



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 Gerrard See
University of Minnesota

Staff Report 568
Revised August 2018

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 August 2018

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 Gerrard See
University of Minnesota

ABSTRACT

This paper examines the reliability of survey data for research on U.S. businesses, including sole proprietorships, partnerships, S corporations, and C corporations. We examine all surveys that ask questions about these businesses and compare outcomes across surveys and with aggregated administrative data. We document large inconsistencies in business incomes, receipts, and number of returns. We highlight problems due to non-representative samples and measurement errors. Non-representativeness is reflected in undersampling of businesses, especially in categories of owners with low total incomes. Measurement errors arise because respondents do not refer to relevant documents when answering survey questions and also because some questions are framed in a manner that is confusing to respondents. Finally, we discuss the implications for statistics of interest, such as returns and valuations of ongoing private businesses, as well as the implications for research that studies positive and normative questions pertaining to businesses.

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

*Bhandari acknowledges support from the Heller-Hurwicz Economic Institute, and McGrattan acknowledges support from the NSF. We thank Joan Gieseke for excellent editorial assistance.

1 Introduction

Representative surveys of households and firms are useful for many macroeconomic questions and have been used extensively by researchers. Relative to administrative data, which have a large number of observations but few details per observation, survey data typically contain much more information at the individual level, which aids in isolating economic mechanisms. This paper examines the reliability of survey data for research on U.S. businesses, including pass-through entities and subchapter C corporations.¹ 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.² Subchapter C corporations account for the remaining half and include all publicly traded firms. We examine data from all surveys that ask questions about these businesses and document issues that arise due to non-representative samples and measurement errors. Finally, we discuss the implications for statistics of interest, such as the returns and valuations of businesses, as well as the implications for research that studies positive and normative questions pertaining to businesses.

Our approach uses publicly available and widely used surveys such as the Survey of Consumer Finances (SCF), Survey of Income and Program Participation (SIPP), Kauffman Firm Survey (KFS), Panel Study of Income Dynamics (PSID), and Panel Study of Entrepreneurial Dynamics (PSED) as well as aggregated data from the Internal Revenue Service (IRS) and national income and product accounts (NIPA). We check the reliability of certain statistics by aggregating survey answers and comparing them with each other and with administrative data totals, if possible by subgroups in the population. We do this in the case of total adjusted gross income, which matches well, and number of returns, net income, and receipts of businesses which do not. We also find that distributions of business incomes do not match comparable administrative data.

An important survey for analyzing business income and wealth is the SCF, which has many detailed questions about business ownership and finances. Households with actively-managed businesses are asked to report net income from a specific line on a tax form.³ This method makes it easy for us to compare the data with the actual incomes reported by the IRS. We find that, depending on the year, the total business income of pass-throughs is overstated in the SCF by factors of 2 to 3 depending on the survey

¹For tax purposes, pass-through entities classify themselves as sole proprietorships, S corporations, and partnerships. They are called “pass-through” because the income earned by such businesses is taxed under the owners’ individual income tax. In contrast, in C corporations, the company itself pays corporate taxes on income.

²Smith 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 private businesses.

³Sole proprietors are asked to report business net income from Form 1040 Schedule C line 31, shareholders of partnerships from Form 1065 line 22, shareholders of S corporations from Form 1120S line 21, and shareholders of C corporations from Form 1120 line 30.

year, while the total business income of C corporations is understated by a factor of one-half on average. When we compare the SCF estimates of incomes to the SIPP, KFS, PSID, and PSED, we find that the other survey results have different measurement issues. For example, while the SCF respondents with pass-through businesses overstate their incomes, the SIPP and KFS respondents understate them. The SCF provides details about the legal status of businesses – for example, whether they incorporated as subchapter S or C or are unincorporated as in the case of a sole proprietorship or a partnership – whereas the PSID does not provide these details and therefore cannot be validated with administrative data. The PSED provides information on annual household income, type of business, and profits or losses from the business. However, only 9 percent of the sample responded to the question that asks about their calculated profits or losses.

To check whether the samples are representative, we compare the number of tax returns by categories of adjusted gross incomes. For pass-through businesses, we find that the total number of returns in surveys is lower than what is reported by the IRS, and the distribution is skewed toward businesses that have owners with relatively large total incomes. From this we conclude that the samples are not representative. For example, in the case of the SCF, the number of business returns for pass-through businesses is low by a factor of 2 in the late 1980s and by a factor of 3 more recently. In the SIPP, returns of unincorporated pass-through businesses are understated by a factor of 2 until the mid-2000s and by a factor of 3 in more recent surveys. In the case of C corporations, we use the SCF and find that the total number returns is lower than IRS returns by a factor of 2.

To quantify measurement errors, we report the extent to which households use supporting documents when answering the survey questions and then check the consistency of answers to related questions. The SCF, for example, provides data on frequency of reference to supporting documents, and the estimates are strikingly low. For example, if we condition on all households, we find that only about 7 percent of households “frequently” reference their tax documents (as opposed to “sometimes,” “rarely,” or “never”). If we ask how many households frequently reference all relevant documents, then the number drops to about 1 percent. If we loosen the criterion by asking how many business-owning households at least rarely reference a tax document, we find that it is 24 percent, implying that 76 percent of business owners never look at their tax documents. In terms of consistency of answers, problems range from respondents not knowing that a sole proprietorship has to file a Schedule C with IRS Form 1040 to not knowing that a net loss implies a negative value for net income.

We show that problems exist with the survey data even if we adjust for tax misreporting. If households underreport incomes to the IRS but accurately report income to the surveyor, we would find an overstatement of incomes, as is the case for pass-through businesses in the SCF. We use tax audit data to correct the administrative data but still find a mismatch with the survey data. Other adjustments, such as correcting

for within-survey inconsistencies in the SCF regarding business ownership and income and for the fact that the SCF only surveys partners who are individuals, do not alleviate the measurement issues.

Finally, we examine survey responses to a question about the value of ongoing businesses, a statistic that is relevant for computing the dispersion in household wealth and business returns. Since there are no measures of total valuations for businesses other than publicly-traded C corporations, we use measures of dividend yields from S&P 500 company data and compare them to closest counterpart in survey data. We find large differences, even for respondents who own and actively manage a C corporation. For example, the value-weighted average for C corporations is in the range of 10 and 26 percent depending on the survey year – much higher than the 2.3 percent dividend yield in the S&P 500 data – with the discrepancies due to problems in measuring both incomes and valuations. The value-weighted average dividend yield for pass-through businesses is around 20 percent, which is also much higher than S&P 500 data.⁴ An even larger gap is present when comparing equally-weighted averages, which is as high as 100 percent for businesses surveyed by the SCF but only 1.9 percent in the S&P 500.

The significant disparity between value-weighted and equally-weighted dividend yields in the SCF indicates the presence of many businesses with unreasonably high returns. For example, when we compute the distribution of pass-through business returns, we find that more than half of all businesses have dividend yields higher than 15 percent, while 90 percent of businesses in the S&P 500 have dividend yields below 6 percent. From these statistics, we conclude that SCF pass-through yields are on average 10 times larger for value-weighted returns and 50 times larger for equally-weighted returns, and far more dispersed in the cross section, than the S&P yields. When analyzing SIPP data, we find that while the median dividend yield is in line with the S&P data, the distribution is not. Furthermore, when we compare distributions of dividend yields in SIPP and SCF, we find large differences, which again demonstrates inconsistencies across surveys.

Measurement problems related to business valuations and returns may be insurmountable without data on business sales or share transactions. First, it is difficult for owners to estimate business valuations when one considers that businesses are heavily invested in intangible assets.⁵ Second, survey respondents answer questions separately about business income and valuations. For example, if the net incomes derive from both capital and labor inputs, while the business valuations are fixed assets owned by the business, then the estimated dividend yields from surveys are not comparable to publicly-traded company data, and possibly not comparable across business owners who may interpret the question differently. The inconsistencies across surveys and the conceptual measurement issues that we highlight suggest that attempts to interpret survey

⁴The magnitude of the SCF yields we compute are comparable to those found by Moskowitz and Vissing-Jorgensen (2002) and Kartashova (2014).

⁵McGrattan 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 building organizations.

based measures of business returns or valuations without a consistent framework for true returns, stocks, and valuations are largely futile.

This paper is organized as follows. Section 2 discusses the implications of our findings on economic research. In Section 3, we show that aggregate income, wages and salaries, and broad business income in the survey data match their respective counterparts in the IRS data reasonably well. Section 4 documents that business income and receipts in the SCF data and in the IRS data are largely inconsistent, and presents problems of survey data regarding non-representativeness and measurement. Section 5 shows that business returns are unreasonably high in the survey data. Section 6 presents robustness checks of our results for the SCF and compares other survey data to the IRS. 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 statistics – and thus raise issues concerning the theoretical developments and policy analyses that have been designed around them.

There is a large and burgeoning empirical literature that documents trends in income and wealth, which 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. Of particular relevance to our work is their finding that business income is one of the main contributors to income inequality and business equity is one of the main contributors to wealth inequality, a finding they document for the history of the SCF surveys between 1989 and 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 answering questions about their business income or business equity and, therefore, we cannot trust the SCF distributions.

Saez and Zucman (2016) document trends in wealth dispersion using administrative tax data. The tax

data do not offer as much information about households as survey data, but are potentially more reliable for distributional information on incomes. (For example, see Smith et al. (2017) for details on trends in S-corporation incomes.) To estimate household wealth, Saez and Zucman (2016) first estimate capitalization factors and use them to capitalize taxable incomes reported to the IRS. Their estimates of capitalization factors are ratios of aggregate flow of funds wealth measures to aggregate IRS incomes. A separate return is computed for each capital income category, and the authors assume that within a given asset class, everyone receives the same return. Saez and Zucman’s (2016) estimates of the wealth distributions are then found by capitalizing the IRS income distributions, which means they multiply each income category by an estimate of the capitalization factor on that income (ignoring any business losses). 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.⁶

There are several reasons why the Saez and Zucman (2016) capitalization method is inappropriate for estimating wealth in business. First, there is no way to validate the procedure. These authors can only invalidate the procedure by conducting the type of exercise we do here, showing in fact that the SCF wealth measures are not an appropriate benchmark. Second, the flow of funds aggregates used to compute capitalization factors for business are not actual data but rather imputations made by the Federal Reserve. 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 that 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. Third, the assumption that returns are the same for everyone 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 these income and wealth shares increasing

⁶They also compare results to estate taxes and foundation records but these data are not informative about most businesses in the United States.

throughout the distribution, and entrepreneurs have high savings rates relative to the population implying much more dispersion in wealth than in income. (See De Nardi, Doctor, and Krane 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 paper is also related to a second strand of the empirical literature, which uses survey data on incomes and wealth to estimate nonpecuniary drivers of entrepreneurship. Researchers have been debating whether estimated returns of private businesses are puzzlingly low given the risks entrepreneurs face. 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. His empirical results suggest that large nonpecuniary benefits of self-employment play an important role in the occupational choice. 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 equity are no higher than the returns to public equity, leading them to agree with Hamilton’s (2000) conclusion that nonpecuniary benefits or differences in risk tolerances could account for their results. In an update to the Moskowitz and Vissing-Jorgensen (2002) study, Kartashova (2014) finds the data may be less puzzling than once thought because she finds a large gap between private and public firms between 1999 and 2007. Hurst and Pugsley (2011, 2017) followed up on this work by incorporating nonpecuniary benefits in a model of entrepreneurship and then studying the impact of small business subsidies. Our results cast a doubt about SIPP and SCF survey data and hence on policy recommendations that arise from frameworks that focus exclusively on nonpecuniary benefits.

We turn next to our investigation of widely-used survey data.

3 Aggregate Income

We first compare total income in the IRS and survey data. We define total income as the sum of wages and salaries; net income from a business, profession, or farm; taxable and non-taxable interest; dividends; capital gains from the sale of capital assets and other property; net income from rental, royalty, estate, and trust; net income from partnerships and S corporations; unemployment compensation; alimony received;

total pensions and annuities; total social security benefits; as well as other income. This corresponds to adding IRS Form 1040 lines 7 to 21, excluding IRA distributions (line 15a) and taxable refunds, credits, or offsets of state and local income taxes (line 10). Data from the IRS are obtained from Individual Income Tax Returns Publication 1304. When collecting data about individual income, the SCF asks respondents to report information from specific lines on their IRS Form 1040. This makes the SCF survey directly comparable to IRS data. To calculate total income from the SCF, we select variables that refer to each subcomponent of IRS Form 1040 that is included in our definition of total income. Figure 1 shows the result of this comparison. The SCF tracks total income in the IRS relatively well in both levels and cyclical trends. Table 1 shows that the percentage errors of SCF total income relative to IRS total income do not exceed 10 percent and average around 2.6 percent.

We also compare aggregate wages and salaries, and a broad measure of business income in the IRS and SCF. Broad business income is defined as income derived from a business or profession (Form 1040 Schedule C) or farm (Form 1040 Schedule F); income from rental real estate, royalties, partnerships, S corporations, estates, trusts (Form 1040 Schedule E); taxable and non-taxable interest income (Form 1040 line 8a and 8b); ordinary and qualified dividends (Form 1040 lines 9a and 9b); and income from gains from the sale of capital and other property (Form 1040 lines 13 and 14). Figure 2 shows that both total wages and salaries, and broad business income in the SCF are roughly close to their respective counterparts in the IRS.⁷

4 Business Income and Receipts

4.1 Business Income and Receipts of Pass-Throughs and C Corporations

The SCF does relatively well in matching aggregates such as total income, wages and salaries, and broad business income. However, when business income is decomposed into income from different businesses by legal structure, we find that the SCF is no longer able to match aggregate data. Figure 3 plots business income in the IRS and SCF by the legal structure of the business. The IRS business income information comes from income reported on Form 1040 Schedule C for sole proprietorships, Form 1065 for partnerships, Form 1120S for S corporations, and Form 1120 for C corporations. We calculate the same statistics from the SCF using variables that exactly correspond to their IRS counterpart. Panel A to C of the figure shows that for pass-through businesses, business income is significantly and consistently overstated in the SCF relative to the IRS. In the year 2006, for example, total S corporation income reported to the IRS was \$297 billion, while aggregated S corporation income in the SCF amounted to \$577 billion, implying that SCF responses

⁷The close match of broad business income in the SCF and the IRS has been documented by Johnson and Moore (2011). However, broad business income includes income categories such as capital gains/losses, rents and royalties, estates and trusts, as well as interest income which are not very informative about pass-through and C-corporation business incomes.

were overstated by 93.8 percent. In contrast, business income of C corporations is largely understated in the SCF relative to the IRS as shown in Panel D. For example, in 2006, C corporations reported an aggregate net income of \$1.12 trillion to the IRS but aggregated C corporation income in the SCF only amounted to \$267 billion. Table 1 shows the percentage errors of reported business income in the SCF compared with the IRS. It shows that the degree of overstatement in the SCF is large for pass-through businesses, with percentage errors averaging 31.5 percent for sole proprietorships, 305.4 percent for partnerships, and 137.1 percent for S corporations. For C corporations, business income is understated in the SCF by 56.3 percent on average. Furthermore, the degree of understatement or overstatement varies considerably over time. For instance, partnership income is overstated by 889.1 percent for tax year 1994 and by 106.7 percent in the next survey conducted for tax year 1997 while C corporation income is understated by only 12.2 percent in tax year 2000 but by 66.7 percent in the next survey for tax year 2003. Finally, Table 2 also demonstrates discrepancies in business receipts reported in the IRS and in SCF. While there is no consistent pattern of overstatement of business receipts in the SCF, percentage errors vary in sign and magnitude significantly over time as well.

4.2 Distribution of Business Income

The SCF also exhibits significant errors in the distribution of business income when compared with the IRS. Figures 4 and 5 compare SCF and IRS business income for businesses that either report a net loss or report a positive net income respectively. Across all legal structures, businesses that report positive income in the SCF overstate their business income relative to the IRS, while businesses that report incurring a net loss in the SCF understate the extent of their losses. For example, in tax year 2006, total net income from partnerships with positive income amounted to \$1103.2 billion in the SCF compared to \$504 billion in the IRS, but total net losses from partnerships that incurred a net loss amounted to only \$21.5 billion in the SCF compared to \$146.9 billion in the IRS. This finding highlights that the SCF's overstatement of business income relative to the IRS is attributable to problems and inconsistencies in the distribution of business income reported in the SCF.

Next, we compare business income reported on individual tax forms when respondents are grouped by their annual gross income (AGI). Figure 6 shows this comparison for income from sole proprietorships (Schedule C), while Figure 7 shows this comparison for income from partnerships, S corporations, rents, royalties, estates, and trusts (Schedule E). We know from the previous section that SCF overstates Schedule C income relative to the IRS. Figure 6 provides the comparison between the SCF and the IRS when respondents are grouped by their AGI. It shows that SCF understates Schedule C income earned by those with low AGI

but severely overstates it for families with high AGI. In contrast, according to Figure 7, Schedule E is overstated in the SCF across all AGI subgroups but more so for low-AGI families. Similar patterns can be observed for S corporation income reported on business tax forms, as shown in Figure 8.

4.3 Non-representativeness and Measurement Error

We now investigate the reasons behind the overestimation of business income for pass-through businesses and the underestimation of business income for C corporations in the SCF relative to the IRS over time. We focus on two potential reasons: i) misreporting of business income by business owners interviewed in the SCF data and ii) the non-representativeness of business owners in the SCF data. In order to understand these candidate problems, we document the number of business returns filed. The hypothesis is that if the number of returns and the distribution of returns across business income percentile groups are similar in both the SCF and the IRS, then the differences of business incomes in the SCF relative to the IRS are more likely to be due only to the misreporting of business income in the SCF. If, however, a comparison of the number of returns reveals a significant difference between the two datasets, this would suggest non-representativeness issues in addition to the misreporting problems.

Figure 9 plots the number of business returns in the IRS and SCF over time for sole proprietorships, partnerships, S corporations, and C corporations. Panel A shows a clear upward trend in the number of sole proprietorship returns in the IRS data, but the number of sole proprietorship returns has been flat for the last two decades according to the SCF. On average, the number of sole proprietorships in the SCF represents only around 35 percent of the aggregate sample in the IRS. When we analyze this result together with the comparison of business income of sole proprietorships in the SCF and IRS in Figure 3, we see that even if the number of returns is significantly lower in the SCF, business income is still significantly overstated. This finding implies that a relatively small number of business owners are clearly overstating their business incomes from sole proprietorships. Panel B shows that the number of partnership returns in the SCF is closer to its counterpart in the IRS. However, as we discussed in relation to Figure 3, the business income of partnerships is larger in the SCF by 305.4 percent on average when compared to the IRS, as documented in Table 1. Hence, this evidence supports the misreporting of business income in the SCF. Panel C demonstrates that the SCF also underrepresents S corporations. In particular, the number of S corporation returns flattens after 2000 and even decreases after 2006 in the SCF, but it keeps increasing in the IRS. In 2012, the number of S corporations in the SCF is less than half of the aggregate sample in the IRS. Recall that the business income of S corporations is larger in the SCF than in the IRS, as shown in Figure 3. Therefore, a small sample of S corporations in the SCF overstate their business income to the

extent that aggregate business income in the SCF exceeds that of the IRS. These results imply that weights in the SCF are low, the severity of which varies across business legal structure. Finally, Panel D shows that while the SCF captures the declining trend in the number of C corporation returns after 1997, it significantly understates the level of the number of C corporations. As we know from Figure 3 that total business income of C corporations is also understated in the SCF. These results suggest that the underrepresentation of the sample of C corporations in the SCF contributes to the underestimation of business income.

We now compare the number of returns for proprietorships, partnerships, S corporations, and C corporations with net income or net loss in the SCF and the IRS. Figure 10 shows that the number of businesses with net losses is clearly underrepresented in the SCF when compared to the IRS. This underrepresentativeness is a more severe problem for sole proprietorships, S corporations and C corporations. For example, in 2012, the number of sole proprietorships with net losses is only one-fourth, the number of returns with net losses of S corporations and C corporations are only one-third of their respective samples in the IRS. While the number of partnerships with net losses exhibits a pattern over time that is similar to its counterpart in the IRS, we still find sizeable differences between the two. When these results are interpreted together with the previous result that business income for all types of businesses with net losses is also underestimated in the SCF compared with the IRS, as seen in Figure 4, we conclude that the underrepresentation of the sample of businesses with net losses contributes to the underestimation of business income with net losses for all types of businesses.

Figure 11 focuses on businesses with net income. The number of sole proprietorships with net income in the SCF is almost constant, and it only represents between one-fourth and one-third of the aggregate number in the IRS. However, as we have documented in Panel A of Figure 5, the income of these businesses is mostly overstated in the SCF. This finding suggests misreporting of business income by the owners of such businesses in the SCF sample. For partnerships with net income, we find that even if the number of these businesses is similar in the SCF and the IRS until 2006, their business income is overstated in the SCF. This finding also support the presence of misreporting of business income in the SCF. Next, the number of S corporations with net income is similar in the SCF and IRS until 2000, but the SCF undersamples such businesses afterward. The income of these businesses, however, is overstated throughout this period in the SCF. Hence, we find that misreporting of business income is a more promising candidate to explain the overestimation of business income for pass-through businesses with net income. For C corporations, both misreporting and non-representativeness contribute to the difference between SCF and IRS. In year 2000 for example, we find that the total number of C corporations with net income in the SCF is close to its counterpart in the IRS, but total business income of such businesses in the SCF is only half of the its level in the IRS. This also suggests misreporting of business income for C corporations is a candidate explanation.

Moreover, after year 2003, the number of these businesses in the SCF is also much lower in the SCF relative to the IRS. Hence, underrepresentation of C corporations with net income during these years in the SCF sample also contributes to the underestimation of business income.

The degree of understatement of the number of returns in the SCF relative to the IRS is non-uniform across businesses with net income or net loss as well as across legal structure, suggesting that weights are not wrong in a consistent manner. If weights were wrong in a consistent manner, then we could simply scale up the number by some constant factor to get the number of returns to match.

To further examine this issue, we rank individual returns according to AGI and classify each return into an AGI subgroup. For each subgroup, we sum the number of Form 1040 Schedule C returns filed with the IRS and compare it with the number of SCF respondents who report owning a sole proprietorship and filing a tax return. Figure 12 reports that the number of sole proprietorship returns in the SCF is significantly understated across all subgroups and over time. More important, the understatement is more severe for lower AGI subgroups. The number of returns associated with the bottom 25 percent of AGI is low by a factor of 7 on average but is only low by a factor of 2 for the top 1 percent of AGI.

One must exercise caution in interpreting results for C corporations. The type of C corporations represented in the SCF is difficult to infer due to the lack of information on whether a C corporation is publicly-traded or closely-held, or the number of shareholders of the business. Furthermore, it is unclear whether respondents who are affiliated with large corporations are knowledgeable about business financial data such as income and valuations. It may also be the case that respondents are unable to distinguish between S corporations and C corporations and thus misreport the legal structure of the business. Nonetheless, regardless of the assumptions one makes about the type of corporations represented in the SCF sample, the estimates we find raise serious concerns. Under the assumption that the SCF predominantly captures closely-held C corporations, net income in the SCF should be understated by a factor much larger than 2 to 5 since only a small fraction of total net income of C corporations are attributable to closely-held businesses. Under the assumption the SCF successfully captures both publicly-traded and closely-held corporations, then the total number of returns and aggregate net income of C corporations is significantly understated relative to aggregate IRS data. Moreover, similar to the case of pass-through businesses, dividend yields are also much larger in the SCF compared to the S&P 500.

To summarize, in this section, we have documented evidence that both non-representativeness of the number of returns and misreporting of business income contribute to the overestimation of business income in the SCF. We then further investigate the reason for misreporting in the SCF and highlight two sources. At the end of the survey, SCF asks respondents i) how frequently they check some type of document while answering the questions during the interview and ii) the type of the document that they checked, if any,

during the interview. Using the answers to these questions, we calculate the fraction of all respondents and business owners who checked various documents during the interview. Table 3 documents these results for the SCF 2016.⁸ Among all respondents, 7.2 percent frequently referred to their income tax returns (as opposed to “sometimes,” “rarely,” or “never”). If we ask how many households frequently reference all relevant documents, then the number drops to about 0.6 percent.⁹ Among respondents who own at least one business, only 1.1 percent checked all necessary documents, 13.2 percent frequently referred to their income tax returns, and 24.1 percent rarely checked their income tax returns. These results suggest that respondents do not refer to their relevant documents while answering the questions in the interview, which is our first explanation for why misreporting is so prevalent in the SCF. We suspect that another reason for misreporting in the SCF is due to the framing of questions that can lead to confusion. Given that businesses with net losses report very small amount of net loss relative to the IRS, respondents may not know that a net loss implies a negative value for net income and may have simply reported zero income instead.

5 Business Returns

We examine survey responses to a question about the value of ongoing businesses, a statistic that is relevant for computing the dispersion in household wealth but difficult to measure given that businesses invest heavily in intangible assets. In this section, we focus on SCF dividend yields, which is defined as the ratio of business income to business net worth of actively managed businesses. In our analysis, we first restrict the sample of these businesses to those with positive net worth. We then exclude businesses with net worth less than the bottom 1 percentile of the net worth distribution, conditional on having positive net worth.

We compute for the dividend yields associated with an index of all businesses, including pass-throughs and C corporations. Figure 13 shows that value-weighted dividend yield from 1989 to 2015 fluctuated between 14 and 31 percent.¹⁰ However, these estimates are significantly higher than any estimate of mean U.S. corporate dividend yields. When we calculate the value-weighted average dividend yield for the businesses in S&P 500 company data, we find an average dividend yield of 2.3 percent. We also compute for average dividend yield when businesses are equally-weighted. The comparison of value-weighted and equally-weighted dividend yields is informative about the tails of the distribution of pass-through business returns. Figure 14 shows the average dividend yields of sole proprietorships, partnerships, S corporations, C corporations, and all

⁸Similar results also hold for other surveys of the SCF.

⁹Here, relevant documents include income tax returns, pension documents, account statements, investment and business records, and loan documents.

¹⁰The magnitude of these returns are comparable to those found by Moskowitz and Vissing-Jorgensen (2002) and Kartashova (2014). Moskowitz and Vissing-Jorgensen (2002) and Kartashova (2014) incorporate capital gains and net business debt owed to owners in the calculation of returns. Given that we obtain similar results, exclusion of these has minor effects on dividend yields.

businesses over time, where business income reported on business tax forms is used. Importantly, we find implausibly high estimated returns. For all businesses, the average dividend yield fluctuates between 66 and 171 percent over the last two decades. Similar patterns are also present when we calculate the average dividend yields across different types of businesses. This result emphasizes two important points. First, it is significantly higher than the equally-weighted average dividend yield for businesses in the S&P500 of 1.9 percent. Notice that this large difference between the average dividend yield in the SCF and the S&P 500 is present even if we have excluded businesses with very small business net worth in the SCF. Second, the significant disparity between value-weighted and equally-weighted dividend yields in the SCF indicates the presence of many businesses with unreasonably high returns.

To understand this, we calculate the distribution of dividend yields for all businesses in the sample described above over time. We find two important results. First, the unreasonably high values of average dividend yields are not driven by a few businesses with very large dividend yields. This is because, as Figure 15 shows, more than half of all businesses have dividend yields higher than 15 percent, and one-fourth have dividend yields that exceed 50 percent. Second, the distribution of dividend yields experiences large leftward or rightward shifts over time. For example, between 1991 and 1997, the median value of dividend yields increases from 13 percent to 25 percent, and it decreases again to 15 percent in 2012. These large swings in the dividend yield distribution should caution theorists who calibrate their models to match cross-sectional moments on business returns for some specific year. Finally, in Table 4, we show that the SCF distribution of dividend yields is largely different from its counterpart obtained from the S&P 500. For example, in 2003, more than a quarter of all pass-throughs (C corporations) had dividend yields higher than 54 percent (13 percent) but even the firm at the 90th percentile of the S&P 500 distribution had dividend yields of only 6.05 percent.

Overall, using dividend yields for the S&P 500 composite as a reference point, SCF yields are around 10 times larger for value-weighted and 50 times larger for equally-weighted returns on average, and far more dispersed in the cross section. Since SCF incomes are overstated by an average factor of 2.2, as we have shown in the previous section, we deduce that reported valuations have to be understated by a factor of about 5 to 20 to rationalize a dividend yield comparable to firms in the S&P 500 composite.

6 Robustness

In this section, we first provide various attempts to further align the SCF with the IRS data, and show that the problems of the SCF remain unresolved. Next, we compare the SCF and IRS estimates with the KFS, SIPP, PSID, and PSED and document other measurement issues within these surveys.

6.1 Adjustments to IRS or SCF

6.1.1 Inconsistencies between answers to similar questions

The SCF asks respondents about business income from sole proprietorships using two different questions. One question asks respondents to report business income from a sole proprietorship or a farm lifted from Form 1040 lines 12 and 18, while the second question asks respondents to report business income from a sole proprietorship lifted from Form 1040 Schedule C line 31. By design, Form 1040 line 12 must be equal to Schedule C line 31. Hence, income reported from Form 1040 lines 12 and 18 must be close to income reported from Schedule C line 31 given that the only difference between the two is the addition of income generated from a farm (on Form 1040 line 18), which is a small amount according to the IRS data. As documented in Figure 17, across answers to both questions, we find large differences that cannot be explained by farm income alone.

We also document the existence of respondents who report non-zero annual net income from a sole proprietorship or farm (Form 1040 lines 12 and 18) yet do not report owning a sole proprietorship. Given this problem in the data, one can calculate the business income of sole proprietorships in two ways: either i) assuming that business ownership information is correct by excluding the reported income of those who do not report business ownership or ii) assuming that the reported income is correct and considering non-zero income as evidence of ownership. In Figure 18, we show the business income and number of returns of sole proprietorships calculated under these assumptions separately. In Panel A, we see that excluding the business income of those who report not owning a sole proprietorship leads to a better match between the total sole proprietorship income in the SCF and the IRS. However, Panel B shows that under this assumption, the number of sole proprietorships is severely understated. Under the alternate assumption that non-zero business income is evidence of business ownership, notice that while the number of returns is understated to a lesser degree in Panel B, we would observe a significant overstatement in business income in Panel A. The conclusion here is that any attempt to reconcile this inconsistency within the survey results in a high discrepancy in either total business income or total business returns for sole proprietorships.

6.1.2 Adjusting for misreporting in the IRS

If households underreport incomes to the IRS but accurately report income to surveys, we would find an overstatement of incomes, as is the case for the SCF. We use data on adjustments for tax misreporting on income tax documents for proprietorships and partnerships published by the U.S. Bureau of Economic Analysis. We add these yearly adjustments to the sum of sole proprietorship and partnership income in the IRS and compare them with estimates of business income for these businesses in the SCF. Figure 16

shows the result of this comparison. We find that the sum of aggregated and adjusted proprietorship and partnership income in the SCF and the IRS are close to each other. One might argue that this adjustment resolves the differences between the SCF and IRS, but this is not the case because for two reasons.

First, Johns and Slemrod (2010) use data from individual income tax reporting noncompliance in the U.S. federal income tax for tax year 2001 and separately report the percentage of the unreported amount of sole proprietorships as well as partnerships/S corporations, estate, and trust income. In particular, they show that 57 percent of the true sole proprietorship income and 18 percent of partnerships/S corporations, estate and trust income are not reported. We use these numbers to adjust the IRS values in tax year 2001 to generate the income adjusted for misreporting of sole proprietorships and partnerships separately.¹¹ The adjusted sum of sole proprietorship and partnership income from the IRS is \$704 billion, while this sum amounts to \$754 billion in the SCF. However, just looking at the sum of adjusted sole proprietorship and partnership income in Figure 16 is misleading. This is because the adjusted sole proprietorship income in the IRS is \$559 billion, but the SCF estimate of sole proprietorship income is \$374 billion; meanwhile, adjusted partnership income in the IRS is \$145 billion, but the SCF estimate of partnership income is \$380 billion. Hence, the SCF understates sole proprietorship income but overstates partnership income relative to tax misreporting-adjusted IRS data. As a result, total sole proprietorship and partnership income would appear to be similar in the SCF and misreporting-adjusted IRS data but is in fact merely a result of offsetting errors.

Second, if there were no misreporting on the part of respondents, we would expect sole proprietorship and partnership income to be lower in the SCF relative to the adjusted IRS data because the SCF undersamples business owners. However, Figure 16 demonstrates that this is not the case. Therefore, if anything, adjusting for misreporting merely alleviates but does not eliminate the overstatement of business income in the SCF relative to aggregate data.

6.1.3 Adjusting for partnerships owned by individual partners

In practice, partnerships can be owned by individuals, other partnerships, and other types of entities. However, the SCF only surveys individuals, and thus it is only able to capture partnerships owned by individuals. Cooper et. al (2016) use administrative tax data from 2011 to analyze the owners of pass-through businesses and calculate the amount of tax they pay. They show that 31.5 percent of total partnership income is generated by individual partners. When we adjust total partnership income in the IRS with this number, the difference between the SCF and IRS becomes even larger. The total partnership income generated by individual partners in 2012 is \$123.55 billion in the IRS, whereas the total partnership income is \$597.74 billion in the SCF 2013, which provides data for tax year 2012.

¹¹The percentage of misreporting of S corporations, estate, and trust income is assumed to be negligible.

6.2 Other survey data

6.2.1 KFS

Gurley-Calvez et al. (2016) compare responses about receipts, expenses, and profits for businesses in the KFS with matched tax forms. They show that the firms in the survey overstate receipts and overstate expenses by even more, implying that the firms understate profits across the distribution. Hence, these findings are for the most part in contrast to the SCF and IRS comparison, as the SCF overstates business income.

6.2.2 SIPP and PSID

Comparisons can also be made between the IRS and other widely-used survey data such as the SIPP and PSID. Both SIPP and PSID contain information about business ownership and income. However, the PSID does not distinguish between sole proprietorships and partnerships (among unincorporated businesses) or between S corporations and other corporations (among incorporated businesses). Moreover, the PSID provides business income information only if the business is unincorporated. Thus, we compare the business income of unincorporated businesses in the PSID and SIPP to the SCF and IRS. Figure 19 shows that while the SCF largely overstates the unincorporated businesses income and the SIPP largely understates it, the PSID is more closely tracks the level of unincorporated business income until the early 2000s where large deviations begin to occur.¹²

In order to understand possible non-representativeness and measurement error problems, it would be insightful to compare the number of returns between the IRS and the survey data. However, the PSID does not provide business ownership share information, which prevents the calculation of number of returns for unincorporated businesses. To get around this problem, we instead compare the number of owners of unincorporated businesses in the PSID, SIPP, SCF, and IRS. Figure 20 shows that all of the survey data significantly understate the number of owners and they all fail to capture the increase in the number of owners after 2000. The close match of unincorporated business income in the PSID and IRS before the 2000s but understatement of the number of owners suggests that business income is overstated by respondents during this period. A similar problem is also present in the SCF because it overestimates income but undersamples the owners. For the SIPP, we conclude that the underrepresentation of owners contributes to the underestimation of income.¹³

One may think that the reason why survey data largely understates the number of business owners of

¹²Before 2004, the SIPP does not provide information about an individual's share of business income from an unincorporated business. Instead, it only contains information about the total income of the business, which without ownership share information is not enough information to calculate the total business income of unincorporated businesses.

¹³We also compare the number of returns for unincorporated businesses between the SIPP and the IRS. We find that the SIPP only represents less than half of the aggregate sample in the IRS. This also suggests that underrepresentation of the sample contributes to the underestimation of business income for sole proprietorships and partnerships in the SIPP.

unincorporated businesses relative to the IRS is that survey data only count individual owners of partnerships but do not count other legal entities (such as corporations) who may also own partnerships. In order to address this concern, we make an adjustment to the IRS. Cooper et. al (2016) document that 73.4 percent of all partnerships are owned by individuals in 2011. We use this number and adjust the IRS level down uniformly across years and show in Figure 20 that this adjustment alone cannot explain the large difference between the IRS and the survey data.

Unlike the SCF, the SIPP and PSID do not ask respondents to refer to specific lines on their tax forms. Thus, it must be emphasized that to begin with, the SIPP and PSID are not directly comparable with the IRS. To demonstrate this, we construct total income in the SIPP by selecting income components that match definition-wise with the components of total income. Here, we include wages and salaries, self-employment earnings, interest income, property or rental income, dividend income, unemployment compensation, social security benefits, alimony, and pensions or annuities. We find that the SIPP understates total income by around 17 percent on average relative to the IRS, implying that the SIPP does not even closely match the total income. The PSID, on the other hand, recently began collecting information about whether the respondents refer to some individual documents while answering the questions in the PSID. In 2015, for example, only 8 percent of the respondents checked their income tax forms during the interview. This is possibly another potential source of mismeasurement in the PSID.

We also calculate the value-weighted average dividend yields in the SIPP, following similar steps above to our calculations from the SCF.¹⁴ Table 5 shows the mean and the distribution of dividend yields for the unincorporated businesses in 2011. We find that the median dividend yield is 1.8 percent for all unincorporated businesses, 2 percent for sole proprietorships, and 0.2 percent for partnerships. These values are significantly smaller from what we have documented from the SCF.

6.2.3 PSED

The PSED provides information about business start-ups using a nationally representative sample. An initial screening survey in the fall of 2005 included 1,214 entrepreneurs. These respondents were asked questions that are relevant for our purposes, such as annual household income, type of business (i.e., sole proprietorship, general or limited partnership, limited liability corporation, S corporation, or general corporation), whether they filed a federal income tax return, and their calculated profits and losses from the business. However, the PSED reflects a measurement issue. For example, among the 1,214 entrepreneurs, only 115 (i.e. 9 percent) responded to the question that asks about their calculated profits and losses for tax year 2006.¹⁵ As a result,

¹⁴PSID does not provide information about the ownership share of the businesses, which prevents us from calculating dividend yields.

¹⁵The fraction of entrepreneurs responding to this question is similar for other years.

the aggregate profit and loss generated from a sole proprietorship is only around \$283,000 in 2006. For this reason, these data are deficient along this dimension.

7 Conclusion

In this paper, we study the reliability of survey data for research on sole proprietorships, partnerships, S corporations, and C corporations. We analyze data for all surveys that ask questions about these businesses and document problems arising from non-representative samples and measurement errors. We document two main sources of measurement error. First, some errors result from respondents not referring to relevant tax documents. For example, among all respondents in the SCF, 7.2 percent frequently referred to their income tax returns. If we ask how many households frequently reference all relevant documents, then the number drops to about 0.6 percent. Among respondents who own at least one business, only 1.1 percent checked all necessary documents, 13.2 percent frequently referred to their income tax returns, and 24.1 percent rarely checked their income tax returns. Second, errors also arise from the framing of questions that can lead to confusion among respondents. These problems range from respondents not knowing that a sole proprietorship has to file a Schedule C with IRS Form 1040 to not knowing that a net loss implies a negative value for net income.

We document that while total adjusted gross income in the survey data matches comparable administrative data well, business net income, receipts, distributions of business net income, number of returns, and business valuations do not match. We show that total pass-through business income is overstated in the SCF by a factor of 2 to 3 depending on the year, whereas it is understated in the SIPP and KFS. In the case of C corporations, the total business income is understated by a factor of one-half on average in the SCF. For number of returns, we find much lower returns, indicating that the samples are not representative. For example, in the case of the SCF, returns for pass-through businesses are low by a factor of 2 in the late 1980s and by a factor of 3 more recently, while returns for C corporations is lower than IRS returns by a factor of 2. For business valuations, using dividend yields for the S&P 500 composite as a reference point, SCF yields for pass-through businesses are around 10 times larger for value-weighted and 50 times larger for equally-weighted returns on average, and far more dispersed in the cross section. Since SCF incomes are overstated by an average factor of 2.2, we deduce that reported valuations have to be understated by a factor of about 5 to 20 to rationalize a dividend yield comparable to firms in the S&P 500 composite. For C corporations, reported valuations can be compared to aggregated U.S. flow of funds data. We find the SCF estimate of the total value of C corporations to be around 7 percent of that reported in the U.S. flow of funds accounts. In contrast, we find small business returns that are comparable to the S&P500 data from

the SIPP.

We show that problems exist with the survey data even if we adjust for tax misreporting. If households underreport incomes to the IRS but accurately report income to the surveyor, we would find an overstatement of incomes, as is the case for the SCF. We use tax audit data to correct the administrative data but still find a mismatch with the survey data. Other adjustments to alleviate the mismeasurement, such as correcting for within-survey inconsistencies in the SCF regarding business ownership and income, and for the fact that the SCF only surveys partners who are individuals, are tried but without success.

Data

The main sources of data reported in the main text are as follows:

- Survey of Consumer Finances of the Board of Governors of the Federal Reserve System
- Survey of Income and Program Participation of the U.S. Census Bureau in the Department of Commerce
- Panel Study of Income Dynamics of the Survey Research Center, Institute for Social Research, University of Michigan - Ann Arbor
- Panel Study of Entrepreneurial Dynamics of the Survey Research Center, Institute for Social Research, University of Michigan - Ann Arbor
- Kauffman Firm Survey of the Kauffman Foundation
- Statistics of Income of the Internal Revenue Service
- National income and product accounts and fixed assets of the Bureau of Economic Analysis in the Department of Commerce

References

- [1] Johns, A., & Slemrod, J. (2010). The distribution of income tax noncompliance. *National Tax Journal*, 63(3): 397–418.
- [2] Bhandari, A., & McGrattan, E. R. (2018). Sweat equity in U.S. private business. Federal Reserve Bank of Minneapolis, Staff Report 560.
- [3] Bohacek, R., & Zubricky, J. (2012). A flat tax Reform in an economy with occupational choice and financial frictions. *Economic Journal*, 122(565): 1313–45.
- [4] Buera, F.J. (2009). A dynamic model of entrepreneurship with borrowing constraints: theory and evidence. *Annals of Finance*, vol.5: 443–464.
- [5] Cagetti, M., & De Nardi, M. (2006). Entrepreneurship, frictions, and wealth. *Journal of Political Economy*, 114(5): 835–870.
- [6] Cooper, M., McClelland, J., Pearce, J., Prisinzano, 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.
- [7] De Nardi, M., Doctor, P., & Krane, 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.
- [8] Gentry, W. M., & Hubbard, R.G. (2004). Entrepreneurship and household saving. *Advances in Economic Analysis and Policy*, 4(1), Article 1.
- [9] 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.
- [10] Hamilton, B. (2000). Does entrepreneurship pay? An empirical analysis of the returns to self-employment. *Journal of Political Economy*, 108(3): 604–631.
- [11] Hurst, E., & Pugsley, B.W. (2011). What do small businesses do?. *Brookings Papers on Economic Activity*, 43(2): 73–142.
- [12] Hurst, E., & Pugsley, B.W. (2017). Wealth, tastes, and entrepreneurial choice, in measuring entrepreneurial businesses: Current Knowledge and Challenges, eds., J. Haltiwanger, E. Hurst, J. Miranda, and A. Schoar (Chicago: National Bureau of Economic Research).

- [13] Kartashova, K. (2014). Private equity premium puzzle revisited. *American Economic Review*, 104(10), 3297–3334.
- [14] Kitao, S. (2008). Entrepreneurship, taxation, and capital investment. *Review of Economic Dynamics*, 11(1): 44–69.
- [15] 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.
- [16] 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.
- [17] McGrattan, E. R., & Prescott, E.C. 2010b. Technology capital and the U.S. current account, *American Economic Review*, 100(4): 1493–1522.
- [18] Meh, C. (2005), Entrepreneurship, wealth inequality, and taxation. *Review of Economic Dynamics*, 8(3): 688–719.
- [19] Moskowitz, T. J., & Vissing-Jørgensen, A. (2002). The returns to entrepreneurial investment: A private equity premium puzzle?. *American Economic Review*, 92(4), 745–778.
- [20] Scheuer, F. (2013). Entrepreneurial taxation with endogenous entry. NBER Working Paper 19235.
- [21] Smith, M., Yagan, D., Zidar, O., & Zwick, E. (2017). Capitalists in the twenty-first century. Working Paper, U.S. Department of the Treasury.
- [22] Quadrini, V. (2000). Entrepreneurship, saving, and social mobility. *Review of Economic Dynamics*, 3(1): 1–40.
- [23] 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.

Table 1: Percentage Errors of Total Income and Business Income in the SCF relative to IRS

Tax Year	Total income	Business income			
		Sole prop.	Partnership	S corp.	C corp.
1988	5.5	53.6	168.6	107.1	-
1991	8.6	68.1	554.2	249.8	-37.0
1994	4.2	3.4	889.1	158.6	-58.2
1997	-1.6	68.4	106.7	145.6	-61.7
2000	-0.2	55.6	219.0	146.7	-12.2
2003	7.9	27.1	239.8	205.4	-66.7
2006	0.6	34.9	202.9	93.8	-76.3
2009	0.5	24.1	316.1	116.3	-61.2
2012	-3.0	-9.4	52.4	10.8	-77.0
2015	3.7	-10.6	-	-	-
Average	2.6	31.5	305.4	137.1	-56.3

Note: This table reports the percentage errors of SCF reported total income and business income when compared with their IRS counterpart. Total income is defined as the sum of wages and salaries; net income from a business, profession, or farm; taxable and non-taxable interest; dividends; capital gains from the sale of capital assets and other property; net income from rental, royalty, estate, and trust; net income from partnerships and S corporations; unemployment compensation; alimony received; total pensions and annuities; total social security benefits; as well as other income. Business income refers to income reported on Form 1040 Schedule C for sole proprietorships, Form 1065 for partnerships, Form 1120S for S corporations, and Form 1120 for C corporations. Percentage error is calculated by dividing the difference between the value in the SCF and the value in the IRS by the value in the IRS and multiplying the result by 100. IRS data for partnerships, S corporations, and C corporations are available only until 2013, and C corporations data starts from 1990 because data for Form 1120 is not available for 1988 and 1989.

Table 2: Percentage Errors of Business Receipts in the SCF relative to IRS

Tax Year	Business receipts			
	Sole prop.	Partnership	S corp.	C corp.
1988	78.54	13.4	-7.8	-
1991	251.8	89.1	-52.4	-57.4
1994	15.8	203.8	6.5	94.2
1997	15.0	-26.0	2.6	-74.9
2000	-4.9	-23.7	10.5	-83.5
2003	-3.8	30.5	-26.4	-88.1
2006	-26.8	5.4	13.5	-79.8
2009	1.3	-6.8	9.9	-82.1
2012	-14.0	-12.0	15.4	-85.2
2015	-39.2	-	-	-
Average	27.5	30.4	-3.1	-57.1

Note: This table reports the percentage errors of SCF reported business receipts when compared with their IRS counterpart. Business receipts refers to gross sales reported on Form 1040 Schedule C for sole proprietorships, Form 1065 for partnerships, Form 1120S for S corporations, and Form 1120 for C corporations. Percentage error is calculated by dividing the difference between the value in the SCF and the value in the IRS by the value in the IRS and multiplying the result by 100. IRS data for partnerships, S corporations, and C corporations are available only until 2013, and C corporations data starts from 1990 because data for Form 1120 is not available for 1988 and 1989.

Table 3: Percentage of Respondents Checking Documents in SCF 2016

	Frequently checked	At least sometimes checked	At least rarely checked
All respondents, all documents	0.6	0.6	0.6
All respondents, income tax returns	7.2	13.9	16
Business owners, all documents	1.1	1.1	1.1
Business owners, income tax returns	13.2	22.4	24.1

Note: This table documents the fraction of all respondents and business owners who frequently check, at least sometimes check, or at least rarely check various documents during the 2016 SCF interview. All documents include income tax returns, pension documents, account statements, investment and business records, and loan documents. A family that actively manages a sole proprietorship, partnership, or S corporation is considered a business owner. Values are in percentages.

Table 4: Dividend Yield Distribution in SCF and S&P 500

	2003				2006			
	Pass-through	C Corp.	All	S&P 500	Pass-through	C Corp.	All	S&P 500
p25	1.00	0	0.15	0	2.00	0	1.60	0
p50	17.36	4.40	14.88	0	22.00	7.50	20.00	0
p75	53.67	12.86	50.00	2.26	80.00	36.00	73.33	2.33
p90	142.00	50.00	133.33	6.05	234.00	133.33	208.00	5.56

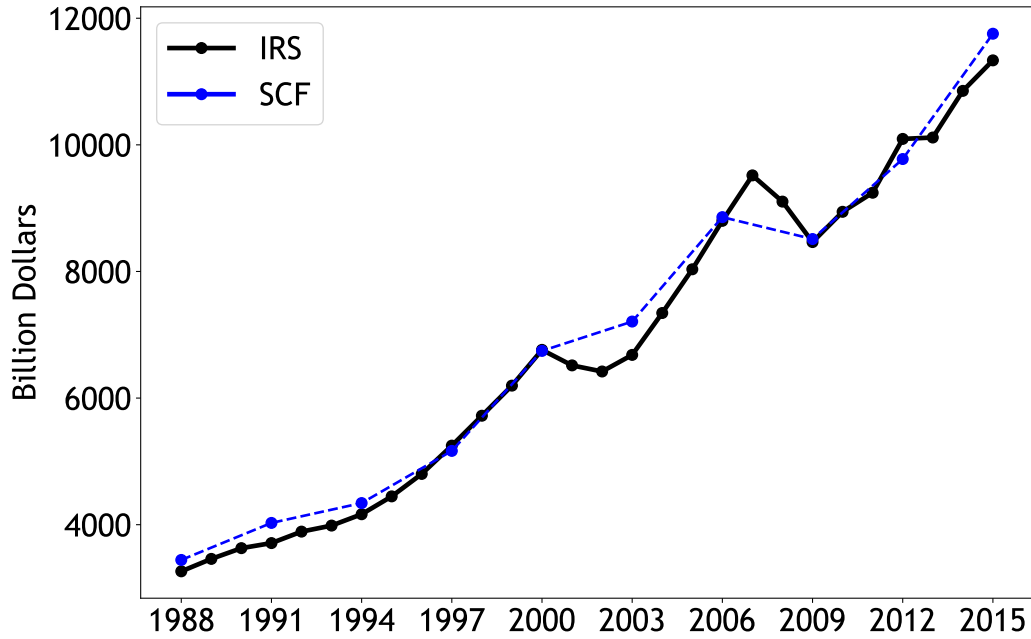
Note: This table compares various percentiles of the dividend yield distribution of pass-throughs, C corporations, and all businesses in the SCF with the S&P 500 in 2003 and 2006. The dividend yield in the SCF is calculated as the business income divided by the net worth of the business. The SCF sample 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.

Table 5: Dividend Yield Distribution in the SIPP, 2011

	Sole Proprietorships	Partnerships	Unincorporated
Value-weighted Ave.	4.1	6.2	4.3
p25	0	0	0
p50	2	0.2	1.8
p75	40	20	37.6
p90	200	240	200

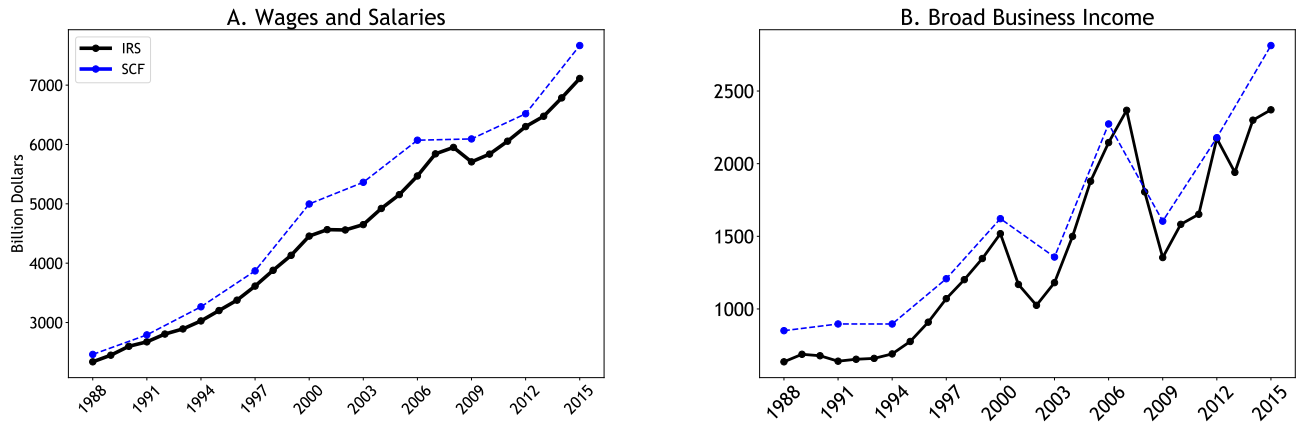
Note: This table shows the value-weighted average and the various percentiles of the dividend yield distribution of sole proprietorships, partnerships, and all unincorporated businesses in the SIPP for the year 2011. Dividend yield in the SIPP is calculated as reported business income divided by reported net worth of the business. The SIPP sample includes businesses with positive net worth and excludes the bottom 1st percentile of these businesses.

Figure 1: Total Income: IRS vs. SCF



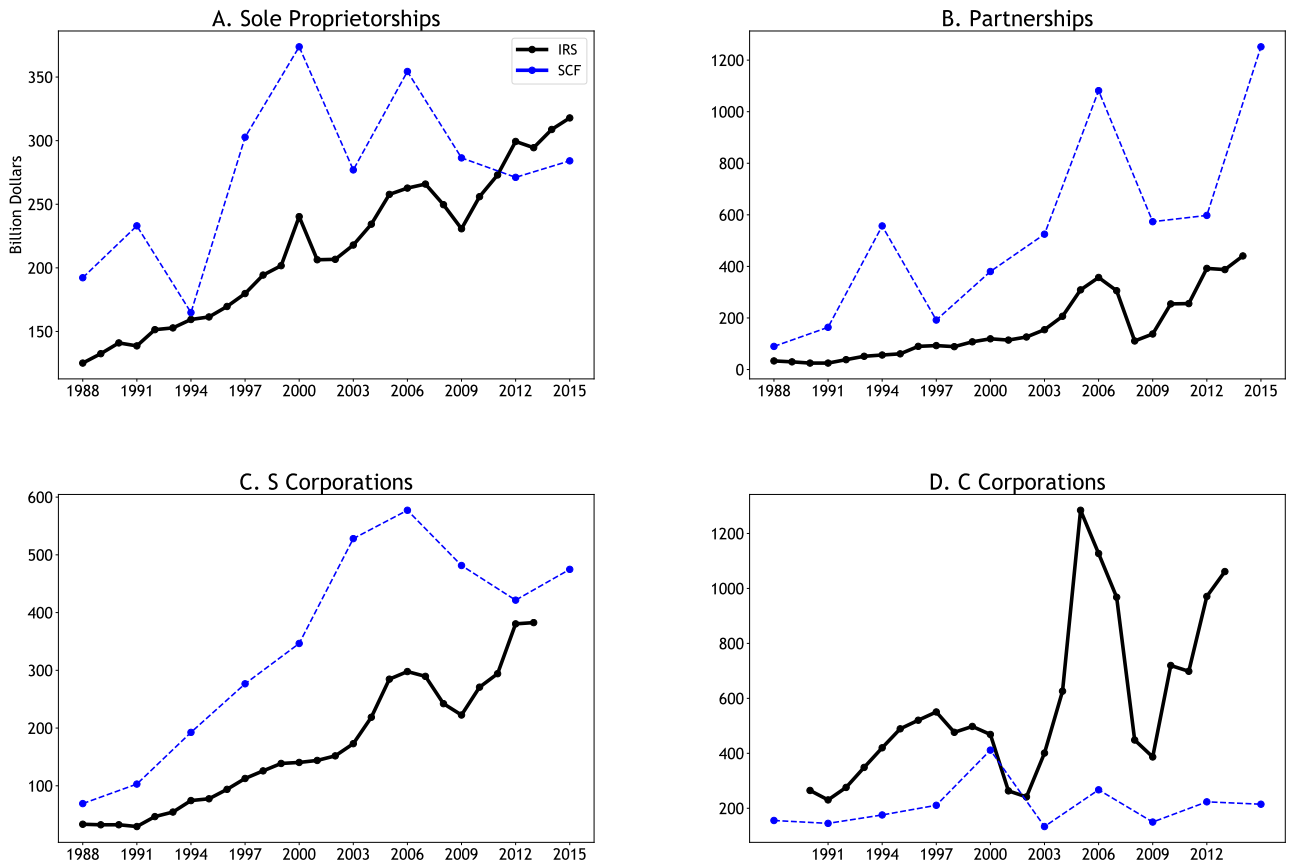
Note: This figure plots total income in the IRS and SCF. Total income is defined as the sum of wages and salaries; net income from a business, profession, or farm; taxable and non-taxable interest; dividends; capital gains from the sale of capital assets and other property; net income from rental, royalty, estate, and trust; net income from partnerships and S corporations; unemployment compensation; alimony received; total pensions and annuities; total social security benefits; as well as other income.

Figure 2: Labor and Business Income in the SCF vs IRS



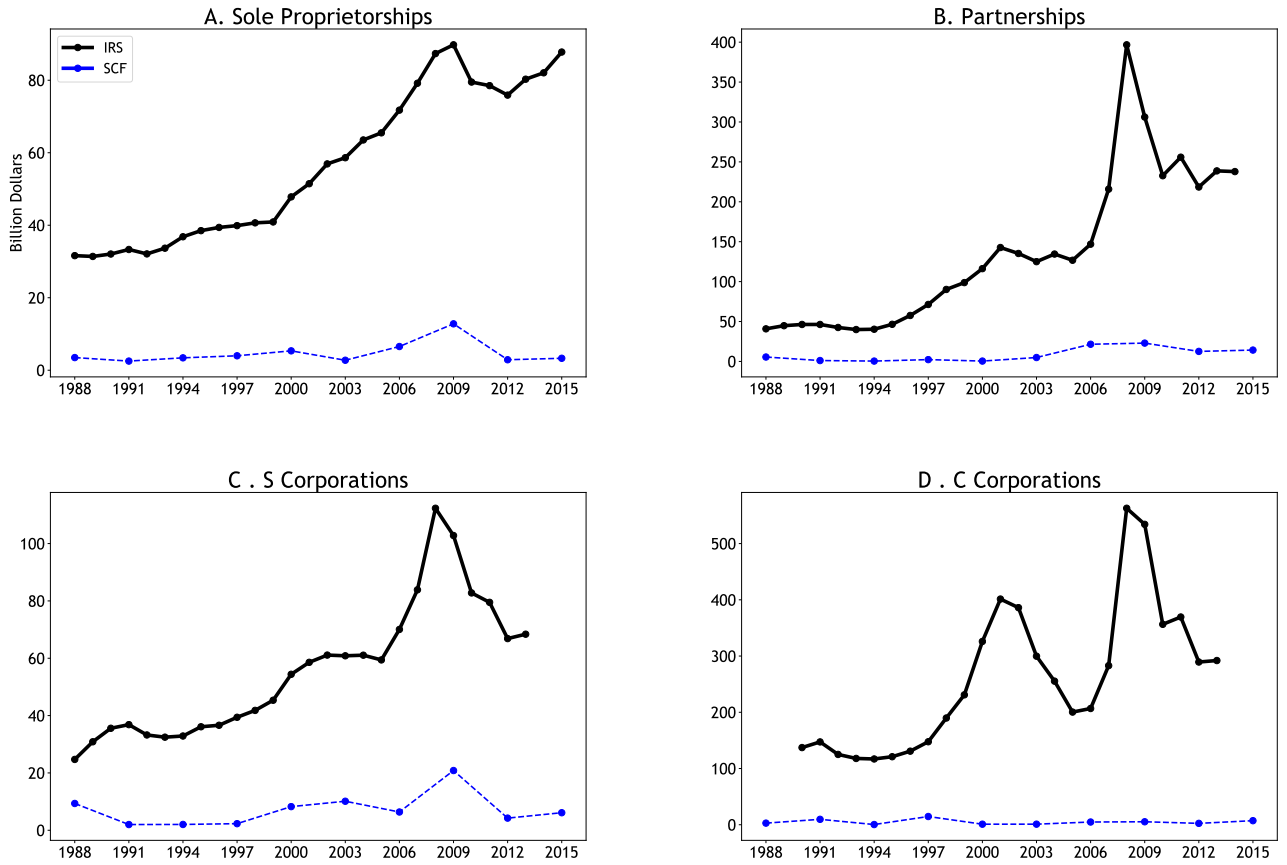
Note: This figure plots total wages and salaries (Panel A) and broad business income (Panel B) in the IRS and the SCF. Broad business income is defined as income derived from a business or profession (Form 1040 Schedule C) or farm (Form 1040 Schedule F); income from rental real estate, royalties, partnerships, S corporations, estates, trusts (Form 1040 Schedule E); taxable and non-taxable interest income (Form 1040 line 8a and 8b); and ordinary and qualified dividends (Form 1040 lines 9a and b); and income from gains from the sale of capital and other property (Form 1040 lines 13 and 14).

Figure 3: Business Income by Legal Entity: SCF vs. IRS



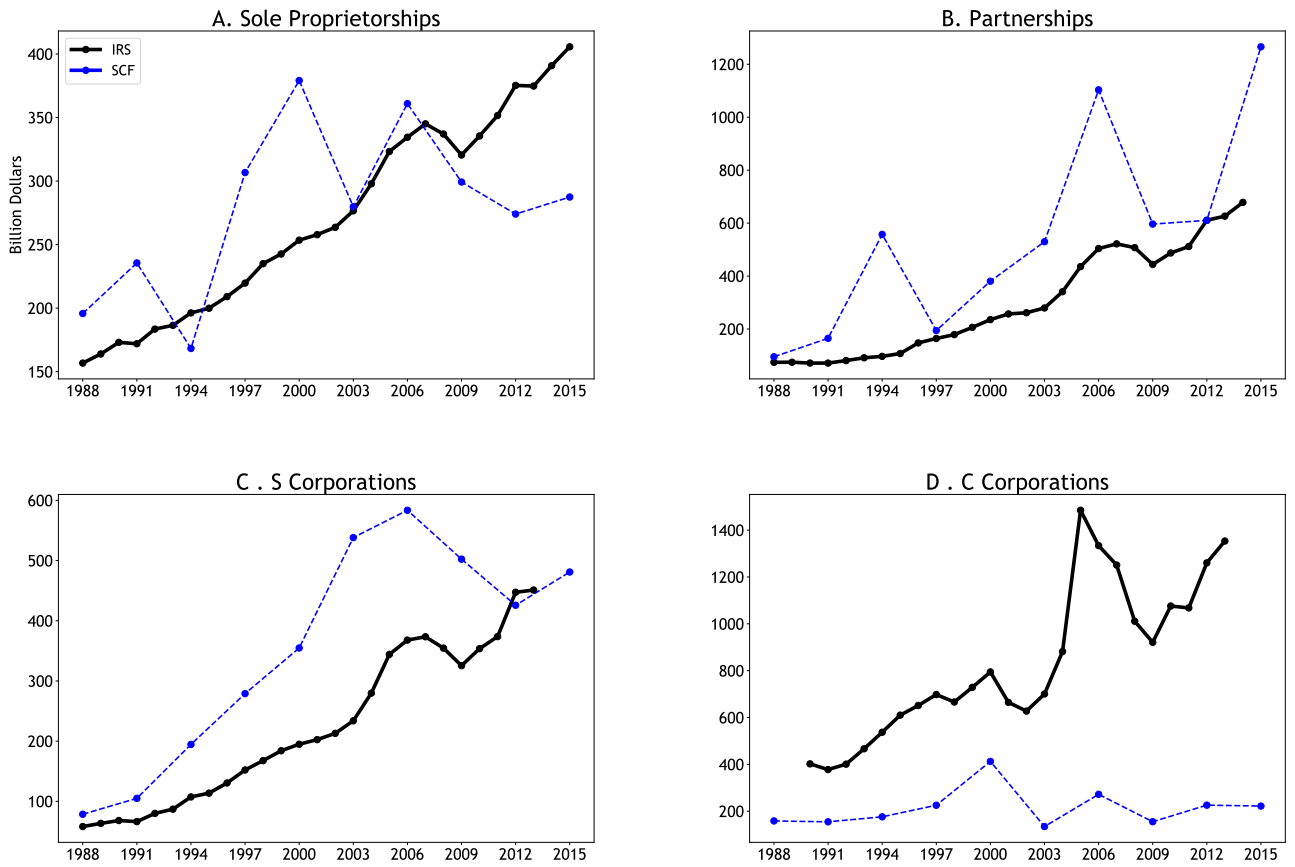
Note: This figure plots business income in the IRS and the SCF as reported on Form 1040 Schedule C for sole proprietorships, Form 1065 for partnerships, Form 1120S for S corporations, and Form 1120 for C corporations. IRS data for partnerships, S corporations, and C corporations are available only until 2013, and C corporations data starts from 1990 because data for Form 1120 is not available for 1988 and 1989.

Figure 4: Business Income of Businesses with Net Loss: SCF vs. IRS



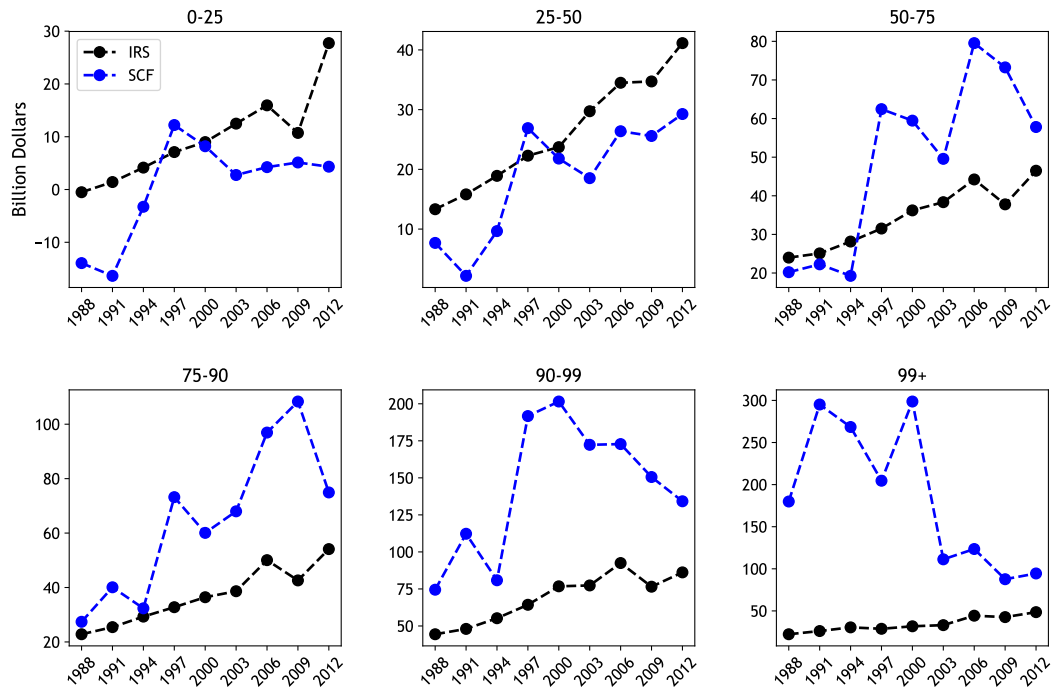
Note: This figure plots business income of businesses that report a net loss in the IRS and the SCF. Business income refers to income reported on Form 1040 Schedule C for sole proprietorships, Form 1065 for partnerships, Form 1120S for S corporations, and Form 1120 for C corporations. IRS data for partnerships, S corporations, and C corporations are available only until 2013, and C corporations data starts from 1990 because data for Form 1120 is not available for 1988 and 1989.

Figure 5: Business Income of Businesses with Net Income: SCF vs. IRS



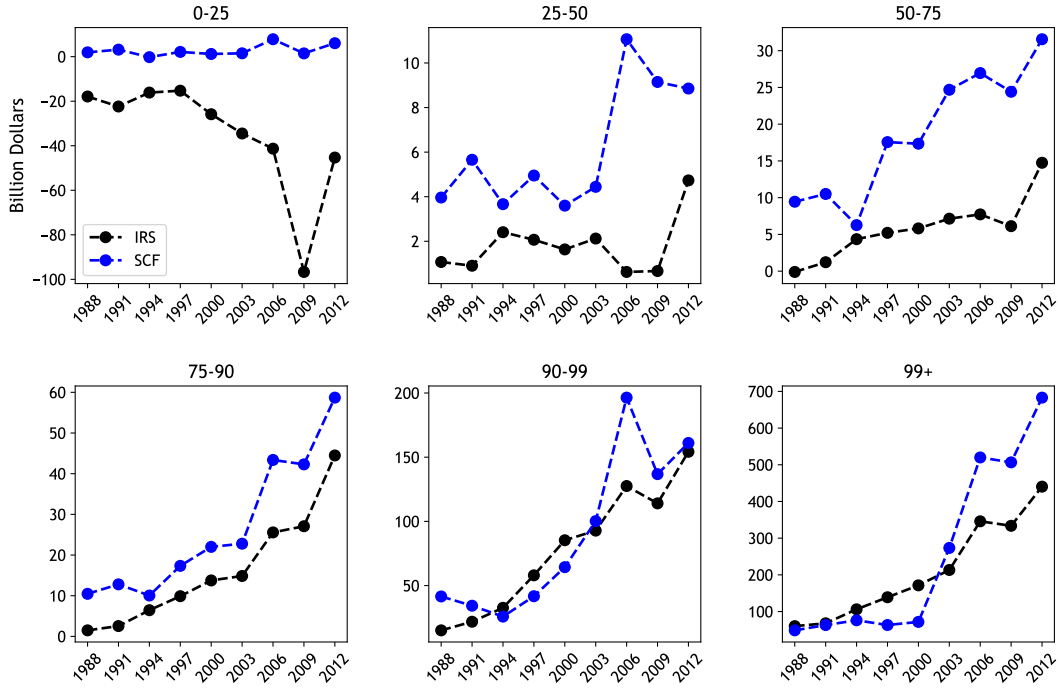
Note: This figure plots business income of businesses that report net income in the IRS and the SCF. Business income refers to income reported on Form 1040 Schedule C for sole proprietorships, Form 1065 for partnerships, Form 1120S for S corporations, and Form 1120 for C corporations. IRS data for partnerships, S corporations, and C corporations are available only until 2013, and C corporations data starts from 1990 because data for Form 1120 is not available for 1988 and 1989.

Figure 6: Distribution of Sole Proprietorship Income SCF vs. IRS



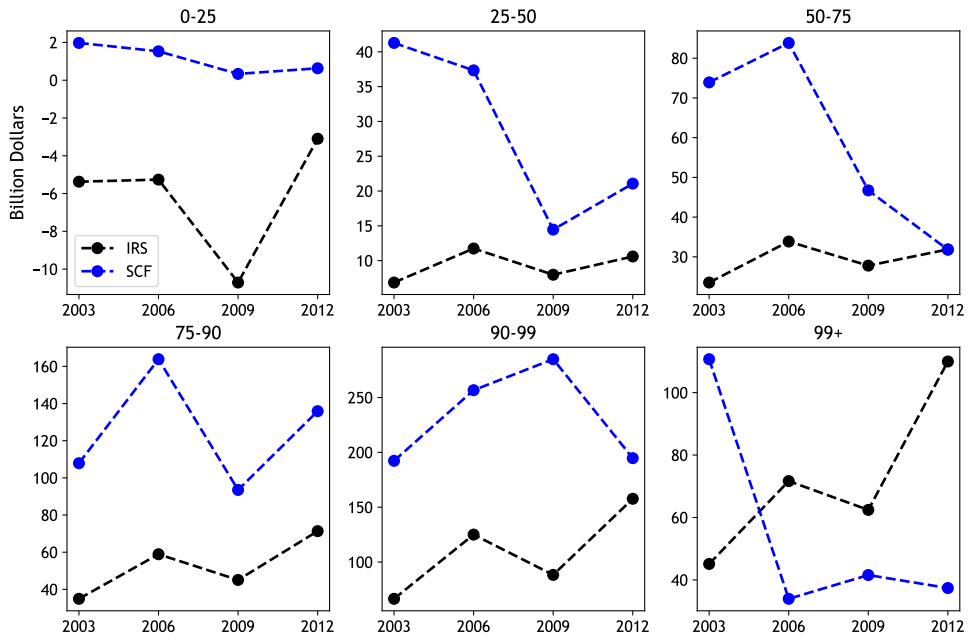
Note: This figure plots the total Schedule C (sole proprietorship) income earned by individuals grouped by their AGI. Data from individual tax returns are first sorted by AGI. Then we take the total Schedule C income reported by individuals within each pre-specified bin. AGI bins include the bottom 25 percent, 25 to 50 percent, 50 to 75 percent, 75 to 90 percent, 90 to 99 percent, and 99 to 100 percent of returns.

Figure 7: Distribution of Partnership/S Corporation Income: SCF vs. IRS



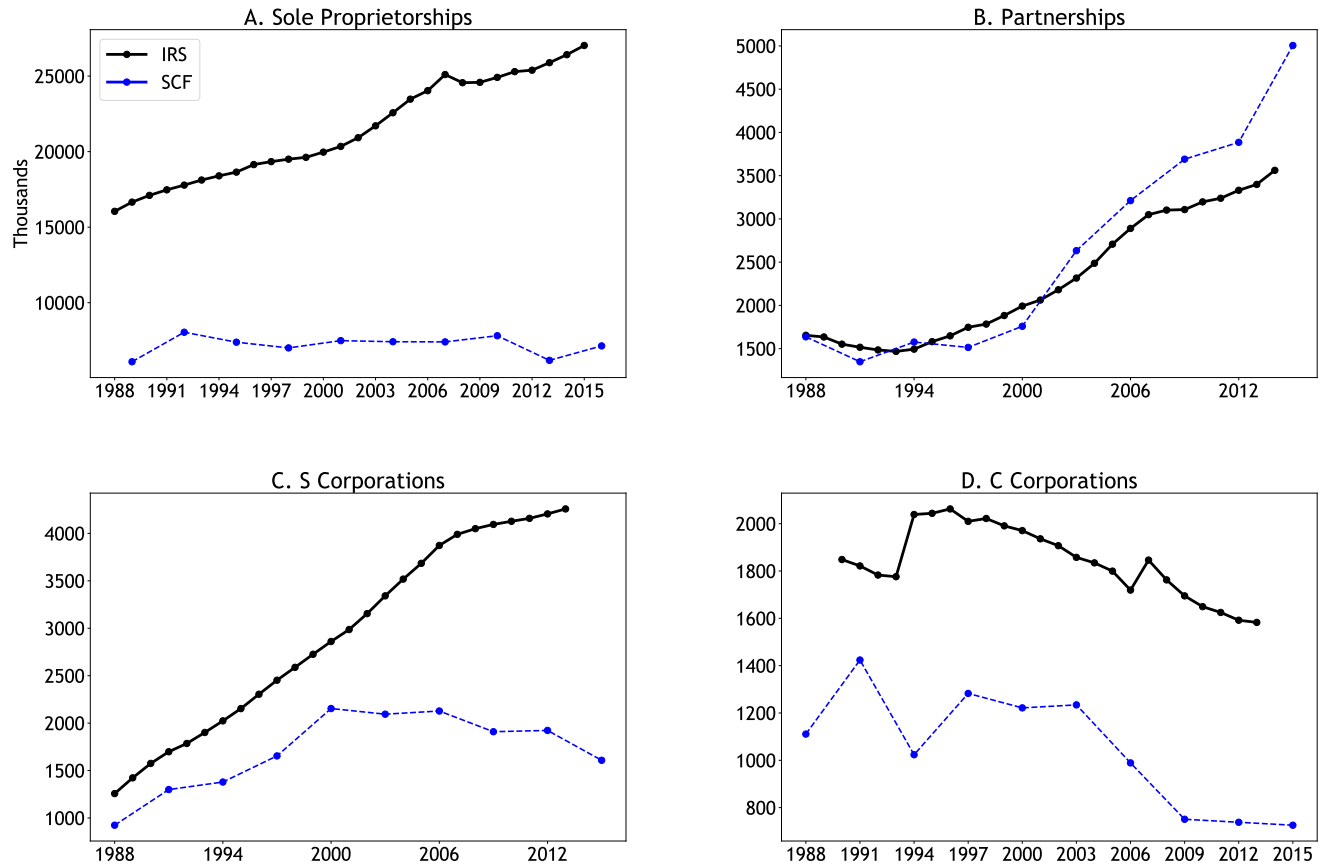
Note: This figure plots the total Schedule E income (from partnerships, S corporations, rents, royalties, estates, and trusts) earned by individuals grouped by their AGI. Data from individual tax returns are first sorted by AGI. Then we take the total Schedule E income reported by individuals within each pre-specified bin. AGI bins include the bottom 25 percent, 25 to 50 percent, 50 to 75 percent, 75 to 90 percent, 90 to 99 percent, and 99 to 100 percent of returns.

Figure 8: Distribution of S Corporation Income (Business Tax Form) SCF vs. IRS



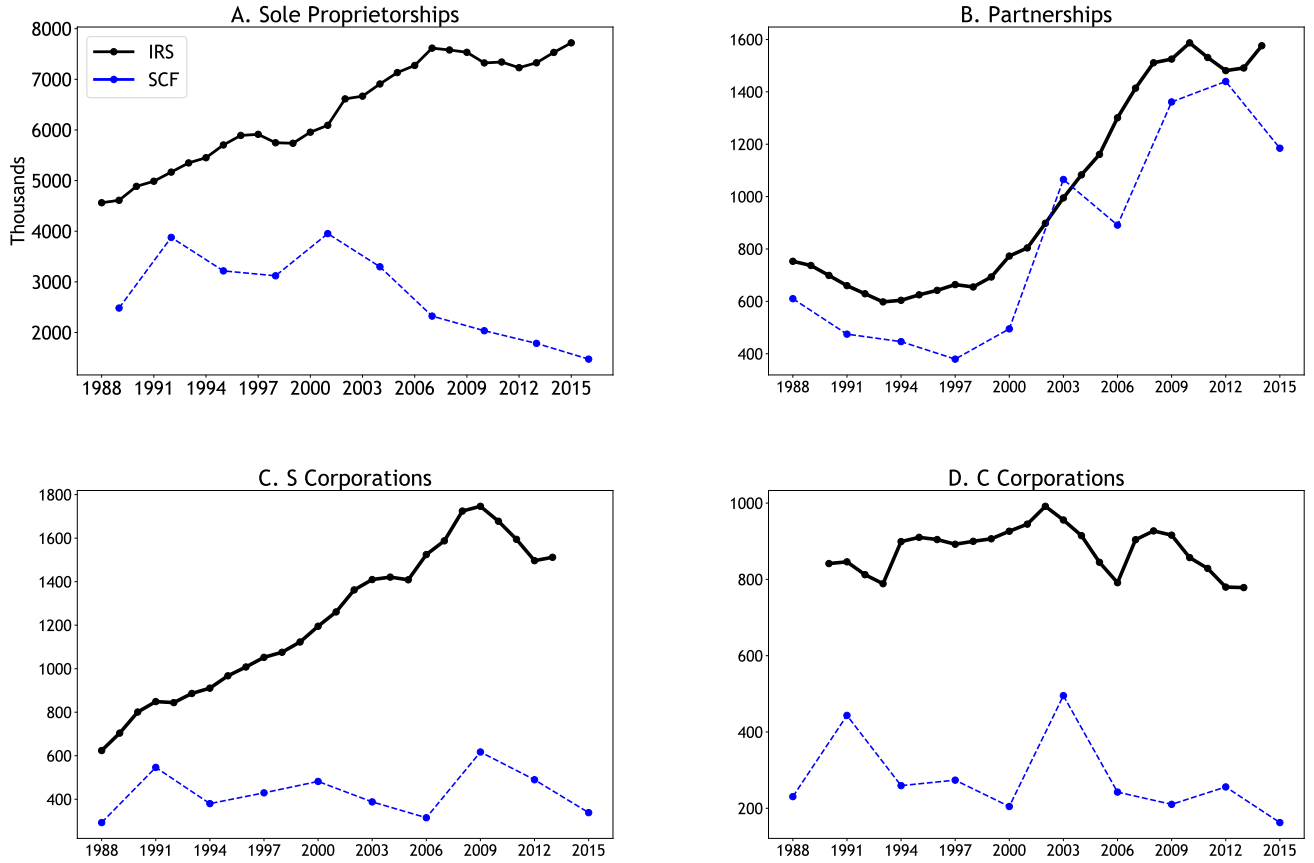
Note: This figure plots S corporation business income as reported on business tax Form 1120S when S corporation returns are grouped by business receipts. Data from S corporation business tax returns are first sorted by business receipts. Then we take the total business income reported by all S corporations within each pre-specified bin. Business receipt bins include the bottom 25 percent, 25 to 50 percent, 50 to 75 percent, 75 to 90 percent, 90 to 99 percent, and 99 to 100 percent of returns.

Figure 9: Number of Returns by Legal Entity SCF vs. IRS



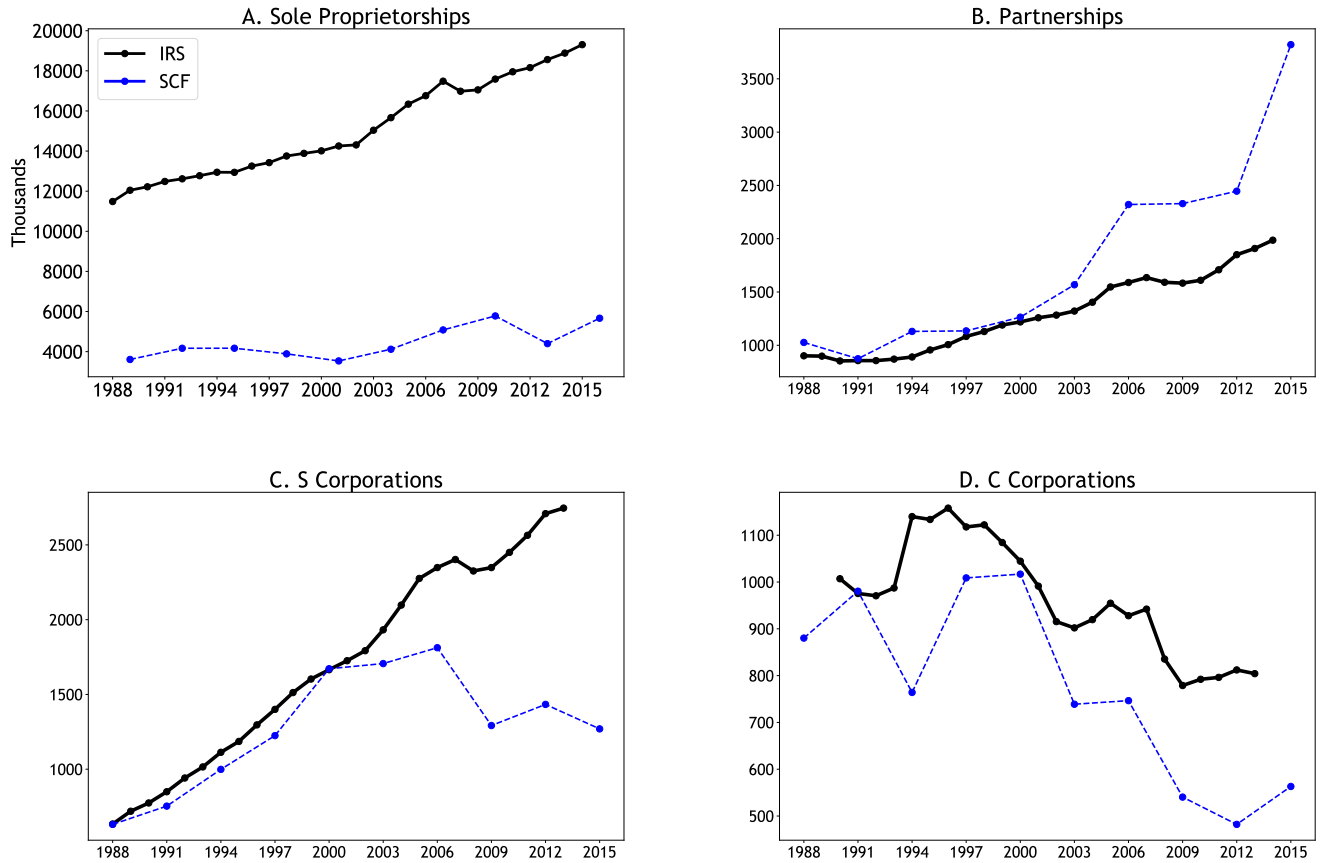
Note: This figure plots the number of business returns of sole proprietorships, partnerships, S corporations, and C corporations over time in the IRS and the SCF. In the SCF, we assume that a proprietor that owns multiple sole proprietorships files one return. This assumption is made to be consistent 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.” (Adrian Dungan, “Sole Proprietorship Returns, Tax Year 2015,” *Statistics of Income Bulletin*, Fall 2017, page 2, note 1). 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. IRS data for partnerships, S corporations, and C corporations are available only until 2013, and C corporations data starts from 1990 because data for Form 1120 is not available for 1988 and 1989.

Figure 10: Number of Returns of Businesses with Net Loss: SCF vs. IRS



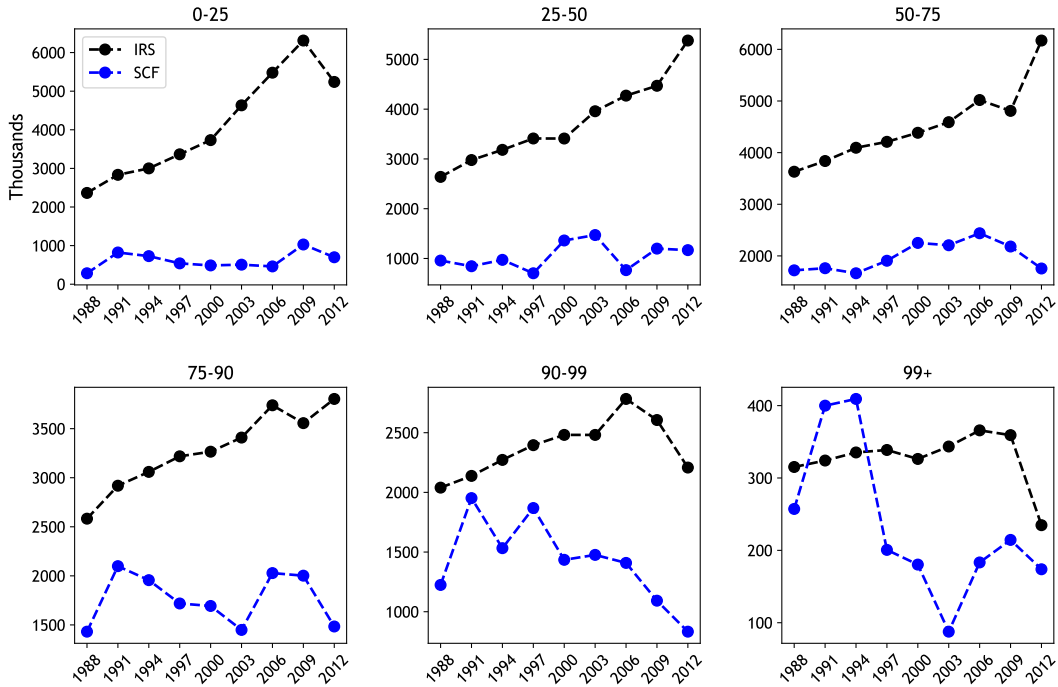
Note: This figure plots the number of returns for businesses with net losses for sole proprietorships, partnerships, S corporations, and C corporations over time in the IRS and the SCF. Break-even returns are included in the group of businesses with net losses to be consistent with IRS calculations. In the SCF, we assume that a proprietor that 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.” (Adrian Dungan, “Sole Proprietorship Returns, Tax Year 2015,” *Statistics of Income Bulletin*, Fall 2017, page 2, note 1). 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. IRS data for partnerships, S corporations, and C corporations are available only until 2013, and C corporations data starts from 1990 because data for Form 1120 is not available for 1988 and 1989.

Figure 11: Number of Returns of Businesses with Net Income: SCF vs. IRS



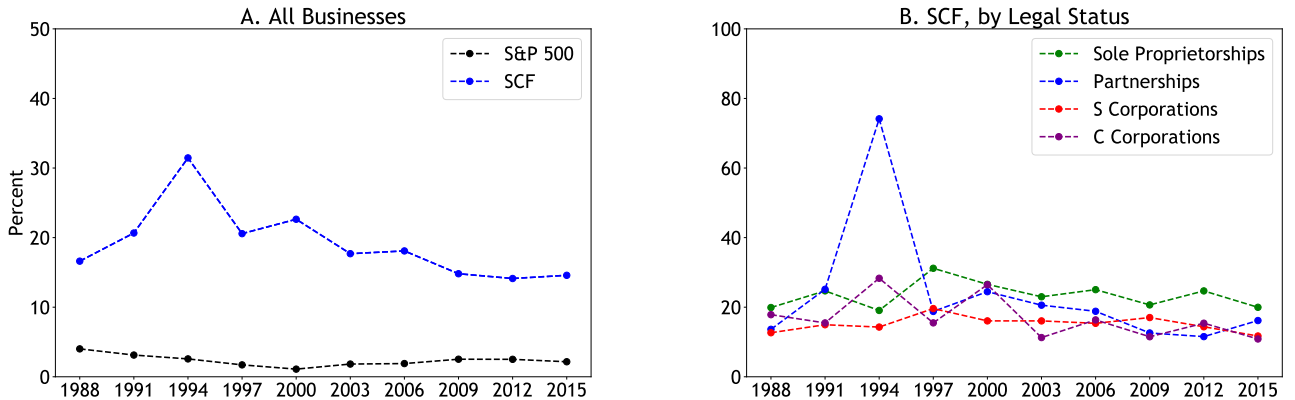
Note: This figure plots the number of returns for businesses with net income for sole proprietorships, partnerships, S corporations, and C corporations over time in the IRS and the SCF. In the SCF, we assume that a proprietor that owns multiple sole proprietorships files one return. This assumption is 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.” (Adrian Dungan, “Sole Proprietorship Returns, Tax Year 2015,” *Statistics of Income Bulletin*, Fall 2017, page 2, note 1). 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. IRS data for partnerships, S corporations, and C corporations are available only until 2013, and C corporations data starts from 1990 because data for Form 1120 is not available for 1988 and 1989.

Figure 12: Distribution of Sole Proprietorship Number of Returns: SCF vs. IRS



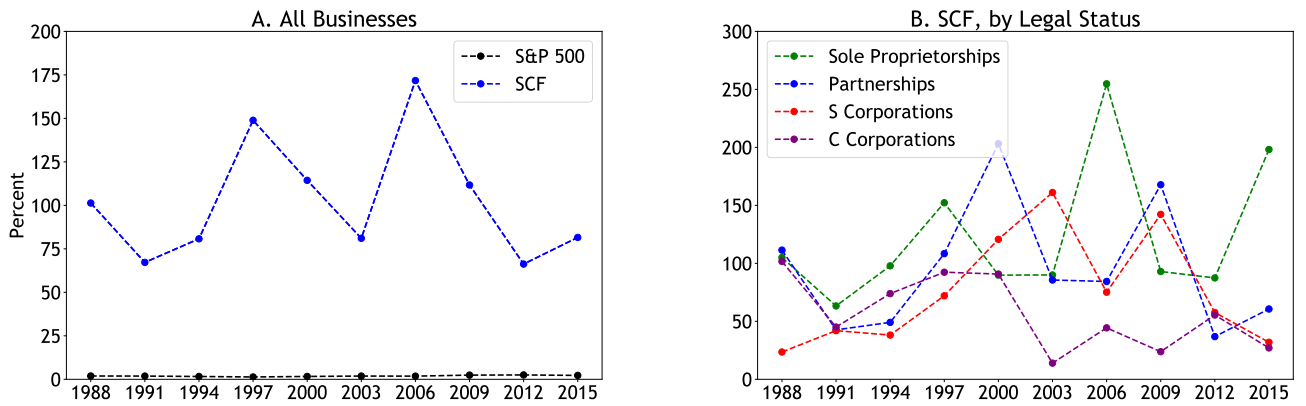
Note: This figure plots the total Form 1040 Schedule C (sole proprietorship) returns filed by individuals grouped by their AGI. Data from individual tax returns are first sorted by AGI. Then we take the total Schedule C returns filed by individuals within each pre-specified bin. AGI bins include the bottom 25 percent, 25 to 50 percent, 50 to 75 percent, 75 to 90 percent, 90 to 99 percent, and 99 to 100 percent of returns.

Figure 13: Value-weighted Average Dividend Yields



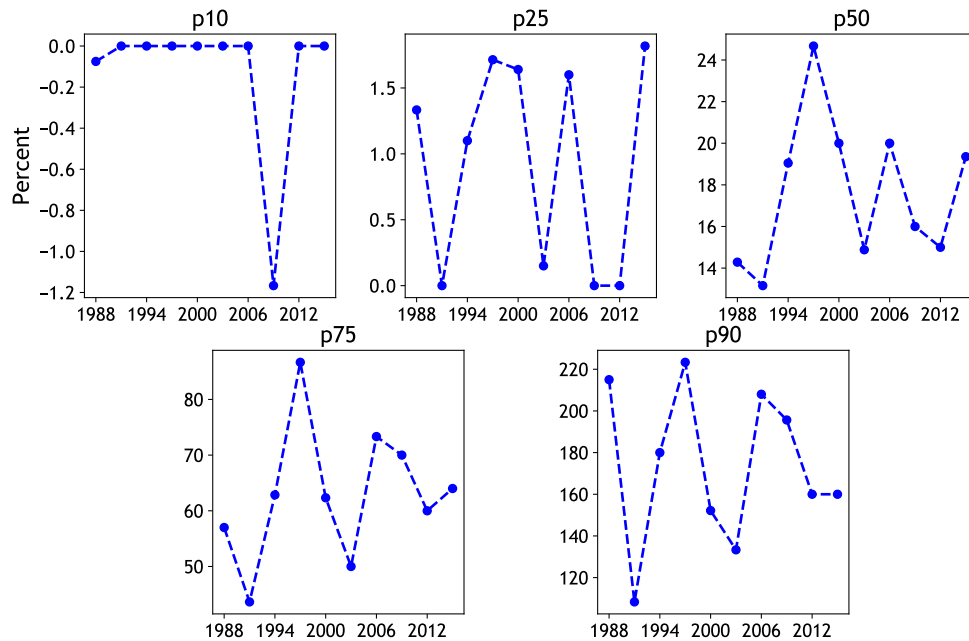
Note: This figure plots average dividend yields - calculated as the business income divided by the net worth of the business — of an index (valued-weighted) of pass-through businesses in the SCF and the corresponding value-weighted average dividend yields for businesses in the S&P 500 data (Panel A) and businesses with different legal status in the SCF (Panel B) over time. 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 14: Equally-weighted Average Dividend Yields



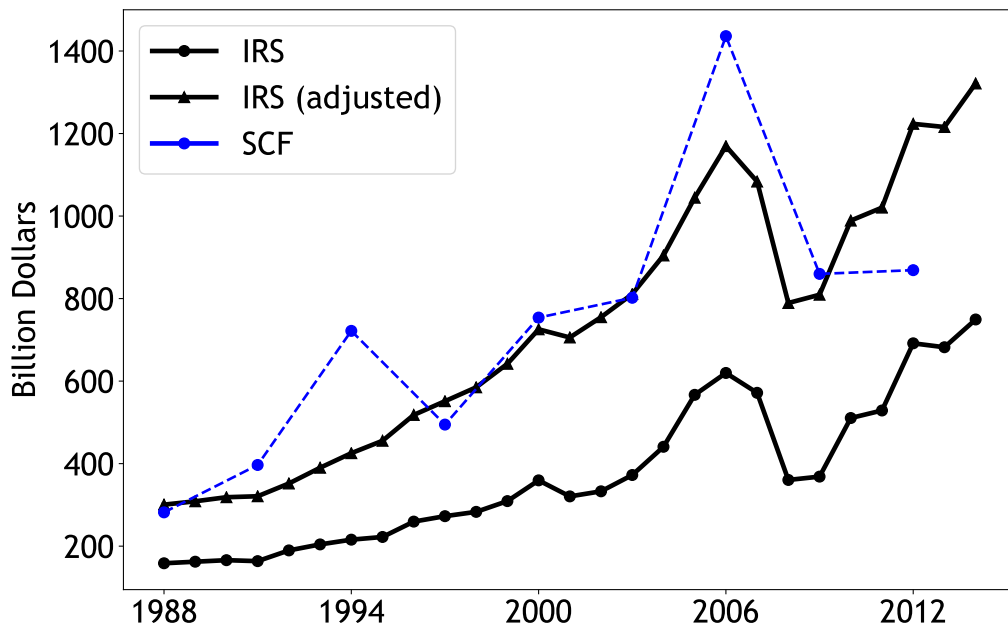
Note: This figure plots equally-weighted average dividend yields - calculated as the business income divided by the net worth of the business — for all businesses in the SCF and the equally-weighted average dividend yields for businesses in the S&P 500 data (Panel A) and businesses with different legal status in the SCF (Panel B) over time. 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 15: Distribution of Dividend Yields in the SCF



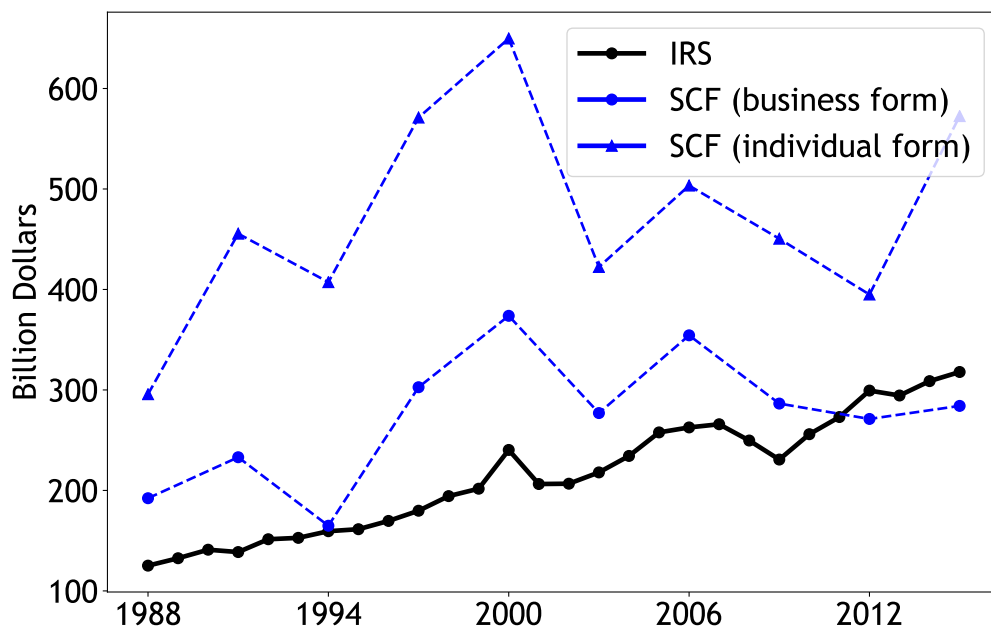
Note: This figure plots the various percentiles of the distribution of dividend yields - calculated as the business income divided by the net worth of the business — for all U.S. pass-through businesses over time. The sample includes businesses with positive net worth and excludes the bottom 1st percentile of these businesses. The business income of each business that the family members own is obtained from SCF variables that correspond to information on business tax forms.

Figure 16: Sole Proprietorship and Partnership Income with Tax Misreporting Adjustments



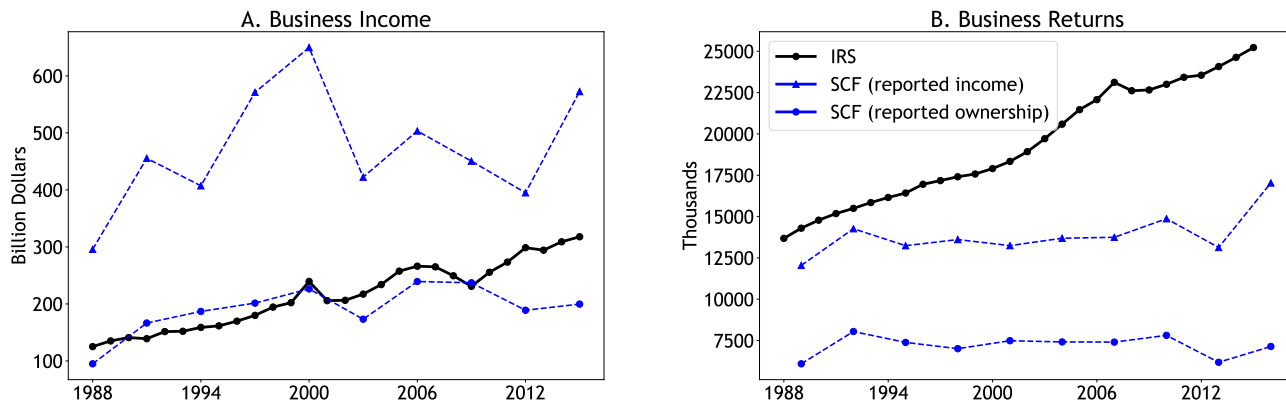
Note: In this figure, we use data on adjustments for tax misreporting on income tax documents for proprietorships and partnerships published by the U.S. Bureau of Economic Analysis. We add these yearly adjustments to the sum of proprietorship and partnership income in the IRS and compare it with estimates of business income for these businesses in the SCF.

Figure 17: Sole Proprietorship Income under Two Different SCF Variables



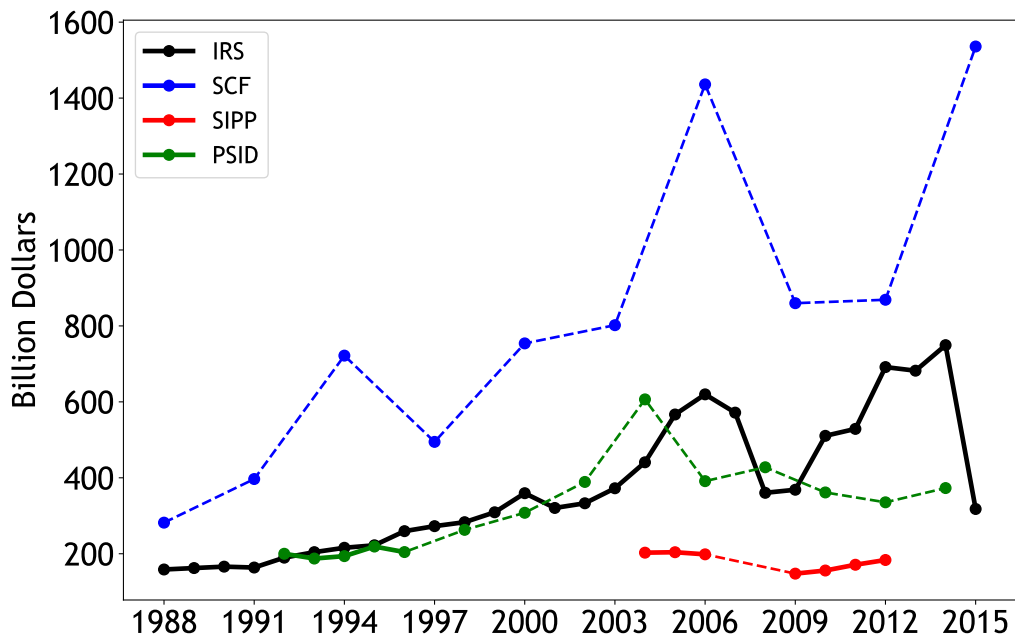
Note: This figure plots the total business income of sole proprietorships in the IRS and the SCF under two different SCF variables: i) variables that correspond to the individual tax form (IRS Form 1040) or ii) variables that correspond to business tax forms (Form 1040 Schedule C for sole proprietorships, Form 1065 for partnerships, and Form 1120S for S corporations). The SCF variable is X5704 for the former and X3132, X3232, and X3332 for the latter (where we use the ownership shares of each business when calculating the total sole proprietorship income).

Figure 18: Sole Proprietorship Income and Number of Returns under Individual Tax Forms with Different Assumptions



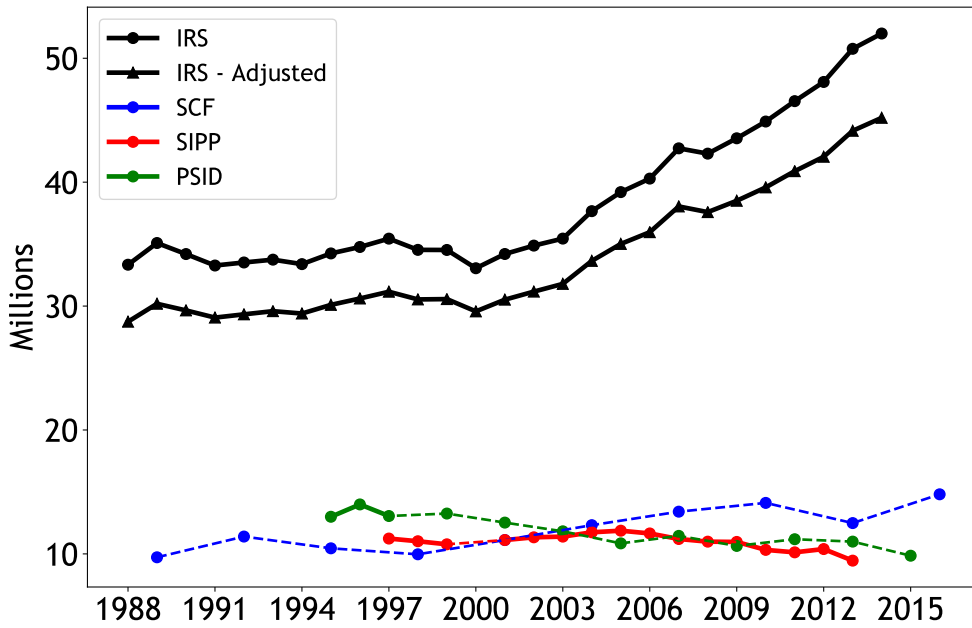
Note: This figure plots the total business income and number of returns of sole proprietorships in the IRS and the SCF. The SCF time series are generated from variables that correspond to individual tax form (IRS Form 1040). We identify that there are some respondents with non-zero net annual income from a sole proprietorship or farm (IRS Form 1040 lines 12 and 18) but at the same time, they do not report owning a sole proprietorship. Given this problem in the data, we calculate the business income of sole proprietorships by assuming either that the reported income is correct (reported income line) or business ownership is correct (reported ownership line). That is, we assume that the respondent does not own a sole proprietorship, and thus we do not incorporate the reported income into the aggregation.

Figure 19: Unincorporated Business Income: SIPP, PSID, SCF vs IRS



Note: This figure plots the total business income of unincorporated businesses in the SIPP, PSID, SCF, and the IRS. Before 2004, the SIPP does not provide information about individual's own share of the business income from the 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.

Figure 20: Number of Unincorporated Business Owners: SIPP, PSID, SCF vs IRS



Note: This figure plots the total number of unincorporated business owners in the SIPP, PSID, SCF, and the IRS. In each data, this value is obtained by the summation of total number of partners and total number sole proprietorship returns. Cooper et. al (2016) document that 73.4 percent of all partnerships are owned by individuals in 2011. Adjusted IRS is obtained by multiplying IRS line by this number to make an alternative comparison between the IRS and the other survey data.