

Measuring Income and Wealth at the Top Using Administrative and Survey Data

Jesse Bricker^{1,2}
Alice Henriques¹
Jacob Krimmel¹
John Sabelhaus¹

April 2015

Abstract

Administrative tax data indicate that U.S. top income and wealth shares are substantial and increasing rapidly (Piketty and Saez 2003, Saez and Zucman 2014). A key reason for using administrative data to measure top shares is to overcome the under-representation of families at the very top that plagues most household surveys. However, using tax records alone restricts the unit of analysis for measuring economic resources, limits the concepts of income and wealth being measured, and imposes a rigid correlation between income and wealth. The Survey of Consumer Finances (SCF) solves the under-representation problem by *combining* administrative and survey data (Bricker et al, 2014). Administrative records are used to select the SCF sample and verify that high-end families are appropriately represented, and the survey is designed to measure comprehensive concepts of income and wealth at the family level. The SCF shows high and rising top income and wealth shares, as in the administrative tax data. However, unadjusted, the levels and growth based on administrative tax data alone appear to be substantially larger. By constraining the SCF to be conceptually comparable, we reconcile the differences, and show the extent to which restrictions and rigidities needed to estimate top income and wealth shares in the administrative data bias up levels and growth rates.

JEL Codes: D31, D63, H2

¹Board of Governors of the Federal Reserve System, Washington, DC. ²Corresponding author, jesse.bricker@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, Kevin B. Moore, Sarah Pack, Max Schmeiser, Jeff Thompson, and Richard Windle. We also thank Diana Hancock, Arthur Kennickell, Wojciech Kopczuk, Victor Rios-Rull, Emmanuel Saez, Gabriel Zucman, and seminar participants at the Federal Reserve Board, the Bank of England, the Bank of Spain, and the HFCN meeting at the European Central Bank for input and comments on earlier versions of this paper. The corresponding author also thanks Olympia Bover and the Bank of Spain for hospitality at the early stages of this paper. Finally, we are grateful to Michael Parisi for providing SOI income growth rate tabulations and to Barry Johnson and the SOI staff for contributions to the SCF sample design.

I. Introduction

Income and wealth are very concentrated in the United States, and top shares have been rising in recent decades, raising both normative and macroeconomic policy concerns. However, the levels and trends observed in household survey data are less dramatic than estimates derived directly from administrative income tax data. For example, Saez and Zucman (2014) use administrative income tax data and estimate that the top 1 percent (by wealth) had a wealth share of 42 percent in 2013, up from 29 percent in 1992. The Survey of Consumer Finances (SCF) shows less than half the increase in the top 1 percent wealth share, rising from 30 percent in 1992 to 36 percent in 2013.¹ Similarly, Piketty and Saez (2003, updated) show that the top 1 percent (by income) had a 23 percent income share in 2012, an increase of 10 percentage points since 1992. The SCF shows a 20 percent income share for the top 1 percent in 2012, an increase of 8 percentage points since 1991.²

In general, administrative data *should* provide better estimates of top income and wealth shares, because traditional random household surveys suffer from underrepresentation of wealthy families.³ Unlike most other household surveys, the SCF is designed to overcome the underrepresentation problem, because administrative data are used to select the sample, and rigorous targeting and accounting for wealthy family participation assures those families are properly represented in the survey data. Thus, the divergence in top income and wealth shares noted above is conceptual, and not attributable to sampling. In particular, we show that the divergence in top shares arises because the SCF uses more appropriate observational units for distributing resources, measures income and wealth more comprehensively, and avoids imposing a rigid correlation between cross-section income and wealth when selecting top end families.

The reconciliation exercise here shows that estimates of top shares derived from administrative data in isolation are inherently biased because of the underlying system parameters from which the data are derived. For example, in the U.S. income tax system,

¹ A slow rise in top wealth shares is also consistent with estimates derived from administrative *estate tax* data (Kopczuk and Saez, 2004).

² SCF income values are for the year preceding the survey.

³ See Sabelhaus, et al (2015) for direct estimates of the relationship between income and unit non-response. Burkhauser, Feng, Jenkins, and Larrimore (2012) show that at least some of the divergence between CPS and administrative incomes is also due to top-coding of very high incomes in the CPS. Attanasio, Hurst, and Pistaferri (2015) use household budget data to study inequality, and in addition to the non-response issues, they find that reporting problems further confound consumption-based inequality estimates. Atkinson, Piketty, and Saez (2011) use administrative data in their multi-national and longer-run view of rising income inequality.

observations are tax filing units, not families. The number of tax units (about 160 million in 2012) is some 30 percent higher than the number of families (122 million in the SCF).⁴ Most of the tax units at the very top are also families, meaning the effect of divergence between the observational units is mainly a factor in the rest of the income and wealth distribution.⁵ Thus, the implication 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.

Administrative income tax data also limits the income concept to what is currently taxable, leading many forms of income to go unmeasured. Although there are various conceptual ways to think about economic resources available to families, the taxable income concept is biased in terms of both levels and trends.⁶ Unmeasured compensation (such as employer provided health care, Social Security, or Medicare contributions) biases down incomes in the middle of the distribution proportionally more than the top, and the relatively rapid growth in untaxed incomes has led to a systematic upward bias in the growth of top shares in tax data. Indeed, we show that the income concept used by Piketty and Saez (2003, updated) fell from 74 percent of National Income and Product Account (NIPA) Personal Income in 1970 to 61 percent by 2012. Although the SCF also provides an incomplete picture of family income, the concept is closer to the NIPA concept than measures derived directly from the tax data.

Using income tax data to estimate top wealth shares introduces additional complications. The inferences about top wealth shares in Saez and Zucman (2014) are based on “capitalizing” wealth using observed capital incomes in the administrative tax data, estimated correlations from surveys (such as the SCF) for wealth components not associated with any taxable income, and decisions about benchmarking distributional estimates against published household sector aggregates from the Financial Accounts (FA). We show that these additional steps generally compound the bias introduced by observational unit and conceptual differences, and thus explains the even wider gap in top wealth shares. One particular driver of the capitalized wealth

⁴ The actual unit of observation in the SCF is the “Primary Economic Unit,” or PEU, which is somewhere between the Census “family” and “household” concepts. See the appendix to Bricker et al (2014) for a precise definition. The number of families in the SCF is benchmarked to that found in the Current Population Survey. The number of tax units includes an estimate of non-filers.

⁵ As discussed later in the paper, even when additional tax units (such as dependent children) do exist in high end families, subtracting those dependent filer incomes from total family income will not substantially alter the primary tax unit resources. Thus, for all intents and purposes, high end families are the same as high end tax units.

⁶ The evolving differences in the concept of income in administrative versus survey data is also emphasized by Burkhauser, Larrimore, and Simon (2012), and Armour et al (2014).

shares at the *very* top in recent years is the inexplicable divergence in the ratio of fixed income assets (in the FA) to taxable interest income (in the administrative tax data). Imposing a more reasonable market-based capitalization relationship between interest and assets lowers top wealth shares substantially.

The conceptual and measurement adjustments described above reconcile most of the divergence between administrative and SCF top share estimates, and thus provide direct estimates (by subtraction) of the bias in top shares derived directly from the administrative data. Identifying *which* families represent the very top helps to explain the residual top share differences, which show up in terms of both levels and volatility. The SCF sampling strategy uses two models to identify top-end families: the gross capitalization model (as in Saez and Zucman, 2014) and an empirical correlation model (based on previous SCF surveys). The SCF's two-pronged sampling strategy and the use of multiple years of administrative data makes it possible to distinguish the *permanently* wealthy from those families who happen to realize very high but transitory incomes in a given year. These two modeling approaches often disagree on predicted wealth rankings and wealth shares at the very top, and the gross capitalization model always predicts higher wealth shares at the top than empirical correlation model. That is, a modeling approach that differs from Saez and Zucman (2014) shows lower wealth shares and a slower wealth trend than even the unadjusted SCF, showing that heterogeneity in models can drastically affect the wealth share results.

The rest of the paper proceeds as follows. The next section describes how the administrative tax data are used to identify wealthy families in the SCF, and how the participant sample can be validated against non-participants in terms of income levels using only the administrative data. The third section focuses on reconciling income shares derived directly from the administrative data with those in the SCF, considering the conceptual differences and then the tax unit versus household effects. Similarly, the fourth section reconciles top wealth shares in the SCF with top shares constructed from the administrative data, with additional adjustments for benchmarking to aggregates and the Forbes 400. The fifth section describes how alternative approaches to identifying high wealth families in the administrative data helps to explain residual differences in the estimated top shares *within* the top 1 percent, based on controlling for transitory income fluctuations and targeting permanently wealth families. The sixth section concludes.

II. Sampling and Surveying Wealthy Families

Economic resources in the U.S. and other industrialized countries are highly concentrated. Measuring and explaining income and wealth concentration has challenged economists at least since Pareto (1896) and Kuznets (1953). Measuring top income and wealth shares using simple random sampling and 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. The Survey of Consumer Finances (SCF) overcomes both problems by oversampling at the top using administrative data derived from tax records, and by verifying that the top is represented using targeted response rates in several high end strata. The administrative data used in the sampling also show that SCF participants are observationally equivalent to non-participants within the high end strata.

SCF Sampling Strategy

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.⁷ The list sample is drawn using statistical records derived from tax returns at the Statistics of Income (SOI) Division of the Internal Revenue Service.⁸ 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 SCF 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 $i=1 \dots N$ 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 $\epsilon(0,1)$. There are six types of income in the SCF model: taxable

⁷ See O’Muircheartaigh et al. (2002) for more information about the NORC national sample.

⁸ 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 Compliance Data Warehouse (CDW) data, where the CDW data are the universe of tax filings (see Statistics of Income, 2012).

interest, 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).⁹

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:

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

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.¹⁰ 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 that are especially prevalent for capital incomes and at the top of the distribution. In contrast to the gross-capitalization model, one key difference is that 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.

The gross capitalization and empirical correlation models generate two independent sets of rankings, or wealth indices. The two wealth indices are blended and 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. The list sample is then selected by a probability proportional to size (PPS)

⁹ Model details are provided in Appendix A, including rates of return. Income is a weighted average of three years of sampling income. Saez and Zucman (2014) use a gross capitalization model to predict wealth from SOI income data, too. In their version, the rate of return for each capital asset type is defined by the ratio of SOI income for type of asset income to the stock of household (and non-profit) assets in the Financial Accounts for each asset type. The end result is that the Saez and Zucman (2014) gross capitalization method allocates wealth according to SOI income and predicted wealth will match the household (and non-profit) wealth in the Financial Accounts. The rates of return used in the SCF are similar to those used by Saez and Zucman (2014) with the exception of the return to interest-bearing assets. This difference is explored in Section IV.

¹⁰ 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.

sampling method, stratifying by the seven wealth strata, with the probability of selection increasing in each stratum.¹¹ 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.

Are SCF Participants Representative Within-Strata?

The sampling mechanism ensures that the oversampled families are representative of the underlying population in the administrative data. But when some families do not respond to the survey, representativeness is no longer guaranteed. That is, although the selected sample is representative *across* strata, differential participation *within* the strata could in principle still lead to biased top share estimates.¹² In this section we show that SCF respondents and sampled non-respondents are generally observationally-equivalent in terms of what can be measured using the administrative sampling data: overall income distributions, mean incomes, and income volatility.

The distributions of total incomes for SCF participants are similar to those of sampled non-respondents (Figure 1A). Moving from the fourth-highest stratum to the highest stratum, one sees the substantial non-linearity of incomes that characterize the top end. The range of incomes in the top four SCF strata completely cover the top 1 percent in an overlapping way, meaning, for example, that the top of the fourth-highest stratum overlaps with the bottom of the third highest stratum, and so on. The capital income distribution of SCF respondents are also similar to those of non-respondents (Figure 1B), and the non-linearity in incomes as one moves from the fourth-highest to the highest stratum is even more dramatic.¹³

¹¹ 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.

¹² See, for example, the discussion in Kennickell and Woodburn (1999).

¹³ Capital income here includes taxable and non-taxable interest, dividends, Schedule C and Schedule E business income, Schedule F farm income, and capital gains.

In general, statistical tests confirm the visual indication that participants and sampled non-participants within strata have very similar income distributions. The null hypothesis is that the two distributions come from the same underlying distribution, and the test statistics generally fail to reject the null using either the Kolmogorov-Smirnov or Wilcoxon Rank-Sum tests. The specific results vary by year and across strata, but in the 2013 sample, the null was rejected for only the second highest stratum for total income.¹⁴

Focusing on the middle of the distributions across strata, average total incomes for both participants and sampled non-participants in the fourth highest stratum are generally around \$500,000, whereas the average total incomes in the highest stratum are above \$50 million (Figure 2A, shown again on a log scale). The averages for total income versus capital income only differ noticeably for the fourth-highest and third-highest stratum (Figure 2B). In the top two stratum, average total income is dominated by and effectively equivalent to capital income. As with differences in the distributions, one can test for differences in the means by income measure, stratum, and year, and in general the tests fail to reject the null that the means for participants and sampled-non-participants are the same.¹⁵

Finally, in terms of observable pre-survey income volatility, SCF participants are also similar to non-respondents for both total income (Figure 3A) and capital income (Figure 3B). Income at the top is known to be much more volatile than in the rest of the income distribution, and the trend seems to be towards higher relative volatility at the top.¹⁶ In the SCF sampling data, for the top four strata covering the top 1 percent, about one-fourth of 2013 families experienced income changes below -50 percent or above + 50 percent. The similarity between SCF respondents and sampled non-respondents means that potential distortionary effects from sampling families with very high or very low transitory income shocks is controlled for using the SCF strategy.

¹⁴ Results across income concepts, strata, and for earlier years are available upon request.

¹⁵ In 2013, the differences for the second-highest stratum were significant at the 5 percent level. Again, results for other years, income measures, and stratum are available upon request.

¹⁶ See, for example, DeBacker et al, (2013), Guvenen, Kaplan, and Song (2014), and Parker and Vissing-Jorgenson (2010).

III. Reconciling Administrative and Survey-Based Top Income Shares

Random-sample surveys such as the Census Bureau's Current Population Survey (CPS) show rising top income shares in recent decades.¹⁷ However, estimates derived directly from administrative tax data suggest that top income shares in surveys such as the CPS are biased down, in large part because of low survey participation rates at the high end.¹⁸ The SCF overcomes the high end unit non-response problem in traditional random-sample household surveys by oversampling at the top of the distribution using administrative data derived from tax records. However, top income shares in the SCF, especially for capital income and at the very top, still diverge from the direct estimates. Systematically reconciling the divergence in top income shares involves focusing on conceptual differences in income measures and households versus tax returns as the unit of observation.

Reconciling Income Concepts

Estimates of top income shares derived from administrative tax data are conceptually limited by the information being collected for tax purposes. The concept of income in the U.S. income tax system is narrower than Personal Income in the National Income and Product Accounts (NIPA), and the gap between the two is growing. The SCF comes closer to the NIPA concept, which helps to explain a substantial part of the gap between estimated top income shares, because a disproportionate share of the missing income in the tax data is received by families outside the top groups.¹⁹

The aggregate income measure used as the base for estimating top income shares in Piketty and Saez (2003 plus updates) is well below aggregate NIPA Personal Income, and the gap has grown over time (Figure 4). The gap is attributable to both conceptual and measurement differences. On the conceptual side, NIPA includes non-taxable compensation paid by employers, government transfers, and retirement saving, none of which is recorded in the administrative tax data. On the measurement side, NIPA builds in estimates of non-compliance, especially for proprietor's income. There are also differences like economic versus accounting profits and imputed rents, that are part conceptual and part measurement. On net, the conceptual

¹⁷ Atkinson, Piketty, and Saez (2011) provide a multi-national and longer-run view of rising income inequality.

¹⁸ Burkhauser, Feng, Jenkins, and Larrimore (2012) show that at least some of the divergence between CPS and administrative incomes is also due to top-coding of very high incomes in the CPS.

¹⁹ Appendix B provides a detailed reconciliation of NIPA, SCF, and administrative tax data income concepts.

and measurement differences drive a nearly 40 percent wedge between the administrative tax data and NIPA Personal Income in 2012 (Figure 4, red line).

Taxable retirement benefits and retirement account withdrawals are measured in the administrative tax data, but are not included in NIPA income, because benefits received are not payments associated with current production.²⁰ Including taxable retirement benefits helps push the administrative tax income concept closer to NIPA, but the overall effect of tax-preferred retirement saving is still pushing down taxable income relative to NIPA. The tax data does not capture new contributions to retirement funds or the interest and dividends earned on those funds. Current benefits paid are well below contributions and income received by retirement funds, which is just another way of saying that retirement funds are substantial net savers, leading to a net increase in the overall gap.

The effect of conceptual differences between the administrative tax data and the NIPA accounts for about half the difference in aggregate totals, as the administrative incomes are about 80 percent of the adjusted NIPA totals in 2012 (Figure 4, blue line). The remaining divergence is mostly measurement-related, though inter-mixed with some remaining conceptual divergence, such as the accounting treatment of various business incomes (depreciation and inventory valuation) and items like imputed rent on owner-occupied housing and the imputed value of financial services.

The NIPA does not count capital gains, as with taxable retirement benefits, because gains are not associated with current production. However, capital gains do show up in the administrative tax data in the year that the gains are realized. Indeed, adding capital gains to the tax data (dotted lines in Figure 4) does raise taxable income noticeably, but in no sense closes the gap in either the unadjusted or adjusted comparisons.²¹

In addition to noting the substantial divergence between NIPA and taxable incomes at each point in time, it is also useful to note that the unmeasured part of income is growing steadily over time. The ratio of aggregate income as derived from administrative tax data in Piketty and

²⁰ In principle one could include Social Security benefits in administrative top income share estimates, at least for tax filers, but in practice only the taxable part of Social Security was well captured by the tax system, until recently. Prior to changes made to instructions in the past decade, there was a dramatic underreporting of non-taxable Social Security benefits, because filers knew the benefits were not taxable, and it was not clear that they were required to enter the amounts in the “A” column of Form 1040.

²¹ The difference between NIPA and tax accounting may shift some economic profits (from a NIPA perspective) into capital gains that are realized in a later year (from a tax perspective).

Saez (2003, plus updates) to NIPA income fell from nearly 75 percent in 1970 to the 60 percent value observed in 2012. Comparison with the adjusted NIPA measure (available starting in 1984) shows less trend, meaning much of the *growing* gap is attributable to the simple conceptual adjustments, that is, employer-provided benefits, government transfers, and net retirement saving.

The SCF captures an income measure that is somewhere between the taxable and NIPA concepts, reflected in the fact that the SCF ‘Bulletin’ income aggregate is always above the taxable aggregate, and that ratio is generally higher in more recent surveys.²² For the past three surveys, the SCF has collected about 15 percent more income than in the administrative tax data (not shown). When one subtracts non-taxable interest, workers’ compensation, and transfers, in order to move towards taxable or ‘Market’ income, the SCF aggregate falls to about 7 percent above aggregate taxable in 2012. These SCF aggregates are estimated with sampling variability, and the series are somewhat noisy (especially the 1988 and 2002 values) but there is a general pattern of rising SCF relative to tax data incomes since the 1990s, commensurate with the decline in taxable relative to NIPA. Thus, the conceptual wedge between SCF Bulletin and taxable, or Market, income can contribute to both the level and trend divergence in top share estimates.

Reconciling Top Share Income Estimates

Differences in income concepts are one source of divergence in top income share estimates, with the unit of observation also contributing. Using these factors, one can reconcile most of the differences in levels and trends between administrative data and SCF top shares, though the extent of the final reconciliation varies between the top 1 percent and top 0.1 income shares. More importantly, the nature of the reconciliation helps to understand why the administrative estimates of top shares are biased upward.

Beginning with the most widely cited number—incomes of the top 1 percent—both the administrative data and the SCF show high and rising top shares (Figure 5). The administrative data (solid black line) shows a 10 percentage point increase in the top 1 percent income share between 1991 and 2012, from 13 to 23 percent. During the same period the SCF Bulletin concept

²² ‘Bulletin’ income derives its name from the fact that this is the consistent series published in the *Federal Reserve Bulletin* after each triennial survey. For the most recent survey, see Bricker, et al, (2014).

(dotted red line) shows an increase from 12 to 20 percent, a similar but smaller increase than what is observed in the administrative data.²³

Adjusting the SCF for conceptual differences, meaning subtracting the (mostly government transfer) incomes not measured in the tax data, pushes the SCF share up in every year, as expected (Figure 5, dotted blue line). This adjustment closes much of the gap in the top share estimates for most years. Indeed, it is difficult to distinguish the top 1 percent income share lines after the conceptual adjustments to the SCF are introduced.

The second step in reconciling top income shares involves adjusting for tax units. In the U.S. income tax system, observations are tax filing units, not families or households, and the number of tax units (about 160 million in 2012) is some 30 percent higher than the number of households (122 million in the SCF). Most of the tax units at the very top are effectively households, because even when other tax units (such as dependent children) exist within those households, removing those units would not substantially alter the resources of the primary observational unit. The only situation in which bias (relative to tax data) arises for top shares is when two high end but separate tax units are observed together in the same household. The implication is that working backwards from the top of the distribution, the sum of incomes for 1 percent of administrative tax units will be larger than the sum of incomes for 1 percent of SCF households.²⁴

It is not possible with available data to reassemble tax units into households using administrative records, but it is possible using the SCF to get a sense of how much bias is introduced into top shares by restricting the unit of observation. The reconciliation involves re-computing SCF top share fractile thresholds using the same count of units as in the administrative data. The new top shares will include more units, 1.6 million instead of 1.2 million in 2012 for the top “1” percent, and thus the income share of that group is higher.

The effect of adjusting for tax units versus households (in addition to the adjustment for Market versus Bulletin incomes) pushes the SCF top 1 percent income shares (just) above the administrative data top shares in every year (purple dotted line in Figure 5). The tax unit

²³ The SCF data make it possible to go back to 1988 with a conceptually consistent time series, but sampling and processing differences for the first survey add a substantial amount of variation around those estimates, so we choose 1991 as the reference point for comparing trends. The differences in the 1989 survey are reflected in the estimated confidence intervals shown later.

²⁴ Appendix C shows the fractile thresholds for total income, capital income, and wealth in administrative and SCF data.

adjustment is not perfect, and because there are at least a few multiple high end units in the same SCF households, the adjustment knowingly introduces some upward bias in the top shares. However, that effect is very likely to be modest.

To be clear, the adjustments to the top shares here are not intended to suggest that the SCF agrees with the administrative tax data that the top 1 percent received just over 23 percent of income in 2012. The SCF value of 20 percent is the preferred estimate for the top 1 percent. The takeaway message is that the three percentage point upward bias in the 1991-2012 growth of administrative estimates can be explained by income concepts and by use of tax units instead of households.

Estimates derived from the SCF have sampling variability, and one can draw confidence intervals around the estimated top shares for any given fractile using the replicate weights made available for every SCF.²⁵ In the case of top 1 percent shares (Panel A of Figure 6), the range for the reconciled shares generally encompasses the point estimates from the administrative data. The only exceptions are in years when the tax data suggest large declines in top shares (2003 and 2009). This pattern is the first evidence of a theme that emerges when looking at higher fractiles, and fully explored below in Section V. The SCF sampling strategy is more likely to capture the incomes and wealth of the *permanently* wealthy, and thus the estimates are not (as) distorted by transitory income fluctuations.

A theme that emerges in work using administrative data is that the very top is growing much faster than even the top 1 percent. Estimates of top 1 percent income shares, decomposed into the top 0.1 percent and the “remaining 1 percent”, excluding the top 0.1, are shown in Figure 6 along with confidence intervals estimated for the final reconciled shares.²⁶ Panel A repeats Figure 5 but includes confidence intervals. Each of the unadjusted SCF, adjusted SCF, and administrative data estimates show slow growth in the income share held by the remaining 1 percent, excluding the top 0.1 (Figure 6, Panel B). For this sub-share, the SCF estimates are slightly larger than those from the administrative data. One explanation for this result is that by selecting households with a more permanent measure of wealth, we are more apt to capture a more permanent measure of income, less distorted by a transitory income shock.

²⁵ Appendix D provides details about how the confidence intervals for fractile shares are constructed.

²⁶ Appendix E shows the point estimates for the top 10 percent and top 0.1 percent fractile shares at each stage of the reconciliation.

The reconciled top 0.1 percent total income shares are also close to the administrative data (Panel C), and except for 2006, the direct tax-based estimates lie within the adjusted SCF confidence intervals. The administrative data shows overall more volatility in top 0.1 percent shares, especially in the 2000 and 2006 levels. The width of the confidence interval for the 2000 estimates speaks clearly to the point about transitory income fluctuations, as the top share estimates in years of volatile (especially capital) incomes has a direct impact on estimated top shares, regardless of data source.

Reconciling Top Capital Income Share Estimates

Conceptual differences between the administrative data and the SCF should not be an issue when comparing capital incomes, though there is some possibility of differences in whether income is recorded as labor income or capital income (for example, realization of stock options). The tax unit adjustment for capital income shares is still relevant, of course, and has the expected effect of reconciling much of the gap between the SCF and administrative data top 1 percent capital income shares in recent surveys (Figure 7). Indeed, the remaining noticeable divergence in top capital income shares goes in the other direction (SCF is *above* the administrative estimates) and the differences are most pronounced in the earlier years of the survey.

The difference in growth is clear when the top 1 percent is decomposed into the top 0.1 percent and the top 1, excluding the top 0.1 (Figure 8, Panels B and C, respectively). The patterns for these top fractiles of capital income generally mirror the results for total incomes in Figure 6, but the tax unit adjustment does not close nearly as much of the gap for the top 0.1 percent gap. This builds on the theme mentioned above and explored in Section V. The SCF sampling strategy is focused on identifying the *permanently* wealthy, and the top 0.1 percent (based on capital income) in any given year in the tax data will be disproportionately populated by the part of the high wealth population who happened to realize very large (but transitory) capital incomes in that year.

The uncertainty about the top 1 percent capital income shares in the early years of the SCF is substantial, as captured in the estimated confidence intervals (Figure 8, Panel A). The significantly wider confidence intervals for the fractile share estimates in the early years are associated with very high values and weights for a relatively small number of observations, and the effect of not including those observations (or bumping their importance relative to the point

estimates in the replicate weighting) is dramatic.²⁷ These issues are largely resolved in the more recent surveys, and there has been little or no trend differential in the reconciled top 0.1 percent capital income share since at least 1997.

Implications of Expanding the Income Concept

The divergence between top share estimates in the SCF and the estimates derived directly from administrative tax data are largely explained by the broader concept of income and household versus tax unit adjustment. In principle, this suggests that one could start with the SCF, add back in the other missing pieces of income, and construct overall superior estimates of top shares and even overall inequality.²⁸ Comprehensively distributing the missing pieces of NIPA Personal Income is beyond the scope of this paper, but it is clear what effect that would have on estimated top shares. The missing income in the administrative data are employer-provided benefits, retirement saving inside and outside Social Security, government transfers, and tax system non-compliance. Those missing incomes are not concentrated at the top like market incomes, and some pieces (government transfers) are primarily received outside the top fractile groups.

This motivates a simple thought experiment using the administrative and NIPA aggregate incomes, along with the estimated fractile shares from administrative data. The administrative data are always missing income relative to NIPA, and the gap is widening (Figure 4). If the missing income is allocated more evenly than the measured incomes, top shares will fall, and possibly quite substantially. Indeed, if *all* of the missing income is allocated equally, the income share of the top 1 percent in 2012 (excluding capital gains) falls from 19 percent to 12 percent, and the growth in the top share after 1970 is reduced from 12 percentage points to 6 percentage points (Figure 9). The thought experiment also draws attention to the large jump in top income shares associated with the Tax Reform Act of 1986, which reinforces the idea that the concept of income being measured is paramount in analysis of trends over time. Allocating all of the

²⁷ On-going SCF work is focused on the role that early case editing practices and weighting may be having on the point estimates for top shares. In the 1989 and 1992 surveys, there are apparent anomalies in the aggregated values for certain capital incomes (relative to known benchmarks) that may be inflating the top share estimates in those years.

²⁸ Expanding income concepts is a theme of Burkhauser, Larrimore, and Simon (2012), Smeeding and Thompson (2011), and others.

missing income equally across tax units would almost certainly bias top shares down, but it is clear that the potential impact of expanding the income measure is substantial.

IV. Reconciling Administrative and Survey-Based Top Wealth Shares

The only wealth tax that exists in the U.S. is an estate tax applied at death, and there is no administrative data system directly associated with measuring the cross-section of wealth at a point in time. Wealth concentration estimated using estate tax data produce show that wealth concentration at the top is increasing slowly, similar to the SCF (Kopczuk and Saez, 2004), though the level of wealth concentration at the top is lower. The inferences about top wealth shares in Saez and Zucman (2014), show more rapidly growing wealth shares at the top. These estimates are based on capitalizing incomes in the administrative tax data, estimating correlations between income and wealth, and benchmarking micro estimates against published household sector aggregates from the Financial Accounts (FA). The unadjusted differentials between capitalized administrative-based and Survey of Consumer Finances (SCF) top wealth shares are even larger than the unadjusted differentials for top income shares. Rising top shares found in the capitalized income data is due to growth in the top 0.1 percent.

As with reconciling top income shares, the limitations imposed by the available administrative data have a substantial impact in terms of both levels and trends for top shares. The observational unit (households versus tax units) differential is effectively the same as with income shares. Differences in wealth concepts involve decisions about what balance sheet items to measure, how to measure those, and whether to benchmark to FA or SCF aggregates. In the case of wealth, there is also an additional sorting effect, as imposing a rigid correlation between income and wealth implies that the highest (inferred) wealth families in a given year are those who realized the highest capital incomes in that year, not necessarily the *permanently* wealthy. Differences between capitalized estimates and the SCF estimates at the top can be reconciled by imposing these restrictions on the SCF, although we do not impose the sorting that results from simply relying on capital income to rank households.

Reconciling Wealth Concepts

Top wealth shares are measured indirectly in the administrative data. Such wealth estimates allocate aggregate household sector net worth in FA based on capital incomes from the

tax data and correlations between income and wealth from other data sources, including the SCF. The FA concept of household net worth does not measure purely household wealth, as it also includes, for example, the assets and liabilities of non-profit organizations. There are also several other adjustments to both FA and SCF concepts needed to establish conceptual equivalence.²⁹ As with the income reconciliation, the adjustments applied to SCF balance sheet categories affect top shares, because changes like subtracting durable goods and adding Defined Benefit (DB) pensions have differential effects across the wealth distribution.

The relationship between reconciled SCF and FA wealth aggregates has evolved over time. Over the past five SCF surveys, the SCF aggregates run between 10 to 20 percent above the FA measures. Given the sorts of fluctuations that have occurred in asset prices in recent years, the fact that the two very different sets of aggregates move together over the past several surveys is reassuring. The recent patterns of SCF relative to FA aggregates are somewhat different than in the early years of the SCF. In the 1989 through 1998 surveys the aggregate value of SCF net worth was roughly the same as FA net worth.

The change in the relationship between SCF and FA aggregates between the early SCF years and recent years is largely driven by differences in the valuation of owner-occupied housing (Figure 10).³⁰ Aggregate household sector liabilities in the SCF have remained fairly steady at around 90 percent of FA liabilities.³¹ Non-housing assets (since 2000) have moved in a fairly tight range above the comparable FA measure, despite large swings in the values of assets such as stocks, privately-held businesses, and commercial real estate. The trend differences in housing values are a first-order issue for top shares, however, because housing is a substantial part of aggregate household sector net worth, and the distribution is much less skewed than is the distribution for financial assets. Thus, the decision about whether to infer top shares by calibrating to SCF or FA aggregates will directly impact top shares.

²⁹ Appendix B provides a detailed discussion of how the reconciled wealth concepts are created from the published FA and underlying SCF measures. See also Henriques and Hsu (2014).

³⁰ The differences in SCF and FA housing stock valuations are driven by the very different methodological approaches. In the aggregate FA data, the housing stock is valued using a perpetual inventory that involves new investment, depreciation, and a national house prices index. In the SCF, house values are owner-reported. Hsu and Henriques (2014) discuss how house values in the SCF compare favorable to other micro-based estimates, such as the American Housing Survey, and Henriques (2013) provides evidence that SCF respondent house valuations generally track local area house price indexes quite well.

³¹ See Brown, et al., (2011) for a discussion of how SCF debt by category tracks relative to Equifax.

Reconciling Top Wealth Share Estimates

The reconciliation of SCF and income tax-based top wealth shares proceeds in much the same way as the income reconciliations, but there are additional steps associated with benchmarking SCF to FA values and the decision about how to treat the Forbes 400. The reconciliation involves five steps (Figure 11). The starting point involves ranking all SCF households by published SCF net worth and computing the share held by each top fractile, and is shown (as in the income reconciliation figures) by the dotted red line. In the most recent survey, the published estimate of the top 1 percent SCF net worth share is about 6 percentage points lower than the administrative estimates from Saez and Zucman (2014), shown by the solid black line.

The first adjustment involves moving to a FA wealth equivalent concept, as described in Appendix B, and indicated by the dotted orange line in Figure 11.³² These top wealth shares are still estimated using SCF households as the unit of observation. This adjustment alone, however, significantly reduces the share of wealth going to the top fractiles, as the FA conceptual adjustment subtracts a few small categories like consumer durables, but more importantly adds trillions of dollars in DB pension wealth, which is more heavily concentrated below the top share cut offs.

The second adjustment accounts for the observed divergence in FA versus SCF balance sheet aggregates by type, as shown in Figure 10. SCF values for owner-occupied real estate, non-housing assets, and liabilities are re-scaled separately to match their FA counterparts, and the effect is to increase the estimated SCF top share (the dotted blue line in Figure 11). The differential rescaling is important, because the divergence in owner-occupied housing aggregates implies that benchmarking administrative data to FA instead of the SCF *lowers* wealth more *below* the top fractiles than *above*. That is, owner-occupied housing is less concentrated at the top than other forms of wealth like financial assets and non-corporate businesses. Also, as shown by the widening of the gap between the orange and blue lines in Figure 11, the fact that house values in SCF have risen relative to FA suggests that the impact of benchmarking imparts a positive bias on administrative top shares estimates and has grown over time.

³² Primarily, we subtract vehicles, miscellaneous financial and nonfinancial assets, cash value of whole life insurance, and miscellaneous debt from the standard SCF balance sheet, then add back in DB pension wealth using an approach suggested by Saez and Zucman (2014).

The third adjustment involves shifting the top fractile cutoffs to account for tax returns versus households as the unit of observation.³³ The adjustment, shown in the green line in Figure 11, is quite dramatic, changes the top 1 percent wealth share substantially, and closes most of the remaining gap between SCF and tax-based estimates. The impact of moving from tax returns to families is disproportionately larger than for the income reconciliation, which reflects the shape of the wealth density (relative to the income density) around the threshold cutoffs.

The final reconciliation adjustment is for the Forbes 400. In 2013, the published Forbes 400 wealth is \$2.021 trillion, accounting for approximately 3 percent of aggregate household sector net worth. The SCF sampling explicitly prohibits interviewing the Forbes 400. They are removed from the sample before the field work begins, and the top stratum is adjusted accordingly. There is some chance, of course, that families with net worth above the Forbes 400 cutoff can make it into the sample, because of evolving wealth or types of wealth (private holdings) that Forbes may not know about.³⁴

As with the income reconciliations, the adjustments applied to resolve differences between the alternative approaches are meant to be informative, but not to suggest that the SCF household-based comprehensive wealth share estimates are somehow deficient. The preferred view of the top 1 percent share of 2013 is the baseline SCF value, though one could argue conceptually for shifting that estimate down to account for DB pensions, and shifting it back up (by about the same amount) to account for the Forbes 400.

The reconciliations for top wealth shares, like those for top income shares, close much of the gap between SCF and administrative tax-based estimates for the top 1 percent, though the tax-based estimates still appear to grow more rapidly (Figure 12, Panel A). However, decomposing the top 1 percent growth, the two data series generally grow at a consistent pace when the top 0.1 percent are omitted from the top 1 percent (Figure 12, Panel B), and the entire difference in growth at the top is due to the top 0.1 percent (Figure 12, Panel C). Similar to the income top share results, the SCF wealth share for the 99th-99.9th percentiles is slightly higher

³³ At this point, because we still exclude the Forbes 400 from our household wealth measure, we subtract 400 from the number of total Piketty-Saez tax units. The reduction of 400 tax units has a miniscule effect on the number of households within top wealth groups and a negligible effect on top wealth shares. When adjusting on a tax unit basis, we use the number of tax units in the year prior to the survey. Sensitivity analysis using number of tax units in survey year shows no significant differences in wealth share level or trend.

³⁴ See also Vermeulen (2014) for a discussion of using Forbes data in conjunction with survey data to re-estimate top wealth shares.

than the estimates from administrative data, likely stemming from classifying households using a more permanent wealth measure.

The remaining unexplained residuals in both levels and trends at the very top are largely related to a difference in asset measurement. Differences in the trend for the top 0.1 share are especially evident in the past decade (Figure 12, Panel C). The next section reconciles this divergence.

Reconciling the divergence in top wealth shares in the past decade

The asset composition of top-end families in the capitalized income data differs from the composition of the top-end families in the SCF (Figure 13). These differences are especially clear since the early 2000s, when the estimated fraction of top share wealth in fixed income assets began to grow substantially. However, the rate of return on fixed-income assets in Saez and Zucman (2014) compares poorly to market rates of return, suggesting interest income in the administrative data is being capitalized at too high of a rate. We show that a slight modification to this rate of return greatly affects the trends in top wealth shares.

About one-fifth of the assets held by the top 0.1 percent of SCF families are fixed-income assets, consisting mostly of bonds, CDs, call accounts, money market accounts, and other savings instruments (Figure 13, Panel A). Businesses and corporate equities comprise the majority of assets held by these families.

In wealth estimates derived from capitalized administrative income data, historically about one-quarter to one-third of assets held by the top 0.1 percent have been fixed-income assets. Notably, though, this composition shifted: for the past decade, nearly half of all assets held by the top 0.1 percent in the capitalized distribution are fixed-income assets (Figure 13, Panel B). Further, nearly all of the growth in recent years in the capitalized top 0.1 share arises from the growth in fixed-income assets.

In a gross capitalization model, assets are predicted by inflating asset income by an asset-specific “capitalization factor,” which is typically the inverse of an asset-specific rate of return. The capitalization factor that generates the black line in Figure 12, for example, is computed from the inverse of the ratio of aggregate SOI asset income to the stock of assets in the Financial Accounts. An alternative capitalization factor is the inverse of a market rate of return. For example, interest *income* can be capitalized into fixed-income *assets* by the inverse of the

Moody's AAA corporate bond yield, the inverse of a 10-year Treasury yield, or, as in Figure 12, by the ratio of FA fixed-income assets to SOI interest income.³⁵

The capitalization factors for these three rates of return can vary dramatically (Figure 14). In 2011, for example, the capitalization factor on interest income is 97 when using the ratio of Financial Accounts fixed-income assets to SOI interest income, 56 when using the 10-year Treasury yield, and 27 when using the Moody's AAA rate of return. In other words, the amount of household wealth implied by interest income could vary by nearly a factor of 4 depending on the capitalization factor. Note that a small change in the rate of return can have a large impact on these capitalization factors: a roughly 1 percent rate of return in the ratio generates the capitalization factor of 97, while a roughly 2 percent return on the 10-year Treasury generates the factor of 56.³⁶

One capitalization factor is applied to all families in the gross capitalization model. Top wealth shares, though, will be smaller if the top wealth families get a higher-than-average rate of return.³⁷ The dashed black line in Figure 15 shows the capitalized wealth share of the top 0.1 percent if the interest income of the top 1 percent is capitalized at the 10-year Treasury rate (the blue line in Figure 14) and the interest income of the bottom 99 percent is capitalized, as in Figure 12, by the ratio of Financial Accounts fixed-income assets to SOI interest income (the red line in Figure 14).³⁸ For comparison, the solid black line from Figure 12 is also displayed.

Notably, the dashed black line generally falls *within* the confidence interval of the SCF top 0.1 wealth share. Further, the growth rate since the year 2000 in the dashed black line is muted relative to the black line and is generally consistent with the SCF trend.³⁹ Thus, the level

³⁵ The gross-capitalization model used in the SCF sampling exercise (Section II) uses the Moody's AAA rate to capitalize SOI interest income.

³⁶ The bond series in the B.101 table of the FA has also been revised downward substantially from the mid-1990s to the current date since the Saez and Zucman (2014) capitalization factors were computed. This revision will shift the red line downward toward the x-axis.

³⁷ The rate of return on assets appears to vary across the wealth distribution in the SCF. In the 2013 SCF, the average rate of return on fixed-income assets (found by the ratio of SCF interest income to SCF fixed-income assets) across all households is about 1 percent, but the average rate of return for the top 1 percent of families is almost 6 percent. The exercise described in this paragraph and in figure 15 assumes only a 2 percent rate of return for the top 1 percent.

³⁸ The dashed black line in this exercise is found in Appendix Table B40 in Saez and Zucman (2014). In the black line in figure 12 and 15, both the top 1 percent and bottom 99 percent are assumed to have the same rate of return on fixed-income assets (the ratio of Financial Accounts fixed-income assets to SOI interest income – the red line in figure 14).

³⁹ This is especially true if the capitalized wealth value found for 2012 is an outlier in the trend. Many wealthy families timed their capital income to be realized in tax year 2012 – before tax rates went up in 2013 – suggesting that the 2012 value may be an outlier (Wolfers, 2015).

and growth of the top 0.1 wealth share in the capitalized administrative data is nearly identical to that observed in the SCF when conceptual differences between the two are alleviated, and one heterogeneous return is allowed.

The composition of fixed-income assets held by the top 0.1 percent in the capitalized model (Figure 13, Panel B) appears to be at odds with the detailed FA tables. Of the total amount of fixed-income assets held by the households sector in the FA (Table B.101), about one-third are held in bonds (\$4.201 trillion in 2012) and two-thirds are held in time and saving deposits (\$7.194 trillion in 2012). The top 0.1 percent, then, would hold roughly \$4 trillion in time and savings deposits in 2012 and, presumably, these savings and deposit accounts held by the top 0.1 percent would have large balances. But Table L.205 in the FA indicates that only \$1.659 trillion of time and savings deposits are held in large balances (amounts greater than \$100,000). These large balance accounts have also not grown recently. Further, the aggregate household stock of bonds in the FA has fallen from \$4.916 trillion in 2010 and the stock of time and saving deposits has grown from \$6.450 trillion in 2010. Thus, the growth observed in fixed-income assets in the FA is due to small balance time and saving deposits, which are unlikely to be held by top-end families (or to be the source of their wealth growth).⁴⁰

V. Alternative Approaches to Identifying Wealthy Families in Administrative Data

Adjusting for conceptual differences and observational unit reconciles most of the differences between Survey of Consumer Finances (SCF) and tax-based top income and wealth shares, but some divergence in both levels and trends remains for the top 0.1 percent of families. These residual top share differences are attributable to exactly *which* families, using the administrative tax data, are identified as representing the very top of each distribution. The SCF sampling strategy combines inferences based on gross capitalization (as in the direct estimates)

⁴⁰ The capitalization factor used in Saez and Zucman (2014) has increased in recent years because the SOI taxable interest income series has fallen while the FA taxable fixed-income assets have increased. Why these two series diverge so much in recent years is open for debate, but one potential story is that low interest rates and small balance growth leads to families' not reporting taxable interest on their tax forms. Generally, financial institutions will send a 1099-INT to an account-holder if taxable interest on the account is greater than \$10; at a 1 percent interest rate, 1099-INTs will be sent if the balance is greater than \$1,000. If all of the increase in FA fixed-income assets is due to small balance deposits, interest rates on savings accounts are near zero, and tax filers only report interest if they receive a 1099-INT, then deposit-account growth (due entirely to small-balances) may be associated with falling interest income observed in SOI data; the implied capitalization factor based on these two series will be too large because small-balance families do not report taxable interest income on their tax returns.

with an empirical correlation approach (based on previous SCF surveys) and multiple years of administrative data to better identify families that are *permanently* wealthy.

The SCF two-pronged sampling strategy makes it possible to distinguish the permanently wealthy from those families who happen to realize very high but transitory (generally capital) incomes in a given year. Families at the top end can experience a great deal of income volatility year-to-year (Figure 3). Using the administrative 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 above.

Models of Wealth and Income

Using administrative income records to identify high *wealth* families further 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, one must rely on annual capital income to infer wealth through a “gross-capitalization” model approach (Greenwood, 1983; Saez and Zucman, 2014).

However, only about half of assets can be inferred from a tax return.⁴¹ 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 (2014) 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 cyclicity 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 large portion of these families’

⁴¹ See Saez and Zucman (2014) Appendix Table A3.

income.⁴² 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).

Model Agreement

Both the gross capitalization model and the empirical correlation model predict wealth from administrative income. 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 correlation model (Table 1). Looking at finer rankings within top 10 percent, though, there are considerable differences.

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 (Table 1). 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.⁴³

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, 2014). The results in

⁴² 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.

⁴³ 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 1 indicate that such identification is not clear, and an alternative model will not necessarily rank families equivalently to the gross capitalizations model.

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 17). 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.

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 (Table 2) and better predicts the SCF wealth ranking (Table 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:

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

The wealth predictions from each model can explain a large portion of the variation in SCF wealth, but the correlations model can explain more (Table 2, row 4). 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 2). 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 capitalizations model (Table 3). 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 capitalizations, 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.

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, 2014). 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 4). 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.

Income Reporting in the SCF

The level and trend in top wealth shares in the SCF and in capitalized income data appears to be about the same after adjusting the SCF and administrative income data (Figure 15), and remaining differences are likely attributable to the model used to capitalize the income data (Figure 17). However, even when the SCF is put in consistent terms with the administrative data, the capital income share of the top 0.1 percent in the administrative data still appears to be higher and rising faster than in the SCF (Figure 8, Panel C). Under-reporting of capital income in the survey itself may be responsible for some of this divergence. Even though it is impossible to evaluate the accuracy of any given SCF respondent's reported income, it is possible to compare the *growth distribution* of incomes reported by SCF respondents to the *growth distribution* observed in the SOI administrative data for families with comparable income levels.⁴⁴

Aggregate total income in the SCF generally matches total aggregate income published by SOI, but the aggregate of some forms of capital income in the SCF appear to be understated (Moore and Johnson, 2009). High income and high wealth families typically have large capital and thus volatile income. For example, in the 2011 SOI data, about 60 percent of the families with AGI greater than \$500,000 realized a decline in income (AGI) in their 2012 tax filing (Figure 16, red bars). At the tails, about 22 percent of the families in 2011 with AGI greater than \$500,000 had a decline in income of 50 percent or more, and about 11 percent had an increase in income of 50 percent or more. However, of the 2011 SOI families with AGI greater than \$500,000 and that responded to the SCF, about 74 percent reported an annual income decline, and nearly 32 percent reported a decline in income of 50 percent or more (Figure 16, blue bars).

Some high income SCF respondents, in other words, may be, on net, underreporting 2012 income. Many high income SCF families haven't filed their taxes at the time of interview so they may be unaware of their actual 2012 income during the interview.⁴⁵ Further, the high *wealth* families that the SCF identifies may not be the same as the families with the highest *capital income* in any given year because of the same income fluctuations shown in Figure 16. Both of these explanations help explain why we cannot fully reconcile capital income in Figure 8 but can reconcile wealth shares in Figure 15.

⁴⁴ We are grateful to the IRS Statistics of Income Division for the unpublished growth rate distributions shown here.

⁴⁵ Almost 19 percent of SCF families in the top two sampling strata have not yet filed their taxes as of interview date but plan to do so; only 4 percent of all other SCF families have not yet filed taxes. Many high-wealth families file the taxes late in the year, after getting an extension.

Overall, though, the SCF appears to meet its goal of providing a representative snapshot of the distribution of U.S. wealth.⁴⁶ When the limitations of the capitalized administrative income data are put to the SCF, the level and trend in top wealth shares are very similar (Figure 12) and when a restrictive assumption (about *one* rate of return) is changed in the capitalization model then even the *very* top wealth share levels and trends are nearly identical (Figure 15). Indeed, the relatively small discrepancies in reported incomes at the very top emphasize the importance of using multiple models to identify the extremely wealthy households.

VI. Conclusions

Rising top income and wealth shares are often cited as a call to action by those who believe government can and should do more about inequality in terms of taxation, spending, regulation, and other market interventions. Rising inequality raises obvious normative concerns, and there is growing belief that recent macroeconomic instability and slow growth may be additional symptoms of the same underlying phenomenon.⁴⁷ These ideas have begun to transcend academic debates, entering the mainstream political arena through best sellers such as Rajan (2010), Stiglitz (2012), and Piketty (2014). Economists disagree about the fundamental causes of rising inequality, as some argue that the trends are associated with free market prices adjusting to equate supply and demand, while at the other extreme some argue that influence wielded by the already wealthy improves their market shares by changing the rules of the game.⁴⁸

The estimates here from the Survey of Consumer Finances (SCF) concur that inequality, at least as reflected in top income and wealth shares, has been rising in recent decades. The levels and trends in top shares are more muted than in recent studies which are based directly on administrative income tax data (Piketty and Saez, 2003, updated, and Saez and Zucman, 2014) but the levels and trends are a bit larger than estimates based on estate tax data (Kopczuk and Saez, 2004). The SCF approach of combining administrative and survey data makes it possible to adjust administrative-based top share estimates for unit of observation, more economically

⁴⁶ The SCF has never covered the Forbes 400 families, though, and including an estimate of their wealth slightly changes top wealth shares in the SCF – see Figure 12 and Vermeulen (2014).

⁴⁷ For a somewhat contrary position on the economic stability effects, see Bordo and Meissner (2012).

⁴⁸ The view that markets underlie rising inequality is well described by Kaplan and Rauh (2010, 2013). See also Jones (2015) for a discussion of how competition among innovators affects top shares.

meaningful income and wealth concepts, and to smooth over the transitory income fluctuations that lead to misclassification of families at the very top.

These differences have important implications for both levels and trends. Conceptual adjustments and unit of observation lower the estimated top 1 percent *wealth* share from 42 percent to 36 percent in 2013, and the growth of that share since 1992 is reduced from 13 to 6 percentage points. The top 1 percent *income* share is reduced from 23 to 20 percent in 2013, and the growth since 1992 is reduced from 10 to 8 percentage points. Residual gaps at the *very* top of the income and wealth distributions arise from SCF restrictions on interviewing Forbes 400 families, the SCF strategy of targeting the *permanently* wealthy, as well divergence in aggregate benchmarks that lead to unrealistic capitalization factors for fixed interest assets.

Although the SCF makes it possible to inform and improve on direct estimates of top shares based on administrative tax data, the survey is still far from capturing comprehensive income and wealth measures. The SCF adds some government transfers into the tax-based income measures, but still misses employer-provided benefits, government in-kind (especially health) transfers, and other forms of income that are both substantial and growing over time. There are also direct analogs in terms of shortcomings in the wealth measures, as (for example) the value of most families' key retirement asset—Social Security—is not measured as part of household net worth. The effect of these omissions is important for understanding top shares, and even more important when looking at inequality across the entire distribution.

The reconciliations here cannot be extended back in time before the development of the modern household surveys, but the specific issues raised draw attention to how changes in government policies and market practices are affecting the measurement of top shares over time. In particular, although the administrative tax data makes it possible to show that top share families are getting increasingly large slices of a particular pie, the overall size of the pie being measured in those data is shrinking relative to more economically-meaningful concepts of income and wealth. The increasingly unmeasured part of the pie is not disappearing, but it is evolving. It may be difficult or even impossible to allocate the missing pieces in the very long historical series, thus any very long trends should also be viewed with an eye towards the conceptual divergence being driven by evolving government policy and economic institutions.

Building on the conceptual measurement theme, the reconciliation of top shares presented here speaks directly to the underlying impetus for—and possible approaches to—public policy

towards income and wealth distribution. The failure to properly measure the effects of government policies and market practices that disproportionately benefit families in the middle and bottom of the resource distribution leads directly to overstatement of top income and wealth shares. Policies and practices such as social insurance, government investment in human capital, workplace regulation, and collective bargaining overcome real market failures, meaning economic surplus is being generated by those policies, and the debate is thus properly focused on the distribution of that economic surplus. If we measure only the costs of such policies and practices, without measuring the benefits, it becomes more difficult to make the case for addressing market failures in future policy debates.

VII. References

- Armour, Philip, Richard V. Burkhauser, and Jeff Larrimore. 2014. "Levels and Trends in U.S. Income and its Distribution: A Crosswalk from Market Income towards a Comprehensive Haig-Simons Income Approach," *Southern Economic Journal*, 81(2): 271-293.
- Atkinson, Anthony B., Thomas Piketty, and Emmanuel Saez. 2011. "Top Incomes in the Long Run of History," *Journal of Economic Literature*. 49(1): 3-71. (March)
- Attanasio, Orazio, Erik Hurst, and Luigi Pistaferri. 2015. "The Evolution of Income, Consumption, and Leisure Inequality in the US, 1980-2010," 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.
- Bordo, Michael D., and Christopher M. Meissner. 2012. "Does Inequality Lead to a Financial Crisis?" National Bureau of Economic Research Working Paper 17896. (March)
- 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)
- Brown, Meta, Andrew Haughwout, Donghoon Lee, and Wilbert van der Klaauw. 2011 (revised 2013). "Do We Know What We Owe? A Comparison of Borrower- and Lender-Reported Consumer Debt," *Federal Reserve Bank of New York Staff Reports*, 523: 1-54.
- Burkhauser, Richard V., Jeff Larrimore, and Kosali I. Simon. 2012. "A 'Second Opinion' on the Economic Health of the American Middle Class." *National Tax Journal*, 65(1): 7-32.
- 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*, 94(2): 371-388.
- 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. 2014. *The Distribution of Household Income and Federal Taxes, 2011*. Washington, DC: Congress of the United States, Congressional Budget Office. (November)
- 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*, 67-122. (Spring)

- 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.
- Jones, Charles I. 2015. "Pareto and Piketty: The Macroeconomics of Top Income and Wealth Inequality," *Journal of Economic Perspectives*, 29(1): 29-46. (Winter)
- Kaplan, Steven N., and Joshua Rauh. 2013. "It's the Market: The Broad-Based Rise in The Return to Top Talent," *Journal of Economic Perspectives*, 27(3): 35-56. (Summer)
- Kaplan, Steven N., and Joshua Rauh. 2010. "Wall Street and Main Street: What Contributes to the Rise in the Highest Incomes?" *Review of Financial Studies*, 23(3): 1004-1050.
- 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.
- Kopczuk, Wojciech. 2015. "What Do We Know About the Evolution of Top Wealth Shares in the United States?" *Journal of Economic Perspectives*, 29(1): 47-66. (Winter)
- Kopczuk, Wojciech, and Emmanuel Saez. 2004. "Top Wealth Shares in the United States, 1916-2000" Evidence from Estate Tax Returns," *National Tax Journal*, 52:445-487.
- Kuznets, Simon. 1953. "Shares of Upper Income Groups in Income and Savings." New York: National Bureau of Economic Research, 1953.
- Moore, Kevin and Barry Johnson. 2009. "Differences in Income Estimates Derived from Survey and Tax Data," In *Proceedings of the Survey Research Methods Section, American Statistical Association*. American Statistical Association, pp. 1495-1503.
- 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.

Pareto, Vilfredo. 1896. *Cours d'Économie Politique*. Geneva: Droz.

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

Piketty, Thomas. 2014. *Capital in the 21st Century*. Harvard University Press. (April)

Piketty, Thomas, and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913-1998," *Quarterly Journal of Economics*, 118(1): 1-39.

Rajan, Raghuram G. 2010. *Fault Lines: How Hidden Fractures Still Threaten the World Economy*, Princeton University Press. (May)

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. 2014. "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data," National Bureau of Economic Research Working Paper 20625. (October)

Smeeding, Timothy M., and Jeffrey P. Thompson. 2011. "Recent Trends in Income Inequality: Labor, Wealth and More Complete Measures of Income," in Immervoll, Herwig, Andreas Peichl and Konstantinos Tatsiramos eds., *Who Loses in the Downturn? Economic Crisis, Employment and Income Distribution. Research in Labor Economics, Vol. 32*. Bingley, U.K.: Emerald, pp. 1-50.

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

Stiglitz, Joseph E. 2012. *The Price of Inequality: How Today's Divided Society Endangers Our Future*, W.W. Norton. (June)

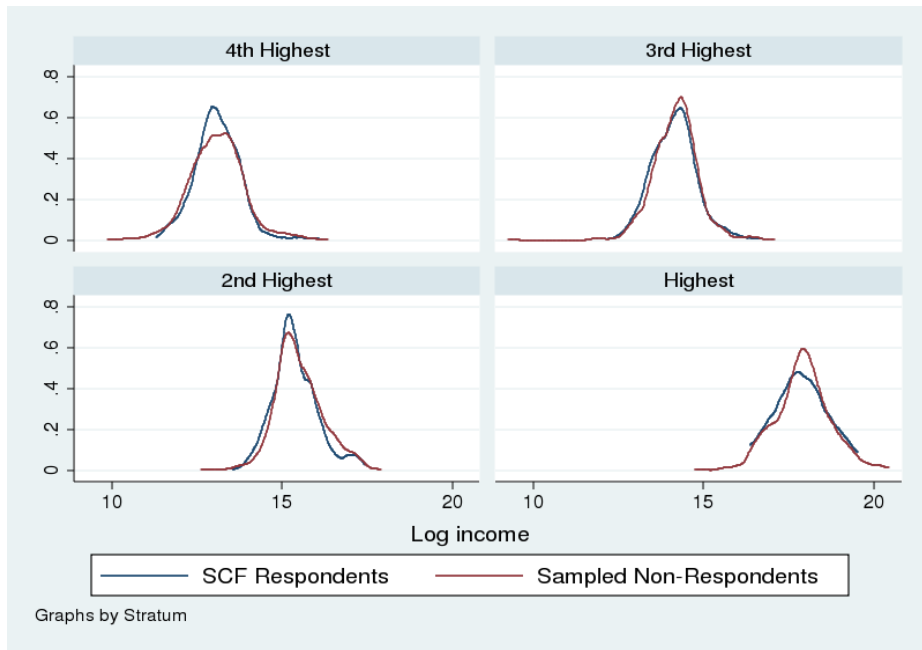
Vermeulen, Phillip. 2014. "How Fat is the Top Tail of the Wealth Distribution?" Directorate General Research, Monetary Policy Research Division, European Central Bank. (September)

Weil, David N. 2015. "Capital and Wealth in the 21st Century," National Bureau of Economic Research Working Paper 20919. (January)

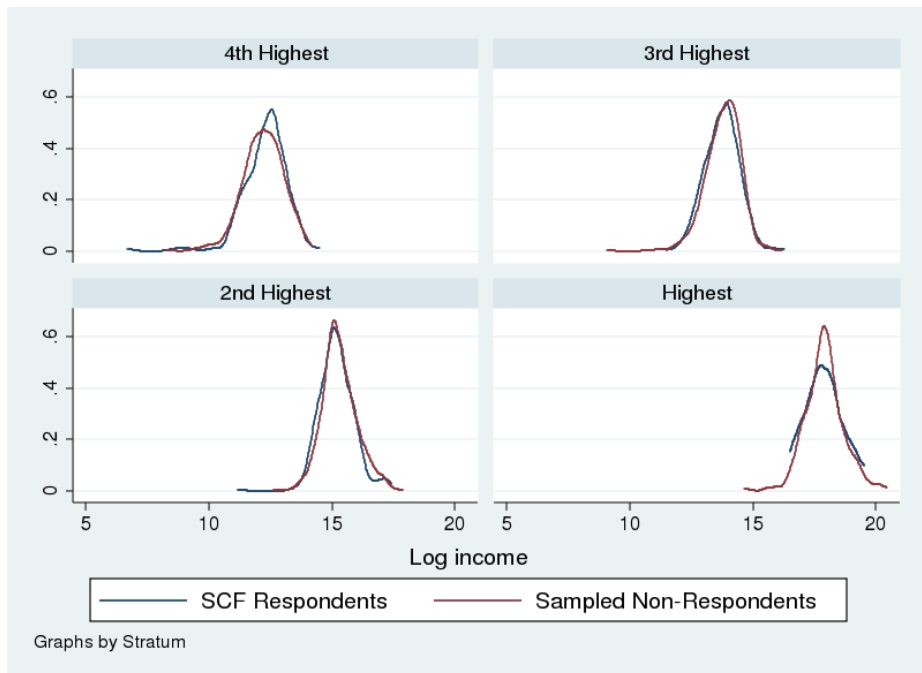
Wolfers, Justin. 2015. "The Gains from the Economic Recovery are Still Limited to the Top One Percent" <http://www.nytimes.com/2015/01/28/upshot/gains-from-economic-recovery-still-limited-to-top-one-percent.html>.

Figure 1. Income Densities for Top Strata SCF Respondents and Non-Respondents

A. Total Income, 2009-2011



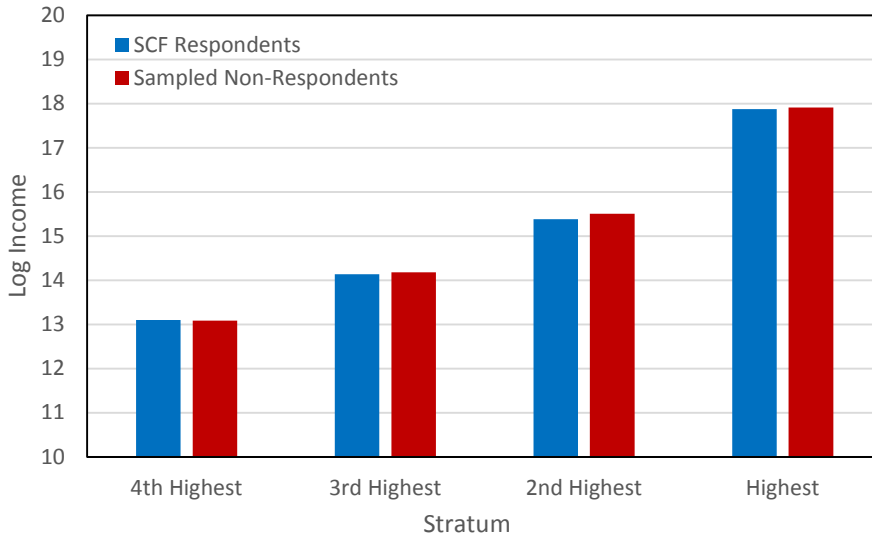
B. Capital Income, 2009-2011



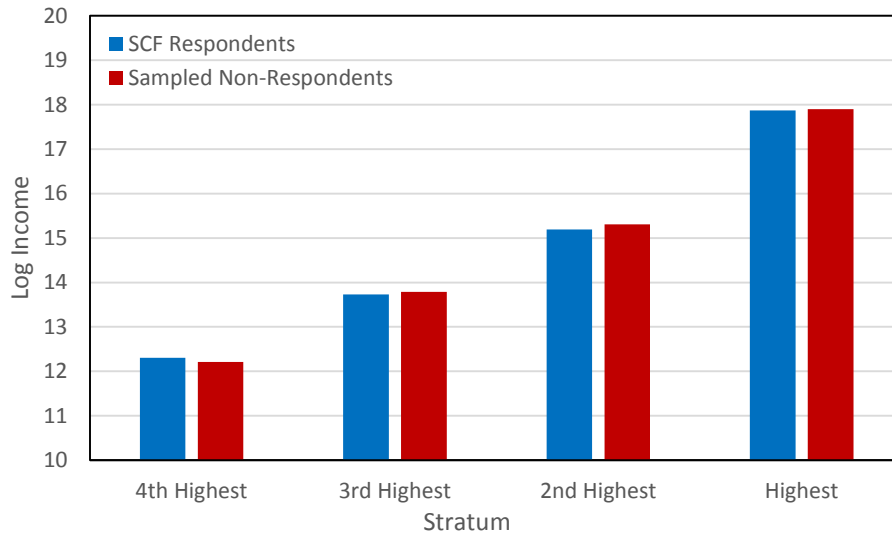
Note: Incomes are 3-year averages and include capital gains. Note: Sample includes the 4 highest strata, which fully encompasses the top 1% of the predicted wealth distribution. Incomes include capital gains. Data for the calendar years 2009-2011 are associated with the sampling for the 2013 SCF.

Figure 2. Mean Incomes for Top Strata SCF Respondents and Non-Respondents

A. Total Income, 2009-2011



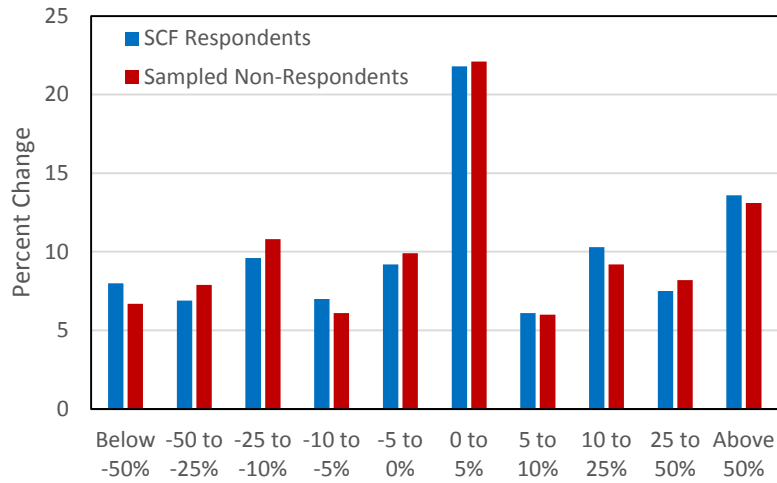
B. Capital Income, 2009-2011



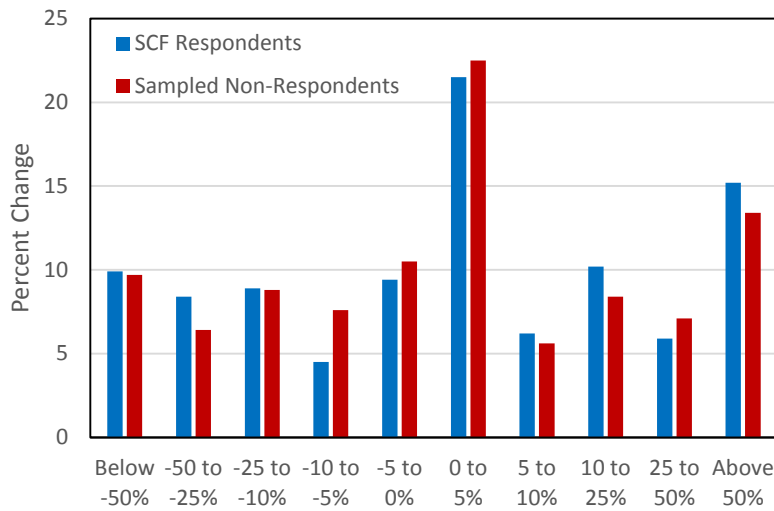
Note: Incomes are 3-year averages include capital gains. Sample includes the 4 highest strata, which fully encompasses the top 1% of the predicted wealth distribution. Incomes include capital gains. Data for the calendar year 2009-2011 are associated with the sampling for the 2013 SCF.

Figure 3. Pre-Survey Income Volatility of Top Strata SCF Respondents and Non-Respondents

A. Percent Change in Total Income, 2010 to 2011



B. Percent Change in Capital Income, 2010 to 2011



Note: Sample includes the 4 highest strata, which fully encompasses the top 1% of the predicted wealth distribution. Incomes include capital gains. Data for the pre-survey calendar years 2010 and 2011 are associated with the sampling for the 2013 SCF.

Figure 4. Ratio of Administrative Data Aggregate Incomes to Alternative NIPA Concepts

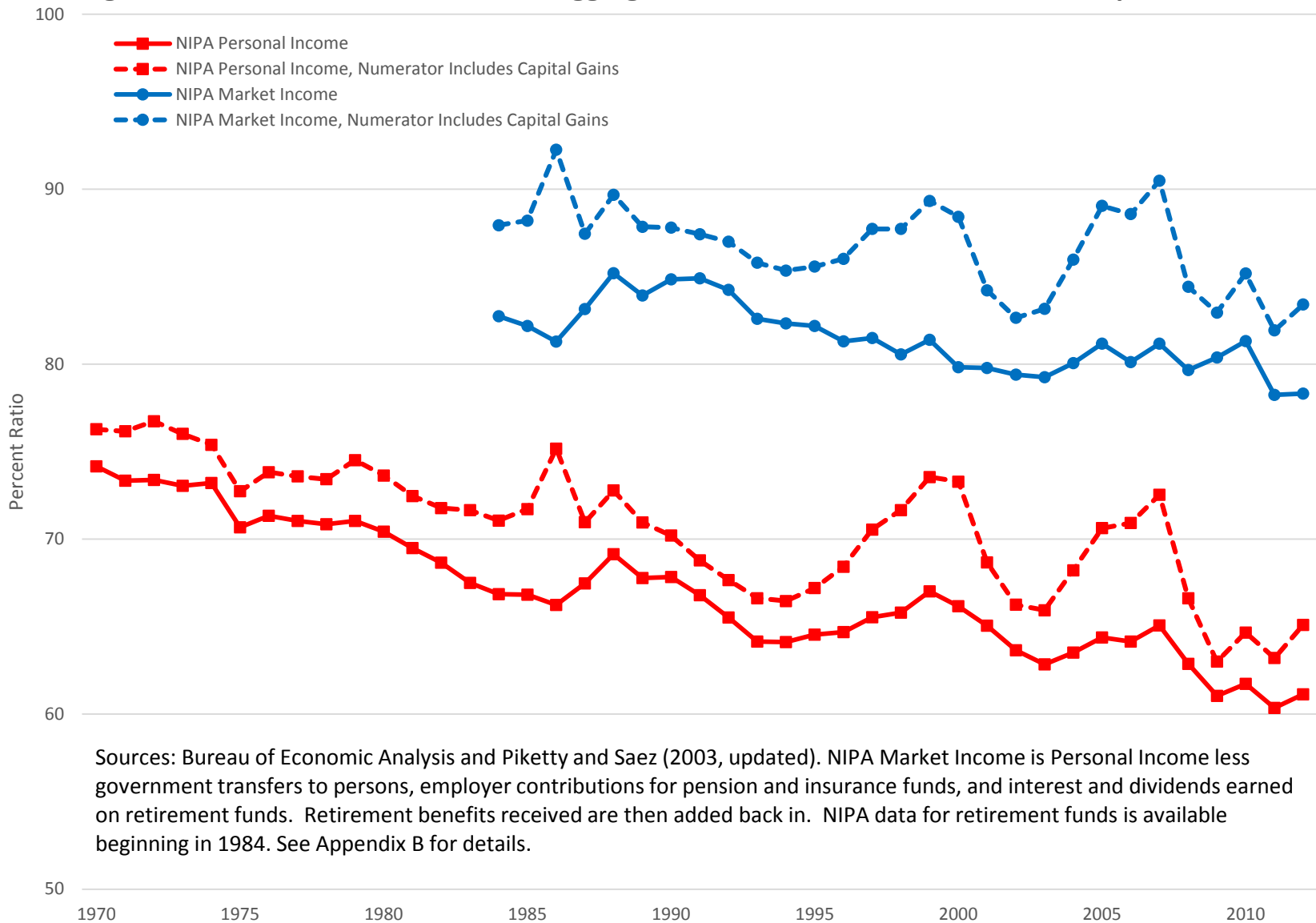


Figure 5. Reconciling Survey of Consumer Finances (SCF) and Administrative Data Top 1% Total Income Shares

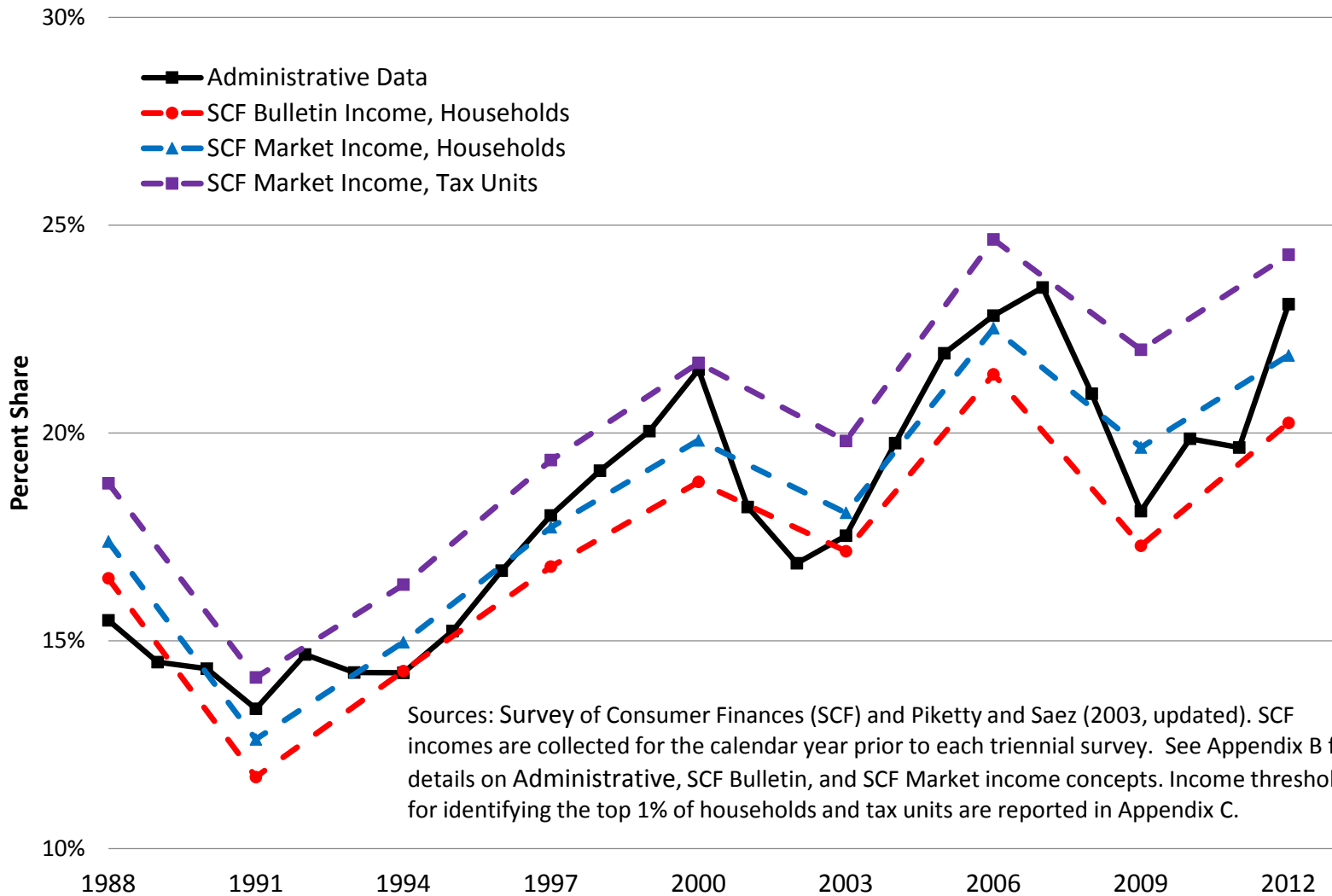


Figure 6. Survey of Consumer Finances (SCF) and Tax-Based Total Income Top Shares

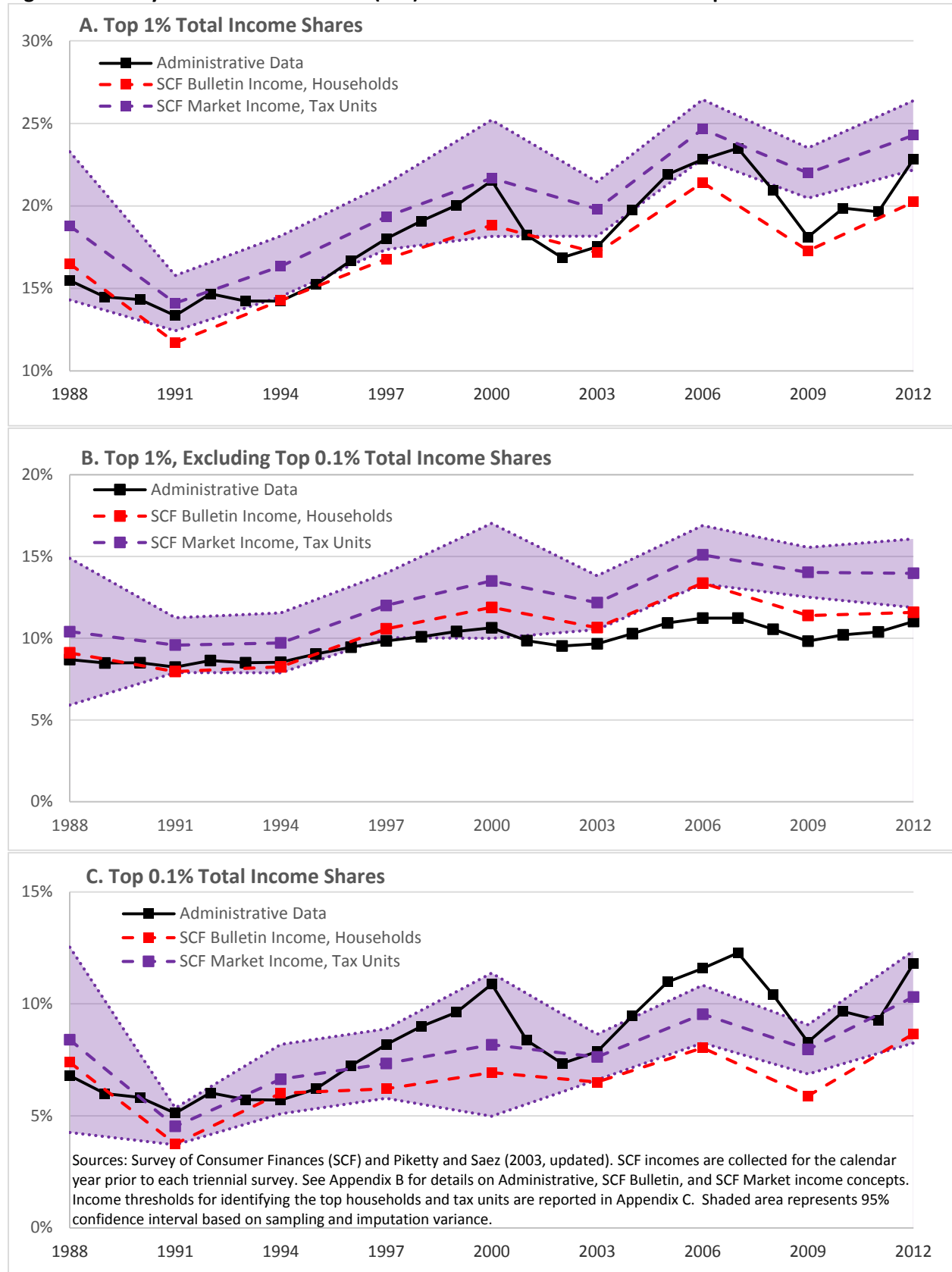


Figure 7. Reconciling Survey of Consumer Finances (SCF) and Administrative Data Top 1% Capital Income Shares

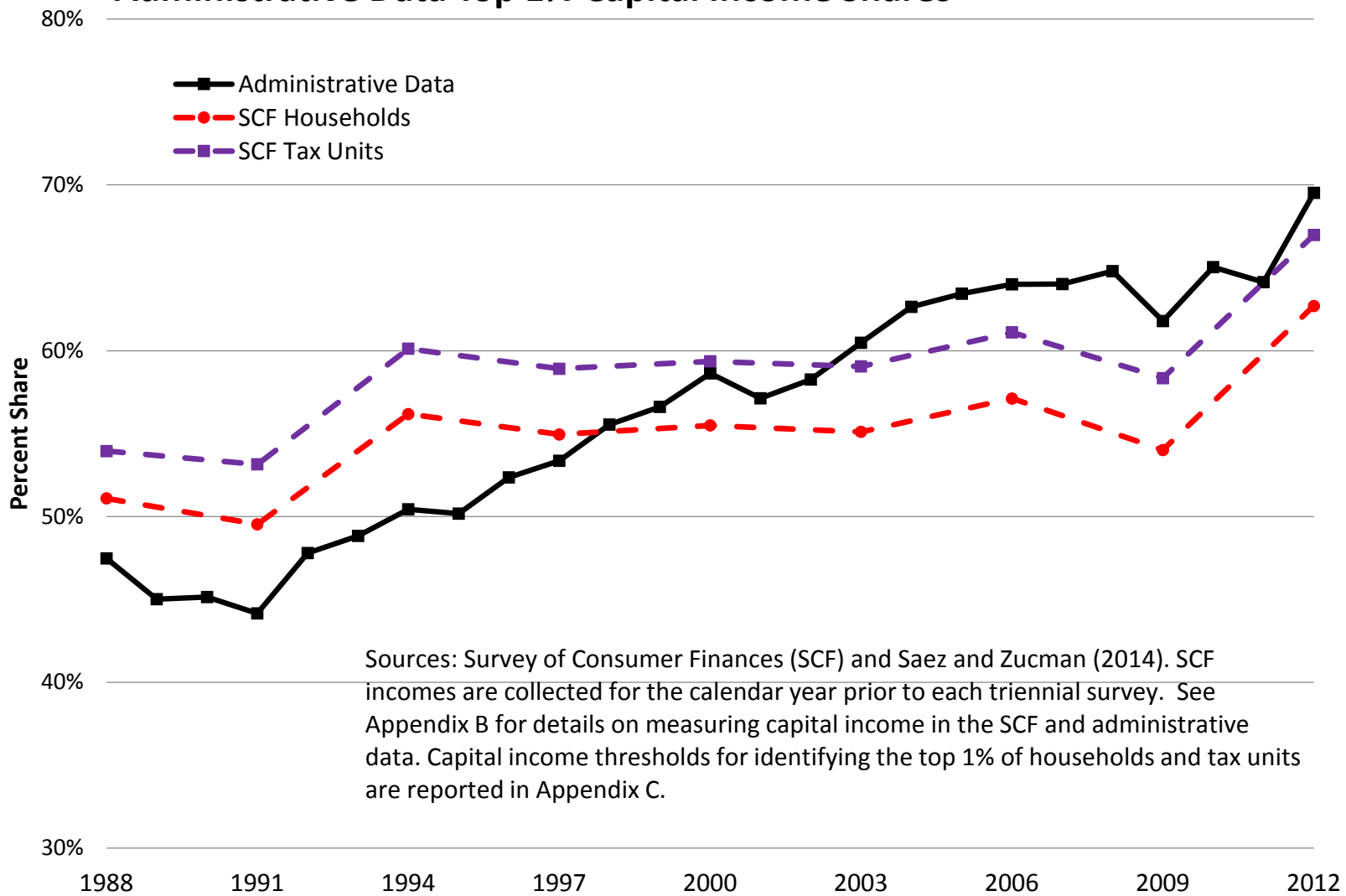


Figure 8. Survey of Consumer Finances (SCF) and Tax-Based Capital Income Top Shares

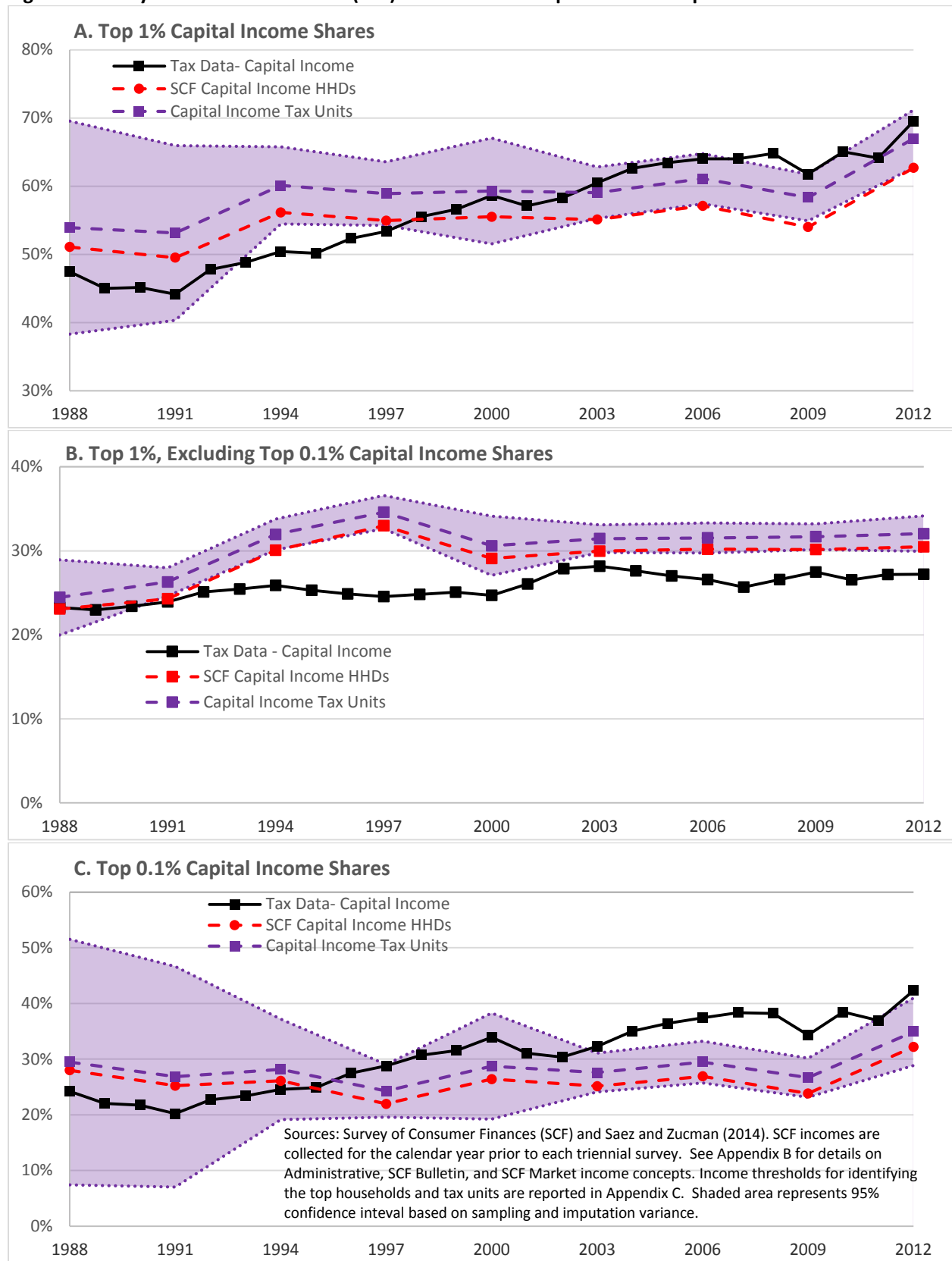


Figure 9. Effect of Allocating Missing Personal Income on Top 1% Income Shares

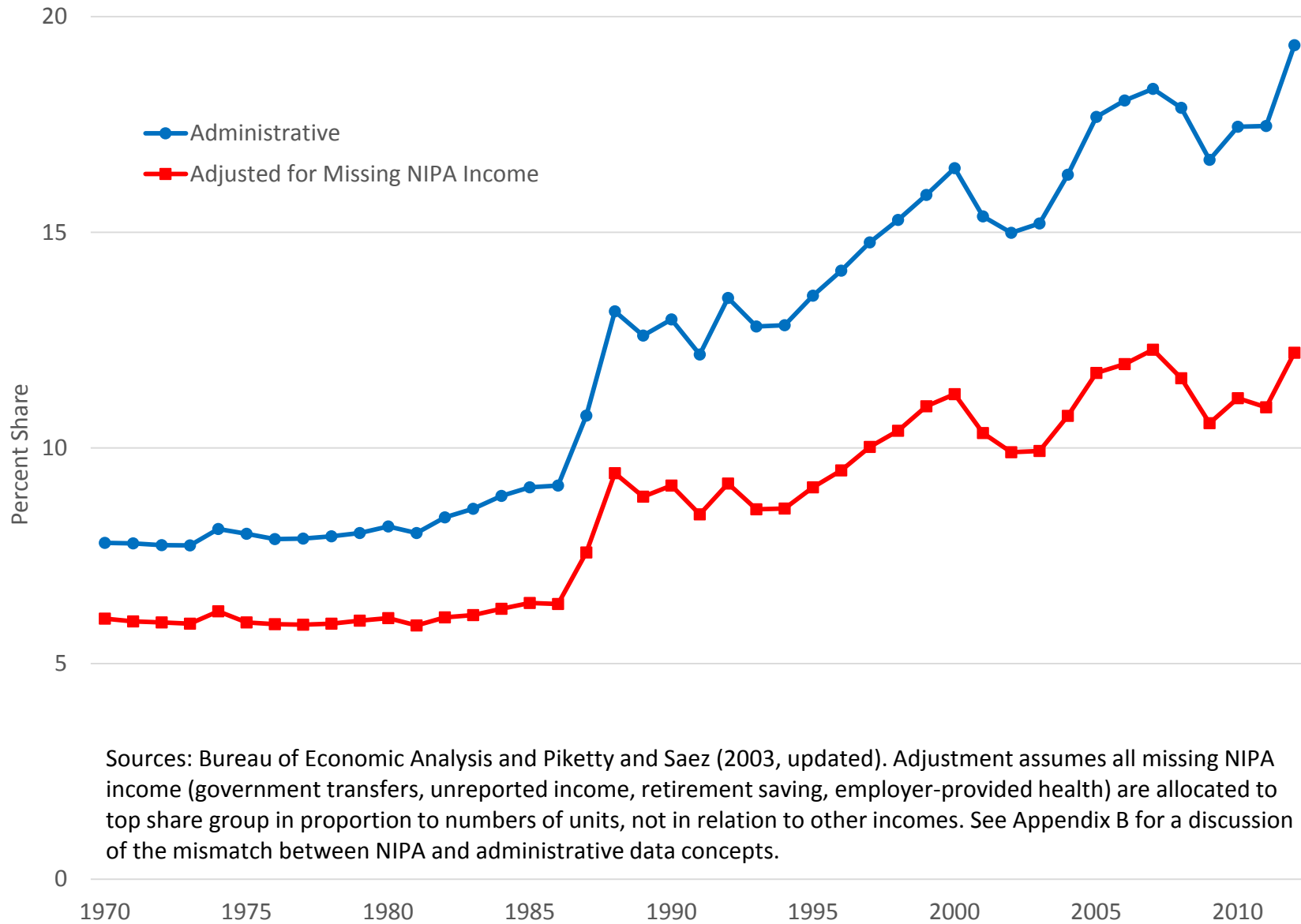
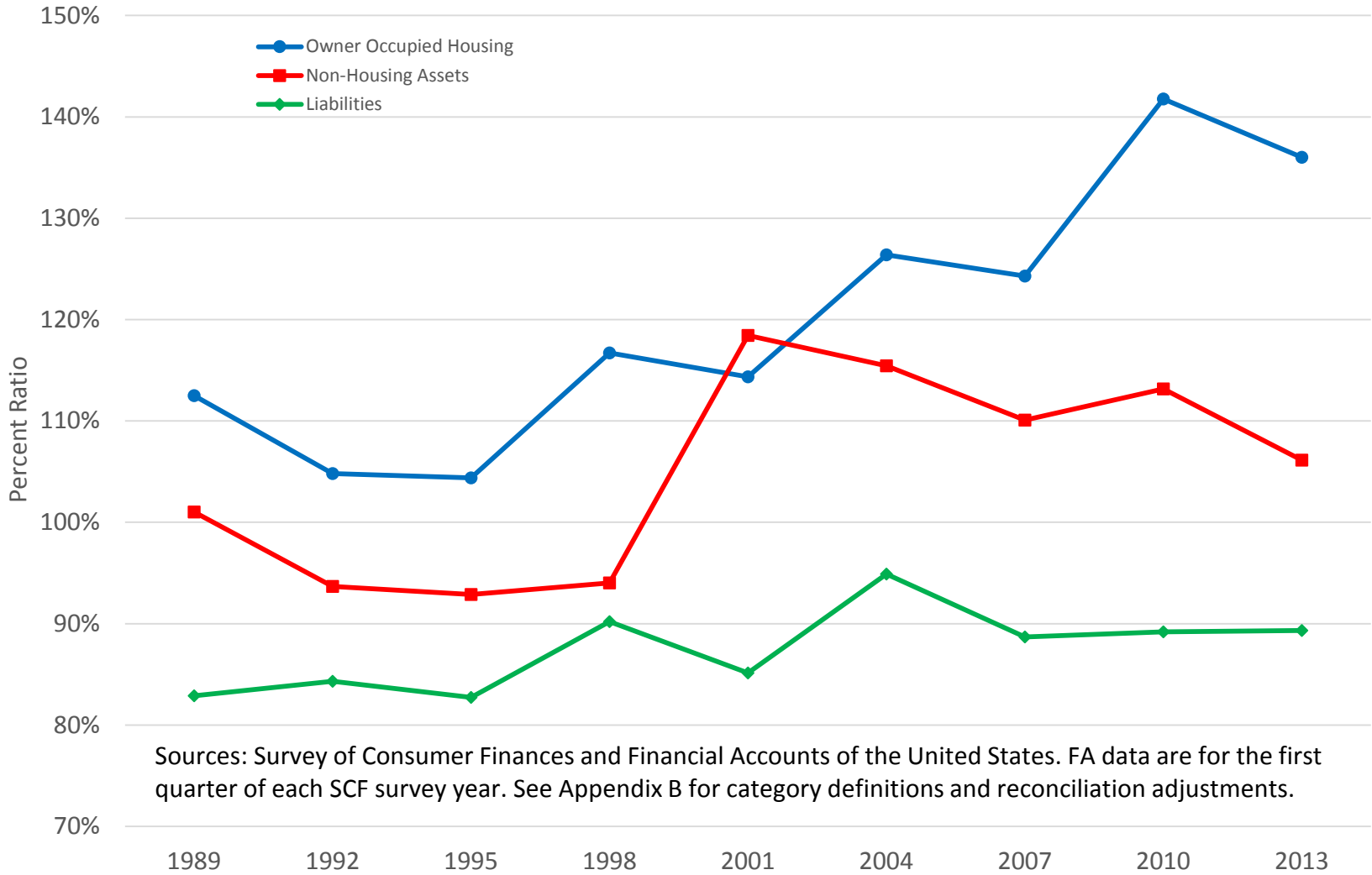


Figure 10. Ratio of SCF Balance Sheet Categories to Comparable FA Aggregates



Sources: Survey of Consumer Finances and Financial Accounts of the United States. FA data are for the first quarter of each SCF survey year. See Appendix B for category definitions and reconciliation adjustments.

Figure 11. Reconciling Survey of Consumer Finances (SCF) and Administrative Data Top 1% Wealth Shares

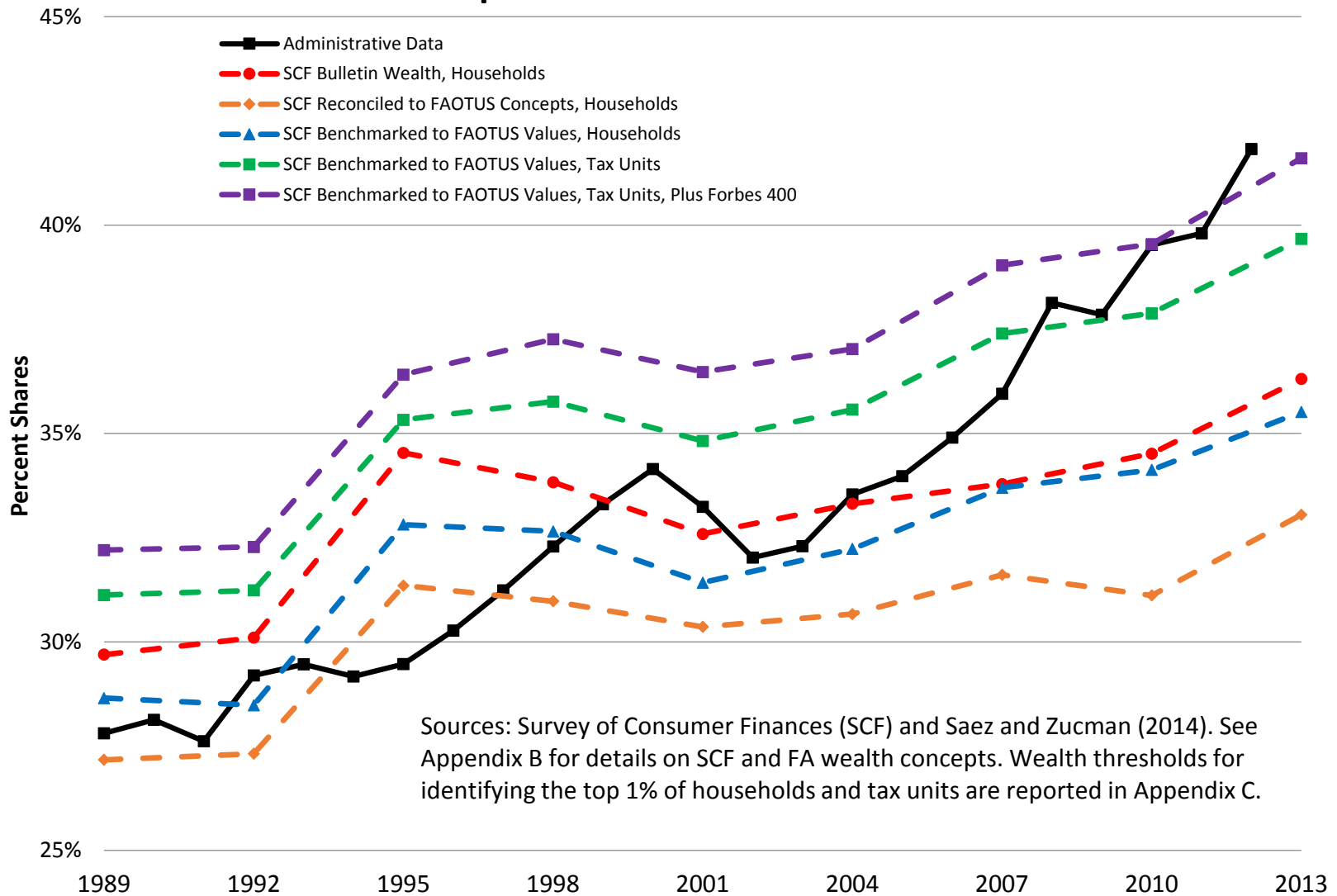


Figure 12. Survey of Consumer Finances (SCF) and Administrative Data Wealth Shares

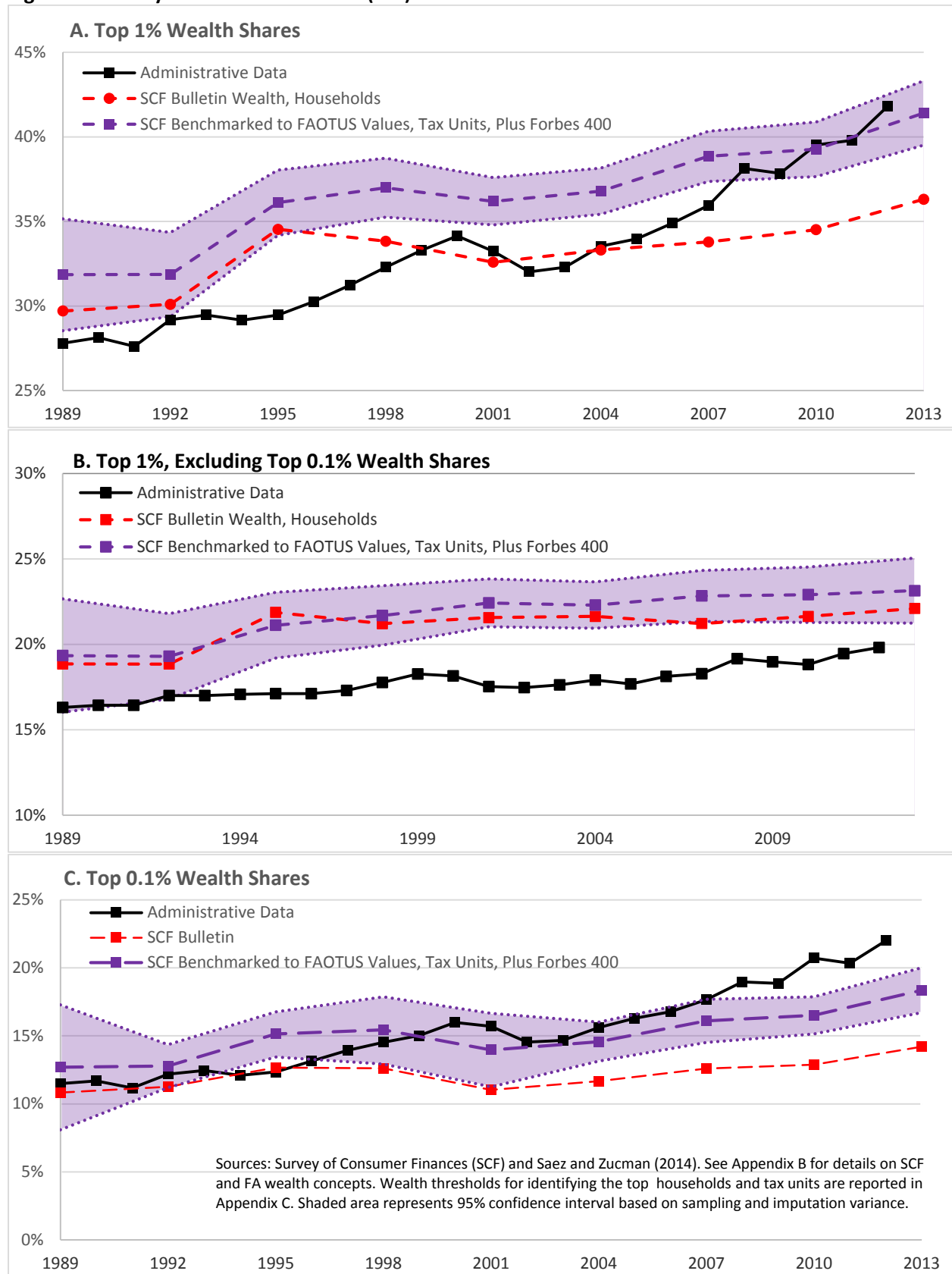
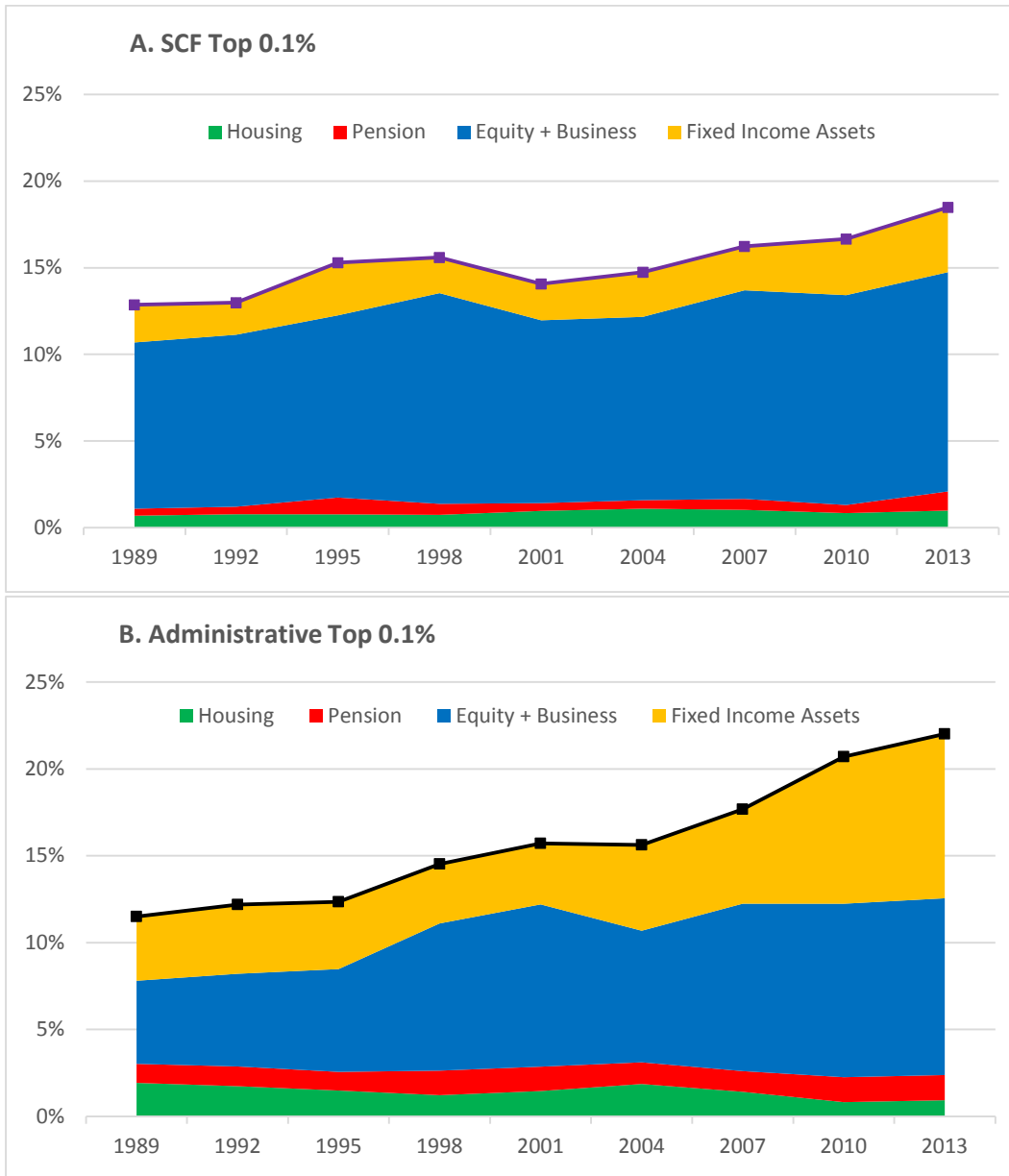
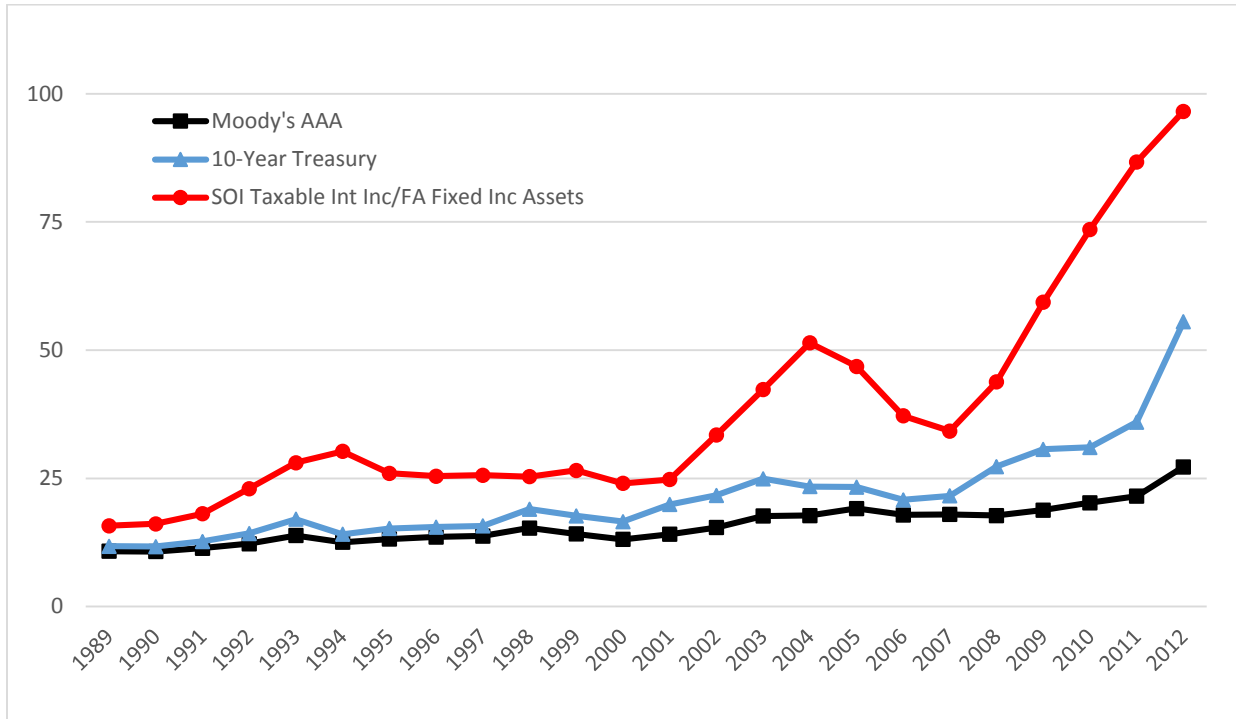


Figure 13. Wealth Composition in the SCF and Capitalized Administrative Income Data, 1989-2013



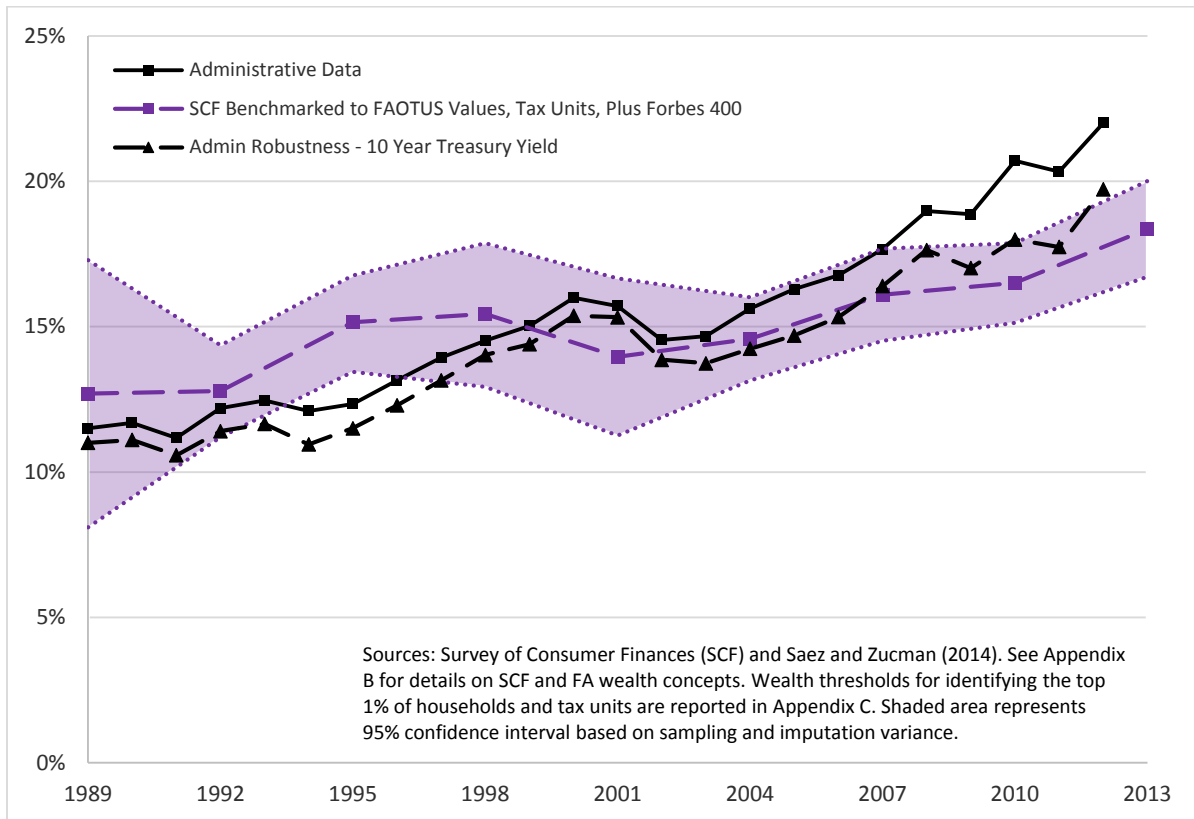
Notes: In panel A, we assume that the assets of Forbes 400, omitted from the SCF, are split proportional to the assets of the top 0.01% according to Saez and Zucman (2014). Administrative data are through 2012, though labelled as 2013. For each year on the x-axis, share of wealth held by the top 0.1 percent of families is broken into four general types of wealth: wealth from housing, from pensions, from corporate equities and private businesses, and from fixed income assets. Fixed income assets are bonds, CDs, savings accounts, and money market funds. Equities and businesses include the net worth of corporate equities, S-Corps, partnerships, and sole proprietorships. The cumulative height of the SCF top 0.1 percent is the SCF net worth benchmarked to FA values, adjusted for tax-units, and including an estimate of the Forbes 400 (i.e. the purple line in figure 12, panel C).

Figure 14. Heterogeneity in Potential Capitalization Factors to Generate Fixed-Income Assets from Interest Income, 1989-2012



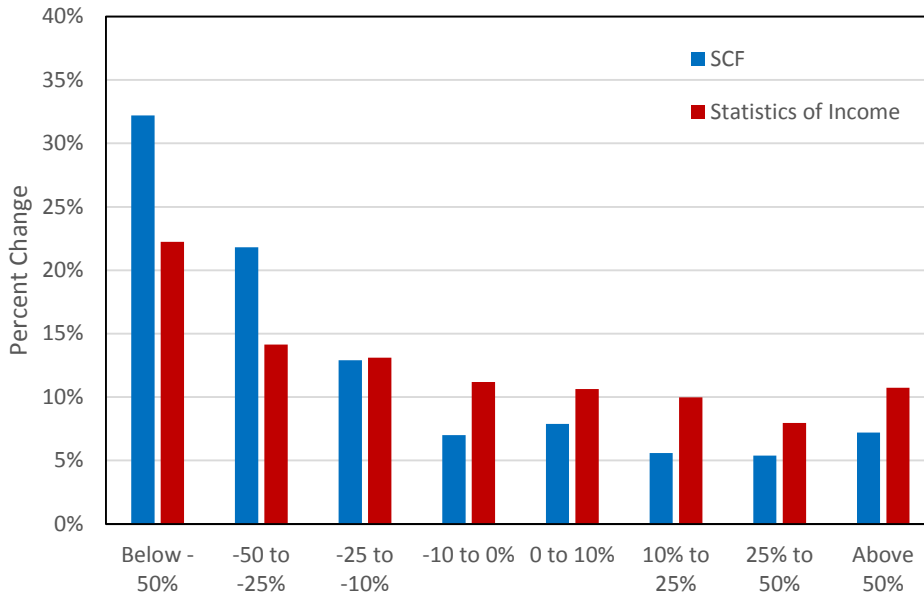
Notes: In a gross capitalization model, the capitalization factor for taxable interest income is the rate at which interest income will be grossed-up to infer fixed-income assets. The Moody's AAA line shows the inverse of the interest rate of the Moody's AAA corporate bond rate (seasoned issue, all industry, annualized) from the Federal Reserve H.15 data series. The 10-year Treasury line shows the inverse of the 10-year Treasury yield, annualized. The red line shows the inverse of the ratio of SOI taxable interest income to the stock of fixed income assets in the Financial Accounts table B.101.

Figure 15. Share of Wealth Held by the Top 0.1 Percent after Heterogeneous Capitalization Factor Allowed in Administrative Income



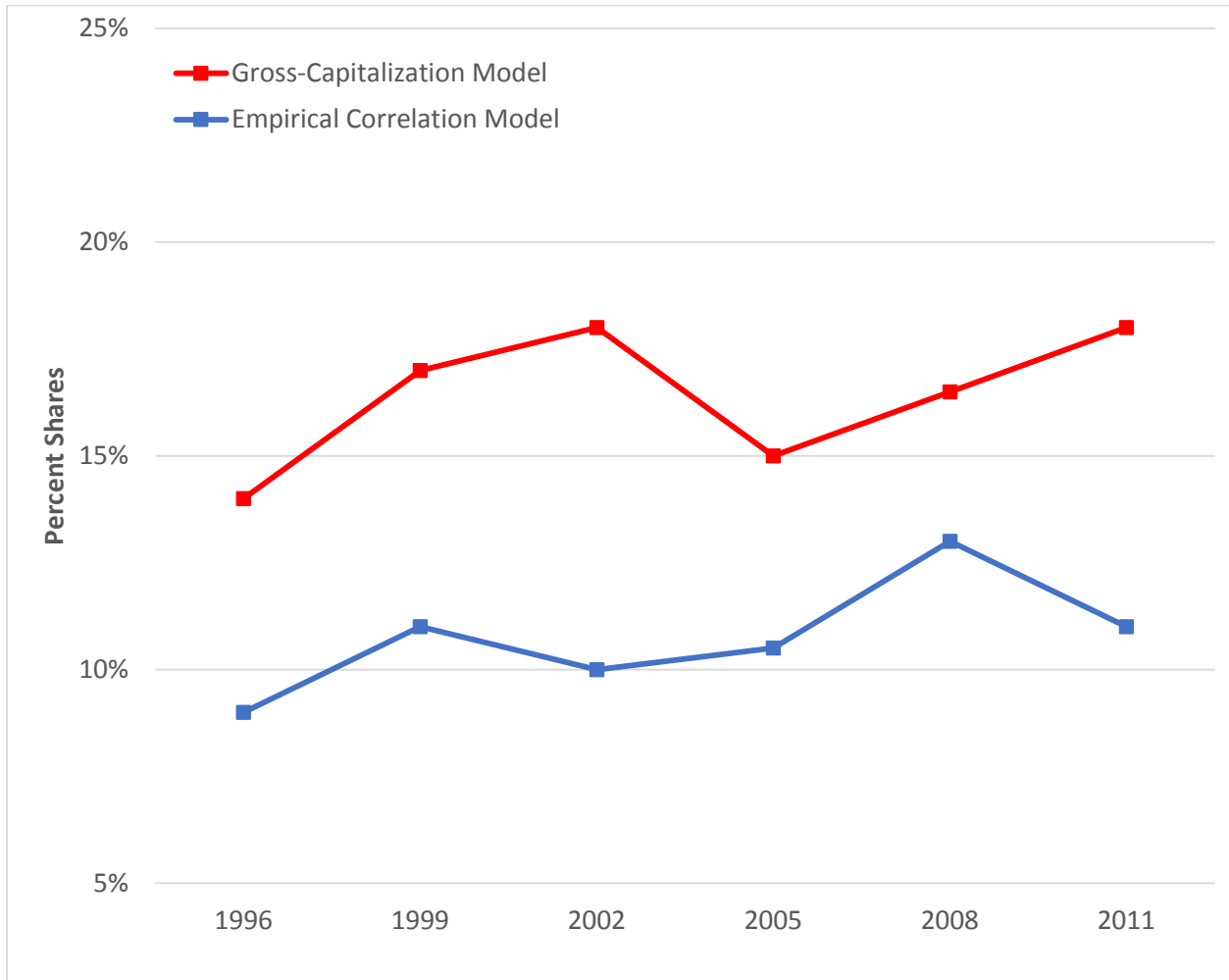
Notes: the purple line shows the share of wealth held by the top 0.1 percent in the SCF (adjusted to match FA asset categories, adjusting away from families and toward tax-units, and adding an estimate of the Forbes 400 wealth) and is identical to the purple line in Figure 12, panel C. The solid black line shows the share of wealth held by the top 0.1 percent in the capitalized income data (Saez and Zucman, 2014) and is identical to the black line from figure 12, panel C. The dashed black line shows a version of the black line where fixed income assets for the top 1 percent of income earners are generated by the inverse of the rate of return on the 10-year Treasury (the blue line in figure 14). In the dashed black line, the fixed income assets for the bottom 99 percent of income earners are still generated, as in the black line, by the inverse of the SOI taxable income to fixed-income assets in the Financial Accounts (the red line in figure 14). Underlying data for the dashed black line can be found in Saez and Zucman’s (2014) Appendix Table B40: “Top wealth shares, higher fixed income yield for top 1%.” In the solid black line, the fixed income assets for both the top 1 percent and the bottom 99 percent are generated by the inverse of the SOI taxable income to fixed-income assets in the Financial Accounts. Underlying data for the solid black line can be found in Saez and Zucman’s (2014) Appendix Table B1: “Top wealth shares.”

Figure 16. Income Change for Families with AGI Greater than \$500,000, 2011-2012



Notes: The red bars show the change in Adjusted Gross Income (AGI) from 2011 to 2012 among all tax returns with AGI over \$500,000 in 2011 (according to unpublished SOI tabulations). The blue bars show the change in AGI from 2011 to 2012 among sampled SCF households with AGI over \$500,000 in the INSOLE data. For SCF households, changes are computed using AGI provided by SOI in 2011 and AGI computed with NBER TAXSIM using household income from the 2013 SCF.

Figure 17. Predicted Top 0.1 Percent Wealth Share from Gross-Capitalization and Empirical Correlation Model in SCF Sampling Exercise



Notes: There are two models used in the SCF sampling process: a gross-capitalization model and an empirical correlation model. The gross capitalization model predicts capital asset wealth from capital income flows; the empirical correlation model uses the past correlation between sampling income and observed survey wealth to predict current wealth. Both models are described in more detail in Section II and Appendix A. Saez and Zucman (2014) use a version of the gross-capitalization model (normed to the FA household wealth) to predict household wealth. In each of the past six SCF sampling exercises, the gross capitalization model predicts higher wealth concentration at the top. The SCF sampling process keys the sampling year on data from two years prior to the survey year (which are the most up-to-date sampling data available).

Table 1. Impact of Ranking Top End Families by an Alternate Model

		Correlation Model Percentile				
		Bottom 90	90-99	(Top 1) 99-99.9	(Top 0.1) 99.9-99.99	(Top 0.01) 99.99+
Gross- capitalization Percentile	Bottom 90	0.89	0.10	0.01	0.00	0.00
	90-99	0.20	0.48	0.28	0.04	0.00
	(Top 1) 99-99.9	0.05	0.22	0.48	0.23	0.02
	(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 CDW panel data.

Table 2. Correlation Between SCF Wealth and Predicted Gross-Capitalization and Empirical Correlation Wealth

	(1)	(2)	(3)
ln(GC model wealth)	0.85 (0.02)	...	0.26 (0.02)
ln(Corr. model wealth)	...	1.02 (0.01)	0.76 (0.03)
Constant	1.57 (0.25)	-0.46 (0.23)	-0.73 (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 ().

Table 3. 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.

Table 4. 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 CDW panel data) compared to 2011 INSOLE data only.

Appendix A. Details on SCF Sampling Strategy

Data

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 INSOLE data, maintained by SOI, are the main data for the list sample selection. Prior to use, the INSOLE 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 CDW 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.

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 two models are briefly described in Section II and are described here in detail in unpublished papers on the SCF website, http://www.federalreserve.gov/econresdata/scf/scf_workingpapers.htm.

The exact form of the gross capitalization model in the SCF when selecting the 2013 SCF was:

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

where, there are where there are $i=1 \dots N$ tax units,

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

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 for a rate of return.

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. 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:

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

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

and $\ln(income_i^{j,2008-06}) = \ln\left(\frac{1}{2} * income_i^{j,2008} + \frac{3}{10} * income_i^{j,2007} + \frac{2}{10} * income_i^{j,2006}\right)$,
for $j=1\dots33$

The $\hat{\alpha}, \hat{\beta}_L, \hat{\beta}_D, \hat{\delta}$ 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:

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

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:

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

The $IQR()$ represents the interquartile range. In past years, the $blend_i$ weighted the empirical correlation model more than the gross capitalization model, in part because of the results shown in table 2 and 3 of this paper. The weight was even in the 2013 selection process.

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

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 selected in a similar fashion, with observations stratified by predicted wealth, and sub-stratified by age and financial income.

Appendix B. 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).

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 SCF and NIPA income are (1) SCF includes capital gains (variable X5712 in the public data) 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.⁴⁹

The income measure in the administrative data (Piketty and Saez, 2003, updated) is a much 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.100 (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. Because FA includes non-profit institutions as part of the household sector, we first remove identifiable non-profit assets and liabilities.⁵⁰ 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.⁵¹ We also remove miscellaneous assets and liabilities

⁴⁹ There are also small accounting differences in the treatment of business incomes, but we do not adjust for those.

⁵⁰ Table B.101 lines 5, 6, 7, 35, 38, and 40.

⁵¹ B.101 line 26 and 39.

from both FA and the SCF.⁵² This adjustment reduces SCF aggregate net worth in 2013Q1 by just under \$1 trillion and FA net worth by about \$1.1 trillion.

Table B.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

We next remove life insurance assets and liabilities from both data sets because of conceptual differences between the SCF and FA.⁵³

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.⁵⁴ 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.⁵⁵ 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

⁵² 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.

⁵³ 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.

⁵⁴ 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.

⁵⁵ 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.

and about \$61.0 trillion in FA in 2013 Q1.⁵⁶ The remaining \$11.2 trillion gap between SCF and FA is roughly twenty percent in 2013 (Figure 11).⁵⁷

Allocating Defined Benefit Pension Wealth to SCF Households

We allocate DB wealth at the household level using the method suggested by Saez and Zucman (2014), supplemented by SCF observations on DB coverage. First, we use the SCF to determine how aggregate FA DB pension wealth (\$10.9 trillion in 2013 Q1, for instance) should be divided between current and future pensioners. For each SCF head of household (the respondent) and spouse, we count the total number people currently receiving DB pension payments and the total number of people with a DB plan that they will draw down in the future. Because of population aging and the fall in prevalence of DB plans, the share of DB plans held by current pensioners rose from about 35 percent in 1989 to 54 percent of plans in 2013. We then use these shares to allocate a total dollar amount to SCF current and future DB pensioners. In 2013, for example, we give just under \$6 trillion to current pensioners and just under \$5 trillion to future pensioners. Next, we use survey responses on pension dollars received for current pensioners and previous year's wages for future pensioners to distribute these allocations across SCF households. More specifically, we assign pension wealth for current pensioners proportional to each household's share of total DB pension benefits received in the past year. We distribute future pension wealth proportional to each household's share of total future pensioner wages.⁵⁸

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, non-housing assets, and liabilities.⁵⁹ These three broad classifications represent the most general categories that are conceptually comparable in the macro and micro data (Table B.2).

⁵⁶ 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.

⁵⁷ The SCF minus FA net worth gap was smaller in the 1989 to 1998 period, see Henriques and Hsu (2014).

⁵⁸ Because we must count head of household and spouse separately, we allocate using x4112 and x4712 (respondent and spouse wage earnings before taxes) rather than total household wage income. Total wages of future DB pensioners was \$1.1 trillion in the 2013 SCF, while current DB benefits received totaled \$543 billion.

⁵⁹ 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

Table B.2. Reconciled SCF and FA Balance Sheet Categories			
	2013Q1 (Trillions)		
	SCF	FA	Difference
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 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 32 percent above the FA estimate, SCF non-housing assets are 7 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.

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.

Appendix C. Thresholds for Households versus Tax Units

Tables C.1, C.2., and C.3 list the fractile thresholds for the 90th, 99th, and 99.9th percentile of income, capital income, and net worth, respectively. Each table shows fractile cutoffs for the published SCF baseline measures, the reconciled concepts of income and wealth as described in Sections III and IV, and the administrative tax data.⁶⁰ The tables display thresholds at each step of the reconciliation process (e.g. Total Income at the SCF household level, SCF Market Income concept at the household level, and SCF Market Income on an adjusted tax unit basis).

In Table C.1., fractile thresholds generally decline moving from total income to market income. Cutoff points fall further once we shift from households to tax units as the unit of observation. Recall that there are about 30 percent more tax units than households, so the tax unit adjustment implicitly draws a threshold level farther down the distribution. Still, even using a conceptually similar income definition and unit of observation, the SCF has higher market income thresholds than administrative data at all fractiles.

Table C.1. Income Fractile Thresholds (Thousands \$)

	1988	1991	1994	1997	2000	2003	2006	2009	2012
90th Percentile									
SCF Total Income, Households (HHDs)	71	76	80	93	116	126	137	140	152
SCF Market Income, HHDs	70	74	79	91	114	125	134	135	150
SCF Market Income, Tax Units (TUs)	63	65	71	83	100	110	120	117	128
Administrative Data	55	62	67	76	90	90	104	106	114
99th Percentile									
SCF Total Income, HHDs	232	221	246	351	500	475	674	604	686
SCF Market Income, HHDs	225	220	237	340	479	472	654	611	679
SCF Market Income, TUs	201	202	205	280	403	399	550	520	570
Administrative Data	154	165	189	244	314	282	376	329	397
99.9th Percentile									
SCF Total Income, HHDs	855	683	885	1,474	2,157	2,124	3,100	2,233	2,475
SCF Market Income, HHDs	855	670	845	1,438	2,157	2,124	3,100	2,350	2,440
SCF Market Income, TUs	723	596	771	1,294	1,800	1,928	2,680	2,017	2,144
Administrative Data	671	619	712	1,076	1,548	1,181	1,910	1,334	1,915

⁶⁰ For our administrative data comparison, we use Piketty and Saez's (2003, updated through 2013, found here: eml.berkeley.edu/~saez/TabFig2013prel.xls) total income fractile thresholds; capital income and net worth thresholds are from Saez and Zucman (2014). For both total income and capital income, we use the administrative data series that include capital gains in both rankings and shares. On wealth, we compare to Saez and Zucman's baseline specification, which ranks tax units by capitalized income excluding capital gains, but uses capital gains to compute shares. SCF income and wealth series include capital gains in both rankings and shares.

Table C.2. Capital Income Fractile Thresholds (Thousands \$)

	1988	1991	1994	1997	2000	2003	2006	2009	2012
90th Percentile									
SCF HHDs	14	14	10	13	16	15	24	20	18
SCF TUs	10	9	7	9	12	9	15	11	11
Administrative Data	11	11	9	13	15	10	15	11	12
99th Percentile									
SCF HHDs	120	145	147	196	226	204	298	222	285
SCF TUs	102	121	121	160	200	170	252	191	214
Administrative Data	82	84	88	118	150	127	190	138	170
99.9th Percentile									
SCF HHDs	764	645	713	1,135	1,325	1,257	2,020	1,332	1,734
SCF TUs	617	535	644	1,000	1,103	1,017	1,750	1,232	1,483
Administrative Data	452	402	466	713	1,011	792	1,406	899	1,319

Table C.3. Net Worth Fractile Thresholds (Thousands \$)

	1989	1992	1995	1998	2001	2004	2007	2010	2013
90th Percentile									
SCF Bulletin Wealth, HHDs	364	357	381	494	740	834	910	953	942
SCF Reconciled to FA, HHDs	401	419	461	562	831	962	1,061	1,106	1,108
SCF Benchmarked to FA, HHDs	375	420	469	548	694	779	887	865	935
SCF Benchmarked to FA, TUs	314	341	387	457	555	644	717	653	729
SCF Benchmarked, TUs + Forbes 400	314	341	387	457	555	644	717	653	729
Administrative Data	329	366	421	520	596	658	790	643	662
99th Percentile									
SCF Bulletin Wealth, HHDs	2,251	2,318	2,460	3,793	5,787	6,356	8,360	6,815	7,880
SCF Reconciled to FA, HHDs	2,159	2,358	2,410	3,789	5,766	6,308	8,155	6,867	7,913
SCF Benchmarked to FA, HHDs	2,082	2,445	2,515	3,928	4,844	5,315	6,965	5,677	7,146
SCF Benchmarked to FA, TUs	1,786	2,045	2,041	3,239	3,787	4,492	6,141	4,763	6,003
SCF Benchmarked, TUs + Forbes 400	1,786	2,045	2,041	3,239	3,787	4,492	6,141	4,763	6,003
Administrative Data	1,510	1,757	2,003	2,770	3,057	3,488	4,312	3,663	3,964
99.9th Percentile									
SCF Bulletin Wealth, HHDs	9,703	9,207	13,898	15,346	20,338	25,109	30,826	27,488	30,894
SCF Reconciled to FA, HHDs	9,638	9,208	13,950	15,090	19,557	24,190	29,893	26,970	30,197
SCF Benchmarked to FA, HHDs	9,495	9,691	14,722	15,742	16,512	20,784	26,895	22,705	27,374
SCF Benchmarked to FA, TUs	7,817	8,238	12,728	13,778	15,413	19,036	22,654	21,111	23,476
SCF Benchmarked, TUs + Forbes 400	7,817	8,238	12,728	13,778	15,413	19,036	22,654	21,111	23,476
Administrative Data	6,452	7,497	8,353	11,631	13,491	15,850	20,165	17,837	20,561

Appendix D. Confidence Intervals for SCF Top Income and Wealth Shares

Figures 6, 8, and 12 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 (http://www.federalreserve.gov/econresdata/scf/scf_workingpapers.htm) that provide further details.

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.

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 Figures 6, 8, and 12 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:

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