Volume Title: Shares of Upper Income Groups in Income and Savings

Volume Author/Editor: Simon Kuznets, assisted by Elizabeth Jenks

Volume Publisher: NBER

Volume ISBN: 0-87014-054-X

Volume URL: http://www.nber.org/books/kuzn53-1

Publication Date: 1953

Chapter Title: Reliability of Estimates

Chapter Author: Simon Kuznets, Elizabeth Jenks

Chapter URL: http://www.nber.org/chapters/c3070

Chapter pages in book: (p. 435 - 468)

Chapter 11

Reliability of Estimates

We cannot measure the probable errors in our estimates directly because our basic data are either byproducts of tax administration or products of censuses, subject to all the imperfections of social records. Some defects are obvious and the adjustments discussed in preceding chapters were designed to correct for them as far as possible. But after all these adjustments, errors inevitably remain, and we are faced with the difficult task of appraising them. This discussion of the reliability of our estimates must necessarily be incomplete and inconclusive. It can be handled under two heads. First, the published data for returns with net income under $5,000 are estimates based upon samples, not actual tabulations of returns. Second, as noted repeatedly, returns with net income, whether over or under $5,000, may be subject to biases due to underreporting or evasion. It is this second problem that will occupy us through most of this chapter. We do not consider here the possible errors in our income denominators, i.e., the countrywide totals, for they are discussed in the publications dealing with national income estimates.

1 *Errors in Sampling Returns with Net Income under $5,000*

The sampling process by which returns with net income under $5,000 are estimated has varied, the most important change being in 1928 from drawing a fairly constant number of Form 1040 and Form 1040A returns filed in each Collector's District to drawing a constant proportion of returns (with a fixed absolute number as the minimum).[1] While the sample throughout was large enough to reduce purely random errors to very narrow limits, it may be worth while to refer to the two published checks which indicate their size and character.

Form 1040A returns filed for 1928 were treated by two methods: first, the items were estimated from the samples in the usual way, i.e., the Bureau of Internal Revenue weighted the samples reported by Collectors' Districts; second, some items were tabulated directly from the returns. For net income the sample estimate exceeded the tabulated total by 4.2 percent;

[1] For a brief description of the sampling procedures see the article by Crum, Harriss, and Keith in *Studies in Income and Wealth, Volume Five* (NBER, mimeo., 1943), Part II, pp. 1-44 to 1-46.

Table 103

Estimates from Samples Compared with Tabulations of Data from Returns
Net Income Classes under $5,000, 1934

		NET INCOME CLASSES					
		Under $1,000	$1,000-2,000	$2,000-3,000	$3,000-4,000	$4,000-5,000	Total under $5,000
		(1)	(2)	(3)	(4)	(5)	(6)
	Number of Returns						
1	Sample estimate (000)	320.5	1,608.1	980.7	533.4	229.2	3,671.8
2	Tabulation (000)	300.5	1,579.6	903.8	515.7	214.9	3,514.7
3	(2) as % of (1)	93.8	98.2	92.2	96.7	93.8	95.7
	Net Income						
4	Sample estimate ($ mil.)	211.1	2,277.7	2,467.9	1,821.2	1,018.1	7,796.0
5	Tabulation ($ mil.)	204.0	2,218.6	2,280.3	1,762.3	954.1	7,419.3
6	(5) as % of (4)	96.7	97.4	92.4	96.8	93.7	95.2
	Salaries, Wages, etc.						
7	Sample estimate ($ mil.)	158.4	1,944.7	2,005.1	1,435.6	756.5	6,300.2
8	Tabulation ($ mil.)	141.1	1,889.1	1,780.8	1,355.1	676.2	5,842.4
9	(8) as % of (7)	89.1	97.1	88.8	94.4	89.4	92.7
	Income from Business & Partnerships (profits minus losses)						
10	Sample estimate ($ mil.)	73.4	273.0	421.7	329.9	196.1	1,294.0
11	Tabulation ($ mil.)	65.0	246.1	382.1	301.6	189.1	1,183.9
12	(11) as % of (10)	88.6	90.1	90.6	91.4	96.4	91.5
	Dividends						
13	Sample estimate ($ mil.)	28.0	75.7	95.5	97.1	84.3	380.6
14	Tabulation ($ mil.)	27.7	65.9	83.7	87.8	80.6	345.7
15	(14) as % of (13)	99.1	87.0	87.6	90.4	95.6	90.8
	Interest						
16	Sample estimate ($ mil.)	42.7	124.9	123.7	97.1	67.3	455.8
17	Tabulation ($ mil.)	39.3	104.2	104.2	82.9	59.9	390.5
18	(17) as % of (16)	92.1	83.4	84.2	85.4	88.9	85.7
	Total of above Income Receipts						
19	Sample estimate ($ mil.)	302.6	2,418.2	2,646.0	1,959.6	1,104.2	8,430.6
20	Tabulation ($ mil.)	273.2	2,305.3	2,350.9	1,827.4	1,005.7	7,762.5
21	(20) as % of (19)	90.3	95.3	88.8	93.3	91.1	92.1

Line

1, 4 *Statistics of Income, 1934,* Part 1, Basic Table 7.

7, 10, 13, 16 *Source Book.*

2, 5, 8, 11, 14, 17 *Statistics of Income Supplement Compiled from Federal Income Tax Returns of Individuals for the Income Year 1934,* Section II, Table 1.

for total income, by 2.5 percent; for dividends, by as much as 24.6 percent; and for wages and salaries, it fell 1.3 percent short.[2]

The large percentage discrepancy for dividends is due chiefly to the fact that persons qualified to file on Form 1040A because their income was chiefly from salaries and wages were not required to report their dividends. But all these percentage discrepancies are important for our analysis only

[2] *Statistics of Income, 1928,* pp. 19-20.

when placed in the more comprehensive picture of all returns and of the cumulated totals of income reported on them. A 10 percent error in income on Form 1040A returns may mean an error of less than 1 percent in the cumulated total of the top 5 or 7 percent of the population, and in the estimated share of the 4th and 5th percentage band merely a shifting of some returns and an improper inclusion of Mr. Jones' instead of Mr. Smith's return. We cannot illustrate the second part of this statement, but from the fact that in 1928 Form 1040A returns accounted for only a fifth of the net income reported on all returns, a fifth, too, of total income, less than a half of wages and salaries, and about a fortieth of dividends, it is evident that in terms of the cumulated total for the top 5 or 7 percent the sampling errors are negligible.

A more significant test can be made for 1934 when returns were tabulated by net income class. While the tabulations covered a somewhat smaller number of returns than the estimates made by the usual sampling procedure, we can assume that the returns not tabulated were, on the whole, similar to those tabulated and that the slightly smaller coverage does not seriously affect the comparison (Table 103).

For all net income classes combined (col. 6) the estimates and the tabulations agree fairly well, especially when we adjust for differences in coverage and thus in effect compare the percentage lines alone (3, 6, 9, etc.). Interestingly enough, as in the comparison for 1928, the estimates are somewhat larger than the tabulations, the difference in the income items being larger on the average than the difference in the number of returns. The percentage differentials are not substantial, however, even for the separate income classes. Only in the case of items whose amounts are relatively small, such as dividends and interest in the lower net income classes, are they at all sizeable. And when we consider their effects on the cumulated totals or on the shares of countrywide totals established by the interpolated lines, it is hard to see how they can be anything but negligible.

2 *Omissions from Income Tax Data — General Considerations*

An important problem in using income tax data is the error caused by what is plausibly assumed to be a tendency for persons to evade reporting or to understate their income. For lack of specific information, we cannot do much about omissions but we can distinguish two types: cases in which a person fails to file a return, and the tax data do not cover him, his dependents, or his income; and cases in which a person files a return but understates his income, so that while he and his dependents are counted fully his income is understated.

In the first type of omission, designated here 'nonfiling', the failure to report may be legal or illegal. In either case it affects our estimates only if

the persons who do not report have a bigger per capita economic income than the lowest per capita income in our lowest percentage band. Thus, if Mr. Jones has a larger economic income than the lowest income person within the 4th and 5th (or 6th and 7th) percentage band but does not file a return, our estimates will be affected since the inclusion of Mr. Jones and his income would have raised the share of income assigned to that percentage band. But if Mr. Jones' income is equal to or less than the lowest under consideration, his failure to file does not affect our estimates.

While there is no evidence from which we could estimate even roughly the extent of nonfiling, we can assume on several grounds that its effect upon our analysis is small. First, since the law is inclusive in its listing of income sources subject to tax, there is little legal excuse for not filing when one receives genuine economic income. In fact, from the standpoint of income sources, the most obvious reason for nonfiling is the illegal character of the activity from which the income comes. But in such cases we could not consider the returns as representing economic income, even though it would be interesting to take them into account in any calculation of disposable income.

Second, in cases where net income is too small to be subject to tax, nonfiling is limited by the requirement since 1921 that gross income equal to or exceeding a specified amount (ranging from $5,000 in 1921-39 to $500 in 1947) be reported regardless how small net income is; and even more by the practice of Collectors' Offices of checking on a person who has once filed a return. Hence, only persons who have never filed or whose gross income is so small as to escape detection by tracing at the source are likely to get away with nonfiling.

Third, because the direct tax burden on the lower income classes is light, persons in these classes have little economic incentive for nonfiling. On the other hand, when the income is large and the income tax burden appreciable, the checking machinery is likely to be more active and the penalty for nonfiling heavier.

Finally, nonfiling affects our analysis only as far as incomes of nonfilers exceed those of filers in the upper percentage bands. Inasmuch as the very large income units are least likely to be among the nonfilers, the proportion of nonfilers will always be much higher than the proportion of income omitted, i.e., than the relative weight of the excess of economic income of nonfilers over the economic income of the lowest units within the upper percentage bands.

Almost all the factors mentioned as limiting nonfiling and its effects upon our estimates apply also to underreporting net income. Another factor is that underreporting may take the form of overstating deductions

rather than of understating gross income; and whenever this is the case, our reinclusion of these deductions fully removes the resulting bias.

However, returns have another source of bias (besides that of underreporting amounts that should be reported): the various items legally exempt both from tax and from reporting. Some of the more important (e.g., interest on certain government obligations, and prior to 1938, salaries of state and local government employees) were adjusted for (even though incompletely in the case of interest). But even a casual perusal of tax manuals or of advisory tax services will bring to notice many other exempt items. Some are exempt because while in a category that suggests current income flow, they are in fact distributions of capital, e.g., dividend distributions 'in liquidation'; others are genuine income flows, e.g., active service pay, up to a specified amount, of armed forces abroad (since 1942). Also, some of the transfers that we included in our countrywide total of employee compensation are, in part or in full, exempt from tax and from reporting. It would be difficult, and not very useful, to list all legal omissions. One need say merely that when exemptions are allowed because the items are not true income flows but capital distributions the items should not be included from our standpoint either; and exempted items included in our denominator for countrywide income are relatively so small at upper income levels as to be negligible.

All this does not mean that the tax data as used in our estimates are free from the downward bias due to nonfiling or understatement; but apparently it is circumscribed by various factors and is likely to be relatively much less among the upper income classes. Moreover, since the effect of nonfiling or underreporting equals the difference between the omitted or understated return and the first return just above our lowest partition line (or the first return, not included, just below the line), the larger the 'pool' of tax return population below the line, the more reliable are our estimates of the shares of the bands above it likely to be. If the pool is big it constitutes a large reserve for compensating or reducing errors of nonfiling or underreporting.

Since biases still remain in the data, even at the upper levels, we may ask what types of income are most likely to be affected, and how they are likely to change over time. The answers can be nothing more than conjectures but we should at least state them as leads to possible interpretation.

As among various types and sources of income one would assume that incomes reported at the source by the distributing agencies and incomes substantial enough not to be deemed negligible by either the recipients or the tax authorities would be those least likely to be nonreported or understated. This means that compensation of employees, particularly fulltime

employees, and large property incomes in the form of dividends and interest are likely to be reported fully. Net income of entrepreneurs, rent, and small amounts of dividends and interest are among the income categories most likely to be understated, or, when quite small, omitted. If for both entrepreneurial income and rent the understatement takes the form of exaggerating deductions, our estimates escape the bias. Nevertheless, all small receipts and incomes arising from individual enterprise, such as entrepreneurial income and rent, are likely to be most affected by nonfiling and underreporting. This conclusion reenforces the one already advanced, that omissions are likely to be relatively more important at the lower income levels, for it is here that subsidiary incomes, though small absolutely, constitute a larger proportion of total income; and here also that entrepreneurial income and rent are so small and recorded under such conditions – primitive bookkeeping, etc. – as to make underreporting easy, and checking by tax authorities difficult and expensive.

Perhaps more important is whether there are short term changes in the relative size of omissions. The answer is even more conjectural than the answer to the other questions but the following considerations seem relevant.

First, during cyclical expansions, when incomes increase, there is more incentive to understate incomes or not to file because the possible saving in tax payments is larger. This presumably would be true even though purchasing power and the marginal value of the dollar to income recipients declined. During contractions, on the contrary, incomes may fall well below the taxable point; then the incentive to understate or not to file is weaker. To the degree this observation is true, the size of omissions would fluctuate cyclically.

Second, according to the evidence for 1919-38, the proportion of employee compensation in aggregate payments (but not that of salaries, taken alone) and the proportion of dividends tend to move on the whole with business cycles. Positive conformity is somewhat less marked for the proportion of entrepreneurial income, and in even greater contrast, the proportions of rent and of interest tend to move counter to business cycles.[3] The first two income types are least subject, the last three, most subject to nonreporting and understatement. Consequently, cyclical shifts in the composition of income by type would, in and of themselves, make the relative magnitude of omissions run counter to business cycles.

Third, the proportion of population covered by income tax returns varies greatly: it is higher during periods of expansion and high income levels and lower during periods of contraction and low income levels. In estimates

[3] *National Income and Its Composition, 1919-1938,* Table 32, p. 251.

that cover a constant top proportion of the population, the lowest partition line would therefore be nearer the bottom of the tax return population 'pool' during contractions than during expansions. As pointed out above, the larger the pool of tax return population below the lowest partition line, the smaller is the effect of omissions on the estimates of shares above it. This size of the tax return population is, then, another factor that would, in and of itself, make the relative magnitude of omissions move counter to business cycles.

The net effect of these factors cannot be measured. We cannot tell whether, on balance, the bias in our estimates is larger or smaller during cyclical expansions than during contractions. All that can be said is that there are influences in both directions and that this might limit short term variations in the relative understatement in the estimates of the type made here. Until further information becomes available, we must leave the reader with this unsatisfactory uncertainty.

3 *Omissions from Income Tax Data — Comparisons with Samples*

Though the possible downward bias in the income tax data, and particularly in our estimates of the shares of upper income groups is probably small, it would be highly desirable to test our estimates against independent data and, by referring to some empirical observations, gain a somewhat more tangible idea of the size of the possible error.

Two bodies of data come to mind. The first would be supplied by the audit of income tax returns. If each and every type of return were equally represented in the audit, if returns in all income classes were checked with the same meticulousness, and if the results were available so that one could study the reported under- and overstatements by types of income and of deduction for a series of years, we would have an adequate basis for measuring the bias in the published unaudited data. But for obvious practical reasons, audits have been confined largely to returns reporting big incomes, have paid more attention to returns that on the surface gave some evidence of noncompliance, and their results have never been released to the public or summarized in enough detail or for a sufficient number of years to permit any satisfactory conclusions. Not until auditing is reorganized in accordance with plans projected for 1948 returns and has been continued for several years will its results be adequate to measure the bias.[4]

[4] See the Use of Audit Reports for Correcting Statistics of Income Compiled from Individual Income Tax Returns by J. R. Turner, submitted to the April 1949 meeting of the Conference on Research in Income and Wealth. Even with the new data, the extent of underreporting in 1919-38 will be difficult to measure. The increase in the relative income tax burden in recent years may well have made for more underreporting. See, however, Section 5 below.

The second body of independent data comprises the sample studies of income by size. Obviously, for our purposes we need countrywide samples, or at least ones covering the nonfarm population, and ones that utilize information independent of tax returns. During the period under analysis there have been several such studies of income size distributions for the entire population. Had their coverage been adequate, comparison with them would test the accuracy of our estimates. Unfortunately, they suffer from three major defects. First, all, no matter how much effort has been devoted to this problem, underrepresent the top income levels – those with which we are most concerned. Second, almost all seem to miss a sizeable proportion of income in addition to the shortage at the top income levels, presumably because of the difficulty of getting accurate amounts by questionnaire or interview if considerable time has elapsed since the income was received. Third, the internal structure of their distribution is distorted in ways hard to gauge. In other words, we are uncertain whether, allowing for the shortage at the top levels, there is relative over- or underrepresentation below, and whether income shortages are bigger at the low, the intermediate, or the levels just below the top.

In view of these major defects of the samples and numerous minor ones, one might well doubt the wisdom of making any comparisons. Indeed, at some stages of preparing this book, I was inclined to discard those already made. They are nevertheless included partly because they do provide some rough check on our estimates, partly because if they were not, other researchers probably would make comparisons – with results that might well be misinterpreted.[5]

a *Comparison with NRC Distributions for 1935-36*

The size distributions of the Study of Consumer Purchases for 1935-36 are available in three versions: the original, published by the NRC; a revision, confined to the distribution of money income, by the OPA, utilizing revised data on number of families, number of individuals, and the aggregate money income of each; and Rufus S. Tucker's modification of the original distributions.

All three versions combine federal income tax data with the consumers' field survey and other data. In calculating the full published version the NRC confronted directly the problem of nonfiling and underreporting.

[5] We do not compare our estimates with the Brookings estimate for 1929 largely because it contains, at the top levels, little evidence independent of the tax returns themselves, partly because it is affected by the inclusion of capital gains and losses. It could be, and was, used for associating savings-income ratios with income levels; but it is of little value in any comparison of size distributions where income tax data are one term of the comparison.

On the basis of "tentative estimates advanced by several authorities who were consulted", it assumes that everyone at family income levels above $20,000 filed a return, and makes an adjustment for failure to file amounting to a 5 percent increase in the number of families with $15,000-20,000 incomes; 15 percent in the number with $10,000-15,000 incomes; and 25 percent in the number with $5,000-10,000 incomes. There are adjustments also for understatement by families filing returns, yielding increases of 15 percent of their aggregate income in the income classes just listed, and of 10 and 5 percent respectively in the $20,000-25,000 and $25,000-50,000 classes. Similar adjustments were made for nonfiling and underreporting by single individuals.[6]

The reason the NRC allowed for nonfiling and underreporting by the upper income classes alone is that for the lower classes independent field sample and other data were available. Its report implies, in fact, that nonfiling and underreporting are likely to be relatively greater among the lower classes, and presumably the entire range is adjusted for possible omissions. Comparison will therefore indicate the understatement in our estimates attributable to the biases due to nonfiling and underreporting.

The NRC distributions of single individuals and of families by income per unit and the number and income of the seven types of institutional resident (shown only in the aggregate, not by income class) had to be converted to one distribution by income per person. This involved converting income per family to income per capita for each size of family group in each size class of family income (App. 6, Sec. A) and calculating income per capita not only for single individuals for each income class but also for each type of institutional resident. Following the order of these per capitas, the income for the three groups was cumulated into one distribution, as was the number of persons. We then drew the upper percentage lines in the customary fashion and derived the shares of the upper groups (Table 104, line 3).

The total income accounted for by the NRC distributions is about 2.8 percent short of the total we used. If we were to assume that the income omitted is distributed proportionately to the income reported, the shares would remain as they are in line 3. But as the shortages may well be largely in the income ranges below the top, we reduce the shares in line 3 by the relative shortage of the NRC income total (line 4).

In converting the OPA version to a per capita basis, we had to utilize

[6] *Consumer Incomes in the United States*, pp. 84 and 87. For a more detailed discussion, see article by Enid Baird and Selma Fine in *Studies in Income and Wealth, Volume Three* (NBER, 1939), pp. 149-203, and comments by A. J. Goldenthal, pp. 204-14.

Table 104

Percentage Shares of Upper Income Groups Based on Income Tax Data Compared with Those Based on NRC Distributions, 1935-1936

	ESTIMATES BASED ON	SHARE OF GIVEN PERCENTAGE BAND: Top 1	2nd & 3rd	4th & 5th	Top 5	6th & 7th	Top 7
		(1)	(2)	(3)	(4)	(5)	(6)
	Income Tax Data						
1	Basic variant, total population	12.6	6.5	5.0	24.0	4.5	28.6
2	Economic income variant (basic variant adj. for imputed rent, compensation of nonfederal employees, family status, & unwarranted inclusions & deductions), total population	14.0	8.2	6.6	28.8	(5.4)	(34.2)
	NRC, as Published						
3	Unadjusted for income base	14.1	9.6	6.3	30.0	5.0	35.0
4	Adjusted for income base	13.7	9.3	6.1	29.2	4.8	34.0
	NRC-OPA Revision: Money Income						
5	Unadjusted for income base	13.6	9.7	6.4	29.7	5.0	34.7
6	Adjusted for income base	12.9	9.2	6.1	28.2	4.7	32.9
	NRC-Tucker Modification						
7	Unadjusted for income base	12.0	8.0	5.6	25.6	4.6	30.2
8	Adjusted for income base	11.6	7.8	5.5	24.9	4.4	29.3

Figures in parentheses are calculated by multiplying line 1 by 1.20, the ratio of line 2 to line 1 for the top 5 percent.

many of the detailed data for family size as published originally (App. 6, Sec. B). The results, unadjusted for the income base, appear in Table 104, line 5. The total income covered is 5.2 percent short of the total we used, largely because income in kind was omitted. Since the latter is received chiefly by the lower income groups, there is some reason to assume that almost all of the missing income is at levels below those covered in line 5, and line 6 is computed on this assumption.

Both the NRC and NRC-OPA distributions may well have allowed too much for nonfiling and underreporting, and other steps in their procedure may have led to an overstatement of the shares of upper income groups. Mr. Tucker has challenged the NRC distributions, and his argument that understatement usually takes the form of exaggerating deductions rather than of concealing or underreporting receipts is particularly telling in the present connection.[7] If he is right, our reinclusion of deductions obviates the need of allowing for this particular element in understatement. Tucker

[7] For an analysis of the NRC distributions see Tucker's articles in the *Review of Economic Statistics*, Nov. 1940, pp. 165-82, and Feb. 1942, pp. 9-21, and in the *Journal of the American Statistical Association*, Dec. 1942, pp. 489-95; this third article is especially useful.

criticized the distributions on other grounds, and constructed a revised distribution of families and of single individuals in which most of the biases in the NRC estimates, as diagnosed by him, are eliminated.

To compare our estimates with those obtained by using Tucker's distributions (Table 104, line 7), we first converted his distribution of families to persons by multiplying by the number per family by income classes as given in the NRC study, then estimated total income, for families and for single individuals separately by multiplying the number of persons by their per capitas, the latter derived from the NRC distributions. Finally, number and income were cumulated from the top down by the order of the per capitas, and the upper percentages interpolated (App. 6, Sec. C). The adjustment for the difference in income bases, identical with that used in passing from line 3 to line 4, yields line 8.

With the shares of upper income groups derived from these three versions of the size distributions for 1935-36 we compare those for the basic variant for total population (line 1) and those for the economic income variant, i.e., after the adjustments for imputed rent, compensation of nonfederal employees, family status, and the maximum effects of unwarranted inclusions and deductions (line 2). Since the adjustments were not, and for some items could not, be made for the bands below the top 5 percent, columns 5 and 6, line 2, are rough approximations and may well underestimate the shares.[8]

First, the shares in the basic variant are smaller than those estimated from the original NRC distributions; and, except that for the top 1 percent, somewhat smaller than the shares estimated from Tucker's distributions. However, the shares in the economic income variant, the latter conceptually more comparable with the original NRC estimates, are, on the whole, the same size as those based on the NRC distributions, and significantly larger than those based on Tucker's. If comparison with the NRC data is valid, the shares in our economic income variant apparently do not contain any significant underestimate even though we do not make any allowance for nonfiling or underreporting.

Second, this conclusion is, on the whole, true whether we adjust the shares estimated from the NRC distributions for a smaller income base, i.e., whether we use line 3 or 4, 5 or 6. However, the agreement is closer if we make the adjustment.

Third, the share of the 2nd and 3rd percentage band derived from the NRC distributions is distinctly larger, and that of the 6th and 7th per-

[8] We calculated these entries by raising the share in the basic variant a fifth, the relative difference between the two variants for the top 5 percent. But for the 4th and 5th percentage band the adjustments raise the share of the basic variant a third.

centage band somewhat smaller than that of the corresponding band of the basic and economic income variants. However, it is doubtful that these differences have much significance.[9]

b *Comparison with BLS-BHE Survey for 1941*

For the full year 1941 and the first quarter of 1942 the Survey of Spending and Saving in Wartime, undertaken jointly by the BLS and the BHE, yields countrywide distributions by size that can be used in our comparisons.[10] Since the data for the first quarter of 1942 may be subject to larger error, and besides cannot be compared with estimates for an entire year, the comparison is confined to 1941.

As in all other cases, the income distribution among consuming units (families and single individuals), had to be converted to a per capita basis (App. 6, Sec. D). Unlike the NRC distributions for 1935-36, the data used here exclude institutional residents, but the omission is not important for the analysis.

Much more important is the fact that the published distributions for 1941 present the sample results as they stand, adjusted for underrepresentation at the top income levels by means of a Pareto curve but not supplemented by income tax data. Hence, unlike the comparison in Table 104, that in Table 105 does not contain any element of spuriousness: the two bodies of data are independent in the sense that they come from completely different sources.

Since these sample data, as published, are unadjusted for income coverage, they are much more short of the totals employed in our analysis than the NRC distributions for 1935-36 – about 13 percent compared with 2.8 percent. The results of the comparison, therefore, depend in large degree upon assumptions concerning the effects of the shortage at different income levels.

[9] The excess in the 2nd and 3rd percentage band may well be due to the use of a rather large top income class in the NRC distributions. Because size of family data were not available for the detailed income classes at the top, we had to use a large open-end class ($10,000 and over; see App. 6, Sec. A) and a single class mean which, divided by the number per family, yielded the per capita income for that class. Such a class mean is adequate for cells that, in fact, do enter the top 1 percent. But for the cells entering the 2nd and 3rd percentage band, it is too high. Clearly, if in our conversion of the NRC data to a per capita basis we had used more detailed income classes at the very top, the calculated share of the top 1 percent would have been somewhat larger and that of the 2nd and 3rd percentage band somewhat smaller.

[10] For a detailed description of the data see Rural Family Spending and Saving in Wartime (Department of Agriculture *Miscellaneous Publication 520*) and Family Spending and Saving in Wartime (BLS *Bulletin 822*).

Table 105

Percentage Shares of Upper Income Groups Based on Income Tax Data Compared with Those Based on the BLS-BHE Survey, 1941

Percentage Band	Estimates Based on Income Tax Data		Estimates Based on BLS-BHE Survey	
	Basic variant, total population	*Economic income variant, total population*	*Unadjusted for income base*	*Adjusted for income base*
	(1)	(2)	(3)	(4)
Top 1	11.4	12.3	6.0	5.2
2nd & 3rd	6.3	7.5	7.6	6.6
4th & 5th	4.2	5.8	5.8	5.1
Top 5	21.9	25.7	19.5	16.9
6th & 7th	3.5	4.4	4.9	4.3
Top 7	25.4	30.1	24.4	21.2
8th-10th	4.6	5.5	6.2	5.4
Top 10	30.0	35.5	30.6	26.6
11th-15th	7.0	(8.3)	9.2	8.0
Top 15	37.0	(43.9)	39.9	34.7
16th-20th	6.6	(7.9)	7.5	6.5
Top 20	43.6	(51.7)	47.3	41.2

Figures in parentheses are calculated by multiplying column 1 by 1.18, the ratio of column 2 to column 1 for the top 10 percent.

One possible assumption is that the shortage is distributed proportionately among all income levels. The shares in column 3 would then be the true shares of the successive groups from the top. But the share of the top 1 percent in column 3 is only 6.0 percent of total income; in our economic income variant (col. 2) it is over twice as large. There must obviously be underrepresentation at the top levels, which must be responsible in part for the over-all income shortage. The implication is that not enough units were included at the top levels, not that the filing units underreported their income. It is perhaps reasonable to assume that the rest of the shortage is due less to underrepresentation of units above the average (or, what is the same thing, overrepresentation of units below the average) than to understating of income by units that did report. In surveys conducted after the income year, which obviously can neither compel the reporting unit to provide full information nor minutely examine the preceding year's income of each covered unit, omissions are highly likely. The important question is whether such underreporting is more significant at upper than at lower income levels.

We assume that underreporting is both more prevalent and relatively more significant at lower income levels – for several reasons. First, underreporting is likely to be greatest among small independent entrepreneurs,

particularly farmers – units that are, on the whole, likely to be well below the lowest partition line, 20 percent, in Table 105. Second, casual and supplementary earnings or small property incomes are likely to be more prevalent at the lower levels – and these are the types of earnings most subject to underreporting. Third, the housewife or other household respondent at the lower levels in a survey is likely to be less informed and hence give less complete information than those reporting at the upper levels. Finally, small amounts are the ones most likely to be overlooked in reporting, and their relative weight in total income is greater at the lower levels. In short, at the upper levels the weakness of the survey tends to lie in underrepresentation, whereas underreporting is more likely at the lower levels.

If this reasoning is accepted, underreporting at the upper levels by units that did report is sufficiently small to be neglected. Hence we adjust column 3 for the full relative disparity between the income bases, getting shares that are a much better approximation to the true level (col. 4). These adjusted shares are, however, a bit on the low side, partly because there may be some underreporting at the upper levels but largely because correction for underrepresentation at the top 1 percent would move all the percentages about a half or three-quarters of a percentage down the population array. For example, if we allow for the omission of a top 0.5 percent, the percentage band now marked 2nd and 3rd would become the 2.5 and 3.5 percentage band; and the new 2nd and 3rd percentage band would receive a share somewhat smaller than the present. Nevertheless, the entries in column 4, adjusted upward a couple of digits beyond the decimal, are probably the fairest approximation we can derive from the 1941 Survey.

With column 4 we compare columns 1 and 2. Because the coverage of the tax return population is not wide enough to permit us to calculate the economic income variant below the 10 percent level the entries in column 2 for the percentage bands below that level are rough approximations.

On the whole, the conclusions are quite similar to those yielded by the comparison for 1935-36. Except for the top 1 percent, our estimates in the basic variant are on the low side, by from about one-tenth to three-tenths. However, the shares in the economic income variant are close to the properly adjusted shares from the Survey, beginning with the 2nd and 3rd percentage band and extending all the way down to the 16th-20th percentage band. One could hardly say that the Survey data confirm in any genuine sense our estimates based on income tax data: if our assumption that underreporting is more common at lower income levels is valid, one could just as fairly say that except for the top 1 percent our estimates confirm the Survey data. Perhaps what is most important, there is no evidence of any downward bias in the shares of upper percentage bands in the economic

income variant, and there is evidence of our much more complete coverage of the top 1 percent.

c *Comparison with Census Samples for 1944, 1945, and 1947*

The third set of size distributions with which we compare our estimates is that by the Bureau of the Census for 1944, 1945, and 1947 based upon its sample studies of some 6,700, 8,700, and 12,000 households respectively.[11] The Census data are for money income alone; they exclude members of the armed forces living on millitary reservations, institutional inmates, and, in some years, all other persons outside the regular household (residents of hotels, noninmate residents of institutions, etc.). The comparisons can, therefore, be made only by dint of some rough assumptions.

We converted the distributions of consuming units to a per capita basis in a manner analogous to that followed for the 1935-36 and the 1941 data (App. 6, Sec. E). A special problem arose because the Census size distributions have a bottom and a top open-end class – below $500, and $10,000 and over – and give neither class totals nor means. The absence of absolute data for the $10,000 and over class was especially bothersome. We assumed two sets of class means. Assumption 1 uses arithmetic means of class intervals, $200 for the bottom open-end class, and $12,500 for the top open-end class. Assumption 2 uses geometric means of class intervals, the same mean ($200) for the lowest open-end class, but $25,000 for the top open-end class, the figure suggested by the average income per family and per single person on tax returns with adjusted gross income of $10,000 and over in 1944. For the Census distributions as given, without any attempt to correct for undercoverage of the upper income groups, Assumption 1 seems more reasonable than 2.

Assumption 2 compensates for the shortage in the income coverage at the top 1 percent level; but there remains a substantial shortage of income in the Census distributions as compared with the income totals we used. Part is explained by the Census exclusion of the armed forces, institutional population, and income in kind. In 1944 and 1945 these exclusions account

[11] For a detailed account of these data see *Family and Individual Money Income in the United States: 1945 and 1944, Family and Individual Money Income in the United States: 1945,* and *Income of Families and Persons in the United States: 1947* (Bureau of the Census, Series P-S, 22, May 8, 1947, Series P-60, 2, March 2, 1948, and Series P-60, 5, Feb. 7, 1949, respectively). We omitted comparison for 1946 when the Census sample data cover nonfarm population alone; and for 1948 because the income tax data became available after most of our computations had been completed and because there were other means of testing our estimates for that year (see Sec. 5).

for large amounts. But even with allowances for these differences in scope, substantial shortages remain.[12]

Of the two sets of columns in Table 106 derived from the Census sample data, the entries adjusted for scope of both population and income coverage (col. 4 and 6) are better approximations to the true levels than the unadjusted (col. 3 and 5). Here again, for reasons indicated in discussing the comparison for 1941, we may assume that the shortages are due partly to underrepresentation at the very top levels and partly to underreporting at the low levels. In this particular case the assumption is strengthened by the omission of income in kind – most prevalent and important at the lower income levels – and by the exclusion of members of the armed forces whose per capita income is presumably mostly below that of the top 20 percent of the population.[13]

As might be expected, the shares in our basic variant run short of those derived from the Census samples, with the exception, as always, of that of the top 1 percent. However, the shares in our economic income variant (those for 1947 roughly estimated from the 1946 ratio of the shares in the economic income variant to those of the corresponding bands in the basic variant) about equal the shares derived from the Census samples on Assumption 1 – that assigning $12,500 as the class mean of the top open-end class. Any overestimate of that mean would affect not only the share of the top 1 percent but also that of the 2nd and 3rd percentage band, since the top open-end class is well over 1 percent (1.7 percent in 1944, 1.4 percent in 1945, and 3.0 percent in 1947) of the total population covered in the Census samples.

For 1944 the shares derived from the Census sample below the top 1 percent tend to run slightly above those in our economic income variant, beginning with the 2nd and 3rd percentage band if we use the higher mean for the top open-end class, and beginning with the 4th and 5th percentage band if we use the lower and more reasonable mean. For 1945 the shares derived from the Census sample tend to run below those in our economic income variant for both the top 1 percent and the 2nd and 3rd percentage

[12] See Selma F. Goldsmith, Appraisal of Basic Data Available for Constructing Income Size Distributions, *Studies in Income and Wealth, Volume Thirteen* (NBER, 1951), pp. 267-377.

[13] This judgment is not contradicted by the fact that the Census totals are particularly short on such items as interest and dividends as well as on entrepreneurial income. These shortages can easily be interpreted as due both to underrepresentation of numbers in the top brackets and to the tendency not to report when the amounts received are small. Such differential shortages do not mean any distortion in weights of groups below the top 1 percent.

Table 106

Percentage Shares of Upper Income Groups Based on Income Tax Data Compared with Those Based on the Census Samples, 1944, 1945, and 1947

PERCENTAGE BAND	ESTIMATES BASED ON INCOME TAX DATA		ESTIMATES BASED ON CENSUS SAMPLE			
			Assumption 1		*Assumption 2*	
	Basic variant, total pop. (1)	Economic income variant, total pop. (2)	Unadj. for pop. & income base (3)	Adj. for pop. & income base (4)	Unadj. for pop. & income base (5)	Adj. for pop. & income base (6)
			1944			
Top 1	8.6	9.0	6.1	4.5	10.6	8.1
2nd & 3rd	4.9	5.7	7.3	5.3	8.5	6.3
4th & 5th	3.2	4.0	5.8	4.1	5.6	4.3
Top 5	16.6	18.7	19.2	13.8	24.7	18.6
6th & 7th	2.7	3.2	4.8	3.5	4.6	3.5
Top 7	19.4	21.9	24.0	17.3	29.3	22.1
8th-10th	3.7	4.2	6.5	4.7	6.1	4.6
Top 10	23.1	26.0	30.5	22.0	35.3	26.7
11th-15th	5.5	5.9	9.0	6.4	8.4	6.3
Top 15	28.6	32.0	39.5	28.3	43.7	33.1
16th-20th	5.1	5.2	7.8	5.6	7.2	5.5
Top 20	33.7	37.2	47.3	34.0	50.9	38.6
			1945			
Top 1	8.8	9.4	5.6	4.0	8.9	6.6
2nd & 3rd	5.3	5.9	6.7	4.8	7.6	5.6
4th & 5th	3.3	4.0	5.4	3.8	5.5	4.0
Top 5	17.4	19.3	17.6	12.5	22.0	16.3
6th & 7th	2.7	3.3	4.7	3.4	4.5	3.3
Top 7	20.1	22.5	22.4	15.9	26.5	19.6
8th-10th	3.6	4.2	6.4	4.5	6.1	4.5
Top 10	23.7	26.7	28.8	20.4	32.6	24.1
11th-15th	5.3	5.8	8.6	6.1	8.2	6.1
Top 15	29.0	32.5	37.4	26.5	40.8	30.1
16th 20th	4.9	5.0	7.8	5.6	7.3	5.4
Top 20	33.9	37.5	45.2	32.1	48.1	35.6
			*1947**			
Top 1	8.5	9.1	5.5	4.1	9.2	7.5
2nd & 3rd	5.4	5.9	6.9	5.2	8.9	7.3
4th & 5th	3.5	4.1	5.4	4.1	5.8	4.7
Top 5	17.4	19.1	17.8	13.5	23.9	19.6
6th & 7th	2.8	3.4	4.9	3.7	4.8	4.0
Top 7	20.2	22.5	22.7	17.2	28.8	23.5
8th-10th	3.7	4.4	6.2	4.7	6.0	4.9
Top 10	24.0	26.9	28.9	21.8	34.8	28.4
11th-15th	5.2	5.7	9.0	6.8	8.4	6.9
Top 15	29.1	32.6	37.9	28.6	43.2	35.3
16th-20th	4.9	5.2	7.8	5.9	7.1	5.8
Top 20	34.0	37.8	45.7	34.5	50.3	41.1

Because of rounding, details may not add to totals.

* Column 2 estimated from the 1946 ratio of the shares in the economic income variant to those of the corresponding bands in the basic variant.

band, even when we use the higher mean for the top open-end class; and the two sets of shares for the percentage bands from the 4th and 5th through the 11th-15th are about the same. For 1947 only the share of the top 1 percent derived from the Census sample, and of the 2nd and 3rd percentage band on Assumption 1, are smaller than in our estimates. But all these differences are minor and cannot be assigned much significance. The general conclusion, as from the preceding comparisons, is that no large biases can be detected in our shares in the economic income variant — at the top income levels and perhaps even down to the 11th-15th percentage band.

d *Comparison with Surveys of Consumer Finances for 1945-47*

Finally, in the Surveys of Consumer Finances for 1945-47 we have size distributions of money income based on some 3,000 returns each year and carried through in connection with a study of liquid asset holdings.[14] Like the other sample distributions, they were converted to a per capita basis, but in this case we had the assistance of those in charge of the Surveys (App. 6, Sec. F). And like all the sample distributions except the 1935-36, these suffer from a shortage of income. Moreover, as with the Census samples, we had to correct also for a shortage in population coverage, i.e., calculate the percentage bands in relation to total population instead of the sample universe (Table 107).

For reasons given repeatedly, the adjusted shares in column 4 are the ones to be compared with our estimates. The comparison shows the shortage in the share of the top 1 percent noted above as characteristic of sample distributions; but the shares of the lower percentage bands of the Survey distributions (except the 2nd and 3rd in 1945), even when reduced as in column 4, exceed our estimates,even those in the economic income variant. The excesses are absolutely small but fairly substantial relatively.

It is difficult to account for them. Perhaps our estimates are on the short side in the percentage bands below the top 1. On the other hand, the trouble may well be in the Survey data; obviously it is not in the overreporting of income by the sample units but in their weights. Analysis suggests that urban families and individuals are overrepresented in the Survey for 1945,[15] and the effort to represent the upper income groups adequately may have resulted in overrepresenting the groups from the 2nd percentage

[14] As in the case of the Census samples, we omitted comparison for 1948 for the reasons indicated in note 11.

[15] See Robert Wasson, Abner Hurwitz, and Irving Schweiger, An Appraisal of Field Surveys of Consumer Income, *Studies in Income and Wealth, Volume Thirteen*, pp. 482-559.

Table 107

Percentage Shares of Upper Income Groups Based on Income Tax Data Compared with Those Based on the Surveys of Consumer Finances, 1945-1947

Percentage Band	Estimates Based on Income Tax Data: *Basic variant, total population* (1)	Estimates Based on Income Tax Data: *Economic income variant, total population* (2)	Estimates Based on Survey of Consumer Finances: *Unadjusted for population & income base* (3)	Estimates Based on Survey of Consumer Finances: *Adjusted for population & income base* (4)
		1945		
Top 1	8.8	9.4	5.9	4.5
2nd & 3rd	5.3	5.9	7.1	5.4
4th & 5th	3.3	4.0	6.0	4.5
Top 5	17.4	19.3	19.0	14.5
6th & 7th	2.7	3.3	5.0	3.8
Top 7	20.1	22.5	23.9	18.3
8th-10th	3.6	4.2	6.4	4.8
Top 10	23.7	26.7	30.3	23.1
11th-15th	5.3	5.8	9.0	6.9
Top 15	29.0	32.5	39.3	30.0
16th-20th	4.9	5.0	7.9	6.0
Top 20	33.9	37.5	47.3	36.0
		1946		
Top 1	9.0	9.6	7.6	6.0
2nd & 3rd	5.7	6.2	8.2	6.5
4th & 5th	3.5	4.2	5.9	4.7
Top 5	18.2	20.0	21.8	17.2
6th & 7th	2.8	3.4	5.1	4.0
Top 7	21.0	23.4	26.8	21.2
8th-10th	3.7	4.3	7.0	5.5
Top 10	24.7	27.7	33.8	26.7
11th-15th	5.3	5.9	8.8	6.9
Top 15	30.0	33.6	42.5	33.6
16th-20th	4.8	5.2	7.7	6.0
Top 20	34.9	38.7	50.2	39.7
		*1947**		
Top 1	8.5	9.1	7.9	6.7
2nd & 3rd	5.4	5.9	9.3	7.9
4th & 5th	3.5	4.1	6.9	5.8
Top 5	17.4	19.1	24.1	20.4
6th & 7th	2.8	3.4	5.5	4.6
Top 7	20.2	22.5	29.6	25.0
8th-10th	3.7	4.4	6.8	5.8
Top 10	24.0	26.9	36.4	30.8
11th-15th	5.2	5.7	9.1	7.6
Top 15	29.1	32.6	45.5	38.4
16th-20th	4.9	5.2	7.1	6.0
Top 20	34.0	37.8	52.7	44.4

Because of rounding, details may not add to totals.
* Column 2 estimated from the 1946 ratio of column 2 to column 1.

downward while still missing a good portion of the very top income units.

A relatively minor error in weighting at an upper income level would

have an effect all the way down. For example, if we assign a weight of 3 (instead of 2) to the 2nd and 3rd, and 4th and 5th percentage bands respectively, and assume that their true shares are as given in column 2 (rather than as in col. 3 or 4), the entries would be about the same or in excess of those in column 4. Yet, in weighting on the basis of a small sample, it is not difficult to overweight by about 2 percentage points, with consequent underweighting elsewhere in the distribution.

One could argue that Table 107 does suggest some shortages in our shares even in the economic income variant. But the evidence is far from impressive. The cumulative share of the upper bands does not fall short of that based on the Survey of Consumer Finances for 1945, and reaches the shortage point at the 20 percent line in 1946 and at the 5 percent line in 1947. In view of the crudity of the comparison, differences of 1 or 2 percentage points cannot be deemed significant. It is, therefore, fair to conclude that the comparison, as far as it goes, does not reveal any serious shortages in our estimates, at least in the top 5 percent range.

4 *The Evidence for 1944-48*

Unless otherwise explained, the smaller per capita income of the population represented on tax returns than of the total population noted in our estimates for 1944-48, when income tax coverage was extended to a high percentage of the total population, indicates serious shortages in income reported on tax returns. In appraising our estimates, therefore, we should pay particular attention to the evidence for these years. Since some earlier years may be affected and it is important to include a prewar year, we go back to 1941.

The puzzle is most manifest in Table 108 where the shares in our basic variant for total population are estimated for lower and lower percentage bands as the tax coverage is extended. If there are no errors in the numerators or denominators, i.e., tax return population and its income on the one hand and total population and its income on the other, the per percentile shares of the successive percentage bands, including those for the lowest group, derived by subtraction, should decline continuously. And this is what we find in Part B for 1941 and 1942. But in 1943 the per percentile share of the lowest group (the 66th-100th percentage band), 0.81, exceeds that of the percentage band just above it, 0.66. This reversal of the downward trend of the per percentile shares is even more striking in 1944-48 when the share of the lowest group (the 81st-100th percentage band) not only exceeds that of the percentage band above it, but exceeds 1, i.e., the average per capita income for the country as a whole, by a fairly wide margin. The reason this puzzling reversal is not evident in 1941 or 1942

Table 108

Percentage Shares of Upper Income Groups, Total and per Percentile, Basic Variant, Total Population, 1941-1948

Percentage Band	*1941*	*1942*	*1943*	*1944*	*1945*	*1946*	*1947*	*1948*
			A TOTAL SHARES					
Top 1	11.39	10.06	9.38	8.58	8.81	8.98	8.49	8.38
2nd & 3rd	6.25	5.33	5.12	4.89	5.30	5.69	5.45	5.57
4th & 5th	4.24	3.55	3.25	3.16	3.28	3.53	3.48	3.68
6th & 7th	3.48	3.04	2.79	2.73	2.71	2.84	2.79	3.07
8th-10th	4.60	4.02	3.74	3.71	3.61	3.66	3.75	4.08
11th-15th	7.03	5.96	5.82	5.55	5.33	5.34	5.19	5.82
16th-20th	6.63	5.45	5.33	5.08	4.88	4.84	4.89	5.05
21st-35th	17.80	14.15	14.20	13.58	13.03	12.67	13.04	13.12
36th-50th		12.40	12.14	11.87	11.60	11.09	11.90	11.53
51st-65th	38.58		9.93	9.72	9.44	9.65	10.43	10.27
66th-80th		36.04	28.30	5.56	5.44	7.07	7.64	7.91
81st-100th				25.56	26.57	24.66	22.96	21.53
			B SHARES PER PERCENTILE					
Top 1	11.39	10.06	9.38	8.58	8.81	8.98	8.49	8.38
2nd & 3rd	3.13	2.67	2.56	2.45	2.65	2.84	2.72	2.79
4th & 5th	2.12	1.77	1.62	1.58	1.64	1.76	1.74	1.84
6th & 7th	1.74	1.52	1.40	1.37	1.35	1.42	1.40	1.54
8th-10th	1.53	1.34	1.25	1.24	1.20	1.22	1.25	1.36
11th-15th	1.41	1.19	1.16	1.11	1.06	1.07	1.04	1.16
16th-20th	1.33	1.09	1.07	1.02	0.98	0.97	0.98	1.01
21st-35th	1.19	0.94	0.95	0.91	0.87	0.84	0.87	0.87
36th-50th		0.83	0.81	0.79	0.77	0.74	0.79	0.77
51st-65th	0.59		0.66	0.65	0.63	0.64	0.70	0.68
66th-80th		0.72	0.81	0.37	0.36	0.47	0.51	0.53
81st-100th				1.28	1.33	1.23	1.15	1.08
			C RECAPITULATION OF PER PERCENTILE SHARES					
16th-20th	1.33	1.09	1.07	1.02	0.98	0.97	0.98	1.01
21st-100th	0.70	0.78	0.81	0.83	0.83	0.81	0.82	0.80
21st-35th	1.19	0.94	0.95	0.91	0.87	0.84	0.87	0.87
36th-100th	0.59	0.75	0.77	0.81	0.82	0.81	0.81	0.79
36th-50th		0.83	0.81	0.79	0.77	0.74	0.79	0.77
51st-100th		0.72	0.76	0.82	0.83	0.83	0.82	0.79

may be that the tax return population is not large enough for us to carry our calculations to the lower reaches of the distribution.

To attribute the reversal to underreporting and omissions in the tax data would be easy. But for at least one reason we suspect they are not the sole cause: the variant in Table 108 is the basic, i.e., unadjusted for family status and unwarranted deductions – two major sources of underestimates in that variant. For 1944, 1945, and 1946 the share of the top 20 percent in the economic income variant was 37.2, 37.5, and 38.7 percent respectively, that in the basic variant, 33.7, 33.9, and 34.9 percent respectively (Tables 106 and 107). We may reasonably assume that this difference between the shares in the two variants would be relatively the same for the top 80 percent also – especially since that for the top 20 percent would be much larger if we excluded the top 1 percent. On this assumption the share of the top 80 percent in the economic income variant would be 82.2 percent in 1944, 81.3 percent in 1945, and 83.5 percent in 1946; and applying the relative difference for 1946 to 1947 and 1948, the share for those years would be 85.4 and 87.0 percent respectively. For all five years then, the share of the top 80 percent would represent a per capita income larger than the countrywide per capita. With this one adjustment the puzzle noted in Chapter 7 would disappear. Consequently, the underestimate of the shares in the basic variant is probably largely due to the failure to make the adjustments called for in the economic income variant; moreover, while the range of these adjustments is only about a tenth, the failure to make them has an increasingly distorting effect on the share of the residual group the smaller the residual group.

Yet we must still consider the possibilities of shortages in the income tax data. Their existence is evident from the fact that even after raising the share of the top 80 percent to allow for the adjustments in the economic income variant, the residual 20 percent of the population has a per percentile share of almost 1 (0.89 in 1944, 0.94 in 1945, 0.82 in 1946, 0.73 in 1947, and 0.65 in 1948) – much larger than some of the bands between the 36th and 80th percentage lines are likely to have.

These shortages may come from three sources of underreporting. The first is failure to report items that are included in our income estimates, used as denominators, but are either not called for or are highly unlikely to appear on income tax returns; e.g., such 'other labor income' as employers' contributions to welfare and pension funds and compensation for injury; and a great deal of nonmoney income of which the economic income variant adjusts for imputed rent alone. Relatively, the total of all such items is not large – at most, 2 percent of countrywide income. But failure to report them produces an error that is relatively large in the residual group.

For example, if of the missing 2 percent, 1.5 percent is not reported on tax returns and the true share of the residual group (81st-100th percentage band) should be, say, 10 percent of countrywide income, 1.5 percent would be added to the true figure, exaggerating it by as much as 15/100.

The second and quantitatively more important source of nonreporting in these years is military pay, as suggested by the big break in the per percentile shares in the lower groups between 1941 and 1942. Thus, for the residual group, 21st-100th percentage band, the income share per percentile which in 1941 is 0.70, jumps to 0.78 in 1942, rises gradually to 1945, then declines slightly. The jump from 1941 to 1942 is even more striking for the per percentile share of the lower 65 percent group (36th-100th percentage band) – from 0.59 to 0.75, over a quarter; yet the rises in the following years are quite gradual and cease in 1945. Even if we did not know anything about the history of these years, Table 108 would make us aware that the break in the distribution, the change that may have caused substantial shortages in the income shares, occurred between 1941 and 1942 and remained.

It was obviously associated with this country's entrance into World War II which caused the withdrawal of a substantial proportion of the population into the armed forces. Noncommissioned personnel of the military or naval forces were allowed to exclude from gross income compensation received for active service during 1942 up to $250 if single, or $300 if married or head of a family, $1,500 during 1943-44, and all service pay received during 1945-48. For 1943-48 commissioned officers were also allowed to exclude from gross income their active service pay up to $1,500. For 1944 and thereafter, mustering-out payments with respect to service in the military or naval forces were also excluded from gross income. In addition, amounts contributed by the government to the serviceman's "monthly family allowance" were regarded as gifts and did not have to be reported as income. Finally, members of the armed forces serving abroad or on sea duty could postpone filing returns and paying taxes until the 15th day of the sixth month following the month in which they returned to the United States, but not beyond June 15, 1948.

To estimate exactly how much shortage to assign to this factor is difficult, but we made an experimental calculation for 1944 and 1945, the two years probably most affected. We had already excluded the government's contributions to military family allowances – $2.5 billion in 1944 and $2.9 billion in 1945 (*Survey of Current Business,* July 1947, National Income Supplement, Table VII, p. 14) – from individuals' total income receipts used as denominators in those years. But we had not excluded the balance of military pay which amounted to as much as $20.8 billion in 1944 and

$22.4 billion in 1945 (*ibid.*, Table 1, p. 19). The total number in military service in these years was 11.4 million and 11.6 million respectively (*Statistical Abstract, 1946,* p. 220). Nonreporting by this group, whose income and numbers are included in the countrywide totals, would therefore seriously affect the completeness of the income tax data. Moreover, the resulting omission would not be in the very lowest brackets of the tax return population. Members of the armed services, particularly abroad, can be considered only as a group of single individuals; and their income, including nonmoney, was in 1944-45 well above $1,000 per capita, i.e., about equal to the per capita income of the total population. This means, in terms of Table 108, that the omitted group should be in the 16th-20th percentage band. And if we assume, again moderately, that 5-7 million in the armed services were outside the country (or roughly 4-5 percent of the total population), the adjustment to be made becomes somewhat clearer. In Table 108 members of the armed services abroad are in our residual group (81st-100th percentage band) in 1944 and 1945, but should be several percentage bands higher. To adjust the residual group in those years we should: (a) subtract 5 percent of its population and income; (b) add 5 percent of the population and income of the lower levels of the band just above it. For (b) we assume that the per percentile share for the lowest third of the 66th-80th percentage band, i.e., the percentages just above the residual band, is 0.30. Thus for 1944 the revised share of the lowest 20 percent would become $25.56 - 5.00 + (0.30 \times 5) = 22.06$; and for 1945, 23.07. This 3.5 percentage point reduction of the lowest band's share, on top of the reduction suggested in passing from the basic to the economic income variant, would bring its share down to 14.3 percent for 1944 and 15.2 percent for 1945. There may be similar if smaller adjustments in the later years. And if we add the possible effects of the other omissions mentioned, the percentage share of the lowest 20 percent of the population, estimated as a residual, would be down to 12-14 percent, or 0.6-0.7 percent of income per percentile of the population.

A third source of underreporting is clearly defined money income by persons filing returns. There are bits of evidence that it too occurs chiefly at the levels well below our upper groups. The first bit is the sharp falling off in the per percentile shares as we pass from the 51st-65th to the 66th-80th percentage band: in 1944 the drop is from 0.65 to 0.37, in 1945, from 0.63 to 0.36, and it is still quite abrupt in the later years. No such break is apparent during equal intervals in passing from the 21st-35th or the 36th-50th percentage band to the next. Indeed, the slope of the general decline strongly suggests that the entries for the 66th-80th percentage band in Table 108, Part B, should have been 0.5 or slightly above in 1944 and

1945; and about 0.6 in the later years. If so, the underreporting in this band is about three-tenths of the true value in 1944 and 1945, and from about a sixth to an eighth in the later years.

Another ground for assuming that underreporting is largely in the lower brackets is that the new entrants into the tax return population are here, and they are the ones most likely to understate their income. Likewise, the groups that have been among those notorious for noncompliance, farmers and small entrepreneurs, are the ones whose incomes perhaps rose most during the war. This combination of heavier weight of new entrants and larger income shares of groups that tend to evade taxation has increased understatement and underreporting largely in the lower income brackets.

Some confirmation comes from the per percentile shares in the various types of income (Table 109). The array for successive percentage bands is by total income per capita, not by income of a given type. Therefore, the shares per percentile in a given type of income may well rise at some point in the array as we descend the percentage bands. Table 109, particularly important for the question of the level at which the shortage in the income tax data is likely to be largest, contains four pieces of evidence:

First, in all types except entrepreneurial income there is the sharp break in the shares from 1941 to 1942, confirming World War II as a major factor in distorting the shares, primarily in the lower bands.

Second, the movement in the shares of employee compensation is not unlike that of total income, for the simple reason that the former accounts for over seven-tenths of the latter. The sharp break in the per percentile shares at the levels near the bottom of the tax return population (see the 36th-50th and 51st-65th percentage bands for 1943-45 and the 51st-65th and 66th-80th percentage bands for 1944-48) lead us to suspect large shortages at these low levels.

Third, for the next large type, entrepreneurial income, the shares at the upper levels (say, above the 5 percent line) rise to 1945, and the shares that decline after 1941 are only those at the intermediate levels (between the 20th and 50th percentage bands). One cannot escape the suspicion that farm and other entrepreneurial incomes, which increased during the war and should have raised or at least held constant the shares at the intermediate levels, are appreciably underreported.

Fourth, in the case of both rent and pure property incomes (interest and dividends) some shares at the upper levels decline to 1944 or 1945. But the steeper declines are at the levels between the 5th percent line and the residual band, which is distorted by shortages. There is no particular reason to believe that the shares at these levels – from the 5th down to, say, the 50th percentage band – actually declined at all; and one suspects that the

Table 109

Percentage Shares per Percentile of Population in Various Types of Income Percentage Bands by Total Income, Basic Variant, Total Population, 1941-1948

Percentage Band	*1941*	*1942*	*1943*	*1944*	*1945*	*1946*	*1947*	*1948*
			A EMPLOYEE COMPENSATION					
Top 1	6.00	4.89	3.75	3.33	3.33	3.76	3.90	3.78
2nd & 3rd	2.57	2.19	1.85	1.68	1.66	1.77	1.80	1.87
4th & 5th	2.22	1.90	1.71	1.61	1.54	1.54	1.55	1.68
6th & 7th	1.97	1.74	1.57	1.54	1.48	1.45	1.46	1.63
8th-10th	1.82	1.60	1.48	1.43	1.38	1.37	1.40	1.55
11th-15th	1.77	1.41	1.40	1.33	1.27	1.27	1.29	1.39
16th-20th	1.68	1.28	1.27	1.24	1.18	1.19	1.22	1.24
21st-35th	1.34	1.13	1.11	1.10	1.06	1.06	1.11	1.10
36th-50th	0.58	0.97	1.01	0.94	0.93	0.93	1.01	0.96
51st-65th		0.67	0.80	0.75	0.74	0.80	0.87	0.83
66th-80th			0.70	0.40	0.41	0.56	0.60	0.60
81st-100th				1.10	1.19	1.00	0.80	0.78
			RECAPITULATION					
16th-20th	1.68	1.28	1.27	1.24	1.18	1.19	1.22	1.24
21st-100th	0.72	0.81	0.85	0.87	0.89	0.88	0.87	0.85
21st-35th	1.34	1.13	1.11	1.10	1.06	1.06	1.11	1.10
36th-100th	0.58	0.74	0.79	0.82	0.85	0.84	0.82	0.79
36th-50th		0.97	1.01	0.94	0.93	0.93	1.01	0.96
51st-100th		0.67	0.73	0.78	0.82	0.81	0.76	0.74
			B ENTREPRENEURIAL INCOME					
Top 1	16.88	18.84	23.35	22.00	23.00	18.28	15.16	15.16
2nd & 3rd	4.14	3.80	4.75	4.89	5.89	5.63	5.11	5.37
4th & 5th	1.68	1.41	1.41	1.53	2.08	2.43	2.33	2.42
6th & 7th	1.18	0.95	0.89	0.88	1.10	1.45	1.34	1.35
8th-10th	0.84	0.75	0.65	0.61	0.72	0.91	0.82	0.88
11th-15th	0.47	0.66	0.49	0.46	0.51	0.62	0.57	0.63
16th-20th	0.47	0.63	0.52	0.39	0.43	0.49	0.44	0.48
21st-35th	0.76	0.49	0.53	0.36	0.37	0.37	0.34	0.36
36th-50th	0.78	0.43	0.26	0.38	0.35	0.31	0.32	0.32
51st-65th		0.93	0.26	0.38	0.33	0.29	0.32	0.35
66th-80th			1.14	0.29	0.23	0.26	0.30	0.36
81st-100th				1.81	1.64	1.80	2.03	1.88
			RECAPITULATION					
16th-20th	0.47	0.63	0.52	0.39	0.43	0.49	0.44	0.48
21st-100th	0.77	0.75	0.69	0.72	0.65	0.68	0.75	0.73
21st-35th	0.76	0.49	0.53	0.36	0.37	0.37	0.34	0.36
36th-100th	0.78	0.81	0.73	0.80	0.71	0.75	0.84	0.82
36th-50th		0.43	0.26	0.38	0.35	0.31	0.32	0.32
51st-100th		0.93	0.87	0.92	0.82	0.89	1.00	0.96

Table 109 concluded:

Percentage Band	*1941*	*1942*	*1943*	*1944*	*1945*	*1946*	*1947*	*1948*
			C RENT					
Top 1	11.35	9.96	9.76	8.94	9.11	10.16	10.98	12.77
2nd & 3rd	3.72	2.80	2.91	2.48	2.71	3.12	3.16	3.35
4th & 5th	2.03	1.33	0.95	0.95	1.14	1.41	1.54	1.56
6th & 7th	1.57	0.99	0.70	0.58	0.66	0.87	0.85	0.95
8th-10th	1.27	0.82	0.59	0.47	0.42	0.59	0.55	0.63
11th-15th	0.90	0.77	0.51	0.37	0.34	0.41	0.39	0.43
16th-20th	0.87	0.74	0.54	0.31	0.29	0.33	0.32	0.31
21st-35th	1.42	0.56	0.51	0.28	0.24	0.26	0.24	0.26
36th-50th		0.71	0.26	0.29	0.24	0.22	0.21	0.20
51st-65th	0.62		0.32	0.29	0.25	0.22	0.24	0.24
66th-80th		1.01	1.65	0.35	0.29	0.31	0.33	0.36
81st-100th				3.00	3.11	2.92	2.87	2.70
			RECAPITULATION					
16th-20th	0.87	0.74	0.54	0.31	0.29	0.33	0.32	0.31
21st-100th	0.77	0.87	0.93	0.98	0.97	0.92	0.91	0.87
21st-35th	1.42	0.56	0.51	0.28	0.24	0.26	0.24	0.26
36th-100th	0.62	0.94	1.02	1.14	1.14	1.07	1.06	1.02
36th-50th		0.71	0.26	0.29	0.24	0.22	0.21	0.20
51st-100th		1.01	1.25	1.40	1.41	1.33	1.32	1.26
			D DIVIDENDS AND INTEREST					
Top 1	44.14	41.55	40.19	38.88	37.80	37.59	37.46	38.56
2nd & 3rd	5.13	4.48	4.75	5.07	5.44	5.54	4.93	5.09
4th & 5th	1.97	1.48	1.27	1.64	1.79	2.08	1.82	1.94
6th & 7th	1.31	0.94	0.77	0.91	0.82	1.21	1.00	1.10
8th-10th	0.87	0.69	0.56	0.64	0.54	0.69	0.54	0.69
11th-15th	0.53	0.57	0.53	0.43	0.40	0.44	0.36	0.45
16th-20th	0.49	0.49	0.43	0.34	0.33	0.34	0.29	0.32
21st-35th	0.74	0.39	0.38	0.29	0.28	0.26	0.22	0.23
36th-50th		0.48	0.22	0.30	0.28	0.22	0.20	0.18
51st-65th	0.31		0.26	0.32	0.27	0.22	0.21	0.21
66th-80th		0.48	0.76	0.29	0.24	0.26	0.27	0.29
81st-100th				1.11	1.25	1.22	1.44	1.28
			RECAPITULATION					
16th-20th	0.49	0.49	0.43	0.34	0.33	0.34	0.29	0.32
21st-100th	0.39	0.47	0.50	0.50	0.51	0.48	0.53	0.49
21st-35th	0.74	0.39	0.38	0.29	0.28	0.26	0.22	0.23
36th-100th	0.31	0.48	0.52	0.55	0.57	0.54	0.60	0.55
36th-50th		0.48	0.22	0.30	0.28	0.22	0.20	0.18
51st-100th		0.48	0.61	0.63	0.65	0.63	0.72	0.66

new entrants into these levels of the tax universe underreport their property incomes, especially as they are probably rather small.

For the years beginning with 1942 or 1943 the distributions could easily

be revised to eliminate the increase in the per percentile share of *total* income: by replacing the rates in Table 109 for employee compensation below the 35 percent line by estimated rates, based on an extrapolation of the decline in the shares of the percentage bands above that line with an allowance for its gradual retardation; and by substituting for the rates for each other income type below the 20 percent line a constant share per percentile based on the average share for the 21st-100th percentage band. Such a calculation for 1943, 1944, and 1945 yields total income shares whose per percentile shares down through the residual group decline continuously. But there is little use in presenting it. We mention it as evidence of the ease with which the shares in Table 108 can be revised on the assumption – supported by the various bits of evidence mentioned above – that the chief shortages in the income tax data occur below the 20 percent line.

This naturally does not mean that there are no shortages above the 20 percent line. As we have seen, the shares in the basic variant for these years must be as short – by perhaps as much as a tenth – of the true shares as they were in other years and there may be other small biases. Also, as the audit study of 1948 returns discussed in the next section shows, income is underreported at upper as well as at lower income levels. But one may reasonably infer from the bits of evidence that the estimates for the top 5 percent in the economic income variant are not subject to much more serious biases for these recent years than they are throughout the period. The rise in the per percentile share of total income in the residual group in 1943-48 may be largely due to the bias in the basic as compared with the economic income variant, and perhaps even more, to the shortages in income tax data affecting the levels below the top, and most conspicuous in the bands near the bottom of the income tax universe.

5 *Sample Audit Study of 1948 Tax Returns*

After this report was written, some results of the sample audit study of 1948 returns, referred to in note 4, became available. Since this is the first audit study that follows the random method of selecting returns for examination and hence permits an unbiased judgment of the errors on returns, it seems worth while to examine whatever general results are available.

The published and tabulated detail distinguishes between Agents' returns – Form 1040 returns with adjusted gross income of $7,000 or more or with gross receipts of $25,000 or more from business or profession – and Collectors' returns – Form 1040A returns, and Form 1040 returns with adjusted gross income of less than $7,000, and with gross receipts of less than $25,000 from business or profession. The failure to tabulate fully by

Table 109a

Summary of Results of Sample Audit Study of 1948 Returns

		AGENTS' RETURNS[a]		
		Under $25,000[b]	$25,000 and over	Total
	TAX CHANGE ($ MILLION)	(1)	(2)	(3)
1	Gross	432	201	633
2	Increase	407	181	588
3	Decrease	25	20	45
4	Net increase	382	161	543
5	% OF TOTAL POPULATION REPRESENTED	4.6	0.5	5.1
6	TOTAL TAX LIABILITY REPORTED ($ MILLION)	3,299	4,239	7,539
	TAX CHANGE AS % OF TOTAL LIABILITY REPORTED			
7	Gross (line 1 ÷ line 6)	13.1	4.7	8.4
8	Net (line 4 ÷ line 6)	11.6	3.8	7.2
9	% OF TAX CHANGE (GROSS) ON RETURNS WITH MAJOR ERROR IN ADJUSTED GROSS INCOME	89	85	88
10	ERRORS PER 100 RETURNS	63	70-72	

Calculated Net Increase in Tax[c] with Increase in Adjusted Gross Income

		ADJUSTED GROSS INCOME LEVEL OF		
		$7,000	$25,000	$100,000
	Increase in Adjusted Gross Income	% Net Increase in Tax		
11	5% of given income level	9	8	7
12	10% of given income level	18	16	14
13	20% of given income level	37	33	29

[a] Form 1040 returns with $7,000 or more of adjusted gross income or with gross receipts of $25,000 or more from business or profession.

[b] Includes returns with $7,000 to $25,000 adjusted gross income; also business returns with less than $7,000 adjusted gross income when gross receipts are $25,000 or more.

[c] For joint return of husband and wife claiming two dependents and deductions of 10 percent of adjusted gross income.

size of income the audit sample results for Collectors' returns renders the available information of little value for the present purpose. Agents' returns are classified into three groups: those with adjusted gross income of $7,000 to $25,000, $25,000 to $100,000, and $100,000 and over. As line 5 of Table 109a indicates, the population represented (calculated from exemptions as reported in the Preliminary Report of *Statistics of Income for 1948* and adjusted to exclude extra exemptions for old age and blindness) is 5.1 percent of total population, returns with incomes over $25,000 accounting for about 0.5 percent, those with incomes of $7,000 to $25,000, for 4.6 percent. It seemed best to study Agents' returns in these two percentage groups – the top 0.5 percent and the next 4.6 percent.

As might have been expected, the audit study disclosed some errors that underestimated and some that overestimated the correct tax liability –

with the former preponderating. The *net* tax increase for all Agents' returns (line 4) amounted to about half a billion dollars. As a percentage of total tax liability reported, this *net* underestimate of tax represented 3.8 percent for the top 0.5 percent of population, 11.6 percent for the next 4.6 percent of population, and 7.2 percent for all Agents' returns combined.

This, however, does not mean that *income* reported on the face of the returns was underestimated by these percentages. First, some of the increase in tax liability resulted from errors other than those in income — of the total tax change (whether increase or decrease) over 10 percent was found on tax returns with the major source of error other than that in income (line 9). We must, therefore, reduce somewhat the relative tax error shown in line 8 if it is to reflect understatement of income voluntarily reported for tax purposes.

A much more important adjustment is involved when one considers that increasing the tax because of understatement of income discovered by auditing involves *marginal* tax rates whereas the total tax liability represents an *average* burden. As line 11 shows, if a joint return of husband and wife claiming two dependents and legitimate deductions amounting to 10 percent of adjusted gross income reported its income as $7,000 whereas it was in fact 5 percent larger, $7,350, the upward adjustment of tax is 9 percent, not 5 percent of the tax liability calculated on the face of the return. Similarly significant differences between assumed changes in adjusted gross income and resulting changes in tax are shown in lines 12 and 13 for income levels of $25,000 and $100,000. Consequently, in translating percentage adjustments in taxes into percentage adjustments in income, the former would have to be scaled down materially if they represent increases.

The underestimate of income can now be approximated from Table 109a. For the top 0.5 percent of population (col. 2) the average net tax increase is 3.8 percent of the reported tax liability and the number of errors per 100 returns about 70. If we may interpret the latter as the percentage of incorrect returns, it is permissible to argue that the understatement of tax on incorrect returns is about 5.4 percent of the reported tax liability. In translating this into understatement of income, we should reduce the 5.4 percent about a seventh to allow for errors other than those in income,[16]

[16] Since the percentage in line 9 takes account of only the tax change on returns whose major source of error has been indicated to be adjusted gross income, and since such returns may have minor errors in itemized or standard deductions, personal exemptions, and arithmetic, we assumed that these minor nonincome errors are more or less offset in their tax effects by the minor income errors on returns whose major error is in nonincome sources.

then perhaps scale down the resulting 4.6 percent at least a third to allow for the difference between marginal and average tax. Hence income on 70 percent of the returns in this group is underestimated about 3.1 percent; and for the group as a whole, probably somewhat over 2 percent. A similar calculation for the next 4.6 percent of the population yields a rough underestimate of income of somewhat less than 6 percent for the group as a whole, and for all Agents' returns, about 4.5 percent.

Underreporting of adjusted gross income can be estimated from the audit study by an alternative, and perhaps somewhat more precise method.[17] We calculate first the average net tax increase per return with the major error in adjusted gross income for each of the three income classes of Agents' returns; then, from the Preliminary Report of *Statistics of Income for 1948,* the average adjusted gross income and tax liability reported per return, again for each of the three income classes of Agents' returns. If we can assume that returns with the major error in adjusted gross income are characterized by the average adjusted gross income and tax liability reported on *all* returns of the given income class, we can, using the assumptions concerning family status, number of dependents, and allowable deductions that were made in Table 109a, lines 11-13, calculate what the net tax increase per return disclosed by audit means in the way of net increase in adjusted gross income per return with the major error in the latter. Then, knowing the number of such returns in each income class, we can estimate the total net increase in adjusted gross income and the percentage it constitutes of adjusted gross income as reported. By this alternative method the increase in adjusted gross income on Agents' returns of $25,000 and over is 2.2 percent; on those under $25,000, 6.6 percent; and on all Agents' returns, 5.0 percent.

These results convey too optimistic a picture of the accuracy of income reporting in the upper brackets, for several reasons. First, the audit study deals with errors on the returns that were received, and is not designed to cope with nonfiling. But the latter is, as already indicated, an exceedingly minor problem at upper income levels. Second, the audit study may not have succeeded in uncovering all missing sources of income: the possibility is always in that direction rather than in the direction of finding incomes that were not in fact received. Third, the year covered is one in which, unlike most years in our study, a substantial part of the tax was withheld at source and errors could, therefore, be made only in the tax voluntarily reported. Finally, the possible understatement of income of some types

[17] Marius Farioletti of the Bureau of Internal Revenue suggested this method. I am greatly indebted to Mr. Farioletti for numerous improvements he made in reviewing this section.

may be much larger proportionately than that of total adjusted gross income. Indeed, the results already published indicate that the errors in reporting net income and loss from business tend to be particularly frequent and relatively sizeable.

Yet with all these qualifications, the results of the audit study do seem to bear out the main assumption of this investigation — that the underestimation of income at the upper income levels is within fairly narrow relative limits. The audit study, as far as the recent results go, warrants an inference that such underestimation is within a 5 percent margin for incomes at the top 1 percent level, and within a 10 percent margin for incomes in the 2nd through 5th percentage bands.

6 *Comparison with Goldenthal's Estimates*

Comparison with Goldenthal's estimates is not intended to shed any light on the reliability of our estimates, for he used the same data and a somewhat similar, if much cruder, procedure. The purpose is rather to show how different approaches may yield fairly similar sets of measures.

In his *Concentration and Composition of Individual Incomes, 1918-1937* (TNEC Investigation of Concentration of Economic Power, Monograph 4, Washington, 1940) Goldenthal compares income on tax returns with countrywide income, both including gains and losses from sales of assets. Treating returns as comparable units, he compares them with income recipients, instead of with total population, distinguishing the upper group of returns as percentages of total recipients. With his share of the top 1 percent, adjusted roughly to exclude gains and losses on sales of assets (Table 110, col. 1), we compare our estimates in both the basic and economic income variants.

The agreement of the shares in our basic variant with Goldenthal's is striking. In only a few years do the two series differ by more than three-tenths of a percent, especially if we exclude the years after the break in 1931. Why should the two sets of estimates be so close when Goldenthal does not adjust tax returns for the number of persons represented and deals with income recipients rather than total population?

The puzzle is solved when we scrutinize Goldenthal's totals of income recipients and our own data on total population and on the number per tax return in the upper brackets. For his total recipients Goldenthal takes the gainfully occupied (see his Table A-4, p. 80), without any allowance for unemployment or for recipients of property incomes who are not gainfully occupied. The number of recipients thus defined is a stable function of the total population. In fact, Goldenthal uses an almost constant percentage of total population as his base — roughly 0.4.

Table 110

Percentage Share of the Top 1 Percent, Our Estimates Compared with Goldenthal's, 1919-1936

	Excluding Gains and Losses from Sales of Assets			Disposable Income	
	Goldenthal	*Basic variant, total population*	*Economic income variant, total population*	*Goldenthal*	*Our variant, total population*
	(1)	(2)	(3)	(4)	(5)
1919	12.7	12.8	14.0	11.9	12.2
1920	12.0	12.3	13.6	11.3	11.8
1921	13.3	13.5	16.2	12.6	14.2
1922	13.4	13.4	15.6	13.1	14.4
1923	12.2	12.3	14.0	12.2	13.1
1924	13.0	12.9	14.7	13.4	14.3
1925	13.7	13.7	15.7	15.6	16.5
1926	14.2	13.9	15.8	15.4	16.3
1927	14.6	14.4	16.5	16.3	17.2
1928	14.8	14.9	17.2	18.1	19.1
1929	14.6	14.5	17.2	17.4	18.9
1930	14.2	13.8	15.6	14.0	15.1
1931	14.2	13.3	15.6	13.1	14.6
1934	12.7	12.0	13.6	11.6	12.4
1935	12.7	12.1	13.6	11.8	12.8
1936	13.2	13.4	14.7	12.2	13.7

Column

1 The shares of total individual income as shown in *Concentration and Composition of Individual Incomes, 1918-1937* (TNEC Monograph 4, Washington, 1940), Table 1, p. 16, are reduced by the percentage accounted for by realized capital gains and losses (*ibid.*, Table 13, p. 40) and increased by the percentage that realized capital gains and losses constitute of total individual income (*ibid.*, Table 11, p. 38). For 1934-36 the shares are further reduced to take account of direct relief and adjusted service compensation not included in total income (*ibid.*, Table 20, p. 67).

2, 3, 5 Tables 116 and 122.

4 *Concentration and Composition of Individual Incomes, 1918-1937*, Table 16, p. 60. The figures for 1935-36 are adjusted by the ratio of the first entry for 1934 to the second.

The upper brackets are dominated by family returns. The average number per family return ranges from 3 to 3.5 in the $10,000 and over and the $5,000-10,000 net income classes (Table 68). Hence, Goldenthal's 1 percent of income recipients is in fact significantly more than 1 percent of the population – roughly 0.4 multiplied by a constant ranging from 3 to 3.5. His shares are, therefore, for a percentage of the population ranging from 1.2 to 1.4 but varying on the whole rather little. The equality of the level of the shares in the first two columns of Table 110 is therefore misleading: our shares for 1 percent of the population are compared with Goldenthal's shares for 1.2 to 1.4 percent. Substantially underestimating the share of the top 1 percent of the population, his estimates obviously have little meaning

as the share of the top 1 percent of recipients. This was to be expected, since failure to allow for number per return would cause an underestimate of the income share of the top percentage of the population.

The similarity of their short term changes despite this hidden substantial difference in the coverage of the two series can also be easily explained. In dealing with a relatively narrow and distinct group, such as that represented by 1.2 to 1.4 percent of the total population, Goldenthal's cruder procedure still manages to reflect the major changes in the income received by the top group as compared with the income of the total population. The differences between the income denominators of the two sets of estimates in Table 110 as well as between year to year changes in the income numerators are minor. The damping of such changes introduced by our reinclusion of deductions may well be roughly matched by the damping introduced by Goldenthal's coverage of a larger proportion of the tax return population for his top 1 percent. However, such offsetting of differences in year to year changes is in precarious balance, and would certainly vanish if the comparison were extended to somewhat wider groups or to different periods.