2-1-1956

An Econometric Investigation of the Lifetime Size Distribution of Average Annual Income

Robert Summers

COWLES FOUNDATION DISCUSSION PAPER, NO. 9

An Econometric Investigation of the Lifetime Size Distribution of Average Annual Income

Robert Summers

February 1, 1956

COWLES FOUNDATION DISCUSSION PAPER, NO. 9

An Econometric Investigation of the Lifetime Size Distribution of Average Annual Income*

§1 Estimates of the size distributions of income for the United States have been derived from a number of data sources and are based upon a variety of different definitions of the various relevant variables. The characteristic common to almost all of these estimates is the accounting period used. Partly because of conceptual difficulties that arise when a longer period is used and partly -- perhaps mostly -- because of inadequate data, the size distributions nearly always relate to one year periods. Even in the exceptional cases where accounting periods of over a year were used, the periods covered were of relatively short duration.** Provided the conceptual difficulties can be straightened

---

* This investigation was based upon unpublished data made available to me by the Survey Research Center of the University of Michigan and the Board of Governors of the Federal Reserve System. The Thomas J. Watson Computing Laboratory of Columbia University permitted me to use one of its electronic calculators for the extensive computation that was performed. The Office of Naval Research supplied funds for desk computation and other expenses. I am grateful to each of these organizations; their material assistance was essential for the completion of this project.

** Frank A. Hanna, "The Accounting Period and the Distribution of Income," Analysis of Wisconsin Income, Vol. IX of Studies in Income and Wealth, (New York: National Bureau of Economic Research, 1948), Part III.

George Katona and Janet A. Fisher, "Postwar Changes in the Income of Identical Consumer Units," Vol. XV of Studies in Income and Wealth, (New York: National Bureau of Economic Research, 1951), Part II.

An Analysis of income tax returns filed in the State of Delaware is now in progress which will cover part of the interwar period.

out, great interest would attach to a satisfactory estimate of the size distribution of average annual income for long periods. An estimate of the size distribution for a period as long as the lifetime of the income-receiving unit would be particularly desirable, for that size distribution is a prerequisite for measuring satisfactorily the degree of income inequality.

§2 The data problem can be seen most clearly if the conceptual problems are assumed away. Let the basic income receiving unit of the United States be the household, a loose designation for a group of persons who live together and together determine how they will use their joint income. Most households fit the usual description of a family and the composition of these households follows the usual family time pattern. Single person households are not ruled out, however. The relevant period of the existence of a household for the purposes of this study is from the time one or more of its members first engages in full-time income-seeking activity until all of its members terminate such activity. "Lifetime" here refers to the income-seeking period. It could be assumed that this income-seeking period is of the same length for all American households and that they maintain a continuous existence with an unchanging composition, except for the coming and going of children, for the entire period. This would make the lifetimes of households homogeneous. The variety of training periods for different occupations and the prevalence of divorce courts and life insurance companies testify to the frequency with which the time patterns of real households deviate from the conditions of this assumption. For the present discussion this is not important.

It would appear that a large number of income histories spanning the whole lifetimes of these households would be sufficient for estimating the size distribution of lifetime average annual income. Suppose that in 1915 a foresighted director of the United States Bureau of the Budget had selected a sizable panel of households at the beginning of their income-seeking periods and arranged for them to make available to the Census when they retired from active income-seeking, say forty years later, a record of their intervening incomes. Would the lifetime size distribution that could be made up from these income histories be useful to us? The economic historian would say yes. This is because the size distribution would be of some help in describing the past. But would it tell us much about the way in which incomes are being distributed in the United States economy today? The structure of the economy has undergone substantial changes which make the United States of today a different world from its predecessor of forty, thirty, or even twenty years ago. The concept of a lifetime size distribution is of use only when it refers to a homogeneous period of time. Three wars and a substantial change in the role of government have certainly made the last forty years non-homogeneous.

A foresighted Census director of today could perhaps make available 40 years from now a set of income histories. One's patience would not be rewarded, though, even if he were willing to wait for it. In a world where non-economic factors like wars, technological revolutions, and underdeveloped country awakenings are important in determining economic activity, surely that set of income histories would be as obsolete when it was finally ready as the ones begun in 1915 would be today. It is inevitable that structural changes --

institutional, technological, and behavioral -- will occur too frequently for the income history method ever to provide satisfactory estimates of lifetime size distributions.

Is the concept of a lifetime size distribution of any use then? During any homogeneous period the economy has an income-distributing potential which, if realized over a long enough period of time, would generate a size distribution of lifetime average annual income. Knowing this latent lifetime size distribution is equivalent to knowing the economy's income-distributing potential, and this is worthwhile for all the reasons that prompt the estimation of a size distribution relating to any accounting period. The latent size distribution refers neither to what has happened nor to what probably will. It is a "maybe" size distribution which has a very, very small probability of actually eventuating. It is the answer to the question: What is the size distribution of lifetime average annual income in prospect for households on the threshold of their income-seeking period, assuming that the households remain intact throughout their lifetimes and that the economy's structure remains unchanged? This study is an attempt to answer this question after it has been qualified to refer to only one sector of the household population. It should be emphasized that the objective of this study is the filling in of numbers for a theoretical construct. Even though all households do not really remain intact throughout their full income-seeking period, the question of what their lifetime size distribution would be like if they did so is still of interest.

Since structural changes in the economy occur too often for the income history method to be feasible, some other kind of data must be collected which will adequately describe the income-distributing aspect of the structure before it changes. Cross-section data on the incomes of many individual households within a short period of time would seem to be suited to this purpose.

The method used in this study is analogous with a technique used by actuaries in constructing an up-to-date mortality table. A table could be based upon the death records of a large group of persons who were, say, twenty-one years of age at the beginning of this century. The probability that a person would die between any particular pair of birthdays could be estimated from two numbers contained in the death records: the number of persons who survived to the first birthday, and the number that survived to the second. The difference between the two numbers, divided by the first one, is an unbiased estimate of the probability. This estimate along with those for birthdays up to the seventy-sixth, after appropriate smoothing, could be put together to form the mortality table.

If this method were used, however, today's medical and sanitation standards would only be reflected in the probability estimates for the aged. To avoid this defect, an alternative method is used. Each of a large number persons is observed over a one year period. An estimate of the probability of death at any particular age is given by the relative frequency of death of persons who were at that age in the group under observation. These estimates, properly smoothed, could also be combined to form a mortality table. This table would be based upon today's mortality experience. And not less important is the fact that it was derived from the observation of death incidence for one year only instead of fifty-five.

§3 It will be helpful to have a tag description for households headed by persons in their full-time income seeking period. They will be referred to as "active" households and the income seeking period will be described as the active period.

The term "size distribution" refers to a frequency table of numbers of economic units. Here it will be synonomous with the statistical terms "relative frequency function" or "density function," depending upon whether a discrete or continuous function is under consideration. From the context of its use, it will be clear which is meant.

§4 The data used in this investigation was collected by the Survey Research Center of the University of Michigan as part of the annual Surveys of Consumer Finances which they have been conducting since 1946 for the Board of Governors of the Federal Reserve System. In 1948 the occupants of about 3000 dwelling units were interviewed about various aspects of their current economic status and about their expectations of the future. The occupants of about one-fifth of these dwelling units were reinterviewed a year later. Another reinterview study of the same sort was conducted in 1952 and 1953. As in all of the annual Surveys, the dwelling units of the reinterview studies were selected on a stratified sampling basis from the population of all dwelling units in the United States. Excluded from the surveys were members of the Armed Forces and civilians living on military reservations; residents of hospitals and religious, educational, and penal institutions; and the floating population of persons living in hotels, large boarding houses; and tourist camps.* Field surveys are never entirely

* For a complete description of the methods used by the Survey Research Center in the Survey of Consumer Finances see "Methods of the Survey of Consumer Finances," Federal Reserve Bulletin, July 1950, pp. 795-809; and John B. Lansing, "Concepts Used in Surveys," Contributions of Survey Methods to Economics, by George Katona, Lawrence R. Klein, John B. Lansing, and James N. Morgan (New York: Columbia University Press, 1954), Chapter I.

free of bias in the selection of households actually interviewed. Quality checks by the Survey Research Center of the representativeness of the households interviewed in their annual Surveys and also of those interviewed in the reinterview surveys suggest that the biases are probably not serious. In any case, for the purposes of regression estimation, a basic tool used in this investigation, it is not necessary for sample observations on the independent variable to be representative of the population. The requirement is only that the sample observations in the dependent and independent variables be linked together in a way which is representative of the population. There is no empirical evidence on this point but it seems likely that, at least approximately, this requirement is complied with in the reinterview surveys.

The Survey Research Center uses as its basic income-receiving unit a person or group of persons related by blood, marriage, or adoption who live together and pool their incomes for major expenditures. Such a group is called a spending unit. The spending unit has been adopted in this study as the empirical counterpart of the economic unit which heretofore has been referred to as the household. The income of a spending unit is defined to be the total of all cash income of the spending unit plus accrued profits (net of inventory valuation adjustment) of any unincorporated enterprise owned by the spending unit, all before personal income taxes. The coverage of the study has been restricted to spending units residing in urban areas containing 2500 or more persons because one of the reinterview studies contained no data on rural spending units.

It is assumed in this study that spending units start out their active periods at the age of twenty-five and retire from income-seeking activity at the age of sixty-five. As a consequence, only data relating to spending units headed by persons between twenty-five and sixty-four were used.

§5 The relationship describing the income-dynamics of an active household can be assumed to be of the general form:

$$(1) \qquad y_t = g(y_{t-1}, y_{t-2}, \ldots, y_1, a_t, u_t)$$

where y_i is the household's income in the i'th year; a_i is the age of the head of the household in the i'th year; and u_i is the value assumed in the i'th year by a stochastic variable which is independent of y_{i-1}, y_{i-2}, ..., y_1, and a_i, and has a particular density function $f(u)$.

Without a_i in the right-hand member of Equation (1), the assertion would be that a household's income in the t'th year is some particular function of the household's income in previous years and of a stochastic variable, and that this function holds whether the household is young or old. Considerable evidence pointing toward age as a significant variable in the income life cycle has been collected, however, so age is included in the equation.

There are many other variables which are relevant to the relationship between a household's incomes in successive years besides age. A list of the more important of these would include: (1) the size of the municipality and (2) the region lived in by the household; (3) the occupation and (4) education of the head of the household; and (5) the net worth of the household. If enough data were available it would be possible to apply methods like those used in this study to groups of household which were homogeneous with respect to the various variables listed above. Instead of disaggregating to this extent, it will be assumed that the influence of these variables is contained in the stochastic variable u. The income-dynamic relationship includes a stochastic variable just because relevant variables, economic and non-economic, have been excluded from it. Using u as a substitute for the omitted variables is justified if the proportions of households in the various categories listed above remain the same through time. Since this investigation deals with one aspect of the implications of an economic structure which remains unchanged, this condition is complied with.

While a household's t'th year income is affected to some degree by its income in all earlier years, it is to be expected that years prior to the immediately preceding ones will have only very slight influence. The null hypothesis that only the most recent year is important enough to be taken into account was tested. It was not possible to reject the null hypothesis, but unfortunately the data available for the test were scanty so the power of the test was low. This result is somewhat at a variance with the findings of other investigations of the dependence of a household's t'th year income on its income in the two immediately preceding years* but it is felt that

---

* Frank A. Hanna, *op. cit.*, Table 16, p. 251.

M. Friedman and S. Kuznets, *Income from Independent Professional Practice* (New York: National Bureau of Economic Research, 1945), p. 308.

---

these other studies examined quite special income recipients. As a consequence, it was assumed that Equation (1) could be rewritten:

$$(2) \qquad y_t = g(y_{t-1}, a_t, u_t)$$

§6 A basic assumption of Equations (1) and (2) is that, as far as income dynamics is concerned, except for age and stochastic variation, all households are exactly alike. This can be made more explicit if Equation (2) is rewritten:

$$(3) \qquad y_{it} = g(y_{i(t-1)}, a_{it}, u_t) \qquad \begin{array}{l} i=1,2, \ldots, n \\ t=2,3, \ldots, T \end{array}$$

where y_{ij} is the i'th household's income in the j'th year and a_{ij} is the age of the head of the i'th household in the j'th year. Equation (3) asserts that (1) the income of every household will change from any year to its successor according to a particular rule; and that (2) any household's income will change from year to year according to that same rule. The rule requires that due allowance be made for the age of the household's head.

Suppose the incomes of all the households in the United States have followed the "rule" of Equation (3) for many generations. If there are now n households and their active periods all last T years, their incomes in the years 1 to T can be arranged in the form of the square array designated (4).

$$\text{(4)} \qquad \begin{Vmatrix} y_{11} & y_{12} & \cdots & y_{1T} \\ y_{21} & y_{22} & \cdots & y_{2T} \\ \vdots & \vdots & & \vdots \\ y_{n1} & y_{n2} & & y_{nT} \end{Vmatrix}$$

where y_{ij} stands for the i'th household's income in the j'th year. For the i'th household, income year by year is given by the entries in the i'th row, and the household's age increases year by year as one moves across the row. The first column consists of the incomes of the n households during the first year of their active period.

Any two consecutive items in a row of (4) are linked together by Equation (3). If we knew the specific form of Equation (3), its parameters could be estimated by applying an appropriate statistical technique to such observations from (4) as we might possess. Any complete income history would consist of all of the entries in one row. From one such complete income history the parameters of Equation (3) could be estimated; better estimates would be obtained, of course, if a set of income histories were available. Suppose, though, that we had only fragments of data from any one row but that there were many such fragments. These could also be used for estimating the parameters. Since Equation (3) contains the variable age, it would be necessary that different columns be represented among the fragments. Using this last kind of data would correspond to a cross-section -- "slice of life" -- approach to the dynamics of household income in contrast with the income history approach.

§7 Assume for the purposes of this section that the form of $g(y_{i(t-1)}, a_{it}, u_t)$ of Equation (3) has been specified and that the parameters in it have been estimated by means of the cross-section approach.

Recall that u_t was a stochastic variable with a density function $f(u_t)$. Until now, because the discussion has been general, nothing has been said about whether or not the u's of different time periods are statistically related. The estimation process appropriate for the cross-section approach does not depend upon this statistical relationship if (as is the case in this study) data from successive years are used. The derivation of the size distribution of lifetime average annual income from Equation (3) does require a specification of the joint density function of the u's, $f(u_1, u_2, \ldots, u_T)$, though. It will be assumed that the u's are independent of each other. Then,

(5) $$f(u_1, u_2, \ldots, u_T) = f(u_1) \cdot f(u_2) \cdot \cdot \cdot f(u_T)$$

For each of a set of households, Equation (3) reduces to (T-1) equations of the form of Equation (2) in the (T-1) different y variables, $y_2, y_3, \ldots, y_T$ and the (T-1) different u's, $u_2, u_3, \ldots, u_T$. These are represented by Equation (2').

(2') $$y_t = g(y_{t-1}, a_t, u_t) \qquad t=2,3, \ldots, T$$

The joint density function of $y_2, y_3, \ldots, y_T$ given that y_1 is equal to some particular value y_1^*, $f(y_2, y_3, \ldots, y_T \mid y_1=y_1^*)$, is obtained from the joint density function of the u's by the use of the Jacobian

$\frac{\partial (u_2, u_3, \ldots, u_T}{\partial (y_2, y_3, \ldots, y_T)}$. Specifically,

(6) $$f(y_2, y_3, \ldots, y_T \mid y_1=y_1^*)$$
$$= f[g^{-1}(u_2)] \cdot f[g^{-1}(u_3)] \cdot \cdot \cdot f[g^{-1}(u_T)] \cdot \frac{\partial (u_2, u_3, \ldots, u_T)}{\partial \ u_2, y_3, \ldots, y_T)}$$

where $g^{-1}(u_j)$ is the value of u_j expressed in terms of y_j, y_{j-1}, and a_j .

Let $\bar{y}$ be lifetime average annual income. Then $\bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t$. In principle, the density function of $\bar{y}$, given that y_1 is equal to y_1^* , $f(\bar{y} \mid y_1=y_1^*)$, can be found by using the following relationships:

$$(7) \qquad f(\bar{y}, y_3, y_4, \ldots, y_T \mid y_1 = y_1^*)$$

$$= f(T\bar{y} - y_1^* - y_3 - \cdots - y_T, y_3, y_4, \ldots, y_T \mid y_1 = y_1^*) \cdot \frac{\partial (y_2, y_3, \ldots, y_T)}{\partial (\bar{y}, y_3, \ldots, y_T)}$$

$$(8) \qquad f(\bar{y} \mid y_1 = y_1^*) = \iint_R \cdots \int f(\bar{y}, y_3, y_4, \ldots, y_T \mid y_1 = y_1^*)\, dy_3\, dy_4 \cdots dy_T$$

where R is the appropriate region of variation of $y_3, y_4, \ldots, y_T$.

Equation (8) gives the size distribution of lifetime average annual income for households which started their active lives with incomes equal to y_1^*. Suppose, though, that the first year incomes of the households were not all the same. If the relative frequency distribution of these incomes is $d(y_1=y^i)$, i'1, ..., m, then $f[\bar{y} \mid d(y_1=y^i)]$ will be a composite density function formed by "blending" together m different density functions. In the synthesis, these components, $f(\bar{y} \mid y_1=y^i)$, $i=1, \ldots, m$, would be weighted $d(y_i=y^1), d(y_1=y^2), \ldots, d(y_1=y^m)$, respectively. Equation (9) is a symbolic statement of the composite density function.

$$(9) \qquad f[\bar{y} \mid d(y_1=y^i)] = \sum_{i=1}^{m} d(y_i=y_i) \cdot f(\bar{y} \mid y_1=y^i)$$

§8 In each reinterview sample, just over 500 urban spending units gave their ages in decade intervals and their incomes in each of two successive years. These data were used to find a particular functional form to replace the general function $g(y_{t-1}, a_t, u_t)$ of Equation (2). It must be recognized that any functional form accepted is necessarily only an empirical formula which best fits the data. Since there were sufficient data for the purpose, instead of using age as an explicit variable in the regression, different regressions between y_t and y_{t-1} were fitted for different age groups. Doing this made it possible to drop the restrictive assumption that the u's were independent of age. Since age data came in four coarse decade-wide intervals, estimating four sets of regression parameters did not necessitate wasting age detail. The following equations were selected as possible functional forms for Equation (2):

$$y_t = \alpha_o^{Ai} + \alpha_1^{Ai} y_{t-1} + u_t^{Ai} \qquad i=1,2,3,4 \tag{10}$$

$$\frac{y_t}{y_{t-1}} = \alpha_o^{Ai} + \alpha_1^{Ai} y_{t-1} + u_t^{Ai} \qquad i=1,2,3,4 \tag{11}$$

$$\ell n\ y_t = \alpha_o^{Ai} + \alpha_1^{Ai}\ \ell n\ y_{t-1} + u_t^{Ai} \qquad i=1,2,3,4 \tag{12}$$

where α_o^{Ai} and α_1^{Ai} are the regression coefficients of the income-dynamics relationship for the i'th age group and $\ell n\ y_j$ is the natural logarithm of y_j. The four age groups were: 25-34, 35-44, 45-54, and 55-64.

Least squares estimates of the parameters of each of 12 regressions (four age groups per functional form and three functional forms) were computed for each of the reinterview samples. Equation (12) was judged to be the best version of Equation (2) because in Equation (12) the residuals from it in both reinterview samples were closest to being independent of the independent variable. This criterion was adopted because the justification for using the Method of Least Squares, the Markoff Theorem on Least Squares, requires such independence. This criterion amounts to a specification that the residuals are homoscedastic and that the regression relationship is really linear in the dependent and independent variables. Tests of homoscedasticity and linearity of the three functional forms were performed on the eight different sets of data (four age groups and two reinterview samples). In the case of Equation (12), six of the eight homogeneity tests and seven of the eight linearity tests were passed acceptably. Since this was a better record than that of either of the other two functional forms, Equation (12) was deemed the best description of the income-dynamics relationship. The regression parameters appear in Table 1. Note that the high correlation coefficients indicate a good fit. The residuals from this regression curve were found to be negatively skewed and quite peaked, so care should be taken in interpreting the standard errors and confidence intervals.

§9 Equation (12), particularized by the parameter estimates in Table 1, was selected to describe the income-dynamics of spending units in 1947-1948 and 1951-1952. Should one infer that there was a change in the income-distributing structure of the United States economy between 1948 and 1952 from the fact that the 1947-1948 and 1951-1952 parameter estimates were not identical? Small differences could easily arise because the estimates were subject to sampling variation. But more than that, it is to be expected that the income-dynamics relationship would depend upon aggregate variables for the economy as a whole and the values of these variables probably changed. The change in a spending unit's income from one period to the next surely depends upon the level and rate of change in aggregate personal income. Perhaps the level and rate of change of prices and the rate of growth of population are also important. Only eight observations (four age groups and two pairs of years) were available for investigating the rôle of these variables, so only a guess could be made about the way in which they appeared in Equation (12). The assumption made was that the true income-dynamics relationship in the economy was of the form given in Equation (13).

$$(13) \qquad \left[\ell n \frac{y_t / p_t}{\bar{Y}_t / p_t} \right] = \beta_o^{Ai} + \beta_1^{Ai} \, \ell n \left[\frac{y_{t-1} / p_{t-1}}{\bar{Y}_{t-1} / p_{t-1}} \right] + v_t^{Ai} \qquad i=1,2,3,4$$

Here $\bar{Y}_j$ = aggregate personal income in the j'th year per spending unit, and p_j is the price level of consumers' goods in the j'th year.

Clearly, Equation (13) is merely a different way of writing Equation (12). The parameters of Equation (12) are related to those of Equation (13) as follows:

$$(14) \qquad \beta_1^{Ai} = \alpha_1^{Ai} \qquad i=1,2,3,4$$

$$(15) \qquad \beta_o^{Ai} = \alpha_o^{Ai} - (1-\alpha_1^{Ai}) \, . \, \ell n \, \bar{Y}_{t-1} - \ell n \, \gamma_t \qquad i=1,2,3,4$$

$$(16) \qquad \sigma_v^{Ai} = \sigma_u^{Ai} \qquad i=1,2,3,4$$

where $\gamma_t = \dfrac{\bar{Y}_t}{\bar{Y}_{t-1}}$.

It is maintained that Equation (13) is a sensible way of incorporating the level and rate of change of aggregate personal income, the level and rate of change of prices, and the rate of change of population into the income-dynamics relationship because of its reasonable implications:

(1) a k per cent rise in Personal Income per spending unit between two years will result in each spending unit's enjoying a k per cent rise in income over what it would have received if there had been no rise in Personal Income per spending unit;

(2) the expected value of a spending unit's income in any year will vary with the level of Personal Income per spending unit of the previous year (though the relationship will not be a proportionate one).

Allowance was made for the differences in the values of the various aggregate variables between 1947-1948 and 1951-1952, and then statistical tests were performed to see if the remaining differences between the parameter estimates could be accounted for by sampling variation. The null hypothesis was not rejected that the standard deviations of the v's for 1947-1948 were within 10 per cent of those of the v's for 1951-1952. Furthermore, for three of the age groups, it was not possible to reject the hypothesis that the population regression coefficients for the two pairs of years were the same. The probability was only .004, however, that the differences in the regression coefficient estimates in the 25-34 age group arose by chance. All in all, this suggests some structural change between 1947 and 1952 but not much. How large can the change be and yet still be economically insignificant? A guiding rule here might be "A difference that makes no difference is no difference." In §13 the significance of the observed degree of structural change will be considered.

§10 It was indicated in §7 that, in principle, if one started with an income-dynamics relationship, a density function, f(u), and an initial income distribution, d(y), the size distribution of lifetime average annual income could be obtained as an analytic function. In practice, however, the ease with which this can be done depends upon the actual form of the income-dynamics relationship and the density function. Unfortunately, the linear-in-the-logarithms form of Equation (12) gives rise to a multiple integral which cannot be evaluated directly so a numerical method must be used. The method most suitable to the overall economic model seemed to be the Monte Carlo technique as applied to distribution sampling. As an alternative to trying to make the intractable integral manageable, a miniature income distribution system exactly like the one described by the economic model was set up. Groups of spending units of age 25 were then "put to work" in the system. An electronic "paymaster" kept track of the incomes of each of the spending units "year by year", and recorded their lifetime average annual incomes after they "retired" on their 65th birthdays. The relative frequency distribution of these lifetime averages was used as an estimate of the size distribution of lifetime average annual income in urban United States. This computational procedure, in which physical phenomena are mimicked with a calculating machine, is perhaps unprecedented in an empirical, as contrasted with methodological, investigation in economics, but it is quite common in the physical sciences.

A particular spending unit's lifetime average annual income was computed as follows:

1. $\ell n\ y_2$ and y_2 were obtained from y_1 by means of Equation (12), outfitted with the regression coefficients for $i = 1$. The value of u was obtained by drawing a number at random from a normal distribution with zero mean and standard deviation equal to the estimated value of σ_u^{Ai}. (The normal distribution was used for computational simplicity even though the residuals from Equation (12) were observed to be non-normal.) This process was repeated eight times, giving in all $y_1, y_2, \dots, y_{10}$.
2. Starting with y_{10}, the calculations of (1) were repeated, using the regression coefficients of Equation (12), $i = 2$. This yielded $y_{11}, y_{12}, \dots, y_{20}$.
3. The calculations of (1) were repeated, using the regression coefficients of Equation (12), $i = 3$ and $i = 4$, to get $y_{21}, y_{22}, \dots, y_{40}$.
4. The arithmetic mean, $\bar{y}$, of $y_1, y_2, \dots, y_{40}$ was calculated.

§13 The Monte Carlo computation was carried out for a large number of spending units. These spending units were to start out their economic lives with incomes distributed like those of actual spending units in the economy on the thresholds of their income-seeking periods. Unfortunately, because of the unavailability of data of this sort, it was necessary to make crude guesses about the parameters of this initial distribution. Four different guesses are described in Table 2. It was found that though the guesses covered a wide range of distributions, the final lifetime size distributions based upon the various guesses were all very much the same. Table 3 gives these final lifetime size distributions. Figure 1 contains graphs of the four cumulative frequency distributions of the lifetime size distributions for the four guesses. Logarithmic probability paper has been used for the graphs so that the distributions can be compared visually with the log-normal distribution. The guess considered best, incidentally, is distribution ⟨2⟩ so a 99 per cent confidence band around the cumulative frequency distribution for ⟨2⟩ was drawn in. The lifetime size distribution clearly is quite insensitive to differences in the initial distribution.

Notice that the headings of Table 3 and Figure 1 both contain references to "Pool" parameters. As reported in §12, some structural change was found between 1948 and 1952. The sensitivity of the lifetime size distribution to this change was investigated by running through the Monte Carlo computation using both the structural parameters observed for 1947-1948 and those for 1951-1952. The Pool parameters are, in a crude sense, averages of the 1947-1948 and 1951-1952 ones. After an adjustment occasioned by the differences in the level and rate of changes in Personal Income per spending unit between the two pairs of years, the data were properly weighted and then pooled into four large samples, one for each age group, and regression parameters were estimated for the pooled data. These Pool parameter estimates provide the best estimates of the true regression parameters. The Monte Carlo computation was run through a third time using the Pool parameters. Tables 3, 4, and 5 give the lifetime size distributions based upon the three sets of parameters for all four different initial distribution guesses. Figure 2 contains the log-probability paper graphs of the cumulative frequency distributions for the three sets of parameters and the 99 per cent confidence band for the Pool parameters. These are all based upon initial distribution ⟨2⟩ . The fact that the three different lifetime size distributions are very close to each other indicates that the lifetime size distribution is insensitive to small structural changes.

§14 The impact of the economy's structure on the size distribution of lifetime average annual income is contained in Equation (13). Thus, specifying that the structure does not change means that throughout the lifetime period, the parameters of Equation (13) do not change. But what should be the dynamic character of the economy as a whole during the lifetime of the individuals under examination? Should the number of households in the economy be expanding? Should income per household be expanding?

The decision was made that the lifetime size distribution should be derived for a stationary economy -- one with a constant income per household and fixed consumers' price level. The argument for basing the lifetime size distribution on an expanding economy model is that this specification is more descriptive of the real world. This kind of realism is not what is wanted, though. The relevant consideration is that the income-distributing potential of the economy at a particular time is best described by the lifetime size distribution derived from a stationary economy model. Aggregate income, population, and the price level in the model should be poised at the corresponding levels of the economy being described.

The question this study is designed to answer then is: What is the size distribution of lifetime annual average income in prospect for urban households on the threshold of their income-seeking periods, assuming that the households remain intact, that the economy's real Personal Income per spending unit remains constant, and that the economy's structure does not change?

SUMMARY AND CONCLUSION

§15 It has been assumed that the income-distributing potential of the American economy is determined by a particular aspect of the structure of the economy. This is the income-dynamics relationship which describes how spending unit incomes change from year to year. A set of four first-order difference equations which are linear in the logarithms of incomes was found to be a satisfactory framework for the income-dynamics relationship. By incorporating in the equations exogeneous variables relating to demographic and aggregate economic conditions, it was possible to assert that this aspect of the economy's structure changed only unimportantly during four years of the post-world War II period. Each of the four equations describes the income-dynamics relationship for ten years of the 40 year period that the typical spending unit spends in active, income-seeking activity. Equations (17), (18), (19), and (20) describe this relationship for urban spending units in the United States in 1951 on the very special assumption that the economy is "slowed down" sufficiently for Personal Income per spending unit to remain constant. The parameter estimates are considered the best of the three sets that were found. The parentheses under the estimates of the coefficients in the equations contain the standard errors of the coefficients and the parentheses under the estimates of the standard deviations contain the 95 per cent confidence intervals of the standard deviations.

(17) $$\ell n\ y_t = \underset{(.035)}{.767}\ \ell n\ y_{t-1} + \underset{(.289)}{1.954} + v$$ 25-34

$$Ev=0 \qquad S_v=.291$$
$$(.274 \text{ to } .309)$$

(18) $$\ell n\ y_t = \underset{(.034)}{.847}\ \ell n\ y_{t-1} + \underset{(.278)}{1.294} + v$$ 35-44

$$Ev=0 \qquad S_v=.367$$
$$(.348 \text{ to } .385)$$

(19) $$\ln y_t = .878 \ln y_{t-1} + 1.032 + v$$ 45-54

$$\underset{(.030)}{} \qquad \underset{(.245)}{}$$

$$Ev=0 \qquad S_v=.358$$

$$(.340 \text{ to } .377)$$

(20) $$\ln y_t = .814 \ln y_{t-1} + 1.533 + v$$ 55-64

$$\underset{(.036)}{} \qquad \underset{(.286)}{}$$

$$Ev=0 \qquad S_v=.454$$

$$(.419 \text{ to } .493)$$

The most noteworthy feature of this system of equations is the size of the coefficients of $\ln y_{t-1}$. Because each of these is less than unity, the influence of a spending unit's initial income on its income in later years becomes weaker and weaker with time.*

---

* The size of the coefficients also has implications for the proposition "The rich get richer and the poor get poorer," but these will not be explored here.

---

§16 The income-dynamics relationship for urban spending units was used to estimate the size distribution of lifetime average annual income for urban spending units starting their active income-seeking periods in 1951. In arriving at the lifetime size distribution, it was assumed that all spending units remained in existence for forty years and that the structure of the American economy remained unchanged throughout the whole period. Furthermore, it was assumed that, unlike the true state of affairs, average spending unit income remained constant during the period.

Table 6 gives a relative frequency table of the best estimate of the size distribution of lifetime annual average income obtained and Figure 3 contains a histogram of it. Figure 4 contains a log-normal probability graph of the cumulative frequency distribution of lifetime average annual income, and a band about it that corresponds, approximately to a 99 per cent confidence band.

§17 One would expect that some of the inequality in the size distribution of income in the whole economy could be explained by the correlation between age and income. Even if no inequality was displayed in the size distributions for each age group, the single amalgamated size distribution for the economy as a whole would display inequality if the medians of the separate distributions for each age group were unequal. Clearly, such inequality would be of little significance, though. Conceivably, the variability of household income over time might be a cause for concern, but there would be no inequality problem in the usual sense. To the extent, however, that the percentile standings of household change in the age size distributions from year to year, the age size distributions also exaggerate lifetime income inequality.

The degree of inequality for the lifetime size distibution of Table 6 was compared with that of the economy as a whole and with those of the various age size distributions for the year 1951. The lifetime size distribution displayed significantly less inequality than any of the other distributions.

Table 1: The least squares estimates of the regression parameters of Equation (12): $\ln y_t = \bar{\alpha}_o + \bar{\alpha}_1 \ln y_{t-1} + u$.

	N		$\bar{\alpha}_1$		$\bar{\alpha}_o$		$\bar{\sigma}_u$		r	
Age	1947-48	1951-52	1947-48	1951-52	1947-48	1951-52	1947-48	1951-52	1947-48	1951-52
25-34	119	133	.830	.718	1.389	2.444	.301	.275	.798	.815
			(.058)	(.045)	(.470)	(.367)	(.27 to .34)	(.24 to .31)	(.73 to .85)	(.74 to .86)
35-44	160	148	.838	.847	1.358	1.352	.398	.341	.764	.853
			(.056)	(.043)	(.461)	(.356)	(.35 to .46)	(.30 to .39)	(.69 to .82)	(.80 to .89)
45-54	151	142	.854	.946	1.271	.474	.406	.312	.864	.870
			(.041)	(.046)	(.328)	(.382)	(.36 to .46)	(.28 to .36)	(.81 to .90)	(.82 to .90)
55-64	96	95	.720	.880	2.292	1.024	.537	.371	.907	.913
			(.062)	(.041)	(.489)	(.329)	(.47 to .63)	(.32 to .43)	(.87 to .94)	(.87 to .94)

Notes to Table 1

1. A Least Squares estimate is indicated by a bar over the Greek letter standing for the parameter estimated.

2. The number in parentheses under each regression coefficient estimate is the standard error of that estimate.

3. The pair of numbers following each $\bar{\sigma}_u$ is the 95 per cent confidence interval of the $\bar{\sigma}_u$, computed on the assumption that $f(u)$ is normally distributed.

4. The pair of numbers following each r, the correlation coefficients, is the 95 per cent confidence interval of the ρ, computed on the assumption that the joint distribution of the dependent and independent variables is bivariate normal.

Table 2: Size distributions of income of 25 year olds

Income	Relative Frequency 1	2	3	4
$750	.0077	.0066	-	-
1250	.0397	.0642	-	.0129
1750	.0817	.1558	.0094	.0983
2250	.1096	.1980	.1057	.1973
2750	.1169	.1780	.2670	.2154
3250	.1124	.1396	.2807	.1746
3750	.0994	.0943	.1880	.1174
4250	.0842	.0632	.0930	.0766
4750	.0673	.0385	.0360	.0445
5250	.0574	.0243	.0134	.0263
5750	.0474	.0148	.0046	.0155
6250	.0361	.0088	.0015	.0090
6750	.0270	.0052	.0005	.0051
7250	.0230	.0035	.0001	.0030
7750	.0180	.0019	.0001	.0015
8250	.0151	.0013	-	.0011
8750	.0105	.0007	-	.0006
9250	.0090	.0005	-	.0003
9750	.0075	.0003	-	.0003
12,500	.0259	.0005	-	.0001
17,500	.0034	-	-	.0002
22,500	.0005	-	-	-
27,500	.0002	-	-	-
Total	1.0000	1.0000	1.0000	1.0000
No. of Spending Units	965	885	665	805
Median	3656	2697	2697	2440
Mean of Natural Logs	8.204	7.900	7.900	7.800
Standard Deviation of Natural Logs	.535	.400	.250	.400

Table 3: Size distributions of lifetime average annual income in urban United States in 1951 derived from Pool regression coefficients

Income Class	Relative Frequency			
	⟨1⟩	⟨2⟩	⟨3⟩	⟨4⟩
$ 0-999	-	-	-	-
1000-1999	-	-	-	-
2000-2999	.047	.055	.057	.057
3000-3999	.176	.192	.199	.197
4000-4999	.277	.298	.293	.305
5000-5999	.203	.195	.200	.189
6000-6999	.102	.078	.070	.070
7000-7999	.069	.062	.061	.062
8000-8999	.053	.055	.062	.056
9000-9999	.025	.022	.019	.020
10,000-10,999	.020	.020	.017	.022
11,000-11,999	.009	.005	.004	.004
12,000-12,999	.003	.002	.001	.001
13,000-13,999	.006	.006	.008	.005
14,000-14,999	.002	.001	.001	.001
15,000-15,999	.007	.009	.008	.011
16,000-16,999	.001	-	-	-
17,000-17,999	-	-	-	-
Total	1.000	1.000	1.000	1.000
No. of Spending Units	965	885	665	805
Size Distribution of Income of 25 year old spending units:				
Median	3656	2697	2697	2440
Mean of Natural Logs	8.204	7.900	7.900	7.800
Standard Deviation of Natural Logs	.535	.400	.250	.400

Table 4: Size distributions of lifetime average annual income in urban United States in 1951 derived from 1947-1948 regression coefficients

Income Class	Relative Frequency			
	1	2	3	4
$ 0-999	-	-	-	-
1000-1999	.006	.008	.005	.011
2000-2999	.067	.077	.084	.079
3000-3999	.217	.240	.254	.243
4000-4999	.274	.292	.283	.299
5000-5999	.170	.157	.157	.151
6000-6999	.100	.080	.080	.071
7000-7999	.075	.070	.069	.071
8000-8999	.037	.029	.028	.027
9000-9999	.022	.023	.019	.024
10,000-10,999	.010	.003	.001	.003
11,000-11,999	.005	.004	.004	.002
12,000-12,999	.005	.002	.001	.001
13,000-13,999	.005	.006	.008	.006
14,000-14,999	.007	.009	.007	.012
15,000-15,999	-	-	-	-
Total	1.000	1.000	1.000	1.000
No. of Spending Units	965	885	665	805
Size Distribution of Income of 25 year old spending units:				
Median	3656	2697	2697	2440
Mean of Natural Logs	8.204	7.900	7.900	7.800
Standard Deviation of Natural Logs	.535	.400	.250	.400

Table 5: Size distribution of lifetime average annual income in urban United States in 1951 derived from 1951-1952 regression coefficients

Income Class	Relative Frequency			
	◇1	◇2	◇3	◇4
$ 0-999	-	-	-	-
1000-1999	.004	.008	.005	.011
2000-2999	.058	.061	.057	.061
3000-3999	.186	.192	.203	.190
4000-4999	.272	.297	.303	.303
5000-5999	.179	.184	.185	.186
6000-6999	.092	.066	.055	.059
7000-7999	.068	.059	.065	.054
8000-8999	.049	.045	.038	.048
9000-9999	.032	.029	.035	.027
10,000-10,999	.017	.019	.018	.020
11,000-11,999	.018	.017	.013	.020
12,000-12,999	.005	.002	.001	.002
13,000-13,999	.003	.003	.004	.002
14,000-14,999	.002	.001	.001	-
15,000-15,999	.006	.008	.008	.010
16,000-16,999	.003	.002	.001	.001
17,000-17,999	.001	.001	.001	-
18,000-18,999	.002	.001	-	-
19,000-19,999	.003	.005	.007	.006
20,000-20,999	-	-	-	-
Total	1.000	1.000	1.000	1.000
No. of Spending Units	965	885	665	805
Size Distribution of Income of 25 year old spending units:				
Median	3656	2697	2697	2440
Mean of Natural Logs	8.204	7.900	7.900	7.800
Standard Deviation of Natural Logs	.535	.400	.250	.400

Table 6: Relative frequency table of the best estimate of the size distribution of lifetime average annual income of urban spending units in the United States in 1951.

Income	Relative Frequency
$ 0-999	-
1000-1999	-
2000-2999	.055
3000-3999	.192
4000-4999	.298
5000-5999	.195
6000-6999	.078
7000-7999	.062
8000-8999	.055
9000-9999	.022
10,000-10,999	.020
11,000-11,999	.005
12,000-12,999	.002
13,000-13,999	.006
14,000-14,999	.001
15,000-15,999	.009
Total	1.000
Mean	$ 5,420
Median	$ 4,880
Standard Deviation	$ 2,190

N = 885

Figure 1: CUMULATIVE FREQUENCY DISTRIBUTIONS
BASED UPON POOL PARAMETER ESTIMATES

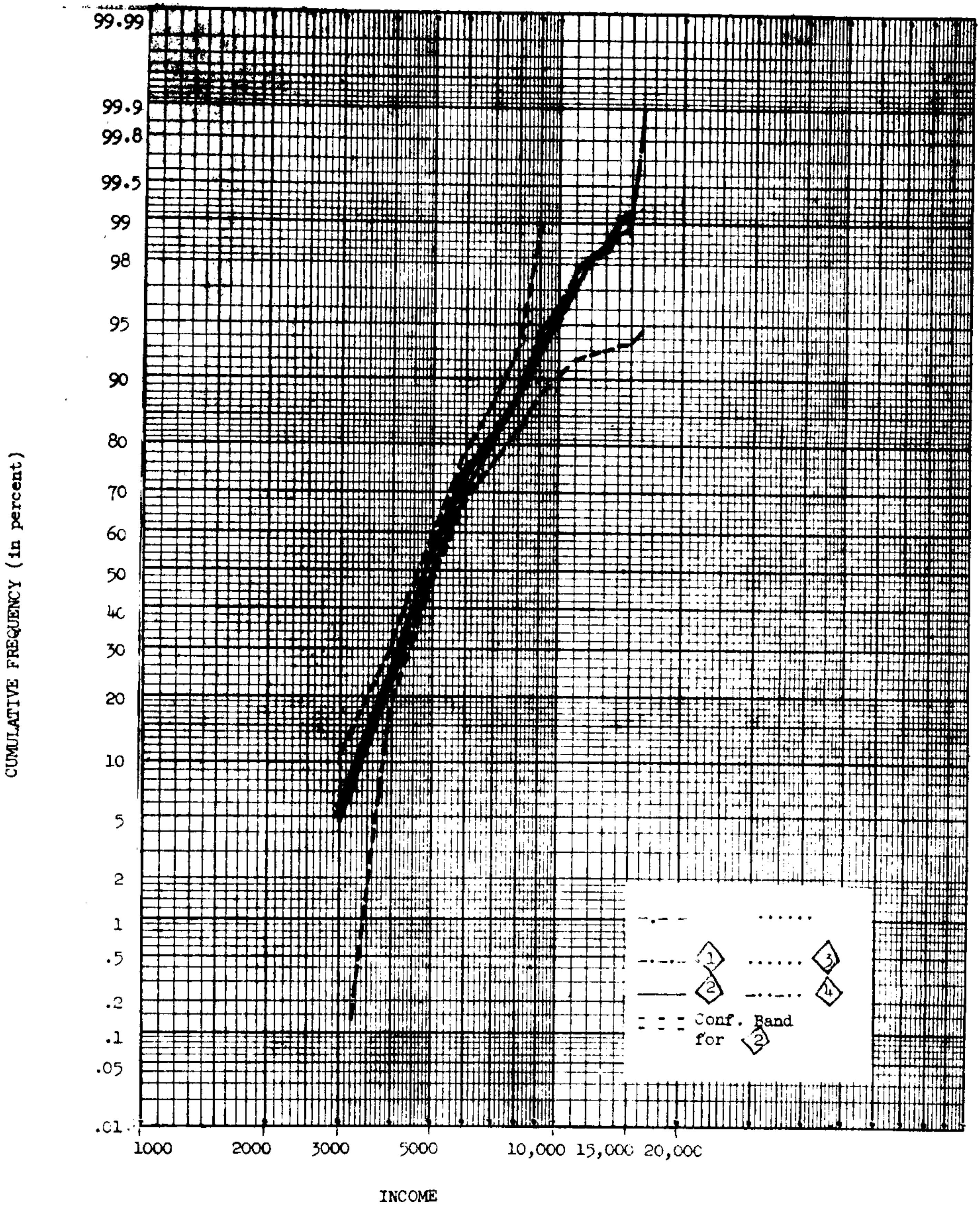

Figure 2: CUMULATIVE FREQUENCY DISTRIBUTIONS
BASED UPON 25 YEAR OLD SIZE DISTRIBUTION ◇2

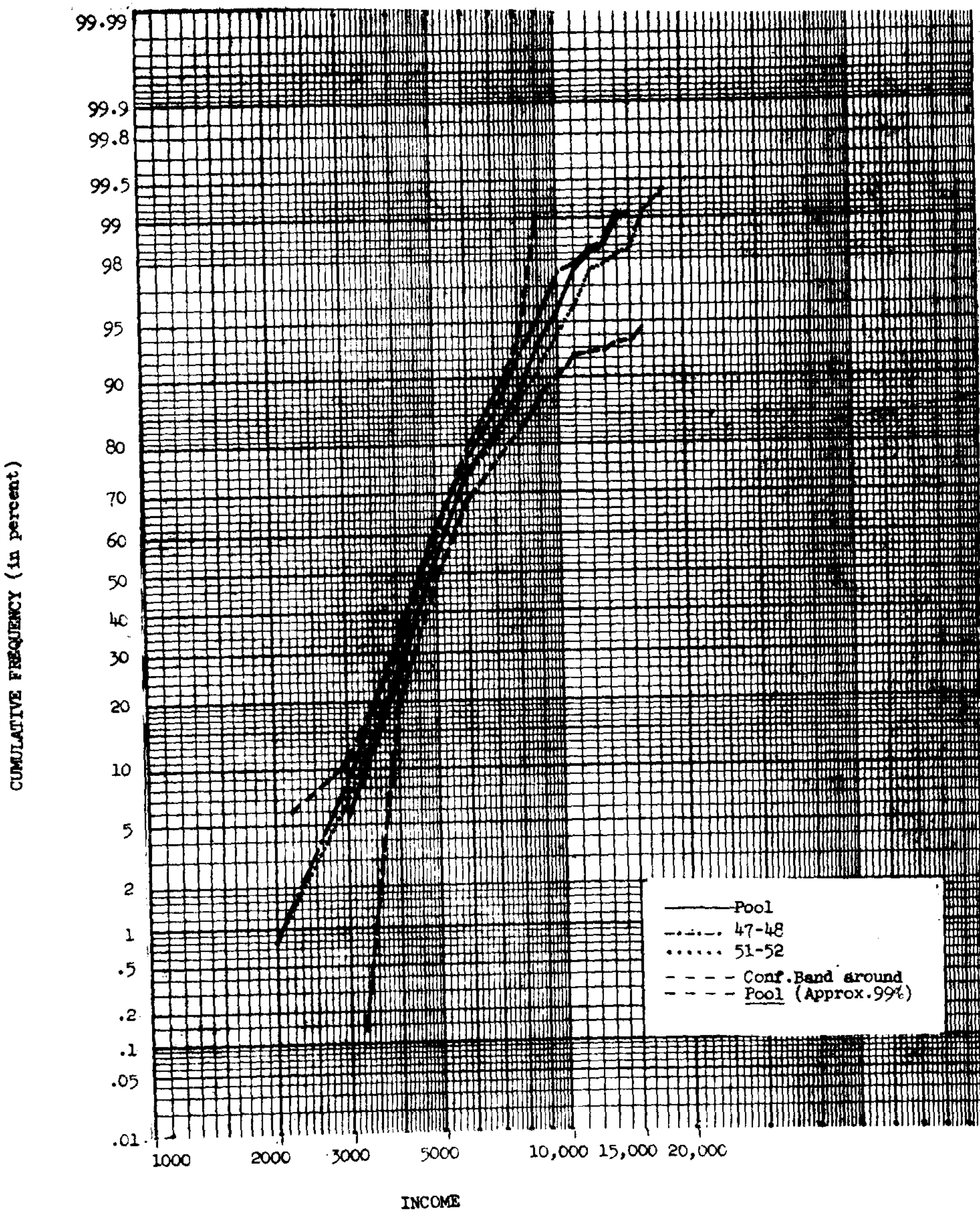

Figure 3: HISTOGRAM OF THE RELATIVE FREQUENCY TABLE OF THE BEST ESTIMATE OF THE SIZE DISTRIBUTION OF LIFETIME AVERAGE ANNUAL INCOME OF URBAN SPENDING UNITS IN THE UNITED STATES IN 1951

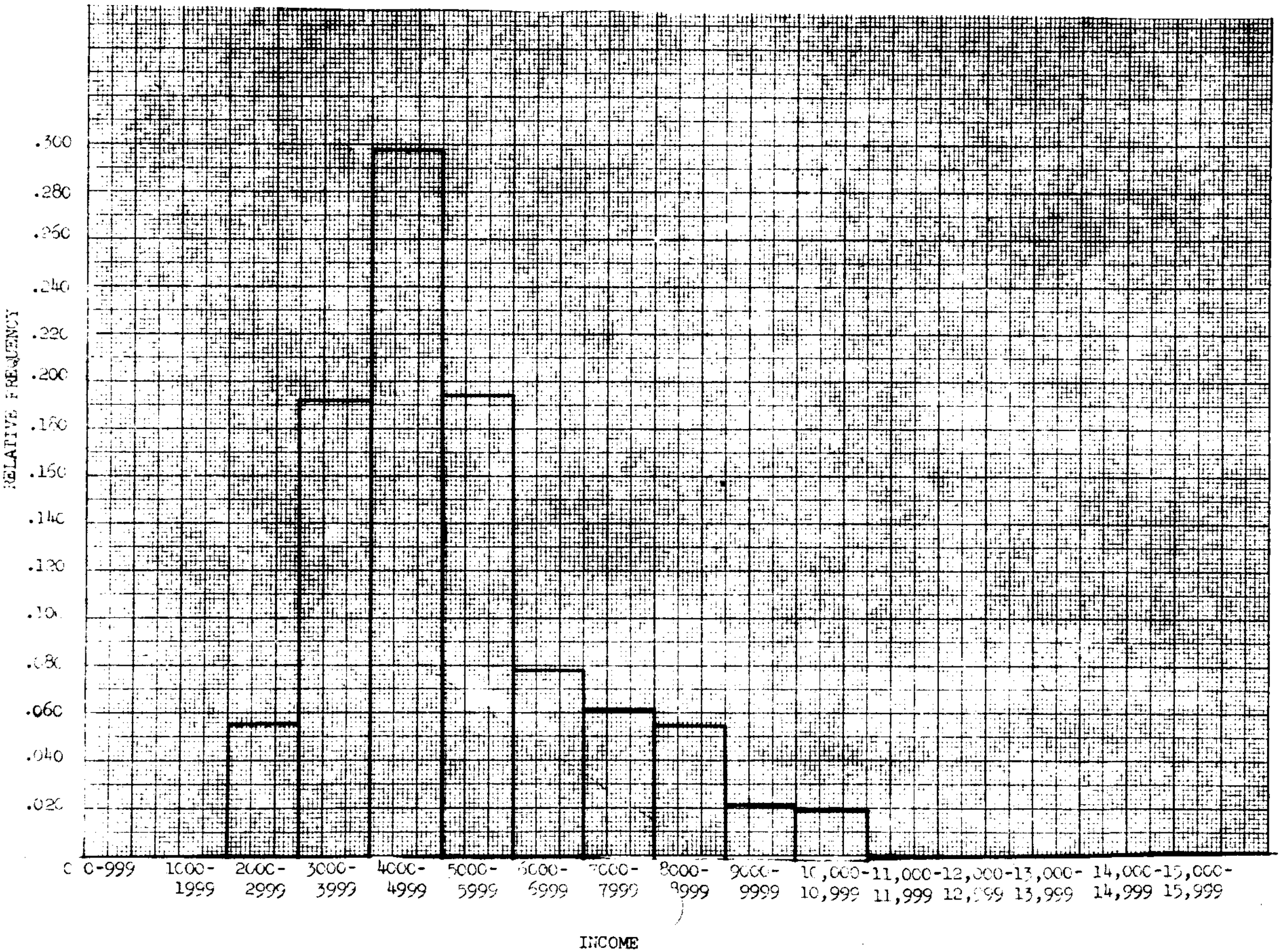

Figure 4: CUMULATIVE FREQUENCY DISTRIBUTION OF LIFETIME AVERAGE ANNUAL INCOME OF URBAN SPENDING UNITS IN THE UNITED STATES IN 1951, BASED UPON POOL PARAMETER ESTIMATES AND SIZE DISTRIBUTION OF 25 YEAR OLD INCOME ⟨2⟩