Volume Title: Modeling the Distribution and Intergenerational Transmission of Wealth

Volume Author/Editor: James D. Smith, ed.

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-76454-0

Volume URL: http://www.nber.org/books/smit80-1

Publication Date: 1980

Chapter Title: Estimates of the 1969 Size Distribution of Household Wealth in the U.S. from a Synthetic Data Base

Chapter Author: Edward N. Wolff

Chapter URL: http://www.nber.org/chapters/c7448

Chapter pages in book: (p. 223 - 272)

6 Estimates of the 1969 Size Distribution of Household Wealth in the U.S. from a Synthetic Data Base

Edward N. Wolff

This paper presents a description, as well as some new estimates, of the size distribution of household wealth for the United States in 1969, from a synthetic data base called MESP. This data base was developed at the National Bureau of Economic Research as part of a project called Measurement of Economic and Social Performance (MESP), under the direction of Richard Ruggles, from October 1972 to October 1977.[1] The data base is the product of three statistical matches and two sets of imputations and contains asset and liability information, as well as detailed demographic data, for a sample of 63,457 households.

Some justification may be required for developing a new (and synthetic) data base for estimating household wealth distributions. There are four major sources of household wealth data. The first consists of administrative records, in particular tax returns required of wealthholders for paying wealth taxes. Unfortunately (or fortunately), the U.S. has not imposed a wealth tax, and such a data source is not available in the U.S. However, Sweden and several other Western European countries do have a general wealth tax and this data source. This type of data is probably the best for wealth distribution analysis. Even so, there are three major problems in using it. First, there is usually a minimum level of wealth required for filing the return; thus the coverage of the popu-

Edward N. Wolff is associate professor in the Department of Economics, New York University.

The original version of this paper was presented at the 1977 Conference on Research in Income and Wealth in Williamsburg, Virginia. This is a revised version and, as such, embodies many of the valuable suggestions of Mr. Vito Natrella, the paper's discussant. Most of the credit for the development of this data base must go to Nancy and Richard Ruggles, who originally conceived the idea and directed its execution. The research for this paper was funded by NSF Grant No. SOC74–21391.

lation is incomplete. Second, not all assets are normally included in this type of tax return (particularly, consumer durables), and for those that are, there are usually problems of underreporting (both from ignorance of current market value and for tax reasons). Thus, the coverage of asset values is normally deficient. Third, there are quite often disclosure problems in releasing this type of data for research use.

The second major source of wealth data consists of estate tax records. These, too, are administrative records, but unlike wealth tax records they cover decedents, not the living. In the United States the use of estate tax data as a means of making wealth distribution estimates has been largely developed by James Smith (see Smith 1974 and Smith and Franklin 1974 for a description of the methodology used). There are five main problems associated with this source of data. First, the sample is limited to the top of the wealth distribution (decedents with gross estates of $60,000 or more in 1969). Second, asset coverage is also limited, with consumer durables and household inventories omitted. In addition, there is a tendency for assets, particularly business equity, to be undervalued for tax reasons. Third, very limited demographic detail is available on the decedent and none on his family. Fourth, developing full population estimates from the sample of decedents depends on assumptions about relative mortality rates (though the overall size distribution estimates are fairly robust with respect to different assumptions). Fifth, there is almost no way of determining the effect of *inter vivos* transfers (gifts before death) and the establishment of trust funds on the size distribution of wealth estimated from this data source.

The third major source of wealth data comes from direct surveys of households. This might come from a full census or from a sample survey. Perhaps the most well known example of this type in the United States is the Federal Reserve Board's 1962–63 Survey of Financial Characteristics of Consumers (see Projector and Weiss 1966). As in all surveys, deficiencies arise because of the limited time and budget allocated to complete them. For this survey, 2,557 consumer units were given questionnaires to report their assets and liabilities, as well as other household information. The asset coverage is fairly complete, except for consumer durables. The main problem with this survey is the severe underreporting of liquid assets and installment debt (Projector and Weiss 1966, p. 61). For example, in comparison with Flow of Funds data, only 51 percent of savings accounts, 55 percent of U.S. Government securities, 39 percent of state and local government securities, and 58 percent of installment debt were reported in the survey. Another problem with this survey is that due to its relatively small sample size, wealth distribution estimates for subgroups of the population, particularly the poor and the rich, are not very reliable.

The fourth major source of wealth data is income flows. Essentially, the technique involves "capitalizing" interest, dividends, business profits, and the like into corresponding asset values. An early example of such a set of estimates for the U.S. is contained in Stewart (1939). A more recent set of estimates is provided in Lebergott (1976). To date, the technique has been used on aggregate income flow data. MESP, in effect, uses the same technique on a micro–data base. There are both advantages and disadvantages to this technique (See Friedman 1939). First, the resulting asset estimates are only as good as the income flow estimates. In Stewart, Lebergott, and the MESP data base, the underlying income flows come from Internal Revenue Service tax returns. This is probably the most accurate source of income information in the U.S., particularly for nonwage income. Moreover, the income data contained in the tax returns are probably far more reliable than survey wealth data. A second advantage is that the resulting wealth imputations automatically balance with the national totals, because the capitalizing ratio is the ratio of the national total for a given asset to the sample total of the corresponding income flow. A possible disadvantage is that the resulting wealth estimates are sensitive to the yield ratios used. In Stewart, Lebergott, and the MESP data base, it was implicitly assumed that the yield on each asset was the same for each income class, race, region of the country, and the like. If there were a systematic relation between yield and some demographic characteristics (for example, higher income classes may receive a higher dividend yield on stock equity), then a bias would be introduced into the wealth imputations. But the advantages outweigh the disadvantages. Since a capitalization procedure is not "tied" to a particular survey or set of administrative records, it can be applied to any sample frame. Thus, as in the MESP data base, full coverage of the population is possible. Also, the technique is a relatively open one, so that assets not normally covered by this approach, like consumer durables, can be added to household portfolios. This approach thus makes possible full coverage of assets and liabilities.

The bulk of this paper will present a description of the techniques used in the construction of the MESP data base (section 6.1). Section 6.2 will present some new estimates of wealth holdings for different social and economic classes in 1969. Section 6.3 will present some concluding remarks as well as cautions in the use of this data base. It should be noted that no attempt has been made here to compare wealth estimates from the MESP data base with those from other sources. This is done in Wolff (1978). However, it might be noted that the results of the comparison are encouraging and indicate reasonably close sets of estimates. Most of the discrepancies that exist can be traced to differences in concept or sampling frame.

6.1 The Formation of the MESP Data Base

The MESP data base was formed by combining information from the 1970 Census Public Use sample with the Internal Revenue Service tax return data and by imputing asset and liability values based on income flows and other available household information. The sample frame of the MESP data base is the 1970 state 15 percent Census 1/1000 Public Use Sample (PUS), which contains personal and household information for a randomly drawn sample of 63,457 households. Statistical matching procedures were used to add household information from three other data sets: the 1970 Internal Revenue Service Tax Model (IRS 70), the 1969 Internal Revenue Service Tax Model (IRS 69), and the 1970 state 5 percent Census 1/1000 Public Use Sample (PUS5). Asset and liability information was then imputed to each household based on its extended set of demographic and income data. Household asset and liability estimates were then adjusted to align with national balance sheet totals of household wealth.

6.1.1 The Statistical Matches

A statistical matching procedure developed by Nancy and Richard Ruggles (see Ruggles and Ruggles 1974; Ruggles, Ruggles and Wolff 1977) was used to combine information from the two census and the two tax return files. In all, three separate matches were performed (See appendix). The first match was between the 1969 and 1970 IRS files. This was done because a special 1970 IRS file had been developed by the Social Security Administration containing the race and age of the head of household on each tax return,[2] as well as more detailed information on the deductions taken in each tax return, particularly mortgage and other interest payments and state, local, sales, and property tax payments, than the 1969 IRS file. For the match, the two files were first divided into single and joint returns. The single filers were then divided into four cohort groups: males under 65, males 65 or over, females under 65, and females 65 or over. The joint filers were also divided into four cohorts: both under 65, both 65 or over, husband under 65 and wife 65 or over, and husband 65 or over and wife under 65. Each of these groups was then subdivided again, depending on the number of children in the family. Tax returns within each of these finely divided groups were then matched between the IRS69 and IRS70 file, depending on how close the two records were with respect to the following thirteen items: adjusted gross income (AGI); wage and salary earnings/AGI; interest income/AGI; long-term capital gains/AGI; rental income/AGI; dividends/AGI; farm income/AGI; trust income/AGI; royalty income/AGI; business and professional earnings/AGI; pension income/AGI; property sale gains/AGI; and total deductions/AGI.

Race, age, and itemized deductions were then transferred from the IRS-70 record to the corresponding IRS69 record.

The second and major match was between this "augmented" IRS69 file and the 1970 PUS file, containing income and earnings information for the year 1969. The purpose of this match was to combine the detailed income information of the IRS69 file with the detailed demographic information of the PUS. Moreover, the PUS contains information on the value of owner occupied housing as well as stocks of durables held. Both sets of information were thus required to construct household balance sheets.

The two files were first divided into cohort groups on the basis of the following four (common) variables: marital status (single vs. married); sex (for singles); age of head of household; and race of head of household. Within each cohort group the two files were matched depending on how close the two records were with respect to the following six characteristics: number of children; homeowner vs. renter; wage and salary earnings; business earnings; farm income; total income. The detailed income information, as well as data on itemized deductions, was then transferred from the IRS69 file to the 1970 PUS file.

The last match was that of the PUS5 file to the PUS. The reason for this match was that only the PUS5 file has information on the televisions, radios, and clothes washers and dryers owned by each household. The two files were first divided into cohorts on the basis of the following five variables: marital status; age of head of household; sex of head of household; race of head of household; homeowner vs. renter. Records from the two files were matched depending on how close they were with respect to the following five characteristics: number of children; value of property or gross monthly rental; wage earnings of head of household; wage earnings of spouse; total family income. Information on the stocks of consumer durables was then transferred from the PUS5 file to the PUS.[3]

6.1.2 Alignment of Income Flows

Since tax returns were imputed to households in the PUS, some error was expected in the total income flows computed from this sample. This is documented in table 6.1, which compares the MESP totals with those of the IRS *Statistics of Income*. The adjusted gross income (AGI) and wage and salary totals were quite close. The interest, dividend, business and professional net income, and rental income totals were all higher in the unadjusted MESP file than in the IRS totals. The main reason for the discrepency is evident from the second column of table 6.1: the matching procedure assigned too many tax returns containing these income items to households in the PUS sample frame.[4]

Table 6.1 **Comparison of 1969 Income Flows between the Unadjusted MESP Totals and Statistics of Income**

Item	National Totals (billions of dollars)			Percent Receiving the Item	
	MESP	IRS	MESP/IRS	MESP	IRS
Adjusted gross income	$629.6	$603.6	1.04		
Wages and salary	573.1	499.0	1.15		
Interest	44.5	19.6	2.27	65.1%	42.3%
Dividends	38.4	16.9	2.27	30.7	16.0
Business and professional net income	42.6	30.4	1.40	32.7	8.0
Partnership net income	−17.2	2.0		5.7	2.7
Farm net income	−10.0	3.6		16.5	4.1
Rental income	4.0	2.6	1.54	20.6	8.4
Estates and trust income	− 1.2	1.4		1.9	0.8

Source: U.S. Internal Revenue Service 1971.

Our fix-up procedure was straightforward. In the case of interest, dividends, business and professional net income, and rental income, we randomly eliminated these entries so that the percent of households receiving each item in the MESP file would be equal to the IRS percent.[5] We then adjusted the remaining income entries by a constant multiple so that they would sum to the IRS total. In the case of partnership, farm, and trust income, where the signs for the totals differed, we used a somewhat different procedure. We randomly eliminated a certain percent of positive entries and a certain (though different) percent of negative entries, so that the percent receiving the income item and the total income flow would equal the IRS total.[6]

6.1.3 Asset and Liability Imputations

Our next step was to "build up" balance sheet information for each household based on the stock and flow data already contained in the (now adjusted) MESP data base. The imputation procedures differed for different assets and liabilities. However, in all the procedures the resultant stock totals were aligned with the national balance sheet totals for the household sector (see table 6.2).

Owner-occupied Housing

House values were provided in the PUS, though they were coded in 11 intervals. The midpoints of each interval were used, except for the last, open-ended interval of $50,000 or more. For this we chose a value of $77,538 so that the total would agree with the aggregate balance sheet.

Consumer Durables

Ownership, though not values, was provided for the following set of durables in the PUS: number of automobiles (0, 1, 2, 3 or more), air conditioning unit, washing machine and clothes dryer, dishwasher, home food freezer, television, and radio. To construct a balance sheet for each household, it was necessary to increase the coverage of durables and to impute a *dollar* value for each durable owned by the household.

Estimates of the total value of consumer durables held by households were obtained from the Bureau of Economic Analysis (see table 6.3). Moreover, from the 1960–61 Bureau of Labor Statistics Consumer Expenditure Survey (CES), information was provided about the annual expenditure by families on each of the following durables for 1960–61: automobiles, washer/dryer combinations, refrigerators, other major ap-

Table 6.2 **Aggregate National Balance Sheet of Household Wealth for the U.S., 1969, by Item (billions of current dollars)**

Item	Value	
Assets	3,612.8	
Tangible Assets	1,220.3	
Owner-occupied housing		635.0
Other real estate		175.8
Automobiles		89.5
Other consumer durables		227.3
Inventories		92.7
Financial Assets	2,392.6	
Demand deposits and currency		104.9
Time and savings deposits		381.4
Federal securities		101.4
State and local governments securities		34.8
Corporate and foreign bonds, mortgages, open market paper, other instruments		85.6
Corporate stock		635.9
Farm business equity		218.1
Unincorporated nonfarm equity		314.5
Trust fund equity		132.8
Insurance and pension reserves		383.1
Liabilities	450.2	
Mortgage debt		276.6
Consumer credit		121.1
Other debt		52.5
Net Worth	3,162.6	

Source: Estimates prepared by Raymond Goldsmith in Ruggles (1977). Consumer durables were split into autos and others from Bureau of Economic Analysis worksheets provided by John Musgrave.

Table 6.3 **Net Stocks of Consumer Durables Held by Households in 1969 (billions of current dollars)**

Item	Value
Automobiles	$ 89.5
Other motor vehicles	9.5
Appliances	30.8
Radios, televisions, phonographs, etc.	30.9
Furniture	52.3
Textiles and other durable home furnishings (excluding china and utensils)	68.4
Other (including china and utensils, jewelry, books, and toys)	35.4
Total	$316.8

Sources: Bureau of Economic Analysis Worksheets, provided by Mr. John Musgrave. (See Young and Musgrave, [1976] for methods.)

pliances, small appliances, televisions, radios and phonographs, furniture, textiles, floor coverings, and housewares. For the imputation of consumer durable values, it was necessary to combine the information contained in these three sources of data. This was done in three successive steps. First, ownership of durables not included in the PUS inventory was imputed to households, and the purchase price and year of purchase of each durable were estimated for each household. Second, the current market value (as of 1969) of each durable was estimated by depreciating the purchase price of the durable according to its age and the life span of the durable. Third, the total value of durables held by households in the sample was aligned to the BEA net stock estimates.

Imputing the Ownership of Durables. Using the CES we computed the percent of households falling within predefined demographic categories who purchased each of eleven durables during the survey year. We initially used nine demographic characteristics. The categories of each of these, as well as the marginal percent who purchased each durable are shown in table 6.4. Using the nine-dimensional breakdown would have resulted in 43,336 ($2 \times 4 \times 2 \times 3 \times 3 \times 2 \times 7 \times 3 \times 7$) categories—far in excess of the 13,728 families in the CES. We therefore chose the three most important demographic characteristics out of the nine—income, age, and urban/rural residence—and added a fourth homeowner/renter category.[7] This resulted in a small enough number of cells to obtain reliable estimates of the proportion who purchased each durable by demographic group.[8]

We treated the proportion of each group purchasing each durable in 1960–61 as the *probability* of each group's purchasing the good in cal-

Table 6.4 **Percent of Families of the Indicated Type Purchasing Each of Eleven Durables in the CES Survey Year**

	Textiles	Furniture	Floor Coverings	Refrigerator	Washing Machine	All Other Major Appliances	Small Appliances	Housewares	Automobile	Television	Radio
Urbanization											
Urban	75.9	45.1	29.9	8.2	9.3	20.3	26.8	68.3	23.2	54.3	46.5
Rural	77.0	39.8	31.9	10.9	11.0	20.3	26.6	74.6	27.0	52.4	37.5
Region											
Northeast	76.4	43.8	32.4	7.2	8.8	18.4	25.4	66.8	23.7	55.1	45.9
North Central	77.1	45.0	31.7	9.1	10.8	21.3	26.7	69.5	24.9	56.0	45.6
South	74.8	41.0	30.2	10.1	9.0	19.4	27.8	74.3	23.5	50.2	36.8
West	77.3	44.4	26.0	9.3	11.0	23.0	26.8	69.3	26.2	54.2	50.0
Sex of head											
Male	79.4	46.6	32.5	9.9	10.9	22.4	28.5	73.2	27.3	56.5	47.0
Female	60.4	27.1	20.6	4.8	4.1	9.8	17.8	55.5	9.5	39.6	27.2
Race of head											
White	77.1	44.2	30.8	9.1	10.1	21.2	26.8	70.2	25.4	55.1	44.6
Negro	70.2	38.2	30.4	8.2	6.7	12.1	28.2	71.2	15.0	42.9	35.3
Other	69.4	32.2	18.2	9.7	10.1	21.7	16.7	70.9	24.0	46.5	48.4
Education of head											
Less than H.S. grad	72.9	37.2	30.1	8.8	9.1	17.5	25.1	67.9	22.3	50.5	34.6
H.S. grad	80.7	48.8	32.1	9.6	10.8	23.0	29.1	72.8	28.2	59.0	52.3
Some college	82.2	55.3	30.7	9.3	10.8	25.4	29.1	74.6	26.6	57.8	61.0
Marital status											
Married	82.4	49.1	34.2	10.3	11.6	23.7	29.7	76.2	28.5	58.5	48.7
Not married	56.3	24.8	18.6	5.0	4.1	9.1	17.2	51.1	11.1	38.1	27.2

Table 6.4—*continued*

	Textiles	Furniture	Floor Coverings	Refrigerator	Washing Machine	All Other Major Appliances	Small Appliances	Housewares	Automobile	Television	Radio
Age of head											
Under 25 years	78.0	59.5	34.8	12.6	18.8	27.2	27.7	79.7	35.8	48.6	46.2
25–34 years	83.2	57.1	35.2	12.0	14.9	27.5	30.0	80.9	32.7	57.4	53.1
35–44 "	85.7	51.0	35.3	10.2	11.5	23.4	32.0	78.6	29.9	60.2	60.0
45–54 "	79.9	44.1	32.5	8.9	9.0	20.1	29.5	72.5	27.0	57.2	50.9
55–64 "	72.5	34.9	26.1	7.8	6.7	16.2	23.9	64.3	18.9	49.7	31.1
65–74 "	60.5	26.4	23.4	5.3	4.5	12.7	18.7	53.5	10.2	47.2	19.4
75 years & over	48.4	17.4	14.6	4.1	3.7	9.3	11.2	43.3	5.4	34.1	12.8
Occupational status											
White collar (empl)	83.3	53.2	32.6	8.9	10.8	24.1	29.3	74.9	27.6	58.7	59.6
Blue collar (empl)	79.4	46.6	33.7	10.4	11.3	21.7	29.2	73.9	28.8	56.2	45.2
Retired	56.1	24.5	19.9	5.3	4.1	11.2	15.8	49.4	8.8	42.8	17.6
Income of head											
Under $1,000	40.6	12.0	12.2	4.6	3.8	8.2	8.4	45.2	4.8	19.5	14.3
$1,000–2,999	58.6	24.5	20.9	6.5	5.3	9.8	17.1	58.0	11.1	39.0	18.3
$3,000–4,999	75.4	39.5	29.6	9.1	10.1	18.0	26.7	71.1	23.3	52.7	35.4
$5,000–6,999	81.5	49.3	33.5	9.9	11.8	22.2	28.0	73.3	27.2	58.3	48.3
$7,000–9,999	86.9	55.1	37.2	10.5	11.5	26.4	32.0	76.7	31.4	63.0	59.6
$10,000–14,999	88.7	58.3	36.0	10.6	12.7	29.2	35.0	79.3	35.1	64.5	69.2
$15,000 & above	89.9	59.7	36.5	9.0	9.0	32.3	38.2	78.8	37.3	66.5	71.5

Source: 1960–61 Bureau of Labor Statistics Consumer Expenditure Survey.

endar year 1960 and all successive years.[9] We let q_{ij} be the probability of demographic group j's purchasing durable i. We obtained information on L_i, the service life of each durable i (table 6.5). Thus, the probability r_{ij} that a person in group j owns durable i is $q_{ij}L_i$, under the assumption that no one in a group purchases a durable until its service life is over.[10] Probability r_{ij} was then computed for each household in the PUS on the basis of its demographic characteristics and for all durables except cars, television, radios, and washer/dryer units.[11] A number s_i between 0 and 1 was randomly picked (using a random number generator) from a uniform distribution for each household and each durable i. If $s_i < r_{ij}$, we assigned ownership of durable i to the household; otherwise no ownership was imputed. The age A_i of durable i was also imputed to households for all durables owned by the household (including those in the PUS inventory). Let $T_{ij} = 1/q_{ij}$. T_{ij} then indicates the average length of ownership of durable i for demographic group j, where, if $T_{ij} > L_i$, the good has zero value in the $T_{ij} - L_i$ years of possession. The age A_i of good i is then given for each household by $A_i = s_iT_{ij} = s_i/q_{ij}$ (as long as $s_i < r_{ij}$).[12]

Table 6.6 compares the percent of households which purchased each of the eleven durables in 1960–61 according to the CES and our estimates of the percent which purchased each durable in 1969 (that is, those durables whose age is less than or equal to 1.0). The imputed purchase estimates were quite close to the actual CES figures for all durables except autos and televisions. Automobile and television purchases in the imputation for calendar year 1969 were probably overstated because of the occurrence of multiple ownership of each item in the PUS.[13]

Imputing the Current Market Value of Durables. The purchase price of each durable owned by households was imputed using regression analysis. Using the CES, we regressed family expenditure on each of the eleven durables, conditional on purchasing the durable, on the following set of variables common to the CES and PUS[14]: family income; years of schooling of head of household; age of head of household; family size; urban/rural/farm residence; region; sex of head of household; race of head of household; marital status; industry of employment of head of household; occupation of head of household; and homeowner/renter.

The regression results were used to impute a purchase price to all households in the PUS owning durables, as follows: for each household owning durable i, we computed $p_i = xb_i$, where b_i are the regression coefficients for durable i and x the set of regressors. The estimate p_i is the mean purchase price (in 1961 dollars) of durable i for households with characteristics x. The variance was added back in by setting p_i, the purchase price of durable i, equal to $p_i \pm t\sigma$ where σ is the estimated standard error for the regression and t is a standard normal variate whose

Table 6.5 **Service Life (in years) of Each of Eleven Consumer Durables**

Item	Service Life (years)
Automobiles	10
Televisions	9
Radios	9
Housewares	11
Small appliances	11
Textiles	10
Furniture	14
Floor coverings	10
Refrigerators	10
Washing machines/dryers	10
Other major appliances	10

Source: Young and Musgrave 1976, table 1, p. 10.

Table 6.6 **Comparison of the Percent of Households Purchasing Durables and the Average Purchase Price between the 1960–61 CES and the Imputed Value for the 1970 PUS**

	% of Households		Average Purchase Price (1961 dollars)		
Item	1960–61 CES	1969 PUS (imputed)	CES	PUS	% Diff.
Textiles	76%	70%	$ 44	$ 70	59%
Furniture	44	44	173	274	58
Floor coverings	30	25	87	183	110
Refrigerators	9	10	240	275	15
Washer/dryer units	10	9	193	207	7
Other major appliances	20	23	135	168	24
Small appliances	27	25	28	37	32
Housewares	70	56	19	41	116
Automobiles	24	40	1,234	1,561	26
Televisions	54	74	71	90	27
Radios & phonographs	44	41	76	148	95

Note: Sample sizes CES 13,728; PUS 63,457.

value was obtained from a standard normal random number generator.[15] Table 6.6 shows the mean purchase price of each of the eleven durables in the CES and the mean (imputed) purchase price for the same durables in the PUS. The PUS mean purchase prices are uniformly higher. This is to be expected, since the PUS imputations use 1969 incomes. (In fact, the mean income in current dollars is about 50 percent higher in 1969 than in 1961.)

To obtain the current market value V_i of durable i, we assumed a straight line depreciation schedule and computed V_i as follows:

$$V_i = (L_i - A_i)\ p_i/L_i$$

where A_i is the imputed age of durable i. In the case of autos and televisions, this valuation was done for each one owned by the household.

Reconciling the Estimates with the BEA Net Stock Totals. The final step was to reconcile our valuation of consumer durables with the aggregate BEA totals of household owned stocks (table 6.3). There are two major sources of error in our estimates. First, the estimates are still in 1961 dollars, though adjusted for 1969 incomes. Second, purchase decisions and expenditure behavior may have changed between 1961 and 1969.

To balance our estimates of the stock of household durables, we applied "adjustment factors," shown in table 6.7, so that the stock of durables in the PUS sample would sum to the BEA totals. Automobile and major appliances required minor adjustment. Furniture and home furnishings required a large adjustment, presumably because the CES coverage of these groups was considerably smaller than the BEA coverage. The television category required a large adjustment, probably because of the introduction of color televisions during the 1960s.

The MESP coverage of consumer durables included all BEA categories except the "other durable" group (china, utensils, jewelry, books, toys, etc.). The PUS coverage thus amounted to $248.4 billion, or 78 percent of the BEA total.

Time Deposits, Bonds (excluding state and local government securities), Notes and Other Interest-Earning Securities

Capitalization techniques were used for the valuation of the remaining assets in the household balance sheet. Ideally, information providing differential yields by demographic and income characteristics of households for different asset types would have been desirable. Thus, for example if we knew that high income households had an average yield of

Table 6.7 **Adjustment Factors for the Alignment of Consumer Durable Totals in the PUS with the BEA Totals**

BEA Group	Adjustment Factor
Automobiles	0.99
Appliances (washer/dryer units, refrigerators, other major appliances, small appliances)	1.37
Televisions, radios, and phonographs	2.49
Furniture	4.04
Home furnishings (textiles, floor coverings, housewares)	2.80

8.0 percent on bonds, and low income households an average yield of 6.0 percent, different capitalization ratios could be provided for low and high income households. Such information, however, was not available except for stock equity (see below). We therefore provided uniform capitalization ratios for each of the remaining assets in the portfolio.

In the case of financial securities, interest on time and savings deposits is not distinguished from that on bonds, notes, mortgages, and other financial securities in the tax return. Time and savings deposits were therefore aggregated with the other financial securities to form one category. Moreover, state and local government bonds were excluded, since interest received from these bonds is nontaxable and, as a result, not recorded in the tax return. The average yield on this group of securities for 1969 was 3.4% (19.6/568.4), which was used to capitalize the interest into stock estimates. There are two offsetting biases in this procedure. First, the fact that savings accounts normally have lower interest rates than bonds and other securities implies that our imputation procedure is overstating the asset values of bond holders relative to those with savings accounts. We are therefore overestimating the financial security holdings of the upper income classes relative to the lower ones. Second, the fact that state and local government bonds have been excluded implies that the financial security holdings of their owners, who are primarily upper income, are being understated.[16]

Corporate Securities

Dividends received from corporate equities are recorded in the IRS tax return data. The average yield was 2.7% (16.9/635.9), which we used to capitalize dividends into corporate stock estimates. In the case of this asset, some information was available on the relation of dividend yield to household income for 1969 (Blume, et al., 1974, p. 26). Dividend yields were found to vary inversely with income. However, average dividend yields by AGI class varied only from 2.78 to 2.51 percent.[17] This range was so small compared with the likely error in the imputation that we ignored this correction.

Investment Real Estate Holdings

Net rental income is reported in the IRS tax return data. A simple capitalization procedure was not possible here, since some of the income reported was negative.[18] In general, gross rents and costs[19] rise with the value of the property. Thus, the greater the discrepancy between gross rents and costs, the higher, in general, the value of the property. We therefore capitalized net rental income into real estate value proportional to the *absolute value* of net rental income. The average "yield" figure was 7.5% (13.2/175.8).

Unincorporated Nonfarm Equity

Net business and professional (including partnership) income is reported in the IRS tax return data. Like net rental income, both positive and negative entries occur. We therefore used the same procedure as for real estate holdings, and capitalized the absolute value of net income into unincorporated nonfarm equity, using an average "yield" figure of 18.7% (58.8/314.5).

Farm Equity

We used the same procedure as above to capitalize the absolute value of farm net income into farm equity value. The average "yield" figure was 4.8% (10.5/218.1).

Mortgage Debt

Considerably more information was available for the imputation of home mortgage debt. In the Public Use Sample, both home value and length of time of ownership ("When Moved In") were provided for each household. From other sources, we obtained information on average interest rates for home mortgages, average maturity of home mortgages, and a price index for residential housing (see table 6.8). Assum-

Table 6.8 Basic Data for Mortgage Debt Imputation

Period	Average Interest Rate on Home Mortgages[a]	Price Index for Residential Structures[b] (1970 = 100)	Average Maturity[c] (months)
1946–49	4.34%	60.8	231
1950–59	4.81	76.9	261
1960–64	5.69	80.9	318
1965–66	5.93	83.5	329
1967	6.56	87.7	334
1968	7.19	91.9	338
1969	8.26	100.0	338

Sources: [a]U.S. Department of Commerce, Bureau of Economic Analysis, *Business Conditions Digest* (February 1976), table C.118, p. 109 (FHA mortgages).

[b]U.S. Department of Commerce, Bureau of Economic Analysis, *National Income and Product Accounts of the United States, 1929–74*, table 7–13, pp. 294–95.

[c]For average maturity, we used a weighted average of FHA and conventional mortgages. Prior to 1964, the source is Guttentag and Beck, *New Series on Home Mortgage Yields*, NBER, 1970 (#92 General Series), tables C–2 and C–3. After 1963, the source is Department of Housing and Urban Development, *Housing and Urban Development Trends: Annual Summary* (May 1970), table A–61.

ing an average down payment of 25 percent and using standard mortgage amortization tables, we computed the outstanding home mortgage for each homeowner based on initial house value (current value multiplied by the price index) and time of ownership. Our initial estimates resulted in a total household mortgage debt of 273.8, compared with the balance sheet total of 276.6. We then adjusted our estimates by 1.0 percent (276.6/273.8).

Other Household Debt

Interest payments for households itemizing their deductions are recorded in the IRS tax return data. In the MESP file, 40.9 percent of all households recorded some interest payment. The Survey of Financial Characteristics of Consumers reported that 56.0 percent of all households in 1962 had some form of debt other than mortgage debt. We assumed that 56.0 percent of all households in 1969 had some consumer debt, and that the remainder (56.0% − 40.9% = 15.1%) were households that did not itemize their deductions. We randomly selected this remaining 15.1% from households that did not itemize deductions and capitalized the resulting interest flows into household debt, using an average interest yield of 7.3% (12.6/173.6).

Asset Coverage

Table 6.9 gives a summary of household information contained in the MESP data base. A comparison with the aggregate balance sheet in table 6.2 reveals the extent of our coverage. Owner occupied housing, other real estate, and automobiles are fully covered. 70% (158.9/227.3) of other consumer durables are included in the MESP data base but there is no coverage of inventories. The MESP coverage of tangible assets thus amounts to 87% (1059.2/1220.3). Coverage of financial assets is also incomplete. Time and savings deposits, federal securities, bonds, mortgages, and other securities, corporate stock, farm business equity, and unincorporated nonfarm equity are fully covered. However, demand deposits and currency, state and local government securities, trust fund equity, and insurance and pension reserves are not included. The coverage of financial assets amounts to 73% (1736.9/2392.6), and that of total assets equals 77% (2796.1/3612.8). Liabilities are fully covered in the data base.

6.2 Estimates of the Size Distribution of Household Wealth

This section presents some new estimates of the distribution of household wealth in the United States in 1969 from the MESP data base. Our basic definitions and concepts follow Goldsmith (as reported in Ruggles

1977). For illustrative purposes we have divided the household portfolio into five categories: owner occupied home (primary home only); automobiles and other consumer durables (excluding the Bureau of Economic Analysis "miscellaneous" category); financial securities, including time and savings deposits, federal securities, corporate and foreign bonds, mortgages, open market paper, other instruments (excluding state and local government bonds), and corporate stock; farm business equity, unincorporated nonfarm equity, and investment real estate (in-

Table 6.9 Summary of Household Information in the MESP Data Base

Demographic Information
Family and household size and composition
Location of household
Age, sex, race, education of each member

Labor Force Information
Employment status of each member
Industry and occupation of employment
Time worked for each member

Income Information
Wage and salary earnings
Self-employment earnings (including partnership and unincorporated business income)
Farm income
Social security income
Pension income
Welfare and public assistance transfers
Royalties
Interest
Dividends
Capital gains
Rental income
Trust income

Balance Sheet Information
Tangible assets
- owner occupied housing
- other real estate
- automobiles
- other consumer durables

Financial assets
- time and savings deposits, bonds (except state and local government), and other securities
- corporate stock
- farm business equity
- unincorporated nonfarm equity

Liabilities
- mortgage debt
- other household debt

cluding second homes); and debt, including mortgage debt, consumer debt, and other personal loans, but excluding debt secured by stock, investment real estate, or business equity. Total assets are the sum of the first four categories. Net worth is equal to total assets less debt (last category).

6.2.1 Estimates for the U.S. Population

The full sample consists of 63,457 households and is representative of the U.S. population as a whole for 1969 (table 6.10). In 1969 mean assets per household were estimated at $44,000 and mean net worth at $40,000. The concentration of ownership, as measured by the Gini coefficient, varied predictably by type of asset.[20] The Gini coefficient for consumer durables was quite low, at 0.30. That for owner occupied housing was 0.68. Financial securities were highly concentrated with a Gini of 0.91. Business equity was the most concentrated of all, as indicated by a Gini coefficient of 0.94. The distribution of total assets was more unequal than that of consumer durables but less unequal than that of financial securities or of business equity; its Gini coefficient was 0.69. The distribution of net worth was more unequal than that of total assets, indicating an overall negative correlation between assets and debt.

6.2.2 The Composition of Wealth by Demographic Group

The interesting differences are found when we disaggregate the sample by demographic group. Table 6.11 shows the composition of total assets for selected demographic groups, as well as debt, as a percentage of total assets. For the population as a whole, 22.7 percent of assets owned

Table 6.10 **1969 Summary Statistics for the Full Sample**

Number of households	63,457
Mean asset value per household (current $)	$44,029
Mean net worth per household (current $)	$39,926
Gini coefficients	
Own home	0.68
Consumer durables	0.30
Financial securities[a]	0.91
Business equity[b]	0.94
Total assets	0.69
Net worth	0.81

Notes: [a]This category includes time and savings deposits, stocks, bonds, government securities, mortgages, and other financial securities.

[b]This category includes both farm and nonfarm business equity and investment real estate.

Table 6.11 **Composition of Wealth by Demographic Group, 1969**

	Number of Households	Own Home	Consumer Durables	Financial Securities[a]	Business Equity[b]	Debt
AGI						
Negative	1,852	7.5%	2.6%	34.1%	55.7%	28.4%
0–$4,999	17,583	25.4	10.5	42.1	22.1	18.4
5,000–9,999	18,377	28.5	14.6	32.7	24.4	23.4
10,000–14,999	14,349	38.3	14.8	29.5	17.6	21.6
15,000–19,999	5,957	33.9	11.8	34.4	19.9	18.9
20,000–24,999	2,346	27.8	9.1	40.1	23.0	14.4
25,000–29,999	1,094	22.7	7.1	41.7	28.5	11.7
30,000–39,999	850	20.2	6.1	43.2	30.5	11.1
40,000–49,999	328	18.0	4.5	42.8	34.7	9.4
50,000–59,999	218	16.7	3.6	48.1	31.6	8.4
60,000–69,999	111	10.6	2.2	38.8	48.4	5.8
70,000–79,999	65	3.7	0.8	52.1	43.5	2.4
80,000–89,999	36	6.2	1.5	57.9	34.4	4.2
90,000–99,999	38	7.0	1.6	47.3	44.1	4.7
100,000 or more	253	1.1	0.3	79.4	19.1	1.2
Age of household						
24 years or less	4,673	12.9	21.5	40.1	25.5	10.4
25–34	11,620	30.9	15.3	32.7	21.1	22.2
35–44	11,788	36.7	12.3	29.0	22.1	22.4
45–54	12,159	27.1	9.5	41.0	22.6	14.1
55–64	10,806	17.9	6.5	49.3	26.4	9.4
65 or more	12,411	12.9	4.5	52.3	30.3	17.9
Race						
White	56,600	23.7	9.0	41.9	25.4	16.4
Nonwhite	6,857	12.8	7.9	54.7	24.7	13.2

Table 6.11—*continued*

	Number of Households	Own Home	Consumer Durables	Financial Securities[a]	Business Equity[b]	Debt
Schooling of household head						
0–8 years	17,068	13.2	6.5	49.2	31.3	15.7
9–11	11,675	19.4	8.9	45.0	26.8	13.2
12	18,125	26.1	10.5	41.7	21.8	15.8
13–15	7,927	30.3	10.8	37.8	21.1	17.5
16 or more	8,662	31.0	9.1	37.4	22.6	18.9
Region						
Northeast	12,505	24.6	9.1	42.7	23.8	17.0
North-Central	10,563	20.6	8.8	45.4	25.3	14.1
South	22,257	21.5	8.7	44.4	25.5	16.0
West	18,132	24.4	9.1	40.5	26.1	16.9
Occupation of household head						
Professional and managerial	13,902	29.5	9.0	38.1	23.6	17.8
Clerical and sales	9,501	26.6	9.9	41.5	22.2	15.9
Craft	10,665	26.6	10.9	39.4	23.2	17.8
Operative	10,218	23.8	12.1	40.8	23.3	17.8
Service and unskilled	10,970	13.4	7.8	49.9	29.0	12.6
Not reported or not employed	8,201	11.6	4.6	52.4	31.5	13.8
Industry of employment of head						
Agriculture	2,743	6.6	5.7	55.4	32.3	9.7
Mining	648	20.4	8.7	53.4	17.5	18.6
Construction	4,729	22.5	9.2	43.5	24.8	16.4
Manufacturing	15,626	30.4	11.8	37.9	20.0	20.0
Transportation	4,460	28.8	11.4	35.2	24.6	17.9
Trade	9,737	23.5	8.9	46.9	20.8	15.3
Finance, insurance and real estate	2,467	29.7	9.4	37.9	23.0	18.2
Services	11,342	21.9	8.3	39.8	30.1	14.1

Table 6.11—*continued*

	Number of Households	Own Home	Consumer Durables	Financial Securities[a]	Business Equity[b]	Debt
Public administration	3,504	34.2	12.5	31.9	21.4	20.4
Not reported or not employed	8,201	11.6	4.6	52.4	31.5	13.8
Household composition						
Single, no children	14,824	11.8	6.1	60.0	22.2	5.9
Single, with children	3,811	12.6	5.8	58.4	23.2	7.7
Married, no children	19,041	19.5	7.9	44.2	28.5	18.5
Married, 1 child	8,240	30.9	12.8	32.4	23.9	17.9
Married, 2 children	8,047	37.3	12.6	26.8	23.4	23.6
Married, 3 or more children	9,494	36.9	11.6	26.3	25.4	23.0
Net worth class						
0–$4,999	17,161	14.4	67.4	11.5	6.8	78.1
5,000–9,999	9,671	49.4	36.2	11.2	3.2	33.5
10,000–14,999	8,441	63.3	20.8	10.8	5.2	36.9
15,000–19,999	5,838	62.7	18.0	13.3	6.1	32.6
20,000–24,999	4,023	55.5	15.9	19.3	9.3	25.9
25,000–29,999	2,797	51.0	13.4	24.4	11.2	22.2
30,000–39,999	3,634	40.8	10.7	28.9	19.6	18.3
40,000–49,999	2,420	31.7	8.5	36.9	23.0	15.0
50,000–59,999	1,745	23.1	6.9	43.3	26.6	11.2
60,000–69,999	1,272	22.6	6.0	47.6	23.8	9.7
70,000–79,999	1,063	20.2	5.2	46.7	27.9	9.3
80,000–89,999	743	18.8	4.8	41.0	35.5	7.9
90,000–99,999	615	14.6	4.3	47.7	33.4	7.8
100,000–199,999	2,640	11.3	3.3	49.8	35.7	5.1
200,000 or more	1,394	2.7	0.8	62.4	34.1	1.4
Full sample	63,457	22.7	8.9	43.0	25.3	16.1

Note: The table shows the value of each asset (or debt) as a percent of the total assets held by the group.
[a]This category includes time and saving deposits, stocks, bonds, government securities, mortgages and other financial securities.
[b]This category includes both farm and nonfarm business equity and investment real estate.

by households was in the form of owner occupied housing, 8.9 percent in the form of durables, 43.0 percent in the form of financial assets, and 25.3 percent in the form of business equity. Moreover, the average debt-to-asset ratio was 16.1 percent. When we disaggregate the population by income class, we find that the share of housing and durables in total assets rose with income to about $15,000 and then fell continuously with income, while the share of financial assets and business equity generally rose with income. Moreover, debt as a fraction of assets rose with income until $10,000 and then declined almost continuously with income level.

The major difference in asset structure between whites and nonwhites was that the share of assets in owner occupied housing for whites was almost twice that for blacks. The debt-asset ratios were about the same. The percent of assets in home and durables rose with schooling level, as did the percent of debt. The asset and debt structure was very similar by region of the country and among occupational groups, except for (low-paid) service and unskilled workers. There was some variation by industry of employment. (Those in agriculture, for example, had a predictably large share of their assets in business equity and a low share in homeownership.) Singles with and without children and those married with no children had a large share of their assets in financial securities, a low share in housing and durables, and a small debt-to-asset ratio. The converse was true for married couples with children.

There was considerable variation in wealthholdings among wealth classes. The share of assets in homeownership increased with wealth through the first four wealth classes and then declined, while that in durables declined almost continuously with net worth. The share in financial securities rose almost continuously with wealth, while that in business equity increased through the first twelve wealth classes and then leveled off. The debt-to-asset ratio declined almost continuously with net worth.

6.2.3 Mean Wealth by Demographic Group

Table 6.12 shows the mean value of total assets and net worth for different groups in 1969. Household wealth rose consistently with household income, except between the first and second income classes and between the thirteenth and fourteenth income classes.[21] Moreover, net worth tended to rise considerably faster with income than income itself, particularly above $50,000 of income. Assets rose with age until age 65 and then leveled off, whereas net worth rose until age 65 and then fell by 9 percent.[22] The biggest increase in net worth occurred between the 35–44 and the 45–54 age groups. The 55–64 age group had mean assets 4.0 times as great as the youngest age group and a mean net worth 4.0 times as great. Mean assets were 20 percent greater for whites than nonwhites, and mean net worth 16 percent greater. There was relatively

Table 6.12 **Mean Household Assets and Net Worth by Demographic Group, 1969 (in current $1,000)**

	Number of Households	Assets	Net Worth
AGI			
Negative	1,852	102.5	73.4
0–$4,999	17,583	22.7	18.6
5,000–9,999	18,377	23.3	17.9
10,000–14,999	14,349	31.6	24.8
15,000–19,999	5,957	49.4	40.1
20,000–24,999	2,346	73.0	62.5
25,000–29,999	1,094	100.7	88.9
30,000–39,999	850	125.4	111.5
40,000–49,999	328	181.2	164.3
50,000–59,999	218	232.0	212.4
60,000–69,999	111	372.4	350.8
70,000–79,999	65	696.7	680.3
80,000–89,999	36	486.3	465.9
90,000–99,999	38	383.3	365.3
100,000 or more	253	1,644.0	1,624.2
Age of household head			
24 years or less	4,673	15.0	13.4
25–34	11,620	26.8	20.8
35–44	11,788	36.7	28.5
45–54	12,159	49.5	42.6
55–64	10,806	59.4	53.8
65 or more	12,411	59.6	48.9
Race			
White	56,600	44.8	37.5
Nonwhite	6,857	37.4	32.4
Schooling of household head			
0–8 years	17,068	45.4	38.3
9–11	11,675	40.6	35.3
12	18,125	38.9	32.8
13–15	7,927	41.9	34.6
16 or more	8,662	58.8	47.7
Region			
Northeast	12,505	43.9	36.4
North-Central	10,563	43.0	37.0
South	22,257	44.0	37.0
West	18,132	44.8	37.3
Occupation of household head			
Professional and managerial	13,902	58.4	48.0
Clerical and sales	9,501	41.3	34.7
Craft	10,665	39.3	32.3
Operative	10,218	30.6	25.1
Service and unskilled	10,970	40.3	35.2
Not reported or not employed	88,201	51.3	44.3

Table 6.12—*continued*

	Number of Households	Assets	Net Worth
Industry of employment of head			
Agriculture	2,743	59.3	53.5
Mining	648	46.6	38.0
Construction	4,729	45.3	37.8
Manufacturing	15,626	36.6	29.3
Transportation	4,460	36.9	30.3
Trade	9,737	46.0	38.9
Finance, insurance, and real estate	2,467	50.3	41.1
Services	11,342	47.4	40.7
Public administration	3,504	35.4	28.2
Not reported or not employed	8,201	51.3	44.3
Household composition			
Single, no children	14,824	41.7	39.3
Single, with children	3,811	45.8	42.3
Married, no children	19,041	54.4	44.3
Married, 1 child	8,240	36.8	30.2
Married, 2 children	8,047	37.8	28.9
Married, 3 or more children	9,494	38.1	29.3
Full sample	63,457	44.0	36.9

little variation in wealth by educational level, except for college graduates, who were considerably richer than other groups.

There was almost no variation of mean wealth by region. Professional and managerial workers were the wealthiest occupational group, while operatives were the poorest. There was considerable variation of wealth by industry of employment, with workers in agriculture by far the wealthiest. Married couples without children and singles with and without children were considerably wealthier than married couples with children. Married couples without children had considerably more assets (and debt) than singles. From this rather cursory analysis it would seem that age and income are the most important determinants of wealth. Household composition was less important than these two factors but more important than the remaining ones.

6.2.4 Inequality of Wealth within Demographic Groups

Our final table (table 6.13) shows the level of inequality in the size distribution of assets and net worth among households *within* each of the indicated groups.[23] The Gini coefficients for the full sample were 0.69 for assets and 0.81 for net worth. Thus, except for the first income class, the level of wealth inequality was lower within income class than for the whole population. Moreover, the level of wealth inequality generally declined over the first four income classes and then remained

Table 6.13 **Gini Coefficients of the Size Distribution of Assets and Net Worth by Demographic Group, 1969**

	Number of Households	Assets	Net Worth
AGI			
Negative	1,852	0.87	0.93
0–$4,999	17,583	0.60	0.73
5,000–9,999	18,377	0.61	0.80
10,000–14,999	14,349	0.51	0.58
15,000–19,999	5,957	0.49	0.58
20,000–24,999	2,346	0.47	0.54
25,000–29,999	1,094	0.47	0.53
30,000–39,999	850	0.48	0.53
40,000–49,999	328	0.48	0.52
50,000–59,999	218	0.55	0.60
60,000–69,999	111	0.49	0.52
70,000–79,999	65	0.42	0.44
80,000–89,999	36	0.46	0.51
90,000–99,999	38	0.47	0.49
100,000 or more	253	0.46	0.47
Age of household head			
24 years or less	4,673	0.69	0.70
25–34	11,620	0.65	0.71
35–44	11,788	0.59	0.66
45–54	12,159	0.64	0.70
55–64	10,806	0.73	0.80
65 or more	12,411	0.72	0.94
Race			
White	56,600	0.68	0.80
Nonwhite	6,857	0.77	0.89
Schooling of household head			
0–8 years	17,068	0.74	0.94
9–11	11,675	0.70	0.81
12	18,125	0.66	0.75
13–15	7,927	0.64	0.71
16 or more	8,662	0.67	0.75
Region			
Northeast	12,505	0.67	0.81
North-Central	10,563	0.71	0.81
South	22,257	0.70	0.83
West	18,132	0.68	0.80
Occupation of household head			
Professional and managerial	13,902	0.66	0.75
Clerical and sales	9,501	0.66	0.75
Craft	10,665	0.66	0.79
Operative	10,218	0.66	0.81
Services and unskilled	10,970	0.74	0.87
Not reported or not employed	8,201	0.73	0.91

Table 6.13—*continued*

	Number of Households	Assets	Net Worth
Industry of employment of head			
Agriculture	8,743	0.78	0.90
Mining	648	0.75	0.94
Construction	4,729	0.71	0.84
Manufacturing	15,626	0.64	0.77
Transportation	4,460	0.65	0.75
Trade	9,737	0.70	0.79
Finance, insurance, and real estate	2,467	0.65	0.74
Services	11,342	0.71	0.79
Public administration	3,504	0.58	0.68
Not reported or not employed	8,201	0.73	0.91
Household composition			
Single, no children	14,824	0.74	0.77
Single, with children	3,811	0.72	0.75
Married, no children	19,041	0.71	0.94
Married, 1 child	8,240	0.64	0.71
Married, 2 children	8,047	0.61	0.70
Married, 3 or more children	9,494	0.61	0.70
Full sample	63,457	0.69	0.81

stable across the rest of the income ladder. The (unweighted) average level of wealth inequality within income class was 0.52 for assets and 0.58 for net worth, both surprisingly large.

The inequality of ownership of assets declined with age until age 45 and then increased with age, while inequality in net worth remained relatively constant until age 55 and then increased. The inequality in net worth was extremely high for those over 65. Wealth inequality was higher for nonwhites than for whites (or for the overall sample). Wealth inequality was somewhat higher for the less educated than for the more educated. The level of wealth inequality showed little variation by region of the country and was close to the overall level in all regions. Wealth inequality was somewhat greater for service and unskilled workers than for other occupational groups. Wealth inequality was greater among married couples with no children and singles than among married couples with children. In general, except within income classes, the level of wealth inequality was at approximately the same level within these demographic groups as in the whole population.

6.4 Conclusions and Cautions

The MESP data base, we believe, provides a valuable new resource tool for the analysis of wealth distribution in the United States. In par-

ticular, the vast array of demographic information made available by it will make possible work focusing on the wealth behavior of small subgroups of the U.S. population. More detailed work on size distributions, the composition of wealth, life cycle accumulation patterns, and simulation models can be undertaken with this new data base.

A word of caution should be noted in the use of this data base, even though general tests of its reliability have proved positive (see Wolff 1978). As with any new data base, there are certain problems and limitations in its use. Some can be corrected for or overcome with additional work and some cannot. In any synthetic data base created through statistical matching techniques, certain *conditional* joint distributions are not reliable. In this case, the joint distributions of noncommon variables in the PUS file and the IRS file conditional on a common variable cannot be used for estimation purposes, because this is the information that is lacking (and the rationale for performing the match). For example, the covariance of education (a PUS variable) and stock equity (an IRS variable) conditional on income (a common variable) will not be reliable in the MESP data base. However, the overall (unconditional) covariance of education and stock equity can be reliably estimated (see Ruggles, Ruggles and Wolff [1977] for more details).

Other deficiencies involve the estimation of household assets and liabilities. These estimates might be improved with additional work. With regard to *Owner occupied housing,* currently, house values are recorded in eleven interval codes; some attempt might be made to "smooth out" the distribution using a random number generator. The estimation procedure for consumer durables might be redone using the recently available 1972–73 Consumer Expenditure Survey; full coverage of durables might also be possible. It would also be desirable to add stocks of semidurables to the household portfolio using the new Consumer Expenditure Survey. The category of currency and demand deposits should also be added to the household portfolio (see Wolff [1978] for one attempt). With regard to financial securities, some attempt might be made to split time and savings deposits from bonds, mortgages, and other financial instruments, since the two groups of assets are currently aggregated into one category; it would also be desirable to add a separate imputation for nontaxed state and local government bonds, though appropriate data may be difficult to locate. Trust fund equity is not currently included in the household portfolio, and it would, of course, be desirable to include this, since much of the wealth of the rich is held in this form. One possible source for this imputation is the entry "trust fund income," which is currently in the IRS tax return data. Before this can be undertaken, the problem of whom to assign the assets of a trust to—whether the current beneficiary, the remainderman, or possibly the trustee—must be resolved. Pensions, too, should be added to the household portfolio,

but here again important conceptual issues must first be resolved. For example, should only vested pensions be assigned to households? Should only redeemable pensions be imputed? How should one handle partially funded pensions? Should Social Security be included in pensions? Finally, the assignment of the cash surrender value of life insurance policies to households poses less serious conceptual problems than that of pensions. Here, the problem of obtaining pertinent data makes this imputation very rough, if not impossible.

Despite its limitations and deficiencies, the MESP data base is still the most complete in coverage of both households and assets of any now currently available. Moreover, unlike survey or administrative data sets, the MESP data base, thanks to its methodology, *allows* continual modification, improvement, and expansion of asset and liability estimates and coverage. Future use, it is hoped, will result in its gradual improvement as a research tool.

Appendix: A Technical Description of the MESP Matches

The Sort-Merge Matching Procedure

Six steps are involved in the sort-merge matching procedure we used in the creation of the MESP data base (see Ruggles and Ruggles 1974 and Ruggles and Wolff 1977). The first step is to select which of the two files is to be used as the sampling frame; the second data set, the "B File," is then matched onto the first data set, the "A File." The next step is to select the unit of the match; in the case of household data, the unit could be the household, the family, the individual, or some other composite. Information is then transferred from the B File to the A File on a unit by unit basis.

The variables in each of the two data sets are then divided into four kinds. The first are the "cohort" variables; the A and B samples are first divided into cohorts and the matches then performed within each cohort. The second are the X or "matching" variables; the values of the X variables are partitioned into intervals and the two files matched on the interval values of these variables. The third are the Y variables, used to construct the intervals of the X variables. The fourth are the remaining variables.

Matching intervals for each X variable are then constructed by running cross-tabulations of Y and X and parsing the X variable such that the conditional (frequency) distribution of Y on X is constant within intervals and different between intervals for each Y variable according

to a predetermined statistical criterion. Y and X are thus conditionally independent within intervals and significantly related across intervals. By varying the statistical criterion, we can generate different sets of matching intervals at different levels of statistical confidence.

Frequently, an X variable will differ in concept or sampling distribution between the A and B files because of differences in definition or differences in sampling frame. Before the match is executed, the X variable in the A file is adjusted or aligned to the corresponding X variable in the B file to reconcile the differences between the two files.

Finally, the two files are each sorted into cohorts and within cohort by the matching intervals of the X variables. Matches are first made at the highest level of statistical confidence. Records that fail to match at this level are then matched at successively lower levels of statistical confidence. This results in a distribution of matches by matching level, which is calibrated. For reasons of optimization, if the distribution is nonuniform over matching levels, new confidence levels are selected, new matching intervals computed, and the sort-merge match redone. A number of iterations may be required bfeore the distribution of matches is approximately uniform.

The Construction of the MESP Data Base

The Internal Revenue Service Tax File 1970–1969 Match

The first match that was executed involved the 1969 and 1970 Internal Revenue Service Tax Files (IRS69 and IRS70). Both are samples of about 100,000 tax returns, heavily stratified on adjusted gross income.[24] The IRS69 file was used as the sample frame; the main purpose of the match was to transfer race and age information contained in the 1970 file to the 1969 file.

The tax return was used as the basic unit in the match. Joint returns from the IRS70 file were matched with joint returns in the IRS69 file, and IRS70 single returns with IRS69 single returns. The cohort, X, and Y variables used in the match are shown in table 6.A.1. Sex was used as a cohort variable for single returns. Both data sets contained information indicating whether the filer(s) was 65 or over in age or less than 65 in age (since the former resulted in an added exemption). In the case of single returns, there were two categories: 65 or over in age; under 65 in age. In the case of joint returns, there were four categories: both filers 65 or over in age; husband 65 or over, wife under 65; husband under 65, wife 65 or over; both filers under 65 in age. The fourth cohort variable was the number of children, which we divided into four categories: zero, one, two, and three or more.

The first X variable was the level of adjusted gross income (AGI). Because of the change in the size distribution of AGI between 1969 and

Table 6.A.1 Structure of the IRS70–IRS69 Match

Cohort Variables
Type of tax return
Sex
Age (over and under 65)
Number of children

X Variables
Adjusted gross income (AGI)
Wage and salary earnings as a percent of AGI
Interest income as a percent of AGI
Long-term capital gains as a percent of AGI
Rental income as a percent of AGI
Dividends as a percent of AGI
Farm income as a percent of AGI
Royalty income as a percent of AGI
Trust income as a percent of AGI
Business and professional earnings as a percent of AGI
Pension income as a percent of AGI
Property sale gain as a percent of AGI
Total deductions as a percent of AGI
(only for those who itemize deductions)

Y Variables (IRS70)
Race (white or nonwhite)
Age (whites)
Age (nonwhites)

1970, it was necessary to align AGI in the two files before the match was executed (see below). The next eleven X variables represented the major components of AGI. Since these items (particularly wage and salary earnings), as well as total deductions, are highly correlated with AGI, it would be redundant to match on the level of these income items as well as on the AGI level. We therefore matched on each income source and total deduction as a percent of AGI. The Y variables are the age and race of the head of household recorded on the tax returns; these items were used as the Y variables since these are the chief data to be transferred by the match.

Three iterations were necessary for a satisfactory match. Table 6.A.2 shows the final matching levels and the number of matching intervals for each X variable by level. The matching intervals were generated on the IRS70 file. The number of matching intervals fell off sharply between the level of greatest confidence (level 1) and the level of least confidence. As is evident from table 6.A.2, the most important X variable in the match was AGI, since it had consistently the highest number of intervals, except for level 6. Wage and salary earnings as a percent of AGI, dividends as a percent of AGI, business and professional income as a per-

Table 6.A.2 **Number of Matching Intervals by Matching Level in the IRS70–IRS69 Match**

	Matching Level					
	6	5	4	3	2	1
X Variable	Correlation (.50)	Correlation (.95)	Correlation (.97)	Correlation (.99)	Chi-square (.995)	Chi-square (.50)
Adjusted gross income (AGI)	1	13	15	16	21	93
Wages and salary/AGI	2	6	6	8	12	29
Interest/AGI	1	1	1	1	7	9
Long-term capital gains/AGI	1	1	1	5	7	13
Rental income/AGI	1	1	1	1	5	8
Dividends/AGI	1	7	7	8	9	24
Farm income/AGI	1	1	1	1	1	10
Royalty income/AGI	1	1	1	1	1	9
Trust income/AGI	1	1	1	1	4	4
Business and professional income/AGI	1	2	3	6	7	19
Pension income/AGI	1	1	1	1	3	13
Property sale gain/AGI	1	1	1	1	1	4
Total deductions/AGI	1	4	10	10	12	24

cent of AGI, and deductions as a percent of AGI were also important matching variables. The remaining X variables "washed out" at either the second, third, or fourth matching level.

As noted above, because of the general increase in income between 1969 and 1970, the IRS 69 and IRS70 files could not be matched directly on AGI level. Some alignment was required first. This was done on the basis of *percentile* rank. This meant, in effect, that the nth percentile AGI level in the IRS 70 file was treated as equivalent to the nth percentile AGI level in the IRS69 file. The matching intervals were then adjusted accordingly. Thus, if the nth percentile AGI level in the IRS70 file fell into matching interval j, the nth percentile AGI level in the IRS69 file was also mapped into matching interval j.[25]

Table 6.A.3 shows the distribution of matches by matching level in the third and final iteration of the sort-merge matching procedure. There were no nonmatches and no matches at the cohort level. Of the 95,288 tax returns in the IRS70 file, 38,211 (or 40.1%) were used in the match.

The following variables were transferred to the IRS69 file from the matched record in the IRS70 file: race of head of household; age of head of household; mortgage interest paid (only returns with itemized deductions); other interest paid (only for itemized deductions); and state and local taxes (only for itemized deductions).

Table 6.A.3 **Final Calibration of the IRS70–IRS69 Match**

Matching Level	Percent of Matches
1. Chi-square (.50)	13.2%
2. Chi-square (.995)	11.8
3. Correlation (.99)	13.6
4. Correlation (.97)	18.7
5. Correlation (.95)	24.7
6. Correlation (.50)	18.1
7. Cohort	0.0

The Internal Revenue Service Tax File 1969—Public Use Sample 1970 Match

The second and major match was between the IRS69 file, augmented with information from the IRS70 file, and the 1970 Census 1/1000 Public Use Sample (PUS).[26] This PUS file is a random sample of the U.S. population, with a sample size of 63,457 households. Both the IRS69 and the PUS file contain income information for calendar year 1969. The purpose of this match was to augment the income information in the PUS with the more detailed income breakdown in the IRS file.

The PUS file was used as the sample frame, and the IRS69 file was matched to the PUS. In effect, tax returns were imputed to households in the PUS. The reason for this is that the PUS is a representative sample of the U.S. population, while the IRS file is heavily stratified on income. By matching the IRS file to the PUS, we could assure that the tax information would be given its appropriate population weight.

We chose the tax return as the unit of the match. This required the creation of tax units from the information in the PUS. In the PUS, the basic unit is the household, but the household is broken down into family and individual observations. By assuming that all married couples file joint returns and all others file single returns, we constructed tax return units from the individuals in the PUS file.

The cohort, X, and Y variables used in the match are shown in table 6.A.4. Both the race and age variables on the IRS69 file were imputed in the IRS70–IRS69 match. Only two categories were used for race: white and nonwhite.

For the number of children, we used the number who were listed as dependents (exemptions) in the IRS file, and the number under age 18 in the PUS file. Homeowner status is directly indicated in the PUS. In the case of the IRS sample, we used the mortgage interest deduction as a proxy for homeownership.[27] Wage and salary earnings of both spouses were summed in the case of married couples, and the sum matched against the corresponding entry in the IRS file. This was likewise done

Table 6.A.4 Structures of the IRS69–PUS Match

Cohort Variables
Type of tax return
Sex of respondent (single returns)
Race of head of household
Age of head of household
X Variables
Number of children
Owner occupied home or rental unit
Wage and salary earnings
Business earnings
Farm income
Total income
Y Variables (PUS file)
Education
Birthplace
Occupation
Industry of employment
Class of worker
Years married (married couples only)
Number of years at current address
Value of property (homeowners only)
Number of automobiles in household

for business and professional earnings. Because the definition of farm income differs so much between the PUS and the IRS files, we used a (0,1) dummy variable for farm income reported or not reported. For the total (personal) income variable, we started with adjusted gross income (AGI) on the IRS file and total income in the PUS file. The two concepts differ considerably. To reconcile them, we first added dividend exclusions and other adjustments to AGI to obtain personal gross income. The two concepts were still not identical, since gross income in the IRS file excluded Social Security and welfare income but included capital gains, whereas total income in the PUS included Social Security and welfare income but excluded capital gains. We therefore subtracted capital gains from gross income in the IRS file and subtracted social security and welfare income from total income in the PUS file.[28]

Six iterations were necessary for a satisfactory match. Table 6.A.5 shows the final matching levels and the number of matching intervals at each level for each X variable. In this match there was also a sharp fall-off in the number of intervals by matching level. The two most important X variables were wage and salary earnings and total income. Business and farm income both washed out at the second level of the match, number of children at the third level, and homeowner status at the fourth level.

Table 6.A.5 Number of Matching Intervals by Matching Level in the IRS69–PUS Match

	Matching Level					
	6	5	4	3	2	1
X Variable	Correlation (.50)	Correlation (.70)	Correlation (.80)	Correlation (.90)	Correlation (.97)	Chi-square (.99)
Number of children	1	1	1	1	4	8
Homeowner status	1	1	1	2	2	2
Wage earnings	3	4	5	9	20	36
Business earnings	1	1	1	1	1	13
Farm income	1	1	1	1	1	2
Total income	2	2	3	6	16	36

Before the match was executed, both business and professional income and total income required alignment. The fact that the distribution of business and professional income differed in the two files was due to differences in concept and in reporting error. The distribution of total income also differed in the two files, even after the adjustments described above were made. This was probably due to differences in reporting error. The alignment was done on the basis of percentile rank. Selected correspondence points at given percentile ranks are shown in table 6.A.6.[29]

IRS values were consistently higher than the corresponding PUS values in the bottom eight deciles of total income and slightly lower in the top two. This may be due to underreporting by low income recipients in the Public Use Sample. The percent difference between the two files declined steadily through the first eight deciles, and corresponding values were quite close in the top four deciles. For business and professional earnings, the PUS values were consistently higher than the corresponding IRS values. This may be due to the fact that costs are offset against earnings in computing business and professional profit or loss in the tax returns but not as a rule in the census questionnaire. The percent difference increased up through the fourth decile and declined thereafter.

Table 6.A.7 shows the distribution of matches by matching level in the final iteration of the match. There were no nonmatches, and only three percent of the records matched at the cohort level. Of the 89,705 tax returns in the IRS69 file, 15,406 (or 17.2 percent) were used in the match. The low percent of IRS records used is not surprising, since the IRS file is heavily stratified toward the upper income levels.

Table 6.A.8 presents some additional statistics used to evaluate how close the match was with respect to three of the *X* variables. The cor-

Table 6.A.6 **Selected Correspondence Points in the Alignment of Total Income and Business and Professional Income in the IRS69–PUS Match**

Percentile	PUS Value	IRS Value	Percent Difference[a]
		Total Income	
12	$ 750	$ 1,145	52.7%
22	1,550	2,311	49.1
31	2,550	3,495	37.1
42	4,150	5,029	21.2
50	5,750	6,409	11.5
60	7,550	8,013	6.1
71	9,750	9,882	1.4
81	12,150	12,119	− 0.3
90	16,150	15,602	− 3.4
		Business and Professional Income (recipients only)	
12	$ 550	$ 430	−21.8%
21	1,150	814	−29.2
31	2,150	1,444	−32.8
41	3,550	2,242	−36.9
53	5,150	3,655	−29.0
60	6,550	4,763	−27.3
70	8,550	6,619	−22.6
80	11,350	9,732	−14.3
90	18,150	16,322	−10.1

Note: [a]Percent difference is defined as (IRS–PUS/PUS) × 100.

Table 6.A.7 **Final Calibration of the IRS69–PUS Match**

Matching Level	Percent of Matches
1. Chi-square (.99)	16.0%
2. Correlation (.97)	18.8
3. Correlation (.90)	30.6
4. Correlation (.80)	14.3
5. Correlation (.70)	12.2
6. Correlation (.50)	6.2
7. Cohort	3.0

Table 6.A.8 **Measures of Closeness of Fit by Matching Level for Selected X Variables in the IRS69–PUS Match**

Matching Level	Wage and Salary Earnings		Business and Professional Income		Total Income	
	Correlation Coefficient	Percent Difference	Correlation Coefficient	Percent Difference	Correlation Coefficient	Percent Difference
1. Chi sq. (.99)	0.95	0.8%	0.92	7.3%	0.96	1.4%
2. Correl. (.97)	0.94	4.1	0.43	17.9	0.96	2.5
3. Correl. (.90)	0.96	3.1	0.36	40.2	0.96	3.1
4. Correl. (.80)	0.92	9.5	0.23	265.1	0.92	9.2
5. Correl. (.70)	0.89	4.7	0.07	384.9	0.91	8.0
6. Correl. (.50)	0.75	11.8	0.04	423.1	0.74	20.3
7. Cohort	0.41	19.2	0.01	485.8	0.57	54.7
Total file	0.96	3.5%	0.50	72.9%	0.97	4.1%

Notes: The correlation coefficient is defined as the correlation of X_A and X_B for matched records occurring in the specified match level, where subscript A refers to the IRS value and subscript B to the PUS value.

Percent difference is defined as $100 \times (\overline{X}_A - \overline{X}_B)/\overline{X}_B$, where the bar indicates the mean value of the X variable in the specified match level.

relation coefficients measure how close the individual X values in the matching records were by matching level, and the percent differences measure how close the mean values of the X variables were in each of the matching levels. As to be expected, the matches were closer in value for matches at higher levels of statistical confidence than for matches at lower levels. Wage and salary earnings entries in the IRS file were quite close to their corresponding entries in the PUS file in the first five match levels, which accounted for 91 percent of the matches. The correlation coefficient was 0.96 for the entire file and the percent difference was 3.5. The same pattern was recorded for the total income variable. The matches were quite close for the first five matching levels and the overall correlation coefficient was 0.97. The fit for business and professional income was decidedly poorer, with an adequate fit occurring only at the first match level.

The following variables were transferred to the PUS file from the matched record in the IRS 69 file: adjusted gross income; wage and salary earnings; interest income; long-term capital gains; short-term capital gains; rental income; dividends; farm income; royalty income; trust income; business and professional income; pension income; property sale gain; income adjustment; mortgage interest expenditure; and other interest expenditure.

The 1970 5% Public Use Sample–1970 15% Public Use Sample Match

The third match was between the 1970 census 1/1000 5% and 15% Public Use Samples (PUS5 and PUS15). The designations 5 and 15 refer to the percent of the population receiving the respective questionnaires. Approximately 80 to 90 percent of the variables are the same in the two samples. Our interest in matching the two files was to transfer consumer durable information present in the 5% sample but not in the 15% sample to the now augmented 15% sample.

The 5% and 15% samples are identical in structure. Since consumer durable ownership is assigned to the household, we used the household as the unit of the match. The cohort, X, and Y variables used are shown in table 6.A.9. Since there was a wide choice of overlapping variables in the two files, we chose for the cohort and X variables those we felt would be significantly related to consumer durable ownership. The Y variables, which were drawn from the PUS 15% sample, consisted of additional demographic and income information, as well as data on automobile ownership.

Table 6.A.9 Structure of the 1970 Public Use Sample 5%–15% Match

Cohort Variables
Marital status (married vs. single)
Age of head of household
Sex of head of household (if single)
Race of head of household (white vs. nonwhite)
Owner occupied home vs. rental unit
X Variables
Number of children in household
Value of property or gross monthly rental
Wage earnings of head of household
Wage earnings of spouse of head of household (if married)
Total family income
Y Variables (PUS15 File)
Education of head of household
Education of spouse of head of household (if married)
Industry of employment of head of household
Occupation of head of household
Place of birth of head of household
Farm income (yes or no)
Professional income (yes or no)
Social Security income (yes or no)
Welfare income (yes or no)
Place of residence five years ago
Place of work of head of household
Number of automobiles owned by the household

Table 6.A.10 Number of Matching Intervals by Matching Level in the PUS5–PUS15 Match

	Matching Level					
	6	5	4	3	2	1
X Variable	Correlation (.80)	Correlation (.90)	Correlation (.93)	Correlation (.97)	Correlation (.98)	Chi-square (.99)
Number of children	1	1	3	3	3	7
Value of property	2	5	5	7	9	11
Gross rental	1	1	1	1	4	6
Wage earnings (head)	4	6	10	16	20	26
Wage earnings (spouse)	1	1	1	3	7	12
Total family income	3	5	8	12	18	41

Three iterations were necessary for a satisfactory match. Table 6.A.10 shows the final matching levels and the number of matching intervals for each *X* variable at each matching level. The dominant *X* variables in this match were total family income and the wage earnings of the head of household. Property value was also an important *X* variable. The other variables washed out after the first few matching levels.

Table 6.A.11 shows the distribution of matches by matching level in the final iteration of the match. There were no nonmatches, and less than 2 percent of the matches occurred at the cohort level. Moreover, of the 63,490 households in the PUS 5% file, 34,623 (or 55 percent) were used in the match.

Table 6.A.12 presents the correlation coefficients and the percent differences for total family income, wage and salary earnings of the head of household, and wage and salary earnings of the spouse. For total family income, the correlation coefficients are quite high for the first five levels but low for the bottom two. For wage and salary earnings of the household head, the correlations are high at all levels except the cohort

Table 6.A.11 Final Calibration of the PUS5–PUS15 Match

Matching Level	Percent of Matches
1. Chi-square (.99)	18.3%
2. Correlation (.98)	19.4
3. Correlation (.97)	25.8
4. Correlation (.93)	17.6
5. Correlation (.90)	13.5
6. Correlation (.80)	3.9
7. Cohort	1.5

Table 6.A.12 Measures of Closeness of Fit by Matching Level for Selected X Variables in the PUS5–PUS15 Match

Matching Level	Total Family Income		Wage and Salary Earnings			
			Head		Spouse	
	Correlation Coefficient	Percent Difference	Correlation Coefficient	Percent Difference	Correlation Coefficient	Percent Difference
1. Chi-sq. (.99)	0.96	−0.2	0.95	0.5	0.99	− 1.1
2. Correl. (.98)	0.96	0.4	0.98	0.2	0.96	0.7
3. Correl. (.97)	0.95	0.3	0.93	0.2	0.88	2.1
4. Correl. (.93)	0.87	0.2	0.88	−1.4	0.16	− 2.9
5. Correl. (.90)	0.70	0.8	0.76	2.3	0.34	− 3.3
6. Correl. (.80)	0.31	−6.5	0.71	3.5	0.32	12.1
7. Cohort	0.19	−4.6	0.14	17.8	0.26	−10.1

Notes: The correlation coefficient is defined as the correlation of X_A and X_B for matched records occurring at the specified match level, where subscript A refers to the PUS5 value and subscript B to the PUS15 value.

Percent difference is defined as $[(\overline{X}_A - \overline{X}_B)/\overline{X}_B] \times 100$, where the bar indicates the mean value of the X variable at the specified match level.

level. The correlation coefficients for wage and salary earnings of the spouse are high at only the top three levels. The mean values of these three X variables are quite close in the two files at all matching levels except for the cohort level for wage and salary earnings of the head and the bottom two levels for wage and salary earnings of the spouse.

The following variables were transferred to the PUS 15% file from the matched record in the PUS 5% file: washing machines; clothes dryer; dishwasher; home food freezer; television set; radio; and second home ownership.

Notes

1. This was only one of several major data bases developed as part of this project. Others included extended national income, product, and capital accounts (R. Ruggles, N. Ruggles, J. Kendrick, R. Eisner, and R. Goldsmith); a micro–data base for the government sector (J. Quigley); a micro-data base for the enterprise sector (R. Lipsey and M. Gort); and an environmental pollution account (H. Peskin).

2. This information is not normally included in the tax return, except when the filer is 65 years of age or older.

3. Since the overlap in demographic information between the two samples was so substantial, this match provided an ideal opportunity to test the reliability of

the matching technique. To do this, we ran two sets of regressions, the first with variables from the 15% PUS and the second with a mix of variables from the two files. In 90 percent of the cases, the regression coefficients in the two sets were not statistically different (see Ruggles, Ruggles, and Wolff [1977] for more details).

4. We expected some upward bias, since the MESP sample is a sample of *households,* which may file more than one tax return.

5. This procedure probably resulted in a slight downward bias in the percent of households receiving the respective income items (see note 7).

6. We determined the percent of positive entries to keep (p_1) *and* the percent of negative entries to keep (p_2) by solving the following simultaneous system:

$$p_1 P + p_2 N = T$$

$$p_1 q + p_2 r = s$$

where P = total positive income in the MESP file;
N = total negative income in the MESP file;
T = total income from the IRS file;
q = percent receiving positive income in the MESP file;
r = percent receiving negative income in the MESP file;
s = percent receiving the income item in the IRS file.

7. Technically, we might have performed a t-ratio test for the difference in means for choosing the pattern of aggregation for each durable. However, from table 6.4, income, age, and residence seemed by far the predominant determinants.

8. An alternative technique would have been to use logit regression to estimate the probability of purchase of each durable as a function of all nine demographic characteristics. Time and cost constraints prevented us from pursuing this course.

9. This procedure introduces two offsetting biases. First, since real income grows over time, the probability of purchasing for a given household will increase between 1960 and 1969. However, the probability of purchasing a durable declines with the age of the head of household (table 6.4), since stocks of durables tend to be acquired early in the life cycle and then gradually replaced (and perhaps upgraded) as the household ages.

10. This is, of course, a very rough assumption. We could have assumed that the decision to purchase durable i is independent of ownership of i to allow multiple purchases. The distribution of the number of times durable i is purchased in a given span of years would then be given by a binomial distribution.

11. These are the durables already included in the PUS inventory.

12. In the case of automobiles and televisions, where the PUS inventory indicates the household owns more than one, the age of each was estimated.

13. This would overstate the probability of purchasing each in a given year, since the decision to purchase the item is treated as independent of the ownership of that item.

14. Regression results are available on request from the author. Our major findings were: (1) Income is a positive determinant of the amount spent on each durable, while the percent of income spent on durables is negatively related to the income level. (2) The amount spent on durables is positively related to the rate of dissavings, particularly for the more costly durables. (3) Homeowners spend more on durables relative to income than renters. (4) Larger families have smaller expenditures on durables.

15. The only restriction was that if p_i was less than zero, p_i was set equal to zero.

16. We ignore the problem of differences in capital gains for different portfolios in the case of financial securities, as well as stocks. We also ignore the problem of both capital and ordinary gains in the case of the other assets. See Lebergott (1976) for a discussion of this problem.

17. The average yield by AGI class was as follows:

AGI CLASS	AVERAGE DIVIDEND YIELD
Under $5,000	2.77%
5,000–9,999	2.76
10,000–14,999	2.78
15,000–24,999	2.75
25,000–49,999	2.65
50,000–99,999	2.56
100,000 +	2.51

18. There is the additional problem that not all investment real estate is rented. This will result in an overstatement in the concentration of investment real estate ownership, though there is no apparent systematic bias with respect to income or wealth.

19. The costs include such items as utilities, repairs and maintenance, mortgage interest, property taxes, and depreciation.

20. The Gini coefficient measure includes both holders and nonholders. The Gini coefficients were considerably lower for owners alone in most asset groups.

21. Households with negative AGI must be rich enough to own stocks and bonds, which they can sell at a loss, or to own a business that can report a (book) loss.

22. This conforms with the predictions of many life cycle models. See Modigliani and Brumberg (1954), for example.

23. The Gini coefficient, which we use to measure the level of inequality, is defined as twice the area between the Lorenz curve and the 45 degree line of perfect equality.

24. This information was added in a special run by the Social Security Administration, which used the actual Social Security numbers on the sample of tax returns to transfer this information.

25. This method of alignment was deemed superior to a simple inflation of 1969 AGI levels by the average increase in AGI between 1969 and 1970. The reason is that different parts of the AGI distribution shifted by different percents between 1969 and 1970.

26. The particular sample used was the "state 15%" sample.

27. This will somewhat understate the level of homeownership in the IRS file, since some homeowners do not have an outstanding mortgage and some do not itemize their deductions.

28. One additional adjustment was made. Because the income entries in the PUS file were truncated at $50,000, we truncated IRS income entries above $50,000.

29. No alignment was necessary for wage and salary earnings, since their distributions were almost identical in the two files.

Comment Vito Natrella

This paper concerns a new synthetic microdata file containing information on wealthholdings of consumers in the U.S. together with income data and a considerable amount of information on demographic characteristics. Edward Wolff describes the methods used to put the data base together and presents comparisons with other files and estimates.

Wolff indicated that there are three data bases containing information on individual wealth. These are the 1962 Federal Reserve Board Survey of Financial Characteristics of Consumers, the 1967 Survey of Economic Opportunity, and the estimates of personal wealth based on Internal Revenue Service estate tax data. The first two of these are one-time surveys based on samples of 2,500 and 30,000 households, respectively. The third is based on a sample of about 50,000 estate tax returns (Internal Revenue Service 1975). This latter file is available approximately every four years and has been used by the Internal Revenue Service to estimate the wealth of the living. Estimates of wealth have also been prepared by Smith and Franklin based on the same data files using somewhat different multipliers (Smith and Franklin 1974). Estimates based on estate tax data cover the population of top wealthholders—those with assets of $60,000 or more—and in 1969 accounted for almost 50 percent of total wealth. As Wolff mentions, the estimates are created under critical mortality rate assumptions which affect significantly the level of the estimates. However, it should be noted that the various mortality assumptions have considerably less effect on the distributions.

There are two other current estimates of household wealth based on aggregate data rather than microdata sets. One consists of the residual estimates from the annual flow of funds data of the Federal Reserve Board. This, together with the modified version developed by Helen Tice and R. W. Goldsmith, was used by Wolff as the source of control aggregates. The other set of estimates was prepared by Stanley Lebergott (1976) on the basis of aggregate income flows obtained from *Statistics of Income* for 1970. These were used to distribute national wealth estimates also developed from the flow of funds. In effect, both Wolff and Lebergott use flow of funds national balance sheet data as a basis for capitalizing income flows, one on a micro basis, the other as applied to aggregates, in order to develop estimates of the distribution of wealth.

Vito Natrella is director of the Statistics Division of the Internal Revenue Service, Washington, D.C.

These comments are based on the original paper presented at the conference in Williamsburg, December 1977. Many of the suggestions made have been incorporated in the revised version published in this volume.

The data file developed by Wolff and known as the MESP file has certain significant advantages over each of the above. For one thing, it covers the whole distribution of wealth, not just the segment of the population with more than $60,000 in assets. It is aligned with national wealth totals and includes data on holdings of consumer durables. Most important, it contains a wealth of demographic data which can be used for analyses not possible up to this time.

Creation of MESP Data File

The sample frame for the MESP file is the 1970 PUS 15 percent Census 1/1000 file and contains approximately 63,400 households. Since this file consists of information for 1969, it was desired to have a 1969 Tax Model file which included data on itemized deductions not ordinarily included in the odd years and all Social Security demographic data. The construction of such a file starts with the 1970 IRS Tax Model, containing data from about 90,000 returns augmented with Social Security demographic data which is matched on a simulated basis with the 1969 IRS Tax Model so that the demographic information and itemized deduction information can be introduced. The tax model file also contains, of course, all income flows as reported on individual tax returns.

Another simulation match is then made between the modified and augmented 1969 Tax Model and the 1970 PUS 15 percent file. This latter file, which contains all housing, durables, overall income, and basic demographic information, comprises the final sample frame. Estimates of income flows and certain itemized deductions based on the 1970 PUS file are compared with published IRS *Statistics of Income* aggregates for 1969. The PUS file is aligned to the *Statistics of Income* totals by reducing the number of households and reducing the aggregate amounts for the remaining households. In this way, the *Statistics of Income* totals constitute the control figures on income flows.

A further match is made with the 1970 5 percent PUS Census file in order to bring in additional information on durable goods. Values for automobiles and other durables are imputed from data contained in the 1960–61 BLS Consumer Expenditure Survey. The totals are then aligned with Bureau of Economic Analysis aggregates.

The final step in the preparation of the file consists of converting income flows to asset holdings. This is done by capitalizing the income flows appearing in the microfile on the basis of the relation of aggregate income from *Statistics of Income* to aggregate assets held by households as developed in national balance sheets. The tangible assets data on holdings were obtained from estimates prepared by R. W. Goldsmith, while the financial assets came from the flow of funds of the Federal Reserve Board.

The paper presents estimates of the income-size distribution of wealth in the form of percent distribution of portfolio at each size level. Estimates of percent portfolio distribution are also shown for various demographic categories. In addition, data are presented on the mean holdings of total assets and net worth by the various demographic categories and by income levels. An analysis of inequality within the various income levels and demographic groups is presented in a table of Gini coefficients. Comparisons are also made with estimates of personal wealth based on estate tax data using information on the percent of total wealth held by the top 1 percent of the population.

The MESP data file makes a distinct contribution to the body of information available for analyses of wealth and its distribution. It contains more demographic characteristics than any other wealth file. It makes it possible to construct size distributions of various kinds, such as by income or by asset holdings. It shows the composition of wealth and includes holdings of consumer durables which generally have been rather meagerly detailed. The file can also be used in analyses of the life cycle accumulation pattern and in simulation for various purposes.

Problems and Deficiencies

There are a number of troublesome shortcomings in the presentation, some of which can be taken care of easily. One is the absence of dollar figures for the population by the various classifications. The data presented in the tables are in the form of percents or means. Since the dollar figures are easily available, I feel that the estimates should be presented in that form so that it is possible for the reader to make comparisons with other similar estimates.

I also think that more information should be included on what was actually done. For instance, it was difficult to determine what was done to align estimates from the MESP file with aggregate income flows from *Statistics of Income*. On the other hand, a considerable amount of detail is presented on the methods used to arrive at the durable goods imputations. Categories used should also be better defined as to what they include, particularly with respect to the various combinations of assets. Finally, I think it is very important that asset size distributions of the new data be presented as well as income size distributions. At present, personal wealth estimates based on the estate tax data of the IRS can be distributed only by asset size. It should be pointed out that in the 1976 personal wealth estimates by the IRS income size distributions will be presented, since income of the decedents for a prior year is being introduced in the file.

In addition, there are some problems with the data base of a more basic nature. In constructing a national balance sheet to which income

flows are capitalized, data from "Flow of Funds Accounts, 1965–73," published in September 1974, were used for financial assets in spite of the fact that revised figures were published in 1976. Also, the Consumer Expenditure Survey for 1960–61 was used in connection with the estimates of durable goods. Although probably not available at the time the paper was first prepared, the 1972–73 Consumer Expenditure Survey could be used in a revision, improving the estimation procedures considerably.

Besides out-of-date sources, a number of the sources used appeared to be inappropriate to their purpose. The aggregate household balance sheet which was used as the benchmark for asset holdings comes from a mixed source. Financial assets are from the flow of funds of the Federal Reserve Board. These figures include both nonprofit organizations and trusts. Tangible asset data in the balance sheet are obtained from estimates prepared by R. W. Goldsmith and appear in Richard Ruggles's "Statement for the Task Force on Distributive Impacts of Budget and Economic Policies of the House Budget Committee" (1977). In the same statement appear estimates of financial assets developed from the flow of funds data eliminating nonprofit organizations and treating trusts as a separate form of wealth. In addition, the estimates are more up to date than the flow of funds data used by Wolff. It would, therefore, have been more consistent as well as more accurate to use the Goldsmith data for the financial assets as well as for the tangible assets.

Mortgage interest available from the 1970 individual tax model was used as a proxy for home ownership. Using this item, of course, has an inherent understatement since individuals owning homes with no mortgage would be underrepresented. A better approach would be to use the itemized deduction for real estate tax which is also available in the 1970 Tax Model.

In connection with the capitalization ratios used by Wolff, he rightly indicates that it would be desirable to use differential yields according to demographic and income characteristics. Research in this area was reported for corporate stock by Blume, Crockett, and Friend (1974). For 1971, they indicate yields of 2.2 percent for persons with adjusted gross income of $50,000 or more compared with yields of about 2.7 percent for persons with lower income and 2.5 percent in total. In view of these data, Wolff's figure of 2.3 percent does not appear to be as out of line as he indicates.

In developing the estimates, Wolff used only business and professional income both in the tax model files and in the *Statistics of Income* alignment procedures. However, I believe that the PUS file includes partnership income in its total income while equity in partnerships is also included in the aggregate assets shown in the household balance sheet. As

used by Wolff, therefore, the capitalizing process implies that equity in partnerships is distributed in the same way as equity in sole proporietorships. This could have been avoided by including partnership income from the tax model in the matching procedure and also in *Statistics of Income* totals for alignment.

Comparisons with Other Wealth Data

As previously mentioned, comparison with other wealth estimates are rather difficult because of the lack of money amounts and the failure to include a distribution of wealth according to size of assets. However, Wolff makes some comparisons with the personal wealth estimates based on estate tax data using the proportion of total wealth held by the top 1 percent of the population.

The indications are that the estate tax data give estimates of higher concentration than do the MESP data. Estimates from the MESP data show the top 1 percent holding 14 percent of tangible assets, 42 percent of financial assets, and 25 percent of total assets. These compare with 21 percent, 46 percent, and 37 percent, respectively, for wealth estimated from estate tax data.[1] These differences were ascribed by Wolff to the truncated house values used in the MESP file and the inclusion of durables. I feel that they could also reflect a basic distortion in the distribution of asset holdings as shown in the estimates derived from the MESP data. In effect, too little wealth may have been allocated to the upper income brackets.

A comparison which Wolff did not make is with the estimates for 1970 prepared in a basically similar way using, however, aggregate data. The Lebergott estimates show a greater proportion of total held in financial form, assets which persons in the upper income brackets are more disposed to hold. These estimates also indicate substantially higher holdings of total assets in the upper income brackets. Total assets held by persons with incomes over $50,000 were 75 percent higher than those shown for the same group by Wolff. On the other hand, for persons with incomes under $50,000, the holdings were about the same in both sets of estimates. On the basis of income size distributions, the Lebergott estimates indicate considerably greater inequality in wealth than the Wolff estimates. Since the asset size distribution is not available for the Lebergott estimates, it was not possible to compare Gini coefficients.

Wolff indicated a Gini coefficient of .66 for the estimates of total assets developed from the MESP file. For that part of the population subject to the federal estate tax, Smith estimated the Gini coefficient at .50. Extending this to the full population could imply a Gini coefficient close to Wolff's figure. However, in view of the comparisons with other data, I would like to know more about the Wolff computation.

Alignment with SOI Income Flows

The simulated match of the 1970 PUS file with the 1969 modified and augmented IRS Tax Model produced a first approximation of the MESP file with the PUS file as the sample frame. Estimates of population income flows were prepared and then compared with similar data from *Statistics of Income.* The results for adjusted gross income and salaries and wages were reasonably close. However, dividends and interest estimated from the MESP file were more than twice, while business and rental income was about one and one half times, the SOI aggregates.

Wolff ascribes these results to oversampling of high income returns in the IRS file. Alignment to the controlling SOI totals was made by reducing the number of imputed returns and reducing the dollar amounts in the remaining returns. I find adjustments of the magnitudes involved particularly disturbing and feel that the findings of comparatively low concentration could have resulted from this alignment procedure. Forcing by such large amounts substantially weakens the differences in the distribution of those assets more likely to be held by persons in the upper brackets. The result is a data file with a serious deficiency in the upper levels.

I feel that the distortion in favor of the lower brackets may reflect the use of the 1970 PUS file as the sample frame. This file consists of an across-the-board random 1-in-1,000 sample. Such a sample is excellent for estimating demographic characteristics. It is very poor for estimating money amounts which are unequally distributed. The sampling variability of such estimates would be quite high. I suggest that the final sample frame should be one whose sample selection rates are higher the higher the income level.

In conducting the match of the 1969 Tax Model with the 1970 PUS file, Wolff indicated that only 17 percent of the Tax Model file was used. This rather low rate is also an indication of possible undersampling in the PUS file. One solution may be to use the 1969 augmented tax model as the sample frame for tax return filers while the 1970 PUS file could be used as the sample frame for nonfilers. This could be achieved by dividing the PUS file into those required to file and those not required to file, using requirements in effect for 1969. Data from the file of 1970 PUS return filers would be merged into the 1969 augmented tax model. Data for non-filers from the 1970 PUS file would be imputed. Estimates of money aggregates would be prepared using the stratified weights from the tax model for filers and the random 1,000 weight for nonfilers.

This procedure should result in income flow estimates needing only small adjustments to align with SOI. They should have much smaller sampling errors and be distributed more accurately to reflect holdings in the upper income levels.

Conclusion

A data file of the MESP type meets a definite need for wealth data associated with income and demographic characteristics. However, as indicated, improvements are needed. Inconsistencies need to be resolved. More of the basic estimates should be presented so that the file can be better evaluated. Distributions should show actual dollar estimates instead of just ratios. Asset size distributions should be presented as well as income size. More careful use of both definitions and sources should be made in regard to balance sheet aggregates.

In spite of these problems and deficiencies, I feel that this approach has great promise for development of an excellent analytic tool.

Note

1. Based on IRS estimates used by Natrella (1975). The more recent aggregate household balance sheet data developed by Goldsmith were used as the base. These ratios are not too different from those used by Wolff based on the Smith estimates if corrections are made to keep them comparable. Cash and deposits (including time deposits) must be included in the Smith figures, while demand deposits and currency must be included in the aggregate assets.

References

Blume, Marshall; Jean Crockett; and Irwin Friend. 1974. "Stockownership in the United States: Characteristics and Trends." *Survey of Current business* 54, no. 11.

Friedman, Milton. 1939. "Discussion on Income Capitalization." *Studies in Income and Wealth* 3. New York: National Bureau of Economic Research.

Internal Revenue Service. 1975. "Statistics of Income, 1972: Personal Wealth Estimated from Estate Tax Returns." Washington, D.C.: U.S. Government Printing Office.

Lebergott, Stanley. 1976. *The American Economy*. Princeton: Princeton University Press.

Modigliani, Franco, and Richard Brumberg. 1954. "Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data." In K. K. Kurihara, ed., *Post-Keynesian Economics*. New Brunswick, N.J.: Rutgers University Press.

Natrella V. 1975. "Wealth of Top Wealthholders, 1972." Proceedings of the American Statistical Association, Washington, D.C.

Projector, Dorothy, and Gertrude Weiss. 1966. *Survey of Financial Characteristics of Consumers*. Washington, D.C.: Federal Reserve Board.

Ruggles, Richard. 1977. "Statement for the Task Force on Distributive Impacts of Budget and Economic Policies of the House Committee on the Budget." Mimeograph.

Ruggles, Richard, and Nancy Ruggles. 1974. "A Strategy for Matching and Merging Microdatasets." *Annals of Economic and Social Measurement* 3, no. 2.

Ruggles, Richard; Nancy Ruggles; and Edward Wolff. 1977. "Merging Microdata: Rationale, Practice, and Testing." *Annals of Economic and Social Measurement* 6 no. 4.

Smith, James D. 1974. "The Concentration of Personal Wealth in America, 1969." *Review of Income and Wealth,* series 20, no. 2.

Smith, James D., and Stephen Franklin. 1974. "The Concentration of Personal Wealth, 1922–1969." *American Economic Review* 64, no. 2.

Stewart, Charles, 1939. "Income Capitalization as a Method of Estimating the Distribution of Wealth by Size Group." *Studies in Income and Wealth* 3. New York: National Bureau of Economic Research.

U.S. Internal Revenue Service. 1971. *Statistics of Income 1969.* Washington, D.C.: Government Printing Office.

Wolff, Edward. 1978. "The Effect of Alternative Imputation Techniques on Estimates of Household Wealth in the U.S. in 1969." Paper presented at the C.R.E.P.-I.N.S.E.E. International Meeting on Wealth Accumulation and Distribution, Paris.

Young, Allan and John Musgrave. 1976. "Estimation of capital stock in the United States." Paper delivered at the Conference on Research in Income and Wealth, Toronto.

This Page Intentionally Left Blank