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

December 2019

Anmol Bhandari
University of Minnesota and NBER

Serdar Birinci
Federal Reserve Bank of St. Louis

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

Kurt See
Bank of Canada

*The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis, the Federal Reserve Bank of St. Louis, the Bank of Canada, or the National Bureau of Economic Research.

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

1 Data Sources

The main data sources are:

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

Besides the main data sources listed above, we also use information from the national income and product accounts and fixed asset tables of the Bureau of Economic Analysis; financial accounts by the Board of Governors of the Federal Reserve System; and the Kauffman Firm Survey (KFS) by the Kauffman Foundation.

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

Table 1: Data Construction

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

Table 1: Data Contstruction (cont.)

Line No.	**Variable**	**Description**	**Sources**	**Remarks**
9	Private C corp. income, receipts, and number of returns (IRS)	Form 1120, lines 30 and 1c for privately held C corporations	SOI Table 16: *Balance Sheet, Income Statement, Tax, and Selected Other Items, by Major Industry* SOI Schedule M3 information *A First Look at [Year] M3 reports*	Data available for 1990–2013
10	C corp. income, receipts, and number of returns (SCF)		See [4] above.	See [4] above.
11	Unincorporated business income and number of owners (IRS)	Income from partnerships and sole proprietorship	See [3] and [5] above.	See [3] and [5] above.
12	Unincorporated business income and number of owners (SCF)		See [4] and [6] above. X4106, X4706 to infer self-employed members of household.	If the head and the spouse are both not self-employed but the household has an actively managed partnership business, then we infer that someone else in the household is self-employed and the owner of that business.
13	Unincorporated business income and number of owners (SIPP)		Business income: TBMSUM1–SUM2 Business shares: TPRFTB1–B2 Legal form: EINCPB1–B2 and EPROPB1–B2	Data available for 2004–2006 and 2009–2013.[1] SIPP reports information about two businesses on a monthly frequency. We average to report annual values.
14	Unincorporated business income and number of owners (PSID)		Code names vary across surveys but can be linked. For 2005 business income: ER65192, business owners: ER60932.	We use longitudinal family weights to compute aggregates.
15	Unincorporated business income and number of owners (CPS)		Business income: INCBUS Class of worker: CLASSWKR	1989-2016 Annual Social and Economic Supplement. Respondents who identify as self-employed running an incorporated business are considered business owners.

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

Table 1: Data Contstruction (cont.)

Line No.	**Variable**	**Description**	**Sources**	**Remarks**
16	Income yield (SCF)	Business Income / Business Value	For business income see [4], [6], [8] above. X3129–X3329 for business valuations.	We exclude businesses with net worth less than the bottom 1st percentile of the net worth distribution, conditional on having positive net worth.
17	Income yield (CRSP/Compustat)		PI/ (csho x prcc_f) where PI: Pretax Income csho: Common Shares Outstanding prcc_f: Price Close - Annual - Fiscal	Sample includes firms incorporated in the US and for 1988–2017. Income yields are passed through a 1 percent winsorizing filter. For the "small" firms, we use bottom 20 percent of the firms by (i) market value=csho x prcc_f, (ii) book value of assets (AT), and (iii) by gross sales (SALE).
18	Income yield (Pratts Stats)		EBT/MVIC where EBT=Earnings before Taxes, MIVC=Total consideration paid to the seller	Sample truncated to exclude transactions where the EBT and MVIC are in the top 1 percent and bottom 1 percent of their respective series.
19	Income yield (PSID)		See [14] above for business income. For business valuations, code ER61736 in 2015 survey.	
20	Income yield (SIPP)		For business income, see [13] above. TVBVA1-VA2 for business valuations of assets. TVBDE1-DE2 for value of debts.	See [13] above.

In addition to the variables listed in Table 1, we use BEA estimates of income misreporting by noncorporate and corporate businesses and General Accountability Office (GAO) estimates of income misreporting by S corporations to adjust IRS business incomes. BEA estimates of income misreporting over time are obtained from NIPA Table 7.14 (line 2) and Table 7.16 (line 2). The GAO estimates are taken from reports GAO 14-453 and 10-195, which summarize the progress of the tax compliance studies conducted by the IRS through the National Research Program. Moreover, given that the survey data collect information on individual business owners, we compare business income of individuals, number of returns filed by individuals, and number of individual owners between the survey data and adjusted IRS data for partnerships. Namely, we adjust IRS partnership income, number of partnership returns, and number of partners using estimates from Cooper et al. (2016), who use administrative US tax data and find that while individual partners constitute 73.9 percent of all partners, they receive 31.5 percent of all partnership income. We further explain how we make these adjustments to the IRS data in detail below.

2 Measurement

In this section, we first explain (i) the variables we use in the SCF and (ii) how we use them to calculate the moments we discuss in the main text and in this online Appendix. We then provide more details on IRS dataset and the adjustments we make to it.

2.1 Definition of variables in the SCF

- *Business legal status*: SCF collects information on the type of actively managed business up to three businesses until 2007 and up to two businesses after 2007. For each business respectively, SCF variable codes X3119, X3219, and X3319 ask "*Is it (the business) a partnership, a sole-proprietorship, an LLC, a subchapter S corporation, another type of corporation, or something else?*"

- *Business net income*: SCF variable codes X3132, X3232, and X3332 ask "*What was the business's total pre-tax net income?*" Moreover, starting with 2004 survey, the respondents are instructed (with a probe) to refer to specific lines on their tax return based on their business legal status:

  - Partnership: Ordinary income/loss — IRS Form 1065, line 22
  - Sole proprietorship: Net profit/loss — IRS Form 1040, Schedule C, line 31
  - S corporation: Net income — IRS Form 1120S, line 21
  - Other corporation: Taxable income before net operating loss deduction — IRS Form 1120, line 30

- *Business ownership share*: SCF variable codes X3128, X3228, and X3328 ask "*What percentage of the business do you (and your family living here) own?*"

- *Self-employment:* The SCF also collects information on employment status of the respondent and the respondent's husband/wife/partner/spouse) separately. Variable codes X4106 and X4706 ask "*(Do you/Does he/Does she/Does he or she) work for someone else, (are you/is he/is she/is he or she) self-employed, or something else?*" We use answers to these questions to infer business ownership of the respondent and the respondent's husband/wife/partner/spouse.

- *Schedule E income:* SCF variable X5713 asks "*Did you (or anyone else) have income from other businesses or investments, net rent, trusts, or royalties?*" with a probe "*Did you file a Schedule E?*" We use this information when we count the number of Schedule E returns that are attached to individual income tax (Form 1040).

- *Filing of income tax return:* SCF variable X5744 asks "*Did you (or your {husband/wife/partner/spouse}) file, or do you expect to file, a Federal Income tax return?*" Also, X5746 asks "*(Did/Will) you and your (husband/wife/partner/spouse) file a joint return, (did/will) you file separately, or (did/will) only one of you file?*" From the answers to these questions, we can understand whether they file jointly or have only one of them file (implying one income tax return) or they file separately (implying two income tax retruns). Thus, we use answers to these questions to calculate the number of individual income tax returns filed within the family.

- *Adjusted gross income:* Several SCF variables measure separate components of AGI. Importantly, each of them refers to different lines of Form 1040:
  - X5702: "*In total, what was your (family's) annual income from wages and salaries, before deductions for taxes and anything else?*" Probe: Form 1040, line 7

- X5704: "*In total, what was your (family's) net annual income from a sole proprietorship or a farm, before deductions for taxes and anything else?*" Probe: Form 1040, lines 12,18
- X5706: "*In total, what was your (family's) annual income from non-taxable investments such as municipal bonds, before deductions for taxes and anything else?*" Probe: Form 1040, line 8b
- X5708: "*In total, what was your (family's) annual income from other interest, before deductions for taxes and anything else?* " Probe: Form, 1040 line 8a
- X5710: "*In total, what was your (family's) annual income from dividends, before deductions for taxes and anything else?*" Probe: Form 1040, line 9a
- X5712: "*In total, what was your (family's) annual income from net gains or losses from mutual funds or from the sale of stocks, bonds, or real estate, before deductions for taxes and anything else?* " Probe: Form 1040, lines 13,14
- X5714: "*In total, what was your (family's) annual income from other businesses or investments, net rent, trusts, or royalties, before deductions for taxes and anything else?* " Probe: Form 1040, line 17
- X5716: "*In total, what was your (family's) annual income from unemployment or worker's compensation, before deductions for taxes and anything else?*" Probe: Form 1040, line 19
- X5718: "*In total, what was your (family's) annual income from child support or alimony which you (or your family here) received in 2009, before deductions for taxes and anything else?*" Probe: Form 1040, line 11
- X5722: "*(Including the retirement income you told me about, in/In) total, what was your (family's) net income from Social Security or other pensions, annuities, or other disability or retirement programs in 2009, before deductions for taxes and anything else? (Please do not include withdrawals from IRAS, 401(k)s and other such retirement accounts.)*" Probe: Form 1040, lines 16a, 20a
- X5724: "*(Other than withdrawals from account-type pensions or IRAs you told me about earlier in the interview, in/In) total, what was your (family's) annual income from any*

other sources in 2009, before deductions for taxes and anything else?" Probe: Form 1040, lines 10, 21

- *Business net worth:* SCF variables X3129, X3229, and X3329 collect information on the net worth of each business (up to three) that households own. They ask "*What is the net worth of (your share of) this business?*" with a probe "*What could you sell it for?*" We use answers to these questions to obtain the value of each business, which is used in business rate of return calculations.

- *Business receipts:* SCF variables X3131, X3231, and X3331 collect information on the business receipts of each business (up to three) that households own. They ask "*What were the gross receipts or gross sales of the business as a whole?*" Moreover, starting with the 2004 survey, the respondents are instructed (with a probe) to refer specific lines of their tax return based on their business's legal status:

  - Partnership: IRS Form 1065, line 1C
  - Sole proprietorship: IRS Form 1040, Schedule C, line 1
  - S corporation: IRS Form 1120S, line 1C
  - Other corporation: IRS Form 1120, line 1C

- *Checking relevant documents:* Since the 1995 survey, SCF collects information on the frequency of checking any documents when answering interview questions coded in X6536 with the question "*Did the respondent refer to any documents to answer any questions?*" with possible answers "*frequently,*" "*sometimes,*" "*rarely,*" or "*never.*" Moreover, variable X7451 informs us about whether the respondent referred to income tax documents, and variables X7452 through X7455 inform us about whether the respondent referred to other financial documents, namely, pension documents, account statements, investment or business records, and loan documents, respectively.

2.2 Calculation of moments from the SCF

- *Number of business tax returns*: Here is the general structure for calculating the total number of business tax returns by legal status type:

$$\sum_i \left[\left(\sum_j \text{Legal status}_{i,j} \times \text{Ownership share}_{i,j} \right) \times \text{Family weight}_i \right],$$

where i denotes the family and j denotes the number of businesses so that $j \in \{1, 2\}$ if $t > 2007$ and $j \in \{1, 2, 3\}$ if $t \leq 2007$. We use X3119, X3219, and X3319 to create a legal status indicator given by Legal status$_{i,j}$, and X3128/10000, X3228/10000, and X3328/10000 to create ownership share for businesses 1, 2, and 3, respectively. Family weight is X42001/5.

In the case of sole proprietorships, to be consistent with the IRS statistics, we assume that a business owner with multiple proprietorships files one return. This means that the number of sole proprietorship returns is calculated as

$$\sum_i \left[\max \left(\text{Sole prop}_{i1}, \text{Sole prop}_{i2}, \text{Sole prop}_{i3} \right) \times \text{Family weight}_i \right],$$

where Sole prop$_{ij} = 1$ if Legal status$_{i,j}$ = Sole proprietorship.

When we have extra conditions (i.e., a business type with a net income or net loss), we will have extra indicator variables incorporated into the above formula. Then, we use X3132, X3232, X3332 to create a net income or net loss type for each business.

- *Business income*: Here is the general structure for calculating the total net income of a legal status type:

$$\sum_i \left[\left(\sum_j \text{Legal status}_{i,j} \times \text{Business income}_{i,j} \times \text{Ownership share}_{i,j} \right) \times \text{Family weight}_i \right]$$

Net income of each business is given by X3132, X3232, X3332.

When we have extra conditions (i.e., a business type with a net income or net loss), we will have extra indicator variables incorporated into the above formula. Then, we use X3132, X3232, X3332 to create a net income or net loss type for each business. Note that we multiply

by ownership share since questions on business income ask for total business income, not share-adjusted income.

- *Number of owners of unincorporated businesses*: We use answers to self-employment questions, together with answers to business ownership questions, to calculate the number of unincorporated business owners in the SCF as follows:

  – For sole proprietorships:

$$\sum_i \Big[\underbrace{\max\left(\text{Sole prop}_{i1}, \text{Sole prop}_{i2}, \text{Sole prop}_{i3}\right)}_{\text{family has a sole prop}} \times \text{Family weight}_i\Big]$$

  – For partnerships:

$$\sum_i \Big[\underbrace{\max\left(\text{Partnership}_{i1}, \text{Partnership}_{i2}, \text{Partnership}_{i3}\right)}_{\text{family has a partnership business}} \times [\text{Self emp. head}_i + \text{Self emp. spouse}_i + \text{Self emp. other member}_i] \times \text{Family weight}_i\Big],$$

    where $\text{Partnership}_{ij} = 1$ if $\text{Legal status}_{i,j} = \text{Partnership}$ and
    $\text{Self emp. head/spouse/other member}_i = 1$ if employment status of the individual is self-employed. Notice that SCF counts at most two self-employed members (respondent and spouse) within the family. If the family has more members that are self-employed, we would miss them. As a result, SCF may understate the number of self-employed. To partially correct for this possibility, we create another indicator called "Self emp. other member indicator" such that if the family owns at least one unincorporated business but both head and spouse are not self-employed, we say that there has to be another person in the PEU who is self-employed. This is a partial solution because more than one family member other than the head and spouse may be self-employed.

  – The number of unincorporated owners is simply the sum of both components above.

- *Schedule E income per return:* Schedule E returns are attached to individual income tax documents (Form 1040). We use the answers to X5713 to count number of Schedule E returns together with X5746 to count the number of income tax returns within the family as

follows:

$$\sum_i \left[\text{Sch. E return}_i \times \text{Number of income tax returns}_i \times \text{Family weight}_i\right],$$

where Sch. E return = 1 if the family files a Schedule E return and the number of income tax returns within the family is calculated using X5744. We also calculate the total Schedule E income as follows:

$$\sum_i \left[\text{Tax return}_i \times \text{Schedule E income}_i \times \text{Family weight}_i\right],$$

where Tax return$_i$ = 1 if the family files an income tax return and Schedule E income$_i$ is the answer to X5714. We then divide total Schedule E income by the total number of Schedule E returns to obtain the Schedule E income per return.

- *Business rate of return (net income yields)*: For each business with positive net worth that the family actively manages, we calculate the business rate of return by the ratio of business income to business net worth. We then calculate the value-weighted and equally weighted means and the distribution of net income yields across business legal status.[2]

- *Adjusted gross income:* Adjusted gross income is the sum of X5702—X5724 from above related to AGI. As we show in the online Appendix, sole proprietorship income from lines 12 and 18 of Form 1040 (answered on X5704) and line 31 of Schedule C (X3132, X3232, and X3332, combined with the response to legal status of the actively managed business with codes X3119, X3219, and X3319) do not align with each other. When calculating AGI, if X5704 is nonzero but the household reported no sole proprietorships (in codes X3119, X3219, and X3319), then we set X5704 to 0. Then, we use this revised X5704 in the AGI calculation.

- *Business receipts:* Here is the general structure for calculating the total receipts of a legal status type:

$$\sum_i \left[\left(\sum_j \text{Legal status}_{i,j} \times \text{Business receipts}_{i,j} \times \text{Ownership share}_{i,j}\right) \times \text{Family weight}_i\right]$$

[2]We exclude businesses with net worth less than the bottom 1st percentile of the net worth distribution, conditional on having positive net worth.

The business receipts of each business is given by X3131, X3231, and X3331.

- *Broad business income:* Broad business income is defined to be income derived from a business or profession (Form 1040, Schedule C) or farm (Form 1040, Schedule F) (X5704); income from rental real estate, royalties, partnerships, S corporations, estates, trusts (Form 1040, Schedule E) (X5714); and income from gains from the sale of capital and other property (Form 1040, lines 13 and 14) (X5712):

$$\sum_i \left[(\text{Schedule C+F income}_i + \text{Schedule E income}_i + \text{Capital income}_i) \times \text{Family weight}_i\right],$$

- *Percentage of respondents checking documents:* If a respondent says that he or she checked the income tax document (X7451=1), we use his or her answers to X6536 to obtain the frequency of checking this document. The respondent did not check the income tax document if either (X7451=5 or X7451=0 or X7451=-7) or (X6536=4). We use the same steps to check the referencing of other financial documents by using X7452–X7455 instead of X7451. We classify a respondent who checks at least one of these four documents as someone who refers to any other tax documents. We then obtain the weighted fraction of the group of respondents who check these two types of documents frequently, sometimes, rarely, or never. Roughly 4 percent of all respondents have nonapplicable responses (NaN). We adjust for this nonresponse rate in the results of the main text so that our fractions sum to 100 percent. Using these indicators, we calculate the distributions of nonreferencing to relevant documents in the SCF by AGI, business income, and business profit or loss, using family weights.

2.3 IRS data

- *Number of business tax returns; business income*; *number of business owners; Schedule E income*; *business receipts*; *broad business income*:
  - Sole proprietorships: SOI Table 2: Nonfarm Sole Proprietorships: Income Statements, by Industrial Sectors Table 1.4: All Returns: Sources of Income, Adjustments, and Tax Items, by Size of Adjusted Gross Income
  - Partnerships: SOI Table 1: All Partnerships: Total Assets, Trade or Business Income

and Deductions, Portfolio Income, Rental Income, and Total Net Income, by Selected Industrial Group

  - S corporations: SOI Table 7: S Corporation Returns: Balance Sheet and Income Statement Items, by Major Industry
  - C corporations: SOI Table 16: Balance Sheet, Income Statement, Tax, and Selected Other Items, by Major Industry

- *Adjusted gross income:* SOI Table 1.4: All Returns: Sources of Income, Adjustments, and Tax Items, by Size of Adjusted Gross Income

Adjustments

- Sole proprietorships: Our IRS lines for sole proprietorship income, number of returns, and income per return are generated using reported IRS data from respective tables. For the adjusted IRS income for sole proprietorships, we use BEA estimates of income misreporting by noncorporate businesses over time, which are available in NIPA Table 7.14 (line 2). We first calculate the ratio of this misreported unincorporated business income to IRS unincorporated business income. We then obtain the adjustment amount of sole proprietorship business income for the IRS by multiplying this ratio with the reported IRS sole proprietorship income. The adjusted IRS sole proprietorship income is the sum of this adjustment and reported IRS income. Then, the adjusted IRS sole proprietorship income per return is the ratio of the adjusted IRS sole proprietorship income to the IRS number of sole proprietorship returns.

- Partnerships:

  - *Business income*: We make two types of adjustments to the IRS partnership income. In the first step, we obtain the partnership income generated by individual partners so that it becomes more comparable with survey data, given that survey data collect information on individual business owners only. To do so, we use an estimate from Cooper et al. (2016), who use administrative US tax data and find that individual partners generate 31.5 percent of all partnership income. Hence, we obtain our IRS partnership income by multiplying the reported IRS partnership income by 0.315. Then,

our adjusted IRS income for partnerships is obtained using the same procedure for adjusting sole proprietorship income using BEA estimates of income misreporting as above. This step gives us misreporting adjusted IRS partnership income, which is our IRS-adjusted partnership income line.

  - *Number of returns*: In the survey data, as we discussed above, we use ownership shares when calculating the number of business returns for each legal status. For comparison, we obtain the number of partnership returns filed by individual partners (i.e., adjusted IRS number of partnership returns, which is used in Figure A.2) by multiplying the IRS number of returns with 0.315.
  - *Business income per return*: The adjusted IRS partnership income per return is the ratio of the adjusted IRS partnership income to the adjusted IRS number of partnership returns.
  - *Number of partners*: When calculating the total number of unincorporated business owners (which is used in our Figure 3B in the main text and online Appendix Figure A.3), we use an adjusted number of partners for the IRS. The IRS reports the total number of partners over time. We adjust this number so that we only count the individual partners. To do so, we use another estimate from Cooper et al. (2016), who find that individual partners constitute 73.9 percent of all partners. Hence, our adjusted IRS number of partners is the reported IRS number of partners multiplied by 0.739.

- S corporations: Our IRS lines for S-corporation income, number of returns, and income per return are generated using reported IRS data from respective tables. We adjust S-corporation business income in the IRS using the estimates from Johns and Slemrod (2010) and the Government Accountability Office (GAO) for S corporations based on tax audit data. These estimates are taken from reports GAO 14-453 and 10-195, which summarize the progress of the tax compliance studies conducted by the IRS through the National Research Program. According to these estimates, the misreported income is 18 percent of the true income. We obtain our adjusted IRS S-corporation income using this number. Then, the adjusted IRS S-corporation income per return is the ratio of the adjusted IRS S-corporation income to the IRS number of S-corporation returns.

- C corporations: C-corporation income and number of returns reported in the IRS data cover both publicly held and privately held C corporations. Given that the survey data only contain information on privately held C corporations, we adjust the IRS data so that both sets of data are comparable. To do so, we use IRS tables to obtain total returns and net income of publicly held C corporations using the SOI reports on M3 information for 10-K filers.[3] Subtracting these values from the total number of returns and the total income of reported IRS data gives us the total number of returns and the total income of privately held C corporations.[4] This process gives us our IRS lines for income, number of returns, and income per return for C corporations. Next, we also adjust the income misreporting for privately held C corporations using data from NIPA Table 7.16 (line 2), which provide the total corporate income misreporting amount. We subract misreported income for S corporations (discussed above) from misreported income for all corporate businesses. This gives us the total misreporting amount for income from C corporations. Using the average ratio of the income of publicly held C corporations to the income of all C corporations (which is 75.6 percent), we attribute 75.6 percent of misreporting amount to publicly held C corporations and the rest to privately held C corporations. Then, adding the misreported income of privately held C corporations to their income gives us the total adjusted income of privately held C corporations, which yields our adjusted IRS income line for privately held C corporations. We obtain our adjusted IRS income per return for privately held C corporations by dividing the adjusted IRS income by the number of returns for privately held C corporations.

When it is possible to make these adjustments, we compare the survey data with the adjusted IRS data. As a result, in the main text, we compare the business income (Figure 1 in the main text), business income per return (Figure 2 in the main text), and unincorporated business income and business income per owner (Figure 3 in the main text) in the SCF to our adjusted IRS data (in addition to our IRS data). For all the results in the online Appendix, we only present results

[3]For example, for year 2012, we use information on Table 1B of the document from this link: `https://www.irs.gov/pub/irs-utl/2012%20M-3%20First%20Look%20and%20MNE%20Type--Boynton-DeFilippes-Legel--Tax%20Note%20%2812-21-2015%29.pdf`. Total number of returns and net income (taxable net income reported by SOI) for public firms are available in Table 1B.

[4]These tables are only available in 2004—2010 and 2011 and 2013. For missing years, we calculate the average ratio of total public returns to total returns from available years excluding 2008 and 2009 (and same for income) and use this ratio to obtain the returns and incomes of publicly held C corporations over time. When calculating this average ratio, we exclude 2008 and 2009 because during the Great Recession, privately held C corporations experienced a negative total income in aggregate, which was an extraordinary outcome.

for our IRS data (where partnership moments are adjusted to incorporate partnerships owned by individual partners and C corporations are adjusted to incorporate privately held C corporations).

3 Additional Results

Next, we report our additional findings on business incomes, number of returns, business receipts, and business rates of return.

3.1 Adjusted gross income

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

However, our focus is on measures that relate to business activity. Of particular relevance are the findings in Kuhn and Rios-Rull (2016) that business income is one of the main contributors to income inequality and that business equity is one of the main contributors to wealth inequality, which they document for the history of the SCF surveys between 1989 and 2013. Bricker et al. (2016) also use the SCF to document the rise in the top share of wealth holdings over time and find that the share of wealth attributable to the top 1 percent rose from 30 percent in 1992 to 36 percent in 2013. Our paper exploits the fact that SCF answers can be compared with administrative data from the IRS and finds that respondents are not reliably or consistently answering questions about their business income or business equity, and therefore we cannot trust the SCF distributions.

3.2 Number of business returns and owners

In Figure 1 of the main text, we show that the SCF overstates the aggregate business income of pass-through businesses by an average of 123 percent relative to the IRS data and 54 percent relative to the adjusted IRS data. Similarly, in Figure 2, we show that the SCF overstates the business income per return for pass-through businesses by an average of 399 percent relative to the IRS data and 179 percent relative to the adjusted IRS data. These results imply that the SCF understates the number

of business returns. To provide more details on this comparison, Figure A.2 plots the number of business returns by legal status over time. We find that the SCF underestimates the number of returns by an average of 65 percent for sole proprietorships, 37 percent for S corporations, and 53 percent for pass-through businesses, but overstates it by an average of 238 percent for partnerships after adjusting the IRS data.

In Figure 3 of the main text, we show the total unincorporated business income and business income per owner for the IRS, SCF, SIPP, CPS, and PSID. Now, we also show the total number of unincorporated business owners for these datasets in Figure A.3. We find that the number of unincorporated business owners is much larger in the IRS data than in the survey data. Moreover, there is a clear positive trend in the IRS, whereas we do not observe any rise in the number of owners in any survey data.

3.3 Business income

3.3.1 Evidence on mismeasurement

In Section 2.2.1 (Figure 3) of the main text, we used a Venn diagram to split sole proprietorship income and counts for the year 2015 into several categories: (i) those who have a nonzero line 12 plus 18 on Form 1040, (ii) those who are actively managing a business and report line 31 of Schedule C, and (iii) those reporting to have an interest in a business without having an active management role. In Table A.1, we provide details for all survey years.

Section 2.2.1 of the main text discusses possible reasons for the overstatement of business income in survey data. A reason to be suspicious about misreported incomes in the SCF is that a very small fraction of respondents refer to their tax documents when responding to questions about the specific line items on tax forms. To verify whether respondents in the SCF check documents, we use variable X6536, which provides information on the frequency of checking any documents when answering interview questions. Variable X7451 informs us about whether the respondent referred to income tax documents, and variables X7452 through X7455 inform us about whether the respondent referred to other financial documents, namely, pension documents, account statements, investment or business records, and loan documents, respectively. If a respondent says that he or she checked the income tax document (X7451=1), we use his or her answers to X6536 to obtain the frequency of

checking this document. The respondent did not check the income tax document if either (X7451=5 or X7451=0 or X7451=-7) or (X6536=4). We use the same steps to check the referencing of other financial documents by using X7452–X7455 instead of X7451. We classify a respondent who checks at least one of these four documents as someone who refers to any other tax documents. We then obtain the weighted fraction of the group of respondents who check these two types of documents frequently, sometimes, rarely, or never. Roughly 4 percent of all respondents have nonapplicable responses (NaN). We adjust for this nonresponse rate in the results of the main text so that our fractions sum to 100 percent.

We calculate the frequency with which business owners referenced either tax or other financial documents in tax year 2015.[5] These tabulations are shown in Table A.2. The first row shows that 75 percent of business owners in the SCF never referenced tax documents, 2 percent rarely did, 9 percent sometimes did, and 14 percent frequently did. The second row shows that 64 percent never referenced any other financial documents, 6 percent rarely did, 15 percent sometimes did, and 15 percent frequently did.

In the main text, we assert that non referencing of tax documents is uniform across business owners. To show this, we use tax year 2015 and group owners by their AGI and by total business income. In all cases, we find the fraction of owners who never reference a tax document to be very high(about 75 percent) and not varying too much across groups. The information is summarized in Table A.3.

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

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

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

profits listed on Form 1040 are farm income, and farm losses exceed profits in many of the years of our sample.

3.3.2 Business income distribution

In Section 2.2.2 of the main text, we discussed the underrepresentation of business whose owners have little income. In Figure A.5, we rank sole proprietors in the SCF by their AGI, assign them to income brackets using the same bins as the IRS, and plot the fractions of business income for owners with below-median AGI and for those with AGIs in the top 1st percentile. For most years, the SCF income shares for these two groups are understated and display large year-by-year variation. For example, the share for those with below-median AGI is nearly doubled or halved from one survey to the next. Since the fractions sum to 100 percent across all AGI groups, the SCF must necessarily overstate incomes for some bins. We find the largest overstatement of shares for those with AGIs between the 50th and 75th percentiles. In Figure A.6, we see that the overstatement of business income per return in the SCF data also varies a lot across years and across AGI bins, with no systematic pattern. In contrast, the incomes per return in the IRS data show little variation over time and vary similarly across AGI bins. Figure A.7 shows the number of sole proprietorship returns with AGIs per return below and above the median. For businesses that have owners with below-median AGIs, the number of IRS returns has risen from about 5 million in 1988 to over 12 million in 2015, but the SCF estimate has remained at roughly 2 million for the entire period. For businesses with above-median AGIs, the number of IRS returns has risen from a little over 8 million to above 12 million, but the SCF estimate has hovered around 5 million. Finally, In Figure A.8, we report the distributional statistics for S corporations. As we noted in the main text, the data for S corporations are only available for limited years, namely 2003–2012, but these data show similar inconsistencies between SCF and IRS data, as was found with sole proprietorships.

In Section 2.2.3, we discussed the distribution of business income by splitting pass-through businesses into two categories: those that make profits and those that incur losses (or make no income). In Figures A.9 and A.10, we plot business income per return by legal status for those making profits and incurring losses, respectively. In Figures A.11 and A.12, we plot the number of returns for the same sets of businesses. In Table A.4, we extend the analysis of decomposing the total percentage error into the overstatement of profits and understatement of losses. Table

A.5 shows the distribution of losses by AGI bins for tax year 2015. We see that 10 out of 19 bins, which account for 23 percent of the total number of returns and 26 percent of the total losses in the IRS, have an aggregate zero (that is, all respondents in those income brackets reported a zero net income) in the SCF data. All of these results show that the distribution of business income in the SCF is largely inconsistent with its counterpart in the IRS and that the inconsistencies vary across survey years.

3.3.3 Misclassification of business income

In Section 2.2.3 of the main text, we discuss that another source of measurement error in the SCF may be due to the respondent's possible confusion about closely related categories of business income. We then argue that even when we compare (i) Schedule E income, (ii) Schedule C, E, and F income, or (iii) broad business income (by adding capital gains which includes nonbusiness income), in the IRS and the SCF, we still find an overstatement of business income in the SCF. Here, we provide these results.

Figure A.13 shows a comparison between the SCF and the IRS for total Schedule E income and Schedule E income per return. We find that the SCF income per return is overstated relative to the IRS, especially in recent years. Figure A.14 shows that the SCF also overstates Schedule C, E, and F income and income per tax return. Finally, in Figure A.15, 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 income), and we still find an overstatement.

3.3.4 KFS data

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

estimates from their study in Table A.6.

3.4 Business receipts

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

3.5 Business valuations and rates of return

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

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

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

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

In the main text, we showed evidence that the SCF income yields are high when compared with CRSP-Compustat or Pratt's Stats. In Table A.7, we provide several additional moments for the distribution of income yields in the SCF. The additional moments show that SCF income yields are high regardless of year or legal structure.

In the main text, we compared the income yields for S and C corporations in the SCF with small

firms in CRSP where we defined "small" as corporations that are in the bottom quintile of the size distribution as measured by the book value of total assets. In Table A.8, we extend the analysis to two alternative definitions of "small": (i) those in the bottom quintile by market value and (ii) those in the bottom quintile by gross sales. Although there are some differences in the magnitudes compared with Table 1 in the main text, the equally weighted and value-weighted yields are still negative in all years, regardless of how we classify the small firms.

Income yields for all businesses as well as non-tech and non-distressed firms obtained from Pratt's Stats were discussed in the main text. We extend this discussion with Table A.9, which reports income yields from Pratt's Stats for all legal forms. We see that sole proprietors have higher yields than other pass-throughs and C corporations. However, since these businesses have much smaller valuations, the value-weighted yield for all businesses is relatively low when compared with SCF data.

In Tables A.10 and A.11, we report the income yields in PSID and SIPP for all years that the data are available for unincorporated businesses. For the PSID, we obtain business net worth using a question that asks the sale value (net of debts) on the business (for example, for 2015, the code for the variable is ER61736). Business income is obtained from the households' income from unincorporated businesses (for example, for 2015, the code for the variable is ER65192). Similarly, for the SIPP, we obtain business net worth from questions asking the value of the business assets (coded TVBVA1 TVBVA2) and liabilities (coded TVBDE1 and TVBDE2) separately. Business income is obtained from the households' share of net business income from each of the two businesses (coded TPRFTB1 and TPRFTB2). We see that value-weighted income yields in the PSID and SIPP are comparable to the SCF, and all yields are much higher than those from Pratt's. The survey estimates are comparable even though income per owner is much lower in the PSID and SIPP than in the SCF. This implies that average business values are even lower in these other surveys. However, if we compare yields across the distribution, we see large differences across surveys, especially in the right tail. These observations point to a lack of representativeness in the PSID and SIPP for the universe of noncorporate businesses as well as their lack of comparability to the SCF.

Finally, in the main text, we compare our result that income yields in survey data are overstated to the result in Moskowitz and Vissing-Jorgensen (2002), who conclude, using SCF data, that pri-

vate business returns were surprisingly low. We show that the difference in our results is explained by Moskowitz and Vissing-Jorgensen's (2002) imputation method used to calculate capital gains. Below, we provide more details on how we reached this conclusion.

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

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

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

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

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

As discussed in the main text, we replicate the exercise of Moskowitz and Vissing-Jorgensen (2002) for our full sample with income yields and capital gains compared separately. We find that the capital gain imputation drives the differences between our findings and theirs. The full results that support the discussion in Section 3.1 of the main text are summarized in Table A.12. The first two columns show estimates of SCF and CRSP-Compustat income yields, R_t^{vw}, in all SCF survey years. The last three columns show estimates of $\tilde{R}_t^{cg}$ for SCF and both R_t^{cg} and $\tilde{R}_t^{cg}$ for the CRSP-Compustat sample.

The table reveals two important results. First, SCF yields are substantially higher than the CRSP-Compustat counterparts for all survey years. Second, the annualized SCF capital gains vary substantially less than those for firms in the CRSP-Compustat gains R_t^{cg} over the sample, which is not surprising given the conceptual differences in the measures and the long interval between

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

survey years.[8] If we were to add R_t^{vw} plus $\tilde{R}_t^{cg}$ for SCF and R_t^{vw} plus R_t^{cg} for CRSP-Compustat firms, we would confound the two discrepancies and conclude that the private and public returns are not very different on average: 26 percent for SCF versus 21 percent for CRSP-Compustat.

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

A Tables

Table A.1: Sole Proprietorships in the SCF

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

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

Table A.2: Percentage of Respondents Checking Documents in SCF 2016

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

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

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

Table A.3: Distribution of Nonreferencing in the SCF

Groups	Fraction of nonreferencing owners
By AGI	
<p25	0.77
p25-p50	0.79
p50-p75	0.75
>p75	0.71
By Business income	
< p25	0.72
p25-p50	0.72
p50-p75	0.80
> p75	0.77
Nonpositive	0.70
Positive	0.76

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

Table A.4: SCF-IRS Business Income Gap by Legal Structure

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

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

Table A.5: Sole Proprietorships with Net Losses in the IRS and SCF by AGI Bins, 2015

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

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

Table A.6: Comparison of KFS and IRS Business Tax Data, 2004–2011

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

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

Table A.7: Net Income Yields in the SCF by Legal Structure

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

Note: This table shows moments of the net income yield distribution of businesses in the SCF by legal structure. The sample includes businesses with positive net worth and excludes the bottom 1st percentile of these businesses. The business income of each business that the family members own in the SCF is obtained from SCF variables that correspond to information on business tax forms.

Table A.8: Income Yield for Small Firms in CRSP

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

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

Table A.9: Income Yield from Pratt's Stats

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

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

Table A.10: Net Income Yields of Unincorporated Businesses in the PSID

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

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

Table A.11: Net Income Yields in the SIPP

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

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

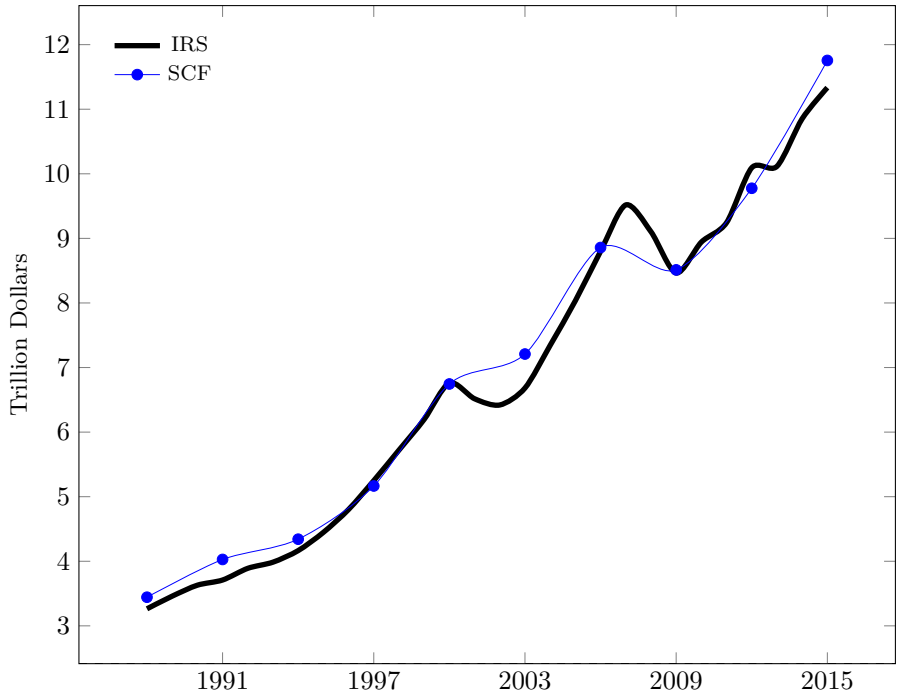
Table A.12: Net Income Yields and Capital Gains

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

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

B Figures

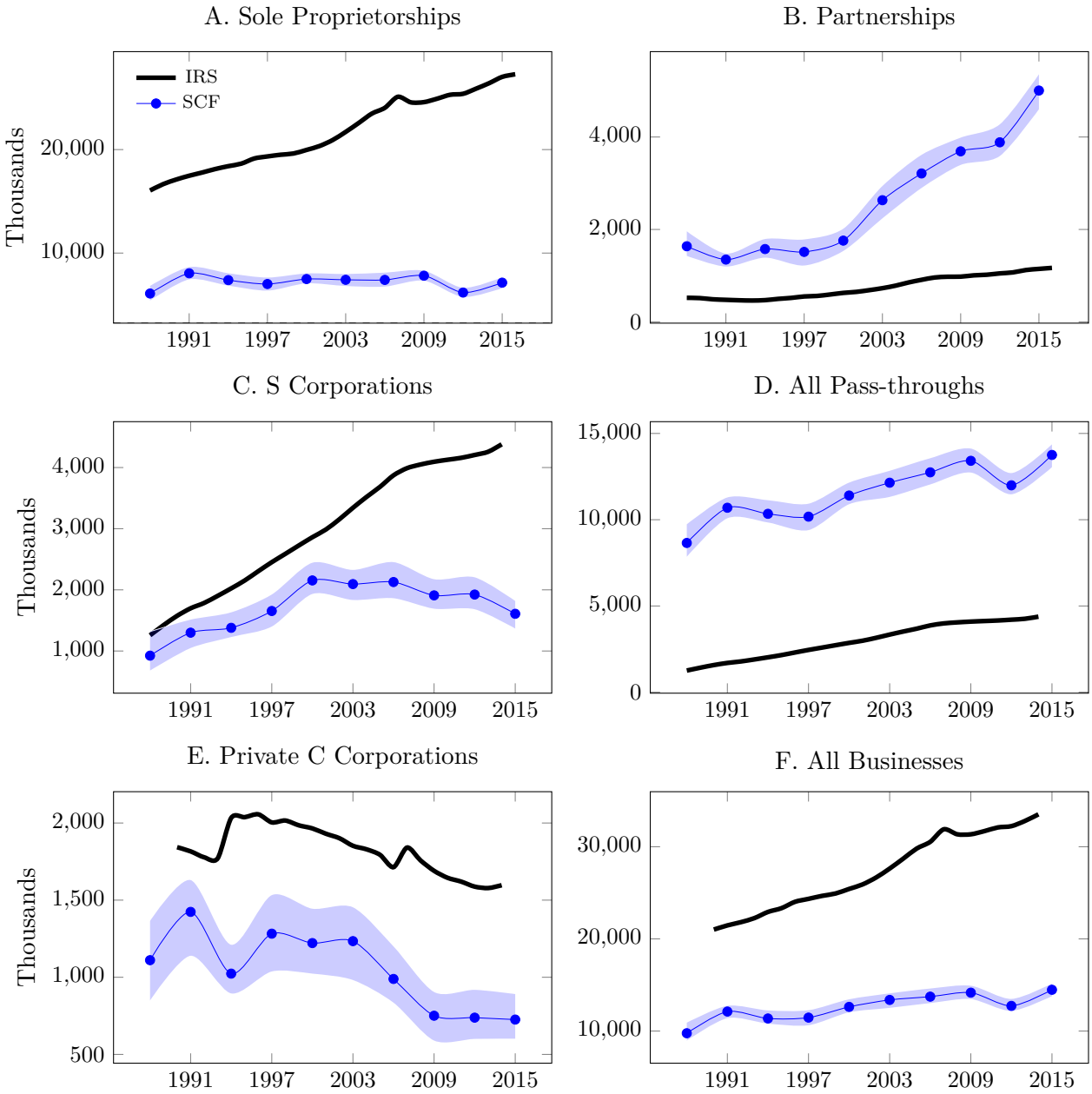
Figure A.1:
Adjusted Gross Income, SCF vs. IRS



Note: For the IRS, adjusted gross income (AGI) is obtained from Form 1040. Please refer to the main text of this Appendix for the definition of AGI.

Figure A.2:

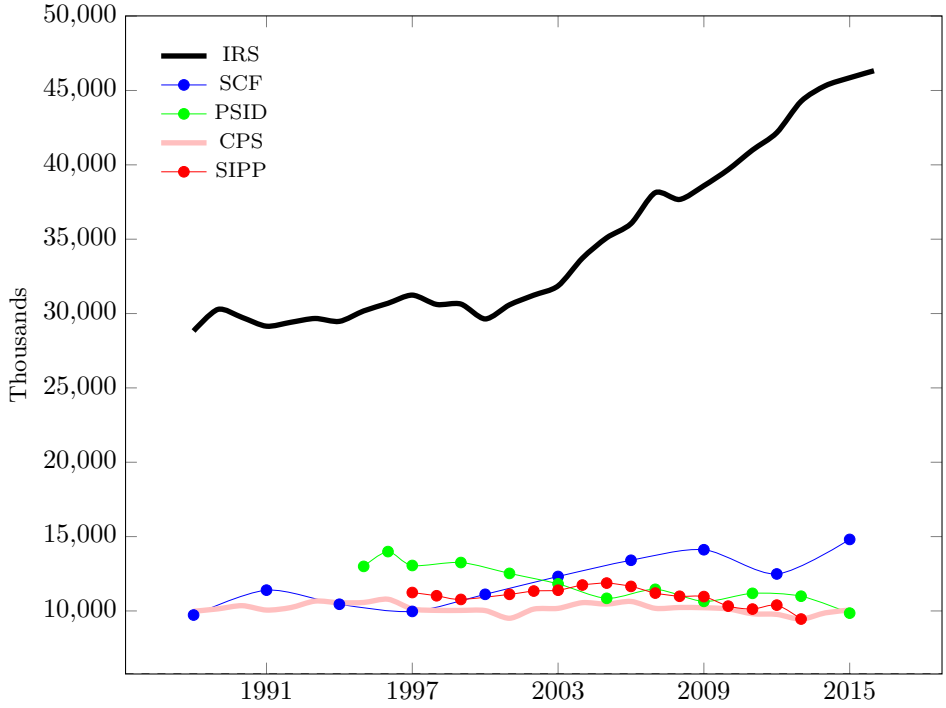
Number of Returns by Legal Entity, SCF vs. IRS



Note: This figure plots the number of business returns of sole proprietorships, S corporations, partnerships, pass-throughs, privately held C corporations, and all businesses over time in the IRS and the SCF.

Figure A.3:

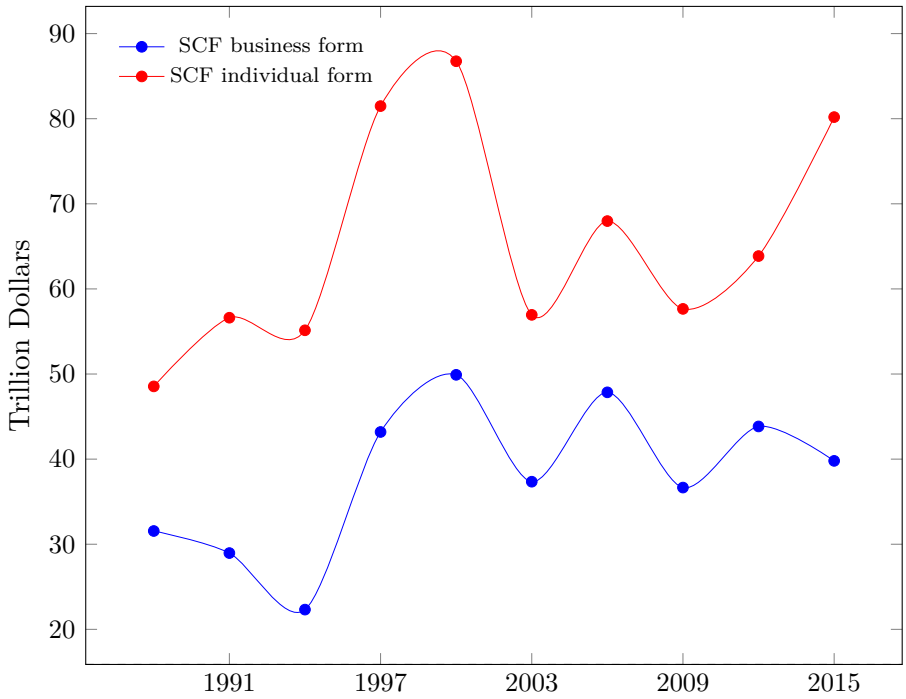
Number of Owners of Unincorporated Businesses



Note: This figure plots the total number of unincorporated business owners in the SCF, CPS, PSID, SIPP, and IRS.

Figure A.4:

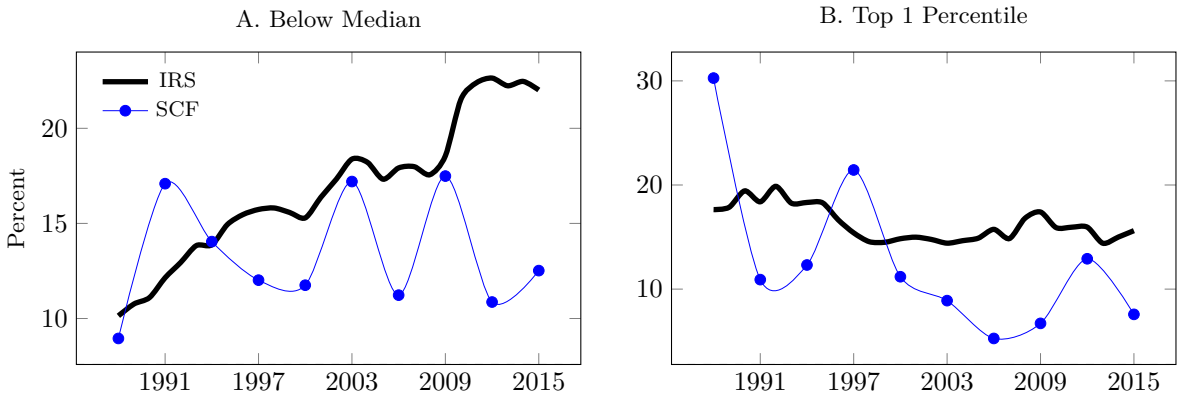
Comparing Proprietors' Individual and Business Incomes in the SCF



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

Figure A.5:

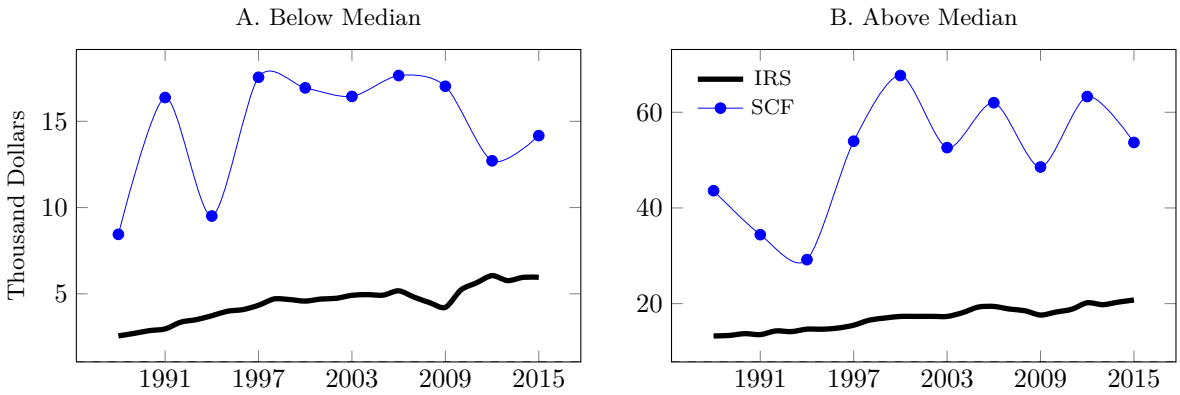
Proprietor Income Shares, SCF vs. IRS



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

Figure A.6:

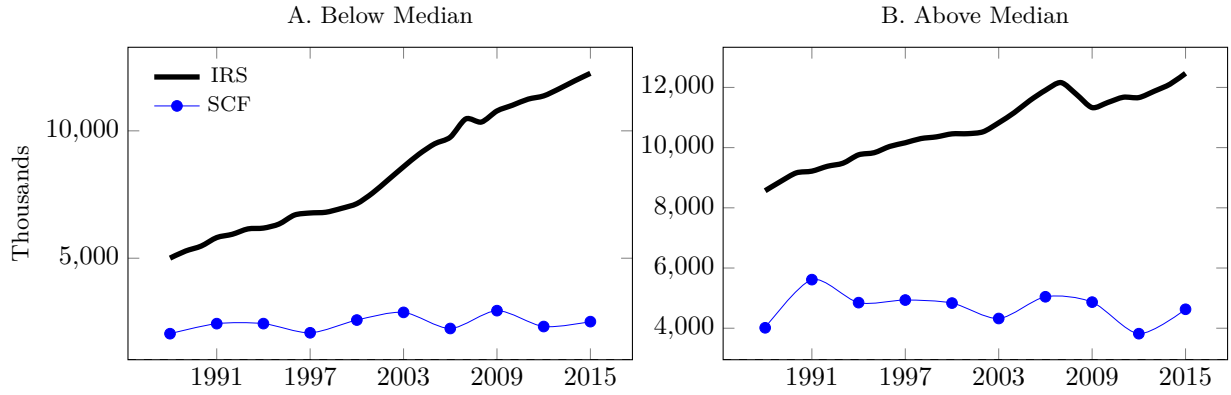
Income Per Return, Proprietors with Below- and Above-Median AGI, SCF vs. IRS



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

Figure A.7:

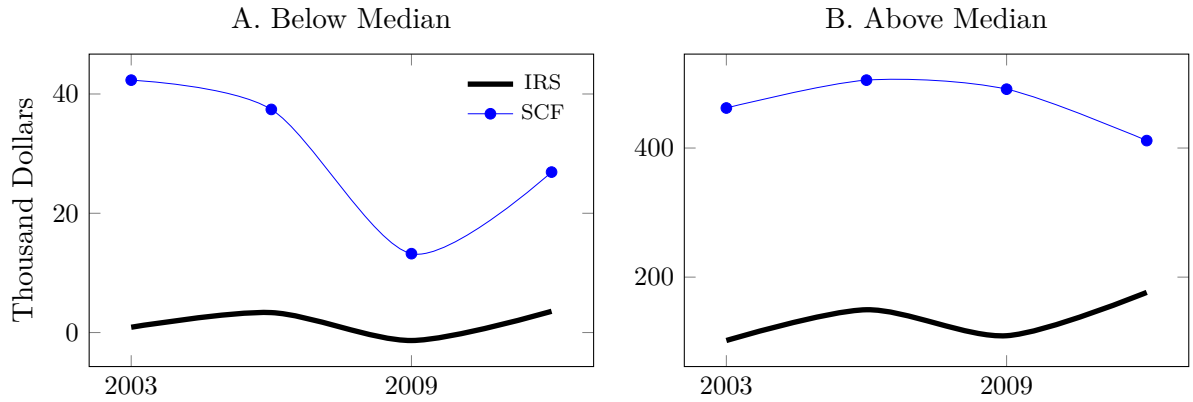
Number of Returns of Proprietors with Below- and Above-Median AGI, SCF vs. IRS



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

Figure A.8:

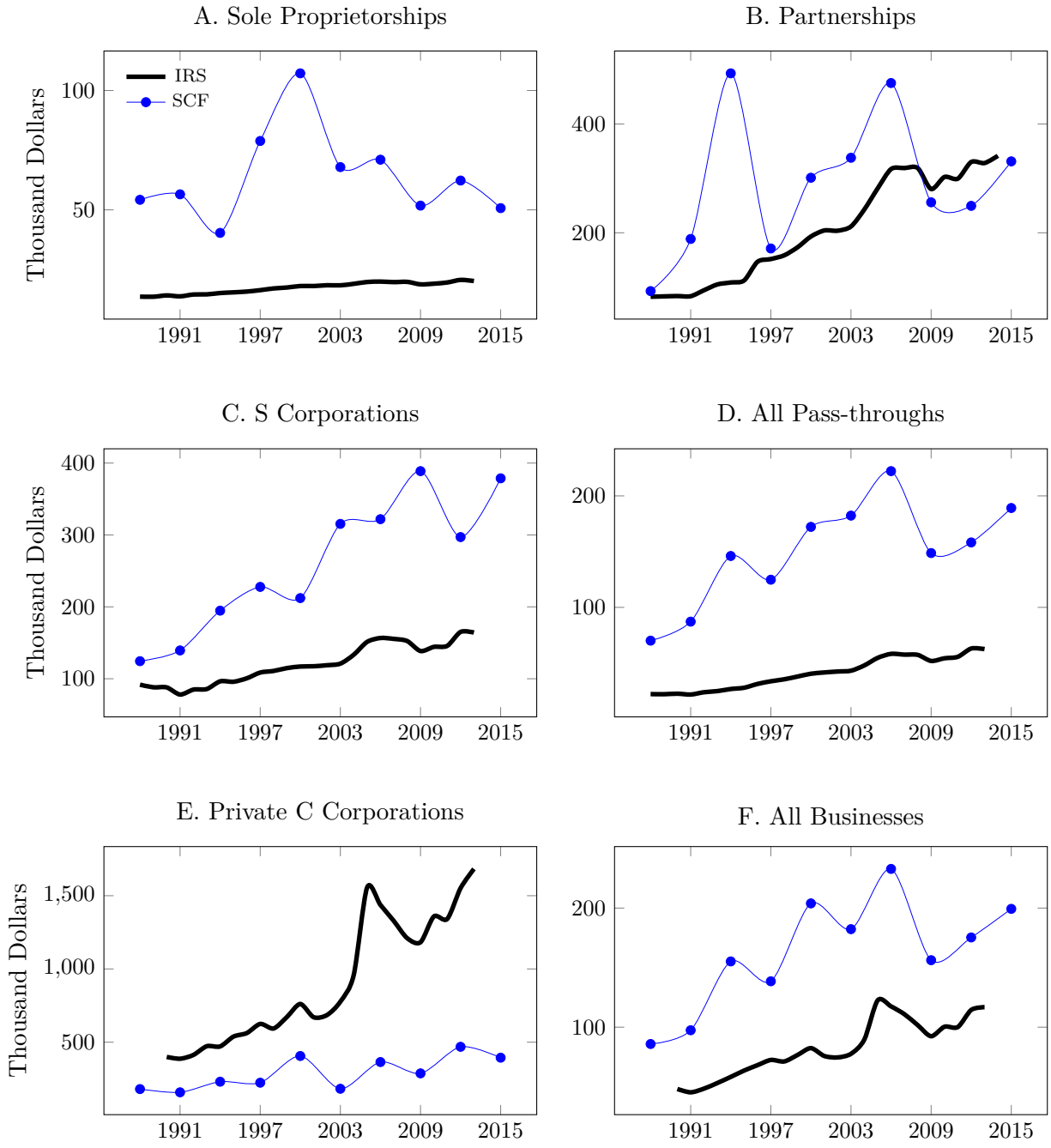
Distribution of S-Corporation Business Income per Return, SCF vs. IRS



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

Figure A.9:

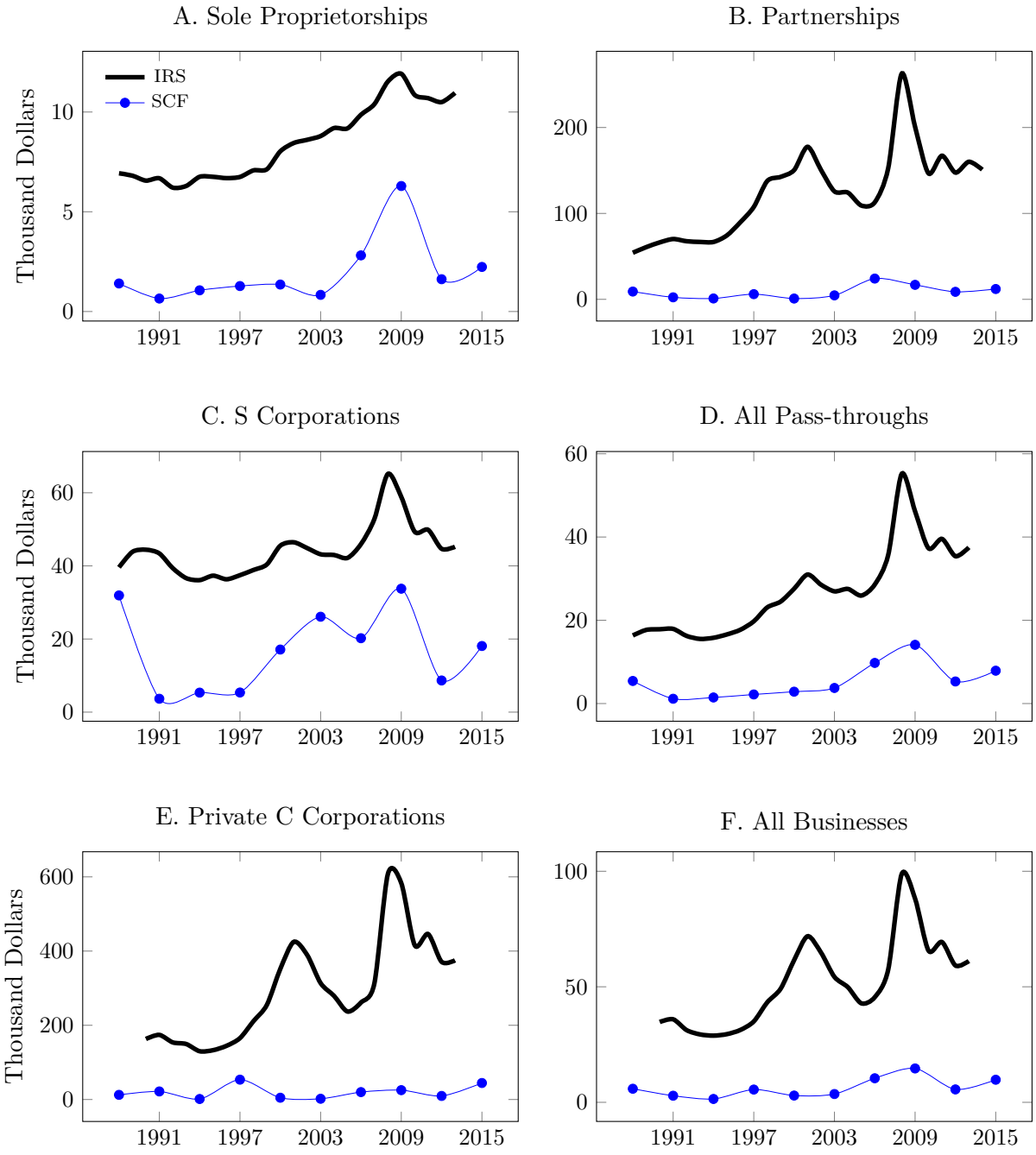
Business Income per Tax Return by Legal Status for Businesses with Net Income, SCF vs. IRS



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

Figure A.10:

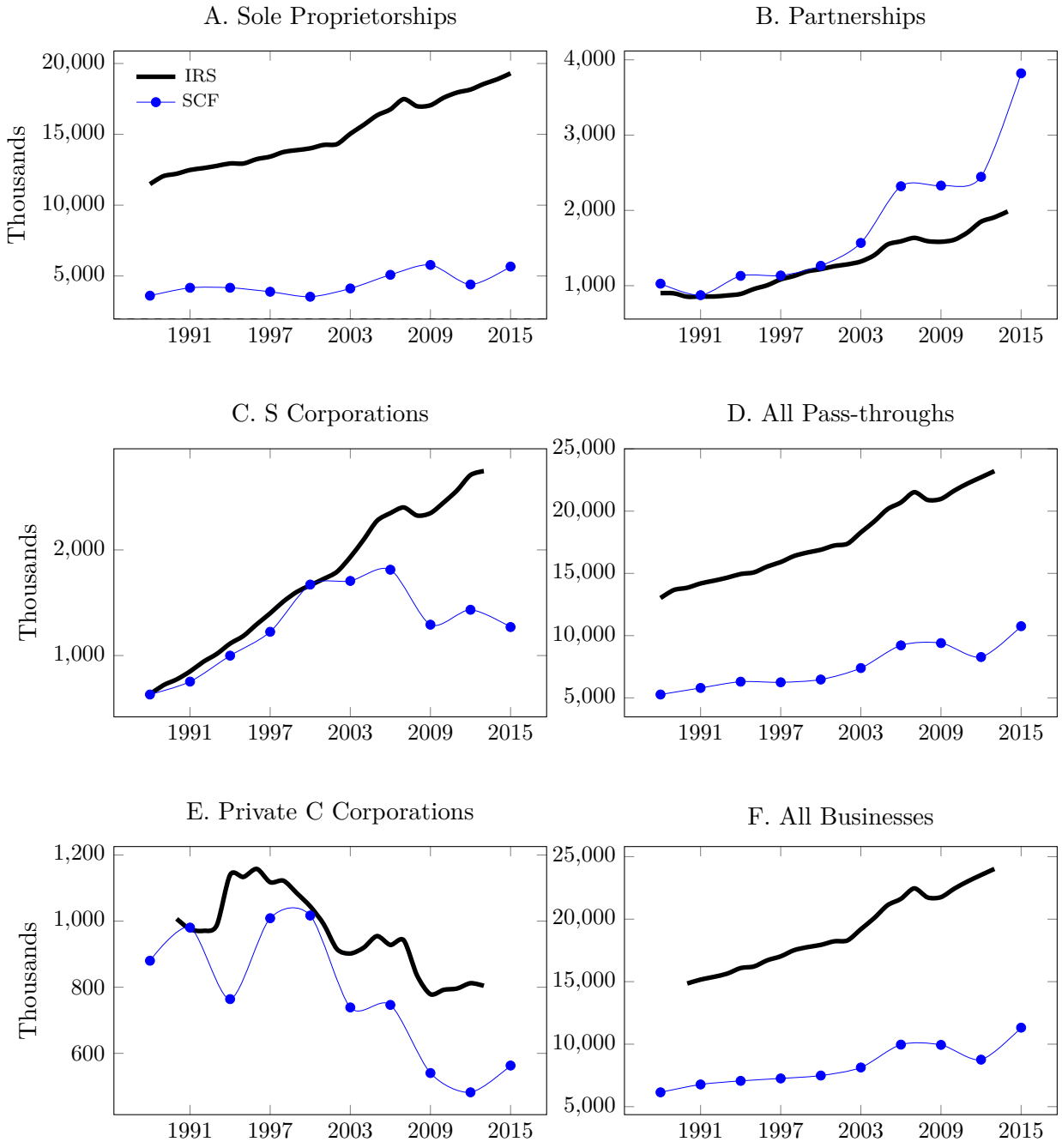
Business Income per Tax Return by Legal Status for Businesses with Net Loss, SCF vs. IRS



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

Figure A.11:

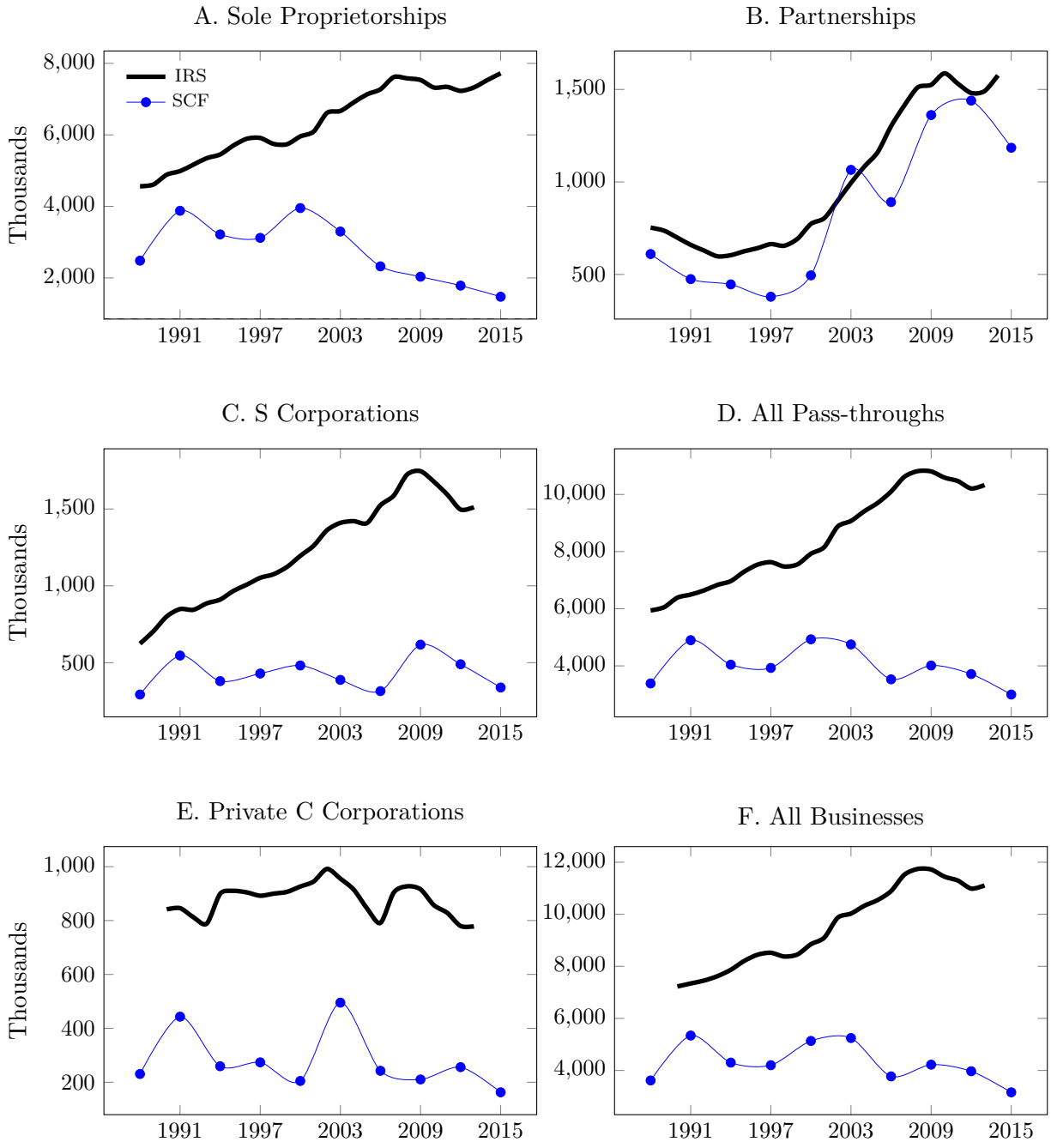
Number of Returns by Legal Status for Businesses with Net Income, SCF vs. IRS



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

Figure A.12:

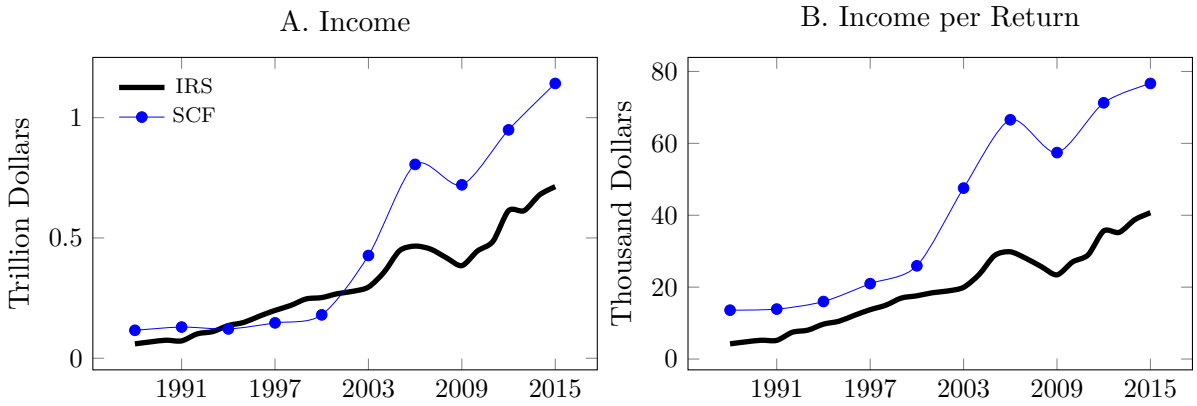
Number of Returns by Legal Status for Businesses with Net Loss, SCF vs. IRS



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

Figure A.13:

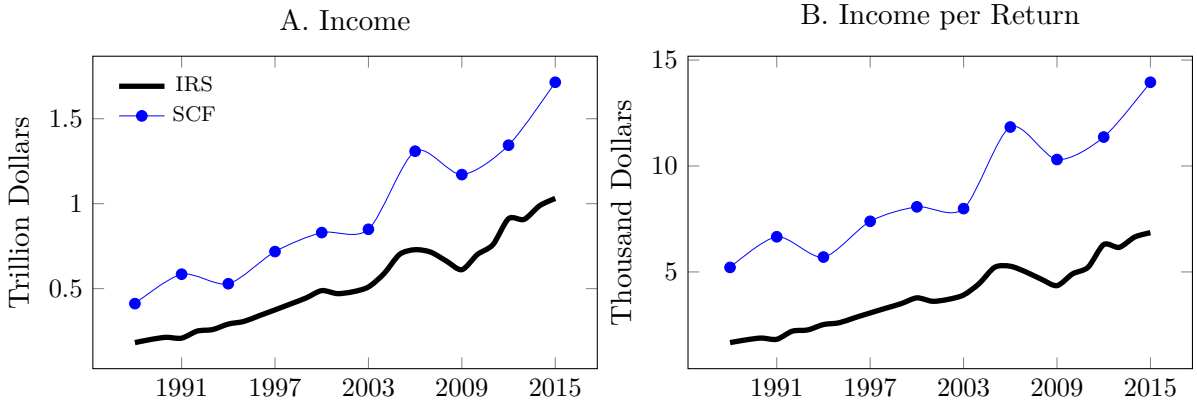
Schedule E Income and Income per Return, SCF vs. IRS



Note: In panels A and B, we respectively plot business income and 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.

Figure A.14:

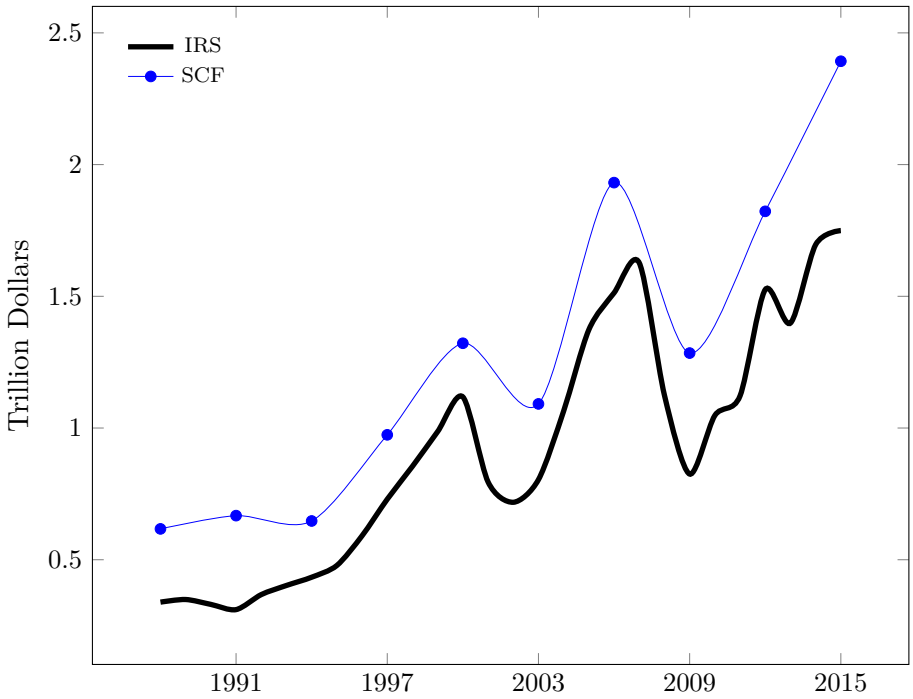
Schedule C, E, and F Income and Income per Return, SCF vs. IRS



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

Figure A.15:

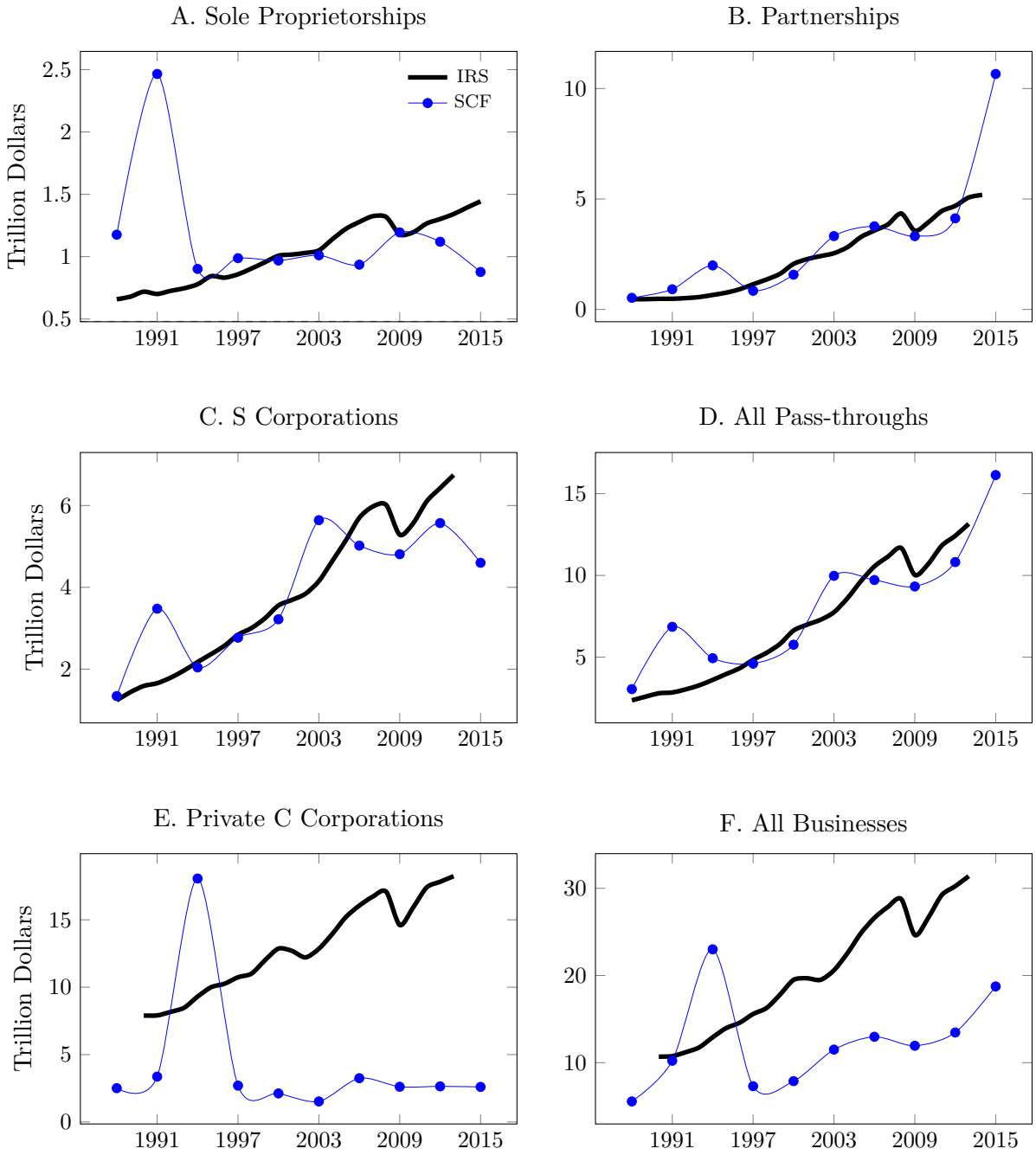
Broad Business Income, SCF vs. IRS



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

Figure A.16:

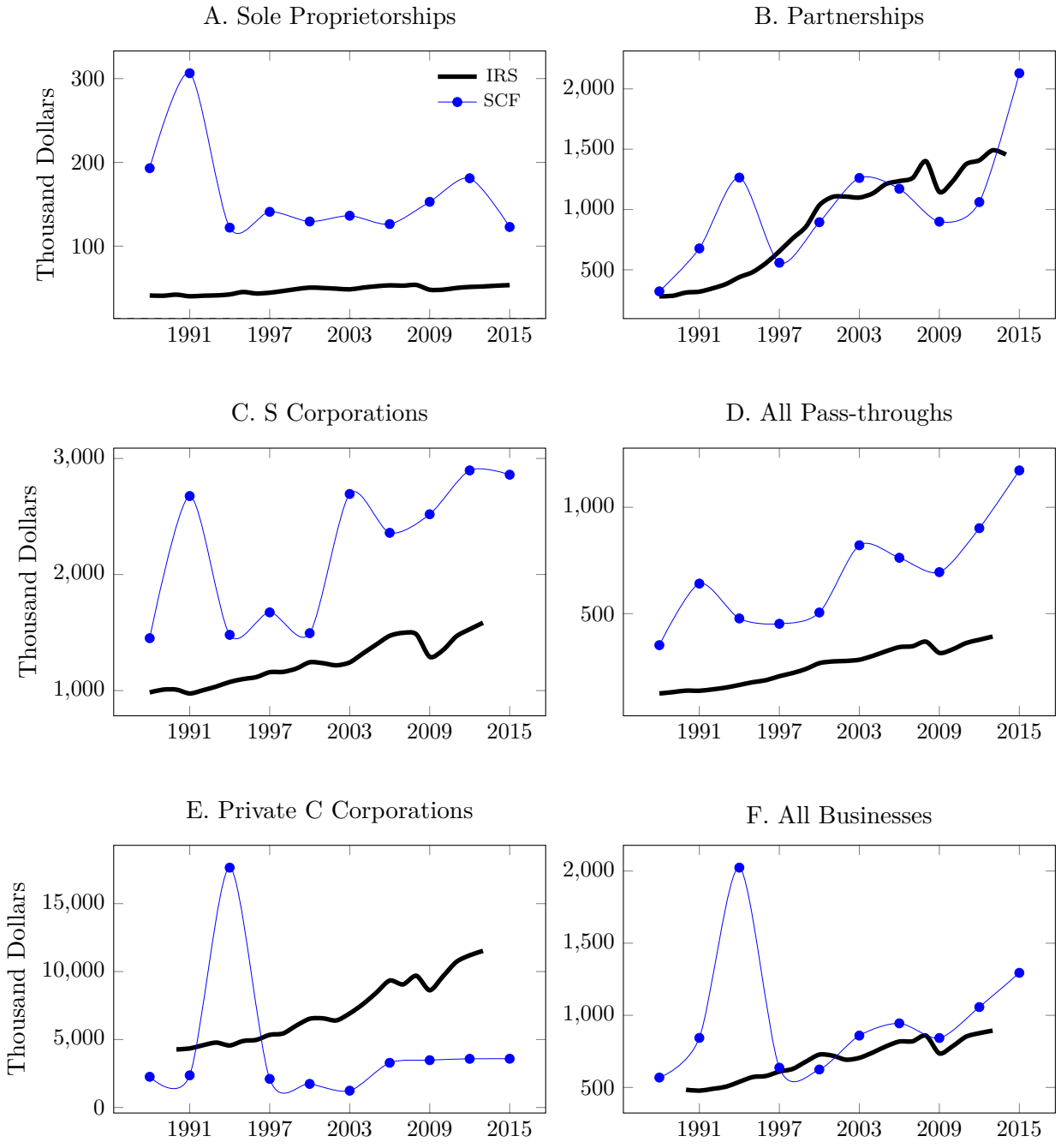
Business Receipts by Legal Status, SCF vs. IRS



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

Figure A.17:

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



Note: This figure plots the business receipts per tax return by legal status in the SCF and IRS. Business receipts refer to gross sales reported on Form 1040, Schedule C, for sole proprietorships, Form 1065 for partnerships, Form 1120S for S corporations, and Form 1120 for C corporations.

References

[1] 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.

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

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

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

[5] 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.

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

[7] 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.

[8] 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.