Volume Title: An Appraisal of the 1950 Census Income Data

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

Volume Publisher: Princeton University Press

Volume ISBN: 0-691-04102-4

Volume URL: http://www.nber.org/books/unkn58-2

Publication Date: 1958

Chapter Title: Income Reported in the 1950 Census and on Income Tax
Returns

Chapter Author: Herman P. Miller, Leon R. Paley

Chapter URL: http://www.nber.org/chapters/c1053

Chapter pages in book: (p. 177 - 204)

Income Reported in the 1950 Census and on Income Tax Returns

HERMAN P. MILLER AND LEON R. PALEY,

BUREAU OF THE CENSUS

About forty years ago, a group of economists and statisticians met to discuss, among other things, the adequacy of income statistics in the United States. Papers were presented to this gathering by Allyn Young, who later became president of the American Economic Association, and by Willford King, one of the early pioneers in income analysis.[1] Both authors lamented the poor quality of the scanty income data available at that time, but both were optimistic about the future.

Since that meeting on the eve of our entry into World War I, a considerable volume of statistics on the size distribution of income has been collected in the United States. There is still a pressing need for more income data, but there is also need for analysis and appraisal of the data already at our disposal. We who participate in collecting such information are duty-bound to investigate its reliability. This paper reports on one phase of such an investigation.

Various procedures were designed to check the accuracy of the income statistics collected in the 1950 Census of Population. The one described here was a matching study of the income reported to the census interviewer and to the Internal Revenue Service (IRS) by a representative sample of the population. The study was conducted solely for the purpose of checking the accuracy of income statistics collected in the census. At every stage of the investigation precautions were taken to keep data pertaining to specific individuals confidential; all tabulations dealt with subgroups of the population, never with individual families.

Description of Census-Tax Return Match

The primary vehicle used in the census-tax return matching study was the Post-Enumeration Survey (PES) sample, a probability sample of about 25,000 households created for the specific purpose

[1] Papers presented at a joint meeting of the American Economic and American Statistical Associations in December 1916: Allyn A. Young, "Do the Statistics of the Concentration of Wealth in the United States Mean What They Are Commonly Assumed to Mean?"; and Willford I. King, "Desirable Additions to Statistical Data on Wealth and Income," *Quarterly Publication of the American Statistical Association*, March 1917, pp. 471–501.

of evaluating the 1950 census results.[2] Theoretically, all households in the sample should have been interviewed in the 1950 census; and one-fifth of them, or about 5,000 households, should have reported on income. In September 1950, about four months after completion of the census, these households were reinterviewed, and the more detailed information obtained was compared with the original census data for the identical households to measure the accuracy of the original data.

In November 1951, a year after the completion of the PES, the identifying information for each of the 5,000 households in the income subsample of the PES sample was copied on a separate form. Each household member fourteen years old or over was identified. If there were more than three such members in the household, additional matching forms were used. Each person's name, address, and social security number were copied from the PES schedule, his age from the census schedule. Income information was not copied from either the census or the PES schedule at this time to eliminate the possibility of disclosure of census income information to the IRS, and to minimize the possibility of bias in matching with income tax returns.

Equipped with this information, Census Bureau clerks (sworn in as IRS employees) attempted to locate the tax returns for the specified persons at IRS offices throughout the country, using a manual of instructions prepared for the purpose. Files were searched for the 1949 tax return filed by the head of each household, and for separate returns filed by other members of the household if their names did not appear on the head of the household's return. The results of all searches and the income information from the tax returns were entered on the matching forms, which were then submitted to the Census Bureau for collation with income and other information obtained in the 1950 census and the PES. A separate punch card was made for each family and unrelated individual (as determined by the standard Census Bureau definition).

To minimize the conceptual differences between the census and IRS definitions of income, the clerks were instructed to transcribe from the tax returns:

1. Wage and salary income
2. Net, not gross, income from a business, farm, or profession, or from rents and royalties

[2] For a detailed discussion of the plan of the Post-Enumeration Survey, see Eli S. Marks, W. Parker Mauldin and Harold Nisselson, "The Post-Enumeration Survey of the 1950 Censuses: A Case History in Survey Design," *Journal of the American Statistical Association,* June 1953, pp. 220–243; see also the paper by Leon Pritzker and Alfred Sands in this volume.

3. Gross, not merely taxable, income from annuities and pensions
4. All other income except capital gains or losses

Even so, some elements of difference remained. Certain kinds of income other than earnings, such as unemployment compensation, veterans payments, dependency allotments, and other types described below (page 189) were not reported on tax returns but were reported to the census interviewer. These items could not be deducted from the census reports because they had been recorded as part of a lump sum that included receipts from all sources other than earnings. Using procedures described below (page 189), separate tabulations were made for families roughly classified as "with taxable income only" so that families whose income was reported according to roughly the same concept in both sources could be identified.

How representative was the PES sample with respect to income? And how successful was the attempt to match this sample in the IRS files? The answer to the first question is indicated in the first two tables. Table 1 shows the 1950 census percentage distribution of families and individuals by income class, as estimated from the full 20 per cent sample used to obtain income estimates in the census and as transcribed for all families and individuals in the PES sample. (It should be emphasized that the income distribution for the PES sample is based on information obtained in the census itself and not in the PES interview.) This information was copied from the basic census schedule after the IRS data had been entered. For both families and individuals, the census income distribution based on the PES sample closely approximates the published census results. Differences are statistically insignificant; for all families, there is only a $60 difference between the two medians; for unrelated individuals, only a $3 difference. A closer look at the distributions in Table 1 indicates that in the two samples the proportion of families at each income level varies by no more than 1 percentage point, and the proportions of unrelated individuals at each income level are almost equally close.

While Table 2 presents less detail by income class than Table 1, it provides better insight into the structure of the sample by showing for each group the proportions of families and individuals by color and residence, as well as their median incomes. Unfortunately, the census data for this comparison had to be obtained from Preliminary Sample Tabulations (PST) of census returns rather than from the full 20 per cent census sample, because these data were not included in the final census tabulations. However, since the

TABLE 1

Income Reported in 1950 Census for Families and Unrelated Individuals:
20 Per Cent Sample and PES Sample

	20 PER CENT SAMPLE [a]		PES SAMPLE [b]	
INCOME IN CENSUS	Families	Unrelated Individuals	Families	Unrelated Individuals
	(number)			
Total in sample	38,310,980	11,051,050	7,131	1,409
Reporting on income	36,439,955	10,049,135	6,730	1,328
	(per cent)			
Total reporting	100.0	100.0	100.0	100.0
Under $ 500	8.3	31.8	8.1	29.5
$ 500– 999	6.4	18.3	6.1	20.5
1,000– 1,499	7.1	12.4	6.8	13.3
1,500– 1,999	7.5	9.2	7.7	9.2
2,000– 2,499	9.7	8.7	9.2	7.7
2,500– 2,999	9.4	6.1	9.0	5.2
3,000– 3,499	11.0	5.0	11.6	5.7
3,500– 3,999	8.4	2.7	9.4	2.9
4,000– 4,499	7.1	1.7	6.7	1.4
4,500– 4,999	5.0	0.9	4.6	1.1
5,000– 5,999	7.8	1.2	8.2	0.6
6,000– 6,999	4.3	0.6	4.8	1.3
7,000– 9,999	4.9	0.6	4.4	0.2
10,000 and over	3.1	0.7	3.2	1.4
	(dollars)			
Median income	3,073	997	3,129	1,000 .

Note: "Income" in this and the following tables refers to total money income unless otherwise qualified.

[a] *1950 Census of Population,* Vol. II, *Characteristics of the Population,* Part 1, *United States Summary,* Table 57.

[b] Income reported in the 1950 census for families and individuals included in the Post-Enumeration Survey sample. The numbers shown are weighted sample cases and not inflated population estimates, that is, each sample case was weighted by a factor required to make the sample self-weighting, but was not inflated to meet population controls. To cite an unrealistic example, if the rural population was sampled at one-half the rate of the urban population, each rural card would have been counted twice and each urban card once but neither card would have been multiplied by the weight (let us assume 5,000) required to inflate it to the independent estimate of the population. The figures exclude the small number of families and individuals (207 weighted sample cases) who were missed in the census but were interviewed in the PES. However, they include the equally small number of families and individuals (113 weighted sample cases) erroneously included in the census.

PST results were close to the final census results wherever comparisons could be made, there is some justification for the use of these data. The striking fact indicated in Table 2 is that the PES and PST results agree closely in both the proportions and the median incomes for each residence and color subgroup. Almost all the sample proportions differ by less than 1 percentage point, and the greatest

TABLE 2

Median Income in 1950 Census for White and Nonwhite Families and Individuals, by Residence: PST Sample and PES Sample

COLOR AND RESIDENCE	PST SAMPLE [a]		PES SAMPLE [b]	
	As % of Sample	Median Income	As % of Sample	Median Income
Families	100.0	$3,068	100.0	$3,129
Nonfarm	85.5	3,245	86.4	3,307
White	78.4	3,379	79.6	3,427
Nonwhite	7.1	1,658	6.8	1,681
Farm	14.5	1,733	13.6	1,648
White	12.9	1,937	11.1	2,053
Nonwhite	1.6	730	2.5	605
Unrelated individuals	100.0	$ 987	100.0	$1,000
Nonfarm	94.2	1,021	93.0	1,070
White	81.2	1,079	80.1	1,115
Nonwhite	13.0	814	13.0	883
Farm	5.8	651	7.0	b
White	5.0	702	6.4	b
Nonwhite	0.7	420	0.6	b

[a] Derived from *1950 Census of Population, Employment and Income in the United States, by Regions, 1950,* Series PC-7, No. 2, Table 9. This report was based on Preliminary Sample Tabulations (PST) of 1950 census returns. In general, the results of this preliminary sample agreed very favorably with the final census tabulations. However, the data by color shown above were never tabulated on the basis of the full 20 per cent sample of census returns which contained income data.

[b] Median not shown where there were fewer than 100 cases in the sample reporting on income in census.

discrepancy between any two medians is about $125—statistically insignificant differences.

The above facts leave little doubt that the PES sample matched with tax returns is an adequate reflection of the universe from which it was drawn. It now remains to be determined how successful the matching was. This question cannot be answered unequivocally because the nonmatched group includes families and individuals whose tax returns could not be located, as well as those not required to file returns because their incomes were too low. The problem is not quite so simple as it was depicted by King Gama in Gilbert and Sullivan's *Princess Ida,* when he said:

"I know everybody's income and what everybody earns;
And I carefully compare it with the income tax returns."

Results of the Matching Process

As a first attempt to measure the degree of success in locating tax returns for the appropriate families and individuals we will first consider the number of returns completely matched, partly matched,

or not matched at all, without regard to whether or not they should have been matched. A complete match means that each person in the family aged fourteen years or over was identified on a tax return; a partial match, one in which some family member aged fourteen years or over could not be located on tax returns; and a nonmatch, one in which no tax return was located for the family. The actual figures for each of these groups are shown in Table 3.

TABLE 3

Families and Unrelated Individuals by Match Status, Actual and Adjusted,
and by Residence

RESIDENCE	ACTUAL MATCH STATUS				ADJUSTED MATCH STATUS		
	Completely Matched	Partly Matched	Not Matched	TOTAL	Completely Matched [a]	Partly Matched	Not Matched
	(number)						
Families	3,903	946	2,242	7,091 [b]	5,340	634	1,117
Nonfarm	3,553	805	1,762	6,120	4,539	568	1,013
Farm	350	141	480	971	801	66	104
Unrelated individuals	334	——	1,002	1,336 [b]	717	——	619
Nonfarm	321	——	923	1,244	c	——	c
Farm	13	——	79	92	c	——	c
	(per cent)						
Families	55.0	13.3	31.6	100.0	75.3	8.9	15.8
Nonfarm	58.1	13.2	28.8	100.0	74.2	9.3	16.6
Farm	36.0	14.5	49.4	100.0	82.5	6.8	10.7
Unrelated individuals	25.0	——	75.0	100.0	53.7	——	46.3
Nonfarm	25.8	——	74.2	100.0	c	——	c
Farm	d	——	d	100.0	c	——	c

—— Equals zero or less than 0.05 per cent in this and the following tables.
[a] Including those implicitly matched; see the discussion in the text for adjustments made.
[b] These numbers are slightly lower than the comparable numbers shown in Table 1 because the figures in the above table exclude families and individuals who were missed in the census as well as those who were erroneously included in the census. The latter were included in Table 1 because they were part of the universe which the PES sample was designed to approximate; however, because of their relatively small number they are excluded from this table and all later tables.
[c] The breakdown into nonfarm and farm was not calculated for this group.
[d] Percentages not shown where there were fewer than 100 cases in the sample base.

About 55 per cent of all the families were completely matched, 13 per cent partly matched, and 32 per cent not matched at all. The matching rate was much higher for nonfarm families (about 58 per cent) than it was for those living on farms (36 per cent). This difference is probably largely due to the lower incomes of the farm families.[3] The matching rate for individuals (25 per cent) was

[3] Part of this difference may also be attributed to possibly greater compliance with tax laws on the part of the nonfarm population, as suggested by F. D.

below that obtained for families. The main reason for the low rate is that unrelated individuals usually receive low incomes, often below the tax filing requirement. They also may be more dependent on nontaxable income, such as pensions. And finally, they move more often than families do, and so their tax returns are harder to find.

The fact that a rather large proportion of the families and individuals could not be matched with tax returns is not in itself sufficient evidence of a shortcoming in the study. The unmatched group includes families with taxable incomes whose tax returns could not be located, as well as other families who did not file a return, apparently either because their income was below the tax-filing requirement or because it was not high enough to require a tax payment, when allowance was made for dependents. Part of this non-filing group could be implicitly matched by adopting certain assumptions.

All of the 3,188 families not matched with tax returns reported on size of family in the census and all but 183 reported their income in the census. If the income reported for these families in the census is assumed correct, and if the standard tax exemption of $600 is allowed for each family member, then incompletely matched or unmatched two-person families with incomes under $1,200, three-person families with incomes under $1,800, and so forth, can be regarded as implicitly matched. This is based on the supposition that most of these families did not file a tax return because they did not have to pay a tax or claim a tax refund. In the case of the self-employed, where taxes are not withheld at the source, many farmers and shopkeepers may regard it as relatively unimportant to file a return if no tax payment is required. Even among wage and salary workers, if no taxes are withheld at the source as a result of allowances for dependents, the question of filing a tax return may often become purely a matter of technical compliance with the law.

Under the foregoing assumptions, the adjusted matching rates shown in Table 3 are considerably greater than the unadjusted rates. Whereas 58 per cent of the nonfarm families were completely matched, about 74 per cent were matched when allowance is made for the implicit matches as defined above. Among farm families, the matching rate was increased from 36 to 83 per cent by the adjustment procedure.

Although the adjusted matching rate for unrelated individuals (54 per cent) was about twice the unadjusted rate, it was still much

Stocker, *The Impact of Federal Income Taxes on Farm People,* Dept. of Agriculture, ARS 43–11, July 1955, p. 29.

lower than the rate for families. The rate for unrelated individuals is lower partly because many with total incomes over $600 had lower taxable incomes and could not be readily identified as implicitly matched. Implicit matches were estimated only to provide a more refined index of matching success; however, they are treated as nonmatches or partial matches in the tabulations by income class in Tables 4 through 6.

The validity of the matching study depends not only on how many matched households there are but also on how representative they are, that is, on how large a proportion is successfully matched at each income class. Since the lowest income groups are not required to file tax returns, one would expect the nonmatched families to have lower incomes than the matched and Table 4 clearly shows this was so. About one-third of the nonmatched families had incomes under $1,000 compared with only 5 per cent of the completely matched group. The median income for nonmatched families

TABLE 4

Income in 1950 Census for Families and Individuals by Match Status

(weighted sample cases)

	FAMILIES				UNRELATED INDIVIDUALS		
INCOME IN CENSUS	TOTAL	*Completely Matched*	*Partly Matched*	*Not Matched*	TOTAL	*Matched*	*Not Matched*
				(number)			
Total in sample	7,091	3,903	946	2,242	1,336	334	1,002
Reporting on income	6,692	3,687	893	2,112	1,259	311	948
				(per cent)			
Total reporting	100.0	100.0	100.0	100.0	100.0	100.0	100.0
None	3.9	1.5	3.4	8.2	15.9	6.4	19.0
Loss	0.1	0.1	—	0.1	0.6	0.6	0.5
$ 1–$ 499	4.2	1.6	2.6	9.3	13.6	8.4	15.3
500– 999	6.1	2.1	4.7	13.4	20.6	15.1	22.4
1,000– 1,499	6.7	4.1	8.0	10.7	13.1	12.2	13.4
1,500– 1,999	7.8	5.7	8.4	11.1	8.6	10.0	8.1
2,000– 2,499	9.2	8.9	11.8	8.8	7.8	13.5	5.9
2,500– 2,999	9.0	10.7	9.2	6.0	4.8	6.8	4.2
3,000– 3,499	11.6	14.4	10.5	7.3	5.7	6.4	5.5
3,500– 3,999	9.4	12.4	6.7	5.3	3.1	7.7	1.6
4,000– 4,499	6.7	8.3	7.1	3.7	1.5	1.9	1.4
4,500– 4,999	4.6	5.1	4.6	3.7	1.1	3.2	0.4
5,000– 5,999	8.2	9.6	11.1	4.6	0.6	0.6	0.6
6,000– 6,999	4.9	6.0	4.9	2.9	1.4	4.2	0.4
7,000– 9,999	4.4	5.6	4.5	2.3	0.2	1.0	—
10,000 and over	3.3	3.8	2.7	2.5	1.4	1.9	1.3
				(dollars)			
Median income	3,133	3,534	3,098	1,868	986	1,863	840

was only $1,868 compared with $3,098 for the partly matched group, and $3,534 for the completely matched. Similarly, the median income for the nonmatched unrelated individuals ($840) was only about one-half that shown for the matched group ($1,863). The matched sample of families had a somewhat higher income than the universe from which it was drawn. This bias, however, is largely eliminated when allowance is made for the fact that a relatively large proportion of the nonmatched group were not required to file tax returns and are therefore implicitly matched. It is more difficult to appraise the representativeness of the matched sample of unrelated individuals because of the large proportion of nonmatched cases, and also because of the difficulty of adjusting the matching rate to include implicitly matched cases.

Another way of examining the income distributions is to compare the proportion of cases matched in each income class. This is done in Table 5. The proportion of completely matched families increased progressively from about 20 per cent for those with incomes under $1,000 to a maximum of 73 per cent for those in the $3,500 to $3,999 income class. In higher income classes the matching rate fluctuated between 60 and 70 per cent. Among families with incomes of $5,000 and over, where there are relatively few

TABLE 5

Match Status of Families by 1950 Census Income Class

(*based on weighted sample cases*)

INCOME IN CENSUS	TOTAL	Completely Matched	Partly Matched	Not Matched
		(*per cent*)		
Total reporting	100.0	55.0	13.3	31.7
None	100.0	21.6	11.6	66.8
Loss	100.0	a	a	a
$ 1–$ 499	100.0	20.9	8.3	70.9
500– 999	100.0	19.5	10.4	70.1
1,000– 1,499	100.0	33.5	15.8	50.7
1,500– 1,999	100.0	40.6	14.4	45.0
2,000– 2,499	100.0	53.1	17.0	29.9
2,500– 2,999	100.0	65.6	13.6	20.9
3,000– 3,499	100.0	68.1	12.1	19.8
3,500– 3,999	100.0	72.9	9.5	17.6
4,000– 4,499	100.0	68.4	14.0	17.6
4,500– 4,999	100.0	60.9	13.4	25.7
5,000– 5,999	100.0	64.3	18.0	17.7
6,000– 6,999	100.0	67.7	13.5	18.8
7,000– 9,999	100.0	70.0	13.5	16.5
10,000 and over	100.0	64.7	11.0	24.3

a Percentages not shown since there were fewer than 100 cases in the sample reporting "Loss" in the census.

implicitly matched cases and where failure to match must be generally attributed to technical difficulties inherent in the matching process, about two-thirds of the families were completely matched. Although the matching rate is not impressively high for any income class, there is no evidence of a bias in the ability to match different groups of families within the income class in which one can reasonably expect to establish a match.

One final factor to be considered in appraising the validity of the matching study is the income level of families and individuals who did not report income in the census, since a marked bias in this group would seriously challenge the representativeness of the matched group. There is little evidence that such a bias exists. The nonresponse rate for income questions in the census as a whole was only 6 per cent, and tax returns were located for 54 per cent of the families in the sample who did not report on income in the census (see Table 6). The median income on tax returns for these families ($3,707) was only slightly higher than that for all matched

TABLE 6

Income on Tax Returns for Families Who Did Not Report on Income in the 1950 Census

(weighted sample cases)

INCOME ON TAX RETURNS	*Families*
	(number)
Total	399
Completely matched	216
Not completely matched	183
	(per cent)
Total completely matched	100.0
Loss	1.9
$ 1–$ 499	0.9
500– 999	1.9
1,000– 1,499	2.3
1,500– 1,999	6.5
2,000– 2,499	13.9
2,500– 2,999	6.0
3,000– 3,499	8.8
3,500– 3,999	19.0
4,000– 4,499	8.3
4,500– 4,999	2.8
5,000– 5,999	9.3
6,000– 6,999	6.5
7,000– 9,999	10.2
10,000 and over	1.9
	(dollars)
Median income	3,707

families ($3,591). The number of unrelated individuals in the sample who did not report income in the census was too small to permit comparisons.

Comparison of Census and Tax Return Income Distributions for Matched Groups

Turning from a discussion of the representativeness of the sample and the degree to which the sample households were matched with tax returns, we can now compare the income information reported in the census and on tax returns by those families and individuals who were completely matched in the tax files. To sharpen the analysis, information is shown separately for all matched groups and for those roughly identified as not having reported any "nontaxable" income in the census.

All families and individuals who reported no income other than earnings were immediately classified in the latter category. Then PES schedules were located for families and individuals who, in the census, either did not report on such income or reported receiving one dollar or more of it. The PES schedule contained separate information on each of the following categories of income other than earnings: (1) unemployment or workmen's compensation; (2) social security benefits and government pensions or assistance; (3) other pensions and allowances; (4) veterans payments; (5) dependency allotments; (6) interest, dividends, and income from estates and trusts; (7) receipts from roomers and boarders; (8) rents and royalties; (9) money for support from persons not living in the household. Families and individuals reporting no income from items 1 through 5 and item 9 were classified as "with taxable income only"; all others were classified as having nontaxable income or as not reporting on taxable income.[4] Certain types of income, such as interest from nontaxable securities and Railroad Retirement pensions, may have been correctly reported in the census (the amount actually reported was very small) and correctly omitted from the tax returns. Similarly, the tax returns may include the cash value of certain types of income "in kind," which would have been

[4] The following additional criterion was used in the classification of families and individuals by taxable income. If there was a difference in type and size of family between the census and PES, the family was excluded from the group under analysis—"with taxable income only," in order to retain in this category only those families who reported the same composition in the PES and the census, and for whom all income information should have been reported for all family members in the census and on tax returns. Only 3 per cent of the matched families showed a difference in type or size of family between the PES and the census.

excluded from the census data. But in spite of many inadequacies, the procedure provides a reasonable, objective basis for roughly identifying families and individuals whose income in the census and on the tax returns was based on approximately the same income concept.

As might have been expected, the matching rate for families with taxable income only was higher (about 59 per cent) than that for other families (55 per cent). The same was true for unrelated individuals, 28 per cent, compared with 25 per cent for all individuals in the sample. Table 7 presents a summary picture of the income distributions in the census and on the tax returns for completely matched families and individuals.

There is a marked similarity between the census and tax return distributions for all families by income class; the maximum difference for any given class was only 2 percentage points and for most classes the difference was less than 1 percentage point. The discrepancy of $57 between the medians is not statistically significant. The same conclusions apply to families with taxable income only. Exclusion of families with nontaxable income tended to raise the level slightly for both distributions but had a greater impact on the tax return than on the census distribution. The explanation of this higher level is that nontaxable income, as defined here, consists largely of veterans payments, social security payments, public assistance payments, and other types of income usually received by low income groups.

The differences between the census and tax return income distributions are much greater for unrelated individuals than for families. Although the difference between the medians for all unrelated individuals is not significant ($90), the differences for given income classes are often quite substantial (Table 7). In part, the differences by income class are due to the relatively small number of unrelated individuals in the matched sample, but many are too great to be attributed to sampling error. For example, the census distribution shows about 6 per cent of matched individuals with no income, while none could appear in the tax return distribution. However, if one considers the proportion of individuals with incomes under $1,000, instead of just the "zero-income" group, the two distributions are more comparable—about 31 per cent in the census distribution, 28 per cent in the tax return distribution. This would indicate that while the zero-income individuals were classified at too low an income level in the census, most of this group clearly seems to have received less than $1,000 during 1949.

The similarity of the income distributions in the census and on

TABLE 7

Matched Families and Unrelated Individuals by Income in 1950 Census and on Tax Returns

(weighted sample cases)

INCOME IN CENSUS OR ON TAX RETURNS [a]	MATCHED FAMILIES [b]				MATCHED INDIVIDUALS [b]			
	Total		With Taxable Income Only [c]		Total		With Taxable Income Only [c]	
	Census	Tax Returns	Census	Tax Returns	Census	Tax Returns	Census	Tax Returns
	(number)							
Total in sample	3,903	3,903	3,286	3,286	334	334	280	280
Reporting on income	3,687	3,903	3,118	3,286	311	334	266	280
	(per cent)							
Total reporting	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
None	1.5	—	1.7	—	6.4	—	6.8	—
Loss	0.1	0.4	0.1	0.5	0.6	1.5	0.8	1.8
$ 1–$ 499	1.6	1.1	1.4	0.9	8.4	9.6	6.8	7.1
500– 999	2.1	3.7	2.0	3.5	15.1	18.6	12.8	18.9
1,000– 1,499	4.1	4.7	3.6	3.7	12.2	11.4	11.7	10.0
1,500– 1,999	5.7	6.7	5.3	6.4	10.0	16.5	11.7	15.0
2,000– 2,499	8.9	8.2	8.9	8.0	13.5	11.7	14.3	13.2
2,500– 2,999	10.7	10.1	10.7	10.6	6.8	9.6	7.9	10.7
3,000– 3,499	14.4	13.3	14.6	13.2	6.4	3.3	7.1	3.6
3,500– 3,999	12.4	10.4	12.8	9.8	7.7	8.7	6.8	9.6
4,000– 4,499	8.3	7.7	7.8	8.1	1.9	2.1	2.3	2.1
4,500– 4,999	5.1	7.0	5.3	7.5	3.2	1.2	3.0	1.4
5,000– 5,999	9.6	10.9	10.0	11.6	0.6	3.6	0.8	4.3
6,000– 6,999	6.0	6.7	6.5	7.3	4.2	—	4.9	—
7,000– 9,999	5.6	5.5	5.7	5.5	1.0	0.6	1.1	0.7
10,000 and over	3.8	3.7	3.7	3.4	1.9	1.8	1.5	1.4
	(dollars)							
Median income	3,534	3,591	3,569	3,669	1,863	1,773	1,984	1,905

[a] The income classes refer to the amount of income in the 1950 census when columns marked Census are used, and by the amount of income on tax returns when the columns marked Tax Returns are used. (This applies also to Tables 8, 9, and 10.)

[b] In this and the following tables "matched" refers to "completely matched" families or unrelated individuals.

[c] As defined in the text. (This applies also to Table 8.)

tax returns for all families masks important differences in the under-
lying figures. One is revealed in Table 8, a comparison of the distri-
butions for farm and nonfarm families. The census and tax return
income distributions for nonfarm families are very similar. The
discrepancy between the medians for all nonfarm families is only
about $100, and for nonfarm families with taxable income only,
about $150. But the differences between the distributions for farm
families are striking. The median income for all farm families was
$2,800 in the census compared with about $2,300 on tax returns.
If the farm families with taxable income only are considered in-
stead, the medians are closer; $2,700 in the census, $2,300 on tax
returns. The differences between the *medians* for farm families are
symptomatic of relatively large differences in the underlying distri-
butions. Thus the census data show 8 per cent of all farm families
with incomes under $500; the tax returns, 2 per cent. If the group
is increased to all under $1,000, the proportions are much closer—
12 per cent for the census and 15 per cent on tax returns. However,
for the $1,000 to $1,999 income class, the difference is again strik-
ing—17 per cent for the census and 27 per cent on tax returns.
Equally large differences were found for several other income
classes. At the upper end of the farm income distribution ($6,000
and over) the census data show about 7 per cent of the families
compared with only 3 per cent on tax returns.

The relatively large difference between the census and tax re-
turn income distributions for farm families is consistent, at least in
direction, with facts obtained in the Audit Control Program (ACP)
conducted by the Internal Revenue Service for a representative
sample of tax returns for 1948.[5] One of the striking facts discovered
in this study was that "mistakes on erroneous tax returns aggregated
about $1.5 billion of tax change, or 10 per cent of the total tax li-
ability voluntarily reported." [6] Even more important for the present
discussion is the fact that about one-half of the $1.5 billion was
made on returns with income or loss from business, farm, or profes-

[5] A description of the procedures and some preliminary results of the Audit
Control Program may be found in: *The Audit Control Program: A Summary
of Preliminary Results, May 1951,* Bureau of Internal Revenue; and in three
papers by Marius Farioletti, "The 1948 Audit Control Program for Federal In-
come Tax Returns" (*National Tax Journal,* June 1949); "Some Results of the
First Year's Audit Control Program of the Bureau of Internal Revenue" (*Na-
tional Tax Journal,* March 1952); and the one in this volume.

[6] M. Farioletti, "Some Results of the First Year's Audit Control Program of the
Bureau of Internal Revenue," p. 66. This article further notes that "about $1.4
billion or more than nine-tenths of the total tax change is estimated to involve
additional assessments, and somewhat less than $100 million involves over assess-
ments."

TABLE 8

Matched Families by Income in 1950 Census and on Tax Returns and by Residence

(weighted sample cases)

INCOME IN CENSUS OR ON TAX RETURNS	MATCHED NONFARM FAMILIES				MATCHED FARM FAMILIES			
	Total		With Taxable Income Only		Total		With Taxable Income Only	
	Census	Tax Returns	Census	Tax Returns	Census	Tax Returns	Census	Tax Returns
	(number)							
Total in sample	3,553	3,553	3,001	3,001	350	350	285	285
Reporting on income	3,351	3,553	2,845	3,001	336	350	273	285
	(per cent)							
Total reporting	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
None	1.4	—	1.5	—	2.4	—	2.9	—
Loss	0.1	0.4	0.1	0.4	0.6	1.1	0.7	1.4
$ 1–$ 499	1.3	1.1	1.1	0.8	4.8	1.1	5.1	1.4
500– 999	1.9	2.8	1.7	2.6	4.2	12.9	4.4	12.6
1,000– 1,499	3.4	3.8	2.9	2.8	10.7	13.4	10.3	12.3
1,500– 1,999	5.6	6.0	5.1	5.6	6.5	13.7	7.0	14.0
2,000– 2,499	8.3	7.9	8.3	7.6	14.6	11.4	15.0	11.9
2,500– 2,999	10.8	9.9	10.8	10.3	10.4	12.3	9.9	13.7
3,000– 3,499	14.1	13.5	14.2	13.4	17.6	10.9	18.7	11.2
3,500– 3,999	12.9	10.8	13.4	10.3	8.0	6.0	7.0	4.6
4,000– 4,499	8.7	7.9	8.0	8.3	5.1	5.1	5.5	5.6
4,500– 4,999	5.5	7.2	5.8	7.8	1.2	4.6	0.7	4.9
5,000– 5,999	9.8	11.5	10.4	12.5	6.8	4.6	6.2	2.8
6,000– 6,999	6.1	7.3	6.7	7.8	4.8	1.1	3.7	1.4
7,000– 9,999	6.1	6.0	6.1	6.0	1.2	0.6	1.5	0.7
10,000 and over	4.1	3.9	3.9	3.6	1.2	1.1	1.5	1.4
	(dollars)							
Median income	3,621	3,718	3,659	3,807	2,800	2,338	2,731	2,346

sional practice.[7] Since nine-tenths of the tax change on returns with business and professional incomes involved additional assessments (probably indicating underreporting of income), it is of interest to compare the census and tax return income distributions for families headed by farm operators and by nonfarm businessmen. In Table 9, all matched families are classified by total income in the census and on tax returns, and by the occupation of the family head.

Although the cross-classification of the sample of matched families by both income and occupation produces distributions with relatively small frequencies and, therefore, relatively large sampling errors, certain significant patterns appear. The medians show very similar income levels in both sources for several groups of families headed by persons who are predominantly wage or salary workers: salaried professional and managerial workers; clerical and sales workers; craftsmen and operatives; laborers and service workers. None of the above pairs of medians were significantly different in a statistical sense (that is, the variations were all within the range of sampling error of the estimates). In contrast, the census median income for farm operator families was about $800 higher than that reported on tax returns, and for families headed by self-employed professional and managerial workers (which includes proprietors of unincorporated businesses), about $500 higher.

The findings noted in the occupation analysis are. generally confirmed by the figures in Table 10, which gives the distributions by type of income in the census and on tax returns. The striking similarity between the two distributions of total income and the median incomes for families headed by wage workers is borne out by the wage or salary income data. Furthermore, the sharp disparity between the self-employment income distributions confirms the differences previously noted in the total income distributions for families headed by farm operators or by self-employed professional and managerial workers.

The census and tax returns distributions of income other than earnings received by all families bear a strong resemblance to each other. According to both sets of data, four-fifths of the families received less than $1,000 and about 4 per cent received over $5,000 in unearned income; the median was the same in both cases—$400. This picture is somewhat deceptive, however, because of the marked differences in the definition of income other than earnings in the census and on tax returns. If one looks at families with taxable income only, a weakness in the census. data becomes apparent. Only

[7] *Ibid.*, p. 77.

TABLE 9

Matched Families by Income in 1950 Census and on Tax Returns and by Occupation of Head in April 1950

(weighted sample cases)

TOTAL INCOME IN CENSUS OR ON TAX RETURNS	SELF-EMPLOYED PROFESSIONAL AND MANAGERIAL WORKERS		SALARIED PROFESSIONAL AND MANAGERIAL WORKERS		FARMERS AND FARM MANAGERS		CLERICAL AND SALES WORKERS		CRAFTSMEN AND OPERATIVES		LABORERS AND SERVICE WORKERS	
	Census	Tax Returns	Census	Tax Returns	Census	Tax Returns	Census	Tax Returns	Census	Tax Returns	Census	Tax Returns
	(number)											
Total in sample	307	307	557	557	230	230	473	473	1,574	1,574	413	413
Reporting on income	268	307	527	557	216	230	445	473	1,521	1,574	380	413
	(per cent)											
Total reporting	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
None	4.9	—	0.2	—	1.9	—	—	—	0.3	—	2.1	—
Loss	—	1.3	—	0.2	—	2.6	0.4	—	—	0.1	0.5	0.7
$ 1–$ 499	0.7	1.6	0.4	0.4	4.6	0.9	0.9	0.8	1.0	0.6	1.1	3.4
500– 999	6.0	7.8	0.4	0.5	6.9	10.9	2.0	1.9	1.1	2.0	1.6	3.9
1,000– 1,499	2.6	3.9	1.1	1.4	10.2	14.3	6.5	6.6	2.9	4.1	6.3	12.1
1,500– 1,999	2.6	5.9	1.5	1.1	4.2	17.0	8.8	9.5	4.7	5.1	8.7	12.3
2,000– 2,499	8.2	10.4	4.6	3.2	11.1	11.7	10.8	8.9	9.1	7.7	15.3	11.1
2,500– 2,999	6.7	10.4	6.3	5.2	11.6	11.7	15.7	16.7	12.9	11.7	16.1	18.2
3,000– 3,499	8.2	11.1	7.0	9.2	19.0	9.6	13.9	13.1	16.8	13.7	19.2	10.2
3,500– 3,999	11.6	3.6	13.9	9.5	2.8	2.2	12.6	8.0	14.5	13.9	10.3	7.3
4,000– 4,499	9.7	3.6	5.7	5.7	5.1	4.3	4.5	5.1	9.7	10.4	5.5	6.3
4,500– 4,999	3.0	3.6	6.8	8.8	0.9	2.6	8.5	12.1	6.3	9.3	4.5	8.4
5,000– 5,999	10.1	13.7	13.9	16.3	10.2	8.7	7.9	6.1	10.1	9.8	4.7	3.9
6,000– 6,999	3.4	3.3	11.8	19.7	7.4	0.9	3.8	6.8	4.9	5.3	3.2	2.2
7,000– 9,999	5.2	5.9	17.5	11.0	2.3	0.9	3.6	4.4	4.7	4.9	0.5	—
10,000 and over	17.2	14.0	9.1	7.7	1.9	1.7	—	—	1.2	1.3	0.5	—
	(dollars)											
Median income	3,935	3,390	5,158	5,291	2,980	2,185	3,673	3,714	3,546	3,679	2,951	3,177

TABLE 10

Matched Families by Type of Income in 1950 Census and on Tax Returns

(*weighted sample cases*)

INCOME IN CENSUS OR ON TAX RETURNS	WAGE OR SALARY INCOME		SELF-EMPLOYMENT INCOME				INCOME OTHER THAN EARNINGS			
			Nonfarm Residents		*Farm Residents*		*All Families*		*Families with Taxable Income Only*	
	Cen-sus	*Tax Re-turns*	*Cen-sus*	*Tax Re-turns*	*Cen-sus*	*Tax Re-turns*	*Cen-sus*	*Tax Re-turns*	*Cen-sus*	*Tax Re-turns*
					(*number*)					
Total in sample	3,918	3,918	3,556	3,556	352	352	3,908	3,908	3,290	3,290
Reporting on type of income	3,742	3,918	3,402	3,556	348	352	3,757	3,908	3,187	3,290
Reporting $1 or more	3,144	3,379	530	642	232	229	1,014	1,006	525	829
					(*per cent*)					
Total reporting $1 or more	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Loss	—	—	2.8	10.9	0.9	5.7	—	5.2	—	4.1
$ 1–$ 499	2.7	2.4	10.4	12.5	12.9	5.2	57.1	62.9	63.0	65.7
500– 999	2.6	3.3	9.6	9.5	9.5	19.2	20.7	13.6	15.2	12.9
1,000– 1,499	4.5	4.8	6.0	10.1	9.9	14.4	9.2	5.4	7.2	5.2
1,500– 1,999	6.6	6.2	6.4	8.7	10.3	10.9	3.8	2.2	2.3	1.9
2,000– 2,499	9.2	7.8	6.0	12.3	11.6	12.2	3.0	2.0	2.9	2.2
2,500– 2,999	11.8	10.4	8.3	7.6	3.9	11.4	1.6	0.9	1.9	0.8
3,000– 3,499	15.3	13.5	11.1	6.1	19.0	5.2	0.1	2.4	0.2	1.7
3,500– 3,999	12.0	12.0	8.1	1.2	3.4	2.6	0.4	0.6	0.4	0.5
4,000– 4,499	7.6	8.2	5.8	4.5	3.4	4.4	0.3	0.7	0.2	0.8
4,500– 4,999	5.7	6.3	3.0	2.8	0.9	—	0.2	0.4	0.4	0.5
5,000– 5,999	8.7	9.6	10.4	5.5	9.1	6.1	1.1	0.8	1.7	0.7
6,000– 6,999	6.7	7.5	2.6	1.1	2.6	—	0.8	0.4	1.1	0.5
7,000– 9,999	4.0	5.4	3.6	3.0	0.9	0.9	0.2	1.0	0.4	1.0
10,000 and over	2.6	2.6	5.7	4.2	1.7	1.7	1.6	1.6	3.0	1.4
					(*dollars*)					
Median income for those reporting income of specified type	3,412	3,570	3,017	1,902	2,278	1,750	439	356	398	350

one-sixth reported to the census receipts of one dollar or more of such income, but one-fourth of the families reported such receipts on tax returns. Although the census indicated a smaller proportion of recipients, the level and distribution of income reported was the same as that reported on tax returns.

*Cross-Classification of Income in Census
and on Tax Returns*

The data presented thus far were used primarily in classifications by the amount of income reported by each family and individual either in the census or on tax returns. Cross-classification by the amount of income reported in both of these sources so far has been discussed only in an examination of the income on tax returns for families who did not report on income in the census and an examination of the income reported in the census for families whose tax returns could not be matched. The cross-classification data helped to answer important questions on a potential source of bias in the census income reports and contributed to an evalution of the present matching study.

But these data can also be used to examine the consistency of response for identical families. In the preceding section a marked similarity between the census and tax return distributions was seen particularly for nonfarm families headed by wage or salary workers. This similarity could be the result of reports either of the same income to the census interviewer and on tax returns, or of different incomes but ones that in balance tended to offset each other. If there are offsetting differences, it is important to find out whether they are random errors attributable to faulty memory or if they are systematically related to the economic or demographic characteristics of the families and so may affect some of the cross-classifications of income with other variables.

Table 11, restricted to completely matched families which reported on income in the census, shows the percentage distribution of families in each income class in the census according to the amount of income reported on tax returns. The diagonal cells (indicated by rules) represent families found in the same income class in both sources. Some families in the diagonal cells may have reported relatively large differences, whereas those in adjacent cells may have reported relatively small ones, but here it is assumed that families in the diagonal cells reported their incomes consistently and that the other families did not. Clearly the variation in response for identical families is considerable. Only about 45 per cent of the families were found in the same income class in both the census and on tax returns. Another 24 per cent of the families were either in one higher or lower adjacent class in both sources. The proportion of families in the same class was greatest in the income classes above $5,000, largely because their minimum size was $1,000 compared with the uniform $500 classes used at the lower levels. This

TABLE 11

Matched Families by Income in 1950 Census and by Income on Tax Returns

(weighted sample cases)

INCOME IN CENSUS [a]	Total	INCOME ON TAX RETURNS [a]														
		Loss	$1 to $499	$500 to $999	$1,000 to $1,499	$1,500 to $1,999	$2,000 to $2,499	$2,500 to $2,999	$3,000 to $3,499	$3,500 to $3,999	$4,000 to $4,499	$4,500 to $4,999	$5,000 to $5,999	$6,000 to $6,999	$7,000 to $9,999	$10,000 and Over
						(per cent)										
$ 1,000–$1,499	100.0	—	—	16.0	38.7	16.0	4.0	6.0	2.7	4.7	1.3	—	1.3	8.0	—	1.3
1,500– 1,999	100.0	—	6.2	10.4	13.7	41.2	10.9	6.6	3.3	2.4	1.9	—	0.5	—	2.8	—
2,000– 2,499	100.0	—	1.2	3.0	6.7	16.2	33.8	13.1	9.5	5.8	1.5	2.1	4.0	0.6	1.8	0.6
2,500– 2,999	100.0	1.0	0.5	—	4.0	6.1	12.4	45.7	15.7	4.5	2.8	2.5	2.5	1.3	1.0	—
3,000– 3,499	100.0	—	0.4	2.8	2.6	3.6	8.1	10.2	46.6	10.4	4.7	4.3	2.8	2.3	0.8	0.4
3,500– 3,999	100.0	—	—	—	1.7	1.7	1.7	8.3	17.4	40.5	8.1	9.8	7.0	1.3	1.5	0.9
4,000– 4,499	100.0	0.7	—	—	1.3	1.0	3.3	2.6	10.7	12.4	43.3	11.4	6.5	3.6	1.3	2.0
4,500– 4,999	100.0	0.5	—	—	—	2.1	2.1	4.8	2.1	4.3	12.8	48.7	14.4	4.3	1.1	2.7
5,000– 5,999	100.0	0.6	1.1	—	—	0.3	3.7	1.1	2.8	4.5	4.8	9.6	55.8	7.9	5.9	1.7
6,000– 6,999	100.0	—	1.8	0.5	—	2.7	1.8	3.6	0.5	2.3	2.3	3.2	20.9	45.0	13.6	1.8
7,000– 9,999	100.0	1.0	—	—	—	1.0	1.4	1.0	2.4	—	3.8	1.9	9.1	30.8	47.1	0.5
10,000 and over	100.0	0.7	—	—	1.4	—	1.4	—	1.4	—	1.4	4.3	2.8	1.4	8.5	76.6

[a] No percentage shown for the first four census income classes— "none", "loss", $1–499, and $500–999—because there were fewer than 100 matched sample cases in each of these classes in the census.

fact would explain the extent of agreement (77 per cent) in the $10,000 and over class. In most income classes below $5,000, between 40 and 45 per cent of the families were in the same class in the census and on tax returns, the only exception being the $2,000–2,499 class, where only 34 per cent were in agreement.

Table 12 highlights the considerable response variation in income reporting by families headed by persons in different occupational

TABLE 12

Matched Families in Same or Different Income Class in 1950 Census and on Tax Returns by Occupation of Head in April 1950

(based on weighted sample cases)

		OCCUPATION OF FAMILY HEAD					
INCOME CLASS IN CENSUS AND ON TAX RETURNS	TOTAL REPORTING ON INCOME	*Professional and Managerial Workers*		*Farmers and Farm Managers*	*Clerical and Sales Workers*	*Craftsmen and Operatives*	*Laborers and Service Workers*
		Self Employed	*Salaried*				
				(per cent)			
Total reporting	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Same class	44.9	36.2	49.1	36.6	47.6	50.0	36.6
Adjacent class	24.3	17.5	31.5	18.5	21.2	23.6	27.1
Census higher	13.9	12.3	22.2	10.2	13.3	11.4	13.2
Tax returns higher	10.4	5.2	9.3	8.3	7.9	12.2	13.9
Difference of 2 or more classes	30.8	46.2	19.3	44.9	31.3	26.4	36.3
Census higher	14.4	27.2	8.3	33.3	11.7	11.7	14.2
Tax returns higher	16.4	19.0	11.0	11.6	19.6	14.7	22.1
				(dollars)			
Median income:							
Census	3,534	3,935	5,158	2,980	3,673	3,546	2,951
Tax returns	3,591	3,390	5,291	2,185	3,714	3,679	3,177

groups. Compared to all families in the same income class or in one higher or lower adjacent class in the census and on tax returns (about 70 per cent), families headed by self-employed persons or by laborers and service workers showed somewhat greater variation. About 35 per cent of these families were in the same income class and another 20 per cent in an adjacent one. In contrast, about 50 per cent of the other families (those headed by salaried professional and managerial workers, clerical and sales workers, and craftsmen and operatives) were in the same class and another 25 per cent in an adjacent one. Occupation groups having the most stable incomes show the least variation in response. However, even in these groups, only one-half of the families were in the same income class in both the census and on tax returns. Despite this extreme variation in response, however, the median incomes reported

in the census and on tax returns for most groups were quite similar. Only in the case of the self-employed did the median vary by more than $250.

All income matching studies conducted to date indicate that the variability of response in income surveys is relatively great.[8] Apparently only about half of the income recipients report receipts in the same income class even when asked a given set of income questions only one month apart. Fortunately, the variations in response elicited in repeated interviews appear to be random and do not introduce any systematic bias into the income distribution. The study of variation in response in field surveys of income does not preclude the existence of a systematic downward bias attributable to faulty memory, misunderstanding of the income concept, or misrepresentation. However, if one assumes that a given respondent uses a constant framework of reference from one month to the next, then his failure to report the same amount of income must be laid to random lapses of memory resulting in overstatements of income balanced by understatements.

Implications of Findings

In this study, income reports obtained in the 1950 census were compared with similar information reported by identical families on tax returns, on the premise that such a comparison would validate the census data. The results indicate a high degree of consistency between income distributions for wage or salary workers based on these two sources of data. Also, differences between the distributions for self-employed persons are consistent, in direction at least, with information obtained in field studies conducted by the Internal Revenue Service. On this basis, one could conclude that the census income data are quite accurate. The validity of the conclusion, however, depends entirely on the validity of the assumption that tax returns are accurate.

There may be some understatement of income on tax returns, though its precise amount is difficult to estimate. According to Selma Goldsmith, tax returns contained about a 14 per cent understatement of income in 1944 to 1946.[9] This figure was obtained by estimating both the aggregate income reported on tax returns and

[8] For a more detailed discussion of the variability of response in income surveys, as indicated by income matching studies, see Herman P. Miller, *Income of the American People*, Wiley, 1955, pp. 143–145, 157–164.

[9] Selma F. Goldsmith, "An Appraisal of Basic Data Available for Constructing Size Distributions," in Volume Thirteen (1951) of Studies in Income and Wealth, p. 302, and also her paper in this volume.

the comparable aggregate derived from the personal income series of the Department of Commerce. If Mrs. Goldsmith's analysis is correct, the underlying assumption that tax returns can be used as a validation check for census reports is questionable. The national income estimates may be more nearly correct, since they are among the most important and the most carefully prepared of all our statistical series. Yet as the national income estimators themselves admit, certain segments of the national income totals, such as entrepreneurial income, rental income, and interest, may be subject to substantial margins of error. And adjustment of the basic national income estimates for comparability with tax returns may be subject to even greater error.

COMMENT

JOSEPH A. PECHMAN, COMMITTEE FOR
ECONOMIC DEVELOPMENT

The matching studies summarized in this volume have long been awaited by the income distribution fraternity. This small band of diligent statisticians and economists has been struggling for years to evaluate and to reconcile the several income distributions appearing periodically, and the studies needed for this purpose have at last been made possible by the Census Bureau, the Internal Revenue Service, the Social Security Administration, the Federal Reserve Board, and the Michigan Survey Research Center. The value of the information thus provided has been enhanced by the release in this volume of some of the results of the 1949 IRS audit study. Our small income distribution fraternity can now go to work not only to evaluate the existing data but also to improve those to be obtained in the future.

I shall confine my remarks to the paper prepared by Herman Miller and Leon R. Paley, which compares the 1949 incomes reported by a representative sample of the population to the Census Bureau and to the IRS.

SUMMARY OF THE MILLER–PALEY FINDINGS

At the risk of condensing too much, I believe that the findings of the study by Miller and Paley may be fairly summarized as follows:

1. There is a marked similarity between the distributions of the matched families as ranked by the income reported to the Census

Bureau and to the IRS, the medians of the IRS being $57 higher (a little more than 1.6 per cent). For unrelated individuals, however, the census median is $90 higher (about 5 per cent).[1] These differences are based on incomes as reported to the two agencies, without correction for differences in income concept.

2. To eliminate the effect of conceptual differences, the sample cases reporting nontaxable income in the census were removed. This increased the medians for both distributions, widened the gap between the medians of families, and narrowed the gap between the medians of unrelated individuals. For families with taxable income only,[2] the IRS median is higher by $100 (2.8 per cent) than the corresponding median based on census reports; the census median for unrelated individuals with taxable income exceeds the corresponding IRS median by only $79 (4.1 per cent).

3. Although the distributions of all families from the census and IRS reports are almost identical, there are marked differences in the distributions for the farm (census median 16 per cent higher) and nonfarm (IRS median 4 per cent higher) sectors and for some of the occupations within the nonfarm sector. Among nonfarm families, the IRS gives higher medians for those headed by wage earners and salaried, professional, and managerial workers, while the census yields higher medians for the self-employed businessmen and professional workers. Miller and Paley note correctly that these results are consistent with what we know about tax returns from the IRS Audit Control Programs for the years 1948 and 1949. The ACP found inaccuracies in reporting income on tax returns by self-employed persons (farm and nonfarm), who tend to overstate deductions and forget receipts, in contrast to the fairly accurate reports of income by wage earners, subject to tax by withholding.[3]

4. One of their most interesting tables (Table 11) distributes the matched families by census incomes and cross-classifies them by IRS incomes. As might be expected, the diagonal cells in this table are the most densely populated, indicating that numerous families report "approximately" equal incomes to the two agencies. But the families in the diagonal cells are by no means the majority

[1] These and subsequent percentage comparisons between census and IRS medians are computed by dividing the absolute difference by the lower of the two figures.

[2] To avoid confusion, I am using the term "taxable" income as it is used by Miller and Paley—i.e., as the income (exclusive of net capital gains) individuals or married couples include in their tax returns, before allowances for deductions and exemptions. The tax law actually defines taxable income as adjusted gross income less deductions and less personal exemptions.

[3] See Marius Farioletti, "Some Results of the First Year's Audit Control Program of the Bureau of Internal Revenue," *National Tax Journal*, March 1952, pp. 65–78, and his paper in this volume, Tables 1–6.

(about 45 per cent below $5,000), and the class intervals in the table are wide enough to mask differences in the two incomes for families in the same cells. This table would have been more reveal-ing if the diagonal cells had been split between those families who reported higher incomes to the census and those who reported higher incomes to the IRS. However, it is quite obvious from the table as it stands that the relatively small net difference previously noted between the medians for families covers up sizeable offsetting differences in reporting.

UNDERREPORTING BY THE CENSUS

The most important facts in the Miller and Paley study are sig-nificant for income distribution work, but the conclusions they seem to draw from these facts are questionable. It is hard to pin down their reasoning, but I believe it proceeds somewhat as follows:

A distribution of randomly selected families ranked by census incomes looks a good deal like a distribution of the same families ranked by their IRS incomes (admitting significant differences in the reports by some segments of the population—particularly for self-employed farm and nonfarm groups). However, the results of this matching study, like others, indicate that "the variations in re-sponse elicited in repeated interviews appear to be random and *do not introduce any systematic bias into the income distribution*" (italics added). Thus the authors point out that "on this basis" the match between the census and IRS data indicates that "the census in-come data are quite accurate." But they realize immediately that they have overstated their case for the accuracy of census data, be-cause such a conclusion rests entirely on "the validity of the assump-tion that tax returns are accurate." Referring to Mrs. Goldsmith's finding that total income calculated from income tax returns is under-stated compared to national income estimates of total income— which the authors consider "most nearly correct"—they reluctantly conclude that "the underlying assumption that tax returns can be used as a validation check for census reports is questionable."

It is surprising indeed that Miller and Paley suggest that there is no systematic bias in the census distribution, and that they are so reluctant to admit that unaudited tax returns cannot be used to check the census data. Conclusive evidence on this point is given in Farioletti's paper on the IRS Audit Control Program. It is well to recall that the estimates of reporting errors described in this study were based on personal interviews conducted by trained internal revenue agents, that the income concept they used was unambigu-ous (except in the rare instances where even the income tax law

can be variously interpreted), that these agents had the force of law behind them, and that they had in their possession at the time of the interviews as many of the information returns as could be found for each individual in the files of the IRS. Farioletti warns us that, despite these important advantages, the IRS was not able to disclose all the errors in its sample returns. Accordingly, it is not a matter of conjecture that unaudited tax returns understate incomes—we know this to be a fact. Since the census median for families is lower than the IRS median based on unaudited tax returns, it follows that the census distribution understates total family income.

It is abundantly clear that the aggregate income covered by the census distributions falls considerably short of the correct total—$42 billion, or 25 per cent, short of the NID aggregate for 1949.[4] While the national income estimates may not be accurate to the nearest billion dollars, few will challenge Mrs. Goldsmith's statement that "no serious student of the national income statistics would suggest that the aggregate money income embodied in the OBE personal income series could be overstated by anything like these orders of magnitude." [5] The evidence of underreporting by individuals in field surveys is so clear that the national income estimator has no choice but to correct for this underreporting when he distributes total income by size.

To take one example: the national income estimate of dividends paid is based on data reported on corporation income tax returns, but comparison of these estimates with dividends reported by individuals on their tax returns (corrected for the known differences in coverage) reveals a gap of some 15 to 20 per cent.[6] Thus, two sets of data published by the same agency yield substantially different results. When the national income estimator finds that survey data yield about the same amount of dividends as that disclosed on individual tax returns, he must attempt to correct for underreporting in estimating the distribution of total income by size.

One of the major arguments in favor of census data has been that, in spite of the known underreporting, they are useful in disclosing internal relationships in the structure of the income distribution (differences in incomes among cities of different size, relationships between earnings and income other than earnings, and relative distributions of families headed by persons belonging to different population groups). However, underreporting in the census surveys

[4] See the paper by Selma F. Goldsmith, in this volume, Table 1.
[5] *Ibid.*, p. 73.   [6] *Ibid.*, Table 3.

is not distributed evenly by size classes, by regions, or by size of city. According to Mrs. Goldsmith's estimates census data cover more than 90 per cent of total wages; 89 per cent of nonfarm entrepreneurial income; 78 per cent of farm income; about two-thirds of rental income, social security, and military payments; and only about 23 per cent of interest and dividends.[7] With such wide variations in coverage, many of the relationships shown by the census surveys must be unreliable. To mention one obvious example, the CPS survey for the year 1954 shows almost fifteen million families and unattached individuals with money incomes of less than $2,000.[8] Since farm families make up a large portion of this group, and since there is more understatement of farm than of urban incomes, it is difficult to judge the magnitude and character of the low-income problem.

CORRECTING FOR UNDERREPORTING

Unfortunately, although a careful analyst can approximate the amount of underreporting in a field survey, he is rarely able to distribute that amount by income level. He needs both what Hart and Lieblein called a "statistical bridge" [9] between the field survey data and the tax return data and estimates of underreporting on tax returns. The papers by Miller and Paley and by Farioletti indicate that it is technically feasible to obtain such information, and I urge that a concerted effort be made to do so. There is sufficient time between now and 1960, when the next decennial census is to be taken, to plan for the necessary tabulations.

I suggest that the census materials will be useful for classifying the population into family units, but the incomes they report should be regarded as only first approximations of their actual incomes. A subsample of the census sample might be used to obtain the necessary statistical bridge between census incomes and the incomes reported on tax returns. This subsample should be large enough to provide statistically reliable cross-classifications of family units by income size classes and by other characteristics, such as occupation, type of income receipts, and size of family. A matching study between the census and IRS data, similar to the study reported on by Miller and Paley, would provide the data to fill in the cells in the statistical bridge. Once the bridge is available, all that would be needed is an audit study of tax returns (based on an entirely differ-

---

[7] *Ibid.,* Table 2.   [8] *Ibid.,* Table 4.

[9] Albert Gailord Hart and Julius Lieblein, "Family Income and the Income Tax Base," in Volume Eight (1946) of Studies in Income and Wealth, pp. 235–262.

ent sample in order to avoid disclosing the names of census respondents to the IRS) to estimate the underreporting of census incomes. The audit study should be designed to provide estimates of underreporting by type of income throughout the income scale, in addition to the administrative data needed by the IRS for evaluating and improving tax enforcement techniques.[10] These tabulations would, of course, be supplementary to the regular census tabulations and would not supplant them.

This undertaking would be costly, but not prohibitively so. If necessary, the number of annual surveys could be reduced, especially since available evidence suggests that the year-to-year changes in the relative distribution of income are minor. Accurate data on income distributions once every two or five years would be more valuable than incomplete data on annual distributions.

[10] The expenditure of funds for an audit study could be justified on the ground that the data would be useful not only for income size distribution work, but also for administrative purposes by the IRS. A follow-up of the 1948–1950 audit studies once every five years is a minimum requirement for effective income tax enforcement.