Household Incomes in Tax Data: Using Addresses to Move From Tax Unit to Household Income Distributions

Jeff Larrimore
Federal Reserve Board

Jacob Mortenson
Joint Committee on Taxation

David Splinter
Joint Committee on Taxation

The analysis and conclusions set forth in this presentation are those of the authors and do not indicate concurrence of the Federal Reserve Board, the Federal Reserve Banks, or their staff. This paper embodies work undertaken for the staff of the Joint Committee on Taxation, but as members of both parties and both houses of Congress comprise the Joint Committee on Taxation, this work should not be construed to represent the position of any member of the Committee.

Background

- IRS tax return data increasingly common for measuring income distributions
  - Income inequality and income mobility (Piketty and Saez, 2003; Chetty, Hendren, Kline, and Saez, 2014)
  - Tax liabilities (JCT 2012, Tax Policy Center 2017)

- Two major limitations
  - 15% of adults do not file a tax return
  - No links to others in household (Atkinson, Rainwater, Smeeding 1995; CBO 2016)

Goals of this paper

1. Overcome these two limitations of tax data
   - Incorporate non-filers
   - Combine tax records into households

2. Determine how the limitation of not observing households affects income inequality measurement

(Focus on 2010 to match Decennial Census)

Data: IRS Compliance Data Warehouse

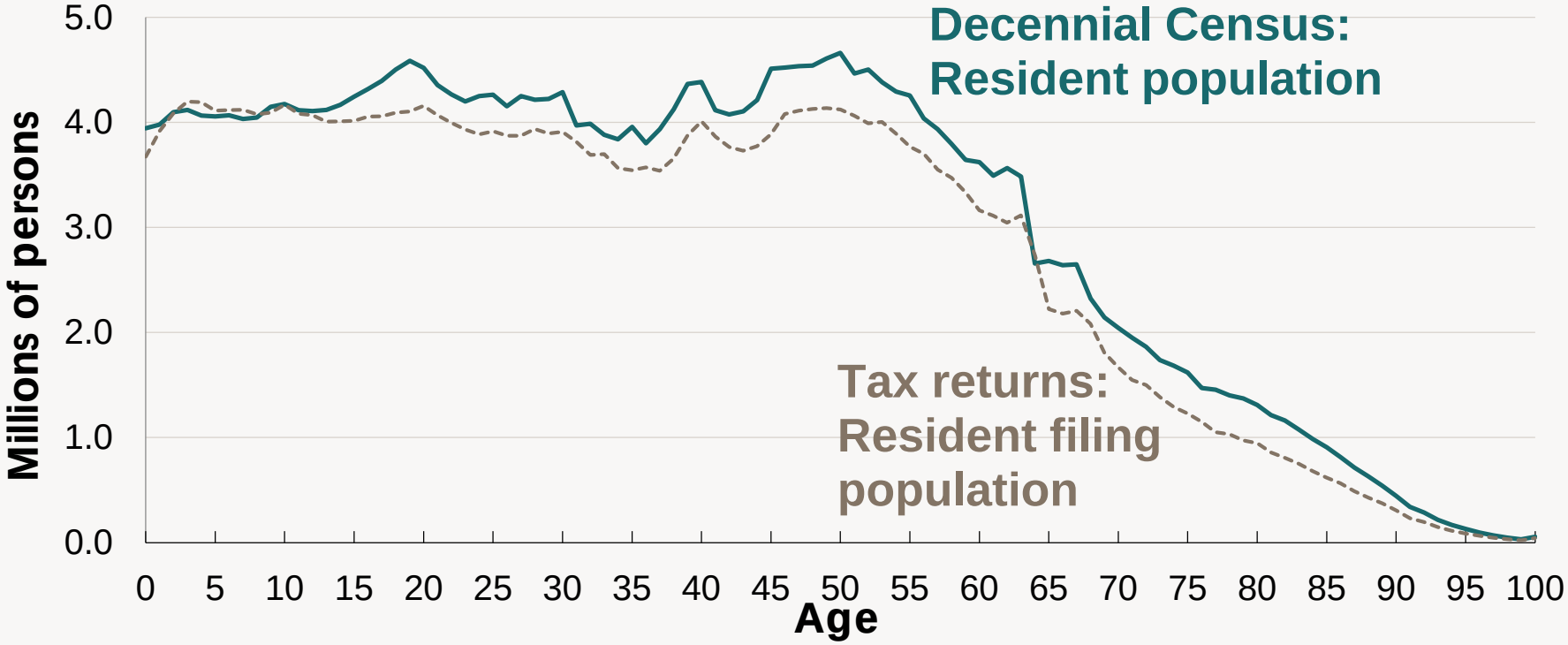
All individual tax forms received by the IRS

- Annual tax returns (e.g. Form 1040)
- Information returns
  - W-2
  - SSA-1099
  - 1099-INT, 1099-DIV, etc
  - 1099-Misc

Population coverage of tax data (2010)

Decennial Census: 308.7 million

IRS (resident filing population only): 281.3 million

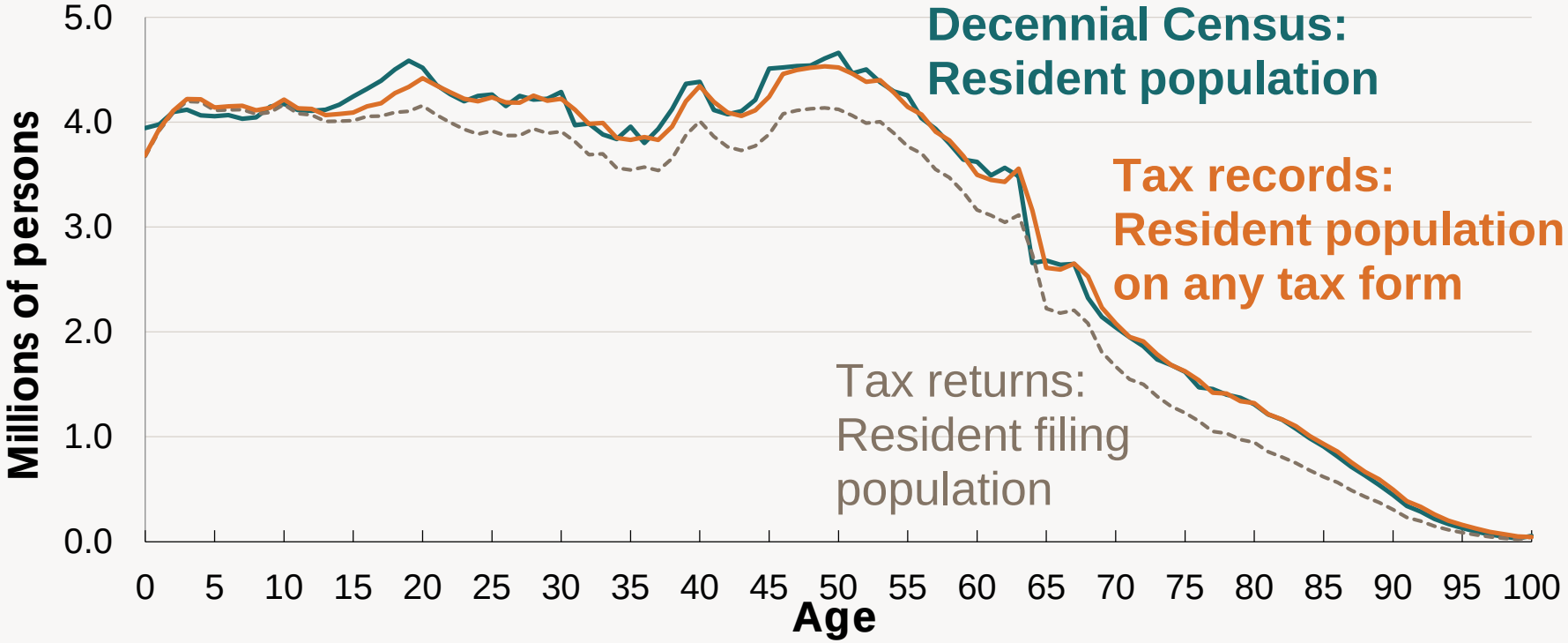


Population coverage of tax data (2010)

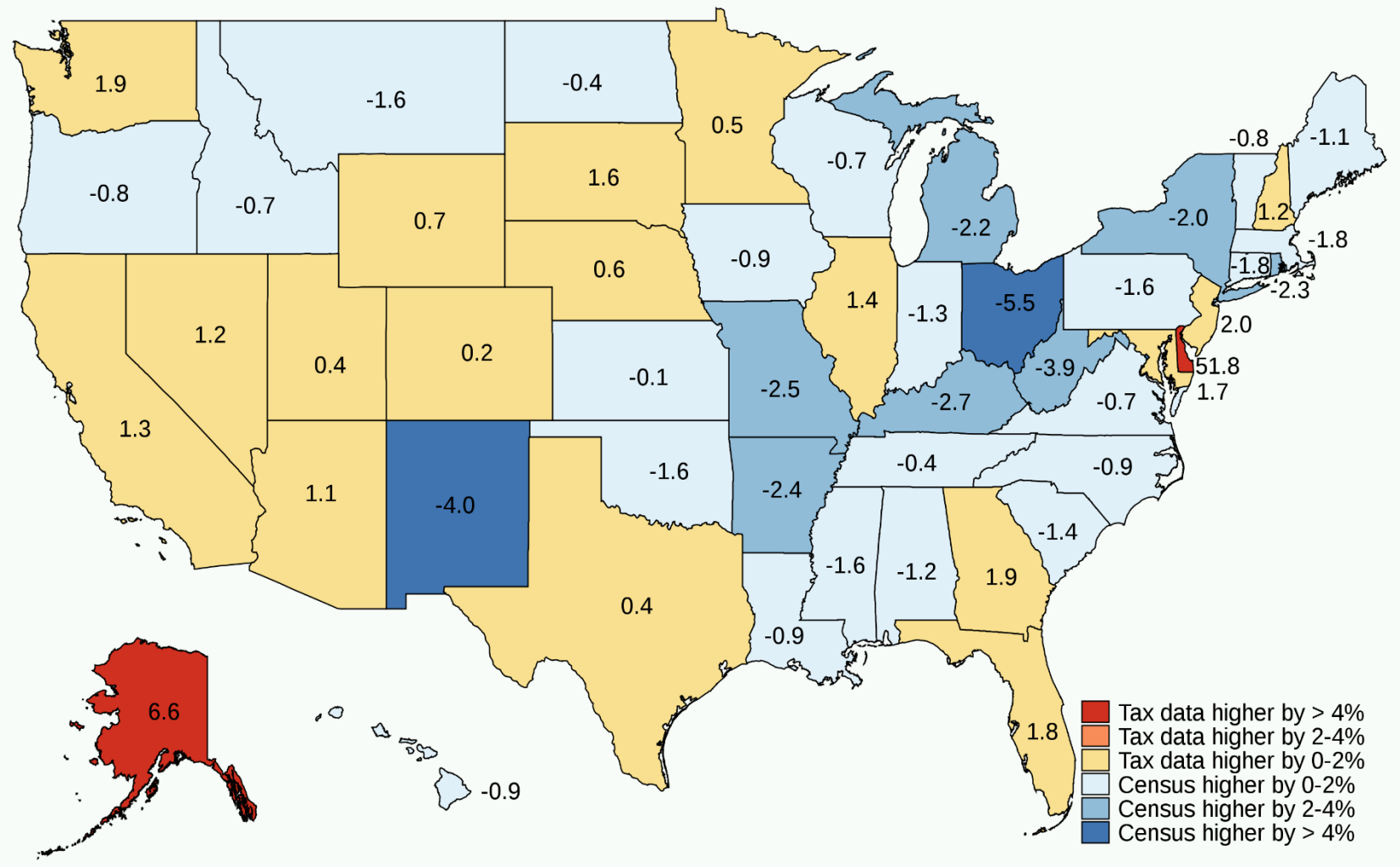
Decennial Census: 308.7 million

IRS (resident filing population only): 281.3 million

IRS (all residents): 307.9 million



State level population



Creating Households

All tax forms contain a mailing address

- Clean addresses to a uniform style
- Link all returns with the same mailing address and ZIP code

Creating Households (continued)

Check validity of unmatched 1-person addresses with a master list of valid street-ZIP code combinations

- Replace invalid street name with neighboring year if similar and valid
- Replace remaining invalid street names with most similar valid street (if any) in ZIP code

Similarity of text strings defined using Levenshtein distance method

- Count number of replacements, insertions, and deletions between text strings
- Example: [Suoth Street] to [South Street] = 2

Household Counts

March CPS: 117.5 million

Decennial Census: 116.7 million

Tax Data: 113.5 million

About 2 million households would be added if dependents with different addresses were counted independently rather than with claimant

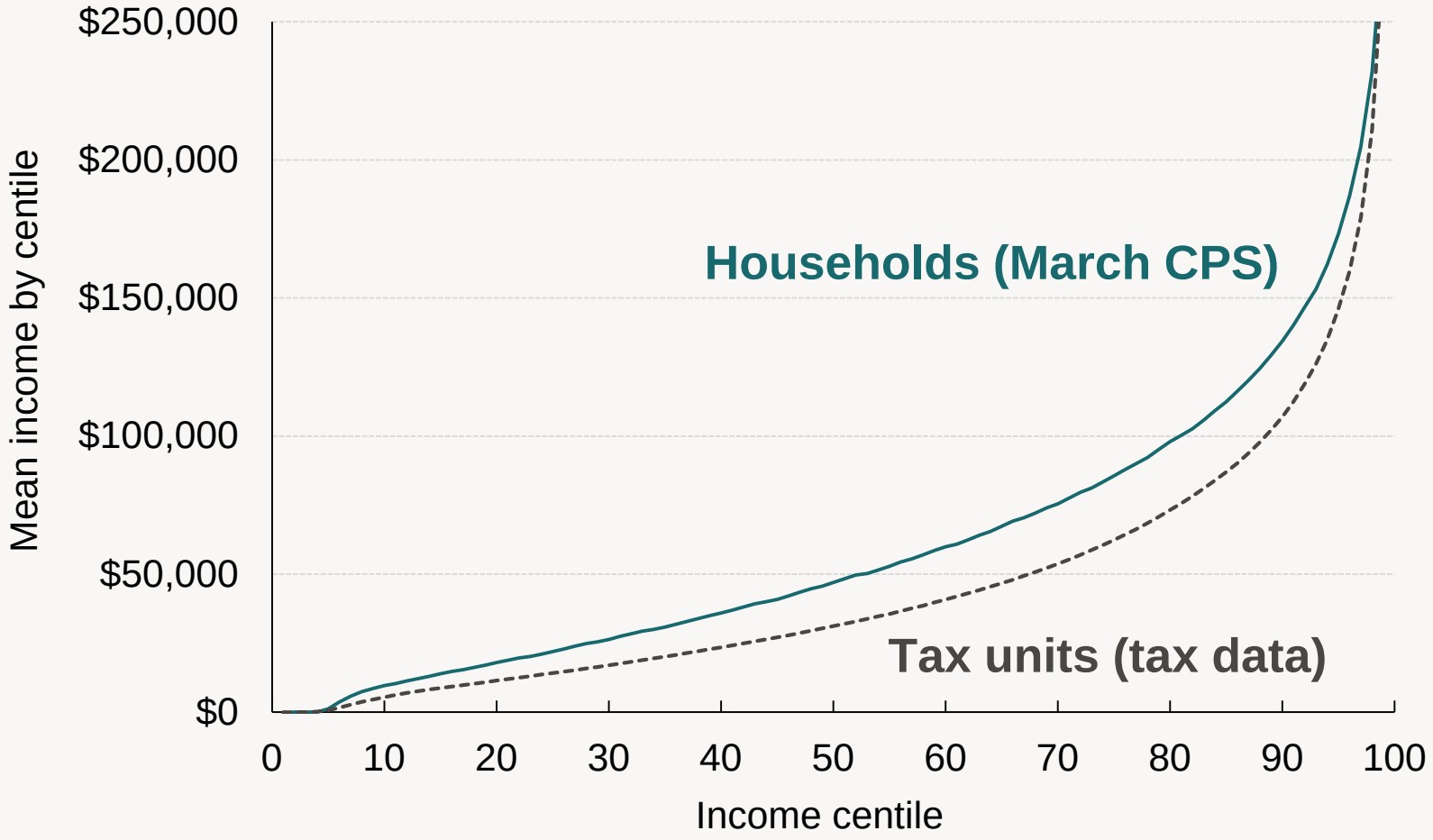
Households vs. Tax Units

Percent of households with each combination of filers and non-filers:

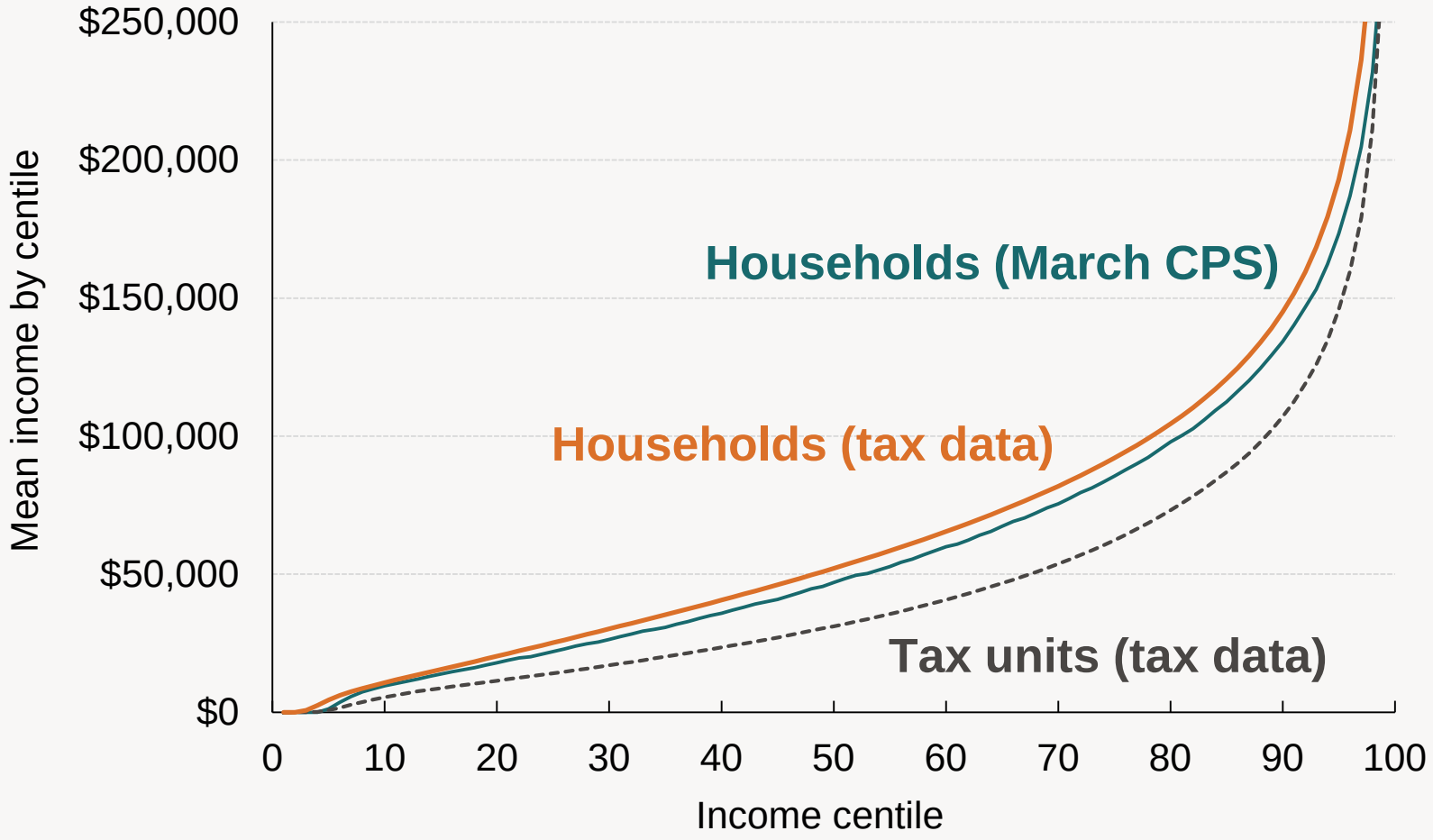
		Non-filing individuals		
		0	1	2+
Filing tax units	0	---	10%	2%
	1	59%	4%	1%
	2+	23%	2%	<1%

Note: dependent non-filers are included as part of the tax-unit who claimed them on their return

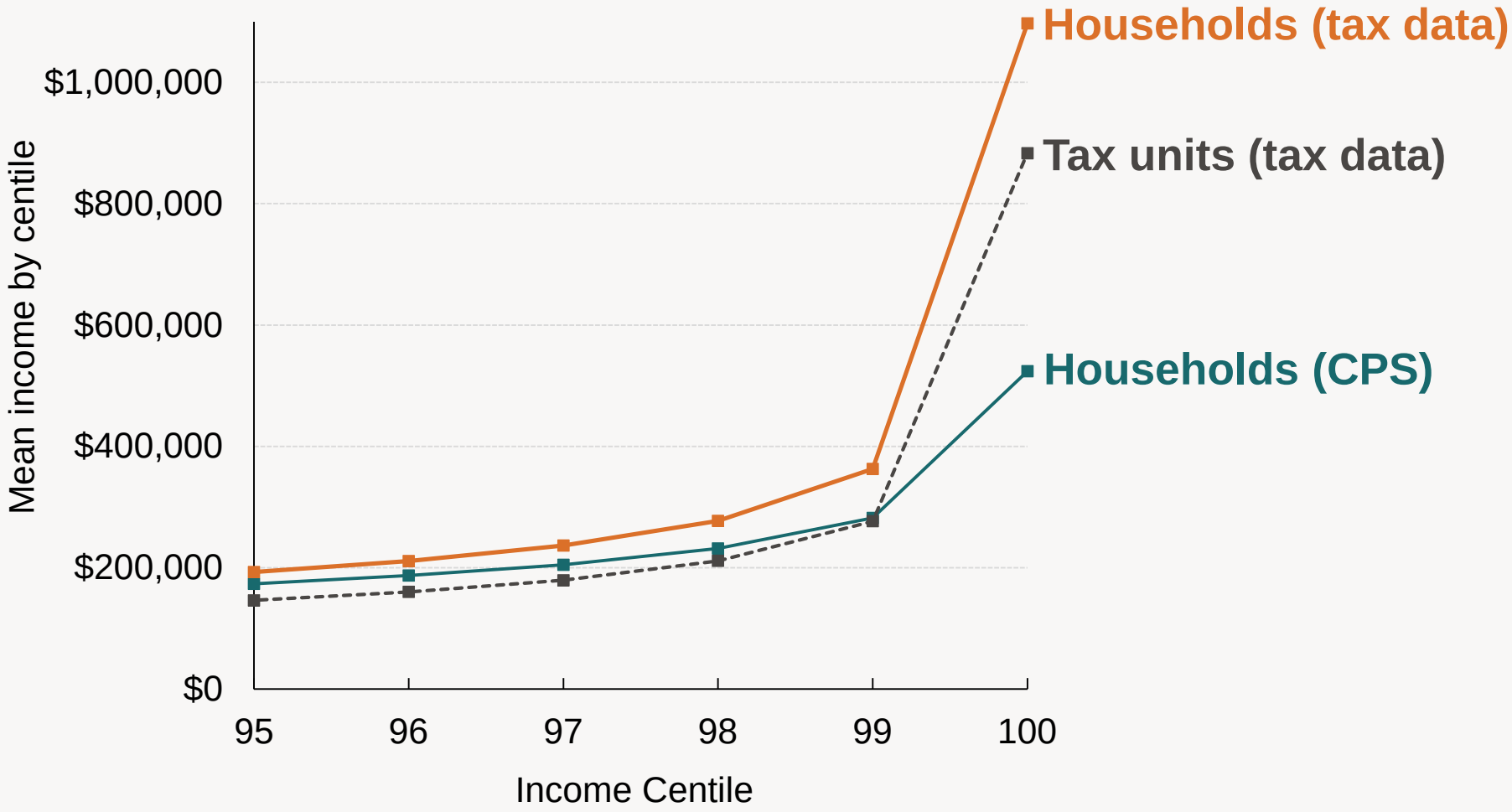
Income distributions



Income distributions



Top 5% distribution



Inequality Statistics

	Tax data (HH)	Tax data (TU)	March CPS (HH)	% difference using tax units	% difference using March CPS
Gini	0.516	0.570	0.483	+10%	-6%
P90/P10	13.2	18.8	13.7	+42%	+4%
Top 20% share	54.6	59.5	51.0	+9%	-7%
Top 5% share	27.9	31.4	21.8	+13%	-22%
Top 1% share	14.0	16.2	---	+16%	---

Conclusions

- Using universe of tax-record data, possible to:
  - Incorporate non-filers
  - Observe complete households

- CPS understates inequality (greatly understates incomes of the top 2%)

- Tax-units are *not* a sufficient proxy for households in tax data – assuming equivalence overstates inequality