Statistical Journal of the IAOS 33 (2017) 59–79
DOI 10.3233/SJI-170349
IOS Press

Lining up: Survey and administrative data estimates of wealth concentration[1]

Arthur B. Kennickell
Board of Governors of the Federal Reserve System (Ret.), Washington, DC 20551, USA
E-mail: Arthur.Kennickell@gmail.com

Abstract. The Survey of Consumer Finances (SCF) has a dual-frame sample design that supplements a standard area-probability frame with a sample of observations drawn from statistical records derived from tax returns. The tax-based frame is stratified on the basis of a "wealth index" constructed largely from observed income flows, with the intent of heavily oversampling wealthy households. Although the SCF is not specifically designed to estimate wealth concentration, the design arguably provides sufficient support to enable such analysis with a reasonable level of credibility. Similar estimates may also be made by using tax-based data directly, as in Saez and Zucman [1], by using a construct very close to a key part of the SCF wealth index. Such an approach has appeal as a way of tapping a much larger set of information to improve SCF estimates. Not surprisingly, there are differences in the two approaches, largely as a result of conceptual differences or complications in the survey implementation. This paper focuses on the top 1 percent of the wealth distribution, the group most intensively covered by the SCF list sample and it explores the stability of the relationship between the patterns of concentration in the survey data and parallel patterns in tax-based estimates and considers how those patterns differ across survey participants, the full sample and the entire survey frame. In addition, the paper makes as series of recommendation for further research on the technical support of the survey.

Keywords: Wealth measurement, sampling, oversampling, nonresponse, skewed distributions

1. Introduction

The Survey of Consumer Finances provides information on household wealth and its composition and relationship to other variables, on a common basis from 1989 to 2013. Because wealth is highly concentrated in the U.S., a survey that did not take special account of that situation would tend to produce very noisy (or even biased) estimates of values strongly influenced by the upper tail of the wealth distribution, such as means or concentration ratios. The SCF addresses this issue through the use of a list sample selected from statistical records derived from individual tax returns, as a supplement to a multistage area-probability sample. The list sample uses a proxy variable that supports oversampling of very wealthy households. That proxy and its relationship to net worth measured in the survey are the principal focus of this paper.

The wealth proxy is a "wealth index" constructed from income flows and other variables available in the frame data for tax filers. The sample design assumes there is a mapping from the observed data to a wealth concept close enough to an actual measure of wealth appropriate for the SCF to be able to provide a statistically efficient sample for wealth measurement and to allow for adjustments for nonresponse correlated with wealth.

The wealth index model, especially one aspect of it, bears a strong resemblance to an approach used by Saez and Zucman [1] to simulate the distribution of wealth in the U.S. for every year since 1913. That paper makes comparisons of their estimates to ones obtained from the SCF and claims that the SCF appears to under-represent increases in the concentration of wealth at the top of the distribution. As dis-

[1]This paper was presented at the 2015 Joint Statistical Meetings in Seattle, Washington; slides from that presentation are available at ww2.amstat.org/meetings/jsm/2015/onlineprogram/AbstractDetails.cfm?abstractid=316175. The current version of the paper is the final one, dated September 30, 2015, as redacted in December 2016.

cussed in [2] there are a variety of problems to be addressed for the mapping they assume from income to wealth to be comparable with the estimates available in the SCF. Further work in that direction seems highly worthwhile, however, if only to enhance the usefulness of the SCF through extrapolating wealth between survey periods in a very cost-efficient way or through supporting lower-variance (and potentially less biased) estimates of the extreme upper tail of the wealth distribution.

The SCF also faces problems in measurement. Any survey, particularly one with such a sensitive and difficult core subject matter as the SCF, would have problems with unit nonresponse and reporting errors. Over time, the SCF has devoted substantial resources to trying to understand and address as well as possible the most pressing problems. This paper fits in that mold. The goal here is to evaluate the performance of the SCF list sample within the top 1 percent of the wealth distribution, as viewed directly from the survey or as inferred from the wealth index models that sustain the sample design. This investigation may also shed light on some concerns about the approach taken by in [1].

The first section of the paper reviews the structural elements of the SCF that are needed to understand the technical support for estimates of the wealth distribution from the survey. The second section discusses the measure of wealth used in the paper and compares values from the survey with seemingly comparable values available in other sources. The third section presents a time series of estimates of wealth distribution and supporting portfolio structures observed in the survey and it highlights the importance of the list sample in the estimates. The fourth section looks more narrowly at the correspondence of measures of concentration at the top of the wealth distribution resulting from the survey measures with comparable estimates using the proxies for wealth estimated using statistical records derived from individual tax returns; the section also addresses the effects of nonresponse and sampling on these measures. The final section concludes and points toward further research.

2. Background on the SCF

It is impossible to have a clear understanding of how reliable estimates with a survey might be, without having at least some knowledge about the supporting technical framework. This need is greater than usual when the estimates in question concern descriptions of a rare population that is difficult to interview and that tends to have unusually complicated circumstances – as is the case for very wealthy households in the SCF. This section reviews much of the technical framework of the SCF, with a particular focus on the special sample that provides the survey with the most information on this population.

The SCF is specifically designed to provide reliable estimates of wealth and income for households.[2] Its questionnaire collects detailed information on assets, liabilities and income, along with a wide variety of related attributes.[3] Questions are carefully sequenced and framed to support a clear understanding of concepts by survey respondents. The implementation of the questionnaire as a CAPI (computer-assisted personal interviewing) program supports a variety of real-time data checks as well as a facility for capturing information on ranges around dollar values when the respondent is unable to provide a more precise estimate. The program also includes a means of recording auxiliary comments helpful in understanding or evaluating the information provided in each interview.

Field work for the survey is typically concentrated between spring of the survey year and the end of the year, though sometimes attempts at data collection for a small fraction of cases continues into the following year. Since 1992, the survey has been conducted for the Federal Reserve Board by NORC at the University of Chicago. Intensive training is undertaken to ensure that interviewer are able to identify and persuade the appropriate respondents, navigate the survey instrument, and assist the respondents in providing reliable data. A typical interview takes about 90 minutes, often over more than one session. Missing data in the survey are addressed with multiple imputation, which allows one

[2] Strictly speaking, the SCF organizes its data collection around a "primary economic unit", not the household. That unit is defined as a person or couple economically dominant within the household (or a specifically designated person or couple assumed to be economically dominant, in the case of the list sample) and all others in the household economically interdependent with them. For people in a household not included in this group (e.g., roommates, economically independent parents or children of the core person or couple, etc.), summary information is collected; that information is not ordinarily included in calculations of wealth from the SCF, as is also the case in this paper. In some cases, partnered survey respondents report that they hold their assets independently of their partner; for SCF purposes, respondents are asked to provide answers that include the partner and to involve the partner in answering the questions, where possible.

[3] See [3] for information on the most recent survey at the time this paper was written.

to estimate the variability of results as a consequence of item nonresponse.[4]

Wealth is highly skewed in the U.S. For that reason, a sample design that did not make some allowance for targeting wealthy households would include few of them, and the consequent sampling variability of wealth estimates dependent on the upper tail of the distribution, such as means or concentration estimates, would tend to be very large. Wealthy households also tend to be more difficult to get to participate in surveys.[5] In some cases, wealthy households express strong reservations about their privacy, but this view is also often expressed by other households. Two of the largest sources of difficulty in interviewing wealthy households are gaining access to be able to persuade the respondent to participate and then finding a time when they will agree to be interviewed. The amount of effort involved in this process of contact and persuasion can be far more than what is required in other cases. A survey that did not have some means of at least approximately segmenting the sample by wealth levels to be able to target effort to gain sufficient participation across all groups would suffer from differential nonresponse; unless there were some means of adjusting for those differences through weighting, upper-tail-dependent estimates would be biased.

Data collected in the SCF are subjected to a thorough review, driven primarily by comments recorded by interviewers and an algorithmic inspection of the data.[6] This editing exercise results in some substantial shifts in the estimated distribution of wealth.[7] The practice of the survey is to flag data edits to allow users of the data to identify values that have been changed in this way. Because editing is so labor intensive, there has never been a systematic attempt to perform sufficient double-blind work to gauge the variability in decisions and the possible consequences for estimates using the data. Nonetheless, it is at least clear that a failure to edit at all would result in many highly distorting data values remaining in the data set.

The SCF employs a dual-frame sample design, including a national area-probability sample to provide robust coverage of households with characteristics broadly distributed in the population, and a specially designed list sample to select wealthy households disproportionately.[8] The area sample derives from a systematic process for selecting housing units and households within units.[9] The list sample is selected from statistical records derived from tax returns, under an interagency agreement with the Statistics of Income Division (SOI) of the Internal Revenue Service, which requires strong provisions aimed at protecting the privacy of sample members and the confidentiality of their information.[10] The records available for selection in the SOI data are tax units – that is, individuals or married couples.[11] Although there is an attempt to adjust for the possibility of multiple tax filers in a given household and to adjust the SOI weights to account for married taxpayers filing separately, the procedure is imperfect; however, the approach appears most reliable among the wealthiest few percent of households

[4] See [4] for a description of the approach to imputation in the survey.

[5] See [5] for an analysis of relevant paradata on the participation of wealthy households in the SCF.

[6] See [6].

[7] See [7].

[8] See [8] for a description of the area-probability sample and [9] for a discussion of the use of address sequences from the U.S. Postal Service for replacing field listing of dwellings for the sample selection. See [10] for a detailed description of the overall sample design for the SCF and how it integrates the area-probability and list samples.

[9] The total number of cases in the area-probability sample (number of participants/out-of-scope sample cases) has varied over time: 1989: NA (2,937/NA); 1992: NA (2,456/NA); 1995: 4,756 (2,780/561); 1998: 4,921 (2,813/651); 2001: 4,993 (2,917/705); 2004: 5,339 (3,007/963); 2007: 5,227 (2,915/701); 2010: 8,651 (5,012/1,338); 2013: 7,900 (4,568/991). For the area-probability sample, out-of-scope cases include ones where the dwelling was vacant (including seasonal dwellings), demolished, no longer used for a residential purposes, or where there was no occupant aged 18 or older or the original listing of dwellings used in the sample selection was in error.

[10] Federal Reserve staff are never given the name of any of the taxpayers; the survey contractor is given name an address information necessary for contacting the sampled unit, but the contractor never receives information about income; SOI is never told who among the sample actually participated and they are given no link between the survey participant and tax data. Unlike the case of respondents in the area-probability sample, list sample respondents are given an opportunity to decline participation before they are ever contacted by an interviewer. After the completion of the survey, the contractor is obliged to destroy all information on the names and addresses of the list sample respondents. The total number of cases in the list sample (number of participants/out-of-scope sample cases) has varied over time: 1989: NA (866/NA); 1992: NA (1,450/NA); 1995: 5,720 1,519/54); 1998: 5,762 (1,496/45); 2001: 5,200 (1,532/34); 2004: 5,162 (1,515/29); 2007: 5,151 (1,507/35); 2010: 5,164 (1,480/53); 2013: 5,155 (1,458/43). Out of scope cases for the list sample include ones where the listed name (or names) is deceased and there is no surviving spouse or the listed name (and spouse, where relevant) is no longer a U.S. resident.

[11] Where the originally selected filer was a couple filing a joint tax return who became divorced or separated by the time of the SCF interview and maintained separate finances, both parties are treated as separate units for statistical purposes and an adjustment is made to their wealth index values to approximate a division of assets.

that are the primary focus of the list sample.[12] Thus, the two samples are complements, each providing strength where the other is weak.

For reasons related to cost control, the area-probability sample is selected to include a limited number of high-level geographic areas, which are selected in a multi-stage process. These areas include the largest metropolitan areas (such as the New York consolidated metropolitan statistical area) with certainty and a probability sample of other metropolitan areas and non-urban counties. Despite the fact that more such areas are selected than there are states in the U.S., not every state is present in the sample; the stratification involved in selecting the areas provides a balanced representation of characteristics of geographic areas within regions.

Because much of the analytical content of this paper turns specifically on the design of the list sample, it is worthwhile here to provide a higher level of detail about its design. The SOI data that serve as the frame for the list sample are themselves based on a sample from the universe of all individual tax returns.[13] Fortunately for SCF purposes, the sampling rate in the SOI file among the population most relevant for the list sample is sufficiently high that sampling error from that source is not a serious concern. The SOI sample that defines the population for the list sample is derived from individual tax returns filed in the year before the SCF.[14] Almost always, the income in the tax return is that for the prior tax year – that is, for the 2013 SCF, the SOI sample used to define the list sample universe predominantly contained income for 2011.[15]

Some initial pruning of the SOI sample is preformed to align it as well as possible with the target population. Filers younger than age 19 are deleted from the sample, under the assumption that they are secondary filers in another household; such deletions are monitored for signs that the filer has substantial income. In addition, only cases corresponding to filers from the 50 states and the District of Columbia are retained and an effort is made to combine records likely to represent members of single household at one address. Again for reasons of cost control, the list sample selection is restricted to the complete set of counties composing the high-level areas selected for the area-probability sample. Earlier research has indicated that wealthy households are not distributed geographically in the same way as other households.[16] Wealthy people appear to be more concentrated in the largest metropolitan areas than the population in general, though there also appear to be areas of disproportionate concentration across the country, often in the location of a formerly important industrial or similar activity or an area usually associated with retirement or entertainment. The list sample selection is always monitored to look for unusual shifts in the geographic distribution of potential sample members across the areas excluded at this stage, and growth in some such areas has led to some additions over time to the areas beyond those included in the area-probability sample.

An additional geographic complication with the list sample concerns the mobility of wealthy people.[17] The available evidence suggests that the wealthiest people are much more likely than others to live in multiple homes or to travel for extended periods for work or pleasure. This tendency further complicates the task of locating a respondent and persuading them to participate in the survey. Such mobility also suggests that sub-national estimates of characteristics strongly influenced by the upper tail of the wealth distribution may be less meaningful than is the case for other types of estimate.

An important definitional consideration for the list sample that might appear particularly important for this paper is the exclusion of members of the *Forbes* list of the 400 wealthiest Americans. This exclusion is motivated by practical concerns. Experience suggests that the wealthier a household, the more likely it is to be difficult to reach for any reason. The wealthiest SCF respondents are often surrounded by a large number of "gatekeepers" – assistants, attorneys, guards, servants,

[12] See [11].

[13] See [12] for a current description of the SOI individual sample design.

[14] The list sample frame does not include households that do not file an individual income tax return. Although some very low-income households do not meet the requirement for filing a return, they may nonetheless file in order to receive a refundable tax credit. Some other with low income, such as some people receiving only income from Social Security may not file a return. While the SCF is designed to cover the full range of filers, its sampling rate at the bottom of the distribution used for sampling is low, and the shape of the wealth distribution of the least wealthy part of the distribution is almost entirely determined by the area-probability sample.

[15] A relatively small number of cases in the SOI file represent returns for earlier years, typically amended returns. In processing the data for the SCF sample, only the most recent return for a given taxpayer is retained.

[16] See [13].

[17] A further complication is that, contrary to IRS instructions to report a home address on an individual tax return, many returns are filed using the address of an accountant, lawyer or some other professional advisor.

and sometimes literal gatekeepers – each of whom is tasked with protecting the privacy of their employer. Not surprisingly, the level of success in contacting such respondents is not high. Attempting to reach people even more protected would be prohibitively expensive, particularly in light of the slender prospect of obtaining an interview. In addition, because members of the *Forbes* list are so well known, it would be extremely difficult to release data from an SCF interview without either compromising their confidentiality, and masking the data to a sufficient extent via statistical techniques would greatly reduce its value. Nonetheless, it is often the case that the SCF includes participants whose wealth would be sufficient to place them in the *Forbes* list, but who were not present in that list.[18]

Table 1 provides a time series of the ratio of the total wealth of the *Forbes* list to the total of household wealth as estimated from the corresponding year of the SCF.[19] The ratio ranges from about 1.5 percent in 1989 to 3.1 percent in 2013. For at least two reasons, these figures may overstate the wealth omitted from the SCF. First, as noted further in the discussion of SCF weighting below, the adjustment of the list sample at that stage takes account of the full distribution of the observations in the SOI file, including any members of the *Forbes* list. Second, the *Forbes* estimates are based on self-reports, public records of ownership and other sources, and the estimates sometimes refer to extended families or other funds held through such structures as foundations; for such reasons, those estimates may differ materially from what ideally would be recorded in the SCF for the *Forbes* list members.

The list sample is stratified by a "wealth index" defined using data available in the SOI data. In the 1989–1995 SCFs, the index (WINDEX0) was based only on a capitalization of income flows in the SOI data, similar to the approach of [1].[20] The model is described in the equation below, where Y_i is capital income of type i, r_i is the corresponding period-specific average rate of return, K is the absolute value of total capital gains or losses, and H is an average housing value estimated for a range of income classes from the previous SCF and adjusted to the year of the SOI data that determines the list sample universe (observation-specific subscripts are suppressed for clarity). For each survey, the rates of return are taken from published period-specific sources. Strictly speaking, capital gains do not have an obvious place in a capitalization model of wealth; the intent is to capture some indication of assets that are held largely for realized gains and that do not yield appreciable dividends or other returns.[21] The model does not include any allowance for debts; because according to SCF data, average debt is a far smaller proportion of average assets among the wealthiest households than is the case for other households (about 2.5 percent for the wealthiest 1 percent, vs. about 14.6 percent overall in 2013), this omission seems unlikely to be a serious source of distortion for the intended purpose.

$$\mathit{WINDEX0} = \sum_{i=1}^{n} \frac{Y_i}{r_i} + K + H$$

Beginning in 1998, a more elaborate version of the wealth index (WINDEXM) was introduced. This model combines the original index with another index (WINDEX1) derived from a regression of wealth observed in the preceding survey on the income and other characteristics available in the SOI data used in the design of that sample.[22] The value of WINDEX1 for the next survey is obtained by projecting the model estimates on the values in the more current SOI data. An advantage of this approach is that it allows for a more complicated modeling allowing implicitly for some differences in rates of returns, patterns of holdings of debts or assets that do not yield returns and other factors, that may be correlated with the information available in the SOI data including the income variables used in the WINDEX0 model and others such as the age of the tax filer or the amount of wage income, property taxes or charitable deductions. A defect of WINDEX1 is that it implicitly embeds rates of return in the estimated model parameters; unlike the case with WINDEX0, there is no direct way of updating rates of return to the time of the later sample selection. Thus, in times of changing rates of return, WINDEX1 would be expected to provide a noisier indication of the relative position of sample elements in the wealth distribution. In addition, WINDEX1 also embeds any

[18] These observations are suppressed in the public version of the data.

[19] See [14] and the October issue of *Forbes* for a discussion of the methodology used in constructing the *Forbes* 400.

[20] [15] followed a similar approach for estimating the wealth distribution from tax-based data.

[21] Obviously, the model would not be able to reflect gains-focused assets that are intended to be bequeathed.

[22] The match of the SOI data with the information from respondents is purely for purposes of estimating this model and it was done in such a way that identifying information was not included in the data set used for the regression. The regression is estimated using logarithms for net worth and other dollar values and converted to levels thereafter.

biases present in the previous survey reports. To mitigate the risks of misclassification in the use of either index, WINDEXM is computed as a weighted combination of the two, where the two indexes are standardized in levels to have the same median and interquartile range.[23] WINDEXM is then divided into seven strata corresponding to fixed percentile points of its distribution.[24] Five of the strata apply to only the top 5 percent of the WINDEXM distribution and three entirely to a group smaller than the 0.5 percent at the top of the WINDEXM distribution. Until 2001, the 100 highest values of the index were excluded from the possibility of being selected into the sample. The motivation for the exclusion was to eliminate at least some number of members of the *Forbes* list. In all years, the full set of selected observations has been reviewed by SOI staff to remove any members of the *Forbes* list and some other prominent people.[25]

Typically, in planning the list sample for each survey, the performance of the model is evaluated against its *ex post* performance in classifying wealth measured in the preceding survey. Wealth observed in the SCF is always highly correlated overall with WINDEX1 and WINDEX0, but it is more strongly correlated with WINDEX1. Although the observed wealth levels of sample elements in each of stratum overlap, sometimes substantially, the available evidence supports the claim that the design does achieve an oversampling of wealthy households. How the sample aligns with distributional characteristics at the top of the wealth distribution is the subject of the fourth section of this paper.

Income may vary from year to year in response to transitory events. Evidence suggests that there is substantial variability of sources of capital income among people who are at least periodically in the upper reaches of the income distribution. Thus, basing the stratification model for the list sample on only a single year of data would tend to be noisier than basing it on more than one year of data so that transitory variations could be smoothed. Beginning with the 2001 survey, the SCF list sample has employed multiple years of data to compute weighted averages of the values used in calculating the wealth indices.[26] In recent surveys, the weighted average has attributed 50 percent of the weight to the most recent data, 30 percent to the previous year of data and the remaining 20 percent to the data from the year preceding that. This choice represents a trade-off between reflecting the most current information available about a case and minimizing distortions as a result of lumpy realization of some types of income.

The SOI sample that underlies the SCF list sample is itself formally cross-sectional. But primarily for reasons of minimizing sampling variation in estimates of changes, the SOI sample is constructed in a way that there is much greater than random persistence of observations across years.[27] This persistence is strong in the population most relevant for the SCF list sample. Thus, there is a chance that the same household could be selected for multiple waves of the SCF. In order to avoid placing undue burden on respondents who tend to have the most lengthy and difficult interviews, participants in either of the two surveys preceding a given SCF are not allowed to be selected; people who refused participation in the preceding survey are also excluded. For example, about 25 percent of the base SOI sample for the 2007 SCF in the highest WINDEXM stratum were interviewed in the 2004 or 2001 surveys or declined to participate in the 2001 survey. The comparable figures for the next two lowest strata were about 6 percent and 2 percent, respectively.

The ability of the list sample to describe the reality of the wealth distribution is perhaps most obviously challenged by nonresponse to the survey.[28] By agree-

[23] Originally, WINDEXM weighted WINDEX0 and WINDEX1 equally. To reflect the better ex post predictive power of WINDEX1, as discussed in the text, the weighing was adjusted starting in 2004 to give more weight to WINDEX1.

[24] The design in 1989 included only 6 strata; the 1992 and subsequent designs further segmented the top of the distribution.

[25] Earlier examination of the relationship of the income and wealth of the Forbes list reported in [10] indicated that many such people did not have the levels of income that might be expected at the level of wealth reported by *Forbes*.

[26] See [16] for details. The 2001 and 2004 surveys employed two years of data and beginning with the 2007 survey the number of years was expanded to three. Although there is an attempt to match observations included in the SOI file that determines the list sample universe with observations in earlier versions of that file, or with information from the IRS Masterfile of individual returns in order to construct an approximate equivalent, sometimes it is not possible to make a match. A given observation may have changed marital status over years, it may not have filed an individual tax return in some year, or there may be errors or other difficulties in the matching. Where a match cannot be made, the base year information is substituted. All dollar values are adjusted to the reference year of the base data. In computing the wealth indices, a weighted average of value of each dollar value is used.

[27] See [12] for a description of the sample design.

[28] Throughout the history of the SCF, it has been recognized that it is critically important to monitor and understand to the extent possible the patterns and implications of nonresponse in the survey. More recently, guidelines in [17] impose a more formal obligation to study potential nonresponse biases in U.S. government surveys.

ment with SOI, all list sample respondents selected are given a right to refuse participation in the survey before they are ever contacted by an interviewer. Since the 1998 SCF, about between 10 to 13 percent of list sample respondents have so refused participation. In earlier years of the survey, the rate of such refusal was much higher – as high as 38 percent in 1989 – perhaps in part due to later changes in the presentation of the option or to changes in the extent to which people open or respond to mail from unrecognized sources. Generally in the surveys beginning with the 1998 SCF, the use of this option to decline participation has been tended to be lower, the higher the stratum of WINDEXM; still, for the highest stratum, the rate since 1998 has ranged from about 7 to 8 percent.

Despite intensive efforts to reach respondents who do not exercise their right of prior refusal, the rate of other incomplete cases (very largely refusals to participate) has been about 80 percent or more for the wealthiest stratum since 1998; taken together with prior refusals, this implies a response rate of between 8 and 12 percent over this period.[29] Although the overall response rate is progressively higher for lower WINDEXM strata, the refusal rate is still substantial; for example, in the 2013 SCF, the response rate aside from the top WINDEXM stratum ranged between about 12 and 56 percent.[30]

To some degree, these low response rates are driven by a system of quotas for completed cases. Each stratum has a target number of completed cases. Although there may be some degree of "work to quota" in the lower WINDEXM strata, it is clear from detailed examination of the associated information in the series of data maintained for each observation on contact attempts, that the overwhelming explanation of nonresponse in other parts of the list sample is the difficulty of reaching and persuading a respondent to participate.

While the SCF has been criticized for having such low response rates among wealthy households, unlike other surveys in the U.S. it also has the means of identifying and making adjustments for at least some aspects of the problem, through the design of the list sample and the associated frame data. Nonetheless, it would be naïve to assume that any set of simple adjustments would be capable of erasing all differences between the population and the set of participants. The more realistic hope is that weighing adjustments may reduce biases in estimates based on measures closely related to wealth to a degree acceptable for the core purposes of the survey.

The weighting process in the SCF entails separate adjustments to the area-probability and list samples; both sets of adjustments involve post-stratification and raking over a small number of variables.[31] The two samples are combined using a post-stratification procedure that weights the contribution of the list sample more heavily at the top end of the wealth distribution and the area-probability sample more heavily elsewhere; this treatment corresponds to the relative strengths of the two samples. The top of the wealth distribution is very largely determined by the list sample. Because this paper focuses on the performance of the list sample in estimates of the highest parts of the wealth distribution, only the construction of the separate weights for the list sample is discussed here.

Ideally, weighting adjustments for a repeating cross-sectional survey would be based on a stable set of characteristics across all surveys and any deviations would be minimal. Particularly when adjusting a population with such extreme skewness, as is the case with the wealth represented by the SCF list sample, small changes may have large effects on distributional estimates. For the 1989 survey, the response patterns in the SCF list sample were examined to look for characteristics that may be correlated with nonresponse or for changes in such characteristics, and these results have been reviewed for subsequent surveys.[32] From this work, a small set of characteristics was selected as the framework for weight adjustments. A set of consistent analysis weights based on this framework is available for all of the cross-sectional surveys beginning with 1989. In brief, the adjustment process includes a ratio recalibration of the initial sampling weights of individual observations within each WINDEXM stratum to the relevant population totals computed from the entire file that served as the frame for the sample; from that base, the weights are further post-stratified to regional totals, and a raking procedure is defined around

[29] Typically, a very small fraction of observations is detected to be ineligible – because they live outside the U.S. or because they are deceased with no surviving spouse.

[30] Although much higher than the response rates for the highest strata, the response rates for the lowest WINDEXM strata are much lower than that of the area-probability sample, the great majority of which corresponds economically to these WINDEXM strata. Part of this difference is explained by the fact that the area-probability cases tend to be much more highly clustered than the list sample cases, which range randomly across an entire metropolitan area. Such dispersion can lead to serious operational inefficiencies.

[31] See [18] for a description of the weighting methodology.

[32] See [11] for a discussion of the first such investigation.

the WINDEXM strata, a set of categories of "financial income" and a categorical variable representing Census regions crossed with a binary measure of a high degree of urbanicity. Financial income, which is defined as the sum of dividend and taxable and nontaxable interest income, has been a consistently important predictor of nonresponse. Certainly, other approaches could be considered. The process deliberately avoids deeper intervention, to support a procedure that can be applied consistently across all the surveys and that does not induce extreme variations in weights that would inflate the sampling error. The set of adjustment factors used at each stage of adjustment is captured in the weighting process and this information is recorded and inspected for each survey; adjustments that differ notably from those in prior surveys are investigated.

The information presented above in this section highlights the sequence of approximations and adjustments made in an attempt to define a frame for the list sample, select the sample and align the set of participants with the a measure of the true population. Because the wealth distribution is so highly skewed in the region most intensively covered by the list sample, small irregularities in that process may lead to relatively large shifts in estimated wealth levels. Comparisons with other sources of data, given in the next section of this paper, provide comfort that the survey estimates and high-level aggregates are in a reasonably near vicinity of each other, but differences remain and the differences fluctuate through time.

Only some of those differences are potentially explained by differences in procedure or concept. Undoubtedly, some of the differences are due to sampling error and various sources of nonsampling error, as well as possible sources of bias in the treatment of the list sample. The SCF provides a means of estimating sampling error through the use of replicate weights.[33] For the SCF, there are 999 replicate samples of the cases actually completed, selected from that set in a way that mimics the high-level selection of the area-probability and list samples; the associated replicate weights for each replicate are constructed for the list sample cases, the area-probability cases and both types of cases. However, the replicate weights (and the main weights as well) condition on the assumptions used in constructing the final frame for the sample selection, the selection process, and a fixed model of weight adjustment aimed at addressing coverage and nonresponse. That is, all variability from the replicate weights stems only from the extent of randomization in the selection from the completed observations.[34] Thus, confidence intervals for estimates strongly affected by the presence of the list sample are likely to under state the true confidence intervals by an unknown amount. Among sources of nonsampling error, there is currently only the means to incorporate directly some element of uncertainty due to item nonresponse, through the use of multiple imputation.

In addition to unmeasured sources of variability, there is always the possibility that undetected bias in some dimension(s) remains in the final representation of the list sample participants after the weighting adjustments. Except in trivial examples, it is not reasonable to think weights can "correct" for deviations from the true population across all dimensions as a result of coverage or response problems, and do so across the entire range of those dimensions. Moreover, the relevant population controls (or similar factors used as weighting targets) also may be subject to error or variability. Such possibilities for error exist for virtually any survey, though again, the amplification effects of extreme skewness in the upper tail of the wealth distribution may make them more worrisome for the SCF. Holding constant as many factors as possible in the chain of production of survey estimates, except where evidence strongly motivates a deliberate change, should help in minimizing extraneous changes in estimates. Thus, any change in design that affects one of the areas of approximation – particularly the definition of the approximation of the population by the choice of base year for the SOI data, the eligibility restrictions for selection, or the weighting adjustments – should be changed only with extreme caution, unless such changes can also be implemented as revisions to the earlier cross-sectional time series. Any such changes should be subjected to rigorous evaluation both before and after implementation.

3. SCF wealth and its comparison with other sources

"Wealth" is a term used frequently as if it had an unambiguous definition.[35] It is sometimes ambiguous

[33] See [19] for a discussion of the most recent revision of the bootstrap weight construction.

[34] An alternative approach would be to sample from the complete set of sample cases, not just the set of participants. Such an approach might provide a better representation of the variability introduced as a result of unit nonresponse.

[35] See [20] for one such framework specialized to what can reasonably be expected to be measurable in a survey.

even whether the term is intended to refer to income or to a measure of assets or net worth. In this paper, wealth is taken to be a particular construction of net worth, defined as assets less liabilities. That construction itself rests on a variety of assumptions about both assets and liabilities, in terms of their scope, definition and valuation.[36]

Here, "assets" is taken to include a set of items of value that are under the control of a household. That set includes checking or savings accounts of a variety of sorts, certificates of deposit, money market accounts, other mutual funds, hedge funds, stocks, bonds, the cash value of life insurance, annuities with a cash value, other "designated retirement assets", real estate, personal businesses, vehicles, and miscellaneous assets. Designated retirement assets include tax-preferred Individual Retirement Accounts, Keogh Accounts and tax-preferred accounts associated with employment (sponsored by an employer, union or similar association) over which the owner has the right to make current or future withdrawals.

Importantly, the measure of assets does not include a number of items of value that are not under the direct present or future control of the person owning or entitled to benefits from the item. Such items include defined-benefit pension plans from which the beneficiary has only the prospect of future income, other annuities that cannot be liquidated, other future income benefits such as Social Security, and other contingent or expected benefits or income that cannot be sold, such as lifetime rights to live in a dwelling without cost or lifetime income rights in a trust. Such excluded items may have a substantial economic value to some people, either in the present or the future. Human capital, the embodiment of future earnings possibilities, might be consider the largest asset for many people. One approach might be to capture such values through a present value calculation. Unfortunately, the amount of information required to for such a computation is often large and inevitably many assumptions, such as future rates of interest or inflation, future work decisions, or life expectancy are needed. Because there is no standard set of such assumptions to rely on, even if the necessary data were available, it would be necessary to consider a variety of possibilities for the necessary assumptions to give a reasonable indication of the sensitivity to the assumptions. The complexity of that task is beyond the scope of this paper. There many other, more peripheral possibilities for valuation, such as future inheritances, the insurance value of social relationships, and participation in public goods (clean environment, safety, transportation systems, etc.) that are also excluded in the definition of assets here.

In principle, survey respondents report the value of each of the assets of their household as of the time of the interview.[37] For some assets, such as bank accounts or publicly traded securities or funds, there is generally an unambiguous current value that would be available to a survey respondent willing to consult the appropriate source. In the 2013 SCF, 42 percent of respondents reported checking records for the report of some value during the interview; among the wealthiest one percent of respondents as estimated in the survey, 54 percent did so. Undoubtedly, many respondents make their own estimates. Such estimates may be particularly difficult where the relevant asset does not have a clear, current market value – as is generally the case, for example, with houses or personal businesses. The survey asks what the household would get if it sold such an asset. To obtain a definitive value, a survey respondent would most often need to take the extreme step of putting the asset up for sale. Although respondents may have a personal expectation of the value, that expectation may be influenced by a variety of factors, and there may be more of a range of subjectively plausible values than a single value. One factor that may introduce variance (or perhaps bias relative to some standards) in the expectation is the underlying search process assumed. For example, a value might be reasonably be taken to be relatively high if a potential seller is willing to wait for someone with a higher than average draw from the distribution of potential bids. For someone with no interest in selling, there might be a relatively high reservation price. In addition, there may be non-pecuniary aspects of such assets or idiosyncratic aspects of value to the owner that would lead them to hold a reservation price that would exceed any reasonably expectable offer price. In some cases, SCF respondents report values as ranges, rather than a single value, and this response behavior may reflect their evaluation of the true underlying range of

[36] The assets and liabilities included in the net worth measure here are more precisely defined in Bricker et al. [3] or more briefly in appendix Table A3.

[37] Respondents are encouraged to consult other household members when that might help in providing more accurate responses. For example, credit card use, work history and pension rights are often thought to be best known at the individual level. Unfortunately, the survey contains no systematic information on when someone other than the respondent provided information during an interview.

possibilities, rather than simply uncertainty about the exact value. Overall, it seems likely that the variability attributed to estimates based on reported values understates the true variability and if there are systematic differences between reported values and values that might be computed as an average on the basis of fundamentals, there might also be bias in the estimates from that perspective.

The term "liabilities" in this paper is taken to encompass debts directly associated with assets – including mortgages, home equity lines of credit, vehicle loans, margin loans, loans against cash value life insurance, and loans against pension accounts – as well as debts not tied to assets. The latter type includes credit card balances, unsecured lines of credit, education loans, other types of consumer loans, and miscellaneous debt, such as loans from friends of family. The SCF also collects information on loans to or from a personal business, but for present purposes, such businesses are treated as having a value net of such loans. Similarly, mortgages for nonresidential real estate are netted against the value of such properties.

With the exception of credit card debt, the value of each debt is taken to be the amount outstanding as of the time of the interview. For credit cards, the value is taken to be the value as of the time after the payment on the most recent bill. For all except the most miscellaneous personal debt, there should usually be a reasonably current statement of loan balances available in principle. However, there is a subtle way in which the treatment of the valuation of debt is not parallel to that for assets, which have a value that is, at least in principle, determined by the current market. One could also consider the value of debt without a floating interest rate to be a function of the current relevant interest rate, as would be appropriate for a debt security. In that scenario, a rise in market rates for a fixed-rate loan held by a household would lead to a decline in the value of the debt. Although interest rates are recorded for most debts in the SCF, the appropriate reference rate equivalent for the period is often not clear. In addition, there are generally not common instruments most consumers would turn to directly in order to realize any gains or losses under this valuation approach, and it is clear that some consumers do not refinance debt even when there appears to be a substantial current benefit to doing so.

As is the case with most surveys, there is no readily available external source of information for validating directly the survey responses on wealth in the SCF. However, something may be learned from comparing results with the estimates of aggregates from national accounts or with estimates from other surveys. [20] compares SCF estimates with aggregate estimates for the household sector in the Financial Accounts of the United States (FAOTUS).[38] The paper finds that SCF and FAOTUS levels of total net worth (defined over comparable items) are very close from 1989 to 1998, but beginning in 2001 the SCF estimated total net worth begins to exceed the FAOTUS estimate. This shift is driven largely by higher SCF estimates of owner-occupied real estate and noncorporate businesses and lower SCF estimates of consumer credit. Conceptual differences explain some of these differences. The value of financial assets was substantially understated in SCF relative to FAOTUS in the years before 2001; the direction of difference reversed in 2001 and then the difference declined to a moderate level of understatement. Holdings of financial assets may be particularly affected by the inclusion of nonprofits in the household sector of FAOTUS. There are more substantial differences in some disaggregated categories, but a variety of factors make it difficult to draw a firm conclusion about the source of the differences. However, the two sources show generally similar trends.

Survey reports of detailed types of income may also provide insights into the extent to which the survey sufficiently represents economic reality. Part of the SCF collects income using a framework that is designed to be comparable to that of an individual income tax return. [23] compares SCF income reports with SOI estimates for the period covered by the 1989 to 2004 SCFs.[39] The paper finds that the SCF tends to over-estimate total income relative to the IRS data by somewhat less than 10 percent, but categories of income other than wage and salary income show more substantial differences. The SCF estimates tend to understate interest and dividend income and income from pensions, annuities and Social Security by varying fractions, and to overstate similarly a broad measure of business income, unemployment insurance and alimony, and a miscellaneous category of other income. Some conceptual differences in the survey, variations in the use of records by SCF respondents, and other misclassifications of income may explain some of these differences.

The ability of the SCF even to come close to aggregate estimates of wealth or income is critically dependent on the list sample. Without the list sample,

[38] See [22] for a critical discussion of the FAOTUS.

[39] Each survey contains information on income for the preceding calendar year.

the SCF would, in effect, provide significant information on the full population less approximately the wealthiest one percent.[40] This wealthiest group holds approximately a third of total net worth, so underrepresentation of that group would yield very large differences and results would be subject to large sampling error and potentially other sources of error. [25] compares a variety of wealth estimates for the SCF and the Panel Study of Income Dynamics. That paper finds that the list sample explains the largest difference between the estimates and that away from the top of the wealth distribution, the surveys give reasonably comparable results.

4. History of wealth and income distributions in the SCF

The SCF data have often been used to study the distribution and composition of wealth and income.[41] This section updates an earlier series of estimates of wealth and income shares and discusses the role of the list sample in supporting those estimates.

For the surveys from 1989 to 2013, Table 2 shows the fraction of households estimated to have negative net worth, zero net worth or positive net worth, using the combined area-probability and list samples or the area-probability sample alone. From 1989 to 2007, the data for the combined samples show a fairly steady proportion of households, from about 7 to 8 percent, followed by a jump to over 11 percent in 2010 and 2013, following the Great Recession. Over the full period covered, the proportion with exactly zero wealth tended to decline. Although there may be some error in classifying households with zero or negative wealth, even when the two groups are taken together, there is still a jump in 2010 that persists in 2013.

Table 3 shows estimates of the proportion of wealth held by various percentile groups over the same period covered by Table 2. The shares of wealth held by the groups covering the least wealthy 90 percent of households declined significantly in 2013 relative to the shares measured in the surveys from 1989 to 2007; there was little change for these groups from 2010 to 2013. A logical consequence of the changes for the least wealthy 90 percent are offsetting changes for the wealthiest 10 percent; changes for the subgroups within that wealthiest group, however, are less clear. For example, only the 1989 and 1992 shares for the wealthiest 1 percent are significantly lower than the corresponding share in 2013 and only the 1995 and 1998 shares of the next wealthiest 9 percent are significantly lower than corresponding share in 2013.

Figure 1 shows a relative quantile-difference plot for net worth from a 1989 baseline to 2007, 2010 and 2013.[42] The figure shows clearly the advances made across the broad middle of the distribution from 1989 to 2007, despite the much larger percentage gain at the top of the distribution. It also shows the diminution or loss of those gains across the middle by 2010 or 2013, but with some gains relative to 1989 still surviving for the group above the median.

The comparable story for total pre-tax household income, shown in Table 4, is similar in some ways, but there are notable differences. This difference from the pattern seen for wealth may reflect systematic factors in the weighting of income components in the wealth indexes, changes in income since the time of the data used for the list sample design, conceptual differences in the ways survey respondents conceptualize or report their incomes, or other technical factors.

Figure 2 shows a relative quantile-difference plot for total household income, comparable to the Fig. 1 for net worth. The data show a U-shaped pattern of largely positive changes from 1989 to 2007, 2010 and 2013, with the figure being roughly nested in chronological order and the change to 2013 being closest to the zero line. The center of the distribution from the 20th to the 60th percentile saw approximately no change from 1989 to 2013. Estimates for the top few percent are strikingly lower for the change from 1989 to 2010.

Underlying the differences in the wealth concentration estimates are many differences in the estimated portfolio structures of households across the wealth spectrum. For each survey, starting with 1989, appendix Tables A1.1–A1.9 break out full-sample estimates of the amount and share of various assets and liabilities for each of the wealth percentile groups in

[40] See [24].

[41] For example, see [1,26–30].

[42] The plot shows the percentage difference in the value at each quantile of the distributions (the value for a given year minus the value for 1989) as a percent of the 1989 value. All dollar figures reported in this paper are given in 2013 dollars, adjusted where necessary by the CPI-URS price series. The figure suppresses the range below the 20th percentile; in this range, small absolute changes yield such large proportional changes as to make the figure difficult to interpret. See [31] for a detailed discussion of the wealth of the lower half of the wealth distribution. The dots around the black line in the figures indicate the 95 percent confidence intervals around selected percentiles for the change from 1989 to 2013.

Table 3 and appendix Tables A2.1–A2.9 provide estimates of the percent of households within each of the wealth groups holding the assets and liabilities.

According to the full-sample estimates, the wealthiest one percent hold a strongly disproportionate share of every type of asset or liability considered, except credit card debt. Business assets figure particularly strongly in the assets of the group: the great majority of the group owns a non-publicly-traded business (80.5 percent in 2013) their holdings are well over half of the total value of such businesses (64.1 percent in 2013). They also hold near half or more of bonds, stocks, mutual funds other than money market funds, and managed assets such as trusts or managed investment accounts. Although the group tends to have the highest rate of ownership of account-type retirement accounts (88.7 percent of the group, compared with 49.2 of families overall in 2013), their share of the total was much less than 20 percent, as might be expected from the limits imposed on contributions to such plans. They have near universal homeownership (97.0 of the group, compared with 65.1 of families overall in 2013). The group is less likely to have debt overall (66.3 percent of the group, compared with 74.5 percent of families overall in 2013) and the share of the total is disproportionately less (5.4 percent in 2013).

Most the remaining assets and debts are held by the rest of the top half of the wealth distribution. Houses become a much larger fraction of assets for these groups. As noted earlier, the least wealthy 10 percent is typically dominated by families with negative net worth; though houses, retirement accounts and vehicles are common assets for that group, they tend to be outweighed by debts.

5. Comparison of SCF and tax-based wealth estimates

There has been a growing sense that as surveys become more expensive and more problematic, particularly in terms of the difficulty of persuading people to participate and to provide good answers, various forms of alternative data, often administrative data, may provide timelier and less expensive estimates to supplement or even replace surveys. But often such sources are based on a different reference population and/or a different conceptual basis for the variables of most interest. Typically, the most immediately pressing questions in using alternative sources concern whether the reference population can be "inverted" or otherwise mapped with minimal error into a desired framework and whether there is some stable or otherwise reliably measurable method of mapping of variables into the desired conceptual framework.

Of relevance to this paper, [1] uses SOI micro data on personal income to estimate a time series of net worth distributions starting in 1913, by applying a set of fixed capitalization factors for each year to apply to the various components of capital income. Taken at face value, their impressive results suggest that the SCF understates the concentration of wealth at the very top of the wealth distribution. However, like the SCF their results rest on a set of approximations and assumptions. The capitalization factors they use are developed from FAOTUS data on household wealth components and SOI income figures. They also argue for a series of adjustments to address wealth not reflected in reported income. In the process of creating their simulated distributions, they do not appear to make any adjustment for the unit of observation. Although the resulting series refer to estimated characteristics of tax filers, the authors compare wealth concentration estimates for with those obtained from the SCF, which provides such estimates on a household basis.[43] The tax-income-based estimates appear to include filers who are not U.S. residents. In addition, as [2] notes, because the number of tax filers is substantially larger than the number of households, the quantile points of the estimated distribution do not necessarily align with those that would be obtained from tax data aggregated at the household level, without further adjustment. As noted earlier in this paper, because very high income filers are much less likely to file their taxes as married filing separately than are other households, the 99th percentile of taxpayers corresponds to a percentile below the 99th percentile of households, and that point may vary over time as a function of changes in filing arrangements. In the revised version of their calculations published, Saez and Zucman account for this difference and continue to find an understatement of wealth concentration at the very top of the wealth distribution as measured in the SCF.

The mapping from income to wealth may also raise issues. Although the authors do an admirable job of testing the robustness of their assumptions, they do not give a measurable indication of the inherent range of plausible assumptions and their effects on the level of estimated variability of estimated wealth values at the

[43] Saez and Zucman approximate the wealth of nonfilers as zero.

observation level or overall. Good estimates of variability may be particularly important in light of the possibilities for very large distortions in capitalizing such highly skewed capital income data. [2] also questions specific adjustments and capitalization factors, some of which may have a substantial effect on estimated wealth values and their distribution. They also note that Saez and Zucman might obtain somewhat different results if they smoothed the data over multiple years, as is the caser for the SCF list sample design.

Whatever problems or concerns there may be with the approach of Saez and Zucman, it does seem intuitively sensible that there is substantial information in the SOI data that should be exploited, especially if that information could be augmented with other data matched at the micro or small-area level that could help with more record-specific estimates of wealth. For such an approach to have greater appeal, further work is needed to test estimates against observable wealth and to develop a means of estimating or approximating the variability of the estimates.

This section of the paper does not pursue such work directly, but it does provide insights into the relationship between net worth measured in the SCF and the wealth indices based on tax data that support the SCF list sample. The WINDEX0 measure that is one component of the list sample stratifier is similar in concept to the Saez and Zucman measure, though the WINDEX0 capitalization factors are taken from various market sources, rather than estimates made indirectly. Unfortunately, the "true" measure or wealth is not known in general. So the results of this section reflect as much on the SCF measurement process (particularly reporting error, nonresponse and sample design considerations) as they do on income-based projections of wealth. The available data at least make it possible to provide sufficient information to suggest some explanations for differences between survey-based and tax-based wealth estimates.

Ideally, the survey would support estimates of the wealth distribution via a sample constructed using observable characteristics that have a mapping with minimal variance to a conceptually unambiguous measure of wealth, and that clear structure would not be disturbed by problems in implementation, such as unit nonresponses or misreporting answers. In practice, we have an approximate indication of wealth from the wealth index models, the appropriate measure of wealth is subject to a variety of interpretations, and unit nonresponse is non-negligible; moreover, as noted earlier reporting errors large enough to have substantial effects on the estimated wealth distribution are not rare. Problems about potential bias and lack of statistical efficiency might well be expected in such circumstances.

It is important to note that the design of the list sample in itself is only a framework for identifying pools of observations that are then sampled. If the sampling mechanism simply misclassifies observations in some way relative to the desired measure of wealth, that on its own would only make estimates of the observed wealth distribution less statistically efficient – it would not directly induce bias. It is possible, however, that imperfect stratification could result in bias due to less adequate adjustment for nonresponse in the construction of the analysis weights, owing to the apparent differential nature of nonresponse over wealth groups.

The most obvious sources from which bias may arise in this context are errors in reporting answers in the survey, mis-coverage of the population (undercoverage, over-coverage, or elements of each), and nonrandom unit nonresponse along with inadequate or inappropriate weighting adjustments.[44] Reporting errors may include mistaken answers, answers based on inattention to the questions, deliberately incorrect answers, or use by respondents of a conceptual framework that differs in important ways from that intended in the survey design. Imputation may fail to address item nonresponse adequately if there are systematic factors related to the act of nonresponse that are correlated with the unreported value. For example, respondents with unusually valuable assets might be differentially unwilling to reveal the fact; unless some other observed factors correlated with this behavior are available for all cases along with correctly reported wealth for some cases, imputations in such circumstances would tend to be too low relative to the true values.

Coverage errors may arise from the dated nature of the list sample frame relative to the time of the interview, from approximations made in the attempts to align the list sample frame as much as possible with a concept of U.S.-resident households, or from errors in filing or processing the individual tax returns that form the basis of the statistical records used for the list sample. There may be undetected instances where incorrect households were interviewed, but treated in the processing of the survey weights as identical to the

[44] There might also be a variety of problems related to administration of the survey in the field or processing of the data after collection.

cases actually selected.[45] As noted earlier, unit nonresponse is very substantial for the SCF list sample, and for good reasons, only a few controls are used in the weighting adjustments intended to ameliorate this problem. If the nonrespondents differ from the survey participants in a systematic way that is not captured through the weighting adjustments – for example, by disproportionately having high levels of wealth in categories that generate no or very small corresponding incomes – there will be bias in the estimated wealth distribution.

Another issue that may be important for the rarefied extreme upper tail of the wealth distribution is the granularity that emerges for that group. Typically, we think of sampling from a population dense with comparable population elements. When similar elements become sufficiently rare, heterogeneous or otherwise "far apart" in the population, it becomes more likely that important areas may be missed misstated, either as a result of sampling or nonresponse. Such problems may be especially severe in the final, open-ended stratum used in the selection of the list sample, for which unit nonresponse is highest.[46] Even aside from the potential for bias, estimates of variability based only on the observed data may be inadequate.

A substantial research paper would be necessary in order to dig into each of these issues in detail. The point of this paper is to look more narrowly at the implications of the design, execution and processing of the SCF list sample on estimates of the distribution of wealth and the variability of those estimates. In what follows, the paper explores differences in the distributions of measured wealth and the wealth indices, in an attempt to identify potential biases or sources of variability that may not currently be accounted for. The focus is on the top 1 percent of distribution of net worth or of one of the wealth indices; this is the region within which the list sample is most dense and thus where it makes its greatest contribution to the combined-sample estimates.[47]

To start, it is useful to compare the ordering of individual observations within the distributions of net worth and the wealth indices. This approach abstracts from any differences in levels across the distributions or the relative size differences within distributions. For each SCF cross-section from 2001 to 2014, Table 5 shows the mean and standard deviation of the percentile of the distribution of the various wealth classification schemes, given that an observation is in a particular one-tenth percentile group of the top one percent of another of the classifiers.[48] For example, the first column of the table shows the mean and standard deviation of the estimated percentile of entities in the distribution of WINDEX0, by its position in the estimated distribution of directly measured net worth. There is clearly substantial agreement in the ordering of observations under the different classifiers overall, in that the great majority of the means are at least above 99. But there are still some notable differences, with some figures substantially below 99 and with a lack of clear monotonic increases in the figures with increasing wealth classification groups.

Generally, the rank of survey net worth and the rank under the wealth indexes appear to be more closely aligned in the top half percent in the distribution of net worth than in the lower half of the group. One might expect that there to be some variability in the ordering between net worth and the wealth indices, if only because net worth is observed at the time of the survey, whereas the income figures used in computing the indices are for an earlier time. However, there is also substantial variability in the classification between WINDEX0 and WINDEX1, which overlap at least in terms of common income inputs.

Some part of the differences in ordering is surely due to deviations in the parameters of the WINDEX0 and WINDEX1 models from the specific values appropriate for the values observed for individual tax filers in the SOI data. Rates of return may be idiosyncratic. In extreme cases, an asset may lose its value entirely or increase in value at a rate far beyond average returns for the asset class. Even well within those extremes, the data collected in the SCF on assets and their associated income indicate great variation in rates of return implied by those related pairs, and this is particularly

[45] In principle, interview fabrication ("curbstoning") is an additional potential source of error, but this appears not to be an important factor for the SCF.

[46] Indeed, at a certain level of population granularity, there may be a strong argument for selecting such units with probability one. However, with unit nonresponse among such cases, it is not clear what, if any, weighting adjustment would be appropriate.

[47] It might be desirable to look more broadly at this group in the context of the entire wealth distribution observed in the list sample, but the number of observations below the estimated 95th percentile is too small for such a comparison to be meaningful.

[48] The analysis in this section focuses on the 2001–2013 SCF cross-sections, because the sample construction is most similar over that period, as discussed in the second section of this paper. The percentiles are computed using the relevant weights, but the means and standard deviations reported in the table are unweighted.

so for relatively wealthy households.[49] As noted earlier, the model underlying WINDEX1 is a regression of wealth observed in the previous survey on income (and other characteristics) from the SOI data used in creating the sample for that survey. As also noted earlier, while flexibility in that regression allows some implicit accounting for variations in rates of return across observations, those returns are particular to the estimation period; because the regressions coefficients are a combination of rates of return and other factors, there is no obvious direct way to "update" them for the succeeding survey. In addition, the rates of return assumed in the WINDEX0 model are specific only to the reference period of the SOI data. Because rates of return are embedded in the fundamentals that underlie many asset values, changes in relative returns between the reference time for the SOI data and the time of the survey may lead to WINDEX0 values that imply a different distribution of wealth than that in the period of the survey, and thus a different ordering.

Of course, even fair agreement across the measures on the ordering of observations does not necessarily imply that the various distributions agree as well on the relative concentrations of amounts. As a primary exploratory device, the paper uses a variety of estimates based on Lorenz curves to examine measurement at the extreme right-hand tail of the wealth distribution. A Lorenz curve plots the cumulative percentage of an item held by a population against the corresponding cumulative percent of the population. Loosely speaking, a Lorenz curve close in its midsection to the 45-degree diagonal reflects a more equal distribution than one further from the diagonal. Like the investigation of ordering, this approach offers the benefit that comparisons are independent of overall differences in levels across the wealth measures. Like the earlier results on ordering, the Lorenz curves shown in Figs 3 and 4 for net worth, WINDEX0, WINDEX1 and WINDEXM for 2001 and 2013 also include data only for the top one percent of list sample participants.[50] The estimates for each wealth measure are made using the nonresponse-adjusted weights computed for the list sample, as described earlier in this paper. Table 6 provides supporting information on the shares of the overall top one percent of net worth and the wealth indexes of the top one percent, attributable to each of the one-tenth percentile groups between the 99th and 100th percentiles.[51]

It is clear from the figures that the curves for net worth and WINDEX1 are most similar and they are mostly distinctly to the left of the curves for WINDEX0 and WINDEXM, which are also similar to each other. Because the WINDEX1 model is based on a regression of observed survey wealth on variables available in the sample frame, it might be expected to be at least somewhat more aligned on a conceptual basis with measured wealth than WINDEX0. The results imply that for the set of survey participants, the distributions described by observed net worth and WINDEX1 tend to be less concentrated than those implied by WINDEX0 and WINDEXM. In addition, there are substantial differences in the spreads of these curves in the two years. Moreover, there does not appear to be any tendency toward a smaller spread of the curves over time, as indicated by inconsistent variation across comparable plots for 2004, 2007 and 2010 (not shown).

A defect of a Lorenz curve is that it displays the relevant information in only a small space of the plot, and it is difficult to make comparisons across curves. To facilitate further exploration of sources of variation relevant for understanding the measurement of the upper tail of the wealth distribution in the SCF, the remaining analysis focuses on the differences between a Lorenz curve for a given wealth index and sample group and the Lorenz curve for net worth for the survey participants ("Lorenz difference curve"). Figures 5 and 6 provide such estimates corresponding to the curves for survey participants, shown in Figs 3 and 4, respectively.[52] By construction, each line in the figure is zero at the 99th percentile and the 100th percentile. Positive (negative) values indicate that the relevant wealth index is more (less) heavily concentrated at higher percentile values than is the distribution of net worth, as estimated for the survey participants. An up-

[49] Of course, some of this variation may reflect reporting error in the survey or instances where the household chose to consume or give away assets since the reference period of the survey measure of income.

[50] The one-percent group is defined separately for each curve.

[51] Appendix Table A4 provides similar information for the combined area-probability and list samples. Note that the population defining the top one percent is different in this case. The figures for the list sample in Table 6 are based on the population of tax filers modified as described above, whereas the population in Table A4 is the full set of households. The standard errors with respect to sampling shown in the table reflect only variability of the shares taking the share of the top one percent as given.

[52] The value shown on the Y-axis is the value of the Lorenz curve for net worth at each percentile point minus the corresponding value of the Lorenz curve for each of the wealth indexes.

ward (downward) sloping curve to the right indicates that the wealth index is becoming more (less) concentrated than net worth. For 2013 (2001), the figure shows that at the 99.5th percentile, about 2.5 (5) percent more of the total of WINDEX0 was held by wealthier households than was the case for net worth. Although the population examined here for each estimate differs somewhat, within each year they all derive from the same set of survey participants and they are very largely the same cases, as is clear from the results above on rank ordering. Thus, sampling error is unlikely to be an important explanation of differences across wealth measures within a given period.

There are three primary potential explanations for the persistent differences in distributions of wealth measured in the survey and the corresponding WINDEX distributions within each of the surveys.

(1) It is possible that conceptual differences or ambiguities in wealth measures are an important source of difference. For example, all the WINDEX measures use a measure of income reported for tax purposes to estimate an underlying wealth component. In practice, as discussed earlier not every type of asset generates income of a sort that would be reported for tax purposes, and information on debts is generally limited at best.[53] In addition, some elements of wealth do not have a well-defined and objective value (such as personal businesses or even houses), short of testing the marketplace with an actual transaction. However, for such factors to explain the differences in concentration would require that conceptual differences have a differentially larger effect on the very highest parts of the distributions of net worth and the wealth indexes. As noted earlier, there also may well be a substantial variation in the returns associated with different assets even within the same asset class. Loosely speaking, for such differences to explain the persistent differences in concentration would require greater understatement of rates of return relative to the true value at the top of the distribution.

(2) At least some of the differences are surely due to deviations in survey responses from precisely correct answers. Some such deviations are ones that may have a random component, as would be the case with ignorable item nonresponse.[54] More difficult to address are the effects of reporting error, in the underlying tax data or in the survey. For present purposes, it seems unlikely that clear errors in the tax data are a dominant factor, particularly given the importance of the tax returns of very high-income taxpayers within the tax collection process. Nonetheless, there may still be interpretations of reportable income derived from specific understandings of the tax code that lead to idiosyncratic reporting behavior that functions in the context of this analysis like reporting error.

Errors in the survey reports almost surely at least contribute to the differences between directly measured wealth and the WINDEX estimates. Because it is not feasible to conduct an exact evaluation study for the SCF, the primary evidence in this area derives from the survey editing. As noted earlier, examination of that process shows that editing leads to very substantial differences in the estimated wealth distribution. While the SCF editing process is based on principles, the execution is often not exact and, consequently, the process adds noise to measurement that is very difficult to quantify. Moreover, it may also introduce errors not initially present, either because of misinterpretation of the situations or because of the application of rules with embedded biases, such as might be the case for rules for selecting values to edit based on the saliency of data items. It should be noted that even if this process introduces no bias, it remains an unaccounted source of variability in estimates.

(3) As discussed earlier in this paper, the date of the survey wealth measure is generally as of a point two years or more after the latest reference period for the relevant income measures. Despite the approach to smoothing values across multiple years of tax-based data, large transient changes in income may still lead to important distortions in WINDEX measures rel-

[53] In addition, very wealthy individuals may have a greater ability than others to determine the extent to which their assets yield taxable income or the timing of income realizations.

[54] In principle, imputation error may be estimated. In practice, it is sufficiently difficult to estimate such error for present purposes that it is beyond the scope of this paper and left for future development. Past experience suggests that error attributable to that source is unlikely to be important relative to the differences observed here. In this paper, the emphasis is on directions of change and their consistency over time and with respect to the sampling or unit response processes.

ative to an ideal measure relative to wealth at the time of the survey. In periods where capital gains or losses occur relatively rapidly and differentially across asset classes, there may also be corresponding distortions in the WINDEX measures. Moreover, in the case of WINDEX1, the model estimates implicitly embed period-specific interest rates that cannot be adjusted in any straightforward way, in contrast to the case of WINDEX0. Nonetheless, the consistency of the general patterns of difference across periods with very different patterns of change in incomes over the preceding period suggests that while such factors may add to variation in the difference over time, they are likely to be less important than other factors in explaining relatively greater concentration of the WINDEX measures compared with the survey measure of net worth.

Figure 7 shows wide variation in the Lorenz difference curve for WINDEXM, the wealth index used directly in the sample selection, over the five surveys between 2001 and 2013. Unlike the case of the plots within Figs 5 and 6, here sampling error may explain some of the differences across years; each of the samples is selected independently.[55] But as indicated by the figures in Table 6 for ownerships shares by one-tenth percentiles of the top one percent, the estimated standard error due to sampling for share estimates for net worth and the wealth indexes averages between about 5 and 8 percent of the estimate (not percentage points of the estimate).[56] Thus, while the estimated sampling error may understate the true sampling error, the results suggest that straightforward sampling explains a relatively small part of the observed variation.

Other possibilities to explain the variation in differences include time-dependent conceptual differences in measured net worth and WINDEXM, nonstationary errors in the survey measurement, or true changes in economic conditions from the time frame relevant for the construction of WINDEXM to the time of the survey measurement. Addressing the first two possibilities would be quite difficult, but at least for the third possibility we may gain some insights from changes in income and asset prices over the relevant periods. The curves show the smallest differences in the Lorenz curves for net worth and WINDEXM in 2001, 2007 and 2013. One might argue that there were fewer disruptions in the relationship between wealth and the income processes that underlie WINDEXM from the reference time of values reported in the frame data to the time of the interview than was the case for the other survey years considered here. The fact that the major stock market declines in 2007 began only after the great bulk of the 2007 SCF interviews were already completed could explain the relatively low difference for that survey. Similarly, the much larger differences in 2010 may reflect disturbance from the Great Recession and the large differences for 2004 could reflect unusual reporting of capital income as a result of changes in the tax code generally seen as favorable to dividends and capital gains.

Ultimately, we care about the concentration of net worth for the entire relevant population, not just the set of participants. Distortions may be introduced by nonresponse and issues related to survey administration or even by artifacts in the sample design that affect the coverage properties of the sample. We have no direct information on net worth for either the entire list sample frame or full set of selected list sample respondents; but we are able to observe WINDEXM for the participants, the members of the full sample, and the members of the entire frame. Comparison of Lorenz difference curves for net worth participants and the wealth index values for the full sample can reveal differences due to nonresponse and comparison with estimates for the entire list frame can show differences relative to the closest approximation to the true target population.[57]

A potentially important difference in comparing the concentration of wealth measure for the survey participants with the WINDEX measures for the full sample and for the entire frame is that the latter two populations potentially contain members of the *Forbes* list of the 400 wealthiest people in the U.S., while the set of participants by construction does not. To address this concern, Table 6 for the survey participants, Table 7 for the full sample, and Table 8 for the full frame include concentration estimates for the top ten one-tenth percentiles groups, excluding the 400 top ele-

[55] There is some small dependence of samples between surveys as a result the exclusion of cases that had contact with interviewers in a preceding survey, as discussed earlier in this paper.

[56] Note that the sampling error estimate here is defined only on the variability within the top one percent. Of course in the context of the broader sample, the share of that group overall has variability.

[57] For this purpose, the entire frame is taken to be the full SOI sample adjusted to provide the basis for sampling, as described in the first section of this paper. The adjustments exclude the restriction applied in the sample selection to allow only cases in pre-selected areas of the U.S.

ments, as well as the corresponding unrestricted estimates.[58] Although the population adjustment has some effect on reducing differences in the concentration estimates based on the participants and those based on the other two groups, the effect is generally relatively small. This result highlights additionally that the overall results are not determined by a small number of extreme outliers. For simplicity, the remaining analysis here focuses on the unrestricted set of cases in each group.

Under both of the broader group definitions and across the all the survey years considered here, WINDEXM remains more concentrated than net worth. As shown in Figs 8 and 9 for 2001 and 2013 respectively, WINDEXM is even more highly concentrated than net worth for the full sample and for the entire frame than is the case for just the set of participants. However, this result does not hold for all other years (not shown): for 2004 and 2010, the difference in the concentration of WINDEXM and net worth is greatest for the set of participants. For all the years of the surveys considered except 2010, the degree of difference of concentration is greater for the entire frame than for the full sample; for 2010, they are about the same, as one might expect to be the case on average overall, absent distortions in the sampling process. This result suggests that it may be worthwhile to reconsider some of the geographic and other restrictions imposed in the sample selection or the approach to weighting.

The time series of differences for the full sample shown in Fig. 10 gives a similar ordering of differences over years as that for the set of participants, as shown in Fig. 7. It also shows a more compressed range of differences across the years. The most notable change from the ordering in the earlier figure is the greatly reduced difference for the 2004 data. In contrast, the comparable estimates for the entire frame, shown in Fig. 11, show larger differences, more on the order of those in Fig. 7; it also yields yet another order of differences, with the greatest difference being between the concentration of 2001 net worth for participants and WINDEXM for the entire frame. Thus, the possible macroeconomic explanation of some differences offered in the case of the participant group does not appear to be sustained as well here.

Although the evidence for consistent differences attributable to unit nonresponse or the sample design is not definitive, the results strongly suggest that there is at least another dimension of variability that may be substantial and that is not fully accounted for in the existing methodology of variance estimation. More work is needed to understand the differences.

Although some of the differences observed in the survey and wealth index distributions examined here are substantial, it should be kept in mind that those differences apply only to the distribution of the wealth of the top 1 percent. Because this group holds roughly one third of the total wealth, according to the SCF full-sample estimates, the differences should be considered in that more reduced context. The list sample appears to make a very substantial contribution toward anchoring the upper tail of the wealth distribution in a way that avoids unbelievably extreme jumps across the surveys. Nonetheless, pending future research to formalize additional sources of variability, the results still seem to argue for modesty in making SCF-based claims about the wealthiest one percent or, *a fortiori*, narrow subgroups of that one-percent group over time.

6. Conclusions and future research

This paper presents a comprehensive review of the SCF list sample, which traditionally has been argued to distinguish the SCF most essentially from other surveys that collect wealth data based on sample designs without a comparable tie to income or wealth. The bulk of the list sample is aimed at the top one percent of the wealth distribution, a group very difficult to include in surveys, but also critically important for the accuracy and precision of estimates of right-tail-dependent statistics related to wealth, such as means or concentration estimates.

The list sample design is based on proxies ("wealth indices") for net worth, generated using statistical records derived from individual income tax returns. One of those proxies is very similar to one used by [1] to make a time series of estimates of the wealth distribution. If a correspondence of such estimates with SCF estimates – or with a closer representation of actual wealth – could be clearly established, it might be possible to use that information to provide better and more regular estimates of wealth at a high level and to increase the accuracy and precision of SCF estimates,

[58] The 400 elements are defined here using the relevant weights to exclude a conceptually more comparable group. As noted earlier, the *Forbes* 400 group is not, in fact, necessarily among the top 400 in terms of net worth of the wealth indices. For the full list sample, the selection weights are used for the estimates reported, and for the full frame, the weights developed for the adjusted SOI data are used.

which are important for understanding the more detailed compositional effects for wealth. Unfortunately, we do not yet have sufficient grasp of the relevant "true values" to be able to say with assurance that a particular path is the right one. More research is clearly needed and would be a valuable contribution.

Based on comparisons of SCF estimates of aggregate wealth with external estimates, it appears that the survey does a good job of reproducing the important outlines of changes in wealth in the economy. However, there are differences in those comparisons and some of them may turn on artifacts of measurement in the very rarefied extreme upper tail of the wealth distribution. In the observed range, the wealth distribution has a fractal-like nature, with a small "wealthiest" fraction holding a very disproportionate share of the total even within progressively smaller slices of the top of the wealth distribution – seemingly up to the point that all that is left is a small group of extraordinarily wealthy people. Because it is clear from patterns of nonresponse in the survey that there is progressively greater nonresponse with greater values of the wealth index used for sampling, it seems very likely that some important share of wealth will surely be not be meaningfully represented – weighting at the very top cannot compensate, because there would be no similar cases to up-weight.

This paper looks in detail at the performance of the survey among the top one percent of the population by several wealth measures. The evidence presented indicates that the survey-based estimates of wealth concentration among the top one percent is generally less concentrated than would be expected from the wealth proxies based in part on capitalized income flows, which serve as the stratification basis for the SCF list sample. Moreover, the results also suggest that unit nonresponse and some aspects of the restrictions imposed in the list sample design may affect the accuracy or precision of concentration estimates.

In this light, the following areas of research seem most important for understanding the underlying differences and for plotting a path toward better estimates. First, it may be helpful to undertake a systematic evaluation of survey cases whose position in the estimated wealth distribution differs substantially from what would be expected on the basis of the wealth index employed in the SCF sample design. This exercise might give additional insights into the nature of reporting errors in the survey as well as a sense of any important distortions in the construction of the wealth index. Improvements might be identified for the data collection or processing efforts or for the design of the wealth index. Second, an effort should be made to obtain the necessary data to compute a version of the wealth index that corresponds to the same year as the survey measurement and potentially the intervening year as well; the data used in the list sample design generally predate the data collection by at least two years. A temporally parallel estimate of the wealth index might help explain further differences in classification in terms of measured wealth and the wealth index used for sampling.

Third, some reconsideration of the survey weighting protocol should be undertaken. The weighting adjustment for the SCF might be thought to bring the estimated wealth distribution of key frame variables for the survey participants into line with the estimates from the entire frame, but it is a virtual impossibility to create a perfect alignment in every dimension. The adjustments made surely do align a number of important characteristics. The extent to which the distribution of net worth and the wealth index are aligned as a result depends critically on the broad conceptual correspondence of measures, the lack of systematic bias in the survey data and the adequacy of the weighting adjustments. The temporally parallel estimate of the wealth index, as described above, might also provide the basis for a further weighting adjustment within the existing framework, with the potential to reduce the variance of estimates and possibly to identify or address some aspects of bias. It is an open question whether it would be advisable to align other aspects of the survey observations more tightly with the frame data – for example, a the relative concentration of the wealth index might be imposed on the survey to yield common concentration estimates. Even if there are such important conceptual differences that complete alignment would not be advisable, there might be some advantage in a least a partial alignment, if only as an experimental alternative to be studied over time.

Fourth, further attention should be given to the effects of some restrictions imposed in the design of the list sample. The exclusion of most geographic areas not included in the SCF area-probability sample and the exclusion of past sample members may be the most pressing points to consider, but it may also be advisable to review the adjustments imposed to transform a file defined in terms of tax units to a frame purporting to represent households containing at least one of filer of an individual tax return.

Fifth, in light of the time series of information available on nonresponse for the list sample, it would be

worthwhile to undertake additional modeling with the frame data together with other information to improve or update understanding of unit nonresponse. Consideration of changes across the multiple years of observations combined in each year of the frame data since the 2001 survey may be especially informative. Some of the resulting information might affect the strategy for adjusting the analysis weights to compensate for nonresponse, but there may also be useful lessons for the survey administration.

Finally, even if no changes are decided for the survey design and execution, the available evidence gives weight to the thought that there may be sources of variability particularly important for the extreme upper tail of the wealth distribution that are not currently captured in the variance estimation methods developed for the SCF. A methodical review should be undertaken to ensure that variability is not seriously misstated for right-tail-dependent SCF estimates or other important estimates.

For any change that would touch as deeply into the heart of the SCF measurements as one in the sample design or related technical matters, it would be important to make a very strenuous effort to bridge any period of change. Ideally, any such bridging would be made available for as many waves of the survey as possible, but at a minimum there should be an overlap of methods for at least one survey. Obviously, any serious changes should be subjected to a rigorous process of evaluation both before and after implementation. But where a change is one that cannot be bridged, even greater care should be exercised.

The results of the paper can be taken to argue that, at least pending additional research, there should be an extra degree of modesty in interpreting SCF concentration estimates and similarly right-tail-dependent estimates, and their changes over time. Because the "truth" is not known, the same point might be argued for wealth estimates simulated directly from tax-based data. Better versions of particular estimates often may be had by pooling or otherwise leveraging alternative approaches or data, if we can understand the respective strengths of the sources so as to be able to weight their contributions appropriately. The SCF has already gone further than most surveys in exploiting alternative data and undoubtedly it can go further.

Acknowledgments

The views expressed in this paper are those of the author alone and they do not necessarily reflect the views of the Board of Governors of the Federal Reserve System or its staff. The author is grateful to all those whose work has made this paper possible, but particularly Fritz Scheuren, who as director of Statistics of Income at IRS first made it possible to use tax-return-derived data for the SCF sample design; Barry Johnson, the current director or SOI who has supported the survey in critical ways over many years; Louise Woodburn, who was a partner in many of the statistical developments underlying the SCF; Steven Heeringa, who made fundamental contributions in his work on the earliest SCFs conducted with the Survey Research Center at the University of Michigan; many staff at the Federal Reserve Board and SOI who did the necessary work to give birth to and sustain the survey; the central office and field staff at NORC who did the very hard work of collecting the data; and the survey respondents, without whose contributions the entire exercise would be pointless. The author is grateful to Emmanuel Saez for comments.

Supplementary data

The supplementary files are available to download from http://dx.doi.org/10.3233/SJI-170349.

References

[1] E. Saez and G. Zucman, Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Data, working paper, University of California Berkeley, Subsequently published in *The Quarterly Journal of Economics* **131**(2) (May 2016), 519–578.

[2] J. Bricker, A.M. Henriques, J.A. Krimmel and J.E. Sabelhaus, Measuring Income and Wealth at the Top Using Administrative and Survey Data, Finance and Economics Discussion Series 2015-30. Washington: Board of Governors of the Federal Reserve System, 2015, http://dx.doi.org/10.17016/FEDS.2015.030.

[3] J. Bricker, L.J. Dettling, A.M. Henriques, J.W. Hsu, K.B. Moore, J.E. Sabelhaus, J. Thompson and R.A. Windle, Changes in U.S. Family Finances from 2010 to 2013: Evidence from the Survey of Consumer Finances, *Federal Reserve Bulletin* **100**(4) (September 2014), 1–40.

[4] A.B. Kennickell, Multiple Imputation in the Survey of Consumer Finances, *Statistical Journal of the IAOS* **33**(1) (2017), 143–151

[5] A.B. Kennickell, Getting to the Top: Reaching Wealthy Respondents in the SCF, *Statistical Journal of the IAOS* **33**(1) (2017), 113–123

[6] J. Bricker and A.B. Kennickell, Shared Understanding and Data Quality in the SCF, *Proceedings of the Survey Research Methods Section*, American Statistical Association, 2013.

[7] A.B. Kennickell, Dirty and Unknown: Statistical Editing and Imputation in the SCF, forthcoming (September) *Statistical Journal of the International Association for Official Statistics* **31**(3) 2015, 435–445. .

[8] R. Tourangeau, R.A. Johnson, J. Qian, H.-C. Shin and M.R. Frankel, Selection of NORC's 1990 National Sample,working paper, National Opinion Research Center at the University of Chicago, Chicago, IL, 1993.

[9] C. O'Muircheartaigh, S. Eckman and C. Weiss, Traditional and Enhanced Field Listing for Probability Sampling, *Proceedings of the Social Statistics Section of the American Association for Public Opinion Research*, 2002, 2563–2567.

[10] C. O'Muircheartaigh, S. Eckman and C. Weiss, Using Income Data to Predict Wealth, *Proceedings of the Survey Research Methods Section*, American Statistical Association, 1999.

[11] A.B. Kennickell and D.A. McManus, Sampling for Household Financial Characteristics Using Frame Information on Past Income, *Proceedings of the Survey Research Methods Section*, American Statistical Association, 1993.

[12] Statistics of Income Division of the Internal Revenue Service, Individual Complete Report, 2014, http://www.irs.gov/pub/irs-soi/08insec2.pdf.

[13] M. Frankel and A.B. Kennickell, Toward the Development of an Optimal Stratification Paradigm for the Survey of Consumer Finances, *Proceedings of the Survey Research Methods Section*, American Statistical Association, 1995.

[14] E.R. Canterbury and E. Joe Nosari, The Forbes Four Hundred: The Determinants of Super-Wealth, *Southern Economic Journal* **51** (April 1985), 1073–1982.

[15] D. Greenwood, An Estimation of U.S. Family Wealth and its Distribution from Microdata, *Review of Income and Wealth* **29** (1983), 23–43.

[16] A.B. Kennickell, Modeling Wealth with Multiple Observations of Income: Redesign of the Sample for the 2001 Survey of Consumer Finances, *Statistical Journal of the IAOS* **33**(1) (2017), 51–58

[17] Office of Management and Budget, Executive office of the President, Standards and Guidelines for Statistical Surveys, 2006, https://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf.

[18] A.B. Kennickell and R. Louise Woodburn, [1997] Consistent Weight Design for the 1989, 1992, and 1995 SCFs, and the Distribution of Wealth, *Review of Income and Wealth* **45**(2) (June 1999), 193–215.

[19] A.B. Kennickell, Revisions to the Variance Estimation Procedure for the SCF, 2000, http://www.federalreserve.gov/econresdata/scf/files/variance.pdf.

[20] OECD, *OECD Guidelines for Micro Statistics on Household Wealth*, OECD Publishing, 2013, http://dx.doi.org/10.1787/9789264194878-en.

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

[22] M. Cagetti, E.B. Holmquist, L. Lynn, S.H. McIntosh and D. Wasshausen, The Integrated Macroeconomic Accounts of the United States, in: *Measuring Economic Sustainability and Progress*, D.W. Jorgenson, J.S. Landefeld and P. Schreyer, Studies in Income and Wealth Volume 72, University of Chicago Press, 2014, pp. 277–321,

[23] B.W. Johnson and K.B. Moore, Differences in Income Estimates Derived from Survey and Tax Data, in *Proceedings of the Survey Research Methods Section*, American Statistical Association, 2009, pp. 1495–1503.

[24] A.B. Kennickell, The Role of Over-sampling of the Wealthy in the Survey of Consumer Finances, in *The IFC's contribution to the 56th ISI Session, Lisbon, August 2007*, **28**, 2008, pp. 403–408.

[25] F.T. Pfeffer, R.F. Schoeni, A. Kennickell and P. Andreski, Measuring Wealth and Wealth Inequality: Comparing Two U.S. Surveys, working paper, Institute for Social Research, University of Michigan, 2014.

[26] W. Kopczuk, What Do We Know about Evolution of Top Wealth Shares in the United States, NBER working paper 20734, 2014, http://www.nber.org/papers/w20734.

[27] T. Picketty, *Capital in the Twenty-First Century*, Belknap Press, 2014.

[28] V. Quadrini, J. Díaz-Giménez and J.-V. Ríos-Rull, Dimensions of inequality: Facts on the US distributions of earnings, income, and wealth, *Federal Reserve Bank of Minneapolis Quarterly Review* **21**(2) (1997), 3–21.

[29] E.N. Wolff, Recent Trends in Household Wealth in the United States: Rising Debt and the Middle-Class Squeeze – an Update to 2007, Working Paper No. 589, Levy Economics Institute of Bard College, 2010.

[30] A.B. Kennickell, Ponds and Streams: Wealth and Income in the U.S., 1989 to 2007, Finance and Economics Discussion Series 2009. Washington: Board of Governors of the Federal Reserve System, 2009, http://www.federalreserve.gov/pubs/feds/2009/200913/200913pap.pdf.

[31] A.B. Kennickell, The Other, Other Half: Changes in the Finances of the Least Wealthy 50 Percent, 2007–2009, *Statistical Journal of the IAOS* **33**(1) (2017), 185–193.