The World Top Incomes Database – WTID: An Assessment

Andrea Brandolini

Bank of Italy, DG Economics, Statistics & Research

Journal of Economic Inequality panel
"Appraising World Income Inequality Databases"

Overview and background

Tax data and inequality

- Tax-based statistics on income inequality always seem to hit a nerve of the economics profession
  - Worldwide debate on Pareto's Law (1895)
  - Kuznets (1953): change in US income distribution "... for its magnitude and persistence ... unmatched in the record" → Burns: "one of the great social revolutions of history"
  - Atkinson, Piketty, Saez (2003-2010) → "The One Percent"

- But not much used for inequality analysis: limited coverage & problems with income definition

- **Surprising it has taken until 2011** to exploit known and accessible data such as tax records to construct an **international inequality database**

The World Top Incomes Database

- International database released in 2011 by Alvaredo, Atkinson, Piketty and Saez, in collaboration with large group of researchers

- **Shares of personal income received by the richest groups of taxpayers**

  – **29 countries** from all continents over a period of **143 years**: 1870 (Denmark) to 2012 (many)

- All WTID figures are computed from **tax records**, except for China and more recent Finnish figures

The WTID project

- **Dynamic**: old series extended forwards and backwards and revised, new series for other countries, plan for new series on earnings and wealth

- **Cooperative**: everyone being "aware of data that have not been exploited yet" or having the possibility of "getting unpublished data for so-far uncovered countries (even if only for a few recent years) is invited to collaborate

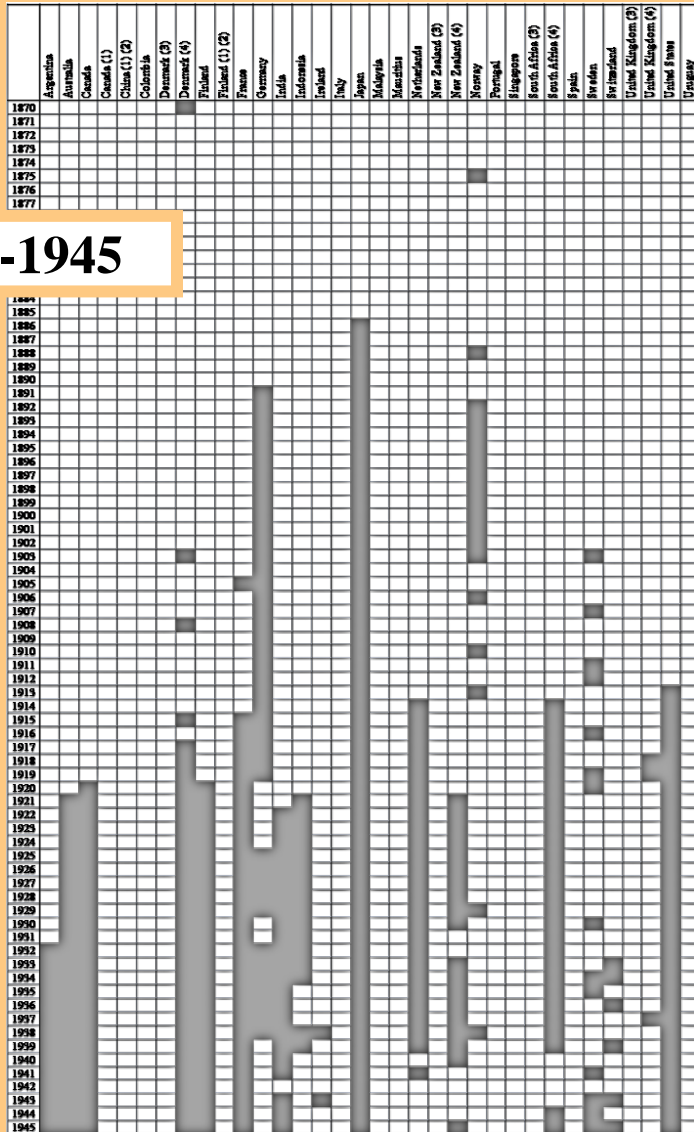- **Easy access**: freely available database hosted by the Paris School of Economics :

  http://topincomes.g-mond.parisschoolofeconomics.eu/#Home:

WTID coverage (September 2014)

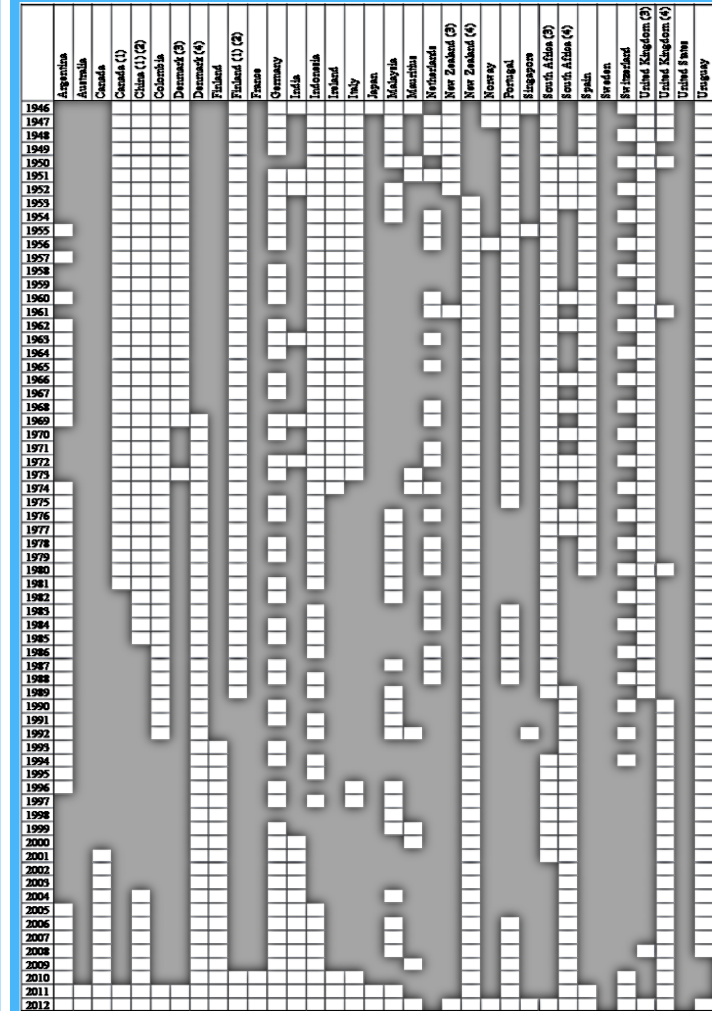available series | work in progress
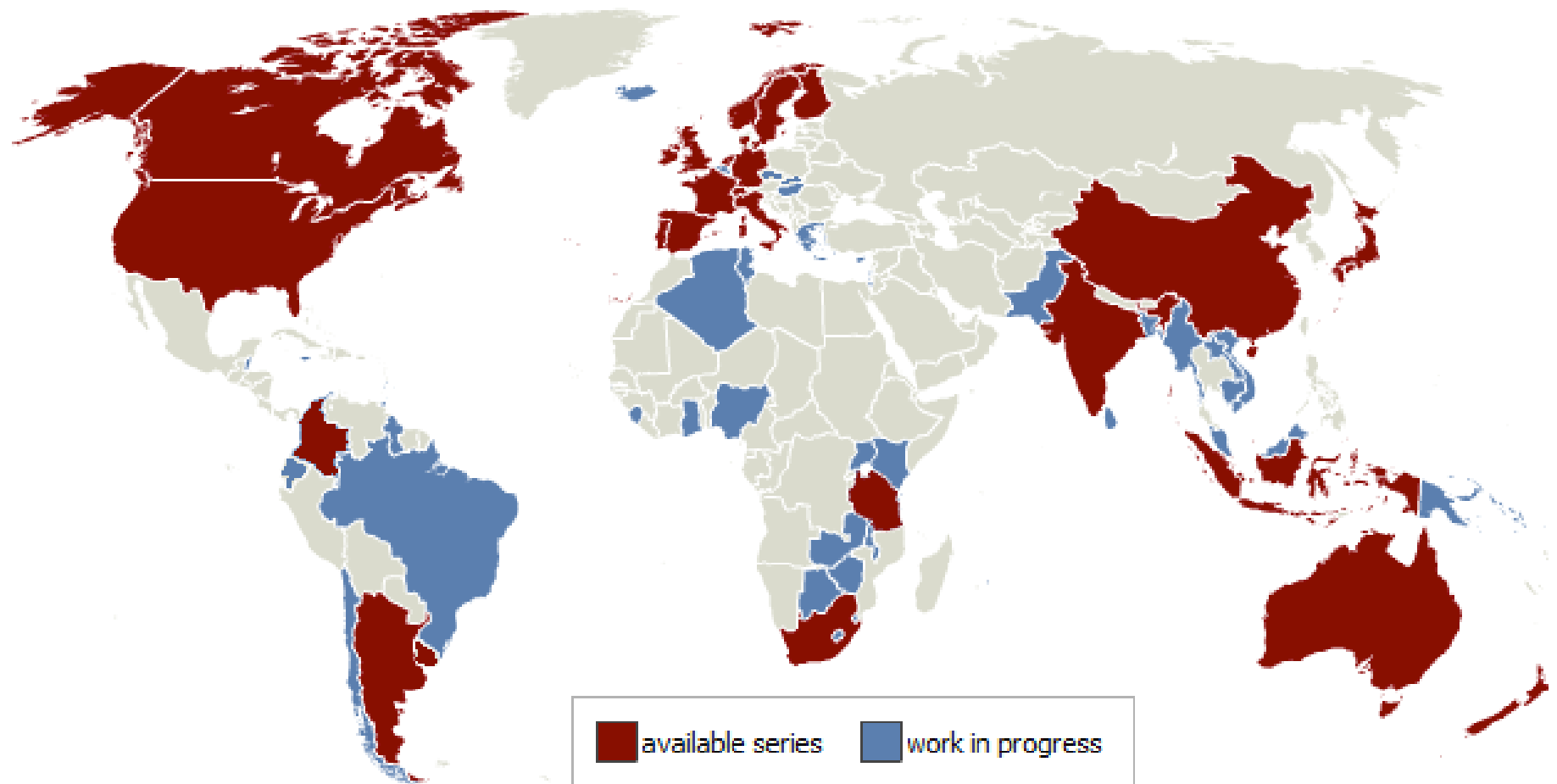
Source: WTID.

WTID top 1% share

1870-1945

1946-2012

WTID: strengths & weaknesses

- Strengths
  - **length**: more than 1/4 of 1718 values for top 1% share are pre-1946
  - **density**: relatively fewer missing values
  - coverage of **highest income earners**

- Weaknesses
  - income definitions reflect **administrative rules**
  - reference unit is **taxpaying unit**
  - **discontinuities** due to changes in tax legislations
  - **tax avoidance** practices
  - coverage of top fractions of population allows estimating **only a class of inequality statistics**, namely income shares of the richest taxpayers

Methods and problems with estimating inequality from tax data

WTID: raw material

- Example of **income tax data**: UK 1911-12
  (from Atkinson-Piketty-Saez, *Journal of Economic Literature* 2010)

Income class		Number of persons	Total income assessed
At least	but less than		
£5,000	£10,000	7,767	£52,810,069
£10,000	£15,000	2,055	£24,765,153
£15,000	£20,000	798	£13,742,318
£20,000	£25,000	437	£9,653,890
£25,000	£35,000	387	£11,385,691
£35,000	£45,000	188	£7,464,861
£45,000	£55,000	106	£5,274,658
£55,000	£65,000	56	£3,295,110
£65,000	£75,000	37	£2,590,606
£75,000	£100,000	56	£4,929,787
£100,000	—	66	£12,183,724
Total		11,953	£148,095,867

WTID: method

- Kuznets, *Shares of Upper Income Groups in Income and Savings,* NBER, New York (1953)

  The basic procedure is to compare the number and income of persons represented on federal income tax returns with the total population and its income receipts.

  Since except for a few recent years, tax returns cover only a small fraction of total population – the fraction at the highest income levels – our estimates of income shares are for only a small upper sector.

- Atkinson, 'Measuring Top Incomes: methodological issues, in Atkinson-Piketty (2007)

WTID: measurement issues

- **Inequality measure**: share of richest population fraction

- **Control totals for population**
  - Individual taxation: no. adults 20+ (lower bound for income share) or 15+ (upper bound)
  - Family taxation: subtract no. of married females

- **Control totals for income**
  - Income tax data + estimated income of 'non-filers'
  - External control total (national accounts)

- **Interpolation** (for data in grouped tabulations)
  - Fit a Pareto distribution, $1 - F(y) = (k/y)^\alpha \ (k > 0, \alpha > 1)$ but there are alternatives
    (e.g. Cowell-Mehta, *Review of Economic Studies* 1982)

WTID: population controls

- Atkinson (2007)
  - Control larger by $(1+c)$ means going further down distribution to locate top $x\%$, so that top income share up by $(1+c)^{1-1/\alpha}$, assuming Pareto distribution

- Applies to **choice of adult population**: e.g. individuals 20+ vs. individuals 15+, etc.

- Applies to **population revisions**: e.g. reconstruction backwards after 2011 Census reduces Italians 20+ by 1.8% in 2009 (but small impact: 1% fall in shares)

- **Variability across countries**, and a few changes within countries

WTID population controls

|	Number of adults	Number of tax units
Argentina	Individuals aged 20+.	Individuals aged 20+.
Australia	Individuals aged 15+.	Individuals aged 15+.
Canada	Up to 1999, individuals aged 20+; from 2000, tax filers.	Up to 1999, individuals aged 20+; from 2000, tax filers.
China		Individuals represented in the urban household surveys.
Colombia	Individuals aged 20+.	Individuals aged 20+.
Denmark	Until 1969: individuals aged 15+. From 1970: tax returns.	Until 1968: individuals aged 15+ minus married females. From 1970: tax returns (individuals aged 15+ plus those individuals below 15 years old who also file a tax return).
Finland	n.a.	Individuals aged 15+ minus married females until 1969; individuals aged 15+ from 1970.
France	Individuals aged 20+.	Families.
Germany		Single individuals aged 20+ plus one half of married individuals.
India		Individuals (40% of total population).
Indonesia		Households.
Ireland		Individuals aged 18+ minus married women.
Italy	Individuals aged 20+.	Individuals aged 20+.
Japan	Individuals aged 20+.	Individuals aged 20+.
Malaysia	Individuals aged 15+.	Individuals aged 15+.
Mauritius	Individuals aged 15+.	Individuals aged 15+ minus the number of married women and those not married but living together.
Netherlands		Number of individuals aged 15+ minus the min(number of married women, number of married men).
New Zealand	Until 1999: individuals aged 15+. From 2000: total individual taxpayers.	Until 1952: individuals aged 15+ minus married women. 1953-1999: individuals aged 15+ from 1953 to 1999. From 2000: total individual taxpayers.
Norway	Individuals aged 16+.	Individuals aged 16+. Despite joint taxation, separate filing increasingly prevalent. From 1998 Statistics Norway ceased to treat married couples with joint taxation as one taxpayer.
Portugal		Individuals aged 20+ minus married women.
Singapore	Individuals aged 15+.	Individuals aged 15+.
South Africa	Individuals aged 15+.	Until 1989: individuals aged 15+ minus married women. From 1990: individuals aged 15+.
Spain	Individuals aged 20+.	Individuals aged 20+.
Sweden		Until 1950: individuals aged 16+ minus married women. 1951-1970: individuals aged 16+ minus married women with low or no income. From 1971: indviduals aged 16+.
Switzerland	Until 1995: individuals aged 20+. From 1996: individuals aged 18+.	Until 1995: individuals aged 20+ minus one half of married men and women. From 1996: individuals aged 18+ minus one half of married men and women.
Tanzania	Individuals aged 15+.	Every tax unit is assumed to be comprised by 1.6 adults (individuals aged 15+).
United Kingdom		Until 1989: individuals aged 15+ minus married females. From 1990: individuals aged 15+.
United States		Families.
Uruguay	Individuals aged 20+.	Individuals aged 20+.

WTID: income controls

- Similar considerations apply to income controls, but impact likely to be larger

  – Variability of covered components: **non-taxable items vary across countries** (transfers, capital gains)

  – **Revisions** in external controls

  – Compare to **common external national accounts benchmark**

WTID income controls: Australia
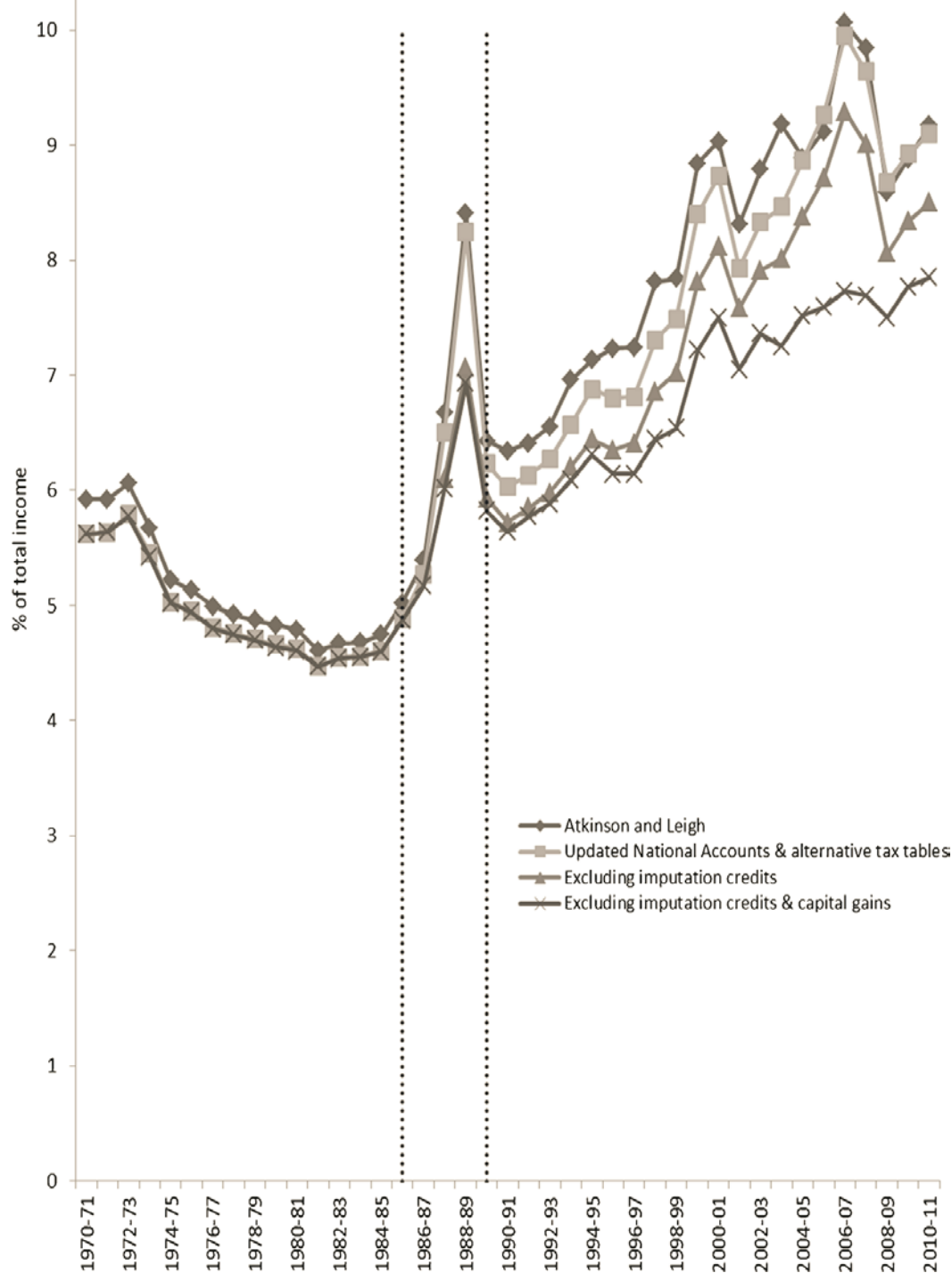
Source:
Burkhauser-Hahn-Wilkins,
Journal of Economic Inequality
forthcoming
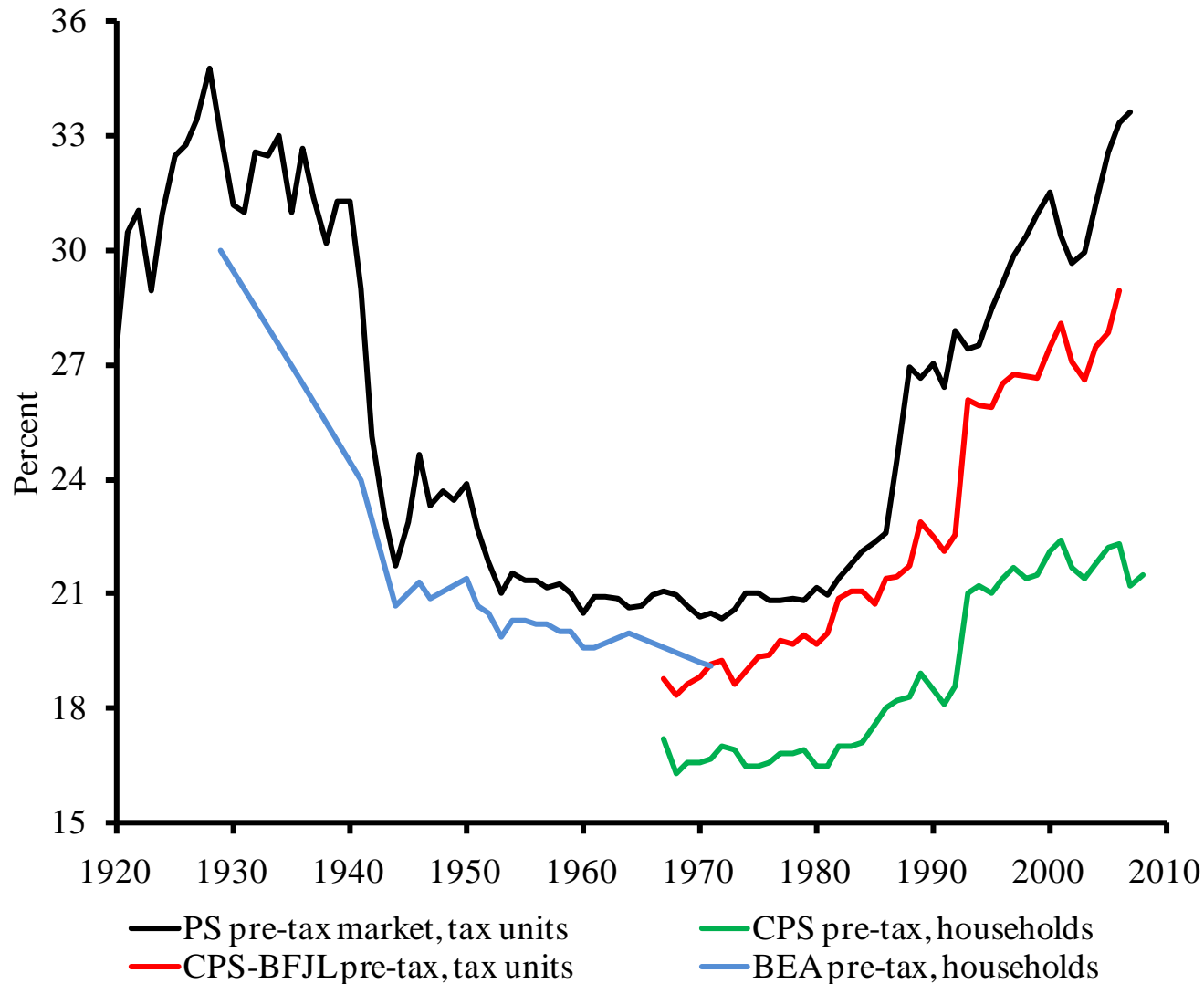
Taxable vs. household pre-tax NA income
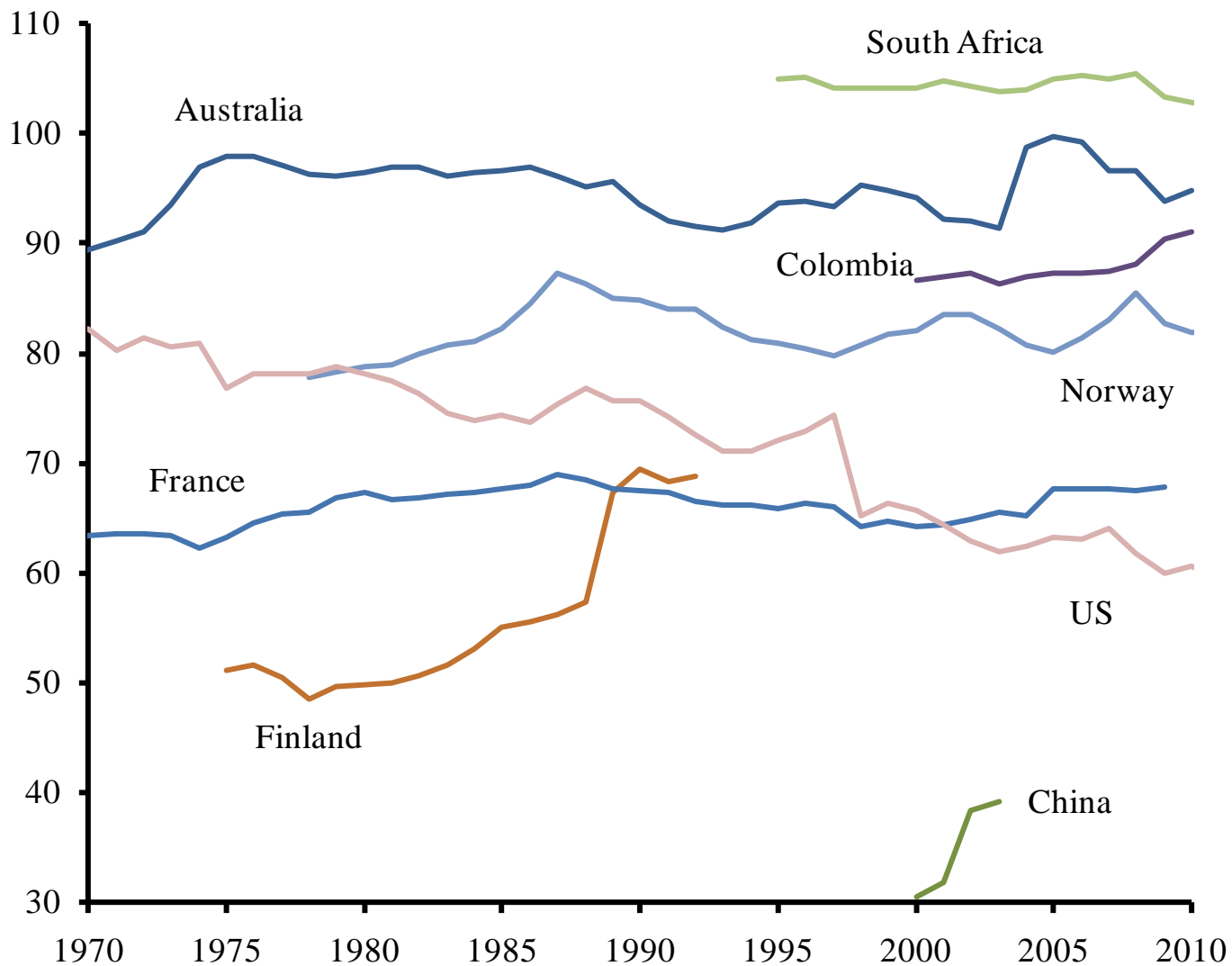
Source: author's elaboration on WTID and OECD data.

About source comparisons
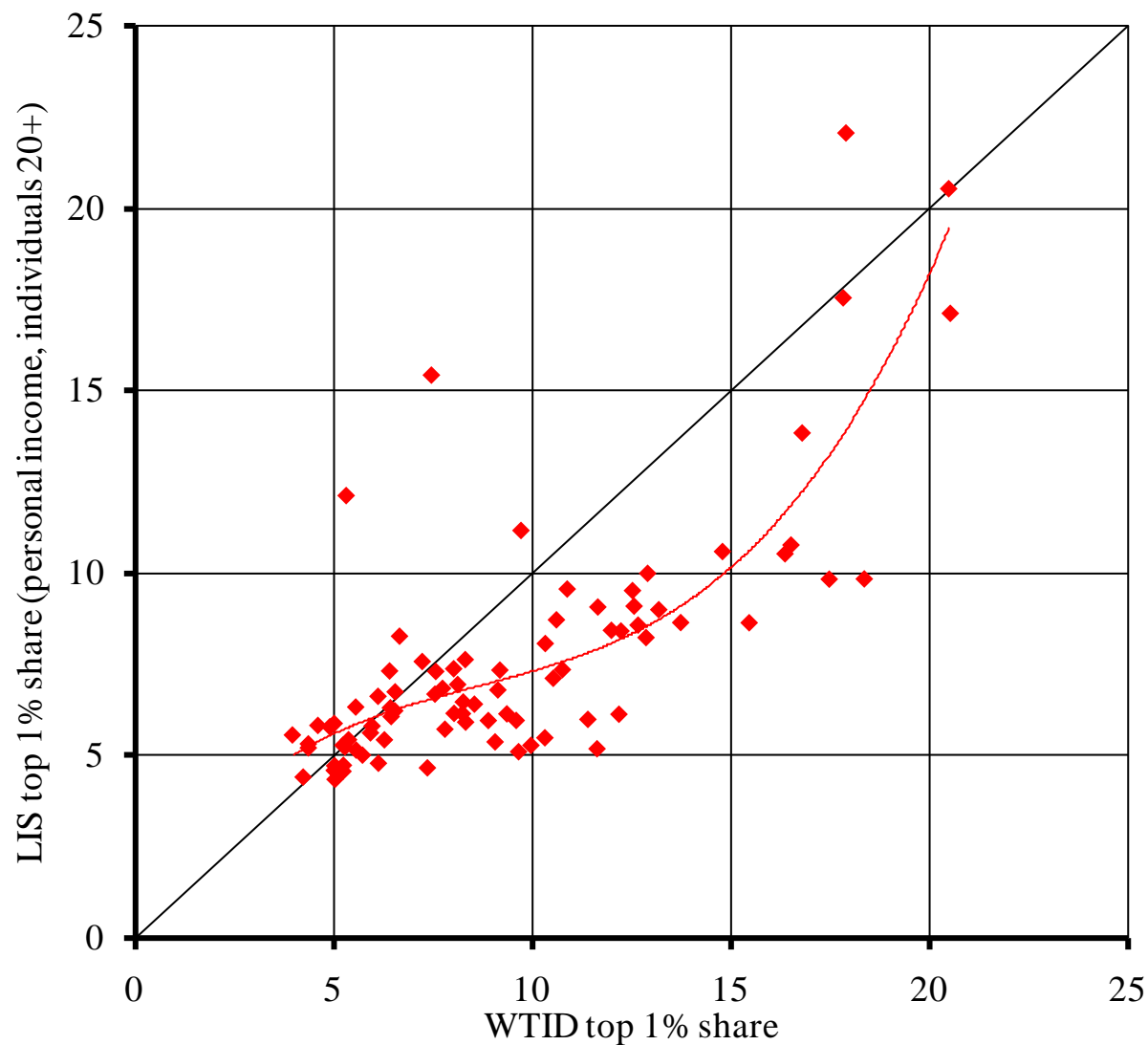
Tax-based vs. survey-based statistics

- Many good reasons to differ
  - Income definition, reference unit, coverage, measurement errors, ...

- **Do they tell different stories?**

- Leigh, *Economic Journal* 2007: strong and significant relationship between top income shares and broader inequality measures (Gini, Theil, etc.)
  - Comparison with measures from WIID (but: household gross/net income) and LIS (but: household disposable income)

- **Comforting, but work on reconciling definitions**

Top 5% share in the US

PS pre-tax market, tax units
CPS pre-tax, households
CPS-BFJL pre-tax, tax units
BEA pre-tax, households

Source: Brandolini, *Politics & Society* 2010.

Top 1% share: WTID vs. LIS (PI)

Source: author's elaboration on WTID and LIS data.

The WTID database,
and how it deals with problems

Using the WTID: website

- Take the perspective of a **user seeking ready-to-use data** – not so mindful to go through the 1,200 pages of the two Atkinson-Piketty et al. volumes

- Do website & database give **enough information?**

- **Website**: friendly and well-designed
    - **Noticeboard** informs on updates, new papers, etc.
    - Nice feature: **Preview** option, which allows users to see immediately a chart with updated time series
    - **Graphics**: user-friendly way to plot main variables
    - **Download**: easy selection of countries/years/ variables; entire dataset quickly downloadable in excel file

Using the WTID: dataset

- Excel file with sources and two different layouts
    - About **400 variables** for 29 countries
- Basic structure (layout B)

Variable code	1110201	3100101	3100201	4100101
	Argentina- Top 5% income share	Argentina- Number of adults	Argentina- Number of tax units	Argentina- Income control
Units	%	Thousand	Thousand	Nominal million Pesos
Notes	Estimates do not include capital gains. Statistics exclude taxpayers with wage income only.	Individuals aged 20+.	Individuals aged 20+.	
1870				
1871				
1872				
1873				

Using the WTID: dataset

- Excel file with sources and two different layouts
  - About **400 variables** for 29 countries
- Basic structure (layout B)

Variable code	1110201	3100101	3100201	4100101
	Argentina- Top 5% income share	Argentina- Number of adults	Argentina- Number of tax units	Argentina- Income control
Units	%	Thousand	Thousand	Nominal million Pesos
Notes	Estimates do not include capital gains. Statistics exclude taxpayers with wage income only.	Individuals aged 20+.	Individuals aged 20+.	
1870				
1871				
1872				
1873				

Codification rules not provided, but would be useful

Using the WTID: dataset

- Excel file with sources and two different layouts
  - About **400 variables** for 29 countries
- Basic structure (layout B)

Variable code	1110201	3100101	3100201	4100101
	Argentina- Top 5% income share	Argentina- Number of adults	Argentina- Number of tax units	Argentina- Income control
Units	%	Thousand	Thousand	Nominal million Pesos
Notes	Estimates do not include capital gains. Statistics exclude taxpayers with wage income only.	Individuals aged 20+.	Individuals aged 20+.	
1870				
1871				
1872				
1873				

One series for each top share (except tax unit/source change)

Using the WTID: dataset

- Excel file with sources and two different layouts
  - About **400 variables** for 29 countries
- Basic structure (layout B)

Variable code	1110201	3100101	3100201	4100101
	Argentina-Top 5% income share	Argentina Number of adults	Argentina Number of tax units	Argentina Income control
Units	%	Thousand	Thousand	Nominal million Pesos
Notes	Estimates do not include capital gains. Statistics exclude taxpayers with wage income only.	Individuals aged 20+.	Individuals aged 20+.	
1870				
1871				
1872				
1873				

Important information on population/income controls

Using the WTID: dataset

- Excel file with sources and two different layouts
  - About **400 variables** for 29 countries
- Basic structure (layout B)

Variable code	1110201	3100101	3100201	4100101
	Argentina-Top 5% income share	Argentina-Number of adults	Argentina-Number of tax units	Argentina-Income control
Units	%	Thousand	Thousand	Nominal million Pesos
Notes	Estimates do not include capital gains. Statistics exclude taxpayers with wage income only.	Individuals aged 20+.	Individuals aged 20+.	
1870				
1871				
1872				
1873				

Notes: important to asses comparability, but meagre

Conclusions

Developing WTID

- Streamline and release variable **codification rules**

- Provide more details on **taxable income items**, to give an idea of how much income is captured

- Measurement choices affect results, but only one series available in WTID: provide **lower and upper limits?**

- Include **original raw data** to stimulate application of alternative assumptions

- **Cross-validation** with other sources

- **Cross-country comparability** deserves further research

Few last general words

- **Exciting time**: a lot of new data on inequality

- Alvaredo-Atkinson-Piketty-Saez have rediscovered a rich gold vein that had been largely neglected since the gold rush of Pareto's time

  - Tax data have problems, but other data do as well

- Look at **sources as complementary**, rather than rival

  - Substantive – not only statistical reasons explain differences

  - Integrate micro information with national accounts

Kuznets, 1953
It is as if one tried to paint a fine picture with thick brushes and large blobs of somewhat mixed colors

But still better than a white page!

Thank you for your attention.