Volume Title: Income Size Distributions in the United States, Part I

Volume Author/Editor: Conference on Research in Income and Wealth

Volume Publisher: NBER

Volume ISBN: 0-870-14161-9

Volume URL: http://www.nber.org/books/unkn43-2

Publication Date: 1943

Chapter Title: Comparability And Deficiencies Of Existing Data And The Construction Of A Size Distribution For The United States

Chapter Author: Conference on Research in Income and Wealth

Chapter URL: http://www.nber.org/chapters/c5717

Chapter pages in book: (p. 68 - 98)

CHAPTER 3

Comparability and Deficiencies of Existing Data and the Construction of a Size Distribution for the United States

TWO IMPORTANT questions in the interpretation and use of income data are: How comparable are the data from the various sources? Can material from diverse sources be combined to give a reasonably reliable estimate of the distribution of income among all individuals or families in the country?

Discussion is confined to the bodies of data described in Part II, since these are the best sources of income distribution we have at present. Some studies of narrower scope may prove useful in filling some of the gaps noted here or in providing bridges to link bodies of data not on a comparable basis. However, it is believed that our conclusions would not be materially altered had it been possible to extend our review beyond its present scope.

I ILLUSTRATIVE INCOME DISTRIBUTIONS

These sample distributions, drawn from the studies described in Part II, reveal both the diversity of the data and many of the problems of comparability and interpretation. The three sets of distributions: (a) distributions of family income in communities of different size; (b) national distributions of family incomes based on different sources; and (c) distributions of individual incomes for a single state illustrate facets of the problem of integrating different bodies of data. No attempt has been made to give a statistically rigorous explanation of the differences among the distributions in each set since the data are presented for illustrative rather than substantive purposes.

A FAMILY INCOME IN URBAN, VILLAGE, AND RURAL AREAS

These distributions illustrate the marked differences among the income distributions of families living in communities of different size in various parts of the country. They are taken from the Consumer Purchases Study and should be comparable in respect of the definition of income. Up to $4,000, the published data are classified by more income intervals than are shown in Table 5 and Chart 1.

TABLE 5

Complete Families, 1935–1936 (Consumer Purchases Study)[1]

	URBAN		VILLAGE		FARM	
INCOME CLASS	Denver	Atlanta[2]	12 Calif. villages	15 Ga. & S. C. villages[2]	2 southern Calif. counties	5 Iowa counties
			Number of families			
All families	8,578	19,851	1,845	3,974	1,159	748
On relief	1,186	3,742	324	709	44	36
Nonrelief families with income of						
Under $250[3]	68	196	14	242	79	38
250- 499	158	851	33	643	72	74
500- 999	1,055	3,665	258	857	197	265
1,000-1,999	3,299	5,627	805	972	365	262
2,000-3,999	2,257	4,760	362	461	309	69
4,000-7,499	470	890	38	80	70	4
7,500 & over	85	120	11	10	23	..
			Percentage of families			
All families	100.0	100.0	100.0	100.0	100.0	100.0
On relief	13.8	18.8	17.5	17.8	3.8	4.8
Nonrelief families with income of						
Under $250[3]	.8	1.0	.8	6.1	6.8	5.1
250- 499	1.8	4.3	1.8	16.2	6.2	9.9
500- 999	12.3	18.5	14.0	21.6	17.0	35.5
1,000-1,999	38.5	28.3	43.6	24.5	31.5	35.0
2,000-3,999	26.3	24.0	19.6	11.6	26.7	9.2
4,000-7,499	5.5	4.5	2.1	2.0	6.0	.5
7,500 & over	1.0	.6	.6	.2	2.0	..

[1] Data from U. S. Department of Labor, Bureau of Labor Statistics, Bulletin 646, I, 124; Bulletin 647, I, 147 and 195; and U. S. Department of Agriculture, Misc. Publication 339, p. 110; Misc. Publication 375, pp. 15 and 93; Misc. Publication 356, p. 21; Misc. Publication 383, p. 21. Number of village and farm families in highest income group furnished by Department of Agriculture.

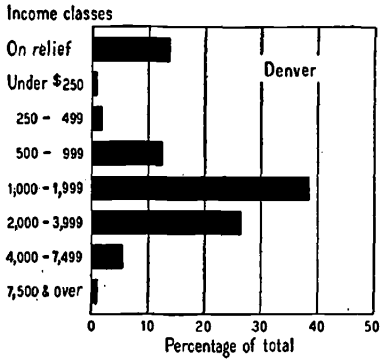
[2] Sum of white and colored families in each income group.

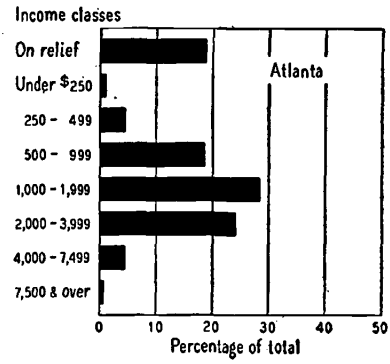
[3] Includes nonrelief families with deficits, i.e., with business losses in excess of other income.

CHART 1

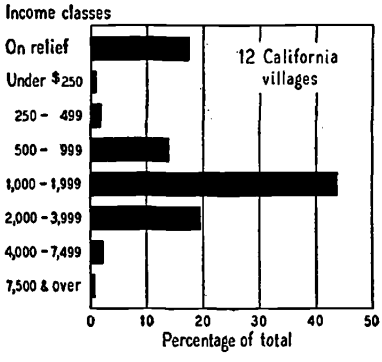
COMPLETE FAMILIES, 1935-1936
(CONSUMER PURCHASES STUDY)

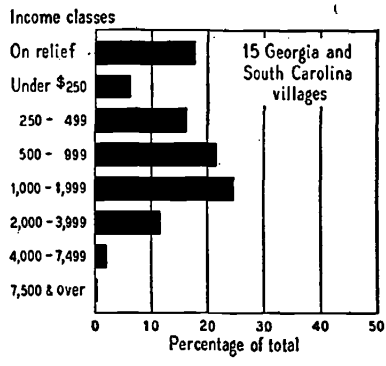
URBAN



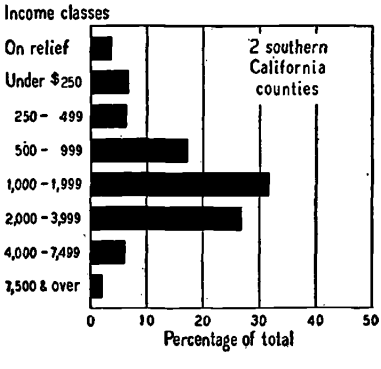


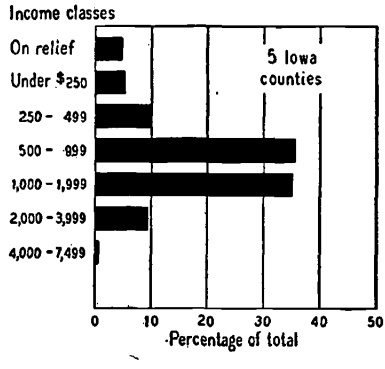
VILLAGE





FARM





B FAMILY INCOME REPORTED ON FEDERAL RETURNS AND IN URBAN AREAS

Of the data described in this volume these are the only three published distributions of family income, or of approximations to family income, that are for areas representing a substantial portion of the nation's population (Table 6 and Chart 2).[1] They well illustrate the difficulties encountered in attempts to compare tabulations from one body of data with those from another.

In order to approximate a family distribution, the tabulation of federal returns uses the combined net income of husbands and wives, as tabulated from matched returns, instead of the net incomes of husbands and of wives reported on separate returns. The only year for which this tabulation is available is 1936. Income is defined as statutory net income, excluding capital gain or loss. Because of the deductions permitted, statutory net income differs in important respects from total income as defined in the National Health Survey and in the Financial Survey of Urban Housing, the other two sources of our data.

In all three bodies of data, individuals living alone are tabulated as 'families'. However, the incomes of persons other than husband and wife related to and living in the family are included in family income in the National Health Survey and in the Financial Survey of Urban Housing, but are omitted from the tax return tabulation or are tabulated separately. While the 'family' in the National Health Survey and in the Financial Survey of Urban Housing are not identical, they are similar.

Tax returns obviously exclude most families with incomes below $2,500. For this reason, the charts on the right side of the page are drawn to make comparison of the three distributions for intervals above $3,000 possible.

All income groups for which data are published by the National Health Survey are given in the table and chart. Data from the Financial Survey of Urban Housing are available for 39 income groups and those from federal returns for 43 income and deficit groups. The intervals presented were chosen to afford as much comparability among the three distributions as possible, though

TABLE 6

Family Income Reported on Federal Income Tax Returns and by Surveys in Urban Areas

FEDERAL RETURNS, 1936 [1]		83 CITIES, *National Health Survey, 1935* [2]		33 CITIES, *Financial Survey of Urban Housing, 1933* [3]	
Income class	Number	Income class	Number	Income class	Number
		Number of families			
All returns	5,219,820	All families	2,402,786	All families	241,207
With net deficit	84,110	Relief	452,913	With income of	
With net income of		Nonrelief with income of		None	12,304
Under $500	74,885	Under $1,000	577,677	$1- 250	26,276
500- 1,000	199,111	1,000-1,499	538,210	251- 450	24,126
1,000- 1,500	1,305,929	1,500-1,999	394,349	451- 949	56,917
1,500- 2,000	725,237	2,000-2,999	279,062	950- 1,450	46,960
2,000- 3,000	1,199,948	3,000-4,999	110,311	1,451- 1,949	32,172
3,000- 5,000	1,001,591	5,000 & over	50,264	1,950- 2,749	23,876
5,000-10,000	407,598			2,750- 4,949	13,998 [4]
10,000-20,000	142,862			4,950- 9,949	3,714
20,000 & over	78,549			9,949-19,949	699
				19,950 & over	165
		Percentage of families			
All returns	100.0	All families	100.0	All families	100.0
With net deficit	1.6	Relief	18.9	With income of	
With net income of		Nonrelief with income of		None	5.1
Under $500	1.4	Under $1,000	24.0	$1- 250	10.9
500- 1,000	3.8	1,000-1,499	22.4	251- 450	10.0
1,000- 1,500	25.0	1,500-1,999	16.4	451- 949	23.6
1,500- 2,000	13.9	2,000-2,999	11.6	950- 1,450	19.5
2,000- 3,000	23.0	3,000-4,999	4.6	1,451- 1,949	13.3
3,000- 5,000	19.2	5,000 & over	2.1	1,950- 2,749	9.9
5,000 & over	12.1			2,750- 4,949	5.8
				4,950 & over	1.9
		Percentage of families with net incomes over $3,000			
All returns	100.0	All families	100.0	All families	100.0
$3,000- 5,000	61.4	$3,000-4,999	68.7	$3,000- 4,949	67.7
5,000-10,000	25.0	5,000 & over	31.3	4,950- 9,949	26.2
10,000-20,000	8.8			9,950-19,949	4.9
20,000 & over	4.8			19,950 & over	1.2

[1] U. S. Treasury Department, *Statistics of Income Supplement, compiled from income tax returns for 1936,* Sec. I, p. 1, and Sec. II, pp. 2-16. Number of matched returns, classified by combined net income, substituted for separate returns of husbands and wives.

[2] U. S. Public Health Service, The Relief and Income Status of the Urban Population of the United States, 1935, preliminary reports, National Health Survey, Population Series, Bulletin C. Figures for cities in four areas combined, without weighting.

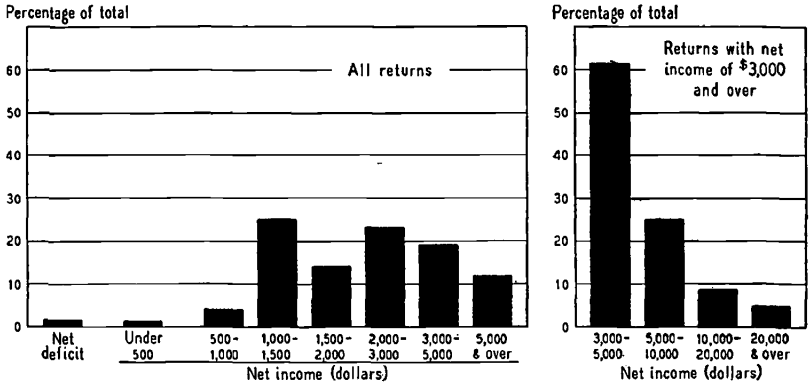
[3] David L. Wickens, *Residential Real Estate* (National Bureau of Economic Research, 1941), p. 146.

[4] It is estimated, by interpolation, that three-fourths of the families in the income group $2,750-3,149, or 4,406 of the families in the income group $2,750-4,949, had incomes under $3,000.

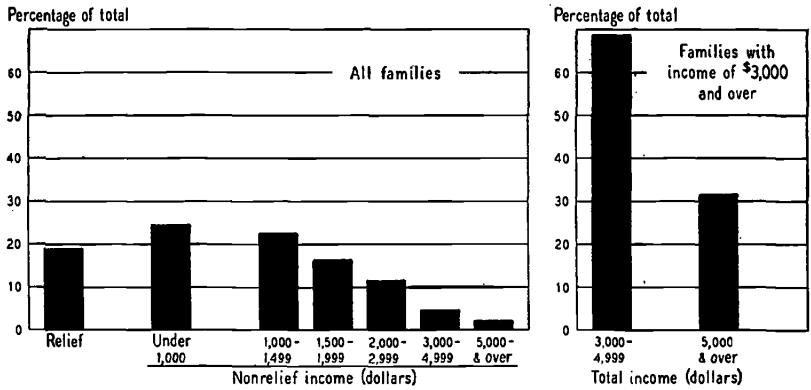
CHART 2

FAMILY INCOME REPORTED ON FEDERAL RETURNS AND IN URBAN AREAS

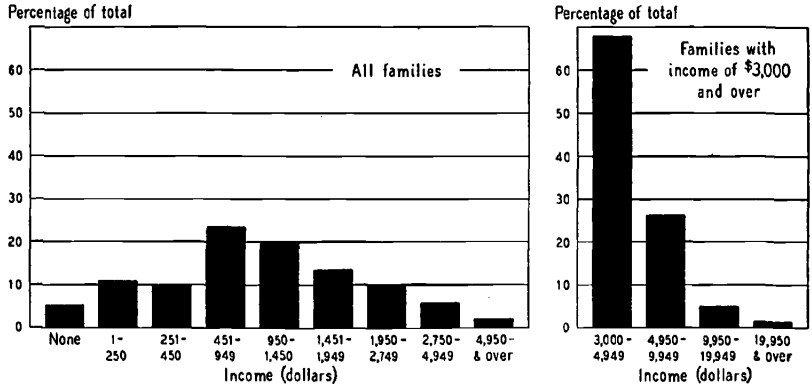
HUSBANDS AND WIVES, AND OTHER INDIVIDUALS, FILING INCOME TAX RETURNS, 1936



FAMILIES IN 83 CITIES, NATIONAL HEALTH SURVEY, 1935, UNWEIGHTED SAMPLE



FAMILIES IN 33 CITIES, FINANCIAL SURVEY OF URBAN HOUSING, 1933, UNWEIGHTED SAMPLE



they do not conform even approximately to any arithmetic or geometric principle.

It is impossible to determine, or even to judge crudely, how much of the difference among the distributions is due to: (a) the difference in the period, (b) the representativeness of the various samples, (c) the difference between urban areas and the entire nation with respect to incomes over $3,000, (d) the difference between family and combined husband and wife income, (e) the difference between statutory net and 'total' income, or (f) the differences in interval limits.

C INDIVIDUAL INCOME, DELAWARE

These distributions illustrate the differences among income distributions derived from data collected each year for the same area and population group. While for the most part the distributions pertain strictly to individual incomes, the tax data, both federal and state, include joint returns of husband and wife (Table 7 and Chart 3). Since none of the distributions is based on a sample, the charts are drawn from actual numbers rather than percentages of totals. Of course, the distributions do not pretend to include all individuals in Delaware.

The differences between the two distributions from federal returns are due to (a) the difference between net income for tax purposes and total income as reported on tax returns; (b) the difference arising from inclusion or omission of statutory net capital gain or loss. The difference between the distributions of total income reported on state and federal returns is due to the difference in coverage, since in Delaware all adults are required to file. The difference between the distribution for 1936 derived from state returns and the distribution for 1937 derived from Old-Age and Survivors Insurance statistics is due primarily to the difference between the total income, on the one hand, and wages in covered employment, on the other, and secondarily, to the difference between income recipients and recipients of wages in covered industries.

TABLE 7

Individual Income, Delaware

INCOME CLASS	FEDERAL RETURNS, 1936: Net income [1]	FEDERAL RETURNS, 1936: Total income [2]	STATE RETURNS, 1936, TOTAL INCOME [2]	OLD-AGE AND SURVIVORS INSURANCE, 1937, WAGES IN COVERED INDUSTRIES [3]
	Number of individuals			
All individuals	13,739	13,197	85,243	83,126
With income of				
Under $0			203	..
0- 200			18,355	22,946
200- 400	3,730	3,028	6,032	9,738
400- 800			13,886	14,129
800- 1,500			23,335	20,548
1,500- 3,000	4,888	4,563	17,705	12,157
3,000- 6,000	3,114	3,555	3,792	3,608
6,000-12,000	1,040	1,137	1,802	
12,000-25,000	515	523	500	
25,000 & over	452	391	353	
	Percentage of individuals			
All individuals	100.0	100.0	100.0	100.0
With income of				
Under $0			0.2	..
0- 200			21.5	27.6
200- 400	27.1	22.9	7.1	11.7
400- 800			16.3	17.0
800- 1,500			27.4	24.7
1,500- 3,000	35.6	34.6	20.8	14.6
3,000- 6,000	22.7	26.9	4.4	4.4
6,000-12,000	7.6	8.6	1.3	
12,000-25,000	3.7	4.0	.6	
25,000 & over	3.3	3.0	.4	

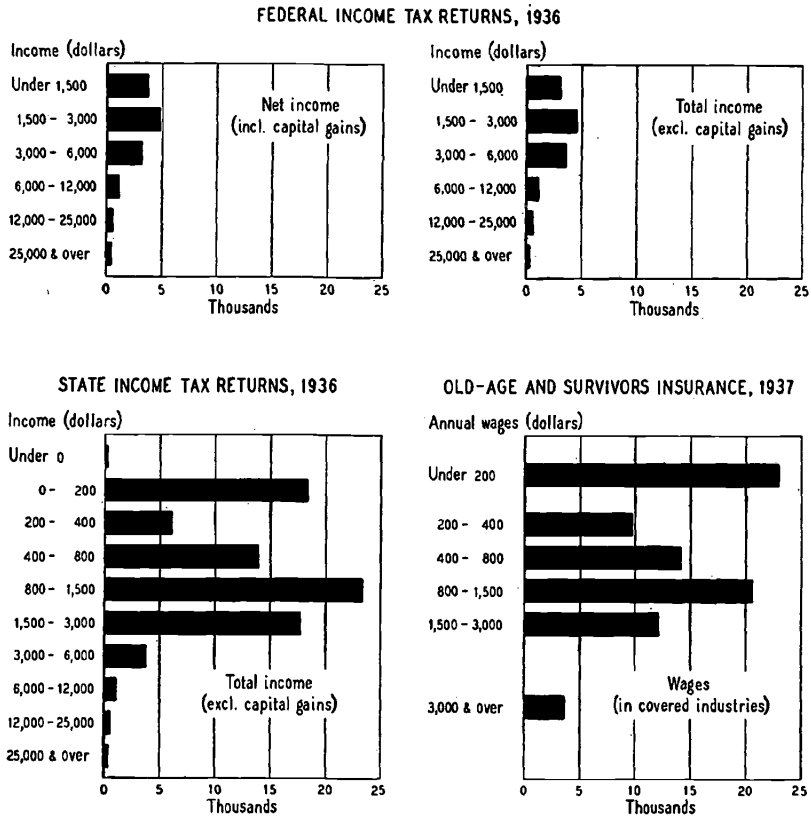
[1] U. S. Treasury Department, *Statistics of Income for 1936,* Part I, p. 111. Net income includes statutory capital gain or loss, but excludes deductions authorized by law.

[2] See Part II, Ch. 4.

[3] Social Security Board, *Old-Age and Survivors Insurance Statistics, Employment and Wages of Covered Workers, 1938,* pp. 208-10.

CHART 3

INDIVIDUAL INCOME, DELAWARE



II INTEGRATION

A INCOME CONCEPTS

A basic obstacle to the integration of income data from diverse sources is the simple but vital fact that the concepts of income differ, sometimes widely. As yet we do not have sufficient uniformity and agreement to determine the significance of the divergencies; and the lack of uniformity in other respects—temporal, geographical, etc.—has prevented the discovery of what is common.

The federal income tax uses at least three concepts of income—total, net, and net taxable—although the second occurs most frequently and is the basis of classification. Included, as mentioned above, are most items ordinarily considered income, but there are enough about which there has been less agreement to impair the usefulness of the figures. Many if not most of these items are on the borderline between income and wealth—capital gains and losses, stock dividends, insurance, annuities, uninsured losses, trust income, corporation and business items, depreciation, inheritances, and gifts. Others, though not easy to classify under a single heading, are no less important for some purposes—interest on obligations of state and local governments, pensions and benefits under the Old-Age and Survivors Insurance program, income in kind and imputed, relief and unemployment compensation, living allowances, and veterans' pensions. Certain deductions allowed in computing net income are also peculiar to the income tax: interest payments, taxes, and contributions are the chief examples. As the federal definition has changed from time to time, data for different years are not always comparable.

Most state income tax laws have concepts more or less parallel to the federal but a few differences may make the figures much less comparable than they seem at first. How income from sources outside the state and how capital gains and losses are treated, and what deductions are allowed, especially taxes, are probably of chief interest.

The other continuous source, Social Security records, yields data on wages and salaries alone. Thus only one, albeit the largest, item of income is covered. Though its definition is in line with that of the comparable income tax item, we have as yet no way of linking the two. The Minnesota Study may forge such a link for Minnesota, but whether observed relationships there are typical of those in other states is doubtful. The Census also specified wages and salaries rather than total income, and its definition parallels that of the Social Security program, but the coverages are not identical. In addition, the 1940 Census collected data on the receipt of $50 or more income from other sources. Although this information

adds to that from Social Security records it is only one short step toward a definition that takes in the chief items usually considered income or reported on tax returns.

The surveys described in Part II varied so much in concept that they are scarcely comparable either among themselves or with tax returns and Social Security records, although the Michigan Unemployment Census definition of income does approximate wages and salaries under the Social Security program. The Financial Survey of Urban Housing and the National Health Survey both defined income more broadly, including that from property as well as wages and salaries. However, since they disregarded, among other things, changes in capital values, did not exclude contributions, taxes, or interest, and treated some items ambiguously, if at all, their concepts diverge from those of tax regulations. Wages and salaries obtained by the Financial Survey were not tabulated as a separate item and cannot, therefore, be compared with Social Security tabulations. The Health Survey entered an income figure for families that had been on relief only if it was volunteered, but the Financial Survey treated relief as a constituent of total family income. At other points also the two diverge from each other and from other sources. Neither included imputed income.

The Consumer Purchases Study was much more specific. Its income concept includes money income from property and the imputed rental value of an owner-occupied house, as well as wages and salaries, and, for farm families, the value of food produced and consumed on the farm. With minor exceptions capital gains and losses were not taken into account, but inheritances and gifts used for living expenses were treated as income. Some pensions were included, some not, but inclusion depended on different factors than those determining the entry on tax returns.

The Minnesota Survey had still different concepts. The federal income tax definition is departed from in many respects—items of imputed income are included, inventory changes, annuities, capital gains and losses, gifts, pensions are treated differently. Definitions and tabulations are detailed, however, and of the data described in Part II probably these can most readily be integrated with others.

B RECIPIENT UNITS

The unit for which income figures are desired is either an individual or some group, ordinarily the family. For most purposes 'individual' can be defined without much trouble. But the laws of eight or nine states divide much of what would in other states be considered the income of one married person so that it is legally the income of two. In some cases the entire income is so divided; in others, none or merely a portion. Tax but not Social Security and Census statistics recognize this arbitrary division. The ordinary concept of individual is blurred when a person owns a business; if the business is a sole proprietorship, the owner and the business are, for tax purposes but again not for Social Security and most surveys, considered one income receiving unit; if there is a partnership, income is attributed according to the partners' respective shares. If the owner or owners incorporate the enterprise, however, it becomes a separate unit, no longer identified with the owners except as they withdraw salaries, dividends, etc. Trusts also are not treated uniformly: in some cases they are considered as separate income units; in others, the income is attributed to a person who may or may not have received it or had control over it; in still others the treatment is obscure.

Social Security records cover only individuals; the 1940 Census and the Michigan Unemployment Survey also compiled data for individuals. The income tax material, however, applies both to individuals and groups without distinguishing between the two sharply enough for fine analysis or highly reliable integration with other data.

'Family' is more troublesome, for, in addition to most of the problems that arise in attempting to define 'individual', there is the question of what individuals are to be grouped together. One criterion is relationship, a second, common use of household facilities, a third, pooling of income. Decisions on borderline cases of one criterion tend to be made by reference to one or both of the others. Table 2, line 1, summarizes the definitions used by the studies described in Part II. Links between these sources are still,

for the most part, missing. For example, although the 1936 federal returns for husbands and wives were matched as well as possible, the basis on which the results are tabulated differs from that regularly used in *Statistics of Income* so that the effects of combining returns cannot be isolated. The Census hopes to make tabulations that will give keys to the relation of all wage and salary income of individuals to (1) the receipt of, but not the amount of, other income, (2) wages and salaries from employment covered by the Social Security program, (3) total wages and salaries and whether the family received other income. However, since a total figure was obtained for neither individual nor family income and maximum salaries of about $5,000 or less were alone covered, the findings can be related only imperfectly to tax return data. From the tabulations of the Minnesota data, which are more detailed, individuals' incomes can be separated and combined more satisfactorily than figures from any other study; yet some limitations in the data cannot be overcome, e.g., all income from property was attributed to one member of the economic unit no matter who owned it, and the income of relatives not living in the household was excluded.

C PERIODS COVERED

The different periods covered constitute another stumbling block to integration. Federal income tax figures exist for each year beginning with 1916, but no comparable continuous series goes so far back. State income tax figures now constitute a continuous annual source but most are for recent years only. The year 1929 is covered by the Wisconsin Study and the Financial Survey of Urban Housing, and 1932 and 1933 by the latter. The Michigan Survey covered 1934, a year for which federal, Wisconsin, and Montana tax returns also were studied; the Health Survey covered a period not clearly delimited and not uniform for all respondents but probably mainly 1935. The Consumer Purchases Study covered a twelve-month period comprising in most cases parts of both 1935 and 1936.

For 1936 there is an unusual abundance of data—studies of federal, Delaware, Wisconsin, and Montana tax returns, and a considerable part of the Consumer Purchases Study. Unfortunately

the Social Security program did not go into effect until the next year. In addition to the continuous sources, federal and state tax returns, we have for years since 1936 studies of Delaware and Montana returns for 1936 and 1938, the Minnesota Study covering income for part of 1938 and part of 1939, and the Population Census for 1939. All except the Minnesota Study apply to calendar years.

Moreover, the wide fluctuations in aggregate income payments during the 'thirties make any stability manifested in distributions for single years suspect. From the data described in Part II we cannot judge adequately the extent to which the distribution by size varied as totals fluctuated; nor, because of temporal, conceptual, and other differences, can we combine data for different periods to get a comprehensive distribution for any single period.

Data tracing the income pattern of a group for more than one year are few. The Financial Survey of Urban Housing collected 1929, 1932, and 1933 income figures for the same families; the Wisconsin Study traced as well as it could changes from 1929 to 1936; the Social Security Board has begun a study of the wage and salary receipts of a sample of workers in covered employment but it will not be completed for several years. A recently initiated study is tracing the income history of a sample of Delaware residents from 1925 to 1936.

D INCOME GROUPS COVERED

The income levels covered also differ. Were all other factors comparable, this would not constitute a real obstacle to integration, but given other differences, it makes comparison more difficult. Social Security records are primarily useful for persons with incomes below $3,000, though some of the income of many persons with large incomes is reported. The upper income classes are more completely represented by tax returns than the lower. Until 1940 the federal and most state exemptions and filing requirements excluded the great majority of the persons who come under the Social Security program. Of course, there was some overlapping, but in general (and to the extent that the two sources covered comprehensively the incomes of the units they include) they applied to

individuals at different income levels. Though less true at the time of writing (October 1942), this generalization still retains a large degree of validity. The Delaware income tax data are an obvious exception.

The surveys, without apparent exception, have not been very successful in obtaining reports on incomes above $7,500. Even the Michigan Census had so few reports on incomes above $3,000 that all were grouped in the '$2,000 and over' class. Although we do not know the exact dividing line, the Financial Survey of Urban Housing, the National Health Survey, the Consumer Purchases Study, and the Minnesota Survey were also less successful in gathering data for the upper income ranges—over perhaps $7,500—than tax authorities. Their coverage of the lower ranges, however, is probably better, certainly of incomes below $3,000. The Census, with complete population coverage, will yield data whose chief value will be for the study of incomes below about $5,000.

E GEOGRAPHIC COVERAGE

The Census, federal income tax, and Social Security records cover the entire United States and in addition a few persons who for some purposes would be considered as belonging to other regions. Incomes can be classified by relatively small geographic areas, although Social Security records are by place of employment, and income tax returns often, but not always, by residence. For some purposes differences between place of employment and residence may create a significant element of incomparability.

The state tax studies have yielded valuable details on income in the respective states, but whether they are applicable to other states is not known. In many respects Delaware, Montana, and Wisconsin may not be typical of the country or resemble any other state. Similarly the Michigan and Minnesota studies each covers a single state with its own peculiarities.

The Financial Survey of Urban Housing and the National Health Survey were confined to urban communities, widely distributed throughout the country. Many were identical with and covered by other sources as well. The Consumer Purchases Study

was carefully planned to represent urban and rural communities in all major regions except the Southwest. However, no part of Delaware fell in the sample, no farm or urban community in Minnesota, no urban community in Michigan, and the Montana rural area covered is grouped with those of two other states in the tabulations. The only cities common to the Consumer Purchases Study, Financial Survey of Urban Housing, and National Health Survey were Atlanta, Ga; Portland, Ore; and Springfield, Mo; no one of which is in either a community property state or a state covered by special studies.

F CLASS INTERVALS

Only three income class intervals are common to all series described in this volume: 'under $1,000', '$1,000-2,000', and 'over $2,000'. By omitting one study, a '$2,000-3,000' class could be added; by omitting also Social Security data, two more intervals would be common. When such broad classes must be used, the sacrifice of information is serious. When intervals overlap, comparability may be achieved by rounding and interpolation, at the price of some error. Obviously, the larger the number and the smaller the range of class intervals the greater the ease of integration and the smaller the error introduced.

G ACCESSIBILITY OF DATA

Several studies have published their findings in great detail with carefully planned tables and explanatory text; some have made or will make additional material accessible under certain conditions to accredited students. Some have preserved schedules, punch cards, transcription sheets, and tabulating tapes. Although the individual student will hardly find it practicable to scrutinize such material, a group or agency may. If it is desired, for example, to bring together data from divers studies to construct a more complete distribution by size or cross-classify income with various socio-economic factors, it may be worth while to go to the source material for data not published or so classified and grouped as to be of substantially less value than the material in the form in which it was collected. Finally, the value of the data depends in part upon

the promptness with which they are made accessible. On this score too both the studies and the continuing sources have varied, as is to be expected in view of the differences in the size of the undertakings and the resources available. Unfortunately, it is improbable that the data from the two chief continuing sources, federal tax returns and Social Security records, can be released more promptly in the near future.

III GAPS

In the hope of facilitating the use and preventing the misuse of data, and, more important, of indicating gaps and perhaps inducing their filling, we discuss the main deficiencies of data on income distributions implied in the preceding sections. In view of the diverse uses to which the data are put, no single practicable set of specifications can be written that, even if met, would assure the solution of all problems. What may be a grave defect for some purposes may be negligible for others.

No continuous source yields a distribution by size of either family or individual income for the United States. A theoretical ideal would be a source showing for each recipient details of all major types of income and at least some types of expense and deduction items, such as those on the borderline between personal and occupational expense, the major types of nonmoney receipts, occupation and industry, residence, age, sex, education, and other socio-economic characteristics. Individuals should be so identified that they could be grouped by families according to at least one generally acceptable definition and their incomes could be traced from one period to another to give size distributions by accounting periods longer than a year. The data should be released at intervals of not more than a year, and promptly tabulated. Those published should be in great detail; those unpublished, made readily accessible. In the absence of complete coverage, a good sample would meet the chief needs. That we are far from such a goal is obvious.

A FARM INCOME

Data by which farm income could be distributed by size are meager. Legal provisions have in effect freed most farmers from

liability for income taxes and administrators have not found it worth while to try to get returns. In Delaware, where the law requires every farmer to file, coverage has by no means been complete; in Wisconsin also many farmers who should have filed have apparently not done so, and the same situation seems to prevail generally in other states that tax personal income. It is still too early to estimate the effect of the 1940 and 1941 lowering of federal filing requirements, but they will certainly widen the coverage of farmers greatly. What remains uncertain is the amount of undetected delinquency, the accuracy with which returns are filled out, the effects of the option of inventory account, and, most important, the distribution among the huge group of farm families with incomes below about $1,500. Numerous accounting problems in connection with farm incomes are still unsolved.

Farmers are exempt from the Social Security Act. Of the surveys described in Part II, two did not cover farmers, but the two that did, the Consumer Purchases and Minnesota, have supplied what are in many respects the best data we have. The former has the advantage of representing a much wider area, the latter of being more detailed. The two treated some items differently and pertained to periods at least three years apart. Though farm families were canvassed in the Michigan survey the data have not been analyzed. The source material of both the Delaware and Wisconsin studies covered farm income incompletely and the Census income questions were not designed to yield detailed data on it. These studies contributed toward filling the gap in information on farm income in the continuous sources, but probably much still remains to be learned before we shall know as much about the income of farmers as about that of industrial workers and persons well covered by income tax laws. Especially must we learn how to integrate the data from special studies with those from continuous sources.

On the outlay side, too, there are unsettled questions, such as how the cost of running an automobile shall be allocated. How much of it is a business expense? If deducted, should the urbanite's commuting fares also be deducted? The basis for distinguishing between wealth and income, discussed below, arises as in any busi-

ness enterprise but may be more vexing because of the importance of land, the paucity of carefully kept records, and, in recent years, the poor marketability of the major item of investment, the farm itself, whose changes in value can seldom be ascertained satisfactorily. In the past, it has been said, a large part of the change in the farmer's net worth was derived from the increase in the value of his farm. Should we evaluate such increments? If we do not, can we adequately portray the farmer's economic position? If we do, how will the distribution of tenant and owner-operator income be affected?

B NONMONEY INCOME

We know little about nonmoney income and its distribution. Clearly, some groups receive more in the way of nonmoney income than others. Farm families, for example, often grow a large portion of the food they consume; much of the expense, together with that of the family dwelling, is a part of the rent paid for the farm as a unit and treated as business expense or, if the farm is owned by the operator, is essentially a nonpecuniary return on the investment. Though the effects on the standard of living are direct and positive and must be taken into account for valid comparison of the family's real income or well-being with that of other families, we have few data making such comparison possible. The Consumer Purchases Study and the Minnesota Survey considered these questions but more is required. How to evaluate the items is especially perplexing; the effects of different methods should be examined.

For nonrural families the contrast between the owner-occupant and the renter introduces an element deserving further study, especially for a consistent use of income tax data, as the present treatment seems quite indefensible in principle. But how large are the values and how much is the distribution affected? The Consumer Purchases Study found in the cities where figures were obtained that the value of home-produced, home-consumed food was insignificant; nevertheless this question may profitably be examined further. Urban dwellers may enjoy forms of nonpecuniary income of which no account is taken in our income figures; many may in

effect be purchases financed from income by tax payments the rural dweller does not make, therefore no actual element of incomparability may exist. When items of real income, however, give rise to no tax or price charge, a real dilemma exists.

No study described in Part II attempted to account for imputed income from consumer durable goods other than residences. Yet the total value of other such items—autos, furniture, clothing—is huge, a figure that cannot be ignored in refined analysis and that may be significant for even more general study. At present, however, we have no basis for accounting for these amounts in distributing income by size. Omission obviously leads to understatement of total real income. But does it affect significantly the distribution by size, age, region, occupation, etc?

Similar puzzles beset attempts to evaluate services performed in the home and for oneself. Our data give no basis for adjustments that can be justified in principle. Failure to take account of this item probably leads to overstating the degree of inequality, but we do not know by how much. We are still further from satisfactory treatment of less tangible but from the viewpoint of real income no less important factors such as economic or emotional security, happy family life, congenial occupation, physical well-being, etc.

C INCOME VS. CHANGES IN THE FORM OF WEALTH

Problems of measuring income and wealth are so intertwined that they cannot be separated, but every income study that pretends to any significant refinement of the concepts must immediately define each term precisely. First of all, income must be differentiated from changes in the form of an individual's wealth. Some receipts of many if not all individuals can be considered return of capital rather than income. Some of the issues are among the most ticklish in economic theory. Though it is beyond the scope of this summary either to mention them all or explore their implications, it is pertinent to point out those no income study can escape, though it may try to ignore them. Most have arisen implicitly in connection with the income tax, perhaps because it covers more extensively than other sources individuals possessing fortunes above the average.

We know, for example, too little about undistributed profits and losses. We have totals for all corporations, and in some years they are tremendous. How would income distributions by size be affected if each stockholder's share of the undistributed gain or loss were treated as a part of his income? Advocates of the undistributed profits tax argued that the failure to distribute all profits resulted in considerable understatement of incomes in the upper ranges, but there are no published data on which estimates can be based. No more is known about who ultimately stands corporation losses. The highly unequal distribution of dividends supports the conclusion that for years in which losses or undistributed profits are substantial their inclusion in the incomes of their owners might considerably alter size distributions. This is not to say that they should be included but only to show that we do not yet have the information requisite to estimate amounts and assess their effects if differently allocated. A related problem destined to become increasingly important is the treatment of accruals on government bonds sold at a discount; the owner may, if he wishes, include each year's appreciation in his current taxable income, but most owners probably do not bother.

The question of undistributed profits and losses ties in closely with that of capital gains and losses. Should they be treated as income items, whether or not 'realized'? Some of the logical and practical issues, such as have arisen in trying to devise a satisfactory treatment for tax purposes, are beside the point here. Tax returns give some data on capital gains and losses, but they relate to merely a portion of the whole and reflect certain arbitrary decisions on inclusion and tabulation. In addition, they are affected by the highly relevant consideration that both the fact and the method or circumstances of inclusion are intimately related to the tax liability of the person reporting. Consequently, we can be sure we do not have a random sample. It is not obvious just what would be most useful. What is a 'capital' item? What is 'gain' or 'loss'? Between what points of time should they be calculated? On what basis? The treatment of inventories is a case in point, an important one. The totals may often be huge and they are distributed in anything but a uniform manner.

Some of the largest personal and family fortunes in this country have been built up from accretions to economic power most of which would never appear as a part of the owner's income in any existing source of income data. The same process may or may not be going on today. Changes in net worth due to savings from wages, interest, dividends, or rent are from items appearing in our size distributions, but not changes in net worth due to unpaid labor, fluctuations in demand or in the interest rate, discoveries, uninsured losses, and the reinvestment of, or drafts upon, profits. These items will probably never be treated satisfactorily for all statistical purposes, though progress can certainly be hoped for. Useful contributions have been made by the Wisconsin and 1936 federal studies, but the limitations of the underlying data have obviously not been overcome.

To determine the income of most business enterprises we must account for changes in inventories—in the number, size, and characteristics of the physical constituents and in their prices. No method is universally accepted. In recent years more attention has been paid to methods that offset the fluctuations in business income arising from fluctuations in prices; some, though probably few, firms have adopted one of these methods, and others have tried to create reserves out of what would be considered income to form a cushion against price declines. Depreciation raises similar problems. Large in total, it can never be estimated precisely during a period as short as a year, and again no method is universally accepted.

Data on the relation between total income, income from property, and the value of the property are few. After exploring the possibility of tabulating ratios, the 1936 Federal Study staff concluded that the entries on tax returns gave an inadequate basis for reliable results. Further exploration and experimentation would probably disclose that the apparent impasse is not hopeless. The importance of these relationships makes the present deficiency of more than passing concern.

It is often difficult to know what to deduct from gross income to get net. From some items the costs of their acquisition have been deducted, from others, not. Business income is ordinarily a net

figure, some attempt being made to deduct costs though often, as indicated above, on a basis that leaves something to be desired. The accounts and records of many small service, retail, and farming enterprises are imperfect. Probably most rent receipts are approximations; depreciation, maintenance, and insurance may be deducted, and a few careful attempts have been made to do so on some generally acceptable basis; other costs, such as interest and taxes, may be deducted from gross rent or grouped with similar items and deducted from total income rather than from the specific item to which they are attributable. It is highly likely that the smaller the income the less good the adjustments.

How are size distributions affected by the inclusion or exclusion of such semi-capital items as life insurance and endowment proceeds, pensions, gifts, and inheritances? For a numerous group, one that is probably growing faster than the population as a whole, these items constitute a main source of receipts spent on current living expenses. No source described in Part II, with the possible exception of the Minnesota Study, yields data from which their size and distribution can be estimated.

D PERSONAL EXPENSES

Equally perplexing is how to treat items on the borderline between personal expense and the cost of getting an income. Only if we could agree on a common philosophy concerning 'living' would it be possible to separate all the costs of getting an income from utilizing it. For example, should costs of going to and from work be deducted as an expense? Should account be taken of extra hazards assumed or special benefits enjoyed? How should costs of training be treated? Perhaps they should be amortized. Can costs of education be divided between those primarily for 'living' and those for 'making a living'? Should union dues be deducted? lunches? clothing? entertainment? contributions? Should any part of traveling allowances be included in income? As yet we are in no position to do more than regret that we have too few data for those who wish to adjust for these items to do so.

E TAXES

The increasing weight taxes have in determining the individual's economic position needs no emphasis. The chief difference for our purposes between taxes and many other drafts on income, such as outlay for food, is that one of the largest, the personal income tax, is directly but by no means uniformly related to size of income. An obligatory payment, it is for some economic units by far the largest item of outlay, but for others nonexistent. Consequently, to deduct it, as would seem reasonable in ascertaining what amount is available for consumption and saving, would alter the distribution by size more than would any other adjustment. Federal returns yield fairly good information, but greater detail in tabulation and publication would be useful, e.g., cross-classification of total or net income (including, if possible, tax exempt income) by net income after deduction of income taxes. States allowing deduction of the federal income tax could help by publishing a distribution of income before the tax is deducted. Surveys should be able to find out how much income tax was paid. However, most people pay the tax on one year's income from the succeeding year's receipts. Should the tax on, or that paid from, the income be shown? The answer will depend largely upon the purposes of the study.

Should the corporation income tax be treated statistically as income of the stockholders which is then taken from them as a tax collected at source? This problem is complicated by lack of knowledge about the distribution of corporate profits among persons in the various income levels and also by the fact that the corporation income tax is not levied at a single flat rate.

The proper adjustment of other taxes and their relation to income are less clear. To what extent do the taxes that are deducted as business expenses tend to diminish incomes received or to increase prices? Other taxes are imposed on consumption items, and still others on property. Which can be said to raise the cost of living? How much? There may be no way to relate tax payments to income without using detailed consumption data. To what extent have taxes been capitalized and what adjustment can be made for those

that have? These unanswered questions have significance for pressing problems of public policy.

F ACCOUNTING PERIOD

Though the calendar year is the most widely used accounting period, it is not the best for every purpose. A much longer period may sometimes be better. From a size distribution of life income some problems could be analyzed more thoroughly than from any data now available, though there is no way of knowing whether the conclusions would differ significantly from those for single years. Since, as seems probable, there is little prospect of obtaining data on life income, distributions covering a shorter period, such as one business cycle, might be made. Whether data for a period shorter than a year would be of much value is open to more doubt because many businesses are seasonal and some types of payment are lumped at certain periods of the year. Many businesses operate on a fiscal year basis, and strong arguments can be advanced for using income data for these more natural economic accounting periods. For a few large groups, such as farmers in large crop areas, a twelve-month period might be selected that accords better with economic reality than the calendar year. For other groups, however, there is such diversity that no one period preferable to the calendar year can be selected. The problems inherent in inventory valuation and depreciation as well as the relations between capital and income in general deserve mention here.

G RELATIONS BETWEEN CONTINUOUS SOURCES

We need links between income tax returns and Social Security records. The obstacles are formidable, in view of the many ways in which these two major sources of continuous data differ, and the margin of inevitable error may remain so great that the best links possible will eventually prove deplorably weak. The task may be simplified, and also made less urgent, by the recent extensions in the coverage of the federal income tax. If the Census plans mature, the tabulations should facilitate the task; for this purpose, however, it is unfortunate that the federal and most state income taxes for the year covered by the Census exempted such a large portion of

wage earners. The analysis of the Minnesota data is awaited with great interest because it is attempting under the most favorable circumstances yet realized in this country to discover whether data from tax returns and Social Security records can be spliced and if so, how.

The proportion of wages and salaries that is covered by the Old-Age and Survivors Insurance program would perhaps be revealed by comparing its annual nationwide reports with those on unemployment compensation in at least the states in which the requirements for size of firm filing are identical. The Census will provide a basis for estimating the proportion that earnings of families with 'covered' family heads constitute of total family earnings. For the sectors of the economy not covered by the Social Security program, we must resort to the Census, surveys, of which the Minnesota is most promising, and Delaware returns.

H SHIFTS IN INCOME

Our information on changes in the distribution of income over time is sketchy; e.g., between prosperity and depression. To what degree are relationships that determine the distribution stable? How rapidly and in what ways do they shift? We have little basis for deciding which of the relationships revealed by the data described in Part II will persist and which will be profoundly altered by the war and its aftermath. Do they have permanent value or are they pertinent merely for the past? Will patterns of change be traceable?

I ASSOCIATIONS AND CAUSAL RELATIONSHIPS

The difficulties of establishing causal relationships among the many variables bearing upon any of the major economic and social aspects of life are discussed in Chapter 1. So little is known about distributions of income that efforts to discover and assess its determinants can hardly be expected to bring rich rewards. Further study of the theoretical issues is urged, for not until they are clarified can we say whether they can be treated statistically or what types of data are most needed.

About the effect of many other factors on distributions of in-

come we are largely in the dark; for example, the size of the family (or economic unit), the age, sex, education, training, military experience, place of birth, citizenship, religion, color, physical condition and medical history, geographical and occupational mobility, expenditure habits, net worth (by type of assets), union and trade association, and similar affiliation of recipients, the type (and cost) of the dwelling, the size and location of the community in which recipients live. Some are obviously more directly related and of vastly greater permanent interest than others. Some can apparently be discovered from small carefully selected samples. At least a few of the more intimate relationships must be established to give the necessary 'controls', and improvements in the basic data make the sampling tool more effective. The Census tabulation program will probably include several factors, for example, size of family, age, sex, and education of head, color of head, and rent or value of dwelling unit.

Finally, we need devices to test and analyze the significance of findings. What, for example, is meant by 'inequality of income distribution'? How can the essential elements of an income distribution best be portrayed?

IV Constructing an Estimate for the United States

One national estimate of the distribution of income by size has been constructed from data described in this volume.[2] Prepared by the Consumption Research Staff of the National Resources Committee under the direction of Hildegarde Kneeland, it relates to the twelve months July 1935 through June 1936. It was made by splicing data from the Consumer Purchases sample study for family and individual incomes under $7,500 with federal tax return data for taxpayers with incomes of more than $5,000. The numerous problems involved in this splicing process, which necessitated numerous more or less arbitrary assumptions, have been described in Volume Three of *Studies in Income and Wealth*.[3] To expand the income sample collected by the Consumer Purchases Study to cover the entire population was also an intricate task, despite the care taken in planning the study to obtain a representative sample of the entire population. The methods used to splice the two

bodies of data and expand the Consumer Purchases Study sample have been criticized, one critic concluding that the income distribution is seriously distorted.[4]

Data from more sources are available for 1936 than for any earlier year. Some of the studies for 1936, particularly of federal and state tax returns, had not been made when the National Resources Committee estimate was prepared. With them, especially if relationships among types and distributions of income were established from them, a better national estimate for 1936 might be made.

However, with the tabulation and release of information from the 1940 Census, far more information on the distribution of the largest element in income, namely, wages and salaries, will be available for 1939 than for any prior year. Since 1939 is likely to be regarded in the future as a relatively normal, prosperous, peacetime year, it will undoubtedly be extensively used as a 'benchmark' for studying changes wrought in the national economy by both war and postwar adjustments. It is therefore highly desirable that an effort be made to prepare as good an estimate as possible of the national distribution of income by size for 1939. We recognize that the gaps in information are serious. Nevertheless, a better estimate can be prepared for 1939 than for any preceding year, and it may be long before as many data are available for another peacetime year.

The incomparability of existing income data suggests that preparation of a reasonably good national estimate of the distribution of income by size for 1939 is a complex task, requiring special tabulations of the data collected, careful analysis of relationships among various groups of data, and astute judgment in combining data from various sources and in the application of equations of relationship to convert certain types of series into other types.

From the 1940 Census we can estimate, for 1939, distributions of wages and salaries under $5,000 for both individuals and families, and also the number of individuals and families whose total income differs by $50 or more from their wages and salaries. If we scrutinize the relationships between wages and salaries, on the one hand, and income from other sources, on the other, as shown by

the Consumer Purchases Study, the Minnesota Field Study, and the studies of federal returns for 1934 and 1936, we may learn how to convert the distribution of wages and salaries, as derived from Census data, into distributions of total income, for both individuals and families. However, business fluctuations are known to affect aggregate wages and salaries and income from other sources somewhat differently, and the year to year changes in the distribution of the two may be sufficiently different to cast doubt upon the validity of assuming that 1934 or 1936 relationships hold for 1939 data. For this reason tabulation and analysis of sources of income reported on federal returns within the income range of approximately $2,500 to $8,000 (the range for which wages and salaries of less than $5,000 are important) would be desirable. This could be a sample study, but the sample should be drawn carefully, since the relations between wages and salaries, on the one hand, and total individual or family income, on the other, vary from region to region and are closely related to occupation.

For incomes above $5,000, both individual and family, federal returns must be relied upon mainly. Yet the tax data as tabulated and published have the disadvantage that husband and wife returns are in part joint and in part separate returns. To overcome it, husband and wife returns must be matched to give as close an approximation as possible to family income. To calculate individual income from joint returns is not feasible, since probably few returns specify whether each item is the income of the husband or of the wife.

The adjustments in net income necessary to obtain an income concept reasonably comparable with the concept used for incomes below $5,000 are such that it is easier to make an entirely new tabulation, based on total income as reported on the return. The most common differences between total and net income arise because of deductions for interest on personal indebtedness, taxes, and contributions, none of which are deducted from family incomes under $5,000. The two important adjustments that would be necessary in tabulating total income as reported on tax returns are a deduction for losses and the addition of income from tax exempt securities.

To provide for adequate overlap in the splicing of tax returns with Census data, it is desirable for tabulations to be made for all returns (or an adequate sample) showing the combined husband and wife income, and the income of single individuals receiving $4,000 or more, and preferably $3,000 or more.

Neither the income questions on the 1940 Census nor the federal returns will yield much information concerning the distribution of income among one large segment of the population, farmers. An attempt should be made to fill this gap with the help of Census information on the value of farm products, together with studies of relationships between it and farm expenses or total farm family income, based on farm management and farm family income studies. Several items on the agricultural schedule of the Census will probably have to be tabulated before size distributions of the value of farm products can be made.

Other sources of income data for 1939 that might be used to check and supplement the estimates prepared by the methods described above are Social Security records of wages and salaries, and state tax returns in states that in this year had a lower exemption limit than the federal.

Additional sets of estimates that should be prepared for 1939 are for direct taxes paid by families and individuals in the various income classes, and of the size distribution of income after payment of taxes. Most direct taxes other than federal income are probably reported on federal returns, though many persons, particularly with incomes under $10,000, probably do not deduct all their taxes. Analysis of the number of tax deductions and comparisons with data from state returns might indicate the proportion of unreported taxes on federal returns.

By subtracting the sum of taxes reported as paid, the estimate of unreported taxes, and the federal income tax liability from the adjusted total income, a distribution of income after deduction for taxes can be estimated roughly. This would not, of course, be precise for 1939, since it would allow for taxes other than federal income taxes paid in 1939 and therefore levied in part on incomes received in 1938, and for federal income tax liability based on income for 1939 but not paid until 1940.

It is hardly likely that personnel can be had during the war to carry out these suggestions for preparation of a national estimate of the distribution of income by size for 1939. However, the records should be preserved and the study planned so that it can be started immediately after.

NOTES

[1] Tabulations from the 1939 Census were not available when this chapter was prepared. The Consumer Purchases Study was not used here because neither an unweighted nor weighted combination of the data collected in the various localities has been published. The national estimate published by the National Resources Committee, and based in large part upon the data collected by the Consumer Purchases Study, is too far removed from the original data to be used in this comparison of data derived from diverse sources.

[2] National Resources Committee, *Consumer Incomes in the United States, Their Distribution in 1935–36* (Washington, 1938).

[3] Enid Baird and Selma Fine, The Use of Income Tax Data in the National Resources Committee Estimate of the Distribution of Income by Size.

[4] Rufus S. Tucker, The National Resources Committee's Report on Distribution of Income, *Review of Economic Statistics,* XXII, 165-82, Nov. 1940.