# Labor Economics, Week 3
# Estimating top income shares, and Distributional decompositions

Maximilian Kasy

Department of Economics, Oxford University

# Takeaways for week 3 - Estimating economic inequality

1) Estimating top income shares

- Inequality has changed dramatically over time, especially at the top of the distribution.
- A key data source showing this are tax records.
- Need to estimate (i) how much income the rich received, and (ii) how large the total income generated in the economy was.
- The distribution of top incomes (and of top wealth holdings) is well approximated by the so-called Pareto distribution.
- This allows us to estimate top shares even if data are only available for tax brackets.

# Takeaways for week 3 - Estimating economic inequality

2) Distributional decompositions and de-unionization

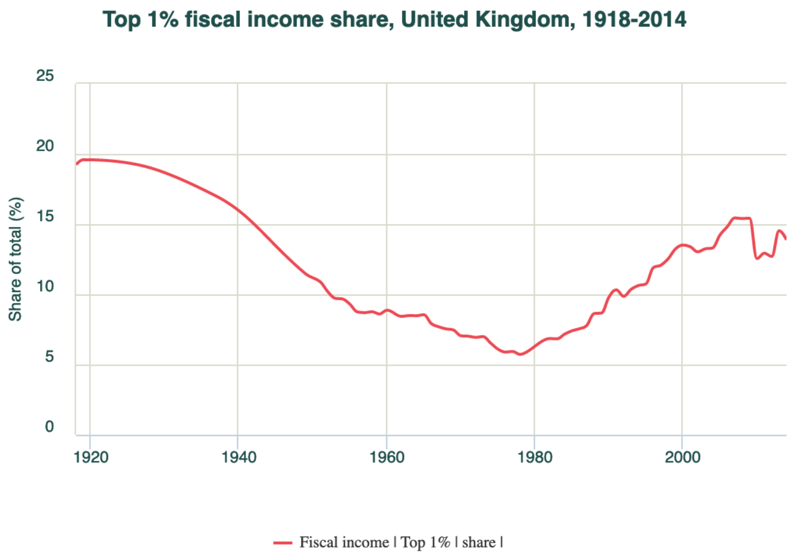
- What is the effect of declining unionization on the distribution of wages?
- Compare the wages of union- and non-union members?
- Problem: They might differ in terms of demographic & economic characteristics.
- Better: Compare people who look similar, and differ only in terms of union membership.
- Hypothetical question: What if
    1. the distribution of demographic covariates (age, gender,...) had stayed the same,
    2. the distribution of wages given demographics and union membership status had stayed the same, but
    3. we consider actual historical changes of union membership for different demographic groups.
- Like matching estimation, but for distributions!

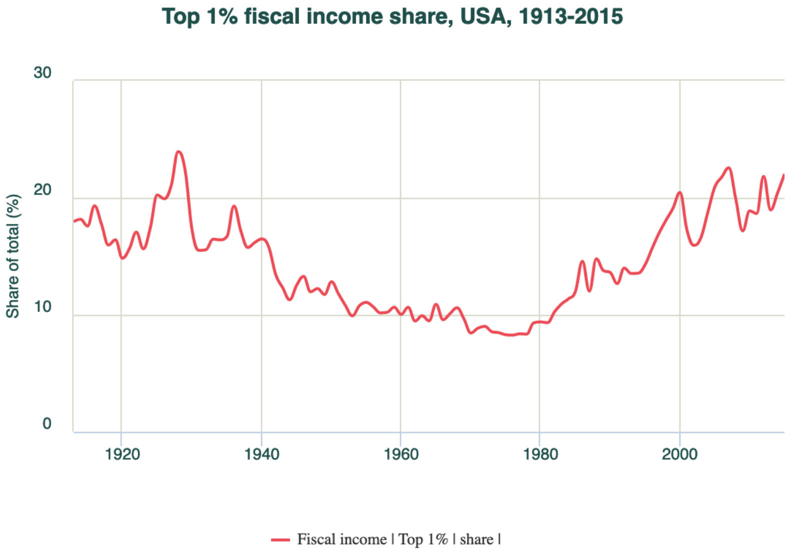
# Roadmap

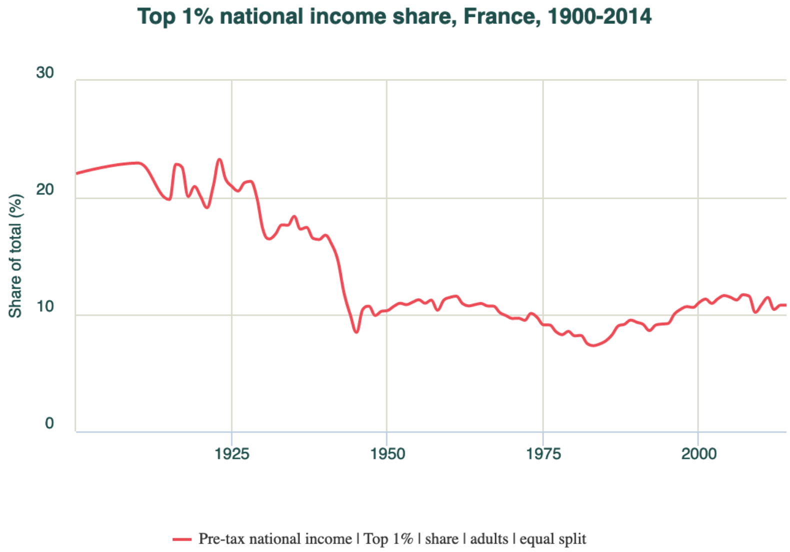
Top income shares

Censored data

Distributional decompositions


Top 1% fiscal income share, United Kingdom, 1918-2014
Share of total (%)
25
20
15
10
5
0
1920
1940
1960
1980
2000
Fiscal income | Top 1% | share |


Top 1% fiscal income share, USA, 1913-2015
Share of total (%)
30
20
10
0
1920
1940
1960
1980
2000
Fiscal income | Top 1% | share |


Top 1% national income share, France, 1900-2014
Share of total (%)
30
20
10
0
1925
1950
1975
2000
Pre-tax national income | Top 1% | share | adults | equal split

# How are these estimated?

- Using income tax data (for numerator) and national accounts (for denominator).
- Available for top incomes since the introduction of income taxes.
- For lower incomes: only since the expansion of income taxes.
- These slides: Econometric issues.
- Readings: Data issues, interpretation, etc.

## The Pareto distribution

- Top incomes are very well described by the Pareto distribution
- Defined by:

$$P(Y > y | Y \geq \underline{y}) = \left(\underline{y}/y\right)^{\alpha_0}$$

for $y \geq \underline{y}$, where $\alpha_0 > 1$.

- Corresponding density:

$$\begin{aligned} f(Y;\alpha_0) &= -\frac{\partial}{\partial y} P(Y > y | Y \geq \underline{y}) \\ &= -\frac{\partial}{\partial y} \left(\underline{y}/y\right)^{\alpha_0} \end{aligned}$$

**Questions for you**

Calculate $f(Y;\alpha_0)$

Answer:

$$f(Y;\alpha_0) = \alpha_0 \cdot \underline{y}^{\alpha_0} \cdot y^{-\alpha_0 - 1}.$$

# Key property

- Pareto distribution satisfies:

$$E[Y|Y \geq y] = \frac{\alpha_0}{\alpha_0 - 1} \cdot y.$$

- This is true for all $y$!!

**Questions for you**

Describe this equation in words.

- We can therefore calculate average incomes of the 1% as:

$$\overline{y}^{1\%} = \frac{\alpha_0}{\alpha_0 - 1} \cdot q^{99},$$

where

$$P(Y \leq q^{99}) = .99$$

- To get top income shares, we need estimates of
    1. $\alpha_0$
    2. $q^{99}$
    3. National income for the denominator
- We will discuss $\alpha_0$.
- Smaller $\alpha_0 \Rightarrow$ fatter tails $\Rightarrow$ more inequality, larger top income shares.

## Key problem

- Standard technique to construct estimators: maximum likelihood.
- Find the number $\alpha_0$ which makes the observed incomes $y_1, \ldots, y_n$ "most likely"

$$\widehat{\alpha}^{MLE} = \underset{\alpha}{\operatorname{argmax}} \prod_{i=1}^{n} f(y_i; \alpha)$$
$$= \underset{\alpha}{\operatorname{argmax}} \sum_{i=1}^{n} \log(f(y_i; \alpha)).$$

- First order condition

$$\frac{\partial}{\partial \alpha} \sum_{i=1}^{n} \log(f(y_i; \alpha)) = 0.$$

**Questions for you**

Solve this first order condition for the Pareto density.

## Answer

- Log density of $y_i$

$$\log(f(y_i;\alpha)) = \log(\alpha\left(\underline{y}/y_i\right)^{\alpha}\cdot y_i^{-1}) = \log(\alpha) + \alpha\log\left(\underline{y}/y_i\right) - \log(y_i).$$

- First order condition

$$\begin{aligned} 0 &= \frac{\partial}{\partial\alpha}\sum_{i=1}^{n}\log(\alpha\left(\underline{y}/y_i\right)^{\alpha}\cdot y^{-1}) \\ &= \sum_{i=1}^{n}\left(\frac{1}{\alpha}+\log\left(\underline{y}/y_i\right)\right). \end{aligned}$$

- Solving for $\alpha$

$$\widehat{\alpha}^{MLE} = \frac{n}{\sum_i \log\left(y_i/\underline{y}\right)}. \tag{1}$$

# Additional problem

- Available data do not list actual incomes,
- just the number of people in different tax brackets $[y_l, y_u]$.
- Technical term: The data are "censored."
- For the Pareto distribution:

$$\begin{aligned} P(Y \in [y_l, y_u] | Y \geq \underline{y}) &= P(Y > y_l | Y \geq \underline{y}) - P(Y > y_u | Y \geq \underline{y}) \\ &= \left(\underline{y}/y_l\right)^{\alpha_0} - \left(\underline{y}/y_u\right)^{\alpha_0}. \end{aligned} \tag{2}$$

# Likelihood for two tax brackets

- Data on $N$ people with incomes above $\underline{y}$
- $N_2$ people in the bracket $[y_l, \infty)$
- Probability of any given individual in the top bracket:

$$p(\alpha_0) = P(Y > y_l | Y > \underline{y}) = \left(\underline{y}/y_l\right)^{\alpha_0}.$$

- Probability of exactly $N_2$ individuals in the top bracket:

$$P(N_2 = n_2 | N = n; \alpha) = \binom{n}{n_2} \cdot p(\alpha_0)^{n_2} (1 - p(\alpha_0))^{n - n_2}.$$

- Remember the binomial distribution?

## Questions for you

Calculate the maximum likelihood estimator for censored data

$$\widehat{\alpha}^{MLE} = \underset{\alpha}{\operatorname{argmax}}\ P(N_2 = n_2 | N = n; \alpha).$$

(Homework)

# References

- *Atkinson, A. B., Piketty, T., and Saez, E. (2011). Top incomes in the long run of history.* Journal of Economic Literature, *49(1):3–71.*
- *Piketty, T. (2014).* Capital in the 21st Century. *Harvard University Press, Cambridge.*
- *(2019). World inequality database.*
  https://wid.world/

# Roadmap

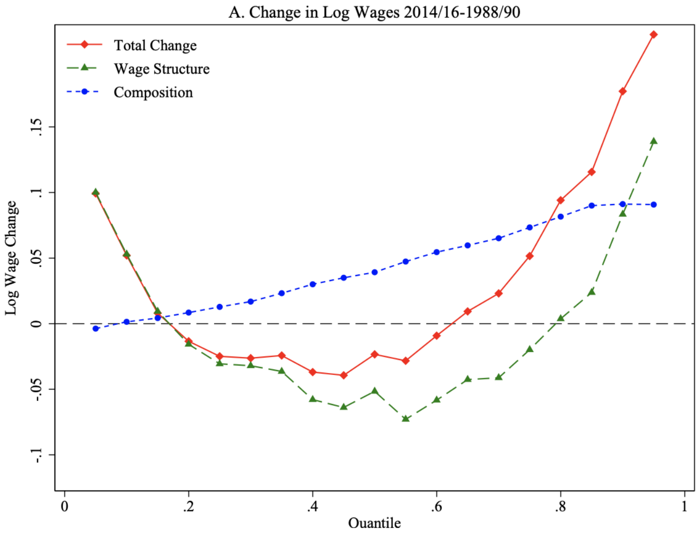
Top income shares

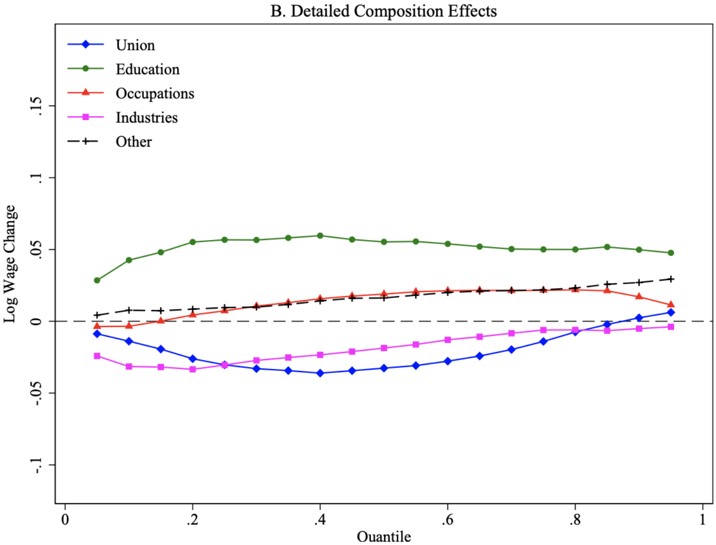
Censored data

Distributional decompositions

# Distributional decompositions

- Next 2 slides:
  Changes of log hourly wages of men in the US across different quantiles of the wage distribution.
- Covariates:
  - six education groups,
  - potential experience (nine groups),
  - union coverage,
  - occupation (17 categories),
  - industry (14 categories),
  - marital status and race.
- Total change: Polarization.
  Explained by composition: Increase in inequality.


A. Change in Log Wages 2014/16-1988/90
Total Change
Wage Structure
Composition
Log Wage Change
.15
.1
.05
0
-.05
-.1
0
.2
.4
.6
.8
1
Quantile


B. Detailed Composition Effects
Union
Education
Occupations
Industries
Other
Log Wage Change
.15
.1
.05
0
-.05
-.1
0
.2
.4
.6
.8
1
Quantile

# Decreasing unionization since the 1980s

- Union wages: higher and less unequal
- Thus: declining unionization
  $\Rightarrow$ increase in inequality?
- Just compare wages of union / non-union members?
- Problem: two groups might be different, in terms of
  - age,
  - education,
  - gender,
  - ethnicity,
  - sector of the economy,
  - state of residence,
  - ...
- Want to compare people who look similar along all these dimensions!

# Distributional decompositions

Hypothetical questions of the form:

- What if
  1. distribution of demographic covariates
     had stayed the same,
  2. distribution of wages
     *given* demographics and union membership status
     had stayed the same, but
  3. we consider actual historical changes of union membership
     given demographics.
- How would the distribution of wages have changed?
- i.e., to what extent is de-unionization responsible for the rise in inequality?

# Setup

- Observe repeated cross-sections
  of draws from the time $t$ distributions $P^t$.
- Variables $(Y, D, X)$
  - $Y$: outcome, e.g., real earnings
  - $X$: demographic covariates, e.g., age, gender, ...
  - $D$: binary “treatment,” e.g., union membership
- Effect of historical changes in $D$
  on the distribution $P(Y)$?
- In particular, on statistics $\nu(P(Y))$?
- Examples for $\nu$: mean, variance, share below the poverty line, quantiles, Gini coefficient, top income shares, ...

# Probability reminder

Let $p(y,x)$ denote a joint probability density.

1. Conditional distribution:
$$p(Y|X) = \frac{p(Y,X)}{p(X)}$$
2. Marginal distribution:
$$p(Y) = \int p(Y,X)dX$$
3. Thus:
$$p(Y) = \int p(Y|X)p(X)dX$$
4. Similarly (law of iterated expectations):
$$E[Y] = E[E[Y|X]]$$

# Counterfactual distribution

- Two distributions $P^0(Y, D, X)$, $P^1(Y, D, X)$ (beginning and end of historical period)
- What would the wage distribution $P^*(Y)$ be, assuming
  1. dist of demographics stayed the same,
  2. dist of wages given demographics, union membership stayed the same
  3. actual historical change of union membership

$$\begin{aligned} P^*(X) &= & P^0(X) \\ P^*(Y \leq y|X, D) &= & P^0(Y \leq y|X, D) \\ P^*(D|X) &= & P^1(D|X). \end{aligned}$$

- Get the counterfactual distribution $P^*(Y)$:

$$P^*(Y \leq y) := \int_{X,D} P^0(Y \leq y|X, D) dP^1(D|X) dP^0(X).$$

# Rewriting the counterfactual distribution

1. Multiply and divide the integrand by $P^0(D|X)$.
2. Rewrite the probability $P^0(Y \leq y|X, D)$ as an expectation $E^0[\mathbf{1}(Y \leq y)|X, D]$.
3. Give the fraction $P^1(D|X)/P^0(D|X)$ a new name: $\theta(D, X)$.
4. Pull $\theta$ into the conditional expectation.
5. Use the “law of iterated expectations” to get an unconditional expectation.

**Questions for you**

Execute these steps, and see what you get!

# Solution

$$
\begin{aligned}
P^*(Y \leq y) &= \int_{X,D} P^0(Y \leq y|X,D)\frac{P^1(D|X)}{P^0(D|X)}P^0(D|X)P^0(X)dDdX \\
&= \int_{X,D} E^0[\mathbf{1}(Y \leq y)|X,D]\theta(D,X)P^0(D|X)P^0(X)dDdX \\
&= E^0[E^0[\mathbf{1}(Y \leq y) \cdot \theta(D,X)|X,D]] \\
&= E^0\left[\mathbf{1}(Y \leq y) \cdot \theta(D,X)\right],
\end{aligned}
$$

where

$$
\theta(D,X) := \frac{P^1(D|X)}{P^0(D|X)}.
$$

**Questions for you**

Interpret this representation of the counterfactual distribution.

## Estimation

- Suppose $X$ is discrete.
- Let $N^t(d,x)$ be the number of observations in period $t$ with $D = d$, $X = x$,
- similar for $N^t(x)$.
- Then we can estimate $\theta(d,x)$ as

$$\widehat{\theta}(d,x) = \frac{N^1(d,x)}{N^1(x)} \Big/ \frac{N^0(d,x)}{N^0(x)}.$$

- Estimate $P^*(Y \leq y)$ as

$$\sum_i \mathbf{1}(Y_i \leq y) \cdot \widehat{\theta}(D_i, X_i) \Big/ \sum_i \widehat{\theta}(D_i, X_i),$$

where the sums are over all observations in period 0.

## Questions for you

Implement this in R!

# References

- *Fortin, N. M. and Lemieux, T. (1997). Institutional changes and rising wage inequality: Is there a linkage?* The Journal of Economic Perspectives*, 11(2):pp. 75–96.*
- *Firpo, S., Fortin, N., and Lemieux, T. (2011). Decomposition methods in economics.* Handbook of Labor Economics*, 4:1–102.*