Improved Wealth Measures in the Health and Retirement Study

Asset Reconciliation and Cross-Wave Imputation

Michael D. Hurd, Erik Meijer, Michael B. Moldoff, and Susann Rohwedder

RAND Labor & Population

WR-1150

April 2016

This paper series is made possible by the NIA funded RAND Center for the Study of Aging (P30AG012815) and the NICHD funded RAND Population Research Center (R24HD050906).

For more information on this publication, visit www.rand.org/pubs/working_papers/WR1150.html

Published by the RAND Corporation, Santa Monica, Calif.

Improved wealth measures in the Health and Retirement Study: Asset reconciliation and cross-wave imputation

Michael D. Hurd[a], Erik Meijer[a,b], Michael Moldoff[a,b], and Susann Rohwedder[a]

[a]*RAND Corporation*
[b]*University of Southern California*

April 2016

Acknowledgements

We thank Mary Beth Ofstedal, David Weir, and participants at various HRS co-PI meetings and HRS data monitoring committee meetings as well as the Fourth Panel Survey Methods Workshop (Ann Arbor, MI, 2014) for numerous comments and suggestions. Furthermore, we thank Chris Chan, Jack Chen, Regan Main, Philip Pantoja, Patty St.Clair, and the RAND HRS team for help with data and coding. This research was supported by grants from the Social Security Administration and the National Institute on Aging.

Summary

The evolution of household wealth before and after retirement is a key topic of interest for economists and policy makers. Wealth is also a key covariate for a large number of studies in economics and other disciplines. Because of its large sample size and its many waves of follow-up, the Health and Retirement Study (HRS) is the main source of information about wealth trajectories at older ages in the U.S. The HRS is a large-scale multidisciplinary longitudinal study of individuals age 50 and over and their spouses of any age, who have been interviewed biennially since 1992, with new cohorts added regularly. The HRS measures wealth in about 20 components, which provides a moderately detailed picture of a household's asset portfolio. The HRS has pioneered several questionnaire innovations that have been shown to improve the quality of the wealth measurements, and as a result, the HRS wealth measures are considered high quality.

However, the wealth data contain missing values due to the respondent not knowing the answer or refusing to answer. Furthermore, despite all efforts to obtain the best possible data, respondents can make a mistake when reporting about their wealth, and interviewers may mis-hear or mis-type the answer. To catch gross errors of this kind, the HRS has implemented an *asset verification* section since 2002, in which respondents are asked to correct or confirm current or previous reports about wealth components if the difference between them is large. RAND provides user-friendly versions of the HRS data, including *imputations* of missing data. Until version L (2013) of these data, however, these did not take the asset verification section into account, and the imputations did not exploit the longitudinal information in the data.

In this report, we present improved wealth measures for the HRS, which aim to reduce the effect of observation error on wealth levels and changes in wealth. The improved measures have been available since version M (2014) of the RAND income and wealth imputation data for the HRS, and are included in the RAND HRS. The new wealth measures take account of the asset verification section in the HRS and use cross-wave information, most notably the value of the same asset in adjacent waves, in the imputation models, so imputed values better preserve serial correlation in the asset values. We document how we dealt with several methodological challenges in the implementation of these improvements and compare distributions of asset values and changes in asset values between wealth measures that use the former methodology and the improved methodology.

Incorporating the asset verification data has small and nonsystematic effects on the means of total household wealth and wealth components, but due to the correction of a few large outliers standard deviations in some waves are reduced substantially (47% in 2006).[1] The corrections from the asset verification data have little effect on the average wave-to-wave *changes*, but systematically reduce the standard deviations of these changes by

[1] However, in one wave (2002), two large values were introduced by this section, leading to a substantial increase in the standard deviation.

substantial amounts (up to 57% for total wealth and more than 60% for primary residence and stock wealth),. These corrections also lead to systematic but modest reductions (up to 7%) in the number of changes in ownership, ownership reversals (a change followed by a change back), and spikes and trenches, which are defined as large changes in value followed by large changes back. We find that trenches are much more common than spikes, which suggests that accidental underreporting is more common than accidental overreporting.

The cross-wave imputations have small and nonsystematic effects (compared to the cross-sectional imputations) on the means and standard deviations of the distributions of total wealth and its components, although occasionally the percentage differences for individual components are larger due to smaller numbers of households who own a specific asset. The effects on the mean changes are again small and nonsystematic, but in most years, the standard deviations of the changes are reduced by a modest amount (generally much less than 10%, and in some cases, the standard deviation increases). With the exception of the first few waves (1992-1998), the cross-wave imputations consistently reduce the number of ownership changes (2-5%) and ownership reversals (3-7%). The most important effect of the cross-wave imputations is on the numbers of spikes and trenches. These are reduced considerably, in most cases between 20 and 30 percent for the spikes and between 10 and 20 percent for the trenches.

Contents

1. Introduction and overview of the HRS asset elicitation

The Health and Retirement Study (HRS; Juster & Suzman, 1995; NIA, 2007) is a large scale multidisciplinary panel survey of individuals over age 50 in the U.S. and their households. The HRS is the primary source of information about the elderly and future elderly in the U.S. and as such plays a key role in scientific studies of this age group and in the evaluation of actual or proposed policies regarding Medicare, Social Security, and many other topics. The RAND HRS (Chien et al., 2015) is a respondent-level data product that contains a large subset of the HRS variables, merges data from all waves, attaches data from the spouse or partner to the respondent's record, includes imputations of missing data on most economic variables, and implements a host of other data coding, recoding, renaming, and processing measures, and adds detailed documentation, in order to facilitate the use of the HRS by researchers from across disciplines.

Among the main interests of economists and policy makers is the economic well-being of households in this age group, and in particular their wealth and how it evolves over this part of the life span. The HRS measures wealth in about 20 categories, see Table 1.1. The HRS has been at the forefront of methodological innovations in economic measurement (e.g., Hurd, Juster, & Smith, 2003) and correspondingly, the quality of its economic data is highly regarded. One of the innovations is the use of follow-up questions after an initial nonresponse, through a sequence of so-called *unfolding brackets*. The typical sequence of questions starts with asking about whether the household *owns* the asset (ownership). If so, the respondent is asked about the *value* (amount). If the respondent indicates owning the asset but does not provide a value, the respondent is asked the unfolding *bracket* questions, which are questions of the form "Is it more than $X, less than $X, or about $X?". One or more of these questions narrow the interval of uncertainty. Although the unfolding bracket questions are not entirely without drawbacks (Hurd, 1999), they considerably increase accuracy of economic measurement and this reduces problems with nonrandom missingness (Juster & Smith, 1997). The information from the unfolding bracket sequence plays an important role in ensuring high quality imputations for HRS income and wealth variables, which are released by RAND both as a standalone data product (Pantoja et al., 2015) and in condensed form as part of the RAND HRS (Chien et al., 2015).

As shown by Juster, Smith, & Stafford (1999), even if cross-sectional wealth measurement is satisfactory, this may still need further improvement if the goal is to use the between-wave changes in wealth to study wealth accumulation (including saving) and decumulation, because the effects of measurement error in levels on empirical estimation of economic models are exacerbated by taking first differences (see, e.g., Wansbeek & Meijer, 2000, pp. 138-146, and the references therein). Both Hill (2006) and Venti (2011) have studied the distribution of wealth changes in the HRS and found some evidence that measurement error may disproportionately affect these distributions. .

Table 1.1: Assets asked in the housing (H) and income and wealth (Q) sections

Asset (code)	Description	Waves	Notes
	Section H		
hou1	Primary residence (not mobile home)		
mobl	Mobile home as primary residence		
mrt1	Mortgage 1 on primary residence		
mrt2	Mortgage 2 on primary residence		
eqcd	Home equity line of credit balance	not in 2A	(1)
eqln	Other home loans on primary residence	not in 2A	(1)
houb	Second home	not in 2A & 3	(2)
mrtb	Mortgages/loans on second home	not in 2A & 3	(2)
	Section Q		
rles	Other real estate		
tran	Vehicles		(3)
trn1	Vehicles (except RV)	only 1	
trn2	Recreational vehicle (RV)	only 1	
bsns	Businesses/farms		
ira	Individual Retirement Accounts (IRAs)	1 and 2	(4)
ira1	First IRA	3 and later	
ira2	Second IRA	3 and later	
ira3	Remaining IRAs (3+)	3 and later	
stck	Stocks and mutual funds		
chck	Checking and savings accounts		
cd	Certificates of deposit (CDs)		
bond	Bonds and bond funds		
trst	Trusts (not reported elsewhere)	not in 2H	
othr	Other assets		
debt	Other debt		

Notes. These are short descriptions. See Pantoja et al. (2015) for detailed descriptions of what is included in each category. 2A = AHEAD wave 1 (1993); 2H = HRS wave 2 (1994). The HRS also measures several other wealth components, most notably pension wealth and the cash value of life insurance, but these are not included in the RAND income and wealth file and therefore not included in this study.

(1) These are asked as one combined category in 2A, but they are not included in the RAND file.

(2) These were asked in all waves, but there were some problems of comparability and therefore these are not included in the RAND files for waves 2A and 3.

(3) In wave 1: Sum of trn1 and trn2 (cross-sectional imputations); ONLY trn1 (cross-wave imputations)

(4) Sum of ira1-3 from wave 3 onward

Hill (2006) reports the results of an experiment in the HRS (performed in 2001) that involved a call-back of households who had large changes in reported asset values (between 1998 and 2000) and asked them to confirm or correct these values. Incorporating the corrections from this call-back led to a drop in the variance of the change in net worth of about 50%. Meijer, Spierdijk, and Wansbeek (2013) showed the effect of measurement error in wealth in the HRS on regression coefficients. They estimated a simple dynamic model for the evolution of wealth and found that the autoregressive coefficient jumps from 0.1 without taking measurement error into account to 0.4 after taking measurement error into account

As a result of the findings in Hill (2006), the HRS questionnaire has included the asset reconciliation (or asset verification) section, section U, since the 2002 wave. When the value of an asset differs more than $50,000 from the value in the previous wave (and total wealth differs more than $150,000 and the financial respondent is the same), the respondent is asked to confirm or correct current wave's and last wave's values of the asset. Chapter 2 contains an overview of section U.

Since Version M of the RAND HRS, we have incorporated information from section U in the wealth measures constructed for the RAND HRS. This consists of three elements: (a) corrections of current wave's asset values using current wave's section U corrections, (b) corrections of current wave's asset values using the next wave's section U corrections, and (c) updated imputations. Because the structure of section U (i.e., the questions and skip patterns) is not identical to the structure of the asset section (Q) and housing section (H) from which the original asset reports are taken, the respondent corrections may result in data inconsistencies or lead to ambiguities and other data problems with the imputations. Chapter 3 discusses these issues in detail and chapter 4 discusses issues that are specific to using the retrospective information from the next wave's section U. Chapter 5 presents the effects of the section U corrections on the distributions of assets in each wave and on the distributions of the asset changes between waves.

When a respondent fails to report a continuous value for an asset, RAND imputes the value. These imputations are generally regarded highly, for two main reasons. First, in the data collection phase, HRS has implemented innovations that increase the quality of the data at the source. In particular, the use of follow-up questions after a "don't know" or "refuse" answer, in order to elicit partial information in the form of so-called *unfolding brackets* greatly enhances the amount of available information and reduces the amount of imputation uncertainty. Second, the imputation method uses the information in the unfolding brackets systematically and thoroughly, imputing ownership, bracket, and value as necessary; it uses a broad set of covariates in the imputation model; and for most value imputations uses a nearest neighbor approach that does not require distributional assumptions, which makes the method robust to model (mis)specification issues. In Version M of the RAND HRS, we have introduced a further refinement of the imputation method, aimed specifically at reducing variance of asset changes induced by the imputation uncertainty. This refinement consists primarily of adding information about the previous wave's and next wave's value of the asset to the set of covariates used in the imputation models. We will refer to these as *longitudinal* or *cross-wave* imputations, as

opposed to *cross-sectional* imputations. Because the values of most assets tend to be highly correlated across time, this is expected to lead to both more accurate imputations with less imputation error and smaller variances of asset changes. Chapter 6 describes the methodology and Chapter 7 describes the results.

In Chapter 8, we present the results of an experiment that was performed in the 2012 HRS. In this experiment, respondents who would not be eligible to enter section U (because the financial respondent was different than in the previous wave) were asked to confirm or correct the reported value of their house, checking and savings accounts, or stocks if these values differed by more than $50,000 from the prior wave reported value for the household.

Chapter 9 concludes.

The results presented in this report have been obtained using the cross-sectional and cross-wave imputations from version N (2014) of the RAND HRS and RAND income and wealth imputations file, and for comparisons cross-sectional imputations using the same methodology but not using the section U information. The latter are not publicly available, but can be obtained from us upon request. Version O (2015) introduced a number of small improvements to the imputations, and especially the underlying demographic data (see Chien et al., 2015, for details). The effects of these on the distributions of the wealth variables are small, and the comparison between the different methodologies presented here should also apply to the version O imputations.

In addition to the work reported here, we have also studied outlier detection and correction. The large variances of asset changes are largely driven by a few extreme values, which are reported values. Some of these are corrected in section U, but others are not. We have developed a Bayesian model for net worth across time, which has a flexible but smooth functional form and allows individual heterogeneity, but also contains a "contamination" component that is thought to represent misreports. Based on the model, we have constructed imputations of the true net worth values that take the model-based estimate of the probability of a contamination into account. We have no plans to include these in the main RAND HRS wealth measures, but we envision releasing this as a separate research data set. This is, however, outside the scope of the current report. This work will instead be reported separately in Burgette, Hurd, & Meijer (2016).

2. Overview of section U

2.1 Eligibility for section U

The basic rules for determining whether a respondent enters section U have been relatively stable since section U was introduced (2002), although there have been some significant changes in the details and in the implementation, which have had profound impact on the effectiveness of section U. Respondents who enter section U typically answer questions about a small subset of their assets: only the ones that changed much since the previous wave. Hence, eligibility is determined at the respondent-asset level. Entering section U for a certain asset requires both global and local eligibility. Global eligibility is a respondent-level criterion, whereas local eligibility is determined at the level of the asset type.

Central to the determination of eligibility is the "previous wave", because eligibility depends on changes in assets in the current wave compared to the "previous wave". In waves 6-10 (2002-2010), the "previous wave" was defined as the last wave prior to the current wave that the financial respondent's household participated in. This could be more than two years earlier if the household did not participate in one or more waves prior to the current wave. For example, if the household responded in wave 4 (1998) but not in waves 5 and 6 (2000 and 2002), then in wave 7 (2004), the "previous wave" was wave 4, and changes in asset values were computed with respect to wave 4's asset values. In wave 11 (2012), the definition of "previous wave" was changed to just the prior wave (wave 10).

With this definition, *local eligibility* is defined as a change in the asset of more than $50,000. That is, a difference between the currently reported value of the asset and the one reported in the "previous wave" of more than $50,000 is considered a large change that is eligible for questioning in section U. If there is no "previous wave", there is nothing to compare to and the respondent does not enter section U. This happens when a household is interviewed for the first time, and in wave 11 it happens if the household did not participate in wave 10.

Missing values slightly complicate the comparisons. When respondents give a "don't know" or "refuse" answer, they enter a sequence of unfolding bracket questions (see appendix D), which often provide some information about the asset value. It is then possible to determine minimum and maximum values that are consistent with the respondent's answers. In the comparisons, these are used by comparing the closest values. The difference criterion is interpreted as a gap of more than $50,000 between the previous wave and current wave intervals, that is, the minimum of one exceeds the maximum of the other by more than $50,000. An exception is "DK own", which is treated as a reported zero for this purpose, except for 2002, in which it was the entire nonnegative real line.

A respondent is *globally eligible* for entering section U if the respondent is the financial respondent of the household and was also the financial respondent in the "previous wave",

and the change in net worth between the "previous wave" and the current wave is more than $150,000. (In waves 6-8, this threshold was $1,000,000 for a random 10% of the sample.) The reason for the financial respondent criterion is confidentiality of the answers that a respondent gives: these should not be revealed to others (such as the spouse). Because section U mentions the previous wave's answer, this can only be asked if it is the same person answering.[2] The rationale for the net worth criterion is to prevent bothering households who rebalance their asset portfolios, for example, by selling their stocks and putting the money in a savings account.

Like the local eligibility criterion, the net worth comparison is also complicated by missing values. For each asset, we can determine minimum and maximum values that are consistent with the respondent's answers, as is done for determining local eligibility, which can then be aggregated to provide a lower and upper bound of net worth. However, if the respondent gives a noninformative answer to one or more asset questions, or an open bracket for both an asset question and a liability (negative asset) question, then these lower and upper bounds may span all nonnegative numbers, or even the whole real line. The result is that we cannot rule out a small change in net worth. This would imply that many respondents whose asset values for one or more assets change significantly would not enter section U. Because the net worth criterion is an auxiliary criterion and the actual net worth underlying it is never mentioned to the respondent, HRS takes a pragmatic approach and replaces unbounded intervals by bounded ones. Specifically, if the maximum of an asset is infinite, it is replaced by 3 times the minimum. This is how a finite minimum and maximum for the previous wave's net worth was computed from waves 6-10. For the current wave, this is further simplified by computing only a rough estimate of net worth. For closed brackets (bounded minimum and maximum), the average of the minimum and maximum is used as the estimate of the asset value for this purpose, whereas for open brackets (unbounded maximum), 3 times the minimum is used as the point estimate. In wave 11, the previous wave's net worth for this comparison was also changed to a single amount: for respondents who were included in the wave 10 early release, this was the aggregate from the RAND imputations of the assets, whereas for respondents who were not in the wave 10 early release, HRS internal imputations were used. Of course, such imputations are not available for the current wave data, so the algorithm for computing this was not changed.

A refusal for an asset in section U leads to a break out of section U, so respondents may not enter section U for all assets for which they are eligible. For example, suppose that a respondent was eligible for entering section U for both "other assets" and primary residence. If the repondent refuses to answer for "other assets", primary residence is also skipped.

The eligibility description here is how it was intended. Unfortunately, however, there have been some problems with the implementation, especially the preloads, which in some waves affected substantial numbers of households. Section 2.3 mentions some of these issues. Hence, some respondents inadvertently entered section U and some

[2] This carries over to proxy interviews: the proxy needs to be the same person as in the previous wave.

respondents inadvertently did not. Nevertheless, if a respondent entered section U and corrected a value, the data are valid whether or not the respondent "should have" entered section U, and therefore, we use these data to make our corrections.

2.2 Assets included in section U and potential comparability issues in their definition

Table 2.1 lists the assets that are potentially questioned in section U. Assets are asked in the order listed. The order matters because in 2002 and 2004, only the first three eligible assets were asked, and in all years, a refusal for one asset means that later assets are not asked about.

Table 2.1: Assets included in section U

Number	Description
1	Other debt
2	Trusts
3	Other assets
4	Vehicles
5	Certificates of deposit
6	Checking and savings accounts
7	Bonds and bond funds
8	Stocks and mutual funds
9	Individual Retirement Accounts
10	Businesses/farms
11	Other real estate
12	Primary residence
13	Mortgage 1 on primary residence
14	Mortgage 2 on primary residence
15	Home equity line of credit balance
16	Mobile home
17	Second home
18	Mortgages/loans on second home

Several assets that are asked in the HRS are not asked in section U: other loans on the first home are not included, the cash value (and face value) of life insurance, and defined contribution pension plan (e.g., 401(k)) balances. Section Q asks about all trusts and trusts that are not reported elsewhere, whereas at most one of these is asked in section U (see below). The other home loans and trusts not reported elsewhere are in the RAND wealth imputations file, and the former is also part of RAND's total wealth measure.

The structure of section U is much simpler than the structure of the housing (H) and asset (Q) sections, which often have several questions and introductions leading in to the question about the value of a specific asset, and more complicated skip patterns. Chapter 3 discusses some of the implementation issues for the corrections that result from these differences in questionnaire structure. One aspect of the simplification is that the

description of assets is often simplified and abbreviated. The descriptions have occasionally been adapted in an attempt to improve their ability to succinctly indicate what is included and what is not, but significant differences between the wording in sections H and Q on the one hand and section U on the other remain.

Some of these wording differences are:

- Most assets in section Q include the qualifier "that we haven't asked about", whereas in section U this qualifier is omitted. It is difficult to imagine how one could introduce such a qualifier in section U and for most assets it will not matter. But for some, such as "other debt", the difference in wording could have an effect on the perceived inclusiveness of the asset category.
- The questions about the values of many assets in section Q ask about the *net* value, using the phrasing "if you paid off anything you owed on them". In section U, no such qualifiers are presented. This may influence the answers for some asset categories, for example, vehicles and other real estate.
- In many cases, asset categories in sections H and Q appear to be more inclusive (broader) than in section U. For example, CDs just mentions "certificates of deposit" in section U, whereas in section Q, it also includes Government Savings Bonds and Treasury Bills. (However, these categories are explicitly excluded from the "bonds" category in section Q and not in section U.)

For the implementation of the section U corrections, we necessarily assume that these differences in wording do not affect the concepts as they are perceived by the respondents, and thus the amounts given in section U refer to the same assets as in sections Q and H.

The preloads for "trusts" were based on the total trust value in 2002-2004 and 2008 data models[3] 1-2; in 2006, trusts were not asked in section U. Starting with 2008 data model 3, both previous and current wave values reference the value of trusts not reported earlier in the interview. In all cases, however, the wording of the question does not make this distinction and just mentions "trusts worth ...". See appendix A for a more detailed discussion of trusts. Because of this ambiguity, we do not implement the section U corrections for trusts.

2.3 Section U survey design changes since inception

2001:

- Experimental call-back of respondents from the 2000 wave, as reported in Hill (2006). These data are not available, but they led to the inclusion of section U in the questionnaire.

2002:

- Section U introduced;

[3] A "data model" is a version of the questionnaire, including skip patterns, and preload variables. Every time when during field work (small) corrections are made to the questionnaire or the preloads, the data model number is increased.

- Up to 3 assets were questioned, in the priority order of Table 2.1;
- "Previous wave" is the latest wave in which the household participated, which may be earlier than 2000;
- Problem with the preloads: changes from no ownership to ownership were never challenged.

2004:

- Change in data structure, but not in the questionnaire;
- The 2002 preload problem was fixed;
- The preloads used for the previous wave's values were the section Q/H values, not the corrected section U values (if applicable);
- The "comma problem" was introduced (see below).

2006:

- The limit on the number of assets questioned was removed;
- Trusts were not asked in section U (see appendix A).

2008:

- Trusts were asked again, but (from data model 3 onward) only the ones not reported elsewhere are used for determining eligibility and presentation to the respondent, although the wording of the question itself has not changed.

2010:

- Preloaded values now incorporate the previous wave's section U corrections for the first correction made in 2008. (The second and later corrections were inadvertently not used and respondents saw the section Q/H values from 2008 as the 2010 preloads for those assets);
- The "comma problem" was fixed.

2012:

- All bracket thresholds are now aligned with sections Q and H;
- The "previous wave" now only means wave 10 (2010): if the household was not present in wave 10, it is not eligible for section U;
- Improved wording of some of the asset categories to better align them with the section Q/H wording;
- Imputations for the 2010 assets are now used to estimate total wealth for the global eligibility criterion instead of the rough method used earlier, although the latter is still used for the 2012 total assets;
- It was intended that preloaded values now incorporate all of the previous wave's section U corrections, but there were some glitches in which still the uncorrected section Q or H preloads were used or the preloads were missing for some respondents. In some cases, the 2010 corrections for 2008 were used as preloads.

2014:

- Preloaded values now incorporate all of the previous wave's section U corrections.

The comma problem

In some years, dollar amounts included thousands separators (commas), either in the field where the interviewer entered them, or in the amounts presented on screen (that had been reported previously), or both, whereas in others, the commas were not always present. The latter is probably related to some amounts that change by a factor 10, for example, a change from 25000000 to 2500000 ($25 million to $2.5 million) in the value of the primary residence between 2004 and 2006. The comma problem affects assets in 2004-2008. It was corrected in 2010.

3. Issues with implementing section U corrections

The structure of the section U questionnaire is much simpler than the structure of the housing section (H) and the income and asset section (Q) questionnaires, and occasionally differs in other respects as well. This poses some challenges for the implementation of the corrections, which are described in this chapter.

3.1 Ownership ambiguity

In sections H and Q, respondents are first asked whether they own a certain asset. If so, they are asked about the value. In section U, respondents are only asked to correct the value. Hence, respondents are not able to correct ownership explicitly. However, if a respondent gives a nonzero amount in section U, the household must own the asset. We interpret a change of the value to $0 as indicating that they do not own the asset, rather than own the asset without it having any economic value. This differs from sections H and Q, where we can distinguish between the two, and where it does happen that respondents report owning the asset and then report a value of $0.

3.2 Inconsistencies in the housing data

Determining ownership for housing wealth assets is more complicated than it is for non-housing assets. Ownership of non-housing assets is self-contained, that is, it does not depend on ownership of other assets. For example, debt ownership is concluded by whether debt values and indicators are present from the HRS questionnaire. However, ownership of housing assets is interrelated: a household would only have a mortgage if that household also owned a house.

Housing ownership allocation before section U corrections

To see how Section U impacts the ownership situation, it is crucial to first understand how the process works without Section U value corrections. Section U includes seven housing wealth assets. They are: primary residence ("house"), mortgage 1, mortgage 2, home equity line of credit, mobile home, second home, and mortgage on the second home. Section H additionally has "other home loans", which is not included in section U. The seven assets can be divided into two sets, where answers to one question in the set potentially affect asset values and ownership indicators of other elements of the set. Set 1 consists of primary residence,[4] mobile home, home equity line of credit, mortgage 1, and mortgage 2. Second home and mortgages and loans on the second home comprise set 2.

[4] Unless otherwise stated, in this report "primary residence" does not include a mobile home. Mobile home as the primary residence is a separate category.

For set 1, ownership is first assessed for house and mobile home. Based on Section H variables for home type (H002) and home ownership (H004, H008, H014), the asset ownership flags for house and mobile home are set to owned, not owned, or undetermined. Next, mortgage 1, mortgage 2, and home equity line of credit are evaluated. For mortgage 1 and mortgage 2, the variable H024 (loan with the property as collateral) is used along with the house and mobile home ownership flags. The value of H024 (which is a "check all that apply" question) is an indicator that represents whether the household has a first mortgage, second mortgage, or other loan on the house or mobile home. If the household does not own the house or mobile home, then mortgages 1 and 2 are deemed not owned. If H024 indicates that a mortgage is owned and either house or mobile home is owned, then the mortgage asset is considered owned. Home equity line of credit is evaluated in the same way, except that instead of H024, the variable H056 is used. If either house or mobile home is owned, and the property loan variable indicates a response of don't know/refused or blank, then the ownership of the loan type is imputed.

Similarly to set 1, the set 2 assets are connected to determine ownership. If the second home is not owned, mortgages and loans on the second home are not owned either. If the second home is owned and H170 (a "check all that apply" question for loans on the second home) indicates a yes or no answer, then second home mortgage or loan ownership can be concluded as yes or no. Else, second home mortgage or loan ownership is imputed.

During the ownership imputation process, primary residence, mobile home, and second home are imputed before the other housing assets. The main reason for this is that if it is determined that both primary residence and mobile home are not owned, the mortgages and home equity line of credit are also classified as not owned. The same goes for second home mortgage or loan if second home is imputed as not owned.

Section U effects on housing assets

Section U potentially leads to inconsistencies in the housing data: combinations of values of variables that would not be allowed in section H. Suppose, for example, that a household reports owning a house and a mortgage in section H, and in section U is questioned about the value of the house but not about the value of the mortgage. Then if the respondent changes the value of the house to zero, which we interpret as not owning, the situation arises in which the house is not owned but the mortgage value is still positive, which would not be possible in section H because of the skip patterns there. If such a situation happens, our imputation programs ignore the value of the mortgage and set the mortgage variable to not owned, thereby enforcing the section H skip pattern logic. The same logic is applied to the other mortgages, loans, and equity line of credit on the first home, and analogously for the loans on the second home versus ownership of a second home.

Another complication that we encounter in the data is that after section U, a household may own both a house and a mobile home as primary residence, whereas the section H skip patterns do not allow this: the primary residence is either a house (or condominium or apartment) or a mobile home. This happens in a handful of cases each wave. We inspect all these cases visually to determine what the most appropriate way to proceed is. In the vast majority of cases, the resulting values of house and mobile home are very similar or even identical, and similar to the value reported for the primary residence (either house or mobile home) in the previous wave. Hence, it seems that in most of these cases, there is some ambiguity over how to classify the primary residence, and the two resulting values reflect the same asset. We then assign the value to what seems the most appropriate asset category (often mobile home). This has no effect on the value of the RAND HRS primary residence variable, which combines the two, but it matters for the separate house and mobile home variables in the Income and Wealth Imputations File. In a few cases, the values of the house and mobile home are very different and the value of one may be similar to the value of the second home, or the previous wave's second home. Therefore, occasionally, we resolve the issue by assigning one of the two values to the second home variable. We also look at the "other real estate" variable, but so far we haven't assigned any of the values to this asset. We also look at the "transportation" variable, because there may be some ambiguity about the distinction between a mobile home and a recreational vehicle (RV), but this has not led to any additional insights so far.

3.3 Multiple IRAs vs. only their total

In section Q, respondents are asked how many IRAs they have. Then they are asked to report the values of the two most important ones separately, and the total of all remaining IRAs. Hence, this results in up to three separate values. RAND imputes these values when they are missing, and in the Income and Wealth Imputations File, the three values are included separately, although in the main RAND HRS file, only their total is included. Section U does not ask for the three IRA values separately, but only their combined total value. Thus, if a respondent makes a correction in section U, we have to decide how to determine the number of IRAs and the allocation of the total to the three IRA values. In most cases, this issue is fairly minor, because it only affects the distribution of a fixed total across the three components, and most users will only be interested in the total. Nevertheless, we need to make a choice (or change the data structure, which is a less attractive option), and in case users are interested in the separate values, we should make a choice that seems reasonable, although we do not aim for optimality here, especially because we would like to implement the section U corrections before we do the imputations.

Based on these considerations, we have implemented the following rules:

1. If the respondent reported a number of IRAs in section Q, then this number is assumed to apply to the section U IRAs as well, unless it conflicts. The latter only happens if there is a positive number of IRAs in section Q and a zero amount in section U, or no ownership in section Q and a positive amount in section U. A zero amount in section U is interpreted as no ownership, so this is straightforward.

With a positive amount in section U and no ownership in section Q, the latter is replaced by ownership but unknown how many (i.e., "1+").

2. If the number of IRAs is unknown, it is imputed using the standard imputation programs.
3. If in section U a bracket response is given for the total amount of the IRAs, the continuous amount needs to be imputed. For imputation purposes, the bracket is assigned to IRA1 with IRAs 2 and 3 fixed to zero, and the resulting imputation of the IRA1 amount is interpreted as an imputation of the total IRA amount.
4. If steps 1-3 result in a positive total IRA amount with only one IRA owned, this total is assigned to IRA1.
5. If steps 1-3 result in a positive total IRA amount and two or more IRAs owned, and continuous amounts were reported for these IRAs in section Q, the new total amount is allocated to the two or three IRA variables proportional to the amounts reported in section Q.
6. If steps 1-3 result in a positive total IRA amount and two or more IRAs owned, and we cannot apply the previous rule, the new total amount is allocated to the two or three IRA variables proportional to the amounts reported in section Q by all respondents who gave only continuous amounts in section Q and who reported the same number of IRAs (i.e., two or three) as the current respondent.

3.4 Nonaligning brackets

For some assets in some waves, in particular stocks in 2004-2010 and other debt in 2010, the threshold values of the unfolding brackets in section U differ from the ones used in section Q or H for the same asset. This affects the imputation procedures. For households that report owning the asset, or for which owning the asset is imputed, but for which neither a continuous amount nor a complete bracket was reported, a complete bracket is imputed. This is done by first estimating an ordered logit model with the complete bracket as dependent variable, depending on a set of covariates (see chapter 6 for more details about the imputation methods), for respondents who have given a complete bracket response. In this model, the cutpoint parameters are estimated, which makes the procedure more robust to misspecification than imposing them (based on the known threshold values), because when estimating them, we are essentially allowing for nonlinear transformations of the underlying latent dependent variable.

The estimated parameters of this model are combined with the covariates of observations for which a bracket needs to be imputed, which results in a probability for each bracket. The imputation then consists of drawing a random number that assigns the bracket, according to the computed probabilities. If a respondent gives an incomplete bracket response, we have some information about the bracket, but the interval that is consistent with the respondent's answer is a combination of two or more complete brackets. Imputation is then done within the set of consistent brackets. That is, the probabilities for the other brackets are set to zero and the probabilities for the consistent brackets are scaled up such that they sum to 1.

The brackets are also used in the imputation of the continuous amount. For this, the amount is imputed within the given bracket. For a closed bracket, this means restricting the set of donors to the observations with answers within the given bracket, whereas for an open (top) bracket, this means drawing a number from a truncated normal distribution that is estimated by a tobit model.

There are several consequences of the nonaligning brackets. First consider the imputation of the continuous amount. The estimation of the imputation models for the amount is not affected by the nonaligning brackets, because the dependent variable is the (observed) continuous amount and the model is not estimated separately for the different brackets. However, the imputation procedure itself is affected. Because the brackets form a mutually exclusive and exhaustive set, the standard imputation program does this by splitting the sample into subsamples per bracket and then imputing within subsamples. With the nonaligning brackets for section U added, the brackets are not mutually exclusive anymore. Thus, in the adapted program we first set aside all observations with nonaligning brackets and impute the ones with the section Q/H brackets. Then we set those aside and impute the amounts for only the observations with nonaligning brackets, using the different set of thresholds, and with the donors split according to the section U brackets.

Second, consider the imputation of the bracket itself. Respondents who report a bracket value in section U that does not correspond to a section Q/H bracket have to be excluded from the estimation sample, because the dependent variable cannot be mapped to the categories used in the model. Then, for observations for which no bracket information was given (i.e., "no value/bracket", "DK ownership", or "no financial respondent"), the situation after section U correction is identical to the situation if there was no bracket information in section Q or H. Hence, for these, we can impute the section Q/H bracket as if this information was given in section Q or H, even if the information (or lack thereof) was extracted from section U. For respondents who gave an incomplete bracket response, we cannot compute the probabilities for the complete brackets, whether these are defined as the section U brackets or the section Q/H brackets. Therefore, these cannot easily be imputed. It turns out that there were only five observations across all waves with an incomplete bracket that did not correspond to a section Q/H bracket. We concluded that it was not worthwhile (or even feasible) to develop special protocols for these cases. Therefore, we take a practical approach and treat these as complete brackets, that is, we directly impute a value within each bracket.

4. Section U corrections of the previous wave

In section U, respondents are confronted with the previous wave's report and the current wave's report. They are asked whether these are correct, and if not, which wave is incorrect. They can indicate that the current wave's value is incorrect, the previous wave's value, or both, and then they are given the opportunity to correct the value(s) of the wave(s) that was (were) claimed to be incorrect. Consequently, we may have up to three reports about an asset value in the current wave: the initial report in the current wave's section Q or H, the section U correction in the current wave, and the section U correction reported in the next wave about the current wave. In the case of multiple reports, we have to decide which value to use.

We always give same-wave corrections in section U priority over the section Q/H value. Thus, we implement all section U corrections that apply to the current wave, subject to the implementation issues discussed in the previous chapter.

For "previous wave" corrections from the next wave's section U, we take into account whether the respondent entered section U in the current wave as well. For example, suppose that a respondent entered section U in 2006 for stocks and corrected the 2006 stocks value, and then entered section U for stocks again in 2008 and corrected the 2006 stocks value again, we tend to give priority to the 2006 (contemporaneous) section U report. We believe that it is likely that respondents know the current value of their assets better (and tend to report it more accurately) than the value two years earlier. The latter is subject to potentially misremembering values, or misremembering the dates of acquisition or sale of assets. However, we inspected the "previous wave" reports and found that they do contain useful information. Therefore, we incorporate this information where this seems reasonable.

The question is under what circumstances to use wave *t*+1's corrections for wave *t*'s asset values. For wave *t*'s asset values, we can distinguish between the cases in Table 4.1. For cells A1, A2, B1, and B2, there is nothing to do. For cell C1, only the contemporaneous section U's answers are relevant, and we have implemented these corrections as discussed above.

Table 4.1: Labeling of combinations of confirming or correcting a value in two consecutive waves.

Wave *t* section U	Wave *t* + 1 section U		
	Did not enter sec.U	Confirmed	Corrected
Did not enter sec.U	A1 (626,623, 95.4%)	A2 (9,354, 1.4%)	A3 (3,229, 0.5%)
Confirmed	B1 (10,155, 1.5%)	B2 (3,954, 0.6%)	B3 (994, 0.2%)
Corrected	C1 (1,810, 0.3%)	C2 (607, 0.1%)	C3 (373, 0.1%)

Note. frequencies and percentages across all assets and waves between parentheses.

For cells C2 and C3, a complication is that the respondent in wave *t*+1 was presented with the uncorrected answers from wave *t* in the earlier waves. In 2014, and in some situations in 2010 and 2012 (see section 2.3), they were presented with the corrected ones.

To find whether corrections from the previous wave seem to be informative we checked in the HRS 2010 data whether, respondents who were given the corrected values were relatively more likely to confirm than correct, compared to when given the uncorrected values. We found that to be the case and decided therefore to implement these corrections of previous wave values as well.

There is a discrepancy between what was reported in section U in wave *t* and what was reported in section U in wave *t*+1 in cell B3 for all respondents, in C2 for those who were given the uncorrected values, and in C3 for those who were given the corrected values, Furthermore, in many cases where the same value was corrected twice (C3 with uncorrected value), the corrections are not identical, and thus there is a discrepancy as well. In all these cases, it seems reasonable to assume that the concurrent (wave *t*) section U report is more reliable than the later (wave *t*+1) one, so we use the concurrent section U report. However, we make one exception: If the wave *t* correction is by an order of magnitude or more, which with a little margin we define as a factor of 9 or more, there may be a decimal place error. We visually inspected these cases and developed general rules for assigning them, but we evaluate all of these and make corrections where we deem those necessary.

Cell A3 is the one that potentially contains the most valuable additional information. It is also much more common than B3, C2, and C3. For example, in 2008 across all assets there were 802 observations in A3, 224 in B3, 147 in C2, and 86 in C3. Surprisingly, most corrections in cell A3 are corrections of asset ownership, and in particular replacing no ownership by a positive value. For example, of the 802 corrections (of 2006 assets) in 2008, there were 312 (39%) in which no asset was reported in 2006 and 48 (6%) in which a positive asset value was corrected to zero. In 2004, 2006, and 2010, the sum of these two categories even exceeds 50% of all corrections. However, in 2012, this dropped to 23%.

One potential explanation of the correction from no asset to a positive value would be that in wave *t*, the household owned the asset, but this was somehow misreported in wave *t* and corrected in wave (*t*+1)'s section U. Another potential explanation would be that in wave *t*, the household did not own the asset, then acquired it at some point between wave *t* and wave *t*+1, and the recollection of the date at which it was acquired is incorrect. There are some indications that the former explanation dominates. For example, if incorrect recollection was the main explanation, we would expect to see the same pattern in cell B3, where the value was confirmed in wave *t*'s section U, as in cell A3, where they did not enter section U in wave *t*. However, in cell B3, the fraction of any ownership corrections is lower than in cell A3 (36% of corrections vs. 45% of corrections in 2008; in 2006 and 2010 and to a lesser extent 2012 the differences are larger and in 2004 the sample size was too small to make a meaningful comparison), and the ratio between corrections from no asset to positive and vice versa is also much lower (46/33 in cell B3 in 2008, compared to 312/48 in A3).

A tentative story behind the difference in corrections from no asset to positive and vice versa would be that, if a respondent indicates no asset in section Q or H in wave *t*, this is

the last question they will be asked about the asset if they do not enter section U for the asset, whereas if they indicate owning the asset, they will be asked follow-up questions about the value. Thus, if inadvertently ownership is misrecorded as no asset, there is no scope for correction within section Q/H, whereas if inadvertently ownership is misrecorded as owning the asset, the follow-up questions might trigger a correction within section Q/H. Hence, misclassification as "no asset" is more likely than misclassification as "owning" the asset. However, it is more complicated than this. Population rates of ownership (prior probabilities) factor also in this (posterior) probability. Furthermore, transition probabilities from ownership to no ownership and vice versa, as well as the distribution of asset values conditional on ownership, further determine the likelihood that the asset will be challenged in wave (t+1)'s section U.

In 5–9% of the observations that did not enter section U in wave t, but corrected the value in section U in wave t+1 (cell A3), the corrected value is equal to the uncorrected, so there is nothing to do, and in another 5–12%, one value is within the bracket of the other report, or they are overlapping brackets. In about 1.5% of the observations (10–18 per wave), the corrected and uncorrected values differ by a factor of exactly 10 or 100, so there seems to be a decimal place error. The remaining 25–40% are cases where a positive value or bracket was corrected to another positive value or bracket that was unequivocally higher or lower. In 2004–2008 and 2012, there are more increases than decreases, whereas in 2010 (reporting about 2008 assets), there were more decreases. The relative differences in these frequencies are not huge, though (e.g., 162 increases vs. 124 decreases in 2008; 61 increases vs. 70 decreases in 2010). This pattern follows the trends in wealth, so this suggests that the corrections are somewhat biased towards the current value. That is, when faced with a discrepancy, respondents may have a tendency to overcorrect the previous wave's value in the direction of the current wave's value. This leads to a small downward bias in the variance of first differences. If there is a linear trend, this downward bias would be partially offset by an upward bias for the difference between t and t-1. For example, assume that the true value of the asset is $100,000 in 2002, $120,000 in 2004, and $140,000 in 2006, but because of the mentioned bias, the value for 2004 is corrected to $130,000 in 2006. This reduces the variance of the 2004-2006 first differences but increases the variance of the 2002-2004 first differences. But similarly, the corrections in 2004 of the 2002 values will have a downward biasing effect of the 2002-2004 first differences, so the net effect on the 2002-2004 differences is ambiguous. However, if there is a trend reversal, as in the 2006-2008-2010 years, where 2008 is higher on average than both 2006 and 2010, the variances of both the 2006-2008 first difference and the 2008-2010 first difference will be slightly biased downward.

Based on this analysis, we implemented the following rules:

1. If a respondent entered section U in wave t for the same asset (cells B2, B3, C2, C3), we use wave t's section U response to correct (or not) the wave t section U asset value and we do not use the wave t+1 correction of that same value.
2. The only exception to the previous rule is when the wave t+1 section U correction of the wave t value differs from the wave t section U result by more than a factor 9, which might indicate a decimal place error (e.g., by the interviewer). We have inspected all cases where the difference was more than a factor 9 and developed

and implemented a set of rules for these cases. We then inspected all these cases and made some ad-hoc corrections where the rules selected what appeared to be the incorrect value. Appendix B gives the rules in detail and shows the exceptions.

3. If the respondent did not enter section U for the same asset in wave *t* (cell A3), we implemented the wave *t*+1 corrections of the wave *t* value.
4. For observations in which the "previous wave" is not the prior wave (2 years earlier) but an earlier wave (because the household did not participate in the prior wave), we do not use the corrections for the "previous wave". It is possible that corrections for the wave 4 years before the current interview are somewhat reliable, but, say, corrections in 2008-2010 for waves 1-4 (1992-1998) are probably highly unreliable. There are only a handful of cases that enter section U with the "previous wave" dating back 4 years or more. Assessing the reliability of those reports would be difficult.

5. Descriptive statistics of section U correction results

This chapter studies what the effect of the section U corrections is on the distributions of asset components and total wealth, as well as changes in these. All analyses are unweighted.

5.1 Frequencies of entering and correcting

Table 5.1 shows the number of households that entered section U for one or more assets by wave. There are a few reasons for the patterns we see (see section 2.3). In 2002, there was an issue with the preloads, which prevented households from entering section U if they had claimed no ownership in 2000. This explains the lower percentage for this year. In 2004-2008, there was the "comma problem" (no thousands separator in the amounts), which may have caused some decimal place errors. In later data models of 2010, the preloads used the 2008 section U corrected values instead of the original section Q or H values. This may have led to fewer discrepancies. The larger absolute number in 2012 is probably due to the larger eligible sample: this is the second wave for the Middle Baby Boomers cohort (age 51-56 in 2010), which could not enter section U in 2010. Another potential explanation would be the incorrect preloads used for some cases in 2012, as mentioned in section 2.3.

Table 5.1: Number of households entering section U

Wave year	Households interviewed	Households entered section U	Percent of hh that entered section U
2002	12,349	711	5.8
2004	13,645	1986	14.6
2006	12,605	2234	17.7
2008	11,897	2362	19.9
2010	15,280	2021	13.2
2012	14,316	2675	18.7

A household may enter section U for more than one asset, Allowing households to contribute multiple records if they were challenged on multiple assets, Table 5.2 shows the number of assets (household-asset combinations) that entered section U in each wave, and whether the respondent confirmed or corrected the value. In most waves, between 60 and 70 percent of the challenged asset values were confirmed, but this number was substantially lower in 2012. Corrections of the previous wave were more common, except in 2002 (and to a much lesser extent in 2004), corrections of the current wave were more common. The latter may be explained by the preload issue mentioned above, because a relatively large fraction of previous wave corrections are corrections from no ownership to a positive value, and these cases did not enter section U in 2002. The relative improvement of the previous wave in 2010 may be due to replacing the section Q or H preloads by their section U corrections (if applicable) in the later data models of 2010.

The high fraction incorrect, in particular previous wave incorrect, in 2012, is likely due to the preload problems in 2012 mentioned in section 2.3.

Table 5.3 shows the same results, but now broken down by asset type rather than wave. Assets that are more common, and that are more likely to have large values, such as the primary residence, are more likely to be subject to a section U challenge. However, the value of the primary residence is also most likely to be confirmed, implying that relatively large changes are considered accurate by respondents even when challenged. "Other assets" is least likely to be confirmed. This asset is elicited as the residual category and it is likely difficult for respondents to remember and tally up all remaining assets they own (or owned) within this category. It may also be difficult for respondent to know what to include in the first place in such a residual category, and to do so consistently across sections of the survey (asset section and later section U) and even across waves several years apart.

Table 5.2: Number of assets entering section U, and percentage confirmed and corrected, by wave

Wave year	Entered sec.U	Confirmed (% of enter)	Incorrect (% of enter)	Prev.wv incorrect (% of incorrect)	Curr.wv. incorrect (% of incorrect)	Both incorrect (% of incorrect)
2002	1478	61.4	32.9	38.2	54.0	0.6
2004	4097	62.9	34.0	41.8	44.2	10.8
2006	5414	69.5	28.7	52.9	36.4	7.7
2008	5640	65.4	32.3	56.5	35.4	6.4
2010	5071	64.5	32.2	45.4	44.2	5.4
2012	6504	48.4	48.2	69.8	22.7	4.2

Note: "Entered section U" is number of household-asset combinations for which section U was entered; categories "DK whether correct" and "DK which wave incorrect" not shown.

5.2 Effect of section U corrections on marginal distributions

Table 5.4 shows the mean and standard deviation of total wealth before and after section U corrections and the resulting imputations. Total wealth as defined here is the RAND HRS variable HwATOTA, which is the sum of the 18 assets from Table 2.1 (with liabilities counted as negative assets), except for trusts, the second home, and any loans on the second home, but including "other home loans" (on the first home), which are not part of section U.[5] We see that section U has a modest impact on the means, but a large impact on the standard deviations in some years. For 2004 and 2006, the standard deviation is much lower, because a few large outliers were corrected in section U. In 2002, the reverse happened: two large outliers were introduced in section U.

[5] The more comprehensive measure HwATOTB also includes the second home and any loans on it. However, the value of the second home and any loans on it cannot be consistently defined for all waves as mentioned in chapter 7. Therefore, we use the HwATOTA measure in chapter 7 and for consistency do the same here.

Table 5.3: Number of assets entering section U, and percentage confirmed and corrected, by asset (2002-2012 waves)

Asset	Entered sec.U	Confirmed (% of enter)	Incorrect (% of enter)	Prev.wv incorrect (% of incorrect)	Curr.wv. incorrect (% of incorrect)	Both incorrect (% of incorrect)
Other debt	294	60.5	36.1	57.5	30.2	10.4
Trusts	959	55.1	41.5	51.0	39.9	5.3
Other assets	1562	42.8	52.6	50.0	36.1	9.5
Vehicles	620	59.2	38.9	61.0	30.3	7.5
CDs	1257	64.4	31.4	45.6	43.8	6.1
Checking/saving	2335	68.4	28.4	48.9	44.3	3.8
Bonds	852	59.5	35.3	47.8	39.9	7.6
Stocks	3366	56.8	39.1	49.9	39.8	6.1
IRAs	3023	60.4	36.1	58.8	31.9	5.8
Business/farms	1645	57.8	40.5	53.1	39.0	6.0
Real estate	2508	55.9	41.0	53.6	34.5	8.4
Primary residence	5463	70.9	27.2	64.6	27.6	4.4
Mortg 1 prim. res.	1601	62.8	35.8	58.5	35.3	3.1
Mortg 2 prim. res.	77	58.4	39.0	56.7	43.3	0.0
Equity line of credit	412	72.8	26.9	82.9	12.6	2.7
Mobile home	107	49.5	48.6	55.8	38.5	5.8
Second home	1769	63.3	34.8	62.6	28.6	6.8
Mortg/ln 2nd home	354	61.3	36.2	46.1	42.2	6.3

Note: "Entered section U" is number of household-asset combinations for which section U was entered; categories "DK whether correct" and "DK which wave incorrect" not shown.

Table 5.4: Effect of section U corrections on total wealth

Wave	mean			s.d.			N
year	Without U	With U	Diff (%)	Without U	With U	Diff (%)	
2000	315,490	308,140	-2.3	972,906	921,939	-5.2	13,214
2002	314,297	325,437	3.5	838,382	1,151,907	37.4	12,349
2004	372,762	361,986	-2.9	1,425,276	1,052,745	-26.1	13,645
2006	469,880	426,678	-9.2	2,151,258	1,139,984	-47.0	12,605
2008	440,711	440,965	0.1	1,255,424	1,175,743	-6.3	11,897
2010	327,664	342,427	4.5	900,611	954,488	6.0	15,276
2012	345,585	338,405	-2.1	1,173,743	979,390	-16.6	14,314

Note: "Total wealth" is the sum of all assets (liabilities viewed as negative assets) except trusts and the net value of the second home. The large increase of the s.d. in 2002 is due to two households reporting very large values for "other assets" in section U: one of $30 million and one of $90 million.

Tables 5.5 and 5.6 show the analogous results for two assets of specific interest, the main home and stocks, respectively. These are common assets with large values, and these are the two assets with the most challenges in section U. Thus these are assets for which we may expect the biggest impact of section U. Apart from their role in total wealth, these

are also the assets that are arguably most interesting for separate study by researchers. The results largely confirm the results for total wealth: the impact on the mean is modest—although noticeably larger for stocks—and the impact on the standard deviations tends to be bigger. Here, the standard deviations are generally lowered. The largest effects are on the main home in 2006 and 2008 and stocks in 2004 and 2006. This is most likely related to the comma problem mentioned earlier.

Table 5.5: Effect of section U corrections on the value of the primary residence

Wave	Mean			s.d.			N
year	Without U	With U	Diff (%)	Without U	With U	Diff (%)	
2000	98,786	98,650	-0.1	191,077	190,248	-0.4	13,214
2002	114,844	114,552	-0.3	186,592	186,090	-0.3	12,349
2004	144,613	142,047	-1.8	357,622	309,348	-13.5	13,645
2006	184,750	171,830	-7.0	863,318	262,000	-69.7	12,605
2008	183,912	176,517	-4.0	623,867	426,395	-31.7	11,897
2010	146,826	145,914	-0.6	244,613	241,051	-1.5	15,276
2012	138,389	137,497	-0.6	256,445	229,036	-10.7	14,314

Note: "Primary residence" as defined here does not include mobile homes, and is set to zero if the household does not own their primary residence.

Table 5.6: Effect of section U corrections on wealth in stocks

Wave	mean			s.d.			N
year	Without U	With U	Diff (%)	Without U	With U	Diff (%)	
2000	63,598	59,931	-5.8	468,567	409,448	-12.6	13,214
2002	48,707	48,381	-0.7	233,563	221,133	-5.3	12,349
2004	74,468	62,595	-15.9	1,044,630	443,469	-57.5	13,645
2006	72,205	64,749	-10.3	852,046	439,230	-48.5	12,605
2008	64,241	65,489	1.9	371,718	348,328	-6.3	11,897
2010	46,571	49,826	7.0	292,523	316,332	8.1	15,276
2012	56,439	53,625	-5.0	578,795	461,707	-20.2	14,314

5.3 Effect of section U corrections on the distributions of first differences

The main impetus for the introduction of section U in the questionnaire was the hope that this would reduce the variance in the wave-to-wave changes, which was thought to be too high because of relatively large measurement errors. With that in mind, we now look at the distribution of wave-to-wave changes (first differences) before and after the section U corrections. Table 5.7 presents the means and standard deviations of the first differences in total wealth (as defined in section 5.2). In Table 5.4, we saw that mean total wealth was between $315,000-470,000. As expected, the mean change is of a much smaller magnitude. The impact of the section U corrections on the mean change is relatively large compared to this smaller base, but modest compared to mean total wealth. In contrast, the

standard deviation of total wealth was between $840,000-$2,100,000 in Table 5.4, and we see that the standard deviation of the change in total wealth is of a similar order of magnitude. In most years, the section U corrections result in a substantial reduction of this standard deviation of the change in wealth. The exception is 2000-2002. The increase in standard deviation there is largely due to the two outliers introduced by section U in 2002, as mentioned in section 5.2.

Table 5.7: Effect of section U corrections on first differences in total wealth

Wave	mean			s.d.			N
years	Without U	With U	Diff (%)	Without U	With U	Diff (%)	
1998-2000	44,652	36,944	-17.3	1,055,494	1,011,108	-4.2	12,843
2000-2002	-8,311	11,127	-233.9	887,338	1,154,082	30.1	11,815
2002-2004	62,587	36,514	-41.7	1,357,598	1,233,760	-9.1	11,144
2004-2006	91,832	58,328	-36.5	1,970,696	851,753	-56.8	12,263
2006-2008	-38,813	1,331	-103.4	1,786,520	957,765	-46.4	11,442
2008-2010	-78,903	-59,592	-24.5	1,087,506	837,102	-23.0	10,550
2010-2012	9,657	-12,362	-228.0	1,001,104	759,602	-24.1	13,941

Note: "Total wealth" is the sum of all assets (liabilities viewed as negative assets) except trusts and the net value of the second home. The large increase of the s.d. in 2002 is due to two households reporting very large values for "other assets" in section U: one of $30 million and one of $90 million.

Tables 5.8 and 5.9 show the effects of the section U corrections on the means and standard deviations of the value of the main home and stock wealth. The differences in the means are again relatively modest (in absolute size), although certainly not trivial. But the effects on the standard deviations are generally much larger: in most years the standard deviations are drastically reduced. Notable exceptions are the 1998-2000 and 2000-2002 changes, for which the standard deviations were already much smaller before section U, and which are not much affected by the section U corrections. This may be due to the absence of the comma problem in these years.

Table 5.8: Effect of section U corrections on first differences in value of the primary residence

Wave	mean			s.d.			N
years	Without U	With U	Diff (%)	Without U	With U	Diff (%)	
1998-2000	8,333	8,226	-1.3	161,271	160,971	-0.2	12,843
2000-2002	13,846	13,804	-0.3	183,124	182,081	-0.6	11,815
2002-2004	23,537	20,155	-14.4	333,785	278,966	-16.4	11,144
2004-2006	38,031	27,919	-26.6	878,291	270,570	-69.2	12,263
2006-2008	-12,463	-1,279	-89.7	865,751	324,300	-62.5	11,442
2008-2010	-38,880	-30,493	-21.6	598,783	375,171	-37.3	10,550
2010-2012	-10,288	-10,470	1.8	200,497	156,531	-21.9	13,941

Note: "Primary residence" as defined here does not include mobile homes, and is set to zero if the household does not own their primary residence.

Table 5.9: Effect of section U corrections on first differences in stock wealth

Wave	mean			s.d.			N
years	Without U	With U	Diff (%)	Without U	With U	Diff (%)	
1998-2000	17,416	13,571	-22.1	451,735	389,704	-13.7	12,843
2000-2002	-16,644	-13,008	-21.8	378,484	320,221	-15.4	11,815
2002-2004	31,641	17,276	-45.4	1,080,786	392,999	-63.6	11,144
2004-2006	-5,115	426	-108.3	851,380	443,162	-47.9	12,263
2006-2008	-8,263	1,614	-119.5	822,700	411,672	-50.0	11,442
2008-2010	-7,791	-5,087	-34.7	330,297	294,017	-11.0	10,550
2010-2012	8,524	2,452	-71.2	571,788	472,261	-17.4	13,941

5.4 Ownership changes and reversals

Changes in ownership form an important component of wave-to-wave changes in wealth. If ownership is recorded incorrectly, this may have a large impact on recorded wealth and especially recorded changes in wealth. Table 5.10 presents the number of ownership changes before and after section U corrections. In most cases, section U reduces the number of ownership changes, although compared to the total number of assets (more than 200,000 household-asset combinations), the net reduction is modest.

Table 5.10: Effect of section U corrections on ownership changes (across all assets)

Wave	own to not own			not own to own			N
years	Without U	With U	Diff (%)	Without U	With U	Diff (%)	
1998-2000	12,474	12,465	-0.1	11,735	11,730	-0.0	269,703
2000-2002	12,783	12,602	-1.4	9,865	9,970	1.1	248,115
2002-2004	10,792	10,615	-1.6	10,112	9,936	-1.7	234,024
2004-2006	12,554	12,229	-2.6	10,628	10,413	-2.0	257,523
2006-2008	11,457	11,174	-2.5	9,605	9,324	-2.9	240,282
2008-2010	11,418	11,141	-2.4	8,663	8,418	-2.8	221,550
2010-2012	14,338	14,114	-1.6	10,461	9,954	-4.8	292,761

Table 5.11 shows analogous results for ownership reversals, that is, households that saw an ownership change and then a reversal of that change in the subsequent wave. Such a reversal could be an indication of an error in the middle wave, especially for assets that are not often acquired or disposed of, such as IRAs (Venti, 2011). These results are in line with the results in Table 5.10, but stronger.

Table 5.11: Effect of section U corrections on ownership reversals (across all assets)

Wave	not own to own to not own			own to not own to own			N
years	Without U	With U	Diff (%)	Without U	With U	Diff (%)	
1995-2000	1,185	1,183	-0.2	963	960	-0.3	70,433
1996-2000	2,563	2,558	-0.2	2,275	2,270	-0.2	111,777
1998-2002	5,560	5,510	-0.9	3,654	3,680	0.7	242,004
2000-2004	4,469	4,401	-1.5	4,020	3,911	-2.7	225,099
2002-2006	4,634	4,422	-4.6	3,195	3,060	-4.2	211,344
2004-2008	4,927	4,699	-4.6	3,803	3,583	-5.8	234,297
2006-2010	4,611	4,351	-5.6	3,442	3,280	-4.7	214,074
2008-2012	4,249	4,037	-5.0	3,187	2,970	-6.8	201,453

Note. In 1995 and 1996, we do not have measures for the second home and any loans or mortgages on it. Therefore, the number of assets compared in the 1995-2000 and 1996-2000 rows are smaller.

5.5 Spikes and trenches in values

Section U was intended to catch conspicuously large changes in asset values. As with ownership reversals, two consecutive large amount changes in opposite directions are indicative of errors in the middle of the three waves involved. To study the frequencies of these phenomena, we define a *spike* as a set of three consecutive positive values for an asset in which the middle value is larger than 10 times the average of the two outer values. Analogously, a *trench* is a set of three consecutive positive values in which the middle value is smaller than one tenth of the average of the two outer values. Table 5.12 shows the frequencies of spikes and trenches by wave triplet. With three exceptions, section U reduces the number of spikes and trenches.

Table 5.12: Effect of section U corrections on spikes and trenches in values of individual assets (across all assets)

Wave	spikes			trenches			N
years	Without U	With U	Diff (%)	Without U	With U	Diff (%)	
1995-2000	156	161	3.2	465	461	-0.9	70,433
1996-2000	251	253	0.8	768	766	-0.3	111,777
1998-2002	542	511	-5.7	1,445	1,493	3.3	242,004
2000-2004	536	508	-5.2	1,489	1,430	-4.0	225,099
2002-2006	509	477	-6.3	1,413	1,362	-3.6	211,344
2004-2008	569	538	-5.4	1,511	1,448	-4.2	234,297
2006-2010	498	461	-7.4	1,352	1,315	-2.7	214,074
2008-2012	383	361	-5.7	1,140	1,063	-6.8	201,453

Note. In 1995 and 1996, we do not have measures for the second home and any loans or mortgages on it. Therefore, the number of assets compared in the 1995-2000 and 1996-2000 rows are smaller.

Table 5.13 shows the effect of the section U corrections on spikes and trenches in total wealth. The results are a bit more mixed than for individual assets, but on the whole suggest that section U has a dampening effect.

Table 5.13: Effect of section U corrections on spikes and trenches in total wealth

Wave	spikes			trenches			N
years	Without U	With U	Diff (%)	Without U	With U	Diff (%)	
1995-2000	22	23	4.5	103	100	-2.9	3,707
1996-2000	17	17	0.0	89	93	4.5	5,883
1998-2002	68	63	-7.4	229	240	4.8	11,524
2000-2004	66	69	4.5	211	192	-9.0	10,719
2002-2006	50	42	-16.0	200	191	-4.5	10,064
2004-2008	63	61	-3.2	232	206	-11.2	11,157
2006-2010	63	51	-19.0	205	196	-4.4	10,194
2008-2012	51	65	27.5	209	186	-11.0	9,593

Note. In 1995 and 1996, we do not have measures for the second home and any loans or mortgages on it. Therefore, the number of assets compared in the 1995-2000 and 1996-2000 rows are smaller.

A common characteristic of these two tables is that trenches are much more common than spikes. Accidental underreporting thus appears more common than accidental overreporting.

6. Cross-wave imputation methodology

Imputations of missing asset values in the HRS are greatly facilitated by the fact that HRS asks follow-up questions in case the respondent does not report a continuous value. to find whether the value is greater or smaller than some threshold and may follow-up with additional thresholds if the respondent is willing to answer these types of questions, so called "unfolding brackets." See chapter 1 and appendix D for more information about these bracket questions. Information elicited in the unfolding bracket sequences, in most cases greatly limits the amount of uncertainty about the value of the asset by narrowing the range within which a value needs to be imputed. RAND provides imputations as part of the RAND HRS. An important feature of these imputations is the multi-stage approach, in which ownership, bracket, and value are imputed sequentially if necessary, and the use of nearest-neighbor imputation within closed brackets, which greatly reduces sensitivity to model specification; in particular distributional assumptions are not necessary and not used for value imputation within closed brackets. (They are, however, used for the open upper bracket, to limit the potential disproportional influence of outliers in the reported data.)

In each of the stages, the model uses the same set of covariates, which are the first 10 principal components of about 30 explanatory variables. (See Chien et al., 2015, section 3.2.4, for a list.) These covariates are predictive of the variation in the dependent variable. Appendix E provides more detail about the imputation methods as implemented for the cross-sectional imputations, most of which also applies to the cross-wave imputations. However, additional predictive power can be gained from incorporating information contained in the values of the asset in adjacent waves. If the primary residence was reported to have been worth $300,000 in the last wave, and $350,000 in the next wave, then these values could be included in the regressions as covariates to help predict what the primary residence was likely worth between these two waves. In this chapter, we develop the details of these cross-wave imputations.

More specifically, the core of each of the imputation models is a linear index function, a linear combination of the covariates. The covariates in the RAND imputations are the first 10 principal components of a larger set of explanatory variables. In the cross-wave imputations, we add four covariates to these: (1) a dummy for the previous wave's ownership of the asset; (2) the inverse hyperbolic sine transform of the previous wave's value of the asset (zero in case of no ownership); (3) a dummy for the following wave's ownership of the asset; and (4) the inverse hyperbolic sine transform of the following wave's value of the asset (zero in case of no ownership). In case any of these four variables are missing, this raises the issue of how to proceed. If a household was present in the adjacent waves but did not report a continuous value of one or more relevant variables, we do have the existing (cross-sectional) imputations of them, so in estimating the cross-wave imputation models and performing the actual imputations we use these cross-sectional imputations as covariates instead. If a household was not present in one of the adjacent waves, we do not have such imputations. Therefore, we estimate two additional imputation models: one with only the previous wave's covariates (1 and 2) added and one with only the following wave's covariates (3 and 4) added. We impute

using the model that uses the most information available for the household. For households for which neither previous wave information nor following wave information is available, we use the cross-sectional imputations.

Changes in couple status have large consequences for household wealth. Therefore, for the cross-wave imputations we add dummies for divorce/separation, widowing, and (re)marriage to the list of covariates (six additional covariates if the household is present in both adjacent waves, three if only in one adjacent wave). However, because some of these changes are not very common, we may not have enough observations to estimate their coefficients for each wave separately. Therefore, we pool the data for all waves and estimate these coefficients for all waves jointly, that is, we assume that these coefficients are the same for all waves. Because we would like to have wave-specific coefficients for all other covariates, and in particular wave-specific auxiliary parameters (such as bracket threshold parameters), this caused some technical complications. As a result, we had to write a new bracket imputation program.

6.1 Model specification

All models and imputations are asset-specific, so we suppress the asset indicator in the following and abstractly talk about "the asset". Let y_{it} be the outcome variable of interest for household i in wave t. This is the inverse hyperbolic sine transform of the value of the asset. Analogously, let the binary variable d_{it} denote whether the household owns the asset. In the cross-sectional imputations the set of covariates x_{it} is the set of the first 10 principal components of a large set of potential covariates. We use the same x_{it} in the cross-wave imputations.

For each wave and each asset, 10 imputation models are estimated. Five of these are used for imputations for households with a financial respondent, and the other five are used when there is no financial respondent. The difference is that the set of variables that is used for computing the principal components differs, and thus the x variables are not comparable. The five models are (1) a binary logit for ownership; (2) an ordered logit for the complete bracket conditional on ownership; (3) a linear regression for the value to be used for nearest neighbor imputation within closed brackets; (4) a tobit model ("tobit25") for the value to be used for imputing in the upper (open) bracket, and (5) a loglinear regression model (called "tobit0" in Appendix E, because that is how it is estimated) for the value to be used for imputing in certain problem sitations, for example, when there are no continuous observations in the reported bracket, so the nearest neighbor method cannot be used. These five models are estimated for two sets of covariates, depending on whether the model is to be used for households with a financial respondent or not. In some cases with few continuous observations, we use a conditional hotdeck method.

The tobit25 model censors values in the lower quartile, which has the effect of better fitting the distribution in the right tail, which leads to better imputations in the upper bracket than if the uncensored linear regression model (3) were used.

Each of these models is based on a single linear index $v_{it} = x_{it}\,\beta_t$ (the coefficients are different for the different models, but we suppress that in the notation here). The ordered logit model additionally estimates threshold parameters, and the tobit and loglinear regression models additionally estimate a variance parameter. Because the models are estimated separately for each wave, these additional parameters are also wave-specific.

For the cross-wave imputations, we augment the linear indexes of these models with a few additional terms: dummies for ownership of the asset in the previous and next wave, the values of the asset in the previous and next wave (zero if no ownership), and dummies for marital status changes. Because the sample sizes for the latter are too small to estimate separate coefficients for all combinations in all waves, the coefficients for marital status change are assumed to be the same for all waves. Thus, the linear indexes we use for the cross-wave imputation models are of the form

$$v_{it} = x_{it}\,\beta_t + d_{i,t\text{-}1}\,\alpha_{1t} + d_{i,t+1}\,\alpha_{2t} + y_{i,t\text{-}1}\,\rho_{1t} + y_{i,t+1}\,\rho_{2t} + m_{it}\,\delta_1 + m_{i,t+1}\,\delta_2, \qquad (1)$$

where m_{it} is a vector of dummies for marital status changes between wave t - 1 and wave t. We classify marital status in the following categories: (1) married or unmarried couple; (2) divorced/separated; (3) widowed; and (4) never married. We use "no change" as the reference category, and include in m_{it} separate dummies for changes from 1 to 2 ("divorce"), 1 to 3 ("widow"), and 2, 3, or 4 to 1 ("(re)marry"), and consider all other potential combinations "no change", so these are three dummies.

Thus, the cross-sectional imputation models estimate (1) with only the first term on the right-hand side included, and the full longitudinal imputation models estimate it with all terms included. Because there are substantial numbers of observations in which the household is present in the next wave but not the previous or vice versa (e.g., when a cohort was introduced, the latest observed wave in the data, or mortality), we also estimate variants of (1) that include the previous wave covariates and not the next wave covariates or vice versa. This gives us four versions of each of the 10 imputation models.

In some cases, the number of reported observations on a variable is insufficient to estimate the imputation model accurately. Then we simply use the marginal distribution, which is the same as estimating the model without covariates. See Appendix E for details.

6.2 Estimation sample

The ownership dummies and asset values for the previous wave or the next wave may be missing themselves. If this is the case, we will use the cross-sectional imputations of those as covariates instead. However, if the household did not respond at all in the previous wave or next wave, the corresponding ownership and asset values have not been imputed, and we will exclude these households from estimation for the models that require them.

We estimate the models for all waves jointly, on the stacked data with all observations included for which the required information is available. By interacting all explanatory variables, except the marital status change dummies, with a set of wave dummies, we allow all coefficients to be wave-specific, except the coefficients of the marital status change dummies. The wave dummies are coded such that the AHEAD and HRS waves (2A and 2H, 3A and 3H) are considered separate waves. Special care needs to be taken to assure that the other parameters in the models are also wave specific. This includes the threshold parameters in the ordered logit models for the brackets and the variance parameter in the tobit models. Also, the tobit25 model must allow a wave-specific censoring threshold.

6.3 Imputation

Cross-wave imputations proceed in almost exactly the same way as the cross-sectional imputations, with the main difference being that the augmented linear index v_{it} is used, adding the previous and/or next wave covariates depending on whether the household was present in the previous and/or next wave. Thus, we take the following steps:

1. If ownership is missing, compute the predicted probability p_{it} of ownership based on the imputation model. Take a draw u_{it} from the uniform distribution on the interval (0, 1]. Assign ownership if $u_{it} \leq p_{it}$ and non-ownership otherwise.
2. If ownership is given or imputed, but no brackets are given, compute the predicted probabilities p_{itj} of being in the j-th bracket based on the imputation model. If the respondent gave some bracket information, but not complete, the probabilities p_{itj} that are consistent with the respondent's answers are scaled such that they sum to 1, and the probabilities of the other brackets are set to zero. Then compute the resulting cumulative probabilities $P_{itj} = \sum_{k=1}^{j} p_{itk}$. Take a draw z_{it} from the uniform distribution on the interval (0, 1] and assign bracket j if $P_{it,j-1} < z_{it} \leq P_{itj}$.
3. If a closed complete bracket is given or imputed, but no continuous value is given, compute the value of the linear index v_{it} of the linear regression model for all households. From the households that (in the same wave) report a value within the bracket of interest, select the household whose linear index is closest to the one of the household whose value must be imputed. Impute the continuous value reported by this (nearest neighbor, donor) household into the household for which we only have the bracket.
4. If the open upper bracket is given or imputed, but no continuous value is given, compute the value of the linear index v_{it} of the tobit (tobit25) model. Let c_t be the threshold value of the upper bracket (i.e., its lower endpoint) and σ_t the estimated standard deviation of the tobit model. Draw y_{it} from the truncated normal distribution with mean v_{it}, standard deviation σ_t, lower truncation point c_t, and no upper truncation. That is, compute $y_{it} = v_{it} + \sigma_t \, \Phi^{-1}[(1 - \varepsilon_{it}) \, \Phi((c_t - v_{it})/\sigma_t) + \varepsilon_{it}]$, where ε_{it} is again a draw from the standard uniform distribution.

There are some alternatives for special cases in which the above methods cannot be used, and a conditional hot-deck is used, or the amount is imputed from the loglinear regression model (tobit0). See Appendix E for a list of these situations and which imputation method is used in which case.

6.4 Missing covariates and differences in which assets are measured

There are three sets of covariates: (i) same-wave covariates; (ii) previous-wave covariates; and (iii) next-wave covariates. The same-wave covariates are the principal components, which are never missing. The previous-wave covariates are previous-wave ownership and value of the asset, and marital status change between the previous wave and the current wave. These are missing if the household did not participate in the previous wave (or if there was no previous wave, in the case of wave 1). Analogously, the next-wave covariates are the next wave's ownership and value of the asset and marital status changes between the current and the next wave, and these are missing if the household did not participate in the next wave (or if there is no next wave yet, in the case of the last available wave).

If ownership or value of the previous or next wave's asset is missing, but the household did participate in those waves, (cross-sectional) imputations of these are available, and we use these instead of observed variables. Marital status is missing for a small number of observations and below we have developed rules to assign the most reasonable value for this (see section 6.5).

With these choices, there are four sets of missingness patterns: (a) all covariates present; (b) previous-wave covariates missing; (c) next-wave covariates missing; and (d) both previous and next-wave covariate missing. Case (a) is the full model that follows the description above. For case (d), only cross-sectional information is available and thus we use the cross-sectional imputations. For cases (b) and (c), we use the partial models that only use the next (b) or previous (c) wave's values as covariates.

There are slightly over 100 HRS-AHEAD overlap households. These are couples in which one member was HRS age-eligible and the other AHEAD age-eligible, which were transfered from the HRS to AHEAD after wave 1. Thus, these were present in wave 1, but after that considered part of the AHEAD cohort. Strictly speaking, this would give us previous-wave information about them in wave 2A (AHEAD wave 1), but because this is such a small sample, we do not use this and treat them as not being in the previous wave when doing the wave 2A imputations. Analogously, we treat them as having no next-wave information in wave 1. Table 6.1 gives the resulting sample sizes (number of households per wave) in each of the four information groups by wave.

For a few assets, previous- or next-wave values are not available for a reason other than unit nonresponse. The RAND HRS does not include values for the second home and any mortgages or loans on it in waves 2A (AHEAD wave 1, 1993) and 3 (HRS wave 3, 1996, and AHEAD wave 2, 1995), because of comparability issues in the raw data, resulting

from problems with the questionnaire and skip patterns (Chien et al., 2015). Therefore, we treat these as missing for all households. Hence, in wave 2H (HRS wave 2), we do not have observations with next-wave information for these assets and in wave 4 we do not have observations with previous-wave information.

Table 6.1: Patterns of availability for cross-wave imputations

	Availability of previous wave and next wave				
Wave year	Both missing	Previous wave missing	Next wave missing	Both present	Total
1992	730	6,972	0	0	7,702
1993	831	5,216	0	0	6,047
1994	0	1	531	6,519	7,051
1995	0	1	806	4,415	5,222
1996	53	186	418	6,157	6,814
1998	374	3,211	1,163	9,647	14,395
2000	85	222	1,361	11,546	13,214
2002	127	390	1,104	10,728	12,349
2004	324	2,159	1,090	10,072	13,645
2006	72	239	1,137	11,158	12,606
2008	102	329	1,256	10,210	11,897
2010	415	4,277	969	9,617	15,278
2012	306	0	14,009	0	14,315
Total	3,419	23,203	23,844	90,069	140,535

In waves 1 and 2, only the total of all IRAs was asked, whereas from wave 3 onward, the respondents are asked about the largest two IRAs separately and then about any remaining IRAs (which we call "IRA3"). For imputing the wave 2 total IRA amount, we use the sum of the three IRA variables in wave 3 as the next-wave value. For imputing the individual wave 3 IRAs, we use the wave 2 total as the previous-wave value for IRA1, but we leave the previous-wave values for IRAs 2 and 3 missing.

Wave 1 does not ask about trusts (see appendix A), so for wave 2H, the previous-wave value is missing.

6.5 Marital status change, unit nonresponse, and source of adjacent-wave covariates

As stated in section 6.1, we include dummies for marital status changes in the models that use adjacent-wave information as covariates. There are some problems with using the respondent-reported marital status variable for determining marital status changes: there are some missings; there are changes from partnered to never married, which could either denote a breakup ("divorce/separation") or the death of the partner ("widowing"); and

other changes that do not seem to reflect the change of interest well. Additional issues arise when the respondent is not present in the current or next wave. Related to this is what the source of the previous or next wave's asset information is if the respondent was not present in the previous or next wave: wealth is a household-level variable, and thus the adjacent-wave wealth information could be obtained from the spouse's record. It turns out that these issues are not easily solved.

We have developed a robust assignment algorithm for marital status changes, and for the adjacent wave asset values. The main element of this is whether the HRS treats the household as a couple household. This not only reflects our impression that this is the most reliable indicator of couple status, but also that it indicates whether the asset questions reflect the assets of only the respondent, or the respondent and the spouse. Thus, a change in couple status reflects a change in couple's assets versus individual's assets. Whether a change from couple to single is a divorce/separation or widowing is then first checked by looking at whether the spouse (who is also a potential respondent) is still alive. If this is unknown or ambiguous, the respondent's reported marital status is checked.

If the respondent was not present in the previous wave, the spouse may have been present in the previous wave, and similarly for the next wave, and information about adjacent-wave's assets may be obtained from the spouse's record. Because the current spouse may not be the adjacent-wave spouse and vice versa, the algorithm checks the records of the person the respondent in the current wave reports as the spouse, but also checks whether another respondent claims the respondent of interest as his or her spouse in the adjacent wave. Note that, for example, due to different timing of interviews, in any given wave, the respondent may not be the spouse's spouse and vice versa. Our algorithm takes all these issues into account and makes educated guesses in cases of ambiguity. The algorithm is described in detail in appendix C.

The imputation models are estimated, and imputations are performed, at the household level. However, marital status change may be different for both spouses from a certain household in a certain wave (e.g., one may remarry, the other not). Therefore, we have to designate a target respondent in each household in each wave. It seems most logical to use the financial respondent as the target respondent. However, in some cases, a household has two financial respondents, or a two-person household has none. Across all waves, there are four cases of the former and 17 of the latter. For these cases, we designate the respondent with the variable HwPICKHH set to 1 (which is the respondent with the lowest "person number" if there is not a unique financial respondent) as the target respondent.

Table 6.2 shows the resulting numbers of households with each type of marital status change by wave and Table 6.3 indicates whose records the adjacent-wave asset information is obtained from.

Table 6.2: Marital status changes (households)

Wave year	No change	Divorce	Widowing	(Re)marry	No info	Total
			Compared to previous wave			
1992	0	0	0	0	7,702	7,702
1993	0	0	0	0	6,047	6,047
1994	6,643	159	145	103	1	7,051
1995	4,832	13	341	35	1	5,222
1996	6,167	141	153	114	239	6,814
1998	10,006	147	527	130	3,585	14,395
2000	12,073	139	547	148	307	13,214
2002	11,032	143	518	139	517	12,349
2004	10,469	118	419	156	2,483	13,645
2006	11,510	154	472	159	311	12,606
2008	10,741	162	432	131	431	11,897
2010	9,822	143	466	155	4,692	15,278
2012	13,121	274	423	191	306	14,315
Total	106,416	1,593	4,443	1,461	26,622	140,535
			Compared to next wave			
1992	6,633	91	145	103	730	7,702
1993	4,832	8	341	35	831	6,047
1994	6,157	97	153	113	531	7,051
1995	4,009	20	352	35	806	5,222
1996	5,996	74	175	98	471	6,814
1998	12,066	87	547	158	1,537	14,395
2000	11,028	79	518	143	1,446	13,214
2002	10,464	69	419	166	1,231	12,349
2004	11,496	97	473	166	1,413	13,645
2006	10,735	97	432	134	1,208	12,606
2008	9,817	89	466	167	1,358	11,897
2010	13,102	169	423	200	1,384	15,278
2012	0	0	0	0	14,315	14,315
Total	106,335	977	4,444	1,518	27,261	140,535

Table 6.3: Source of asset information for the adjacent waves

Wave year	No info	Respondent	Current wave spouse	Adjacent wave spouse	Total
		From previous wave			
1992	7,702	0	0	0	7,702
1993	6,047	0	0	0	6,047
1994	1	7,029	20	1	7,051
1995	1	5,209	4	8	5,222
1996	239	6,537	29	9	6,814
1998	3,585	10,770	32	8	14,395
2000	307	12,843	51	13	13,214
2002	517	11,815	4	13	12,349
2004	2,483	11,144	12	6	13,645
2006	311	12,264	23	8	12,606
2008	431	11,442	19	5	11,897
2010	4,692	10,550	27	9	15,278
2012	306	13,944	52	13	14,315
Total	26,622	113,547	273	93	140,535
		From next wave			
1992	730	6,860	112	0	7,702
1993	831	5,079	137	0	6,047
1994	531	6,380	140	0	7,051
1995	806	4,267	149	0	5,222
1996	471	6,205	138	0	6,814
1998	1,537	12,540	318	0	14,395
2000	1,446	11,482	286	0	13,214
2002	1,231	10,872	244	2	12,349
2004	1,414	11,967	263	1	13,645
2006	1,209	11,133	264	0	12,606
2008	1,358	10,277	262	0	11,897
2010	1,384	13,633	261	0	15,278
2012	14,315	0	0	0	14,315
Total	27,263	110,695	2,574	3	140,535

7. Descriptive statistics of cross-wave imputations

7.1 Frequencies of imputations

A first indication of the potential for improved imputations having a noticeable effect is given in Table 7.1. It shows the number of all household-asset level observations per wave by imputation status. The vast majority of assets are reported as not owned or as a continuous answer. Among the remaining reports, a sizable fraction result in a complete bracket, that is, the respondent completed the unfolding bracket sequence. This greatly limits the amount of imputation uncertainty and thus the scope for improved imputations to make a big difference. However, there are also a sizable number of observations in which no value or bracket was given (but the respondent reports owning the asset) or in which it is not known whether the household owns the asset. In these cases, the imputation uncertainty is very large, and improvements in imputation method may have a large impact. This is especially the case when considering cross-wave differences in asset values: asset values are often strongly serially correlated, and not taking this into account may lead to much larger differences in the imputations. The cross-sectional imputations are serially correlated because the covariates used are serially correlated, but the cross-wave imputations are intended to reflect the serial correlation better because they also allow for *conditional* correlation, after controlling for the covariates.

Table 7.1: Number of household/asset level observations by imputation status

Wave year	No asset	Continuous	Complete brkt	Incomplete brkt	No value or brkt	DK own/ no finR
1992	92,698	29,517	4,069	264	1,575	2,811
1993	70,073	13,299	3,992	330	865	2,146
1994	85,011	27,223	3,840	184	897	2,712
1995	77,711	12,877	2,626	1,857	1,559	2,588
1996	90,984	28,446	4,202	246	1,855	3,676
1998	229,807	50,188	10,178	738	4,046	7,338
2000	210,573	47,237	8,345	790	4,476	6,073
2002	197,899	43,221	7,172	620	4,851	5,566
2004	217,606	49,103	8,368	671	5,190	5,607
2006	201,645	46,230	7,032	655	4,599	4,544
2008	190,231	43,230	6,613	704	4,121	4,938
2010	242,073	53,888	6,760	983	3,783	13,309
2012	234,323	48,560	6,076	812	3,464	7,359

7.2 Effect of cross-wave imputations on marginal distributions

In the remainder of this chapter, we compare distributions in the HRS data using cross-sectional imputations (after section U corrections as described in chapter 5) with the corresponding distributions using cross-wave imputations (also incorporating the section U corrections).

Table 7.2 shows the effect of cross-wave imputation on the marginal distribution of total wealth. There are some modest differences in means and standard deviations, but there does not seem to be a systematic pattern in sign or magnitude of these. This confirms our expectations: the cross-wave imputations are expected to increase the serial correlation, but not necessarily the marginal distributions.

Table 7.2: Effect of cross-wave imputations on total wealth

	mean			s.d.			N
Wave year	Before cross-wave	After cross-wave	Diff (%)	Before cross-wave	After cross-wave	Diff (%)	
1992	193,778	185,424	-4.3	430,782	438,150	1.7	7,702
1993	156,624	154,978	-1.1	363,241	364,318	0.3	6,047
1994	220,009	222,602	1.2	482,419	517,794	7.3	7,051
1995	234,111	230,704	-1.5	855,267	835,696	-2.3	5,222
1996	247,321	246,444	-0.4	565,812	579,417	2.4	6,811
1998	265,818	268,749	1.1	984,536	997,645	1.3	14,395
2000	308,140	303,461	-1.5	921,939	896,656	-2.7	13,214
2002	325,437	322,249	-1.0	1,151,907	1,153,152	0.1	12,349
2004	361,986	364,538	0.7	1,052,745	1,080,659	2.7	13,645
2006	426,678	431,312	1.1	1,139,984	1,212,187	6.3	12,605
2008	440,965	434,116	-1.6	1,175,743	1,156,295	-1.7	11,897
2010	342,427	338,436	-1.2	954,488	962,490	0.8	15,276
2012	338,405	341,797	1.0	979,390	1,007,010	2.8	14,314

Note: "Total wealth" is the sum of all assets (liabilities viewed as negative assets) except trusts and the net value of the second home.

Table 7.3 shows the analogous comparisons for the value of the primary residence. The effects on the means are negligible. For some years, the standard deviations are modestly higher after cross-wave imputation and for others, the effect is minimal. Table 7.4 presents the analogous results for stock wealth. The percentage differences are larger for stocks, especially in the earliest waves, but again, without a systematic pattern. Part of the reason why the percentage differences are larger is that fewer households own stocks, so sample sizes for the positive amounts are smaller, and the distribution is more skewed, which shows in the tables as smaller means and larger standard deviations.

Table 7.3: Effect of cross-wave imputations on the value of the primary residence

	mean			s.d.			N
Wave year	Before cross-wave	After cross-wave	Diff (%)	Before cross-wave	After cross-wave	Diff (%)	
1992	78,326	78,472	0.2	96,509	97,175	0.7	7,702
1993	64,898	65,395	0.8	121,023	124,335	2.7	6,047
1994	84,878	84,960	0.1	119,370	119,785	0.3	7,051
1995	64,569	65,433	1.3	117,505	124,664	6.1	5,222
1996	89,289	89,788	0.6	110,710	117,018	5.7	6,811
1998	87,305	87,348	0.0	151,340	151,705	0.2	14,395
2000	98,650	98,984	0.3	190,248	191,194	0.5	13,214
2002	114,552	114,484	-0.1	186,090	188,596	1.3	12,349
2004	142,047	142,946	0.6	309,348	314,622	1.7	13,645
2006	171,830	173,407	0.9	262,000	268,537	2.5	12,605
2008	176,517	176,557	0.0	426,395	426,537	0.0	11,897
2010	145,914	146,274	0.2	241,051	252,732	4.8	15,276
2012	137,497	137,534	0.0	229,036	229,415	0.2	14,314

Note: "Primary residence" as defined here does not include mobile homes, and is set to zero if the household does not own their primary residence.

Table 7.4: Effect of cross-wave imputations on wealth in stocks

	Mean			s.d.			N
Wave year	Before cross-wave	After cross-wave	Diff (%)	Before cross-wave	After cross-wave	Diff (%)	
1992	17,658	16,476	-6.7	103,819	88,824	-14.4	7,702
1993	20,210	18,367	-9.1	129,371	110,797	-14.4	6,047
1994	23,391	25,809	10.3	109,160	159,140	45.8	7,051
1995	55,032	51,625	-6.2	656,659	642,642	-2.1	5,222
1996	33,520	33,679	0.5	180,386	192,846	6.9	6,811
1998	48,124	49,312	2.5	351,968	358,908	2.0	14,395
2000	59,931	58,028	-3.2	409,448	385,840	-5.8	13,214
2002	48,381	47,568	-1.7	221,133	218,062	-1.4	12,349
2004	62,595	62,592	-0.0	443,469	470,911	6.2	13,645
2006	64,749	67,957	5.0	439,230	492,687	12.2	12,605
2008	65,489	62,818	-4.1	348,328	312,843	-10.2	11,897
2010	49,826	50,548	1.4	316,332	360,342	13.9	15,276
2012	53,625	55,289	3.1	461,707	484,971	5.0	14,314

7.3 Effect of cross-wave imputations on the distributions of first differences

Table 7.5 shows the effect of cross-wave imputation on the first differences of total wealth. We would not expect an effect on the mean first difference, but we would expect a reduction in the standard deviation. By and large, the table confirms this: percentage-wise, the differences in the means can be large, but this is mainly because the means are much closer to zero than the levels, and sometimes very close to zero (small denominator). With only one exception, the standard deviations of the first differences are smaller with cross-wave imputation, but the magnitudes of the effect of cross-wave imputation compared to cross-sectional imputation are modest. The cross-wave imputations decrease the variability of the changes, but not by much.

Table 7.5: Effect of cross-wave imputations on first differences in total wealth

	Mean			s.d.			N
Wave years	Before cross-wave	After cross-wave	Diff (%)	Before cross-wave	After cross-wave	Diff (%)	
1992-1994	24,450	35,731	46.1	370,663	358,724	-3.2	7,029
1993-1995	67,703	67,221	-0.7	786,058	760,311	-3.3	5,209
1994-1996	25,363	22,414	-11.6	406,645	385,269	-5.3	6,534
1995-1998	-9,588	-7,680	-19.9	576,200	527,158	-8.5	4,419
1996-1998	45,779	51,687	12.9	1,004,359	997,685	-0.7	6,351
1998-2000	36,944	28,268	-23.5	1,011,108	978,886	-3.2	12,843
2000-2002	11,127	11,588	4.1	1,154,082	1,134,003	-1.7	11,815
2002-2004	36,514	44,041	20.6	1,233,760	1,228,908	-0.4	11,144
2004-2006	58,328	59,357	1.8	851,753	839,579	-1.4	12,263
2006-2008	1,331	-11,258	-946.1	957,765	968,656	1.1	11,442
2008-2010	-59,592	-56,556	-5.1	837,102	796,151	-4.9	10,550
2010-2012	-12,362	-3,802	-69.2	759,602	717,772	-5.5	13,941

Note: "Total wealth" is the sum of all assets (liabilities viewed as negative assets) except trusts and the net value of the second home.

Tables 7.6 and 7.7 present the comparisons for the first differences in the primary residence and stocks. Again, because the average changes are small, the percentage differences between them are sometimes large. The differences in the standard deviations are generally modest, and without clear pattern.

Table 7.6: Effect of cross-wave imputations on first differences in value of the primary residence

	Mean			s.d.			N
Wave years	Before cross-wave	After cross-wave	Diff (%)	Before cross-wave	After cross-wave	Diff (%)	
1992-1994	5,686	5,604	-1.4	88,698	85,627	-3.5	7,029
1993-1995	-2,381	-2,079	-12.7	64,207	69,870	8.8	5,209
1994-1996	3,402	3,499	2.8	94,765	95,445	0.7	6,534
1995-1998	846	370	-56.3	153,744	152,896	-0.6	4,419
1996-1998	5,691	5,115	-10.1	65,159	63,646	-2.3	6,351
1998-2000	8,226	8,498	3.3	160,971	158,572	-1.5	12,843
2000-2002	13,804	13,564	-1.7	182,081	175,957	-3.4	11,815
2002-2004	20,155	20,714	2.8	278,966	278,706	-0.1	11,144
2004-2006	27,919	28,393	1.7	270,570	267,308	-1.2	12,263
2006-2008	-1,279	-2,477	93.7	324,300	324,449	0.0	11,442
2008-2010	-30,493	-30,302	-0.6	375,171	375,553	0.1	10,550
2010-2012	-10,470	-10,625	1.5	156,531	160,430	2.5	13,941

Note: "Primary residence" as defined here does not include mobile homes, and is set to zero if the household does not own their primary residence.

Table 7.7: Effect of cross-wave imputations on first differences in stock wealth

	Mean			s.d.			N
Wave years	Before cross-wave	After cross-wave	Diff (%)	Before cross-wave	After cross-wave	Diff (%)	
1992-1994	5,374	8,897	65.6	124,637	132,837	6.6	7,029
1993-1995	32,147	31,989	-0.5	643,405	631,559	-1.8	5,209
1994-1996	9,453	8,008	-15.3	156,754	176,810	12.8	6,534
1995-1998	1,641	7,051	329.8	309,703	281,760	-9.0	4,419
1996-1998	11,670	13,566	16.2	252,424	257,805	2.1	6,351
1998-2000	13,571	10,089	-25.7	389,704	357,171	-8.3	12,843
2000-2002	-13,008	-12,962	-0.4	320,221	303,791	-5.1	11,815
2002-2004	17,276	18,893	9.4	392,999	418,434	6.5	11,144
2004-2006	426	3,277	668.7	443,162	400,535	-9.6	12,263
2006-2008	1,614	-5,402	-434.6	411,672	412,312	0.2	11,442
2008-2010	-5,087	-1,424	-72.0	294,017	302,403	2.9	10,550
2010-2012	2,452	3,734	52.3	472,261	452,145	-4.3	13,941

7.4 Ownership changes and reversals

Table 7.8 shows the number of ownership changes across all assets. As with the first differences, we would expect that imputations that take adjacent-wave information into account result in fewer changes than imputations that do not take this into account. The table confirms this: The cross-wave imputations result in fewer ownership changes across the board, but again, the effect is relatively small.

Table 7.8: Effect of cross-wave imputations on ownership changes (across all assets)

	own to not own			not own to own			N
Wave years	Before cross-wave	After cross-wave	Diff (%)	Before cross-wave	After cross-wave	Diff (%)	
1992-1994	7,672	7,621	-0.7	7,007	6,900	-1.5	126,522
1993-1995	3,761	3,638	-3.3	4,554	4,466	-1.9	83,344
1994-1996	6,629	6,508	-1.8	5,772	5,662	-1.9	104,544
1995-1998	4,085	4,052	-0.8	2,550	2,500	-2.0	83,961
1996-1998	7,029	6,890	-2.0	5,592	5,491	-1.8	120,669
1998-2000	12,465	12,181	-2.3	11,730	11,430	-2.6	269,703
2000-2002	12,602	12,320	-2.2	9,970	9,631	-3.4	248,115
2002-2004	10,615	10,361	-2.4	9,936	9,659	-2.8	234,024
2004-2006	12,229	11,956	-2.2	10,413	10,157	-2.5	257,523
2006-2008	11,174	10,942	-2.1	9,324	9,095	-2.5	240,282
2008-2010	11,141	10,866	-2.5	8,418	8,066	-4.2	221,550
2010-2012	14,114	13,591	-3.7	9,954	9,441	-5.2	292,761

Note. In 1995 and 1996, we do not have measures for the second home and any loans or mortgages on it. Therefore, the number of assets compared in rows spanning these years is smaller.

Table 7.9 presents the number of ownership reversals across three consecutive waves. This also shows across the board reductions. Comparing Table 7.9 with Table 7.8, we see that the percent effect of the cross-wave imputations on ownership reversals is larger than on wave-to-wave ownership changes. Thus, the ownership changes that are affected are more often ones that are reversed in the next wave, that is, the ones that are more likely to be incorrect.

Table 7.9: Effect of cross-wave imputations on ownership reversals (across all assets)

	not own to own to not own			own to not own to own			N
Wave years	Before cross-wave	After cross-wave	Diff (%)	Before cross-wave	After cross-wave	Diff (%)	
1992-1996	3,056	2,970	-2.8	2,335	2,258	-3.3	104,128
1993-1998	1,929	1,894	-1.8	851	821	-3.5	70,464
1994-1998	2,632	2,542	-3.4	1,860	1,828	-1.7	98,048
1995-2000	1,183	1,139	-3.7	960	925	-3.6	70,433
1996-2000	2,558	2,456	-4.0	2,270	2,188	-3.6	111,777
1998-2002	5,510	5,262	-4.5	3,680	3,498	-4.9	242,004
2000-2004	4,401	4,177	-5.1	3,911	3,727	-4.7	225,099
2002-2006	4,422	4,211	-4.8	3,060	2,943	-3.8	211,344
2004-2008	4,699	4,469	-4.9	3,583	3,428	-4.3	234,297
2006-2010	4,351	4,160	-4.4	3,280	3,133	-4.5	214,074
2008-2012	4,037	3,748	-7.2	2,970	2,799	-5.8	201,453

Note. In 1995 and 1996, we do not have measures for the second home and any loans or mortgages on it. Therefore, the number of assets compared in rows spanning these years is smaller.

7.5 Spikes and trenches in values

Tables 7.10 and 7.11 show the effects of cross-wave imputation on the number of spikes and trenches as defined in chapter 5, that is, instances in which the amount is positive in three consecutive waves, but the middle one differs by more than a factor of 10 from the average of the two outer waves. Spikes and trenches are relatively rare, but they may have a large impact on statistics such as means and standard deviations, especially of the distribution of change scores. Furthermore, if they are due to imputation that takes only cross-sectional information into account, it is likely that these misrepresent the distribution of interest. Tables 7.10 and 7.11 show that the cross-wave imputations lead to a systematic and substantial reduction in the number of spikes and trenches.

Table 7.10: Effect of cross-wave imputations on spikes and trenches in values of individual assets (across all assets)

	spikes			trenches			N
Wave years	Before cross-wave	After cross-wave	Diff (%)	Before cross-wave	After cross-wave	Diff (%)	
1992-1996	271	220	-18.8	575	477	-17.0	97,620
1993-1998	187	186	-0.5	404	348	-13.9	66,060
1994-1998	209	164	-21.5	717	610	-14.9	91,920
1995-2000	161	141	-12.4	461	390	-15.4	70,433
1996-2000	253	231	-8.7	766	653	-14.8	111,777
1998-2002	511	402	-21.3	1,493	1,220	-18.3	242,004
2000-2004	508	398	-21.7	1,430	1,222	-14.5	225,099
2002-2006	477	335	-29.8	1,362	1,092	-19.8	211,344
2004-2008	538	417	-22.5	1,448	1,238	-14.5	234,297
2006-2010	461	354	-23.2	1,315	1,034	-21.4	214,074
2008-2012	361	254	-29.6	1,063	928	-12.7	201,453

Note. In 1995 and 1996, we do not have measures for the second home and any loans or mortgages on it. Therefore, the number of assets compared in rows spanning these years is smaller.

Table 7.11: Effect of cross-wave imputations on spikes and trenches in total wealth

	spikes			trenches			N
Wave years	Before cross-wave	After cross-wave	Diff (%)	Before cross-wave	After cross-wave	Diff (%)	
1992-1996	18	17	-5.6	88	81	-8.0	6,508
1993-1998	44	43	-2.3	86	71	-17.4	4,404
1994-1998	15	12	-20.0	77	64	-16.9	6,128
1995-2000	23	20	-13.0	100	93	-7.0	3,707
1996-2000	17	13	-23.5	93	79	-15.1	5,883
1998-2002	63	45	-28.6	240	198	-17.5	11,524
2000-2004	69	43	-37.7	192	168	-12.5	10,719
2002-2006	42	32	-23.8	191	158	-17.3	10,064
2004-2008	61	40	-34.4	206	192	-6.8	11,157
2006-2010	51	45	-11.8	196	172	-12.2	10,194
2008-2012	65	40	-38.5	186	160	-14.0	9,593

Note: "Total wealth" is the sum of all assets (liabilities viewed as negative assets) except trusts and the net value of the second home.

8. The experiment with confirmation checks in sections H and Q in 2012

As we have seen in chapter 5, section U has been successful in catching and correcting incorrect amounts. However, not all households are eligible for being questioned in section U. Specifically, because of privacy issues, HRS does not administer section U if the financial respondent in a wave differs from the financial respondent in the previous wave. Therefore, in 2012, HRS conducted an experiment to investigate whether some of the section U ideas can still be implemented in a limited way for these households. In these households, if the respondent reported a continuous amount that differed more than $50,000 from the amount (or nearest bracket threshold, in the case of an unfolding bracket response in the previous wave) given in 2010, the respondent was immediately asked to confirm or correct the *current* amount, that is, the one just reported, without mentioning the previous wave's report. This was done for three types of assets: the value of the primary residence, the balance in checking and savings accounts, and the net value of stocks and mutual funds (directly held, i.e., outside retirement accounts).

Unfortunately, the results of this have been disappointing. Table 8.1 shows that, out of 744 cases, only six were not confirmed. Table 8.2 shows these six cases. In only two of these cases is the correction of an order of magnitude for which this question was designed. The benefits of these two corrections do not seem to outweigh the additional burden on the much larger set of respondents who are asked this question.

Table 8.1: Households who were asked the experimental questions and their responses

Asset	Asked question	Confirmed	Corrected
Stocks	108	107	1
Checking/savings	148	147	1
Primary residence	488	484	4
Total	744	738	6

Table 8.2: Households that did not confirm

Asset	Preload value (from 2010)	Initial report	Corrected value
Stocks	215,000	160,000	233,000
Checking/savings	2,500	600,000	2,000
Primary residence	600,000	1,000,000	(refused)
Primary residence	1,000,000	800,000	800,000
Primary residence	(missing)	82,000	110,000
Primary residence	(missing)	285,000	300,000

9. Conclusions and discussion

In this report, we present revised wealth measures for the HRS that have been made available to the research community as part of the RAND HRS data (version M forward), which aim to reduce the effect of observation error on wealth and changes in wealth. The improvements include incorporating corrections from the asset verification section and using cross-wave information in the imputation procedures.

Since 2002, the HRS has included the asset verification section (section U), which questions respondents about large changes in asset values and allows them to correct the current and previous waves' values. We found that both corrections to the current wave's values and corrections to the previous wave's values usually seem plausible, and thus we incorporated all of these. However, when a respondent was questioned about the same asset in two consecutive waves, and thus we have two potentially corrected reports plus the original for the same asset in the same wave, we generally give precedence to the contemporaneous section U result. The exception is when there appears to be a decimal place error. Although most of the corrections from section U have been implemented in computer code, there are several situations in which we review the cases manually and hard code the decisions we made on the basis of this. This happens with the double corrections mentioned above, and with corrections that lead to data inconsistencies (a household owning both a mobile home and a house as their primary residence).

The cross-wave imputations, which include the asset's values in the previous and next wave among the covariates, are based on the premise that an asset's value in one wave is highly informative about the value in an adjacent wave, and that this strong serial correlation would be represented better by cross-wave imputations than by cross-sectional imputations, which only use information from the current wave, and which thus rely on serial correlation in the explanatory variables to capture the serial correlation in the wealth measures. Challenges with the cross-wave imputations arise when households only have partial information in the adjacent waves or when marital status changes, after which we do expect large changes in wealth. For the former, we use the adjacent-wave cross-sectional imputations when necessary and we have separate imputation procedures for households that were completely absent in an adjacent wave. Marital status changes are included as additional covariates.

Both the section U corrections and the cross-wave imputations have the expected (and intended) effect: with a few exceptions, they reduce the variability of wealth changes. The effect on the variability of first differences is stronger for the section U corrections than for the cross-wave imputations, but they complement each other. Conversely, cross-wave imputations reduce the number of "spikes" and "trenches" (in which a wave's asset value differs more than a factor of 10 from the average of the two adjacent waves) more than section U does.

The improved measures have been available since version M (2014) of the RAND HRS data. The main RAND HRS data only include the imputations that include both

improvements, but the RAND income and wealth imputations file also includes the cross-sectional imputations (still incorporating the section U corrections).

References

Burgette, L. F., Hurd, M. D., & Meijer, E. (2016). Asset smoothing in the Health and Retirement Study: A Bayesian nonparametric approach. (in preparation)

Chien, S., et al. (2015) *RAND HRS data documentation, Version O*. Santa Monica, CA: RAND Corporation.

Groves, R. M., Fowler, F. J., Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). New York, NY: Wiley.

Hill, D. H. (2006). Wealth dynamics: Reducing noise in panel data. *Journal of Applied Econometrics*, *21*, 845-860.

Hurd, M. D. (1999). Anchoring and acquiescence bias in measuring assets in household surveys. *Journal of Risk and Uncertainty*, *19*, 111-136.

Hurd, M., Juster, F. T., & Smith, J. P. (2003). Enhancing the quality of data on income: Recent innovations from the HRS. *Journal of Human Resources*, *38*, 758-772.

Juster, F. T., & Smith, J. P. (1997). Improving the quality of economic data: Lessons from the HRS and AHEAD. *Journal of the American Statistical Association*, *92*, 1268-1278.

Juster, F. T., Smith, J. P., & Stafford, F. (1999). The measurement and structure of household wealth. *Labour Economics*, *6*, 253-275.

Juster, F. T., & Suzman, R. (1995). An overview of the Health and Retirement Study. *Journal of Human Resources*, *30*, S7-S56.

Lee, J., Meijer, E., & Phillips, D. (2015). *The effect of using different imputation methods for economic variables in aging surveys*. Working Paper No: 2015-019. Los Angeles, CA: University of Southern California, Center for Economic and Social Research.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York, NY: Wiley.

Marsden, P. V., & Wright, J. D. (Eds.). (2010). *Handbook of survey research* (2nd ed.). Bingley, UK: Emerald.

Meijer, E., Spierdijk, L., & Wansbeek, T. (2013). Measurement error in the linear dynamic panel data model. In B. C. Sutradhar (Ed.), *Longitudinal data analysis subject to measurement error, missing values and/or outliers*. New York, NY: Springer.

National Institute on Aging (2007). *Growing older in America: The Health and Retirement Study* (NIH Publication No. 07-5757). Bethesda, MD: Author.

Pantoja, P., et al. (2015). *RAND HRS income and wealth imputations, Version O*. Santa Monica, CA: RAND Corporation.

Venti, S. (2011). Economic measurement in the Health and Retirement Study. *Forum for Health Economics & Policy*, 14(3), Article 2.

Wansbeek, T., & Meijer, E. (2000). *Measurement error and latent variables in econometrics*. Amsterdam, Netherlands: North-Holland.

Appendix

Appendix A: Trusts

A trust is not an asset in itself, but a legal vehicle that its creator puts assets into for the benefit of others, typically the children of the creator after his or her death. Often, the trust can still be managed and changed by the creator and thus the assets in it would still be owned by the creator until his or her death. In all waves except wave 1 (1992), the HRS (and AHEAD) asks whether the household has put any assets into a trust. If so, in most waves, the HRS then asks about the value of the assets in the trust, and whether all of the assets in the trust were already reported earlier in the interview. If the latter is not the case, the HRS then asks about the value of the assets that had not been reported earlier. The exceptions are the 1994-1996 waves (HRS waves 2 and 3, AHEAD wave 2), in which the total value is not asked, but only the value of assets not reported earlier.

In addition to trusts created by the respondent and his or her spouse, the HRS in most waves includes "rights in a trust [...] that you haven't already told me about" among the examples of "other assets" in the question that asks about those. Thus, this refers to trusts created by others in which the respondent (and/or the spouse) is a beneficiary. The exceptions here are AHEAD wave 1 (1993) and HRS wave 2 (1994), in which trusts were not mentioned in the question text.

In this report, as in the HRS variable labels and the RAND HRS categorization, "trusts" refers to the trusts that were created by the respondent and spouse, and "other assets" includes the trusts in which the respondent or spouse are beneficiaries.

Section U uses the same categorization, but never includes trusts in the list of examples for "other assets" (or "personal items of value" as it is called since 2010). The preloads for "trusts" were based on the total trust value in 2002-2004 and 2008 data models 1-2; in 2006, trusts were not asked in section U. Starting with 2008 data model 3, both previous and current wave values reference the value of trusts not reported earlier in the interview. In all cases, however, the wording of the question does not make this distinction and just mentions "trusts worth ...".

Trusts are not included in the main RAND HRS file; neither as a separate asset category, nor as part of one of the wealth aggregates. However, trusts are included in the RAND Income and Wealth Imputations file. Specifically, this includes the "trusts that have not been reported elsewhere".

Because of the ambiguity in the meaning of the trusts variable, we do not use the section U corrections for trusts.

Appendix B: Correction of the previous wave's values when in the previous wave the respondent also entered section U

Suppose in wave *t* a respondent enters section U for an asset *a*. Thus, there is a large difference between the value of *a* in wave *t* and in wave *t* - 1. The respondent then confirms or corrects the values for wave *t* - 1 and wave *t* in the wave *t* section U. Then, in wave *t* + 1, if the respondent enters section U again, the values for the asset for waves *t* and *t* + 1 are confirmed or corrected in wave *t* + 1. The question we now address is what to do when in this situation, the respondent corrects the wave *t* value in wave *t* + 1's section U. Thus, we have an original wave *t* answer (section H or Q), a confirmation or correction of this value in wave *t*'s section U, and then a correction of this same value in wave *t* + 1's section U.

Given that the respondent explicitly confirmed or corrected the original value in wave *t*'s (contemporaneous) section U, we generally give precedence to the outcome resulting from wave *t*'s section U over the section U correction two years later. Thus, we do not use the wave *t* + 1 correction for these cases unless there is strong evidence that the second correction (i.e., wave *t* + 1) is "better" than the first correction or confirmation (in wave *t*). In particular, we only look at order of magnitude differences and inspect these individually. The idea is that there may be typos that lead to a decimal place error. To allow for "almost exact" decimal place error corrections, we inspected all corrections in which the correction in wave *t* + 1's section U differs from the result after wave *t*'s section U by more than a factor 9. We looked at the values in waves *t* - 1, *t*, and *t* + 1 after incorporating contemporaneous section U corrections and compared the wave *t* + 1 section U corrections to wave *t*'s values against these. In the case of a correction in wave *t*'s section U for wave *t*'s value, we also looked at the original uncorrected values, but found that these did not affect our assessments, so the discussion here will ignore them. Thus we look at a sequence of three values and a correction to the middle one, proposed at the same time as reporting the most recent one. Two of us independently scored all the cases for the 2008-2010 waves (*t* = 2008, *t* + 1 = 2010) with an agree/disagree to correct. Then we discussed the findings and developed computer-implementable rules that we hoped would mimic our assessments in nearly all cases. We subsequently ran the rules on all the relevant waves (2000-2012) and checked to what extent we agreed with the computer-generated decisions, and whether the cases of disagreement showed any discernable pattern that we could implement in an additional rule. After two iterations of this process, there were only a handful of cases of disagreement, and no obvious generally applicable rule that would cover them. Thus, these cases have been hardcoded. Below, we describe both the rules and the disagreements in detail.

While we believe the rules and the few exceptions give reasonable results, we should note that there are a considerable number of cases where we were very unsure of what the "right" answer was, even though we were only looking at order of magnitude differences. For example, there is an observation (hhidpn = 210526010) with reported "other debt" in 2006 of $100,000, in 2008 of $500,000, and in 2010 of $30,000. In the 2010 section U,

the respondent corrected the 2008 number to $28,000. The $28,000 is close to the $30,000, but it was reported in the same interview, so the respondent may have been biased by the current value. The $28,000 is not close to the $100,000 of 2006 and there does not seem to be a relationship with the $500,000 reported in 2008 itself, and confirmed in the 2008 section U.

Rule 1: ownership changes
If either the wave *t* value or its proposed correction is 0/no ownership, then do not correct.

Ownership changes are discrete events that should be less susceptible to typos than continuous values. This is especially the case when the respondent confirmed or explicitly corrected the value in wave *t*'s section U. Inspection of these cases showed that there rarely is an indication that the proposed change in wave $t + 1$ is better than the result after wave *t*'s section U. However, the converse is also often not the case. We do not change this because we give precedence to the contemporaneous section U results over the ones reported two years later.

Rule 2: no useful information from adjacent waves
When both the wave $t - 1$ result and the wave $t + 1$ result are 0/no ownership or "no value/bracket" or otherwise completely missing, and both the wave *t* result and its proposed correction are strictly positive continuous values, accept the correction if and only if it differs from the wave *t* result by a factor of exactly 10, 100, 1/10, or 1/100.

In this case we have no useful information whatsoever from the adjacent waves about what would be a reasonable order of magnitude, and therefore we put the weight on the wave *t* (contemporaneous) section U result. But if in wave $t + 1$, this result is changed by a factor of exactly 10 or 100, this is an indication of a decimal place error in wave *t*: If it was a decimal place error due to a typo in wave $t + 1$, the respondent would not have needed to correct the value, so the vast majority of these corrections must be intended corrections. Therefore, we do correct these.

Rule 3: useful information from one adjacent wave and an order of magnitude difference
(3a) When the wave $t - 1$ result is 0/no ownership or "no value/bracket" or otherwise completely missing and the wave $t + 1$ value is positive, and the wave *t* value is more than a factor 10 larger or smaller than the wave $t + 1$ value, and the proposed correction is closer to the wave $t + 1$ value than the wave *t* value is to the wave $t + 1$ value, then accept the correction.
(3b) When the wave $t + 1$ result is 0/no ownership or "no value/bracket" or otherwise completely missing and the wave $t - 1$ value is positive, and the wave *t* value is more than a factor 10 larger or smaller than the wave $t - 1$ value, and the proposed correction is closer to the wave $t - 1$ value than the wave *t* value is to the wave $t - 1$ value, then accept the correction.

In this case, we have information about the asset value from only one adjacent wave, so the only criterion we have is whether the correction is somehow "closer" to the adjacent wave value than the uncorrected value is. However, changes in asset values do occur, so a blanket rule that always selects the value that's closest to the adjacent wave's value is likely to dampen changes in asset values too much. Thus, we only implement the correction if there is a large discrepancy, which makes it more reasonable to assume it's a typo.

Rule 4: following the trend
If the wave t - 1, t, and t + 1 values are all positive, and the proposed corrected value is positive as well, then apply the correction if the wave t value is more than 2 times the largest of the wave t - 1 and wave t + 1 values or less than 50% of the smallest of these two, and the corrected value is between 80% of the wave t - 1 value and 120% of the wave t + 1 value, or vice versa. In formula: let a_{t-1}, a_t, and a_{t+1} be the wave t - 1, t, and t + 1 values, respectively. Then the correction c is accepted if (i) $a_{t-1} > 0$, $a_t > 0$, $a_{t+1} > 0$, and $c > 0$; (ii) $a_t > 2 * \max(a_{t-1}, a_{t+1})$ or $a_t < \min(a_{t-1}, a_{t+1})/2$; and (iii) $0.8\ a_{t-1} \leq c \leq 1.2\ a_{t+1}$ or $0.8\ a_{t+1} \leq c \leq 1.2\ a_{t-1}$.

Thus, if after the correction, the three values follow a monotonic trend (with a tolerance of 20%) whereas without the correction, the middle value is off by a factor 2 or more from the closest adjacent value, the correction seems more reasonable than the wave t value.

For rules 3 and 4, if a bracket answer was given, we take the average of the lower and upper bracket thresholds as the value. However, for the cases studied here, very few bracket answers were given.

After implementation of the corrections for the previous wave, we again check whether there were any cases that now had both a house and a mobile home. There was 1 such case for the 2002 base wave, 6 in 2004, 4 in 2006, 5 in 2008, and 9 in 2010. In all of the cases up to 2008, we decided to stick with the result after the contemporaneous section U corrections and not accept the further corrections, but for the 2010 wave, we accepted two corrections of residence type (from mobile home to house) and a few value corrections as well, although the latter tended to be small.

B.1 Disagreements

There is one observation (hhidpn = 173152020, which carries over to 173152010) in which there were several consecutive decimal place corrections of the value of the house. Table B.1 shows the data. We conclude that the correct value of the house is around $1 million, but the confirmation of the $11 million in 2006 section U and the subsequent correction to $1 million in 2008 section U, and conversely the correction of the $12 million in 2008 section U to $1.2 million and re-correction to $12 million in 2010

section U make this a difficult case to catch in general rules. The second line from the bottom indicates the result after all the general rules have been applied in the order (1) implementation of contemporaneous section U corrections; (2a) implementation of previous wave section U corrections if the respondent did not enter section U for the asset in the previous wave itself; (2b) implementation of previous wave section U corrections if the respondent entered section U for the asset in the previous wave, using the rules 1-4 above. The bottom line indicates the result after manual intervention.

Table B.1: Multiple decimal place corrections

		Year about which reported					
Wave year	Section	2000	2002	2004	2006	2008	2010
2000	H	750,000					
	(no U)						
2002	H		850,000				
	U	did not enter section U					
2004	H			1,200,000			
	U		did not enter section U				
2006	H				11,000,000		
	U			900,000	confirmed		
2008	H					12,000,000	
	U				1,000,000	1,200,000	
2010	H						1,200,000
	U					12,000,000	confirmed
Rules		750,000	850,000	900,000	1,000,000	12,000,000	1,200,000
Manual		750,000	850,000	900,000	1,000,000	1,200,000	1,200,000

The cases in the Table B.2 are cases where the rules do not accept the corrections, but we decided to accept the corrections anyway. Here and in the following, the "base wave" is wave t, for which the corrections are intended, the "following wave" is wave $t + 1$, the wave in which the corrections were reported by the respondent, and the "preceding wave" is wave $t - 1$.

Table B.2: Cases in which the rules do not accept the correction, but manual intervention did

hhidpn	base wave year	asset	preceding wave[a]	base wave[a]	following wave[a]	following wave sec U correction of base wave
Confirmed in base wave section U						
38535010	2004	stocks	90,000	400,000	35,000	25,000
147799020		real estate	470,000	0	800,000	500,000
183406010		primary residence	0	360,000	37,000	33,000
20978010	2006	real estate	.	400,000	2,501 - 124,999	40,000
90685030		real estate	300,000	600	125,001-499,999	120,000
501845010	2008	other debt	5,000	400,000	4,000	10,000
501463010		other assets	0	340,000	35,000	34,000
212585010		second home	200,000	1,600,000	0	25,000
Corrected in base wave section U						
74715010	2004	business/farm	650,000	32,500	650,000	410,000

[a]After contemporaneous corrections from section U, if any.

B.2 Descriptive statistics

Table B.3 shows in how many cases each of the four rules applies, broken down by whether the case was confirmed or corrected in the base wave section U and whether the rule prescribes accepting the correction or not (so the manual interventions from section B.1 are not taken into account).

Table B.3: Number of cases to which each rule applies (base wave = 2002-2010)

	Confirmed in base wave U		Corrected in base wave U	
Rule	No correction	Correction	No correction	Correction
1	98	0	29	0
2	11	6	0	0
3	7	20	2	3
4	25	41	9	7
Total	141	67	40	10

Thus, most cases are Rule 1, that is, ownership corrections, which are not implemented. This is followed by Rule 4, where all four values are positive. Most of these corrections are accepted. There are many more cases that were confirmed in the base wave section U than cases that were corrected, which suggests that confirmations are less careful than corrections.

Table B.4 breaks the cases down by base wave year. This shows an increasing pattern, which is probably largely due to only three assets per respondent being eligible for section U in the earlier years, but a drop with 2010 as base wave year.

Table B.4: Number of cases with section U entry in the base wave and section U correction of the base wave's value in the next wave, by base wave year

Base wave year	Confirmed in base wave U	Corrected in base wave U
2002	3	1
2004	35	6
2006	62	14
2008	64	18
2010	44	11
Total	208	50

In Table B.5, the cases are broken down by asset type. A noteworthy characteristic of this table is the large number of cases for "other assets" (called "personal items of value" in later waves). We suspect that this category is less well defined, so it's less clear which items belong in this category and which do not. Also, it would be easier to forget certain items in this category.

Table B.5: Number of cases with section U entry in the base wave and section U correction of the base wave's value in the next wave, by asset type

Asset	Confirmed in base wave U	Corrected in base wave U
1.Other debt	3	0
2.Trusts	0	0
3.Other assets	49	8
4.Vehicles	1	0
5.CDs	7	1
6.Checking/savings	15	2
7.Bonds	6	2
8.Stocks	26	6
9.IRAs	20	6
10.Business/farm	9	2
11.Real estate	24	9
12.Primary residence	19	12
13.Mortg 1 prim. res.	4	0
14.Mortg 2 prim. res.	0	0
15.Equity line of credit	0	0
16.Mobile home	2	0
17.Second home	23	2
18.Loans 2nd home	0	0
Total	208	50

Appendix C: Algorithm used for classifying marital status changes

As discussed in chapter 6, for the cross-wave imputations we use information about marital status changes. In most cases, it is easy to ascertain whether there was a marital status change and, if so, what kind, but some situations are more ambiguous, for example, when a respondent's marital status changes from "partnered" to "never married". We have developed an algorithm to assign unique codes for marital status changes in all circumstances. In this appendix, we describe this algorithm in detail. In the algorithm, we use the following variables from the RAND HRS (w = 1, ..., 11 is the wave number):

HHIDPN	Respondent identifier
RwIWSTAT	Interview status
HwCPL	Whether a couple
RwMSTAT	Marital status in wave w
SwHHIDPN	Identifier of the respondent's spouse

Throughout, we assume we have identified a respondent in wave w for whom we need to impute an asset value, or who may be a potential donor for imputation, or who is used in estimating the imputation model. Let $p = w - 1$ be the previous wave and $n = w + 1$ be the next wave. Let r denote the target respondent. We use brackets to indicate the respondent whose record contains the value of a variable, so HHIDPN[r] = r and RwMSTAT[r] is the value of RwMSTAT as reported by respondent r, i.e., respondent r's reported marital status in wave w.

In addition to the respondent's record, information about the previous wave's marital status and asset value may be found in the record of the wave w spouse and wave w - 1 spouse. Our target respondent will be someone who was a respondent in wave w, and if this respondent had an identified spouse in wave w, this person's identifier is stored in the variable SwHHIDPN[r]. Let s denote this person. Information about r's marital status in wave p may be contained in variables like HpCPL[s] and RpMSTAT[s].

It is also possible that there is a wave p respondent who is the wave p spouse of r but not equal to s (e.g., in case of a divorce or widowing) and the record of this person may also contain valuable additional information about r's wave p marital status and asset values, or the marital status change. Let u = SpHHIDPN[r] be the wave p spouse of the target respondent. We will also need information from the spouse if r is not in wave p. In this case, SpHHIDPN[r] does not contain the identifier of this spouse. Instead, we define the wave p spouse for these purposes as the respondent q for whom SpHHIDPN[q] = r, i.e., the respondent who reports that our target respondent r is his or her spouse in wave p. If there is no such respondent, q is missing

Note that SwHHIDPN is not symmetric: it is possible (and occasionally happens in the data, although it is rare) that SwHHIDPN[r] = s but SwHHIDPN[s] is zero or missing. This happens in the case of marital status changes (divorce, widowing) in between the

interviews of *r* and *s*. In principle, it would even be possible that S*w*HHIDPN[*s*] is the respondent identifier of another respondent who is not *r*, but this does not happen in the data. However, in waves 2 and 3 it does happen that S*w*HHIDPN[*s*] is a seemingly valid number different from HHIDPN[*r*], but this occurs because this *r*'s identifier was changed later on due to remarriage to another respondent. (See Chien et al., 2015, appendix A.) Thus, S*w*HHIDPN[*s*] is what HHIDPN[*r*] used to be at the time but is not anymore. This is best solved by changing S*w*HHIDPN[*s*] to the current value of HHIDPN[*r*], so the data are consistent again.

Also, note that because of the asymmetry, the "S" variables in the RAND HRS are not necessarily equal to the "R" variables of the spouse, e.g., S*w*MSTAT[*r*] ≠ R*w*MSTAT[*s*]. Moreover, this carries over to the "H" variables: in cases where S*w*HHIDPN is not symmetric, one of the spouse's variables reflects the situation of the couple (jointly), whereas the other spouse's variable reflects the situation after breakup/widowing. Hence, we would have H*w*CPL[*r*] = 1, H*w*CPL[*s*] = 0 (or missing), H*w*HHID[*r*] ≠ H*w*HHID[*s*], and asset values will be different as well. Here and below, "≠" is used as Stata interprets it, i.e., if *X* is missing and *Y* not, then *X* ≠ *Y*. If both have the same missing value, *X* = *Y*, so . = ., .m = .m, but . ≠ .m, etc.

After these preliminaries, here is the algorithm for determining marital status changes and sources of previous wave's asset values:

I. The target respondent is present in wave *p* (R*p*IWSTAT[*r*] = 1). Previous wave's asset values are taken from the respondent's record after cross-sectional imputations. (These asset values may have been extracted from *r*'s spouse's record if the spouse was the financial respondent in the previous wave.)
    a. H*p*CPL[*r*] = 0, H*w*CPL[*r*] = 0. The respondent is single in both waves: no marital status change.
    b. H*p*CPL[*r*] = 1, H*w*CPL[*r*] = 1. The respondent is married/partnered in both waves: no marital status change.
    c. H*p*CPL[*r*] = 0, H*w*CPL[*r*] = 1. "(Re)marriage".
    d. H*p*CPL[*r*] = 1, H*w*CPL[*r*] = 0.
        i. The wave *p* spouse is a respondent (and thus alive) in wave *w* (R*w*IWSTAT[*u*] = 1; note that we use *u* here and not *q*, although in practice these will often be the same, but *q* may be missing when *u* is not). In this case, the change must have been a breakup/divorce/separation.
        ii. The wave *p* spouse is dead in wave *w* (R*w*IWSTAT[*u*] = 5 or 6). In this case, the change must have been a "widowing".
        iii. The wave *p* spouse is not a respondent in wave *w* and not known to be dead (R*w*IWSTAT[*u*] = 4, 7, or missing, or *u* is not a valid identifier, i.e., 0 or missing).
            1. The respondent reports being divorced or separated in wave *w* (R*w*MSTAT[*r*] = 4, 5, 6). We assume the change was a divorce/separation.

2. The respondent reports being widowed in wave w (RwMSTAT[r] = 7). We assume the change was a widowing.
3. Most of the remaining cases have RwIWSTAT[u] = 4 (NR, alive). There are some indications that this classification may not be accurate and these respondents may actually be dead. This follows from the date of death in the variables RADYEAR and RANYEAR, but it is impractical to use these or look more than one wave ahead, because this would imply that imputations may have to be changed after every new release. Therefore, we classify these respondents as alive for our current purposes, and assume the change was a divorce/separation.

II. The target respondent is not present in wave p, but the wave w spouse is (RpIWSTAT[r] ≠ 1, RpIWSTAT[s] = 1). Previous wave's asset values are taken from the spouse's (s) record after cross-sectional imputations. Note that the wave w spouse does not have to be a respondent in wave w.
   a. HpCPL[s] = 0, HwCPL[r] = 0. This is not possible because s would not be defined with HwCPL[r] = 0.
   b. HpCPL[s] = 1, HwCPL[r] = 1. Married/partnered in both waves: no marital status change. Note that it is possible that the spouse was married to a different person than r in wave p, and r may or may not have been married in wave p, and s may be treated as single in wave w. But because we use the spouse's previous assets, those assets reflect the "married" status, and because we do not distinguish between "married to the same spouse" and "married to a different spouse", "no change" is the appropriate category.
   c. HpCPL[s] = 0, HwCPL[r] = 1. "(Re)marriage". Note that it is possible that the target respondent was married (to a different spouse) in wave p, but because we use the (wave w) spouse's previous (wave p) assets, those assets reflect the "single" status, and remarriage is the appropriate category.
   d. HpCPL[s] = 1, HwCPL[r] = 0. This is not possible because s would not be defined with HwCPL[r] = 0.

III. The target respondent is not present in wave p, and neither is the wave w spouse, but r is the spouse of a wave p respondent (RpIWSTAT[r] ≠ 1; RpIWSTAT[s] ≠ 1 or s is not a valid identifier; SpHHIDPN[q] = r; note that we need to use q and not u here, because u is missing). Previous wave's asset values are taken from the wave p's spouse's (i.e., q's) record after cross-sectional imputations.
   a. HpCPL[q] = 0, HwCPL[r] = 0. This is not possible because r would not be defined as q's wave p spouse with HpCPL[q] = 0.
   b. HpCPL[q] = 1, HwCPL[r] = 1. The respondent is married/partnered in both waves: no marital status change. Note that it is possible that the respondent was married to a different person than q in wave w, and q may or may not have been married in wave w. But because we use the respondent's wave w assets and we do not distinguish between "married to

the same spouse" and "married to a different spouse", "no change" is the appropriate category.

c. H*p*CPL[*q*] = 0, H*w*CPL[*r*] = 1. This is not possible because *r* would not be defined as *q*'s wave *p* spouse with H*p*CPL[*q*] = 0.

d. H*p*CPL[*q*] = 1, H*w*CPL[*r*] = 0.

i. The wave *p* spouse is a respondent (and thus alive) in wave *w* (R*w*IWSTAT[*q*] = 1). In this case, the change must have been a breakup/divorce/separation.

ii. The wave *p* spouse is dead in wave *w* (R*w*IWSTAT[*q*] = 5 or 6). In this case, the change must have been a "widowing".

iii. The wave *p* spouse is not a respondent in wave *w* and not known to be dead (R*w*IWSTAT[*q*] = 4, 7, or missing).

1. The respondent reports being divorced or separated in wave *w* (R*w*MSTAT[*r*] = 4, 5, 6). We assume the change was a divorce/separation.
2. The respondent reports being widowed in wave *w* (R*w*MSTAT[*r*] = 7). We assume the change was a widowing.
3. For reasons explained under I.d(iii)3, we classify the remaining former spouses as alive for our current purposes, and assume the change was a divorce/separation.

IV. The target respondent is not present in wave *p*, and neither is the wave *w* spouse, and *r* is not the spouse of a wave *p* respondent (R*p*IWSTAT[*r*] ≠ 1, R*p*IWSTAT[*s*] ≠ 1, there does not exist a *q* with S*p*HHIDPN[*q*] = *r*). This includes as a major special case the situation in which *w* is the baseline wave of the cohort (e.g., wave 4 for the CODA/WB cohorts). In this case we have no information about marital status change and previous wave's assets, and the respondent is only part of the imputations and models that do not require this information.

For the next wave's assets and the marital status change between the current and next wave, a similar but slightly different algorithm is used. Define the two wave *n* spouse identifiers *v* = S*n*HHIDPN[*r*] and *x* is such that S*n*HHIDPN[*x*] = *r* if possible, and missing otherwise.

I. The target respondent is present in wave *n* (R*n*IWSTAT[*r*] = 1). Next wave's asset values are taken from the respondent's record after cross-sectional imputations. (These asset values may have been extracted from *r*'s spouse's record if the spouse was the financial respondent in the next wave.)

a. H*w*CPL[*r*] = 0, H*n*CPL[*r*] = 0. The respondent is single in both waves: no marital status change.

b. H*w*CPL[*r*] = 1, H*n*CPL[*r*] = 1. The respondent is married/partnered in both waves: no marital status change.

c. H*w*CPL[*r*] = 0, H*n*CPL[*r*] = 1. "(Re)marriage".

d. H*w*CPL[*r*] = 1, H*n*CPL[*r*] = 0.

i. The wave *w* spouse is a respondent (and thus alive) in wave *n* (R*n*IWSTAT[*s*] = 1). In this case, the change must have been a breakup/divorce/separation.
ii. The wave *w* spouse is dead in wave *n* (R*n*IWSTAT[*s*] = 5 or 6). In this case, the change must have been a "widowing".
iii. The wave *w* spouse is not a respondent in wave *n* and not known to be dead (R*n*IWSTAT[*s*] = 4, 7, or missing, or *s* is not a valid identifier, i.e., 0 or missing).
1. The respondent reports being divorced or separated in wave *n* (R*n*MSTAT[*r*] = 4, 5, 6). We assume the change was a divorce/separation.
2. The respondent reports being widowed in wave *n* (R*n*MSTAT[*r*] = 7). We assume the change was a widowing.
3. For reasons explained under I.d(iii)3 for the previous wave, we classify the remaining former spouses as alive for our current purposes, and assume the change was a divorce/separation.

II. The target respondent is not present in wave *n*, but the wave *w* spouse is (R*n*IWSTAT[*r*] ≠ 1, R*n*IWSTAT[*s*] = 1). Next wave's asset values are taken from the spouse's (*s*) record after cross-sectional imputations. Note that the wave *w* spouse does not actually have to be a respondent in wave *w*.
a. H*w*CPL[*r*] = 0, H*n*CPL[*s*] = 0. This is not possible because *s* would not be defined with H*w*CPL[*r*] = 0.
b. H*w*CPL[*r*] = 1, H*n*CPL[*s*] = 1. Married/partnered in both waves: no marital status change. Again, it is possible that the spouse was married to a different person than *r* in wave *n*, but because we do not distinguish between "married to the same spouse" and "married to a different spouse", "no change" is the appropriate category.
c. H*w*CPL[*r*] = 0, H*n*CPL[*s*] = 1. This is not possible because *s* would not be defined with H*w*CPL[*r*] = 0.
d. H*w*CPL[*r*] = 1, H*n*CPL[*s*] = 0.
i. The respondent is present (and thus alive) in wave *n* (R*n*IWSTAT[*r*] = 1). This is not the case and thus not possible.
ii. The respondent is dead in wave *n* (R*n*IWSTAT[*r*] = 5 or 6). In this case, the change must have been a "widowing". Note that this is the change for the spouse and not for the respondent, but because we are looking at the respondent's and spouse's joint assets, this is immaterial and looking at the spouse's next wave assets uses the most information.
iii. The target respondent is not a respondent in wave *n* and not known to be dead (R*n*IWSTAT[*r*] = 4, 7, or missing.
1. The spouse reports being divorced or separated in wave *n* (R*n*MSTAT[*s*] = 4, 5, 6). We assume the change was a divorce/separation.

2. The spouse reports being widowed in wave *n* (R*n*MSTAT[*s*] = 7). We assume the change was a widowing.
3. For reasons explained under I.d(iii)3 for the previous wave, we classify the remaining respondents as alive for our current purposes, and assume the change was a divorce/separation.

III. The target respondent is not present in wave *n*, and neither is the wave *w* spouse, but *r* is the spouse of a wave *n* respondent (R*n*IWSTAT[*r*] ≠ 1; R*n*IWSTAT[*s*] ≠ 1 or *s* is not a valid identifier; S*n*HHIDPN[*x*] = *r*; note that we need to use *x* and not *v* here, because *v* is missing). Next wave's asset values are taken from the wave *n*'s spouse's (i.e., *x*'s) record after cross-sectional imputations.

a. H*w*CPL[*r*] = 0, H*n*CPL[*x*] = 0. This is not possible because *r* would not be defined as *x*'s wave *n* spouse with H*n*CPL[*x*] = 0.
b. H*w*CPL[*r*] = 1, H*n*CPL[*x*] = 1. Married/partnered in both waves: no marital status change. Again, see the notes on potentially changed spouses in earlier similar cases.
c. H*w*CPL[*r*] = 0, H*n*CPL[*x*] = 1. "(Re)marriage".
d. H*w*CPL[*r*] = 1, H*n*CPL[*x*] = 0. This is not possible because *r* would not be defined as *x*'s wave *n* spouse with H*n*CPL[*x*] = 0.

IV. The target respondent is not present in wave *n*, and neither is the wave *w* spouse, and *r* is not the spouse of a wave *n* respondent (R*n*IWSTAT[*r*] ≠ 1, R*n*IWSTAT[*s*] ≠ 1 or *s* is not a valid identifier, and there does not exist a *x* with S*n*HHIDPN[*x*] = *r*). This includes as a major special case the situation in which *w* is the latest wave (wave 11 for versions N and O of the RAND income and wealth imputations data). In this case we have no information about marital status change and next wave's assets, and the respondent is only part of the imputations and models that do not require this information.

Appendix D: Missing values and imputation flags

Missing values are a pervasive problem in survey research, and books on survey research (e.g., Groves et al., 2009; Marsden & Wright, 2010) tend to devote one or more chapters to it. They lead to loss of information, and therefore to less precise estimates. Furthermore, if missingness is related to the value of interest (e.g., if it is more likely to give a "don't know" or "refuse" answer if the true value is larger), missing data may lead to biased estimates.

For the asset data in the HRS, the problem is particularly salient. The HRS asks wealth in a moderately large number of components. The advantage of asking about many detailed components is that respondents are less likely to omit important components of wealth, and thus total wealth will be measured more accurately, with generally higher total amounts (Juster, Smith, & Stafford, 1999). The drawback, however, is that each of these components can lead to a missing value, and if there are more components, the chance that at least one of them is missing is generally higher, and thus the chance that the resulting aggregate measure of total wealth will be missing is also higher, often substantially so.

To reduce the amount of uncertainty about the asset values, the HRS has pioneered the use of *unfolding brackets* after an initial missing value. A typical sequence of questions about an asset starts with a question about whether the household owns the asset. If the answer is yes, the respondent is asked about its value. If the respondent fails to report a continuous value, the HRS asks a question of the form "Is it more than X, less than X, or about X?", where the X is a certain amount. Often, respondents give an answer to this, which narrows down the range of uncertainty. If respondents give a "more" or "less" answer, they are asked another question of the same sort, but with a different amount (higher if they answered more and lower if they answered less). If the respondent again answers this question, this further narrows the range of uncertainty. If the respondent answers both questions and does not give a "don't know" or "refuse" answer, this is classified as a *complete bracket*, and if only the first is answered, this is an *incomplete bracket*. If none of these questions are answered, this is a *no value/bracket*. The use of unfolding bracket sequences greatly reduces the range of uncertainty and thus improves wealth measurement and alleviates problems with nonrandom missingness (Juster & Smith, 1997).

We can also classify the result of an unfolding bracket sequence as an *open bracket*, which does not have a finite upper bound, or a *closed bracket*, which does have a finite upper bound. Generally it is assumed that the lower bound is zero, although in principle negative values would be possible for certain assets, for which the net value is asked (e.g., vehicles), after subtracting any loans on the asset.

Like most surveys, the HRS *imputes* missing asset values. The imputations are performed by RAND. They take the information from the unfolding brackets into account, as well as information about variables (covariates) that are related to the missing values. The procedure is described in chapter 6. Note that imputation is not the same as *prediction*.

The goal of the latter is to find the value that is the best estimate of the value of an asset of a given household, whereas imputation involves drawing from an estimated conditional distribution. The imputations are not generally the best possible predictions at the level of the individual household, but they lead to better estimates of marginal and joint distributions, and consequently of parameters of interest that depend on these distributions. By using a rich set of covariates, imputation allows unbiased estimation even if the missingness is selective, as long as it can be considered random *conditional on* the covariates. This is called *missing at random* (MAR), whereas unconditionally random missingness is called *missing completely at random* (MCAR), which is a much stronger assumption. See, for example, Little and Rubin (2002) for a discussion of these terms, their implications, and the relative virtues of imputation versus prediction and of different imputation methods.

RAND disseminates the imputations in the income and wealth imputations file (Pantoja et al., 2015), and in more aggregate form as part of the main RAND HRS file (Chien et al., 2015). To indicate the amount of information available for a given case, most asset variables are accompanied by one or more imputation flag variables. For asset values that are asked directly, these flags have the following categories:

1.continuous value
2.complete bracket
3.incomplete bracket
5.no value/bracket
6.no asset
7.DK ownership
9.no Fin Resp

Some assets in the RAND HRS combine information from multiple questions. For example, the RAND HRS variable HwAIRA combines information about up to three IRAs, each of which may have missing information. This involves a more extensive sequence of questions, for which it is more difficult to classify the information into the imputation categories above. Hence, for these variables, a simpler imputation flag is provided, which has the following categories:

0.no asset
1.no imputations
2.some imputation
9.no Fin Resp

Broad aggregates of wealth do not have their own imputation flags, but the RAND HRS indicates how they are aggregated from components and subaggregates, which do have their own imputation flags.

Appendix E: Wealth imputation models

In this appendix, we describe the imputation models in more detail. The description here applies to both cross-sectional and cross-wave imputation, as they use the same methods. There are two important differences between the two: the set of covariates and the sample. Regarding the covariates, the cross-wave imputations add the ownership and transformed value of the asset in the previous and next wave as additional covariates, as well as marital status change dummies, as descibed in chapter 6. Regarding the sample, the cross-wave imputations involve three estimation and imputation methods, depending on whether information is available from the previous wave, the next wave, or both. Again, this is described in chapter 6. Below, the description of the samples should be interpreted as being the mentioned subset of the total sample, restricted to households with the necessary adjacent-wave information.

E.1 Ownership

If the respondent did not report whether or not the household owns a given asset, ownership is imputed. The standard ownership imputation method is taking a random draw from a binary logit model. The sample on which this model is estimated is the set of all households that reported whether or not they own the asset. However, if fewer than 50 respondents reported owning the asset, the model is replaced by the unconditional fraction who reported owning the asset, which is mathematically equivalent to a logit model with no covariates, except that in principle probabilities of 0 and 1 are possible with the marginal probabilities.

An exception to this imputation method occurs when non-ownership of an asset can be deduced from non-ownership of another asset. For example, if the household does not own a home, they do not have a mortgage either. In the HRS questionnaire, this is reflected by a skip pattern: if the respondent reports not owning the home, the mortgage question is skipped. In this case, we perform "logical imputation" by setting the mortgage to not owned. When the respondent answers "don't know" or refuses to answer the home ownership question, the mortgage question is also skipped. Then imputation flags of both the home and the mortgage are set to "DK own". When the home is imputed as owned, mortgage ownership is also imputed. However, when the home is imputed as not owned, we follow the logic of the HRS questionnaire and set the mortgage also to not owned.

There are multiple home loan variables for which this dependence occurs: first mortgage, second mortgage, other home loans, and equity line of credit. Having a second mortgage depends in a similar way also on having a first mortgage: if first mortgage is reported or imputed as not being owned, second mortgage is also set to not owned. A similar situation happens with first, second, and remaining IRAs. For the primary residence, we have two types of asset: mobile home or "house", where the latter also includes other types of residence, like apartment. Usually, it is known what type of primary residence the household lives in, but sometimes this has not been ascertained. In this case, home

ownership is first imputed for the "house". If that results in owning, "mobile" is set to not owned; otherwise, "mobile" (non-)ownership is imputed.

E.2 Bracket

If the respondent reported owning the asset but did not report a continuous value, the respondent is asked a set of unfolding bracket questions. If the sequence of unfolding bracket questions is not completed (i.e., ends prematurely with a don't know or refuse answer), the result is an "incomplete bracket", or if no information is obtained through the unfolding bracket sequence, it is classified as a "no value/bracket" response. In these cases, we do know ownership, but we do not have complete bracket information, and we first impute a complete bracket. Analogously, if we have *imputed* that the household owns the asset, we have no bracket information and we impute a complete bracket.

Imputation of a complete bracket is achieved by taking a draw from an ordered logit model. The sample on which this model is estimated is the set of all households that reported a complete bracket. If the size of this sample is less than 50, the model is replaced by the unconditional fractions in each of the brackets, which is mathematically equivalent to an ordered logit model with no covariates. However, these marginal probabilities are only used to impute complete brackets from (informative) incomplete brackets. With fewer than 50 complete bracket reporters, no complete bracket is imputed for "no value/no bracket" and "DK own" cases. These cases are then also treated differently in the amount imputations, see below.

There are a few other situations in which no bracket is imputed. First, there are some assets (in some waves) for which no unfolding bracket questions were asked. For example, the amount currently owed on a home equity line of credit does not have an unfolding bracket sequence in all waves. Without bracket questions, we cannot impute brackets either. Second, for questions where bracket questions were asked, but no-one reported a complete bracket, or all respondents who reported a complete bracket reported *the same* complete bracket, we do not impute a bracket either. Finally, if complete brackets were reported, but there is no overlap between the reported complete brackets and the reported incomplete brackets, then we cannot impute complete brackets either. Again, these cases are treated differently in the amount imputations, see below.

E.3 Amount

Several imputation methods are used for imputing continuous amounts, depending on the following factors:

- *Bracketed item*: As mentioned in section E.2, some assets in some waves did not have unfolding bracket questions, that is, they are not bracketed items.
- *Reported or imputed bracket*: We make a distinction between a *closed* bracket, which has a finite lower and upper bound, and the *top* bracket, which has a lower bound but no upper bound. Furthermore, as mentioned in section E.2, in some

cases, an incomplete bracket was reported, or no value/bracket, or ownership was imputed, but no complete bracket was imputed.

- *Number of continuous reporters*: Number of households that reported a continuous value.
- *Number of compatible donors*: Number of continuous reporters with a reported amount within the (reported or imputed) bracket of the household that needs imputation. Even if no complete bracket was imputed, the number of compatible donors still takes into account the bounds of the incomplete bracket, if this is available. Otherwise, this is the same as the number of continuous reporters.
- *Max incomplete* ≤ *Max complete*: Whether the highest upper bound of the reported incomplete brackets (excluding "no value/no bracket" and "DK ownership") does not exceed the highest upper bound of the reported complete brackets.

In the unfolding bracket sequence, each question is of the form "Does it amount to less than $X , more than $X , or what?", where X is a numerical value called the *breakpoint*. The response categories are "Less than $X", "More than $X", and "About $X". The "about" answers are treated as a complete bracket in the bracket imputation and in determining the number of continuous reporters for the amount imputation. This means that, for example, in the ordered logit model, "About $5,000" is the bracket ordered between (say) $0-$4,999 and $5,001-$9,999. For the amount imputation, the amount $X is imputed for households who report this bracket. Note that this is different from treating it as a continuous report. In particular, these observations are not in the estimation sample of the models that use continuous reports for estimating a model for the values, and not in the potential donor pool of the nearest-neighbor imputations. Conversely, they are (necessarily) part of the estimation sample of the ordered logit model.

The most common imputation method is the **nearest neighbor** approach. This is primarily used for closed brackets. In this method, the model estimated is a linear regression model in which the dependent variable is the inverse hyperbolic sine (IHS) of the value. The sample on which this model is estimated is the set of all households that reported a continuous value. For the imputation, the predicted value (on the IHS scale) of the household that has a missing value (the recipient) is compared with the predicted values of the households that reported a continuous value (the donors). For each recipient, the set of donors is first limited to those households that reported a continuous value in the bracket that was reported by (or imputed for) the recipient household. The selected donor is the household whose predicted value is closest to the predicted value of the recipient. This donor is called the nearest neighbor. The imputed value for the recipient is the *reported* value of the nearest neighbor. If there are multiple donor households that are equally near (typically because they have the same predicted value), one of them is randomly selected as the donor.

If there are fewer than 50 continuous reporters, the regression model is not estimated and a (conditional on bracket information) **hot deck** is used, provided there are at least two potential donors for the recipient. Here, all recipients and donors with the same bracket are randomly ordered, and for each recipient the "most recent" donor before the recipient

is selected, and the value reported by this donor is imputed. Technically, the hot deck is the same as a nearest neighbor method without covariates.

The second most common amount imputation method is the **tobit25** method. This is used for imputing amounts in the (open) top bracket. The nearest-neighbor method is not used in this case, because it was found that this tends to impute too many outliers. (See Lee et al., 2015, who also find evidence for this phenomenon.) Instead, a draw from a model is imputed. The sample on which this model is estimated is the set of all households that reported a positive continuous value. The model is a loglinear regression model with homoskedastic normally distributed errors. However, to minimize the impact of the left tail of the distribution on the estimates of the model (this is for imputing in the top bracket, so the right tail is more relevant), observations between zero and the first quartile (25th percentile, which in the data is always lower than the lower threshold of the top bracket) of the observed reported values are censored at the first quartile, and a censored regression (tobit) model is estimated instead of an ordinary linear regression. In the imputation stage, a draw is taken from the estimated distribution, conditional on the given (top) bracket. Because the log of the value is taken to be conditionally normally distributed, the amount is conditionally lognormally distributed. Thus, the draw is from a left-truncated lognormal distribution with truncation point equal to the lower bracket threshold of the top bracket.

If there are less than 100 continuous reporters, the tobit model is estimated without covariates, which amounts to estimating a marginal lognormal distribution, and imputations are done from this distribution, again truncated to make sure the imputations are within the top bracket. However, if the number of continuous reporters is less than 50, the hot deck is used as in the closed brackets.

The **tobit0** method is similar to the tobit25 method, except that at the estimation stage it does not censor observations at the 25th percentile, but at $0. This makes the model equivalent to a loglinear regression model with normally distributed homoskedastic errors (on the log scale). At the imputation stage, a draw is taken from the estimated distribution, conditional on the given bracket information. Because the log of the value is taken to be conditionally normally distributed, the value is conditionally lognormally distributed. Thus, the draw is from a possibly doubly-truncated lognormal distribution with truncation points equal to the minimum and maximum value that are consistent with the reported bracket information. This method is used in some cases where no complete bracket was imputed and the reported incomplete brackets have a higher maximum threshold than the reported complete brackets.

There are a few rare cases in which not enough information is available to use one of the mentioned methods. In such cases, an ad hoc method is used, which varies between waves and assets. This may, for example, be imputing from a normal distribution for which the mean and variance are taken from a "similar" asset.

Table E.1 details which imputation method is used for which cases. Note that there are two cases where the current programs do not impute anything. These cases have not occurred in the data, but when they do, these will be imputed in an ad-hoc way as well.

Table E.1: Amount imputation method used

Bracketed item	Reported or imputed complete bracket	Number of continuous reporters	Number of compatible donors	Max incomplete ≤ Max complete	Amount imputation method
No		0	0		Ad hoc
No		1	1		(None)
No		2–49	2–49		Hot deck
No		50+	50+		Nearest neighbor
Yes	No	0	0		Ad hoc
Yes	No	1–49	0, 1	Yes	(None)
Yes	No	1–49	2–49	Yes	Hot deck
Yes	No	1–49		No	Tobit0 without covariates
Yes	No	50+		Yes	Nearest neighbor
Yes	No	50–99		No	Tobit0 without covariates
Yes	No	100+		No	Tobit0
Yes	About				Breakpoint value
Yes	Closed	0	0		Ad hoc
Yes	Closed	1–49	0, 1		Ad hoc
Yes	Closed	1–49	2–49		Hot deck
Yes	Closed	50+	0, 1, 2		Ad hoc
Yes	Closed	50+	>2		Nearest neighbor
Yes	Top	0	0		Ad hoc
Yes	Top	1–49	0, 1		Ad hoc
Yes	Top	1–49	2–49		Hot deck
Yes	Top	50–99			Tobit25 without covariates
Yes	Top	100+			Tobit25

E.4 No financial respondent

Occasionally, it happens that there is no financial respondent in the household. It also happens that the designated financial respondent does not report anything in the income and assets section (section Q), for example, when the interview is prematurely broken off. In this case, the household is also classified as "no financial respondent" for the income and asset components reported in section Q. Analogously, if the financial respondent

does not report anything in the housing section (section H), the household is classified as "no financial respondent" for the housing section.

Imputations for households without a financial respondent are done separately from the imputations with a fincancial respondent. The procedure is identical to the one described above, except that the list of explanatory variables that enter the principal components analysis differs, and thus all imputation models are estimated separately for this set of covariates. For financial respondent imputations, the covariates for the models include demographics, education, health, cognition, bequest expectations, and income variables. For no financial respondent cases, the income variables are omitted.

For the cross-sectional imputations, there is also a difference in the estimation samples and donor pools. In the tobit25 model, the first quartile is now based on the sample that includes both the reported values and the values imputed for households with a financial respondent. Furthermore, the households with a financial respondent but without a reported value are now also part of the estimation samples of the tobit25, tobit0, and nearest-neighbor models, in which the imputed values are used as if they were reported values. These households are also available as potential donor households in the nearest-neighbor and hotdeck methods. For the cross-wave imputations, the estimation samples and donor pools are the same for the non-financial respondent imputations as for the financial respondent imputations; previously imputed households do not enter the estimation sample and donor pool.