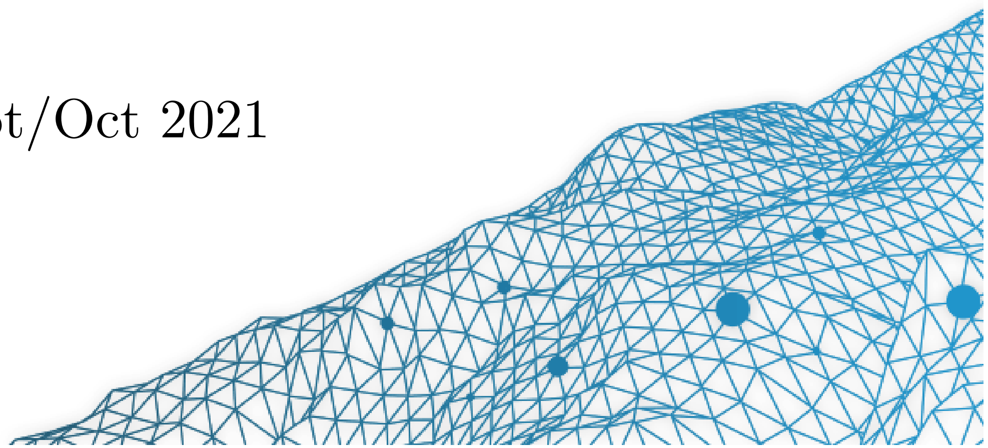
Extended Pareto distribution and applications

Arthur Charpentier, Université du Québec à Montréal

(joint work with Emmanuel Flachaire, AMSE)

Emeritaat Jan Beirlant, KU Leuven, Sept/Oct 2021



Emeritaat Jan Beirlant

(Source: KU Leuven - Rob Stevens)

Regularly Varying Functions (part 1/3)

A measurable function $g : \mathbb{R}_+ \to \mathbb{R}_+$ is regularly varying at infinity if there exists a function h such that, for all $x > 0$,

$$\lim_{t\to\infty} \frac{g(tx)}{g(t)} = h(x)$$

Then necessarily $h(x) = x^\beta$ for some $\beta \in \mathbb{R}$, and we write $g \in RV_\beta$.

A measurable function $\mathcal{L} : \mathbb{R}_+ \to \mathbb{R}_+$ is slowly varying at infinity if, for all $x > 0$,

$$\lim_{t\to\infty} \frac{\mathcal{L}(tx)}{\mathcal{L}(t)} = 1$$

Any regularly varying function with index β can be written $g(x) = x^\beta \mathcal{L}(x)$

The Pareto survival function $\overline{F}$ is regularly varying with index $-\alpha$, $\overline{F}(x) = Ax^{-\alpha}$.

Extended Regular Variation (part 2/3)

One can prove (see de Haan & Ferreira (2010)) that regular variation is obtained if there is a positive function a such that

$$\lim_{t\to\infty}\frac{g(tx)}{a(t)}$$

exists, for all $x>0$.

A extension (extended regular variation) is obtained if

$$\lim_{t\to\infty}\frac{g(tx)-g(t)}{a(t)}$$

exists, and is denoted h. In that case,

$$\lim_{t\to\infty}\frac{g(tx)-g(t)}{a(t)}=c\frac{x^\gamma-1}{\gamma}$$

and we write $g\in ERV_\gamma$. If $\gamma>0$, f is necessarily regularly varying with tail index γ (theorem 2.B.2).

Second-Order Regularly Varying Functions (part 3/3)

Assume that there is a function b such that

$$\lim_{t\to\infty}\frac{1}{b(t)}\left[\frac{g(tx)-g(t)}{a(t)}-\frac{x^\gamma-1}{\gamma}\right]$$

exits (and is denoted h). This is second order regular variation.

de Haan & Stadtmüller (1996) obtained a general expression for h, related to some index ρ. In a nutshell, following Drees (1998) and Cheng & Jiang (2001), the limit can be

$$\lim_{t\to\infty}\frac{1}{b(t)}\left[\frac{g(tx)-g(t)}{a(t)}-\frac{x^\gamma-1}{\gamma}\right]=\frac{x^{\gamma+\rho}-1}{\gamma+\rho}$$

with $\rho<0$ (theorem B.3.10).

The Extended Pareto Distribution $\mathcal{EPD}(u,\delta,\tau,\alpha)$

For the Pareto distribution (with tail index $-\gamma$),

$$\lim_{t\to\infty} x^{-\gamma}\frac{\overline{F}(tx)}{\overline{F}(t)}=1$$

Consider the following extension

$$\lim_{t\to\infty} x^{-\gamma}\frac{\overline{F}(tx)}{\overline{F}(t)}=1+\frac{x^\rho-1}{\rho},\text{ for some }\rho\leq 0$$

or, up to some affine transformation, $\overline{F}(x)=cx^{-\gamma}[1+x^\rho\ell(x)]$

Since $(1+u)^b\sim(1+bu)$, define (changing γ in α, ρ in τ)

$$\overline{F}(x)=\mathbb{P}[X>x]=\left[\frac{x}{u}\left(1+\delta-\delta\left(\frac{x}{u}\right)^\tau\right)\right]^{-\alpha},\text{ for }x\geq u$$

where $\tau\leq 0$ and $\delta>\max(-1,1/\tau)$. Write $X\sim\mathcal{EPD}(u,\delta,\tau,\alpha)$,

as in Beirlant, Joossens & Segers (2009)

Inequality Measures and Income Distributions

The top p-% income share can be defined as follows,

$$\text{TS}_p = \frac{p\,\mathbb{E}(X|X > Q(1-p))}{\mathbb{E}(X)}$$

where $Q(t)$ is the quantile function, $Q(t) = \inf\{x \in \mathbb{R} : t \leq F(x)\}$.

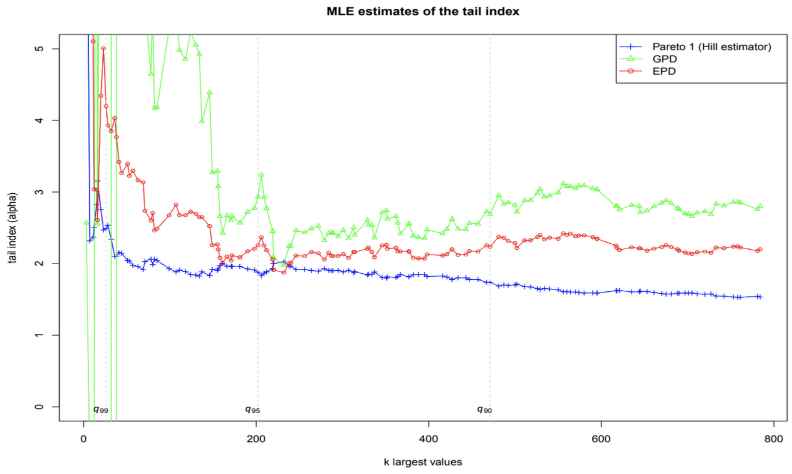
The top p-% income share from a distribution being Pareto $\mathcal{EPD}$ in the tail, that is, on the top q-% of the distribution, is

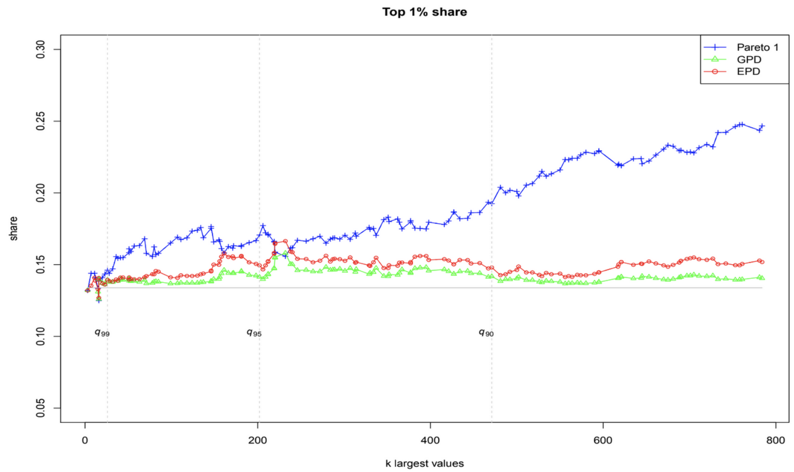
$$\text{TS}_{p,q}^{(\mathcal{EPD})} = \begin{cases} \dfrac{pu' + qE_{u'}}{(1-q)\bar{x}_q + q(u + E_u)} & \text{if } p \leq q \\ 1 - \dfrac{(1-p)\,\bar{x}_p}{(1-q)\bar{x}_q + q(u + E_u)} & \text{if } p > q \end{cases}$$

where $E_{u'} = \displaystyle\int_{u'}^{\infty} \overline{F}(x)dx = \int_0^{1/u'} \frac{1}{x^2}\,\overline{F}\left(\frac{1}{x}\right) dx.$

Inequality Measures and Income Distributions

Estimation of tail index ($\widehat{\alpha}$) and implied 1%-top share





(Source: South Africa Income Survey, 2012)

Pareto Models, to measure Inequalities and Risks

It is widely believed that $\widehat{\alpha}$ is upward-biased in surveys

"*The Pareto parameter is estimated using the ratio of the top 5 percent income share to the top decile income share (...). Because those top income shares are often based on survey data (and not tax data), they likely underestimate the magnitude of the changes at the very top*" Atkinson *et al.* (2011)

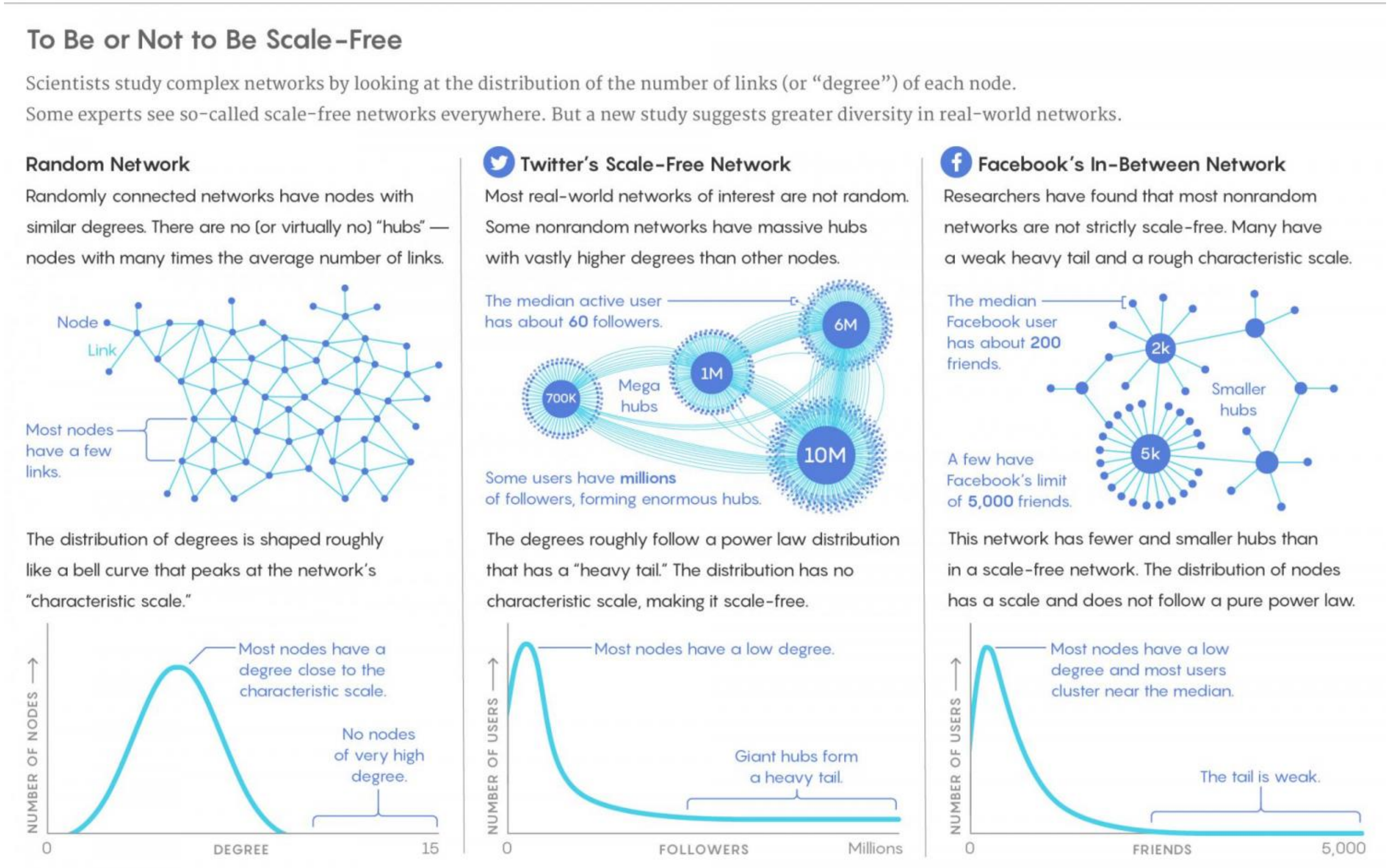
It is widely believed that $\widehat{\alpha}$ is much more reliable in tax data,

but sensitive to misspecification bias (the bias doesn't disappear in huge sample !)

Fitting (strict) Pareto model is not always good...

See C. & Flachaire (2019a)

Scale-free networks

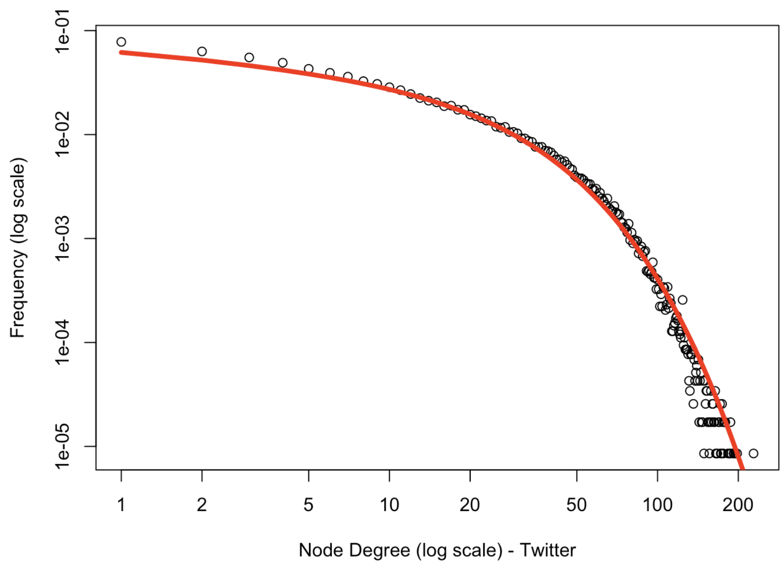


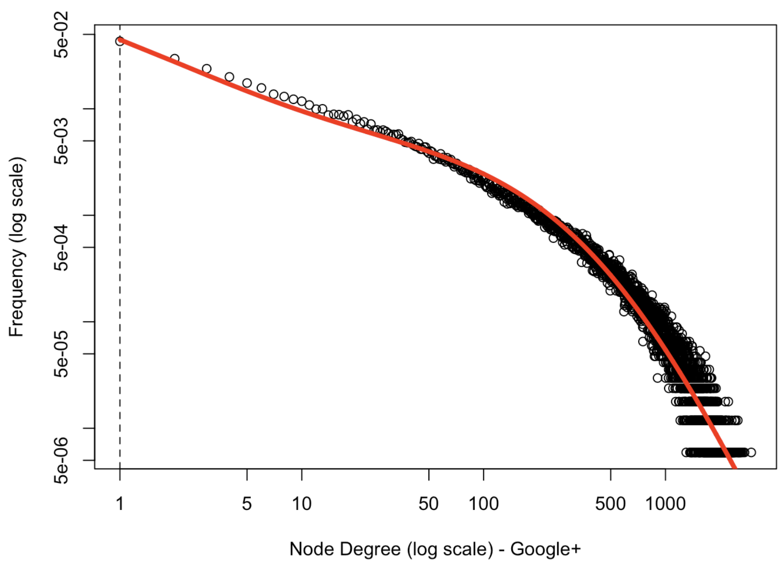
(Source: https://www.colorado.edu/biofrontiers/2018/02/15/scant-evidence-power-laws-found-real-world-networks)

(Extended) Scale-free networks

See Broido & Clauset (2019)'s "*scale-free networks are rare*"

Echoing "*subnets of scale-free networks are not scale-free*" Stumpf, Wiuf & May (2005) (sampling properties of networks)





(Source: Stanford Large Network Dataset Collection)

See C. & Flachaire (2019b)

Emeritaat Jan Beirlant

(Source: KU Leuven - Rob Stevens)