Inequality and Institutions: **Global Inequality**

Ignacio Flores

PSE Summer School

June, 2026

The challenge of measurement

Contents of section 3

3. The challenge of measurement

3.1 The mathematics of extreme events
- The essence of power laws
- The fractal nature of power laws
- Pareto's version of a power law

3.2 Sources of information
- National Accounting
- Households surveys
- Administrative data

3.3 The inequality toolbox
- Pareto interpolations
- Microdata combination

3.4 Measuring national wealth
- Taking stock: aggregate wealth
- Direct and indirect measurement methods

The mathematics of extreme events

Modelling extreme events

Many models assume that random variables behave normally:

$$X \sim \mathcal{N}(\mu, \sigma^2).$$

- Extreme deviations from the mean are unlikely.
- Averages stabilize with growing samples (LLN).

But many natural and social systems behave very differently:

$$P(X > x) \propto x^{-\alpha}.$$

- Where a small number of observations can dominate totals.
- And the mean or variance may not even exist.

Power laws describe such systems.

Power Laws in Nature (I): Biology & Ecology

Kleiber's Law (Metabolism).
Metabolic rate scales with body mass across species. This pattern is often linked to the fractal structure of vascular networks and resource distribution. (Kleiber, 1932; West, Brown, and Enquist, 1997)

Forest Patch Sizes.
The distribution of forest patch areas and fire sizes follows a heavy-tailed pattern. (Malamud, Morein, and Turcotte, 1998)

Power Laws in Nature (II): Earth & Space

Earthquakes (Gutenberg–Richter Law).
The frequency of earthquakes declines as a power of their magnitude: large events are rare but far more common than under thin-tailed models. (Gutenberg and Richter, 1944)

Initial Mass Function of Stars (Salpeter IMF).
The masses of newly formed stars follow a power-law tail, with a small number of very massive stars contributing disproportionately to total stellar mass. (Salpeter, 1955)

Power Laws in Society

City Size Distributions (Zipf's Law).
City sizes follows a Zipf law. This regularity appears across many countries and time periods. (Zipf, 1949)

Firm Sizes.
Firm employment and sales exhibit a Pareto-like upper tail. (Axtell, 2001)

Income and Wealth Distributions.
The upper tails of income and wealth follow Pareto laws: a small fraction of individuals holds a large share of aggregate resources. (Pareto, 1897; Gabaix, 2009; Benhabib, Bisin, and Zhu, 2011)

Power Laws and Fractals

If a system has no characteristic scale (self-similar), then many observables must transform in a simple multiplicative way (Mandelbrot, 1982). This leads to relationships of the form

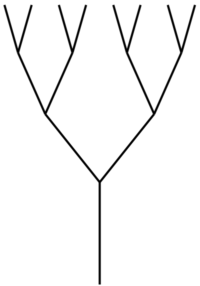
$$P(X > x) \propto x^{-\alpha},$$

Fractals describe the *geometry* of self-similar structures; power laws describe the *statistics* of their sizes and events.

Power Laws and Fractals

Self-similarity.
Fractals are objects that look “similar” at different magnifications: zooming in reveals structures that resemble the whole.

Scale invariance and power laws.
In scale-free systems, there is no characteristic size: rescaling distances simply multiplies observables by a constant factor.



A simple fractal-like tree: zooming in repeats the same pattern.

Defining power laws

A power law is a functional relationship of the form:

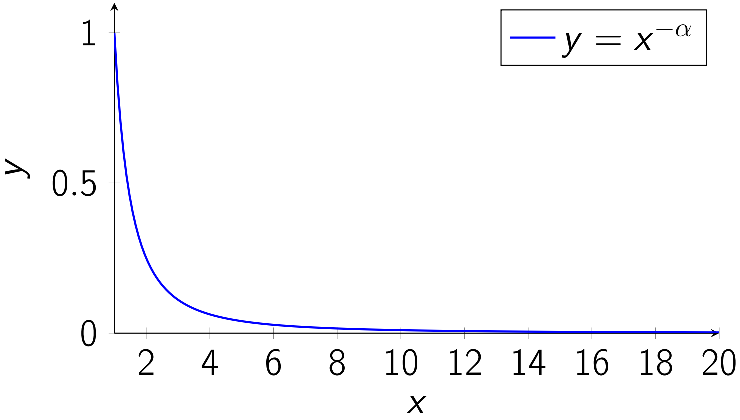
$$y = Cx^{-\alpha}, \qquad C > 0,\ \alpha > 0.$$

Taking logarithms:

$$\log y = \log C - \alpha \log x,$$

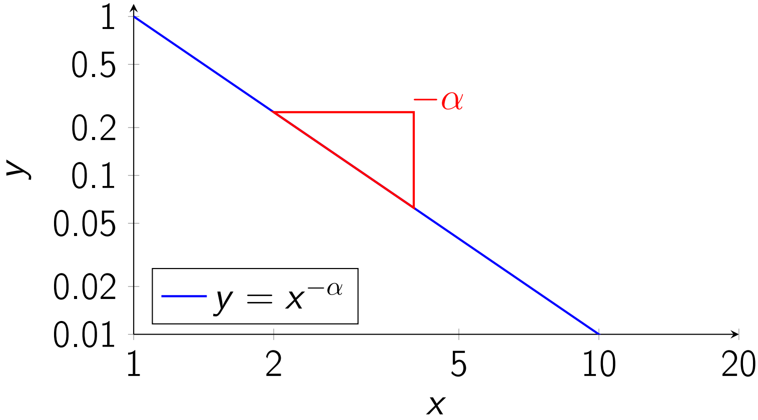
which is a straight line on log–log scales.

Power laws on linear scale ($\alpha = 2$)



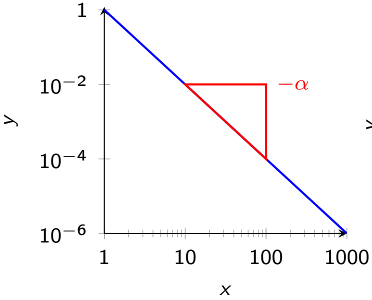
Power laws decay slowly: the tail remains thick for large x.

Power laws on log–log scale ($\alpha = 2$)

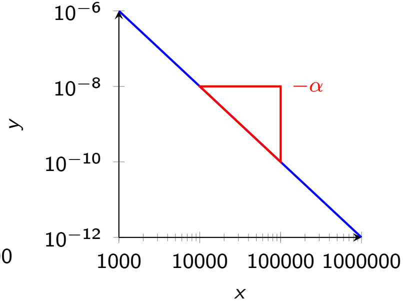


It becomes straight in logarithmic coordinates.

Power laws are fractal in nature (self-similarity)



Global view (1 to 10^6)



Zoom on the tail

A power law is scale-free: zooming in preserves its shape.

Pareto's version of a power law

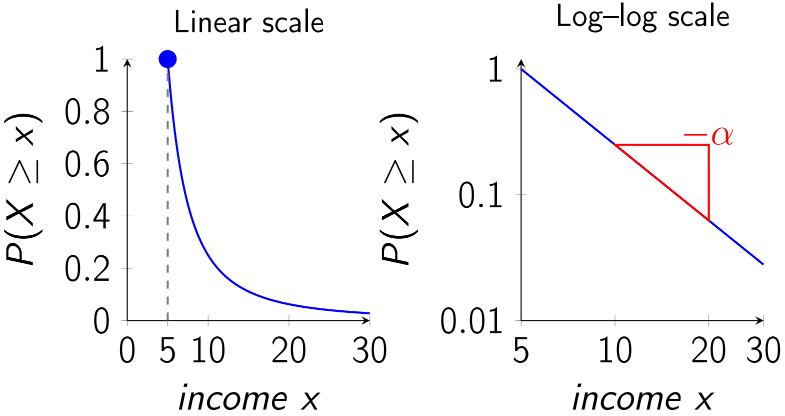
- Pareto, 1897: the upper tail of income follows a power law.
- He expressed this through the survival function (complementary CDF). Defining the fraction of people who earn at least x as:

$$P(X \geq x) = \left(\frac{x_{\text{min}}}{x}\right)^{\alpha}, \qquad x \geq x_{\text{min}}, \ \alpha > 0.$$

This is a direct application of the power-law form:

$$P(X \geq x) = Cx^{-\alpha}, \quad C = x_{\text{min}}^{\alpha}.$$

Example ($\alpha = -2, x_{min} = 5$)



In the Pareto distribution, the constant C becomes x_{min}^{α}, with the α exponent ensuring $P = 1$ at $x = x_{min}$, effectively shifting the distribution across incomes.

The Pareto Distribution

Moments:

$$\mathbb{E}[X] = \begin{cases} \frac{\alpha x_{\text{min}}}{\alpha - 1}, & \alpha > 1, \\ \text{undefined}, & \alpha \leq 1, \end{cases}$$

$$\mathrm{Var}(X) = \begin{cases} \frac{\alpha x_{\text{min}}^2}{(\alpha-1)^2(\alpha-2)}, & \alpha > 2, \\ \text{infinite}, & 1 < \alpha \leq 2. \end{cases}$$

Thus:

- Mean exists only if $\alpha > 1$.
- Variance exists only if $\alpha > 2$.

Sources of information

Making sense of contradicting sources

- **Observation:** Large discrepancies across sources (trends, levels, compositions)
- **Problem:** Can we make meaningful comments on the inequality-growth relation if we only capture a fraction of growth?
- **Challenge:** Find a way to use the strengths of each source to compensate for the weaknesses of others

National Accounts

Pros

- The main reference for macroeconomic statistics
- Accounts for the whole economy and its sectors
- Multiple sources consistently combined
- International standard

Cons

- Only reports aggregates, no granular distribution (yet)
- The level of detail is often lower in developing countries

The distribution of primary income, Households

Uses			Resources		
			B2n	**Operating surplus, net**	**69**
			B3n	**Mixed income, net**	**53**
D4	Property income	41	D1	Compensation of employees	1154
D41	*Interest*	*14*	*D11*	*Wages and salaries*	*954*
D45	*Rents*	*27*	*D12*	*Employers' social contributions*	*200*
			D4	Property income	126
			D41	*Interest*	*14*
			D42	*Distributed income of corporations*	*20*
			D43	*Reinvested earnings on foreign direct investment*	*3*
			D44	*Other investment income*	*30*
			D441	*Investment income attributable to insurance policyholders*	*20*
			D442	*Investment income payable on pension entitlements*	*8*
			D443	*Investment income attributable to collective investment fund share holders*	*2*
			D45	*Rents*	*21*
				Income from offshore tax havens	*3*
B5n	**Primary income, net**	**1361**			

- **Notes.** Each sector's resources are recorded at the right hand. They follow the sequence from income generation to savings and capital formation. Source adapted from the SNA 2008

The secondary distribution of income

Uses			Resources		
			B5n	**Primary income, net**	**1361**
D5	Current taxes on income, wealth, etc.	178	D62	Social benefits other than social transfers in kind	384
D61	Net social contributions	333	*D621*	*Social security benefits in cash*	53
D611	*Employers' actual social contributions*	181	*D6211*	*Social security pension benefits*	45
D6111	*Employers' actual pension contributions*	168	*D6212*	*Social security non-pension benefits in cash*	8
D6112	*Employers' actual non-pension contributions*	13	*D622*	*Other social insurance benefits*	279
D612	*Employers' imputed social contributions*	19	*D6221*	*Other social insurance pension benefits*	250
D6121	*Employers' imputed pension contributions*	18	*D6222*	*Other social insurance non-pension benefits*	29
D6122	*Employers' imputed non-pension contributions*	1	*D623*	*Social assistance benefits in cash*	52
D613	*Households' actual social contributions*	129			
D6131	*Households' actual pension contributions*	115			
D6132	*Households' actual non-pension contributions*	14			
D614	*Households' social contribution supplements*	10			
D6141	*Households' pension contribution supplements*	8			
D6142	*Households' non-pension contribution supplements*	2			
B6n+D7	**Secondary income, net**	**1234**			

- **Notes.** Income taxes, contributions and social transfers are recorded at this stage. Source adapted from the SNA 2008

Household surveys

Pros

- Rich micro-level source with an array of variables
- Less than 1 pct of the population sampled
- Comparability increased over time, since 1960s
- Harmonized sources available (EU-SILC, LIS, WIID)
- Somewhat captures the informal economy

Cons

- Subject to both sampling and non-sampling biases
- Captures a fraction of national income at best

Household survey limitations

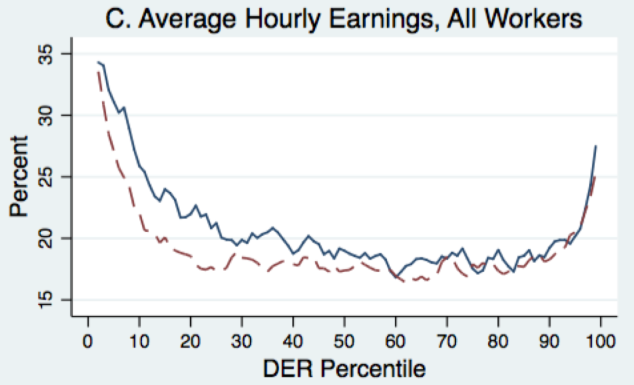
Sampling

- Small sample bias (Taleb and Douady, 2015)

Non-Sampling

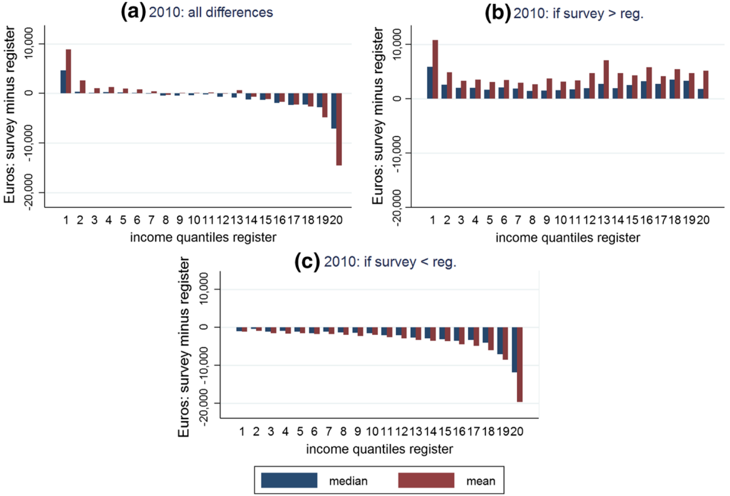
- Misreporting (Bound and Krueger, 1991; Bollinger, 1998; Paulus, 2015; Angel, Heuberger, and Lamei, 2018)
- Heterogeneous response rates across the income distribution (Korinek, Mistiaen, and Ravallion, 2006; Johansson and Klevmarken, 2007; Bollinger et al., 2014; Chenevert, Klee, and Wilkin, 2016)

U-shaped nonresponse



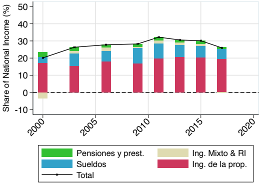
- **Notes.** Men are in blue and women in red. (Bollinger et al., 2019). Compares declared incomes in the CPS and Social Security records in the US. Accounting for 1/3-1/2 of the measured gap in inequality

Over-reporting bottom and under-reporting top

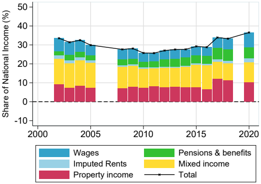


- **Notes.** Misreporting is mean reverting but non deterministic. Austrian SILC (survey), individuallly matched to income tax declarations. Source: Angel, Heuberger, and Lamei, 2018

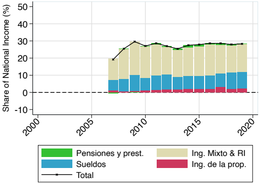
Aggregate income under-coverage in surveys



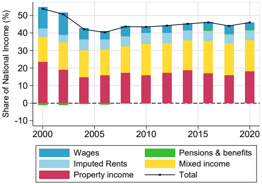
(a) Chile



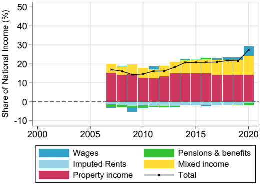
(b) Colombia



(c) Ecuador



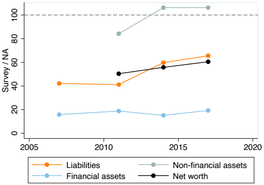
(d) México



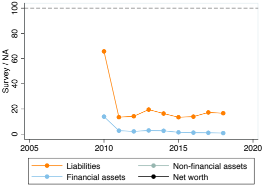
(e) Perú

- **Notes.** Comparing Latin American surveys to national accounts aggregates unveils large gaps, especially in capital income measurement, which are particularly relevant for inequality analysis. Source: Alvaredo et al. (2022)

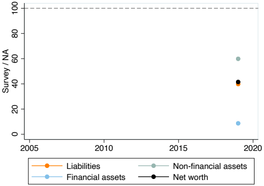
Aggregate wealth survey under-coverage



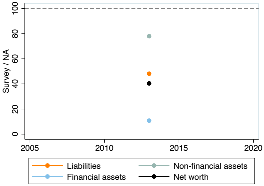
(f) Chile



(g) Colombia



(h) Mexico



(i) Uruguay

- **Notes.** A similar phenomenon is observed in wealth surveys, where gaps are even larger. Financial wealth is measured with the largest discrepancy across sources. Source: Carranza, De Rosa, and Flores (2023)

Administrative data (1/2)

Pros

- Used historically to estimate income proportions of affluent groups (Pareto, 1896; Kuznets and Jenks, 1953; Atkinson and Harrison, 1978, Alvaredo et al., 2013)
- High time coverage, and precision at the very top
- Direct measurements from the universe, rarely sampled
- Enforced by law

Cons

- Limited population coverage in developing countries and in historical series
- Changes in the tax code can influence measurement
- Economic incentives to under-declare

Administrative data (2/2)

Practical considerations

- Tax return information is confidential, with strict security.
- Data aggregation is common to protect individual privacy.
- Interpolation techniques are needed.

Applications in Developing Countries

- They rarely provide complete population coverage.
- Research focuses on estimating top income shares.
- Methodological challenge: low coverage and large informality.
- Importance of publishing data for greater fiscal transparency.

Brazilian administrative data

Figure: Brazilian income tax declarations - 2020 - tabulated format

Receita Federal
CETAD

GRANDES NÚMEROS DIRPF 2021 - ANO-CALENDÁRIO 2020

Tabela 9 - Resumo das Declarações Por Faixa de Rendimentos Totais (em salários-mínimos)

Valores em R$ milhões

Faixa SM Mensal	Qtde Declarantes	Rendim. Tribut.	Rendim. Tribut. Exclus.	Rendim. Isentos	Imposto Devido	Dedução de Incentivo	Imposto Devido 1	Imposto a Pagar	Doações Efetuadas
Até 1/2	1,604,246	1,110.77	146.77	263.62	0.02	0.00	0.02	0.01	957.11
De 1/2 a 1	578,381	4,975.59	242.51	699.00	0.15	0.00	0.15	0.08	184.88
De 1 a 2	2,755,245	43,655.42	1,623.49	6,499.02	1.13	0.00	1.13	0.61	562.12
De 2 a 3	5,014,845	135,856.84	8,796.73	15,271.23	404.80	0.00	404.80	83.86	872.45
De 3 a 5	8,262,962	322,633.21	26,223.85	54,106.30	6,348.92	0.43	6,348.49	1,014.70	1,783.86
De 5 a 7	4,316,910	245,948.76	22,023.01	50,870.07	12,338.48	1.73	12,336.75	1,923.85	1,759.58
De 7 a 10	3,219,144	249,138.21	24,438.66	62,356.49	22,211.34	5.88	22,205.47	3,468.77	31,466.02
De 10 a 15	2,526,189	269,857.80	29,508.51	85,199.97	35,293.13	18.76	35,274.37	4,368.64	2,858.15
De 15 a 20	1,167,190	165,207.33	20,683.24	66,463.50	26,005.52	23.44	25,982.07	2,616.15	2,245.76
De 20 a 30	1,073,869	197,198.67	28,396.73	100,128.29	34,473.24	44.63	34,428.62	2,954.61	3,728.53
De 30 a 40	462,066	110,538.55	18,533.01	70,142.24	20,747.34	37.83	20,709.51	1,698.10	28,512.49
De 40 a 60	340,832	94,559.36	21,735.02	88,543.26	18,465.74	41.61	18,424.13	1,764.24	8,756.65
De 60 a 80	117,987	35,199.37	12,391.79	53,803.75	7,119.31	18.50	7,100.81	896.25	97,696.69
De 80 a 160	122,641	42,143.33	21,567.30	102,302.42	8,894.52	25.53	8,868.98	1,487.84	5,081.24
De 160 a 240	30,762	13,990.22	10,278.35	50,327.53	3,038.27	11.36	3,026.91	682.28	3,046.41
De 240 a 320	13,567	7,779.53	7,135.74	31,942.87	1,710.26	7.54	1,702.72	422.01	1,915.07
Mais de 320	28,007	35,640.27	77,873.98	257,190.17	7,784.50	75.42	7,709.08	1,880.57	24,834.56
Total	**31,634,843**	**1,975,433.24**	**331,598.71**	**1,096,109.71**	**204,836.68**	**312.66**	**204,524.02**	**25,262.58**	**216,261.57**

- **Notes.** Observations are not accessible at the individual level (unless you work at the tax agency). The level of detail varies across countries and time, both in observations and in the composition of declared incomes/assets.

The inequality toolbox

A truncated distribution - Chilean tax data

		Consolidado		
Año Comercial	**Tramo de Rentas**	**N° de Personas**	**Renta Determinada**	**Impuesto Determinado**
2019	Tramo 1 - 0 a 13,5 UTA (Exento)	7 749 288	23 169 595	9 861
2019	Tramo 2 - 13,5 a 30 UTA (Tasa 4%)	1 761 945	20 501 037	252 113
2019	Tramo 3 - 30 a 50 UTA (Tasa 8%)	499 887	11 301 031	383 221
2019	Tramo 4 - 50 a 70 UTA (Tasa 13,5%)	183 369	6 412 004	374 275
2019	Tramo 5 - 70 a 90 UTA (Tasa 23%)	89 156	4 189 123	369 942
2019	Tramo 6 - 90 a 120 UTA (Tasa 30,4%)	67 326	4 133 703	537 303
2019	Tramo 7 - Más de 120 UTA (Tasa 35%)	86 178	12 155 716	3 040 973

p	thr	bracketavg
0,45423	0	3139515
0,859628	8038926	12209919
0,951524	17864280	23729774
0,977634	29773800	36805552
0,987234	41683320	49826944
0,991909	53592840	66006266
0,99544	71457120	1,56E+08

- **Notes.** Tabulations give precise estimates of cumulative densities at arbitrary points in the distribution. But how can we get a smooth Lorenz curve with that information?

Connecting the dots (1/3)

Simple Pareto Interpolation

- The Pareto distribution follows a power law density function.
- The shape parameter (α) defines the slope at the top.
- Higher α value suggests more equitable income distribution.
- Smooths income distributions with continuous estimates from aggregated data, allowing estimation at specific brackets.

Density Function

$$f(x) = \frac{\alpha x_0^\alpha}{x^{\alpha+1}}$$

- x_0 is the minimum value. α is the shape parameter.

Connecting the dots (2/3)

Cumulative Distribution Function

$$1 - F(x) = \left(\frac{x_0}{x}\right)^{\alpha} \quad \text{where } x_0 > 0, \alpha > 1$$

From Atkinson, Piketty, and Saez (2011):

- Constant relation between a value x and the average above it.

$$\frac{x^*(x)}{x} = \beta, \quad \text{where } \beta = \frac{\alpha}{\alpha - 1}$$

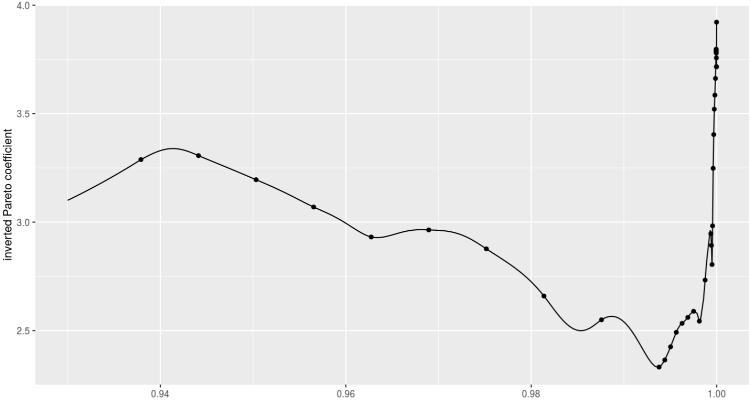
- β is an intuitive alternative to α.
- If $\beta = 2$, the average income above \$100,000 is \$200,000.

Connecting the dots (3/3)

Generalized Pareto interpolation (Blanchet, Fournier, and Piketty, 2022)

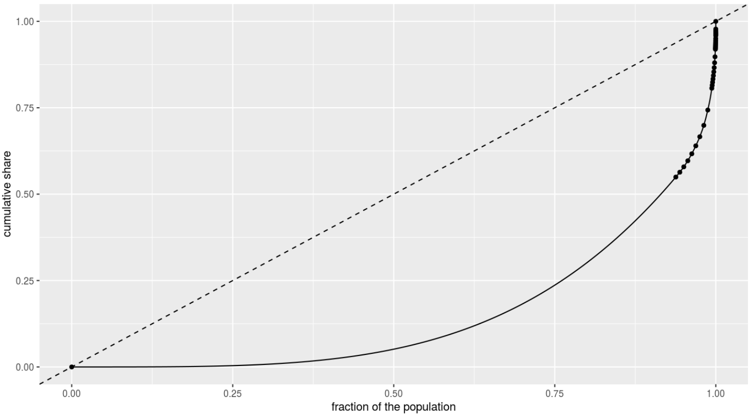
- Atkinson (2017) and Jenkins (2017) note potential inaccuracies with simple interpolations, especially with the lower bound selection.
- Key feature: Adjusts the β coefficient along the distribution.
- Captures heterogeneity more effectively.
- Tested with microdata from the US and France (1962-2014).
- Reduces error rate, especially at the top.

U-shaped Beta coefficients - Brazil 2000



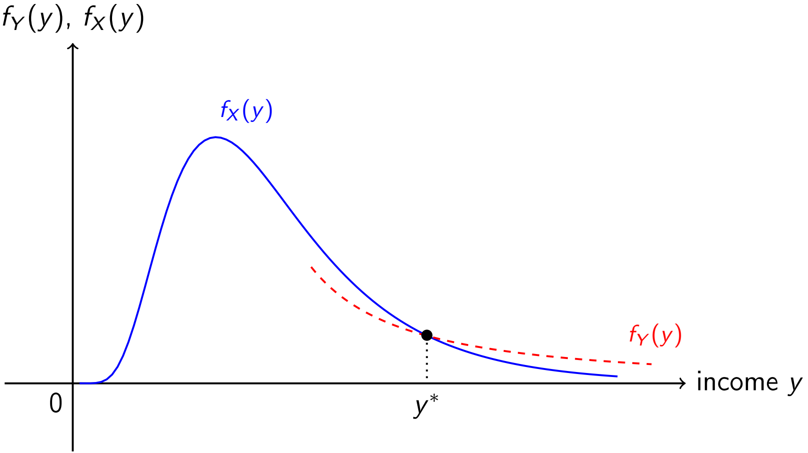
- **Notes.** A higher Beta coefficient implies higher inequality within a given bracket. When data are sufficiently detailed, the coefficients often increase sharply at the top

Interpolated Lorenz curve - Brazil 2000



- **Notes.** The top section of the Lorenz curve (≈top 10%) was successfully, by overlapping sections of Pareto distributions between actual observations of cumulative densities. Extrapolations are also possible, yet they rest on more acrobatic assumptions

Comparing sources: a stylized view



- **Notes.** The blue solid line represents the distribution in a household survey, theoretically covering the whole distribution (log-normal). The red dashed line represents the top tail of the distribution, as captured by income tax records (Pareto). Source: Blanchet, Flores, and Morgan, 2022

Understanding the differences: a formal approach

Non-Response

- $f_Z(y) = f_Y(y)\theta(y) \quad \Rightarrow \quad \theta(y) = f_Z(y)/f_Y(y)$

Misreporting

- $f_Z(y) = f_Y(y)(1 - p(y)) + f_M(y)\bar{p}$

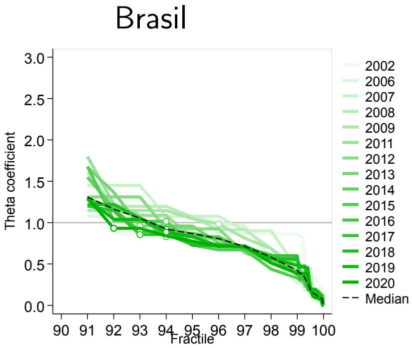
Non-Response & Misreporting

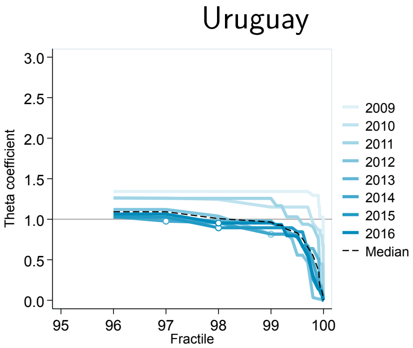
- $f_Z(y)/f_Y(y) = \theta(y)(1 - p(y)) + f_M(y)/f_Y(y)\theta(y)\bar{p}$

Conclusions

- If both biases exist, they are indistinguishable *ex-post*.
- Misreporting only can be solved by matching declarations at the individual level

The Θ coefficient





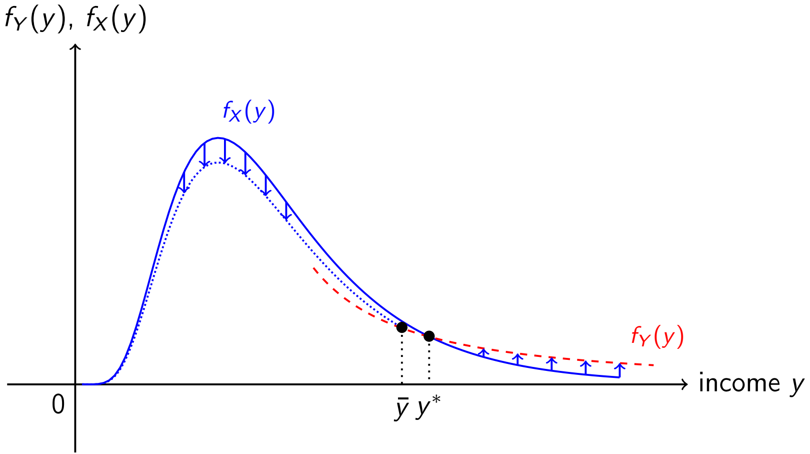
- **Notes.** θ coefficients measure the discrepancy of densities between surveys and a reference distribution. The underestimation of richer populations seems to aggravate with time (more examples in the paper). Source: De Rosa, Flores, and Morgan, 2022

The BFM approach: theoretical requirements

Two assumptions

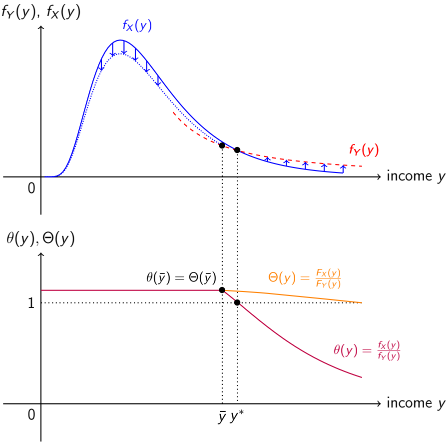
- Income tax declarations are a lower bound of frequencies
- We can trust self declared items in the survey

The BFM approach: Intuition (1/2)



- **Notes.** The adjustment consists in adjusting the weights of observations with income $> \bar{y}$ to fit the red dashed line. Observations below that point will be lowered proportionally to keep the total population unchanged. Source: Blanchet, Flores, and Morgan, 2022

The BFM approach: Intuition (2/2)



- **Notes.** The merging point is defined in $\bar{y}$, not in y^*, to preserve the continuity of denisity functions. It is defined as the point where θ and Θ meet. Source: Blanchet, Flores, and Morgan, 2022

Survey calibration

- Traditional methodology (Deville and Särndal, 1992)
- Survey variable $x_1, \ldots, x_n$, with *design weights* $d_1, \ldots, d_n$
- We need a new set of weights $w_1, \ldots, w_n$ consistent with totals from external sources:

$$\sum_{k=i}^{n} w_i x_x = T$$

- Condition: minimize the distance between original and adjusted weights:

$$\min_{w_1,\ldots,w_n} \sum_{i=1}^{n} \frac{(w_i - d_i)^2}{d_i}$$

$\Rightarrow$ The solution is interpreted as a model of *non-response*.

Alternative methods

- **Reweighting**
  - ⇒ Korinek, Mistiaen, and Ravallion (2006); Hlasny and Verme (2017; 2018); Alvaredo (2011, for Argentina)...
- **Replacing**
  - ⇒ Burkhauser et al. (2016); Piketty, Yang, and Zucman (2017); Chancel and Piketty (2017); Czajka (2017); DWP (2015); Alvaredo (2011, for U.S.); Burkhauser et al. (2018); Jenkins (2017)...
- **Hybrids**
  - ⇒ Bourguignon (2018); Medeiros et al. (2018)
- **Optimal transport**
  - ⇒ Each observation from database A is matched one-to-one with an observation in another one, while minimizing the sum of distances over overlapping variables between matched observations. (Blanchet, Saez, and Zucman, 2022)

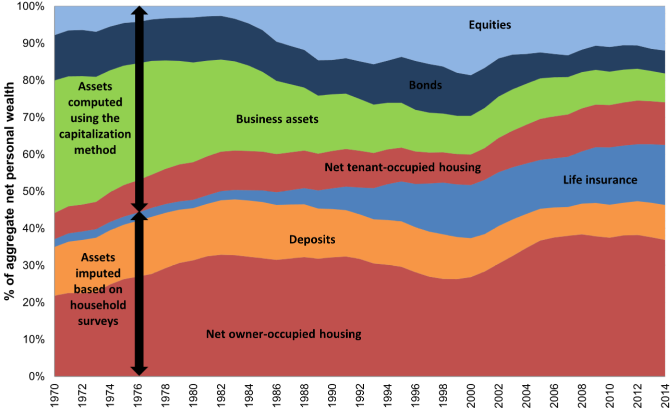
Measuring national wealth

The balance sheet

	Gross personal wealth	4689
AN, S14	Non-financial assets owned by households	1429
	Housing assets of households	*1174*
AN111, S14	*Dwelings owned by households*	*681*
AN21111, S14	*Land underlying dwellings owned by households*	*493*
	Business and other non-financial assets of households	*255*
AN2112, S14	*Agricultural land of households*	*23*
	Other domestic capital of households	*232*
AF, S14	Financial assets owned by households	3260
AF2+AF3+AF4+AF7+AF8, S14	*Currency, deposits, bonds and loans of households*	*1120*
AF5, S14	*Equity and investment fund shares of households*	*1749*
AF6, S14	*Life insurance and pension funds of households*	*391*
AF, S14	Minus: Liabilities of households	189
	Equals: Net personal wealth	**4500**

- **Notes.** As happens for income, wealth aggregates are also recorded in the national accounting system, this time as a stock instead of a flow. Data is much scarcer in developing countries, especially for non-financial wealth

Wealth composition in France 1970-2014



- **Notes.** Modern national accounting systems report long series with detailed compositions. In France, we can observe the surge of financialization since the 1980s. Equity, and life insurance increased as a share of personal wealth with respect to business assets. Land prices explain most of the surge of housing value. Source: Garbinti, Goupille-Lebret, and Piketty, 2021

Wealth distribution: direct methods

Wealth taxes

- Wealth taxes generate administrative data on individual wealth holding (Norway, Spain, Switzerland)
- Only a few countries have them, and most of those who have them have started dismantling them (France).
- The taxable population might be limited

Wealth surveys

- Subject to the same limitations than income surveys (tend to underestimate financial wealth), but they are less common
- In theory, they cover the whole population

Wealth distribution: indirect methods

Estate multiplier

- Uses inheritance data to study the wealth of the deceased
- Assumes dead population of a given year is a sample of the living. Weights are adjusted to represent different probabilities of dying, given socio-economic factors (i.e., age, gender).
- Might be sensitive to assumptions on randomness of mortality rates (conditional on demographics) and extreme cases

Capitalization method

- Uses capital income reported and information on asset-specific returns to compute the underlying asset behind the flow.
- Biased if taxable rates of return vary with wealth