Strengths & Limitations of Tax Data for Distributional Estimates

1. **Population in Tax Data**
2. **Income in Tax Data**
3. **Expanded Income, Transfers, & Taxes**

Jeff Larrimore
Jacob Mortenson
David Splinter

July 12, 2022
National Academy of Science CNSTAT Meeting

The Population in Tax Data

Jeff Larrimore

July 2022

The views in this presentation reflect those of the authors and should not be attributed to the Federal Reserve Board, the Federal Reserve System or other members of the research staff.

Key points I'll cover

1. Types of tax data and who is included/excluded from each

2. Linking records into households (and comparisons to Census data)

3. Demographic information in tax data

Multiple types of tax data

- Based on IRS Form 1040 (or similar) filings
  - SOI individual tax files (Bryant, Czajka, Ivsin, and Nunns 2014)

- Based on broader array of tax records
  - Population tax file (Larrimore, Mortenson, and Splinter 2021)
  - IRS databank (Chetty, Friedman, Saez, and Yagan 2018)
  - Census merge-file with select line-items (Bee and Rothbaum 2019)
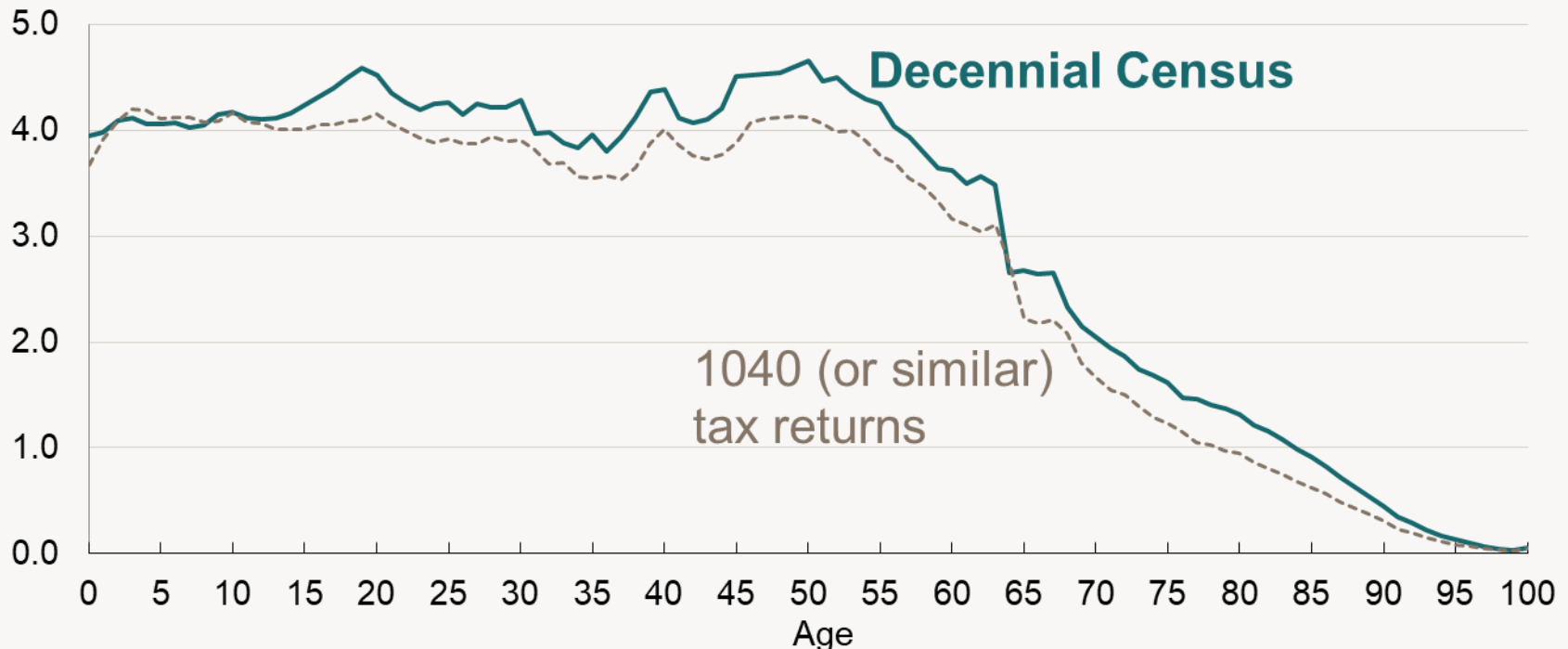
Who are in "1040 filings" data?

- People who file a tax return, or are claimed as a dependent on someone else's return

- **Reasons people excluded from data**
  - Non-compliance
  - Low-incomes (details from IRS Publication 501)
    - In 2021, most have no filing requirement if married under age 65 with gross income <$25,100
    - Gross-income generally excludes Social Security
    - However, must file if self-employment income >$400
  - Children whose parents have no incentive to file or claim as a dependent
  - Undocumented immigrants (although 2/3 appeared prior to TCJA and about 1/2 after TCJA)

Population coverage of 1040s (2010)

Decennial Census: 308.7 million

IRS (1040 tax returns): 281.3 million (91%[1])

[1]The share is somewhat lower if focused only on adults. See e.g. Molloy, Smith, and Wozniak (2011)

Who are in "population file" data?

- In the 1040 filing data (filers, spouses, & dependents) *or* has third-party reporting such as:
    - Wages (W-2)
    - Some self-employment (1099-MISC / 1099-NEC)
    - Social Security (1099-SSA)
    - Interest or dividends (1099-INT / 1099-DIV)
    - Retirement withdrawals (1099-R)
    - Unemployment income (1099-G)
    - Business income (K1s from partnership or S-corp)
    - Higher education (1099-8 and 1098-T)

- **Reasons people excluded from data**
    - No income (or other activity) with reporting requirements
    - Only under-the-table earnings or from non-taxable sources

Population coverage of tax data (2010)

Decennial Census: 308.7 million

IRS (1040 tax returns only): 281.3 million

IRS (all residents): 306 million+ (about 99%)
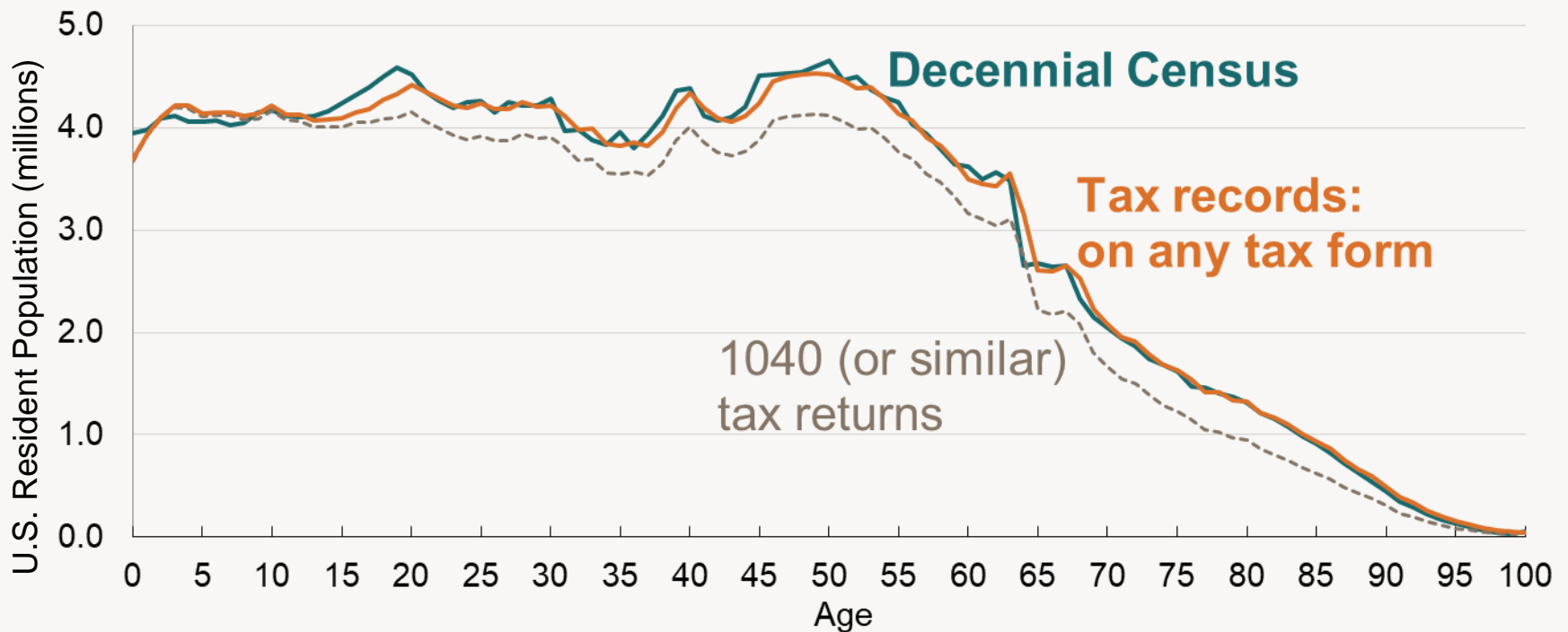
Best practices

Recommendations

- Use 1040 data when available

- Supplement with information returns to observe non-filers (adjusting for unobserved self-employment expenses)
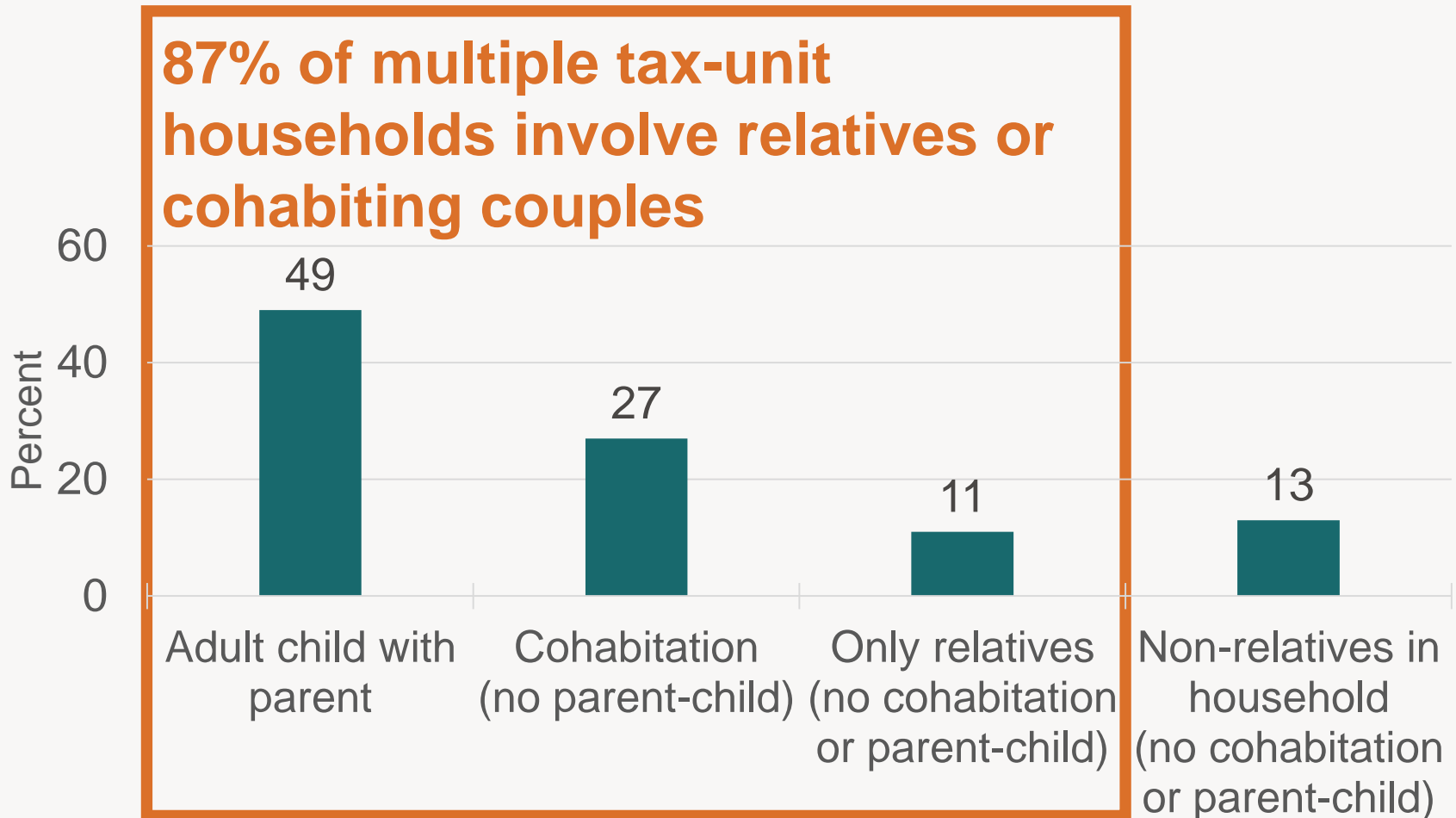
Reasons

- 1040 data validated by taxpayer and contains self-reported information like business deductions

  - Best quality tax data for those who file

- But 1040 data misses many low-income individuals

Creating sharing units

- Tax Units ≠ Households
  - About 35-40% more *Tax Units* than *Households*

- Filers: only observe those who file a tax return together, not complete households (or Census families)
- Non-filers: observe no family/household links

- **However, all tax forms contain a mailing address**
  - Can use address to link households (Larrimore, Mortenson, and Splinter 2021)

Relationships in multi-TU households

87% of multiple tax-unit households involve relatives or cohabiting couples

Chart values (Percent):
- Adult child with parent: 49
- Cohabitation (no parent-child): 27
- Only relatives (no cohabitation or parent-child): 11
- Non-relatives in household (no cohabitation or parent-child): 13

Source : Larrimore, Mortenson, and Splinter (2021) using 2011 March CPS

Challenges and solutions for household links

Challenge: Typos/inaccurate addresses

Solution: Compare to valid address lists (Census may do better than we can). Use prior-year addresses and fuzzy matches to correct non-existent addresses

**Challenge:** Outdated addresses on information returns

Solution: Where possible, use 1040 addresses (most consistently updated)

Challenge: Taxpayer omits apartment information

Solution #1: Separate all tax units with no apartment # at addresses where others list apartments

Solution #2: Define large households (11+) as group quarters

IRS results consistently between the ACS and CPS household counts
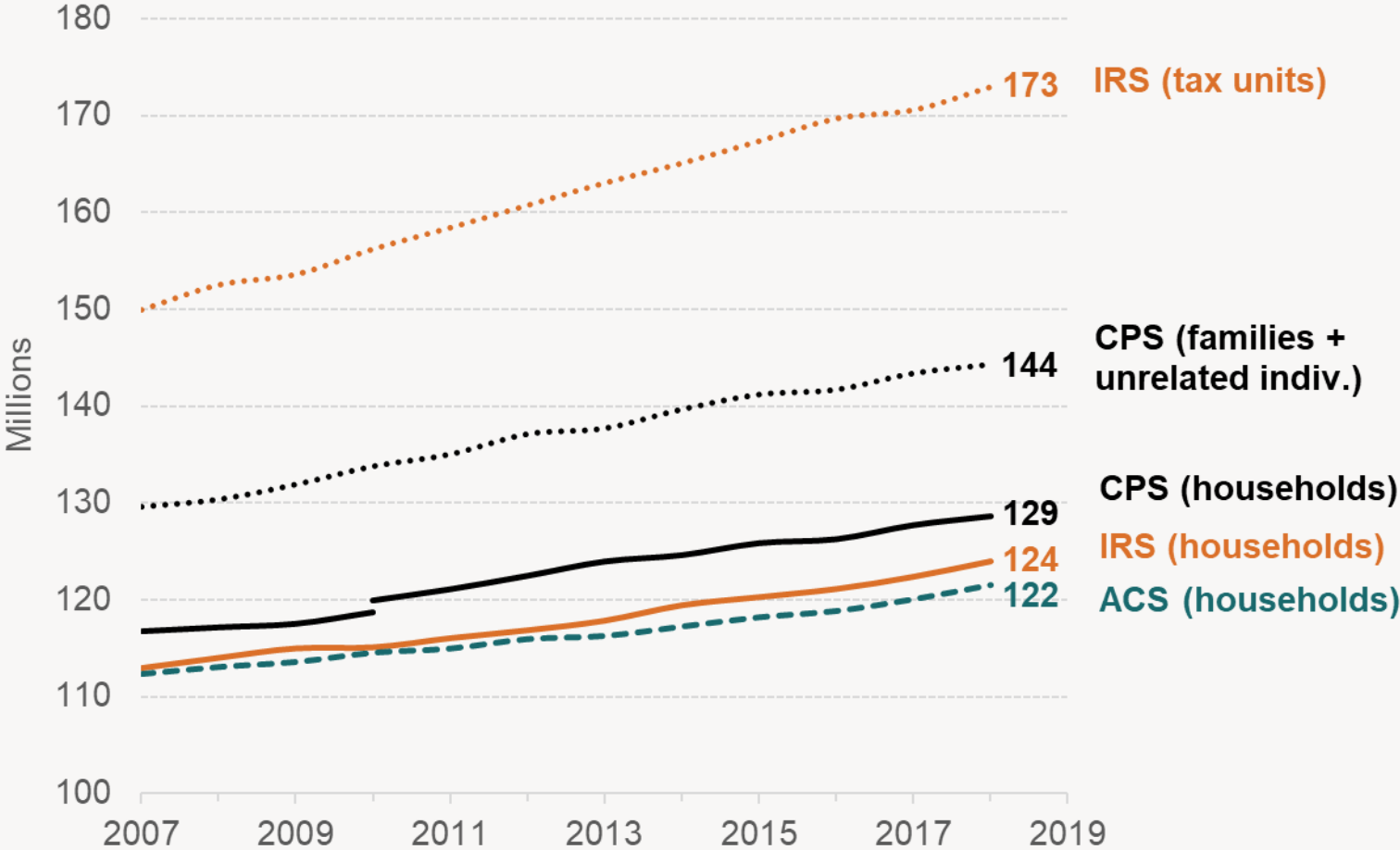
Source: Larrimore, Mortenson, and Splinter (2020)

Imputing Demographic Information

Tax system observes age, sex, and geography

It does not observe race or ethnicity

Options to include race/ethnicity:

    Imputation: Geography + Surname

    Survey link: Decennial linked to tax data

    (Survey links can only be done by Census due to one-way data sharing under Title 13/Title 26)

Conclusions

- US resident population well-captured in the IRS data *if* using information returns as well as 1040 data

- Tax-units are not households or families so should not be treated as such
  - Adjustments to the sharing units are necessary prior to comparing to survey data
  - Using the address information in tax data is successful for merging into households

- We don't link in data from outside sources due to Title 13 and (lack of) Data Use Agreements

Income in Tax Data

Jacob Mortenson

Joint Committee on Taxation, U.S. Congress

July 2022

What is Tax Data (2019)?

Information Returns: Both Filers and Non-Filers

Income type	Form	Forms (millions)	Amounts (billions)
Wages	Form W-2	276	$9,054
Retirement Withdrawals	1099-R	110	$1,236
Social Security	SSA-1099	68	$1,064
Business Income	K-1s, S-Corps/Partners	30	$634
Self-Employment	1099-MISC/NEC	42	$512
Dividends	1099-DIV	95	$269
Payment Processor	1099-K	3	$207
Interest Income	1099-INT	169	$79
Unemployment Comp.	1099-G	78	$30

Tax Returns: Filers Only

	Form	Forms (millions)	Amount (billions)
Tax Unit Composition Total Income	1040 Family	142	$12,112

Time Coverage and Timeliness

Time Periods Covered

Information Returns	1999–Present
Population of Tax Returns	1994–Present
Samples of Tax Returns	1960–2014 (public use); 1979–Present (SOI)

Timeliness

Information Returns	Spring of following year; Complete in one year
Tax Returns	Spring of following year; Never fully "complete"
SOI Data	Two year lag

Strengths and Weaknesses of Tax Data

MORE Accurate:	*LESS* Accurate:
With Third-Party Reporting	When there are Mis-Reporting Incentives
Federal Tax Liability & Credits	State Tax Liability & Credits
Income with Clear Timing *Wages, Self-Emp., Interest, Dividends*	Less Clear Timing *Capital Gains, Stock Options*
Taxable or Deductible Income	Non-Taxable and Non-Deductible Income *SNAP, Section 8, TANF, Imputed Rent*

Income Coverage by Source

Comparison of Income Items in 2010 ($billions)

	Tax Data	March CPS	Difference
Retirement Income	$930	$369	$561
Business Income, Self-Employment	$838	$451	$387
Wages and Salaries	$5,896	$6,133	-$237
Interest and Dividends	$381	$256	$126
Social Security and Disability	$696	$594	$101
Other Form 1040 Income	$87	–	$87
Unemployment compensation	$141	$97	$44
Alimony	$8	$5	$3
Market Income + UI/SS	$8,979	$7,859	**$1,119**
Transfers not on tax returns (excluded)	–	$201	**-$201**

Source: Larrimore, Mortenson, and Splinter (2021)

Retirement Income

Bee and Mitchell (2017) find that CPS poorly captures DC/DB/IRA withdrawals.

What does the tax system observe?
- Individual contributions to IRAs and DC plans
- IRA balances
- Withdrawals from IRAs, DCs + DBs (but not separately)

What does the tax system **not** observe?
- DC or DB balances $\rightarrow$ annual returns in either
- individual-specific employer contributions

Comparing Data Sources, 2010: Tax Units $\neq$ Households
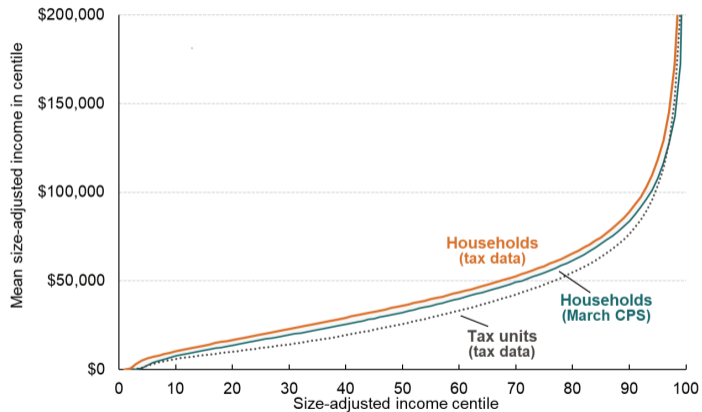
Source: Larrimore, Mortenson, and Splinter (2021)

Counts:
- 157.5m tax units
- 117.5m CPS households
- 115.3m Tax households

Gini Coefficients:
- 0.54 (tax units)
- 0.45 (CPS)
- 0.48 (tax households)

Comparing Data Sources, 2010: Top Incomes are Higher in Tax Data
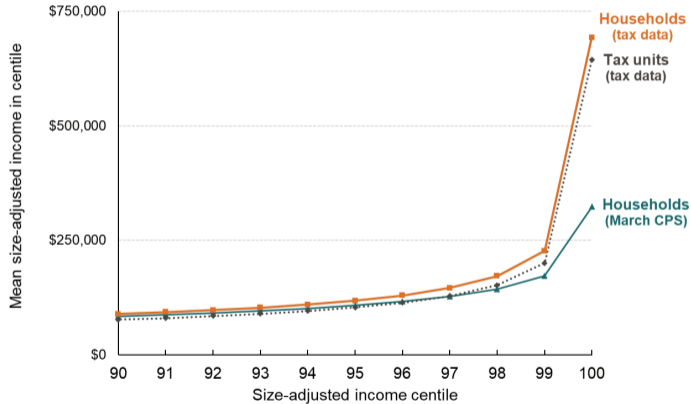
Top 5 Percent Share:
- 30% (tax units)
- 21% (CPS)
- 27% (tax households)

Source: Larrimore, Mortenson, and Splinter (2021)

Additional References

Line item estimates (2019):
- Publication 5385
- Publication 4801

Non-filers: Sailer, Weber, & Yau (1993); Sailer & Weber (1998); Mortenson et al. (2009); Cilke (2014)

Measuring retirement income: Goodman, Mackie, Mortenson, & Schramm (2021)

Measuring top labor income: Eisfeldt, Falato, & Xiaolan (2021)

Measuring top incomes: Auten & Splinter (2022); Piketty, Saez & Zucman (2018)

Income, Transfers, and Taxes:
Strengths & Limitations of Tax Data

David Splinter

July 2022
National Academy of Science CNSTAT Meeting

Outline

Can tax data go…
…from reported to full economic income?
…from economic income to disposable income?

Market Income
Strength: tax data > survey data
Limitation: tax data < national accounts

Transfers
Strength: accurate measures in tax data
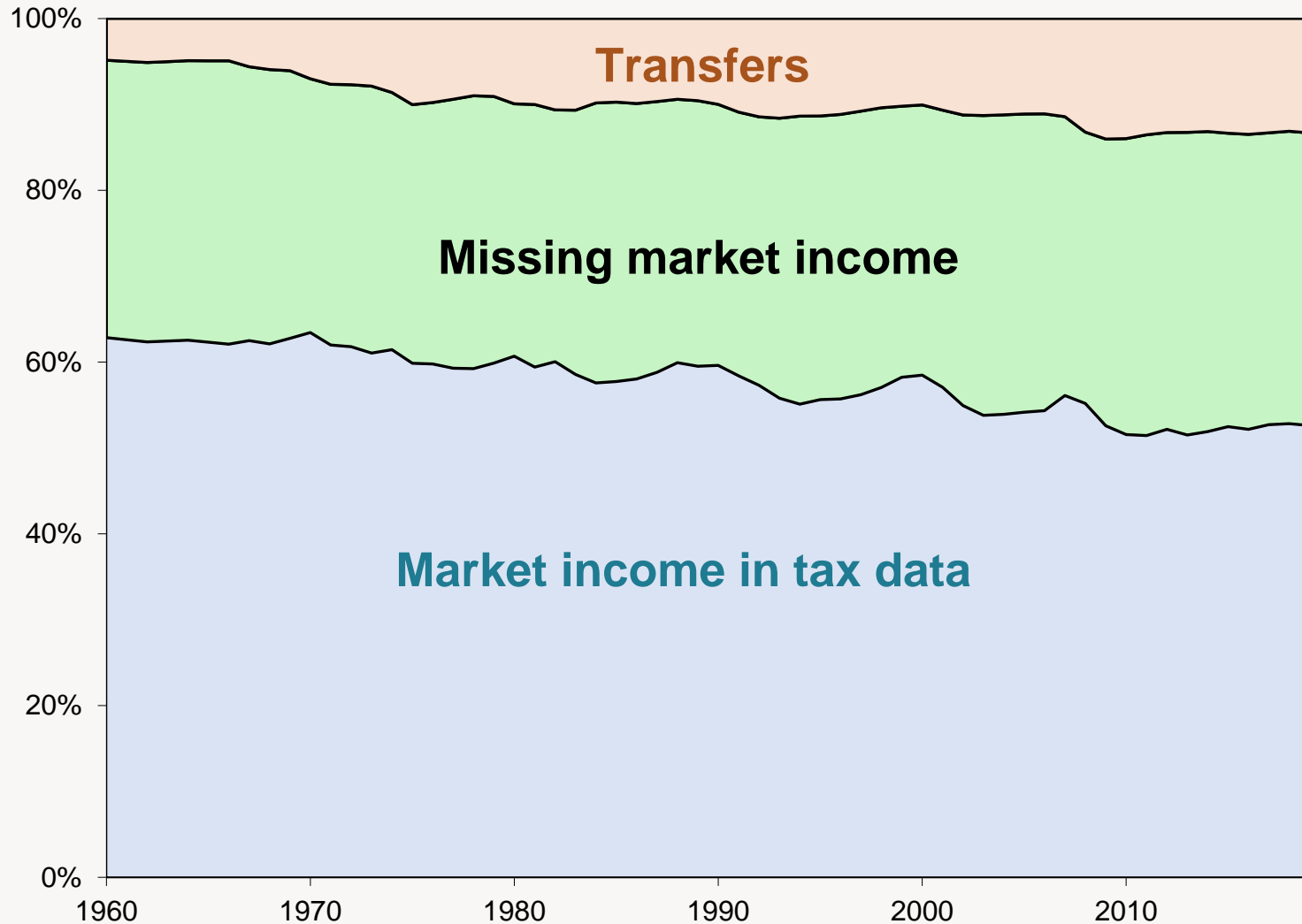**Limitation**: but misses many transfers

Taxes
Strength: accurate federal taxes in tax data
Limitation: but incomplete state/local taxes

Can tax data go towards consumption?

Missing market income and size of Transfers
1960-2019 (relative to national income + transfers)

Jeff Larrimore, Jacob Mortenson, and David Splinter

Source: Auten and Splinter (2022)

Market Income

Market Income, 2010 (trillions)
~$8t tax returns > ~$7t CPS, b/c has less retirement/business income

~$8t tax returns < ~$12t NIPA Personal Income (no transfers, with payroll tax)

NIPA Personal Income larger because
- Depreciation gaps ($0.2t), Misreporting ($0.8t), Imputed Rents ($0.5t), Retire. accounts/Health Insur. ($1.1t), Employer payroll taxes ($0.5t)
  [National Income: corporate retained earnings ($0.8t), government/non-profits]

Accelerated depreciation in tax data
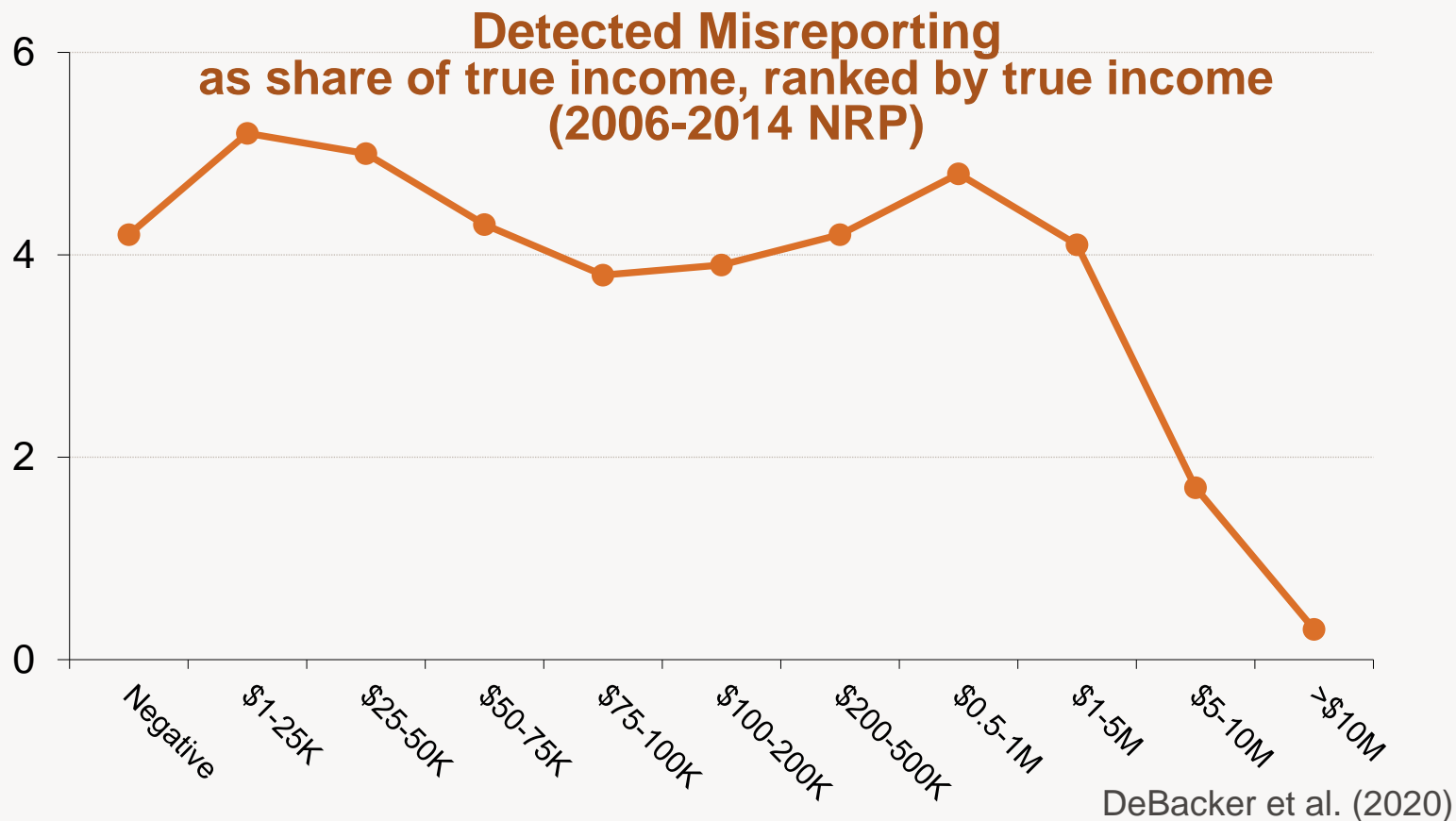- Can allocate by expensed amounts in tax data

Misreporting = underreporting – overreporting
- Cannot use operational audits to allocate misreporting
- Use special audit studies with random selection: non-filers, tips, S corps, & individual tax returns (NRP)
- NRP has higher sampling rates for higher incomes

Market Income

Misreporting rates similar over income distribution

- Misreporting rates fall for very high incomes (>$5M)
  but missing some offshore income in these earlier years

**Detected Misreporting
as share of true income, ranked by true income
(2006-2014 NRP)**

DeBacker et al. (2020)

Market Income

Little impact on income shares from adding misreporting

Detailed NRP estimates can be used to impute detected and undetected misreporting: Auten and Splinter ([2021](#), [2022](#))
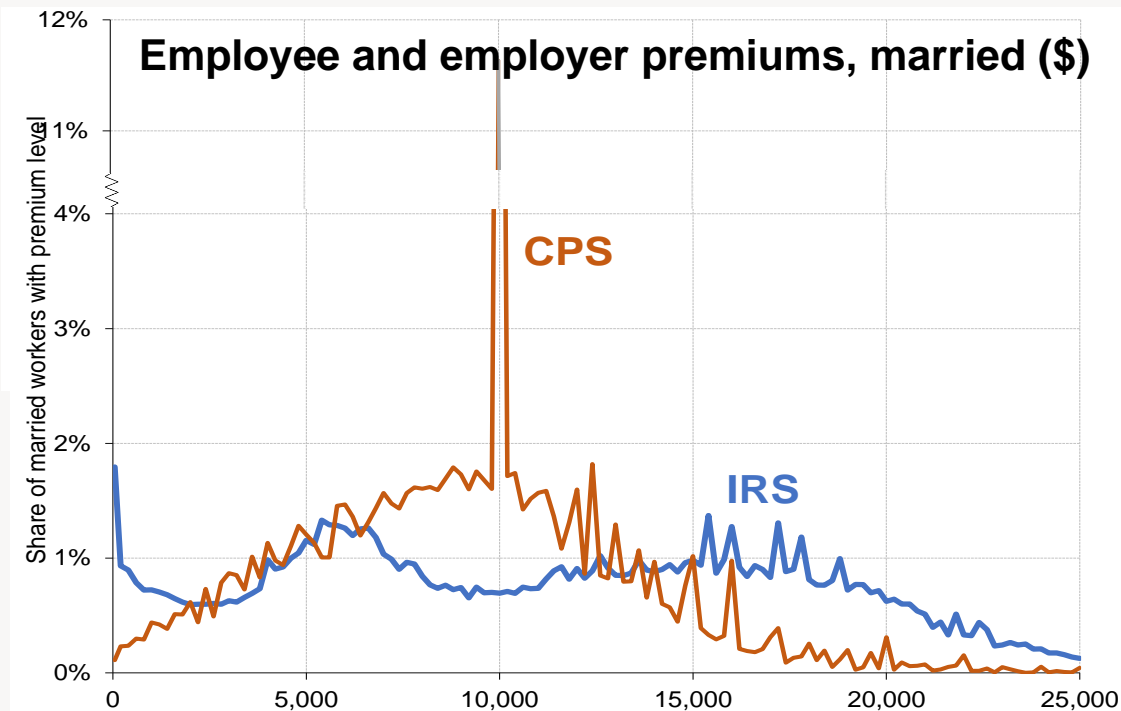

Johns and Slemrod (2010)
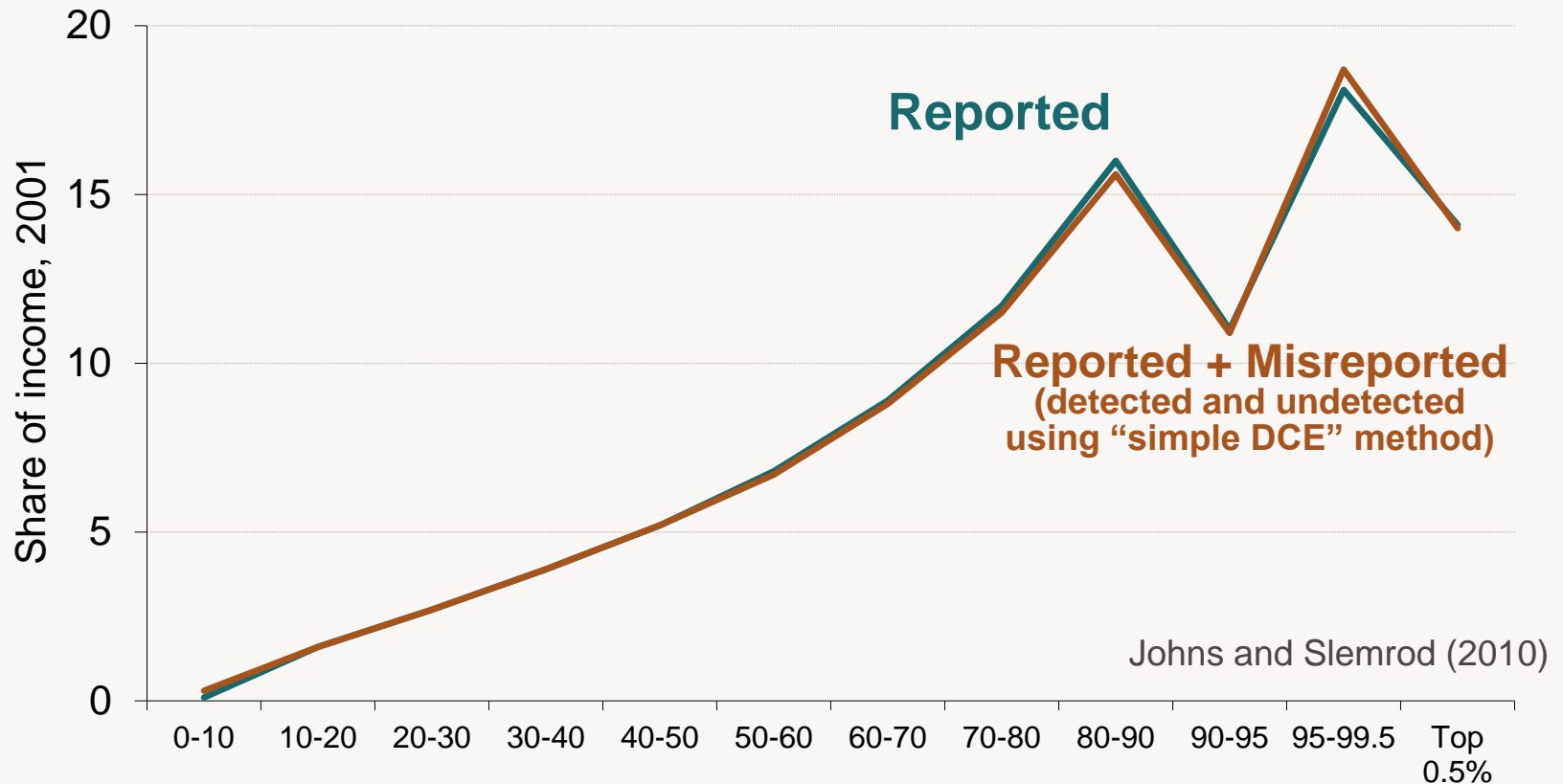
Market Income

Retirement account income
- Timing issues: undistributed vs. distributed income
- Allocate using SCF-based information retirement wealth

Private Health Insurance
- SE on tax returns, most employer insurance on W-2s
- Tax data captures expensive plans, top-coded in CPS

Employee and employer premiums, married ($)

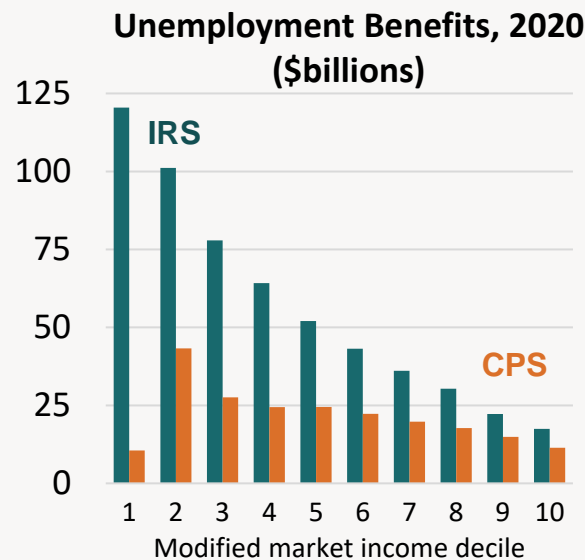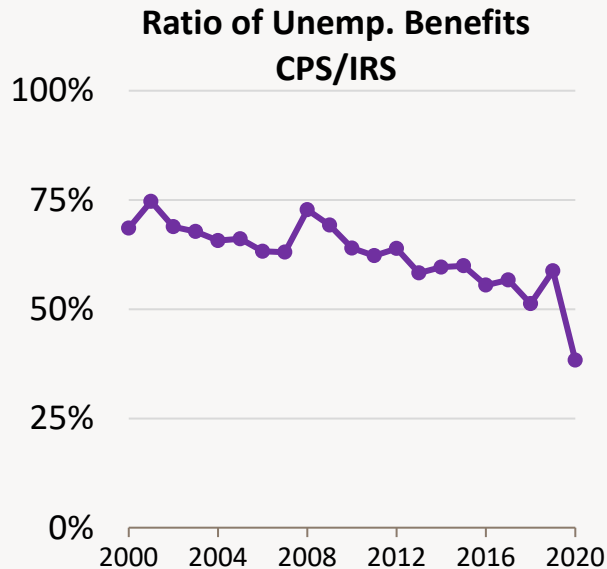Source: Larrimore, and Splinter ([2019](#))

Transfers

Social Security and Disability Benefits (full)
- Form SSA-1099 (and reported on tax returns)
- Substantial share missing in CPS

Unemployment Benefits (full)
- Form 1099-G (and reported on tax returns)
- Falling share reported in CPS, missing for low incomes

Ratio of Unemp. Benefits CPS/IRS

Unemployment Benefits, 2020 ($billions)

IRS

CPS

Modified market income decile

Source: Larrimore, Mortenson, and Splinter (2022)

Transfers

ACA exchanges
- Premium tax credits on tax returns
- Health insur. marketplace coverage: Form 1095-A

Medicare/Medicaid & adult dependents
- Impute by demographic and income groups
- Individual coverage on Form 1095-B

- 1095 Forms: See work by Lurie et al.

Transfers

Other cash transfers

- In tax data: Stimulus Payments, ~$1 trillion over Covid-era
- Missing from tax data: SNAP, SSI, TANF, Workers' compensation, Veterans' benefits, educational assistance

Imputing to tax data

- CPS to tax data: Share receiving and amount received across income and demographic groups
- Larrimore et al. (2021 JPE) and CBO (Habib 2018)
- Linked Survey-Admin data: Rothbaum et al.; Meyer et al.

Taxes

Federal individual income taxes and credits (full)
- $158b refundable tax credits in 2019
  Main child anti-poverty programs, underestimated in CPS (Meyer et al. 2020)
  Children reassigned for credits: Jones & O'Hara (2016); Larrimore, Mortenson, & Splinter (2017)

Federal Payroll taxes (full)
- Calculate from wages, self-empl. taxes reported on tax returns

State/local indiv. income & home property taxes (partial)
- Deducted amounts observed for itemizers
  ~90% deducted in 2017, ~40% since 2018 (post-TCJA)
- Income taxes: estimate with historical data or state tax calculators
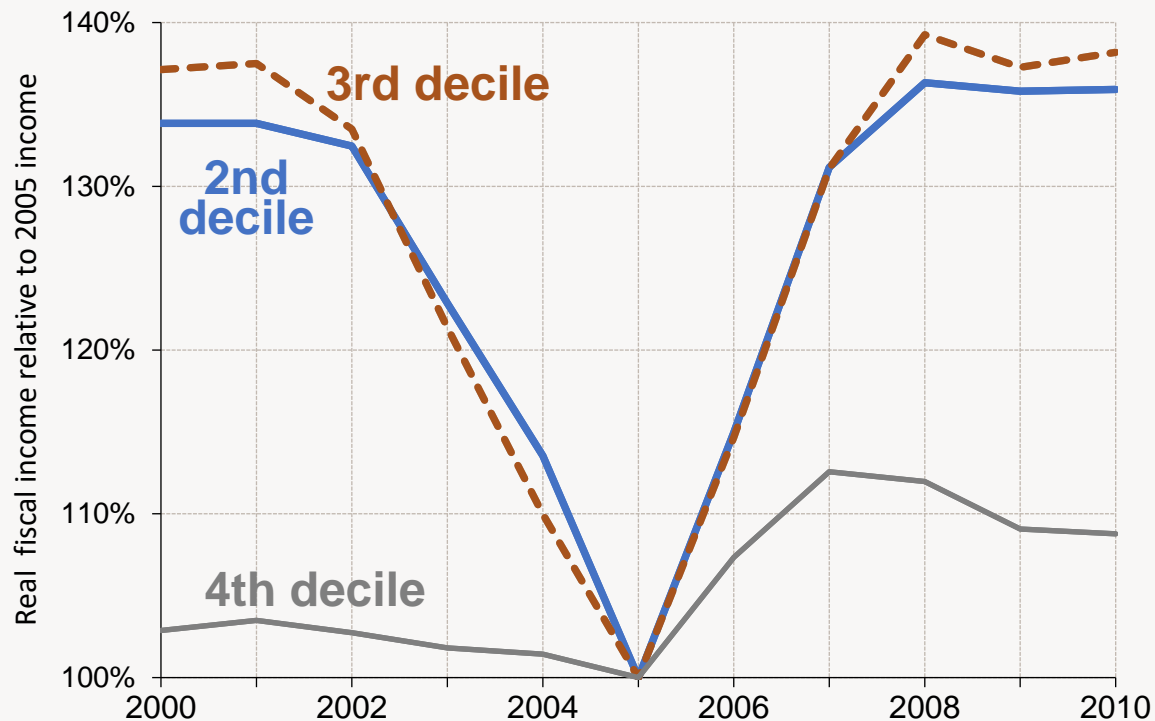- Property taxes/Imputed rents: impute with mortgage info & survey data

Sales and excise taxes (missing)
- Allocate by household consumption?

Towards Consumption

Permanent income hypothesis

- Consumption inequality < Disposable income inequality
- Tax data panels can provide multi-year incomes
- Market incomes are volatile at bottom of annual distribution

Towards Consumption

Retirement distributions
- Likely for consumption (exclude rollovers & required min. distributions)

Sharing across tax units
- Households can be constructed in tax data
- Private transfers: non-resident dependents support, some charitable contributions & alimony
- Missing most private transfers: child support, in-kind, etc.

Limited info. on gifts/bequests in tax data
- High threshold for annual gifts: $15,000 in 2021
- High threshold for bequests: $11.7 million in 2021

Conclusion

Population in tax data

Tax data include about 99% of resident population

Income, Transfers, and Taxes in tax data

Tax income > survey income, but similar distribution
Tax income < NIPA income, missing sources & misreporting
Accurate transfers for SS/UI, but missing many transfers
Accurate federal taxes, but incomplete state/local taxes

Augment tax data with other sources…

Piketty, Saez, & Zucman (2018); Auten & Splinter (2022), matched Census-Admin data

…adjust surveys with augmented tax data

Example for unemployment benefits: Larrimore, Mortenson, & Splinter (2022)
Allows for an open-source method