Notes and Brief Reports

Estimating Distributions of Workers and Taxable Wages Under OASDHI*

What taxable maximum under the old-age, survivors, disability, and health insurance program would be needed in 1970 to cover fully the earnings of the same proportion of wage and salary workers as the proportion in 1938 with all their wages taxed? How high would the taxable maximum have to be in 1975 to reach the same percentage of total wages and salaries as that in 1959 under the $4,800 maximum? What would the revenue under the Federal Insurance Contributions Act for 1971–75 amount to, if 95 percent of the wage and salary workers had their earnings fully covered and the combined tax rate for employers and employees were raised to 14 percent?

The Social Security Administration needs reliable projective information to find answers to questions like those in the preceding paragraph about wage and salary workers covered under OASDHI in the future. The needed information can be arrived at through the joint efforts of two divisions of the Office of Research and Statistics—the Division of Economic and Long-Range Studies and the Division of Statistics.

Projections of the total number of wage and salary workers and of aggregate wages to be covered under the program are developed by economists in the Division of Economic and Long-Range Studies. Deriving distributions and related estimates based on those aggregates is the responsibility of the Division of Statistics. This note describes the various techniques used to develop the following estimates, essential for studying the effects of the existing taxable maximum and the potential effects of alternative maximums:

(1) The number of workers whose wages and salaries are fully covered and the proportion of all covered workers that these workers represent;

(2) the amount of taxable wages and salaries and the proportion of total wages and salaries that this amount represents.

---

* Prepared by Gilda J. Garrett, Division of Statistics, Office of Research and Statistics.

DEVELOPMENT OF WORKER DISTRIBUTION

Ratio Method

Obtaining the estimates needed calls for development of an earnings distribution of workers based on projections for the specific year. The desired distribution is developed by assuming that the ratio of the average total wages and salaries for the future year and the year preceding it reflects a shift of workers in the income distribution as well as a shift in the level of earnings. If, for example, the average for total wages and salaries in 1968 is 1.05 times the 1967 average, it is assumed that (1) workers earnings at least $600 in 1968 would have earned at least $571 in 1967 and (2) the percentage of workers who did earn at least $571 was the same as the proportion with at least $600 in 1968.

Sample data for past years indicate that the rate of increase in earnings declines as one progresses in the income array. However, the "and over" frequency distribution of workers by $600 earnings intervals up to the existing taxable maximums is developed on the assumption that the ratio for each "and over" dollar amount would be equal to the aggregate ratio at most. It is felt that any improvement in this part of the distribution derived from adjusting the aggregate ratio would be small. Nevertheless, the ratio must be adjusted to obtain the "and over" frequencies at three of the dollar amounts above the maximum (one and one-third, two, and four times the maximum). Because these frequencies are relatively high in the income array, failure to reflect the decline in the ratio would result in too much distortion.

Polynomial Method

The "and over" distribution of workers between the maximum and four times the maximum is assumed to be of the form $Y = 1/(A+Bx+Cx^2+Dx^3)$, with Y workers earning at least x dollars. An exact fit of the equation through the "and over" frequencies at the existing taxable maximum and at the three amounts above the maximum yields the curve for a particular distribution.

ESTIMATING TOTAL WAGES AND SALARIES

Wages and salaries below the existing taxable maximum are calculated, interval by interval, by multiplying the frequencies between each "and over" interval by an amount a little less than the midpoint of the interval.

Income between the maximum and open-end tail is computed by 100 intervals based on the equation $x(Y) = \int^b_a xdy$, where $x(Y)$ indicates earnings of x dollars for Y workers in the interval (a,b). The distribution of workers in the open-end tail—workers with wages at least four times the maximum—is assumed to conform to the Pareto distribution $Y = \frac{k}{\alpha - 1} x^{1-\alpha}$ indicating Y workers earning at least x dollars. It can be shown that average wages and salaries for this group is x dollars times $((\alpha-1)/(\alpha-2))$. Thus, the estimate of aggregate earnings for Y workers who earned at least four times the maximum, say x_4, is simply Y_4 times the average earnings computed on the basis of the Pareto assumption.

ESTIMATING TAXABLE WAGES

The amount of taxable wages and salaries up to the taxable maximum is assumed to be equal to the total earnings up to that maximum. However, the average taxable wages of Y workers who earned at least x, the maximum amount, is assumed to be greater than the maximum because of multi-employment. Taxable wages for workers who earned at least the maximum are therefore equal to R_9xY where R_9 is an adjustment factor to reflect taxable multi-employment earnings. The total amount of taxable wages and salaries for the maximum is then $T_9 + R_9xY$, where T_9 represents the taxable earnings up to the maximum.

SUMMARY

The "and over" frequency distribution of wage and salary workers, by $600 intervals up to the maximum and three broad intervals above the maximum, is developed on the computer by the ratio method. Estimates of workers who earned at least the maximum amount, the percentage of workers whose earnings are fully covered, and the amount and percentage of the total amount that is taxable under the maximum are also generated by the computer, for the existing taxable maximum as well as alternative maximums above the existing maximum. Before any details are generated above the existing taxable maximum, however, a consistency check is made by the computer. The distribution, determined by the four points obtained from the ratio method, is tested to see if:

(a) The aggregate wage and salary income for the distribution is within ½ of 1 percent of the control estimate;

(b) the Pareto average for workers in the open-end tail of the distribution is within 1 percent of a specified average.

If the test is negative, the original input points are adjusted by the computer via the Newton-Raphson Method[1] until the ending conditions are met.

Technical Note

Computing "and over" frequency distribution of workers (ratio method):

The ratio method is best explained by an example.

Given:
(a) Aggregate workers, 1968 = 85,200,000
(b) Average wages and salaries, 1968 = $4,837
(c) Average wages and salaries, 1967 = $4,563
(d) "And over" % distribution of workers, 1967:

Interval	*Workers* (percent of total)
Total	100.000
$600 and over	85.219
1,200 and over	76.119
1,800 and over	68.406
2,400 and over	61.995
3,000 and over	55.633
3,600 and over	49.732
4,200 and over	43.990
4,800 and over	38.431
5,400 and over	33.273
6,000 and over	28.735
6,600 and over	24.331
7,200 and over	20.460
7,800 and over	17.019
10,400 and over	6.911
15,600 and over	1.962
31,200 and over	.394

Find: A comparable "and over" frequency distribution for 1968.

---

[1] See Kaj L. Nielsen, *Methods in Numerical Analysis* (2d Edition), Macmillan, 1956 and 1964, pages 205–215.

Solution: Ratio of average earnings $= R$

$$\frac{\text{1968 average}}{\text{1967 average}} = \frac{\$4{,}837}{\$4{,}563} = 1.060$$

1967 interval equivalent to

$$\text{1968 interval} = I' = \frac{I}{R}$$

$$= \frac{\text{1968 "and over" interval amount}}{R}$$

$$= \frac{\$600+}{1.060} = \$566+$$

Let $P(x)$ = percentage of workers with x in 1968
Let $P'(x)$ = percentage of workers with x in 1967

$$\begin{aligned} P(I) &= P'(I') \\ &= \{(I-I')/[I-(I-\$600)]\}[P'(I-\$600)-P'(I)]+P'(I) \\ &= [(\$600-\$566)/\$600](100.000-85.219)+85.219 \\ &= [(34/600)14.781]+85.219 = 0.838+85.219 \\ &= 86.057 \end{aligned}$$

Therefore in 1968, 86.057 percent of all workers is assumed to have earned at least $600.

Computations for 1968 distribution are summarized below:

I	R	I'	$P'(I)$	$P'(I') = P(I)$	
1968 intervals	Ratio of average earnings	1967 equivalent interval	Percent of all workers		1968 frequency distribution (in thousands)
			1967	1968	
Total	$R = 1.060$		100.000	100.000	85,200
Less than $600		Less than $566	14.781	13.943	11,879
600 and over		566 and over	85.219	86.057	73,321
1,200 and over		1,132 and over	76.119	77.150	65,732
1,800 and over		1,698 and over	68.406	69.717	59,399
2,400 and over		2,264 and over	61.995	63.448	54,058
3,000 and over		2,830 and over	55.633	57.436	48,935
3,600 and over		3,396 and over	49.732	51.738	44,081
4,200 and over		3,962 and over	43.990	46.268	39,420
4,800 and over		4,528 and over	38.431	40.951	34,890
5,400 and over		5,094 and over	33.273	35.904	30,590
6,000 and over		5,660 and over	28.735	31.307	26,674
6,600 and over		6,226 and over	24.331	27.076	23,068
7,200 and over		6,792 and over	20.460	23.092	19,674
7,800 and over		7,358 and over	17.019	19.554	16,660
10,400 and over	.61R	10,033 and over	6.911	8.338	7,104
15,600 and over	.50R	15,146 and over	1.962	2.394	2,040
31,200 and over	.28R	30,703 and over	0.394	0.446	380

Fitting cubic equation (polynomial method):

We have to fit
$Y = 1/(A+Bx+Cx^2+Dx^3)$
to our estimates of the number of workers earning at least the amount of four specified income levels. Our data is of the form indicated below.

Income level (in thousands)	Workers
$x_1 = K$ = Existing maximum	Y_1
$x_2 = 1.333K$	Y_2
$x_3 = 2K$	Y_3
$x_4 = 4K$	Y_4

First we consider

$$\frac{1}{Y} = Y' = A + Bx + Cx^2 + Dx^3$$

Then we substitute our data in the equation and obtain four linear equations in four unknowns:

$$\begin{aligned} Y'_1 &= A+Bx_1+Cx_1^2+Dx_1^3 \\ Y'_2 &= A+Bx_2+Cx_2^2+Dx_2^3 \\ Y'_3 &= A+Bx_3+Cx_3^2+Dx_3^3 \\ Y'_4 &= A+Bx_4+Cx_4^2+Dx_4^3 \end{aligned}$$

The solution of this system of equations is straightforward, and the results are summarized below.

$$\begin{aligned} D &= (-16/Y_1+27/Y_2-12/Y_3+1/Y_4)/(16K^2) \\ C &= (2/Y_1-3/Y_2+1/Y_3-K_6D)/(2K^2/3) \\ B &= (-1/Y_1+1/Y_2-K_7C-K_9D)/(K/3) \\ A &= 1/Y_1-KB-K^2C-K^3D \end{aligned}$$

where:

$$\begin{aligned} K_6 &= 2.889K^3 \\ K_7 &= .778K^2 \\ K_9 &= 1.37037K^3 \end{aligned}$$

Then, by substituting any value x between x_1 and x_4 into our equation we can obtain an estimate of the number of workers earning at least that amount.

Computing earnings between the maximum and four times the maximum:

We have to integrate

$$x\,dy = \frac{(Bx+2Cx^2+3Dx^3)dx}{(A+Bx+Cx^2+Dx^3)^2}$$

Since the $\int_a^b x\,dy$ cannot be calculated by conventional methods, we approximate the value of our definite integral, using the Gauss formula for numerical integration[2]

$$I = \frac{b-a}{2}\, G_i \psi(V_i),$$

where G_i and $\psi(V_i)$ are based on the function to be evaluated. The principle of Gauss's formula, based on Legendre polynomials, is to obtain the best subdivision of the interval (a,b), the value of the function at these points, and the coefficients to multiply the functional values to yield the value of the definite integral.

Estimating earnings in the open-end tail of the distribution (Pareto method):

The assumed distribution function is

$$Y = \frac{k}{\alpha-1} x^{1-\alpha}$$

In logarithmic notation, we have
$\log Y = \log k - \log(\alpha-1) + (1-\alpha)\log x$

where k and α are unknown constants, using the following notation:

Income level (in thousands)	Workers
$x_3 = 2K$	Y_3
$x_4 = 4K$	Y_4

where K is the existing maximum.

Substituting the above values in our log equation, we have

$$\begin{aligned} \log Y_3 &= \log k - \log(\alpha-1) + (1-\alpha)\log x_3 \\ \log Y_4 &= \log k - \log(\alpha-1) + (1-\alpha)\log x_4 \end{aligned}$$

Subtracting, we have

$$\begin{aligned} \log Y_3 - \log Y_4 &= (1-\alpha)\log x_3 - (1-\alpha)\log x_4 \\ \log (Y_3/Y_4) &= \log (x_3/x_4) + \alpha \log (x_4/x_3) \end{aligned}$$

$$\alpha = \frac{\log(Y_3/Y_4)}{\log(x_4/x_3)} + 1 = \frac{\log(Y_3/Y_4)}{\log(4K/2K)} + 1 = \frac{\log(Y_3/Y_4)}{\log 2} + 1$$

[2] *Ibid.*, pages 129–132.

We know that total earnings for workers in this part of the distribution is

$$T(y)=\int_x^\infty x\,dy = \frac{k}{\alpha-2}x^{2-\alpha}$$

Then the average wage paid to workers with at least x_4 dollars is

$$\text{Average}=\frac{\text{Total wages}}{\text{Total workers}}=\frac{\frac{k}{\alpha-2}x_4{}^{2-\alpha}}{\frac{k}{\alpha-1}x_4{}^{1-\alpha}}=\frac{\alpha-1}{\alpha-2}x_4$$

Note, then, that we may calculate the average wage using only known quantities.

Then the total wages for Y_4 workers may be expressed as:

$$\text{Total wages}=Y_4\frac{(\alpha-1)}{(\alpha-2)}x_4$$

Federal Coal Mine Health and Safety Act of 1969*

On December 30, 1969, President Nixon signed the Federal Coal Mine Health and Safety Act of 1969 (P.L. 91–173). The law is primarily designed to establish nationwide health and safety standards for the coal-mining industry. It also includes an income-maintenance provision that is of unusual interest since it gives the Federal Government a temporary responsibility in the area of workmen's compensation. Under title IV of the new law, monthly cash benefits are provided for coal miners who are "totally disabled" because of pneumoconiosis ("black lung" disease) and for their dependents and survivors.

Two Federal agencies—the Department of Health, Education, and Welfare and the Department of Labor—will be involved in administering the cash benefit provisions. As the result of modifications made by the House-Senate conference committee in the bills originally passed by each House, the Department of Health, Education, and Welfare (through the Social Security Administration) will be responsible for the payment and administration of benefit claims filed before January 1, 1973. The Department of Labor will have the responsibility for claims filed after December 31, 1972.

The monthly benefits payable by the Social Security Administration to a miner disabled by pneumoconiosis or to the widow of a miner who died with the disease will be a flat amount—about \$136 at the present. (The amount of the benefit will be equal to 50 percent of the minimum monthly payment to which a totally disabled Federal Government employee in the first step of grade GS–2 would be entitled under the Federal Employees' Compensation Act.) For one dependent (wife or child[1]) an additional 50 percent of the miner's benefit will be payable and the total payment will thus be about \$204. For two dependents, the additional amount will be 75 percent of the benefit (a total of \$238), and for three or more it will be 100 percent (a total of \$272). If the deceased miner does not leave an eligible widow, however, no survivor benefits are payable (even when there are surviving children who were paid the supplemental benefits while the miner was alive).

Benefit payments to a miner or his widow will be reduced if the beneficiary is also receiving payments under the workmen's compensation, unemployment insurance, or disability insurance[2] programs of a State on account of the disability of the miner. Benefits paid to miners (but not widows) will also be subject to an earnings test, the provisions of which will be the same as the retirement test provisions under the Social Security Act. The law specifies that benefit payments will not be subject to Federal income tax.

To be eligible for the benefits paid by the Social Security Administration, the disabled miner must file his claim before January 1, 1973. A widow's claim must be filed within 6 months after the death of her husband or by December 31, 1972, whichever is later. Benefits are payable to a widow of a miner if he was receiving "black lung" benefits before his death or if he died from the disease.

In addition, the program provides that if a claim is filed after December 31, 1971, but before January 1, 1973, the claimant can receive benefits from the Social Security Administration only through December 31, 1972. The benefits will con-

* Prepared in the Interprogram Studies Branch, Division of Economic and Long-Range Studies, Office of Research and Statistics.

[1] The definition of child follows that in the Federal Employees' Compensation Act. Thus, benefits are payable to an unmarried child who is under age 18, totally disabled, or a full-time student under age 23.

[2] Temporary disability insurance laws are in operation in California, Hawaii, New Jersey, New York, Rhode Island, and Puerto Rico.