USES OF TAX FILES COMBINED WITH FIELD SURVEYS

BENJAMIN A. OKNER AND JOSEPH A. PECHMAN

The Brookings Institution*

For a long time, income-distribution analysts have been hampered in their research because the available data have not represented the entire income-receiving population or have failed to include all of the income known to have been received by that population. On the one hand, distributions of income based on tax returns omit persons who do not file; on the other, distributions based on field surveys, which provide demographic and other data not available on tax returns, omit a large fraction of total income because of underreporting by respondents. With the advent of the computer, it is now possible to combine the best information on tax files and field surveys so that the two sources can be used together for research purposes. The purpose of this paper is to describe the methods we at Brookings have used to merge the information in two such files for the calendar year 1966, to report briefly on the distribution of income that emerged, and to outline our plans for future research on the basis of the MERGE data file.

In creating the MERGE File, we combined information on 30,000 families and single persons included in the 1967 Survey of Economic Opportunity (SEO) conducted by the U.S. Census Bureau for the Office of Economic Opportunity, and a file containing information from 90,000 U.S. federal individual income-tax returns. Thus, the MERGE File contains data for low-income SEO families who are not in the tax-filing population, as well as the more complete—and, we believe, more accurate—income tax information for higher-income individuals. In addition, we corrected the income information in the MERGE File for non-reporting and underreporting, so that—with the appropriate weights applied to the sample units—the file accounts for the total income (on almost any desired definition of income) estimated to have been received in the United States in 1966.

The most important characteristic of the file is that calculations can be made on the basis of individual records at great speed and with a high degree of accuracy. Moreover, it is no longer necessary to make assumptions regarding the average characteristics of an entire income class or population cell in a cross-classification. The availability of information for individual families permits us to provide answers about a much wider variety of economic and social questions than has been possible heretofore.

*This study was financed under a research grant to the Brookings Institution from the U.S. Office of Economic Opportunity. The views presented in this paper are those of the authors and not necessarily those of the officers, trustees, or other staff members of the Brookings Institution, or of the Office of Economic Opportunity. All programming and computer operations described in the paper were performed at the Brookings' Social Science Computation Center. Stephen W. Kidd and Robert Wallace were responsible for the computer programming and we gratefully acknowledge their efforts on our behalf.

Creating the MERGE Data File[1]

Since the SEO income reporting units are a sample of the entire U.S. population and the returns in the Tax File are a sample of only the tax-filing population, we based the final MERGE File on the demographic information for the families in the SEO File.[2] However, we substituted the income data in the Tax File for the corresponding information in the SEO File to take advantage of the superior income reporting on tax returns (including the information on capital gains that is excluded from the SEO-Census income concept). This was done by first estimating (on the basis of reported SEO information) the kind of tax return or returns that would have been filed by each family and, then, for tax-filers, by matching each "SEO tax unit" with a tax return selected from the Tax File.

The ideal method of matching the SEO data with the tax data would have been to obtain the tax information directly from the Internal Revenue Service. But this was not practical because neither the Census Bureau nor the Internal Revenue Service permits others to use their files, even for statistical purposes. In place of an exact one-to-one match, a less satisfactory—but feasible—means of simulating a match was developed. In effect, we randomly selected from the Tax File a return "similar" to the SEO return and then substituted the income data in the tax record for the information in the SEO record. Since close to 30,000 matches had to be made, the selection and linking of returns in the SEO and Tax Files was performed on a computer.[3]

For most families, the final MERGE File contains the demographic data and information on receipts of nontaxable income from the SEO File plus taxable income figures from the return or returns assigned to it from the Tax File. For SEO units deemed to be nonfilers, the MERGE File includes no tax return information. Since there are very few high-income units in the SEO File, the upper "tail" of the Tax File (returns with incomes above $30,000) was substituted *in toto* for the SEO tail. For this group, which represents less than 2 percent of the entire population, the MERGE File does not contain any SEO demographic data.

The basic definition of income in the MERGE File is adjusted family income (AFI),[4] a concept which was developed for the tax-burden study described below. The basic data for estimating AFI were obtained from the Office of Business Economics (OBE) personal income accounts, individual income-tax information

[1] For a detailed description of the methods described in this section, see Benjamin A. Okner, "Constructing a New Data Base From Existing Microdata Sets: The 1966 MERGE File," *Annals of Economic and Social Measurement*, Vol. 1 (July 1972).

[2] In this paper, the term "families" refers to both unrelated individuals (one-person families) and the conventional Census family consisting of two or more persons, related by blood, marriage, or adoption.

[3] The characteristics used to link the two files were (1) marital status, (2) age of head of the unit, (3) number of dependents, (4) pattern of income, and (5) major and minor sources of income. The basic rule was to match a SEO unit with a tax unit having the same characteristics and major source income within 2 percent of the major source income reported in the SEO survey.

[4] "Adjusted family income" as used in this paper corresponds to an augmented national income concept. Since this paper was completed the authors have decided that an income concept corresponding to augmented net national product is more appropriate for measuring effective tax burdens. In later work, therefore, the concept referred to here as "adjusted family income" is renamed "family income." "Adjusted family income" in subsequent work is equal to family income plus indirect business taxes.

from the Internal Revenue Service, and other government records, which were adjusted—where necessary—to take account of differences in income concept and of population covered. The AFI concept is intended to correspond as closely as practicable to an economic concept of income, i.e., it is equal to consumption plus tax payments plus (or minus) the net increase (or decrease) in the value of assets during the year. AFI includes only income which accrues *directly* to individuals and families; as a consequence, it does not include the income of fiduciaries and other recipients not represented in the SEO population.[5]

After substituting tax return data for the SEO income data, the total income accounted for by units in the MERGE File amounted to 93 percent of the AFI computed for 1966 (see Table 1). The next step in creating the MERGE File involved adjusting the SEO and Tax File income data to correspond with national aggregates. As Table 1 indicates, the aggregates for wages and salaries were very

TABLE 1

COMPARISON BETWEEN ADJUSTED FAMILY INCOME AND MERGE FILE INCOME BEFORE ADJUSTMENT, BY SOURCE OF INCOME, 1966

[dollar figures in billions]

Source of Income	Adjusted Family Income (1)	MERGE File Income[1] (2)	Difference (3)=(1)−(2)	MERGE File Income as Percent of Adjusted Family Income (4)=(2)÷(1)
Wages, salaries, and other labor	$423	$415	$ 8	98%
Nonfarm proprietors	43	46	−3	107
Farm proprietors	14	6	8	43
Rents and royalties	20	16	4	80
Personal interest	24	21	3	88
Corporate earnings	64[2]	60	4	94
Transfer payments	34	25	9	74
Accrued capital gains on inventories, farm assets, and nonfarm real estate	37	27	10	73
Total	$660	$616	$33	93%

Note: Details may not add to totals because of rounding.

[1] MERGE File income excludes adjustments for nonreporting and underreporting of income.

[2] Includes corporation income tax and undistributed profits.

close. On the other hand, reported farm proprietors' income was only 43 percent of the expected AFI amount, and there were less serious, but significant, discrepancies between the expected and reported amounts of interest, rent, and transfer payments. Some of the discrepancies were due to the partial coverage of the Census money-income concept, which was used in the field survey; the remainder was due to nonreporting and underreporting of income by respondents.

Although nonreporting and underreporting are conceptually separable, in practice it is difficult to distinguish these two types of response errors. On the

[5] For a detailed description of how the AFI figures were derived, see Benjamin A. Okner, "Adjusted Family Income: Concept and Derivation," Brookings Technical Working Paper II, for the Distribution of Federal, State, and Local Taxes Research Program, March 1971 (revised, mimeographed), which is available on request.

basis of data from other sources, we believe that most of the differences between the reported and AFI aggregate factor-payment amounts resulted from underreporting, while transfer payments were understated primarily because of nonreporting.

For income components where we believed the discrepancies were due to underreporting, the MERGE File data were adjusted to the AFI aggregates on the assumption that the underreporting was not related to other characteristics of the survey unit. A single ratio was therefore applied to the reported incomes of all units to increase them to the aggregate adjusted family income amounts. In the case of nonreporting, we imputed missing amounts stochastically to MERGE File units, based on various other characteristics of the survey units.

In addition to the adjustments for underreporting and nonreporting, several imputations were made to add information to the MERGE File which was not available—because it was not collected—in either the SEO or the Tax Files. These included imputed rent on owner-occupied homes, employer supplements to wage and salary income, tax-exempt interest on state and local bonds, and accrued capital gains on assets.[6]

The final MERGE File records each contain the original demographic, income (corrected for underreporting), employment, education, and other data derived from the SEO[7] plus one or more tax segments containing the income tax data for these families. We have also prepared a 10 percent sample of the file for use in rapidly checking out computer programs and estimation techniques. The complete file and the sample are stored on disc packs for rapid calculations on the Brookings' PDP-10 computer, but they are also available on magnetic tape. The time required to obtain a simple tabulation of several characteristics of families classified by, say, 30 income classes is roughly ten minutes on the sample and one hour on the complete file. Using our "tax calculator program," calculations of federal tax liabilities under the present tax law, or under several variants, can be completed in less than two hours on the entire file. We have efficient cross-tabulation programs available for use on the file and a recently written output package that provides us with a high degree of flexibility for printing tables in virtually free-form format. In addition, we have a Calcomp 565 digital plotter and, with the software developed for its use, we have the ability to produce graphic as well as tabular displays of our results.

Distribution of Income

Before proceeding to income distributions derived from the current MERGE File, we think it would be useful to describe more fully the relationship between the SEO-Census and adjusted family-income concepts.

[6] Imputed rent was allocated on the basis of the equity in owner-occupied homes reported by respondents. Wage supplements were based on the occupational, industrial, and wage characteristics reported by the survey units. State-local bond interest was based on the distribution of state-local bond ownership from the Federal Reserve Board's 1963 Survey of Financial Characteristics. Accrued gains on assets were based largely on realized capital gains and property income reported on tax returns. Details concerning these imputations are reported in Benjamin A. Okner, "The Imputation of Missing Income Information," Brookings Technical Working Paper III, for the Distribution of Federal, State, and Local Taxes Research Program, April 1971 (mimeographed), which is available on request.

[7] Even though they contain very limited data inferred from the tax return, the MERGE File does contain a demographic record segment for each high-income "upper-tail" tax return.

SEO-Census money income is essentially a total money receipts concept (except that receipts from the sales of capital assets are excluded). AFI is an accrued income concept. Therefore, in order to go from SEO-Census to adjusted family income, it is necessary to: (1) subtract money receipts that do not represent current income; and (2) add income not counted as current receipts by the Census.

The derivation involves the following steps:

(figures in billions of dollars)		
SEO-Census money receipts		524
Less:		
Federal government pensions	4	
State and local government pensions	2	
Veterans' life insurance	1	
Subtotal		−7
Plus:		
Employee wage supplements	40	
Net imputed rent	12	
Imputed interest	6	
Retained corporate profits	22	
Corporate income tax	26	
Accrued capital gains on inventories, farm assets, and nonfarm real estate	37	
Subtotal		143
Equals adjusted family income		660

In essence, adjusted family income is equal to national income (as defined in the national income accounts)[8] plus transfer payments plus accrued gains on farm assets and nonfarm real estate. In keeping with the national income concept, AFI includes corporation incomes before tax. The portion of corporate income distributed as dividends is included in money receipts and is not shown separately in the derivation above. However, undistributed profits and corporation tax liability must be added to income to derive AFI. This procedure has the advantage not only of consistency but also of providing a complete account of the accrued income claims of the household sector. Retained earnings of corporations, which are thus automatically included in adjusted family income, may be regarded as an approximation of accrued capital gains on corporate stock during the year.[9]

[8] The only departure from the official definition of income is the omission of interest imputed to individuals for the services rendered to them by the banking system.

[9] We used this approximation because the annual fluctuations in the value of corporate stock are high and even three-to-five-year averages may not give an adequate representation of accrued capital gains. Martin J. Bailey and Martin David have shown that over very long periods, capital gains on corporate securities are roughly equal to retained earnings. See Martin J. Bailey, "Capital Gains and Income Taxation" in Arnold C. Harberger and Martin J. Bailey, eds., *The Taxation of Income from Capital*, Brookings Institution, 1969, pp. 15–26; and Martin David, *Alternative Approaches to Capital Gains Taxation*, Brookings Institution, 1968, pp. 242–246.

Even after substituting tax return data for the income reported by the SEO respondents, total SEO-Census money income in the MERGE File totaled only $489 billion, or about $35 billion less than the amount expected.[10] The adjustments made to correct for underreporting and nonreporting of income raised the median money income from its initial level of $7,508 to $8,592 after correction.

TABLE 2
COMPARISON OF SHARES OF SEO-CENSUS MONEY INCOME RECEIVED BY EACH FIFTH OF FAMILIES BEFORE AND AFTER ADJUSTMENT FOR NONREPORTING AND UNDERREPORTING OF INCOME

Families Ranked from Lowest to Highest	Before Adjustment		After Adjustment	
	Income Range (dollars)	Percent of Income Received	Income Range (dollars)	Percent of Income Received
Lowest fifth	Under 2,823	4.3	Under 3,261	3.4
Second fifth	2,823– 5,416	10.9	3,261– 6,057	10.7
Middle fifth	5,416– 7,878	17.4	6,057– 8,747	17.0
Fourth fifth	7,878–11,000	24.6	8,747–12,500	23.8
Highest fifth	11,000 and over	42.7	12,500 and over	45.1
Top 5 percent	16,922 and over	16.4	20,227 and over	19.1
Top 1 percent	28,333 and over	5.5	44,792 and over	6.8

In Table 2, we show the share of income received by each fifth of the families, when they are ranked from lowest to highest, before and after the income adjustments. Before correction, the lowest fifth of the families had incomes under $2,823 and received 4.3 percent of total income. The highest fifth of the families had incomes of $11,000 or more and received 42.7 percent of the total. After adjustment, the poorest fifth of the families had incomes under $3,261 and received 3.4 percent of the total; the highest fifth moved up to $12,500 and received 45.1 percent of total income.

Although the upward shift can be seen all along the income distribution, the effect is most pronounced among those at the very top. Before adjustment, the top 5 percent included families with incomes of $16,922 and over and they received 16.4 percent of total money income. After adjustment, the top 5 percent included families with incomes of $20,227 and over and this group received 19.1 percent of the total money income. The share of the total received by the top 1 percent of all families increased from 5.5 percent to 6.8 percent after adjustment.[11] This large change in the relative distribution of income mainly reflects the addition of high-income family units which were omitted from the original SEO population.

We now turn to the presentation of the MERGE data classified by still another income concept—money factor income (MFI). As shown in Table 3A, MFI is equal to the sum of wages, farm and nonfarm proprietors' income, rents

[10] This exceeds the $33 billion difference shown in Table 1 because of conceptual differences between items in the SEO-Census and adjusted family income concepts.

[11] It should be noted that we have chosen the SEO-Census money income concept for comparison purposes only because it is the most comprehensive one that is available on a before adjustment basis in the SEO. However, we have retained all the detailed income components in the MERGE File for maximum flexibility. Thus, the researcher is free to define income any way he wishes to suit his own particular needs.

TABLE 3A

DERIVATION OF 1966 TOTAL MONEY FACTOR INCOME IN THE FAMILY MERGE FILE, BY COMPONENT

[amounts in millions]

Money Factor Income (000's)	Wage & Salary Income	Nonfarm Proprietors' Income	Farm Proprietors' Income	Rent & Royalty Income	Dividend Income	Monetary Interest Income	Total Money Factor Income
Under 1	2,489.023	−1,135.015	−3,015.265	−1,679.803	330.265	957.598	−2,053.197
1–2	3,539.452	180.010	6.655	154.335	321.126	899.536	5,101.113
2–3	5,460.715	538.434	160.322	308.002	305.282	874.992	7,647.746
3–4	9,297.470	784.080	92.171	220.297	346.643	874.872	11,615.533
4–5	13,521.933	1,227.001	257.821	346.056	241.525	839.732	16,434.068
5–6	20,803.325	1,240.478	440.048	81.511	224.297	660.978	23,450.637
6–7	22,474.012	1,239.152	422.834	205.781	306.262	702.837	25,350.880
7–8	29,823.461	1,627.226	292.097	231.719	234.753	705.991	32,915.248
8–9	34,218.634	1,699.995	611.873	99.672	262.307	805.379	37,697.857
9–10	28,880.215	1,425.900	547.015	163.020	335.240	619.602	31,970.993
10–11	30,711.309	2,070.669	357.582	155.244	544.639	683.615	34,523.056
11–12	27,330.153	1,761.201	462.961	365.668	340.639	597.007	30,857.627
12–13	23,457.244	2,530.382	385.534	217.522	370.665	633.419	27,594.767
13–14	20,684.459	2,215.073	445.764	351.263	459.247	632.021	24,787.827
14–15	17,166.195	2,185.835	406.394	238.448	225.261	522.705	20,744.840
15–20	49,085.511	5,750.749	2,519.856	1,288.877	1,150.266	1,831.223	61,626.478
20–25	17,396.036	2,897.263	1,606.761	889.643	768.906	792.584	24,351.198
25–50	21,731.693	9,836.769	4,651.733	3,014.206	2,641.823	1,694.471	43,570.689
50–100	5,410.753	4,598.426	1,008.093	1,438.451	2,407.229	962.263	15,825.211
100–500	2,042.475	1,738.098	363.844	690.308	2,634.702	882.822	8,352.249
500–1,000	60.671	76.571	1.176	24.373	401.207	36.658	600.655
1,000+	20.916	42.582	−3.324	15.719	453.445	19.836	549.174
Total	385,605.660	44,530.879	12,021.945	8,820.311	15,305.729	17,230.140	483,514.650

TABLE 3B

NUMBER OF FAMILIES WITH 1966 TOTAL MONEY FACTOR INCOME IN THE FAMILY MERGE FILE, BY COMPONENT

[population in thousands]

Money Factor Income (000's)	Wage & Salary Income	Nonfarm Proprietors' Income	Farm Proprietors' Income	Rent & Royalty Income	Dividend Income	Monetary Interest Income	Total Money Factor Income
Under 1	2,457.957	511.323	430.889	616.838	678.343	3,435.113	9,357.429
1–2	2,382.836	330.967	214.146	544.397	586.357	1,701.577	3,447.173
2–3	2,379.826	354.548	307.496	393.006	468.173	1,409.837	3,074.655
3–4	2,851.253	435.630	164.838	363.616	461.004	1,411.806	3,316.938
4–5	3,178.775	549.791	186.630	336.985	387.038	1,592.963	3,645.549
5–6	3,990.247	493.874	174.154	266.499	502.543	1,929.612	4,253.833
6–7	3,668.725	411.803	150.254	258.433	505.253	2,050.490	3,886.884
7–8	4,174.483	496.773	158.357	330.647	652.752	2,395.625	4,386.948
8–9	4,273.991	479.613	170.591	345.919	753.630	2,610.897	4,442.586
9–10	3,241.476	335.722	127.362	271.748	604.806	2,113.139	3,364.848
10–11	3,136.919	428.016	226.588	283.802	788.207	2,192.650	3,285.386
11–12	2,590.504	333.780	284.356	305.400	665.965	1,932.792	2,685.208
12–13	2,059.316	346.676	219.937	224.526	607.219	1,524.591	2,206.159
13–14	1,717.743	331.400	211.707	177.300	645.901	1,360.911	1,837.235
14–15	1,352.595	311.139	192.464	187.700	489.588	1,120.242	1,432.035
15–20	3,344.962	700.826	509.311	565.174	1,419.797	2,843.125	3,607.209
20–25	991.349	264.326	187.384	288.932	615.657	933.868	1,108.326
25–50	1,043.833	538.435	303.546	350.025	826.737	1,136.851	1,330.302
50–100	157.164	133.689	48.593	97.118	196.791	227.472	238.876
100–500	35.190	31.548	11.684	25.146	49.659	53.785	54.633
500–1,000	0.529	0.598	0.223	0.502	0.878	0.893	0.900
1,000+	0.168	0.169	0.077	0.174	0.280	0.275	0.283
Total	49,029.842	7,820.648	4,280.586	6,233.886	11,906.578	33,978.514	60,963.395

TABLE 3C

AVERAGE AMOUNT OF 1966 TOTAL MONEY FACTOR INCOME PER RECIPIENT FAMILY IN THE FAMILY MERGE FILE, BY COMPONENT

Money Factor Income (000's)	Wage & Salary Income	Nonfarm Proprietors' Income	Farm Proprietors' Income	Rent & Royalty Income	Dividend Income	Monetary Interest Income	Total Money Factor Income
Under 1	1,013	−2,220	−6,998	−2,723	487	279	−219
1–2	1,485	544	31	283	548	529	1,480
2–3	2,295	1,519	521	784	652	621	2,487
3–4	3,261	1,800	559	606	752	620	3,502
4–5	4,254	2,232	1,381	1,027	624	527	4,508
5–6	5,214	2,512	2,527	306	446	343	5,513
6–7	6,126	3,009	2,814	796	606	343	6,522
7–8	7,144	3,276	1,845	701	360	295	7,503
8–9	8,006	3,545	3,587	288	348	308	8,486
9–10	8,910	4,247	4,295	600	554	293	9,501
10–11	9,790	4,838	1,578	547	691	312	10,508
11–12	10,550	5,277	1,628	1,197	511	309	11,492
12–13	11,391	7,299	1,753	969	610	415	12,508
13–14	12,042	6,684	2,106	1,981	711	464	13,492
14–15	12,691	7,025	2,112	1,270	460	467	14,486
15–20	14,674	8,206	4,948	2,280	810	644	17,084
20–25	17,548	10,961	8,575	3,079	1,249	849	21,971
25–50	20,819	18,269	15,325	8,611	3,195	1,490	32,752
50–100	34,427	34,397	20,746	14,811	12,232	4,230	66,249
100–500	58,042	55,094	31,140	27,452	53,056	16,414	152,879
500–1,000	114,690	128,044	5,272	48,552	456,955	41,050	667,394
1,000+	124,502	251,964	−43,172	90,337	1,619,446	72,130	1,940,543

and royalties, dividends, and monetary interest. About 80 percent of the $483.5 billion total is from wage and salary income; 11 percent is proprietors' income; and the remaining 9 percent of MFI is income from property.

The average amount of each component of MFI received by MERGE File families is shown in Table 3C. With but one exception, the average amount received of each component rises as income increases. The exception is farm proprietors' income where the average per recipient family rises with income over most of the income range, but then drops sharply at the very highest income levels. This is consistent with other findings and results from the large losses of very wealthy "hobby farmers."

The next group of tables illustrates the distribution of various employer supplements to wages and salaries. As shown in Table 4A, contributions for private pension and welfare funds and for social security account for $29.1 billion, or 73 percent of the $39.7 billion total. The average contributions for each wage supplement component are shown in Table 4C. The averages for social security are particularly interesting since the maximum employer (and employee) payment in 1966 was $377. Yet, the average for recipient families in all the income classes between $15,000 and $50,000 exceeds the $377 maximum. The reason for this is that in these classes, there are numerous families with more than one earner with wages subject to social security.

More than 63 percent of total transfer payment income is derived from social security benefits (Table 5A); these benefits are fairly evenly distributed among families all along the income scale. Although far smaller than social security in magnitude, the same fairly even distribution is found for veterans' disability payments and the work-related workmen's compensation and unemployment insurance benefits. While such payments do play a role in maintaining income for families for short periods of disability, their wide distribution over the entire income scale suggests that they do not play a major role in improving the lot of the very poor. On the other hand, we find that over 70 percent of all public assistance payments go to families with money factor income under $1,000. The average amount of public assistance received is about $1,000 at the very lowest income levels and falls to about $650 per recipient family at $9,000 to $10,000 of MFI (Table 5C).[12]

Finally, we show some of our preliminary tax-distribution results in Table 6. Federal personal income taxes were derived directly from the tax segments in the MERGE File; the federal payroll taxes are equal to the sum of employee and employer contributions for social security, unemployment insurance, and workmen's compensation. Summarizing very briefly, we find that: (1) transfers as a percent of total income before transfers start out greater than 100 percent (i.e., they exceed nontransfer income) and then drop sharply as a percent of income as income rises; (2) the effective income tax rate rises steadily with income (except at the very bottom and top of the income scale where the relationship is distorted

[12]Those who are familiar with the stringent requirements for receiving public assistance in the United States may wonder about the units in the $5,000 to $10,000 income range who are shown as benefit recipients. The number of such recipient families is quite small and misreporting could explain some of these cases. In addition, a large proportion of these anomalous cases result from conceptual differences between the SEO reporting unit and the public-assistance recipient unit.

TABLE 4A

DERIVATION OF 1966 TOTAL EMPLOYER WAGE SUPPLEMENT CONTRIBUTIONS IN THE FAMILY MERGE FILE, BY COMPONENT

[amounts in millions]

Money Factor Income (000's)	Private Pensions & Welfare Funds	Social Security	Workmen's Compensation	Unemployment Insurance	Civilian Govt. Retirement	Total Wage Supplements
Under 1	116.180	82.645	14.348	28.076	25.174	266.424
1–2	88.297	138.127	20.143	52.920	25.691	325.178
2–3	166.930	218.630	37.397	103.868	32.352	559.176
3–4	308.809	371.450	63.174	163.215	71.308	977.957
4–5	507.937	540.699	95.288	205.137	109.618	1,458.680
5–6	835.147	817.468	148.858	277.478	198.807	2,277.759
6–7	947.097	859.963	161.414	268.237	261.387	2,498.097
7–8	1,350.520	1,086.905	214.850	331.338	332.933	3,316.547
8–9	1,531.295	1,175.635	245.164	368.140	374.194	3,694.428
9–10	1,343.361	954.988	205.519	305.630	265.161	3,074.659
10–11	1,403.553	977.354	216.491	303.782	295.804	3,196.983
11–12	1,195.591	824.431	192.354	256.980	333.292	2,802.649
12–13	987.764	703.200	165.573	220.585	265.760	2,342.883
13–14	827.701	598.500	141.845	182.321	238.409	1,988.777
14–15	764.536	492.470	122.448	150.255	185.673	1,715.383
15–20	1,937.397	1,315.225	333.408	391.692	670.004	4,647.724
20–25	655.948	402.829	118.633	110.468	229.462	1,517.340
25–50	1,375.267	391.531	112.894	119.285	180.244	2,179.220
50–100	532.773	37.932	11.772	11.356	3.562	597.394
100–500	204.231	8.857	2.827	2.673	0.000	218.585
500–1,000	6.067	0.133	0.042	0.040	0.000	6.281
1,000+	2.092	0.039	0.012	0.012	0.000	2.155
Total	17,088.493	11,999.012	2,624.456	3,853.489	4,098.837	39,664.278

TABLE 4B

NUMBER OF FAMILIES WITH 1966 TOTAL EMPLOYER WAGE SUPPLEMENT CONTRIBUTIONS IN THE FAMILY MERGE FILE, BY COMPONENT

[population in thousands]

Money Factor Income (000's)	Private Pensions & Welfare Funds	Social Security	Workmen's Compensation	Unemployment Insurance	Civilian Govt. Retirement	Total Wage Supplements
Under 1	1,072.682	2,145.439	1,258.784	1,202.734	269.477	2,233.013
1–2	1,205.976	2,257.495	1,626.985	1,424.428	315.946	2,338.243
2–3	1,481.001	2,297.935	1,945.995	1,755.864	268.033	2,377.388
3–4	1,901.549	2,734.466	2,435.797	2,192.666	428.099	2,842.449
4–5	2,139.169	3,028.648	2,815.052	2,469.359	509.718	3,162.998
5–6	2,873.048	3,789.565	3,576.395	3,215.683	723.082	3,982.800
6–7	2,684.104	3,400.618	3,273.008	2,945.586	800.160	3,660.742
7–8	3,200.066	3,947.927	3,809.620	3,427.562	915.807	4,165.629
8–9	3,321.654	4,025.991	3,934.519	3,612.908	930.475	4,263.599
9–10	2,631.721	3,104.561	3,031.827	2,822.850	661.493	3,232.480
10–11	2,481.685	3,005.705	2,867.422	2,673.599	694.160	3,129.486
11–12	2,054.205	2,440.003	2,398.662	2,188.453	729.019	2,570.747
12–13	1,664.540	1,971.490	1,916.791	1,755.938	526.952	2,045.777
13–14	1,350.345	1,644.423	1,599.718	1,478.623	471.191	1,711.200
14–15	1,052.629	1,317.798	1,259.844	1,144.284	358.398	1,340.987
15–20	2,650.982	3,192.479	3,064.216	2,835.589	1,110.003	3,342.923
20–25	735.124	954.493	897.274	815.902	305.234	986.681
25–50	832.545	1,010.771	916.910	877.777	199.631	1,035.403
50–100	148.403	153.461	147.123	148.086	1.417	153.461
100–500	35.112	35.188	34.798	35.061	0.000	35.188
500–1,000	0.529	0.529	0.528	0.529	0.000	0.529
1,000+	0.168	0.168	0.168	0.168	0.000	0.168
Total	35,517.236	46,459.153	42,811.435	39,023.648	10,218.294	48,611.889

TABLE 4C

AVERAGE AMOUNT OF 1966 TOTAL EMPLOYER WAGE SUPPLEMENT CONTRIBUTIONS PER RECIPIENT FAMILY IN THE FAMILY MERGE FILE, BY COMPONENT

[dollars]

Money Factor Income (000's)	Private Pensions & Welfare Funds	Social Security	Workmen's Compensation	Unemployment Insurance	Civilian Govt. Retirement	Total Wage Supplements
Under 1	108	39	11	23	93	119
1–2	73	61	12	37	81	139
2–3	113	95	19	59	121	235
3–4	162	136	26	74	167	344
4–5	237	179	34	83	215	461
5–6	291	216	42	86	275	572
6–7	353	253	49	91	327	682
7–8	422	275	56	97	364	796
8–9	461	292	62	102	402	867
9–10	510	308	68	108	401	951
10–11	566	325	76	114	426	1,022
11–12	582	338	80	117	457	1,090
12–13	593	357	86	126	504	1,145
13–14	613	364	89	123	506	1,162
14–15	726	374	97	131	518	1,279
15–20	731	412	109	138	604	1,390
20–25	892	422	132	135	752	1,538
25–50	1,652	387	123	136	903	2,105
50–100	3,590	247	80	77	2,515	3,893
100–500	5,817	252	81	76	0	6,212
500–1,000	11,469	251	79	76	0	11,874
1,000+	12,450	234	74	72	0	12,829

TABLE 5A
DERIVATION OF 1966 TOTAL TRANSFER PAYMENTS INCOME IN THE FAMILY MERGE FILE, BY COMPONENT
[amounts in millions]

Money Factor Income (000's)	Social Security Income	Public Assistance Income	Veterans' Disability Compensation	Workmen's Compensation	Unemployment Insurance Income	Total Transfer Payments
Under 1	8,647.367	3,008.471	1,412.064	263.073	139.805	13,470.782
1–2	2,492.056	379.298	289.048	206.163	125.267	3,491.831
2–3	1,704.710	219.548	263.931	110.262	166.961	2,465.413
3–4	1,400.755	195.552	254.366	179.365	169.468	2,199.506
4–5	1,065.517	117.557	181.372	113.935	163.022	1,641.403
5–6	1,019.246	103.719	230.804	185.951	193.173	1,732.893
6–7	744.477	38.761	208.995	155.483	168.234	1,315.950
7–8	644.688	42.843	221.764	121.898	195.404	1,226.596
8–9	601.159	54.878	178.054	136.137	124.147	1,094.375
9–10	417.606	41.054	96.064	111.383	126.556	792.663
10–11	559.525	0.000	133.919	81.927	128.987	904.358
11–12	324.767	0.000	172.054	50.947	80.481	628.249
12–13	376.466	0.000	130.638	43.764	67.380	618.248
13–14	369.463	0.000	76.946	45.492	51.920	543.820
14–15	218.866	0.000	81.102	58.628	37.398	395.994
15–20	609.433	0.000	173.020	156.087	93.951	1,032.491
20–25	178.769	0.000	23.885	31.924	30.131	264.710
25–50	174.449	0.000	41.355	14.219	16.022	246.045
50–100	2.430	0.000	1.527	0.068	0.000	4.026
100–500	0.000	0.000	0.000	0.000	0.000	0.000
500–1,000	0.000	0.000	0.000	0.000	0.000	0.000
1,000+	0.000	0.000	0.000	0.000	0.000	0.000
Total	21,551.747	4,201.683	4,170.911	2,066.707	2,078.305	34,069.354

TABLE 5B

NUMBER OF FAMILIES WITH TRANSFER PAYMENTS INCOME IN THE FAMILY MERGE FILE, BY COMPONENT

[population in thousands]

Money Factor Income (000's)	Social Security Income	Public Assistance Income	Veterans' Disability Compensation	Workmen's Compensation	Unemployment Insurance Income	Total Transfer Payments
Under 1	6,769.122	2,986.344	996.499	99.251	206.914	7,733.105
1–2	1,832.159	415.831	177.482	102.722	234.073	2,202.947
2–3	1,318.674	269.072	184.991	88.883	284.162	1,695.896
3–4	1,045.731	195.276	157.925	128.497	299.968	1,501.279
4–5	858.767	164.369	146.042	154.959	319.557	1,362.928
5–6	788.343	147.790	187.696	194.729	366.522	1,398.285
6–7	603.194	75.688	126.048	151.610	352.044	1,155.268
7–8	584.207	80.723	188.414	182.035	426.341	1,274.392
8–9	493.563	90.705	177.452	156.658	285.563	1,060.857
9–10	375.869	62.791	122.708	127.589	285.692	882.547
10–11	406.227	0.000	127.141	137.269	263.707	845.968
11–12	290.776	0.000	115.428	91.999	165.720	617.771
12–13	265.014	0.000	89.835	76.552	156.283	529.597
13–14	258.445	0.000	78.372	55.795	86.857	433.748
14–15	171.693	0.000	87.787	71.412	74.420	338.929
15–20	473.560	0.000	158.318	146.100	205.816	903.449
20–25	171.752	0.000	41.990	26.749	40.220	258.908
25–50	129.841	0.000	41.324	13.755	36.654	203.572
50–100	2.288	0.000	1.186	0.580	0.000	2.869
100–500	0.000	0.000	0.000	0.000	0.000	0.000
500–1,000	0.000	0.000	0.000	0.000	0.000	0.000
1,000+	0.000	0.000	0.000	0.000	0.000	0.000
Total	16,839.227	4,488.589	3,206.638	2,007.147	4,090.513	24,402.314

TABLE 5C

AVERAGE AMOUNT OF 1966 TOTAL TRANSFER PAYMENTS INCOME PER RECIPIENT FAMILY IN THE FAMILY MERGE FILE, BY COMPONENT

[dollars]

Money Factor Income (000's)	Social Security Income	Public Assistance Income	Veterans' Disability Compensation	Workmen's Compensation	Unemployment Insurance Income	Total Transfer Payments
Under 1	1,277	1,007	1,417	2,651	676	1,742
1–2	1,360	912	1,629	2,007	535	1,585
2–3	1,293	816	1,427	1,241	588	1,454
3–4	1,339	1,001	1,611	1,396	565	1,465
4–5	1,241	715	1,242	735	510	1,204
5–6	1,293	702	1,230	955	527	1,239
6–7	1,234	512	1,658	1,026	478	1,139
7–8	1,104	531	1,177	670	458	962
8–9	1,218	605	1,003	869	435	1,032
9–10	1,111	654	783	873	443	898
10–11	1,377	0	1,053	597	489	1,069
11–12	1,117	0	1,491	554	486	1,017
12–13	1,421	0	1,454	572	431	1,167
13–14	1,430	0	982	815	598	1,254
14–15	1,275	0	924	821	503	1,168
15–20	1,287	0	1,093	1,068	456	1,143
20–25	1,041	0	569	1,193	749	1,022
25–50	1,344	0	1,001	1,034	437	1,209
50–100	1,062	0	1,288	118	0	1,403
100–500	0	0	0	0	0	0
500–1,000	0	0	0	0	0	0
1,000+	0	0	0	0	0	0

TABLE 6

RELATION BETWEEN 1966 DIRECT FEDERAL TAX AND TRANSFER PAYMENTS AND TOTAL INCOME BEFORE TRANSFERS, BY INCOME CLASS

[amounts in millions]

Money Factor Income (000's)	Total Income Before Transfer Payments (1)	Total Transfer Payments (2)	Federal Income Tax (3)	Federal Payroll Taxes (4)	Transfers as Percent of Income (2)/(1)	Income Tax as Percent of Income (3)/(1)	Payroll Tax as Percent of Income (4)/(1)
Under 1	3,881.510	13,470.782	248.704	214.282	347.050	6.407	5.521
1–2	8,682.402	3,491.831	188.838	364.062	40.217	2.175	4.193
2–3	11,017.625	2,465.413	301.107	609.570	22.377	2.733	5.533
3–4	15,340.025	2,199.506	674.403	1,013.779	14.338	4.396	6.609
4–5	20,662.634	1,641.403	1,089.878	1,453.981	7.944	5.275	7.037
5–6	28,744.909	1,732.893	1,703.583	2,133.883	6.029	5.927	7.424
6–7	31,069.325	1,315.950	1,973.744	2,216.097	4.236	6.353	7.133
7–8	39,977.253	1,226.596	2,813.446	2,807.400	3.068	7.038	7.022
8–9	45,556.143	1,094.375	3,375.071	3,041.322	2.402	7.409	6.676
9–10	38,922.802	792.663	3,014.736	2,480.918	2.037	7.745	6.374
10–11	42,622.948	904.358	3,324.364	2,559.856	2.122	7.799	6.006
11–12	37,514.156	628.249	3,100.028	2,170.638	1.675	8.264	5.786
12–13	33,475.612	618.248	2,916.985	1,884.066	1.847	8.714	5.628
13–14	30,606.612	543.820	2,678.464	1,599.781	1.777	8.751	5.227
14–15	24,977.261	395.994	2,273.389	1,330.879	1.585	9.102	5.328
15–20	75,496.215	1,032.491	7,133.639	3,520.281	1.368	9.449	4.663
20–25	30,527.806	264.710	3,153.279	1,101.154	0.867	10.329	3.607
25–50	58,552.694	246.045	6,859.060	1,124.772	0.420	11.714	1.921
50–100	25,192.633	4.026	4,222.095	121.263	0.016	16.759	0.481
100–500	17,593.022	0.000	3,023.974	27.604	0.000	17.188	0.157
500–1,000	1,894.540	0.000	262.995	0.402	0.000	13.882	0.021
1,000+	1,991.324	0.000	264.264	0.114	0.000	13.271	0.006
Total	624,299.460	34,069.354	54,596.046	31,776.102	5.457	8.240	4.914

by negative incomes and the very large amount of capital gains and other income subject to preferential rates, respectively) but never reaches more than 17 percent of total income before transfers in any MFI class; and (3) the effective payroll-tax rate is roughly constant up to the $7,000 MFI level, where the taxable earnings maximum is reached, and then it declines as income rises. Thus, in terms of a comprehensive income concept, transfers and the individual income tax are progressive while the payroll taxes are regressive.

Uses of the MERGE File

The initial purpose of the MERGE File was to provide the basis for estimating the distribution of federal, state, and local taxes by income levels. But the file has also been useful for a number of other purposes—mainly tax calculations—which require information not now available on individual income-tax returns. We have only just begun to exploit the many uses of the file; and in this section, we present a number of examples to illustrate the versatility of the file and the types of analyses that can be made with it.

Distribution of Tax Burdens

Approximately the same methodology has been used for the last thirty-five years in the United States and other countries to estimate the distribution of tax burdens by income classes. Essentially, the method is to allocate individual taxes to broad income classes on the basis of a large number of statistical series which are proxies for the tax distributions. Thus, for example, sales taxes are allocated on the basis of the distribution of consumption (adjusted when necessary, for items which are not taxable), payroll taxes are allocated on the basis of the distribution of payrolls, and so on.[13]

The major disadvantage of this methodology is that it distributes taxes on the basis of the average income and behavior of all households in a particular income class, rather than on the basis of the income and behavior of the individual micro-units in each class. This means that it is impossible to differentiate among households for the numerous differences (e.g., income, consumption patterns, marital status, living arrangements) that may lead to relatively large differences in tax payments among families with approximately the same amount of income.

Although we cannot make all the distinctions that are relevant to the estimation of tax liabilities, the MERGE File is the richest source of information developed thus far for this purpose. Among the characteristics that are particularly important for estimating tax payments are sources of income; marital status and family composition; home ownership and mortgage debt; and state and local tax payments. Unfortunately, the SEO survey did not obtain consumption data, but this gap was filled by simulation techniques, using a survey for an earlier year.[14]

[13] The classic study along these lines is by Richard A. Musgrave and others, "Distribution of Tax Payments by Income Groups: A Case Study for 1948," *National Tax Journal*, Vol. 4 (March 1951).

[14] The basic source was the 1960–61 Consumer Expenditure Survey, conducted by the U.S. Bureau of Labor Statistics in connection with its revision of the weights for the preparation of the official consumer price index.

In addition, whenever it is necessary to make assumptions about the economic behavior of households, we are not limited to a single assumption for all families in a given income class. The availability of the computer permits us to attribute characteristics to individual units in substantial detail through simulation techniques. For example, we have already prepared some twenty-odd multivariate regression equations for various consumption items in order to estimate sales and excise tax payments for each unit in the file. While these techniques will not insure absolute accuracy, they will, at least, permit us to depart from the assumption of uniformity which has been the hallmark of all previous tax-burden studies.

Aside from this major improvement in methodology, the MERGE File permits us to prepare distributions of tax burdens on the basis of numerous alternative assumptions of the incidence of various taxes. In the past, the number of incidence combinations has been limited by the sheer magnitude of the computational job. The computer gives us much greater flexibility and scope in this respect. Furthermore, it will be possible to classify the tax burden distributions not only by size of income, but also by family size; age, sex, and education of family head; housing status (homeowners versus renters); and many other characteristics. These classifications will provide new insights into the impact of the tax system on different socioeconomic groups in the population.

Reforming the Payroll Tax

In most countries, the social security system is financed by a payroll tax levied at a flat rate, without exemptions or deductions. There is often a limit on the earnings which are subject to tax, so that the payroll tax becomes regressive for those with earnings above the limit. The use of a regressive tax is justified primarily on the grounds that the social security system is a system of insurance, which requires separate financing on the basis of an earmarked tax, and which merits some contribution even by wage earners who are acknowledged to be poor.

The insurance rationale for social security has come under increasing attack as the burden of the payroll tax has increased. Many economists have pointed out that the insurance elements of social security are extremely tenuous, and that it is cruel to impose heavy tax burdens on persons with low incomes on this ground. According to this view, the social security system should be regarded as a tax-transfer system, which should be financed out of general revenues, just as other transfers are financed. The U.S. social security system distributes benefits to persons who experience a sharp decline in income at retirement or if they become disabled, but the amount of their tax contributions is not even approximately related to the eventual benefits they receive. It can be shown that in a country with rising per capita income and a growing population, each generation can afford to pay much higher benefits to the disabled and retired persons, without increasing tax rates.[15]

In the United States, there is great interest, inside and outside of Congress, in developing new methods of financing social security that will bear less heavily on low-income earners than does the present system. We have used the MERGE

[15] For further development of these ideas, see Joseph A. Pechman, Henry J. Aaron, and Michael K. Taussig, *Social Security: Perspectives for Reform*, The Brookings Institution, 1968.

File to illustrate the effect on the tax rate and on tax liabilities of introducing personal exemptions into the payroll-tax base. We have also made estimates of the rate required to replace the payroll tax on employees by a flat tax on total income less the personal exemptions. Since the Tax File does not include the earnings of nonfilers and only very limited occupational information, it was necessary to use the MERGE File for these calculations.

Our calculations show that the flat payroll tax paid by wage and salary earners can be replaced by a mildly progressive tax on total income or on earnings, at reasonably moderate rates. The progressive tax would relieve those who earn less than the officially defined "poverty lines" from making any contribution to social security out of their inadequate incomes; and it would reduce the taxes of the vast majority of income recipients, while raising taxes only for the top 10 or 15 percent of earners. The merits of these alternative methods of financing social security are just being recognized, and the public debate is already underway.[16]

Developing a Comprehensive Income Tax

Much has been said in the United States about the "erosion" of the tax base resulting from the numerous exclusions, exemptions, and deductions permitted under various provisions of the Internal Revenue Code. The extent of the erosion has been estimated in aggregate terms, but reliable estimates of the differential impact of the special provisions at various income levels have never been available. The Tax File has been used to make some of the estimates, but, of necessity, they have been confined to the items that appear on tax returns. The MERGE File now permits us to make these estimates on the basis of the adjusted-family-income concept, which is a close approximation to the concept of "economic income."[17]

The computer program used to make these calculations provides us with estimates of the tax base and tax liability under the current law by income classes, and by marital status, and with similar data after the following successive tax-law revisions: (1) elimination of the rate advantages of income splitting; (2) treatment of capital gains as ordinary income; (3) constructive realization of capital gains at gift or death; (4) taxation of net imputed rent on owner-occupied houses and elimination of the deductions for mortgage interest and property taxes; (5) taxation of transfer payments as ordinary income; (6) elimination of most of the personal deductions; and (7) substitution of a flat standard deduction of $1,300 for the present standard deduction of 15 percent of income up to a maximum of $2,000.

After the tax basis and tax liabilities are calculated, it is relatively simple to estimate the lower tax rates that would yield the same revenue as is now collected from the income tax, after each of the changes is made. To make the estimates relevant to the current scene, we have also developed projection techniques to raise the incomes in the MERGE File to the expected 1972 levels.[18]

[16] Senators Mondale and Muskie introduced legislation, S. 2656, incorporating features similar to these, in the U.S. Senate on October 5, 1971.

[17] See pp. 68–69 above.

[18] The estimates for 1972 were based on projections of income from the 1966 base, assuming that the percentage change in individual income sources will be the same as the estimated change in the personal income components.

On the basis of MERGE File calculations we have estimated that the 1972 tax yield on such a comprehensive tax base would have been $77 billion higher than under existing law. Conversely, average tax rates could have been reduced by 43 percent without reducing the yield of the individual income tax.[19]

The MERGE File provides a mine of information for analytical work on the characteristics of income recipients at all income levels. In addition to tax analysis, the new file will be useful for making estimates of alternative income-maintenance programs. Other uses will doubtless be developed as we gain more experience with the use of the file and develop a more complete library of computer programs for its use. We hope that other analysts will be able to develop similar files on the basis of the tax and survey information in their own countries. Our experience indicates that the benefits will be well worth the costs.

[19] See Joseph A. Pechman and Benjamin A. Okner, "Individual Income Tax Erosion by Income Classes" in *The Economics of Federal Subsidy Programs*, A Compendium of Papers Prepared for the Use of the Joint Economic Committee, 92 Cong. 2nd session (1972) (Brookings Reprint No. 230).