Volume Title: Income in the United States, Its Amount and Distribution, 1909-1919, Volume II: Detailed Report

Volume Author/Editor: Wesley Clair Mitchell, editor

Volume Publisher: NBER

Volume ISBN: 0-87014-001-9

Volume URL: http://www.nber.org/books/mitc22-1

Publication Date: 1922

Chapter Title: The Construction of a Frequency Curve for All Income Recipients

Chapter Author: Frederick R. Macaulay

Chapter URL: http://www.nber.org/chapters/c9429

Chapter pages in book: (p. 424 - 425)

CHAPTER 33

THE CONSTRUCTION OF A FREQUENCY CURVE FOR ALL INCOME RECIPIENTS

The direct and only adequate method of discovering what is the frequency distribution of income in the United States would be to define very carefully the terms *income* and *income recipient* and then have a carefully planned census taken by expert enumerators upon the basis of these definitions. The returns brought in by the enumerators should moreover be sworn to by the persons making them and heavy penalties attached to the making of false or inaccurate returns. A less satisfactory method but one which would probably give excellent results would be to have a large number of truly random samples taken by such a census. The results of either procedure could then be adjusted in the light of other statistical information concerning the National Income and also in the light of theoretical conclusions derived from the data themselves.

Constructing an income frequency distribution for all income recipients in the United States from the existing data, a few of whose peculiarities have been noted in the preceding chapters, necessarily involves an extremely large amount of pure guessing. It is only because of the practical value of even the roughest kind of an estimate that any statistician would think of attacking the problem. The method followed in the actual construction of the income frequency distribution has been outlined in volume I.[1] This method contains one assumption after another that is open to question. Moreover we feel in many cases quite unable to estimate the probable errors involved in these assumptions. Their only excuse is their necessity. What is the amount of under-reporting for income tax and how is it distributed? What is the effect upon the returns of "legal evasion?" To what extent is the "bulge" on the income-tax returns in the region under about $5,000 in 1918 the result of the "intensive drive?" What is the relation between wages and total income by wage intervals? What is the relation between wage rates and earnings in any particular industry? Etc., etc. These are all questions which must be answered over and over again and yet they are questions the answers to which must be, in many instances, almost pure guesses. And, to repeat, the margin of possible error is often large.

In view of the sparsity and inadequacy of the data, our first approach to the problem was an attempt to discover, if possible, some general mathematical law for the distribution of income. Were we to get any very defin-

[1] *Income in the United States*, Vol. I, pp. 122-139.

ite and reliable clues as to the mathematical nature of the frequency distribution of income from small sample income distributions and from wages distributions, etc., such clues might of course be invaluable in checking the results obtained from piecing together existing wage distributions, income distributions, and other scattered information. We would be in the position of the astronomer who is able to "adjust" the results of his observations in the light of some known mathematical law. It soon became clear, however, that it is quite impossible to discover any essential peculiarities of the income frequency distribution. The available material is not only insufficient for purposes of such generalizations, but moreover the distribution from year to year is so dissimilar, that any generalization of this nature is too vague to be of any practical value.

The method finally used for the construction of the income curve has therefore, we are sorry to say, practically all the weaknesses of the data from which it has been constructed. The occupations of the country were tabulated and to each occupation was assigned those wage and income distributions which seemed applicable with the least strain. We had then a series of income and wage distributions which nominally covered nearly all the income recipients in the United States, though for some occupations the inadequacy of the wage and income samples was little short of absurd. The wage distributions were converted into income distributions on the assumption that the smaller the wage the larger is its percentage of total income. Beyond this simple assumption the particular functional relationships used for many industries were almost pure guess work. Moreover, not only was there the danger of error in moving from wage distribution to income distribution and the danger of error resulting from estimating a wage distribution for a particular industry in a particular locality from a similar though not identical industry in a different locality, but also there was the danger of error resulting from estimating a wage distribution for one year from a wage distribution for another.

The final results are probably not quite so bad as they might have been had we not had a number of collateral estimates with which roughly to check up and otherwise adjust the first results of our estimates. For example, such independent information as Mr. King's estimate of the total income of the country and Mr. Knauth's estimate of the total amount of income from dividends were pieces of information with which the results of the frequency curve calculations were made to agree.

Some hypothetical reasoning is inevitable in such a statistical study as the present one. The investigator must not lose heart. Sir Thomas Browne in his rolling periods sagely remarks that "what song the Syrens sang, or what name Achilles assumed when he hid himself among women, though puzzling questions, are not beyond all conjecture!"