Tail Mobility

Matthieu Gomez[*]

First version: April 2026
This version: April 2026
[latest version]

Abstract

Random growth models capture a central fact about inequality: the right tail of the distribution is well described by a Pareto law. Little is known, however, about what these models imply for the dynamics of the right tail between two steady states. I show that the transition takes the form of a traveling Pareto tail: the new Pareto exponent first takes hold at moderate incomes and moves upward at a constant speed, with the very top still reflecting the old exponent. The speed of this traveling tail is also the rate at which successful entrants climb to the top, formalizing the link between two well-known observations in the literature: standard models imply low global rates of convergence (Gabaix et al., 2016) and old age at the top (Luttmer, 2011). This speed — which corresponds to the derivative of the cumulant generating function of income growth evaluated at the tail index — captures a dimension of the right tail that the Pareto exponent alone misses: its underlying dynamism. Holding this exponent fixed, the speed can vary by an order of magnitude across otherwise-comparable models, implying transition horizons — and ages at the top — from years to centuries.

Introduction

Top wealth levels can respond to a permanent shock within a year. The local slope of the right tail — the Pareto exponent one would estimate from the same data — typically moves more slowly, over decades. The random growth literature characterizes the initial and final Pareto exponents, but it is silent on what the right tail looks like over the transition.

The main result of this paper is easy to state verbally: the transition takes the form of a *traveling Pareto tail*. After a shock, the new Pareto exponent does not rotate the tail everywhere at once. It first becomes visible at moderate incomes and then advances upward through the distribution at a constant speed, while the very top continues to reflect the old steady state for a long time. Formally, this speed of transition is given by $\Lambda'(\zeta)$, where Λ is the cumulant generating function of (log) income growth and ζ is the Pareto exponent of the new distribution. A shock therefore does not rotate the tail everywhere at once. It changes the local slope only once post-shock paths have had enough time to reach that part of the distribution.

[*]Columbia University, mg3901@columbia.edu.

The same object has a direct mobility interpretation. Successful post-shock entrants observed at log-income level x are typically age $x/\Lambda'(\zeta)$. So slow tail transitions and old age at the top are not two separate failures of standard random growth models; they are the same mechanism viewed in the cross section and in the life cycle. Models calibrated to the same stationary Pareto exponent can imply transition horizons ranging from decades to centuries, depending on whether growth is driven by drift, diffusion, jumps, or persistent heterogeneity. The inverse object $1/\Lambda'(\zeta)$ is also the semi-elasticity of Pareto inequality $1/\zeta$ to permanent shift in growth rates.

This perspective also sharpens model comparison. Holding fixed the stationary Pareto exponent, diffusion raises transition speed only modestly, while jumps or persistent high-growth states can raise it substantially. The relevant question is therefore not only whether a model matches the stationary tail, but also whether it delivers plausible horizons for the tail to adjust and for successful entrants to populate the top.

The unifying contribution is that a single object, $\Lambda'(\zeta)$, answers three questions that have been treated separately. It characterizes the full spatial profile of the transitory tail, not just the decay rate of global distance metrics. It gives that speed a mobility interpretation as the typical rate at which successful post-shock entrants climb through the distribution, so slow tail transitions and old age at the top become two readings of the same mechanism. And it delivers tractable comparative statics across diffusions, jump processes, and Markov-modulated growth models.

Literature review. The stationary Pareto tail of random growth models has been characterized in a large literature, starting with Kesten (1973) and surveyed in Gabaix (2009). The tail exponent is determined by the equation $\Lambda(\zeta) = 0$. The closest paper to mine is Gabaix et al. (2016) who study convergence between steady states and characterize its speed through the second eigenvalue of the generator — a rate that governs how fast the density converges globally. I provide a finer characterization focused on the right tail: which parts of it have already reached their new steady state by time t. This speed turns out to equal the rate at which successful entrants climb to the top, connecting formally the observation that standard models imply slow transitions (Gabaix et al., 2016) with the observation that they imply old age at the top (Luttmer, 2011). My characterization covers the case where the dynamics of wealth is modulated by a Markov process, as in Beare and Toda (2022) and Hansen and Scheinkman (2009).

Methodologically, the paper uses large deviation theory (Touchette, 2009, 2018; Chetrite and Touchette, 2015). This also builds on Gomez and Gouin-Bonenfant (2024), who apply similar tools to characterize stationary Pareto tails under aggregate-state dependence. Here I use them to characterize the full transitory path of the tail between steady states.

1 Transition Dynamics of the Right Tail

This section establishes the main technical result. I first work through the Steindl model, where the transition dynamics are available in closed form and the three-region structure can be read off by inspection. I then show that the same pattern holds for general growth processes, with the transition speed pinned down by the derivative of the scaled cumulant generating function of log-income growth evaluated at the tail exponent.

1.1 Steindl Benchmark

To fix ideas, let us start with the Steindl model.[1] Assume that initially, log income x_t grows at rate μ_0 with death rate δ. At time $t = 0$, the initial cross-sectional distribution is Pareto with exponent ζ_0:

$$p_0(x) = \zeta_0 e^{-\zeta_0 x}$$

with $\zeta_0 = \delta/\mu_0$. Starting from time $t = 0$, the income drift is μ while the death rate is δ. The new stationary density is Pareto:

$$p_\infty(x) = \zeta e^{-\zeta x}$$

with exponent $\zeta = \delta/\mu$. Moreover, in this simple case, the density along its transitory path is known analytically:[2]

$$p_t(x) = \begin{cases} \zeta e^{-\zeta x} & \text{if } x \in (0, \mu t] \\ \zeta_0 e^{-\zeta_0 x + (\zeta_0 \mu - \delta)t} & \text{if } x \in (\mu t, +\infty) \end{cases}$$

In words, in the region $(0, \mu t]$, the density of income is locally Pareto with the new exponent ζ, and, in the region $(\mu t, \infty)$, the density of income is locally Pareto with the old exponent ζ_0.

This gives useful intuition. The new tail effectively travels through the distribution with cutoff $x = \mu t$: below that line the slope has already changed, while above it the old slope remains in place. If $\mu = 2\%$, for instance, it takes 100 years for this cutoff to move by two log points. Equivalently, a given income region switches to the new Pareto exponent exactly when the post-shock cohort has had enough time to reach it. This also clarifies the distinction between level effects and tail-shape effects: high quantiles can move up immediately, but the local tail exponent at a given x does not change until x falls below μt. In the calibration of Figure 2 in Gabaix et al. (2016), where $\delta = 1/30$, this implies roughly $\log(100)/\delta \approx 140$ years for the change in local slope to reach the top 1% cutoff, and $\log(1000)/\delta \approx 210$ years for the top 0.1% cutoff.

[1] See Jones (2015) for a textbook exposition of this class of models.
[2] See, for instance, Lemma 1 of Gabaix et al. (2016).

1.2 The General Case

The intuition is more general. Over an arbitrarily short horizon, all incumbents are exposed to the same post-shock growth law, so the far-right tail initially shifts more than it rotates. The local slope only starts changing once post-shock entrants or sufficiently successful post-shock paths reach that part of the distribution. This section states this rigorously.

Assumption 1. Agents reset with hazard rate δ, and the reinjection distribution has tails thinner than exponential. Conditional on survival, log income evolves as

$$x_t = x_0 + a_t,$$

where $(a_t)_{t\geq 0}$ is the cumulative change in log income.

Economically, Assumption 1 says that entry itself does not create the far-right tail on impact. Entrants are reinjected far enough below the top that the only way to reach the tail is through subsequent growth.

The central object we need from the growth process (a_t) is the long-run growth rate of cross-sectional moments. Since a_t is cumulative log-income growth, $e^{\xi a_t}$ is the ξ-th power of the income ratio e^{x_t}/e^{x_0}. Averaging it over the cross section asks whether that moment grows or decays over long horizons. I summarize the process by

$$\Lambda(\xi) := \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}\left[e^{\xi a_t}\right] - \delta,$$

the exponential growth rate of the ξ-th moment of income, net of the death rate δ. I call Λ the *scaled cumulant generating function* (SCGF) of log-income growth.

The object is familiar in the leading examples. For a Brownian motion $a_t = \mu t + \sigma W_t$, Λ is the quadratic $\Lambda(\xi) = \mu\xi + \sigma^2\xi^2/2 - \delta$; for a compound-Poisson jump process with jump-size MGF M_G, $\Lambda(\xi) = \mu\xi + \phi(M_G(\xi) - 1) - \delta$. More generally, when a_t is an additive functional of a Markov process, Λ is the principal eigenvalue of a tilted generator (see Beare and Toda (2022); Hansen and Scheinkman (2009)).

Three values of Λ matter economically. First, $\Lambda(0) = -\delta < 0$: the total mass of the distribution shrinks at rate δ because of deaths. Second, $\Lambda(\zeta) = 0$ pins down the stationary Pareto exponent: ζ is the unique order at which the corresponding income moment neither grows nor decays under the combined effect of growth and deaths. For the Steindl model, $\Lambda(\xi) = \mu\xi - \delta$ gives $\zeta = \delta/\mu$. Third, the slope $\Lambda'(\zeta)$ at that point will turn out to be the transition speed: it tells us how fast the new tail becomes visible farther up the distribution. Figure 1 illustrates these three quantities.

Assumption 2. The limit $\Lambda(\xi)$ exists and is finite for every $\xi \in \mathbb{R}$, the function Λ is convex, and it has a unique positive root $\zeta > 0$ at which it is differentiable.

This assumption is mild in the processes I consider. Convexity is automatic from Hölder's inequality. Finiteness rules out cases where one-period log changes are themselves so fat-tailed that moment growth is not well defined. Differentiability at ζ holds in all the leading examples.

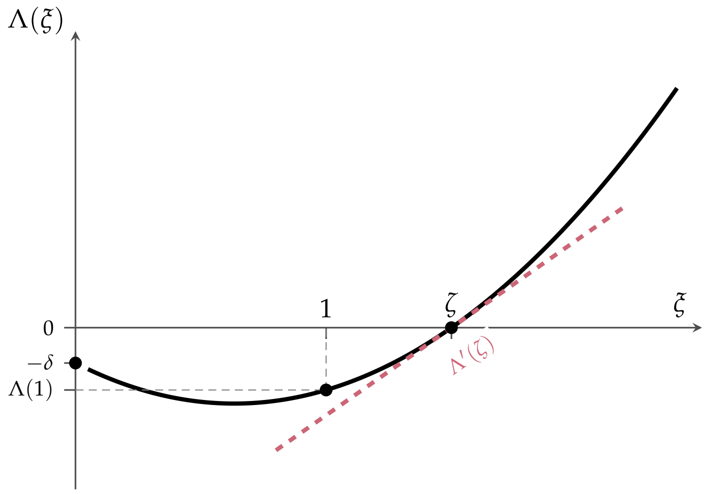


Figure 1: Key quantities read off the SCGF Λ.

Notes: $\Lambda(0) = -\delta$ is the death rate; $\Lambda(1)$ is the growth rate of aggregate income net of δ and its sign determines whether mean income is finite in the stationary distribution; $\Lambda(\zeta) = 0$ defines the tail index; $\Lambda'(\zeta)$ is the transition speed.

Assumption 3. $x_0 \geq 0$, x_0 is independent of a, and x_0 has Pareto upper-tail exponent ζ_0, i.e.

$$\lim_{x\to\infty} \frac{1}{x} \log \mathbb{P}(x_0 \geq x) = -\zeta_0.$$

I assume moreover that Λ is differentiable at ζ_0.

This says that the initial cross section already has a Pareto right tail with exponent ζ_0, and that initial position in the distribution is not mechanically tied to post-shock growth shocks.

Under these assumptions, the cross-sectional distribution converges to a stationary law with a Pareto right tail of exponent ζ. In terms of upper tails,

$$\lim_{x\to\infty} \lim_{t\to\infty} \frac{1}{x} \log \mathbb{P}(x_t \geq x) = -\zeta.$$

If instead one first moves arbitrarily far into the tail at a fixed date, and only then lets time pass, one recovers the initial exponent:

$$\lim_{t\to\infty} \lim_{x\to\infty} \frac{1}{x} \log \mathbb{P}(x_t \geq x) = -\zeta_0.$$

There is no contradiction. At any finite date, the extreme far-right tail is still inherited from the initial cross section; the new tail only becomes visible at income levels that post-shock growth has had time to reach. Letting $t \to \infty$ before $x \to \infty$ gives the eventual exponent ζ; letting $x \to \infty$ before $t \to \infty$ gives the initial exponent ζ_0.

To resolve this order-of-limits issue, I study what happens along the joint scale $x = \beta t$ for $\beta > 0$ fixed. Here β is the ratio of log-income to time: it parametrizes how deep into the tail one looks, measured in log points per unit of time. At level $x = \beta t$, two kinds of paths can place an agent: either the agent was born after the shock, at some age $\alpha t \in (0, t]$, and grew all the way from zero to βt—requiring average growth β/α; or the agent was alive at date 0, started the transition already in the initial Pareto tail at some level $(\beta - \nu)t$, and then grew by

νt after the shock. Proposition 1 shows that the tail at level βt is produced by whichever of these two mechanisms has the higher probability, and characterizes the regions in which each dominates.

Proposition 1. *Assume Assumptions 1, 2, and 3, and assume moreover that $\zeta_0 \geq \zeta$, where $\zeta > 0$ solves $\Lambda(\zeta) = 0$.*[3] *Then, for every $\beta > 0$,*

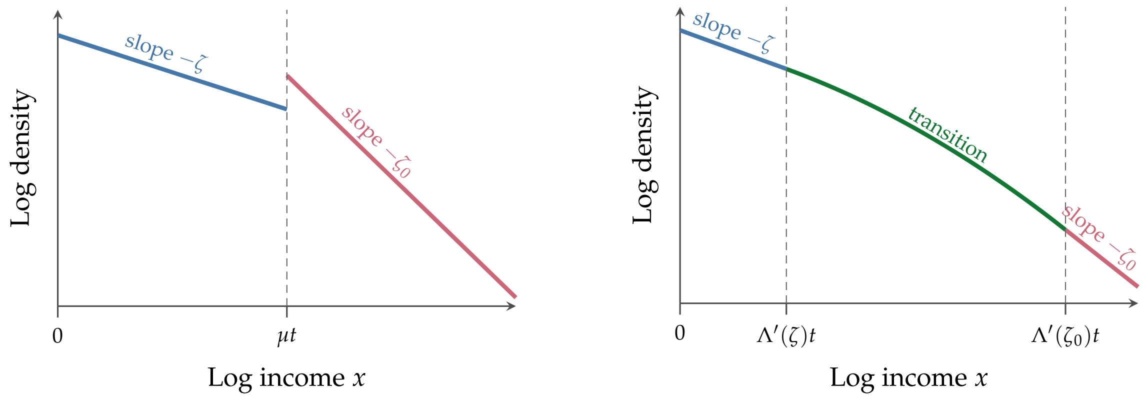
$$\lim_{\varepsilon \downarrow 0} \lim_{t \to \infty} \frac{1}{t} \log \mathbb{P}\left(\left| \frac{x_t}{t} - \beta \right| < \varepsilon \right) = -I(\beta) \tag{1}$$

where

$$I(\beta) = \begin{cases} \zeta\beta & \text{if } \beta \in (0, \Lambda'(\zeta)], \\ \Lambda^*(\beta) & \text{if } \beta \in [\Lambda'(\zeta), \Lambda'(\zeta_0)], \\ \zeta_0\beta - \Lambda(\zeta_0) & \text{if } \beta \in [\Lambda'(\zeta_0), \infty). \end{cases} \tag{2}$$

In economic terms, the proposition implies a simple three-region pattern: below $\Lambda'(\zeta)t$ the local exponent already equals ζ, above $\Lambda'(\zeta_0)t$ it still equals ζ_0, and between the two lies a transition region.

Figure 2, panel (b), shows the three distinct regions of the transitory right tail: a region $(0, \Lambda'(\zeta)t)$ where the local Pareto exponent is ζ, an intermediate region $(\Lambda'(\zeta)t, \Lambda'(\zeta_0)t)$ in which the distribution is not locally Pareto, and a region $(\Lambda'(\zeta_0)t, +\infty)$ where the local Pareto exponent is ζ_0. This is consistent with the Steindl benchmark in panel (a): there $\Lambda(\zeta) = \mu\zeta - \delta$, so $\Lambda'(\zeta) = \Lambda'(\zeta_0) = \mu$, and there is no intermediate region.



(a) Steindl model

(b) General model

Figure 2: Transition of the right tail: Steindl benchmark and general case.

Notes: Panel (a) shows the Steindl benchmark, where the change in local slope occurs at μt. Panel (b) shows the general case, where an intermediate transition region appears between $\Lambda'(\zeta)t$ and $\Lambda'(\zeta_0)t$.

Gabaix et al. (2016) examine the same transition problem through the convergence of the density in $\mathbb{L}^1$ and weighted $\mathbb{L}^1$ norms, or equivalently through Laplace transforms. These objects characterize temporal convergence rates in the tail, but not the spatial profile of the

[3] In the opposite case $\zeta_0 < \zeta$ (inequality declining over the transition), the same computation yields a two-region decomposition with a single crossing at the β where $\zeta\beta = \zeta_0\beta - \Lambda(\zeta_0)$.

tail along the transition. The three-region decomposition of Proposition 1 complements their analysis: it shows which parts of the income distribution have already switched to the new local Pareto slope by time t.

In particular, the region where the local Pareto exponent equals ζ extends up to $\Lambda'(\zeta)t$. Proposition 2 shows that the same object also gives the typical speed at which successful post-shock entrants climb the right tail.

To make that local-slope interpretation precise, compare the tail over two nearby levels, βt and $(\beta+h)t$. When a density exists, this is just the slope of the log-density curve over a short interval. In probability terms, the corresponding local Pareto exponent is

$$\zeta_t^{\text{loc}}(\beta;h) := -\frac{\log \mathbb{P}(x_t \approx (\beta+h)t) - \log \mathbb{P}(x_t \approx \beta t)}{ht}.$$

Proposition 1 therefore gives

$$\lim_{t\to+\infty} \zeta_t^{\text{loc}}(\beta;h) = \frac{I(\beta+h)-I(\beta)}{h}.$$

Hence,

$$\lim_{h\downarrow 0}\lim_{t\to+\infty} \zeta_t^{\text{loc}}(\beta;h) = \begin{cases} \zeta & \text{if } 0<\beta<\Lambda'(\zeta), \\ (\Lambda')^{-1}(\beta) & \text{if } \Lambda'(\zeta)<\beta<\Lambda'(\zeta_0), \\ \zeta_0 & \text{if } \beta>\Lambda'(\zeta_0). \end{cases}$$

So the transition can be summarized in terms of the local slope: ζ for low enough β, ζ_0 for high enough β, and a continuous interpolation in between.

2 Interpreting Transition Speed

The object $\Lambda'(\zeta)$ has already appeared once, as the speed at which the new Pareto tail advances through the cross section. This section shows that the same object also has two complementary economic interpretations. It measures how fast successful post-shock entrants climb the distribution, and, under additive growth shifts, it is the inverse semi-elasticity of the stationary Pareto exponent. Together, these two readings link transition dynamics, mobility at the top, and stationary comparative statics.

2.1 Transition Speed and Mobility

This subsection shows that transition speed is not just a reduced-form description of the cross section. It is exactly the speed at which successful post-shock entrants climb the distribution. Fix $t>0$ and draw an individual from the cross section at time t. Let B_t denote the event that this individual was born after date 0. On B_t, let $A_t \in [0,t]$ denote the individual's age, that is, the time elapsed since birth.

The next proposition shows that, below the cutoff $\Lambda'(\zeta)t$, the right tail is populated by agents born after date 0 and that these successful entrants move through the distribution at speed $\Lambda'(\zeta)$.

Proposition 2 (Age and speed of successful entrants)**.** *Assume the hypotheses of Proposition 1. Fix $\beta \in (0, \Lambda'(\zeta))$ and define*

$$\alpha_\beta := \frac{\beta}{\Lambda'(\zeta)} \in (0,1).$$

Then:

(i) the individuals observed around level βt at time t are asymptotically born after date 0:

$$\lim_{\eta \downarrow 0} \lim_{t\to\infty} \mathbb{P}\Big(B_t \,\Big|\, \frac{x_t}{t} \in (\beta-\eta, \beta+\eta)\Big) = 1;$$

(ii) for every $\varepsilon > 0$, their age concentrates at $\alpha_\beta t$:

$$\lim_{\eta \downarrow 0} \lim_{t\to\infty} \mathbb{P}\left(\left|\frac{A_t}{t} - \alpha_\beta\right| > \varepsilon \,\middle|\, B_t,\ \frac{x_t}{t} \in (\beta-\eta, \beta+\eta)\right) = 0.$$

Equivalently, under the same conditioning,

$$\frac{x_t}{A_t} \longrightarrow \Lambda'(\zeta) \qquad \textit{in probability.}$$

In particular, if $\beta = \alpha\Lambda'(\zeta)$ with $\alpha \in (0,1)$, then the agents observed around level βt at time t were typically born around date $(1-\alpha)t$. Equivalently, successful entrants observed at log-income level x are typically age

$$\frac{x}{\Lambda'(\zeta)}.$$

This is the precise sense in which the transition speed $\Lambda'(\zeta)$ measures both the speed of the transition in the cross section and the speed at which new individuals reach the top. Equivalently, $1/\Lambda'(\zeta)$ is the amount of time needed, asymptotically, to move one unit of log income through the upper tail.

2.2 Transition Speed and Tail Elasticity

This subsection shows that the same slope also governs stationary comparative statics. The inverse transition speed $1/\Lambda'(\zeta)$ is exactly the semi-elasticity of Pareto inequality $1/\zeta$ to permanent shifts in growth rates, under additive growth shifts. So the object that governs how fast the new tail moves through the cross section also governs how strongly the stationary tail reacts to permanent changes in growth.

Consider a one-parameter family of post-shock growth processes indexed by θ, with SCGF $\Lambda(\xi, \theta)$ (where Λ' denotes $\partial_\xi \Lambda$ throughout). Let $\zeta(\theta) > 0$ denote the corresponding stationary tail exponent, defined by

$$\Lambda(\zeta(\theta), \theta) = 0.$$

Proposition 3 (Sensitivity of the stationary tail)**.** *Suppose that, for θ in a neighborhood of θ_0,* *(i)* $(\xi, \theta) \mapsto \Lambda(\xi, \theta)$ *is continuously differentiable;*

(ii) *for each such* θ*, the map* $\xi \mapsto \Lambda(\xi, \theta)$ *is finite and convex, with* $\Lambda(0, \theta) < 0$*;*

(iii) *there exists a unique positive* $\zeta(\theta)$ *such that* $\Lambda(\zeta(\theta), \theta) = 0$*.*

Then

$$\zeta'(\theta_0) = -\frac{\partial_\theta \Lambda(\zeta(\theta_0), \theta_0)}{\Lambda'(\zeta(\theta_0), \theta_0)}.$$

In particular, if the parameter shifts log growth additively so that

$$\Lambda(\xi, \theta) = \Lambda_0(\xi) + \theta\xi,$$

then

$$\partial_\theta \log \zeta(\theta_0) = -\frac{1}{\Lambda'(\zeta(\theta_0), \theta_0)}.$$

Equivalently,

$$\partial_\theta \log \frac{1}{\zeta(\theta_0)} = \frac{1}{\Lambda'(\zeta(\theta_0), \theta_0)}.$$

This gives a second interpretation of transition speed. In the additive-shift benchmark, the inverse transition speed $1/\Lambda'(\zeta)$ is exactly the semi-elasticity of Pareto inequality $1/\zeta$ with respect to the growth shift. Thus, models with small $\Lambda'(\zeta)$ have three linked properties: slow tail adjustment, old successful entrants at the top, and stationary tail exponents that react strongly to permanent growth shifts.

3 Quantifying Transition Speed

I now quantify transition speed. In the first part, I do so in standard models. In the second part, I look directly at the data for indications of its magnitude.

3.1 Across Canonical Models

This subsection quantifies transition speed across standard models using a simple horizon metric: how long it takes the new local Pareto slope to reach the cutoff of the top p%. The answer is that models matched on the same stationary Pareto exponent can still imply transition horizons ranging from years to centuries.

Consider the cutoff of the top p%. Even holding fixed the stationary tail exponent and the death rate, different microeconomic mechanisms can imply very different horizons for the new local slope to reach that point in the distribution.

If the right tail is locally Pareto with exponent ζ, the cutoff for the top p% is approximately

$$x_p \approx \frac{\log(1/p)}{\zeta},$$

so successful entrants observed at that cutoff are typically age

$$\frac{x_p}{\Lambda'(\zeta)} \approx \frac{\log(1/p)}{\zeta\Lambda'(\zeta)}.$$

This puts all models on a common scale: the horizon for the new tail to reach a given top cutoff.

Table 1 summarizes the canonical examples. The key object is always the same — the slope $\Lambda'(\zeta)$ at the tail exponent — but the formula for that slope depends sharply on the underlying microeconomic mechanism.

Table 1: Canonical examples for the transition speed

Model	$\Lambda(\xi)$	Tail exponent ζ	Transition speed $\Lambda'(\zeta)$
Steindl	$\mu\xi - \delta$	$\zeta = \delta/\mu$	δ/ζ
Brownian motion	$\mu\xi + \frac{\sigma^2}{2}\xi^2 - \delta$	$\mu\zeta + \frac{\sigma^2}{2}\zeta^2 = \delta$	$\frac{\delta}{\zeta} + \frac{\sigma^2}{2}\zeta$
Jumps	$\mu\xi + \phi(M_G(\xi) - 1) - \delta$	$\mu\zeta + \phi(M_G(\zeta) - 1) = \delta$	$\frac{\delta}{\zeta} + \phi\left(M'_G(\zeta) - \frac{M_G(\zeta)-1}{\zeta}\right)$
Two-type model	$\max(\mu_H\xi - \tau, \mu_L\xi) - \delta$	$\zeta = (\delta + \tau)/\mu_H$ if the H-state governs the tail	μ_H

Notes: In the jump row, f_G denotes the jump-size density and $M_G(\xi) := \int e^{\xi g} f_G(g)\, dg$ its moment generating function. The appendix derives the closed-form examples used in the numerical comparison below.

Table 2 reports an apples-to-apples comparison for the four illustrative models used in Figure A1 in the appendix. Throughout that table, I hold fixed the stationary tail exponent $\zeta = 2$ and the death rate $\delta = 1/30$.

Table 2: Transition horizons in a common-parameter comparison

Model	$\Lambda'(\zeta)$	Top 1%	Top 0.1%	Top 0.01%
Steindl	0.017	138 years	207 years	276 years
Brownian motion	0.079	29 years	44 years	58 years
Jumps	0.154	15 years	22 years	30 years
Two-type model	0.150	15 years	23 years	31 years

Notes: All models share $\zeta = 2$ and $\delta = 1/30$. Brownian: $\sigma = 0.25$. Jumps: degenerate jump-size distribution with mass at $G = 1.6$, intensity $\phi = 0.005$. Two-type: $\mu_H = 0.15$, with transition rate $\tau = \mu_H\zeta - \delta$. Horizons are $\log(1/p)/[\zeta\Lambda'(\zeta)]$ for the top p fraction. See Figure A1 for the corresponding SCGFs.

The magnitudes vary enormously even though the stationary tail is the same. In this common-parameter comparison, a model that matches the same Pareto exponent can imply that the new tail reaches the top 1% in 15 years or in 138 years. That difference is economically first-order: it changes the predicted persistence of tail inequality and the age composition of those who occupy the top.

Steindl benchmark. In the pure-drift model, $\Lambda(\xi) = \mu\xi - \delta$, so $\zeta = \delta/\mu$ and

$$\Lambda'(\zeta) = \mu = \frac{\delta}{\zeta}.$$

Hence the typical age of successful entrants at the top $p\%$ cutoff is $\log(1/p)/\delta$. With $\delta = 1/30$, this implies roughly 140 years for the top 1% cutoff and 210 years for the top 0.1% cutoff.

Brownian motion. If a_t is Brownian motion with drift μ and volatility σ, then

$$\Lambda(\xi) = \mu\xi + \frac{\sigma^2}{2}\xi^2 - \delta.$$

Using $\Lambda(\zeta) = 0$ to eliminate μ yields

$$\Lambda'(\zeta) = \mu + \sigma^2\zeta = \frac{\delta}{\zeta} + \frac{\sigma^2}{2}\zeta.$$

The key intuition is selection. Ex ante, every incumbent has expected log-growth rate μ. But ex post, the people who actually make it to the top are precisely those who realized unusually favorable Brownian shocks, so their average realized growth rate is $\mu + \zeta\sigma^2 = \Lambda'(\zeta)$. Diffusion therefore raises transition speed relative to the Steindl benchmark, but only gradually. In the common-parameter comparison of Table 2, I set $\zeta = 2$, $\delta = 1/30$, and $\sigma = 0.25$, which gives $\Lambda'(\zeta) \approx 8\%$ and therefore horizons of about 29 years for the top 1% cutoff and 44 years for the top 0.1% cutoff. Note that the Brownian calibration in Section 4.3 of Gabaix et al. (2016) gives $\Lambda'(\zeta) \approx 4\%$, implying about 69 years to the top 1% cutoff and 104 years to the top 0.1% cutoff.

Jump growth. For a compound Poisson process with drift and jump-size density f_G, let

$$M_G(\xi) := \int e^{\xi g} f_G(g)\, dg$$

denote the moment generating function of jump size. Then

$$\Lambda(\xi) = \mu\xi + \phi(M_G(\xi) - 1) - \delta.$$

Eliminating μ using $\Lambda(\zeta) = 0$ gives

$$\Lambda'(\zeta) = \frac{\delta}{\zeta} + \phi\left(M_G'(\zeta) - \frac{M_G(\zeta) - 1}{\zeta}\right),$$

which is increasing in the jump intensity ϕ when jump sizes are nonnegative. Thus, among models with the same stationary Pareto exponent, more jump-driven growth generates faster tail transitions and younger successful entrants at the top. For the numerical comparison in Table 2, I specialize to the degenerate case with jump size $G = 1.6$.

Persistent high-growth types. In the two-type model, when the high-growth state governs the tail,

$$\zeta = \frac{\delta + \tau}{\mu_{\mathrm{H}}} \qquad \text{and} \qquad \Lambda'(\zeta) = \mu_{\mathrm{H}}.$$

Transition speed is therefore pinned down by the drift of the high-growth type. Increasing μ_{H} makes successful entrants younger and transitions faster, while increasing τ raises the stationary tail exponent without changing transition speed.

These formulas make the connection between Gabaix et al. (2016) and Luttmer (2011) operational. A model that matches the stationary Pareto exponent but predicts old successful entrants at the top also predicts slow tail transitions. Conversely, any mechanism that generates rapid changes in tail inequality must also generate young successful entrants among those who rise to the top.

Global measures can look fast even when the top tail adjusts slowly. In the Brownian calibration of Gabaix et al. (2016), the overall half-life is about 20 years, but the local Pareto slope still takes roughly 69 years to change at the top 1% cutoff and 104 years at the top 0.1% cutoff. The reason is that global metrics average over the whole distribution, whereas $\Lambda'(\zeta)$ asks when a particular part of the tail actually changes slope.

Appendix E collects the corresponding derivations for the closed-form examples used in the numerical comparison.

3.2 Empirical Estimates in the Forbes 400 and CRSP

This subsection asks whether the transition speeds implied by the theory look plausible in the data. To answer that question, I estimate $\Lambda'(\zeta)$ in two ways — a Brownian benchmark and an exponentially tilted-growth proxy — and compare both with reduced-form mobility measures in the Forbes 400 wealth panel and among CRSP-listed firms. The main message is simple: in both samples, the two model-based estimates are very close to each other, and both lie strictly between the recent-speed and lifetime-speed proxies. Table 3 summarizes the results, placing the implied speed at roughly 9.7% per year in the Forbes 400 and 17.1% per year in CRSP.

Moment-based estimators. Let $\tilde{g}_{it} := \log W_{i,t+1} - \log W_{it}$ denote winsorized one-year forward log wealth growth, and let μ_t and σ_t be its yearly cross-sectional mean and standard deviation. I use a wider sample here than for the speed proxies below because the growth moments require more data than a top-100 average: the top 400 in Forbes and the top 1000 in CRSP. From these moments we build two estimators,

$$\widehat{\Lambda}'(\zeta)_t^{\text{Brownian}} := \mu_t + \zeta\sigma_t^2, \qquad \widehat{\Lambda}'(\zeta)_t^{\text{direct}} := \frac{\overline{\tilde{g}_{it}^{(h)}\, e^{\zeta h_{it}\tilde{g}_{it}^{(h)}}}}{\overline{e^{\zeta h_{it}\tilde{g}_{it}^{(h)}}}},$$

where the bar denotes the yearly cross-sectional average and $\tilde{g}_{it}^{(h)} := (\log W_{i,t+h} - \log W_{it})/h$ is per-unit-time log growth over horizon h. The Brownian estimator imposes the Brownian SCGF $\Lambda(\xi) = \mu\xi + \frac{1}{2}\sigma^2\xi^2 - \delta$ and recovers $\Lambda'(\zeta)$ from the first two one-year growth moments. The direct estimator is an exponentially tilted growth proxy: in the independent-increment benchmarks that motivate my closed-form examples, it coincides with the derivative of the horizon-h log moment-generating function per unit of time.

The two estimators differ in the horizon they use. In Forbes, I set $h_{it} = 1$ throughout. In CRSP, for any fixed long horizon, the firms observed near the end of the panel would be systematically dropped. Because these late-panel firms are disproportionately recent entrants into the top 1000 — exactly the high-growth observations that the exponential tilt upweights — such a truncation would bias the direct estimator downward. I therefore use the largest available forward horizon $h_{it} \in \{3, \dots, 10\}$ for which $\log W_{i,t+h}$ is observed. Under per-unit-time stationarity of log growth, mixing horizons across firms leaves the tilted estimator consistent for the same object. The floor $h \geq 3$ also separates the direct estimator from the Brownian one, which uses one-year moments: if the direct estimator also used $h = 1$, it would coincide with $\mu_t + \zeta\sigma_t^2$ under an exact Brownian specification, and agreement between the two would be near-tautological rather than informative.

Reduced-form speed proxies. On the top 100 of each annual cross-section I also compute two model-free speed proxies. *Recent speed* is the mean of backward five-year log growth, $(\log W_{it} - \log W_{i,t-5})/5$, which asks how fast currently top-100 units climbed over the past five years. *Lifetime speed* is the mean of $\log(W_{it}/\underline{W})/\text{age}_{it}$, where $\underline{W}$ is one million 2020 dollars and age is years beyond 21 for Forbes individuals and firm age for CRSP.[4] Lifetime speed has a direct theoretical interpretation: Proposition 2 states that successful post-shock entrants observed at log wealth x are asymptotically of age $x/\Lambda'(\zeta)$, so $x/\text{age} \to \Lambda'(\zeta)$ in probability as $x \to \infty$. Under stationarity, recent and lifetime speed should therefore both approximate $\Lambda'(\zeta)$ and be of comparable magnitude.

Table 3: Empirical transition speed in the Forbes 400 and CRSP

					$\widehat{\Lambda}'(\zeta)$	
Sample	Recent speed	Lifetime speed	μ	σ	Brownian	Direct
Forbes 400 ($\zeta = 1.5$)	5.6%	23.9%	-0.8%	25.1%	9.7%	9.7%
CRSP ($\zeta = 1.1$)	4.1%	5.0%	4.5%	32.2%	17.1%	16.9%

Notes: All entries are averages of yearly moments, reported in annual percentage terms. Recent speed is the mean backward five-year log-growth rate among the top 100 units in each year. Lifetime speed is log wealth (above a one-million-dollar threshold in 2020 dollars) divided by age, averaged among the top 100 using post-2010 years; age is years beyond 21 for Forbes individuals and firm age for CRSP. μ and σ are the sample mean and standard deviation of winsorized one-year forward log growth on the top 400 (Forbes) or top 1000 (CRSP). Brownian $\widehat{\Lambda}'(\zeta) = \mu + \zeta\sigma^2$. Direct is the exponentially tilted mean of forward growth: in Forbes, $\overline{\tilde{g}\, e^{\zeta\tilde{g}}}/\overline{e^{\zeta\tilde{g}}}$; in CRSP, $\overline{\tilde{g}^{(h)} e^{\zeta h \tilde{g}^{(h)}}}/\overline{e^{\zeta h \tilde{g}^{(h)}}}$ using the first available horizon $h \in \{3, \dots, 10\}$.

The Brownian approximation is accurate. The Brownian and tilted-growth estimates are extremely close in both samples: 9.7% versus 9.7% in Forbes, and 16.9% versus 17.1% in CRSP. The diffusion approximation therefore captures most of the variation that these tilted growth moments are picking up.

The estimates sit between the two speed proxies. The direct and Brownian estimates both lie strictly between recent speed and lifetime speed: 5.6% < 9.7% < 23.9% in Forbes, and 4.1% < 17.1% < 5.0% in CRSP. So the theoretical speed is not disciplined by only one of the two proxies; it falls in a sensible range relative to both.

[4] Lifetime speed is computed on post-2010 years for both samples, where age coverage is adequate. Because the five-year backward lag is unavailable at the start of the Forbes sample, I also set recent speed to missing before 1988.

Recent versus lifetime: a finite-horizon gap. Lifetime speed substantially exceeds recent speed in both samples. Under stationarity and the asymptotics of Proposition 2, the two should be of comparable magnitude. The gap is consistent with either top units having accumulated faster earlier in life than over the most recent five years, or having entered the sample already above the normalized one-million-dollar threshold, so that $\log(W/\underline{W})/\text{age}$ overstates the asymptotic speed. These proxies should therefore be read as reduced-form consistency checks rather than as structural estimates of $\Lambda'(\zeta)$.

4 Conclusion

This paper shows that the transition of tail inequality after a permanent shock is organized by a single object: the slope of the scaled cumulant generating function at the new Pareto exponent, $\Lambda'(\zeta)$. Three questions that have often been treated separately in the literature — how fast the new exponent becomes visible at higher incomes, how old the successful entrants at the top typically are, and how sensitive the stationary tail is to permanent growth shifts — all have the same answer.

A Proofs

Lemma 1 (Convex-duality identities). *Fix $c > 0$ such that Λ is finite and differentiable at c, and let $v_c := \Lambda'(c)$.*

If $\Lambda(c) = 0$, then for every $\beta > 0$,

$$\inf_{0<\alpha\leq 1} \alpha\Lambda^*\left(\frac{\beta}{\alpha}\right) = \begin{cases} c\beta & \text{if } \beta \leq v_c, \\ \Lambda^*(\beta) & \text{if } \beta \geq v_c. \end{cases} \tag{A1}$$

If $\beta < v_c$, the minimizer is unique and equal to $\alpha^* = \beta/v_c$.

For every $\beta > 0$,

$$\inf_{\nu\leq\beta} \Big\{ c(\beta - \nu) + \Lambda^*(\nu) \Big\} = \begin{cases} \Lambda^*(\beta) & \text{if } \beta \leq v_c, \\ c\beta - \Lambda(c) & \text{if } \beta \geq v_c. \end{cases} \tag{A2}$$

Proof. The map $\alpha \mapsto \alpha\Lambda^*(\beta/\alpha)$ is the perspective of the convex function Λ^*, hence convex on $(0, \infty)$. If $\Lambda(c) = 0$, Fenchel's inequality gives

$$\alpha\Lambda^*\left(\frac{\beta}{\alpha}\right) \geq \alpha\left(c\frac{\beta}{\alpha} - \Lambda(c)\right) = c\beta.$$

Since Λ is differentiable at c, equality holds if and only if $\beta/\alpha = \Lambda'(c) = v_c$. Hence, if $\beta \leq v_c$, the feasible point $\alpha^* = \beta/v_c \in (0, 1]$ attains the lower bound, proving (A1); if $\beta < v_c$, uniqueness follows from the equality condition. If instead $\beta > v_c$, the global minimizer $\alpha^* = \beta/v_c$ lies to the right of 1, so convexity implies that the minimum over $(0, 1]$ is attained at the boundary point $\alpha = 1$, which gives $\Lambda^*(\beta)$.

Next, define $g(\nu) := c(\beta - \nu) + \Lambda^*(\nu)$. The map g is convex, and Fenchel's inequality implies

$$g(\nu) \geq c\beta - \Lambda(c)$$

for every ν, with equality if and only if $\nu = \Lambda'(c) = v_c$. Thus v_c is the global minimizer of g. Restricting to the feasible set $\{\nu \leq \beta\}$ proves (A2): if $\beta \geq v_c$, the constrained minimizer remains v_c and the value is $c\beta - \Lambda(c)$; if $\beta \leq v_c$, convexity implies that the constrained minimum is attained at the boundary point $\nu = \beta$, yielding $\Lambda^*(\beta)$. ☐

A.1 Proof of Proposition 1

Proof of Proposition 1. Fix $\beta > 0$ and $\varepsilon > 0$. Conditioning on the age of the agent at time t gives

$$\mathbb{P}\left(\left|\frac{x_t}{t} - \beta\right| < \varepsilon\right) = e^{-\delta t}\,\mathbb{P}\left(\left|\frac{x_0 + a_t}{t} - \beta\right| < \varepsilon\right) + \int_0^t \delta e^{-\delta s}\,\mathbb{P}\left(\left|\frac{Y + a_s}{t} - \beta\right| < \varepsilon\right) ds, \tag{A3}$$

where Y is a reinjection draw. The first term is the contribution of agents who were already alive at $t = 0$ and survive until t; the integral is the contribution of agents born after time 0.

Because the reinjection distribution has thinner-than-exponential tails, Y does not affect the exponential rate. Let $\tilde{\Lambda}(\xi) := \Lambda(\xi) + \delta = \lim_{t\to\infty} t^{-1} \log \mathbb{E}[e^{\xi a_t}]$ denote the gross cumulant generating function of a_t, so that $\tilde{\Lambda}^* = \Lambda^* - \delta$. Since $\tilde{\Lambda}$ is finite on $\mathbb{R}$ and convex, the Gärtner–Ellis theorem (e.g., Dembo and Zeitouni, 2010, Thm. 2.3.6) yields an exponential upper bound for a_s/s with rate $\tilde{\Lambda}^*$ at every point, and a matching lower bound at every point of the form $\gamma = \Lambda'(\xi)$ where Λ is differentiable at ξ. By Lemma 1, the minimizer of $\alpha\Lambda^*(\beta/\alpha)$ places the ratio β/α at $\Lambda'(\zeta)$, which is such a point (take

$\xi = \zeta$, a differentiability point by Assumption 2), so the lower bound is available where it is needed. Writing $s = \alpha t$ in the second term of (A3), the survival factor $e^{-\delta s}$ combines with $\tilde{\Lambda}^*$ to produce Λ^* at the level of the exponential rate, and Laplace's principle gives

$$\lim_{\varepsilon \downarrow 0} \lim_{t\to\infty} \frac{1}{t} \log \int_0^t \delta e^{-\delta s}\, \mathbb{P}\left(\left| \frac{Y + a_s}{t} - \beta \right| < \varepsilon \right) ds = -I_{\text{new}}(\beta), \tag{A4}$$

with

$$I_{\text{new}}(\beta) = \inf_{0<\alpha\le 1} \alpha \Lambda^*(\beta/\alpha). \tag{A5}$$

By Lemma 1 with $c = \zeta$,

$$I_{\text{new}}(\beta) = \begin{cases} \zeta\beta & \text{if } \beta \le \Lambda'(\zeta), \\ \Lambda^*(\beta) & \text{if } \beta \ge \Lambda'(\zeta). \end{cases} \tag{A6}$$

For the first term in (A3), I first extract the local-window rate of the initial condition from the upper-tail assumption. For every fixed $\gamma > 0$ and every $\rho \in (0, \gamma)$,

$$\mathbb{P}\Big(\frac{x_0}{t} \in (\gamma - \rho, \gamma + \rho) \Big) = \mathbb{P}(x_0 \ge (\gamma - \rho)t) - \mathbb{P}(x_0 \ge (\gamma + \rho)t).$$

By Assumption 3, the two terms on the right have exponential rates $-\zeta_0(\gamma - \rho)$ and $-\zeta_0(\gamma + \rho)$, respectively, so the second is exponentially negligible relative to the first. Therefore

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{P}\Big(\frac{x_0}{t} \in (\gamma - \rho, \gamma + \rho) \Big) = -\zeta_0(\gamma - \rho),$$

and letting $\rho \downarrow 0$ yields the local-window rate $-\zeta_0\gamma$. At $\gamma = 0$, the same rate is 0 because $x_0/t \to 0$ almost surely. This one-sided rate is sufficient here because the convolution below only uses $\gamma = \beta - \nu \ge 0$. Independence of x_0 and a_t therefore implies, again by Laplace's principle (with the survival factor $e^{-\delta t}$ absorbed into Λ^*),

$$I_{\text{old}}(\beta) = \inf_{\nu \le \beta} \Big\{ \zeta_0(\beta - \nu) + \Lambda^*(\nu) \Big\}. \tag{A7}$$

By Lemma 1 with $c = \zeta_0$,

$$I_{\text{old}}(\beta) = \begin{cases} \Lambda^*(\beta) & \text{if } \beta \le \Lambda'(\zeta_0), \\ \zeta_0\beta - \Lambda(\zeta_0) & \text{if } \beta \ge \Lambda'(\zeta_0). \end{cases} \tag{A8}$$

Finally, the two contributions in (A3) add, so the exponential rate of the sum is

$$I(\beta) = \min\{ I_{\text{new}}(\beta), I_{\text{old}}(\beta) \}. \tag{A9}$$

If $\zeta_0 \ge \zeta$, then $\Lambda'(\zeta_0) \ge \Lambda'(\zeta)$ and comparing (A6)–(A8) yields the three-region formula in the statement: for $\beta \le \Lambda'(\zeta)$ the newborn term dominates, for $\beta \in [\Lambda'(\zeta), \Lambda'(\zeta_0)]$ the two rates coincide and equal $\Lambda^*(\beta)$, and for $\beta \ge \Lambda'(\zeta_0)$ the initial-cohort term dominates. □

A.2 Proof of Proposition 2

Proof of Proposition 2. Since births are governed by a Poisson process with rate δ, the event $\{B_t,\ A_t/t \in (\alpha-\eta, \alpha+\eta)\}$ contributes the survival factor $e^{-\delta\alpha t}$, i.e. has exponential rate $-\delta\alpha$. Conditional on this event, $x_t \overset{d}{=} Y + a_{\alpha t}$ up to an $o(t)$ correction, where Y is an independent reinjection draw with thinner-than-exponential tails. By the same Gärtner–Ellis argument as in the proof of Proposition 1, which only uses the LDP at $\gamma = \Lambda'(\zeta)$ and relies on differentiability of Λ at ζ, the conditional rate for $\{x_t/t \in (\beta-\eta, \beta+\eta)\}$ is $-\alpha\tilde{\Lambda}^*(\beta/\alpha)$, where $\tilde{\Lambda}^* = \Lambda^* - \delta$. Adding survival and conditional rates, the $\alpha\delta$ terms cancel and the joint large-deviation cost of observing age αt and log income βt is

$$J_\beta(\alpha) := \alpha\Lambda^*\left(\frac{\beta}{\alpha}\right).$$

More precisely, for each fixed $\alpha \in (0,1]$,

$$\lim_{\eta\downarrow 0}\lim_{t\to\infty}\frac{1}{t}\log\mathbb{P}\left(B_t,\ \frac{A_t}{t}\in(\alpha-\eta,\alpha+\eta),\ \frac{x_t}{t}\in(\beta-\eta,\beta+\eta)\right) = -J_\beta(\alpha).$$

By Lemma 1 with $c = \zeta$, the unique minimizer of J_β on $(0,1]$ is

$$\alpha_\beta = \frac{\beta}{\Lambda'(\zeta)} \in (0,1),$$

and

$$\inf_{\alpha\in(0,1]} J_\beta(\alpha) = J_\beta(\alpha_\beta) = \zeta\beta.$$

This identifies the leading contribution of the new cohort. By Proposition 1,

$$\lim_{\eta\downarrow 0}\lim_{t\to\infty}\frac{1}{t}\log\mathbb{P}\left(\frac{x_t}{t}\in(\beta-\eta,\beta+\eta)\right) = -\zeta\beta.$$

At the same time, the contribution of individuals already alive at date 0 has strictly larger cost when $\beta < \Lambda'(\zeta)$; indeed its rate is

$$\inf_{\nu\le\beta}\{\zeta_0(\beta-\nu) + \Lambda^*(\nu)\} = \Lambda^*(\beta) > \zeta\beta,$$

where the equality is Lemma 1 with $c = \zeta_0$, and the strict inequality follows because Fenchel equality at $\xi = \zeta$ holds only at $\beta = \Lambda'(\zeta)$. This proves (i).

Fix now $\varepsilon > 0$ and define

$$F_\varepsilon := \{\alpha \in (0,1] : |\alpha - \alpha_\beta| \ge \varepsilon\}.$$

Since α_β is the unique minimizer of J_β,

$$\inf_{\alpha\in F_\varepsilon} J_\beta(\alpha) > \zeta\beta.$$

Therefore,

$$\limsup_{\eta\downarrow 0}\limsup_{t\to\infty}\frac{1}{t}\log\mathbb{P}\left(B_t,\ \frac{A_t}{t}\in F_\varepsilon,\ \frac{x_t}{t}\in(\beta-\eta,\beta+\eta)\right) < -\zeta\beta.$$

Comparing with the denominator, whose rate is exactly $-\zeta\beta$, yields age concentration. Since, under the same conditioning, $x_t/t \to \beta$ by construction and $A_t/t \to \alpha_\beta$ in probability with $\alpha_\beta = \beta/\Lambda'(\zeta)$, it also follows that

$$\frac{x_t}{A_t} \to \frac{\beta}{\alpha_\beta} = \Lambda'(\zeta)$$

in probability. This proves (ii). $\square$

A.3 Proof of Proposition 3

Proof of Proposition 3. Fix θ_0 and write

$$\zeta_\star := \zeta(\theta_0)$$

(I use $\zeta_\star$ rather than ζ_0 to avoid collision with the initial-exponent notation of the main text). By assumption,

$$\Lambda(\zeta_\star, \theta_0) = 0 \qquad \text{and} \qquad \zeta_\star > 0.$$

Step 1: show that $\Lambda'(\zeta_\star, \theta_0) > 0$. Since $\xi \mapsto \Lambda(\xi, \theta_0)$ is convex, its derivative is nondecreasing. Suppose, for contradiction, that

$$\Lambda'(\zeta_\star, \theta_0) \leq 0.$$

Then, for every $\xi \in [0, \zeta_\star]$, monotonicity of the derivative gives

$$\Lambda'(\xi, \theta_0) \leq \Lambda'(\zeta_\star, \theta_0) \leq 0.$$

Hence $\xi \mapsto \Lambda(\xi, \theta_0)$ is nonincreasing on $[0, \zeta_\star]$, which implies

$$0 = \Lambda(\zeta_\star, \theta_0) \leq \Lambda(0, \theta_0) < 0.$$

This is impossible. Therefore

$$\Lambda'(\zeta_\star, \theta_0) > 0.$$

Step 2: differentiate the implicit equation defining the tail exponent. Define

$$F(\xi, \theta) := \Lambda(\xi, \theta).$$

By Step 1, the partial derivative of F with respect to ξ is nonzero at $(\zeta_\star, \theta_0)$. Since F is continuously differentiable, the implicit function theorem implies that $\theta \mapsto \zeta(\theta)$ is differentiable in a neighborhood of θ_0. Differentiating

$$F(\zeta(\theta), \theta) = 0$$

with respect to θ gives

$$\partial_\xi F(\zeta(\theta), \theta)\zeta'(\theta) + \partial_\theta F(\zeta(\theta), \theta) = 0.$$

Returning to Λ, this becomes

$$\Lambda'(\zeta(\theta), \theta)\zeta'(\theta) + \partial_\theta \Lambda(\zeta(\theta), \theta) = 0.$$

Evaluating at $\theta = \theta_0$ and rearranging yields

$$\zeta'(\theta_0) = -\frac{\partial_\theta \Lambda(\zeta(\theta_0), \theta_0)}{\Lambda'(\zeta(\theta_0), \theta_0)}.$$

This proves the first formula.

Step 3: specialize to an additive growth shift. Now suppose

$$\Lambda(\xi, \theta) = \Lambda_0(\xi) + \theta\xi.$$

Then

$$\partial_\theta \Lambda(\xi, \theta) = \xi$$

for every (ξ, θ). Plugging this into the formula from Step 2 gives

$$\zeta'(\theta_0) = -\frac{\zeta(\theta_0)}{\Lambda'(\zeta(\theta_0), \theta_0)}.$$

Dividing both sides by $\zeta(\theta_0) > 0$ yields

$$\partial_\theta \log \zeta(\theta_0) = \frac{\zeta'(\theta_0)}{\zeta(\theta_0)} = -\frac{1}{\Lambda'(\zeta(\theta_0), \theta_0)}.$$

Finally,

$$\partial_\theta \log \frac{1}{\zeta(\theta_0)} = -\partial_\theta \log \zeta(\theta_0) = \frac{1}{\Lambda'(\zeta(\theta_0), \theta_0)}.$$

This proves the additive-shift formulas. □

B Local Pareto tail

Define, for fixed $h > 0$,

$$D_t(\beta, h, \varepsilon) := -\frac{1}{ht} \log \frac{\mathbb{P}\left(\left|\frac{x_t}{t} - (\beta + h)\right| < \varepsilon\right)}{\mathbb{P}\left(\left|\frac{x_t}{t} - \beta\right| < \varepsilon\right)}.$$

Then Proposition 1 gives

$$\lim_{\varepsilon \downarrow 0} \lim_{t \to \infty} D_t(\beta, h, \varepsilon) = \frac{I(\beta + h) - I(\beta)}{h}.$$

Now let $h \downarrow 0$. Wherever I is differentiable,

$$\lim_{h \downarrow 0} \lim_{\varepsilon \downarrow 0} \lim_{t \to \infty} D_t(\beta, h, \varepsilon) = I'(\beta).$$

Since the rate function is

$$I(\beta) = \begin{cases} \zeta\beta, \\ \Lambda^*(\beta), \\ \zeta_0\beta - \Lambda(\zeta_0), \end{cases}$$

its derivative is

$$I'(\beta) = \begin{cases} \zeta & \beta < \Lambda'(\zeta), \\ (\Lambda')^{-1}(\beta) & \Lambda'(\zeta) < \beta < \Lambda'(\zeta_0), \\ \zeta_0 & \beta > \Lambda'(\zeta_0). \end{cases}$$

B.1 Exponential tilting interpretation

The convex-duality formulas above admit a standard large-deviation interpretation. Conditioning on the rare event $x_t \approx \beta t$ selects growth histories that are atypical under the original law but typical under an exponentially tilted law. Under a tilt with parameter ξ, the selected paths move at typical speed $\Lambda'(\xi)$. Hence, in the transition region, the local Pareto slope is the unique tilt $\xi = (\Lambda')^{-1}(\beta)$ that makes speed β typical. For $\beta < \Lambda'(\zeta)$, the optimizer is pinned at the boundary $\xi = \zeta$, which is why the new Pareto exponent appears immediately below the moving cutoff. This is the same change-of-measure logic emphasized in the large-deviation literature; see especially Touchette (2009) on exponential tilting and Touchette (2018) on tilted generators for Markov processes.

The same idea can be seen in a more concrete one-step calculation in the i.i.d. case. Fix a horizon $h > 0$, let $A_h = e^{a_h}$ denote the gross income ratio over that horizon, and write g for the stationary density. In the Pareto tail we have heuristically

$$\frac{g(x - a_h)}{g(x)} \approx e^{\zeta a_h} = A_h^{\zeta},$$

so conditioning on being observed around a large income level x reweights the last growth shock by the factor A_h^{ζ}. The selected short-run log-growth rate is therefore the ζ-tilted mean

$$\frac{\mathbb{E}[\log A_h \, A_h^{\zeta}]}{h \, \mathbb{E}[A_h^{\zeta}]} = \frac{d}{d\xi} \left(\frac{1}{h} \log \mathbb{E}[A_h^{\xi}] \right) \Big|_{\xi=\zeta}.$$

When a_t has stationary independent increments, the term in parentheses is the gross SCGF, so this selected growth rate is exactly $\Lambda'(\zeta)$ because the death term drops out of the derivative. In that sense, $\Lambda'(\zeta)$ is the average short-run growth rate of the histories selected into the far right tail.

C Ergodic Case

In the main text, I obtained a stationary distribution from a quasi-random walk a_t using death. Another way to obtain a stationary distribution would be to add a small positive force at low levels of x. This "ergodic" case can be treated with very similar methods. Throughout this section I take $\delta = 0$, so that the SCGF reduces to the gross Laplace exponent, $\Lambda(\xi) = \lim_{t\to\infty} t^{-1} \log \mathbb{E}[e^{\xi a_t}]$, and in particular $\Lambda(0) = 0$.

Assumption 4 (Strengthened ergodic case)**.** Assume that:

(i) Income follows the continuous-time analogue of a Kesten process:

$$e^{x_t} = e^{x_0 + a_t} + \int_0^t e^{a_t - a_s} b_s \, ds,$$

where b_s is deterministic and bounded below and above by positive constants.

(ii) The rescaled paths

$$A_t(s) := \frac{a_{ts}}{t}, \qquad s \in [0, 1],$$

satisfy an LDP on $C([0, 1])$ with good rate function

$$J(\phi) = \begin{cases} \int_0^1 \Lambda^*(\dot{\phi}(u)) \, du & \text{if } \phi \in AC([0, 1]) \text{ and } \phi(0) = 0, \\ +\infty & \text{otherwise.} \end{cases}$$

(iii) x_0 is independent of a and has a local Pareto tail with exponent ζ_0, i.e. for every $\gamma > 0$ and every fixed $\varepsilon > 0$,

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{P}\Big(\frac{x_0}{t} \in (\gamma - \varepsilon, \gamma + \varepsilon) \Big) = -\zeta_0 \gamma.$$

(iv) There exists a unique $\zeta > 0$ such that $\Lambda(\zeta) = 0$, and Λ is differentiable at 0 and at ζ.

Since $\Lambda(0) = 0$ and Λ has a unique positive zero at ζ, convexity and differentiability at 0 imply $\Lambda'(0) < 0$. Hence the typical drift of a_t is negative, which is exactly the ergodic case of interest.

Proposition 4. *Assume Assumptions 2, 3, and 4, and assume $\zeta_0 \geq \zeta$, where $\zeta > 0$ is the unique positive solution to $\Lambda(\zeta) = 0$. Then, for every $\beta > 0$,*

$$\lim_{\varepsilon \downarrow 0} \lim_{t \to \infty} \frac{1}{t} \log \mathbb{P}\left(\left| \frac{x_t}{t} - \beta \right| < \varepsilon \right) = -I(\beta), \tag{A10}$$

where

$$I(\beta) = \begin{cases} \zeta\beta & \text{if } \beta \in (0, \Lambda'(\zeta)], \\ \Lambda^*(\beta) & \text{if } \beta \in [\Lambda'(\zeta), \Lambda'(\zeta_0)], \\ \zeta_0\beta - \Lambda(\zeta_0) & \text{if } \beta \in [\Lambda'(\zeta_0), \infty). \end{cases} \tag{A11}$$

This is exactly the same formula as Proposition 1, with $\delta = 0$.

It also yields the right analogue of Proposition 2. In the ergodic case there are no literal post-shock entrants, because the lower-region term

$$Z_t = \int_0^t e^{a_{t-u}} b_u \, du$$

receives contributions continuously. But the proof below shows that, for $\beta < \Lambda'(\zeta)$, the minimizer of the "new" term is

$$\tau_\beta = \frac{\beta}{\Lambda'(\zeta)}.$$

So the lowest-cost way to generate $x_t \approx \beta t$ through the additive term is to amplify lower-region contributions over a horizon $\tau_\beta t$. Equivalently, the amount of time needed for mass flowing in from the lower region to reach log-income level x is again $x/\Lambda'(\zeta)$. Thus $\Lambda'(\zeta)$ still measures the speed at which the new local tail behavior becomes visible, even though the entrant-age interpretation no longer applies literally.

Proof of Proposition 4. Write

$$Z_t := \int_0^t e^{a_{t-u}} b_u \, du, \qquad Y_t := \frac{1}{t} \log Z_t, \qquad R_t := \frac{x_0 + a_t}{t}.$$

Then

$$\frac{x_t}{t} = \frac{1}{t} \log\left(e^{tY_t} + e^{tR_t}\right),$$

and therefore

$$\left| \frac{x_t}{t} - \max\{Y_t, R_t\} \right| \leq \frac{\log 2}{t}.$$

So it is enough to identify the local rate of $\max\{Y_t, R_t\}$.

Step 1: the "old" term. The same Laplace-principle argument as in Proposition 1 (with $\delta = 0$) gives

$$\lim_{\varepsilon \downarrow 0} \lim_{t \to \infty} \frac{1}{t} \log \mathbb{P}(R_t \in (\beta - \varepsilon, \beta + \varepsilon)) = -I_{\text{old}}(\beta), \tag{A12}$$

with

$$I_{\text{old}}(\beta) = \inf_{\nu \leq \beta} \left\{ \zeta_0(\beta - \nu) + \Lambda^*(\nu) \right\}. \tag{A13}$$

By Lemma 1 with $c = \zeta_0$, this gives

$$I_{\text{old}}(\beta) = \begin{cases} \Lambda^*(\beta) & \text{if } \beta \le \Lambda'(\zeta_0), \\ \zeta_0\beta - \Lambda(\zeta_0) & \text{if } \beta \ge \Lambda'(\zeta_0). \end{cases} \tag{A14}$$

Step 2: the "new" term. Define, for $\phi \in C([0,1])$,

$$F_t(\phi) := \frac{\log t}{t} + \frac{1}{t}\log\int_0^1 e^{t\phi(1-u)} b_{tu}\,du, \qquad F(\phi) := \sup_{s\in[0,1]} \phi(s). \tag{A15}$$

Then $Y_t = F_t(A_t)$. I claim that $F_t \to F$ uniformly on compact subsets of $C([0,1])$. Indeed, fix a compact $K \subset C([0,1])$ and $\eta > 0$. Since K is equicontinuous, there exists $\delta > 0$ such that

$$|\phi(s) - \phi(r)| \le \eta \qquad \text{whenever } |s-r| \le \delta,\ \phi \in K.$$

For every $\phi \in K$,

$$F_t(\phi) \le \sup_s \phi(s) + \frac{\log(t\bar{b})}{t}, \tag{A16}$$

$$F_t(\phi) \ge \sup_s \phi(s) - \eta + \frac{\log(\delta \underline{b})}{t}. \tag{A17}$$

Hence $\sup_{\phi\in K} |F_t(\phi) - F(\phi)| \to 0$. By the extended contraction principle, $Y_t = F_t(A_t)$ satisfies an LDP with rate

$$I_{\text{new}}(\beta) = \inf\{J(\phi) : \sup_{s\in[0,1]} \phi(s) = \beta\}. \tag{A18}$$

I now compute this infimum. Let ϕ be absolutely continuous with $\sup\phi = \beta$, and let

$$\tau := \inf\{s \in [0,1] : \phi(s) = \beta\}.$$

Since $\phi(0) = 0$,

$$\beta = \int_0^\tau \dot{\phi}(u)\,du.$$

By Jensen's inequality,

$$J(\phi) \ge \int_0^\tau \Lambda^*(\dot{\phi}(u))\,du \ge \tau\Lambda^*\left(\frac{\beta}{\tau}\right). \tag{A19}$$

Therefore

$$I_{\text{new}}(\beta) \ge \inf_{0<\tau\le 1} \tau\Lambda^*\left(\frac{\beta}{\tau}\right). \tag{A20}$$

Conversely, let $m := \Lambda'(0) < 0$. For every $\tau \in (0,1]$, define

$$\phi_\tau(s) = \begin{cases} \frac{\beta}{\tau}s & \text{if } 0 \le s \le \tau, \\ \beta + m(s-\tau) & \text{if } \tau \le s \le 1. \end{cases}$$

Then $\sup \phi_\tau = \beta$, and since $\Lambda^*(m) = 0$,

$$J(\phi_\tau) = \tau \Lambda^* \left(\frac{\beta}{\tau} \right).$$

Hence

$$I_{\text{new}}(\beta) = \inf_{0<\tau\leq 1} \tau \Lambda^* \left(\frac{\beta}{\tau} \right). \tag{A21}$$

By Lemma 1 with $c = \zeta$,

$$I_{\text{new}}(\beta) = \begin{cases} \zeta\beta & \text{if } \beta \leq \Lambda'(\zeta), \\ \Lambda^*(\beta) & \text{if } \beta \geq \Lambda'(\zeta). \end{cases} \tag{A22}$$

Step 3: combine the two terms. Let $M_t := \max\{Y_t, R_t\}$. Since $|x_t/t - M_t| \leq (\log 2)/t$, x_t/t and M_t have the same local rate. For the upper bound, note that

$$\{M_t \in (\beta - \varepsilon, \beta + \varepsilon)\} \subseteq \{Y_t \in (\beta - \varepsilon, \beta + \varepsilon)\} \cup \{R_t \in (\beta - \varepsilon, \beta + \varepsilon)\},$$

so

$$\limsup_{\varepsilon \downarrow 0} \limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}(M_t \in (\beta - \varepsilon, \beta + \varepsilon)) \leq -\min\{I_{\text{new}}(\beta), I_{\text{old}}(\beta)\}. \tag{A23}$$

For the lower bound, choose the cheaper of the two mechanisms. If $I_{\text{new}}(\beta) \leq I_{\text{old}}(\beta)$, use a path neighborhood of a minimizer ϕ_τ in (A21); this neighborhood forces $Y_t \in (\beta - \varepsilon, \beta + \varepsilon)$ and also keeps $R_t \leq \beta + \varepsilon$, hence $M_t \in (\beta - \varepsilon, \beta + \varepsilon)$ with the same exponential cost. If $I_{\text{old}}(\beta) \leq I_{\text{new}}(\beta)$, use the same construction as in Proposition 1: choose x_0/t near $\beta - \nu$ and a path of a with terminal slope ν, where ν minimizes $I_{\text{old}}(\beta)$. Taking the path linear keeps its running maximum below β, so again $M_t \in (\beta - \varepsilon, \beta + \varepsilon)$ with exponential cost $I_{\text{old}}(\beta)$. Therefore

$$\lim_{\varepsilon \downarrow 0} \lim_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left(\frac{x_t}{t} \in (\beta - \varepsilon, \beta + \varepsilon) \right) = -\min\{I_{\text{new}}(\beta), I_{\text{old}}(\beta)\}. \tag{A24}$$

Finally, if $\zeta_0 \geq \zeta$, then $\Lambda'(\zeta_0) \geq \Lambda'(\zeta)$, and comparing (A14) and (A22) yields the three-region formula in the statement. □

D Relationship to Gabaix et al. (2015)

Gabaix et al. (2016) study the same transition problem: the cross-sectional distribution $p(x,t)$ converges from an initial Pareto steady state (exponent ζ_0) to a new one (exponent ζ). They propose two measures of the speed of this transition.

Average speed of convergence. Their first measure (Proposition 1) is the rate λ at which the L^1 distance $\|p(\cdot,t) - p_\infty\| := \int |p(x,t) - p_\infty(x)|\, dx$ decays exponentially:

$$\lambda := -\lim_{t\to\infty} \frac{1}{t} \log \|p(\cdot,t) - p_\infty\|.$$

They show that λ equals the spectral gap (i.e., the second eigenvalue) of the Kolmogorov Forward operator $\mathcal{A}^*$. In the baseline model with death rate δ and no lower bound on income, $\lambda = \delta$. With a reflecting barrier and $\mu < 0$, $\lambda = \frac{\mu^2}{2\sigma^2} + \delta$. For the GBM calibration in their Section 4.3 ($\delta = 1/30$), this implies a half-life $\log(2)/\lambda \approx 20$ years.

Speed of convergence in the tail. Their second measure (Proposition 3) refines the first by placing exponential weight on the tail. Define the ξ-weighted L^1 norm:

$$\|p(\cdot, t) - p_\infty\|_\xi := \int |p(x,t) - p_\infty(x)| \, e^{-\xi x} \, dx.$$

They show that this norm decays at rate $\lambda(\xi) := -\lim_{t\to\infty} \frac{1}{t} \log \|p(\cdot, t) - p_\infty\|_\xi$, which, in the baseline model without jumps, equals

$$\lambda(\xi) = \xi\mu - \xi^2 \frac{\sigma^2}{2} + \delta. \tag{A25}$$

Since $\xi \leq 0$ puts weight on the right tail, $\lambda(\xi)$ is decreasing in $|\xi|$: convergence is slower the further into the tail one looks. This is the content of their Figure 3.

Connection to the SCGF. Without a lower bound on income, their Laplace-transform formula implies

$$\lambda(\xi) = -\Lambda(-\xi). \tag{A26}$$

With a reflecting barrier, the identity is replaced by a piecewise formula, so (A26) should be read as the clean no-lower-bound benchmark. In particular, $\lambda(0) = -\Lambda(0) = \delta$ recovers the average speed. As $\xi \to -\zeta$ (i.e., as one puts more and more weight on the right tail), the convergence rate vanishes: $\lambda(-\zeta) = -\Lambda(\zeta) = 0$. The rate at which it vanishes is

$$\left.\frac{d\lambda}{d\xi}\right|_{\xi=-\zeta} = \Lambda'(\zeta),$$

which is precisely the transition speed in Proposition 1. So my transition speed can be read as the slope of their tail-specific temporal rate exactly at the point where it hits zero.

Comparison. The two approaches measure different aspects of the transition. The spectral gap λ is a *temporal rate* (units: 1/time); it describes how fast a global distance metric decays. The transition speed $\Lambda'(\zeta)$ is a *spatial speed* (units: log-income/time); it describes which parts of the income distribution have already switched to the new local slope by time t.

Shape of the transitory density. The convergence rate $\lambda(\xi)$ tells us how fast the distribution converges, but not what it looks like along the way. Proposition 1 delivers a complete spatial profile: the local Pareto exponent equals ζ for $x < \Lambda'(\zeta)t$, transitions continuously for $\Lambda'(\zeta)t < x < \Lambda'(\zeta_0)t$, and equals ζ_0 for $x > \Lambda'(\zeta_0)t$.

Mobility interpretation. The transition speed $\Lambda'(\zeta)$ has a direct economic interpretation (Proposition 2): it equals the typical speed at which successful new entrants climb through the distribution. The spectral gap λ has no analogous individual-level counterpart.

Generality. Because the SCGF is defined for general processes satisfying Assumption 2, the transition speed $\Lambda'(\zeta)$ can be computed for models with jumps, type heterogeneity, or other departures from GBM, without requiring closed-form solutions to the Kolmogorov Forward equation.

Insensitivity to the lower bound. The transition speed $\Lambda'(\zeta)$ depends only on the individual-level income process (through Λ) and the tail exponent ζ. In particular, it is the same whether the stationary distribution is sustained by death and reinjection or by a reflecting barrier at $x = 0$. This is intuitive: the cutoff $\Lambda'(\zeta)t$ lies far from the lower bound and is driven entirely by the fastest-growing individuals, who never interact with the barrier. In contrast, the spectral gap λ of Gabaix et al. (2016) *does* depend on the lower bound: it equals δ without a barrier but $\frac{\mu^2}{2\sigma^2} + \delta$ with a reflecting barrier. This is because the spectral gap is a global average over the entire distribution, and the barrier creates additional churning near $x = 0$ that speeds up convergence there, pulling up the overall rate. Note that Gabaix et al.'s tail-specific measure $\lambda(\xi)$ shares this insensitivity for $\xi < \mu/\sigma^2$: in the deep-tail region, their formula is identical with or without barrier, consistent with the relationship (A26).

E Closed-form derivations

This section derives the closed-form expressions for the specific examples used in the numerical comparison of Section 3.1. Figure A1 plots the SCGF for four models, all calibrated to the same tail exponent ζ and death rate δ. The slope at ζ — the transition speed $\Lambda'(\zeta)$ — varies substantially across models. The Steindl model (pure drift) has the lowest speed. Adding diffusion (BM) increases the speed modestly. Jumps and type dependence can generate much faster transitions.

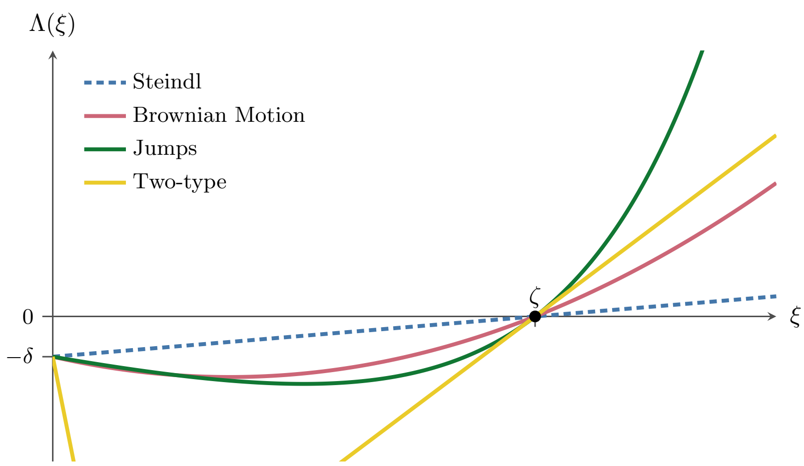


Figure A1: SCGF for four models with the same ζ and δ.

Notes: The slope at ζ is the transition speed $\Lambda'(\zeta)$. Parameters are as in Table 2.

E.1 Brownian motion

Suppose a_t is a Brownian motion with drift μ and volatility σ. The SCGF is

$$\Lambda(\xi) = \mu\xi + \frac{\sigma^2}{2}\xi^2 - \delta.$$

The tail exponent ζ solves the quadratic $\mu\zeta + \frac{\sigma^2}{2}\zeta^2 = \delta$, giving $\zeta = \left(-\mu + \sqrt{\mu^2 + 2\sigma^2\delta}\right)/\sigma^2$, and the transition speed is $\Lambda'(\zeta) = \mu + \sigma^2\zeta$.

To compute the Legendre transform $\Lambda^*(\beta) = \sup_\xi\{\xi\beta - \Lambda(\xi)\}$, the first-order condition $\beta = \Lambda'(\xi) = \mu + \sigma^2\xi$ gives $\xi^* = (\beta - \mu)/\sigma^2$. Substituting:

$$\Lambda^*(\beta) = \frac{(\beta-\mu)^2}{2\sigma^2} + \delta.$$

The rate function from Proposition 1 therefore takes the form

$$I(\beta) = \begin{cases} \zeta\beta & \text{if } \beta \le \mu + \sigma^2\zeta, \\ \dfrac{(\beta-\mu)^2}{2\sigma^2} + \delta & \text{if } \mu + \sigma^2\zeta \le \beta \le \mu + \sigma^2\zeta_0, \\ \zeta_0(\beta-\mu) - \dfrac{\sigma^2\zeta_0^2}{2} + \delta & \text{if } \beta \ge \mu + \sigma^2\zeta_0. \end{cases}$$

In the Brownian motion case, the transitory density is available in closed form, which provides an independent verification of the rate function. Gabaix et al. (2016) (Proposition 12) show that, without a lower bound on income,

$$p(x,t) = p_\infty(x) + e^{-\delta t}\,\mathbb{E}[p_0(x - G_t) - p_\infty(x - G_t)]\,, \tag{A27}$$

where $G_t := \mu t + \sigma Z_t$, Z_t is a standard Brownian motion, and p_0 and p_∞ are extended by 0 on $(-\infty, 0)$. I now recover $I(\beta)$ from this formula by applying Laplace's method to the Gaussian expectation.

Set $x = \beta t$ with $\beta > 0$. The Gaussian density of G_t satisfies

$$\frac{1}{t}\log\varphi_{\mu t,\sigma^2 t}(\gamma t) \longrightarrow -\frac{(\gamma-\mu)^2}{2\sigma^2}.$$

Writing $g = \gamma t$ and using $p_0(\beta t - g) = \zeta_0 e^{-\zeta_0(\beta t - g)}\mathbf{1}_{g \le \beta t}$, Varadhan's lemma applied to the expectation gives

$$\frac{1}{t}\log\mathbb{E}[p_0(\beta t - G_t)] \longrightarrow \sup_{\gamma\le\beta}\left\{-\zeta_0(\beta-\gamma) - \frac{(\gamma-\mu)^2}{2\sigma^2}\right\}.$$

The unconstrained maximizer is $\gamma^* = \mu + \sigma^2\zeta_0$. If $\beta \ge \mu + \sigma^2\zeta_0$ the constraint $\gamma \le \beta$ is slack, the maximum equals $-\zeta_0\beta + \zeta_0\mu + \sigma^2\zeta_0^2/2 = -\zeta_0\beta + \Lambda(\zeta_0) + \delta$, and adding the $e^{-\delta t}$ prefactor gives

$$\frac{1}{t}\log\left(e^{-\delta t}\mathbb{E}[p_0(\beta t - G_t)]\right) \longrightarrow -(\zeta_0\beta - \Lambda(\zeta_0)).$$

If instead $\beta < \mu + \sigma^2\zeta_0$, the constrained maximum is attained at the boundary $\gamma = \beta$, where the exponent equals $-(\beta-\mu)^2/(2\sigma^2)$, and the rate becomes $(\beta-\mu)^2/(2\sigma^2) + \delta$.

The same argument applied to $e^{-\delta t}\mathbb{E}[p_\infty(\beta t - G_t)]$ (replacing ζ_0 by ζ and using $\Lambda(\zeta) = 0$) gives unconstrained rate $\zeta\beta$ when $\beta \ge \mu + \sigma^2\zeta$ and constrained rate $(\beta-\mu)^2/(2\sigma^2) + \delta$ when $\beta < \mu + \sigma^2\zeta$. Finally, the pure $p_\infty(\beta t)$ term on the right-hand side of (A27) has rate $\zeta\beta$.

Collecting these contributions, the rate of $p(\beta t, t)$ is the minimum of the three:

$\beta \le \mu + \sigma^2\zeta$:	$p_\infty(\beta t)$ and $e^{-\delta t}\mathbb{E}[p_\infty]$ coincide and dominate; rate $\zeta\beta$.
$\mu + \sigma^2\zeta \le \beta \le \mu + \sigma^2\zeta_0$:	boundary saddle in $\mathbb{E}[p_0 - p_\infty]$; rate $\frac{(\beta-\mu)^2}{2\sigma^2} + \delta$.
$\beta \ge \mu + \sigma^2\zeta_0$:	interior saddle of $\mathbb{E}[p_0]$ dominates; rate $\zeta_0\beta - \Lambda(\zeta_0)$.

This recovers the rate function $I(\beta)$ stated above.

This Brownian case also allows a direct comparison with the calibration in Section 4.3 of Gabaix et al. (2016), where $\delta = 1/30$, σ rises from 0.1 to $\sqrt{0.025} \approx 0.158$, and μ is kept fixed at $\delta \times 0.39 - 0.01/(2 \times 0.39) \approx 0.00018$. Under the new parameters, the stationary tail exponent is $\zeta \approx 1.63$, so my closed-form transition speed is

$$\Lambda'(\zeta) = \mu + \sigma^2\zeta \approx 0.041.$$

That is, the new Pareto tail advances at about 4.1% log-income per year. Using the stationary Pareto approximation $x_p \approx \zeta^{-1}\log(1/p)$, the new exponent reaches the top 1% cutoff after about 69 years and the top 0.1% cutoff after about 104 years. This is consistent with their numerical Figure 4 and with the half-life calculations in their Figure 3; the difference is that here these magnitudes come directly from a closed-form transition speed rather than from simulated paths of moments or shares.

E.2 Compound Poisson process with fixed jump size

Suppose $a_t = \mu t + \sum_{i=1}^{N_t} G$ where N_t is a Poisson process with arrival rate ϕ and $G > 0$ is a fixed jump size. The SCGF is

$$\Lambda(\xi) = \mu\xi + \phi(e^{\xi G} - 1) - \delta.$$

The tail exponent ζ solves $\mu\zeta + \phi(e^{\zeta G} - 1) = \delta$, and the transition speed is $\Lambda'(\zeta) = \mu + \phi G e^{\zeta G}$.

To compute the Legendre transform, the first-order condition $\beta = \Lambda'(\xi) = \mu + \phi G e^{\xi G}$ gives $\xi^* = \frac{1}{G}\log\left(\frac{\beta-\mu}{\phi G}\right)$. Substituting:

$$\Lambda^*(\beta) = \frac{\beta-\mu}{G}\left(\log\frac{\beta-\mu}{\phi G} - 1\right) + \phi + \delta.$$

The rate function is therefore

$$I(\beta) = \begin{cases} \zeta\beta & \text{if } \beta \le \Lambda'(\zeta), \\ \dfrac{\beta-\mu}{G}\left(\log\dfrac{\beta-\mu}{\phi G} - 1\right) + \phi + \delta & \text{if } \Lambda'(\zeta) \le \beta \le \Lambda'(\zeta_0), \\ \zeta_0\beta - \Lambda(\zeta_0) & \text{if } \beta \ge \Lambda'(\zeta_0). \end{cases}$$

This example clarifies the difference between the comparative statics in Gabaix et al. (2016) and the one relevant here. Their weighted-L^1 rate $\lambda(\xi)$ asks how fast moments converge for a fixed stochastic process; varying jump intensity there also changes the stationary tail exponent ζ. Here the natural question is instead: holding fixed the stationary Pareto exponent, do more jumps make the tail adjust faster across income levels? Using the restriction $\Lambda(\zeta) = 0$ to eliminate μ, the transition speed can be written as

$$\Lambda'(\zeta) = \frac{\delta}{\zeta} + \phi\left(Ge^{\zeta G} - \frac{e^{\zeta G} - 1}{\zeta}\right),$$

which is increasing in ϕ for $G > 0$. So, among models with the same tail exponent, more jump-driven growth generates faster *spatial* transitions, even though Gabaix et al. (2016)'s *temporal* convergence rate need not increase.

E.3 Two-type model

Suppose agents are born in a high-growth state with drift μ_{H} and switch to a low-growth state with drift $\mu_{\mathrm{L}} < \mu_{\mathrm{H}}$ at rate τ, with no switching back (L is absorbing). The tilted generator is the triangular matrix

$$A(\xi) = \begin{pmatrix} \mu_{\mathrm{H}}\xi - \tau & 0 \\ \tau & \mu_{\mathrm{L}}\xi \end{pmatrix},$$

whose eigenvalues are the diagonal entries $\mu_{\mathrm{H}}\xi - \tau$ and $\mu_{\mathrm{L}}\xi$. The SCGF is therefore

$$\Lambda(\xi) = \max(\mu_{\mathrm{H}}\xi - \tau,\ \mu_{\mathrm{L}}\xi) - \delta.$$

For $\xi > \tau/(\mu_{\mathrm{H}} - \mu_{\mathrm{L}})$, the H-type eigenvalue dominates and $\Lambda(\xi) = \mu_{\mathrm{H}}\xi - \tau - \delta$. In this regime, the tail exponent solves $\mu_{\mathrm{H}}\zeta - \tau - \delta = 0$, giving

$$\zeta = \frac{\delta + \tau}{\mu_{\mathrm{H}}},$$

and the transition speed is $\Lambda'(\zeta) = \mu_{\mathrm{H}}$. The speed is therefore determined entirely by the drift of the high-growth type. Increasing μ_{H} simultaneously increases the speed and decreases the tail exponent (fattens the tail), while increasing τ (shorter high-growth spells) increases ζ without affecting the speed.

Since Λ is piecewise linear, the Legendre transform is

$$\Lambda^*(\beta) = \begin{cases} -\mu_{\mathrm{L}}\beta + \delta & \text{if } \beta \le \mu_{\mathrm{L}}, \\ +\infty & \text{if } \mu_{\mathrm{L}} < \beta < \mu_{\mathrm{H}}, \\ -\mu_{\mathrm{H}}\beta + \tau + \delta & \text{if } \beta \ge \mu_{\mathrm{H}}. \end{cases}$$

The fact that $\Lambda^* = +\infty$ for $\beta \in (\mu_{\mathrm{L}}, \mu_{\mathrm{H}})$ reflects the absence of diffusion: income grows at exactly μ_{H} or μ_{L}, so intermediate growth rates are impossible. The rate function is

$$I(\beta) = \begin{cases} \zeta\beta & \text{if } \beta \le \mu_{\mathrm{H}}, \\ \zeta_0\beta - \Lambda(\zeta_0) & \text{if } \beta \ge \mu_{\mathrm{H}}. \end{cases}$$

There is no intermediate region: the transition speed $\Lambda'(\zeta) = \mu_{\mathrm{H}}$ coincides with the only possible growth rate in the H-state.

References

Beare, Brendan K and Alexis Akira Toda, "Determination of Pareto exponents in economic models driven by Markov multiplicative processes," *Econometrica*, 2022, *90* (4), 1811–1833.

Chetrite, Raphaël and Hugo Touchette, "Nonequilibrium Markov processes conditioned on large deviations," *Annales Henri Poincaré*, 2015, *16* (9), 2005–2057.

Dembo, Amir and Ofer Zeitouni, *Large Deviations Techniques and Applications*, 2 ed., Berlin: Springer, 2010.

Gabaix, Xavier, "Power laws in economics and finance," *Annual Review of Economics*, 2009, *1*, 255–294.

_ , Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll, "The dynamics of inequality," *Econometrica*, 2016, *84* (6), 2071–2111.

Gomez, Matthieu and Emilien Gouin-Bonenfant, "Wealth inequality in a low rate environment," *Econometrica*, 2024, *92* (4), 1225–1260.

Hansen, Lars Peter and José A Scheinkman, "Long-term risk: An operator approach," *Econometrica*, 2009, *77* (1), 177–234.

Jones, Charles I., "Pareto and Piketty: The macroeconomics of top income and wealth inequality," *Journal of Economic Perspectives*, 2015, *29* (1), 29–46.

Kesten, Harry, "Random difference equations and renewal theory for products of random matrices," *Acta Mathematica*, 1973, *131*, 207–248.

Luttmer, Erzo GJ, "On the mechanics of firm growth," *The Review of Economic Studies*, 2011, *78* (3), 1042–1068.

Touchette, Hugo, "The large deviation approach to statistical mechanics," *Physics Reports*, 2009, *478* (1–3), 1–69.

__ , "Introduction to dynamical large deviations of Markov processes," *Physica A: Statistical Mechanics and its Applications*, 2018, *504*, 5–19.