Geometrically stopped Markovian random growth processes and Pareto tails

Brendan K. Beare and Alexis Akira Toda

Department of Economics, University of California, San Diego

December 6, 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 independent and identically distributed 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 formulas for Pareto exponents. We present two applications: (i) we show that the tails of the wealth distribution in a heterogeneous-agent dynamic general equilibrium model with idiosyncratic endowment risk decay exponentially, unlike models with investment risk where the tails may be Paretian, and (ii) we show that a random growth model for the population dynamics of Japanese prefectures is consistent with the observed Pareto exponent but only after allowing for Markovian dynamics.

Keywords: exponential tails, Gibrat's law, Pareto tails, power law, random growth, Tauberian theorem.

JEL codes: C46, C65, D30, D52, D58, R12.

1 Introduction

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

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

We thank Alberto Bisin, Xavier Gabaix, Émilien Gouin-Bonenfant, Makoto Nirei, Werner Ploberger, Jonathan Weinstein, and seminar participants at Queen's University, UCSB, UCSD, the University of Toronto, Washington University in St. Louis, and the 2017 Summer Workshop in Economic Theory at the Cowles Foundation for helpful comments and suggestions.

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 independent and identically distributed (iid) and T has mean $1/p$, the upper tail exponent α 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 the innovations conditional on the state variable and its transition probabilities; see equation (3.2) below.

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.[1] 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's law of proportional growth (Gibrat, 1931) and an exponential age distribution.[2] Suppose that the size S_t of an individual unit at age $t \geq 0$ follows a geometric Brownian motion initialized at some fixed 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 μ and σ 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

[1]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, 2017b), and wealth (Klass et al., 2006; Vermeulen, 2017).

[2]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).

distributed time. Reed (2001) showed that the pdf of this quantity is given by

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

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

$$\frac{\sigma^2}{2}\zeta^2 + \left(\mu - \frac{\sigma^2}{2}\right)\zeta - \eta = 0.\text{[3]} \tag{1.4}$$

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.[4]

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 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

[3]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 α on the first page, in which the numerator and the denominator are flipped.)

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

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),[5] 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). In Theorem 3.5, we derive comparative statics for the Pareto exponent with respect to perturbations in lifespan, growth, volatility, and persistence.

As one 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, both with Markovian 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 alone have difficulty in matching the (empirically fat) upper tail of the wealth distribution (see Benhabib et al.,

[5] 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*).

2017, for an excellent discussion of this issue). The latter case with idiosyncratic investment risk has already be known to generate Pareto tails under specific conditions (typically iid Gaussian shocks). We extend those results by permitting non-Gaussian shocks dependent on a Markov state variable and thus considerably expand the toolkit for applied economists.

As a second application of our results, in Section 5 we ask whether the historical population dynamics of Japanese prefectures are consistent with a Pareto cross-sectional size distribution. Using the population of 46 Japanese prefectures (excluding Okinawa) since 1873, we document that the power law emerged only around 1920 and the Pareto exponent has decreased over the century from about 3 in 1920 to about 1.3 in 2015 (except for a brief break during World War II). We fit a random growth model to the panel of prefecture populations. Using an iid random growth model, the implied Pareto exponent is 56.7, which is far larger than the value directly estimated from the cross-section (1.3). However, we show that by estimating a Markov switching model with three states, the implied Pareto exponent becomes 1.6, which is close to what we see in the data. This shows that incorporating Markovian dynamics into random growth models can be crucial for explaining tail exponents observed in data. Our theorems provide a technical tool to study such models.

1.1 Related literature

Important early contributions on generative mechanisms for power laws were made by Champernowne (1953), who proposed a model of income dynamics in which a power law emerges as a steady state equilibrium. Wold and Whittle (1957) found that a constant growth rate of wealth together with an exponential age distribution generates a Pareto upper tail, which anticipates the mechanism of Reed (2001). 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.[6]

[6]da Saporta (2005), Roitershtein (2007), and others have extended the Kesten (1973) result to Markovian environments using renewal theory. Our results are different primarily because (i) growth is proportional ($B_t = 0$), consistent with Gibrat's law, and (ii) we study the tails of a geometrically stopped nonstationary process, not the tails of the invariant distribution of a stationary (Kesten) process. Furthermore, our results are more generally applicable in economics because proportional growth arises in a wide variety of dynamic models, whereas the Kesten process is more rare.

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. Gabaix showed that augmenting the geometric Brownian motion law for city size evolution with a lower reflecting boundary generates a power law exponent slightly above one in the upper tail of the city size distribution, consistent with Zipf's law. Note that this generative mechanism is distinct from that of Reed (2001). 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).[7]

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 solutions to this problem was to introduce birth and death. In particular, Rutherford (1955) assumes an exponential age distribution. Since the income shocks are Gaussian in his model, it is a discrete-time version of the model 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 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 $-\beta \leq 0 \leq \alpha$ 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,

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

[7] 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. However, such processes were introduced by Lévy (1937) subsequent to the publication of Gibrat (1931).

exists if and only if

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

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.

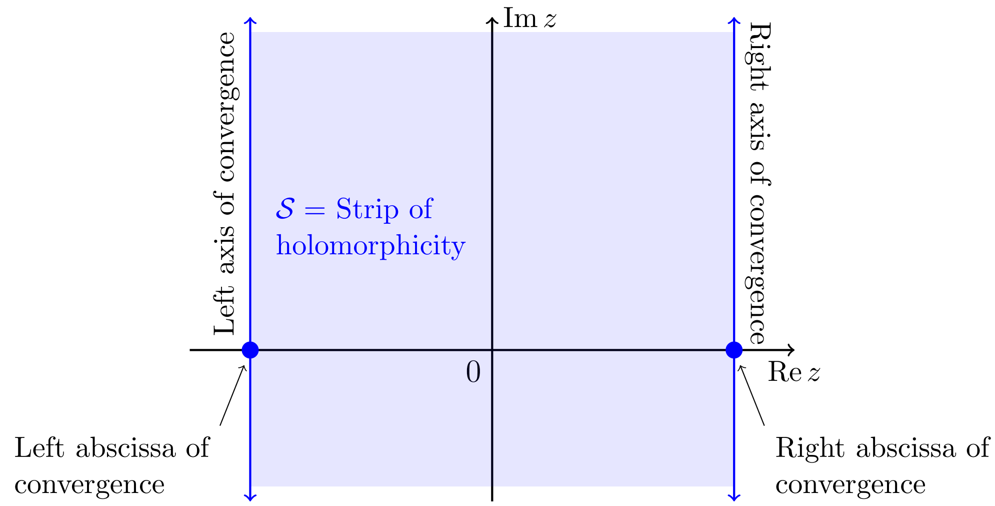


Figure 2.1: Region of convergence of the Laplace transform $M(z) = \int_{-\infty}^{\infty} e^{zx} \, dF(x)$.

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

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

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 $e^{\alpha x} P(X > x)$ and

$e^{\beta x}P(X < -x)$ 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 Appendix A (see Remark 2.5 below).

Theorem 2.1. *Let X be a real random variable and $M(z) = E[e^{zX}]$ its Laplace transform, with right abscissa of convergence α 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

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

where the bounds should be read as A/α 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

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

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 is given by

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

on its strip of holomorphicity $\mathcal{S}$. The zeros of the denominator $1-(1-p)e^z$ are $z_k = \alpha + 2\pi ik$, $k \in \mathbb{Z}$, where $\alpha = -\log(1-p) > 0$. The right abscissa of convergence is therefore α and the strip of holomorphicity is $\mathcal{S} = \{z \in \mathbb{C} : \operatorname{Re} z < \alpha\}$. The zeros z_k are singularities of $M(z)$ lying along its axis of convergence $\operatorname{Re} z = \alpha$. 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 α 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 $n \leq x < n+1$ 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

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

for all $x \geq x_0$. 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 α 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 introduced 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

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 $x_1, \ldots, x_N$ then it is a hidden Markov process: just set $Y_t \equiv (x_1, \ldots, x_N)^\top$ (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 $\pi_{nn'} = \mathrm{P}(J_1 = n' \mid J_0 = n)$. We assume that Π is irreducible. Define the set

$$\mathcal{I} := \left\{ s \in \mathbb{R} : (\forall n \in \mathcal{N})\, \mathrm{E}\left[\mathrm{e}^{sX_1} \,\middle|\, J_1 = n \right] < \infty \right\}, \tag{3.1}$$

and for $s \in \mathcal{I}$, let $D(s)$ be the $N \times N$ diagonal matrix with n-th diagonal entry equal to $\mathrm{E}\left[\mathrm{e}^{sX_1} \,\middle|\, J_1 = n \right]$. 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 empirically plausible conditions the geometric sum W_T has exponential tails, and characterizes their decay rates in terms of Π, D and p. That characterization is illustrated in Figure 3.1. For a square matrix A, we let $\rho(A)$ denote the spectral radius of A.

Theorem 3.1. *The set $\mathcal{I}$ is convex and contains zero. There can be at most one positive value of $s \in \mathcal{I}$ that solves the equation*

$$\rho(\Pi D(s)) = \frac{1}{1-p}. \tag{3.2}$$

Suppose that a positive solution α to (3.2) exists in the interior of $\mathcal{I}$. Then

$$\lim_{w \to \infty} \frac{1}{w} \log \mathrm{P}(W_T > w) = -\alpha. \tag{3.3}$$

Similarly, there can be at most one negative value of $s \in \mathcal{I}$ that solves equation (3.2). Suppose that a negative solution $-\beta$ exists in the interior of $\mathcal{I}$. Then

$$\lim_{w \to \infty} \frac{1}{w} \log \mathrm{P}(W_T < -w) = -\beta. \tag{3.4}$$

Proof. We divide our proof into five steps, stated in italics as claims to be proved.

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

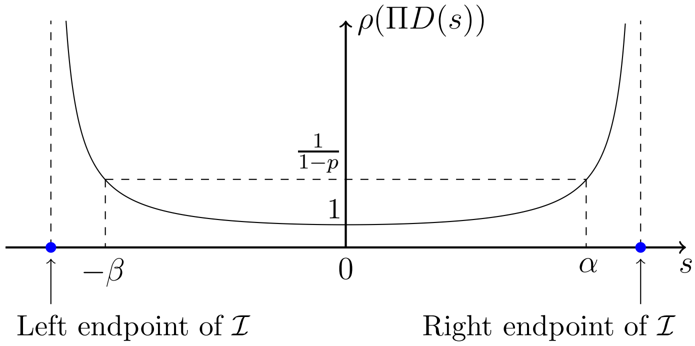


Figure 3.1: Determination of α and β.

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

$$\begin{aligned}
&\mathrm{E}\left[\mathrm{e}^{((1-\theta)s_1+\theta s_2)X_1} \,\middle|\, J_1 = n\right] \\
&\qquad \le \left(\mathrm{E}\left[\mathrm{e}^{s_1 X_1} \,\middle|\, J_1 = n\right]\right)^{1-\theta} \left(\mathrm{E}\left[\mathrm{e}^{s_2 X_1} \,\middle|\, J_1 = n\right]\right)^{\theta} < \infty \qquad (3.5)
\end{aligned}$$

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 (3.5) into a matrix we obtain

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

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(ii) and B.1, we obtain

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

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

To show the uniqueness of α, 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 λ 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 $\mathcal{S}$.

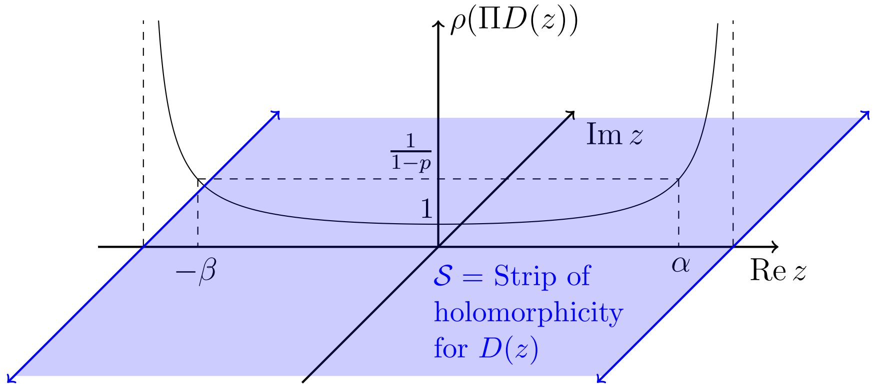


Figure 3.2: Definition of $\mathcal{S}$.

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

$$\begin{bmatrix} \mathrm{E}\left[\mathrm{e}^{zW_t} \,\middle|\, J_0 = 1\right] \\ \vdots \\ \mathrm{E}\left[\mathrm{e}^{zW_t} \,\middle|\, J_0 = N\right] \end{bmatrix} = (\Pi D(z))^t e. \tag{3.6}$$

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 ω_0 is the $N \times 1$ vector of probabilities $\mathrm{P}(J_0 = n)$, $n = 1, \dots, N$.

Equality (3.6) 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 $J_0, J_1, \dots$ 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 (3.6) 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 (3.6) 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) $A(z) = I - (1-p)\Pi D(z)$.

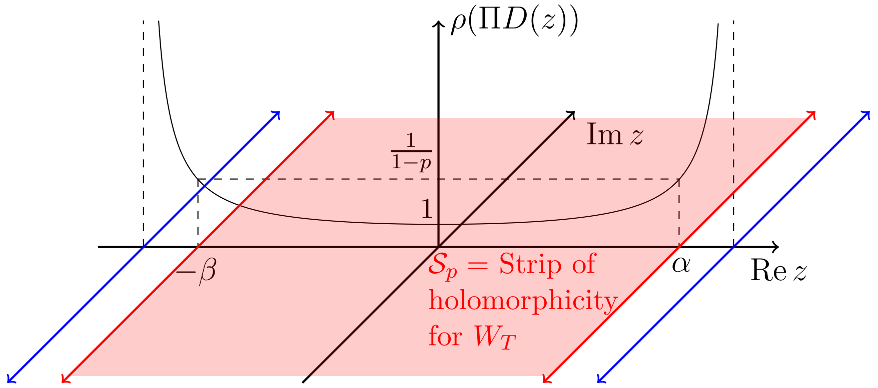


Figure 3.3: Definition of $\mathcal{S}_p$.

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

$$\mathrm{E}[\mathrm{e}^{zW_T}] = p\omega_0^\top A(z)^{-1}\Pi D(z)e. \tag{3.7}$$

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 Proposition B.2(ii) and the second from the definition of $\mathcal{S}_p$. Hence by Proposition B.2(i), $(1-p)^k \left\|(\Pi D(z))^k\right\|$ 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 (3.7).

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 on 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 Π 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 (Ahlfors, 1979, p. 127), and therefore $A(z)$ is holomorphic on 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 on 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 (3.7) is well-defined and holomorphic on 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 α of $A(z)^{-1}$ is given by the $N \times N$ matrix

$$R = x(y^\top A'(\alpha)x)^{-1}y^\top =: cxy^\top, \tag{3.8}$$

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 α 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 (3.7) to holomorphically extend $\mathrm{E}[\mathrm{e}^{zW_T}]$ to a punctured neighborhood of α, which is a singularity. Let us show that this singularity is a simple pole by showing that

$$\lim_{z\to\alpha}(z-\alpha)p\omega_0^\top A(z)^{-1}\Pi D(z)e \tag{3.9}$$

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 (3.9) 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 (3.8) again, it follows that

$$(3.9) = p\omega_0^\top \frac{1}{1-p}Re = c\frac{p}{1-p}(\omega_0^\top x)(y^\top e) = c\frac{p}{1-p}\omega_0^\top x \neq 0 \tag{3.10}$$

since the entries of y sum to 1, $\omega_0 > 0$, and $x \gg 0$.[8] Therefore our holomorphic

[8] Given real vectors $a = (a_n)$ and $b = (b_n)$, we say that $a \geq b$ if $a_n \geq b_n$ for all n, that $a > b$

extension of $\mathrm{E}[\mathrm{e}^{zW_T}]$ has a simple pole at $z = \alpha$.

The limit (3.3) now follows from Theorem 2.1 and Remark 2.3. A symmetric argument establishes the analogous result for the lower tail. $\square$

Remark 3.1. Equation (3.2) is the most important equation in our paper. It provides an implicit characterization of the tail exponents α and β in terms of the parameters Π, D, and p. In Section 3.3 we use the implicit function theorem to investigate the response of the tail exponents to perturbations in these parameters.

Remark 3.2. 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 the equations

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

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

Remark 3.3. One of the assumptions of Theorem 3.1 is that there exists $\alpha > 0$ 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 $s > \alpha$.[9] 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 α 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 3.4. 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) $\mathrm{E}\left[\mathrm{e}^{sX_1} \,\middle|\, J_1 = n\right] < \infty$ for all $s > 0$ and all $n \in \mathcal{N}$ and (ii) $\pi_{nn} > 0$ and $\mathrm{P}(X_1 > 0 \mid J_1 = n) > 0$ for some $n \in \mathcal{N}$. To see this, note that under condition (i) the matrix $D(s)$ is well-defined for all

if $a \geq b$ and $a \neq b$, and that $a \gg b$ if $a_n > b_n$ for all n.

[9] 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 $s > \alpha$.

s. Let $M(s)$ be a matrix whose n-th diagonal entry is $\pi_{nn} \mathrm{E}\left[\mathrm{e}^{sX_1} \,\middle|\, J_1 = n\right]$ for one n satisfying condition (ii) and 0 for all other entries. Clearly $\Pi D(s) \geq M(s)$. By Proposition B.2(ii), we have

$$\infty > \rho(\Pi D(s)) \geq \rho(M(s)) = \pi_{nn} \mathrm{E}\left[\mathrm{e}^{sX_1} \,\middle|\, J_1 = n\right] \to \infty$$

as $s \to \infty$ by condition (ii), so $\rho(\Pi D(s))$ crosses $1/(1-p)$.[10] A similar sufficient condition applies to the lower tail exponent.

Remark 3.5. 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

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

The discrete-time analog of this process is

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

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 η. Then in the discrete-time analog, the birth/death probability is $p = 1 - \mathrm{e}^{-\eta\Delta}$. By Remark 3.2, W_T has exponential tails with exponents $\zeta = \alpha, -\beta$ solving the equation $\mathrm{E}[\mathrm{e}^{\zeta X_1}] = 1/(1-p)$. But with $X_1 \sim N((\mu - \sigma^2/2)\Delta, \sigma^2\Delta)$ this equation can be rewritten as

$$\mathrm{e}^{(\mu - \sigma^2/2)\Delta\zeta + \sigma^2 \Delta \zeta^2 / 2} = \mathrm{e}^{\eta\Delta},$$

which is equivalent to (1.4). Therefore with iid Gaussian shocks, the tail exponents

[10] There may not be α such that $\rho(\Pi D(\alpha)) = 1/(1-p)$ if condition (ii) fails. A counterexample is the two-state Markov chain that takes the values ± 1 with transition probability matrix $\Pi = [(0,1)^\top, (1,0)^\top]$. Then $\Pi D(s) = [(0, \mathrm{e}^{-s})^\top, (\mathrm{e}^{s}, 0)^\top]$, whose spectral radius is identically equal to 1. In this case $W_T \in \{-1, 0, 1\}$ (bounded), so it does not have exponential tails.

are identical in the discrete-time and continuous-time models.

Example 3.4 (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

$$\begin{aligned}\rho(\Pi D(s)) &= \rho\left(\begin{bmatrix}\pi_{11}\mathrm{e}^{sx_1} & \pi_{12}\mathrm{e}^{sx_2}\\ \pi_{21}\mathrm{e}^{sx_1} & \pi_{22}\mathrm{e}^{sx_2}\end{bmatrix}\right)\\ &= \frac{1}{2}\left(\pi_{11}\mathrm{e}^{sx_1} + \pi_{22}\mathrm{e}^{sx_2} + \sqrt{(\pi_{11}\mathrm{e}^{sx_1} - \pi_{22}\mathrm{e}^{sx_2})^2 + 4\pi_{12}\pi_{21}\mathrm{e}^{s(x_1+x_2)}}\right). \end{aligned} \tag{3.11}$$

Setting this quantity equal to $1/(1-p)$ and solving for s gives a unique positive solution α provided that $\pi_{nn} > 0$ and $x_n > 0$ for some $n \in \{1, 2\}$, and a unique negative solution $-\beta$ provided that $\pi_{nn} > 0$ and $x_n < 0$ for some n (Remark 3.4).

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 (3.3) and (3.4). 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 for some $b > 0$ 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*

$$\frac{2\pi C/b}{\mathrm{e}^{2\pi\alpha/b} - 1} \leq \liminf_{w\to\infty} \mathrm{e}^{\alpha w}\mathrm{P}(W_T > w) \leq \limsup_{w\to\infty} \mathrm{e}^{\alpha w}\mathrm{P}(W_T > w) \leq \frac{2\pi C/b}{1 - \mathrm{e}^{-2\pi\alpha/b}}, \tag{3.12}$$

where

$$C = \frac{p\omega_0^\top x}{(1-p)^2 y^\top \Pi D'(\alpha) x}. \tag{3.13}$$

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

$$\lim_{w\to\infty} \mathrm{e}^{\alpha w}\mathrm{P}(W_T > w) = \frac{C}{\alpha}. \tag{3.14}$$

Proof. Let $I_{\alpha,b} = \{z = \alpha + it : |t| < b\}$. 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 $I_{\alpha,b}\backslash\{\alpha\}$ then we may deduce that $A(z)$ is holomorphic on an open set containing $I_{\alpha,b}$ with a unique point of non-invertibility 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 $I_{\alpha,b}\backslash\{\alpha\}$. In view of (3.7) the latter function constitutes a continuous extension of the mgf of W_T to the union of its strip of holomorphicity and $I_{\alpha,b}\backslash\{\alpha\}$; moreover, it was shown in the proof of Theorem 3.1 that the singularity at α 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 (3.12).

If $z = \alpha$ is the unique point of noninvertibility of $A(z)$ on the axis $\operatorname{Re} z = \alpha$, then we can take b arbitrarily large. Letting $b \to \infty$, both sides of (3.12) converge to C/α and we obtain (3.14). □

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 dependent on a Markov state variable; 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*

$$\lim_{s\to\infty} \frac{\log \mathrm{P}(S > s)}{\log s} = -\alpha. \tag{3.15}$$

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

$$\lim_{s\to\infty} s^\alpha \mathrm{P}(S > s) = \frac{C}{\alpha}\mathrm{E}[S_0^\alpha], \tag{3.16}$$

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 α. 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 α. Therefore the mgf of $\log S$ is also holomorphic on the strip $0 < \operatorname{Re} z < \alpha$ with a simple pole at α, 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 (3.15) 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 $\lim_{s\to\infty} s^\alpha \mathrm{P}(\mathrm{e}^{W_T} > s) = C/\alpha$. The upper tail of the random variable e^{W_T} is therefore regularly varying with index $-\alpha$, and so by applying the well-known lemma of Breiman (1965) 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. If $A(\alpha + i\tau)$ is noninvertible for some $\tau \neq 0$, then there exist $a_1, \dots, a_N \in \mathbb{R}$ such that*

$$\operatorname{supp}(X_1 \mid J_1 = n) \subset a_n + \frac{2\pi}{\tau}\mathbb{Z}$$

for all $n \in \mathcal{N}$, with $a_n = 0$ if $\pi_{nn} > 0$. Conversely, if

$$\operatorname{supp}(X_1 \mid J_1 = n) \subset \frac{2\pi}{\tau}\mathbb{Z}$$

for some $\tau \neq 0$ and 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 $\operatorname{supp} X \subset a + (2\pi/\tau)\mathbb{Z}$ 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

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

Fix any support point $a \in \operatorname{supp} X$. Since the triangle inequality holds with equality, it must be the case that $0 \le \mathrm{e}^{(\alpha+i\tau)x}/\mathrm{e}^{(\alpha+i\tau)a} = \mathrm{e}^{(\alpha+i\tau)(x-a)}$ 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 $\operatorname{supp} X \subset a + (2\pi/\tau)\mathbb{Z}$.

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

$$\sum_{m\in\mathbb{Z}} p_m \mathrm{e}^{\alpha(a+2\pi m/\tau)} = \mathrm{E}[\mathrm{e}^{\alpha X}] = \mathrm{E}[\mathrm{e}^{(\alpha+i\tau)X}]$$
$$= \sum_{m\in\mathbb{Z}} p_m \mathrm{e}^{(\alpha+i\tau)(a+2\pi m/\tau)} = \mathrm{e}^{i\tau a} \sum_{m\in\mathbb{Z}} p_m \mathrm{e}^{\alpha(a+2\pi m/\tau)}.$$

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 $\operatorname{supp}(X_1 \mid J_1 = n) \subset a_n + (2\pi/\tau)\mathbb{Z}$ *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 $1/(1-p) \le \rho(D(\alpha + i\tau))$. Since $D(\alpha + i\tau)$ is a diagonal matrix whose n-th diagonal entry is $\mathrm{E}\left[\mathrm{e}^{(\alpha+i\tau)X_1} \,\middle|\, J_1 = n\right]$, by the triangle inequality we obtain $|D(\alpha + i\tau)| \le D(\alpha)$ entry-wise. Since Π is a nonnegative matrix, by Proposition B.2(ii) we obtain

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

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 $\Theta = \operatorname{diag}(\mathrm{e}^{i\theta_1}, \dots, \mathrm{e}^{i\theta_N})$ for some $\theta_1, \dots, \theta_N \in \mathbb{R}$. Comparing the (m,n)-th entry, we have

$$\pi_{mn} \mathrm{E}\left[\mathrm{e}^{(\alpha+i\tau)X_1} \,\middle|\, J_1 = n\right] = \pi_{mn} \mathrm{e}^{i(\theta_m - \theta_n)} \mathrm{E}\left[\mathrm{e}^{\alpha X_1} \,\middle|\, J_1 = n\right] \tag{3.17}$$

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

Since Π is irreducible, for each $n \in \mathcal{N}$ there exists $m \in \mathcal{N}$ such that $\pi_{mn} > 0$. Taking the absolute value of (3.17) 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 $a_1, \dots, a_N \in \mathbb{R}$ such that $\operatorname{supp}(X_1 \mid J_1 = n) \subset a_n + (2\pi/\tau)\mathbb{Z}$ for all $n \in \mathcal{N}$.

If $\pi_{nn} > 0$, setting $m = n$ in (3.17) 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 $\operatorname{supp}(X_1 \mid J_1 = n) \subset (2\pi/\tau)\mathbb{Z}$, so we can take $a_n = 0$.

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

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

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

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 α, so it must also be noninvertible at $\alpha + i\tau$. $\square$

Remark 3.6. The condition $\operatorname{supp}(X_1 \mid J_1 = n) \subset (2\pi/\tau)\mathbb{Z}$ 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 $\operatorname{supp}(X_1 \mid J_1 = n) \subset a_n + (2\pi/\tau)\mathbb{Z}$ for all n, then (3.17) (which is sufficient for noninvertibility) 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 $\operatorname{supp}(X_1 \mid J_1 = n) \subset \pm 1/\pi + \mathbb{Z}$.

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 3.5 (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 $A(z) = 1 - (1 - p)\mathrm{e}^{\mu z}$, and the mgf of $W_T = \mu T$ is given by

$$\mathrm{E}[\mathrm{e}^{zW_T}] = \sum_{t=1}^{\infty} p(1 - p)^{t-1}\mathrm{e}^{\mu z t} = \frac{p\mathrm{e}^{\mu z}}{1 - (1 - p)\mathrm{e}^{\mu z}}. \tag{3.18}$$

The right abscissa of convergence is $\alpha = -\log(1 - p)/\mu > 0$. Setting the denominator in (3.18) 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 $S = \mathrm{e}^{W_T}$. 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 $1/(1-p)$ as $s \to \infty$, so the tail is not Paretian. These limits are precisely the bounds we obtain in (3.12) by setting $C = p/(\mu(1-p))$ and $b = 2\pi/\mu$.

3.3 Comparative statics

Theorems 3.1–3.3 show that geometrically stopped random growth processes have Pareto tails under empirically plausible conditions. Our characterization of the Pareto exponents is implicit, in the sense that they are given by the positive and negative solutions to (3.2). It may be desirable to provide an explicit characterization of how the exponents vary when we perturb the exogenous parameters Π, D, and p. In this section, we derive such comparative statics.

We consider perturbations that involve the stopping probability p, location-scale transformations of the increments X_t, and the overall persistence of the hidden Markov state J_t. We interpret increases (decreases) in p as decreases (increases) in lifespan. Regarding the location-scale transformations of X_t, let $Y_1, \dots, Y_N$ be zero mean random variables and suppose that the distribution of X_t conditional on $J_t = n$ is parametrized as $\mu_n + \sigma_n Y_n$, where $\mu_n \in \mathbb{R}$ and $\sigma_n > 0$. Let $\mu = (\mu_1, \dots, \mu_N)^\top$ and $\sigma = (\sigma_1, \dots, \sigma_N)^\top$, and let $D(s; \mu, \sigma)$ be the $N \times N$ diagonal matrix with n-th diagonal entry equal to $\mathrm{E}[\mathrm{e}^{s(\mu_n + \sigma_n Y_n)}]$. We interpret increases (decreases) in μ_n and σ_n as increases (decreases) in growth and volatility in state n. Regarding persistence, given our fixed irreducible transition probability matrix Π we define $\Pi(\tau) = \tau I + (1-\tau)\Pi$ for $\tau \in [0,1]$. We interpret increases (decreases) in τ as increases (decreases) in persistence. Note that since Π is irreducible, so is $\Pi(\tau)$ for $\tau \in (0,1)$.

The following theorem shows that increasing the lifespan, growth, volatility, or persistence makes the upper Pareto exponent smaller (the upper tail heavier), which is intuitive. Its proof is an application of the implicit function theorem. Analogous results hold for the lower tail.

Theorem 3.5. *Let $\theta = (p, \mu, \sigma, \tau) \in (0,1) \times \mathbb{R}^N \times \mathbb{R}^N_{++} \times (0,1)$ be the parameters we perturb. Suppose that the conditions of Theorem 3.1 are satisfied for some parameter value θ^0. Then there exists an open neighborhood Θ of θ^0 and a unique continuously differentiable function $\alpha : \Theta \to (0, \infty)$ such that*

$$\rho(\Pi(\tau)D(\alpha(\theta); \mu, \sigma)) = \frac{1}{1-p}. \tag{3.19}$$

The partial derivatives of α satisfy the following inequalities at θ^0.

(i) $\partial\alpha/\partial p > 0$*: Longer lifespan implies a smaller Pareto exponent.*

(ii) $\partial\alpha/\partial\mu_n \leq 0$*: Higher growth implies a smaller Pareto exponent.*

(iii) $\partial\alpha/\partial\sigma_n \leq 0$*: Higher volatility implies a smaller Pareto exponent.*

(iv) $\partial\alpha/\partial\tau \leq 0$*: Higher persistence implies a smaller Pareto exponent.*

Proof. Define $F : (0,1) \times \mathbb{R}^N \times \mathbb{R}^N_{++} \times (0,1) \times (0,\infty) \to \mathbb{R} \cup \{\infty\}$ by

$$F(\theta, s) = \rho(\Pi(\tau)D(s; \mu, \sigma)) - \frac{1}{1-p},$$

where the spectral radius is understood to be ∞ whenever one or more elements of $D(s; \mu, \sigma)$ are infinite. Since the conditions of Theorem 3.1 are satisfied at the parameter value $\theta^0 = (p^0, \mu^0, \sigma^0, \tau^0)$, there exists $\alpha^0 > 0$ in the interior of $\mathcal{I}$ such that $F(\theta^0, \alpha^0) = 0$. Let us show that there exists an open neighborhood U of (θ^0, α^0) such that F is finite on U. Clearly F is finite if and only if $\mathrm{E}[\mathrm{e}^{s(\mu_n + \sigma_n Y_n)}] < \infty$ for all n. Since $\mathrm{E}[\mathrm{e}^{s(\mu_n + \sigma_n Y_n)}] = \mathrm{e}^{s\mu_n} \mathrm{E}[\mathrm{e}^{s\sigma_n Y_n}]$, it suffices to show that $\mathrm{E}[\mathrm{e}^{s\sigma_n Y_n}] < \infty$ if (s, σ_n) is sufficiently close to (α^0, σ^0_n). However, this follows trivially from the definition of $\mathcal{I}$ in (3.1) and the fact that α^0 lies in the interior of $\mathcal{I}$.

To prove our claim we apply the implicit function theorem. Proposition B.3 tells us that the spectral radius of a nonnegative irreducible square matrix varies holomorphically with local perturbations to the elements of that matrix. Therefore, noting that $\Pi(\tau^0)D(\alpha^0; \mu^0, \sigma^0)$ is nonnegative and irreducible and that the elements of $D(s; \mu, \sigma) = \mathrm{diag}((\mathrm{E}[\mathrm{e}^{s(\mu_n + \sigma_n Y_n)}])_{n=1}^N)$ and $\Pi(\tau)$ are continuously differentiable with respect to μ, σ, τ, s, we find that F is continuously differentiable on U. In Step 1 of the proof of Theorem 3.1 we established that $\rho(\Pi D(s))$ is convex in s. It follows that $F(\theta^0, s)$ is convex in s, and so

$$0 > -\frac{p}{1-p} = 1 - \frac{1}{1-p} = F(\theta^0, 0) - F(\theta^0, \alpha^0) \geq \frac{\partial F}{\partial s}(\theta^0, \alpha^0)(0 - \alpha^0),$$

implying that $\partial F/\partial s > 0$ on a neighborhood of (θ^0, α^0). We may therefore apply the implicit function theorem, which guarantees the existence of an open neighborhood Θ of θ^0 and a unique continuously differentiable function $\alpha : \Theta \to (0, \infty)$ such that (3.19) holds. The partial derivatives of α on Θ are then given by

$$\nabla\alpha(\theta) = -\frac{1}{\partial F/\partial s}\nabla_\theta F.$$

It remains only to show that the partial derivatives of F with respect to p, μ_n, σ_n, τ have the appropriate sign at (θ^0, α^0). We consider them in turn.

(i) Noting that $\partial F/\partial p = -1/(1-p)^2 < 0$, we deduce that $\partial\alpha/\partial p > 0$ at θ^0.

(ii) $F(\theta, s)$ depends on μ_n only through the n-th diagonal element of $D(s; \mu, \sigma)$, which we denote by $M_n(s; \mu_n, \sigma_n) = \mathrm{E}[\mathrm{e}^{s(\mu_n+\sigma_n Y_n)}]$. Since for $s > 0$ we have

$$\frac{\partial M_n}{\partial \mu_n} = s\,\mathrm{E}[\mathrm{e}^{s(\mu_n+\sigma_n Y_n)}] > 0,$$

M_n is increasing in μ_n. Since by Proposition B.2(ii) the spectral radius of a nonnegative matrix is increasing in its elements, it follows that $\partial F/\partial \mu_n \geq 0$ at (θ^0, α^0). Therefore $\partial\alpha/\partial\mu_n \leq 0$ at θ^0.

(iii) Since by assumption Y_n has zero mean,

$$\left.\frac{\partial M_n}{\partial \sigma_n}\right|_{\sigma_n=0} = s\,\mathrm{E}[\mathrm{e}^{s(\mu_n+\sigma_n Y_n)}Y_n]\big|_{\sigma_n=0} = 0.$$

Since the exponential function is convex, M_n is convex in σ_n. Therefore $\partial M_n/\partial\sigma_n \geq 0$ for $s, \sigma_n > 0$. By the same argument as the previous case, we have $\partial F/\partial\sigma_n \geq 0$ at (θ^0, α^0) and hence $\partial\alpha/\partial\sigma_n \leq 0$ at θ^0.

(iv) Proposition B.4 guarantees that

$$\frac{\partial F}{\partial \tau}(\theta^0, \alpha^0) = \left.\frac{\partial}{\partial \tau}\rho((\tau I + (1-\tau)\Pi)D(\alpha^0))\right|_{\tau=\tau^0} \geq 0,$$

implying that $\partial\alpha/\partial\tau \leq 0$ at θ^0. □

Example 3.6 (Two-state Markov chain)**.** Consider the random growth model where the growth rate process $\{X_t\}_{t=1}^\infty$ is a two-state Markov chain as in Example 3.4. Let $X_t = \mu_n > 0$ if $J_t = n$, where $n = 1, 2$. One can interpret this model as the deterministic growth model of Wold and Whittle (1957) except that the growth rate has two regimes.

Let $\eta = -\log(1-p)$ be the death rate. Suppose that the transition probability matrix is

$$\Pi(\tau) = \tau I + (1-\tau)\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \tau & 1-\tau \\ 1-\tau & \tau \end{bmatrix}$$

and let $\alpha(\tau)$ be the Pareto exponent corresponding to τ. By Theorem 3.5, $\alpha(\tau)$ is decreasing in τ. Using (3.11), we obtain

$$\alpha(0) = \frac{\eta}{\bar{\mu}} \quad \text{and} \quad \lim_{\tau \uparrow 1} \alpha(\tau) = \min_n \frac{\eta}{\mu_n},$$

where $\bar{\mu} = (\mu_1 + \mu_2)/2$.

As a numerical example, suppose that $(\mu_1, \mu_2) = (0.03, 0.01)$ and $\eta = 0.04$, so units grow by 3% and 1% in each state and the average life span is about 25 periods. The blue curve in Figure 3.4 shows the Pareto exponent. In this case $\bar{\mu} = 0.02$, so $\alpha(0) = \eta/\bar{\mu} = 2$ and $\alpha(1) = \eta/\mu_1 = 4/3$. The red line corresponds to the case where $\{X_t\}_{t=1}^{\infty}$ is iid with $X_t = \mu_n$ with probability $1/2$. We can see that the two Pareto exponents are similar when the persistence τ of the Markov chain is not so high, say $\tau \in [0, 0.8]$. However, increasing τ beyond 0.8 quickly makes the Pareto exponent small. In the limit as $\tau \to 1$, the exponent converges to the case of highest growth $\max_n \mu_n$.

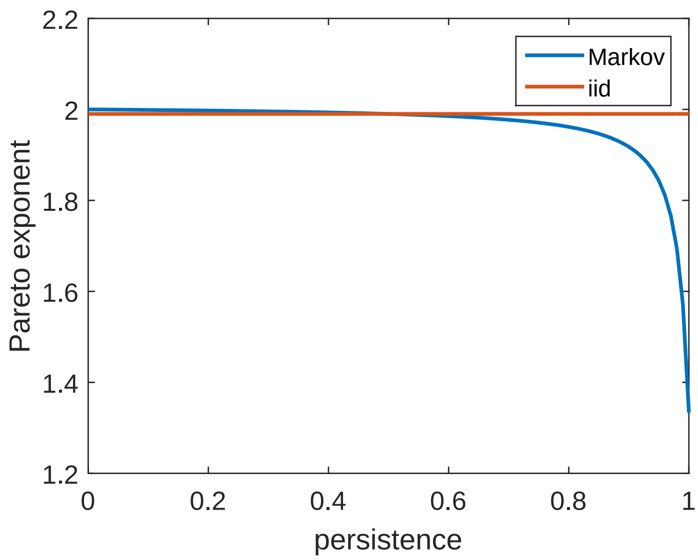


Figure 3.4: Pareto exponent of the two-state random growth model.

4 Tails of wealth in heterogeneous-agent models

As an application of Theorems 3.1–3.3, in this section we characterize the tails of the wealth distribution in heterogeneous-agent models with non-Gaussian, Markovian shocks. We consider two cases, models with idiosyncratic endowment and investment risks. All omitted proofs in this section may be found in Appendix D.

4.1 Wealth distribution in CARA Huggett economies

The first application is to 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 Markovian shocks that exploits constant absolute risk aversion (CARA) preferences. We first consider the single agent problem (partial equilibrium), which we subsequently embed into a general equilibrium model.

4.1.1 Single agent problem

Consider an agent with additive CARA utility

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

where $c_t \in \mathbb{R}$ is consumption at period t, the discount factor is $\beta \in (0,1)$, and the period utility function $u(c) = -\mathrm{e}^{-\gamma c}/\gamma$ exhibits constant absolute risk aversion $\gamma > 0$.[11] The agent can borrow or save at an exogenous gross risk-free rate $R > 1$ and is subject to uninsurable idiosyncratic income risk.[12] Letting y_t be the income at t and w_t the financial wealth at the beginning of period t excluding the current income y_t, the budget constraint of the agent is

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

[11] We focus on CARA preferences because they are tractable with additive shocks. For example, the income process is iid Gaussian in Calvet (2001), non-Gaussian AR(1) in Wang (2003), and non-Gaussian VAR(1) in Toda (2017a). 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.

[12] The assumption $R > 1$ is without loss of generality: since the expected present value of future income is infinite when $R \leq 1$, without borrowing constraints (as in the current setup) the income fluctuation problem does not have a solution. In the subsequent general equilibrium analysis, we prove that a stationary equilibrium with $R > 1$ always exists.

Given the initial wealth w_0 and the income process $\{y_t\}_{t=0}^{\infty}$, the agent's objective is to maximize the utility (4.1) subject to the budget constraint (4.2) for all $t \geq 0$.

Suppose that the income process $\{y_t\}_{t=0}^{\infty}$ is a time-homogeneous finite-state Markov chain. Let $s \in \mathcal{S} = \{1, \dots, S\}$ denote the states, $P = (p_{ss'})$ be the (irreducible) transition probability matrix, and $y = (\bar{y}_s)_{s=1}^{S}$ be the vector of incomes. By merging redundant states, without loss of generality we may assume $\bar{y}_1 < \cdots < \bar{y}_S$. Due to time-homogeneity, the state variables of the utility maximization problem are the current wealth $w = w_t$ and the exogenous state $s = s_t$, so let $V_s(w)$ be the value function in state s. The Bellman equation is

$$V_s(w) = \max_c \left\{ u(c) + \beta \operatorname{E}\left[V_{s'}(R(w - c + \bar{y}_s)) \mid s \right] \right\}, \tag{4.3}$$

where s' is next period's state.

The following proposition characterizes the optimal consumption rule in closed-form up to a system of S nonlinear equations in S unknowns.

Proposition 4.1. *Suppose that $R > 1$. Then the utility maximization problem has a unique solution. 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

$$b = \left(1 - \frac{1}{R}\right) y - \frac{1}{\gamma R} \log\left(\beta R P \mathrm{e}^{-\gamma b}\right).[13] \tag{4.5}$$

4.1.2 General equilibrium and wealth distribution

Next we embed the above single agent problem into the general equilibrium. Consider a Huggett (1993) economy, where there are infinitely many 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.

Before formally stating the model, we note that we cannot obtain stationarity just by replicating agents, and therefore we need to introduce additional assumptions. To see this, use the budget constraint (4.2), the consumption rule (4.4b),

[13] For notational simplicity, we apply functions entry-wise. For example, for $b = (b_s)_{s=1}^{S}$, we have $\mathrm{e}^{-\gamma b} = (\mathrm{e}^{-\gamma b_s})_{s=1}^{S}$.

and the equation $a = 1 - 1/R$ to write

$$w_{t+1} = R(w_t - (aw_t + b_{s_t}) + \bar{y}_{s_t}) = w_t + R(\bar{y}_{s_t} - b_{s_t}). \tag{4.6}$$

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 $\tilde{\beta} = \beta(1 - p)$.

The formal model works as follows. There are countably infinite agents indexed by $i \in \mathbb{N}$. Agent i's consumption, income, and wealth at time t are denoted by c_{it}, y_{it}, and w_{it}. All agents have the same utility function (4.1). When an agent is born, their initial wealth is 0 and their income is $\bar{y}_s$, where the initial state s is drawn from the stationary distribution $\pi = (\pi_s)_{s=1}^S$ of the Markov chain with transition probability matrix $P = (p_{ss'})$ independently across agents. By the strong law of large numbers, the limits of average quantities such as

$$\frac{1}{I}\sum_{i=1}^{I} c_{it}, \quad \frac{1}{I}\sum_{i=1}^{I} y_{it}, \quad \frac{1}{I}\sum_{i=1}^{I} w_{it}$$

exist almost surely as $I \to \infty$. We refer to such quantities as (per capita) aggregate consumption, income, and wealth and denote by capital letters C_t, Y_t, W_t, etc.

Because agents die with probability p each period, we need to specify what happens to the risk-free asset position of deceased agents. Suppose that there are perfectly competitive insurance companies that offer annuities and life insurances. Let R be the gross risk-free rate and $\tilde{R}$ 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_t be aggregate savings, by accounting we obtain $RA_t = (1 - p)\tilde{R}A_t + p0$, and thus deduce that $\tilde{R} = R/(1 - p)$.

Now we can formally define the stationary equilibrium.

Definition 4.1 (Stationary equilibrium of CARA Huggett economy)**.** A *stationary equilibrium* of a CARA Huggett economy consists of a gross risk-free rate $R > 0$ and a sequence of consumption and wealth $\{(c_{it}, w_{it})_{i \in \mathbb{N}}\}_{t=0}^{\infty}$ such that

(i) (Optimization) for all $i \in \mathbb{N}$, $\{(c_{it}, w_{it})\}_{t=0}^{\infty}$ maximizes utility (4.1) subject to the budget constraint (4.2), where β, R are replaced with $\tilde{\beta} = \beta(1 - p)$

and $\tilde{R} = R/(1-p)$,

(ii) (Market clearing) the risk-free asset market clears, i.e., the aggregate savings is 0, and

(iii) (Stationarity) all aggregate quantities are constant over time almost surely.

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, the aggregate consumption is

$$C = aW + \pi^\top b = \pi^\top b,$$

where $\pi = (\pi_s)_{s=1}^S$ is the stationary distribution of the Markov chain. Furthermore, aggregate consumption must equal aggregate income, so $C = Y = \pi^\top y$. Combining these two equations, the equilibrium condition is

$$\pi^\top b = \pi^\top y. \tag{4.7}$$

The following theorem shows the existence of a stationary equilibrium.

Theorem 4.2. *There exists a stationary equilibrium in the CARA Huggett economy. The gross risk-free rate satisfies* $1 - p < R \leq 1/\beta$.

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

Theorem 4.3. *Let* $P = (p_{ss'})$ *be the transition probability matrix with positive diagonal entries,* $y = (\bar{y}_1, \dots, \bar{y}_S)^\top$ *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* α_1 *and lower tail exponent* α_2*.*

Proof. By (4.6) with R replaced by $\tilde{R}$, individual wealth increases by $\tilde{R}(\bar{y}_s - b_s)$ in state s. Since P has positive diagonal entries and the vector of moment generating functions is $M(z) = \mathrm{e}^{z\tilde{R}(y-b)}$, assuming $y - b$ has a positive element, by Theorem 3.1 (see Remark 3.4) 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.

Therefore to complete the proof it remains to show that $y-b$ has both positive and negative elements. Since by the equilibrium condition we have $\pi^\top(y-b)=0$, it suffices to show that $y \neq b$. Suppose on the contrary that $y=b$. Then by (4.5) we obtain $\gamma y = -\log(\beta R P \mathrm{e}^{-\gamma y})$, or equivalently

$$Pv = \frac{1}{\beta R} v$$

with $v = \mathrm{e}^{-\gamma y} \gg 0$. Therefore v is an eigenvector of P with eigenvalue $1/(\beta R)$. Since P is an irreducible nonnegative matrix with spectral radius 1, and $\beta R \leq 1$ by Theorem 4.2, it must be the case that $\beta R = 1$ and $Pv = v$. By the Perron-Frobenius theorem, v must be a scalar multiple of the Perron vector, which is $\mathbf{1}$. Therefore $\bar{y}_1 = \cdots = \bar{y}_S$, which contradicts the assumption $\bar{y}_1 < \cdots < \bar{y}_S$. □

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 4.3 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.

4.1.3 Numerical example

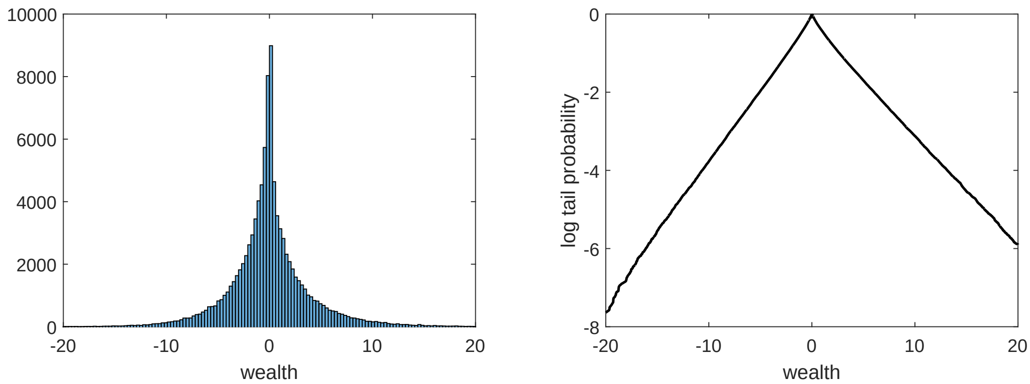
As a numerical example, let the discount factor be $\beta = 0.96$, absolute risk aversion $\gamma = 3$, death probability $p = 0.025$, and suppose that log income $x_t = \log y_t$ is a centered stationary Gaussian AR(1) process,

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

where $\rho = 0.9$ and $\sigma = 0.1$. Since our theory concerns a finite-state Markov chain, instead of the original AR(1) process we take a discrete-state analog as the true process. For this purpose we apply the method in Farmer and Toda (2017) to discretize the AR(1) process $\{x_t\}$ using an even-spaced grid with $S = 9$ points. Letting $\bar{x}_s$ be the grid point in state s, the income in state s is defined by $\bar{y}_s = \mathrm{e}^{\bar{x}_s}$. 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 4.3. 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% of observations in each tail), we obtain $\widehat{\alpha}_1 = 0.2805$ and $\widehat{\alpha}_2 = 0.3657$, 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

Heterogeneous-agent models with idiosyncratic investment risk have recently become quite popular for studying size distributions, but most papers assume iid shocks.[14] Benhabib et al. (2011) consider an overlapping generations model where the returns on wealth and income remain constant over the life cycle of the agent but change when the generation turns over. This way they are able to solve for the optimal consumption rule during the lifetime but allow for Markovian dynamics across generations using the results in da Saporta (2005). Toda (2014) considers a heterogeneous-agent model with Markovian shocks but the Pareto tail results hold only in the continuous-time limit, where shocks are Gaussian. Cao and Luo (2017) consider a continuous-time Markov switching model with two states. Using our main results (Theorems 3.1–3.4), applied researchers may characterize the

[14]See, for example, Luttmer (2007), Nirei and Souma (2007), Benhabib et al. (2015, 2016), Acemoglu and Cao (2015), Toda and Walsh (2015), Gabaix et al. (2016), Nirei and Aoki (2016), and Aoki and Nirei (2017).

tail behavior of the size distribution provided that the agents solve a homogeneous problem (i.e., maximizing a homothetic function subject to proportional constraints such as an optimal consumption-portfolio problem) in a Markovian setting. Due to space considerations we do not discuss a concrete example, but interested readers may refer to the above cited papers.

5 Power law in Japanese prefectures

As an empirical illustration, we apply the random growth model to a panel of Japanese prefecture populations. The main question is whether the time series properties of population dynamics estimated from the panel are consistent with a Pareto index estimated from the cross-section.[15]

5.1 Cross-sectional estimation

We use annual population data for 46 Japanese prefectures (excluding Okinawa) since 1920. Only for the cross-sectional estimation, we also use the census data in 1873, 1884, 1893, 1908, and 1913. See Appendix E.1 for a more detailed description of the data.

Figure 5.1 displays estimates of the upper Pareto exponent of the size distribution of prefectures in different years. The blue curve is the maximum likelihood estimate of the Pareto-lognormal (PlN) distribution, which is the product of independent lognormal and Pareto random variables (see Appendix E.2 for details). The red curve is obtained by running the $\log(\text{Rank} - 1/2)$ regression as in Gabaix and Ibragimov (2011) using the largest 15 prefectures in each year.

We can make a few observations. First, the two curves in Figure 5.1 are similar after 1920 except during WWII, which suggests the robustness of the estimation. Since the maximum likelihood estimation of PlN using the whole sample is more efficient than the $\log(\text{Rank} - 1/2)$ regression using the tail observations and does not require a cutoff selection, below we focus on the former, although the latter may be more robust to misspecification.

Second, the Pareto exponent estimated by maximum likelihood is very large before 1920 and during WWII. Since PlN nests the lognormal distribution as a limiting case ($\alpha \to \infty$), this suggests that the prefecture distribution was closer to lognormal before 1920 and during WWII. Figure 5.2a shows the p-value of the

[15] Davis and Weinstein (2002) examine historical population data from Stone Age Japanese regions and reject the random growth hypothesis, but this is because they narrowly interpret "random growth" as "iid growth".

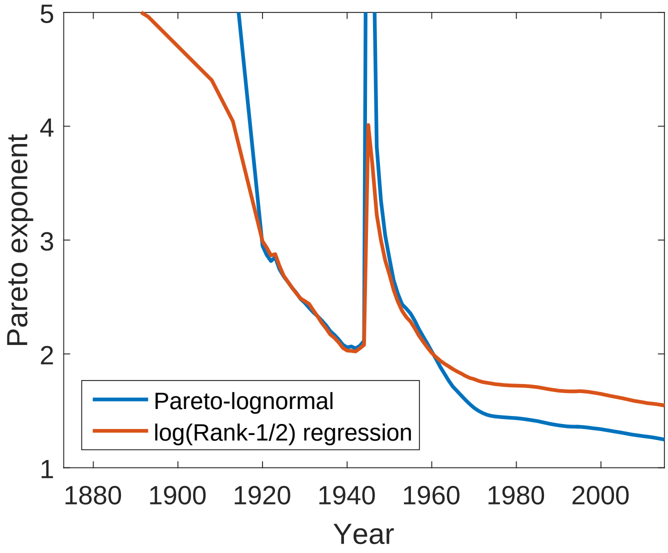
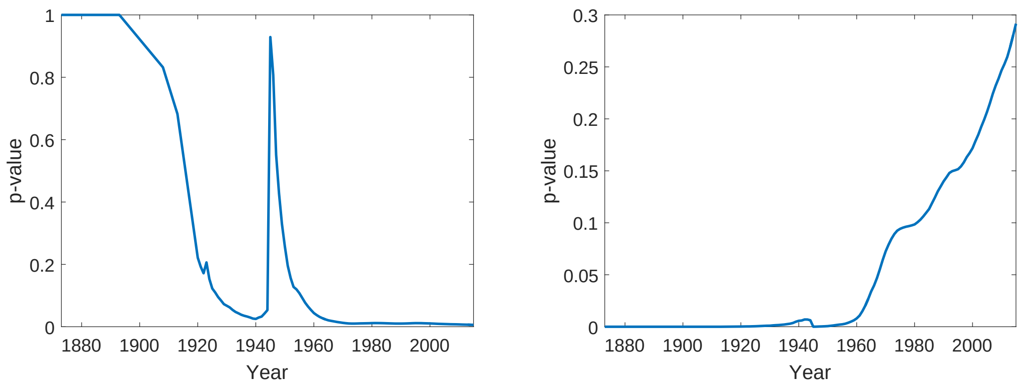


Figure 5.1: Pareto exponent of Japanese prefectures.

likelihood ratio test of the lognormal distribution ($\alpha = \infty$) against the Pareto-lognormal ($\alpha < \infty$). The lognormal distribution is rejected at significance level 0.05 for the years 1933–1943 and all years since 1960. It is not surprising that the power law breaks down during WWII since much of the urban population migrated to rural areas in order to avoid the air raid by the United States.



(a) Test of lognormality ($\alpha = \infty$). **(b)** Test of Zipf's law ($\alpha = 1$).

Figure 5.2: p-values for likelihood ratio tests.

Third, except for the break during WWII, the Pareto exponent is roughly monotonically decreasing and approaches one, which is consistent with Zipf's law. Figure 5.2b shows the p-value of the likelihood ratio test of Zipf's law ($\alpha = 1$) against the unrestricted Pareto-lognormal. Zipf's law fails to be rejected at significance level 0.05 for all years since 1968.

Figure 5.3 displays log-log plots of the rank size distribution of prefectures for several years, together with the values predicted from the Pareto-lognormal and lognormal distributions using the maximum likelihood estimates. In 1873, the size distribution was approximately lognormal. The power law pattern starts to emerge in 1920, and the Pareto exponent drifts downward toward one until 1944. Dislocation and destruction caused by WWII halved the population of Tokyo in 1945, which made the size distribution close to lognormal again. The power law pattern reemerges around 1950 and stabilizes by 1970.

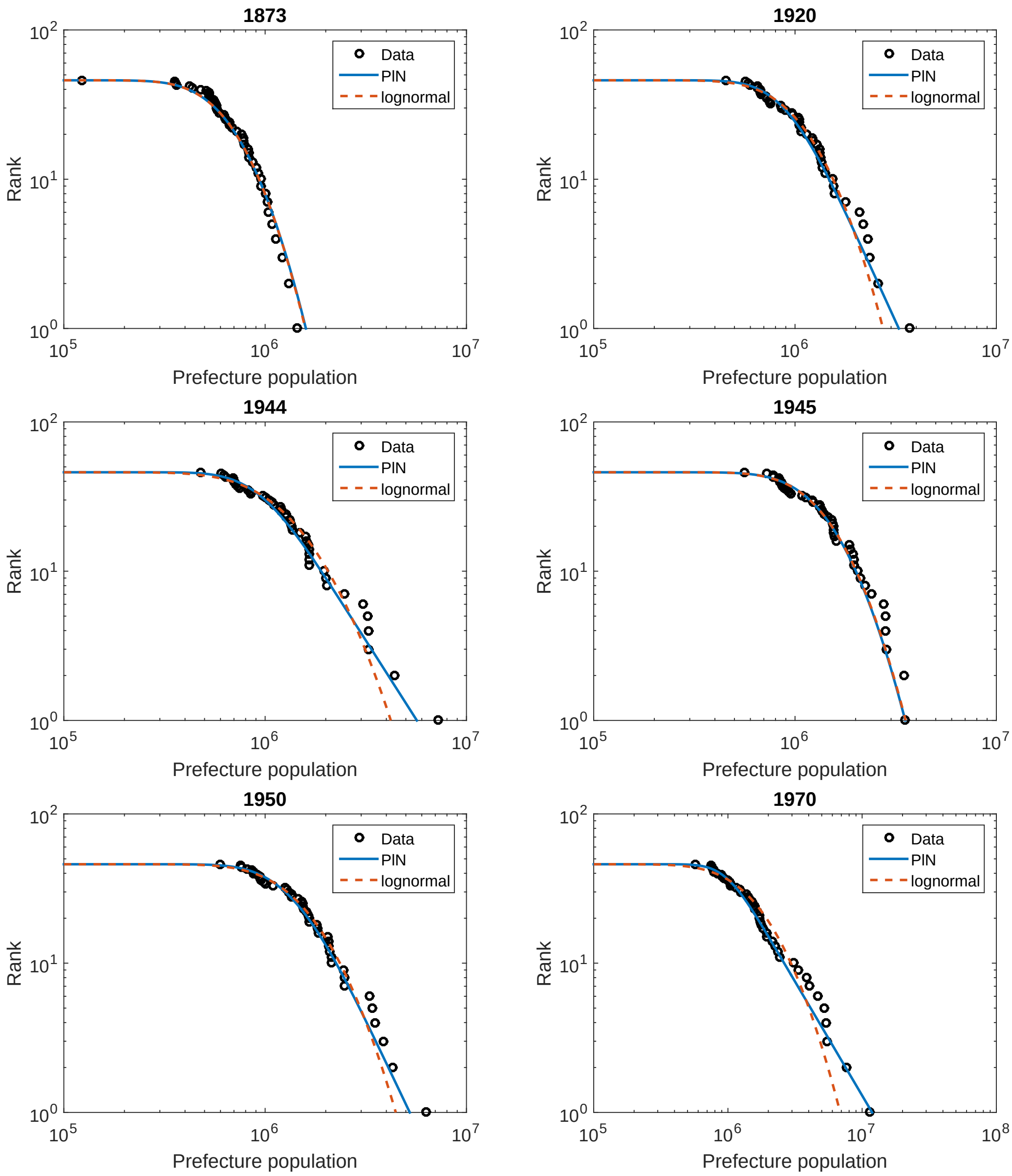


Figure 5.3: Log-log plot of rank size distribution of prefectures.

5.2 Panel estimation

Are the time series properties of the population dynamics consistent with a power law in the cross-section? To answer this question, we estimate a random growth model for the population dynamics and compute the theoretical Pareto exponent. Due to data quality concerns and the structural break during WWII, we only use post war data (1946–2015). We assume that the relative size (population of a prefecture divided by the total population) S_{it} of prefecture i in year t follows the random growth process

$$S_{i,t+1} = G_{i,t+1} S_{it},$$

where $G_{i,t+1}$ is the gross growth rate of the relative size of prefecture i from year t to $t+1$. Following the setting in Section 3, suppose that there are N states indexed by $n = 1, \dots, N$ and in state n the growth rate is lognormally distributed, so

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

where n_{it} is the state of prefecture i in year t. For the number of states, we consider $N = 1, 2, 3$ and estimate the model using the Hamilton (1989) filter (see Appendix E.3 for details). After estimating the transition probability matrix $\Pi = (\pi_{nn'})$, conditional mean $\mu = (\mu_1, \dots, \mu_N)^\top$, and standard deviation $\sigma = (\sigma_1, \dots, \sigma_N)^\top$, we compute the implied Pareto exponent $\alpha > 0$ by solving the equation

$$\rho(\Pi \operatorname{diag}(\mathrm{e}^{\mu_1 s + \sigma_1^2 s^2/2}, \dots, \mathrm{e}^{\mu_N s + \sigma_N^2 s^2/2})) = \frac{1}{1-p}, \tag{5.1}$$

which corresponds to (3.2) in Theorem 3.1. Since some regions in Japan were developed much earlier than others (for example, the region around Kyoto and Osaka were well-developed by the 6th century, while Hokkaido was developed only in the late 19th century), the tail in the age distribution is long. We therefore set $p = 0$ when computing the upper tail exponent.[16] Table 5.1 shows the results.

When the model is iid ($N = 1$), the implied Pareto exponent $\alpha = 56.7$ is wildly inconsistent with the value estimated from the cross-section (about 1.3–1.5 in 2015 according to Figure 5.1). When we expand the model to two states ($N = 2$), the log-likelihood increases dramatically, so the model with a single state is clearly misspecified. The expected growth rate does not change much (only from -0.0035 to -0.0030) but there are now two volatility regimes with identical growth rate. The implied Pareto exponent $\alpha = 26.8$ decreases by half

[16] Empirically, we find that the spectral radius in (5.1) is equal to one at $s = 0$ and at a positive value of s; we take the latter value to be the implied upper Pareto exponent.

Table 5.1: Estimation of random growth model for Japanese prefectures.

Parameter	Number of states (N) 1	2	3
Π	1	$\begin{bmatrix} 0.9754 & 0.0246 \\ 0.0283 & 0.9717 \end{bmatrix}$	$\begin{bmatrix} 0.9439 & 0.0561 & 0.0000 \\ 0.0145 & 0.9671 & 0.0184 \\ 0.0210 & 0.0141 & 0.9649 \end{bmatrix}$
$\mu^\top$	-0.0035	$\begin{bmatrix} -0.0030 & -0.0030 \end{bmatrix}$	$\begin{bmatrix} -0.0122 & -0.0022 & 0.0084 \end{bmatrix}$
$\sigma^\top$	0.0111	$\begin{bmatrix} 0.0029 & 0.0169 \end{bmatrix}$	$\begin{bmatrix} 0.0053 & 0.0026 & 0.0199 \end{bmatrix}$
$\log L$	9,925	11,638	12,388
α	56.7	26.8	1.61

Note: The structural parameters are Π: transition probability matrix, μ: conditional mean, σ: conditional standard deviation. $\log L$ is the log-likelihood. α is the upper Pareto exponent obtained by solving (5.1) for s.

compared to the case $N = 1$ but is still very large. With three states ($N = 3$), the log-likelihood increases dramatically again and the Pareto exponent $\alpha = 1.61$ becomes similar to the cross-sectional estimate. Therefore the emergence of a power law in the Japanese prefecture sizes seems to be consistent with a random growth model, but only by allowing for Markovian dynamics. Note that in each case the estimated transition probability matrix has large diagonal elements, so the process is very persistent. This is another reason why our non-iid results in Section 3 are useful.

6 Conclusion

It has long been conjectured that random growth models robustly generate power law tails. For example, in his well-known paper on Zipf's 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 results that affirmatively resolve the robustness conjecture of power law tails. It is not essential that the growth rate process is Gaussian or iid. 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 Pareto under mild conditions.

Applying a random growth model to historical Japanese prefecture population data, we have shown that the time series properties of the growth process and the tail behavior of the cross-sectional distribution of levels can only be reconciled by introducing a hidden Markov state variable with several regimes. Because our main theorems provide an analytical tool to study such non-iid processes, we believe they have potentially a wide range of applications.

References

Acemoglu, D. and D. Cao (2015): "Innovation by Entrants and Incumbents," *Journal of Economic Theory*, 157, 255–294.

Ahlfors, L. V. (1979): *Complex Analysis*, New York: McGraw-Hill, third ed.

Aiyagari, S. R. (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *Quarterly Journal of Economics*, 109, 659–684.

Aoki, S. and M. Nirei (2017): "Zipf's Law, Pareto's Law, and the Evolution of Top Incomes in the U.S." *American Economic Journal: Macroeconomics*, 9, 36–71.

Arata, Y. (2015): "Essays on Stochastic Macroeconomics," Ph.D. thesis, University of Tokyo.

Arkolakis, C. (2016): "A Unified Theory of Firm Selection and Growth," *Quarterly Journal of Economics*, 131, 89–155.

Axtell, R. L. (2001): "Zipf Distribution of U.S. Firm Sizes," *Science*, 293, 1818–1820.

Beare, B. K., J. Seo, and W.-K. Seo (2017): "Cointegrated Linear Processes in Hilbert Space," *Journal of Time Series Analysis*, 38, 1010–1027.

Benhabib, J. and A. Bisin (2017): "Skewed Wealth Distributions: Theory and Empirics," *Journal of Economic Literature*, forthcoming.

Benhabib, J., A. Bisin, and M. Luo (2017): "Earnings Inequality and Other Determinants of Wealth Inequality," *American Economic Review Papers and Proceedings*, 107, 593–597.

Benhabib, J., A. Bisin, and S. Zhu (2011): "The Distribution of Wealth and Fiscal Policy in Economies with Finitely Lived Agents," *Econometrica*, 79, 123–157.

——— (2015): "The Wealth Distribution in Bewley Economies with Capital Income Risk," *Journal of Economic Theory*, 159, 489–515.

——— (2016): "The Distribution of Wealth in the Blanchard-Yaari Model," *Macroeconomic Dynamics*, 20, 466–481.

Bewley, T. F. (1983): "A Difficulty with the Optimum Quantity of Money," *Econometrica*, 51, 1485–1504.

Blackwell, D. (1965): "Discounted Dynamic Programming," *Annals of Mathematical Statistics*, 36, 226–235.

Blanchard, O. J. (1985): "Debt, Deficits, and Finite Horizons," *Journal of Political Economy*, 93, 223–247.

Breiman, L. (1965): "On Some Limit Theorems Similar to the Arc-Sin Law," *Theory of Probability & Its Applications*, 10, 323–331.

Calvet, L.-E. (2001): "Incomplete Markets and Volatility," *Journal of Economic Theory*, 98, 295–338.

Cao, D. and W. Luo (2017): "Persistent Heterogeneous Returns and Top End Wealth Inequality," *Review of Economic Dynamics*, 26, 301–326.

Castañeda, A., J. Díaz-Giménez, and J.-V. Ríos-Rull (2003): "Accounting for the U.S. Earnings and Wealth Inequality," *Journal of Political Economy*, 111, 818–857.

Champernowne, D. G. (1953): "A Model of Income Distribution," *Economic Journal*, 63, 318–351.

Coad, A. (2010): "The Exponential Age Distribution and the Pareto Firm Size Distribution," *Journal of Industry, Competition and Trade*, 10, 389–395.

da Saporta, B. (2005): "Tail of the Stationary Solution of the Stochastic Equation $Y_{n+1} = a_n Y_n + b_n$ with Markovian Coefficients," *Stochastic Processes and their Applications*, 115, 1954–1978.

Davis, D. R. and D. E. Weinstein (2002): "Bones, Bombs, and Break Points: The Geography of Economic Activity," *American Economic Review*, 92, 1269–1289.

Elsner, L., C. R. Johnson, and J. A. Dias da Silva (1988): "The Perron Root of a Weighted Geometric Mean of Nonneagative Matrices," *Linear and Multilinear Algebra*, 24, 1–13.

Engle, R. F. and C. W. J. Granger (1987): "Co-integration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251–276.

FARMER, L. E. AND A. A. TODA (2017): "Discretizing Nonlinear, Non-Gaussian Markov Processes with Exact Conditional Moments," *Quantitative Economics*, 8, 651–683.

GABAIX, X. (1999): "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics*, 114, 739–767.

——— (2009): "Power Laws in Economics and Finance," *Annual Review of Economics*, 1, 255–293.

——— (2016): "Power Laws in Economics: An Introduction," *Journal of Economic Perspectives*, 30, 185–206.

GABAIX, X. AND R. IBRAGIMOV (2011): "Rank−1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents," *Journal of Business and Economic Statistics*, 29, 24–39.

GABAIX, X., J.-M. LASRY, P.-L. LIONS, AND B. MOLL (2016): "The Dynamics of Inequality," *Econometrica*, 84, 2071–2111.

GIBRAT, R. (1931): *Les Inégalités Économiques*, Paris: Librairie du Recueil Sirey.

GIESEN, K. AND J. SUEDEKUM (2014): "City Age and City Size," *European Economic Review*, 71, 193–208.

GIESEN, K., A. ZIMMERMANN, AND J. SUEDEKUM (2010): "The Size Distribution Across All Cities—Double Pareto Lognormal Strikes," *Journal of Urban Economics*, 68, 129–137.

GRAHAM, S. W. AND J. D. VAALER (1981): "A Class of Extremal Functions for the Fourier Transform," *Transactions of the American Mathematical Society*, 265, 283–302.

HAMILTON, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

HORN, R. A. AND C. R. JOHNSON (1985): *Matrix Analysis*, New York: Cambridge University Press.

HOWLAND, J. S. (1971): "Simple Poles of Operator-Valued Functions," *Journal of Mathematical Analysis and Applications*, 36, 12–21.

HUGGETT, M. (1993): "The Risk-Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies," *Journal of Economic Dynamics and Control*, 17, 953–969.

——— (1996): "Wealth Distribution in Life-Cycle Economies," *Journal of Monetary Economics*, 38, 469–494.

JOHANSEN, S. (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551–1580.

——— (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford, UK: Oxford University Press.

Kalecki, M. (1945): "On the Gibrat Distribution," *Econometrica*, 13, 161–170.

Kamihigashi, T. (2014): "Elementary Results on Solutions to the Bellman Equation of Dynamic Programming: Existence, Uniqueness, and Convergence," *Economic Theory*, 56, 251–273.

Kapteyn, J. C. (1903): *Skew Frequency Curves in Biology and Statistics*, Groningen: P. Noordhoff.

Karlin, S. (1982): "Classifications of Selection-Migration Structures and Conditions for a Protected Polymorphism," in *Evolutionary Biology*, ed. by M. K. Hecht, B. Wallace, and G. T. Prance, New York: Plenum, vol. 14, chap. 2, 61–204.

Kesten, H. (1973): "Random Difference Equations and Renewal Theory for Products of Random Matrices," *Acta Mathematica*, 131, 207–248.

Klass, O. S., O. Biham, M. Levy, O. Malcai, and S. Solomon (2006): "The Forbes 400 and the Pareto Wealth Distribution," *Economics Letters*, 90, 290–295.

Korevaar, J. (2004): *Tauberian Theory: A Century of Developments*, no. 329 in Grunclehren der mathematischen Wissenschaften, Berlin: Springer.

Kotz, S., T. J. Kozubowski, and K. Podgórski (2001): *The Laplace Distribution and Generalizations*, Boston: Birkhäuser.

Lévy, P. (1937): *Théorie de l'Addition des Variables Aléatoires*, Paris: Gauthier-Villars.

Lindeberg, J. W. (1922): "Eine Neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung," *Mathematische Zeitschrift*, 15, 211–225.

Lukacs, E. (1970): *Characteristic Functions*, London: Griffin, second ed.

Luttmer, E. G. J. (2007): "Selection, Growth, and the Size Distribution of Firms," *Quarterly Journal of Economics*, 122, 1103–1144.

Manrubia, S. C. and D. H. Zanette (1999): "Stochastic Multiplicative Processes with Reset Events," *Physical Review E*, 59, 4945–4948.

Mimica, A. (2016): "Exponential Decay of Measures and Tauberian Theorems," *Journal of Mathematical Analysis and Applications*, 40, 266–285.

Mitzenmacher, M. (2004): "A Brief History of Generative Models for Power Law and Lognormal Distributions," *Internet Mathematics*, 1, 226–252.

NAKAGAWA, K. (2007): "Application of Tauberian Theorem to the Exponential Decay of the Tail Probability of a Random Variable," *IEEE Transactions on Information Theory*, 53, 3239–3249.

NIREI, M. AND S. AOKI (2016): "Pareto Distribution of Income in Neoclassical Growth Models," *Review of Economic Dynamics*, 20, 25–42.

NIREI, M. AND W. SOUMA (2007): "A Two Factor Model of Income Distribution Dynamics," *Review of Income and Wealth*, 53, 440–459.

PARETO, V. (1896): *Cours d'Économie Politique*, Lausanne: F. Rouge.

REED, W. J. (2001): "The Pareto, Zipf and Other Power Laws," *Economics Letters*, 74, 15–19.

——— (2002): "On the Rank-Size Distribution for Human Settlements," *Journal of Regional Science*, 42, 1–17.

——— (2003): "The Pareto Law of Incomes—An Explanation and an Extension," *Physica A*, 319, 469–486.

REED, W. J. AND B. D. HUGHES (2002): "From Gene Families and Genera to Incomes and Internet File Sizes: Why Power Laws Are So Common in Nature," *Physical Review E*, 66, 067103.

——— (2003): "On the Distribution of Family Names," *Physica A*, 319, 579–590.

REED, W. J. AND M. JORGENSEN (2004): "The Double Pareto-Lognormal Distribution—A New Parametric Model for Size Distribution," *Communications in Statistics—Theory and Methods*, 33, 1733–1753.

ROITERSHTEIN, A. (2007): "One-Dimensional Linear Recursions with Markov-Dependent Coefficients," *Annals of Applied Probability*, 17, 572–608.

RUTHERFORD, R. S. G. (1955): "Income Distributions: A New Model," *Econometrica*, 23, 277–294.

SCHUMACHER, J. M. (1986): "Residue Formulas for Meromorphic Matrices," in *Computational and Combinatorial Methods in Systems Theory*, ed. by C. I. Byrnes and A. Lindquist, North-Holland, 97–111.

——— (1991): "System-Theoretic Trends in Econometrics," in *Mathematical System Theory: The Influence of R. E. Kalman*, ed. by A. C. Antoulas, Berlin: Springer, 559–577.

SHAKAI KOUGAKU KENKYUJO (1974): *Nihon Rettou Ni Okeru Jinkou Bunpu No Chouki Jikeiretsu Bunseki (Long Time Series Analysis of Population Distribution in Japan)*, Shakai Kougaku Kenkyujo (Social Engineering Research Institute).

Simon, H. A. and C. P. Bonini (1958): "The Size Distribution of Business Firms," *American Economic Review*, 48, 607–617.

Steinberg, S. (1968): "Meromorphic Families of Compact Operators," *Archive for Rational Mechanics and Analysis*, 31, 372–379.

Sutton, J. (1997): "Gibrat's Legacy," *Journal of Economic Literature*, 35, 40–59.

Toda, A. A. (2011): "Income Dynamics with a Stationary Double Pareto Distribution," *Physical Review E*, 83, 046122.

——— (2012): "The Double Power Law in Income Distribution: Explanations and Evidence," *Journal of Economic Behavior and Organization*, 84, 364–381.

——— (2014): "Incomplete Market Dynamics and Cross-Sectional Distributions," *Journal of Economic Theory*, 154, 310–348.

——— (2017a): "Huggett Economies with Multiple Stationary Equilibria," *Journal of Economic Dynamics and Control*, 84, 77–90.

——— (2017b): "A Note on the Size Distribution of Consumption: More Double Pareto than Lognormal," *Macroeconomic Dynamics*, 21, 1508–1518.

Toda, A. A. and K. Walsh (2015): "The Double Power Law in Consumption and Implications for Testing Euler Equations," *Journal of Political Economy*, 123, 1177–1200.

Toda, A. A. and K. J. Walsh (2017): "Fat Tails and Spurious Estimation of Consumption-Based Asset Pricing Models," *Journal of Applied Econometrics*, 32, 1156–1177.

Vahrenkamp, R. (1976): "Derivatives of the Dominant Root," *Applied Mathematics and Computation*, 2, 29–39.

Vermeulen, P. (2017): "How Fat is the Top Tail of the Wealth Distribution?" *Review of Income and Wealth*.

Wang, N. (2003): "Caballero Meets Bewley: The Permanent-Income Hypothesis in General Equilibrium," *American Economic Review*, 93, 927–936.

Widder, D. V. (1941): *The Laplace Transform*, Princeton, NJ: Princeton University Press.

Wold, H. O. A. and P. Whittle (1957): "A Model Explaining the Pareto Distribution of Wealth," *Econometrica*, 25, 591–595.

Yaari, M. E. (1965): "Uncertain Lifetime, Life Insurance, and the Theory of the Consumer," *Review of Economic Studies*, 32, 137–150.

Appendices

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 define the continuous functions $K_\lambda(x), k_\lambda(x)$ by

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

for all real x that are not integer multiples of ω. 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_λ and k_λ satisfy the following properties.*

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

(ii) *K_λ and k_λ 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 $|E(\sigma(y-x))\mathrm{e}^{\sigma y}| \le 1 \vee \mathrm{e}^{\alpha x}$ and $|E(\alpha(y-x))\mathrm{e}^{\sigma y}| \le \mathrm{e}^{\alpha x}$, two applications of the dominated convergence theorem with dominating function $1 \vee \mathrm{e}^{\alpha x}$ (constant as a function of y) reveal that

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

Therefore,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $a_{k_1k_2}, a_{k_2k_3}, \dots, a_{k_{p-1}k_p}$ 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 followings 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 $B = \mathrm{e}^{i\phi}DAD^{-1}$, 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). □

Proposition B.3. *Let A be a square, nonnegative, and irreducible matrix, and B be a complex matrix of the same dimension. Then $\rho(A + zB)$ depends holomorphically on z in a neighborhood of $z = 0$.*

Proof. See Vahrenkamp (1976). □

Proposition B.4 (Theorem 5.2, Karlin, 1982)**.** *Let Π be a nonnegative, irreducible matrix with rows summing to one, D be a positive diagonal matrix, and $f(t) = \rho((tI + (1 - t)\Pi)D)$ for $t \in [0, 1]$. Then $f'(t) \geq 0$.*

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 Ω 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 Ω, 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). $\square$

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. We divide the proof into three steps.

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

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

The first-order condition with respect to c is

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

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

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

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 to demonstrate the optimality of the consumption rule (4.4b), it remains to verify 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).[17] Using the formula for the value function (4.4a) and the budget constraint (4.2), the continuation value in the next period is

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

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 Theorem 4.2. 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 β, R. We divide the proof into three steps.

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

By (4.5), we have

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

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

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

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

where the last inequality is element-by-element. Left-multiplying the stationary distribution $\pi \gg 0$ as an inner product, it follows that

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

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

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

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

Step 2. Let $b(R)$ be the value of $b = (b_s)_{s=1}^S$ implied by (4.5). Then $\min_s b_s(R) \to \infty$ as $R \downarrow 1$.

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

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

as $R \downarrow 1$ because $\beta < 1$.

Step 3. An equilibrium exists. The gross risk-free rate satisfies $1 - p < R \le 1/\beta$.

Let $b(R)$ be the value of $b = (b_s)_{s=1}^S$ implied by (4.5). By (4.7), the equilibrium condition is $g(R) := \pi^\top b(R) - \pi^\top y$. 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 the previous step, we have $\lim_{R \downarrow 1} g(R) = \infty$. Letting $R = 1/\beta$ in (D.5), we obtain

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

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 $1 < \tilde{R} \leq 1/\tilde{\beta} \iff 1 - p < R \leq 1/\beta$. $\square$

E Japanese population data and estimation

E.1 Historical background and data

During the Edo era (1603–1868), Japan was divided into provinces called *han*, which were controlled by feudal lords called *daimyō*. The movement of people across regions was severely restricted during this period. In 1868, the government power was transferred from the Tokugawa shogunate to the Emperor (Meiji Restoration) and Japan transitioned to a modern nation with a market-based economy. In 1871 the Meiji government abolished the *han* system (*haihan-chiken*). *Han*s were reorganized into prefectures. Initially (August 1871) there were 305 prefectures (302 *ken*s plus Tokyo, Osaka, Kyoto), but they were consolidated into 75 by November 1871. The prefectures were further consolidated into 72 in 1872, 63 in 1873, 62 in 1875, and 38 in 1876. In 1889, some of them were divided and the total number became 46 plus Hokkaido, which is the current number.

There are 47 modern prefectures in Japan. Since 1920, the census has been conducted every five years. The early census was irregular, conducted in 1873, 1884, 1893, 1908, and 1913. Special census was conducted in 1944, 1945, 1946, and 1947 due to WWII and its aftermath. Statistics Japan[18] computes prefecture population data for all years (as of October 1) since 1920 by aggregating the residence registration data in each city. Since completing residence registration is a prerequisite for receiving government service, the quality of this data is high. The population data come in spreadsheets for 1920-2000, 2000-2010, and each year thereafter. Combining these spreadsheets, we can construct a balanced panel of

[18] http://www.stat.go.jp/data/jinsui/2.htm

annual prefecture population data from 1920 to 2015 except for the prefecture of Okinawa, which was occupied by U.S. from 1945 to 1972.

Before 1920, the boundaries of prefectures changed several times. Using the census data in 1873, 1884, 1893, 1908, and 1913, the Japanese urban planning book Shakai Kougaku Kenkyujo (1974) has imputed the prefecture population data with the 1920 boundaries. Due to data quality concerns, we use this data only for the cross-sectional estimation and not the panel one (which is our focus). Figure E.1 plots the population of selected prefectures over time.

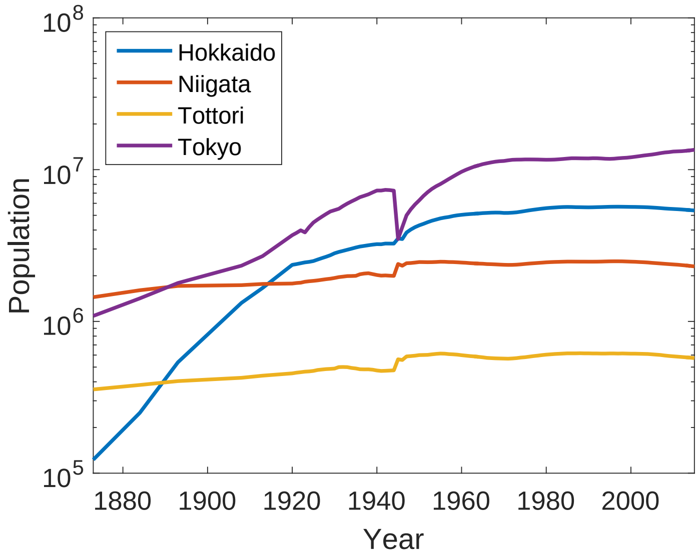


Figure E.1: Population of selected prefectures.

The prefectures are Hokkaido (smallest in 1873), Niigata (largest in 1873), Tottori (smallest in 2015), and Tokyo (largest in 2015). Hokkaido (the northern island) experienced high growth before WWII, which is largely due to migration. (Before the Meiji Restoration in 1868, Hokkaido was an undeveloped territory inhabited by the *Ainu* people, who were hunter-gatherers. After 1868 the Japanese government encouraged migration and development.) Tokyo's population has two dips, one in 1923 (due to the earthquake on September 1 that killed more than 100,000) and another in 1945 (due to refugees and death during WWII). Niigata is a north-central prefecture whose main industry is agriculture (rice). It was the largest prefecture in 1873 but has experienced little growth (except in 1945 when it received war refugees from other urban areas). The graph for Tottori (a rural western prefecture) is almost parallel to Niigata.

E.2 Cross-sectional estimation

A Pareto-lognormal (PlN) random variable is the product of a lognormal variable and an independent Pareto variable with minimum size 1 (normalization) and Pareto exponent $\alpha > 0$ (Reed and Jorgensen, 2004). The distribution of the logarithm of a PlN variable is the convolution of the normal and exponential distributions, which is known as the normal-exponential distribution. It has three parameters, the mean μ and standard deviation $\sigma > 0$ of the normal component and the exponent $\alpha > 0$ of the exponential component. The density of the normal-exponential distribution is given by

$$f(x; \mu, \sigma, \alpha) = \alpha \mathrm{e}^{-\alpha(x-\mu)+\alpha^2\sigma^2/2} \Phi\left(\frac{x-\mu}{\sigma} - \alpha\sigma\right),$$

where $\Phi(\cdot)$ is the cdf of $N(0,1)$. Maximum likelihood estimation is straightforward using this density.

E.3 Panel estimation

We estimate the hidden Markov model using the Hamilton (1989) filter. The following algorithm computes the log-likelihood. (The algorithm considers the case of one time series, but the panel case is similar.)

(i) Given the transition probability matrix Π, compute the stationary distribution π as the right Perron vector of Π that sums to 1. Set $\xi_{0|0} = \pi$.

(ii) For $t = 1, \dots, T$, do the following:

   (a) Let $\xi_{t|t-1} = \Pi^\top \xi_{t-1|t-1}$ be the one period ahead prior forecast.

   (b) Define the vector of conditional likelihoods $\eta_t = (\eta_{nt})_{n=1}^N$ by

   $$\eta_{nt} = \frac{1}{\sqrt{2\pi\sigma_n^2}} \mathrm{e}^{-\frac{(\log G_t - \mu_n)^2}{2\sigma_n^2}},$$

   where $\log G_t$ is the realized log growth rate.

   (c) Update the one period ahead forecast by $\xi_{t|t} = (\eta_t \odot \xi_{t|t-1})/(\eta_t^\top \xi_{t|t-1})$, where $\odot$ denotes the Hadamard (element-wise) product.

(iii) Compute the log likelihood as $\log L = \sum_{t=1}^T \log(\eta_t^\top \xi_{t|t-1})$.

We then estimate the parameters Π, $\mu = (\mu_1, \dots, \mu_N)^\top$, and $\sigma = (\sigma_1, \dots, \sigma_N)^\top$ by maximizing the log-likelihood.