Econ 133 – Global Inequality and Growth

Inequality between individuals

Gabriel Zucman

zucman@berkeley.edu

What we've learned so far:

Trends in the functional distribution of income

- The capital share is rising, the labour share falling
- What theories can account for this evolution

Now we move to the interpersonal distribution of income, starting with the tools

Roadmap

1. Data sources to study inequality between individuals

2. Metrics: Gini coefficient, Pareto-Lorenz coefficient, top shares

3. Unit of observation

1 Data sources for interpersonal inequality

1.1 Survey data

- Surveys are a popular data source to study inequality:
  - Ask a sample of families about their income, wealth...
  - Lots of socio-demographic characteristics
  - Revolutionized empirical research in second half of 20th century

- Numerous household surveys now available:
  - Luxembourg income study (40 countries, 1968–)
  - Luxembourg wealth studies (12 countries, 1994–)
  - World Bank Living Standard Measurement Studies (39 countries, 1985–).
- Survey data are useful, but insufficient:
  - Large gap between surveys and macro totals
  - Practical pbs: non-response & under-reporting at the top

1.2 Tax data

- Tax administrations have published tabulations of income by size of income since beginning of income tax (usually early 20th century)
- In recent decades, availability of micro-samples of tax returns
- Kuznets (1953) first to use tax data to compute top income shares
- Extended by Atkinson, Piketty, Saez (2011) and others

Limits of tax data:

- Miss tax evasion
- Miss legally tax-exempt income
- Ex: US tax data only capture 60% of US national income
- Incomplete information on distribution within bottom 90%

1.3 Distributional national accounts

DINAs = decompositions of national account aggregates such that:

- Distributions of income, wealth, saving, taxes, transfers... are consistent with what survey/tax data show

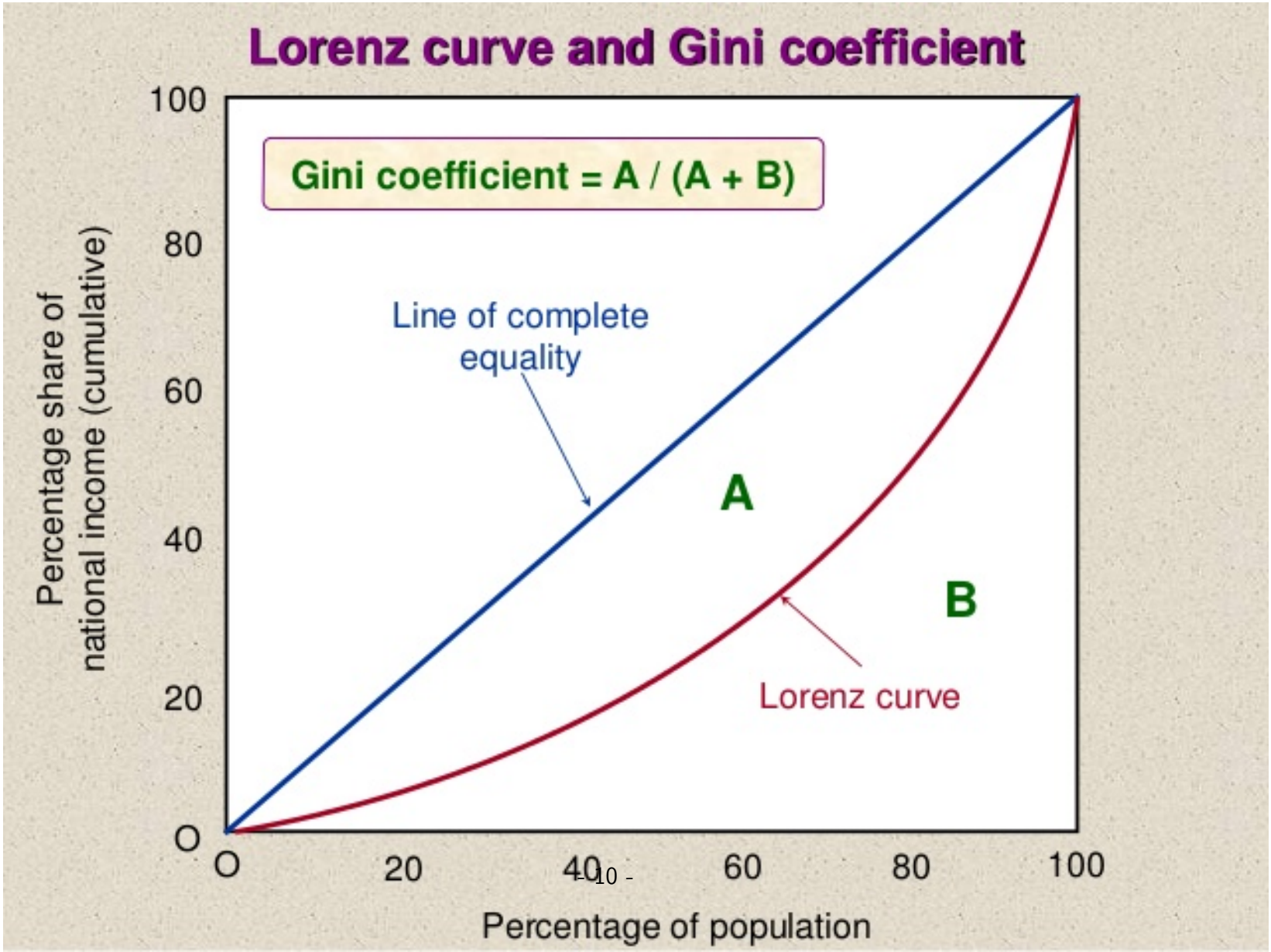
- Totals match macro aggregates

- Current attempt to compile DINAs throughout the world: `http://WID.world`

2 How to quantify inequality?

2.1 Gini coefficient

- Inequality often summarized by Gini coefficient G
- Lorenz curve shows % of income earned by people below fractile p
- G $=$ 2 x area between 45 degree line and Lorenz curve
- G =0 means Lorenz curve is the 45 degree line $=$ perfect equality


Lorenz curve and Gini coefficient
Gini coefficient = A / (A + B)
Line of complete equality
A
B
Lorenz curve
Percentage share of national income (cumulative)
Percentage of population
100
80
60
40
20
O
O
20
40
60
80
100

2.2 Income and wealth shares

- Problem of Gini: quite abstract & requires lots of data

- Shares are more concrete ("the top 1% income share")

What is the link between the Gini coefficient and top shares?

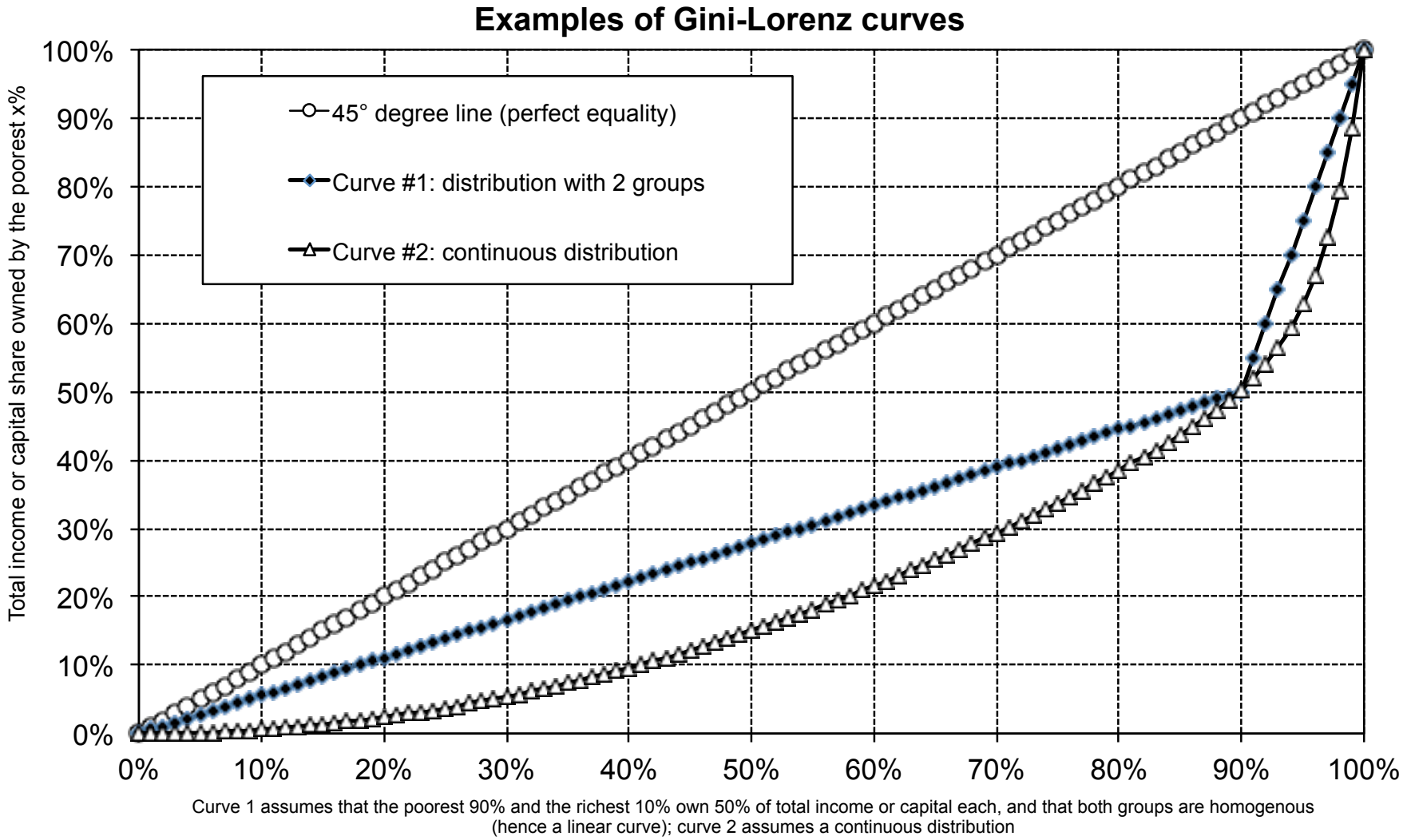
- Let's consider a finite number of income groups

- Individuals below percentile p_1 own a share s_0 of income, individuals between p_1 and p_2 own a share s_1, etc.

- Ex: Assume there are 2 groups, and that both groups are homogenous

- Ex: p_1 =0.9, s_0 =0.5, s_1 =0.5. I.e., the bottom 90% and the top 10% both own 50% of total income

- With two homogenous groups, geometrically easy to show that $G = s_1 + p_1 - 1$



Curve 1 assumes that the poorest 90% and the richest 10% own 50% of total income or capital each, and that both groups are homogenous (hence a linear curve); curve 2 assumes a continuous distribution

2.3 Pareto coefficients

- Another useful metric of inequality is the Pareto coefficient
- At the top, income & wealth well approx. by Pareto distributions
- Pareto distributions have a probability density function

$$f(y) = \frac{ac^a}{y^{1+a}}$$

- and a cumulative distribution function $1 - F(y) = (c/y)^a$
- with $c =$ constant and $a =$ Pareto coefficient

- Key property of Pareto distributions: ratio average/threshold = constant

- Note $y^*(y)$ average income of pop. above threshold y. Then:

$$y^*(y) = y\frac{a}{a-1} = yb$$

- b is called the inverted Pareto-Lorenz coefficient

- If a=2, b=2: average income above \$100,000 = \$200,000; average income above \$1 million = \$2 million, etc.

- US 1970s, income: $b = 1.7$–1.8 ($a = 2.2$–2.3)

- US 2010s, income: $b = 2.2$–2.5 ($a = 1.7$–1.8)
- For wealth distributions, b can be larger than 3
- $b =$ index of concentration
- Pareto coefficients are easy to estimate using tabulations
- See Kuznets 1953, and Atkinson, Piketty and Saez 2011 for graphs on b coeff over time & across countries

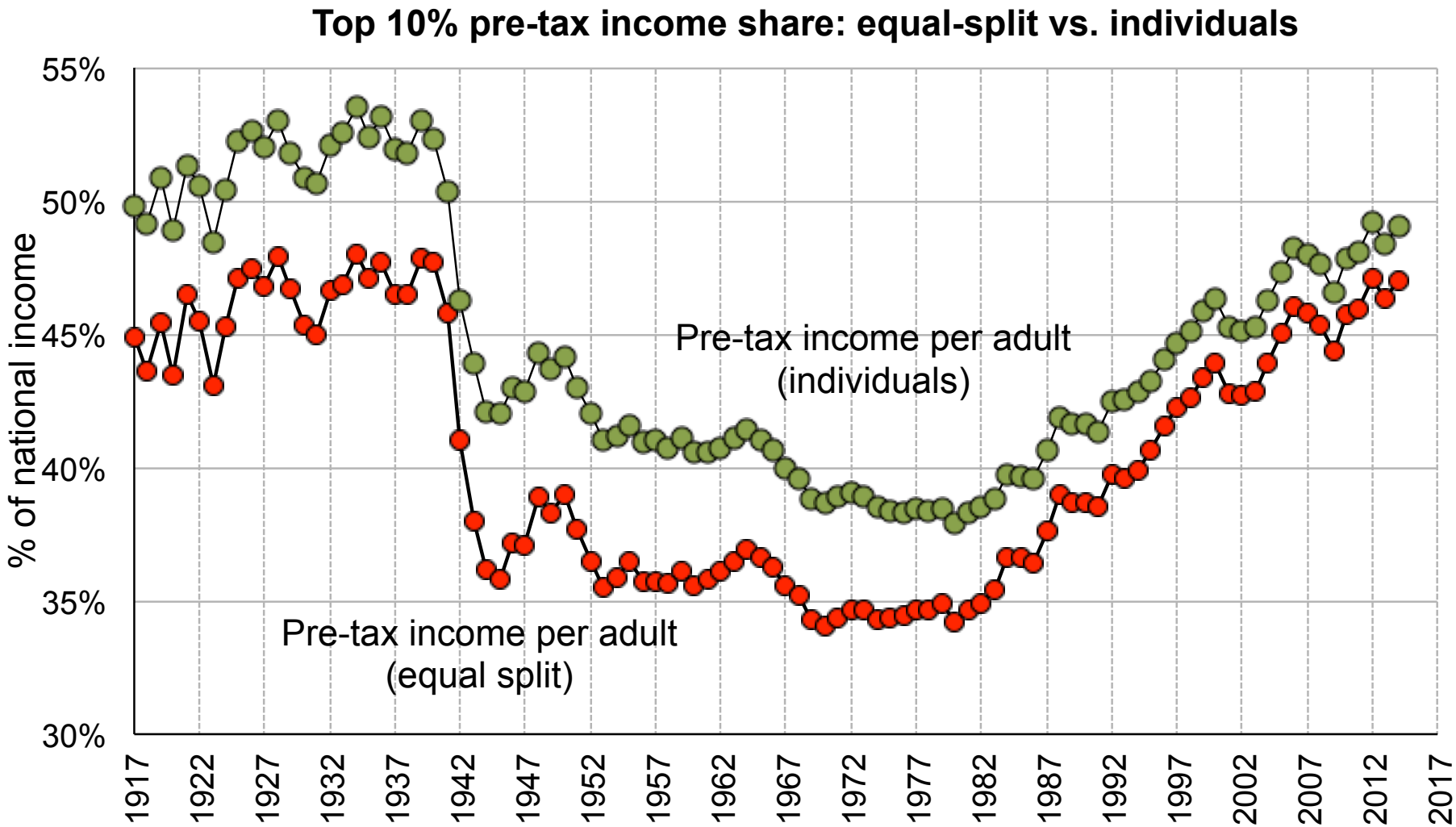
The Pareto coefficient a is $a = 1.7$ in country A and $a = 1.5$ in country B. Therefore, income is more concentrated at the top in country A than in country B.

A — Yes

B — No

3 Unit of observation

- Individual adult: assumes no sharing of resources between spouses
- Equal-split adults: assumes full sharing of resources
- Tax unit $\approx$ households: relevant for tax policy simulations



Source: Appendix Table II-B9.

References

Alvaredo, Facundo, "A Note on the Relationship between Top Income Shares and the Gini Coefficient", *Economics Letter*, 2011 (web)

Atkinson, Anthony, Thomas Piketty, and Emmanuel Saez "Top Incomes in the Long-Run of History", *Journal of Economic Literature*, 2011 (web)

King, Gregory, *Natural and Political Observations and Conclusions Upon the State and Condition of England*, 1696, 45p.

Kuznets, Simon *Shares of Upper Income Groups in Income & Saving*, 1953