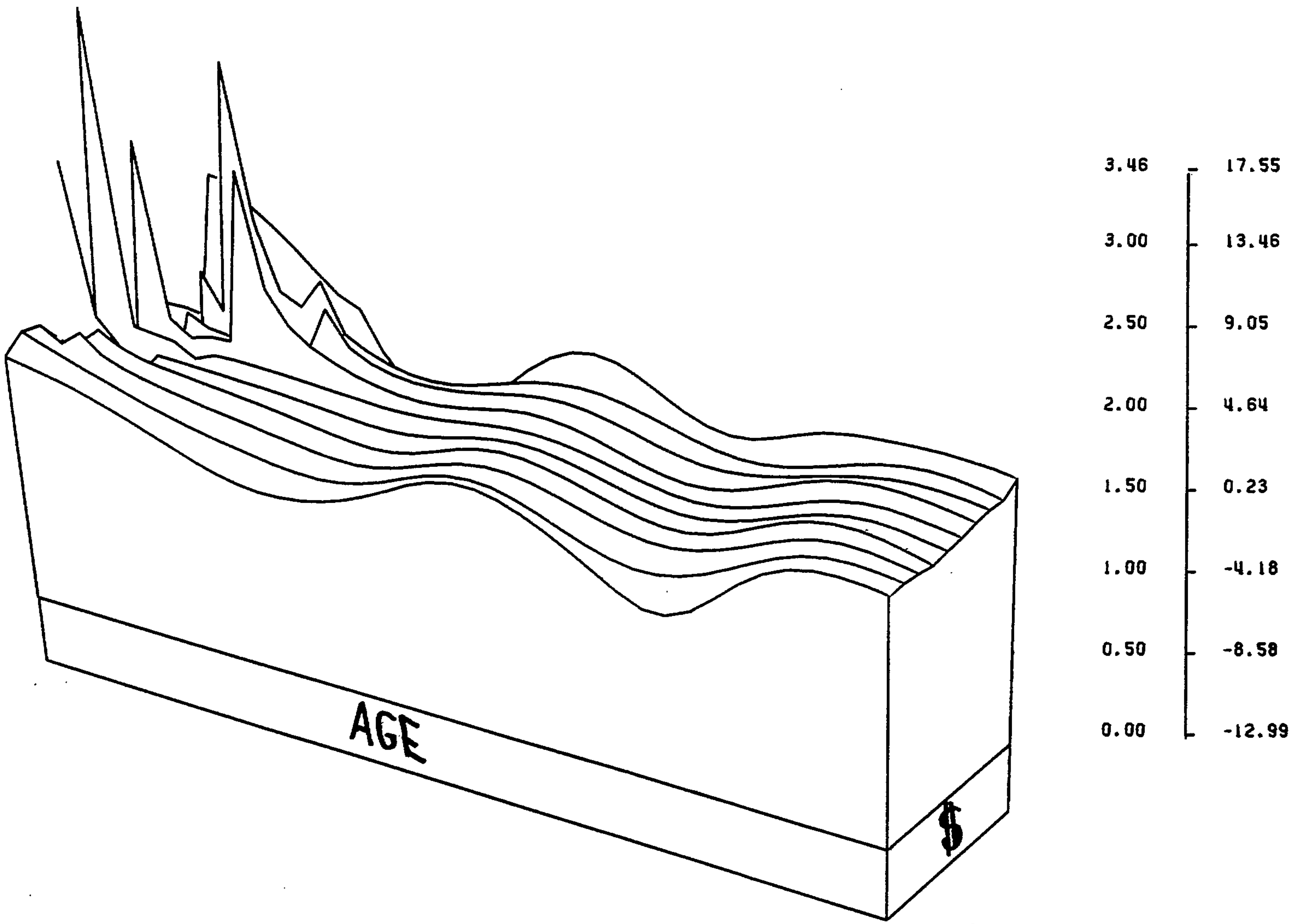
$a_5(t)$ = Fifth Component from Real Data



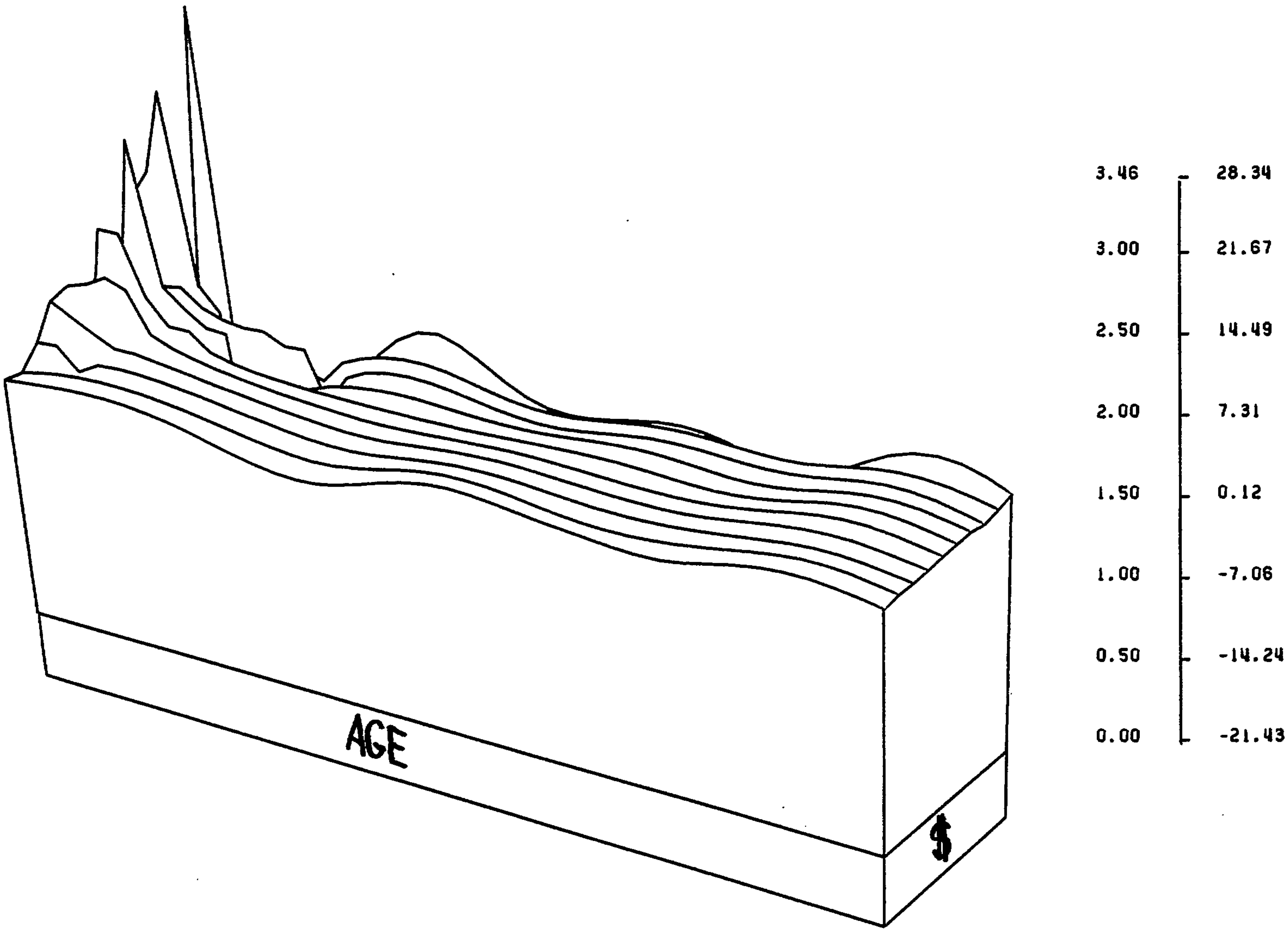
COMPONENTS OF MU FROM ACTUAL DATA - COMPONENT 1



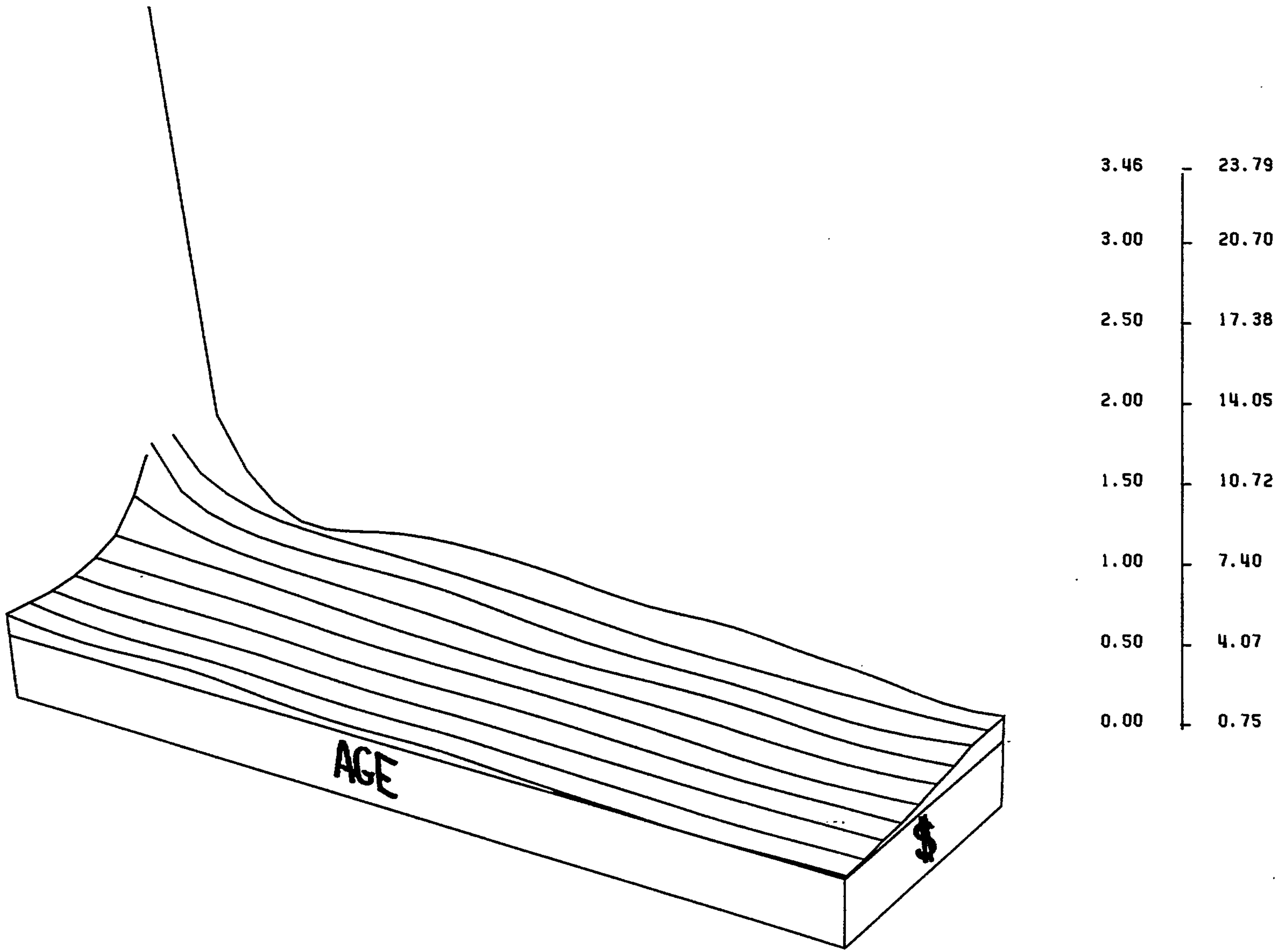
COMPONENTS OF MU FROM ACTUAL DATA - COMPONENT 2



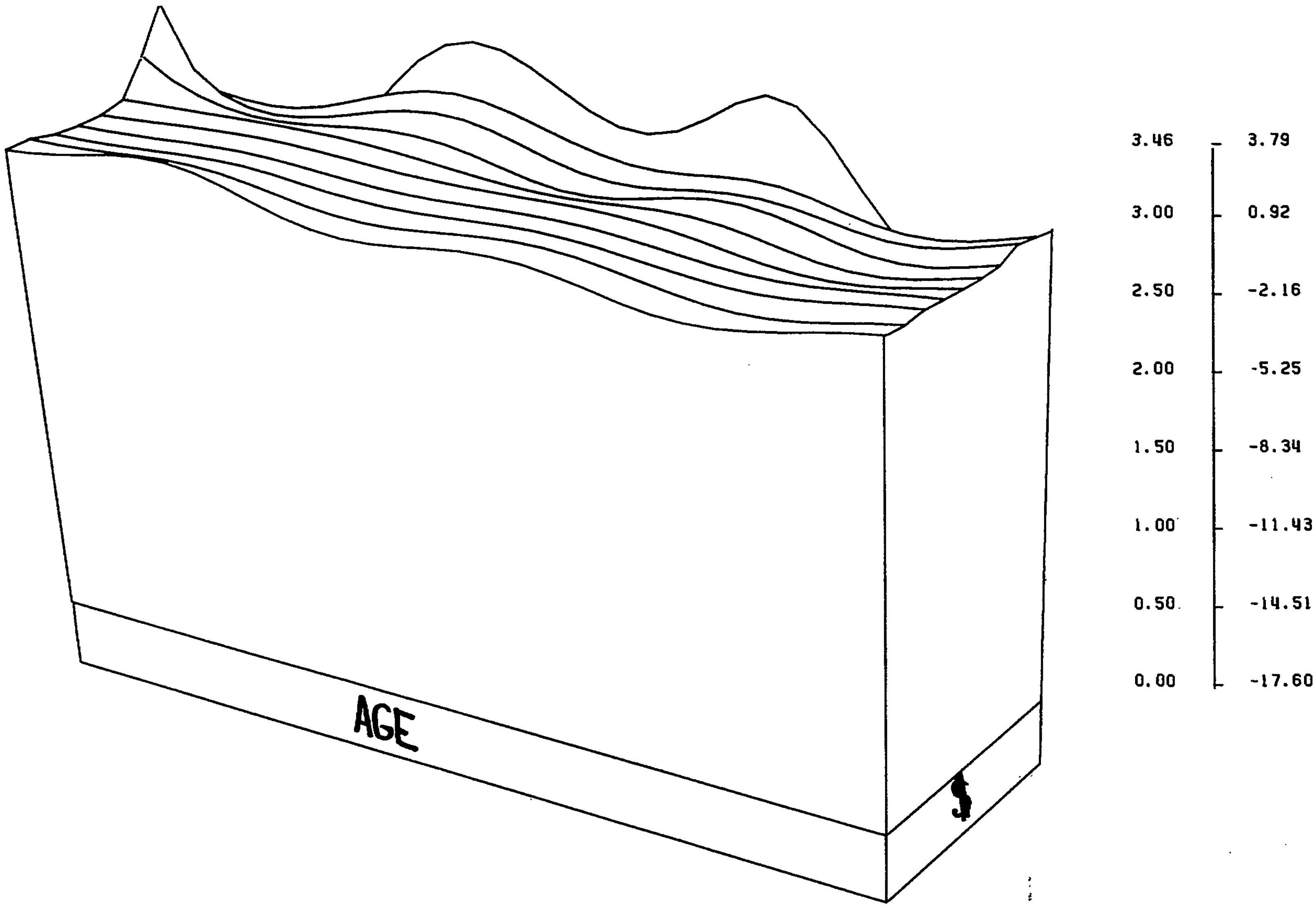
COMPONENTS OF MU FROM ACTUAL DATA - COMPONENT 3



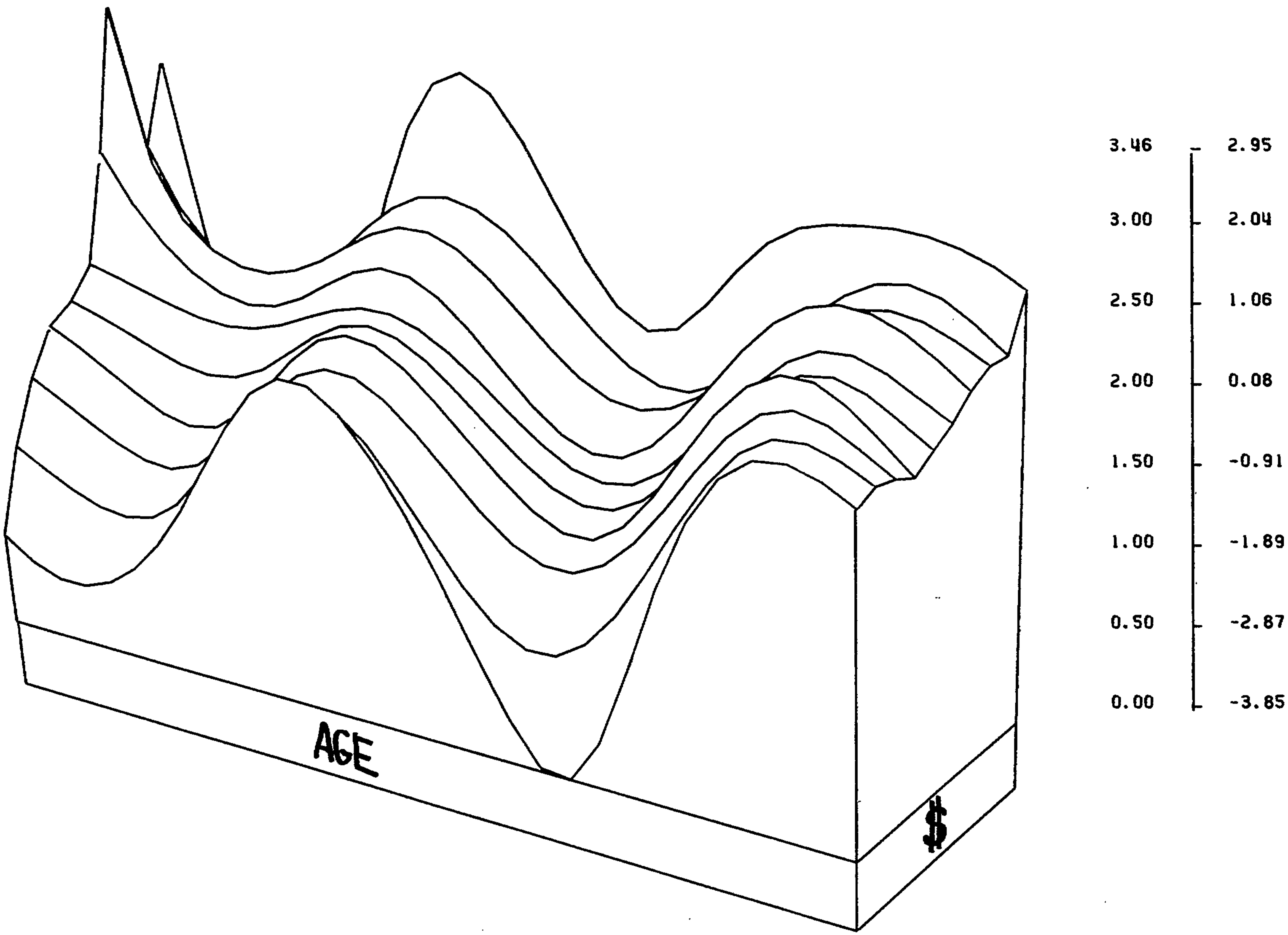
COMPONENTS OF MU FROM ACTUAL DATA - COMPONENT 4



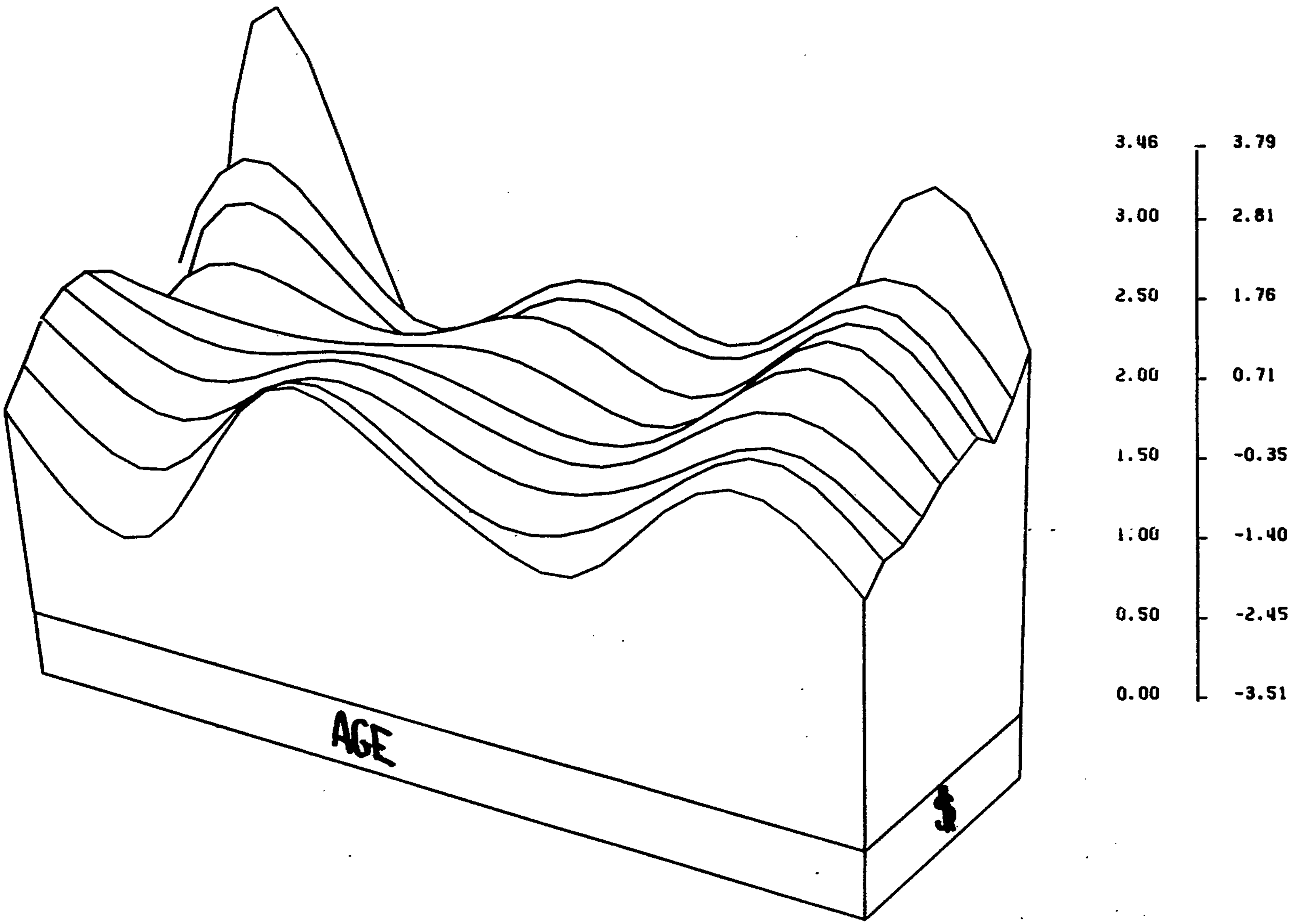
COMPONENTS OF MU FROM DATA - ELEVEN YEARS AGES 27 - 56 COMPONENT 1



COMPONENTS OF MU FROM DATA - ELEVEN YEARS AGES 27 - 56 COMPONENT 2



COMPONENTS OF MU FROM DATA - ELEVEN YEARS AGES 27 - 56 COMPONENT 3



COMPONENTS OF MU FROM DATA - ELEVEN YEARS AGES 27 - 56 COMPONENT 4

The computed results show the first two time components persistent under the bordering procedure. The plots of the age-income components also show extremely rough results on the low end of the age span. (Plots start at age 17, normally, even though the estimation data is bordered to exclude the tails of the distribution.) Since these high spikes dominate the scaling of the plot routine, the section from ages 27 to 56 was replotted to bring out greater detail.

In view of the high data aggregation, it is difficult to draw valid conclusions from the above results. The results suggest the presence of systematic components in the data; the results of our numerical experiments suggest that the results computed above may not be an accurate representation of what systematic terms are present. The only way to resolve the question is to process a less highly aggregated version of the data; we were not able to find a source of such data and have no knowledge of whether or not such data may be obtained (at an acceptable cost).

Finally, the estimation program included in the appendix is constructed to process sequential segments of the data in a manner analogous to the fertility parameter estimation package. The data we have obtained is of too short a duration to permit much work in this direction. With eleven years of data available, we have made two consecutive ten year runs; the results were inconclusive.

IV. Dynamic Interaction Models

A General Discussion

Once the coefficients of the partial differential equations in the case of the model have been estimated, the problem of determining a dynamical model for the evolution of these coefficients must be considered.

As mentioned in our interim report [3], it is clear that the evolution of these coefficients is affected by both economic levels and less easily quantifiable factors which might be described as sociological effects. Since it is evident that trying to model the latter must be a process subject to a wide range of possible approaches, it is of interest to seek some possible means of separating the two effects. The approach that has been adopted here is to attempt this separation on the basis of "time scales". In view of the often-noted inertia in social attitudes, it may be hoped that the dependence of the coefficient time functions on these factors is on a "much slower" time scale than the dependence on the more variable economic levels.

The use of this distinction is suggested by two sources. One is the use of the "method of multiple time scales", which has found success in numerous areas of applied mathematics. Another source is the widespread practice in econometric time-series work of "subtracting a trend line" from the available data. The usual interpretation of this procedure is that the trend represents an exogenous effect not included in the model. As will be seen below, the method proposed in a sense represents an alternative to this procedure usable in the case of relatively short data runs. In fact, it was the scarcity of the available data which has led us to attempt the approach below rather than to use conventional time-series fitting programs.

It is of course difficult to define precisely the meaning of the phrase "much slower" in the above, and one may no doubt cite instances of "rapid change" in social attitudes. In the practical use of the method described below, the allowable rate of change from exogenous effects is in effect specified by the user of the package. Whether or not the time scales are widely separated is evident in the results of the computation. One also expects that a model must be estimated by use of a "sliding subset" of the available data (in a manner analogous to the estimation procedures of sections II and III above). Periods of "rapid change" affecting the variables being modelled would probably appear as rapid variation in the estimated model coefficients.

The model and estimation procedure described in detail in Appendix B below is based on a linear model (state variable model) of the general form

$$x_{k+1} = F\, x_k + G\, u_k$$

$$y_k = H\, x_k$$

In the above, the input sequence $\{u_k\}$ represents the economic time series driving the model, while the output sequence $\{y_k\}$ represents the variable coefficient functions of the partial differential equations in the model core. In the usual control theory literature, it is assumed that the above system represents a so-called <u>minimal realization</u> of the system being modelled. This assumption means roughly that there is no component of the output sequence which is unaffected by the inputs. In the present case this assumption would rule out the presence of any exogenous component in the observed values of $\{y_k\}$, and in effect would preclude the inclusion of other than economic effects on the core model coefficients.

As shown below, it has been found possible to include the possibility of exogenous effects in the model above, and to weaken the requirement of minimality of the model. This modification makes it possible to fit models (based on the idea of a time scale difference in the essential dynamics) without resort to the necessity of introducing an ad-hoc subtraction of a trend to account for exogenous effects. The structure of the estimation algorithm described below is such that the exogenous component is implicitly determined by the algorithm at the same time that the model parameters are estimated.

The problem as formulated is that of finding an appropriate linear time-invariant model. One may either regard a linear model as appropriate for relatively small values of the variables involved, or alternatively, one may regard the fitting of a succession of linear models as a first step in the process of determining a possibly more realistic non-linear model. One may always hope that a relatively simple linear (or quasi-linear) model will turn out to be appropriate in the situation under study. Common sense dictates that the simplest possibilities be explored first, and this is the approach which has been taken.

A detailed description of the model formulation, and derivation of the estimation procedures is somewhat technical and is given below in Appendix B.

A serious issue in modelling the interaction effects is the treatment of the stochastic elements of the data. The presence of stochastic elements in the available economic indicator data is quite evident; in addition, the time functions produced by the estimation algorithms described above must contain errors as well. While it is felt that the algorithms used for coefficient estimation have a "smoothing effect", and that the level of noise on the final estimates (for the fertility data processing) is not excessive, it appears very difficult to produce quantitative estimates of its magnitude. The small amount of available data (relative to the estimates being made) makes it virtually impossible to formulate and estimate a statistical model of the disturbances.

While it is possible to formulate statistical estimation procedures in connection with models of the general type described below, these procedures fall basically into two types [19], [16]. The first are "correlation based" techniques, which are essentially based on "large sample" averaging procedures. The second type is a Bayesian or maximum likelihood estimator, the use of which requires a reliable statistical model for the sampling distribution.

Both of these procedures seem inapplicable in the present circumstance; the first because of the lack of a "large sample" over which to average, and the second because of the lack of sufficient data to estimate a useful sampling distribution.

The only remaining alternative appears to be the formulation of the problem in a deterministic manner, and this is what has been done below. Since a certain amount of noise in the actual data is inevitable, it is essential that the users of the package have a certain amount of appreciation for the effects of noise on the performance of the deterministic estimator developed below. It is probably only with a reasonable amount of numerical experience with the package that an ability to "see a reasonable fit" in the results can be developed. We expect that successful use of this package will require considerable experimentation, and that the results obtained will depend upon the experience and judgement of the user of the programs.

One may hope that the data runs are relatively noise free, although our experience indicates that this is not the case in connection with the economic data.

As a means of combatting the noise, one may suppose that the model to be fitted is of sufficiently large dimension to provide "internal filtering" of the noise in the economic indicator inputs. However, increasing the dimension of the model increases the data requirement for each parameter estimation. This decreases the number of estimates possible from the limited available data; the smaller number of estimates severely limits the user's ability to see consistency in fits produced from successive subsets of the data. Given the amount of presently available data, this does not appear to be a useable alternative.

One may also attempt to reduce the noise in the data by more or less ad-hoc methods. One such method is to replace the economic series with "moving averages". This has an interpretation similar to the large dimensional model mentioned above. The essential difference is that the filtering dynamics would in this case be supplied by the user of the program, while in the former case the filtering dynamics would be estimated from the data. Since the data requirement for the latter case is smaller, a larger number of estimates can be made from the available data in this case. This method may provide a way around the problem of a limited data base, but would require considerable numerical experimentation. There is no guarantee that a useful result could be obtained; even if consistent fits were obtained, there is a danger of lack of confidence in the results on the basis that nearly anything may be fit with enough free parameters available.

Another possible approach is to make an attempt to "remove the noise" from the economic time series before their use in an estimation algorithm. One possible rationale of such a procedure is to assume the existence of smooth "true" economic fluctuations, and to regard the observed series as consisting of the time series plus an additive noise term. If it is supposed that the "true" smooth series are the appropriate driving terms for this section of the model, then the smooth series are the appropriate ones to use in the identification procedure.

A problem with this approach, obviously, is that of extracting the "true" series from the noisy observations. It is conceivable that a Kalman-Bucy algorithm might be useful for this purpose; it is unclear, however, that the available data is sufficient to determine the required statistical parameters for this approach. Several runs (see below) have been made with data smoothed using the ad-hoc device of a least-squares polynomial fit based on a visual display of the data involved. Needless to say, this device is somewhat difficult to justify on a systematic basis. It also seems difficult to support use of an alternative in the case of the limited amount of available data.

A somewhat more mathematical treatment of the approach adopted is given in Appendix B. It is probably the case that the ability to successfully use the algorithms developed is dependent on understanding the material in Appendix B. The problems involved are not entirely elementary, and require a fair grasp of linear algebra and the theory of difference equations on the part of the reader.

B Tactical Use of the Algorithm

The problem of fitting linear models from input-output data must be regarded as a process subject to the use of a certain amount of judgement on the part of the individuals attempting the task. Using real data there will never be an exact fit, and one must recognize both the uncertainties in the estimation algorithm as well as the proposed subsequent uses of the model in evaluating the results of computations using real data.

At the heart of the estimation algorithms described in Appendix B below (and, indeed, essentially all algorithms for estimation of linear models) lies the numerical problem of solving a system of linear equations. As is well known, such a process may turn out to be numerically ill conditioned, in the sense that small changes in problem parameters may produce large variation of the computed answers.

The result given in Appendix B below essentially guarantees that (almost always) it is possible to solve the estimation equations for the estimated values of the model parameters. The result does not ensure that the computation is well-conditioned. Since the coefficient matrix involved is constructed from the observed data sequence, it is possible that the matrix occuring in a given computation is ill conditioned. The effect of this is to exaggerate errors (noise) in the data, and to produce inaccurate parameter estimates. The least-square algorithms employed in the programs of the appendix are "flagged" to alert the user to possible ill conditioning; it is still possible, however, for the package to make computations which may be unacceptable from this point of view. The user of the programs should be aware of this possibility. More than this, a "feel" for the conditioning of the computations is an invaluable asset in evaluating the results of attempted model fits.

In this context, just as in the case of estimation of the partial differential equation coefficients, one is faced with the problem of confidence in the computed results. One might take the whole of the available data, and perform a one-time computation of the parameters based on the algorithms described above. Of course, if the data were exact, the model dimensions correct, the system truly linear, and the computation well conditioned, this would produce the correct answer. Unfortunately, even if the model is incorrect, or the data noisy, there is a danger that the algorithm may produce a reasonable set of (essentially useless) parameters.

To get around this problem, we adopt a procedure analogous to that of the previous estimation problem. That is, we make the estimates on the basis of a <u>subset</u> of the available data; <u>consistency</u> in the results as <u>sequential subsets</u> of the data are processed is then taken to be indicative of a successful fit.

Of course, consistency in the results is somewhat a matter of judgement, and is closely connected with the issues of computational conditioning and data error discussed above. Before one concludes that the situation is truly desperate, however, it may be mentioned that numerical experiments attempting to force incorrect models on computer generated data sets typically generate <u>wildly inconsistent</u> sequential estimations. It may well be the case that while one may never by sure that a hypothesized model is <u>correct</u>, it is usually evident that a hypothesized model is <u>incorrect</u>.

There is a further constraint on an identified model in the present context arising from physical consideration. This constraint is that the "driven part" of the model should represent a <u>stable</u> system, i.e. that the eigenvalues of the matrix A_1 should be inside the unit disc of the complex plane. If this is not the case, the "free response" of the model will diverge; this is a situation not to be expected on a "physical" basis. (The FORTRAN and APL versions of the estimation program ESTIMATEA incorporate an eigenvalue calculation along with the estimation.)

In the derivation of the estimation equations in Appendix B, it is assumed that the system dimension n, and the characteristic polynomial corresponding to the exogenous output data drift are supplied by the user of the program. In principle, in the case of exact data, and a "true" linear system, both of these may be calculated from the given data.

In practice with real data, these must essentially be determined on a trial-and-error basis by the user of the program. These parameters, as well as the time series chosen as inputs, must be regarded as variables to be manipulated in order to achieve a reasonable sequentially consistent fit.

In the deterministic case, the system dimension n is essentially determined by the fact that the matrix inversion required in the parameter estimation becomes impossible when a model fit of dimension greater than n is attempted. In the case of noisy data, ill conditioning may be indicative of the same thing although large data errors may hide this.

As a matter of practical tactics, it is obviously advisable to attempt to fit lower dimensional models first. With a finite available observation record, this produces a greater number of sequential estimates on which consistency may be checked; since the linear system requiring solution is essentially of dimension $n(m+1)$, with n the state dimension and m the number of inputs, computations on lower dimensional models are more likely to be well conditioned. It is thus possible to have considerably more confidence in a low dimensional model than in one of higher dimension.

The choice of the dynamical character of the exogenous drift is a matter requiring some judgement and experience. In effect, the choice of the "drift dynamics" and the selection of the time series to be used as inputs must be made together. With some experience, it is possible to select likely combinations on the basis of graphs of the input-output data. The typical response characteristics of first and second order difference equations are well known; this knowledge can be used to advantage here.

For example, inspection of data graphs may suggest that the output to be modelled is the response of a single-input second order system offset by a straight line. Since the generation of an exogenous "straight line" requires two dimensions, (see below) one would attempt a model fit of state dimension four, with a single input and output. More complicated situations undoubtedly involve more trial and error.

It may be mentioned that it is essentially always possible to achieve a consistent fit by this procedure (for "smooth" output data) simply by supplying an exogenous drift containing enough linearly independent time functions to fit the observed output well. (Polynomials of sufficiently high degree in principle would suffice.) The identification scheme should then identify an essentially zero input matrix, and attribute <u>all</u> of the observed output to the drift term. In a qualitative sense, large dimension of the drift block required to obtain a fit may be taken as an indication that the whole output is nearly exogenous, i.e. not due to the supposed input series. As one may see, there is considerable leeway available to the user of the program in this regard. The rationale of this whole approach, however, demands "conservative" treatment of the drift term.

Since the exogeneous drift is supposed to operate on a slower time scale than the economic (or other) interaction effects being modelled, one should, in principle, try slowly varying functions as the exogenous component. If it is suspected on other grounds that a relatively rapidly varying exogenous component is present in the data, however, this suggests use of an exogenous drift of similar character. Such decisions are essentially a matter of judgement, and it is possible to give no precise rules in this regard.

The computer program ESTIMATEA requires as input the q (non-leading) coefficients of the characteristic polynomial of the exogenous drift matrix **D** . These coefficients are chosen on the basis of the time functions allowed to be present in the exogenous drift; the connection between allowable drift terms and required polynomial follows from the theory of z-transforms [14].

In general, any (discrete) time function with rational z-transform is allowable as an exogenous component. If it is supposed that the drift is representable in the form

$$d_k = \sum_{i=1}^{m} \alpha_i \varphi_k^{(i)} \quad ,$$

where the unknown scalars $\{\alpha_i\}$ are determined in the identification process) then the input to the program is the vector of non-leading coefficients of the least common denominator of the z-transforms $\{\hat{\varphi}^{(i)}(z)\}$ of the functions $\{\varphi_k^{(i)}\}$. For reference, a chart of common drift forms is given below. It should be emphasized that the algorithm requires only (a guess of) the functional form of the exogenous term. The coefficients α , β , γ , etc. are effectively computed by the algorithm, and are not supplied by the program user.

Drift Chart

Description	Time Function Form	Characteristic Polynomial	Program Input
constant:	α	$\lambda-1$	-1
linear "trend":	$\alpha + \beta k$	$(\lambda-1)^2$	$-2 \quad 1$
parabola:	$\alpha + \beta k + \gamma k^2$	$(\lambda-1)^3$	$-3 \quad 3 \quad -1$
sinusoid:	$\alpha \cos wk + \beta \sin wk$	$(\lambda^2-2 \cos w\lambda+1)$	$-2 \cos w \quad 1$
exponential:	$a \gamma^n$	$\lambda-\gamma$	$-\gamma$
trend + sinusoid	$\alpha + \beta k + \gamma \sin(wk+\theta)$	$(\lambda-1)^2(\lambda^2-2 \cos w\lambda+1)$	$-(2+2 \cos w) \quad (2+4 \cos w) \quad -(2+2 \cos w) \quad 1$
general:	$\sum_{i=1}^{m} \alpha_i \varphi_{(k)}^{(i)}$	l.c.d of $\{\varphi^{(i)}(\lambda)\}$ $= \lambda^q + d_q \lambda^{q-1} + \dots + d_1$	$d_q \; d_{q-1} \cdots d_1$

C Real Data Experiments

We have run some experiments with the estimation algorithm discussed above in an attempt to determine the interaction between economic indicator levels and the fertility equation coefficients estimated above. Since the output generated by the estimation algorithm program is somewhat long, only abbreviated versions of the results may be presented in this report. It is hoped, however, that these examples will provide useful guidance for the users of the algorithm.

As mentioned above, the fertility coefficients have been estimated on a yearly basis; processing the available data provides estimates for the years 1958-1970. The available data thus consists essentially of only thirteen consecutive values for which one expects a consistent level of error.

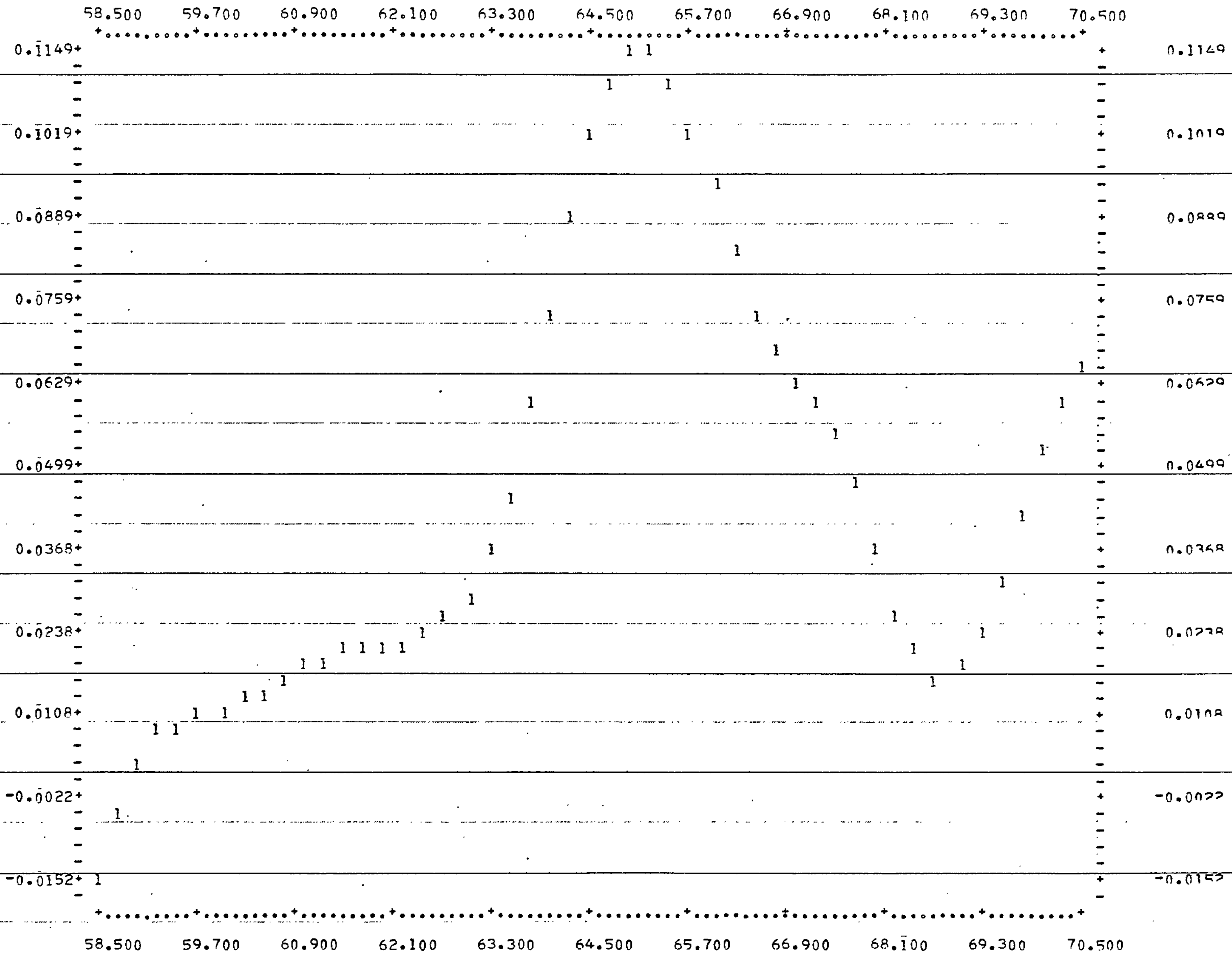
Econometric time series data is commonly available on a quarterly basis. The approach that has been taken here is to assume smoothness (and a low level of error) in the estimated fertility coefficients. A spline interpolation has been used to generate values on a quarterly basis. Econometric data on a seasonally adjusted basis has been used wherever available. The time series used together with a numerical code utilized in the program are listed in the following table.

Input Series Table

Number	Description
1	Industrial Wages
2	Manufactoring Wages
3	Housing Completions
4	Personal Income
5	Disposable Income
6	Participation Rate, Total
7	Participation Rate, Females
8	Unemployment Rate
9	Industrial Production
10	Housing Completions - Four Quarter Average
11	Housing Completions - 4th Degree Polynomial Fit
12	Participation Rate - 4th Degree Polynomial Fit
13	Unemployment - 6th Degree Polynomial Fit

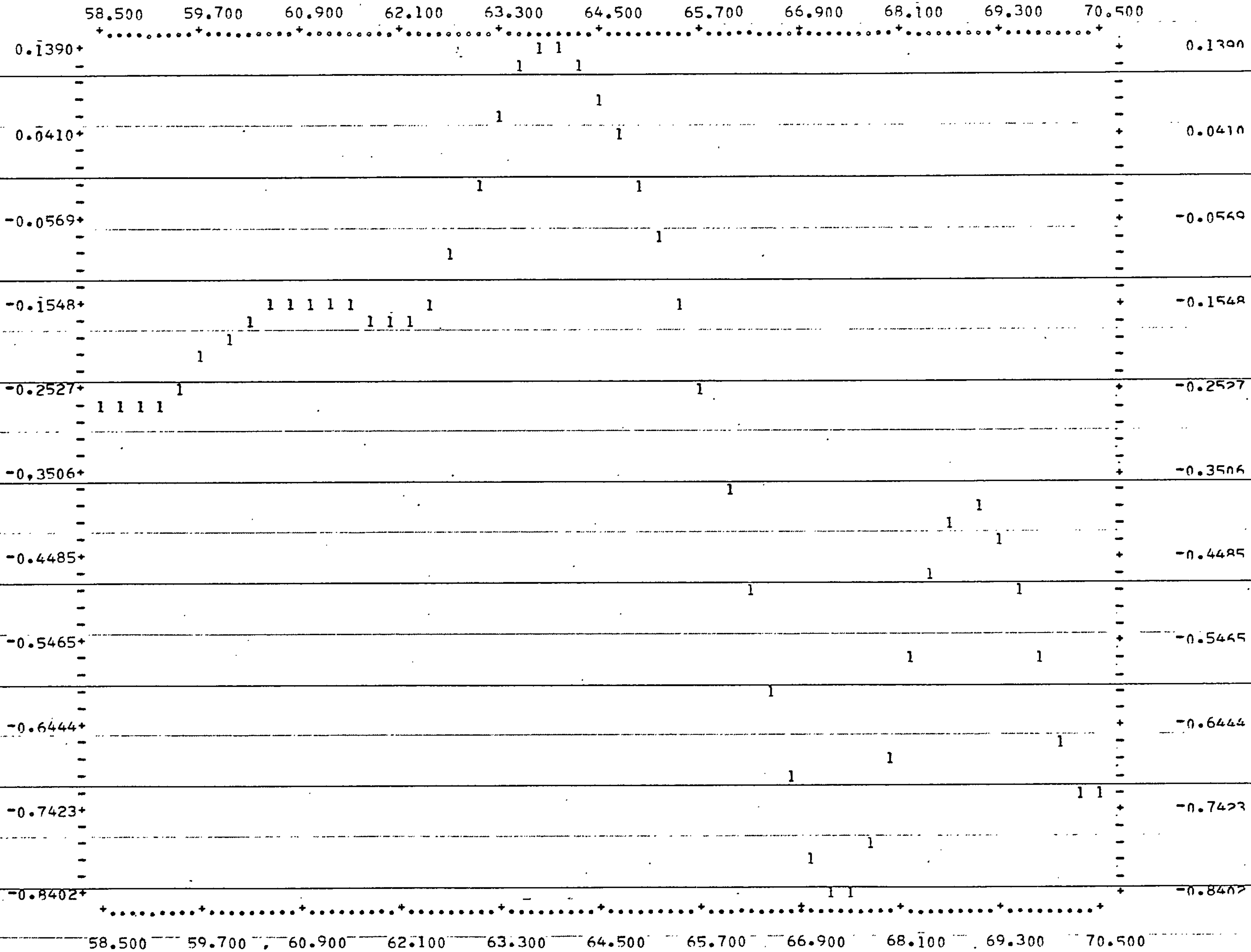
The four identified fertility coefficient series (on a quarterly basis) and the above econometric series are displayed in graphical form below.

CHART 21



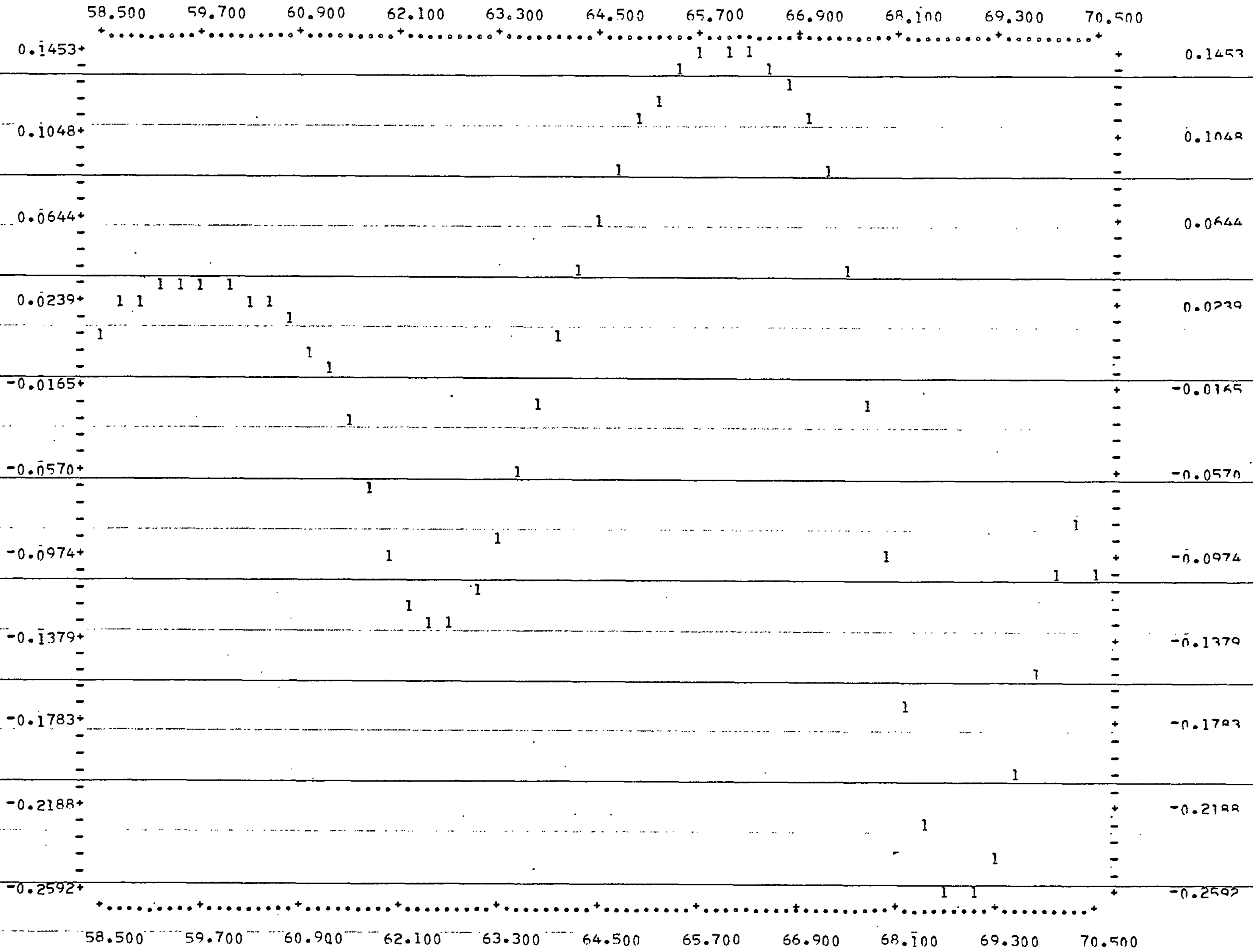
b(t) - Rate of Change of "Family Size"

CHART II



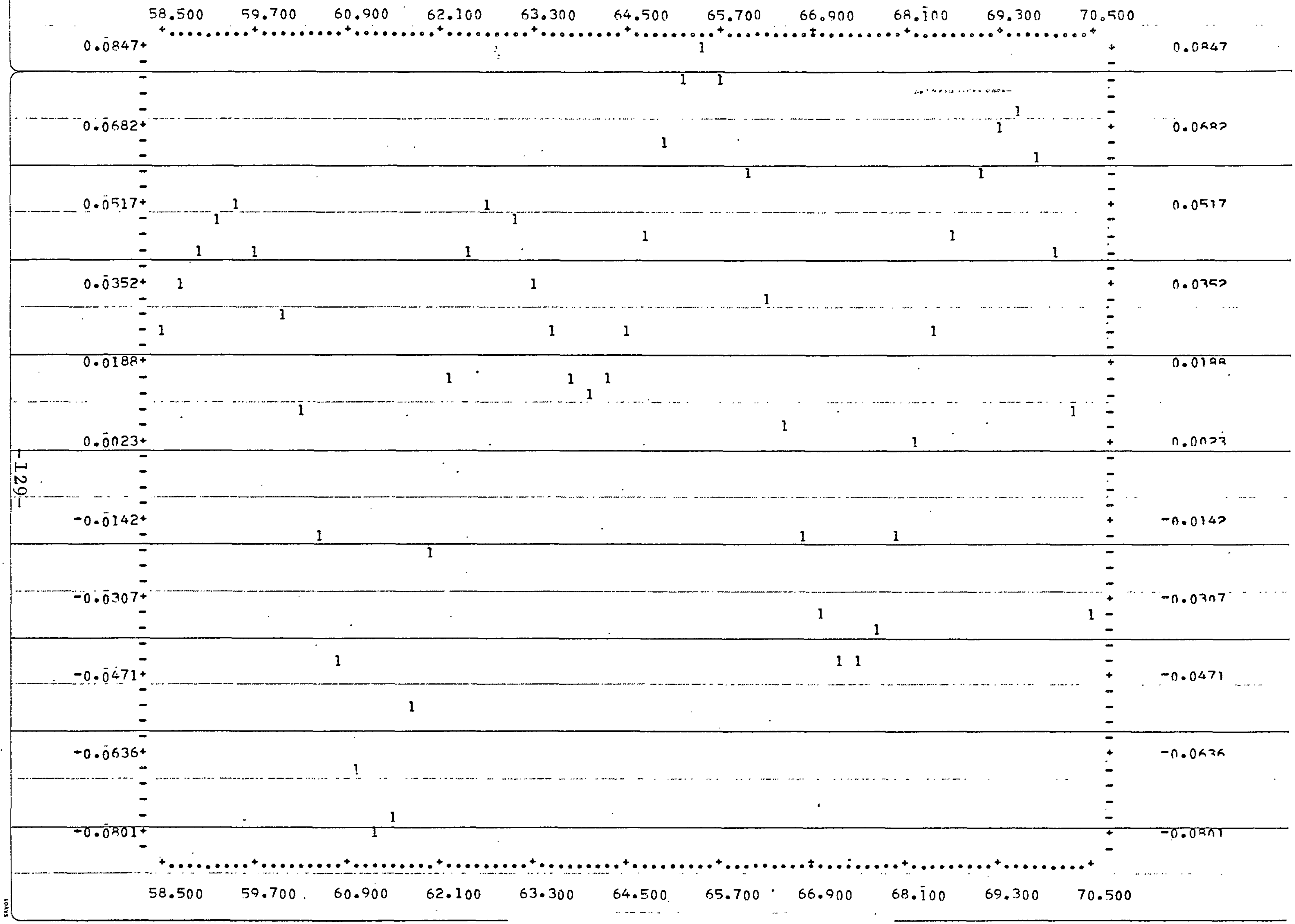
$a_1(t)$ - First Fertility Time Component

CHART 12



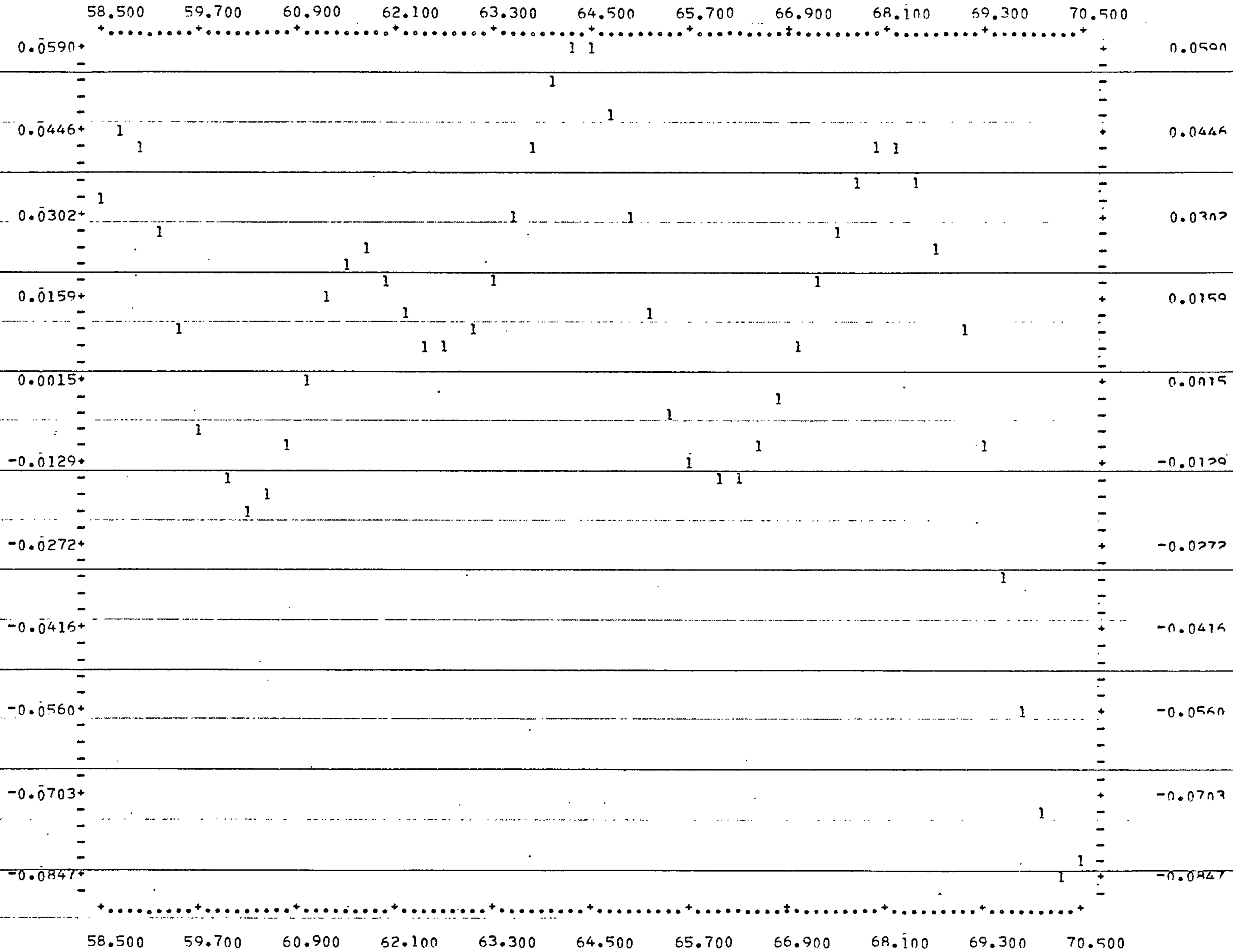
$a_2(t)$ - Second Fertility Time Component

CHART 13



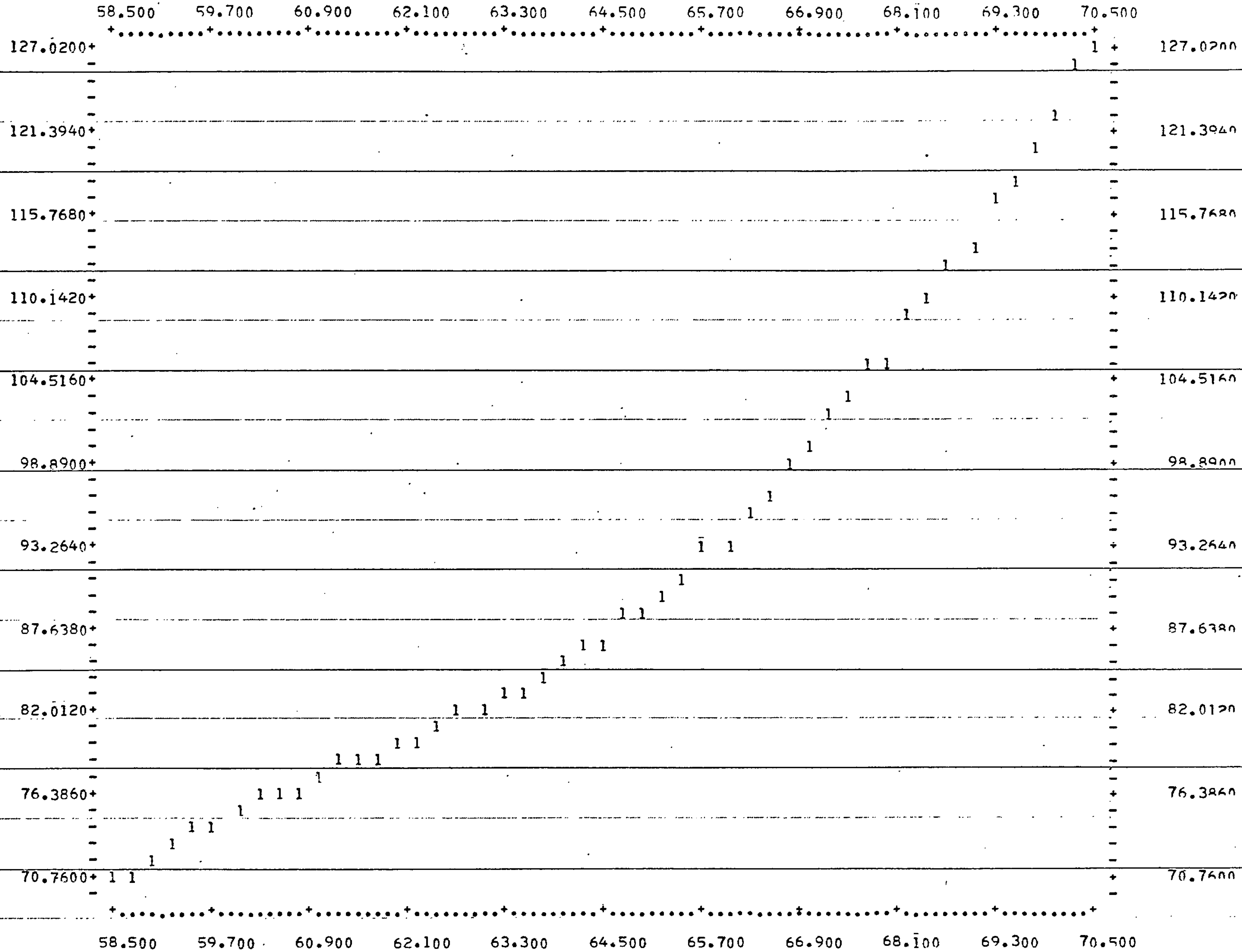
$a_3(t)$ - Third Fertility Time Component

CHART 14



$a_4(t)$ - Fourth Fertility Time Component

CHART 1



Industrial Wages

CHART 2

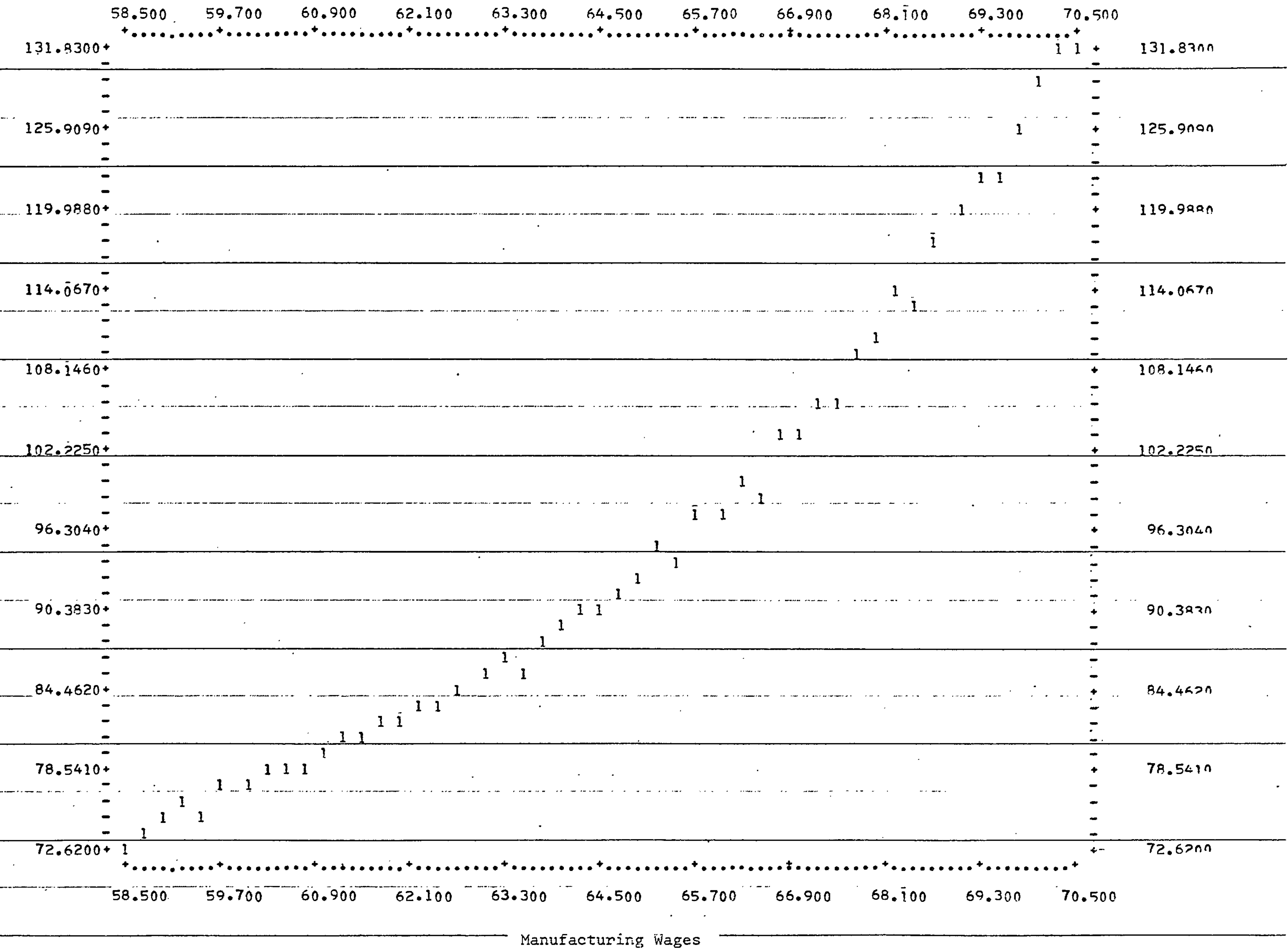
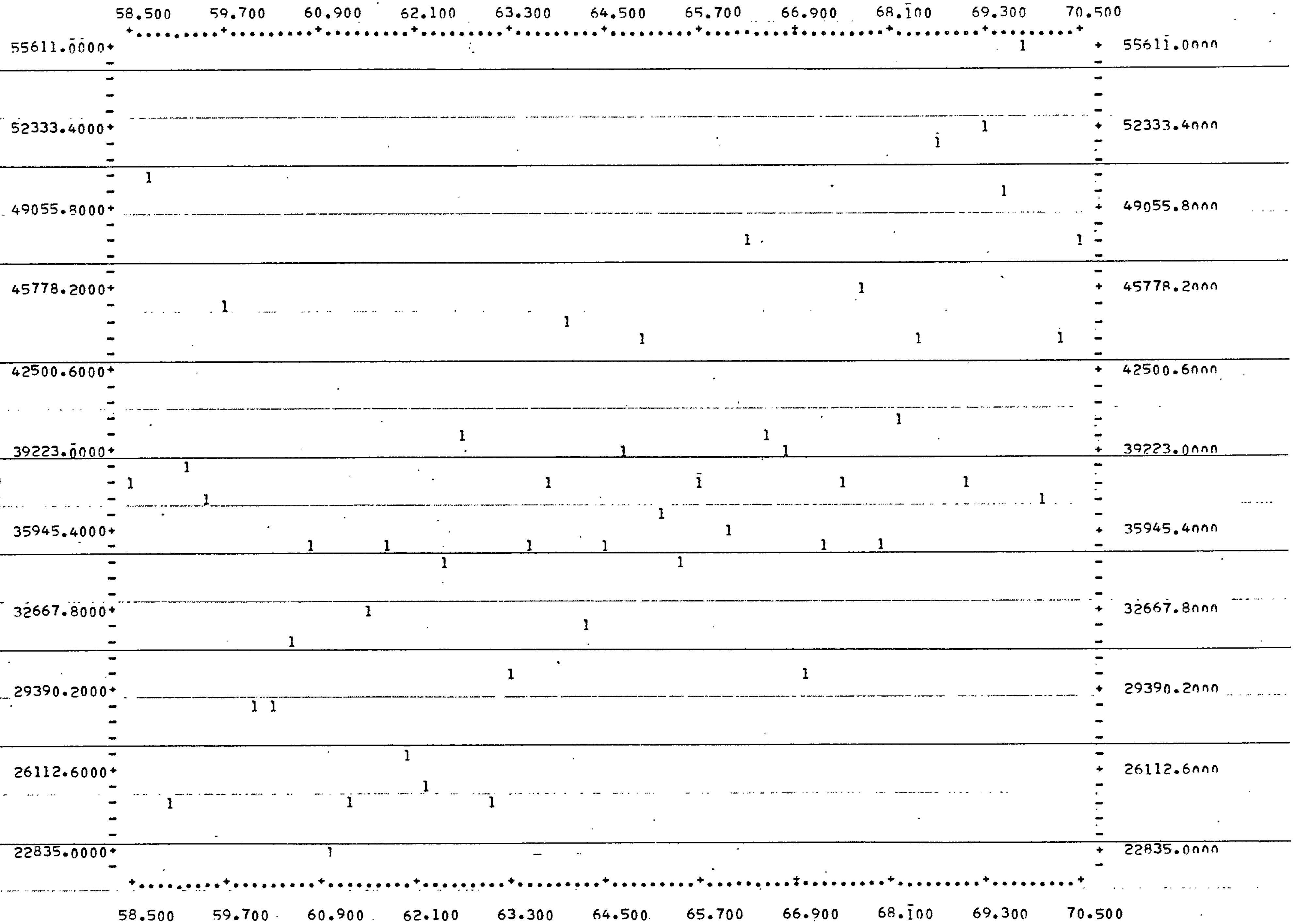
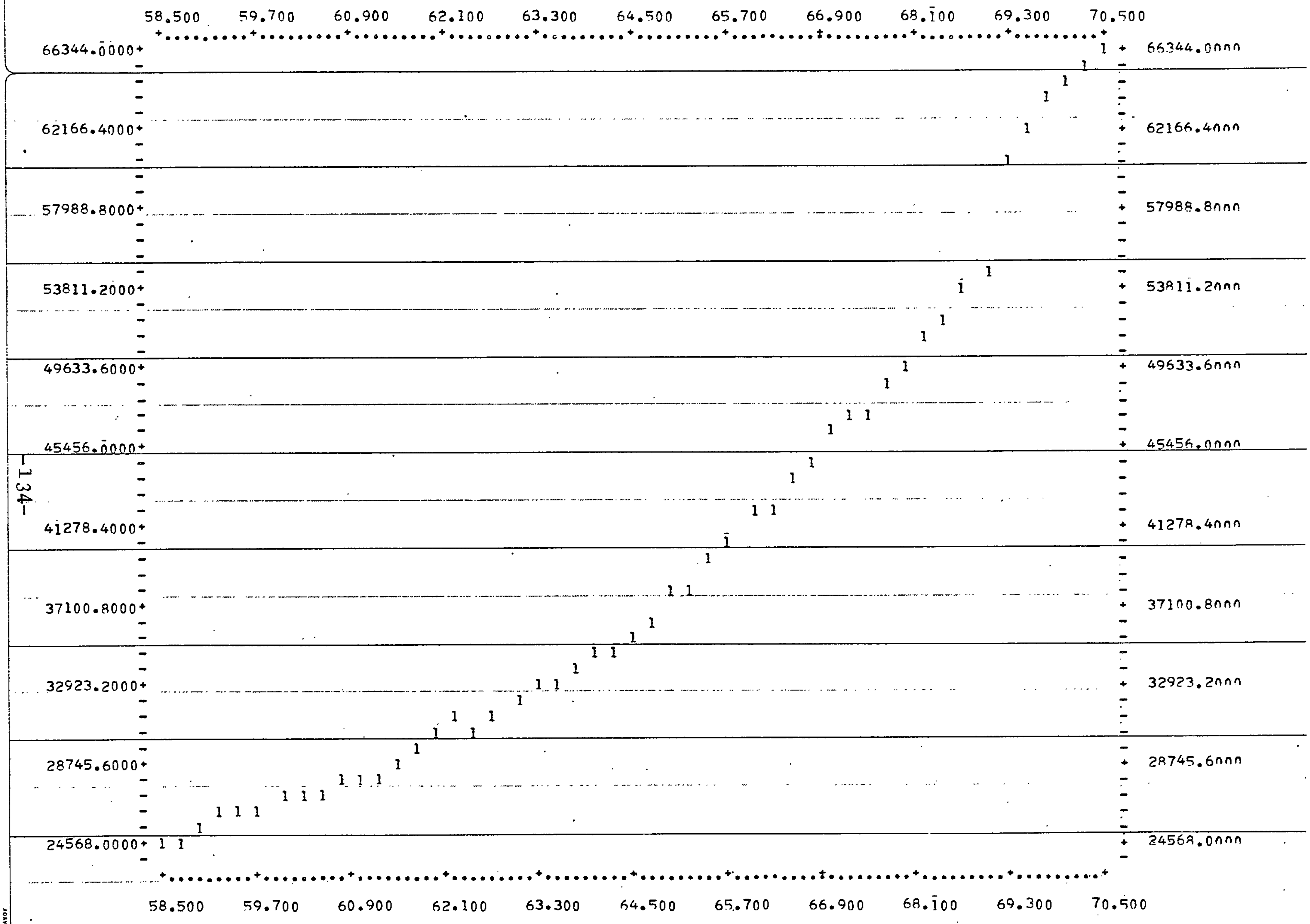

58.500
59.700
60.900
62.100
63.300
64.500
65.700
66.900
68.100
69.300
70.500
131.8300
125.9090
119.9880
114.0670
108.1460
102.2250
96.3040
90.3830
84.4620
78.5410
72.6200
Manufacturing Wages

CHART 3



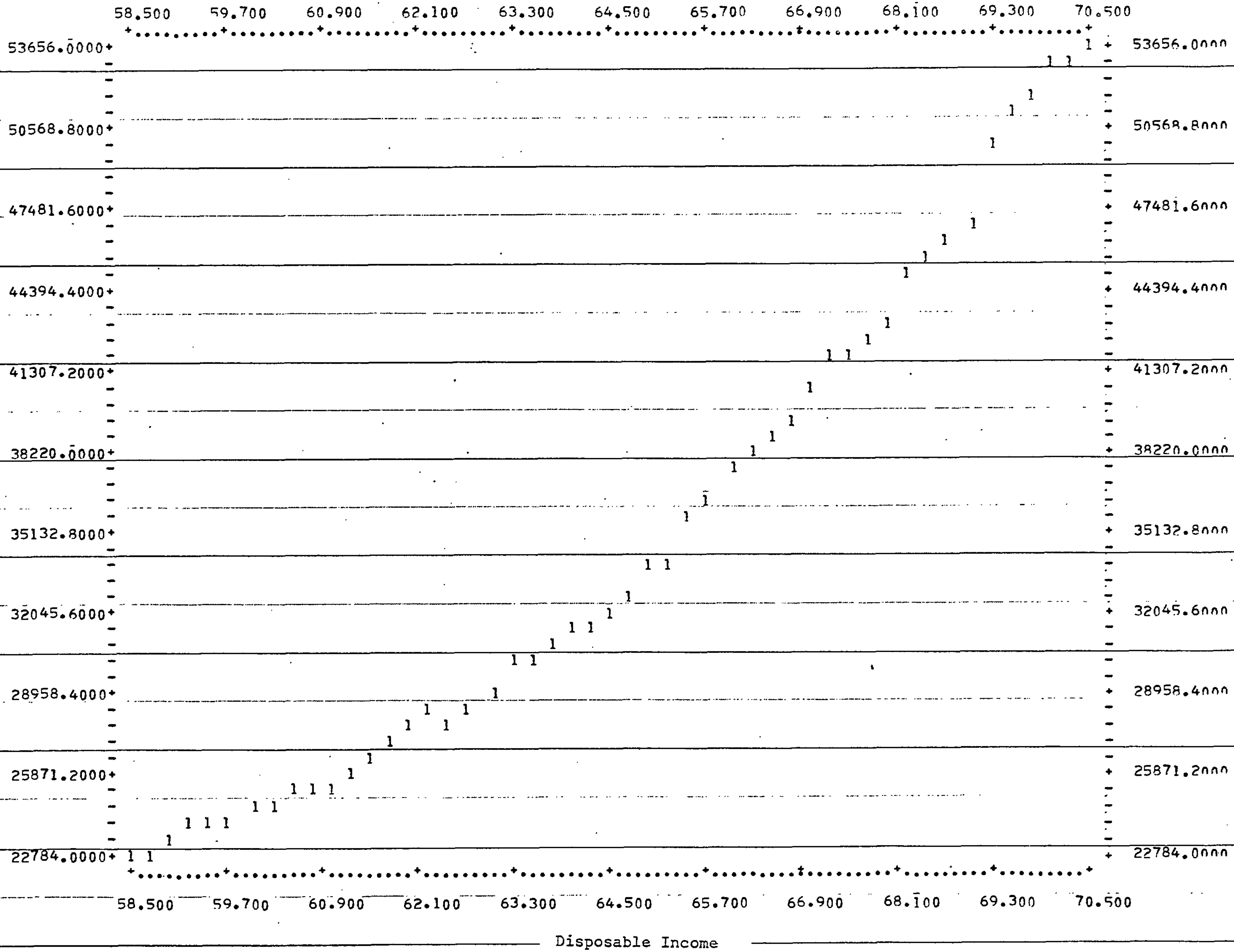
Housing Completions

CHART 4



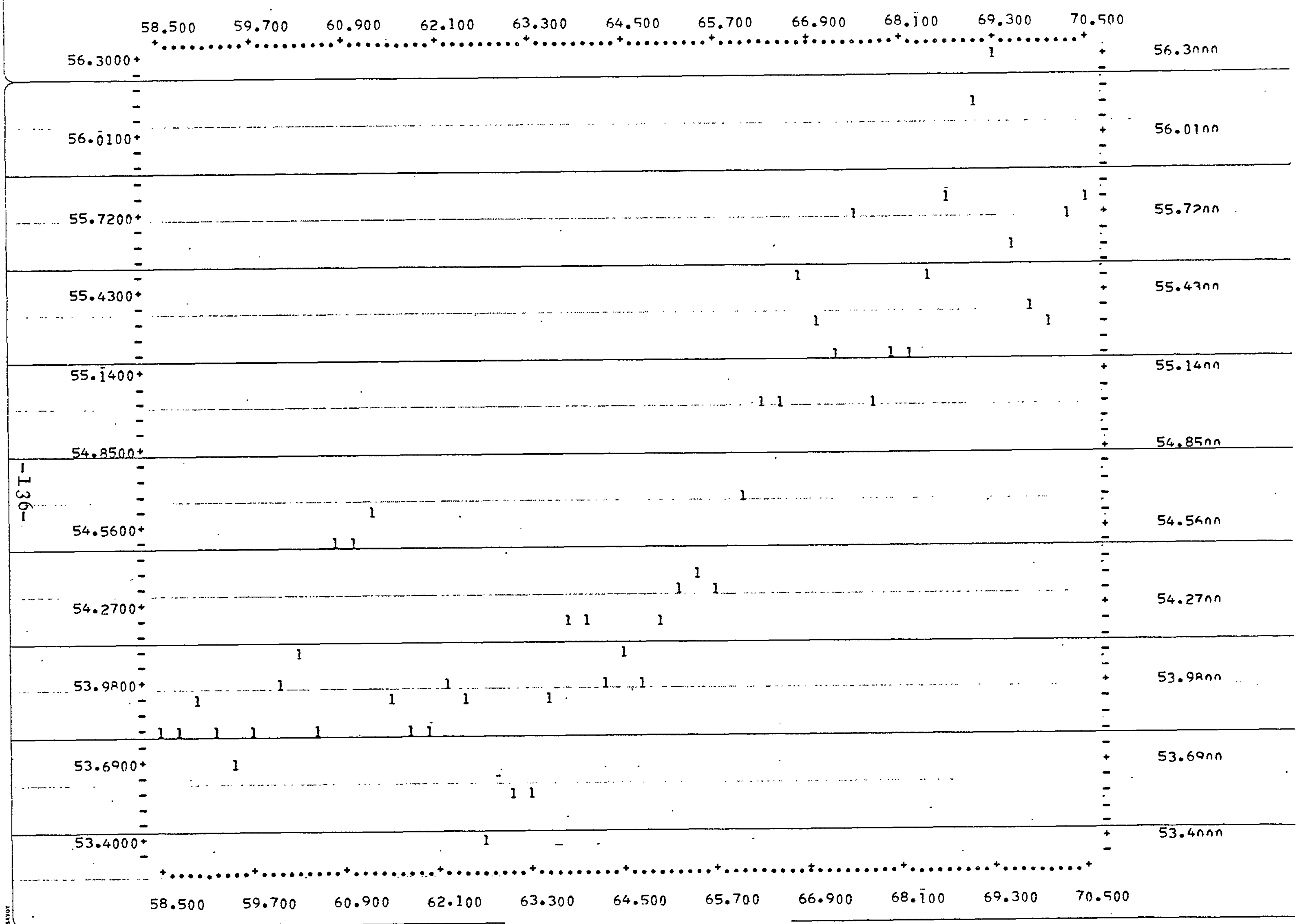
Personal Income

CHART 5


58.500
59.700
60.900
62.100
63.300
64.500
65.700
66.900
68.100
69.300
70.500
53656.0000
50568.8000
47481.6000
44394.4000
41307.2000
38220.0000
35132.8000
32045.6000
28958.4000
25871.2000
22784.0000
Disposable Income

-135-

CHART 6



Participation Rate - Total

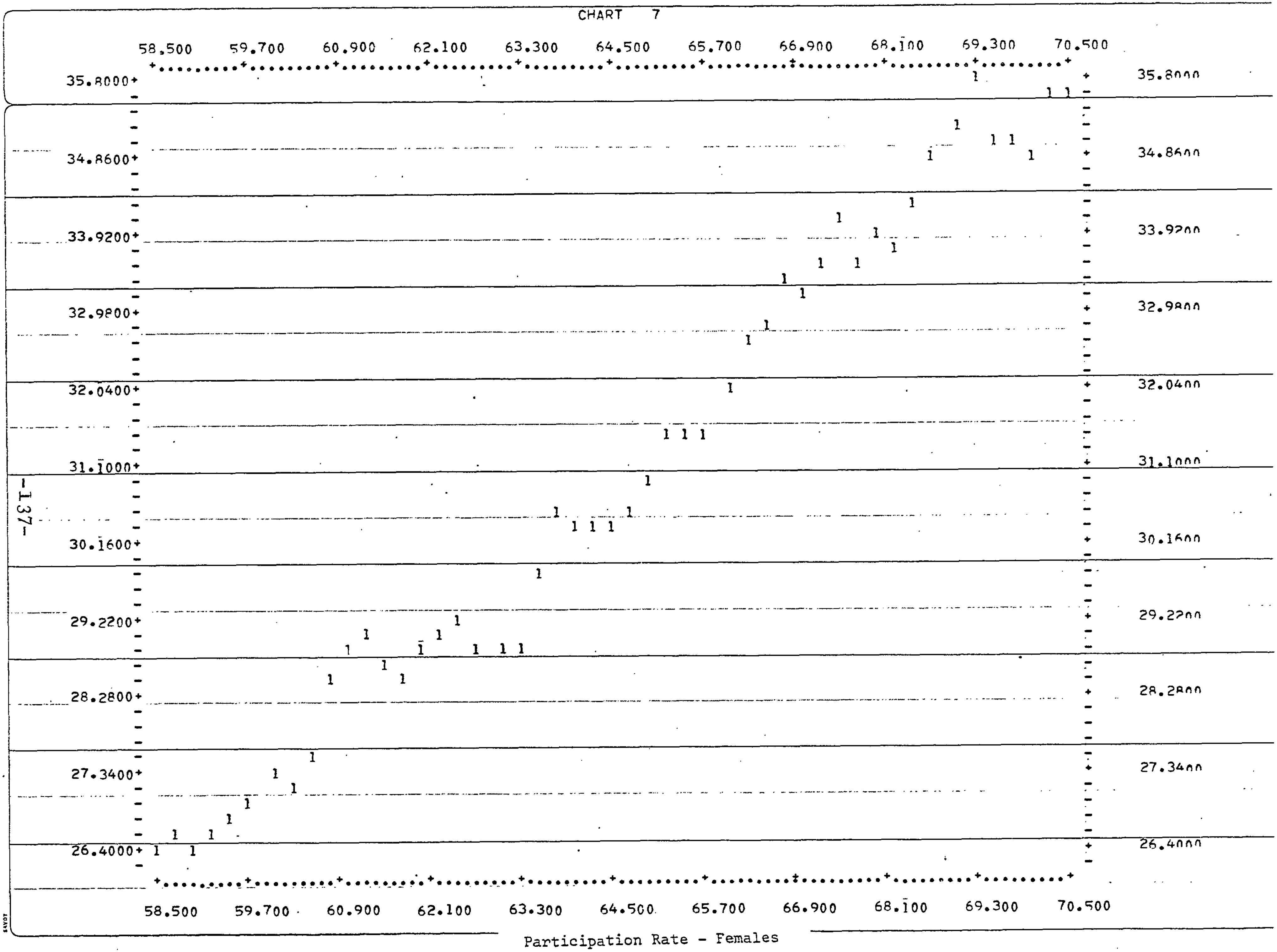
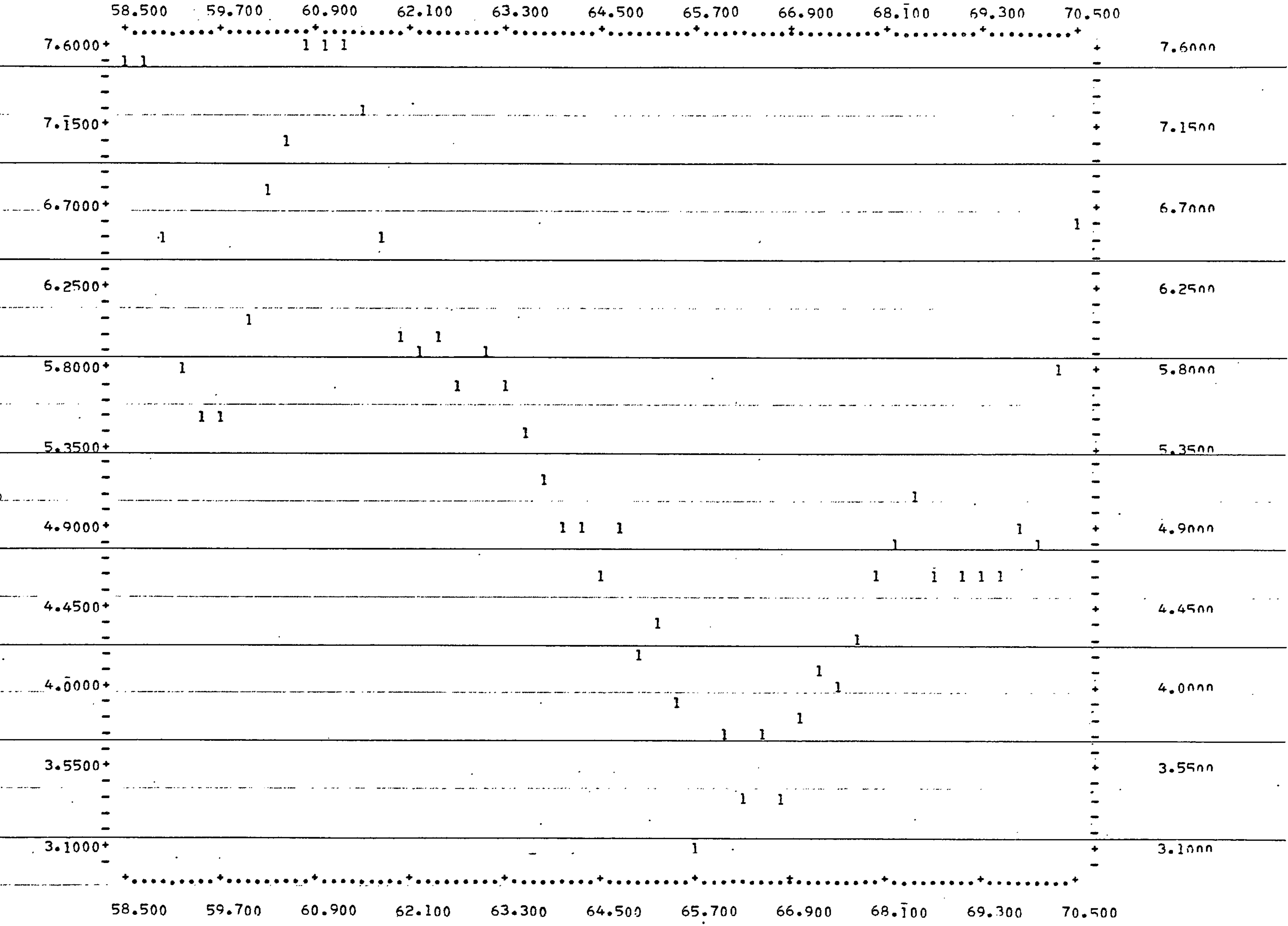


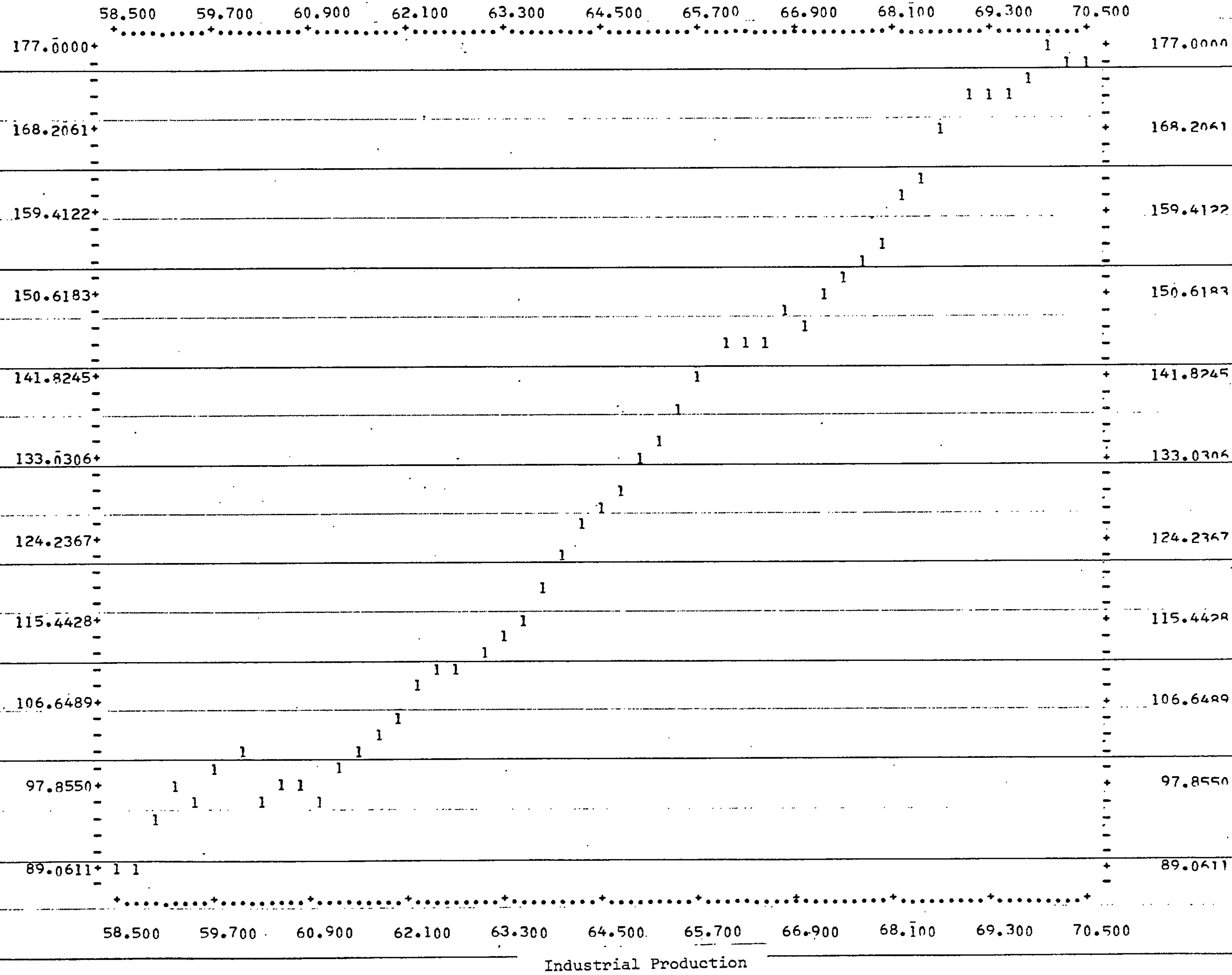
CHART 7
58.500
59.700
60.900
62.100
63.300
64.500
65.700
66.900
68.100
69.300
70.500
35.8000
34.8600
33.9200
32.9800
32.0400
31.1000
30.1600
29.2200
28.2800
27.3400
26.4000
Participation Rate - Females

CHART 8



Unemployment Rate

CHART 9



58.500
59.700
60.900
62.100
63.300
64.500
65.700
66.900
68.100
69.300
70.500
177.0000
168.2061
159.4122
150.6183
141.8245
133.0306
124.2367
115.4428
106.6489
97.8550
89.0611
Industrial Production

-139-

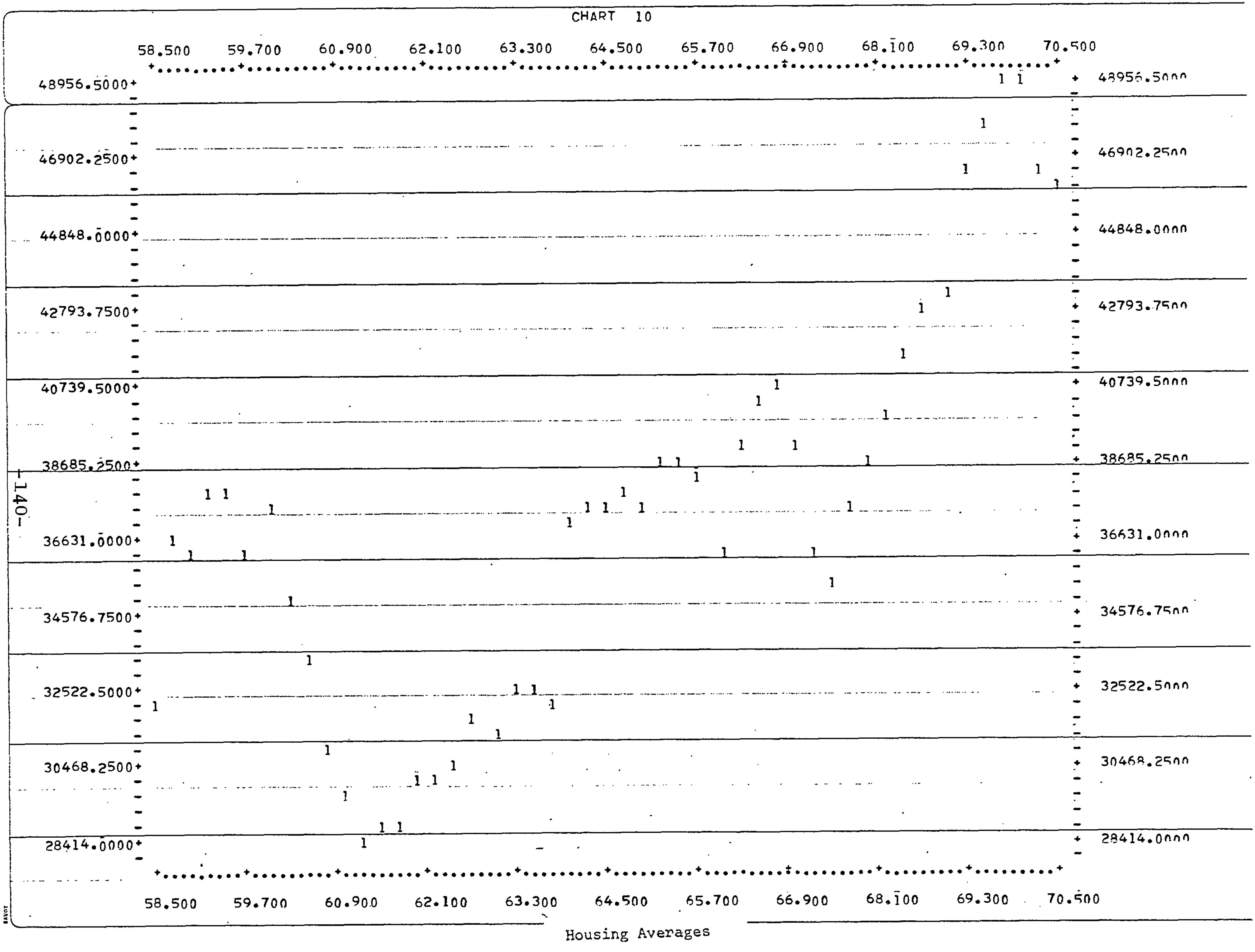
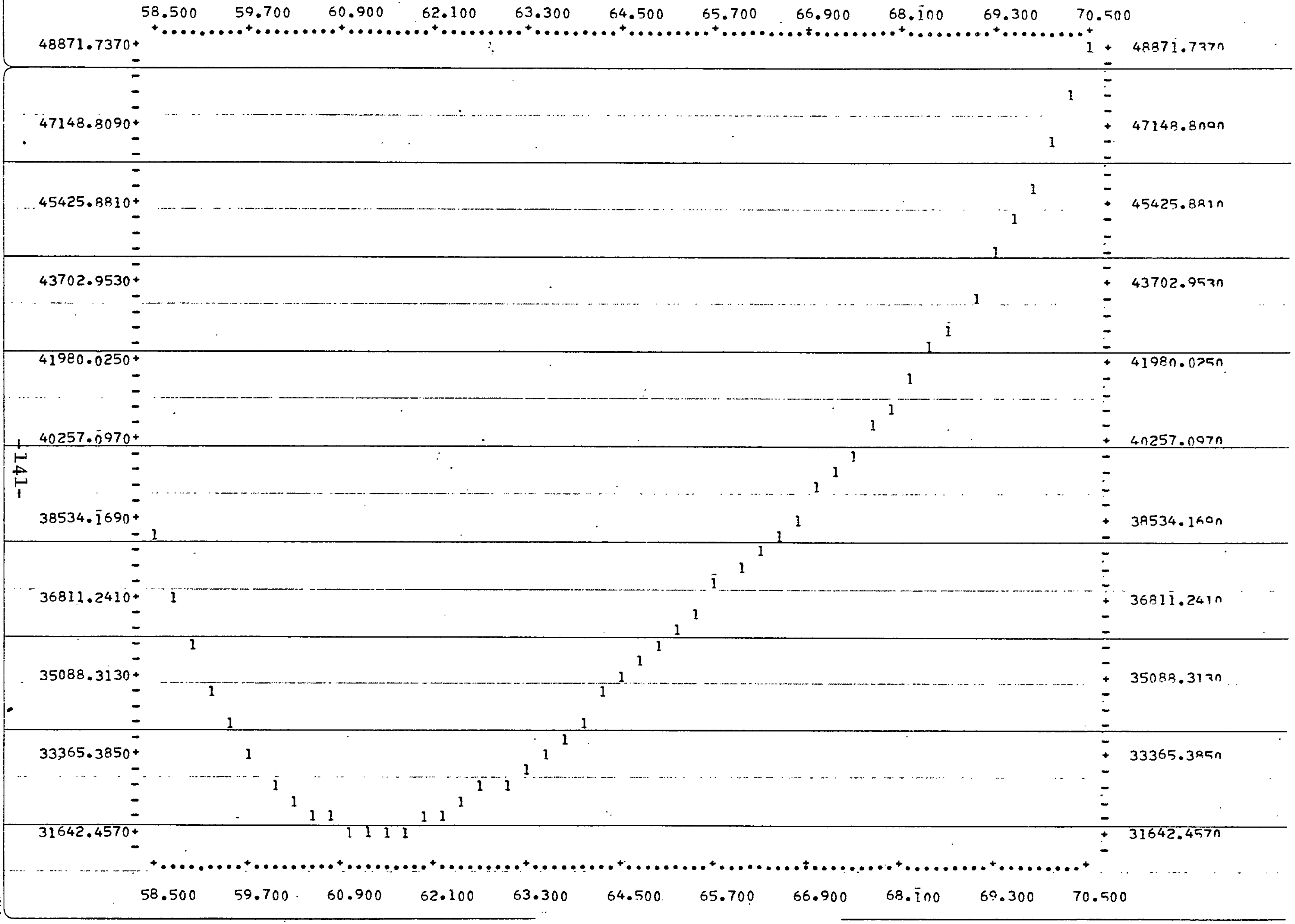


CHART 10
58.500
59.700
60.900
62.100
63.300
64.500
65.700
66.900
68.100
69.300
70.500
48956.5000
46902.2500
44848.0000
42793.7500
40739.5000
38685.2500
36631.0000
34576.7500
32522.5000
30468.2500
28414.0000
Housing Averages

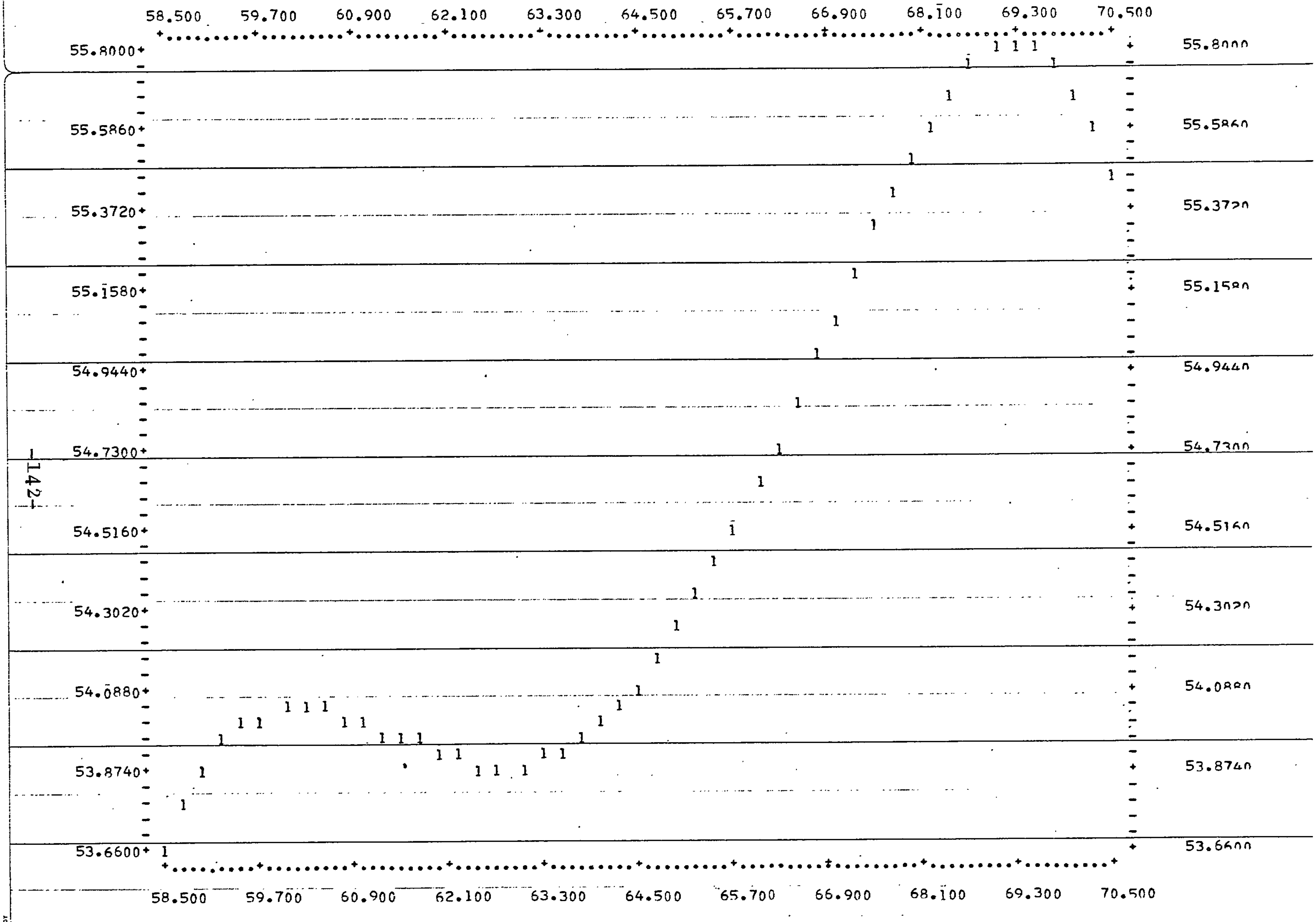
-140-

CHART 11



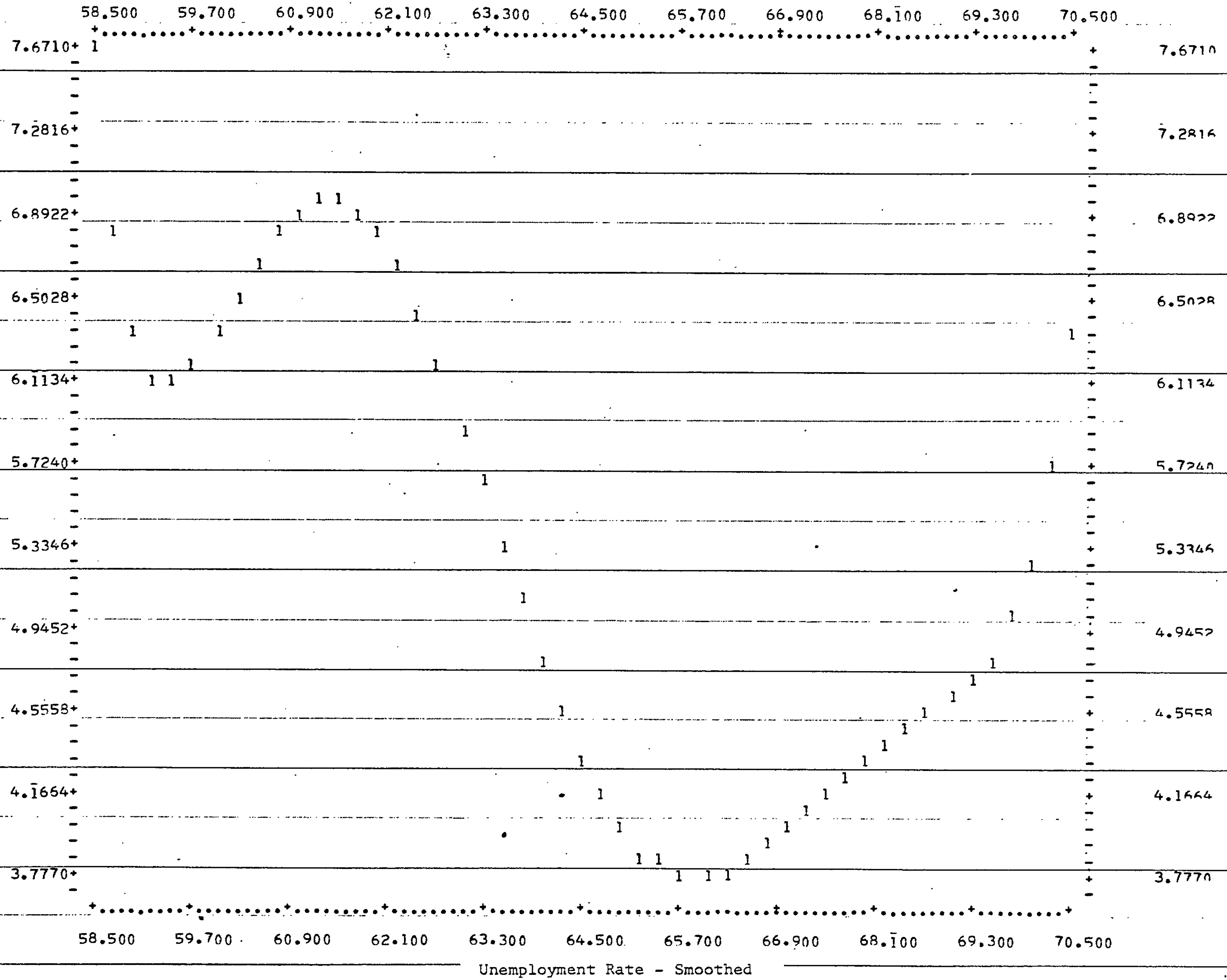
Housing Averages - Smoothed

CHART 12



Participation Rate - Total - Smoothed

CHART 13



58.500 59.700 60.900 62.100 63.300 64.500 65.700 66.900 68.100 69.300 70.500
7.6710
7.2816
6.8922
6.5028
6.1134
5.7240
5.3346
4.9452
4.5558
4.1664
3.7770
Unemployment Rate - Smoothed

It is evident from the plots above that the time series one might judge the most likely candidates for the series driving the fertility coefficients are also the series with the highest apparent level of noise. All of the problems mentioned above in the context of the use of the proposed algorithm with noisy data therefore arise. As a working hypothesis, we suppose that the fertility coefficient series determined above are essentially deterministic. Since the series reproduce the observed behaviour with reasonable accuracy, we also assume a relatively low level of error here. (The available data makes error estimates essentially impossible; there is little apparent alternative to this assumption.) Since the coefficient $b(t)$ ("rate of change of average family size") is essentially directly measurable, this series undoubtedly has the smallest error.

In the tables below we list the results of computer runs made using $b(t)$ and $a_1(t)$ is output. Almost all of the runs were made with use of either constant, linear, or parabolic exogenous drift, and a variety of the techniques mentioned above were employed in an attempt to get a reasonable fit.

b(t) Estimation Runs

State Dimension	Inputs	Drift	Result
5	4, 8, 9, 10	linear	Inconsistent F
5	7, 5, 3, 8	"	"
5	6, 7, 8, 9	parabolic	"
5	6, 8, 9, 1	"	Some consistency in F , just two columns of G .
5	7, 5, 3, 8	"	Fairly consistent F , smooth variation in G .
6	3, 8, 9	"	F somewhat variable
5	5, 7, 8, 9	"	Fairly consistent F ; second, third columns of G fair.

$a_1(t)$ Estimation Runs

State Dimension	Inputs	Drift	Results
3	8^*	constant	Fair F , very inconsistent G
3	13^*	"	Some consistency in F , G variable
3	12^*, 13^*	linear	Variable F , G .
4	8	constant	F , G inconsistent
4	1^*	linear	Occasional consisting in F , G variable
4	12^*, 13^*	linear	Fair F , very inconsistent G .
4	7, 2, 9	linear	Fair F , inconsistent G .
5	13^*	parabolic	Occasional consistency in F , G variable

Dimension	Inputs	Drift	Results
5	7, 8	parabolic	Spotty consistency
5	10, 13*	parobolic	Fair F , variable G
5	12*, 13*	parabolic	F , G both inconsistent
6	8	linear	Inconsistent
6	7, 8	parabolic	Inconsistent
6	10, 13*	parabolic	Scattered
6	10, 8, 9	parabolic	Inconsistent
9	8	constant	Possible consistency
10	8	constant	Fairly inconsistent

In the above tables, * indicates time series <u>delayed</u> by six quarters. Also, in the above (as well as in the computer programs) F refers to the model coefficient matrix, G to the input matrix, and H to the output matrix.

One surprise in the above is that the $b(t)$ estimation seems better behaved than that of $a_1(t)$. This was not anticipated by the authors.

The lack of consistent results for $a_1(t)$ is somewhat discouraging, as one intuitively expects more economic effect on the $a_i(t)$ than on $b(t)$.

It may be the case that use of a larger or more rapidly varying exogenous drift would produce an acceptable fit; on the other hand, it may be that the level of randomness in economic time series makes estimation on a deterministic basis an impossibility. If this is the case, attempts to estimate the interaction effects must await the accumulation of considerably more data than is currently available.

V. Conclusions

As mentioned in our Interim Report [3], we regard the process of model construction as one subject to a great deal of experimentation. Our intention in this project has been to develop some techniques and models useful in such experimentation. Further, we regard it as essential that the models constructed and employed be useable with available data in order that model parameters may be estimated and model accuracy verified. As should be evident from the above discussion, we have found that the utilization of the available data required considerable effort. The required data base is not easy to obtain, and the available records seem to cover a far shorter time interval than is desirable. This problem is particularly acute with regard to the income distribution data. Here the increased numerical difficulties in the estimation procedure demand detail (and accuracy) in the distribution data, and the subsequent interaction modelling will no doubt require data from a larger span of time than we have been able to obtain.

The results of our work in modelling the behaviour of the fertility curve seem fairly satisfactory. The algorithm developed for estimating the partial differential equation coefficients has produced consistent results, and simulation using the estimated coefficients reproduces the observed data to an acceptable degree of accuracy. The data required for the estimation procedures is available in sufficient detail; the relatively short time span for which the detailed data is available, however creates severe difficulties in the area of modelling the dynamical behaviour of the coefficients.

From the point of view of subsequent use, the estimation of the economic mobility function $\mu(x,s,t)$ and determination of its dynamical relation to other economic factors is probably of more interest. Our experimentation has shown that, while the fundamental methods of the estimation problem for $\mu(x,s,t)$ are similar to the methods used for estimating the "shift dynamics" of the fertility curve, the additional dimension involved causes the numerical difficulties to be somewhat more delicate. This is more the case in estimation of the age-income dependent components of the economic mobility; the algorithm appears quite robust as far as estimation of the time dependent components is concerned.

Initial experience with use of the algorithm in conjunction with real data led us to suspect that the level of aggregation of the available data might be too great to allow reliable mobility estimates to be made.

Subsequent experiments with varying aggregation levels were made utilizing simulated data. The results show a severe degradation in the reliability of the computed results at the level of aggregation present in the real data; the results also suggest that reliable estimates may be recovered from data aggregated to approximately half the extent of the data we were able to obtain.

In our work in estimating the interaction effects between economic factors and the estimated model coefficients, great difficulties have been caused by the scarcity of the available data. In fact, the income distribution data we have obtained covers such a short period of time and has been aggregated to such an extent that we have not attempted to estimate a dynamical model of the economic mobility.

The problems which arise in this area are formidable. Concisely stated, the process being modelled contains an unknown exogenous component, the available data is corrupted by noise of an unknown statistical nature, and the length of the available data record is short. Our approach has been to develop an algorithm capable of fitting models on the basis of short data and with a variety of possible exogenous components. This is based on a deterministic approach; the potential user of the package must be familiar with the effect of random data errors on the computed results in order to make effective use of the routine. We hope that our discussion of the ideas behind this approach and the details of the model and algorithms involved will be of some assistance in this regard.

Appendix A: Model Invariance and Inflating Income Scales

As discussed above, the fact that reported income data, subject to inflationary effects, appears to be the only available annual income distribution data leads one to consider the problem of estimating economic mobility on the basis of this data. The original derivation of the model equation for evolution of the distribution is easier to interpret on the basis of a time-invariant income scale, although there is no explicit assumption to this effect in the model derivation. One is led, therefore, to consider the problem of determining whether or not the available data ought to be deflated in some fashion before an estimation attempt is made. This is not an attractive prospect, for several reasons. One is that the effects of inflation on reported income varies across income levels (in a manner difficult to estimate). A second reason is that inflation affects wage and salary levels in various sectors of the economy with differing amounts of delay (again a circumstance difficult to quantify). These problems make it preferable, if possible, to formulate the model in terms of some "current" income measure. It turns out, as shown below, that the form of the governing equations is retained even if the distribution is referred to a "current" income measure (such as "reported income"). This may be shown as follows.

Recall that the model equation derived in [3] in terms of age x, income s and time t is given by

$$\frac{\partial p}{\partial t} = -\frac{\partial p}{\partial x} - \frac{\partial}{\partial s}(\mu(x,s,t)p) + i(x,s,t) - \gamma(x,s,t)p \quad .$$

where p is population density, i is (net) immigration, and $\gamma(x,s,t)$ the death rate. p has the interpretation of the number of people per unit age, per unit income interval, at time t.

We now consider the effect of introducing a time-variable change of income scale, of the general form

$$\sigma = \Sigma(s,t) \quad .$$

σ may be interpreted as current income (or current disposable, or reported, etc. income), while s represents a fixed income scale.

This change includes as a special case, the change which might be described as "uniform inflation", i.e.

$$\sigma = c(t)s$$

as well as various models incorporating the non-uniformities mentioned above.

It is convenient to assume that Σ has continuous second partial derivatives, and that it is monotone in s for each fixed t, so that

$$\sigma = \Sigma(s,t)$$

may be smoothly inverted to give

$$s = S(\sigma,t)$$

(i.e. s in terms of σ) so that we have the identity

$$\sigma = \Sigma(S(\sigma,t),t) \quad .$$

From the above identity follow the useful formulae

$$0 = \frac{\partial\Sigma}{\partial s} \cdot \frac{\partial S}{\partial t} + \frac{\partial\Sigma}{\partial t}$$

(from differentiating the above partially with respect to t, holding σ constant) and

$$1 = \frac{\partial\Sigma}{\partial s} \cdot \frac{\partial S}{\partial \sigma} \quad .$$

(from differentiating partially with respect to σ).

We now consider the change of independent variables x , s , t to the variables ξ , σ , τ according to $T : (x,s,t) \to (\xi,\sigma,\tau)$, where

$$\tau = t$$

$$\xi = x$$

$$\sigma = \Sigma(s,t) \quad .$$

(τ,ξ represent again time and age). This induces a corresponding change of any function f of (x,s,t) according to

$$\bar{f}(\xi,\sigma,\tau) = f(x(\xi,\sigma,\tau) , s(\xi,\sigma,\tau) , t(\xi,\sigma,\tau))$$

$$= f(\xi,S(\sigma,\tau),\tau) \quad .$$

The "chain rule" for partial differentiation gives

$$\frac{\partial f}{\partial t} = \frac{\partial \bar{f}}{\partial \sigma} \cdot \frac{\partial \Sigma}{\partial t} + \frac{\partial \bar{f}}{\partial \tau} \cdot 1 \cdot + 0 \quad .$$

$$\frac{\partial f}{\partial s} = \frac{\partial \bar{f}}{\partial \sigma} \cdot \frac{\partial \Sigma}{\partial s} + 0 + 0$$

$$\frac{\partial f}{\partial x} = \frac{\partial \bar{f}}{\partial \xi} \cdot 1 + 0 + 0 \quad .$$

Writing out the model partial derivatives in terms of the new variables (ξ,σ,τ) gives

$$\frac{\partial\bar{p}}{\partial\tau} + \frac{\partial\bar{p}}{\partial\sigma}\left[\frac{\partial\Sigma}{\partial t}\right] = -\frac{\partial\bar{p}}{\partial\xi} - \frac{\partial}{\partial\sigma} - \frac{\partial}{\partial\sigma}(\overline{\mu p})\left[\frac{\partial\Sigma}{\partial s}\right]$$

$$-\bar{\gamma}\,\bar{p} + \bar{i} \quad .$$

The appropriate formula for the age-income density in terms of the new variables (ξ,σ) is determined by noting that the expression

$$\iint_B p(x,s,t)\; dx\; ds$$

represents the population with ages and incomes in the arbitrary region B at time t . By the Jacobian rule for change of variable in a double integral, this is the same as

$$\iint_{TB} \bar{p}(\xi,\sigma,t)\;\left|\frac{\partial(x,s)}{\partial(\xi,\sigma)}\right|\; d\xi\; d\sigma$$

which, evaluating the Jacobian, is

$$\iint_{TB} \bar{p}(\xi,\sigma,t)\;\left|\frac{\partial S}{\partial\sigma}\right|\; d\xi\; d\sigma \quad .$$

Using the identity above, this becomes

$$\iint_{TB} \frac{\bar{p}(\xi,\sigma,t)}{\left|\frac{\partial\Sigma}{\partial s}\right|} \, d\xi \; d\sigma \quad .$$

Since TB is an arbitrary region of the new variable space, this identifies the appropriate density as

$$\bar{p}(\xi,\sigma,t) = \frac{\bar{p}(\xi,\sigma,t)}{\frac{\partial\Sigma}{\partial s}(S(\sigma,t),t)} \quad .$$

(The absolute value is dropped by the monotonicity assumption). Multiplying the governing equation by

$$\frac{\partial\Sigma}{\partial s}(S(\sigma,\tau),\tau))^{-1}$$

and re-arranging gives

$$\frac{\partial}{\partial\tau} \quad \frac{1}{\frac{\partial\Sigma}{\partial s}} \, \bar{p} \quad - \bar{p} \, \frac{\partial}{\partial\tau}\left[\frac{\partial S}{\partial\sigma}\right] \; +$$

$$\frac{\partial}{\partial\sigma}\left[\frac{\partial\Sigma}{\partial t} \cdot \frac{\bar{p}}{\frac{\partial\Sigma}{\partial s}}\right] - \bar{p} \, \frac{\partial}{\partial\sigma}\left[\frac{\frac{\partial\Sigma}{\partial t}}{\frac{\partial\Sigma}{\partial s}}\right] =$$

$$- \frac{\partial}{\partial\xi}\left[\frac{1}{\frac{\partial\Sigma}{\partial s}} \, \bar{p}\right] \; - \; \frac{\partial}{\partial\sigma}\left[\bar{\mu} \, \frac{\partial\Sigma}{\partial s} \cdot \; \frac{\bar{p}}{\frac{\partial\Sigma}{\partial s}}\right]$$

$$- \frac{\bar{r}}{\frac{\partial \Sigma}{\partial s}} \bar{p} + \frac{\bar{i}}{\frac{\partial \Sigma}{\partial s}}$$

Using the fact that

$$\frac{\frac{\partial \Sigma}{\partial t}}{\frac{\partial \Sigma}{\partial s}} = - \left(\frac{\partial S}{\partial t}\right)_\sigma = - \left(\frac{\partial S}{\partial \tau}\right)_\sigma$$

we see that the second and fourth terms cancel in the above, and defining

$$\tilde{p}(\xi,\sigma,\tau) = \frac{\bar{p}}{\frac{\partial \Sigma}{\partial s}}$$

as the "new" density, we see that

$$\frac{\partial \tilde{p}}{\partial \tau} = - \frac{\partial \tilde{p}}{\partial \xi} - \frac{\partial}{\partial \sigma} (\tilde{\mu}(\xi,\sigma,\tau)\tilde{p}) - \tilde{r}\, \tilde{p} + \tilde{\imath}$$

with

$$\tilde{\mu}(\xi,\sigma,\tau) = \bar{\mu}(\xi,\sigma,\tau) \cdot \frac{\partial \Sigma}{\partial s} + \frac{\partial \Sigma}{\partial t} \quad .$$

The above defines the appropriate mobility in terms of the "current" income scale, and shows the invariance of the model equations under such a change. As a practical matter, this removes the necessity of deflating the available data (available on a "current" basis) and prevents the addition of additional errors from such processing.

Appendix B: Mathematical Description - Linear Model Estimation Algorithm

As mentioned above, an algorithm has been constructed for the estimation of linear model coefficients on the basis of input-output data. There exists a considerable amount of literature on this topic, covering both stochastic and deterministic approaches. The work reported below is based most closely on the references [16], [17] and [19], but differs from this work in a way which has required some modification and extension of the results in those references.

The main impediment to the use of [16], [17] and [19] (as well as most other estimation algorithms known to the authors) in the present case is that no provision is made to handle systematically the presence of any exogenous components in the output data record. These references are concerned exclusively with the estimation of coefficients in minimal realizations. As noted above, this restriction in effect requires that all outputs be affected by the inputs.

The mathematical content of the following discussion consists essentially of linear algebra and difference equations. The framework is essentially that of discrete-time linear control theory.

Since the integration schemes used for the model core essentially reduce to difference equations, and it is discrete time data (typically yearly or quarterly values) which is available, it is natural to formulate the interaction section of the model in terms of difference equations.

The basic linear model may be described by the system of equations

$$z^{(1)}_{(k+1)} = A\, z^{(1)}_{(k)} + \sum_{i=1}^{m} b_i\, u_i(k)$$

$$z^{(2)}_{(k+1)} = D\, z^{(2)}(k)$$

$$y_j(k) = \langle c_j^{(1)}, z^{(1)}(k)\rangle + \langle c_j^{(2)}, z^{(2)}(k)\rangle \quad .$$

$$j = 1 \;, \;\ldots$$

In the above, $z^{(1)}$ and $z^{(2)}$ are elements of finite dimensional vector spaces over R , A and D are linear mappings, b_i vectors, and c_j elements of an appropriate dual space. By choosing bases in the vector spaces, the system may be represented in terms of components with respect to the bases in the form

$$x_{k+1} = \begin{bmatrix} A_1 & 0 \\ 0 & D \end{bmatrix} x_k + \begin{bmatrix} B_1 \\ 0 \end{bmatrix} u_k$$

$$y_k = \begin{bmatrix} C^{(1)} & C^{(2)} \end{bmatrix} \begin{bmatrix} x_k^{(1)} \\ x_k^{(2)} \end{bmatrix} = C\, x_k \quad .$$

The physical interpretation placed on the above model is as follows. The x is referred to as the state vector. The first "half" of the state vector, $x^{(1)}$, represents the part of the state driven by the input u, and so carries the effect of the input on the output. The second section of x, $x^{(2)}$, is entirely decoupled from the inputs; we visualize this as representing a "drift term" present in the observed output.

It is this component which accounts for the exogenous terms in the output. It is also the presence of this term which makes it necessary to modify the results of the references mentioned above, as the system model above is clearly not controllable.

In the use of this algorithm, the matrix D is essentially specified by the user. From the properties of linear difference equations, it follows that arbitrary linear combinations of polynomials, discrete exponential functions, discrete sinusoids (among others) may be included as possible exogenous (drift) terms in the model. As will be clear from what follows, it is not necessary to pre-specify the magnitudes of these terms; the estimation algorithm determines the magnitudes involved in the course of the computation. This property of the approach may make stochastic versions of the results obtained worth pursuit in connection with other applications, even though the available data limits their usefulness here.

Before our assumptions for the estimation model are listed, it may be worth emphasizing the fact that the state vector in the above representation is far from unique. In fact, a change of basis in the state space according to

$$x' = Px$$

leads to the system

$$x'_{k+1} = P \begin{bmatrix} A_1 & \\ & D \end{bmatrix} P^{-1} x'_k + P \begin{bmatrix} B_1 \\ 0 \end{bmatrix} u_k$$

$$y_k = C\,P^{-1} x'_k \quad .$$

This "new" system in fact has the same input-output behaviour as the original, and is indistinguisable on the basis of input-output records. This invariance is basic in the theory of linear systems, and is regularly exploited with the use of canonical forms to derive efficient estimation algorithms.

The standing assumptions which we make on the model are:

1. That the triple $[C^1, A_1, B_1]$ is a <u>minimal realization</u> of the transfer function $C^1(Iz-A_1)^{-1} B_1$;

2. That the pair $[C, \begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix}]$ is <u>observable</u>;

3. That the coefficient matrix $\begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix}$ is non-derogatory, i.e. that the characteristic and minimal polynomials are equal;

and 4. That the characteristic polynomial of the matrix D is brown.

Assumptions 1 and 2 above are made with no loss of generality in the ability to construct models of the sort considered. The assumption 3 (also referred to as the assumption that the system is cyclic) entails some loss of generality, but includes a wide enough class of systems to be useful in practice. Removal of assumption 3 forces the use of more complicated canonical forms, and requires knowledge of certain "structural indices" associated with the system being modelled. While in principle these indices may be determined from the available data, the tests involved require matrix rank calculations. As a result, they are for practical purposes unuseable in a situation of short, noisy data records.

A related issue is that of the determination of the dimension of the state vector in the above model. In principle, again, the system dimension may be determined (or estimated in a statistical manner) on the basis of rank tests on the available data. In practice, with imperfect measurements, the system order must in effect be guessed; at any rate, it is assumed that at least a condidate for the dimension of the model is available. Further discussion of this practical point is given in Section B of Chapter IV above.

For the sake of reference, it is assumed that the state dimension is N, that the output dimension is p, that the input dimension is m, and that dimension of the drift matrix is $q \times q$. We also adopt the notation defined by

$$A = \begin{bmatrix} A_1 & 0 \\ 0 & 0 \end{bmatrix} , \quad B = \begin{bmatrix} B_1 \\ 0 \end{bmatrix} .$$

For completeness, we outline the basic derivation of the parameter estimation equations. The basic techniques are quite similar to those of the references [16] [17] and [19], although certain complications arise through our use of the exogenous drift term in the model.

With the system equations written as

$$x_{k+1} = Ax_k + Bu_k \ , \ y_k = Cx_k$$

one may obtain by iterating the system equations the relation

$$\bar{y}_k = Mx_k + R^* \bar{u}_k$$

where

$$\bar{y}_k = \begin{bmatrix} y_k \\ \vdots \\ \\ y_{k+n-1} \end{bmatrix} , \quad \bar{u}_k = \begin{bmatrix} u_k \\ \vdots \\ \\ u_{k+n-1} \end{bmatrix} ,$$

$$M = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \quad \text{and} \quad R^* = \begin{bmatrix} 0 & & & \\ CB & & & \\ \vdots & & & \\ CA^{n-2}B & \cdots & CB & 0 \end{bmatrix}$$

Next, select a processing matrix P with the property that

$$T \overset{\Delta}{=} PM$$

is an invertible matrix. With the assumption that the system model is cyclic, and that the system is observable, it follows that defining the $n \times nm$ dimensional matrix P by

$$P = \begin{bmatrix} \alpha' & 0 & \cdots \\ 0 & \alpha' & \vdots \\ 0 & \vdots & \\ & & 0 \\ 0 & & \alpha' \end{bmatrix}$$

results in invertible matrix T for almost all choices of the m-dimensional vector α.

In effect, this replaces the given observed (vector) outputs by a single linear combination (with coefficients defined by α) of the outputs. The assertion about the invertibility of T is simply that the cyclic observable system is observable from almost all linear combinations of outputs. It may be noted that some choices of α may produce better conditioned matrices in the subsequent calculations than other choices, although there seems to be no way to determine an a-priori optimal choice of α.

With any choice of processing matrix P, as long as $T = PM$ is non-singular, the equations above may be manipulated to yield

$$P\bar{y}_{k+1} = Ap\ P\ \bar{y}_k + [P\tilde{R} - Ap\ P\ R^*]\ \bar{u}_k \quad ,$$

where $Ap = TAT^{-1}$. With obvious notation, the above may be written as

$$P\bar{y}_{k+1} = [A_p \quad R_p] \begin{bmatrix} P\bar{y}_k \\ \bar{u}_k \end{bmatrix} .$$

It also follows that with a partition of the matrix R_p in the form

$$R_p = [R_n \mid R_{n-1} \quad \cdots \mid R_1]$$

the matrix TB may be recovered by

$$TB = R_n + A_p R_{n-1} + \ldots + A_p^{n-1} R_1 .$$

The relation

$$P\bar{y}_{k+1} = [A_p \quad R_p] \begin{bmatrix} P\bar{y}_k \\ \bar{u}_k \end{bmatrix}$$

involves only the (unknown) system parameters, and the (observed) values of the inputs and outputs, and may be used to determine the system parameters.

One should note that solving the above equations for A_p, R_p, and solving the analogous equation to determine the output matrix C, (Assuming, as turns out to be true, that the rank of the resulting coefficient matrix allows this) in effect produces the realization triple $[CT^{-1}, TAT^{-1}, TB]$; that is, it produces one of the equivalent realizations of the input-output relation defined by the model.

In the interest of simplifying computation as much as possible, it if of course desirable that the above system of equations involve as few unknown parameters as possible. Since A_p is an $n \times n$ matrix, and R_p is of dimension $n \times m \cdot n$, as many as $n^2(m+1)$ parameters might be involved. In fact, with the choice of $P = S\alpha$, we have

$$T = \begin{bmatrix} \alpha' C \\ \alpha' CA \\ \\ \alpha' CA^{n-1} \end{bmatrix}$$

and a direct calculation shows that

$$A_p = \begin{bmatrix} 0 & 1 & & 0 \\ 0 & 0 & 1 & \\ & & & 0 \\ 0 & 0 & & 1 \\ p_1 & & & p_n \end{bmatrix}$$

where the characteristic polynomial of A is given by

$$p_A(\lambda) = \lambda^n - \sum_{i=1}^{n} p_i \lambda^{i-1} \quad .$$

A further direct calculation shows that

$$R_p = \begin{bmatrix} 0 \\ r' \end{bmatrix}$$

So that the estimation equations reduce to the system

$$y_\alpha(k+n) = [p' \quad r'] \begin{bmatrix} \bar{y}_\alpha(k) \\ \bar{u}(k) \end{bmatrix} \quad .$$

for the $n(m+1)$ unknown parameters in p' and r'.

Writing the above equation for (at least) $n(m+1)$ successive values of the time parameter k gives a system of equations which we symbolically write as

$$\begin{bmatrix} Data_y & Data_u \end{bmatrix} \begin{bmatrix} P \\ r \end{bmatrix} = \mathcal{Y}$$

Further manipulations (see [19] for example) lead to a similar system of equations (with the same coefficient matrix) for the entries in the output matrix CT .

To this point, the formal manipulations leading to the estimation equations are the same as employed in [16], [17] and [19]. These references rely on the assumption of controllability of the model to obtain the crucial result that the coefficient matrix

$$\begin{bmatrix} Data_y & Data_u \end{bmatrix}$$

has (almost surely) rank $n(m+1)$. In the present situation, the fact that almost all q-vectors are cyclic vectors for the matrix D , combined with the controllability assumption on the "non-exogenous" block of the model allows one to extend the argument in appendix A.2 of reference [16] to conclude that the coefficient matrix again (almost surely) has full rank.

While in the noise-free case it suffices to consider data lengths of only $n(m+1)$, in the actual implementation of the algorithm much more numerical stability results from using longer data strings and least squares solutions for the resulting system of equations. Essentially the situation is that although almost all data segments provide a full rank coefficient matrix, some data segments may (and do in practice) produce ill-conditioned coefficient arrays.

We note that the above algorithm in effect identifies the dynamics of the drift along with the forced component of the model. It neither utilizes the fact that the characteristics of the exogenous term are assumed known a-priori, nor does the method enforce (or check) the supposition that the exogenous terms are "decoupled" from the inputs. In the case of (artificially generated) noise free data, these characteristics automatically appear in the identified model. In the case of sparse or noisy data, however, there is the possibility that this additional information relating to the structure of the model may be utilized.

One modification which has been made to the above algorithm is to introduce the fact that the characteristic polynomial corresponding to the exogenous terms is assumed to be supplied by the user of the algorithm. If this is the case, then not all of the coefficients of the minimal

(=characteristic) polynomial of the full matrix A are independent. In fact if $P_A(\lambda)$, $p_a(\lambda)$ and $p_d(\lambda)$ denote the characteristic polynimials of the matrices A , A_1 , and D respectively, we have that

$$p_A(\lambda) = p_a(\lambda)\, p_d(\lambda) \quad .$$

Writing out the above in terms of the polynomial coefficients gives a relation of the form

$$p = b + L\, a$$

where p is a vector constructed from the coefficients of the characteristic polynomial of A , a the same for A_1 , and the vector b and matrix L are determined from the characteristic polynomial of D . Substituion of the above into the estimation equation gives

$$[\,\mathrm{Data}_y \cdot L \qquad \mathrm{Data}_u\,] \begin{bmatrix} a \\ \\ r \end{bmatrix} = [\,\mathcal{Y} - \mathrm{Data}_y \cdot b\,] \quad .$$

The above reduces the number of unknown parameters to $(n-q) + m \cdot n$, and so reduces the computational burden of the algorithm. Using the observation that postmultiplication of $Data_y$ by L effectively removes the exogenous component from the output data, one may check that (almost surely) the modified coefficient matrix has full rank. The modified estimation equations may then be solved for the coefficients of $p_a(\lambda)$, the characteristic polynomial $p_A(\lambda)$ may be computed according to the above, and the calculation of the input matrix TB proceeds as before.

While the above modification of the basic algorithm should serve to improve the numerical conditioning of the computations, there is still no guarantee that the estimates satisfy the decoupling condition implicit in the original system model. For that matter, we have yet to outline a procedure by which the decoupling may be verified after the computations have been made, even if the constraint is not incorporated in the original algorithm.

It appears necessary to have a computation of the characteristic polynomial $p_a(\lambda)$ corresponding to the driven components of the state in order to perform the decoupling computations.

In order to verify the decoupling conditions, one must obtain, in effect, a basis for the state space with respect to which the coefficient matrix will appear in the required block diagonal form. It is useful to keep in mind that it must be possible to carry out the computations required entirely on the basis of the available input-output record, so that the required basis must in effect be implicitly rather than explicitly constructed. Finally, recall that it has been shown above that the characteristic polynomials $p_a(\lambda)$ and $p_d(\lambda)$ are computable in terms of the available data under the assumptions of the model.

On the assumption that $p_a(\lambda)$ and $p_d(\lambda)$ are known (i.e. have been determined in a previous preliminary calculation) a suitable basis may be determined as follows.

From the assumption that the coefficient matrix is non-derogatory, it follows that the monic polynomials $p_a(\lambda)$ and $p_d(\lambda)$ are relatively prime, and hence that there exist polynomials $r(\lambda)$ and $s(\lambda)$ such that

$$r(\lambda)\, p_d(\lambda) + s(\lambda)\, p_a(\lambda) = 1 \quad .$$

The required polynomials are computable from the Euclidean algorithm, or by direct solution of the resulting system of linear equations.

A standard argument in the theory of matrix canonical forms shows that the matrices

$$r(A)\, p_d(A)$$

and $$s(A)\ p_a(A)$$

represent projection operators onto the complementary subspaces corresponding to the forced (i.e. A_1 block) and drift (i.e. D block) portions of the state space. In order to construct the required form, we construct a basis in the following way.

Define the vector c by

$$c = (\alpha' C)'$$

The basis used is implicitly dependent on the choice of the "selector coefficient" vector α. We construct (candidates for) basis vectors in the <u>dual</u> of the state space by

$$u_1 = \gamma(A')\ p_d(A')c$$

$$u_{n-q} = \gamma(A')\ p_d(A')\ A'^{\,n-q-1}\ c$$

$$u_{n-q+1} = s(A')\ p_a(A')c$$

$$u_n = s(A')\ p_a(A')\ A'^{\,q-1}\ c \quad .$$

That the above actually are linearly independent follows from

1. the observation that $\{u_1 \ldots u_{n-q}\}$ and $\{u_{n-q+1} \ldots u_n\}$ lie in mutually disjoint subspaces,

and

2. the fact that A is assumed non-derogatory implies that(for almost all choices of α) that $\gamma A')\ p_d(A')c$ and $s(A')\ p_d(A')c$ are both cyclic vectors for A'.

With the above definition of the basis vectors, an explicit calculation of the matrix representing A with respect to the dual of the constructed basis shows that the matrix is

$$\begin{bmatrix} 0 & 1 & & & & & & \\ & & 1 \ddots & 1 & & & 0 & \\ a_1 & \cdots & & a_{n-q} & & & & \\ & & & & 0 & 1 & & \\ & 0 & & & 0 & 0 & 1 & \\ & & & & & & & 1 \\ & & & & d_1 & \cdots & & d_q \end{bmatrix},$$

which is in the desired black diagonal form.

In order to incorporate or check the decoupling constraint, it is necessary to determine the matrix representing the change of basis from the basis used in the original estimation algorithm to the block diagonal basis constructed above. This can be determined by recalling that the matrix T such that

$$TAT^{-1} = \begin{bmatrix} 0 & 1 & & \\ 0 & 0 & 1 & \\ & & & 1 \\ p_1 & & & p_n \end{bmatrix}$$

in the original algorithm is just

$$T = \begin{bmatrix} c \\ c'A \\ \vdots \\ c'A^{n-1} \end{bmatrix}$$

while the matrix π producing the block diagonal form

$$\pi A \pi^{-1} = \begin{bmatrix} 0 & 1 & & & & & & \\ 0 & 0 & 1 & & & 0 & & \\ a_1 & & & a_{n-q} & & & & \\ & & & & 0 & 1 & & \\ & 0 & & & 0 & 0 & 1 & \\ & & & & d_1 & & & d_q \end{bmatrix}$$

can be written as

$$\pi = \begin{bmatrix} c' \rho(A) \\ c' \rho(A)\, A \\ \vdots \\ c' \rho(A)\, A^{n-q-1} \\ c' \sigma(A) \\ c' \sigma(A) \cdot A \\ \vdots \\ c' \sigma(A)\, A^{q-1} \end{bmatrix},$$

where $\rho(A) = r(A)\, p_d(A)$, and $\sigma(A) = s(A)\, p_a(A)$.

By use of the Caley-Hamilton theorem, the matrix polynomials in π may be reduced to combinations of the first $n - 1$ powers of the matrix A . Defining the scalars $p_j^{(i)}$ and $\sigma_j^{(i)}$ by

$$\sum_{j=0}^{n-1} \rho_j^{(i)} A^j = \rho(A) \cdot A^{i-1} \pmod{P_A(A)}$$

$$\sum_{j=0}^{n-1} \sigma_j^{(i)} A^j = \sigma(A)\, A^{i-1} \pmod{P_A(A)}$$

we see that with Q defined by

$$Q = \begin{bmatrix} \rho_0^{(1)} & \cdots & \rho_{n-1}^{(1)} \\ \vdots & & \vdots \\ \rho_0^{(n-q)} & \cdots & \rho_{n-1}^{(n-q)} \\ \sigma_0^{(1)} & \cdots & \sigma_{n-1}^{(1)} \\ \vdots & & \vdots \\ \sigma_0^{(q)} & \cdots & \sigma_{n-1}^{(q)} \end{bmatrix}$$

we have

$$\pi = QT \quad .$$

Note that Q^{-1} exists, since Q effects a change of basis.

From the original model formulation, we see that the condition that the exogenous drift subspace be decoupled from the inputs translates simply into the condition that the last q rows of the matrix πB should vanish. Using the above formula for π, we get the condition

$$w_i' \, QTB \equiv 0 \quad , \qquad i = 1, \ldots q$$

with w_i' defined as a row vector with 1 in position $n - q + i$, zeroes elsewhere.

The above result may be used in two ways. The first is as a check that the exogenous subspace is decoupled after a computation has been performed. Since the estimation algorithm described above produces (in the noise free case) the matrices

$$TAT^{-1}$$

$$TB$$

$$CT^{-1} \quad ,$$

the results do not display the decoupling. To check this, one need only compute Q as described above, and form Q times the (estimated) matrix TB. For consistency with the model assumptions, this should return zeroes in the last q rows. In the practical case where inaccurate data has been processed, and in which one does not expect an exact model fit in any case, the last q rows of the result should be "relatively small".

A second use for this result is to incorporate the decoupling constraint in a "two stage" estimation algorithm. As seen from the above, in the noise-free case Q may be calculated from the available data. In the case of inaccurate data, one may still calculate an estimated value for Q , or a range of values consistent with the available data.

The condition

$$w_i' \, QTB = 0$$

may be combined with the formula for TB , i.e.

$$TB = R_n + A_p R_{n-1} + \dots + A_p^{n-1} R_1$$

where, recall, in general

$$R_p = [R_n \quad R_{n-1} \quad \dots \quad R_1]$$

and for the specific choice $P = S\alpha$

$$R_p = \begin{bmatrix} 0 \\ \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} r_n' & r_{n-1}' & \dots & r_1' \end{bmatrix}$$

Hence, we have

$$w_i' Q[R_n + A_p R_{n-1} - \dots + A_p^{n-1} R_1] = 0$$

$$w_i'[QR_n+QA_pQ^{-1}QR_{n-1}+\ldots+QA_p^{n-1}\ Q^{-1}QR_1] = 0$$

which reduces to the condition

$$\sum_{j=0}^{n-1} \begin{bmatrix} 0 & 1 & & \\ 0 & 0 & 1 & \\ & & & 1 \\ d_1 & & & d_q \end{bmatrix}^j \tilde{q}\ r'_{n-j} = 0$$

where $\tilde{q}$ is the column vector consisting of the last q entries of

$$Q\begin{bmatrix} 0 \\ \\ 0 \\ \\ 1 \end{bmatrix} .$$

The above system of equations ($q\cdot m$ equations in all) represents the decoupling constraint referred to the coordinate system of the original estimation algorithm, and further, expressed as a linear constraint on the intermediate parameter vector r involved in the algorithm. If we represent the above as a system of linear equations of the form

$$v_i'\ r = 0 \quad , \quad i = 1,\ldots m\cdot q\ .$$

Then the original estimation equations

$$\left[\mathrm{Data}_y \cdot L \quad \mathrm{Data}_u \right] \begin{bmatrix} a \\ \\ r \end{bmatrix} = \left[y - \mathrm{Data}_y \, b \right]$$

may be augmented to give the estimation equations <u>incorporating</u> the decoupling constraint

$$\begin{bmatrix} \mathrm{Data}_y \cdot L & \mathrm{Data}_u \\ & v_1' \\ 0' & \\ & v_{qm}' \end{bmatrix} \begin{bmatrix} a \\ \\ r \end{bmatrix} = \begin{bmatrix} y - \mathrm{Data}_y \cdot b \\ 0 \\ 0 \end{bmatrix} .$$

There seems to be some numerical evidence to suggest that in the original estimation algorithm, the computation of the characteristic polynomial is more stable with regard to data errors than is the calculation of the input matrix TB . This is probably the case because calculation of TB essentially involves <u>powers</u> of the estimated TAT^{-1} matrix, compounding errors in A with those in the parameter vector r . If this is indeed the case, then there is reason to believe that the two stage algorithm suggested, consisting of a calculation of Q, followed by a re-calculation of TAT^{-1} and TB based on the augmented system above, will prove numerically more stable.

The discussion above omits most of the detailed calculations involved in the actual use of the algorithms proposed. It is intended, however, to provide the reader with enough information to understand the basic nature of the method. Such understanding is probably essential in order to gain a reasonable ability to make effective use of the algorithm with the inevitable inaccuracies of real data.

The three basic algorithms outlined above (the "straight estimation", the estimation incorporating a user-supplied exogenous drift polynomial, and the "two stage" algorithm incorporating the decoupling constraint) have been implemented in the APL programming language, and are listed for reference in the appendix as ESTIMATE, ESTIMATEA, and ESTIMATEB. ESTIMATEA has been translated into FORTRAN (also listed in the Appendix), and sample experiments using ESTIMATEA have been run using real data. (See below).

APPENDIX C: Remarks on Long-Term Sequential Estimation

As mentioned above, the structure of the estimation algorithm for the determination of both the fertility parameters and the economic mobility in such that orthogonal components are returned by the computer output. That is, if we represent the data matrix M in the form

$$M = \sum_i d_i \, a_i'$$

(which includes both estimation algorithms), then the result of the computations is to essentially select a basis with respect to which both $\{a_i\}$ and $\{d_i\}$ are orthogonal sets.

When the process of sequential estimation described in Chapter II is carried out, processing the data beginning at year τ essentially involves the orthogonalization of the functions $\{a_i(\tau+t)\}$ on the interval $0 \leq t \leq L$. Here L is the estimation data length. As a result, the computed basis vectors vary with the starting time τ.

Strictly speaking, the results of the computations for various starting times τ ought to be referred to a common (stationary) basis for comparison. As a practical matter, however, with smooth time variation in the $\{a_i(t)\}$ this effect of the "skewing" of the coordinate system is slow, and largely swamped out by computational and model errors. It is only in the case where a long (relative to the estimation length L) run of data is available that this effect becomes significant. In such a case the programs should be modified accordingly.

APPENDIX D: Interim Report

The final report above has been written on the assumption that readers are familiar with the contents of the interim report [3]. The final report makes numerous references to and use of results and methodology presented in [3].

In order to make the final report essentially self contained, we include here a copy of the interim report.

An Interim Report to the Ministry of State for Science and Technology

on

THE QUEEN'S MATHEMATICS DEPARTMENT POPULATION MODEL

by

J. Davis and J. Verner

Queen's University, Kingston, Ontario

Queen's Mathematical Preprints No. 1974-8

April 15, 1974

Contents

Appendices

Abstract

The Queen's Mathematics Department population model is a dynamic model for simulating the evolution of a population distribution as a function of age and income level.

The basic structure of the model is such that birth-rates may be generated within the model as endogenous variables. This allows the inclusion of feedback effects from the population distribution to birth and immigration rates, and so provides a capability for simulations valid over longer time intervals than are possible with exogenous birth and immigration rate variables.

The model has been constructed with two main issues in view. The first is that of compatability of this model with other models with which it might be combined. This requirement dictates a modular structure described in this report. The second issue is the problem of parameter estimation in the model. The model has been formulated in such a way that estimation is made possible.

Effective numerical algorithms for these estimations based on available data formats are also reported.

A description of work that remains to be done in order to complete development of the model is also included.

I. Introduction

Models of population growth form an essential part of any attempt at large scale socio-economic modelling. The age and economic level structure of the population has a direct bearing on various government service requirements ranging from elementary schools to pension plans, as well as the economic base available to support such programs. For this reason, it is essential that population models capable of simulating behaviour over a reasonable length of time be investigated.

Traditional demographic methods project population estimates forward in time by means of an aggregation procedure, followed by a linear extrapolation procedure based essentially on a Markov-chain type of model. Such methods are reasonably accurate over the relatively short term; however, the model structure is such that the fertility curve (the age-specific distribution of the birth-rates) is treated as an "exogenous variable" which must be specified for each run. Some attempts (the so-called "cohort method") have been made to include in the model the observed fact that birth-rates do vary over time, but the problem of extrapolation birth-rates forward in time in order to increase the length of time that model results are valid remains.

It is clear that <u>many</u> factors affect birth-rates: economic conditions, perceptions of future economic conditions, ecological concerns, a host of other factors affect birth-rates to a greater or lesser effect. It is also clear that present population structure affects in turn the economic climate, and the general environment. The present population is in turn the result of past birth-rates (and immigration).

The conclusion of the above observations is that it is impossible to decouple the dynamics of the birth-rates from those of the population structure without compromising the long term validity of the model simulation. In effect, there exists a feedback path from population structure to birth-rates which may not be ignored over the long term. (This does not imply that such decoupling, based on assumptions that certain factors "vary slowly with time", detracts from the usefulness of models intended for use over relatively short time periods).

The model discussed in this report represents an approach to the problem of including the dynamic feedback effect mentioned above in a simulation model. More specifically, this report contains the results of some work on what we regard as the <u>basic</u> structural elements and problems associated with models of this sort.

The structure and "philosophy" of the model is discussed more fully in Section II.

It was determined early in our investigation that partial differential equations were an appropriate component of the model - in fact, it is hard to consider the effect of the "baby boom" without coming to the conclusion that a wave equation occupies a central position in a model of population distribution. In work on any dynamical model it is necessary to determine numerical values for parameters occurring in the model equations before any simulation may be carried out. At worst, these parameters may have to be guessed; obviously it is much more desirable that the parameters be estimated from historical records of the phenomenon being modelled, if possible. The latter procedure provides an indirect means of assessing the validity of the model.

In the case of models governed by partial differential equations this estimation problem is even more severe, as it is often necessary to estimate not just a finite set of parameters, but a function of one or more independent variables. Aspects of this problem are reported in Section III and Appendix C.

Once parameters and functions have been estimated from the available data, it is possible to simulate the system on a digital computer. This, of course, involves the solution of coupled systems of ordinary and partial differential equations by numerical methods. It is necessary to investigate the effects of the numerical methods used on the accuracy of the results obtained, in order to ensure that the behaviour of the model is a result of the actual "dynamics" of the model itself, and not the result of instability caused by inaccurate numerical methods. The difficulty of this problem is again increased by the fact that partial differential equations are involved. The work in this area has been checked by use of certain exact solutions to the governing equations (Appendix A) and is described in Section V and Appendix B.

II. Structure of the Model

It is helpful in describing the structure of the model presented here to explain briefly the general philosophy of "modelling" that the authors of this report hold, and which has had a strong effect on the structure adopted for the model discussed here.

In the first place, we feel that a main product of any modelling and simulation effort should be insight into the behaviour of the phenomena being modelled. Perhaps the worst fate that can befall any model is that it be used to generate one set of trajectories which are then canonized as "the predictions" of the model (or worse yet, of the computer used to generate the output). Rather, the use of a model should itself be a dynamic process. It is certain that better data regarding the variables involved in a model will become available in the future, and it is only prudent that this data, if possible, be used to improve any "forecasts" made using the model.

It is also rather likely that there are alternative opinions regarding the actual structure of some sections of any given model. In this situation, it is essential that simulations be run incorporating these alternative opinions, rather than selecting one arbitrarily and incorporating it permanently into the model. It is only by simulating each of the reasonable alternatives (a matter of judgement is involved here) that any true insight into the behaviour of the system as a whole can be gained; this includes an appreciation of the range of results which might be expected under reasonable alternative models.

These considerations suggest at least that a useful model must have sufficient flexibility of structure to accommodate changes of the sort mentioned above. In order to build flexibility of this sort into a model, it is necessary to identify a basic dynamical core around which the model may be constructed.

The basic core of the model in this paper consists of equations for the evolution of the population distribution, and for the evolution of the fertility curve over time.

As was mentioned in the previous section, it is clear that economic conditions interact with the current population distribution and other factors to produce the current instantaneous birth-rate. It is also clear that the exact nature of these interactions is complicated and probably poorly understood in total, although some progress in this direction may be made by various methods. On the other hand, the evolution of the population distribution may be described (see the following section) by a partial differential equation of the conservation law type. Also, by looking at birth-rate records, it is possible to argue that the evolution of the fertility curve is also adequately modelled by a relatively simple partial differential equation. Furhter, the structure of these sections of the model is independent of the details of economic and other interactions which combine to affect birth-rates.

These considerations have led us to the decision to base the framework of the model on the dynamics of the population distribution and of the fertility curve. This leads to an overall model structure which may be represented in the "block diagram" form illustrated in Figure 1.

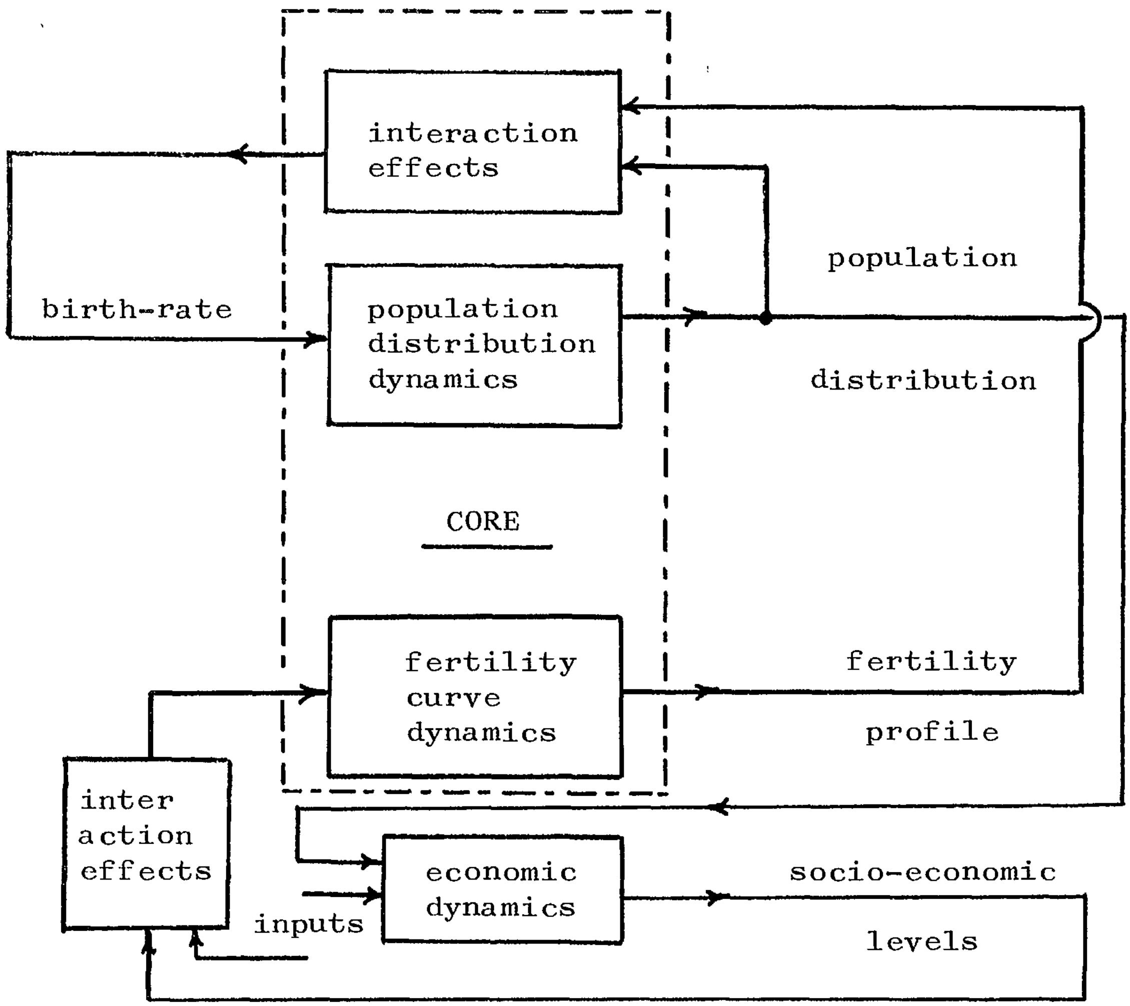


Figure 1.

This block diagram includes a section labelled economic dynamics. It is judged that the exact form of this section will be the subject of some debate, and that repeated simulations with varying socio-economic models will be required. It is expected that models with a relatively long time horizon and moderately high level of aggregation will be found most appropriate. In particular, models of the "Candide" type with high levels of detail and relatively short (eg. ten year) time spans are not felt to be appropriate. As work up to the time of the writing of this report has been concentrated on problems associated with core section of the model, problems in this particular area require further study.

III. Derivation of Equations of Dynamic Core

The core of the model consists of two partial differential equations: one for the evolution of the population density as a function of time, age, and income level, and one for the evolution of the fertility curve (i.e. the curve of age and income specific birth-rates). These two equations are coupled in a non-linear fashion, although the non-linearity appears only in the boundary conditions for the population equation. This fact is of considerable use in connection with the estimation problems discussed in the following section, and makes the derivations presented below simpler than might otherwise be the case.

A. Population Distribution Evolution

The model presented below is formulated as a basically deterministic model, and processes are modelled as occurring continuously in time on a macroscopic level, even though on a microscopic level the events may occur at discrete intervals of time. In this connection, a first step is to recognize that an averaging process is taking place whenever what are essentially discrete events are "smeared out" and modelled continuously in time. This process is illustrated by the use of death-rates in population models, decay rates in radioactive decay problems, and, in the derivation below, of an economic mobility μ . In these cases, the use of such rates essentially distinguishes between deterministic and stochastic modelling approaches.

The equation governing the population distribution may be derived readily from what are essentially counting or bookkeeping methods. This is most easily demonstrated by the derivation of a simple model of population as a function of age x , neglecting death-rates, immigration and any other variables. In this case, the appropriate counting argument is essentially that the number of people at age x at time t is the same as the number of people at age $x - \Delta t$ at a time Δt units earlier: In terms of population density p, this becomes

$$\int_{x-\frac{\Delta x}{2}}^{x+\frac{\Delta x}{2}} p(x,t)dx = \int_{x-\frac{\Delta x}{2}-\Delta t}^{x+\frac{\Delta x}{2}-\Delta t} p(x,t-\Delta t)dx ,$$

which for smooth densities p is essentially

$$p(x,t) = p(x-\Delta t, t-\Delta t)$$

or $$p(x,t) - p(x,t-\Delta t) = p(x-\Delta t,t-\Delta t) - p(x,t-\Delta t) .$$

Dividing the above by Δt and letting $\Delta t \to 0$ results in the partial differential equation

$$\frac{\partial p}{\partial t} = - \frac{\partial p}{\partial x} .$$

As is well known, the general form of solution to the above is $f = \varphi(t-x)$, with φ an arbitrary function which must be evaluated from the boundary conditions appropriate to the problem. The appropriate boundary condition is that

$$p(x=0,t) = \beta(t)$$

where $\beta(t)$ is the birth-rate of time t. That this is the appropriate boundary condition may be verified by noting that this gives the solution

$$p(x,t) = \beta(t-x) \quad ,$$

which says essentially that the number of people at age x at time t is the number of people born at time $t - x$, i.e. x years before time t. This of course is entirely evident from the assumptions made above.

The model considered in this paper includes a partial differential equation for the population density $p(x,s,t)$ as a function of the three variables age x, income s, and time t. As will be shown below in Appendix A, it is unnecessary to specify at this point the units involved in the income scale s, that is, whether s represents net income, disposable income, or some other measure. This is so because the form of the governing equation is invariant under a (non-linear) change of income scale, so that the units involved become an issue only during the processing of data for estimation purposes. This fact is a pleasant surprise which naturally arises out of the structure of the model equations.

To derive an equation for the population density on a realistic basis, it is necessary to account for effects neglected in the simplified model above, in particular to introduce terms

$$i(x,s,t)$$

representing the immigration rate (as a function of age, income level, and time), and the death-rate

$$r(x,s,t) \quad .$$

It is also necessary to introduce a term which accounts for the change of income level of various segments of the population over time. To accomplish this, we introduce an economic mobility function,

$$\mu(x,s,t) \quad .$$

Even though income levels of individuals on a microscopic scale undergo changes at discrete instants of time, perhaps modelled by a Poisson process, on the macroscopic scale of its influence on the income distribution we model the effect as one of a continuous flow across income levels. With this effect in mind, a term of the form

$$\mu(x,s,t) \cdot \Delta t$$

has an interpretation as the fraction of people at income level and age x crossing through level s in the time interval from t to $t + \Delta t$.

With the above definition of terms, it is easy to use a "counting argument" entirely similar to the one above to arrive at an equation representing the evolution of population density. The result is

$$\frac{\partial p}{\partial t} = -\frac{\partial p}{\partial x} - \frac{\partial}{\partial s}(\mu(x,s,t)p) - r(x,s,t)p + i(x,s,t)$$

Just as in the above derivation, it can be seen that the appropriate boundary condition for this equation is again

$$p(x=0) = \text{birth-rate}.$$

There is a technical problem associated with this boundary condition, since taken literally it demands the assignment of an income level to babies at birth. In fact, the model as formulated above is capable of propagating income level migration through childhood. It is clear that such a procedure makes little sense; however, the problem may be avoided rather easily by the following device. The income distribution at $x = 0$ may be set equal to that at the age of entry into the labour market. If the economic mobility is equal to zero for values of age x less than the labour market entry age, then the income distribution will remain constant for ages less than entry at the values of the entry distribution. Income distribution data generated by simulation runs may then be considered only for ages greater than an age of entry into the labour market, and no further modification of the model is required. The income distribution at entry age must be generated as part of the economic section of the model, and this effect comprises one of the feedback paths from the economic to population distribution sector shown on Figure 1.

Other effects of this sort, for example, an effect of p on the economic mobility μ , are implicitly contained in the time dependence of μ . As will be seen below in the section dealing with estimation problems associated with the model, there are substantial theoretical and practical benefits which follow from modelling the income migration process as above. In particular, it is then possible to devise numerical methods to estimate μ from observed population distribution data.

B. Fertility Curve Dynamics

Although the observation that socio-economic conditions, social attitudes, and so on, exert an effect on birth-rates is a common one, there seems to have been little effort made to quantify these effects in a dynamic model. Undoubtedly, a major reason for this is that it appears impossible to "derive" such a set of relations in the sense of the derivation outlined above for the population density dynamics.

For this reason, we have decided to approach this problem as one in system identification. That is, we attempt to formulate the problem in such a way that the problem is reduced to that of estimating a dynamical relationship between a relatively small number of variables. This in itself is a major reduction, since in principle a fertility curve is an element requiring an infinite number of numbers for its specification.

This first reduction may be obtained by examining typical historical records of the behaviour of fertility curves over time (See Figure 2). A first observation is that the curves are all of roughly the same shape. An examination of their differences shows that their peaks slide from age to age over time, and that the area under the curve, representing the total birth-rates to be expected from a uniformly distributed population, varies over time.

A simple partial differential equation capable of reproducing this observed behaviour has been adopted as the basis for the fertility curve dynamics. This is

$$\frac{\partial f}{\partial t} = - a(t) \frac{\partial}{\partial x} (d(x)f) - b(t)f \quad .$$

The first term in this equation produces the effect of the shifting peak, while integrating the equation with respect to x shows that b(t) is the percentage change of the area under the fertility curve per unit time. <u>An equivalent interpretation is that it represents the percentage rate of change of average family size.</u>

The justification of the representation of the fertility curve dynamics by the above equation may be carried on in several ways. In the first place, the interpretation given b(t) guarantees the presence of the term b(t)f in virtually any such equation. The appropriateness of the term representing the "shifts" may be supported on the basis of a time scale argument, combined with the fact that the model fits the observed data reasonably well. The "shifts" occur in the data on a time scale considerably faster than that of the dynamics of the population section of the model. In fact, the shifts appear correlated with variations in the economic climate, recessions, rising and falling unemployment, and the like. Since these effects are expected to be introduced into the model most likely on the basis of "standard" econometric and business cycle models, it is anticipated that it will be possible to include the function a(t) and its dynamics in this section of the model. The dynamics of a(t) are to be identified by means of either the usual econometric model identification techniques, or more recent work in the area of control theory. Since this identification problem presupposes knowledge of the term $d(x)$, work in this area is dependent on solving the problem of estimating $d(x)$, and applying the algorithm for this purpose is described in the following section.

Comments similar to the above also apply to the problem of determining the dynamics governing the term $b(t)$, although it is suspected that this will be even more difficult than the above process. This is so because $b(t)$ is dominated more by social attitudes, education, and other effects much less easy to quantify than economic ones. It is felt that this area represents an example of the need for alternative sub-models and repeated simulations discussed above in Section II in connection with the overall structure of the model.

While the above discussion has been carried through as though the fertility curve were independent of income, an entirely similar derivation is possible on the basis of an income dependent fertility curve. If one also allows the possibility that the economic interactions occur unevenly across income levels, then the appropriate equation is

$$\frac{\partial t}{\partial t} = -\frac{\partial}{\partial x}(a(x,s,t)f) - b(s,t)f \quad .$$

Because of the meaning of a fertility curve (or surface, if s is included as an independent variable) as an age (and income) specific birth-rate, the formula for the total birth-rate is simply

$$b(t) = \int f(x,s,t)\, p(x,s,t)dxds \quad .$$

No mention has been made in the above derivations of any geographical aspect of the problem. There are, however, some restrictions implicit in the derivations of the model equations. It is clear that certain of the quantitites involved in the above equations vary with geographical locality. From this, it is obvious that the model must be applied separately over geographical areas between which the relevant quantities vary. To obtain an overall model, then, internal migrations must then be included in the immigration rates of the models for each geographical region.

There is also implicit in the model derivation as assumption of a sufficiently large sample population, so that the modelling of the immigration, death, birth, and economic migration processes as continuous is valid.

IV. Model Estimation Methods

The dynamical equations governing the core section of the model derived above involve various auxiliary functions, namely: death and immigration rates, an economic mobility function $\mu(x,s,t)$, and functions $a(t)$, $b(t)$ and $d(x)$ determining the evolution of the fertility curve. Before it is possible to produce any simulation runs with the model, it is necessary to determine suitable estimates of these functions.

Generally speaking, this problem of parameter and function estimation is one of the most difficult ones involved in the construction of any model. Consideration of the conventional techniques of econometric modelling makes obvious the amount of effort which is expended in this area. In fact, with a certain amount of injustice one might view much econometric modelling as consisting of the development of schemes for the recursive estimation of parameters for short term (often linear) extrapolation models. This view ignores the effort involved in determining the extrapolation model whose parameters are to be estimated, but the fact remains that there continues to be much work on the development of regression - estimation methods in this area.

At practically the opposite end of this problem stand models of the sort proposed by Forrester and his associates. One of the most consistent criticisms levelled at Forrester's World and Urban Dynamics models is that practically no attempt has been made to estimate the parameters and functions involved in the models in any "realistic" fashion.

This apparent gulf between the Forrester models and conventional econometric models is, in our view, a large contributing factor to the hostile reaction Forrester's models have received in some quarters. It is also a gulf that is not easily overcome by philosophic discussions about differences of purpose between the two approaches.

In the case of the present model, it happens that considerable progress can be made in estimating the functions that are involved in the model of the core dynamics. Of course, this is not entirely unexpected, since an effort has been made to formulate the dynamics of the core in terms of variables which may be readily measured. Also, our definition of what constitutes the core dynamics of the model virtually assures that it must be possible to produce useful quantitative estimates of the functions involved.

The functions $r(s,s,t)$ and $i(x,s,t)$ in the population model are just death and immigration rates, so there is no problem in obtaining historical records of these. Similarly, the function $b(t)$ may be readily determined on the basis of its interpretation in terms of area under the fertility curve.

This leaves just the terms $\mu(x,s,t)$ in the population equation and $a(t)\,d(x)$ in the fertility equation to be determined. It can be seen that each term enters its equation in an analogous way, so that an estimation method can be derived which can be used to estimate <u>both</u> the economic mobility $\mu(x,s,t)$ and the term $a(t)\,d(x)$ in the fertility equation.

It is shown in Appendix A that the form of the partial differential equations is such that an integrating factor may be introduced to reduce the problem to that of estimating $\mu(x,s,t)$ and $a(t)\,d(x)$ in the equations

$$\frac{\partial \bar{p}}{\partial t} = -\frac{\partial \bar{p}}{\partial x} - \frac{\partial}{\partial s}\left(\mu(x,s,t)\bar{p}\right)$$

$$\frac{\partial \bar{f}}{\partial t} = -\frac{\partial}{\partial x}\left(a(t)d(x)\bar{f}\right) \quad .$$

In the modified population equation, integrate between the limits of s and infinity. There results

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right) \int_s^\infty \bar{p} \, ds = \mu(x,s,t) \cdot \bar{p}(x,s,t) \quad ;$$

here $\int_s^\infty \bar{p} \, ds$ has the interpretation of the number of people at age x , income s , and time t having an income greater than s . Solving this for μ gives

$$\mu(x,s,t) = \frac{1}{\bar{p}(x,s,t)} \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right) \int_s^\infty \bar{p}(x,s,t) ds \quad .$$

This provides an estimate of μ wherever $\bar{p}(x,s,t) \neq 0$. Since $\bar{p}(x,s,t) \neq 0$ except on the "tails of the distribution", the above formula may be used to determine μ throughout the age and income brackets containing the great bulk of the population. On the tails of distribution (e.g., at very high income levels) the interpretation of μ makes it clear that μ must tend to zero, so that the fact that the above formula is less useful there is of little concern.

Carrying out exactly the same procedure with the modified fertility curve results in

$$a(t)\ d(x) = \frac{1}{\bar{f}(x,t)} \frac{\partial}{\partial t} \int_x^\infty \bar{f}(x,t)dx \quad .$$

This determines $a(t)\ d(x)$ over those portions of the age scale which $\bar{f}(x,t) \neq 0$. Again $\bar{f}$ tends to zero only on the tails of the fertility distribution. Recalling that the term $a(t)\ d(x)$ was introduced to account for changes in the age distribution of fertility we see that intuitively $a(t)\ d(x)$ reflects the effects of shifts in "planned births" for the most part. Since births arising in the extremes of the fertility distribution do not fall into that category, it is clear that $d(x)$ must approach zero at these extremes. Hence it is again true that the fact that the formula derived is less useful in regions where $\bar{f}$ is close to zero is of small consequence.

Once $\mu(x,s,t)$ and $a(t)\ d(x)$ have been estimated, further estimation problems remain. One problem is that of determining numerically the values of $a(t)$ alone for use in identifying the interactions between the economic sector and the fertility curve. A second related problem is that of isolating the time dependence in $\mu(x,s,t)$ in such a way that a similar interaction analysis may be carried out. These problems are of a somewhat more technical nature, so our work on them is reported in Appendix C below.

A further technical complication arises in connection with practical use of the estimation formulas above. This arises from the fact that the actual population density data is not available; rather figures are available for, say, the number of persons between ages 25 and 29 with income between eight and ten thousand dollars. This amounts to the data

$$\int_{8,000}^{10,000} \int_{25}^{29} p(x,s,t)\, dx\, ds$$

at a fixed value of t .

We have expended a moderate amount of effort to develop accurate numerical algorithms with which $u(x,s,t)$ and $a(t)\, d(x)$ may be determined from aggregated data of the sort mentioned above. The method devised uses somewhat delicate application of numerical spline techniques. This work is also described in Appendix C.

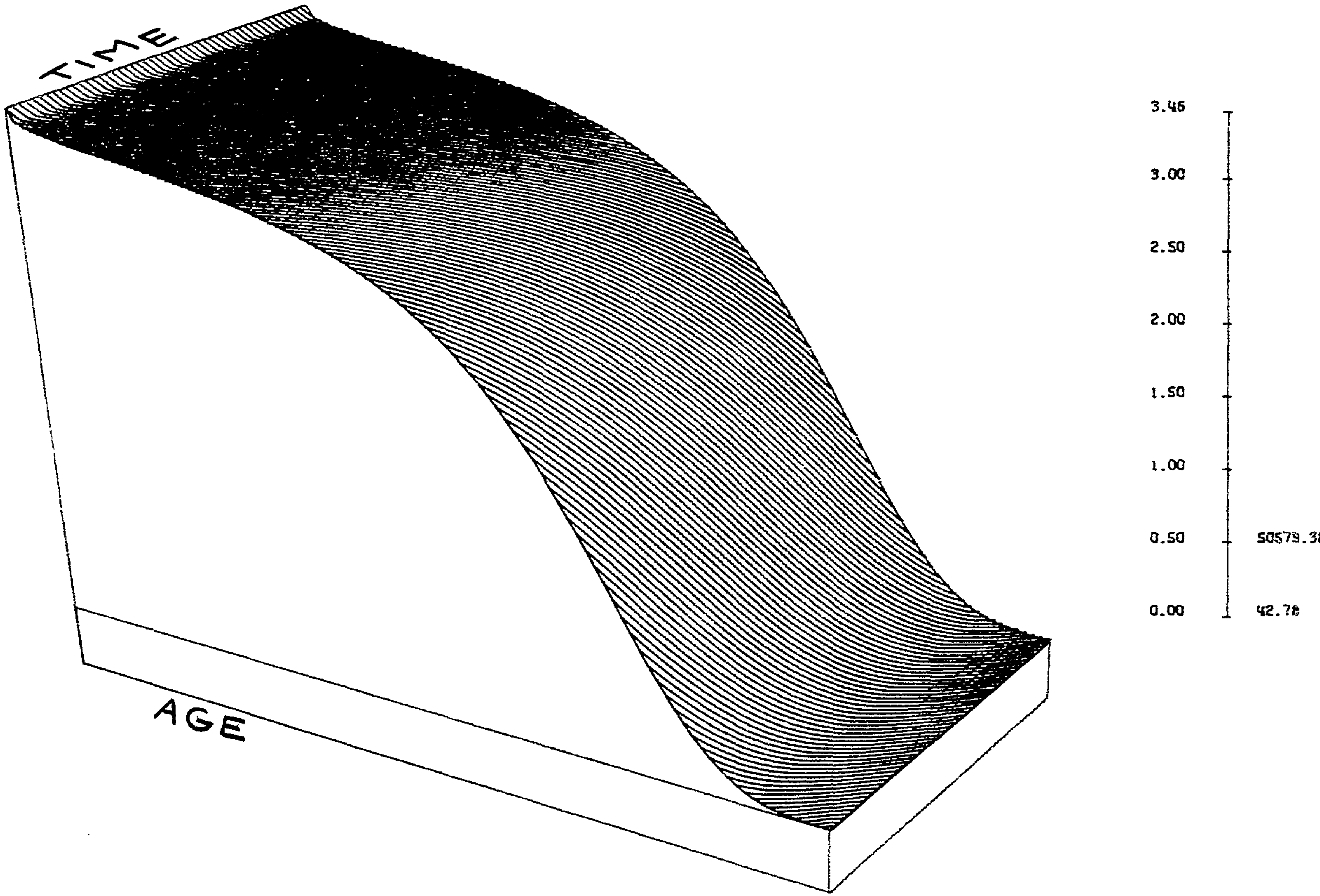
V. Simulation Trial Examples

Simulation runs have been made in order to test the algorithm for numerical solution of the coupled system of ordinary and partial differential equations which constitute the model

Since there is a rather large amount of numerical data associated with each simulation run, the results are produced by the simulation program in a visual as well as numerical format. This is accomplished through a plotting routine which constructs perspective drawings of the three dimensional surfaces generated by a simulation run.

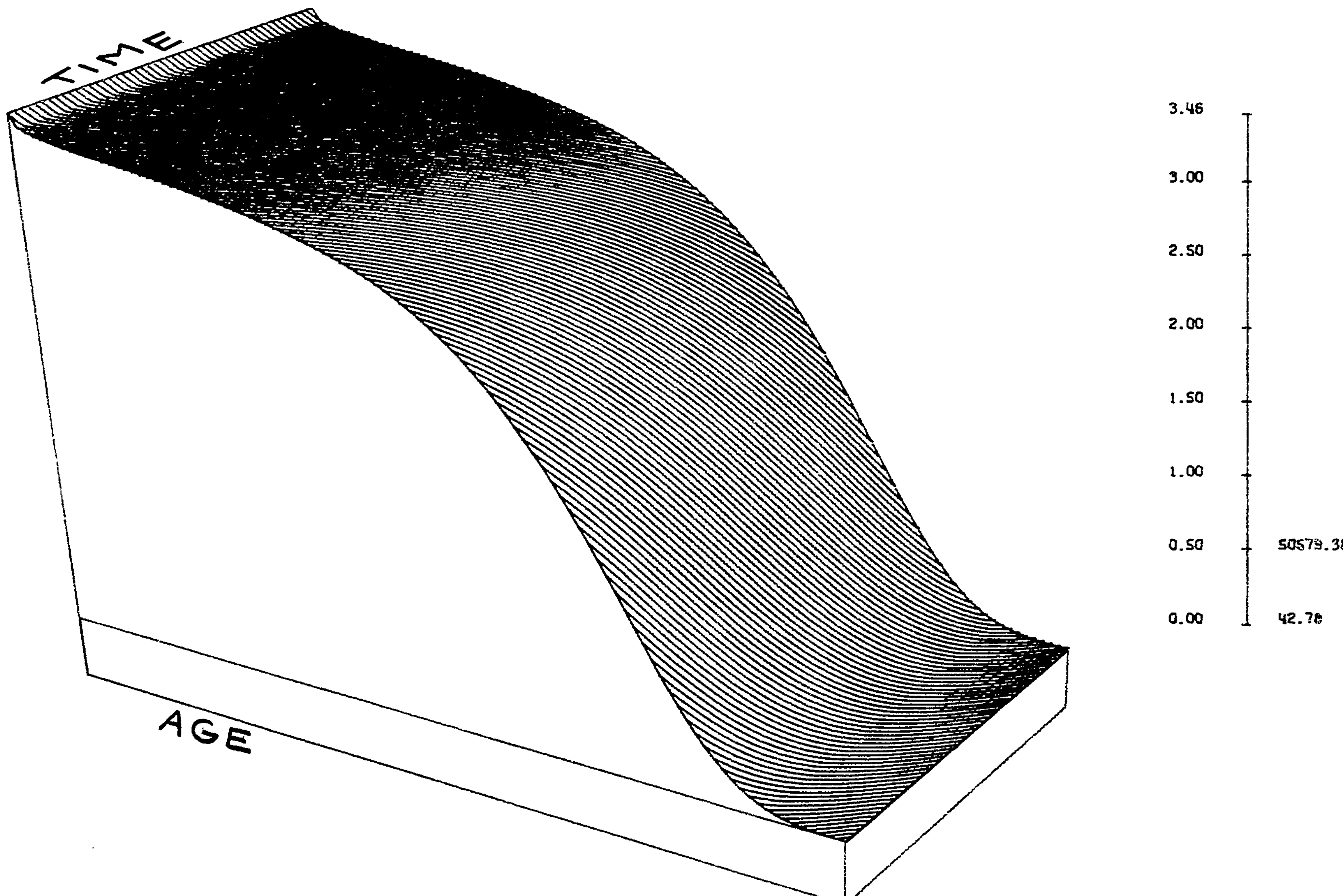
The (steady state) age distribution which results from a constant birth-rate and an absence of immigration is illustrated in Figure 2. The age distribution which results in this case is of course determined solely by the death-rate.

The wave-like nature of the solutions of the governing equations may be clearly observed in a simulation which creates a rise in the fertility curce, starting from an initial condition of the steady state illustrated in Figure 3. Since the dynamics governing $a(t)$ and $b(t)$ have not yet been determined, a simulation has been carried out by introducing $b(t)$ as an exogenous variable; $a(t)$ was determined through the



EVOLUTION OF POPULATION OVER 50 YEARS: CONSTANT PROFILE

Figure 2.



EVOLUTION OF POPULATION OVER 50 YEARS: CONSTANT PROFILE

Figure 2.

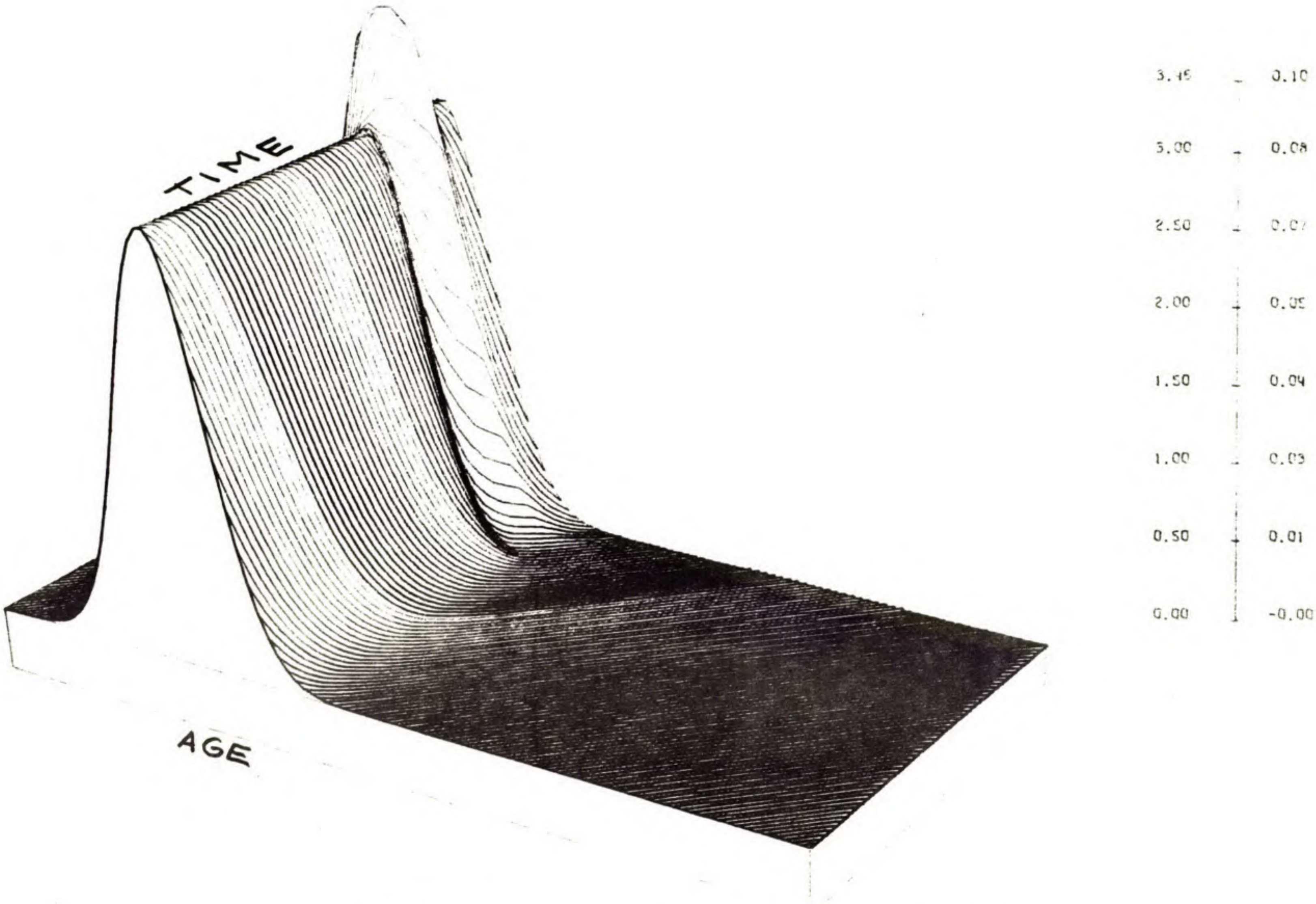
dynamic equation

$$\left(\frac{d}{dt} + 1\right)^2 a(t) = b(t) \quad .$$

The functions $b(t)$ and $a(t)$ have been determined so that $\int_0^T a(t)\, dt = \int_0^T b(t)\, dt = 0$, so that the fertility curve returns to its original value. This produces the response to Figure 4 in the fertility curve, which corresponds to a "baby boom" of duration approximately five years. The effect of this rise in the fertility curve on the age distribution is illustrated in Figure .

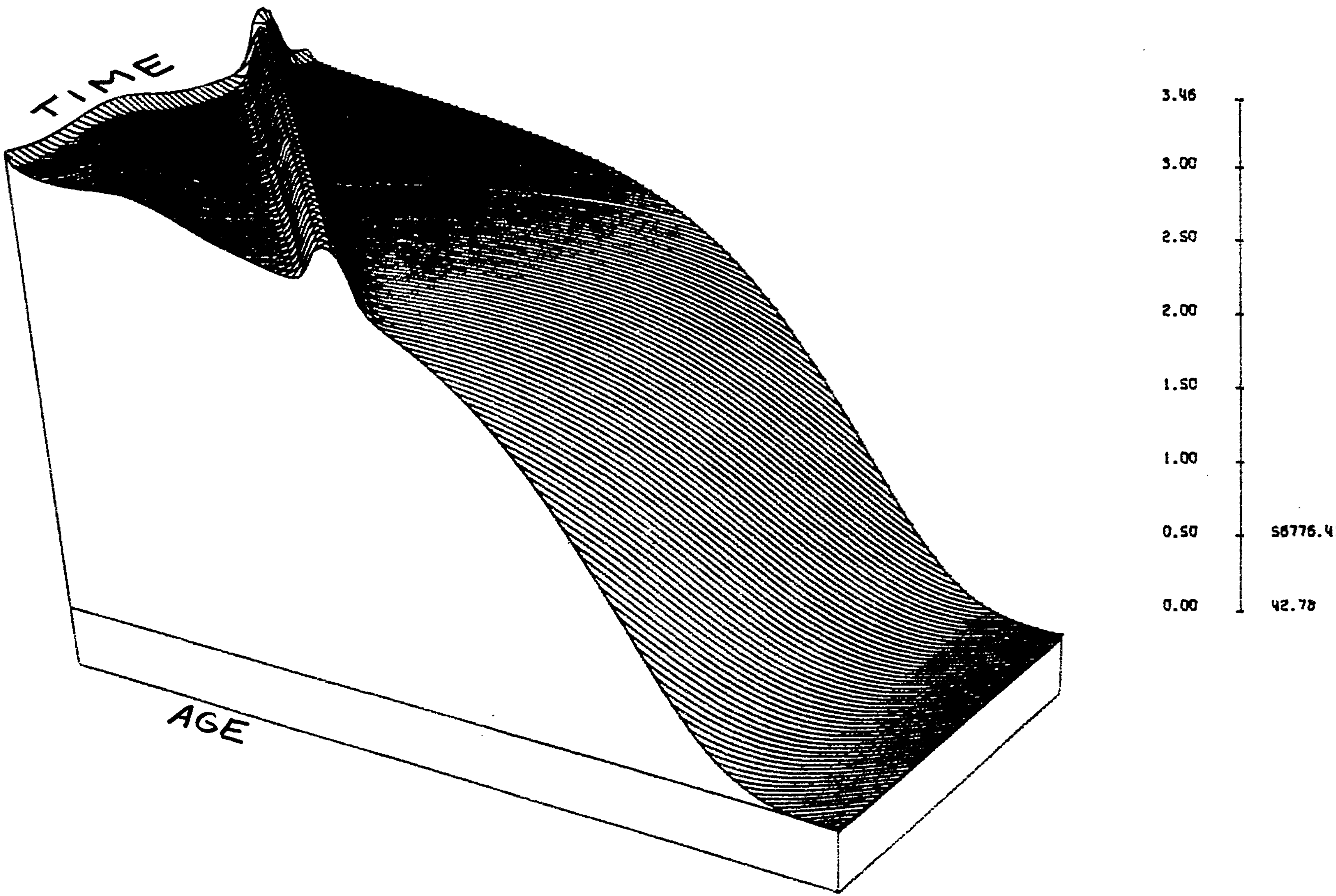
The varying total birth-rate may be clearly seen at the back edge of the figure; the secondary rise in the birth-rate which occurs as the original "offspring" of the boom pass through the childbearing ages is plainly visible. It is also easy to see the original boom passing as a wave through the age structure.

Both the wave nature of the solutions and the birth-rate variations which occur due to a non-uniform age distribution are illustrated in Figure 5 . This output results from an initial age distribution which is significantly different from the steady-state distribution. Such a distribution might be viewed as



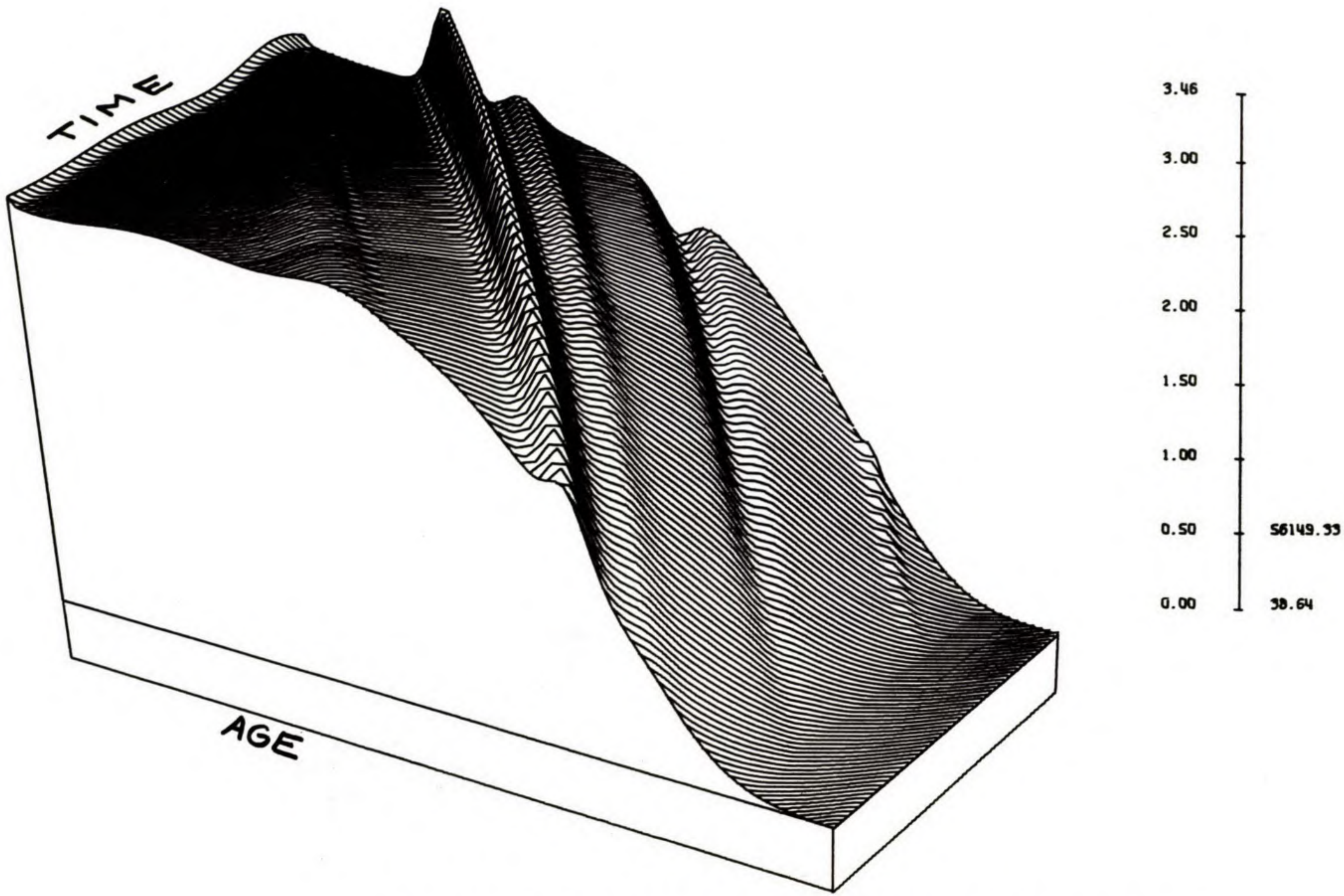
EVOLUTION OF FERTILITY OVER 50 YEARS: WITH FERTILITY PROGRESSION

Figure 3.



EVOLUTION OF POPULATION OVER 50 YEARS: WITH FERTILITY PROGRESSION

Figure 4.



EVOLUTION OF POPULATION OVER 50 YEARS: WITH UNCHANGING FERTILITY

Figure 5.

the result of variations in fertility and immigration which have occurred previous to the time interval covered by the simulation. In this simulation the fertility curve has been held constant, so that the birth-rate variations which occur are due to the varying number of people in the child-bearing age brackets.

VI. Future Work

As of the time of writing of this report, the model has progressed to the point that the basic structure has been established, and the crucial numerical problems associated with the model are well in hand. In particular, numerical methods have been devised for the simultaneous integration of the partial differential equations involved in the model core dynamics and the accuracy of the method has been tested by means of comparison with explicit special case solutions of the model equations obtained by analytical means. Also, a considerable amount of effort has been expended on the problem of devising efficient numerical procedures for extracting estimates of the functional coefficients of the partial differential equations from date available in the form of a histogram. This algorithm has been tested again by use of explicit solutions of the governing equations, and has produced accurate results in the tests.

With these two obstacles removed, the next step in the development of the model is to begin the process of modelling the dynamics of the interaction between the parameters occurring in the core section of the model and various economic variables. The first step in this procedure is to apply the estimation algorithms to the actual historical data in order to determine the time history of $a(t)$, $b(t)$ and the variations associated with the economic mobility μ . Once these functions have been extracted from the data, various approaches to establishing the interaction can be started.

The problem of determining the interaction will first be treated by conventional time-series techniques, that is, correlation analysis based on the assumption of a linear dynamical system of finite dimension as the dynamical intermediary between the economic variables and those in the model core. More recent methods associated with input-output analysis of control systems and identification from operating records will also be tried if the time-series methods are found unsuitable. The relative effectiveness of these two techniques will probably depend on the "actual" location of the stochastic noise element in the real system, length of the operating records, and other factors which are difficult to predict in advance. Non-linear regression methods and techniques of non-linear system identification are held in reserve in case the above methods prove incapable of modelling the interactions.

After a suitable dynamic model of the interaction effects has been determined, full scale model simulations may begin. This requires models to generate economic variables as mentioned above in Section II, and it is currently planned to adapt standard economic models to this purpose. It is also expected that in this context stochastic as well as deterministic simulations will be carried out. This is desirable for two reasons: first, it is a means of assessing the sensitivity of the overall model; second, it is clearly more realistic to model economic behaviour to include random fluctuations if possible.

Finally, we mention that there is a considerable amount of additional work which should be carried out in connection with an investigation of this model. In this area, we mention here only two possibilities which might be considered. The first is described here only because of its possible relevance to the problem of interaction identification discussed above.

The model as it currently exists has been formulated on a "macroscopic" level, that is, the processes which transfer people from one level to another as well as birth and death processes have been modelled as occurring continuously in time. On the microscopic level of an individual, these processes obviously occur at discrete instants of time and are most suitably modelled as a stochastic process. Such modelling will involve the determination of the probability of the occurances of the various "elementary events" which occur on the microscopic level. In such a model, the various interaction effects which are to be estimated in the continuous model appear in the form of dependence of transition probabilities on the current state of the other variables involved in the model. On this level, there is then the possibility of estimating these transition probabilities and their dependence on the other model variables, and thus modelling the interaction effects directly. Close examination of the resulting stochastic process model should then shed light on the form of the interaction in the continuous model formulation of the problem.

A second area where very useful work may be done is in the area of the construction of highly effecient numerical methods for the solution of the governing equations. In our work so far relatively standard numerical techniques have been used for numerical integration of the evolution equations. It may be quite possible to make use of the special forms of the equations to construct more efficient methods. Some preliminary investigation indicates that methods based on Lie algebraic techniques hold promise in this regard. It will be especially important to the usefulness of the final simulation programs that program execution time be kept as low as possible in order that the required number of repeated simulations may be carried out at a reasonable cost.

Appendix A: Analytical Properties of the Governing Equations

In this Appendix we report some of the analytical properties of the partial differential equations governing the dynamics of the population distribution and the fertility curve. The study of these analytical properties in itself provides considerable insight into the problems of population dynamics, as well as providing material essential for the testing of the accuracy of numerical methods developed for use in the model.

It was mentioned in Section III that the population distribution equation enjoys an invariance property which makes it unnecessary to specify in the model formulation the exact measure of income represented by the variable s .

This can be readily demonstrated mathematically as follows. Suppose that instead of considering the distribution function $p(x,s,t)$ as a function of the income scale s , governed by

$$\frac{\partial p}{\partial t}(x,s,t) = -\frac{\partial p}{\partial x}(x,s,t) - \frac{\partial}{\partial s}(\mu(x,s,t)p(x,s,t))$$

$$+ i(x,s,t) - r(x,s,t)\, p(x,s,t)$$

we ask for the evolution of the distribution expressed as a function of the income measure σ . Here the new scale σ is related to the scale s according to

$$\sigma = \varphi(s)$$

where φ is a monotone, smooth (non-linear) function otherwise arbitrary.

By the Chain Rule,

$$\frac{\partial}{\partial s} = \frac{\partial \sigma}{\partial s} \cdot \frac{\partial}{\partial \sigma} = \varphi'(s)\, \frac{\partial}{\partial \sigma}$$

so that the evolution equation for p becomes

$$\frac{\partial}{\partial t} p(x,\varphi^{-1}(\sigma),t) = -\frac{\partial}{\partial x} p(x,\varphi^{-1}(\sigma),t)$$

$$- \varphi'(\varphi^{-1}(\sigma)) \frac{\partial}{\partial \sigma} (\mu(x,\varphi'(\varphi^{-1}(\sigma)),t) \cdot$$

$$p(x,\varphi^{-1}(\sigma),t)) + i - r\, p(x,\varphi^{-1}(\sigma),t) \quad .$$

The Jacobian rule shows that the population density in terms of x , σ and t is given by

$$\tilde{p}(x,\sigma,t) = \frac{1}{\varphi'(\varphi^{-1}(\sigma))} p(x,\varphi^{-1}(\sigma),t) \quad .$$

Rearranging the previous equation to introduce $\tilde{p}$ gives

$$\frac{\partial \tilde{p}}{\partial t} = -\frac{\partial \tilde{p}}{\partial x} - \frac{\partial}{\partial \sigma} (\tilde{\mu}(x,\sigma,t)\tilde{p}) + \tilde{i} - \tilde{r}\, \tilde{p}$$

with $\tilde{\mu}(x,\sigma,t) = \mu(x,\varphi^{-1}(\sigma),t) \cdot \varphi'(\varphi^{-1}(\sigma)) \quad .$

$$\tilde{i}(x,\sigma,t) = \frac{i(x,\varphi^{-1}(\sigma),t)}{\varphi'(\varphi^{-1}(\sigma))}$$

$$\tilde{r}(x,\sigma,t) = r(x,\varphi^{-1}(\sigma),t) \quad .$$

This identifies the transformation law of the economic mobility, and shows the invariance of the governing equation under such a change of scale.

While the coupled system consisting of the population anf fertility evolution equation is a non-linear one, the non-linear interaction occurs only in the calculation of the instantaneous birth-rate (so long as μ , $a(t)$ and $b(t)$ are treated as exogoneous variables). Since the birth-rate enters only as a boundary condition, it is possible to get useful results from explicit solutions of the equations.

Both the population and fertility equations fall into the class of evolution equations governed by first order partial differential equations. While the equations in general have variable coefficients, they are linear in the dependent variable; hence, in principle, the method of characteristics is applicable.

This observation does not dispose of the problem, however. A principal reason for carrying out the investigation into the analytical properties of the equations is to obtain if possible explicit solutions to the equations. By explicit solutions, we mean solutions obtained in closed form analytically.

These solutions have been used to test the accuracy of the numerical methods used both to integrate the evolution equations and to estimate the functional coefficients of the equations $\mu(x,s,t)$ and $a(t) \cdot d(x)$. In the absence of explicit solutions, only lengthy (and expensive) trial runs with varying step sizes can be employed to attempt to estimate accuracy; with explicit solutions available, it is far easier to estimate the step sizes required for a given level of numerical accuracy.

The above remarks pertain to evaluation of the integration scheme; in the case of the estimation problem, the unavailability of explicit solutions would force one to the use of the integration routine to generate the data on which to test the estimation algorithm. In the case of inaccurate results, it then becomes tedious to determine whether the inaccuracy arises from the estimation scheme, or from the numerically generated data.

It is this need for explicit solutions, at least in particular cases, that has led to the work reported here below. The method of characteristics in general produces a solution in <u>implicit</u> form; it is essentially impossible to carry out the required function inversions numerically with enough control on accuracy to make such implicit solutions useful for our purposes.

Fertility Equation

An explicit solution to the fertility equation may be obtained by the method of characteristics. For

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial x}(a(t)d(x)f) - b(t)\,f$$

introduction of an integrating factor of $e^{\int_0^t b(t)dt}$ reduces the problem to

$$\frac{\partial \bar{f}}{\partial t} = -\frac{\partial}{\partial x}(a(t)d(x)\bar{f}) \quad .$$

Solution of the above by the method of characteristics gives

$$\bar{f}(x,t) = \frac{1}{d(x)}\,\phi\,(h^{-1}(h(x)-\int_0^t a(s)dx))$$

$$\cdot\; d(h^{-1}(h(x) - \int_0^t a(s)ds)) \;,$$

with $h(x) - h(x_o) = \int_{x_o}^{x} \frac{1}{d(x)} dx$,

and $\emptyset(x) = \bar{\bar{f}}(x,0)$.

Population Equation

As may be seen from the above example, explicit solutions are generally very involved in form. For this reason, explicit solutions of the population equation will not be exhibited here. We remark that such solutions may be found; the case in which the death-rate varies linearly with age is one example of use in connection with the estimation problem for $\bar{\mu}$. (Results from this example allow the removal of the death-rate term from the governing equations by means of an integrating factor.)

The use of explicit solutions has some potential use beyond evaluation of numerical methods. This is in the area of decreasing the size and cutting down the execution time for the simulation of the model. This may become important in later phases of development of the model, and will have an effect on the frequency of use of the completed model.

The key to such reduction of time expenditure is the observation that an explicit solution reduces the problem of evolution over an arbitrary time interval to a single function evaluation. This is to be contrasted with the repeated evaluations involved in a numerical integration. Of course, the full benefit of this discrepancy is available only if the interaction effects are specified exogenously. In the case of the full model, however, it seems likely that explicit solutions could be used together with extrapolation methods to improve simulation execution time.

This leads naturally to the question of which classes of coefficient functions tive rise to explicit solution formulae. Of particular interest is the problem of explicit solutions to models in which the coefficient functions appear in "separable form" (see Appendix B below), so that the equation has the form

$$\frac{\partial p}{\partial t} = -\frac{\partial p}{\partial x} - \frac{\partial}{\partial s}\left(\sum_{i=1}^{N} \eta_i(t)\nu_i(x,s)p\right) - \xi(t)\ r(x,s)\ p$$

Progress in the direction of explicit solutions to the above equation may be made by recourse to the theory of Lie Algebras. In particular, if the Lie algebra generated by the partial differential operators on the right side of the above equation is solvable, then (global) explicit expressions are possible. Other conditions on the Lie algebra lead to (local) results which may prove useful.

Appendix B:

Numerical Methods for Partial and Ordinary Differential Equations

In Section III of the report, the population and fertility are dynamically modelled by a pair of partial differential equations:

$$\frac{\partial p}{\partial t} = -\frac{\partial p}{\partial x} - \frac{\partial}{\partial s}(\mu(x,s,t) \cdot p) - r(x,s,t)p + i(x,s,t)$$

$$\frac{\partial f}{\partial t} = -a(t)\frac{\partial}{\partial x}(d(x)f) - b(t)f \quad .$$

In Section IV, techniques for estimating the parameters $\mu(x,s,t)$, $a(t)$, $d(x)$ are discussed and there, it is pointed out that full simulation of the overall model requires dynamic modelling of these functions using partial and/or ordinary differential equations.

Since this appendix deals with aspects of the actual simulation of the model, it is assumed that ordinary differential equations for $a(t)$ and $b(t)$ have been obtained, estimated values of $d(x)$ and $u(x,s,t)$ have been obtained by use of the estimation procedures described elsewhere in this report, and that initial fertility and population distributions are known: $f(x,s,o)$ and $p(x,s,o)$. Values of $p(x,s,t)$ and $f(x,s,t)$ are required, and these are simulated using numerical techniques. The techniques have been chosen in order to be consistent with the conservation law character of the governing equations, to attain a reasonable accuracy in the simulated values subject to restrictions on the size of data groupings which are expected in currently available data, and to balance these with economy of computation.

In the numerical simulation which has been carried out up to the time of writing of this report, the income level dependence of the population density has been suppressed. As well as yielding computational efficiencies during the development of the model, this procedure has made the analysis of numerical problems arising in the modelling considerably easier. The extension of the numerical methods developed so far to include the income variable s is expected to cause no significant difficulty, as the problems which arise should parallel those already encountered.

Numerical Integration of the Fertility Equation

The fertility equation has been transformed by an integration factor to

$$\frac{\partial \bar{f}}{\partial x}(x,t) = -a(t)\frac{\partial}{\partial x}(d(x)\bar{f}(x,t))$$

where

$$f(x,t) = e^{-\int_o^t b(t)dt}\,\bar{f}(x,t) \quad .$$

An approximation F_{ij} to $\bar{f}(ih,jk)$ is obtained using

$$F_{ij+1} = F_{ij} - \frac{k}{2h}a_j(d_{i+1}F_{i+1j} - d_{i-1}F_{i-1j})$$

$$+\frac{1}{2}\left(\frac{ka_j}{h}\right)^2 [d_{i+\frac{1}{2}}(d_{i+1}F_{i+1j} - d_iF_{ij})$$

$$- d_{i-\frac{1}{2}}(d_iF_i - d_{i-1}F_{i-1j})]$$

and $f(ih,jk)$ is estimated by F_{ij} using numerical integration for the integration factor (see below). This scheme is almost second order, and has two desirable properties: For $d(x) = d$ constant, it is numerically stable provided that the step size ratio is chosen to satisfy

$$\frac{k}{h} \leq \left|\frac{1}{d \cdot a_j}\right| \quad .$$

Also this scheme has the property that it removes a distortion of the fertility profile when the effect causing the distortion is removed. This property is exhibited in Figure 3.

The dynamical equations governing the variables $a(t)$ and $b(t)$ must of course be integrated simultaneously with the partial differential equations. Numerical approximations are currently calculated using a modified Euler method over time steps of length k .

While this procedure may be easily replaced by a more accurate process, this method was selected in view of the decision to use simple routines initially as an aid to algorithmic development, and later to replace these by more sophisticated routines as dictated by accuracy and economy in large simulations.

Since dynamic modelling of $a(t)$ and $b(t)$ has not yet been carried out, the dynamics

$$(D+1)^2 \; a(t) = b(t)$$

have been assumed in order to verify the integration methods. In this case

$$\underline{a}(t^*) = a(t) + h\underline{f}(\underline{a}(t))$$

$$\underline{a}(t+k) = \underline{a}(t) + \frac{h}{2}[f(\underline{a}(t)) + f(\underline{a}(t^*))]$$

where

$$\underline{f} \begin{bmatrix} I\ b(t) \\ I\ a(t) \\ a(t) \\ a^1(t) \end{bmatrix} = \begin{bmatrix} b(t) \\ a(t) \\ a^1(t) \\ b(t) - a(t) - 2a^1(t) \end{bmatrix}$$

describes the modified Euler method.

Numerical Integration of the Population Equation

Initially only the age-time dynamics of population have been considered; hence the equation is

$$\frac{\partial p}{\partial t}(x,t) = -\frac{\partial p}{\partial x}(x,t) - r(x,t)\cdot p(x,t) \quad .$$

To solve this numerically, we approximate $p(x,t)$ by $P(x,t)$ where

$$P(x+h,t+h) = (1-r(x,t))\, P(x,t) \quad .$$

The fertility is used to estimate the population birth-rate

$$p(o,t+h) = \int_0^{\infty} p(x,t)\, f(x,t)\, dx \quad ,$$

and this is approximated numerically by

$$P(o,t+h) = \sum_{i=1}^{100} P(x,t)\, F(x,t+\frac{h}{2})$$

where $F(x,t+\frac{h}{2})$ is obtained from the numerical approximations of the fertility curve.

The low accuracy method for simulating the population is reasonably accurate for that section of the profile where the death-rates are almost constant. It is expected that improvements will be possible after additional work. Improvements in the simple scheme used for estimating $P(o,x+h)$ would lead only to a change in scale of values, but not their dynamics.

In conclusion we point out that certain portions of the model are particularly sensitive to errors - that is small errors may lead to very inaccurate simulations of the dynamics, whereas other portions of the model are not so sensitive. For this reason, it is possible (and economically reasonable) to tailor the accuracy of the methods used to the sensitivity of that part of the model being simulated.

Appendix C: Numerical Determination of Partial Differential Equation Coefficients

In Section IV it is shown that integration of the partial differential equations leads to analytic formulas for the estimation of $\mu(x,s,t)$ from the population equation, and $a(t) \cdot d(x)$ from the fertility equation. To use these formulas, available data must be used to estimate the quantities required. In particular, it is required that histogram data be used

1. to generate (continuous) density functions, that
2. partial derivatives of these density functions be estimated, and that
3. the required integrals be estimated.

The distributions involved appear to be very smooth, and as a result piecewise approximation by polynomials with continuous first derivatives is necessary; additional smoothness is desirable. The algorithm employed is described below for the problem of estimating the economic mobility $\mu(x,s,t)$. The procedure for estimating the term $a(t)\ d(x)$ in the fertility equation is entirely similar.

1. A function which might be best described as a fourth-order spline (having three continuous derivatives) is determined so that its integrals over the appropriate intervals are equal to the given values from the histogram data.

2. Differentiation of the fourth-order spline with respect to the x-variable provides an estimate of $\frac{\partial p}{\partial x}$; determination of an additional cubic spline in the t-variable followed by a t-differentiation provides an estimate of $\frac{\partial p}{\partial t}$.

3. Finally, the required integral is estimated by integration of the result of 2. above.

In the use of spline methods in approximation problems, it is necessary to provide additional boundary conditions beyond the requirement that the spline interpolate the appropriate sample points. Unfortunately, in the present application, use of the so-called "natural boundary conditions" was found to give particularly poor estimates near the boundary of the region involved. Further, it was found that these errors were quite sensitive to the values assigned.

After considerable experimentation, it was determined that adequate results could be obtained through estimation of the third derivatives near the endpoints by third order finite differences, and use of this data to determine the boundary conditions. The scheme for approximating the distribution function requires the solution of a system of linear equations including three different types

$$\text{(a)} \quad r''_i - 2r''_{i-1} + r''_{i-2} - \frac{r''''_i}{2} - \frac{r''''_{i-1}}{2} = 0 \qquad i = 2, \ldots, 100$$

$$\text{(b)} \quad r_i - 2r_{i-1} + r_{i-2} - \frac{r''_i}{6}$$

$$- \frac{2}{3} r''_{i-1} - \frac{r''_{i-2}}{6} + \frac{r''''}{24} \ddagger \frac{r''''}{24} = 0 \qquad i = 2, \ldots, 100$$

$$\text{(c)} \quad r_i + r_{i-1} - \frac{r''_i}{12} - \frac{r''_{i-1}}{12} + \frac{r''''_i}{60} = 2P_i \qquad i = 1,\ldots,100$$

where P_i is the number of persons between ages i and $i + 1$, and the boundary conditions are

$$r''_0 = r''_{100} = 0 \quad , \quad r'''_0 = 0 \Rightarrow -r''_0 + r''_1 - \frac{r''''_1}{2} = 0$$

$$r''''_0 = 0 \ .$$

Here r_i is interpreted as the population distribution at age i years, and it is assumed that values P_i are available for $i = 0,1,\ldots,99$.

To solve this system, a reduction method for a sparse matrix is used, and the equations are ordered so that coefficients of moderate size are maintained on the diagonal. For a test distribution, the error in regenerating the histogram was less than 1 percent.

Spline Approximation

A standard analysis for cubic spline approximation represents this function in terms of estimates for the second derivative at nodes. These are obtained as solutions of the system of equations

$$\frac{h_i M_{i-1}}{h_i+h_{i+1}} + 2M_i + \frac{h_{i+1} M_{i+1}}{h_i+h_{i+1}} = 6\,\frac{\dfrac{r_{i+1}-r_i}{h_{i+1}} - \dfrac{r_i - r_{i-1}}{h_i}}{h_i + h_{i+1}}$$

to interpolate $\{r_i\}$ with a spacing $\{h_i\}$ (which for our model is either h or k constant), and boundary conditions used are

$$\frac{M_1-M_o}{h} = \frac{1}{h^3}\,(-r_o+3r_1-3r_2+r_3)$$

$$\frac{M_n-M_{n-1}}{h} = \frac{-1}{h^3}\,(-r_{n-3}+3r_{n-2}-3r_{n-1}+r_n)$$

With this approximation, errors in $\mu(x,s,t)$ and $a(t)$ $d(x)$ obtained using the estimation procedures are less than 1 percent on the interior of the domain. Although errors are large where $P(x,s,t)$ is small, values of $\mu(x,s,t)$ there are not crucial (see Section IV).

Separability of Coefficient Functions

It was mentioned in Section III that one aspect of the structure of the model was that it was formulated in such a way as to make it possible to model the feedback effects on the core section of the model in terms of a finite number (even a small finite number) of functions of time. This was illustrated in Section III in the hypothesis that the effect of the rest of the world on the fertility curve could be adequately modelled by

$$\frac{\partial f}{\partial t} = -a(t)\frac{\partial}{\partial x}(d(x)f) - b(t)f \quad .$$

In this formulation, the world affects f only through $a(t)$ and $b(t)$. However, it was also mentioned that

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial x}(a(x,s,t)f) - b(t,s)f$$

might well represent a more realistic model, and it is easy to verify that the estimation procedure described in Section IV and Appendix C above will equally well produce an estimate of the coefficient function $a(x,s,t)$. A problem that arises, then, is that of distinguishing between $a(x,s,t)$ and $a(t)\cdot d(x)$ at the stage of the output of the estimation algorithm. From the point of view of subsequent modelling effort, it might well be hoped that the result had the form $a(t)\ d(x,s)$, or perhaps even $a(t) \cdot d(x) \cdot c(s)$.

A similar problem occurs in the case of estimation of $\mu(x,s,t)$, where the validity of a representation of the form

$$\mu(x,s,t) = \eta_1(t)\ \nu_1(x,s) + \ldots + \eta_N(t)\ \nu_N(x,s) \quad ,$$

with N a "reasonably small" integer is at least a practical requirement for the success of any attempt to model the interaction affects.

Given the implicit smoothness assumptions on the coefficient functions of the model and the fact that the ranges of the x , s , and t variables involved are finite, there is no problem in applying standard approximation theorems to deduce that μ may be closely approximated by a function of the above form. (A similar remark obviously applies to $a(x,s,t)$). For convenience, we refer to the above form as a "separable representation for μ ".

Since we have shown above that separable representations exist, the only problems which remain are those of the number N of terms to be expected in the representation, and the numerical determination of N and $\eta_i(t)$ from the available data.

Consideration of the effects that the t variation in μ (and a) is intended to model, and the probable variability of these effects across age and income brackets suggests strongly that N is small. It would be surprising if N were greater than 3 in the case of the estimation of $\mu(x,s,t)$, and it appears entirely possible that a single term will suffice in the case of the estimation of $a(x,s,t)$.

It remains to show the feasibility of determining separability of the representation numerically. To distinguish a separable μ from a non-separable one we proceed as follows:

A smooth function $\mu(x,s,t)$, defined for $s \in S$, $t \in T$, $x \in X$, with S, T, X compact subsets of R' defines the kernel of a compact linear operator L mapping from $L^2(X \times S) \to L^2(T)$ according to the formula

$$Lf(t) = \int\int_{X \times S} \mu(s,s,t)\, f(x,s)\, dxds \quad .$$

Now a <u>separable</u> μ is distinguished by the fact that the associated L is an operator of finite dimensional range, and this observation essentially solves the problem.

When $\mu(s,s,t)$ has been estimated numerically, μ is not obtained as a continuous function. What is obtained in fact is a set of sample values $\{\mu(x_i, s_j, t_k)\}$, with $x_i \in \tilde{X}$, $s_j \in \tilde{S}$, $t_k \in \tilde{T}$. Here $\tilde{X}$, $\tilde{S}$, $\tilde{T}$ are each Euclidean space of dimension equal to the number of sample points in each of the independent variables. The discrete analog of the definition of L is to use the above three dimensional array to define a linear mapping (matrix) $\tilde{L} : R^{\dim(\tilde{X}\times\tilde{S})} \to R^{\dim \tilde{T}}$.

The problem of finding the $\eta_i(t)$ is now equivalent to determining the range space of $\tilde{L}$, and N is simply the rank of $\tilde{L}$.

The problem is simplified still further by invoking the fact that

$$\text{Range } \tilde{L} = \text{Range } \tilde{L}\,\tilde{L}^* ,$$

where $\tilde{L}^*$ is the adjoint (actually transpose in this case) of the matrix $\tilde{L}$. This reduces the problem to the entirely standard one of an eigenvector/eigenvalue analysis of a symmetric matrix, and hence effectively solves it.

Appendix D Computer Program Listing

In this appendix we list the computer programs developed up to the time of this report for use with the model. Included below are both the programs used for numerical integration of the governing evolution equations in simulation runs, and the programs designed to estimate model coefficients from the available data.

The programs listed here are written in FORTRAN. Given the relatively large arrays of data which must be handled in connection with this model, it is clear that FORTRAN is not the most convenient language in which to program the numerical algorithms required. With a view to future uses of the model, however, such factors as the wide availability of FORTRAN compilers, the existence of the I.B.M. CSMP (Continuous System Modelling Package) which is FORTRAN compatible, and of FORTRAN packages for the Calcomp plotter used to produce output data plots make FORTRAN a reasonable language choice.

Program I

Simulation of population and fertility propagation over time. Initial age-specific profiles of population, fertility and mortality, and dynamics for $a(t)$ and $d(x)$ are required. Here, the values for population are taken as the number of live individuals at age x in the population as given by the Commissioner's 1941 Standard Ordinary Mortality Table. Values for mortality are also taken from this table.

Values for the fertility are given by the artificial distribution:

$$f(x,o) = C\,e^{-\frac{(x-24.5)^2}{8}}$$

where the constant C is chosen so that

$$\int_0^{100} f(x,o)\,p(x,o)\,dx = p(o) \quad .$$

Values of $d(x)$ are assumed from (the artificial distribution)

$$d(x) = \frac{\left(1-e^{-\frac{(x+1)^2}{600}}\right)}{1+e^{\frac{(x-24.5)^2}{4}}}$$

The parameter $a(t)$ is determined by numerical integration of the differential equation

$$(D+1)^2 \, a(t) = b(t)$$

where

$$b(t) = -\frac{\sin(e^{-|t-5|} - e^{-|t-10|})}{15} \qquad 0 \leq t < 15$$

$$= 0 \qquad t > 15$$

```
      $JOB    ACCT-NUM,'VERNEP',TIME=60
              REAL X(100)
              IC=100
              IF=0
              DO 9 IR=1,10
              CALL DATA(X,IR,IC,IF)
        9         CONTINUE
              STOP
              END

              SUBROUTINE DATA(X,IR,IC,IF)
              REAL D(100),F(100),P(100),X(100),R(100),US(100),D5(100),F1(100)
              REAL HH,KK,SUM,IV(10),DLY(10)
              REAL PLTPER,PRTPER,I1
              INTEGER YRS,AGES,YOUNG,NN
      C       ************************************************************
      C       SET IFORM=0 IF YOU WANT PLOT,    =1 IF YOU WANT PRINT
      C       ************************************************************
              IFORM=1
              PL=1E-5
              PR=PL
              TINT-1E-4
              ICM1-IC-1
              IF (IR.GT.1) GO TO 31
      C       RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
      C       HH IS AGE STEP   -   KK IS TIME STEP
              READ(5,101) HH,KK
        101   FORMAT(2F10.6)
      C       PLTPER IS PLOT INTERVAL   -  PRTPER IS PRINT INTERVAL
              READ(5,101) PLTPER,PRTPER
      C       NDE IS THE NUMBER OF FIRST ORDER ODE'S AFTER TRANSFORMATION
      C       IF NDE > 10 CHANGE REAL DIMENSION IN DATA AND $STEP
      C       NN IS THE NUMBER OF INTEGRATION STEPS IN KK FOR SOLVING THE ODES
              READ(5,103) NDE, NN
        103   FORMAT(2I4)
      C       PUT INITIAL VALUES OF ODES ONE TO A CARD
              DO 97 I=1,NDE
                  READ(5,107) DLY(I)
        107       FORMAT(F10.4)
        97        CONTINUE
      C       RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
              PR=PR+KK
              T=0.0
      C       ************************************************************
      C                           READ DEATH RATES
      C                         READ POPULATION DENSITY -
      C       ************************************************************
      C       RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
              READ(5,100) (R(I),I=1,IC)
        100   FORMAT(5F10.6,30X)
              READ(5,104) (P(I),I=1,IC)
        104   FORMAT(10F8.0)
      C       RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
              SUM=0.0
              DO 99 I=1,IC
                  SUM=SUM+P(I)
        99    CONTINUE
              DO 93 I=1,IC
                  P(I)=P(I)*22000000./SUM
              X(I)=P(I)
```

```
   98   CONTINUE
C       PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
        IF(IFORM.NE.1) GOTO 61
        TPG=.
        WRITE(6,109)
  109   FORMAT('1','POPULATION FOR AGES  1  TO 20  IS')
        WRITE (6,102) (P(I),I=1,20)
  102   FORMAT(5F10.0/)
        WRITE (6,110)
  110   FORMAT(' ','AT BEGINNING OF YEAR              0, AND')
        WRITE (6,111)
  111   FORMAT(' ','-----------------------------------')
   61   CONTINUE
C       PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
        DO 70 I=1,IC
             I1=FLOAT(I)
             D(I)=(1-EXP(-I1*I1/600.))/(1+((I1-25.5)**2)/4.)
   70        CONTINUE
        YOUNG=0
        F1(1)=EXP(-(.125*HH*(YOUNG-24.5))**2)
          SUM=F1(1)*P(1)
        DO 14 I=2,IC
             F1(I)=EXP(-(.125*HH*(I+YOUNG-25.5))**2)
             SUM=SUM+F1(I)*P(I)
             D5(I)=.5*(D(I)+D(I-1))
   14        CONTINUE
        D05=.5*(3.*D(1)-D(2))
        DT5=.5*(3.*D(IC)-D(IC-1))
        SUM=P(1)/SUM
        DO 15 I=1,IC
             F1(I)=F1(I)*SUM
   15        CONTINUE
        PA=KK/HH
        PA1=0.0
        P2=0.0
        FA=1
        GO TO 33
   31   CONTINUE
        PR=P2+KK
C       ****************************************************************
C                             UPDATE COEFFICIENTS OF FERTILITY EQUATION
C       ****************************************************************
        PA1=DLY(2)
        P2=DLY(2)*DLY(2)*PA
        FA=EXP(-DLY(4))
C       PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
        IF (ORTPER.GT.PR.OR.IFORM.NE.1) GOTO 62
        WRITE(6,106) T
  106   FORMAT(' AT TIME',F6.2)
        WRITE(6,112) DLY(1)
  112   FORMAT('0','DELAY IN PEAK OF FERTILITY CURVE IS  ',F9.6)
   62   CONTINUE
C       PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
C       ****************************************************************
C                             MULTIPLY FERTILITY BY INTEGRATING FACTOR
C       ****************************************************************
   33   CONTINUE
        SUM=0.0
        DO 42 I=1,IC
             SUM=SUM+F1(I)
```

```
                F(I)=FA*F1(I)
   42           CONTINUE
C     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
      IF (PRTPER.GT.PR.OR.IFORM.NE.1) GOTO 63
      WRITE(6,113) SUM
  113 FORMAT(' ','     INTEGRAL OF NORMALIZED FERTILITY IS        ',F9.6)
      WRITE(6,114) FA
  114 FORMAT(' ','         FRACTION OF INITIAL FERTILITY IS
     2,F9.6)
      WRITE(6,115)
  115 FORMAT('0','VALUES OF FERTILITY CURVE FOR AGES  21 TO 30  ARE')
      WRITE(6,108)  (F(I), I=21,30)
  108 FORMAT(5F10.6/)
   63 CONTINUE
C     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
C     *******************************************************************
C                             INCREMENT POPULATION BY ONE YEAR
C     *******************************************************************
      TINT=TINT+KN*2
      IF (HH.GT.TINT) GOTO 45
      SUM=0.
      SUM1=0.
      DO 44 I=1,IC1
                J=IC-I
                SUM=SUM+P(J+1)*F(J+1)
                P(J+1)=P(J)*(1-R(J))
                SUM1=SUM1+P(J+1)
                X(J+1)=P(J+1)
   44           CONTINUE
      SUM=SUM+P(1)*F(1)
      P(1)=SUM
      X(1)=P(1)
      SUM1=SUM1+P(1)
      TINT=-1E-4
   45 CONTINUE
C     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
      IF (PRTPER.GT.PR.OR.IFORM.NE.1) GOTO 64
      WRITE(6,116) SUM1
  116 FORMAT('0','TOTAL POPULATION IS   ',F10.0)
      PR=+1E-5
      WRITE(6,117)
  117 FORMAT('0','**************************************************
     1*************************')
   64 CONTINUE
C     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
C     *******************************************************************
C                             INCREMENT FERTILITY BY ONE YEAR
C     *******************************************************************
      DO 34 I=1,IC
                IF (1E-20.GT.F1(I)) F1(I)=0.0
                US(I)=P(I)*F1(I)
   34           CONTINUE
      UQ=2.*US(1)-US(2)
      UT=2.*US(IC)-US(IC-1)
      F1(IC)=F1(IC)-.5*PA*((PA1*(UT-US(IC-1)))-P2*(DT5*UT+D5(IC)*US
     2(IC-1)-(DT5+D5(IC))*US(IC)))
      F1(1)=F1(1)-.5*PA*((PA1*(US(2)-UQ))-P2*(D5(2)*US(2)+DQ5*UQ-(D5
     2(2)+DQ5)*US(1)))
      DO 35 I=2,IC1
                F1(I)=F1(I)-.5*PA*((PA1*(US(I+1)-US(I-1)))-P2*(D5(I+1)*US
```

```
      2(I+1)+D5(I)*US(I-1)-(D5(I+1)+D5(I))*US(I)))
   35    CONTINUE
C     ******************************************************************
C                 SOLVE ODES FOR DELAY PARAMETERS AND INCREMENT TIME
C     ******************************************************************
      DO 36 I=1,NDE
         IV(I)=DLY(I)
   36    CONTINUE
      CALL $RUN(IV,DLY,T,KK,NDE,NN)
      PL=PL+KK
      IF (PLTPER.GT.PL) GOTO 31
C     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
      IF (IFORM.NE.1) GOTO 65
      IPG=IPG+1
      IF (IPG.NE.2) GOTO 79
      IPG=0
      WRITE(6,121)
  121 FORMAT('1')
   79 CONTINUE
      WRITE(6,118)
  118 FORMAT(' ','POPULATION FOR AGES  1 TO 20  IS')
         WRITE(6,102) (P(I), I=1,20)
      WRITE(6,119) IR
  119 FORMAT(' ','AT BEGINNING OF YEAR',I8,' AND')
      WRITE(6,120)
  120 FORMAT(' ','-----------------------------------')
   65 CONTINUE
C     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
      RETURN
      END

      SUBROUTINE $RUN(IV,Z,T,KK,NDE,NN)
C     ******************************************************************
C     THIS SUBROUTINE GIVES A SIMULTANEOUS SIMULATION OF VALUES
C     Z=(IA(T),A(T),A'(T),IB(T),B(T)) AT  T+H  FROM INITIAL VALUES  IV
C     OF THE SAME VARIABLES AT T.  FOR EACH DIFFERENT SIMULATION THE
C     SUBROUTINE  $DE  MUST BE REWRITTEN TO EVALUATE THE DERIVATIVES
C     D(I)  AS A FUNCTION OF TIME  T  AND SOLUTION VALUES  Y(I)
C     ******************************************************************
      REAL IV(NDE),Z(NDE),T,T2,KK,K1
      K1=KK/FLOAT(NN)
      T2=T
      T=T+KK
   28 CONTINUE
      CALL $STEP(IV,Z,T2,K1,NDE)
      T2=T2+K1
      DO 29 I=1,NDE
         IV(I)=Z(I)
   29    CONTINUE
      IF (T.GT.T2+K1/2) GOTO 28
      RETURN
      END

      SUBROUTINE $STEP(IV,Z,T,KK,NDE)
      REAL IV(NDE),Z(NDE),Z1(10),D(10),D1(10),KK
      CALL $DE(T,IV,NDE,D)
      DO 11 I=1,NDE
         Z1(I)=IV(I)+KK*D(I)
   11    CONTINUE
      TT=T+KK
```

```
      CALL $DE(TT,Z1,NDE,D1)
      DO 12 I=1,NDE
           Z(I)=IV(I)+.5*KK*(D(T)+D1(I))
12         CONTINUE
      RETURN
      END

      SUBROUTINE $DE(T,Y,NDE,D)
      REAL Y(NDE),D(NDE),KK
      D(1)=Y(2)
      D(2)=Y(3)
      T1=0.
      IF (T.GT.15.) GOTO 78
      TT1=ABS(T-5.)
      TT2=ABS(T-10.)
      T1=(-SIN(EXP(-TT1)-EXP(-TT2)))/15.
78    CONTINUE
      D(3)=15.*T1-Y(2)-2.*Y(3)
      D(4)=T1
      D(5)=0.
      RETURN
      END

   $ENTRY
```

```
POPULATION FOR AGES  1  TO 20  IS
  350170.    342263.    340289.    338880.    337734.

  336725.    335795.    334919.    334091.    333320.

  332613.    331958.    331324.    330688.    330033.

  329350.    328642.    327922.    327184.    326431.

AT BEGINNING OF YEA              0, AND
-----------------------------------
AT TIME  0.75

DELAY IN PEAK OF FERTILITY CURVE IS  -0.000373
   INTEGRAL OF NORMALIZED FERTILITY IS          1.089507
      FRACTION OF INITIAL FERTILITY IS                  1.000501

VALUES OF FERTILITY CURVE FOR AGES  21 TO 30  ARE
 0.056025  0.063485  0.069726  0.074224  0.076578

 0.076574  0.074216  0.069721  0.063482  0.056023

TOTAL POPULATION IS   22000110.

********************************************************************************
POPULATION FOR AGES  1 TO 20  IS
  350215.    342263.    340289.    338880.    337734.

  336724.    335795.    334919.    334091.    333320.

  332613.    331958.    331324.    330688.    330033.

  329350.    328642.    327922.    327184.    326431.

AT BEGINNING OF YEA        1,AND
-----------------------------------
AT TIME  1.75

DELAY IN PEAK OF FERTILITY CURVE IS  -0.004762
   INTEGRAL OF NORMALIZED FERTILITY IS          1.089505
      FRACTION OF INITIAL FERTILITY IS                  1.002134

VALUES OF FERTILITY CURVE FOR AGES  21 TO 30  ARE
 0.056127  0.063609  0.069877  0.074401  0.076738

 0.076675  0.074285  0.069795  0.063563  0.056101

TOTAL POPULATION IS   22000510.

********************************************************************************
```

```
POPULATION FOR AGES  1 TO 20  IS
  350562.   342307.   340289.   338880.   337734.

  336725.   335795.   334919.   334092.   333320.

  332613.   331958.   331324.   330688.   330033.

  329350.   328642.   327922.   327184.   326431.

AT BEGINNING OF YEAR        2,AND
---------------------------------
AT TIME  2.75

DELAY IN PEAK OF FERTILITY CURVE IS  -0.020237
   INTEGRAL OF NORMALIZED FERTILITY IS         1.089506
      FRACTION OF INITIAL FERTILITY IS                 1.006584

VALUES OF FERTILITY CURVE FOR AGES  21 TO 30  ARE
 0.056417  0.063967  0.070328  0.074940  0.077210

 0.076924  0.074418  0.069958  0.063759  0.056301

TOTAL POPULATION IS   22001840.

*******************************************************************************
POPULATION FOR AGES  1 TO 20  IS
  351507.   342646.   340332.   338880.   337734.

  336724.   335795.   334919.   334091.   333320.

  332613.   331958.   331324.   330687.   330033.

  329350.   328642.   327922.   327184.   326432.

AT BEGINNING OF YEAR        3,AND
---------------------------------
AT TIME  3.75

DELAY IN PEAK OF FERTILITY CURVE IS  -0.064413
   INTEGRAL OF NORMALIZED FERTILITY IS         1.089506
      FRACTION OF INITIAL FERTILITY IS                 1.018705

VALUES OF FERTILITY CURVE FOR AGES  21 TO 30  ARE
 0.057215  0.064960  0.071590  0.076457  0.078518

 0.077577  0.074741  0.070374  0.064276  0.056834

TOTAL POPULATION IS   22005710.

*******************************************************************************
```

POPULATION FOR AGES 1 TO 20 IS

354085.	343570.	340669.	338923.	337734.
336724.	335795.	334919.	334091.	333320.
332613.	331958.	331324.	330688.	330033.
329350.	328642.	327922.	327184.	326432.

AT BEGINNING OF YEA 4,AND

AT TIME 4.75

DELAY IN PEAK OF FERTILITY CURVE IS -0.184397
INTEGRAL OF NORMALIZED FERTILITY IS 1.089502
FRACTION OF INITIAL FERTILITY IS 1.050882

VALUES OF FERTILITY CURVE FOR AGES 21 TO 30 ARE

0.059358	0.067649	0.075042	0.080616	0.082016
0.079207	0.075498	0.071418	0.065613	0.058225

TOTAL POPULATION IS 22016510.

****

POPULATION FOR AGES 1 TO 20 IS

361071.	346090.	341588.	339259.	337778.
336725.	335795.	334919.	334091.	333320.
332613.	331958.	331324.	330688.	330033.
329349.	328642.	327922.	327184.	326432.

AT BEGINNING OF YEA 5,AND

AT TIME 5.75

DELAY IN PEAK OF FERTILITY CURVE IS -0.480408
INTEGRAL OF NORMALIZED FERTILITY IS 1.089503
FRACTION OF INITIAL FERTILITY IS 1.098618

VALUES OF FERTILITY CURVE FOR AGES 21 TO 30 ARE

0.062964	0.072470	0.081761	0.089994	0.088098
0.080292	0.074753	0.071688	0.066822	0.059826

TOTAL POPULATION IS 22043640.

****

```
POPULATION FOR AGES  1 TO 20  IS
  377589.   352918.   344093.   340174.   338112.

  336768.   335795.   334919.   334091.   333320.

  332613.   331958.   331324.   330688.   330033.

  329350.   328641.   327922.   327184.   326432.

AT BEGINNING OF YEAR        6,AND
-----------------------------------
AT TIME  6.75

DELAY IN PEAK OF FERTILITY CURVE IS  -0.909306
   INTEGRAL OF NORMALIZED FERTILITY IS          1.089499
      FRACTION OF INITIAL FERTILITY IS                    1.118620

VALUES OF FERTILITY CURVE FOR AGES  21 TO 30  ARE
 0.065618  0.076765  0.088974  0.098260  0.092387

 0.076925  0.069761  0.068678  0.065408  0.059325


TOTAL POPULATION IS   22082090.

*********************************************************************************
POPULATION FOR AGES  1 TO 20  IS
  389333.   369063.   350881.   342668.   339024.

  337101.   335838.   334919.   334091.   333320.

  332613.   331958.   331324.   330688.   330033.

  329350.   328642.   327922.   327184.   326432.

AT BEGINNING OF YEAR        7,AND
-----------------------------------
AT TIME  7.75

DELAY IN PEAK OF FERTILITY CURVE IS  -1.277342
   INTEGRAL OF NORMALIZED FERTILITY IS          1.089500
      FRACTION OF INITIAL FERTILITY IS                    1.121859

VALUES OF FERTILITY CURVE FOR AGES  21 TO 30  ARE
 0.067241  0.079908  0.094899  0.105441  0.093917

 0.072075  0.064586  0.065381  0.063390  0.058125


TOTAL POPULATION IS   22123950.

*********************************************************************************
```

```
POPULATION FOR AGES  1 TO 20  IS
  393105.   380542.   366934.   349429.   341510.

  338010.   336170.   334962.   334091.   333320.

  332613.   331958.   331324.   330688.   330033.

  329350.   328642.   327922.   327184.   326431.

AT BEGINNING OF YEAR          8,AND
-----------------------------------
AT TIME  8.75

DELAY IN PEAK OF FERTILITY CURVE IS  -1.496449
   INTEGRAL OF NORMALIZED FERTILITY IS        1.089500
      FRACTION OF INITIAL FERTILITY IS                1.111391

VALUES OF FERTILITY CURVE FOR AGES  21 TO 30  ARE
 0.067541  0.081096  0.097754  0.108647  0.093259

 0.067985  0.060862  0.062787  0.061503  0.056761

TOTAL POPULATION IS   22164480.

*******************************************************************************
POPULATION FOR AGES  1 TO 20  IS
  392049.   384229.   378346.   365415.   348248.

  340489.   337077.   335293.   334134.   333320.

  332613.   331958.   331324.   330687.   330033.

  329350.   328642.   327922.   327184.   326431.

AT BEGINNING OF YEAR          9,AND
-----------------------------------
AT TIME  9.75

DELAY IN PEAK OF FERTILITY CURVE IS  -1.532525
   INTEGRAL OF NORMALIZED FERTILITY IS        1.089499
      FRACTION OF INITIAL FERTILITY IS                1.078118

VALUES OF FERTILITY CURVE FOR AGES  21 TO 30  ARE
 0.065753  0.079162  0.095780  0.106416  0.090456

 0.065073  0.058294  0.060436  0.059348  0.054862

TOTAL POPULATION IS   22197900.

*******************************************************************************
```

```
POPULATION FOR AGES  1 TO 2[illegible]  IS
  385079.    383196.    382112.    376780.    364179.

  347206.    339549.    336197.    334465.    333362.

  332613.    331958.    331324.    330688.    330033.

  329350.    328642.    327922.    327184.    326431.

AT BEGINNING OF YEA[illegible]       1[illegible],AND
---------------------------------

CORE USAGE        OBJECT CODE=     9856 BYTES,ARRAY AREA=     3400 BYTES,TOTAL AREA

DIAGNOSTICS          NUMBER OF ERRORS=        0, NUMBER OF WARNINGS=        0, NUMBE

COMPILE TIME=      3.98 SEC,EXECUTION TIME=     14.47 SEC,   QUEEN'S WATFOR VERSIO

 COST FOR THIS PROGRAM IS  $     1.42             RUN IN HSC     MAR  19,  1974
```

The above output simulates population and fertility over ten years with "Baby boom" dynamics - wide peak of intermediate height. See Figures 3 and 4 in Section V.

Program Ib

This is a copy of the program which was used to plot the profile in Figure 6.

```
$JOB   STF55952,'VER FR'
// EXEC SYMVU,TIME=
//COMP.SYSIN DD *
      SUBROUTINE DATA(X,IR,IC,IF)
      REAL D(100),F(100),P(100),X(100),R(100),US(100),D5(100),F1(100)
      REAL HH,KK,SUM,IV(10),OLY(10)
      REAL PLTPER,PRTPER,TI
      INTEGER YRS,AGES,YOUNG,NN
C     ******************************************************************
C     SET IFORM=0 IF YOU WANT PLOT,    =1 IF YOU WANT PRINT
C     ******************************************************************
      IFORM=1
      IFORM=0
      PL=1E-5
      PR=PL
      TINT=1E-4
      ICM1=IC-1
      IF (IR.GT.1) GO TO 31
C     RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
C     HH IS AGE STEP  -  KK IS TIME STEP
      READ(5,101) HH,KK
 101  FORMAT(2F10.6)
C     PLTPER IS PLOT INTERVAL  -  PRTPER IS PRINT INTERVAL
      READ(5,101) PLTPER,PRTPER
C     NDE IS THE NUMBER OF FIRST ORDER ODE'S AFTER TRANSFORMATION
C     IF NDE > 10 CHANGE REAL DIMENSION IN DATA AND $STEP
C     NN IS THE NUMBER OF INTEGRATION STEPS IN KK FOR SOLVING THE ODES
      READ(5,103) NDE, NN
 103  FORMAT(2I4)
C     PUT INITIAL VALUES OF ODES ONE TO A CARD
      DO 97 I=1,NDE
           READ(5,107) OLY(I)
 107       FORMAT(F10.4)
 97        CONTINUE
C     RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
      PR=PR+KK
      T=0.0
C     ******************************************************************
C                        READ DEATH RATES
C                        READ POPULATION DENSITY - OBTAINED FROM SPLINE
C     ******************************************************************
C     RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
      READ(5,100) (D(I),I=1,IC)
 100  FORMAT(5F10.6,30X)
      READ(5,104) (P(I),I=1,IC)
 104  FORMAT(10F8.0)
C     RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
      SUM=0.0
      DO 99 I=1,IC
           SUM=SUM+P(I)
 99   CONTINUE
      DO 98 I=1,IC
           P(I)=P(I)*22000000./SUM
      X(I)=P(I)
 98   CONTINUE
C     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
      IF(IFORM.NE.1) GOTO 61
      IPG=0
      WRITE(6,109)
 109  FORMAT('1','POPULATION FOR AGES  1  TO 20  IS')
      WRITE (6,102) (P(I),I=1,20)
 102  FORMAT(5F10.0/)
      WRITE(6,110)
 110  FORMAT(' ','AT BEGINNING OF YEAR            0, AND')
```

```
  111   FORMAT( [illegible]
   61   CONTINUE
C       PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
        DO 70 I=1,IC
             I1=FLOAT(I)
             D(I)=(1-EXP(-I1*I1/600.))/(1+((I1-25.5)**2)/4.)
   70        CONTINUE
        YOUNG=0
        F1(1)=EXP(-(.125*HH*(YOUNG-24.5))**2)
              SUM=F1(1)*P(1)
        DO 14 I=2,IC
             F1(I)=EXP(-(.125*HH*(I+YOUNG-25.5))**2)
             SUM=SUM+F1(I)*P(I)
             D5(I)=.5*(D(I)+D(I-1))
   14        CONTINUE
        D05=.5*(3.*D(1)-D(2))
        DT5=.5*(3.*D(IC)-D(IC-1))
        SUM=P(1)/SUM
        DO 15 I=1,IC
             F1(I)=F1(I)*SUM
   15        CONTINUE
        PA=KK/HH
        PA1=0.0
        P2=0.0
        FA=1
        GO TO 33
   31   CONTINUE
        PR=PR+KK
C       *********************************************************************
C                               UPDATE COEFFICIENTS OF FERTILITY EQUATION
C       *********************************************************************
        PA1=DLY(2)
        P2=DLY(2)*DLY(2)*PA
        FA=EXP(-DLY(4))
C       PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
        IF (PRTPER.GT.PR.OR.IFORM.NE.1) GOTO 62
        WRITE(6,106) T
  106   FORMAT(' AT TIME',F6.2)
        WRITE(6,112) DLY(1)
  112   FORMAT('0','DELAY IN PEAK OF FERTILITY CURVE IS  ',F9.6)
   62   CONTINUE
C       PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
C       *********************************************************************
C                               MULTIPLY FERTILITY BY INTEGRATING FACTOR
C       *********************************************************************
   33   CONTINUE
        SUM=0.0
        DO 42 I=1,IC
             SUM=SUM+F1(I)
             F(I)=FA*F1(I)
   42        CONTINUE
C       PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
        IF (PRTPER.GT.PR.OR.IFORM.NE.1) GOTO 63
        WRITE(6,113) SUM
  113   FORMAT(' ','   INTEGRAL OF NORMALIZED FERTILITY IS          ',F9.6)
        WRITE(6,114) FA
  114   FORMAT(' ','        FRACTION OF INITIAL FERTILITY IS
      2,F9.6)
        WRITE(6,115)
  115   FORMAT('0','VALUES OF FERTILITY CURVE FOR AGES  21 TO 30  ARE')
        WRITE(6,108) (F(I), I=21,30)
  108   FORMAT(5F10.5)
   63   CONTINUE
C       PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
C       *********************************************************************
```

```
C     ********************************************************************
      TINT=TINT+KK*?
      IF (HH.GT.TINT) GOTO 45
      SUM=0.
      SUM1=0.
      DO 44 I=1,IC-1
           J=IC-I
           SUM=SUM+P(J+1)*F(J+1)
           P(J+1)=P(J)*(1-R(J))
           SUM1=SUM1+P(J+1)
           X(J+1)=P(J+1)
 44        CONTINUE
      SUM=SUM+P(1)*F(1)
      P(1)=SUM
      X(1)=P(1)
      SUM1=SUM1+P(1)
      TINT=-1E-4
 45   CONTINUE
C     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
      IF (PRTPER.GT.PR.OR.IFORM.NE.1) GOTO 64
      WRITE(6,116) SUM1
 116  FORMAT('0','TOTAL POPULATION IS  ',F10.0)
      PR=+1E-5
      WRITE(6,117)
 117  FORMAT('0','*****************************************************
     I*************************')
 64   CONTINUE
C     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
C     ******************************************************************
C                              INCREMENT FERTILITY BY ONE YEAR
C     ******************************************************************
      DO 34 I=1,IC
           IF (1E-20.GT.F1(I)) F1(I)=0.0
           US(I)=D(I)*F1(I)
 34        CONTINUE
      UQ=2.*US(1)-US(2)
      UT=2.*US(IC)-US(IC-1)
      F1(IC)=F1(IC)-.5*PA*((PA1*(UT-US(IC-1)))-P2*(DT5*UT+D5(IC)*US
     2(IC-1)-(DT5+D5(IC))*US(IC)))
      F1(1)=F1(1)-.5*PA*((PA1*(US(2)-UQ))-P2*(D5(2)*US(2)+DQ5*UQ-(D5
     2(2)+DQ5)*US(1)))
      DO 35 I=2,IC-1
           F1(I)=F1(I)-.5*PA*((PA1*(US(I+1)-US(I-1)))-P2*(D5(I+1)*US
     2(I+1)+D5(I)*US(I-1)-(D5(I+1)+D5(I))*US(I)))
 35        CONTINUE
C     ******************************************************************
C                  SOLVE ODES FOR DELAY PARAMETERS AND INCREMENT TIME
C     ******************************************************************
      DO 36 I=1,NDE
           IV(I)=DLY(I)
 36        CONTINUE
      CALL $RUN(IV,DLY,T,KK,NDE,NN)
      PL=PL+KK
      IF (PLTPER.GT.PL) GOTO 31
C     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
      IF (IFORM.NE.1) GOTO 65
      IPG=IPG+1
      IF (IPG.NE.2) GOTO 79
      IPG=0
      WRITE(6,121)
 121  FORMAT('1')
 79   CONTINUE
      WRITE(6,118)
 118  FORMAT(' ','POPULATION FOR AGES   1 TO 20   IS')
           WRITE(6,102) (P(I), I=1,20)
```

```
119   FORMAT(' ','AT BEGINNING OF YEAR ',I5,' AND')
      WRITE(6,120)
120   FORMAT(' ','-----------------------------------')
65    CONTINUE
C     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
      RETURN
      END
      SUBROUTINE $RUN(IV,Z,T,KK,NDE,NN)
C     *********************************************************
C     THIS SUBROUTINE GIVES A SIMULTANEOUS SIMULATION OF VALUES
C     Z=(IA(T),A(T),A'(T),IB(T),B(T)) AT  T+H  FROM INITIAL VALUES  IV
C     OF THE SAME VARIABLES AT T.  FOR EACH DIFFERENT SIMULATION THE
C     SUBROUTINE  $DE  MUST BE REWRITTEN TO EVALUATE THE DERIVATIVES
C     D(I)  AS A FUNCTION OF TIME  T  AND SOLUTION VALUES  Y(I)
C     *********************************************************
      REAL IV(NDE),Z(NDE),T,T2,KK,K1
      K1=KK/FLOAT(NN)
      T2=T
      T=T+KK
28    CONTINUE
      CALL $STEP(IV,Z,T2,K1,NDE)
      T2=T2+K1
      DO 29 I=1,NDE
           IV(I)=Z(I)
29         CONTINUE
      IF (T.GT.T2+K1/2) GOTO 28
      RETURN
      END
      SUBROUTINE $STEP(IV,Z,T,KK,NDE)
      REAL IV(NDE),Z(NDE),Z1(10),D(10),D1(10),KK
      CALL $DE(T,IV,NDE,D)
      DO 11 I=1,NDE
           Z1(I)=IV(I)+KK*D(I)
11         CONTINUE
      TT=T+KK
      CALL $DE(TT,Z1,NDE,D1)
      DO 12 I=1,NDE
           Z(I)=IV(I)+.5*KK*(D(I)+D1(I))
12         CONTINUE
      RETURN
      END
      SUBROUTINE $DE(T,Y,NDE,D)
      REAL Y(NDE),D(NDE),KK
      D(1)=Y(2)
      D(2)=Y(3)
      T1=0.
      IF (T.GT.15.) GOTO 78
      TT1=ABS(T-5.)
      TT2=ABS(T-10.)
      T1=(-SIN(EXP(-TT1)-EXP(-TT2)))/15.
78    CONTINUE
      D(3)=15.*T1-Y(2)-2.*Y(3)
      D(4)=T1
      D(5)=0.
      RETURN
      END
//GO.SYSIN DD *
EVOLUTION OF POPULATION OVER 50 YEARS:  WITH FERTILITY PROGRESSION
  50 100    2   4         0                                   1         3
 30.0330.0  7.0          4.0                        20.0
1.0         .25
1.0          1.0
    5    1
0.0
0.0
```

```
0.0
0.1
.02258     .00577     .00414     .00338     .00299
.00276     .00261     .00247     .00231     .00212
.00197     .00191     .00192     .00198     .00207
.00215     .00219     .00225     .00230     .00237
.00243     .00251     .00259     .00268     .00277
.00288     .00299     .00311     .00325     .0034
.00356     .00373     .00392     .00412     .00435
.00459     .00486     .00515     .00546     .00581
.00618     .00659     .00703     .00751     .00804
.00861     .00923     .00991     .01064     .01145
.01232     .01327     .01430     .01543     .01665
.01798     .01943     .021       .02271     .02457
.02659     .02878     .03118     .03376     .03658
.03964     .04296     .04656     .05046     .0547
.0593      .06427     .06966     .0755      .08181
.08864     .09602     .10399     .11259     .12186
.13185     .1426      .15416     .16657     .17988
.19413     .20937     .22563     .243       .26144
.28099     .30173     .32364     .34666     .371
.39621     .44719     .54826     .72467     1.0
1023102.1000000. 994230.990114. 986767. 983817. 981102. 978541. 976124. 973869.
971804. 969890. 968038. 966179. 964266. 962270. 960201. 958098. 955942. 953743.
951483. 949171. 945789. 944337. 941806. 939197. 936492. 933692. 930788. 927763.
924609. 921317. 917980. 914282. 910515. 906554. 902393. 898007. 893382. 888504.
883342. 877883. 872[illegible]98. 865967. 859464. 852554. 845214. 837413. 829114. 820292.
810900. 800910. 790282. 778981. 766961. 754191. 740631. 726241. 710990. 694843.
677771. 659749. 640761. 620782. 599824. 577882. 554975. 531133. 506403. 480850.
454548. 427593. 400112. 372240. 344136. 315982. 287973. 260322. 233251. 206989.
181765. 157799. 135297. 114440. 95378.  78221.  63036.  49838.  38593.  29215.
21577.  15514.  10833.  7327.   4787.   3011.   1818.   1005.   454.    125.
/*
//


******** JOB DELETED BY HASP OR CANCELLED BY OPERATOR BEFORE EXECUTION ********

 HASP-II JOB STATISTICS --    296 CARDS READ --        0 LINES PRINTED --      0
```

Estimation of $\mu(x,s,t)$: this program uses values of the population density to estimate $\mu(x,s,t)$ through differentiation of two cubic spline approximations and integration of a subsequent spline approximation.

As a test problem, a separable economic mobility was chosen

$$\mu(x,s,t) = \alpha(x)\,\beta(s) \quad ,$$

$$\alpha(x) = \frac{(x+20)(80-x)}{10000} \quad , \quad \beta(s) = \frac{s+10}{200} \quad .$$

With an initial population density

$$p(x,s,o) = p_0(x,s) \equiv e^{-\frac{x}{200}} \frac{(2s^3-405s^2+21000s)}{540}$$

the population density without deaths

$$\bar{p}(x,s,t) = [e^{\frac{w}{2000000}} p_0(x-t, (s+10)e^{\frac{w}{2000000}}-10)] \quad ,$$

$$w = -\frac{t^3}{3} + t^2(30-x) + t(x-80)(x+20) \quad ,$$

evolves, and with a death-rate of

$$r(x,s,t) = (1-.01t)\,(.0003x+.0006)$$

the density with deaths is

$$p(x,s,t) = \bar{p}(x,s,t)e^{-R} \quad ,$$

$$R = \left(t - \frac{.01t^2}{2}\right)(.0003x+.0006) - .0003\left(\frac{t^2}{2} - \frac{.01t^3}{6}\right).$$

```
$JOB    ACCT-NUM,'VERNER',TIME=60,PAGES=20
C
C       **************************************************************************
C       THIS PROGRAM IS USED TO ESTIMATE THE ECONOMIC MOBILIITY
C       PARAMETER MU.  AGE- TIME- AND INCOME-SPECIFIC VALUES OF THE
C       POPULATION ARE REQUIRED.
C       FOR A TEST RUN WE ASSUME THAT  4D'7#S,T)=ALPHA(X)*BETA(S).
C       FOR THIS PARTICULAR EXAMPLE THE EIGENVECTOR TECHNIQUE USED
C       TO ESTIMATE  A(T) AND D(X)  IN THE FERTILITY EQUATION COULD
C       HAVE BEEN USED.
C       **************************************************************************
C
        REAL P(7,11,11),MU(7,11,11),ALPHA(7),BETA(11),X,S,T
        REAL H(11),A(3,11),COF(4,11)
        REAL K(2,11),D(11),TOT,W,W1,Y
        INTEGER DX,DS,DT,XE,SM,TM,XX,SS,TT
        XE=7
        SM=11
        TM=11
        DX=10
        DS=10
        DT=2.
        E=2.71828182846
C
C       **************************************************************************
C       DEFINE FUNCTION ALPHA
C       **************************************************************************
C
        WRITE(6,112)
  112   FORMAT('1','  THE ECONOMIC MOBILITY IS  ALPHA(X)*BETA(S) WHERE')
        WRITE(6,106)
  106   FORMAT('0')
        DO 10 IX=1,XE
        XX=DX*(IX-1)
        ALPHA(IX)=(XX+20.)*(80.-XX)/10.**4.
C
C       PRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRIN
        WRITE(6,100) XX,ALPHA(IX)
  100   FORMAT('        ALPHA(',I3,') IS  ',F10.5)
   10   CONTINUE
        WRITE(6,106)
C       PRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRIN
C
C
C       **************************************************************************
C       DEFINE FUNCTION BETA
C       **************************************************************************
C
        DO 11 IS=1,SM
        SS=DS*(IS-1)
        BETA(IS)=(SS+10.)/200.
C
C       PRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRIN
        WRITE(6,101) SS,BETA(IS)
  101   FORMAT('        BETA(',I3,') IS  ',F10.5)
   11   CONTINUE
C       PRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRIN
C
C       **************************************************************************
C       FIND ANALYTIC SOLUTION TO THE PDE
```

```
C             DP/DX + DP/DT = D/DS(MU*P) - R*P
C        ************************************************************
C
         DO 12 LT=1,TM
         T=DT*(LT-1)
         DO 12 LX=1,XE
         X=DX*(LX-1)
         W=(.0003*X +.0006)*(T-.005*T*T)-.0003*(T*T/2-(.005*T**3)/3)
         W1=E**(-W)
         W2=T*((X+20.)*(X-80.)+T*((30.-X)+T/3.))*(-1.)
         DO 12 LS=1,SM
         SS=DS*(LS-1)
         S=SS+10.
         W=S*E**(W2/2000000.)-10.
         Y=W*(21000.+W*((-405.)+2.*W))
         W=(W+10.)/S
         P(LX,LS,LT)=(W*Y/540.)*E**((T-X)/200.)*W1
   12    CONTINUE
C
C        ************************************************************
C            SPLINE  P(X,S,T)  AGAINST  X  AND DIFFERENTIATE THE SPLINE
C               MU(X,S,T)  IS  DP/DX
C        ************************************************************
C
         DO 20 IX=1,XE
         X=DX*(IX-1)
         K(1,IX)=X
   20    CONTINUE
         CALL SETUP(XE,A,H,K)
         DO 21 IS=1,SM
         DO 21 IT=1,TM
         DO 22 IX=1,XE
         K(2,IX)=P(IX,IS,IT)
  22     CONTINUE
         CALL SOLVE(A,K,H,XE,D,COF)
         DO 21 IX=1,XE
         X=DX*(IX-1)
         MU(IX,IS,IT)=COF(3,IX)+X*(2.*COF(2,IX)+X*3.*COF(1,IX))
   21    CONTINUE
C
C        ************************************************************
C            SPLINE  P(X,S,T)  AGAINST  T  AND DIFFERENTIATE THE SPLINE
C              MU(X,S,T)  IS  DP/DX+DP/DT+R(X,S,T)*P(X,S,T)
C        ************************************************************
C
         DO 30 IT=1,TM
         T=DT*(IT-1)
         K(1,IT)=T
   30    CONTINUE
         CALL SETUP(TM,A,H,K)
         DO 31 IX=1,XE
         X=DX*(IX-1)
         DO 31 IS=1,SM
         DO 32 IT=1,TM
         K(2,IT)=P(IX,IS,IT)
   32    CONTINUE
         CALL SOLVE(A,K,H,TM,D,COF)
         DO 31 IT=1,TM
         T=DT*(IT-1)
         MU(IX,IS,IT)=MU(IX,IS,IT)+COF(3,IT)+T*(2.*COF(2,IT)+
```

```
     &T*3.*COF(1,IT))+(1.-.01*T)*(.0003*X+.0006)*P(IX,IS,IT)
   31 CONTINUE
C
C     ***************************************************************
C        SPLINE  DP/DX+DP/DT+R(X,S,T)*P(X,S,T)   AGAINST  S  AND
C        INTEGRATE THE SPLINE FROM  S  TO THE MAXIMUM VALUE OF  S.
C                 MU(X,S,T) IS  END CONDITION PLUS THIS INTEGRAL
C     ***************************************************************
C
      DO 40  IS=1,SM
      S=DS*(IS-1)
      K(1,IS)=S
   40 CONTINUE
      CALL SETUP(SM,A,H,K)
      DO 41 IX=1,XE
      DO 41 IT=1,TM
      DO 42 IS=1,SM
      K(2,IS)=MU(IX,IS,IT)
   42 CONTINUE
      CALL SOLVE(A,K,H,SM,D,COF)
      MU(IX,SM,IT)=0.
      S2=DS*SM
      S1=S2-DS
      DO 41 IS=2,SM
      S1=S1-DS
      S2=S2-DS
      I=SM+1-IS
            MU(IX,I,IT)=MU(IX,I+1,IT)+(((S2*COF(1,I)/4.+COF(2,I)/3.)*
     &      S2+COF(3,I)/2.)*S2+COF(4,I))*S2-(((S1*COF(1,I)/4.
     &      +COF(2,I)/3.)*S1+COF(3,I)/2.)*S1+COF(4,I))*S1
   41 CONTINUE
C
C     PRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRIN
      WRITE(6,110)
  110 FORMAT('1',' VALUES ARE % ERROR IN MU, S AND T GIVEN, X ACROSS')
      WRITE(6,114)
  114 FORMAT('0','      S       T    X=     0         10         20         30
     & 40         50         60')
C     PRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRIN
C
      DO 55 KS=1,SM
      WRITE(6,106)
      SS=DS*(KS-1)
      DO 55 KT=1,TM
      TT=DT*(KT-1)
      DO 50 KX=1,XE
      IF (P(KX,KS,KT).EQ.0.) GOTO 51
      MU(KX,KS,KT)=(BETA(SM)*ALPHA(KX)*P(KX,SM,KT)-MU(KX,KS,KT))/
     &P(KX,KS,KT)
      GOTO 52
   51    MU(KX,KS,KT)=ALPHA(KX)*BETA(KS)
   52 CONTINUE
      MU(KX,KS,KT)=(MU(KX,KS,KT)/(BETA(KS)*ALPHA(KX))-1)*100.
   50 CONTINUE
C
C     PRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRIN
      WRITE(6,113) SS,TT,(MU(KX,KS,KT), KX=1,XE)
  113 FORMAT('  ',2F5.1,'  ',7F9.2)
   55 CONTINUE
C     PRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRIN
```

```
      WRITE(6,105)
  105 FORMAT('1')
      STOP
      END
C
      SUBROUTINE SETUP(N,A,H,K)
C
C     ****************************************************************
C     THIS ROUTINE SETS UP A TRIDIAGONAL MATRIX OF THE SPLINE EQUATION
C     IN A 3 BY N ARRAY  -  FIRST ROW IS SUPERDIAGONAL
C                           SECOND ROW IS DIAGONAL
C                           THIRD ROW IS SUBDIAGONAL
C     AND THEN DECOMPOSES A TO  LU  SO THAT
C        FIRST ROW IS SUPERDIAGONAL OF U, SECOND ROW IS DIAGONAL OF U
C       THIRD ROW IS SUBDIAGONAL OF L, DIAGONAL OF L IS UNITY
C     ****************************************************************
C
      REAL A(3,N),H(N),K(2,N)
      NM1=N-1
      H(2)=K(1,2)-K(1,1)
      DO 10  I=2,NM1
         H(I+1)=K(1,I+1)-K(1,I)
         A(1,I)=H(I+1)/(H(I+1)+H(I))
         A(2,I)=2.0
         A(3,I)=1-A(1,I)
   10    CONTINUE
      A(1,1)=-2.
      A(2,1)=2.
      A(3,1)=0.
      A(1,N)=0.
      A(2,N)=2.
      A(3,N)=-2.
      DO 11 I=2,N
         A(3,I)=A(3,I)/A(2,I-1)
         A(2,I)=A(2,I)-A(3,I)*A(1,I-1)
   11    CONTINUE
      RETURN
      END
C
      SUBROUTINE SOLVE(A,K,H,N,D,COF)
C
C     ****************************************************************
C     THIS ROUTINE CALCULATES THE SECOND ORDER FINITE DIFFERENCES OF
C     THE SPLINE, AND THEN SOLVES  AM=D,  (BY FORWARD AND BACKWARD
C     SUBSTITUTION), PLACING M(VECTOR OF SECOND DERIVATIVES) IN D
C     END POINT CONDITIONS NOW USE THIRD ORDER FINITE DIFFERENCES
C     TO ESTIMATE THE THIRD ORDER DERIVATIVES AT  X0+H/2  AND XN-H/2.
C     ****************************************************************
C
      REAL H(N),D(N),A(3,N),K(2,N),COF(4,N)
      D(2)=(K(2,2)-K(2,1))/H(2)
      NM1=N-1
      DO 12 I=2,NM1
         D(I+1)=(K(2,I+1)-K(2,I))/H(I+1)
         D(I)=6*(D(I+1)-D(I))/(H(I+1)+H(I))
   12    CONTINUE
      D(1)  =-2.*(-K(2,1)+K(2,4)+3*(K(2,2)-K(2,3)))/(H(2)*H(2))
      D(N)=-2.*(-K(2,N)+K(2,N-3)+3.*(K(2,N-1)-K(2,N-2)))/(H(N-1)**2)
```

```
      DO 13 I=2,N
         D(I)=D(I)-A(3,I)*D(I-1)
 13      CONTINUE
      D(N)=D(N)/A(2,N)
      DO 14 I=2,N
         J=N+1-I
         D(J)=(D(J)-A(1,J)*D(J+1))/A(2,J)
 14      CONTINUE
      CALL POLLY(N,D,K,H,COF)
      RETURN
      END
C

      SUBROUTINE POLLY(N,M,K,H,COF)
C
C     **********************************************************************
C     THIS ROUTINE COMPUTES THE COEFFICIENTS OF THE SPLINE POLYNOMIAL
C       ON EACH SUBINTERVAL
C     K IS THE ARRAY OF DATA POINTS
C     H IS THE VECTOR OF SUBINTERVAL LENGTHS
C     M IS THE SOLUTION VECTOR TO THE EQUATION AM=D
C     **********************************************************************
C
      REAL M(N),K(2,N),H(N),COF(4,N)
      NM1=N-1
      DO 11 I=1,NM1
      COF(1,I)=(M(I+1)-M(I))/(6.*H(I+1))
      COF(2,I)=(K(1,I+1)*M(I)-K(1,I)*M(I+1))/(2.*H(I+1))
      DD=M(I+1)*K(1,I)*K(1,I)-M(I)*K(1,I+1)*K(1,I+1)
     &+2.*K(2,I+1)-2.*K(2,I)
      COF(3,I)=(DD/(2.*H(I+1)))+H(I+1)*(M(I)-M(I+1))/6.
      DD=M(I)*(K(1,I+1)**3)-M(I+1)*(K(1,I)**3)+6.*K(1,I+1)*K(2,I)
     &-6.*K(1,I)*K(2,I+1)+K(1,I)*M(I+1)*(H(I+1)**2)-K(1,I+1)*M(I)
     &*(H(I+1)**2)
      COF(4,I)=DD/(H(I+1)*6.)
 11   CONTINUE
      DO 16 J=1,4
 16   COF(J,N)=COF(J,NM1)
      RETURN
      END

$ENTRY
```

THE ECONOMIC MOBILITY IS ALPHA(X)*BETA(S) WHERE

ALPHA(0)	IS	0.15000
ALPHA(10)	IS	0.21000
ALPHA(20)	IS	0.24000
ALPHA(30)	IS	0.25000
ALPHA(40)	IS	0.24000
ALPHA(50)	IS	0.21000
ALPHA(60)	IS	0.15000

BETA(0)	IS	0.05000
BETA(10)	IS	0.10000
BETA(20)	IS	0.15000
BETA(30)	IS	0.20000
BETA(40)	IS	0.25000
BETA(50)	IS	0.30000
BETA(60)	IS	0.35000
BETA(70)	IS	0.40000
BETA(80)	IS	0.45000
BETA(90)	IS	0.50000
BETA(100)	IS	0.55000

VALUES ARE % ERROR IN MU. S AND T GIVEN. X ACROSS

S	T	X = 0	10	20	30	40	50	60
0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	2	403.75	359.65	75.85	115.51	170.63	160.18	803.37
0	4	-41.60	18.10	-20.49	27.85	-27.13	-15.82	83.30
0	6	-203.99	-7.42	2.34	18.17	13.93	33.79	99.21
0	8	-133.41	-1.23	16.22	-3.06	6.51	-2.07	100.96
0	10	-131.42	25.24	10.08	-3.89	-11.09	-1.89	107.53
0	12	-125.68	66.76	-4.97	0.09	17.15	-8.16	95.22
0	14	-85.40	18.42	-1.60	5.88	8.73	-15.13	105.93
0	16	-95.43	18.09	6.82	7.75	10.38	-11.17	103.78
0	18	-237.38	27.41	-10.80	-3.91	10.89	-40.15	86.25
0	20	-343.99	18.83	-38.13	17.28	1.92	16.66	283.49
10	0	-0.70	-0.86	-0.28	-0.44	-0.51	-1.00	-1.71
10	2	0.36	0.43	0.08	0.17	0.22	0.15	0.75
10	4	-0.07	0.07	-0.03	0.07	-0.07	-0.05	0.23
10	6	-0.54	-0.02	-0.01	0.05	0.06	0.11	0.30
10	8	-0.44	0.00	0.06	0.01	0.05	-0.03	0.49
10	10	-0.53	0.15	0.06	0.00	-0.06	-0.04	0.74
10	12	-0.65	0.43	-0.03	-0.03	0.13	-0.06	0.75
10	14	-0.51	0.16	-0.01	0.07	0.07	-0.15	0.95
10	16	-0.61	0.18	0.04	0.11	0.13	-0.18	1.10
10	18	-1.37	0.28	-0.13	-0.06	0.17	-0.53	1.15
10	20	-1.89	0.24	-0.37	0.30	-0.02	0.29	3.78
20	0	-0.34	-0.37	-0.15	-0.22	-0.25	-0.45	-0.71
20	2	0.13	0.15	0.02	0.05	0.07	0.04	0.27
20	4	-0.03	0.04	-0.01	0.03	-0.02	-0.00	0.13
20	6	-0.26	-0.01	-0.01	0.01	0.03	0.03	0.13
20	8	-0.19	0.01	0.02	0.01	0.01	-0.02	0.24
20	10	-0.22	0.07	0.02	0.01	-0.02	-0.02	0.34
20	12	-0.33	0.15	-0.02	-0.02	0.05	-0.02	0.33
20	14	-0.28	0.09	-0.01	0.02	0.02	-0.06	0.41
20	16	-0.32	0.11	-0.00	0.05	0.07	-0.11	0.51
20	18	-0.64	0.13	-0.06	-0.02	0.08	-0.23	0.62
20	20	-0.79	0.14	-0.08	0.14	-0.06	0.12	1.60
30	0	-0.24	-0.25	-0.09	-0.14	-0.18	-0.29	-0.43
30	2	0.05	0.07	-0.01	0.02	0.04	0.01	0.15
30	4	-0.02	0.03	-0.01	0.02	-0.00	0.01	0.11
30	6	-0.17	0.00	-0.01	0.00	0.01	0.01	0.08
30	8	-0.12	0.01	0.01	0.02	0.00	-0.02	0.16
30	10	-0.12	0.05	0.02	0.01	-0.01	-0.02	0.24
30	12	-0.25	0.09	-0.02	-0.02	0.03	-0.01	0.22
30	14	-0.22	0.06	-0.00	0.01	0.00	-0.04	0.26
30	16	-0.24	0.09	-0.01	0.03	0.05	-0.09	0.36
30	18	-0.45	0.10	-0.04	-0.01	0.07	-0.14	0.46
30	20	-0.54	0.08	0.01	0.09	-0.09	0.06	1.03
40	0	-0.16	-0.20	-0.06	-0.11	-0.14	-0.22	-0.30
40	2	0.00	0.05	-0.01	0.00	0.03	-0.00	0.10
40	4	-0.01	0.02	-0.01	0.02	0.01	0.00	0.11

40	6	-0.13	0.01	-0.01	0.00	0.00	-0.00	0.06
40	8	-0.08	0.01	0.00	0.01	-0.00	-0.02	0.11
40	10	-0.08	0.03	0.01	0.02	-0.01	-0.02	0.21
40	12	-0.22	0.07	-0.02	-0.02	0.02	-0.00	0.15
40	14	-0.19	0.04	0.00	0.01	-0.01	-0.03	0.20
40	16	-0.22	0.08	-0.01	0.03	0.04	-0.08	0.30
40	18	-0.37	0.09	-0.04	0.00	0.07	-0.11	0.39
40	20	-0.46	0.05	0.06	0.07	-0.11	0.03	0.79
50	0	-0.14	-0.17	-0.05	-0.11	-0.09	-0.19	-0.23
50	2	-0.02	0.04	-0.02	0.01	0.01	-0.01	0.07
50	4	0.01	0.02	-0.01	0.01	0.00	0.01	0.10
50	6	-0.11	0.00	-0.01	-0.01	0.00	-0.01	0.05
50	8	-0.06	0.01	-0.00	0.01	-0.01	-0.03	0.10
50	10	-0.05	0.02	0.01	0.02	-0.00	-0.01	0.18
50	12	-0.22	0.05	-0.01	-0.02	0.02	0.01	0.13
50	14	-0.18	0.05	-0.00	0.00	-0.01	-0.02	0.16
50	16	-0.22	0.06	-0.02	0.02	0.04	-0.08	0.26
50	18	-0.35	0.08	-0.04	0.01	0.05	-0.09	0.37
50	20	-0.39	0.05	0.07	0.06	-0.11	0.03	0.62
60	0	-0.12	-0.11	-0.03	-0.11	-0.05	-0.16	-0.17
60	2	-0.02	0.02	-0.03	0.01	-0.01	-0.02	0.05
60	4	0.01	0.02	-0.01	0.01	-0.00	0.03	0.08
60	6	-0.09	0.00	-0.01	-0.01	0.01	-0.01	0.03
60	8	-0.05	0.01	-0.01	0.02	-0.01	-0.04	0.09
60	10	-0.05	0.02	0.02	0.03	-0.00	-0.02	0.16
60	12	-0.24	0.04	0.00	-0.01	0.01	0.01	0.12
60	14	-0.19	0.05	-0.02	-0.00	-0.01	0.00	0.14
60	16	-0.23	0.08	-0.02	0.02	0.04	-0.07	0.23
60	18	-0.37	0.08	-0.02	0.01	0.04	-0.09	0.36
60	20	-0.35	0.05	0.07	0.08	-0.10	0.05	0.51
70	0	-0.10	-0.09	-0.03	-0.07	-0.02	-0.14	-0.12
70	2	-0.01	0.02	-0.01	0.01	-0.01	-0.00	0.04
70	4	-0.01	0.01	-0.01	-0.00	-0.00	0.02	0.05
70	6	-0.07	0.00	-0.01	-0.00	0.01	-0.02	0.04
70	8	-0.05	0.01	-0.01	0.01	-0.02	-0.03	0.07
70	10	-0.04	0.03	-0.00	0.02	0.00	-0.02	0.12
70	12	-0.21	0.04	0.00	0.00	0.01	0.00	0.13
70	14	-0.21	0.04	-0.01	-0.01	-0.00	-0.01	0.15
70	16	-0.27	0.08	-0.03	0.02	0.03	-0.07	0.22
70	18	-0.37	0.10	-0.02	0.01	0.03	-0.08	0.33
70	20	-0.34	0.08	0.05	0.05	-0.06	0.03	0.48
80	0	-0.10	-0.09	-0.01	-0.04	-0.00	-0.12	-0.08
80	2	-0.00	0.03	0.00	0.01	-0.02	0.01	0.04
80	4	-0.02	0.01	-0.01	-0.00	0.01	0.01	0.02
80	6	-0.04	0.00	-0.01	-0.00	0.00	-0.03	0.05
80	8	-0.06	0.01	0.00	-0.00	-0.02	-0.01	0.06
80	10	-0.13	0.02	-0.02	0.01	0.01	-0.01	0.09
80	12	-0.19	0.05	-0.00	0.00	0.01	-0.01	0.14
80	14	-0.22	0.04	-0.00	-0.01	0.01	-0.02	0.15
80	16	-0.31	0.08	-0.03	0.03	0.02	-0.05	0.19
80	18	-0.36	0.11	-0.02	0.01	0.00	-0.07	0.27
80	20	-0.37	0.09	0.01	-0.02	-0.01	-0.01	0.44

90	0	-0.05	-0.05	0.00	-0.02	-0.01	-0.07	-0.06
90	2	-0.09	0.02	-0.00	-0.00	-0.01	-0.01	0.01
90	4	-0.01	0.01	-0.01	-0.00	0.00	0.02	0.02
90	6	-0.03	0.01	-0.01	0.00	0.00	-0.02	0.04
90	8	-0.03	0.01	-0.01	-0.01	-0.02	-0.02	0.02
90	10	-0.10	0.01	0.00	0.01	0.01	0.00	0.03
90	12	-0.17	0.05	0.00	0.00	-0.00	-0.01	0.09
90	14	-0.14	0.04	-0.01	-0.01	0.00	-0.01	0.10
90	16	-0.25	0.06	-0.02	0.02	0.02	-0.03	0.12
90	18	-0.31	0.09	-0.02	0.01	-0.01	-0.05	0.18
90	20	-0.35	0.03	-0.01	-0.01	-0.01	-0.01	0.23
100	0	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
100	2	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
100	4	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
100	6	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
100	8	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
100	10	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
100	12	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
100	14	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
100	16	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
100	18	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
100	20	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00

```
CORE USAGE        OBJECT CODE=   10216 BYTES,ARRAY AREA=    7332 BYTES,TOTAL AREA
DIAGNOSTICS       NUMBER OF ERRORS=        0, NUMBER OF WARNINGS=        0, NUMBE
COMPILE TIME=     4.78 SEC,EXECUTION TIME=    41.80 SEC,   QUEEN'S WATFOR VERSIO

 COST FOR THIS PROGRAM IS  $     3.21            RUN IN HSC     MAR  26,  1974
```

Program III

Estimation of the parameters $a(t)$ and $d(x)$ from values of the normalized fertility. The normalized fertility might be estimated from the fertility curve and the parameter $b(t)$ which represents the family size.

Here we chose as a test case:

$$d(x) = \frac{x}{100} \qquad x = 0,10,\ldots,50$$

$$a(t) = 1+.4\sin(\frac{t}{6}) \qquad t = 0,2,\ldots,20$$

```
$JOB    ACCT-NUM,'VERNER',TIME=60
C
C       ****************************************************************
C       THIS PROGRAM ESTIMATES THE PARAMETERS  A(T)  AND  D(X) FOR THE
C       FERTILITY DIFFERENTIAL EQUATION.  DATA IN THE FORM OF VALUES
C       OF THE NORMALIZED FERTILITY ARE REQUIRED, AND THESE MAY BE
C       ESTIMATED FROM VALUES OF THE FERTILITY CURVE AND THE PARAMETER
C       B(T) - THE FAMILY SIZE.
C       TO TEST THE PROGRAM A CLOSED FORM SOLUTION FOR A SPECIAL CASE
C       IS USED.
C       ****************************************************************
C
        COMMON K,A,H,COF,D
        REAL FBAR(51,11),FBARST(51,11),K(2,51),G(51,11),GT(11,51),ID(51)
        REAL AT(11),D(51),PROD(11,11)
        REAL A(3,51),H(51),COF(4,51)
        REAL NU(51)
C       ****************************************************************
C       OBSERVE THAT DYNAMIC PARAMETERS ARE BEING USED WITH COMMON AND
C       AN ERROR MAY OCCUR AS A RESULT OF MIXED INDEXING - INSURE THAT
C       COLUMNS ARE COMPLETELY FILLED ON USE OF A DYNAMIC INDEX
C       ****************************************************************
        NT=11
        NX=51
        CALL ANAL(NX,NT,FBAR,AT,D)
        D2=D(2)
        AT2=AT(2)
        CALL OBS(NX,NT,FBAR,FBARST,G)
        CALL TRANS(NX,NT,G,GT)
        CALL MULT(11,51,GT,G,PROD)
        CALL EIGEN(PROD,11,AT,ID)
C       ****************************************************************
C       TO EXHIBIT ERRORS IN THIS APPROACH WE CALCULATE A(1) EXACTLY
C       AND MULTIPLY THE OTHER COMPONENTS BY THE APPROPRIATE FACTOR.
C       ****************************************************************
        AT2=AT2/AT(2)
        DO 82 IT=1,NT
           AT(IT)=AT(IT)*AT2
   82      CONTINUE
        DO 85 IX=1,NX
           SUM=AT(1)*G(IX,1)
           SUM1=AT(1)*AT(1)
           DO 84 IT=2,NT
              SUM=SUM+AT(IT)*G(IX,IT)
              SUM1=SUM1+AT(IT)*AT(IT)
   84         CONTINUE
           D(IX)=SUM/SUM1
   85      CONTINUE
C
C       PRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRIN
        WRITE(6,114)
  114   FORMAT('0',' A(T)  OBTAINED FROM  11 BY 11  MATRIX IS')
        WRITE(6,104) (AT(I),I=1,11)
  104   FORMAT('0',5F12.6)
        WRITE(6,115)
  115   FORMAT('0','  D(X)  OBTAINED AS  (G(X,T),A(T))/(A(T),A(T))')
        WRITE(6,104) (D(I),I=1,51)
C       PRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRIN
C
        STOP
```

```
      END
C
C
      SUBROUTINE ANAL(NX,NT,FBAR,AT,D)
      REAL FBAR(NX,NT),AT(NT),D(NX)
C
C     ******************************************************************
C     SUPPOSE  D(X)=X/100
C     SUPPOSE  A(T)=1.+.4SIN(T/6.)
C     SUPPOSE FOBAR(X)=(1-COS(X/8))*X/100
C     THEN THIS SUBROUTINE COMPUTES THE CORRESPONDING ANALYTIC SOLUTION
C     TO THE FERTILITY PDE
C          D(FBAR)/DT = -A(T)*D(D(X)*FBAR)/DX
C     IF  A(T) < 0  THERE IS AN ADVANCE IN THE FERTILITY
C     IF  A(T) > 0  THERE IS A DELAY IN THE FERTILITY
C     ******************************************************************
C
      DT=2.
      DO 8  IT=1,NT
           T=DT*(IT-1)
           AT(IT)=1+.4*SIN(T/6)
           W=(2.4+T-2.4*COS(T/6.))/400.
           W=EXP(-W)
           DO 8  IX=1,NX
                X=FLOAT(IX-1)
                D(IX)=X/400.
                FBAR(IX,IT)=.01*X*W*W*(1-COS(X*W/8.))
    8           CONTINUE
C
C     PRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRIN
      WRITE(6,116)
  116 FORMAT('1',' A(T)  FROM THE CLOSED FORM IS')
      WRITE(6,104) (AT(I),I=1,11)
  104 FORMAT('0',5F12.6)
      WRITE(6,117)
  117 FORMAT('0',' D(X)  FROM THE CLOSED FORM IS')
      WRITE(6,104) (D(I),I=1,51)
      WRITE(6,105)
  105 FORMAT('0')
C     PRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRIN
C
      RETURN
      END
C
      SUBROUTINE OBS(NX,NT,FBAR,FBARST,G)
C     ******************************************************************
C     THIS ROUTINE COMPUTES THE 'OBSERVATIONS' OF THE FORM
C     INTEGRAL FROM 0 TO X OF -D/DT(FBAR(X,T)) ALL DIVIDED BY FBAR(X,T)
C      THE ARRAY G STORES THE OBSERVATIONS
C     ******************************************************************
      COMMON K,A,H,COF,D
      REAL K(2,51),A(3,51),H(51),COF(4,51),D(51)
      REAL FBAR(NX,NT),FBARST(NX,NT),G(NX,NT)
      DT=2.
      DO 5  I=1,NT
    5      K(1,I)=DT*(I-1.)
      CALL SETUP(NT,A,H,K)
      DO 25  IX=1,NX
```

```
                DO 15 IT=1,NT
   15               K(2,IT)=FBAR(IX,IT)
                CALL SOLVE(A,K,H,NT,D,COF)
                DO 24 J=1,NT
                    T=DT*(J-1.)
                    FBARST(IX,J)=3.*COF(1,J)*T*T+2.*COF(2,J)*T+COF(3,J)
   24               CONTINUE
   25           CONTINUE
            DO 35 I=1,NX
                K(1,I)=I-1.
   35           CONTINUE
            CALL SETUP(NX,A,H,K)
            DO 55 IT=1,NT
                DO 45 IX=1,NX
   45               K(2,IX)=FBARST(IX,IT)
                CALL SOLVE(A,K,H,NX,D,COF)
                G(1,IT)=0.
                X=0.
                NXM1=NX-1
                DD=0.
                DO 55 J=1,NXM1
                    DD=DD-X*(COF(4,J)+X*(COF(3,J)/2.+X*(COF(2,J)/3.+
     2              X*COF(1,J)/4.)))
                    X=X+1.
                    DD=DD+X*(COF(4,J)+X*(COF(3,J)/2.+X*(COF(2,J)/3.+
     2              X*COF(1,J)/4.)))
                    G(J+1,IT)=-DD/FBAR(J+1,IT)
   55               CONTINUE
            RETURN
            END
C

            SUBROUTINE SETUP(N,A,H,K)
C
C           ****************************************************************
C           THIS ROUTINE SETS UP A TRIDIAGONAL MATRIX A OF THE SPLINE EQUATION
C           IN A  3 BY N  ARRAY  -  FIRST ROW IS SUPERDIAGONAL
C                                   SECOND ROW IS DIAGONAL
C                                   THIRD ROW IS SUBDIAGONAL
C           AND THEN DECOMPOSES A TO  LU  SO THAT
C              FIRST ROW IS SUPERDIAGONAL OF U, SECOND ROW IS DIAGONAL OF U
C             THIRD ROW IS SUBDIAGONAL OF L, DIAGONAL OF L IS UNITY
C           ****************************************************************
C
            REAL A(3,N),H(N),K(2,N)
            NM1=N-1
            H(2)=K(1,2)-K(1,1)
            DO 10 I=2,NM1
                H(I+1)=K(1,I+1)-K(1,I)
                A(1,I)=H(I+1)/(H(I+1)+H(I))
                A(2,I)=2.0
                A(3,I)=1-A(1,I)
   10           CONTINUE
            A(1,1)=-2.
            A(2,1)=2.
            A(3,1)=0.
            A(1,N)=0.
            A(2,N)=2.
            A(3,N)=-2.
            DO 11 I=2,N
```

```
               A(3,I)=A(3,I)/A(2,I-1)
               A(2,I)=A(2,I)-A(3,I)*A(1,I-1)
   11          CONTINUE
          RETURN
          END
C

          SUBROUTINE SOLVE(A,K,H,N,D,COF)
C
C         ******************************************************************
C         THIS ROUTINE CALCULATES THE SECOND ORDER FINITE DIFFERENCES OF THE
C         SPLINE. AND THEN SOLVES  AM=D  , (BY FORWARD AND BACKWARD
C         SUBSTITUTION), PLACING M(VECTOR OF SECOND DERIVATIVES) IN D
C         END POINT CONDITIONS NOW USE THIRD ORDER FINITE DIFFERENCES
C         TO ESTIMATE THIRD ORDER DERIVATIVES AT X0+H/2 AND XN+H/2
C         ******************************************************************
C
          REAL H(N),D(N),A(3,N),K(2,N),COF(4,N)
          D(2)=(K(2,2)-K(2,1))/H(2)
          NM1=N-1
          DO 12 I=2,NM1
               D(I+1)=(K(2,I+1)-K(2,I))/H(I+1)
               D(I)=6*(D(I+1)-D(I))/(H(I+1)+H(I))
   12          CONTINUE
          D(1)  =-2.*(-K(2,1)+K(2,4)+3*(K(2,2)-K(2,3)))/(H(2)*H(2))
          D(N)=-2.*(-K(2,N)+K(2,N-3)+3.*(K(2,N-1)-K(2,N-2)))/(H(N-1)**2)
          DO 13 I=2,N
               D(I)=D(I)-A(3,I)*D(I-1)
   13          CONTINUE
          D(N)=D(N)/A(2,N)
          DO 14 I=2,N
               J=N+1-I
               D(J)=(D(J)-A(1,J)*D(J+1))/A(2,J)
   14          CONTINUE
          CALL POLLY(N,D,K,H,COF)
          RETURN
          END
C

          SUBROUTINE POLLY(N,M,K,H,COF)
C
C         ******************************************************************
C         THIS ROUTINE COMPUTES THE COEFFICIENTS OF THE SPLINE POLYNOMIAL
C            ON EACH SUBINTERVAL
C         K IS THE ARRAY OF DATA POINTS
C         H IS THE VECTOR OF SUBINTERVAL LENGTHS
C         M IS THE SOLUTION VECTOR TO THE EQUATION AM=D
C         ******************************************************************
C
          REAL M(N),K(2,N),H(N),COF(4,N)
          NM1=N-1
          DO 11 I=1,NM1
          COF(1,I)=(M(I+1)-M(I))/(6.*H(I+1))
          COF(2,I)=(K(1,I+1)*M(I)-K(1,I)*M(I+1))/(2.*H(I+1))
          DD=M(I+1)*K(1,I)*K(1,I)-M(I)*K(1,I+1)*K(1,I+1)
         &+2.*K(2,I+1)-2.*K(2,I)
          COF(3,I)=(DD/(2.*H(I+1)))+H(I+1)*(M(I)-M(I+1))/6.
          DD=M(I)*(K(1,I+1)**3)-M(I+1)*(K(1,I)**3)+6.*K(1,I+1)*K(2,I)
         &-6.*K(1,I)*K(2,I+1)+K(1,I)*M(I+1)*(H(I+1)**2)-K(1,I+1)*M(I)*
         &(H(I+1)**2)
```

```
      COF(4,I)=DD/(H(I+1)*6.)
11    CONTINUE
      DO 16 J=1,4
16    COF(J,N)=COF(J,NM1)
      RETURN
         END
C
      SUBROUTINE TRANS(NX,NT,G,GT)
C
C     ***************************************************************************
C     THIS ROUTINE TRANSPOSES A MATRIX
C     ***************************************************************************
C
      REAL G(NX,NT),GT(NT,NX)
      DO 65 IX=1,NX
           DO 65 IT=1,NT
                GT(IT,IX)=G(IX,IT)
 65             CONTINUE
      RETURN
      END
C
      SUBROUTINE MULT(NA,NB,A,B,PROD)
C
C     ***************************************************************************
C     THIS ROUTINE MULTIPLIES TWO MATRICES
C     ***************************************************************************
C
      REAL        A(NA,NB),B(NB,NA),PROD(NA,NA)
      DO 67 IX=1,NA
           DO 67 IT=1,NA
                SUM=A(IX,1)*B(1,IT)
                DO 66 K=2,NB
                SUM=SUM+A(IX,K)*B(K,IT)
 66                  CONTINUE
                PROD(IX,IT)=SUM
 67             CONTINUE
      RETURN
      END
C
      SUBROUTINE EIGEN(PROD,NA,NU,EV)
C
C     ***************************************************************************
C     THIS ROUTINE FINDS THE DOMINANT EIGENVECTOR OF THE  NA BY NA
C     MATRIX  PROD  AND PLACES IT IN  EV  WHICH IS NORMALIZED BY THE
C     1-NORM.  ITERATION  RUNS UNTIL ROUNDING ERROR IS DOMINANT.
C     ***************************************************************************
C
      REAL PROD(NA,NA),EV(NA),NU(NA)
      SUM1=0.
      DO 77 IX=1,NA
           SUM=PROD(IX,1)
           DO 76 IT=2,NA
                SUM=SUM+PROD(IX,IT)
 76             CONTINUE
           NU(IX)=SUM
           SUM1=SUM1+SUM
 77        CONTINUE
```

```
      ERROR=1.
   89 CONTINUE
      SUM2=0.
      DO 87 IX=1,NA
         SUM=PROD(IX,1)*NU(1)
         DO 86 IT=2,NA
            SUM=SUM+PROD(IX,IT)*NU(IT)
   86       CONTINUE
         EV(IX)=SUM/SUM1
         SUM2=SUM2+EV(IX)
   87    CONTINUE
      SUM1=0.
      BOUND=.5*ERROR
      ERROR=0.
      DO 97 IX=1,NA
         SUM=PROD(IX,1)*EV(1)
         DO 96 IT=2,NA
            SUM=SUM+PROD(IX,IT)*EV(IT)
   96       CONTINUE
         NU(IX)=SUM/SUM2
         SUM1=SUM1+NU(IX)
         ERROR=ERROR+ABS(EV(IX)-NU(IX))
   97    CONTINUE
      IF (ERROR.LT.BOUND) GOTO 89
      RETURN
      END

   $ENTRY
```

A(T) FROM THE CLOSED FORM IS

1.000000	1.130877	1.247348	1.336588	1.388775
1.398163	1.363718	1.289234	1.182909	1.056447
0.923773				

D(X) FROM THE CLOSED FORM IS

0.000000	0.002500	0.005000	0.007500	0.010000
0.012500	0.015000	0.017500	0.020000	0.022500
0.025000	0.027500	0.030000	0.032500	0.035000
0.037500	0.040000	0.042500	0.045000	0.047500
0.050000	0.052500	0.055000	0.057500	0.060000
0.062500	0.065000	0.067500	0.070000	0.072500
0.075000	0.077500	0.080000	0.082500	0.085000
0.087500	0.090000	0.092500	0.095000	0.097500
0.100000	0.102500	0.105000	0.107500	0.110000
0.112500	0.115000	0.117500	0.120000	0.122500
0.125000				

A(T) OBTAINED FROM 11 BY 11 MATRIX IS

0.997756	1.130877	1.248731	1.339316	1.390486
1.399143	1.365165	1.290611	1.183847	1.058285
0.921801				

D(X) OBTAINED AS (G(X,T),A(T))/(A(T),A(T))

0.000000	0.002437	0.004990	0.007491	0.009990
0.012487	0.014985	0.017483	0.019980	0.022478
0.024975	0.027473	0.029970	0.032468	0.034965
0.037463	0.039960	0.042457	0.044955	0.047452
0.049950	0.052447	0.054945	0.057442	0.059939
0.062437	0.064934	0.067432	0.069929	0.072426
0.074924	0.077421	0.079918	0.082415	0.084913
0.087410	0.089907	0.092403	0.094899	0.097395

0.099892	0.102386	0.104882	0.107379	0.109874
0.112368	0.114861	0.117357	0.119841	0.122342
0.124579				

```
CORE USAGE        OBJECT CODE=   12528 BYTES,ARRAY AREA=   12156 BYTES,TOTAL AREA
DIAGNOSTICS        NUMBER OF ERRORS=       0, NUMBER OF WARNINGS=       0, NUMBER
COMPILE TIME=     4.75 SEC,EXECUTION TIME=    20.16 SEC,   QUEEN'S WATFOR VERSION

 COST FOR THIS PROGRAM IS  $     1.83            RUN IN HSC     MAR  21,  1974
```

Program IV

It is expected that available data will be in the form of a histogram: that is, for population the number of individuals between the ages of x and $x + h$ years is known. To generate a density function for this histogram, a fourth order spline approximation routine is used. To investigate the accuracy of the scheme, the histogram is regenerated by integrating the spline constructed.

```
$JOB      ACCT-NUM,VERNER,TIME=60
C         *****************************************************************
C         THIS ROUTINE USES A POPULATION HISTOGRAM TO GENERATE A POPULATION
C         DENSITTY FUNCTION USING A FOURTH ORDER SPLINE APPROXIMATION
C         *****************************************************************
          REAL A(303,9),P(100),Q(303),PO
          INTEGER I,J,K,TN,TO,TT
          INTEGER T3
C         RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
          READ(5,100) TN
  100     FORMAT(I3)
          READ(5,101) (P(I), I=1,TN)
  101     FORMAT(10F8.0)
C         RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
C         PRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINT
          WRITE(6,106)
  106     FORMAT('1')
          PRINT, 'THE GIVEN POPULATION IS'
          PRINT, ' '
          WRITE(6,102) (P(I),I=1,TN)
  102     FORMAT(' ',10F8.0)
C         PRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINT
          TO=0
          T3=3*(TN-TO+1)
          CALL ISPLIN(A,TO,TN,T3,P,Q)
          CALL CHECK(P,Q,T3,TN)
          P(1)=Q(3)
          P(2)=Q(5)
                DO 25 I=3,TN
                P(I)=Q(3*I-2)
  25            CONTINUE
C         PRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINT
          PRINT, ' '
          PRINT, ' '
          PRINT, 'THE POPULATION DENSITY  AT AGES  0,1,2,...,99  IS'
          PRINT, ' '
          WRITE(6,102) (P(I),I=1,TN)
          WRITE(6,106)
C         PRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINT
C         PUNCHPUNCHPUNCHPUNCHPUNCHPUNCHPUNCHPUNCHPUNCHPUNCHPUNCHPUNCHPUNCH
C         WRITE(7,101) (P(I),I=1,TN)
C         PUNCHPUNCHPUNCHPUNCHPUNCHPUNCHPUNCHPUNCHPUNCHPUNCHPUNCHPUNCHPUNCH
          STOP
          END

          SUBROUTINE CHECK(P,Q,T3,TN)
C         *****************************************************************
C         THIS SUBROUTINE REGENERATES THE POPULATION HISTOGRAM FROM THE
C         DENSITTY FUNCTION BY EXACT INTEGRATION OF THE FOURTH ORDER SPLINE.
C         HENCE THE ERRORS ARE THOSE SUFFERED AS A RESULT OF ROUND-OFF ERRO
C         *****************************************************************
          INTEGER T3,TN
          REAL Q(T3),P(TN)
          P(1)=(Q(3)+Q(5))/2.-(Q(2)+Q(4))/24.+Q(6)/120
          P(2)=(Q(7)+Q(5))/2.-(Q(8)+Q(4))/24.+Q(9)/120.
          DO 26 I=3,TN
           P(I)=(Q(3*I+1)+Q(3*I-2))/2.-(Q(3*I+2)+Q(3*I-1))/24.+Q(3*I+3)/120
  26            CONTINUE
C         PRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINT
          PRINT,' '
```

```
      PRINT, ' '
      PRINT,'THE POPULATION REGENERATED FROM THE FOURTH ORDER SPLINE IS
      PRINT, ' '
      WRITE(6,102) (P(I),I=1,TN)
  102 FORMAT(' ',10F8.0)
C     PRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTPRINTP
      RETURN
      END

      SUBROUTINE CREATE(A,T0,TN,T3,P,Q)
C     ******************************************************************
C     THIS ROUTINE CREATES THE INTSPLINE MATRIX FOR APPROXIMATING
C     A DENSITY FUNCTION GIVEN SUBINTERVAL INTEGRALS
C     THE UNKNOWN VALUES Q(I) ARE THE VALUES  R4(0),R2(0),R(0),R2(1),
C     R(1),R4(1),R(2),R2(2),R4(2),...,R(I),R2(2),R4(I),...
C     R(T0),R2(T0),R4(T0).
C     BOUNDARY CONDITIONS ARE R''(0)=R'''(0)=R(100)=R''(100)=0
C     ******************************************************************
      INTEGER I,J,K,TN,T0,TT
      INTEGER T3
      REAL A(T3,9),P(TN),Q(T3)
      DO 30 I=1,13
C     RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
      READ(5,103) D
  103 FORMAT(F10.4)
            READ(5,104) (A(I,J),J=1,9)
  104 FORMAT(9F8.5)
      DO 3  J=1,9
      A(I,J)=A(I,J)/D
   30       CONTINUE
C     RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
      TT=TN-T0-1
      DO 31 I=4,TT
            DO 31 J=1,9
                  A(2+3*I,J)=A(11,J)
                  A(3+3*I,J)=A(12,J)
                  A(4+3*I,J)=A(13,J)
   31             CONTINUE
      TT=TT+1
      READ(5,104) (A(3*TT+2,J),J=1,9)
      READ(5,104) (A(3*TT+3,J),J=1,9)
      N1=T0+2
      DO 32 I=N1,TT
            Q(3*I-1)=0.
            Q(3*I)=0.
            Q(3*I+1)=2*P(I)
   32       CONTINUE
      Q(1)=0.
      Q(2)=0.
      Q(3)=2*P(T0+1)
      Q(4)=0.
      Q(3*TT+2)=0.
      Q(3*TT+3)=0.
      RETURN
      END

      SUBROUTINE ISPLIN(A,T0,TN,T3,P,Q)
C     ******************************************************************
C     THIS ROUTINE SOLVES THE INTEGRAL SPLINE PROBLEM FOR A POPULATION
C     ******************************************************************
```

```
      INTEGER T0 ,TN,I,J,K ,T1
      INTEGER T3
      REAL A(T3,9),P(TN),Q(T3)
      CALL CREATE(A,T0,TN,T3,P,Q)
      DO 33 I=1,T3
         Q(I)=Q(I)/A(I,5)
      DO 34 J=6,9
            A(I,J)=A(I,J)/A(I,5)
 34      CONTINUE
      A(I,5)=1
      DO 33 K=1,4
         IF((I+K).GT.T3) GOTO 33
         Q(I+K)=Q(I+K)-A(I+K,5-K)*Q(I)
         DO 35 J=2,5
               A(I+K,4-K+J)=A(I+K,4-K+J)-A(I+K,5-K)*A(I,4+J)
 35         CONTINUE
             A(I+K,5-K)=0
 33      CONTINUE
C     ************************************************************
C     BACKSUBSTITUTION
C     ************************************************************
      N=3*(TN-T0+1)
      Q(N-1)=Q(N-1)-Q(N)*A(N-1,6)
      Q(N-2)=Q(N-2)-(Q(N)*A(N-2,7)+Q(N-1)*A(N-2,6))
      Q(N-3)=Q(N-3)-(Q(N)*A(N-3,8)+Q(N-1)*A(N-3,7)+Q(N-2)*A(N-3,6))
      DO 36 I=5,N
         K=N-I+1
         DO 36 J=1,4
            Q(K)=Q(K)-A(K,5+J)*Q(K+J)
 36      CONTINUE
      RETURN
      END

$ENTRY
```

THE GIVEN POPULATION IS

1023102.	1000000.	994230.	990114.	986767.	983817.	981102.	978541.	976124.	973869.
971804.	969890.	968038.	966179.	964266.	962270.	960201.	958098.	955942.	953743.
951483.	949171.	946789.	944337.	941806.	939197.	936492.	933692.	930788.	927763.
924609.	921317.	917880.	914282.	910515.	906554.	902393.	898007.	893382.	888504.
883342.	877883.	872098.	865967.	859464.	852554.	845214.	837413.	829114.	820292.
810900.	800910.	790282.	778981.	766961.	754191.	740631.	726241.	710990.	694843.
677771.	659749.	640761.	620782.	599824.	577882.	554975.	531133.	506403.	480850.
454548.	427593.	400112.	372240.	344136.	315982.	287973.	260322.	233251.	206989.
181765.	157799.	135297.	114440.	95378.	78221.	63036.	49838.	38593.	29215.
21577.	15514.	10833.	7327.	4787.	3011.	1818.	1005.	454.	125.

THE POPULATION REGENERATED FROM THE FOURTH ORDER SPLINE IS

1023102.	1000000.	994230.	990114.	986766.	983816.	981102.	978541.	976124.	973869.
971804.	969890.	968038.	966178.	964266.	962270.	960201.	958098.	955942.	953743.
951483.	949171.	946789.	944337.	941806.	939197.	936492.	933692.	930788.	927763.
924609.	921317.	917880.	914282.	910515.	906554.	902393.	898007.	893382.	888504.
883342.	877883.	872098.	865967.	859464.	852554.	845214.	837413.	829114.	820292.
810900.	800910.	790282.	778981.	766961.	754191.	740631.	726241.	710990.	694843.
677771.	659749.	640761.	620782.	599824.	577882.	554975.	531133.	506403.	480850.
454548.	427593.	400112.	372240.	344136.	315982.	287973.	260322.	233251.	206989.
181765.	157799.	135297.	114440.	95378.	78221.	63036.	49838.	38593.	29215.
21577.	15514.	10833.	7327.	4787.	3011.	1818.	1005.	454.	125.

THE POPULATION DENSITY AT AGES 0,1,2,...,99 IS

1038460.	1008672.	995297.	992581.	988103.	985351.	982383.	979819.	977299.	974970.
972804.	970830.	968960.	967114.	965234.	963284.	961243.	959156.	957028.	954851.
952623.	950337.	947992.	945576.	943085.	940516.	937861.	935108.	932259.	929297.
926209.	922987.	919625.	916109.	912429.	908568.	904509.	900240.	895735.	890988.
885972.	880664.	875047.	869092.	862781.	856079.	848958.	841394.	833348.	824794.
815694.	806008.	795705.	784748.	773094.	760704.	747547.	733577.	718762.	703069.
686464.	668919.	650419.	630936.	610466.	589017.	566587.	543207.	518911.	493758.
467817.	441169.	413929.	386229.	358212.	330052.	301936.	274070.	246671.	219966.
194185.	169555.	146288.	124581.	104599.	86475.	70297.	56107.	43896.	33602.
25119.	18298.	12960.	8901.	5914.	3786.	2334.	1363.	688.	262.

```
CORE USAGE        OBJECT CODE=     6336 BYTES,ARRAY AREA=   12520 BYTES,TOTAL AREA

DIAGNOSTICS           NUMBER OF ERRORS=         0, NUMBER OF WARNINGS=        0, NUMBE

COMPILE TIME=      2.31 SEC,EXECUTION TIME=    12.33 SEC,   QUEEN'S WATFOR VERSIO

 COST FOR THIS PROGRAM IS  $      1.19           RUN IN HSC      MAR  19,  1974
```