

Estimating Top Income and Wealth Shares: Sensitivity to Data and Methods

By JESSE BRICKER, ALICE HENRIQUES, JACOB KRIMMEL AND JOHN SABELHAUS*

* Bricker, Board of Governors of the Federal Reserve, 20th & C St, NW, Washington, DC 20551 (e-mail: jesse.bricker@frb.org); Henriques: Board of Governors of the Federal Reserve, 20th & C St, NW, Washington, DC 20551 (e-mail: alice.m.henriques@frb.gov); Krimmel: The Wharton School, University of Pennsylvania, 1400 Steinberg Hall-Dietrich Hall, Philadelphia, PA 19104 (e-mail: jkrimmel@wharton.upenn.edu); Sabelhaus: Board of Governors of the Federal Reserve, 20th & C St, NW, Washington, DC 20551 (e-mail: john.e.sabelhaus@frb.gov). 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 made this research possible: Lisa Dettling, Sebastian Devlin-Foltz, Joanne Hsu, Lindsay Jacobs, Kevin B. Moore, Sarah Pack, Jeff Thompson, and Richard Windle. We also thank Jeff Larrimore, our conference discussant.

Estimates of top income and wealth shares in U.S. household survey data are generally lower than the estimates from U.S. administrative income tax data. However, these top share estimates are sensitive to the unit of analysis, the income concept being measured, and, in the case of wealth, to assumptions about the correlation between income and wealth. We constrain a household survey – the 2010 Survey of Consumer Finances (SCF) – to be conceptually comparable to tax records and are able to reconcile the cross-sectional difference between the survey and administrative estimates.

A survey with sample coverage at the top of the distribution – like the SCF – has

measurement advantages over the tax data. Surveys, in general, measure a wider concept of income than does the income tax data, and wealth estimates from the income tax data are heavily dependent on rates of return. Overall, top share estimates derived from income tax data generally overstate income and wealth concentration levels relative to SCF survey data.

I. Measurement: Methods and Data

Both household surveys and administrative data can be used to measure US income and wealth distributions. Comparisons between these are difficult, however, as these two sources of data often differ in the conceptual measurement of income and wealth and the population available for measurement. Each of these differences can impact distributional share estimates.

Typically, administrative tax data offer nearly universal coverage at the top of the distributions, as tax filing is virtually compulsory for those at the top. However, tax data are often missing information on those that do not have to file (nearly 20 million units in

recent years). And the unit of observation in tax records are *tax units*, an arbitrary unit defined by the tax code rather than by economic theory.

Survey data typically come from a random sample of *families*, a more economically-meaningful unit of observation. Many tax units may choose to reside within one family and pool economic resources.

Surveys, though, often suffer from low response and from measures of income and wealth that are too low at the top (Burkhauser et al., 2012; Bricker et al., 2015).

We use the 2010 SCF, which has an oversample of wealthy families and a weighting scheme that corrects for under-coverage at the top. Using sampling frame data, wealthy SCF non-respondents are observationally equivalent to wealthy SCF respondents even among the highest wealth families (Bricker et al., 2015).¹

A. Income measurement

The concept of income being measured differs between the survey data and tax data. Income in the tax data is conceptually limited by the information being collected for tax

purposes, while surveys can choose to include a more expansive set of income measures.

For instance, the income concept in the US tax system is narrower than that found in the SCF or in other sources, such as the National Income and Product Accounts (NIPA). Tax data may capture as little as 60 percent of 2010 NIPA Personal Income, and the SCF has typically collected about 15 percent more income than the tax aggregate in recent years (Bricker et al., 2015). The SCF, for example, collects information on transfer income that the tax system does not.

B. Wealth measurement

Estimates of the US wealth distribution can come from a household survey with a wealthy oversample (such as the SCF – Bricker et al., 2015), estate tax data (Kopczuk and Saez, 2004), or by capitalizing income tax data (Saez and Zucman, 2015). The concepts are generally similar across datasets – including all assets and debts – but measurement differs.

Household survey data.—The SCF survey directly measures assets and debt values by querying the family.² These survey measures of

¹ Because the SCF oversample is sampled from administrative records derived from tax returns, non-respondents can be compared to respondents. Kennickell and Woodburn (1999) describe the SCF weighting process.

² The SCF questionnaire asks detailed questions about the value of the family's assets held in housing, businesses, vehicles, other non-financial assets, financial transaction accounts, privately-held stocks,

bonds, mutual funds and trust accounts, as well as assets held in retirement accounts. Information of receipt of defined-benefit pension accounts is collected but not typically included in asset estimates. There are also detailed questions about the balances owed on mortgages, credit cards, lines of credit, household installment debts, and pension loans. The difference between the assets and debts is the SCF estimate of family net worth.

wealth compare favorably to external aggregates (Detting et al., 2015). Among cases in the oversample – for whom sampling frame data are known – the income and predicted wealth distribution of the responding families matches that of the non-responding families (Bricker et al., 2015).

Estate tax data.—Data on estate tax filers are the only direct measurement of wealth in the tax data; estate tax filings occur at death and for very few families. The SCF, then, can directly measure wealth for a much wider set of families than administrative tax data can. Estimates from estate data are rely on mortality models.

Income tax data.—Wealth can also be estimated from income tax data (Saez and Zucman, 2015, Kennickell and Woodburn, 1999, Greenwood, 1983). In contrast to survey data, these wealth estimates are measured indirectly with the help of a model that “capitalizes” taxable income into wealth, and calibrates to external wealth aggregates. These wealth estimates are sensitive to model parameters and to taxable income concepts.

The most straightforward capitalization model involves estimating financial wealth by capitalizing k types of capital income from the

tax data by a general rate of return on assets associated with that income (r_k), then adding an estimate of non-financial wealth.³ For family i :

$$(1) \widehat{wealth}_i = \widehat{nonfin}_i + \sum_{\forall k} \frac{income_{i,k}}{r_k}.$$

The choice of data and rates of return are the main inputs into the model. Estimates of non-financial wealth (mainly for housing and defined-benefit (DB) pensions) can be estimated by known aggregates such as the Financial Accounts of the United States (FA) (Saez and Zucman, 2015) or though survey data (Bricker et al., 2015).

Rates of return for income type k can be derived using the ratio of aggregate tax income of type k to its related concept in the FA data (Saez and Zucman, 2015). For example, the rate of return on interest income is the ratio of total interest income in the tax data to total stock of interest-bearing assets in the FA.

Alternatively, rates of return can be estimated using annual market rates of return. Wealth estimates derived from income tax data are heavily dependent on rates of return (Bricker et al., 2015; Kopczuk, 2015).

³ See Greenwood (1983), Bricker et al. (2015) and Saez and Zucman (2015) for three examples. In Saez and Zucman (2015) eight different types of capital income are included in the model. Bricker et

al. describe the oversampling model in the SCF, which is based on Greenwood (1983) and similar to Saez and Zucman (2015), though using nine income sources and market-based rates of return.

C. Top share measurement

Aside from differences in coverage and measurement, the differences in unit of measurement between income tax data and survey data impact the measurement of top income and wealth shares. Top share estimates using the income tax data in 2010 are based on 156 million *tax units*, but SCF estimates are based on 117 million *families*. Most tax units in the top 1 percent are families, but many families at lower percentiles are split into multiple tax units.⁴ If a tax unit is always a family at the top, then identifying the top 1 percent of tax units (1.56 million tax units) is equivalent to identifying roughly the top 1.3 percent of families. The unadjusted tax data, then, are predisposed to estimating more concentration at the top relative to a household survey.

The SCF provides good coverage of the entire wealth distribution up to the Forbes 400 families, which the SCF is precluded from sampling. Many families in the SCF are as wealthy as these families, though, so the SCF provides coverage even within the top 400. SCF top share estimates need to be augmented for the missing families, though.⁵

⁴ In the 2013 SCF, less than 3 percent of families in the top 1 percent had multiple tax units in a family, while nearly 20 percent of the bottom 99 percent of families had multiple tax units. The measure of 156 million tax units includes an estimate of nearly 20 million non-filers.

The tax unit issue is common to both income and wealth measurement. In the next two sections we describe these and other steps needed to reconcile survey and administrative estimates of top income and wealth shares.

II. Reconciling income concepts and measurement

Top income share estimates differ between the tax data and the SCF. The top 1 percent of families in the SCF held 17.3 percent of 2009 income, while the top 1 percent of tax units held 18.1 percent of total income in the administrative income tax data. Similarly, the top 0.1 percent of SCF families held 5.9 percent of 2009 income, while the top 0.1 percent of tax units held 8.3 percent of total income in the administrative income tax data (figure 1).

The concept of income, though, differs between the two data sources (Section IA). However, when the SCF household survey data are constrained to have the same income concepts, the top 1 percent and top 0.1 percent income shares increase to 19.6 and 7.1 percent, respectively (figure 1), making the SCF top 1 percent share slightly more concentrated than the tax data and eliminating much of the

⁵ Augmenting the SCF estimates to include missing Forbes 400 wealth typically add about 2 percentage points to the SCF top share estimates (figure 2), similar to Vermeulen (2016).

difference between the SCF and tax data estimates of the top 0.1 percent share.

[Insert Figure 1 Here]

Next, we adjust the SCF family level data to be comparable to a tax unit level in the tax data (see Section I) If the top 1.56 million SCF families are considered the top 1 percent, as in the tax data, then the SCF top 1 and top 0.1 estimates increase to 22 percent and 8 percent, respectively (figure 1). Thus, in comparable terms the SCF top 1 percent share is *more* concentrated than the tax data and nearly identical to the tax data estimates of the top 0.1 percent share.

III. Reconciling wealth concepts and measurement

Top wealth share estimates differ between the SCF and the capitalized income tax data (figure 2, first and third bars).⁶ The top 1 percent of families in the SCF held 34.5 percent of 2010 wealth, while the top 1 percent of tax units held 39.5 percent of total wealth in the capitalized income tax data. Similarly, the top

0.1 percent of SCF families held 13 percent of 2010 wealth, while the top 0.1 percent of tax units held about 21 percent of total wealth in the capitalized income tax data.

[Insert Figure 2 Here]

A. Rates of return in modeled wealth estimates

The estimates based on income tax data are sensitive to even small deviations in rates of return, and about 40 percent of the gap between SCF and income tax estimates are explained by a small change in the rate of return on interest-bearing assets. The 1 percent estimate drops to 37.3 percent and the top 0.1 percent estimate drops to 18 percent when a conservative market-based rate of return on interest-bearing assets – the 10-year Treasury rate – is used to capitalize interest income and all other rates of return are unchanged (figure 2, second bar).⁷

B. Reconciling SCF and modeled income tax wealth estimates

As in the case of income, the capitalized tax data effectively use different wealth concepts

⁶ The estate tax data estimates end in the year 2000 but the focus of this article is on recent wealth estimates. Thus, we concentrate on reconciling the 2010 SCF to the 2010 capitalized income tax estimates. The level of wealth concentration among the top 1 percent in the estate tax data (20 percent in the year 2000) is lower than in the either the SCF (33 percent in 2001) or the capitalized income tax data (33 percent in 2001). However, the time trend in the SCF estimates more closely resemble the time trend in the estate tax data.

⁷ The estimates in Saez and Zucman (2015) and presented as the light red bar in figure 2 use a 1.46 percent return on interest-bearing assets while the 10-year Treasury rate is about 3 percent. The average

AAA corporate bond rate for 2010 was about 4.5 percent. Kopczuk (2015) and Bricker et al. (2015) have noted that the interest-bearing asset rate of return in Saez and Zucman (2015) is different from what both estate tax data imply and what market rates imply. Why this is the case is unclear, but could be due to IRS reporting rules: families with less than \$10 in interest may not receive a 1099-INT from their financial institution and then may not claim this income on their tax filing, thus making the tax data have too little income for the assets in the FA. When rates of return are so close to zero, small changes in the rate can have large implications for a capitalization model in equation (1).

than the SCF. The capitalized income tax data are calibrated to match the FA household data, which includes an estimate of aggregate DB pension wealth. The SCF wealth estimate does not include DB pension wealth.

Further, the *composition* of assets differs somewhat between the SCF and the FA, even though aggregate asset totals are similar. Notably, the SCF estimate of housing wealth as a share of assets is larger than that in the FA. The wealth estimates from capitalized income in figure 2, then, bias down housing wealth – a key middle-class asset – relative to the SCF.

We adjust the SCF wealth *concept* to match the FA concept by including an estimate of DB pension wealth, which lowers the SCF top share. We also adjust SCF wealth to match the *values* in the FA, as is done in the income tax estimates, which raises the estimated top share. These changes mostly offset each other (figure 2, fourth bar).

Adjusting the SCF family level data to be comparable to a tax unit level in the tax data leads to larger top share estimates: 37.9 percent for the top 1 percent and 14 percent for the top 0.1 percent. The top 1 percent is now reconciled with the alternative capitalized income data, though the top 0.1 percent is still a few percentage points below (figure 2, fifth bar).

The SCF, recall, is precluded from sampling families in the Forbes 400. Incorporating an estimate of the wealth held by these families further increases top wealth shares in the SCF. Wealth share levels are about equal in the augmented SCF and capitalized income tax estimates (figure 2, sixth bar).

REFERENCES

- Bricker, Jesse, Alice Henriques, Jacob Krimmel, and John Sabelhaus. 2015. “Measuring Income and Wealth at the Top Using Administrative and Survey Data” *Brookings Papers on Economic Activity*, forthcoming.
- Burkhauser, Richard V., Shuaizhang Feng, Stephen P. Jenkins, and Jeff Larrimore. 2012. “Recent Trends in Top Income Shares in the United States: Reconciling Estimates from March CPS and IRS Tax Return Data.” *Review of Economics and Statistics*, Vol. 94, No. 2, pp. 371-388.
- Dettling, Lisa, Sebastian Devlin-Foltz, Jacob Krimmel, Sarah Pack, and Jeffrey Thompson. 2015. “Comparing Micro and Macro Sources for Household Accounts in the United States” FEDS Working Paper 2015-086.
- 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.
- Kopczuk, Wojciech. 2015. “What do We Know about the Evolution of Top Wealth Shares in the United States?” *Journal of Economic Perspectives*, Vol. 29, No. 1, pp. 47-66.
- Kopczuk, Wojciech, and Emmanuel Saez. 2004. “Top Wealth Shares in the United States, 1916-2000” Evidence from Estate Tax Returns,” *National Tax Journal*, Vol. 52, pp. 445-487.
- Piketty, Thomas, and Emmanuel Saez. 2003. “Income Inequality in the United States, 1913-1998,” *Quarterly Journal of Economics*, Vol. 118, No. 1, pp. 1-39.
- Saez, Emmanuel, and Gabriel Zucman. 2015. “Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data,” *Quarterly Journal of Economics*, *forthcoming*.
- Vermeulen, Philip. 2016. “How Fat is the Top Tail of the Wealth Distribution?” *American Economic Review* (*this edition*).

You may enter your figures on this page of the template. At publication, your figure must fit within the width of one column of this template (2 5/8 inches or 6.8 cm).



FIGURE 1. ESTIMATED SHARE OF INCOME HELD BY TOP PERCENTILES IN 2010

Note: Data for red bars drawn from administrative tax data for income earned in 2009. Data in blue bars drawn from 2009 income reports in the 2010 SCF. The light blue bar labeled “SCF (1)” represents the SCF estimate of the share of total income held by the top 1 or top 0.1 percent of families. The medium shaded blue bar labeled “SCF (2)” shows the top share estimate when the SCF income concepts are reconciled to the administrative data income concepts. The final dark blue bar labeled “SCF (3)” shows the top share estimate when the SCF estimate is further modified: from a household-level estimate to a tax-unit-based estimate. The dark line in the final bar shows a 95 percent confidence interval.

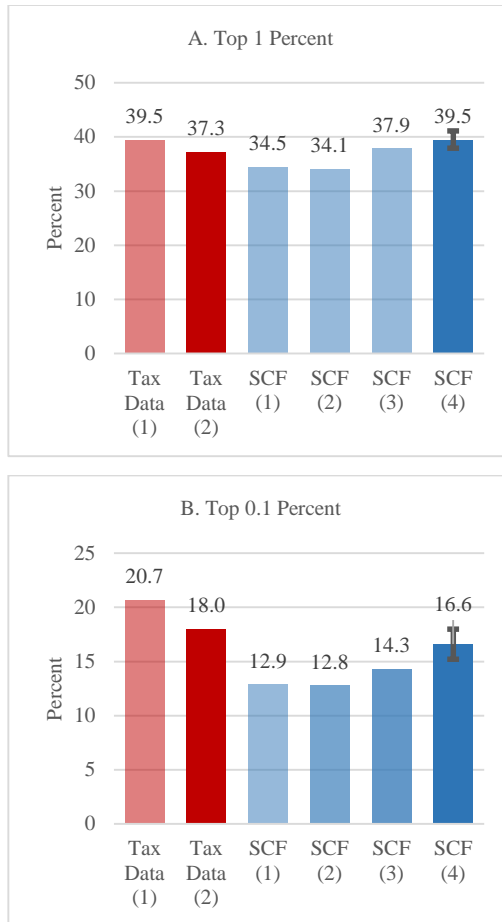


FIGURE 2. ESTIMATED SHARE OF WEALTH HELD BY TOP PERCENTILES IN 2010

Note: data for red bars drawn from administrative tax data for income earned in 2010 and capitalized and distributed as in Saez and Zucman (2015). Light red bar labeled “Tax Data (1)” uses the base rates of return in Saez and Zucman, where rates of return are based on the ratio of capital income flows – measured in the IRS Statistics of Income (SOI) data – to the stock of wealth measured in the Financial Accounts (FA) of the United States. Namely, taxable interest return is 1.46%, non-taxable interest is 5.09%, dividends 2.15%, rent and royalties 4.54%. Data in dark red labeled “Tax Data (2)” uses same rates of return *except* for interest, which uses 10-year Treasury yield (3.22% in 2010). Data in blue bars drawn from 2010 wealth reports in the 2010 SCF. The light blue bar labeled “SCF (1)” represents the SCF estimate of the share of wealth held by the top 1 or top 0.1 percent of families. The next shaded blue bar labeled “SCF (2)” shows the top share estimate when the SCF wealth concepts are reconciled to the Financial Accounts values and concepts (as in Saez and Zucman, 2015). The next blue bar labeled “SCF (3)” shows the top share estimate when the SCF estimate is further modified: from a household-level estimate to a tax-unit-based estimate. The final blue bar labeled “SCF (4)” shows the SCF estimate of top wealth shares when the further augmented to include an estimate of the wealth held by the Forbes 400, a group that the SCF is legally obligated to not sample. The dark line in the final bar shows a 95 percent confidence interval.