Household Incomes in Tax Data: Using Addresses to Move from Tax Units to Households

Jeff Larrimore[†], Jacob Mortenson[*], and David Splinter[*]

† U.S. Federal Reserve Board

∗ Joint Committee on Taxation, U.S. Congress

April 2018

Disclaimer: This research 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. 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.

Tax Units and Households

Tax Units:

- Those appearing together on an annual tax return
- Non-filers – 15% of the adult population

157.5 million tax units in 2010

Households:

- Individuals residing together

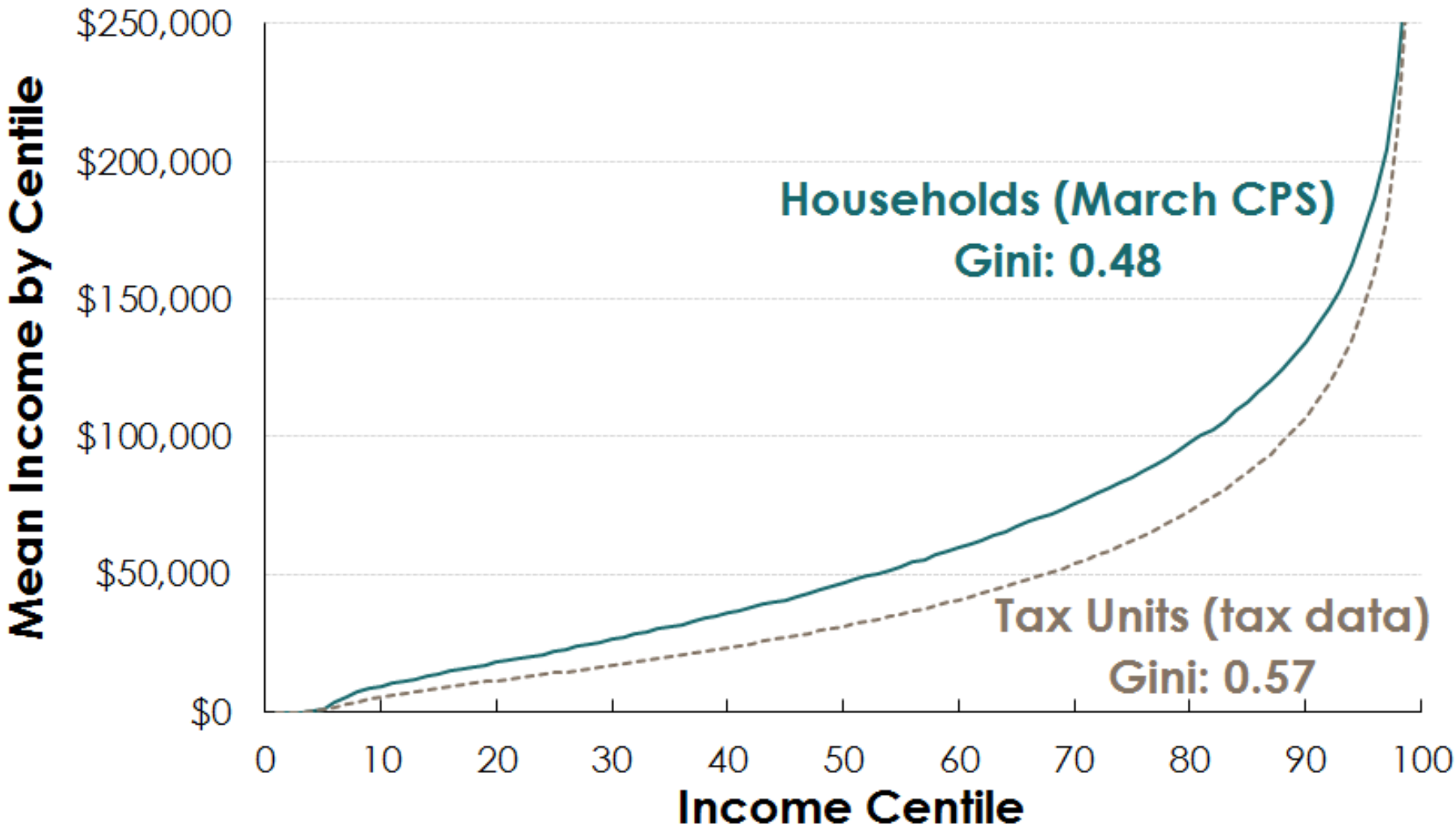
117.5 million households in March 2010 CPS

Tax Units and Households

Tax units often used as unit of analysis, or as proxy for household.

- Income inequality and income mobility
  - Piketty and Saez, 2003
  - Chetty, Hendren, Kline, and Saez, 2014
  - Chetty, Hendren, Jones, and Porter, 2018
- Tax liabilities
  - JCT
  - OTA
  - Tax Policy Center

Tax Units and Households in 2010



Creating Households from Tax Records

Data: Population of tax records in 2010

- Annual tax returns
- Information returns (e.g. W-2, 1099 series)

Process: aggregate by mailing address

1. Clean addresses.
   - "EIGHTH" to "8TH"
   - "MOUNTAIN" to "MTN"
2. Choose a single address for each individual.
3. Link all returns with the same street address and ZIP code.

Creating Households from Tax Records

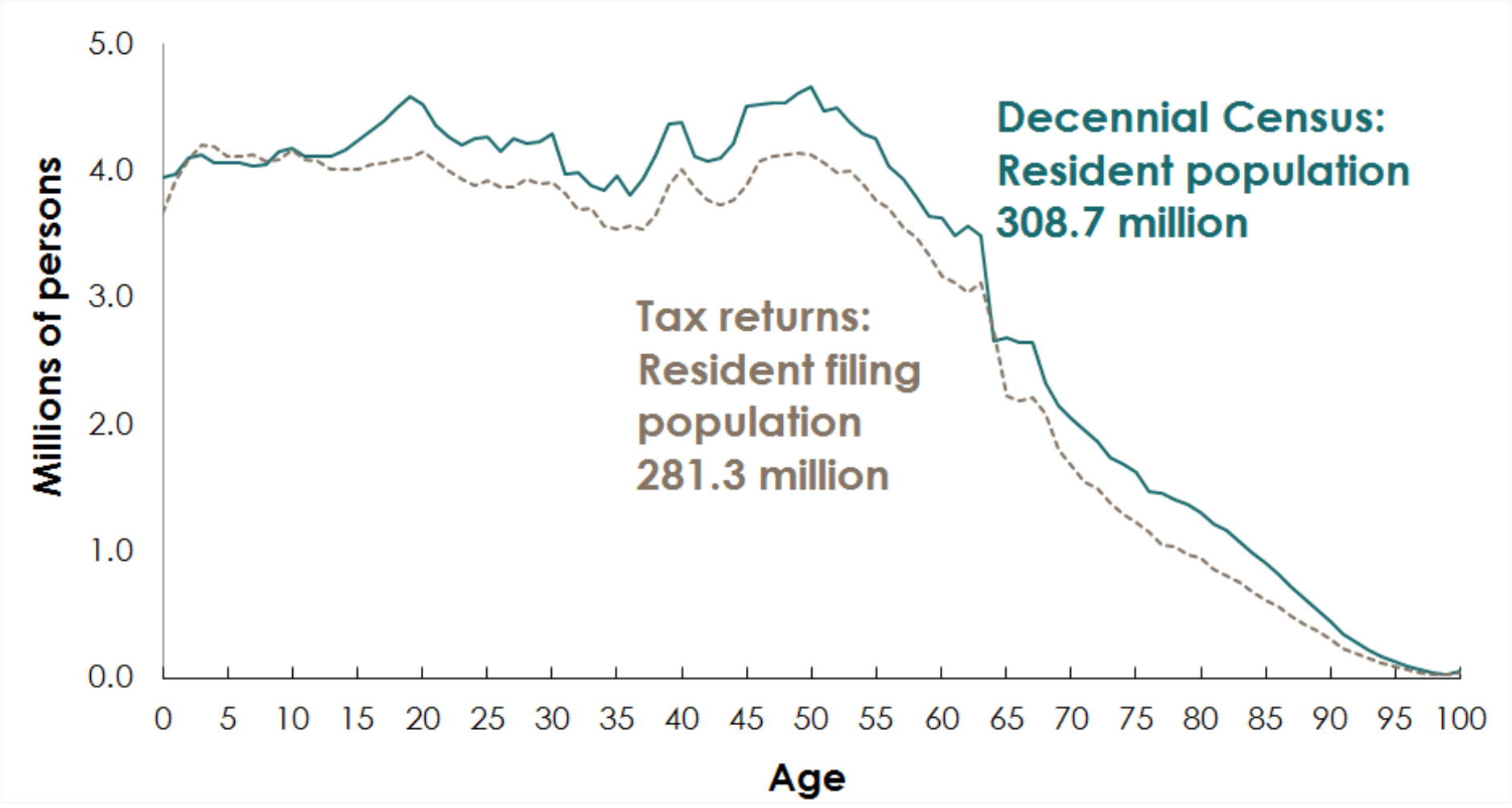
Unmatched 1-person households:

1. Check validity with an address master list.
2. Replace invalid street name with neighboring year if similar, valid.
3. Replace remaining invalid street names with most similar valid street (if any) in ZIP code [*Fuzzy Match*].

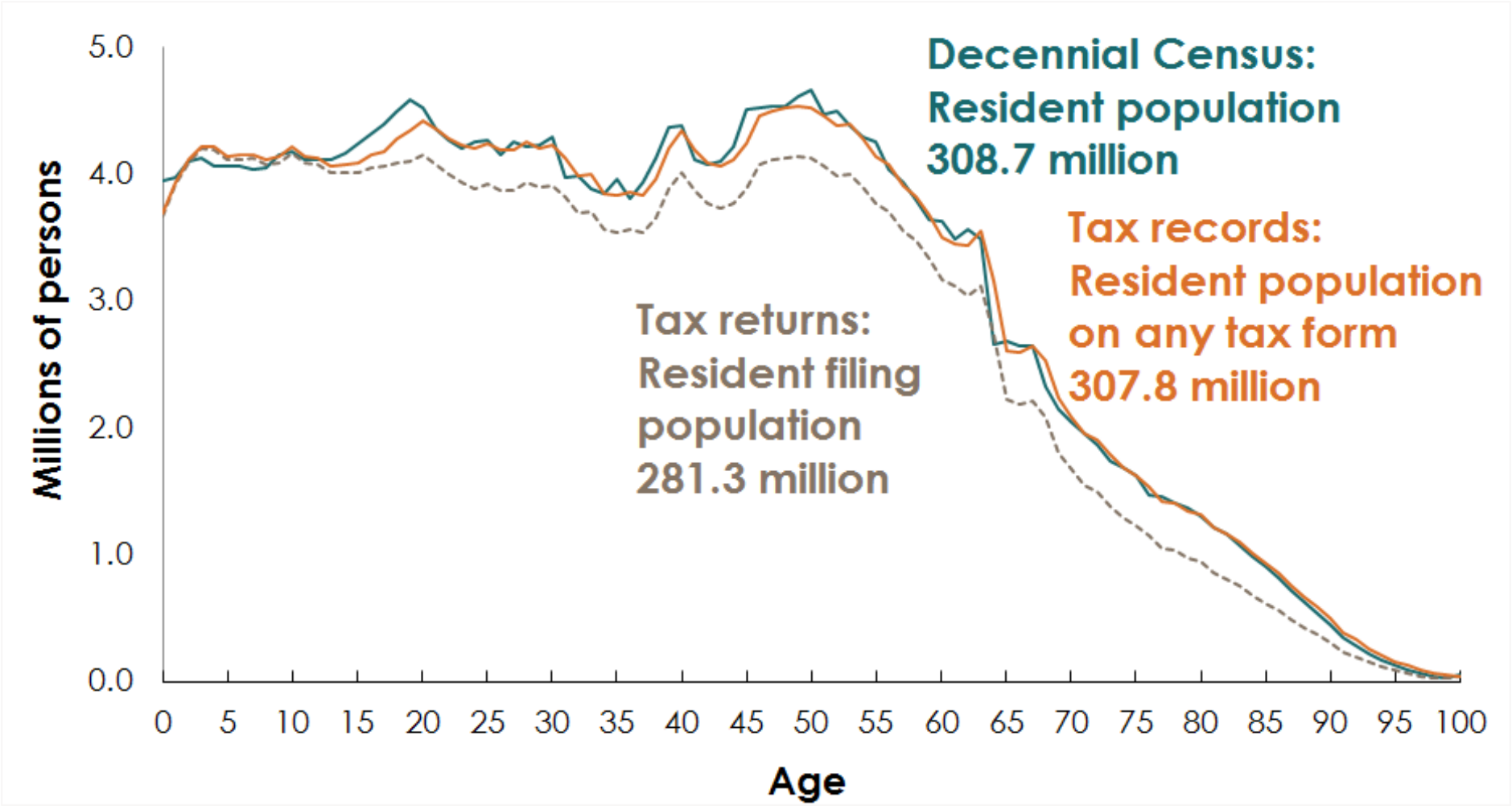
Fuzzy Match

- Similarity of strings defined using Levenshtein distance method
- Count replacements, insertions, and deletions between strings
- Example: [S**uo**th Street] to [S**ou**th Street] = 2

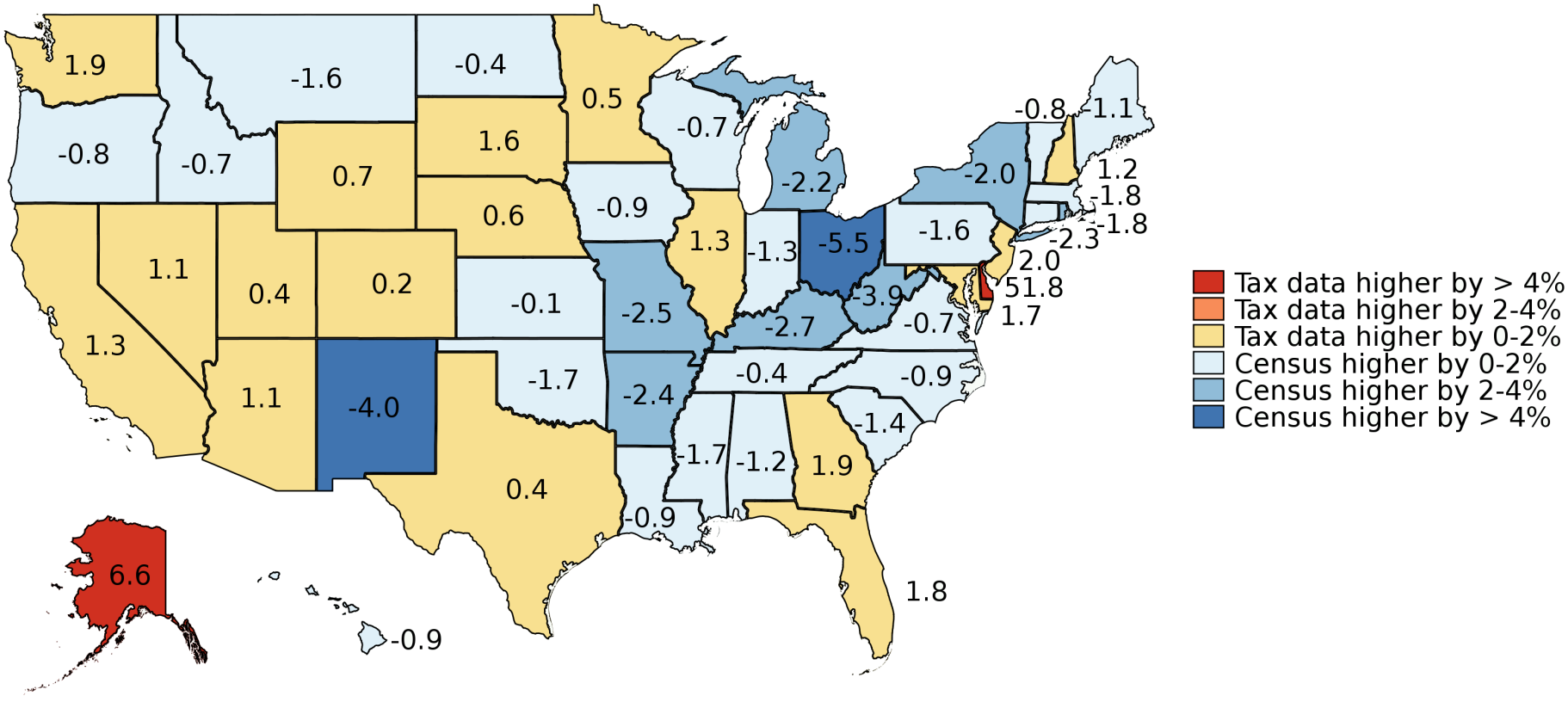
Age Distribution: Census v. Tax Filers



Age Distribution: Census v. Tax Population



Percent Difference in **Individuals** by State: Tax v. Census



Household Counts in 2010

March CPS:	117.5 million
Decennial Census:	116.7 million
American Community Survey:	114.5 million
Housing Vacancy Survey:	112.9 million
Tax Households:	**113.5 million**

Different treatment of young adult dependents away-from-home:

- Young adult dependents are counted in their parents' household
- Census bases on residence when surveyed

Results in around 1.9 million *fewer* tax households

Potential False Matches

Potential sources of false positives:

- Filing using accountant's address
- Group housing
- Missing apt/suite/number

Potential sources of false negatives:

- PO Boxes
- Missing apt/suite/number

Those with $>$10 household members are considered to be group quarters and dropped.

Distributions of Household Size

Household Size	Tax Units	Decennial Census	Tax Hholds	Tax v. Census
1	73,811	31,205	32,557	4%
2	43,017	38,243	32,787	-14%
3	18,184	18,758	18,359	-2%
4	14,259	15,625	15,587	0%
5	5,741	7,538	7,768	3%
6	1,752	3,075	3,675	19%
7 or more	752	2,272	2,799	23%
Total	157,515	116,716	113,532	-3%

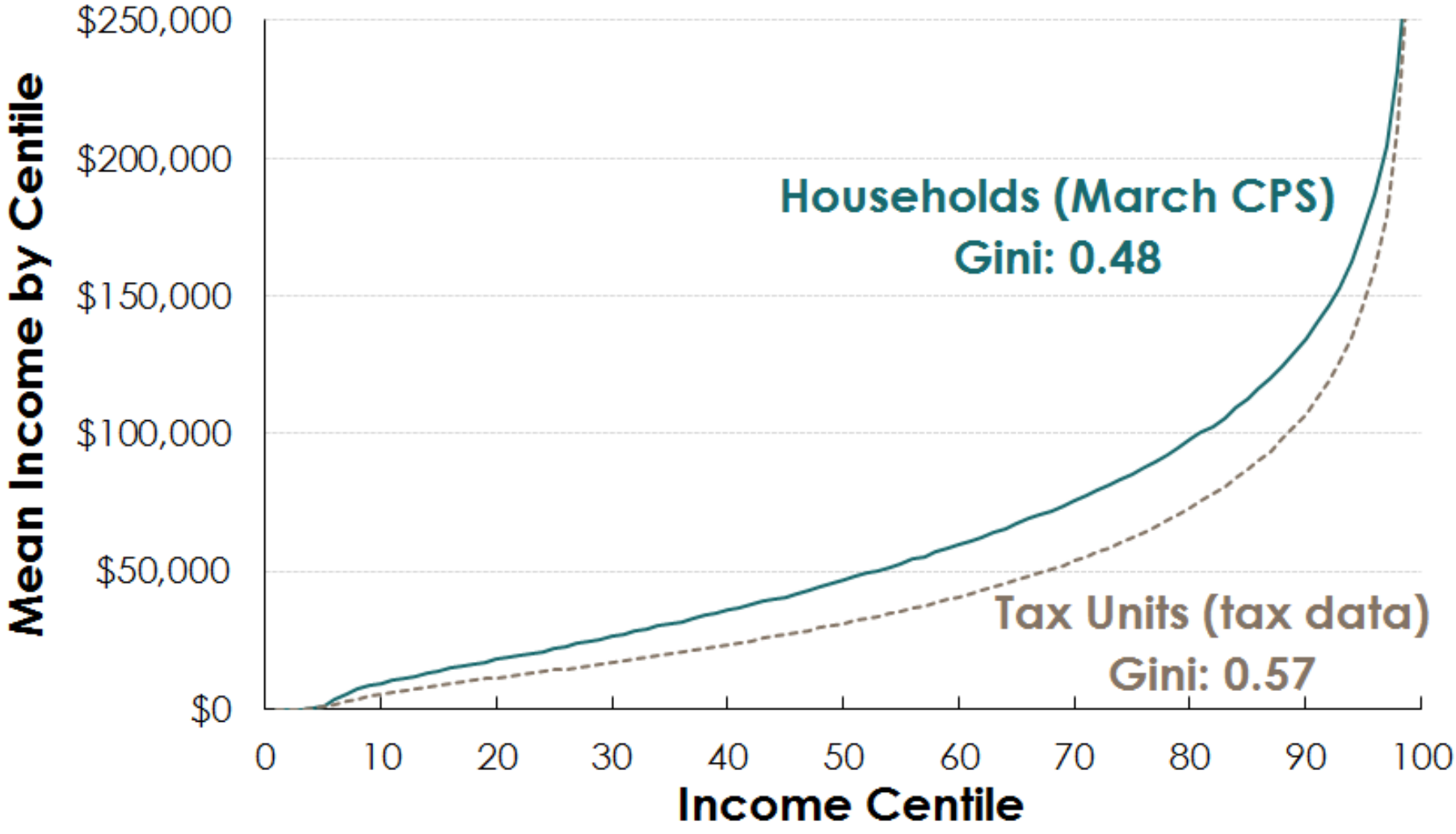
Tax Unit Types Comprising Households

		Non-filing individuals		
		0	1	2+
Filing Tax Units	0	–	**10%**	2%
	1	**59%**	4%	1%
	2+	23%	2%	<1%

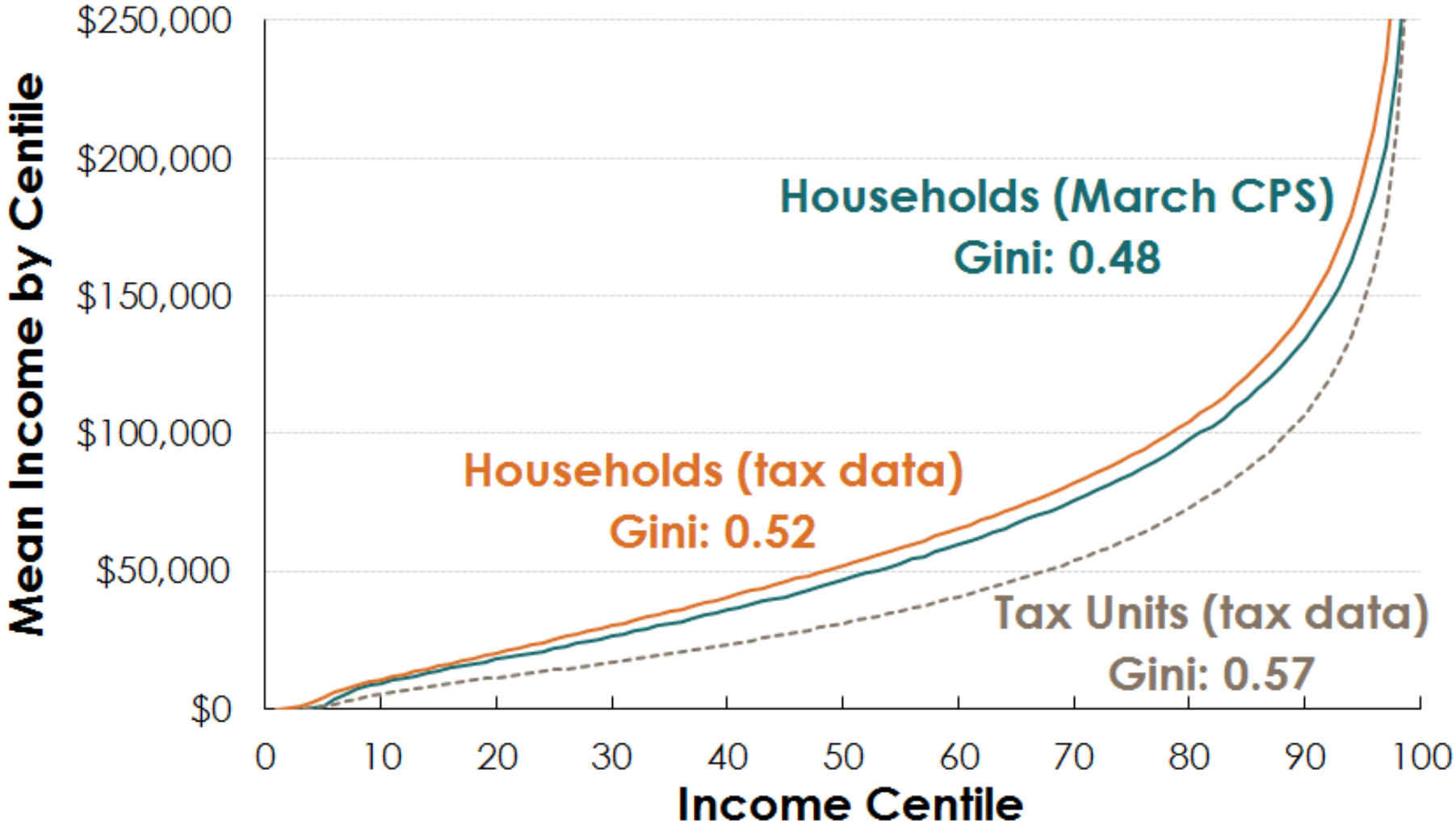
Income Distributions

Income Definition: Pre-tax, excluding capital gains

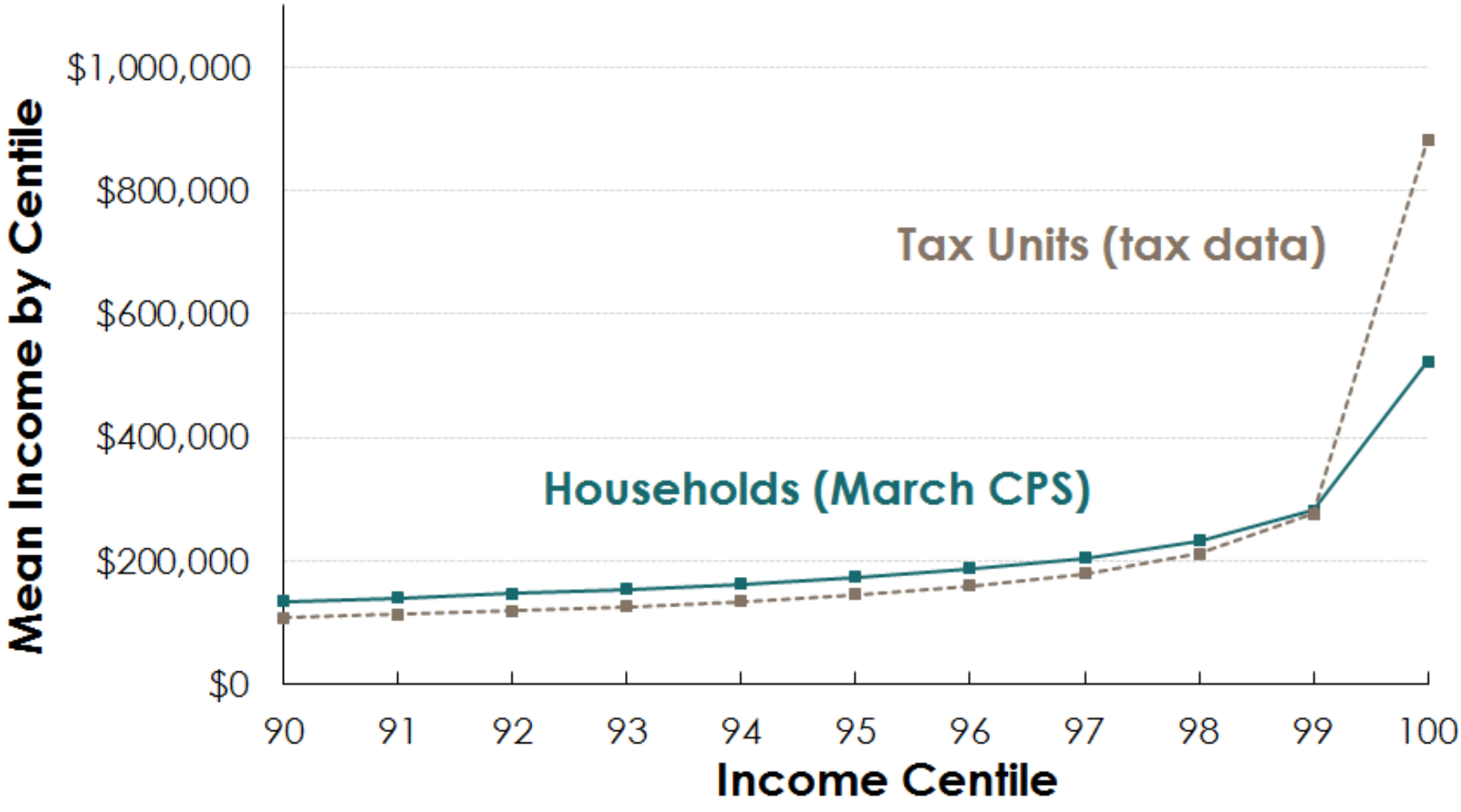
Income Distributions



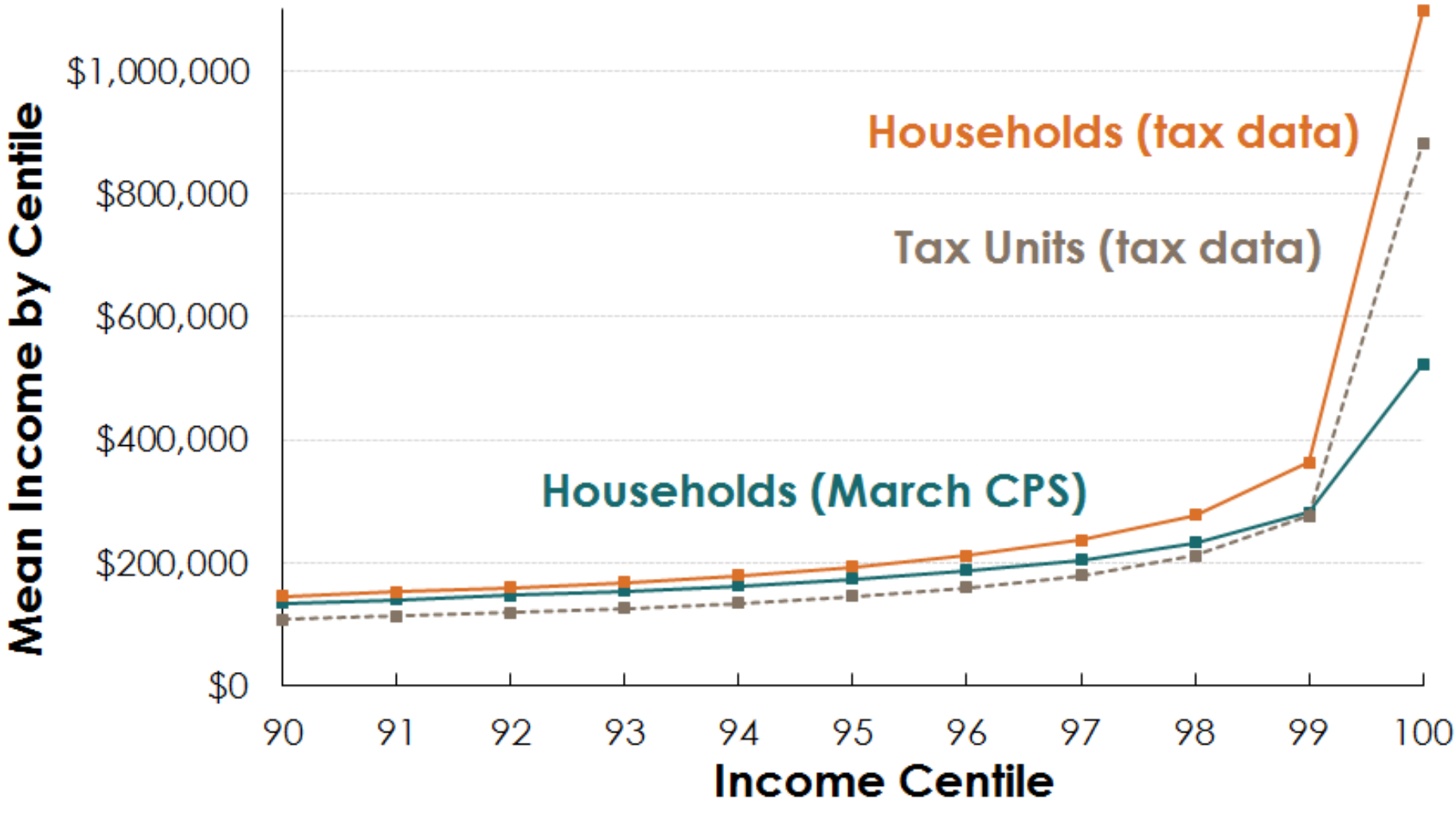
Income Distributions



Top 10% Distribution



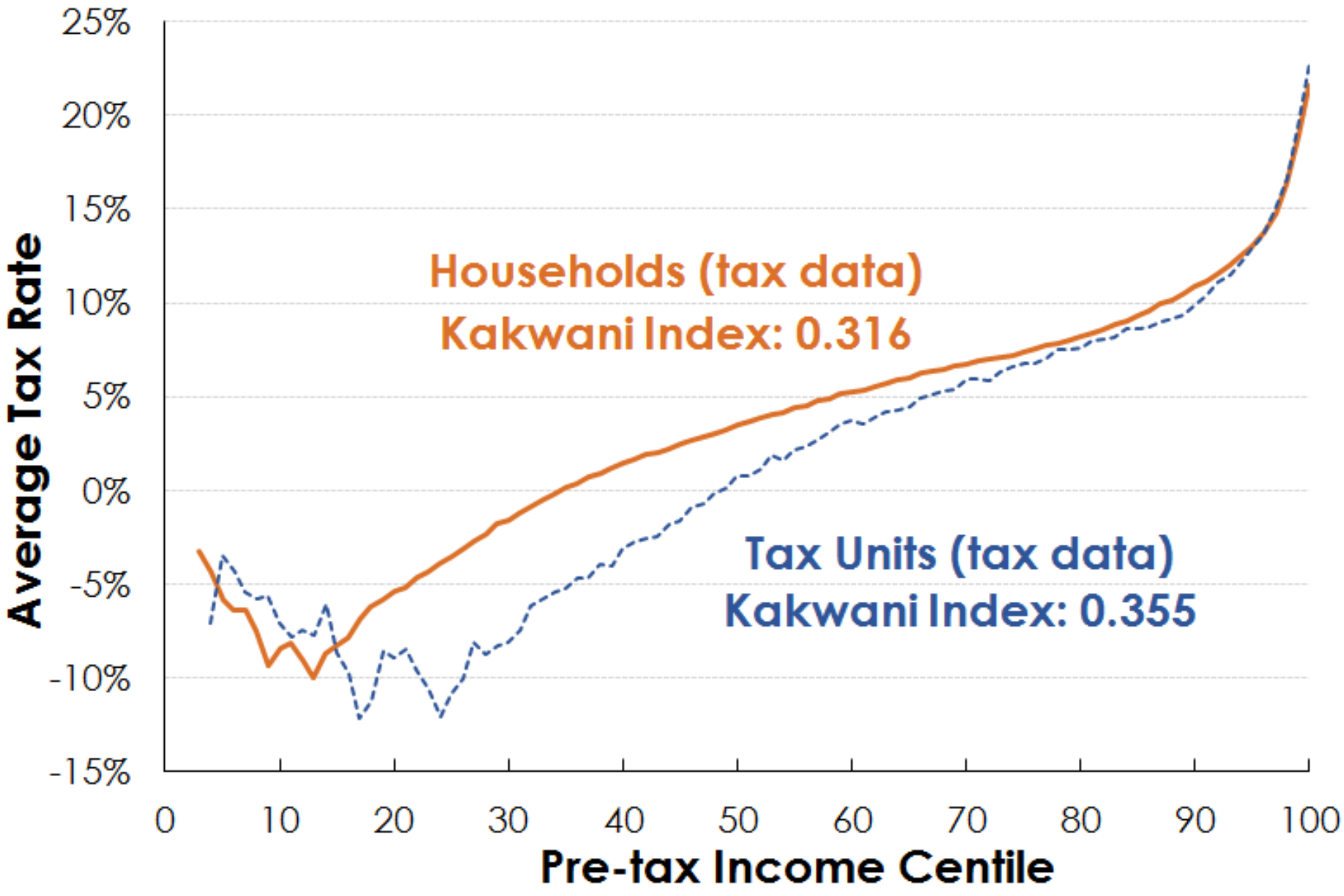
Top 10% Distribution



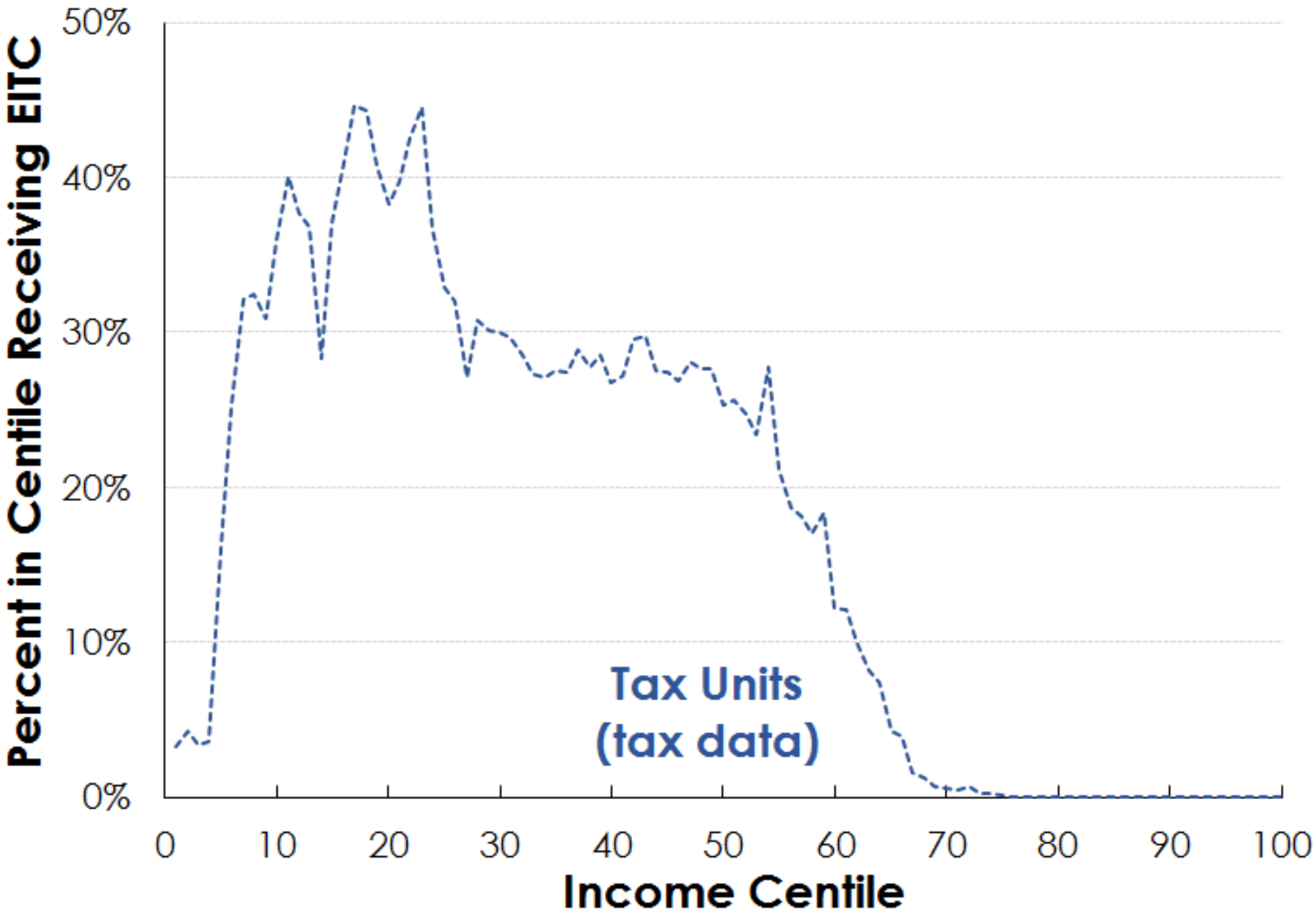
Inequality Statistics

				Relative to Tax Households...	
	Tax Households	Tax Units	March CPS	Tax Units	March CPS
Gini	0.52	0.57	0.48	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%	—

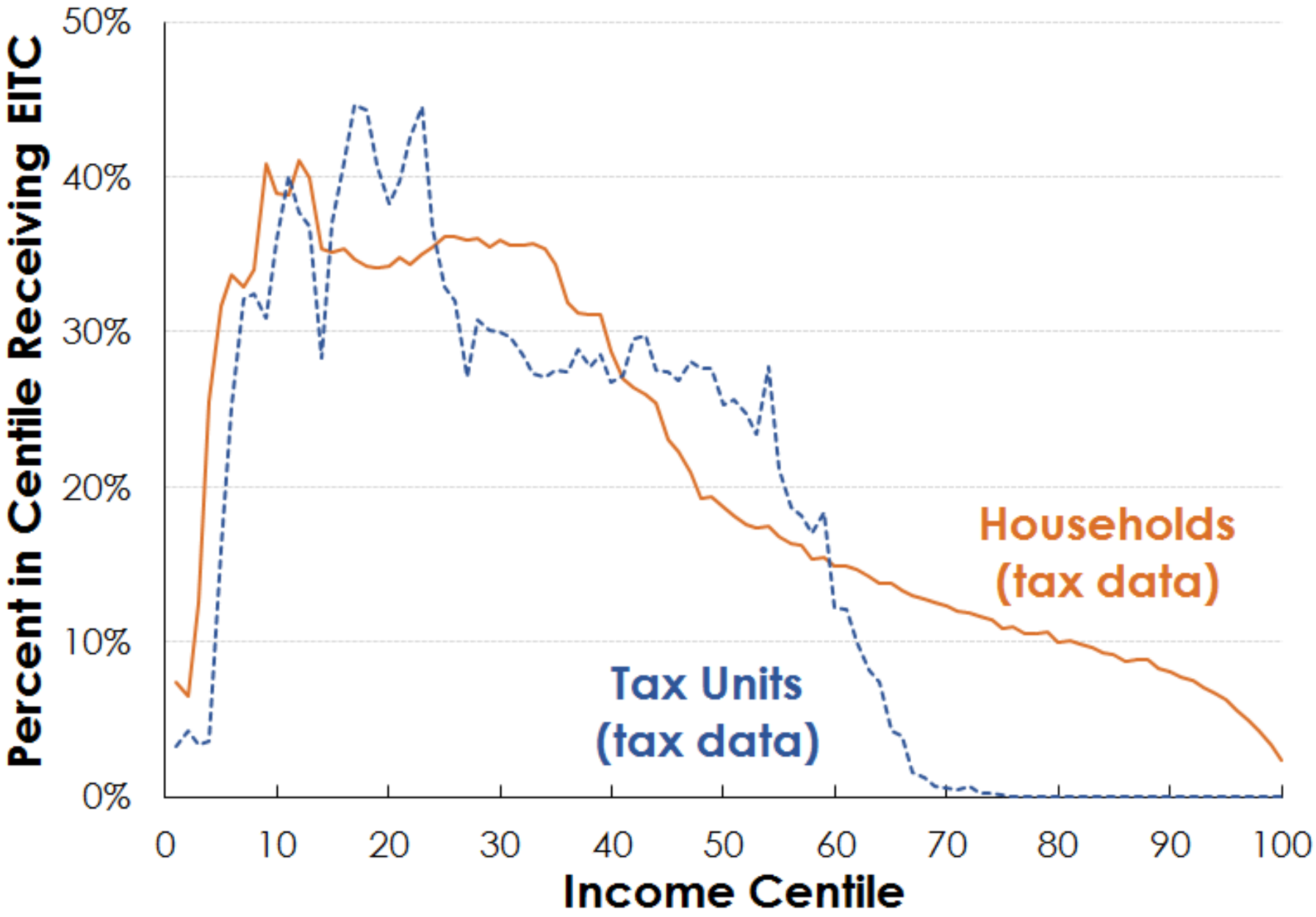
Tax Progressivity



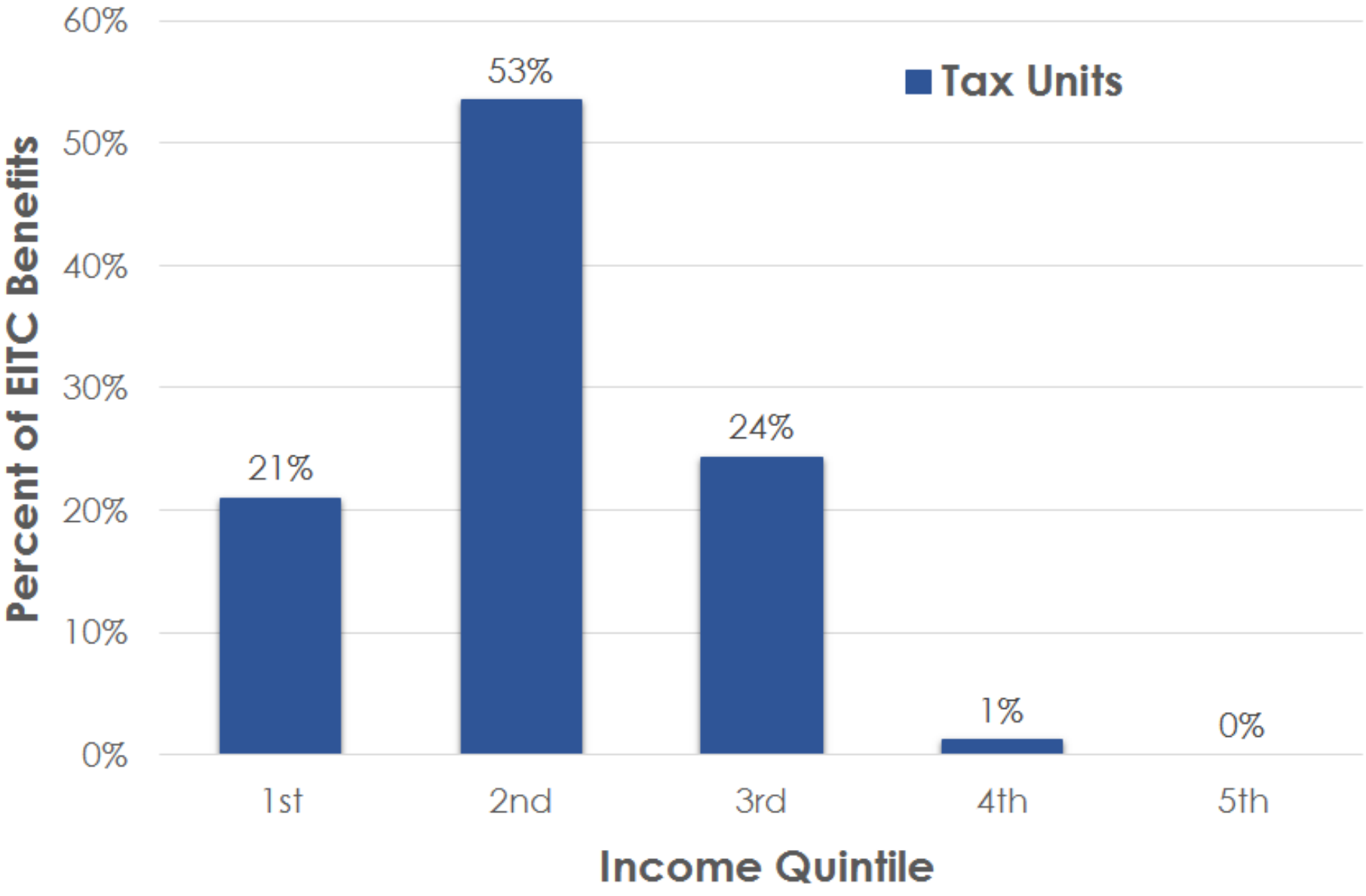
EITC Recipiency



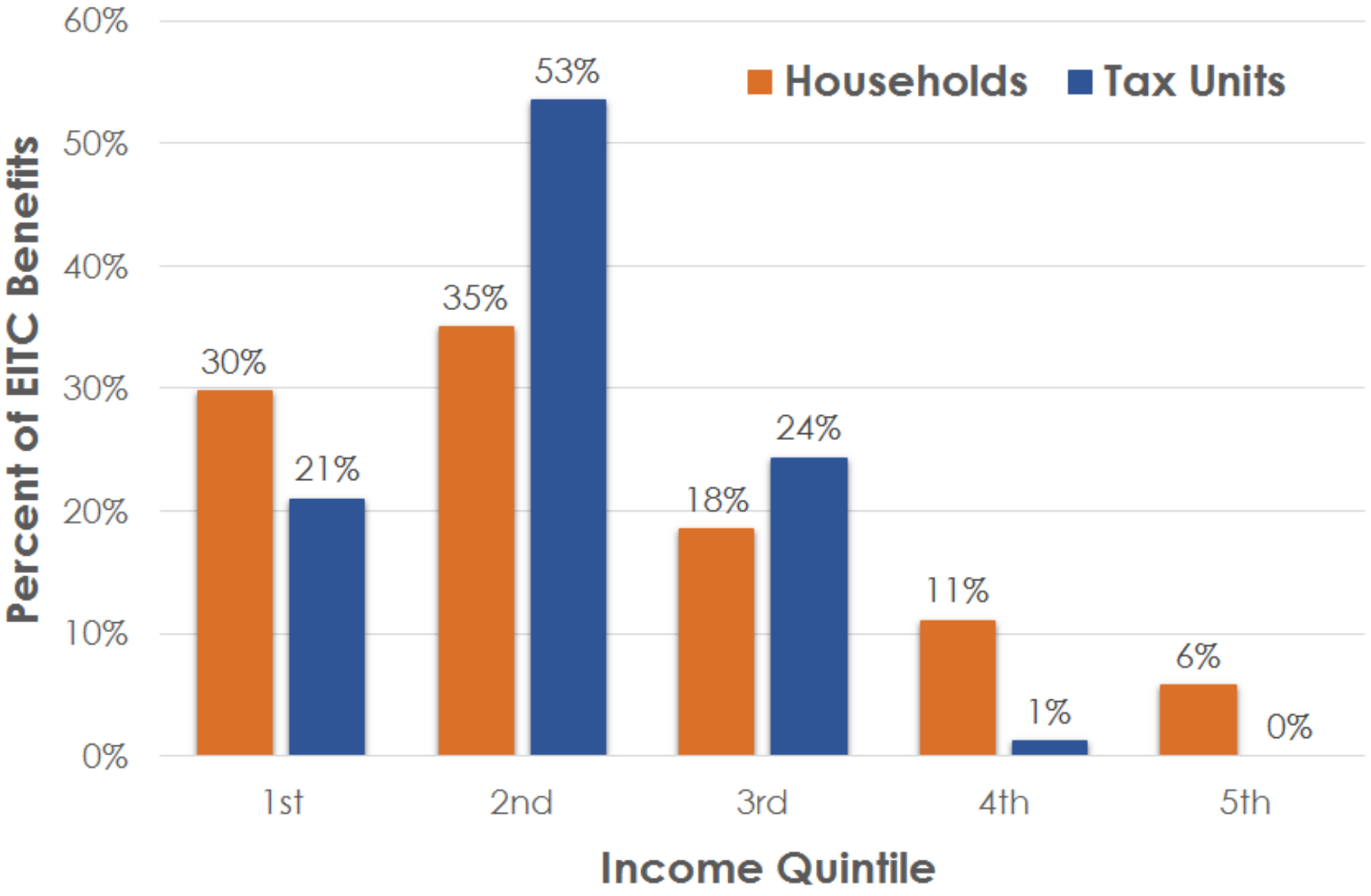
EITC Recipiency



EITC Benefits



EITC Benefits



Discussion

- CPS *understates* household inequality
- Using tax units as a proxy *overstates* inequality
- Progressivity is weaker when using households
- EITC is less well targeted than previously understood

APPENDIX

Income Definition

Pre-tax income excluding capital gains.

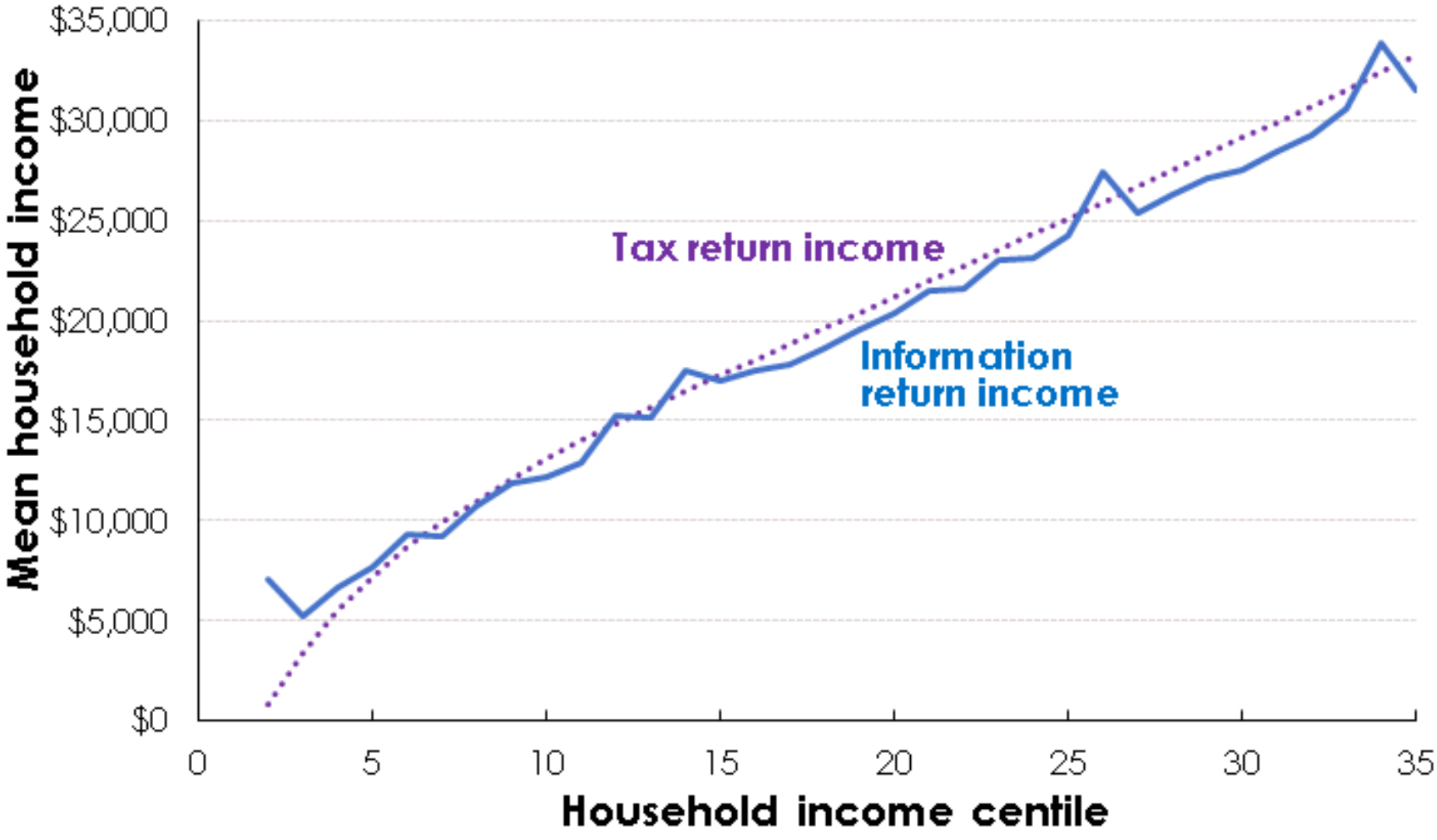
Tax filers:

- Total income + Tax exempt interest - Capital gains realizations

Non-filers:

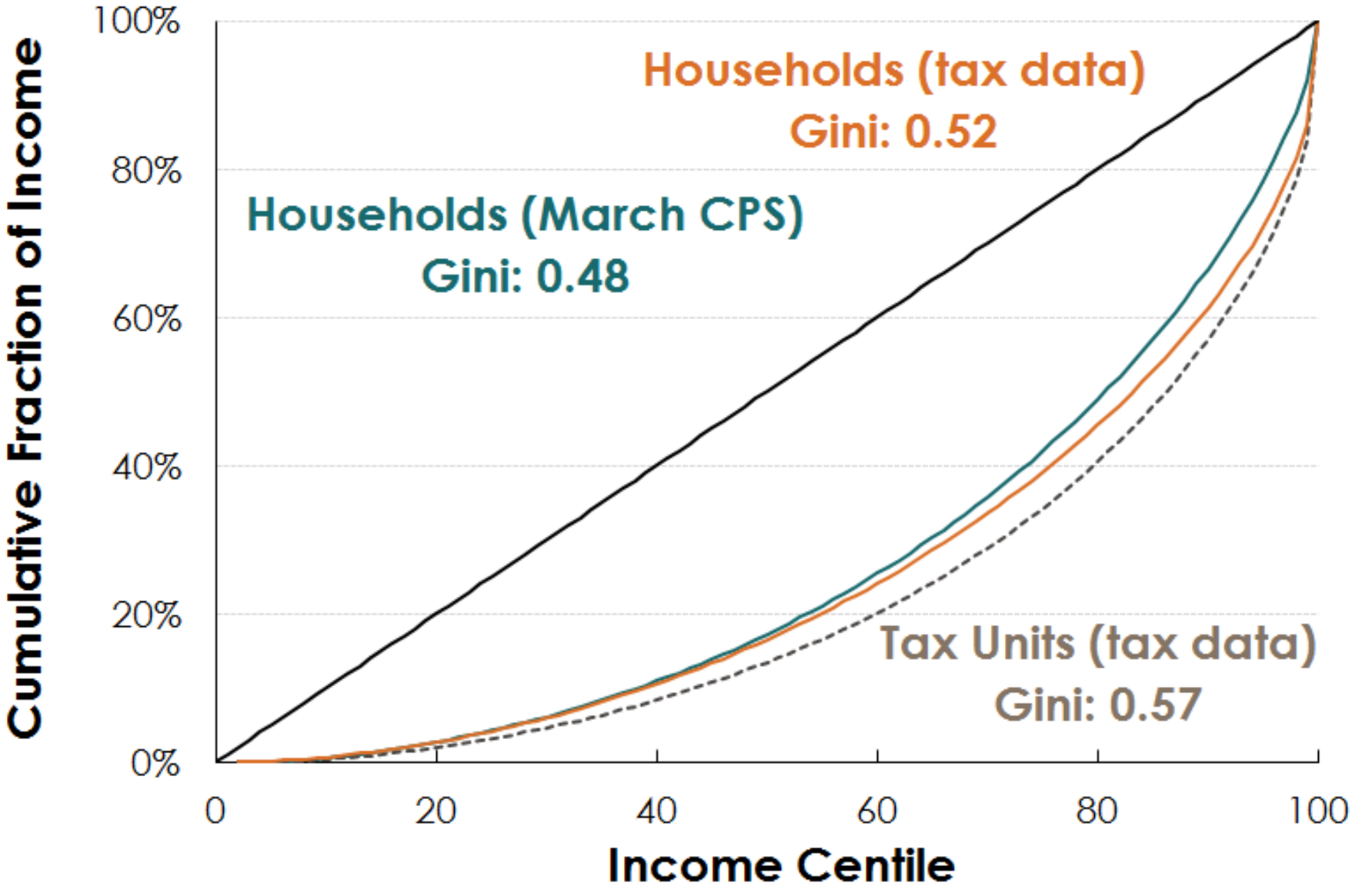
- Wages (Form W-2) and Unemp. Comp. (Form 1099-G)
- SSA and DI (Form 1099-SSA)
- Interest income (Form 1099-INT) and dividends (Form 1099-DIV)
- Gross private retirement income (Forms 5498 and 1099-R)
- 30% of 1099-MISC

Tax Return Income and Information Return Income

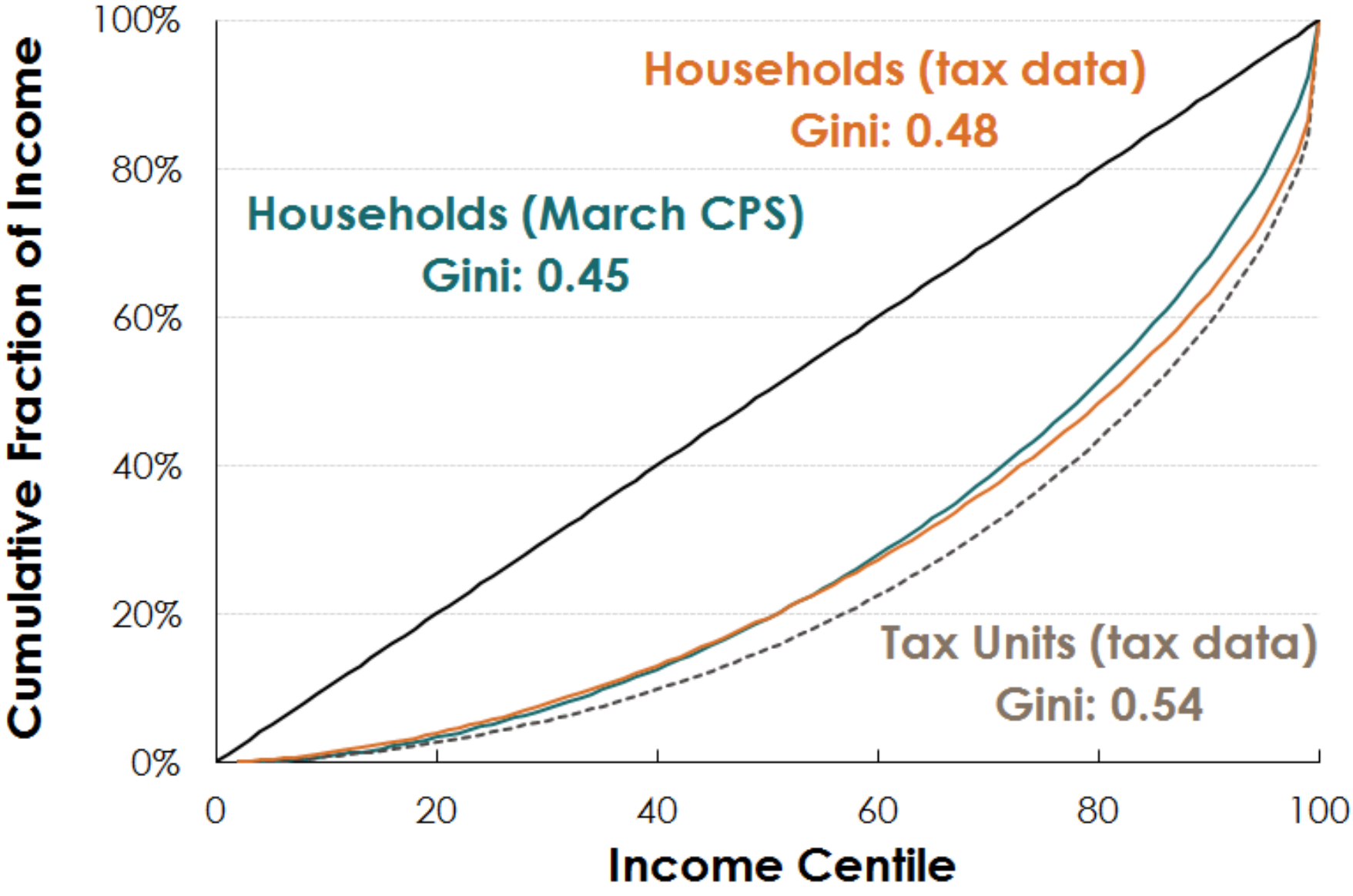


Back

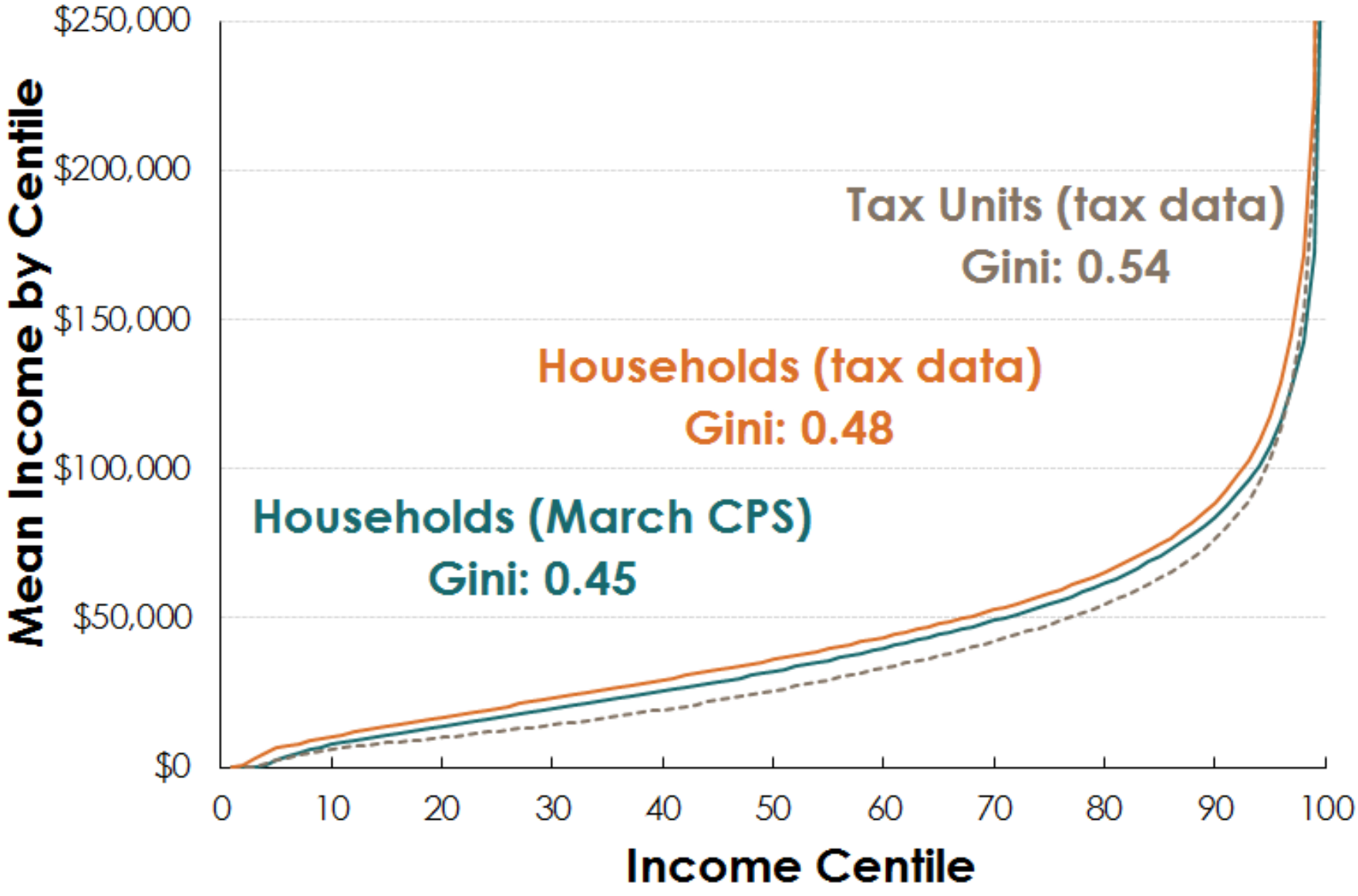
Lorenz Curve



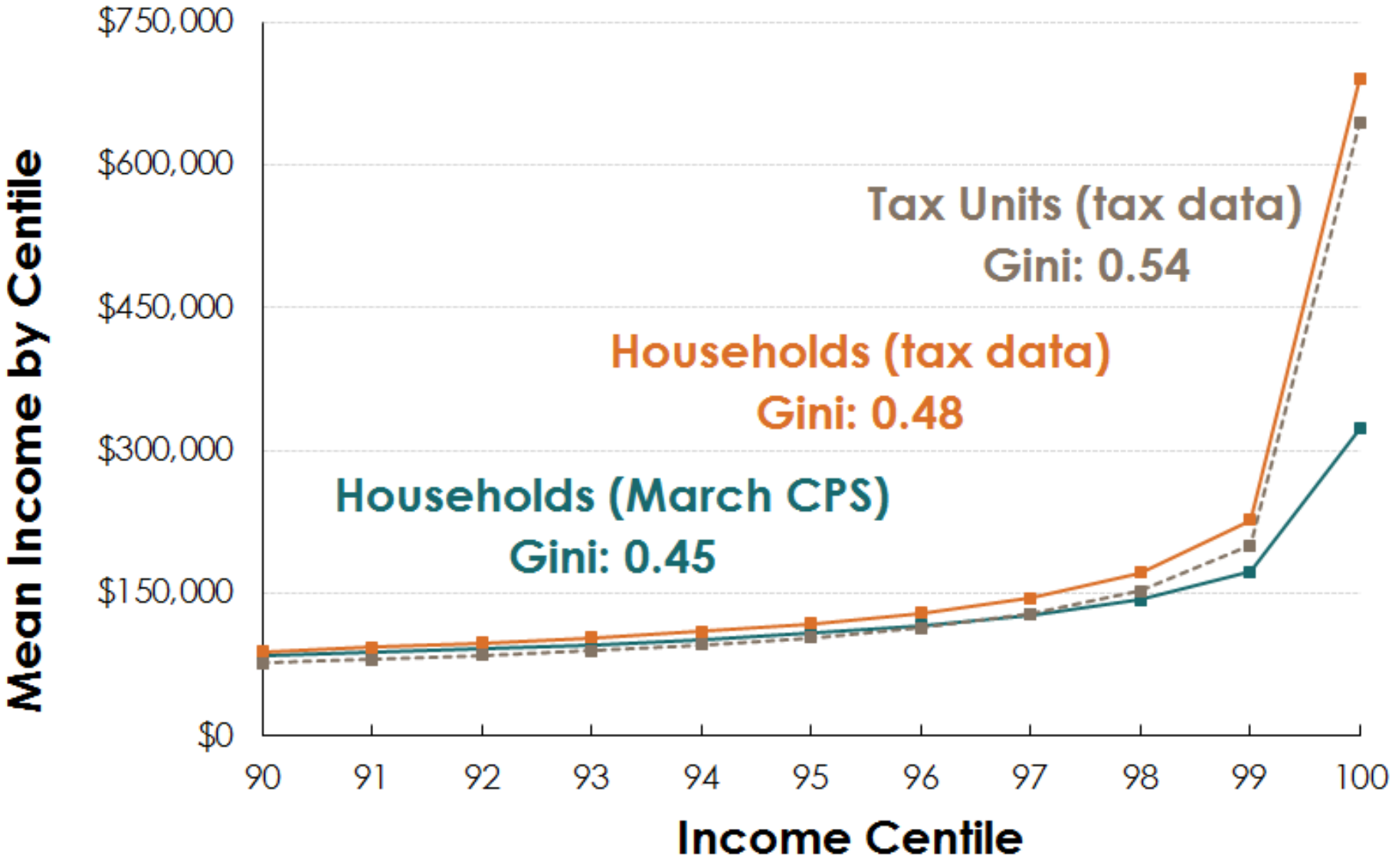
Size-Adjusted Lorenz Curve



Size-Adjusted Income Distributions



Size-Adjusted Top 10% Distribution



Percent Difference in **Households** by State: Tax v. Census

