ANALYZING THE EFFECTS OF AGE AND HOUSEHOLD SIZE ON THE DISTRIBUTION OF HOUSEHOLD WEALTH IN THE U.S. USING INCOME TAX MICRODATA

Daphne T. Greenwood, University of Colorado at Colorado Springs and
Statistics of Income Division, Internal Revenue Service

1. INTRODUCTION

This paper presents an analysis of personal wealth data generated from a large sample of merged income tax and census records. Wealth data on individuals (or families) must be estimated, since there is no regular mechanism for direct collection of these statistics. Prior estimates of wealth and its distribution have generally relied on estate tax records or personal surveys [1] in contrast to the method used here. The distribution of a nation's wealth--its capital stock, natural resources, land, structures, and liquid assets--is a critical variable on both the demand and supply sides of the economic equation. Wealth, as well as income, is important for measuring the economic position of the worker, consumer, or taxpayer. On the supply side, ownership of assets often determines how they will be utilized in the production of goods or services.

Estimates of Wealth Distribution

Net wealth (assets minus debts) has been estimated for the U.S. population as a whole from a sample of 45,030 families on a microdata file of merged income tax and census returns, using methods outlined in an earlier article [2]. Arraying shares of wealth held by various percentiles of the 1973 U.S. population, it was shown that the upper 1 percent of families held 33 percent of net wealth and the upper 10 percent held roughly 70 percent. The lower 50 percent of families held approximately 1 percent of private household wealth.

Life-Cycle Effects on Wealth Distribution

According to economic theory, variations in wealth holdings are due to differences in income, in savings rates, in rates of return on assets, and in inheritance. An important reason for not expecting complete equality in wealth distribution is that individuals (and thus, households) vary in their position in the life-cycle. Income tends to increase with experience, training, and seniority, so that it has a strong positive correlation with age both in cross-sectional and longitudinal studies. Inheritances tend to be received by persons in their fifties or above. In addition, savings rates vary considerably over the life-cycle as young persons first "save" by purchasing houses and consumer durables, begin to acquire financial assets in mid-life, and subsequently may dissave during retirement years to generate income [3].

Although age has traditionally been used to indicate life-cycle position, individuals (or families) may reach the same point in their life-cycle at different ages [4]. Length of education, military service, cultural norms, and economic resources influence decisions to marry, bear children, and sometimes to divorce. These events in turn will influence savings, indebtedness, and net wealth accumulation. Although age is an indicator of life-cycle position, it is an imperfect one. Household size, indicating the presence of a spouse and/or children in the home, will be used here in part to improve the definition of the position of the household in its "life-cycle."

Criticism of Traditional Inequality Measures

Several years ago, the accuracy of measuring inequality with Lorenz curves and Gini coefficients was challenged. In an article in the American Economic Review Paglin asserted that while our conceptual notions of justice or equity assume higher income and wealth as age increases, and intergenerational inequality resulting from economic growth, the Lorenz and Gini measures are based on measuring deviation from a line of "perfect equality" which would require that all households have equal wealth regardless of the number and age of individuals belonging to the household. Paglin suggested a method for computing the degree of inequality which was due only to age differences and subtracting it from the standard Gini coefficient [5].

Paglin's suggestions quickly generated quite a few criticisms, primarily on the grounds that he had underestimated inequality of income distribution and overestimated its post-war decline [6]. Critics also pointed out that his method of calculating an "age-Gini" to handle life-cycle effects was incorrectly calculated and included some inequality which was not due to life-cycle position. In this paper, I will use an adjusted Gini coefficient which follows a modification to Paglin's approach suggested recently in two different articles, by Formby and Seaks and by Atack and Bateman [7]. I also incorporate differences in household size as well as the age of the head of household in my adjusted measure.

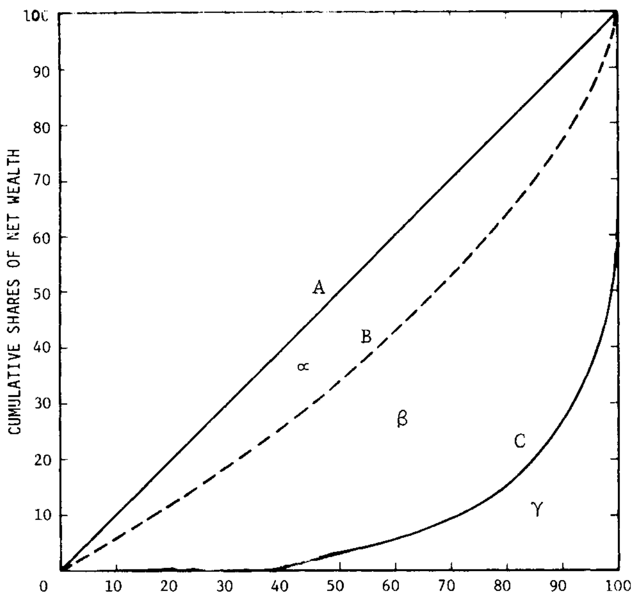
2. SUMMARY OF FINDINGS

In order to explore the effects of life-cycle position on the distribution of net wealth, I have analyzed the relationship between net wealth to age and household size via multiple regression, computed standard Gini coefficients for six different age groups, and used household size as well as age of the head of household to construct an adjusted Gini coefficient from the 1973 microdata file.

The conventional Gini coefficient, based on the area between the empirical Lorenz curve (line C) and the 45 degree line of equality, as in Figure 1 below, was 0.82, whereas the adjusted Gini coefficient with effects of age and household size removed was 0.76. The age-group Gini coefficients ranged from 0.89 in the lowest age group (25 and under) to 0.75 in the 56-65 age group, indicating substantial within-group inequality, which for the most part, decreased with age. The regression relationship produced significant coefficients on both age and household size but an adjusted R^2 of only 0.06. All three calculations support the conclusion that variations in age and household size are <u>not</u> a major source of observed variations in wealth.

Figure 1

ILLUSTRATION OF CONVENTIONAL AND MODIFIED LORENZ CURVES



A - Lorenz's 45° line of equality
B - modified comparison line
C - conventional Lorenz curve

α = area between A and B
β = area between β and C
γ = area below C

3. METHODOLOGY AND RESULTS

Attributable Variation in Regression

A random sample of 2 percent (861 households) was drawn from the file of tax/census microdata and used to estimate by multiple regression the relationship between net wealth, age, and household size. Number of children was added to the number of adults present in the household, not including the head of household since each household has, by definition, at least one member. The age variable represents age in years of the head of household minus 18, the assumed minimum age for heads of household. An interaction term created by multiplying the two together performed better than either of the individual variables.

Various forms of regression were tested. The raw value of net wealth regressed on the independent variables discussed above yielded an R^2 of only 0.02. The following regression (equation 1) relating the natural log of net wealth to age in excess of 18 and household size in excess of 1, plus their interaction term, yielded an R^2 of 0.06 (the highest of any regression) with residuals closer to a normal distribution than in any other equation. T-statistics are in parentheses below each estimated coefficient.

$$(1)\ LNW = \underset{(216.0)}{6.48} - \underset{(76.23)}{.0007\ Age^2} - \underset{(46.06)}{.65\ HH} + \underset{(39.15)}{.04\ HHA}$$

where LNW = natural log of net wealth
Age^2 = (age of household head - 18)^2
HH = household size less one
HHA = HH * (Age of Head - 18)

Age-Specific Gini Coefficients

The full data set was sorted into six age classes based on age of head of household, and standard Gini coefficients were run on each group to assess the extent of within-group inequality. The Gini coefficient measures the proportion of the area below the 45 degree line of perfect equality (line A, in Figure 1) which lies above the Lorenz curve (line C) and is equal to the value of $\frac{\alpha + \beta}{\alpha + \beta + \gamma}$. The more bowed the Lorenz curve, the larger the Gini coefficient, indicating greater inequality in the distribution of wealth. Table 1, below, indicates the boundaries of each group, its mean wealth, share of 1973 population, share of total household wealth, and Gini coefficient. Mean wealth rises with age, as does the share of wealth by age class. Concentration, as measured by the Gini coefficient, is lowest in the middle age groups but quite high for every group.

Table 1
VARIATIONS IN U.S. WEALTH INEQUALITY
BY AGE CLASS, 1973

Head of Household	Mean Wealth	Share of Households (percent)	Share of Net Wealth (percent)	Standard Gini Coefficient
25 and under	$ 9,763	8	2	.89
26 - 35	24,096	20	13	.84
36 - 45	36,454	17	16	.81
46 - 55	43,669	18	21	.78
56 - 65	48,068	17	20	.75
Over 65	50,855	20	27	.84
All House holds	$37,711	100	100	.82

Adjustments to the Gini Coefficient

The conventional Lorenz curve, plotted as line C in Figure 1, indicates the cumulative shares of wealth held by percentiles of the population (see Table 2), beginning with the lowest wealth households. Since roughly 40% of the population holds no measurable net wealth the standard Lorenz curve does not depart from the horizontal axis until this point. The conventional Gini, measuring the entire area between line A, (the 45 degree line of perfect equality,) and line C does not separate out age and size effects on wealth distribution.

The adjusted Gini computed here follows the modification to Paglin suggested by both Atack and Bateman, and Formby and Seaks [7] but expands it further to remove effects of household size as well as age. Household size improves the accuracy of age as a life-cycle proxy as well as being a relevant variable in its own right. In Figure 1, the adjusted Gini measures the areas between line B (rather than the 45 degree line of perfect equality) and the conventional Lorenz curve. The modified comparison line, B, was constructed by dividing the population of U.S. households into 30 groups, based on the age of the head of household and the number of persons in the household. Table 3 orders the groups by mean wealth and shows the percentage of households which they compose, and the share of net worth which they hold.

Table 2

CUMULATIVE PERCENTAGES OF NET WEALTH BY WEALTH CLASSES, 1973

Net Wealth Percentile	Cumulative Share of Net Wealth
0-35	0.0
36-40	0.1
41-45	0.3
46-50	1.0
51-55	2.1
56-60	3.6
61-65	5.5
66-70	8.0
71-75	11.2
76-80	15.3
81-85	20.9
86-90	29.2
91-95	42.5
96	46.4
97	51.3
98	57.6
99	67.4
100	100.0

For example, single person households over age 65 (Group 17) constituted 8.3 percent of all households in 1973. They held a mean net wealth of $36,095 and 7.9 percent of private net wealth. Two person households in the same age group were the highest mean wealth group at $62,553. They accounted for 9.2 percent of households and held 15.3 percent of net wealth.

Information from Table 3 was cumulated and used to plot line B in a manner similar to that of the standard Lorenz curve. I began with single person households 25 years of age or under, as they have the lowest mean net worth of any age/size grouping. Their share of population and share of wealth form the first point on line B. The cumulative shares of the next to the lowest wealth group are then plotted. Line B thus indicates measured inequality which is directly related to differences in age and household size, and the area between lines A and B ( ) divided by 0.5 measures an "age-size" Gini coefficient, which in this case is .24.

Another look at Table 3 suggests that the factor of age is considerably more important than household size, even though the latter was statistically significant in the regression analysis. The lowest five groups are all headed by someone age 25 or younger, and four of the next five are in the second age group of persons 26-35. The highest eight household groups, in terms of mean net wealth, are all headed by persons age 56 or older.

Single person households rank lowest in mean wealth for their age group in a majority of cases. Large households (five or more persons) show no consistent pattern of ranking and account for very small percentages of the population for some age categories. While a larger household may mean more "pooling" of wealth, it can also indicate a greater drain on income for necessary consumption expenditures which would limit asset accumulation.

Paglin's original formulation understates inequality by computing the "P-Gini" as the ratio of the area between B and C (β) to the entire area below the 45 degree line. The MP-Gini is the ratio of the area between B and C (β) to the area below B ($\beta + \gamma$).

Where the conventional Gini is measured as

$$(2)\ \text{Gini} = \frac{\alpha + \beta}{\alpha + \beta + \gamma}$$

the Paglin-Gini would be

$$(3)\ \text{P-Gini} = \frac{\beta}{\alpha + \beta + \gamma}$$

and has now been more correctly identified as

$$(4)\ \text{MP-Gini} = \frac{\beta}{\beta + \gamma}$$

The modified Gini is 0.76 for the 1973 wealth distribution versus the 0.82 of the standard Gini.

4. COMPARISON TO PRIOR RESULTS

These results are consistent with the general thrust of empirical research in wealth distribution: namely, that inequality is quite prevalent and that a large proportion of it is unrelated to life-cycle effects. Simulation models specified to allow accumulation of wealth over time have found that life-cycle effects generate only a fraction of observed inequality [8].

A number of researchers have used age in regression models to explain differences in net wealth and also found it explaining only a small portion of the differences [9]. Wolff recently demonstrated the improvement which can be achieved in regressing wealth on age not only by adding other explanatory variables (in his case, a lifetime earnings estimate), but primarily by omitting the very wealthy from the sample [10]. The R^2 was 0.075 for the botton 95 percent of urban whites in the wealth distribution, compared to 0.015 for the full sample. The model was also much stronger in explaining differences, in what Wolff termed "life-cycle wealth"--the sum of own home, durables, cash, and demand deposits less mortgage debt,--than in explaining differences in

Table 3

WEALTH BY AGE-HOUSEHOLD SIZE GROUP, 1973

	Mean Wealth	Age Class	Household Size	Share of Households (percent)	Share of Net Wealth (percent)
1.	$ 4,171	25 and under	1	1.6	0.2
2.	7,669	25 and under	5+	0.3	0.1
3.	9,301	25 and under	3	2.2	0.5
4.	11,584	25 and under	4	0.9	0.3
5.	12,345	25 and under	2	3.5	1.2
6.	14,407	26-35	1	2.4	0.9
7.	17,476	46-55	1	2.3	1.1
8.	21,826	26-35	2	4.2	2.4
9.	21,993	26-35	5+	4.0	2.4
10.	25,739	26-35	4	5.4	3.7
11.	30,448	56-65	1	3.5	2.8
12.	30,851	36-45	4	4.1	3.3
13.	31,292	26-35	3	4.5	3.7
14.	32,574	36-45	2	1.8	1.6
15.	34,829	36-45	1	1.3	1.2
16.	34,883	over 65	5+	0.4	0.4
17.	36,095	over 65	1	8.3	7.9
18.	37,797	36-45	3	2.3	2.3
19.	39,747	46-55	3	3.7	3.9
20.	40,430	36-45	5+	7.3	7.8
21.	43,669	46-55	5+	4.3	5.0
22.	47,993	46-55	2	4.7	6.0
23.	50,118	56-65	5+	1.1	1.5
24.	51,460	56-65	2	7.3	9.9
25.	53,720	56-65	3	2.9	4.1
26.	59,067	56-65	4	3.5	5.5
27.	60,239	over 65	4	0.5	0.7
28.	62,253	over 65	3	1.5	2.5
29.	62,428	56-65	4	1.2	2.1
30.	62,553	over 65	2	9.2	15.3

NOTE: Percentages do not add to 100 due to rounding.

"capital wealth"--the sum of savings and time deposits, stocks and bonds, investment real estate and business equity less other debt. This was an excellent empirical demonstration of the limited applicability of the life-cycle hypothesis.

5. IMPLICATIONS

The modified Gini computed here may be considered a way of decomposing the measured inequality in the standard Gini coefficient into "within-group variation" and "between group variation." The variation between the mean net wealth of different household sizes and different ages of head of household is excluded from the Gini. Where I differ from Paglin is in comparing the remaining inequality--the "within-group variation"--not to the entire area below the 45 degree line, but to the area below the modified comparison line. This is the relevant comparison since the area between the 45 degree line and the modified comparison line is attributable to between-group variation (age and household size differences) and is not a possible source of within-group variation.

Even with these adjustments to the conventional Gini coefficient, most of the substantial inequality in the distribution of wealth is found to be due to factors other than age or household size--it is "within-group inequality." Regression models which attempt to explain variation in net wealth do very poorly on this data set, as they have on earlier examples. The next step should be the very difficult one of attempting to measure the other more elusive factors which we know cause inequality in wealth distribution: differential patterns of saving, investment, and rate of return and inheritance patterns, as well as the effects of the tax system as a whole on people of different wealth levels.

NOTES AND REFERENCES

[1] Smith, James D., and Stephen D. Franklin. "The Concentration of Personal Wealth, 1922-1969," American Economic Review 64 (May 1974): 162-167; Gallman, Robert E. "Trends in the Size Distribution of Wealth in the Nineteenth Century: Some Speculations" in Lee Soltow, ed., Six Papers on the Size Distribution of Wealth and Income, Studies in Income and Wealth, Vol. 33, New York: Columbia University Press, 1969.

[2] Greenwood, Daphne, "A Method of Estimating the Distribution of Wealth Among American Families in 1973," Papers and Proceedings of the 1981 Meetings of the American Statistical Association, 472-5; see also "An Estimation of U.S. Family Wealth and Its Distribution from Microdata, 1973," Review of Income and Wealth, March 1983, pp. 23-43.

[3] Ando, A. and Modigliani F., "The 'Life Cycle' Hypothesis of Saving: Aggregate Implications and Test," American Economic Review 53, Mar 1963, pp. 55-84.

[4] Lydall, Harold, "The Life Cycle in Income, Saving, and Asset Ownership," Econometrica 23, No. 2, April 1955, pp. 131-150.

[5] Paglin, Morton, "The Measurement and Trend of Inequality: A Basic Revision," American Economic Review 65:4, Dec. 1975, pp. 598-609.

[6] Nelson, Eric, "The Measurement and Trend of Inequality: Comment," American Economic Review 67, June 1977, pp. 497-501; Johnson, William, R., "The Measurement and Trend of Inequality: Comment," American Economic Review 67, June 1977, pp. 502-4; Danziger, Sheldon, Haveman, Robert and Smolinsky, Eugene "The Measurement and Trend of Inequality: Comment," American Economic Review 67, pp. 505-512; Minarik, Joseph J., "The Measurement and Trend of Inequality: Comment," American Economic Review 67, pp. 513-516; Kimen, John C., "The Measurement and Trend of Inequality: Comment," American Economic Review 67, pp. 517-519.

[7] Formby, John P. and Seaks, Terry G. "Paglin's Gini Measure of Inequality: A Modification," American Economic Review 70:3, June 1980, pp. 479-482; Atack, J. and Batemen, F., "The Measurement and Trend of Inequality: An Amendment to a Basic Revision," Economics Letters 4:4, 1979, pp. 389-93.

[8] See, for example, the following literature on simulation models of wealth accumulation: Atkinson, A.B., "The Distribution of Wealth and the Individual Life-Cycle," Oxford Economic Papers; and White, B.B., "Empirical Tests of the Life Cycle Hypothesis," American Economic Review 68:4, Sept. 1978, pp. 547-60.

[9] Projector, Dorothy and Weiss, Gertrude, Survey of Financial Characteristics of Consumers, Federal Reserve Technical Paper, Washington, D.C., Board of Governors of the Federal Reserve System, 1966; Smith, James D., "White Wealth and Black People: The Distribution of Wealth in Washington, D.C. in 1967" in James D. Smith, ed., The Personal Distribution of Income and Wealth, New York, Columbia University Press, 1975; and Mirer, Thad W., "The Wealth-Age Relationship Among the Aged," American Economic Review 69:3, June 1979, pp. 435-443.

[10] Wolff, Edward N., "The accumulation of Household Wealth Over the Life-Cycle: A Microdata Analysis", The Review of Income and Wealth 27:1, March 1981, pp. 75-96.