
Tax Data:

Possibilities and Limitations for Joint Measures of Household Income, Consumption, and Wealth

David Splinter and Jacob Mortenson
Joint Committee on Taxation

December 2024
CNSTAT Meeting

These comments embody 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.

Why care about Income, Consumption, and Wealth together

Economic prosperity not fully measured by one

- **Income insufficient**

Retiree with substantial assets but low income

- **Wealth insufficient**

Early career worker: student loan debt & high income

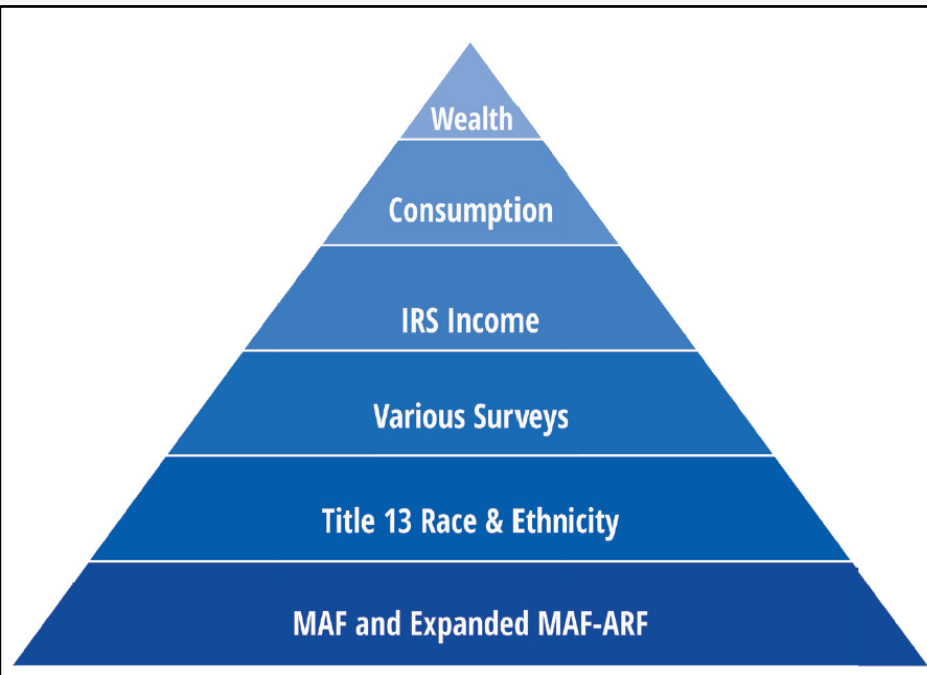
- **Consumption insufficient**

Individual supporting current consumption by accumulating debt

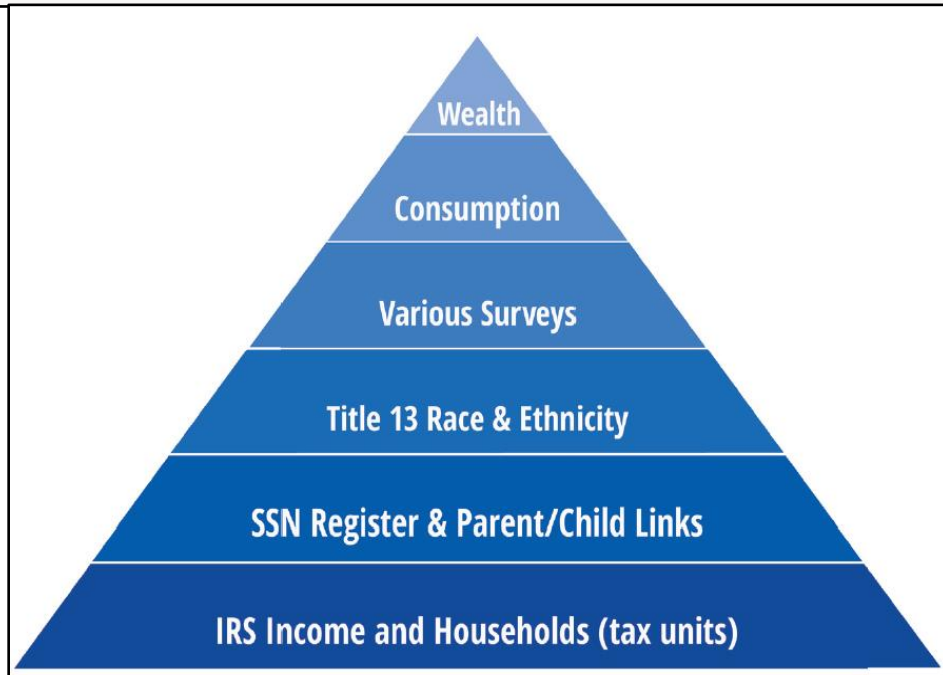
Potential “Spines” for data

CNSTAT ([2024](#)) considers Census or IRS data spines that capture many measures of economic prosperity/need. We discuss possibilities and limitations of a tax spine.

Census MAF Spine



Tax Data (IRS) Spine



Tax data spine: Pros & Cons

Pros

- Population coverage is nearly complete
 - Most institutional population (students 1098-T, elderly SSA-1099, incarcerated Garin et al. [2024](#))
- Administrative data
 - Health insurance, Transfers (e.g., no unemployment insurance underreporting)
- Third-party income reporting (wages, 1099-NEC, dividends, etc.)
- Modest data lag
- Panel data
 - Short-term (Larrimore, Mortenson, & Splinter [2017](#); [2022a](#); [2022b](#); [2023](#))
 - Long-term/Intergenerational (Hoynes & Schanzenbach [2018](#); Chetty, Hendren, & Katz [2016](#))

Cons

- Missing if no IRS contact or third-party reports
- Missing non-tax transfer programs (SSI, SNAP, etc.)
- No monthly data (except health insurance)
- Limited data access: IRS now has Joint Statistical Research Program
- Limited demographic variables

Who is in the tax data?

In 1040 tax return data (filers, spouses, & dependents) **or** has third-party information return such as:

- Wages (W-2)
- Some self-employment (1099-MISC / 1099-NEC)
- Social Security (SSA-1099)
- Interest or dividends (1099-INT / 1099-DIV)
- Retirement withdrawals (1099-R)
- Unemployment income (1099-G)
- Business income (K1s from partnership or S-corps)
- Higher education (1098-E and 1098-T)

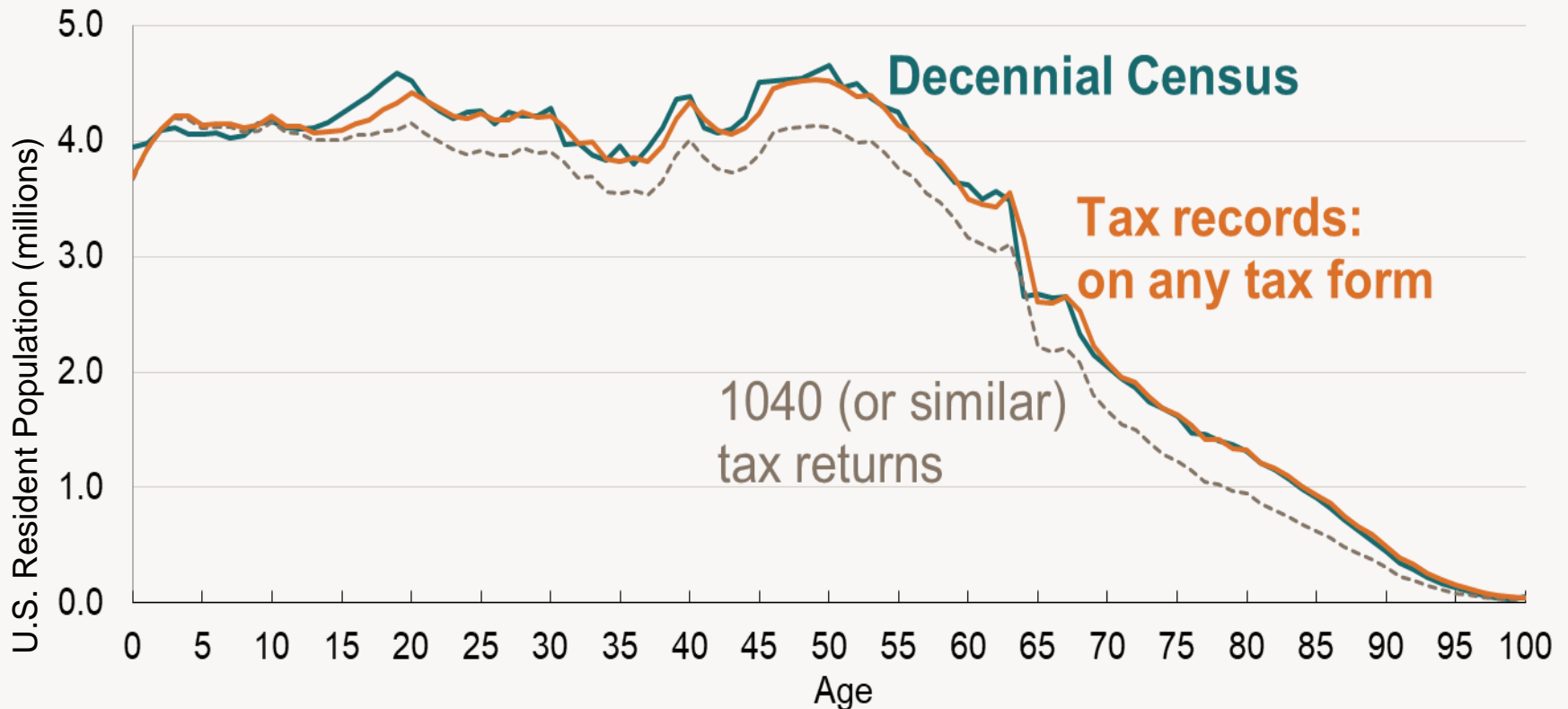
Reasons people excluded from data

- No income (or other activity) with reporting requirements
- Only under-the-table earnings
- Teenagers not claimed as a dependent by any filer

Population coverage (2010)

Decennial Census: 309 million

Tax data (all residents): >306 million → about 99%



Creating sharing units

Tax Units \neq Households

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

Filers: observe those filing a tax return together, not complete households (or Census families)

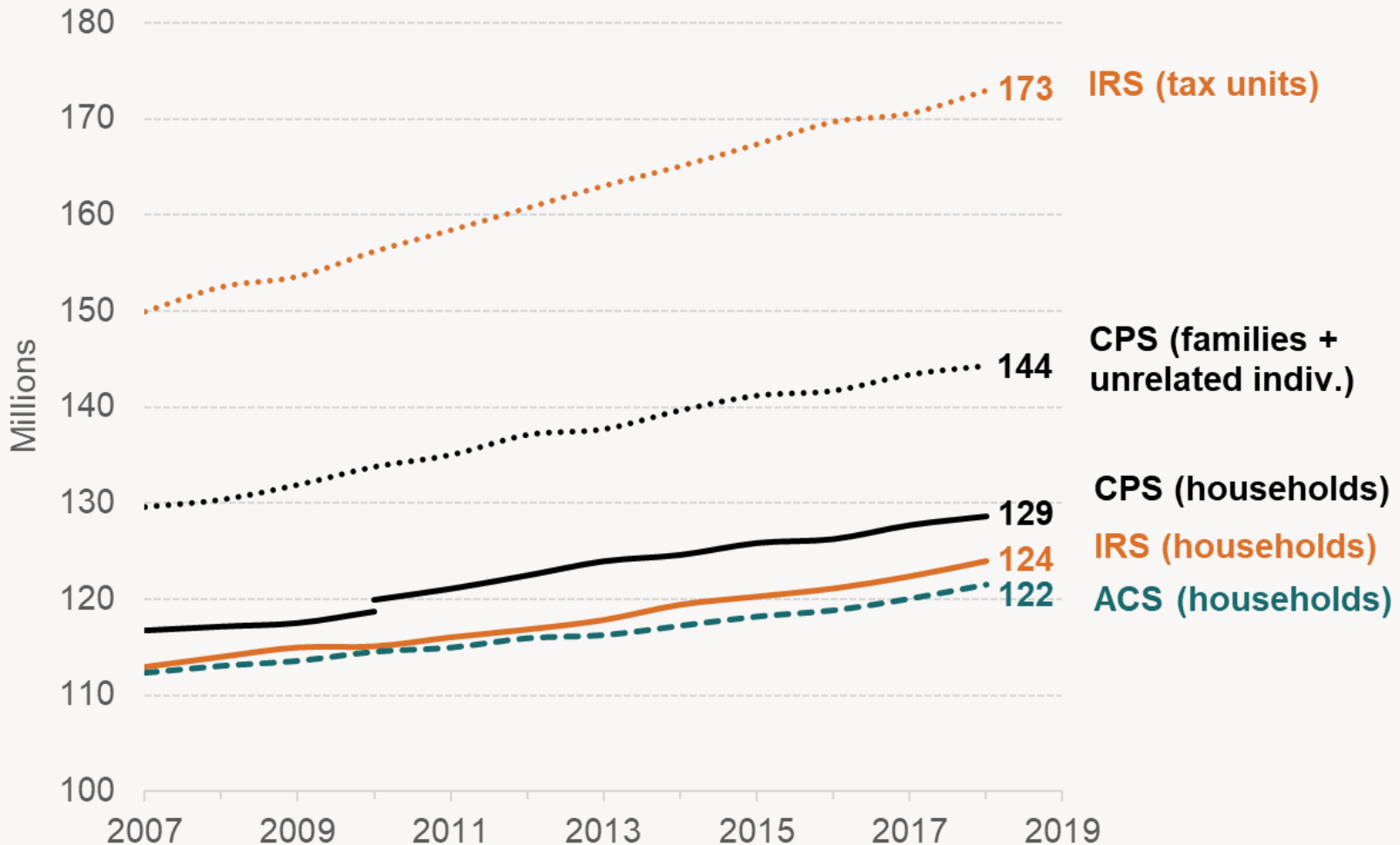
Non-filers: observe no household or family links

Fortunately, all tax forms have a mailing address

Can use address to link households

(Larrimore, Mortenson, and Splinter [2021](#))

Address links in tax data gives HH counts between ACS and CPS



Source: Larrimore, Mortenson, and Splinter (2022)

Market income in tax data to disposable income (after transfers & taxes)

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

Market income in tax data

Market Income, 2010 (trillions)

\$8t tax returns < \$12t NIPA Personal Income (no transfers)

NIPA Personal Income larger because

Employer insurance (\$0.7t), Employer payroll taxes (\$0.5t),
Imputed rents for owner-occupied housing (\$0.5t),
Underreported income (\$0.8t), Retirement income

Underreported Income

Impute using special random audit studies
(Auten and Langetieg [2023](#), Auten and Splinter [2021](#))

Disposable Income: Add transfers

Social Security and Disability Benefits

- Form SSA-1099 and reported on tax returns

Unemployment Benefits

- Form 1099-G and reported on tax returns
- Growing underreporting in CPS (Larrimore, Mortenson, & Splinter [2022](#))

ACA health insurance exchanges

- Premium tax credits in tax data
- Health insurance coverage (Form 1095-A)

Medicare/Medicaid

- Impute by demographic and income groups
- Individual coverage (Form 1095-B)

Disposable Income: Add transfers

Other cash transfers

- Tax data: Stimulus checks
- Missing from tax data: SNAP, SSI, Workers' comp., Veterans' benefits, most educational/childcare assistance, WIC, etc.
- Tax data currently linked to few outside sources due to Title 13 and (lack of) data use agreements

Missing transfers approaches

- CPS imputation to tax data: Share receiving and amount received by income/demographic groups (Larrimore et al. [2021](#); CBO [2018](#))
- Linked Survey-Admin data: Rothbaum et al.; Meyer et al.

Disposable Income: Remove taxes

Federal individual income taxes and credits

- Refundable tax credits underestimated in CPS (Meyer et al. [2020](#))

Federal payroll taxes

- Calculate from wages & self-empl. taxes on tax returns

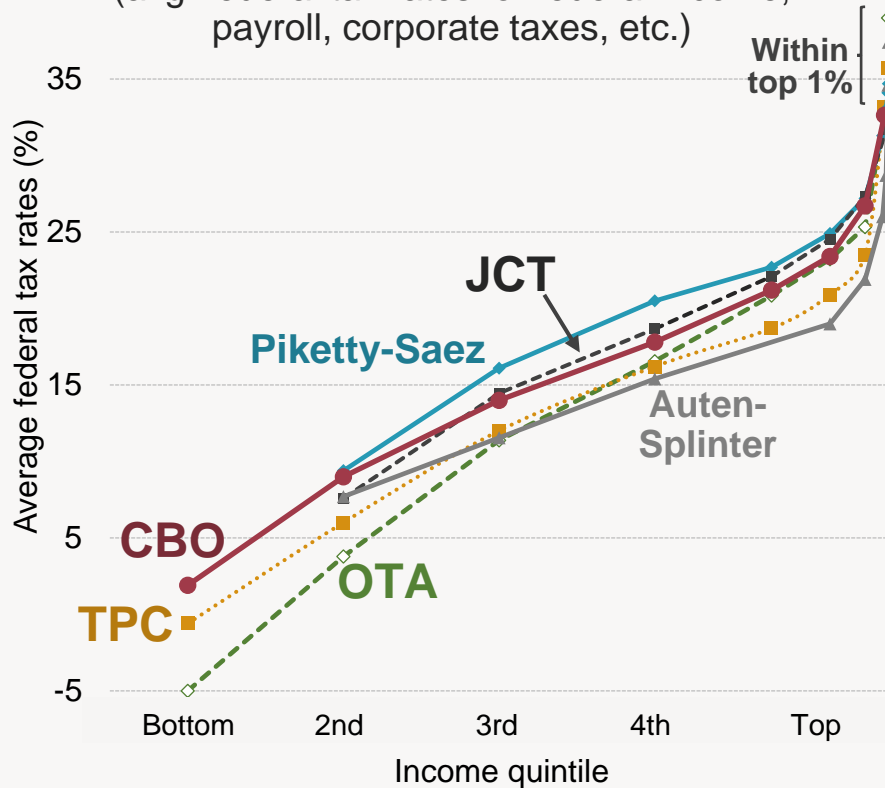
State/local indiv. income & home property taxes

- Deducted amounts observed for itemizers
 - ~90% deducted in 2017, ~40% since 2018 (post-TCJA)
- Income taxes: historical deductions or state tax calculators
- Property taxes: historical deductions or impute w/mortgage data

Disposable income is more equal

Federal taxes are progressive

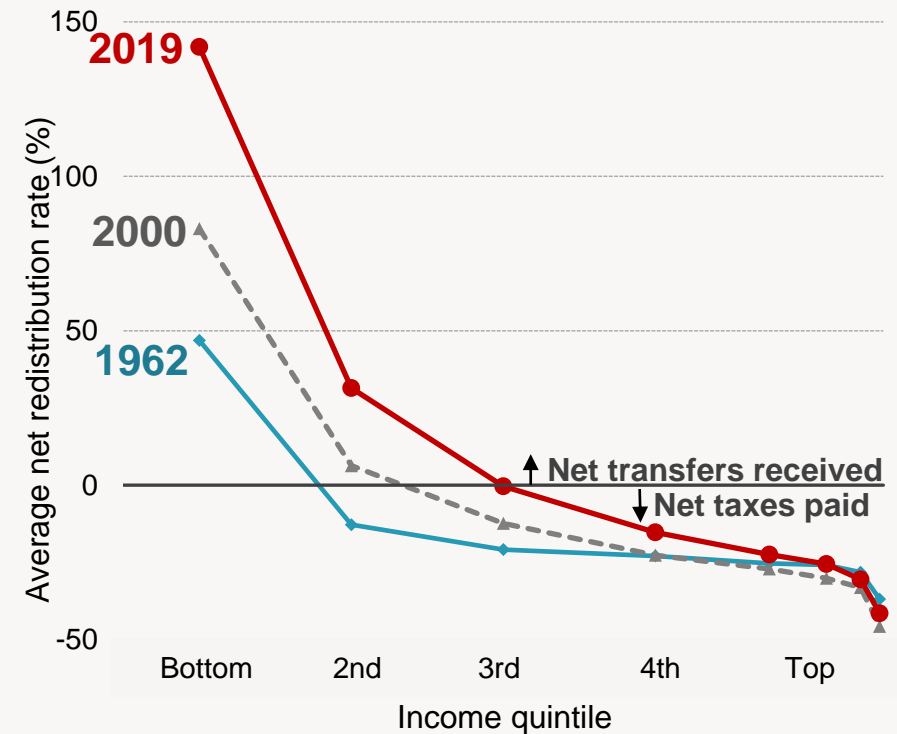
(avg. federal tax rates for federal income, payroll, corporate taxes, etc.)



Source: Splinter (2020). Notes: Average tax rates are taxes divided by income, defined by Piketty-Saez as fiscal income plus payroll and corporate taxes, AS as pre-tax/after-transfer national income, JCT as expanded income, TPC and OTA as expanded cash income, and CBO as market income plus social insurance benefits. Rates are for 2014, but OTA and JCT for 2015 and Piketty-Saez for 2004. See online data for additional details.

All taxes and transfers are progressive

(net redistribution rates for all federal, state, and local tax/transfer programs)



Source: Auten and Splinter (2024). Notes: Average net redistribution rates are all cash and non-cash transfers (no government consumption) less taxes (federal, state, and local taxes, including payroll taxes) divided by pre-tax income of each income group. The top quintile is divided into four groups: P80-90, P90-95, P95-99, and the top 1%.

Wealth: Capitalize income in tax data

Capitalization usually scales up capital income sources

- Can approximate wealth distribution in Survey of Consumer Finances

Most wealth is from homes, corp. equity, & retirement

- Home value by **property taxes**, debt by **mortgage interest** (itemizers/1098)
- Corporate equity by **dividends** and **capital gains**
- Retirement: IRA wealth reported (Form 5498)
but DC/DB wealth less clear (only observe distributions and rollovers)

Top wealth sources

- Fixed-income assets from **interest income** (equal-returns problematic)
- Non-corporate business assets: **passthrough balance sheets** and **K1s**
- Research focuses on top shares: Saez & Zucman ([2016](#)); Smith, Zidar, & Zwick ([2023](#))

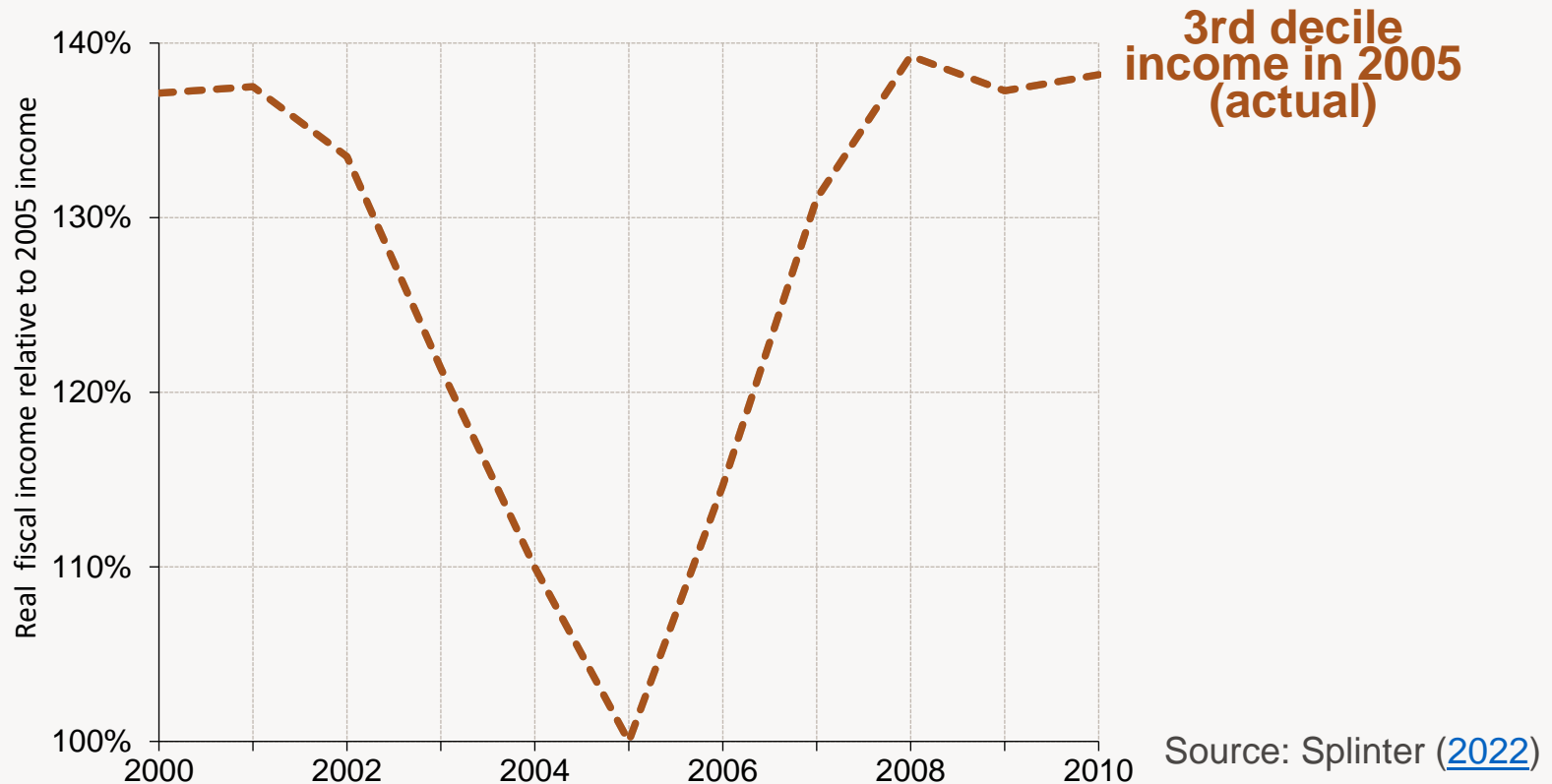
Bottom net worth harder to observe in tax data

- Durable goods (cars), cash deposits, consumer debt, etc.
- Social Security wealth (Sabelhaus & Volz 2020; Catherine, Miller, & Sarin 2023)

Consumption in tax data

Permanent income hypothesis

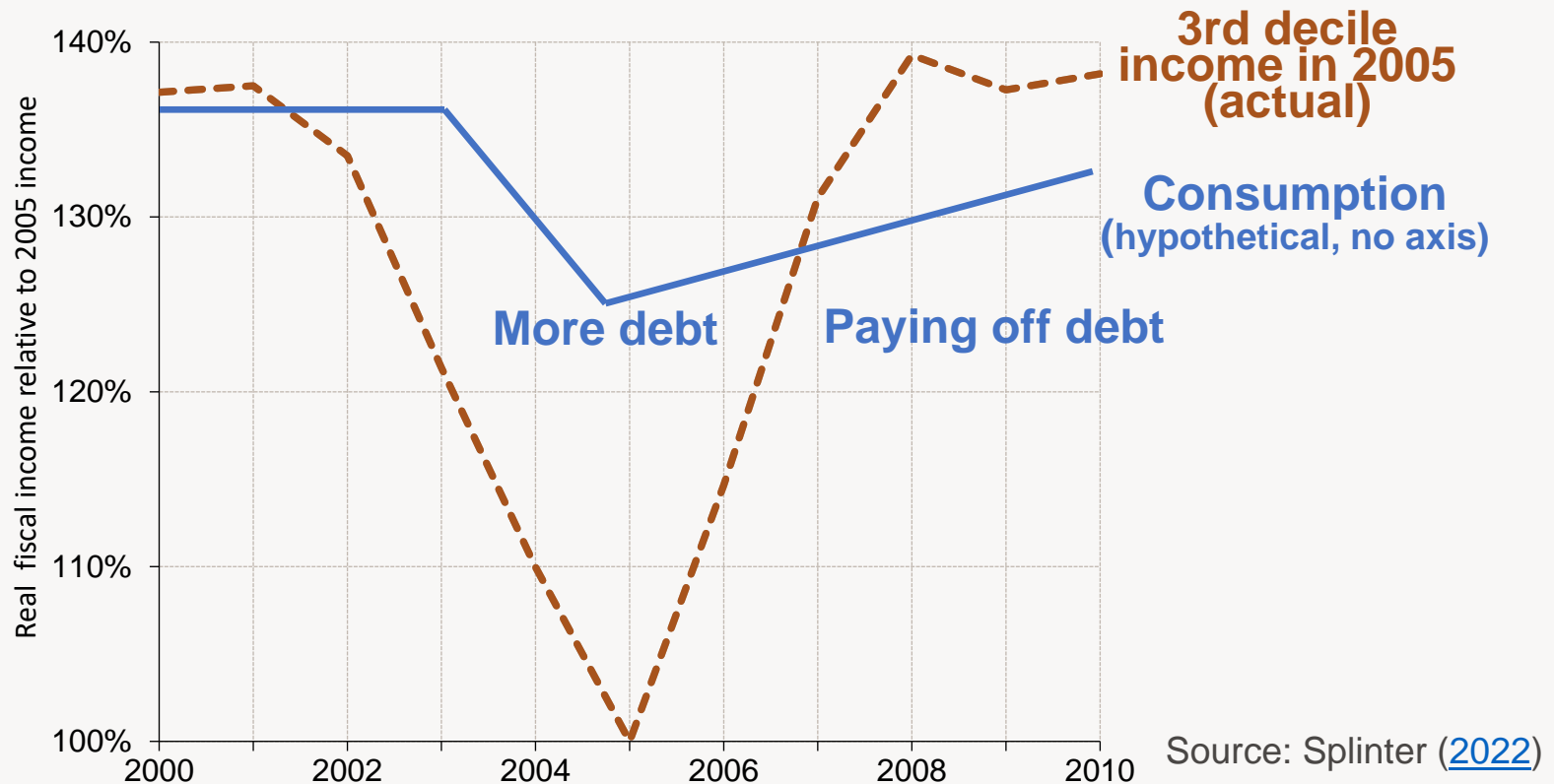
- Market incomes are volatile at bottom of annual distribution
- Consumption should be less volatile



Consumption in tax data

Permanent income hypothesis

- Market incomes are volatile at bottom of annual distribution
- Consumption should be less volatile



Consumption in tax data: Problematic

Consumption from disposable income & Δ wealth

- Consumption = Disposable Income – Savings
- Savings (Δ wealth) is problematic

Don't know individual-level rates of return/savings/debt

Retirement account distributions (dissavings in tax data)

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

Sharing of resources

- Households captures some sharing across individuals
- Some private transfers in tax data: non-resident dependents & alimony
- Missing most private transfers: child support, in-kind, gifts, etc.

Limited gifts/bequests in tax data

- High threshold for reporting gifts: \$15,000 in 2021
- High threshold for reporting bequests: \$11.7 million in 2021

Demographic Information in Tax Data

Date of birth, date of death, sex, and address
... but not race/ethnicity

Options to include race/ethnicity:

- Knowable only to Census: survey + tax data
- Blurred race/ethnicity file from Census
- BIFSG: Derby, Dowd, Mortenson ([2024](#))

Bayesian Improved First name Surname Geocoding (BIFSG in Derby, Dowd, Mortenson [2024](#))

Inputs: Distribution of race/ethnicity across ...

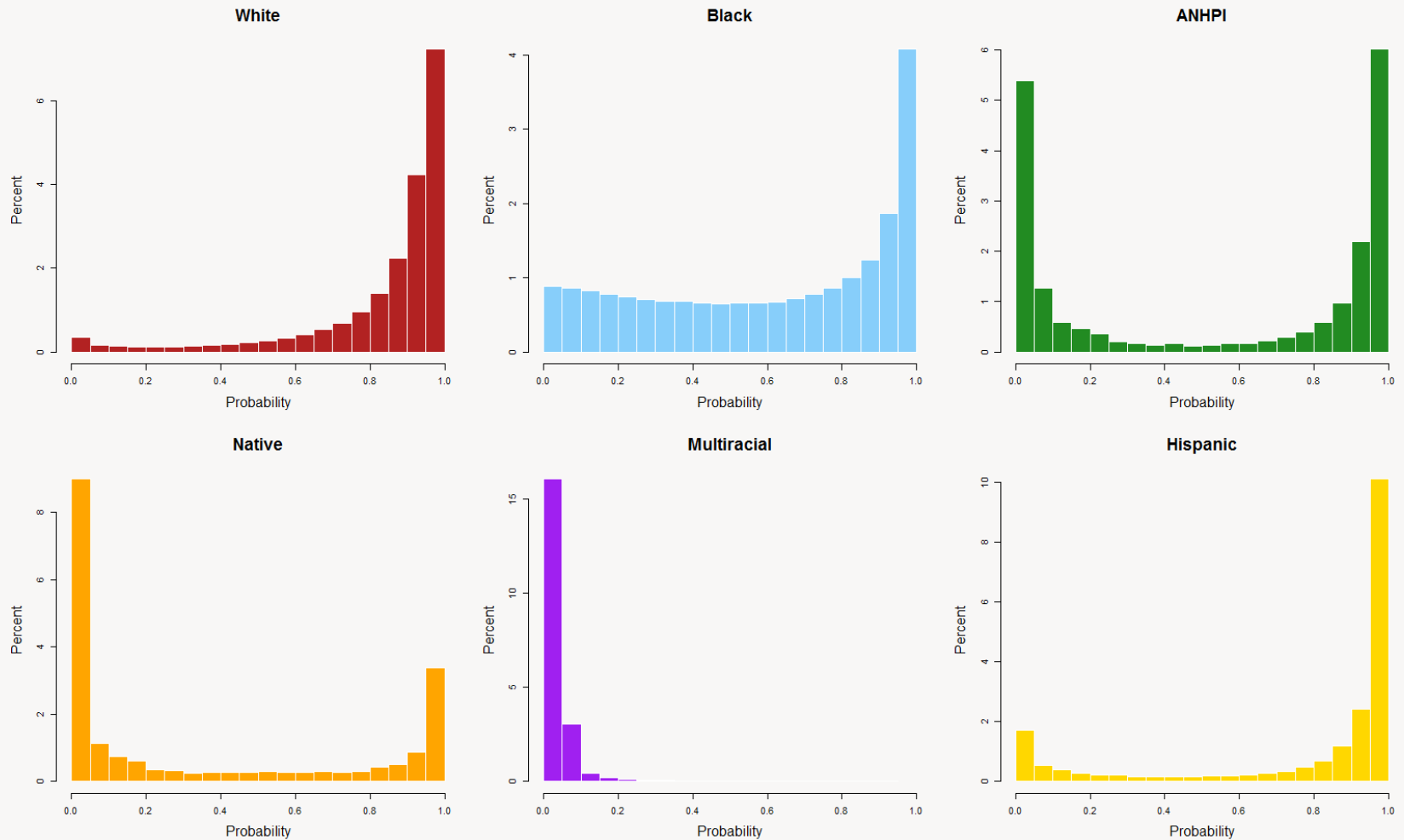
- **First names** (voter files; Rosenman, et al.)
- **Surnames** (Census)
- **Geography** (5 year ACS)

Bayesian combination of this information

Output: set of probabilities, that an individual belongs to one of six groups...

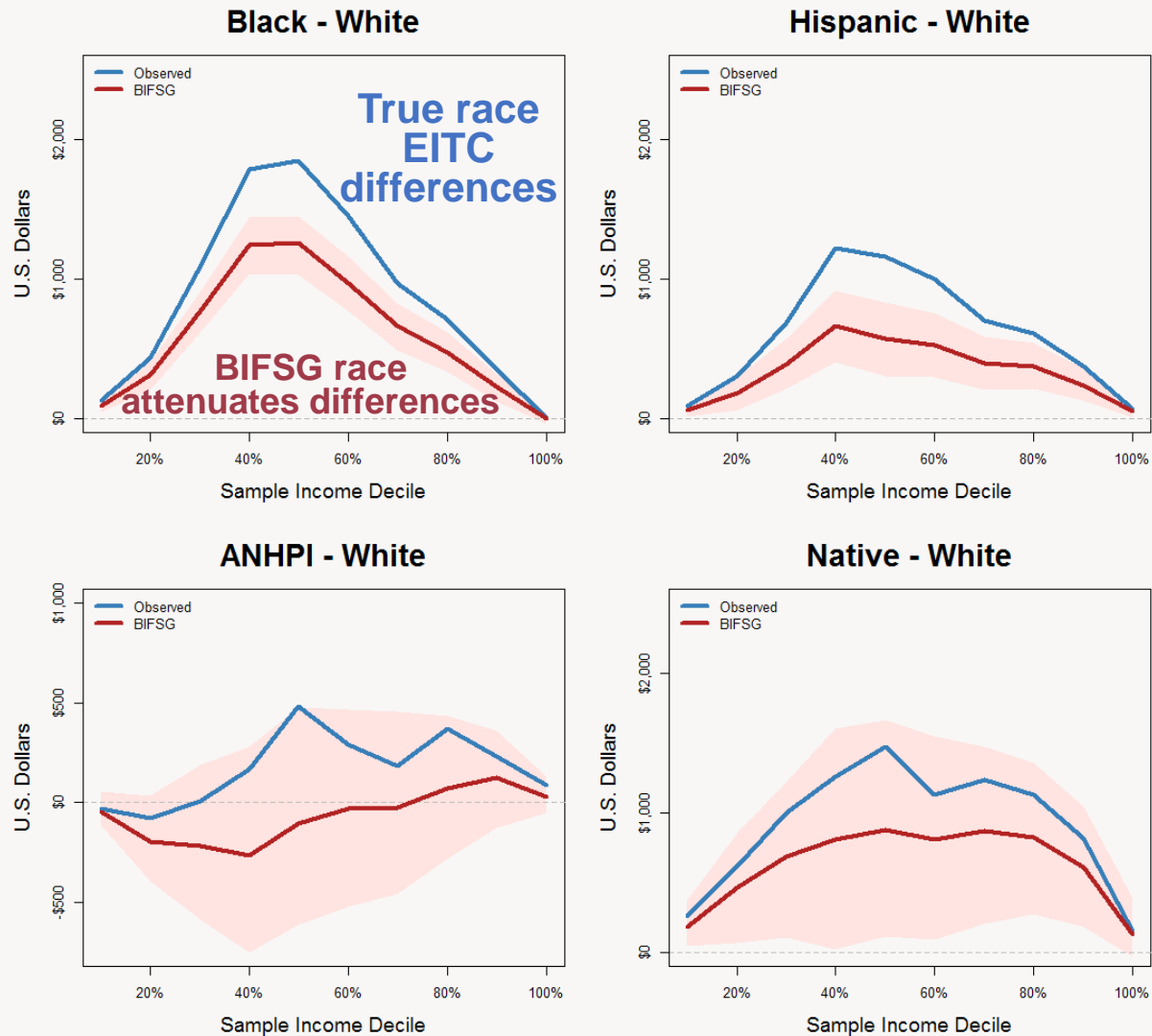
- Hispanic
- Non-Hispanic (NH) White (“White”)
- NH Black or African American (“Black”)
- NH Asian, Native Hawaiian, or Pacific Islander (“ANHPI”)
- NH American Indian or Alaskan Native (“Native”)
- NH Multiracial or Other (“Multiracial”)

How Well Does BIFSG Do, When We Observe Race?



Source: Derby, Dowd, Mortenson ([2024](#))

Using BIFSG to Estimate EITC Disparities Across Groups



Source: Derby, Dowd, Mortenson (2024)

Conclusion

Population in tax data

Tax data include about 99% of resident population
Can merge tax units into households with addresses

Income, Transfers, and Taxes in tax data

Tax income < NIPA income, can calculate/impute missing
Accurate transfers for SS/UI, but missing many transfers
Accurate federal taxes, but incomplete state/local taxes

Wealth and Consumption in tax data

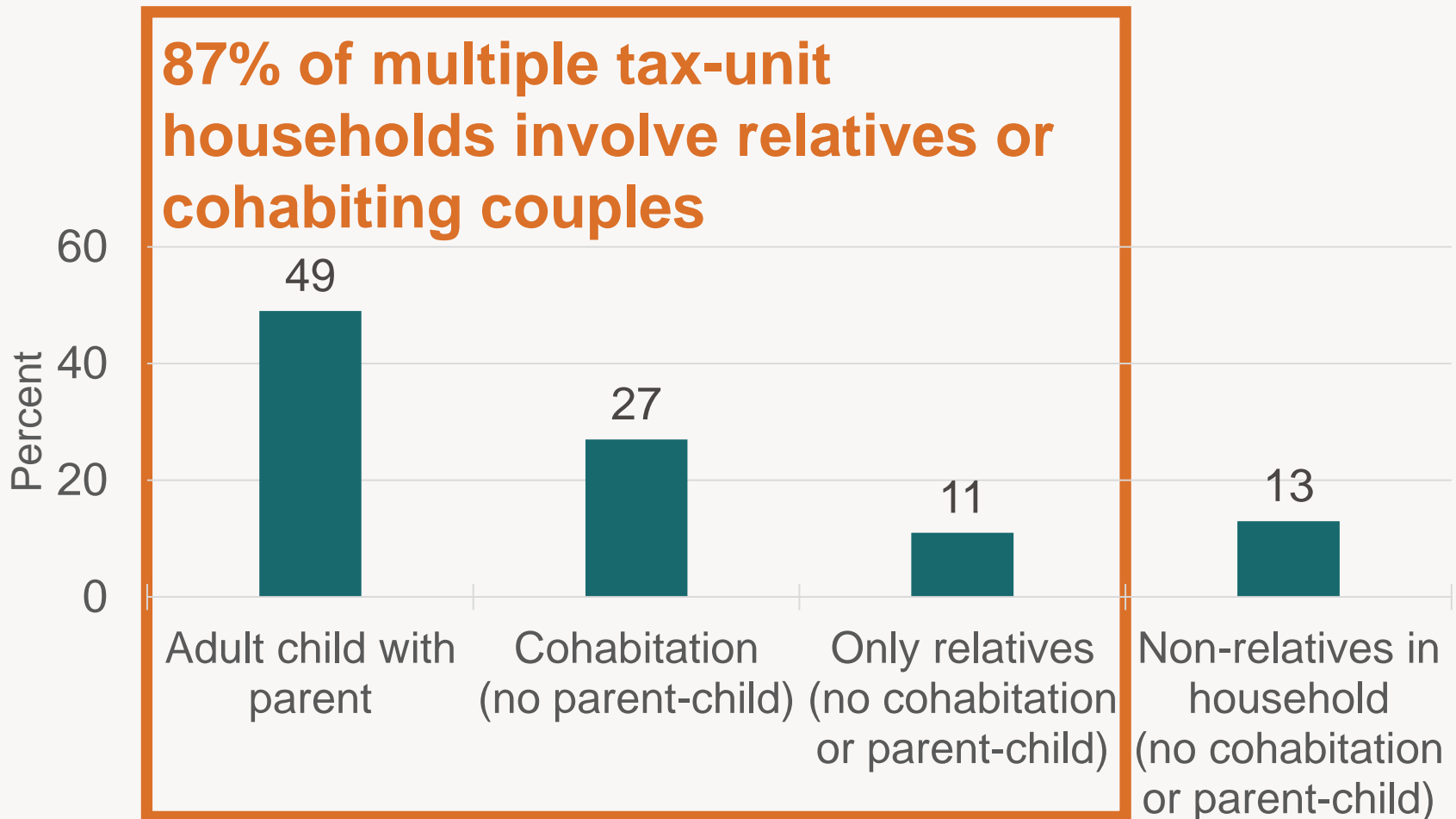
Wealth from capitalized income: Noisy at individual level
Consumption from income plus dissavings: Problematic

Race/Ethnicity imputations

Be careful (and incorporate uncertainty)

Appendix

Relationships in multi-TU households



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

Challenges and solutions for household links in tax data

Challenge: Typos/inaccurate addresses

Solution: Compare to valid address lists. 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

Unit of analysis / sharing unit

Report recommends the **household as unit of analysis**

Can use **household** as sharing unit (for equivalence scales)
....but **individual** as unit of analysis

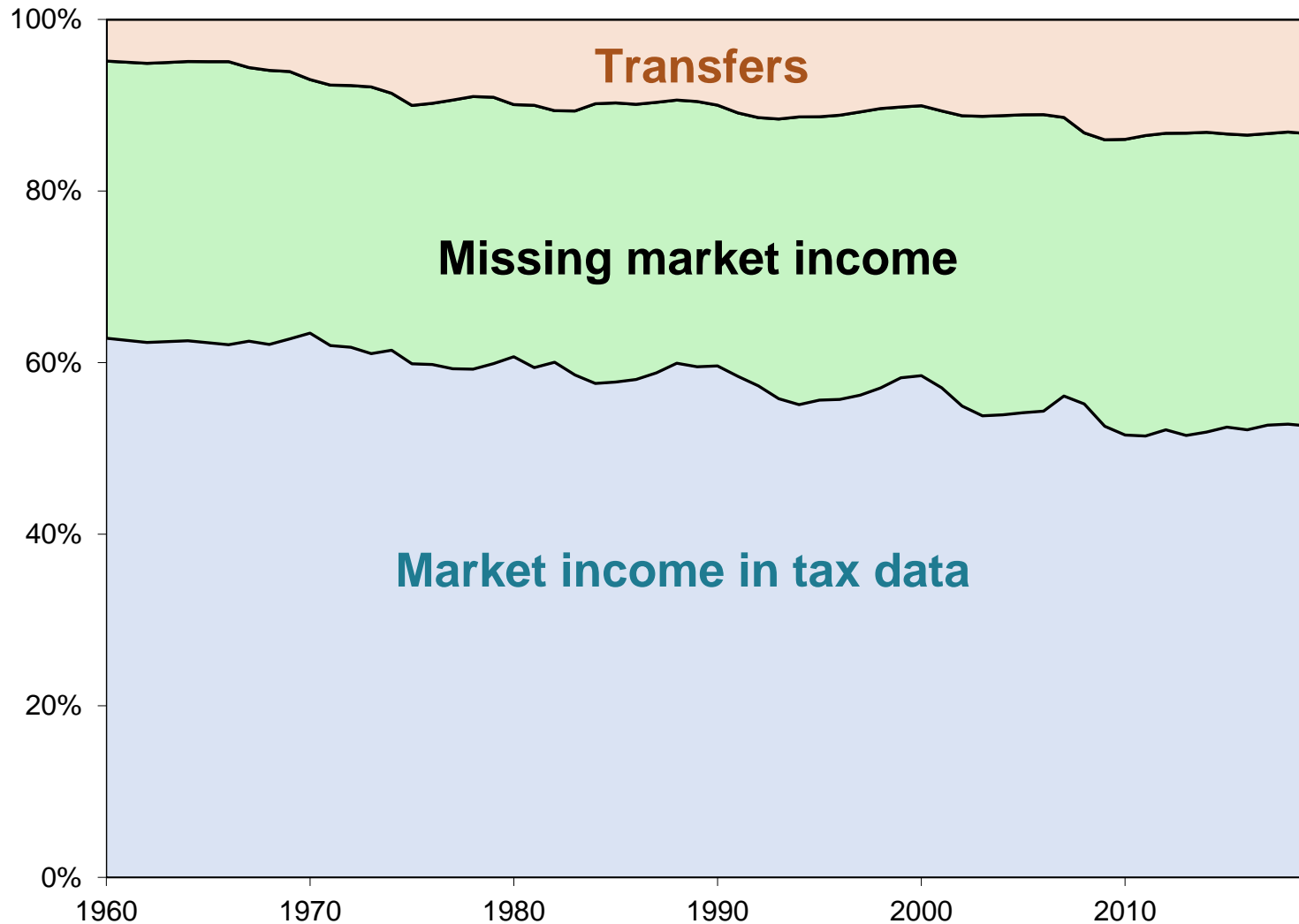
- Care about distribution across people, not households

Avg. HH size fell from 2.4 to 2.0 in 1950-2023 ([CPS, via FRED](#))

- HH size changes affect distributional measures
(Auten and Splinter [2024](#))

Missing market income in tax data

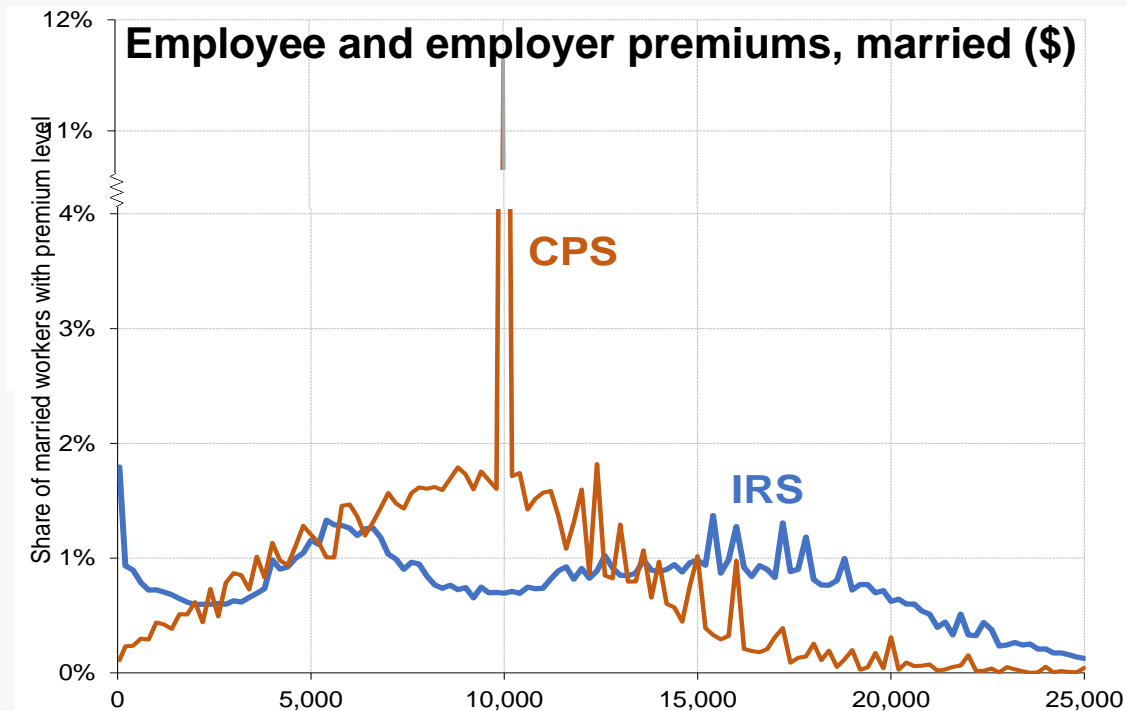
1960-2019 (relative to national income + transfers)



Market income in tax data

Private Health Insurance

- Most employer insur. on W-2s & self-emp. on tax returns
- Tax data better captures expensive insur. plans than CPS



Source: Larrimore, and Splinter (2019)

Integrating data from multiple sources

Neither administrative data nor survey data are perfect

- Survey data have reporting error and non-response
 - Acute in tails of the distribution (Bollinger et al. [2018](#))
 - Under-captures sources like retirement (Bee and Mitchell [2017](#)) and unemployment insurance (Larrimore, Mortenson, and Splinter [2023](#))
- Administrative data changes based on laws/regulations and misreporting incentives
 - Sharp increase in top incomes after TRA-86 lowered individual rates below corporate rates
 - Bunching around tax kinks (Mortenson and Whitten 2020)

Tax data should not necessarily always be default data

Need a plan if administrative data availability changes