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 Author: Frederick R. Macaulay

Chapter URL: http://www.nber.org/chapters/c9428

Chapter pages in book: (p. 418 - 423)

CHAPTER 32

WAGE DISTRIBUTIONS

There is in all an immense amount of American wage data. On the other hand, as an investigator gets into his subject, he begins to realize that the material is more remarkable for its fragmentary nature than for its amount —great as that may be. For no recent year can he obtain wage distributions for more than about 8 per cent. of those gainfully employed. Of course, if these 8 per cent. were scattered over the different types of employment and localities in any truly random fashion, and if their wages were uniformly reported, much might be done with the material. As things are, however, whole occupations as important as agricultural labor and trade are almost unrepresented. Moreover, as we are interested in the amount of wages actually received during the year, it is rather discouraging to find that this is the one type of distribution which practically never occurs. Distributions of amounts actually earned in a month are almost as rare. There are a few distributions of amounts actually earned in a week or fortnight, but the great majority of wage distributions are distributions of wage *rates*—figures by the *hour* being the commonest—or of hypothetical earnings, generally known as full-time earnings per week.

Now it is in general impossible to construct a wage distribution for earnings from a distribution of rates. Earnings depend, of course, not only on rates but also on hours worked. However, we seldom know anything about the distribution of hours worked and almost never do we know anything about the relation between *rates* and hours worked. Chart 32A illustrates how violent may be the difference in shape of the *earnings* and *rates* curves for the same individuals.[1] The *earnings* distribution in this particular case shows not only a much greater scatter than the *rates* distribution but is of an entirely different shape, as may be seen from Chart 32B where the data are drawn on a double log scale. Chart 32C shows the distribution of hours worked in a week for the same individuals. Now, though the slaughtering and meat packing industry may be an extreme example, what evidence we have suggests that distributions of *rates* and of *earnings* are rarely in close agreement. Moreover the relation of the one distribution to the other changes as we pass from industry to industry.[2]

[1] 43,063 Male Employees in the Slaughtering and Meat Packing Industry in 1917. Bureau of Labor Statistics, *Bulletin 252*. For purposes of comparison the two distributions are so placed that the frequency curves show the same arithmetic means and areas.

[2] Resulting largely, of course, from the varying types of distributions of hours-worked-in-

CHART 32 A

FREQUENCY DISTRIBUTIONS OF RATES OF WAGES PER HOUR AND EARNINGS PER WEEK FOR 43,063 MALE EMPLOYEES IN THE SLAUGHTERING AND MEAT PACKING INDUSTRY IN THE U. S. IN 1917.

SOURCE: BUREAU OF LABOR STATISTICS BULLETIN 252

SCALES NATURAL

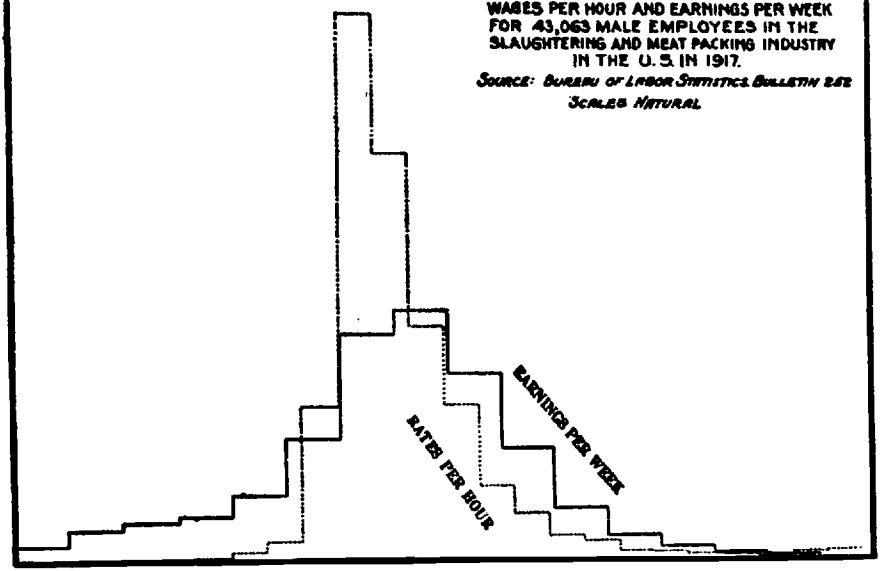
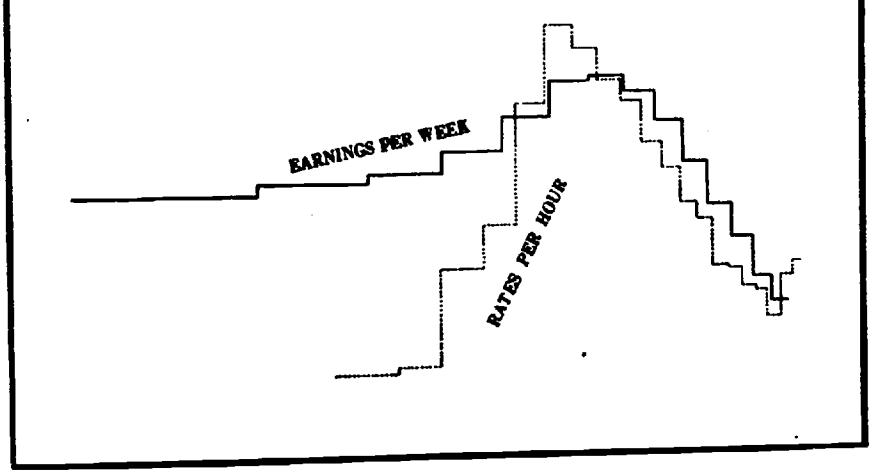


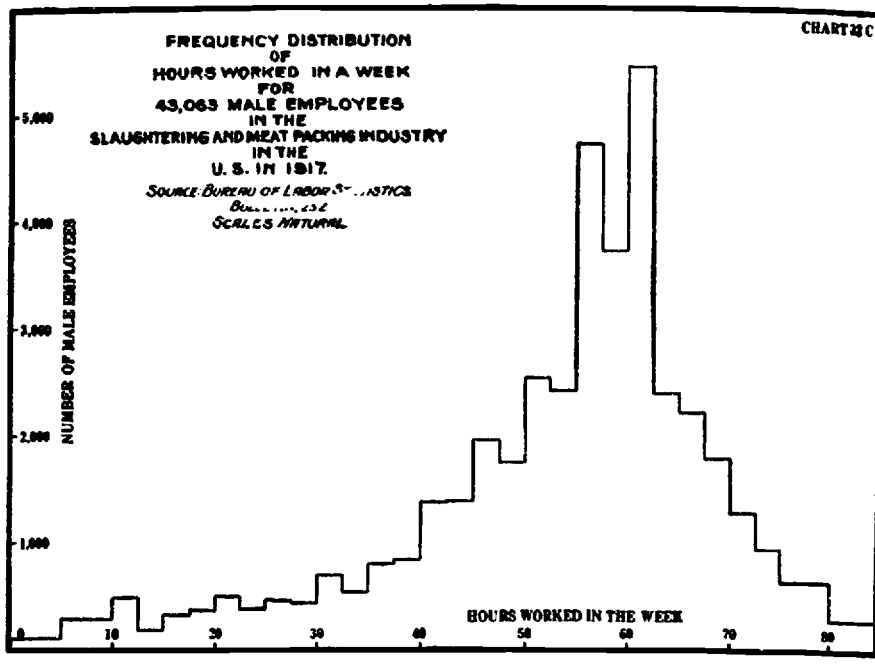
CHART 32B

FREQUENCY DISTRIBUTIONS OF RATES OF WAGES PER HOUR AND EARNINGS PER WEEK FOR 43,063 MALE EMPLOYEES IN THE SLAUGHTERING AND MEAT PACKING INDUSTRY IN THE U. S. IN 1917.

SOURCE: BUREAU OF LABOR STATISTICS BULLETIN 252

SCALES LOGARITHMIC.





The same difficulty as we find in any attempt to estimate the distribution of earnings per week from the distribution of rates per hour seems inherent in any attempt to estimate the distribution of earnings in a year from the distribution of earnings in a week. The unknown distribution of weeks worked in the year must seriously affect our results.[1]

Estimating the frequency distribution of wages earned in a year for an industry from the frequency distribution of wages earned in another year in the same industry, if we had such data, would involve us in a similar difficulty. Even though we knew the total number of individuals gainfully employed and their total wage bill each year and also the frequency distribution of earnings for one of the years, estimating the frequency distribution for the other year would be hazardous. While some *rates* distributions for the same industry in the same locality show symptoms of not changing in shape very radically from year to year,[2] this does not seem

the-week (month or year) in different industries. Illustrations of lack of uniformity in the relation between rates and earnings of the same persons for the same period but in different industries were worked up from Professor Davis R. Dewey's Special Report on Employees and Wages for the 12th Census.

[1] We have no distributions of amounts earned in a week and in a year for the same industry, with which to illustrate this point directly.

[2] For example, the distribution curve for wages per week among Massachusetts factory workers shows a moderate degree of similarity of shape from year to year.

Professor H. L. Moore (*Political Science Quarterly*, vol. XXII, pp. 61–73) discussed the fluctuation from 1890 to 1900 in the variability of wage rates in a total made up of thirty

a sufficient reason for assuming the same of *earnings* distributions. The shape of the distribution representing hours or days worked in the year may be expected to change greatly from year to year with alternations of prosperity and depression.[1]

What little evidence we possess suggests that wage distributions [2] for individuals of the same sex in the same industry at the same date, but in different *localities*, though generally more dissimilar in shape than distributions for the same industry in the same place but at different *dates*, are less unlike one another than distributions for different *industries* though in the same place and at the same time. The variation in shape of such distributions for different industries is often extreme.[3]

selected manufacturing industries. These distributions (for 1890 and 1900) illustrate both the similarity and the difference in *rates* distributions between the two years.

[1] For example, what little information we have points to the "scatter" of the days-worked-in-a-year distribution being much greater in a year of depression than in a year of prosperity. The extreme variations in shape of the *income* distributions for the same 1240 individuals in the years 1914 to 1919 as seen in the *Statistics of Income*, 1919, page 30, are interesting in this connection.

[2] Whether earnings or rates.

[3] Examples of this are numerous. Charts 32D and 32E show the distribution of wages per week among Massachusetts males working in (a) the boot and shoe industry and (b) the paper and wood pulp industry. For purposes of comparison the two distributions are so placed on the natural scale chart that the frequency curves show the same arithmetic means and areas. The double log chart is based directly upon the natural scale chart. It was necessary to break up the "over $35" interval before calculating the arithmetic means.

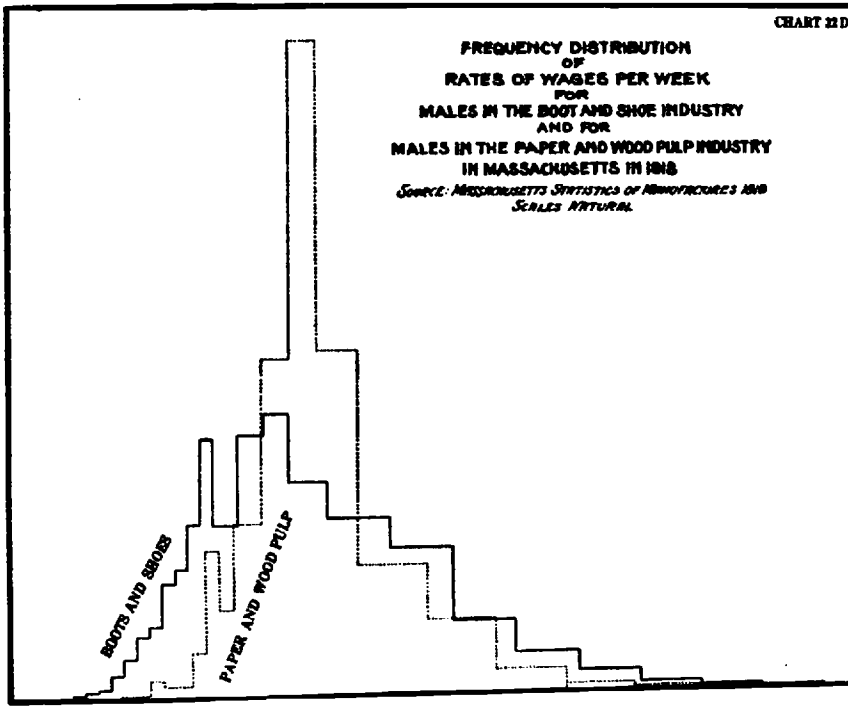
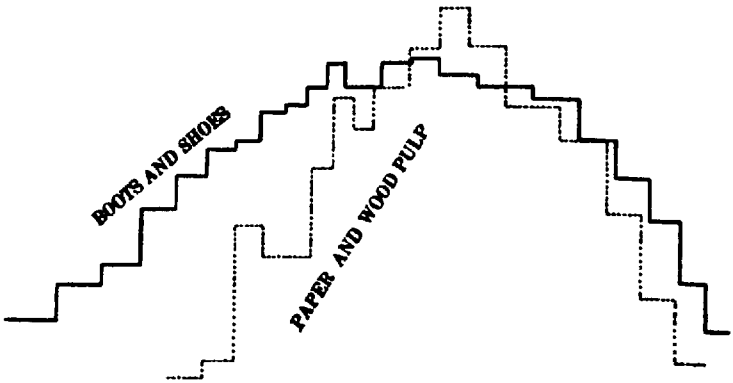


CHART 32E

FREQUENCY DISTRIBUTION OF RATES OF WAGES PER WEEK FOR MALES IN THE BOOT AND SHOE INDUSTRY AND FOR MALES IN THE PAPER AND WOOD PULP INDUSTRY IN MASSACHUSETTS IN 1918.

SOURCE. MASSACHUSETTS STATISTICS OF MANUFACTURES 1918
SCALES LOGARITHMIC



In conclusion, the order of importance of the variables as affecting the shape of the distribution curve seems to be—industry, place, time.

We have but little basis for estimating total income from earnings. In the preceding chapter on Income Distributions from other Sources than Income Tax Returns attention was drawn to the difficulty of arriving at any reliable statement of relationship between earnings and income from such distributions because of the way in which the data were selected. It is even less possible to discover the nature of any such relationship from the income-tax material. Though there is no such apparent "selection" in the income-tax data as in the case of non-tax income distributions, the material is not arranged to answer our particular question.

The non-statistical reader examining Charts 30D, 30E and 30F, on which are plotted average total income and average income from wages in each income interval, might think that it would be quite simple to estimate the probable average total income of persons having any specified wage. However there is a profound statistical fallacy involved in the use of this material for any such purpose. As given in the official tables, income is the independent variable, wages the dependent. This condition cannot be reversed without retabulation of the original returns. The statistical student recognizes the problem as one involving the impossibility of deriving one regression line from the other when neither the nature of the

equation representing the regression line [1] nor the degree of relationship (correlation in the broad, non-linear sense) is known. Even if we knew that the average *net income* of those persons reporting in 1918 in the $5,000 to $6,000 net income class was $5,474 and the average *wage* obtained by these persons was $2,192, we would be quite unwarranted in concluding that the average *income* of persons receiving $2,192 per annum *wages* was $5,474. If no wage earner received income from any other source than wages we still would have a condition where the average *income* in the income class would be greater than the average *wage*. Total wages would be necessarily less than total income, because in the income class are included not only wage earners but capitalists and entrepreneurs. But both total wages and total income are divided by the same number to get an average—namely total number of persons in that income class.

This suggests a technical criticism of the material contained in the *Statistics of Income*. All data concerning the relation between two variables are always there published in such a manner as to give information concerning only one of the regression lines and no information whatever concerning the "scatter." If such data were published in the form of "correlation tables" the increase in usefulness for statistical analysis would be very great. Such "correlation tables" keep closer to the original data than the usual type of statistical tables. Freer use of them is much to be desired, particularly in cases where it is difficult to anticipate all the problems for whose solutions investigators will go to the tabulated materials.

[1] The difficulty of the problem is, if possible, increased in this particular case because of the fact that the regression is radically non-linear.