What Do Survey Data Tell Us about US Businesses?

A. Bhandari, S. Birinci, E. McGrattan, and K. See

EUI Lecture 2, Feburary 2019

Motivation

- Survey data used extensively to study

    - Entrepreneurial choice

    - Wealth inequality

- How reliable are these data?

Motivation

- Survey data used extensively to study

  - Entrepreneurial choice

  - Wealth inequality

- How reliable are these data?

  - The short answer: not very

Main Evidence

- Survey responses:

    ○ Business incomes, receipts, and valuations

    ○ Number of returns and owners

- Compared to:

    ○ IRS incomes, receipts, owners from tax data & audits

    ○ SEC income yields from corporate filings

    ○ Private business valuations from broker data (Pratt's)

Which Surveys?

- Survey of Consumer Finances (SCF)

- Panel Surveys of Income Dynamics (PSID)

- Survey of Income and Program Participation (SIPP)

- Consumer Population Survey (CPS)

Main Findings

- Relative to IRS/SEC/Pratt's counterparts, survey

    ○ Incomes and incomes per owner too high

    ○ Income to value ratios too high

    ○ Numbers of returns and owners too low

- Over- and understatements vary widely

    ○ Across years

    ○ Across surveys

Problems with Measurement

- Sampling issues

  - Nonrepresentative samples

- Measurement issues

  - Documents not referenced

  - Questions framed in confusing ways

Main Take-aways

- Future survey recommendations:

    ○ Link responses to administrative data

    ○ Limit questions to verifiable queries

    ○ Ensure representative samples for all business types

- Current quantitative research:

    ○ Use tax and financial data directly

Aggregate Income

Standard Check of Survey Data

- Make sure aggregates are similar (eg, AGI)

Standard Check of Survey Data

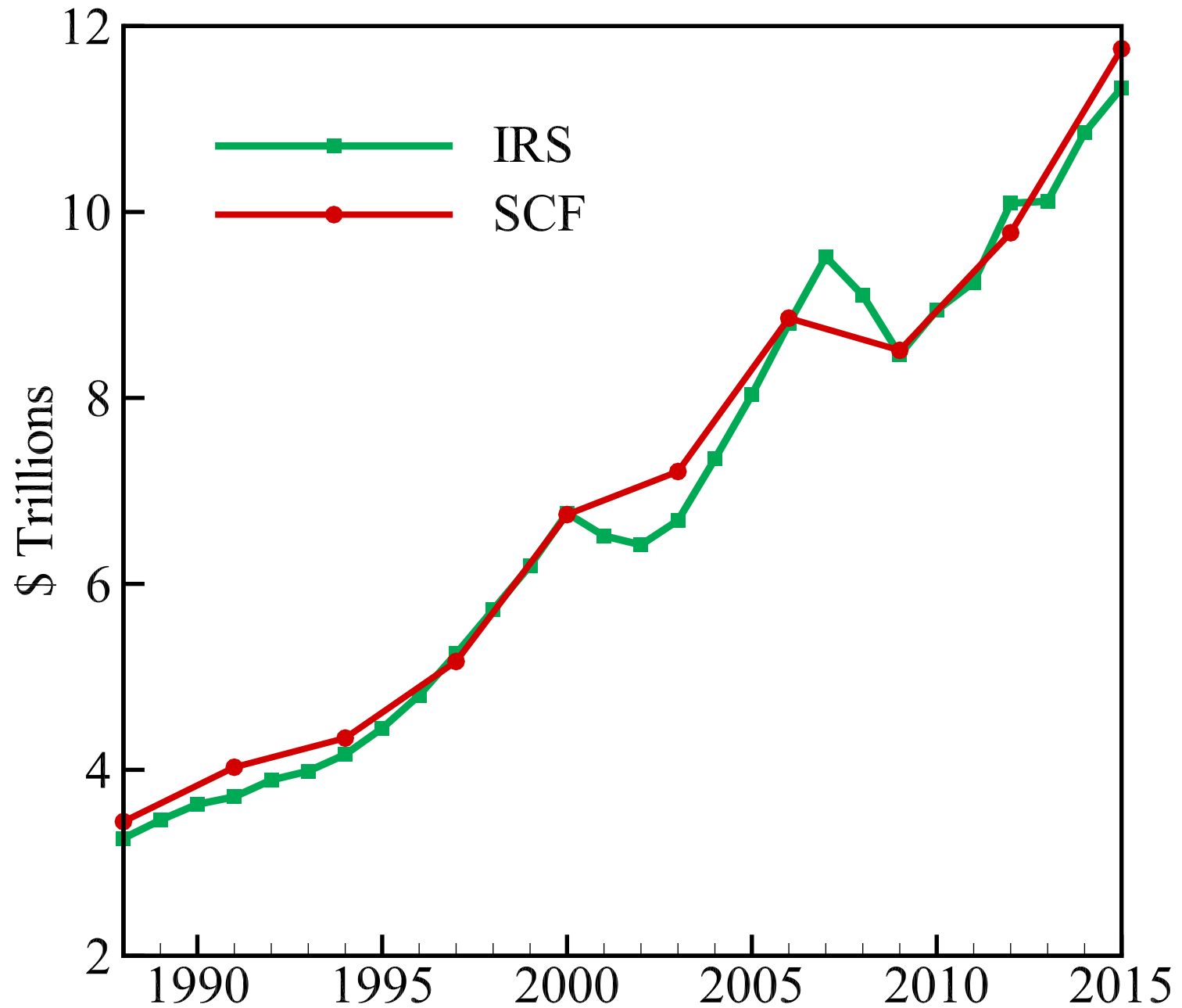- Make sure aggregates are similar (eg, AGI)

- Consider case of SCF

Standard Check of Survey Data

- Make sure aggregates are similar (eg, AGI)

- Consider case of SCF

    - IRS vs SCF AGI are close

AGI: IRS vs SCF

Standard Check of Survey Data

- Make sure aggregates are similar (eg, AGI)

- Consider case of SCF

    - IRS vs SCF AGI are close

    - But, business total or per-return incomes are not

Background

US Business types

- Pass-through entities:
  - Sole proprietor (Form 1040, Schedule C):
    - Unincorporated business with 1 owner

  - Partnership (Form 1065):
    - Unincorporated business with 2 or more owners
    - Owners can be people or corporations

  - S corporation (Form 1120S):
    - Incorporated business with ≤ 100 shareholders
    - Shareholders must be US citizens or resident aliens

- C corporation (Form 1120):
  - Incorporated business without limit on shareholders
  - Business must pay corporate income tax

Growth of Pass-Through Entities

- Net income, 1980:

    - 20% to pass-through entities

    - 80% to C corporations

- Net income, 2015:

    - 50% to pass-through entities

    - 50% to C corporations

$\Rightarrow$ Harder to measure business activity

Private Business Growing

Percent of net income

Sole Proprietors

Partnerships

S Corporations

Schedule C Corporations

Note: C-corps mostly public, others mostly private
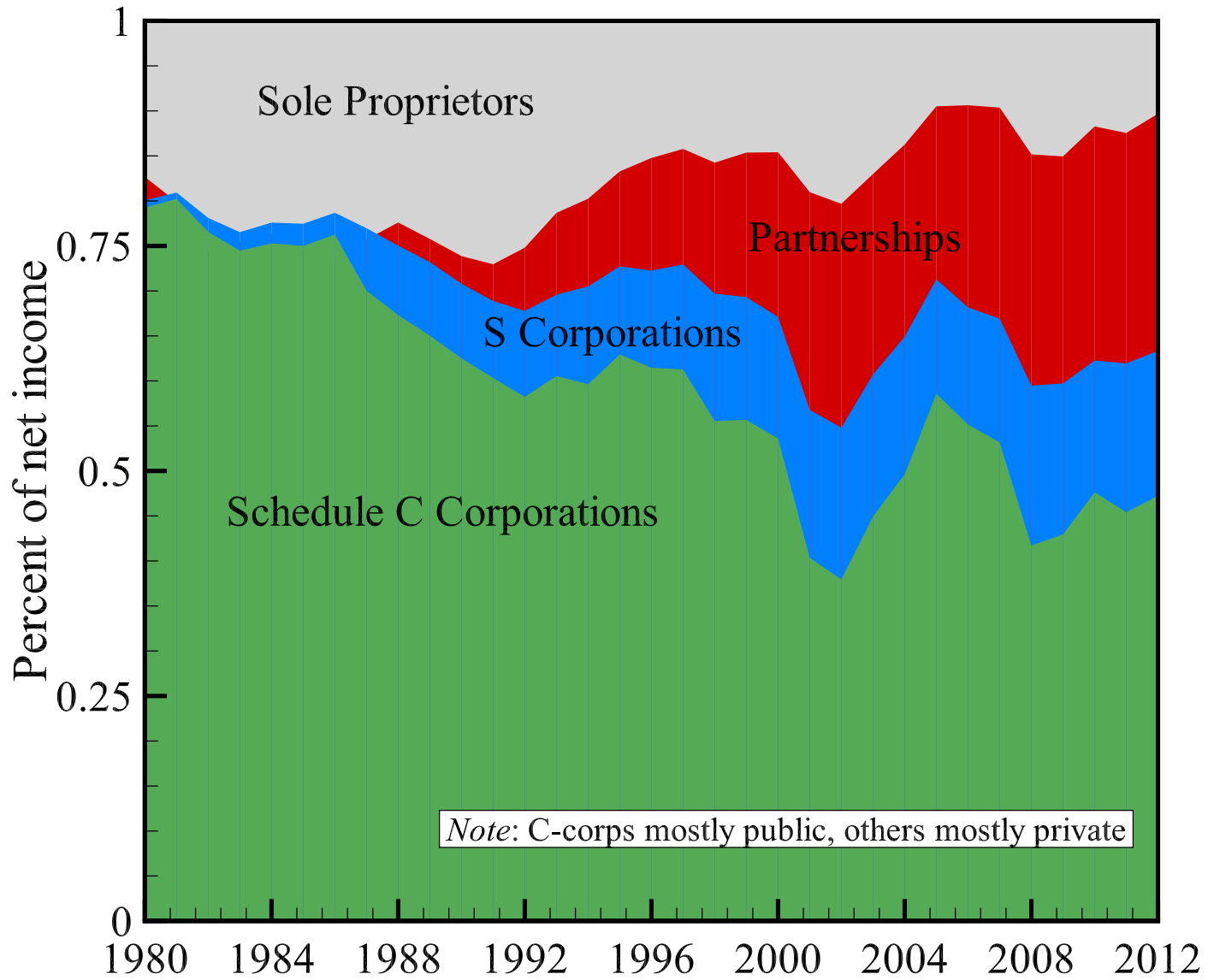
What Do Data Tell Us about Sole Proprietors?

Sole Proprietors

- In 2015, IRS reported

    ○ 27 million tax returns filed

    ○ 74% of all business returns

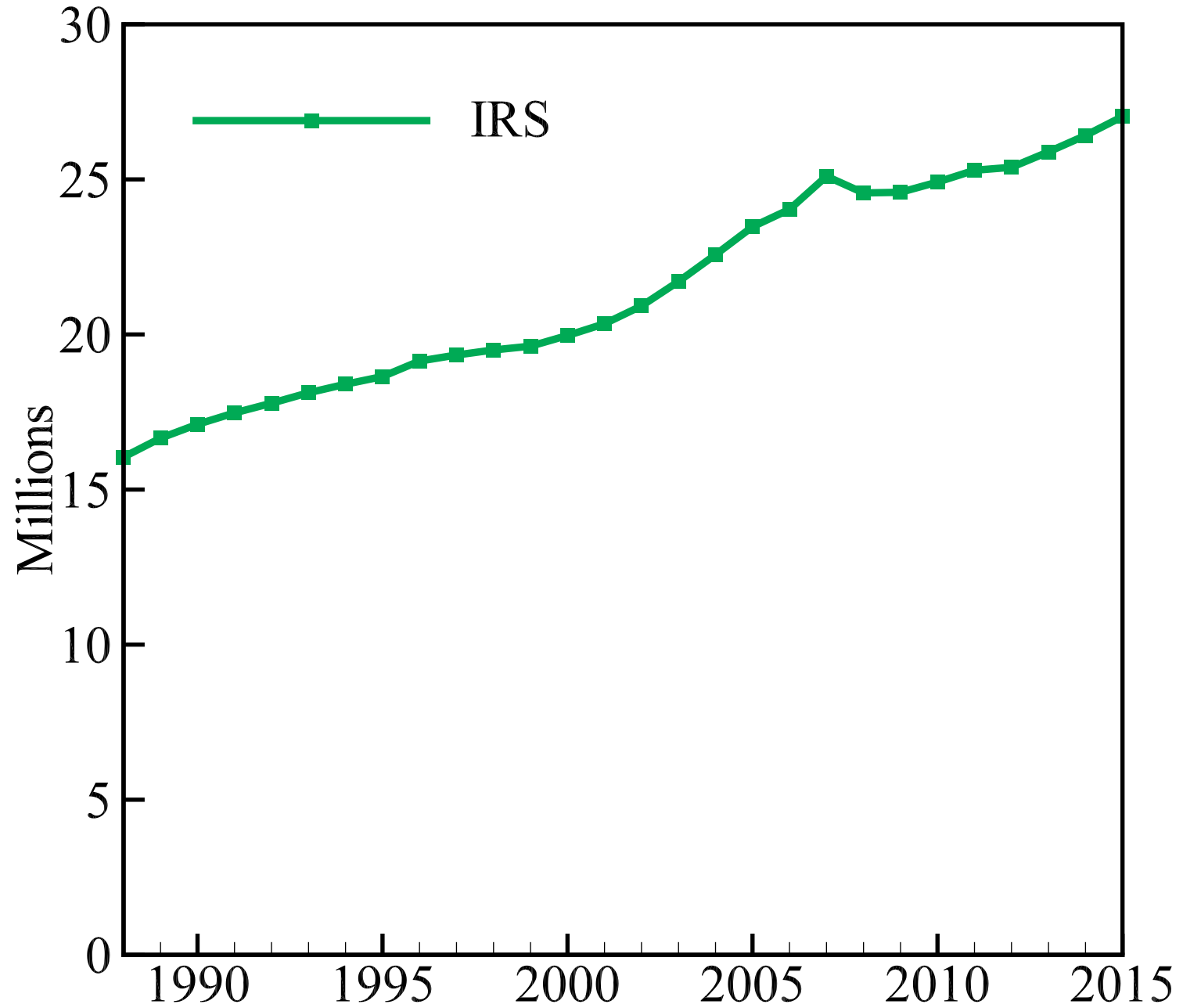   ○ $11,800 of *reported* net income per return
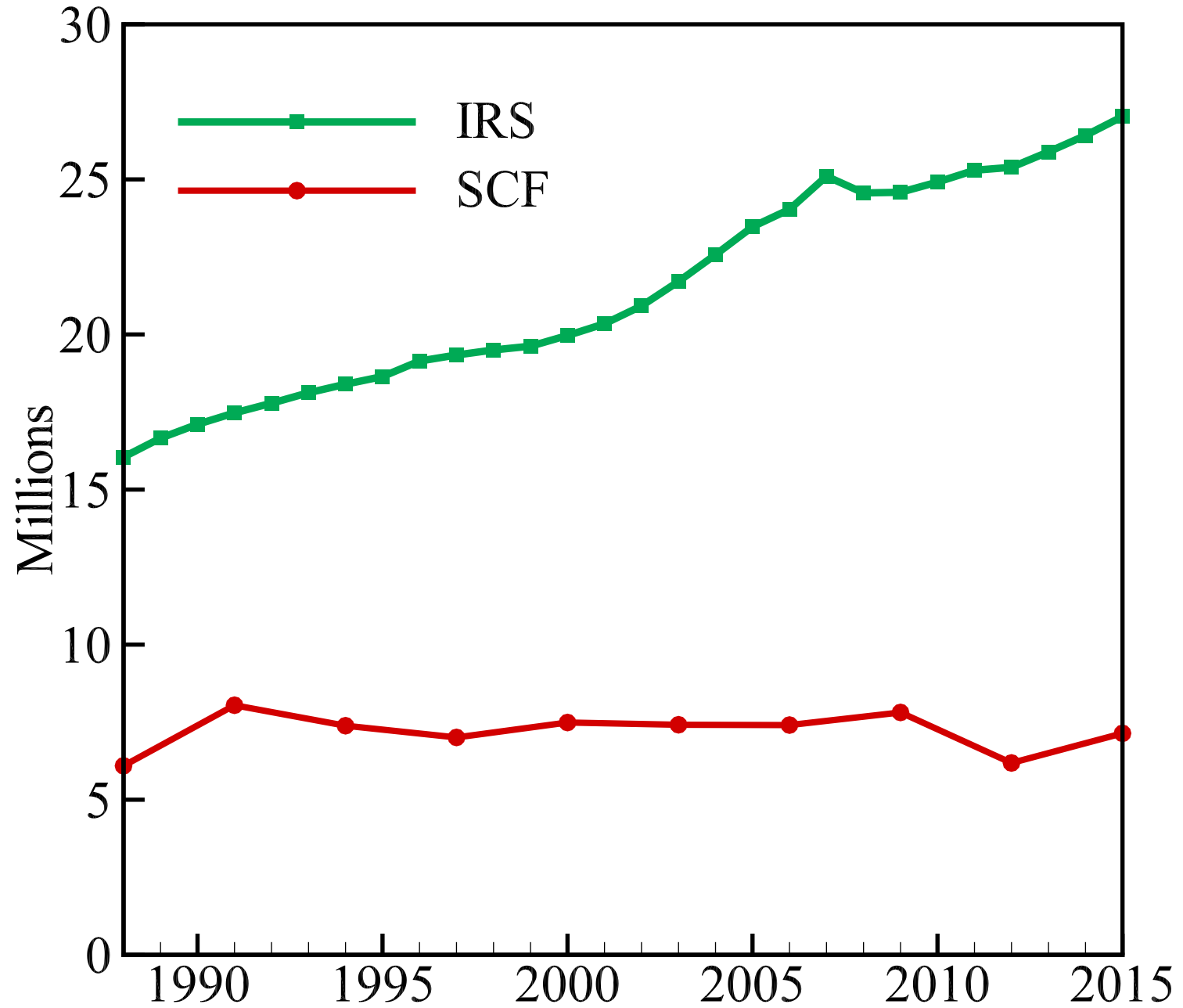
- What do SCF data tell us?

Sole Proprietor Returns
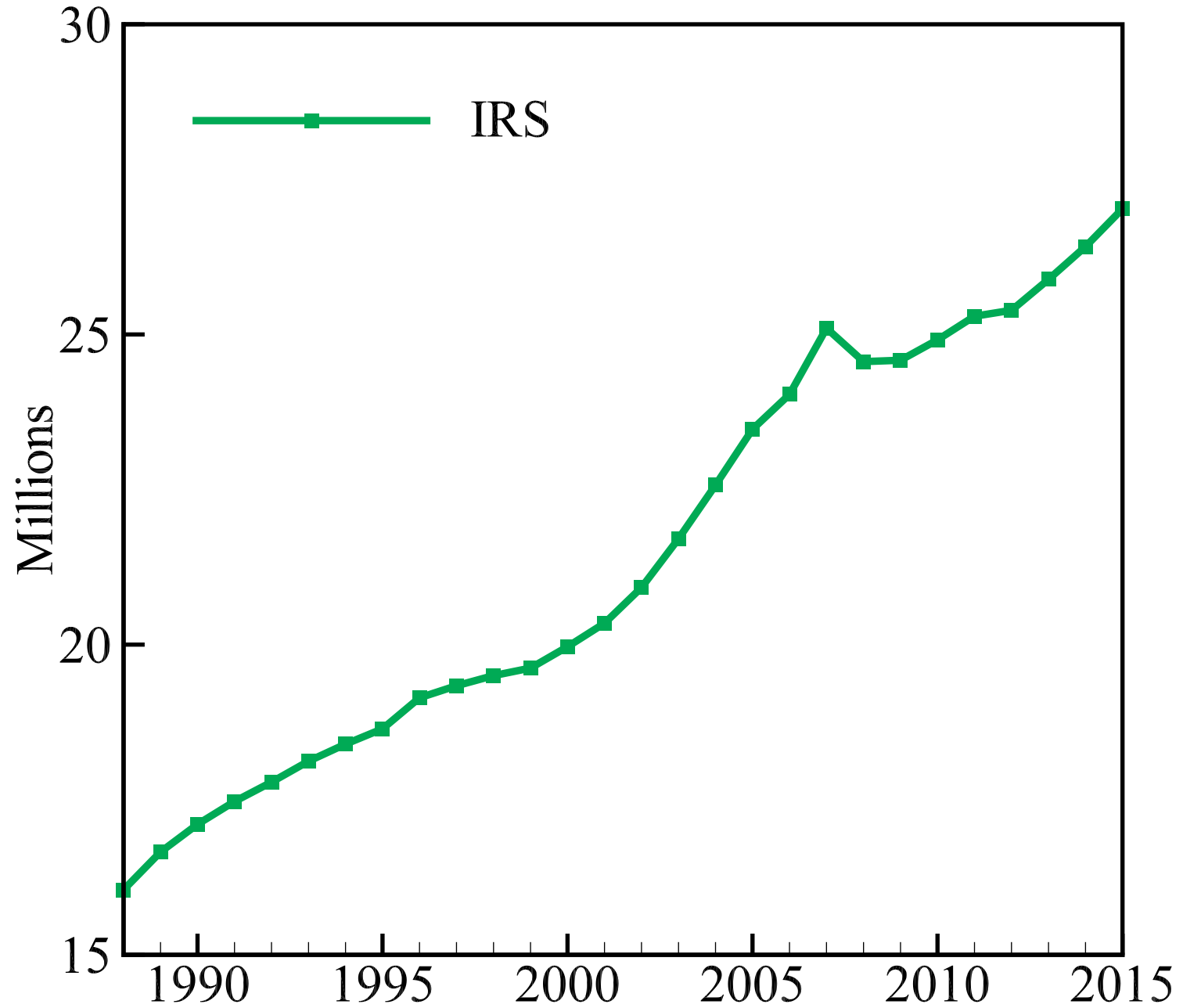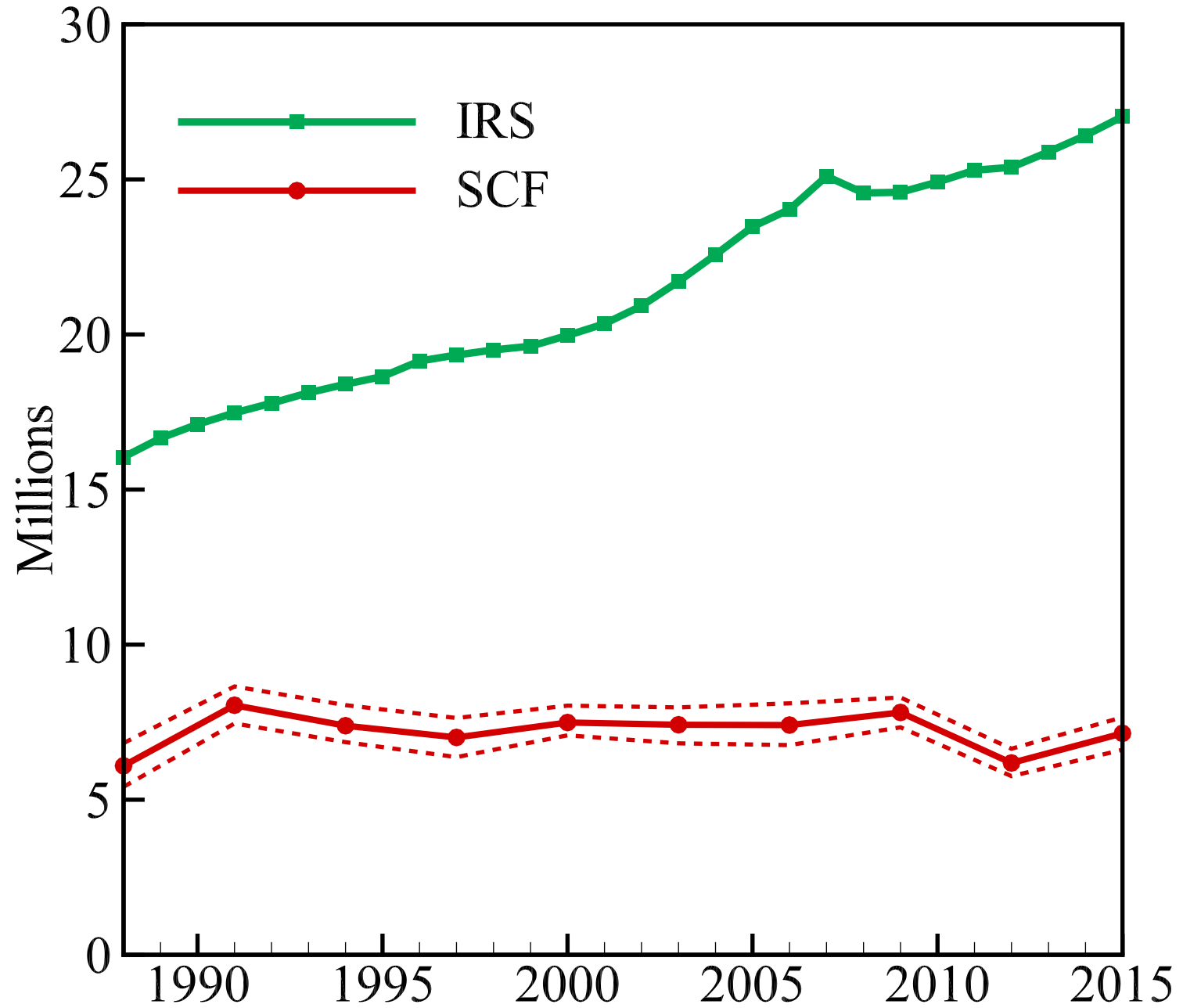
Sole Proprietor Returns

Sole Proprietor Returns
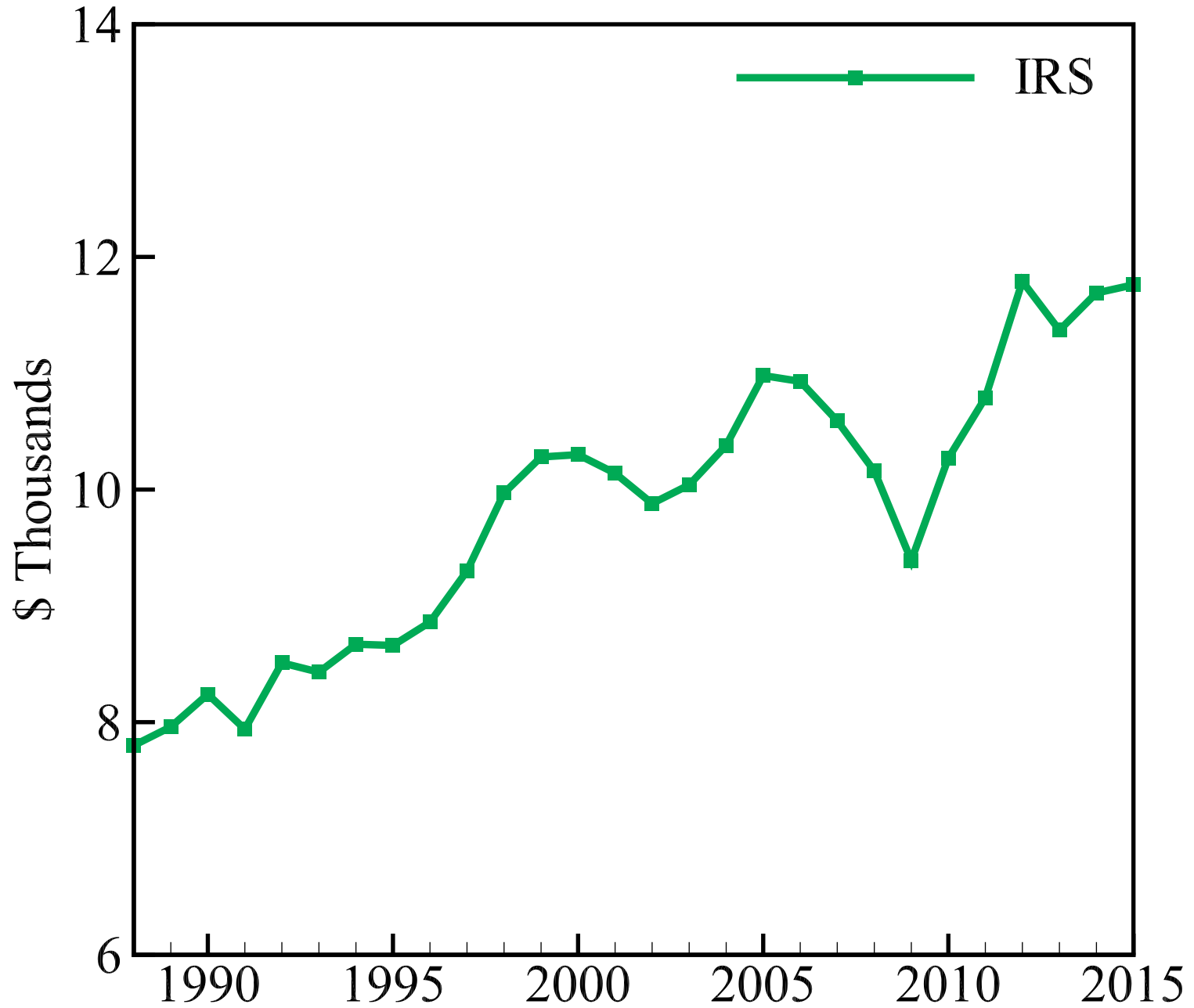
Sole Proprietor Returns

Sole Proprietor Returns

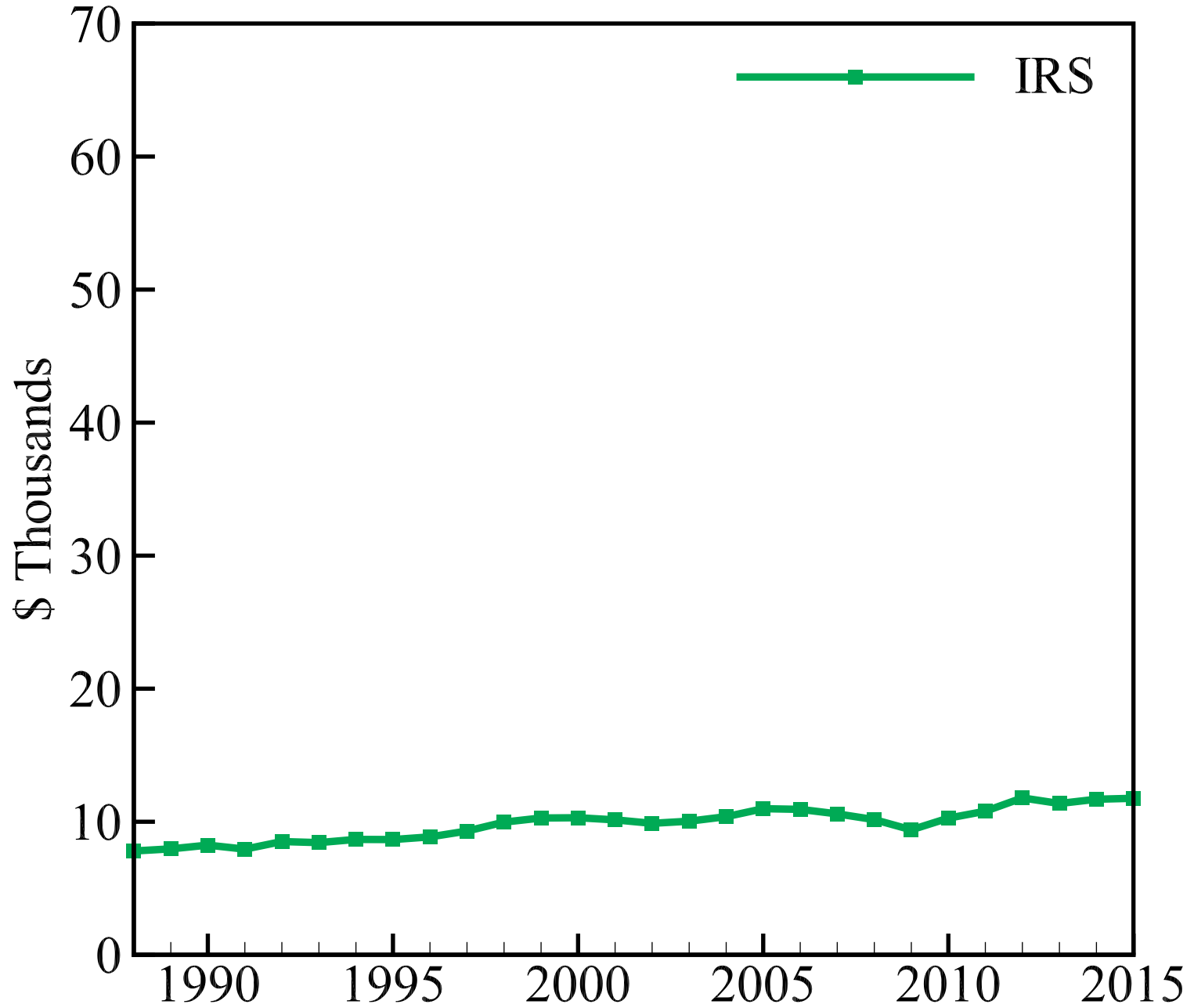$\Rightarrow$ Proprietors significantly underrepresented

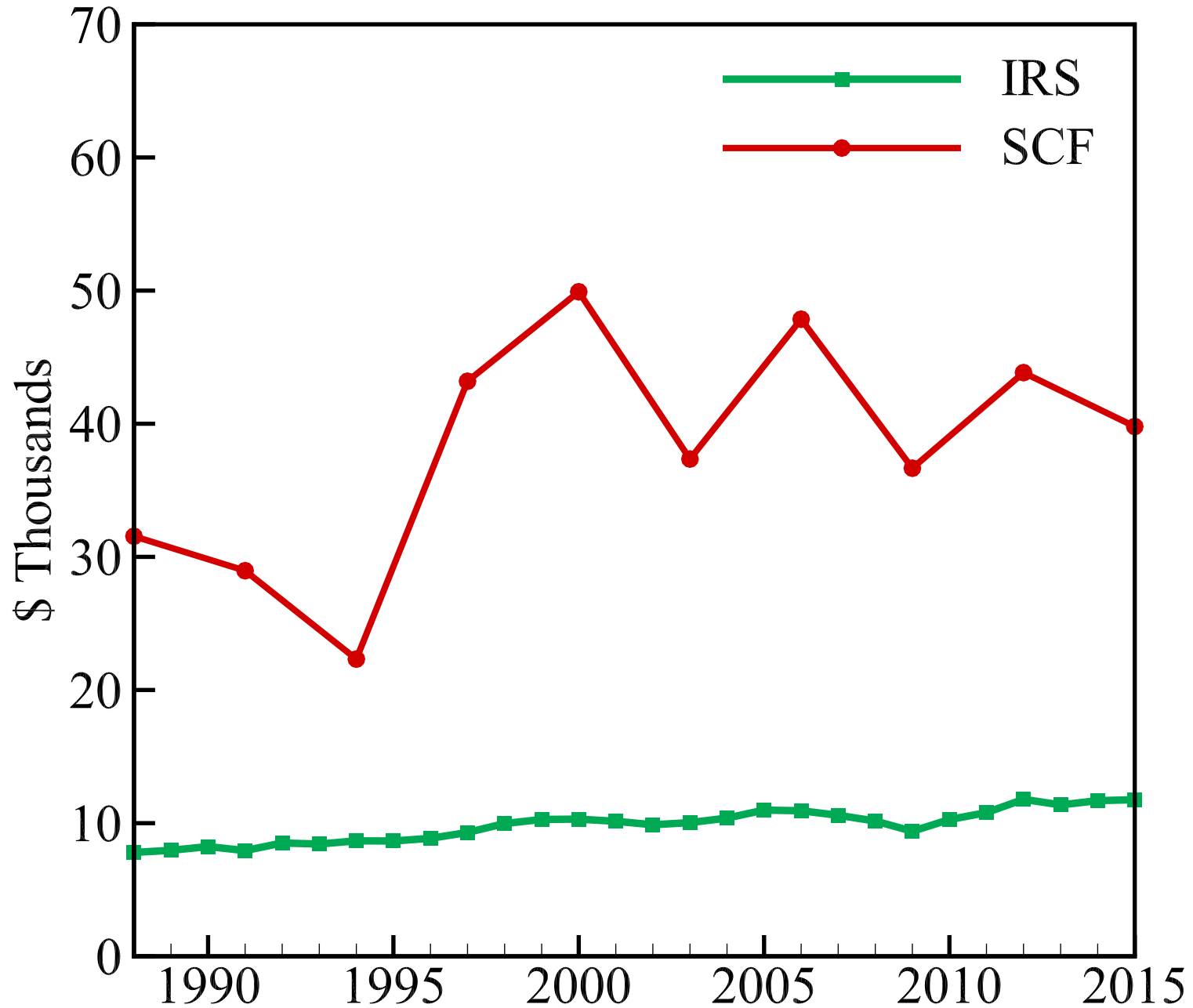# Sole Proprietor Income per Return

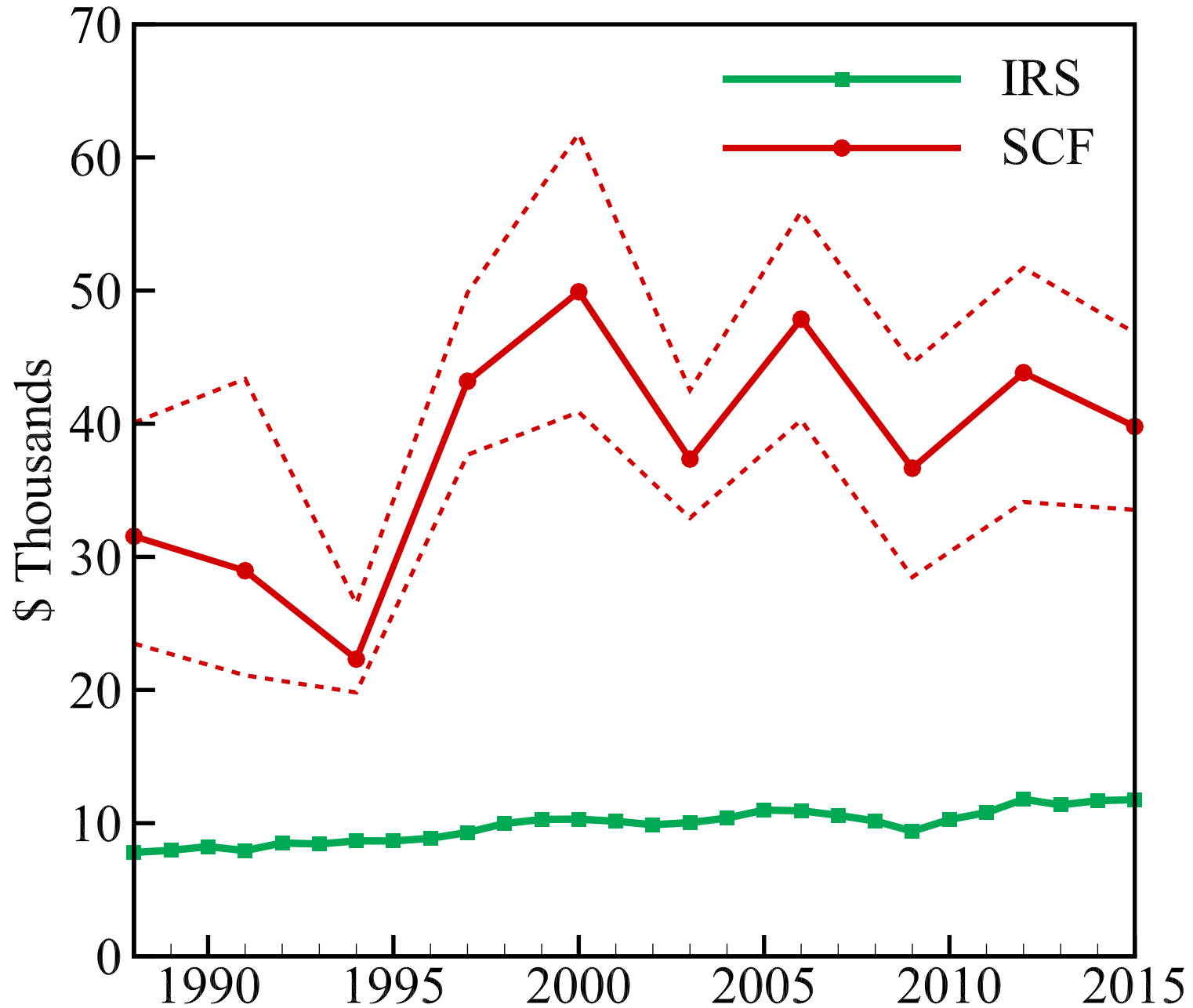# Sole Proprietor Income per Return

Sole Proprietor Income per Return

Sole Proprietor Income per Return

Sole Proprietor Income per Return

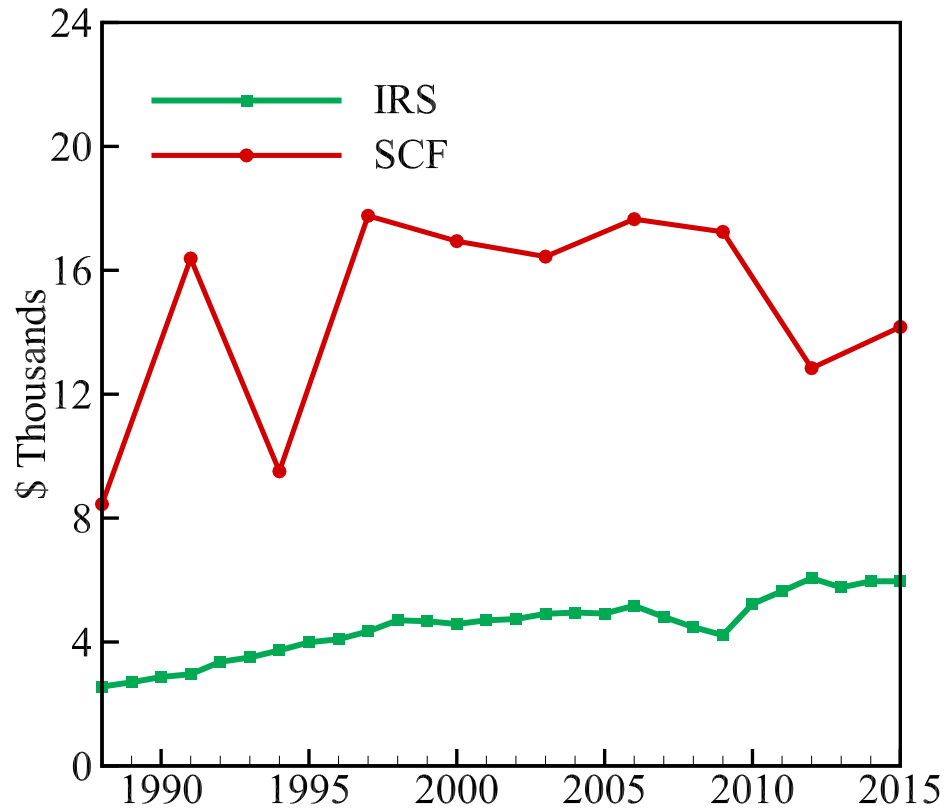$\Rightarrow$ Income significantly overstated

Sole Proprietor Income per Return

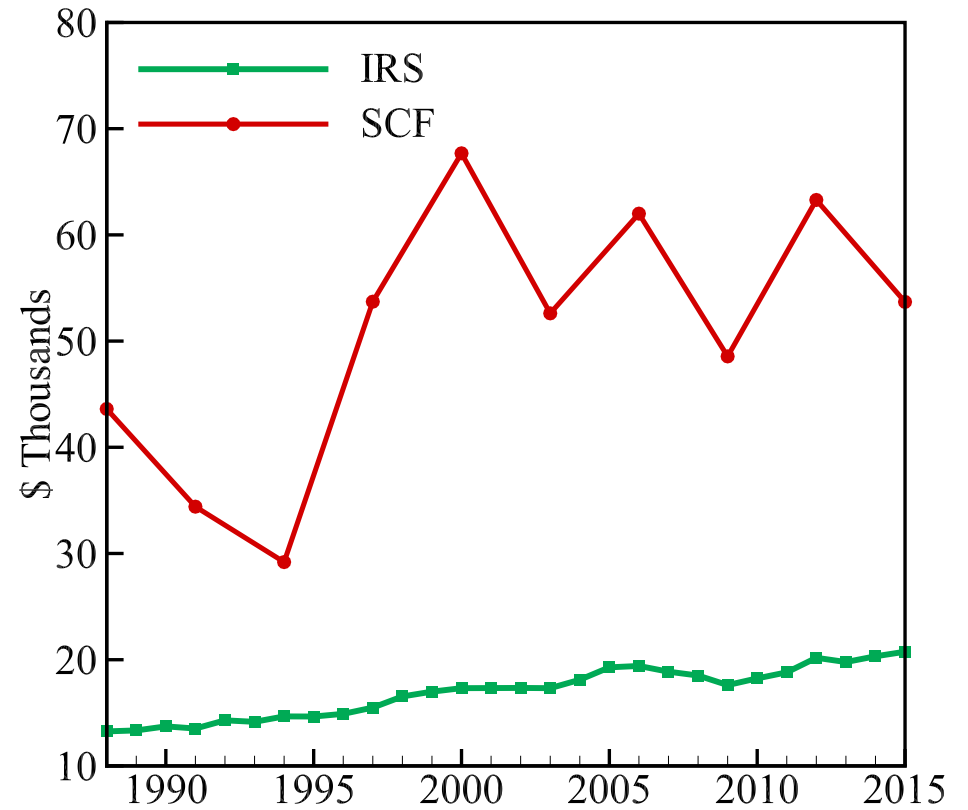$\Rightarrow$ Income significantly overstated

Let's consider the cross-section...

Sole Proprietor Income per Return

with AGI: Below-Median

Above-Median

Sole Proprietor Returns
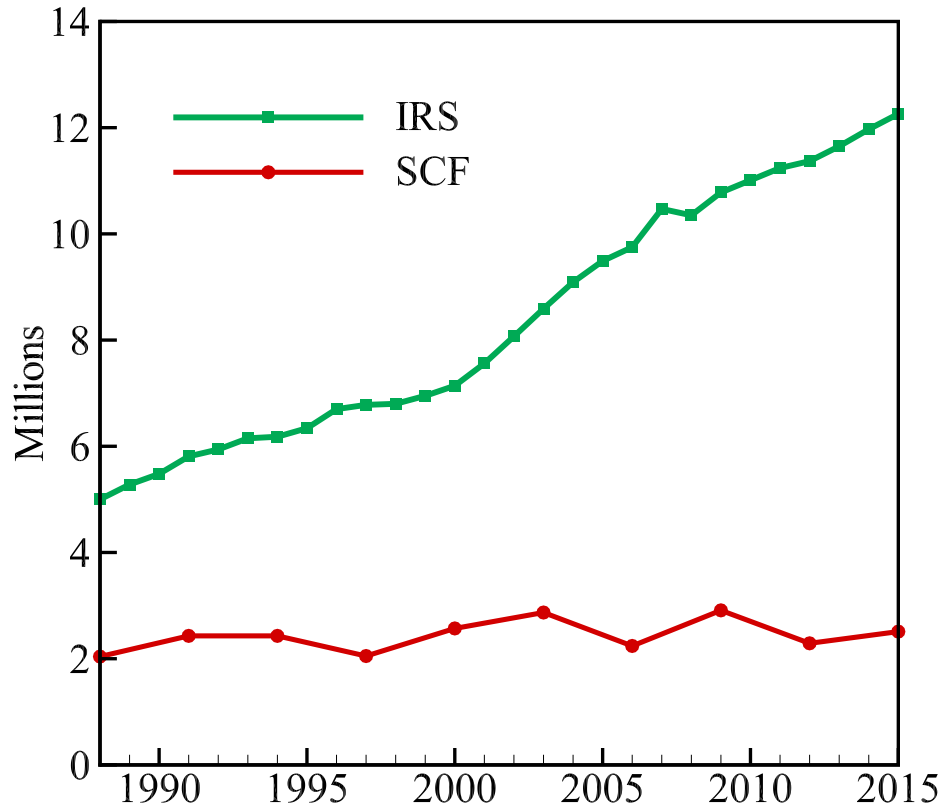
with AGI: Below-Median
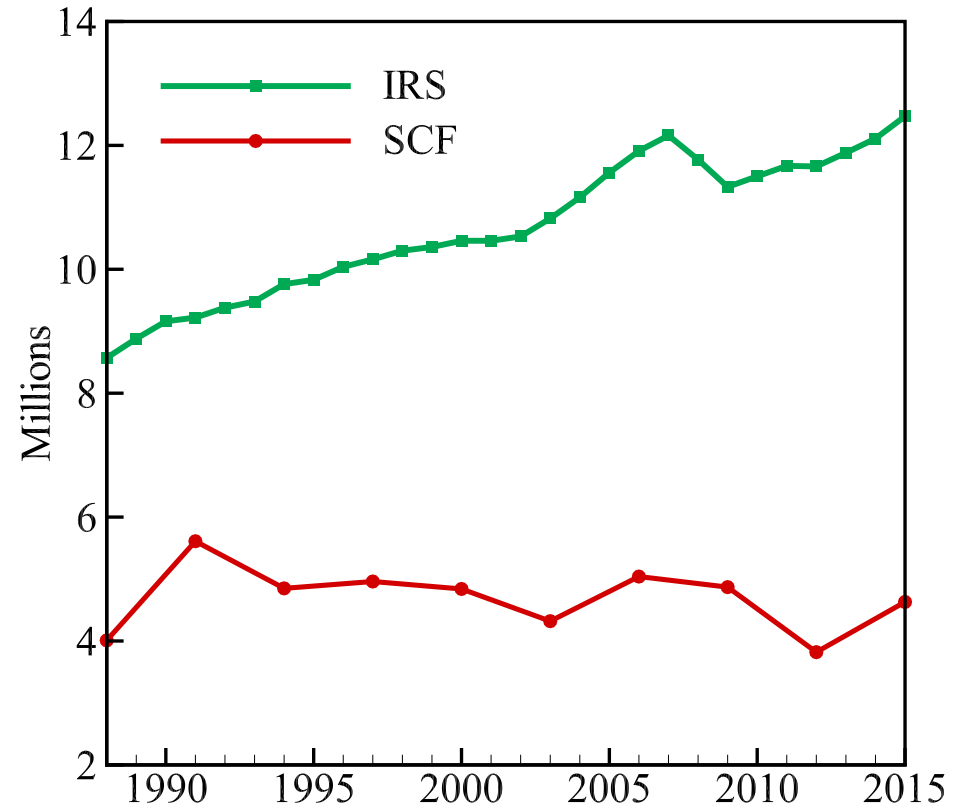
Above-Median

Sole Proprietor Returns

$\Rightarrow$ Underrepresentation of below-median businesses

      $\circ$ 25% in SCF

      $\circ$ 43% in IRS

And likely reason for overstatement of income

An Inconsistency in SCF

- 2 questions of sole proprietors:

  - What is on Form 1040, lines 12+18?

  - What is on Form 1040, Schedule C, line 31?

  $\Rightarrow$ Difference is Schedule F farm income

An Inconsistency in SCF

- 2 questions of sole proprietors:

    ○ What is on Form 1040, lines 12+18?

    ○ What is on Form 1040, Schedule C, line 31?

    $\Rightarrow$ Difference is Schedule F farm income

- But, difference is too large to be farm income

An Inconsistency in Income per Return

$ Thousands

Individual Form
Business Form
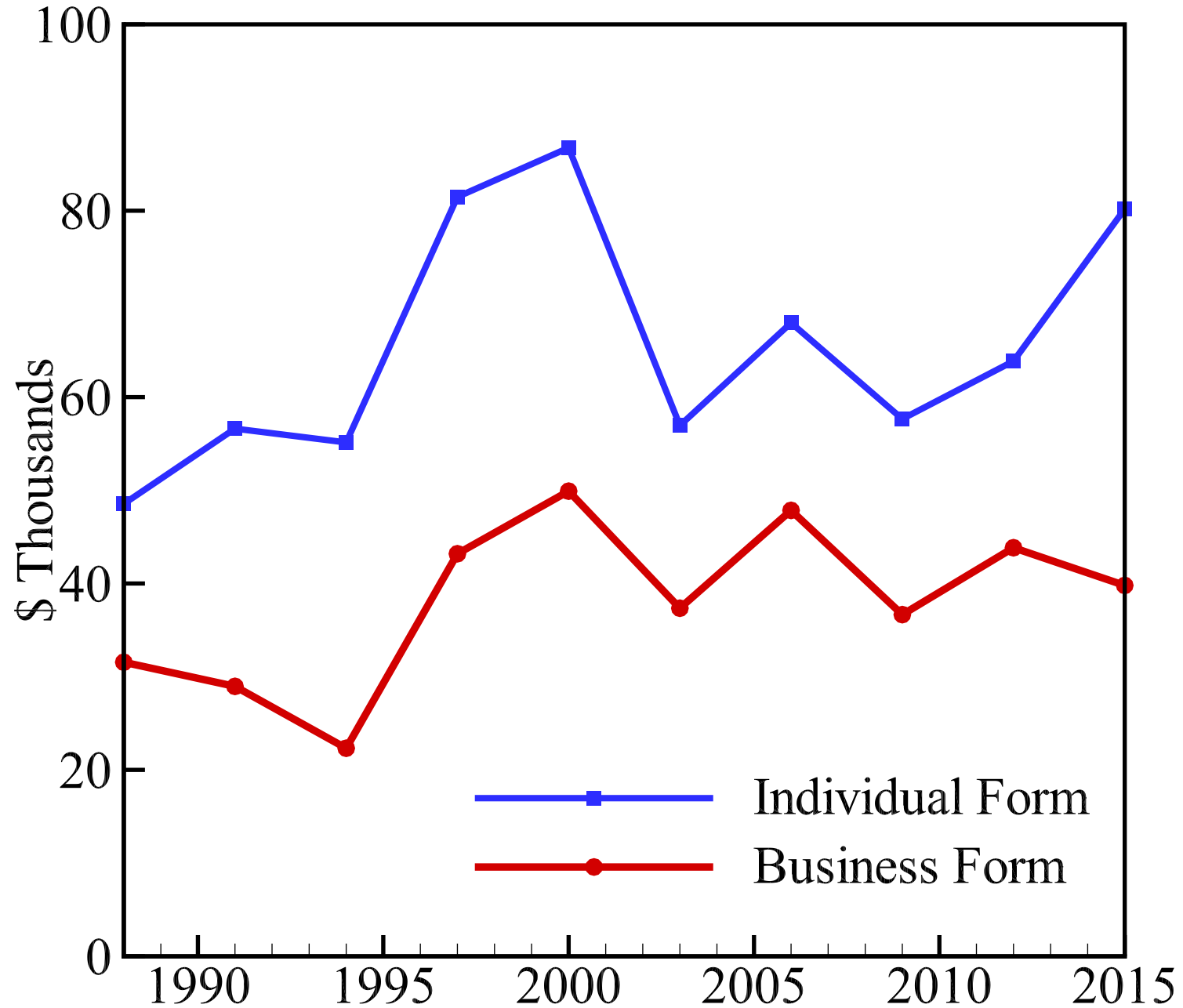
A Related Inconsistency in SCF

- 16 million reported nonzero proprietor income

  - 6 million said they actively managed a business

  - 10 million said they did not (eg, no Sched C)

  $\Rightarrow$ But, they earned 65% of the income

Recap for Sole Proprietors

- Proprietor returns significantly underrepresented

- Proprietor incomes significantly overstated

- Inconsistencies in respondent answers

What Do Data Tell Us about S Corporations?

S Corporations

- In 2013, IRS reported

    - 4.3 million tax returns filed

    - 12% of all business returns

    - $89,900 of *reported* net income per return

- What do SCF data tell us?

S Corporation Returns

S Corporation Income per Return

S Corporations Recap

$\Rightarrow$ As in case with sole proprietors

- S corporations significantly underrepresented

- Income per return significantly overstated

And aggregates are off in both cases

Partnerships

- In 2014, IRS reported

  - 3.6 million tax returns filed

  - 10% of all business returns

  - $123,700 of *reported* net income per return

  - $\approx 68\%$ of income, 27% of partners are corporations

- SCF surveys individual partners

  - What do data tell us?

Partnership Returns

Partnerhsip Income per Return
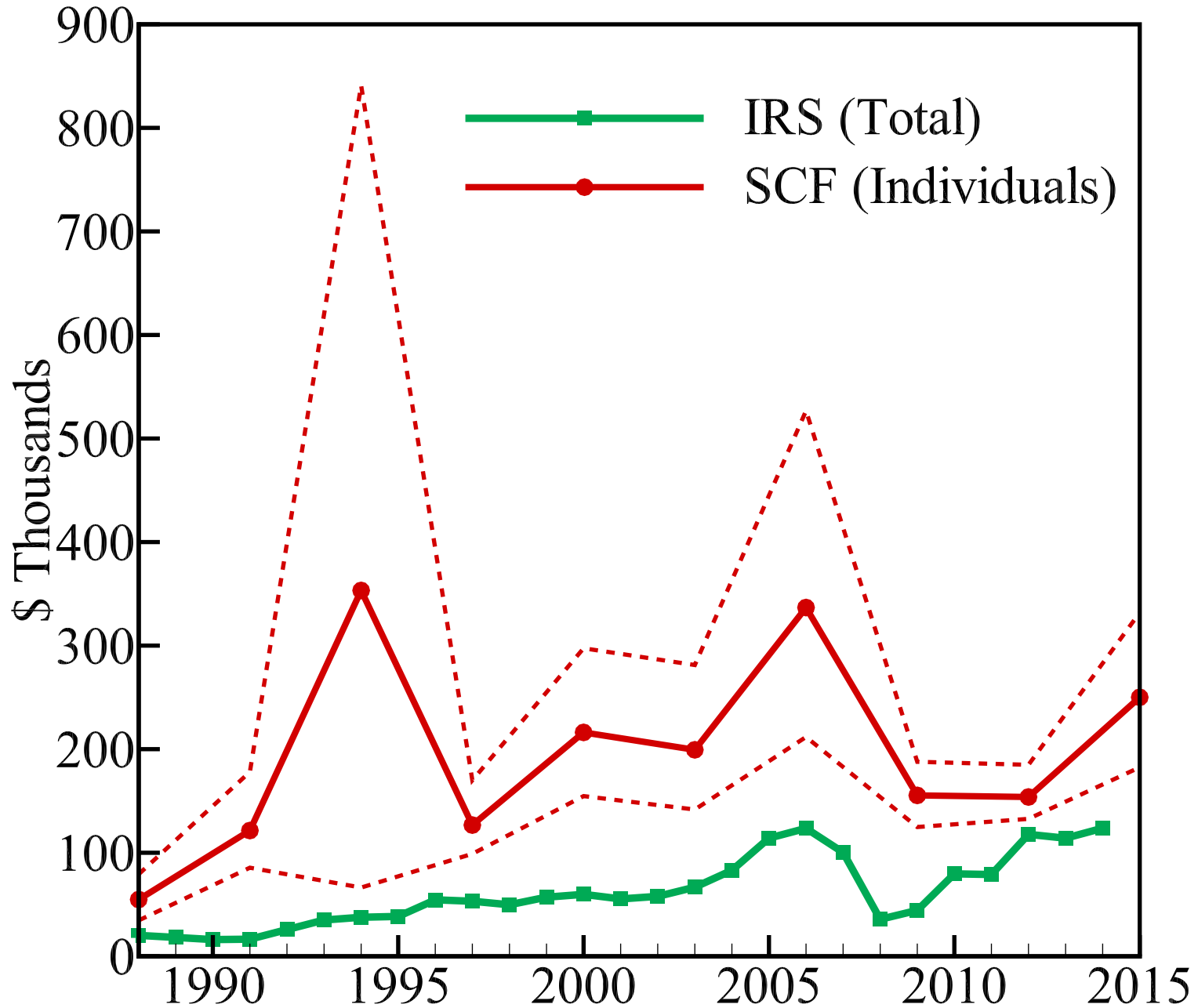
Partnership Recap

$\Rightarrow$ Unlike sole proprietors and S corporations

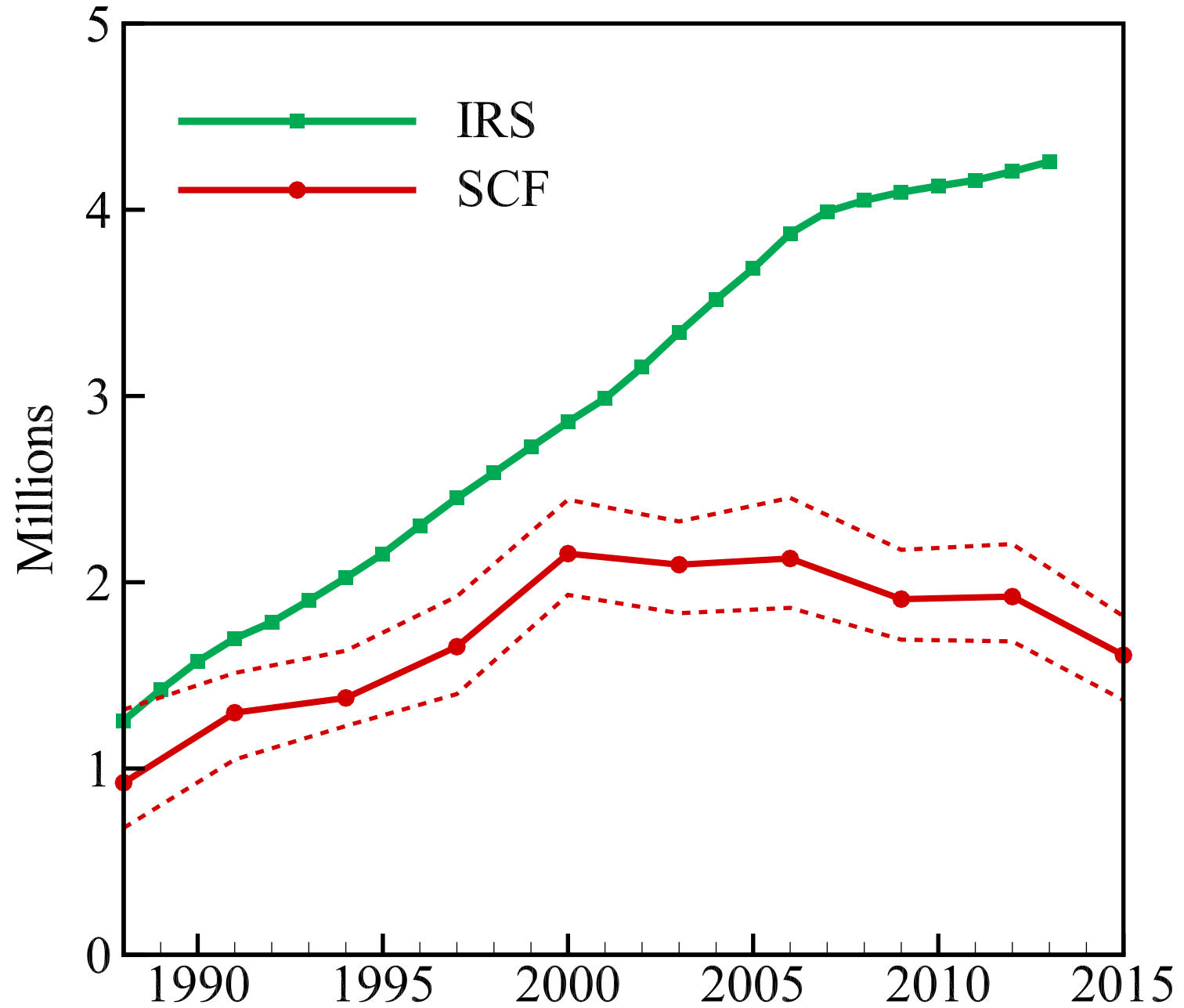 $\circ$ Numbers of returns and incomes *both* overstated

C Corporations

- In 2013, IRS reported

    - 1.6 million tax returns filed

    - 4% of all business returns

    - $670,600 of *reported* net income per return

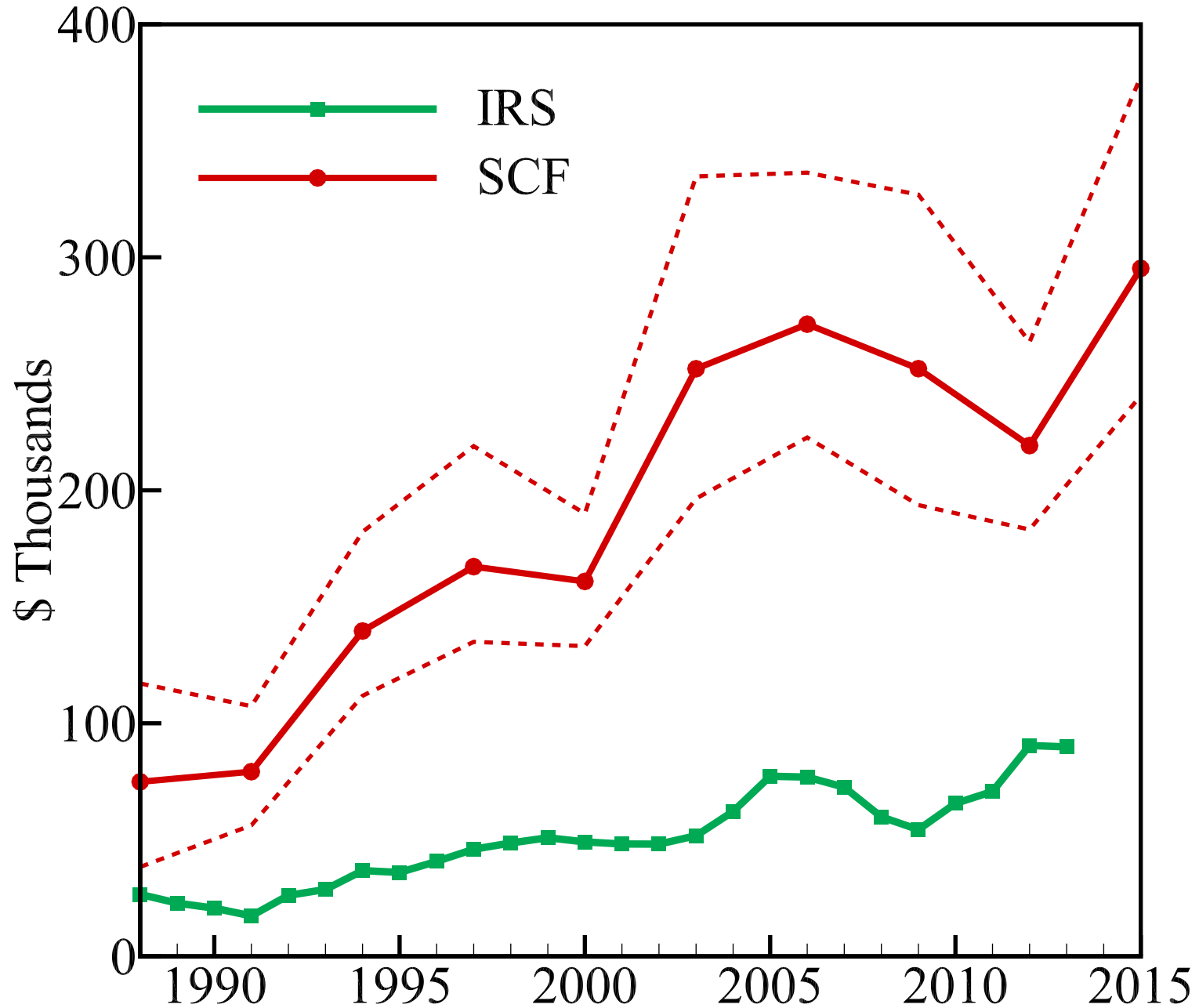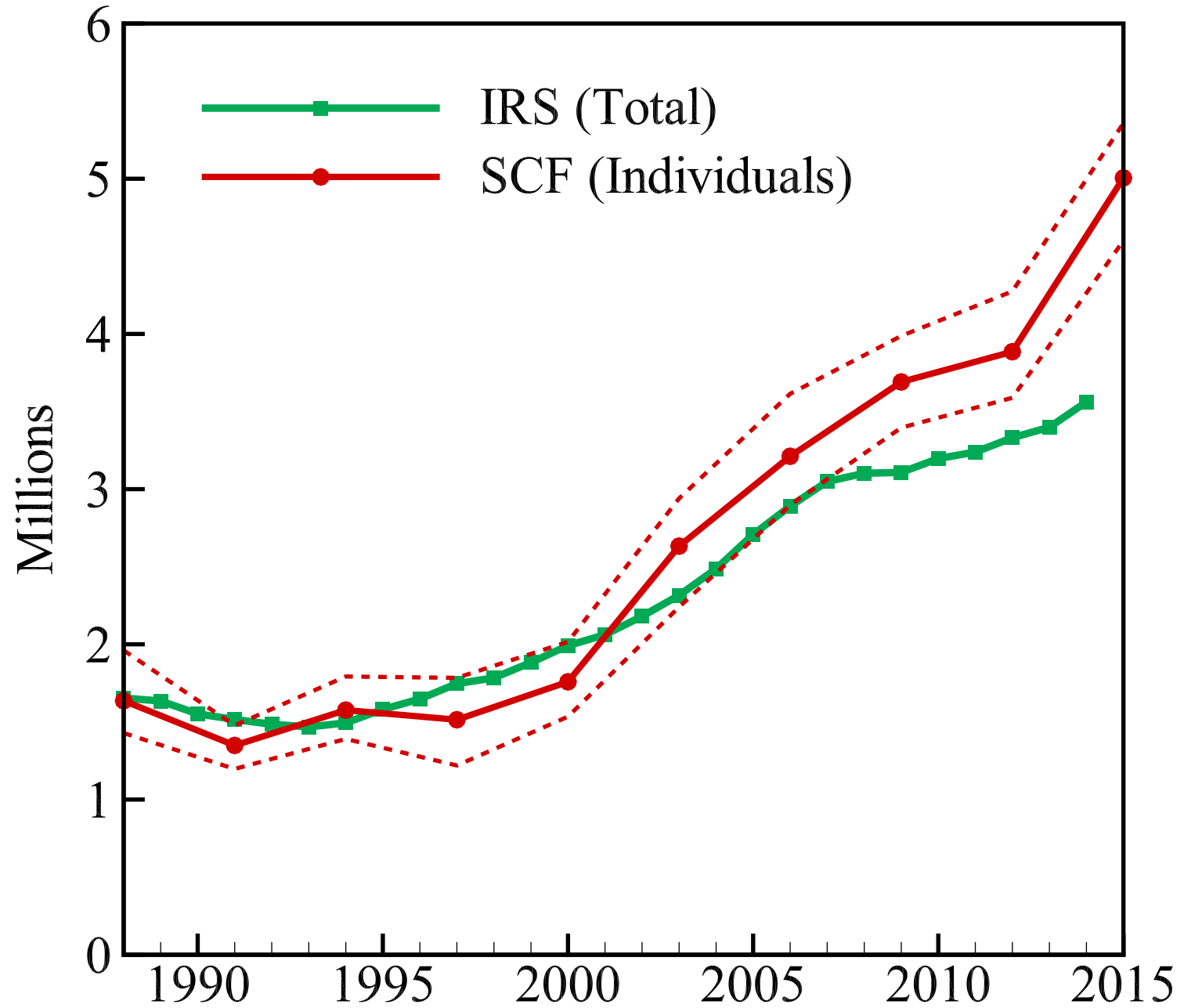- SCF surveys non-publicly traded businesses

    - What do data tell us?

C Corporation Returns
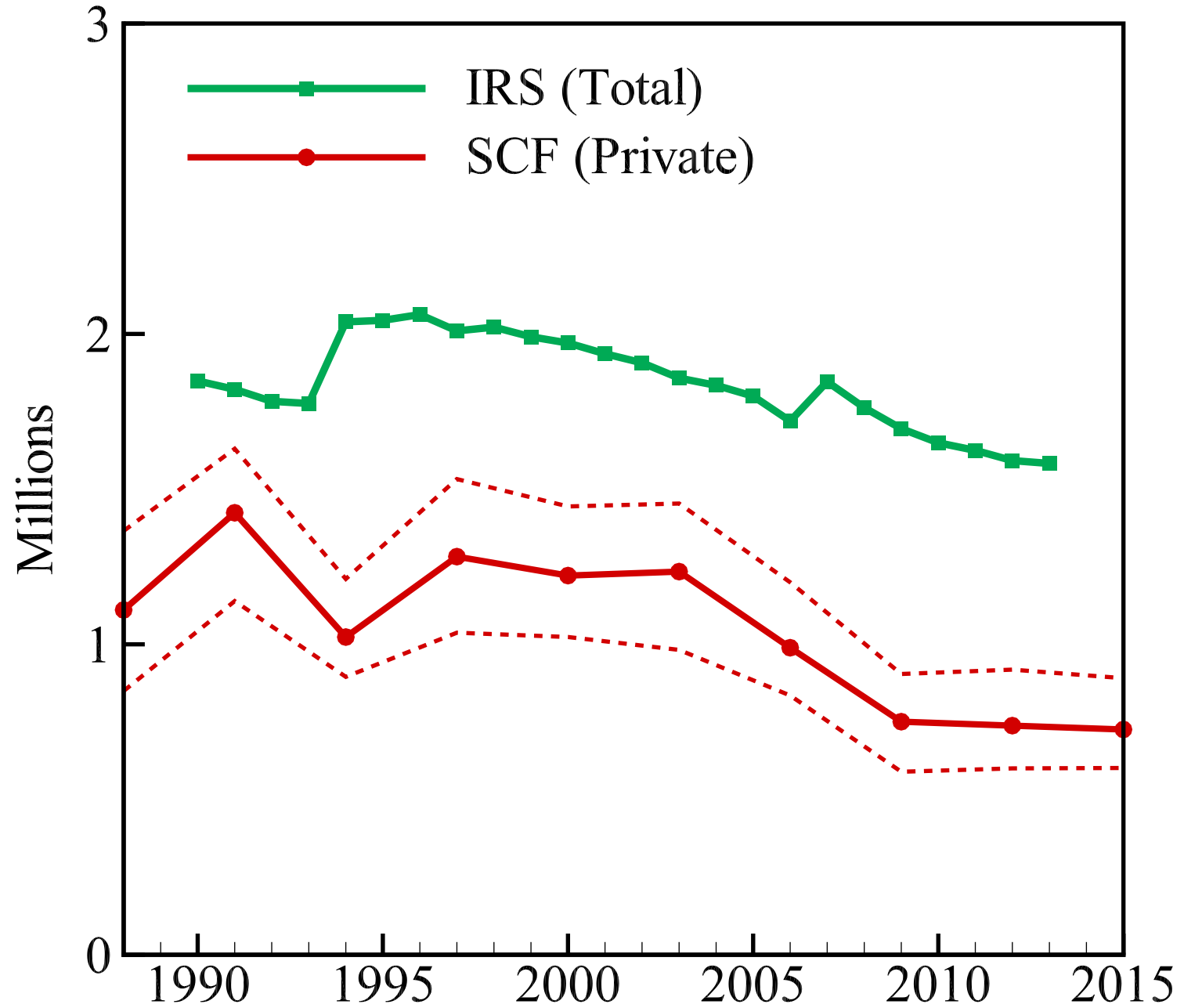
C Corporation Income per Return
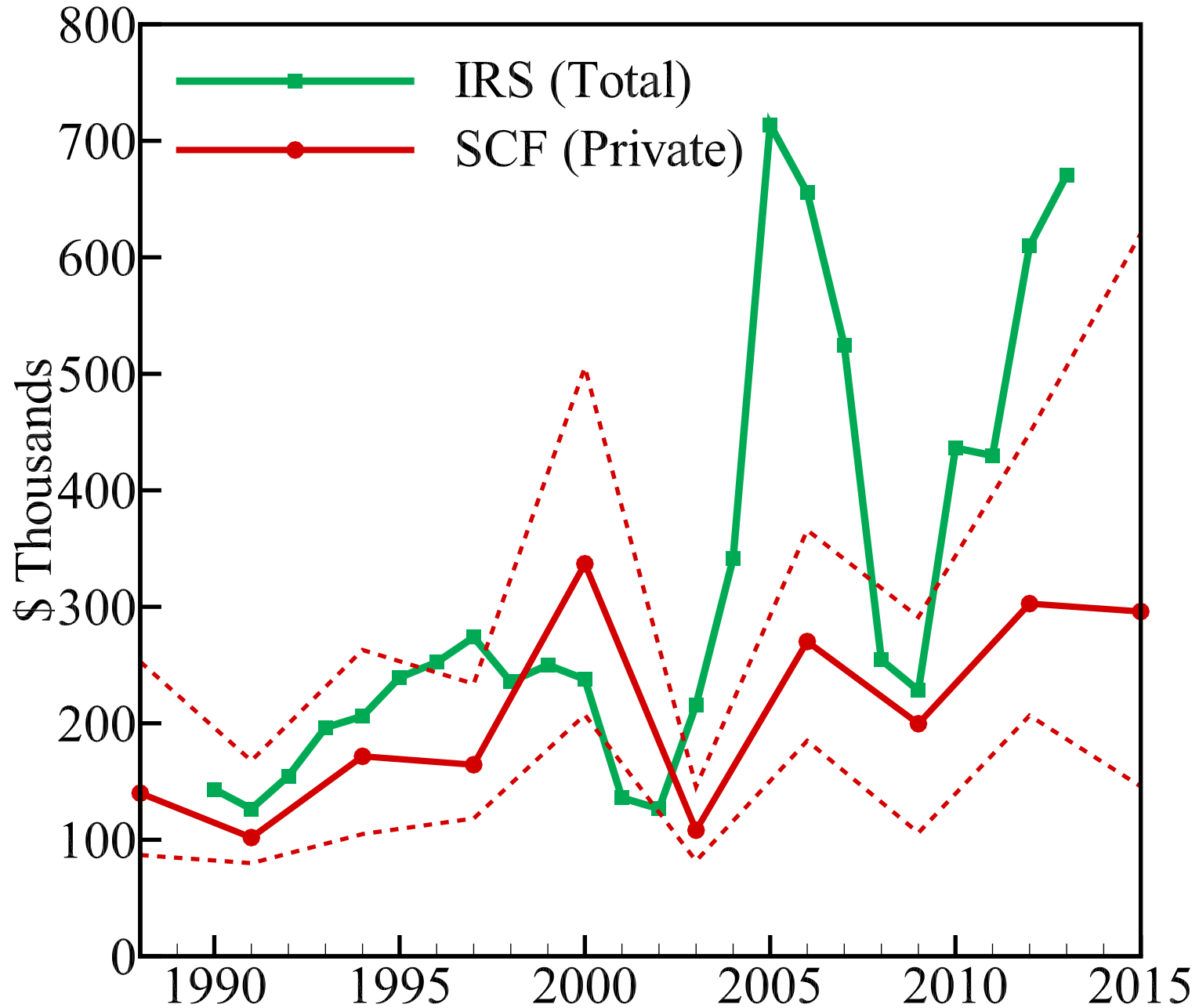
C Corporations Recap

$\Rightarrow$ As in case with sole proprietors, S corporations

- Private C corps significantly underrepresented

- Income per return significantly overstated

Sources of SCF-IRS Discrepancies

Sources of SCF-IRS Discrepancies

- Discrepancy between survey (S) and IRS (I) aggregate:

$$X^S - X^I = \underbrace{\left(\sum_i \frac{(\omega_i^S - \omega_i^I)}{\omega_i^I} \right) X^I + \sum_i \omega_i^I X_i^I \left(\frac{\omega_i^S}{\omega_i^I} - \frac{\sum_i \omega_i^S}{\sum_i \omega_i^I} \right)}_{\text{sampling errors}}$$

$$+ \underbrace{\sum_i \omega_i^S (X_i^S - X_i^I)}_{\text{measurement errors}}$$

Sources of SCF-IRS Discrepancies

- Discrepancy between survey (S) and IRS (I) aggregate:

$$X^S - X^I = \underbrace{\left(\sum_i \frac{(\omega_i^S - \omega_i^I)}{\omega_i^I} \right) X^I + \sum_i \omega_i^I X_i^I \left(\frac{\omega_i^S}{\omega_i^I} - \frac{\sum_i \omega_i^S}{\sum_i \omega_i^I} \right)}_{\text{sampling errors}}$$

$$+ \underbrace{\sum_i \omega_i^S (X_i^S - X_i^I)}_{\text{measurement errors}}$$

- Next, consider evidence for sampling errors $(\omega_i^S \neq \omega_i^I)$

Sampling Errors

- Evidence for $\sum_i \omega_i^S \neq \sum_i \omega_i^I$ shown earlier:

    ○ Underrepresentation of proprietors and corporations

    ○ Overrepresentation of partnerships

- Evidence for $\omega_i^S / \omega_i^I \neq \sum_i \omega_i^S / \sum_i \omega_i^I$ shown earlier

    ○ Underrepresentation of low-income businesses

Sources of SCF-IRS Discrepancies

- Discrepancy between survey (S) and IRS (I) aggregate:

$$X^S - X^I = \underbrace{\left(\sum_i \frac{(\omega_i^S - \omega_i^I)}{\omega_i^I} \right) X^I + \sum_i \omega_i^I X_i^I \left(\frac{\omega_i^S}{\omega_i^I} - \frac{\sum_i \omega_i^S}{\sum_i \omega_i^I} \right)}_{\text{sampling errors}}$$

$$+ \underbrace{\sum_i \omega_i^S (X_i^S - X_i^I)}_{\text{measurement errors}}$$

- Next, consider evidence for $X_i^S \neq X_i^I$

Measurement Errors

- Evidence for $X_i^S \neq X_i^I$:

    - Inconsistency in answers (shown earlier)

    - Problematic framing of questions

    - Little referencing of documents

- Next, consider evidence for misinterpreted questions

Measurement Errors: Framing

- Question: *What is your net income?*

- Possible answers for those with losses

    ○ Report actual loss

    ○ Report zero since there was no income

- Evidence of the latter...

Proprietors with Net Losses

- Sort proprietors by AGI into 19 bins

- How many reported no losses?

    - IRS: none

    - SCF: 10 of 19 bins (26% of losses)

Proprietors with Net Losses

- Sort proprietors by AGI into 19 bins

- How many reported no losses?

    - IRS: none

    - SCF: 10 of 19 bins (26% of losses)

  Next, what's known about referencing documents

Measurement Errors: Referencing

- Question: *Were documents referenced?*

- Answers for income tax documents

    - 75% never

    - 2% rarely

    - 9% sometimes

    - 14% frequently

Measurement Errors: Referencing

- Question: *Were documents referenced?*

- Answers for financial documents

    - 64% never

    - 6% rarely

    - 15% sometimes

    - 15% frequently
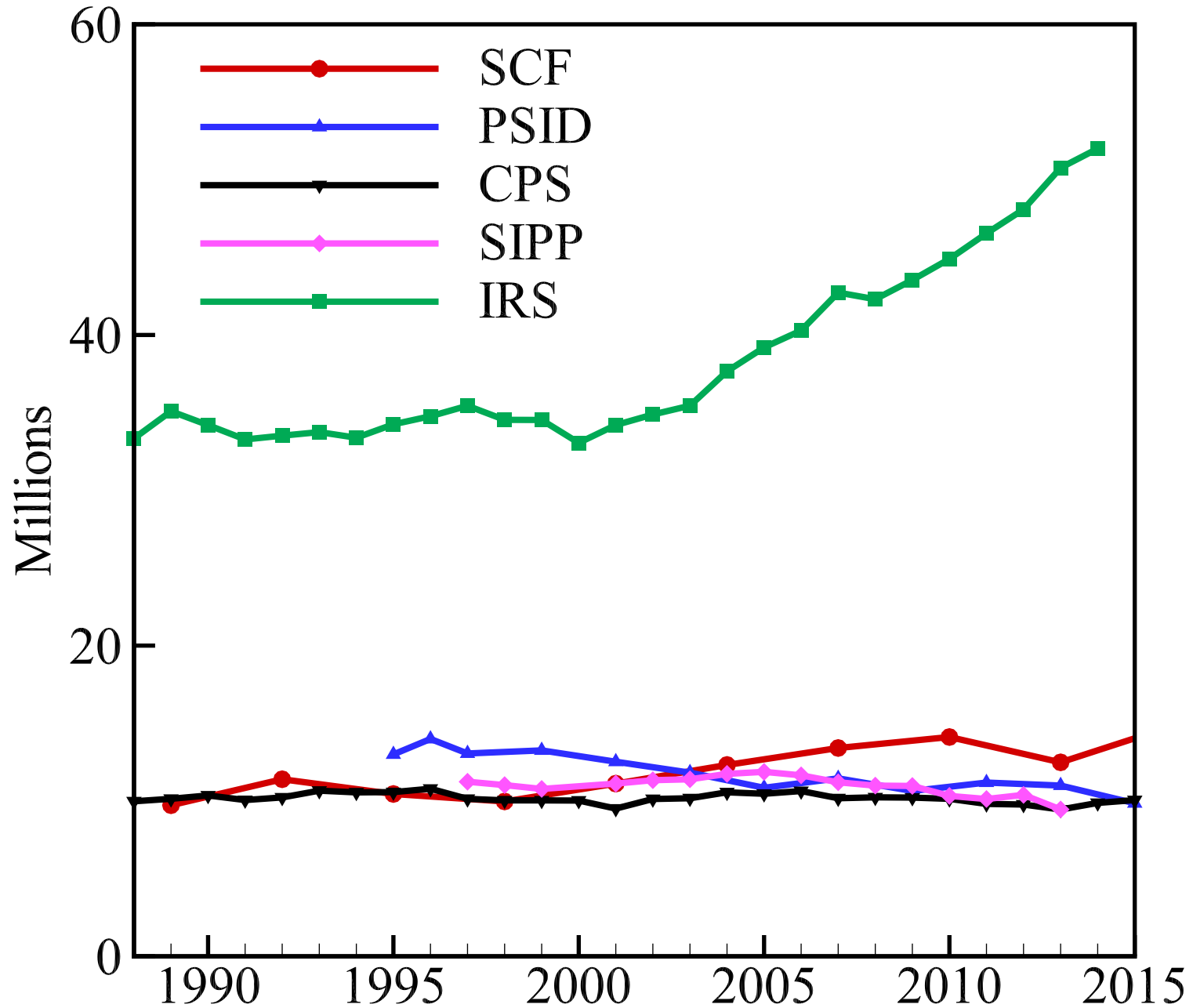
Other Surveys: PSID, SIPP, CPS

Other Surveys

- Relative to SCF:

    ○ Legal form of business not distinguished

    ○ Incomes not connected to line items on tax forms

- To compare across surveys, restrict attention to

    ○ Numbers of owners

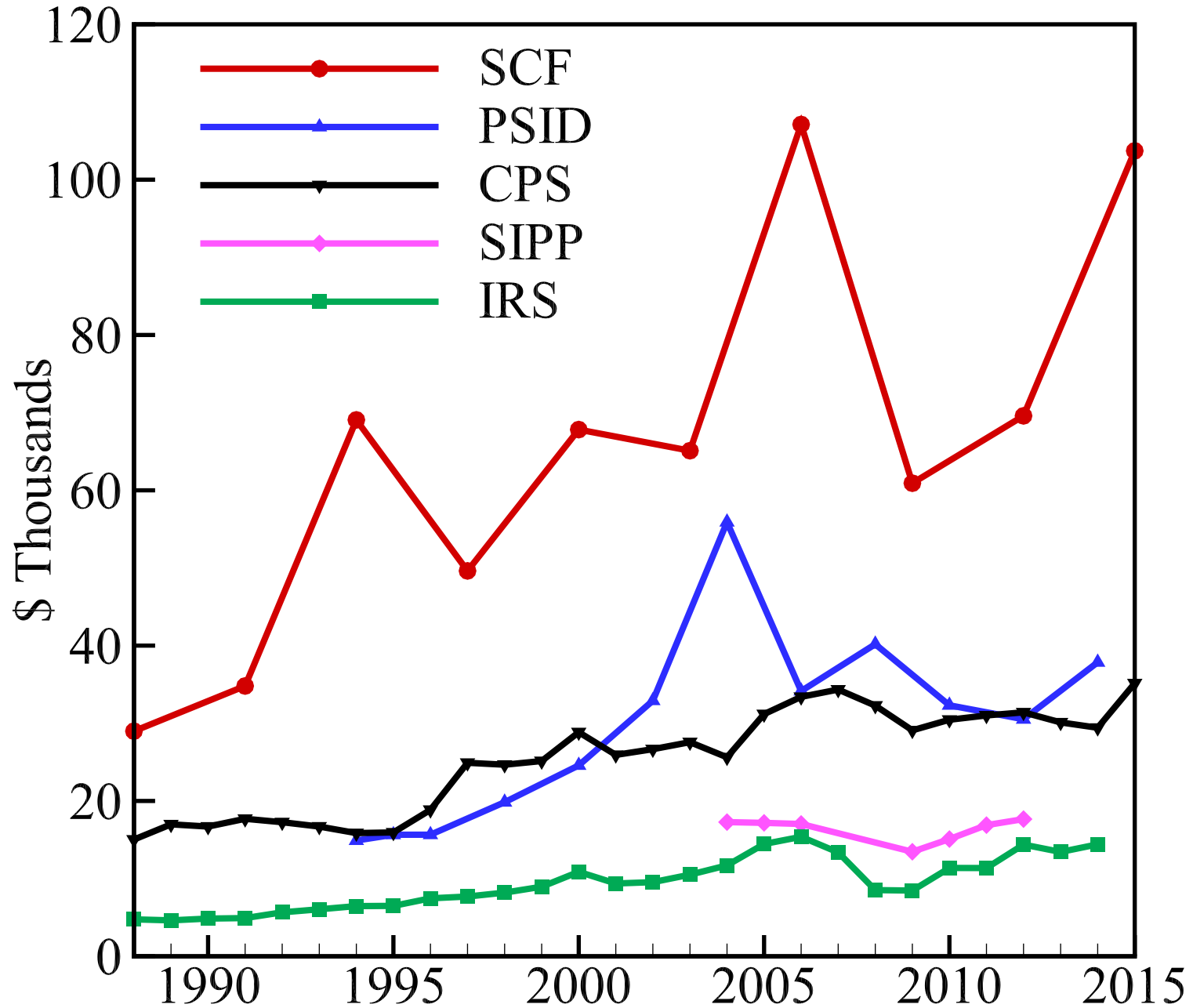   ○ Unincorporated business incomes

Unincorporated Business Owners

Millions

- SCF
- PSID
- CPS
- SIPP
- IRS

Unincorporated Business Income per Owner

Other Survey Recap

- Inconsistencies between all surveys and IRS

- Inconsistencies across surveys:

  - Driven by differences in aggregate incomes

  - Not driven by numbers of owners

Business Valuations and Returns

Business Valuations

- Survey valuations:

    ○ Self reported by ongoing business

    ○ No empirical counterpart for private businesses

    $\Rightarrow$ Focus on net income yields from CRSP, Pratt's

Value-Weighted Net Income Yields

- All businesses

    - 19%, SCF

    - 7%, CRSP

    - 2%, Pratt's brokered sales data

- Small businesses (ranked by assets)

    - 15%, SCF S corporations

    - -9%, CRSP bottom 20%

- Unincorporated businesses

    - 15%, SCF

    - 18%, SIPP

    - 15%, PSID

Value-Weighted Net Income Yields

$\Rightarrow$ Income yields significantly overstated

Total Rates of Return

- Need estimates of capital gains

- Not available for privately-held firms

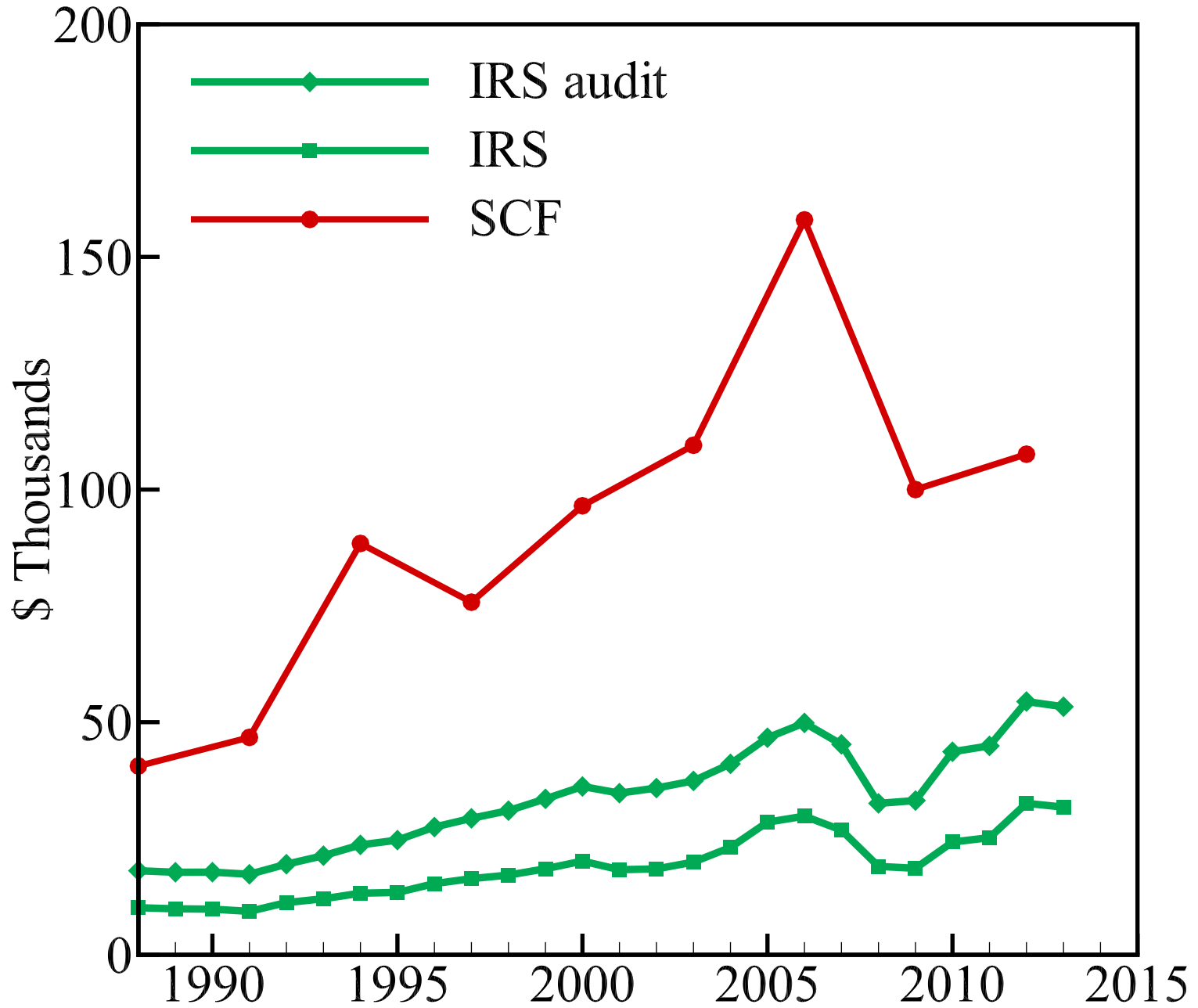$\Rightarrow$ Not possible to compare total returns in SCF and CRSP

Robustness

- Adjust for income misreporting to IRS

- Check for miscategorization of business income

Income Misreporting

Check Miscategorization

- Do owners confuse business income categories, eg,

  - Schedule C (proprietors)

  - Schedule E (rents, royalties, pass-throughs)?

- Would overstate one and understate other

- But, find overstatement in both

What Do Survey Data Tell Us about Aggregate Capital Income?

Aggregate Capital Income

- Given measurement issues, tempting to:

  - Combine all nonwage income categories

  - Use result as measure of "capital income"

- But, this nonwage residual is bad measure of

  - Business income: includes significant nonbusiness inc

  - Capital income: misses significant untaxed income

Main Take-aways

- Future survey recommendations:

  - Link responses to administrative data

  - Limit questions to verifiable queries

  - Ensure representative samples for all business types

- Current quantitative research:

  - Use tax and financial data directly