



## **What Do Survey Data Tell Us about U.S. Businesses?**

Anmol Bhandari  
University of Minnesota

Serdar Birinci  
University of Minnesota

Ellen R. McGrattan  
University of Minnesota and Federal Reserve Bank of Minneapolis

Kurt See  
University of Minnesota

**Staff Report 568**  
**Revised January 2019**

DOI: <https://doi.org/10.21034/sr.568>

Keywords: Survey data; Intangibles; Business taxes and valuation

JEL classification: C83, E22, H25

The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

Federal Reserve Bank of Minneapolis  
Research Department Staff Report 568  
Revised January 2019

## What Do Survey Data Tell Us about U.S. Businesses?\*

Anmol Bhandari  
University of Minnesota

Serdar Birinci  
University of Minnesota

Ellen R. McGrattan  
University of Minnesota and Federal Reserve Bank of Minneapolis

Kurt See  
University of Minnesota

---

### ABSTRACT

This paper examines the reliability of widely used surveys on U.S. businesses. We compare survey responses of business owners with administrative data and document large inconsistencies in business incomes, receipts, and the number of owners. We document problems due to nonrepresentative samples and measurement errors. Nonrepresentativeness is reflected in undersampling of owners with low incomes. Measurement errors arise because respondents do not refer to relevant documents and possibly because of framing issues. We discuss implications for statistics of interest, such as business valuations and returns. We conclude that predictions based on current survey data should be treated with caution.

Keywords: Survey data, intangibles, business taxes and valuation  
JEL classification: C83, E22, H25

---

\*An online appendix is available at [www.minneapolisfed.org/research/sr/sr578.pdf](http://www.minneapolisfed.org/research/sr/sr578.pdf). Bhandari and See acknowledge support from the Heller-Hurwicz Economic Institute, and McGrattan acknowledges support from the NSF. For helpful comments, we thank seminar and conference participants at Colorado, Minnesota, Yale, Autonoma, ASU, NYU, Columbia, UCSB, SITE, CMU, Midwest Macro, LAMES, Michigan, and WU in St Louis. For excellent editorial assistance, we thank Joan Gieseke. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

# 1 Introduction

Representative surveys of households and firms have become an important source of data on business owners and their activities, and are now used extensively in studies of wealth inequality and entrepreneurial choice. This paper examines the reliability of these data for research on U.S. businesses, including pass-through entities and subchapter C corporations.<sup>1</sup> Pass-through businesses account for roughly half of business net income in the United States and have been a focus of recent tax reforms and debates about income inequality.<sup>2</sup> Subchapter C corporations account for the remaining half and include all publicly traded firms. We document issues arising from nonrepresentative samples and measurement errors in survey data and discuss the implications of the errors for economic research.

We start by examining the reliability of measures of business incomes, receipts, and valuations in the Federal Reserve’s Survey of Consumer Finances (SCF), which is a publicly available and widely used triennial cross-sectional survey of U.S. households. Households with actively managed businesses are asked to report business receipts and net income from specific lines on their tax forms.<sup>3</sup> This aspect of the survey design makes it easy for us to compare the household responses with administrative data from the IRS *Statistics of Income* (SOI).<sup>4</sup> Averaging across survey years, we find that the SCF overstates pass-through business income per tax return by 400 percent and business re-

---

<sup>1</sup>For tax purposes, pass-through entities classify themselves as sole proprietorships, S corporations, or partnerships. They are called “pass-through” because the income earned by such businesses is taxed under the owners’ individual income tax. In contrast, C corporations pay corporate taxes on income earned, and individual shareholders pay dividend or capital gains taxes when profits are distributed.

<sup>2</sup>Smith et al. (2017) use tax audit data to conclude that rising business income accounts for all of the growth in the top 1 percent income share since 2000. Furthermore, the majority of rising top business income resulted from rising income of pass-through businesses.

<sup>3</sup>Sole proprietors are asked to report business gross receipts and net income from Form 1040 Schedule C (lines 1 and 31), shareholders of partnerships from Form 1065 (lines 1c and 22), shareholders of S corporations from Form 1120S (lines 1c and 21), and shareholders of C corporations from Form 1120 (lines 1c and 30).

<sup>4</sup>The IRS compiles the SOI based on stratified probability samples of income tax returns and other forms. We use information from the SOI Corporation Income Tax Returns, Individual Income Tax Returns, and Partnership Returns that are available in the historical data tables from [www.irs.gov/statistics](http://www.irs.gov/statistics).

ceipts per return by 169 percent. For C corporations, net income and receipts are on average understated in the SCF by 26 and 21 percent, respectively, but the SCF does not include publicly traded corporations, whereas the IRS does. Since publicly traded corporations have much higher receipts and net incomes per tax return than private corporations, including them would result in a significant overstatement for all business entities.

The overstatement of incomes and receipts in the SCF varies in the cross section and year by year, making it difficult, if not impossible, to systematically correct for the errors. To demonstrate this, we provide evidence of both sampling and measurement errors. Sampling errors arise from an understatement or overstatement of the reported number of tax returns in the SCF relative to the IRS. We find that the SCF significantly understates the number of sole proprietorships, S corporations, and C corporations and significantly overstates the number of partnerships, with the degree of under- or overstatement varying across the income distribution. Looking at data in the cross section, our findings suggest a significant underrepresentation of low-income businesses driving the overstatement of business incomes. This may be attributable to measurement error resulting from how the questions are framed. For example, there are many IRS businesses with net losses but few in the SCF, possibly because the respondents answered that they had no net income rather than a negative net income. Another measurement issue we document is that the frequency of referencing supporting documents is strikingly low. For example, if we condition on all business owners in the SCF, we find that 75 percent never referenced any tax document.

The SCF survey has also been used extensively to study the level and dispersion of business wealth. Households with actively managed businesses are asked to assess the value of their business, net of all loans, if they were to sell it. Since there are no measures of total valuations for ongoing businesses other than publicly traded C corporations, we construct net income-to-value ratios in the SCF and compare them to available income yields from brokered private business sales recorded by Pratt's Stats and publicly traded companies, both small and large, recorded by the Center for Research in Security Prices (CRSP) (merged with Compustat). For virtually all subsamples and all years, the SCF

income yields are significantly higher than comparable measures from the other datasets. The overstatement in yields is even greater than for incomes, which is suggestive of an understatement in business valuations. For example, the SCF average value-weighted income yield is 19 percent, much higher than the Pratt's Stats estimates of 2 percent or the CRSP estimates of 7 percent for all businesses, and  $-9$  percent for those in the bottom quintile when firms are ranked by total assets. We also find that the SCF distributions are more right-skewed than those based on Pratt's Stats or CRSP-Compustat data.

For unincorporated businesses, we can compare the SCF estimates of business incomes per owner and, if available, income yields to those of three other widely used surveys: the Current Population Survey (CPS), the Panel Study of Income Dynamics (PSID), and the Survey of Income and Program Participation (SIPP).<sup>5</sup> We show that all four surveys overstate incomes per owner relative to the IRS data, but by differing amounts. Averaging across survey years, business income per owner for unincorporated businesses is overestimated by 586 percent in the SCF, 179 percent in the CPS, 185 percent in the PSID, and 34 percent in the SIPP. Average value-weighted income yields calculated for the PSID and SIPP are high relative to Pratt's Stats and CRSP data in all cases but are not very different from those in the SCF. The main differences in yields across surveys are found when we compare the distributional statistics.

An overstatement of business incomes in the survey data relative to the IRS could be the result of misreporting in the IRS or miscategorization of incomes by survey respondents. To check misreporting, we use auxiliary tax audit data to adjust the IRS data but still find a significant mismatch with the survey data. For example, for pass-through businesses, the SCF overstates the average income per return by 178 percent relative to the tax audit data. To check miscategorizations, we use a broader definition of business income. For example, business owners might confuse business incomes on Schedule C,

---

<sup>5</sup>We also compare the SCF results to the results of Gurley-Calvez et al. (2016), who match responses of new businesses in the Kauffman Firm Survey (KFS) with IRS tax data and find that these firms understate business incomes. For some other surveys of businesses, such as the Panel Study of Entrepreneurial Dynamics (PSED), we find the response rates of business-related questions to be too low to be reliable.

E, and F, overstating one category and understating another. When we combine these categories into a broader concept of business income, we still find incomes to be significantly overstated. Respondents are not miscategorizing the incomes but rather are often overstating all categories of business income. Other adjustments, such as correcting for within-survey inconsistencies regarding business ownership and income and correcting for the fact that the survey only includes individual owners, do not alleviate the measurement issues.

The inconsistencies across surveys and the conceptual measurement issues that we highlight suggest that the “facts” about business income and wealth based on current survey data should be treated with caution. Measurement problems related to business incomes are surmountable given that respondents are asked about specific lines on the tax forms. Measurement problems related to business valuations and returns may be insurmountable without data on actual business sales transactions. First, it is difficult for owners to estimate business valuations when one considers that businesses are heavily invested in intangible assets.<sup>6</sup> Second, survey business owners answer questions separately about income and valuations. For example, if the net incomes derive from both capital and labor inputs, while the business valuations are based on fixed assets owned by the business, then the estimated income yields from surveys may not be comparable across owners who interpret the question differently. Interpreting survey-based measures of business returns or valuations requires a consistent framework for true returns, stocks, and valuations. Given current measurement issues, such interpretations may not be possible.

This paper is organized as follows. Section 2 discusses the related literature and the implications of our findings for theoretical and applied research on businesses. In Section 3, we document that business incomes and receipts measured from the SCF survey data are largely inconsistent, and we discuss problems regarding nonrepresentativeness and measurement errors. Section 4 studies business valuations and rates of return. Section 5 presents robustness

---

<sup>6</sup>McGrattan and Prescott (2010a,b) and Bhandari and McGrattan (2018) both find estimates of the value for intangible assets to be close to estimates of tangible assets used by businesses. Intangible assets come in the form of research and development, software, advertising, brands, and investments in customer lists, goodwill, and other forms of investments in building organizations.

checks. Section 6 compares the SCF results with other widely used surveys. Finally, Section 7 concludes.

## 2 Related Literature

Our findings have implications for three active areas of economic research. The first area is the empirical literature that documents levels and trends in the dispersion of income and wealth and emphasizes the role of entrepreneurs in wealth accumulation. The second area, which is motivated by and builds upon the first, is the theoretical literature developing models of entrepreneurial choice, which are specifically designed to fit the “stylized facts” of the empirical literature. The third area includes quantitative policy analyses that use the empirical findings and theoretical developments of the first two literatures as their laboratory for the study of counterfactual policies. Our findings cast doubt on the facts that have been uncovered in the empirical literature—specifically documenting that survey data are unreliable for business-related statistics—and thus raise issues concerning the theoretical developments and policy analyses that have been designed around them.

A large and burgeoning empirical literature documents trends in income and wealth and has been particularly focused on increased dispersion over time. Greater dispersion is attributed to top earners, and therefore researchers work primarily with survey data from the SCF or administrative tax data from the IRS. For example, Kuhn and Rios-Rull (2016) provide a nearly exhaustive summary of distributional facts about U.S. earnings, income, and wealth based on the SCF. A starting point for several papers in this literature is the observation that, for broad income categories, aggregated SCF responses match up well to the aggregated IRS data. In Figure 1, we replicate the time series plot for adjusted gross income (AGI) from the SCF and plot it against the corresponding data from the IRS. We see that the SCF tracks the level and cyclical trends for AGI in the IRS.

Our focus is on measures that relate to business activity. Of particular relevance are the findings in Kuhn and Rios-Rull (2016) that business income is one of the main contributors to income inequality and that business equity

is one of the main contributors to wealth inequality, which they document for the history of the SCF surveys between 1989 and 2013. Bricker et al. (2016) also use the SCF to document the rise of the top share of wealth holdings over time and find that the share of wealth attributable to the top 1 percent rose from 30 percent in 1992 to 36 percent in 2013. Our paper exploits the fact that SCF answers can be compared to administrative data from the IRS and finds that respondents are not reliably or consistently answering questions about their business income or business equity, and therefore we cannot trust the SCF distributions.

Given issues with measuring business incomes, researchers might be tempted to combine all nonwage income into a residual “capital income” category, since SCF aggregates match up well with aggregated IRS data. Here, we argue that this capital income measure would not be appropriate for either research on U.S. businesses or research on U.S. capital. For research on businesses, the residual income measure would be inappropriate because significant non-business income is included with interest payments, capital gains, pensions and annuities, alimony, trusts, and government transfers. Furthermore, as we noted earlier, there is evidence that owners are not miscategorizing income categories, and therefore using broader categories of income would do little to ameliorate the measurement issues. For research on capital, the nonwage income in AGI would be inappropriate because a significant fraction of capital income is untaxed and the corresponding assets are held by fiduciaries. Furthermore, as we show later, the majority of respondents do not reference financial documents, making it nearly impossible to have reliable estimates of their total capital income or wealth.

Saez and Zucman (2016) document trends in wealth dispersion by capitalizing incomes from administrative tax data. They compare their results to the SCF and find similar levels and trends for wealth in the top 10 percent of the distribution but differences for the top 1 percent.<sup>7</sup> The Saez and Zucman (2016) capitalization method is inappropriate for estimating wealth in business for several reasons. First, there is no way to validate the procedure

---

<sup>7</sup>They also compare results to estate taxes and foundation records, but these data are not informative about most businesses in the United States.



except by comparing to survey data, which we find are unreliable. Second, the U.S. flow of funds aggregates used to compute capitalization factors include ad hoc imputations made by the Federal Reserve for private businesses.<sup>8</sup> Third, the assumption that returns are the same for everyone and constant over time is hard to reconcile with the fact that there is significant entry into and exit out of business (see Bhandari and McGrattan (2018)).

Because of the problems with data from the SCF and the capitalized IRS incomes, the main message of our findings for the theoretical literature is a cautionary one, namely, that these data tell us little about business valuations or returns, and therefore theorists should not insist on models that replicate “stylized facts,” which are not actually facts. The most popular stylized facts are that entrepreneurs, as a group, own a substantial share of household wealth and income, with shares increasing throughout the distribution, and that entrepreneurs have high savings rates relative to the population, implying much more dispersion in wealth than in income (see De Nardi, Doctor, and Karen (2007) and Gentry and Hubbard (2004)). These findings have led researchers to model entrepreneurs as overcoming significant market frictions to run highly risky businesses with the expectation of earning high returns and amassing significant wealth (see, for example, Quadrini (2000), Cagetti and De Nardi (2006), and Buera (2009)). Furthermore, the theoretical frameworks parameterized to match the survey data have been used as a laboratory for policy work, especially when considering tax policy reform (see, for example, Meh (2005), Kitao (2008), Bohacek and Zubricky (2012), and Scheuer (2013)). Our results cast doubt on the survey data underlying the models of financial frictions and, hence, the subsequent policy recommendations.

Our paper is also related to a second strand of the empirical literature,

---

<sup>8</sup>For example, when constructing capitalization factors, Saez and Zucman (2016) use aggregate flow of funds wealth measures for closely held corporations (both subchapter C and S) and unincorporated businesses. These businesses are not publicly traded and thus have no market valuations. The Federal Reserve imputes market values for closely held corporations by taking a ratio of market value to revenues for publicly traded companies and then applying that ratio to private businesses with similar industry, employment, and revenue profiles—after arbitrarily adjusting the estimate downward by 25 percent to reflect the lack of liquidity of closely held shares. Valuations for unincorporated businesses are based on balance sheet data reported to the IRS, which are historical-cost accounting measures, not market valuations.

which reaches different conclusions about the payoffs to entrepreneurial activity.<sup>9</sup> Hamilton (2000) uses survey data from the 1984 SIPP and finds that self-employed individuals—who could be running an incorporated or unincorporated business—have lower median earnings than similar individuals in paid employment. Moskowitz and Vissing-Jorgensen (2002) extend his analysis and work with SCF data, allowing for a more comprehensive treatment of equity returns and including adjustments for firm entry and exit. They find that returns to private businesses are no higher than returns to publicly traded firms and thus puzzlingly low given the risks entrepreneurs face.<sup>10</sup> Using PSED data, Hurst and Pugsley (2011) report that more than 50 percent of new business owners cite flexible hours and being one’s own boss as the primary reason for starting their own business. These findings have led researchers to conclude that the nonpecuniary benefits of self-employment play an important role in occupational choice.

We document that survey data *overstate* business incomes, and the overstatement leads to income yields for private businesses that are significantly higher than those for publicly traded companies. In other words, we find that the private returns computed with survey data are puzzlingly high, not puzzlingly low. When comparing our results to Moskowitz and Vissing-Jorgensen (2002), we find that the main difference is the concept of return: they add an imputation for capital gains that drive their results. Regrettably, neither the income yield nor the capital gain imputation are reliable estimates, leaving us with little to say about whether private returns are low or high relative to public returns. This finding is relevant for policy discussions related to business taxation and subsidization. For example, Hurst and Pugsley (2017) followed up on the work of Hamilton (2000) and Moskowitz and Vissing-Jorgensen (2002) by incorporating nonpecuniary benefits in a model of entrepreneurship

---

<sup>9</sup>Later, we document inconsistencies across surveys and across time that could lead to different empirical insights.

<sup>10</sup>Moskowitz and Vissing-Jorgensen (2002) used samples over the period 1989–1998. Karashova (2014) extended their analysis to 2010 and documented that for the longer sample, the private equity premium is about 6 percent as compared to about 0 percent in the shorter sample. Since the difference in means is swamped by the variability of the CRSP index returns, which has a standard deviation of 20 percent over the period 1988–2015, we would argue that their estimated private returns are not significantly different.

and then analyzed the impact of small business subsidies. Our results cast doubt on SIPP and SCF survey data and hence on policy recommendations that arise from frameworks that incorporate nonpecuniary benefits of business entrepreneurs.

### 3 Business Incomes

In this section, we compare SCF pretax business incomes that correspond to specific lines on U.S. tax forms with incomes reported to the IRS. We first document that the SCF responses—both in the aggregate and across the distribution—are statistically different and argue that the magnitudes of the differences are economically important. We then explore reasons for the mismatch.<sup>11</sup>

#### 3.1 Aggregates

We start with aggregated business incomes and show that, while the SCF does well in matching the IRS total AGI (as shown in Figure 1), there are large discrepancies between the survey and tax data for businesses. For pass-through businesses, business income per return is significantly and consistently overstated in the SCF relative to the IRS. For C corporations, the average per-return business income in the SCF is not very different from the IRS but should be much smaller given that the survey excludes publicly traded companies.

To demonstrate that there is a significant discrepancy between SCF and IRS data, we start by defining *business income* as gross receipts from sales minus expenses (including depreciation) incurred in running the business. Information on business incomes is obtained from the respective business tax forms: Form 1040, Schedule C (line 31) for sole proprietors, Form 1065 (line 22) for partnerships, Form 1120S (line 21) for S corporations, and Form 1120 (line 30) for C corporations. In each survey year, we use the SCF sampling weights and ownership information to compute the aggregate business income

---

<sup>11</sup>In our online appendix cited henceforth as Bhandari, et al. (2019), we provide a comprehensive collection of statistics for all survey years and subsamples.

and the aggregate number of business tax returns by legal form of the business.

Figure 2 plots aggregated business income divided by the number of business tax returns using the SCF and the data actually reported to the IRS for tax years between 1988 and 2015.<sup>12</sup> All data are reported in thousands of current dollars. The shaded region in Figure 2 shows the 90 percent confidence interval of the average per-return business incomes.<sup>13</sup> Panel A of Figure 2 shows the results for sole proprietorships. For this group, average incomes reported in the IRS are about \$8,000 per return in 1988 and rise gradually to \$12,000 per return by 2015. Average incomes per return reported in the SCF are significantly higher, rising from \$32,000 in 1988 to \$40,000 by 2015, and display large year-to-year variation. If we construct a percentage error (that is,  $100(\text{SCF}-\text{IRS})/\text{IRS}$ ) in each year, we find the average error is 289 percent and ranges from 158 percent to 384 percent across all years. We see a similar result in Panel B, which shows the data for S corporations. The SCF S-corporation incomes per return are significantly higher than the IRS analogues for the entire sample period. The errors in this case average 273 percent and range from 142 percent to 387 percent.

In the case of partnerships, a sampling issue implies that the discrepancy between the SCF and IRS income could be even larger than that shown in Panel C of Figure 2. The SCF only surveys owners of partnerships who are individuals, whereas the IRS includes information on partnerships owned by individuals and other legal entities such as corporations.<sup>14</sup> The exclusion of corporate partners in the SCF should lead to an understatement of aggregate

---

<sup>12</sup>In the SCF, we assume that a business owner who owns multiple sole proprietorships files one return. This assumption is made to be consistent with IRS statistics that state: “For purposes of the statistics, if a proprietor owned more than one business, the statistics for each business were combined with those of the proprietor’s dominant business and included in the industrial group for that business activity” (see Dungan (2017, p. 2)). For partnerships, S corporations, and C corporations, we calculate the number of returns taking into account the ownership share of the family from each reported business.

<sup>13</sup>Standard errors are computed using a bootstrap method. For each survey year, the SCF provides a set of 1,000 replicate weights. We use these weights to compute 1,000 values for the relevant statistic, for example, business income per return, and then we compute the confidence intervals using the 5th and 95th percentiles of these 1,000 draws.

<sup>14</sup>For the year 2011, Cooper et al. (2016) estimate that 32 percent of total business income from partnerships is generated by individual partners, who account for 73 percent of all partners.

business income but, in principle, should not affect the business income per return.<sup>15</sup> However, we find that both aggregate income and per-return income are overstated for partnerships. Aggregate business income from partnerships in the SCF is *higher* than the IRS by 305 percent on average, with the range of errors between 52 percent and 889 percent across survey years. Per-return income is also overstated by 300 percent on average, with errors between 31 percent and 837 percent, as shown in Panel C of Figure 2.

If we include all pass-through businesses in one category (using data in Panels A through C of Figure 2), we find that the SCF error is 400 percent on average and ranges from 230 percent to 568 percent for business income per return. Contrast this with business incomes per return for C corporations, shown in Panel D of Figure 2. For these businesses, we find that in most years, the average SCF business income per return is actually understated by about 26 percent as compared to the IRS data. However, the IRS data include publicly traded corporations, whereas the SCF data do not. Publicly traded C corporations are typically much larger than their private counterparts. If we could include the incomes from these publicly traded corporations with the SCF estimates, we would find that the SCF total incomes would be significantly higher than the IRS estimates, as is the case for the pass-through businesses. Despite these measurement issues, we compare business incomes of C corporations in the SCF with the IRS and interpret the results with the understanding that the SCF will underrepresent large businesses.

While incomes per return are overstated in the SCF relative to the IRS, the number of returns filed by businesses are significantly understated for sole proprietors and corporations in all years. Figure 3 plots the number of business returns filed as reported by the IRS and the SCF, over time and by legal entity, with shading marking the 90 percent confidence interval. In the case of sole proprietors and S corporations shown in Panels A and B, the understatement has worsened over time as the number of IRS filings has grown and the number reported in the SCF has not. In Panel C, we see that the number of partnership

---

<sup>15</sup>Here, we are operating under the assumption that the SCF is representative and partnerships with all corporate partners (which are entirely missed by the SCF) either are small in terms of their share of aggregate business income or else are not systematically different from the rest of the partnerships.

returns in the SCF is undercounted in only a few years and not by as much as in the case of the other business types. However, as mentioned before, the SCF data only include partners who are individuals, implying that the SCF significantly *overstates* the number of returns for partnerships owned by individuals. The number of C-corporate returns is shown in Panel D of Figure 3. In this case, part of the difference between the IRS and SCF results is the exclusion of publicly traded companies in the SCF, but these businesses only account for about 5,000 out of the roughly 1.6 million C corporations and therefore cannot account for the large understatement of returns shown in the figure.

One possible reason for understated returns is that the SCF data may not include owners that earn very little business income (for example, part-time Uber drivers or AER referees), while the IRS includes all business tax filers. If this were indeed the case, then the aggregate business incomes—found by multiplying values in Figure 3 by values in Figure 2—would be similar for the IRS and SCF. We find that this is not the case: aggregate business incomes are significantly overstated in the SCF. For example, in the case of pass-through businesses, we find average overstatements of 34, 137, and 305 percent for sole proprietors, S corporations, and partnerships, respectively, with a large range in the errors over time. (See Bhandari et al. (2019) for details.) The large overstatement of aggregate incomes, especially in S corporations and partnerships, is clearly inconsistent with the hypothesis that nonactive business owners explain the differences between the SCF and IRS results.

Finally, we extend the analysis to business receipts and again find large and variable errors in the SCF responses when compared to the IRS counterparts. (Full details are provided in Bhandari et al. (2019).) For example, in the case of pass-through businesses, we find that the average error in business receipts per return over the period 1988–2015 is 169 percent, with errors over the period in the range of 89 percent to 367 percent. Thus, our main finding is an overstatement of aggregated business incomes and receipts in the SCF across all legal forms, with large variation in the discrepancies across survey years.

## 3.2 Distributions

Next, we show that the overstatement of business income per return documented in the previous section varies in the cross section and year by year, making it impossible to systematically correct the SCF responses. The discrepancies between the SCF and IRS statistics are so large and variable as to make the cross-sectional moments based on the survey data unusable for applied work.

To demonstrate this, we compute percentage errors by grouping businesses in two different ways. First, we group those with positive business income (profits) separately from those with negative business income (losses). For pass-through businesses, the SCF overstates business income per return for those that have profits by 277 percent on average, with the year by year errors in the range of 151 percent to 446 percent. The SCF understates the business income per return for those that have losses by about 82 percent, with the errors in the range of 66 percent to 94 percent. Both the overstatement of profits and the understatement of losses affect the errors in cross-sectional statistics in quantitatively important ways. We demonstrate this in Table 1 for pass-through businesses by decomposing the percentage errors in the aggregate business income into the fraction that arises due to overstatement of profits and the fraction that arises due to understatement of losses. For example, in tax year 1988, the overstatement of profits and understatement of losses each account for about 50 percent of the aggregate error. In subsequent years, these fractions vary but are nontrivial in both the overstatement of profits and understatement of losses. (See Bhandari et al. (2019) for results across all legal forms and survey years.)

A second method of grouping businesses is by ranking them according to total income of the owners. Here, we do this in the case of sole proprietorships since we have comparable IRS data in all SCF survey years. (See Bhandari et al. (2019) for a more limited analysis of S corporations.) Specifically, we rank sole proprietors in the SCF by their AGI and then assign them to income brackets using the same bins as the IRS. In Figure 4, we plot the fractions of business income for owners with below-median AGI and for those with AGIs in the top 1st percentile. For most years, the SCF income shares for these two

groups are understated and display large year by year variation. For example, the share for those with below-median AGI is nearly doubled or halved from one survey to the next. Since the fractions sum to 100 percent across all AGI groups, the SCF must necessarily overstate incomes for some bins. We find the largest overstatement of shares for those with AGIs between the 50th and 75th percentile.

In Figure 5, we see that the overstatement of business income per return in the SCF data also varies a lot across years and across AGI bins, with no systematic pattern. The panels of this figure can be compared to the aggregate data for sole proprietorships in Panel A of Figure 2. In contrast, the incomes per return in the IRS data show little variation over time and vary similarly across AGI bins.

### 3.3 Nonrepresentativeness and Measurement Error

We now investigate the reasons behind the discrepancies in business incomes between the SCF and IRS and provide evidence for two types of errors in the SCF: nonrepresentativeness of business owners and misreporting of business income by business owners. The evidence of both types of errors again demonstrates that there is no easy correction for the survey data.

To fix ideas, we decompose the difference between a survey aggregate,  $X^S$ , and an IRS aggregate,  $X^I$ , for some measure into three terms as follows:

$$X^S - X^I = \left( \frac{\sum_i (\omega_i^S - \omega_i^I)}{\sum_i \omega_i^I} \right) X^I + \sum_i \omega_i^I X_i^I \left( \frac{\omega_i^S}{\omega_i^I} - \frac{\sum_i \omega_i^S}{\sum_i \omega_i^I} \right) + \sum_i \omega_i^S (X_i^S - X_i^I), \quad (1)$$

where sums are taken over household types indexed by  $i$ . The average  $X$  for type  $i$  is denoted by  $X_i^s$ , and the population weight for type  $i$  is denoted by  $\omega_i^s$ , with  $s \in \{S, I\}$ . The first two terms capture differences in weights,  $\omega_i^S \neq \omega_i^I$ , which we refer to as “sampling errors,” and the last term captures differences in averages,  $X_i^S \neq X_i^I$ , which we refer to as “measurement errors.”

With linked survey-IRS data, one can fully decompose the difference on the left-hand side of (1) into the sampling and measurement error components. For the SCF, we do not have access to such linked data but can provide



evidence that is strongly indicative that both sampling and measurement errors are nontrivial. We start with evidence on the sampling errors. Earlier, we provided evidence based on the total number of business returns that the first term in (1) is significantly different from zero. Recall that we found a significant underrepresentation of sole proprietors and corporations and an overrepresentation of partnerships, indicating severe sampling issues.

To shed light on the second term in (1), we analyze how the returns are distributed in the cross section. We define the groups of businesses as we did previously in Section 3.2, first on the basis of whether they earned profits or losses and second by ranking them according to their owners' AGI. We then compare ratios of population weights,  $\omega_i^S/\omega_i^I$ , across types and across time. For example, if we compare these ratios for pass-through businesses that have profits with those that have losses, we find significant differences in most years, with the highest difference being 35 percentage points. Similarly, if we compare ratios for sole proprietors in different AGI bins, we find significant differences across AGI bins. Figure 6 shows this in the case of sole proprietors with AGIs per return below and above the median. For businesses that have owners with below-median AGIs, the number of IRS returns ( $\omega_i^I$ ) has risen from about 5 million in 1988 to over 12 million in 2015, but the SCF estimate ( $\omega_i^S$ ) has remained at roughly 2 million for the entire period. For businesses with above-median AGIs, the number of IRS returns has risen from a little over 8 million to above 12 million, but the SCF estimate has hovered around 5 million. Comparing these data to the full sample in Figure 6, we find that the share of sole proprietorship returns below the median AGI is around 25 percent in the SCF and 43 percent in the IRS. These findings suggest a significant underrepresentation of low-income businesses, which leads to an overstatement of business incomes if business income is positively correlated with AGI.

Finally, we provide evidence on the last term in (1), which summarizes the measurement error. As mentioned earlier, without linked records, it is impossible to directly validate measurement errors, but we document several aspects of the SCF survey results that suggest they are nontrivial. The first is related to the aforementioned observation that a significant part of the overstatement of income is due to an understatement of losses. This could

arise from the framing of the question “What is your net income?” which could be misinterpreted as being a question about positive net income. For instance, consider the distribution of losses by AGI bins for tax year 2015 as shown in Table 2. We see that 10 out of 19 bins, which account for 23 percent of the total number of returns and 26 percent of the total losses in the IRS, have an aggregate zero (that is, all respondents in those income brackets reported a zero net income) in the SCF data.

A second reason to be suspicious about misreported incomes in the SCF is that a very small fraction of respondents refer to their tax documents when responding to questions about the specific line items on tax forms. At the end of the survey, SCF interviewers note how frequently respondents accessed particular documents while answering questions and the type of documents they referenced, if any. Using this information, we calculate the frequency with which business owners referenced either tax or other financial documents in tax year 2015.<sup>16</sup> These tabulations are shown in Table 3. The first row shows that 75 percent of business owners in the SCF never referenced tax documents, 2 percent rarely did, 9 percent sometimes did, and 14 percent frequently did. The second row shows that 64 percent never referenced any other financial documents, 6 percent rarely did, 15 percent sometimes did, and 15 percent frequently did.

To provide further evidence on measurement errors, we show that the SCF fails a simple consistency check by comparing answers to two closely related questions. In the case of sole proprietors, respondents are asked to report incomes listed on lines 12 and 18 of their Form 1040, which correspond to Schedule C and F incomes, respectively. Separately, they are asked about business income from a sole proprietorship and told it is listed on line 31 of Schedule C.<sup>17</sup> By design, the difference in responses to these two questions must be farm income from Schedule F. In Figure 7, we see that the differences across

---

<sup>16</sup> Other financial documents include account statements, investment and business records, loan documents, and pension documents. If any of these documents are referenced, we assume all are.

<sup>17</sup>The first answer is coded as X5704 and the second as X3132, X3232, and X3332, combined with the response to legal status of the actively managed business with codes X3119, X3219, and X3319.

the two answers vary between \$17,000 and \$40,000 per return, considerably more than could be attributable to farm incomes. In a typical year, only 4 percent of business profits listed on Form 1040 are farm income, and farm losses exceed profits in many of the years of our sample.

A related exercise is to check whether there are SCF respondents who report that they do not own or actively manage a sole proprietorship but still report nonzero income from a sole proprietorship or farm. For example, in 2015, of the 16 million who reported a nonzero income (coded as X5704), only 6 million reported that they actively manage a sole proprietorship (coded as X3119, X3219, or X3319), while 10 million reported that they did not. More importantly, the fraction of misreported income is significant. According to SCF data, 65 percent of the business income from Schedule C and F was earned by those reporting that they did not actively manage a sole proprietorship.

We turn next to measures of business valuations and rates of return, which are key for measuring wealth inequality as well as disciplining theories of entrepreneurial activity.

## 4 Business Valuations and Rates of Return

A challenge in estimating valuations and returns for privately held businesses is that they are not frequently traded, and for this reason, most researchers use the SCF to study the distributional aspects of business wealth and returns.<sup>18</sup> SCF business valuations are self-reported evaluations of the current net worth of the business if sold. We can use these estimates along with net incomes to construct SCF income yields that are comparable to available yields from brokered private business sales and publicly traded firms, both small and large. We find that for virtually all subsamples and all years, the SCF yields are significantly higher and more right-skewed than comparable measures from the brokered sales and public firms. We relate this finding to a previous empirical literature that has documented a puzzlingly small private equity premium and

---

<sup>18</sup>Some studies use aggregate business valuations from the Federal Reserve's flow of funds, which imputes values for privately held businesses using estimates of private business revenues, publicly traded business revenue-to-value ratios, and an estimate of the liquidity premium on public versus private business.

show that the earlier estimates are driven by an inconsistent imputation of capital gains on private businesses.

We start by describing the measurement of SCF income yields, which will be compared to yields based on broker data from Pratt’s Stats and public firm data from the CRSP-Compustat merged database. The SCF income yield, which is computed for each business, is the ratio of total pretax net income from businesses divided by the self-reported total net worth of businesses. Let  $\{\omega_{i,t}\}$  be the SCF population weights for survey year  $t$ . We compute an equally weighted and value-weighted mean yield for  $t$ , denoted as  $R_t^{ew}$  and  $R_t^{vw}$ , respectively:

$$R_t^{ew} = \sum_i \omega_{i,t} \left( \frac{NI_{i,t}}{V_{i,t}} \right), \quad R_t^{vw} = \sum_i \left( \frac{\omega_{i,t} V_{i,t}}{\sum_i \omega_{i,t} V_{i,t}} \right) \left( \frac{NI_{i,t}}{V_{i,t}} \right), \quad (2)$$

where  $NI$  is total pretax net income and  $V$  is the self-reported total business value. In Figure 8, we plot time series of yields for all businesses by legal form across years. Across SCF survey years 1989–2016, the average equally weighted yield,  $R_t^{ew}$ , is 102 percent for all businesses, 104 percent for pass-through businesses, and 57 percent for C corporations. The average value-weighted yield,  $R_t^{vw}$ , is 19 percent for all businesses, 20 percent for pass-through businesses, and 17 percent for C corporations. Yields vary significantly across surveys. For example, in the case of C corporations,  $R_t^{ew}$  is in the range of 14 to 102 percent, and  $R_t^{vw}$  is in the range of 11 to 28 percent.

Next, we compare the SCF income yields to comparable statistics from Pratt’s Stats and show that the SCF yields are much higher and more right-skewed. The Pratt’s Stats database has transaction-level data on sales of private and public businesses over the period 1988–2017. The dataset includes financial information about the target business and other attributes of the sale including payment terms, purchase price allocations, and employment agreements. We compute the income yield in Pratt’s Stats by dividing the pretax income earned by the business in the year before the sale by the sale price. The advantage of Pratt’s Stats is that it records the price at which the business was actually sold, thus, it is conceptually close to the ideal answer to the SCF question on business valuation. The results of this comparison

are shown in Table 4. In the first two rows and columns, we report equally weighted and value-weighted mean yields for all businesses in the SCF dataset and the Pratt's Stats dataset. The differences are dramatic: the Pratt's Stats equally weighted yield is 27 percent as compared to 102 percent for the SCF, and the Pratt's Stats value-weighted yield is 2 percent as compared to 19 percent for the SCF. The fact that there is a larger discrepancy in the equally weighted yield than in the value-weighted yield suggests that there are also discrepancies in the distribution of yields. In the last three rows of Table 4, we report percentiles of the income yield distribution across data sources. Here, we see that the 75th percentile yield in the SCF is substantially higher than the counterpart in Pratt's Stats. This result suggests that the SCF overestimates the right-skewness of the cross-sectional distribution of business returns.

An obvious concern about the broker data is its representativeness. There may be a bias in business returns arising from the comparison of ongoing concerns (in the SCF) and a possibly selected subset of businesses that were sold. For instance, if businesses with higher rates of return also have a higher probability of finding potential buyers, then Pratt's Stats returns will be biased upward. On the other hand, there could be factors that lead to a downward bias in the Pratt's Stats returns. Sales triggered by distress, say because of health-related issues facing the owner, would imply a lower yield. Similarly, sales of technology- or research-intensive businesses would imply lower yields because of the significant expensing done by these firms. We can partially correct for the downward bias by ignoring transactions in which the target company is in technology- and research-intensive sectors (that is, with NAICS codes 51, 5415, or 5417) and those for which the stated reason for the sale was health related. In the third column of Table 4, we report the data for this subset of firms and find the results are nearly the same as for all businesses.

Next, we compare SCF income yields to those of publicly traded firms in the CRSP-Compustat database. This comparison has the advantage that we have a sample of businesses that are active, but it has the disadvantage that the typical firm in CRSP is much larger than the typical firm in the SCF. To address this issue, we restrict our attention to larger firms in the SCF (S and C corporations) and compare them to both the full sample of CRSP-Compustat

firms and a subset of small firms. Our baseline definition of “small” is a firm that belongs to the bottom quintile of firms ranked by the book value of assets.<sup>19</sup> The income yield for a firm in CRSP-Compustat is computed by dividing business income (before taxes and retained earnings) by the firms’ market capitalization. The equally weighted and value-weighted means are computed as in (2) where we use the end-of-year market capitalization as a measure of business value. In Table 5, we see that the equally weighted income yield is negative for both the full sample (−9 percent) and the subsample of small firms (−27 percent) in the CRSP-Compustat database, whereas the yield is large and positive for both C corporations (57 percent) and S corporations (76 percent) in the SCF. Similarly, SCF value-weighted yields are significantly higher than CRSP yields, although less dramatically different. Considering the distribution, we again find that the SCF yields are more right-skewed than those in CRSP-Compustat. For example, at the 75th percentile, the SCF C-corporate yield is 36 percent, while the CRSP yield for all businesses is 10 percent.

Our results thus far would appear to be inconsistent with Moskowitz and Vissing-Jorgensen (2002), who constructed private business returns using SCF data and concluded that they were surprisingly low when compared to those of publicly traded firms. We use a longer sample than they do, but we know from Kartashova (2014) that this would account for only about a 6 percentage point difference in the SCF estimates. The more important difference for the quantitative results is the concept of return. The earlier results are based on a measure of return equal to the sum of a value-weighted income yield and an imputed capital gain. In theory, one would need a panel of firm valuations to compute a value-weighted capital gain, namely,

$$R_{t+1}^{cg} = \sum_i \left( \frac{\omega_{i,t} V_{i,t}}{\sum_i \omega_{i,t} V_{i,t}} \right) \left( \frac{V_{i,t+1}}{V_{i,t}} \right), \quad (3)$$

using survey weights  $\{\omega_{i,t}\}$  and valuations  $\{V_{i,t}\}$  for each firm  $i$  in year  $t$ . Given that the SCF survey is triennial with virtually no panel aspect (other than two

---

<sup>19</sup>In Bhandari et al. (2019), we also report results separately for each survey year and for different definitions of “small,” for example, based on gross sales or market capitalization.

surveys), there is no way to compute  $V_{i,t+1}/V_{i,t}$  firm by firm. Moskowitz and Vissing-Jorgensen (2002) instead compute their capital gains measure using the following annualized index:

$$\tilde{R}_{t+3}^{cg} = \left( \frac{\sum_i \omega_{i,t+3} V_{i,t+3}}{\sum_i \omega_{i,t} V_{i,t}} \right)^{\frac{1}{3}} - 1. \quad (4)$$

Their concept of rate of return is given by  $R_t^{vw} + \tilde{R}_t^{cg}$ , where  $R_t^{vw}$  is defined in (2). They adjust the SCF net income by subtracting imputed measures of taxes and retained earnings and compare their measure of return to the value weighted mean holding period return on the CRSP index portfolio.<sup>20</sup> This procedure generates private returns that are similar in magnitude to the CRSP returns. Considering the higher risk for private businesses, Moskowitz and Vissing-Jorgensen (2002) conclude that there is a puzzle as to why individuals become entrepreneurs. Their preferred explanation is that there are non-pecuniary benefits of running a business.<sup>21</sup>

Replicating the exercise of Moskowitz and Vissing-Jorgensen (2002) for our full sample with income yields and capital gains compared separately, we find that the capital gain imputation drives the differences between our findings and theirs. The results are summarized in Table 6. The first two columns show estimates of SCF and CRSP-Compustat income yields,  $R_t^{vw}$ , in all SCF survey years. The last three columns show estimates of  $\tilde{R}_t^{cg}$  for SCF and both  $R_t^{cg}$  and  $\tilde{R}_t^{cg}$  for the CRSP-Compustat sample. The table reveals several noteworthy patterns. First, consistent with our findings for the average income yields, the yearly SCF yields are substantially higher than the CRSP-Compustat counterparts for all survey years. Second, the annualized SCF capital gains vary substantially less than those for firms in the CRSP-Compustat gains  $R_t^{cg}$  over the sample, which is not surprising given the conceptual differences in the measures and the long interval between survey years.<sup>22</sup> If we were to add

---

<sup>20</sup>Since the assumptions underlying the imputations of taxes and retained earnings are ad hoc, we measure  $R_t^{vw}$  using pretax income in both the SCF and CRSP samples.

<sup>21</sup>See also Hamilton (2000) and Hurst and Pugsley (2011), who reach a similar conclusion using data from the SIPP and the PSED, respectively.

<sup>22</sup>Incidentally, the time variation in the capital gains components explains why Moskowitz and Vissing-Jorgensen (2002) and Kartashova (2014) estimate different average returns for

$R_t^{vw}$  plus  $\tilde{R}_t^{cg}$  for SCF and  $R_t^{vw}$  plus  $R_t^{cg}$  for CRSP-Compustat firms, we would confound two discrepancies and conclude that the private and public returns are not very different on average: 26 percent for SCF versus 21 percent for CRSP-Compustat. If we were to restrict attention to comparable measures, either  $R_t^{vw}$  or  $R_t^{vw} + \tilde{R}_t^{cg}$ , we would instead conclude that the private business yields and the imputed total returns are relatively high for private businesses when compared to public returns, not low as previously thought.

As we discussed in Section 3.3, we find that SCF returns are relatively high because there are significant measurement and sampling errors. In the case of self-reported business valuations, respondents would understate values of intangible assets or nontransferable human capital (for example, reputation of the owner). Given that the SCF incomes are verifiably overstated, an understatement of business value would bias the SCF returns upward even more. Furthermore, a rate of return is the ratio of two terms, both of which have measurement errors. This injects a much larger error in the ratio and can cause large discrepancies in the distribution of measured returns with no obvious correction. A constructive way to deal with the measurement issues in the SCF and estimate aggregate and distributional statistics for business valuations is to rely more heavily on a theory that is disciplined by the flows measured from the IRS and business sales data such as Pratt's Stats. A theory featuring business sales would take a stand on the selection bias and also provide a way to impute the valuations for ongoing concerns.<sup>23</sup>

## 5 Robustness

In this section, we show that the overstatement of business incomes in the SCF is robust to potential misreporting in the IRS and to potential miscategorization by SCF respondents across closely related categories of business income.

---

the different sample periods they study.

<sup>23</sup>An attempt in this direction is some ongoing work in Bhandari and McGrattan (2018).



## 5.1 Adjusting for misreporting in the IRS

One explanation for the overstated business incomes in the SCF is that individuals might report true incomes in the surveys but underreport their incomes to the tax authorities. In principle, this should not be a concern for our analysis because the SCF specifically asks them to report what they *wrote* on their tax forms. Nevertheless, we analyze data from several sources on the extent of misreporting on tax forms to evaluate this hypothesis and conclude that tax misreporting is far from sufficient to close the gap between the SCF and IRS business incomes.

The Bureau of Economic Analysis (BEA) estimates tax misreporting for noncorporate income (that is, incomes from sole proprietorships and partnerships) to reconcile the national accounts with the data from tax audits. For the years 1988–2015, the BEA estimates that reported noncorporate tax incomes are lower by roughly 50 percent because of misreporting. These BEA estimates are in line with studies such as Johns and Slemrod (2010), who used tax audit data for the year 2001 and document that Schedule C income is underreported by 54 percent. For S- and C-corporate business incomes, measures of tax misreporting are hard to obtain. Johns and Slemrod (2010) document underreporting of 18 percent for Schedule E income, which includes all supplemental income from S corporations, partnerships, rental real estate, royalties, estates, trusts, and farm rentals. Based on reviews of documents stemming from the National Research Program at the IRS, the Government Accountability Office (GAO) estimates net misreporting margins for S-corporate incomes on the order of 15 to 20 percent. We construct a measure of adjusted IRS pass-through income by adding back the BEA estimates of misreporting for noncorporate incomes, along with an adjustment of 18 percent for income from S corporations based on the study of Johns and Slemrod (2010) and the reports of the GAO. In Figure 9, we compare the SCF business incomes per return to the adjusted IRS incomes per return and find that they are still significantly higher. Computing the SCF errors as before, we find that the average error with the tax audit adjustment is 178 percent, with a range of errors of 98 percent to 274 percent over the sample.

## 5.2 Adjusting for categorization of business income

Another source of measurement error in the SCF is the respondent’s possible confusion about closely related categories of business income. For example, when asked about income from a sole proprietorship appearing on line 31 of Schedule C, business owners might also include income appearing on Schedule E, which includes income from real estate, royalties, partnerships, S corporations, estates, and trusts. From our previous analysis, we know that business incomes from Schedule C are overstated in the SCF. If the overstatement was due to classification errors, we should see an understatement in categories of income corresponding to Schedule E. In Table 7, we report the percentage errors for Schedule E income and document that they are overstated for most survey years.

Johnson and Moore (2008) conducted a similar exercise but constructed an even broader category of business income by including capital gains and losses (lines 13 and 14 of Form 1040) to the incomes individuals report on Schedules C, E, and F. Neither the IRS nor SCF data allow us to isolate the capital gains for business owners or for business-related assets. Hence, we did not include these data in our baseline analysis of business income. In Bhandari et al. (2019), we replicate the findings of Johnson and Moore (2008) and extend them to all survey years. We find that, although the capital gains in the SCF are lower when compared to the IRS gains, the Johnson and Moore (2008) measure of broader business income in the SCF is still larger in all years than its counterpart in the IRS. The average error is 47 percent, with a range of 18 percent to 115 percent across survey years.

## 6 Other Surveys

In this section, we review evidence from other surveys, namely, the CPS, PSID, SIPP, KFS, and PSED. These surveys contain information about businesses and have been widely used by researchers. When comparing business incomes and valuations across these surveys and with the SCF, we find that there are significant inconsistencies but similar concerns related to sampling and

measurement.

We start with the CPS, PSID, and SIPP. All three surveys contain questions about business incomes and organizational forms (that is, whether they are incorporated or unincorporated). The PSID and SIPP additionally contain self-reported estimates of business valuations.<sup>24</sup> Unlike the SCF, the surveys have less detailed information on the legal form of the businesses. For example, these surveys do not distinguish among types of pass-through businesses, and the questionnaires do not specifically connect responses to line items on tax forms. In order to compare across surveys, we focus on business income per owner and income yields for unincorporated businesses.

In Figure 10, we plot incomes per owner for four surveys (SCF, CPS, PSID, and SIPP) and the IRS in Panel A and the number of owners for all surveys in Panel B.<sup>25</sup> As with the SCF, the CPS, PSID, and SIPP have higher business income per owner than is reported by the IRS, but the magnitudes are statistically different across surveys. The SCF is highest with estimates in the range of \$29,000–\$100,000, the PSID is next with a range of \$15,000–\$55,000, the CPS after that with a range of \$15,000–\$35,000, and the SIPP is lowest with a range of \$13,000–\$18,000. All are higher than the IRS, which has a range of \$5,000–\$15,000.<sup>26</sup> The inconsistencies between surveys are driven primarily by differences in aggregate business incomes. The number of owners across these surveys are not significantly different from each other—on the order of 10 to 13 million and stable across years—but are far lower than the IRS, which reports roughly 35 million owners in 1988 and over 50 million by 2015.<sup>27</sup>

---

<sup>24</sup>See Bhandari et al. (2019) for more details.

<sup>25</sup>Our sample in the PSID starts in 1992 and provides annual data until 1996 and biennially after that until 2014. The SIPP reports business incomes every four months for the years 2004–2006 and 2009–2012, and valuations are reported once a year for 2004, 2005, and 2009–2011 depending on when the “topical” modules are available.

<sup>26</sup>Hurst, Li, and Pugsley (2014) combine spending data from the Consumer Expenditure (CE) survey with the PSID and estimate that self-employed individuals underreport income by about 25 percent relative to an imputed measure of true income. The imputation relies on estimating the relationship between expenditures and incomes for wage and salary workers and using it along with food expenditures for the self-employed to infer “true” income of the self-employed. We instead compare survey responses directly to IRS data.

<sup>27</sup>As in the SCF, these surveys only account for partners who are individuals. However, as we mentioned before, using estimates from Cooper et al. (2016), this fact alone does not

Next, we use the responses on self-reported business valuations to compute income yields, as we did for the SCF in Section 4. In Table 8, we see that value weighted income yields in the PSID and SIPP are comparable to the SCF even though business income per owner is lower than that in the SCF by a factor of two or three. This implies that average business values are even lower in these other surveys. However, if we compare yields across the distribution, we see large differences, especially in the right tail. These observations point to the lack of representativeness in the PSID and SIPP for the universe of unincorporated businesses as well as their lack of comparability to the SCF.

For KFS, Gurley-Calvez et al. (2016) compare responses about receipts, expenses, and profits with matched tax forms for an eight-year panel of new businesses beginning in 2004. They match responses from Form 1040, Schedule C for sole proprietorships, Form 1065 for partnerships, and Form 1120S or 1120 for corporations. Eighty percent of firms are matched to tax files, and the matched data file includes 3,940 firms. They find that the businesses in the survey overstate receipts and overstate expenses by even more, implying that the businesses understate profits across the distribution. These findings are for the most part in contrast to the SCF and IRS comparison, as the SCF overstates business income, while the KFS firms understate business income. We report estimates from their study in Table 9 for ease of comparison.

The PSED provides information about business start-ups using a nationally representative sample. However, from the perspective of our study, the PSED suffers from a critical measurement issue: the response rates for business-related questions is very low in all years of the survey. For example, among the 1,214 entrepreneurs in the 2005 panel, only 115 (that is, 9 percent) responded to the question that asks about calculated profits and losses during a follow-up interview for tax year 2006. Thus, we would caution against any use of statistics for quantitative research on entrepreneurial activity from this survey given the small sample size.

In summary, we find severe measurement issues with other surveys currently being used to study U.S. businesses. Key statistics drawn from these surveys are inconsistent with administrative data from the IRS and are incon-  

---

help to account for the massive understatement in the number of owners.

sistent with each other.

## 7 Conclusion

This paper has examined the reliability of widely used survey data for studies of U.S. businesses. We compared key statistics for net incomes and receipts to counterparts in administrative data from the IRS and found large sampling and measurement errors. In all surveys examined, we found that incomes are significantly overstated relative to IRS data, even when respondents are asked to provide incomes from specific lines on their tax forms. The errors we document are large on average and vary wildly across years and across surveys. We provide evidence that the overstatements of income may be due to the non-representativeness of business owners with lower incomes and to the fact that the majority of respondents do not reference any tax or financial documents. We also consider the implications for key statistics used in economic research, such as the level and dispersion of wealth and the return on businesses.

We hope and expect that our analysis will lead to improved measurement in future surveys. Studies of wealth inequality, entrepreneurial choice, and business taxation are using current surveys as predictive tests for economic theory. Our findings suggest that the current data should be treated with great caution, but we hope improvements in sampling will lead to improvements in quantitative predictions in the future. Attempts should be made to link responses to administrative data where possible. Questions should be limited to queries that are verifiable. In the case of businesses, care should be taken to ensure representative samples of all types of legal organizations.

## References

- [1] Bhandari, A., & McGrattan, E. R. (2018). Sweat equity in U.S. private business. Federal Reserve Bank of Minneapolis, Staff Report 560.
- [2] Bhandari, A., Birinci, S., McGrattan, E. R., See, K. (2019). Data appendix: What do survey data tell us about U.S. businesses? Federal Reserve Bank of Minneapolis, Staff Report 578.
- [3] Bohacek, R., & Zubricky, J. (2012). A flat tax reform in an economy with occupational choice and financial frictions. *Economic Journal*, 122(565): 1313–1345.
- [4] Bricker, J., Henriques, A., Krimmel, J., & Sabelhaus, J. (2016). Measuring income and wealth at the top using administrative and survey data. *Brookings Papers on Economic Activity*, 2016(1): 261–331.
- [5] Buera, F. J. (2009). A dynamic model of entrepreneurship with borrowing constraints: Theory and evidence. *Annals of Finance*, 5(3–4): 443–464.
- [6] Cagetti, M., & De Nardi M. (2006). Entrepreneurship, frictions, and wealth. *Journal of Political Economy*, 114(5): 835–870.
- [7] Cooper, M., McClelland, J., Pearce, J., Prinsziano, R., Sullivan, J., Yagan, D., Zidar, O., & Zwick, E. (2016). Business in the United States: Who owns it, and how much tax do they pay? *Tax Policy and the Economy*, 30(1): 91–128.
- [8] De Nardi, M., Doctor, P., & Karen, S. D. (2007). Evidence on entrepreneurs in the United States: Data from the 1989–2004 Survey of Consumer Finances. *Federal Reserve Bank of Chicago Economic Perspectives*, 4:18–36.
- [9] Dungan, A. (2017). Sole proprietorship returns, tax year 2015. *Statistics of Income Bulletin*, 37(2): 2–24.
- [10] Gentry, W. M., & Hubbard, R. G. (2004). Entrepreneurship and household saving. *Advances in Economic Analysis and Policy*, 4(1), Article 1.
- [11] Gurley-Calvez, T., Bruce, D., Reedy, E. J., & Russell, J. (2016). Comparing survey data and tax data: Differences in reporting across businesses. Working Paper, *Statistics of Income*.
- [12] Hamilton, B. (2000). Does entrepreneurship pay? An empirical analysis of the returns to self-employment. *Journal of Political Economy*, 108(3): 604–631.
- [13] Hurst, E., Li, G., & Pugsley, B. W. (2014). Are household surveys like tax forms? Evidence from income underreporting of the self-employed. *The Review of Economics and Statistics*, 96(1): 19–33.
- [14] Hurst, E., & Pugsley, B. W. (2011). What do small businesses do? *Brookings Papers on Economic Activity*, 43(2): 73–142.

- [15] Hurst, E., & Pugsley, B. W. (2017). Wealth, tastes, and entrepreneurial choice. In *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges*, ed. J. Haltiwanger, E. Hurst, J. Miranda, and A. Schoar. Chicago: University of Chicago Press.
- [16] Johns, A., & Slemrod, J. (2010). The distribution of income tax noncompliance. *National Tax Journal*, 63(3): 397–418.
- [17] Johnson, B. W., & Moore, K. (2008). Differences in income estimates derived from survey and tax data. In *Proceedings of the Joint Statistical Meetings, Survey Research Methods Section*. Alexandria: American Statistical Association.
- [18] Kartashova, K. (2014). Private equity premium puzzle revisited. *American Economic Review*, 104(10): 3297–3334.
- [19] Kitao, S. (2008). Entrepreneurship, taxation, and capital investment. *Review of Economic Dynamics*, 11(1): 44–69.
- [20] Kuhn, M. & Rios-Rull, J. V. (2016). 2013 update on the U.S. earnings, income, and wealth distributional facts: A view from macroeconomics. *Federal Reserve Bank of Minneapolis Quarterly Review*, 37(1): 2–73.
- [21] McGrattan, E. R., & Prescott, E. C. (2010a). Unmeasured investment and the puzzling U.S. boom in the 1990s. *American Economic Journal: Macroeconomics*, 2(4): 88–123.
- [22] McGrattan, E. R., & Prescott, E. C. (2010b). Technology capital and the U.S. current account, *American Economic Review*, 100(4): 1493–1522.
- [23] Meh, C. (2005), Entrepreneurship, wealth inequality, and taxation. *Review of Economic Dynamics*, 8(3): 688–719.
- [24] Moskowitz, T. J., & Vissing-Jorgensen, A. (2002). The returns to entrepreneurial investment: A private equity premium puzzle? *American Economic Review*, 92(4): 745–778.
- [25] Quadrini, V. (2000). Entrepreneurship, saving, and social mobility. *Review of Economic Dynamics*, 3(1): 1–40.
- [26] Saez, E. & Zucman, G. (2016). Wealth inequality in the United States since 1913: Evidence from capitalized income tax data. *Quarterly Journal of Economics*, 131(2): 519–578.
- [27] Scheuer, F. (2013). Entrepreneurial taxation with endogenous entry. NBER Working Paper 19235.
- [28] Smith, M., Yagan, D., Zidar, O., & Zwick, E. (2017). Capitalists in the twenty-first century. Working Paper, U.S. Department of the Treasury.

Table 1: Decomposition of SCF-IRS Pass-Through Business Income Gap

Tax Year	SCF–IRS \$ Bill.	Percentage of Gap	
		Overstatement of Profit	Understatement of Loss
1988	159	50	50
1991	307	64	36
1994	624	83	17
1997	386	63	37
2000	635	68	32
2003	785	71	29
2006	1,096	77	23
2009	750	41	59
2012	218	–56	156
2015	1,693	96	4
Mean	665	56	44

*Note:* This table shows the difference (gap) between aggregated business income of all pass-through businesses in the SCF and the IRS. The gap is then decomposed into the fraction attributable to an overstatement of profits or that attributable to an understatement of losses.



Table 2: Sole Proprietorships with Net Losses in the IRS and SCF by AGI Bins, 2015

AGI Bins	IRS		SCF	
	Returns '000	Losses \$ Mil.	Returns '000	Losses \$ Mil.
No adjusted gross income	426.0	12.2	91.4	0.2
\$1 under \$5,000	138.3	0.9	39.7	0.2
\$5,000 under \$10,000	185.7	1.5	33.3	0.0
\$10,000 under \$15,000	270.8	2.4	10.6	0.0
\$15,000 under \$20,000	344.3	3.5	47.9	0.0
\$20,000 under \$25,000	351.4	3.1	60.0	0.2
\$25,000 under \$30,000	316.8	3.0	77.5	0.2
\$30,000 under \$40,000	533.0	3.9	102.2	0.6
\$40,000 under \$50,000	469.3	3.4	62.8	0.0
\$50,000 under \$75,000	833.7	5.8	159.3	0.1
\$75,000 under \$100,000	626.4	4.3	199.5	0.8
\$100,000 under \$200,000	1047.9	7.7	216.2	0.8
\$200,000 under \$500,000	312.4	3.7	71.6	0.4
\$500,000 under \$1,000,000	50.4	1.3	0.0	0.0
\$1,000,000 under \$1,500,000	11.6	0.6	0.6	0.0
\$1,500,000 under \$2,000,000	5.3	0.4	0.0	0.0
\$2,000,000 under \$5,000,000	8.4	1.0	0.1	0.0
\$5,000,000 under \$10,000,000	2.3	0.5	0.7	0.0
\$10,000,000 or more	1.8	1.3	36.6	0.0

*Note:* This table shows the number of business returns that report a net loss and the corresponding amount of these net losses across various AGI bins for tax year 2015.

Table 3: Percentage of Respondents Checking Documents in SCF 2016

	Never	Rarely	Sometimes	Frequently
Income tax document	75	2	9	14
Other financial documents	64	6	15	15

*Note:* This table shows the fraction of business owners that refer to their income tax documents or other relevant financial documents in varying frequency. A respondent who referred to account statements, investment/business records, or loan documents is considered to have checked other financial documents.

Table 4: Net Income Yields in the SCF and Pratt's Stats

Moments	SCF	Pratt's Stats	
	All businesses	All businesses	Non-tech & nondistressed
Equally weighted mean	102.5	27.4	29.3
Value weighted mean	19.1	1.9	3.5
p25	0.9	3.8	5.0
p50	17.6	21.7	23.1
p75	63.0	46.8	48.3

*Note:* This table shows moments of the net income yield distribution from the SCF and Pratt's Stats. For Pratt's Stats, we also consider income yields for a subset of businesses that excludes those in technology- and research-intensive sectors (NAICS codes 51, 5415, or 5417) and those for which the stated reason for the sale was health related.

Table 5: Net Income Yields in the SCF and CRSP-Compustat

	SCF		CRSP-Compustat	
	C Corps	S Corps	All businesses	Small businesses
Equally weighted mean	56.8	76.4	-9.2	-26.6
Value weighted mean	16.9	15.2	7.3	-8.5
p25	1.3	2.2	-5.5	-29.0
p50	10.6	14.2	5.4	-7.7
p75	36.2	50.5	10.4	4.0

*Note:* This table shows moments of the net income yield distribution from the SCF and the CRSP-Compustat database. For the CRSP-Compustat sample, small businesses refer to publicly traded firms in the CRSP database that belong to the bottom 20 percent when ranked by total assets.

Table 6: Net Income Yields and Capital Gains

Tax Year	Net Income Yields		Capital Gains		
	SCF	CRSP	SCF	CRSP-Compustat	
				$(t-1) \rightarrow t$	$(t-3) \rightarrow t$
1988	16.6	12.4	—	—	—
1991	20.7	6.2	0.2	26.9	13.2
1994	31.5	9.8	5.3	-3.2	8.5
1997	20.6	6.2	11.4	30.2	29.7
2000	22.6	4.6	11.7	3.7	13.8
2003	17.7	6.2	6.6	28.6	-4.8
2006	18.1	8.0	15.9	10.3	8.9
2009	14.8	5.7	-7.9	21.6	-8.6
2012	14.1	8.0	2.9	12.0	9.6
2015	14.6	5.4	12.8	-3.0	10.7
Mean	19.1	7.3	6.6	14.6	9.0

*Note:* This table shows estimates of income yields and capital gains for businesses in the SCF and CRSP-Compustat firms. For the SCF, capital gains are computed using Equation 4 found in the main text, as in Moskowitz and Vissing-Jorgensen (2002). For the CRSP-Compustat firms, we report two measures of capital gains. The column  $(t-1) \rightarrow t$  measures the realized capital gains using Equation 3 for year  $t$  where  $t$  corresponds to the fiscal year for which income is reported in the SCF. The column  $(t-3) \rightarrow t$  measures a geometric mean of the capital gains for the index over the past three periods using equation 4.

Table 7: Schedule E Income Comparison

Tax Year	IRS	SCF	Error
	\$ Bill.	\$ Bill.	%
1988	57.3	116.1	102.6
1991	69.9	129.6	85.6
1994	133.0	121.8	-8.4
1997	195.3	147.0	-24.7
2000	249.0	180.3	-27.6
2003	292.7	427.1	45.9
2006	463.1	805.6	74.0
2009	380.8	720.7	89.3
2012	613.3	949.3	54.8
2015	713.2	1142.1	60.1

*Note:* This table shows aggregated Schedule E income from the IRS and respondents' reported Schedule E income in the SCF. Dollar amounts are in billions.

Table 8: Income Yield Distribution of Noncorporate Businesses in the SCF, SIPP, and PSID

	SCF	SIPP	PSID
Value-weighted mean	14.5	17.7	14.9
p25	0.0	2.2	3.2
p50	18.0	33.2	27.0
p75	79.2	230.1	114.9

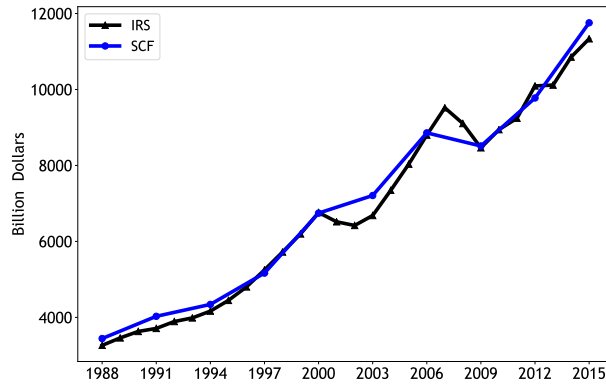
*Note:* This table shows moments of the income yield distribution for noncorporate businesses for the SCF, SIPP, and PSID. The columns average the income yields for all years that the data are available. The SCF is available triennially between 1989 and 2016, the SIPP for the years 2004–2005 and 2009–2011, and the PSID biennially between 1988 and 2014.

Table 9: Comparison of KFS and IRS Business Tax Data, 2004–2011

Statistic	Receipts			Expenses			Profit		
	KFS '000	IRS '000	Error %	KFS '000	IRS '000	Error %	KFS '000	IRS '000	Error %
Mean	552	417	32	369	188	96	30	169	−82
Median	92	66	29	57	36	57	5	24	−79
p25	21	11	74	1	12	−1,400	−3	1	−700
p75	350	281	25	236	152	55	31	142	−78
p99	11,500	7,434	55	7,450	2,680	178	810	2,478	−67

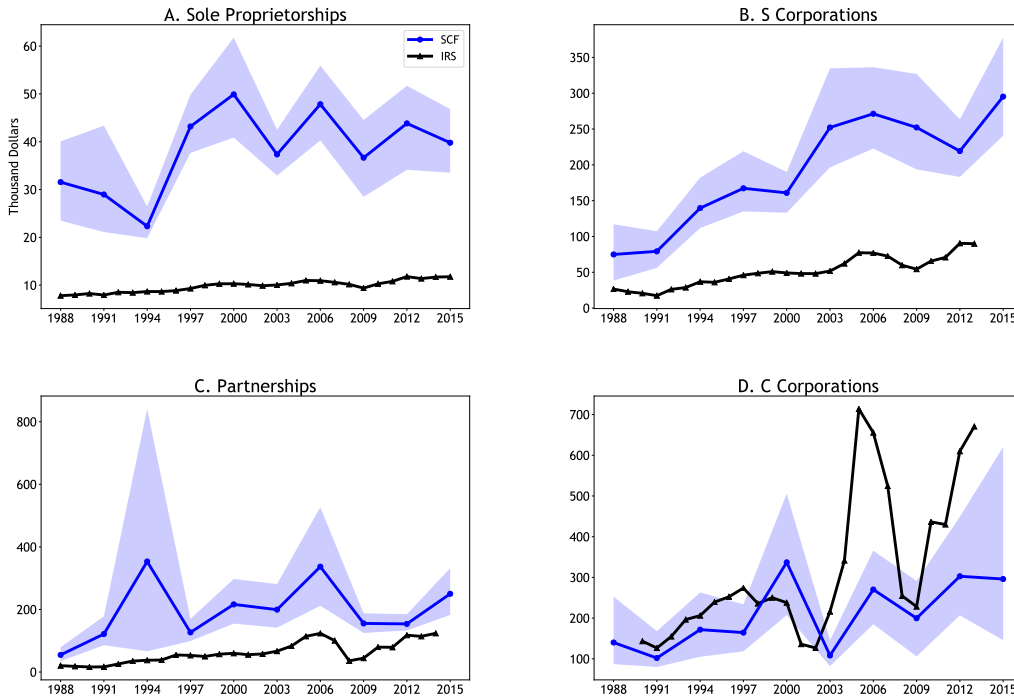
*Note:* The source of statistics is Gurley-Calvez et al. (2016).

Figure 1: Adjusted Gross Incomes: SCF vs. IRS



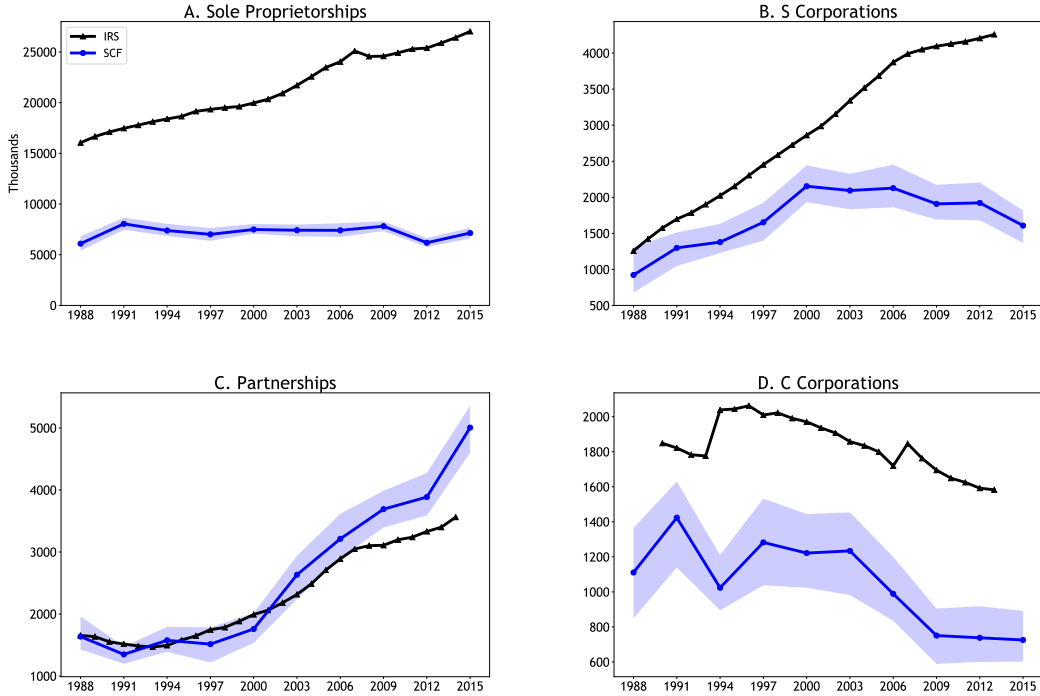
Note: For the IRS, adjusted gross income is obtained from Form 1040. For the SCF, if AGI is not available, we construct it by adding the appropriate income categories.

Figure 2: Business Income per Return by Legal Entity: SCF vs. IRS



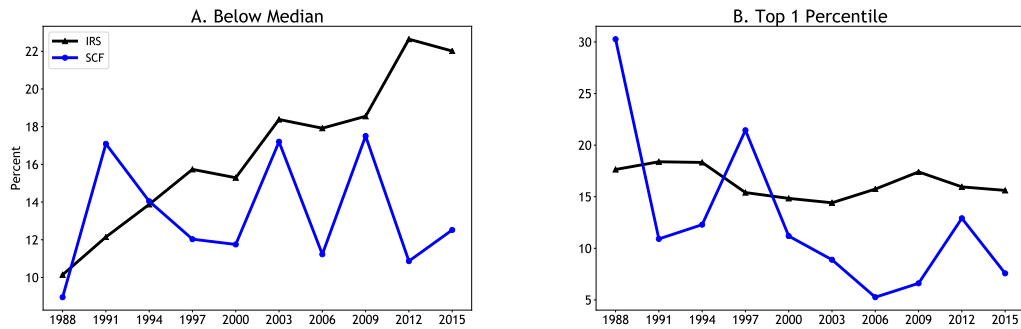
Note: This figure plots business income per business tax return in the IRS and the SCF as reported on Form 1040 Schedule C for sole proprietorships, Form 1120S for S corporations, Form 1065 for partnerships, and Form 1120 for C corporations. IRS data for partnerships, S corporations, and C corporations are available only until 2013. IRS data for C corporations exclude data for those filing 1120A, 1120F, 1120L, 1120PC, 1120REIT, 1120RIC. Prior to 1990, only consolidated information is available and thus is not comparable to the series plotted here. The shaded region for the SCF shows the 90 percent confidence interval.

Figure 3: Number of Returns by Legal Entity: SCF vs. IRS



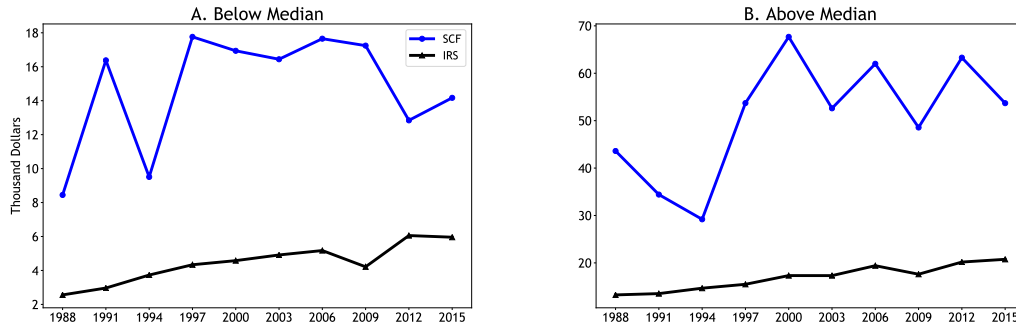
*Note:* This figure plots the number of business returns of sole proprietorships, S corporations, partnerships, and C corporations over time in the IRS and the SCF. IRS data for partnerships, S corporations, and C corporations are available only until 2013, and C-corporation data are unavailable prior to 1990 when only consolidated information is available and thus is not comparable to the series plotted here. The shaded region for the SCF shows the 90 percent confidence interval.

Figure 4: Proprietor Income Shares: SCF vs. IRS



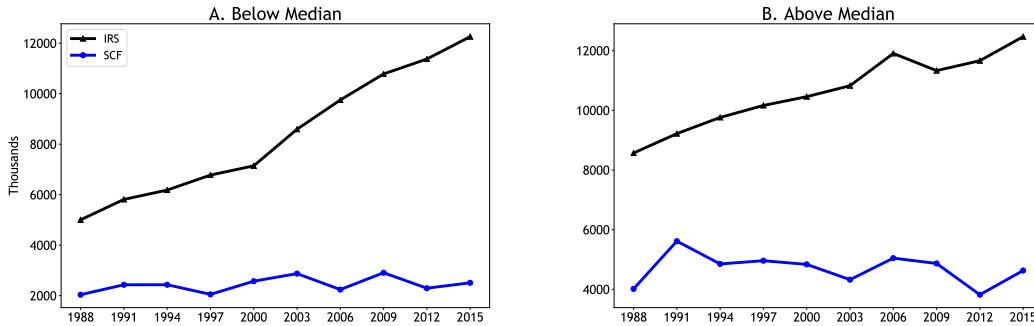
*Note:* This figure plots the fraction of business income from sole proprietorships attributable to returns with AGI below the median and above the 99th percentile.

Figure 5: Income Per Return, Proprietors with Below- and Above-Median AGI: SCF vs. IRS



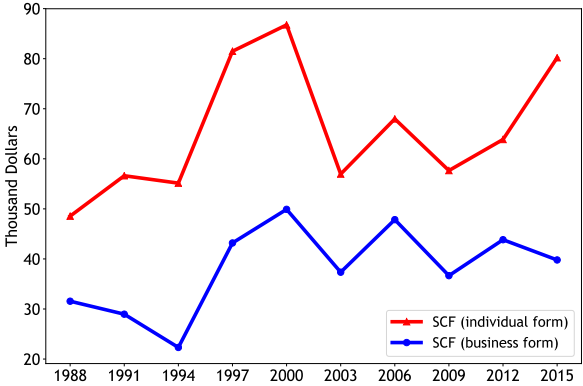
Note: This figure plots sole proprietorship business income per return for those with below- and above-median AGI.

Figure 6: Number of Returns, Proprietors with Below- and Above-Median AGI: SCF vs. IRS



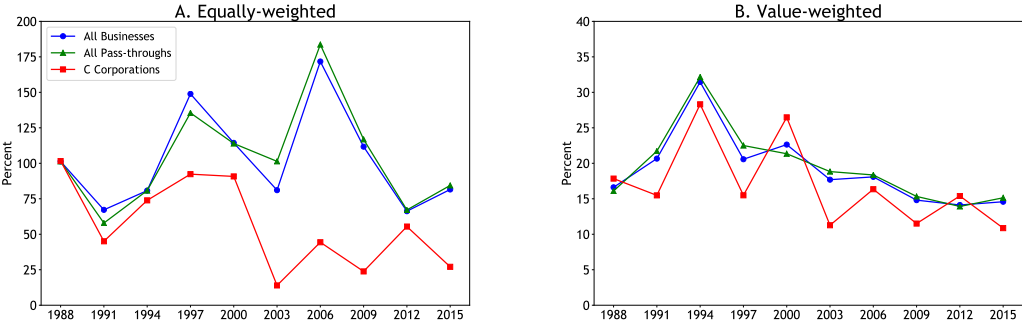
Note: This figure plots the number of sole proprietorship returns (Form 1040 Schedule C) filed by business owners with below- and above median AGI.

Figure 7: Comparing Proprietors' Individual and Business Incomes, SCF



Note: This figure plots business income per return in the SCF for questions that ask respondents to report individual incomes listed on Form 1040, lines 12 plus 18, and business income on Schedule C of 1040, line 31.

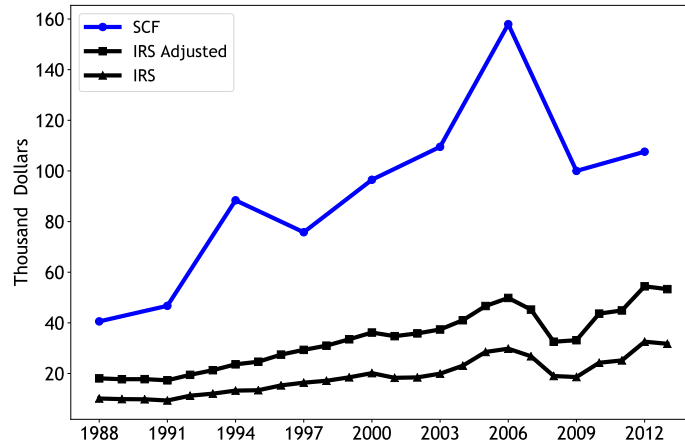
Figure 8: Equally and Value-Weighted Average Net Income Yields, SCF



Note: This figure plots equally weighted and value-weighted average net income yields. The SCF sample includes businesses with positive net worth and excludes the bottom 1st percentile of these businesses. The business income of each business that the family members own in the SCF is obtained from SCF variables that correspond to information on business tax forms.

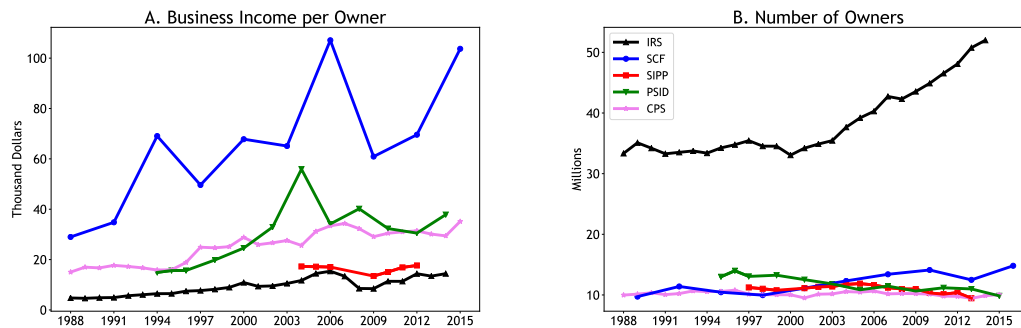


Figure 9: Unincorporated Business Income per Return with Tax Misreporting Adjustments, SCF vs. IRS



Note: In this figure, we use BEA estimates for misreporting of noncorporate business incomes and reports from the Government Accountability Office (GAO) on misreporting of S-corporation business incomes to adjust IRS pass-through business income per return. We add these yearly adjustments to the sum of pass-through income in the IRS, calculate total business income per tax return, and compare it with estimates from the SCF.

Figure 10: Unincorporated Business Income per Owner and Number of Owners



Note: This figure plots the total business income per owner of unincorporated businesses (Panel A) and total number of unincorporated business owners (Panel B) in the SCF, CPS, PSID, SIPP, and the IRS. Before 2004, the SIPP does not provide information about an individual's own share of business income from an unincorporated business. Instead, it contains information about the total income of the business, which is not enough information to calculate the total business income of unincorporated businesses.