University of Michigan Retirement Research Center

Working Paper

WP 2002-030

Ensuring Time-Series Consistency in Estimates of Income and Wealth

F. Thomas Juster, Joseph P. Lupton, and Honggao Cao

MRRC Project # UM02-Q3

"Ensuring Time-Series Consistency in Estimates of Income and Wealth"

F. Thomas Juster
University of Michigan

Joseph P. Lupton
University of Michigan

Honggao Cao
University of Michigan

July 2002

Michigan Retirement Research Center
University of Michigan
P.O. Box 1248
Ann Arbor, MI 48104
www.mrrc.isr.umich.edu
(734) 615-0422

Acknowledgements

This work was supported by a grant from the Social Security Administration through the Michigan Retirement Research Center (Grant # 10-P-98358-5). The opinions and conclusions are solely those of the authors and should not be considered as representing the opinions or policy of the Social Security Administration or any agency of the Federal Government.

1. Introduction

The last decade has seen substantial progress in improving the quality of micro-data on both income and wealth. Some of these developments are documented in recent papers by Juster and Smith (1997), Juster, Lupton, Smith and Stafford (under review, 2001), and Hurd, Juster and Smith (under review, 2001). These papers explore a number of quality enhancements: the use of unfolding brackets for income or wealth components that convert “don’t know” or “refusal” responses into quantitative imputations that contain measurement error but little or no bias; the use of improved estimates of changes over time in wealth and active saving to generate measures of capital gains or losses; the use of a merged questionnaire sequence that integrates survey questions about asset holdings and income flows from these assets to reduce the bias in estimates of income from capital; and finally, matching of the periodicity specified in income questions to the actual periodicity of income receipts as a way to enhance the quality of reports for certain income categories.

These enhancements of survey data on income and wealth, while substantially improving the quality of the cross section data, do not come without a cost. A major problem associated with any change made to the methodology used in a panel survey is that they tend to produce time series inconsistencies. By definition, quality improvements reduce the bias and/or measurement error of the cross section point estimate but, by doing so, introduce a bias in the estimate of the change over time.

One way to avoid producing such a time series inconsistency is to freeze the survey technology, thus eliminating any quality enhancement. As a long run strategy, this is clearly a bad idea – robust empirical findings cannot be obtained from poor data. A preferred alternative would be to develop methods of recovering time series consistency in the face of data enhancements. In this paper, we explore methods of recovering time series consistency in the measurement of income from capital in the Health and Retirement Survey (HRS).

Respondents in both Waves 1 (1992) and 2 (1994) of the HRS were asked to report all sources of income in a stand-alone series of questions. The conventional view is that these questions should be reported together since they all have the characteristic of being resource flows. In a separate set of questions, the value of household assets and liabilities were obtained. Again, the idea was that these are all stock values and thus should be grouped together. However, while this classification of flows and stocks into separate groups is useful from the perspective of the researcher, it may not be the optimum question structure from the viewpoint of the survey respondent. Given that the source of asset income is the asset itself, it makes sense to integrate stocks and flows in a way that allows the survey respondent to consider these dollar amounts at the same time. This innovation was implemented in the HRS beginning in Wave 3 (1996) and continues to be the methodology used in all following waves including Wave 4 (1998) and Wave 5 (2000). Hurd, Juster and Smith (2001) examine the effect of this data collection enhancement and find that the income from capital almost doubles between Waves 2 and 3, suggesting the reduction of a serious bias resulting from the stock/flow separation of asset amounts and income. And as noted in that paper, other surveys, such as the Current Population Survey, also suggest a serious underestimate of income from assets using the conventional survey design that has income from assets reported in one module and the assets reported in a separate module.

Although clearly indicating a substantial improvement in the measurement of asset income, the mean doubling between Waves 2 and 3 of the HRS is problematic for researchers wishing to utilize the panel aspect of the survey. The results of any time series study of HRS asset income will be dominated by this technology change in data collection. To correct the problem we propose a strategy that utilizes the distribution of the rates of return obtained in the unbiased data. Random imputation of asset income rates of return in Waves 1 and 2, using Wave 3 as the donor distribution, are used with the asset values of Waves 1 and 2 to generate an estimate of asset income.

Two crucial assumptions are required if this strategy is to be successful. First, it is assumed that although there is a time series inconsistency in the estimate of asset income, the estimates of asset values are not contaminated by this bias. We provide evidence that the measurement of asset values is indeed consistent over time and that the major source of bias in the rate of return to assets stems from the measurement of asset income. Second, the donor distribution must be an adequate representation of the true distribution in the time period where the imputations are being made. To determine how robust our strategy is to this assumption, we provide imputed estimates based on donor distributions coming from HRS Waves 3, 4 and 5. The stability of the imputed estimates across donor distributions is noteworthy.

In the next section, we examine the HRS data on household financial wealth and income flowing from that wealth. We discuss the possibility of various sources of measurement error in the time series across Waves 1 through 5 and provide the mean rate of return to financial assets in these years. In Section 3, we outline three imputation procedures and discuss their relative advantages and disadvantages. These procedures are applied to the HRS asset income data and the results are reported in Section 4. The robustness of each procedure is examined as are the various imputation strategies. Concluding remarks are provided in Section 5.

2. Survey Structure Induced Bias in the HRS Financial Asset Income

Financial wealth in the HRS is defined as the sum of four components: checking, saving and money market accounts; CD's, savings bonds and Treasury bills; publicly traded corporate equities and equity mutual funds; and corporate bonds. Each of these potentially yields some amount of asset income. Data from the 1992 and 1994 data are based on the conventional survey format while the 1996, 1998 and 2000 data are based on the revised format that integrates questions about asset holdings with questions about income from assets. In the conventional format, respondents are asked whether they own any of the four financial assets, or any investment real estate or business or farm equity, and how much they own if they report owning any. In a later section of the questionnaire, respondents are asked about income from a variety of sources (wages or salary, workers compensation, veterans benefits, business income, rent, Social Security, pensions, interest or dividends, etc.). In the revised question sequence, households are asked whether they have each of the four asset components noted above. If the respondent claims to own a particular asset, they are asked about its value and, if greater than zero, whether they received any dividend or interest income from that asset. If they claim to have asset income, they are ask how much and how often. Similar question sequences are asked for each of the four types of financial assets.

Gross differences in the reporting of financial assets and income from those assets across the five waves are enormous. These are shown in Table 1. In 1992 and 1994, only about a third of the sample reported income from financial assets while almost two-thirds reported zero income from assets. These proportions were approximately reversed in 1996, 1998 and 2000, with almost two-thirds reporting income from assets and a bit more than one-third reporting zero income from assets. Interestingly enough, the proportion of the sample reporting ownership of financial assets is essentially the same on all five waves: the fraction owning financial assets is a bit over 80% in 1992, and goes up slightly in each later year as one would expect during a vigorous economic expansion.

Another way to look at the linkage between assets and income from assets is to examine the proportion of the sample reporting zero income from assets within different asset percentiles across survey years. This is provided in Table 2. In the lowest asset category (zero to the 25th percentile), the proportion of the sample reporting zero income from assets is over 90% in all five survey years, although it is a bit higher in 1992 and 1994 than in 1996, 1998, or 2000. The differences by year become substantial when we look at higher asset percentiles. For example, in the 90th percentile and above, the 1992 and 1994 proportions of households reporting zero income from assets are, respectively, 22% and 36%, extraordinarily high numbers for households in the upper 10% of the financial asset distribution. Integrating the survey questions on asset income into the asset and liabilities module reduces the proportion of households reporting zero income from assets to about 3% in that percentile group. Substantial differences in the fraction of households reporting zero income from assets also show up in the 25th-50th percentile, in the 50th-75th percentile, and in the 75th through the 90th percentile. In the 25th-50th percentile, the fraction of households reporting zero income from assets goes from about 80% using the conventional survey format to between 40 and 50% using the revised format The fraction reporting zero goes from over 50% in the conventional mode to about 12% in the revised mode in the 50th-75th percentile, while going from about a third in 1992 and 1994 to around 5% in 1996, 1998 and 2000 in the 75th-90th percentile.

Tables 3a, 3b and 3c contain a more detailed picture of the change in income from financial assets and in asset holdings over the five survey years and over the percentile distribution of financial asset holdings. The pattern of the data in these tables is very consistent. In Table 3a, which has mean income from financial asset holdings by percentiles of financial asset holdings, the full sample means in 1992 and 1994 are roughly 50% of the means in 1996, 1998 or 2000. This across year mean difference stems largely from differences among households whose financial asset holdings are in the 75th percentile or higher. For example, in the 90th-100th percentile, mean asset income is about $8,000 in 1992 and 1994, but about $18,000, $22,000 and $23,000 in 1996, 1998 and 2000, respectively – roughly a three-fold increase. In contrast, in the 50th-75th percentile, the 1996, 1998 and 2000 data look to be about the same size as the 1992 data, all of which are higher than the 1994 mean.

Table 3b contains mean financial asset holdings across asset holding percentile groups. No pattern difference is evident between the 1992-1994 data and the 1996-1998-2000 data. By year, the mean grows substantially, as one would expect during a period of economic prosperity with substantial capital gains. In the 50th-74th percentile, the mean grows from roughly $15,000 to slightly over $22,500 – a 50% increase over the eight-year period. In the 90th+ percentile group, the mean grows from around $300,000 in 1992 to about $650,000 in 2000 – roughly a two-fold increase. Thus the pattern that one would expect in the absence of any survey innovation is exactly what one finds in Table 3b. Mean financial asset holdings grows steadily and substantially over the 1992 to 2000 period with no indication that the growth rate is affected by the transition from conventional survey methods to the revised method. Generally speaking, the growth rates over the entire period tend to average about 9% per year with growth being larger in the higher percentiles than in the lower ones.

The effect of the revised survey format conditioned on asset holdings is presented in Table 3c which provides the mean of the average rate of return to financial assets, defined as the ratio of financial asset income to financial assets. Note that this is a mean of individual rates rather then the ratio of the means from Tables 3a and 3b. The mean average rate of return over all households increases by roughly 50% from the conventional format to the revised format. This pattern can be seen across the asset groups as well. For households with financial assets above the 90th percentiles, the mean of the average rate of return jumps from 3.3% and 2.4% in 1992 and 1994, respectively, to 4.6% in 1996 after which it stays relatively constant. As evident from Table 3c as well as Table 3a, the asset income data from 1994 seems to be particularly anomalous. One could also argue that the mean average rates of return in 1996 seem to be anomalously above the values in 1998 and 2000. This possibility is considered in more detail below.

The data displayed in Tables 1, 2 and 3 make it clear that time-series analysis of the effect of income change on various types of behavior would be greatly aided if the income component that reflected the return on financial assets could be adjusted to ensure consistency. The problem is that all datasets using the conventional HRS survey design (asking about a long set of income components, including dividends and interest income) will seriously underestimate income flows from financial assets and hence overstate the change across the conventional and revised survey years.

There are at least two potentially important ways in which biased measurement error is introduced into reported financial income from 1992 and 1994 – error in reporting having any asset income, and error in reporting the value of asset income conditional on having any at all. As indicated in Table 2, a striking features of the quality enhancement in measuring income from capital is that the merged question sequence converts the proportion of respondents who report zero asset income from 71% in Wave 2 to 38% in Wave 3. Even more striking is that the merged module converts the proportion of households with financial assets above the 90th percentile who reported zero interest or dividend income from 36% in Wave 2 to 2.3% in Wave 3. Thus, one possibility could be that the bias in reported financial income is generated solely by households who actually have but report no asset income. This would imply no bias among households who reported asset income and thus require the imputation of only those households who report owning assets but no asset income. If we limit comparisons to households reporting some asset income in each year we might find the same degree of time series consistency that we find in the level of asset holdings from Table 3b. If that were true, we could focus on devices for imputing values to households that reported owning financial assets with no asset income in Waves 1 and 2 based on relationships observed in wave 3.

To examine this hypothesis, Table 4 reproduces Table 3c for households who report positive income from financial assets. Among all households, the mean average rate of return for 1992 and 1994 seems much more in line with those from the later waves. However, this masks some remaining time series inconsistencies across the financial asset distribution. The average rate of return for households with financial assets above the 90th percentile, households with by far the most asset income on average (Table 3a), remains roughly 40% lower in 1992 and 1994 than in 1996, 1998 or 2000. Thus, while the elimination of households who report no asset income alleviates some of the time series inconsistency, it fails to do so for the most relevant households, i.e. households with significant asset income. This is strong evidence against the hypothesis that the only survey induced bias is among households reporting no asset income. The existence of survey structure induced bias appears to be present both in households reporting positive asset income as well as in those reporting zero asset income.

3. Imputation Strategy

The average rates of return reported in Table 3c are not only evidence of the measurement error in asset income from Waves 1 and 2 of the HRS, but also suggest a possible solution to correcting the problem. As noted above, there is a high degree of consistency in financial wealth across all waves in the HRS. The time series consistency is a product of the fact that the survey instrument did not change over the years. Furthermore, the use of a follow-up sequence of unfolding bracket questions for respondents reporting 'don't know' or 'refuse' in the collection of asset and liability data, combined with random imputation within brackets, greatly minimizes any bias in the measurement of financial wealth. The result is that, while the time series consistency of financial asset income is clearly suspect, the reliability of measured financial wealth is strong.[1] It is thus possible to use the rates of return computed for the 1996 data to assign a rate of return to households in 1992 and 1994. These rates of return can then be combined with the financial wealth data for those households to impute a reliable measure of financial asset income.

To implement this strategy, a number of issues must first be resolved. The first issue involves specifying which households should be assigned a new rate of return. Throughout, we restrict attention to those households who report owning some financial assets. While this neglects households who may have owned financial assets at some point over the survey year but sold them prior to the survey date these cases are likely to be rare and we see no simple way of handling them. We consider two strategies for

[1] Note that unfolding brackets were implemented in the collection of asset income in all waves except Wave 1. This makes the reliability of asset income in Wave 1 even more suspect.

imputing financial asset income to households with positive financial assets. As suggested above, one strategy (A) would be to assign a new rate of return only to those who report no asset income. However, this does nothing about the survey induced bias for households who do report asset income. An alternative strategy (B) would be to impute a rate of return to all households including those that report asset income. This completely replaces the asset income from Waves 1 and 2 with imputed data. Strategies (A) and (B) represent two extremes. We present results from both.

The second issue is what rate of return to assign each household. The simplest imputation method is to assign the mean or median rate of return from Wave 3 households to households in Waves 1 and 2 using either strategy (A) or (B). However, this has at least one serious drawback. Assigning the same rate of return eliminates all heterogeneity in the rate of return. The average rate of return to financial wealth is a product of portfolio choice across different asset groups (equities, bonds, checking and saving) as well as the choice and performance of the chosen individual assets within each asset group. Assigning the mean rate of return neglects this important individual choice variation. A better approach is to assume that individual choice regarding portfolio selection remains relatively constant and to apply each individual household's financial asset income rate of return in Wave 3 to the financial assets held in Wave 1 and 2. This is problematic for households that have financial assets in Wave 1 or 2 but do not in Wave 3. To impute a rate of return to these households while still maintaining the empirical heterogeneity of the donor distribution, a rate of return is randomly drawn (with replacement) from the donor distribution for each household. This is the approach we take. For strategies (A) and (B), all households being imputed receive a randomly drawn rate of return. A third strategy (C) is to impute a rate of return to all households as in (B) but use the household's actual rate of return from Wave 3 if one is available and randomly impute if no Wave 3 rate of return is available.

The implementation of a random imputation procedure raises the issue of what donor distribution to use. The imputation procedure used to impute missing values for assets and liabilities relies on the donor distribution from the bracket in which the respondent claims their asset value resides.[2] No such information is available regarding the rate of return to financial assets. One approach is to use the entire rate of return distribution from Wave 3. However, this is problematic for several reasons. Foremost is the fact that along with actual rates of return, the zero's must be included in the donor distribution since households reporting zero asset income in Waves 1 and 2 are a large source of the bias that needs to be corrected. The probability of having zero asset income is larger for households with small amounts of financial wealth since this wealth is less likely to have large fractions of high yielding assets such as equities and bonds. Furthermore, the result of classical measurement error is greatly magnified for low financial wealth households since these values are in the denominator of the variable of interest, i.e. the average rate of return. Imputing a high rate of return to a large asset value would grossly overestimate the true asset income value. Finally, one could make a behavioral argument that households with higher levels of financial wealth are more likely to have portfolios dominated by equities and bonds, both of which have higher yields than checking and savings accounts. For these reasons and for the fact that reliable financial wealth data is observable in all waves of the data, donor distributions of the rate of return are computed for various financial wealth groups and applied to the same groups in the data to be imputed. The asset groups considered in this paper are the first

[2] The number of households refusing to not knowing the bracket information is surprisingly low. See Juster and Smith (1997) for more details.

three quartiles of financial wealth along with the 75th to 90th and 90th and above percentile groups.[3] As a result of the donor distributions varying by asset level, strategy (C) only uses the household's actual Wave 3 rate of return if its asset value in Wave 1 or 2 falls in the same asset group as Wave 3.

The fourth issue that needs to be considered is the treatment of outliers. Although our results rest on the assumption that the reported financial wealth from all waves and asset income values from Waves 3 and later are unbiased, classical measurement error is still a problem. These errors yield unrealistic rates of return in Wave 3 which could, in turn, get imputed to households in Waves 1 and 2. The standard treatment of outliers in empirical work is to trim. In the present case, this would entail dropping some values from the top of each financial asset group's donor distribution. However, by trimming the donor distribution, the result will yield yet another time series inconsistency since the donor data have not been trimmed in any such way. Since the goal is to achieve time series consistency, we make no attempt to treat outliers and thereby keep them in the donor distributions. An alternative which we also consider is to trim the donor distributions and apply them for the imputation of not only Waves 1 and 2 but also the dropped outliers of Waves 3 and later.

The final issue is robustness. As noted in the introduction, a crucial assumption for the validity of the imputation procedure is that the rate of return distribution, within financial asset groups, is the same over time. This may not be true for several reasons. First, there have been changes in the way in which certain assets pay out income. For instance, there has been a trend for equities to pay out less in dividends in favor of capital gains. This suggests a shift downward in the rate of return distribution. Second,

[3] Note that these donor groups require that households with no financial wealth in the donor wave be dropped since it is not possible to compute a rate of return.

households could be changing the way in which they allocate their financial wealth among assets. The increased household participation in financial markets over the past decade suggests a shift up in the rate of return distribution. Conversely, if this increase has been the result more of a shift from bonds to equities than from checking and savings accounts to either bonds or equities, then this would imply a shift downward in rates of return. Finally, the past decade has experienced tremendous growth. Although most of this has been reflected in large capital gains, returns to capital in all forms has increased suggesting higher rates of return. The net effect of these phenomena is ambiguous. While it seems most plausible to use the donor distribution from data collected nearest the collection date of the data requiring imputation, i.e. Wave 3 data, robustness is verified by applying the same imputations using donor data from Waves 4 and 5 of the HRS.

An outline of the imputation strategies and procedures considered in this paper are provided in Figure 1. We now turn to the results of implementing these procedures.

4. Imputation Results

The imputation procedures used in this paper rely heavily on the distribution of the rate of return in 1996. The central assumption is that the rate of return distributions for Wave 1 and 2 of the HRS are biased downward while the Wave 3 distribution, although not free of measurement error, has no such bias. The rate of return distribution for Waves 1, 2 and 3 are provided in Table 5a, 5b and 5c, respectively. These rates are computed only for households who have financial assets. However, it is important to note that there are many households who have a zero average rate of return.

The survey induced bias is clear by comparing the Wave 3 distribution with that of Wave1 and 2. The median rate of return for all households in 1996 is 2.4%. This value is zero for households in 1992 and 1994. Moving up the rate of return distribution, the bias remains. The average rate of return in 1996 is 5.7% at the 75th percentile while only being 3.3% and 1.5% in 1992 and 1994, respectively. Not surprisingly, average rates at a given percentile are smaller for lower values of financial assets. This is largely a result of the fact that the number of households with zero asset income increases. Households with small amounts of financial assets are more likely to have a portfolio that yields little to no asset income. For households in the lowest asset group, the median rate of return is zero in all years of the survey. Nevertheless, the pattern of the bias is consistent. The median rate of return for households with financial assets in the 50th to 75th percentile is 3.1% in 1996. In 1992 and 1994, this value remains at zero.

The importance of stratifying by financial assets is also made clear by Table 5. The distributions vary quite substantially by asset group within each year. As noted, this is largely influenced by households with zero asset income. This is the dominant effect in the distributions across financial asset levels up through the 75th rate of return percentile. However, by the 90th percentile of the average rate of return, classical measurement error in the denominator is seen to dominate. Financial assets are unlikely to yield estimates of income flows in the neighborhood of 25% or more, and the cases that fall into these categories are almost certainly ones in which there is a very small amount of assets combined with a moderate amount of income flow, resulting in an extremely high estimate of the rate of return. If one were to look at the details of the cases falling into the 25% or more rate of return category, one would find a great many cases where the average rate of return amounted to several hundred percent or even several thousand percent – cases where asset holdings were reported to be a small number like $10, and income flows reported to be a moderate amount like $500 or $1000. In general, errors that take the form of incorrect recording of the number of zeros are quite likely to result in extremely high rate of return estimates. In 1996, the 90th percentile of households in the lowest asset group is 20%. This is more than twice as large as the 90th percentile for households in the top asset group. The rate of return triples for the lowest group at the 95th percentile while only increasing by less than 50% for the highest asset group. Clearly, imputing a 60% rate of return to households with large levels of financial assets would lead to gross outliers in imputed asset income. These large differences in the empirical rate of return distribution across financial assets make it crucial that the random imputations stratify on financial assets.

The main results of this paper are found in Table 6. This table reports mean financial asset income by financial asset group using each of the three imputation strategies outlined in Figure 1. The un-imputed means are reported in the first row of each data year from Table 3a for the purposes of comparison. The imputation method for these values is labeled as ‘None’. Recall that Strategy (A) randomly imputes a rate of return only to households who report positive financial assets and zero income from those assets. The effect on the mean across all households is substantial. Financial asset income increases by 36% in 1992 from $1,876 to $2,543. The effect is even larger for the 1994 data. Imputation strategy (A) increases reported asset income in 1994 by 76% from $1,481 to $2,600. Not surprisingly, the largest gains from the imputation in both 1992 and 1994 go to those with the most financial asset wealth. However, the proportionate increase is roughly the same for households with financial assets above the 25th percentile – between 70 and 90%.

The third row of each data year in Table 6 reports the results of implementing Strategy (B). All households with positive financial assets are randomly imputed an average rate of return within financial asset groups. This argument for Strategy (B) over

Strategy (A) rests on the results from Table 4 which indicated a bias in the mean average rate of return time series even among households who reported some financial income. Given our priors that the survey induced bias acts to reduce reported asset income, it is not surprising that the implementation of Strategy (B) increases mean financial asset income from that of Strategy (A). However, the results are not that different for the mean across all households. Mean financial asset income is only increased an additional 3.5% in 1992 from $2,543 under Strategy (A) to $2,633 under Strategy (B). The 15% increase in 1996 is slightly larger.

The dominant effect of the imputations on the overall mean appears to be a result of imputing an average rate of return to households who report zero asset income. However, as in Table 4, the overall means mask large differences across the financial wealth distribution. The largest difference between Strategy (A) and (B) is evident for households with financial wealth above the 90th percentile. For both 1992 and 1994, mean imputed asset income for high wealth households is 31% larger under strategy (B). Clearly, the survey induced bias not only increases the number of households reporting zero asset income but also significantly reduces the amount of reported asset income. It is interesting to note however, that while the imputations under Strategy (B) increase mean income by 9% in 1994 over Strategy (A) for households with financial wealth in the 75th to 90th percentile, the procedure actually reduces the mean in 1992. Of course, both strategies increase the mean from the value with no imputations.

Strategies (A) and (B) reflect two extremes in the way measurement error enters reported asset income over the five waves of the HRS. While (A) assumes a reporting error only among households that report no asset income and leaves reports of positive asset income unchanged, (B) assumes reported asset income of all households is contaminated. While Strategy (C) is closer to (B) in that it imputes asset income for all households (with positive financial assets), it uses each particular households rate of return from 1996 under the assumption that this rate of return reflects the portfolio allocation behavior of the household. The household’s 1996 rate of return is used only if they have asset income in both 1996 and the imputation year and if the levels of financial wealth in both years are in the same asset group. Otherwise, the method of random imputation within asset groups is used. Within each asset group, roughly 50% of the cases under Strategy (C) utilize the households’ own 1996 rate of return to impute an asset income value in either 1992 or 1994.

The results from implementing Strategy (C) are reported in the fourth row of each data year in Table 6. Relative to the increase from the original data, there is little difference between any of the strategies in the overall means of imputed financial asset income. The imputation strategies increase the mean by roughly 35-50% in 1992 and by 75-100% in 1994. Within asset groups, Strategy (B) and (C) are more similar with each other than with Strategy (A). The higher the level of financial wealth, the more the results for Strategy (A) differ from Strategies (B) and (C). Given the presumed theoretical advantages of using the within household portfolio allocation information along with the relatively stability between Strategies (B) and (C), Strategy (C) is the imputation procedure of choice.

The goal of the exercise in this paper is to create time series consistency in the values of reported financial asset income. The biennial overall mean change in the original data over the eight-year period is –21%, 115%, 17% and 7.6%, respectively between 1992 and 2000. The seam problem between 1994 and 1996 is glaring. In addition, the large fall in asset income between 1992 and 1994 also seems anomalous.

Focusing on Strategy (C), the biennial overall mean change of the imputed data between 1992 and 2000 is 2.6%, 7.7%, 17% and 7.6%. This general upward trend is much more consistent with the upward trend in financial assets than is the original data.

The results presented in Table 6 rely on random imputations using the 1996 distribution as the donor distribution. To verify the robustness of these results, the same imputation strategies are re-done using either the 1998 data or 2000 data as the donor distribution. These results are found in Table 7. The table reports the percentage difference using the 1998 or 2000 donor distribution from the respective value using the 1996 donor distribution. Differences in the means across all households are small for strategy (A) and (B) but are on the order of 10 to 18% in 1992. The differences are trivial in 1994 for the overall mean. The differences become larger for lower asset levels. This is to be expected as the base values become smaller. Overall the imputation results appear quite robust to the donor distribution. Nevertheless, using the imputations based on the 1996 distribution seems most advisable since it is the year closest to the years being imputed.

Finally, it is over a broader macroeconomic interest to examine the time series of financial asset income net of the effects of outliers. Outliers are handled by dropping the top five percent of the donor distributions used in the random imputations. To maintain time series consistency, outliers that are trimmed also get imputed using the donor distribution from the respective year. The results of this exercise are reported in Table 8. Mean financial asset income across all households under Strategies (A) and (B) appear less consistent than under Strategy (C). The mean under Stategy (B) in 1992 is $2,080 and then increases by 2.3%, 42.9%, 10.3% and 2.4% biennially over the following eight years. The seam problem still seems apparent between the 1994 to 1996 survey years.

Using individual household rate of return information in Strategy (C), mean income in 1992 is $2,177 and then rises by 9.1% to $2,376 in 1994 and then by an additional 10.3% to $2,584 in 1996. There is much more heterogeneity across the financial wealth distribution but the conclusion seems to be same: the seam problem introduced by the new survey technology in 1996 is eliminated most effectively in the imputed and cleaned data under Strategy (C).

IV. Conclusion

In this paper we note the substantial effects of asking survey respondents about asset income in a merged asset/income module in which the income question sequences directly followed after the asset sequences rather than being asked in a separate income module. The inability of many surveys to ascertain accurate asset income data is certainly a product of this phenomenon. We go on to note that the improvements made by correcting this survey flaw do not come without a cost. This cost is a substantial seam problem between the years in which the survey technology is improved. In an attempt to improve cross-year consistency in the financial asset income series of the Health and Retirement Survey, we propose a number of imputation strategies that take advantage of the fact that cross-year consistency is maintained in the levels of financial assets.

Using various schemes to impute an average rate of return to households in 1992 and 1994, we are able to establish a time series of financial asset income with similar consistency to that of financial wealth. The strategy that yields the best results is one which combines a household's own portfolio allocation information from later waves of the data with random imputation of rates of return within various financial asset groups where the donor distributions come from the 1996 survey year. These results are notably robust to replacing the 1996 donor distribution with that of either the 1998 or 2000 survey years. A version of this imputation procedure that also accounts for gross outliers in the average rate of return yields a time series of financial asset income that is consistent with macroeconomic trends.

Future work will include correcting the income from privately owned business farms and real estate. Income from these two assets shares the same time series inconsistency as the financial asset income examined in this paper since it was also asked in a separate model from the value of the assets. The bias in business, farm and real estate asset returns is more difficult to correct since the rates of return are far more idiosyncratic than they are for financial wealth. Nonetheless, once these issues are adequately resolved, a superiod measure of total household income will be made available.

Until then, the results presented here should be a warning to surveyors that respondents provide far more accurate measures of financial asset income when preceded by questions regarding the assets which generate that income.

Reference List

Hurd, Michael F., Juster, Thomas, Smith, James P., "Ehancing the Quality of Data on Income: Recent Innovations from the HRS." *RAND Working Paper*, 2001.

Juster, Thomas F., Lupton, Joseph P., Smith, James P., Stafford, Frank P., "The Decline in Household Saving and the Wealth Effect." *University of Michigan Working Paper*, May 1999.

Juster, Thomas F., Smith, James P., "Improving the Quality of Economic Data: Lessons from HRS and AHEAD." *Journal of the American Statistical Association*, 1997, *92* (440), 1268-1278.

Table 1: HRS Financial Income and Asset Ownership Across Waves: Percent Reporting Income from Financial Assets

Financial	1992			1994			1996			1998			2000		
Assets	Yes	No	Total	Yes	No	Total	Yes	No	Total	Yes	No	Total	Yes	No	Total
Yes	35.7	45.4	81.1	29.2	53.2	82.2	62.3	21.0	83.3	63.0	21.3	84.3	60.1	25.1	85.2
No	0.2	18.8	20.0	0.2	17.4	17.6	0.1	16.7	16.7	0.0	15.8	15.8	0.1	14.8	14.8
Total	35.9	63.2	100.0	29.4	70.6	100.0	62.3	37.7	100.0	63.0	37.1	100.0	60.1	39.9	100.0
Observations	7359			6976			6736			6530			6220		

Table 2: Percent Reporting Zero Income from Assets by Asset Percentiles

Year	Percentile Group of Financial Assets					All
	[0-25%)	[25-50%)	[50-75%)	[75-90%)	[90-100%]	Households
1992	98.0	79.1	51.5	32.0	26.2	63.2
1994	97.7	81.9	61.1	47.6	36.3	70.6
1996	93.7	42.5	11.1	5.2	4.5	37.7
1998	92.8	43.7	10.8	5.8	1.3	37.1
2000	92.9	45.7	16.3	6.7	2.3	39.9

Table 3a: Mean Income from Financial Assets by Percentiles of Financial Asset Holdings (1996 Dollars)

Year	Financial Asset Percentile					All Households
	[0-25%)	[25-50%)	[50-75%)	[75-90%)	[90-100%]	
1992	25	360	1,081	2,882	8,776	1,876
1994	16	311	706	1,883	7,683	1,481
1996	11	143	1,070	4,680	18,451	3,190
1998	6	163	1,057	4,643	22,545	3,740
2000	31	284	1,015	4,889	23,307	4,024

Table 3b: Mean Financial Asset Holdings in Dollars by Percentiles of Financial Asset Holdings (1996 Dollars)

Year	Financial Asset Percentile					All Households
	[0-25%)	[25-50%)	[50-75%)	[75-90%)	[90-100%]	
1992	13	1,960	14,723	62,493	318,749	51,197
1994	34	2,793	19,047	71,070	369,886	60,887
1996	41	2,479	19,335	80,113	454,030	70,656
1998	30	2,190	18,909	85,009	589,991	88,957
2000	54	2,674	22,550	100,480	649,099	100,539

Table 3c: Mean Average Rate of Return by Percentiles of Financial Asset Holdings (Percent)

Year	Financial Asset Percentile					All Households
	[0-25%)	[25-50%)	[50-75%)	[75-90%)	[90-100%]	
1992	1.3	3.3	4.7	4.1	3.3	3.7
1994	0.6	2.3	2.7	2.4	2.7	2.3
1996	3.9	5.0	5.2	5.8	4.6	5.0
1998	3.0	5.0	5.0	4.6	4.5	4.6
2000	2.8	4.5	4.3	4.6	4.1	4.2

Note: Table 3c presents the mean of individual average rates of return, defined as the ratio of financial asset income to financial assets. This requires all households with no financial wealth to be dropped from the sample. In addition, ratios above one are trimmed in the calculation. This drops roughly one percent of the sample in each year with most coming from the first quartile (about 3% dropped in the first quartile).

Table 4: Mean Average Rate of Return by Percentiles of Financial Asset Holdings, Only Households with Positive Asset Income (Percent)

Year	Financial Asset Percentile					All Households
	[0-25%)	[25-50%)	[50-75%)	[75-90%)	[90-100%]	
1992	14.5	9.0	6.0	4.8	3.2	8.2
1994	11.3	6.0	4.4	5.1	3.3	6.4
1996	9.4	5.9	6.0	5.8	4.6	6.6
1998	9.6	5.9	5.2	4.8	4.6	6.3
2000	8.6	5.3	4.9	4.6	4.3	5.8

Note: Table 4 presents the mean of individual average rates of return, defined as the ratio of financial asset income to financial assets. This requires all households with no financial wealth to be dropped from the sample.

Table 5a: Distribution of Rate of Return to Financial Assets (Percent), HRS Wave I (1992)

Precentile	Financial Asset Percentile					All Households
	[0-25%)	[25-50%)	[50-75%)	[75-90%)	[90-100%]	
5th	0.0	0.0	0.0	0.0	0.0	0.0
10th	0.0	0.0	0.0	0.0	0.0	0.0
25th	0.0	0.0	0.0	0.0	0.0	0.0
50th	0.0	0.0	0.8	1.6	1.7	0.0
75th	0.0	1.7	4.7	5.0	4.5	3.3
90th	2.6	10.8	13.9	10.0	7.4	10.0
95th	44.4	35.7	25.0	16.3	10.0	23.1

Table 5b: Distribution of Rate of Return to Financial Assets (Percent), HRS Wave II (1994)

Precentile	Financial Asset Percentile					All Households
	[0-25%)	[25-50%)	[50-75%)	[75-90%)	[90-100%]	
5th	0.0	0.0	0.0	0.0	0.0	0.0
10th	0.0	0.0	0.0	0.0	0.0	0.0
25th	0.0	0.0	0.0	0.0	0.0	0.0
50th	0.0	0.0	0.0	0.4	0.8	0.0
75th	0.0	0.8	2.4	2.5	2.9	1.5
90th	0.0	8.0	7.1	5.7	5.7	5.8
95th	10.0	16.3	12.5	8.8	8.9	12.0

Table 5c: Distribution of Rate of Return to Financial Assets (Percent), HRS Wave III (1996)

Precentile	Financial Asset Percentile					All Households
	[0-25%)	[25-50%)	[50-75%)	[75-90%)	[90-100%]	
5th	0.0	0.0	0.0	0.0	0.1	0.0
10th	0.0	0.0	0.0	0.6	0.4	0.0
25th	0.0	0.1	1.0	2.0	1.8	0.0
50th	0.0	1.8	3.1	4.0	3.6	2.4
75th	3.0	5.3	6.2	6.2	6.0	5.7
90th	20.0	12.9	12.0	11.9	9.8	12.1
95th	60.0	21.8	17.0	16.9	14.1	21.5

Table 6: Mean Income from Financial Assets by Imputation Method (1996 dollars)

Data Year	Imputation Method	Financial Asset Percentile [0-25%)	[25-50%)	[50-75%)	[75-90%)	[90-100%]	All Households
1992	None	25	360	1,081	2,882	8,776	1,876
	(A)	11	565	1,446	4,011	11,675	2,543
	(B)	2	272	734	3,745	15,306	2,633
	(C)	19	202	958	4,443	18,901	2,886
1994	None	16	311	706	1,883	7,683	1,481
	(A)	11	524	1,293	3,622	12,960	2,600
	(B)	4	252	996	3,976	17,010	2,984
	(C)	6	240	993	3,693	17,256	2,961
1996	None	11	143	1,070	4,680	18,451	3,190
1998	None	6	163	1,057	4,643	22,545	3,740
2000	None	31	284	1,015	4,889	23,307	4,024

Table 7: Alternative Donor Distributions, Percent Difference from Results Using 1996 Donor Distribution

Data Year	Baseline Distribution	Imputation Method	Financial Asset Percentile [0-25%)	[25-50%)	[50-75%)	[75-90%)	[90-100%]	All Households
1992	1998	(A)	9.1	14.3	-1.0	7.3	2.3	5.3
		(B)	50.0	38.2	-4.1	25.5	-6.7	2.3
		(C)	5.3	-30.7	7.5	20.4	19.8	18.0
	2000	(A)	-81.8	-12.0	2.8	6.7	2.8	2.9
		(B)	-450.0	-6.6	5.7	14.5	-2.9	1.8
		(C)	-31.6	-260.9	5.5	19.1	15.5	10.8
1994	1998	(A)	18.2	16.4	2.9	5.9	-1.7	1.7
		(B)	25.0	31.0	2.4	16.8	-16.8	-6.4
		(C)	50.0	29.2	-2.4	9.7	-1.5	1.4
	2000	(A)	-336.4	-28.1	4.9	5.9	0.5	0.7
		(B)	-925.0	-60.3	11.0	12.6	-3.4	0.1
		(C)	-66.7	-94.6	13.2	1.5	-3.2	-2.6

Table 8: Mean Income from Financial Assets by Imputation Method, Imputed Outliers (1996 dollars)

Data Year	Imputation Method	Financial Asset Percentile [0-25%)	[25-50%)	[50-75%)	[75-90%)	[90-100%]	All Households
1992	(A), trim	6	135	802	3,114	10,855	1,811
	(B), trim	5	110	742	3,176	13,642	2,080
	(C), trim	4	112	755	3,273	14,419	2,177
1994	(A), trim	2	115	763	2,685	9,614	1,808
	(B), trim	2	94	742	3,215	11,681	2,128
	(C), trim	2	98	775	3,056	13,967	2,376
1996	trim	1	84	769	3,455	15,732	2,584
1998	trim	0	67	706	3,206	17,996	2,850
2000	trim	1	72	666	3,611	17,805	2,917

Figure 1: Imputation Strategies