Generalized Pareto Curves: Theory and Applications

Thomas Blanchet [1]
Juliette Fournier [2]
Thomas Piketty [1]

WID.world conference
December 15th, 2017

[1]Paris School of Economics

[2]Massachusetts Institute of Technology

The Pareto distribution

- The distribution of income and wealth is well approximated by the Pareto (1896) distribution.
- Minimum income $x_0 > 0$.
- Linear relationship between log(rank) and log(income):

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

- Hence the "power law":

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

- $\alpha =$ Pareto coefficient

Beyond Pareto

- We introduce **generalized Pareto curves** to characterize power law behavior in a flexible way.
- We use them as the basis for an improved empirical method for using censored tax data.
- We connect their shape to the existing literature explaining Pareto behavior of income and wealth.

Generalized Pareto Curves

Generalized Pareto curves

- Assume a Pareto distribution with coefficient α.
- Quantile function $Q : p \mapsto x_0(1-p)^{-1/\alpha}$.
- Then:

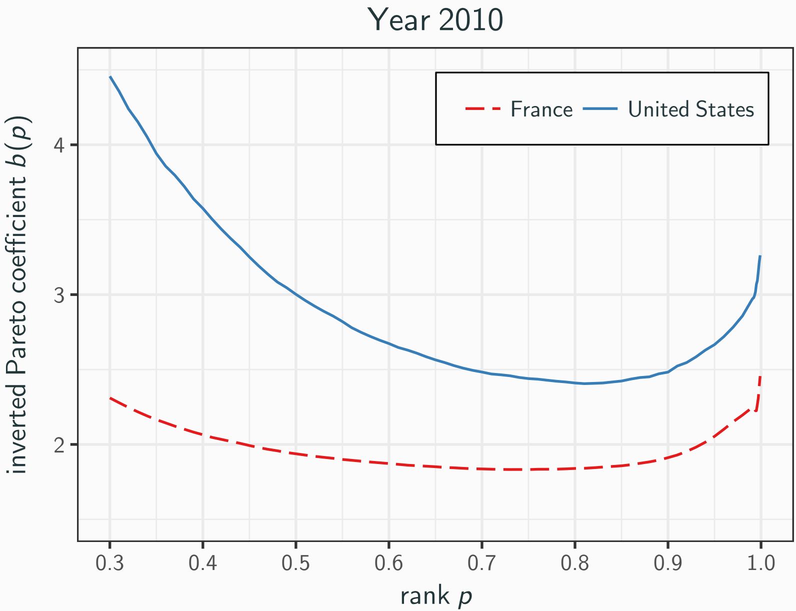
$$\frac{\mathbb{E}[X|X > Q(p)]}{Q(p)} = \frac{\alpha}{\alpha - 1} = b$$

  is constant.
- b is the **inverted Pareto coefficient**.
- More generally, we can define:

$$b(p) = \frac{\mathbb{E}[X|X > Q(p)]}{Q(p)}$$

- $p \mapsto b(p)$ is the **generalized Pareto curve**.

Pre-tax national income



Generalized Pareto Interpolation

An example of tabulated data

Tax data is typically available as:

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

- We know $Q(p)$ and $b(p)$ for a few values of p.
- How can we reconstruct an entire distribution based on that information?

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:

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

- Its derivative is directly linked to $b(p)$:

$$\varphi'(x) = 1/b(p)$$

- The initial problem is equivalent to interpolating the function φ, knowing, at the interpolation points:
  - its **value**
  - its **first derivative**

Spline interpolation

- Connect the dots with piecewise polynomials.
- Mathematical interpretation: find the "most regular" curve that fit the constraints of the problem.
- Polynomials of degree 5 (quintic splines) necessary to get enough regularity.
- Solving a linear system of equations.

Making sure the quantiles are increasing

- No guarantee yet that the quantile function is increasing. (Although most of the time it is.)
- Derive additional constraints to ensure increasing $Q(p)$.
- Add degrees of freedom and then solve the problem numerically.

Testing the method

Alternative method #1

Method M1: piecewise constant $b(p)$

- assumes $b(p)$ 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

- uses $\log(1 - F(x)) = A - B \log(x)$ within each bracket.
- only uses threshold information, not shares.

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

Alternative method #3

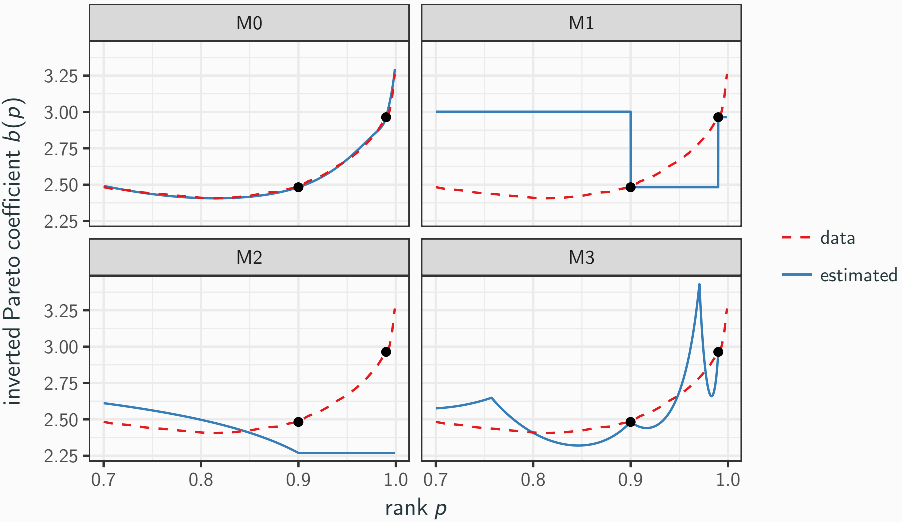
Method M3: mean-split histogram

- divide each bracket in two parts.
- defines a uniform distribution on each part.
- the breakpoint is the mean income inside the bracket.

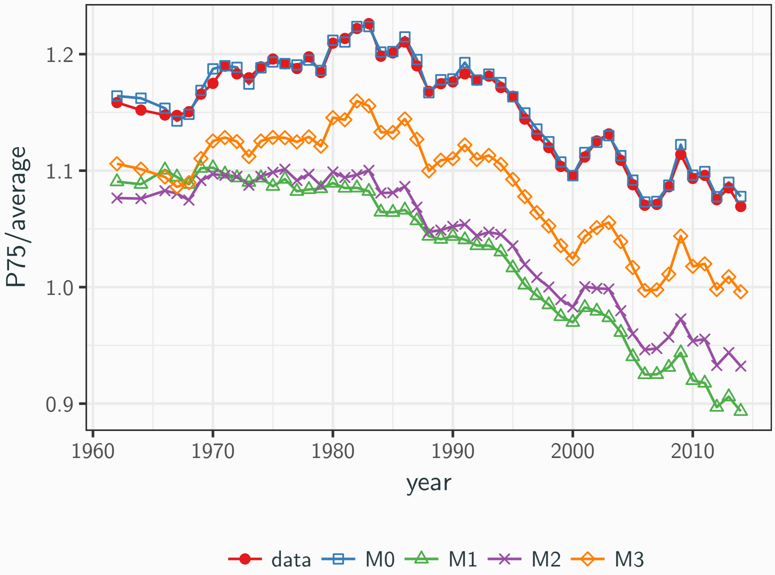
Test dataset

- Data from France and the United States coming from exhaustive or quasi-exhaustive microdata:
  - 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.

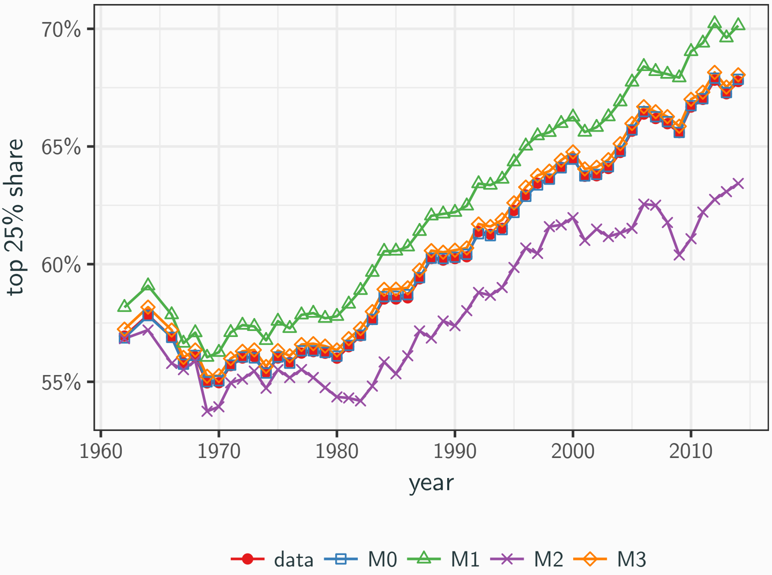
US pre-tax national income, 2010: generalized Pareto curve



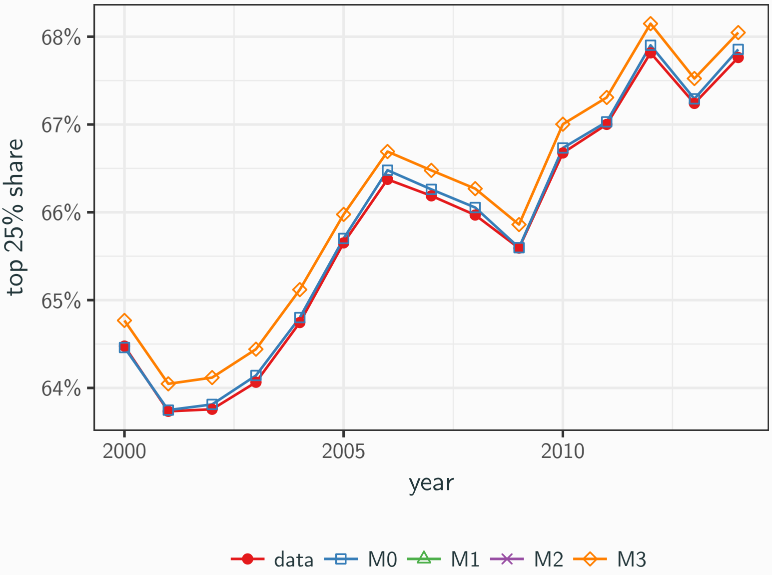
US pre-tax national income, P75/average



US pre-tax national income, top 25% share



US pre-tax national income, top 25% share (2000–2014)



Explaining the shape of wealth and income distributions

The basics

- Pareto distributions emerge through **proportional random growth** with frictions.
- Champernowne (1953), Simon (1955), Whittle (1957), Piketty and Zucman (2015), Gabaix et al. (2016), etc.
- Example: Kesten (1973) process.

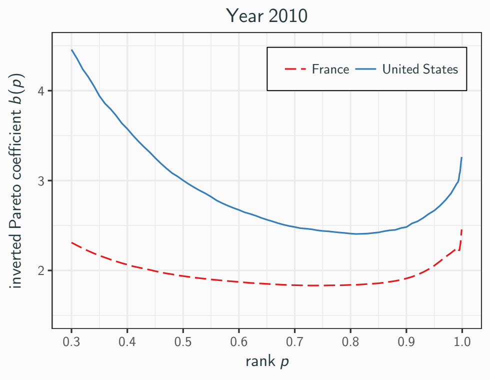
$$X_t = \omega_t X_{t-1} + \varepsilon_t$$

- Power-law tail: $1 - F(x) \sim Cx^{-\alpha}$ with

$$\mathbb{E}[|\omega_t|^\alpha] = 1$$

Limitations

It is hard to explain with those models why $b(p)$ increases toward the top of the distribution.



Need to make the variance and/or mean of multiplicative shocks vary with income.

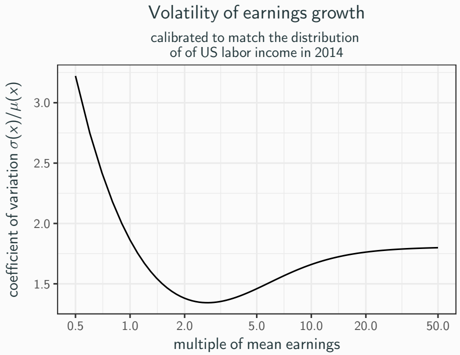
Generalized model

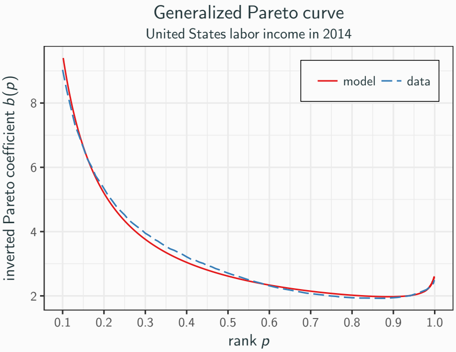
Two main results:

- Empirical literature noticed a U-shaped profile of the variance of income shocks (Chauvel and Hartung, 2014; Hardy and Ziliak, 2014; Bania and Leete, 2009; Guvenen et al., 2015).
- The U-shaped profile of $b(p)$ can be explained by a U-shaped profile of variance.

Simple calibration

US labor income, 2014. We assume constant mean, changing variance.

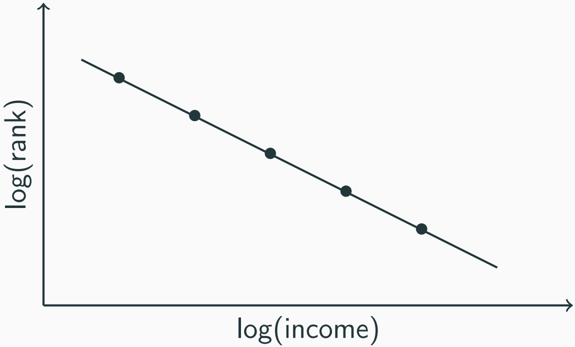




☞ wid.world/gpinter

Additional slides

The Pareto distribution



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
  - ...

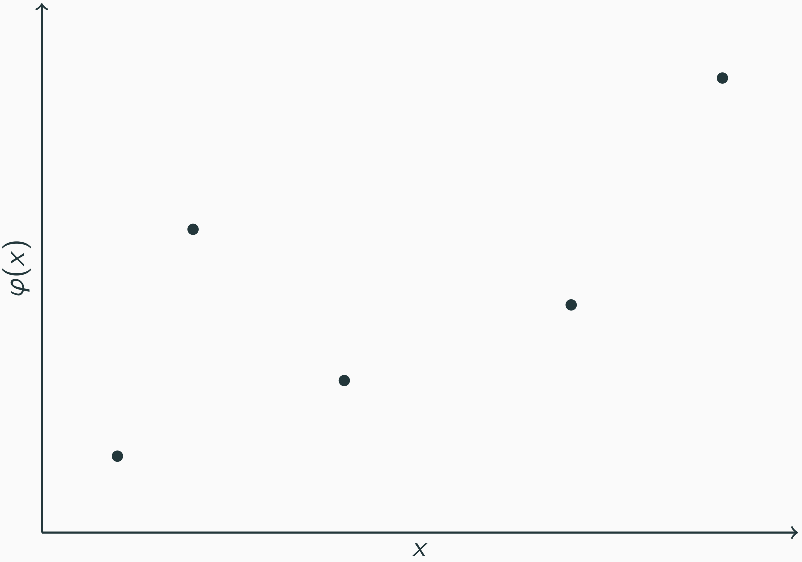
back

A simple typology of distributions

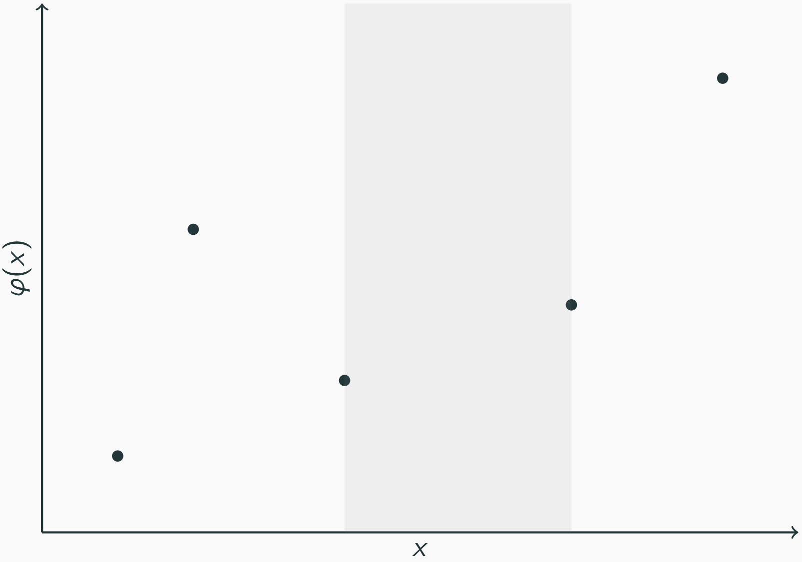
Category	Examples	$b(p)$ behavior
Power laws	Pareto Student's t Dagum	$\lim_{p \to 1} b(p) > 1$
Thin tails	Normal Log-normal Exponential	$\lim_{p \to 1} b(p) = 1$
Pathological cases	none	oscillates indefinitely (no convergence)

back

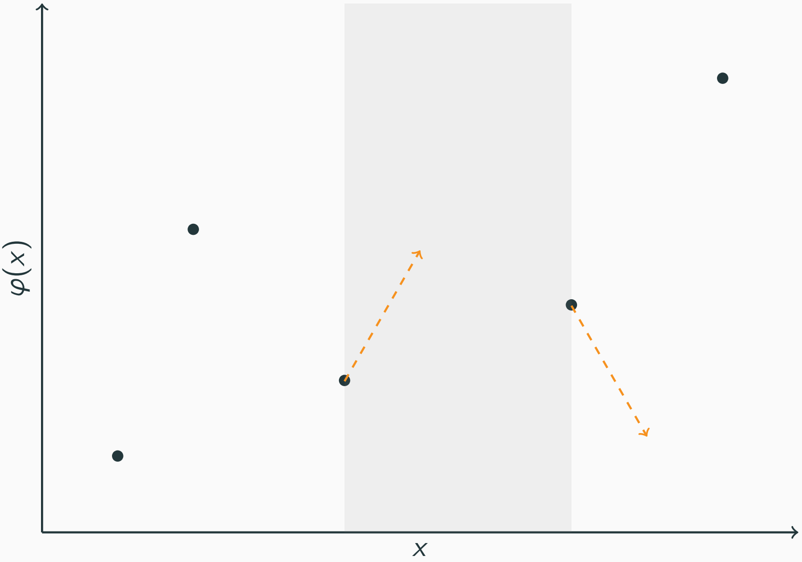
Cubic spline interpolation



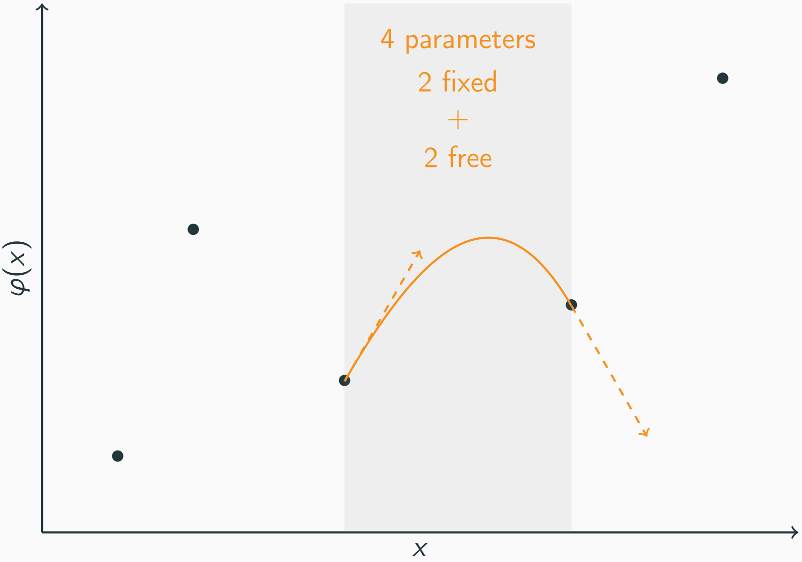
Cubic spline interpolation



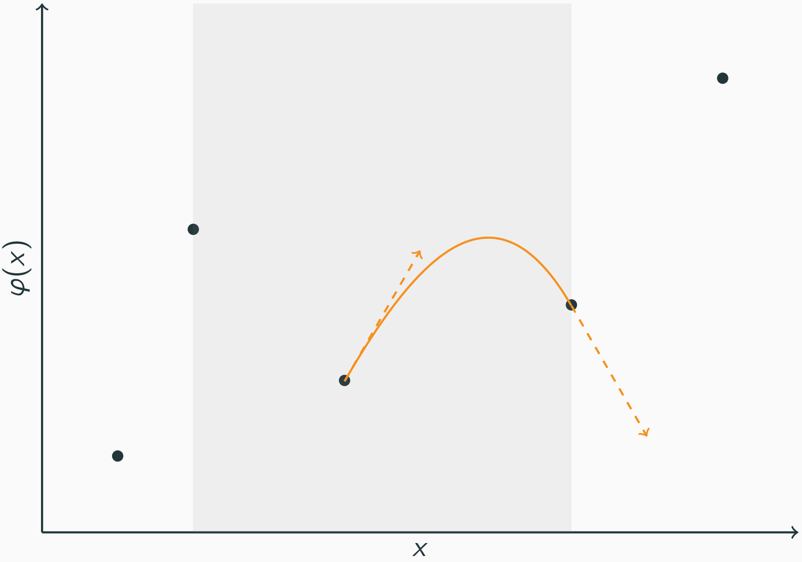
Cubic spline interpolation



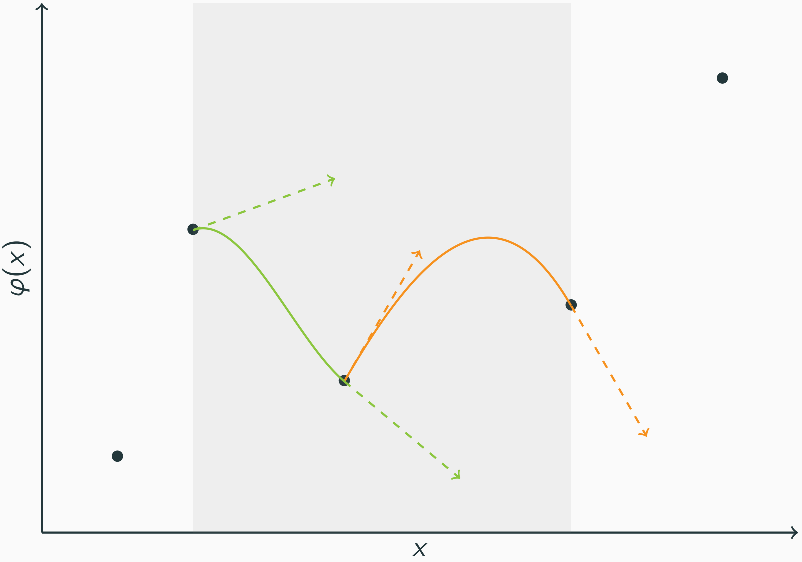
Cubic spline interpolation



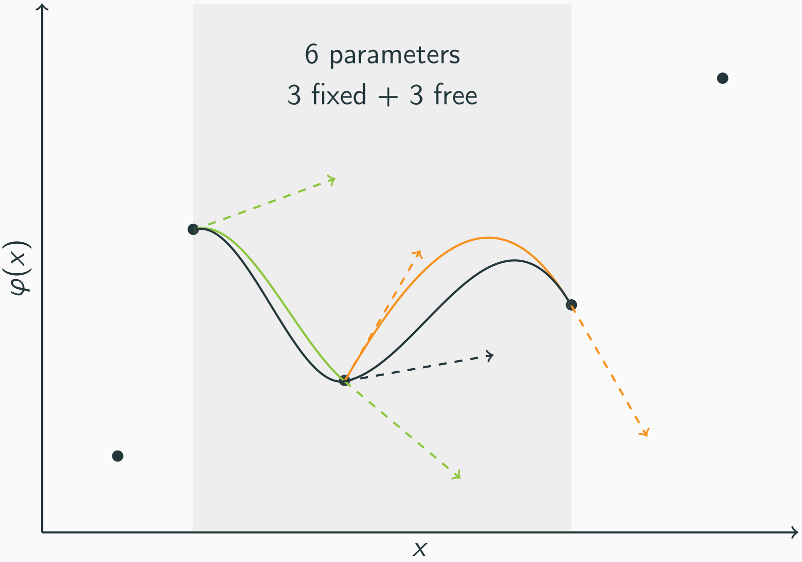
Cubic spline interpolation



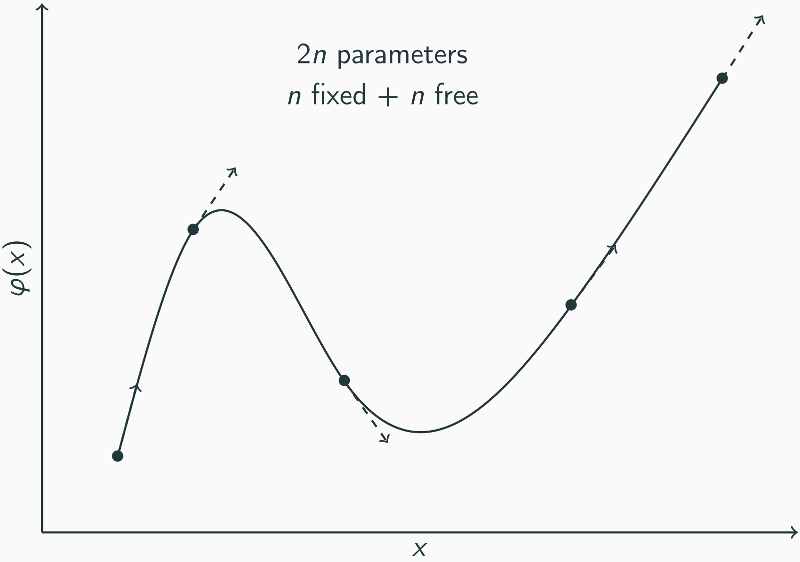
Cubic spline interpolation



Cubic spline interpolation



Cubic spline interpolation



Sufficient conditions for an increasing quantile function

Cargo and Shisha (1966)

Let $P(x) = c_0 + c_1 x_1 + \cdots + c_n x^n$ be a polynomial of degree $n \geq 0$ with real coefficients. Then:

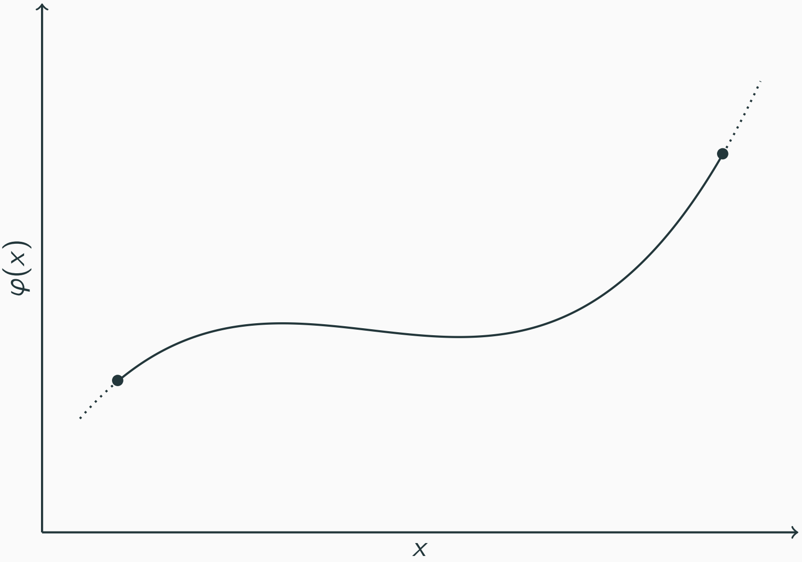
$$\forall x \in [0,1] \qquad \min_{0 \leq i \leq n} b_i \leq P(x) \leq \max_{0 \leq i \leq n} b_i$$

where:

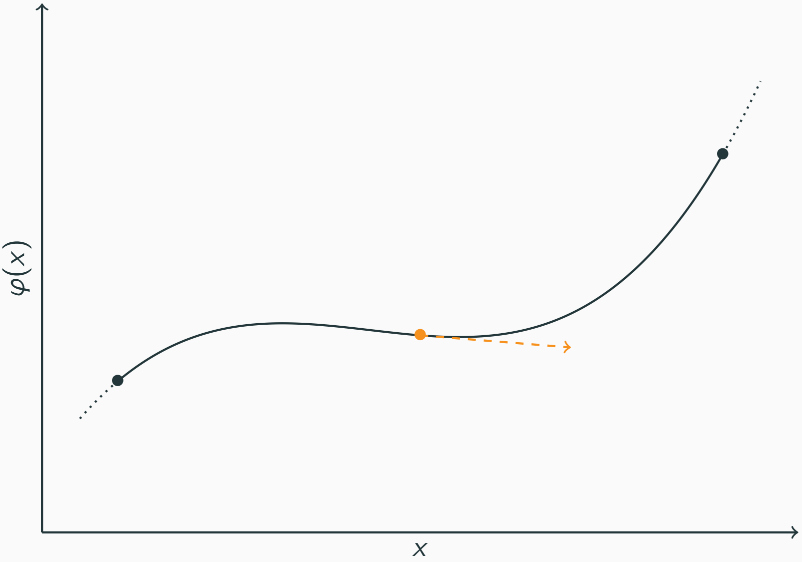
$$b_i = \sum_{r=0}^{n} c_r \binom{i}{r} \Big/ \binom{n}{r}$$

back

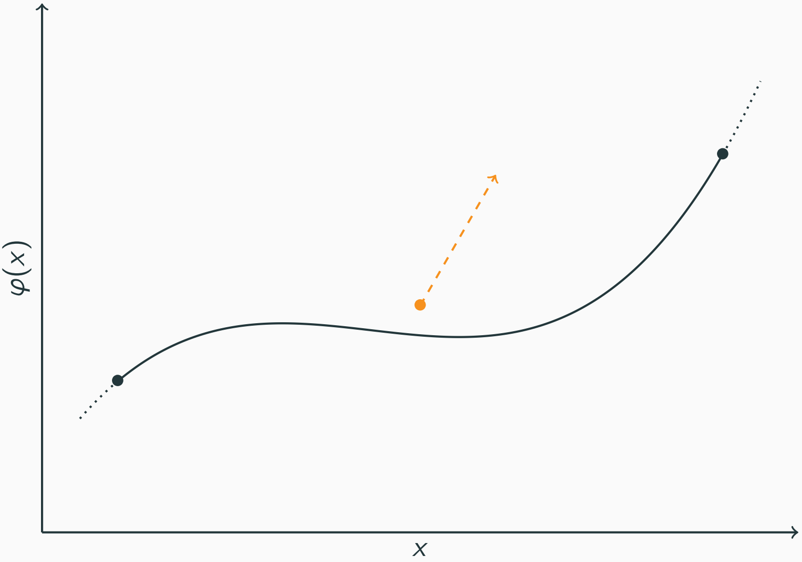
Making sure the quantiles are increasing



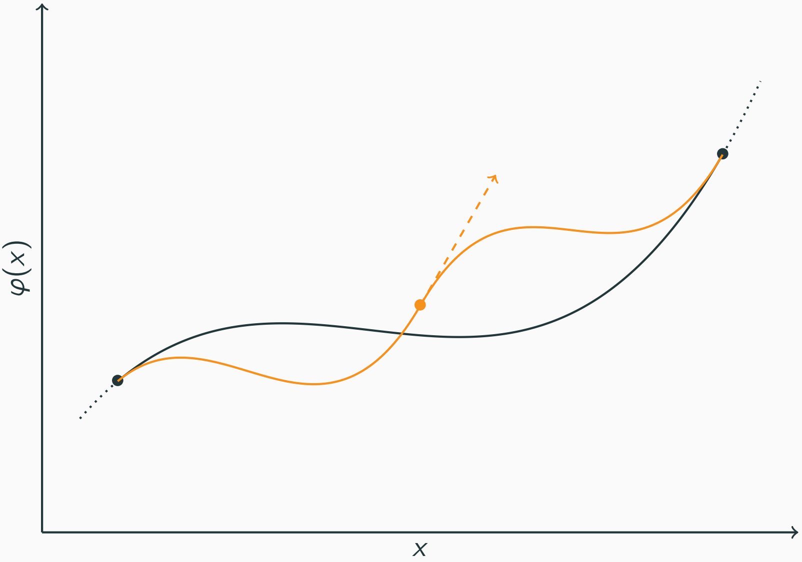
Making sure the quantiles are increasing



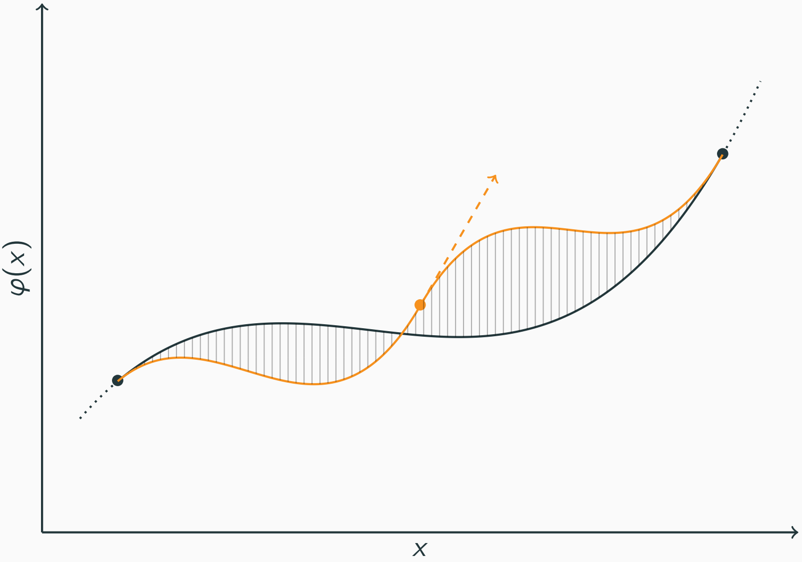
Making sure the quantiles are increasing



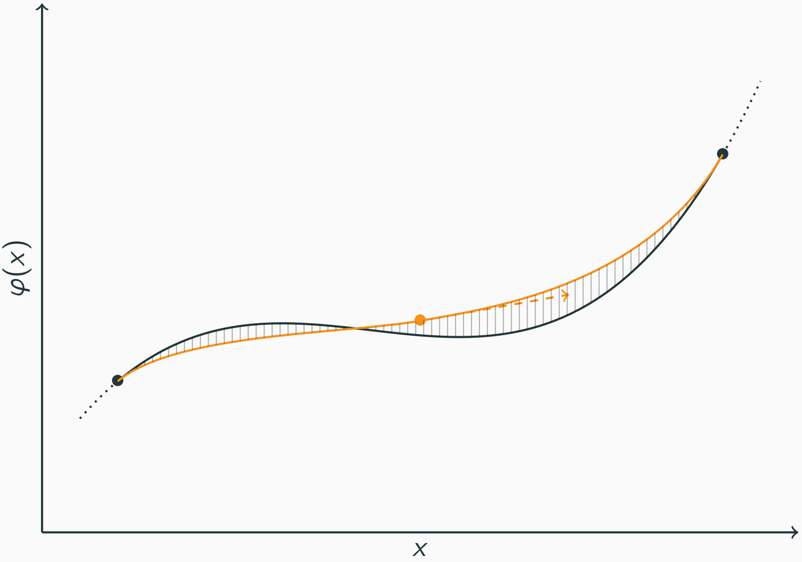
Making sure the quantiles are increasing



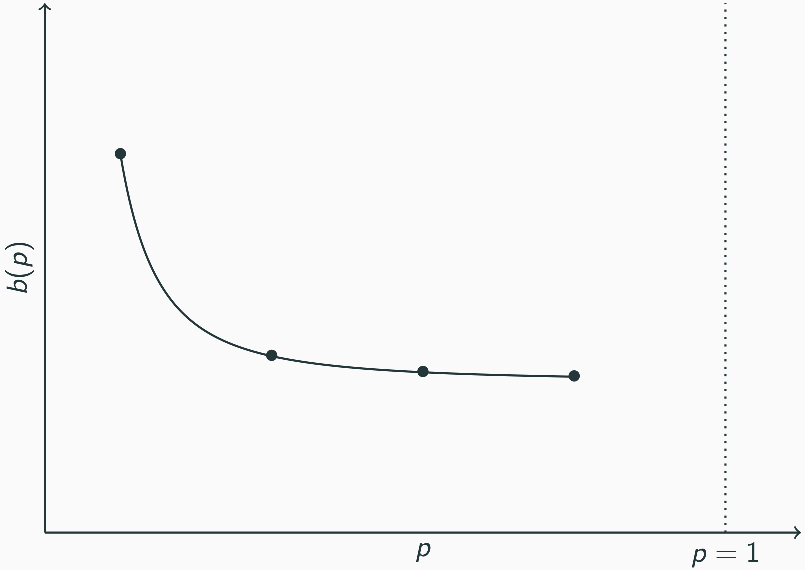
Making sure the quantiles are increasing



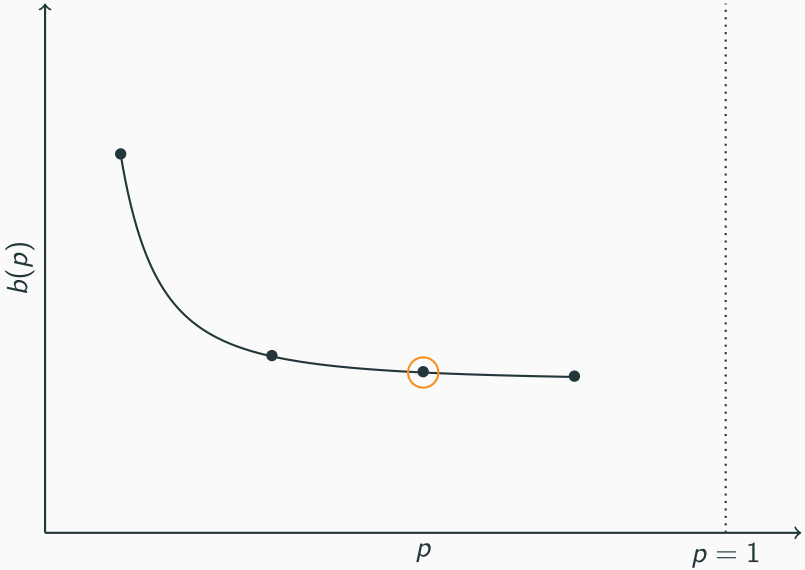
Making sure the quantiles are increasing



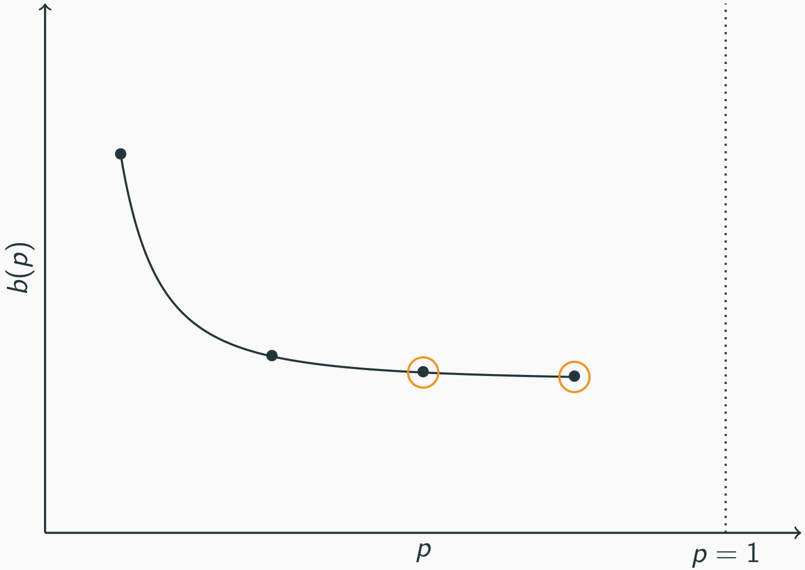
Extrapolation beyond the last threshold



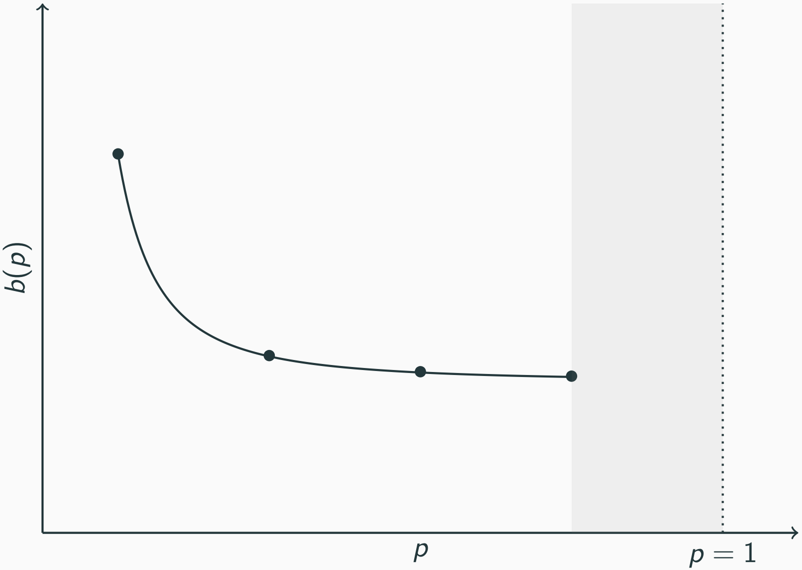
Extrapolation beyond the last threshold



Extrapolation beyond the last threshold



Extrapolation beyond the last threshold



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)

Compare with a survey

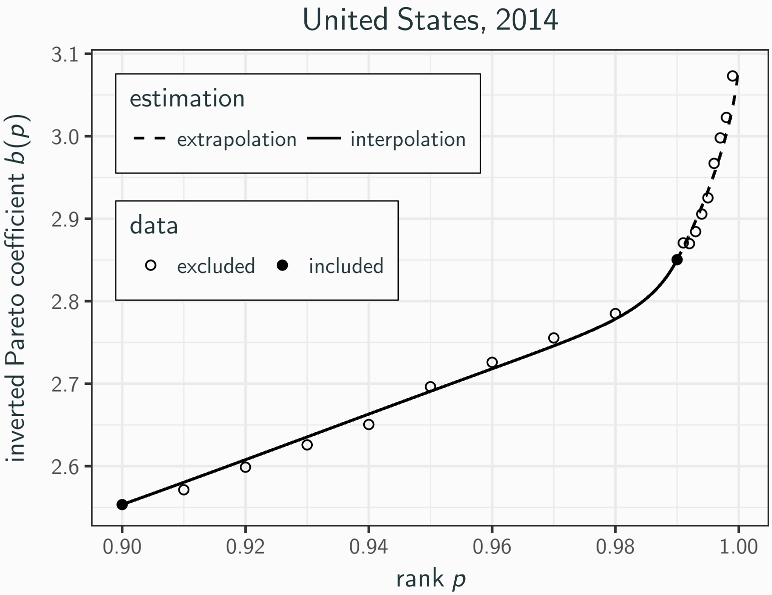
Take the example of the US distribution of pre-tax national income.
What precision can we expect using subsamples of the data?

	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%

Which error?

- Two possible definitions of "the error"
  - the error with respect to the **actual population value**.
  - the error with respect to the **underlying statistical model**.
- The first error is more relevant for causal inference.
- We consider the second kind of error.

Extrapolation: generalized Pareto curve



Extrapolation: comparison table

Estimating 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)

Two components

- What if the population was infinite (no sampling variability)?
  - There would still be an error because the actual distribution doesn't match our interpolation exactly.
  - **misspecification error**
- What if the actual distribution matched the functional forms we use to interpolate?
  - There would still be an error due to sampling variablity.
  - **sampling error**
- The total error is the sum of both.
- We'll focus on the unconstrained estimation (explicit results, still covers the overwhelming majority of cases).

Sampling error

- Finite variance case:
  - Standard approach (CLT-type results + delta method).
  - Asymptotic normality.
- Infinite variance case:
  - Generalized CLT (Gnedenko and Kolmogorov, 1968).
  - Converge to a stable distribution.
- In both cases: sampling error negligible.

Misspecification error: theory

- Explicit expression for this error (Peano kernel theorem).
- Depends on two elements:
  - the interpolation percentiles.
  - φ'''
- Think of φ''' as a residual. It captures all the features of the distribution not taken into account by the interpolation method.

Misspecification error: applications

- We estimate φ''' when we have access to microdata.
- We can plug-in these estimates in the error formula to:
  - get bounds on the error in the general case.
  - solve the inverse problem: how to place thresholds optimally?

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%