Generalized Pareto Curves

Theory and Applications

Juliette Fournier[1] (joint with Thomas Blanchet[2] and Thomas Piketty[2])

April 5, 2018

[1]MIT

[2]Paris School of Economics

Outline

Introduction

Goal

Goal

→ Estimating income/wealth distributions using tax data

→ Producing statistics about income/wealth inequality

- Individual micro-data on taxpayers available in a few countries for recent years
  - 1962-2018 in the U.S.
  - 1994-2018 in France
- **Problem:** tax data often censored

  ⇒ for most countries and years, only **tabulated tax data**

An example of tabulated tax data

Typical tax data

1. Brackets of the income/wealth tax schedule
2. Number of taxpayers per bracket
3. Average income/wealth in the bracket

Income bracket	Bracket size	Bracket average income
From 0 to 1000	300 000	500
From 1000 to 10 000	600 000	5 000
From 10 000 to 50 000	80 000	30 000
More than 50 000	20 000	200 000

The Pareto distribution

Empirical observation: Top income and wealth distributions well approximated by the Pareto (1896) distribution

- Minimum income $x_0 > 0$
- Linear relationship between log(rank) and log(income):

$$\forall x \geq x_0, \quad \log \mathbb{P}\{X > x\} = -\alpha \log(x/x_0)$$

- Hence the "power law":

$$\forall x \geq x_0, \quad \mathbb{P}\{X > x\} = (x/x_0)^{-\alpha}$$

- $\alpha =$ **Pareto coefficient**

Existing interpolation methods

- **Pareto interpolation methods** exploit this observation
- Typical assumption = exact Pareto distribution within each tax bracket
  → Kuznets, 1953; Feenberg and Poterba, 1992; Piketty, 2001; Piketty and Saez, 2003; etc.

Limitations

1. Only valid for the top of the distribution (highest revenues/estates)
2. Imprecise even at the top
3. Don't exploit all available information
4. Not internally consistent

Beyond Pareto: Generalized Pareto interpolation

1. Introduce **generalized Pareto curves**
    - characterize power law behavior in a flexible way
    - allow to visualize and estimate distributions of income and wealth
2. Develop **generalized Pareto interpolation**
    - empirical method exploiting tabulated tax data that preserves deviations from strict Paretian behavior
    - more precise
    - can estimate the entire distribution

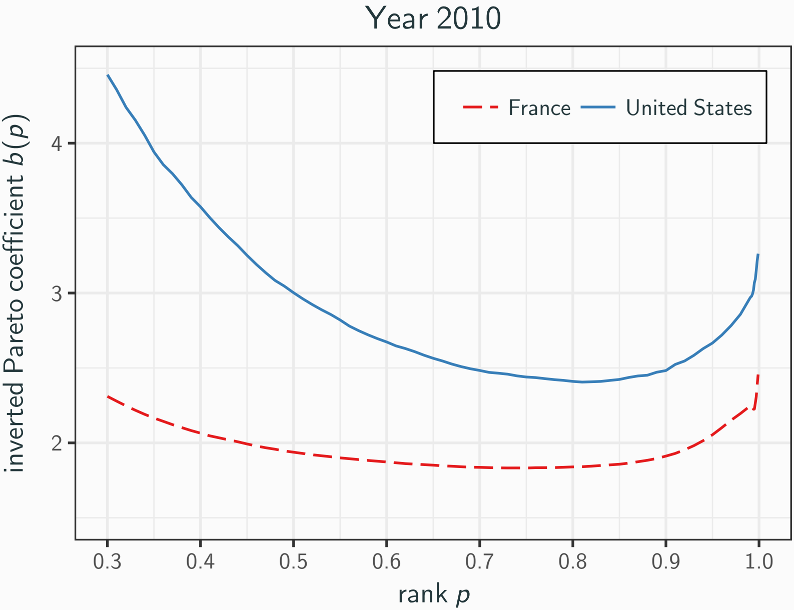
Generalized Pareto Curves: Definition and Theory

Definition of the generalized Pareto curve

- Consider a Pareto distribution with coefficient α
- Quantile function: $Q : p \mapsto x_0(1-p)^{-1/\alpha}$
- Then:
$$\frac{\mathbb{E}\left[X|X > Q(p)\right]}{Q(p)} = \frac{\alpha}{\alpha - 1} \equiv b$$
is constant

  $\rightarrow$ b is the **inverted Pareto coefficient**
- More generally, define for an arbitrary random variable X:
$$b(p) = \frac{\mathbb{E}\left[X|X > Q(p)\right]}{Q(p)}$$
  $\rightarrow$ the function $p \mapsto b(p)$ is the **generalized Pareto curve**

Generalized Pareto curve of pre-tax national income



Interpretation

- More flexible definition of power laws ▸ definition
  - → based on Karamata's (1930) theory of regular variations
- Two types of distributions, excluding pathological cases:

Power laws	Thin tails
Pareto	Normal
Student's t	Log-normal
...	...
$\lim_{p \to 1} b(p) > 1$	$\lim_{p \to 1} b(p) = 1$

Relationship with the quantile function

Proposition

If X positive random variable with quantile function Q, and Pareto curve b, then:

$$Q(p) = Q(\bar{p})\frac{(1-\bar{p})b(\bar{p})}{(1-p)b(p)}\exp\left(-\int_{\bar{p}}^{p}\frac{1}{(1-u)b(u)}\mathrm{d}u\right).$$

$\Rightarrow$ Quantile function entirely pinned down by:

1. the Pareto curve
2. a scale factor:
    - value of Q at one point $\bar{p}$
    - or average income/wealth

Relationship with the quantile function

Proposition

If X positive random variable with quantile function Q, and Pareto curve b, then:

$$Q(p) = Q(\bar{p})\frac{(1-\bar{p})b(\bar{p})}{(1-p)b(p)} \exp\left(-\int_{\bar{p}}^{p} \frac{1}{(1-u)b(u)} \mathrm{d}u\right).$$

$\Rightarrow$ Quantile function entirely pinned down by:

1. the Pareto curve
2. a scale factor:
   - value of Q at one point $\bar{p}$
   - or average income/wealth

Relationship with the quantile function

Proposition

If X positive random variable with quantile function Q, and Pareto curve b, then:

$$Q(p) = Q(\bar{p})\frac{(1-\bar{p})b(\bar{p})}{(1-p)b(p)} \exp\left(-\int_{\bar{p}}^{p} \frac{1}{(1-u)b(u)} \mathrm{d}u\right).$$

$\Rightarrow$ Quantile function entirely pinned down by:

1. the Pareto curve
2. a scale factor:
    - value of Q at one point $\bar{p}$
    - or average income/wealth

Generalized Pareto Interpolation

Back to the interpolation problem

Income bracket	Bracket size	Bracket average income
From 0 to 1000	300 000	500
From 1000 to 10 000	600 000	5 000
From 10 000 to 50 000	80 000	30 000
More than 50 000	20 000	200 000

- **Known:** $Q(p_k)$ and $b(p_k)$ for a few percentiles p_k
- *How to reconstruct the entire distribution based on that information?*

The naive approach

Naive approach:

1. Interpolate the Pareto curve using the $b(p_k)$'s
2. Determine the scale of the quantile function using one $Q(p_{k_0})$

Two problems

1. Won't be consistent with the $Q(p_k)$'s, $k \neq k_0$
2. No guarantee that Q is increasing

The interpolation problem

Problem: the interpolation must be constrained so that $Q(p)$ and $b(p)$ are consistent *with the data* and *with each other*

- With $x = -\log(1-p)$, define the function:

$$\varphi(x) = -\log \int_{1-e^{-x}}^{1} Q(u)\,\mathrm{d}u$$

- φ directly related to $Q(p)$ and $b(p)$:

$$\begin{cases} \varphi(x) = -\log(1-p)Q(p)b(p) \\ \varphi'(x) = 1/b(p) \end{cases}$$

Reformulation

Initial problem

$\iff$ Interpolation of φ knowing at the interpolation points:

- its **value**
- its **first derivative**

$+$ condition ensuring that Q is increasing

Spline interpolation

- **Spline:** function defined by piecewise polynomials
  - $\rightarrow$ commonly used in interpolation problems
- **Cubic splines:** piecewise polynomials of degree 3
- Hermite's basis:

$$h_{ij}^{(k)}(x) = \begin{cases} 1 & \text{if } (i,j) = (k,l) \\ 0 & \text{otherwise} \end{cases}$$

- Interpolate $x \in [x_k, x_{k+1}]$ by:

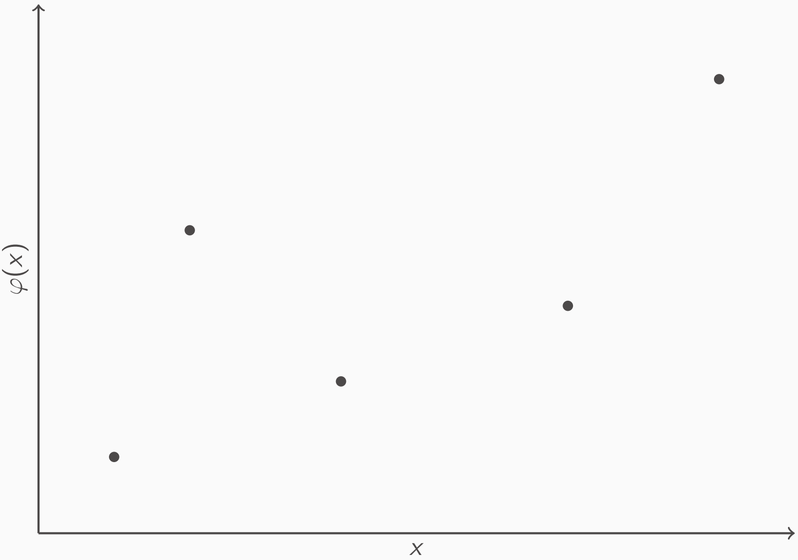
$$\hat{\varphi}(x) = h_{00}(t)y_k + h_{10}(t)(x_{k+1} - x_k)s_k + h_{01}(t)y_{k+1} + h_{11}(t)(x_{k+1} - x_k)s_{k+1}$$

with $t = (x - x_k)/(x_{k+1} - x_k)$ to have:

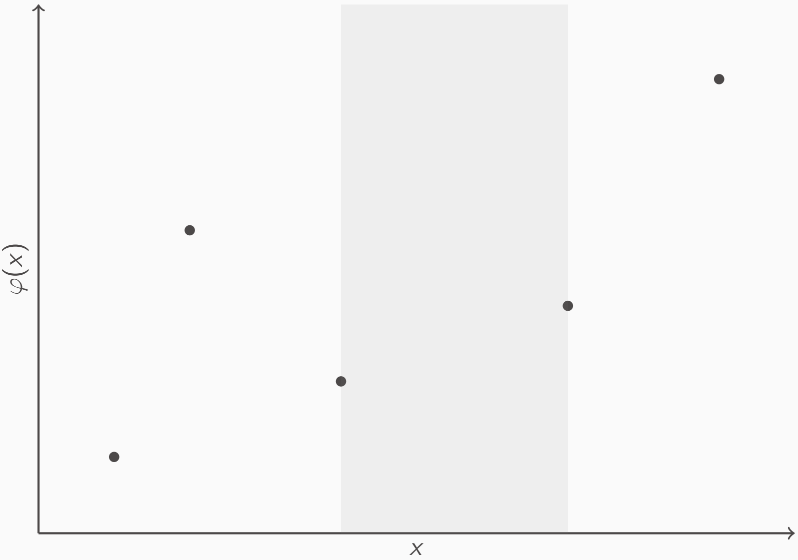
$$\forall k, \quad \hat{\varphi}(x_k) = y_k, \quad \hat{\varphi}'(x_k) = s_k$$

  - $\rightarrow$ with cubic splines, the y_k's and the s_k's determine the interpolant

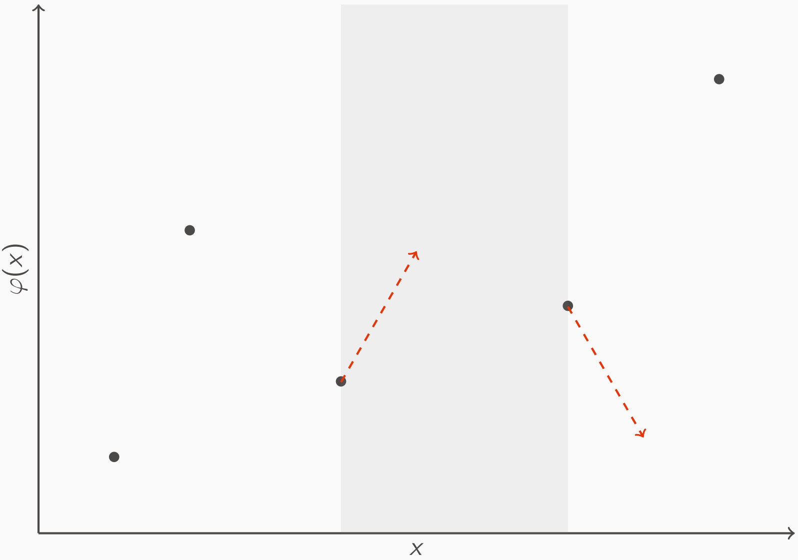
Cubic spline interpolation



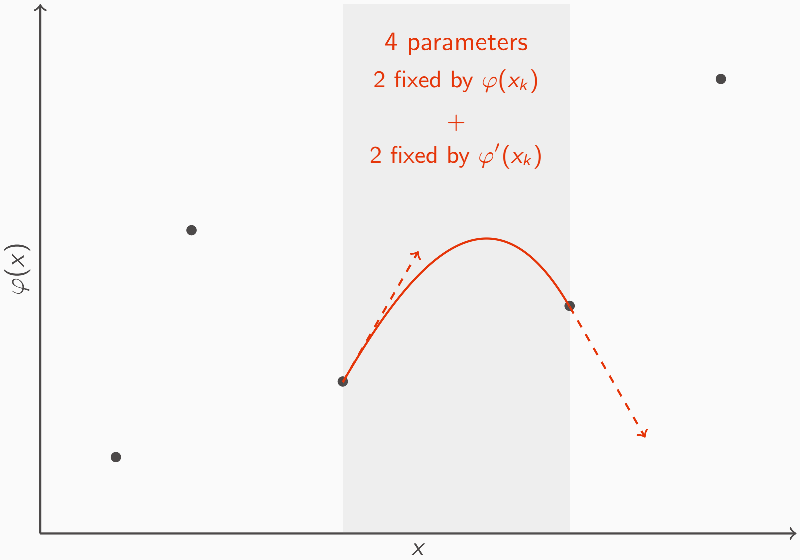
Cubic spline interpolation



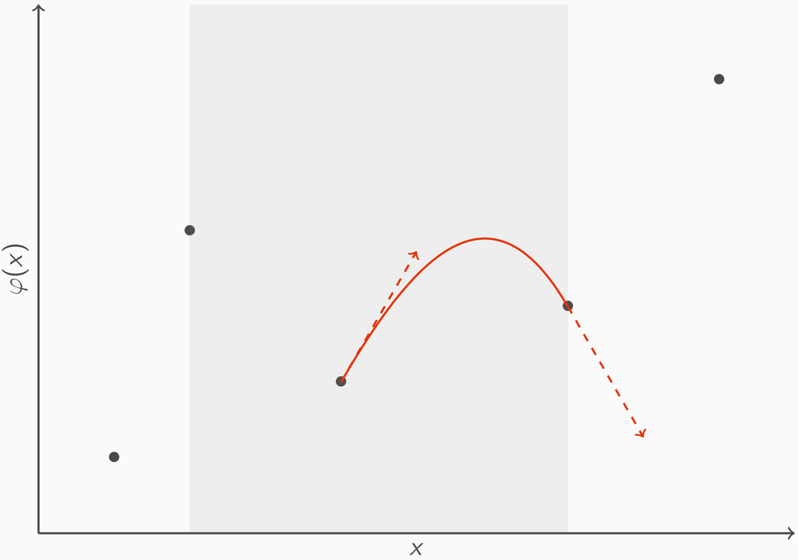
Cubic spline interpolation



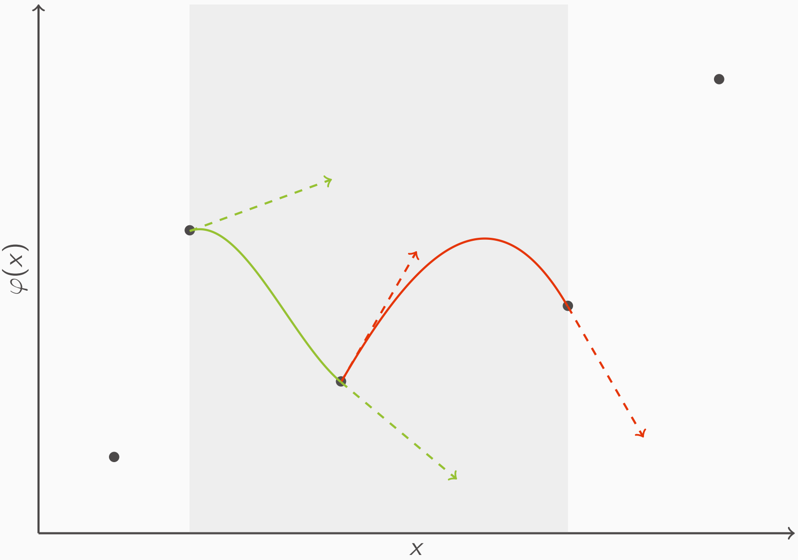
Cubic spline interpolation



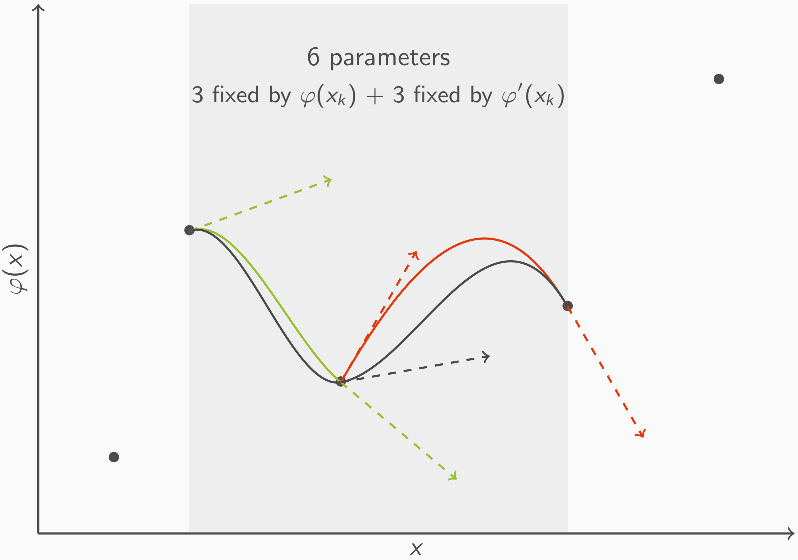
Cubic spline interpolation



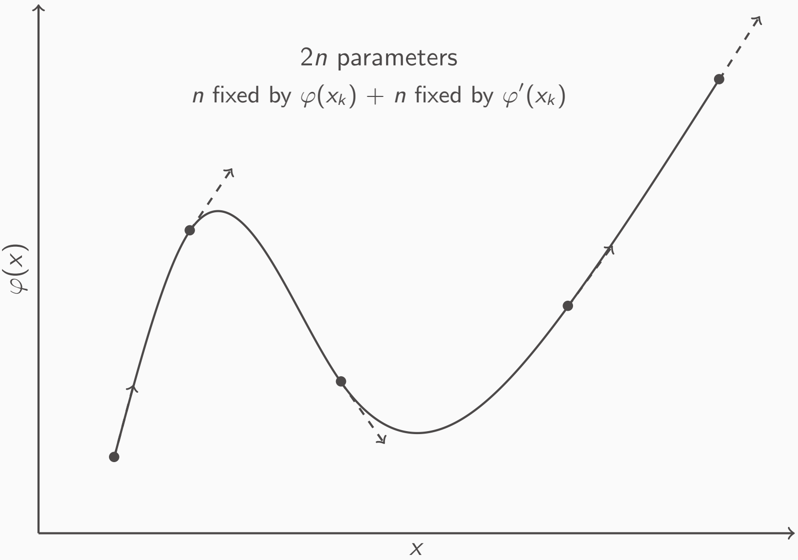
Cubic spline interpolation



Cubic spline interpolation



Cubic spline interpolation



Quintic spline interpolation (I)

- Need $\hat{\varphi}$ to be *twice* continuously differentiable
  - $\rightarrow$ necessary for a continuously differentiable Pareto curve
- Could have $\hat{\varphi}$ twice differentiable with cubic splines if the s_k's were free parameters
- But we *already know* the function and its first derivative at interpolation points
  - $\Rightarrow$ Cubic splines? Not smooth enough
- **Quintic splines:**
  - n extra degrees of freedom
  - one extra level of differentiability
  - same principle as cubic spline, but applied to the derivative
  - $\Rightarrow$ $3n$ parameters, $2n$ fixed and n free

Quintic spline interpolation (II)

- Determine the n free parameters by looking for the "most regular" curve
- Two **equivalent** approaches:
    1. Enforce continuity of **second** derivative at the jointures (+ two boundary conditions)
    2. Minimize the **curvature** of the overall curve:

$$\min \int_{x_1}^{x_K} (\hat{\varphi}''(x))^2 \, \mathrm{d}x$$

$\rightarrow$ requires solving a linear system of equations

Making sure that the quantiles are increasing

- No guarantee yet that the quantile function is increasing
- Quantile function increasing if and only if:

$$\hat{P}(x) \equiv \hat{\varphi}''(x) + \hat{\varphi}'(x)(1 - \hat{\varphi}'(x)) > 0$$

- Derive fairly general sufficient conditions on the parameters of the splines:

Cargo and Shisha (1966)

If $P = c_0 + c_1 X + \cdots + c_n X^n$, then:

$$\forall x \in [0,1], \quad \min_{0 \leq i \leq n} b_i \leq P(x) \leq \max_{0 \leq i \leq n} b_i$$

with:

$$b_i = \sum_{r=0}^{n} c_r \binom{i}{r} \Big/ \binom{n}{r}$$

Algorithm to obtain an increasing quantile function (I)

Define constrained estimate $\tilde{\varphi}$ as follows:

1. Start with the unconstrained estimate $\hat{\varphi}$
2. Set $\tilde{\varphi}''(x_k) = \tilde{\varphi}'(x_k)(1 - \tilde{\varphi}'(x_k))$ if $\hat{P}(x_k) < 0$ to have $\tilde{P}(x_k) \geq 0$ for all k
3. Check whether $\tilde{P}(x) \geq 0$ over each $[x_k, x_{k+1}]$

Algorithm to obtain an increasing quantile function (II)

4. If not, choose $x_k < x_1^* < \ldots < x_L^* < x_{k+1}$ and define $\tilde{\varphi}$ as:

$$\tilde{\varphi}_k(x) = \begin{cases} \varphi_0^*(x) & \text{if } x_k \leq x < x_1^* \\ \varphi_l^*(x) & \text{if } x_l^* \leq x < x_{l+1}^* \\ \varphi_L^*(x) & \text{if } x_l^* \leq x < x_{k+1} \end{cases}$$

with φ_l^*'s quintic splines such that:

$$\varphi_l^*(x_l^*) = y_l^*, \quad (\varphi_l^*)'(x_l^*) = s_l^*, \quad (\varphi_l^*)''(x_l^*) = a_l^*$$

$$\varphi_l^*(x_{l+1}^*) = y_{l+1}^*, \quad (\varphi_l^*)'(x_{l+1}^*) = s_{l+1}^*, \quad (\varphi_l^*)''(x_{l+1}^*) = a_{l+1}^*$$

$+$ boundary constraints at x_k and x_{k+1}

$\rightarrow$ y_l^*, s_l^*, a_l^* $(1 \leq l \leq L)$ parameters to be adjusted

Algorithm to obtain an increasing quantile function (III)

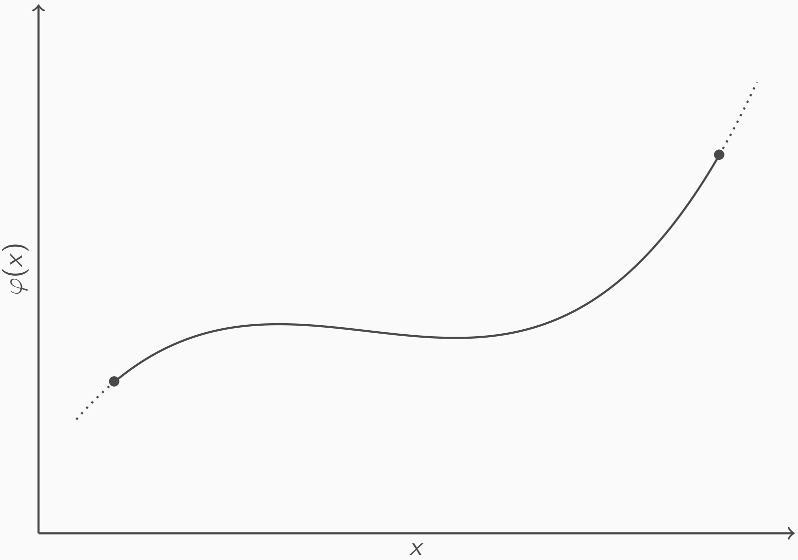
5. Set y_l^*, s_l^*, a_l^* to minimize the $\mathcal{L}^2$ distance to unconstrained estimate $\hat{\varphi}$ subject to positivity constraints:

$$\min_{\substack{y_l^*, s_l^*, a_l^* \\ 1 \leq l \leq L}} \int_{x_k}^{x_{k+1}} (\hat{\varphi}_k(x) - \tilde{\varphi}_k(x))^2 \mathrm{d}x$$

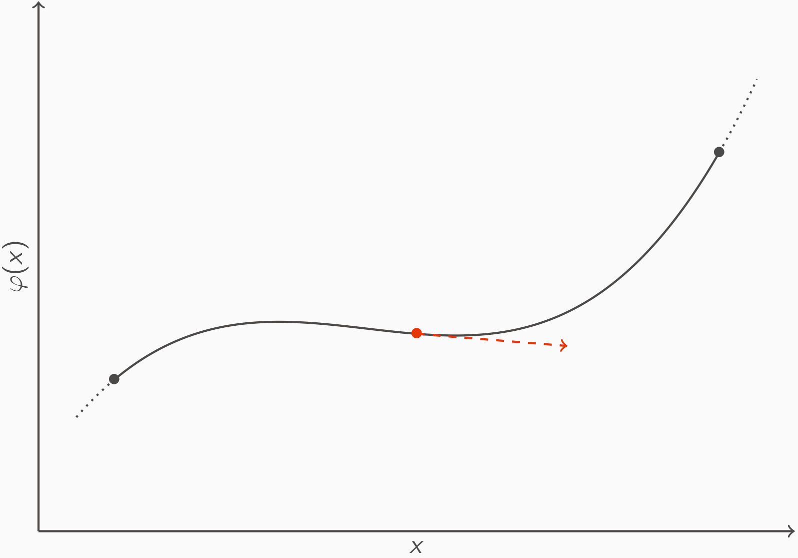
   subject to:

$$b_i^l \geq 0 \quad (0 \leq i \leq 8,\, 0 \leq l \leq L)$$

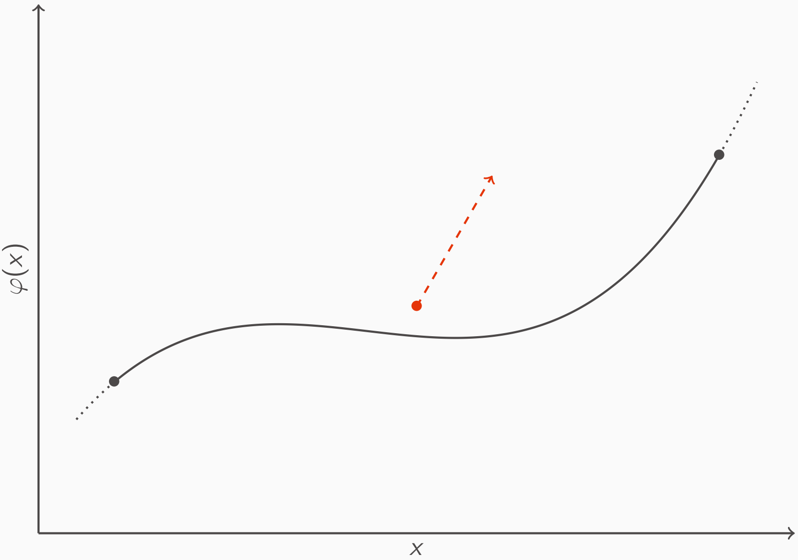
Making sure the quantiles are increasing



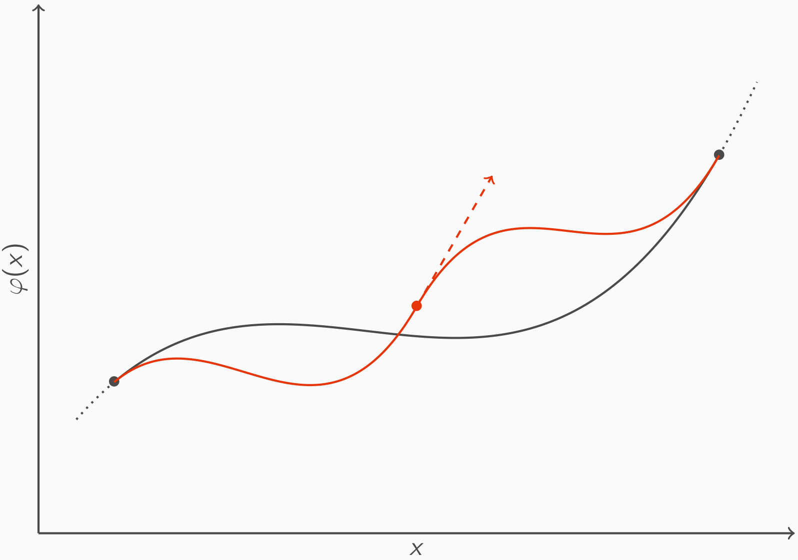
Making sure the quantiles are increasing



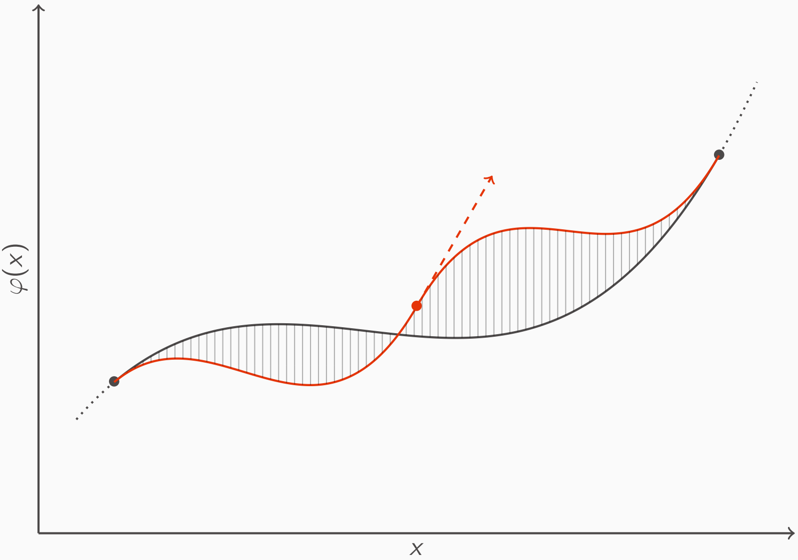
Making sure the quantiles are increasing



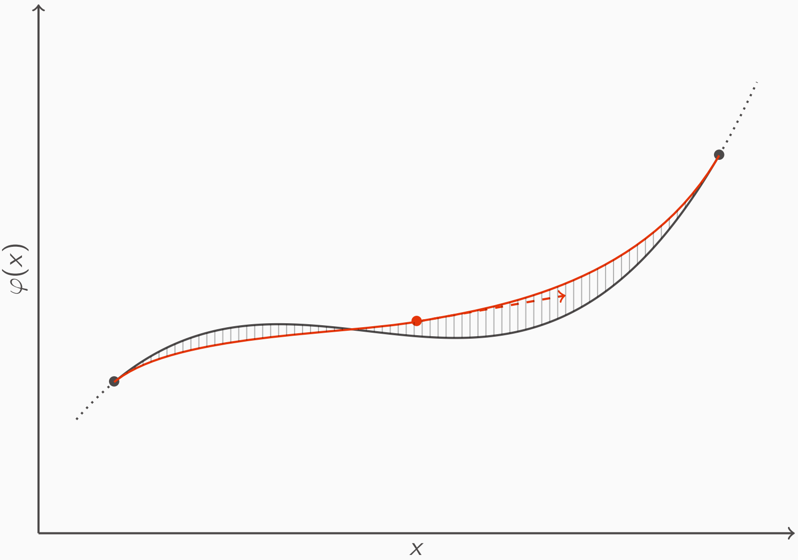
Making sure the quantiles are increasing



Making sure the quantiles are increasing

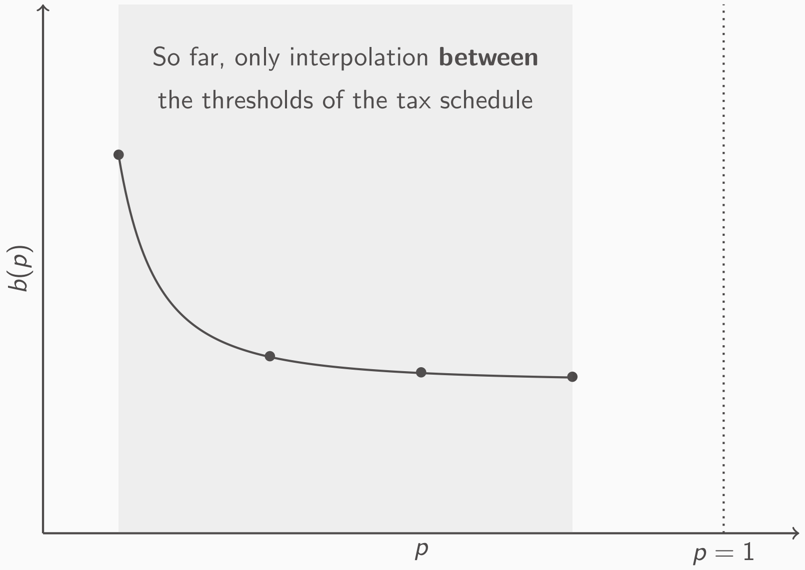


Making sure the quantiles are increasing

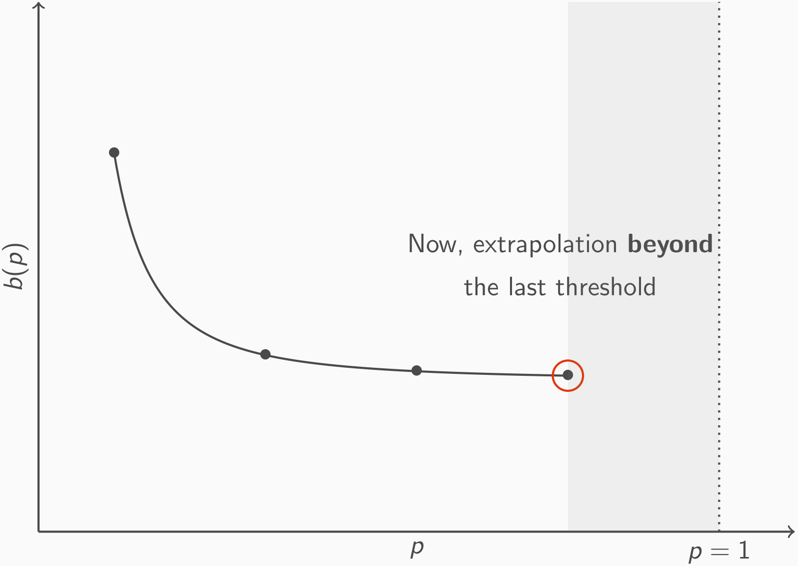


Extrapolation Beyond the Last Threshold

Extrapolation beyond the last threshold



Extrapolation beyond the last threshold



Extrapolation beyond the last threshold (I)

- **Generalized Pareto distribution**

$$\mathbb{P}(X \leq x) = GPD_{\mu,\sigma,\xi}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x-\mu}{\sigma}\right)^{-1/\xi} & \text{for } \xi \neq 0 \\ 1 - e^{-(x-\mu)/\sigma} & \text{for } \xi = 0 \end{cases}$$

- Asymptotic Pareto coefficient:

$$\lim_{p \to 1} b(p) = 1/(1-\xi)$$

  → If $\mu\xi - \sigma = 0$, strict power law
  → If $\mu\xi - \sigma > 0$ (resp. $<$), $b(p)$ converges from below (resp. above)

- Assume that the distribution follows a GPD for $p > p_K$
- Three parameters, identified using:
  1. last threshold
  2. last inverted Pareto coefficient
  3. differentiability condition at the jointure

Empirical Test

Alternative method #1

Method M1: Piecewise constant $b(p)$

- Assume $b(p)$ constant within each bracket
- Not entirely consistent with the input data
- Not always self-consistent

See Piketty (2001), Piketty and Saez (2003)

Alternative method #2

Method M2: Piecewise Pareto distribution

- Use $\log(1 - F(x)) = A - B \log(x)$ within each bracket
- Only use threshold information, not shares

See Kuznet (1953), Feenberg and Poterba (1992)

Alternative method #3

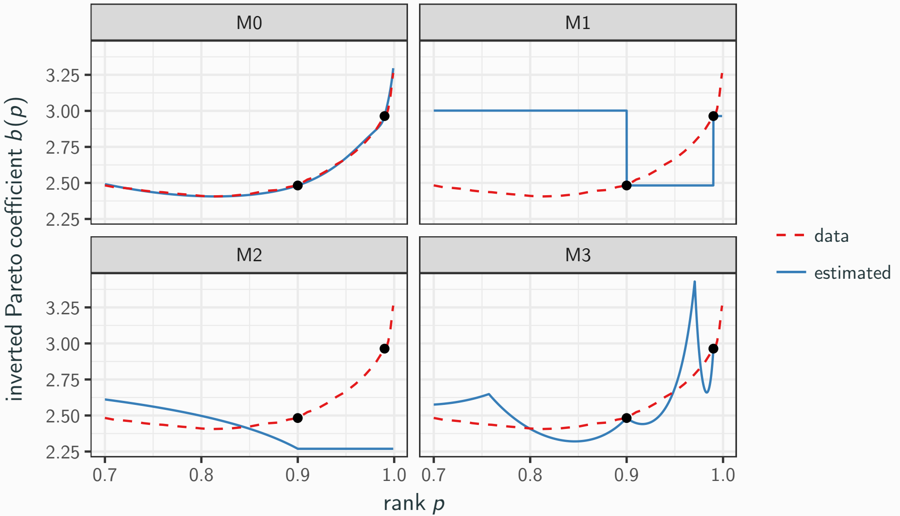
Method M3: Mean-split histogram

- Divide each bracket in two parts
- Define a uniform distribution on each part
- The breakpoint is the mean income inside the bracket

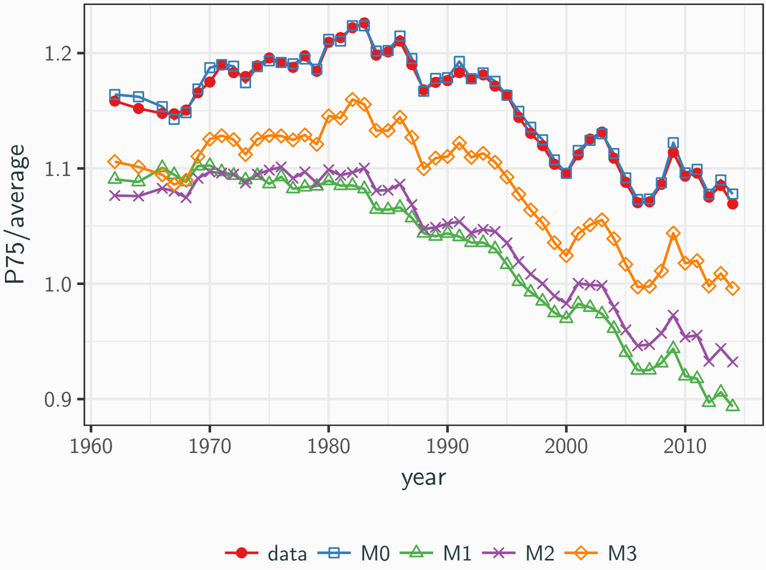
Test dataset

- Data from France and the United States coming from exhaustive or quasi-exhaustive micro-data:
  - France: Garbinti, Goupille-Lebret and Piketty (2016)
  - United States: Piketty, Saez and Zucman (2016)
- Create a tabulation $p = 10\%, 50\%, 90\%, 99\%$
- Compare estimated and actual value

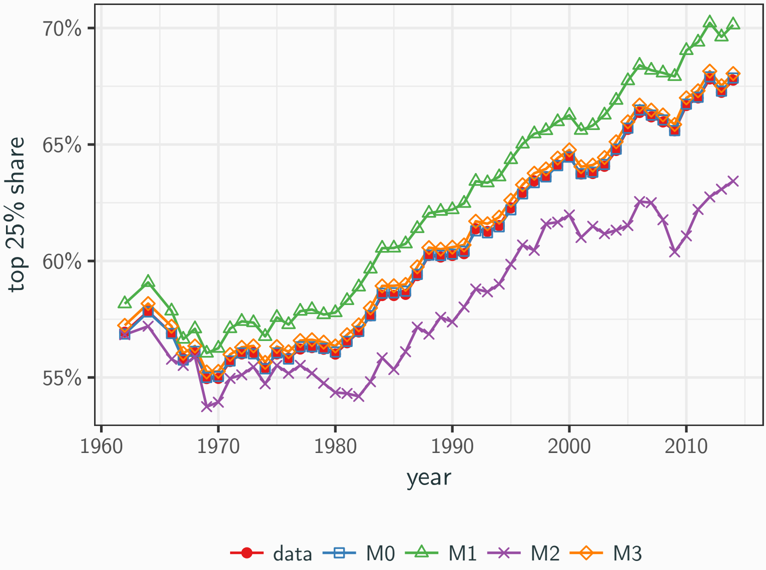
U.S. pre-tax national income, 2010: Generalized Pareto curve



U.S. pre-tax national income, P75/average



U.S. pre-tax national income, top 25% share



U.S. pre-tax national income, top 25% share (2000–2014)

Comparison table

		mean percentage gap between estimated and observed values			
		M0	M1	M2	M3
United States (1962–2014)	Top 70% share	0.059% (ref.)	2.3% (×38)	6.4% (×109)	0.054% (×0.92)
	Top 25% share	0.093% (ref.)	3% (×32)	3.8% (×41)	0.54% (×5.8)
	Top 5% share	0.058% (ref.)	0.84% (×14)	4.4% (×76)	0.83% (×14)
	P30/average	0.43% (ref.)	55% (×125)	29% (×67)	1.4% (×3.3)
	P75/average	0.32% (ref.)	11% (×35)	9.9% (×31)	5.8% (×18)
	P95/average	0.3% (ref.)	4.4% (×15)	3.6% (×12)	1.3% (×4.5)

Comparison table

		mean percentage gap between estimated and observed values			
		M0	M1	M2	M3
United States (1962–2014)	Top 70% share	0.059% (ref.)	2.3% (×38)	6.4% (×109)	0.054% (×0.92)
	Top 25% share	0.093% (ref.)	3% (×32)	3.8% (×41)	0.54% (×5.8)
	Top 5% share	0.058% (ref.)	0.84% (×14)	4.4% (×76)	0.83% (×14)
	P30/average	0.43% (ref.)	55% (×125)	29% (×67)	1.4% (×3.3)
	P75/average	0.32% (ref.)	11% (×35)	9.9% (×31)	5.8% (×18)
	P95/average	0.3% (ref.)	4.4% (×15)	3.6% (×12)	1.3% (×4.5)

Comparison table

		mean percentage gap between estimated and observed values			
		M0	M1	M2	M3
United States (1962–2014)	Top 70% share	0.059% (ref.)	2.3% (×38)	6.4% (×109)	0.054% (×0.92)
	Top 25% share	0.093% (ref.)	3% (×32)	3.8% (×41)	0.54% (×5.8)
	Top 5% share	0.058% (ref.)	0.84% (×14)	4.4% (×76)	0.83% (×14)
	P30/average	0.43% (ref.)	55% (×125)	29% (×67)	1.4% (×3.3)
	P75/average	0.32% (ref.)	11% (×35)	9.9% (×31)	5.8% (×18)
	P95/average	0.3% (ref.)	4.4% (×15)	3.6% (×12)	1.3% (×4.5)

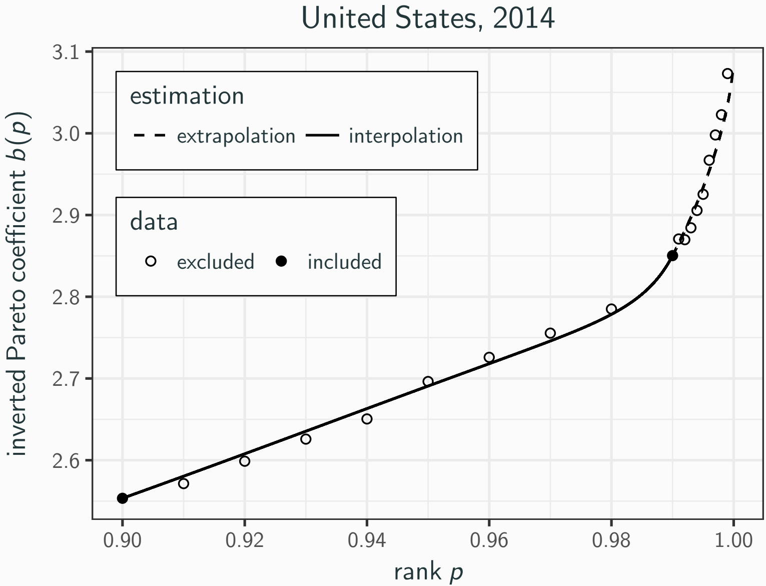
Comparison table

		mean percentage gap between estimated and observed values			
		M0	M1	M2	M3
United States (1962–2014)	Top 70% share	0.059% (ref.)	2.3% (×38)	6.4% (×109)	0.054% (×0.92)
	Top 25% share	0.093% (ref.)	3% (×32)	3.8% (×41)	0.54% (×5.8)
	Top 5% share	0.058% (ref.)	0.84% (×14)	4.4% (×76)	0.83% (×14)
	P30/average	0.43% (ref.)	55% (×125)	29% (×67)	1.4% (×3.3)
	P75/average	0.32% (ref.)	11% (×35)	9.9% (×31)	5.8% (×18)
	P95/average	0.3% (ref.)	4.4% (×15)	3.6% (×12)	1.3% (×4.5)

Comparison table

		mean percentage gap between estimated and observed values			
		M0	M1	M2	M3
United States (1962–2014)	Top 70% share	0.059% (ref.)	2.3% (×38)	6.4% (×109)	0.054% (×0.92)
	Top 25% share	0.093% (ref.)	3% (×32)	3.8% (×41)	0.54% (×5.8)
	Top 5% share	0.058% (ref.)	0.84% (×14)	4.4% (×76)	0.83% (×14)
	P30/average	0.43% (ref.)	55% (×125)	29% (×67)	1.4% (×3.3)
	P75/average	0.32% (ref.)	11% (×35)	9.9% (×31)	5.8% (×18)
	P95/average	0.3% (ref.)	4.4% (×15)	3.6% (×12)	1.3% (×4.5)

Extrapolation: Generalized Pareto curve



Extrapolation: Comparison table

→ Estimation of the top 1% from the top 10% and the top 5%

		mean percentage gap between estimated and observed values		
		M0	M1	M2
United States (1962–2014)	Top 1% share	0.78% (ref.)	5.2% (×6.7)	40% (×52)
	P99/average	1.8% (ref.)	8.4% (×4.7)	13% (×7.2)

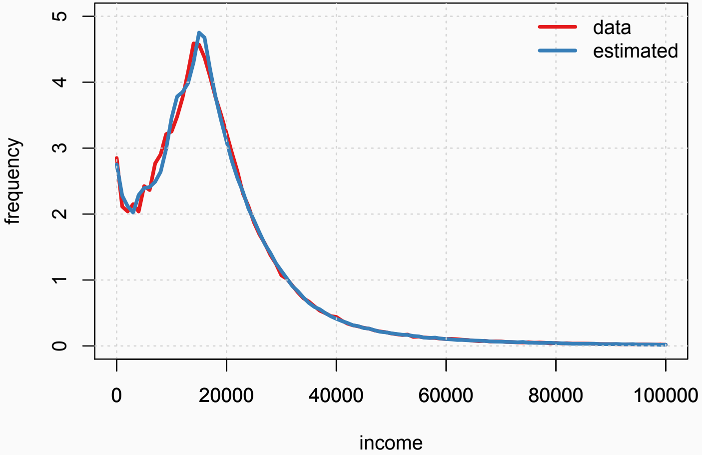
Comparison with a survey

What precision can we expect using subsamples of the data?

- U.S. distribution of pre-tax national income
- Mean percentage gap on the top 5% share:
  1. **Generalized Pareto interpolation:** 0.058%
     (tabulation: $p = 10\%, 50\%, 90\%, 99\%$)
  2. **Sample of** 10^7 **out of** 10^8**:** 0.44%

	mean percentage gap between estimated and observed values for a survey with simple random sampling and sample size n out of 10^8					
	$n = 10^3$	$n = 10^4$	$n = 10^5$	$n = 10^6$	$n = 10^7$	$n = 10^8$
Top 5% share	13.40%	6.68%	3.34%	1.34%	0.44%	0%
Top 1% share	27.54%	14.51%	7.39%	2.98%	0.97%	0%
Top 0.1% share	51.25%	33.08%	17.89%	7.41%	2.43%	0%

France pre-tax national income, 2006: Density comparison



(tabulation: $p = 10\%, 30\%, 50\%, 90\%, 99\%$)

Error Estimation

Which error?

- Two possible definitions of "the error"
  1. the error with respect to the **actual population value**
     - $\rightarrow$ we try to estimate the income/wealth in the population, approximated by a continuous distribution
  2. the error with respect to the **underlying statistical model**
     - $\rightarrow$ we only observe the realization of some underlying statistical model that we try to estimate
- In practice, sampling error and approximation error are negligible
  - $\Rightarrow$ errors 1 and 2 $\simeq$ same magnitude
- To fix ideas, focus on the second kind of error

Two components

- *What if the population was infinite?*
  - $\rightarrow$ there would still be an error because the actual distribution doesn't match our interpolation exactly
  - $\rightarrow$ **misspecification error**
- *What if the actual distribution matched the functional forms we use to interpolate?*
  - $\rightarrow$ there would still be an error due to sampling variability
  - $\rightarrow$ **sampling error**
- Total error $=$ misspecification error $+$ sampling error
- For simplicity, focus on unconstrained estimation

Sampling error

1. Finite variance case:
   - Standard approach (CLT-type results + delta method)
   - Asymptotic normality
2. Infinite variance case:
   - Generalized CLT (Gnedenko and Kolmogorov, 1968)
   - Convergence to a stable distribution

$\Rightarrow$ negligible sampling error in both cases

Misspecification error: theory

- Explicit expression for this error (Peano kernel theorem)

$$\text{misspecification error} = \int_{x_1}^{x_K} \varepsilon(x,t)\varphi'''(t)\mathrm{d}t$$

- Depends on two elements:
  1. the interpolation percentiles
  2. φ'''
- $\varphi''' \simeq$ residual
  - $\rightarrow$ captures all the features of the distribution not accounted for by the interpolation method

Misspecification error: applications

- Estimate φ''' when we have access to micro-data
- Plug-in these estimates in the error formula to:
  1. Get bounds on the error in the general case
  2. Solve the inverse problem: *how to place thresholds optimally?*

     optimal thresholds

Conclusion

- Generalized Pareto curves
  1. characterize and visualize the distributions of income or wealth
  2. estimate those distributions
- Generalized Pareto interpolation
  1. largely outperforms commonly used methods
  2. method applied to construct the World Inequality Database
  3. R package available:
     ☞ wid.world/gpinter
- What mechanisms?
  1. random growth models? can't account for the increasing inequality at the top
  2. simple deviations from them: the very top experiences higher growth and/or more risk
     $\Rightarrow$ the processes generating income and wealth distributions are not fully scale invariant

Thank you!

Additional slides

Power laws: Karamata's (1930) definition

- For some α, write $1 - F(x)$ as:

$$1 - F(x) = \mathbb{P}\{X > x\} = L(x)x^{-\alpha}$$

- X is an **asymptotic** power law if L is **slowly varying**:

$$\forall\lambda > 0 \qquad \lim_{x\to+\infty} \frac{L(\lambda x)}{L(x)} = 1$$

- Includes cases where $L(x) \to$ constant, but also (say) $L(x) = (\log x)^{\beta}$, $\beta \in \mathbb{R}$.

back

Thin tails: rapidly varying functions

- Otherwise, $1 - F$ may be **rapidly varying**, meaning:

$$\forall \lambda > 1 \qquad \lim_{x \to +\infty} \frac{1 - F(\lambda x)}{1 - F(x)} = 0$$

- That corresponds to **thin tailed** distributions:
    - Normal
    - Log-normal
    - Exponential
    - ...

▸ back

A simple typology of distributions

Category	Examples	$b(p)$ behavior
Power laws	Pareto Student's t Dagum	$\lim_{p\to 1} b(p) > 1$
Thin tails	Normal Log-normal Exponential	$\lim_{p\to 1} b(p) = 1$
Pathological cases	none	oscillates indefinitely (no convergence)

back

Optimal position of thresholds

	3 brackets	4 brackets	5 brackets	6 brackets	7 brackets
optimal placement of thresholds	10.0% 68.7% 95.2% 99.9%	10.0% 53.4% 83.4% 97.1% 99.9%	10.0% 43.0% 70.4% 89.3% 98.0% 99.9%	10.0% 36.8% 60.7% 80.2% 93.1% 98.6% 99.9%	10.0% 32.6% 53.3% 71.8% 86.2% 95.4% 98.9% 99.9%
maximum relative error on top shares	0.91%	0.32%	0.14%	0.08%	0.05%

▸ back