Recent Advances in the Theory of Power Law and Applications

Alexis Akira Toda

University of California San Diego

March 1, 2024

This talk

- We study the tail behavior of

$$W_T = \sum_{t=1}^{T} X_t,$$

where
  - $\{X_t\}_{t=1}^{\infty}$: some stochastic process,
  - T: some stopping time.
- Main result: W_T has exponential tails under fairly mild conditions; simple formula for the tail exponent α.
- Example: if $\{X_t\}_{t=1}^{\infty}$ is IID and T is geometric with mean $1/p$, then

$$(1-p)\,\mathrm{E}[\mathrm{e}^{\alpha X}] = 1.$$

Why this problem is interesting

- Many empirical size distributions obey power laws (e.g., city size (Gabaix, 1999), firm size (Axtell, 2001), income, consumption (Toda and Walsh, 2015), wealth, etc.)

$$\mathrm{P}(S > s) \sim s^{-\alpha},$$

  where S: size.
- Popular explanation is "random growth model": $S_t = G_t S_{t-1}$, where G: gross growth rate.
- Taking logarithm and setting $W_t = \log S_t$, $X_t = \log G_t$, we obtain the random walk

$$W_t = W_{t-1} + X_t.$$

  Hence if $W_0 = 0$, we have $W_T = \sum_{t=1}^T X_t$.

Questions

- Most existing explanations using random growth assume IID Gaussian environment (geometric Brownian motion; Reed, 2001).
- Given ubiquity of power law distributions in empirical data (likely non-IID and non-Gaussian), generative mechanism should be robust (not depend on IID Gaussian assumptions).

Questions:

1. Do non-Gaussian, Markovian random growth processes generate Pareto tails?
2. If so, how is Pareto exponent determined?

Contribution

- Characterize tail behavior of random growth models with non-Gaussian, Markovian shocks.
  1. Analytical determination of Pareto exponent.
  2. Comparative statics.
- Two applications:
  1. Estimate random growth model using Japanese prefecture/municipality population data. Model consistent with observed Pareto exponent but *only after* allowing for Markovian dynamics.
  2. Estimate random growth model using US county daily COVID case data. Model consistent with observed Pareto exponent.

Basic setup of Beare and Toda (2022)

Object of interest:

- We seek to characterize the behavior of tail probabilities

$$\mathrm{P}(W_T > w) \quad \text{and} \quad \mathrm{P}(W_T < -w)$$

as $w \to \infty$, where...

Basic setup of Beare and Toda (2022)

Object of interest:

- We seek to characterize the behavior of tail probabilities

$$\mathrm{P}(W_T > w) \quad \text{and} \quad \mathrm{P}(W_T < -w)$$

as $w \to \infty$, where...

Markov additive process:

- $\{W_t, J_t\}_{t=0}^{\infty}$ is a Markov additive process, which means...

Basic setup of Beare and Toda (2022)

Object of interest:

- We seek to characterize the behavior of tail probabilities

$$\mathrm{P}(W_T > w) \quad \text{and} \quad \mathrm{P}(W_T < -w)$$

  as $w \to \infty$, where...

Markov additive process:

- $\{W_t, J_t\}_{t=0}^{\infty}$ is a Markov additive process, which means...

Hidden Markov state:

- $\{J_t\}_{t=0}^{\infty}$ is a time homogeneous Markov chain taking values in $\mathcal{N} = \{1, \dots, N\}$.
- The transition probability matrix is $\Pi = (\pi_{nn'})$, where $\pi_{nn'} = \mathrm{P}(J_1 = n' \mid J_0 = n)$.
- Initial condition: ϖ is the $N \times 1$ vector of probabilities $\mathrm{P}(J_0 = n)$, $n = 1, \dots, N$.

Basic setup of Beare and Toda (2022)

Increment process:

- $W_0 = 0,\ W_t = \sum_{s=1}^{t} X_s$.
- Distribution of increment $X_t = W_t - W_{t-1}$ depends only on $(J_{t-1}, J_t) = (n, n')$.
- Special cases:
  1. If $N = 1$, then $\{X_t\}_{t=1}^{\infty}$ is IID.
  2. If $X_t =$ constant conditional on J_t, then $\{X_t\}_{t=1}^{\infty}$ is a finite-state Markov chain.

Stopping time:

- $\{W_t\}_{t=0}^{\infty}$ stops with state-dependent probability.
- $\upsilon_{nn'} = \mathrm{P}(T > t \mid J_{t-1} = n, J_t = n', T \geq t)$: conditional survival probability.
- $\Upsilon = (\upsilon_{nn'})$: survival probability matrix.

Basic setup of Beare and Toda (2022)

Conditional moment generating function:

- For $s \in \mathbb{R}$, define $\psi_{nn'}(s) = \mathrm{E}\left[\mathrm{e}^{sX_1} \,\middle|\, J_0 = n, J_1 = n'\right] \in (0, \infty]$.
- $\Psi(s) = (\psi_{nn'}(s))$: $N \times N$ matrix of conditional MGFs.

Region of convergence:

- We define

$$\mathcal{I} = \left\{ s \in \mathbb{R} : \psi_{nn'}(s) < \infty \text{ for all } n, n' \in \mathcal{N} \right\}.$$

- $\mathcal{I}$ is an interval containing zero, with possibly infinite endpoints.
- $\mathcal{I}$ is the intersection of the N^2 regions of convergence of the conditional moment generating functions of X_t given $(J_{t-1}, J_t) = (n, n')$.

Assumption

Assumption

1. *The matrix* $\Upsilon \odot \Pi$ *is irreducible.*
2. *There exists a pair* (n, n') *such that* $\upsilon_{nn'} < 1$ *and* $\pi_{nn'} > 0$.

Assumption

Assumption

1. *The matrix $\Upsilon \odot \Pi$ is irreducible.*
2. *There exists a pair (n, n') such that $\upsilon_{nn'} < 1$ and $\pi_{nn'} > 0$.*

- $\Upsilon \odot \Pi$ is Hadamard (entry-wise) product.
- A matrix A is irreducible if for any pair (n, n'), there exists k such that $|A|^k_{nn'} > 0$.
- Intuitively, irreducibility of $\Upsilon \odot \Pi$ means we can transition from n to n' eventually without stopping.
- $\upsilon_{nn'} < 1$ and $\pi_{nn'} > 0$ guarantees $T < \infty$ almost surely.
- $\rho(A)$: spectral radius (largest absolute value of all eigenvalues) of A.

Main result

Theorem

As a function of $s \in \mathcal{I}$, the spectral radius $\rho(\Upsilon \odot \Pi \odot \Psi(s))$ is convex and less than 1 at $s = 0$. There can be at most one positive $\alpha \in \mathcal{I}$ such that

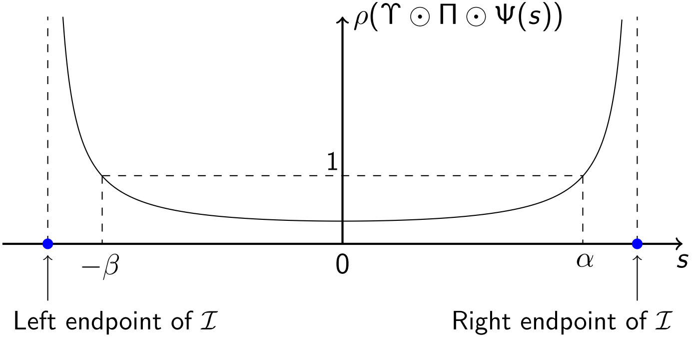
$$\rho(\Upsilon \odot \Pi \odot \Psi(\alpha)) = 1,$$

and if such α exists in the interior of $\mathcal{I}$ then

$$\lim_{w \to \infty} \frac{1}{w} \log \mathrm{P}(W_T > w) = -\alpha.$$

- Similar statement holds for lower tail ($-\beta < 0$ instead of $\alpha > 0$).

Determination of α and β



Refinement

Theorem

Let everything be as above. Then there exist $A, B > 0$ such that

$$\lim_{w\to\infty} \mathrm{e}^{\alpha w}\mathrm{P}(W_T > w) = A,$$
$$\lim_{w\to\infty} \mathrm{e}^{\beta w}\mathrm{P}(W_T < -w) = B$$

except when there exist $c > 0$ and $a_{nn'} \in \mathbb{R}$ such that

$$\mathrm{supp}(X_1 | J_0 = n, J_1 = n') \subset a_{nn'} + c\mathbb{Z}$$

for all $n, n' \in \mathcal{N}$. (We can take $a_{nn} = 0$ if $\upsilon_{nn}\pi_{nn} > 0$.)

Geometrically stopped random growth processes

Theorem

Let everything be as above. Let $S_0 > 0$ be a random variable independent of W_T satisfying $\mathrm{E}[S_0^{\alpha+\epsilon}] < \infty$ *for some* $\epsilon > 0$*, and define the random variable* $S = S_0 \mathrm{e}^{W_T}$*. Then there exist numbers* $0 < A_1 \le A_2 < \infty$ *such that*

$$A_1 = \liminf_{s \to \infty} s^{\alpha} \mathrm{P}(S > s) \le \limsup_{s \to \infty} s^{\alpha} \mathrm{P}(S > s) = A_2,$$

with $A_1 = A_2 = A$ *unless there exist* $c > 0$ *and* $a_{nn'} \in \mathbb{R}$ *such that* $\operatorname{supp}(X_1 | J_0 = n, J_1 = n') \subset a_{nn'} + c\mathbb{Z}$ *for all* $n, n' \in \mathcal{N}$*.*

- S has a Pareto upper tail with exponent α.

Proof of main result

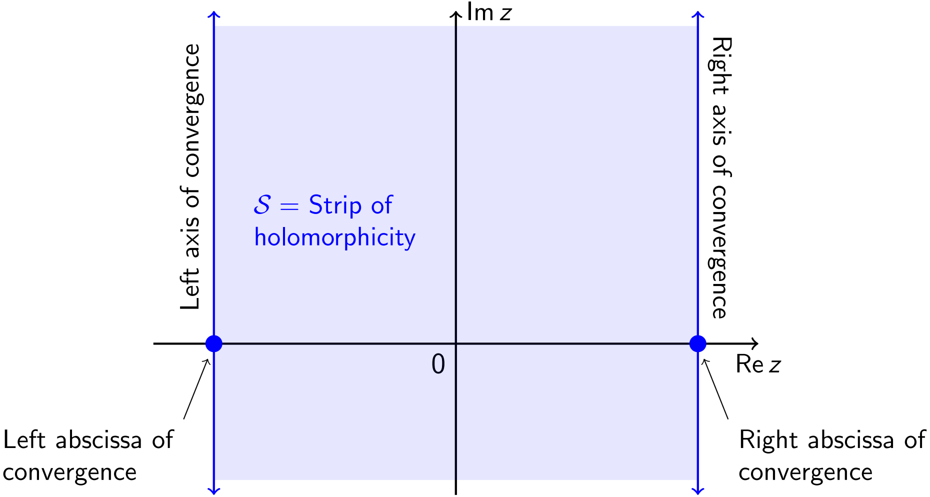
- The proof uses several mathematical results:
    1. Nakagawa (2007)'s Tauberian Theorem and its refinement
    2. Convex inequalities for spectral radius
    3. Perron-Frobenius Theorem
    4. Residue formula for matrix pencil inverses
- For the IID case, we can avoid 2–4 above.

Skip proof

Laplace transform

- For a random variable X with cdf F, let
$$\psi(s) = \mathsf{E}[\mathrm{e}^{sX}] = \int_{-\infty}^{\infty} \mathrm{e}^{sx} \, \mathrm{d}F(x)$$
be its moment generating function (mgf), which is also known as the (two-sided) Laplace transform.
- Since e^{sx} convex in s, so is $\psi(s)$; hence its domain $\mathcal{I} = \{s \in \mathbb{R} : \psi(s) < \infty\}$ is an interval. Let $-\beta \le 0 \le \alpha$ be boundary points (may be 0 or $\pm\infty$).
- For $z \in \mathbb{C}$, by definition of Lebesgue integral,
$$\psi(z) = \mathsf{E}[\mathrm{e}^{zX}] = \int_{-\infty}^{\infty} \mathrm{e}^{zx} \, \mathrm{d}F(x)$$
exists and finite if and only if $\operatorname{Re} z \in \mathcal{I}$. $\psi(z)$ holomorphic on strip of analiticity $\mathcal{S} = \{z \in \mathbb{C} : -\beta < \operatorname{Re} z < \alpha\}$.

Strip of holomorphicity



Tauberian theorem

Theorem (Essentially, Theorem 5* of Nakagawa, 2007)

Let X be a real random variable and $\psi(z) = \mathrm{E}[\mathrm{e}^{zX}]$ its Laplace transform with right abscissa of convergence $0 < \alpha < \infty$ and strip of holomorphicity $\mathcal{S}$. Suppose $A := \lim_{s \uparrow \alpha} (\alpha - s)\psi(s)$ exists, and let B be the supremum of all $b > 0$ such that $\Psi(z) + A(z-\alpha)^{-1}$ continuously extends to $\mathcal{S}_b^+ = \mathcal{S} \cup \{z \in \mathbb{C} : z = \alpha + it, |t| < b\}$. Suppose that $B > 0$. Then we have

$$\frac{2\pi A/B}{\mathrm{e}^{2\pi\alpha/B} - 1} \le \liminf_{x\to\infty} \mathrm{e}^{\alpha x} \mathrm{P}(X > x) \le \limsup_{x\to\infty} \mathrm{e}^{\alpha x} \mathrm{P}(X > x) \le \frac{2\pi A/B}{1 - \mathrm{e}^{-2\pi\alpha/B}},$$

where the bounds should be read as A/α if $B = \infty$.

Discussion

- By previous result, taking logarithm and letting $x \to \infty$, we get
$$\lim_{x\to\infty} \frac{\log \mathrm{P}(X > x)}{x} = -\alpha,$$
which is Nakagawa (2007)'s main result.
- Example: mgf of exponential distribution with exponent α is
$$\psi(z) = \int_0^\infty \alpha \mathrm{e}^{-\alpha x} \mathrm{e}^{zx} \,\mathrm{d}x = \frac{\alpha}{\alpha - z},$$
so we can take $A = \alpha$ and $B = \infty$.

Proof of main result for IID case

- Let $\{X_t\}_{t=1}^{\infty}$ be IID with mgf $\psi_X(z) = \mathrm{E}[\mathrm{e}^{zX}]$.
- mgf of $W_T = \sum_{t=1}^{T} X_t$ when T is geometric with mean $1/p$ is

$$\psi_W(z) = \sum_{k=1}^{\infty}(1-p)^{k-1}p(\psi_X(z))^k = \frac{p\psi_X(z)}{1-(1-p)\psi_X(z)}.$$

- Since $\psi_X(z)$ holomorphic, pole of $\psi_W(z)$ satisfies $\psi_X(z) = \frac{1}{1-p}$.
- Using convexity of $\psi_X(s+it)$ with respect to s, easy to show pole is simple.
- Hence assumption of Tauberian theorem satisfied. Tail exponents satisfy

$$\mathrm{E}[\mathrm{e}^{\alpha X}] = \mathrm{E}[\mathrm{e}^{-\beta X}] = \frac{1}{1-p}.$$

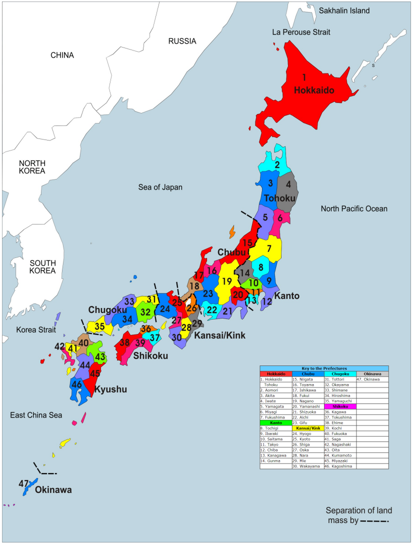
Application 1: Power law in Japanese municipalities

- Main question: are time series properties of population dynamics estimated from panel consistent with a stationary Pareto distribution estimated from cross-section?
- Estimate either at
  - 47 prefecture level (1873-) or
  - 1741 municipality level (1970-)

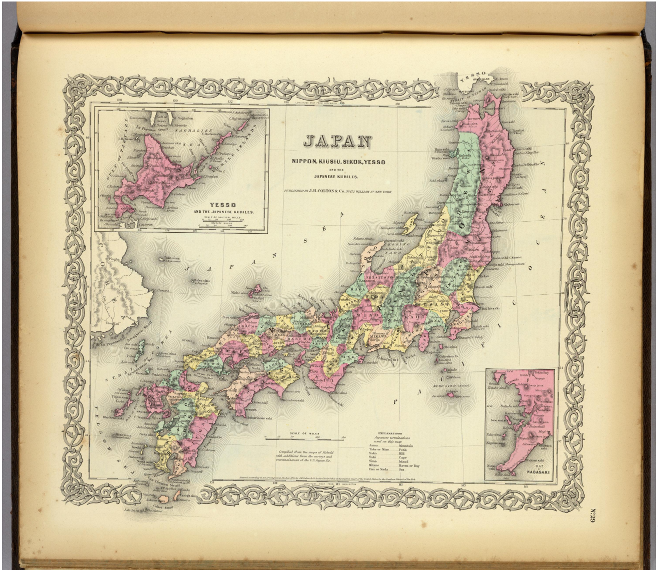
Historical background

- Edo era: 1603-1868. Japan was divided into provinces called *han*, which were controlled by feudal lords called *daimyō*. No free movement of people across regions.
- 1868: Meiji Restoration. Free movement of people.
- 1871: Abolition of the *han* system (*haihan-chiken*). Number and boundary of prefectures settled by 1889
- Boundaries of modern prefectures largely follow those of *ryoseikoku* (province) established in the Nara era (8th century)

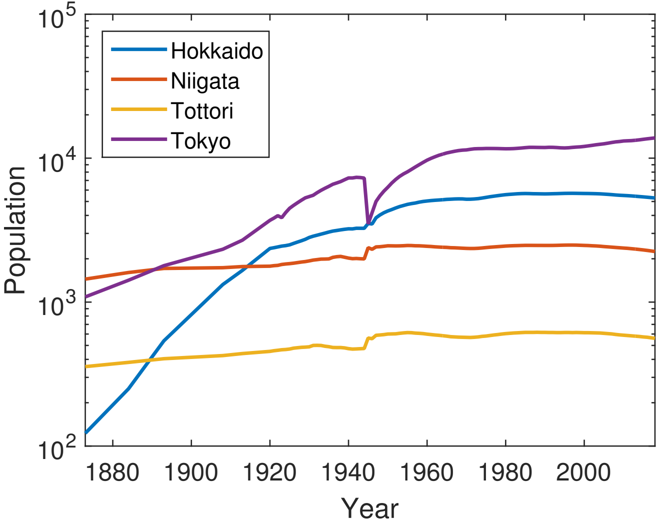
Modern prefectures



Ryoseikoku


JAPAN
NIPPON, KIUSIU, SIKOK, YESSO
AND THE
JAPANESE KURILES.
YESSO
AND THE JAPANESE KURILES.
NAGASAKI

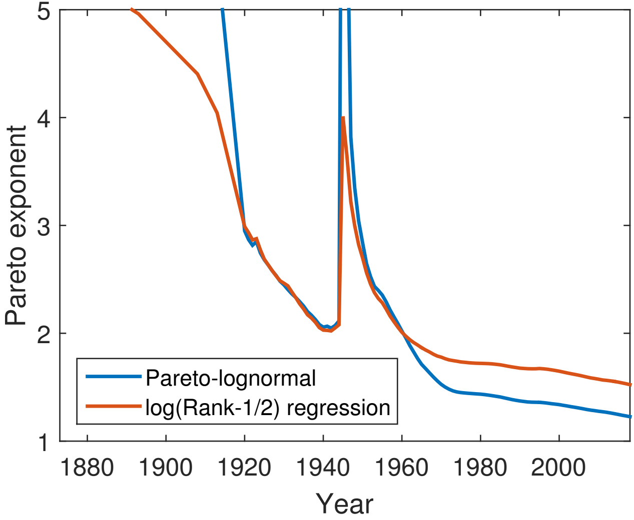
Population of selected prefectures



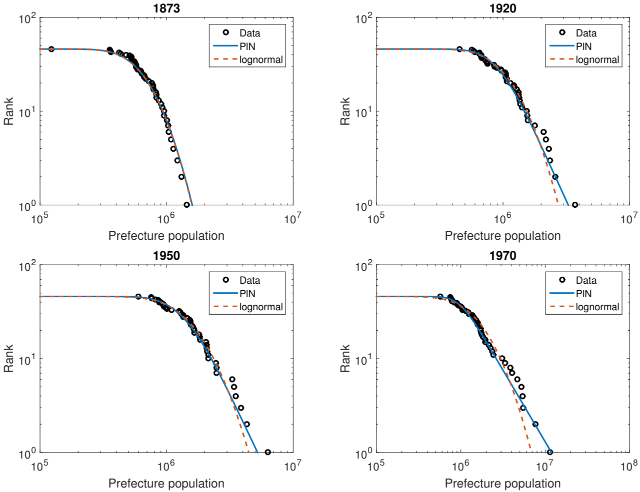
Cross-sectional estimation

- For each year, assume that the cross-sectional distribution of prefecture population is Pareto-lognormal (product of independent Pareto and lognormal distributions).
- Three parameters (μ, σ, α), mean and standard deviation of lognormal component and Pareto exponent.
- Lognormal is special case by setting $\alpha = \infty$.

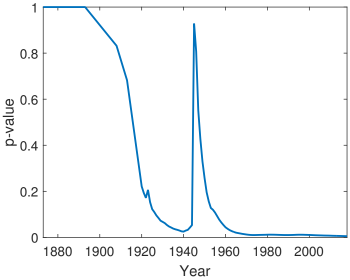
Pareto exponents



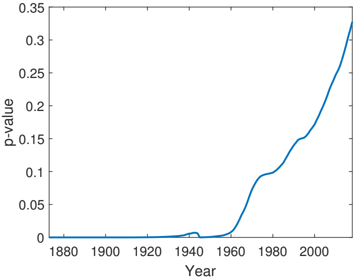
Log-log plot


1873
1920
1950
1970
Data
PIN
lognormal
Rank
Prefecture population
10^0
10^1
10^2
10^5
10^6
10^7
10^8

Likelihood ratio tests



(a) Test of lognormality ($\alpha = \infty$).



(b) Test of Zipf's law ($\alpha = 1$).

Panel estimation

- Assume relative size S_{it} of prefecture i in year t follows random growth process $S_{i,t+1} = G_{i,t+1}S_{it}$, where $G_{i,t+1}$: gross growth rate between year t and $t+1$.
- N-state Markov switching model with conditionally Gaussian shocks:

$$\log G_{i,t+1} \mid n_{it} = n \sim N(\mu_n, \sigma_n^2),$$

  where state n_{it} evolves as a Markov chain with transition probability matrix Π.
- Consider $N = 1, 2, 3$; estimate parameters from post war data by maximum likelihood using Hamilton (1989) filter.
- Compute implied Pareto exponent by solving

$$\rho(\Pi \operatorname{diag}(\mathrm{e}^{\mu_1 s + \sigma_1^2 s^2/2}, \dots, \mathrm{e}^{\mu_N s + \sigma_N^2 s^2/2})) = \frac{1}{1-p}.$$

Estimation of random growth model

N	1
Π	1
μ	-0.0035
σ	0.0111
$\log L$	9,925
α	**56.7**
α_{2015}	1.3

Estimation of random growth model

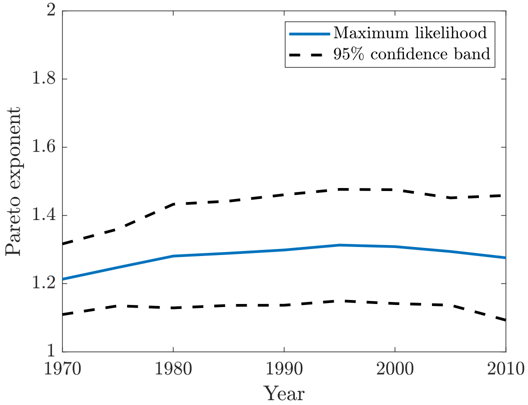
N	1	2
Π	1	$\begin{bmatrix} 0.9754 & 0.0246 \\ 0.0283 & 0.9717 \end{bmatrix}$
μ	-0.0035	$\begin{bmatrix} -0.0030 & -0.0030 \end{bmatrix}^\top$
σ	0.0111	$\begin{bmatrix} 0.0029 & 0.0169 \end{bmatrix}^\top$
$\log L$	9,925	11,638
α	**56.7**	**26.8**
α_{2015}	1.3	1.3

Estimation of random growth model

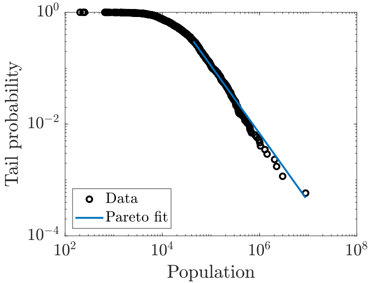
N	1	2	3
Π	1	$\begin{bmatrix} 0.9754 & 0.0246 \\ 0.0283 & 0.9717 \end{bmatrix}$	$\begin{bmatrix} 0.9439 & 0.0561 & 0.0000 \\ 0.0145 & 0.9671 & 0.0184 \\ 0.0210 & 0.0141 & 0.9649 \end{bmatrix}$
μ	-0.0035	$\begin{bmatrix} -0.0030 & -0.0030 \end{bmatrix}^\top$	$\begin{bmatrix} -0.0122 & -0.0022 & 0.0084 \end{bmatrix}^\top$
σ	0.0111	$\begin{bmatrix} 0.0029 & 0.0169 \end{bmatrix}^\top$	$\begin{bmatrix} 0.0053 & 0.0026 & 0.0199 \end{bmatrix}^\top$
$\log L$	9,925	11,638	12,388
α	**56.7**	**26.8**	**1.61**
α_{2015}	1.3	1.3	1.3

Cross-sectional estimation for municipalities

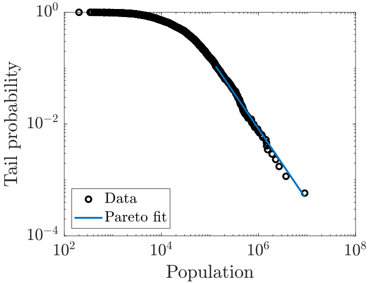
- Estimate Pareto exponent by maximum likelihood (Hill estimator).



Cross-sectional estimation for municipalities



(a) 1970.



(b) 2010.

Panel estimation for municipalities

- Consider $N = 1, \dots, 5$; estimate parameters by maximum likelihood using Hamilton (1989) filter and expectation-maximization algorithm.
- Compute implied Pareto exponent by solving

$$(1-p)\rho(\Pi \operatorname{diag}(\mathrm{e}^{\mu_1 s+\sigma_1^2 s^2/2}, \dots, \mathrm{e}^{\mu_N s+\sigma_N^2 s^2/2})) = 1.$$

Panel estimation for municipalities

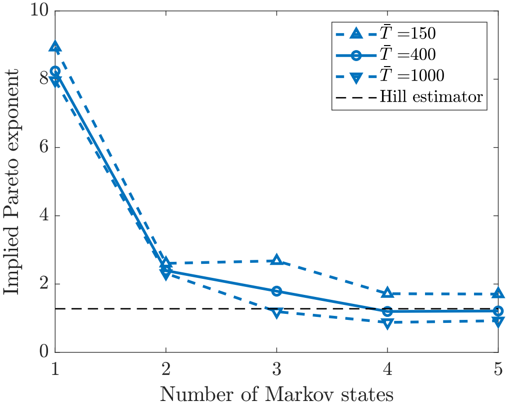
- Consider $N = 1, \dots, 5$; estimate parameters by maximum likelihood using Hamilton (1989) filter and expectation-maximization algorithm.
- Compute implied Pareto exponent by solving

$$(1-p)\rho(\Pi \operatorname{diag}(\mathrm{e}^{\mu_1 s+\sigma_1^2 s^2/2}, \dots, \mathrm{e}^{\mu_N s+\sigma_N^2 s^2/2})) = 1.$$

- Choosing mean age $\bar{T} = 1/p$:
  - Meiji Restoration is in 1868, so lower bound $\bar{T} = 150$.
  - Kamakura Shogunate started in 1185, so upper bound $\bar{T} = 1000$.
  - Tokugawa Shogunate started and moved capital to Tokyo in 1603, so $\bar{T} = 400$ reasonable.
  - Hence consider $p = 1/1000, 1/400, 1/150$.

Implied Pareto exponent

- With $N = 1$ (IID), $\alpha \approx 8 \gg 1$.



Application 2: Power law in COVID-19 cases

- Main question: are growth dynamics and random stopping consistent with Pareto exponent estimated from cross-section?
- Analysis from Beare and Toda (2020)
- Data:
  - Daily COVID-19 case data from January 2020 to March 2020
  - US counties (2,121 counties with at least one case out of 3,243 counties)
  - Merge 5 boroughs of New York City as "New York"

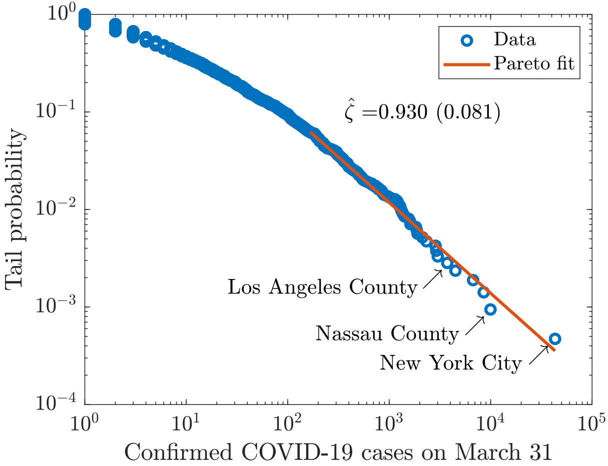
SIR model

- Susceptible-Infected-Recovered (SIR) model:

$$\begin{aligned}\dot{S} &= -\beta S I,\\ \dot{I} &= \beta S I - \gamma I,\\ \dot{R} &= \gamma I,\\ S + I + R &= 1\end{aligned}$$

- At beginning of epidemic, we have $S \approx 1$, $I \ll 1$, $R \approx 0$
- Easy to show that cumulative cases $C := I + R$ grows at rate $\beta - \gamma$
- In practice, cases grow randomly

Cases on 3/31/2020



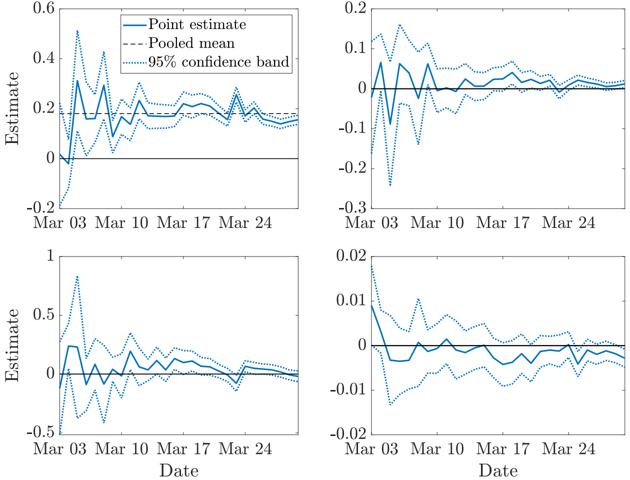
Testing Gibrat's law

- If Gibrat's law holds, growth rate of cases should be independent of current cases
- For each date t, estimate cross-sectional regression

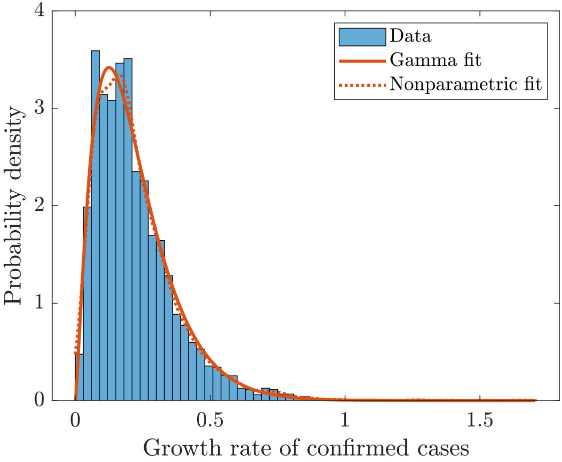
$$\Delta \ln c_{i,t+1} = \beta_{0t} + \beta_{1t} \ln c_{it} + \beta_{2t} \Delta \ln c_{it} + \beta_{3t} D_{it} + \varepsilon_{it}$$

- Here
  - c_{it}: cumulative cases in country i on date t
  - D_{it}: number of days elapsed since first case reported
  - ε_{it}: error term
- Gibrat's law holds if $\beta_{1t} = \beta_{2t} = \beta_{3t} = 0$

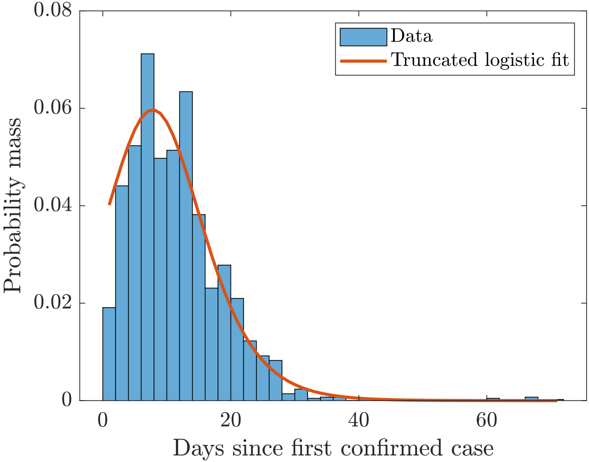
Daily estimates of $\beta_{0t}, \beta_{1t}, \beta_{2t}, \beta_{3t}$



Distribution of growth rate of cases



Distribution of days since first case



Implied Pareto exponent

- Distribution of growth rate is mixture of point mass at 0 and gamma:

$$f(x) = \pi\delta(0) + (1-\pi)\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}$$

with $(\pi, \alpha, \lambda) = (0.128, 2.30, 10.4)$

- Distribution of days since first case is truncated logistic:

$$\mathrm{P}(T = n) = \frac{(1+\phi)(1-q)q^{n-1}}{(1+\phi q^{n-1})(1+\phi q^n)}$$

with $(q, \phi) = (0.825, 4.06)$

Implied Pareto exponent

- MGF of log cases is

$$M_Y(z) = \sum_{n=1}^{\infty} \mathrm{P}(T = n) M(z)^n,$$

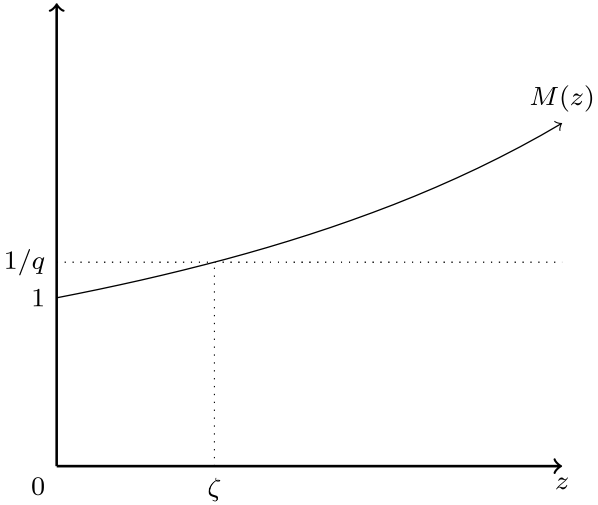
where

$$M(z) = \pi + (1 - \pi)(1 - z/\lambda)^{-\alpha}$$

- Can show $M_Y(z)$ has pole ζ with $M(\zeta) = 1/q$, which gives Pareto exponent
- Solving equation, get

$$\zeta = \lambda \left[1 - \left(\frac{1 - \pi}{1/q - \pi} \right)^{1/\alpha} \right] = 0.928$$

Implied Pareto exponent



Conclusion

- Determination of Pareto exponent under
  - Markov modulation
  - Random stopping
- Many data sets known to obey power law, but generative mechanism has not been tested often
- Evidence for
  - Japanese population dynamics
  - COVID dynamics

References

Axtell, R. L. (2001). "Zipf Distribution of U.S. Firm Sizes". *Science* 293.5536, 1818–1820. DOI: 10.1126/science.1062081.

Beare, B. K. and A. A. Toda (2020). "On the Emergence of a Power Law in the Distribution of COVID-19 Cases". *Physica D: Nonlinear Phenomena* 412, 132649. DOI: 10.1016/j.physd.2020.132649.

Beare, B. K. and A. A. Toda (2022). "Determination of Pareto Exponents in Economic Models Driven by Markov Multiplicative Processes". *Econometrica* 90.4, 1811–1833. DOI: 10.3982/ECTA17984.

Gabaix, X. (1999). "Zipf's Law for Cities: An Explanation". *Quarterly Journal of Economics* 114.3, 739–767. DOI: 10.1162/003355399556133.

References

Hamilton, J. D. (1989). “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle”. *Econometrica* 57.2, 357–384. DOI: 10.2307/1912559.

Nakagawa, K. (2007). “Application of Tauberian Theorem to the Exponential Decay of the Tail Probability of a Random Variable”. *IEEE Transactions on Information Theory* 53.9, 3239–3249. DOI: 10.1109/TIT.2007.903114.

Reed, W. J. (2001). “The Pareto, Zipf and Other Power Laws”. *Economics Letters* 74.1, 15–19. DOI: 10.1016/S0165-1765(01)00524-9.

Toda, A. A. and K. Walsh (2015). “The Double Power Law in Consumption and Implications for Testing Euler Equations”. *Journal of Political Economy* 123.5, 1177–1200. DOI: 10.1086/682729.