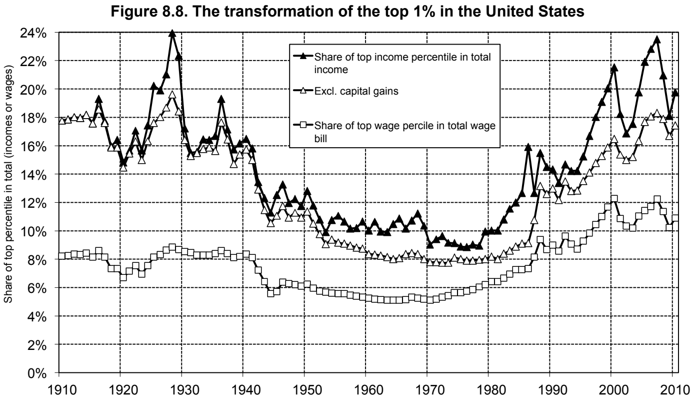
Empirical Research on Economic Inequality
Estimating top income shares

Maximilian Kasy

Harvard University, fall 2015

Top 1% income share in the US



The rise in the top 1% highest incomes since the 1970s is largely due to the rise in the top 1% highest wages. Sources and series: see piketty.pse.ens.fr/capital21c.

Reproduced from Piketty (2014)

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
- Student presentation: 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. α_0
  2. q^{99}
  3. National income for the denominator
- We will discuss α_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 α_0 which makes the observed incomes $y_1, \ldots, y_n$ "most likely"

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

- 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 α

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

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.*

Atkinson, A. B. and Morelli, S. (2015). Chartbook of economic inequality.

`http://www.chartbookofeconomicinequality.com/`