Emmanuel Saez
(UC Berkeley)

September 2017

Distributional National Accounts: Public Use Micro-Files

Large disconnect between the study of inequality and macro growth

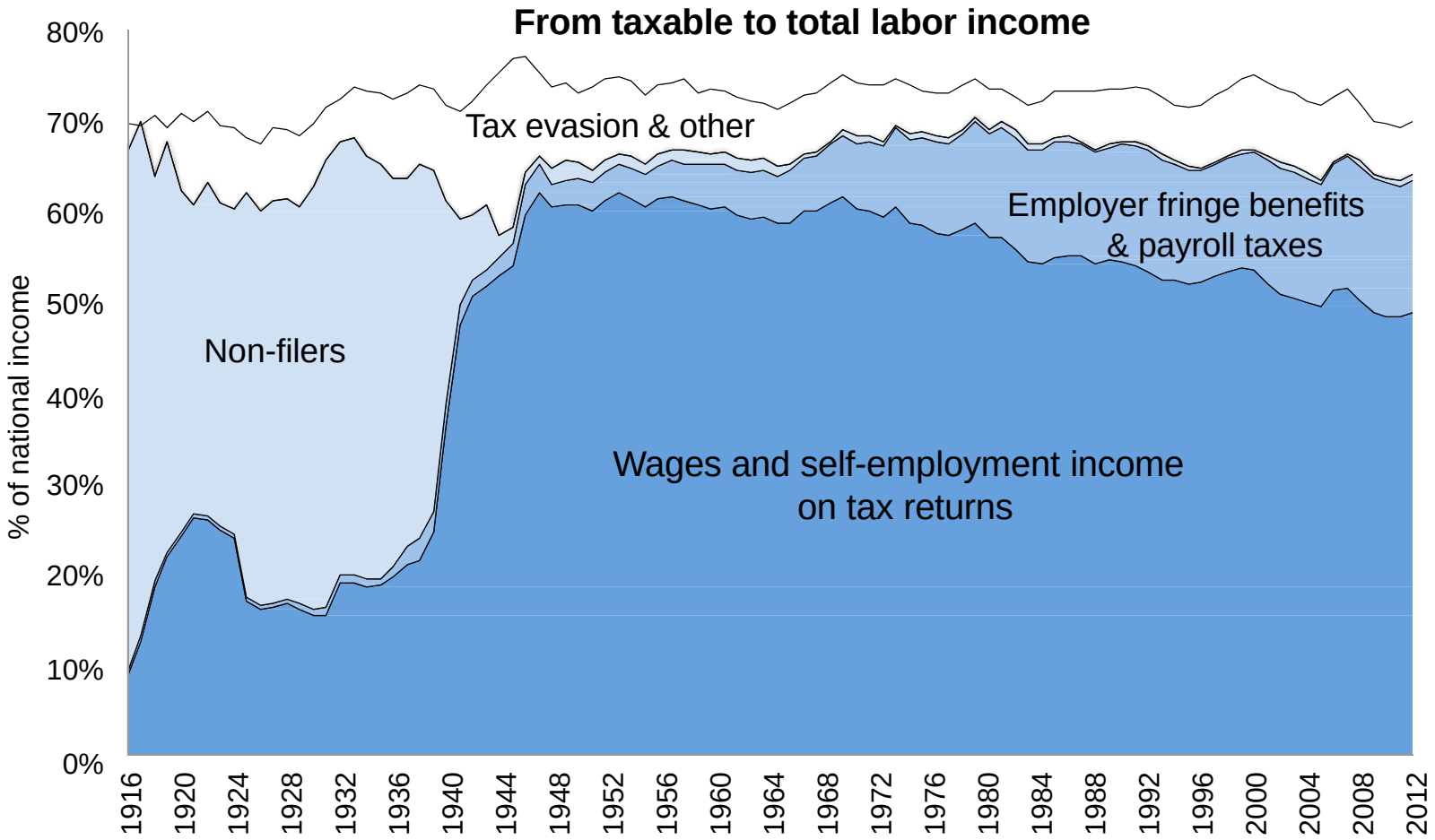
Macro: national accounts, no distributional information

Inequality: use survey & tax data, inconsistent with macro totals: Total AGI is only 2/3 of National Income

This gap makes it hard to know how growth is distributed

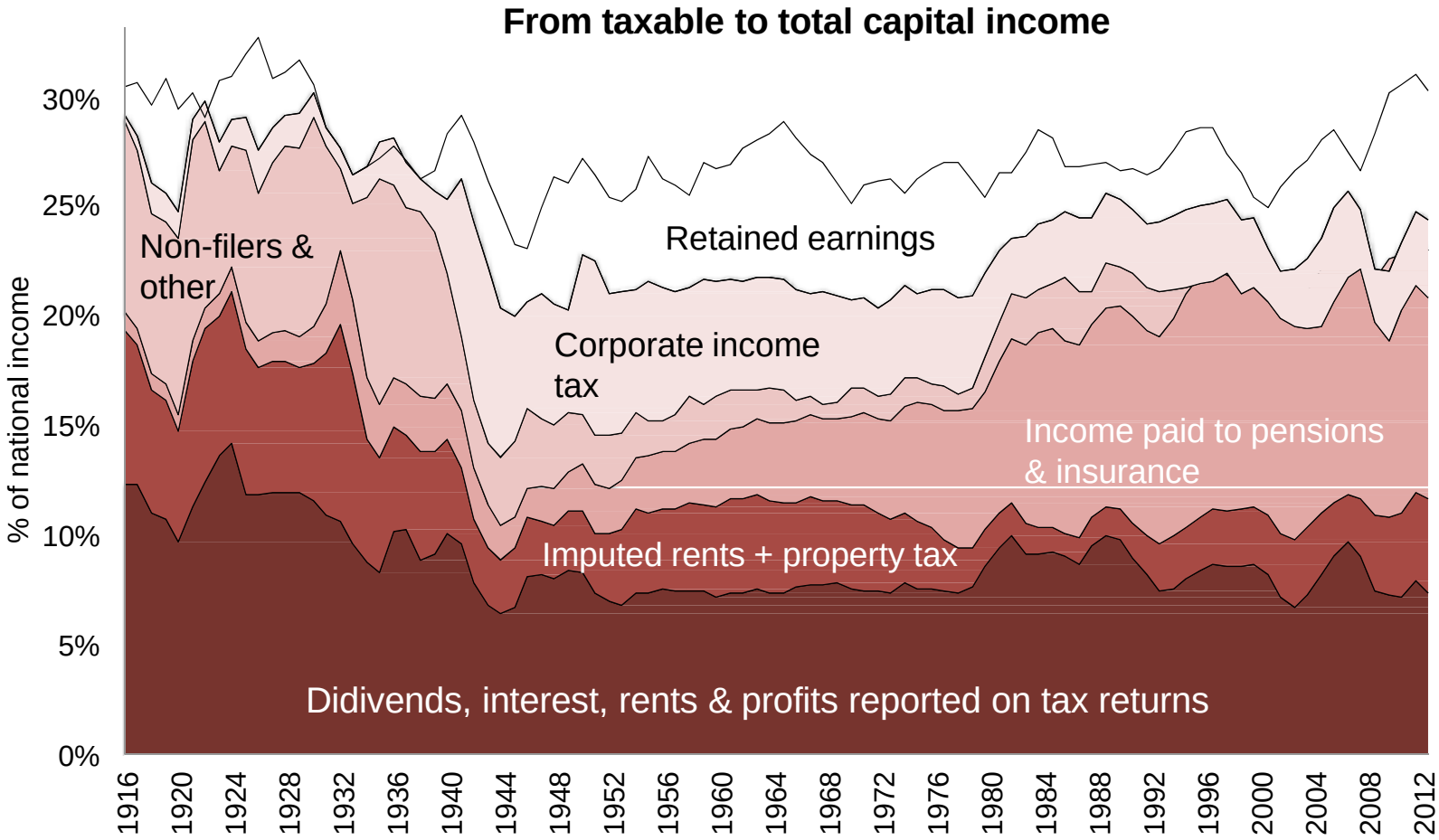
- Growth of bottom 50%, middle 40%, top 10% vs. average growth?
- How do taxes and govt spending affect distribution of growth?
- How to compare growth and inequality across countries?

1/4 of labor income in National Income missed by tax data



Source: Appendix Table I-S.A8b.

2/3 of capital income in National Income missed by tax data



Source: Appendix Table I-S.A8.

Piketty-Saez-Zucman (2016) construct Distributional National Accounts (DINA)

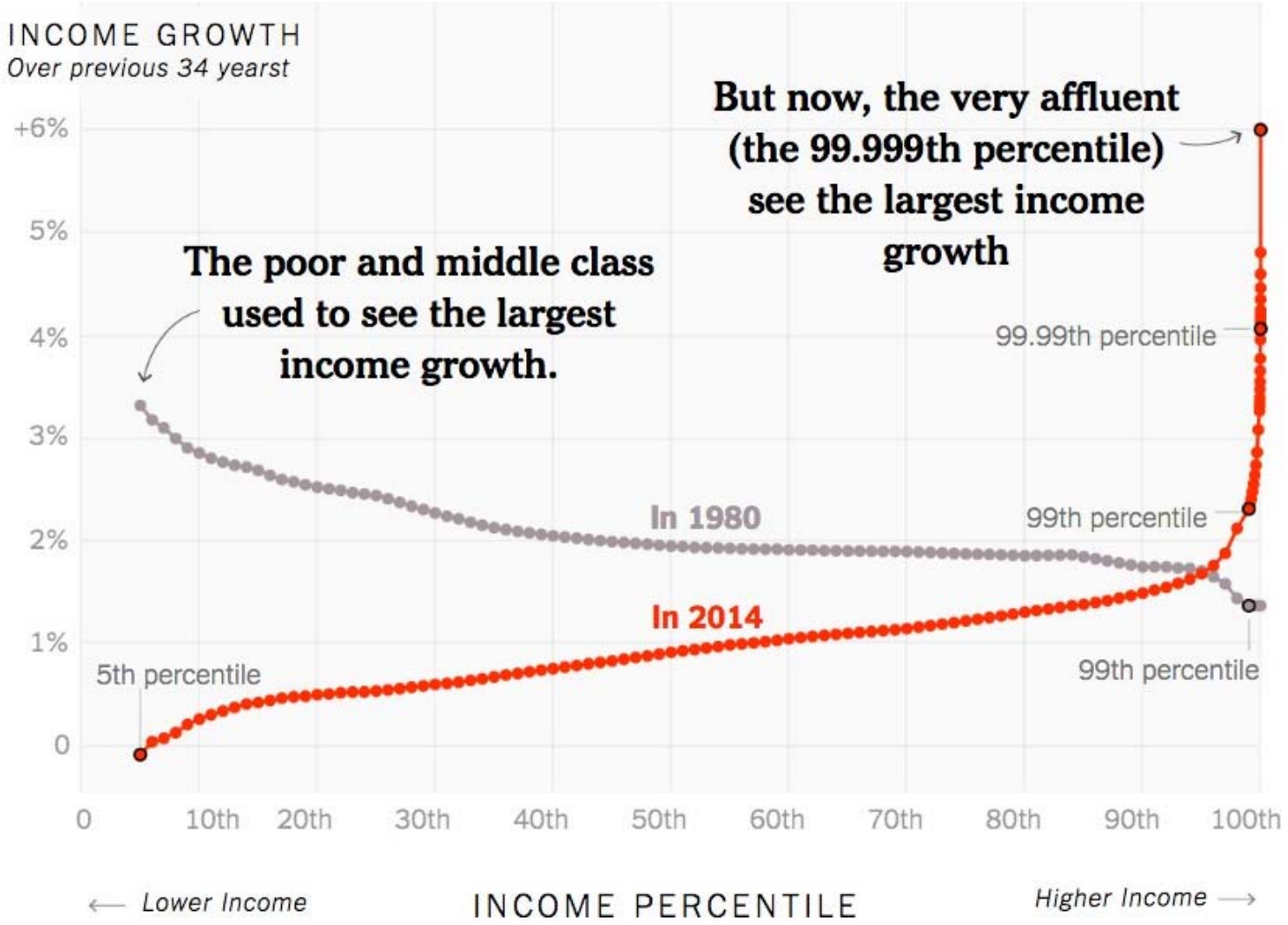
Annual **micro databases** since 1962 (tabulations for 1913-1961) of income, wealth, taxes and transfers consistent with national accounts totals in the US:

- Data cover full population and 100% of national income
- Growth statistics of income per adult by quantiles consistent with macro growth
- Assess redistributive effects of govt taxes, transfers, and spending by computing both pre-tax and post-tax incomes

Changes in standards of living per adult in the United States since 1946

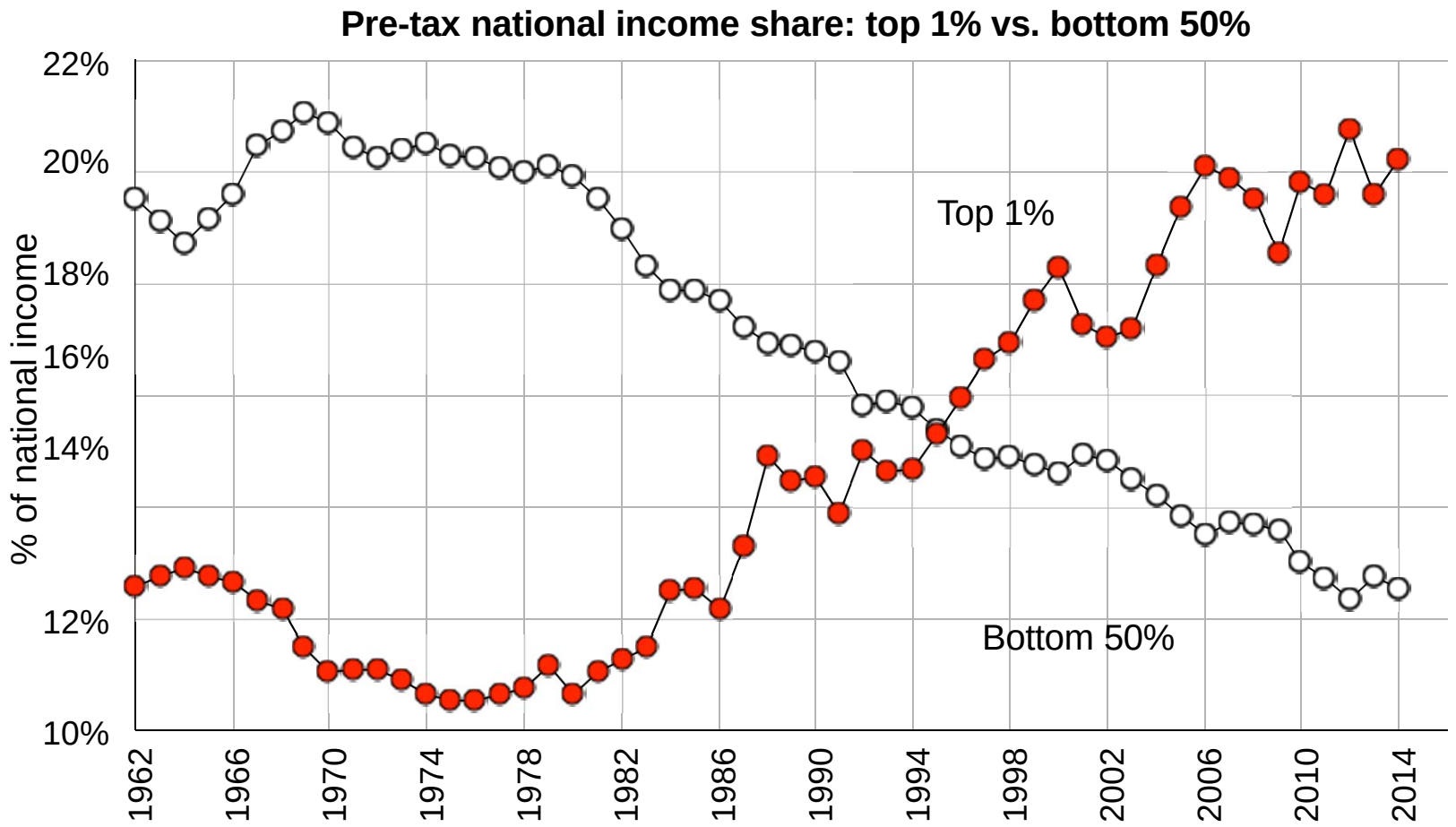
	Pre-tax income growth		Post-tax income growth	
Income group	**1980-2014**	**1946-1980**	**1980-2014**	**1946-1980**
Full Population	61%	95%	61%	95%
Bottom 50%	1%	102%	21%	130%
Middle 40%	42%	105%	49%	98%
Top 10%	121%	79%	113%	69%
Top 1%	205%	47%	194%	58%
Top 0.1%	321%	54%	299%	104%
Top 0.01%	454%	75%	424%	201%
Top 0.001%	636%	57%	617%	163%

NYT: Post-tax Annual Income Growth by Percentile, 1946-1980 vs. 1980-2014



Note: Inflation-adjusted annual average growth using income after taxes, transfers and non-cash benefits.

Fall of bottom 50% income share mirrors rise of top 1% income share



Source: Appendix Table II-B1

Broad methodology

1. Start with individual tax files created by SOI

- High quality annual records since 1962 that oversample top

2. Use additional tax data

- Age and gender information since 1979 from DM1
- Earnings split for couples & benefits on W2s since 1999

3. Impute missing income and nonfilers

- Non-taxable capital income (pension funds, imputed rents): SCF
- Government transfers: imputed based on CPS
- Records for nonfilers created based on CPS

Creating a Public Use Version

DINA database most useful and influential if there is a **distributable public use version** of the micro-files

We have constructed an external version of the micro-files using exclusively public use sources: t

- PUF tax files (1962-2011) SOI published tabulations
- Special additional tabulations from SOI working paper Saez and Zucman (2016) on age, gender, earnings split within couples

External files generate statistics highly comparable to internal files (Saez and Zucman, 2016)

Disclosing a Public Use Version

Next step: generate synthetic records from the external files to create freely disclosable files

On-going work supervised by John Czajka through his SOI contract keeping Barry Johnson and Mike Weber at SOI informed of the plan

Goal is to apply sufficient blurring/synthesizing of PUF records to meet modern disclosure rules for tax statistics publications

We hope to be able to complete this work in the Fall 2017

Steps To Produce Public Use Version

1. **Select about 20 key variables** in each PUF homogeneous across years 1962-2011. Key demographic variables and main income and tax variables
2. **Apply 2011 PUF disclosure avoidance procedures** retrospectively to the older PUF files
   - Subsampling of high income records
   - Aggregate records grouping together returns with extreme values
   - Blurring of some variables
3. **Aggregate individual records in groups of 5** chosen to be similar along some key family status and income variables

Next Step: testing and aging files

1. **Creating and testing disclosable DINA files** Repeat DINA estimations with disclosable files and test systematically statistics against stats based on internal data
2. **Create 2012-2015 files** by projecting 2011 PUF records using SOI tabulations
   - Partition records by AGI × status cells in 2011
   - Blow up 2011 records to match 2015 counts and $ averages for all income variables
   - Simple method respects totals and basic distributional properties
3. **Create 2016 file** by projecting from CDW preliminary tabulations

Next Step: State level files

State level DINA files: Easiest way to compute National Income and its distribution across residents

INSOLE files are too small to produce representative files at State level for small States and PUF files no longer have State indicators

CDW databank could be used to produce state level DINA files (aggregating by cells of 10 records)

State files could be of great use for simulating state level policy changes