Economic Measurement in the Health and Retirement Study

Steven F. Venti
Dartmouth College and NBER
January 21, 2011

Abstract

The Health and Retirement Study (HRS) is widely use for research on the well-being of the elderly. This paper assesses the quality of economic and financial variables in the HRS. I find the coverage is comprehensive and the quality of the data is uniformly high. Thus the HRS has earned its position as the most widely used data source for research on retirement, saving adequacy, pension policy and a host of other aging-related topics. I identify two general areas that continue to merit special attention. The first is measurement error, particularly errors arising from item non-response and from inaccurate respondent reports of the ownership and level of assets. The second is the quality of the pension data. Where appropriate, I make suggestions for improving economic measures in the HRS.

This review was prepared for the National Institute on Aging HRS Data Review Committee. The author thanks Al Gustman, Nahid Tabatabai and David Wise for comments and many discussions and an anonymous referee for useful comments.

Population aging presents formidable challenges to countries all over the world. To develop policies to deal with the consequences of population aging researchers need to understand the economic, health and social implications of aging. The Health and Retirement Study has pioneered the collection of data that has improved the scientific understanding of the aging process and the HRS has become the role model for surveys in many countries desiring to achieve the same goals. Among economists it has become the dominant data source for research on important policy questions concerning retirement, income, saving adequacy, pensions and other age-related issues. The HRS has earned this position. It is well designed in the sense that the survey content corresponds closely with the major policy questions in the economics of aging.

On the whole I find the economic questions in the HRS to be clear and comprehensive. I can think of no significant economic variables (other than consumption variables that appear in CAMS) that are not present in the survey. As a consequence, I recommend no major changes in the question sequences or other aspects of the survey design. The current design is working well and the cost of losing longitudinal consistency by modifying the existing questionnaire far outweighs the gain from redesign. Nonetheless, the HRS, like any other survey, has not achieved perfection. Below I discuss some remaining data quality problems with the economic measures in the HRS and make some recommendations. These recommendations, for the most part, preserve the integrity of the current survey and involve either adding questions or "post-processing" the data to improve the accuracy of responses.

Values for most economic variables are obtained in the *Assets and Income* and *Housing* sections of the survey. The data obtained in these sections have clearly benefitted from the innovative use of unfolding brackets and from a questionnaire sequence that integrates questions about stocks and flows for each asset. The data are of uniformly high quality - non-response is low and the values obtained are mostly "sensible", but troublesome outliers are still plentiful enough to confound some analyses

Other economic variables are obtained from the *Employment* section. Here data quality is perhaps not quite as good. There is more missing information and, particularly in the pension realm, a rather high level of internally inconsistent or contradictory information provided by respondents. The fault lies not so much in the sample design or the execution of the survey, but is instead a consequence of the extraordinary lack of knowledge displayed by respondents. This unawareness, particularly of pension features, has been well documented elsewhere (Gustman and Steinmeier (2004), Gustman, Steinmeier and Tabatabai (2008), Dushi and Honig (2008), etc.). Obtaining accurate pension information from respondents presents a formidable challenge and the HRS team has spent a great amount of effort on this in the past (using administrative records, having respondents provide financial statements, experimenting with alternative ways of identifying plan type, etc.). There still is more work to be done in this area. Below I will make

some recommendations to improve the quality of the pension data, particularly the data on 401(k)s.

My main focus will be data quality issues that arise using the panel or longitudinal data properties of the HRS, although some of the data quality issues discussed also arise in the cross-section. I focus on measurement errors in the data, particularly those arising from item nonresponse and from inaccurate respondent reports of the ownership and level of assets. Some of these data quality problems, such as item non-response, have already been dealt with in a way that is more than satisfactory for cross-section analyses. However, these data quality problems pose more serious concerns for longitudinal analyses. Indeed, some quality improvements that reduce measurement error in the cross-section analyses may enhance the effect of measurement error in longitudinal analyses.

Section 1 provides two examples to illustrate the problems caused by measurement error. Section 2 provides some suggestions to improve the quality of the non-pension economic measures. Section 3 discusses some additional issues concerning pensions.

1. Measurement Error Due to Inaccurate Reports and Item Non-response

There are two related sources of measurement error that cause difficulty. One is item non-response. The HRS has greatly reduced non-response through the innovative use of unfolding brackets and by using a questionnaire sequence that integrates questions about stocks and flows for each asset. In the HRS, values missing due to non-response are imputed and the imputation procedures themselves may be a source of measurement error, particularly in longitudinal analyses. The other source of measurement error is an inaccurate report by a respondent (or the rare coding or processing error). In this case the HRS takes a "what you see is what you get" approach - all reported values appear in the data files regardless of how questionable they may be. Despite the extraordinary efforts of the HRS team, measurement error remains a serious problem for investigators using the HRS. In this section I provide two examples to illustrate the severity of these problems. My concern is not the statistical properties of the estimates when there is measurement error in the data nor with estimation techniques that may minimize the effect of measurement error. Instead, these examples are intended to show how data quality problems can affect research outcomes and to highlight the need to get the data "right."

1.a. The cross-section distribution of health care costs

Recent work on the distribution of out-of-pocket health care costs provides an illustration of the importance of measurement error and how its treatment by investigators can have enormous effects on research outcomes.[1] Interest in the distribution of health care costs -

[1] The source of measurement error here is potentially inaccurate reports by respondents, but much of what is said here would also apply to measurement error resulting from item non-response.

particularly the upper tail of the distribution- has been sparked in part by a number of recent studies attempting to determine if households are adequately prepared for retirement (Scholz, Seshadri and Khitatrakun (2006), Scholz and Seshadri (2009), Hurd and Rohwedder (2006, 2009b), Munnell, Soto, Webb, Sass and Muldoon (2008), De Nardi, French and Jones (2010)). In particular the study by De Nardi et. al. argues that households need to maintain high levels of assets in retirement because there is a small probability (increasing with age) that the household will face catastrophic health care expenses in late life. Thus the upper tail, or the "outliers", in health care spending is perhaps the key feature of the distribution of health care costs. In most analyses outliers are a concern because they may distort estimates of the mean. Here the outliers are the story so there is a premium on ascertaining their accuracy.

De Nardi, French and Jones (2010) use all values reported in the AHEAD data to construct their distribution of health care costs. Other researchers do not believe all outliers are "true" values and take steps to minimize their influence. Marshall, Skinner and McGarry (2009) express concern about the "potential for erroneously high values" and impose caps or limits on individual components of health care costs. Webb and Zhivan (2010) recode some expenditure categories that they judge to be "implausibly large in relation to the household's income and assets." Hurd and Rohwedder (2009a) compare the distribution of costs in the HRS to the distributions reported in other surveys and conclude that "the level and risk of out-of-pocket spending on health care are exaggerated in the HRS." They suggest that users of the HRS data either examine outliers on a case-by-case basis to determine if reported expenditures are consistent with household resources or use methods to "shrink" reported values in the HRS to levels consistent with external benchmarks.

It is difficult to reach consensus among research studies if each author must arbitrarily decide whether to exclude, censor, or impute particular observations. In this example of recent research on saving adequacy, the conclusions different authors reach may hinge on the accuracy of a small number of outliers.

1.b. Longitudinal estimates of IRA assets

Inaccurate respondent reports of asset values and (to a lesser extent) item non-response also cause difficulties in longitudinal analyses that consider wave-to-wave "changes" in economic variables. Indeed, many of the problems are more severe in longitudinal data. Most economic variables are persistent, that is, they are positively serially correlated. Persistence implies that values in adjacent waves tend to be similar, so wave-to-wave changes tend to be small even if the levels of the variables are substantial. Thus a value measured with error in a particular wave may be quite different from the waves before and after. For example, if a homeowner fails to report ownership in wave T then the HRS records a zero home value in that wave. Thus unusually large "changes" in home value (between waves T-1 and T and between waves T and T+1) are spuriously generated by the erroneous report. The extent of bias or efficiency loss will depend on whether home ownership is used as an independent or dependent

variable, as well as on the properties of the measurement error. At the very least the "changes" created by the inaccurate reports of ownership or inaccurate reports of asset balances are typically much larger than changes that do not involve misreporting and thus the noise can overwhelm the signal.

"Noise" created by inaccurate reports and non-response is a problem common to all asset variables in the HRS. I use data on one particular asset, balances held in Individual Retirement Accounts (IRAs), to illustrate the seriousness of this problem. Table 1 shows IRA balances for a few households (that were selected to illustrate the problem) for the years 2000 through 2006. Each row in the table pertains to a household. The top panel contains households suspected of misreporting ownership of an IRA in at least one year. The bottom panel shows households that may have inaccurately reported their balance in at least one year.[2] It is possible that reporting errors may wash out in the cross-section, but it is clear that in the panel these apparent errors introduce a great deal of noise as judged by the wave-to-wave variation that they create.

Table 1. Selected observations showing potential misreporting of IRA ownership and IRA balance (given ownership), 2000 to 2006

(each row is a household)

2000	2002	2004	2006
Suspected of misreporting ownership			
65,000	0	0	500,000
555,000	500,000	0	430,000
155,000	0	150,000	170,000
110,000	120,000	0	172,000
840,000	568,000	0	500,000
20,000	0	322,000	213,000
1,300,000	0	1,810,000	11,150,000
35,000	0	30,000	30,000
165,000	0	199,000	216,000
Suspected of misreporting balance (for IRA owners)			
140,800	171,000	475,000	137,000
225,000	305,000	135,000	525,000
900,000	805,000	15,700,000	2,450,000
570,000	1,550,000	500,000	435,000
800,000	27,000	460,000	440,000
290,000	400,000	55,000	700,000
284,000	215,000	90,000	380,000
466,000	171,000	520,000	806,000

[2] In both cases it is possible that the year-to-year changes are reported accurately. One possibility is that respondents make transfers between DC plans and IRAs when changing employers.

How common are inaccurate respondent reports of the ownership of assets and the balances held in assets? To estimate the frequency of respondent misreports I restrict the sample to households that answered the ownership and asset value questions in all years (i.e. drop all households with a DK, NR, or bracketed amount in any year) . I also restrict the sample to households reporting a positive IRA balance in at least one year. The results, shown in Table 2, suggest that the misreporting of IRA ownership is quite common. This table shows sequences of IRA ownership for 1992 through 2006. Each sequence contains eight "1"s or "0"s that indicate whether the household is an owner ("1") or is a non-owner ("0") in each of the eight waves of the HRS. To summarize the many possible sequences that can arise, I categorize them as belonging to one of five different patterns. The first three patterns are what we expect to observe in the absence of misreporting (given what we know about the persistence of IRA ownership). These are: 1) households that own in all years, 2) households that do not own in 1992 but become and remain owners in future years, 3) households that own in 1992 but become and remain non-owners in subsequent years. The final two patterns are more likely to be consistent with misreporting: 4) households that both begin and end ownership within the eight wave period and especially 5) households that exhibit "gaps" in ownership.

Table 2. How common are innaccurate reports of IRA ownership, 1992-2006?

	pattern	examples	percent of cases
1	own in all years	11111111	37.5
2	begin and continue ownership after 1992	01111111 00111111 etc.	10.8
3	end ownership before 2006	11100000 11110000 etc.	11.2
4	begin and end ownership	00111100 00001100 etc.	15.6
5	gaps in ownership	11100111 00110011 etc	**24.9**

Note: a "1" indicates ownership and a "0" indicates no ownership in each year
Universe: all households who own an IRA in at least one year and have no missing data

Over 40 percent of the sample exhibit a pattern that is consistent with misreporting and about one-quarter claim they go in and out of IRA ownership. Of course, some of these cases may be legitimate, but it is unlikely that younger respondents terminate an IRA because of the penalty on withdrawals before age 59 and 1/2 and older respondents cannot restart an IRA if they are retired (no earned income) or if they are over age 70 and 1/2 (even with earned income). I emphasize also that the sample is restricted to households that answered the ownership question, so none of this apparent misreporting error is attributable to imputation for non-responses. The results are even more dramatic for other asset components - for example, over 80 percent of households holding bonds in at least one year display the last pattern (although there is no reason why holding bonds should be as persistent as holding IRAs).

Given ownership, the misreporting of balances is also common. As a rough measure of the likelihood of inaccurate reports I identify households who report either positive or negative "spikes" in IRA balances. Of course, some spikes may legitimately occur if respondents hold undiversified portfolios, particularly if IRA assets are held in their employer's stock. The sample includes all households that owned an IRA in all years 1992 through 2006 and report ownership and balance in each year. I find that 31.4 percent of these households report at least one instance of either a negative spike (the IRA balance falls by more than half between one wave and the next and then more than doubles in the subsequent wave) or a positive spike (the IRA balance more than doubles from one wave to the next and then falls by more than half in the subsequent wave). I emphasize again the problem for longitudinal analysis is that these potential misreports create "spurious" wave-to-wave changes in assets that are common and typically much larger than "legitimate" wave-to-wave changes.

I have used the term "apparent misreports" to describe the unusual patterns of ownership and the spikes in balances described above. As I have tried to emphasize, we do not know if these patterns really are the consequence of inaccurate reports. A possible explanation is that the large changes identified above are the consequence of asset reclassification, that is, the respondent may report some stocks as a 401(k) in one wave, an IRA in another wave, and a non-tax-advantaged stock account in another wave. If this is the case, then the large changes in IRA balances that are suspect should be offset by opposing changes in other components of wealth. I have looked for this on a case-by-case basis and found little evidence for this possibility. The large changes in IRA balances for the most part carry through to large changes in total wealth. Smith (1995) found similar results for a broad range of assets in his analysis of measurement error in the first two waves of the HRS. Another possibility is that households really did experience large capital gains or losses. This explanation is difficult to evaluate without additional information on the assets held in IRA portfolios, but the large changes we observe do not appear to be associated with the dot.com bubble between waves 5 and 6. Still another possibility is that the large changes in IRA balances are legitimate transactions - households really did buy and sell and buy and sell IRA assets. Given legal restrictions, this is unlikely to be true. Moreover, as a general rule the buying and selling of assets reported in the *Asset Change*

section does not account for most of the large changes in assets reported in the *Assets and Income* section[3]

2. Some suggestions for improving the quality of economic measures

If measurement error is a potential serious problem, what can the HRS do to reduce it? I offer a few suggestions below that I believe should be considered.

2.a Make better use of asset call-back procedures

A promising, but thus far unexploited, solution to the problem of inaccurate reports is to identify households that report dramatic changes in assets and then "call back" these households to reconcile conflicting reports. This has been done in the *Asset Verification* (or Asset Reconciliation) section.[4] For example, in 1998-2000 waves households were called back if the wave-to-wave change exceeded $50,000 for an individual asset or if the wave-to-wave change exceeded $150,000 for total wealth. Eighteen asset components were considered. A complete description of the methodology can be found in Hill (2004). The call-backs revealed that for 50 to 60 percent of the asset changes flagged, either the 1998 or the 2000 report was erroneous. The RMSE of models with the change in total net worth as the dependent variable was reduced by 30 percent after the inaccurate reports were corrected. There appears to be much potential for these call back methods to improve the quality of the data. The data have been obtained in each wave since 2002 but, to the best of my knowledge, have never been used to "correct" outliers in a research study. Also, reconciled values based on the Asset Verification modules have not been integrated in any systematic way into the data products made available by the HRS or RAND. It would be useful to make reconciled values available in public release data files (to supplement to the unreconciled values). It would probably also be better if the Asset Verification section used percentage thresholds (accompanied by some other criteria to exclude changes that are small in absolute terms), rather than absolute thresholds, to flag questionable asset changes.

2.b Add a round-up section that flags longitudinally inconsistent responses

For some asset components reconciliation of longitudinally inaccurate reports can be done "on-the-fly" by pre-loading prior wave values in the computer assisted interview system. The technology is already in place to do this. To preserve longitudinal consistency the respondent should only be prompted to resolve conflicts after the traditional survey is completed. Moreover, both the original and reconciled responses should be released.

---

[3] The *Asset Change* section of the HRS inquires about "active saving" by households, that is, whether they have bought or sold assets since the last wave. Thus large changes in assets reported in the *Assets and Income* section should match-up with purchases or sales in the *Asset Change* section. I have looked at this for stocks and bonds and found that reported sales and purchases did not "explain" large changes in asset balances. Purchases and sales of IRAs are not included in the *Asset Change* section so I could not perform this comparison for IRAs.

[4] The HRS staff may also make direct inquiries to verify reported values. Wave 9 includes a respondent who claims to receive a pension of $60,000,000 annually. The HRS RAND documentation states "The accuracy of this amount has been confirmed by HRS."

2.c Use external information to validate or reconcile potential misreports

Validation studies have been widely used to gauge the extent of measurement error in survey data. Bound, Brown and Mathiowetz (2001) survey validation studies for several different economic variables. Bricker and Engelhardt (2008) compared respondent-reported earnings in the HRS to W-2 records. Unfortunately, validation is likely to be more difficult to implement for financial assets because, unlike earnings, there is no central registry of financial records. Federal tax returns provide only limited information on asset ownership and income from assets. In any event, I suspect that obtaining IRS permission to link to tax records is almost impossible, although respondents could be asked to refer to or provide copies of their tax filings. It may be more practical to ask respondents (or perhaps to ask respondents who appear to have misreported balances) for copies of financial statements. This has already been done for a small subsample of HRS respondents for 2004, 2006 and 2008 for financial assets held in pension plans. In these years some respondents were asked to provide copies of pension plan documents they received from their pension plan or from the financial institution administering their pension. Although the response rate was low and the resulting sample size small (less than 200 cases) the request for statements did improve the accuracy of responses. In particular, about a third of those providing documents showing they had a balance did not report having a balance in the core survey.

Another possibility is to attempt to determine the accuracy of outliers by comparing the distribution of values for a particular asset in the HRS to the distribution of values obtained from an external source. This is the approach adopted by Hurd and Rohwedder (2009b) to "correct" the upper tail of the distribution of out-of-pocket health care costs. For example, the distribution of financial assets in the HRS might be compared to the data from the Survey of Consumer Finances (that over samples high income households), to data obtained from industry sources, or to data reported to the Federal Reserve Board.

2.d Create a set of longitudinal imputations for non-response

Missing values are imputed for most of the wealth and income variables in the HRS.[5] There are many different levels of missing data: a household may not report ownership for particular asset, the household may report ownership but not the value, or the household may report ownership and provide either complete or incomplete bracket information. In all cases imputations are made based on information from the wave for which the data are missing. The RAND imputations use a regression based method that includes a long list of contemporaneous-wave covariates and the HRS uses a hotdeck method in which a missing value is replaced with a similar record from the same wave. No cross-wave information is used in either the RAND or the HRS imputations.

[5] There are two sets of imputations available. The imputations in the HRS public release files do not use a procedure that is consistent across waves. The imputations in the RAND release are consistent across wave.

As noted above, many economic variables are persistent, making values in adjacent waves similar. Cross-section imputations do not use the persistent component and thus increase the "noise" in wave-to-wave changes. It would be useful to encourage exploration of alternative methods of imputation that make use of both longitudinal and cross-section information. In most survey innovations the HRS is the leader of the pack. Not in this case. Both the CPS and the SIPP have already experimented with longitudinal imputation.

2.e Other changes to keep the HRS up-to-date

Over the years the HRS has done an excellent job updating the questionnaire to reflect changes in the broader economic environment. The HRS may want to consider (if not already incorporated in the 2010 wave) adding to or modifying questions inquiring about reverse mortgages, mortgage foreclosure and bankruptcy, and to separately distinguish between Roth and Traditional IRAs (because of different tax treatment the balances in these accounts are not comparable). In addition, the recent financial crisis has highlighted the need to know household exposure to the stock and housing markets. The HRS has housing pretty well covered. The information on exposure to equity markets is less complete for two reasons. One is that many households now own balanced funds or target-date (lifecycle) funds that combine stocks and bonds. These funds - being neither stocks nor bonds - are now large enough to deserve their own place in the list of assets owned. The second reason is that we have only limited information on stock holdings in IRAs and 401(k)s. For example, the survey asks what proportion of the IRA is invested in "stocks or mutual funds" (less that 50%, 50%, or greater than 50%). A more precise percentage would be useful, as would a more detailed list of asset categories. The current wording presents "stocks and mutual funds" as a single category of investment but, given the growth of bond mutual funds, an investment in this category could be all stocks, all bonds or a mix.

3. Pension Wealth Issues

Pension wealth is one of the largest assets on the household balance sheet and is important for understanding retirement decisions and financial security. It is also the most poorly measured component of wealth in most household surveys. The HRS is the pioneer among household surveys in the collection of detailed pension information needed to calculate pension wealth. The pension sequence remains one of the unique features of the HRS.

Despite an enormous amount of effort, obtaining consistent and accurate data on pension plan characteristics has been an elusive objective. This is not surprising. No previous household survey had attempted to gather pension plan provisions that were as detailed as those incorporated in the original HRS in 1992. The original HRS pension sequence was a milestone. It was also the start of a learning experience. Since its inception the HRS team has continued to make small modifications to the survey to improve the accuracy of responses. In some cases they have been swimming upstream since many respondents do not know enough about their

pensions to provide more accurate information. Since 1992 the pension sequence has also been modified (or in most cases "added to") to keep pace with changes in the ever-evolving pension landscape. In recent years the HRS has added questions about holdings of company stock, whether respondents have changed their portfolio allocations, whether respondents can choose how their (or their employer's) share of the contribution is invested, if automatic enrollment is a feature, and whether the respondent receives a quarterly statement. In addition, feedback from several experimental modules resulted in a substantial redesign in 2008. The biggest change was to minimize the number of questions that were conditioned on plan type. All in all, the HRS team has continually improved the quality of the pension data. It still provides the best household survey data available to calculate pension wealth.

3.a Defined benefit plans

Over the past 30 years the number of workers relying on DB plans for retirement income has declined dramatically in the U.S. This economy-wide trend is driven by younger cohorts and is not yet reflected in the HRS data. Gustman, Steinmeier and Tabatabai (2010b) find that in 2006 DB plans are still the dominant pension plan (as measured by assets), even for the youngest HRS cohort, the Early Boomers. The *Employment* section provides respondent reports about the features of the DB pension plan. Respondents are asked if their current (or former) employer offers a plan and then about the plan type (DB, DC, or a combination) . If the respondent indicates a DB plan (or a combination plan) additional questions are asked about plan characteristics that determine future benefits. In principle, pension wealth can be calculated from these responses. As a practical matter the process is fraught with error. In Venti and Wise (1999) I calculated DB pension wealth for all respondents in 1992 (the estimates are available on the HRS web site) and had to impute one or more key characteristics for most of the respondents. There are now several studies that show that a large minority of respondents can't even correctly identify the type of plan they have. In retrospect it was overly optimistic to expect respondents to be able to provide details of their pensions. Indeed, some of the most enlightening research that has come out of these data are studies showing how little people know about their pensions (Dushi and Honig (2008), Gustman, Steinmeier and Tabatabai (2009, 2010a)).

The HRS team has made an effort to improve the quality of these self-reported data on DB plans. An experimental module tried to infer pension type from a series of questions without asking respondents to define the type. In this module respondents were asked if they had an account balance in their plan, whether their plan sent then a quarterly statement, if they could control the asset allocation, if they could take a loan, if they make any contributions, if they expect to receive periodic payments after retiring for as long as they live, and other questions that could be used to infer plan type. Surprisingly, even these features have only a small effect on our ability to correctly identify pension type from self-reported information. My sense is that nothing that has been tried thus far has resulted in more accurate self-reports so, at least for DB plans, more emphasis should be placed on matched administrative data.

Since its inception the HRS has obtained the name and address of the employer for the current and last jobs held by the respondent. The HRS then attempts to obtain the Summary Plan Descriptions (SPD) from each employer. As long as the correct SPD is linked to each respondent (some companies offer over 20 different pension plans and thus may provide 20 SPDs) then the pension plan provisions are known with great accuracy and are complete enough to make reliable calculations of DB wealth. There are still some shortcomings with these data that the HRS is making an effort to address. Perhaps the most serious shortcoming is that only about two-thirds of the respondents who report pensions on their current jobs were matched to their SPD in 1992 and the match rate has fallen below 50 percent in both 1998 and 2004. The match rates for previous jobs are well below 50 percent. Respondents may fail to correctly provide their employer information, employers may fail to provide SPDs, or in some cases employers did not send all of their SPDs, that is, an employer sent SPDs for their DC plans but not their DBs. The HRS is taking aggressive steps to increase the match rate by obtaining SPDs from other sources, including public records available for most public sector employees. It is also obtaining administrative data on pension plans by matching respondent pensions to Form 5500 records.

3.b 401(k) plans[6]

Almost all information in the HRS on 401(k) balances is obtained in the *Employment* section. For each current or past job, the survey inquires whether the respondent still has a 401(k) with the employer and, if so, what the balance is. For the respondent's *current* job the most serious problem with this approach is that respondents have trouble identifying whether they have a DB or DC pension. If they get the pension type wrong - say misclassify their DC as a DB - then the reported value of the 401(k) balance is zero for the wave. This type of misreporting is common. There are many cases where the 401(k) balance is positive in one wave, zero in the next and then positive again in the subsequent wave. There are also many respondents who say that neither their job nor pension changed between waves, but they classify their pension as one type in the beginning wave and as another type in the ending wave. Table 3 measures the frequency of these apparent misclassifications. The sample for these calculations includes all currently employed respondents who report that both their job and their pension plan remained unchanged in adjacent waves. The Table shows the percent of workers who report a different pension type in adjacent waves. For each wave-to-wave transition, 20 to 30 percent of respondents report a change in plan type (DB vs. DC) despite also reporting that neither their job nor their pension changed. As a result, the time series of DC pension balances for these persons contains many "zeros".[7] Such patterns caused by potential inaccurate reporting of plan type complicate longitudinal analyses based on wave-to-wave changes in asset.[8]

---

[6] 401(k) plans are by far the largest employer-based tax-deferred saving plan. There are also traditional DC plans, 457 plans, 403(b) plans and similar saving plans. I refer to all of these collectively as "401(k) plans."

[7] Gustman, Steinmeier and Tabatabai (2010) look at respondents who remained employed, did not change jobs, and report that their pension did not change between 1992 and 2004. They find that only 37 percent of the

Table 3. Percent of respondents reporting a different answer in adjacent waves to: Do you have at least one DC/401(k) plan?	
wave 1 to wave 2	23%
wave 2 to wave 3	34%
wave 3 to wave 4	31%
wave 4 to wave 5	24%
wave 5 to wave 6	27%
wave 6 to wave 7	27%
wave 7 to wave 8	19%
Universe: currently employed respondents reporting that both their job and pension plan remained unchanged in adjacent waves.	

Many respondents also continue to hold balances in 401(k) plans that originated with *past* employers. Roughly half of the employees with non-negligible 401(k) balances who leave a job keep their funds with the original 401(k) provider. A small fraction (typically with low balances) cash out and the remainder rollover their balances into an IRA. Over the years the HRS has had serious problems obtaining *current* 401(k) balances for accounts associated with *past* employers. The original AHEAD and CODA cohorts - mostly retired at the initial survey- were not asked any questions about 401(k) plans. The HRS/WB/EBB cohorts were asked for 401(k) balances on all current and past jobs in the first interview. In subsequent waves information on 401(k) account ownership is preloaded and each respondent is asked for the current status of each 401(k) plan owned in the past. Some variant of the following question is used:

> Our records show that in (.) you had a pension account from a job where you worked from approximately (.) until (.).
>
> (1) Do you still have the pension account?
>
> (2) Did you withdraw the money?
>
> (3) Did you roll it over into an IRA?
>
> (4) Did you convert the account into an annuity?
>
> (5) Did you transfer it to a new employer or what?

respondents who self-report coverage by only a DC plan in 1992 also report being covered by only a DC plan in 2004.

[8] A possible alternative to self-reports is to calculate 401(k) balances from the 401(k) plan features contained in the employer matched SPD. This approach may be more likely to get plan type correct. This approach doesn't work as well for DC plans as it does for DB plans. To calculate the balance from plan features one would need to know the number of years the respondent contributed, the amount contributed each year, and investment returns.

If the respondent indicated he/she still had an account, then the respondent is prompted for the 401(k) balance.

In principle, each respondent's total holdings of 401(k) assets can be obtained by summing the balances obtained for each past and current pension. Unfortunately, several factors have diminished the usefulness of this series of questions. First, between 1994 and 2002 the preloads were either not done or incomplete because of implementation problems, so balances held in past pensions are severely underestimated in these years (although a correctly implemented preload may have picked these pensions up again beginning in 2004). Even after 2002 there is no preload if the account is acquired as a spousal beneficiary. Second, if a respondent has a 401(k) from a past employer and incorrectly reports an outcome other than "Do you still have the pension account?", then the pension will not be preloaded in subsequent waves and the respondent will not be prompted for the balance in the 401(k) account in the next wave.[9] Although respondents may provide multiple answers, it is probably not uncommon for respondents making withdrawals to check off only that response. A case-by-case analysis reveals that 401(k) balances are "lost" over time, that is, I can identify 401(k)s that vanish, but cannot find the assets that were in the 401(k) anywhere else in the household balance sheet. The cumulative effect of losing these 401(k)s is that 401(k) balances are underreported and, within each cohort, the magnitude of underreporting becomes more serious over time.[10]

The 401(k) questions are currently part of the *Employment* section. There is good reason for this. A major purpose of the survey is to understand retirement decisions and linking each 401(k) to a particular job/employer is critical to understanding retirement incentives. I don't think this feature of the survey should be changed. Yet having 401(k) questions in the *Employment* section creates complexity that, as the above discussion illustrates increases the likelihood that 401(k) balances are under-reported. In the current pension sequence we obtain the correct 401(k) balance only if the respondent correctly identifies plan type (DB vs. DC) in the initial interview, correctly reports that the pension is still active in all subsequent interviews and the preload system is implemented correctly.

---

[9] For all other asset components, including IRAs, a respondent who says they do not own the asset in a particular wave is still asked if they own the asset in the next wave.

[10] In 2006 a "catch-all" question was added to the Employment section to capture balances in 401(k)s that were not preloaded. Respondents were asked "do you have any (other) pension plans from former jobs from which you are not currently receiving any income payments and that you have not already told me about?" Unfortunately this question was only asked of respondents less than age 71 and only then for respondents "not currently receiving any income". In 2006 there were only 392 responses to this question. There is also a question in the *Assets and Income* section that is supposed to pick up 401(k) balances for accounts for respondents "currently receiving other income from retirement pensions", but it contains almost no data (50 responses in 2006), perhaps because more than three-quarters of 401(k) owners under the age of 71 do not make withdrawals (receive income) from their 401(k)s. Additional changes to the question sequence were made in 2008. These changes were aimed at obtaining 401(k) balances from households that didn't know they had a 401(k) or were not asked about a particular 401(k).

The 401(k) balances calculated from the HRS are too low in comparison to other household surveys (SCF or SIPP) and to Investment Company Institute and EBRI provider surveys. The most direct way to obtain 401(k) balances that are not currently reported is to deal with 401(k)s the same way as IRAs are presently handled - ask for 401(k) balances in the *Assets and Income* section. Respondents would just be asked for balances on their first, second, and third 401(k)s without having to first pass screens based on plan type and without having to link each 401(k) account to a particular past or present employer.[11] This is the method currently used in the ELSA to obtain balances in DC plans. The 401(k) balances obtained in this way are likely to be more accurate than those obtained through the more complex employer-based design in the *Employment* section and should adequately serve the needs of most researchers. It is still important to retain the employer-based design of the *Employment* section because it is needed to value DB pensions and to understand the effects of DB pensions on retirement. It is also still useful to know the features of the 401(k) plan associated with the respondent's *current* employer. It is less clear why researchers would need to know which *former* employer originated each 401(k) account, particularly since the identity of the originating employer of many 401(k)s is lost anyway when they are rolled over into IRAs. Finally, if a 401(k) sequence is added to the *Assets and Income* section, the results can be used to judge how serious underreporting of 401(k) balances is in the *Employment* section.

Summary

For very good reasons the HRS has become the most widely used survey for research on the well-being of the elderly. The coverage is comprehensive and the quality of the data on economic variables is uniformly high. I have tried to identify the areas that deserve extra attention in the future. I have argued that the misreporting of asset balances and the imputations procedures currently used have made it difficult for researchers to use the HRS in longitudinal applications. I hope my assessment and recommendations will further enhance the usefulness of the HRS. I want to emphasize that the current HRS question sequence is excellent for most applications. Thus most of the changes I recommend involve post-processing (rather than changing the current survey design) to improve the overall quality of the data when used in longitudinal applications. These changes should be in addition to - rather than a replacement for - the current question sequences and data products. Maintaining consistency with previously released data products is paramount. Any major changes that are made should be released alongside existing data.

I have tried to identify areas in the HRS where data quality issues continue to exist. In some cases I have made recommendations, but in other cases I have done no more than identify shortcomings . Of course, focusing on areas where there is room for improvement rather than on

[11] Note that respondents are being asked for the balances in all 401(k)s including those already mentioned in the *Employment* section. A downside to this strategy is the possible confusion that might results from asking respondents twice about the same accounts - once in the *Employment* section and again in the *Assets and Income* section. It should be made clear to respondents that this is being done.

areas of strength gives a very distorted view of my appreciation of the HRS. The HRS has achieved its dominant position because it is superior to other surveys that could be used in aging-related research. The HRS has played a critical role moving research forward in the past decade. Without it there would be a tremendous void in our knowledge. I cannot imagine what the current state of research on retirement, saving adequacy, pension policy and a host of other aging-related topics would be without the HRS. Overall, the HRS does an excellent job obtaining data on economic variables. It has become well-known as the "gold standard" in aging research. I commend the HRS team for their efforts.

References

Bound, John, Charles Brown and Nancy Mathiowetz. 2001. "Measurement Error in Survey Data," in *Handbook of Econometrics*. Elsevier Inc. vol. 5. p. 3705-3843

Bricker, Jesse and Gary Engelhardt. 2008. "Measurement Error in Earnings Data in the Health and Retirement Study," *Journal of Economic and Social Measurement*. vol. 33. p. 39-61.

De Nardi, Mariacristina, Eric French and John Bailey Jones. 2019. "Why Do the Elderly Save? The Role of Medical Expenses," *Journal of Political Economy*. vol. 118, no. 1, pp39-75.

Dushi, Irena and Marjorie Honig. 2008. "How Much Do Respondents in the Health and Retirement Study Know About Their Tax-deferred Contribution Plans? A Cross-cohort Comparison." Michigan Retirement Research Center working paper no. 2008-201.

Gustman, Alan and Thomas Steinmeier. 2004. "What People Don't Know About Their Pension and Social Security." In W. Gale, J. Shoven and M. Warshawsky (ed.) *Private Pensions and Public Policies*. Brookings Institution. p. 88-113.

Gustman, Alan, Thomas Steinmeier and Nahid Tabatabai. 2008. "Do Workers Know Their Pension Plan Type? Comparing Workers' and Employer's Pension Information." In A. Lusardi (ed.) *Overcoming the Saving Slump*. University of Chicago Press. p. 47-81.

Gustman, Alan, Thomas Steinmeier and Nahid Tabatabai. 2010a. *Pensions in the Health and Retirement Study*. Harvard University Press.

Gustman, Alan, Thomas Steinmeier and Nahid Tabatabai. 2010b. "What the Stock Market Decline Means for the Financial Security and Retirement Choices of the Near-Retirement Population," Journal of Economic Perspectives. vol. 24, No. 1. p. 161-182.

Hill, Daniel. 2006. "Wealth Dynamics: Reducing Noise in Panel Data,." *Journal of Applied Econometrics*. Vol 21. p. 845-860.

Hurd, Michael and Susann Rohwedder. 2006. "Life-Cycle Consumption and Wealth Paths at Older Ages," RAND Manuscript, presented at IZA Conference.

Hurd, Michael and Susann Rohwedder. 2009a. "The Level and Risk of Out-of-Pocket Health Care Spending." Michigan Retirement Research Center Working Paper No. 2009-218.

Hurd, Michael and Susann Rohwedder. 2009b. "The Adequacy of Resources in Retirement, Then and Now." Presented at the AEA Meetings, Atlanta. December.

Marshall, Samuel, Kathleen McGarry, and Jonathan Skinner. 2010. "Out of Pocket Medical Expenditures and Retirement Security in the United States," NBER Working Paper 16170.

Munnell, Alicia, Mauricio Soto, Anthony Webb, Francesca Golub-Sass and Dan Muldoon. 2008. "Health Care Costs Drive Up the National Retirement Risk Index," *Issue Brief 8-3*. CRR, Boston College.

Scholz, John Karl, Ananth Seshadri and Surachai Khitatrakun. 2006. "Are Americans Saving Optimally for Retirement." *Journal of Political Economy*. vol. 114, No. 4, pp. 607-643.

Scholz, John Karl, William Gale Ananth Seshadri. 2009. "Are All Americans Saving Optimally for Retirement." presented at the AEA meetings, Atlanta.

Smith, James. 1995. “Wealth Accumulation Between the Waves,” paper presented at University of Michigan Workshop on HRS Surveys, November

Venti, Steven and David Wise. 1999. "Lifetime Earnings, Saving Choices, and Wealth at Retirement," J. Smith and R. Willis (ed.) *Wealth, Work and Health*. University of Michigan Press. p. 87-120.