Social Security Wealth, Inequality, and Lifecycle Saving

John Sabelhaus and Alice Henriques Volz

Appendix A. Benchmarking Aggregate SSW to OCACT Estimates

The SCF-based SSW estimates reported in the text have a published counterpart for years going back to the 2001 survey (Appendix Table 1). The OCACT benchmarks are a byproduct of disaggregating the "infinite horizon" unfunded liabilities of the Social Security system. The starting point is to compute the discounted present value of all future taxes and all future benefits, take the difference, then add the current Trust Fund balance to measure the overall aggregate shortfall in present value terms. The decomposition in tables (available only since 2001) allocates the PDVs across past, current, and future participants. The groups closest to the SCF population is current participants, which is the population 15 and older in the year for which the calculations are being made.

The SCF population we consider in this paper is a subset of the OCACT current participants group, because we only compute SSW for SCF respondents and spouse/partners, which means we are excluding dependents and other persons 15 and older (the "Non-Primary Economic Unit," or "NPEU" in SCF parlance) living in SCF households. Adult members of SCF households we miss include children still living with their parents, roommates, parents, or other older relatives living with respondents and spouse/partners. It may be feasible to study those individuals using the rudimentary SCF information that is collected, but for now, it is clear that we expect to calculate PDVs for taxes and benefits that are below OCACT, even if we have the earnings profiles right for respondents and spouse/partners.

In general, the comparison of our estimated PDVs against published OCACT values is reassuring, though the effect of alternative discounting and differences between the 62 and older and younger than 62 populations are notable and warrant further investigation. On the discounting front, there is little difference between simply using a 2.8 percent real discount rate and using the year-by-year OCACT discount factors through 2013. Between the 2013 and 2016 surveys, in addition to continued gradually lowering of assumed long-run real discount rates, OCACT moved to an alternative time path for closing the gap between the current (and persistently low) real discount rates and their long run values, which has the effect of dramatically increasing SSW relative to earlier years. The differences in PDVs between the retirement age and pre-retirement populations is also notable and warrants further investigation, especially the low PDV of estimated taxes for the 62 and older population.

Appendix Table 1. Survey of Consumer Finances (SCF) and Office of the Chief Actuary (OCACT) Present Discounted Values

	All Current Participants						
	2001	2004	2007	2010	2013	2016	2019
PDV Expected Social Security							
OCACT Published	$ 10,542	$ 12,552	$ 16,265	$ 19,735	$ 26,100	$ 31,400	$ 37,566
SCF, OCACT Discount Rates	$ 8,489	$ 10,327	$ 13,230	$ 14,891	$ 18,561	$ 27,575	$ 32,846
SCF, Fixed Real Rate=2.8%	$ 9,507	$ 11,529	$ 14,177	$ 16,180	$ 19,382	$ 21,589	$ 23,931
= PDV Expected Benefits							
OCACT Published	$ 23,200	$ 27,351	$ 34,257	$ 40,321	$ 51,600	$ 62,000	$ 72,721
SCF, OCACT Discount Rates	$ 16,333	$ 19,012	$ 23,238	$ 25,484	$ 30,273	$ 42,710	$ 50,484
SCF, Fixed Real Rate=2.8%	$ 17,614	$ 20,549	$ 24,407	$ 27,168	$ 31,139	$ 34,246	$ 38,254
- PDV Expected Taxes							
OCACT Published	$ 12,658	$ 14,799	$ 17,992	$ 20,586	$ 25,500	$ 30,600	$ 35,155
SCF, OCACT Discount Rates	$ 7,844	$ 8,685	$ 10,008	$ 10,593	$ 11,712	$ 15,135	$ 17,637
SCF, Fixed Real Rate=2.8%	$ 8,107	$ 9,021	$ 10,230	$ 10,988	$ 11,757	$ 12,657	$ 14,323
	Current Participants, Ages <62						
	2001	2004	2007	2010	2013	2016	2019
PDV Expected Social Security							
OCACT Published	$ 6,595	$ 8,030	$ 10,413	$ 12,311	$ 16,000	$ 19,100	$ 22,223
SCF, OCACT Discount Rates	$ 5,117	$ 6,610	$ 8,361	$ 8,815	$ 10,769	$ 16,170	$ 19,180
SCF, Fixed Real Rate=2.8%	$ 6,041	$ 7,688	$ 9,214	$ 9,900	$ 11,587	$ 11,964	$ 12,722
= PDV Expected Benefits							
OCACT Published	$ 18,944	$ 22,418	$ 27,928	$ 32,225	$ 40,600	$ 48,400	$ 55,826
SCF, OCACT Discount Rates	$ 12,926	$ 15,257	$ 18,290	$ 19,300	$ 22,360	$ 31,152	$ 36,596
SCF, Fixed Real Rate=2.8%	$ 14,113	$ 16,670	$ 19,365	$ 20,779	$ 23,224	$ 24,478	$ 26,839
- PDV Expected Taxes							
OCACT Published	$ 12,349	$ 14,388	$ 17,515	$ 19,914	$ 24,600	$ 29,300	$ 33,603
SCF, OCACT Discount Rates	$ 7,809	$ 8,647	$ 9,930	$ 10,485	$ 11,590	$ 14,982	$ 17,415
SCF, Fixed Real Rate=2.8%	$ 8,072	$ 8,982	$ 10,151	$ 10,878	$ 11,637	$ 12,514	$ 14,117
	Current Participants, Ages 62+						
	2001	2004	2007	2010	2013	2016	2019
PDV Expected Social Security							
OCACT Published	$ 3,947	$ 4,522	$ 5,852	$ 7,424	$ 10,100	$ 12,300	$ 15,343
SCF, OCACT Discount Rates	$ 3,372	$ 3,717	$ 4,869	$ 6,076	$ 7,792	$ 11,405	$ 13,666
SCF, Fixed Real Rate=2.8%	$ 3,466	$ 3,840	$ 4,963	$ 6,280	$ 7,795	$ 9,625	$ 11,209
= PDV Expected Benefits							
OCACT Published	$ 4,256	$ 4,933	$ 6,329	$ 8,096	$ 11,000	$ 13,600	$ 16,895
SCF, OCACT Discount Rates	$ 3,407	$ 3,755	$ 4,948	$ 6,185	$ 7,914	$ 11,558	$ 13,888
SCF, Fixed Real Rate=2.8%	$ 3,501	$ 3,879	$ 5,043	$ 6,389	$ 7,915	$ 9,768	$ 11,415
- PDV Expected Taxes							
OCACT Published	$ 309	$ 411	$ 477	$ 672	$ 900	$ 1,300	$ 1,552
SCF, OCACT Discount Rates	$ 35	$ 38	$ 79	$ 108	$ 122	$ 153	$ 222
SCF, Fixed Real Rate=2.8%	$ 35	$ 39	$ 79	$ 110	$ 120	$ 143	$ 206

Notes: OCACT values for 2013 and 2016 from Trustees Report appendix tables, prior years and 2019 are from various tables in the Financial Report of the US Government. SCF values based on "expected" SSW concept as described in text.

Appendix B. Details on Differential Mortality Correction

The computation of Social Security (and Defined Benefit pension) wealth requires conditional survival probabilities for every SCF respondent and spouse/partner from their current age through age 100. The starting point for individual mortality is the average mortality rate by gender, cohort, and age from the Social Security Administration Office of the Chief Actuary website: www.ssa.gov/OACT/HistEst/Death/2019/DeathProbabilities2019.html. The second step is a differential mortality adjustment based on Chetty et al (2016) that assigns relative mortality by income within each gender and cohort group.

The Chetty et al (2016) analysis of differential mortality uses linked income tax and Social Security death records for the period 2001 through 2014. The specific data file we use has number of deaths and populations by gender, income percentile (1 to 100), age (40 to 76), and year (2001 to 2014). The data are available at https://healthinequality.org/data/. There are various ways to process the differential mortality estimates to make them suitable for linking to a data file such as the SCF. We use a method that automatically preserves average mortality within gender, age, and birth cohort groups.

The first step in processing the Chetty et al (2016) mortality data is to compute death rates for each year, gender, age, and income percentile group. We then average over the fourteen years of data, which means we are not attempting to capture and project any trends over time in relative mortality. The third step is to compute overall average mortality for each gender and age group, and the fourth step involves dividing percentile group mortality by average mortality to compute *relative* mortality for each income percentile within the gender and age group. By working with relative mortality rates in this way, we know that the weighted average (across income percentiles) mortality within a given gender and age group will match the average mortality when we merge to a data set with mortality by gender and age across birth cohorts, so long as the matching data has 100 equally weighted income percentile groups. We also further smooth relative mortality in an additional step by regressing relative mortality on a cubic in income percentile. Those smoothed relative mortality rates are then multiplied by the Social Security cohort by gender and age mortality rates to provide absolute mortality. Values of relative mortality for ages between 76 and 100 are linearly interpolated to converge to one at age 100.

Appendix B. Improving SCF Lifecycle Earnings Imputations Using HRS-SSA Linked Administrative Earnings Records

The Survey of Consumer Finances (SCF) labor force and earnings modules have provided the key inputs needed to estimate lifecycle earnings and thus Social Security Wealth (SSW) in this research to date. Our ongoing research is focused on using Health and Retirement Study (HRS) data linked to Social Security Administration (SSA) earnings records to improve the lifecycle earnings imputations in the SCF. The approach involves creating comparable populations in the SCF and linked HRS along two dimensions. The first dimension is basic demographics, and the second is patterns of lifecycle labor force attachment within each of the demographic groups. After the data sets are reconciled, the last step is to explore various ways to use the lifecycle earnings details in the HRS-SSA data to impute year-by-year earnings for SCF respondents.

Linked HRS-SSA Administrative Earnings File

The SCF data from the current job, labor force history, and work expectations modules used to estimate lifecycle earnings is described above (Main text, Section 3). Before describing how we are using HRS-SSA linked data to improve the SCF lifecycle earnings estimates, we discuss how we are developing the various public-use and restricted data sets available from HRS. Our starting point is the public-use HRS Rand longitudinal file, which pools and links information across survey respondents for HRS interview years 1992 through 2016. The Rand file is the source of the key demographic linking variables.

We merge the HRS Rand file with processed restricted-use SSA administrative earnings data for 1951 through 2012 (using the HRS produced SSA "imputations" file) and then add in subsequent year by year records for those individuals from the detailed earnings file for years 2013 through 2016. The current HRS Rand file has something like 42,000 records overall, and roughly 24,000 records have links to (at least some) SSA earnings data. However, many of the 42,000 observations are out-of-scope for our purposes because they were born before our first analysis cohort (1930-39).

Changes in respondent permission requirements for linking SSA earnings records imply that some observations have incomplete earnings—the respondent agreed to have their earnings linked in one survey wave and then changed their mind. With that in mind, we created the concept of "complete earnings" to indicate that the HRS respondent has SSA earnings available through the minimum of (1) their current age, (2) age 61, or (3) their year of death. With that restriction, we have about 14,000 longitudinal observations with complete earnings within our analysis population. The fraction of observations with complete earnings is just over 50 percent in the 1930s and 1940s birth cohorts, then drops to around 40 percent in the 1950s and 1960-64 birth cohorts. There are roughly 5,000 usable observations in the 1930s cohort, but that drops to about 4,000 in the 1940s cohort, about 3,500 in the 1950s cohort, and about 1,700 in the 1960-64 birth cohort group.

Reconciling Demographic Types Across Data Sets

The choice of demographic variables for disaggregating and aligning the HRS-SSA and SCF populations is based on the tradeoff between variability in earnings levels and lifecycle profiles versus sample sizes across the demographic groups. The four key demographics we are currently working with include (1) birth cohort, (2) gender, (3) education, and (4) occupation. Our on-going work is focused on how to collapse/combine groups to efficiently preserve sample size in one or both data sets being reconciled.

As indicated in the main text, the focus of our output is on ten-year birth cohorts born 1930-39 and 1970-79, but that does not mean we have to reconcile the samples by those same birth years. The HRS-SSA linked sample is generally limited to birth years before 1965 (there are a handful of spouse/partners born in 1965 and later, but those are not useful for our purposes). One key overarching question is the extent to which earnings levels and profiles have shifted across birth cohorts, especially for sub-groups (men versus women, college educated versus others). Our current approach is focused on using 3 to 4 birth cohorts to disaggregate the samples at the birth cohort level. Within each cohort, the second (and binary) disaggregation is by gender. Given those two top-level choices, each of the disaggregated cohort by gender data sets is still well populated, averaging about 1,500 longitudinal observations.

Grouping by education and occupation are also key, because of well-known differences in earnings levels, lifecycle patterns, and variability across those groups (after controlling for birth cohort and gender). Our on-going work is focused on balancing lost information (from aggregating) against sampling variability (from creating groups that are too small). In terms of education, the key distinction is college educated versus all others. In terms of occupation, the SCF is somewhat limited to begin with (there are six occupational categories in the public-use data) and we also distinguish wage earners versus self-employed workers. Thus, we collapse occupation in the SCF and HRS into 5 comparable groups: (1) managerial and professional, (2) sales and office, (3) service, (4) construction, maintenance, production, transportation, and (5) self-employed.

In simplest terms, this means the average 1,500 observations in each birth cohort/gender group are further divided into 2 education and 5 occupation groups. On average, the longitudinal sample size is still fairly robust (150 per birth cohort/gender/education/occupation group on average) but differences in the share of the populations in each demographic group mean that some of the samples are much smaller. Our current focus is thus on disaggregating in ways that preserve as much of the heterogeneity across groups as possible, by identifying and collapsing along dimensions where there are no statistically discernible differences.

Within-Demographics Variation in Labor Force Attachment

Disaggregating the HRS-SSA and SCF samples by demographics (birth cohort, gender, education, and occupation) is the first step in reconciling the data sets. However, estimating individual lifecycle earnings involves controlling for the heterogeneity in lifecycle labor supply within those groups. The demographic splits tell us how much the average person of a particular demographic type earned at a particular age, but average earning includes individuals with a range of labor force attachments. This step can best be thought of as assigning work status for every individual at every age, where work status is (1) part time, (2) full time, or (3) not working.

The SCF labor force module meta data (see Section 3 in the main text) already provides what we need to identify the overall degree of labor force attachment for every individual. We know, for example, the number of years the individual has worked full-time and part-time. Although we do not know at exactly which years (meaning ages) the individual worked full-time, in most cases it is straight forward to make the inference. For example, most age 62 males will answer that they have worked full-time for something like 40-44 years, and the answers are generally consistent with other known lifecycle outcomes such as whether they report having gone to college. Some labor force participation patterns are more complicated to recover, but the various answers respondents provide about their current job or longest career job are generally sufficient to piece together a reasonable lifecycle pattern.

On the HRS-SSA side there is less meta data in one sense because the HRS only asks about the total number of years worked. Part-time and full-time years worked are not distinguished in the HRS retrospective question. However, there is much more data in another sense because we know the exact earnings at every age from the linked SSA data. We are currently focused on allocating HRS person-level labor force attachment into part- and full-time and reconciling those (within demographic groups) with the SCF meta data. The first step (within the HRS-SSA linked file) is to reconcile the survey-based number of years worked variable with the count of non-zero earnings years in the SSA records. The second step is to then reconcile total years worked with the sum of part- and full-time years in the SCF by demographic group. The third step is to use the SCF part-time/full-time split and the actual earnings values in the linked HRS-SSA data to reconcile lifecycle labor force attachment measures by age.

Imputing Annual Earnings Over the Lifecycle

Disaggregating the SCF and HRS-SSA samples by both demographics and patterns of lifecycle labor force attachment is the first step in estimating lifecycle earnings for SCF respondents. The second step is modeling within-group individual earnings heterogeneity conditional on the lifecycle labor force attachment measures. One aspect of earnings heterogeneity is how to control for differences between observed SCF earnings (in the year(s) for which we observe SCF earnings) relative to the average within the demographic group at the same point in the lifecycle. The earnings heterogeneity adjustment also involves adjusting for the type of job at the specific phase of the lifecycle, meaning pre-career, career, and post-career (or "bridge") jobs.

The SCF provides up to two distinct observations on individual earnings. If the individual is currently working, the SCF collects earnings details for their current and any second jobs. Regardless of whether the individual is working and has this "main" job information, they are also asked about their longest lifetime job (or second longest, if their current job is their longest). If that "long" job is longer than five years the respondent is asked detailed questions about that "long" job, including what they were earning when they left the job.

This combination of questions about "main" and "long" jobs gives us up to two annual earnings observations per individual. In our current SCF lifecycle earnings imputations, those two data points (combined with estimated earnings profiles by gender and education) are used to fill in earnings at every point if the lifecycle—both backward and forward relative to the respondent's current age and age at which they left their "long" job. Those imputations of course reflect the labor force attachment variables (not working, part-time, and full-time).

The same SCF earnings observations are the starting point when introducing heterogeneity using the linked HRS-SSA earnings data, but there are of course more options for exactly how to construct person-level year by year earnings imputations. The simplest approach is just to treat the differential between the SCF earnings and the HRS-SSA demographic group average earnings as a fixed effect and preserve that differential at all other points in the lifecycle. That approach is similar to what we are currently doing with the SCF data, with the added information that comes from comprehensive lifecycle earnings profiles estimated using HRS-SSA across demographic types. More advanced approaches will involve multiple imputations that capture realistic variability around the lifecycle averages, using a standard permanent versus transitory shock decomposition, where those moments are estimated using the HRS-SSA longitudinal data. Again, as with our current approach, the actual earnings imputation builds on the individual's labor force attachment by age already inferred from the SCF meta data.

Finally, the imputation involves adjusting for the type of job at the specific phase of the lifecycle, meaning pre-career, career, and post-career (or "bridge") jobs. Unadjusted earnings profiles have the well-known pattern of relatively rapid growth at young ages, followed by a gradual slowing of earnings growth, and eventually a decrease in earnings at older ages. These patterns vary by the demographic variables we are controlling for, but more importantly, the averages reflect moving from pre-career (often part-time) to career (usually full-time) jobs, then moving back to post-career jobs (again, often part-time). This distinction matters for our purposes because (for example) an individual who continues working in their career job at older ages generally does not experience an earnings decline and is thus accruing Social Security benefits at a different rate than a similar individual who has moved to a post-career job. Both the SCF and HRS-SSA data sets make it possible to identify these sorts of switches, and thus we can impute lifecycle earnings more efficiently for individuals based on where they are in terms of lifecycle jobs.