Why is Europe Less Unequal than the United States?

Evidence From Distributional National Accounts, 1980–2017

Thomas Blanchet[1,2] Lucas Chancel[2,3] Amory Gethin[2]

International Macro History Seminar
March 3, 2021

[1]Paris School of Economics – EHESS

[2]World Inequality Lab

[3]IDDRI

What Do We Know about Inequality in Europe?

Unlike some other parts of the world, there is quite a lot of data available on the distribution of income in Europe.

So why write a paper on the topic?

- Not because of a lack of data *per se*...
- ...but because existing data is scattered across a variety of sources, so it can be hard to make sense of it.

Disparate set of indicators:

- hard to compare
- hard to aggregate
- hard to tell consistent stories

Many Questions Lack a Clear Answer

⇒ **Literature has struggled to answer simple questions:**

- Is Europe as a whole more or less unequal than the United States?
- Is the difference between European and US inequality driven by pre-tax incomes or redistribution?
- Is European inequality driven by the distribution of income *between* or *within* countries?
- Which parts of the distribution have benefited the most from European growth?

⇒ **Difficulties monitoring of internationally agreed goals**

- European institutions' Pillar of Social Rights (2017)
- Sustainable Development Goals adopted by all European countries (2015)

This Paper: Distributional National Accounts

This paper is an attempt to address this problem by constructing **Distributional National Accounts (DINA)** for Europe since 1980.

Part of larger ongoing project in which we construct similar estimates for as many geographical areas as possible, over as long as possible.

This paper contributes to that project in several respects:

- Production of new estimates of the income distribution within the DINA framework.
- Methodological advances to build such estimates in spite of data limitations.

Conceptual Framework

A Short History of DINA

- Early 2000s: Piketty (2003) for France and Piketty & Saez (2003) for the United States revived an old literature (Kuznets, 1953) that used tax data to estimate inequality.
- Work extended to large number of countries by many researchers, and collected in two collective volumes edited by Piketty & Atkinson (2007, 2010).
- Data collected into the World Top Income Database (WTID), which then became the **World Inequality Database (WID.world)**, maintained by the World Inequality Lab.
- Estimated based on raw tax data have many qualities, but also many drawbacks:
  - Inconsistent concepts
  - Only covers the top of the distribution
  - Does not take redistribution into account

DINA in Broad Strokes

- The DINA project was created to address those criticisms.
- Set of guidelines for measuring inequality:
    - consistent concepts.
    - consistent methodology.
    - consistent with the framework put forth by the System of National Accounts (SNA).
- Combine existing sources (surveys, tax data, national accounts).
- Distribute consistently 100% of national income.
- This paper: construct DINA estimates for 38 countries over 1980-2017

DINA concept: net national income

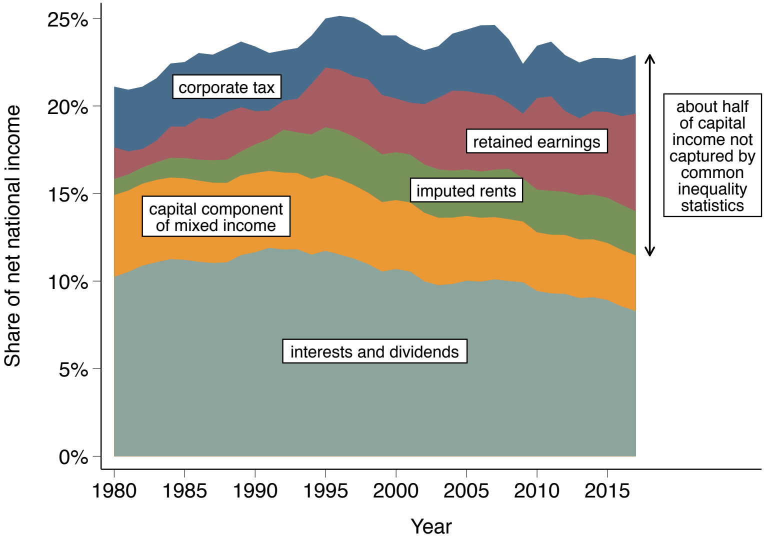
Distributing all of national income means accounting for income that may never explicitly appear on the bank account of any household, yet is still part of "economic growth".

net national income = GDP - depreciation + net foreign income

We consider income from all sectors of the economy:

- Income of the household sector (incl. imputed rents)
- Income of the government:
    - before taxes: production taxes
    - after taxes: taxes - transfers + government consumption
- Income of corporations (retained earnings)

Standard Inequality Statistics Miss a Large Part of Income



WID.WORLD

Yet "missing income" is still income

Not accounting for income outside of the household sector sometimes yields undesirable results:

- The owners of corporation can choose to distribute income to themselves arbitrarily.
- Some taxes are accounted for (e.g. income tax) but not others (e.g. VAT).
- Countries with strong provision of public goods appear to have poorer households.

We make simple and transparent assumptions to distribute these types of income. Still a lot of room for improvement.

DINA concepts: pre-tax and post-tax income

- **pre-tax income:** income after the operation of social insurance systems (pension and unemployment insurance), but before other types of transfers. We add pensions and unemployment benefits, but remove the social contributions that pay for them. Conceptually similar to "taxable income" in many countries.
- **post-tax:** income after the operation of all government redistribution. We remove all taxes, the remaining social contributions, and then add all transfers and government consumption.

Methodology

A myriad of income inequality datasets in Europe

Macro data:

- UN, OECD, Eurostat

Survey microdata:

- Eurostat surveys: SILC, ECHP, HBS
- Luxembourg Income Study (LIS)

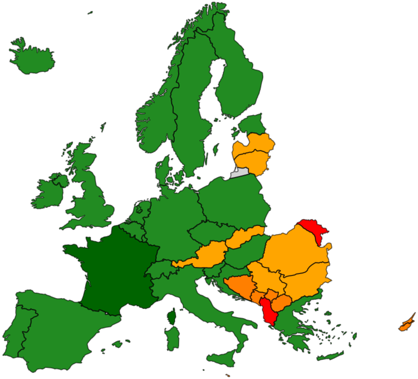
Survey tabulations:

- Diverse sources, compiled in WIID by the UN.
- We turn them into complete distributions using generalized Pareto interpolation (Blanchet, Fournier & Piketty, 2018)

Tax data:

- WID.world
- Couple of formerly unused sources: Iceland, East Germany.

Data availability: good coverage for most of the population



Low (survey tabulations, low time coverage)	Medium high (survey tabulations + microdata + tax data)
Medium low (survey tabulations, good time coverage)	High (distributional national accounts)
Medium (survey tabulations + microdata)	

Different factors explain discrepancies between sources

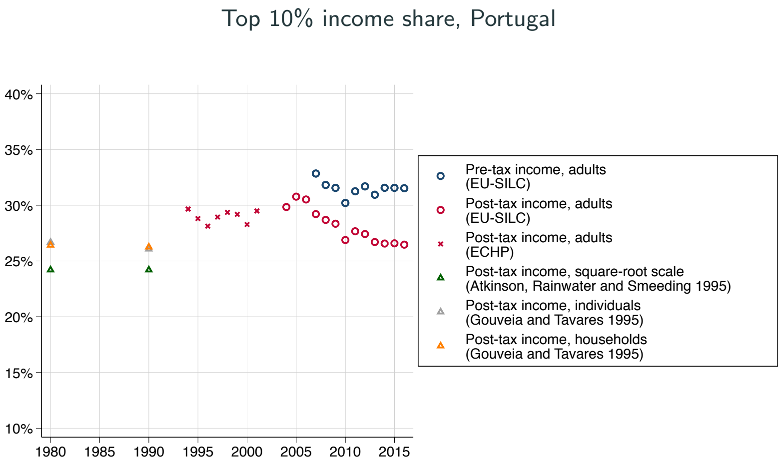
Three reasons why these data sources yield different results:

- **Conceptual differences:** data refers to different concepts:
  - income concepts
  - statistical unit: individuals, households, per capita, per adults, equivalence scales
- **Underestimation of top incomes:** heterogeneous non-response in surveys, misreporting, sampling error.
- **Missing income**: retained earnings, imputed rents.

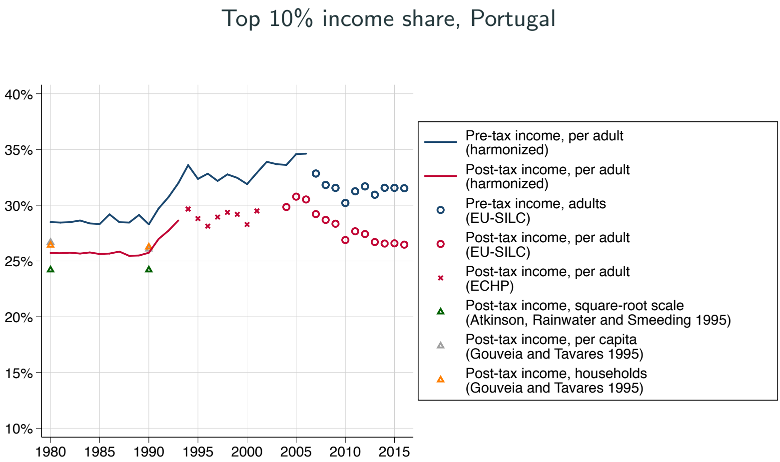
1) Harmonizing concepts using machine learning

- Use existing survey data to establish correspondences between the different concepts used in the literature:
  - **3 income concepts:** pre-tax income, post-tax income, consumption
  - **5 statistical units:** households, per capita, per adult, OECD equivalence scale, square root equivalence scale
  - $3 \times 5 =$ **15 distributions by year and country**
- We train a **machine learning algorithm** (tree boosting) to map how all these different concepts relate to one another, and then use it to correct for systematic biases.

Harmonization of Concepts: Example



Harmonization of Concepts: Example



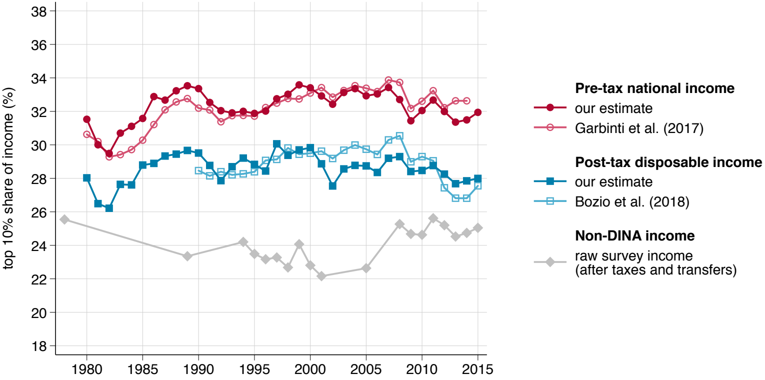
2) Correcting for misreporting and nonresponse using tax data

- Adapt the standard survey calibration framework (age, gender, ...) to improve the representation of the rich (Blanchet, Flores & Morgan, 2018).
- Adjust weights to enforce constraints on the top income shares while minimizing distortion from the original survey:
- Empirically, the reweighting depends on the inequality in the survey.

3) Distribution of incomes not captured by survey or tax data

- **Imputed rents**: statistical matching using the distribution of imputed rents recorded in EU-SILC.
- **Retained earnings**: proportional to stock ownership using statistical matching with data from HFCS.
- **Taxes on products and production**: proportionally to consumption using statistical matching with data from HBS.
- **Corporate income tax**: similarly to undistributed profits.
- **In-kind transfers**: public health spending in a lump-sum way; other types of in-kind transfers (police, infrastructure, etc.) proportionally in our benchmark series, as in earlier DINA studies.

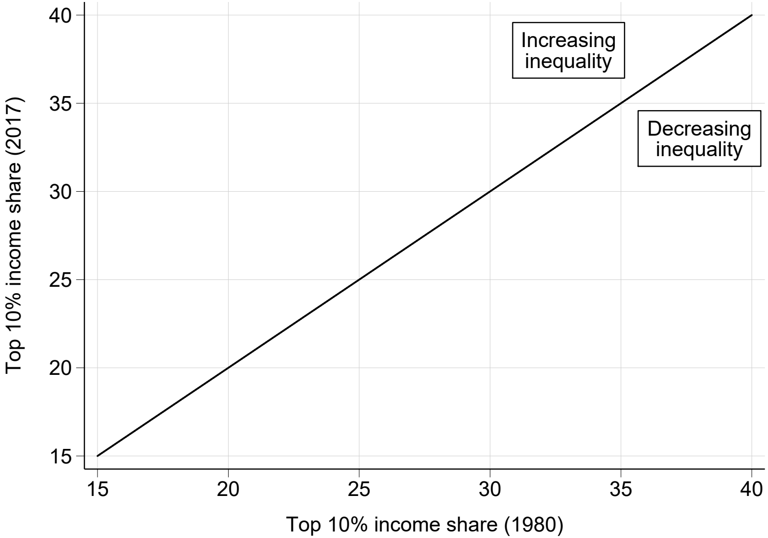
Validation: top 10% share in France



Results - The evolution of pretax income inequality

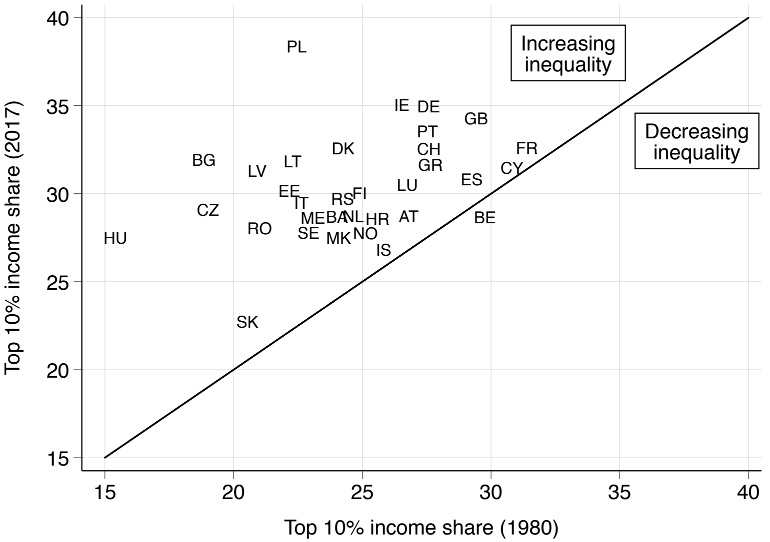
Inequalities within European countries: 1980 vs. 2017

Top 10% income shares across European countries: 1980 vs. 2017

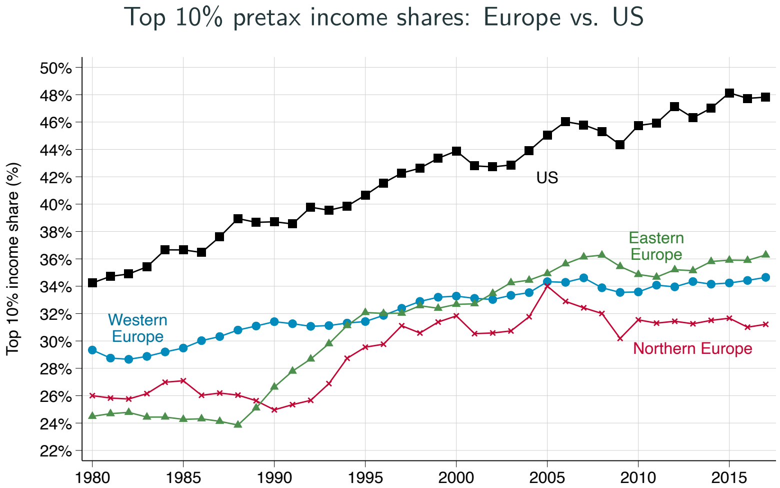


Inequalities increased in nearly all countries...

Top 10% income shares across European countries: 1980 vs. 2017

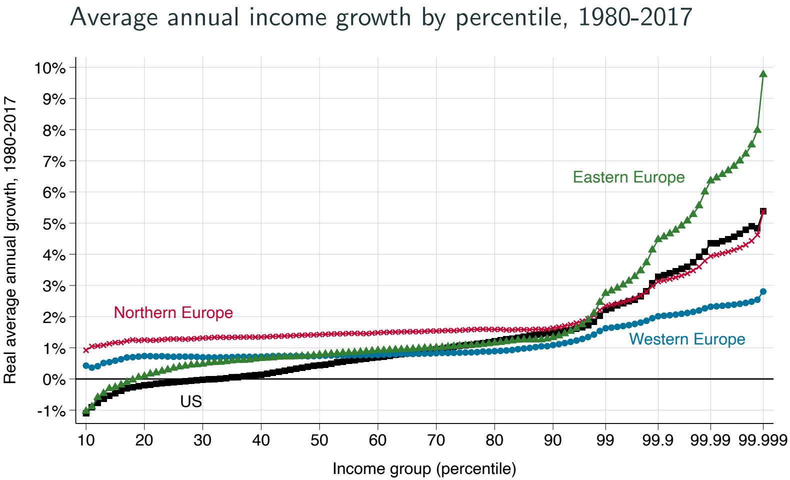


… but much less than in the US



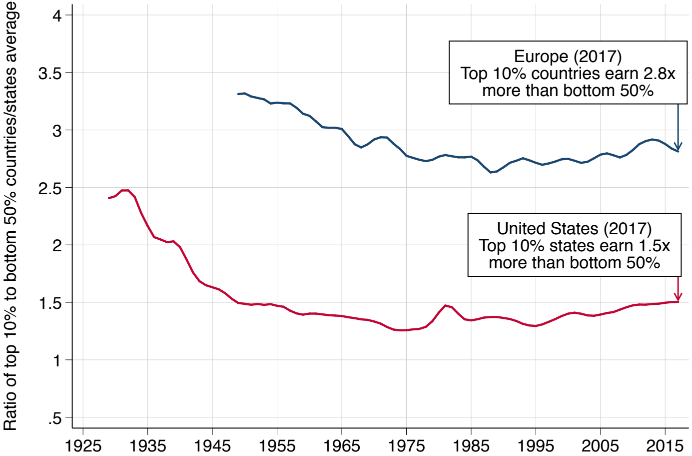
WID.WORLD

The distribution of pretax income growth



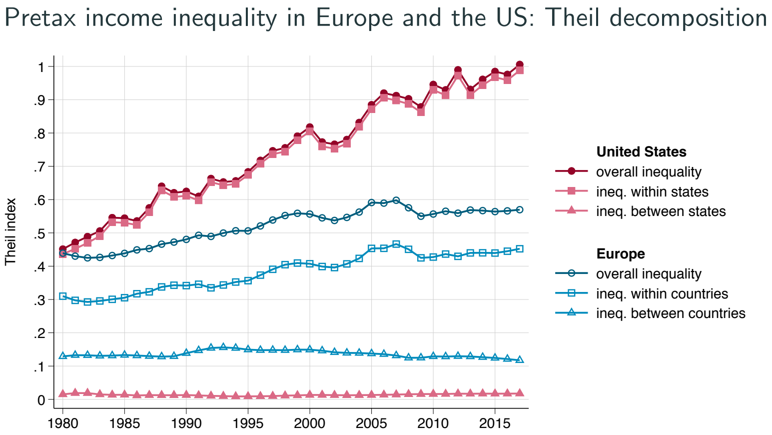
Spatial inequalities remain stronger in Europe as a whole than in the US...

US vs. Europe ratio of top 10% to bottom 50% average state / country incomes, 1929-2017



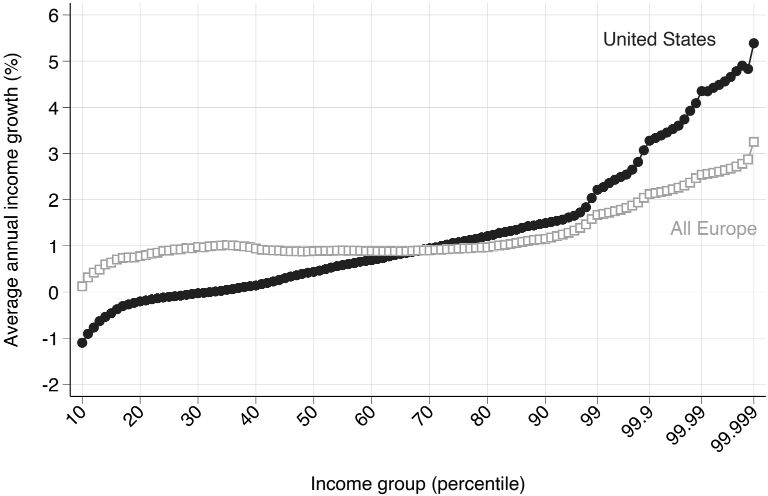
WID.WORLD

... yet inequality is still much lower in Europe as a whole than in the United States



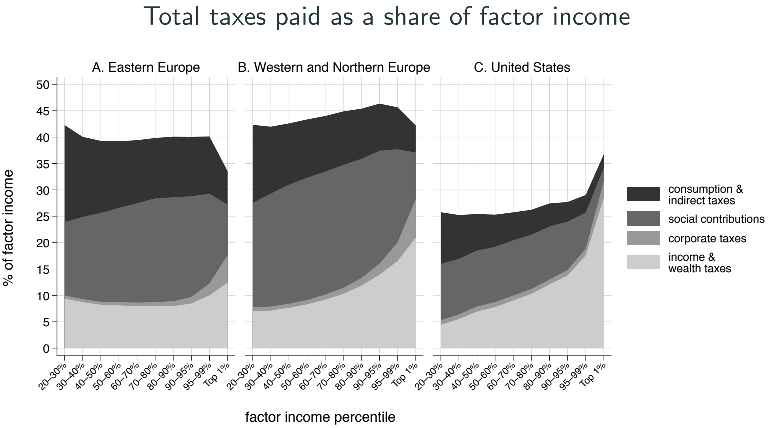
The distribution of pretax income growth: All Europe vs. US

Average annual income growth by percentile, 1980-2017



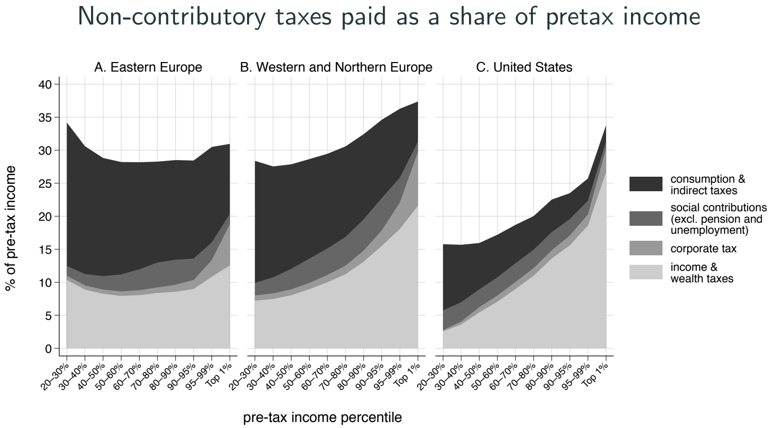
Results - The distributional incidence of taxes and transfers

Taxes in the US are lower... but not less progressive!



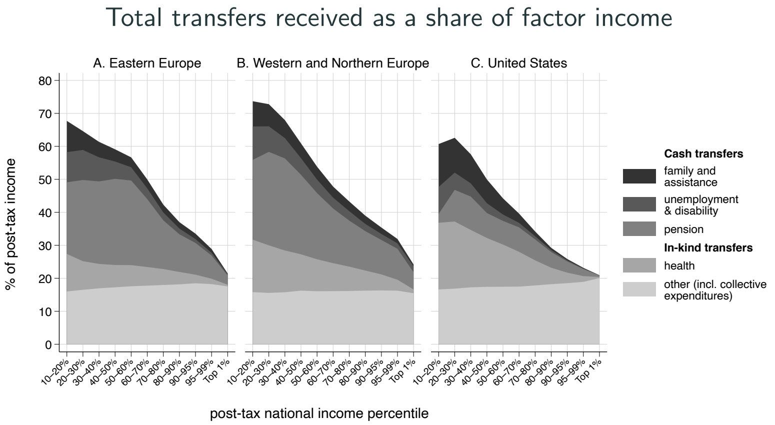
WID.WORLD

Taxes in the US are lower... but not less progressive!

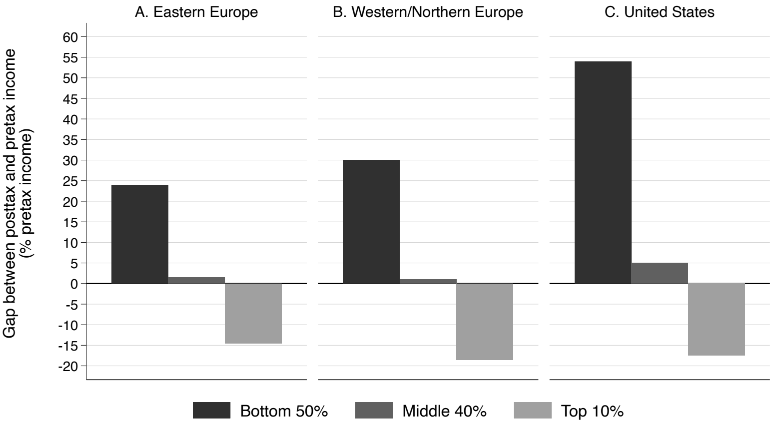


WID.WORLD

Transfers in the US are slightly lower... but not less progressive!

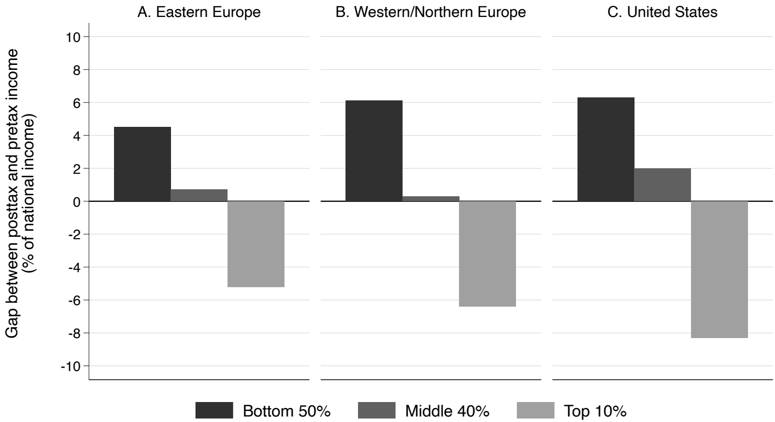


Net redistribution (% of group average income)



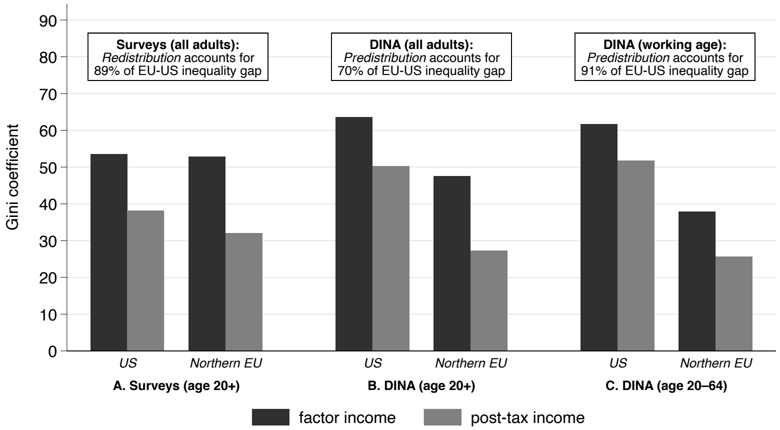
WID.WORLD

Net redistribution (% of national income)



WID.WORLD

Revisiting the basic facts: the US is more unequal than Europe because of pretax inequality, NOT redistribution



Conclusion

- A new public database to study income inequality in Europe fully consistent with national accounts.
- Inequality increased in nearly all EU countries since 1980, but at different speeds.
- Despite rising income disparities, EU still much more equal than the US.
- Contrary to a commonly held view, the EU-US gap can be explained by predistribution, *not* redistribution.
- Calls for renewed attention to policies reducing market income inequality (education, minimum wages, etc.), although tax system may also affect the distribution of pretax income (incentives, bargaining, etc.).

$\Rightarrow$ https://wid.world/

Appendix

Harmonization of Concepts: Methodology

- Discretize the quantile function of distributions, normalized by the average: $Q_1, \ldots, Q_n$
- Statistical model between distributions A and B:

$$Q_k^A = f(p_k, Q_1^B, \ldots, Q_n^B, Z) + \varepsilon_k$$

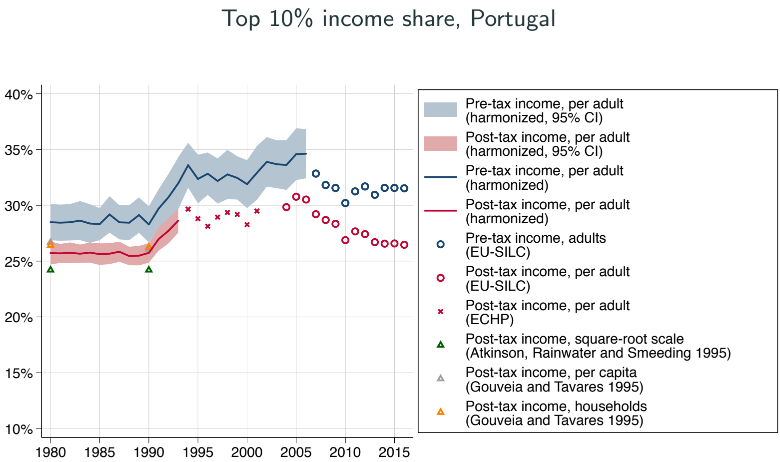
where Z is a set of additional variables to improve prediction (regional dummy, average income, demographic composition, tax schedule, etc.)

The problem is hard statistically because:

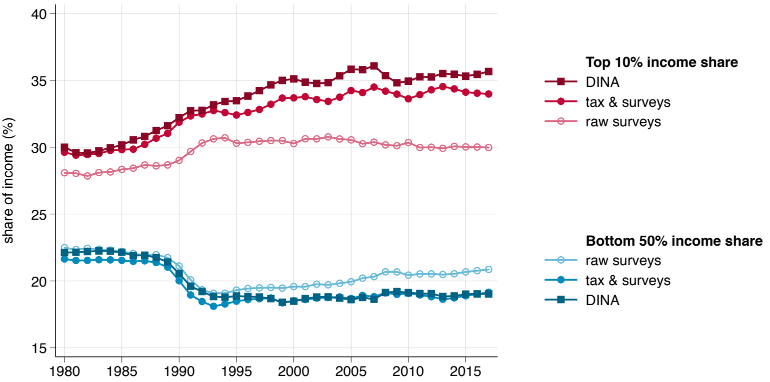
- Non-parametric functional form.
- Large number of highly correlated predictors.
- Monotonicity constraint on Q_k^A.

$\Rightarrow$ machine learning is very good at dealing with this type of problem: we use a standard, state-of-the-art algorithm known as gradient tree boosting

Harmonization of Concepts: Example



From surveys to DINA: pretax inequality in Europe, 1980-2017



WID.WORLD