Zipf's Law: A Microfoundation

Alexis Akira Toda*

This version: October 7, 2016

Abstract

Existing explanations of Zipf's law (Pareto exponent approximately equal to 1) in size distributions require strong assumptions on growth rates or the minimum size. I show that Zipf's law naturally arises in general equilibrium when individual units solve a homogeneous problem (*e.g.*, homothetic preferences, constant-returns-to-scale technology), the units appear and disappear at a small constant rate (*e.g.*, rare disasters), and at least one production factor in the economy is in limited supply. My model explains why Zipf's law is empirically observed in the size distributions of cities and firms, which consist of people, but not in other quantities such as wealth, income, or consumption, which all have Pareto exponents well above 1.

Keywords: Gibrat's law, homogeneous problem, power law, rare disasters

JEL codes: D30, D52, D58, L11, R12

1 Introduction

Zipf's law is an empirical regularity that holds in the size distributions of cities and firms, stating that the frequency of observing a unit larger than the cutoff x is approximately inversely proportional to x:

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

where the Pareto (power law) exponent ζ is slightly above 1. This relationship holds regardless of the choice of countries or time periods.[1] To get a sense of how the empirical size distribution looks like, Figure 1 shows a log-log plot of employment cutoffs and the number of firms larger than the cutoffs (essentially the ranks) using the 2011 U.S. Census Small Business Administration (SBA) data. Consistent with a power law, the figure shows a straight-line pattern up to small firms with as few as 10 employees. The estimated Pareto exponent is $\widehat{\zeta} = 1.0972$ with standard error 0.0788. We obtain similar patterns for all years from 1992 to 2011 for which data is available. Figure 2 shows the estimated Pareto exponent over the period 1992–2011. In all years the Pareto exponent is slightly above 1.

*Department of Economics, University of California San Diego. Email: atoda@ucsd.edu. I thank Eli Berman, Jim Rauch, seminar participants at Fudan, Keio, Kobe University Research Institute for Economics and Business Administration, Osaka, and UCSD for comments and feedback. Miles Berg and Xuan Fang has provided excellent research assistance.

[1] For empirical studies documenting Zipf's law, see Zipf (1949), Rosen and Resnick (1980) (cities), and Axtell (2001) (firms), among others.

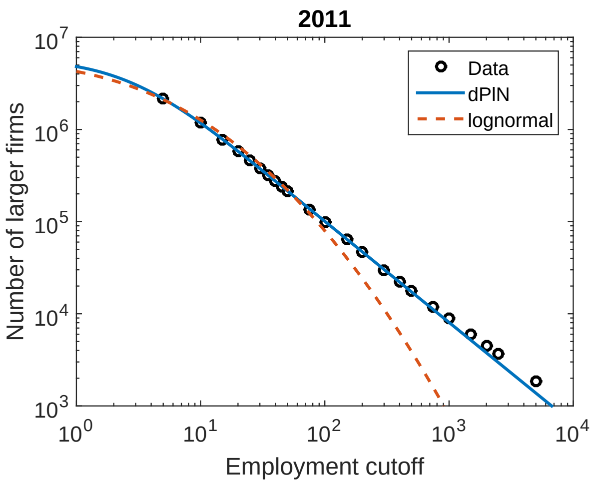

Figure 1: Log-log plot of firm size distribution.

Note: The figure plots employment cutoffs and the number of firms larger than the cutoffs (ranks). dPlN stands for *double Pareto-lognormal*, which is a distribution arising from the theoretical model in the paper. The straight-line pattern is consistent with a power law, with estimated exponent $\widehat{\zeta} = 1.0972$ and standard error 0.0788 using maximum likelihood with binned data. Source: 2011 U.S. Census Small Business Administration data.

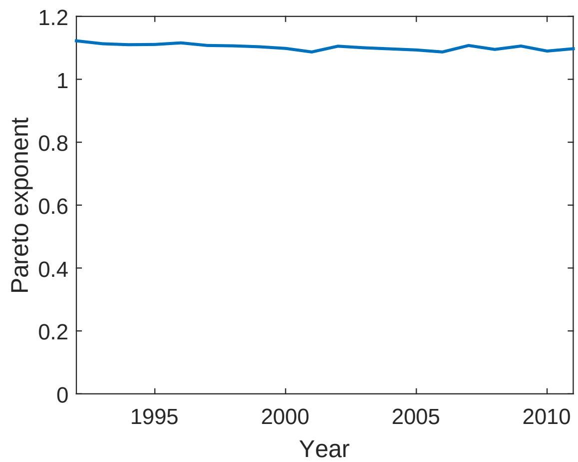

Figure 2: Time series of estimated Pareto exponent.

Note: The figure plots the estimated Pareto exponent $\widehat{\zeta}$ from 1992 to 2011. Source: U.S. Census Small Business Administration data.

In a seminal paper, Gabaix (1999) has shown that Zipf's law arises when individual units follow Gibrat (1931)'s law of proportional growth and there is some small minimum size that the units must meet. His work has generated a large subsequent literature on power laws in economics and finance as well as models that explain Zipf's law.[2] Despite the considerable advances in the theory

[2]The theoretical literature is too large to review here. Examples include Luttmer (2007), Nirei and Souma (2007), Rossi-Hansberg and Wright (2007), Benhabib et al. (2011, 2015,

of power laws in size distributions made during the past decade or so, the explanation of Zipf's law (Pareto exponent very close to 1) remains incomplete: one of the conditions for Zipf's law—that there is a small minimum size, or equivalently the expected growth rate of existing units is small in absolute value—has often been assumed without proper justifications. In this paper, I show that in a certain class of dynamic general equilibrium models, Zipf's law holds without introducing ad hoc assumptions, and hence provide a microfoundation for Zipf's law.

My theory is surprisingly simple, and essentially relies on the following three elements: (i) Gibrat's law of proportional growth, (ii) reset events that occur with small probability ("idiosyncratic rare disasters"), and (iii) existence of a production factor in limited supply. Conditions (i) and (ii) have already been known to be sufficient to generate Pareto tails (Reed, 2001), but Zipf's law (Pareto exponent close to 1) holds only in the knife-edge case in which the expected growth rate of units is small in absolute value. My contribution is thus in showing that condition (iii)—the existence of a production factor in limited supply—limits aggregate growth, which *in equilibrium* also limits individual growth and hence generates Zipf's law. Note that the low growth condition is an endogenous outcome, not an assumption.

To illustrate these points in the simplest possible way, I first construct a stylized model of the population dynamics in cities (villages). In the model, there are a continuum of villages and households. The village authorities produce a single good ("potato") using a constant-returns-to-scale technology and hiring labor. Households migrate across villages freely without any cost. Villages are hit by two types of idiosyncratic shocks—technological shocks and rare disasters ("famine"). When a famine occurs, the potatoes in the village are wiped out, but the village authority receives deliveries of potatoes from other villages because they have a mutual insurance. This simple model has all the ingredients sufficient to generate Zipf's law: (i) with multiplicative technological shocks and constant-returns-to-scale technology, we obtain Gibrat's law for individual villages, (ii) famines are reset events and generate a stationary distribution with Pareto tails, and (iii) the inelastic labor supply endogenously forces the expected population growth rate in individual villages to be small in equilibrium, generating Zipf's law.

The intuition for this simple model carries over to more general models. Consider a dynamic general equilibrium model which consists of several agent types, and suppose that we are interested in the size distribution of an economic variable of a particular type (*e.g.*, firm size distribution measured by the number of employees). The main result of this paper, Theorem 3.6, shows that if agents of this type solves a homogeneous problem (*e.g.*, homothetic preferences, constant-returns-to-scale technology, proportional constraints), the agents appear and disappear at a constant rate $\eta > 0$, and at least one factor of production is in limited supply, then Zipf's law holds in the stationary equilibrium as $\eta \to 0$.

Because the main theorem is an asymptotic result, the Pareto exponent need not be close to 1 for particular models or parameter configurations. To address the quantitative validity of my theory, I construct a model of entrepreneurship and firm size distribution. The economy is populated by entrepreneur-CEOs and

2016), Gabaix (2011), Toda (2014), Toda and Walsh (2015), Arkolakis (2016), Gabaix et al. (2015), Nirei and Aoki (2016), and Aoki and Nirei (2016), among others. See Gabaix (2009, 2016) and Jones (2015) for reviews.

household-workers. Each entrepreneur operates a firm using a constant-returns-to-scale technology and hiring labor, and makes consumption-saving-portfolio-hiring decisions optimally. Entrepreneurs are subject to idiosyncratic investment risk and bankruptcy. Workers supply labor inelastically but make consumption-saving decisions optimally. In this setting under mild conditions I prove that a unique stationary equilibrium exists and characterize the equilibrium in closed-form. I prove that the stationary firm size distribution obeys Zipf's law when the bankruptcy rate is small. I calibrate the model to the U.S. economy and find that the Pareto exponent is close to 1, consistent with Zipf's law. To show its robustness, I generate random parameter configurations drawn from a uniform distribution with a large support, and for each case I compute the equilibrium Pareto exponent. For this particular model I find that the 95 percentile of the Pareto exponent is 1.13, so Zipf's law holds even for quite extreme parameter configurations, confirming its robustness.

2 Existing explanations and difficulties

In this section I review the existing explanations of Zipf's law based on random growth models[3] and point out their difficulties.

2.1 Geometric Brownian motion with minimum size

Suppose that the size of individual units (*e.g.*, population of cities, number of employees in firms, etc.) satisfies Gibrat (1931)'s law of proportional growth: the growth rate of units is independent of their sizes.[4] The simplest of all such processes is the geometric Brownian motion (GBM)

$$\mathrm{d}X_t = gX_t\,\mathrm{d}t + vX_t\,\mathrm{d}B_t, \tag{2.1}$$

where X_t is the size of a typical unit,[5] g is the expected growth rate, $v > 0$ is the volatility, and B_t is a standard Brownian motion that is independent across units. As is well known, the geometric Brownian motion leads to the lognormal distribution whose log variance increases linearly over time, and hence does not admit a stationary distribution.

In order to obtain a stationary distribution, a common practice in the literature is to introduce a minimum size $x_{\min} > 0$ below which individual units cannot operate.[6] Mathematically, we are considering the geometric Brownian motion with a reflective barrier at $x_{\min}$. Assuming that the growth rate is negative ($g < 0$), it is well known (see Appendix A) that the system converges to

[3]I focus on the random growth model because (i) it is the earliest model to explain power laws (Champernowne, 1953; Simon, 1955), and (ii) almost all existing explanations rely on this mechanism one way or another. An exception is Geerolf (2016), who studies the production decision within an organization in a static setting. The Pareto exponent is exactly equal to 2 when there are two layers in the organization (*e.g.*, managers and workers). He also shows that Zipf's law obtains as the number of layers tends to infinity.

[4]See Sutton (1997) for a review of the empirical literature on Gibrat's law.

[5]X_t is sometimes interpreted as the size of a typical unit relative to the cross-sectional average. In that case, g and v are the expected growth rate and volatility relative to the average, and $x_{\min}$ below is the minimum relative size.

[6]Such assumptions are made in Levy and Solomon (1996), Gabaix (1999), Malcai et al. (1999), Luttmer (2007), Rossi-Hansberg and Wright (2007), Córdoba (2008), and Aoki and Nirei (2016), among others.

the unique stationary distribution

$$P(X > x) = \left(\frac{x}{x_{\min}}\right)^{-\zeta}, \tag{2.2}$$

which is a Pareto distribution with minimum size $x_{\min}$ and Pareto exponent

$$\zeta = 1 - \frac{2g}{v^2} > 1. \tag{2.3}$$

Thus we obtain Zipf's law ($\zeta \approx 1$) when the growth rate is small in absolute value relative to the variance: $|g| \ll v^2$. Another way to formulate the condition for Zipf's law is to compare the minimum size $x_{\min}$ to the average size $\bar{x}$. Using the distribution function (2.2), the average size is

$$\bar{x} = \int_{x_{\min}}^{\infty} x\zeta x_{\min}^{\zeta} x^{-\zeta-1}\,\mathrm{d}x = \frac{\zeta}{\zeta-1}x_{\min} \iff \zeta = \frac{1}{1 - x_{\min}/\bar{x}}. \tag{2.4}$$

Hence Zipf's law is also equivalent to $x_{\min} \ll \bar{x}$: the minimum size is small relative to the average. The intuition is that the minimum size is small relative to the average when the latter is large, which occurs precisely when the expected growth rate g is large, or when it is close to zero since it must be negative.

2.2 Geometric Brownian motion with birth/death

Next, consider a purely mechanistic model as in Reed (2001), where the size of individual units X_t evolves according to the geometric Brownian motion (2.1) but with a constant probability of birth/death.[7] Unlike in the previous example, there is no minimum size but new units are constantly born at rate $\eta > 0$, with initial size x_0, and existing units die at the same rate η.[8] It is well known (see Appendix A) that regardless of the parameter values, the size distribution of units always has a unique stationary distribution, with a density of the form

$$f(x) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta} x_0^{\alpha} x^{-\alpha-1}, & (x \ge x_0) \\ \frac{\alpha\beta}{\alpha+\beta} x_0^{-\beta} x^{\beta-1}, & (0 < x < x_0) \end{cases} \tag{2.5}$$

which is known as *double Pareto*. $\alpha, \beta > 0$ are called Pareto (or power law) exponents. Given the parameters g, v, η of the stochastic process, the exponents $\zeta = \alpha, -\beta$ are the solutions to the quadratic equation

$$\frac{v^2}{2}\zeta^2 + \left(g - \frac{v^2}{2}\right)\zeta - \eta = 0. \tag{2.6}$$

[7]Wold and Whittle (1957) is one of the earliest examples that shows that random birth/death can generate Pareto tails. Working in continuous-time is convenient for tractability. Although the results in this section are exact only in continuous-time, Toda (2014) shows that it is also approximately true in discrete-time under general Markov processes.

[8]For cities it may be unreasonable to assume that they disappear at a constant rate. However, this assumption is not important because we obtain the exact same result if cities are infinitely lived, new cities are created at rate η, and the total population also grows at rate η. Also it is not important that the average size of cities is constant over time. If there is population growth, we obtain the same conclusion by considering the balanced growth path. See the discussion in Reed (2001) for details.

Solving (2.6), we obtain the Pareto exponents

$$\alpha, \beta = \frac{1}{2}\left(\sqrt{\left(1 - \frac{2g}{v^2}\right)^2 + \frac{8\eta}{v^2}} \pm \left(1 - \frac{2g}{v^2}\right)\right). \tag{2.7}$$

As is clear from this formula, Zipf's law ($\alpha \approx 1$) arises when $g, \eta \ll v^2$, *i.e.*, when the growth rate and death probability are small compared to the variance.

2.3 Difficulties

Although the above models are purely mechanical, they underly the mechanism of generating Zipf's law in virtually all papers. Of course, in order to make it an economic model, one needs to provide mechanisms that generate Gibrat's law of proportional growth. However, this is not difficult if we assume that individual units solve a homogeneous problem (*e.g.*, homothetic preferences, constant-returns-to-scale production, proportional constraints).[9] The more difficult part is to explain why there is a minimum size, and why the growth rate is small. These are the difficulties in existing explanations.

First, in many models a minimum size is often introduced as an ad hoc assumption. While a minimum size may be justified in some cases (*e.g.*, positive integer constraint, fixed cost of operation, borrowing constraints), in the presence of a minimum size, fully optimizing agents will typically behave differently depending on whether they are close to the lower boundary or not. Since Zipf's law is a statement about the upper tail, and large agents are likely not affected much by the lower boundary, it is reasonable to expect that the size distribution is similar in models where (i) agents behave rationally in the presence of an ex ante minimum size,[10] and (ii) agents ignore the minimum size but it is imposed ex post, at least in the upper tail. Therefore the assumption of a minimum size is not really an issue, although characterizing the stationary distribution with fully optimizing agents in the presence of a minimum size is challenging.

The second issue, which is more problematic, is the condition that the growth rate or the minimum size must be small in absolute value in order to obtain Zipf's law, which is a knife-edge case. Since the growth rate g is an endogenous variable in any fully specified economic model, there is no obvious reason why we should expect it to be close to zero. In order to obtain this condition, one usually needs to pick very particular parameter values.[11]

[9] See, for example, Saito (1998), Krebs (2003), Angeletos (2007), Benhabib et al. (2011, 2016), Toda (2014), and Toda and Walsh (2015), among others.

[10] For example, Benhabib et al. (2015) consider a Bewley model with capital income risk and show that the optimal consumption rule is asymptotically linear (*i.e.*, the lower boundary does not matter) as agents become rich. As a result, they show that the stationary wealth distribution exhibits a Pareto upper tail.

[11] For example, Simon and Bonini (1958) consider a random growth model of firm size based on Simon (1955) and show that Zipf's law obtains when the net growth attributed to new firms relative to that of existing firms approaches zero. Luttmer (2007) studies a general equilibrium model of firms with monopolistic competition and entry/exit, and shows that Zipf's law holds when the technology improvement rates of entrants is slightly above those of incumbents. In both of these cases, incumbents will grow slightly slower than the average, and after subtracting the average rate, we obtain the low growth condition $|g| \ll v^2$. Córdoba (2008) studies a model of city size distribution and shows that Zipf's law holds when the elasticity of substitution between goods is exactly 1. See Gabaix (1999) for a review of mechanisms suggested in the earlier literature, which all require a fine-tuning of parameters.

To summarize, the explanation of Zipf's law remains incomplete until we provide a fully specified economic model with optimizing agents in which (i) there is no ad hoc minimum size, and (ii) the low growth condition emerges endogenously as an equilibrium outcome. I provide such models in the following sections.

3 Homogeneity and limited factor yield Zipf

In this section I show that whenever (i) individual units solve a dynamic optimization problem that is homogeneous in the state variable (size) as well as all control variables, (ii) individual units are reset at a constant Poisson rate $\eta > 0$, and (iii) at least one production factor is in limited supply, we obtain Zipf's law in the limit $\eta \to 0$. To motivate the general result I first provide a minimal model of population dynamics and city size distribution to highlight the ingredients that give rise to Zipf's law.

3.1 A simple model of city size distribution

Environment Consider an economy consisting of a continuum of villages and households. The mass of villages and households is normalized to 1 and N, respectively. There is a single consumption good, potato. Each household supplies 1 unit of labor inelastically and consumes the entire wage ("hand-to-mouth" behavior). Households migrate across villages freely without any moving costs; therefore in equilibrium, all villages must offer the same competitive wage. Each village authority uses its stock of potatoes and hires labor to produce new potatoes using a constant-returns-to-scale technology.

Each village is subject to two types of idiosyncratic shocks. First, the stock of potatoes is subject to a productivity shock coming from a Brownian motion. Second, each village is occasionally hit by a rare disaster—famine—which arrives at a (small) Poisson rate $\eta > 0$. When a famine hits a village, the entire stock of potatoes perishes. However, there is a mutual insurance agreement across villages: a village hit by a famine receives a delivery of potatoes from other villages and starts over at size $\kappa > 0$ times the aggregate stock of potatoes; this delivery is financed by contributions from other villages proportional to their stock of potatoes.

A stationary equilibrium is defined by a wage ω and size distributions of village population and stock of potatoes such that (i) profit maximization: given the wage and stock of potatoes, each village authority demands labor to maximize profits,[12] (ii) market clearing: for each village, population equals labor demand, and (iii) stationarity: the size distributions are invariant over time.

Population dynamics of individual villages Let ω be the equilibrium wage and x_t be the stock of potatoes in a typical village. Then the resource constraint when there is no famine is

$$\mathrm{d}x_t = (F(x_t, n_t) - \omega n_t)\,\mathrm{d}t - \eta\kappa x_t\,\mathrm{d}t + v x_t\,\mathrm{d}B_t, \tag{3.1}$$

[12] To keep the analysis as simple as possible, in this model I assume that the village authority maximizes profits point-by-point, without specifying fundamentals on the behavior (*e.g.*, utility function). One way to justify this behavior is to assume that the village authority (dictator) has an additive CRRA utility in the stock of potatoes (*i.e.*, gets utility from looking at potatoes) and the dictator gets replaced whenever a famine occurs.

where n_t is the labor input (population of the village in equilibrium), F is the production function (which is homogeneous of degree 1 since it exhibits constant-returns-to-scale), v is volatility, and B_t is a standard Brownian motion. $F(x_t, n_t) - \omega n_t$ is the amount of potatoes the village authority retains after paying the wage. The term $-\eta\kappa x_t$ reflects the delivery of potatoes to a village hit by a famine (in a short period of time Δt, there are $\eta\Delta t$ such villages, and each village gets κx_t, where $\kappa > 0$ is the constant of proportionality). The term $vx_t \,\mathrm{d}B_t$ is the technological shock to the stock of potatoes. The village authority maximizes the profit, so chooses n_t such that

$$n_t = \arg\max_n (F(x_t, n) - \omega n).$$

Let $f(x) = F(x, 1)$.[13] Since by assumption F is homogeneous of degree 1, we have $F(x, n) = nf(x/n)$. By the first-order condition, we obtain

$$\omega = f(y) - yf'(y), \tag{3.2}$$

where $y = x_t/n$ is the potato per capita. Hence given the wage ω and the stock of potatoes x_t, the labor demand is $n_t = x_t/y$, where y is determined by (3.2). The profit rate per unit of potato is then

$$\mu = \frac{F(x_t, n) - \omega n}{x} = \frac{1}{y}(f(y) - (f(y) - yf'(y))) = f'(y). \tag{3.3}$$

Substituting the profit (3.3) into the resource constraint (3.1), we obtain

$$\mathrm{d}x_t = (\mu - \eta\kappa)x_t \,\mathrm{d}t + vx_t \,\mathrm{d}B_t. \tag{3.4}$$

Therefore the stock of potatoes in each village evolves according to a geometric Brownian motion until a famine hits. Since $n_t = x_t/y$ is proportional to x_t, the village population n_t also obeys the same geometric Brownian motion (3.4).

Equilibrium To compute the equilibrium, we need to derive the dynamics of the aggregate stock of potatoes, X_t (which is constant in steady state). Consider what happens to the stock of potatoes in each village during a short period of time Δt. If the village does not experience a famine (which occurs with probability $1 - \eta\Delta t$), then by (3.4) the stock of potatoes grows at rate $\mu - \eta\kappa$ on average. If the village is hit by a famine (which occurs with probability $\eta\Delta t$), the potatoes are wiped out, and the village receives a delivery of κX_t from other villages according to the mutual agreement. Hence aggregating the stock of potatoes across villages and using the law of large numbers, we obtain

$$\begin{aligned} X + \Delta X &= \underbrace{(1 - \eta\Delta t)(1 + (\mu - \eta\kappa)\Delta t)X}_{\text{Aggregate potatoes of non-famine villages}} + \underbrace{(\eta\Delta t)(\kappa X)}_{\text{Aggregate potatoes of famine villages}} \\ &= (1 + (\mu - \eta)\Delta t)X + \text{higher order terms}. \end{aligned}$$

Subtracting X from both sides and letting $\Delta t \to 0$, we obtain

$$\mathrm{d}X = (\mu - \eta)X \,\mathrm{d}t. \tag{3.5}$$

[13] A typical example is the Cobb-Douglas production function $F(x, n) = Ax^\alpha n^{1-\alpha} - \delta x$, so $f(x) = Ax^\alpha - \delta x$, where δ is the depreciation rate.

In steady state, since by definition the aggregate stock of potatoes is constant, we must have $\mathrm{d}X = 0$ and hence

$$\mu = \eta. \tag{3.6}$$

Combining (3.3) and (3.6), the equilibrium potato per capita y is determined by $f'(y) = \eta$. The equilibrium wage is then determined by (3.2). Substituting (3.6) into the equation of motion (3.4) of potatoes in each village (and hence the population), we obtain

$$\mathrm{d}x_t = \eta(1-\kappa)x_t \,\mathrm{d}t + v x_t \,\mathrm{d}B_t. \tag{3.7}$$

Note that (3.7) is a special case of the mechanistic model (2.1) with $g = \eta(1-\kappa)$. Since η is small, so is g, and hence we can expect Zipf's law to hold. In fact, we can show the following proposition.

Proposition 3.1. *The stationary city size distribution is double Pareto. The upper tail Pareto exponent ζ is given by α in (2.7) with $g = \eta(1-\kappa)$, which satisfies the bound*

$$1 < \zeta < 1 + \frac{2\eta\kappa}{v^2}. \tag{3.8}$$

As $\eta \to 0$, we obtain Zipf's law $\zeta \to 1$.

3.2 General theory

The above stylized model can be generalized as follows.

Consider a dynamic optimization problem with one positive state variable (called "size") denoted by $x > 0$, finitely many control variables denoted by $y \in \mathbb{R}^{d_y}$, and finitely many parameters denoted by $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$. Some parameters may be exogenous (*e.g.*, preference and technology parameters), while others are endogenous (*e.g.*, prices). Furthermore, the parameters may vary over time. Let $\Gamma(x;\theta) \subset \mathbb{R}^{d_y}$ be the constraint set of the control y given the state variable x and parameter θ, and $V(\{x_t, y_t; \theta_t\})$ be the objective function to be maximized. In this paper I introduce the following definition.

Definition 3.2. The dynamic optimization problem is *homogeneous* if for each parameter $\theta \in \Theta$ the followings hold:

1. the constraint function $\Gamma(\cdot;\theta) : \mathbb{R}_+ \rightrightarrows \mathbb{R}^{d_y}$ is homogeneous of degree 1, so for all $\lambda > 0$ we have $y \in \Gamma(x;\theta) \implies \lambda y \in \Gamma(\lambda x;\theta)$,

2. the equation of motion for the state variable is a diffusion with homogeneous coefficients, so

$$\mathrm{d}x_t = g(x_t, y_t; \theta_t)\,\mathrm{d}t + v(x_t, y_t; \theta_t)\,\mathrm{d}B_t, \tag{3.9}$$

   where B_t is a standard Brownian motion and g, v are drift and volatility, which are homogeneous of degree 1 in (x, y),

3. the objective function is homothetic, so for all $\lambda > 0$ and feasible $\{x_t, y_t\}_{t\geq 0}$ and $\{x'_t, y'_t\}_{t\geq 0}$, we have

$$V(\{x_t, y_t; \theta_t\}) \geq V(\{x'_t, y'_t; \theta_t\}) \implies V(\{\lambda x_t, \lambda y_t; \theta_t\}) \geq V(\{\lambda x'_t, \lambda y'_t; \theta_t\}).$$

Example 1. A typical example of a homogeneous problem is a Merton (1969)-type optimal consumption-portfolio problem. In this problem the investors maximize the expected utility

$$\mathrm{E}_0 \int_0^\infty \mathrm{e}^{-\rho t} \frac{c_t^{1-\gamma}}{1-\gamma} \,\mathrm{d}t$$

subject to the budget constraint

$$\mathrm{d}x_t = (rx_t + (\mu - r)s_t - c_t)\,\mathrm{d}t + \sigma s_t \,\mathrm{d}B_t,$$

where x_t is total wealth, s_t is the amount of wealth invested in the risky asset (stock), c_t is consumption, r is the risk-free rate, μ is the expected return on stocks, and σ is volatility. In this case the control variable is $y = (c, s)$ and the parameter is $\theta = (\rho, \gamma, r, \mu, \sigma)$. Since consumption is nonnegative, the constraint set is $y = (c, s) \in \mathbb{R}_+ \times \mathbb{R} = \Gamma(x, \theta)$, which is homogeneous of degree 1. Clearly the objective function is homogeneous of degree $1-\gamma$ in $\{c_t\}$, and the drift and volatility

$$g(x, y; \theta) = rx + (\mu - r)s - c, \quad v(x, y; \theta) = \sigma s$$

are homogeneous of degree 1 (in fact, linear) in (x, y).

As is well known, the solution to a homogeneous problem scales with the state variable.

Lemma 3.3. *If $\{y_t\}$ solves a homogeneous problem, then there exists a function $\alpha_t : \Theta \to \mathbb{R}^L$ such that $y_t = \alpha_t(\theta_t)x_t$.*

Proof. By homogeneity, if y is the optimal control given the state $x > 0$ and parameter θ, λy is the optimal control given the state λx and parameter θ. Letting $\lambda = 1/x$, y/x is the optimal control given the state 1 and parameter θ, which we can denote by $\alpha_t(\theta) \in \mathbb{R}^L$. Therefore $y = \alpha_t(\theta_t)x$. □

Consider the class of dynamic general equilibrium models that consist of one or several types of agents and feature only idiosyncratic risks. I define two notions of equilibria.

Definition 3.4. An *aggregate steady state* consists of endogenous parameters and decision rules of all agent types such that (i) agents optimize, (ii) markets clear, and (iii) all endogenous parameters and decision rules are time-invariant. If in addition the cross-sectional distributions of all agent types are time-invariant, the aggregate steady state is called a *stationary equilibrium*.

Suppose that in a dynamic general equilibrium model, a particular agent type solves a homogeneous problem. Since there is only idiosyncratic risks, the Brownian motion in (3.9) is i.i.d. across all agents.

The following lemma shows that if the aggregate supply of at least one positive control variable is bounded, then in a steady state the cross-sectional size distribution has a finite mean.

Lemma 3.5. *Suppose that a dynamic general equilibrium model has an aggregate steady state, and that one agent type solves a homogeneous problem. If the aggregate supply of at least one positive control variable is bounded, then the cross-sectional size distribution of that type has a finite mean. Furthermore, the size of individual units obeys some geometric Brownian motion*

$$\mathrm{d}x_t = g(\theta)x_t \,\mathrm{d}t + v(\theta)x_t \,\mathrm{d}B_t. \tag{3.10}$$

Using Lemmas 3.3 and 3.5, we can prove the main result: homogeneity, limited supply, and a (small) constant rate of Poisson birth/death yield Zipf's law.

Theorem 3.6. *Let everything be as in Lemma 3.5 and suppose that individual units of that particular type are born and disappear at a constant Poisson rate $\eta > 0$, and new units are drawn from some initial size distribution $x_0 \sim F(x; \theta, \eta)$ with finite mean. Assume that a stationary equilibrium exists and let $\theta(\eta) \in \Theta$ be all exogenous and endogenous parameters in stationary equilibrium given $\eta > 0$,*

$$\kappa(\eta) = \frac{\int_0^\infty x F(\mathrm{d}x; \theta(\eta), \eta)}{\mathrm{E}[x_t]} > 0$$

be the average initial size relative to the cross-sectional mean, and $v(\eta) := v(\theta(\eta)) > 0$ be the volatility. Then the cross-sectional size distribution has a Pareto upper tail with exponent ζ that satisfies the bound (3.8). In particular, if

$$\lim_{\eta \to 0} \frac{\eta \kappa(\eta)}{v(\eta)^2} = 0, \tag{3.11}$$

then $\zeta \to 1$ as $\eta \to 0$, so we obtain Zipf's law.

Theorem 3.6 is quite powerful since we obtain Zipf's law regardless of the details of the model ("detail-free"). All we need are that (i) individual units solve a homogeneous problem,[14] so the size variable obeys the geometric Brownian motion, (ii) individual units appear and disappear at a constant Poisson rate, so the cross-sectional distribution is double Pareto, and (iii) there is a factor in the economy that is in limited supply, so in equilibrium all aggregate variables remain bounded, which forces the growth rate of GBM to be small in absolute value and makes the Pareto exponent close to 1.

Of course, Theorem 3.6 assumes that a stationary equilibrium exists and the technical condition (3.11) holds. In general, for a given model we need to verify these conditions on a case-by-case basis. In the simple model of city size distribution in Section 3.1, these conditions are trivial since we can construct an equilibrium analytically and $\kappa, v > 0$ are exogenous constants.

3.3 Robustness

In this section I show that the assumptions of Theorem 3.6 are satisfied in a wide variety of models and that the assumptions can be weakened further.

3.3.1 Elastic labor supply

In the city size example in Section 3.1, households supply labor inelastically. This assumption is inessential, since village authorities still solve a homogeneous problem regardless of whether labor supply is inelastic or not, and therefore the assumptions of Theorem 3.6 hold. Even if households make some labor-leisure choice, the conclusion of Proposition 3.1 remains valid because the total population is bounded and hence so is the total labor supply.

[14] Clearly, it is not necessary that *all* agent types solve homogeneous problems. All we need is that individual units of a particular type whose distribution we are interested in solve a homogeneous problem.

In other models, such as Angeletos and Panousi (2009, 2011), there is a single type of agents (entrepreneur-workers) that operates a constant-returns-to-scale technology while choosing labor supply and demand. In this case the individual problem is not homogeneous because labor-leisure choice is bounded. However, after computing the present value of wage and fixing the labor-leisure choice at the optimum, the remaining problem (optimal consumption-portfolio choice) becomes a homogeneous problem. Therefore Zipf's law still holds in this case.

3.3.2 Balanced growth equilibrium

In the city size example in Section 3.1, I assumed that the total population is constant at N, and hence bounded. Boundedness of some factor is sufficient for Zipf's law, but not necessary. Suppose, for example, that population grows (or shrink) at a constant rate ν, so $N_t = N_0 \mathrm{e}^{\nu t}$. Since the equation of motion for the aggregate stock (3.5) still holds, we have a balanced growth equilibrium if and only if

$$\mu - \eta = \nu.$$

In this case the growth rate of individual cities relative to the mean is

$$g - \nu = (\mu - \eta\kappa) - \nu = \eta(1 - \kappa),$$

which is exactly the same as in the case with no population growth. Therefore in the balanced growth equilibrium, the mean of the cross-sectional distribution will grow at rate ν, but the upper tail Pareto exponent will still satisfy the bound (3.8). Hence we obtain Zipf's law as $\eta \to 0$.

3.3.3 Random initial size

When the initial size of newborn units is constant, by Proposition 3.1 the cross-sectional distribution is exactly double Pareto. Since the double Pareto distribution has a kink at the mode, it is unlikely to be observed in the data. Reed (2002) and Giesen et al. (2010) suggest that the entire size distribution of cities is closer to the *double Pareto-lognormal* (dPlN) distribution, which has two Pareto tails with a lognormal body (Reed, 2003). It is straightforward to obtain dPlN in my model: instead of assuming that the initial size after the reset event is constant, if the initial size distribution is lognormal, we obtain dPlN. Therefore my model can explain simultaneously why the size distribution of cities is close to dPlN and obeys Zipf's law. More generally, as long as the initial size distribution is thin-tailed, the initial size does not affect the upper tail of the cross-sectional distribution since the latter is governed by the distribution of relative size, which is fat-tailed.

3.3.4 Multiple types

In the empirical literature on firm sizes (Sutton, 1997), it is well known that Gibrat's law of proportional growth does not quite hold: small firms tend to grow faster but also exit at a higher rate. My theory is not necessarily inconsistent with these empirical facts. Suppose, for instance, that firms consist of several types, indexed by $j = 1, \dots, J$. Suppose that all firm types solve a (type-specific) homogeneous problem, and hence by Lemma 3.3, in a stationary equilibrium the size of type j firms evolve according to a geometric Brownian motion with growth rate g_j and volatility $v_j > 0$. Suppose also that type j firms either go bankrupt or transition to a different type at rate $\eta_j > 0$. Letting $\kappa_j > 0$ be the average initial size of new type j firms relative to the average existing type j firms, it follows from Theorem 3.6 that the cross-sectional size distribution of type j firms has a Pareto exponent ζ_j that satisfies

$$1 < \zeta_j < 1 + \frac{2\eta_j \kappa_j}{v_j^2}.$$

Therefore the entire cross-sectional firm size distribution has a Pareto exponent ζ that satisfies

$$1 < \zeta \le \min_j \zeta_j < 1 + \min_j \frac{2\eta_j \kappa_j}{v_j^2}.$$ [15]

Hence Zipf's law holds if $\eta_j \kappa_j / v_j^2$ is small for *at least one* type j.

Note that in this model the cross-sectional distributions are distinct across types. Hence if the firm type is imperfectly observed to the econometrician, the probability that a firm is of a particular type conditional on its size will generally depend on the size. The empirical fact that small firms tend to grow and exit faster need not be a violation of Gibrat's law but simply because firm types are imperfectly observed: a firm type that grows and exits fast may just happen to have a small average size.

3.3.5 Coexistence of Zipf and non-Zipf distributions

So far I have shown that under the assumptions of Theorem 3.6, Zipf's law for the size distribution is possible and robust. Is this theory consistent with the fact that empirically Zipf and non-Zipf distributions coexist? For example, while Zipf's law empirically holds for cities and firms, the Pareto exponent for household income is around 1.5–3 (Reed, 2003; Toda, 2012) and 4 for consumption (Toda and Walsh, 2015; Toda, 2016).

My theory can explain why Zipf's law holds for some size distributions but not for others. As an example, consider the simple model of city size distribution in Section 3.1. Instead of assuming that households are infinitely lived, suppose that they are born and die at constant Poisson rate $\delta > 0$. Assume that newborn households are endowed with 1 unit of human capital, but the human capital evolves according to a geometric Brownian motion with growth rate $\mu < \delta$ and volatility $\sigma > 0$ over the life cycle. Letting H be the aggregate stock of human capital in steady state, by accounting we have

$$0 = \frac{\mathrm{d}H}{\mathrm{d}t} = (\mu - \delta)H + \delta N \iff H = \frac{\delta}{\delta - \mu} N > 0.$$

Suppose that a household with human capital h supplies h units of labor services inelastically. Since aggregate human capital H is bounded, by the same argument as in Section 3.1, assuming that migration occurs independent of household income, the cross-sectional city size distribution obeys Zipf's law as $\eta \to 0$. Since the household human capital also satisfies a geometric Brownian

[15] Note that tails are fatter the smaller the Pareto exponent is, so the mixture of several distributions with Pareto upper tails has a Pareto tail with exponent equal to the minimum among its mixture components.

motion, the cross-sectional household income and consumption distributions will be double Pareto. However, the upper tail exponent $\alpha > 0$ will satisfy

$$\frac{\sigma^2}{2}\alpha^2 + \left(\mu - \frac{\sigma^2}{2}\right)\alpha - \delta = 0, \tag{3.12}$$

so α is solely determined by household characteristics (μ, σ, δ) and need not be close to 1.

As a numerical illustration, Deaton and Paxson (1994) report that within cohorts, the cross-sectional variance of household log consumption increases linearly over time at a rate 0.0069 per annum. Toda and Walsh (2015) find that the entire cross-sectional distribution of household consumption has a Pareto exponent around 3–4. Hence setting $\mu = 0$ (cohort effects are controlled), $\sigma^2 = 0.0069$, and $\alpha = 3, 4$ in (3.12), the implied Poisson rate is $\delta = 0.0207, 0.0414$ (mean lifespan $1/\delta = 48.3, 24.1$), which is reasonable since households are economically active for about 30–40 years.

4 A model of firm size distribution

Because Theorem 3.6 is an asymptotic result, the Pareto exponent need not be close to 1 for particular models or parameter configurations. To address the quantitative validity of my theory, in this section I construct a model of entrepreneurship and firm size distribution. The model builds on the continuous-time version of Angeletos (2007).

4.1 Environment

Consider an economy populated by two types of agents, household-workers and entrepreneur-CEOs. There are a continuum of both types, and entrepreneurs and workers have mass 1 and N, respectively. There is a single consumption good produced by the firms operated by the entrepreneurs, which can also be used as capital.

Households are infinitely lived and supply 1 unit of labor inelastically in a perfectly competitive labor market. They are infinitely risk averse, so they only borrow or lend at the market risk-free rate up to the natural borrowing limit and make consumption-saving decisions optimally.

Entrepreneurs die (go bankrupt) and are born at Poisson rate $\eta > 0$ (Yaari (1965)-Blanchard (1985) perpetual youth model). When an entrepreneur dies, his capital is wiped out and his firm disappears. Each entrepreneur is born with one "idea". Upon birth, she converts her "idea" to physical capital one-for-one[16] and starts to operate a constant-returns-to-scale technology with idiosyncratic investment risk. Entrepreneurs use their own physical capital and hire labor in a competitive market to carry out production. Markets are incomplete, so entrepreneurs may only invest in their own firms but can borrow or lend at the market risk-free rate.

A stationary equilibrium is defined by a wage ω, risk-free rate r, aggregate capital stock K, households' risk-free asset position X, households' consumption choice, entrepreneur's consumption-portfolio-saving-hiring choice, and size

[16]Since capital is wiped out when an entrepreneur goes bankrupt and entrepreneurs are born with capital, it is more appropriate to interpret capital as organization capital.

distributions of firms' capital and employment such that (i) households make optimal consumption-saving choice and entrepreneurs make optimal consumption-portfolio-saving-hiring choice, (ii) markets for labor and risk-free asset clear, and (iii) all aggregate variables and size distributions are invariant over time.

4.2 Individual decisions

Workers The utility function of a worker is

$$U_t = \int_0^\infty \mathrm{e}^{-\rho s} \frac{c_{t+s}^{1-1/\varepsilon}}{1-1/\varepsilon} \,\mathrm{d}s,$$

where $\rho > 0$ is the discount rate and $\varepsilon > 0$ is the elasticity of intertemporal substitution. Since workers hold only the risk-free asset, the budget constraint is

$$\mathrm{d}x_t = (rx_t + \omega_t - c_t)\,\mathrm{d}t,$$

where x_t is the financial wealth (which is entirely invested in the risk-free asset) and $\omega_t = \omega$ is the (constant) wage. Letting

$$h_t = \int_0^\infty \mathrm{e}^{-rs}\omega_{t+s}\,\mathrm{d}s = \frac{\omega}{r}$$

be the human wealth (present discounted value of future wages) and $w_t = x_t + h_t$ be the effective total wealth, we have

$$\mathrm{d}w_t = (rw_t - c_t)\,\mathrm{d}t. \tag{4.1}$$

The problem thus reduces to a standard Merton (1969, 1971)-type optimal consumption-saving problem. A solution exists if and only if $\rho\varepsilon + (1-\varepsilon)r > 0$, in which case the optimal consumption rule is

$$c = (\rho\varepsilon + (1-\varepsilon)r)w = (\rho\varepsilon + (1-\varepsilon)r)(x + \omega/r). \tag{4.2}$$

Entrepreneurs Entrepreneurs have Epstein-Zin preferences with discount rate ρ, relative risk aversion γ, and elasticity of intertemporal substitution ε.

Let k_t be the physical capital, b_t be the corporate bond holdings, and $x_t = k_t + b_t$ be the financial wealth (net worth) of a typical entrepreneur. The budget constraint is

$$\mathrm{d}x_t = (F(k_t, n_t) - \omega n_t + (r+\eta)b_t - c_t)\,\mathrm{d}t + \sigma k_t\,\mathrm{d}B_t, \tag{4.3}$$

where n_t is the labor input, c_t is consumption, F is a constant-returns-to-scale production function net of capital depreciation, $\sigma > 0$ is the volatility of the idiosyncratic shock, and B_t is a standard Brownian motion that is independent across entrepreneurs. Note that the effective risk-free rate faced by entrepreneurs is not r, but $r + \eta$, reflecting the fact that they go bankrupt at Poisson rate $\eta > 0$ and hence are charged an insurance premium $\eta > 0$ on their borrowing (they get annuities at the same rate if they are lending). η can also be interpreted as the spread of corporate bonds over the risk-free asset.

Because labor appears only in the budget constraint and can be chosen freely, letting $f(k) = F(k,1)$, as in (3.2) the capital-labor ratio $y = k_t/n_t$ satisfies

$\omega = f(y) - yf'(y)$. The labor demand is $n_t = k_t/y$, and as in (3.3) the profit rate per unit of capital is $\mu = f'(y)$. Substituting into the budget constraint (4.3), we obtain

$$\mathrm{d}x_t = (r_e + (\mu - r_e)\theta - m)x_t \,\mathrm{d}t + \sigma\theta x_t \,\mathrm{d}B_t, \tag{4.4}$$

where $r_e = r+\eta$ is the effective risk-free rate faced by entrepreneurs, $\theta = k_t/x_t$ is the leverage (the fraction of wealth invested in the physical capital, so $k_t = \theta x_t$ and $b_t = (1-\theta)x_t$), and $m = c_t/x_t$ is the propensity to consume out of wealth. Therefore this problem also becomes a Merton (1971)-type optimal consumption-saving-portfolio problem. According to Svensson (1989), the solution for the case with Epstein-Zin utility is

$$\theta = \frac{\mu - r_e}{\gamma\sigma^2}, \tag{4.5a}$$

$$\begin{aligned} m &= (\rho+\eta)\varepsilon + (1-\varepsilon)\left(r_e + (\mu - r_e)\theta - \frac{1}{2}\gamma\sigma^2\theta^2\right) \\ &= (\rho+\eta)\varepsilon + (1-\varepsilon)\left(r_e + \frac{(\mu - r_e)^2}{2\gamma\sigma^2}\right), \end{aligned} \tag{4.5b}$$

provided that these θ, m are positive. Substituting these rules into the budget constraint (4.4), we obtain

$$\mathrm{d}x_t = gx_t \,\mathrm{d}t + vx_t \,\mathrm{d}B_t, \tag{4.6}$$

where the drift g and volatility v are given by

$$g = (r-\rho)\varepsilon + (1+\varepsilon)\frac{(\mu - r_e)^2}{2\gamma\sigma^2}, \tag{4.7a}$$

$$v = \sigma\theta = \frac{\mu - r_e}{\gamma\sigma}. \tag{4.7b}$$

4.3 Equilibrium

Next I characterize the equilibrium. So far I have implicitly assumed that the discount rate ρ and EIS ε are common across agent types, but this is not necessary. Hence let ρ_W, ε_W be the parameter values for the workers, and let the symbols without subscripts be those of the entrepreneurs. Throughout the rest of the paper I assume that the production function $f(x) = F(x,1)$ satisfies the usual conditions $f(0) = 0$, $f' > 0$, $f'' < 0$, $f'(0) = \infty$, and $f'(\infty) \le 0$.

Define $0 < y_0 < y_1 < y_2$ by

$$f'(y_0) = \rho_W + \eta + \gamma\sigma^2, \quad f'(y_1) = \rho_W + \eta, \quad f'(y_2) = \eta, \tag{4.8}$$

which uniquely exist by the Inada condition.

Depending on the discount rate of workers, in equilibrium workers may consume a positive amount or zero. The following theorem characterizes the equilibrium.

Theorem 4.1. *A stationary equilibrium exists if and only if*

$$\left(1 - \frac{1}{y_2 N}\right)\eta > -\rho\varepsilon. \tag{4.9}$$

The equilibrium falls into exactly one of the following two categories.

1. *If*
$$\left(1-\frac{1}{y_1 N}\right)\eta > (\rho_W-\rho)\varepsilon, \tag{4.10}$$
then the equilibrium is unique, the risk-free rate equals the discount rate of workers: $r=\rho_W$*, and the capital-labor ratio* $y=K/N$ *is the unique solution in* $(0,y_1)$ *to*
$$\left(1-\frac{1}{yN}\right)\eta = (r-\rho)\varepsilon + (1+\varepsilon)\frac{(f'(y)-r-\eta)^2}{2\gamma\sigma^2}. \tag{4.11}$$
In equilibrium workers consume a positive amount.

2. *If* (4.10) *fails, then the equilibrium capital-labor ratio* y *and risk-free rate* r *satisfy* (4.11) *and*
$$\frac{r}{r+f(y)/y-f'(y)} = \frac{f'(y)-r-\eta}{\gamma\sigma^2}. \tag{4.12}$$
In equilibrium workers consume zero. Furthermore, $y_0<y<y_2$ *and* $0<r<\rho_W$*.*

In either case, the net worth x_t *of individual entrepreneurs evolves according to the geometric Brownian motion*
$$\mathrm{d}x_t = \eta(1-\kappa)x_t\,\mathrm{d}t + vx_t\,\mathrm{d}B_t, \tag{4.13}$$
where $\kappa=\frac{1}{K}=\frac{1}{yN}$ *is the ratio between the initial and the steady state capital and* $v=\frac{f'(y)-r-\eta}{\gamma\sigma}>0$ *is volatility.*

It immediately follows that an equilibrium exists if η is sufficiently small.

Corollary 4.2. *An equilibrium exists if* η *is sufficiently small. If* $\rho_W<\rho$ *(*$\rho_W\ge\rho$*), then in equilibrium workers consume a positive (zero) amount.*

Proof. Since $0<(f')^{-1}(\rho_W+\eta)\le y_1<y_2$, it follows that y_1,y_2 are bounded away from 0 as $\eta\to 0$. Therefore the left-hand sides of (4.9) and (4.10) converge to 0 as $\eta\to 0$. Since $\rho\varepsilon>0$, for small enough $\eta>0$ (4.9) holds, so by Theorem 4.1 a stationary equilibrium exists. If $\rho_W<\rho$, then for small enough $\eta>0$ (4.10) holds, so in equilibrium workers consume a positive amount. Otherwise ($\rho_W\ge\rho$), workers consume zero. □

Since this model satisfies the assumptions of Theorem 3.6, the upper tail Pareto exponent ζ satisfies the bound (3.8). However, since κ, v are endogenous, it is not immediately clear whether Zipf's law holds as $\eta\to 0$. Nevertheless, we can show that the technical condition (3.11) holds, and so does Zipf's law.

Theorem 4.3 (Zipf's law)**.** *As* $\eta\to 0$*, we obtain Zipf's law* $\zeta\to 1$*.*

Theorem 4.3 is an asymptotic result, and hence for any given parameters the upper tail Pareto exponent need not be close to 1, although the bound (3.8) is always true. Whether ζ is close to 1 or not is therefore a quantitative question, which I address in the numerical example below.

4.4 Numerical example

In this section I compute a numerical example of the model of firm size distribution. For the production function, I assume the Cobb-Douglas form $F(k, n) = Ak^\alpha n^{1-\alpha} - \delta k$, where A is a constant (normalized to $A = 1$), α is the capital share, and δ is the capital depreciation rate.

4.4.1 Calibration

The model is completely specified by the parameters $(\rho_W, \rho, \gamma, \varepsilon, \alpha, \delta, \sigma, \eta, N)$.[17] I calibrate the model at the annual frequency. Following Angeletos (2007), I set $\rho = 0.04$, $\varepsilon = 1$, $\alpha = 0.36$, $\delta = 0.08$, and $\sigma = 0.2$, which are all relatively standard values. Since in steady state the risk-free rate r equals the discount rate of the workers ρ_W when they have positive consumption, I set $\rho_W = 0.01$ so that the risk-free rate is 1%, which is about the historical value in U.S. For N, which is the average number of workers per firm, according to 2011 *U.S. Census Small Business Administration* (SBA) data,[18] 5,684,424 firms employed 113,425,965 workers, which implies an average of 19.95 employees per firm. Therefore I set $N = 20$.

The parameters that may be controversial are the relative risk aversion γ and the bankruptcy rate η. Based on SBA data for 1988–2006, Luttmer (2010) reports that the average exit rate is 10.4% per annum for firms with fewer than 20 employees and 2.5% for firms with 500 or more employees. If we take the model literally, η is also the spread of (defaultable) corporate bond over the risk-free asset. Based on a monthly 1990–2008 sample of 899 publicly traded non-financial firms (mostly large firms) covered by the *Center for Research in Security Prices* (CRSP), Gilchrist et al. (2009) find that the mean spread of corporate bonds is 192 basis points (1.92%), which is comparable to the exit rate of large firms. Since I am interested in the upper tail behavior (large firms), I set $\eta = 0.025$ or 2.5% spread, which implies an average lifespan of $1/\eta = 40$ years. However, since by Theorem 4.3 Zipf's law obtains when η is small, it is interesting to know the Pareto exponent under larger values of η, for which the bound (3.8) may not be so informative. Therefore I also consider the cases $\eta = 0.05$ (5% spread or 20 years lifespan) and $\eta = 0.1$ (10% spread or 10 years lifespan). One can think of the case $\eta = 0.025$ as a CEO operating a blue-chip firm, and the case $\eta = 0.05, 0.1$ as a young entrepreneur operating a start-up company.

For the relative risk aversion, it is reasonable to assume that the rich CEOs of large firms are not so risk averse, so I set $\gamma = 1$.[19] As a robustness check, I also consider the cases $\gamma = 0.5, 2$.

4.4.2 Results

By Theorem 4.1, computing the equilibrium with positive consumption reduces to solving a single nonlinear equation (4.11). If the existence condition (4.10) fails, we need to look for an equilibrium with zero consumption, in which case

[17] Note that the elasticity of intertemporal substitution for the workers, ε_W, is irrelevant for the steady state, so there is no need to specify it.

[18] `https://www.sba.gov/advocacy/firm-size-data`

[19] Aoki and Nirei (2016) also assume $\gamma = 1$ (log utility), but the reason is for tractability for solving the entire transitional dynamics.

we need to solve a system of two nonlinear equations (4.11) and (4.12). Table 1 shows the results, which are all equilibria with positive consumption. The private equity premium, leverage (fraction of own physical capital to entrepreneur net worth), and volatility are all reasonable numbers, roughly in line with U.S. stock returns. In each case, the upper tail Pareto exponent ζ is close to 1, in agreement with Zipf's law.

Table 1: Parameters and endogenous variables in steady state.

Quantity	Symbol	Values				
Risk aversion	γ	1	0.5	2	1	1
Bankruptcy rate (%)	η	2.5	2.5	2.5	5	10
Capital-labor ratio	y	3.49	4.01	2.93	2.58	1.65
Wage	ω	1.004	1.055	0.942	0.900	0.767
Private premium (%)	$\mu - r_e$	4.68	3.31	6.61	5.62	7.13
Equity premium (%)	$\mu - r$	7.18	5.81	9.11	10.62	17.13
Leverage	θ	1.17	1.65	0.83	1.41	1.78
Volatility (%)	v	23.4	33.1	16.5	28.1	35.6
Pareto exponent	ζ	1.007	1.004	1.011	1.011	1.019

Note: the table shows the values of endogenous variables in steady state. The capital-labor ratio is $y = K/N$, where K is the aggregate capital. The private premium is the expected return on capital in excess of the effective risk-free rate faced by entrepreneurs, $\mu - r_e$, where $\mu = f'(y)$ and $r_e = r + \eta = \rho_W + \eta$ is the effective risk-free rate (true risk-free rate plus spread). The equity premium is the expected return on capital in excess of the risk-free rate $r = \rho_W$ conditional on survival. The leverage $\theta = \frac{\mu - r_e}{\gamma\sigma^2}$ is the ratio between entrepreneur's own physical capital to net worth. $v = \sigma\theta$ is the volatility of entrepreneur's net worth (which is also the market capitalization of the firm). ζ is the upper tail Pareto exponent computed as in Proposition 3.1.

As we make the environment riskier (larger γ or η), the private equity premium goes up, the capital-labor ratio goes down, which also suppresses the wage. However, the mechanism is very different depending on whether we increase risk aversion γ or the bankruptcy rate η. When γ increases, the entrepreneurs become less willing to invest capital, so they leverage less (portfolio effect). Since there is less investment in the high return capital, the aggregate capital goes down. On the other hand, when η increases, aggregate capital goes down just because there is more bankruptcy and hence destruction of capital (resource effect). Since capital is more scarce, the risk premium goes up, and entrepreneurs leverage more to take advantage.

It is not surprising that the upper tail Pareto exponent ζ is close to 1 regardless of the parameter specification. The reason is that, according to (3.8), we always have the bound

$$1 < \zeta < 1 + \frac{2\eta\kappa}{v^2}.$$

As a rough estimate, the bankruptcy rate η has order of magnitude about 10^{-1} or 10^{-2} and the volatility v has order of magnitude about 10^{-1}. Hence the upper bound of ζ is $1 + \frac{2\eta\kappa}{v^2} \approx 1 + \kappa$. Since κ is the ratio of the initial capital of new firms to that of the average firm, it is reasonable to expect that κ is quite small. Therefore ζ must be close to 1.

4.4.3 Sensitivity analysis

How robust is Zipf's law? In this section, I conduct two robustness checks.

First, I fix the parameter values $(\rho_W, \rho, \gamma, \varepsilon, \alpha, \delta, \sigma, \eta, N)$ at the baseline specification and vary one parameter at a time up to ten-fold increase or decrease. (For the capital share α, I consider all values in $(0, 1)$.) For example, since at the baseline we have $\gamma = 1$, I consider $\gamma \in [0.1, 10]$. Figure 3 shows the results. We can see that in all cases the Pareto exponent ζ is slightly above 1 regardless of the parameter values (which can be quite extreme), and in most cases below 1.1, consistent with Zipf's law.

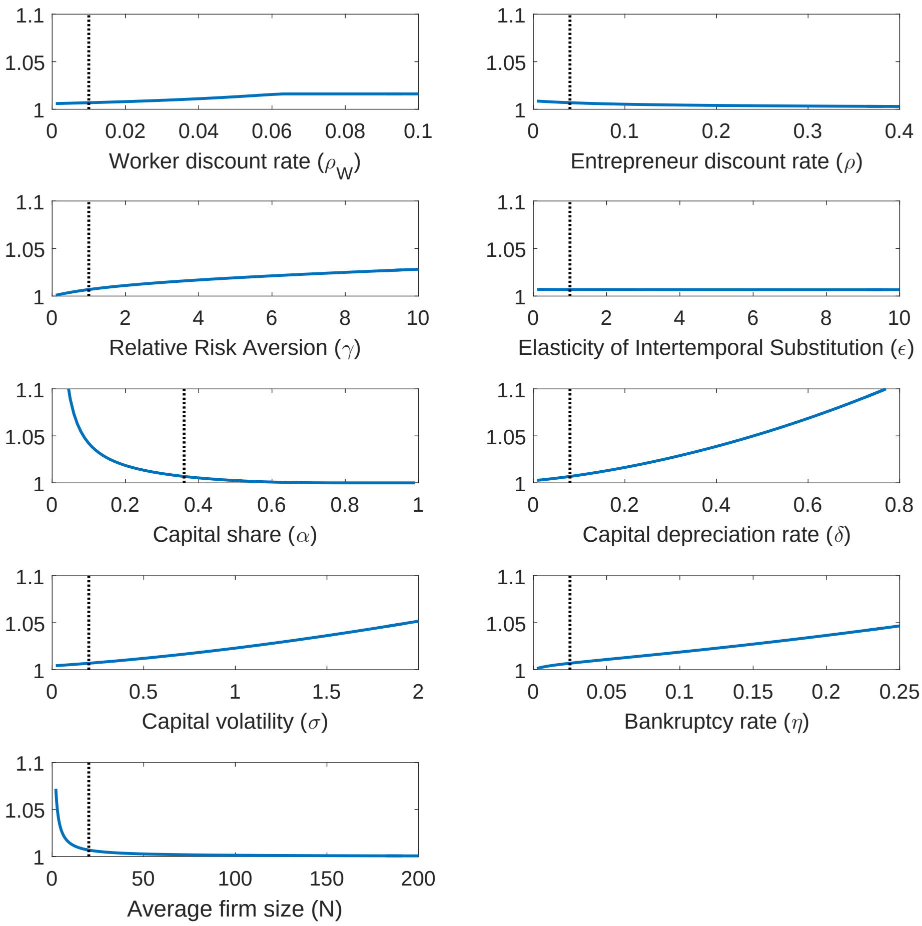

Figure 3: Sensitivity of Pareto exponent ζ with respect to model parameters.

Note: The figure plots the upper tail Pareto exponent ζ computed as in Proposition 3.1. The dotted vertical lines indicate the baseline parameter values.

In the second robustness check, I generate 10,000 random parameter configurations and compute the Pareto exponent for each simulation. For this experiment, I consider up to five-fold change in the parameters, so in each simulation a parameter is 5^U times the baseline value, where U is uniformly drawn from $[-1, 1]$ independently across all parameters and simulations. (For the capital share α, it is uniformly drawn from $[0.1\alpha, 1.9\alpha]$.)

Figure 4 shows the histogram of the Pareto exponent ζ in the range $[1, 1.1]$. The mean, median, and the 95% percentile are 1.0312, 1.0089, and 1.1313, respectively. Again Zipf's law is quite robust.

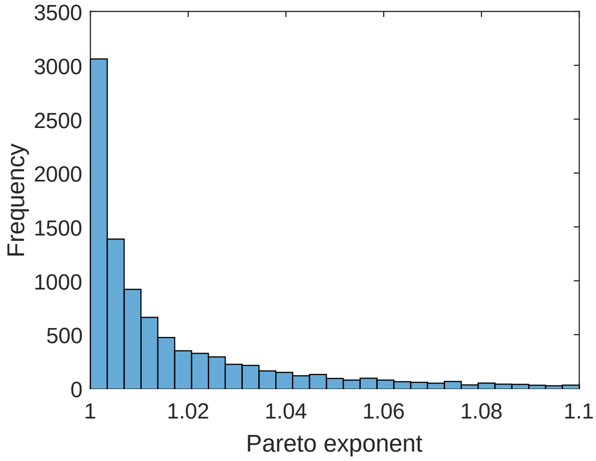

Figure 4: Histogram of Pareto exponent ζ with random parameter configurations.

A Fokker-Planck equation

In this appendix, I derive the *Fokker-Planck equation*, also known as the *Kolmogorov forward equation*, which is useful in characterizing the cross-sectional distribution in general settings. A good introduction is Gabaix (2009).

A.1 Derivation of Fokker-Planck equation

Proposition A.1. *Consider the diffusion*

$$\mathrm{d}X_t = g(t, X_t)\,\mathrm{d}t + v(t, X_t)\,\mathrm{d}B_t, \tag{A.1}$$

where B_t is standard Brownian motion. Let $p(x,t)$ be the density of X_t at time t. Then

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(gp) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(v^2 p), \tag{A.2}$$

which is known as the Fokker-Planck (Kolmogorov forward) equation.

Proof. The proof is based on the following (unintuitive) calculation.

First, fix $t_1 < t_2$ and let $F(t,x)$ be a smooth function such that $F(t_1,x) = F(t_2,x) = 0$ and $F(t,x), F_x(t,x) \to 0$ as $x \to \pm\infty$. There are plenty of such functions, for example

$$F(t,x) = (t-t_1)(t-t_2)f(x)$$

with $f(x) > 0$ and $f(x), f'(x) \to 0$ as $x \to \pm\infty$.

By Itô's formula, we get

$$\begin{aligned}
\mathrm{d}F(t,X(t)) &= F_t\,\mathrm{d}t + F_x\,\mathrm{d}X_t + \frac{1}{2}F_{xx}(\mathrm{d}X_t)^2 \\
&= F_t\,\mathrm{d}t + F_x(g\,\mathrm{d}t + v\,\mathrm{d}B) + \frac{1}{2}F_{xx}v^2\,\mathrm{d}t \\
&= \left(F_t + F_x g + \frac{1}{2}F_{xx}v^2\right)\mathrm{d}t + F_x v\,\mathrm{d}B.
\end{aligned}$$

Taking expectations and using the martingale property of the Brownian motion, we get

$$\begin{aligned}
\mathrm{E}[\mathrm{d}F(t,X(t))] &= \mathrm{E}\left[\left(F_t + F_x g + \frac{1}{2}F_{xx}v^2\right)\mathrm{d}t\right] \\
&= \int_{-\infty}^{\infty}\left(F_t + F_x g + \frac{1}{2}F_{xx}v^2\right)p(x,t)\,\mathrm{d}t\,\mathrm{d}x.
\end{aligned}$$

Integrating from $t = t_1$ to t_2 and using $F(t_1,x) = F(t_2,x) = 0$, we get

$$\begin{aligned}
0 &= \mathrm{E}[F(t_2,X(t_2)) - F(t_1,X(t_1))] \\
&= \int_{-\infty}^{\infty}\int_{t_1}^{t_2}\left(F_t + F_x g + \frac{1}{2}F_{xx}v^2\right)p(x,t)\,\mathrm{d}t\,\mathrm{d}x =: I_1 + I_2 + I_3.
\end{aligned}$$

Integrating by parts, we get

$$\begin{aligned}
I_1 &= \int_{-\infty}^{\infty}\int_{t_1}^{t_2}\frac{\partial F}{\partial t}p(x,t)\,\mathrm{d}t\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty}\left(F(t_2,x) - F(t_1,x) - \int_{t_1}^{t_2}F\frac{\partial}{\partial t}p(x,t)\,\mathrm{d}t\right)\mathrm{d}x \\
&= -\int_{t_1}^{t_2}\int_{-\infty}^{\infty}F\frac{\partial}{\partial t}p(x,t)\,\mathrm{d}x\,\mathrm{d}t,
\end{aligned}$$

where I have used $F(t_1,x) = F(t_2,x) = 0$ and Fubini's theorem. By similar calculations, we get

$$\begin{aligned}
I_2 &= -\int_{t_1}^{t_2}\int_{-\infty}^{\infty}F\frac{\partial}{\partial x}(gp(x,t))\,\mathrm{d}x\,\mathrm{d}t, \\
I_3 &= \int_{t_1}^{t_2}\int_{-\infty}^{\infty}F\frac{\partial^2}{\partial x^2}\left(\frac{1}{2}v^2 p(x,t)\right)\mathrm{d}x\,\mathrm{d}t.
\end{aligned}$$

Putting all the pieces together, we get

$$0 = I_1 + I_2 + I_3 = \int_{t_1}^{t_2}\int_{-\infty}^{\infty}F\left[-\frac{\partial p}{\partial t} - \frac{\partial}{\partial x}(gp) + \frac{\partial^2}{\partial x^2}\left(\frac{1}{2}v^2 p\right)\right]\mathrm{d}x\,\mathrm{d}t.$$

Since F is (nearly) arbitrary, the integrand must be identically zero.[20] Therefore we obtain the (parabolic) partial differential equation (PDE) (A.2). □

[20] To see this more rigorously, set

$$F = (t-t_1)(t-t_2)\left[-\frac{\partial p}{\partial t} - \frac{\partial}{\partial x}(gp) + \frac{\partial^2}{\partial x^2}\left(\frac{1}{2}v^2 p\right)\right].$$

The Fokker-Planck equation (A.2) holds if the diffusion (A.1) holds at all times. However, we can consider situations in which the process is occasionally reset. For example, if X_t in (A.1) describe individual wealth, since the individual will die eventually, we need to specify what happens when an individual dies. If there is influx $j_+(x,t)$ and outflux $j_-(x,t)$ per unit of time at location x at time t, then the Fokker-Planck equation (A.2) must be modified as

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(gp) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(v^2 p) + j_+ - j_-.$$

For example, if the units die at constant probability η per unit of time (Poisson rate η) and is reborn at location x_0, then the FPE becomes

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(gp) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(v^2 p) + \eta\delta(x - x_0) - \eta p,$$

where $\delta(x - x_0)$ is the Dirac delta function located at x_0.

A.2 Stationary density

If the diffusion has time-independent drift $g(x)$ and variance $v(x)$ and admits a stationary distribution $p(x)$, then we get

$$0 = -\frac{\mathrm{d}}{\mathrm{d}x}(gp) + \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2}(v^2 p).$$

Integrating with respect to x and using the boundary condition $p(x), p'(x) \to 0$ as $x \to \pm\infty$, we get

$$0 = -g(x)p(x) + \frac{1}{2}(v(x)^2 p(x))'.$$

Letting $q(x) = v(x)^2 p(x)$ and solving the ODE, we get

$$\begin{aligned} q' = \frac{2g}{v^2} q \iff & \frac{q'}{q} = \frac{2g}{v^2} \\ \iff & \log q(x) = \int \frac{q'(x)}{q(x)}\,\mathrm{d}x = \int \frac{2g(x)}{v(x)^2}\,\mathrm{d}x \\ \iff & q(x) = \exp\left(\int \frac{2g(x)}{v(x)^2}\,\mathrm{d}x\right). \end{aligned}$$

Therefore the stationary density is

$$p(x) = \frac{q(x)}{v(x)^2} = \frac{1}{v(x)^2}\exp\left(\int \frac{2g(x)}{v(x)^2}\,\mathrm{d}x\right), \tag{A.3}$$

where the constant of integration is determined by the condition $\int_{-\infty}^{\infty} p(x)\,\mathrm{d}x = 1$ since $p(x)$ is a density.

If there is a constant probability of death η, the stationary density is the solution of the second-order ODE

$$0 = -\frac{\mathrm{d}}{\mathrm{d}x}(gp) + \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2}(v^2 p) - \eta p,$$

which holds at every point except x_0.

A.2.1 Geometric Brownian motion with minimum size

As examples, consider the geometric Brownian motion with minimum size $x_{\min}$ or constant Poisson rate η of birth/death with reset size x_0. In the former case, setting $g(x) = gx$ (with $g < 0$) and $v(x) = vx$ in (A.3), the stationary density is

$$p(x) = \frac{1}{(vx)^2} \exp\left(\int \frac{2gx}{(vx)^2} \,\mathrm{d}x\right) = Cx^{\frac{2g}{v^2}-2}$$

for some constant $C > 0$. Since the minimum size is $x_{\min}$ and the probability must add up to 1, it follows that

$$1 = C\int_{x_{\min}}^{\infty} x^{\frac{2g}{v^2}-2} = \frac{C}{1-\frac{2g}{v^2}} x_{\min}^{-1+\frac{2g}{v^2}}.$$

Therefore

$$p(x) = \zeta x_{\min}^{\zeta} x^{-\zeta-1}$$

for $\zeta = 1 - 2g/v^2$, which is the probability density function of the Pareto distribution (2.2) with exponent $\zeta > 1$.

A.2.2 Geometric Brownian motion with Poisson birth/death

Next, consider the geometric Brownian motion with birth/death at Poisson rate $\eta > 0$ and reset size x_0. In this case, it is easier to solve in logs. Using Itô's lemma, $Y_t = \log X_t$ obeys the Brownian motion

$$\mathrm{d}Y_t = \left(g - \frac{1}{2}v^2\right)\mathrm{d}t + v\,\mathrm{d}B_t.$$

The Fokker-Planck equation in the steady state is

$$0 = -\left(g - \frac{1}{2}v^2\right)p(y)' + \frac{1}{2}v^2 p(y)'' - \eta p(y)$$

except at $y_0 := \log x_0$, where I used the fact that g, v are constant. Since this is a linear second-order ODE with constant coefficients, the general solution is

$$p(y) = C_1 \mathrm{e}^{-\lambda_1 y} + C_2 \mathrm{e}^{-\lambda_2 y},$$

where $\lambda_1 > 0 > \lambda_2$ are solutions to the quadratic equation

$$\frac{1}{2}v^2\xi^2 + \left(g - \frac{1}{2}v^2\right)\xi - \eta = 0,$$

which is (2.6). Since the PDF must be continuous and integrate to 1, letting $\alpha = \lambda_1 > 0$ and $\beta = -\lambda_2 > 0$, it follows that

$$p(y) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta}\mathrm{e}^{-\alpha|y-y_0|}, & (y \geq y_0) \\ \frac{\alpha\beta}{\alpha+\beta}\mathrm{e}^{-\beta|y-y_0|}, & (y \leq y_0) \end{cases}$$

which is the asymmetric Laplace distribution with mode y_0 and exponents α, β. Taking the exponential, we obtain the double Pareto distribution (2.5).

B Proofs

Proof of Proposition 3.1. Since the equation of motion (3.7) is a special case of the mechanistic model (2.1), it suffices to show the bound (3.8). Let

$$q(\zeta)=\frac{v^2}{2}\zeta^2+\left(\eta(1-\kappa)-\frac{v^2}{2}\right)\zeta-\eta$$

be the quadratic function that determines the Pareto exponent as in (2.6). Since

$$\begin{aligned}
q(1)&=\frac{v^2}{2}+\eta(1-\kappa)-\frac{v^2}{2}-\eta=-\eta\kappa<0,\\
q\left(1+\frac{2\eta\kappa}{v^2}\right)&=\frac{v^2}{2}\left(1+\frac{2\eta\kappa}{v^2}\right)^2+\left(\eta(1-\kappa)-\frac{v^2}{2}\right)\left(1+\frac{2\eta\kappa}{v^2}\right)-\eta\\
&=\frac{v^2}{2}\left(1+\frac{4\eta\kappa}{v^2}+\frac{4\eta^2\kappa^2}{v^4}\right)+\left(\eta-\eta\kappa-\frac{v^2}{2}\right)\left(1+\frac{2\eta\kappa}{v^2}\right)-\eta\\
&=\frac{2\eta^2\kappa}{v^2}>0,
\end{aligned}$$

the solution satisfies $1<\zeta<1+\frac{2\eta\kappa}{v^2}$, which is (3.8). □

Proof of Lemma 3.5. Suppose that a positive control variable y_l has a finite aggregate supply $0<e_l<\infty$. By lemma 3.3, the demand of a unit (of a particular type that solves a homogeneous problem) with size x_t is $y_{lt}=\alpha_l(\theta)x_t$. Since other types may also demand that variable, taking the cross-sectional expectation, by market clearing we have

$$\infty>e_l\ge \mathrm{E}[y_{lt}]=\mathrm{E}[\alpha_l(\theta)x_t]=\alpha_l(\theta)\,\mathrm{E}[x_t]>0.$$

Since $\alpha_l(\theta)>0$, we have $0<\mathrm{E}[x_t]<\infty$.

Substituting the optimal control $y_t=\alpha(\theta)x_t$ into the equation of motion (3.9), we obtain

$$\begin{aligned}
\mathrm{d}x_t&=g(x_t,\alpha(\theta)x_t;\theta)\,\mathrm{d}t+v(x_t,\alpha(\theta)x_t;\theta)\,\mathrm{d}B_t\\
&=g(1,\alpha(\theta);\theta)x_t\,\mathrm{d}t+v(1,\alpha(\theta);\theta)x_t\,\mathrm{d}B_t\\
&=:g(\theta)x_t\,\mathrm{d}t+v(\theta)x_t\,\mathrm{d}B_t,
\end{aligned}$$

where I have used the homogeneity of g,v. □

Prof of Theorem 3.6. By Lemma 3.5, the size of individual units evolves according to the geometric Brownian motion (3.10). Let $X=\mathrm{E}[x_t]$ be the cross-sectional average, which is positive and finite by Lemma 3.5 and constant over time by stationarity. Since individual units grows at rate $g(\theta)$, disappear at rate η, and newborn units have average size $\kappa(\eta)X$, it follows that

$$0=\frac{\mathrm{d}X}{\mathrm{d}t}=(g(\theta)-\eta)X+\eta\kappa(\eta)X\iff g(\theta)=\eta(1-\kappa(\eta))$$

at $\theta=\theta(\eta)$. Substituting into (3.10), we obtain the same equation as (3.7). Hence by Proposition 3.1, the cross-sectional size distribution relative to initial size x_0 is double Pareto with an upper tail exponent ζ that satisfies (3.10). The upper tail exponent of the (unconditional) cross-sectional size distribution also satisfies (3.10) since the initial size distribution $F(\cdot;\theta,\eta)$ either does not affect the tail (if F is thin-tailed) or makes the tail even fatter (if F is fat-tailed with exponent smaller than ζ). □

Proof of Theorem 4.1. Since the proof is long and tedious, I break it down into several steps.

Step 1. If an equilibrium exists, the net worth x_t of individual entrepreneurs evolves according to (4.13).

Since (4.6) holds and $k_t = \theta x_t$, where θ is given by (4.5a), individual capital k_t also obeys the same geometric Brownian motion: $\mathrm{d}k = gk\,\mathrm{d}t + vk\,\mathrm{d}B_t$. To derive the dynamics of aggregate capital K_t (which is constant in steady state), consider what happens to individual capital during a short period of time Δt. If the entrepreneur survives (which occurs with probability $1-\eta\Delta t$), then the capital grows at rate g, so it becomes $(1+g\Delta t)k_t$. If the entrepreneur goes bankrupt (which occurs with probability $\eta\Delta t$), the capital is wiped out, and a new agent is born with 1 unit of capital. Hence by accounting we obtain

$$\begin{aligned}K+\Delta K &= \underbrace{(1-\eta\Delta t)(1+g\Delta t)K}_{\text{Aggregate capital of surviving agents}} + \underbrace{\eta\Delta t\times 1}_{\text{Aggregate capital of newborn agents}}\\ &= (1+(g-\eta)\Delta t)K+\eta\Delta t+\text{higher order terms.}\end{aligned}$$

Subtracting K from both sides and letting $\Delta t\to 0$, we obtain

$$\mathrm{d}K = (g-\eta)K\,\mathrm{d}t + \eta\,\mathrm{d}t.$$

In steady state, aggregate capital is constant, so it must be

$$(g-\eta)K+\eta = 0 \iff g = (1-\kappa)\eta, \tag{B.1}$$

where $\kappa = 1/K$ is the amount of initial capital relative to the steady state value. Substituting this g into (4.6), we obtain (4.13).

Step 2. If a stationary equilibrium exists, then $r>0$. The propensity to consume out of wealth, m in (4.5b)*, is positive. The volatility of entrepreneur's wealth is given by $v = \frac{f'(y)-r-\eta}{\gamma\sigma} > 0$.*

If $r\le 0$, then the present value of a worker's wage $\int_0^\infty \mathrm{e}^{-rt}\omega\,\mathrm{d}t$ is infinite, so the utility maximization problem does not have a solution. Therefore if an equilibrium exists, it must be $r>0$.

If an equilibrium exists, by (4.5a) the fraction of wealth invested in physical capital is

$$0<\theta = \frac{\mu-r_e}{\gamma\sigma^2} = \frac{f'(y)-r-\eta}{\gamma\sigma^2},$$

where $\mu = f'(y)$ and $r_e = r+\eta$. By (4.7b), we have $v = \frac{f'(y)-r-\eta}{\gamma\sigma} > 0$. To show that the propensity to consume is positive, note that by (4.4), (4.5a), and (4.13), we have

$$g = r_e + \frac{(\mu-r_e)^2}{\gamma\sigma^2} - m = (1-\kappa)\eta.$$

Since $r_e = r+\eta$, $\mu = f'(y)$, and $r,\kappa,\eta>0$, it follows that

$$m = r+\kappa\eta+\frac{(f'(y)-r-\eta)^2}{\gamma\sigma^2} > 0.$$

Step 3. If a stationary equilibrium exists, the capital-labor ratio $y = K/N$ and risk-free rate r satisfy (4.11)*, and* (4.9) *must hold.*

By (B.1) and (4.7a), we must have

$$g = (1-\kappa)\eta = (r-\rho)\varepsilon + (1+\varepsilon)\frac{(\mu - r_e)^2}{2\gamma\sigma^2}.$$

Substituting $\kappa = \frac{1}{yN}$, $\mu = f'(y)$, and $r_e = r + \eta$, the equilibrium capital-labor ratio y must satisfy (4.11). To show (4.9), let

$$\phi(y,r) = \left(1 - \frac{1}{yN}\right)\eta - (r-\rho)\varepsilon - (1+\varepsilon)\frac{(f'(y) - r - \eta)^2}{2\gamma\sigma^2} \tag{B.2}$$

be the left-hand side minus the right-hand side of (4.11). If an equilibrium exists, since capital investment must be positive we have

$$\theta > 0 \iff f'(y) - r - \eta > 0 \iff y < (f')^{-1}(r+\eta).$$

If $y > 0$ satisfies this inequality, then

$$\frac{\partial \phi}{\partial y}(y,r) = \frac{\eta}{y^2 N} - (1+\varepsilon)\frac{f'(y) - r - \eta}{\gamma\sigma^2} f''(y) > 0,$$

so ϕ is strictly increasing in y. Since $\phi(0,r) = -\infty$ and ϕ is continuous, there exists $y \in (0, (f')^{-1}(r+\eta))$ such that (4.11) holds if and only if

$$\psi(r) := \phi((f')^{-1}(r+\eta), r) = \left(1 - \frac{1}{(f')^{-1}(r+\eta)N}\right)\eta - (r-\rho)\varepsilon > 0.$$

Since $f'' < 0$ and $\eta, \varepsilon > 0$, clearly $\psi(r)$ is strictly decreasing. Since ψ is continuous, there exists $r > 0$ such that $\psi(r) > 0$ if and only if

$$\psi(0) > 0 \iff \left(1 - \frac{1}{(f')^{-1}(\eta)N}\right)\eta + \rho\varepsilon > 0,$$

which is exactly (4.9).

Step 4. A stationary equilibrium in which workers consume a positive amount exists if and only if (4.10) *holds. In this case the equilibrium is unique and $r = \rho_W$.*

In steady state with positive consumption of workers, their wealth must be a positive constant. Setting $\mathrm{d}w/\,\mathrm{d}t = 0$ in (4.1), we have $c = rw$. Comparing to the optimal consumption rule (4.2), we obtain

$$r = \rho_W \varepsilon_W + (1 - \varepsilon_W) r \iff r = \rho_W. \tag{B.3}$$

In this case $(f')^{-1}(r+\eta) = (f')^{-1}(\rho_W + \eta) = y_1$, so by Step 3 an equilibrium exists if and only if

$$0 < \psi(\rho_W) = \phi(y_1, \rho_W) = \left(1 - \frac{1}{y_1 N}\right)\eta - (\rho_W - \rho)\varepsilon,$$

which is exactly (4.10). Since ϕ is strictly increasing in y, the capital-labor ratio $y = K/N$ is unique.

So far we have shown that (4.9) is necessary for equilibrium existence, and that (4.10) is necessary and sufficient for the existence of a stationary equilibrium in which workers consume a positive amount. Therefore it remains to show that if (4.9) holds but (4.10) fails, then there exists a stationary equilibrium in which workers consume zero.

Step 5. A stationary equilibrium in which workers consume zero exists if and only if there exist $y, r > 0$ such that (4.11) *and* (4.12) *hold.*

By Steps 2 and 3, $r > 0$ and (4.11) are necessary for equilibrium. Letting $y = K/N$ be the capital-labor ratio and $\theta = \frac{f'(y)-r-\eta}{\gamma\sigma^2} > 0$ be entrepreneurs' portfolio share of capital investment, their aggregate net worth is $K/\theta = \frac{yN}{\theta}$. Since they invest fraction $1-\theta$ in the risk-free asset, its market capitalization is $B = \frac{1-\theta}{\theta} yN$. If workers consume zero in equilibrium, since they have zero net worth, all the wage must be used for interest payments on debt. Therefore the equilibrium condition is

$$rB = \omega N \iff r\frac{1-\theta}{\theta}y = f(y) - yf'(y),$$

which is equivalent to (4.12). Conversely, if $y, r > 0$ satisfy (4.11) and (4.12), aggregate capital is constant and workers have zero net worth and consumption, so it is an equilibrium.

Step 6. If (4.9) *holds but* (4.10) *does not, then a stationary equilibrium in which workers consume zero exists. The capital-labor ratio $y > 0$ and risk-free rate $r > 0$ satisfy* (4.11) *and* (4.12)*. Furthermore, $y_0 < y < y_2$ and $0 < r < \rho_W$.*

Since (4.9) holds but (4.10) fails, we have

$$\psi(0) = \left(1 - \frac{1}{y_2 N}\right)\eta + \rho\varepsilon > 0 \geq \left(1 - \frac{1}{y_1 N}\right)\eta + (\rho_W - \rho)\varepsilon = \psi(\rho_W).$$

Since ψ is strictly decreasing, there exists a unique $\bar{r} \in (0, \rho_W]$ such that $\psi(\bar{r}) = 0$. For any $0 < r \leq \bar{r}$, we have $\psi(r) \geq 0$, so by the above argument there exists a unique $y \in (0, (f')^{-1}(r+\eta)]$ such that (4.11) holds. Denote this y by $y(r)$. By the definition of ψ and $\bar{r}$, we have $0 = \psi(\bar{r}) = \phi((f')^{-1}(\bar{r}+\eta), \bar{r})$, so

$$y(\bar{r}) = (f')^{-1}(\bar{r}+\eta) \iff f'(y(\bar{r})) - \bar{r} - \eta = 0. \tag{B.4}$$

Let

$$\varphi(r) = \frac{r}{r + f(y)/y - f'(y)} - \frac{f'(y) - r - \eta}{\gamma\sigma^2}$$

be the left-hand side minus the right-hand side of (4.12), where $y = y(r) > 0$. Note that φ is well-defined for all $0 \leq r \leq \bar{r}$. To see this, since f is strictly concave and $f(0) = 0$, for $y > 0$ we have

$$f(0) - f(y) < f'(y)(0 - y) \iff f(y)/y - f'(y) > 0,$$

so the denominator of the first term of φ is always positive. Since $f'(y(r)) - r - \eta > 0$, we have

$$\varphi(0) = -\frac{f'(y(0)) - 0 - \eta}{\gamma\sigma^2} < 0.$$

Furthermore, by (B.4) we have

$$\varphi(\bar{r}) = \frac{\bar{r}}{\bar{r} + f(y)/y - f'(y)} > 0,$$

where $y = y(\bar{r})$. Since φ is continuous, by the intermediate value theorem there exists $r \in (0, \bar{r})$ such that $\varphi(r) = 0$. Since $y = y(r)$ and $r > 0$ satisfy (4.11) and (4.12), an equilibrium exists. In this equilibrium $0 < r < \bar{r} \leq \rho_W$.

It remains to show that $y_0 < y < y_2$. Since $0 < r < \rho_W$, $f(y)/y - f'(y) > 0$, and (4.12) holds, it follows that

$$\begin{aligned}
&\theta = \frac{f'(y) - r - \eta}{\gamma\sigma^2} > 0 \\
\Longrightarrow\ & y < (f')^{-1}(r + \eta) < (f')^{-1}(\eta) = y_2, \\
&\theta = \frac{f'(y) - r - \eta}{\gamma\sigma^2} = \frac{r}{r + f(y)/y - f'(y)} < 1 \\
\Longrightarrow\ & y > (f')^{-1}(r + \eta + \gamma\sigma^2) > (f')^{-1}(\rho_W + \eta + \gamma\sigma^2) = y_0. \qquad \square
\end{aligned}$$

Proof of Theorem 4.3. Since by Theorem 3.6 the bound (3.8) holds, in order to show $\zeta \to 1$ as $\eta \to 0$, it suffices to show that $\kappa > 0$ is bounded above and $v > 0$ is bounded away from 0.

Case 1: $\rho_W < \rho$. In this case (4.10) holds as $\eta \to 0$, so in equilibrium workers consume a positive amount and $r = \rho_W$.

Fix any $\underline{y} > 0$ such that

$$-(\rho_W - \rho)\varepsilon - (1 + \varepsilon)\frac{(f'(\underline{y}) - \rho_W)^2}{2\gamma\sigma^2} < 0,$$

which exists by the Inada condition $f'(0) = \infty$. Let $\phi(y, \rho_W; \eta)$ be $\phi(y, r)$ in (B.2) with $r = \rho_W$, given $\eta > 0$. Then we have

$$\lim_{\eta \to 0} \phi(\underline{y}, \rho_W; \eta) = -(\rho_W - \rho)\varepsilon - (1 + \varepsilon)\frac{(f'(\underline{y}) - \rho_W)^2}{2\gamma\sigma^2} < 0.$$

Since ϕ is strictly increasing in y and $\phi(y, \rho_W; \eta) = 0$ in equilibrium, it follows that for sufficiently small η we have $y > \underline{y}$. Therefore $\kappa = \frac{1}{yN} < \frac{1}{\underline{y}N}$ is bounded.

By Theorem 4.1, the equilibrium condition (4.11) is equivalent to

$$(1 - \kappa)\eta = (\rho_W - \rho)\varepsilon + \frac{1 + \varepsilon}{2}\gamma v^2 \iff v^2 = \frac{2\varepsilon(\rho - \rho_W) + 2(1 - \kappa)\eta}{\gamma(1 + \varepsilon)}.$$

Since κ is bounded and $\rho_W < \rho$, v is bounded away from 0 as $\eta \to 0$.

Case 2: $\rho_W \geq \rho$. In this case (4.10) fails as $\eta \to 0$, so in equilibrium workers consume zero and $r < \rho_W$.

By Theorem 4.1, we have $0 < (f')^{-1}(\rho_W + \eta + \gamma\sigma^2) = y_0 < y$. Therefore as $\eta \to 0$ we have

$$\kappa = \frac{1}{yN} < \frac{1}{(f')^{-1}(\rho_W + \eta + \gamma\sigma^2)N} \to \frac{1}{(f')^{-1}(\rho_W + \gamma\sigma^2)N} < \infty,$$

so κ is bounded. To show that v is bounded away from 0, using (4.7b), (4.11), and (4.12), we have

$$\left(1 - \frac{1}{yN}\right)\eta - (r-\rho)\varepsilon - \frac{1+\varepsilon}{2}\gamma v^2 = 0, \tag{B.5a}$$

$$\frac{r}{r + f(y)/y - f'(y)} = \frac{v}{\sigma}. \tag{B.5b}$$

If $v \to 0$ as $\eta \to 0$, by (B.5b) we have $r \to 0$ since y is bounded away from 0. But letting $\eta \to 0$ (and hence $v, r \to 0$) in (B.5a), we obtain $\rho\varepsilon = 0$, which is a contradiction. Therefore v is bounded away from 0 as $\eta \to 0$. □

References

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

George-Marios Angeletos and Vasia Panousi. Revisiting the supply side effects of government spending. *Journal of Monetary Economics*, 56(2):137–153, March 2009. doi:10.1016/j.jmoneco.2008.12.010.

George-Marios Angeletos and Vasia Panousi. Financial integration, entrepreneurial risk and global dynamics. *Journal of Economic Theory*, 146 (3):863–896, May 2011. doi:10.1016/j.jet.2011.02.001.

Shuhei Aoki and Makoto Nirei. Zipf's law, Pareto's law, and the evolution of top incomes in the U.S. 2016. URL `http://ssrn.com/abstract=2426418`.

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

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

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

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

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

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

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

Juan-Carlos Córdoba. On the distribution of city sizes. *Journal of Urban Economics*, 63(1):177–197, January 2008. doi:10.1016/j.jue.2007.01.005.

Angus Deaton and Christina Paxson. Intertemporal choice and inequality. *Journal of Political Economy*, 102(3):437–467, June 1994.

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

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

Xavier Gabaix. The granular origins of aggregate fluctuations. *Econometrica*, 79(3):733–772, May 2011. doi:10.3982/ECTA8769.

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

Xavier Gabaix, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. The dynamics of inequality. NBER Working Paper 21363, 2015. URL `http://www.nber.org/papers/w21363`.

François Geerolf. A static theory of Pareto distributions. 2016. URL `http://www.econ.ucla.edu/fgeerolf/geerolf-pareto.pdf`.

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

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

Simon Gilchrist, Vladimir Yankov, and Egon Zakrajšek. Credit market shocks and economic fluctuations: Evidence from corporate bond and stock market. *Journal of Monetary Economics*, 56(4):471–493, May 2009. doi:10.1016/j.jmoneco.2009.03.017.

Charles I. Jones. Pareto and Piketty: The macroeconomics of top income and wealth inequality. *Journal of Economic Perspectives*, 29(1):29–46, Winter 2015. doi:10.1257/jep.29.1.29.

Tom Krebs. Human capital risk and economic growth. *Quarterly Journal of Economics*, 118(2):709–744, 2003. doi:10.1162/003355303321675491.

Moche Levy and Sorin Solomon. Power laws are logarithmic Boltzmann laws. *International Journal of Modern Physics C*, 7(4):595–601, August 1996. doi:10.1142/S0129183196000491.

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

Erzo G. J. Luttmer. Models of growth and firm heterogeneity. *Annual Review of Economics*, 2:547–576, 2010. doi:10.1146/annurev.economics.102308.124410.

Ofer Malcai, Ofer Biham, and Sorin Solomon. Power-law distributions and Lévy-stable intermittent fluctuations in stochastic systems of many autocatalytic elements. *Physical Review E*, 60(2):1299–1303, August 1999. doi:10.1103/PhysRevE.60.1299.

Robert C. Merton. Lifetime portfolio selection under uncertainty: The continuous-time case. *Review of Economics and Statistics*, 51(3):247–257, August 1969. doi:10.2307/1926560.

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

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

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

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

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

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

Kenneth T. Rosen and Mitchel Resnick. The size distribution of cities: An examination of the Pareto law and primacy. *Journal of Urban Economics*, 8 (2):165–186, September 1980. doi:10.1016/0094-1190(80)90043-1.

Esteban Rossi-Hansberg and Mark L. J. Wright. Urban structure and growth. *Review of Economic Studies*, 74(2):597–624, 2007. doi:10.1111/j.1467-937X.2007.00432.x.

Makoto Saito. A simple model of incomplete insurance: The case of permanent shocks. *Journal of Economic Dynamics and Control*, 22(5):763–777, May 1998. doi:10.1016/S0165-1889(97)00077-8.

Herbert A. Simon. On a class of skew distribution functions. *Biometrika*, 42 (3/4):425–440, December 1955. doi:10.2307/2333389.

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

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

Lars E. O. Svensson. Portfolio choice with non-expected utility in continuous time. *Economics Letters*, 30(4):313–317, October 1989. doi:10.1016/0165-1765(89)90084-0.

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

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

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

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

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

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

George K. Zipf. *Human Behavior and the Principle of Least Effort*. Addison Wesley, Cambridge, MA, 1949.