

Federal Reserve Bank of Minneapolis
Research Department Staff Report 578

Revised July 2019

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

Anmol Bhandari
University of Minnesota

Serdar Birinci
University of Minnesota

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

Kurt See
University of Minnesota

*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.

In this appendix, we provide details on the data sources and construction of variables for our analysis in “What Do Survey Data Tell Us about U.S. Businesses?” We also include the auxiliary tables and figures omitted from the main text.

1 Data Sources

The main data sources are:

- Statistics of Income of the Internal Revenue Service (SOI);
- Survey of Consumer Finances of the Board of Governors of the Federal Reserve System (SCF);
- Survey of Income and Program Participation of the U.S. Census Bureau in the Department of Commerce (SIPP);
- Panel Study of Income Dynamics of the Survey Research Center, Institute for Social Research, University of Michigan (PSID);
- Current Population Survey at the Bureau of Labor Statistics (CPS);
- Center for Research in Security Prices and Compustat (CRSP);
- Pratt’s Stats (now renamed as DealStats) from Business Valuation Resources.

Besides the main data sources listed above, we also use information from the national income and product accounts and fixed asset tables of the Bureau of Economic Analysis; financial accounts of the Board of Governors of the Federal Reserve System; Panel Study of Entrepreneurial Dynamics of the Survey Research Center, Institute for Social Research, University of Michigan; and the Kauffman Firm Survey of the Kauffman Foundation.

Table 1 lists the main variables used in our analysis: business incomes, the number of returns or owners, and business rates of return. The four columns are: (i) the variable name, (ii) the measurement concept, (iii) the database codebook or publication reference, and (iv) other remarks. In lines 1–15, we list variables that are used to construct business incomes and numbers of returns and owners from the IRS, SCF, SIPP, PSID, and CPS. In lines 16–20, we describe the variables used to construct income yields from the SCF, CRSP, and Pratt’s Stats database.

Table 1: Data Construction

Line No.	Variable	Description	Sources	Remarks
1	Adjusted gross income (IRS)	Sum of wages and salaries; net income from a business, profession, or farm; taxable and nontaxable 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; and other income.	SOI Table 1.4: <i>All Returns: Sources of Income, Adjustments, and Tax Items, by Size of Adjusted Gross Income</i> X5702+X5704+X5708+X5710 +X5712+X5714+X5716+X5718 +X5724	Data available for 1988–2015.
2	Adjusted gross income (SCF)			Data available for 1988–2015 (triennial). We reset X5704=0 for a household that does not own any actively managed sole proprietorship as identified through X3119, X3219, and X3319.
3	Sole proprietorship income, receipts, and number of returns (IRS)	Schedule C, Form 1040 [line 31] and Schedule F (farm), of Form 1040 [line 34]	SOI Table 2: <i>Nonfarm Sole Proprietorships: Income Statements, by Industrial Sectors</i> Table 1.4: <i>All Returns: Sources of Income, Adjustments, and Tax Items, by Size of Adjusted Gross Income</i>	See [1] above.
4	Sole proprietorship income, receipts, and number of returns (SCF)		Business incomes: X3132–X3232 Business receipts: X3131–X3331 Legal status: X3119–X3318 Ownership shares: X3128–X3328	Data available for 1988–2015 (triennial). For years until 2007, SCF provides all required information for three businesses, and after 2007 it provides information for two businesses.
5	Partnership income, receipts, and number of returns (IRS)	Form 1065, lines 22 and 1c	SOI Table 1: <i>All Partnerships: Total Assets, Trade or Business Income and Deductions, Portfolio Income, Rental Income, and Total Net Income, by Selected Industrial Group</i>	Data available for 1988–2014. We exclude foreign partnerships that file 1065 or 1065-B using Table 9a.
6	Partnership income, receipts, and number of returns (SCF)		See [4] above.	See [4] above.
7	S corp. income, receipts, and number of returns (IRS)	Form 1120S, lines 21 and 1c	SOI Table 7: <i>S Corporation Returns: Balance Sheet and Income Statement Items, by Major Industry</i>	See [1] above.
8	S corp. income, receipts, and number of returns (SCF)		See [4] above.	See [4] above.
9	C corp. income, receipts, and number of returns (IRS)	Form 1120, lines 30 and 1c	SOI Table 16: <i>Balance Sheet, Income Statement, Tax, and Selected Other Items, by Major Industry</i>	Data available for 1990–2013
10	C corp. income, receipts, and number of returns (SCF)		See [4] above.	See [4] above.

Table 1: Data Construction (cont.)

11	Unincorporated business income and number of owners (IRS)	Partnerships and proprietor income	See [3] and [5] above.	See [3] and [5] above.
12	Unincorporated business income and number of owners (SCF)		See [4] and [6] above. X4106, X4706 to infer self-employed members of household.	If the head and the spouse are both not self-employed but the household has an actively managed partnership business, then we infer that someone else in the household is self-employed and the owner of that business.
13	Unincorporated business income and number of owners (SIPP)		Business income: TBMSUM1-SUM2 Business shares: TPRFTB1-B2 Legal form: EINCPB1- B2 and EPROP1-B2	Data available for 2004-2006 and 2009-2013. ¹ SIPP reports information about two businesses on a monthly frequency. We average to report annual values.
14	Unincorporated business income and number of owners (PSID)		Code names vary across surveys but can be linked. For 2005 business income: ER65192, business owners: ER60932.	We use longitudinal family weights to compute aggregates.
15	Unincorporated business income and number of owners (CPS)		Business income: INCBUS Class of worker: CLASSWKR	1989-2016 Annual Social and Economic Supplement. Respondents who identify as self-employed running an incorporated business are considered business owners.
16	Income yield (SCF)	Business Income / Business Value	For business income see [4], [6], [8] above. X3129-X3329 for business valuations.	We exclude businesses with net worth less than the bottom 1st percentile of the net worth distribution, conditional on having positive net worth.
17	Income yield (CRSP/Compustat)		PI/ (csho x prcc_f) where PI: Pretax Income csho: Common Shares Outstanding prcc_f: Price Close - Annual - Fiscal	Sample includes firms incorporated in the US and for 1988-2017. Income yields are passed through a 1% winsorizing filter. For the “small” firms, we use bottom 20% of the firms by (i) market value=csho x prcc_f, (ii) book value of assets (AT), and (iii) by gross sales (SALE).
18	Income yield (Pratts Stats)		EBT/MVIC where EBT=Earnings before Taxes, MIVC= Total consideration paid to the seller	Sample truncated to exclude transactions where the EBT and MVIC are in the top 1% and bottom 1% of their respective series.
19	Income yield (PSID)		See [14] above for business income. For business valuations, code ER61736 in 2015 survey.	
20	Income yield (SIPP)		For business income, see [13] above. TVBVA1-VA2 for business valuations of assets. TVBDE1-DE2 for value of debts.	See [13] above.

¹ Prior to SIPP panel 2004, TBMSUM1 and TBMSUM2 asked respondents about total business income. However, ownership share information is only asked once a year at most in the SIPP’s topical modules. Hence, we are unable to recover aggregate business income for SIPP panels prior to 2004.

In addition to the variables listed in Table 1, we use BEA estimates of income misreporting by noncorporate businesses and General Accountability Office (GAO) estimates of income misreporting by S corporations to adjust IRS pass-through business income. BEA estimates of income misreporting over time are obtained from NIPA Table 7.14 (line 2). The GAO estimates are taken from reports GAO 14-453 and 10-195, which summarize the progress of the tax compliance studies conducted by the IRS through the National Research Program.

2 Additional Results

Next, we report on our auxiliary tables and figures that relate to our findings on business incomes, receipts, and business rates of return.

2.1 Adjusted Gross Income

A starting point for several papers in the literature is the observation that, for broad income categories, aggregated SCF responses match up well to the aggregated IRS data. In Figure 1 we construct the time series plot for adjusted gross income (AGI) from the SCF and plot it against the corresponding data from the IRS. We see that the SCF tracks the level and cyclical trends for AGI in the IRS.

2.2 Business Income

2.2.1 Aggregate business income

In Section 2.1 of the main text, we provide evidence on an overstatement of business income per return and an understatement of the number of returns across years and legal forms in the SCF relative to the IRS. In Figure 2, we report aggregate business incomes in the SCF and the IRS and show that they are also overstated in the SCF. In Figure 3, we also compare the aggregate business income for noncorporate businesses from the SCF with other surveys, namely, the SIPP, PSID, and CPS, and extend the analysis from Section 2.1 of the main text. Across all surveys, we document large discrepancies with the IRS.

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

in contrast to the SCF and IRS comparison, as the SCF overstates business income, while the KFS firms understate business income. We report estimates from their study in Table 4.

2.2.2 Business income per owner

In Section 2.1 of the main text, we discussed the accuracy of estimates of business income per owner of noncorporate businesses in the CPS, PSID, and SIPP. Figure 4 plots incomes per owner for noncorporate businesses for four surveys (SCF, CPS, PSID, and SIPP) and the IRS in Panel A and the number of owners for all surveys in Panel B.² As with the SCF, the CPS, PSID, and SIPP have higher business income per owner than is reported by the IRS, but the magnitudes are statistically different across surveys. The SCF is highest with estimates in the range of \$29,000–\$100,000, the PSID is next with a range of \$15,000–\$55,000, the CPS after that with a range of \$15,000–\$35,000, and the SIPP is lowest with a range of \$13,000–\$18,000. All are higher than the IRS, which has a range of \$5,000–\$15,000.³ The inconsistencies between surveys are driven primarily by differences in aggregate business incomes as shown in Figure 3. The number of owners across these surveys are not significantly different from each other—on the order of 10 to 13 million and stable across years—but are far lower than the IRS, which reports roughly 35 million owners in 1988 and over 50 million by 2015.⁴

2.2.3 Business income distribution

In Section 2.2.2 of the main text, we discussed the underrepresentation of business whose owners have little income. In Figure 5, we rank sole proprietors in the SCF by their AGI, assign them to income brackets using the same bins as the IRS, and plot the fractions of business income for owners with below-median AGI and for those with AGIs in the top 1st percentile. For most years, the SCF income shares for these two groups are understated and display large year by year variation. For example, the share for those with below-median AGI is nearly doubled or halved from one survey to the next. Since the fractions sum to 100 percent across all AGI groups, the SCF must necessarily overstate incomes for some bins. We find the largest overstatement of shares for those with AGIs between the 50th and 75th percentile. In Figure 6, we see that the overstatement of business income per return in the SCF data also varies a lot across years and

²Our sample in the PSID starts in 1992 and provides annual data until 1996 and biennially after that until 2014. The SIPP reports business incomes every four months for the years 2004–2006 and 2009–2012, and valuations are reported once a year for 2004, 2005, and 2009–2011 depending on when the “topical” modules are available.

³Hurst, Li, and Pugsley (2014) combine spending data from the Consumer Expenditure (CE) survey with the PSID and estimate that self-employed individuals underreport income by about 25 percent relative to an imputed measure of true income. The imputation relies on estimating the relationship between expenditures and incomes for wage and salary workers and using it along with food expenditures for the self-employed to infer “true” income of the self-employed. We instead compare survey responses directly to IRS data.

⁴As in the SCF, these surveys only account for partners who are individuals. However, as we mentioned before, using estimates from Cooper et al. (2016), this fact alone does not help to account for the massive understatement in the number of owners.

across AGI bins, with no systematic pattern. In contrast, the incomes per return in the IRS data show little variation over time and vary similarly across AGI bins. Finally, Figure 7 shows the number of sole proprietorship returns with AGIs per return below and above the median. For businesses that have owners with below-median AGIs, the number of IRS returns has risen from about 5 million in 1988 to over 12 million in 2015, but the SCF estimate has remained at roughly 2 million for the entire period. For businesses with above-median AGIs, the number of IRS returns has risen from a little over 8 million to above 12 million, but the SCF estimate has hovered around 5 million.

In Section 2.2.3, we discussed the distribution of business income by splitting pass-through businesses into two categories: those that make profits and those that make losses (or no income). In Figures 8 and 9, we plot business income per return by legal status for those making profits and losses, respectively. In Figures 10 and 11, we plot the number of returns for the same sets of businesses. In Table 5, we extend the analysis of decomposing the total percentage error into the overstatement of profits and understatement of losses. Table 6 shows the distribution of losses by AGI bins for tax year 2015. We see that 10 out of 19 bins, which account for 23 percent of the total number of returns and 26 percent of the total losses in the IRS, have an aggregate zero (that is, all respondents in those income brackets reported a zero net income) in the SCF data. In Figure 12, we report the distributional statistics for S corporations. As we noted in the main text, the data for S corporations are only available for limited years, namely 2003–2012, but these data show similar inconsistencies between SCF and IRS data, as was found with sole proprietorships. All of these results show that the distribution of business income in the SCF is largely inconsistent with its counterpart in the IRS, and that the inconsistencies vary across survey years.

2.2.4 Breakdown of proprietorships in the SCF

In Figure 3 of the main text we used a Venn diagram to split sole proprietorship income and counts for the year 2015 in several categories: (A) those who have a non-zero line 12 plus 18 on Form 1040 (B) those who are actively managing and report line 31 of Schedule C and (C) those reporting to have an interest in business without an active-management role. In Table 2 we provide details for all survey years.

2.2.5 Misclassification of business income

In Figure 13, we extend the analysis of Section 2.2.5 in the main text by plotting for all years a measure of broad business income consistent with Johnson and Moore (2008). 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). As

we noted in the main text, the SCF estimates are still larger in all years than the IRS counterpart even with the broader concept of income. Meanwhile, Figure 14 shows that the same conclusion holds when we restrict our definition of broad business income to exclude capital gains and include only Schedule C, F, and E income. These findings imply that miscategorization of income across different types does not explain the overstatement of business income in the survey data.

2.2.6 Evidence on mismeasurement

Section 2.2 of the main text discusses possible reasons for the overstatements of business income in survey data. A reason to be suspicious about misreported incomes in the SCF is that a very small fraction of respondents refer to their tax documents when responding to questions about the specific line items on tax forms. To verify whether respondents in the SCF check documents, we use variable X6536, which provides information on the frequency of checking any documents when answering interview questions. Variable X7451 informs us about whether the respondent referred to income tax documents, and variables X7452 through X7455 inform us about whether the respondent referred to other financial documents, namely, pension documents, account statements, investment or business records, and loan documents, respectively. If a respondent says that he or she checked the income tax document ($X7451=1$), we use his or her answers to X6536 to obtain the frequency of checking this document. The respondent did not check the income tax document if either ($X7451=5$ or $X7451=0$ or $X7451=-7$) or ($X6536=4$). We use the same steps to check referencing of other financial documents by using X7452–X7455 instead of X7451. We classify a respondent who checks at least one of these four documents as someone who refers to any other tax documents. We then obtain the weighted fraction of the group of respondents who check these two types of documents frequently, sometimes, rarely, or never. Roughly 4 percent of all respondents have nonapplicable responses (NaN). We adjust for this nonresponse rate in the results of the main text so that our fractions sum to 100 percent.

We calculate the frequency with which business owners referenced either tax or other financial documents in tax year 2015.⁵ These tabulations are shown in Table 7. The first row shows that 75 percent of business owners in the SCF never referenced tax documents, 2 percent rarely did, 9 percent sometimes did, and 14 percent frequently did. The second row shows that 64 percent never referenced any other financial documents, 6 percent rarely did, 15 percent sometimes did, and 15 percent frequently did.

In the main text, we assert that non referencing of tax documents is uniform across business owners. To show this we use tax year 2015 and group owners by their AGI and by total business income. In all cases, we find the fraction of owners who never reference a tax document to be very high about 75 percent and not

⁵ Other financial documents include account statements, investment and business records, loan documents, and pension documents. If any of these documents are referenced, we assume all are.

varying too much across groups. The information is summarized in Table 3.

To provide further evidence on measurement errors, we show that the SCF fails a simple consistency check by comparing answers to two closely related questions. In the case of sole proprietors, respondents are asked to report incomes listed on lines 12 and 18 of their Form 1040, which correspond to Schedule C and F incomes, respectively. Separately, they are asked about business income from a sole proprietorship and told it is listed on line 31 of Schedule C.⁶ By design, the difference in responses to these two questions must be farm income from Schedule F. In Figure 15, we see that the differences across the two answers vary between \$17,000 and \$40,000 per return, considerably more than could be attributable to farm incomes. In a typical year, only 4 percent of business profits listed on Form 1040 are farm income, and farm losses exceed profits in many of the years of our sample.

2.3 Business Receipts

In this section, we extend the analysis in the main text to business receipts. Figure 16 and 17 compare aggregate business receipts and business receipts per return across legal forms and across years respectively in the SCF and the IRS. We again find large and variable errors in the SCF responses when compared to the IRS counterparts. For example, in the case of pass-through businesses, we find that the average error in business receipts per return over the period 1988–2015 is 169 percent, with errors over the period in the range of 89 percent to 367 percent. Thus, our main finding is an overstatement of aggregated business incomes and receipts in the SCF across all legal forms, with large variation in the discrepancies across survey years.

2.4 Business Valuations and Rates of Return

In this section, we provide additional details for the comparison of the income yields in SCF to CRSP-Compustat, Pratt’s Stats, and other surveys to augment the analysis in Section 3 of the main text.

We begin by formally describing the measurement of SCF income yields. The SCF income yield, which is computed for each business, is the ratio of total pretax net income from businesses divided by the self-reported total net worth of businesses. Let $\{\omega_{i,t}\}$ be the SCF population weights for survey year t . We compute an equally weighted and value-weighted mean yield for t , denoted as R_t^{ew} and R_t^{vw} , respectively:

$$R_t^{ew} = \sum_i \omega_{i,t} \left(\frac{NI_{i,t}}{V_{i,t}} \right), \quad R_t^{vw} = \sum_i \left(\frac{\omega_{i,t} V_{i,t}}{\sum_i \omega_{i,t} V_{i,t}} \right) \left(\frac{NI_{i,t}}{V_{i,t}} \right), \quad (1)$$

where NI is total pretax net income and V is the self-reported total business value.

⁶The first answer is coded as X5704 and the second as X3132, X3232, and X3332, combined with the response to legal status of the actively managed business with codes X3119, X3219, and X3319.

In the main text, we showed evidence that the SCF income yields are high when compared to CRSP-Compustat or Pratt’s Stats. In Table 8, we provide several additional moments for the distribution of income yields in the SCF. The additional moments show that SCF income yields are high regardless of year or legal structure.

In the main text, we compared the income yields for S and C corporations in the SCF to small firms in CRSP where we defined “small” as corporations that are in the bottom quintile of the size distribution as measured by the book value of total assets. In Table 9, we extend the analysis to two alternative definitions of “small”: (i) those in the bottom quintile by market value and (ii) those in the bottom quintile by gross sales. Although there are some differences in the magnitudes compared to Table 1 in the main text, the equally weighted and value-weighted yields are still negative in all years, regardless of how we classify the small firms.

Income yields for all businesses as well as non-tech and non-distressed firms obtained from Pratt’s Stats were discussed in the main text. We extend this discussion with Table 10 which reports income yields from Pratt’s Stats for all legal forms. We see that sole proprietors have higher yields than other pass-throughs and C corporations. However, since these businesses have much smaller valuations, the value-weighted yield for all businesses is relatively low when compared to SCF data.

In the main text, we noted that the average yields are comparable across the SCF, PSID, and SIPP, while the distributions are not. In Tables 11 and 12, we report the income yields in PSID and SIPP for all years that the data are available. These tables more clearly demonstrate this finding.

Finally, in the main text, we compare our result that income yields in survey data are overstated to Moskowitz and Vissing- Jorgensen (2002) who conclude using SCF data that private business returns were surprisingly low. We show that the differences in our results is explained by Moskowitz and Vissing- Jorgensen’s (2002) imputation method used to calculate capital gains. Below, we provide more details on how we reached to this conclusion.

In theory, one would need a panel of firm valuations to compute a value-weighted capital gain, namely,

$$R_{t+1}^{cg} = \sum_i \left(\frac{\omega_{i,t} V_{i,t}}{\sum_i \omega_{i,t} V_{i,t}} \right) \left(\frac{V_{i,t+1}}{V_{i,t}} \right), \quad (2)$$

using survey weights $\{\omega_{i,t}\}$ and valuations $\{V_{i,t}\}$ for each firm i in year t . Given that the SCF survey is triennial with virtually no panel aspect (other than two surveys), there is no way to compute $V_{i,t+1}/V_{i,t}$ firm by firm. Moskowitz and Vissing-Jorgensen (2002) instead compute their capital gains measure using the following annualized index:

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

Their concept of rate of return is given by $R_t^{vw} + \tilde{R}_t^{cg}$, where R_t^{vw} is defined in (1). They adjust the SCF net income by subtracting imputed measures of taxes and retained earnings and compare their measure of return to the value weighted mean holding period return on the CRSP index portfolio.⁷ This procedure generates private returns that are similar in magnitude to the CRSP returns.

As discussed in the main text, we replicate the exercise of Moskowitz and Vissing-Jorgensen (2002) for our full sample with income yields and capital gains compared separately. We find that the capital gain imputation drives the differences between our findings and theirs. The full results that support the discussion in Section 3.1 of the main text are summarized in Table 13. The first two columns show estimates of SCF and CRSP-Compustat income yields, R_t^{vw} , in all SCF survey years. The last three columns show estimates of \tilde{R}_t^{cg} for SCF and both R_t^{cg} and \tilde{R}_t^{cg} for the CRSP-Compustat sample.

The tables reveals two important results. First, SCF yields are substantially higher than the CRSP-Compustat counterparts for all survey years. Second, the annualized SCF capital gains vary substantially less than those for firms in the CRSP-Compustat gains R_t^{cg} over the sample, which is not surprising given the conceptual differences in the measures and the long interval between survey years.⁸ If we were to add R_t^{vw} plus \tilde{R}_t^{cg} for SCF and R_t^{vw} plus R_t^{cg} for CRSP-Compustat firms, we would confound two discrepancies and conclude that the private and public returns are not very different on average: 26 percent for SCF versus 21 percent for CRSP-Compustat.

If we were to restrict attention to comparable measures, either R_t^{vw} or $R_t^{vw} + \tilde{R}_t^{cg}$, we would instead conclude that the private business yields and the imputed total returns are relatively high for private businesses when compared to public returns, not low as previously thought.

⁷Since the assumptions underlying the imputations of taxes and retained earnings are ad hoc, we measure R_t^{vw} using pretax income in both the SCF and CRSP samples.

⁸Incidentally, the time variation in the capital gains components explains why Moskowitz and Vissing-Jorgensen (2002) and Kartashova (2014) estimate different average returns for the different sample periods they study.

Table 2: Sole Proprietorships in the SCF

Tax Year	Incomes					Counts				
	A	B	C	$A \cap B$	$A \cap C$	A	B	C	$A \cap B$	$A \cap C$
1988	297.3	176.0	75.7	95.0	75.3	10.2	6.1	1.2	4.3	1.0
1991	456.8	283.0	171.5	167.0	160.8	11.1	8.0	0.9	4.9	0.7
1994	409.7	229.0	56.3	187.0	55.7	10.6	7.4	0.9	4.8	0.7
1997	575.6	329.0	77.8	203.0	73.6	11.1	7.0	0.8	4.5	0.7
2000	652.7	405.0	89.5	228.0	88.7	10.1	7.5	0.9	4.3	0.8
2003	424.2	249.6	57.3	174.0	57.2	11.2	7.4	0.6	4.9	0.5
2006	506.5	270.0	51.0	239.0	50.5	12.3	7.4	0.5	6.0	0.5
2009	452.5	241.4	NA	237.0	33.5	14.0	7.8	NA	6.9	0.5
2012	401.6	256.4	NA	189.0	37.6	12.0	6.2	NA	5.1	0.5
2015	583.0	229.0	NA	206.0	74.0	16.2	7.1	NA	6.3	0.7

Note: This table shows business income and counts for three sets of sole proprietors: (A) those who have a non-zero line 12 plus 18 on Form 1040, (B) those who report to be actively managing, and (C) those reporting having an interest in business without an active-management role. Some of the groups overlap and the columns with headings $A \cap B$ and $A \cap C$ list the intersection of the overlapping sets. The NAs for tax years 2009, 2012, and 2015 column with heading C are missing information because the SCF stopped identifying legal form of organization for passive owners after 2007.

Table 3: Distribution of non referencing in the SCF

Groups	Fraction of non referencing owners
By AGI	
<p25	0.77
p25-p50	0.79
p50-p75	0.75
>p75	0.71
By Business income	
< p25	0.72
p25-p50	0.72
p50-p75	0.80
> p75	0.77
non positive	0.70
positive	0.76

Note: This table summarizes non referencing for survey year 2016. Households owning an actively-managed business are ranked by their AGI and by their total business income into 4 bins with p25, p50, p75 representing the 25th percentile, the 50th percentile, and the 75th percentile. For each bin, we compute fraction of households that did not check their income tax form. The row “non positive” are households that actively manage a business and have total business income less than or equal to zero. The row “positive” are households that actively manage a business and have total business income greater than zero.

Table 4: Comparison of KFS and IRS Business Tax Data, 2004–2011

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

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

Table 5: SCF-IRS Business Income Gap by Legal Structure

Tax Year	SCF-IRS Gap (\$)	Percentage of Gap from	
		Overstatement of Profits (%)	Understatement of Losses (%)
Sole Proprietorship			
1988	67.09	58	42
1991	94.36	67	33
1994	5.44	-515	615
1997	122.91	71	29
2000	168.09	75	25
2003	59.06	5	95
2006	91.66	29	71
2009	55.72	-38	138
2012	-28.22	359	-259
2015	-33.74	350	-250
Mean	60.24	46	54
Partnership			
1988	56.28	37	63
1991	138.70	67	33
1994	500.59	92	8
1997	99.05	30	70
2000	261.03	56	44
2003	370.45	68	32
2006	724.62	83	17
2009	435.59	35	65
2012	205.51	0	100
Mean	310.20	52	48
S Corporation			
1988	35.78	57	43
1991	73.53	53	47
1994	118.07	74	26
1997	163.99	77	23
2000	206.06	78	22
2003	355.15	86	14
2006	279.35	77	23
2009	258.94	68	32
2012	41.06	-53	153
Mean	170.21	57	43
C Corporation			
1991	-85.35	261	-161
1994	-244.42	148	-48
1997	-339.64	139	-39
2000	-57.00	670	-570
2003	-267.37	212	-112
2006	-859.87	123	-23
2009	-236.88	323	-223
2012	-747.36	138	-38
Mean	-354.74	252	-152

Note: This table shows the difference (gap) between aggregated business income by legal structure in the SCF and IRS. The gap is then decomposed into the fraction attributable to an overstatement of profits or an understatement of losses. Dollar amounts are in billions.

Table 6: Sole Proprietorships with Net Losses in the IRS and SCF by AGI Bins, 2015

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

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

Table 7: Percentage of Respondents Checking Documents in SCF 2016

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

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

Table 8: Net Income Yields in the SCF by Legal Structure

Tax Year	Sole Proprietorship					Partnership					
	Value-Weighted Mean	Equally Weighted Mean	p25	p50	p75	Value-Weighted Mean	Equally Weighted Mean	p25	p50	p75	
1988	19.9	105.0	3.2	20.0	80.0	13.6	111.4	0.0	8.0	50.0	
1991	24.7	63.3	0.2	15.0	52.0	25.1	42.6	0.0	4.4	24.1	
1994	19.1	97.8	2.0	24.0	74.0	74.1	49.1	0.3	10.7	42.3	
1997	31.2	152.2	2.2	29.5	100.0	18.8	108.4	0.8	16.4	60.0	
2000	26.6	89.8	0.9	25.5	75.0	24.5	203.1	0.1	11.9	40.0	
2003	23.0	90.0	3.0	25.0	70.0	20.6	85.6	0.0	5.0	30.0	
2006	25.0	254.8	2.3	32.0	100.0	18.8	84.4	0.1	10.0	40.0	
2009	20.7	92.9	1.6	27.2	93.3	12.6	167.8	0.0	4.5	40.0	
2012	24.7	87.4	0.0	23.2	82.4	11.5	36.8	0.0	5.4	33.7	
2015	20.0	198.2	2.6	32.5	100.0	16.2	60.6	1.0	12.0	48.8	
Mean	23.5	123.1	1.8	25.4	82.7	23.6	95.0	0.2	8.8	40.9	
		S Corporation					C Corporation				
1988	12.7	23.5	0.5	6.0	37.5	17.8	101.7	3.2	16.7	30.5	
1991	15.0	42.0	0.0	11.2	43.6	15.5	45.1	0.0	9.0	32.0	
1994	14.3	38.1	0.9	11.7	40.0	28.3	73.9	0.4	8.0	41.1	
1997	19.6	72.0	0.1	15.8	76.0	15.5	92.4	5.3	20.8	62.2	
2000	16.1	120.7	4.4	18.4	40.0	26.5	90.8	2.9	15.8	46.0	
2003	16.1	161.1	4.0	14.2	40.0	11.3	13.9	0.0	4.4	12.9	
2006	15.4	75.1	3.8	16.7	80.0	16.3	44.4	0.0	7.5	36.0	
2009	17.0	142.3	0.0	13.3	58.1	11.5	23.8	0.0	5.4	23.3	
2012	14.4	57.6	2.7	15.2	52.2	15.4	55.4	0.0	9.0	41.3	
2015	11.7	31.9	5.9	19.8	37.5	10.9	27.1	1.3	9.7	36.5	
Mean	15.2	76.4	2.2	14.2	50.5	16.9	56.8	1.3	10.6	36.2	
		All Pass-throughs					All Businesses				
1988	16.1	101.3	1.2	13.3	62.5	16.6	101.3	1.3	14.3	57.0	
1991	21.7	57.9	0.0	13.3	50.0	20.7	67.2	0.0	13.2	43.6	
1994	32.2	80.8	1.1	20.0	64.0	31.5	80.8	1.1	19.0	62.9	
1997	22.5	135.5	1.1	24.5	93.0	20.6	148.9	1.7	24.7	86.7	
2000	21.3	113.9	1.3	21.0	62.9	22.6	114.4	1.6	20.0	62.3	
2003	18.8	101.4	1.0	17.4	53.7	17.7	81.1	0.2	14.9	50.0	
2006	18.4	183.7	2.0	22.0	80.0	18.1	171.7	1.6	20.0	73.3	
2009	15.3	116.9	0.0	17.5	75.0	14.8	111.7	0.0	16.0	70.0	
2012	13.9	67.1	0.0	15.0	60.0	14.1	66.2	0.0	15.0	60.0	
2015	15.1	84.4	2.0	20.0	65.0	14.6	81.5	1.8	19.4	64.0	
Mean	19.5	104.3	1.0	18.4	66.6	19.1	102.5	0.9	17.6	63.0	

Note: This table shows moments of the net income yield distribution of businesses in the SCF by legal structure. 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 in the SCF is obtained from SCF variables that correspond to information on business tax forms.

Table 9: Income Yield for Small Firms in CRSP

Tax Year	by Market Capitalization					by Sales				
	EW	VW	p25	p50	p75	EW	VW	p25	p50	p75
1988	-43.6	-27.0	-52.3	-14.3	6.1	-27.2	-8.8	-26.3	-8.6	1.1
1991	-72.9	-49.0	-72.4	-15.9	5.1	-31.7	-6.0	-23.3	-5.6	1.5
1994	-23.3	-14.2	-34.1	-4.1	9.3	-18.1	-9.2	-24.8	-6.6	4.0
1997	-29.9	-19.2	-43.2	-8.5	7.1	-21.1	-8.5	-25.4	-8.0	2.7
2000	-104.1	-71.8	-103.4	-16.4	10.4	-52.8	-12.4	-42.2	-10.7	2.2
2003	-14.2	-9.2	-21.0	-0.9	7.8	-9.5	-7.2	-15.2	-3.3	5.5
2006	-12.1	-8.1	-20.8	-0.2	7.6	-11.9	-8.6	-18.6	-5.1	4.7
2009	-65.0	-47.3	-72.4	-22.5	4.7	-32.6	-11.0	-34.6	-10.8	3.0
2012	-22.7	-12.6	-35.6	-3.8	10.4	-17.1	-5.7	-22.7	-5.4	6.7
2015	-59.6	-35.6	-55.4	-11.5	6.3	-37.6	-11.5	-35.8	-11.9	1.9
Mean	-44.7	-29.4	-51.1	-9.8	7.5	-25.9	-8.9	-26.9	-7.6	3.3

Note: This table shows estimates of income yields for small businesses in CRSP-Compustat firms. The column “EW” reports the equally weighted average, the column “VW” reports the value-weighted average, the column “p25” reports the 25th percentile, the column “p50” reports the 50th percentile, and the column “p75” reports the 75th percentile.

Table 10: Income Yield from Pratt’s Stats

Legal Form	EW	VW	p25	p50	p75
Sole Proprietorship	41.3	31.6	13.3	36.7	61.5
Partnership	26.6	4.8	2.7	20.5	48.8
S Corporation	30.3	6.9	6.5	23.3	47.8
C Corporation	6.8	-2.1	-2.3	6.5	29.8

Note: This table shows estimates of income yields from the Pratt’s Stats database. The column “EW” reports the equally weighted average, the column “VW” reports the value-weighted average, the column “p25” reports the 25th percentile, the column “p50” reports the 50th percentile, and the column “p75” reports the 75th percentile.

Table 11: Net Income Yields of Unincorporated Businesses in the PSID

Tax Year	Value-Weighted Mean	Equally Weighted Mean	p25	p50	p75
1998	5.2	136.4	0.0	12.5	75.0
2000	21.7	182.4	0.0	7.5	73.3
2002	21.8	187.0	0.0	33.3	139.5
2004	22.2	287.7	3.9	36.9	140.0
2006	20.6	630.1	10.0	42.5	222.2
2008	10.9	175.8	2.7	28.8	125.0
2010	13.9	110.3	3.9	25.0	75.9
2012	10.7	90.8	3.3	23.0	83.3
2014	6.9	182.9	4.8	33.3	100.0
Mean	14.9	220.4	3.2	27.0	114.9

Note: This table shows moments of the net income yield distribution of unincorporated businesses in the PSID. The sample includes businesses with positive net worth and excludes the bottom 1st percentile of these businesses.

Table 12: Net Income Yields in the SIPP

Tax Year	Value-Weighted Mean	Equally Weighted Mean	p25	p50	p75
Sole Proprietorship					
2004	20.2	545.0	6.8	44.8	240.0
2005	19.4	727.7	4.5	41.2	240.0
2009	13.0	3043.1	0.2	24.0	203.3
2010	15.8	5916.6	0.2	31.0	240.0
2011	14.9	8878.2	0.5	29.2	188.0
Mean	16.7	3822.1	2.4	34.0	222.3
Partnership					
2004	25.1	605.9	0.6	29.2	220.0
2005	19.9	1271.2	0.3	22.6	189.1
2009	17.4	853.4	0.0	7.4	108.0
2010	21.3	2128.0	0.0	22.5	204.0
2011	18.9	1551.7	0.0	11.8	190.7
Mean	20.5	1282.0	0.2	18.7	182.4
Unincorporated					
2004	22.0	2936.2	6.4	45.7	260.0
2005	19.8	12590.7	4.0	40.4	250.0
2009	14.0	15353.1	0.1	22.5	202.5
2010	17.2	38737.5	0.1	30.8	240.0
2011	15.3	7971.4	0.3	26.7	197.8
Mean	17.6	15517.8	2.2	33.2	230.1

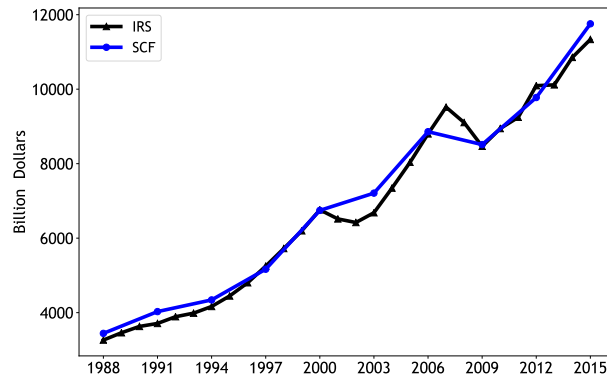
Note: This table shows moments of the net income yield distribution of sole proprietorships, partnerships, and unincorporated businesses in the SIPP 2004 and 2008 panels. Statistics are calculated for years where asset topical modules are available. The sample includes businesses with positive net worth and excludes the bottom 1st percentile of these businesses.

Table 13: Net Income Yields and Capital Gains

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

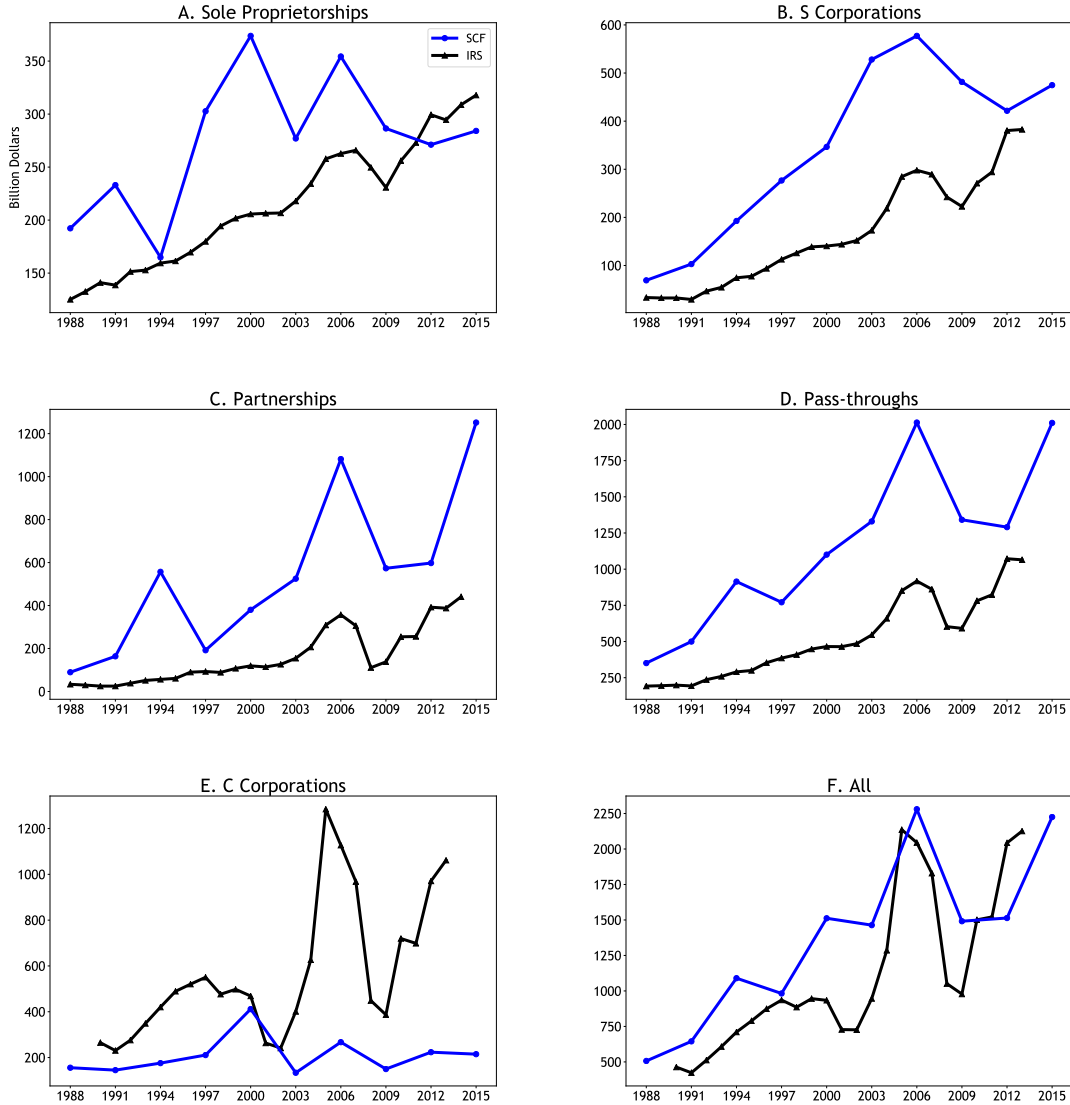
Note: This table shows estimates of income yields and capital gains for businesses in the SCF and CRSP-Compustat firms. For the SCF, capital gains are computed using Equation 3 found in the main text, as in Moskowitz and Vissing-Jorgensen (2002). For the CRSP-Compustat firms, we report two measures of capital gains. The column $(t-1) \rightarrow t$ measures the realized capital gains using Equation 2 for year t where t corresponds to the fiscal year for which income is reported in the SCF. The column $(t-3) \rightarrow t$ measures a geometric mean of the capital gains for the index over the past three periods using equation 3.

Figure 1: Adjusted Gross Incomes: SCF vs. IRS



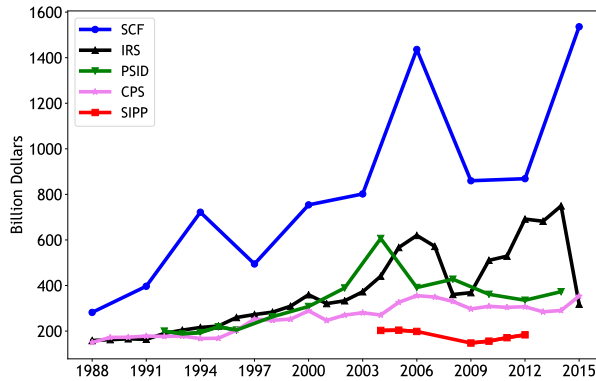
Note: For the IRS, adjusted gross income is obtained from Form 1040. For the SCF, if AGI is not available, we construct it by adding the appropriate income categories.

Figure 2: Business Income by Legal Status, SCF vs. IRS



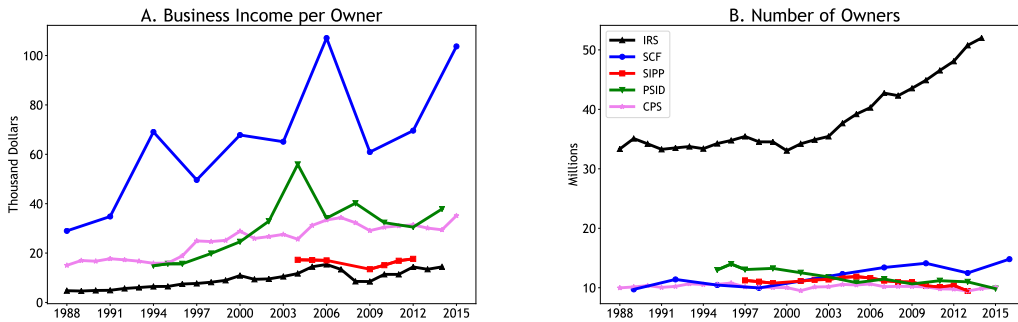
Note: This figure plots the total business income by legal status in the SCF and the IRS. Business income refers to income reported on Form 1040 Schedule C for sole proprietorships, Form 1065 for partnerships, Form 1120S for S corporations, and Form 1120 for C corporations. IRS data for partnerships, S corporations, and C corporations are available only until 2013, and C-corporation data start from 1990 because data for Form 1120 are not available for 1988 and 1989.

Figure 3: Total Unincorporated Business Income in SCF, SIPP, PSID, and CPS vs. IRS



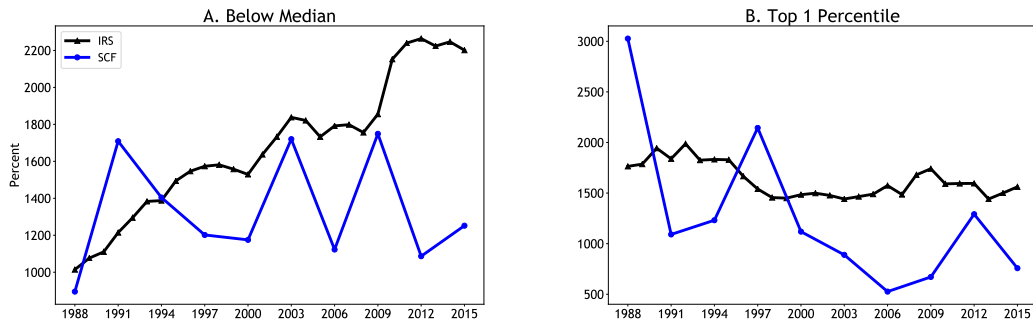
Note: This figure plots the total business income of unincorporated businesses in the SCF, SIPP, PSID, CPS, and IRS. Before 2004, the SIPP does not provide information about an individual's own share of business income from an unincorporated business. Instead, it contains information about the total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

Figure 4: Unincorporated Business Income per Owner and Number of Owners



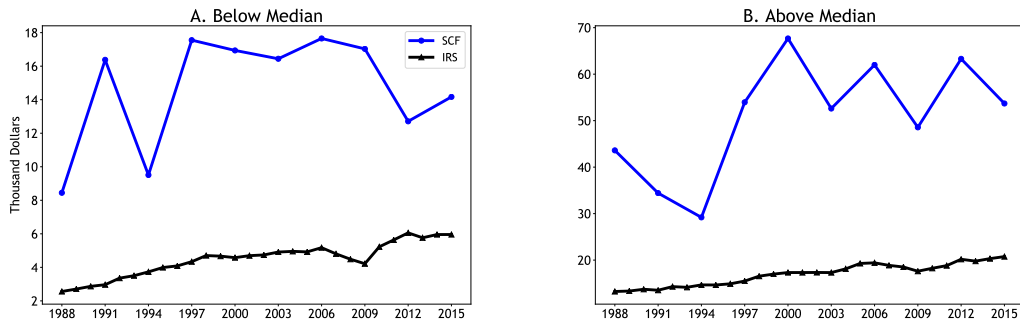
Note: This figure plots the total business income per owner of unincorporated businesses (Panel A) and total number of unincorporated business owners (Panel B) in the SCF, CPS, PSID, SIPP, and the IRS. Before 2004, the SIPP does not provide information about an individual's own share of business income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

Figure 5: Proprietor Income Shares: SCF vs. IRS



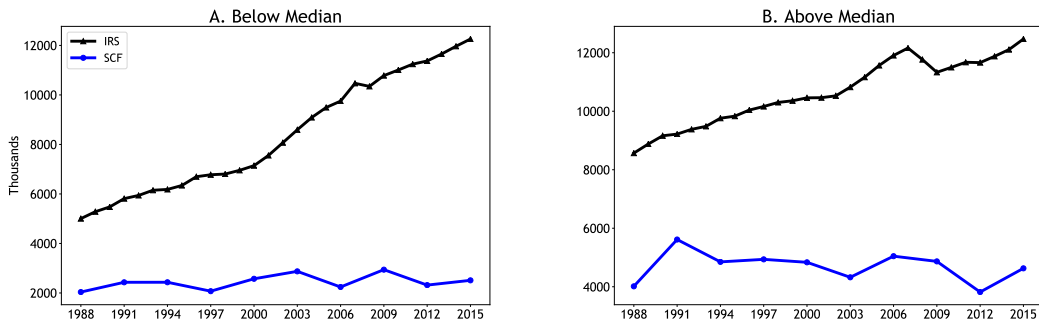
Note: This figure plots the fraction of business income from sole proprietorships attributable to returns with AGI below the median and above the 99th percentile.

Figure 6: Income Per Return, Proprietors with Below- and Above-Median AGI: SCF vs. IRS



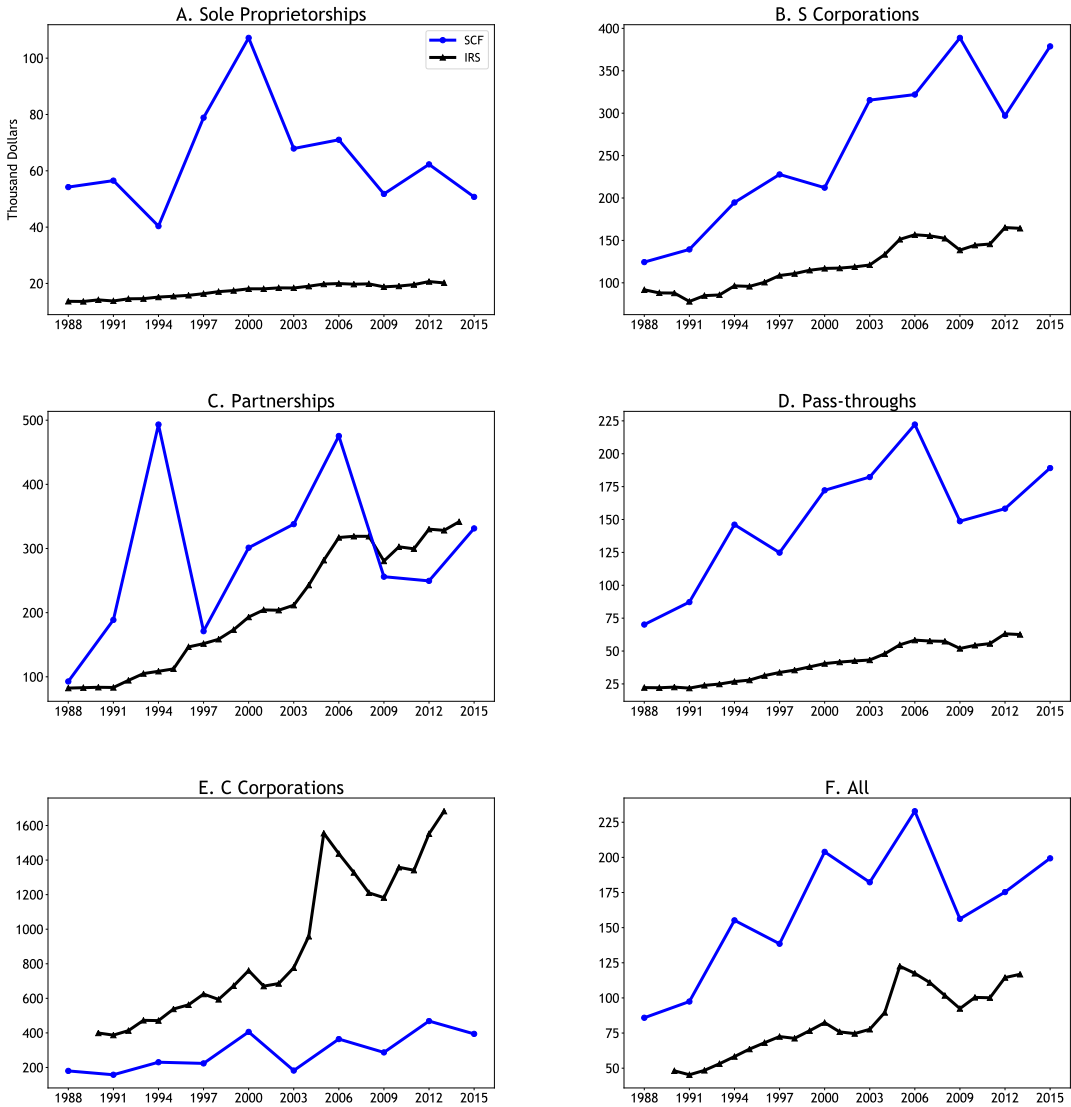
Note: This figure plots sole proprietorship business income per return for those with below- and above-median AGI.

Figure 7: Number of Returns, Proprietors with Below- and Above-Median AGI: SCF vs. IRS



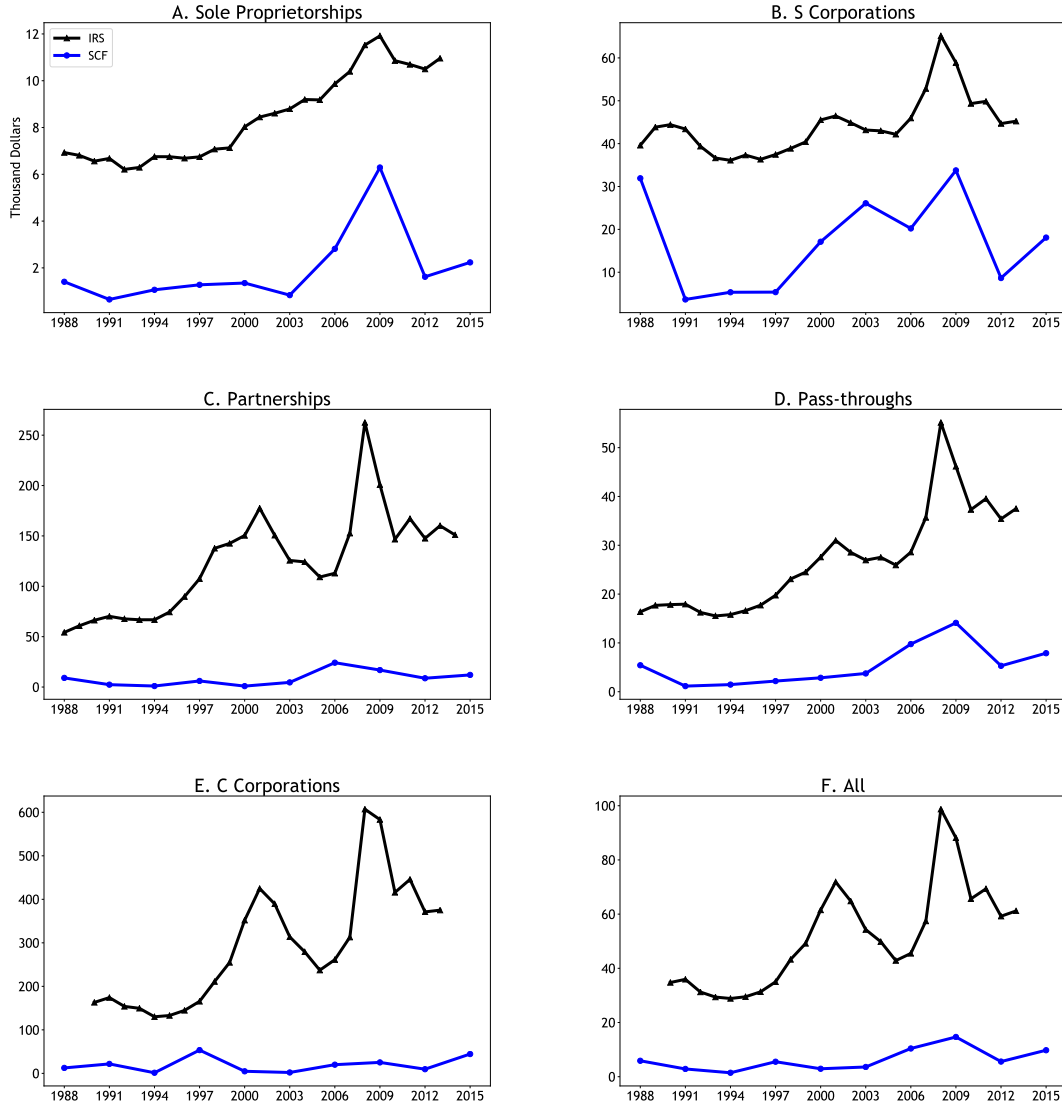
Note: This figure plots the number of sole proprietorship returns (Form 1040 Schedule C) filed by business owners with below- and above median AGI.

Figure 8: Business Income per Tax Return by Legal Status for Businesses with Net Income, SCF vs. IRS



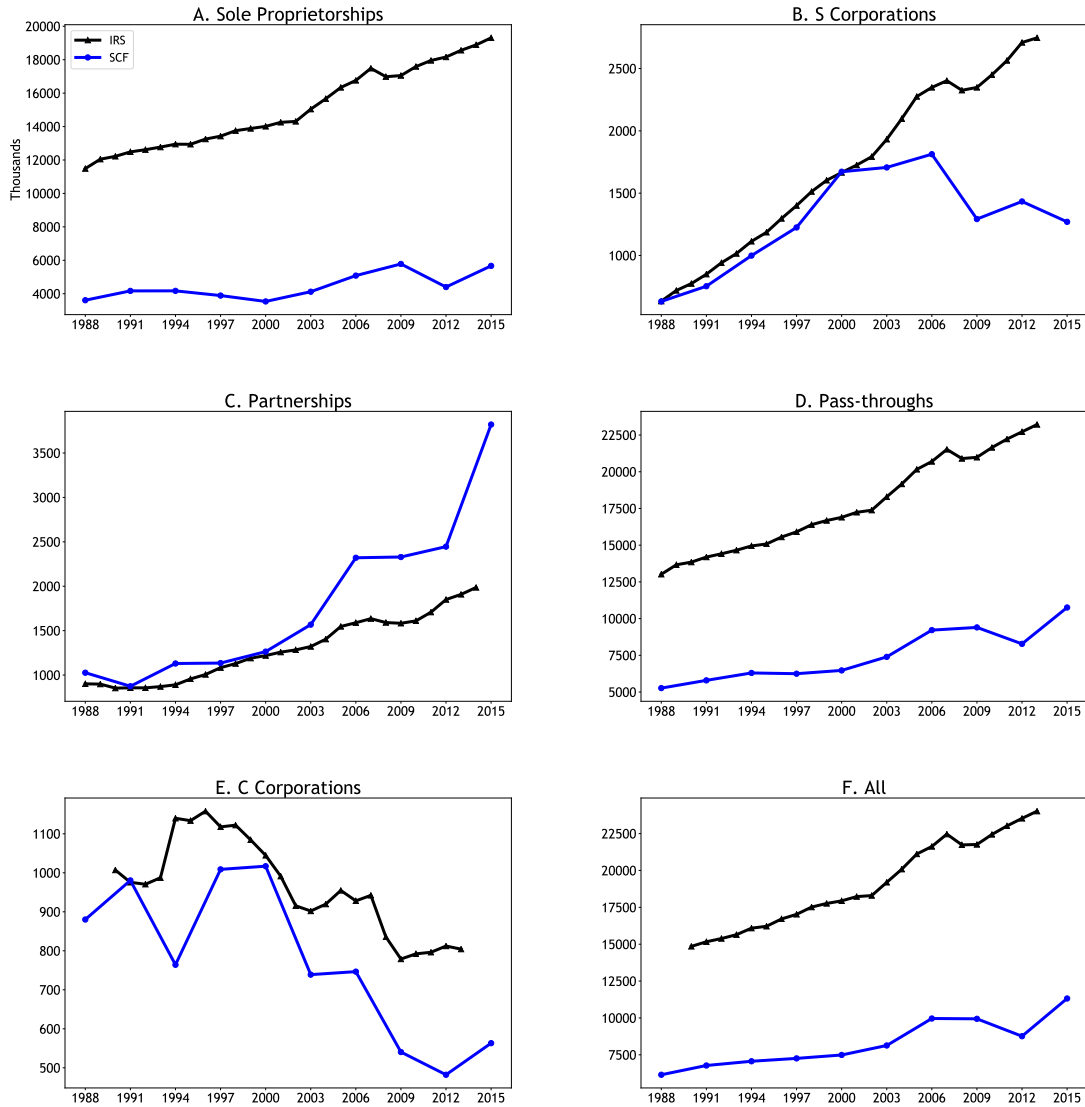
Note: This figure plots the business income per tax return by legal status for businesses with net income in the SCF and IRS. Business income refers to income reported on Form 1040 Schedule C for sole proprietorships, Form 1065 for partnerships, Form 1120S for S corporations, and Form 1120 for C corporations. IRS data for sole proprietorships, partnerships, S corporations, and C corporations are available only until 2013, and C-corporation data start from 1990 because data for Form 1120 are not available for 1988 and 1989.

Figure 9: Business Income per Tax Return by Legal Status for Businesses with Net Loss, SCF vs. IRS



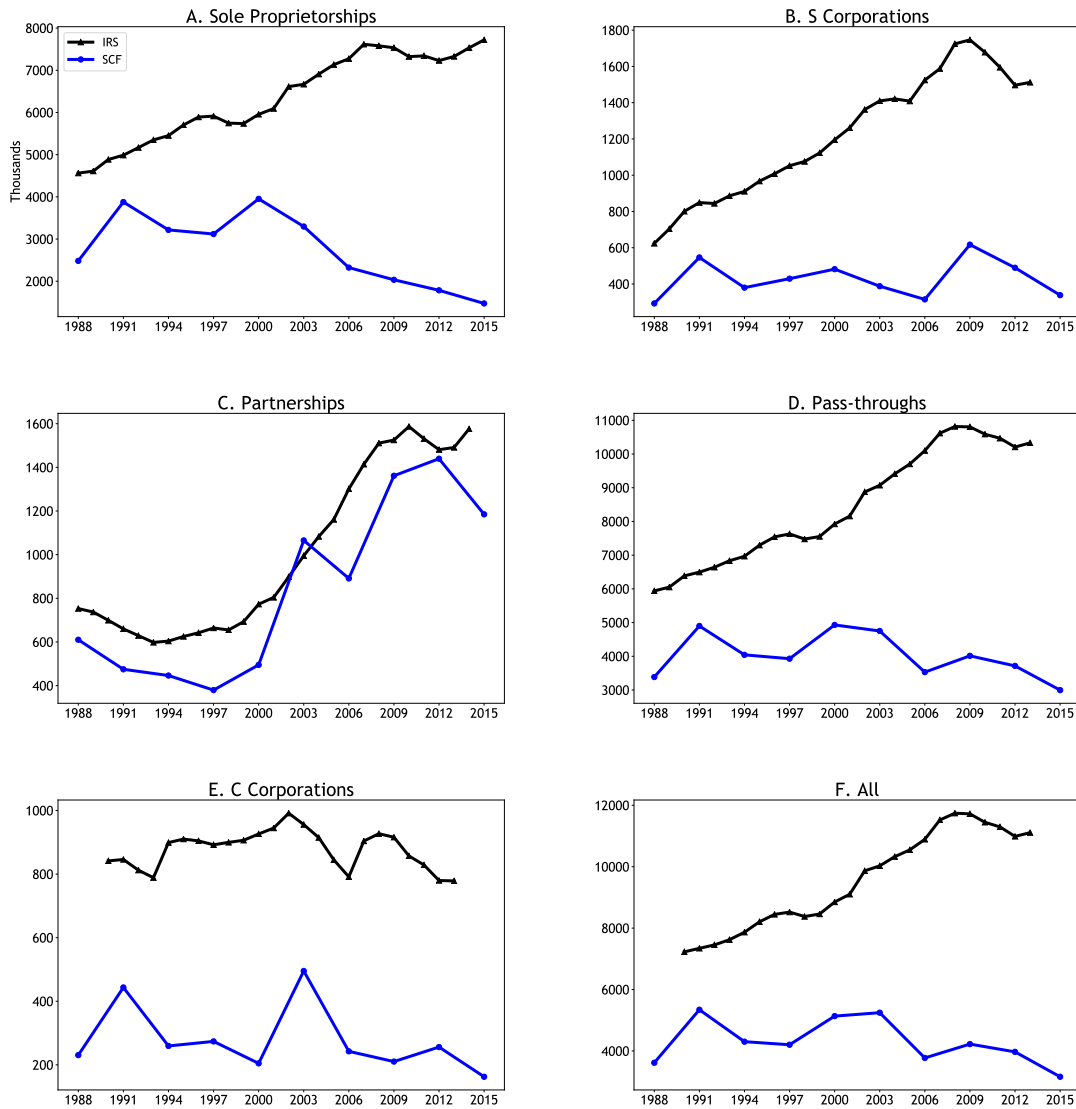
Note: This figure plots the business income per tax return by legal status for businesses with net loss in the SCF and IRS. Business income refers to income reported on Form 1040 Schedule C for sole proprietorships, Form 1065 for partnerships, Form 1120S for S corporations, and Form 1120 for C corporations. IRS data for sole proprietorships, partnerships, S corporations, and C corporations are available only until 2013, and C-corporation data start from 1990 because data for Form 1120 are not available for 1988 and 1989. Businesses with zero net income are included with those that have net losses.

Figure 10: Number of Returns by Legal Status for Businesses with Net Income, SCF vs. IRS



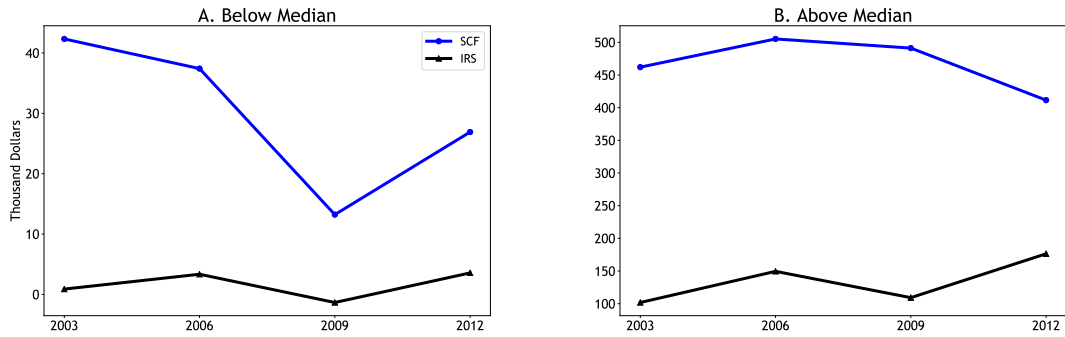
Note: This figure plots the number of business tax returns by legal status for business with net income in the SCF and the IRS. Business income refers to income reported on Form 1040 Schedule C for sole proprietorships, Form 1065 for partnerships, Form 1120S for S corporations, and Form 1120 for C corporations. IRS data for sole proprietorships, 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 by Legal Status for Businesses with Net Loss, SCF vs. IRS



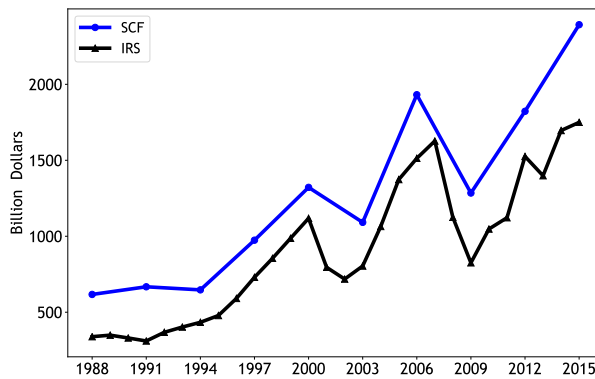
Note: This figure plots the number of business tax returns by legal status for businesses with net loss in the SCF and IRS. Business income refers to income reported on Form 1040 Schedule C for sole proprietorships, Form 1065 for partnerships, Form 1120S for S corporations, and Form 1120 for C corporations. IRS data for sole proprietorships, partnerships, S corporations, and C corporations are available only until 2013, and C- corporation data start from 1990 because data for Form 1120 are not available for 1988 and 1989. Businesses with zero net income are included with those that have net losses.

Figure 12: Distribution of S-Corporation Business Income per Return, SCF vs. IRS



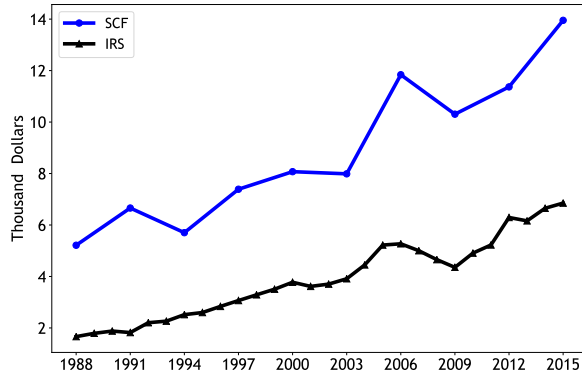
Note: This figure plots S-corporation business income per return for those with below- and above-median business receipts.

Figure 13: Broad Business Income, SCF vs. IRS



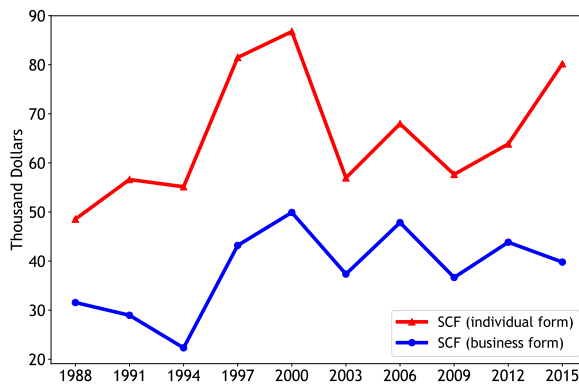
Note: This figure compares a broader measure of business income in the SCF and IRS. Broad business income is defined to be income derived from a business or profession (Form 1040 Schedule C) or farm (Form 1040 Schedule F); income from rental real estate, royalties, partnerships, S corporations, estates, trusts (Form 1040 Schedule E); and income from gains from the sale of capital and other property (Form 1040, lines 13 and 14).

Figure 14: Schedule C, E, F Income per Individual Tax Return



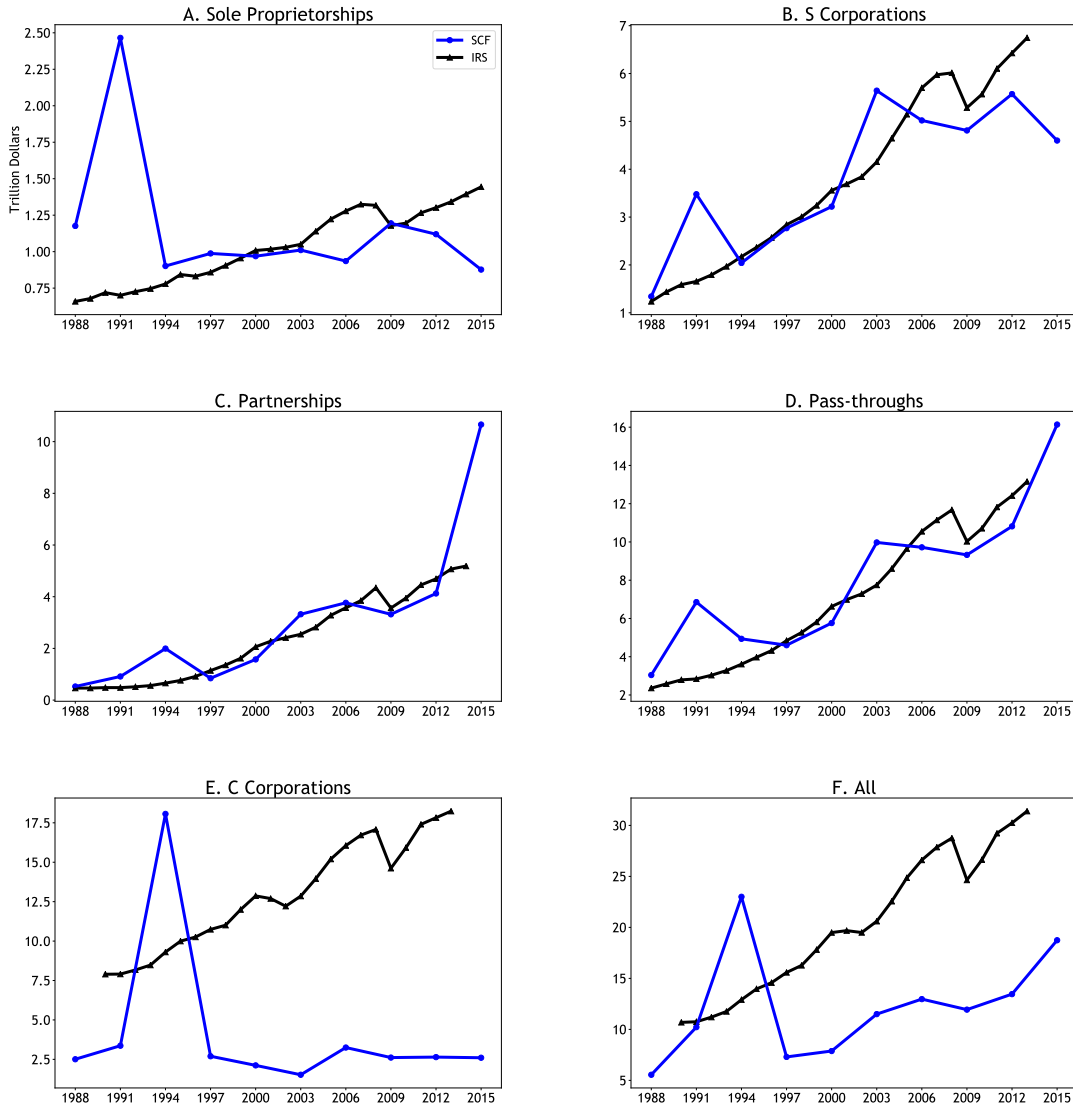
Note: This figure plots Schedule C, E, and F income per Form 1040 return. Schedule C comprises of income derived from a business or profession, Schedule F comprises of farm income, while Schedule E comprises of income earned from rental real estate, royalties, partnerships, S corporations, estates, trusts.

Figure 15: Comparing Proprietors' Individual and Business Incomes, SCF



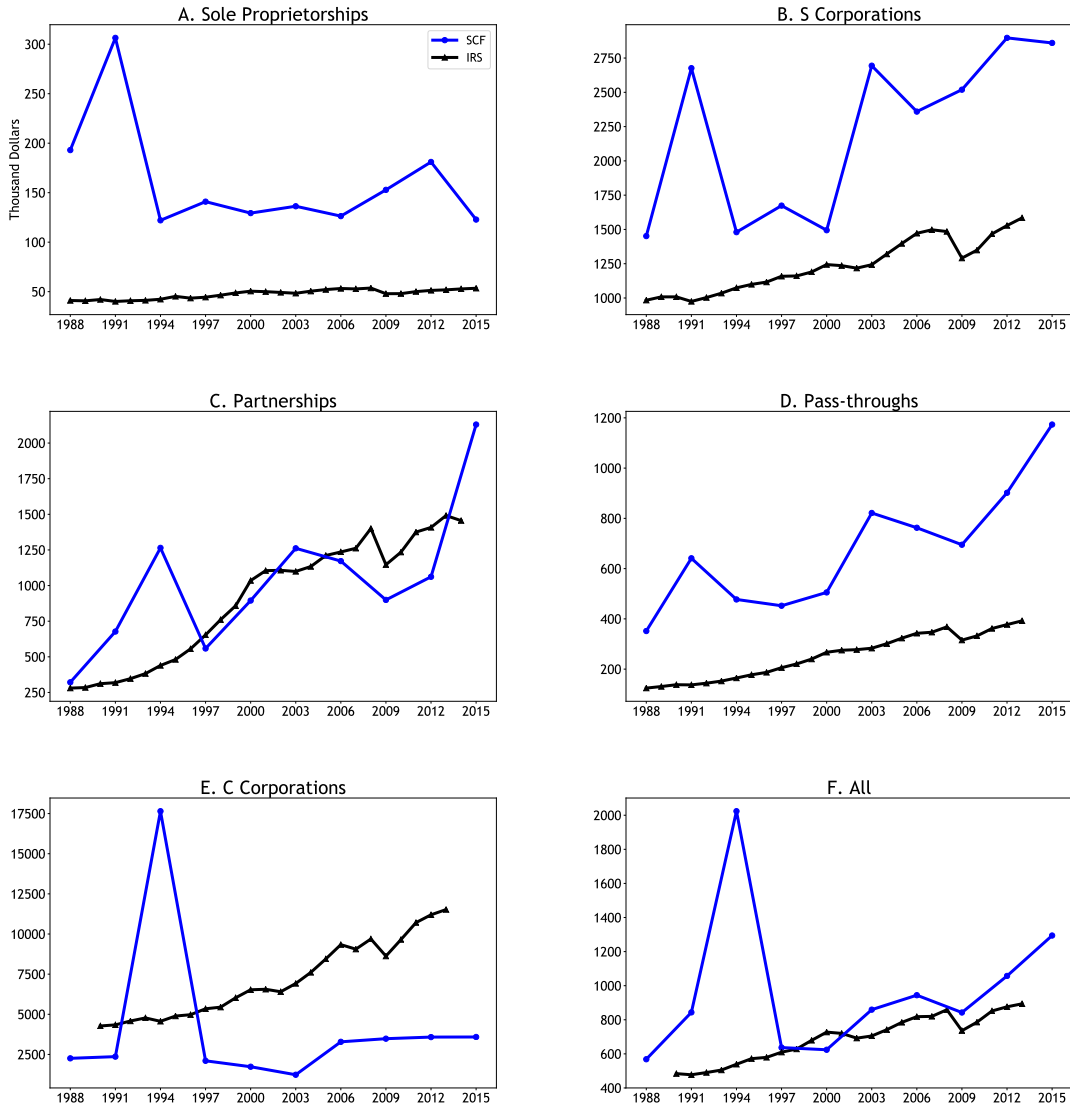
Note: This figure plots business income per return in the SCF for questions that ask respondents to report individual incomes listed on Form 1040, lines 12 plus 18, and business income on Schedule C of 1040, line 31.

Figure 16: Business Receipts by Legal Status, SCF vs. IRS



Note: This figure plots the total business receipts by legal status in the SCF and IRS. 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. IRS data for partnerships, S corporations, and C corporations are available only until 2013, and C-corporation data start from 1990 because data for Form 1120 are not available for 1988 and 1989.

Figure 17: Business Receipts per Tax Return by Legal Status, SCF vs. IRS



Note: This figure plots the business receipts per tax return by legal status in the SCF and IRS. 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. IRS data for partnerships, S corporations, and C corporations are available only until 2013, and C-corporation data start from 1990 because data for Form 1120 are not available for 1988 and 1989.