On-line Appendix for:

Measuring Income and Wealth at the Top Using Administrative and Survey Data[1]

Jesse Bricker
Alice Henriques
Jacob Krimmel
John Sabelhaus

Federal Reserve Board

April 1, 2016

Table of Contents

---

[1]This appendix accompanies the paper prepared for the Brookings Panel on Economic Activity, March 10th and 11th, 2016. Contacts: jesse.bricker@frb.gov, alice.m.henriques@frb.gov, jkrimmel@wharton.upenn.edu, john.sabelhaus@frb.gov. This paper was written while Jacob Krimmel was a research assistant at the Federal Reserve Board. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors.

A. SCF Sampling Strategy

The Survey of Consumer Finances (SCF) begins with a traditional household-level Area Probability (AP) sampling frame, and then supplements that sample using administrative data derived from tax records.[2] The administrative data is used primarily to develop the so-called "List" sample of wealthy families, which makes it possible to overcome the problems with thin samples and unit non-response among wealthy families that plague traditional random samples.[3] The List sample is drawn from tax records based on *predicted* wealth, where those predictions are based on income and other observable family characteristics. This section discusses in detail the methodology of the SCF List sample approach, including the strengths and weakness of various models of wealth prediction, and how these models can affect conclusions about measured wealth concentration.

Sampling Overview

The SCF List sampling strategy uses two methods of predicting wealth from income. The first is a gross-capitalization model, generated by inflating the tax unit's asset-based income by an asset-specific rate of return and adding a predicted housing value (Greenwood, 1983). The general form of the SCF model is:

$$\widehat{wealth}_i^{GC} = \widehat{house}_i + \sum_{\forall k} [\overline{Income}_i^k / r^k],$$

where there are i=1…N tax units, K types of income and $r_k$is the rate of return on the k-th type of income, and r^k is typically $\epsilon(0,1)$ and K=6.

The second model uses the empirical correlation between wealth collected in the SCF and income from the administrative sampling data. The basis for this "empirical correlation model" is a regression of observed SCF wealth from the most recent SCF on the administrative income used to generate the SCF List sample for that survey year. The most recent SCF is denoted here as T-3 and the base sampling income data are from two years prior to that:

$$\ln(SCF\ wealth_i^{T-3}) = \ln(\overline{Income}_i^{T-5})\beta + \varepsilon_i.$$

[2] The appendix to Bricker et al (2014) provides a high-level overview of the SCF sampling strategy, along with some summary statistics on participation in the Area Probability and List samples. This material is also covered in depth in the FEDS Working Paper version of this paper, Bricker et al (2015).
[3] See, for example, Sabelhaus et al (2015).

The matrix of sampling income, ($\overline{Income}_i^{T-5}$), includes many more types of income than the gross capitalization model, not necessarily based on a physical asset, allows rates of return to vary across different types of families, and permits the inclusion of some basic demographic data.[4]

The $\hat{\beta}$vector from this regression model is then applied to the current administrative sampling data to obtain a predicted wealth index:

$$\widehat{wealth}_i^{ECorr} = f(\overline{Income}_i; \hat{\beta}).$$

Both the empirical correlation and gross capitalization models use multiple years of administrative data in order to identify wealthy individuals, which helps to smooth over the effects of transitory income fluctuations and distinguish the permanently wealthy from those who happen to realize a very high, but transitory, income in a given year. These shocks are especially prevalent for capital incomes and at the top of the distribution.

Sampling Data

The SCF combines a standard nationally-representative area probability (AP) sample with a "List" sample derived from administrative data based on information from tax returns.[5] The List sample is drawn using statistical records derived from tax returns at the Statistics of Income (SOI) Division of the Internal Revenue Service.[6]

Since 1992, the Federal Reserve Board (FRB) has contracted the SCF field work to NORC at the University of Chicago and for more than thirty years the SCF has partnered with the Statistics of Income (SOI) Division of the Internal Revenue Service to select a "List" oversample of expectedly wealthy families. The process of selecting the List sample has evolved since the current SCF began in 1989, as more refined models for selecting wealthy respondents have been introduced, including moving from cross-section to panel-based administrative records in order to better control for transitory income fluctuations. The INSOLE data, maintained by SOI, are the main data for the List sample selection. Prior to use, the INSOLE

[4] As in the gross capitalization model, income is a weighted average of three years of sampling income. The variables in the empirical correlation model are selected by a stepwise model selection method.
[5] See O'Muircheartaigh et al. (2002) for more information about the NORC national sample.
[6] At the time the sample is drawn, the most recent complete administrative data are those from two years prior to the survey year. The sample includes individual and sole proprietorship tax filings from the IRS administrative tax (see Statistics of Income, 2012).

data are statistically edited by SOI to support policy work of Congressional and US Treasury staff (Statistics of Income, 2012).

The INSOLE file from the year prior to the survey (which describes the income from two years prior to the survey) are the main sampling data. Two years of panel data are attached to these records. Often the panel data are from the two previous years of INSOLE data, but sometimes they are from the IRS administrative tax data. For the 2013 SCF, the sampling data were anchored in 2011, but included 2010 and 2009 panel data on the 2011 INSOLE records.

The INSOLE data used for SCF sampling are anonymized and a great degree of security is involved with this sampling procedure. A formal contract governs the agreement between the FRB (who are responsible for selecting the List sample), SOI, and NORC. None of the three entities will ever know all of the sampling, contacting, and survey information. NORC needs to know the contacting information and collects the survey information but will never know the sampling information. SOI knows the contacting and sampling information but not the survey information. And the FRB knows the sampling and survey information but not the contacting information.

Sample Selection

A probability proportional to size (PPS) method is used to select the sample. PPS sampling can be described through the following example. A statistician wishes to select 100 families from a set of 1,000 families. The families are order from 1 to 1,000 and a sampling interval equal to 10 (=1000/100) is computed, which bins off the families into 100 bins of 10 families. Find a random number between 1 and 10; if the number is 6 then select the 6th family, the 16th family, the 26th family, etc… until 100 families are selected.

If each family has a sampling weight associated with it (as the INSOLE data do) then the example changes a bit. Assume that the first seven-hundred and fifty families have a weight of 1 and the next 249 have a weight of 10 and the final family has a weight of 60. Instead of 1,000 total families, the statistician actually picks from a weighted total of 3,400. The statistician still want to select 100 families, so the sampling interval is 34 (=3400/100) and there are 100 bins of 34 families. The families are ordered from highest weight to lowest then the family with weight of 100 is selected with certainty. Draw a random number between 1 and 34, say 31, then select the 31st family (which is the family with weight of 60), then the 62nd family, the 93rd family, etc… until 100 families are selected.

The List sample is also selected by a probability proportional to size (PPS) sampling method, stratifying by the seven wealth strata, sub-stratified by age and financial income, with the probability of selection increasing in each stratum.[7] In total, about 5,100 List sample cases are selected; the majority are from strata that capture the top one percent of expected wealth.

Wealthy families are much less likely to respond to a survey (Sabelhaus et al., 2015) and response rates in the List sample vary across strata in an expected manner. The response rate in the wealthiest SCF stratum is around 12 percent, increasing to about 25 percent in the second-wealthiest stratum, 30 percent in the third-wealthiest stratum, 40 percent in the fourth- and fifth-wealthiest and then about 50 percent in the two least-wealthy strata. These response rates are considerably lower than the roughly 70 percent response rate observed in the SCF AP sample.

Gross Capitalization Model

The data used to select the 2013 List sample were anchored in 2011 but included 2010 and 2009 panel data on the 2011 records. More weight is given to the income from the most recent tax year (as seen below). These data are read into two models which predict wealth from income. The exact form of the gross capitalization model in the SCF when selecting the 2013 SCF is:

$$\widehat{wealth}_i^{GC,T} = \frac{max(0,|taxable\ interest_i|)}{ror^{taxable\ interest}} + \frac{max(0,|non\ taxable\ interest_i|)}{ror^{non\ taxable\ interest}} + \frac{max(0,|dividends_i|)}{ror^{dividends}} + \frac{max(0,|rent\ \&\ royalties_i|)}{ror^{rent\&royalties}} + \frac{(|partnerships\ \&\ S\text{-}corps_i| + |estates\ \&\ trusts_i|)}{(ror^{dividends} + ror^{non\ taxable\ interest})/2} + \frac{(|schedule\ C\ gross\ income_i| + |gross\ farm\ income_i|)}{(ror^{dividends} + ror^{non\ taxable\ interest})/2} + net\ capital\ gains_i + \widehat{house}_i\ ,$$

where, there are where there are i=1…N tax units,

$$inc\ concept_i = \frac{1}{2} * inc\ concept_i^{2011} + \frac{3}{10} * inc\ concept_i^{2010} + \frac{2}{10} * inc\ concept_i^{2009},$$

[7] Within the seven strata there are nine financial income sub-strata and four age sub-strata. Sub-strata are arranged (head-to-tail) so that the PPS mechanism selects a good number of cases for each financial income and age bin.

and:

$$ror_i^{inc\ concept} = \frac{1}{2} * ror_i^{inc\ concept,2011} + \frac{3}{10} * ror_i^{inc\ concept,2010} + \frac{2}{10} * ror_i^{inc\ concept,2009},$$

for:

$$inc\ concept_i = taxable\ interest, non\ taxable\ interest, dividends, rent\ \&\ royalties, partnerships\ \&\ S - corps, estates\ \&\ trusts, schedule\ C\ gross\ income, gross\ farm\ income, net\ capital\ gains.$$

The rate of return on taxable interest is based on the Federal Reserve H.15 data series on the AAA corporate bond rate (seasoned issue, all industry). The rate of return on non-taxable interest is based on the H.15 data series on Moody's June rate on AAA state and local 20-year bonds. The rate of return on dividends is based on the S&P dividend price ratio, and the return on rent and royalties is based on the effective yield from a 30-year conventional mortgage from the H.15 data series. The rate of return on businesses, estates, trusts, and farms is estimated to be the mean of the rate of return of taxable interest and dividends. Capital gains are not adjusted.

Predicted home equity is based on finding the median house value within that tax unit's income range from the most recent SCF; the 2010 SCF data were used in selecting the 2013 List sample (Table A.1). Tax units are grouped into those with less than $60,000 in income (in $1989), between $60,000 and $120,000, between $120,000 and $250,000, between $250,000 and $1,000,000, between $1,000,000 and $5,000,000, and greater than $5 million in income.

Table A.1. Predicted home equity for gross-capitalization model

	Median value in 2010 SCF
Less than $60,000 in income ($1989)	$114,140
Between $60,000 and $120,000 in income ($1989)	$354,125
Between $120,000 and $250,000 in income ($1989)	$703,400
Between $250,000 and $1,000,000 in income ($1989)	$1,300,605
Between $1,000,000 and $5,000,000 in income ($1989)	$2,416,087
More than $5,000,000 in income ($1989)	$6,085,780

Empirical Correlation Model

The second model uses the empirical correlation between past SCF wealth and sampling data to predict a wealth ranking in the current sampling data. In selecting the 2013 List sample, the 2010 SCF wealth was linked to the sampling data for the 2010 SCF; these sampling data are the panelized version of the 2008 INSOLE file. A special dispensation granted by SOI allows this link for the purpose of selecting the List sample.

The sampling data contain many sources of income. The first step in the empirical correlation modelling process begins by finding the sampling variables that are most correlated with wealth. The sampling variables can describe income or certain deductions.

The process begins with a simple regression of logged SCF wealth on logged dollar values of sampling data and dummies for positive values of each income type; a stepwise selection process is used to determine which of these variables are most highly correlated with SCF wealth. In a stepwise selection criteria, the most variables most highly correlated with SCF wealth are sequentially added until all highly correlated variables are included; once a variable is added, the process also removes the variables that lose their correlation with wealth once the added variable is included in the model. The criterion for inclusion in the model is a p-value of 0.35. Some theoretically-relevant variables are added even if they are not selected in the stepwise selection process.

Thirty-three income variables in total are selected for the model, along with several geography dummies, marital and filing status, and age variables. These variables are included in a final first step model to find the correlation between SCF wealth and sampling data:

$$\ln(SCF\ wealth_i^{2010}) = \alpha + \beta_L^1 \ln\left(income_i^{1,2008-06}\right) + \beta_D^1\, \mathrm{I}\left(income_i^{1,2008-06} > 0\right) + \cdots + \beta_L^{33} \ln\left(income_i^{33,2008-06}\right) + \beta_D^{33}\, \mathrm{I}(income_i^{33.2008-06} > 0) + X_i^{2008-06}\delta + \varepsilon_i,$$

where $X = [geography, marital, filing, age]$,

and $\ln\left(income_i^{j,2008-06}\right) = \ln(\left|\frac{1}{2} * income_i^{j,2008} + \frac{3}{10} * income_i^{j,2007} + \frac{2}{10} * income_i^{j,2006}\right|)$,

for j=1…33

The $\hat{\alpha}, \widehat{\beta_L}, \widehat{\beta_D}, \hat{\delta}$ vector from this regression model is then applied to the current administrative sampling data (for which the same income variables are available) to get a predicted wealth index, which we denote here as the "empirical correlation" prediction:

$$\widehat{wealth}_i^{ECorr,2013} = \alpha + \hat{\beta}_L^1 \ln\left(income_i^{1,2011-09}\right) + \hat{\beta}_D^1 \mathrm{I}\left(income_i^{1,2011-09} > 0\right) + \cdots + \hat{\beta}_L^{33} \ln\left(income_i^{33,2011-09}\right) + \hat{\beta}_D^{33} \mathrm{I}\left(income_i^{33,2011-09} > 0\right) + X_i^{2011-09}\hat{\delta}.$$

Final rankings

The two predictions are blended together and used to rank the INSOLE families from highest to lowest expected wealth. In the 2013 selection process, the blend was:

$$blend_i^{2013} = \frac{1}{2}\left\{\frac{\widehat{wealth}_i^{ECorr,2013} - median(\widehat{wealth}_i^{ECorr,2013})}{IQR(\widehat{wealth}_i^{ECorr,2013})} + \frac{\widehat{wealth}_i^{GC,2013} - median(\widehat{wealth}_i^{GC,2013})}{IQR(\widehat{wealth}_i^{GC,2013})}\right\}.$$

The IQR( ) represents the interquartile range. In past years, the $blend_i$ weighted the empirical correlation model more than the gross capitalization model. The weight was even in the 2013 selection process. With the blended ranking, the sampling data are ordered from least wealthy to most wealthy. Seven wealth strata are created; the wealth of filers in the lowest stratum is often comparable to the AP sample while the top four strata fully cover the top one percent.

Families in the Forbes 400 and other families who finances are too unique for public data disclosure are removed from the sample.

Model Comparisons

Both the gross capitalization model and the empirical correlation model predict wealth from administrative income. Using the administrative SCF sampling data, it can be shown that the empirical correlation and gross capitalization approaches often disagree on predicted wealth rankings at the very top. The empirical correlation approach and using multiple years of data generate lower predicted top capital income shares, and thus by construction, lower predicted top wealth shares, than gross capitalization alone. Indeed, the differences in predicted top 0.1 percent shares for both wealth and capital income are larger than the residual gaps for the top 0.1 percent identified in the main body of the paper.

Using administrative income records to identify high *wealth* families requires strong assumptions both about the link between taxable income and wealth and about the distribution of wealth components that have no taxable income. To identify the wealthy from income tax data alone, one must rely on annual capital income to infer wealth through the gross-capitalization approach (Greenwood, 1983; Saez and Zucman, 2016).

However, only about half of assets can be inferred from a tax return.[8] When using tax return data, then, the value of these assets must be estimated and benchmarked to aggregate data. The most important "middle-class" assets, like housing and pensions, are typically not included on an income tax return. Saez and Zucman (2016) combine what information can be gathered from tax returns, for example, property tax deductions and current pension payments, with external data, like the SCF, to estimate these types of asset holdings for each tax unit.

Further, the annual *capital* income that is used to estimate wealth from the tax return also has permanent and transitory components. The variance and cyclicality of transitory income has also increased at the top end in recent years (Parker and Vissing-Jorgenson, 2010; Guvenen, Kaplan, and Song, 2014); capital income typically makes up a larger portion of these families' income.[9] An example of this increased variance is seen in the choice of many high-end families to receive capital income in the 2012 tax year in response to increased rates in 2013; predicting wealth using this one-year snapshot will overstate top wealth shares (Wolfers, 2015).

The wealth rankings of each model, using the SCF approach to gross capitalization, are compared here. About 89 percent of families that are predicted to be in the bottom 90 percent in the gross capitalization model are also predicted to be in the bottom 90 percent in the empirical

[8] See Saez and Zucman (2016) Appendix Table A3.
[9] Castaneda, Diaz-Gimenez, and Rios-Rull (2003) look at the dynamic relationship between income and wealth from a theoretical perspective, in the context of a lifecycle model calibration exercise.

correlation model (Table A.2). Looking at finer rankings within top 10 percent, though, there are considerable differences.

Table A.2. Impact of Ranking Top End Families by an Alternate Model

		Correlation Model Percentile				
				(Top 1)	(Top 0.1)	(Top 0.01)
		Bottom 90	90-99	99-99.9	99.9-99.99	99.99+
	Bottom 90	0.89	0.10	0.01	0.00	0.00
Gross-	90-99	0.20	0.48	0.28	0.04	0.00
capitalization	(Top 1) 99-99.9	0.05	0.22	0.48	0.23	0.02
Percentile	(Top 0.1) 99.9-99.99	0.03	0.10	0.31	0.46	0.10
	(Top 0.01) 99.99+	0.01	0.03	0.11	0.39	0.47

Notes: Rows sum to 1. Table describes where a family ranked in gross capitalization model would be ranked in the empirical correlation model. For example, in the last row, of families ranked in top 0.01 percentile in the gross capitalizations model, 1 percent of families are ranked in the bottom 90 percentiles by the correlation model, 3 percent are ranked between the 90-99th percentiles by the correlation model, 11 percent are ranked between the 99th-99.9th percentile by the correlation model, 39 percent are ranked between the 99.9th and 99.99th percentile by the correlation model, and 47 percent are ranked in the top 0.01 percent by the correlation model. Source: 2011 INSOLE data, supplemented with two years of INSOLE or IRS administrative tax panel data.

Within the top 10 percent, slightly less than half of records ranked by the gross-capitalizations model are ranked in the same percentile in the correlation model. Only 47 percent of those ranked in the top 0.01 percent in the gross-capitalizations model are also ranked in the top 0.01 percent in the correlation model. The agreement is at a similar level for the top 0.1 percent (but not in the top 0.01), the top 1 percent (but not the top 0.1), and the top 10 percent (but not the top 1 percent): only 46, 48 and 48 percent, respectively, of those ranked by the gross-capitalizations model are ranked similarly by the correlation model. And viewed another way, 41 percent of families ranked in the top 0.1 percent by the gross capitalizations model are not ranked in the top 0.1 by the correlation model.[10]

Often, the disagreement between the two models in terms of ranking is not large. Of the 53 percent of records ranked in the top 0.01 percent by the gross-capitalizations model that are *not* similarly ranked by the correlation model, 39 percentage points are in the top 0.1 percent (excluding the top 0.01 percent) and only 4 percentage points are out of the top 1 percent when ranked by the correlation model. These classification disagreements are at very fine levels. But often the case for using administrative data are that the sample size allows for the identification of these "top 0.01 percent" or "top 0.1 percent" families (Saez and Zucman, 2016). The results in Table A.2 indicate that such identification is not clear.

Model Fit

The predicted wealth rankings and wealth shares differ between the gross capitalizations and correlation models. The sampling process in the SCF uses both models, and then the survey collects wealth data on these families. The SCF sampling and survey data, therefore, allow a unique opportunity to assess the performance of both models by seeing how well the model predicted wealth correlates to survey-collected wealth. The correlation model performs better: wealth predictions from the correlation model are more strongly associated with SCF wealth and better predicts the SCF wealth ranking (Table A.3.).

Each model is assessed by regressing the natural log of SCF wealth on the predicted wealth level of each wealth index. This model can be run for all N SCF List sample respondents:

$$\ln(SCF\ wealth_i) = \alpha + \beta \ln\left(\widehat{wealth}_i^m\right) + \varepsilon_i \text{ for } m \epsilon \{GC, Corr\} \text{ and } i=1,\ldots,N.$$

[10] Further, the amount of wealth held by families in disagreement between the models is substantial. About 54 percent of the wealth of families ranked in the top 0.1 percent by the gross capitalizations model is held by families ranked below the top 0.1 percent by the regression model.

Table A.3. Correlation Between SCF Wealth and Predicted Gross-Capitalization and Empirical Correlation Wealth

	(1)	(2)	(3)
ln(GC model wealth)	0.85	…	0.26
	(0.02)	…	(0.02)
ln(Corr. model wealth)	…	1.02	0.76
	…	(0.01)	(0.03)
Constant	1.57	-0.46	-0.73
	(0.25)	(0.23)	(0.22)
R^2	0.69	0.78	0.80
Obs.	1,450	1,450	1,450
Predicted ln(wealth) at mean:	15.42	15.43	15.35

Notes: Regression of log of SCF family net worth in 2013 on log of predicted wealth of gross capitalization model (col. 1), correlation model (col. 2), and both (col. 3). Data from first implicate of SCF survey data matched to the wealth predictions that were used to stratify the List sample. Standard error in ().

The wealth predictions from each model can explain a large portion of the variation in SCF wealth, but the correlations model can explain more (row 4, Table A.3.). About 69 percent of the variation in SCF wealth is explained by variation in the gross-capitalization model, but about 78 percent is explained by the correlation model. After accounting for the correlation model, the gross capitalizations model adds very little explanatory power. When the gross capitalizations index is also included as an explanatory variable, the explained variation increases only slightly from 78 percent to 80 percent (column 3, Table A.3.). And much of the correlation observed in column 1 between gross capitalizations wealth and SCF wealth is absorbed by the correlation model in column 3.

The gross capitalizations model explained less variation in wealth in other survey years than it did in 2013. In other survey years the explained variation from the correlation model is unchanged when gross capitalizations is also included. Across the survey years, by itself, the gross capitalizations model is able to explain 59 to 69 percent of the variation in SCF wealth while the correlation model explains between 73 percent and 81 percent of the variation in SCF wealth; in each year the explained variation is about 15 percent higher than the gross capitalizations model.

The correlation model also rank-orders the tax units better than the gross capitalization model (Table A.4.). The Spearman correlation between the level of correlation and SCF net worth is usually around 0.90, and about 0.1 higher than the Spearman correlation found for gross capitalizations. The Pearson correlation coefficient describes the linear correlation between SCF net worth and each wealth index. Here, the story is more muted. The correlation coefficient of both indices tend to hover around 0.50, and both have hit a high value in the 0.73 to 0.77 range. The gross capitalization model, though, has a higher variance across years: the high in the 2013 SCF of 0.73 was preceded by a low of 0.14 in 2010.

Table A.4. Pearson and Spearman Correlations: SCF Wealth and Predicted Gross-Capitalization and Empirical Correlation Wealth

	Spearman correlations				
	2013	2010	2007	2004	2001
Gross-capitalization model	0.83	0.82	0.83	0.82	0.78
Empirical correlation model	0.90	0.91	0.90	0.91	0.87
	Pearson correlations				
	2013	2010	2007	2004	2001
Gross-capitalization model	0.73	0.14	0.39	0.46	0.53
Empirical correlation model	0.49	0.77	0.43	0.64	0.42

Notes: Data from first implicate of SCF survey data matched to wealth indices used to stratify the List sample.

Gain from Using Multiple Years

In selecting the List sample, the SCF uses three years of panel administrative income data to alleviate the effect of transitory income changes on the wealth rankings. The sampling data begins with the income records from two years prior to the SCF survey year (these are the most up to date income records possible), and a three year panel dataset is created using the two years prior to the initial sample. Alternate studies of identifying wealthy families use just one year of data (Saez and Zucman, 2016). In either model, about 90 percent of families identified in the top 0.01 percent using three years of data are also ranked in the top 0.01 when using one year of data (Table A.5.). Though these results generally point to consistency in wealth *rankings* across time, they also show that about 10 percent of the top 0.01 wealthy families (a very select group) in a given year are misclassified, presumably from a transitorily-high income shock.

Table A.5. Impact of Using Multiple Years of Data to Classify Families

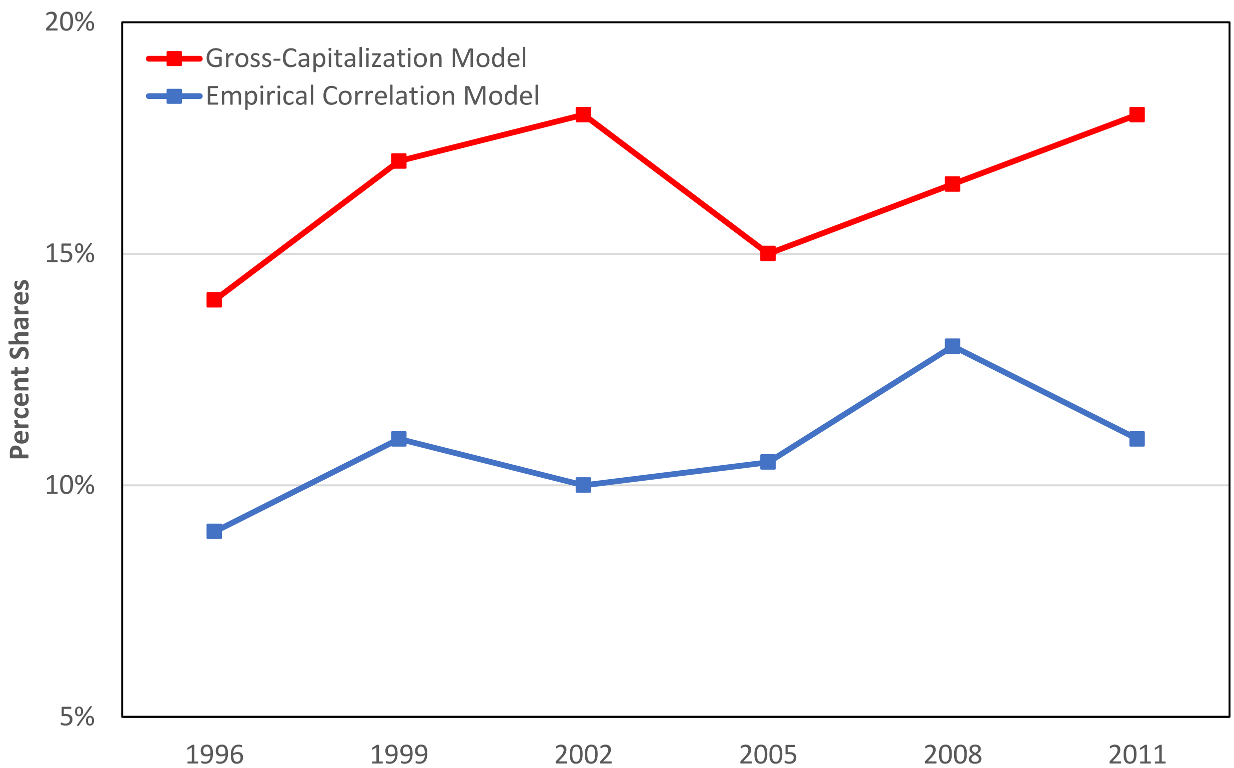
		2011-only gross capitalization model				
				(Top 1)	(Top 0.1)	(Top 0.01)
		Bottom 90	90-99	99-99.9	99.9-99.99	99.99+
2011-2009 gross-capitalization model	Bottom 90	0.98	0.02	0.00	0.00	0.00
	90-99	0.04	0.93	0.03	0.00	0.00
	(Top 1) 99-99.9	0.00	0.06	0.89	0.05	0.00
	(Top 0.1) 99.9-99.99	0.00	0.00	0.06	0.90	0.04
	(Top 0.01) 99.99+	0.00	0.00	0.01	0.05	0.94
		2011-only correlation model				
				(Top 1)	(Top 0.1)	(Top 0.01)
		Bottom 90	90-99	99-99.9	99.9-99.99	99.99+
2011-2009 correlation model	Bottom 90	0.97	0.03	0.00	0.00	0.00
	90-99	0.07	0.87	0.06	0.00	0.00
	(Top 1) 99-99.9	0.00	0.08	0.86	0.06	0.00
	(Top 0.1) 99.9-99.99	0.00	0.00	0.11	0.84	0.05
	(Top 0.01) 99.99+	0.00	0.00	0.00	0.14	0.86

Notes: Rows sum to 1. Tables show the impact of using 3 years of administrative data (2011, 2010, and 2009) versus 1 year of data (2011) to organize top end families and are organized similarly to table 1. Source: 2011 INSOLE data (supplemented with two years of INSOLE or IRS administrative tax panel data) compared to 2011 INSOLE data only.

__Predicted Top Shares by Model__

The gross capitalizations model also predicts larger wealth shares than does the correlations model and has shown larger growth in recent years (Figure A.1.). The top 0.1 percent in the SCF gross capitalizations model hold about 18 percent of predicted wealth, while the top 0.1 percent in the empirical correlation model hold about 11 percent of predicted wealth. The SCF top share falls between these two values. Over the most recent 6-year period, the top 0.1 share has grown by 20 percent in the gross capitalizations model, while the correlation model share has grown very little. In general, the levels and muted growth pattern shown in the empirical correlation model are consistent with the SCF levels and trends.

Figure A.1. Predicted Top 0.1 Percent Wealth Share from Gross-Capitalization and Empirical Correlation Model in SCF Sampling Exercise



B. Measurement Error and Gross Capitalization Factors

Our reconciliation of top wealth shares in the SCF with the estimates based on the gross capitalization approach in Saez and Zucman (2016, hereafter SZ) involved showing that the implied gross capitalization factors for interest bearing assets have become implausibly large in recent years, leading to upward bias in the share of wealth owned at the very top. In the paper we listed a few reasons why the estimated gross capitalization factors may be biased up. A complete decomposition of the bias in gross capitalization factors is beyond the scope of the paper, in large part because there is no comprehensive micro data (other than the SCF) which allows one to compute capitalization factors. The capitalization factors in SZ are derived as the ratio of two aggregate variables, household sector holdings of the given asset in the Financial Accounts (FA) divided by the Statistics of Income (SOI) measure of taxable income generated by the given asset In this appendix we look more closely at possible sources of measurement error in the numerator and denominator, and the implications for estimated gross capitalization factors.

Non-profits

Capitalization factors are computed using the ratio of FA assets to SOI income in the SZ approach. However, the FA assets include assets owned by non-profits as well as households. We first mention that a capitalization factor that includes financial assets held by families and non-profits in the numerator but financial income from families only in the denominator will be biased upward (as long as financial assets are positive):

$$CF = \frac{Assets_{families+non-profits}}{Income_{\ families}} > CF^{*} = \frac{Assets_{families}}{Income_{\ families}}.$$

The FA data can isolate the value of non-financial assets held by non-profits (real estate, equipment, and intellectual property), though isolating financial assets held by non-profits is not possible in the FA framework. However, the Center on Nonprofits and Philanthropy at the Urban Institute (McKeever, 2015, hereafter CNP) produces estimates of *total* assets held by the non-profit sector. In 2013, CNP estimated that nonprofits held about $5.17 trillion in total assets while the FA estimated that non-profits held about $3.10 trillion in non-financial assets. The net of these two estimates implies that nonprofits held about $2.07 trillion in financial assets in 2013. Total financial assets in the FA was about $65 trillion in 2013.

How much bias might these extra financial assets impart on estimates of capitalization factors? It is hard to say without an accounting of what types of assets are held by nonprofits, but results from a related exercise can help. The capitalization factors used in SZ are derived from FA data posted in early 2014 and prior to a major revision in FA bond assets that occurred later in 2014. For example, the FA data used in SZ contain about $1.5 trillion more household bond assets in the year 2013 than do the post-revision data. We estimate that the fixed-income capitalization factor is biased-up by about 10 percent just by including this extra $1.5 trillion. Thus, including an extra $2.07 trillion in non-profit financial assets could impart a similar bias. However, without an accounting of the portfolio of the $2.07 trillion, it is hard to say which assets classes are being biased upward or how much they are being biased upward.

Tax filing rules

Another potential bias for the SZ fixed-income asset capitalization factor comes from tax filing rules for some forms of financial income. A tax filer must claim all income received during a tax year and this is often done with the reminders: a mailed W-2 for wage income, or a mailed form 1099-INT for interest income. However, financial institutions are not required to mail to 1099-INT for an account that did not generate at least $10 of interest income, and these families may be less likely to report without the reminder and guidance of the 1099-INT. As interest rates have fallen in recent years to historically low levels, one can assume that a larger share of accounts generate less than $10 of interest income.

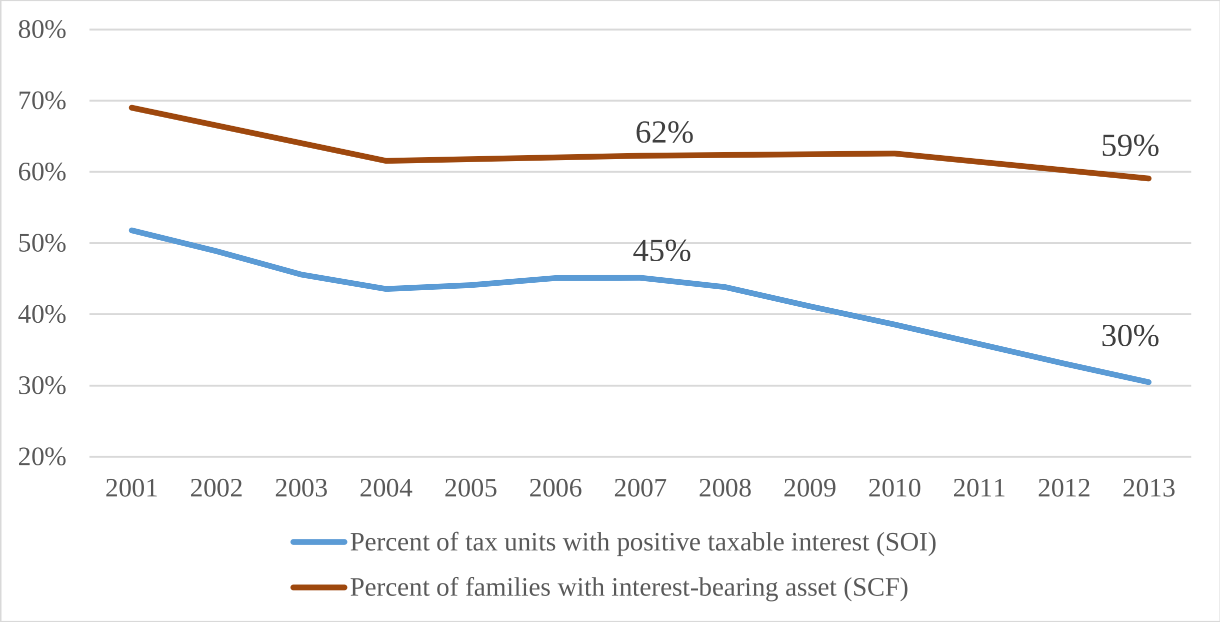
Thus, the amount of interest income reported to US tax authorities is presumably a lower bound on the true amount of interest income received by US families. At the same time, the FA estimate of fixed-income assets is unaffected by such reporting requirements meaning that the capitalization factor estimated by SZ is greater than the true factor (CF^*):

$$CF = \frac{Assets_{all\ families}}{Income_{\ some\ families}} > CF^* = \frac{Assets_{all\ families}}{Income_{\ all\ families}}.$$

Though the SZ estimated fixed income capitalization factor has become unmoored from other fixed-income capitalization factors over the past 15 years, it is really since 2007 that their factor experiences the most divergence from the other measures. This is also the year that taxable interest income reached its peak, and taxable interest income in 2013 has fallen to about 50% of its 2007 value.

There is substantial evidence to suggest that interest income tax filing rates have decreased more than asset fundamentals would suggest. First, the fraction of tax units that filed a return with positive taxable interest has fallen from 45% in 2007 to 30% in 2013. At the same time, the fraction of SCF families that reported having an account that produces interest income (savings and money market accounts, CDs, or bonds) stayed level at about 60% (Figure B.1.).

Figure B.1. Families with taxable interest and interest-bearing accounts, 2001-2013



Both the SCF and the FA aggregates show growth in interest-bearing accounts during this time, too, and both show roughly a 10 to 15% growth from 2007-2013 (Figure B.2.). Thus, it is likely that the share of SCF families with an account that produces interest income (Figure B.1.) is a good estimate. Further, these assets do not appear to be any more unequally distributed in the 2013 SCF than they were in the 2007 SCF (Figure B.3.).

Figure B.2. Growth in interest-bearing asset accounts, FA and SCF 2007-2013 (2007=1)

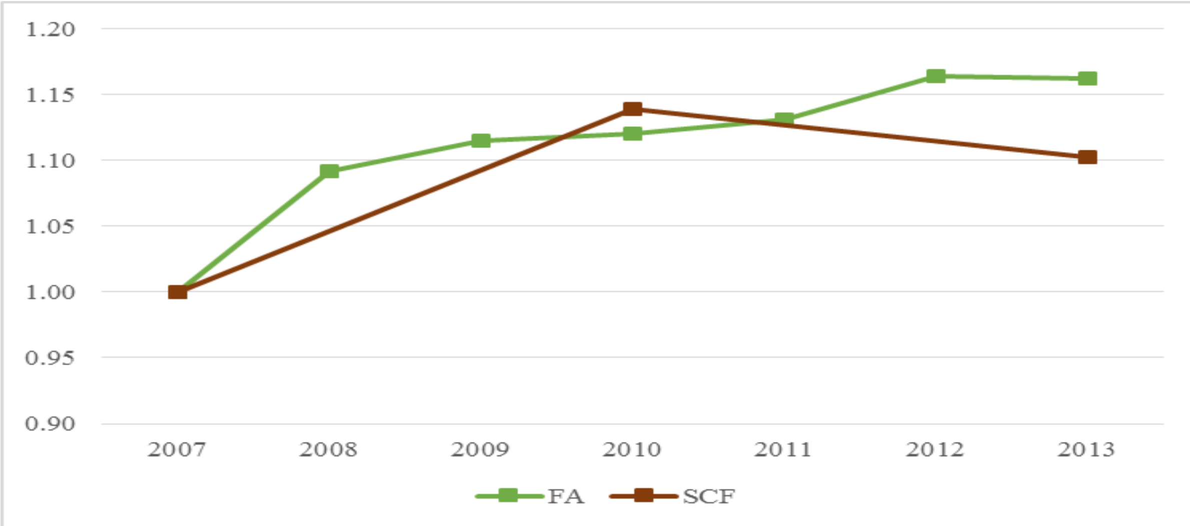
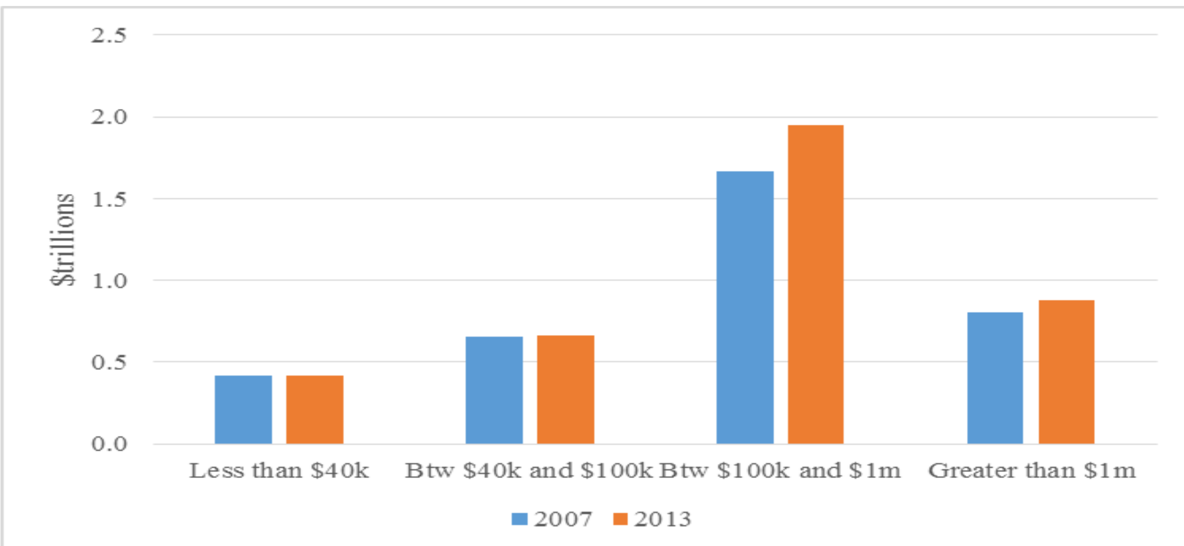


Figure B.3. Balances in interest-bearing accounts, by income level, 2007 and 2013 SCF

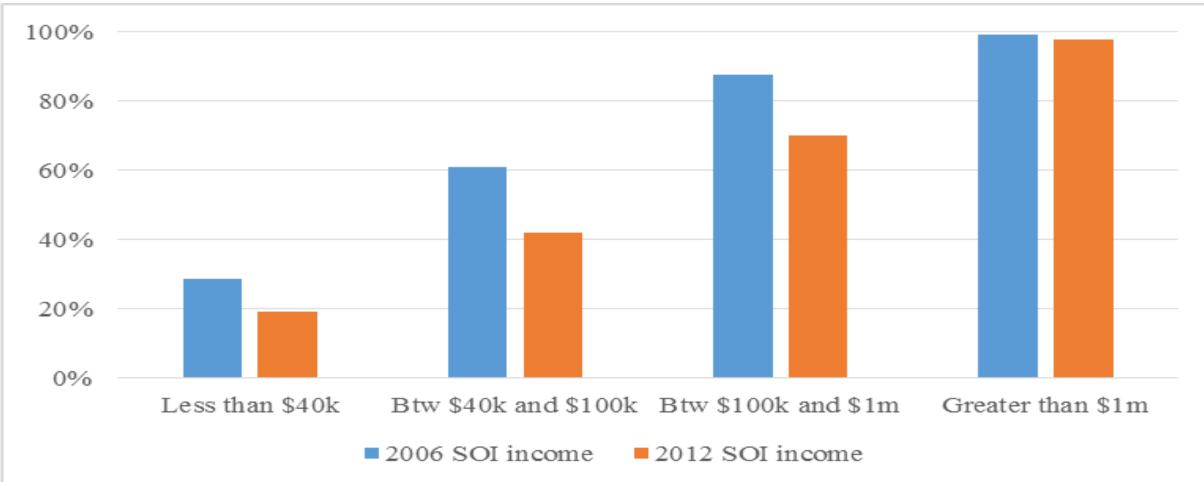


Though balances are steady or growing across income groups, the share of families in each group filing tax returns with interest income is falling (Figure B.4.), especially at the lower income levels.[11] Low-to-middle income families saw the greatest decline in tax filings with interest

[11] The decline in interest income in the SCF from 2007-2013 is 33%, similar to the SOI data which show a 49% decrease, not shown.

income reported (55-65% decline in the share filing taxable income from 2006 to 2012), even though their interest bearing assets declined no more than higher-income families (Figure B.3.).

Figure B.4. Fraction with interest income, 2006 and 2012 (SOI Table 1.4)



Aggregate taxable interest income also fell for fundamental reasons, namely because the returns on interest-bearing assets fell, on average, between 2007 and 2013 (Figure B.5). If we assume that the decline in aggregate taxable interest income among families with income of $1 million or more was due only to fundamentals (such as lower returns) while the decline in taxable interest income among families with lower income was due to fundamentals *plus being more likely to have interest income fall below the below $10 threshold* then we can estimate what the aggregate taxable income of lower income families would have been had they only been affected by fundamentals.

Figure B.5. Aggregate interest income, by AGI of tax filers, 2006 and 2012 (SOI Table 1.4)

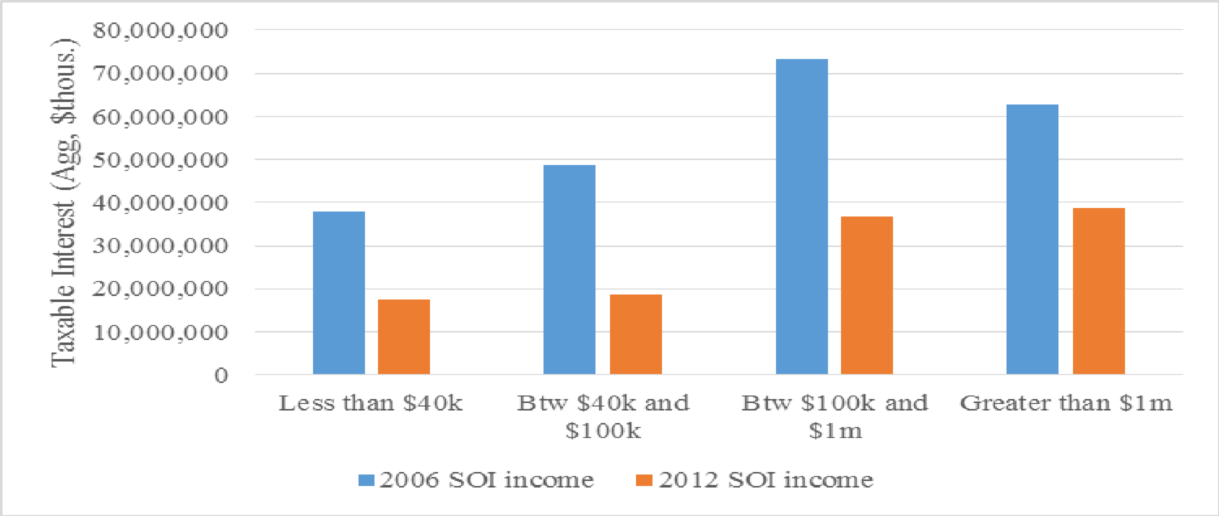


Table B.1. Aggregate taxable interest income (2006, 2012 SOI), predicted taxable interest income (2012), by AGI of tax filers (SOI Table 1.4)

	2006*	2012*	Decline	2012 (pred.)*
Less than $40k	38,011,674	17,589,395	54%	23,569,774
Btw $40k and $100k	48,689,979	18,680,075	62%	30,191,036
Btw $100k and $1m	73,392,308	36,695,605	50%	45,508,128
Greater than $1m	62,613,484	38,824,538	38%	38,824,538
Total	*222,707,445*	*111,789,613*	*50%*	*138,093,476*

* In $thousands. From SOI Table 1.4 2006 and 2012.

2012 prediction (final column) assumes that taxable interest income declined by 38% for all income groups, the same decline as observed by highest income group.

Families with greater than $1 million in income saw a 38% decline in taxable interest income between 2006 and 2012. Had families with income less than $1 million also seen a 38% decline in taxable interest income then total taxable interest income would be about 23% larger ($137 billion instead of $112 billion) than reported in 2012 (table B.1). This effect alone would bias up the capitalization factor in reported in SZ for fixed-income assets in 2012 by almost 25 percent.

C. Reconciling Income and Net Worth Concepts with Published Aggregates

Reconciling concepts of income and net worth is a key step in understanding differences in both aggregate values and distributional estimates. This appendix describes how the micro-level income concepts in the SCF and administrative data relate to aggregate Personal Income in the National Income and Product Accounts (NIPA), and how the micro wealth concepts relate to the household sector balance sheet estimates in the Financial Accounts (FA). This section is largely a summary of the reconciliations in Henriques and Hsu (2014) and Dettling et al (2015).

NIPA Incomes

Personal income is reported in Table 2.1 of the NIPA. In general, the NIPA income concept is a comprehensive measure of incomes received by households, except for capital gains. The published SCF total income concept (or "Bulletin" income) includes income from: wages and salaries; sole proprietorship and farms; other businesses or investments, net rent, trusts, and royalties; nontaxable bonds; interest and dividends; capital gains; unemployment insurance and worker's compensation; child support and alimony; Social Security and other pension income (including pension account withdrawals); government transfers such as TANF, SNAP, and SSI; and other miscellaneous income.

The key differences between published SCF and NIPA income are (1) SCF includes capital gains (variable x5712) while NIPA does not, (2) NIPA includes employer- and government-provided health insurance, while SCF does not, and (3) SCF captures retirement income only as it is being received, while NIPA captures the retirement income as it is being accrued.[12]

Since the SCF income concepts falls a bit short of NIPA in terms of population (missing the Forbes 400) and concepts (in-kind transfers), we modify the SCF measure to account for these shortcomings. We distribute the employer-provided health insurance and Medicare spending across all SCF families. This is accomplished using information collected on types of health insurance coverage each family has. For other in-kind transfers (e.g. SNAP and housing vouchers), we allocate the full NIPA total to the bottom 99% of the income distribution. To

[12] There are also small accounting differences in the treatment of business incomes, but we do not adjust for those.

estimate income for the Forbes 400, we use the median of the income to wealth ratio among the top strata in the SCF to estimate aggregate income flowing to these families.

The income measure in the administrative data (Piketty and Saez, 2003, updated) is a more narrow "market" income concept. In addition to the differences between SCF and NIPA, it also excludes government transfers and nontaxable interest. Thus, the equivalent NIPA "market" income concept shown in the text begins with Personal Income (Table 2.1, line 1), subtracts government social benefits to persons (line 17), and subtracts employer contributions for employee pension and insurance funds (line 7). The remaining adjustment for retirement income (employer contributions are already removed) is based on NIPA Table 7.20, which tracks contributions, interest and dividend earnings, and payments from retirement funds. The payments from retirement funds (except Social Security) are largely taxable, and therefore captured in the administrative tax data (and the SCF). The specific adjustment to NIPA personal income involves adding benefit payments and withdrawals (Table 7.20, line 27) and subtracting the income receipt on assets (line 11).

The SCF equivalent to market income begins with SCF total income, then subtracts nontaxable bonds (x5706), government transfer income (x5716, x5718, x5720), and retirement income specifically from Social Security (x5306, x5311). The administrative data market income also excludes business losses and minimizes capital losses at $3,000. As such, we also add back in losses to businesses and any capital losses greater than $3,000. The capital income concepts in the SCF and administrative data are conceptually equivalent: positive profits from sole proprietorships and farms (x5704); positive profits of other businesses, investments, rent, trusts, and royalties (x5714); taxable interest income (x5708); divided income (x5710); and capital gains (x5712) (with losses capped at $3,000).

Financial Accounts Net Worth

The Financial Accounts (FA, formerly known as the Flow of Funds Accounts) produces quarterly estimates of aggregate assets and liabilities held by the household sector, though the FA concept of net worth reported in table B.101 (Balance Sheet of Households and Nonprofit Organizations) diverges conceptually from the SCF in several ways. In creating an equivalent version of household net worth, we remove irreconcilable asset and liability categories from both FA and the SCF to put the two data sources on level footing.[13] Because FA includes non-profit institutions as part of the household sector, we first remove identifiable non-profit assets and liabilities.[14] This reduces published FA household net worth by $2.1 trillion in 2013 Q1 (Table B.1). Next, we remove from FA asset and liability categories involving security credit, which is not well-measured at the household level.[15] We also remove miscellaneous assets and liabilities from both FA and the SCF.[16] This adjustment reduces SCF aggregate net worth in 2013Q1 by just under $1 trillion and FA net worth by about $1.1 trillion.

Table C.1. Reconciling SCF and FA Aggregates

	2013Q1 ($ Trillions)		
	SCF	**FA**	**Difference**
Published Household Net Worth	65.5	72.3	-6.8
- Less Identifiable Nonprofit Net Worth		2.1	
- Less Security Credit, miscellaneous assets and liabilities	1.0	1.1	
- Less Life Insurance	0.8	1.2	
+ Plus DB Pensions	10.9		
- Less Durables	2.4	4.9	
- Less Forbes400 Net Worth		2.0	
= Conceptually Equivalent Net Worth	72.2	61.0	11.2

[13] The reconciliation relies on the FA release from September 18, 2014, as these are the values used by Saez and Zucman (2016).
[14] Table B.101 lines 5, 6, 7, 35, 38, and 40.
[15] B.101 line 26 and 39.
[16] B.101 lines 30, 36, and 37 and, from the SCF, bulletin variables OTHFIN, OTHNFIN, and ODEBT. We remove miscellaneous assets and liabilities for several reasons. First, there is potential misclassification between FA and the SCF. Second, miscellaneous assets and liabilities in the SCF includes money owed between households, which would net out in the FA aggregate household balance sheet.

We next remove life insurance assets and liabilities from both data sets because of conceptual differences between the SCF and FA.[17]

The SCF does not measure the value of defined benefit (DB) pensions, but collects information on current DB payments to retirees and workers currently enrolled in DB pension plans. This allows us to allocate DB wealth in FA across SCF households, so we add the value of DB pensions in FA, about $10.9 trillion in 2013 Q1, to SCF household net worth.[18] Next, we remove durables from both the SCF and FA because the SCF only captures the vehicles part of durables stocks, and it is not possible to separate vehicles from other durables in FA.[19] Finally, we subtract the wealth of the Forbes 400 list from FA aggregate household net worth because the SCF is explicitly forbidden from sampling any household identifiable by its high wealth. We arrive at conceptually equivalent aggregate net worth figures of about $72.2 trillion in the SCF and about $61.0 trillion in FA in 2013 Q1.[20] The remaining $11.2 trillion gap between SCF and FA is roughly twenty percent in 2013 (Figure 11).[21]

Grouping by Asset and Liability Categories

After reconciling total household net worth in the SCF and FA, we group the respective assets and liabilities into three comparable balance sheet categories: owner-occupied housing,

---

[17] We remove the net of B.101 line 27 less B.101 line 41 from FA and the variable CASHLI from the SCF. FA measures term life insurance reserves less deferred and unpaid life insurance, while SCF net worth includes the cash value of whole life insurance. Because the two are conceptually different, we remove all assets and liabilities related directly to life insurance plans.

[18] We estimate DB pensions as the portion of Total Pension Entitlements (B.101 line 28) not found in Defined Contribution pension assets (Table L.116 line 26) and annuities held in IRAs at life insurance companies (Table L.115 line 24), which are both captured in detail in the SCF. We explain how the residual, DB pensions, is allocated across households later in this section.

[19] The SCF captures only the value of vehicles, while FA includes all consumer durable goods according to the National Income and Product Accounts. Thus, we remove the variable VEHIC from SCF net worth and B.101 line 8 from FA net worth.

[20] SCF surveys are conducted throughout the year, and thus choosing any given quarter for benchmarking against FA aggregates is problematic, especially in periods of rapidly rising or falling asset prices. We benchmark to the first quarter FA levels in each survey year, because logic (and the data itself) indicates that survey answers are anchored to the pre-survey period for which the respondent has account statements and/or awareness of relevant market transactions for assets like housing. This is also consistent with the principle that SCF questions for items like household income, are deliberately focused on the calendar year preceding the survey.

[21] The SCF minus FA net worth gap was smaller in the 1989 to 1998 period, see Henriques and Hsu (2014).

non-housing assets, and liabilities.[22] These three broad classifications represent the most general categories that are conceptually comparable in the macro and micro data (Table C.2).

Table C.2. Reconciled SCF and FA Balance Sheet Categories			
	2013Q1 (Trillions)		
	SCF	**FA**	**Ratio**
Owner-Occupied Real Estate	$24.6	18.1	1.36
Total Non-Housing Assets	58.7	55.3	1.06
Liabilities	11.1	12.4	0.89
Net Worth	72.2	61.0	1.18

Disaggregating into further subcategories is problematic, especially when attempting to match specific types of assets in the SCF to their counterparts in FA. For example, SCF businesses show up in a number of FA sub-series, depending on how the respondent reports the business. Previous SCF-FA reconciliation projects show that this uncertainty and potential for cross-classification yield significant variation in the SCF-to-FA scaling ratios across more detailed asset subcategories (Henriques and Hsu, 2014). If the level of reconciliation is not conceptually consistent, adjusting the assets and liabilities of SCF households to match macro-level aggregates (as in a gross-capitalization benchmarking exercise) introduces a large variation in scaling ratios that re-shuffles the distribution of household wealth in undesirable ways. Limiting ourselves to three general categories minimizes asset misclassification and thus minimizes

[22] In the SCF, owner-occupied real estate includes the value of all primary residences, plus the value of secondary residences for which the household does not receive rental income. Non-Housing Assets includes: rental residential real estate, net equity in non-residential real estate, non-corporate business, transaction accounts and certificates of deposit, all assets in bonds and corporate equities, mutual funds and other managed assets, and retirement liquidity (including DB pensions). Liabilities in the SCF include debt secured by primary and other residences (including home equity loans), installment loans, credit card balances, and other lines of credit. In FA, owner-occupied real estate is given by line 4 of table B.101. Non-housing assets equals the sum of lines 10 (Deposits), 15 (Credit Market Instruments), 24 (Corporate Equities), 25 (Mutual Funds), 28 (Pension Entitlements) and 29 (Non-corporate Business) of B.101. We then reduce FA non-housing assets by the wealth of the Forbes 400, for whom we assume have negligible owner-occupied housing wealth and no liabilities. Thus, another strength of using a three-category balance sheet is that we make minimal assumptions when allocating Forbes 400 wealth. FA liabilities includes lines 33 (Mortgages) and 34 (Consumer Credit) of Table. B.101.

changes to the wealth distribution caused by scaling the SCF balance sheet to match the macro-level aggregates.

Benchmarking SCF Relative to FA Balance Sheet Categories

The impact on top wealth *shares* from benchmarking only arises because of *differentials* in the reconciled balance sheet categories. In 2013, SCF housing was 36 percent above the FA estimate, SCF non-housing assets are 6 percent above FA, and SCF liabilities 11 percent below FA. These differentials have both systematic and trend components that will affect levels and trends in top wealth shares in benchmarked relative to unadjusted survey-based estimates.

The impact of benchmarking SCF to FA on top wealth shares at any point in time depends on *where* in the wealth distribution one finds any given type of wealth. Housing is more middle-class wealth, while non-housing assets are concentrated at the top. Therefore, in 2013, lowering SCF housing values by 32 percent and non-housing assets by 7 percent will mechanically increase estimates of top wealth shares, because middle-class wealth is being pulled down by more than top wealth. As the gap between the SCF to FA ratios for housing and non-housing assets has widened in the past few surveys, the effect of benchmarking on raising top wealth shares has increased.

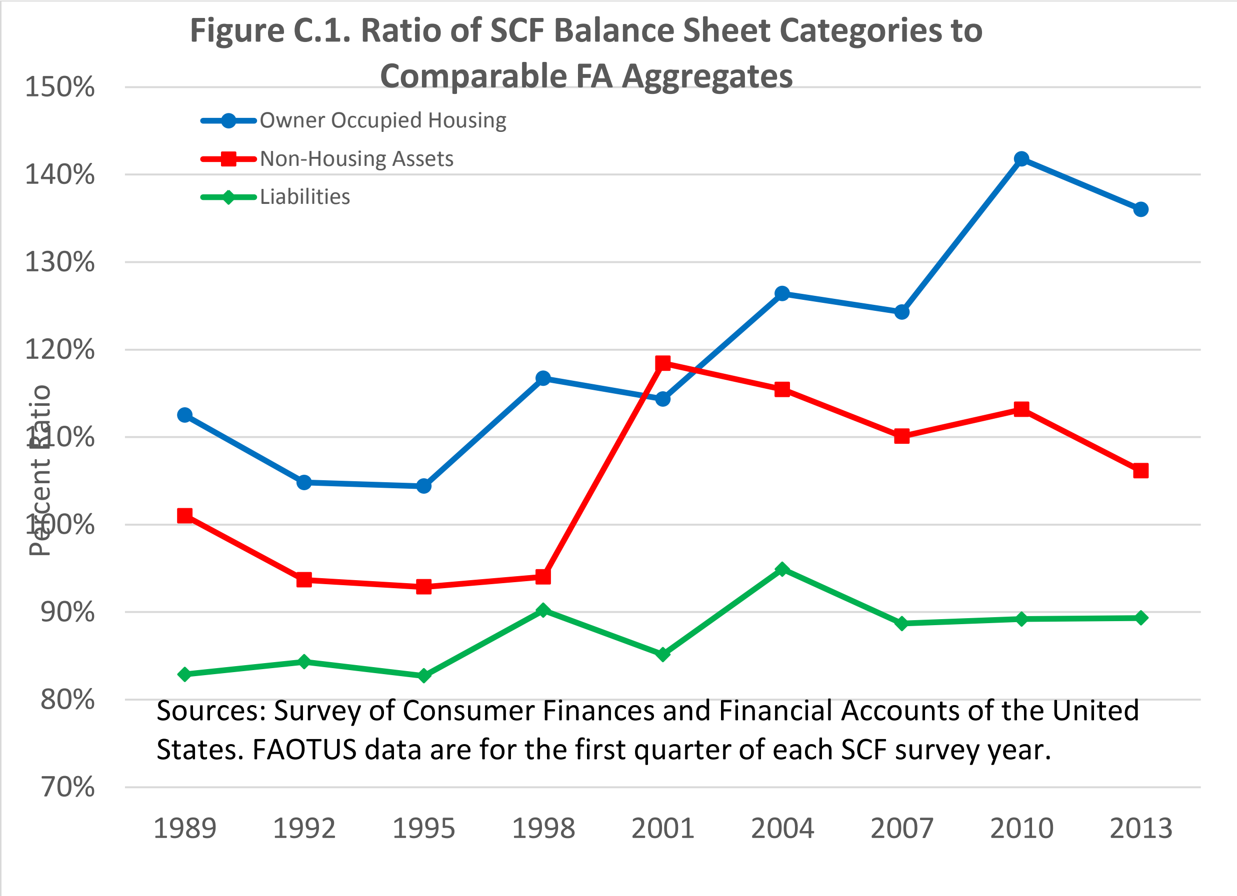


Figure C.1. Ratio of SCF Balance Sheet Categories to Comparable FA Aggregates
150%
140%
130%
120%
110%
100%
90%
80%
70%
Percent Ratio
Owner Occupied Housing
Non-Housing Assets
Liabilities
Sources: Survey of Consumer Finances and Financial Accounts of the United States. FAOTUS data are for the first quarter of each SCF survey year.
1989
1992
1995
1998
2001
2004
2007
2010
2013

D. Confidence Intervals for SCF Top Income and Wealth Shares

Several figures in the main paper show a confidence interval around SCF point estimates. This confidence interval is an estimate of both sampling and imputation variance that is present in SCF data. In addition to the descriptions below, there are a number of unpublished working papers on the SCF website that provide further details.[23]

Sampling variation

The SCF is based on a sample of families; sampling is used because taking a census of the outcome of interest is typically too costly. Because we will never observe the non-sampled families, the sampling process introduces "sampling error" to the survey estimates.

Sampling error can be estimated, and the SCF has typically produced a set of replicate weights to estimate sampling variability (Kennickell and Woodburn, 1999). The replicate weights are derived from resampling the SCF respondents along the dimensions of the SCF sample design; the resampling is done 999 times and weights are generated for each family in each resample. The final result is a set of 999 "bootstrap replicate weights" from which 999 SCF point estimates can be computed. The SCF sampling variation is estimated from these 999 estimates.

Imputation variance

Unit nonresponse occurs when a family decides not to respond to a survey. Section II considers the implications of unit nonresponse in the SCF. However, even when a family responds to the SCF, they are not required to answer all questions; *item* nonresponse describes this situation. Considering only the "completed cases" and ignoring the cases with item nonresponse will lead to selection bias, especially if families of certain types are more likely to have item nonresponse. The SCF uses a multiple imputation technique to impute data to the questions with item nonresponse; 5 "implicates" are imputed for each missing value. Multiple imputation is used in the SCF to acknowledge that any imputation model can only recover some distribution of the underlying missing data.

[23] http://www.federalreserve.gov/econresdata/scf/scf_workingpapers.htm

The full SCF data, then, is actually five datasets put together, each identified by their implicate number. Because imputed data vary across implicates, each dataset may arrive at a slightly different estimate. The variance across the five implicate datasets is called the imputation variance.

Confidence intervals

The confidence intervals shown in the paper describe an estimate of both the sampling and imputation variance of the SCF estimates. The combined standard error due to both sampling and imputation is described by the formula:

$$SE^{Overall} = \left(Var^{Sampling} + \frac{6}{5} * Var^{Imputation}\right)^{1/2}.$$

E. Distributing Aggregate DB Pension Assets

The SCF does not ask respondents about the present value of expected future defined benefit (DB) pensions, but the survey does collect information about current DB payments of retirees and about the expected future claims of workers currently enrolled in DB pension plans. Various papers have used the SCF to estimate household-level DB wealth for distributional and other purposes, and there are a number of methodological issues to be addressed in order to generate these distributional estimates using the data elements available in the survey.

The first decision involves micro aggregation versus using control totals for aggregate DB pension assets. In this paper, the aggregate value of DB assets by year is taken from the Federal Reserve Board's Financial Accounts (FA) of the United States.[24] DB pension wealth is the portion of Total Pension Entitlements (B.101 line 28) not found in Defined Contribution pension assets (Table L.116 line 26) and annuities held in IRAs at life insurance companies (Table L.115 line 24). In the first quarter of 2013, this amounted to $10.9 trillion, or roughly one-sixth of total FA household sector net worth.[25]

Aggregate DB wealth is distributed across households is a series of steps. We build off the approach used by Bricker, Henriques, Krimmel, and Sabelhaus (2015), which in turn was largely based on an approach introduced by Saez and Zucman (2016). The algorithm we use is still very rough, and does not make use of all of the available information in the SCF. However, that simplicity is also useful because it minimizes the number of behavioral assumptions one needs in order to implement the micro-level allocations.

The first phase of the micro allocation involves splitting aggregate pension wealth between SCF respondents already receiving benefits, and those who are or were covered by DB plans but not yet receiving benefits. We effectively assume that current beneficiaries have a first claim to plan assets, as we solve for the present value of promised benefits for those currently

[24] Financial Accounts data is available on the Federal Reserve Board's web site, in the quarterly Z1 release. The data can be accessed at http://www.federalreserve.gov/releases/z1/.

[25] Dettling et al (2015) show how total SCF net worth compares to the conceptually equivalent FA measures, but they do not discuss DB assets because there is no direct measure available in the SCF. One of the SCF values that lines up quite well with FA estimates is the total value of DC balances (including IRAs and other individually held tax-preferred assets). The fact that DC balances track FA assets very well means that we are not introducing any "calibration" distortion by using FA assets as the control total for DB while using aggregated survey values for DC.

receiving benefits, and subtract that amount from total plan assets to solve for the share to be distributed to those not yet receiving benefits. The present value of benefits for those already receiving is based on the respondent-reported values for those benefits, life tables from the Social Security Administration, and an assumed three percent real discount factor.

The number of SCF households currently receiving DB benefits increases across the 1989 to 2013 period (Table E.1., column 2) while the number of households with promised future benefits decreases (Table E.1., column 3). The first trend is clearly a function of demographics, as the aging of the Baby Boom and increase in life expectancy has led to systematically more DB recipients. The second trend reflects the shift from DB to DC, as there are fewer current workers in the queue to receive DB benefits after they retire.

The top-level allocation of assets between current and future beneficiaries is not as obvious, however, because the level of DB assets (Table E.1., column 1) has grown fast enough that the share of aggregate plan assets we assign to current beneficiaries is actually slightly lower now than it was at the beginning of the sample period (Table E.1., column 4). That is, if we assume current beneficiaries have first claim to plan assets, and measure those claims using observed benefits, life tables, and an assumed three percent real return, there is still a rising level of plan assets left over to be distributed among those who have not yet begun to receive benefits.

Some of the increase in aggregate DB plan assets may be attributable to changes in DB funding principles, but there is also (again) a key demographic component, and that underlies how we allocate the remaining DB assets among those not yet receiving benefits. The algorithm we use assigns each future recipient a share of the residual DB plan assets (the amount left over after current beneficiaries claim their share) based on their earnings and the number of years they have been in the plan (to reflect how DB plans generally work) and then discounts those claims relative to a typical benefit commencement age (we use age 60). The approach is meant to roughly capture how pension actuaries would compute the present value of the obligation. For example, given two observationally equivalent people (in terms of salary and number of years in plan) the actuaries would hold much more in assets for (say) a 60 year old than they would for a 40 year old. Indeed, using the same three percent discount rate, those differences in asset holdings are quite large. And, when one acknowledges that the age distribution of those who are expecting but not yet receiving benefits has shifted towards retirement as Baby Boomers have aged and new labor force entrants are less likely to be covered by DB plans, it is clear why (even without a change in funding principles) DB plans are holding much more in assets per future recipient than they did in the past.

The algorithm we use for distributing DB assets among those not yet receiving benefits is not based on SCF respondent-reported expected DB benefits. A more elaborate approach involves a sequence of assumptions about workers' continued participation in their current plans, retirement/claim ages, and life expectancy that one needs to make, in order to bring to bear all of the relevant information in the SCF. In addition to the behavioral assumptions, one also needs to assume that workers have a good understanding of their plan parameters. Future work should focus on sorting this out, and ideally, one would construct micro-level expected DB benefits that (appropriately discounted) track well with aggregate plan assets in the FA.

Table E.1. Total DB Wealth, Total Number of Households Receiving DB Wealth, and Share of DB Wealth by Source 1989-2013.

Year	Total DB Wealth ($Billions)	Number of HHs Currently Receiving DB Benefits (Thousands)	Number of Households with Future DB Claims (Thousands)	Share of Total DB Wealth Allocated to Current DB Recipients	Share of Total DB Wealth Allocated to those with Future DB Claims
1989	2,733	15,366	24,873	67%	33%
1992	3,419	14,772	23,126	61%	39%
1995	4,174	15,978	19,034	64%	36%
1998	4,895	15,561	18,221	56%	44%
2001	5,841	15,390	18,283	50%	50%
2004	6,931	17,342	18,419	58%	42%
2007	8,317	17,727	16,214	50%	50%
2010	9,529	18,412	17,518	53%	47%
2013	10,981	21,047	16,657	59%	41%

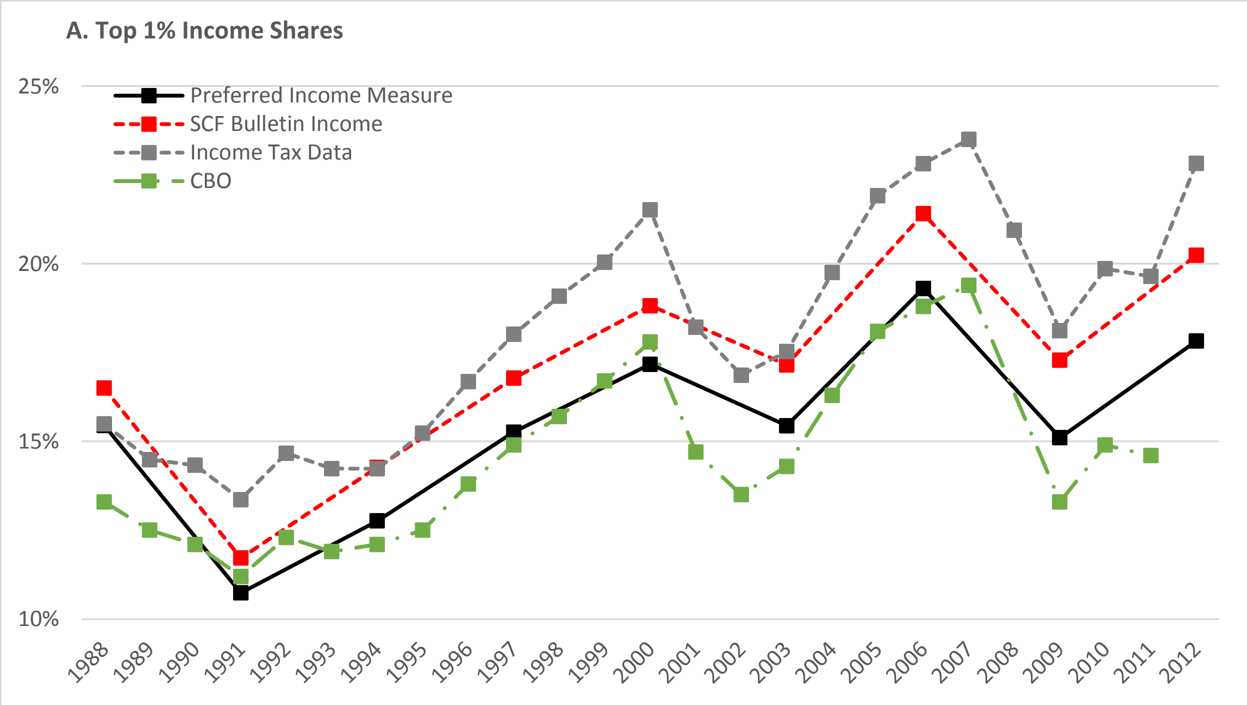
F. Comparison with CBO Income Top Shares

The Congressional Budget Office (CBO) produces an income series using income from tax filers from SOI and statistically matching to survey respondents to the Current Population Survey's (CPS) Annual Social and Economic Supplement (ASEC) (CBO 2010, 2014). By merging the two data sources, CBO combines a data source with good coverage of income at the top (SOI) with one that has good coverage at the bottom (CPS).

The CBO income series includes market income (labor income, business income, capital gains, capital income, income received in retirement for past services, other income, and government transfers (cash payments and in-kind benefits from social insurance and other government assistance programs).

The preferred measure of income used in this paper (plotted in black in Figure F.1) is comparable in level and trend to the CBO income series.

Figure F.1 Top Income Shares, Including CBO



Note: CBO series ends in 2011. This figure is a replica of Figure 2, though augmented to include the CBO series.

G. References

Board of Governors of the Federal Reserve System. 2015. "Financial Accounts of the United States". Retrieved from http://www.federalreserve.gov/releases/z1/current/z1.pdf.

Bricker, Jesse, Lisa J. Dettling, Alice Henriques, Joanne W. Hsu, Kevin B. Moore, John Sabelhaus, Jeffrey Thompson, and Richard A .Windle. 2014. "Changes in U.S. Family Finances from 2010 to 2013: Evidence from the Survey of Consumer Finances," *Federal Reserve Bulletin*, 100(4): 1-40. (September)

Bricker, Jesse, Alice Henriques, Jacob Krimmel, and John Sabelhaus. 2015. "Measuring Income and Wealth at the Top Using Administrative and Survey Data," Federal Reserve Board, *Finance and Economics Discussion Series*, 2015-30. (April)

Castaneda, Ana, Javier Diaz-Gimenez, and Jose-Victor Rios-Rull. 2003. "Accounting for the U.S. Earnings and Wealth Inequality." *Journal of Political Economy,* 111(4): 818-857.

Congressional Budget Office. 2010. *Average Federal Tax Rates in 2007.* Washington, DC: Congress of the United States, Congressional Budget Office

Congressional Budget Office. 2014. *The Distribution of Household Income and Federal Taxes, 2011.* Washington, DC: Congress of the United States, Congressional Budget Office

Dettling, Lisa J., Sebastian Devlin-Foltz, Jacob Krimmel, Sarah J. Pack, and Jeffrey P. Thompson. 2015. "Comparing Micro and Macro Sources for Household Accounts in the United States: Evidence from the Survey of Consumer Finances," Federal Reserve Board, *Finance and Economics Discussion Series*, 2015-86. (June)

Greenwood, Daphne. 1983. "An Estimation of US Family Wealth and its Distribution from Microdata" *Review of Income and Wealth* (March) pp 23-44.

Guvenen, Fatih, Greg Kaplan, and Jae Song. 2014. "How Risky are Recessions for Top Earners," *American Economic Review*, 104(5): 148–153. (May)

Henriques, Alice M. (2013). "Are Homeowners in Denial about their House Values? Comparing Owner Perceptions with Transaction-Based Indexes," Federal Reserve Board, *Finance and Economics Discussion Series*, 2013-79.

Henriques, Alice M., and Joanne W. Hsu. 2014. "Analysis of Wealth using Micro and Macro Data: A Comparison of the Survey of Consumer Finances and Flow of Funds Accounts," in Jorgenson, Dale W., J. S. Landefeld and Paul Schreyer eds., *Measuring Economic Sustainability and Progress*, Studies in Income and Wealth, Volume 72. Cambridge, MA: National Bureau of Economic Research.

Kennickell, Arthur B., and R. Louise Woodburn. 1999. "Consistent Weight Design for the 1989, 1992 and 1995 SCFs, and the Distribution of Wealth," *Review of Income and Wealth,* 45(2): 193-215.

McKeever, Brice. 2015. “The Nonprofit Sector in Brief in 2015: Public Charities, Giving, and Volunteering.” Washington DC: Urban Institute.

O'Muircheartaigh, Colm, Stephanie Eckman, and Charlene Weiss. 2002. “Traditional and Enhanced Field Listing for Probability Sampling,” *American Association for Public Research 2002: Strengthening Our Community - Social Statistics Section*, 2563-2567.

Parker, Jonathan A., and Annette Vissing-Jorgensen. 2010. "The Increase in Income Cyclicality of High-Income Households and Its Relation to the Rise in Top Income Shares." *Brookings Papers on Economic Activity*, 1-55. (Fall)

Sabelhaus, John, David Johnson, Stephen Ash, Thesia Garner, John Greenlees, Steve Henderson, and David Swanson. 2015. “Is the Consumer Expenditure Survey Representative by Income?” Forthcoming in *Improving the Measurement of Consumer Expenditures*, Christopher D. Carroll, Thomas Crossley, and John Sabelhaus (Eds.), Studies in Income and Wealth, Volume 74. Cambridge, MA: National Bureau of Economic Research.

Saez, Emmanuel, and Gabriel Zucman. 2016. “Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data,” Forthcoming, *Quarterly Journal of Economics*. Accessible version: National Bureau of Economic Research Working Paper 20625. (October 2014)

Statistics of Income. 2012. *Individual Income Tax Returns.* Washington, DC: Internal Revenue Service.

Wolfers, Justin. 2015. “The Gains from the Economic Recovery are Still Limited to the Top One Percent” http://www.nytimes.com/2015/01/28/upshot/gains-from-economic-recovery-still-limited-to-top-one-percent.html.