

NBER WORKING PAPER SERIES

WHAT DO SURVEY DATA TELL US ABOUT US BUSINESSES?

Anmol Bhandari  
Serdar Birinci  
Ellen McGrattan  
Kurt See

Working Paper 26063  
<http://www.nber.org/papers/w26063>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
July 2019

An online appendix (Bhandari et al. 2019) is available at [www.minneapolisfed.org/research/sr/sr578.pdf](http://www.minneapolisfed.org/research/sr/sr578.pdf). Bhandari and See acknowledge support from the Heller-Hurwicz Economic Institute, and McGrattan acknowledges support from the NSF (#1657891). For helpful comments, we thank seminar and conference participants at Colorado, Minnesota, Yale, Autonoma, ASU, NYU, Columbia, UCSB, SITE, CMU, Midwest Macro, LAMES, Michigan, WU in St. Louis, and Stanford. For excellent editorial assistance, we thank Joan Gieseke. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research, Federal Reserve Bank of Minneapolis, or the Federal Reserve System.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Anmol Bhandari, Serdar Birinci, Ellen McGrattan, and Kurt See. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

What Do Survey Data Tell Us about US Businesses?

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

NBER Working Paper No. 26063

July 2019

JEL No. C83,E22,H25

### **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.

Anmol Bhandari  
Department of Economics  
University of Minnesota  
1925 Fourth Street South  
Minneapolis, MN 55455  
and NBER  
bhandari@umn.edu

Serdar Birinci  
University of Minnesota  
One Federal Reserve Bank Plaza  
St. Louis, MO 63102  
birin001@umn.edu

Ellen McGrattan  
Department of Economics  
University of Minnesota  
4-101 Hanson Hall  
1925 4th St S  
Minneapolis, MN 55455  
and NBER  
erm@umn.edu

Kurt See  
University of Minnesota  
234 Wellington Street  
Ottawa K1A 0G9  
Canada  
seexx028@umn.edu

A data appendix is available at <http://www.nber.org/data-appendix/w26063>  
Replication files are available at <http://users.cla.umn.edu/~erm/data/bbms/>

# 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 IRS.

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. Averaging 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 business incomes per return is that survey data omit owners with little business at-

tachment.<sup>1</sup> 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 business 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.<sup>2</sup> For example, Johnson and Moore (2008) have noted that

---

<sup>1</sup>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.

<sup>2</sup>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

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 non-pecuniary 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 cast doubt on the facts that have been uncovered

---

and IRS, and Kennickell, Kwast, and Pogach (2017) for comparisons of business counts in the SCF and US Census Bureau.

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.<sup>3</sup>

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,  $100(\text{SCF} - \text{IRS})/\text{IRS}$ ), 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 years. For S corporations (Panel B), the average error is 273 percent with a

---

<sup>3</sup>To be consistent with the IRS statistics, we assume that a business owner with multiple proprietorships files one return.

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, 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 as the number of IRS filings has grown and the number reported in the SCF has not.<sup>4</sup> 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 number of SCF C corporations should be lower in the SCF than the IRS

---

<sup>4</sup>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).



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, 137, 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 are driven by owners with little business attachment. 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 3 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.<sup>5</sup> 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 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 being actively managing. In sum, the evidence from 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, to link future surveys to administrative 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,

---

<sup>5</sup>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.

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.<sup>6</sup> 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 profits and losses separately. For pass-through businesses, the SCF overstates the income per return for profitable businesses by an average of 277 percent,

---

<sup>6</sup>We found virtually no difference in behavior when conditioning on income.

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 might 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 on 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 to 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 respondent’s 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, 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 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 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—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.<sup>7</sup> 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

---

<sup>7</sup>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.

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 as 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 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 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, 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, 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 few differences—we use a longer sample and comparable measures of pretax earnings in SCF and CRSP as against imputing retained earnings for pass-through firms, but the more important difference for the quantitative results is the concept of return.<sup>8</sup>

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 to the value-weighted mean holding period return on the CRSP index portfolio they find that 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 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 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

---

<sup>8</sup>Kartashova (2014) also finds that a longer sample raises the SCF-CRSP return gap, but uses the same concept of return.



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 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 quintile of firms ranked by the book value of assets.<sup>9</sup> Given we have data on all busi-

---

<sup>9</sup>The results are similar for other definitions of “small,” for example, based on gross sales

ness 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.<sup>10</sup>

## 4 Conclusion

The need for accurate data on private businesses is especially urgent given 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 caution. We hope and expect that our analysis will lead to improved measurement in future surveys. Measurement problems related to business incomes

---

or market capitalization.

<sup>10</sup>An attempt in this direction is ongoing work in Bhandari and McGrattan (2019).

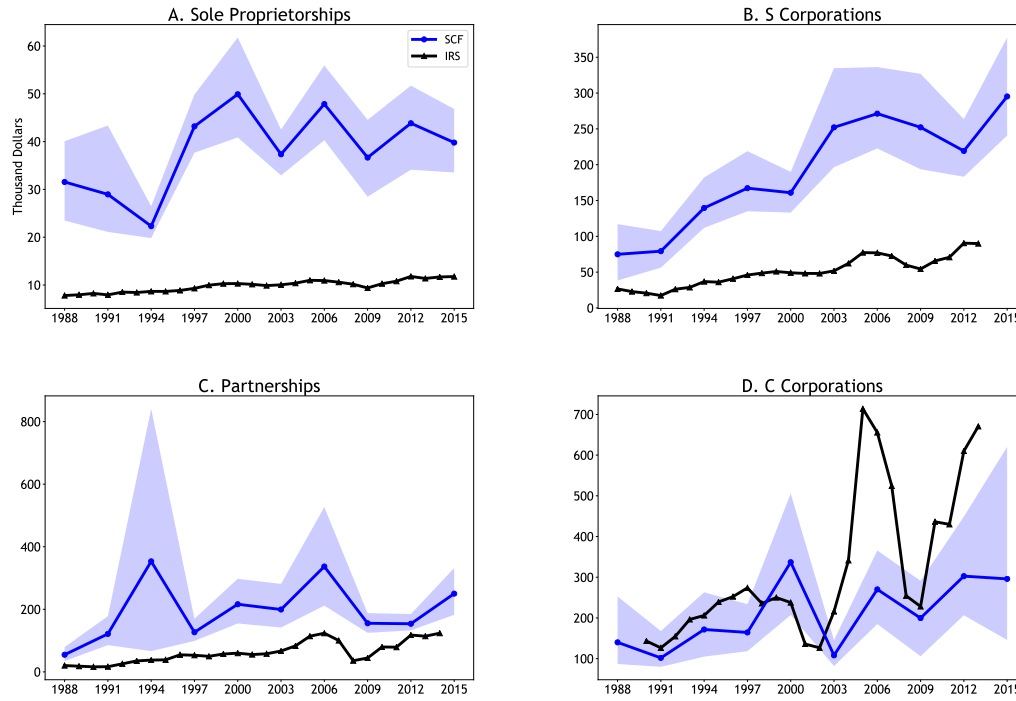
are surmountable given that respondents are asked about specific lines on tax forms. Measurement problems related to business valuations and returns may be insurmountable without data on actual business sales transactions or a theoretical framework and a method of indirect inference.

## 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.

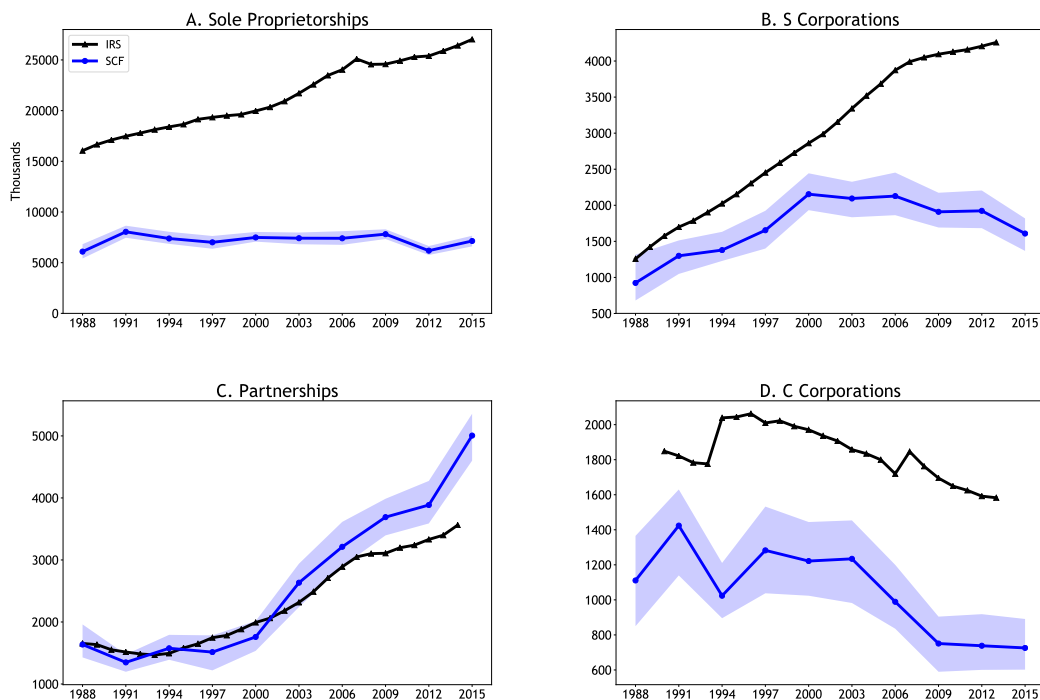
- [13] Kuhn, Moritz, and José-Victor Rios-Rull. 2016. “2013 Update on the US Earnings, Income, and Wealth Distributional Facts: A View from Macroeconomics.” *Federal Reserve Bank of Minneapolis Quarterly Review* 37(1): 2–73.
- [14] Moskowitz, Tobias J., and Annette Vissing-Jorgensen, 2002. “The Returns 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



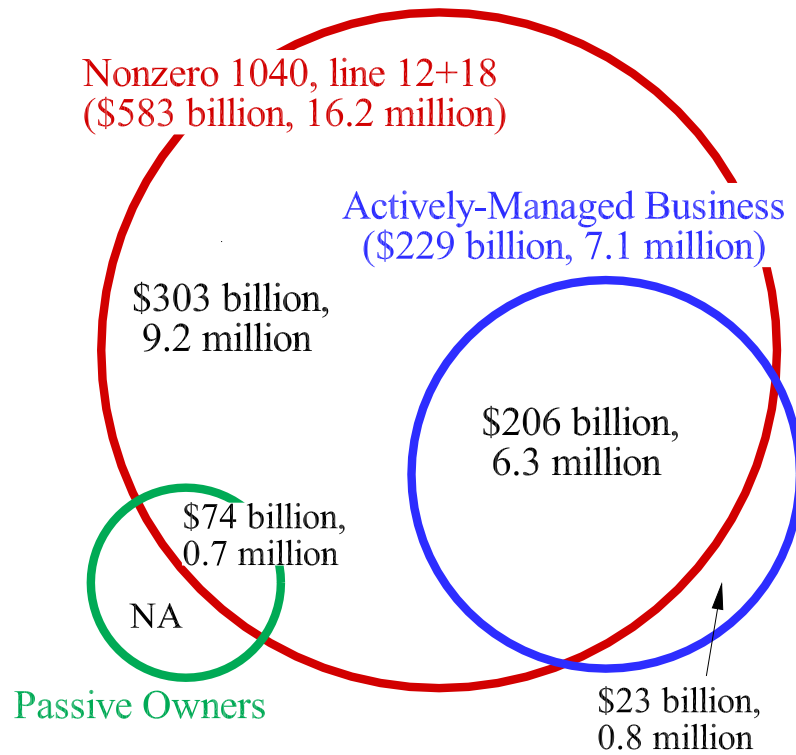
*Note:* This figure plots business income per business tax return in the IRS and the SCF as reported on Form 1040 Schedule C for sole proprietorships, Form 1120S for S corporations, Form 1065 for partnerships, and Form 1120 for C corporations. IRS data for partnerships, S corporations, and C corporations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

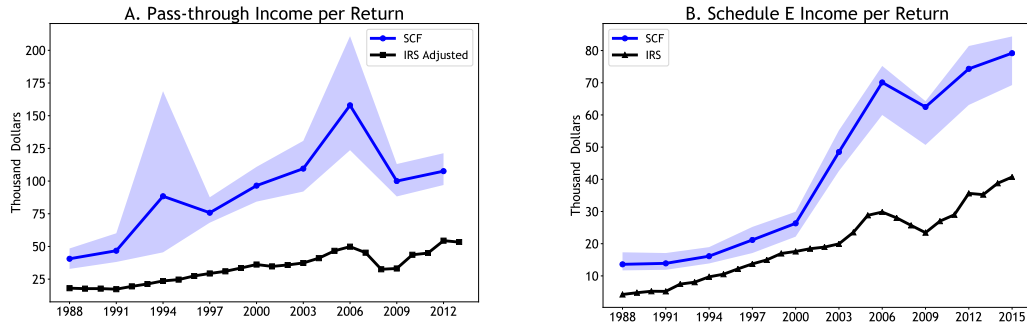
Figure 3  
Sole Proprietor Responses in the SCF



*Note:* This figure shows incomes and counts for three sets of sole proprietors. In red are those who have a 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, line 12 plus 18. The counts and incomes of passive owners without income on Form 1040 lines 12+18 is not available (NA) because the SCF stopped identifying legal form of organization for passive owners after 2007. In the online 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	−5.5	5.4	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 for different subgroups in the SCF (1988–2015), PSID (1998–2014), SIPP (2004, 2005, 2009–2011), Pratt's Stats (1993–2017) and CRSP-Compustat (1988–2015). 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.