11 Power law, firm distribution and Gibrat's Law

1 Power law

1.1 Def.

In statistics, a power law is a functional relationship between two quantities, where a relative change in one quantity results in a proportional relative change in the other quantity, independent of the initial size of those quantities:

one quantity varies as a power of another.

1.2 Properties

1.2.1 Scale invariance

1. Given a relation $f(x) = ax^{-k}$, scaling the argument x by a constant factor c causes only a proportionate scaling of the function itself, i.e.,

$$f(cx) = a(cx)^{-k} = c^{-k} f(x) \propto f(x) \tag{1}$$

   - where $\propto$ denotes direct proportionality.
   - (1) implies that scaling by a constant c simply multiplies the original power-law relation by the constant c^{-k}.
2. Thus, it follows that all power laws with a particular scaling exponent are equivalent up to constant factors, since each is simply a scaled version of the others.
   - This behavior is what produces the linear relationship when logarithms are taken of both $f(x)$ and x.
     - The straight-line on the log-log plot is often called the *signature of a power law*.
   - With real data, such straightness is a **necessary**, but not sufficient condition for the data following a power-law relation.
     - There are many ways to generate finite amounts of data that mimic this signature behaviour, but in their asymptotic limit, are not true power laws.
       - e.g., if the generating process of some data follows a **log-normal distribution**.
         - a log-normal distribution is a continous probability distribution whose logarithm is normally distributed.
           - If X is log-normally distributed, then $Y = \ln(X)$ has a normal distribution.

1.2.2 Lack of well-defined average value

1. A power-law function x^{-k}
   - has a well-defined mean over $x \in [0, \infty)$ only if $k > 2$;
   - has a finite variance only if $k > 3$.
2. Most identified power laws in nature have exponents such that the mean is well-defined by the variance is not, implying they are capable of black swan behavior.
   - Black swan theory is a metaphor that describes an event that comes as a surprise, has a major effect, and is often inappropriately rationalised after the fact with the benefit of hindsight.
   - Two effects:
     - This makes it incorrect to appy traditional statistics that are based on variance and standard deviation (such as regression analysis).
     - This allows for cost-efficient interventions.
3. The median does exist, however:
   - for a power law x^{-k}, with exponent $k > 1$, it takes the value $2^{1/(k-1)} x_{\min}$,
     - $x_{\min}$ is the minimum value for which the power law holds.

1.2.3 Universality

1. The equivalence of power laws with a particular scaling exponent can have a deeper origin in the dynamical processes that generate the power-law relation.
2. Formally, this sharing of dynamics is referred to as universality.
   - systems with precisely the same critical exponents are said to belong to the same universality class.

1.3 Power-law functions

1. Scientific interest in power-law relations stems partly from the ease with which certain general classes of mechanisms generate them.
   - The demonstration of a power-law relation in some data can point to speicific kinds of mechanisms that might underlie the natural phenomenon in question, and can indicate a deep connection with other, seemingly unrelated systems.
   - In complex systems, power laws are often thought to be signatures of hierarchy or of specific stochastic processes.
     - e.g.:
       - Pareto's law of income distribution
     - Research on the origins of power-law relations, and efforts to observe and validate them in the real world, is an active topic of research in many fields of science.
2. Much of the recent interest in power laws comes from the study of probability distributions:
   - This is because the distributions of a wide variety of quanities seem to follow the power-law form, at least in their upper tail (large events).

    - The behavior of these large events connexts thesee quantities to the study of theory of large deviations,
        - It considers the frequency of extremely rate events.
3. In empirical contexts, an approximation to a power-law $o(x^k)$ often includes a deviation term ε, which can represent uncertainty in the observed values

$$y = ax^k + \varepsilon \tag{2}$$

4. Mathematically, a strict power function cannot be a probability distribution, but a distribution that is a truncated power function is possible:

$$p(x) = Cx^{-\alpha} \; for \; x > x_{\min} \tag{3}$$

    - where
        - the exponent α is greater than 1.
            - Otherwise, the tail has infinite area.
        - The minimum value $x_{\min}$ is needed.
            - Otherwise the distribution has infinite area as x approaches 0;
        - The constant C is a scaling factor to ensure that the total area is 1, as required by a probability distribution.

2 Pareto distribution

2.1 Def. (Type I)

1. If X is a random variable with a (Type I) Pareto distribution, then the probability that X is greater than some number x, i.e., the survival function (tail function), is given by

$$\bar{F}(x) = P(X > x) = \begin{cases} (\frac{x_m}{x})^\alpha & x \geq x_m \\ 1 & x < x_m \end{cases} \tag{4}$$

    - where
        - x_m: the (necessarily positive) minimum possible value of X;
        - α: a positive parameter.
2. The Pareto Type I distribution is characterized by a scale parameter x_m and a shape parameter α.
    - α: the tail index.
    - When this distribution is used to model the distribution of wealth, then the parameter α is called the Pareto index.

2.2 Properties

2.2.1 Cumulative distribution function

The cumulative distribution function a Pareto random variable with parameter α and x_m is

$$F_X(x) = \begin{cases} 1 - (\frac{x_m}{x})^\alpha & x \geq x_m \\ 0 & x < x_m \end{cases} \tag{5}$$

2.2.2 Probability density function

It follows by differentiation from (5) that the probability density function is

$$f_X(x) = \begin{cases} \frac{\alpha x_m^\alpha}{x^{\alpha+1}} & x \geq x_m \\ 0 & x < x_m \end{cases} \tag{6}$$

- when plotted on linear axes, the distribution assumes the familiar J-shaped curve.
  - which approaches each of the orthogonal axes asymptotically.
  - All segments of the curve are self-similar subject to appropriate scalling factors.
- When plotted in a log-log plot, the distribution is represented by a straight line.

2.2.3 Moments and characterization function

1. The expected value of a random variable following a Pareto distribution is

$$E(X) = \begin{cases} \infty & \alpha \leq 1 \\ \frac{\alpha x_m}{\alpha - 1} & \alpha > 1 \end{cases} \tag{7}$$

2. The variance of a random variable following a Pareto distribution is

$$Var(X) = \begin{cases} \infty & \alpha \in (1, 2] \\ (\frac{x_m}{\alpha-1})^2 \frac{\alpha}{\alpha-2} & \alpha > 2 \end{cases} \tag{8}$$

  - If $\alpha \leq 1$, the the variance does not exist.

2.3 Relation to Zipf's law

1. The Pareto distribution is continuous probability distribution, but Zipf's law, also sometimes called the zeta distribution, is a discrete distribution, separating the values into imple ranking.
   - Both are a simple power law with a negative exponents, scaled so that their cumulative distributions equal 1.
   - Zipf's can be derived from the Pareto distribution if the x values (income) are binned into N ranks sothat the number of people in each bin follows a $1/rank$ pattern.

3 A generalized Gibrat's law

3.0 Abstract

1. Many economic and non-economic variables, such as income, wealth, firm size, or city size often distribute Pareto in the upper tail.
   - It is well established that Gibrat's law can explain this phenomenon,
   - but Gibrat's law often does not hold.

2. This note characterizes a class of processes,
   - One that includes Gibrat's law as a special case, that can explain Pareto distributions.
   - Of particular importance is a parsimonoious generalization of Gibrat's law that allows **size** to affect the **variance** of the growth process but **not its mean**.
3. This note also shows that under **plausible conditions**,
   - **Zipf's law** is equivalent to **Gibrat's law**.

3.1 Motivation

1. The **upper tail** of the **cross-sectional** distribution of various economic and non-economic variables, such as wealth, income, firm size, city size, or publications per author in *Econometrica*, often conforms well to a **Pareto distribution**, or a **power law**.
   - Champernowne (1953) and Simon (1955) showed that
     - **Pareto distributions** arise naturally if the **time series behavior** of the variable in question satisfies what is known as **Gibrat's law**:
       - the **current position of the variable** does **not influence** its **expected rate of growth** nor the **variance of its growth rate**.
   - For concreteness, consider a discrete Markov process $\{X_t\}$ taking values in the set $X = \{x_i \equiv (1+g)^i\}_{i=1}^N$, where $g > 0$ and $N \leq \infty$ is a positive integer.
     - Suppose that the conditional distribution of X_{t+1}, given X_t, is described by the Markov matrix

$$\pi = \begin{pmatrix} a_1+b_1 & c_1 & 0 & 0 & \cdots & 0 & 0 \\ a_2 & b_2 & c_2 & 0 & \cdots & 0 & 0 \\ 0 & a_3 & b_3 & c_3 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & a_{N-1} & b_{N-1} & c_{N-1} \\ 0 & 0 & 0 & 0 & 0 & a_N & b_N+c_N \end{pmatrix} \tag{9}$$

       - where
         - $\pi_{i,j} = P\{X_{t+1} = x_j | X_t = x_i\}$;
         - $a_i + b_i + c_i = 1$;
         - $[a_i, b_i, c_i] \gg 0$.
       - Notice that $\{X_t\}$ has a lower barrier and an upper barrier if $N < \infty$. (**because of infinite numbers???**)
       - Denoting μ_i the **conditional expected growth rate** of X_t and σ_i^2 its **conditional variance**, (**Q1: ??????**)

$$\begin{aligned} \mu_i &\equiv E[\frac{X_{t+1}-X_t}{X_t}|X_t = x_i] \\ &= gc_i - \frac{g}{1+g}a_i, for\ i = 2, \ldots, N-1, \end{aligned} \tag{10}$$

       and

$$\sigma_i^2 \equiv E[(\frac{X_{t+1}-X_t}{X_t}-\mu_i)^2|X_t=x_i] \tag{11}$$
$$= a_i(\frac{g}{1+g}-\mu_i)^2 + b_i\mu_i^2 + c_i(g-\mu_i)^2, for\ i=2,\dots,N-1.$$

    - Champernowne (1953) showed that
        - if $a_i = a, b_i = b, c_i = c, a > c$, and $N = \infty$ (**Q2: Too strong here???**), then the invariant distribution of π has the Pareto form

$$P\{X_t \geq x\} = Mx^{-\delta} \tag{12}$$

        - This characterization implies that
            - **off the boundaries** (**?**), $\{X_t\}$ satisfies Gibrat's law:
                - as μ_i and σ_i^2 are independent of i. (**???**)
  - For diverse economic and non-economic problems, Gibrat's law provides an unsatisfactory characterisation of the underlying dynamics generating power laws.
    - e.g.:
        - evidence suggests that
            - small firms face higher proportiional **risk** than large firms (Evans, 1987),
            - richer individuals take more proportional **risk** than that poorer individuals (Carroll, 2000),
            - large diversified cities experience **lower growth volatility** than smaller cities (Fujita et al., 1999).
    - Yet the distributions of firm sizes, incomes, wealth, and city sizes are often well described by Pareto distributions particularly in the upper tail (see e.g., Axtell, 2011; Dragulescu and Yakovenko, 2001; Champernowne, 1953; Zipf, 1949).
2. This note makes three contributions;
  - First, it **characterizes** Markov processes of the class illustrated by π in (9).
    - i.e., with a **quasi-diagonal** structure, leading to **Pareto distributions**.
        - They show that if $a_i = a\mu^i$ and $c_i = c\mu^i$, where $0 < \mu \leq 1$, then the invariant distribution associated to π is Pareto.
        - This finding generalizes Sims' and Chapernowne's results in that
            - **Gibrat's law** is included as the **special of** $\mu = 1$.
    - Importantly, this class includes diffusions that are Markov processes with a continuous state-space and in continuous time.
  - Second, they show that if a **weak version of Gibrat's law** holds, one that requires **only mean growth** to be independent of size, then the Markov process just described is the **only process supporting Pareto distributions**.
    - They motivate the importance of this weak restriction on the fact that
        - Many variables in economics grow over time and
        - balanced growth conditions are often required on theoretical or empirical grounds.
    - Their generalized Gibrat's law provides a general process that is scale invariant.

- Finally, they establish that along balanced growth paths, **Zipf law** can **only** result from **Gibrat's law**.

3.2 Main results

Setup

1. Let
    - $p_t = [p_{1t}, p_{2t}, \ldots, p_{Nt}]$ be the unconditional distribution of X_t at time t ;
    - π , defined by (9), be its conditional distribution.
    - Thus, $p_{t+1} = p_t \cdot \pi$.
2. An invariant distribution p satisfies $p = p \cdot \pi$.
3. Given that $[a_i, b_i, c_i] \gg 0$ and p exists,
    - p is also unique and $p = \lim_{t\to\infty} p_t$ if N is finite. (**??**)
4. For $N = \infty$, (**??**)
    - the additional restriction $a_i > c_i$ guarantees **existences**, **uniqueness**, and **convergence**.
5. By the **ergodic theorem** (**Q3: Same we had in finite Markov chain???**), p can also be interpreted as the **cross-sectional distribution of population** across **states of a closed system without entry or exit**.
6. Notice that p satisfies

$$p_i = a_{i+1}p_{i+1} + b_i p_i + c_{i-1}p_{i-1}, for\ i = 2, \ldots, N-1, \tag{13}$$

   and

$$p_1 = a_2 p_2 + (a_1 + b_1)p_1\ and\ p_N = (b_N + c_N)p_N + c_{N-1}p_{N-1} \tag{14}$$

The following result is due to Champernowne (1953):

Proposition 1

Suppose $a_i = a, b_i = b$, and $c_i = c$.

Then the **only solution** to the system (13) and (14) is

$$p_i = M\rho^i = Mx_i^{-\delta}, 0 < \rho < 1 \tag{15}$$

- where
    - M: a constant,
    - $\rho = \frac{c}{a}$, and
    - $\delta = \frac{\ln(a/c)}{\ln(1+g)}$.

Moreover,

$$P\{X_t \geq x_i\} = \frac{1}{1-(c/a)} Mx_i^{-\delta}\ if\ N = \infty \tag{16}$$

Proof. (Prop.1 is a special case of Prop.2)

7. Prop. 1 formalizes the idea that
    - Gibrat's law explain Pareto distributions.
    - Notice, however, that
        - Gibrat's law can induce other invariant distributions depending on boundary conditions.
8. To **uncover** additional Markov chains leading to Pareto distributions, consider the system of (13) and (14) restricted so that p_i adopts the Pareto form

$$p_i = M\rho^i, 0 < \rho < 1 \tag{17}$$

    - In that case, (13) and (14) can be written as (**TBC???: Checked on the paper**)

$$a_{i+1}\rho^{i+1} - (a_i + c_i)\rho^i + c_{i-1}\rho^{i-1} = 0, for\ i = 2, \ldots, N-1, \tag{18}$$

    and

$$c_1 = \rho a_2, and\ a_N = c_{N-1}/\rho \tag{19}$$

    - For given ρ, (18) and (19) is a system of N equations in $2N$ unknowns.
        - It has **potentially multiple solutions**.
        - The following is one of them:

Lemma 1

For given ρ, a solution to the system (18) and (19) is

$$a_i = a\theta^i, c_i = c\theta^i, for\ i = 1, \ldots, N, \tag{20}$$

- where
    - $\theta \equiv \frac{1}{\rho}\frac{c}{a}$, a and c are strictly positive parameters satisfying

$$\theta \leq 1, and\ a + c < 1 \tag{21}$$

Proof. (TBC)

To check whether (18) is satisfied, plug (20) into the LHS of (18),

$$\begin{aligned} & a_{i+1}\rho^{i+1} - (a_i + c_i)\rho^i + c_{i-1}\rho^{i-1} \\ &= a(\theta\rho)^{i+1} - (a+c)(\theta\rho)^i + c(\theta\rho)^{i-1} \\ &= (\theta\rho)^{i-1}[a(\theta\rho)^2 - (a+c)(\theta\rho) + c] \end{aligned} \tag{22}$$

Since we have $\theta \equiv \frac{1}{\rho}\frac{c}{a}$, we obtain

$$\theta\rho = \frac{c}{a} \tag{23}$$

Therefore, substitute (23) into (22), we have

$$
\begin{aligned}
& a_{i+1}\rho^{i+1} - (a_i + c_i)\rho^i + c_{i-1}\rho^{i-1} \\
&= a(\theta\rho)^{i+1} - (a+c)(\theta\rho)^i + c(\theta\rho)^{i-1} \\
&= (\theta\rho)^{i-1}[a(\theta\rho)^2 - (a+c)(\theta\rho) + c] \\
&= (\theta\rho)^{i-1}[a(\frac{c}{a})^2 - (a+c)(\frac{c}{a}) + c] \\
&= (\theta\rho)^{i-1}[\frac{c^2}{a} - c - \frac{c^2}{a} + c] \\
&= (\theta\rho)^{i-1} \cdot 0 \\
&= 0
\end{aligned} \tag{24}
$$

Thus, (18) is satisfied.

Similarly, to check whether (19) is satisfied, plug (20) into the LHS of first expression in (19), we have

$$
c_1 = c\theta \tag{25}
$$

Since we have $\theta \equiv \frac{1}{\rho}\frac{c}{a}$, we obtain

$$
c = \theta\rho a \tag{26}
$$

By (20), plug (26) into (25), (25) becomes

$$
\begin{aligned}
c_1 &= c\theta \\
&= \theta\rho a\theta \\
&= \theta^2 a\rho \\
&= a_2\rho
\end{aligned} \tag{27}
$$

Thus, the first expression of (19) is satisfied.

Plug (20) into the LHS of the second expression in (19), we have

$$
a_N = a\theta^N \tag{28}
$$

Since we have $\theta \equiv \frac{1}{\rho}\frac{c}{a}$, we obtain

$$
a = \frac{c}{\rho\theta} \tag{29}
$$

By (20), plug (29) into (28), (28) becomes

$$
\begin{aligned}
a_N &= a\theta^N \\
&= \frac{c}{\rho\theta}\theta^N \\
&= \frac{c}{\rho}\theta^{N-1} \\
&= \frac{c_{N-1}}{\rho}
\end{aligned} \tag{30}
$$

Thus, the second expression of (19) is satisfied.

Proof Done

9. The restrictions on a and c in **Lemma 1** guarantee that $0 < a_i + c_i < 1$ for $i = 1, \ldots, N$.
   - **Lemma 1** takes p as given to derive a particular Markov chain π.
   - **Prop. 2** takes π as given to derive p.

Proportion 2

Let the Markov chain π defined by (9) satisfying

$$a_i = a\theta^i, c_i = c\theta^i, for\ i = 1, \ldots, N, \tag{31}$$

- where
  - $0 < c + a < 1$,
  - $0 < \theta \leq 1$ And,
  - $\frac{c}{a\theta} < 1$.

Then

$$p_i = M\rho^i = Mx_i^{-\delta} \tag{32}$$

- where
  - M: a constant,
  - $\rho \equiv \frac{c}{a\theta}$ and,
  - $\delta = \frac{\ln(1/\rho)}{\ln(1+g)} > 0$.

is the unique absolute convergent solution to (13) and (14).

Moreover,

$$P\{X_t \geq x_i\} = \frac{M}{1-\rho} x_i^{-\delta}\ if\ N = \infty \tag{33}$$

Proof. (TBC???: Checked as the following.)

By **Lemma 1**, we know that

$$p_i = M\rho^i \tag{34}$$

satisfy equations (13), (14) and (18), (19) for $\rho \equiv \frac{1}{\theta}\frac{c}{a}$.

By the definition of x_i, we have

$$x_i = (1+g)^i, for\ i = 1, \ldots, N \tag{35}$$

From (35), we have

$$i = \frac{\ln x_i}{\ln(1+g)} \tag{36}$$

Plug (36) into (34), and since $\delta = \frac{\ln(1/\rho)}{\ln(1+g)} = \frac{-\ln(\rho)}{\ln(1+g)}$, we obtain

$$
\begin{aligned}
p_i &= M\rho^i \\
&= M\rho^{\frac{\ln x_i}{\ln(1+g)}} \\
&= Me^{\frac{(\ln x_i)(\ln \rho)}{\ln(1+g)}} \\
&= Me^{(\ln x_i)\frac{\ln \rho}{\ln(1+g)}} \\
&= Mx_i^{\frac{\ln \rho}{\ln(1+g)}} \\
&= Mx_i^{-\delta}
\end{aligned} \tag{37}
$$

Uniqueness and absolute convergence follow from a standard result in stochastic processes (Cox and Miller, 1965, pp. 108-10, **Q4: TBC???**).

Finally, by (34) and (37),we have

$$
\begin{aligned}
P\{X_t \geq x_i\} &= \sum_{j=i}^{\infty} p_j \\
&= \sum_{j=0}^{\infty} p_i (\rho)^j \\
&= p_i \sum_{j=0}^{\infty} (\rho)^j \\
&= p_i \frac{1}{1-\rho} \\
&= Mx_i^{-\delta} \frac{1}{1-\rho}
\end{aligned} \tag{38}
$$

Proof Done.

10. Notice that **Prop. 1** is obtained as the speical case of **Prop. 2** in which $\theta = 1$.
11. The Markov chain described by **Prop. 2** is just one of many chains that can explain Pareto distributions.
    - Suppose π is further restricted to satisfy a weak version of Gibrat's law:
      - the expected growth rate of X_t be independent of the position of X_t.
    - Some examples in which this is a natural restriction are provided below by **Theorem 1**.
    - In that case, one can show that the
      - only Markov chain π supporting Pareto distributions is of the form described by **Prop. 2**.
    - The following theorm (**Theorem 1**) **normalises the common expected growth rate** to 0.
      - But this requirement can be relaxed.

Theorem 1

Suppose that unconditional invariant distribution of π has the Pareto form

$$p_i = M\rho^i = Mx_i^{-\delta} \tag{39}$$

- where
  - M: a constant,
  - $\rho < 1$ and,
  - $\delta = \frac{\ln(1/\rho)}{\ln(1+g)} \leq 1$.

Suppose, moreover, that

- $\mu_i = 0$ for $i = 2, 3, \ldots$ and
- $N = \infty$.

Then π must be of the form

$$a_i = a\theta^i, c_i = c\theta^i, for\ i = 2, 3, \ldots, N(?), \tag{40}$$

- where
  - $0 < c + a < 1$,
  - $0 < \theta = \frac{1}{\rho}\frac{1}{1+g} \leq 1$ And,
  - $a = (1+g)c$.

Moreover,

$$\sigma_i^2 = A\theta^i \tag{41}$$

- where
  - $A \equiv a(\frac{g}{1+g})^2 + cg^2$.

Proof. (TBC???: Checked as the following.)

From (10), we have

$$\mu_i = 0 \iff a_i = (1+g)c_i \tag{42}$$

Plug (42) into (18) produces

$$\begin{gathered} a_{i+1}\rho^{i+1} - (a_i + c_i)\rho^i + c_{i-1}\rho^{i-1} \\ = (1+g)c_{i+1}\rho^{i+1} - (2+g)c_i\rho^i + c_{i-1}\rho^{i-1} \\ = (1+g)z_{i+1} - (2+g)z_i + z_{i-1} \\ = 0, \\ where\ z_i \equiv c_i\rho^i,\ for\ i = 2, \ldots, N-1, \end{gathered} \tag{43}$$

The associated characteristic equation for (43):

$$(1+g)z^2 - (2+g)z + 0 = 0 \tag{44}$$

has two roots:

$$z_1 = 1 \; or \; z_2 = \frac{1}{1+g} \tag{45}$$

Therefore, since $z_i \equiv c_i \rho^i$, the solution for z_i and c_i are (**Q5: Why this form??? Refer to the notes 3 of ECON8026?**)

$$z_i = k_1 (\frac{1}{1+g})^i + k_2 \tag{46}$$

and

$$\begin{aligned} c_i &= z_i \rho^{-i} \\ &= [k_1 (\frac{1}{1+g})^i + k_2] \rho^{-i} \\ &= k_1 (\frac{1}{\rho(1+g)})^i + k_2 \rho^{-i} \end{aligned} \tag{47}$$

- where $k_1 \geq 0$ and $k_2 \geq 0$. (**Q6: ??? TBC**)

So unless $k_2 = 0$, for other cases when $k_2 > 0$, we have

$$c_i > 1, \, for \; large \; i \tag{48}$$

Thus, (**??? In order to let c_i be some numbers smaller than 1.**) let $k_2 = 0$, we have

$$c_i = k_1 (\frac{1}{\rho(1+g)})^i = c\theta^i \tag{49}$$

- where
  - $c = k_1$;
  - $\theta = \frac{1}{\rho(1+g)}$.

Similarly, for a_i, since $a_i = (1+g)c_i$, we obtain (**??? TBC**)

$$\begin{aligned} a_i &= (1+g)c_i \\ &= (1+g)k_1 (\frac{1}{\rho(1+g)})^i \end{aligned} \tag{50}$$

- where
  - $a = k_1(1+g)$;
  - $\theta = \frac{1}{\rho(1+g)}$.

Finally, plug (40) into (11), and since $\mu_i = 0$, we have

$$\begin{aligned}\sigma_i^2 &\equiv E[(\frac{X_{t+1}-X_t}{X_t}-\mu_i)^2|X_t=x_i] \\ &= a_i(\frac{g}{1+g}-\mu_i)^2+b_i\mu_i^2+c_i(g-\mu_i)^2 \\ &= a_i(\frac{g}{1+g})^2+c_ig^2 \\ &= a\theta^i(\frac{g}{1+g})^2+c\theta^ig^2 \\ &= \theta^i[a(\frac{g}{1+g})^2+cg^2] \\ &= A\theta^i, \\ &for\ i=2,\ldots,N-1.\end{aligned} \tag{53}$$

- where $A=a(\frac{g}{1+g})^2+cg^2$.

Proof Done.

12. The restriction $\delta \leq 1$ in the **Theorem 1** is required to guarantee that

$$c_i+a_i\leq 1 \tag{52}$$

    - for all i.
13. **Theorem 1** provides an even **more parsimonious process** than **Prop. 2** to explain Pareto distributions.
    - This process is relevant when it is reasonable to expect that
        - a variable's position does not influence its expected rate of growth.
    - e.g.:
        - models of economic growth typically impose such restriction in the form of "balanced growth" conditions (**?**).
            - A stochastic definition of a balanced growth is precisely that
                - the expected growth rate of a variable is
                    - constant through time, and therefore,
                    - Independent of the size of the variable.

3.3 An application: city size distribution

1. A case in which a balanced growth condtion is expected to be satisfied is that of the evolution of U.S. cities.