# Tail behavior of geometrically stopped random growth processes and wealth distributions

Brendan K. Beare and Alexis Akira Toda

Department of Economics, University of California, San Diego

September 7, 2017

**Abstract**

Many empirical studies document power law behavior in size distributions of economic interest such as cities, firms, income, and wealth. One mechanism for generating such behavior combines iid Gaussian additive shocks to log-size with a geometric age distribution. We generalize this mechanism by allowing the shocks to be non-Gaussian (but light-tailed) and dependent upon a Markov state variable. Our main results provide sharp bounds on tail probabilities and simple analytical formulas for Pareto exponents. As an application, we characterize the tails of the wealth distribution in heterogeneous-agent models, and show that models with idiosyncratic endowment risk exhibit exponential tails, while those with investment risk exhibit Pareto tails.

# 1 Introduction

In this paper we are concerned with the tail behavior of geometric sums of the form

$$W_T = \sum_{t=1}^{T} X_t, \tag{1.1}$$

where $\{X_t\}_{t=1}^{\infty}$ is a sequence of random innovations and $T$ is a geometrically distributed random variable. The innovations $\{X_t\}_{t=1}^{\infty}$ may depend contemporaneously on a time homogeneous Markov state variable (i.e., $X_t$ is a "hidden Markov process"; see Definition 3.1 below). We may view $W_T$ in (1.1) as a geometrically stopped random walk with possibly serially dependent innovations. The main result of our paper shows that under quite general conditions the distribution of $W_T$ has exponential tails. Further, we provide a simple formula characterizing the tail decay rates. For instance, when $\{X_t\}_{t=1}^{\infty}$ is iid and $T$ has mean $1/p$, the upper tail exponent $\alpha$ of $W_T$ solves the equation

$$\mathrm{E}[\mathrm{e}^{\alpha X}] = \frac{1}{1-p}. \tag{1.2}$$

In more general settings, the left-hand side of (1.2) is replaced with the spectral radius of a matrix determined by the law of $\{X_t\}_{t=1}^{\infty}$.

Our problem is motivated by the observed characteristics of size distributions in economics and other fields, where variables of interest are often known to exhibit power law behavior in the upper tail. Such variables include city sizes (Gabaix, 1999; Reed, 2002; Giesen et al., 2010), firm sizes (Axtell, 2001; Luttmer, 2007), and household income (Pareto, 1896; Reed, 2003; Reed and Jorgensen, 2004; Toda, 2011, 2012), consumption (Toda and Walsh, 2015; Toda, 2016), and wealth (Klass et al., 2006; Vermeulen, 2017). A variety of explanations for the emergence of these power laws have been proposed in the economics literature.

One particular mechanism for generating power laws proposed by Reed (2001), and to some extent anticipated by Wold and Whittle (1957) and Simon and Bonini (1958), combines two main ingredients: Gibrat (1931)'s law of proportional growth and an exponential age distribution.[1] Suppose that the size $S_t$ of an individual unit at age $t \geq 0$ follows a geometric Brownian motion initialized at some fixed

[1]There is a large empirical literature documenting that Gibrat's law of proportional growth is a good first approximation; see Sutton (1997) and the references therein. The exponential age distribution has drawn much less attention but there is some evidence both for firms (Coad, 2010) and cities (Giesen and Suedekum, 2014).

$S_0$, so that

$$\mathrm{d}S_t = \mu S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}B_t, \tag{1.3}$$

where $B_t$ is a standard Brownian motion and $\mu$ and $\sigma$ are the drift and volatility parameters. If the distribution of the age $T$ of a unit randomly selected at a given point in time is exponential with parameter $\eta > 0$, so that it has probability density function (pdf) $f_T(t) = \eta \mathrm{e}^{-\eta t}$ for $t \geq 0$, then the size of a randomly selected unit is given by $S_T$, our geometric Brownian motion evaluated at an exponentially distributed time. Reed showed that the pdf of this quantity is given by

$$f_{S_T}(s) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta} S_0^{-\beta} s^{\beta-1} & \text{for } 0 \leq s < S_0 \\ \frac{\alpha\beta}{\alpha+\beta} S_0^{\alpha} s^{-\alpha-1} & \text{for } S_0 \leq s < \infty, \end{cases}$$

where $\zeta = \alpha, -\beta$ are the positive and negative roots of the quadratic equation

$$\frac{\sigma^2}{2}\zeta^2 + \left(\mu - \frac{\sigma^2}{2}\right)\zeta - \eta = 0.^{2} \tag{1.4}$$

He dubbed this distribution the *double Pareto* distribution. Reed's mechanism is notable in that it generates power law behavior in not only the upper tail of a distribution, but also the lower tail: we have

$$\lim_{s\uparrow\infty} As^{\alpha}\mathrm{P}(S_T > s) = 1 \quad \text{and} \quad \lim_{s\downarrow 0} Bs^{-\beta}\mathrm{P}(S_T < s) = 1 \tag{1.5}$$

for some positive constants $A$ and $B$. This is the case even though the distribution of a geometric Brownian motion evaluated at a fixed point in time is lognormal and therefore has tails decaying more rapidly than a power law. Reed's mechanism has recently been applied in economics to characterize the tail behavior of size distributions in heterogeneous-agent models in continuous-time with Brownian shocks.[3]

Does Reed's mechanism also generate power law tails when applied to more general stochastic processes? Given that power law distributions are empirically so common, and that realistic alternatives to Brownian motion may involve non-Gaussian, non-independent increments, it is natural to conjecture that more general random growth processes with geometric age distributions give rise to power

[2]Note that assuming $\mu > 0$ and letting $\sigma \to 0$, we obtain $\alpha = \frac{\eta}{\mu}$, which is exactly the result of Wold and Whittle (1957). (Their paper contains a typographical error in the formula for $\alpha$ on the first page, in which the numerator and the denominator are flipped.)

[3]See, for example, Toda (2014), Toda and Walsh (2015, 2017), Arkolakis (2016), Benhabib et al. (2016), and Gabaix et al. (2016).

law tails, or exponential tails after taking the logarithm. Some affirmative results in this direction are available in the physics literature for the geometric sum (1.1) in the special case where $\{X_t\}_{t=1}^{\infty}$ is iid. Manrubia and Zanette (1999) give a heuristic derivation of (1.2), which corresponds to equation (10) in their paper, and provide supporting evidence from numerical simulations. Reed and Hughes (2002) observe that the tails of their "killed discrete multiplicative process" are characterized by (1.2). In the economics literature, (1.2) appears in Proposition 5 of Nirei and Aoki (2016), who appeal to Manrubia and Zanette (1999) to characterize the tail exponent of the wealth distribution in a heterogeneous-agent model with iid shocks.

The main results of our paper, Theorems 3.1–3.4 below, provide a formal justification for the formula (1.2) in the iid case and extend it to a wider class of processes. The key to proving them is a Tauberian theorem due to Nakagawa (2007),[4] which we discuss in Section 2 and Appendix A. Nakagawa's theorem provides sharp bounds on the tail probabilities of a random variable whose Laplace transform has a pole at its abscissa of convergence. In the case of a simple pole, which is the relevant one for our purposes, those bounds depend on the residue of the pole and the distance to other singularities along the axis of convergence. To obtain the residue in the non-iid case we rely on the Perron-Frobenius theory of nonnegative irreducible matrices (Appendix B) and some results on simple poles of matrix pencil inverses (Appendix C). When the distribution of $X_t$ conditional on the current Markov state is not concentrated on an evenly spaced grid, $S = \mathrm{e}^{W_T}$ has Pareto tails as in (1.5). Without this condition, the limits in (1.5) may not exist, but we have sharp bounds for the corresponding limits superior and inferior. Those bounds imply (Remark 2.3) that the tail probabilities satisfy

$$\lim_{s\uparrow\infty} \frac{\log \mathrm{P}(S > s)}{\log s} = -\alpha \quad \text{and} \quad \lim_{s\downarrow 0} \frac{\log \mathrm{P}(S < s)}{\log s} = \beta,$$

which is a weaker property than (1.5).

As an application of our results, in Section 4 we characterize the tail behavior of the wealth distributions in heterogeneous-agent macroeconomic models. We consider two cases, economies with idiosyncratic endowment and investment risks,

[4] Reed and Hughes (2003, p. 588) suggested that Tauberian theorems may be useful to characterize tail probabilities in the related class of Galton-Watson branching processes. The Online Appendix of Gabaix et al. (2016) appeals to a Tauberian theorem of Mimica (2016, Corollary 1.4), slightly more general than a result of Nakagawa (2007, Theorem 3), to characterize the tail decay rate of the wealth distribution. We appeal to a sharper result (Theorem 2.1 below) essentially obtained by Nakagawa (2007, Theorem 5*).

both with Markov shocks. In the former case, we prove that the wealth distribution has exponential tails, which are thin. This theoretical result explains the numerical findings in the literature that the so-called Bewley (1983)-Huggett (1993)-Aiyagari (1994) models with labor income risk have difficulty in matching the (empirically fat) upper tail of the wealth distribution (see Benhabib et al., 2017, for an excellent discussion of this issue). In the latter case with idiosyncratic investment risk, we show that the wealth distribution has Pareto tails. This type of model has recently become quite popular for studying size distributions; see, for example, Luttmer (2007), Nirei and Souma (2007), Benhabib et al. (2011, 2015, 2016), Toda (2014), Acemoglu and Cao (2015), Toda and Walsh (2015, 2017), Gabaix et al. (2016), Nirei and Aoki (2016), and Aoki and Nirei (2017). All of these papers assume iid shocks, with the exception of Toda (2014), whose non-iid result only holds in the continuous-time limit, where shocks are Gaussian. Our results permit non-Gaussian Markov shocks and thus considerably expand the toolkit for applied economists.

## 1.1 Related literature

Important early contributions on generative mechanisms of power laws were made by Champernowne (1953), who proposed a model of income dynamics that is mean-reverting in which a power law emerges as a steady state equilibrium. Wold and Whittle (1957) found that a constant growth of wealth together with the exponential age distribution generates a Pareto upper tail, which anticipates Reed (2001)'s mechanism. Simon and Bonini (1958) observed that Gibrat's law in combination with exponential growth in the number of firms could lead to a power law in the upper tail of the firm size distribution. Kesten (1973) studied the random difference equation $X_t = A_t X_{t-1} + B_t$ and showed that the stationary distribution exhibits a power law tail.

More recently, a paper by Gabaix (1999) providing an explanation for Zipf's law in city sizes sparked renewed interest in power laws amongst economists. Note that Reed (2001)'s mechanism differs from that of Gabaix (1999), who augmented the geometric Brownian motion law for city size evolution with a lower reflecting boundary in order to obtain a power law exponent slightly above one in the upper tail of the city size distribution, consistent with Zipf's law. Survey papers by Gabaix (2009, 2016) discuss much of the subsequent economics literature. Another survey paper by Mitzenmacher (2004) discusses generative mechanisms for power laws proposed across a wider range of disciplines, including biology, computer science, networks, and physics. A more recent survey by Benhabib and Bisin (2017) focuses on the wealth distribution.

Given the long tradition of random growth models, a little historical digression may be justified. By the early 20th century, it was recognized as a puzzle amongst statisticians that variables of empirical interest that were considered to be accumulating numerous small independent shocks frequently exhibited skewness or excess kurtosis, despite the central limit theorem. A simple explanation was suggested by Kapteyn (1903): if small shocks accumulate to a function $F(x)$ of a quantity of interest $x$, rather than $x$ itself, then one can generate skewed distributions by applying the change of variable formula to the normal distribution.

Gibrat (1931, ch. 5) applied this argument to $F(x) = \log x$ and arrived at his celebrated "la loi de l'effect proportionnel" (the law of proportional effect). In Chapter 6, Gibrat elaborates further in the context of the firm size distribution and mentions that necessary and sufficient conditions for obtaining "formula A" (which is essentially the lognormal distribution) are:

(i) "Les causes de fluctuation du personnel sont nombreuses" (there is a large number of shocks to the fluctuation of employees),

(ii) "Leur effet *relatif* sur le nombre d'ouvriers (ou leur effet absolut sur le *logarithme*), ne dépend pas de ce nombre d'ouvriers." (the relative effect on the number of workers (or the absolute effect on the logarithm) does not depend on the number of workers),

(iii) "L'effet de chaque cause de fluctuation est petit vis-à-vis de l'effet de toutes" (the effect of each shock is small relative to the total effect).

Essentially, he is applying the central limit theorem to the shocks to the logarithm of the firm size measured by the number of employees (and cites Lindeberg, 1922, for a justification).[5]

Gibrat's argument has one pitfall, as pointed out by Kalecki (1945): with a random growth model with infinitely lived agents, the cross-sectional distribution is approximately lognormal but the log mean and variance increase linearly overtime, and hence a stationary distribution does not exist. One of the very first

[5]Note that nowhere in Gibrat's original argument is the assumption of Gaussian shocks. Using the geometric Brownian motion to represent a random growth process is merely a mathematical convenience, which may not be supported empirically. For example, Arata (2015, ch. 2) provides evidence that the changes in log firm size are not Gaussian and argues that the more general class of Lévy processes may be empirically preferable. The foundational theory for infinitely divisible distributions and Lévy processes was developed by de Finetti, Lévy, Khintchine and Kolmogorov during the late 1920s and 1930s (Mainardi and Rogosin, 2006).

solutions to this problem was to introduce birth and death. In particular, Rutherford (1955) mentions that "[the] life table for most populations appears to be such that it is a reasonable approximation to assume the number of survivors at period $(t+r)$ of an entry of $\alpha_t$ at time $t$ is given by $\alpha_t \mathrm{e}^{-Kr}$, where $1/K$ is the expectation of life," thus assumes an exponential age distribution. Since the income shocks are Gaussian in his model, it is exactly the discrete-time version of Reed (2001). However, Rutherford (1955) did not characterize the tail behavior; for this we had to wait until Wold and Whittle (1957) and Reed (2001).

## 2 Exponential tails via the Laplace transform

Theorems that allow us to deduce limit properties of a probability distribution or other function from limit properties of its Laplace transform are called Tauberian theorems. In this section we state a version of a Tauberian theorem of Nakagawa (2007) which will be used to prove our main results on the tail probabilities of geometric sums of hidden Markov processes. First we briefly review the Laplace transform. For more details see Widder (1941) and Lukacs (1970, ch. 7).

Given a cumulative distribution function (cdf) $F$, let

$$M(s) = \int_{-\infty}^{\infty} \mathrm{e}^{sx} \,\mathrm{d}F(x) \in \mathbb{R} \cup \{\infty\}$$

be its moment generating function (mgf). Since $\mathrm{e}^{sx}$ is convex in $s$, the domain $\mathcal{I} = \{s \in \mathbb{R} : M(s) < \infty\}$ is convex, and hence an interval. Clearly $0 \in \mathcal{I}$, so there exist boundary points $-\beta \le 0 \le \alpha$ of $\mathcal{I}$ (which may be 0 or $\pm\infty$). The numbers $\alpha, -\beta$ are called the right and left abscissae of convergence.

We obtain the Laplace transform of $F$ by extending the domain of $M$ into the complex plane. Suppose that $-\beta < \alpha$ and let $z = s + it \in \mathbb{C}$. By the definition of the Lebesgue-Stieltjes integral,

$$M(z) = \int_{-\infty}^{\infty} \mathrm{e}^{zx} \,\mathrm{d}F(x) \tag{2.1}$$

exists if and only if

$$\int_{-\infty}^{\infty} |\mathrm{e}^{zx}| \,\mathrm{d}F(x) = \int_{-\infty}^{\infty} \left|\mathrm{e}^{(s+it)x}\right| \mathrm{d}F(x) = \int_{-\infty}^{\infty} \mathrm{e}^{sx} \,\mathrm{d}F(x) = M(s) < \infty.$$

Therefore by the definition of $\alpha, -\beta$, the value $M(z)$ is well-defined for $z \in \mathbb{C}$ with $\operatorname{Re} z \in \mathcal{I}$. Let $\mathcal{S} = \{z \in \mathbb{C} : -\beta < \operatorname{Re} z < \alpha\}$. Using the dominated convergence theorem, it is easy to see that $M(z)$ is holomorphic on $\mathcal{S}$, which is called the strip of holomorphicity (Figure 2.1). The lines $\operatorname{Re} z = \alpha, -\beta$ comprising the boundary of $\mathcal{S}$ are called the right and left axes of convergence. In this paper we refer to (2.1) as the (two-sided) *Laplace transform* of $F$, or of any real random variable $X$ with cdf $F$. We also use "Laplace transform" and "moment generating function" interchangeably.

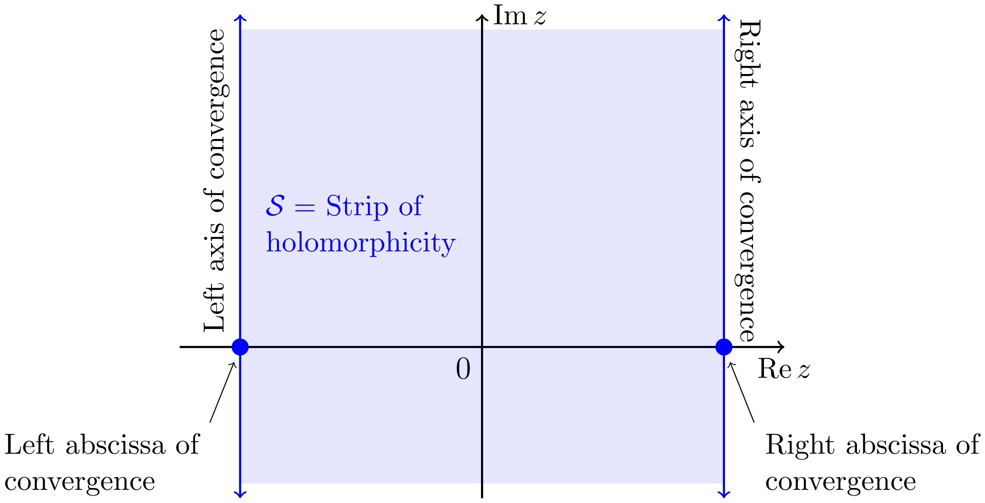



**Figure 2.1:** Region of convergence of the Laplace transform $M(z) = \int_{-\infty}^{\infty} \mathrm{e}^{zx}\,\mathrm{d}F(x)$.

There is a close relationship between the tail probabilities of a cdf and the abscissae of convergence of its Laplace transform. In general, we have

$$\limsup_{x\to\infty} \frac{1}{x} \log \mathrm{P}(X > x) = -\alpha, \quad \limsup_{x\to\infty} \frac{1}{x} \log \mathrm{P}(X < -x) = -\beta \tag{2.2}$$

whenever the relevant abscissa is nonzero (Widder, 1941, pp. 42–43, 241; Lukacs, 1970, p. 194). Moreover, each abscissa is a singularity of the Laplace transform (Widder, 1941, p. 58). A Tauberian theorem of Nakagawa (2007) establishes a tighter relationship between the abscissae of convergence and the tail probabilities that depends on the nature of the singularities at the abscissae: if either singularity is a pole then the relevant limit superior in (2.2) may be replaced with an ordinary limit. In addition, bounds for the limits superior and inferior of $\mathrm{e}^{\alpha x}\mathrm{P}(X > x)$ and $\mathrm{e}^{\beta x}\mathrm{P}(X < -x)$ become available which depend upon the order of the poles at the abscissae, the leading Laurent coefficients, and the location of other singularities along the axes of convergence. For our purposes, it will be enough to consider the case where the singularities at the abscissae are simple poles. The following result is similar to Theorem 5* of Nakagawa (2007), specialized to the case of a simple pole. We provide a proof in the Appendix (see Remark 2.5 below).

**Theorem 2.1.** *Let $X$ be a real random variable and $M(z) = \mathrm{E}[\mathrm{e}^{zX}]$ its Laplace transform, with right abscissa of convergence $\alpha$ satisfying $0 < \alpha < \infty$, and strip of holomorphicity $\mathcal{S}$. Fix $A > 0$, and let $B$ be the supremum of all $b > 0$ such that the function $M(z) + A(z-\alpha)^{-1}$ can be continuously extended from $\mathcal{S}$ to the set*

$$\mathcal{S}_b^+ = \mathcal{S} \cup \{z \in \mathbb{C} : z = \alpha + it, |t| < b\} . \tag{2.3}$$

*Suppose that $B > 0$. Then we have*

$$\frac{2\pi A/B}{\mathrm{e}^{2\pi\alpha/B} - 1} \leq \liminf_{x\to\infty} \mathrm{e}^{\alpha x}\mathrm{P}(X > x) \leq \limsup_{x\to\infty} \mathrm{e}^{\alpha x}\mathrm{P}(X > x) \leq \frac{2\pi A/B}{1 - \mathrm{e}^{-2\pi\alpha/B}}, \tag{2.4}$$

*where the bounds should be read as $A/\alpha$ if $B = \infty$. These bounds are sharp.*

**Remark 2.1.** Applying Theorem 2.1 to the random variable $-X$ can yield bounds analogous to (2.4) for the lower tail probabilities. Specifically, given $A > 0$, if the left abscissa $-\beta$ satisfies $-\infty < -\beta < 0$ and $B$ is the supremum of all $b > 0$ such that $M(z) - A(z+\beta)^{-1}$ can be continuously extended from $\mathcal{S}$ to the set

$$\mathcal{S}_b^- = \mathcal{S} \cup \{z \in \mathbb{C} : z = -\beta + it, |t| < b\} ,$$

then, if $B > 0$, we have the bounds

$$\frac{2\pi A/B}{\mathrm{e}^{2\pi\beta/B} - 1} \leq \liminf_{x\to\infty} \mathrm{e}^{\beta x}\mathrm{P}(X < -x) \leq \limsup_{x\to\infty} \mathrm{e}^{\beta x}\mathrm{P}(X < -x) \leq \frac{2\pi A/B}{1 - \mathrm{e}^{-2\pi\beta/B}}.$$

**Remark 2.2.** To see why the bounds in (2.4) are sharp, let $X$ be a geometrically distributed random variable with mean $1/p \in (0,\infty)$. The Laplace transform of $X$ has right abscissa of convergence $\alpha = -\log(1-p) > 0$ and strip of holomorphicity $\mathcal{S} = \{z \in \mathbb{C} : \mathrm{Re}\, z < \alpha\}$, on which it is given by

$$M(z) = \mathrm{E}[\mathrm{e}^{zX}] = \sum_{n=1}^{\infty} p(1-p)^{n-1}\mathrm{e}^{zn} = \frac{p\mathrm{e}^z}{1-(1-p)\mathrm{e}^z}.$$

The zeros of the denominator $1-(1-p)\mathrm{e}^z$ are $z_k = \alpha + 2\pi ik$, $k \in \mathbb{Z}$, all of which lie on the axis of convergence of $M(z)$. These zeros are singularities of $M(z)$. By l'Hôpital's rule we have

$$\lim_{z\to\alpha} \frac{(z-\alpha)p\mathrm{e}^z}{1-(1-p)\mathrm{e}^z} = -\lim_{z\to\alpha} \frac{p\mathrm{e}^z + (z-\alpha)p\mathrm{e}^z}{(1-p)\mathrm{e}^z} = -\frac{p}{1-p},$$

which shows that the singularity at $\alpha$ is a simple pole with residue $-p/(1-p)$.

The singularities along the axis of convergence are separated by gaps of $2\pi i$, so the assumptions of Theorem 2.1 are satisfied with $A = p/(1-p)$ and $B = 2\pi$, and thus (2.4) holds with the lower and upper bounds

$$\frac{2\pi A/B}{\mathrm{e}^{2\pi\alpha/B}-1} = 1, \quad \frac{2\pi A/B}{1-\mathrm{e}^{-2\pi\alpha/B}} = \frac{1}{1-p}.$$

On the other hand, if $n \le x < n+1$ for some $n \in \mathbb{N}$, then

$$\mathrm{P}(X > x) = \sum_{k=n+1}^{\infty} p(1-p)^{k-1} = (1-p)^n = (1-p)^{\lfloor x \rfloor}$$

by direct computation. Therefore

$$\mathrm{e}^{\alpha x}\mathrm{P}(X > x) = (1-p)^{-x}\mathrm{P}(X > x) = (1-p)^{\lfloor x \rfloor - x},$$

which oscillates between 1 and $1/(1-p)$ as $x \to \infty$. The bounds in (2.4) are therefore sharp.

**Remark 2.3.** When the bounds in (2.4) are satisfied, for any $\epsilon > 0$ there exists $x_0 < \infty$ such that

$$\frac{2\pi A/B}{\mathrm{e}^{2\pi\alpha/B}-1} - \epsilon \le \mathrm{e}^{\alpha x}\mathrm{P}(X > x) \le \frac{2\pi A/B}{1-\mathrm{e}^{-2\pi\alpha/B}} + \epsilon$$

for all $x \ge x_0$. Taking logarithms, dividing by $x$, and letting $x \to \infty$, we obtain

$$\lim_{x\to\infty} \frac{1}{x} \log \mathrm{P}(X > x) = -\alpha, \tag{2.5}$$

which improves upon the characterization of right tail probabilities in (2.2).

**Remark 2.4.** In the case $B = \infty$, which corresponds to the simple pole at $\alpha$ being the unique singularity on the axis of convergence, we may rewrite (2.4) as

$$\lim_{s\to\infty} s^\alpha \mathrm{P}(S > s) = \frac{A}{\alpha}, \tag{2.6}$$

where $S = \mathrm{e}^X$. In this sense, the right tail probabilities of $S$ are Paretian. Property (2.6) implies regular variation of the right tail probabilities of $S$ with index $-\alpha$, under which $s^\alpha \mathrm{P}(S > s)$ is required to be slowly varying at infinity. Property (2.6) also implies (2.5), which will be satisfied for any $B > 0$. On the other hand, (2.5) does not imply (2.6). For instance, if $X$ is a geometric random variable as in Remark 2.2, then (2.5) holds but (2.6) fails to hold because

$$1 = \liminf_{s\to\infty} s^\alpha \mathrm{P}(S > s) < \limsup_{s\to\infty} s^\alpha \mathrm{P}(S > s) = \frac{1}{1-p}.$$

**Remark 2.5.** Aside from the fact that it deals only with nonnegative random variables (which is unimportant), Theorem 5* of Nakagawa (2007) is more general than Theorem 2.1 above, as it provides bounds on tail probabilities that apply when the pole at the right abscissa of convergence is of arbitrary order. This additional generality makes the proof quite complicated—and in fact it is given in detail only for the case of a second order pole. For the reader's convenience, in Appendix A we provide a simpler proof of Theorem 2.1 that applies in our more restrictive setting. Following Nakagawa, we draw on a technique used by Graham and Vaaler (1981) to prove a refinement of the Wiener-Ikehara Tauberian theorem, a helpful account of which may be found in the monograph of Korevaar (2004, ch. 5). Theorem 2.1 is in fact very similar to the Tauberian theorem proved by Graham and Vaaler (1981), with the only real difference being that it concerns exponentially decaying tail probabilities rather than exponential growth.

# 3 Main results

## 3.1 Tail behavior of geometric sums of hidden Markov processes

In this section we characterize the tail behavior of the geometric sum (1.1) when $\{X_t\}_{t=1}^\infty$ is a hidden Markov process.

**Definition 3.1** (Hidden Markov process)**.** A sequence of real-valued random variables $\{X_t\}_{t=1}^\infty$ is called a *hidden Markov process* if there exists a natural number $N$, a time homogeneous Markov process $\{J_t\}_{t=0}^\infty$ taking values in $\mathcal{N} = \{1, \dots, N\}$, and an iid sequence of $\mathbb{R}^N$-valued random variables $\{Y_t\}_{t=1}^\infty$ (where $Y_t = (Y_{1t}, \dots, Y_{Nt})^\top$) such that $X_t = Y_{J_t,t}$ for all $t \in \mathbb{N}$.

**Example 3.1.** If $\{X_t\}_{t=1}^\infty$ is iid then it is a hidden Markov process with $N = 1$.

**Example 3.2.** If $\{X_t\}_{t=1}^\infty$ is a finite-state time homogeneous Markov chain taking the values $x_1, \dots, x_N$ then it is a hidden Markov process: just set $Y_t \equiv (x_1, \dots, x_N)^\top$ (a constant vector).

Throughout this section, let $\{X_t\}_{t=1}^\infty$ be a hidden Markov process with the underlying states denoted by $\{J_t\}_{t=0}^\infty$. Let $\Pi = (\pi_{nn'})$ be the transition probability matrix, where $\pi_{nn'} = \mathrm{P}(J_1 = n' \mid J_0 = n)$. Let $W_t$ be the cumulative sum of $X_t$, so $W_0 = 0$ and $W_t = W_{t-1} + X_t$ for $t \in \mathbb{N}$. Finally, let $T$ be a geometrically distributed random variable with mean $1/p$ and independent of the $X, J$ processes.

The following theorem is our main result. It shows that under weak assumptions the geometric sum $W_T$ has exponential tails, and characterizes their decay rates in terms of $p$, $\Pi$, and the conditional mgf of $X_1$ given $J_1$. That characterization is illustrated in Figure 3.1.

**Theorem 3.1.** *Let everything be as above and suppose that $\Pi$ is irreducible. Then the set*

$$\mathcal{I} := \left\{s \in \mathbb{R} : (\forall n \in \mathcal{N})\, \mathrm{E}\left[\mathrm{e}^{sX_1} \,\middle|\, J_1 = n\right] < \infty\right\}$$

*is convex and contains zero. For $s \in \mathcal{I}$, let $D(s)$ be the $N \times N$ diagonal matrix with $n$-th diagonal entry equal to $\mathrm{E}\left[\mathrm{e}^{sX_1} \,\middle|\, J_1 = n\right]$. There can be at most one positive real number $\alpha \in \mathcal{I}$ for which $\Pi D(\alpha)$ has spectral radius $1/(1-p)$, and if such $\alpha$ exists in the interior of $\mathcal{I}$ then*

$$\lim_{w\to\infty} \frac{1}{w} \log \mathrm{P}(W_T > w) = -\alpha. \tag{3.1}$$

*Similarly, there can be at most one negative real number $-\beta \in \mathcal{I}$ for which $\Pi D(-\beta)$ has spectral radius $1/(1-p)$, and if such $-\beta$ exists in the interior of $\mathcal{I}$ then*

$$\lim_{w\to\infty} \frac{1}{w} \log \mathrm{P}(W_T < -w) = -\beta. \tag{3.2}$$

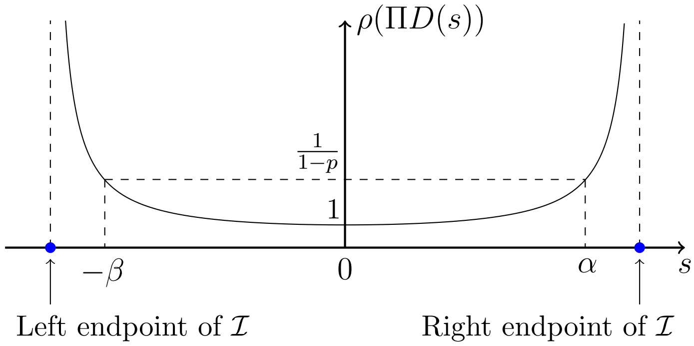



**Figure 3.1:** Determination of $\alpha$ and $\beta$

*Proof.* For a square matrix $A$, let $\rho(A)$ denote the spectral radius of $A$. We divide our proof into five steps.

*Step 1. $\rho(\Pi D(s))$ is convex in $s \in \mathcal{I}$. The number $\alpha$ in the statement of Theorem 3.1, if it exists, is unique.*

For each $n \in \mathcal{N}$ define $\mathcal{I}_n = \left\{s \in \mathbb{R} : \mathrm{E}\left[\mathrm{e}^{sX_1} \,\middle|\, J_1 = n\right] < \infty\right\}$. Applying Hölder's inequality, we obtain

$$\begin{aligned}\mathrm{E}&\left[\mathrm{e}^{((1-\theta)s_1+\theta s_2)X_1} \,\middle|\, J_1 = n\right] \\ &\leq \left(\mathrm{E}\left[\mathrm{e}^{s_1X_1} \,\middle|\, J_1 = n\right]\right)^{1-\theta}\left(\mathrm{E}\left[\mathrm{e}^{s_2X_1} \,\middle|\, J_1 = n\right]\right)^{\theta} < \infty \end{aligned} \tag{3.3}$$

for all $\theta \in (0,1)$, all $s_1, s_2 \in \mathcal{I}_n$, and all $n \in \mathcal{N}$. This shows that each $\mathcal{I}_n$ is convex and contains zero, implying that $\mathcal{I} = \bigcap_{n\in\mathcal{N}} \mathcal{I}_n$ is convex and contains zero. It also shows that each diagonal entry of the matrix $D(s)$ is a log-convex function of $s \in \mathcal{I}$. Letting $\odot$ denote the Hadamard (entry-wise) product, collecting (3.3) into a matrix we obtain

$$\Pi D((1-\theta)s_1 + \theta s_2) \leq \Pi D(s_1)^{(1-\theta)} D(s_2)^{(\theta)} = (\Pi D(s_1))^{(1-\theta)} \odot (\Pi D(s_2))^{(\theta)}$$

entry-wise for all $\theta \in (0,1)$ and all $s_1, s_2 \in \mathcal{I}$, where for a nonnegative matrix $A = (a_{mn})$ we define $A^{(\theta)} = (a_{mn}^{\theta})$. By Propositions B.2 and B.1, we obtain

$$\begin{aligned}\rho(\Pi D((1-\theta)s_1 + \theta s_2)) &\leq \rho\left((\Pi D(s_1))^{(1-\theta)} \odot (\Pi D(s_2))^{(\theta)}\right) \\ &\leq \rho(\Pi D(s_1))^{1-\theta} \rho(\Pi D(s_2))^{\theta},\end{aligned}$$

which shows that $\rho(\Pi D(s))$ is log-convex (and hence convex).

To show the uniqueness of $\alpha$, suppose on the contrary that there are two numbers $0 < \alpha_1 < \alpha_2$ such that $\lambda(s) := \rho(\Pi D(s))$ equals $1/(1-p)$ for $s = \alpha_1, \alpha_2$. Let $\theta = \alpha_1/\alpha_2 \in (0,1)$. Since $\lambda(0) = \rho(\Pi D(0)) = \rho(\Pi) = 1$ (a nonnegative matrix whose rows sum to one has spectral radius one—see *e.g.* Horn and Johnson, 1985, p. 547) and $1 < 1/(1-p)$, it follows from the convexity of $\lambda$ that

$$\begin{aligned}\frac{1}{1-p} &= \lambda(\alpha_1) = \lambda(\theta\alpha_2 + (1-\theta)0) \\ &\leq \theta\lambda(\alpha_2) + (1-\theta)\lambda(0) = \frac{\theta}{1-p} + 1 - \theta < \frac{1}{1-p},\end{aligned}$$

which is a contradiction.

Define the strip $\mathcal{S} = \{z \in \mathbb{C} : \operatorname{Re} z \in \mathcal{I}\}$ (Figure 3.2). For each $z \in \mathcal{S}$ and $n \in \mathcal{N}$ we have $\mathrm{E}\left[\left|\mathrm{e}^{zX_1}\right| \,\middle|\, J_1 = n\right] = \mathrm{E}\left[\mathrm{e}^{(\operatorname{Re} z)X_1} \,\middle|\, J_1 = n\right] < \infty$, and may therefore extend the domain of definition of $D(s)$ from $\mathcal{I}$ to $\mathcal{S}$ by letting $D(z)$ be the $N \times N$ diagonal matrix with $n$-th diagonal entry equal to $\mathrm{E}\left[\mathrm{e}^{zX_1} \,\middle|\, J_1 = n\right]$, a holomorphic function of $z$ on the interior of $\mathcal{S}$.

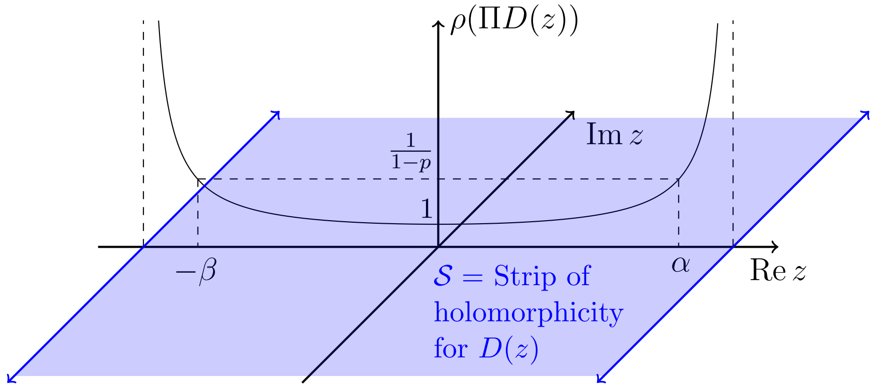



**Figure 3.2:** Definition of $\mathcal{S}$

*Step 2. Let $e$ be the $N \times 1$ vector of ones. For $t \in \{0\} \cup \mathbb{N}$ and $z \in \mathcal{S}$, the conditional mgf of $W_t$ given $J_0$ is given by*

$$\begin{bmatrix} \mathrm{E}\left[\mathrm{e}^{zW_t} \,\middle|\, J_0 = 1\right] \\ \vdots \\ \mathrm{E}\left[\mathrm{e}^{zW_t} \,\middle|\, J_0 = N\right] \end{bmatrix} = (\Pi D(z))^t e. \tag{3.4}$$

*Consequently, the mgf of $W_t$ is given by $\mathrm{E}[\mathrm{e}^{zW_t}] = \omega_0^\top (\Pi D(z))^t e$ for $z \in \mathcal{S}$, where $\omega_0$ is the $N \times 1$ vector of probabilities $\mathrm{P}(J_0 = n)$, $n = 1, \ldots, N$*

Equality (3.4) is trivially true for $t = 0$ since $W_0 = 0$ by definition. To deal with the case $t \geq 1$ we use the law of iterated expectations and the conditional independence of $W_{t+1}$ and $J_0$ given $J_1$ to write

$$\mathrm{E}\left[\mathrm{e}^{zW_{t+1}} \,\middle|\, J_0 = n\right] = \sum_{n'=1}^{N} \mathrm{P}(J_1 = n' \mid J_0 = n)\, \mathrm{E}\left[\mathrm{e}^{zW_{t+1}} \,\middle|\, J_1 = n'\right].$$

Since $J_0, J_1, \ldots$ is a time homogeneous Markov chain, the law of $W_{t+1} - X_1$ given $J_1 = n'$ is the same as the law of $W_t$ given $J_0 = n'$. Thus, using the conditional independence of $X_1$ and $W_{t+1} - X_1$ given $J_1$ we have

$$\begin{aligned} \mathrm{E}\left[\mathrm{e}^{zW_{t+1}} \,\middle|\, J_1 = n'\right] &= \mathrm{E}\left[\mathrm{e}^{zX_1} \,\middle|\, J_1 = n'\right] \mathrm{E}\left[\mathrm{e}^{z(W_{t+1} - X_1)} \,\middle|\, J_1 = n'\right] \\ &= \mathrm{E}\left[\mathrm{e}^{zX_1} \,\middle|\, J_1 = n'\right] \mathrm{E}\left[\mathrm{e}^{zW_t} \,\middle|\, J_0 = n'\right]. \end{aligned}$$

It follows that

$$\mathrm{E}\left[\mathrm{e}^{zW_{t+1}} \,\middle|\, J_0 = n\right] = \sum_{n'=1}^{N} \mathrm{P}(J_1 = n' \mid J_0 = n)\, \mathrm{E}\left[\mathrm{e}^{zX_1} \,\middle|\, J_1 = n'\right] \mathrm{E}\left[\mathrm{e}^{zW_t} \,\middle|\, J_0 = n'\right].$$

If (3.4) is true for some $t \geq 0$, this last equation tells us that $\mathrm{E}\left[\mathrm{e}^{zW_{t+1}} \,\middle|\, J_0 = n\right]$ is equal to the $n$-th row of $\Pi D(z)(\Pi D(z))^t e$. Thus (3.4) is true for all $t \geq 0$ by induction.

Define the set

$$\mathcal{I}_p = \left\{ s \in \mathcal{I} : \rho(\Pi D(s)) < \frac{1}{1-p} \right\}.$$

The set $\mathcal{I}_p$ is convex and contains zero, due to the previously established convexity of $\rho(\Pi D(s))$ and the fact that $\rho(\Pi D(0)) = 1 < 1/(1-p)$. Define the strip $\mathcal{S}_p = \{z \in \mathbb{C} : \operatorname{Re} z \in \mathcal{I}_p\}$ (Figure 3.3). Let $I$ be the $N \times N$ identity matrix, and for $z \in \mathcal{S}$ define the matrix-valued complex function (matrix pencil) $A(z) = I-(1-p)\Pi D(z)$.

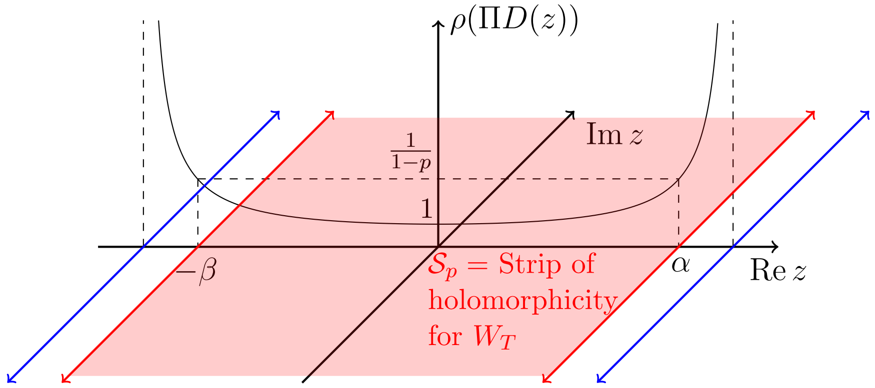



**Figure 3.3:** Definition of $\mathcal{S}_p$

*Step 3. For $z \in \mathcal{S}_p$, $A(z)$ is invertible and the mgf of $W_T$ is given by*

$$\mathrm{E}[\mathrm{e}^{zW_T}] = p\omega_0^\top A(z)^{-1}\Pi D(z)e. \tag{3.5}$$

For $z \in \mathcal{S}_p$, we have $\rho(\Pi D(z)) \leq \rho(\Pi D(\operatorname{Re} z)) < 1/(1-p)$, where the first inequality follows from Property (ii) of Proposition B.2 and the second from the definition of $\mathcal{S}_p$. Hence by Property (i) of Proposition B.2, $(1-p)^k \left\|(\Pi D(z))^k\right\|$ decays to zero at an exponential rate as $k \to \infty$ for any matrix norm $\|\cdot\|$ and any $z \in \mathcal{S}_p$. Hence the geometric series $\sum_{k=0}^\infty (1-p)^k(\Pi D(z))^k$ converges on $\mathcal{S}_p$. However,

$$(I-(1-p)\Pi D(z))\sum_{k=0}^{K}(1-p)^k(\Pi D(z))^k = I-(1-p)^{K+1}(\Pi D(z))^{K+1} \to I$$

as $K \to \infty$, so $A(z) = I-(1-p)\Pi D(z)$ is invertible on $\mathcal{S}_p$ and

$$A(z)^{-1} = \sum_{k=0}^{\infty}(1-p)^k(\Pi D(z))^k.$$

Therefore, using the fact that $\mathrm{E}[\mathrm{e}^{zW_t}] = \omega_0^\top(\Pi D(z))^t e$ for $z \in \mathcal{S}$ and hence for $z \in \mathcal{S}_p$, we obtain

$$\mathrm{E}[\mathrm{e}^{zW_T}] = \sum_{k=1}^{\infty} p(1-p)^{k-1}\omega_0^\top(\Pi D(z))^k e = p\omega_0^\top A(z)^{-1}\Pi D(z)e$$

for $z \in \mathcal{S}_p$, which is (3.5).

*Step 4. Suppose there exists $\alpha > 0$ in the interior of $\mathcal{I}$ such that $\rho(\Pi D(\alpha)) = 1/(1-p)$. Then $A(z)^{-1}$ is well-defined and holomorphic in a punctured neighborhood of the singularity at $z = \alpha$, which is a simple pole.*

Suppose there exists $\alpha > 0$ in the interior of $\mathcal{I}$ such that $\rho(\Pi D(\alpha)) = 1/(1-p)$. By assumption $\Pi$ is nonnegative and irreducible, and $D(\alpha)$ is a diagonal matrix with positive diagonal elements, so $\Pi D(\alpha)$ is also nonnegative and irreducible. The Perron-Frobenius theorem (Appendix B) therefore implies that the spectral radius $1/(1-p)$ is an eigenvalue of $\Pi D(\alpha)$, called the Perron root. Let $x, y$ be the right and left Perron vectors corresponding to the Perron root. Since $x, y$ are positive, we can normalize them so that the entries of $x, y$ sum to 1.

Since $1/(1-p)$ is an eigenvalue of $\Pi D(\alpha)$, we have $\det A(\alpha) = 0$. On the other hand, since $A(z)$ is invertible for $z \in \mathcal{S}_p \subset \mathcal{S}$, we have $\det A(z) \neq 0$ for some $z \in \mathcal{S}$. Moreover, $A(z)$ and $\det A(z)$ inherit from $D(z)$ the property of holomorphicity on the interior of $\mathcal{S}$. Since $\det A(z)$ is nonconstant and holomorphic on the interior of $\mathcal{S}$ with a zero at $z = \alpha$, that zero must be isolated, and therefore $A(z)$ is holomorphic in a neighborhood of $z = \alpha$ with an isolated point of noninvertibility at $z = \alpha$. Consequently, Proposition C.1 tells us that the inverse $A(z)^{-1}$ is holomorphic in a punctured neighborhood of the singularity at $z = \alpha$, and that this singularity is a simple pole if and only if the algebraic and geometric multiplicities of the eigenvalue $1/(1-p)$ of $\Pi D(\alpha)$ are equal. We know from the Perron-Frobenius theorem that both multiplicities are one; $A(z)^{-1}$ therefore has a simple pole at $z = \alpha$.

*Step 5. The right-hand side of (3.5) is well-defined and holomorphic in a punctured neighborhood of the singularity at $z = \alpha$, which is a simple pole.*

Proposition C.1 implies that the residue $R$ of the simple pole $\alpha$ of $A(z)^{-1}$ is given by the $N \times N$ matrix

$$R = x(y^\top A'(\alpha)x)^{-1}y^\top =: cxy^\top, \tag{3.6}$$

where $A'(z)$ is the matrix of derivatives of $A(z)$, $x, y$ are the Perron vectors introduced above, and $c = (y^\top A'(\alpha)x)^{-1}$ is a nonzero scalar. Since $x, y$ are right and left eigenvectors of $\Pi D(\alpha)$ with corresponding eigenvalue $1/(1-p)$, it follows that $A(\alpha)R = RA(\alpha) = 0$.

By assumption $\alpha$ belongs to the interior of $\mathcal{S}$, and we observed earlier that $D(z)$ is holomorphic on the interior of $\mathcal{S}$, so we may use the right-hand side of (3.5) to holomorphically extend $\mathrm{E}[\mathrm{e}^{zW_T}]$ to a punctured neighborhood of $\alpha$, which is a singularity. Let us show that this singularity is a simple pole by showing that

$$\lim_{z\to\alpha}(z-\alpha)p\omega_0^\top A(z)^{-1}\Pi D(z)e \tag{3.7}$$

exists and is nonzero. Since $A(z)^{-1}$ has a simple pole at $z = \alpha$ with residue $R$ we know that

$$\lim_{z\to\alpha}(z-\alpha)A(z)^{-1} = R,$$

and since $D(z)$ is continuous at $z = \alpha$ we know that $\lim_{z\to\alpha} D(z) = D(\alpha)$. Therefore the limit in (3.7) exists and is equal to $p\omega_0^\top R\Pi D(\alpha)e$. Since $RA(\alpha) = 0$, we have

$$R\Pi D(\alpha)e = -\frac{1}{1-p}R(I-(1-p)\Pi D(\alpha))e + \frac{1}{1-p}Re = \frac{1}{1-p}Re.$$

Using (3.6) again, it follows that

$$(3.7) = p\omega_0^\top \frac{1}{1-p}Re = c\frac{p}{1-p}(\omega_0^\top x)(y^\top e) = c\frac{p}{1-p}\omega_0^\top x \neq 0 \tag{3.8}$$

since the entries of $y$ sum to 1, $\omega_0 > 0$, and $x \gg 0$.[6] Therefore our holomorphic extension of $\mathrm{E}[\mathrm{e}^{zW_T}]$ has a simple pole at $z = \alpha$.

The limit (3.1) now follows from Theorem 2.1 and Remark 2.3. A symmetric argument establishes the analogous result for the lower tail. □

**Remark 3.1.** If $N = 1$, so that $\{X_t\}_{t=1}^\infty$ is iid, then $\rho(\Pi D(s)) = \mathrm{E}[\mathrm{e}^{sX_1}]$, and the tail exponents are determined by

$$\mathrm{E}[\mathrm{e}^{\alpha X_1}] = \frac{1}{1-p} \quad \text{and} \quad \mathrm{E}[\mathrm{e}^{-\beta X_1}] = \frac{1}{1-p},$$

whenever such $\alpha$ and/or $-\beta$ exist in the interior of $\mathcal{I}$. This explains (1.2).

**Remark 3.2.** One of the assumptions of Theorem 3.1 is that there exists $\alpha > 0$

[6] Given real vectors $a = (a_i)$ and $b = (b_i)$, we write $a > b$ when $a_i \geq b_i$ for all $i$ and $a_i > b_i$ for some $i$, and we write $a \gg b$ when $a_i > b_i$ for all $i$.

in the interior of $\mathcal{I}$ such that $\rho(\Pi D(\alpha)) = 1/(1-p)$. This assumption may not always be satisfied. As a counterexample, fix $\alpha > 0$ and suppose that $\{X_t\}_{t=1}^{\infty}$ is iid ($N = 1$) with an mgf $M(s)$ such that $M(\alpha) < \infty$ and $M(s) = \infty$ for $s > \alpha$.[7] Then the mgf of $W_T$ is infinite for $s > \alpha$ and satisfies

$$\mathrm{E}[\mathrm{e}^{sW_T}] = \sum_{k=1}^{\infty} p(1-p)^{k-1} M(s)^k = \frac{pM(s)}{1-(1-p)M(s)} < \infty$$

for $s \in [0, \alpha]$ if $p \in (0, 1)$ is sufficiently close to 1 because $M(\alpha) < \infty$. Thus the right abscissa of convergence of $W_T$ is $\alpha$ and recalling (2.2) we have

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

but there is no $s > 0$ that satisfies $\rho(\Pi D(s)) = M(s) = 1/(1-p)$, so we may not appeal to Theorem 3.1 to strengthen the above limit superior to an ordinary limit.

**Remark 3.3.** A sufficient condition for there to exist $\alpha > 0$ in the interior of $\mathcal{I}$ such that $\rho(\Pi D(\alpha)) = 1/(1-p)$ is that (i) $\mathrm{P}(X_1 > 0 \mid J_1 = n) > 0$ for some $n \in \mathcal{N}$, and (ii) $\mathrm{E}\left[\mathrm{e}^{sX_1} \,\middle|\, J_1 = n\right] < \infty$ for all $s > 0$ and all $n \in \mathcal{N}$. This is because (i) ensures that the convex function $\rho(\Pi D(s))$ diverges to infinity as $s \to \infty$ and (ii) ensures that $\rho(\Pi D(s)) < \infty$ for all $s > 0$. A similar sufficient condition applies to the lower tail exponent.

**Remark 3.4.** Theorem 3.1 is somewhat related to the (asymmetric) Laplace distribution, which is the logarithm of the double Pareto distribution. When $\{X_t\}_{t=1}^{\infty}$ is iid Laplace and $T$ is a geometric random variable with mean $1/p$, then the geometric sum $W_T = \sum_{t=1}^{T} X_t$ is also Laplace (Kotz et al., 2001, p. 151). When $\{X_t\}_{t=1}^{\infty}$ is iid with a general distribution with finite variance, then $W_T$ (properly scaled) weakly converges to the Laplace distribution as $p \to 0$ (Kotz et al., 2001, pp. 152–155). Toda (2014, Theorem 15) proves the same for the non-iid case provided that the central limit theorem holds for $\{X_t\}_{t=1}^{\infty}$.

**Example 3.3** (Gaussian distribution)**.** Consider the geometric Brownian motion (1.3). Using Itô's lemma, we obtain

$$\mathrm{d} \log S_t = \left(\mu - \frac{\sigma^2}{2}\right) \mathrm{d}t + \sigma \, \mathrm{d}B_t.$$

[7] One such example is provided by the density $f(x) = \mathbb{1}(x \geq 1) C x^{-\kappa-1} \mathrm{e}^{-\alpha x}$, where $\kappa, \alpha > 0$ and $C > 0$ is a number such that $\int f(x) \, \mathrm{d}x = 1$. The corresponding mgf is finite for $s \leq \alpha$ and infinite for $s > \alpha$.

The discrete-time analog of this process is

$$W_t = W_{t-1} + X_t, \quad X_t \sim \text{iid} N((\mu - \sigma^2/2)\Delta, \sigma^2\Delta),$$

where the choice of $\Delta > 0$ depends on the unit of time. Suppose that in the continuous-time model, the age distribution is exponential with parameter $\eta$. Then in the discrete-time analog, the birth/death probability is $p = 1 - \mathrm{e}^{-\eta\Delta}$. By Remark 3.1, $W_T$ has exponential tails with exponents $\zeta = \alpha, -\beta$ satisfying

$$\begin{aligned} \mathrm{E}[\mathrm{e}^{\zeta X_1}] = \frac{1}{1-p} &\iff \mathrm{e}^{(\mu-\sigma^2/2)\Delta\zeta + \sigma^2\Delta\zeta^2/2} = \mathrm{e}^{\eta\Delta} \\ &\iff \frac{\sigma^2}{2}\zeta^2 + \left(\mu - \frac{\sigma^2}{2}\right)\zeta - \eta = 0, \end{aligned}$$

which is exactly (1.4). Therefore with iid Gaussian shocks, the tail exponents are identical in the discrete-time and continuous-time models.

**Example 3.4** (Gamma distribution)**.** Suppose that $\{X_t\}_{t=1}^{\infty}$ is a sequence of iid random variables with the gamma density

$$f(x) = \frac{\lambda^{\kappa}}{\Gamma(\kappa)} x^{\kappa-1} \mathrm{e}^{-\lambda x}.$$

The gamma mgf is $M(z) = (1 - z/\lambda)^{-\kappa}$. Solving $M(\alpha) = \frac{1}{1-p}$ as in Remark 3.1, we see that the geometric sum $W_T = \sum_{t=1}^{T} X_t$ has an exponential upper tail with decay rate $\alpha = \lambda(1 - (1-p)^{1/\kappa})$.

**Example 3.5** (Two-state Markov chain)**.** Suppose that $\{X_t\}_{t=1}^{\infty}$ is a two-state time homogeneous Markov chain, as in Example 3.2 with $N = 2$. This is the simplest example of a non-iid process satisfying the assumptions of Theorem 3.1. The cumulated process $W_t$ is the "Markov trend in levels" studied by Hamilton (1989). By the Perron-Frobenius theorem, $\rho(\Pi D(s))$ is the maximum eigenvalue of $\Pi D(s)$, and so we compute

$$\begin{aligned} \rho(\Pi D(s)) &= \rho\left(\begin{bmatrix} \pi_{11}\mathrm{e}^{sx_1} & \pi_{12}\mathrm{e}^{sx_2} \\ \pi_{21}\mathrm{e}^{sx_1} & \pi_{22}\mathrm{e}^{sx_2} \end{bmatrix}\right) \\ &= \frac{1}{2}\left(\pi_{11}\mathrm{e}^{sx_1} + \pi_{22}\mathrm{e}^{sx_2} + \sqrt{(\pi_{11}\mathrm{e}^{sx_1} - \pi_{22}\mathrm{e}^{sx_2})^2 + 4\pi_{12}\pi_{21}\mathrm{e}^{s(x_1+x_2)}}\right). \end{aligned}$$

Setting this quantity equal to $1/(1-p)$ and solving for $s$ gives a unique positive solution $\alpha$ provided that $\max\{x_1, x_2\} > 0$, and a unique negative solution $-\beta$ provided that $\min\{x_1, x_2\} < 0$ (Remark 3.3).

## 3.2 Refinements

In the proof of Theorem 3.1 we applied Theorem 2.1 in conjunction with Remark 2.3 to establish the limits (3.1) and (3.2). By exploiting the sharp bounds in (2.4), we can improve Theorem 3.1 as follows. (A similar statement holds for the lower tail, which we omit.)

**Theorem 3.2.** *Let everything be as in Theorem 3.1, and let $x, y$ be the right and left Perron vectors of $\Pi D(\alpha)$ whose entries sum to 1. Let $z = s + it$ and suppose that the matrix pencil $A(z) = I - (1-p)\Pi D(z)$ is invertible on the axis of convergence $\operatorname{Re} z = \alpha$ for $t \in (-b, b)$ except at $t = 0$. Then*

$$\frac{2\pi C/b}{\mathrm{e}^{2\pi\alpha/b} - 1} \leq \liminf_{w\to\infty} \mathrm{e}^{\alpha w}\mathrm{P}(W_T > w) \leq \limsup_{w\to\infty} \mathrm{e}^{\alpha w}\mathrm{P}(W_T > w) \leq \frac{2\pi C/b}{1 - \mathrm{e}^{-2\pi\alpha/b}}, \tag{3.9}$$

*where*

$$C = \frac{p\omega_0^\top x}{(1-p)^2 y^\top \Pi D'(\alpha) x}. \tag{3.10}$$

*In particular, if $z = \alpha$ is the unique point of noninvertibility of $A(z)$ on the axis of convergence $\operatorname{Re} z = \alpha$, then*

$$\lim_{w\to\infty} \mathrm{e}^{\alpha w}\mathrm{P}(W_T > w) = \frac{C}{\alpha}. \tag{3.11}$$

*Proof.* We showed in the proof of Theorem 3.1 that $A(z)$ is holomorphic on the interior of $\mathcal{S}$ with an isolated point of noninvertibility at $z = \alpha$. If also $A(z)$ is invertible on $\{z = \alpha + it : 0 < |t| < b\}$ then we may deduce that $A(z)$ is holomorphic on an open set containing $\{z = \alpha + it : |t| < b\}$ with a unique point of noninvertibility at $z = \alpha$. Therefore, Proposition C.1 implies that $A(z)^{-1}$ and thus $p\omega_0^\top A(z)^{-1}\Pi D(z)e$ are holomorphic on an open set containing $\{z = \alpha + it : 0 < |t| < b\}$. In view of (3.5) the latter function constitutes a continuous extension of the mgf of $W_T$ to the union of its strip of holomorphicity and $\{z = \alpha + it : 0 < |t| < b\}$; moreover, it was shown in the proof of Theorem 3.1 that the singularity at $\alpha$ is a simple pole with residue $-C$. By applying Theorem 2.1 with $A = C$ (note that the lower and upper bounds in (2.4) are increasing and decreasing respectively in $B$, so that if they are valid for $B > b$ then they are valid for $b$) we obtain (3.9).

If $z = \alpha$ is the unique point of noninvertibility of $A(z)$ on $\operatorname{Re} z = \alpha$, then we can take $b$ arbitrarily large. Letting $b \to \infty$, both sides of (3.9) converge to $C/\alpha$ and we obtain (3.11). $\square$

The following theorem characterizes the upper tail behavior of a geometrically stopped random growth process. It is more generally applicable than (a discrete-time reformulation of) the main result of Reed (2001) because the growth rate process is permitted to be non-Gaussian and Markovian; on the other hand, we only characterize the upper tail of the stopped process, not its entire distribution. A similar statement holds for the lower tail, which we omit.

**Theorem 3.3.** *Let everything be as in Theorem 3.1. Let $S_0 > 0$ be a random variable independent of $W_T$ satisfying $\mathrm{E}[S_0^{\alpha+\epsilon}] < \infty$ for some $\epsilon > 0$, and define the random variable $S = S_0 \mathrm{e}^{W_T}$. Then*

$$\lim_{s \to \infty} \frac{\log \mathrm{P}(S > s)}{\log s} = -\alpha. \tag{3.12}$$

*If in addition $z = \alpha$ is the unique point of noninvertibility of $A(z)$ on the axis of convergence $\operatorname{Re} z = \alpha$, then*

$$\lim_{s \to \infty} s^\alpha \mathrm{P}(S > s) = \frac{C}{\alpha} \mathrm{E}[S_0^\alpha], \tag{3.13}$$

*where $C$ is defined as in Theorem 3.2.*

*Proof.* Since $S_0$ is independent of $W_T$, the mgf of $\log S$ is the product of the mgfs of $\log S_0$ and $W_T$. The moment condition $\mathrm{E}[S_0^{\alpha+\epsilon}] < \infty$ ensures that the mgf of $\log S_0$ is holomorphic on the strip $0 < \operatorname{Re} z < \alpha + \epsilon$, and clearly it cannot have a zero at $\alpha$. The proof of Theorem 3.1 establishes that the mgf of $W_T$ is holomorphic on the strip $0 < \operatorname{Re} z < \alpha$ with a simple pole at $\alpha$. Therefore the mgf of $\log S$ is also holomorphic on the strip $0 < \operatorname{Re} z < \alpha$ with a simple pole at $\alpha$, and applying Theorem 2.1 in conjunction with Remark 2.3 we find that $\lim_{x \to \infty} x^{-1} \log \mathrm{P}(\log S > x) = -\alpha$. The limit (3.12) follows by substituting $x = \log s$.

If $z = \alpha$ is the unique point of noninvertibility of $A(z)$ on $\operatorname{Re} z = \alpha$ then from Theorem 3.2 we have $\lim_{s \to \infty} s^\alpha \mathrm{P}(e^{W_T} > s) = C/\alpha$. The upper tail of the random variable $\mathrm{e}^{W_T}$ is therefore regularly varying with index $-\alpha$, and so by applying Breiman (1965)'s lemma we obtain

$$\lim_{s \to \infty} s^\alpha \mathrm{P}(S > s) = \left( \lim_{s \to \infty} s^\alpha \mathrm{P}(\mathrm{e}^{W_T} > s) \right) \left( \lim_{s \to \infty} \frac{\mathrm{P}(S_0 \mathrm{e}^{W_T} > s)}{\mathrm{P}(\mathrm{e}^{W_T} > s)} \right) = \frac{C}{\alpha} \mathrm{E}[S_0^\alpha],$$

as claimed. □

Theorem 3.3 provides a stronger characterization of the tail behavior of $S$ when

$z = \alpha$ is the unique point of noninvertibility of $A(z)$ on the axis of convergence $\operatorname{Re} z = \alpha$. The following theorem shows that this is the case except when the support of $X_1$ is a subset of an evenly-spaced grid. For a scalar $c$, let $c\mathbb{Z} = \{cm : m \in \mathbb{Z}\}$, where $\mathbb{Z}$ is the set of integers.

**Theorem 3.4.** *Let everything be as in Theorem 3.3 and $\tau \neq 0$. If $A(\alpha + i\tau)$ is noninvertible, then there exist $a_1, \dots, a_N \in \mathbb{R}$ such that $\operatorname{supp}(X_1 | J_1 = n) \subset a_n + \frac{2\pi}{\tau}\mathbb{Z}$ for all $n \in \mathcal{N}$, with $a_n = 0$ if $\pi_{nn} > 0$. Conversely, if $\operatorname{supp}(X_1 | J_1 = n) \subset \frac{2\pi}{\tau}\mathbb{Z}$ for all $n \in \mathcal{N}$, then $A(\alpha + i\tau)$ is noninvertible.*

*Proof.* We divide our proof into three steps.

*Step 1. Let $X$ be a random variable and suppose that $\left|\mathrm{E}[\mathrm{e}^{(\alpha+i\tau)X}]\right| = \mathrm{E}[\mathrm{e}^{\alpha X}] < \infty$ for some $\tau > 0$. Then $\operatorname{supp} X \subset a + \frac{2\pi}{\tau}\mathbb{Z}$ for some $a \in \mathbb{R}$. If $\mathrm{E}[\mathrm{e}^{(\alpha+i\tau)X}] = \mathrm{E}[\mathrm{e}^{\alpha X}]$, then we can take $a = 0$.*

Using the triangle inequality, we obtain

$$\mathrm{E}[\mathrm{e}^{\alpha X}] = \left|\mathrm{E}[\mathrm{e}^{(\alpha+i\tau)X}]\right| \leq \mathrm{E}\left[\left|\mathrm{e}^{(\alpha+i\tau)X}\right|\right] = \mathrm{E}[\mathrm{e}^{\alpha X}].$$

Fix any support point $a \in \operatorname{supp} X$. Since the triangle inequality holds with equality, it must be the case that

$$0 \leq \frac{\mathrm{e}^{(\alpha+i\tau)x}}{\mathrm{e}^{(\alpha+i\tau)a}} = \mathrm{e}^{(\alpha+i\tau)(x-a)}$$

for all $x \in \operatorname{supp} X$. Therefore, for each $x \in \operatorname{supp} X$, there exists $m \in \mathbb{Z}$ such that $\tau(x - a) = 2\pi m$, so $\operatorname{supp} X \subset a + \frac{2\pi}{\tau}\mathbb{Z}$.

If $\mathrm{E}[\mathrm{e}^{(\alpha+i\tau)X}] = \mathrm{E}[\mathrm{e}^{\alpha X}]$, letting $p_m = \mathrm{P}(X = a + 2\pi m/\tau)$ we obtain

$$\begin{aligned}\sum_{m=-\infty}^{\infty} p_m \mathrm{e}^{\alpha(a+2\pi m/\tau)} &= \mathrm{E}[\mathrm{e}^{\alpha X}] = \mathrm{E}[\mathrm{e}^{(\alpha+i\tau)X}] \\ &= \sum_{m=-\infty}^{\infty} p_m \mathrm{e}^{(\alpha+i\tau)(a+2\pi m/\tau)} = \mathrm{e}^{i\tau a} \sum_{m=-\infty}^{\infty} p_m \mathrm{e}^{\alpha(a+2\pi m/\tau)}.\end{aligned}$$

Therefore $\mathrm{e}^{i\tau a} = 1$, so there exists $m_0 \in \mathbb{Z}$ such that $a = 2\pi m_0/\tau$. We thus obtain

$$\operatorname{supp} X \subset a + \frac{2\pi}{\tau}\mathbb{Z} = \frac{2\pi}{\tau}(m_0 + \mathbb{Z}) = \frac{2\pi}{\tau}\mathbb{Z}.$$

*Step 2. If $A(\alpha + i\tau)$ is noninvertible, then there exist $a_1, \dots, a_N \in \mathbb{R}$ such that $\operatorname{supp}(X_1 | J_1 = n) \subset a_n + \frac{2\pi}{\tau}\mathbb{Z}$ for all $n \in \mathcal{N}$, with $a_n = 0$ if $\pi_{nn} > 0$.*

Without loss of generality we may assume $\tau > 0$. If $A(\alpha + i\tau)$ is noninvertible then $1/(1-p)$ is an eigenvalue of $\Pi D(\alpha + i\tau)$. Therefore $1/(1-p) \leq \rho(D(\alpha + i\tau))$. Since $D(\alpha + i\tau)$ is a diagonal matrix whose $n$-th diagonal element is $\mathrm{E}\left[\mathrm{e}^{(\alpha+i\tau)X_1} \,\middle|\, J_1 = n\right]$, by the triangle inequality we obtain $|D(\alpha + i\tau)| \leq D(\alpha)$ entry-wise. Since $\Pi$ is a nonnegative matrix, by Proposition B.2(ii) we obtain

$$\frac{1}{1-p} \leq \rho(\Pi D(\alpha + i\tau)) \leq \rho(\Pi \, |D(\alpha + i\tau)|) \leq \rho(\Pi D(\alpha)) = \frac{1}{1-p}.$$

Since all inequalities hold with equality and $\Pi D(\alpha)$ is irreducible, and noting that $1/(1-p)$ is an eigenvalue of $\Pi D(\alpha + i\tau)$, by Proposition B.2(iii) we have $\Pi D(\alpha + i\tau) = \Theta \Pi D(\alpha) \Theta^{-1}$, where $\Theta = \mathrm{diag}(\mathrm{e}^{i\theta_1}, \ldots, \mathrm{e}^{i\theta_N})$ for some $\theta_1, \ldots, \theta_N \in \mathbb{R}$. Comparing the $(m, n)$-th entry, we have

$$\pi_{mn} \, \mathrm{E}\left[\mathrm{e}^{(\alpha+i\tau)X_1} \,\middle|\, J_1 = n\right] = \pi_{mn} \mathrm{e}^{i(\theta_m - \theta_n)} \, \mathrm{E}\left[\mathrm{e}^{\alpha X_1} \,\middle|\, J_1 = n\right] \tag{3.14}$$

for all $m, n \in \mathcal{N}$.

Since $\Pi$ is irreducible, for each $n \in \mathcal{N}$ there exists $m \in \mathcal{N}$ such that $\pi_{mn} > 0$. Taking the absolute value of (3.14) and dividing by $\pi_{mn} > 0$, we obtain $\left|\mathrm{E}\left[\mathrm{e}^{(\alpha+i\tau)X_1} \,\middle|\, J_1 = n\right]\right| = \mathrm{E}\left[\mathrm{e}^{\alpha X_1} \,\middle|\, J_1 = n\right]$ for all $n \in \mathcal{N}$. It now follows from the previous step that there exist $a_1, \ldots, a_N \in \mathbb{R}$ such that $\mathrm{supp}(X_1 | J_1 = n) \subset a_n + \frac{2\pi}{\tau}\mathbb{Z}$ for all $n \in \mathcal{N}$.

If $\pi_{nn} > 0$, setting $m = n$ in (3.14) and dividing by $\pi_{nn} > 0$, we obtain $\mathrm{E}\left[\mathrm{e}^{(\alpha+i\tau)X_1} \,\middle|\, J_1 = n\right] = \mathrm{E}\left[\mathrm{e}^{\alpha X_1} \,\middle|\, J_1 = n\right]$. Hence by the previous step we have $\mathrm{supp}(X_1 | J_1 = n) \subset \frac{2\pi}{\tau}\mathbb{Z}$, so we can take $a_n = 0$.

*Step 3. If* $\mathrm{supp}(X_1 | J_1 = n) \subset \frac{2\pi}{\tau}\mathbb{Z}$ *for all* $n \in \mathcal{N}$*, then* $A(\alpha + i\tau)$ *is noninvertible.*

Fix any $n \in \mathcal{N}$ and let $p_m = \mathrm{P}(X_1 = 2\pi m/\tau | J_1 = n)$ for $m \in \mathbb{Z}$. Then

$$\mathrm{E}\left[\mathrm{e}^{(\alpha+i\tau)X_1} \,\middle|\, J_1 = n\right] = \sum_{m=-\infty}^{\infty} p_m \mathrm{e}^{(\alpha+i\tau)2\pi m/\tau} = \sum_{m=-\infty}^{\infty} p_m \mathrm{e}^{2\pi\alpha m/\tau} = \mathrm{E}\left[\mathrm{e}^{\alpha X_1} \,\middle|\, J_1 = n\right].$$

Since $n$ is arbitrary, we have $D(\alpha + i\tau) = D(\alpha)$, and consequently $A(\alpha + i\tau) = A(\alpha)$. We know that $A(z)$ is noninvertible at $\alpha$, so it must also be noninvertible at $\alpha + i\tau$. $\square$

**Remark 3.5.** The condition $\mathrm{supp}(X_1 | J_1 = n) \subset \frac{2\pi}{\tau}\mathbb{Z}$ for all $n \in \mathcal{N}$ is sufficient but not necessary for $A(\alpha + i\tau)$ to be noninvertible. To see this, note that if $\mathrm{supp}(X_1 | J_1 = n) \subset a_n + \frac{2\pi}{\tau}\mathbb{Z}$ for all $n$, then (3.14) (which is sufficient for non-invertibility) is equivalent to $\pi_{mn} \mathrm{e}^{i\tau a_n} = \pi_{mn} \mathrm{e}^{i(\theta_m - \theta_n)}$. This equation holds, for example, if $N = 2$, $\pi_{11} = \pi_{22} = 0$, $\pi_{12} = \pi_{21} = 1$, $\tau = 2\pi$, $\theta_1 = 1$, $\theta_2 = -1$, $a_1 = -1/\pi$, and $a_2 = 1/\pi$. Then $\operatorname{supp}(X_1|J_1 = n) \subset \pm 1/\pi + \mathbb{Z}$.

The following example shows that if $A(z)$ is noninvertible at multiple points on the axis of convergence $\operatorname{Re} z = \alpha$, the upper tail of $S$ is not necessarily Paretian.

**Example 3.6** (Deterministic growth)**.** Let $X_t = \mu > 0$, a constant. Wold and Whittle (1957) used a continuous-time model with deterministic growth and mortality rates to investigate the tail behavior of wealth distributions. In this case we have $A(z) = 1 - (1-p)e^{\mu z}$, and the mgf of $W_T = \mu T$ is given by

$$\mathrm{E}[e^{zW_T}] = \sum_{t=1}^{\infty} p(1-p)^{t-1} e^{\mu z t} = \frac{pe^{\mu z}}{1-(1-p)e^{\mu z}}. \tag{3.15}$$

The right abscissa of convergence is $\alpha = -\log(1-p)/\mu > 0$. Setting the denominator in (3.15) equal to zero, we obtain poles at $z = \alpha + 2\pi i k/\mu$ for $k \in \mathbb{Z}$. These poles are the points of noninvertibility of $A(z)$ on the axis of convergence $\operatorname{Re} z = \alpha$. Let $S = e^{W_T}$. If $t \le w/\mu < t + 1$, then

$$\mathrm{P}(W_T > w) = \mathrm{P}(T > w/\mu) = (1-p)^t = (1-p)^{\lfloor w/\mu \rfloor}.$$

Therefore

$$\mathrm{P}(S > s) = \mathrm{P}(\log S > \log s) = \mathrm{P}(W_T > \log s) = (1-p)^{\lfloor (\log s)/\mu \rfloor}.$$

Clearly $\log \mathrm{P}(S > s)/\log s \to \log(1-p)/\mu = -\alpha$ as $s \to \infty$, consistent with Theorem 3.3. However,

$$s^{\alpha}\mathrm{P}(S > s) = s^{-\log(1-p)/\mu}(1-p)^{\lfloor (\log s)/\mu \rfloor} = (1-p)^{\lfloor (\log s)/\mu \rfloor - (\log s)/\mu}$$

oscillates between 1 and $\frac{1}{1-p}$ as $s \to \infty$, so the tail is not Paretian. These limits are precisely the bounds we obtain in (3.9) by setting $C = p/(\mu(1-p))$ and $b = 2\pi/\mu$.

# 4 Wealth distribution with heterogeneous agents

As an application of Theorems 3.1–3.3, in this section we characterize the tail behavior of the wealth distribution in heterogeneous-agent models with non-Gaussian, Markovian shocks. We consider two cases, models with idiosyncratic endowment and investment risks.

## 4.1 Wealth distribution in CARA Huggett economies

The first application is a standard Huggett (1993) economy, where agents are subject to uninsurable endowment risk and trade a risk-free asset. We present an analytical solution to this model with Markov shocks that exploits constant absolute risk aversion (CARA) preferences.

### 4.1.1 Model

There is a continuum of agents with additive CARA utility

$$\mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t), \tag{4.1}$$

where $0 < \beta < 1$ is the discount factor and $u(c) = -\mathrm{e}^{-\gamma c}/\gamma$ has constant absolute risk aversion $\gamma > 0$.[8] Agents can borrow or save at gross risk-free rate $R > 1$ determined in equilibrium and are subject to uninsurable idiosyncratic income risk. The income at period $t$ is denoted by $y_t$, which evolves according to some Markov process that is iid across agents. Letting $w_t$ be the financial wealth at the beginning of time $t$ excluding current income, the budget constraint of a typical agent is

$$w_{t+1} = R(w_t - c_t + y_t). \tag{4.2}$$

The Bellman equation is

$$V(w, y) = \max_c \{u(c) + \beta\, \mathrm{E}[V(R(w - c + y), y')]\}. \tag{4.3}$$

A stationary equilibrium consists of a risk-free rate and quantities (distributions of consumption and risk-free assets) such that (i) agents optimize, (ii) risk-free bond market clears, and (iii) all aggregate quantities are constant over time.

### 4.1.2 Individual decision

Under an AR(1) specification for the income process, Wang (2003) obtains a closed-form solution of the utility maximization problem. Below, we consider the case in which the income follows an arbitrary finite-state Markov chain.

[8] We focus on CARA preferences because it is tractable with additive shocks (Calvet, 2001; Wang, 2003, 2007; Angeletos and Calvet, 2005, 2006). Also, since the CARA utility is defined on the entire real line, as is common with these models we assume that consumption can be negative.

Suppose that the income process $\{y_t\}_{t=0}^{\infty}$ is exogenous and follows an arbitrary finite-state Markov chain. Let $s \in S = \{1, \dots, S\}$ denote the states, $P = (p_{ss'})$ be the (irreducible) transition probability matrix, and $y = (y_s)_{s=1}^{S}$ be the vector of incomes. In this case the state variables are wealth $w$ and the exogenous state $s$, so let $V_s(w)$ be the value function in state $s$. The following proposition characterizes the optimal consumption rule in closed-form up to a system of $S$ nonlinear equations in $S$ variables.

**Proposition 4.1.** *The value function and optimal consumption rule are*

$$V_s(w) = -\frac{1}{\gamma a} \mathrm{e}^{-\gamma(aw+b_s)}, \tag{4.4a}$$

$$c_s(w) = aw + b_s, \tag{4.4b}$$

*where $a = 1 - 1/R$ and $b = (b_s)_{s=1}^{S}$ uniquely solves*

$$b = \left(1 - \frac{1}{R}\right) y - \frac{1}{\gamma R} \log\left(\beta R P \mathrm{e}^{-\gamma b}\right). \tag{4.5}$$

Although closed-form expressions for $b$ do not generally exist, we can still show the following comparative statics in the Markov chain economy.

**Proposition 4.2.** *Let $b(\beta, \gamma, y, R)$ be the constant terms $b = (b_s)_{s=1}^{S}$ in Proposition 4.1 corresponding to the discount factor $\beta$, absolute risk aversion $\gamma$, income levels $y \in \mathbb{R}^S$, and gross risk-free rate $R > 1$. Then the followings are true.*

(i) *If $\beta_1 < \beta_2$, then $b(\beta_1, \gamma, y, R) \gg b(\beta_2, \gamma, y_2, R)$, that is, the agent saves more if he is more patient.*

(ii) *If $\beta R < 1$ and $\gamma_1 < \gamma_2$, then $b(\beta, \gamma_1, y, R) \gg b(\beta, \gamma_2, y, R)$, that is, higher risk aversion implies larger precautionary savings.*

(iii) *If $y_1 \ll y_2$, then $b(\beta, \gamma, y_1, R) \gg b(\beta, \gamma, y_2, R)$, that is, the agent consumes more if income increases in all states.*

(iv) *For all $s$, we have $b_s(\beta, \gamma, y, R) \to \infty$ as $R \downarrow 1$.*

The intuition that $b_s \to \infty$ as $R \downarrow 1$ is straightforward. When the gross-risk free rate approaches 1 (net interest rate approaches 0), the present discounted value of future labor income diverges to infinity. Since the agent becomes infinitely rich, he wishes to consume an infinite amount.

### 4.1.3 General equilibrium and wealth distribution

Next we consider the general equilibrium. Consider a Huggett (1993) economy, where there are a continuum of independent agents receiving income and trading a risk-free asset in zero net supply. We restrict the analysis to a stationary equilibrium, where the risk-free rate is constant over time.

Using the budget constraint (4.2), the consumption rule (4.4b), and $a = 1 - 1/R$, individual wealth evolves according to

$$w' = R(w - (aw + b_s) + y_s) = w + R(y_s - b_s). \tag{4.6}$$

Since wealth is a random walk in levels, with infinitely lived agents there is no stationary wealth distribution. In order to obtain a stationary distribution, we assume that agents enter/exit the economy at constant probability $p$ as in Yaari (1965) and Blanchard (1985). Because agents survive each period with probability $1-p$, the effective discount factor is $\tilde{\beta} = \beta(1-p)$. Suppose that there are perfectly competitive insurance companies that offer annuities and life insurances. Let $\tilde{R}$ be the effective risk-free rate that agents face. If an agent saves or borrows 1, the position grows to $\tilde{R}$ next period if the agent survives, and 0 if he dies (an agent who dies with positive assets surrender to the insurance company; the debt of an agent who dies with negative assets is covered by life insurance). Letting $A$ be aggregate savings, by accounting we obtain

$$RA = (1-p)\tilde{R}A + p0 \iff \tilde{R} = \frac{R}{1-p}.$$

To derive the market clearing condition, let $C, W$ be aggregate consumption and wealth in the stationary equilibrium. Since there is no aggregate savings in a Huggett economy, we have $W = 0$. Aggregating the optimal consumption rule (4.4b) across agents, aggregate consumption is

$$C = aW + \pi' b = \pi' b.$$

Furthermore, aggregate consumption must equal aggregate income, so $C = \pi' y$. Combining these two equations, the equilibrium condition is

$$\pi' b = \pi' y. \tag{4.7}$$

The following theorem shows the existence of equilibrium.

**Theorem 4.3.** *There exists an equilibrium in the CARA Huggett economy. The gross risk-free rate satisfies* $1 - p < R < 1/\beta$.

The following theorem shows that the stationary wealth distribution in a CARA Huggett economy has exponential tails.

**Theorem 4.4.** *Let* $P = (p_{ss'})$ *be the transition probability matrix,* $y = (y_1, \dots, y_S)^\top$ *be the vector of incomes,* $b = (b_1, \dots, b_S)^\top$ *be as in Proposition 4.1, and* $\tilde{R} = \frac{R}{1-p}$ *be the effective risk-free rate in equilibrium. For any* $z \in \mathbb{R}$*, let* $\lambda(z) > 0$ *be the Perron root of the matrix* $P(\operatorname{diag} \mathrm{e}^{z\tilde{R}(y-b)})$*. Then there exist unique* $\alpha_1, \alpha_2 > 0$ *such that* $\lambda(\alpha_1) = \lambda(-\alpha_2) = \frac{1}{1-p}$*. The stationary wealth distribution has exponential tails with upper tail exponent* $\alpha_1$ *and lower tail exponent* $\alpha_2$*.*

*Proof.* By (4.6) with $R$ replaced by $\tilde{R}$, the individual wealth increases by $\tilde{R}(y_s - b_s)$ in state $s$. Since the vector of moment generating functions is $M(z) = \mathrm{e}^{z\tilde{R}(y-b)}$, by Theorem 3.1, the wealth distribution has an exponential upper tail with exponent $\alpha_1 > 0$ that satisfies $\lambda(\alpha_1) = \frac{1}{1-p}$, where $\lambda(z)$ is the Perron root of $P(\operatorname{diag} M(z))$. The argument for the lower tail is similar. □

In the quantitative macro literature, it is well known that heterogeneous-agent models with idiosyncratic income risk alone have difficulty in matching the wealth distribution (Huggett, 1996; Castañeda et al., 2003). Theorem 4.4 provides a theoretical explanation of these numerical findings: idiosyncratic income risk (additive shocks) alone can only generate exponential tails—not Pareto tails—when the income process is light-tailed. Benhabib et al. (2017) show that when the income process is fat-tailed, then the tail exponent of the wealth distribution is identical to that of the income process, which is counterfactual. These theoretical results suggest that researchers need to go beyond models with idiosyncratic income risk in order to understand the upper tail of the wealth distribution.

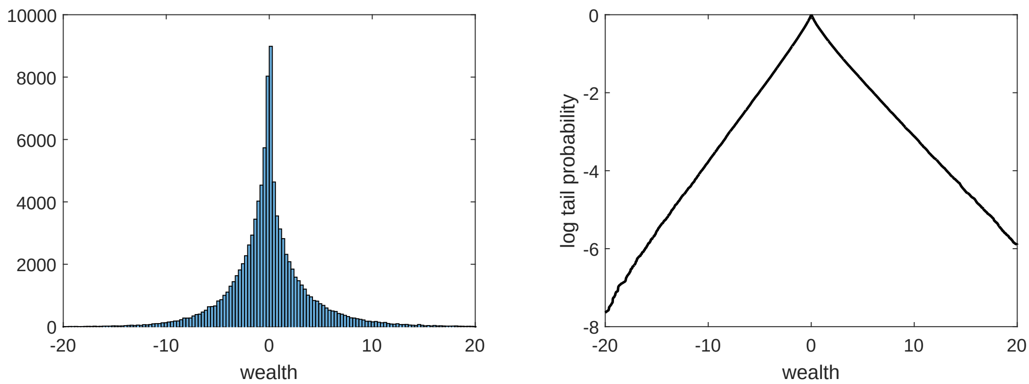
#### 4.1.4 Numerical example

As a numerical example, let the discount factor be $\beta = 0.96$, absolute risk aversion $\gamma = 3$, death probability $p = 0.025$, and suppose that log income $x_t = \log y_t$ is AR(1) with mean zero,

$$x_t = \rho x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2),$$

where $\rho = 0.9$ and $\sigma = 0.1$. To convert this process into a finite state Markov chain, we apply the discretization method in Farmer and Toda (2017) to discretize the AR(1) process $\{x_t\}$ using an even-spaced grid with $S = 9$ points. Letting $x_s$ be the grid point in state $s$, the income in state $s$ is defined by $y_s = \mathrm{e}^{x_s}$. We then solve for the equilibrium risk-free rate $R = 1.0350$ using (4.5) and (4.7). Finally, we compute the tail exponents of the wealth distribution using Theorem 4.4. The theoretical values are $\alpha_1 = 0.2857$ and $\alpha_2 = 0.3669$.

Figure 4.1a shows the histogram of the wealth from a simulation with 100,000 agents. The wealth distribution seems to have exponential tails. Letting $F(x)$ be the empirical cumulative distribution function, Figure 4.1b shows the tail probability ($\frac{F(x)}{F(0)}$ for $x < 0$, and $\frac{1-F(x)}{1-F(0)}$ for $x > 0$) in a semi log plot. Since Figure 4.1b shows straight line patterns, the tails are exponential. Estimating the tail exponents by maximum likelihood (using the largest 5% observations in each tail), we obtain $\widehat{\alpha}_1 = 0.2805$ and $\widehat{\alpha}_2 = 0.3657$, which are close to the theoretical values.




**(a)** Histogram. **(b)** Log tail probability.

**Figure 4.1:** Wealth distribution of the CARA Huggett economy.

## 4.2 Wealth distribution with idiosyncratic investment risk

As a second application, we consider a neoclassical growth model with idiosyncratic investment risk similar to Angeletos (2007) but with Markov shocks as in Krebs (2006) and Toda (2014).

### 4.2.1 Model

There is a continuum of households. A household consists of an entrepreneur (wife) and a worker (husband). The worker supplies one unit of labor ineastically and earns a competitive wage. The entrepreneur uses her own capital, hires labor, and produces outputs using a constant-returns-to-scale technology.

Households have Epstein-Zin utility

$$U_t = \begin{cases} \left((1-\beta)c_t^{1-1/\varepsilon} + \beta\,\mathrm{E}_t[U_{t+1}^{1-\gamma}]^{\frac{1-1/\varepsilon}{1-\gamma}}\right)^{\frac{1}{1-1/\varepsilon}}, & (\varepsilon \neq 1) \\ \exp\left((1-\beta)\log c + \beta \log\left(\mathrm{E}_t[U_{t+1}^{1-\gamma}]^{\frac{1}{1-\gamma}}\right)\right), & (\varepsilon = 1) \end{cases} \tag{4.8}$$

where $c_t$ is consumption, $U_t$ is continuation utility, $\beta \in (0,1)$ is the discount factor, $\gamma > 0$ is the relative risk aversion coefficient, and $\varepsilon > 0$ is the elasticity of intertemporal substitution.

The shock to the individual household follows a finite state Markov chain, which is iid across agents. Let $s \in S = \{1, \dots, S\}$ denote the Markov state with (irreducible) transition probability matrix $P = (p_{ss'})$. Let $F_s(k,n)$ be the production function in state $s$ net of depreciation, which is assumed to be homogeneous of degree 1.[9] Letting $R_f > 1$ be the gross risk-free rate and $\omega > 0$ be the wage (which are constant in stationary equilibrium), the household budget constraint is

$$c + k' + b' = F_s(k,n) - \omega n + R_f b + \omega, \tag{4.9}$$

where $c$ is consumption, $k$ is capital, $b$ is bond holdings, and $n$ is labor demand.

The timing is as follows. At the beginning of period $t$, the current state $s_t$ realizes and the entrepreneur chooses the labor demand $n_t$ and production takes place. The household then chooses consumption $c_t$ and the next period's capital and bond holdings $k_{t+1}, b_{t+1}$. The household's objective is to maximize the recursive utility (4.8) subject to the budget constraint (4.9). A stationary equilibrium consists of prices (risk-free rate and wage) and quantities (distributions of consumption, capital, and bonds) such that (i) agents optimize, (ii) labor market clears, (iii) risk-free bond market clears, and (iv) all aggregate quantities are constant over time.

### 4.2.2 Individual decision

We can characterize the household's optimal decision rule by reducing the model to a consumption-portfolio problem. Let

$$h = \omega \sum_{t=1}^{\infty} \frac{1}{R_f^t} = \frac{\omega}{R_f - 1}$$

[9] A typical example is the Cobb-Douglas production function $F_s(k,n) = A_s k^\alpha n^{1-\alpha} + (1-\delta_s)k$, where $A_s > 0$ is productivity and $\delta_s \leq 1$ is depreciation rate in state $s$.

be the present discounted value of future wages, or human capital. Adding $h$ to both sides of (4.9) and letting $h' = h$, we obtain

$$c + k' + b' + h' = F_s(k, n) - \omega n + R_f(b + h) =: w, \tag{4.10}$$

where $w$ is total wealth. Since labor demand $n$ can be chosen after observing the current state $s$, we have $n = \arg\max_n [F_s(k, n) - \omega n]$. Since $F_s$ is homogeneous of degree 1, letting $y = k/n$ be the capital-labor ratio and $f_s(y) = F_s(y, 1)$, by the first-order condition we obtain

$$\omega = f_s(y) - y f_s'(y). \tag{4.11}$$

Define the portfolio share of capital by $\theta = \frac{k'}{w-c} \geq 0$. Then the portfolio share of risk-free assets (bonds and human capital) is $\frac{b'+h'}{w-c} = 1-\theta$. Using the homogeneity of $F$ and letting $y_s$ be the capital-labor ratio $y$ that solves (4.11), the next period's wealth is

$$\begin{aligned} w' &= F_{s'}(k', n') - \omega n' + R_f(b' + h') = k' f_{s'}'(y_{s'}) + R_f(b' + h') \\ &= (f_{s'}'(y_{s'})\theta + R_f(1 - \theta))(w - c) =: R_{s'}(\theta)(w - c), \end{aligned} \tag{4.12}$$

where $R_{s'}(\theta)$ is the gross portfolio return realized in state $s'$ given $\theta$. Therefore the household's problem reduces to maximizing the recursive utility (4.8) subject to the budget constraint (4.12), which is a Merton (1971)-type optimal consumption-portfolio problem. The following proposition characterizes its solution.

**Proposition 4.5.** *Consider the problem of maximizing the recursive utility (4.8) subject to the budget constraint (4.12). Then the value function is linear, $V_s(w) = b_s w$, where the coefficient $b_s > 0$ satisfies*

$$b_s = \begin{cases} \left((1-\beta)^\varepsilon + \beta^\varepsilon \left(\max_\theta \mathrm{E}\left[(b_{s'} R_{s'}(\theta))^{1-\gamma} \mid s\right]^{\frac{1}{1-\gamma}}\right)^{\varepsilon-1}\right)^{\frac{1}{\varepsilon-1}}, & (\varepsilon \neq 1) \\ (1-\beta)^{1-\beta} \left(\max_\theta \mathrm{E}\left[(b_{s'} R_{s'}(\theta))^{1-\gamma} \mid s\right]^{\frac{1}{1-\gamma}}\right)^{\beta}. & (\varepsilon = 1) \end{cases} \tag{4.13}$$

*The optimal portfolio $\theta_s$ and capital-labor ratio $y_s$ are the argmax. The optimal consumption rule is $c = (1-\beta)^\varepsilon b_s^{1-\varepsilon} w$.*

*Proof.* Similar to Toda (2014, Corollary 7). $\square$

### 4.2.3 General equilibrium and wealth distribution

Next we consider the general equilibrium. As before, assume that households are born and die with probability $p \in (0,1)$ each period. Letting $R_f$ be the risk-free rate, the effective risk-free rate is $\tilde{R}_f = \frac{R_f}{1-p}$. Newborn households are endowed with no capital. Therefore the initial wealth of newborn households consists of the current and all future discounted wages, which is

$$w_0 = \omega + h = \omega \sum_{t=0}^{\infty} \frac{1}{\tilde{R}_f^t} = \frac{\omega \tilde{R}_f}{\tilde{R}_f - 1}. \tag{4.14}$$

The initial state of newborn households is drawn from the ergodic distribution, which is denoted by $\pi = (\pi_1, \dots, \pi_S)^\top$. In order to characterize the equilibrium, let us introduce some notations.

Given the wage $\omega > 0$ and the effective gross risk-free rate $\tilde{R}_f > 1$, let $b_s, \theta_s, y_s$ be the coefficient of value function, optimal portfolio, and capital-labor ratio determined as in Proposition 4.5. Let

$$R_{s'}(\theta_s) = f'_{s'}(y_{s'})\theta_s + \tilde{R}_f(1-\theta_s)$$

be the portfolio return from state $s$ to $s'$,

$$G_{ss'} = (1-(1-\beta)^\varepsilon b_s^{1-\varepsilon}) R_{s'}(\theta_s)$$

be the gross growth rate of household wealth from state $s$ to $s'$, and $G = (G_{ss'})$ be the $S \times S$ matrix of growth rates. Let $w_s > 0$ be the aggregate wealth held by households in state $s$ in the steady state. Since agents die with probability $p$ and newborn agents start with wealth $w_0$ and fraction $\pi_{s'}$ of them are in state $s'$, by accounting we have

$$w_{s'} = p\pi_{s'} w_0 + (1-p) \sum_{s=1}^{S} p_{ss'} G_{ss'} w_s, \quad s' = 1, \dots, S.$$

Letting $w = (w_1, \dots, w_S)^\top$ and collecting the above equations into a vector, we get

$$w^\top = p w_0 \pi^\top + (1-p) w^\top (P \odot G) \iff w = p w_0 (I - (1-p)(P \odot G)^\top)^{-1} \pi \tag{4.15}$$

provided that $(1-p)\rho(P \odot G) < 1$.

The following theorem characterizes the stationary equilibrium.

**Theorem 4.6.** *Let everything be as above. Then the equilibrium conditions are the first-order condition with respect to labor* (4.11) *($S$ equations),* (4.13) *($S$ equations), the first-order conditions of* (4.13) *with respect to $\theta$ ($S$ equations), and*

$$w_0 = \sum_{s=1}^{S} (1 - (1-\beta)^{\varepsilon} b_s^{1-\varepsilon}) w_s (1-\theta_s), \tag{4.16a}$$

$$1 = \sum_{s=1}^{S} (1 - (1-\beta)^{\varepsilon} b_s^{1-\varepsilon}) w_s \frac{\theta_s}{y_s}, \tag{4.16b}$$

*where $w_0$ is the initial wealth* (4.14) *and $w_1, \dots, w_S$ are determined as in* (4.15)*.*

Equations (4.16a) and (4.16b) are the market clearing conditions for the risk-free asset and labor, respectively. The unknown variables are $b = (b_s)_{s=1}^S$, $\theta = (\theta_s)_{s=1}^S$, $y = (y_s)_{s=1}^S$, $\omega$, and $\tilde{R}_f$, of which there are $3S+2$. Since there are $3S+2$ equilibrium conditions, we can compute the equilibrium by numerically solving this system of nonlinear equations.

Finally, we characterize the tails of the stationary wealth distribution. Because the wealth growth rate $G_{ss'}$ depends on two consecutive states, the conditional independence assumption in Theorem 3.1 does not hold. However, we can still apply this result by expanding the state space. For notational simplicity, suppose that there are two Markov states ($S = 2$). (The general case is similar.) Let $N = S^2 = 4$ and label the states as follows:

| $n$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $(s, s')$ | $(1,1)$ | $(1,2)$ | $(2,1)$ | $(2,2)$ |

Define the $N \times N$ transition probability matrix $\Pi$, vector of growth rates $g$, and matrix of conditional moment generating function of log growth rates $D(z)$ by

$$\Pi = \begin{bmatrix} p_{11} & p_{12} & 0 & 0 \\ 0 & 0 & p_{21} & p_{22} \\ p_{11} & p_{12} & 0 & 0 \\ 0 & 0 & p_{21} & p_{22} \end{bmatrix}, \qquad g = \begin{bmatrix} G_{11} \\ G_{12} \\ G_{21} \\ G_{22} \end{bmatrix}, \qquad D(z) = \operatorname{diag}(\mathrm{e}^{z \log g}).$$

To understand why $\Pi$ is the correct transition probability matrix, consider the $(1,2)$ element of $\Pi = (\pi_{nn'})$. This element $\pi_{12}$ is the transition probability from $1 = n_t = (s_{t-1}, s_t) = (1,1)$ to $2 = n_{t+1} = (s_t, s_{t+1}) = (1,2)$, which is $p_{12}$, the transition probability from $s_t = 1$ to $s_{t+1} = 2$. The other elements are similar. By

Theorem 3.3, the stationary wealth distribution has Pareto tails with upper and lower tail exponents $\alpha_1, \alpha_2 > 0$ such that

$$\rho(\Pi D(\alpha_1)) = \rho(\Pi D(-\alpha_2)) = \frac{1}{1-p}. \tag{4.17}$$

### 4.2.4 Numerical example

As a numerical example, let the discount factor be $\beta = 0.95$, relative risk aversion $\gamma = 2$, elasticity of intertemporal substitution $\varepsilon = 1$, and death probability $p = 0.02$. The production function is

$$F_s(k, n) = Ak^\alpha n^{1-\alpha} + (1 - \delta_s)k,$$

where $A = 1$, $\alpha = 0.38$, and $\log(1 - \delta_s)$ is a discrete AR(1) process with unconditional mean $-\delta = -0.08$, persistence $\rho = 0.3$, and conditional volatility $\sigma = 0.2$. As before, we use the Farmer and Toda (2017) method to discretize this process ($S = 5$ points).

Since $A$ does not depend on $s$, the optimal capital-labor ratio satisfies

$$0 = \frac{\partial}{\partial n}[F_s(k, n) - \omega n] = A(1 - \alpha)y^\alpha - \omega \iff y = \left(\frac{\omega}{A(1-\alpha)}\right)^{\frac{1}{\alpha}},$$

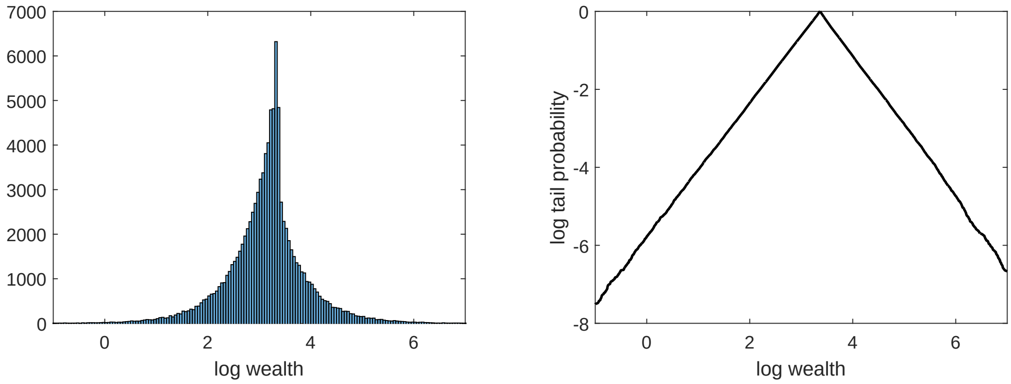
which does not depend on $s$. Given $\omega$ and $\tilde{R}_f$, we can solve for $(b_s)$ and $(\theta_s)$ using Proposition 4.5. Therefore the equilibrium conditions reduce to two equations (4.16a) and (4.16b) in two unknowns, $\omega$ and $\tilde{R}_f$.

The equilibrium wage is $\omega = 1.4694$, the effective gross risk-free rate is $\tilde{R}_f = 1.0535$ ($R_f = 1.0324$), and the optimal portfolio is

$$(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5) = (0, 0, 0, 0.5497, 1.3446).$$

($\theta_s = 0$ for $s = 1, 2, 3$ means that low productive households do not operate the technology and invest 100% in the risk-free asset.) The Pareto exponents computed from (4.17) are $\alpha_1 = 1.7540$ and $\alpha_2 = 1.7144$.

Figure 4.2a shows the histogram of the log wealth from a simulation with 100,000 agents. The log wealth distribution seems to have exponential tails. Letting $F(x)$ be the empirical cumulative distribution function, Figure 4.2b shows the tail probability ($\frac{F(x)}{F(0)}$ for $x < 0$, and $\frac{1-F(x)}{1-F(0)}$ for $x > 0$) in a semi log plot. Since Figure 4.2b shows straight line patterns, the tails are exponential. Estimating the tail exponents by maximum likelihood (using the largest 5% observations in each tail), we obtain $\widehat{\alpha}_1 = 1.8305$ and $\widehat{\alpha}_2 = 1.7141$, which are close to the theoretical values.




**(a)** Histogram. **(b)** Log tail probability.

**Figure 4.2:** Log wealth distribution of the Angeletos economy.

# 5 Conclusion

It has long been conjectured that random growth models robustly generate power law tails. For example, in the now classic paper on power law, Gabaix (1999, footnote 13) writes:

> Let each city grow at an arbitrary mean rate. [. . . ] [I]t does not matter if this mean rate is time varying[, which] is a conjecture that we firmly believe to be true. [. . . ] However, we could not find any argument in the mathematical literature—here we deal with Markov chains with time-varying transition matrices—to help us establish this rigorously.

In this paper we have presented a theorem that affirmatively resolves the robustness conjecture of power law tails. It does not matter whether the growth rate process is non-Gaussian or Markovian. While the details of the underlying process matter *quantitatively* in the sense that they affect the magnitude of the Pareto exponent, they do not matter *qualitatively*: the tail behavior is always Pareto.

# A Proof of Tauberian theorem

The proof of Theorem 2.1 uses a technique developed in Graham and Vaaler (1981). Define

$$E(x) = \begin{cases} \mathrm{e}^{-x} & \text{for } x \geq 0, \\ 0 & \text{for } x < 0. \end{cases}$$

Fix $\lambda > 0$, let $\omega = 2\pi/\lambda$, and let $K_\lambda(x)$ and $k_\lambda(x)$ be continuous functions of $x$ satisfying

$$\begin{aligned} K_\lambda(x) &= \left(\frac{\sin \lambda x/2}{\lambda/2}\right)^2 \left[\sum_{n=0}^{\infty} \frac{\mathrm{e}^{-n\omega}}{(x-n\omega)^2} - \sum_{n=1}^{\infty} \mathrm{e}^{-n\omega}\left(\frac{1}{x-n\omega} - \frac{1}{x}\right)\right], \\ k_\lambda(x) &= K_\lambda(x) - \left(\frac{\sin \lambda x/2}{\lambda x/2}\right)^2, \end{aligned}$$

for all real $x$ that are not integer multiples of $\omega$. Note that this validly defines $K_\lambda(x)$ and $k_\lambda(x)$ for all real $x$ since $\sin \lambda x/2 \sim (-1)^n \lambda(x - n\omega)/2$ around $x = n\omega$ for each integer $n$. We have the following lemma.

**Lemma A.1.** *$K_\lambda$ and $k_\lambda$ satisfy the following properties.*

(i) *$k_\lambda(x) \leq E(x) \leq K_\lambda(x)$ for all real $x$.*

(ii) *$K_\lambda$ and $k_\lambda$ are absolutely integrable, with Fourier transforms $\widehat{K}_\lambda$ and $\widehat{k}_\lambda$ supported on $[-\lambda, \lambda]$ and satisfying*

$$\widehat{K}_\lambda(0) = \frac{\omega}{1 - \mathrm{e}^{-\omega}}, \quad \widehat{k}_\lambda(0) = \frac{\omega}{\mathrm{e}^{\omega} - 1}.$$

*Proof.* See Korevaar (2004, Proposition 5.2). □

The following proof of Theorem 2.1 is an adaptation of arguments appearing in Graham and Vaaler (1981), Korevaar (2004, ch. 5) and Nakagawa (2007).

**Proof of Theorem 2.1.** Let $F$ be the cdf of $X$, and fix $\sigma \in (0, \alpha)$. Since $E(\sigma(y - x))\mathrm{e}^{\sigma y} = \mathrm{e}^{\sigma x} \mathbb{1}(y \geq x)$, we have

$$\mathrm{e}^{\sigma x} \mathrm{P}(X \geq x) = \int_{-\infty}^{\infty} E(\sigma(y - x))\mathrm{e}^{\sigma y}\, \mathrm{d}F(y).$$

Noting that $|E(\sigma(y - x))\mathrm{e}^{\sigma y}| \leq 1 \vee \mathrm{e}^{\alpha x}$ and $|E(\alpha(y - x))\mathrm{e}^{\sigma y}| \leq \mathrm{e}^{\alpha x}$, two applications of the dominated convergence theorem with dominating function $1 \vee \mathrm{e}^{\alpha x}$

(constant as a function of $y$) reveal that

$$\lim_{\sigma\uparrow\alpha}\int_{-\infty}^{\infty} E(\sigma(y-x))\mathrm{e}^{\sigma y}\,\mathrm{d}F(y) = \int_{-\infty}^{\infty} E(\alpha(y-x))\mathrm{e}^{\alpha y}\,\mathrm{d}F(y)$$
$$= \lim_{\sigma\uparrow\alpha}\int_{-\infty}^{\infty} E(\alpha(y-x))\mathrm{e}^{\sigma y}\,\mathrm{d}F(y).$$

Therefore,

$$\mathrm{e}^{\alpha x}\mathrm{P}(X \geq x) = \lim_{\sigma\uparrow\alpha}\int_{-\infty}^{\infty} E(\alpha(y-x))\mathrm{e}^{\sigma y}\,\mathrm{d}F(y). \tag{A.1}$$

For any $\lambda > 0$, Lemma A.1(i) implies that

$$\int_{-\infty}^{\infty} E(\alpha(y-x))\mathrm{e}^{\sigma y}\,\mathrm{d}F(y) \leq \int_{-\infty}^{\infty} K_\lambda(\alpha(y-x))\mathrm{e}^{\sigma y}\,\mathrm{d}F(y). \tag{A.2}$$

Lemma A.1(ii) implies that $K_\lambda(\alpha(y-x))\mathrm{e}^{\sigma y}$, viewed as a function of $x$, is absolutely integrable with Fourier transform

$$\int_{-\infty}^{\infty} K_\lambda(\alpha(y-x))\mathrm{e}^{\sigma y}\mathrm{e}^{-itx}\,\mathrm{d}x = \frac{1}{\alpha}\int_{-\infty}^{\infty} K_\lambda(w)\mathrm{e}^{\sigma y}\mathrm{e}^{-it(y-w/\alpha)}\,\mathrm{d}w$$
$$= \frac{1}{\alpha}\widehat{K}_\lambda(t/\alpha)\mathrm{e}^{(\sigma-it)y}.$$

Therefore, applying Fubini's theorem, we find that the Fourier transform of the upper bound in (A.2) is given by

$$\int_{-\infty}^{\infty} \frac{1}{\alpha}\widehat{K}_\lambda(t/\alpha)\mathrm{e}^{(\sigma-it)y}\,\mathrm{d}F(y) = \frac{1}{\alpha}\widehat{K}_\lambda(t/\alpha)M(\sigma-it).$$

This Fourier transform has support $[-\alpha\lambda, \alpha\lambda]$ by Lemma A.1(ii) so, applying the Fourier inversion theorem, we find that the upper bound in (A.2) satisfies

$$\int_{-\infty}^{\infty} K_\lambda(\alpha(y-x))\mathrm{e}^{\sigma y}\,\mathrm{d}F(y) = \frac{1}{2\pi}\int_{-\alpha\lambda}^{\alpha\lambda} \frac{1}{\alpha}\widehat{K}_\lambda(t/\alpha)M(\sigma-it)\mathrm{e}^{itx}\,\mathrm{d}t. \tag{A.3}$$

In deriving (A.3) we have only assumed that the Laplace transform of $F$ has right abscissa of convergence $\alpha \in (0,\infty)$. It must therefore be valid for the cdf

$$\tilde{F}(x) = \begin{cases} 1-\mathrm{e}^{-\alpha x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0, \end{cases}$$

whose Laplace transform is $\int_{-\infty}^{\infty} \mathrm{e}^{zx}\,\mathrm{d}\tilde{F}(x) = \frac{\alpha}{\alpha - z}$. In this case (A.3) specializes to

$$\int_{-\infty}^{\infty} K_\lambda(\alpha(y-x))\mathrm{e}^{\sigma y}\,\mathrm{d}\tilde{F}(y) = \frac{1}{2\pi}\int_{-\alpha\lambda}^{\alpha\lambda} \frac{1}{\alpha}\widehat{K}_\lambda(t/\alpha)\frac{\alpha}{\alpha-\sigma+it}\mathrm{e}^{itx}\,\mathrm{d}t. \tag{A.4}$$

Combining (A.2), (A.3) and (A.4), we obtain

$$\begin{aligned}\int_{-\infty}^{\infty} E(\alpha(y-x))\mathrm{e}^{\sigma y}\,\mathrm{d}F(y) &\leq \frac{1}{2\pi}\int_{-\alpha\lambda}^{\alpha\lambda} \frac{1}{\alpha}\widehat{K}_\lambda(t/\alpha)\left(M(\sigma-it)+\frac{A}{\sigma-it-\alpha}\right)\mathrm{e}^{itx}\,\mathrm{d}t \\ &\quad + \frac{A}{\alpha}\int_{-\infty}^{\infty} K_\lambda(\alpha(y-x))\mathrm{e}^{\sigma y}\,\mathrm{d}\tilde{F}(y) \\ &=: I_1(\sigma,x)+I_2(\sigma,x).\end{aligned}$$

Let $H(z)$ denote the continuous extension of $M(z)+A(z-\alpha)^{-1}$ to $\mathcal{S}_B^+$, which exists due to the definition of $B$. Fix $b\in(0,B)$. Since $H(z)$ is uniformly continuous on the compact set

$$\{z\in\mathbb{C} : 0\leq \operatorname{Re} z\leq\alpha, -b\leq\operatorname{Im} z\leq b\},$$

it must be the case that

$$\lim_{\sigma\uparrow\alpha}\sup_{t\in[-b,b]}|H(\sigma-it)-H(\alpha-it)| = 0.$$

Consequently, setting $\lambda = b/\alpha$ we obtain

$$\lim_{\sigma\uparrow\alpha} I_1(\sigma,x) = \frac{1}{2\pi}\int_{-b}^{b}\frac{1}{\alpha}\widehat{K}_{b/\alpha}(t/\alpha)H(\alpha-it)\mathrm{e}^{itx}\,\mathrm{d}t,$$

and the Riemann-Lebesgue lemma then yields $\lim_{x\to\infty}\lim_{\sigma\uparrow\alpha} I_1(\sigma,x) = 0$. Next, applying the dominated convergence theorem we obtain

$$\begin{aligned}\lim_{\sigma\uparrow\alpha} I_2(\sigma,x) &= \frac{A}{\alpha}\int_{-\infty}^{\infty} K_\lambda(\alpha(y-x))\mathrm{e}^{\alpha y}\,\mathrm{d}\tilde{F}(y) \\ &= A\int_0^{\infty} K_\lambda(\alpha(y-x))\,\mathrm{d}y = \frac{A}{\alpha}\int_{-\alpha x}^{\infty} K_\lambda(w)\,\mathrm{d}w.\end{aligned}$$

Letting $x\to\infty$ and noting that $\int_{-\infty}^{\infty} K_\lambda(w)\,\mathrm{d}w = \widehat{K}_\lambda(0) = \omega/(1-\mathrm{e}^{-\omega})$ by Lemma A.1(ii), we obtain

$$\lim_{x\to\infty}\lim_{\sigma\uparrow\alpha} I_2(\sigma,x) = \frac{A}{\alpha}\frac{\omega}{1-\mathrm{e}^{-\omega}} = \frac{2\pi A/b}{1-\mathrm{e}^{-2\pi\alpha/b}}.$$

Recalling (A.1), this establishes that

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

Since $b$ can be chosen arbitrarily close to $B$, the claimed upper bound in (2.4) follows. The proof of the lower bound is similar. Sharpness of the bounds was demonstrated in Remark 2.2. □

# B Properties of nonnegative matrices

In this appendix we collect properties of nonnegative matrices that are used throughout the paper. For a square (complex) matrix $A$, let $\rho(A)$ denote its spectral radius, i.e., the largest modulus of all eigenvalues, $|A|$ be the matrix obtained by taking the modulus of each element of $A$, and $\|A\|$ be any matrix norm.

For matrices $A, B$ of the same size, let $A \odot B$ denote the Hadamard (entry-wise) product, so for $A = (a_{mn})$ and $B = (b_{mn})$, we have $A \odot B = (a_{mn} b_{mn})$. For a nonnegative matrix $A = (a_{mn})$, let $A^{(\alpha)} = (a_{mn}^{\alpha})$ denote the matrix of entry-wise power. The following proposition shows that the spectral radius has a convexity property with respect to the Hadamard product.

**Proposition B.1** (Theorem 1, Elsner et al., 1988)**.** *Let $A, B$ be nonnegative square matrices of the same size and $\theta \in (0, 1)$. Then*

$$\rho(A^{(1-\theta)} \odot B^{(\theta)}) \le \rho(A)^{(1-\theta)} \rho(B)^{\theta}.$$

We call an $N \times N$ square matrix $A = (a_{mn})$ *irreducible* if for any $m, n$, there exist numbers $m = k_1, k_2, \dots, k_p = n$ $(1 \le p \le N)$ such that the entries $a_{k_1 k_2}, a_{k_2 k_3}, \dots, a_{k_{p-1} k_p}$ are all nonzero. There are many different ways to characterize irreducibility: see Theorem 6.2.24 of Horn and Johnson (1985).

**Proposition B.2.** *For $N \times N$ complex matrices $A, B$, the following are true.*

(i) $\rho(A) = \lim_{k\to\infty} \left\|A^k\right\|^{\frac{1}{k}}$.

(ii) *If $|B| \le A$, then $\rho(B) \le \rho(|B|) \le \rho(A)$.*

(iii) *If $A$ is nonnegative and irreducible, $|B| \le A$, $\rho(A) = \rho(B)$, and $\lambda = \mathrm{e}^{i\phi} \rho(B)$ is an eigenvalue of $B$, then there exist $\theta_1, \dots, \theta_N \in \mathbb{R}$ such that $B = \mathrm{e}^{i\phi} D A D^{-1}$, where $D = \mathrm{diag}(\mathrm{e}^{i\theta_1}, \dots, \mathrm{e}^{i\theta_N})$.*

*Proof.* Property (i) (which is known as Gelfand's spectral radius formula) is Corollary 5.6.14 of Horn and Johnson (1985). Property (ii) is Theorem 8.1.18 of Horn and Johnson (1985). Property (iii) is Theorem 8.4.5 of Horn and Johnson (1985). □

**Perron-Frobenius theorem.** Let $A$ be a square, nonnegative, and irreducible matrix. Then

(i) $\rho(A) > 0$,

(ii) $\rho(A)$ is an eigenvalue of $A$ (which is called the Perron root),

(iii) there exist positive vectors $x, y$ (called the right and left Perron vectors) such that $Ax = \rho(A)x$ and $y^\top A = \rho(A)y^\top$, and

(iv) $\rho(A)$ is an algebraically (and hence geometrically) simple eigenvalue of $A$.

*Proof.* See Theorem 8.4.4 of Horn and Johnson (1985). □

# C Simple poles of matrix pencil inverses

In this appendix we collect some properties of matrix pencil inverses that are used to prove the main results of our paper. A matrix pencil is a square (complex) matrix-valued function of a complex variable.

**Proposition C.1.** *Let $A(z)$ be an $n \times n$ matrix pencil depending holomorphically on $z \in \Omega$, where $\Omega$ is some open and connected subset of the complex plane. Suppose that $A(z)$ is invertible for some $z \in \Omega$. Then $A(z)$ has a meromorphic inverse $A(z)^{-1}$ on $\Omega$, with poles at the points of noninvertibility of $A(z)$. If $A(z_0)$ has rank $r < n$ for some $z_0 \in \Omega$, so that $A(z)^{-1}$ has a pole at $z = z_0$, then the following four conditions are equivalent.*

(i) *The pole at $z_0$ is simple.*

(ii) *The geometric and algebraic multiplicities of the unit eigenvalue of $I - A(z_0)$ are equal.*

(iii) *The $r \times r$ matrix $y^\top A'(z_0)x$ is invertible, where $x$ and $y$ can be any $n \times r$ matrices of full column rank such that $A(z_0)x = 0$ and $y^\top A(z_0) = 0$.*

(iv) *The complex vector space $\mathbb{C}^n$ is the direct sum of the column space of $A(z_0)$ and the image of the null space of $A(z_0)$ under $A'(z_0)$.*

*Under any of these equivalent conditions, the residue of $A(z)^{-1}$ at the simple pole at $z = z_0$ is equal to $x(y^\top A'(z_0)x)^{-1}y^\top$.*

*Proof.* Meromorphicity of $A(z)^{-1}$ when $A(z)$ is somewhere invertible was proved by Steinberg (1968). The equivalence of (i), (ii) and (iv) was proved by Howland (1971). Note that both of the authors just cited worked in a more general Banach space setting. The equivalence of (i) and (iii) and the residue formula were proved by Schumacher (1986); see also Schumacher (1991, pp. 562–563). □

**Remark C.1.** Proposition C.1 is closely related to the Granger-Johansen representation theorem (Engle and Granger, 1987; Johansen, 1991). The connection was first commented upon by Schumacher (1991), who observed that the condition used by Johansen (1991) to guarantee that the solution to a vector autoregressive equation is integrated of order one, or I(1), corresponded to condition (iii). Johansen (1995, Corollary 4.3) gave a reformulation of the I(1) condition corresponding to condition (ii), while Beare et al. (2017) gave a reformulation of the I(1) condition corresponding to condition (iv), and used it to develop an extension of the Granger-Johansen representation theorem to a general Hilbert space setting. Condition (iv) is not used to prove any of the results in this paper but we have included it for the sake of completeness.

# D Proofs of results in Section 4

### Proof of Proposition 4.1.

*Step 1. If (4.5) has a solution, then the policy (4.4) satisfies the Bellman equation.*

We prove by guess-and-verify. Substituting (4.4a) into the Bellman equation (4.3), we obtain

$$-\frac{1}{\gamma a}\mathrm{e}^{-\gamma(aw+b_s)} = \max_c \left\{-\frac{1}{\gamma}\mathrm{e}^{-\gamma c} - \frac{\beta}{\gamma a}\,\mathrm{E}\left[\mathrm{e}^{-\gamma(aR(w-c+y_s)+b_{s'})}\,\middle|\, s\right]\right\}. \tag{D.1}$$

The first-order condition with respect to $c$ is

$$\mathrm{e}^{-\gamma c} - \beta R\,\mathrm{E}\left[\mathrm{e}^{-\gamma(aR(w-c+y_s)+b_{s'})}\,\middle|\, s\right] = 0. \tag{D.2}$$

Substituting (D.2) into (D.1), we obtain

$$-\frac{1}{\gamma a}\mathrm{e}^{-\gamma(aw+b_s)} = -\frac{1}{\gamma a}\left(a+\frac{1}{R}\right)\mathrm{e}^{-\gamma c}. \tag{D.3}$$

Comparing the coefficients, (D.3) trivially holds if $a = 1 - 1/R$ and $c = aw + b_s$. In this case (D.1) holds, and so does the Bellman equation (4.3). To determine $b = (b_s)_{s=1}^S$, by the first-order condition we have

$$\begin{aligned}
\text{(D.2)} \iff \mathrm{e}^{-\gamma(aw+b_s)} &= \beta R\,\mathrm{E}\left[\mathrm{e}^{-\gamma(aw+(R-1)(-b_s+y_s)+b_{s'})} \,\middle|\, s\right] \\
\iff 1 &= \beta R\,\mathrm{E}\left[\mathrm{e}^{-\gamma(-Rb_s+(R-1)y_s+b_{s'})} \,\middle|\, s\right] \\
\iff b_s &= \left(1 - \frac{1}{R}\right) y_s - \frac{1}{\gamma R} \log \mathrm{E}\left[\beta R \mathrm{e}^{-\gamma b_{s'}} \,\middle|\, s\right] \\
\iff b_s &= \left(1 - \frac{1}{R}\right) y_s - \frac{1}{\gamma R} \log \left(\beta R \sum_{s'=1}^{S} p_{ss'} \mathrm{e}^{-\gamma b_{s'}}\right).
\end{aligned}$$

Expressing in matrix form, we obtain (4.5).

*Step 2. The system of equations (4.5) (for $s = 1, \dots, S$) admits a unique solution $b = (b_s)_{s=1}^S$.*

For $b \in \mathbb{R}^S$, define $T : \mathbb{R}^S \to \mathbb{R}^S$ by

$$Tb = \left(1 - \frac{1}{R}\right) y - \frac{1}{\gamma R} \log\left(\beta R P \mathrm{e}^{-\gamma b}\right).$$

Then (4.5) is equivalent to $Tb = b$, i.e., $b$ is a fixed point of $T$. To show that $T$ has a unique fixed point, we show that $T$ is a contraction using the sufficient condition in Blackwell (1965, Theorem 5). Clearly $b_1 \le b_2 \implies Tb_1 \le Tb_2$, so monotonicity holds. Furthermore, for any $t \ge 0$, we have

$$\begin{aligned}
T(b + t\mathbf{1}) &= \left(1 - \frac{1}{R}\right) y - \frac{1}{\gamma R} \log\left(\beta R P \mathrm{e}^{-\gamma(b+t\mathbf{1})}\right) \\
&= \left(1 - \frac{1}{R}\right) y - \frac{1}{\gamma R} \log\left(\beta R P \mathrm{e}^{-\gamma b}\right) + \frac{t}{R}\mathbf{1} \\
&= Tb + \frac{t}{R}\mathbf{1}.
\end{aligned}$$

Since $R > 1$, $T$ satisfies discounting, and hence by Blackwell's condition $T$ is a contraction. By the contraction mapping theorem, $T$ has a unique fixed point.

*Step 3. The policy (4.4) characterizes the solution to the utility maximization problem.*

Since the value function (4.4a) satisfies the Bellman equation, in order for the consumption rule (4.4b) to be optimal, it remains to show the transversality condition

$$\lim_{t \to \infty} \beta^t \,\mathrm{E}_0[V_{s_t}(w_t)] = 0, \tag{D.4}$$

where $w_t$ is the wealth at time $t$ determined by the budget constraint (4.2) and the consumption rule (4.4b).[10] Using the formula for the value function (4.4a) and the budget constraint (4.2), the continuation value in the next period is

$$V_{s'}(w') = -\frac{1}{\gamma a}\mathrm{e}^{-\gamma(aw'+b_{s'})} = -\frac{1}{\gamma a}\mathrm{e}^{-\gamma(aR(w-c+y_s)+b_{s'})}.$$

Taking the expectation conditional on $s$, by (D.2) we obtain

$$\mathrm{E}\left[V_{s'}(w') \mid s\right] = -\frac{1}{\gamma a\beta R}\mathrm{e}^{-\gamma c} = -\frac{1}{\gamma a\beta R}\mathrm{e}^{-\gamma(aw+b_s)} = \frac{1}{\beta R}V_s(w).$$

Iterating this equation from $t=0$ to $t$, we obtain

$$\beta^t\,\mathrm{E}_0[V_{s_t}(w_t)] = \frac{1}{R^t}V_{s_0}(w_0).$$

Since $R>1$, letting $t\to\infty$ we obtain the transversality condition (D.4). $\square$

**Proof of Proposition 4.2.** (i) Since $\gamma, y, R$ are fixed, let us suppress them. Let $T(\beta)$ be the contraction mapping in Proposition 4.1 associated with discount factor $\beta$, which is the right-hand side of (4.5). Then $T(\beta)b$ is (element-by-element) strictly decreasing in $\beta$. Take any $\beta_1<\beta_2$. Repeatedly applying $T(\beta_1), T(\beta_2)$ to the zero vector, by induction for all $n\ge 1$ we have $T^n(\beta_1)0 \gg T^n(\beta_2)0$. Letting $n\to\infty$ and using the contraction mapping theorem, we obtain $b(\beta_1)\ge b(\beta_2)$. Therefore

$$b(\beta_1) = T(\beta_1)b(\beta_1) \gg T(\beta_2)b(\beta_1) \ge T(\beta_2)b(\beta_2) = b(\beta_2).$$

(ii) Let $T(\gamma)$ be the contraction mapping in Proposition 4.1 associated with absolute risk aversion $\gamma$. Then

$$(T(\gamma)b)_s = \left(1-\frac{1}{R}\right)y_s - \frac{1}{\gamma R}\log(\beta R) + \frac{1}{R}\log \mathrm{E}\left[\mathrm{e}^{-\gamma b_{s'}} \,\middle|\, s\right]^{-1/\gamma}.$$

Since by assumption $\beta R<1$, we have $\log(\beta R)<0$, so the second term is strictly decreasing in $\gamma$. Letting $X_{s'} = \mathrm{e}^{b_{s'}}$, the third term equals $\frac{1}{R}\mathrm{E}\left[X_{s'}^{-\gamma} \,\middle|\, s\right]^{-1/\gamma}$, which is decreasing in $\gamma$ by Lemma D.1. Therefore $T(\gamma)b$ is (element-by-element) strictly decreasing in $\gamma$. The rest of the proof is similar.

(iii) Let $T(y)$ be the contraction mapping in Proposition 4.1 associated with

[10] See Kamihigashi (2014) and references therein for the sufficiency of the transversality condition.

income levels $y$. Then $T(y)b$ is (element-by-element) increasing in $y$. The rest of the proof is similar.

(iv) Let $\underline{b} = \min_s b_s$ and $\underline{y} = \min_s y_s$. Using $R > 1$ and the monotonicity of $T$, it follows from (4.5) that

$$\begin{aligned}
\underline{b} &\geq \left(1 - \frac{1}{R}\right) \underline{y} - \frac{1}{\gamma R} \sum_{s'=1}^{S} p_{ss'} \log(\beta R \mathrm{e}^{-\gamma \underline{b}}) \\
&= \left(1 - \frac{1}{R}\right) \underline{y} - \frac{1}{\gamma R} (\log(\beta R) - \gamma \underline{b}) \\
\iff \underline{b} &\geq \underline{y} - \frac{1}{\gamma(R-1)} \log(\beta R) \to \infty
\end{aligned}$$

as $R \downarrow 1$. □

**Proof of Theorem 4.3.** For notational simplicity, assume $p = 0$. The case $p > 0$ is completely analogous by using $\tilde{\beta} = \beta(1-p)$ and $\tilde{R} = \frac{R}{1-p}$ instead of $\beta, R$.

*Step 1. If an equilibrium exists, then $\beta R < 1$.*

By (4.5), we have

$$b_s = \left(1 - \frac{1}{R}\right) y_s - \frac{1}{\gamma R} \log \left(\beta R \sum_{s'=1}^{S} p_{ss'} \mathrm{e}^{-\gamma b_{s'}}\right).$$

Since $(p_{ss'})_{s'=1}^{S}$ are conditional probabilities, they are nonnegative and sum to 1. Since $-\log(\cdot)$ is strictly convex, it follows that

$$\begin{aligned}
b_s &\leq \left(1 - \frac{1}{R}\right) y_s - \frac{1}{\gamma R} \sum_{s'=1}^{S} p_{ss'} \log(\beta R \mathrm{e}^{-\gamma b_{s'}}) \\
&= \left(1 - \frac{1}{R}\right) y_s - \frac{1}{\gamma R} \left(\log(\beta R) + \sum_{s'=1}^{S} p_{ss'}(-\gamma b_{s'})\right) \\
\iff b &\leq \left(1 - \frac{1}{R}\right) y - \frac{1}{\gamma R} \left(\log(\beta R)\mathbf{1} - \gamma P b\right),
\end{aligned}$$

where the last inequality is element-by-element. Since $P$ is ergodic, at least one conditional distribution is non-degenerate, so at least one inequality is strict. Left multiplying the stationary distribution $\pi \gg 0$ as an inner product, it follows that

$$\begin{aligned}
&& \pi' b &< \left(1 - \frac{1}{R}\right) y - \frac{1}{\gamma R}(\log(\beta R) - \gamma \pi' P b) \\
&& &= \left(1 - \frac{1}{R}\right) y - \frac{1}{\gamma R}(\log(\beta R) - \gamma \pi' b) \qquad\qquad (\because \pi' = \pi' P) \\
&\iff& \pi' b &< \pi' y - \frac{1}{\gamma(R-1)} \log(\beta R). \qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(D.5)}
\end{aligned}$$

By the equilibrium condition (4.7), we have $\pi' b = \pi' y$, so

$$0 < -\frac{1}{\gamma(R-1)} \log(\beta R) \iff \beta R < 1$$

since $R > 1$ is necessary for equilibrium existence.

*Step 2. An equilibrium exists. The gross risk-free rate satisfies* $1 - p < R < 1/\beta$.

Let $b(R)$ be the value of $b = (b_s)_{s=1}^S$ implied by (4.5). By (4.7), the equilibrium condition is $g(R) := \pi' b(R) - \pi' y$. By Proposition 4.1, $b(R)$ is well-defined for $R > 1$, and it is smooth by the implicit function theorem. Therefore $g(R)$ is well-defined and continuous for $R > 1$. By Proposition 4.2, we have $\lim_{R \downarrow 1} g(R) = \infty$. Letting $R = 1/\beta$ in (D.5), we obtain

$$g(1/\beta) < -\frac{1}{\gamma(1/\beta - 1)} \log(1) = 0.$$

By the intermediate value theorem, there exists $R \in (1, 1/\beta)$ such that $g(R) = 0$. If $p > 0$, by the same argument as above we obtain $1 < \tilde{R} < 1/\tilde{\beta} \iff 1 - p < R < 1/\beta$. $\square$

**Lemma D.1.** *Let $X$ be an almost surely positive random variable. Suppose that* $\mathrm{E}[X^{r_i}] < \infty$ *for $i = 1, 2$, where $r_1 < 0 < r_2$. Then* $\mathrm{E}[\log X]$ *is finite. Letting*

$$\phi(r) = \begin{cases} \mathrm{E}[X^r]^{\frac{1}{r}}, & (r \neq 0) \\ \exp(\mathrm{E}[\log X]), & (r = 0) \end{cases}$$

*$\phi$ is finite, increasing, and continuous on $r \in [r_1, r_2]$.*

*Proof.*

*Step 1.* $\mathrm{E}[\log X]$ *is finite.*

Fix any $x > 0$. Since $f(r) = x^r$ is convex, we have

$$f(r) - f(0) \geq f'(0)(r - 0) \iff x^r - 1 \geq (\log x)r.$$

If $x \geq 1$, setting $r = r_2$ we obtain

$$0 \leq \log x \leq \frac{x^{r_2} - 1}{r_2}.$$

If $0 < x < 1$, setting $r = r_1$ we obtain

$$\frac{x^{r_1} - 1}{r_1} \leq \log x < 0.$$

Combining these two inequalities, we obtain

$$|\log x| \leq \max\left\{\frac{x^{r_1} + 1}{|r_1|}, \frac{x^{r_2} + 1}{|r_2|}\right\}.$$

Since $X > 0$ almost surely, setting $x = X$ and taking expectations, we obtain $\mathrm{E}[|\log X|] < \infty$ because $\mathrm{E}[X^{r_i}] < \infty$ for $i = 1, 2$.

*Step 2. $\phi(r)$ is finite, increasing, and continuous on $r \in [r_1, 0)$ and $r \in (0, r_2]$.*

Suppose that $0 < r < s \leq r_2$. Since $y^{s/r}$ is convex in $y$, applying Jensen's inequality to $Y = X^r$, we obtain

$$\mathrm{E}[X^r]^{\frac{s}{r}} = \mathrm{E}[Y]^{\frac{s}{r}} \leq \mathrm{E}[Y^{\frac{s}{r}}] = \mathrm{E}[X^s] \iff \mathrm{E}[X^r]^{\frac{1}{r}} \leq \mathrm{E}[X^s]^{\frac{1}{s}}.$$

Therefore $\phi(r) \leq \phi(s) \leq \phi(r_2) < \infty$. Considering the case $X \geq 1$ and $0 \leq X < 1$ separately, by the dominated convergence theorem $\mathrm{E}[X^r]$ is continuous in $r \in (0, r_2]$. Therefore $\phi(r)$ is continuous. A similar proof applies on $[r_1, 0)$ by considering $r_1 \leq s < r < 0$.

*Step 3. $\phi(r)$ is continuous on $r \in [r_1, r_2]$.*

It suffices to show that $\phi(r)$ is continuous at $r = 0$. Fix $x > 0$ and let

$$f(r) = \begin{cases} \frac{x^r - 1}{r}, & (r \neq 0) \\ \log x, & (r = 0) \end{cases}$$

which is continuous. Let us show that $f(r)$ is increasing on $r \in [r_1, r_2]$. If $0 < r <$

$s \le r_2$, since $x^r - 1$ is convex in $r$ and $r = (1 - r/s)0 + (r/s)s$, we have

$$x^r - 1 \le (1 - r/s)(x^0 - 1) + (r/s)(x^s - 1) \iff \frac{x^r - 1}{r} \le \frac{x^s - 1}{s}.$$

If $r_1 \le s < r < 0$, since $r = (1 - r/s)0 + (r/s)s$, we have

$$x^r - 1 \le (1 - r/s)(x^0 - 1) + (r/s)(x^s - 1) \iff \frac{x^s - 1}{s} \le \frac{x^r - 1}{r}.$$

Therefore $f$ is increasing.

Setting $x = X$, since $\mathrm{E}[X^{r_i}] < \infty$ for $i = 1, 2$, by the dominated convergence theorem we have

$$\lim_{r \to 0} \mathrm{E}\left[\frac{X^r - 1}{r}\right] = \mathrm{E}\left[\lim_{r \to 0} \frac{X^r - 1}{r}\right] = \mathrm{E}[\log X].$$

Therefore $\mathrm{E}[X^r] = 1 + r(\mathrm{E}[\log X] + o(1))$, so

$$\lim_{r \to 0} \phi(r) = \lim_{r \to 0} (1 + r(\mathrm{E}[\log X] + o(1)))^{\frac{1}{r}} = \exp(\mathrm{E}[\log X]). \qquad \square$$

**Proof of Theorem 4.6.** Consider a typical agent with portfolio $\theta$, wealth $w$, and consumption $c$. By definition, the risk-free asset holdings is $b' + h' = (1 - \theta)(w - c)$. Since the optimal consumption rule in state $s$ is $c = (1-\beta)^\varepsilon b_s^{1-\varepsilon} w$ and the aggregate wealth held by agents in state $s$ is $w_s$, the aggregate risk-free asset holdings held by agents in state $s$ is $(1 - (1 - \beta)^\varepsilon b_s^{1-\varepsilon}) w_s (1 - \theta_s)$. Adding this quantity across all states, since the risk-free asset is in zero net supply and aggregate human capital is $w_0$, we obtain (4.16a).

Similarly, the aggregate capital held by agents in state $s$ is $(1-(1-\beta)^\varepsilon b_s^{1-\varepsilon}) w_s \theta_s$. Since the capital-labor ratio in state $s$ is $y_s$, the aggregate labor demand is

$$\sum_{s=1}^{S} (1 - (1 - \beta)^\varepsilon b_s^{1-\varepsilon}) w_s \frac{\theta_s}{y_s},$$

which must be equal to the aggregate labor supply 1 in equilibrium. Therefore (4.16b) holds. $\square$

# References


Daron Acemoglu and Dan Cao. Innovation by entrants and incumbents. *Journal of Economic Theory*, 157:255–294, May 2015. doi:10.1016/j.jet.2015.01.001.

S. Rao Aiyagari. Uninsured idiosyncratic risk and aggregate saving. *Quarterly Journal of Economics*, 109(3):659–684, 1994. doi:10.2307/2118417.

George-Marios Angeletos. Uninsured idiosyncratic investment risk and aggregate saving. *Review of Economic Dynamics*, 10(1):1–30, January 2007. doi:10.1016/j.red.2006.11.001.

George-Marios Angeletos and Laurent-Emmanuel Calvet. Incomplete-market dynamics in a neoclassical production economy. *Journal of Mathematical Economics*, 41(4-5):407–438, August 2005. doi:10.1016/j.jmateco.2004.09.005.

George-Marios Angeletos and Laurent-Emmanuel Calvet. Idiosyncratic production risk, growth and the business cycle. *Journal of Monetary Economics*, 53: 1095–1115, 2006. doi:10.1016/j.jmoneco.2005.05.016.

Shuhei Aoki and Makoto Nirei. Zipf's law, Pareto's law, and the evolution of top incomes in the U.S. *American Economic Journal: Macroeconomics*, 9(3):36–71, July 2017. doi:10.1257/mac.20150051.

Yoshiyuki Arata. *Essays on Stochastic Macroeconomics*. PhD thesis, University of Tokyo, 2015.

Costas Arkolakis. A unified theory of firm selection and growth. *Quarterly Journal of Economics*, 131(1):89–155, 2016. doi:10.1093/qje/qjv039.

Robert L. Axtell. Zipf distribution of U.S. firm sizes. *Science*, 293(5536):1818–1820, September 2001. doi:10.1126/science.1062081.

Brendan K. Beare, Juwon Seo, and Won-Ki Seo. Cointegrated linear processes in Hilbert space. *Journal of Time Series Analysis*, 2017. doi:10.1111/jtsa.12251. Forthcoming.

Jess Benhabib and Alberto Bisin. Skewed wealth distributions: Theory and empirics. *Journal of Economic Literature*, 2017. Forthcoming.

Jess Benhabib, Alberto Bisin, and Shenghao Zhu. The distribution of wealth and fiscal policy in economies with finitely lived agents. *Econometrica*, 79(1): 123–157, January 2011. doi:10.3982/ECTA8416.

Jess Benhabib, Alberto Bisin, and Shenghao Zhu. The wealth distribution in Bewley economies with capital income risk. *Journal of Economic Theory*, 159 (A):489–515, September 2015. doi:10.1016/j.jet.2015.07.013.

Jess Benhabib, Alberto Bisin, and Shenghao Zhu. The distribution of wealth in the Blanchard-Yaari model. *Macroeconomic Dynamics*, 20:466–481, March 2016. doi:10.1017/S1365100514000066.

Jess Benhabib, Alberto Bisin, and Mi Luo. Earnings inequality and other determinants of wealth inequality. *American Economic Review Papers and Proceedings*, 107(5):593–597, May 2017. doi:10.1257/aer.p20171005.

Truman F. Bewley. A difficulty with the optimum quantity of money. *Econometrica*, 51(5):1485–1504, September 1983. doi:10.2307/1912286.

David Blackwell. Discounted dynamic programming. *Annals of Mathematical Statistics*, 36(1):226–235, February 1965. doi:10.1214/aoms/1177700285.

Olivier J. Blanchard. Debt, deficits, and finite horizons. *Journal of Political Economy*, 93(2):223–247, April 1985. doi:10.1086/261297.

Leonard Breiman. On some limit theorems similar to the arc-sin law. *Theory of Probability & Its Applications*, 10(2):323–331, 1965. doi:10.1137/1110037.

Laurent-Emmanuel Calvet. Incomplete markets and volatility. *Journal of Economic Theory*, 98(2):295–338, June 2001. doi:10.1006/jeth.2000.2720.

Ana Castañeda, Javier Díaz-Giménez, and José-Víctor Ríos-Rull. Accounting for the U.S. earnings and wealth inequality. *Journal of Political Economy*, 111(4): 818–857, August 2003. doi:10.1086/375382.

David G. Champernowne. A model of income distribution. *Economic Journal*, 63 (250):318–351, June 1953. doi:10.2307/2227127.

Alex Coad. The exponential age distribution and the Pareto firm size distribution. *Journal of Industry, Competition and Trade*, 10(3-4):389–395, 2010. doi:10.1007/s10842-010-0071-4.

Ludwig Elsner, Charles R. Johnson, and José António Dias da Silva. The Perron root of a weighted geometric mean of nonneagative matrices. *Linear and Multilinear Algebra*, 24(1):1–13, November 1988. doi:10.1080/03081088808817892.

Robert F. Engle and Clive W. J. Granger. Co-integration and error correction: Representation, estimation and testing. *Econometrica*, 55(2):251–276, March 1987. doi:10.2307/1913236.

Leland E. Farmer and Alexis Akira Toda. Discretizing nonlinear, non-Gaussian Markov processes with exact conditional moments. *Quantitative Economics*, 8 (2):651–683, July 2017. doi:10.3982/QE737.

Xavier Gabaix. Zipf's law for cities: An explanation. *Quarterly Journal of Economics*, 114(3):739–767, 1999. doi:10.1162/003355399556133.

Xavier Gabaix. Power laws in economics and finance. *Annual Review of Economics*, 1:255–293, 2009. doi:10.1146/annurev.economics.050708.142940.

Xavier Gabaix. Power laws in economics: An introduction. *Journal of Economic Perspectives*, 30(1):185–206, Winter 2016. doi:10.1257/jep.30.1.185.

Xavier Gabaix, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. The dynamics of inequality. *Econometrica*, 84(6):2071–2111, November 2016. doi:10.3982/ECTA13569.

Robert Gibrat. *Les Inégalités Économiques*. Librairie du Recueil Sirey, Paris, 1931.

Kristian Giesen and Jens Suedekum. City age and city size. *European Economic Review*, 71:193–208, October 2014. doi:10.1016/j.euroecorev.2014.07.006.

Kristian Giesen, Arndt Zimmermann, and Jens Suedekum. The size distribution across all cities—double Pareto lognormal strikes. *Journal of Urban Economics*, 68(2):129–137, September 2010. doi:10.1016/j.jue.2010.03.007.

Sidney West Graham and Jeffrey D. Vaaler. A class of extremal functions for the Fourier transform. *Transactions of the American Mathematical Society*, 265(1): 283–302, May 1981. doi:10.1090/S0002-9947-1981-0607121-1.

James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384, March 1989. doi:10.2307/1912559.

Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, 1985.

James S. Howland. Simple poles of operator-valued functions. *Journal of Mathematical Analysis and Applications*, 36(1):12–21, October 1971. doi:10.1016/0022-247X(71)90015-1.

Mark Huggett. The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of Economic Dynamics and Control*, 17(5-6):953–969, September-November 1993. doi:10.1016/0165-1889(93)90024-M.

Mark Huggett. Wealth distribution in life-cycle economies. *Journal of Monetary Economics*, 38(3):469–494, December 1996. doi:10.1016/S0304-3932(96)01291-3.

Søren Johansen. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59(6):1551–1580, November 1991. doi:10.2307/2938278.

Søren Johansen. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford, UK, 1995.

Michal Kalecki. On the Gibrat distribution. *Econometrica*, 13(2):161–170, April 1945. doi:10.2307/1907013.

Takashi Kamihigashi. Elementary results on solutions to the Bellman equation of dynamic programming: Existence, uniqueness, and convergence. *Economic Theory*, 56(2):251–273, 2014. doi:10.1007/s00199-013-0789-4.

Jacobus Cornelius Kapteyn. *Skew Frequency Curves in Biology and Statistics*. P. Noordhoff, Groningen, 1903.

Harry Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131(1):207–248, 1973. doi:10.1007/BF02392040.

Oren S. Klass, Ofer Biham, Moshe Levy, Ofer Malcai, and Sorin Solomon. The Forbes 400 and the Pareto wealth distribution. *Economics Letters*, 90(2):290–295, February 2006. doi:10.1016/j.econlet.2005.08.020.

Jacob Korevaar. *Tauberian Theory: A Century of Developments*. Number 329 in Gruncllehren der mathematischen Wissenschaften. Springer, Berlin, 2004.

Samuel Kotz, Tomasz J. Kozubowski, and Krzysztof Podgórski. *The Laplace Distribution and Generalizations*. Birkhäuser, Boston, 2001.

Tom Krebs. Recursive equilibrium in endogenous growth models with incomplete markets. *Economic Theory*, 29(3):505–523, 2006. doi:10.1016/S0165-1889(03)00062-9.

Jarl Waldemar Lindeberg. Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 15(1):211–225, 1922.

Eugene Lukacs. *Characteristic Functions*. Griffin, London, second edition, 1970.

Erzo G. J. Luttmer. Selection, growth, and the size distribution of firms. *Quarterly Journal of Economics*, 122(3):1103–1144, 2007. doi:10.1162/qjec.122.3.1103.

Francesco Mainardi and Sergei Rogosin. The origin of infinitely divisible distributions: From de Finetti's problem to Lévy-Khintchine formula. *Mathematical Methods in Economics and Finance*, 1:37–55, 2006.

Susanna C. Manrubia and Damián H. Zanette. Stochastic multiplicative processes with reset events. *Physical Review E*, 59(5):4945–4948, 1999. doi:10.1103/PhysRevE.59.4945.

Robert C. Merton. Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4):373–413, December 1971. doi:10.1016/0022-0531(71)90038-X.

Ante Mimica. Exponential decay of measures and Tauberian theorems. *Journal of Mathematical Analysis and Applications*, 40(1):266–285, August 2016. doi:10.1016/j.jmaa.2016.03.042.

Michael Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2):226–252, 2004. doi:10.1080/15427951.2004.10129088.

Kenji Nakagawa. Application of Tauberian theorem to the exponential decay of the tail probability of a random variable. *IEEE Transactions on Information Theory*, 53(9):3239–3249, September 2007. doi:10.1109/TIT.2007.903114.

Makoto Nirei and Shuhei Aoki. Pareto distribution of income in neoclassical growth models. *Review of Economic Dynamics*, 20:25–42, April 2016. doi:10.1016/j.red.2015.11.002.

Makoto Nirei and Wataru Souma. A two factor model of income distribution dynamics. *Review of Income and Wealth*, 53(3):440–459, September 2007. doi:10.1111/j.1475-4991.2007.00242.x.

Vilfredo Pareto. *Cours d'Économie Politique*. F. Rouge, Lausanne, 1896.

William J. Reed. The Pareto, Zipf and other power laws. *Economics Letters*, 74 (1):15–19, December 2001. doi:10.1016/S0165-1765(01)00524-9.

William J. Reed. On the rank-size distribution for human settlements. *Journal of Regional Science*, 42(1):1–17, February 2002. doi:10.1111/1467-9787.00247.

William J. Reed. The Pareto law of incomes—an explanation and an extension. *Physica A*, 319(1):469–486, March 2003. doi:10.1016/S0378-4371(02)01507-8.

William J. Reed and Barry D. Hughes. From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature. *Physical Review E*, 66(6):067103, December 2002. doi:10.1103/PhysRevE.66.067103.

William J. Reed and Barry D. Hughes. On the distribution of family names. *Physica A*, 319:579–590, March 2003. doi:10.1016/S0378-4371(02)01455-3.

William J. Reed and Murray Jorgensen. The double Pareto-lognormal distribution—a new parametric model for size distribution. *Communications in Statistics—Theory and Methods*, 33(8):1733–1753, 2004. doi:10.1081/STA-120037438.

R. Stuart G. Rutherford. Income distributions: A new model. *Econometrica*, 23 (3):277–294, July 1955. doi:10.2307/1910384.

Johannes M. Schumacher. Residue formulas for meromorphic matrices. In Christopher I. Byrnes and Anders Lindquist, editors, *Computational and Combinatorial Methods in Systems Theory*, pages 97–111. North-Holland, 1986.

Johannes M. Schumacher. System-theoretic trends in econometrics. In Athanasios C. Antoulas, editor, *Mathematical System Theory: The Influence of R. E. Kalman*, pages 559–577. Springer, Berlin, 1991. doi:10.1007/978-3-662-08546-2_32.

Herbert A. Simon and Charles P. Bonini. The size distribution of business firms. *American Economic Review*, 48(4):607–617, September 1958.

Stanly Steinberg. Meromorphic families of compact operators. *Archive for Rational Mechanics and Analysis*, 31(5):372–379, December 1968. doi:10.1007/BF00251419.

John Sutton. Gibrat's legacy. *Journal of Economic Literature*, 35(1):40–59, March 1997.

Alexis Akira Toda. Income dynamics with a stationary double Pareto distribution. *Physical Review E*, 83(4):046122, 2011. doi:10.1103/PhysRevE.83.046122.

Alexis Akira Toda. The double power law in income distribution: Explanations and evidence. *Journal of Economic Behavior and Organization*, 84(1):364–381, September 2012. doi:10.1016/j.jebo.2012.04.012.

Alexis Akira Toda. Incomplete market dynamics and cross-sectional distributions. *Journal of Economic Theory*, 154:310–348, November 2014. doi:10.1016/j.jet.2014.09.015.

Alexis Akira Toda. A note on the size distribution of consumption: More double Pareto than lognormal. *Macroeconomic Dynamics*, 2016. doi:10.1017/S1365100515000942.

Alexis Akira Toda and Kieran Walsh. The double power law in consumption and implications for testing Euler equations. *Journal of Political Economy*, 123(5): 1177–1200, October 2015. doi:10.1086/682729.

Alexis Akira Toda and Kieran James Walsh. Fat tails and spurious estimation of consumption-based asset pricing models. *Journal of Applied Econometrics*, 2017. doi:10.1002/jae.2564.

Philip Vermeulen. How fat is the top tail of the wealth distribution? *Review of Income and Wealth*, 2017. doi:10.1111/roiw.12279.

Neng Wang. Caballero meets Bewley: The permanent-income hypothesis in general equilibrium. *American Economic Review*, 93(3):927–936, June 2003. doi:10.1257/000282803322157179.

Neng Wang. An equilibrium model of wealth distribution. *Journal of Monetary Economics*, 54(7):1882–1904, October 2007. doi:10.1016/j.jmoneco.2006.11.005.

David Vernon Widder. *The Laplace Transform*. Princeton University Press, Princeton, NJ, 1941.

Herman O. A. Wold and Peter Whittle. A model explaining the Pareto distribution of wealth. *Econometrica*, 25(4):591–595, October 1957. doi:10.2307/1905385.

Menahem E. Yaari. Uncertain lifetime, life insurance, and the theory of the consumer. *Review of Economic Studies*, 32(2):137–150, April 1965. doi:10.2307/2296058.