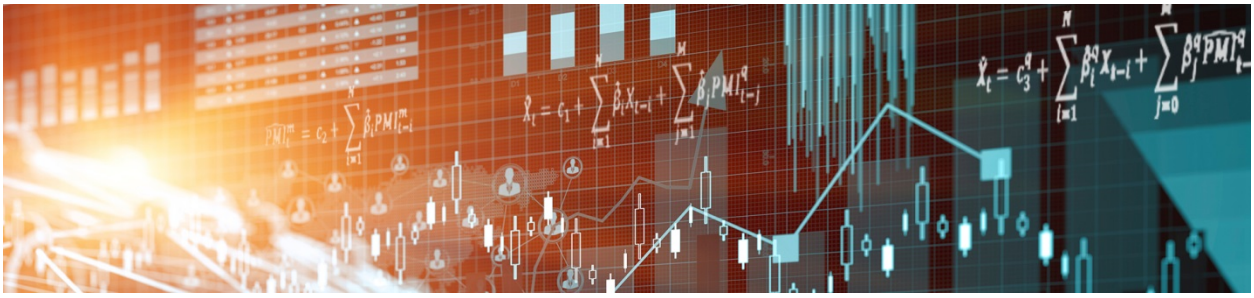
BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper / Document de travail du personnel 2019-45

What Do Survey Data Tell Us About US Businesses?



by Anmol Bhandari, Serdar Birinci, Ellen McGrattan and Kurt See

Bank of Canada Staff Working Paper 2019-45

December 2019

What Do Survey Data Tell Us About US Businesses?

by

Anmol Bhandari[1], Serdar Birinci[2], Ellen McGrattan[3] and Kurt See[4]

[1,3] Department of Economics
University of Minnesota
Minneapolis, Minnesota, United States
bhandari@umn.edu
erm@umn.edu

[2]Research Department
Federal Reserve Bank of St. Louis
St. Louis, Missouri, United States
Serdar.Birinci@stls.frb.org

[4]Canadian Economic Analysis Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
seek@bankofcanada.ca

ISSN 1701-9397

Acknowledgements

Bhandari and See acknowledge support from the Heller-Hurwicz Economic Institute, and McGrattan acknowledges support from the NSF (#1657891). For their helpful comments, we thank seminar and conference participants at the University of Colorado, University of Minnesota, Yale University, Universitat Autònoma de Barcelona, Arizona State University, New York University, Columbia University, University of California, Santa Barbara, Stanford Institute for Theoretical Economics, Carnegie Mellon University, Midwest Macro, Latin American Meeting of the Econometric Society, University of Michigan, Washington University in St. Louis and Stanford University. For their excellent editorial assistance, we thank Joan Gieseke. The views expressed herein are those of the authors and not necessarily those of the Bank of Canada, the Federal Reserve Bank of Minneapolis, the Federal Reserve Bank of St. Louis, the Federal Reserve System or the National Bureau of Economic Research.

Abstract

This paper examines the reliability of survey data on business incomes, valuations, and rates of return, which are key inputs for studies of wealth inequality and entrepreneurial choice. We compare survey responses of business owners with available data from administrative tax records, brokered private business sales, and publicly traded company filings and document problems due to nonrepresentative samples and measurement errors across all surveys, subsamples, and years. We find that the discrepancies are economically relevant for the statistics of interest. We investigate reasons for these discrepancies and propose corrections for future survey designs.

Bank topics: Firm dynamics; Business fluctuations and cycles
JEL codes: C83, E22, H25

Résumé

La présente étude traite de la fiabilité des données d'enquête sur les revenus, la valeur et les taux de rendement des entreprises, qui sont des données importantes pour étudier les inégalités de richesse et les choix entrepreneuriaux. Nous comparons les réponses d'enquêtes effectuées auprès de propriétaires d'entreprises avec des données tirées de dossiers fiscaux administratifs, de ventes d'entreprises privées réalisées par des courtiers et de déclarations de revenus de sociétés cotées en bourse. De plus, nous réunissons des informations sur les problèmes découlant d'échantillons non représentatifs et d'erreurs de mesure repérés dans l'ensemble des enquêtes et des sous-échantillons, pour toutes les années étudiées. Nous constatons que, pour les statistiques qui nous intéressent, les divergences trouvées sont significatives du point de vue économique. Enfin, nous étudions les raisons qui expliquent ces divergences et proposons des corrections pouvant être apportées lors de la conception d'enquêtes futures.

Sujets : Dynamique des entreprises; Cycles et fluctuations économiques
Codes JEL : C83, E22, H25

Non-Technical Summary

This paper examines the reliability of data on business incomes, valuations, and rates of return, which are key statistics for studies of wealth inequality and entrepreneurial choice.

Our analysis covers four widely used US surveys: Survey of Consumer Finances (SCF), Current Population Survey (CPS), Panel Study of Income Dynamics (PSID), and Survey of Income and Program Participation (SIPP). The SCF has the most suitable survey design for our analysis because households with actively-managed businesses are asked to refer to specific lines on their business tax forms when responding to survey questions. This provides a valuable test of the survey irrespective of whether misreporting of taxable income occurs. We have less detailed information from the other surveys but can compare their statistics for noncorporate businesses with the SCF and Internal Revenue Service (IRS) statistics.

We find that the SCF significantly overstates business income per tax return for all business types when compared with IRS statistics. Restricting attention to noncorporate business income per owner, for which we can compute comparable statistics across all surveys, we again find an overstatement of business incomes relative to IRS data, although by differing amounts. Averaged across survey years, business income per owner for noncorporate businesses is overestimated by 586 percent in the SCF, 179 percent in the CPS, 185 percent in the PSID, and 34 percent in the SIPP.

We also assess the accuracy of the survey responses on business valuation and returns by constructing net income-to-value ratios and comparing them with available income yields from brokered private business sales (from Pratt's Stats) and publicly traded companies (from the Center for Research in Security Prices—CRSP). For virtually all subsamples, all years, and all surveys that ask about valuations, we find that the income yields are significantly higher than comparable measures from Pratt's and CRSP. The overstatement

in yields is even greater than for incomes, implying an understatement in business valuations.

These problems arise due to nonrepresentative samples and measurement errors across all surveys, subsamples, and years. We compare business incomes after ranking households by total income and find low-income businesses are in fact underrepresented, which leads to an overstatement of business incomes if total and business incomes are positively correlated. We also provide evidence of measurement errors that arise due to the framing of questions. For example, there are many IRS businesses with net losses but few in the survey data, possibly because the respondents answered that they had no net income rather than a negative net income.

1 Introduction

Representative surveys of households and firms have become an important source of data on business owners and their activities and are now used extensively in economic research. This paper examines the reliability of data on business incomes, valuations, and rates of return based on surveys—key statistics for studies of wealth inequality and entrepreneurial choice. To do this, we first compare responses to questions about business incomes, receipts, and owner counts with corresponding administrative tax data and document problems due to nonrepresentative samples and measurement errors, both in the aggregate and across the distribution. We then demonstrate that the discrepancies are economically relevant for the statistics of interest but not systematically correctable given current survey designs.

The scope of our analysis is four widely used surveys: Survey of Consumer Finances (SCF), Current Population Survey (CPS), Panel Study of Income Dynamics (PSID), and Survey of Income and Program Participation (SIPP). The SCF is the best survey design for our analysis; it asks households with actively managed businesses to report both the legal form of their business and specific lines from the relevant business tax forms—thus providing a valuable test of the survey regardless of whether misreporting of taxable income occurs. The survey includes questions for pass-through entities (sole proprietorships, S corporations, and partnerships) and privately-held C corporations. We have less detailed information from the other surveys but can compare statistics for noncorporate businesses with the SCF and Internal Revenue Service (IRS) statistics.

Averaging across survey years, we find that the SCF significantly overstates business income per tax return for all business types. If we consolidate pass-through entities, we find an overstatement of 400 percent. In the case of C corporations, the SCF does not include publicly traded companies, while the IRS does. If we were to append the SCF estimates to include them, we would again find a significant overstatement of incomes. We also find a significant overstatement of aggregate business incomes and an understatement of the number of tax returns across most business forms. Importantly, the

overstatement of per-return and aggregate business incomes relative to IRS counterparts varies significantly in the cross-section and year by year. If we restrict attention to noncorporate business income per owner, for which we can compute comparable statistics across all surveys, we again find an overstatement of business incomes relative to IRS data, although by differing amounts. Averaged across survey years, business income per owner for noncorporate businesses is overestimated by 586 percent in the SCF, 179 percent in the CPS, 185 percent in the PSID, and 34 percent in the SIPP.

More relevant for economic research are possible issues with survey-based estimates of business valuations, since there are no measures of total valuations for ongoing businesses other than publicly traded C corporations. To assess the accuracy of the survey responses, we construct net income-to-value ratios and compare them with available income yields from brokered private business sales recorded by Pratt's Stats and publicly traded companies—both small and large—recorded by the Center for Research in Security Prices (CRSP) (merged with Compustat). The valuations from the survey data are based on self-reports of the value of their share of the business, net of all loans, if the owners were to sell. For virtually all subsamples, all years, and all surveys that ask about valuations, the income yields are significantly higher than comparable measures from Pratt's and CRSP. The overstatement in yields is even greater than for incomes, implying an understatement in business valuations. For example, the SCF average value-weighted income yield is 19 percent, much higher than the Pratt's estimates of 2 percent and the CRSP estimates of 7 percent for all businesses or -9 percent for those in the bottom quintile when firms are ranked by total assets. We also find that the SCF distributions are more right-skewed than those based on Pratt's or CRSP data. Average value-weighted income yields calculated for the PSID and SIPP are also high relative to Pratt's and CRSP data in all cases, but are not very different from those in the SCF. The main differences in yields across surveys are found when we compare the distributional statistics.

Given the significant issues with business income and wealth in the surveys, we investigate the likely sources of sampling and measurement errors and possible corrections. A natural hypothesis for the overstatement of busi-

ness incomes per return is that survey data omit owners with little business attachment.[1] This hypothesis is not consistent with the finding that aggregate incomes are overstated in the survey data or with direct evidence from our investigation of proprietorships in the SCF. For proprietorships, in which the household and business unit is identical, the SCF records the business incomes twice: once in response to questions about individual incomes (Form 1040) and again in response to questions of owners with actively managed businesses about their incomes (Schedule C). We find more business income—in the aggregate and per return—for proprietors who do not report that they have Schedule C income. While puzzling in its own right, this observation leads us to reject the view that owners with little business attachment drive our findings of overstatement of business incomes in survey data. The findings call for a survey redesign that makes the notion of actively managing a business precise and verifiable and enforces an internal consistency check across survey responses wherever possible.

A hypothesis for the overstatement of incomes that cannot be rejected is that the incomes and numbers of unsuccessful businesses are understated. We compare business incomes after ranking households by total income and find low-income businesses are in fact underrepresented, which leads to an overstatement of business incomes if total and business incomes are positively correlated. We also provide evidence of measurement errors that arise due to the framing of questions. For example, there are many IRS businesses with net losses but few in the survey data, possibly because the respondents answered that they had no net income rather than a negative net income. These issues affect cross-sectional statistics, which are key inputs for studies of income and wealth inequality. A survey redesign that links questions about tax forms to administrative data would alleviate problems of framing. Such a redesign is also warranted given that most owners never reference tax or financial documents when surveyed.

We also investigate proposed corrections for measurement errors discussed in previous work.[2] For example, Johnson and Moore (2008) have noted that

[1] If true, the overstatement of incomes per return would be inconsequential for applied research on businesses. See Kennickell, Kwast, and Pogach (2017) for such a view.

[2] See Johnson and Moore (2008) for comparisons of business incomes in the SCF and IRS

an overstatement of income could be the result of misreporting income to the IRS or miscategorizing income in the surveys. We provide evidence that misreporting and miscategorization do not explain the large discrepancies that we find. In the case of misreporting, we use auxiliary tax audit data to adjust the IRS data but still find a significant mismatch with surveys. For example, for pass-through businesses, the SCF overstates the average income per return by 178 percent relative to the tax audit data. In the case of miscategorizations, we use a broader definition of business income, as recommended by Johnson and Moore (2008). For example, business owners might confuse business incomes on Schedules C, E, and F, overstating one category and understating another. When we combine these categories into a broader concept of business income, we still find incomes to be significantly overstated. Respondents are not miscategorizing incomes but are often overstating all categories of business income. Here again, we would argue for a survey redesign that links questions about tax forms to administrative data.

Our findings have implications for several active areas of economic research. Survey data on businesses are a central input to studies of wealth inequality since rising business incomes account for most of the growth in the top 1 percent share. (See Bricker et al. 2016, Kuhn and Rios-Rull 2016, Saez and Zucman 2016, and Smith et al. 2017.) Even for researchers that use administrative tax data and capitalize incomes, survey data serve as the only check on distributional assumptions and capitalization factors. Survey data on businesses are also a central input to studies of entrepreneurial choice. Based on empirical findings from surveys, studies have come to different conclusions about the payoffs to entrepreneurial activities. For example, some find puzzlingly low payoffs and conclude there are large nonpecuniary benefits of entrepreneurship (see Hamilton 2000, Moskowitz and Vissing-Jorgensen 2002, and Hurst and Pugsley 2011), while others find high payoffs and conclude there are severe financial constraints hindering entrepreneurship (see Evans and Jovanovic 1989, Quadrini 2000, Cagetti and De Nardi 2006, and Buera 2009). Our findings

data, Saez and Zucman (2016) for comparisons of a broader capital income measure in the SCF and IRS data, and Kennickell, Kwast, and Pogach (2017) for comparisons of business counts in the SCF and US Census Bureau.

cast doubt on the facts that have been uncovered in the empirical literature—specifically those related to business incomes and rates of returns—and thus raise issues concerning the theoretical developments and policy analyses that have been designed around them.

2 Business Incomes

In this section, we compare data on business incomes from surveys with corresponding data from the IRS. We document significant overstatements of business income per tax return across all legal forms of organization and show that these discrepancies vary across years. We then evaluate five reasons for the overstatements and suggest possible corrections.

2.1 Evidence

We start with a comparison of SCF and IRS *business incomes*, defined as gross receipts from sales minus expenses (including depreciation) incurred in running the business. Information on business incomes is obtained from the respective business tax forms: Form 1040, Schedule C (line 31) for sole proprietors; Form 1065 (line 22) for partnerships; Form 1120S (line 21) for S corporations; and Form 1120 (line 30) for C corporations. In each survey year, we use the SCF sampling weights and ownership shares for multiowner businesses to compute the aggregate business income and the aggregate number of business tax returns by legal form of the business.[3]

Figure 1 plots aggregated business incomes divided by the number of business tax returns for each business type along with the data actually reported to the IRS for tax years 1988–2015. The shaded regions for the survey data are 90 percent bootstrapped confidence intervals using SCF replicate weights. If we construct percentage errors (that is, 100(SCF−IRS)/IRS), we find that they are large and significantly different from zero. For sole proprietorships (Panel A), the average error is 289 percent with a range of 158 to 384 percent across

[3]To be consistent with the IRS statistics, we assume that a business owner with multiple proprietorships files one return.

years. For S corporations (Panel B), the average error is 273 percent with a range of 142 to 387 percent. In the case of partnerships (Panel C), the SCF excludes part of the sample—namely, partnerships owned by corporations—but in principle this should not affect the income per return. The average error in this case is 300 percent with a range of 31 to 837 percent. Interestingly, when we compare aggregate business incomes for partnerships, the SCF estimate is still higher than the IRS data even though part of the sample is excluded.

Our headline estimate for all pass-through businesses shown in Figure 1, Panels A through C, is an average error of 400 percent with a range of 230 to 568 percent. Contrast this with business incomes per return for C corporations (Panel D). For these businesses, we find that in most years, the average SCF business income per return is understated by about 26 percent compared with the IRS data. The IRS data include publicly traded corporations; however, the SCF data do not. Publicly traded C corporations are typically much larger than their private counterparts. If we were to include the incomes from these publicly traded corporations in the SCF estimates, we would find that the SCF incomes per return would be significantly higher than the IRS estimates, as is the case for pass-through businesses.

While incomes per return are overstated in the SCF relative to the IRS data, the number of returns filed by businesses is significantly understated for all business forms except partnerships. Figure 2 plots the number of business returns in the IRS and the SCF, over time and by legal entity, with shading again marking the 90 percent confidence interval. In the case of sole proprietors and S corporations shown in Panels A and B, the understatement has worsened over time because the number of IRS filings has grown and the number reported in the SCF has not.[4] In Panel C, we see that the number of partnership returns in the SCF is undercounted in only a few years and not by as much as in the case of the other business types. However, as mentioned before, the SCF data only include partners who are individuals, implying that the SCF *overstates* the number of returns for partnerships owned by individuals. Similarly, the

[4]The SCF reports ownership shares for only up to three actively managed businesses (and two after 2007), but the fraction of households with three businesses is tiny (roughly 0.4 percent).

number of C corporations should be lower in the SCF than the IRS data because publicly traded businesses are not included. However, publicly traded businesses account for about 5,000 of the roughly 1.6 million C corporations. Thus, the large difference in number of returns in Panel D means the SCF is also underrepresenting private C corporations.

If we restrict attention to noncorporate businesses per owner, we can evaluate the accuracy of estimates in the CPS, PSID, and SIPP and compare the results with the SCF. 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 errors in the range of 384 to 969 percent when compared to the IRS, PSID next with 110 to 378 percent errors, CPS after that with 104 to 279 percent errors, and finally SIPP with 11 to 59 percent errors. The inconsistencies between surveys are driven primarily by differences in aggregate business incomes.

2.2 Hypotheses

Next, we consider possible reasons for the overstatements of income in survey data and possible corrections.

2.2.1 Owners with little attachment are excluded

One possible reason for understated returns and overstatements of income per return is that the survey data may not include owners earning very little business attachment (for example, part-time Uber drivers or AER referees), while the IRS includes all business tax filers. If this were indeed the case, then aggregate business incomes—found in the SCF by multiplying values in Figure 1 by values in Figure 2—would be similar for the IRS and the survey data because these owners would have little business income. We find, however, that this is not the case: aggregate business incomes are significantly overstated. For example, in pass-through businesses in the SCF, we find average overstatements of 34 percent, 137 percent, and 305 percent for sole proprietors, S corporations, and partnerships, respectively, with a large range in the errors over time. The large overstatement of aggregate incomes, especially in

S corporations and partnerships, is clearly inconsistent with the hypothesis that nonactive business owners explain the differences between the SCF and IRS results.

We can exploit the SCF design to further investigate the hypothesis that the overstatement of incomes per return is driven by owners with little business attachment. Sole proprietorships in the SCF are asked to report their income from business twice—first as a part of questions about the components of their individual tax forms (Form 1040, lines 12 plus 18 coded as X5704) and then again for those reporting they actively manage a business as a part of questions about line items on the business tax form (Schedule C, line 31 coded as X3119, X3219, X3319). Households that earn business income but have low business attachment should reply negatively to the question on active management. Using a Venn diagram in Figure 3 and data for 2015, we see that the SCF reports that 16.2 million proprietors reported nonzero business or farm income on Form 1040, lines 12 plus 18, but only 6.3 million of them reported that they actively managed a business, with most of the rest (9.2 million) reporting neither actively managing nor having an "interest" in a business without an active management role.[5] Kennickell, Kwast, and Pogach (2017) explain that "there is often not a clear distinction between self-employment and business ownership," which might make sense if the 9.2 million earned very little business income. But as we show in Figure 3, the SCF reports that the sole proprietors who do not claim to be managing a business earned *more than half* of all proprietorship income for 2015—$303 billion out of the $583 billion—and have comparable per-return incomes to those who report being actively managing. In sum, the evidence from the SCF does not support the hypothesis that exclusion of owners with low business attachment drive the overstatement of income in survey data.

These observations call for a survey redesign that makes precise the notion of actively managing a business and implements internal consistency checks mid-interview, and, if possible, that links future surveys to administrative

[5]Respondents are asked separately if they are "self-employed" without any reference to tax filings. The self-employed who also claim to have proprietorship income account for roughly 90 percent of income coded as X5704.

data. This is particularly warranted since only a tiny fraction of respondents refer to their tax documents or any business financial statements. In 2015, for example, 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.[6] In the case of other financial documents, 64 percent never referenced any other financial documents, 6 percent rarely did, 15 percent sometimes did, and 15 percent frequently did.

2.2.2 Owners with little income are underrepresented

Next, we investigate if the overstatement is due to an underrepresentation of low-income owners. This hypothesis can be tested by ranking businesses according to owners' total income. Consider the case of sole proprietorships. We have comparable IRS data in all SCF survey years to compute both population and income shares for subgroups after ranking them by their adjusted gross income (AGI). For example, we can split the sample into below- and above-median AGI groups. If we do, we find that the SCF estimate for the number of returns filed by the below-median group is roughly 2 million for the entire period and equal to about 25 percent of the total population, while IRS data show a rise from 5 million in 1988 to over 12 million in 2015, with the group accounting for roughly 43 percent of the total population. These findings suggest a significant underrepresentation of low-income businesses, which leads to an overstatement of business incomes if business income is positively correlated with AGI. Furthermore, if we analyze income shares over time, we find the share of income for those with below-median AGI is nearly doubled or halved from one survey to the next, while the IRS share has steadily grown.

Correcting such problems requires an expansion of the sampling frame in future surveys for better representation of the population of business owners.

2.2.3 Owners misreport business losses in surveys

Business incomes can be overstated if losses are understated. We find this to be the case for the SCF. To demonstrate this, we group businesses with

[6] We found virtually no difference in behavior when conditioning on income.

profits and losses separately. For pass-through businesses, the SCF overstates the income per return for profitable businesses by an average of 277 percent, with the year-to-year errors in the range of 151 to 446 percent. The SCF understates the losses per return for businesses with negative net incomes by an average of 82 percent, with the year-to-year errors in the range of 66 to 94 percent. Both the overstatement of profits and the understatement of losses affect the errors in cross-sectional statistics in quantitatively important ways. If we decompose the percentage errors in aggregate business income into the fractions arising from overstatements of profits and understatements of losses, we find nontrivial errors for both in all survey years.

Part of the problem may be in the framing of questions about business incomes. For example, the question "What is your net income?" could be misinterpreted as being a question about positive net income. As evidence, consider the distribution of losses by AGI bins. In tax year 2015, the IRS data show all AGI subgroups had nontrivial net losses, while the SCF data show 10 out of 19 AGI subgroups—accounting for 23 percent of total returns and 26 percent of all IRS losses—recorded an aggregate net loss of *exactly* zero. Such framing issues can easily be corrected in future surveys by clarifying that net income could be negative.

2.2.4 Owners misreport incomes to tax authorities

Although the SCF questions ask about amounts on specific lines of IRS tax forms, Johnson and Moore (2008) have noted that the overstatement of income should be expected if owners misreport income to tax authorities but truthfully report income in SCF interviews. To test this hypothesis, we adjust the IRS data by adding back estimates of unreported income. The estimates we use are based on imputations from the Bureau of Economic Analysis (BEA) for noncorporate businesses and estimates from Johns and Slemrod (2010) and the Government Accountability Office (GAO) for S corporations based on tax audit data. For the years 1988–2015, the BEA estimates that reported noncorporate tax incomes are lower by roughly 50 percent because of misreporting. Johns and Slemrod (2010) document underreporting of 18 percent for Schedule E income, which includes all supplemental income from S corporations,

partnerships, rental real estate, and royalties. The GAO estimates misreporting margins for S-corporation incomes in the order of 15 to 20 percent. We construct a measure of adjusted IRS pass-through income by adding back the BEA estimates of misreporting for noncorporate incomes, along with an adjustment of 18 percent for S-corporation income. In Panel A of Figure 4, we compare the SCF pass-through business incomes per return with the adjusted IRS incomes per return and find that they are still significantly higher. Computing the SCF errors as before, we find that the average error with the tax audit adjustment is 178 percent, with a range of errors of 98 to 274 percent over the sample.

2.2.5 Owners misclassify business incomes

Another source of measurement error in the SCF is the respondents' possible confusion about closely related categories of business income. For example, when asked about income from a sole proprietorship appearing on Schedule C, business owners might also include income appearing on Schedule E, which includes income from real estate, royalties, partnerships, S corporations, estates, and trusts, or farm income on Schedule F. From our previous analysis, we know that business incomes from Schedule C are overstated in the SCF. If the overstatement was due to classification errors, we should see an understatement in categories of income corresponding to Schedules E or F.

In Panel B of Figure 4, we plot Schedule E income per return and again find the SCF income per return is overstated relative to IRS data, especially in recent years. The average error is 121 percent, with a range of 50 to 221 percent over the sample. Furthermore, if we sum all business incomes from Schedules C, E, and F and compare the totals with IRS data, we find errors averaging 90 percent, with a range of of 47 to 180 percent over the sample. If we follow the recommendation of Johnson and Moore (2008) and broaden even further by adding capital gains (which includes nonbusiness incomes), we still find an overstatement: the average error is 47 percent, with a range of 18 to 115 percent over the sample.

One could further broaden the concept of business income to include all nonwage income, thus lowering the discrepancies between datasets. But such

aggregation is not useful for applied work—either research on US businesses or research on US capital. For research on businesses, the residual income measure would be inappropriate because significant nonbusiness income is included with interest payments, capital gains, pensions and annuities, alimony, trusts, and government transfers. For research on capital, the nonwage income in AGI would be inappropriate because a significant fraction of capital income is untaxed and the corresponding assets are held by fiduciaries.

3 Business Valuations and Rates of Return

Overreporting of incomes would lead to an upward bias in estimates of business rates of return. In this section, we combine net income with self-reported business valuations to construct income yields that are comparable to available yields from brokered private business sales and from small and large publicly traded firms. We find that for virtually all subsamples and all years, the survey yields are significantly higher and more right-skewed than comparable measures from the brokered sales and public firms. We then evaluate two reasons for the discrepancies.

3.1 Evidence

In Table 1, we report statistics for net income yields using different universes of businesses and different data sources (listed in rows). In columns, we report mean yields using two weighting schemes, namely, equally weighted and value weighted, and we report various percentiles of the income yield distribution.

We start by comparing the survey yields with yields based on broker data from Pratt's Stats.[7] We compute the income yield in Pratt's by dividing the pretax income earned by the business in the year before the sale by the sale price. The advantage of Pratt's is that it records the price at which the business was actually sold; thus, it is conceptually close to the ideal answer to

[7]The Pratt's database has transaction-level data on sales of private and public businesses over the period 1988–2017. The dataset includes financial information about the target business and other attributes of the sale, including payment terms, purchase price allocations, and employment agreements.

the survey questions on business valuation.

First, we can compare yields for all businesses in the SCF and Pratt's data. The results of this comparison are shown in Table 1. The differences are dramatic: the Pratt's equally weighted yield is 27 percent compared with 102 percent for the SCF, and the Pratt's value-weighted yield is 2 percent as compared with 19 percent for the SCF. The larger discrepancy in the equally weighted yield relative to the value-weighted yield suggests the presence of discrepancies in the distribution of yields. This can also be seen by directly comparing the percentiles of the income yield distribution across data sources. Here, we see that the 75th percentile yield in the SCF is substantially higher than the counterpart in Pratt's. In other words, the SCF overestimates the right-skewness of the cross-sectional distribution of business returns.

A direct implication of the overstated SCF yield is an understated SCF valuation. If the Pratt's yields are representative of pass-through businesses, then the SCF average net income yields are too high by a factor of 10. We documented earlier that the net incomes for pass-through businesses (in the numerator of the yield) are high by a factor of roughly 2, implying that self-reported valuations are significantly underestimated.

If we restrict attention to noncorporate businesses, we can compare yields across the SCF, SIPP, and PSID surveys. In Table 1, we see that value-weighted income yields in the PSID and SIPP are comparable to the SCF, ranging from 15 to 23 percent, and all yields are much higher than those from Pratt's. The survey estimates are comparable even though income per owner is much lower in the PSID and SIPP than in the SCF. This implies that average business values are even lower in these other surveys. However, if we compare yields across the distribution, we see large differences across surveys, especially in the right tail. These observations point to a lack of representativeness in the PSID and SIPP for the universe of noncorporate businesses as well as their lack of comparability to the SCF.

Next, we compare income yields for all businesses in the SCF and CRSP databases. The income yield for a firm in CRSP is computed by dividing pretax income by the firms' end-of-year market capitalization. In Table 1, we see that the CRSP equally weighted income yield for the full sample is

actually negative (-9 percent), and the CRSP value-weighted income yield (7 percent), while positive, is much lower than that for the SCF. Considering the distribution, we again find that the SCF yields are more right-skewed than those in CRSP. For example, at the 75th percentile, the SCF C-corporation yield is 36 percent, while the CRSP yield for all businesses is 10 percent.

These findings appear to be inconsistent with Moskowitz and Vissing-Jorgensen (2002), who constructed private business returns using SCF data and concluded that they were surprisingly low when compared with those of publicly traded firms. There are a few differences in our procedures—we use a longer sample and comparable measures of pretax earnings in the SCF and CRSP rather than imputing retained earnings for pass-through firms—but the more important difference for the quantitative results is the concept of return.[8]

Earlier results are based on a measure of return equal to the sum of a value-weighted income yield and an imputed capital gain. In theory, one would need a panel of firm valuations to compute a value-weighted capital gain. Given that the SCF survey is triennial with virtually no panel aspect (other than two surveys), there is no way to compute the change in value firm by firm. Moskowitz and Vissing-Jorgensen (2002) instead compute the ratio of aggregated firm values across consecutive surveys and then annualize it to obtain a measure of capital gains. Comparing their return with the value-weighted mean holding-period return on the CRSP index portfolio, they find private returns that are similar in magnitude to the returns on listed public firms. In view of the higher risk for private businesses, Moskowitz and Vissing-Jorgensen (2002) conclude that there is a puzzle as to why individuals become entrepreneurs. Their preferred explanation is that running a business has non-pecuniary benefits. (See also Hamilton 2000 and Hurst and Pugsley 2011.)

Replicating the exercise of Moskowitz and Vissing-Jorgensen (2002) for our full sample with income yields and capital gains compared separately, we find that the capital gain imputation drives the differences between our findings and theirs. First, consistent with our findings for the average income yields, the yearly SCF yields are substantially higher than the CRSP counterparts for

[8]Kartashova (2014) also finds that a longer sample can raise the SCF-CRSP return gap by about 6 percentage point.

all survey years. Second, the annualized SCF capital gains vary substantially less than those for firms in the CRSP sample, which is not surprising given the conceptual differences in the capital gain measures and the long interval between survey years. Combining overstated yields and imputed gains from the survey confounds two discrepancies and results in similar estimates for private and public returns: 26 percent for SCF and 21 percent for CRSP. However, if we were to compute capital gains in CRSP and SCF in a comparable manner, we would find a lower average return of 16 percent for CRSP.

Thus restricting attention to income yields or comparable total returns, we conclude that the rates of returns are significantly higher for private businesses when compared with public returns, not low as previously thought.

3.2 Hypotheses

Next, we consider possible reasons for the overstatements of income yields—and implied understatements of business valuations—in survey data relative to yields of private businesses that were sold or publicly traded businesses.

3.2.1 Yields in Pratt's and CRSP are biased downward

Returns on businesses in Pratt's broker data could be biased downward if sales are triggered by distress, say because of health-related issues facing the owner. Similarly, sales of technology- or research-intensive businesses would imply lower yields because of the significant expensing done by these firms. We test this hypothesis by ignoring transactions in which the target company is in technology- and research-intensive sectors (that is, with NAICS codes 51, 5415, or 5417) and those for which the stated reason for the sale was health related. In Table 1, we report the data for this subset of firms and find the results are nearly the same as for all businesses.

CRSP yields could also be biased downward because the typical firm in CRSP is much larger than the typical firm in the survey data. To test this hypothesis, we compute yields for small firms in CRSP that are more comparable to those in the survey samples and report them in the last row of Table 1. Our baseline definition of "small" is a firm that belongs to the bottom quin-

tile of firms ranked by the book value of assets.[9] Given we have data on all business types for the SCF, we also compute yields for S and C corporations since they are most similar to businesses in CRSP. Here again, the differences are dramatic. The equally weighted income yield for small firms in CRSP is -27 percent, whereas the yield is large and positive for both C corporations (57 percent) and S corporations (76 percent) in the SCF. Interestingly, yields for the small business subsample in CRSP are lower than the full sample across the distribution, implying an even larger discrepancy between survey and CRSP yields for small businesses.

3.2.2 Survey valuations understate intangible assets

Our main hypothesis for the upward bias in the rates of return is that respondents underreport valuations of intangible assets used in businesses. From Pratt's data, we know that roughly 60 percent of the purchase price upon sale is the value of intangible assets. (See Bhandari and McGrattan 2019.) If business owners do not include the value of these assets when reporting business net worth, the returns would be biased upward by even more than the incomes. A constructive way to deal with the measurement issues in the SCF and estimate aggregate and distributional statistics for business valuations is to rely more heavily on a theory that is disciplined by the flows measured from the IRS and business sales data such as Pratt's Stats. A theory featuring business sales would take a stand on the selection bias and also provide a way to impute the valuations for ongoing concerns.[10]

4 Conclusion

The need for accurate data on private businesses is especially urgent given that the US Census has recently discontinued its *Survey of Business Owners.* It is especially urgent given that theories and policies are being developed around survey "facts," but the evidence suggests these should be treated with great

[9]The results are similar for other definitions of "small," for example, based on gross sales or market capitalization.

[10]An attempt in this direction is ongoing work in Bhandari and McGrattan (2019).

caution. We hope and expect that our analysis will lead to improved measurement in future surveys. Measurement problems related to business incomes are surmountable given that respondents are asked about specific lines on tax forms. Measurement problems related to business valuations and returns may be insurmountable without data on actual business sales transactions or a theoretical framework and a method of indirect inference.

References

[1] Bhandari, Anmol, Serdar Birinci, Ellen R. McGrattan, and Kurt See. 2019. "Data Appendix: What Do Survey Data Tell Us about US Businesses?" Federal Reserve Bank of Minneapolis Staff Report 578.

[2] Bhandari, Anmol, and Ellen R. McGrattan. 2019. "Sweat Equity in US Private Business." Federal Reserve Bank of Minneapolis Staff Report 560.

[3] Bricker, Jesse, Alice Henriques, Jacob Krimmel, and John Sabelhaus. 2016. "Measuring Income and Wealth at the Top Using Administrative and Survey Data." *Brookings Papers on Economic Activity* 2016(1): 261–331.

[4] Buera, Francisco J. 2009. "A Dynamic Model of Entrepreneurship with Borrowing Constraints: Theory and Evidence." *Annals of Finance* 5(3–4): 443–464.

[5] Cagetti, Marco, and Mariacristina De Nardi. 2006. "Entrepreneurship, Frictions, and Wealth." *Journal of Political Economy* 114(5): 835–70.

[6] Evans, David S., and Boyan Jovanovic. 1989. "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints." *Journal of Political Economy* 97(4): 808–27.

[7] Hamilton, Barton H. 2000. "Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment." *Journal of Political Economy* 108(3): 604–31.

[8] Hurst, Erik, and Benjamin W. Pugsley. 2011. "What Do Small Businesses Do?" *Brookings Papers on Economic Activity* 43(2): 73–142.

[9] Johns, Andrew, and Joel Slemrod. 2010. "The Distribution of Income Tax Noncompliance." *National Tax Journal* 63(3): 397–418.

[10] Johnson, Barry W., and Kevin Moore. 2008. "Differences in Income Estimates Derived from Survey and Tax Data." In *Proceedings of the Joint Statistical Meetings, Survey Research Methods Section.* Alexandria: American Statistical Association.

[11] Kartashova, Katya. 2014. "Private Equity Premium Puzzle Revisited." *American Economic Review* 104(10): 3297–334.

[12] Kennickell, Arthur B., Myron L. Kwast, and Jonathan Pogach. 2017. "Small Businesses and Small Business Finance during the Financial Crisis and the Great Recession: New Evidence from the Survey of Consumer Finances." In *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges*, edited by John Haltiwanger, Erik Hurst, Javier Miranda, and Antoinette Schoar, 291–340. Chicago: University of Chicago Press.

[13] Kuhn, Moritz, and José-Victor Rios-Rull. 2016. "2013 Update on the US Earnings, Income, and Wealth Distributional Facts: A View from Macroeconomics." *Federal Reserve Bank of Minneapolis Quarterly Review* 37(1): 2–73.

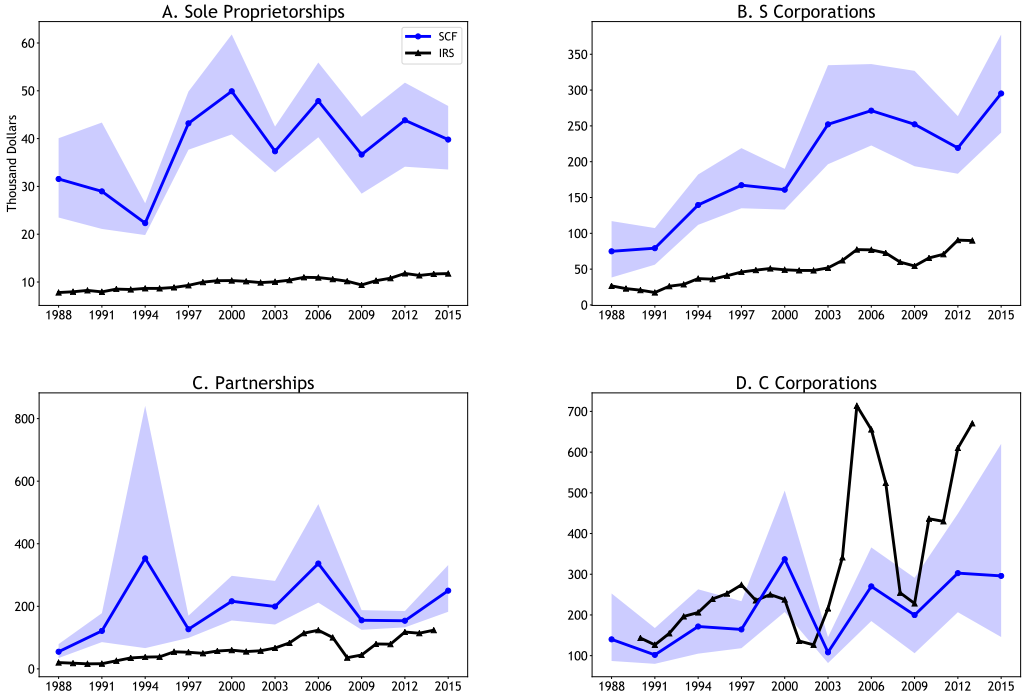
[14] Moskowitz, Tobias J., and Annette Vissing-Jorgensen, 2002. "The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle?" *American Economic Review* 92(4): 745–78.

[15] Quadrini, Vincenzo. 2000. "Entrepreneurship, Saving, and Social Mobility." *Review of Economic Dynamics* 3(1): 1–40.

[16] Saez, Emmanuel, and Gabriel Zucman. 2016. "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data." *Quarterly Journal of Economics* 131(2): 519–78.

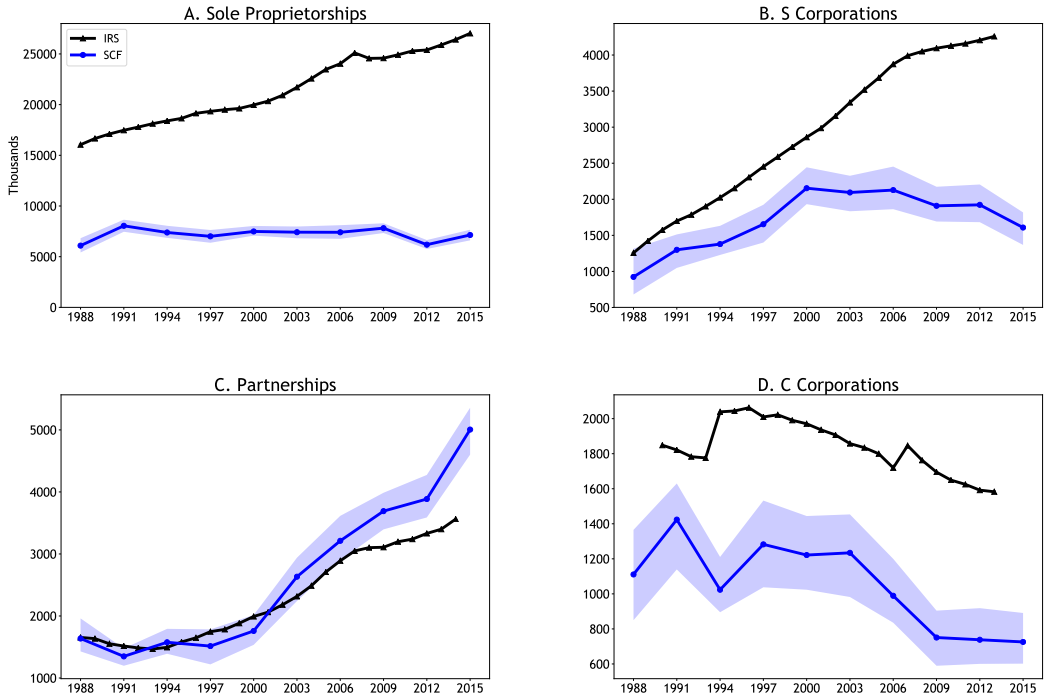
[17] Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick. 2017. "Capitalists in the Twenty-First Century." NBER Working Paper 25442.

Figure 1
Business Income per Return by Legal Entity, SCF vs. IRS



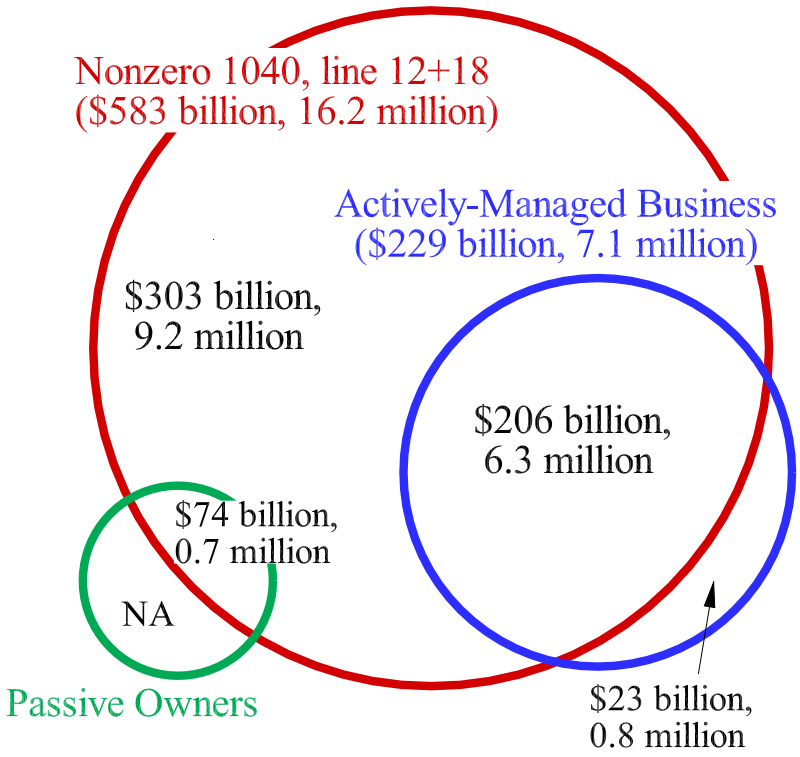
Note: This figure plots business income per business tax return in the IRS and the SCF as reported on Form 1040 Schedule C for sole proprietorships, Form 1120S for S corporations, Form 1065 for partnerships, and Form 1120 for C corporations. IRS data for partnerships, S corporations, and C corporations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

Figure 2
Number of Returns by Legal Entity, SCF vs. IRS



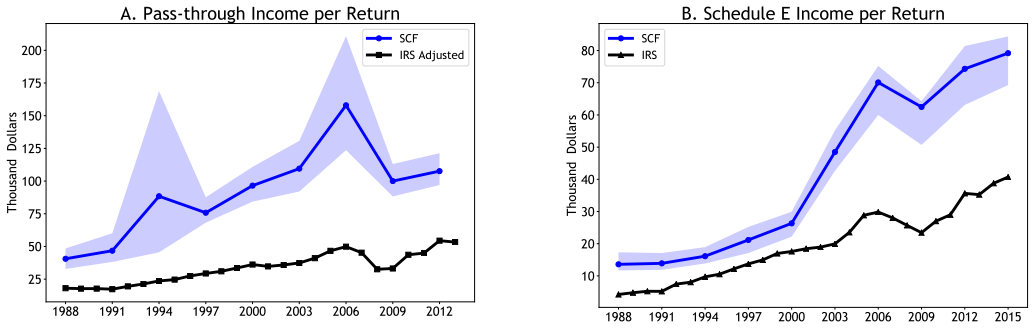
Note: This figure plots the number of business returns of sole proprietorships, S corporations, partnerships, and C corporations over time in the IRS and the SCF. IRS data for partnerships, S corporations, and C corporations are available only until 2013. For C corporations, prior to 1990 only consolidated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

Figure 3
Sole Proprieter Responses in the SCF



Note: This figure shows incomes and counts for three sets of sole proprietors. In red are those who have a non-zero line 12 plus 18 on Form 1040, in blue are those who report to be actively managing, and in green are those who report owning or having an interest in business without an active-management role. Some of the groups overlap and in such cases we use incomes from Form 1040 lines 12+18. The NA for the counts and incomes of the passive owners (green) is because the SCF stopped identifying legal form of organization for passive owners after 2007. In the appendix we provide details for other survey years.

Figure 4
Comparison Assuming Misreports and Misclassifications, SCF vs. IRS



Note: In Panel A, we use BEA estimates for misreporting of noncorporate business incomes and reports from the GAO on misreporting of S-corporation business incomes to adjust IRS pass-through business income per return. We add these yearly adjustments to the sum of pass-through income in the IRS, calculate total business income per tax return, and compare it with estimates from the SCF. In Panel B, we plot business income per business tax return in the IRS and SCF as reported on Form 1040 Schedule E, which includes income and losses from real estate, royalties, partnerships, S-corporations, trusts, and estates. The shaded region for the SCF shows the 90 percent confidence interval.

Table 1
Net Income Yields Comparison

	Means		Percentiles		
	Equal Weights	Value Weights	p25	p50	p75
SCF					
All businesses	102.5	19.1	0.9	17.6	63.0
S corporations	76.4	15.2	2.2	14.2	50.5
C corporations	56.8	16.9	1.3	10.6	36.2
Noncorporate	107.6	22.6	0.8	19.6	70.6
PSID Noncorporate	220.4	14.9	3.2	27.0	114.9
SIPP Noncorporate	*	17.7	2.2	33.2	230.1
Pratt's Stats					
All businesses	27.4	1.9	3.8	21.7	46.8
Nontech, nondistressed	29.3	3.5	5.0	23.1	48.3
CRSP-Compustat					
All businesses	−9.2	7.3	−5.4	5.5	10.4
Small businesses	−26.6	−8.5	−29.0	−7.7	4.0

Note: This table shows moments of the net income yield distribution from the SCF, PSID, SIPP, Pratt's Stats, and CRSP-Compustat for different subgroups. For the "nontech and nondistressed" businesses in Pratt's Stats, we exclude those in technology- and research-intensive sectors (NAICS codes 51, 5415, or 5417) and those for which the stated reason of sale was health related. For the CRSP-Compustat sample, small businesses refer to publicly traded firms in the CRSP database that belong to the bottom 20 percent when ranked by total assets. For the equally-weighted SIPP mean, we found the estimate to be over 15,000 and driven by outliers.

Appendix

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 US 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	SOI Table 1.4: *All Returns: Sources of Income, Adjustments, and Tax Items, by Size of Adjusted Gross Income*	Data available for 1988–2015.
2	Adjusted gross income (SCF)	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.	X5702+X5704+X5708+X5710 +X5712+X5714+X5716+X5718 +X5724	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: *Nonfarm Sole Proprietorships: Income Statements, by Industrial Sectors* Table 1.4: *All Returns: Sources of Income, Adjustments, and Tax Items, by Size of Adjusted Gross Income*	See [1] above.
4	Sole proprietorship income, 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: *All Partnerships: Total Assets, Trade or Business Income and Deductions, Portfolio Income, Rental Income, and Total Net Income, by Selected Industrial Group*	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*: S Corporation Returns: Balance Sheet and Income Statement Items, by Major Industry*	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: *Balance Sheet, Income Statement, Tax, and Selected Other Items, by Major Industry*	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 EPROPB1–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 data 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 data.

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 data.

For the Kauffman Firm Survey (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 data 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 data, 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

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

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

[4]As in the SCF, these surveys only account for partners who are individuals. However, as we mentioned before, using estimates from Cooper et al. (2016), this fact alone does not 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 data, 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 inconsistencies between SCF and IRS data that are similar to those 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.

2.3 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 nonzero line 12 plus 18 on Form 1040, (B) those who are actively managing a business and report line 31 of Schedule C, and (C) those reporting to have an interest in a business without an active-management role. In Table 2 we provide details for all survey years.

Table 2: Sole Proprietorships in the SCF

Tax	Incomes					Counts				
Year	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 a business, and (C) those reporting having an interest in a 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.

2.3.1 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 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, or 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.

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

[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.

Table 3: Distribution of Nonreferencing 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
Nonpositive	0.70
Positive	0.76

Note: This table summarizes nonreferencing 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, and p75 representing the 25th percentile, the 50th percentile, and the 75th percentile. For each bin, we compute the 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.

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.

2.4 Business Receipts

In this section, we extend the analysis in the main text to business receipts. Figures 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 data. We again find large and variable errors in the SCF responses when compared with 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

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

the 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.

2.5 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:

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

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

In the main text, we showed evidence that the SCF income yields are high when compared with 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 with 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 with 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 nontech and nondistressed 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 with 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 with 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 are explained by Moskowitz and Vissing-Jorgensen's (2002) imputation method used to calculate capital gains. Below, we provide more details on how we reached this conclusion.

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

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

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:

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

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 with 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

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

[8] Incidentally, 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.

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 with public returns, not low as previously thought.

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

	Receipts			Expenses			Profit		
	KFS	IRS	Error	KFS	IRS	Error	KFS	IRS	Error
Statistic	’000	’000	%	’000	’000	%	’000	’000	%
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 (%)	Percentage of Gap from 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

	IRS		SCF	
AGI	Returns	Losses	Returns	Losses
Bins	'000	$ Bil.	'000	$ 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

	Sole Proprietorship					Partnership				
Tax Year	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

	By Market Capitalization					By Sales				
Tax Year	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

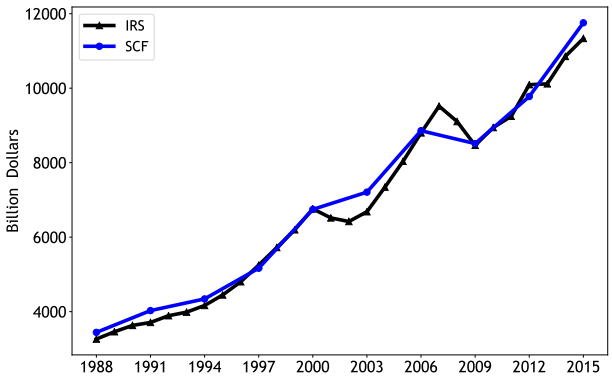
Note: This table shows moments of the net income yield distribution of sole proprietorships, partnerships, and unincorporated businesses in the SIPP 2004 and 2008 panels. Statistics are calculated for years where asset topical modules are available. The sample includes businesses with positive net worth 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	
				$(t-1) \to t$	$(t-3) \to t$
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

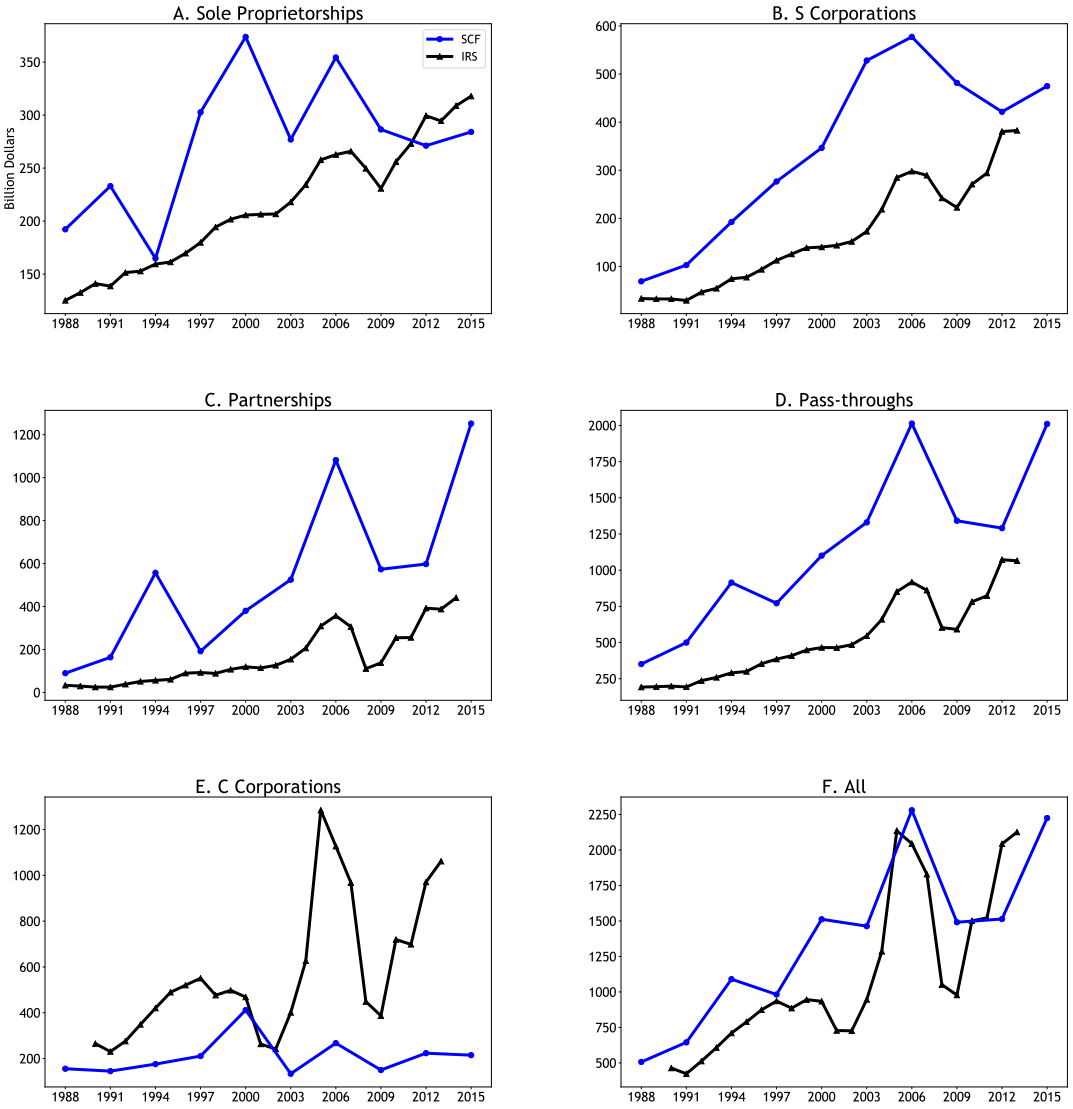
Note: This table shows estimates of income yields and capital gains for businesses in the SCF and CRSP-Compustat firms. For the SCF, capital gains are computed using Equation 3 found in the main text, as in Moskowitz and Vissing-Jorgensen (2002). For the CRSP-Compustat firms, we report two measures of capital gains. The column $(t-1) \to t$ 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 $(t-3) \to t$ 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



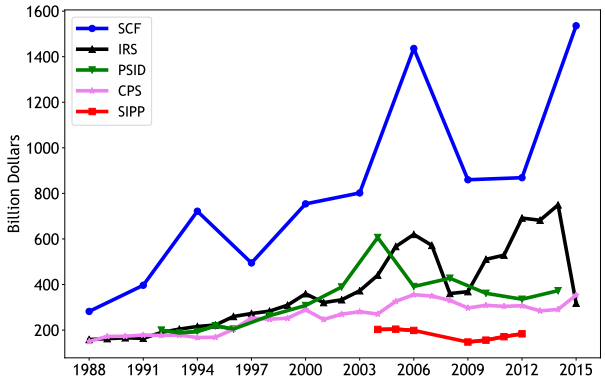
Note: For the IRS, adjusted gross income is obtained from Form 1040. For the SCF, if AGI is not available, we construct it by adding the appropriate income categories.

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



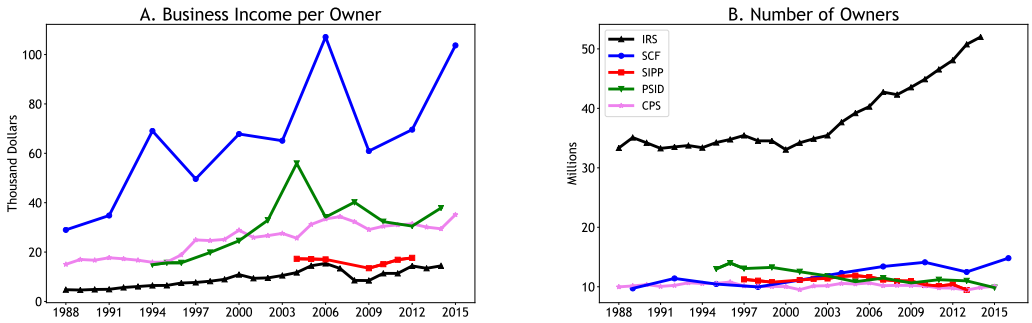
Note: This figure plots the total business income by legal status in the SCF and the IRS data. Business income refers to income reported on Form 1040, Schedule C, for sole proprietorships; Form 1065 for partnerships; Form 1120S for S corporations; and Form 1120 for C corporations. IRS data for partnerships, 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 data. 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

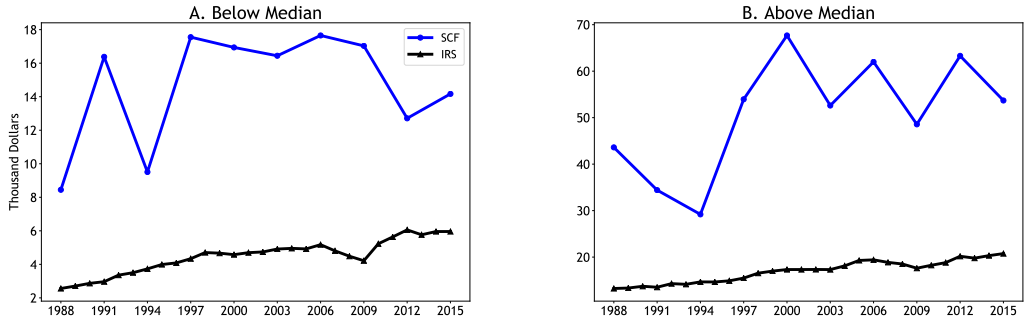


Note: This figure plots the total business income per owner of unincorporated businesses (Panel A) and total number of unincorporated business owners (Panel B) in the SCF, CPS, PSID, SIPP, and IRS data. Before 2004, the SIPP does not provide information about an individual's own share of business income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

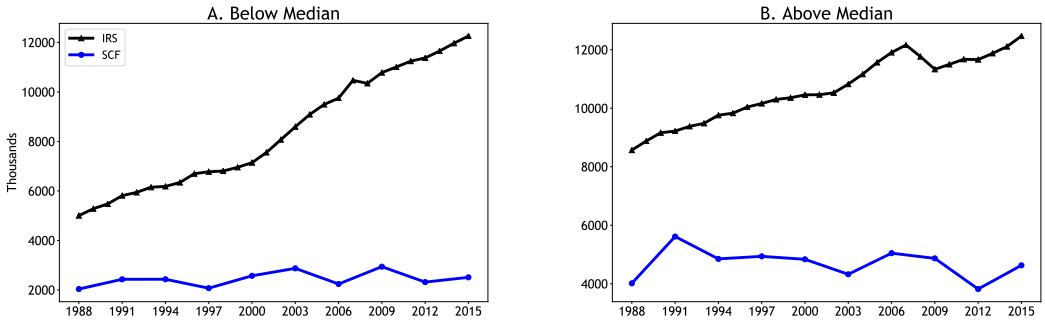
Note: This figure plots the fraction of business income from sole proprietorships attributable to returns with AGI below the median and above the 99th percentile.

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



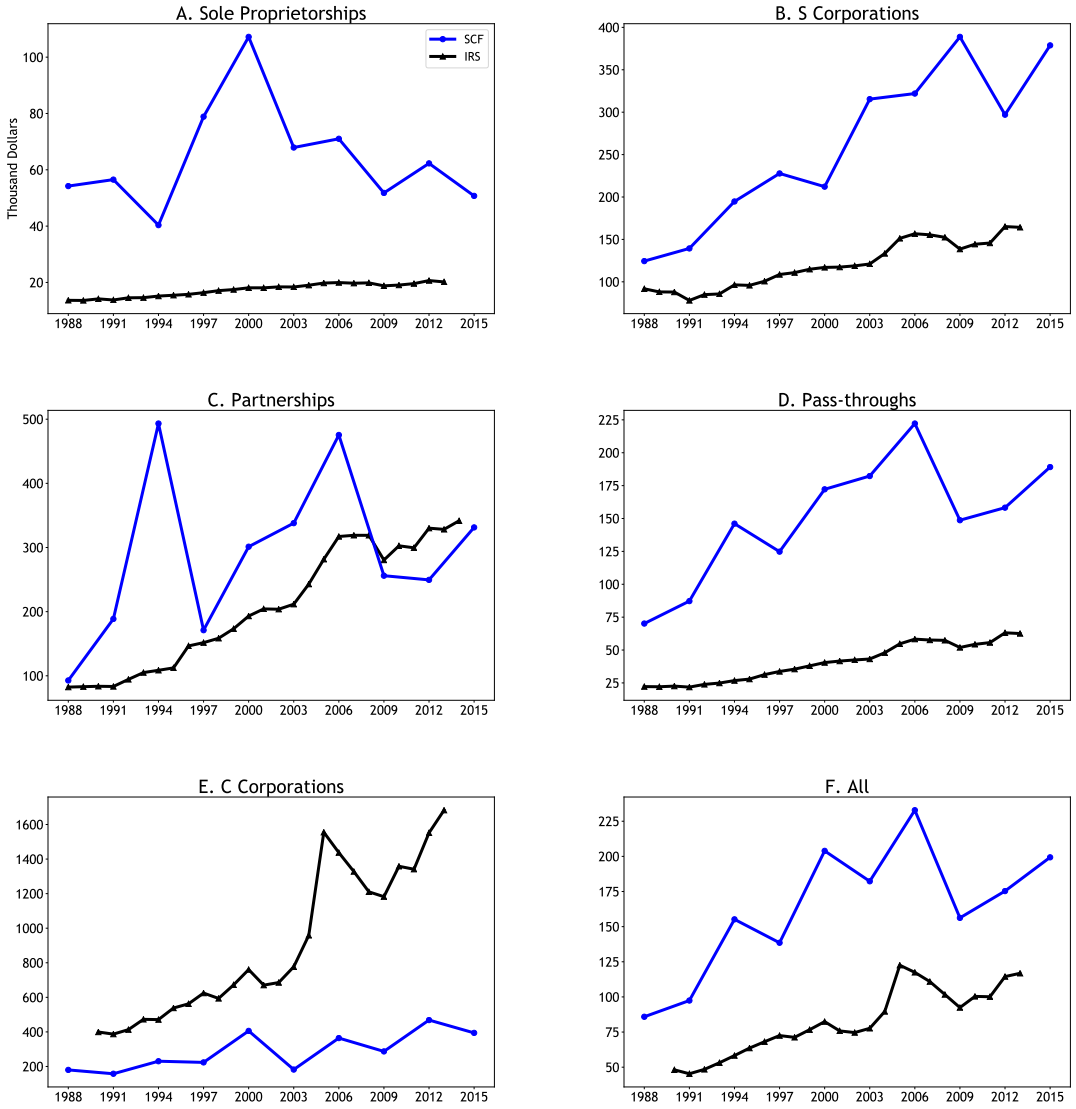
Note: This figure plots sole proprietorship business income per return for those with below- and above-median AGI.

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



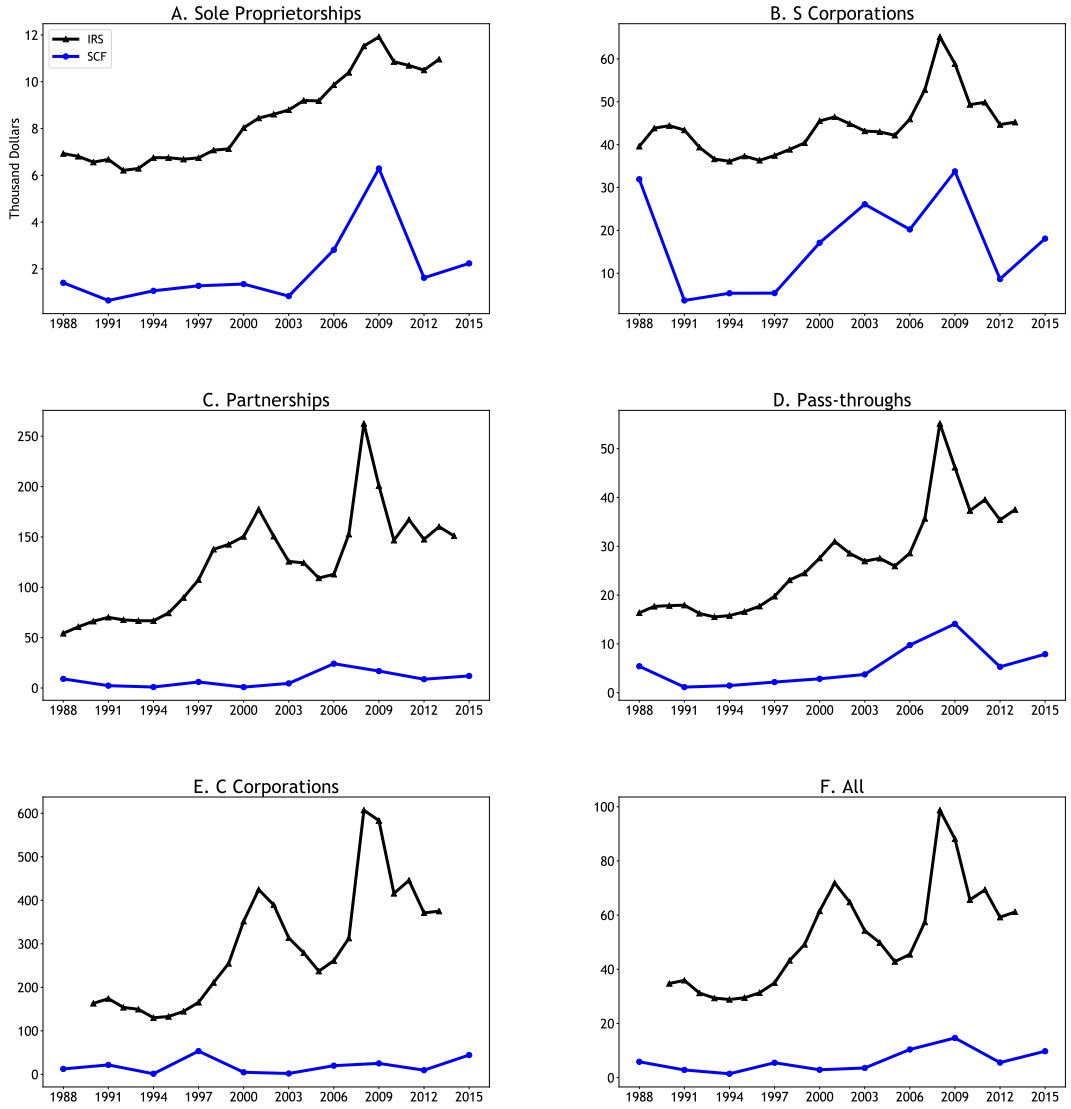
Note: This figure plots the number of sole proprietorship returns (Form 1040, Schedule C) filed by business owners with below- and above-median AGI.

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



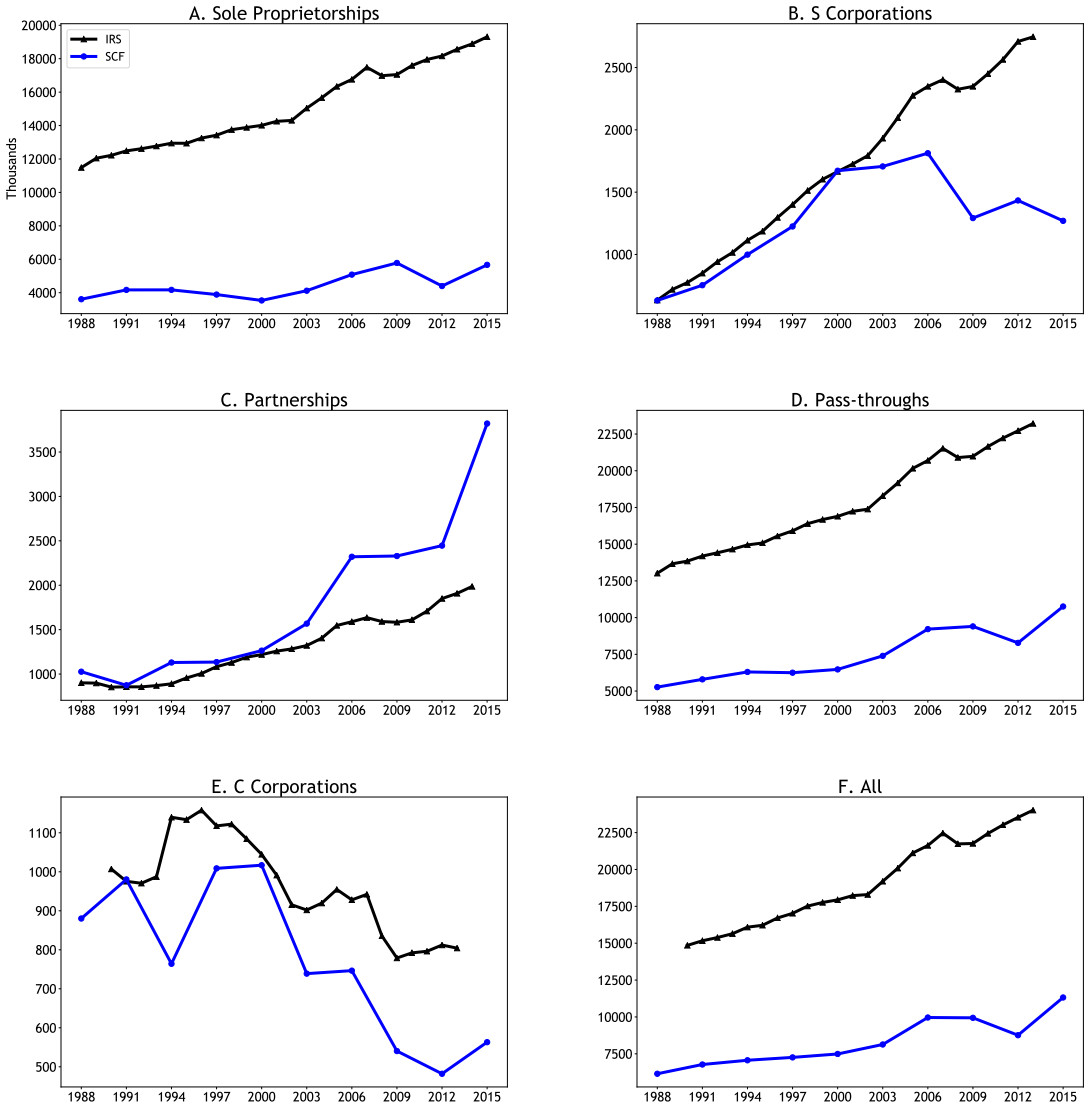
Note: This figure plots the business income per tax return by legal status for businesses with net income in the SCF and IRS. Business income refers to income reported on Form 1040, Schedule C, for sole proprietorships, Form 1065 for partnerships, Form 1120S for S corporations, and Form 1120 for C corporations. 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



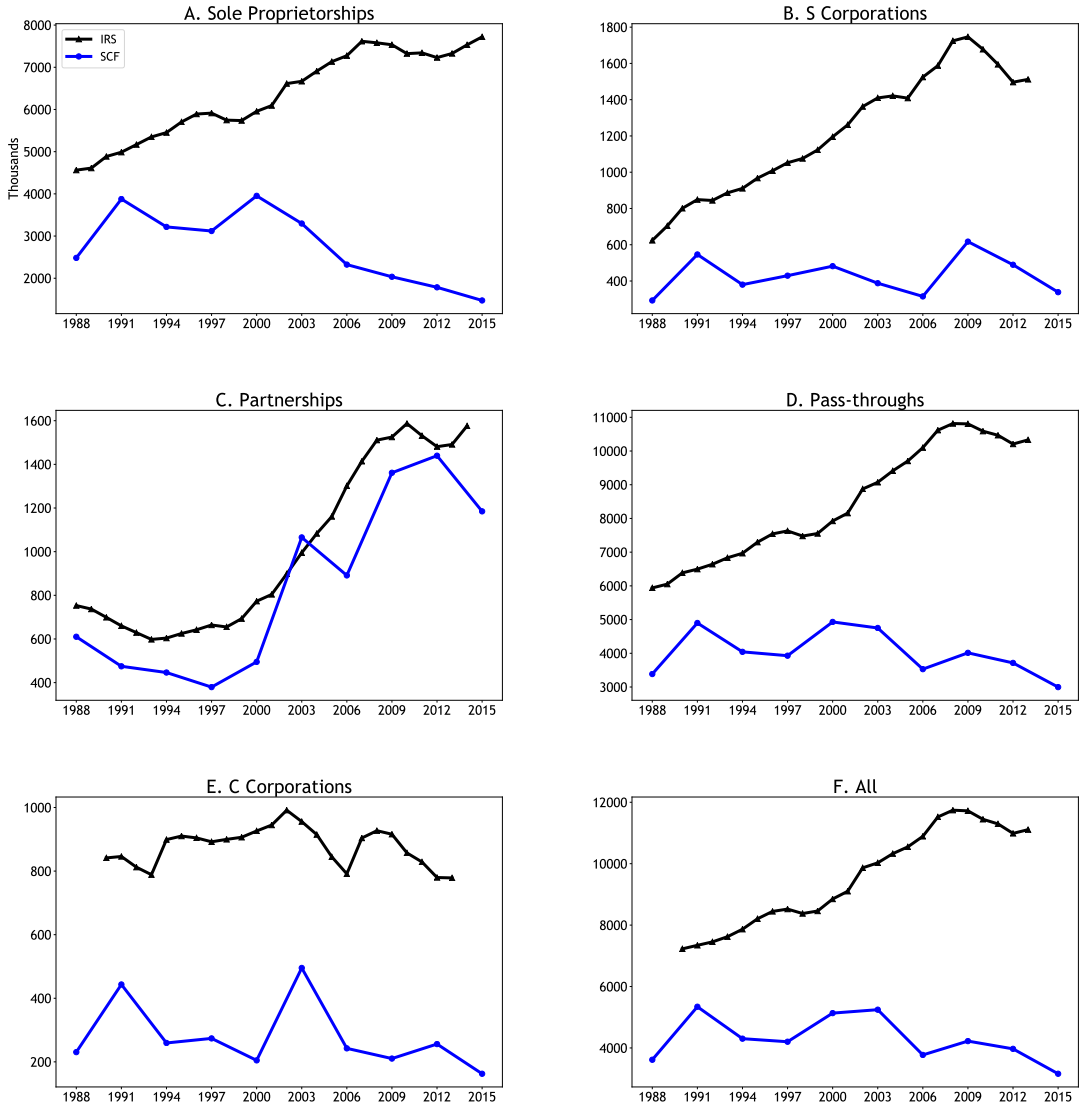
Note: This figure plots the business income per tax return by legal status for businesses with net loss in the SCF and IRS. Business income refers to income reported on Form 1040, Schedule C, for sole proprietorships, Form 1065 for partnerships, Form 1120S for S corporations, and Form 1120 for C corporations. 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



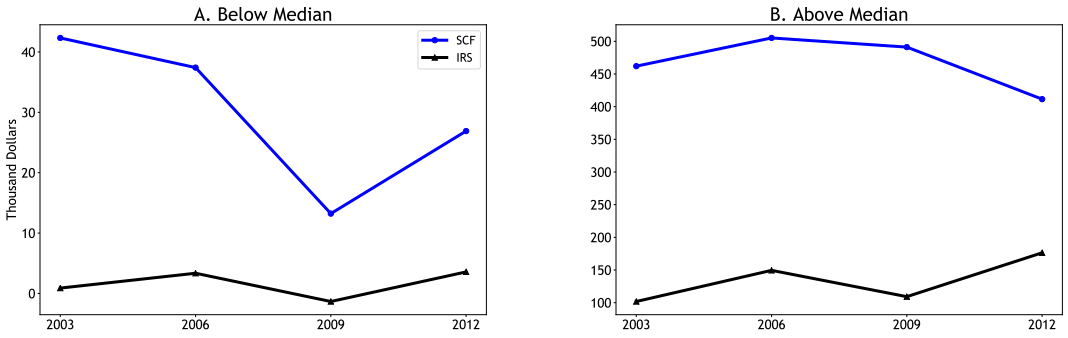
Note: This figure plots the number of business tax returns by legal status for business with net income in the SCF and the IRS. Business income refers to income reported on Form 1040, Schedule C, for sole proprietorships, Form 1065 for partnerships, Form 1120S for S corporations, and Form 1120 for C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

Figure 11: Number of Returns by Legal Status for Businesses with Net Loss, SCF vs. IRS



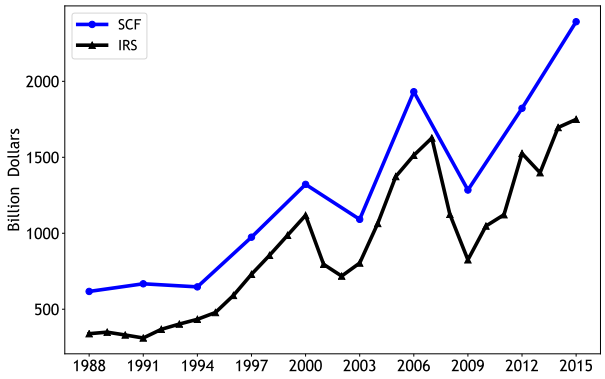
Note: This figure plots the number of business tax returns by legal status for businesses with net loss in the SCF and IRS. Business income refers to income reported on Form 1040, Schedule C, for sole proprietorships, Form 1065 for partnerships, Form 1120S for S corporations, and Form 1120 for C corporations.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



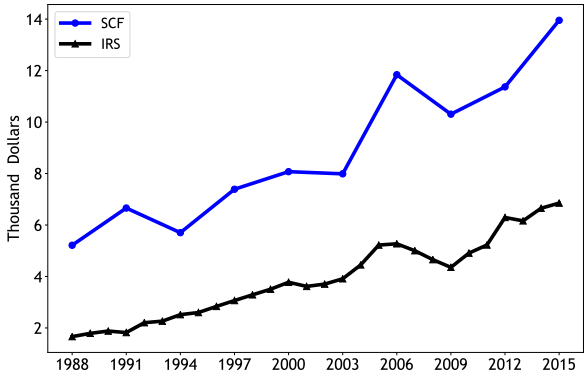
Note: This figure plots S-corporation business income per return for those with below- and above-median business receipts.

Figure 13: Broad Business Income, SCF vs. IRS



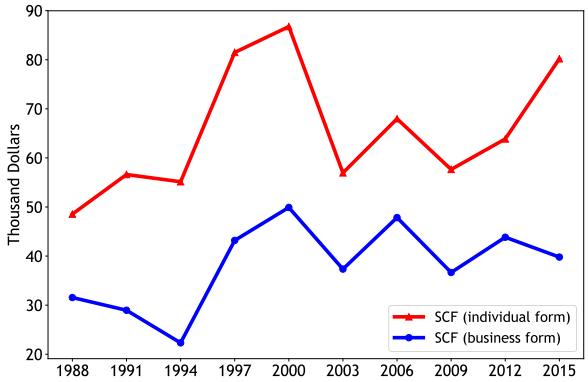
Note: This figure compares a broader measure of business income in the SCF and IRS. Broad business income is defined as income derived from a business or profession (Form 1040, Schedule C) or farm (Form 1040, Schedule F); income from rental real estate, royalties, partnerships, S corporations, estates, or 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



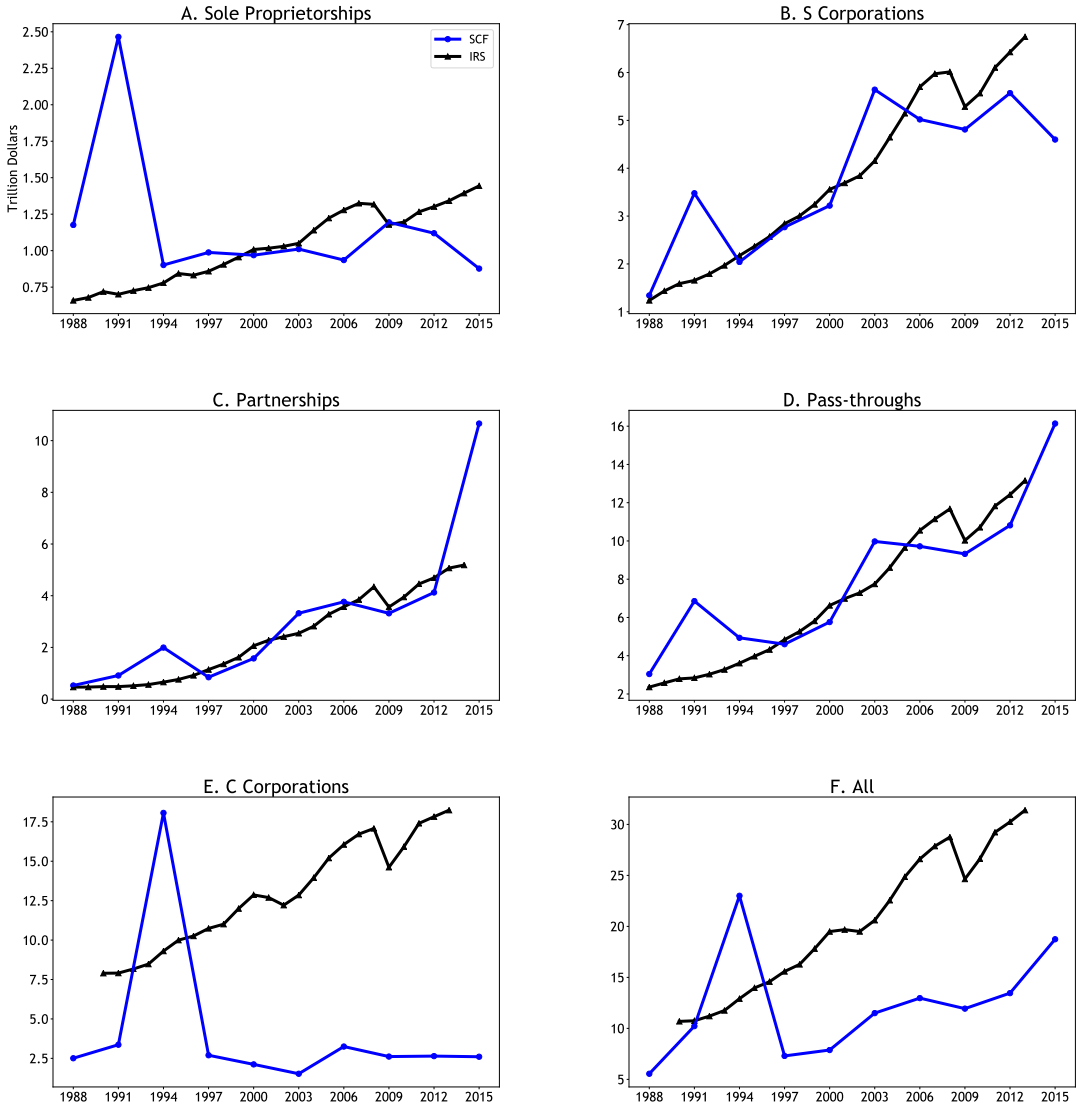
Note: This figure plots Schedule C, E, and F income per Form 1040 return. Schedule C comprises income derived from a business or profession, Schedule F comprises farm income, while Schedule E comprises income earned from rental real estate, royalties, partnerships, S corporations, estates, or trusts.

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



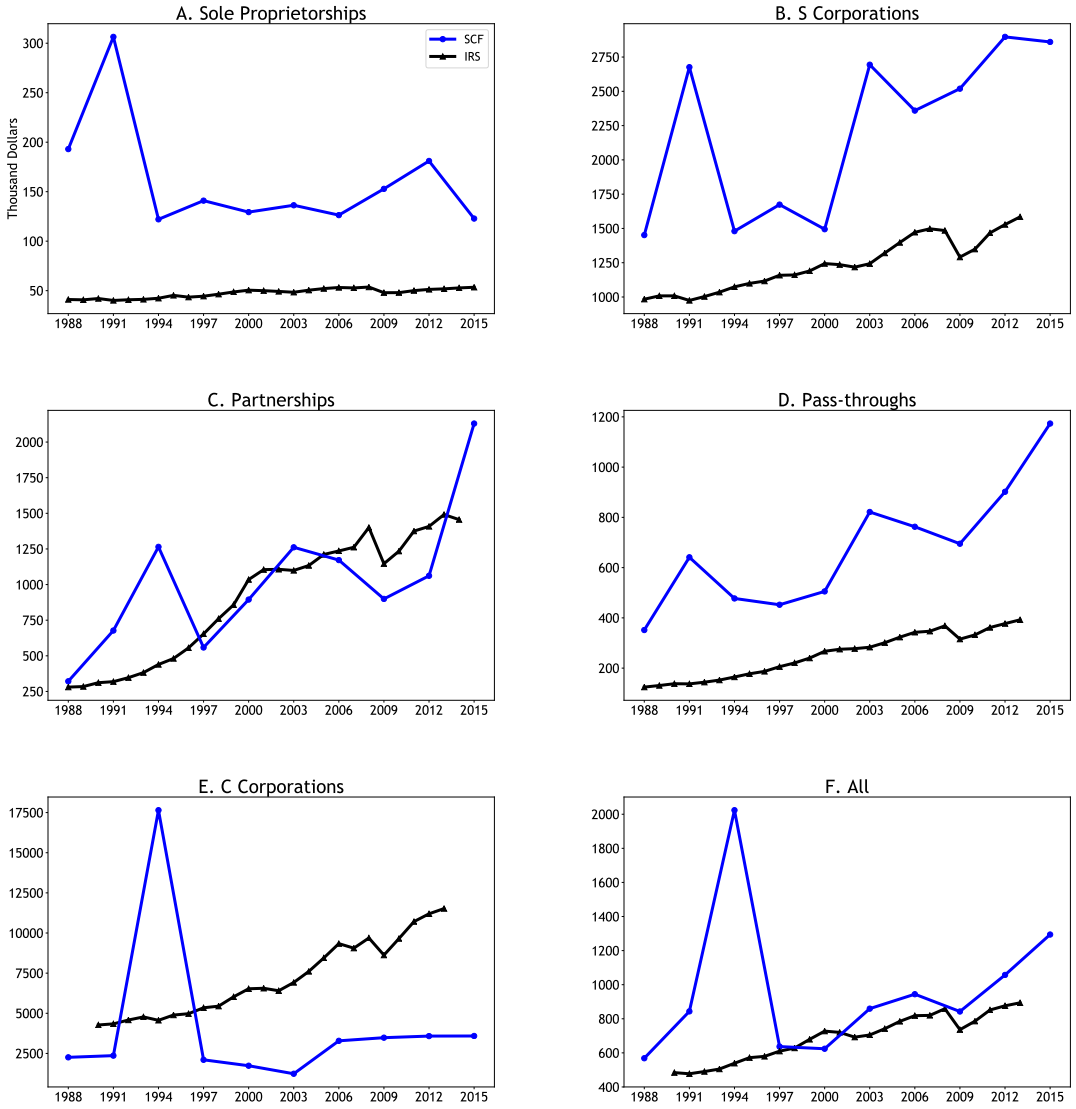
Note: This figure plots business income per return in the SCF for questions that ask respondents to report individual incomes listed on Form 1040, lines 12 plus 18, and business income on Schedule C of 1040, line 31.

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.

References

[1] Cooper Michael, John McClelland, James Pearce, Richard Prisinzano, Joseph Sullivan, Danny Yagan, Owen Zidar, and Eric Zwick. 2016. "Business in the United States: Who Owns It, and How Much Tax Do They Pay?" *Tax Policy and the Economy* 30: 91–128.

[2] Gurley-Calvez, Tami, Donald Bruce, E. J. Reedy, and Josh Russell. 2016. Comparing Survey Data and Tax Data: Differences in Reporting Across Businesses. Working Paper, Statistics of Income.

[3] Hurst, Erik, Geng Li, and Benjamin W. Pugsley. 2014. "Are Household Surveys Like Tax Forms? Evidence from Income Underreporting of the Self-Employed." *The Review of Economics and Statistics* 96(1): 19–33.

[4] Johnson, Barry W., and Kevin Moore. 2008. "Differences in Income Estimates Derived from Survey and Tax Data." In *Proceedings of the Joint Statistical Meetings, Survey Research Methods Section.* Alexandria: American Statistical Association.

[5] Kartashova, Katya. 2014. "Private Equity Premium Puzzle Revisited." *American Economic Review* 104(10): 3297–334.

[6] Moskowitz, Tobias J., and Annette Vissing-Jorgensen, 2002. "The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle?" *American Economic Review* 92(4): 745–78.