How much has wealth concentration grown in the United States? A re-examination of data from 2001-2013 [1]

Jesse Bricker[a], Alice Henriques[a], Peter Hansen[a]

[a] Federal Reserve Board
jesse.bricker@frb.gov
alice.m.henriques@frb.gov
peter.hansen@frb.gov

Wealth concentration can be estimated from several imperfect sources. After accounting for these imperfections, we show that there is little to no difference in the recent growth of wealth concentration in the United States (US). This stands in contrast to prior research showing that wealth concentration estimates modeled from US income tax data are growing more rapidly than in both US estate tax data and in the Survey of Consumer Finances (SCF). The model that serves as the basis for this research "capitalizes" taxable income, and even minor permutations to the model are enough to alter the conclusions about rapid growth. For example, using the 2011 income tax data we estimate the wealthiest top 1 percent hold 40.3% of wealth when using a baseline model, 38% when a heterogeneous rate of return on fixed income assets is applied to the top 1 percent by *income*, but only 33.7% when the same heterogeneous rate of return on fixed income assets is applied to the *wealthiest* top 1 percent. Far from rapid recent growth, this model shows no growth in wealth concentration since 2008. The model permutations are consistent with several auxiliary data sources on heterogeneous rates of return across the wealth distribution.

[1]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. We would like to thank our colleagues on the SCF project who helped make this research possible: Lisa Dettling, Sebastian Devlin-Foltz, Joanne Hsu, Lindsay Jacobs, Elizabeth Llanes, Kevin Moore, Sarah Pack, John Sabelhaus, Jeff Thompson, and Richard Windle. We would like to thank helpful discussions with Tom Crossley, Steven Pedlow, Katharine Abraham and participants at the 2016 meeting of the Society of Economic Measurement, the 2016 NBER CRIW Summer Institute, and the 2017 ESRA conference. We also thank Arthur Kennickell for inherited knowledge of the SCF sampling process, and have greatly benefited from coordination with the staff at the Statistics of Income, especially Barry Johnson, David Paris, Michael Parisi, Lori Hentz, and Lisa Russ.

1. Introduction

Income and wealth concentration is increasingly viewed as a potential source of political and macroeconomic instability (Piketty, 2014, Stiglitz, 2012). A flurry of research showns that *income* is increasingly concentrated, that the increase in income concentration arises from both permanent and transitory factors, and that the tax system can help alleviate income concentration (Piketty and Saez, 2003; DeBacker, Heim, Panousi, Ramnath, and Vidangos, 2012; Auten, Gee, and Turner, 2013; CBO, 2014). Much of the recent research uses micro-level income tax data because income tax filing is nearly universal at the top. In the U.S., though, there are no micro *wealth* data with comparable coverage. Wealth concentration estimates are instead based on household survey data, "capitalized" income tax data, estate tax data, and Pareto interpolations.[2]

Though the same forces that increase income concentration may also increase wealth concentration—for example, through saved income—estimates of the growth in wealth concentration are inconclusive thus far (Bricker et al., 2016, Saez and Zucman, 2016, Kopczuk and Saez, 2004).[3] For example, wealth concentration has rapidly grown from 32 to 42 percent over the past decade when inferred from capitalized income tax data, but has grown modestly—from 33 to 36 percent—when measured in the Survey of Consumer Finances (SCF) (figure 1). Estate tax data show a more modest increase in wealth concentration.

However, we show in this paper that under reasonable assumptions there is little to no difference in wealth concentration growth estimates. We begin by showing that wealth concentration estimates derived by "capitalizing" income tax data with a model are highly

[2]The only wealth tax that exists in the U.S. is an estate tax applied at death and to very few families (less than 0.5 percent of the population, currently).

[3]Theory is also inconclusive. Wealth concentration can grow if wealthier families have higher returns on capital assets, and can fall if they have lower returns. Inheritances also have a theoretically ambiguous impact on concentration. In the data, wealthier families tend to realize higher returns on assets (Fagareng, Guiso, Malacrino, and Pistaferri, 2016)—serving to increase concentration—and inheritances appear to reduce concentration (Elinder, Erixson, and Waldenstrom, 2016; Boserup, Kopczuk, and Kleiner, 2016).

variable.[4] In a baseline income-to-wealth model (whereby all families are assumed to have the same rate of return on assets) the wealthiest 1 percent of US families hold 40.3 percent of wealth in 2011. Saez and Zucman (2016) show a model with a heterogeneous rate of return, whereby the interest income of the top 1 percent by *total income* is capitalized with the 10-year Treasury yield, *cet. par.* Wealth concentration in 2011 falls to about 38 percent in this model (figure 2). However, the estimate of wealth concentration falls to 33.7 percent when the interest income of either the *wealthy* top 1 percent or the top 1 percent of *interest income* is capitalized with the 10-year Treasury yield (figure 2).[5]

The full time series of these permuted wealth models shows a muted growth in wealth concentration relative to the baseline (figure 2). Growth in wealth concentration is similar to the SCF and estate tax estimates (figure 3).[6] Overall, then, wealth concentration estimates in a "capitalization" model can vary dramatically from slight deviations to the basic model. In this case, from a small change in *one* rate of return, applied to only one percent of the population—and it is even more sensitive to *which* one percent it is applied.[7]

The pace of wealth concentration when income tax data are "capitalized", then, depends on the plausibility of heterogeneous rates of return, and the correlation between rate of return on wealth (figure 2). Using population-wide Norwegian registry data, Fagareng, Guiso, Malacrino, and Pistaferri (2016) show that family-level asset rates of return are positively correlated with wealth rank. The evidence is more limited in the United States, but both the SCF and matched estate tax-to-income tax data indicate that wealthier families have

[4] We do not include a detailed estate tax data analysis, as we do not have access to these data and there are more concerns about their representativeness as fewer and fewer families file the estate tax over time (Saez and Zucman, 2016).

[5] Uncertainty also arises in these estimates from using annual income instead of permanent income to predict wealth, though this is a small part of the total variability.

[6] In fact, it appears that *only* the baseline model presented in that paper supports rapid growth in wealth concentration.

[7] The modelling needed to transform income to wealth is sensitive to small changes in assumed rates of return, especially when the return is low (Kopczuk, 2015). For example, in such a model, a one percent rate of return on assets means that we infer $100 of assets for every $1 of income. In this example, a 1 percentage point increase in the rate of return (to two percent) implies inferred wealth falls by half (to $50) from the same $1 of income

higher rates of return on interest-bearing assets—the focus of this analysis.[8] Applying these rates of return to the top of the wealth distribution in the "capitalization" model also leads to muted wealth concentration growth (figure 4).

Estimates from household surveys are variable for well-known sampling and non-sampling reasons. We estimate sampling and non-sampling errors in the SCF and show that these sources of survey variability are *smaller* than the modeling variability in the capitalization model when estimating wealth concentration. Ostensibly, the benefit of using income tax data to infer wealth is the nearly universal coverage at the top that the tax data provide. But, in the results shown here, this benefit is overwhelmed by the variability of using a model used to predict wealth.

The main sources of variability that we can estimate in the SCF come from sampling, item non-response, unit non-response, and coverage. The variability in wealth concentration estimates in the SCF is relatively small because the SCF survey design includes an oversample of wealthy families that can credibly represent the top of the wealth distribution–save the *Forbes* 400–and because non-response bias at the top is ignorable in the SCF (Bricker et al., 2016, Kennickell, 2011). We estimate coverage error arising from the exclusion of the *Forbes* 400 in two ways. In our preferred estimates, the coverage error variability is small. But we expect variability arising from coverage error in other household surveys without a wealthy oversample to be larger.

The missing *Forbes* 400 can be modeled by a Pareto distribution, while the rest of the wealth distribution is measured with SCF survey data (as in Vermeulen, 2015, and Dalitz, 2016). Pareto parameters tend to be highly variable, wealth concentration estimates estimated at least in part by Pareto parameters can be higher variable. But we focus on a second augmented wealth concentration estimate that uses a new set of SCF weights (figure

[8]Intuitively, it makes sense that wealthier families would have higher rates of return. Wealthy families have access to investment vehicles–such as private equity, hedge funds, and loaning money to a closly held business–that less wealthy families cannot typically access. This may be especially true in interest income where a typical family has access to low-yield savings accounts while wealthier families can invest in bonds. On the other hand, hedge funds have underperformed the market in recent years.

5). These weights take advantage of the overlap between the wealth distribution of SCF respondents and the wealth distribution of the Forbes 400 (as shown in, e.g., figure 4 of Vermeulen, 2015).

The rest of the paper is laid out as follows. In section II we describe a set of administrative records derived from income tax data. A version of these data are used in Saez and Zucman (2016) to infer wealth concentration estimates and in the SCF sampling process. In this section we also describe how to model ("capitalize") wealth from these administrative income data.

In section III, we describe how to estimate sampling and non-sampling variability in the SCF. We also document that respondents to the SCF are similar to non-respondents, and that the SCF provides good distributional coverage up to and including the wealthiest 400 U.S. families. We also show how to estimate wealth at the top of the distribution by assuming that it follows a Pareto distribution. We describe several tests of this assumption, too.

In Section IV we show wealth concentration estimates under various assumptions and we conclude in Section V.

2. Income Tax Data and Capitalization Models

Measuring and explaining wealth concentration has challenged economists at least since Pareto (1896). The only official wealth record that exists in the U.S. comes from an estate tax applied at death, so there is no administrative data system directly associated with measuring the cross-section of wealth at a point in time. Measuring top wealth shares using simple random sampling in household surveys is not a viable solution, because thin tails at the top lead to enormous sampling variability, and disproportional non-participation at the top biases down top share estimates. Later, we describe how the the Survey of Consumer Finances (SCF) overcomes both problems by oversampling at the top using administrative data derived from income tax records. In this section we describe efforts to infer wealth directly from these administrative income data by "capitalizing" taxable income.

2.1 Administrative tax data

The Individual and Sole Proprietor (INSOLE) data are a set of administrative records derived from income tax returns. The INSOLE file consists of a sample of individuals and sole proprietorship tax filings from Internal Revenue Service (IRS) administrative tax data, and are statistically edited for quality by the Statistics of Income (SOI) at the IRS (see, for example, Statistics of Income, 2012). These data describe the distribution of income and income sources, deductions, and taxes paid in a tax year.

The Public Use File (PUF) is a modifed version of the INSOLE files that are available for public use, either through NBER or directly from SOI.[9] In most of our work we use the PUF file for the years 2001 to 2011 (the last year they are available), though we sometimes use the INSOLE files (years 2011, 2008, 2005 and 2002).

Though the files contain several hundred-thousand observations, they weight up to the full population of tax units. In 2011, for example, there were about 135 million tax units. For comparison, there are about 110 million *families* measured in the 2011 March CPS. There are fewer families than tax units because some families can choose to file taxes separately.

2.2 Capitalization models

One can predict wealth from income by "capitalizing" (or inflating) asset income by an asset-specific rate of return (Greenwood, 1983, Kennickell, 2001, Saez and Zucman, 2016). However, many forms of wealth do not appear on a tax form, and a capitalization model will also include an estimate of these non-financial sources of wealth. The general form of such a model is:

$$\widehat{wealth}_i^{CAP} = non\widehat{financial}_i + kg_i + \sum_{k=1}^{K}(income_i^k / r^k), \tag{1}$$

[9]Though these data are avaiable, they are restricted-use. There is a fee associated with obtaining the PUF files available directly from SOI, though the NBER version is free to associated researchers.

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 (0,1). Capital gains may or may not be included in such a model.

2.2.1 Saez and Zucman (2016) model

Saez and Zucman (2016) use a version of the gross capitalization model described above (equation 1) to predict the stock of wealth held by each tax filing unit. Their estimated rate of return cleverly distributes the household asset stock of the Financial Accounts of the United States (FA) based on how realized capital income is distributed.

Income and assets are organized into seven classes: (1) *taxable* interest bearing assets, (2) *non-taxable* interest bearing assets, (3) dividend-producing assets (e.g. from publicly-traded companies), (4) pass-through business assets that produce Schedule E income, (5) corporate business assets that produce Schedule C income, (6) privately-held pension assets (IRAs, 401(k)s), and (7) workplace defined benefit pensions.

An asset-class rate of return is found by the ratio of INSOLE income to the FA asset stock for each of the seven types ($ror^{K,SZ} = (\sum_{i=1}^{N} income_i^{INSOLE^k})/FA^k$). Applying that rate of return to INSOLE income blows up that income into an estimate of wealth.

$$\widehat{wealth}_i^{CAP,SZ} = non\widehat{financial}_i + kg_i + \sum_{k=1}^{K}(income_i^{INSOLE^k}/ror^{K,SZ}) + \widehat{pensions}_i, \quad (2)$$

2.2.2 Heterogeneous rates of return

In the capitalization models decribed above, all families realize identical rates of return by asset. However, good evidence suggests that wealthier families get higher rates of return (Fagareng et al, 2016). And the inputs to the capitalization model in Saez and Zucman (2016) yields a rate of return on interest-bearing assets that is much lower than market-based rates such as the 10-year Treasury yield or Moody's Aaa corporate bond, espcially in the late 2000s when interest rates fell (Bricker et al, 2016, Kopczuk, 2015). For example, in 2011 the ratio of income to FA assets yields a rate of return on taxable interest of 1.15%. The 10-year Treasury yield is 2.78%, on average, in 2011.

But the capitalization model can be modified to allow some families to get higher rates of return than others. The FA shows that households own about $10.9 trillion in taxable bonds, deposits and other fixed income assets in 2011. Taxable interest income of the families at the top end—say, the top 1 percent—can be capitalized at the 2.78% rate of the 10-year Treasury, and the remaining income and assets are capitalized at a lower rate (to keep total interest assets equal to the FA total of $10.9 trillion).

In a capitalization model, when the rate of return for top-end families is higher than the average return then the share of wealth held by the top-end will fall. This is especially true when rates of return are low (Kopczuk, 2015)

2.3 Sources of variability in capitalized income data

When wealth is estimated from income in the capitalization model, wealth shares will depend on the estimated rates of return for assets, on the income data, and on the FA asset data. We will call "modeling error" the changes in estimated wealth share due to deviations in the rates of return in the capitalization model. The main source of variability in wealth share estimates from the income tax data are due to modeling error. It is well-known that surveys can suffer from sampling and non-sampling error. It is less-known that the same sources of error can afflict the administative income tax data. In this paper we do not estimate any of these sources of variability in the income tax data, but do note their existence. As such, any estimates of variability can be considered lower bounds.

2.3.1 Modeling variability This is the main source of variability we find for predicting wealth from the income tax data. When income is capitalized into wealth (as in equation 2) there is one main set of parameters to the model–rate of return–and one main input–income. We can vary the rate of return, though, and see how wealth predictions vary in order to estimate modeling error.

One can also think about varying the income inputs into the model. Annual income, as in (Saez and Zucman, 2016), can be a mixture of permanent icnome and transitory income. Appendix A describes why a tax-unit panel of taxable income (to approximate permanent income) may be preferrable to annual income of tax units. We vary annual and permanent income to see how wealth predictions vary.

2.3.2 Sampling and non-sampling variability The INSOLE file is a sample from the IRS administrative records.[10] As such, there is sampling error associated with the use of these data. We do not have access to the full sampling strategy used by SOI, but in principle one could create bootstrapped standard errors, replicating the sampling strategy (and approximating sampling error in the data).

The INSOLE data cover the full population of tax filers, but users of the data need to apporximate the number of non-filers (Saez and Zucman, 2016). One realization of this number may be 20 million, but varying this estimate can approximate the coverage error inherent in these income tax data. Unit nonresponse in the tax data overlaps with coverage error, as many self-employer filers may claim less income than the filing level and thus not file.

Item nonresponse may occur in the tax data when a family does not claim positive income on a type of income when the family does, in fact, have positive income. For example, a 1099-INT is automatically generated by a financial institution when interest income is greater than \$10, but when interest income is less than \$10 then the family will not get this reminder to claim interest income on their tax filing.[11]

The SOI data may have measurement error when families do not accurately file. Some estimates of business filing show significant under-reporting of business income. Concept

[10] See https://www.irs.gov/pub/irs-soi/sampling.pdf

[11] Appendix B of Bricker et al (2016) shows that the number of returns with positive interest income has fallen over the past decade while the number of families with interest-bearing accounts has stayed constant. A decline in interest rates on savings accounts may mean that fewer families will get an automatic 1099-INT reminder.

validity error can occur on the tax data when a family is confused about where to report certain income or deductions. Statistical editing helps alleviate this, and SOI does this. The statistical editing process is good in that it alleviates measurement error and concept validity error, but can introduce processing error. Weights are used in the INSOLE data, too, which can lead to adjustment error.

3. Survey of Consumer Finances data

The SCF is a cross-section survey, conducted every three years by NORC on behalf of the Federal Reserve Board (FRB) and with the cooperation of SOI.[12] The SCF provides the most comprehensive and highest quality survey microdata available on U.S. household wealth. SCF families respond to questions about financial and nonfinancial assets, debts, employment, income, and household demographics.

As noted before, measuring income and wealth at the top using simple random sampling is not viable due to the concentrated nature of economic resources. Thin tails at the top lead to large sampling variability, and disproportional non-participation at the top biases down top share estimates. Both make measuring wealth concentration extremely difficult. The Survey of Consumer Finances (SCF) overcomes both problems by oversampling at the top using administrative data derived from income tax records, and by verifying that the top is represented using targeted response rates in several high end strata. The list sample ensures that the SCF has adequate representation of the upper tail of the wealth distribution and to ensure adequate representation of sparsely held assets.

[12]See Bricker, et al (2014) for results from the most recent triennial SCF. A great degree of security is involved with this sampling procedure and formal contract govern the agreement between the FRB, NORC and SOI. The FRB selects the sample from an anonymized data file. The FRB sends the sampled list to SOI, who remove the famous families and passes along the list to NORC for contacting. NORC collects the survey information and sends to FRB. Thus, the FRB never knows any contacting information, SOI never knows any survey responses, and NORC never knows anything more than survey responses and location information.

3.1 SCF sampling

The SCF combines a geographically-stratified and nationally-representative area probability (AP) sample with a list sample (LS), an oversample of households that are likely to be wealthy. In the 2013 SCF, there are about 6,000 families surveyed by the SCF, of which about 1,500 are from the list sample. The AP sample is drawn by NORC at the University of Chicago and provides a nationally-representative sample of families.[13] The list sample is drawn using a frame of statistical records derived from tax returns—the INSOLE data described above.[14]

The list sample frame data are typically based on income tax returns filed the year prior to the SCF survey year, meaning that the income was earned two years prior to the SCF survey year. In the 2013 SCF, for example, the frame data were derived from tax returns covering income earned in 2011. However, since 2001 the SCF list sample is drawn using multiple years of income for the returns in the frame. In the 2013 SCF, for example, the 2011 income data in the frame were supplemented with 2010 and 2009 income.[15]

3.1.1 SCF sampling models. The SCF sampling strategy uses two methods of predicting wealth from income. The first is a gross-capitalization model, described in equation 1 in the previous section. There are six types of income in the SCF model: taxable interest,

[13]See Tourangeau, et al. (1993), O'Muircheartaigh et al. (2002) for more information about the NORC national samples.

[14]The unit of observation in the INSOLE data is a tax unit while the SCF unit of observation is a family. In practice, there are millions more tax units than families because several members of a family can file distinct tax returns; without a correction, these multi-filer families would have a disproportionately large chance of being selected. To account for this in the SCF LS sampling process, the INSOLE sampling weight of tax units that filed married filing separately is divided in half. Further, all filers below the age of 18 are dropped (a family headed by someone less than age 18 is ineligible for the SCF). Still, to a certain extent, the discrepancy between tax units and families remains in the adjusted INSOLE sampling frame.

[15]The INSOLE file is not designed to be a panel, though certainty sampling of high income families (among others) means that families with consistently high incomes are often in the sample year over year. Filers with total income of at least $5 million, filers with total income of less than negative $5 million, filers with $50 million of Schedule C receipts, and filers with at least $200,000 of AGI but zero tax liability are all sampled with certainty (Czajka, Sukasih, and Kirwan, 2014; Bryan, 2015). Filers with at least $2 million (or less than negative $2 million) in income are sampled at about a 50 percent rate. The file is also sampled in a Keyfitz method, meaning that there is a strong overlap between adjoining year files.

non-taxable interest, dividend income, rents and royalties (in absolute value), business, farm, and estate income (in absolute value), and capital gains (in absolute value).[16]

The general form of the SCF capitalization model is:

$$\widehat{wealth}_i^{GC} = \widehat{house}_i + kg_i + \sum_{k=1}^{K=6} (income_i^{INSOLE^k}/r^k), \tag{3}$$

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

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 by year T and the base sampling income data are from two years prior to that:

$$ln(SCFwealth)_i^T = ln(\overline{income}_i^{T-2})\beta + varepsilon_i, \tag{4}$$

The matrix of sampling income for the previous SCF $\overline{income}_i^{T-5}$ consists of more than 30 logged income variables and a dummy indicating the presence of such income for that tax unit, plus some basic demographic data.[17] The β vector from this regression model is then applied to the current administrative sampling data to obtain a predicted wealth index:

$$\widehat{wealth}_i^{ECorr} = f(\overline{income}_i^T; \hat{\beta}), \tag{5}$$

Recall that both models use a blend of multiple years of income data, which helps to smooth over transitory income fluctuations that are especially prevalent for capital incomes and

[16]Model details are provided in Appendix A, including rates of return. Income is a weighted average of three years of sampling income.

[17]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; complete details are provided in Appendix A.

at the top of the distribution. In contrast to the gross-capitalization model, the empirical correlation model allows a variety of income variables that are not necessarily based on a physical asset and allows rates of return to vary across different types of families.

Each model creates an independent sets of rankings, or wealth indices. Each index is normalized and the two normalized indices are blended together to create a wealth ranking from which the LS sample is created.[18] About 5,100 LS cases are selected and the majority are from sampling strata that capture the top 1 percent of expected wealth. Response rates at the top end strata ranged from 33 percent to about 12 percent in the 2013 SCF. The SCF sample weights adjust for nonresponse and thus account for this variability. Bricker et al (2016) show that the income of the responding families is similar to that of the non-responders.

3.2 Sources of variability in SCF survey data

The sources of variability in surveys are well-established (see, for example, Weisberg, 2005), and fall under two main headings: sampling error and non-sampling error. We will describe each below, but in this paper we can estimate sampling error, coverage error, unit nonresponse error, and item nonresponse error but we will not be able to estimate measurement error, adjustment error, concept validity error, or processing error.

3.2.1 Sampling error. Sampling theory (see, for example, Neyman, 1934) allows a household survey to sample and interview a few households but make inferences about the population of households. The sample mean ($\hat{\overline{x}}$), for example, is an unbiased estimate of the true population mean ($\overline{x}$). The benefit is that population inferences can be made without the cost of conducting a census. The cost, though, is that inferences are based on one realization of a sample rather than the population. Any given sample realization may be different from the true population mean, and different sample realizations may have different

[18]Typically, the blend is a 50/50 split, although in recent years the split has favored the windex-1 model, due to the strengths that we will observe later.

sample means. The difference between the true population mean and the sample mean is called the sampling error.

Sampling error can be estimated, and the SCF staff have produced a set of replicate weights to estimate sampling variability (described in Kennickell and Woodburn, 1999).[19] The replicate weights are derived by resampling the SCF respondents along the dimensions of the SCF sample design; the resampling is done 999 times and a unique set of weights are calculated eacg time. 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.

3.2.2 Unit nonresponse

Unit nonresponse is one of the non-sampling errors possible in a survey (along with item nonresponse error, coverage error, measurement error, concept validity error, processing error, and adjustment error).

Unit nonresponse is when a sampled family does not participate in the survey. The SCF wealth share estimates would be biased if the participants were different than the non-participants (i.e. biased down the poorest LS families participate but the wealthiest do not). We show in earlier work (Bricker et al 2016) that this is not the case. By sampling strata we show that the family income and capital income of respondents comes from the same distribution as the non-respondents. So we argue that unit non-response is ignorable in the SCF top wealth share estimates.

3.2.3 Item nonresponse

Item nonresponse describes the situation where a responding SCF family refuses (or cannot) answer all of the survey questions. Considering only the "complete cases" and ignoring the cases with item nonresponse will lead to selection bias 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 (Kennickell 1995). Five "implicates" are imputed for each missing value. Multiple imputation is used in the SCF to

[19] The FRB provides bootstrap replicates for the public SCF data on the SCF website.

acknowledge that any imputation model can only recover some distribution of the underlying missing data. 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.

3.2.4 Coverage error occurs when the sampling frame cannot cover the entire population. The AP sample is derived from an address-based sample and covers the entirety of the U.S.[20] The list sample covers the upper tail of the wealth distribution, allowing the SCF to have coverage at the top. However, the SCF is not allowed to sample the *Forbes* 400 families; these missing 400 imply coverage error in the SCF.[21]

We describe two ways to handle this error. First, we can augment the SCF sample weights to include the missing *Forbes* 400 wealth. Second, we can estimate coverage error by assuming a that the missing *Forbes* 400 wealth follows a Pareto distribution. Pareto distribution has been used in other studies to augment European household survey data (Vermeulen, 2015, Dalitz, 2016, Eckerstorfer et al, 2016).

Weights correction Our preferred treatment of coverage error will involve adjusting the SCF sample weights at the top and including a weighted version of the *Forbes* 400 wealth. We do so in a "combining samples" weighting approach by leveraging the overlap between the *Forbes* wealth and the wealth of some SCF respondents (O'Muircheartaigh and Pedlow, 2002).[22] The *Forbes* list relies, in part, on public knowledge of wealth (through public filings for publicly-traded companies, or through self-promotion). Privately-held forms of wealth, for example, can evade such public knowledge.

[20]Save some very remote areas that are too hard to contact in person, but less than 0.1 percent of the population lives in these areas.

[21]These families are too easily identifiable to be released in a public dataset.

[22]We do so in a similar way to how the AP and list sample weights are are woven together to create final weights for the SCF (Kennickell and Woodburn, 1999). See, for example, Vermeulen (2015) for a visual of the overlap in the 2010 SCF, and Kennickell (2001). This overlap exists in every survey year used in this analysis.

We begin by creating three wealth bins (\$1-\$2 billion, \$2-\$5 billion, and \$5 billion or more) and counting the number of SCF and *Forbes* cases–weighted and unweighted–in the bins.[23] In each bin (b), we find the relative frequency (RF) of SCF and Forbes cases by the formula

$$RF_{b,t} = (n_{b,t}/N_{b,t})/[(n_{b,SCF}/N_{b,SCF}) + (n_{b,Forbes}/N_{b,Forbes})] \tag{6}$$

for $t = \{SCF, Forbes\}$, $b = \{\$1 - \$2bill., \$2 - \$5bill., \$5 + bill.\}$, where n is an unweighted count in bin b, N is a weighted count in bin b, and $RF_{b,t}$ is defined in [0,1].

The combinded and adjusted weight is $adjusted_{wgt} = RF_{b,SCF} * SCF_{wgt} + RF_{b,Forbes} * Forbes_{wgt}$, where RF depends on b. With this weight we can use wealth information in the SCF and *Forbes*, weighted properly for the overlap in the two datasets.[24] Using this weight allows us to treat coverage error in the SCF wealth data and still use bootstrap replication techniques (described above) to estimate sampling variability.

Pareto The Pareto (or power law) distribution can describe the distribution of many social and natural phenomena–from the population of cities to the frequency of last names in a country (Clauset et al 2009). Its earliest use was to describe the distribution of wealth (Pareto, 1893). Thus, other household surveys that suffer from a larger degree of under-coverage at the top have turned to modelling top tail wealth in a population with the Pareto distribution. The Pareto distribution is described by the equation

$$p(x) = x^{-\alpha}, \tag{7}$$

for values of x that are greater than x_{min}. When using this distribution, then, the value of x_{min} needs to be determined (through assertion or estimation) and the value of α is

[23] We assume that Forbes families are self-representing with weight of one so the number fo weighted cases is equal to the number of unweighted cases. We use the SCF survey weight when considering the SCF cases.

[24] When SCF families with wealth greater than the minimum *Forbes* wealth have a sample weight greater than one, they represent not just themselves but other families with their wealth level. These are presumably families in the *Forbes* list. Thus, the SCF sample weights prior to this weight correction represent some of *Forbes* families.

estimated from the data using a MLE method. Once $\hat{\alpha}$ is estimated, the amount of wealth at the top of the distribution–that is, above x_{min}–can be estimated from the CDF of the Pareto distribution.

In our work, we will estimate the missing wealth at the top of the SCF by assuming that the value of x_{min} is the minimal value of the *Forbes* 400. In every SCF survey there is overlap between the *Forbes* list and SCF respondents' wealth (Vermeulen, 2015). Because there is variation from estimating the value of α, this exercise will add variation to the SCF top share estimates. But we will also show that varying the value of x_{min} can easily increase the variation in this exercise.

3.2.5 Other errors include measurement error, concept validity error, processing error, and adjustment error. We will not be able to estimate variability for any of these errors. Measurement error happens when a respondent gives a value that does not reflect the true condition of the family. Some of this happens in the SCF—like any other survey—though it is mitigated by extensive case review. SCF values match up to external aggregates fairly well, too (Dettling et al 2015). Concept validity error occurs when the respondent does not understand the question (e.g. confuses IRAs with 401(k)s). Statistical editing helps alleviate this, and the SCF does this. The statistical editing process is good in that it alleviates measurement error and concept validity error, but can introduce other error: processing error. Adjustment error occurs when weighting goes bad. But external aggregates that the SCF rake and post-stratify to are the best around.

4. Results

Baseline trends wealth concentration for the capitalized income data and the SCF are shown in figure 1. In both cases, the concentration estimates are around 33 percent in 2001, and grow steadily until 2007. After 2007, concentration estimates from the SCF continue to grow steadily, but the estimates from the capitalization model grow rapidly. Overall, the

SCF estimates grow about 4 percentage points from 2001-2013, but concentration estimates from modeled income tax data grow almost 9 percentage points from 2001-2012 (a 30% increase from the 2001 level).

The baseline capitalized wealth concentration assumes homogeneous rates of return across families for all types of assets. The SCF concentration number does not include the uncovered *Forbes* 400, either. How do the trends in wealth concentration change once uncertainty is taken into account?

4.1 Variability in wealth concentration from capitalized wealth

The red line in Figure 1 relies on the baseline baseline income-to-wealth model from Saez and Zucman (2016) and relies on an assumption of homogeneous rates of return across families, by asset. The baseline model, though, predicts that nearly half of the wealth of the wealthiest families is held in fixed income instruments. In fact, nearly all of the growth in wealth concentration since the mid-2000s is due to the growth in fixed-income assets held by the wealthiest families.

As rates of return go to zero, there is a non-linear increase in predicted wealth in the capitalization model (Kopczuk, 2015). For example, when the rate of return is 3%, then the model predicts that $1 of interest income is associated with $33.33 of wealth. That same dollar is associated with $50 of wealth when the return falls by one percentage point to 2%, but is associated with $100 when rates fall another percentage point to 1%.

In the baseline capitalization model, the fixed-income rate of return is quite low–just over one percent in the year 2011, for example. This is far lower than the 10-year Treasury yield or Moody's Aaa corporate bond yield. One potential robustness then, is allowing the top-end families to have a rate fo return on fixed-income assets equal to the 10-year Treasury yield.

But which "top-end" families should get this higher return? Saez and Zucman (2016) capitalize the interest income of the top 1 percent by *total income* with the 10-year Treasury yield, while all other asset classes are capitalized as described in Section II. Wealth concentration falls from about 40 percent to 38 percent in 2011 (figure 2).

However, the top 1 percent by income are often the "working rich". These families hold a lot of interest income (about 35 percent). But, the top 1 percent of interest income holders hold about 70 percent of interest income (the top 1 by wealth hold a similar share becasue there is strong overalp between the top 1 of interest income and top 1 of wealth).

When the higher rate of return is applied only to the top 1 by income, then one has to believe that most of the wealthiest families still earn a tiny rate of return on interest bearing assets. If one believes that this group is invested in Treasuries–instead of savings accounts–then we should apply the 10-year Treasury retrun to interest income of the *wealthiest* 1 percent (or the top 1 percent of *interest income*). Estimates of wealth concentration falls to 33.7 percent in 2011 in these alternate heterogeneous return models (figure 2).

The full time series of these permuted wealth models shows a muted growth in wealth concentration relative to the baseline (figure 2). In the baseline model, the increase in wealth concentration is close to 9 percentage points. When returns are correlated by wealth, the increase is about 3.5 percentage points: from 30 percent 2001 to 33.7 in 2011.

The mid-2000s mark the beginning of the current low interest-rate environment. At that time, rates of return on interest-bearing assets began to fall, a decline that can also be seen in the ratio of taxable interest income to FA assets. That said, the rates of return implied by this ratio—for example 1.15% in 2011—are clearly an average of returns held by low-wealth and high-wealth families. The 10-year Treasury yield (used above) is a plausible proxy for the rate of return on interest-bearing assets available to wealthy families during this time. Traditional savings accounts—available to most other saving families—had much lower yields duting this time.

4.1.1 Evidence for heterogeneous rates of return in the US There is good evidence from Scandanavia that rates of return are positively correlated to wealth (Fagereng et al 2016). In the US there are two data sources that we can use to examine the heterogeneity in returns: the estate tax data and the SCF. Both have limitations. The estate tax data are a non-random set of decedents in a given year. The SCF represents the US population, but is a sample of only 6,500 families. But both data sources support a positive correlation between wealth and rates of return on interest-bearing assets.

The baseline rate of return on interest-bearing assets and that implied by the estate tax data is described in Saez and Sucman (2016, appendix table C6b), and reproduced in table 1. The overall rate of return and that for wealthy SCF families is also shown in table 1. A time series of the returns for wealthy families in the estate-tax data and SCF data is shown in figure 6. The average returns are remarkably consistent over time: both increase from the early to the mid-2000s and then decrease from the mid-2000s onward.

The growth in wealth concentration is muted relative to the baseline when interest income of wealthy families is capitalized with these rates of return (figure 4). Wealth concentration in the estate-tax model grows steadily from 2002-2011, and 3.5 percentage points from 2007-2011. But concentration barely grows when wealthy SCF respondents' returns are applied to the model.

Though the two lines are similar in the early 2000s, they diverge somewhat later in the period. This divergence highlights the sensitivity of the capitalization model when rates of return are low. Wealth concentration in the SCF and estate-tax models is nearly identical in 2007, but is several percentage points higher in the estate-tax model in 2010. The only difference is that interest income of the wealthiest families is capitalized at 1.7% return in the estate tax model but at 2.4% in the SCF model.

Another piece of data support the idea of heterogeneous rates of return: the Financial Accounts assets data themselves. If wealthy families realized a lower-than-market rate on interest assets (as the homogeneous model implies in the late 2000s-early 2010s), that would imply that wealthy families were holding considerable balances in savings or checking accounts—not higher-yielding bonds—and that these balances would be growing during the late 2000s-early 2010s.[25] But large-balance savings accounts measured in the FA did not grow during the late 2000s-early 2010s, making it improbable that these wealthy families were increasing their holdings of these low-yield assets.[26] All of the rapid growth in household ownership of non-bond fixed-income assets, then, is due to growth in small balance accounts.[27]

Overall, then, the baseline capitalization model—one with homogeneous returns—is the only model tested here that shows a dramatic rise in wealth inequality. Any of the models with heterogeneous returns show much lower growth in wealth concentration from 2001-2011 and most show no growth since the late 2000s (when the current low interest-rate environment began).

4.2 Variability in SCF wealth concentration estimates

Overall, the variability in the SCF top 1 percent wealth share is about plus or minus 2 to 3 percentage points in each survey year (symmetric about the point estimate). We first discuss how adding an estimate of the *Forbes* 400 affects the SCF top wealth share, and then estimate sampling error and item nonresponse error. In other work, we show that unit non-response at the top-end in the SCF can be ignored (Bricker et al 2016, figure 3).

4.2.1 Combining Forbes and SCF information Though the SCF is precluded from sampling from the *Forbes* 400, some SCF respondents are as wealthy as *Forbes* families. As described in Section III, we developed weights to incorporate the wealth of these omitted families into the SCF wealth totals. SCF wealth concentration is after the weight adjustment is mostly a level shift up by about 2.5 percentage points (figure 5).

4.2.2 Sampling and non-sampling variation We can estimate the sampling variability by resampling our data through bootstrap techniques described in Section III. We

[25] Recall that the 10-year Treasury yield in 2011 was 2.78% and the Moody's Aaa average yield was 4.5% in 2011.

[26] Large balance is defined as accounts with $100,000 or more. See table L.205 in the Z1 release for the FA.

[27] See table B.101 in the Z1 release for the FA.

resample the SCF data 999 times (replicating the sample design) and recalculate top wealth shares in each resample to compute an estimate of sampling error. The SCF top 1 percent wealth share varies by 1 or 2 percentage points each year due to this variability.

It may be surprising that sampling variability is so small for the top 1 percent share in a household survey. But recall that the SCF has a heavy oversample, where between 1/6th and 1/8th of the overall sample is expected to be in the top 1 percent. Such a heavy oversample surely helps keep sampling variability low for the top 1 percent wealth share.

Item non-response variance can be estimated by the variance across the five implicates of (multiply-imputed) SCF data. 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 of the top 1 percent wealth share due to both sampling and imputation is described by the formula: $SE^{overall} = (var^{sampling} + (6/5)var^{imputation})^{0.5}$. The shaded band in figure 5 describes this variation. Of note: the variability estimated for the SCF survey is smaller than the range of estimates possible from tweaks to the capitalization model–what we have called "model variability" (figure 7).

4.3 Overall wealth concentration trend

The capitalized income data and SCF survey data also measured in different units and have slightly different wealth concepts. The units of analysis are tax units in the income tax data and the family in a household survey. There are many more tax units than families (161 million versus 122 million in 2012, for example). Families in the bottom 99 percent are often split into multiple tax units, but a tax unit in the top 1 percent is almost always a family. Counting the top 1 percent (1.61 million) of tax units, then, effectively includes more families than counting the top 1 percent (1.22 million) of families in a survey. The capitalized income data includes an estimate of DB pension wealth owed to families, while the baseline SCF does not.[28]

[28] These ideas are explored in detail in Bricker et al (2016). Here we provide just a summary treatment.

We modify the SCF to include an estimate of DB pension wealth and adjust the family-level SCF to tax-units (figure 7). The feasible set of capitalized wealth concentration values (the red shaded area in figure 7) maps the spread of potential wealth concentrations in figure 3. The blue dashed line shows the SCF, including the *Forbes* coverage correction, including an estimate of DB pension wealth,[29] and a tax-unit of analysis correction.[30] The range of estimates allowed in the SCF and the capitalized income data overlap nearly every year.

The final figure plots the estimated wealth share inferred from estate-tax data itself (figure 8).[31] There is no increase in wealth inequality in this data series from 2001-2008 (when the updated series ends). These estimates are at the individual level, so the level of wealth concentration is considerably lower than the capitalized income tax (tax unit-level) or SCF (family-level)

5. Conclusion

Measuring wealth is difficult and each data source considered here has benefits and costs, which we have tried to explore. But we should also remember that the datasets have considerable commonality between them. The top-end estimates in the SCF are based on identifying wealthy families from administrative income tax data, and then asking that family the value of their assets in a structured interview. The capitalized income tax data also identify a set of wealthy families from the same administrative income tax data, and predict a value for their wealth. The estate-tax data are based on an assessment of the value of assets of wealthy decendents. These values are often determined after a structured negotiation between tax lawyers.

[29] Detailed information on DB pensions—whether at a current job or a past job—are collected in the SCF. We use this information and life tables to allocate the DB wealth across families.

[30] A concentration estimate based on tax units will be biased up compared to an estimate based on families. In the 2010 SCF, for example, fewer than 3 percent of coupled families in the top 1 percent filed separately, while about 17 percent of couples in families in the bottom 99 percent filed separately. The implication, then, is that any top share fractile estimate is effectively based on a population that may include 30 percent more family units than the fractile suggests.

[31] See appendix table C4 of Saez and Zucman (2016) for this update.

Not surprisingly, then, wealth concentration—as measured by the share of wealth owned by the wealthiest one percent of families—is broadly consistent across the variety of estimates and data sources considered here. Wealth concentration was growing at a moderate pace prior to 2001, and based on the estimates shown here, it has in all probability continued to grow at a moderate pace since then.

One set of estimates indicates robust growth in wealth concentration: capitalized income tax data with an assumption homogeneous rates of return. However, both the estate tax data and the SCF data on wealthy families supports heterogeneous rates of return (for interest-bearing assets) and a positive correlation between wealth and the rate of return.

Regardless of the trend or level of wealth concentration, though, wealth concentration is increasingly viewed as a potential source of political and macroeconomic instability (Piketty, 2014, Stiglitz, 2012), and is increasingly on the minds of US households.

References

[1] 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*, vol. 100 (September), pp. 1–41.

[2] Bricker, Jesse, Alice Henriques, Jacob Krimmel, and John Sabelhaus (2016) "Measuring Income and Wealth at the Top Using Administrative and Survey Data," *Brookings Papers on Economic Analysis*, Spring.

[3] Bryan, Justin. 2015. "High-Income Tax Returns for 2012", Statistics of Income Bulletin, Summer, pp. 1-60.

[4] Czajka, John, Amang Sukasih, and Brendan Kirwan. 2014. "An Assessment of the Need for a Redesign of the Statistics of Income Individual Tax Sample" mimeo.

[5] Clauset, Aaron, Cosma Rohilla Shalizi, and M.E.J. Newman. 2009. "Power-law Distribution in Empirical Data", *SIAM Review*, vol. 54, no. 2, pp. 661-703.

[6] Congressional Budget Office. 2014. "The Distribution of Household Income and Federal Taxes, 2011." Washington. https://www.cbo.gov/publication/49440

[7] Dalitz, Christoph. 2016. "Estimating Wealth Distribution: Top Tail and Inequality" Hochschule Niederrhein Technical Report 2016-01.

[8] Eckerstorfer, Paul, Johannes Halak, Jakob Keppler, Bernhard Schutz, Florian Springholz and Rafael Wildauer. 2016. "Correcting for the MIssing Rich: an Application to Wealth Survey Data." *The Review of Income and Wealth*, vol. 62, no. 4, pp. 605–627.

[9] Debacker, Jason, Bradley Heim, Vasia Panousi, Shanthi Ramnath, and Ivan Vidangos. 2013. "Rising Inequality: Transitory or Persistent? New Evidence from a Panel of U.S. Tax Returns." *Brookings Papers on Economic Activity*, Spring: 67–122.

[10] Dettling, Lisa J., Sebastian J. 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." Finance and Economics Discussion Series, no. 2015-086. Washington: Board of Governors of the Federal Reserve System.

[11] Fagereng, Andrea, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. 2016. "Heterogeneity in Returns to Wealth and the Measurement of Wealth Inequality." *American Economic Review*, forthcoming

[12] Greenwood, Daphne. 1983. “An Estimation of U.S. Family Wealth and Its Distribution from Microdata, 1973.” *Review of Income and Wealth* vol. 29, no. 1, pp. 23–44.

[13] Kennickell, Arthur. 2007. “The Role of Over-sampling of the Wealthy in the Survey of Consumer Finances” mimeo.

[14] 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*, vol. 45, no. 2, pp. 193–215.

[15] Kopczuk, Wojciech. 2015. “What Do We Know about the Evolution of Top Wealth Shares in the United States?” *Journal of Economic Perspectives* 29, no. 1: 47–66.

[16] Kuznets, Simon. 1953. “Shares of Upper Income Groups in Income and Savings.” New York: National Bureau of Economic Research.

[17] O’Muircheartaigh, Colm, Stephanie Eckman, and Charlene Weiss. 2002. “Traditional and Enhanced Field Listing for Probability Sampling.” In Proceedings of the Joint Statistical Meetings, Survey Research Methods Section. Alexandria: American Statistical Association.

[18] O’Muircheartaigh, Colm, Steven Pedlow. 2002. “Cumlulating Cases versus Combing Samples.” In Proceedings of the Joint Statistical Meetings, Survey Research Methods Section. Alexandria: American Statistical Association.

[19] Pareto, Vilfredo. 1896. “Cours d’conomie Politique.” Geneva: Droz.

[20] Piketty, Thomas. 2014. “Capital in the Twenty-First Century.” Belknap Press.

[21] Piketty, Thomas, and Emmanuel Saez. 2003. “Income Inequality in the United States, 1913-1998.” *Quarterly Journal of Economics*, vol. 118, no. 1: 1–39.

[22] Saez, Emmanuel, and Gabriel Zucman (2016) “Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data,” *Quarterly Journal of Economics*, vol. 131, no. 2, pp. 519–578.

[23] Statistics of Income. 2012. “Individual Income Tax Returns.” Washington, DC: Internal Revenue Service

[24] Stiglitz, Joseph E. 2012. “The Price of Inequality: How Today’s Divided Society Endangers Our Future.” New York: W. W. Norton.

[25] Vermeulen, Philip. 2015. “How Fat Is the Top Tail of the Wealth Distribution?” Working Paper no. 1692. European Central Bank.

Figure 1: Wealth concentration in SCF and modeled from income tax data (2001-2013)

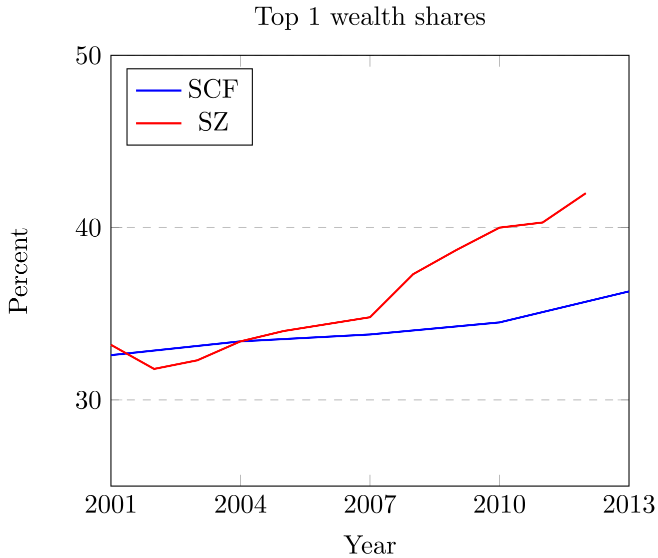


Figure 2: Modeled wealth concentration is flat in permuted model (2002-2011)

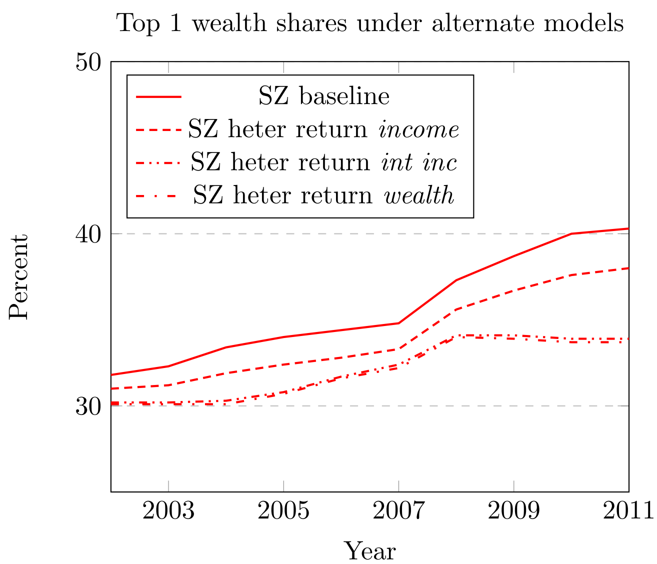


Figure 3: Growth in wealth concentration (2001-2013)

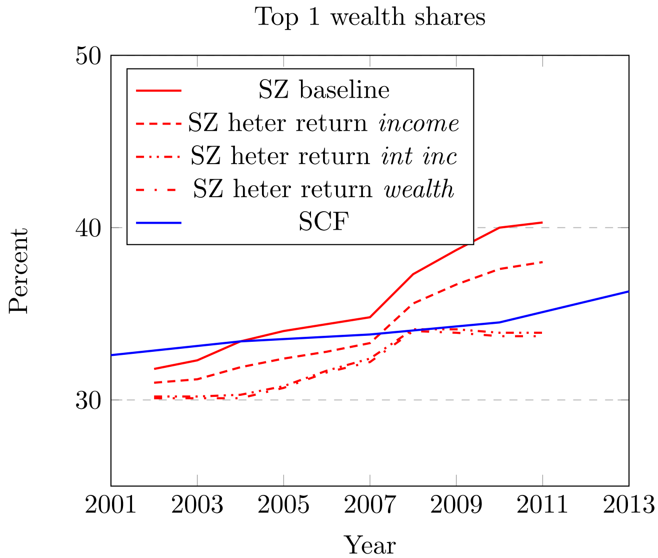


Figure 4: Wealth concentration in cap. model: estate and SCF interest rates of return

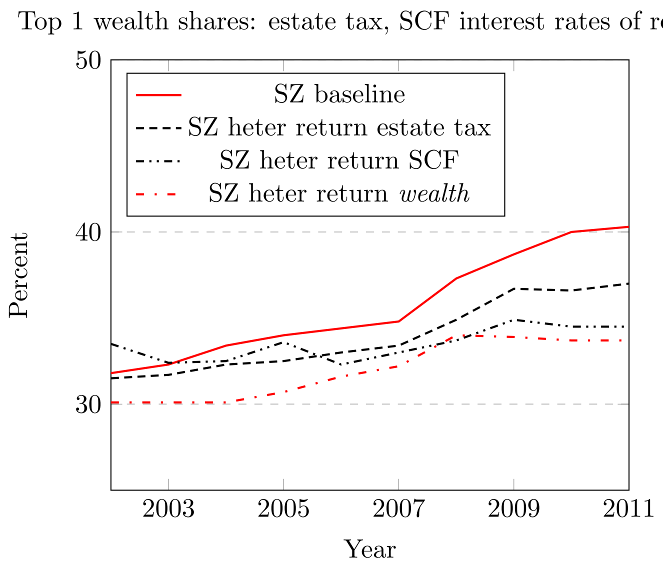


Figure 5: Concentration estimates including *Forbes* wealth

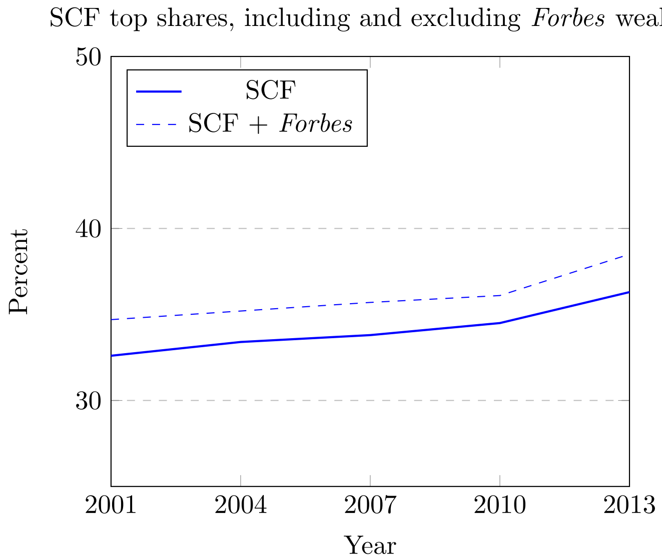


Figure 6: Rates of return on interest-bearing assets, SCF and estate-tax filings

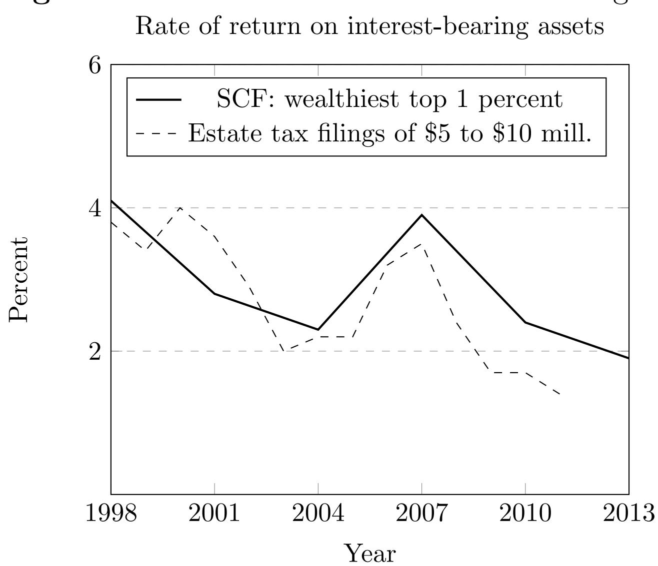


Figure 7: Wealth concentration with uncertainty

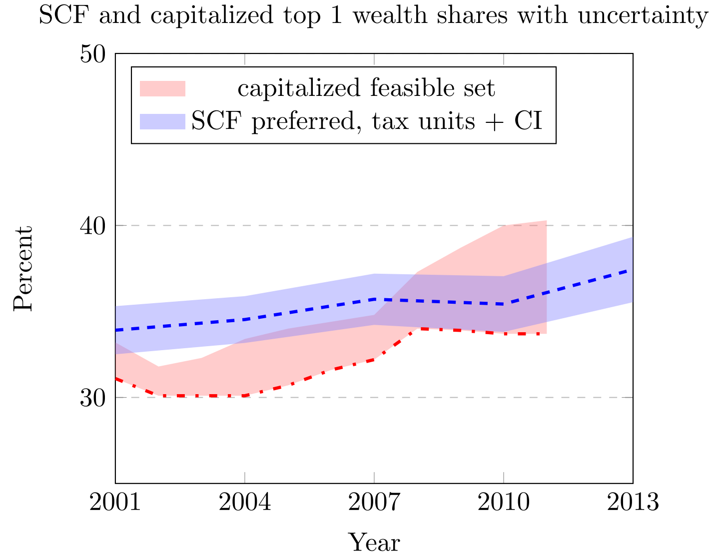
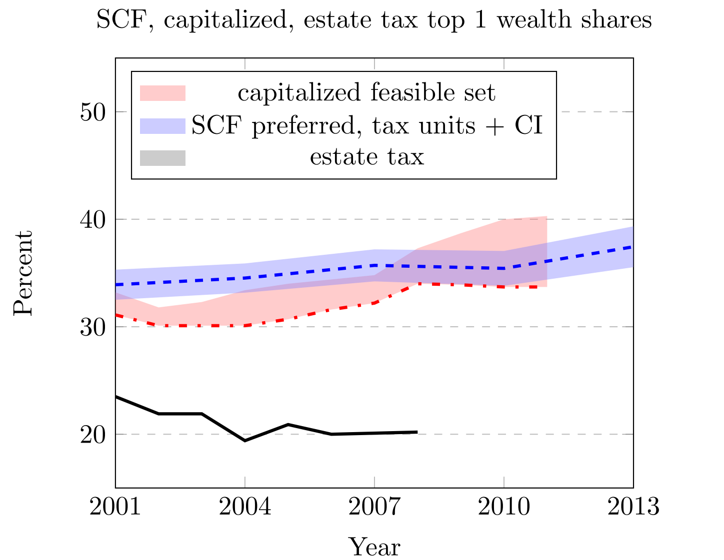


Figure 8: Wealth concentration across three sources



6. Appendix

6.1 Appendix A: Time series properties of earnings

The time series properties of household income deserve some additional consideration in this decision, too. Current income of family i in year t (y_{it}) is often modeled as a function of observables (z_{it}) and an unexplained stochastic term (u_{it}): $y_{it} = \beta z_{it} + u_{it}$. The unexplained term is often modeled as the sum of a persistent component (p_{it}) and a transitory innovation (e_{it}) to income: $u_{it} = p_{it} + e_{it}$. The transitory innovation has the form $e_{it} = \varepsilon_{it} + \delta_1 \varepsilon_{it-1} + \delta_2 \varepsilon_{it-2} + ...$ and the persistent component has the form $p_{it} = \rho_1 p_{it-1} + \mu_{it}$, where μ_{it} is a white noise error term.

If ρ_1=1 then the stochastic term of income is said to have a unit root (or to follow a random walk) and changes to income over time are governed by white noise (μ_{it}) and transitory changes from the lagged values of ε_{it}. If the correlation between current transitory innovations to income and past transitory innovations (the δ terms) dies out quickly then income can be described as having a unit root with a low order moving average (MA) component. Income changes are often modeled as a random walk with a low order MA component (Meghir and Pistaferri, 2004), and empirical support for this model with MA(2) has been found (Abowd and Card, 1989, MaCurdy, 1982).

If income follows a random walk then the most current set of income data are the most valuable, as these income contain all the information needed to predict future income. In this case, averaging multiple years of income would not help in our wealth estimation, and moving the sampling data back in time would also cost us a lot. However, more recent estimates of the time series properties of income indicate that $\rho_1 < 1$, meaning that income does not follow a random walk (Baker, 1997; DeBacker, Panousi, Ramnath, and Vidangos, 2013; Altonji, Smith and Vidangos, 2014).

Long time series of income are needed to tease out these time series properties: with 10 years of data, earnings follow a random walk (Abowd and Card, 1989, MaCurdy, 1982) but with 20 years of earnings data on the same people a random walk is not found (Baker, 1997). Our 4-year panel is likely not able to identify these time series properties. But DeBacker, et al. (2013) use a 23 year time series of SOI household income and find estimates of $\rho_1 < 1$, indicating a lack of random walk in household income.