QUANTIFYING UNCERTAINTIES IN ESTIMATES OF INCOME AND WEALTH INEQUALITY

ONLINE SUPPLEMENTARY MATERIAL

MARTA BOCZOŃ

1st January, 2021

Appendix A. Technical Appendix on the SCF

A.1. **The SCF sampling design.** The SCF employs a dual-frame sample design consisting of a multi-stage national area-probability (AP) sample, designed to provide a good coverage of variables widely distributed in the population (such as balances on checking accounts) and a list sample, designed to provide sufficient coverage of variables largely concentrated in the upper tail of the wealth distribution (such as corporate stocks).

A.1.1. *AP sampling.* As discussed in Tourangeau et al. (1993) and Kennickell (2000), the AP sample is selected in two main stages. In the first stage, the Primary Sampling Units (PSUs) are selected, and in the second stage, secondary units are selected within each of the pre-selected PSUs.

In the first stage of the AP sampling procedure, all 50 states and the District of Columbia are divided into counties (or county-like units) from which PSUs are formed. Once the PSUs

Key words and phrases. Survey of Consumer Finances, Statistics of Income's Public-Use Microdata Files, Total Survey Error. *JEL*: C63, C82, C83, G50, G51.

Author: Boczoń: Department of Economics, University of Pittsburgh, 230 South Bouquet Street, Wesley W. Posvar Hall, Pittsburgh PA, 15213. E-mail: martaboczon@pitt.edu. Foremost, I would like to express my sincere gratitude to my supervisor Jean-François Richard, to whom I would like to dedicate this project, for his continuous support throughout my Ph.D. studies and particularly for his dedication, encouragement, and immense knowledge. He has been not only my primary advisor but also my mentor. Besides my supervisor, I would like to thank the rest of my thesis committee: Arie Beresteanu, Roman Liesenfeld, and Alistair Wilson for their motivation and insightful comments. My completion of this project could not have been possible without the support of Jesse Bricker, John Czajka, Daniel Feenberg, Kevin Moore, William Peterman, and Allison Schertzer. For helpful feedback, I also thank Claudia Sahm, Janine Carlock, and the University of Pittsburgh Writing Center. Lastly, I am grateful to the Federal Reserve Board of Governors Dissertation Internship program for research support. Remaining errors are my sole responsibility.

are formed, a Probability Proportional to Size (PPS) sampling procedure is applied in order to determine the subset of PSUs to be included in the sample.

For illustration, consider a universe of 2,500 PSUs from which 100 PSUs are to be selected. Assuming a total number of housing units across all PSUs to be equal to 102,263,678, the sampling interval is given by 1,022,636.78 (which is 102,263,678 divided by 100). Elaborating on this example, consider a metropolitan PSU comprised of 7,042,994 housing units (i.e., New York Consolidated Metropolitan Statistical Area as of 1990). Since this PSU has a measure of size exceeding the sampling interval ($7,042,994 > 1,022,637$), it would be repeatedly selected under the PPS sampling scheme, which would reduce the total number of distinct PSUs included in the sample. Hence, all such *self-representing* PSUs with a measure of size larger than the sampling interval are automatically selected for the second stage AP sampling (to be discussed further below). In the next step, a new sampling interval is determined that accounts for the number of self-representing PSUs already included in the sample and their combined measure of housing units. Continuing with the above example, assume that 12 self-representing PSUs with a joint measure of 30,455,539 housing units were selected, which implies a new sampling interval of 816,001.6 (which is 102,263,678 less 30,455,539 divided by 88). After determining this new sampling interval and selecting a random starting point of the PPS sampling scheme, PSUs are sorted (according to the set of sorting variables, such as geographic area, minority composition, and per capita income) and next, sampled until the remaining 88 *non-self-representing* PSUs are selected for the second stage of the AP sampling procedure.[1]

In the second stage of the AP sampling procedure, second-stage sampling units, referred to as *segments*, are selected within each PSU. First, all PSUs are divided into segments formed from one or more contiguous census blocks located in the same census tract (or block numbering area). In most cases, populous census blocks constitute a single segment while undersized blocks are linked with neighboring blocks to form segments with a sufficient number of housing units. Once segments have been formed, a sampling interval is determined independently for each PSU by dividing the total number of housing units by a predetermined number of segments to be selected. Next, a random starting point for the PPS sampling procedure is selected and census blocks within each census tract are sorted with respect to their socioeconomic status and ethnic composition. As in the first stage, this stratification has the objective of enhancing the sampling precision of the overall design. Finally, segments are selected from each PSU using the PPS sampling design.

[1] If a measure of size of any of the PSUs in the list still exceeds the sampling interval, step 1 is repeated: another set of self-representing PSUs is included in the sample and removed from the list.

A.1.2. *List sampling.* In the present section, I characterize the other sample selection scheme for the SCF, which is list sampling. As emphasized by Kennickell (2008), list sampling has two main objectives, which are to facilitate estimation of wealth in the far right tail of the wealth distribution and correct for non-response bias in estimates of wealth.

For the SCF, the list sample is selected in a two-stage procedure, aided by a pre-existing sample of individual income tax returns commonly referred to as the INSOLE.[2] In the first stage, tax returns are divided into two sets on the basis of the geographic area the returns were filed in. Those filed in one of the PSUs selected for the AP sample are identified as eligible for the list sample, while those filed in one of the PSUs not included in the AP sample are identified as ineligible. In the second stage, the set of pre-determined eligible tax returns is classified into one of seven strata defined using percentiles of a wealth index constructed using the INSOLE sample. Next, each stratum is sampled using a systematic random sampling, with wealthier strata being sampled at higher rates.[3]

A.2. **Partially missing values in the SCF.** Since 1983, the SCF has gradually allowed for the possibility of reporting partial (range) information about dollar amounts in an effort to reduce the number of completely missing cases. Beginning with the 1995 SCF, respondents were offered the option of providing a single dollar amount, volunteering a range of values, answering *don't know*, or refusing to answer the question.

If a respondent answers with a complete dollar amount, the Computer-Assisted Personal Interviewing (CAPI) program proceeds to the next question. If a respondent volunteers to reveal partial information, she has the option of constructing bounds of her choice (in some cases the upper or lower bound may be missing, for example when a respondent answers *greater than a million dollars*) or selecting an interval from a standardized list consisting of twenty intervals ranging from between $1 and $100 to $100 million and more. If a respondent either does not know the answer or refuses to answer the question, she is asked to reveal partial information; if a respondent continues to refuse, the CAPI program proceeds to the next question; if a respondent agrees to reveal partial information, she is asked about a dollar range with the option of providing her own or selecting one from the standardized list (see above); if the answer is *don't know*, a respondent is asked a series of questions within a decision tree designed to guide her into a range response. Intervals in a decision tree are defined in terms of seven monotonically increasing values (dollar breaks) that vary by question, for example, $100, $250, $500, $1,000, $2,000, $10,000, and $50,000. This results in

[2]See Section B.1.1 below for a detailed description of the INSOLE sample.

[3]The list sample excludes people who are listed on the Forbes 400 list due to concerns about endangering their confidentiality and low response rates.

a total of twenty two unique outcomes for each out of nine different versions of the decision tree.

As a result of this comprehensive data collection process, each observation falls into one of three main categories: not missing (or originally missing but inferred with high confidence from other information, approximately 20 subcategories), partially missing (i.e., range responses, approximately 600 subcategories), and completely missing (approximately 30 subcategories), where the large range of categories reflects a sizable number of plausible realizations of the decision tree algorithm.

A.3. **Rubin's variance estimator.** Variance of a multiple imputation estimator of θ, say $\hat{\theta}_{MI}$, can be rewritten as

$$Var\left(\hat{\theta}_{MI}\right) = Var\left(\hat{\theta}_n\right) + Var\left(\hat{\theta}_{MI} - \hat{\theta}_n\right) + Cov\left(\hat{\theta}_{MI} - \hat{\theta}_n, \hat{\theta}_n\right), \tag{A.1.1}$$

where $\hat{\theta}_n$ denotes the complete sample estimator of θ.

As indicated in Yang and Kim (2016) "Rubin's variance estimator is asymptotically unbiased if $Cov\left(\hat{\theta}_{MI} - \hat{\theta}_n, \hat{\theta}_n\right) \approx 0$, which is called the congeniality condition by Meng (1994)." Specifically, "the uncongeniality arises when the analyst and the imputer have access to different amounts and sources of information, and have different assessments (e.g., explicit model, implicit judgment) about both responses and nonresponses" (Meng, 1994). In the context of the SCF, since details regarding the imputation model are strictly confidential, the congeniality condition is not met, and consequently, Rubin's variance estimator is asymptotically biased.

A.4. **Measurement error in the SCF.** Respondent-related measurement error occurs when respondents provide untrue or incomplete information, either intentionally or unknowingly. In order to better understand the nature of this type of error is it useful to refer to one of the existing and well-established models of the survey response process, discussed in great detail in Groves et al. (2009) and Biemer and Lyberg (2003). The model consists of six major components: *encoding and record formation* (a process of learning information), *comprehension* (a process of interpreting the question and determining the set of permissible answers), *retrieval* (a process of retrieving from memory information relevant to answering the question), *judgment and estimation* (a process of combining, summarizing, and supplementing what the respondent has retrieved), *response formatting* (a process of formulating the answer in the form required by the survey question), and *editing and communication* (a process of deciding whether or not to alter the already formulated answer out of social

desirability, fear of disclosure, or other factors). In the following, I characterize respondent-related measurement error by identifying various problems that are likely to occur at each stage of the aforementioned survey response process.

Encoding and record formation. The first among six major problems known to undermine the survey response process is failure to encode the information sought. As summarized in Groves et al. (2009) "people cannot provide information they do not have; if people never encoded the information in the first place, then no question, no matter how cleverly framed, is going to elicit accurate responses."

In the context of the SCF, failure to encode the information sought may occur when the household member selected for the interview has limited information about the family's finances. Specifically, problems are likely to occur when respondents are not in charge of household finances or handle only a marginal portion of a household's financial accounts. Additionally, difficulties may arise when respondents maintain finances separate from a spouse or partner, are not aware of certain financial decisions made by the spouse (for example regarding life insurance policies, risky portfolio allocations, or excessive credit card debt), or leave their household's finances to accountants and/or tax advisors unavailable for consultation during the SCF interview. Lastly, problems can arise if respondents participate in automatic loan payment programs as participation in such programs means respondents may lack the opportunity to encode the information related to the number of payments left on the loan. Similarly, those who sign up for credit card automatic payments and pay a minimum balance due each month may routinely fail to encode the amount of revolving balance.

Comprehension. The second major issue that can impair the response process is misinterpretation of the survey question. As pointed out in Groves et al. (2009), "even if the respondents know the answers to the questions, they are unlikely to report them if they misunderstand the questions." Such misinterpretations are most likely to result from grammatical ambiguity, excessive complexity, and vague concepts used in the interview process.

For instance, respondents with low levels of financial literacy may lack the ability to quickly comprehend various financial terms used during the SCF interview (even if previously explained by the interviewer). Consequently, they may confuse or misuse terms such as face and market value or revolving and non-revolving credit, or fail to differentiate between different financial products say, term and cash-value life insurance policies or 401(k) and IRS pension plans. In these situations, it is likely that answers provided by the respondents will be incorrect and, in fact, pertain to a different set of questions.

Retrieval. Another potential source of error is related to memory failure.. Memory failures are likely to occur "when the terms the respondent uses to encode an event differ so markedly

from the terms used in the question that the question does not call to mind the intended memories" (Groves et al., 2009). On top of that respondents may face difficulties when reconstructing memories or attempting to retrieve information from the past.

For example, SCF respondents may experience problems when attempting to reconstruct memories regarding exact balances on numerous checking accounts used on a day-to-day basis by multiple family members. Difficulties may also arise when retrieving information about student loans in deferment or long-term investments, such as real estate or the 30-year Treasury Bond.

Judgment and estimation. The fourth major issue that can interfere negatively with the response process is flawed judgment with respect to attitude and behavioral questions. As emphasized by Groves et al. (2009) "some respondents may have a preexisting judgment (...) that they are ready to report in response to a question (...), but most respondents probably have to put together a judgment on the fly (...) With both attitude and behavioral questions, the need for respondents to put together judgments on the spot can lead to errors in the answers."

For example, SCF respondents may experience difficulties answering questions about home value, typical payment amounts, or the amount of time needed to pay off the loan. Such questions often require respondents to make rapid evaluations of their current financial situation and/or (ad hoc) assumptions about the future state of the economy, as well as carry out a series of (back-of-the-envelope) calculations within a very short period of time.

Response formatting. Another potential source of error is related to problems in formatting an answer in a situation where respondents are faced with multiple response options from which one or more answers need to be selected. It has been found that "most researchers believe that respondents are likely to consider the response options one at a time and to select the first one that seems to provide an adequate answer. Using the terminology of Krosnick (1999), respondents satisfice [to pursue the minimum satisfactory condition or outcome] rather than optimize; they pick an answer that is good enough, not necessarily the one that is best" (Groves et al., 2009).

For example, when asked about sources of information used prior to investment decision making, respondents may choose the first answer that seems adequate (say friends and relatives) instead of considering the complete list of available options and arriving at the most suitable answer (say financial planner).

Editing and communication. Finally, besides facing cognitive difficulties when attempting to answer a survey question, respondents may deliberately choose to misreport. "Sometimes,

refusing to answer a question may be more awkward than simply underreporting some embarrassing behavior" (Groves et al., 2009). As such, survey respondents often underreport socially undesirable behaviors, such as drug use, alcohol consumption, smoking, or abortion. On the other hand, "respondents may also be reluctant to admit that they have not done something when they feel they should have" (Groves et al., 2009). This often results in overreporting certain socially-desirable behaviors, such as voting, practicing religion, exercising, owning a library card, wearing a seat belt, giving to charities, and participating in safe-sex practices. Among other plausible reasons for respondents' deliberate dishonesty are fear of disclosure, self-serving lying aimed at building one's self-esteem, making an attempt to please data analysts by providing them with a set of responses that support their underlying hypotheses, and manipulating targeted at influencing survey outcomes in one's favor.

For instance, in order to avoid embarrassment in the eyes of the interviewer, the SCF respondents at the bottom of the wealth distribution may overreport earnings and underreport debts, whereas the wealthiest households may underreport their wealth status due to security concerns or fear of increased taxation.

Table A.1.1. Frequency of referring to any documents to answer any question

Frequently	**Sometimes**	**Rarely**	**Never**
10.8%	18.7%	13.2%	57.4%
(10.1% – 11.5%)	(17.7% – 19.6%)	(12.4% – 14.0%)	(56.1% – 58.7%)

95% confidence intervals in the parentheses. Computed using the 2016 SCF.

Many of the above problems arising in the response process (such as encoding and record formation, comprehension, retrieval, and judgment and estimation) could be mitigated if respondents consulted supplemental documents (such as income tax returns, pension documents, account statements, inheritance papers, or real estate records) in order to seek information they could not provide otherwise. However, as shown in Table A.1.1, only 11 percent of respondents frequently referred to any kind of financial or non-financial documentation throughout their 2015 SCF interview, with almost 60 percent never referring to any auxiliary sources.

Furthermore, as indicated in Table A.1.2, among the roughly 40 percent of respondents who consulted auxiliary sources of information at least once throughout their 2015 interview, only half referred to more than a single class of documents. For example, I have found that among the credit card holders (bank-, store-, or company-branded), only 0.62 percent referred to their credit card statements at any time during the interview.

Table A.1.2. Number of different types of documents referred to during the interview

1 document	2 documents	3 documents	$\geq$ 4 documents
50.6%	25.1%	16.0%	8.4%
(49.0% – 52.2%)	(23.6% – 26.5%)	(14.7% – 17.3%)	(7.5% – 9.4%)

95% confidence intervals in the parentheses. Computed using the 2016 SCF.

A.5. **Residual autocorrelation in the SCF.** Even though the SCF is a cross-sectional survey and not a panel, the data may suffer from serial correlation. Bell and Wilcox (1993) explain that "in general, sampling errors in repeated surveys can be autocorrelated if (1) time series for individual population units are autocorrelated and (2) samples at different time points are not drawn independently, e.g., if they have specified overlapping segments."

In the first stage of the list sampling, observations in PSUs drawn for the AP sample are selected. One could be concerned that since some of the PSUs formed for the AP sample are *self-representing* across all survey years, and therefore somewhat fixed over time, (aside from some population increases/decreases such as for the NYC Metropolitan Area), there may be an overlap among richest households in the list sample from one survey year to another, causing errors to be correlated. However, one key element of the list sample design should dampen any potential correlation that may arise among the wealthiest households. Families interviewed for the list sample in one survey are not eligible to be included in the sample for the next two surveys, meaning that any list sample household interviewed in 2009 would not be eligible for sampling in 2012 or 2015. This, in combination with the three-year gap in between any two surveys, should minimize any potential serial correlation at the top of the wealth distribution.[4]

Appendix B. Technical Appendix on the PUF

B.1. **The PUF sampling design.** In order to set the scene for the estimation of the PUF standard errors, I will first describe the key features of the INSOLE and PUF sampling designs. The reason for characterizing both designs is that the PUF is sub-sampled from the INSOLE, as opposed to being drawn directly from the underlying population of tax returns. Therefore, in order to fully characterize the PUF sampling design, it is necessary to start with a brief description of the INSOLE sample.

B.1.1. *INSOLE sampling.* The INSOLE is a disproportionate (highly) stratified probability sample of individual income tax returns. The returns are stratified using three variables:

[4]Since the probability of drawing the same AP household in back-to-back surveys is marginal, the AP sample is unlikely to suffer from serial correlation.

gross positive or gross negative income, presence or absence of special forms and schedules, and the return's potential usefulness for tax policy modeling (referred to as a return's *degree of interest*).[5] Within each stratum, returns are selected for the sample using two sampling techniques, both of which are based upon the primary filer's Social Security Number (SSN).[6,7]

The first sampling technique focuses on the last four digits of the SSN. In this method, a return is selected for the INSOLE if and only if the last four digits of the primary filer's SSN match one of ten four-digit numbers chosen from the Social Security Administration's Continuous Work History Sample (CWHS).[8] This method, which I refer to as the CWHS selection, gives all taxpayers approximately a 1 in 1,000 chance of being selected, regardless of the strata they are assigned to.

The second sampling technique relies upon the "SSN transform," which is a uniformly distributed five-digit random number generated from the taxpayer's SSN. In this method, a return is selected for the INSOLE if and only if the taxpayer's SSN transform is less than or equal to the stratum-specific "sample number" given by

$$\kappa_j = 100,000 \times (s_j - 0.001 + 0.001 \times s_j) - 1, \tag{B.1.1}$$

where s_j denotes the prescribed sampling rate in stratum j.[9]

As indicated in equation (B.1.1), the sample number chosen to yield a $100s_j$ percent sample is set with an allowance for the CWHS selection. For illustration, consider stratum $j^\star$ with the prescribed sampling rate of $s_j^\star = 0.10$. It follows that returns in $j^\star$ have a CWHS selection probability of 0.001 and a probability of being selected based on the SSN transform being equal to 0.0991 ($= 0.10 - 0.001 + 0.10 \times 0.001$). This results in the sample number $\kappa_{j^\star}$ equal to $9,909$ ($= 100,000 \times 0.0991 - 1$). Consequently, all returns from stratum $j^\star$ with an SSN transform less than or equal to 9,909 are selected into the sample.

[5]See Section A.4 in Appendix A of the Online Supplementary Material for a detailed description of the stratification of the INSOLE sample.

[6]Note that joint returns and married filing separately (MFS) returns have both a primary and secondary SSN, whereas all other types of returns have only a primary SSN.

[7]The SSN is a nine-digit national identification number issued to all US citizens as well as permanent and temporarily residents by the Social Security Administration (SSA). The SSNs issued prior to 2011 consist of three parts: the Area Number, the Group Number, and the Serial Number. The Area Number (the first three digits of the SSN) is determined by the zip code of the mailing address shown on the application for the SSN; the Group Number (digits four and five) is assigned based on the SSN's issuance date; lastly, the Serial Number (the last four digits) is chosen at random from the set of integers ranging from 1 to 9,999. Following a reform of the SSN Numbering Scheme from June 2011, all nine digits of the SSN are assigned randomly.

[8]Prior to 2005 there were only five CWHS endings in the INSOLE.

[9]The "-1" is a correction term that accounts for the fact that the range of the sample number is between 0 and 99,999. Otherwise, a sampling rate of 100 percent would yield a sample number of 100,000, which is outside the range.

Note that one of the most important features of the SSN transform is that it is constant across different tax years for a given taxpayer. This implies that if once selected for the INSOLE, the taxpayer continues to be selected for as long as he or she remains a primary filer and qualifies for a stratum with the same or higher sampling rate. On the contrary, if a taxpayer drops from a stratum with a 10 percent selection probability to one with a 5 percent selection probability, the probability of him or her being retained in the sample is equal to 50 percent.[10]

B.1.2. *PUF sampling.* The PUF is a disproportionate (highly) stratified probability sample of individual income tax returns. It is obtained by sub-sampling the INSOLE, which in turn is drawn directly from the underlying population of tax returns (see Section B.1.1). The INSOLE is sub-sampled for the PUF at different rates depending on two factors: the stratum from which a return was initially selected and the method used in the sampling process (the CWHS selection versus selection based on the SSN transform). Accordingly, the sub-sampling rates vary considerably within and across strata and range from zero (exclusion from the PUF) to one (drawn with certainty).

Across all strata, returns that are sampled for the INSOLE using the CWHS selection (as opposed to selection based on the SSN transform) are sub-sampled at a rate of 30 percent (or 70 percent following the PUF redesign from 2009).[11] In the first step of the two-step sub-sampling process, three (seven) numbers are drawn at random from a set of four-digit integers designated for the CWHS selection of the INSOLE. An important feature of this design is that once drawn at random, the three (seven) CWHS endings are retained in the sample. This implies that the same three (seven) four-digit integers were considered for the CWHS sub-sampling process every year between 1991 and 2008 (2009 and 2012). In the second step, a return is selected for the PUF if and only if the last four digits of the primary filer's SSN match one of the three (seven) numbers drawn in step one. This sub-sampling process gives all taxpayers a 3 in 9,999 (7 in 9,999) chance of being selected into the PUF, irrespective of the stratum. This compares to a 1 in 1,000 chance of being sampled for the INSOLE based on the CWHS selection process.

Returns that are sampled for the INSOLE using selection based on the SSN transform are sub-sampled for the PUF at different rates across different strata. The sub-sampling rates vary from 0 to 1 and can be classified into one of three main categories: drawn with certainty, sub-sampled at the rate that yields a 10 percent PUF sampling rate, and excluded from the

[10] More information on the INSOLE sampling design can be found in Czajka et al. (2014).

[11] For more details regarding the redesign see Bryant et al. (2014).

PUF.[12] For example, in year 2008, strata with gross positive or gross negative income between \$250,000 and \$1 million (in 1991 US dollars) were sampled with certainty; those with gross positive or gross negative income between \$1 and \$5 million were sub-sampled at rates that imply a 1 in 10 PUF sampling rate; and, returns with extreme values from strata with gross positive or gross negative income of over \$1 million were excluded from the sample.

B.2. **Stratification of the INSOLE sample.** In the following, I present details about the stratification of the INSOLE sample. Specifically, in Section B.2.1, I discuss stratification by income, in Section B.2.2, stratification by the presence or absence of special forms and schedules, and in Section B.2.3 stratification by a return's degree of interest. Section B.2.4 characterizes two INSOLE priority strata, and Section B.2.5 concludes.

B.2.1. *Stratification by income.* As detailed in Czajka et al. (2014), the income measure used for stratification, say φ, is defined as the maximum between taxpayer's gross positive income and the absolute value of a taxpayer's gross negative income. Gross positive income is calculated as the sum of (i) twelve strictly positive items from Form 1040 and Schedule E and (ii) eleven items from Form 1040 and Schedules C, D, and F included only if the number is positive (see left panel of Table B.1.1). Similarly, gross negative income is calculated as the sum of (i) seven losses reported on Forms 1040 and 3903 and Schedules C and E, (ii) eleven items from Form 1040 and Schedules C, D, and F included (in absolute value) only if the number is negative, (iii) two deduction items from Schedules C and F, and (iv) three negative income adjustment items from Schedules C, E, and F (see right panel of Table B.1.1). Based on the constructed income measure φ, tax returns are stratified into one of nineteen income categories ranging from negative \$10,000,000 or less to positive \$10,000,000 or more, where the income classes are deflated using the Chain-Type Price Index for the Gross Domestic Product as of 1991.

The nine negative income classes are: \$10 million or more (income level 1); \$5–\$10 million (income level 2); \$2–\$5 million (income level 3); \$1 –\$2 million (income level 4); \$0.5–\$1 million (income level 5); \$250,000 –\$500,000 (income level 6); \$120,000–\$250,000 (income level 7); \$60,000–\$120,000 (income level 8); under \$60,000 (income level 9). The ten positive income classes are: under \$30,000 (income level 10); \$30,000–\$60,000 (income level 11); \$60,000–\$120,000 (income level 12); \$120,000–\$250,000 (income level 13); \$250,000–\$500,000 (income level 14); \$0.5–\$1 million (income level 15); \$1–\$2 million (income level 16); \$2–\$5 million (income level 17); \$5–\$10 million (income level 18); and \$10 million or more (income level 19).

[12]Prior to 2005, any stratum with an INSOLE sampling rate in excess of 1 in 3 was subsampled for the PUF at a rate of 1 in 3. Thus, if a stratum had an INSOLE rate of 35 percent, the PUF sampling rate would be a little over 1 in 9.

Table B.1.1. Items contributing to gross positive and negative income

Gross Positive Income		Gross Negative Income	
Item	Source	Item	Source
Strictly Positive Items		Loss Items	
1. Wage amount	1040	1. Partnership, s corporation loss	E
2. Tax exempt interest	1040	2. Estate and trust loss	E
3. Taxable dividends	1040	3. Total expenses all property amount	E
4. Alimony received	1040	4. Total depreciation all property amount	E
5. Pension amount	1040	4. Alimony paid	1040
6. IRA distribution	1040	6. Form 3903 moving expense amount	3903
7. Unemployment compensation	1040	7. Business-at-home expenses	C
8. Social Security	1040		
		Business Loss Items (if negative)	
Strictly Gain Items		1. Schedule C-1 gross profit/loss	C
1. Total rental payments amount	E	2. Schedule C-2 gross profit/loss	C
2. Total royalty payments amount	E	3. Schedule C-3 gross profit/loss	C
3. Partnership, s corporation income	E	4. Schedule F-1 gross profit/loss	F
4. Estate and trust income	E	5. Schedule F-2 gross profit/loss	F
Business Income Items (if positive)		Net Items	
1. Schedule C-1 gross profit/loss	C	1. Supplemental gains/losses	1040
2. Schedule C-2 gross profit/loss	C	2. Other income amount	1040
3. Schedule C-3 gross profit/loss	C	3. Farm/rent income/loss	1040
4. Schedule F-1 gross profit/loss	F	4. Taxable interest income	1040
5. Schedule F-2 gross profit/loss	F	5. Net short-term gain/loss amount	D
		6. Net long-term gain/loss amount	D
Net Items		Deduction Items	
1. Supplemental gains/losses	1040	1. Total deductions	C
2. Other income amount	1040	2. Total farm expenses	F
3. Farm/rent income/loss	1040		
4. Taxable interest income	1040	Adjustment Items	
5. Net short-term gain/loss amount	D	1. Negative income adjustment	C
6. Net long-term gain/loss amount	D	2. Negative income adjustment	E
		3. Negative income adjustment	F

The letters C, D, E, and F denote a source tax schedule, and the numbers 1040 and 3903 a source tax form. Form 1040 is the standard IRS form used for filing purposes whereas Form 3903 summarizes moving expenses. Schedule C summarizes profits and losses from businesses, Schedule D capital gains and losses, Schedule E supplemental income and losses, and Schedule F profits and losses from farming. Source: Czajka et al. (2014).

B.2.2. *Stratification by degree of interest.* In addition to stratification by income, tax returns are further stratified by degree of interest, where "within the same income class, returns are considered more useful (or 'interesting') if less common income sources or deductions are prominent" (Czajka et al., 2014). The degree of interest ranges from 1 (the least "interesting") to 4 (the most "interesting") and sub-stratifies four income classes with a gross positive income less than $250,000. The income class with less than $30,000 is sub-stratified into three classes whereas the other three income categories (i.e., $30,000−$60,000,

$\$60,000 - \$120,000$, $\$120,000 - \$250,000$) are sub-stratified into two. This yields a total of 24 distinct levels of income by degree of interest.

B.2.3. *Stratification by special forms.* The third and final dimension of stratification depends on the presence or absence of the following forms and schedules: Form 2555 (*Foreign Earned Income*), Form 1116 (*Foreign Tax Credit*), Schedule C, and Schedule F. In most years that are commonly referred to as *non-foreign study years*, returns are classified into one of four mutually exclusive and collectively exhaustive categories based on a form type. The first category consists of returns with either Form 2555, Form 1116, or both; the second category consists of returns without Forms 2555 and 1116, but with Schedule C; the third category consists of returns without Form 2555, Form 1116, or Schedule C, but with Schedule F; finally, the fourth category consists of all other returns. In the remaining years (i.e., those ending in either 1 or 6, and commonly referred to as *foreign study years*), the first category is subdivided into two, one with returns with Form 2555 and the other with returns with Form 1116 but without Form 2555. This distinction between Forms 2555 and 1116 has the objective of over-sampling returns with Form 2555 in order to facilitate the use of the INSOLE in foreign income studies.

B.2.4. *Priority strata.* In addition to regular strata (i.e., strata generated by the combination of income level, degree of interest, and form type), there are two *priority* strata that take precedence over all regular strata and are sampled with certainty. This implies that a return is to be sampled from a regular stratum if and only if it does not qualify for any of the priority strata.

The first priority stratum (stratum 101) comprises returns with Adjusted Gross Income (AGI) or expanded income of at least $200,000 (in current dollars) and with no income tax liability after subtracting all credits. Therefore, this stratum is often referred to as one with "high-income non-taxable" returns.[13] Since the $200,000 income threshold is not deflated (using the Chain-Type Price Index for the 1991 Gross Domestic Product) but instead expressed in current dollars, the number of returns in this priority stratum has increased over time from 2,757 in 1991 to 4,114 in 2000 to 35,067 in 2012. The second priority stratum (stratum 201) comprises returns with business or profession receipts exceeding $50,000 (in current dollars). Even though the number of returns in this stratum increased over time (from 46 in 1991 through 1,025 in 2000 to 323 in 2012), it continues to be negligible, especially when compared to the number of returns in the other priority stratum.

[13]Expanded income is defined as AGI with "tax-exempt interest, nontaxable Social Security benefits, the foreign-earned income exclusion, and tax preference items used to calculate the alternative minimum tax", less "unreimbursed employee business expenses, moving expenses, investment interest expenses up to the value for investment income, and miscellaneous itemized deductions below the 2 percent of AGI floor" (Czajka et al., 2014).

Table B.1.2. Number of tax returns in the population, INSOLE, and PUF samples by stratum for 2008

Gross income	Degree of interest	Form 1040, with Form 1116 or Form 2555			Form 1040, with Schedule C but without Form 1116 or Form 2555			Form 1040, with Schedule F but without Schedule C, Form 1116 or Form 2555			Other forms			Total		
		Population	INSOLE	PUF	Population	INSOLE	PUF	Population	INSOLE	PUF	Population	INSOLE	PUF	Population	INSOLE	PUF
Negative income																
>=$10M	All	456	456	46	1,029	1,029	123	144	144	14	1,182	1,182	117	2,811	2,811	300
$5–$10M	All	860	860	78	1,724	1,724	175	260	260	16	2,202	2,202	231	5,046	5,046	500
$2–$5M	All	3,590	1,182	350	6,543	2,172	670	959	336	101	7,834	2,699	785	18,926	6,389	1,906
$1–$2M	All	7,462	1,203	756	13,451	2,126	1,318	2,358	388	241	15,571	2,436	1,492	38,842	6,153	3,807
$0.5–$1M	All	16,922	552	543	32,988	1,138	1,120	5,730	200	194	36,138	1,145	1,128	91,778	3,035	2,985
$250–$500K	All	33,620	350	325	74,321	715	658	11,769	115	106	80,568	806	757	200,278	1,986	1,846
$120–$250K	All	60,376	287	239	155,147	788	680	19,156	120	102	172,644	886	759	407,323	2,081	1,780
$60–$120K	All	70,010	196	154	201,319	649	479	20,470	72	54	249,179	785	589	540,978	1,702	1,276
<=$60K	All	61,200	108	70	473,691	888	535	28,417	51	30	800,133	1,469	884	1,363,441	2,516	1,519
Positive income																
<=$30K	**1**	⋆⋆	⋆⋆	⋆⋆	⋆⋆	⋆⋆	⋆⋆	⋆⋆	⋆⋆	⋆⋆	31,485,356	31,416	9,378	31,485,356	31,416	9,378
<=$30K	**2**	274,562	270	82	3,046,298	2,977	951	83,688	77	20	28,296,966	28,208	8,445	31,701,514	31,532	9,498
<=$30K	3-4	212,182	315	186	4,805,497	7,413	4,047	106,177	192	95	6,355,132	9,747	5,283	11,478,988	17,667	9,611
$30–$60K	1-2	681,524	678	209	1,978,448	2,028	585	171,290	168	46	21,950,763	21,915	6,598	24,782,025	24,789	7,438
$30–$60K	3-4	528,015	846	477	3,758,970	5,841	3,250	244,505	401	215	6,194,112	10,003	5,570	10,725,602	17,091	9,512
$60–$120K	1-3	1,085,559	1,091	336	2,301,810	2,299	672	217,320	242	73	11,353,034	11,257	3,373	14,957,723	14,889	4,454
$60–$120K	**4**	650,212	966	524	2,483,674	3,822	2,047	174,911	225	126	2,812,052	4,279	2,305	6,120,849	9,292	5,002
$120–$250K	1-3	337,092	634	413	401,367	756	503	82,868	188	111	1,265,376	2,473	1597	2,086,703	4,051	2,624
$120–$250K	4	822,634	2,741	2172	1,337,403	4,564	3656	88,221	296	220	1,816,816	5,888	4,628	4,065,074	13,489	10,676
$250–$500K	All	511,639	3,662	3,316	463,861	3,366	3,019	73,527	517	471	592,030	4,199	3,788	1,641,057	11,744	10,594
$0.5–$1M	All	230,080	5,675	5,513	134,302	3,397	3,316	27,545	642	628	157,455	3,908	3,801	549,382	13,622	13,258
$1–$2M	All	93,022	11,189	9,174	35,861	4,417	3,584	7,029	829	667	45,529	5,638	4,610	181,441	22,073	18,035
$2–$5M	All	44,770	14,457	4,512	11,754	3,807	1,189	1,836	590	172	16,142	5,175	1,609	74,502	24,029	7,482
$5–$10M	All	11,812	11,812	1,215	2,336	2,336	251	299	299	24	3,174	3,174	322	17,621	17,621	1,812
>=$10M	All	7,917	7,917	793	1,088	1,088	102	140	140	11	1,487	1,487	144	10,632	10,632	1,050
Total		5,745,516	67,447	31,483	21,722,882	59,340	32,930	1,368,619	6,492	3,737	113,710,875	162,377	68,193	142,547,892	295,656	136,343
Priority stratum 101														32592	32592	3278
Priority stratm 201														382	382	30
Grand total														142,580,866	328,630	139,651

Table summarizes information for 93 regular strata (top panel) and 2 priority strata (bottom panel). ⋆⋆ denotes either that the data were combined or deleted or that no returns in the population had the characteristic or the characteristic was so rare that it did not appear on any of the sampled returns. $1M denotes 1 million US dollars and $1K 1 thousand US dollars. Sources: Internal Revenue Service Research (2010, 2012).

B.2.5. *Conclusion.* The combination of 24 levels of income by degree of interest and 4 form types yields 96 regular strata ($= 24 \times 4$). However, since among returns with gross positive income less than $\$30,000$ only those without any special form or schedule can be assigned a degree of interest equal to 1 (which indicates the least "interesting" return), effectively, the number of regular strata in non-foreign study years is equal to 93 ($= 23 \times 4 + 1$) and in foreign study years to 116 ($= 23 \times 5 + 1$). To conclude, the INSOLE is sampled from a total of 95 ($= 93 + 2$) strata in non-foreign study years, and from a total of 118 ($= 116 + 2$) strata in the years 1991, 1996, 2001, and 2006. For illustration, see Table B.1.2, where I present the population, INSOLE, and PUF counts for each of the 95 strata for tax year 2008.

B.3. **TSE in the PUF.** Sampling error in the PUF arises due to the fact that the data constitute a sample of the universe of income tax filing. For example, for tax year 2008, the number of tax returns exceeds 140 million, whereas the number of PUF (INSOLE) sample observations is equal to 140,000 (143,000). This implies a sampling fraction of 0.1 percent and a 1 in a 1,000 chance of a random taxpayer entering the sample. In addition to sampling error, the PUF (as well as the INSOLE) is subject to nonsampling errors, such as processing, frame, nonresponse, and measurement errors. Note that specification error in the PUF is not likely to occur since the PUF—as a sample of individual-income tax returns—is an ideal representation of the concept intended to be measured in the data.

The principal source of processing error in the PUF includes four nondisclosure procedures applied separately to low- and high-income returns with the objective of protecting identities of taxpayers on either end of the income distribution. Low-income returns are defined as those with an AGI below $200,000 (in current dollars) and high-income returns as those with an AGI greater than or equal to $200,000 (in current dollars). The first nondisclosure procedure is blurring, where individual values of certain variables are replaced with their average values. For example, among high-income returns, blurring is applied to salaries and wages, state and local income taxes, and real estate taxes. Another technique for reducing disclosure risk is rounding, where all dollar amounts, such as business profits or unemployment benefits, are rounded to the four most significant digits. Consequently, dollar amounts under 10,000 (in absolute value) are not rounded and those between 1 million and 10 million are rounded to the nearest thousand. The third nondisclosure procedure, deleting, is applied only to certain variables of high-income returns, such as state codes, state sales tax deductions, and alimony paid and received. As summarized in Bryant et al. (2014), "these variables are deleted because they may be publicly available from other sources so, in combination with other variables on the PUF, could enable identification of a high-income taxpayer and therefore disclose the other items reported on their return." The last nondisclosure procedure that may lead to processing error is modifying. Similar to deleting, modifying pertains only

to high-income returns and a handful of variables such as marital status, the number of dependents, personal exemption amounts, and the number of children for which the child tax credit is claimed.[14]

Frame error occurs because neither the PUF nor the INSOLE covers the entire population of tax units. "A tax unit is an individual, or a married couple, that files a tax return or would file a tax return if their income were high enough, along with all dependents of that individual or married couple."[15] Thus, the population of tax units comprises both filers and non-filers, where filers are obligated by law to file a tax return and non-filers are exempt from filing requirements due to low-enough incomes.[16] Since the PUF is a sample of individual income tax returns, it does not collect any information on non-filers. In order to account for this source of error, the number of non-filers as well as their income levels need to be either imputed or estimated using other data sources such as the Current Population Survey or the Statistics of Income Division databank.

Nonresponse error may occur as, analogous to survey data, and in particular the SCF, the PUF is subject to both unit and item nonresponse. An excellent example of item nonreponse can be found in Bricker et al. (2018):

> Item nonresponse may occur in the tax data when a family does not claim positive income on a type of income when the family does, in fact, have positive income. For example, a 1099-INT is automatically generated by a financial institution when interest income is greater than $10, but when interest income is less than $10 then the family will not get this reminder to claim interest income on their tax filing. Appendix B of Bricker et al. (2016) shows that the number of returns with positive interest income has fallen over the past decade while the number of families with interest-bearing accounts has stayed constant. A decline in interest rates on savings accounts may mean that fewer families will get an automatic 1099-INT reminder.

Unit nonresponse in the PUF occurs mostly due to taxpayers failing to file. As defined by the IRS, a taxpayer fails to file a required tax return when the return is filed after the filing deadline (or a valid extension thereof) or is not filed at all. According to the most recent

[14]For specific information of how the four variables are modified see Bryant et al. (2014).

[15]See the Urban-Brookings Tax Policy Center at https://www.taxpolicycenter.org/resources/tpcs-microsimulation-model-faq (accessed on May 7, 2020).

[16]As stated by the IRS, tax units with gross income below a certain threshold are not required to file a tax form, where the cutoff values are determined each year and vary depending on individuals' marital status, filing status, and age. For example, single individuals under 65 years old (including those divorced and legally separated) with gross income less than $8,950 were exempt from filing requirements for tax year 2008.

report on the Federal tax compliance for tax years 2011–2013 published by the Internal Revenue Service Research (2019), nonfiling accounts for 10 percent of the gross tax gap, which is defined as a difference between the total true tax liability and tax paid voluntarily and in a timely fashion.

The last type of nonsampling error, measurement error, arises when taxpayers misreport on their tax returns. Intentional misreporting is a criminal act whereby taxpayers deliberately misrepresent the state of their finances with the objective of reducing their tax liability. This form of tax evasion occurs when taxpayers intentionally underreport their incomes, overreport deductions, overvalue non-cash donations, or claim a wrong filing status. For example, waiters may underreport their cash tips, self-employed individuals may overreport business deductions, and married filers may falsely claim a spouse with vision impairment. Altogether, this form of tax evasion accouts for 78 percent of the gross tax gap estimated for tax years 2011–2013 (Internal Revenue Service Research, 2019).

In addition to intentional misreporting, taxpayers may provide incorrect information unintentionally. Unintended errors on tax returns can be traced back to computation errors, insufficient financial knowledge, and ignorance of the law, as well as problems differentiating between different tax forms and schedules.

B.4. **Nonsampling errors in the PUF.** In this section, I present qualitative evidence on why processing, measurement, and nonresponse errors in the PUF can be considered marginal.

First, note that before the PUF is released to the general public, all identifiable errors made by taxpayers as well as transcription errors introduced in the IRS processing are corrected by IRS staff members. This significantly helps to reduce both processing and measurement errors.

In addition, measurement error is likely to be fairly negligible since taxpayers are threatened with punishment for misreporting on their tax returns. Both intentional and unintentional misreporting is either a criminal or a civil act punishable by law. Specifically, offenders "must repay the taxes with an expensive fraud penalty and possibly face jail time of up to five years."[17]

Lastly, since the PUF includes returns filed late (after April 15) but within the same calendar year, unit nonreponse is driven mostly by taxpayers who file after at least several months delay or do not file at all. In order to reduce this error, the PUF includes late returns from

[17]See https://www.hrblock.com/tax-center/irs/tax-responsibilities/prision-for-tax-evasion/ (accessed on June 18, 2020).

previous tax years, which are viewed as substitutes for returns from the current tax year that are to be filed with a considerable delay (if at all). By way of example, consider the PUF data for tax year 2008. In addition to returns that were filed in 2009 for tax year 2008, the sample includes late returns filed in 2009 for tax years 2007 and before. In the most recent PUFs, late returns are limited to the prior three tax years, while in the earlier PUFs, late returns could be older than three years.

References

Bell, William R. and David W. Wilcox, "The effect of sampling error on the time series behavior of consumption data," *Journal of Econometrics*, 1993, *55*, 235–265.

Biemer, Paul and Lars Lyberg, *Introduction to survey quality*, Hoboken, New Jersey: John Wiley & Sons, 2003.

Bricker, Jesse, Alice Henriques, and Peter Hansen, "How much has wealth concentration grown in the United States? A re-examination of data from 2001-2013," *Finance and Economics Discussion Series*, 2018, *2018-024*, 1–52.

_____, _____, **Jacob Krimmel, and John Sabelhaus**, "Measuring income and wealth at the top using administrative and survey data," *Brookings Papers on Economic Activity*, 2016, *Spring*, 261–331.

Bryant, Victoria L., John L. Czajka, Georgia Ivsin, and Jim Nunns, "Design changes to the SOI Public Use File (PUF)," *Prepared for the "New Resources for Microdata-Based Tax Analysis" Session of the 107th Annual Conference on Taxation, National Tax Association Santa Fe, New Mexico, November 15*, 2014, pp. 1–19.

Czajka, John L., Brendan Kirwan, and Amang Sukasih, "An assessment of the need for a redesign of the Statistics of Income Individual Tax Sample," *Mathematica Policy Research*, 2014, pp. 1–95.

Groves, Robert M., Floyd F. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau, *Survey methodology*, Hoboken, New Jersey: John Wiley & Sons, 2009.

Internal Revenue Service Research, Applied Analytics & Statistics, "Federal tax compliance research: Tax gap estimates for tax years 2011–2013," *Publication 1415 (Rev. 9-2019)*, 2019, pp. 1–28.

Internal Revenue Service Research, Statistics of Income, "Individual Income Tax Returns 2008," *Publication 1304 (Rev. 07-2010)*, 2010, pp. 1–295.

_____, "General description booklet for the 2008 Public Use Tax File," 2012, pp. 1–42.

Kennickell, Arthur B., "Revisions of the variance estimation procedure for the SCF," *Working Paper of the Board of Governors of the Federal Reseve System*, 2000.

_____, "The role of over-sampling of the wealthy in the Survey of Consumer Finances," *Irving Fisher Committee Bulletin*, 2008, *28.*

Meng, Xiao-Li, "Multiple-Imputation inferences with uncongenial sources of input," *Statistical Science*, 1994, *9* (4), 538–558.

Tourangeau, Roger, Robert A. Johnson, Jiahe Qian, Hee-Choon Shin, and Martin R. Frankel, "Selection of NORC's 1990 National Sample," *Working Paper, National Opinion Research Center at the University of Chicago*, 1993.

Yang, Shu and Jae K. Kim, "A note on multiple imputation for method of moments estimation," *Biometrika*, 2016, *103* (1), 244–251.

Appendix C. Tables and Figures

Table C.1. Summary statistics on the PUF strata before and after clustering

Year	Total		Mean		Min		Median		Max	
	Before	After	Before	After	Before	After	Before	After	Before	After
1991	119	28	971	4,128	2	120	265	2,891	7,722	16,175
1992	145	27	643	3,454	6	201	114	1,626	8,721	12,661
1993	98	23	961	4,096	12	248	320	3,527	8,617	11,869
1994	98	23	984	4,191	12	267	343	3,622	8,329	12,452
1995	93	23	1,109	4,483	10	252	366	3,651	8,994	14,478
1996	114	28	984	4,007	10	224	306	1,968	9,787	13,852
1997	91	23	1,220	4,829	23	192	340	3,023	9,368	17,098
1998	92	23	1,333	5,332	24	206	357	3,522	8,787	21,119
1999	93	23	1,421	5,744	21	191	368	3,648	8,933	24,674
2000	92	23	1,583	6,333	24	208	398	4,229	11,145	30,757
2001	115	28	1,245	5,115	16	192	395	2,475	8,975	23,315
2002	95	23	1,382	5,709	26	174	519	4,943	8,821	19,895
2003	95	23	1,415	5,845	23	178	548	4,731	8,925	21,765
2004	95	23	1,579	6,524	22	181	538	4,780	10,785	30,318
2005	95	23	1,398	5,772	15	173	465	4,527	9,149	22,357
2006	108	27	1,351	5,404	13	155	458	3,172	10,297	26,518
2007	95	23	1,507	6,224	10	183	468	4,336	11,714	31,011
2008	95	23	1,470	6,072	11	184	543	4,970	9,378	24,523
2009	94	23	1,623	6,632	5	185	447	3,447	20,934	22,330
2010	94	23	1,700	6,947	13	203	457	3,558	21,035	23,706
2011	94	23	1,742	7,121	7	202	473	3,615	22,235	24,420
2012	94	23	1,834	7,496	13	194	442	3,638	22,531	28,852

Table C.2. Population and sample counts in the INSOLE, PUF, and the SCF

Year	Population: tax returns	INSOLE	PUF	$\frac{\text{PUF}}{\text{INSOLE}}$%	SCF	$\frac{\text{SCF}}{\text{PUF}}$%
1989	-	-	-	-	3,143	-
1990	-	-	-	-	-	-
1991	115,438,445	125,923	115,594	91.8	-	-
1992	115,068,332	103,516	93,262	90.1	3,906	4.2
1993	115,699,026	104,605	94,202	90.1	-	-
1994	116,878,243	108,861	96,384	88.5	-	-
1995	118,650,252	114,609	103,117	90.0	4,299	4.2
1996	120,917,968	126,420	112,186	88.7	-	-
1997	123,045,360	124,913	111,065	88.9	-	-
1998	125,037,636	164,340	122,625	74.6	4,305	3.5
1999	127,321,626	176,966	132,108	74.7	-	-
2000	129,644,980	196,149	145,663	74.3	-	-
2001	130,571,421	191,975	143,221	74.6	4,442	3.1
2002	130,540,073	175,566	131,307	74.8	-	-
2003	131,291,334	182,810	134,431	73.5	-	-
2004	133,189,982	200,778	150,047	74.7	4,519	3.0
2005	134,494,443	292,966	132,766	45.3	-	-
2006	138,485,335	321,006	145,898	45.5	-	-
2007	153,832,380	336,236	143,142	42.6	4,417	3.1
2008	142,580,866	328,630	139,651	42.5	-	-
2009	142,580,266	295,133	152,525	51.7	-	-
2010	143,170,763	308,946	159,789	51.7	6,482	4.1
2011	145,601,196	333,106	163,786	49.2	-	-
2012	145,021,073	338,475	172,411	50.9	-	-
2013	-	-	-	-	6,015	-
2014	-	-	-	-	-	-
2015	-	-	-	-	-	-
2016	-	-	-	-	6,248	-
2017	-	-	-	-	-	-
2018	-	-	-	-	-	-
2019	-	-	-	-	5,777	-

The decrease in ratio of the PUF and the INSOLE after 2004 is due to the INSOLE doubling the number of CWHS endings while the PUF number remained unchanged and to the reduction in the PUF subsampling rates for high income returns.

Table C.3. Number of observations in the upper tail of income distribution

Year	PUF	SCF	PUF	SCF	PUF	SCF	PUF	SCF	PUF	SCF	PUF	SCF
	10%		5%		1%		0.5%		0.1%		0.01%	
1988	-	908	-	703	-	416	-	323	-	166	-	24
1989	-	-	-	-	-	-	-	-	-	-	-	-
1990	-	-	-	-	-	-	-	-	-	-	-	-
1991	59,646	1,211	52,501	962	37,385	582	31,441	453	18,719	251	4,812	60
1992	42,541	-	37,850	-	28,915	-	24,834	-	15,925	-	3,721	-
1993	42,581	-	37,796	-	28,224	-	24,117	-	15,261	-	3,620	-
1994	43,994	1,240	38,985	972	28,849	566	24,406	446	15,453	218	3,555	45
1995	48,726	-	43,897	-	33,125	-	28,176	-	17,790	-	3,819	-
1996	56,049	-	50,597	-	36,143	-	30,285	-	18,692	-	3,892	-
1997	54,974	1,256	49,929	1,014	38,429	557	32,995	438	20,728	206	3,933	62
1998	63,094	-	57,824	-	44,480	-	37,941	-	23,690	-	4,106	-
1999	70,896	-	65,280	-	49,722	-	42,527	-	25,932	-	4,142	-
2000	81,870	1,289	75,463	1,035	57,374	599	49,108	447	28,066	214	4,244	65
2001	76,500	-	69,911	-	50,984	-	43,132	-	26,013	-	4,263	-
2002	65,514	-	59,602	-	44,458	-	37,784	-	23,707	-	4,161	-
2003	67,719	1,350	61,695	1,101	45,567	706	38,718	591	23,998	329	4,282	97
2004	79,600	-	72,794	-	53,759	-	45,930	-	27,207	-	4,360	-
2005	63,864	-	56,943	-	38,108	-	30,730	-	12,899	-	1,338	-
2006	73,814	1,389	65,209	1,115	41,674	670	33,335	544	12,731	247	1,390	100
2007	71,457	-	63,789	-	42,883	-	34,703	-	13,588	-	1,391	-
2008	67,010	-	59,795	-	39,378	-	31,835	-	13,598	-	1,439	-
2009	52,388	1,428	45,614	1,080	34,429	579	27,831	453	13,506	238	1,269	92
2010	57,660	-	50,476	-	37,878	-	30,761	-	14,086	-	1,323	-
2011	59,933	-	52,629	-	39,246	-	31,765	-	14,256	-	1,366	-
2012	68109	1,410	60,404	1,054	44,546	592	36,089	483	14,253	293	1,351	79
2013	-	-	-	-	-	-	-	-	-	-	-	-
2014	-	-	-	-	-	-	-	-	-	-	-	-
2015	-	1,453	-	1,099	-	626	-	487	-	231	-	48
2018	-	1,455	-	1,105	-	644	-	490	-	263	-	84

Without loss of generality, for the SCF, I report the number of observations when using the first implicate. Note that this number varies across the five imputations as well as across the 999 bootstrapped sample replicates. This number also varies in the PUF across the 999 bootstrapped sample replicates.

Table C.4. Number of observations in the upper tail of wealth distribution

Year	PUF	SCF	PUF	SCF	PUF	SCF	PUF	SCF	PUF	SCF	PUF	SCF
	10%		5%		1%		0.5%		0.1%		0.01%	
1989	-	980	-	770	-	417	-	329	-	169	-	52
1990	-	-	-	-	-	-	-	-	-	-	-	-
1991	60,378	-	51,464	-	36,064	-	30,241	-	17,496	-	4,069	-
1992	47,201	1,333	40,982	1,091	29,589	644	25,158	538	14,260	326	3,302	98
1993	47,473	-	40,886	-	29,152	-	24,482	-	14,181	-	3,181	-
1994	49,005	-	41,996	-	29,899	-	25,060	-	14,168	-	3,266	-
1995	53,537	1,461	46,452	1,167	33,512	664	28,106	495	14,853	258	2,817	92
1996	57,480	-	48,997	-	35,080	-	29,730	-	16,680	-	3,476	-
1997	59,233	-	52,127	-	38,951	-	33,151	-	18,197	-	3,598	-
1998	68,417	1,380	60,872	1,130	45,712	631	38,814	510	20,540	278	3,778	77
1999	76,450	-	68,280	-	51,147	-	42,897	-	22,213	-	3,847	-
2000	87,742	-	79,069	-	59,336	-	49,588	-	24,327	-	4,003	-
2001	80,057	1,406	70,556	1,156	51,657	659	43,480	535	21,788	324	3,912	88
2002	72,153	-	63,517	-	46,616	-	39,232	-	20,388	-	3,793	-
2003	74,145	-	65,154	-	47,323	-	39,522	-	20,201	-	3,899	-
2004	87,195	1,434	77,066	1,164	56,587	708	47,161	581	23,677	352	4,183	120
2005	69,094	-	58,873	-	38,155	-	29,051	-	10,872	-	1,332	-
2006	75,026	-	63,584	-	40,845	-	31,192	-	11,321	-	1,396	-
2007	76,032	1,504	65,627	1,185	43,387	684	33,610	569	12,116	349	1,390	98
2008	72,343	-	62,001	-	40,817	-	31,673	-	11,955	-	1,433	-
2009	57,682	-	49,601	-	33,801	-	26,180	-	10,682	-	1,269	-
2010	62,863	1,481	54,293	1,124	37,051	655	28,652	526	11,063	315	1,313	98
2011	65,170	-	56,204	-	37,794	-	29,060	-	11,020	-	1,347	-
2012	73,346	-	63,754	-	43,040	-	32,548	-	11,595	-	1,309	-
2013	-	1,499	-	1,132	-	651	-	538	-	312	-	93
2014	-	-	-	-	-	-	-	-	-	-	-	-
2015	-	-	-	-	-	-	-	-	-	-	-	-
2016	-	1,556	-	1,191	-	680	-	563	-	325	-	96
2017	-	-	-	-	-	-	-	-	-	-	-	-
2018	-	-	-	-	-	-	-	-	-	-	-	-
2019	-	1,561	-	1,223	-	700	-	564	-	334	-	86

Without loss of generality, for the SCF, I report the number of observations when using the first implicate. Note that this number varies across the five imputations as well as across the 999 bootstrapped sample replicates. Also, in the PUF, this number varies across bootstrapped sample replicates.

Table C.5. Regression results of weighted least squares for six top-decile income shares between 1991 and 2012

Regressor	**Income share of the top k%**					
	10%	**5%**	**1%**	**0.5%**	**0.1%**	**0.01%**
	SCF					
Intercept	0.377***	0.268***	0.122***	0.088***	0.040***	0.011***
	(57.566)	(33.073)	(14.856)	(12.119)	(8.115)	(5.450)
Linear time trend	0.078***	0.080***	0.061***	0.048***	0.025**	0.009**
	(7.536)	(6.162)	(4.497)	(3.861)	(3.005)	(2.615)
R^2	0.904	0.864	0.771	0.713	0.601	0.533
	PUF					
Intercept	0.386***	0.271***	0.127***	0.093***	0.045***	0.016***
	(110.837)	(71.995)	(44.011)	(38.843)	(30.516)	(22.649)
Linear time trend	0.060***	0.057***	0.042***	0.037***	0.025***	0.012***
	(11.258)	(9.632)	(7.976)	(7.884)	(8.001)	(7.938)
R^2	0.864	0.823	0.761	0.757	0.762	0.759

This table summarizes estimation results from 12 weighted linear regressions of the top-decile income shares on a constant and linear time trend. The time variable is defined as 0 for the first observation (1991) and as 1 for the last observation (2012). Consequently, the intercept is the estimate of income share at time zero and the slope is the estimate of the difference in income shares between the first and last observation. I report estimated coefficients and t-statistics in parentheses. "***" denotes statistical significance at the 99 percent significance level, "**" at the 95 percent significance level, and "*" at the 90 percent significance level.

Table C.6. Regression results of weighted least squares for the wealth shares within the top 10 percent

Regressor	Wealth share of the top k%					
	10%	**5%**	**1%**	**0.5%**	**0.1%**	**0.01%**
	SCF					
Intercept	0.665***	0.546***	0.318***	0.241***	0.114***	0.037***
	(97.556)	(103.860)	(30.716)	(19.407)	(14.223)	(12.592)
Slope	0.067***	0.062***	0.021	0.006	0.004	-0.001
	(6.345)	(7.728)	(1.312)	(0.308)	(0.341)	(-0.123)
R^2	0.890	0.923	0.256	0.019	0.023	0.003
	PUF: Homogeneous					
Intercept	0.648***	0.498***	0.261***	0.195***	0.101***	0.039***
	(168.748)	(122.827)	(69.998)	(57.326)	(37.499)	(19.015)
Slope	0.062***	0.071***	0.075***	0.070***	0.053***	0.030***
	(9.708)	(10.076)	(10.645)	(10.401)	(9.089)	(6.347)
R^2	0.847	0.857	0.870	0.864	0.829	0.703
	PUF: Heterogeneous 10YT					
Intercept	0.647***	0.488***	0.243***	0.172***	0.069***	0.023***
	(112.073)	(71.806)	(32.204)	(29.304)	(22.351)	(14.451)
Slope	0.046***	0.057***	0.065***	0.064***	0.064***	0.034***
	(5.402)	(5.659)	(5.746)	(6.829)	(10.653)	(8.691)
R^2	0.632	0.653	0.660	0.733	0.870	0.816
	PUF: Heterogeneous Moody's					
Intercept	0.653***	0.497***	0.259***	0.187***	0.072***	0.022***
	(99.320)	(62.154)	(28.731)	(28.983)	(21.484)	(13.044)
Slope	0.043***	0.051***	0.055***	0.054***	0.063***	0.035***
	(4.525)	(4.447)	(4.223)	(5.425)	(10.139)	(8.240)
R^2	0.546	0.538	0.512	0.634	0.858	0.800

This table summarizes estimation results from 24 weighted linear regressions of the top-decile wealth shares on a constant and linear time trend. The time variable is defined as 0 for the first observation (1992) and as 1 for the last observation (2010). Consequently, the intercept is the estimate of income share at time zero and the slope is the estimate of the difference in income shares between the first and last observation. I report estimated coefficients and t-statistics in parentheses. "***" denotes statistical significance at the 99 percent significance level, "**" at the 95 percent significance level, and "*" at the 90 percent significance level.

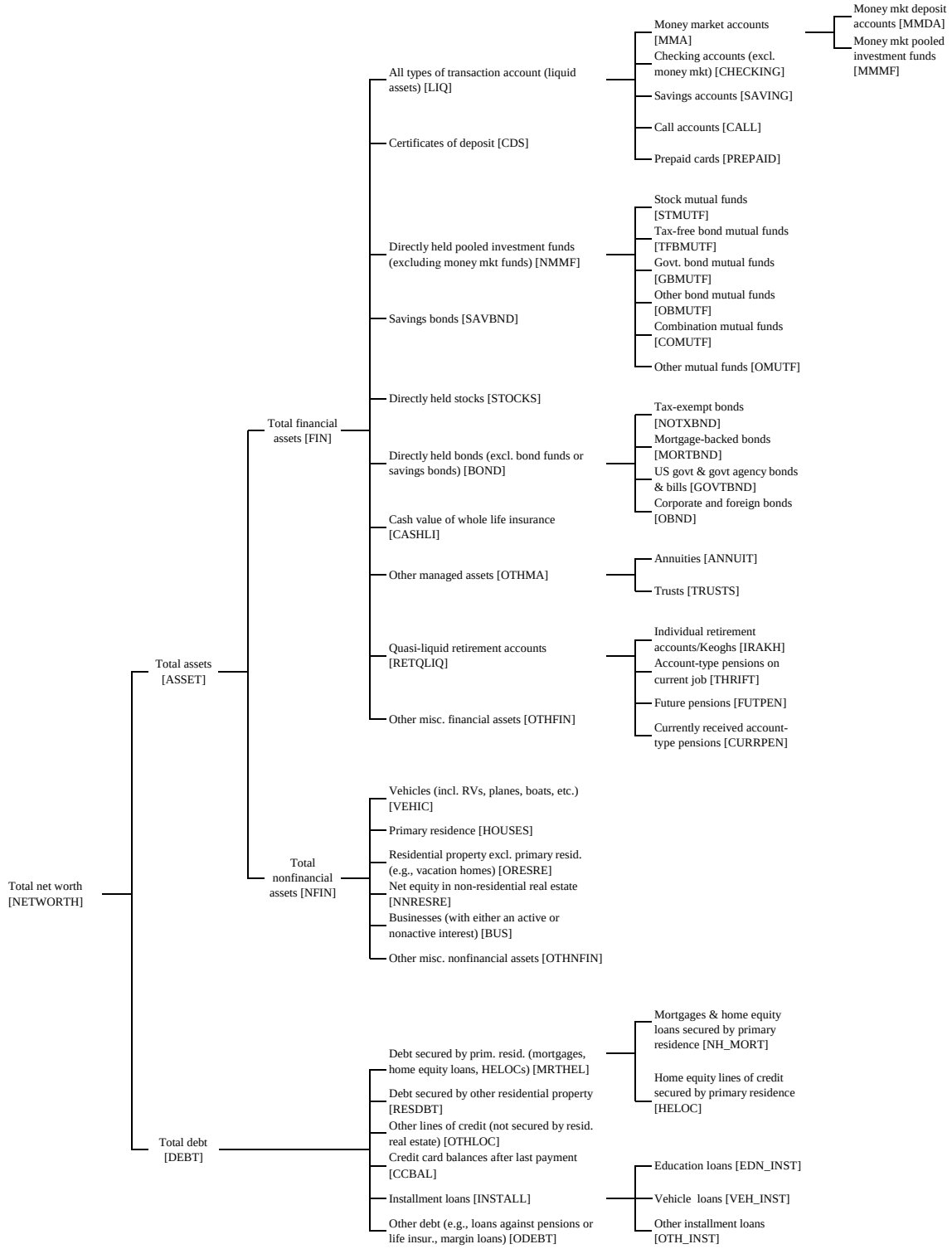

FIGURE C.1. Definition of wealth in the SCF. Names in brackets refer to variables in the SCF Bulletin extract data. Source: The SCF Documentation provided by the Board https://www.federalreserve.gov/econres/scfindex.htm (accessed on October 13, 2019).