Zipf's Law: A Microfoundation

Alexis Akira Toda*

This version: July 12, 2016
Very preliminary

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 operate a constant-returns-to-scale technology with multiple inputs, one of which is in inelastic 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.

Keywords: 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 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 firms as small 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.

In a seminal paper, Gabaix (1999) has shown that Zipf's law robustly 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

*Department of Economics, University of California San Diego. Email: atoda@ucsd.edu.

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

Electronic copy available at: http://ssrn.com/abstract=2808237

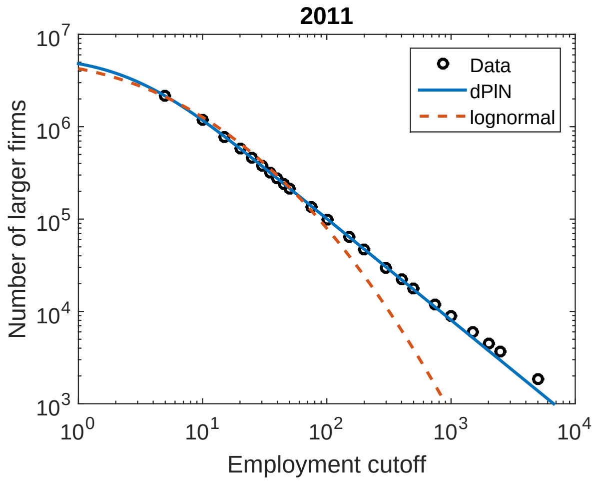

Figure 1: Log-log plot of firm size distribution.

Note: The figure plots employment cutoffs and the number of firms larger than the cutoffs. 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.

as well as models that explain Zipf's law.[2] Despite the considerable advances in the theory 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—has often been assumed without proper justifications. In this paper, I propose a dynamic general equilibrium model that explains Zipf's law without ad hoc assumptions, and hence provide a microfoundation for Zipf's law.

My model 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 inelastic 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. My contribution is thus in showing that condition (iii)—the existence of a production factor in inelastic 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. Thus my theory provides a microfoundation for Zipf's law.

To illustrate these points, I construct a stylized model of the population dynamics in cities (villages) as well as a more realistic model of entrepreneur-

[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, 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.

Electronic copy available at: http://ssrn.com/abstract=2808237

ship and firm size. In the first 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.

In the second model, I consider an economy consisting of entrepreneur-CEOs and 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 between 1 and 1.02 even under bankruptcy rates as high as 10%, replicating Zipf's law.

2 Difficulties with existing explanations

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

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, 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

[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.

cannot operate.[5] 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 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 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), we obtain

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

Hence Zipf's law is also equivalent to $x_{\min} \ll \bar{x}$: the minimum size is small relative to the average.

Although this model is purely mechanical, it underlies the mechanism of generating Zipf's law in most 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 constant-returns-to-scale production, multiplicative idiosyncratic risks, and homothetic preferences.[6] 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 to be 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,[7] and (ii) agents ignore the minimum size but it is imposed ex post. However, 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 order to obtain Zipf's law, which is a knife-edge case. Since the growth rate g is an endogenous variable in any fully

[5]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.

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

[7]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.

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.[8]

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 a model in the following sections.

3 A simple model of city size distribution

In this section I provide a minimal model of population dynamics and city size distribution to highlight the ingredients that give rise to Zipf's law.

3.1 Geometric Brownian motion with birth/death

First 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.[9] 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.[10] It is well known (see Appendix A) that the size distribution of units has a unique stationary density

$$f(x) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta} x_0^{\alpha} x^{-\alpha-1}, & (x \geq x_0) \\ \frac{\alpha\beta}{\alpha+\beta} x_0^{-\beta} x^{\beta-1}, & (0 < x < x_0) \end{cases} \tag{3.1}$$

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

$$\frac{v^2}{2}\zeta^2 + \left(g - \frac{v^2}{2}\right)\zeta - \eta = 0. \tag{3.2}$$

[8] 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 requires a fine-tuning of parameters.

[9] 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.

[10] 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 (3.2), we obtain the Pareto exponents

$$\alpha, \beta = \frac{1}{2}\left(\sqrt{\left(1 - \frac{2g}{v^2}\right)^2 + \frac{8\eta}{v^2}} \pm \left(1 - \frac{2g}{v^2}\right)\right). \tag{3.3}$$

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. This is a purely mathematical result, and of course there is no reason to expect that the growth rate of units is small. In order to explain Zipf's law, we need to introduce some economics, which I turn to next.

3.2 Size distribution with production and inelastic factor

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: each village gives out fraction $\kappa \in (0, 1)$ of its stock of potatoes to a village hit by a famine.

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, (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

$$\mathrm{d}x_t = (F(x_t, n_t) - \omega n_t)\,\mathrm{d}t - \eta\kappa x_t\,\mathrm{d}t + v x_t\,\mathrm{d}B_t, \tag{3.4}$$

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). The term $v x_t\,\mathrm{d}B_t$ 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 $f(x) = F(x, 1)$.[11] Since by assumption F is homogeneous of degree 1, we have $F(x, n) = nf(x/n)$. By the first-order condition, we obtain

$$\omega = f(y) - yf'(y), \tag{3.5}$$

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 (3.5). The profit rate per unit of potato is then

$$\mu = \frac{F(x_t, n) - \omega n}{x} = \frac{1}{y}(f(y) - (f(y) - yf'(y))) = f'(y). \tag{3.6}$$

Substituting the profit (3.6) into the resource constraint (3.4), we obtain

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

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 (3.7).

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 (3.7) 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

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

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

$$\mu = \eta. \tag{3.8}$$

Combining (3.6) and (3.8), the equilibrium potato per capita y is determined by $f'(y) = \eta$. The equilibrium wage is then determined by (3.5). Substituting (3.8) into the equation of motion (3.7) of potatoes in each village (and hence the population), we obtain

$$\mathrm{d}x_t = \eta(1 - \kappa)x_t \,\mathrm{d}t + vx_t \,\mathrm{d}B_t. \tag{3.9}$$

Note that (3.9) 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.

[11] 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.

Proposition 3.1. *The stationary city size distribution is double Pareto. The upper tail Pareto exponent ζ is given by α in* (3.3) *with $g = \eta(1-\kappa)$, which satisfies*

$$1 < \zeta < 1 + \frac{2\eta\kappa}{v^2}. \tag{3.10}$$

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

3.3 Discussion

The above model is highly stylized (the next section presents a model of the firm size distribution with fully optimizing agents). However, it is useful because it highlights the minimal sufficient conditions that give rise to Zipf's law, without going into the details of a more realistic but elaborate model. There are essentially two ingredients.

First, we have the mechanistic model (geometric Brownian motion and Poisson death), which gives us the double Pareto distribution for free. It turns out that the upper tail exponent α is close to 1 (Zipf's law) if the growth rate and the birth/death rate of individual units are small compared to the volatility.

Second, it is economics that gives us the low growth necessary to obtain Zipf's law. In this example, total population is fixed, and the production function exhibits decreasing returns with respect to one factor (potato). Therefore in steady state, by definition the aggregate stock of potatoes must be constant, which means that it is impossible for individual villages to grow on average faster than the rate of rare disasters, which is small. The equilibrium wage adjusts so that this low growth condition is satisfied endogenously.

Thus we can conclude that Zipf's law obtains whenever the following conditions hold: (i) individual units follow Gibrat's law of proportional growth (possibly due to multiplicative idiosyncratic shocks, constant-returns-to-scale technology, and homothetic preferences); (ii) individual units are reset at a (small) constant Poisson rate; (iii) there is a factor of production in inelastic supply, which limits aggregate growth.

I must emphasize that the purpose of the model in this section is to show that it is *possible* to explain Zipf's law without any ad hoc assumptions. There are obviously many important features that are left out in the model. First, in my model villages (cities) are exogenously fixed, but one might wonder why there are cities in the first place. As Rossi-Hansberg and Wright (2007) argue, the existence of cities (urban structure) suggests local increasing returns, while balanced growth suggests constant returns at the aggregate level. They propose a model with these features and explain the endogenous emergence of cities. A similar model can be developed in my setting since the mathematical structure (multiplicative shocks, constant-returns-to-scale, and homothetic preferences) is the same.

Second, my model predicts that the size distribution of cities is double Pareto, which has a kink at the mode and so 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 fixed, 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.

4 A model of firm size distribution

In the previous section I showed that it is possible to obtain Zipf's law without ad hoc assumptions. In this section I develop a more realistic model of the firm size distribution with fully optimizing agents. 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[12] 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 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

[12] 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.

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 (3.5) 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 (3.6) 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

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

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

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

So far I have implicitly assumed that the discount rate ρ and EIS ε are common across agent types, but this is not necessary. Allowing for heterogeneous parameters across types is also useful for calibration, and in fact, necessary for the existence of equilibrium when the bankruptcy rate η is small. Hence let ρ_W, ε_W be the parameter values for the workers, and let the symbols without subscripts be those of the entrepreneurs. The following theorem characterizes the equilibrium.

Theorem 4.1. *Suppose that $f(x) = F(x,1)$ satisfies the usual Inada conditions $f' > 0$, $f'' < 0$, $f'(0) = \infty$, and $f'(\infty) \le 0$. Then a stationary equilibrium exists if and only if*

$$\left(1 - \frac{1}{\bar{y}N}\right)\eta > (\rho_W - \rho)\varepsilon, \tag{4.8}$$

where $\bar{y} > 0$ is the (unique) number such that $f'(\bar{y}) = \rho_W + \eta$. The steady state is unique and satisfies the following properties.

1. *The risk-free rate equals the discount rate of workers: $r = \rho_W$.*
2. *The capital-labor ratio $y = K/N$ is the unique solution in $(0, \bar{y})$ to*

$$\left(1 - \frac{1}{yN}\right)\eta = (\rho_W - \rho)\varepsilon + (1+\varepsilon)\frac{(f'(y) - \rho_W - \eta)^2}{2\gamma\sigma^2}. \tag{4.9}$$

3. *The net worth x_t of individual entrepreneurs evolves according to the geometric Brownian motion*

$$\mathrm{d}x_t = \eta(1-\kappa)x_t\,\mathrm{d}t + vx_t\,\mathrm{d}B_t, \tag{4.10}$$

where $\kappa = \frac{1}{K} = \frac{1}{yN}$ is the ratio between the initial and the steady state capital and $v = \frac{f'(y) - \rho_W - \eta}{\gamma\sigma} > 0$ is volatility.

Condition (4.8) says that entrepreneurs must be sufficiently impatient. This assumption is natural, for otherwise their wealth will grow indefinitely and there will be no steady state. In order for (4.8) to hold for small enough $\eta > 0$, it is necessary and sufficient that $\rho_W < \rho$. Therefore in order to consider the limit where the bankruptcy rate converges to zero, we need to assume that workers are more patient than entrepreneurs (otherwise an equilibrium does not exist).

Note that the equation of motion (4.10) is exactly the same as (3.9). Hence by Proposition 3.1, the upper tail Pareto exponent ζ satisfies the bound (3.10). However, since κ, v are endogenous unlike in the previous model, it is not immediately clear whether Zipf's law holds as $\eta \to 0$. We can nevertheless show that Zipf's law holds in this case, too.

Theorem 4.2. *Suppose that $\rho_W < \rho$, so the equilibrium existence condition (4.8) holds for small enough $\eta > 0$. As $\eta \to 0$, we obtain Zipf's law $\zeta \to 1$.*

Note that Theorem 4.2 is an asymptotic result, and hence for any given parameters the upper tail Pareto exponent need not be close to 1, although the bound (3.10) is always true. Whether ζ is close to 1 or not is therefore a quantitative question, which I address in the numerical example below.

5 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 depreciation rate.

5.1 Calibration

The model is completely specified by the parameters $(\rho_W, \rho, \gamma, \varepsilon, \alpha, \delta, \sigma, \eta, N)$.[13] 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, 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,[14] 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.

[13] 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.

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

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 4.2 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 (3.10) 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 $\gamma = 1$.[15] As a robustness check, I also consider the cases $\gamma = 0.5, 2$.

5.2 Results

By Theorem 4.1, computing the equilibrium reduces to solving a single nonlinear equation (4.9). Table 1 shows the results. 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 by (3.10).

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

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

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 (3.10), 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.

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 reference 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.[16] Therefore we obtain the (parabolic) partial differential equation (PDE) (A.2). □

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

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

[16] To see this more rigorously, set

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

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 (3.2). 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 (3.1).

B Proofs

Proof of Proposition 3.1. Since the equation of motion (3.9) is a special case of the mechanistic model (2.1), it suffices to show the bound (3.10). 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 (3.2). 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 (3.10). □

Proof of Theorem 4.1. First I prove the properties of the equilibrium assuming existence, and later I show existence and uniqueness.

Step 1. The equilibrium risk-free rate is $r = \rho_W$.

In steady state, the wealth of the worker cannot grow. 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

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

Step 2. The net worth x_t of individual entrepreneurs satisfies (4.10)*.*

Since (4.6) holds and $k_t = \theta x_t$, there 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

$$(g-\eta)K + \eta = 0 \iff g = (1-\kappa)\eta, \tag{B.2}$$

where $\kappa = 1/K$ is the amount of initial capital relative to the steady state value. Substituting this g into (4.6), we obtain (4.10).

Step 3. There exists a (unique) steady state if and only if (4.8) *holds. The equilibrium capital-labor ratio $y = K/N$ solves* (4.9)*.*

By (B.2) 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 $\kappa = \frac{1}{yN}$, $r = \rho_W$, $\mu = f'(y)$, and $r_e = r + \eta = \rho_W + \eta$, the equilibrium capital-labor ratio y must satisfy (4.9). Let

$$\phi(y) = \left(1 - \frac{1}{yN}\right)\eta - (\rho_W - \rho)\varepsilon - (1+\varepsilon)\frac{(f'(y) - \rho_W - \eta)^2}{2\gamma\sigma^2} \tag{B.3}$$

be the left-hand side minus the right-hand side of (4.9). Let us show that ϕ is strictly increasing on $(0, \bar{y})$. By simple algebra, we have

$$\phi'(y) = \frac{\eta}{y^2 N} - (1+\varepsilon)\frac{f'(y) - \rho_W - \eta}{\gamma\sigma^2} f''(y).$$

For $y < \bar{y}$, since $f'' < 0$ we have $f'(y) > f'(\bar{y}) = \rho_W + \eta$. Therefore $\phi'(y) > 0$.

Now note that (4.9) is equivalent to $\phi(y) = 0$. If the condition (4.8) does not hold, then $\phi(\bar{y}) \le 0$. Also, clearly $\phi(0) = -\infty$. Since ϕ is strictly increasing, we would have $\phi(y) < 0$ for all $y \in (0, \bar{y})$, and therefore a steady state does not exist. Conversely, if (4.8) holds, then $\phi(\bar{y}) > 0$, so by the intermediate value theorem there exists $y \in (0, \bar{y})$ such that $\phi(y) = 0$. Since ϕ is strictly increasing, y is unique.

Step 4. The propensity to consume out of wealth, m in (4.5b), *is positive. The volatility of entrepreneur's wealth is given by* $v = \frac{f'(y) - \rho_W - \eta}{\gamma\sigma} > 0$.

By the construction of y, we have $f'(y) > \rho_W + \eta$. By (4.7b), $\mu = f'(y)$, $r_e = r + \eta$, and $r = \rho_W$, we have $v = \frac{f'(y) - \rho_W - \eta}{\gamma\sigma} > 0$. By (4.5a), in equilibrium the fraction of wealth invested in physical capital is

$$\theta = \frac{\mu - r_e}{\gamma\sigma^2} = \frac{f'(y) - \rho_W - \eta}{\gamma\sigma^2}.$$

To show that the propensity to consume is positive, note that by (4.4), (4.5a), and (4.10), we have

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

Since $r_e = \rho_W + \eta$ and $\mu = f'(y)$, it follows that

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

Proof of Theorem 4.2. Since the bound (3.10) 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. 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 $\phi(y; \eta)$ be $\phi(y)$ in (B.3), given $\eta > 0$. Then we have

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

Since ϕ is strictly increasing in y and $\phi(y, \eta) = 0$ in steady state, 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 (4.9) 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 $\eta \to 0$. $\square$

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.

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.

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. 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 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.