ECONOMIC RESEARCH
FEDERAL RESERVE BANK OF ST. LOUIS

What Do Survey Data Tell Us about US Businesses?

Federal Reserve Bank of St. Louis, Research Division, P.O. Box 442, St. Louis, MO 63166

What Do Survey Data Tell Us about US Businesses?

By Anmol Bhandari, Serdar Birinci, Ellen R. McGrattan, Kurt See[*]

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 several 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.
JEL: C83,E22,H25
Keywords: Survey data, business incomes, valuations, taxes, intangibles

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 survey responses to questions about businesses 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 amounts

on 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 IRS.

Averaging across survey years, we find that the SCF significantly overstates aggregate business incomes and business incomes per tax return for all business types. If we consolidate all businesses, we find an overstatement in total SCF aggregate business income of 156 percent relative to the IRS data and an overstatement of 434 percent when we compare business incomes per return. We also find a significant understatement of the number of tax returns across most business forms, which is worsening over time in the case of sole proprietorships and S corporations. Importantly, the overstatement of per-return and aggregate business incomes relative to IRS counterparts varies significantly in the cross section and year by year. We document that large differences between the SCF and IRS incomes exist even after adjusting the IRS data using information on tax audits to account for underreporting of incomes to the tax authorities. Averaging across all business types and all years, we again find an overstatement of aggregate business income and business income per return in the SCF relative to the IRS by 57 percent and 229 percent, respectively.

If we restrict attention to noncorporate businesses, we can compute comparable statistics across all four surveys. We find large differences in business income per owner across surveys in every available survey year. Futhermore, no survey provides reliable estimates for aggregate business income or per-owner income, regardless of whether we adjust the IRS data for misreporting. For example, if we do adjust for misreporting, the range of errors between IRS and survey incomes per owner is −16 percent (SIPP) to 481 percent (SCF).

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 in the Pratt's Stats database and from publicly traded company reports—both small and large—in the Center for Research in Security Prices (CRSP) database. 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 ranked by total assets. We also find that the SCF distributions are more right-skewed than those based on Pratt's or CRSP data.

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 business incomes per return is that the coverage of businesses in the surveys is different, and perhaps intentionally so, from the administrative data. To begin with, such a view is clearly inconsistent with our observation that aggregate business incomes are overstated in the SCF relative to the IRS for all legal forms and most years. However, we further explore this hypothesis by exploiting a unique feature of the SCF survey design: business incomes are recorded twice for many proprietorships— once in response to questions about individual incomes on Form 1040 asked of all tax filers and again in response to questions about business incomes on Schedule C asked of owners with actively managed businesses. 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 finding of an overstatement of business income per return 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 that 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 from 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 in 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.[1] For example, Johnson and Moore (2008) have noted that an overstatement of income could be the result of miscategorizing income in the surveys. We provide evidence that miscategorization does not explain the large discrepancies that we find by using the broader definition of business income recommended by Johnson and Moore (2008). For example, business owners might

---

[1] See Johnson and Moore (2008) for comparisons of business incomes in the SCF and IRS, Saez and Zucman (2016) for comparisons of a broader capital income measure in the SCF and IRS, and Kennickell, Kwast and Pogach (2017) for comparisons of business counts in the SCF and US Census Bureau.

confuse 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, and Smith et al. 2019.) Even for researchers that use administrative IRS data and capitalize taxable incomes, survey data serve as the only check on distributional assumptions and capitalization factors. (See Saez and Zucman 2016.) We find issues with both and therefore conclude that there is currently no reliable estimate of wealth inequality for the United States.

Survey data on businesses are a central input to studies of entrepreneurial choice, for both model calibrations and predictions. In many of these studies, parameter estimates—including key estimates for entrepreneurial span of control and borrowing constraints—are based solely on survey data. (See, for example, Evans and Jovanovic 1989, Quadrini 2000, Cagetti and De Nardi 2006, and Buera 2009.) Across different surveys and samples, there is a wide range of parameter estimates and, as a result, a wide range of predictions for business income shares and self-employment rates—key statistics when addressing normative issues. (See, for example, Meh 2005, Kitao 2008, Bruggemann 2017, and Imrohoroglu, Kumru and Nakornthab 2018, who use the survey-based calibrations of the earlier literature.)

Users of survey data have also come to very different conclusions about the payoffs to entrepreneurial activities. For example, some find puzzlingly low payoffs and conclude that entrepreneurships has large non pecuniary benefits (see Hamilton 2000, Moskowitz and Vissing-Jorgensen 2002, and Hurst and Pugsley 2011), whereas others find high payoffs and conclude that severe financial constraints hinder entrepreneurship (see Evans and Jovanovic 1989, Quadrini 2000, Cagetti and De Nardi 2006, Buera 2009). Our findings cast doubt on the facts that have been uncovered in the empirical literature—specifically those related to business incomes and rates of return—and thus raise issues concerning the theoretical developments and policy analyses that have been designed around them.

I. Business Incomes

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

A. Evidence

We start with a comparison of SCF and IRS. All SCF respondents are asked about incomes on the first page of the Individual Income Tax Form 1040 (Codes X5702–X5751). Respondents that own privately held businesses are asked if they have an active management role in any business (code X3104). If so, they are asked to report the legal form organization for up to three businesses (codes X3119, X3219, X3319), gross receipts (codes X3131, X3231, X3331), net incomes before taxes (codes X3132, X3232, X3332), and ownership shares (codes X3128, X3228, X3328). Information on business receipts and incomes is obtained from the respective business tax forms: Form 1040, Schedule C, for sole proprietorships, Form 1065 for partnerships, Form 1120S for S corporations, and Form 1120 for C corporations. In the SCF interviews, specific instructions are provided on which tax forms and line numbers to report. 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 weighted by ownership shares of the business.[2]

Figure 1 plots aggregate business incomes for tax years 1988–2015. The shaded regions for the survey data are 90 percent bootstrapped confidence intervals using SCF replicate weights. The solid lines are the comparable IRS aggregates. If we construct percentage errors (that is, $100(\text{SCF}-\text{IRS})/\text{IRS}$), we find that they are large, significantly different from zero, and vary considerably across all business types and all years. For all businesses (panel F), the average error is 156 percent with a range of 40 to 208 percent across survey years. In the case of pass-through businesses (panel D), which include sole proprietorships, partnerships and S corporations, the average error is 168 percent with a range of 61 to 263 percent across survey years.

Although the SCF asks respondents to report on specific lines of IRS tax forms, we also compare survey data with IRS business incomes adjusted to include estimates of unreported income. The estimates we use are based on imputations from the Bureau of Economic Analysis (BEA) and estimates from Johns and Slemrod (2010) and the Government Accountability Office (GAO) based on tax audit data. For the years 1988–2015, the BEA estimates that reported noncorporate tax incomes are lower by roughly 50 percent and corporate tax incomes lower by 23 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 on the order

---

[2]To be consistent with the IRS statistics, we assume that a business owner with multiple proprietorships files one return. The SCF only surveys owners of partnerships who are individuals, whereas the IRS includes other legal entities such as corporations. Therefore, we adjust the IRS data using an estimate of 32 percent for the fraction of total income generated by individual partners based on the Cooper et al. (2016) study of 2011 administrative tax data. Finally, the SCF only surveys privately held C corporations, and we use information from Schedule M3/M1 filed by publicly listed corporations to obtain the IRS counterparts of business income and number of returns for privately held C corporations.
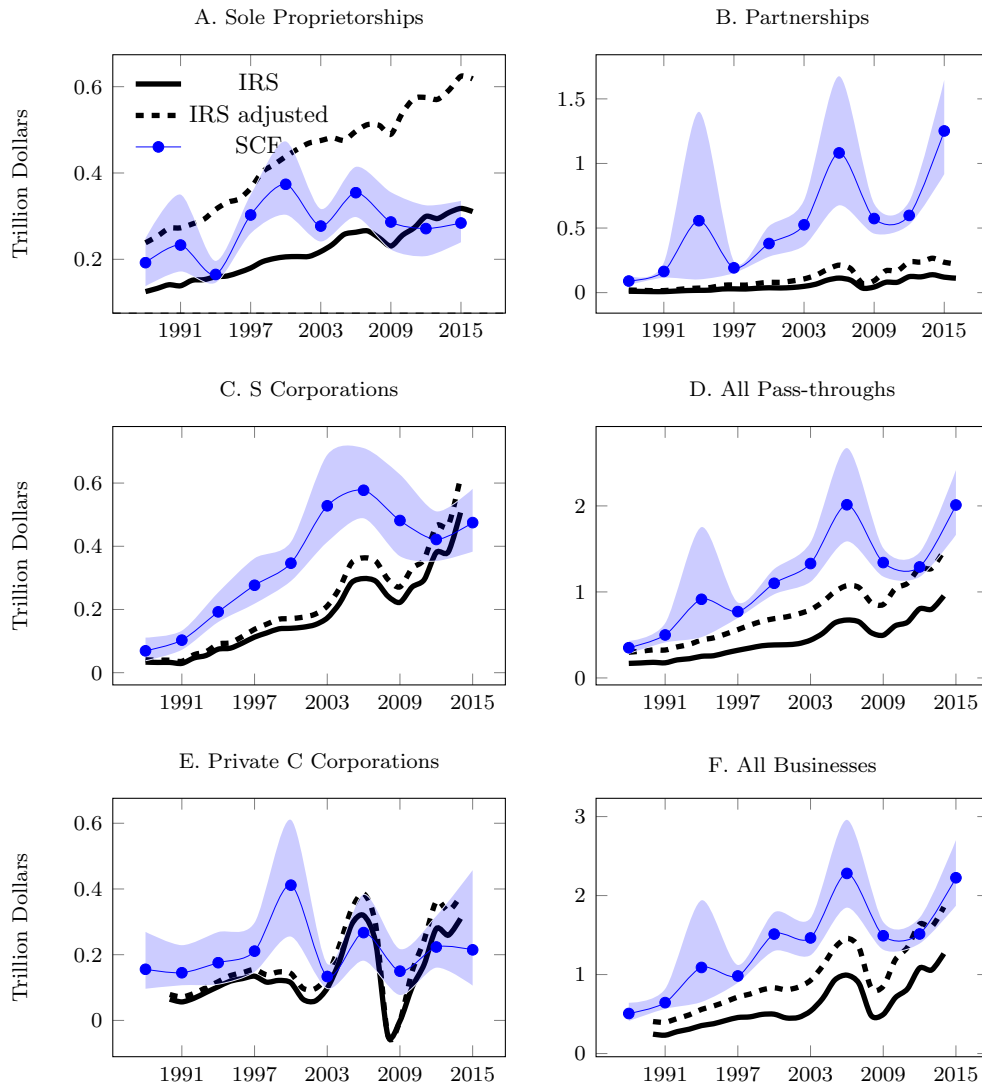
FIGURE 1. AGGREGATE BUSINESS INCOME BY LEGAL ENTITY, SCF VS. IRS

Note: This figure plots aggregate business income using the SCF, IRS. and IRS data adjusted for misreporting. In the SCF, business income is obtained by the weighted sum of households' share (X3128, X3228, X3328) of net business income (X3132, X3232, X3332) for each legal status (X3119, X3219, X3319) over time. For partnerships, the IRS data (solid black line) is adjusted based on the Cooper et al. (2016) study of administrative tax data so that it only covers income generated by individuals. For all business types, we calculate the IRS income adjusted for misreporting (dashed black line) by using the imputations from the Bureau of Economic Analysis (BEA) and estimates from Johns and Slemrod (2010) and the Government Accountability Office (GAO) based on tax audit data. The shaded region for the SCF shows the 90 percent confidence interval.

of 15 to 20 percent. We use this information to construct a measure of adjusted IRS income by adding back estimates of misreporting separately for each business form and year. The adjustments mitigate the overstatement of survey incomes, but large differences between the SCF and IRS data remain. The average error for all businesses in this case is 57 percent with a range of -8 to 95 percent across years. Except for sole proprietorships, the overstatements are present across legal forms and years. In the case of sole proprietorships we find a significant understatement of income in the SCF when we compare it with audit-corrected IRS data. Overall, the SCF answers are not consistent with either incomes reported to the IRS or actual incomes.

For studies relying on per capita information, a more relevant comparison is with business income per tax return. In Figure 2, we plot aggregate business incomes divided by the number of business tax returns for each business type for tax years 1988–2015, along with comparable data from the IRS and the audit-corrected measure that adds back underreported income. For all categories of businesses, we find a significant overstatement of SCF income per return and large year-to-year variation.[3] Our headline estimate for all businesses as shown in Figure 2, panel F, is an average error of 434 percent with a range of 254 to 567 percent when compared with data actually reported to the IRS. If we adjust the IRS incomes using audit data, the error is 229 percent, with a range of errors of 134 to 293 percent over the sample.

Our finding that errors in per-return incomes are higher than errors in total incomes is indicative of the fact that the number of returns filed by businesses is understated in the SCF, which turns out to be true for all business types except partnerships. In the case of sole proprietorships and S corporations, the understatement is significant and has worsened over time as the number of IRS filings has grown and the number reported in the SCF has not.[4]

If we restrict attention to noncorporate businesses, we can evaluate the accuracy of estimates in the CPS, PSID, and SIPP and compare the results with the SCF. In this case, we compare the results with the IRS data adjusted for misreporting since only SCF respondents are asked about specific lines on their tax returns. In Figure 3, we plot total noncorporate incomes in panel A and noncorporate income per owner in panel B. The results show large inconsistencies across surveys. The SCF overstates total noncorporate income relative to the IRS data adjusted for misreporting, whereas the CPS, PSID, and SIPP data understate by varying degrees. Differences across surveys and across time are large—with errors ranging from -71 to 106 percent—and may reflect differences in intended focus across surveys (for example, high versus low income or wealth households). When

---

[3]In principle, the exclusion of corporate partners in the SCF should not affect the business income per return. Here, we are operating under the assumption that the SCF is representative and that partnerships with all corporate partners (which are entirely missed by the SCF) either are small in terms of their share of aggregate business income or are not systematically different from the rest of the partnerships.

[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).

A. Sole Proprietorships

B. Partnerships

C. S Corporations

D. All Pass-throughs

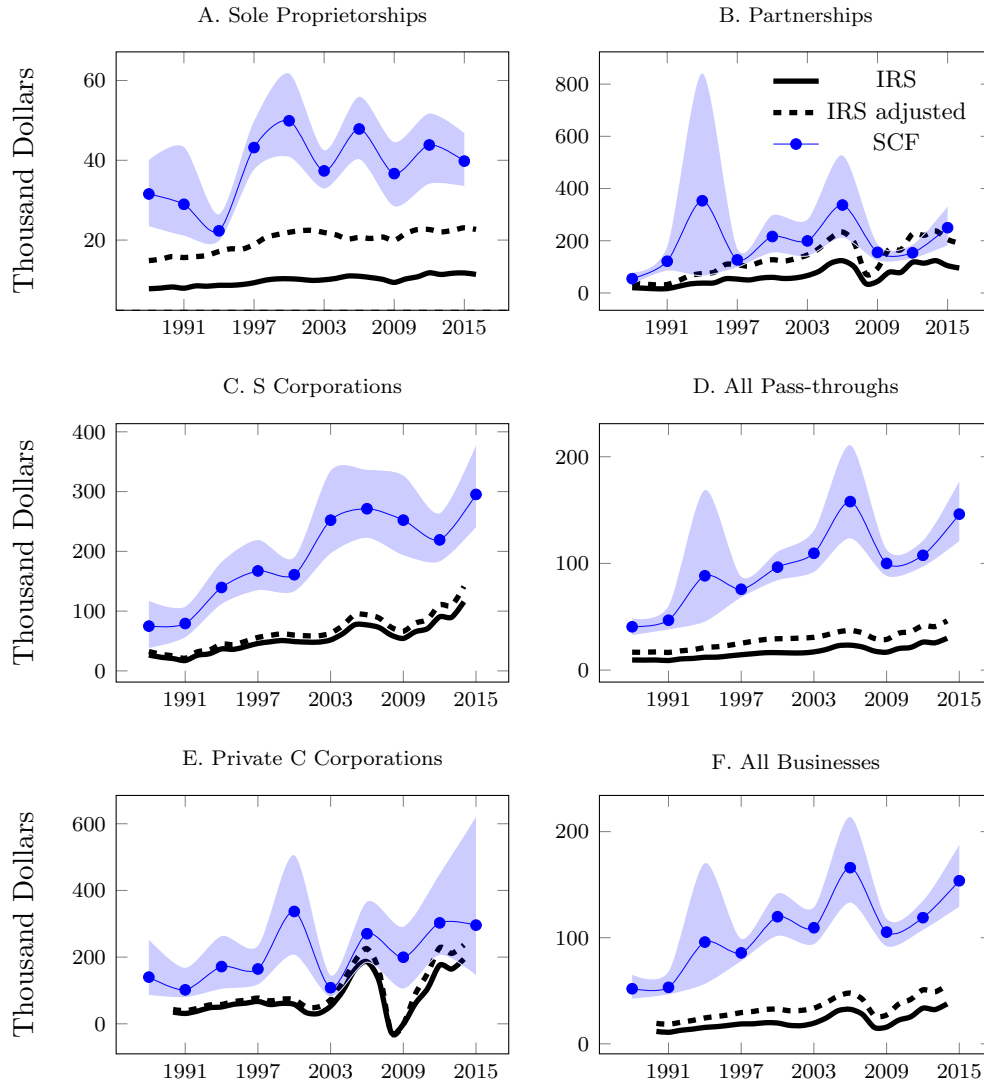E. Private C Corporations

F. All Businesses

FIGURE 2. BUSINESS INCOME PER RETURN BY LEGAL ENTITY, SCF VS. IRS

Note: This figure plots business income per tax return using the SCF, IRS, and IRS data adjusted for tax misreporting. In the SCF, business income is obtained by the weighted sum of households' share (X3128, X3228, X3328) of net business income (X3132, X3232, X3332) for each legal status (X3119, X3219, X3319) over time. Similarly, the number of business tax returns is calculated by the weighted sum of households' share of each legal status over time. For partnerships, the IRS data (solid black line) is adjusted based on the Cooper et al. (2016) study of administrative tax data so that it only covers income generated by individuals and their number of tax returns. Finally, to be consistent with the IRS statistics, we assume in the SCF that a business owner with multiple proprietorships files one return. For all business types, we calculate the IRS incomes adjusted for misreporting (dashed black line) using the imputations from the Bureau of Economic Analysis (BEA) and estimates from Johns and Slemrod (2010) and the Government Accountability Office (GAO) based on tax audit data. The shaded region for the SCF shows the 90 percent confidence interval.

we divide by numbers of owners, we again find large differences across surveys and relative to the IRS data adjusted for misreporting. In this case, all surveys overstate business incomes, with the exception of the SIPP. Errors range from −16 to 481 percent when compared with IRS data adjusted for misreporting.[5]
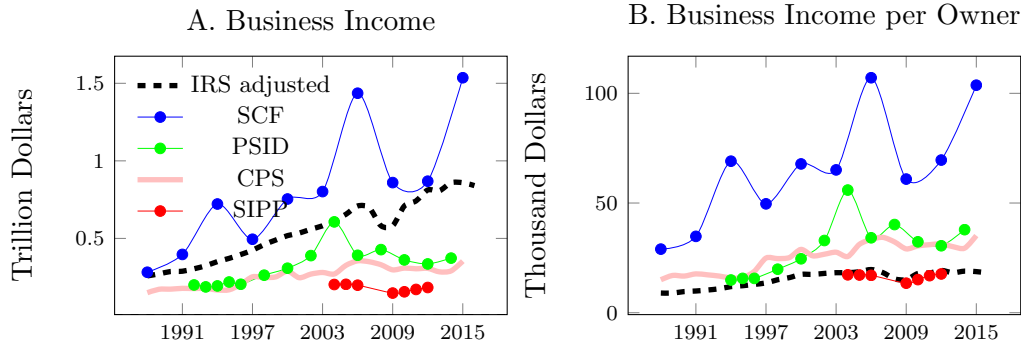

FIGURE 3. UNINCORPORATED BUSINESSES IN SCF, SIPP, PSID, AND CPS VS. IRS

Note: This figure plots the aggregate business income and business income per owner for unincorporated businesses in the SCF, PSID, CPS, SIPP, and IRS adjusted for tax misreporting. For partnerships, the IRS data (dashed black line) is adjusted based on the estimate of Cooper et al. (2016) that 73.9 percent of the partners are individuals and the rest are corporations that are not surveyed in the SCF. Finally, to be consistent with the IRS statistics, we assume in the SCF that a business owner with multiple proprietorships files one return. 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 aggregate business income of unincorporated businesses.

B. Hypotheses

Next, we consider possible reasons for the large differences between survey and IRS data and some possible corrections.

OWNERS WITH LOW INCOME ARE UNDERREPRESENTED

The overstatement of income per return and understatement of business returns in the SCF could arise because of differences in coverage between the SCF and the administrative IRS data. Suppose owners with little business attachment—say, part-time Uber drivers—might respond that they have no active management role in a business when interviewed. The survey would undercount the number of businesses and overstate income per return if the owners with an active

---

[5]Hurst, Li and Pugsley (2014) infer estimates of "true" income for households with self-employed male heads using Engel curve estimates for wage and salary workers and expenditure data for the self-employed. For the PSID, they find evidence that income per household of self-employed heads is underreported in the survey relative to their measure of true income, whereas we find that business income per owner is overreported relative to the IRS data adjusted for misreporting.

management role earn higher incomes. If this was intended in the design of the survey, such an overstatement of incomes per return would be inconsequential for applied research on businesses.[6] Although such a view is clearly inconsistent with the systematic overstatement of aggregate business incomes across years and legal forms documented in the previous section, we investigate this issue for sole proprietorships in more detail by exploiting the structure of questions in the SCF.

Sole proprietors 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). The implicit intent of the design was to gather more information from households with a high business attachment. Using a Venn diagram and SCF data for 2015, we show in Figure 4 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 (9.3 million) reporting neither actively managing a business nor having an "interest" in a business without an active management role.[7] 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 9.3 million earned very little business income. But as we show in Figure 4, the SCF reports that these 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 actively managing a business.

Although the evidence in Figure 4 clearly shows that incomes per return for those actively managing a business are not higher than average, it does not rule out underrepresentation of low-income owners more generally. In the case of sole proprietors, we further investigate by computing both numbers of returns 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 this for those claiming an active management role, 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, whereas IRS data show a rise from 5 million in 1988 to over 12 million in 2015. Even if we compute the statistics for all sole proprietors in the SCF, that is, those who claim to be actively managing a business and those who do not, (using code X5704), the total below-median group is only 6.2 million in 2015, roughly half of the IRS data. These findings suggest a significant underrepresentation of low-income businesses, which leads to an overstatement of business incomes if business income is positively correlated

---

[6]See Kennickell, Kwast and Pogach (2017) for such a view.

[7]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.

Nonzero 1040, lines 12+18
($583 billion, 16.2 million)

Actively Managed Business
($229 billion, 7.1 million)

$303 billion,
9.2 million

$206 billion,
6.3 million

$74 billion,
0.7 million

NA

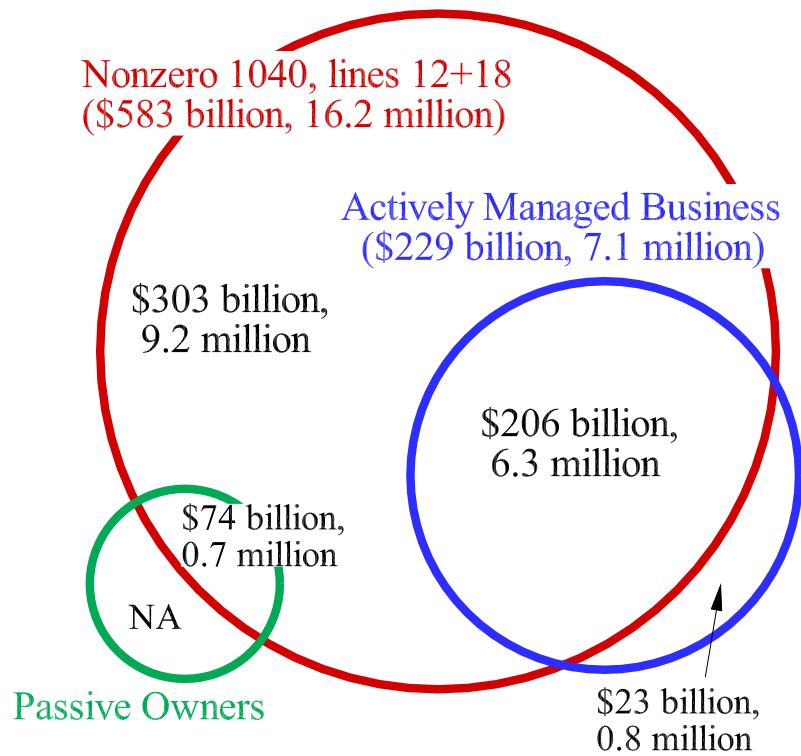Passive Owners

$23 billion,
0.8 million

FIGURE 4. SOLE PROPRIETOR RESPONSES IN THE SCF

Note: This figure shows incomes and counts for three sets of sole proprietors. In red are those who have a nonzero line 12 plus 18 on Form 1040, in blue are those who report to be actively managing a business, and in green are those who report owning a business or having an interest in a business without having an active management role. Some of the groups overlap and in such cases we use incomes from Form 1040, lines 12 plus 18. The counts and incomes of passive owners without income on Form 1040, lines 12 plus 18, are not available (NA) because the SCF stopped identifying the legal form of organization for passive owners after 2007. In the online Appendix, we provide details for other survey years.

with AGI. If we analyze income shares over time, we find that the share of income for those with below-median AGI is nearly doubled or halved from one survey to the next, whereas the IRS share has steadily grown.

These observations call for several survey redesigns. The SCF should make precise the notion of actively managing a business, implement internal consistency checks mid-interview, and, if possible, link future surveys to administrative data. This redesign is especially 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.[8] 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. To correct the problem of underrepresentation, the SCF needs to expand the sampling frame in future surveys.

Owners misreport business losses

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 profits and losses separately. For pass-through businesses, the SCF overstates the income per return for profitable businesses by an average of 278 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, whereas 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 might be negative.

Owners misclassify business incomes

Another source of measurement error in the SCF is the respondent's possible confusion about closely related categories of business income. For example, when

---

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

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 they might include 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.

To test this hypothesis, we compare Schedule E incomes per return and again find that the SCF estimates are overstated relative to the 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 the IRS data, we find errors averaging 90 percent with a range of 47 to 180 percent over the sample. If we follow the recommendation of Johnson and Moore (2008) and broaden the concept of business income even further by adding capital gains (which includes non business 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—neither for research on US businesses nor for 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.

II. 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 first document that for virtually all subsamples and all years, the survey yields are significantly higher and more right-skewed than comparable measures. We then evaluate possible reasons for the discrepancies.

A. 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

Table 1—Net Income Yields Comparison

	Means		Percentiles		
	Equal weights	Value weights	25th	p50	p75
SCF					
All businesses	102.5	19.1	0.9	0.9	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	220.4	22.6	0.8	19.6	70.6
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	−9.3	7.3	−5.3	5.4	10.4
Small businesses	−26.6	−8.5	−29.6	−7.8	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.

Pratt's Stats.[9] 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 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 to 102 percent for the SCF, and the Pratt's value-weighted yield is 2 percent compared to 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

[9]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 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.[10]

Next, we compare income yields for all businesses in the SCF with those for the CRSP database (merged with Compustat). 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, while positive, is much lower than that for the SCF (7 percent). 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, whereas the CRSP yield for all businesses is 10 percent.

These findings appear to be inconsistent with those of 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. The main difference is the concept of return. Their measure is equal to the sum of a value-weighted income yield and an imputed capital gain.[11] 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 estimates to the value-weighted mean holding period returns on the CRSP index portfolio, they find private returns that are similar in magnitude to the returns for listed public firms.

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. (See Bhandari et al. 2019.) 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

[10]The results are similar for other surveys if we restrict attention to noncorporate businesses. (See Bhandari et al. (2019).)

[11]Kartashova (2014) also finds that a longer sample raises the SCF-CRSP return gap but uses the same concept of return as in Moskowitz and Vissing-Jorgensen (2002).

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

B. 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 sold or publicly traded businesses.

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 that the results are nearly the same as for all businesses.

CRSP yields could 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 quintile of firms ranked by the book value of assets.[12] Given that 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.

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

---

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

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.

III. Conclusion

Theories and policies are being developed around survey "facts," but the evidence suggests that these should be treated with great 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

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

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.

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.

Bruggemann, Bettina. 2017. "Higher Taxes at the Top: The Role of Entrepreneurs." Working Paper, McMaster University.

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

Cagetti, Marco, and Mariacristina De Nardi. 2006. "Entrepreneurship, Frictions, and Wealth." *Journal of Political Economy*, 114(5): 835–870.

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(1): 91–128.

Evans, David S., and Boyan Jovanovic. 1989. "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints." *Journal of Political Economy*, 97(4): 808–827.

Hamilton, Barton H. 2000. "Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment." *Journal of Political Economy*, 108(3): 604–631.

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

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.

Imrohoroglu, Ayse, Cagri Kumru, and Arm Nakornthab. 2018. "Revisiting Tax on Top Income." Working Paper, University of Southern California.

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

Johnson, Barry W., and Kevin Moore. 2008. "Differences in Income Estimates Derived from Survey and Tax Data." SOI Paper Series.

Kartashova, Katya. 2014. "Private Equity Premium Puzzle Revisited." *American Economic Review*, 104(10): 3297–3334.

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." *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges*, ed. John Haltiwanger, Erik Hurst and Antoinette Schoar, 291–340. Chicago: University of Chicago Press.

Kitao, Sagiri. 2008. "Entrepreneurship, Taxation, and Capital Investment." *Review of Economic Dynamics*, 11(1): 44–69.

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.

Meh, Cesaire A. 2005. "Entrepreneurship, Wealth Inequality, and Taxation." *Review of Economic Dynamics*, 8(3): 688–719.

Moskowitz, Tobias J., and Annette Vissing-Jorgensen. 2002. "The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle?" *American Economic Review*, 92(4): 745–778.

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

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): 519–578.

Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick. 2019. "Capitalists in the Twenty-First Century." *Quarterly Journal of Economics*, 134(4): 1675–1745.