Emergence of Zipf's Law in United States: Importance of Markov Dependence

Brendan K Beare[1] Alexis Akira Toda[2]

[1]University of Sydney

[2]Emory University

SAET Conference
July 16, 2026

Power laws and Zipf's law

- Size distributions of many economic variables obey power law:

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

  - income & wealth (Pareto, 1897), $\alpha \approx 1.5$-3
  - cities (Auerbach, 1913; Lotka, 1925; Zipf, 1949), $\alpha \approx 1$
  - firms (Simon, 1955; Axtell, 2001), $\alpha \approx 1$
  - consumption (Toda and Walsh, 2015), $\alpha \approx 4$
  - ...
- Special case: power law with $\alpha \approx 1$ is called Zipf's law and empirically holds for cities and firms

Existing explanations

- Krugman (1996):
  "There must be a compelling explanation of the astonishing empirical regularity."
- Most existing explanations use random growth model
  - Simon (1955), Gabaix (1999), ...
- However, all require assumptions coming outside of model, such as
  - minimum size
  - small expected growth rate of existing units
- Furthermore, simple IID growth models cannot quantitatively explain Zipf's law (Ioannides and Overman, 2003)
- Hence explanation of Zipf's law remains incomplete

Questions

1. Why should random growth produce an exponent near 1?
2. Does estimated dynamics reproduce observed tail quantitatively?

This paper

- **Theory:** random multiplicative growth, latent Markov states, geometric age, entry proportional to contemporaneous average
- **Necessity:** stationary size distribution with Pareto upper tail $\alpha(p) > 1$, with

$$\alpha(p) \to 1 \qquad \text{as mean age } 1/p \to \infty$$

- **Quantitative test:** hidden Markov models for US commuting zone growth, comparing panel-implied and cross-sectional exponents

Main message

Random multiplicative growth can theoretically explain Zipf, but can quantitatively do so only by incorporating Markov dependence

Literature

- Random growth and power laws: Yule (1925), Simon (1955), Gabaix (1999), Reed (2001)
- Size-independent urban growth: Ioannides and Overman (2003), Eeckhout (2004)
- Persistent, history-dependent urban growth: Eaton and Eckstein (1997), Davis and Weinstein (2002)
- Tail exponents under Markov multiplicative processes: Beare and Toda (2022), Beare et al. (2022)
- Consistency between cross-section and panel: Beare and Toda (2020)

Zipf and Gibrat: familiar heuristic (Gabaix, 1999)

- Let S_t denote size normalized by cross-sectional average, with

$$S_t = G_t S_{t-1}, \qquad G_t \text{ independent of } S_{t-1}$$

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

$$\begin{aligned} \mathrm{P}(S_t > s) &= \mathrm{E}\left[\mathrm{P}(S_{t-1} > s/G_t \mid G_t)\right] \\ &\sim \mathrm{E}[G^\alpha] ys^{-\alpha} \implies \mathrm{E}[G^\alpha] = 1 \end{aligned}$$

- Normalization suggests $\mathrm{E}[G] = 1$, implying $\alpha = 1$

Limitations

- Exact proportional growth nonstationary
- Minimum size ad hoc and not testable
- IID restriction unrealistic

Age and entry yield stationarity (Reed, 2001)

- Standard stationary device: reset at birth or entry

$$S_t = \begin{cases} G_t S_{t-1}, & \text{with probability } 1-p \\ \kappa, & \text{with probability } p \end{cases}$$

- Age distribution geometric, mean $1/p$
- Under IID growth, Pareto exponent solves

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

  - Random growth plus random age: power law
  - Equation permits essentially any $\alpha > 0$
  - Economic normalization needed for α near one

Markov-modulated growth (Beare and Toda, 2022)

- Let $J_t \in \{1, \dots, N\}$ follow an irreducible Markov chain, transition matrix $\Pi = (\pi_{nn'})$
- Conditional on $(J_{t-1}, J_t) = (n, n')$, gross growth distributed as $G_{nn'} > 0$
- Define matrix of conditional MGF by $\Psi(z)_{nn'} = \mathrm{E}[G_{nn'}^z]$
- For joint-tail conjecture

$$\mathrm{P}(S_t > s, J_t = n) \sim y_n s^{-\alpha},$$

tail coefficients obey

$$y^\top = y^\top (\Pi \odot \Psi(\alpha))$$

- Perron-Frobenius theorem implies $\rho(\Pi \odot \Psi(\alpha)) = 1$

Pareto exponent under Markov dependence

- With reset probability p, positive exponent solves

$$(1-p)\rho\left(\Pi \odot \Psi(\alpha)\right) = 1$$

- Note:
  - $N = 1$ recovers familiar IID formula $(1-p)\,\mathrm{E}[G^{\alpha}] = 1$
  - $\rho(\Pi \odot \Psi(\alpha))$ aggregates likelihood and magnitude of persistent growth paths
  - Greater persistence $\implies$ longer favorable runs $\implies$ lower α and fatter tail

Quantitative implication

One-period marginal growth can sharply overstate Pareto exponent

Economic ingredients implying Zipf

- Let "average capital" $K_t > 0$ be exogenous
- New location i starts proportional to current average

$$K_{i,t_i} = \kappa K_{t_i},$$

where t_i: time of birth

- Key ingredients
  - Incumbents: multiplicative growth
  - Entrants: satellites of existing economy, initial scale tied to aggregate scale through $\kappa > 0$
  - Relative size: $S_{it} = K_{it}/K_t$

Transparent continuous-time model

- Let $\kappa, \lambda > 0$
- Let "average capital" $K_t > 0$ be given
- "Capital" at location i of age $a := t - t_i$:

$$K_{it} = \kappa K_t \exp\left((\lambda(1-\kappa) - \sigma^2/2)a + \sigma B_i(a)\right), \qquad a \sim \mathrm{Exp}(\lambda)$$

- By construction, $\mathrm{E}[K_{it} \mid K_t] = K_t$

Entry

Initial capital: κ times contemporaneous average

Growth

Relative size: geometric Brownian motion

Age

Sampling age: exponential

Why simple model converges to Zipf

- Relative size K_{it}/K_t follows a double-Pareto distribution (Reed, 2001)
- Upper-tail exponent: positive root of

$$\frac{\sigma^2}{2}z^2 + \left(\lambda(1-\kappa) - \frac{\sigma^2}{2}\right)z - \lambda = 0$$

- Hence

$$\alpha \to 1 \quad \text{as} \quad \lambda \to 0 \quad \text{or} \quad \sigma \to \infty$$

  - $\lambda \to 0$: locations long-lived relative to entry flow
  - Large σ: idiosyncratic risk dominates turnover
- If production function $F(K, L)$ neoclassical, free migration $\implies$ wage equalization $\implies$ $L_{it} \propto K_{it}$ $\implies$ same tail exponent

General discrete-time model

- As before, let $J_t \in \{1, \ldots, N\}$ follow an irreducible Markov chain, gross growth G_t distributed as $G_{nn'} > 0$ conditional on $(J_{t-1}, J_t) = (n, n')$
- Normalized size

$$S_t = \begin{cases} c(p) G_t S_{t-1}, & \text{with probability } 1 - p \\ \kappa, & \text{with probability } p \end{cases}$$

with $c(p)$ chosen for $\mathrm{E}[S_t] = 1$

  - p: reset probability, mean age $1/p$
  - κ: entry size relative to average
  - $c(p)$: normalization of incumbent growth
  - G_t: positive, light-tailed conditional growth

- Nests broad state-dependent processes; no Gaussian shocks or IID growth required

Necessity of Zipf's law

Theorem

If (a) $p\kappa < 1$ (initial size not too large), (b) for any (n, n') and $z > 0$, we have $\mathrm{E}[G_{nn'}^z] < \infty$ (growth rate distributions light-tailed), and (c) there exists n such that $\pi_{nn} > 0$ and G_{nn} has unbounded support (persistent growth possible), then

1. *There exists unique stationary Markov multiplicative process with reset described above*
2. *Size distribution of S_t has Pareto upper tail with exponent $\alpha(p) > 1$ in the sense that*

$$0 < \liminf_{s\to\infty} s^{\alpha(p)} \mathrm{P}(S_t > s) \leq \limsup_{s\to\infty} s^{\alpha(p)} \mathrm{P}(S_t > s) < \infty$$

3. *Zipf's law emerges as $p \to 0$: $\lim_{p\to 0} \alpha(p) = 1$*

Sketch of proof

- For fixed p, tail exponent solves

$$F_p(z) := (1-p)c(p)^z \rho\left(\Pi \odot \Psi(z)\right) = 1,$$

  with $c(p)$ uniquely determined
- As $p \to 0$, can show (this is where work needed)

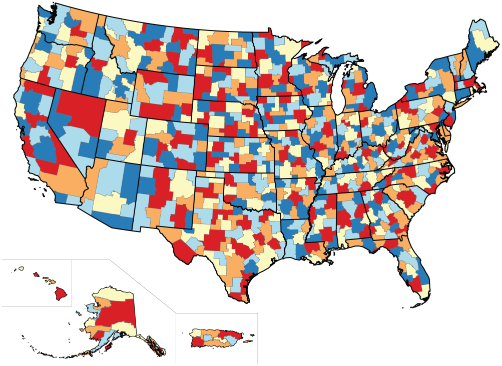
$$c(p) \to c := \frac{1}{\rho(\Pi \odot \Psi(1))}$$

- Limiting equation

$$c^z \rho(\Pi \odot \Psi(z)) = 1$$

  with roots $z = 0$ and $z = 1$
- Strict convexity allows at most two roots; $\alpha(p) > 1$ forces limit 1

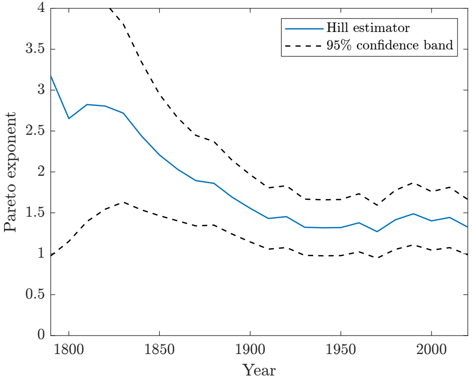
Data: US commuting zones, 1790–2020



- 588 commuting zones under 2020 definition, Puerto Rico excluded
- NHGIS county and state census populations, aggregated to fixed 2020 boundaries
- Cross-sections: positive observations, 1790–2020
- Panel: 1900 onward, avoiding earlier boundary and census artifacts

data quality

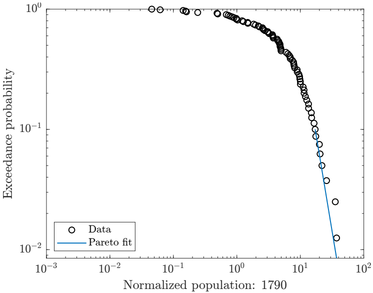
Emergence of Zipf's law over US history



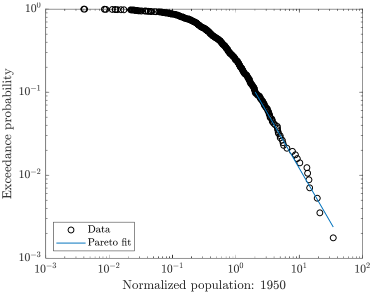
- Hill estimator, largest 10% of commuting zones
- $\widehat{\alpha}$ near 3 around 1790
- Below 1.5 by 1910
- 2020: 1.33

Zipf emerges gradually as US urban system ages and develops

Cross-sectional upper tail becomes nearly linear



1790

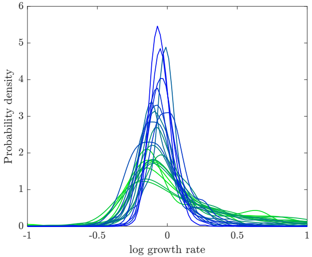


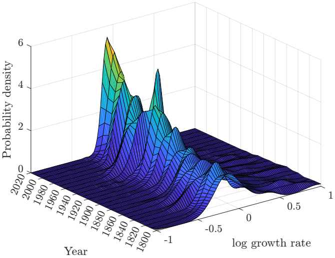
1950

- Population normalized by average commuting-zone population
- 1790: thin, curved upper tail
- By twentieth century: close Pareto line

Growth dynamics after normalization

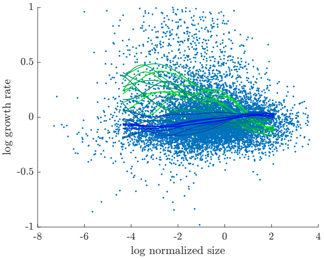
$$S_{it} = \frac{\text{population of commuting zone } i}{\text{average commuting-zone population}}, \qquad g_{it} = \log \frac{S_{it}}{S_{i,t-1}}$$

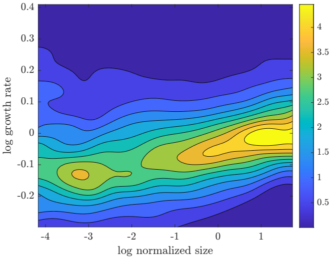




Unconditional growth distribution bell-shaped and broadly stable after 1900; normalization removes aggregate population growth

Growth not IID with respect to current size





- Central 95% of sizes: conditional mean and mode rise with size
- Literal one-state IID Gibrat model rejected
- Markov model: size correlated with latent state through past regime exposure

Hidden Markov model for population growth

- Consider Gaussian hidden Markov model

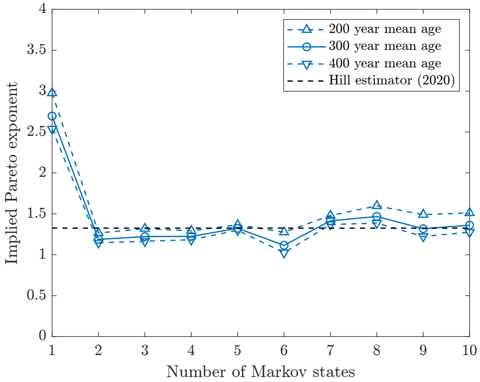
$$S_{i,t+1} = G_{i,t+1} S_{it},$$
$$\log G_{i,t+1} \mid J_{it} = n \sim N(\mu_n, \sigma_n^2), \quad n = 1, \dots, N$$

- Estimate by Baum–Welch algorithm (Rabiner, 1989)
- Implied Pareto exponent solves

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

  - Mean commuting-zone age $\bar{T} = 1/p$: 200, 300, or 400 years
    age
  - Cross-sectional target: 2020 Hill estimate 1.33

Markov dependence closes tail-exponent gap



- $N = 1$: $\alpha \approx 3$
- $N \geq 2$: $\alpha \approx$ 1.1–1.7
- Stable across mean ages 200–400 years
- Centered on cross-sectional estimate 1.33

Panel matches observed tail only with Markov dependence

Two-state estimate

$$\mu = \begin{bmatrix} -0.0740 \\ 0.1541 \end{bmatrix}, \qquad \sigma = \begin{bmatrix} 0.0992 \\ 0.3061 \end{bmatrix}, \qquad \Pi = \begin{bmatrix} 0.9881 & 0.0119 \\ 0.2125 & 0.7875 \end{bmatrix}$$

- State 1: low mean, low volatility; State 2: high mean, high volatility
- High-growth state stationary probability about 5%
- Conditional duration in State 2: about 4.7 census intervals, roughly 47 years
- $N \geq 2$: second largest eigenvalue near 0.85, substantial persistence

Intuition

Rare persistent high-growth episodes: modest effect on typical one-period observation, large effect on far upper tail

Reconciling Gibrat debate

- Multiplicative growth conditional on latent regime
- Size correlated with regime through past growth histories
- Spectral radius captures persistent sequences governing tail

Size dependence rejects iid Gibrat growth, not necessarily random multiplicative growth

Conclusion

- Random multiplicative growth plus random age generate power laws, not Zipf exponent ($\alpha \approx 1$)
- Entry proportional to contemporaneous average plus long-lived system $\implies$ $\alpha \to 1$ under weak conditions
- US cross-section: $\alpha \approx 1.3$
- IID panel: $\alpha \approx 3$; hidden Markov models: close to observed value

Takeaway

Persistence central to tail thickness

References

Auerbach, F. (1913). "Das Gesetz der Bevölkerungskonzentration". *Petermanns Geographische Mitteilungen* 59, 74–76. URL: http://www.mpi.nl/publications/escidoc-2271118.

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

Beare, B. K., W.-K. Seo, and A. A. Toda (2022). "Tail Behavior of Stopped Lévy Processes with Markov Modulation". *Econometric Theory* 38.5, 986–1013. DOI: 10.1017/S0266466621000268.

References

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.

Davis, D. R. and D. E. Weinstein (2002). "Bones, Bombs, and Break Points: The Geography of Economic Activity". *American Economic Review* 92.5, 1269–1289. DOI: 10.1257/000282802762024502.

References

Eaton, J. and Z. Eckstein (1997). "Cities and Growth: Theory and Evidence from France and Japan". *Regional Science and Urban Economics* 27.4–5, 443–474. DOI: 10.1016/S0166-0462(97)80005-1.

Eeckhout, J. (2004). "Gibrat's Law for (All) Cities". *American Economic Review* 94.5, 1429–1451. DOI: 10.1257/0002828043052303.

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

Ioannides, Y. M. and H. G. Overman (2003). "Zipf's Law for Cities: An Empirical Examination". *Regional Science and Urban Economics* 33.2, 127–137. DOI: 10.1016/S0166-0462(02)00006-6.

References

- Krugman, P. (1996). "Confronting the Mystery of Urban Hierarchy". *Journal of the Japanese and International Economies* 10.4, 399–418. DOI: 10.1006/jjie.1996.0023.
- Lotka, A. J. (1925). *Elements of Physical Biology*. Baltimore, MD: Williams & Wilkins Company.
- Pareto, V. (1897). *Cours d'Économie Politique*. Vol. 2. Lausanne: F. Rouge.
- Rabiner, L. R. (1989). "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition". *Proceedings of the IEEE* 77.2, 257–286. DOI: 10.1109/5.18626.
- 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.

References

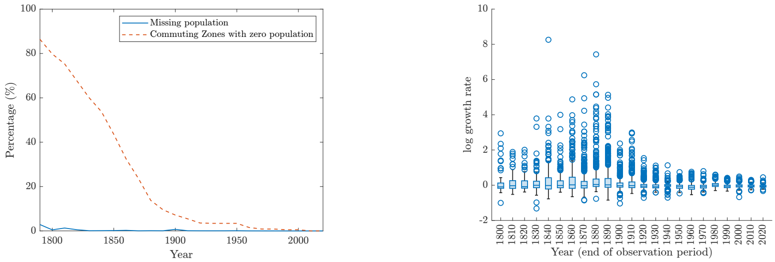
Simon, H. A. (1955). "On a Class of Skew Distribution Functions". *Biometrika* 42.3/4, 425–440. DOI: 10.2307/2333389.

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.

Yule, G. U. (1925). "A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S.". *Philosophical Transactions of the Royal Society B: Biological Sciences* 213.402-410, 21–87. DOI: 10.1098/rstb.1925.0002.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison Wesley.

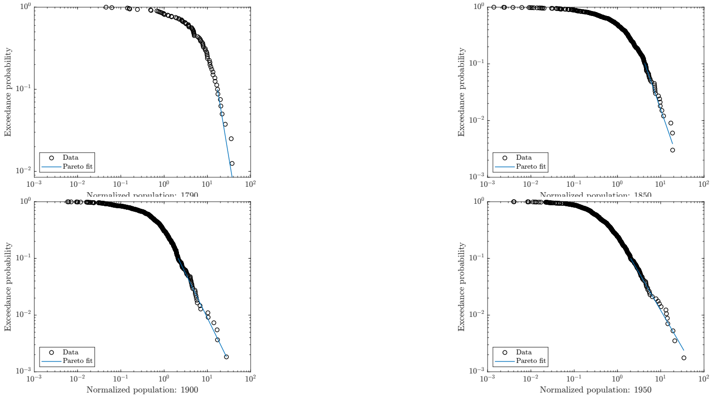
Data coverage and post-1900 panel



- Missing aggregate population below 0.1% through most of twentieth century
- Zero-population commuting-zone share below 10% after 1900
- Pre-1900 growth contains large artifacts from census coverage and boundary changes

return

Full rank-size plots



Mean-age calibration

- Fixed 2020 geography: observed mean age of positive-population commuting zones, 182 years
- Census starts in 1790 with early missing observations, motivating 200-year lower benchmark
- European settlement predates census by roughly two centuries, motivating 400-year upper benchmark
- Implied exponent decreases with mean age; $\bar{T} \in \{200, 300, 400\}$ gives interpretable sensitivity range

Empirical robustness

All three age choices: moving from one state to two or more states produces major exponent drop

◂ return