A Theory of Pareto Distributions

NBER EF&G Meeting

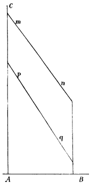
François Geerolf
UCLA

October 21, 2016

Pareto distributions

Schedule D — Année 1893-94.

x £	N: GREAT BRITAIN	N: IRELAND
150	400 648	17 717
200	234 185	9 365
300	121 996	4 592
400	74 041	2 684
500	54 419	1 898
600	42 072	1 428
700	34 269	1 104
800	29 311	940
900	25 033	771
1000	22 896	684
2000	9 880	271
3000	6 069	142
4000	4 161	88
5000	3 081	68
10000	1 104	22



- 1890s, tax tabulations: Pareto plots N of people with incomes $\geq x$:

$$\log N = \log A - \alpha \log x.$$

- Same α: England, Ireland, Prussia, Saxe, and Peru.

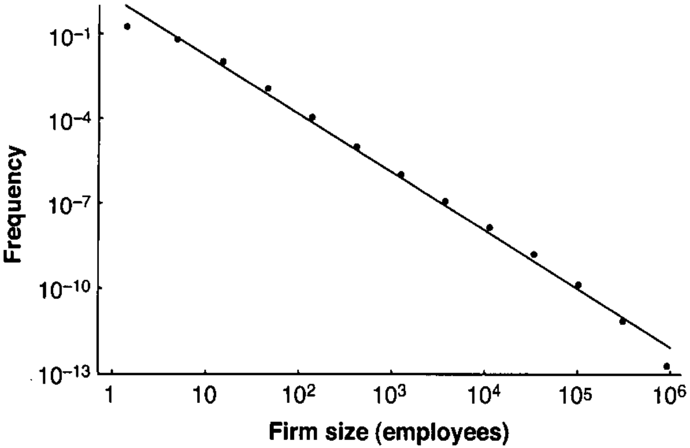
- With Pareto:
  - High Heterogeneity (unbounded).
  - No scale. US: $y_{50} = \$51,939 < y_{av} = \$72,641$.
  - Long tails. 5σ, 10σ draws are very frequent. Top 1% gets $\approx$ 20% of pre-tax income.

- Pareto $\neq$ bell-shaped curve. Few empirical regularities in economics.

Zipf's law for firm sizes

Distribution of US firm sizes. Source: Axtell (2001).

Slope: 2.059 (density) $\Rightarrow$ Tail coeff: $\boxed{\mathbf{1.059}}$. "**Zipf's law**".



Theories of Pareto distributions in Economics

Why Pareto? May reflect some fundamental economic principle:

1. **Pareto distributed primitives.** Explain one Pareto with another Pareto.
   - Lucas (1978), Chaney (2008), Gabaix, Landier (2008), etc.
2. **Paretos from random growth models.**
   - Champernowne (1953), Simon, Bonini (1958), Kesten (1973), Gabaix (1999), Benhabib, Bisin, Zhu (2011), Gabaix, Lasry, Lions, Moll (2016), etc.
3. **New from this paper: Paretos from production functions.** Assignment models with positive sorting, with a special form of production function.
   - Presentation: Garicano (2000) model.
   - Property of the production function, not of specific microfoundations.
   - Another example: Geerolf (2015).

This paper

- Production function derives from **a particular version** of **Garicano (2000)**. Under limited assumptions on the skill distribution:
  - **L layers of hierarchy** = **Pareto tail** for span of control with coefficient:

    $$\boxed{\alpha_L = 1 + \frac{1}{L-1}}, \qquad \boxed{\alpha_2 = 2}, \qquad \boxed{\alpha_{+\infty} = 1}.$$

    $\Rightarrow$ a new theory of **Zipf's law for firm sizes**.
  - **Pareto tail for labor incomes**, with $\beta_L \in [1, +\infty]$, when top skills are scarce enough.
- Data supports these predictions: French matched employer-employee / known US data.
- Taking competitive assignment models to the extreme, where wages are a convex function of skills. (Sattinger (1975)) Here: **wages are Pareto with a bounded support for skills.**

Environment: A Garicano (2000) Economy

- Agents: continuum, measure 1. 1 unit of time.
- 1 good. 1 unit of time $\rightarrow$ 1 good.
- Agents: different exogenous **skills**. Agent with skill x can solve "problems" in $[0, x]$.
- Distribution of skills x: c.d.f. $F(.)$, density $f(.)$ on $[1 - \Delta, 1]$.
  $\boxed{\Delta\textbf{: Heterogeneity in Skills.}}$
  $\boxed{F(.)\textbf{: Skill Distribution.}}$
- Workers encounter **problems** in production. Draw a unit continuum of different problems on $[0, 1]$ in c.d.f. $G(.)$, uniform w.l.o.g. :
  - <u>When they know the solution</u>: produce 1 unit of the good.
  - <u>When they don't</u>: can ask someone else for a solution.
    $h < 1$: manager's time cost to listen to one problem.
    $\boxed{h\textbf{: Helping Time.}}$

Imposing 2 layers

- **Planner's problem.** Planner maximizes total output.
- Occupational cutoff: z_2 splits managers (high x) and workers (low x).
- Workers x fail to solve $1 - x$ problems. Time supervising worker x: $h(1-x)$. Span of control of a manager hiring workers with skill x:

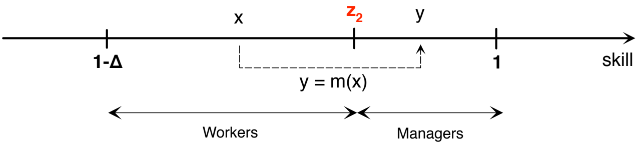
$$\boxed{n = \frac{1}{h(1-x)}}$$

- Output $Q(x, y)$ jointly produced by manager with skill y hiring workers with skill x:

$$Q(x,y) = \frac{y}{h(1-x)} \quad \Rightarrow \quad \frac{\partial^2 Q(x,y)}{\partial x \partial y} = \frac{1}{h(1-x)^2} > 0$$

- Complementarities $\Rightarrow$ **Positive sorting**. $y = m(x)$, $m'(x) > 0$.

Uniform distribution



- $m(.)$ ensures market clearing for time:

  $$f(y)dy = h(1-x)f(x)dx \quad \Rightarrow \quad f(m(x))m'(x) = h(1-x)f(x).$$

- z_2, $m(.)$ unknowns. Boundary value problem:

  $$m(1-\Delta) = z_2, \qquad m(z_2) = 1.$$

- Assume for a moment that $f(x) = 1/\Delta$ on $[1-\Delta, 1]$. Then 1-x is a uniform distribution on $[1-z_2, \Delta]$. What is the distribution of span of control:

  $$n(y) = \frac{1}{h(1-x)}.$$

Mathematical Result: Inverse of a Uniform on $[\Delta^2, \Delta]$

Lemma

If $\mathbf{U} \sim$ *Uniform* $([\Delta^2, \Delta])$, *then*
$\mathbf{1/U} \sim$ *Truncated Pareto* $(1, 1/\Delta, 1/\Delta^2)$.

- Assume $f_U(u) = 1/(\Delta - \Delta^2)$ on $[\Delta^2, \Delta]$. The "tail function" (complementary c.d.f) of $1/U$ is:

$$\begin{aligned}
\bar{F}_{1/U}(x) \equiv 1 - F_{1/\mathrm{U}}(x) &= \mathbb{P}\left[\frac{1}{U} \geq x\right] \\
&= \mathbb{P}\left[U \leq \frac{1}{x}\right] \\
&= \int_{\Delta^2}^{1/x} f_U(u)du \\
\bar{F}_{1/U}(x) \equiv 1 - F_{1/\mathrm{U}}(x) &= \frac{\frac{1}{x} - \Delta^2}{\Delta - \Delta^2}.
\end{aligned}$$

- Inverse of a Uniform on $[0, \Delta]$ = **full Pareto** with tail coefficient 1.

Mathematical Result 2: Inverse of a Uniform on $[\Delta^2, \Delta]$

- Span of control of manager y hiring workers with skill x:

$$n(y) = \frac{1}{h(1-x)}$$

- If $f(.)$ is uniform, $1-x$ is a uniform distribution over $[1-z_2, \Delta]$.
- I show that:

$$1 - z_2 = \frac{\sqrt{1 + h^2\Delta^2} - 1}{h} \sim_{\Delta \to 0} \frac{h}{2}\Delta^2.$$

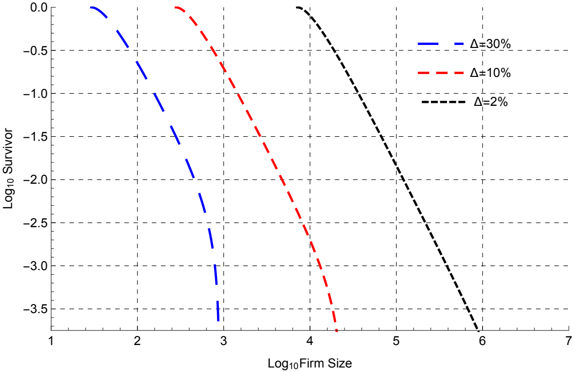
- Thus the **size-biased distribution** is a Truncated Pareto (1).
- Size-biased distribution: a firm with 100 employees is counted 100 times. $\Rightarrow$ Overstating fattailedness.
- **Size-biased distribution** is Truncated Pareto (1) $\Rightarrow$ **distribution** is Truncated Pareto (2).

Non-uniform distribution

- "blowing up" of the denominator $\Rightarrow$ under some regularity conditions on $f(.)$, works also if not uniform.
- If $f_X(0) \neq 0$, then **Pareto tail**:

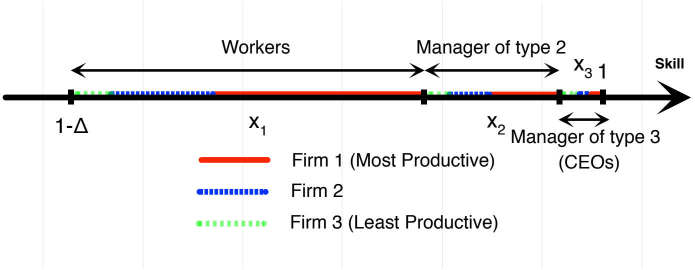
$$1 - F_{1/X}(x) = \int_0^{1/x} f_X(u)du \sim_{+\infty} \frac{f_X(0)}{x}.$$

- Example with a linear increasing density.



Firm with $L = 3$ layers

- Positive Sorting.



- Generalizing $\alpha_2 = 2$ by iteration, the tail exponent:

$$\boxed{\alpha_L = 1 + \frac{1}{L-1}}.$$

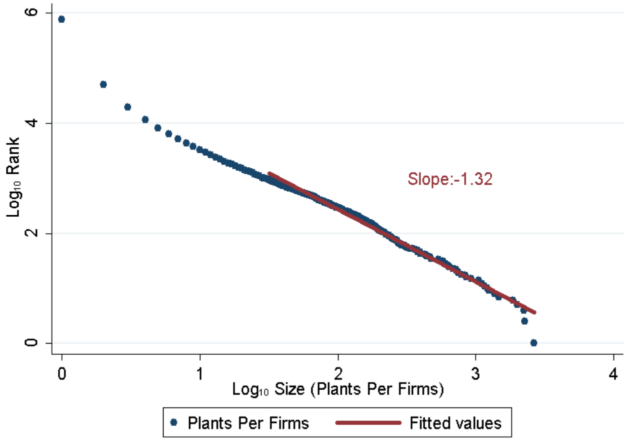
- When $L \to \infty$, **Zipf's law** for firm sizes:

$$\boxed{\alpha_{+\infty} = 1}.$$

- Again, true for any density.

French DADS - establishments per firms

- Empirics: French matched employer-employee data, US data.
- Example: number of establishments per firms. (France)



Assignment equation

- Skill prices $w(.)$ decentralizing optimal allocations:

$$w(y) = \max_x \frac{y - w(x)}{h(1-x)}.$$

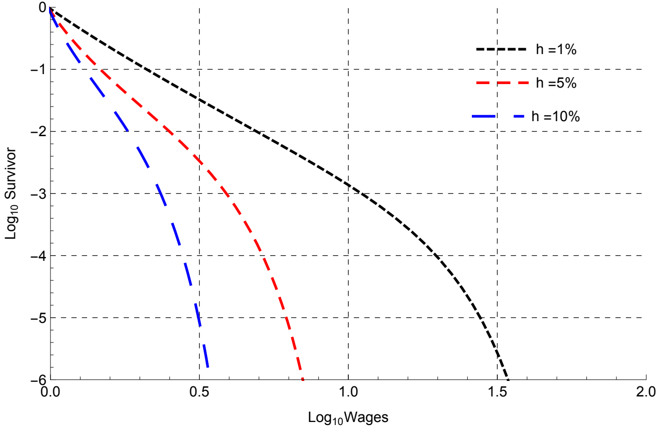
- Envelope condition:

$$w'(y) = \frac{1}{h(1-x)} = n(y) \quad \Rightarrow \quad \boxed{\underbrace{\frac{dw(y(n))}{dn}}_{\Delta \text{Wages}} = \underbrace{n(y(n))}_{\text{Size}} \; \underbrace{y'(n)}_{\Delta \text{Talents}}}.$$

- Comparison:
  - Gabaix, Landier (2008). **Small** differences in talent across managers, **large and Pareto** firm sizes $\Rightarrow$ Large differences in pay.
  - This paper: **Small** differences in talents across **workers and managers** $\Rightarrow$ Large differences in pay. (through endogenous large and Pareto firm sizes)

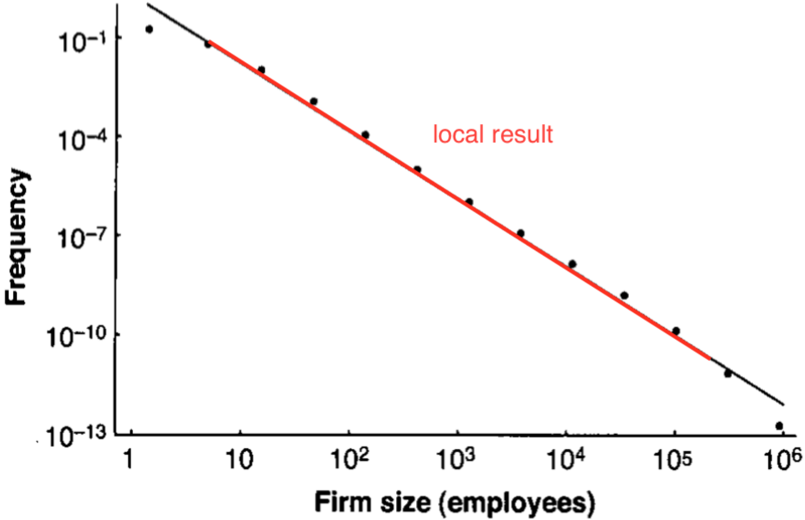
Labor income distribution: effect of a decrease in h (IT?)

- Gabaix, Landier (2008): if skill distribution does not change, Pareto coefficient does not change.
- Not true in this paper when h diminishes (IT?).



- Empirically: $\alpha = -3$ in 1970s to $\alpha = -1.8$ in 2010s.

Conclusion: coming back to Axtell (2001)



Conclusion

- **Main takeaways**:
  - **Maths**:
    - U is Uniform $(0,\Delta) \Rightarrow 1/U$ is Pareto $(1, 1/\Delta)$.
    - X goes through the origin $\Rightarrow 1/X$ has a Pareto tail.
  - Stylized model accounts for **Pareto firm size and labor income distribution**, regardless of the ability distribution.
  - New intuition for why firm sizes and labor incomes are so heterogenous **despite small observable differences**: "power law change of variable near the origin".
  - Endogenous "economics of superstars".
- **Future work**:
  - **Other microfoundations** for power-law production functions.
  - **In applied work**, potential alternative to:
    - <u>Optimal taxation</u>: Pareto distributed skills.
    - <u>Trade</u>: Pareto distributed firm productivities.
    - <u>Misallocation</u>: Pareto distributed manager/firm productivities.