Constant focus: Engaging to measure wealth

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

Abstract. This essay is intended to place the selection of my papers in this issue of the *SJIAOS* in the context of other research on wealth measurement and to provide an overview of their part in the history of the constant focus in the Survey of Consumer Finances to engage in measuring wealth more effectively. It also makes some recommendations for future work on that survey and possibly other systems of measurement.

Keywords: Wealth measurement, wealth distribution, survey methodology

The papers collected in this volume are a selection from work I did over the past few decades aimed at trying to understand household wealth and its measurement, largely using the Survey of Consumer Finances (SCF), an interviewer-mediated household survey conducted by the Federal Reserve Board (FRB) with NORC at the University of Chicago.[1] In making this selection, my hope is to give a reflection of my view of surveys as complex and evolving systems, where the continuous improvement framework offers the most meaningful hope of survival and progress. None of the individual papers is the last word on anything, but rather they represent to me points of departure for further improvements. All of them try to point toward a future beyond themselves.

Among my favorite of Deming's famous "14 Points" is the one about the importance of constancy of purpose.[2] Of course, we also must recognize that the practical understanding of purpose may evolve as a result of such continuing focus, and sometimes what appeared earlier to be sense was, in fact, another level of confusion or error. As I noted in my interview in this issue with Katherine Condon, a favorite quote of mine is one from the work of William Blake: "If others had not been foolish, we should be so." In that vein, I humbly confess that I have many times been my own "other". In my work, I have made a conscious attempt to make the evolution of my ignorance as transparent as I could, in hopes that others may benefit from my mistakes or be clever enough to find solutions to problems I could not solve. I hope that intent will be clear in these papers.

Household wealth matters for a wide range of reasons. For an individual household, wealth may provide a hedge against future shocks or uncertainty, support in retirement, a down payment on a home, the basis for other future spending such as college fees, a legacy to others, support for a private business activity, or other functions. Viewed from the perspective of the macro economy, accumulated wealth supports investments in productive capital, influences the level of spending that drives production, and figures in a variety of other processes. Wealth and its composition matter for financial markets and financial stability. One could also look at wealth from a variety of other perspectives. The composition of wealth varies with the goals, expectations, financial literacy and opportunities of households. Of particular importance, the *distribution* of wealth profoundly affects the ways that macroeconomic shocks propagate through the economy or the population – in a relatively direct way through individual portfolios or by influencing expectations or perceptions of risk at many levels that affect economic activity. Distribution also has implications for social equity, which in

---

[1] Prior to 1992, data for the survey were collected by the Survey Research Center at the Institute for Social Research at the University of Michigan.
[2] See https://deming.org/management-system/fourteenpoints for the entire set.

turn has implications for social stability. Understanding wealth, its composition and its distribution is important for informed policy making and other planning at the societal level.

At least since Adam Smith, scholars have been interested in understanding the wealth of nations. Measurement of wealth is very difficult, because many different concepts or measures of wealth and individual wealth items may be potentially appropriate and because values may be difficult to determine when specific items of wealth are not routinely traded in markets. In addition, wealth is not always even "visible" through direct observation or through traces in observable data. Pioneering work by Raymond Goldsmith and others [1–3] led to the development of systematic national-level estimates of balance sheets for the economy, using available information from a variety of sources; the descendant of this work is the Financial Accounts of the United States maintained at the Federal Reserve and similar sets of accounts elsewhere.[3]

Most of the early sources of distributional data on wealth rest on the use of the "estate multiplier method" proposed by Bernard Mallet [4] for estimating the distribution of the wealth of the living population from the wealth of decedents, as captured on reports made for purposes of determining the taxes to be paid by their estates. The most comprehensive implementation of this method in the U.S. was a large study by Fritz Scheuren [5] in 1962 using data from the U.S. Internal Revenue Service on the estates of decedents, building on the earlier work of Robert Lampman [6]. Inferring the wealth of the living from that of the dead requires that there not be unmanageable "mismeasurement" or "selection bias" in the wealth of decedents. The reported wealth may differ significantly from the value before a final illness, some assets such as trusts are not captured at all, and some others may be stated at a value different from a market value. The population obviously also differs from the overall population in that all the decedents have wealth of at least as much as the reporting threshold for an estate tax return; in addition, a variety of data indicate that relatively wealthy people have a lower age-specific probability of death (equivalent to a selection probability, in this case) than less wealthy people. There are ways of coping with such problems to provide estimates that have at least a reasonably common basis over time. A large attraction of this approach is that the costs of "data collection"

are covered by an administrative function for tax purposes. The most serious limitations are the general lack of contextual information about the wealth holders and the complete lack of detail for less wealthy households.

As discussed in more detail below, direct measurement of household wealth raises many other issues, not least the cost and complexity of collecting data from many individuals. The first attempt I am aware of to use such data to describe the U.S. wealth distribution is work by G.K. Holmes [7] using data on tangible wealth collected in the 1890 Census of Population. Among other findings, he reported that an estimated 91 percent of the population owned no more than 29 percent of wealth, a figure in the range of current estimates based on broader measures of wealth from the SCF.

Although there were subsequent survey efforts to understand the distribution of household wealth, the next major step forward was in 1962, when the FRB conducted the Survey of Financial Characteristics of Consumers (and the 1963 follow-up survey with respondents to the 1962 survey), under the leadership of Dorothy Projector. An article about the survey, "Survey of Financial Characteristics of Consumers", from the 1964 *Federal Reserve Bulletin* is reproduced in the on-line supplement to this paper.[4] This survey made a comprehensive attempt to address all household assets and liabilities, with the principal exception of assets that consist of a contingent right to an income stream, such as defined-benefit pensions. A very important distinction between this survey and the sample survey work that came earlier was the use of data from the U.S. Internal Revenue Service to design the sample, to ensure adequate representation of the upper tail of the income distribution and adjust for nonresponse. Enormous care was taken in the design and processing of the survey – even to the extent that individual holdings of publicly traded stocks were listed by respondents and the values were looked up in the *Wall Street Journal*. However, an unfortunate result of the necessary level of effort and the slow speed of the computers of the period was that production of results required substantial time and a single-minded focus of the survey team. Certainly, those factors reduced the flexibility to use the data for policy work and research and reportedly figured in the decision to cease such work for

---

[3]See https://www.federalreserve.gov/releases/z1/current/default. htm for a description of the Financial Accounts of the Unites States.

[4]The supplementary files are available to download from http://dx.doi.org/10.3233/SJI-170347.

20 years, until the 1983 SCF revived and reformulated the earlier approach.[5]

Contrary to some beliefs, I did not start the SCF. It was begun by my former colleague Robert Avery and others. I came to the FRB after the field work had been completed for that survey. I dived into the project from the point where nonresponse adjustment was the main problem and I got involved with the analysis. I stayed wrapped up in this work for about another 30 years. The SCF has had nearly every interesting problem a survey can have, and I see no end to that – this is the way of the practical world of measurement. Many times, the survey could easily have been drowned, and that provided a very compelling incentive for research to address those problems. The first paper in this collection, "Wealth Measurement in the SCF", provides a more detailed history of the survey and it puts the evolution of problems and solutions in perspective as of the year 2000.[6] I hope the spirit of the early critical evaluation and improvement is clear in this paper and the later ones that built upon the earlier work.

In principle, if a survey sample is randomized over the full target population, executing the survey is equivalent to drawing one realization of a Monte Carlo process. If its questionnaire is also unambiguously designed to support its goals, the resulting estimates are unbiased and the degree of variability of the estimates is a solely a function of the sample design. Obviously, there are many other potential sources of nonsampling error that may greatly complicate analysis beyond that. However, as indicated by the large number of papers related to sample design included here, that subject matters a great deal for the SCF.

Wealth is highly concentrated in the U.S. – it appears that about one-third of total household wealth is held by the wealthiest one percent of households and holdings appear to be highly skewed even within that rarefied group. The implications for statistical efficiency of estimates influenced by the upper tail, such as means or concentration estimates, alone argue for serious attention to sample design.

Following the earlier practice of Dorothy Projector, from the start the SCF sample made use of ad-

ministrative data – statistical records derived from individual tax returns by the Statistics of Income Division (SOI) of the U.S. Internal Revenue Service. Both the work of Projector and the 1983 SCF combined a geographically-based sample with a list sample selected from the administrative data. Each survey used a relatively simple income-based approach to stratification. Using tax-based information for any purpose beyond narrow tax administration or forecasting was still highly controversial in 1983. It was thanks largely to the efforts of Fritz Scheuren, then director of SOI, that data were available for sampling and that the obstacles to providing any information for nonresponse adjustment could be sufficiently overcome.[7]

As the channel to the SOI data for the SCF became clearer and less controversial and as support for the use of the SCF in tax policy analysis grew, the possibilities for improvement expanded. Many members of the SOI staff were very helpful, engaged and supportive, as were staff in the Office of Tax Analysis and the Congressional Joint Committee on Taxation. For the 1989 survey, the relationship of trust and security was well enough established that it was possible to use SOI data more directly to capitalize various categories of capital income to create a proxy for wealth, which was then used for stratification. The capitalization involved applying the inverse of an average rate of return for a given type of asset to the corresponding income category. The income was for returns filed in 1988 (with few exceptions, for income received in 1987), the most recently available information at the time. Wealth and overall income are often not tightly related; the separate capitalization, which amounts to differential weighting of income types, was shown to be an improvement when we evaluated the performance of the list sample.

However, this more detailed approach to developing the stratification variable still yielded noticeable misclassification. Some of the differences may have been attributable to the lag between the reference period for income and the reference period for wealth. But some appeared likely to be a result of the approach to income capitalization, which allowed for no heterogeneity across sample units other than variations in income. Rates of return and even rates of income realization from assets may vary greatly across even the up-

---

[5]A notable effort at the Banca d'Italia, Bilanci delle Famiglie Italiane (Survey of Household Income and Wealth, as it is known in English) began soon afterwards and it continues to the present. See www.bancaditalia.it/statisitice/tematiche/indagini-famiglie-impresse/bilanci-famiglie/index.html?com.dotmarketing.htmlpage. language=1.

[6]The figures and tables for this article and all the others in this issue may be found online at the location indicated in each article.

[7]Only many years later, was the detailed information on the sample design given to Federal Reserve staff. There remains an as yet unexercised potential to exploit that information for reweighting the 1983 data to be more comparable to the later surveys.

per tail of the wealth distribution. For the 1998 SCF, SOI gave permission to project staff at the FRB to link anonymized wealth data from the cases in the tax-derived part of the 1995 SCF sample with the original SOI frame data for that survey.

As described in the paper "Using Income to Predict Wealth", the matched data were used to estimate a more flexible model of observed wealth using a broad set of available characteristics and appropriate transformations. Conditioning the model in this way made it possible to capture a greater degree of heterogeneity in the underlying choices and situations that shape the relationship between wealth and income. One obvious defect of this approach is that it implicitly fixed the rates of return as of a period five years before the 1998 survey, a problem not shared by the simple capitalization model. Thus, applying the estimated coefficients to the comparable data from three years later could distort the projection of wealth, if rates of return or other aspects of the income-generating process changed substantially. To capture both the potentially greater specificity of the estimated model and the stability of the capitalization model, the two approaches were combined. Although *ex post* evaluation indicated that including the estimated model improved the classification of households, it also revealed that the variability of the incomes used in the two models was a significant source of misclassification. The paper "Modeling Wealth with Multiple Years of Income" addresses the effect of pooling multiple years of income to smooth out the more transitory fluctuations in annual income reported for tax purposes, the approach then adopted for the SCF.

The most recent paper included in this issue, "Lining Up: Survey and Administrative Date Estimates of Wealth Concentration" (a redacted version), covers a number of sample-related issues as well as some related issues I will raise later in this essay. The results in the paper support the great power of the list sample in providing a credible basis for a broad view of the wealth distribution. As shown in the presentation slides of a version of the paper I presented at the 2015 Joint Statistics Meetings, estimates of the share of the wealthiest one percent computed from the area-probability sample alone show very large and highly implausible period-to-period variation.[8] In contrast, es-

timates incorporating the list sample show a fairly stable pattern; those estimates also show reasonably stable and narrow confidence intervals.

Nonetheless, the SCF has never claimed to represent the extreme upper tail of the wealth distribution. Indeed, the list sample explicitly excludes members of the *Forbes* list of the 400 wealthiest people in the U.S, though it does usually select a small number of households missed in the *Forbes* list, which are deleted from the public version of the SCF data. The supporting argument for the exclusion has been that the extremely wealthy households in the Forbes list would be highly unlikely to participate, reaching them would be very costly, and their data could not be sufficiently masked in the public version of the data to avoid meaningful risk of compromising their confidentiality. The paper also delves more deeply into issues of representation within the wealthiest one percent of households, the range in which the list sample is most dense. Ideally, the composite wealth proxy used for stratification would rank people very closely to their ordering in terms of actual wealth, and the sampling and the nonresponse adjustments stages would neutrally tend to align the observed data with that ranking. But as the paper shows, there appears to be meaningful variation in the extent of alignment that was not previously clear, and the approach to both sampling and nonresponse adjustment appear to have distorting affects. Arguably, such variation should somehow be incorporated into confidence intervals for the estimates, but it seems even more important to understand more deeply the sources of differences and to address them more directly, as I propose later in this essay in a discussion of possible future directions.

Nonsampling error is painful. As the late Arnold Zellner once remarked in a panel I organized at the Joint Statistical Meetings on item nonresponse, focusing on adjustments is an acknowledgement of failure: rather we should focus on avoiding problems. I agree to my heart with this proposition. It is much the same idea as the distinction between the importance of maintaining a high-quality emergency room versus maintaining an effective public health system – it is better to be proactive to stay healthy than to get patched up. We fail when we are unable to see situations clearly enough to design effective measures, or worse, when we do not act on what we already know. We must strive continually to see better, create systems that can reveal more to us, and engage everyone involved. Making errors or weaknesses as transparent as possible creates the focus to evaluate their importance – and, for many people, it creates an incentive to eliminate them.

---

[8]See ww2.amstat.org/meetings/jsm/2015/onlineprogram/Abstract Details.cfm?abstractid=316175 and related information in "The Role of Over-sampling of the Wealthy in the Survey of Consumer Finances" [8].

Most aspects of nonsampling error in an interviewer-mediated survey can be attributed to at least one of the following: design or implementation errors by the survey designers, administration errors by the interviewer or response errors by the respondent. There should be a high ethical premium placed on eliminating all errors by the survey designers that can be detected in advance and on structuring work as much as possible for detecting others for future resolution. To reach the respondents, the designers must go through the interviewer and the surrounding administrative process. In the end, the interviewer is alone to perform the very difficult task of making the survey instrument and other protocols work with the variety of respondents they face. The designers should do all they can to prepare the interviewer for this role and to capture important traces of the interaction that would allow for greater understanding of weak points in that process and for making future improvements.

Unit nonresponse is a serious problem in the SCF and experience has shown clearly that response is differentially lower among wealthy households. Typically, only about 70 percent of the in-scope area-probability sample cases participate. But for the list sample, only between about 40 and 10 percent participate, with response rates declining with higher levels of the wealth proxy used for stratification. If post-survey adjustments for nonresponse were not able to account for the differentially lower response rate among wealthy households, serious nonresponse bias would result in estimates sensitive to the upper tail of the wealth distribution. The SCF list sample design itself makes possible important, though imperfect, adjustments for unit nonresponse, without which many SCF estimates would have no credibility.

Ultimately, unit nonresponse is a function of how well the entire survey team can act to gain the attention and cooperation of a respondent. The principal focus should be there, and again, even during the necessary ex post "emergency room" repairs, there should be a serious commitment to understanding the failure. Many of the other papers included here were aimed at understanding the process of nonresponse from a specific direction and thinking what might be done to reshape the field work or otherwise support interviewers in their very difficult work.

In the early stage of field work on most surveys, nonresponse does not yet seem like a problem. However, as the more willing respondents are "used up", success becomes more difficult. Physical barriers (e.g., locked doors or gates) or institutional barriers (e.g.,

doormen, assistants), and busy or uninterested respondents who have been contacted and not firmly refused become more prominent. As the locations of the remaining cases become more scattered, the cost of visiting them in person increases. Groves and Heeringa [9] have proposed a potentially very useful approach to field management, "responsive design", intended to focus effort where success is more likely and to reduce bias in the final estimates. Achieving the latter requires knowledge about the distribution of relevant characteristics across a population that are generally unobserved in a cross-sectional survey, though the approach may do as well as post-stratification on observable characteristics, while sparing some expensive effort. On the other hand, understanding the structure of the field work in the late stages may inform us better about how to address obstacles, and to cope with them where necessary. The papers "What Do the "Late" Cases Tell Us? Evidence from the 1998 Survey of Consumer Finances", "The Bitter End? The Close of the 2007 SCF Field Period", "Darkness Made Visible: Field Management and Nonresponse in the 2004 SCF" and "Getting to the Top: Reaching Wealthy Respondents in the SCF" are stages in an attempt to understand more deeply the situations of respondents who respond relatively late in their exposure to field effort and what might be done to structure work to be more effective in reaching and persuading them.

Another related issue addressed to some extent in these papers is the decision to apply effort at all. Often the study of paradata takes the application of effort as exogenous. But without a way to incorporate the decision to apply effort or when to apply it, it is not possible to identify respondent characteristics that may affect response propensities. If interviewers have case-specific knowledge about the likelihood of response, that may wind up shaping the attention those respondents get, thus potentially exaggerating the importance of the characteristics of the "likeliest" respondents in understanding, addressing or coping with nonresponse. Capturing the expectations of interviewers about response likelihood is one approach to addressing that endogeneity problem, as described in the paper "What's the Chance? Interviewers' Expectations of Response in the 2010 SCF".

Unfortunately, even when a respondent is willing to complete an interview that is no guarantee that the interview will be complete. Questions may be written in a way that makes them hard for respondents to answer, respondents may not have the information to answer some questions, questions may have answers in

some situations that do not lend themselves to a single response or a response within a given code frame, and respondents may be unwilling to answer some of the questions. Any of these problems may lead to item nonresponse. Such nonresponse may be reduced if interviewers are able to continue persuading or supporting respondents throughout the interview or allow some more flexible ways of answering questions.

The migration of the 1995 SCF to computer-assisted personal interviewing (CAPI) from a paper-and-pencil mode made it possible to present questions to the interviewer in a less burdensome way, freeing the interviewer to focus more clearly on maintaining rapport with the respondent. It also allowed a more flexible way of answering questions with answers denominated in dollars, as discussed in the paper "Using Range Techniques with CAPI in the 1995 Survey of Consumer Finances". This work built an earlier history at the University of Michigan offering the option of using "range cards" or "decision trees" as alternative means of answering dollar-denominated answers. The change for the SCF was successful, in that it reportedly made it easier for respondents to answer questions where the amounts involved were genuinely uncertain, and it appeared largely to have replaced "don't know" responses with a range. Its effects on refusals to answer questions were more muted. Despite this improvement, some questions without dollar responses remained unanswered. In addition, the range responses for dollar responses were not directly useful for most analytical purposes. Imputation is one way to address missing, or "partially missing" (range), data.

Beginning with the 1989 survey, the SCF has employed a multiple imputation (MI) technique, Federal Reserve Imputation Technique Zeta (FRITZ), following the general framework proposed by Donald Rubin, as described in the paper "Multiple Imputation in the Survey of Consumer Finances". It was among the first implementations of MI for a large, complex survey. FRITZ is an iterative model with chained variable-by-variable imputations within iterations. Possibly the most complicated aspects of imputation for the SCF are the specification of constraints on the imputations, based on range information or prior structural information, and the derivation of the "implications" of each imputation for other imputations later in the chain.

MI is helpful not just for "filling in the missing data" but also for providing a measure of the value of the missing information, in terms of both first and second moments. These metrics have guided attention to survey questions where missing amounts are a relatively large fraction of the total or where the imputation variance is relatively high. In addition, once there was a complete model for imputing all the survey variables, it could then be applied as a tool in disclosure reduction to simulate data in a way that provides greater protection of respondents' confidentiality, while also maintaining correlations across variables.

Even when respondents answer the survey questions, their answers may not always align with the intent of the survey designers in creating the questions or with the reality of the respondents' own situations given their understanding of the questions. Such reporting errors can have seriously distorting effects on analysis of the data, as shown for the estimated wealth distribution, in the paper "Dirty and Unknown: Statistical Editing and Imputation in the SCF" [10] published in an earlier issue of this journal. In addition, that paper attempted to develop a model to guide the application of effort in editing. The case-by-case approach to editing traditionally applied in the SCF is very grueling work. Because a completely edited SCF data set was available as a baseline, it was possible to create various mixtures of edited and unedited data to examine the effects of leaving some sets of cases unedited on the estimated wealth distribution. The results of the paper indicate that a more limited but targeted approach, based on factors observable before editing, may be almost as effective with much less effort.

Editing may help to repair the data, to the extent that is possible, but it may also replace correct data with incorrect data or with a missing value. A virtue of the process, which would remain even with the more targeted approach noted above, is that is brings the editors into close contact with the way questions are answered and to some extent the ways that different interviewers approach them. Systematizing such information during editing can be used as a basis for improvement. The paper "Interviewers and Data Quality: Evidence from the 2001 Survey of Consumer Finances" in this issue was an attempt to use the traces of editing to identify questions with systematic problems. The results of that paper and related work led to revisions in many survey questions or protocols and to changes in the support for interviewers.

Interviewer training and continuing education are critically important means to pass the intentions and understandings of the survey designers to the important people who do the work. A great deal of effort has also been focused on those activities for the SCF. Among the means tried is case-level feedback from the FRB editors to interviewers about issues identified in edit-

ing. For the continuous improvement process to function meaningfully in this dimension, knowledge about problems detected in editing needs to be passed to the point where it could influence action in the field; reactions and clarifications rooted in the exigencies of the interviewing situation need to go the other way.

The initial approach was for the editing staff to write messages directly to the interviewer for each edited case and to ask for responses from interviewers via their managers. Although some interviewers appreciated the personal and direct attention, others felt they were being unfairly criticized for doing the best they could in difficult circumstances. Of course, the intent was always to be supportive and to work together to resolve misunderstandings or other gaps in understanding on either side. But it became clear that there was a serious degree of disconnection of understanding through the administrative chain to the interviewers, so effective support was not possible, despite the good intentions. An apt analogy is a joke sequentially translated into multiple languages and finally back to a meaningless reflection of itself in its original language.

One of the costs of going from a paper to an electronic questionnaire is that the content of the interview becomes more of an abstraction or artifact of memory for the interviewer and much of the rest of the field staff, because it cannot be seen as a concrete whole. That view is largely restricted to the staff processing or analyzing the data. Ideally, everyone should have the same understanding of the survey goals and the constraints. Recreating a more concrete sense of the survey responses and their analytical implications was a goal of the paper "Shared Understanding and Data Quality in the SCF" with my colleague Jesse Bricker. The paper describes an attempt to engage part of the administrative chain in clarifying, diagnosing and resolving some interview problems that could be detected with reasonably high probability by algorithmic means. By creating a deeper understanding at that level, the hope was that we could begin to construct a clearer communication channel. Interviewers deserve the best possible support, and there is very much still to be learned from them – far beyond the knowledge typically collected as paradata. I hope that work will go far beyond where we left it for the SCF.

Much of what we can hope to achieve in a survey turns on what we can achieve in creating a healthy common culture among all those doing the work and in passing the "flame of respect" to respondents. Survey specialists speak often of the "culture of confidentiality", which is indeed a very important aspect of culture.

But culture that supports mutual respect and an aspiration for continuous improvement is far broader. An earlier issue of this journal was devoted to the subject of interview fabrication, arguably the very worst form of nonsampling error. I argue in the paper "Curbstoning and Culture" [11], which was included in a 2015 issue of this journal, that fabrication is the outcome of an unhealthy survey culture – one so unhealthy if fabrication is present to any material degree, that the entire survey effort should be called into question. The mutual respect in a healthy culture is a worthy outcome on its own, but it is critically important in enabling the analytical needs for data to be satisfied in a survey.

Aside from humanistic motives, the guiding concern of the designers of a survey is, of course, the production of data to serve an analytical need. To represent that interest in this collection of papers, I have chosen two that address the wealth distribution. I have written many papers on the U.S. wealth distribution, often with special attention to portfolio composition, the role of inheritances, racial disparities, or other issues. Here, I have chosen "Tossed and Turned: Wealth Dynamics of U.S. Households 2007–2009", which considers changes in wealth composition and concentration over the financial crisis during that time, and "The Other, Other Half: Changes in the Finances of the Least Wealthy 50 Percent, 2007–2009", which focuses more particularly on the least wealthy half of the population. The analysis uses panel data, the first such information available from the SCF since the 1983–1989 panel, which suffered from such serious problems of various sorts that it was not much used for longitudinal analysis. Indeed, the earlier experience was so unsatisfying and difficult that it created, at least in me, the sense that the level of respondent burden and the effects of nonsampling error made it infeasible to collect panel data in the SCF that would be fit for use. Citing the obvious analytical benefits, SCF users had repeatedly listed panel data as their highest priority. But I resisted. It was only the strength of the need for information on the state of households during the financial crisis that overcame that resistance. The success of that effort convinced me that I had been wrong to resist.

The 2009 re-interview was designed to follow, as closely as possible, the framework of the baseline 2007 questionnaire, but with much less granular detail. This approach was intended to maintain conceptual coherence while reducing respondent burden and making the data processing manageable – particularly the case-level examination during the editing, as described in the paper "Look Again: Editing and Imputation of SCF

Panel Data". Because both the baseline and panel measures were examined concurrently during the editing (and some additional edits were even incorporated in the baseline 2007 data), there is a reasonable level of confidence that the measured changes largely represent real changes. In addition, as discussed in the paper "Try, Try Again: Response and Nonresponse in the 2009 SCF Panel", it appears that the nearly 90 percent response rate was common across so many groups that nonresponse did not introduce important biases.

The panel survey was followed by the already scheduled regular 2010 SCF cross-sectional survey. As a result, the project team needed to begin intense preparations for a second survey while still facing the arduous tasks needed to complete the processing of the panel survey. That collision, I am afraid, seriously dampened the necessary enthusiasm to carry on with the collection of panel data, making allowances for all that was learned from the 2007–2009 panel. I continue to believe that a *long-running* SCF panel would be both unmanageable because of the cumulative difficulties in managing nonsampling error, and insufficiently meaningful because of changes in the composition of individual households that would quickly confound analysis of wealth changes at the household level. However, I hope that in the future a sequence of *short* panels can be instituted before the practical lessons of the 2007–2009 panel are lost. Some cost estimates support the view that it would be approximately cost neutral to change the frequency of the new SCF cross-sectional surveys to a quadrennial basis from a triennial basis, with a panel re-interview like that in 2009, two years after each of the cross-sectional surveys. If the effort required for processing cannot be sufficiently managed, the intervals between cross-sectional and panel surveys could be expanded by a year, almost certainly resulting in overall cost savings if the scale of the panel interview were sufficiently similar to that in 2009.

Finally, I would like to offer a vision of a new direction I think is important for the SCF. The work I undertook in the paper "Lining Up: Survey and Administrative Data Estimates of Wealth Concentration" discussed above was spurred by an early draft of the paper "Wealth Inequality in the United States since 1913, Evidence from Capitalized Income Tax Data" by Saez and Zucman [12]. As suggested by the title, this important paper explores the possibilities for using income data deriving from personal income tax returns to estimate the distribution of wealth; they also compare their estimates with SCF estimates. Their paper came to me at a time when I was thinking hard about how to or-

ganize nontraditional data, including government administrative data, for analytical use. Some of the differences between results in the Saez and Zucman paper and in SCF estimates of wealth distribution could be attributed to accounting or similar definitional differences, as discussed in Bricker et al. in "Measuring Income and Wealth at the Top Using Administrative and Survey Data" [13]. However, other aspects of their work suggested to me that there might be important issues related to the way the SCF list sample captures the very top of the wealth distribution. As discussed above, I found support in the data to question the representation of the survey in that region of the distribution. I highlight here some of the additional work I believe is needed.

Wealth has many possible definitions, and many possibilities for valuation.[9] Where there are multiple approaches, as in the Saez and Zucman and the SCF approaches, it is helpful to understand the sources of differences. In the case of these approaches, I believe there may also be important potential synergies. As noted earlier in this essay, the SCF already takes a model-based approach to the rank ordering of the wealth distribution for purposes of sample stratification, using a model similar to that of Saez and Zucman. Wealth measured in the survey may differ from such projected wealth due to conceptual differences or reporting errors. An exact match of the list sample respondents in the SCF to income data concurrent with the wealth measurement would allow a more detailed reconciliation of differences.

Currently, it is not possible to make such a direct linkage. Wealth is measured in the survey as of the time of the interview, but the available tax-based sample data date from a time at least two years before each survey was executed.[10] Matching data more nearly concurrent with the wealth measures would require both an expanded agreement with SOI and an environment for matching that would protect the statistical nature of the data and the confidentiality of the respondents.

If the two sources could be sufficiently reconciled both conceptually and empirically, it might then be possible to create alternative wealth projections exploiting the correlations of wealth components measured in the survey with the income and other mea-

---

[9]See, for example, my paper "Ponds and Streams: Wealth and Income in the U.S, 1989 to 2007" [14].

[10]I understand that for the 2016 SCF, a decision was made to push the time back by an additional year and consider an additional stage of weight adjustment to compensate.

sures in the tax data. Such projections would clearly dominate simple extrapolations of baseline survey data made by applying external changes observed at a more highly aggregated level – such as changes in overall stock market valuation applied uniformly to total baseline stock values in each observation. Moreover, the contextual data available in the survey would enable research that would not be possible with the projections alone. As surveys, including the SCF, appear to grow ever more costly, in substantial part owing to efforts intended to counter the increasing difficulties of gaining cooperation with respondents, such data-based projections may even have scope to support analysis during an extended interval between surveys.

Acknowledgments

The depth of my gratitude to others is deep and wide. Warmth comes to me from thinking of the names of the people who have helped me on my way. Of course, the very kind people who contributed comments to this volume are high on that list. Here I can touch on only a small number of others. I have specifically thanked many others in the individual papers included in this issue and elsewhere.

Dorothy Projector, the director of the predecessor of the SCF, set a high mark of quality that would be very difficult to reach in any other wealth survey. Although I only once heard her voice and then only in a phone call, her work and that of her associates was an inspiration to me. Fritz Scheuren has been my longtime friend and colleague. When he was Director of Statistics of Income at the Internal Revenue Service, he provided the data needed for the SCF to oversample wealthy households. He served me as a sounding board for the SCF and many other ideas since then. Fritz does not know the meaning of "no". In particular, he has been very adept over time in pushing me and others to do things that might not otherwise have been done, but that in retrospect seem wise to have undertaken. This volume is a special instance of his arm twisting, which I hope will work out as well for others. Barry Johnson, now Director of Statistics of Income, has been my close friend and colleague for nearly as long as Fritz. His support for the survey and his positive attitude made it possible to pass many hurdles that would otherwise have been impassable. Robert Avery was my predecessor as director of the SCF and I am grateful to him for first allowing me to come into this fascinating work. Gerhard Fries worked with me for almost the entire time of my association with the SCF. His dedication and loyalty, especially during the time when there were only to the two of us to carry on the SCF work at the FRB, were incomparable.

I have been very lucky to be able to work with many practical and creative survey experts who have shared their insights with me. I can name only a small fraction of them here: Catherine Haggerty, Micah Sjoblom and their "heavenly" team in the central office at NORC and members of the field management staff, including Maureen Bonner, Lee Brandon, Lynn Gallagher, Vivian Howard, Ella Kemp, Sandra Pitzer, Nina Walker and many others. I have cited a much more comprehensive list in the SCF codebooks.[11] Many of the hundreds of SCF interviewers I have met were open with me and changed my views on many things; the interviewers also did the hardest work, bravely and creatively coping with the imperfections in what they were given. I have very deep gratitude to all the respondents to the SCF, without whom this work would not exist.

Starting from a relatively late point in my career, my thinking has been very much enriched by association with a variety of international survey efforts, especially in the Chile, China, Europe and the Philippines. My longest and deepest such association has been with the Household Finance and Consumption Survey, which is coordinated by the European Central Bank and implemented in a federated way by the euro-area national central banks. Having been involved from the very start of the design stage for that survey, I feel I have witnessed the birth and further development of a brilliant addition to economic measurement in Europe. That work has also taught me more about the advantages of federated systems for resisting monocultures that suppress the diversity of views, a view I wish could be shared more broadly in institutions. Navigating differences in such a framework, though sometimes difficult, makes for stronger results when people of good will are involved, as in this case. Although perhaps seemingly far from the papers included here, my work with the Global Legal Entity Identifier System,[12] a global effort to bring greater transparency and stability to financial markets, has also deeply shaped my thoughts about complex systems of measurement in an international context and I believe it has benefitted from my survey work.

---

[11]See the acknowledgement section of the SCF codebooks found on the project website, https://www.federalreserve.gov/econresdata/scf/scfindex.htm.

[12]See https://www.gleif.org/en/lei-system/global-lei-system-a-network-of-federated-services.

Thanks to Kirsten West, editor of the SJIAOS, and Greta Cherry, her editorial assistant, together with Kim Willems and Xiaoxiao at IOS Press who have done so much of the work to produce this volume. Thanks also to Katherine Condon for her very kind representation of a talk we had together.

Finally, thanks to my friends and my husband Joseph Holbach and his wonderful family for tolerating my work obsessions and my consequent absences and distractions. Without their support, I would have been a lot less productive and a lot less happy.

Supplementary data

The supplementary files are available to download from http://dx.doi.org/10.3233/SJI-170347.

References

[1]   R.W. Goldsmith, *Studies in the National Balance Sheet of the United States*, National Bureau of Economic Research **2** (1963).

[2]   R.W. Goldsmith, *The National Wealth of the United States in the Postwar Period*, Princeton, N.J.: Princeton University Press, 1962.

[3]   R.W. Goldsmith, R.E. Lipsey and M. Mendelson, *Studies in the National Balance Sheet of the United States*. Princeton, N.J.: Princeton University Press, 1963.

[4]   B. Mallet, A Method of Estimating Capital Wealth from Estate-Duty Statistics, *Journal, Royal Statistical Society of London* **71** (March 1908), 64–84.

[5]   F. Scheuren, Personal Wealth Estimates from Tax Returns Filed during Calendar Year 1963, Internal Revenue Service, https://www.irs.gov/pub/irs-soi/13pwcestatereturn62.pdf, 1962.

[6]   R.J. Lampman, The share of top wealthholders in national wealth, 1922-56. Princeton, N.J.: Princeton University Press, 1962.

[7]   G.K. Holmes, The Concentration of Wealth, *Political Science Quarterly* **3** (1893), 589–600.

[8]   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*, vol. 28, pp. 403–408.

[9]   R.M. Groves and S.G. Heeringa, Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs, *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **169**(3) (2006), 439–457.

[10] A.B. Kennickell, Dirty and Unknown: Statistical Editing and Imputation in the SCF, *Statistical Journal of the IAOS* **31**(3) (2015), 435–445.

[11] A.B. Kennickell, Curbstoning and Culture, *Statistical Journal of the IAOS* **31**(2) (2015), 237–240.

[12] E. Saez and G. Zucman, Wealth Inequality in the United States since 1913, Evidence from Capitalized Income Tax Data, *Quarterly Journal of Economics* (2016), 519–578,

[13] J. Bricker, A. Henriques, J. Krimmel and J. Sabelhaus, Measuring Income and Wealth at the Top Using Administrative and Survey Data, FEDS Working Paper, 2015-30, Board of Governors of the Federal Reserve System, 2015.

[14] A.B. Kennickell, "Ponds and Streams: Wealth and Income in the U.S, 1989 to 2007" Finance and Economic Discussion Series, 2009-13, www.federalreserve.gov/pubs/feds/2009/200913/200913pap.pdf.