

A Theory of Pareto Distributions

UZH Macroeconomics Seminar

François Geerolf
UCLA

May 3, 2017

Pareto distributions

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

Schedule D — Année 1893-94.

x £	N	
	GREAT BRITAIN	IRELAND
150	400 648	17 717
200	234 485	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

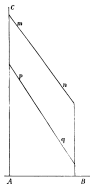
$$\log N_{\text{income} \geq x} = C - \alpha \log x.$$

- ▶ Same log linear relationship, differing $\alpha \in [1, 3]$:
 - ▶ Semifeudal Prussia
 - ▶ Victorian England
 - ▶ Capitalist but highly diversified Italian cities
 - ▶ Communist-like regime of the Jesuits in Peru under Spanish rule, etc.

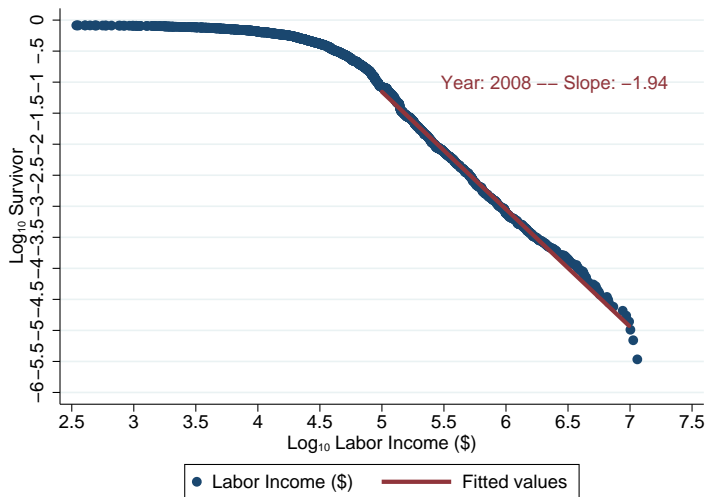
- ▶ With Pareto:

- ▶ No scale. US: $y_{50} = \$51,939 < y_{av} = \$72,641$.
- ▶ Long tails. Top 1% gets $\approx 20\%$ of pre-tax income.
- ▶ Constant elasticity: $d \log N_{\geq x} / d \log x = -\alpha$

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



Pareto tail for US labor incomes, 2008



Source: Statistics of Income, Public Use Sample

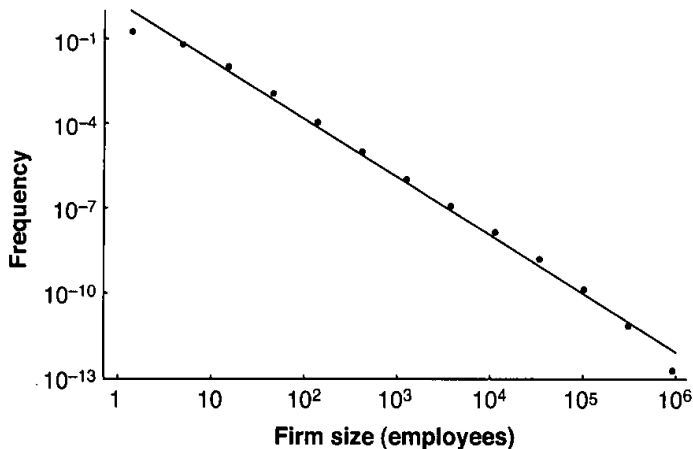
Pareto tail for US labor incomes, 1968



Zipf's law for firm sizes

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

Slope: 2.059 (density) \Rightarrow Tail coeff: **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), Kortum (1997), Melitz / Chaney (2008), Gabaix, Landier (2008), etc.

2. **Pareto from random growth models.**

- ▶ Champernowne (1953), Simon, Bonini (1958), Kesten (1973), Gabaix (1999), Gabaix, Lasry, Lions, Moll (2016), Jones, Kim (2016), etc.

3. **New from this paper: Pareto 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:

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

⇒ 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**.

Literature

- ▶ **Pareto distributions.** Pareto (1896), Zipf (1949).
- ▶ **Competitive assignment models.** Roy (1950), Becker (1973, 1974), Rosen (1981), Sattinger (1975), Kremer (1993), Terviö (2008), Gabaix, Landier (2008).
- ▶ **Span of control.** Lucas (1978), Rosen (1981), Rosen (1982), Rossi-Hansberg, Wright (2007).
- ▶ **Organizational structure.** Calvo, Wellisz (1978,1979), Garicano (2000), Garicano, Rossi-Hansberg (2004, 2006), Antras, Garicano, Rossi-Hansberg (2006), Caliendo, Monte, Rossi-Hansberg (2015).
- ▶ **Literature in Physics.** Sornette (2002), Newman (2005), Sornette (2006).
- ▶ **Random growth.** Champernowne (1953), Simon, Bonini (1958), Kesten (1973), Sutton (1997), Gabaix (1999), Axtell (2001), Luttmer (2007), Gabaix, Lasry, Lions, Moll (2016).

Overview

Environment

Span of control with 2 layers

Span of control with L layers - Zipf's law

Empirics

Labor income distribution

Conclusion

Environment

Span of control with 2 layers

Span of control with L layers - Zipf's law

Empirics

Labor income distribution

Conclusion

A Garicano (2000) Economy 1/2

- ▶ 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. $F(\cdot)$, density $f(\cdot)$ on $[1 - \Delta, 1]$.

Δ : Heterogeneity in Skills.

$F(\cdot)$: Skill Distribution.

- ▶ Workers encounter **problems** in production. Draw a unit continuum of different problems on $[0, 1]$ in c.d.f. $G(\cdot)$, uniform w.l.o.g. :
 - ▶ When they know the solution: produce 1 unit of the good.
 - ▶ When they don't: can ask someone else for a solution.
 $h < 1$: manager's time cost to listen to one problem.
 h : Helping Time.

A Garicano (2000) Economy 2/2

- ▶ Assumption 1: x unknown.
- ▶ Assumption 2: h low enough: always hierarchies.
- ▶ Assumption 3: one manager with time 1 at the top.

Environment

Span of control with 2 layers

Span of control with L layers - Zipf's law

Empirics

Labor income distribution

Conclusion

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 :

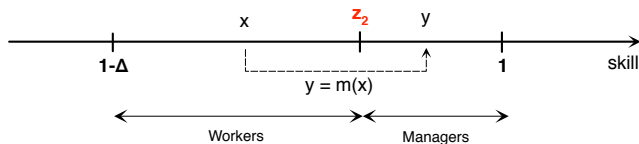
$$n = \frac{1}{h(1 - x)}$$

- ▶ 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



- ▶ $m(\cdot)$ ensures market clearing for time:

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

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

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

- ▶ 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 $\mathbf{U} \sim \text{Uniform}([\Delta^2, \Delta])$, then

$\mathbf{1/U} \sim \text{Truncated Pareto}(1, 1/\Delta, 1/\Delta^2)$.

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

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

- Inverse of a Uniform on $[0, \Delta] = \text{full Pareto}$ 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 $f(\cdot)$ is uniform, $1-x$ is a uniform distribution over $[1-z_2, \Delta]$.
- ▶ I show that:

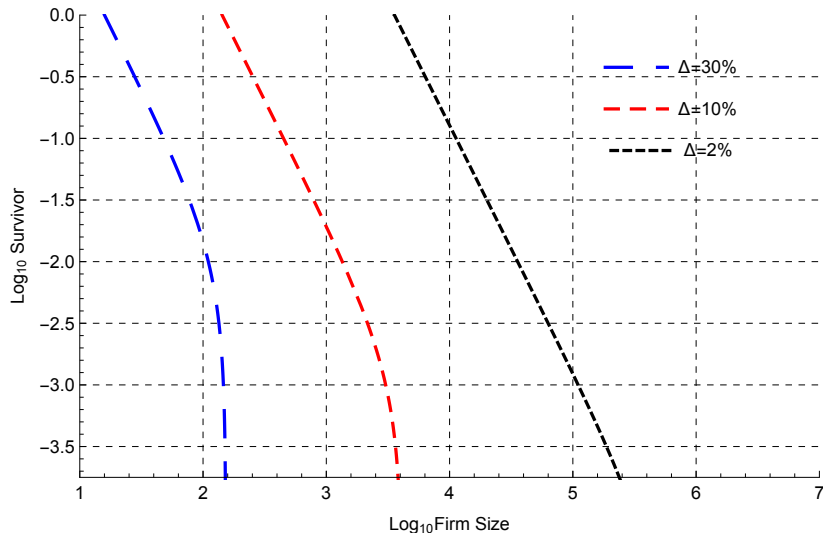
$$1 - z_2 = \frac{\sqrt{1 + h^2 \Delta^2} - 1}{h} \sim_{\Delta \rightarrow 0} \frac{h}{2} \Delta^2.$$

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

$$f_S(x) \sim \frac{1}{x^2} \quad \text{and} \quad f_S(x) \sim x f(x) \quad \Rightarrow \quad f(x) \sim \frac{1}{x^3}.$$

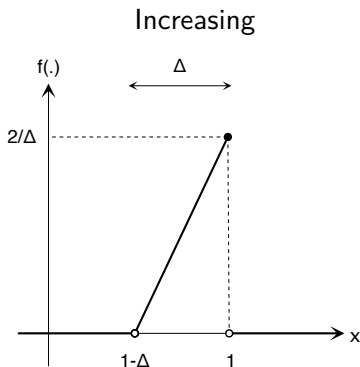
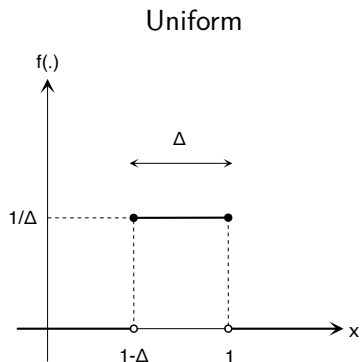
Pareto plot

Example: $h = 70\%$, $f(\cdot)$ uniform with $\Delta = 30\%, 10\%, 2\%$.



Non-uniform distribution

- ▶ What happens if f is not uniform?
- ▶ Example with an increasing distribution.

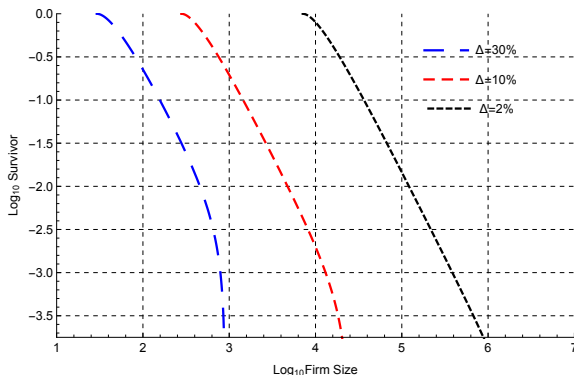


Non-uniform distribution

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

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

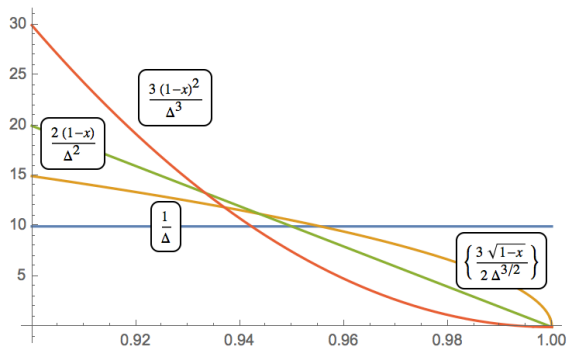
- ▶ Example with a linear increasing density.



Relaxing $f(1) > 0$

- If $f(1) = 0$. Illustration: polynomial functions:

$$f(x) = \frac{\rho + 1}{\Delta^{\rho+1}} (1 - x)^{\rho} \quad \text{if } x \in [1 - \Delta, 1]$$



Relaxing $f(1) > 0$

- ▶ Closed-form for span of control. Truncated Pareto($2 + \rho$):

$$n(y) = \frac{1}{h} \left[\frac{1}{h} \frac{\rho + 2}{\rho + 1} (1 - y)^{\rho+1} + (1 - z_2)^{\rho+2} \right]^{-\frac{1}{\rho+2}}.$$

- ▶ Do not appear in the upper tail, as smaller however.
Maximum size \bar{n} is such that:

$$\frac{\bar{n}(\rho > 0)}{\bar{n}(\rho = 0)} = \Delta^{\frac{\rho}{\rho+1}} \rightarrow_{\Delta \rightarrow 0} 0.$$

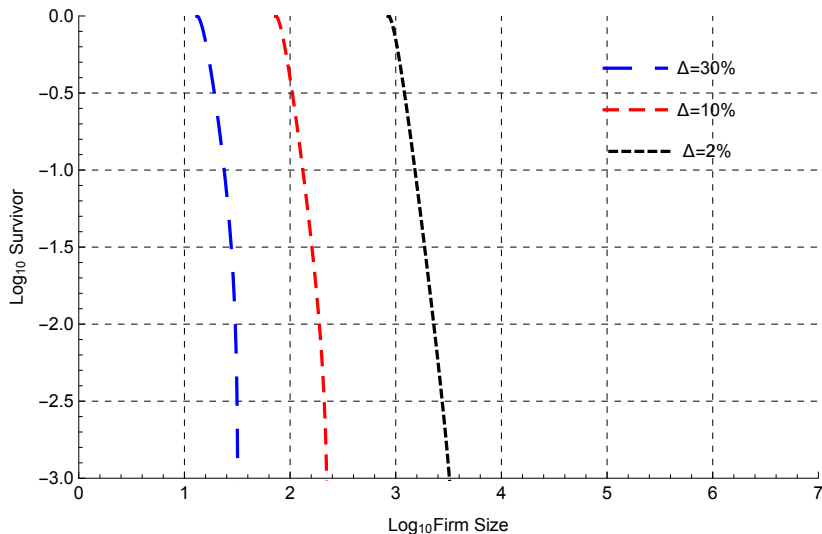
- ▶ If $f(1) = 0$, sufficiently regular, i.e. Taylor with $\rho < +\infty$:

$$f(x) = A(1 - x)^\rho + O((1 - x)^{\rho+1}), \quad \text{with } A > 0.$$

then similarly, weak form of truncated Pareto($2 + \rho$). Also smaller.

Relaxing $f(1) > 0$ - example: Beta(1,1)

- Higher tail coefficients, but smaller firms which do not appear in the upper tail.



Environment

Span of control with 2 layers

Span of control with L layers - Zipf's law

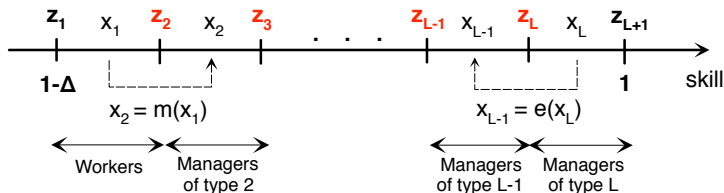
Empirics

Labor income distribution

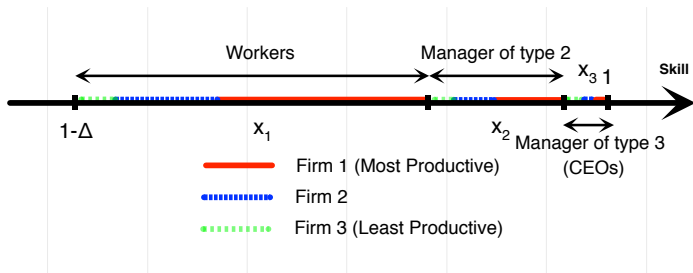
Conclusion

Occupational choice, given L

- In equilibrium, agents split into L types according to their skills in $[1 - \Delta, 1] = [z_1, z_{L+1}]$, and form a hierarchical organization:



Firm with $L = 3$ layers



- ▶ Positive Sorting.
- ▶ Span of control of manager of type 2 x_2 (same):

$$n_{2 \rightarrow 1}(x_2) = \frac{1}{h(1-x_1)} \Rightarrow f(x_2)dx_2 = h(1-x_1)f(x_1)dx_1.$$

- ▶ **Intermediary** Span of control of manager of type 3 x_3 :

$$n_{3 \rightarrow 2}(x_3) = \frac{1}{h \frac{1-x_2}{1-x_1}} \Rightarrow f(x_3)dx_3 = h \frac{1-x_2}{1-x_1} f(x_2)dx_2.$$

Firm with $L = 3$ layers

- ▶ **Total** span of control of manager of type 3 x_3 :

$$n_{3 \rightarrow 1}(x_3) = n_{3 \rightarrow 2}(x_3)n_{2 \rightarrow 1}(x_2) = \frac{1}{h^2(1-x_2)}.$$

- ▶ Previously we had:

$$f(x_2)dx_2 = h(1-x_1)f(x_1)dx_1 \quad \Rightarrow \quad 1-x_2 \sim (1-x_1)^2.$$

- ▶ Now we have:

$$f(x_3)dx_3 = h \frac{1-x_2}{1-x_1} f(x_2)dx_2 \quad \Rightarrow \quad 1-x_3 \sim (1-x_2)^{3/2}.$$

- ▶ Intuitively, exponent on the matching function gives the tail index of the Pareto, thus:

$$\alpha_3 = \frac{3}{2}.$$

Any L

- ▶ First layer always special with:

$$m'(x_1)f(m(x_1)) = h(1 - x_1)f(x_1).$$

- ▶ Subsequent layers $l \in [2, \dots, L - 1]$ with conditional probability:

$$m'(x_l)f(m(x_l)) = h \frac{1 - x_l}{1 - m^{-1}(x_l)} f(x_l).$$

- ▶ Matching the more skilled and less skilled:
 - ▶ $L - 1$ initial conditions.
 - ▶ $L - 1$ equations for occupational cutoffs.
- ▶ Equilibrium number of layers: fixed cost, or indivisibility with a discrete number N of agents:

$$L = \max_L \left\{ L \quad \text{s.t.} \quad 1 - z_L \geq \frac{1}{N} \right\}.$$

Zipf's law for firm sizes

- ▶ Total span of control $n(x_L) \equiv n_{L \rightarrow 1}(x_L)$ is given by:

$$n(x_L) = n_{L \rightarrow L-1}(x_L) * n_{L-1 \rightarrow L-2}(x_{L-1}) * \dots * n_{2 \rightarrow 1}(x_2)$$

$$n(x_L) = \prod_{l=1}^{L-1} n_{l+1 \rightarrow l}(x_l).$$

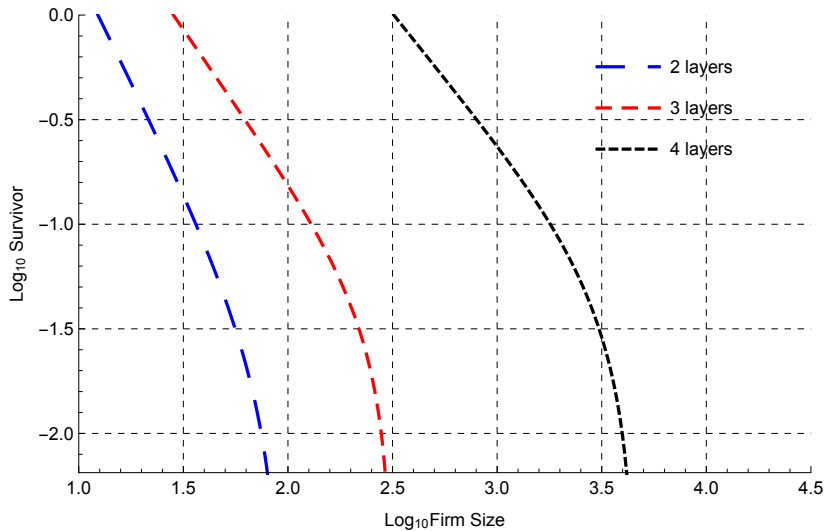
- ▶ Generalizing $\alpha_2 = 2$ and $\alpha_3 = 3/2$ by iteration, the tail exponent for $n(x_L)$ is:

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

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

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

Many layers



Environment

Span of control with 2 layers

Span of control with L layers - Zipf's law

Empirics

Labor income distribution

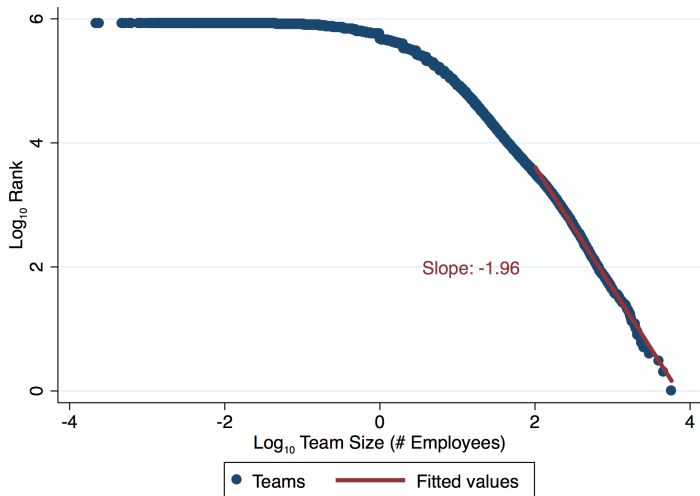
Conclusion

Higher level of disaggregation

Using French matched employer-employee data, and Caliendo, Monte, Rossi-Hansberg's (JPE, 2015) methodology.

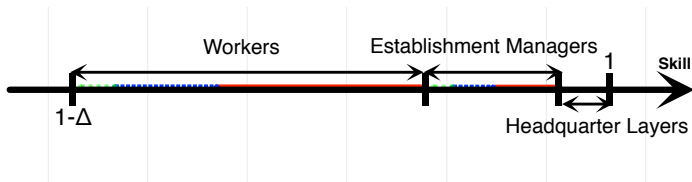
- ▶ Use of "PCS-ESE": Profession Catégorie Socioprofessionnelle.
- ▶ First Digit Corresponds to one of 6 categories:
 1. **Farmers**
 2. **Self-employed / Owners:** *Plumbers, film directors, CEOs.*
 3. **Senior staff or top management positions:** *CFOs, heads of HRs, purchasing managers.*
 4. **Employees at the supervisor level:** *Quality control technicians, sales supervisors.*
 5. **Clerical, white-collar employees:** *Secretaries, HR or accounting, sales employees.*
 6. **Blue-collar workers:** *Assemblers, machine operators, maintenance workers.*
- ▶ Form "teams" in establishments, dividing the # of employees in a layer by the # of employees in the layer above.

French DADS - Distribution of "teams"

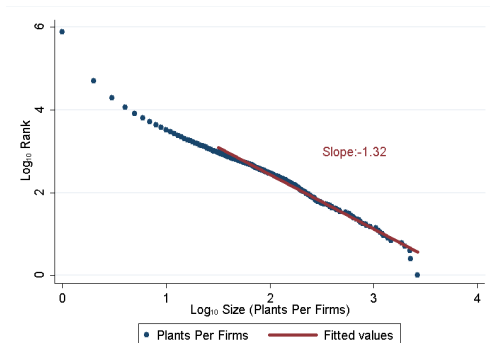


- Data lends support to Zipf's law as **compounding of elementary Pareto (2) \neq Random Growth** .

French DADS - establishments per firms

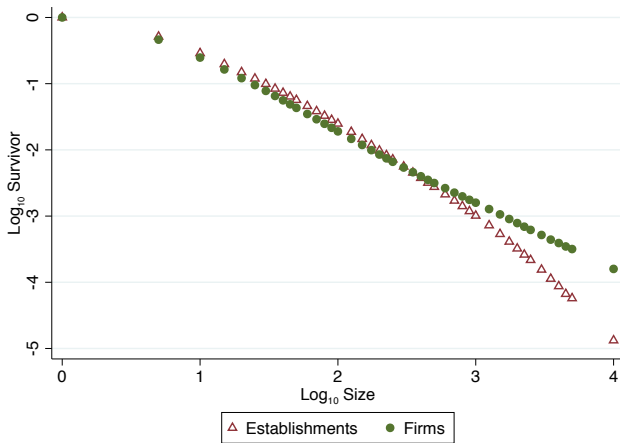


- Pareto on most of the range consistent with the model: the uniform distribution = better approximation locally.



Distribution of US firms and establishments. Source: Census bureau.

- Equivalent for the US? Establishment Level.



Firms: 1.01

Establishments: 1.33

Environment

Span of control with 2 layers

Span of control with L layers - Zipf's law

Empirics

Labor income distribution

Conclusion

Assignment equation

- ▶ Skill prices $w(\cdot)$ 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)

Integrating truncated Pareto distributions

- ▶ Slight difference: Zipf's law is truncated \Rightarrow hypergeometric functions instead of exact Pareto distributions.
- ▶ Example:

$$f(x) = \begin{cases} A_1 & \text{if } x \in [1 - \Delta_1 - \Delta_2, 1 - \Delta_2] \\ A_2(\rho + 1)(1 - x)^\rho & \text{if } x \in [1 - \Delta_2, 1] \end{cases}$$

- ▶ Comparative statics shown 2 layer case, where this is an hypergeometric function:

$$w(y) = w(z_2) + \int_{z_2}^y \frac{du}{h \sqrt{(1 - z_2)^2 + \frac{2}{h} \frac{A_2}{A_1} (1 - u)^{\rho+1}}}.$$

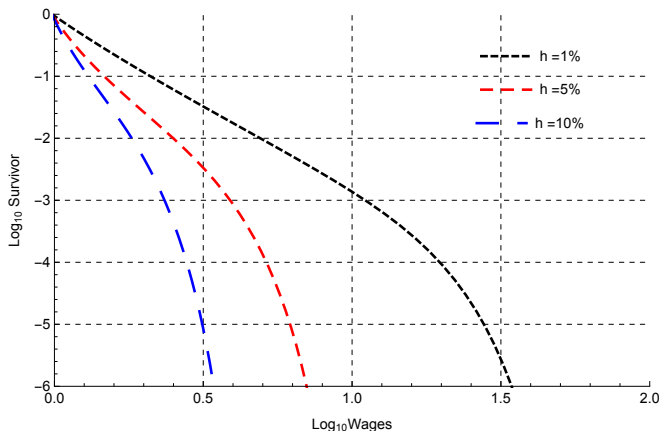
Reduced form VS full model

Apart from positive aspects is a reduced form approach sufficient?
Not always.

- ▶ **Calculate all wages.** "Trickle-down" effects.
- ▶ **Relate change in firm sizes to deep parameters.** Here h and Δ shift the distribution out.
- ▶ And: truncation is key for comparative statics of the Pareto distribution:
 - ▶ Gabaix and Landier (2008) attribute the 5x increase in CEO compensation to a 5-fold in the scale. h or Δ .
 - ▶ Difficulty: $\alpha = -3$ in 1970s to $\alpha = -1.8$ now. In Gabaix and Landier (2008), α is constant.

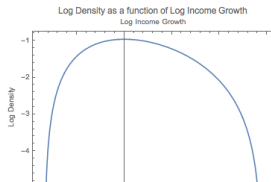
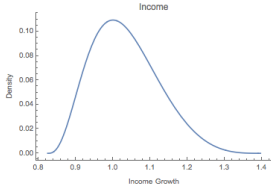
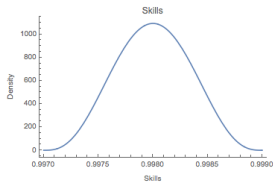
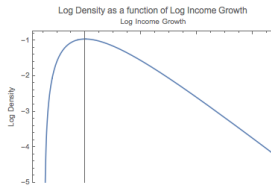
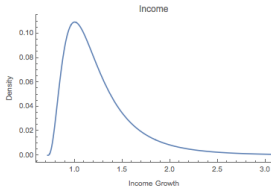
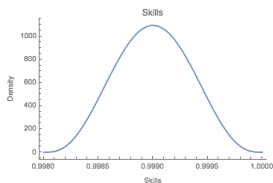
Labor income distribution: effect of a decrease in h (IT?)

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

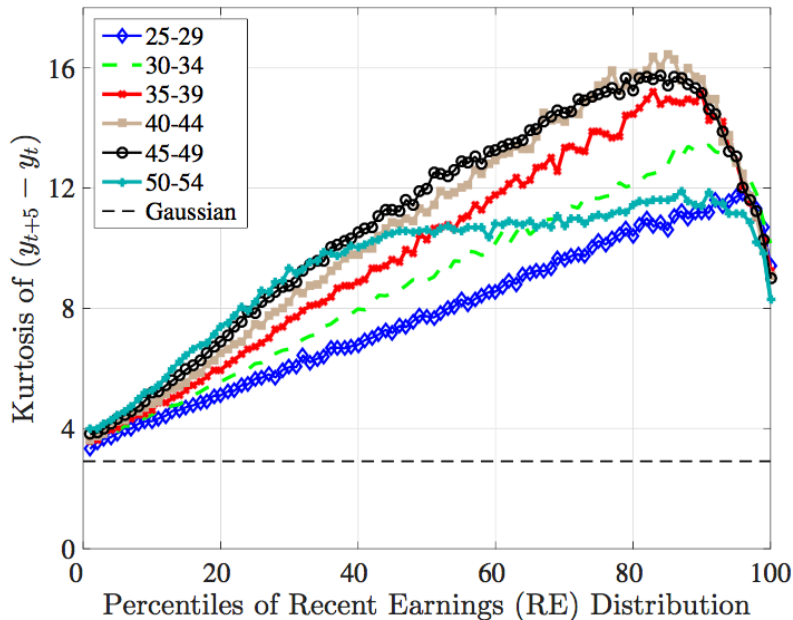


Telling theories of income apart: Dynamics

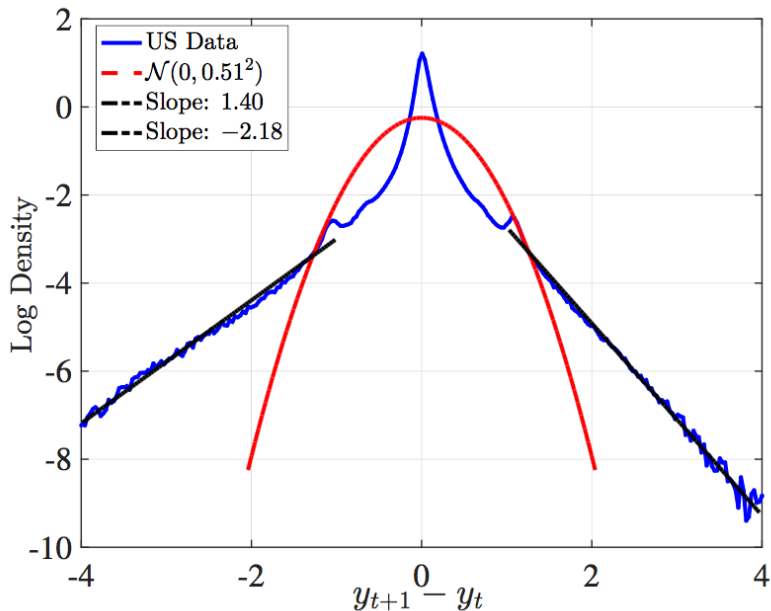
1. Pareto distributed primitives (Lucas (1978), Mirrlees (1971)).
But Francis Galton: mental abilities normally distributed.
2. Paretos from random growth models. (Gabaix, Lasry, Lions, Moll (2016))
3. Paretos from production functions. Small shocks to y .
Non-linear mapping $w(y)$.



Guvenen, Karahan, Ozkan, Song (2016)



Guvenen, Karahan, Ozkan, Song (2016)



Environment

Span of control with 2 layers

Span of control with L layers - Zipf's law

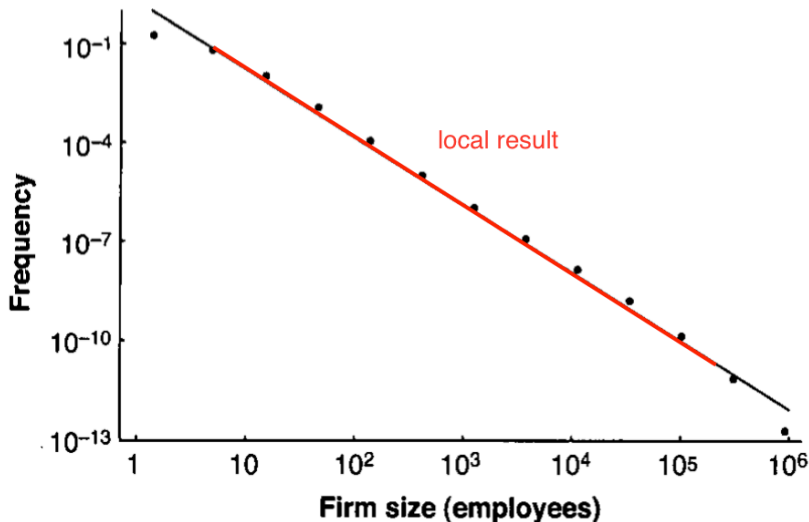
Empirics

Labor income distribution

Conclusion

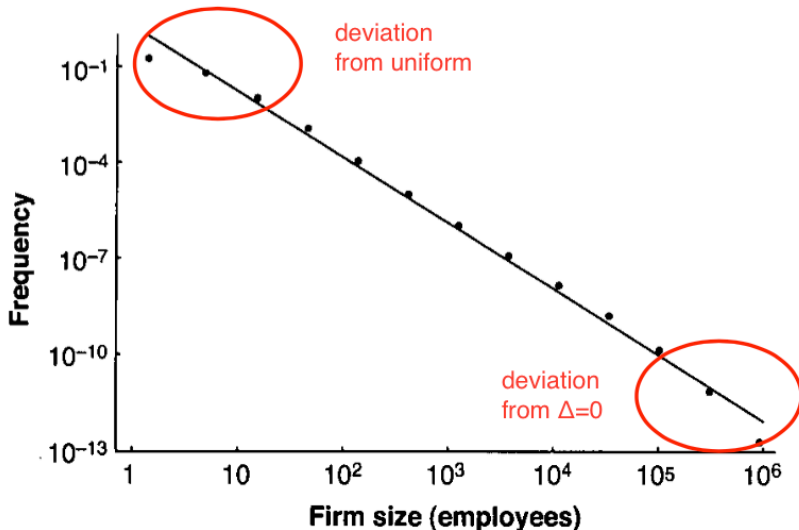
Conclusion: coming back to Axtell (2001)

What does not matter for heterogeneity under a power law production function



Conclusion: coming back to Axtell (2001)

What matters for heterogeneity under a power law production function



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 heterogeneous **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:
 - Optimal taxation: Pareto distributed skills.
 - Trade: Pareto distributed firm productivities.
 - Misallocation: Pareto distributed manager/firm productivities.