



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 July 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

July 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 measurement issues for statistics of interest, such as returns and valuations of ongoing private businesses, that are inherently latent and cannot be recovered using either survey or administrative data.

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 extensively used 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. We also discuss measurement issues for statistics of interest, such as the returns and valuations of ongoing private businesses, that are inherently latent and cannot be recovered using either survey or administrative data.

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 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, 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 a factor of 2 to 3, 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

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 the corporation derives.

²Smith et al. (2017) use tax audit data to conclude that rising business income accounts for all the growth in the top 1% income share since 2000. Furthermore, the majority of rising top business income resulted from rising income from 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.

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 unincorporated as a sole proprietorship or a partnership – whereas the PSID does not provide these details and therefore cannot be compared 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 anywhere from 3 to 8 times depending on the survey year.

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 but conceptually difficult to measure when one considers that private businesses are heavily invested in intangible assets.⁴ As an evidence to this difficulty, we find large differences between average business returns obtained from the SCF and from S&P 500 company data. For example, the value-weighted average for pass-through businesses is around 20 percent in the SCF compared to 2.3 percent in the S&P 500 data. The value-weighted average for C corporations, which should be more in line with S&P data than the pass-throughs, falls between 10 and 26 percent depending on the survey year.⁵ 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, and one-fourth have dividend yields that exceed 50 percent. In contrast, 90 percent of businesses in the S&P 500 have dividend yields below 6 percent.

Overall, using dividend yields for the S&P 500 composite as a reference point, SCF pass-through 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, 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.

2 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

⁴Bhandari and McGrattan (2018) find that aggregate sweat equity – the value of time to build customer bases, client lists, and other intangible assets – is 0.65 times GDP, which is close to the estimate of fixed assets used by private businesses. Intangible assets also come in the form of research and development, software, advertising, brands, and investments in building organizations.

⁵The magnitude of the SCF returns we compute are comparable to those found by Moskowitz and Vissing-Jorgensen (2002) and Kartashova (2014).

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 to be 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); and income from gains from the sale of capital and other property (Form 1040 lines 13 and 14). Figure 2 shows that the SCF matches aggregate wages and salaries quite well but consistently overstates broad business income. This suggests that while the SCF matches up well to aggregates such as total income, problems arise when total income is broken down into its subaggregates, especially in relation to business income.

Comparisons can also be made between other survey data such as the SIPP and the IRS. Unlike the SCF, the SIPP does not ask respondents to refer to their tax forms. Thus, we construct total income 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. Figure 1 shows that the SIPP understates total income by around 17 percent on average relative to the IRS. What this finding emphasizes, however, is that a large degree of inconsistencies can be found even across survey data.

3 Pass-Through Businesses

3.1 Business Income and Receipts

While the SCF does relatively well in matching aggregates such as total income or wages and salaries, it significantly overstates business income. 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, and Form 1120S for S corporations. We

calculate the same statistics from the SCF using variables that exactly correspond to their IRS counterpart. The figure shows that across all legal structures and across time, business income is largely 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 the SCF responses were overstated by 93.8 percent. Table 1 reports 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, with percentage errors averaging 31.5 percent for sole proprietorships, 305.4 percent for partnerships, and 137.1 percent for S corporations. Furthermore, the degree of 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. Finally, Table 1 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 significantly over time as well.

3.1.1 Distribution of Business Income

The SCF also exhibits significant errors in the distribution of business income when compared with the IRS. Figures 5 and 4 compare SCF and IRS business income for businesses that either report a net loss or report a positive net income. 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). Figure 6 shows that while the SCF overstates Schedule C income relative to the IRS, it understates Schedule C income earned by those with low AGI but severely overstates it for families with high AGI. In contrast, 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. Table 2 reports Schedule C income by AGI subgroups in the SIPP, and unlike the SCF, it understates sole proprietorship income across all groups, which results in an overall

percentage error of 42 percent. These findings highlight once again that relative errors between business income in survey data and aggregate IRS data are non-uniform across subgroups.

3.1.2 Non-representativeness and Measurement Error

We now investigate the reasons behind the overestimation of business income in the SCF relative to the IRS over time. We focus on two potential reasons for the overestimation: 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, the overestimation of business income in the SCF relative to the IRS is 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, and S 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.

We now compare the number of returns for proprietorships, partnerships, and S 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 and S corporations, as the number of returns for both of these types of businesses is either flat or even decreasing over time, while their respective counterparts in the IRS show clear, positive trends. In 2012, the number of sole proprietorships with net losses is only one-fourth and the number of S corporations with net losses is 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. We find that misreporting of business income is a more promising candidate to explain the overestimation of business income for businesses with net income across different business types. For example, the number of sole proprietorships with net income in the SCF is almost constant in the SCF, and it only represents between one-fourth and one-third of the aggregate number in the IRS. However, as we have documented in 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 supports the presence of misreporting of business income in the SCF. Finally, 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. Again, this finding suggests misreporting of business income in the SCF.

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.

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 2016.⁶ 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.⁷ 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.

Finally, we show that the type of problems we have documented about the SCF are also present in other surveys. To illustrate this, in Table 4, we compare the number of returns and business income for sole proprietorships and partnerships in the SIPP and IRS before and after the Great Recession. Interestingly, we find that the SIPP actually understates business income relative to the IRS, which is the opposite of the SCF. Moreover, the SIPP only represents less than half of the aggregate sample in the IRS. This suggests that underrepresentation of the sample contributes to the underestimation of business income for sole proprietorships and partnerships in the SIPP. These results also highlight once again that the business income of different types of businesses are not even comparable across the two survey datasets. This finding is problematic because theoretical models on businesses that match the SCF or SIPP would yield completely different implications.

⁶Similar results also hold for other surveys of the SCF.

⁷Here, relevant documents include income tax returns, pension documents, account statements, investment and business records, and loan documents.

3.2 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 pass-through businesses. Figure 13 shows that value-weighted dividend yield from 1989 to 2015 fluctuated between 14 and 32 percent.⁸ 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 pass-through 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, and all 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 58 and 184 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 pass-through businesses over time in the sample described above. 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

⁸The 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.

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. In Table 5, we show that the SCF distribution of dividend yields is largely different from its counterpart obtained from the S&P 500.

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.

4 C Corporations

We now compare statistics between C corporations in the SCF and the IRS. In the SCF, one of the possible answers to a question about the type of actively managed businesses is “other corporations including C corporations”. We use this answer to identify C corporations, which implies an upper bound for the estimates we will discuss below. We then compute for total number of returns, net income, and business returns using similar calculations made for pass-through businesses in the previous section.

We find that the total number of returns reported in the SCF are understated by factors of 3 to 8 times relative to IRS data depending on the survey year. For example, total number of returns in the IRS in 2012 was 5.8 million whereas it was only 738 thousand in the SCF. Total net income is also understated in the SCF relative to the IRS by a factor of 2 to 5. In 2012, the total net income of C corporations in the IRS was 1.052 trillion dollars, but only 223 billion dollars in the SCF. Dividend yields of C corporations, on the other hand, are found to be large when compared to the S&P 500, as was the case with pass-through businesses.⁹ Value-weighted average dividend yields range between 10 to 26 percent, while equally-weighted average dividend yields range between 13 and 100 percent over the last two decades. In terms of the distribution, more than half of C corporations have dividend yields higher than 20 percent on average.

However, one must exercise caution in interpreting these results. 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

⁹Similar to our calculations for pass-through business returns, we restrict the sample of C corporations 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.

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.

5 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 of incomes with the KFS, PSID, and PSED and document that the other survey results have different measurement issues.

5.1 Adjustments to IRS or SCF

5.1.1 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.

5.1.2 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.¹⁰ 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

¹⁰The percentage of misreporting of S corporations, estate, and trust income is assumed to be negligible.

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.

5.1.3 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.

5.2 Other survey data

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.

The PSID has data on the business ownerships and income of unincorporated businesses. Moreover, from these data we can identify incorporated or unincorporated businesses. However, the PSID does not distinguish between sole proprietorships and partnerships (among unincorporated businesses) or between S corporations and other corporations (among incorporated businesses). Furthermore, the PSID provides business income information only if the business is unincorporated.

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 2006.¹¹ As a result, 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.

6 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 anywhere from 3 to 8 times depending on the survey year. 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

¹¹The fraction of entrepreneurs responding to this question is similar for other years.

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.

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. Working Paper No. 24520, National Bureau of Economic Research.
- [3] 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.
- [4] 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*.
- [5] Kartashova, K. (2014). Private equity premium puzzle revisited. *American Economic Review*, 104(10), 3297-3334.
- [6] Moskowitz, T. J., & Vissing-Jørgensen, A. (2002). The returns to entrepreneurial investment: A private equity premium puzzle?. *American Economic Review*, 92(4), 745-778.
- [7] 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: Percentage Errors of Total Income, Business Income, and Business Receipts in the SCF relative to IRS

Tax Year	Total income	Business income			Business receipts		
		Sole prop.	Partnership	S corp.	Sole prop.	Partnership	S corp.
1988	5.5	53.6	168.6	107.1	78.54	13.4	-7.8
1991	8.6	68.1	554.2	249.8	251.8	89.1	-52.4
1994	4.2	3.4	889.1	158.6	15.8	203.8	6.5
1997	-1.6	68.4	106.7	145.6	15.0	-26.0	2.6
2000	-0.2	55.6	219.0	146.7	-4.9	-23.7	10.5
2003	7.9	27.1	239.8	205.4	-3.8	30.5	-26.4
2006	0.6	34.9	202.9	93.8	-26.8	5.4	13.5
2009	0.5	24.1	316.1	116.3	1.3	-6.8	9.9
2012	-3.0	-9.4	52.4	10.8	-14.0	-12.0	15.4
2015	3.7	-10.6	-	-	-39.2	-	-
Average	2.6	31.5	305.4	137.1	27.5	30.4	-3.1

Note: This table reports the percentage errors of SCF reported total income, business income, and business receipts 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 (business receipts) refers to income (gross sales) reported on Form 1040 Schedule C for sole proprietorships, Form 1065 for partnerships, and Form 1120S for S 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.

Table 2: Distribution of Sole Proprietorship Income: SCF vs. SIPP vs. IRS (tax year 2006) (in billions of U.S. dollars)

Income Percentiles	IRS Data	Survey Data			
		SIPP	% Error	SCF	% Error
0-25	15.9	3.6	-77.4	4.23	-73.4
25-50	34.5	19.6	-3.2	26.4	-23.5
50-75	4.2	39.8	-9.9	79.5	79.8
75-90	50.1	51.9	3.7	96.9	93.6
90-99	92.5	45.8	-50.4	172.8	86.8
99+	44.3	2.8	-93.7	123.5	178.7
0-100	281.5	163.6	-41.9	503.4	78.8

Note: This table compares total Schedule C (sole proprietorship) income earned by individuals grouped by their AGI in the IRS, SCF, and SIPP for tax year 2006. 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. The row with range 0 to 100 is total income in billions of dollars. Percentage error is calculated by dividing the difference between the value in the SCF/SIPP and the value in the IRS by the value in the IRS and multiplying the result by 100.

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: Percentage Errors of Number of Returns and Business Income in SIPP Relative to IRS

Number of Returns (in thousands)						
Tax Year	Sole Proprietorships			Partnerships		
	IRS	SIPP	% Error	IRS	SIPP	% Error
2005	23,472	9,789	-58	2,708	1,245	-54
2011	25,294	8,447	-67	3,239	1,109	-66

Business Income (in billions of U.S. dollars)						
Tax Year	Sole Proprietorships			Partnerships		
	IRS	SIPP	% Error	IRS	SIPP	% Error
2005	257.75	174.38	-32	308.98	29.62	-90
2011	273.05	142.49	-48	255.75	28.52	-89

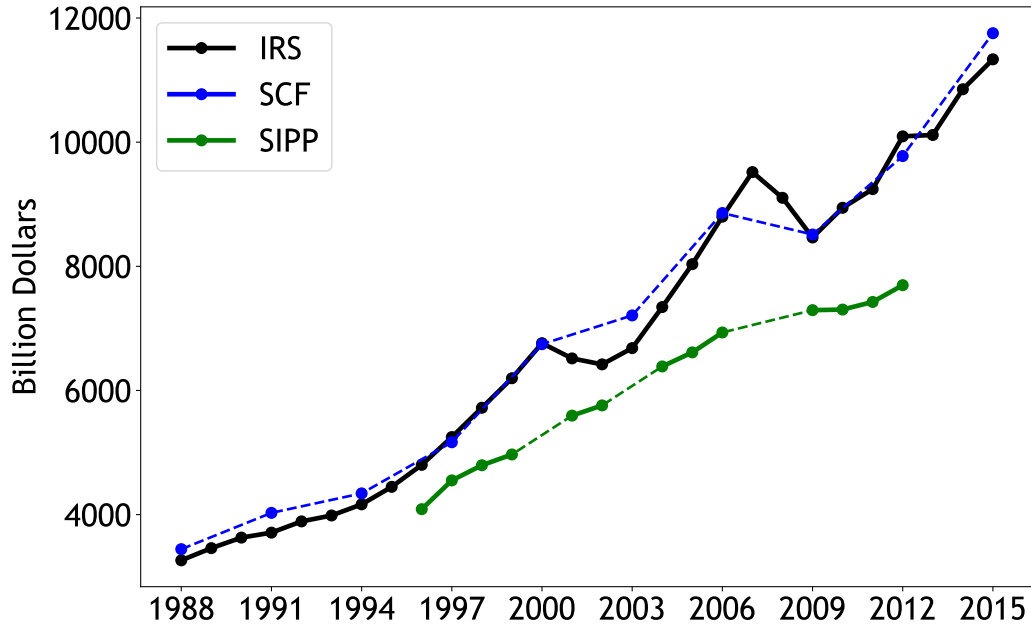
Note: This table compares the number of returns and business income for sole proprietorships and partnerships in the SIPP and the IRS before and after the Great Recession. Percentage error is calculated by dividing the difference between the value in the SIPP and the value in the IRS by the value in the IRS and multiplying the result by 100.

Table 5: Dividend Yield Distribution in SCF and S&P 500

	2003		2006	
	SCF	S&P 500	SCF	S&P 500
p25	2.08	0	2.56	0
p50	16.67	0	22.12	0
p75	50.00	2.26	78.95	2.33
p90	133.33	6.05	207.50	5.56

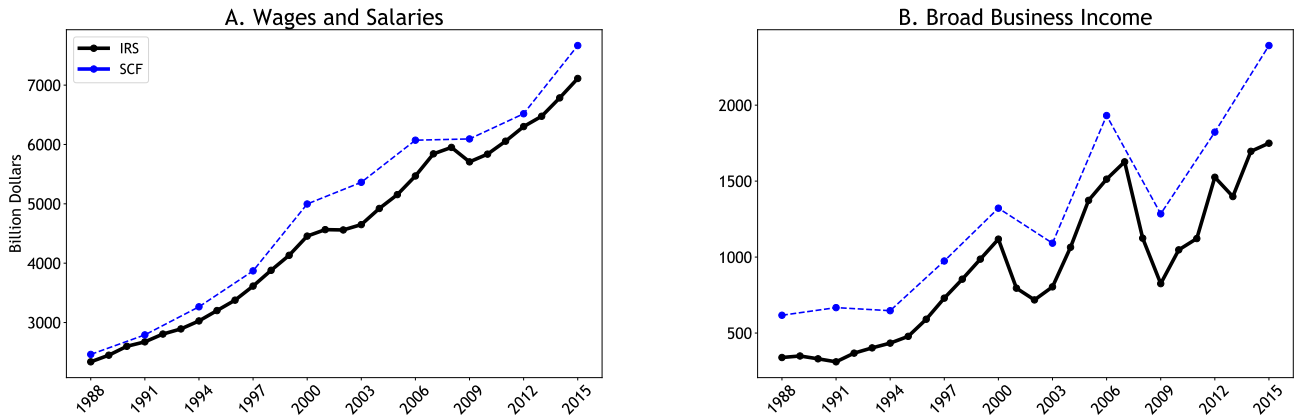
Note: This table compares various percentiles of the dividend yield distribution in the SCF and S&P 500 in 2003 and 2006. The dividend yield in the SCF is calculated as the business income divided by the net worth of the business for all businesses in the SCF. The SCF sample includes businesses with 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 1: Total Income: IRS vs. SCF vs. SIPP



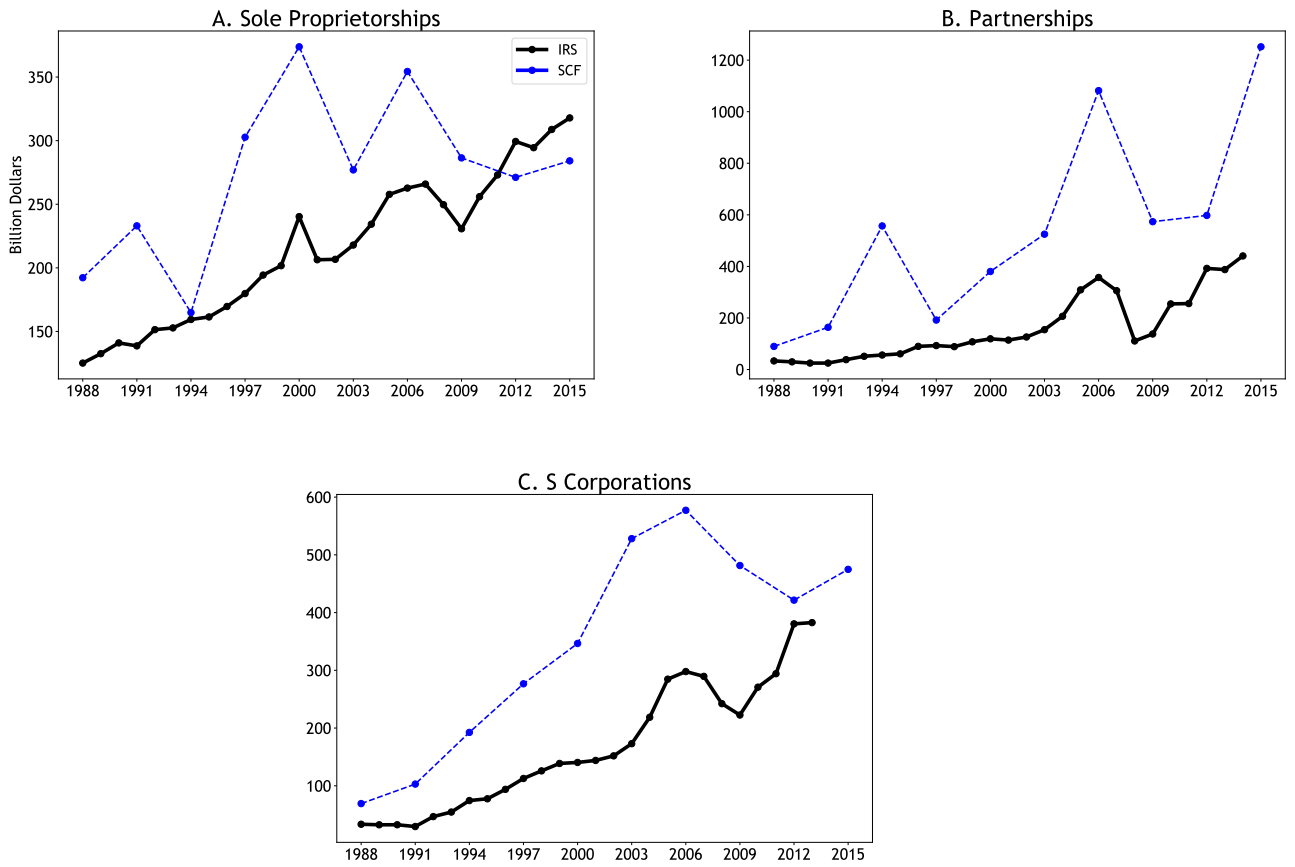
Note: This figure plots total income in the IRS, SCF, and SIPP. Total income is defined as the sum of wages and salaries; net income from a business, profession, or farm; taxable and non-taxable interest; dividends; capital gains from the sale of capital assets and other property; net income from rental, 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



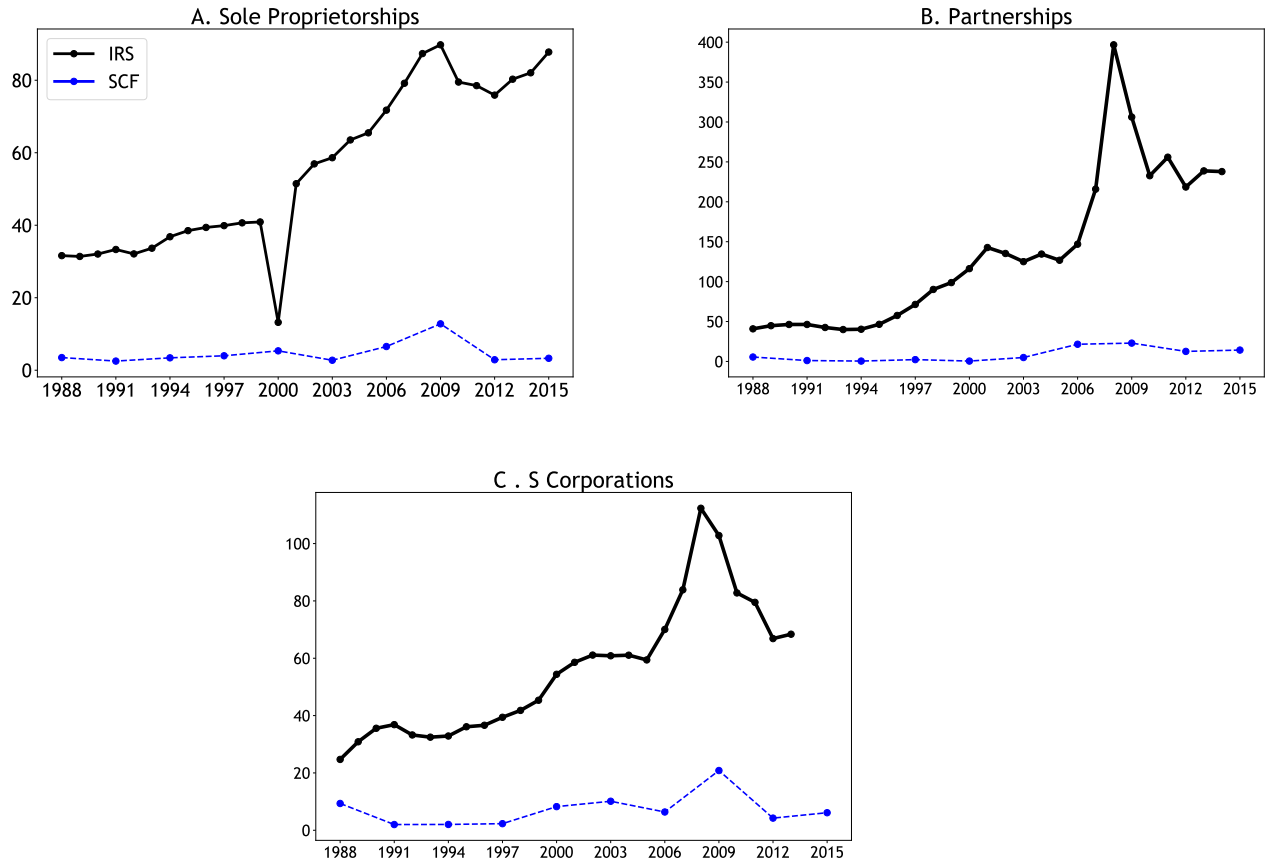
Note: This figure plots total wages and salaries (Panel A) and broad business income (Panel B) in the IRS and the SCF. Broad business income is defined as income derived from a business or profession (Form 1040 Schedule C) or farm (Form 1040 Schedule F); income from rental real estate, royalties, partnerships, S corporations, estates, trusts (Form 1040 Schedule E); 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



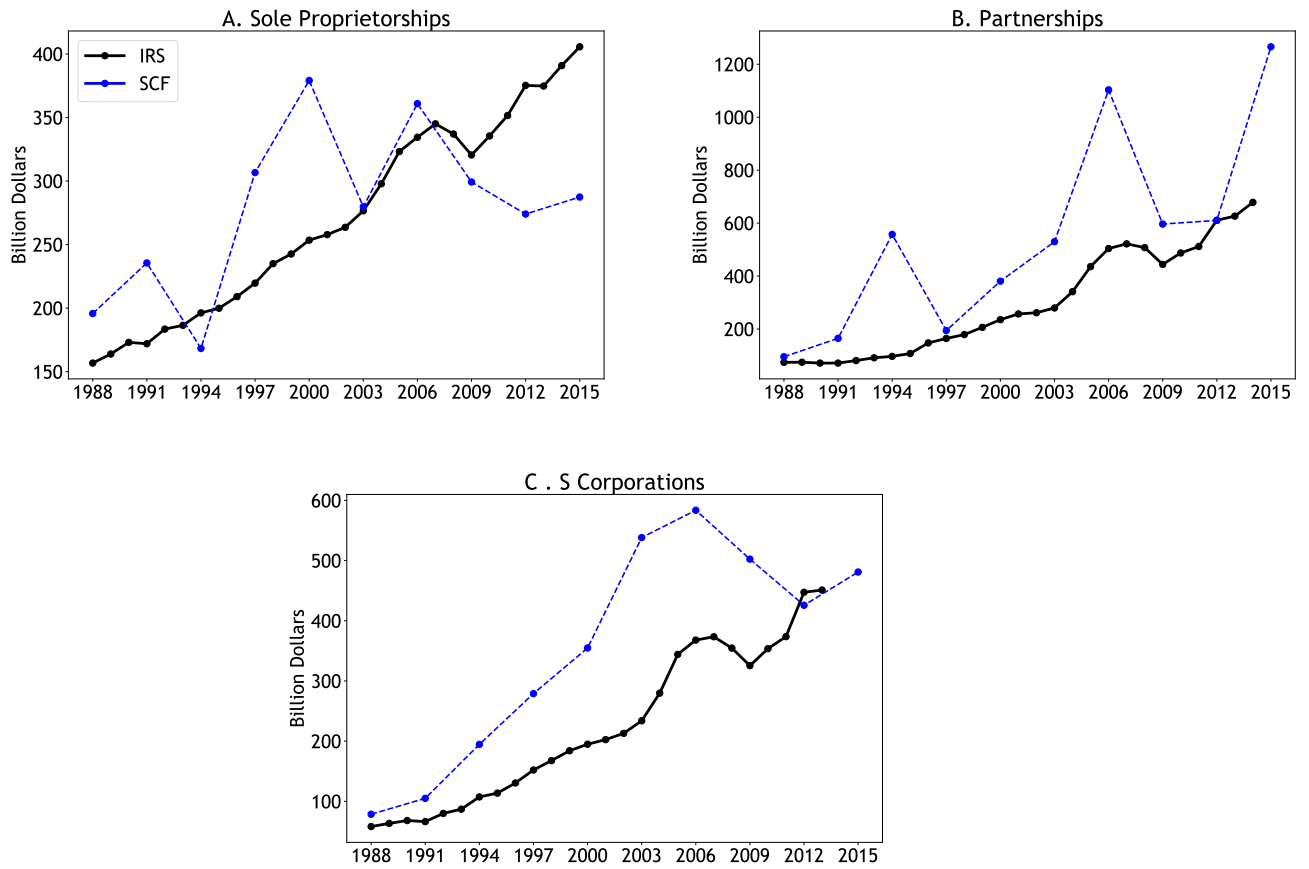
Note: This figure plots business income in the IRS and the SCF as reported on Form 1040 Schedule C for sole proprietorships, Form 1065 for partnerships, and Form 1120S for S corporations.

Figure 4: Business Income of Pass-Throughs with Net Loss: SCF vs. IRS



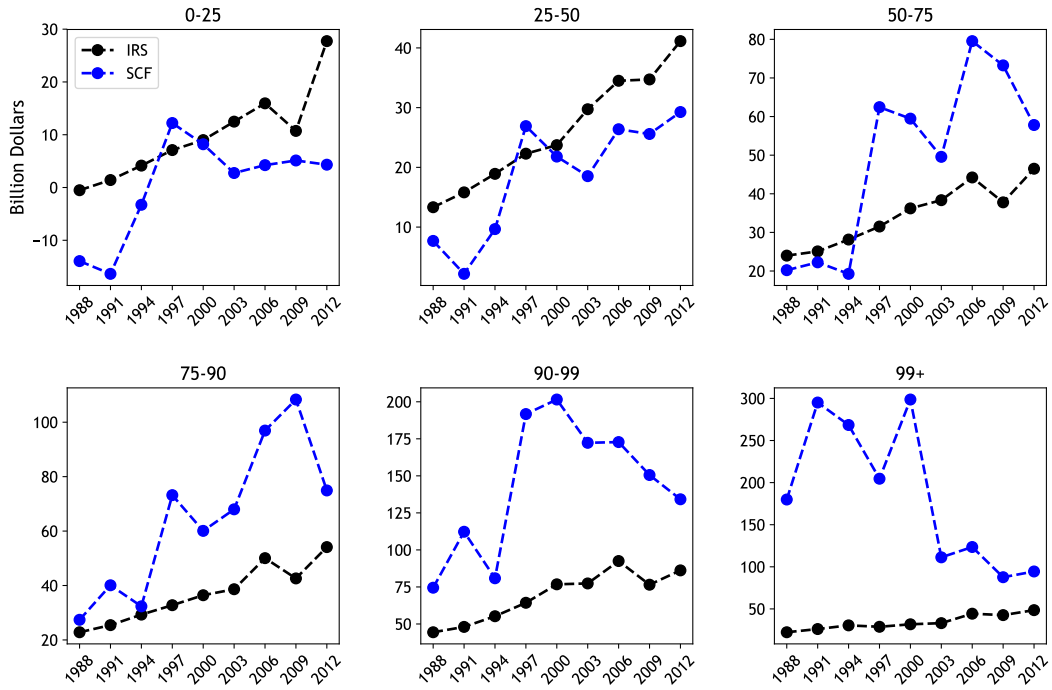
Note: This figure plots business income of businesses that report a net loss in the IRS and the SCF. Business income refers to income reported on Form 1040 Schedule C for sole proprietorships, Form 1065 for partnerships, and Form 1120S for S corporations.

Figure 5: Business Income of Pass-Throughs with Net Income: SCF vs. IRS



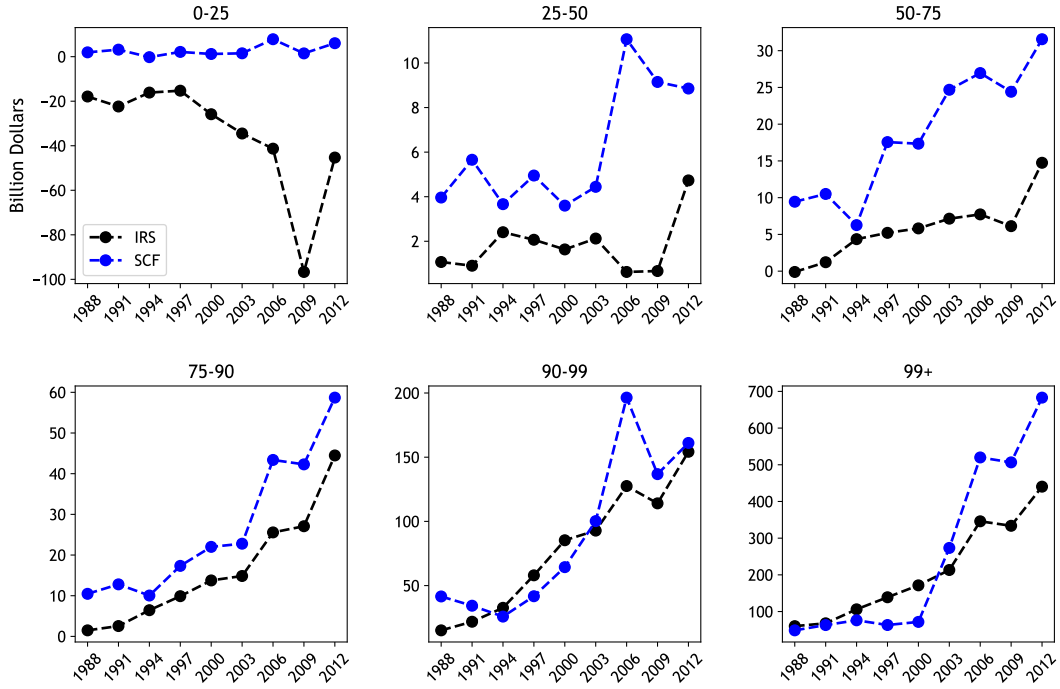
Note: This figure plots business income of businesses that report a positive net income in the IRS and the SCF. Business income refers to income reported on Form 1040 Schedule C for sole proprietorships, Form 1065 for partnerships, and Form 1120S for S corporations.

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



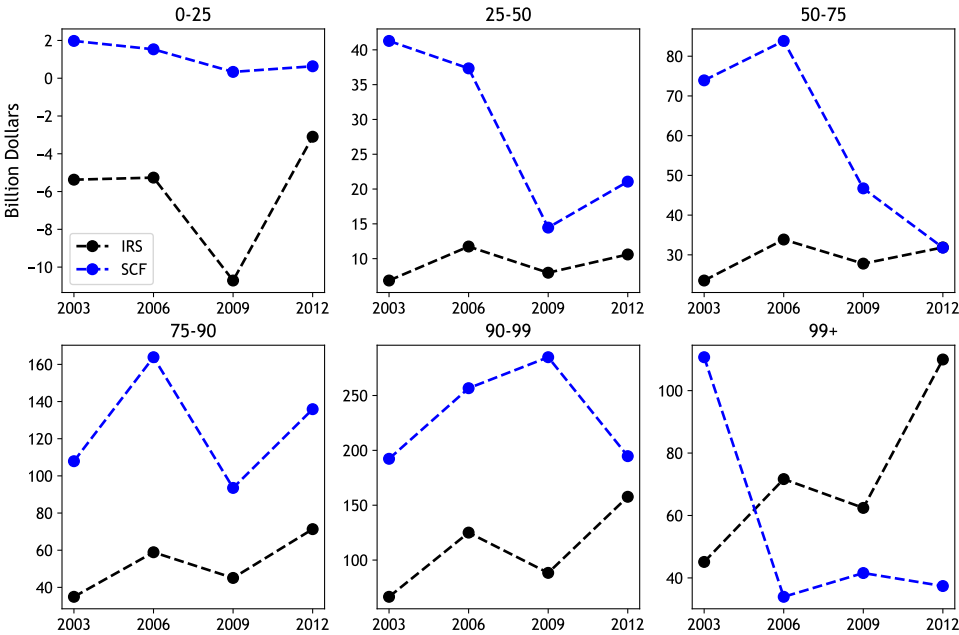
Note: This figure plots the total Schedule C (sole proprietorship) income earned by individuals grouped by their AGI. Data from individual tax returns are first sorted by AGI. Then we take the total Schedule C income reported by individuals within each pre-specified bin. AGI bins include the bottom 25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



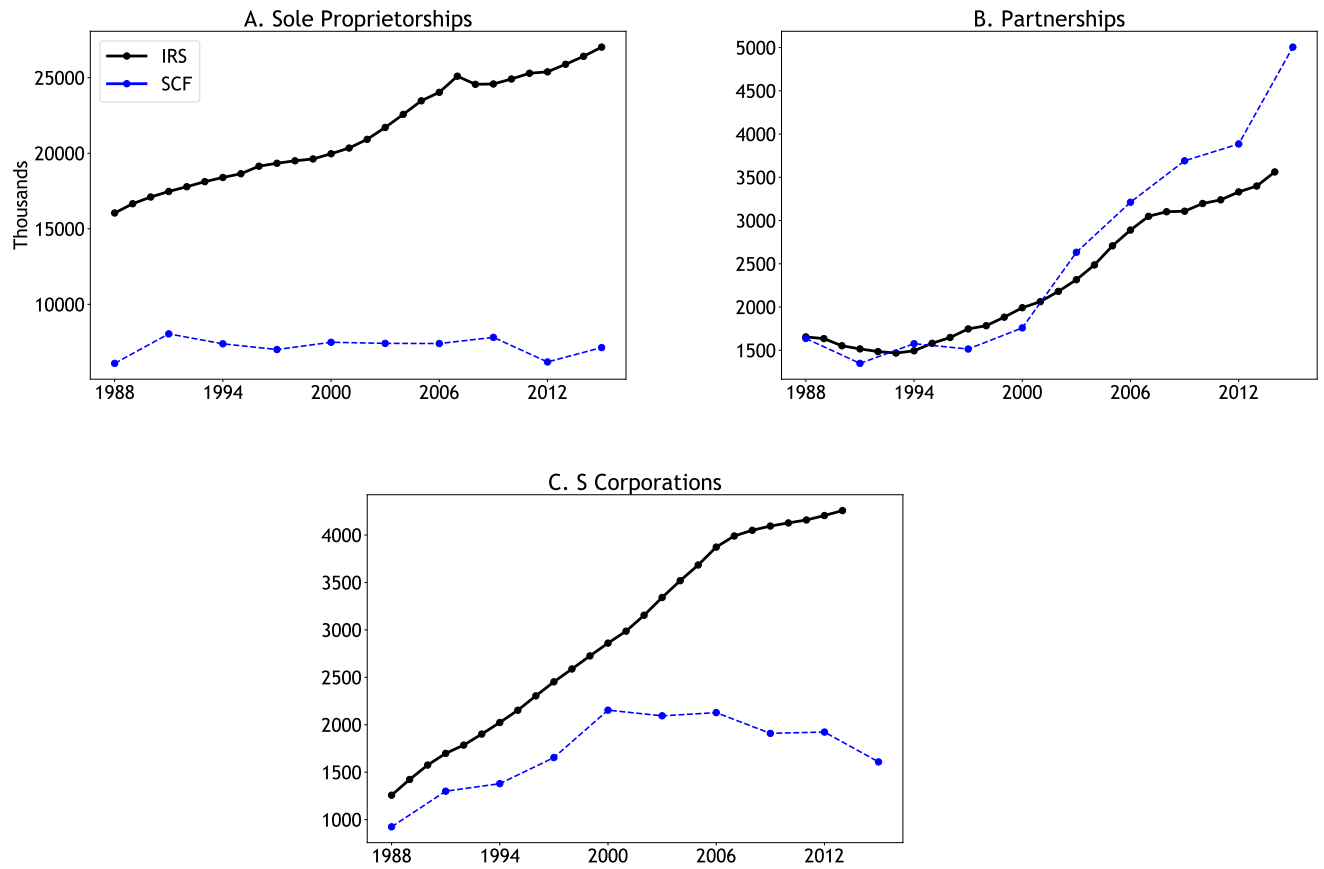
Note: This figure plots the total Schedule E income (from partnerships, S corporations, rents, royalties, estates, and trusts) earned by individuals grouped by their AGI. Data from individual tax returns are first sorted by AGI. Then we take the total Schedule E income reported by individuals within 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



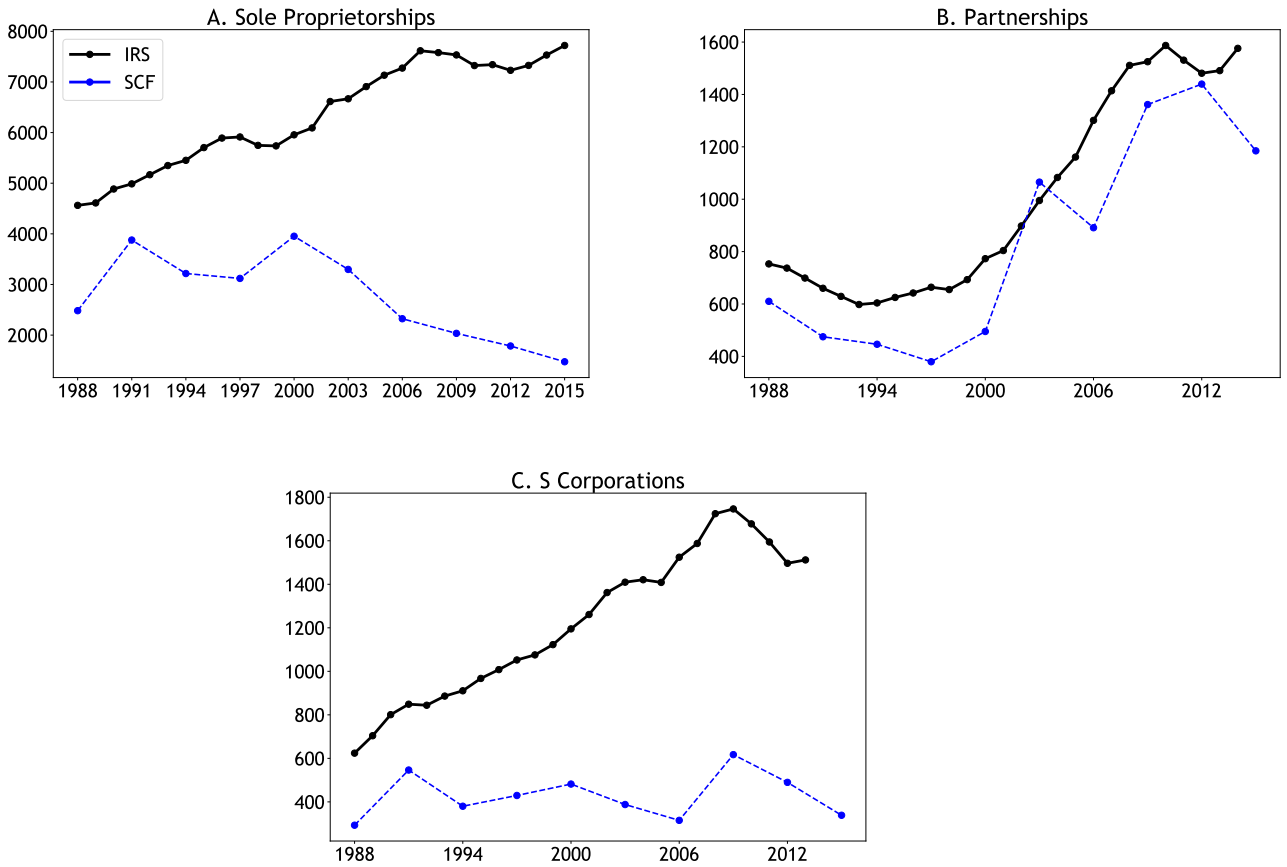
Note: This figure plots S corporation business income as reported on business tax Form 1120S when S corporation returns are grouped by business receipts. Data from S corporation business tax returns are first sorted by business receipts. Then we take the total business income reported by all S corporations 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



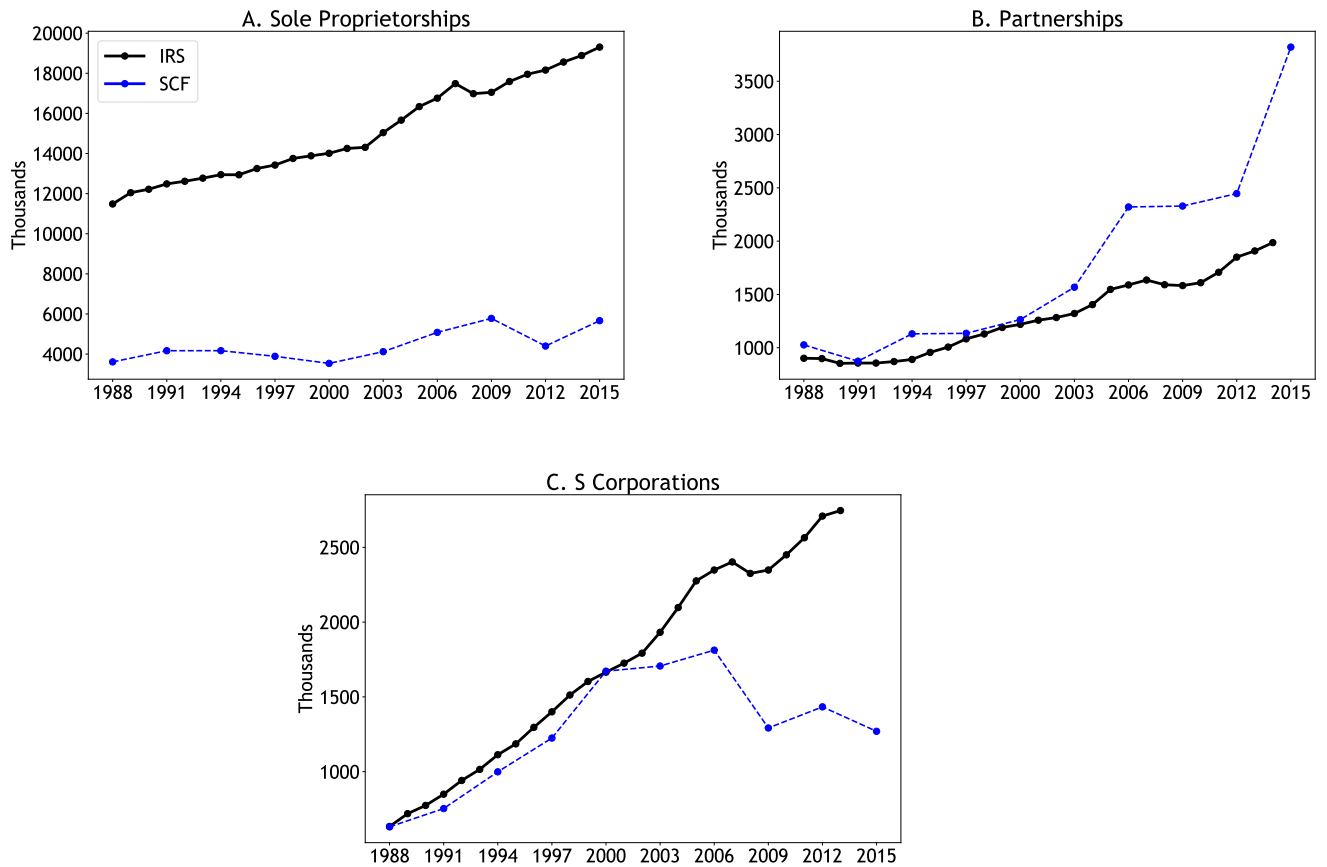
Note: This figure plots the number of business returns of sole proprietorships, partnerships, and S corporations over time in the IRS and the SCF. In the SCF, we assume that a proprietor that owns multiple sole proprietorships files one return. This assumption is made to be consistent with IRS statistics 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 and S corporations, we calculate the number of returns taking into account the ownership share of the family from each reported business.

Figure 10: Number of Returns of Pass-Throughs with Net Loss: SCF vs. IRS



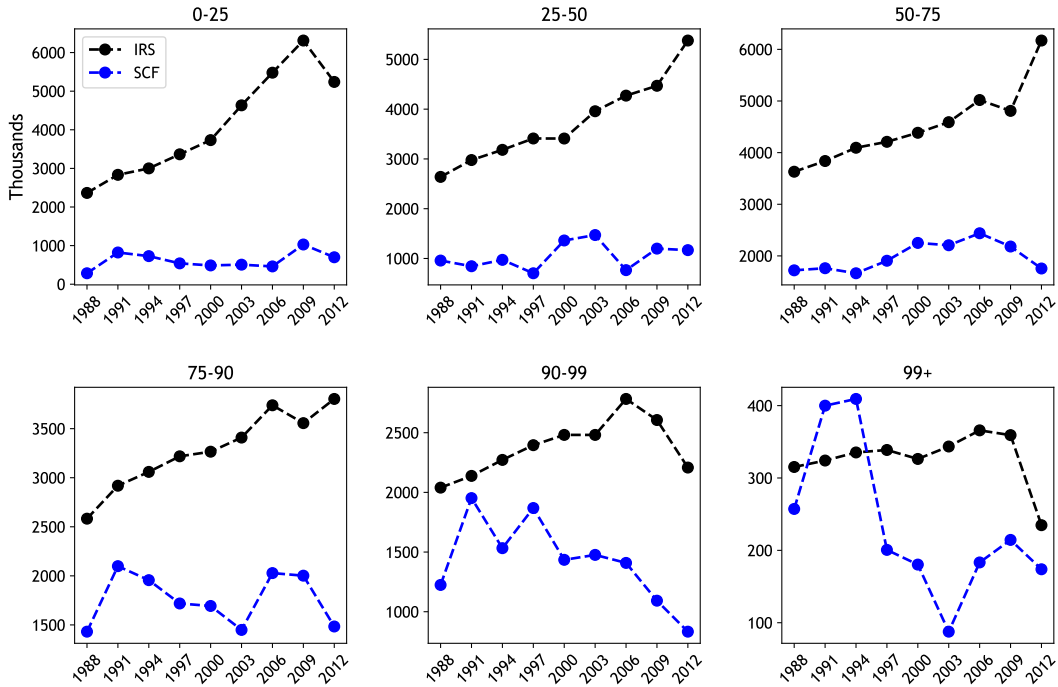
Note: This figure plots the number of returns for businesses with net losses for sole proprietorships, partnerships, and S corporations over time in the IRS and the SCF. Break-even returns are included in the group of businesses with net losses to be consistent with IRS calculations. In the SCF, we assume 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 and S corporations, we calculate the number of returns taking into account the ownership share of the family from each reported business.

Figure 11: Number of Returns of Pass-Throughs with Net Income: SCF vs. IRS



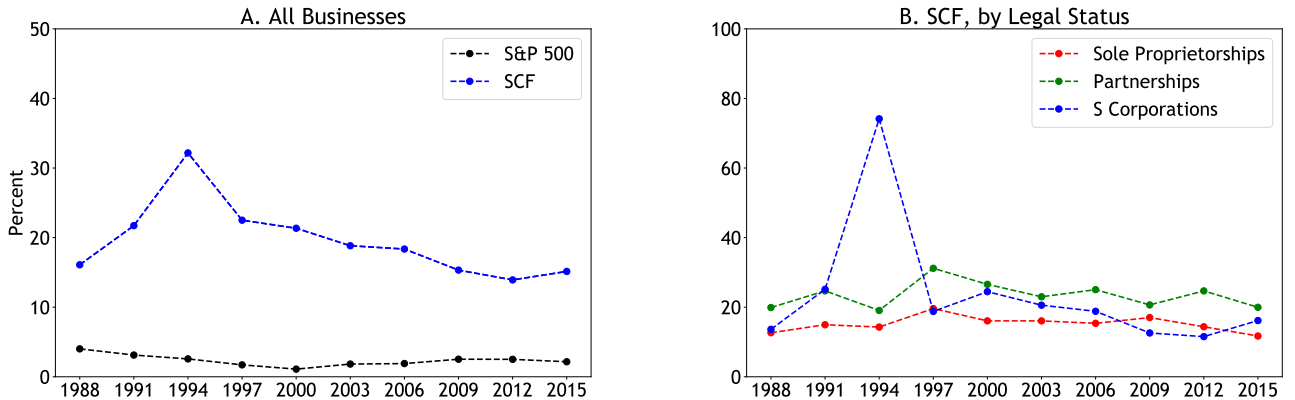
Note: This figure plots the number of returns for businesses with net income for sole proprietorships, partnerships, and S corporations over time in the IRS and the SCF. In the SCF, we assume that a proprietor that owns multiple sole proprietorships files one return. This assumption is made to be consistent 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 and S corporations, we calculate the number of returns taking into account the ownership share of the family from each reported business.

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



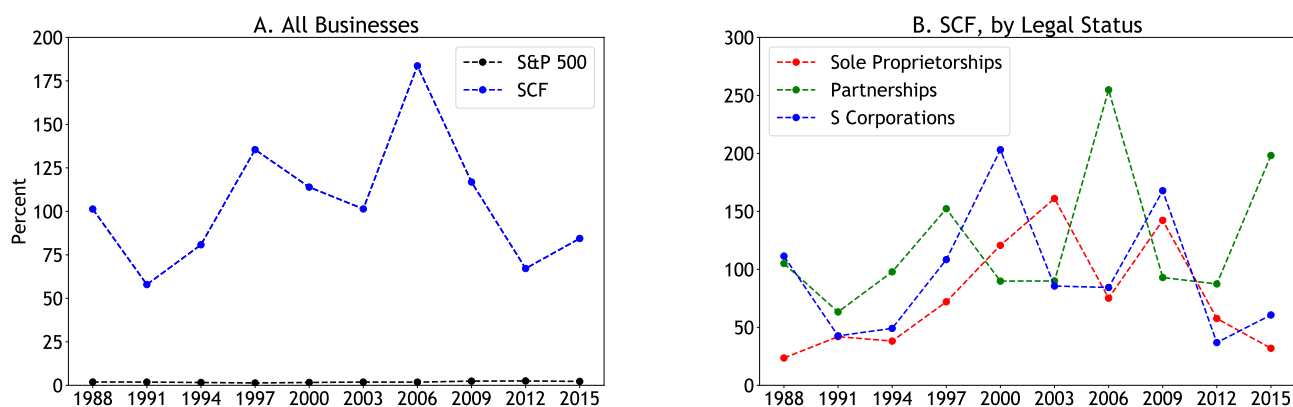
Note: This figure plots the total Form 1040 Schedule C (sole proprietorship) returns filed by individuals grouped by their AGI. Data from individual tax returns are first sorted by AGI. Then we take the total Schedule C returns filed by individuals within each pre-specified bin. AGI bins include the 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



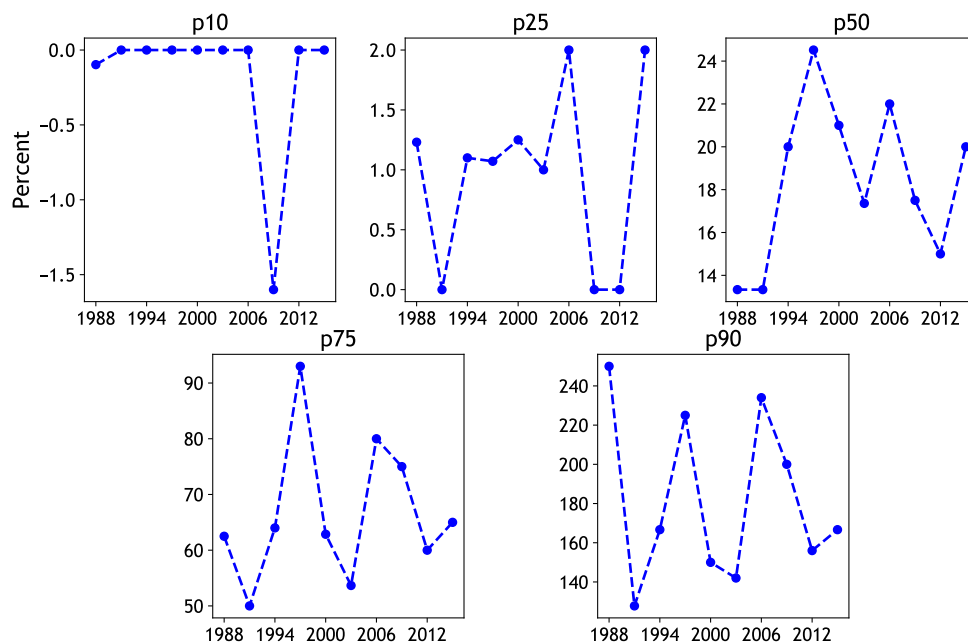
Note: This figure plots average dividend yields - calculated as the business income divided by the net worth of the business — of an index (value-weighted) of pass-through businesses in the SCF and the corresponding value-weighted average dividend yields for businesses in the S&P 500 data (Panel A) and 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



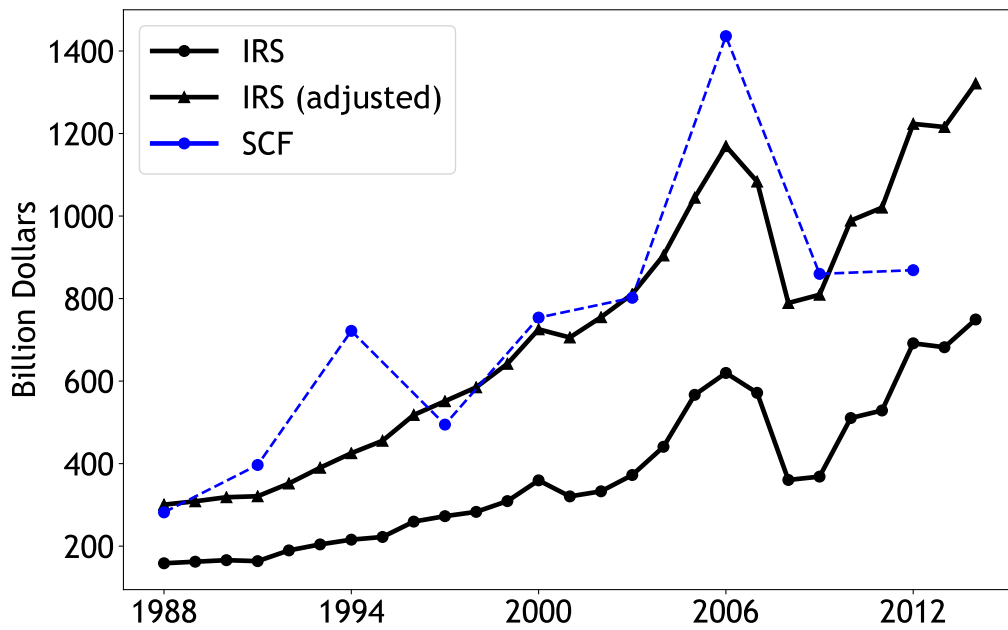
Note: This figure plots equally-weighted average dividend yields - calculated as the business income divided by the net worth of the business — for all businesses in the SCF and the equally-weighted average dividend yields for businesses in the S&P 500 data (Panel A) and businesses with different legal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



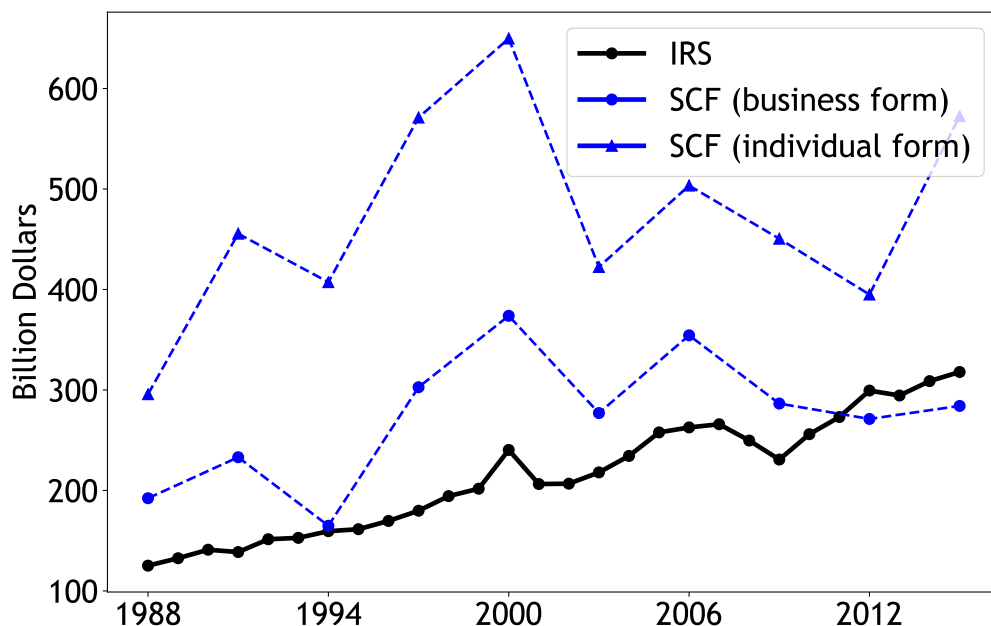
Note: This figure plots the various percentiles of the distribution of dividend yields - calculated as the business income divided by the net worth of the business — for all U.S. pass-through businesses over time. The sample includes businesses with positive net worth and excludes the bottom 1st percentile 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



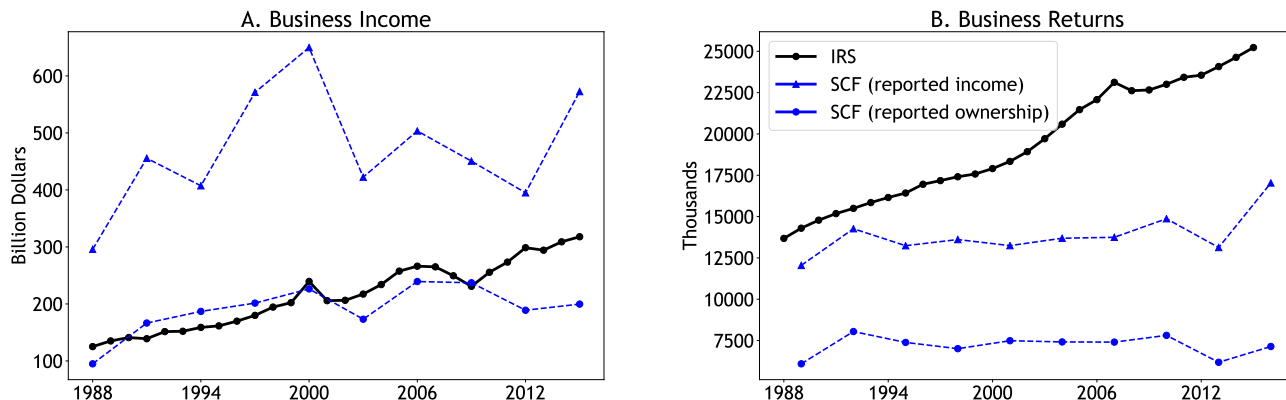
Note: In this figure, we use data on adjustments for tax misreporting on income tax documents for proprietorships and partnerships published by the U.S. Bureau of Economic Analysis. We add these yearly adjustments to the sum of proprietorship and partnership income in the IRS and compare it with estimates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