Review of Income and Wealth
Series 45, Number 2, June 1999

YOUNG BABY BOOMERS' WEALTH

BY JAY L. ZAGORSKY

Ohio State University

Researchers know relatively little about the beginnings of wealth accumulation. This paper analyzes the wealth of young baby boomers, individuals born from 1957 to 1964, using a previously ignored wealth data set, the National Longitudinal Survey of Youth 1979 (NLSY79). First, a detailed data quality evaluation is performed. Findings suggest that not cleaning NLSY79 wealth data causes nonsensical results, but there are no other serious problems. Analyzing the cleaned wealth data quantifies many stylized facts. For example, the typical baby boomer's wealth holdings increase by more than $2,000 a year. Married females hold more wealth than either married or unmarried males. Finally, while young boomers start with a majority of their wealth in illiquid holdings such as automobiles and possessions, they rapidly shift their wealth holdings into homes as they grow older.

I. INTRODUCTION

Wealth is a topic of great interest to the public, policy makers and researchers. Almost all wealth research to date has focused on either the very rich or the elderly, groups who have finished accumulating the majority of their wealth. This research examines a different group—young baby boomers, individuals born from 1957 to 1964. Unlike the rich or the elderly, this group is just beginning their wealth accumulation. Focusing on this stage of life provides clues into how different individuals build wealth over time. Additionally, understanding young baby boomers' wealth is important since their assets may be their primary support in old age if Federal programs, like Social Security and Medicare, are scaled back in the future.

There are only a few micro level data sets that analyze the U.S. experience (Wolff, 1996). This research investigates one of these sources, the National Longitudinal Survey of Youth 1979 cohort. This collection of surveys, referred to as the NLSY79, provides detailed information on individuals born late in the baby boom generation. Detailed survey questions on assets and liabilities, asked in the 1985 to 1996 surveys, enable creation of a longitudinal wealth series for almost seven thousand individuals.

To date, a complete data quality investigation does not exist for the NLSY79s wealth information.[1] To remedy this defect, the paper overviews the wealth data collection and analyzes its strengths and weaknesses. The quality of the data is also investigated by checking the importance of common problems, such as non-response, and how NLSY79 results compare with other data sets. Additionally, a longitudinal net asset series is created and wealth results are analyzed by age, sex and composition.

Note: I wish to thank Randy Olsen, Jean Haurin and Rosella Gardecki for their comments. Jeff Yankow provided excellent research assistance. All remaining errors are mine.

[1]Some data quality aspects are examined in Haurin, Hendershott and Wachter (1996).

The rest of the paper is organized as follows. Section II describes the NLSY79 data generally and then the asset section in particular. A demographic profile of NLSY79 respondents is presented and the strengths and weaknesses of the asset data are described. Section III examines problems and other issues associated with the data such as top coding, out-of-range values, missing answers and non-interviews. Section IV creates the net asset series and analyzes wealth holdings. Section V examines the data's consistency with outside sources. Lastly, the conclusion summarizes the paper's findings.

II. General NLSY79 Data Description

Data used in this research are from the National Longitudinal Survey of Youth 1979 cohort (NLSY79). In 1979, the U.S. Department of Labor began funding a nationwide panel survey of young adults. The first survey contains information on 12,686 individuals between the ages of 14 and 22. Subsequent surveys provide details on how these individuals' lives are changing over time.

While the primary focus of the NLSY79 is tracking the employment, training and educational attainment of the respondents, the survey records a host of other information. For example, data collected include military (ASVAB) test scores, receipt of government assistance, fertility measures, criminal activity, child care use, alcohol use, and health indicators to cite only a few topics.

The NLSY79 panel comprises three groups: a nationally representative sample of 6,111 youths; a supplemental sample of 5,295 poor white, black and Hispanic youth; and 1,280 young members of the military. Over time, funding cutbacks have eliminated most of the military and all poor white oversample youths from the interview, resulting in a sample of 9,964.

The survey is primarily conducted as face-to-face interviews. These interviews were done using paper and pencil questionnaires from 1979 to 1992. Beginning in 1993 the survey switched to a computer assisted personal interview (CAPI) using laptops. This switch reduces the amount of error in the survey's data by automating much of the interview. This paper focuses on the NLSY79 asset questions asked in the surveys conducted from 1985 to 1996.

A. *Wealth Data Summary*

Some NLSY79 respondents have answered wealth questions since the first survey year. The first four surveys (1979, 1980, 1981 and 1982) asked if the respondent had any savings, owned a vehicle or owned a home. Unfortunately, these early questions never asked for a dollar amount of savings or the value of the home and vehicles. Additionally, wealth questions were included only if respondents met one of the following criteria: age 18 or older, had a child, enrolled in college, married, or lived outside their parents' home. NLSY79 data for this time period show few individuals answered the questions until they turned 18 years old. Since the selection process filtered out most of the respondents, this research does not examine the early data.[2]

[2]For example, in 1979, only 5 percent of those under 18 answered asset questions.

Except for the question on home ownership, asset questions were dropped during 1983 and 1984. Then in 1985, after the youngest NLSY79 respondents turned 18 years old, a detailed wealth section was introduced. This section follows a simple pattern: Respondents are first asked if they own an asset or have a debt. If they answer yes, the interviewer asks them to state the exact amount or value.

Initially, respondents answered a maximum of 20 questions about their wealth holdings. This first set of questions asked about home ownership; cash savings; farm, business and real estate holding; vehicles; possessions and major debts. In 1988, the section was expanded by including two new sets of questions, the first about stock and bond holdings and the second on estates and trusts. In 1990, the asset section was expanded again by asking the respondents to estimate their net wealth. Finally, questions were added in 1994 to measure certificate of deposit holdings and retirement accounts such as IRAs, 401Ks and 403Bs.

Two breaks occur in the wealth time series, one in 1991 and the other in 1995. Budgetary restrictions in the survey's 13th round resulted in all wealth questions being eliminated in 1991 for one round of questioning. There are no data for 1995 since the NLSY79 switched from interviewing respondents every year to every other year, beginning in 1994, to lower the survey's cost and reduce respondent burden.

B. *Advantages and Disadvantages of Wealth Questions*

The NLSY79 wealth section has several important strengths. First, it is one of the few longitudinal data sets containing detailed wealth information for a large number of individuals. Another major advantage is the frequent collection intervals, enabling wealth information to be examined in ten out of the twelve years covered by this research.

The NLSY79 also maintains extremely high participation rates. After 17 rounds of interviewing, 86.7 percent of the respondents still answer the survey and of this pool a significant proportion (72.1 percent) have answered the survey every round. Finally, there are no plans to stop interviewing respondents, ensuring that additional wealth and savings information is forthcoming.

The data, unfortunately, also contain deficiencies. The main NLSY79 interview is designed to take approximately one hour. The wealth module is intentionally placed at the survey's end, to ensure that all other information is captured even if wealth questions offend the respondent and result in a refusal to continue. After answering questions for an hour, it is highly probable that respondents are tired and not answering as precisely as earlier. Unfortunately, no information is available to quantify the placement effect.

NLSY79 wealth data have missing values caused by respondents refusing to answer or not knowing answers. Additionally, some data are missing because interviewers accidentally skipped survey questions. While these problems are not unique to NLSY79 these issues are minimized in this analysis by cleaning the data and imputing missing values.

The questionnaire both aggregates and fails to mention a number of wealth categories. One of the most important examples of aggregation is business activities. In a single question, respondents are asked if they own investment real estate,

a personal business or a farm. This aggregation hides important information about individuals who own multiple businesses or investments. Moreover, the questionnaire does not explicitly ask about a number of important wealth categories such as the value of vacation homes, recreational vehicles or stock options.

Finally, unlike other surveys, NLSY79 wealth data focus primarily on the respondent, not the household. Except for information on a spouse or partner, the questions are designed to exclude wealth held by other household members. Parental assets information is not gathered, even if the sample member still lives at home.

C. *Sample Selection, Weighting and Demographics*

Since only 72.1 percent of all eligible NLSY79 respondents answered the survey every year from 1979 to 1996, many possible selection rules can be used to choose which respondents to investigate. This research's results are based on respondents who answered at least one wealth question in *every* survey from 1985 to 1996. This selection rule provides answers based on a "consistent sample" of respondents over time. A previous version of this paper (Zagorsky, 1997) provides results for both this sample and for individuals who participate less frequently. Overall, only small differences exist when comparing results from the two groups. Members of the consistent sample were slightly richer and did not know or refused to answer questions slightly less often. Nevertheless, both samples show similar magnitudes, changes and trends in wealth accumulation over time.

Most results in this research are reported in weighted forms and the places where weights are not used are clearly marked. Every NLSY79 data release contains a weight series that adjusts for the oversampling of poor and minority individuals and accounts for missing respondents. Use of the weights means the figures and tables are nationally representative of young baby boomers. This research uses the 1996 adjustment factors for all years being analyzed.

Table 1 provides a demographic overview of the NLSY79 respondents who meet the participation criteria for this research. Whites comprise 52.9 percent of the sample, while blacks and Hispanics comprise 29.2 percent and 17.9 percent, respectively. The high proportion of blacks and Hispanics, much larger than in the U.S. population, results from the use of an oversample in the survey's design. As the table shows, using weights compensates for this oversampling, and the weighted proportions for race/ethnicity are much closer to the actual population numbers. Since whites holds more assets than blacks and Hispanics, NLSY79 data must be weighted to avoid severely underestimating wealth results.

While the weights compensate for oversampling, Hispanics are still slightly underrepresented in this survey. Since the sample was drawn in 1978, it does not include the recent large influx of Hispanic immigrants. This observation is consistent with the table's statistic that shows 96.3 percent of the NLSY79 respondents were born in the United States.

The sex composition of the sample is split almost evenly between males (49.3 percent) and females (50.7 percent). The table shows many NLSY79 respondents have become highly educated; only 8.9 percent of the sample did *not* obtain a high school or GED degree. While the modal respondent (42.1 percent) completed

TABLE 1

DEMOGRAPHIC CHARACTERISTICS OF NLSY79 RESPONDENTS IN 1966

	Number of Cases	Weighted Value
Race		
White	3,602	81.0%
Black	1,991	13.1%
Hispanic	1,217	5.9%
Sex		
Male	3,274	49.3%
Female	3,536	50.7%
Education		
No degree	874	8.9%
High School Degree	2,951	42.1%
Some College	1,180	16.2%
Junior College Degree	499	8.1%
BA/BS	1,018	18.3%
Advanced Degree	288	6.4%
Born in U.S.A.	6,395	96.3%
Average Age	6,810	34.8 yr.

Note: Total number of respondents examined is 6,810.

high school or obtained a GED, almost half of the sample (49 percent) attended college. Lastly, the spread of ages narrowly focuses on young baby boomers, with ages ranging from 31 to 39 years; the typical respondent in 1996 is about 35 years old.

III. DATA ISSUES AND PROBLEMS

Before using any data set, it is very important to examine the data's quality. Without knowing low-level details, researchers can not judge the accuracy and trustworthiness of results. Moreover, a detailed quality review enables future users to begin using the data armed with a map outlining problem areas. This section first shows that using raw NLSY79 asset data produces nonsensical results. It then examines the effects of top coding, out-of-range values, missing answers, non-interviews and other issues. Overall, the section shows that the primary problem with NLSY79 wealth data is out-of-range values, which must be eliminated before analysis begins.

A. *Problems With Raw Data*

Examining how wealth varies over time is a simple method for determining the accuracy of raw NLSY79 wealth data. Life cycle models of wealth and savings (e.g. Ando and Modigliani, 1963) theoretically show that wealth increases until retirement age. This hypothesis is borne out in many data sets. For example, Kennickell and Starr-McCluer (1994) show wealth increases with age until the mid-50s, while Eller and Fraser (1995) show wealth increasing until the mid-70s.

Since the NLSY79 cohort ages each survey round, one test examines whether net assets increase over time. Unfortunately, raw NLSY79 data fail this test.

The simplest method of creating a net wealth measure for each respondent extracts all wealth values from the NLSY79 CD-ROM for each year and inserts the data into the following formula:

$$
\begin{aligned}
(1) \quad \text{NET ASSETS} = {} & \text{HOME VALUE} - \text{MORTGAGE} \\
& - \text{PROPERTY DEBT} + \text{CASH SAVING} \\
& + \text{STOCK HOLDING} + \text{TRUST} \\
& + \text{BUSINESS EQUITY} - \text{BUSINESS DEBT} \\
& + \text{CAR VALUE} - \text{CAR DEBT} \\
& + \text{POSSESSIONS} - \text{OTHER DEBT} \\
& + \text{IRA} + \text{401K} + \text{CD}.
\end{aligned}
$$

Nonsensical results are obtained when computing this formula.[3] Averaging the net asset terms for every respondent results in the following series; \$10,002 (1985), \$11,561 (1986), \$16,492 (1987), \$20,237 (1988), –\$163,941 (1989), \$39,882 (1990), \$49,859 (1992), \$67,914 (1993) and \$56,873 (1994). While there is a general upward trend in this series, the *negative* \$163,941 value in 1989 and a drop in average wealth of over \$11,000 from 1993 to 1994 show serious problems with using uncleaned NLSY79 wealth data.

A few very large outliers of questionable veracity cause the wild swings in the series. For example, dropping the non-mortgage property debt series in 1989, which contains 160 suspicious values, changes the average from a negative \$163,941 to positive \$39,537. The next sections examine the data problems in more detail.

B. *Top Coding*

One potential problem in NLSY79 wealth data is top coding. The NLSY79 survey takes many measures to protect the confidentiality of respondents. One method of ensuring confidentiality, "top coding" unusually high values is problematic for a number of reasons. First, top coding hides the values of the rich, who are often the very individuals researchers are trying to understand. Second, as the next section shows, many top codes are upwardly biased by nonsensical out of range values.

Finally, top coding causes rich individuals' reported wealth value to fluctuate over time. A simple example shows this fluctuation. Imagine a respondent who owns a house worth half a million dollars in both 1990 and 1992. In 1990, the NLSY79 home top code value was \$250,590. However, in 1992 the value fell to \$225,242. Due to top coding this respondent appears to have lost over \$25,000 in home equity even though the actual home value did not change.

The NLSY79 has used two top coding algorithms. From 1979 to 1988 every NLSY79 asset question that elicited a response that was above the cutoff value, such as \$100,000, was changed into the truncation value plus one dollar, such as \$100,001. Unfortunately this algorithm results in a sharp downward bias in the

[3]Before summing, skips, don't knows, refusals and non-interviews were eliminated.

mean value of NLSY79 asset holdings since the entire right hand tail is truncated. To fix this problem, beginning in 1989 a new algorithm was introduced. The new top code algorithm replaces all values above the cutoff with the average of all outlying values.

TABLE 2

PERCENTAGE OF NLSY79 RESPONDENTS WHOSE ASSETS WERE TOP CODED AND TOP CODE LEVELS, 1985 TO 1996 (UNWEIGHTED)

	Top Code Limit	1985	1986	1987	1988	1989	1990	1992	1993	1994	1996
Home	$150,000	0.3	0.8	1.7	3.2	4.7	6.2	7.9	9.0	11.2	15.0
Mortgate	$150,000	0.1	0.2	0.3	0.0	1.2	1.6	2.4	2.5	3.5	4.5
Property Debt	$150,000	0.0	0.0	0.0	0.0	2.6	0.0	0.0	0.0	0.0	0.0
Cash Savings	$500,000	0.0	0.0	0.0	0.1	0.2	0.0	0.1	0.0	0.0	0.1
Stock/Bonds	$100,000				0.2	0.6	0.3	0.5	0.6	0.6	1.4
Trusts	$100,000				1.2	1.1	1.3	1.3	1.3	1.2	1.7
Business	$500,000	0.2	0.1	0.3	0.3	0.4	0.5	0.5	0.5	0.5	0.8
Business Debt	$500,000	0.0	0.0	0.0	0.1	0.3	0.1	0.1	0.1	0.2	0.2
Car Debt	$30,000	0.0	0.0	0.2	0.2	0.2	0.2	0.2	0.4	0.6	1.2
Car Value	$30,000	0.0	0.0	1.0	1.2	1.5	1.7	2.2	2.4	3.5	5.9
Possessions	$150,000	0.1	0.1	0.1	0.2	0.1	0.0	0.3	0.2	0.1	0.4
Other Debt	$150,000	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.1
Number of Respondents Top Coded		32	53	151	279	545	465	579	635	775	1,106

Note: In 1994 no respondent had an IRA, 401K or CD answer top coded. Data for 1996 are estimated by author based on applying the top code levels to raw asset data. All other years are based on calculations using data on the public release CD-ROM.

Table 2 shows both the NLSY79 top code levels and the extent of top coding each year. The table clearly shows the wide variation in top coding. The vast majority of top coding occurs in three places: home, mortgage and vehicle values. For example, in 1985 only 0.3 percent of NLSY79 respondents owned homes worth more than $150,000; by 1996, 15 percent crossed this threshold.

The bottom line of the table shows that increasing numbers of respondents are top coded over time. While in 1985 slightly more than thirty respondents were top coded, by 1994 well over a thousand had suppressed values. This increasing trend in the number of top coded NLSY79 respondents over time is not surprising given that top code values were fixed in 1985. Unfortunately, the increasing number of top coded individuals means less usable data each year for research.

C. *Out-of-Range Values*

Another major data issue, particularly in 1989, is out-of-range values. Out-of-range values occur when the asset or debt value is greater than permitted by the questionnaire. These values are problematic because the top code computations, since 1989, average all very high values, including those which are out-of-range. For example, in 1989 four respondents' automobile values were coded out-of-range. These four contributed almost one million dollars each to the top code calculation.

In 1985, zero respondents had out-of-range values. In 1986, 1987 and 1988 there were one, two and one respondent(s) coded out-of-range respectively.

Unfortunately, in 1989 there were 194 respondents with out-of-range values.[4] In 1990 and 1992, the number of out-of-range cases dropped back to one and five respectively. Out-of-range values do not occur after 1992 since the computerized questionnaires used in 1993 and beyond prevent interviewers from entering out-of-range values. If exceptional values are encountered, however, the NLSY79 software allows interviewers to enter these numbers plus an explanation in a comment field.

The most problematic area in 1989 occurs in OTHER PROPERTY DEBT, where a suspiciously high 160 respondents were coded out-of-range. This category is designed to capture long-term expenditures such as assessments for street, water and sewer improvements or home remodeling and repairs. Modal values in most years are between $10,000 and $15,000. Moreover, except for 1989 when out-of-range respondents were given a value of almost $10 million, few individuals in other years have reported more than $50,000 of debts in this category.

There are a number of potential reasons why so many out-of-range answers appear in the 1989 data: field interviewers may miscode a respondent's answer, the respondent's answer truly could be greater than the maximum allowed, the questionnaire's maximum field widths can be too small, interviewers could be improperly trained, or the 1989 CAPI experiment could have gone awry. Investigation suggests that none of these reasons are the problem; the most likely reason is that support staff incorrectly transcribed respondents' answers from the paper questionnaire to the computer.[5] Transcription errors are possible since out-of-range values (9999996) are very similar to refusals (9999997) and does not know (9999998). To avoid using these suspicious values, this research imputes answers for all out-of-range data.

D. *Missing Answers*

Another major data quality concern is bias caused by missing answers. Missing assets reduce estimates of people's net worth, while missing debts increase estimates. There are three possible reasons why answers are missing in the NLSY79: invalid skips, don't knows and refusals. Very few individuals have missing answers for questions which determine whether the respondent has a given asset or liability. However, response rates are lower for questions that determine the exact amounts of assets and liabilities. In particular, a large percentage of NLSY79 respondents do not know the value of trust or retirement accounts. Results in this subsection are presented in unweighted form to provide the reader with a better view of the actual number of respondents not answering questions. Weighting the results reduces the problem of item non-response.

Missing values are caused by both the interviewer and interviewee. From 1985 to 1992 interviewers using paper and pencil questionnaires occasionally followed the question flow incorrectly and skipped asset questions. The simplicity of the asset section made such incorrect skips fairly infrequent. Fewer than 10 respondents were incorrectly skipped in the asset section during any year. Since

[4]The number of out-of-range individuals are: property value 2, mortgage value 16, property debt 160, cash 14, stocks 10, trusts 3, business value 2, business debt 10, vehicle value 4 and other debt 4.

[5]This idea was suggested by current NLSY79 data archivist Karima Nagi.

the introduction of CAPI in 1993, it is very difficult for interviewers to incorrectly skip a question. Moreover, invalid skips do not introduce systematic bias, since the skips are correlated with interviewer rather than respondent characteristics.

The more important case is respondent-generated skips. These skips are highly correlated with respondent characteristics (Ross and Reynolds, 1996). This subsection breaks down non-response based on the two styles of wealth questions asked in the NLSY79. The first style is the general question which asks if the respondent has a particular asset or debt. If the respondent answers affirmatively, then the second question probes for a specific amount.

Few respondents refuse or don't know the answer to the ownership questions. Examining ownership questions by year reveals that in eight out of the ten NLSY79 surveys used in this research, unweighted response rates exceed 99 percent. The lowest ownership response year is 1994, when 98.8 percent of the respondents answered the wealth ownership questions. Looking at ownership questions by type of asset also shows few problems. Average responses rates over the ten surveys are: home 99.9 percent, cash saving 99.8 percent, IRA 98.2 percent, 401K 98.4 percent, CD 96.1 percent, stocks/bonds 99.0 percent, trusts 96.5 percent, business 99.6 percent, car ownership 100 percent, car debt 99.9 percent, possessions 99.9 percent and other debts 99.8 percent.

Response rates are lower for the second question about each asset category, which asks the amount of the asset or debt. Table 3 shows the trends over time in refusals and don't knows. Each entry in the table is calculated by dividing the number of refusals or don't knows by the number of respondents answering yes to the relevant ownership question.

TABLE 3

PERCENTAGE REFUSING OR NOT KNOWING HOW MUCH A DEBT OR ASSET IS WORTH (UNWEIGHTED)

Type	Refusal	Don't Know	Type	Refusal	Don't Know
Home	0.1%	1.6%	Trusts	2.3%	25.8%
Mortgage	0.4%	2.3%	Business	1.0%	7.3%
Property Debt	0.3%	1.2%	Business Debt	0.9%	6.4%
Cash Savings	1.6%	2.5%	Car Debt	0.2%	1.5%
IRAs	2.3%	14.2%	Car Value	0.1%	3.3%
401Ks	1.8%	15.8%	Possessions	0.3%	2.1%
CDs	5.5%	10.0%	Other Debt	0.3%	1.5%
Stock/Bond	1.6%	15.5%			

Note: Averages from 1985 to 1996. Entries are calculated by dividing the number of refusals or don't knows by the number of respondents stating yes to the question on ownership.

Examining the table shows that response rate problems are predominantly caused by respondents not knowing the answer rather than by refusals. For example, an average of only 0.1 percent of the respondents refused to divulge their home's value, while 1.6 percent did not know the value. Looking at the worst case, 2.3 percent of the respondents refused to state the value of their trust fund and 25.8 percent did not know its value. Over all ten wealth surveys, the typical respondent refused 0.25 times but did not know the answer 2.79 times. Clearly, large numbers of respondents do not have precise knowledge of their asset and liability values.

Overall, the table shows different response rates based on the type of wealth. Respondents know and will divulge information about their real or physical wealth. This is seen by low refusal and don't know rates on home, mortgage, vehicles, possessions and other debt categories. Financial wealth is very different. Respondents have both higher refusal and higher don't know rates on their cash savings, stocks, bonds, retirement plans, trusts and businesses.[6]

Previously, Juster and Kuester (1991) examined the NLS of Mature Men, the Retirement History Survey and the 1983 Survey of Consumer Finances. They found "there is some significant relation between item non-response and wealth." Although there is a relatively low correlation between average net assets and total refusals (0.037) or don't knows (0.163), this subsection suggests the major reason for young baby boomers' response rate problems is that respondents do not have a clear idea of their wealth's worth.

E. *Non-interviews*

While the NLSY79 maintains very high response rates, some individuals have left the survey; if relatively richer people drop out of the survey, wealth calculations will underestimate the population's true wealth. Conversely, if poorer individuals drop out of the survey, wealth is overestimated. Although non-interviews are *not* a major problem in the NLSY79—the survey has maintained response rates of close to 90 percent, and approximately half (48.3 percent) of the respondents dropped from this research missed only one interview—the effects of non-interviews bias must be examined.

A simple method for understanding the ramifications of non-participation compares the assets of respondents who participate irregularly with respondents who consistently participate. Respondents were broken into two groups. The first group consists of respondents who answer the survey every year. The second group consists of all other NLSY79 respondents who were not dropped for funding reasons, or death and who answered at least one survey since 1985. For both groups, a simple net asset series was calculated using equation (1). The only data cleaning done to the series was to set all out-of-range values to missing because of the issues discussed previously. Unlike the final net asset series, discussed below, imputation of missing values was not done to ensure that estimated values would not affect the comparison.

Table 4 shows the asset percentiles by year for both groups. The table clearly shows that respondents participating every year have more assets than irregular participants. Out of the 70 cells, irregular participants have higher values in three negative net worth cells. In all other cases, consistent sample members in each percentile outrank their less faithful counterparts. The median assets of respondents who answer in every year are 54.8 percent higher than those who do not consistently participate.

[6]The high ratio of don't know to refusals may also be due to respondents using don't know as a more polite method to avoid answering the question than refusing.

TABLE 4

NET ASSETS BY PERCENTILE OF NLSY79 RESPONDENTS WHO ANSWER EVERY YEAR AND RESPONDENTS WHO DO NOT (IN CURRENT DOLLARS, WEIGHTED)

Answer Every Year	1985	1986	1987	1988	1989	1990	1992	1993	1994	1996
Percentile 95%	61,000	72,000	99,400	146,700	179,700	201,000	209,000	230,000	270,500	391,500
90%	35,500	43,300	62,000	87,200	100,000	116,300	133,000	144,500	171,000	231,000
75%	13,000	15,900	25,000	33,300	42,200	51,000	62,000	70,600	83,200	109,000
Median	4,000	4,700	7,500	9,500	12,000	15,000	20,000	24,002	30,400	41,200
25%	600	800	1,200	1,500	1,884	2,200	3,000	3,500	5,100	7,000
10%	−500	−900	−800	−800	−800	−700	−400	−100	0	0
5%	−2,895	−3,750	−4,300	−4,000	−3,800	−4,600	−4,900	−4,000	−3,692	−4,100
Weighted Mean	17,642	16,785	27,572	40,890	48,679	64,894	66,508	68,703	82,005	110,692
Not In Every Year	1985	1986	1987	1988	1989	1990	1992	1993	1994	1996
Percentile 95%	49,000	56,000	81,000	120,000	130,500	156,000	178,220	179,000	200,000	285,000
90%	27,500	33,050	42,500	61,800	74,000	90,000	103,000	104,000	123,000	161,000
75%	9,100	10,100	15,000	19,000	24,900	29,000	39,000	43,000	50,700	62,600
Median	2,500	2,800	3,500	4,500	5,500	6,000	7,820	9,500	11,300	14,100
25%	100	51	200	500	300	400	600	500	800	1,100
10%	−500	−1,050	−1,400	−1,400	−2,000	−2,000	−1,300	−1,350	−835	−1,000
5%	−2,000	−3,200	−5,000	−4,100	−5,798	−6,300	−5,200	−5,600	−5,000	−5,398
Weighted Mean	12,937	12,710	17,500	35,562	40,955	38,249	44,777	43,865	51,588	72,746

Note: Net assets are calculated using equation (1) in the paper. Out-of-range values are set to zero. Imputation and inflation adjustments are not performed on these data. The number of respondents who answer every year is 6,810. The number of respondents with valid wealth data in a given year who did not answer every year are; 2,679 (1985), 2,490 (1986), 2,334 (1987), 2,302 (1988), 2,432 (1989), 2,315 (1990), 2,386 (1992), 2,364 (1993), 2,267 (1994) and 2,038 (1996).

Another test examines mean net assets. Table 4 also contains the weighted means for both types of respondents. Examining the means shows that consistent respondents have net assets averaging 29.8 percent more than irregular participants, with a range of between 13 and 41 percent.

Overall, these results suggest that NLSY79 net asset data overestimates the wealth of U.S. young baby boomers. Comparing the cells indicates that respondents who answer every year have roughly one-third more assets each year than those who do not respond every year. Since the consistent sample comprises three-quarters of all respondents, this suggests consistent sample estimates of wealth are approximately 8.33 percent higher than a "true" national wealth sample would find for individuals in the NLSY79 age range.[7]

Unlike Juster and Kuester's (1991) finding that the NLS of Mature Men and the Retirement History Survey lost wealthy individuals over time, the NLSY79 appears to be retaining wealthier and losing poorer members, causing a slight upward bias to wealth results.

F. *Other Data Quality Issues*

There are three other data quality issues that deserve attention: living outside the U.S., aggregation bias and measurement error. This section explains the issues that each problem poses for researchers.

[7]The full sample is comprised 75 percent of the consistent sample and 25 percent of irregulars. The 25 percent hold 1/3 less assets than consistent sample members. Hence, the bias is approximately 25% * 33% = 8.3%.

Living outside the U.S., after the initial 1979 interview, does not preclude a respondent from being interviewed.[8] For example, in 1993, 66 respondents lived abroad. Before 1993, if the respondent could not report their wealth in U.S. dollars, then wealth values were reported as "invalid skips." Beginning in 1993, interviewers were instructed to use the questionnaire's comment field to indicate the asset's value in foreign currency. NLSY79 support staff then multiply the foreign asset's value by the prevailing currency conversion rate to determine an approximate value in U.S. dollars. These converted values are inserted into the public use data as if the respondent had provided the asset's value in dollar terms. This issue is not a major data concern since only a tiny fraction of wealth is reported in foreign currency.[9]

A second data quality issue is aggregation bias. The NLSY79 questionnaire lumps many categories together. For example, stocks, bonds and mutual funds are all considered in a single question. For respondents who hold multiple financial assets, important information could be lost when the respondent mentally sums their holdings. NLSY79 questionnaire designers are cognizant of this and new questions are added over time as respondents start building up assets. The primary problem with increasing the wealth section is greater respondent burden. NLSY79 surveys currently take approximately one hour to complete and there is a strong desire to avoid increasing its length.

Finally, is the problem of measurement error. Due to the concern over the survey's length, NLSY79 respondents are not explicitly encouraged to get out their financial records when answering wealth questions. Recent Census Bureau research (Marquis, 1995) finds that people are much more accurate when encouraged to use financial records in reporting income, although the payoffs do not occur the first time the survey is changed. Moreover, using records is very costly to implement since each interview takes much longer. Examination of the raw NLSY79 data shows clear lumping around round figures like $10,000 and $100,000. While excessive rounding clearly shows measurement imprecision, the overall effects on net asset data are unknown.

IV. Wealth Data Analysis

How much wealth have young baby boomers accumulated? How many have savings and how many are in debt? To answer these questions the first part of this section describes what proportion of respondents own various types of assets. The second part explains how a cleaned NLSY79 net asset series is constructed. Finally, summary statistics are presented which show how wealth varies based on age, sex and time.

A. *Ownership Changes Over Time*

Table 5 examines how asset ownership has changed over time. The table shows the percentage of young baby boomers with a particular asset or debt each

[8]Being institutionalized also does not preclude being interviewed. In recent rounds more than 150 respondents have been interviewed in jail.

[9]For example, the raw 1993 data contained only 5 asset values not listed in U.S. dollars.

survey year. Many asset and debt categories show clear trends in ownership. For example, 20.7 percent of respondents owned homes during 1985, when most respondents were in their early twenties. Home ownership continuously rises over time, with 62.9 percent of respondents owning their residence by 1996.

TABLE 5

PERCENT OF YOUNG BABY BOOMERS WHO OWN ASSETS OR HAVE DEBTS BY YEAR (WEIGHTED)

Asset/Debt	1985	1986	1987	1988	1989	1990	1992	1993	1994	1996
Home	20.7%	25.1%	31.7%	37.4%	41.0%	46.2%	51.7%	54.7%	59.0%	62.9%
Cash Savings	72.5%	73.8%	74.1%	63.1%	77.6%	77.1%	77.4%	77.2%	77.0%	77.7%
IRAs	—	—	—	—	—	—	—	—	20.9%	24.3%
401Ks	—	—	—	—	—	—	—	—	30.3%	35.1%
CDs	—	—	—	—	—	—	—	—	4.8%	7.1%
Stock Bond	—	—	—	19.1%	19.9%	20.0%	21.0%	22.8%	21.1%	22.4%
Trusts	—	—	—	4.4%	4.3%	4.6%	4.8%	5.2%	4.6%	5.2%
Business	6.8%	7.7%	8.7%	9.3%	10.9%	11.2%	11.8%	12.2%	12.7%	13.1%
Own Car	78.2%	82.0%	83.9%	86.0%	87.5%	87.8%	87.9%	88.7%	89.7%	89.7%
Car Payments	39.9%	43.3%	44.6%	48.7%	50.0%	49.5%	43.2%	43.5%	47.0%	47.0%
Possessions	58.5%	62.0%	64.4%	69.4%	69.0%	70.5%	69.0%	68.9%	70.3%	71.3%
Other Debt	33.7%	37.3%	37.9%	40.7%	41.4%	42.6%	43.7%	40.4%	39.2%	37.5%

While time series data are limited, IRAs, 401Ks and CD holdings also appear to be growing rapidly. The line labeled "business" refers to ownership of either a business, farm or investment real estate; these data show that business ownership rates have nearly doubled, rising from 6.8 percent in 1985 to 13.1 percent in 1996. Car ownership slowly rises over time from the high 70 percent range to the high 80s. Not surprisingly, the percentage of individuals who have outstanding debt on these vehicles climbs at the same time to 47 percent in 1996. Finally, while around 58.5 percent of young baby boomers possessed items of high value (worth more than $500) such as jewelry, furniture or stereos in 1985, by 1996 71.3 percent had accumulated high-value possessions.

Other question series are started too late for researchers to understand when individuals begin to accumulate specific assets or debts. For example, the NLSY79 does not indicate at what age savings begins. Cash savings rates vary around the mid-70 percent range with no clear trend. About 20 percent of baby boomers hold stocks and bonds, while 5 percent have trusts during the years data are available.[10] Finally, depending on the year, 30 to 40 percent of respondents have other debts of more than $500.

B. *Net Assets Series Creation*

While knowing the percentage of young baby boomers who own a particular asset is important, of more interest is how much these assets are worth. This section describes how a net asset series for each respondent is created. Briefly, the

[10]Recent cognitive testing suggests that some respondents are confused by the trust fund question and include both accounts set up legally as a trust funds and money they are informally "trusted" to keep, such as gifts they hold for their children.

net asset series was created by removing top coding, removing out-of-range values, imputing for missing values, summing the components of net wealth and then adjusting for inflation. The exact details are described below.

The first step in building the series was to extract the raw un-top coded data from the original data tapes. These raw data were then merged with the filtered information on the NLSY79 public use CD-ROM. The data needed to be merged because both sources of data have different types of information. The original data tapes have full information on asset values, particularly for high wealth individuals. The public use data tapes have codes explaining why asset data are not present for a particular individual. A simple merge algorithm was used to combine the two sources. Assets below the top code limit were taken from the public use CD-ROM, while assets at or above the top code limit were taken from the raw tapes.

The next step was to find all valid skips (-4) in the data. Valid skips mean the respondent does not own the asset, and these items were given a value of zero. This step ensures that individuals with no wealth have zero assets. Then, all other problem flags, marking invalid skips, refusals, don't knows and out-of-range, were set to minus one. This value flagged the asset as a candidate for the imputation algorithm.

The third step was to impute missing values. While many imputation algorithms are available (e.g. Little and Rubin, 1987), the longitudinal aspect of the NLSY79 data provides a simple but effective solution. Data were linearly interpolated if bracketing values were available. This algorithm is a slight refinement to the procedure used in the Netherlands Socio-Economic Panel (Camphuis, 1993) and is based on the assumption that wealth changes are primarily low frequency trend movements.[11] This imputation choice causes some data smoothing because of the interpolation. However, no matter what algorithm is chosen, the high response rates mean NLSY79 data need little imputation.

Finally, the data are summed using equation (1) and adjusted to 1990 dollars to enable inter-temporal comparisons.[12] These steps eliminate most of the problems in the NLSY79 data and provide a full dataset in real 1990 dollars that enables researchers to understand how the wealth of young baby boomers changes over time.

C. *General Wealth Holdings*

This section first examines the percentage of respondents with positive, negative and no wealth holdings and shows that a significant fraction of young baby boomers have negative or no wealth holdings. Table 6 shows that roughly, 5 percent of all individuals each year have no wealth, 10 percent are in debt and around 85 percent have positive wealth values. The table shows the number of

[11]Camphuis used the value reported in an adjacent survey.

[12]Net assets are rebased using the CPI-W with 1990 as the base year. The adjustment factors used are; 0.829 (1985), 0.841 (1986), 0.871 (1987), 0.906 (1988), 0.950 (1989), 1.000 (1990), 1.040 (1991), 1.070 (1992), 1.101 (1993), 1.129 (1994), 1.161 (1995) and 1.194 (1996).

positive net worth individuals rises by half a percentage point yearly. Nevertheless, a significant fraction of late baby boomers have no wealth in all survey years.

TABLE 6

PERCENTAGE OF YOUNG BABY BOOMERS WITH NEGATIVE, ZERO AND POSITIVE NET WEALTH (WEIGHTED)

Year	Negative	Zero	Positive
1985	11.9%	7.6%	80.5%
1986	12.5%	5.8%	81.7%
1987	11.9%	5.7%	82.4%
1988	11.6%	4.6%	83.8%
1989	11.7%	4.5%	83.8%
1990	11.0%	4.1%	84.9%
1992	10.0%	4.5%	85.5%
1993	9.7%	4.5%	85.8%
1994	8.4%	4.3%	87.3%
1996	8.5%	4.0%	87.5%

Given the uncertainty surrounding the future of government entitlement programs such as Social Security, the results in Table 6 are cause for concern. By 1996, when NLSY79 respondents were in their early to late 30s, almost 13 percent of the respondents had no wealth. Assuming the increase in the number of positive net worth individuals continues at its half percent climb, the last negative or zero net worth respondent will begin to save in 26 years, just as the cohort enters retirement. If this trend continues, many individuals will have few assets upon retiring and are relying on government programs to support them during their old-age.

D. *Age*

Overall, the net wealth of young baby boomers increases linearly with age. Figure 1 examines how median net assets from 1985 to 1996 are distributed by age. Life cycle theory (Ando and Modigliani, 1963) suggests that wealth should follow a hump-shaped pattern. Wealth is expected to rise each year until retirement and then decline. Figure 1 roughly matches this theory by showing net assets generally increasing with age.

Median assets range from a low of almost $1,500 dollars for 28 year olds in 1985 to a high of approximately $50,000 dollars for 35 year olds in 1996. Each age slice shows a steady growth in asset holdings over time. Holding age constant, median asset holdings grow each year for all ages by an average of $2,394. For example, median wealth for 30 year olds increased from $3,237 in 1986 to $5,740 in 1987.

In any particular year older respondents typically have more wealth than younger ones. Holding the year constant shows that each additional year of age increases median wealth by $2,576.[13] For example, in 1988 median wealth for 32 year olds was $11,037 but 33 year olds held $13,686.

[13]Wealth increased by age except for the last group of 36 year olds. This age group may not be truly representative since it contains only half the number of respondents found in other age groups.

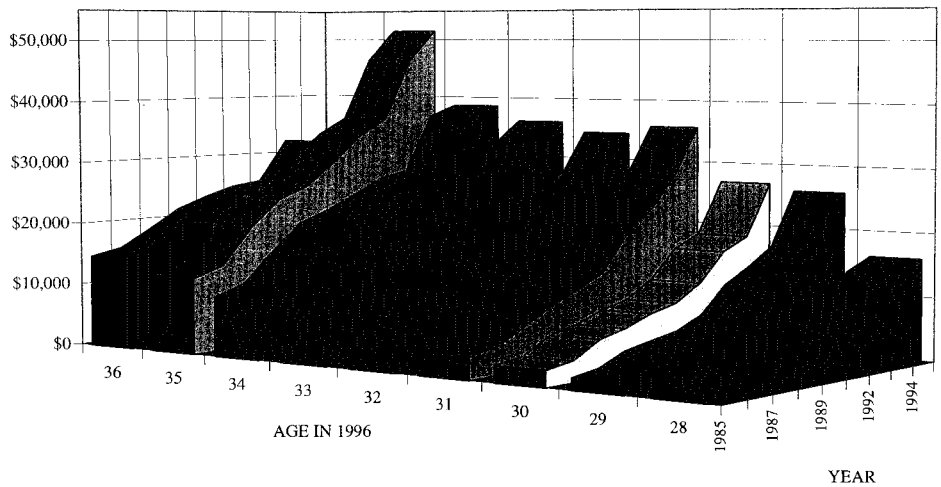

Figure 1. Young Baby Boomer's Median Net Assets by Age (1990 Dollars, Weighted)

Figure 1 also shows a compression in the range of wealth over time. In 1985 (front part of the graph), the gap in wealth between older and younger respondents was more than a factor of six, but by 1996 (back part of the graph) the maximum gap is less than twofold. Younger respondents quickly build up wealth, narrowing the gap between the ages.

Finally, for the period under investigation, wealth does not grow in the upwardly curving pattern suggested by the life cycle theory. Wealth regressions were run with both Age and Age squared explanatory variables. Adding in an Age squared variable to capture curvature rarely resulted in this new term having a statistically significant coefficient and often resulted in both age terms becoming statistically insignificant.[14] The lack of an upwardly curving pattern, however, does not invalidate the life-cycle theory. Rapid asset accumulation should occur close to retirement, which for the typical young baby boomer is still far away.

F. *Sex*

Wealth patterns broken down by sex are a topic of much interest. The NLSY79 makes it difficult to perfectly analyze wealth by sex since married respondents include assets held by their spouse. Nevertheless, examining mean net assets by sex from 1985 to 1996 shows that women hold higher amounts of wealth than men. Women hold an average of $6,494 more assets (roughly 13 percent) than men. Broken down by year male and female wealth holdings were quite similar in 1985, 1986 and 1987, but female wealth grew rapidly beginning in 1988. Finding that women hold more wealth than men is rather surprising since, women typically have lower overall earnings than men. Median data tell a different story than mean data. Median female wealth is very similar to male

[14]The only years in which the age^2 term was significant were 1986, 1988 and 1990. Other explanatory variables in the regressions included race, sex and schooling terms.

wealth over time. Men have slightly more wealth in 8 out of the 10 years of wealth observations and no giant gap appears between the sexes.

The cause of the conflicting results is simple. Figure 2 displays net wealth broken down by both sex and marital status from 1985 to 1996. This picture shows that married females hold higher amounts of wealth than married males. Unmarried males hold more wealth than unmarried females. Median calculations are picking up the large number of unmarried, relatively poor females, which drag the median downward. The mean calculations are picking up the smaller number of relatively rich married females, which pulls the mean upward.

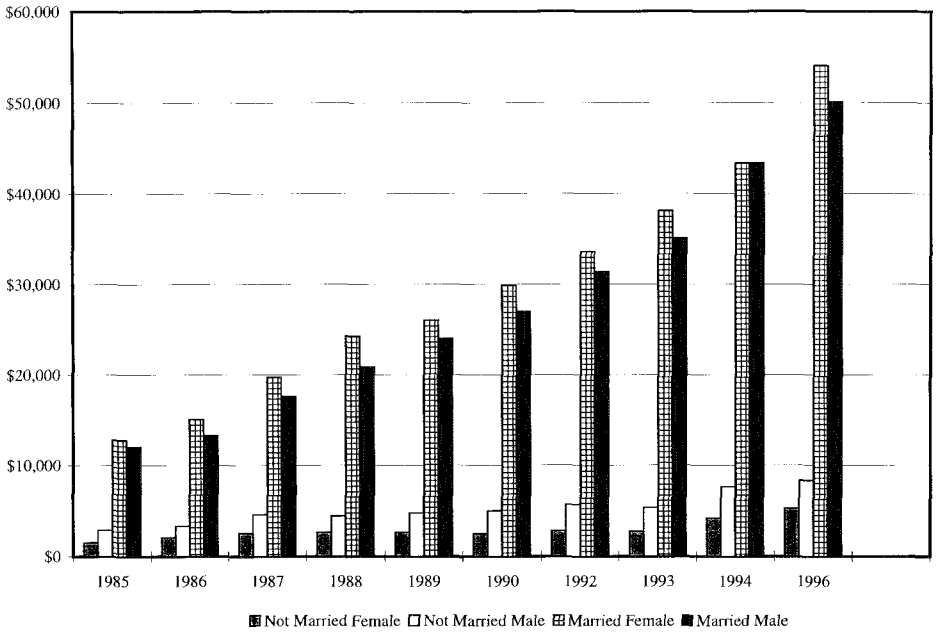

Figure 2. Young Baby Boomer's Median Net Assets by Sex and Marital Status (1990 Dollars, Weighted)

The remaining puzzle is why married females in the NLSY79 have more wealth than married males. The answer is that the NLSY79 wealth section provides wealth data based on both the respondent and their spouse. Females in the sample tend to have older spouses, while married males in sample tend to have younger spouses. The average female respondent married a man born in 1958, while the average male married a female born in 1962. Since wealth holdings, as seen above, are directly related to age, married females have higher wealth than married males because their partners have had, on average, 4 more years to save.

H. *Composition of Wealth*

This section investigates how the composition of wealth changes over time. The composition of young baby boomers' wealth is examined by breaking down assets and debts into three major categories: financial, or liquid wealth holdings;

home, or residential equity; and illiquid, which includes assets difficult to sell quickly.

Financial wealth is calculated by equation (2). This category measures the highly liquid assets held by each person and consists of cash, stocks, bonds, CDs, trusts and retirement accounts.

$$\text{(2)} \qquad \text{FINANCIAL} = \text{CASH SAVING} + \text{STOCK HOLDING} + \text{TRUST} + \text{IRA} + 401\text{K} + \text{CD}.$$

Home, the second category, is calculated by equation (3), which subtracts mortgages and property debt from the residence's current market value. As Wolff (1994) points out, the most important asset in most U.S. portfolios is the individual's primary residence.

$$\text{(3)} \qquad \text{HOME} = \text{HOME VALUE} - \text{MORTGAGE} - \text{PROPERTY DEBT}.$$

Illiquid wealth, the last category, is calculated by equation (4). This collection of variables contains assets that are either difficult to sell quickly, such as a business, or assets that typically have large spreads between asking and selling prices, such as cars. This group of variables comprises net business worth, net vehicle worth, possessions, and other debts.

$$\text{(4)} \qquad \text{ILLIQUID} = \text{BUSINESS EQUITY} - \text{BUSINESS DEBT} + \text{CAR VALUE} - \text{CAR DEBT} + \text{POSSESSIONS} - \text{OTHER DEBT}.$$

Table 7 examines the wealth composition of young baby boomers over time. The three highlighted columns in each table, labeled Financial, Home and Illiquid, contain the average percentage of wealth held in each category. Each highlighted column is followed by the mean and median value of the category for that year.

TABLE 7

COMPOSITION OF YOUNG BABY BOOMER'S WEALTH (1990 DOLLARS, WEIGHTED)

	Financial	Mean	Median	**Home**	Mean	Median	**Illiquid**	Mean	Median
1985	**21.3%**	$3,738	$483	**14.2%**	$5,383	$0	**65.6%**	$10,907	$2,412
1986	**18.7%**	$4,114	$594	**14.3%**	$6,233	$0	**67.8%**	$8,762	$2,734
1987	**18.7%**	$5,630	$803	**18.9%**	$9,429	$0	**63.0%**	$15,497	$3,673
1988	**19.6%**	$15,462	$551	**20.9%**	$12,437	$0	**60.1%**	$15,668	$4,415
1989	**19.9%**	$19,375	$1,052	**24.5%**	$14,407	$0	**56.3%**	$17,936	$4,210
1990	**20.8%**	$25,474	$1,200	**27.5%**	$17,837	$0	**52.4%**	$20,570	$4,500
1992	**24.4%**	$25,758	$1,869	**30.2%**	$17,486	$0	**46.4%**	$18,586	$4,672
1993	**21.0%**	$23,877	$1,816	**23.7%**	$19,145	$1,816	**56.2%**	$19,704	$5,449
1994	**24.0%**	$27,786	$2,657	**31.8%**	$21,950	$5,314	**45.0%**	$23,585	$6,111
1996	**26.4%**	$38,235	$4,020	**34.6%**	$26,774	$10,050	**40.7%**	$25,249	$6,700

Note: The percentage of wealth held in financial, home and illiquid assets adds up to slightly more than 100% due to rounding errors.

Overall, financial assets comprise between a fifth and a quarter of wealth holdings. While there is little upward or downward trend in the percentage of wealth held as financial assets over time, the amount of financial holdings dramatically increases over time. Mean financial wealth increased by more than

$34,500 over the 11 years, while median financial assets increased by approximately $3,500.

In contrast to the relative stability of the percentage devoted to financial holdings, home equity as a percentage of assets clearly increases. Home wealth climbed from 14.2 percent in 1985 to 34.6 percent by 1996.[15] Over the entire 11 years, growth is clearly seen in mean home assets, which increased from $5,383 to $26,774. While the median young baby boomer did not own a home during the 1980s and early 1990s, the typical individual did purchase a home by the mid 1990s. Since the mid 1990s median home assets have increased quickly, jumping almost $5,000 from 1994 to 1996.

Growth in housing assets occurs at the expense of illiquid holdings. Wealth held in illiquid assets, such as personal possessions and automobiles, fell from 65.6 percent in 1985 to 40.7 percent in 1996. As young baby boomers age, they are shifting their portfolios away from illiquid holdings and toward their primary residence. Nevertheless, even though the percentage shrinks, the dollar value of illiquid assets grows. The mean of illiquid assets grew by approximately $15,000 from 1985 to 1996, and the median grew by $4,200. Growth in this category, however, is the slowest of all three groups.

V. External Validity: Comparing Net Assets with Outside Sources

This section compares the NLSY79 with other wealth data sources to check the accuracy of net wealth results. Unfortunately, while surveys like the Federal Reserve Board's Survey of Consumer Finances (SCF), the Institute for Social Research's Panel Study of Income Dynamics (PSID) and the Census Bureau's Survey of Income and Program Participation (SIPP) all provide wealth data, there is little consistency between them and the NLSY79. Wolff (1996) clearly describes the comparison problem: "Wealth data—insofar as it has been analyzed—has often produced contradictory and inconsistent estimates." Evidence presented in this section suggests that NLSY79 wealth data are roughly similar to other sources, but that overall values are higher.

The premier U.S. wealth survey is the Federal Reserve Board's SCF. This survey, run every three years, provides the central bank with a detailed picture of the assets and liabilities of both a cross-section of the population and of those who are very rich. The SCF probes the respondent using detailed and highly specific questions. This ensures even small debts and wealth holdings are recorded and suggests that SCF wealth estimates should be higher than NLSY79 estimates.

Kennickell and Starr-McCluer's (1994) Table 3 provides figures on family net worth of SCF respondents based on selected characteristics such as age and race for the years 1989 and 1992. Rebasing their data from 1992 into 1990 dollars for families whose head was under 35 results in a 1989 mean of $56,450 and median of $7,850. The NLSY79 sample's mean in 1989 is $53,779 and median is $12,315. Mean SCF wealth in 1992 is $56,260, with median wealth at $9,720. This

[15]The percentage of wealth held in housing fell dramatically in 1993 due to a few cases with almost zero net worth. For example, one respondent had housing assets of almost –$36,000 which were offset by an illiquid total of approximately $37,000. Removing just this case raises the home percentage from 23.7 percent to 25.3 percent.

compares with the NLSY79 sample wealth's 1992 mean of $63,711 and median of $18,598. Overall, NLSY79 wealth data are similar to but slightly higher than comparable SCF information.

This unexpected result may be due to the longitudinal nature of the NLSY79. Research by Ferber (1959) suggests respondents who develop a long-term relationship with the interviewer report more wealth than those who do not. As the SCF is cross-sectional, its respondents may tend to under-report wealth to the interviewer while NLSY79 respondents may be more truthful. Another factor causing this difference is the attrition of poorer NLSY79 respondents, discussed earlier, which pushes up NLSY79 wealth estimates.

Wealth data have been collected on three occasions in the PSID. Wealth supplements, in 1984, 1989 and 1994, provide a longitudinal look at family wealth dynamics. Unfortunately, this presents a major mismatch with the NLSY79 since the latter survey asks only about wealth holdings of individuals plus spouses or partners. The PSID, however, asks about all family members living in the household. Under a PSID format, NLSY79 respondents living with their parents would report parental assets and liabilities. This is not a trivial difference, since in 1985 over a quarter of NLSY79 respondents lived with their parents. While many have left home over time, in 1994 nine percent of NLSY79 respondents still resided in parental households.

Hurst, Luoh and Stafford's (1998) Table 9 provides wealth holdings in 1989 by age cohort for the PSID. In 1990 dollars, PSID 25 to 34 years olds have a mean wealth of $36,997 and a median wealth of $10,549, compared to the NLSY79 consistent sample's mean of $53,779 and median of $12,315.[16] While NLSY79 and SCF mean wealth values are much higher than PSID figures, all the three surveys provide roughly comparable median values.

SIPP data provide another benchmark for comparing NLSY79 data. This survey is primarily designed to capture a household's income and participation in government assistance programs. Unfortunately, because of the sample design Eller and Fraser (1995) state that "the SIPP sample frame contains few observations for high income households." SIPP median 1993 wealth estimate for household heads less than 35 years old, deflated to 1990 dollars, is $5,255 (Eller and Fraser, 1995, Table D). This is far lower than the NLSY79 sample estimate of $21,616

Overall, this section suggests that NLSY79 data provide comparable results when compared to other data sources. While NLSY79 wealth data are in similar ranges with other U.S. micro-surveys, the actual net wealth values are greater. This confirms the upward bias calculated earlier in this research.

VI. Conclusions

Prior to this research, little was known about the wealth of young baby boomers, individuals born from 1957 to 1964. Baby boomers are important because they provide a picture of how wealth accumulation begins. Additionally,

[16]Extracting PSID wealth data so that the age range matches the NLSY79 results in a mean value of $27,892 and a median of $9,212 in 1990 dollars.

they are the primary group affected by funding crises in retirement programs such as Social Security. Hence, analysis of their wealth holdings helps policy makers to understand the impact of changing these programs.

This paper begins with a complete data quality investigation of the NLSY79, which shows that using NLSY79 wealth data from the public use CD-ROM without cleaning produces nonsensical results. Particularly in 1989, a small number of very large outliers of questionable veracity cause wild swings in wealth. To eliminate the problem, this research recommends that all out-of-range asset values be treated as mistakes and eliminated when using the data.

The data quality analysis also shows an increasing number of respondents have been top-coded, resulting in a growing loss of usable data over time. Further, response rates reveal that most missing wealth answers result from respondents not knowing the answer. As poorer individuals have left the NLSY79 survey over time, the remaining respondent pool has developed a slight upward wealth bias. Finally, the NLSY79 respondents appear to hold more assets than respondents in other U.S. data sets, possibly reflecting this upward bias plus respondent–interviewer interaction.

A cleaned net asset series was then created. According to this series, the percentage of young baby boomers having either no wealth or negative wealth has fallen from 20 percent to 13 percent since 1985. Nevertheless, calculations suggest some individuals will have almost no savings at retirement. Examining wealth by age shows the typical young baby boomer accumulates more than $2,000 of wealth each year. Overall, females appear to hold slightly more wealth than males. Finally, as young baby boomers age they shift assets away from illiquid holdings, like automobiles, and into their primary residence.

The net asset series created and analyzed in this research is useful in answering other wealth research questions. For example, comparing the net asset series with net assets held by cohorts of other ages will reveal if uncertainty over the future of government public retirement programs is causing young baby boomers to alter their wealth accumulation patterns. No matter what the answer, the wealth holdings of the typical young baby boomer are currently growing steadily.

References

Ando, A. and F. Modigliani, The "Life Cycle" Hypothesis of Saving: Aggregate Implications and Tests, *American Economic Review*, 53, 55–84, March 1963.

Camphuis, H., *Checking, Editing and Imputation of Wealth Data of the Netherlands Socio-Economic Panel for the Period '87–'89*, VSB-CentER Savings Project Progress Report 10, Tilburg University, 1993.

Eller, T. J. and W. Fraser, *Asset Ownership of Households: 1993*, U.S. Bureau of the Census, Current Population Reports, P70–47 Washington, DC, 1995,

Ferber, R., *Collecting Financial Data by Consumer Panel Techniques: A Pilot Study*, Bureau of Economic and Business Research, University of Illinois, Urbana, 1959.

Haurin, D. R., P. H. Hendershott, and S. M. Wachter, Wealth Accumulation and Housing Choices of Young Households: An Exploratory Investigation, *Journal of Housing Research*, 7(1), 33–57, 1996.

Hurst, E., M. C. Luoh, and F. P. Stafford, Wealth Dynamics of American Families, 1984–1994, *Brookings Papers on Economic Activity*, Forthcoming 1998.

Juster, F. T. and K. A. Kuester, Difference in the Measurement of Wealth, Wealth Inequality and Wealth Composition Obtained from Alternative U.S. Wealth Surveys, *Review of Income and Wealth*, Series 37, Number 1, 33–62, March 1991.

Kennickell, A. B. and M. Starr-McCluer, Changes in Family Finances from 1989 to 1992: Evidence from the Survey of Consumer Finances, *Federal Reserve Bulletin*, 861–82, October 1994.

Little, R. J. A. and D. B. Rubin, *Statistical Analysis with Missing Data*, John Wiley & Sons, New York, 1987.

Marquis, K. H., *The SIPP Measurement Quality Experiment and Beyond: Basic Results and Implementation*, paper presented at the Census Bureau 1995 Annual Research Conference, 1995.

Ross, C. E. and J. R. Reynolds, The Effects of Power, Knowledge, and Trust on Income Disclosure in Surveys, *Social Science Quarterly*, 77(4), 899–911, 1996.

Wolff, E. N., *The Size Distribution of Wealth in the United States: A Comparison Among Recent Household Surveys*, New York University, Department of Economics, Working Paper.

———, Trends in Household Wealth in the United States, 1962–83 and 1983–89, *Review of Income and Wealth*, Series 40, Number 2, 143–74, June 1994.

Zagorsky, J. L. *Young Baby Boomer's Wealth*. The Ohio State University, Center For Human Resource Research, Working Paper.