Volume Title: Inquiries in the Economics of Aging

Volume Author/Editor: David A. Wise, editor

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-90303-6

Volume URL: http://www.nber.org/books/wise98-2

Publication Date: January 1998

Chapter Title: Household Wealth of the Elderly under Alternative Imputation Procedures

Chapter Author: Hilary W. Hoynes, Michael D. Hurd, Harish Chand

Chapter URL: http://www.nber.org/chapters/c7088

Chapter pages in book: (p. 229 - 257)

8 Household Wealth of the Elderly under Alternative Imputation Procedures

Hilary Hoynes, Michael Hurd, and Harish Chand

8.1 Introduction

Although many reach retirement with few resources except housing equity and a claim to social security and Medicare, financial wealth, nonetheless, makes an important contribution to the economic status of many of the elderly. Most of our up-to-date information about the wealth of the elderly is based on the Survey of Income and Program Participation (SIPP), which sometimes adds an asset module to its core survey. As in many surveys of assets, the rate of missing data on individual asset items is high, about 30 to 40 percent among those with the asset. This raises the issue of the reliability of SIPP wealth measures because respondents who refuse or are unable to give a value to an asset item may not be representative of the population. Indeed, in the Health and Retirement Survey (HRS) it is clear that asset data are not missing at random. Through the use of bracketing methods, which we will discuss below, the HRS was able to reduce the rate of missing asset data substantially, and the data that were added in this way increased mean wealth in the HRS by about 40 percent (Smith 1995). Furthermore, because the additional data increased the mean so much, they undoubtedly increased measures of wealth inequality.

Because of the extensive use of bracketing to reduce the rate of nonresponse to asset items and because of its large sample size, the Asset and Health Dynamics among the Oldest Old (AHEAD) survey is likely to produce better

Hilary Hoynes is assistant professor at the University of California, Berkeley, and a faculty research fellow of the National Bureau of Economic Research. Michael Hurd is professor at the State University of New York at Stony Brook, a researcher at RAND, and a research associate of the National Bureau of Economic Research. Harish Chand is a student at the University of California, Berkeley.

The authors thank Li Gan, Dan McFadden, Jim Smith, and seminar participants for helpful suggestions. Financial support from the National Institute on Aging is gratefully acknowledged. Computing support was provided by the Econometrics Laboratory at UC Berkeley.

estimates of the distribution of the wealth of the elderly than other data sets. Even with bracketing, however, imputation of amounts is required, and the imputation method may well influence both the level and distribution of total wealth. In this paper we report the effects of a number of imputation methods on components of wealth and total wealth. In particular we extend the imputation techniques of Chand and Gan (1994) and of Smith (1995). Our methods will preserve covariation among measures of economic status to a greater extent than the previous methods, and this should provide a more accurate description of the wealth holdings and degree of wealth inequality of the elderly.

The paper is organized as follows. Section 8.2 describes the AHEAD data, with particular attention to the use of bracketing to reduce the incidence of missing data. Section 8.3 describes our imputation methodology. Section 8.4 presents the results of the imputation process for selected asset groups. Section 8.5 presents estimates of the distribution of imputed wealth. Section 8.6 concludes.

8.2 The AHEAD Data

Our data come from the survey of the Asset and Health Dynamics among the Oldest Old (AHEAD). This is a biennial panel of individuals born in 1923 or earlier and their spouses. The panel data set began in 1993 with a survey of 8,222 individuals representative of the community-based population except for the oversampling of blacks, Hispanics, and Floridians. The response rate in this first year of the survey was 80.6 percent. The second wave of the panel was fielded in October 1995. The results of this paper are based on the first wave of the panel.

The main goal of AHEAD is to provide panel data from the three broad domains of economic status, health, and family connections so that their coevolution can be studied. At baseline the survey elicited information about demographics, health, cognition, family structure and connections, health care and costs, housing, job status and history, expectations, income, and assets and insurance (Soldo et al. 1997). We are particularly interested in the data on asset holdings, which we will discuss in detail below.

AHEAD contains considerable detail about income and work history. Among the income components are social security benefits, pensions and annuities, asset income (with disaggregation as to type), earnings, and other transfer income such as supplemental security income. Measured income in AHEAD has been found to aggregate to the levels that are found in Current Population Survey data (Soldo et al. 1997).

Health in AHEAD is measured in a number of ways such as the ability to perform tasks, limitations on activities of daily living and instrumental activities of daily living, disease conditions and severity, and by self-assessment. AHEAD measures cognitive status in a battery of questions that aim to test a number of domains of cognition (Herzog and Wallace 1997).

8.2.1 Estimation Data Set

The AHEAD sample consists of 6,052 households and 8,222 individuals. In a husband-wife household information on income, assets, and insurance were asked only of the financial respondent, that person said to be the most knowledgeable about the household's finances. The husband was the financial respondent in 59 percent of the couple households. A few households did not complete the asset module of the survey, which reduced our sample to 5,973 households. About 38 percent of the households are married couples, 13 percent single men, and 49 percent single women.

Table 8.1 has the mean values of selected categorical variables for the estimation sample. Unless otherwise indicated, all variables correspond to the characteristics of the financial respondent. About a third of the sample has heads between ages 70 and 74. Whites account for 81 percent of the households. Widows and widowers account for almost 50 percent of the observations. About 27 percent of heads are college graduates, and 32 percent have completed high school. We use self-assessed health status for the head and spouse as an overall summary measure of the health of the household. We use a summary measure of cognitive ability to generate an indicator that cognitive performance is in the lowest third of the distribution. We imagine that low cognitive functioning will be reflected in a diminished ability to give informed answers to questions about income and assets.

8.2.2 Wealth Data in AHEAD

The AHEAD data contain information on household debt and 10 types of household assets: checking and savings accounts, CDs, stocks, bonds, individual retirement (IRA) and Keogh accounts, housing, transportation, other real estate, business equity, and other assets.

It is quite common in household surveys that the response to questions about asset value is "don't know" (DK) or "refused" (RF). For example, in the SIPP the rate of missing values among owners is 30 to 40 percent on asset values.[1] These missing values are usually imputed from a model of asset holdings that is fitted over observed values. The HRS and AHEAD use bracketing methods to reduce the rate of missing data. In a typical sequence a respondent would be asked about, for example, stock ownership and, if an owner, the value of stock holdings. A follow-up to a DK or RF about the value of stock holdings is "Would it amount to $25,000 or more?" If the response to that question is yes, the follow-up is "Would it amount to $100,000 or more?" but if the answer is no, the follow-up is "Would it amount to $5,000 or more?" By this sequence, stock holdings were assigned to one of five intervals. Other assets were bracketed in a similar way except that the bracket intervals differed by asset type because of differences in the distributions of each asset in the population.

1. See table 8.4, which we will discuss below.

Table 8.1 **Means of Covariates in Estimation Sample (N = 5,973)**

Dummy Variable	Definition	Mean
Age of head	70–74	0.34
	75–79	0.27
	80–84	0.19
	85–89	0.09
	90 or over	0.04
Gender	Female	0.64
Race/ethnicity	Nonwhite	0.14
	Hispanic	0.05
Marital status	Divorced/separated	0.06
	Widowed	0.50
	Never married	0.04
	Married	0.39
Education of head	College graduate	0.27
	High school graduate	0.32
Education of spouse	College graduate	0.11
	High school graduate	0.15
Occupation of head	Professional/managerial	0.13
Occupation of spouse	Professional/managerial	0.05
Occupation of former spouse	Professional/managerial	0.04
Work history of head	Worked 10–20 years	0.52
	Worked 20–30 years	0.18
	Worked 30 or more years	0.23
Work history of spouse	Worked 10–20 years	0.20
	Worked 20–30 years	0.07
	Worked 30 or more years	0.08
Work history of former spouse	Worked 10–20 years	0.45
	Worked 20–30 years	0.01
	Worked 30 or more years	0.03
Cognition/proxy interview	Low cognitive score	0.33
	Missing cognition score	0.03
	Proxy interview	0.07
Health status of head	Excellent	0.11
	Very good	0.23
	Good	0.31
	Fair	0.23
	Poor	0.12
Change in health status (2 yr)	Better	0.13
	Same	0.65
	Worse	0.22
Health status of spouse (0 if no spouse)	Excellent	0.05
	Very good	0.09
	Good	0.12
	Fair	0.09
	Poor	0.05
Change in health status (2 yr)	Better	0.05
	Same	0.27
	Worse	0.08
Income receipt indicator	Veteran's benefits	0.06
	Pension	0.44
	Annuities	0.06

Table 8.1 (continued)

Dummy Variable	Definition	Mean
	IRA	0.09
	Stock	0.16
	Earnings	0.01
	Other	0.10
	Savings account	0.32
	Rental property	0.07
	Investment/trusts	0.01
	Relatives	0.00
High school education of parents	Head's mother	0.47
	Head's father	0.44
	Spouse's mother	0.20
	Spouse's father	0.18
Housing information	Low-income housing	0.04
	Duplex	0.20
	Mobile home	0.03
	Apartment	0.00
	Townhouse	0.01
	Other housing	0.09
Condition of dwelling	Excellent	0.23
	Very good	0.33
	Good	0.29
	Poor	0.11
Safety of neighborhood	Excellent	0.26
	Very good	0.32
	Good	0.28
	Poor	0.09
Likelihood of leaving bequest	≥50 percent	0.24

Source: Authors' calculations from AHEAD. Unless otherwise stated, all calculations are weighted.

The importance of bracketing comes from the highly skewed distribution of many types of assets: knowing that an individual has stock holdings of, say, $5,000 to $25,000 provides much better information about the stock holdings of that individual than could be found from imputing stock holdings because the covariates used in the imputation have rather low explanatory power.

Table 8.2 summarizes the response status of families in AHEAD by type of asset. Some respondents either refuse to say or do not know whether they have a particular asset, resulting in missing data on asset ownership. A general conclusion is that the great majority of respondents can and do say whether they own a particular asset. Missing data on ownership averages only about 0.5 to 3 percent of the sample. The rate of ownership varies greatly by asset type: roughly three-quarters have a checking or savings account, just 5.7 percent own bonds, 19.4 percent own common stock, and the rate of home ownership is around 71 percent. These asset ownership rates are comparable to those found for the elderly in other data sets such as the SIPP.

Table 8.2 Asset Ownership

A. Distribution of Households (percent)

		Ownership Reported		
Type of Asset	Ownership Missing	Not Owner	Owner	All
Checking and savings	2.2	24.4	75.6	100
CDs	2.8	78.7	21.3	100
Stocks	1.9	80.6	19.4	100
Bonds	2.3	94.3	5.7	100
IRA/Keogh	1.2	83.8	16.2	100
Housing	0.5	29.1	70.9	100
Other real estate	1.1	81.0	19.0	100
Business	0.4	95.6	4.4	100
Other assets	1.7	89.8	10.2	100
Debts	1.3	86.0	14.0	100

B. Distribution of Owners (percent)

Type of Asset	Continuous Value	Fully Bracketed[a]	Incomplete Bracket[b]	No Bracket[c]	All
Checking and savings	67.4	22.4	2.3	7.8	100
CDs	60.8	23.3	3.3	12.6	100
Stocks	53.8	31.7	2.6	12.0	100
Bonds	57.2	25.4	3.9	13.5	100
IRA/Keogh	72.6	16.4	1.0	10.0	100
Housing	77.7	19.0	1.2	2.1	100
Other real estate	65.9	22.9	1.5	9.7	100
Business	55.3	32.3	0.9	11.5	100
Other assets	69.2	21.3	0.5	9.1	100
Debts	83.3	10.1	0.5	6.1	100

Source: Authors' calculations from AHEAD. Unless otherwise stated, all calculations are weighted. Households missing data on entire wealth section are dropped.

[a]Asset value is within some bracketed range.

[b]Individuals did not complete bracketing sequence, but partial information is available.

[c]Ownership known, but in response to bracketing questions for value of asset, individual either refused to answer (RF) or did not know (DK).

Respondents who indicated ownership of a particular asset can be divided into four groups depending on their responses to follow-up questions. The bottom panel of table 8.2 shows the distribution of owners of each asset. Beginning from the leftmost column, we have what we call "continuous" values: these come from respondents who stated an actual dollar value for the amount of an asset. Thus, about 67 percent of owners of checking or savings accounts reported a dollar amount. "Fully bracketed" are those respondents who completed the sequence of bracketing questions: 22.4 percent in the case of checking or savings accounts. A few respondents gave some bracketing information but did not complete the sequence. These "incomplete bracket" respondents

Table 8.3 **Missing Data Rates by Type of Asset**

Type of Asset	Percentage of Owners with Missing Data: Using Brackets[a]	Percentage of Owners with Missing Data: Not Using Brackets[b]	Overall Missing Data Rates: Using Brackets[c]	Overall Missing Data Rates: Not Using Brackets[d]
Checking and savings	7.8	32.5	8.1	26.8
CDs	12.6	39.2	5.5	11.1
Stocks	12.0	46.3	4.2	10.9
Bonds	13.5	42.8	3.1	4.7
IRA/Keogh	10.0	27.4	2.8	5.6
Housing	2.1	22.3	2.0	16.3
Other real estate	9.7	34.1	2.9	7.6
Business	11.5	44.7	0.9	2.4
Other assets	9.1	30.9	2.6	4.9
Debts	6.1	16.7	2.2	3.6

Source: Authors' calculations from AHEAD. Unless otherwise stated, all calculations are weighted.

[a]Includes observations with no brackets.

[b]Includes observations with incomplete, complete, or no brackets.

[c]Includes observations missing ownership or with no brackets.

[d]Includes observations missing ownership or with incomplete, complete, or no brackets.

(2.3 percent for checking and savings) answered at least the first of the bracketing questions but answered either with a RF or DK on one of the follow-up bracketing questions.[2] The last group, "no bracket," gave no bracketing information at all, answering either RF or DK to the first of the bracket questions. The table shows that there is a great deal of variation in the responses by asset type. Individuals are more likely to give continuous responses to questions about the value of housing and debts and less likely about the value of stocks, bonds, and business assets.

The use of bracketing substantially decreases the rate of missing data. Table 8.3 summarizes the importance of the bracketing questions by comparing the missing data rate in AHEAD with the missing data rate that would result if no bracketing information were used. The first two columns present missing data rates among owners, while the second two columns give overall missing data rates, including missing data on ownership.[3] Without using brackets, as shown in the second column, the rate of missing data among owners would have been

2. E.g., when asked whether the value of the asset was greater or less than $25,000 the respondent said greater. When asked whether the amount was greater or less than $100,000 the respondent answered with DK or RF. This can result in an open or closed interval.

3. For the column labeled "using brackets," missing data among owners consist of those observations without any bracketing information. Ignoring the bracketing questions, the missing data would also include those with incomplete and complete brackets. The overall missing data rates multiply the missing data rate among owners by the ownership rate and add to that the rate of missing data on ownership.

17 to 46 percent. This is reduced to 2 to 14 percent by using bracketing. For example, among owners of common stock, the rate of missing data is reduced from 46 percent to 12 percent because of the bracketing questions. A particularly important example is housing because of its importance in the portfolios of the elderly: the rate of missing data among owners was reduced from 22.3 percent to 2.1 percent by bracketing.

Table 8.3 shows that there is a great deal of variation in the overall missing data rates across asset types, with checking and savings accounts having relatively high missing data rates while housing has relatively low missing data rates. The low rate of missing housing values is especially notable because of the very high ownership rate.[4]

The missing data rates compare favorably to the rates in the SIPP. Table 8.4 shows rates from AHEAD both with and without bracketing information and from the SIPP by age and martial status. The initial rate of nonresponse is about the same, as seen by comparing the SIPP with the AHEAD "not using brackets." For example, for checking accounts among older singles, 38 percent in the SIPP and 34 percent in AHEAD gave an initial nonresponse as to amount. But bracketing in AHEAD reduced this to 7.9 percent. For stock holdings among singles, 66 percent in the SIPP and 52 percent in AHEAD gave an initial nonresponse as to value, but in AHEAD bracketing reduced this to 14 percent. There was a similar reduction in nonresponse among couples in AHEAD from bracketing. We conclude that even though the AHEAD population is quite elderly the use of bracketing reduces item nonresponse to rather low levels.

As shown in table 8.5 the likelihood of asset ownership and of item nonresponse varies with personal characteristics. Those who report owning assets have lower rates of cognitive impairment, are younger, and are more likely to be married. The table shows that the two types of nonresponse correspond to individuals with different characteristics on average. Those who respond DK are more likely to have higher levels of cognitive impairment, are less likely to be married, and are more likely to be over age 80 than those who respond RF.[5] Those who respond with continuous values are younger and have lower levels of cognitive impairment than either kind of nonrespondent. These simple tabulations suggest that the different forms of response display fairly distinct patterns, which will be potentially useful in a model-based imputation procedure. These characteristics also suggest that the option of providing brackets does not crowd out more accurate responses from an able population but rather allows information to be obtained from those unsure about their holdings.

4. Of course, knowing an interval for an asset value is not the same as knowing the exact amount, but even continuous reports are not exact amounts. Indeed, a large percentage of continuous reports tend to give a "focal" point answer, suggesting that a substantial amount of rounding occurs even in continuous responses (Chand and Gan 1994).

5. The DK and RF refer to the initial response to a question about asset value.

Table 8.4 **Comparison of Missing Data Rates in SIPP and AHEAD Surveys (percent of owners with missing data)**

A. 1993 SIPP

	All Persons (16+)		Persons Aged 60–69		Persons Aged 70+	
Type of Asset	Married	Single	Married	Single	Married	Single
N	22,491	30,463	2,897	1,152	2,127	2,172
Checking (own)	37.6	30.0	40.3	29.9	47.3	38.1
Checking (joint)	30.8	–	37.6	–	40.4	–
Stocks (own)	35.6	48.4	33.3	60.7	47.1	65.9
Stocks (joint)	42.7	–	50.7	–	50.0	–
IRA	32.4	30.8	35.0	30.1	41.8	31.3
Keogh	46.1	47.3	46.4	27.8	48.6	42.4
Housing	24.2	29.0	28.3	30.0	27.8	35.1

B. AHEAD

	Married		Single	
Type of Asset	Using Brackets[a]	Not Using Brackets[b]	Using Brackets[a]	Not Using Brackets[b]
Checking and savings	7.7	30.8	7.9	33.9
CDs	12.4	38.2	12.8	40.1
Stocks	10.7	41.7	13.6	52.0
Bonds	10.9	33.7	16.8	54.7
IRA/Keogh	9.4	25.9	11.6	31.0
Housing	1.2	14.9	2.9	29.1
Other real estate	8.7	29.4	11.1	40.3
Business	9.7	41.6	15.6	51.7
Other assets	7.8	27.3	10.7	35.4
Debts	6.3	15.4	6.0	17.8

Source: Authors' calculations from AHEAD and SIPP. Unless otherwise stated, all calculations are weighted.

[a]Includes observations with no brackets.

[b]Includes observations with incomplete, complete, or no brackets.

8.3 Wealth Imputation in AHEAD

For those who provide a complete bracket, only an amount within a bracket will need to be imputed. Individuals who do not report whether they own the asset will potentially require ownership, then bracket, and finally amount to be imputed. Because of the relationships between personal characteristics and wealth item nonresponse (table 8.5), the imputations will use covariates. The descriptive tables suggest that the determinants of nonresponse differ between DK and RF. Therefore, whenever possible, we will differentiate between these two sources of nonresponse.

Table 8.5 Personal Characteristics by Asset Ownership Status

Type of Asset and Response	N[a]	Low Cognitive Score[b] (%)	Proxy Interview (%)	Married (%)	Over Age 80 (%)
Checking and savings					
Nonowner	1,568	0.58	0.12	0.31	0.38
Missing ownership	133	0.47	0.12	0.33	0.38
DK value	875	0.40	0.07	0.36	0.35
RF value	507	0.29	0.06	0.48	0.31
Gave value	2,890	0.25	0.04	0.44	0.29
Stocks					
Nonowner	4,796	0.41	0.08	0.35	0.34
Missing ownership	113	0.45	0.16	0.41	0.39
DK value	368	0.21	0.03	0.47	0.29
RF value	117	0.18	0.03	0.62	0.26
Gave value	579	0.09	0.01	0.61	0.21
Housing					
Nonowner	1,797	0.51	0.13	0.17	0.45
Missing ownership	30	0.52	0.91	0.79	0.40
DK value	911	0.50	0.05	0.32	0.39
RF value	45	0.47	0.02	0.46	0.30
Gave value	3,190	0.24	0.03	0.53	0.24

Source: Authors' calculations from AHEAD. Unless otherwise stated, all calculations are weighted.

[a]These are unweighted observation counts and do not match the weighted percent distribution of observations that are provided in table 8.2.

[b]Includes also those who do not complete cognition battery.

8.3.1 Imputation of Ownership

As was shown in table 8.2, a small percentage of people did not give information about asset ownership. For these people, we imputed ownership based on logistic estimation. Using the sample of those whose ownership status is known, we estimated $P(O) = L(X'\beta)$, the probability of ownership (O) given observations on the covariates X, which include demographic variables (age, race, marital status), education, work history, profession, cognitive impairment indicators, and reported sources of income, and the logistic function L. Then, for someone whose ownership status is unknown, we imputed ownership based on the estimated probability $P^* = L(X'\beta^*)$ by making a random drawing from a binomial distribution with a probability of success of P^*. Covariates will increase the precision of the imputation because of the variation in ownership by personal characteristics (table 8.5).

8.3.2 Imputation of Brackets

After imputing ownership, we allocate the imputed owners and those with no brackets or incomplete brackets to one of the complete brackets. This was done with ordered logistic estimation. Among those with complete brackets we estimated $P_j(X'\alpha)$, the probability of being in the jth bracket. The covariates, X, include the demographic and other variables used in ownership imputation, supplemented by ownership of other assets and brackets of other assets. Then, for someone with missing bracket information, we imputed a complete bracket based on the fitted probabilities $P_j(X'\alpha^*)$, by making a random assignment according to a drawing on a multinomial random variable with probabilities $P_j(X'\alpha^*)$.

8.3.3 Imputation of Amounts

The final step in the imputation is to assign values to all those who either report a complete bracket or who have been imputed into a bracket. Amounts are imputed through a nearest neighbor approach similar to that in Chand and Gan (1994) and Little, Sande, and Scheuren (1988). For each individual to be imputed, a nearest neighbor is selected from among the continuous reporters who are in the same bracket. The selection is based on a regression of asset amount on individual characteristics. First, we fit over the continuous reporters in bracket j, $S = X'\gamma_j$ where S is the value of the asset for those in bracket j. Then S is predicted over all continuous and bracketed reporters in bracket j using the estimated value of γ_j. For each individual to be imputed from bracket j, the nearest neighbor is that continuous reporter in bracket j whose fitted value is closest to the fitted value of that individual. The value assigned is the actual value of the nearest neighbor, not the fitted value. For this imputation step we use the same covariates as in the imputation to the brackets.

This method is a generalization of the "hotdeck" procedure in which a few characteristics such as education and sex are used to stratify the sample of continuous reporters. Then an imputation for an individual with a missing value is made at random from the cell corresponding to that person's characteristics. If we consider a bracket to be a characteristic, our method is hotdeck with complete stratification by bracket and partial stratification by other characteristics. The advantage of our method is that we can use many more covariates than in a traditional hotdeck, which is limited because of empty cells. Our method has the further advantage of preserving the covariances between the asset value and our covariates within the limits of the functional form $X'\gamma$.

This imputation method contains several differences from that of Chand and Gan (1994), who also use a nearest neighbor approach to impute asset amount. First, whereas the Chand-Gan approach uses the nearest neighbor metric to assign amounts, our method breaks the imputation of bracket and amount into two distinct steps. Second, we impute brackets based on only those observations who provide a complete bracket, while Chand and Gan include observations who provided a continuous amount. We consider this an improvement because in the HRS the distribution of households across brackets is different for the continuous respondents than for the bracketed respondents (Smith 1995). We believe that the respondents who did not complete a bracketing sequence are more like those who were bracketed than those who gave a continuous amount. Third, those who gave no bracketing information at all are imputed to a bracket based on the distribution of those who completed the bracketing and who initially gave the same type of response (DK or RF). This procedure is based on the observation that DK or RF responses have different distributions across asset brackets, with refusers typically falling into the higher brackets. Finally, in contrast to Chand and Gan (1994), greater use of financial information was made in the imputation of asset brackets and amounts. Dummy variables for the ownership of other assets and the brackets of total income and other assets were used as additional covariates to preserve some of the interasset structure of wealth in the imputations.

8.3.4 Implementing Imputation Procedure

Since there are 10 components of wealth, each having either four or five brackets, we use stepwise model selection to choose the explanatory variables (the X) for ownership probability, bracket probability, and asset level within a bracket. We experimented with several significance levels for entering a variable into the statistical model. This is discussed briefly in the appendix, where we give tables with descriptive information on the characteristics of the imputations of two representative assets, stocks and housing, at three different significance levels.

The nearest neighbor approach, in common with all hotdeck procedures, has a stochastic component, which could cause random variation in asset values. For example, in the top bracket, which is open ended, selecting several times the highest observed continuous asset amount would affect the mean of the distribution substantially. To reduce the influence of this stochastic component, the entire imputation procedure was repeated several times. The models without covariates, which exhibit the highest amount of stochastic variation, were repeated nine times, and the models with covariates were repeated four times. In each case, the imputed amount was assigned to be the average across the repetitions.

8.4 Results

8.4.1 Ownership Imputation

Table 8.6 shows the results from imputing ownership. The first column of the table shows the asset ownership rates for those who report ownership. The second column gives the imputed ownership rates for those whose ownership rate is unknown. With the exception of bonds, the rate of ownership is lower where ownership is imputed. With stock ownership, for example, 14.7 percent of those with missing ownership are imputed to own stocks, compared to 19.4 percent among those who report ownership. Note that if covariates were not used in the imputation, the rate of ownership would be, on average, the same in both columns. Lower ownership rates among those with missing data occur

Table 8.6 **Percentage of Households Owning Assets: Actual and Imputed**

Type of Asset	Actual	Imputed[a]
Checking and savings	75.6	73.3
CDs	21.3	15.8
Stocks	19.4	14.7
Bonds	5.7	6.6
IRA/Keogh	16.2	7.4
Housing	70.9	66.3
Other real estate	19.0	12.6
Business	4.4	3.7
Other assets	10.2	6.8
Debts	14.0	19.8

Source: Authors' calculations from AHEAD. All calculations are weighted.

[a]Ownership imputed by using estimates from ownership regression using the sample of those with ownership known. See text for details.

because those who do not report ownership have characteristics that tend to be similar to those of nonowners, such as older age and higher rates of cognitive impairment. This was found earlier in the descriptive analysis in section 8.2. This suggests that ownership nonresponse tends to occur more for reasons of informational uncertainty than for privacy concerns.[6]

8.4.2 Bracket Imputation

Because of the large number of assets to be imputed, we will concentrate the discussion of the results of the imputations on two important and very different assets: stocks and housing wealth. Stocks are illustrative of assets with low ownership and high missing value rates but exhibit a very skewed distribution (large upper tail). Housing is important because it comprises a large proportion of individual wealth holdings.

Table 8.7 shows the effects of bracket imputations for stocks. Each column of the table gives the percentage distribution of observations across the five stock brackets. The first column reports the percentage distribution among those giving continuous responses; the next two columns give the distributions for those with complete brackets and those with imputed brackets.[7] Those completing the bracketing sequence tend to have higher stock values than those providing continuous responses. For example, 18 percent of respondents who gave continuous amounts have from zero to $4,999 in stock equity compared with just 14 percent of those who completed the bracketing sequence. The effect of using covariates in the imputation process can be seen by comparing

6. This makes the AHEAD population different from the HRS population, where nonresponse on assets is typically associated with an unwillingness to reveal large amounts: the imputations increase ownership rates substantially in HRS.

7. Observations requiring imputation of brackets include those missing ownership and those with incomplete brackets.

Table 8.7 **Distribution of Owners of Stocks (percent)**

		Brackets		
Range (thousand $)	Continuous	Complete	Imputed[a]	All
0.0–4.9	18	14	11	16
5.0–24.9	24	25	23	25
25.0–99.9	28	37	29	29
100.0–499.9	25	19	30	23
500+	4	6	6	5
All	100	100	100	100

Source: Authors' calculations from AHEAD. Unless otherwise stated, all calculations are weighted. Columns may not sum to 100 percent because of rounding.

[a]Includes observations with no brackets and incomplete brackets.

Table 8.8 **Distribution of Owners of Housing (percent)**

		Brackets		
Range (thousand $)	Continuous	Complete	Imputed[a]	All
0.0–49.9	26	42	46	30
50.0–99.9	39	35	33	38
100.0–199.9	25	16	15	23
200.0+	10	7	6	9
All	100	100	100	100

Source: Authors' calculations from AHEAD. Unless otherwise stated, all calculations are weighted. Columns may not sum to 100 percent because of rounding.

[a]Includes observations with no brackets and incomplete brackets.

the distribution of the complete brackets with the imputed brackets: since the imputation is based on the sample of complete brackets, if no covariates were used, on average they would be the same. We see that the covariates shift the distribution to higher values, implying that those who give incomplete responses have greater socioeconomic status: we estimate that 14 percent of those who completed the bracketing sequence have stock holdings between zero and $4,999 compared with just 11 percent of those with imputed brackets.

Table 8.8 shows the results of imputing housing brackets. This table shows that, in contrast to results for stocks (and most other assets), those with incomplete responses on housing have personal characteristics that make them more likely to have low housing values. Continuous reports are systematically greater than the bracket reports, and the covariates used to impute brackets reduce the bracketed distribution even further. The differences are large: 35 percent of the continuous reports have housing equity of $100,000 or more, whereas just 21 percent of the imputed bracket cases are in that range.

Table 8.9 **Average Stock Wealth within Brackets (thousand dollars)**

Range (thousand $)	Continuous	Brackets
0.0–4.9	1.2	1.2
5.0–24.9	13.5	13.9
25.0–99.9	49.3	50.0
100.0–499.9	185.6	190.7
500+	862.3	751.0

Source: Authors' calculations from AHEAD. Unless otherwise stated, all calculations are weighted.

Table 8.10 **Average Stock Holdings by Type of Observation**

Type of Observation	Number of Observations	Average Value (thousand $)
Continuous	579	97.9
Complete bracket	359	90.2
Incomplete bracket	126	148.8
No bracket	16	163.6

8.4.3 Imputing Amounts

Table 8.9 shows the imputations of amounts of stock holdings within each bracket. The first column gives the average value for stock wealth among all households providing continuous responses in the given range. The second column gives the average imputed value within a bracket. Although the differences are not large, the average amount within the bracketed range tends to be higher than the average continuous amount. For example, within the $100,000–$499,999 range, the average over continuous reporters is about $185,600, which is what the average imputed amount would be if no covariates were used. Among those with brackets, the covariates increase the average amount to about $190,700. This implies that imputed individuals have covariates that are associated with higher stock holdings than those of continuous reporters.

However, this table does not show differences across the subgroups of imputed observations. Those requiring imputation of values within brackets include those who have incomplete brackets, those with complete brackets, and those with no brackets. These groups appear to be very different. On average, those with no brackets have covariates associated with higher levels of stocks relative to those with continuous values, while those with complete brackets have lower values. This can be seen from average stock values for these groups in table 8.10. The figures in that table reflect differences in the distribution of observations across brackets as well as differences in average values within brackets.

Table 8.11 shows the results of imputing the value of housing wealth within

Table 8.11 Average Housing Wealth within Brackets (thousand dollars)

Range (thousand $)	Continuous	Brackets
0.0–49.9	26.7	26.2
50.0–99.9	64.3	66.2
100.0–199.9	121.4	122.6
200.0+	332.8	291.8

Source: Authors' calculations from AHEAD. Unless otherwise stated, all calculations are weighted.

Table 8.12 Average Housing Wealth by Type of Observation

Type of Observation	Number of Observations	Average Value (thousand $)
Continuous	3,190	95.8
Complete bracket	816	75.4
Incomplete bracket	140	68.6
No bracket	23	68.8

Table 8.13 Mean Asset Values by Nonresponse Status, by Type of Asset

		Nonresponse Status	
Type of Asset	Continuous Amount Reported	Bracket Reported	Bracket Imputed
Checking and savings	21.9	22.2	22.6
CDs	42.4	33.6	44.5
Stocks	97.9	90.2	150.4
Bonds	62.4	88.6	116.7
IRA	48.3	63.5	61.4
Housing	95.8	75.4	68.6
Other real estate	117.0	158.7	126.1
Business	101.7	225.9	348.3
Other assets	28.4	29.6	32.0
Debt	6.0	4.7	13.8

Source: Authors' calculations from AHEAD. Unless otherwise stated, all calculations are weighted.

brackets. No pattern emerges here: the averages are about the same, indicating that there is little systematic difference in the covariates that explain housing value between the continuous respondents and the bracketed respondents.

Adding in the differences in the distribution across brackets changes these results substantially as shown in table 8.12. Respondents who provide brackets have lower housing wealth than respondents who give continuous amounts, in contrast to holdings of stocks.

Differences between stock and housing wealth illustrate one of the important findings from this study: there are differences in the character of nonresponse across asset types and nonresponse categories. This is shown in table 8.13, which has a summary description of the results of the bracket imputation. The average amounts for three different types of responses are shown. Because those who provide continuous amounts tend to fall into lower brackets than those who provide brackets, those who provide continuous amounts generally have the lowest average wealth components. The notable exception is housing wealth, where those who provide a continuous amount have the highest average housing value. However, the effect of covariates in imputing brackets varies considerably. Some assets have bracket imputation resulting in higher average amounts such as for stocks, bonds, and business, while other assets such as housing, real estate, and IRAs display the opposite tendency.

8.5 The Distribution of Wealth and the Importance of Bracketing and Imputation

The results presented in the previous section show that imputed wealth differs significantly by type of nonresponse. This suggests more generally that the use of brackets to reduce missing data may lead to significant changes in the estimates of the distribution of household wealth. Our imputation methodology stresses not only the importance of bracketing but also the importance of using covariates at each stage of estimation. To explore the importance of these issues, tables 8.14–8.16 show how the distributions of nonhousing, housing, and total wealth differ under progressively more complicated imputation methods. The different methods vary along two main dimensions: how the bracketing information is used and whether covariates are used in the imputation procedure.

In all three tables, we show the distribution of wealth under seven imputation procedures. The imputation method becomes increasingly complex with each successive row in the table. The following summarizes the methods:

1. Assign ownership by the probability of ownership among that population where ownership is known. Impute amounts from unconditional draws from the continuous amounts. No covariates or bracketing information is used. This is known as unconditional hotdeck.
2. Same as method 1 except impute amounts to those in the complete brackets from the continuous amounts within brackets.
3. Same as method 2 except impute incomplete brackets from pool of completed brackets.
4. Same as method 3 except impute incomplete brackets from pool of completed brackets who provided the same response (DK or RF) to the initial question about amount.
5. Same as method 4 except impute ownership using covariates.
6. Same as method 5 except impute brackets using covariates.
7. Same as method 6 except use covariates to find nearest neighbor for imputation of amount.

Table 8.14 Effects of Imputation on Distribution of Nonhousing Wealth (thousand dollars)

	Nonhousing Wealth			
Imputation Method[a]	Mean	10th Percentile	Median	90th Percentile
1. Unconditional hotdeck	94.7	0.0	24.0	251.5
2. Bracketed hotdeck	97.7	0.0	21.0	255.6
3. Imputing brackets without covariates	97.1	0.0	21.0	258.0
4. Imputing brackets without covariates, stratify by DK/RF	99.9	0.0	21.2	262.0
5. Adding covariates to ownership imputation	99.5	0.0	21.0	264.8
6. Adding covariates to bracket imputation	101.7	0.0	20.0	261.0
7. Adding covariates to level imputation	100.8	0.0	20.0	260.0

Source: Authors' calculations from AHEAD. All calculations are weighted.

[a]Imputation methods described in text. Each successive method nests the method before it. For example, the stratification by don't know (DK) vs. refused (RF) in method 4 is also used in methods 5–7.

Table 8.14 shows the effects of the different methods on estimated values for nonhousing wealth.[8] Going from method 1 to method 2 increases the mean of nonhousing wealth by about 3 percent, which is caused by the brackets. That is, simply knowing what bracket someone falls into increases the estimate of mean wealth. At the same time, the median is reduced, implying that the entire distribution is affected by the brackets. Further, because the 90th percentile increases only marginally, some of the influence on the mean must be coming from the very wealthy. Method 4 shows that differentiating between DK and RF is important, shifting up the distribution at all points. Methods 5 through 7, which vary primarily by the extent of the use of covariates, affect estimates of the distribution of wealth only minimally.

Table 8.15 shows the effects of the different methods on housing wealth averaged over both owners and nonowners. Here the bracketing and covariates all reduce the mean and median. The changes accumulate to be fairly large on the mean: the entry for method 7 is about 5 percent less than for method 1.

Table 8.16 has similar results for total wealth.

We found that the value of stock holdings differed if the response was DK rather than RF, which we attribute to differences in personal characteristics such as cognition. Thus we would expect that individuals answering DK about

8. Nonhousing wealth includes all categories of wealth except housing (checking, CDs, stocks, bonds, IRAs, other real estate, business, and other assets).

Table 8.15 **Effects of Imputation on Distribution of Housing Wealth (thousand dollars)**

Imputation Method[a]	Housing Wealth: Mean if Greater Than Zero	Mean	10th Percentile	Median	90th Percentile
1. Unconditional hotdeck	95.8	67.9	0.0	45.0	150.0
2. Bracketed hotdeck	92.8	65.7	0.0	40.0	150.0
3. Imputing brackets without covariates	92.9	65.6	0.0	40.0	150.0
4. Imputing brackets without covariates, stratify by DK/RF	93.4	65.3	0.0	40.0	150.0
5. Adding covariates to ownership imputation	92.5	65.3	0.0	40.0	150.0
6. Adding covariates to bracket imputation	93.0	65.3	0.0	40.0	150.0
7. Adding covariates to amount imputation	91.2	64.4	0.0	40.0	150.0

Source: Authors' calculations from AHEAD. All calculations are weighted.

[a]Imputation methods described in text. Each successive method nests the method before it. For example, the stratification by don't know (DK) vs. refused (RF) in method 4 is also used in methods 5–7.

Table 8.16 **Effects of Imputation on Distribution of Total Wealth (thousand dollars)**

Imputation Method[a]	Total Wealth: Mean	10th Percentile	Median	90th Percentile
1. Unconditional hotdeck	162.6	0.6	88.5	375.0
2. Conditional hotdeck	163.3	0.5	80.0	378.0
3. Imputing brackets without covariates	162.7	0.5	80.0	379.0
4. Imputing brackets without covariates, stratify by DK/RF	165.2	0.5	80.0	384.5
5. Adding covariates to ownership imputation	164.8	0.5	80.0	387.0
6. Adding covariates to bracket imputation	167.1	0.5	79.0	382.5
7. Adding covariates to level imputation	165.2	0.5	77.2	380.0

Source: Authors' calculations from AHEAD. All calculations are weighted.

[a]Imputation methods described in text. Each successive method nests the method before it. For example, the stratification by don't know (DK) vs. refused (RF) in method 4 is also used in methods 5–7.

Table 8.17 Effects of Imputation on Distribution of Wealth by Response Type (thousand dollars)

		Nonhousing Wealth		Housing Wealth		Total Wealth	
Imputation Method[a]	Response	Mean	Median	Mean	Median	Mean	Median
1. Unconditional hotdeck	DK	109.1	33.0	189.2	108.0	200.3	115.0
	RF	98.8	41.0	168.9	115.0	182.7	123.5
	Both	133.2	58.0	218.2	142.0	230.4	158.2
2. Bracketed hotdeck	DK	109.0	27.0	183.5	93.0	186.0	93.0
	RF	126.1	42.0	196.4	108.2	200.7	111.0
	Both	156.2	49.5	239.7	125.0	247.2	129.0
4. Imputing brackets without covariates, stratify by DK/RF	DK	110.6	27.0	184.6	92.0	187.2	91.0
	RF	132.6	43.0	202.7	110.0	233.8	107.2
	Both	182.6	59.5	267.8	139.0	285.2	140.0
6. Adding covariates to bracket imputation	DK	114.3	25.5	188.3	92.5	182.7	91.5
	RF	139.1	35.0	209.1	100.0	203.9	101.0
	Both	184.5	49.0	271.6	130.5	262.3	130.0
7. Adding covariates to level imputation	DK	110.9	25.3	183.3	90.6	183.3	90.6
	RF	135.8	35.0	205.3	100.0	205.3	100.0
	Both	193.1	45.2	275.7	128.2	275.7	128.2

Source: Authors' calculations from AHEAD. All calculations are weighted.

[a]Imputation methods described in text. Each successive method nests the method before it. For example, the stratification by don't know (DK) vs. refused (RF) in method 4 is also used in methods 6 and 7.

one type of asset would have differences in overall wealth from individuals who answer RF. Table 8.17 compares the distribution of wealth across selected imputation methods for the various types of nonresponse. For each method, households who require any imputation are divided into three categories: those who answered DK to at least one asset question, those who answered RF to at least one asset question, and those who answered DK to at least one and RF to at least one asset question. In the unconditional hotdeck method, the mean is lower for RF observations than for DK or both.[9] However, all other methods produce greater total wealth among the RF than among the DK. The main difference comes from using brackets, method 2. The implication is that the RF tend to be in higher brackets.

9. Observations with both an RF and a DK may have larger mean asset values since they, by definition, correspond to individuals who hold at least two assets.

Table 8.18 **Imputation Results by Nonresponse Type**

	Mean Amount (thousand $)		
Type of Asset	Amount Reported	DK	RF
Checking and savings	21.9	19.8	26.2
CDs	42.4	34.7	42.2
Stocks	97.9	81.5	183.9
Bonds	62.4	85.2	123.6
IRA	48.3	75.1	48.2
Housing	95.8	74.2	76.2
Other real estate	117.0	138.6	210.0
Business	101.7	226.8	340.4
Other assets	28.4	25.4	50.5
Debt	6.0	8.1	2.8

Source: Author's calculations. All calculations are weighted. Results are from the preferred imputation method (method 7).

As shown in table 8.18, the importance of differentiating across DK and RF responses holds in almost all asset types, with the exception of IRAs. Imputed wealth for those who refuse to answer the question about asset value is consistently higher than for those who respond that they do not know.

We summarize our results in table 8.19, which shows mean and median wealth by various personal characteristics. At the median the divorced or separated have the lowest wealth. Wealth declines sharply with age and with worse health. A low cognition score is associated with substantially lower wealth.

8.6 Conclusion

We have studied the effects of a number of imputation methods on aggregate measures of wealth such as the median, mean, and percentiles. There are many conclusions that emerge from this study. First, using bracketing in survey design can dramatically reduce the rate of missing data and increase the quality of asset data. Second, using covariates in the imputation process affects the distributions of individual asset holdings substantially. The net effects are minimal, however, in that aggregate wealth is not significantly affected by the introduction of covariates. An implication is that imputation based on covariates may provide an important gain in assigning assets at the individual level even though the effect on the population may not be large. Third, missing data can be the result of the respondent's not knowing (DK) or refusing to answer (RF). We find that these two groups are very different; DK respondents typically have characteristics like those with lower asset levels and RF respondents have characteristics like those with high asset levels. Differentiating between these two groups in the imputation process has important effects on the distribution of wealth.

Table 8.19 Mean and Median Wealth by Demographic Characteristics

		Nonhousing Wealth		Total Wealth	
Characteristic	*N*	Mean	Median	Mean	Median
Marital status					
Married	2,324	161.8	50.0	250.1	132.0
Divorced/separated	399	58.8	7.0	104.4	28.0
Widowed	3,015	58.1	10.3	107.5	52.5
Never married	235	84.7	12.1	124.8	50.8
Age					
70–74	1,975	133.6	32.5	208.4	98.5
75–79	1,567	101.2	18.3	171.1	80.0
80–84	1,217	69.1	12.7	120.0	60.8
85+	892	54.4	7.5	94.7	35.8
Health					
Excellent/very good	2,004	151.0	42.0	237.4	113.0
Good	1,842	96.1	24.7	159.5	87.0
Fair/poor	2,121	54.0	7.0	96.8	40.4
Not married					
Male	735	96.1	18.0	154.7	61.0
Female	2,914	51.0	8.9	96.8	47.0
Cognition					
Normal	3,655	129.6	40.2	207.8	110.0
Low or missing	2,318	49.4	4.6	89.1	32.7

Source: Author's calculations. All calculations are weighted. Results are from the preferred imputation method (method 7).

Our analysis uses a single cross section from the AHEAD data for all imputations. Because of the unique combination of sample size, measures of health, economic status, and family connections in the AHEAD data, many researchers will use similar cross-sectional samples of the data. We hope that our imputations will be helpful in this context. AHEAD is, however, a panel data set, and future work should extend this imputation procedure to utilize the panel nature of the data.

Appendix

Appendix tables 8A.1A and 8A.1B provide details of the imputation procedure for housing and stocks. Stocks represent an asset with high rates of missing values and a skewed distribution. Housing has lower missing value rates and a more uniform distribution. The format of the two tables is identical. Column (1) gives the number of "donors" (continuous responses that are used to match to the missing data), and column (2) gives the number of observations missing

Table 8A.1A **Characteristics of Imputation Matches for Stocks**

					Percentage of Observations with		
F-Value for Entry	Range (thousand $)	No. of Donor Observations (1)	No. of Imputed Observations (2)	No. of Covariates Entered (3)	No Match (4)	Unique Match (5)	Multiple Matches (6)
0.15	0.0–4.9	107	60	21	51.7	25.0	23.3
	5.0–24.9	137	122	18	53.3	27.1	19.7
	25.0–99.9	158	179	13	58.1	29.1	12.9
	100.0–499.9	154	109	13	48.6	33.0	18.4
	500+	24	31	12	32.3	5.6	41.9
	All	580	501	–	52.5	28.7	18.8
0.25	0.0–4.9	107	59	43	93.2	6.8	0.0
	5.0–24.9	137	120	51	100.0	0.0	0.0
	25.0–99.9	158	166	35	95.2	4.8	0.0
	100.0–499.9	154	111	34	97.3	2.7	0.0
	500+	24	45	12	31.1	22.2	46.7
	All	580	501	–	90.8	5.0	4.2
0.50	0.0–4.9	107	63	101	100.0	0.0	0.0
	5.0–24.9	137	119	46	99.2	0.8	0.0
	25.0–99.9	158	172	60	100.0	0.0	0.0
	100.0–499.9	154	116	61	100.0	0.0	0.0
	500+	24	35	12	31.4	22.9	45.7
	All	580	505	–	95.0	1.8	3.2

Source: Authors' calculations from AHEAD. Unless otherwise stated, all calculations are weighted.

Table 8A.1B Characteristics of Imputation Matches for Housing

F-Value for Entry	Range (thousand $)	No. of Donor Observations (1)	No. of Imputed Observations (2)	No. of Covariates Entered (3)	Percentage of Observations with: No Match (4)	Unique Match (5)	Multiple Matches (6)
0.15	0.0–49.9	925	435	30	68.5	21.2	10.3
	50.0–99.9	1,227	331	26	49.2	21.8	29.0
	100.0–199.9	748	146	19	19.2	19.9	61.0
	200+	294	67	26	73.1	20.9	6.0
	All	3,194	979	–	55.0	21.1	23.9
0.25	0.0–49.9	925	432	44	87.5	10.7	1.9
	50.0–99.9	1,227	337	43	92.6	7.1	0.3
	100.0–199.9	748	144	36	88.2	11.8	0.0
	200+	294	63	40	98.4	1.6	0.0
	All	3,194	976	–	90.1	9.0	0.9
0.50	0.0–49.9	925	439	64	96.6	3.4	0.0
	50.0–99.9	1,227	338	68	100.0	0.0	0.0
	100.0–199.9	748	140	51	98.6	1.4	0.0
	200+	294	61	59	100.0	0.0	0.0
	All	3,194	978	–	98.3	1.7	0.0

Source: Authors' calculations from AHEAD. Unless otherwise stated, all calculations are weighted.

asset values, for each bracketed range.[10] Column (3) gives the number of covariates used in estimating the regression for the particular bracket given the F-value criterion in the stepwise regression. Columns (4) through (6) give the percentage distribution of the imputed observations by the type of match. A multiple exact match represents the case where two or more donor observations have the same fitted value as the observation requiring imputation. No match corresponds to the case where no donor observation has the same fitted value as the observation requiring imputation, while the unique and multiple match cases correspond to the cases where one or more than one donor has the same fitted value.[11] As expected, changing the significance level from 0.15 to the lowest significance level (0.50) dramatically increases the number of covariates selected into the model, usually more than doubling the number. As a result, the probability that an exact match of fitted values will be found is greatly decreased. Among stocks, at a significance level of 0.15, about 25 percent have a unique exact match while 23 percent have multiple exact matches. Lowering the significance level drastically reduces the probability that an exact match will be found. In order to avoid the potential of overfitting the imputation model, an F-value of 0.15 was used for ordinary least squares regressions and 0.05 for logistic regressions, unless noted otherwise.[12] While the character of the matches varies across the F-significance levels, the distribution of wealth does not change dramatically.[13]

References

Chand, Harish, and Li Gan. 1994. Wealth imputation in the AHEAD. Berkeley: University of California. Mimeograph.

Herzog, A. Regula, and Robert Wallace. 1997. Measure of cognitive functioning in the AHEAD study. *Journal of Gerontology* 52B:37–48.

Little, Roderick J., I. G. Sande, and Fritz Scheuren. 1988. Missing-data adjustments in large surveys. *Journal of Business and Economic Statistics* 6 (3): 117–31.

Smith, James P. 1995. Racial and ethnic differences in wealth. *Journal of Human Resources* 30:S158–S183.

10. Note that, due to randomness in the (previous stages of) assignment of ownership and brackets, the number of observations across brackets is not constant across significance levels.

11. Because the regression consists of categorical variables, an exact match is possible.

12. Since the logistic regressions are used to form probabilities for Bernoulli draws rather than for nearest neighbor fits, a more conservative F-value is warranted to preserve the empirical distribution.

13. In a previous version of this paper, we did not use stepwise regression to limit the number of regressors in the imputation models. As a consequence, estimation of brackets with small samples resulted in perfect fitting of the model. In the case of the highest (open) bracket for stocks, this resulted in a very large value for stocks ($6 million) being imputed to more than one observation, leading to a large increase in mean wealth. We feel that using the stepwise regression procedure generates more robust imputation results.

———. 1997. Wealth inequality among older Americans. *Journal of Gerontology* 52B:74–81.
Soldo, Beth, Michael Hurd, Willard Rodgers, and Robert Wallace. 1997. Asset and health dynamics among the oldest old: An overview of the AHEAD study. *Journal of Gerontology* 52B:1–20.

Comment James P. Smith

Hoynes, Hurd, and Chand (hereafter HHC) have written an excellent paper using the recently released data from the survey on Asset and Health Dynamics among the Oldest Old (AHEAD). Their paper makes two important contributions, one methodological and the other substantive. The methodological contribution presents a new method of imputing missing asset data in social science surveys.

HHC deal with the implications of an important recent survey innovation—follow-up bracket questions—that was used extensively in both the Health and Retirement Survey (HRS) and AHEAD. When respondents did not answer a question on the value of an asset, instead of simply going on to the next question as most surveys do, both HRS and AHEAD asked a series of follow-up questions to determine whether the unknown asset value lay above or below certain selected amounts. I can only agree with HHC's bottom line conclusion on the importance of follow-up brackets for the imputation of missing values. As HHC show, the value of these follow-up brackets is that they substantially reduce item nonresponse to asset questions. Using an illustration from their paper, nonresponses to questions about the value of housing are reduced by almost 80 percent by the use of follow-up bracket questions.

The second reason why brackets matter so much is that they substantially reduce the estimation error in predicting the missing asset amount. It is one thing to try to assign a missing business value when all one knows are the characteristics of the owner. It is a much less daunting problem when one also knows that the value of the business lies between \$50,000 and \$100,000. HHC's basic results on the value of follow-up brackets are quite consistent with those I obtained with both the HRS (Smith 1995) and AHEAD (Smith 1997). Follow-up brackets are an important survey innovation that I predict will be adopted extensively in other surveys. While I agree with the major points in HHC's paper, I do have two quarrels with how they estimate their imputations. The first deals with missing values on asset ownership and the second with the sensitivity of their estimates to outliers.

James P. Smith is senior economist at RAND and holds the RAND Chair in Labor Markets and Demographic Studies.

Missing Value on Owners

The first step in their imputation procedure involved assigning missing values for cases in which respondents were uncertain or refused to say whether they had the asset. Since it affects only roughly 2 percent of the sample, imputations for this subsample of uncertain owners will not be very important in the overall scheme of things. However, it does caution us against too mechanistic an approach in our missing value algorithms. To assign a missing value, HHC estimate a logistic function for the probability of ownership using the full sample of nonmissing value respondents. Covariates in their logistic model included a rather standard and noncontroversial list of demographic and other characteristics. Imputed missing asset ownership was assigned based on a prediction from this model with a random draw from the residual distribution.

Their predictions actually imply a somewhat lower rate of asset ownership among nonresponses than was observed in the full sample, implying that, on average, characteristics of nonresponses on asset ownership are tilted toward those attributes reducing the odds of ownership. Since asset ownership is relatively rare in this age group, HHC end up assigning very low rates of ownership to these missing values. I would like to caution against this conclusion, largely because it relies on too mechanistic an approach to the entire imputation exercise. Before estimating missing values, we must first step back and ask what the nature of the process leading to nonresponse is. The approach HHC follow assumes that the forces producing nonresponses to ownership questions are basically identical (after stratifying by characteristics in the imputation algorithm) to the factors that distinguish owners and nonowners of asset in the full sample. This assumption is unlikely to be true.

There are actually two distinct reasons why respondents have missing values on whether they even own an asset. These nonresponses filter from those respondents who either said they did not know or those respondents who refused to reply. Given the relative simplicity of the question (do you know whether you have an asset?), the don't know responses in part include the cognitively impaired or those who are simply confused about the meaning of the question. This category also includes respondents who have already decided that they do not want to participate in this survey but are too polite to terminate the interview. The quickest way to get through the survey is to answer "I do not know." Supporting evidence for this view is that more than half of wave 1 respondents in the companion HRS who said that they did not know whether they had an asset had attrited from the HRS by wave 2. On average, these attriters were high wealth holders, implying that many of these respondents who said that they did not know whether they had an asset were likely to have it. Similarly, refusals represent those respondents generally quite sensitive to income or wealth questions. In most cases, such respondents probably do have the asset in question. A nonresponse is an excellent way of telling the interviewer that their wealth holdings are not his or her business. The upshot of these arguments

Table 8C.1 Distribution of Open-Ended Cases (Stocks) among Continuous Reporters

Asset Value ($)	Number of Cases
500,000	11
600,000	6
700,000	1
750,000	2
800,000	2
3,000,000	1
5,000,000	1

is that ownership rates among nonresponses to asset questions are likely to be much higher than observed for the full AHEAD sample and considerably higher than HHC predict.

Missing Data on Amounts

My second quarrel with the HHC imputations is far more critical since it can significantly affect the mean imputations of missing values. HHC impute missing values by first assigning those with missing bracket information a bracket category using ordered logistic regressions. Within-bracket imputations of exact amounts were then assigned based on a regression over those respondents with continuous amount data within the bracket. Using what they label "the nearest neighbor approach," HHC impute each individual from a continuous reporter whose fitted value is closest to the fitted value for the missing amount individual. The value assigned is the actual value of the nearest neighbor, not the fitted value.

Based on their methodology, imputation had a large impact on estimated missing values. For stocks, their estimates imply that imputed values for bracketed respondents were more than two and one-half times the amount for continuous reporters. Virtually all of this difference stems from amounts imputed in the open-ended interval for stocks (more than $500,000). In this range, mean values among continuous reporters were $871,000, compared to $2,613,000 among those whose values were imputed. Virtually all of this difference stems from the use of covariates in the imputation algorithm. Roughly similar results were obtained for other forms of nonhousing wealth. In summary, HHC's results imply that imputed values with brackets had a reasonably large impact on estimates of nonhousing wealth, particularly among those at the very top of the wealth distribution. This impact largely flowed from the use of personal covariate information in the upper open-ended brackets.

How much confidence should we place on these results? I would like to urge considerable caution due to their sensitivity to a few outlier observations. The reasons for my caution are illustrated in table 8C.1, which illustrates a typical distribution of continuous reporter cases in the open-ended interval—in this case for stocks. There were only 24 continuous reported cases with values in this open-ended interval, only two of which had values that exceeded \$1,000,000. The mean across these 24 continuous reported cases was \$871,000. Yet, HHC assign a mean of \$2,614,000 to the 21 bracket cases for stocks. The only way that this could happen is that virtually all of the missing value cases were matched to those continuous reporter cases at the top of the open-ended interval. For example, if 11 of the 21 cases were matched to the \$800,000 case and the remainder divided between the \$3 and \$5 million cases, we will still be below their estimated mean of \$2,614,000. It is clear then that in the open-ended interval HHC are matching most missing value respondents with the highest value cases. Instead of the "nearest neighbor approach," their algorithm might be more aptly titled the "richest neighbor approach."

References

Smith, James P. 1995. Racial and ethnic differences in wealth. *Journal of Human Resources* 30:S158–S183.

———. 1997. Wealth inequality among older Americans. *Journal of Gerontology* 52B (May): 74–81.