

# 14.999: Topics in Inequality, Lecture 8

## Pareto Income and Wealth Distributions

Daron Acemoglu

MIT

April 1, 2015.

# Introduction

- The right tail of income and wealth distributions often resemble Pareto.
- The Pareto distribution emerges in many settings, both in economics (firm size distribution, sizes of cities and perhaps income distribution) and in other social sciences (language, family names, popularity, certain social network patterns, crime per convict), and even in various physical environments (e.g., sizes of large earthquakes, power outages, etc.).
- Pareto distributions might also provide some insights on the relationship between interest rates and growth rate and top inequality, as claimed by Piketty (2013).
- In this lecture, we will discuss why Pareto distributions (or “Pareto tails”) could emerge under certain circumstances, and why income and wealth might take this form.

# Pareto Distribution

- Many quantities in economics, other social sciences and physical sciences appear to be well approximated by Pareto distribution.
- Pareto distribution or the power law has the following counter-cumulative distribution function:

$$\mathbb{G}(y) \equiv 1 - \Pr[\tilde{y} \leq y] = \Gamma y^{-\lambda},$$

where  $\lambda \geq 1$  is the shape parameter.

- When the literature refers to the Pareto or the power law distribution, this generally means that the distribution has Pareto tails, meaning that it takes this form for  $y$  large.

# Zipf Distribution

- The Zipf distribution is a special case with  $\lambda = 1$  (or sometimes it is used for the case where  $\lambda$  is approximately equal to 1).
- Empirical city size distribution and firm size distributions appear to be well approximated by this distribution, and for city size distributions, this is generally referred to as “Zipf’s Law”.
- This is also equivalent to a relationship of slope -1 between log rank of the city (according to city size) and log of the population, or:

$$\ln \text{Rank}_j = c - \ln y_j,$$

where  $y$  is size. Thus

$$y_j = \frac{C}{\text{Rank}_j}.$$

- Rewriting this

$$\Pr(\tilde{y} > y_j) = \frac{C}{y_j},$$

- This may apply exactly or only in the “tails,” i.e.,  $y_j$  large enough.

# Firm Size Distribution

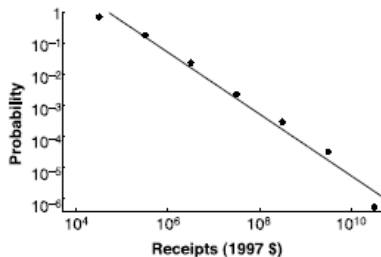
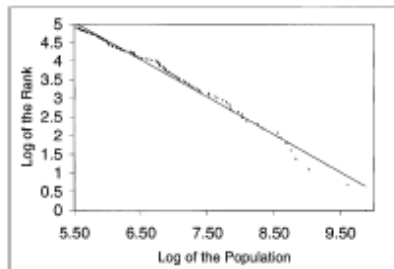


Fig. 2. Tail cumulative distribution function of U.S. firm sizes, by receipts in dollars. Data are for 1997 from the U.S. Census Bureau, tabulated in bins whose width increases in powers of 10. The solid line is the OLS regression line through the data and has slope of 0.994 (SE = 0.064; adjusted  $R^2 = 0.976$ ).

Axtell 2001

# For US Metropolitan Areas in the 1990s



## Top Inequality and Pareto

- If income distribution is Pareto, then one can derive simple expressions for the share of top 1%, or top 10% etc.
- For example, suppose that the entire income distribution is given by Pareto with a shape parameter  $\lambda$  ( cdf  $1 - \Gamma y^{-\lambda}$ ).
- Then the top  $q$ th percentile's share of total income can be derived as:

$$\left( \frac{q}{100} \right)^{\frac{\lambda-1}{\lambda}} . \quad (1)$$

- This expression makes it clear that a lower  $\lambda$  corresponds to a “thicker tail” of the Pareto distribution and thus to a greater share of total income being captured by individuals/households at higher percentiles of the distribution. For example, with  $\lambda = 2$ , the top 1%'s share is 10%, and with  $\lambda = 3$ , it is 4%.

## Top Inequality and Pareto (continued)

- To derive this expression, note that the income of people with income greater than some level  $y'$  is

$$\int_{y'}^{\infty} \lambda \Gamma y^{-\lambda-1} \times y \times dy$$

- Then the share of total income accruing to those above the  $q$ th percentile can be written as

$$\frac{\int_{y_q}^{\infty} \lambda \Gamma y^{-\lambda} dy}{\int_{y_{\min}}^{\infty} \lambda \Gamma y^{-\lambda} dy} = \frac{-\frac{\lambda}{\lambda-1} \Gamma y^{-(\lambda-1)} \Big|_{y_q}^{\infty}}{-\frac{\lambda}{\lambda-1} \Gamma y^{-(\lambda-1)} \Big|_{y_{\min}}^{\infty}} = \left( \frac{y_q}{y_{\min}} \right)^{-(\lambda-1)}.$$

The  $q$ th percentile satisfies  $\Gamma y_q^{-\lambda} = q/100$ , i.e.,  $y_q = (100\Gamma/q)^{1/\lambda}$ , and the lower support of the distribution thus satisfies  $y_{\min} = \Gamma^{1/\lambda}$ . Substituting these, we obtain (1).



## Where Might the Pareto Distribution Come From?

- Many stochastic processes lead to the Pareto distribution. The most famous and most convenient is the so-called **Kesten process**, which takes the form

$$y_{t+1} = \gamma_t y_t + z_t,$$

where  $\gamma_t$  and  $z_t$  are independent random variables, with  $\mathbb{E}\gamma_t < \infty$  and  $\mathbb{E}z_t < \infty$ .

- Suppose that  $y$  has a stationary distribution  $\mathbb{G}$  (this is not trivial as we will see, so this assumption makes life much more straightforward).

## Pareto Distribution (continued)

- $\mathbf{G}(y) \equiv 1 - \Pr[\tilde{y} \leq y] = \Gamma y^{-\lambda}$  can be written as

$$\begin{aligned} \Pr[y_{t+1} \geq y] &= \mathbb{E}[\mathbf{1}_{\{y_{t+1} \geq y\}}] \\ &= \mathbb{E}[\mathbf{1}_{\{\gamma_t y_t + z_t \geq y\}}] \\ &= \mathbb{E}[\mathbf{1}_{\{y_t \geq (y - z_t)/\gamma_t\}}], \end{aligned}$$

where  $\mathbf{1}_{\{\mathcal{P}\}}$  is the indicator function for the event  $\mathcal{P}$ .

- Then, by the definition of a stationary distribution  $\mathbf{G}$ , we have

$$\mathbf{G}(y) = \mathbb{E}_{\gamma, z} \left[ \mathbf{G} \left( \frac{y - z}{\gamma} \right) \right].$$

- The solution, which exists by assumption, will be Pareto in the tail:

$$\mathbf{G}(y) = \Gamma y^{-\lambda}$$

for large  $y$ , and moreover,  $\lambda$  will be given by  $\mathbb{E}_{\gamma} \gamma^{\lambda} = 1$ .

## Pareto Distribution (continued)

- To see this, consider the special case where  $z_t = z$ . Then

$$\mathbf{G}(y) = \mathbb{E}_\gamma \left[ \mathbf{G} \left( \frac{y-z}{\gamma} \right) \right].$$

Suppose  $\mathbf{G}(y) = \Gamma y^{-\lambda}$  for large  $y$ . Then, again for large  $y$ ,

$$\Gamma y^{-\lambda} = \mathbb{E}_\gamma \left[ \Gamma (y-z)^{-\lambda} \gamma^\lambda \right], \text{ or}$$

$$\Gamma y^{-\lambda} = \Gamma (y-z)^{-\lambda} \mathbb{E}_\gamma \gamma^\lambda,$$

where the term on the left and the first term on the right are approximately equal for large  $y$ , giving the desired result.

- When  $z$  is random, same reasoning applies for large  $y$ .

## When Does a Limiting Distribution Exist?

- Take the process

$$y_{t+1} = \gamma_t y_t,$$

with  $\gamma_t$  independent and mean one. This clearly does not have a limiting distribution, as the empirical distribution (as a function of time) will keep on expanding.

- Essentially for a limiting distribution to exist, the  $z_t$  term needs to make sure that there aren't too many small observations (in the exact Pareto case,  $y_t$  has to be greater than  $\Gamma$ ).
- One way of achieving this is to have a hard or soft lower **reflecting barrier**.

## When Does a Limiting Distribution Exist? (continued)

- The following is a well-known theorems from stochastic processes.

### Theorem

*Suppose that  $y$  follows the continuous time reflected geometric Brownian motion process*

$$\frac{dy_t}{y_t} = \begin{cases} \gamma dt + \sigma dB_t & \text{if } y_t > y_{\min} \\ \max\{\gamma dt + \sigma dB_t, 0\} & \text{if } y_t \leq y_{\min}, \end{cases}$$

*where  $\gamma < 0$ , then  $y_t$  converges in distribution to the Pareto distribution with shape parameter*

$$\lambda = \frac{1}{1 - y_{\min}/\bar{y}},$$

*where  $\bar{y}$  is its stationary distribution average.*

- This implies that if the lower barrier  $y_{\min}$  is sufficiently small, the exponent is approximately 1.

## Pareto Wage Distribution

- We already seen how superstar phenomenon can generate Pareto distributions.
- In the Gabaix-Landier model, when the value that CEO creates is multiplicative in firm size and CEO skill, when the firm size distribution is Pareto, when skill of CEOs comes from an extreme value distribution, and when  $\alpha\gamma - \beta > 0$ , then the wage of the CEO working in a firm with  $q$ th percentiles size is

$$w(q) = \frac{A^\gamma B \Gamma}{\alpha\gamma - \beta} q^{-(\alpha\gamma - \beta)}.$$

- Assuming that CEOs make up the right tail of income distribution, we can read off from this equation when a Pareto tail might obtain.
- First, suppose that  $\gamma = 0$ , so that there is no positive assortative matching between firms and managers. Then a Pareto distribution only obtains if  $\beta < 0$ . This is the case in which skills have a Pareto distribution, and the wage distribution simply reflects this.

## Pareto Wage Distribution (continued)

- More interestingly, the same equation implies that when  $\gamma > \beta/\alpha$ , this right tail is Pareto, with a shape parameter  $\alpha\gamma - \beta$ .
- In this case, more interestingly, it is the assignment of high-skill managers to large firms (or firms with hire marginal product) that leads to the Pareto distribution (even when  $\beta \geq 0$  the distribution of skills is far from Pareto, in fact has an upper bound).

## Pareto Wage Distribution (continued)

- There might be reasonable (or at least reasonable-looking) ways in which skills might have a Pareto-like distribution.
- Jones and Kim (2014) suggest one way. Suppose that an individual's earning potential (e.g., entrepreneurial skill) grows proportionately, say at the rate  $\mu$ , as long as he remains in that status.
- Then an individual who has been an entrepreneur for  $t$  periods, will have income (today) of

$$y(t) = y(0)e^{\mu t}.$$

- Now consider the fraction of individuals having income greater than some amount  $y$  in the economy. This is simply the fraction of individuals in the economy who have been entrepreneurs for longer than  $t(y)$  periods where

$$t(y) = \frac{1}{\mu} \log \left( \frac{y}{y(0)} \right).$$



## Pareto Wage Distribution (continued)

- Suppose also that individuals leave entrepreneurship permanently (die or retire) at the rate  $\delta$ . Then the fraction of individuals with entrepreneurial experience greater than  $t(y)$  is simply

$$\begin{aligned}\Pr[\text{Experience} > t(y)] &= e^{-\delta t(y)} \\ &= \left(\frac{y}{y(0)}\right)^{-\frac{\delta}{\mu}},\end{aligned}$$

so that experience has a Pareto distribution.

- But since

$$\Pr[\text{Income} > y] = \Pr[\text{Experience} > t(y)],$$

income has a Pareto distribution with shape parameter  $\delta/\mu$ .

## Pareto Wage Distribution (continued)

- This is in fact a general property:
  - If skill has a proportional growth rate and Poisson exit (stoppage), it will have an exponential distribution.
  - If skill is remunerated exponentially (log income being proportional to it), then we end up with a Pareto tail.
- These are potential explanations for a Pareto distribution coming from wage income (and their plausibility will be discussed later). We next turn to the implications of capital income.

## Interest Rate and Inequality in Kaldor-Type Models

- Before working with Pareto distributions, it might be useful to visit the baseline arguments used for drawing a link between  $r - g$  and top inequality, which rely on Kaldor-type models, in which only one group of agents (“capitalists”) save and have capital income, while remaining agents (“workers”) only have labor income.
- Kaldor also assumed that capitalists have no income from labor, though this is not important).
- Consider a continuous-time economy, and denote the capital stock held by capitalists by  $K_C$ .
- The fraction of capitalists in the population is  $m$ .
- Ignore social mobility between capitalists and workers to start with.

## Income Distribution: Capitalists

- Assuming that capital depreciates at the rate  $\delta$  and the interest rate is  $r$ , total income of capitalists is

$$I_C = (r + \delta)K_C.$$

- Following Kaldor (and Piketty), assume a constant saving rate for capitalists, say  $s_C$ , so that

$$\dot{K}_C = s_C I_C - \delta K_C = [s_C(r + \delta) - \delta]K_C.$$

- Then the growth rate of capitalist's income is

$$g_C^I = \frac{\dot{K}_C}{K_C} + \frac{\dot{r}}{r + \delta} = s_C(r + \delta) - \delta + \frac{\dot{r}}{r + \delta}.$$

# Income Distribution: Workers

- Workers' income is

$$\begin{aligned} I_W &= (r + \delta)K_W + wL \\ &= Y - (r + \delta)K_C, \end{aligned}$$

where  $K_W$  is the capital stock held by workers,  $w$  the real wage,  $L$  total employment.

- Then the growth rate of their income is

$$g_W^I = \frac{\frac{\dot{Y}}{Y} - \frac{\dot{K}_C}{K_C} \frac{I_C}{Y} - \frac{\dot{r}}{r+\delta} \frac{I_C}{Y}}{1 - \frac{I_C}{Y}}.$$

## Income Distribution: Comparison

- Denote the fraction of national income accruing to capitalists by  $\phi$  ( $= I_C / Y$ ) (e.g., if capitalists correspond to the richest 1 percent in the population, then  $\phi$  is the top 1 percent share).

- Then

$$g_W^I = \frac{g - [s_C(r + \delta) - \delta]\phi - \frac{\dot{r}}{r + \delta}\phi}{1 - \phi}.$$

- And thus we obtain a condition for divergence between capitalists' and workers' incomes:

$$g_C^I > g_W^I \text{ if and only if } s_C(r + \delta) > g + \delta - \frac{\dot{r}}{r + \delta}.$$

- So  $r - g > 0$  could create a force towards divergence between capitalists' and workers' incomes, but only if it's sufficiently large.
  - In particular, thus saving rates of capitalists as well as changes in interest rates matter.

# Introducing Social Mobility

- More interesting is the impact of social mobility.
- To start with, let us model social mobility in a simple fashion, but we will enrich this soon.
- Suppose that at some flow rate  $\nu$ , a capitalist is hit by a random shock and becomes a worker, inheriting the worker's labor income process and saving rate.
- At this point, he (or she) of course maintains his current income, but from then on his income dynamics follows those of other workers.
- Simultaneously, a worker becomes a capitalist (also at the flow rate  $\nu$ ), keeping the fraction of capitalists in the population constant at  $m \in (0, 1)$ .
- Also set  $\delta = 0$  and assume the interest rate  $r$  is constant for simplicity.

## Social Mobility (continued)

- The dynamics of the total income of capitalists can be written as

$$\dot{I}_C = s_C r I_C - v \left[ \frac{I_C}{m} - \frac{I_W}{1-m} \right]. \quad (2)$$

- The key observation here is that, on average, a capitalist leaving the capitalist class has income  $I_C/m$  (total capitalists' income divided by the measure of capitalists), and a worker entering the capitalist class has, on average, income  $I_W/(1-m)$ .
- Major simplification: social mobility is divorced from the income level of a capitalist or a worker (in practice a worker with income close to that of capitalists is more likely to be upwardly mobile than a poor worker).
- Incorporating this feature will require more complicated analysis as we will see.



## Social Mobility (continued)

- Dividing both sides by  $l_C$ , we have

$$\begin{aligned} g_C^l &= s_C r - v \left[ \frac{1}{m} - \frac{1}{1-m} \frac{l_W}{l_C} \right] \\ &= s_C r - v \left[ \frac{1}{m} - \frac{1}{1-m} \frac{1-\phi}{\phi} \right]. \end{aligned}$$

- With a similar reasoning, the growth rate of the total income of workers is

$$g_W^l = \frac{g - s_C r \phi}{1 - \phi} + v_W \left[ \frac{1}{m} \frac{\phi}{1 - \phi} - \frac{1}{1 - m} \right].$$

## Social Mobility (continued)

- Combining these expressions and rearranging terms, we can write

$$g_C^l > g_W^l \text{ if and only if } s_C r - g > v \frac{\phi - m}{\phi m (1 - m)}.$$

- The term on the right-hand side is strictly positive since  $\phi > m$  (i.e., the share of top 1 percent in national income is greater than 1%).
- Thus even when  $s_C r - g > 0$ , inequality between capitalists and workers may decline if there is sufficient social mobility. Whether it does will depend on the extent of social mobility. In fact, the quantitative implications of social mobility could be quite substantial as we next illustrate.

## How Important Is Social Mobility?

- From Chetty, Hendren, Kline and Saez's data, the likelihood that a child with parents in the top 1 percent will be in the top 1 percent is 9.6%.
- If we take the gap between generations to be about 30 years, this implies an annual rate of exiting the top 1 percent approximately equal to 0.075 (7.5%).
- This number corresponds to  $\nu/m$  in our model (the probability that a given capitalist is hit by a shock and becomes a worker), so we take  $\nu = 0.00075$ .
- There are many reasons why this may be an overestimate, and some that will lead to underestimation.

## How Important Is Social Mobility? (continued)

- Using the top 1 percent's share as 20%, we can compute that the right-hand side of this relationship is approximately 0.072 (7.2%).
- This implies that for the left-hand side to exceed the right-hand side, the interest rate would have to be very high. For example, with a saving rate of 50% and a growth rate of 1%, we would need the interest rate to be greater than 15%.

## How Important Is Social Mobility? (continued)

- Alternatively, if we use the top 10 percent so as to reduce exits that may be caused by measurement error, the equivalent number from Chetty, Hendren, Kline (2014) is 26%, implying an annual exit rate equal to 4.4%.
- Using a share of 45% of income for the top 10 percent, the right-and side is now 0.038, again making it very difficult for realistic values of  $r - g$  to create a natural and powerful force for the top inequality to increase. For example, using again a saving rate of 50% and a growth rate of 1%, the interest rate would need to be over 8.5%.
- So this exercise suggests that even modest amounts of social mobility can nullify significant forces towards divergence of incomes between the top and the middle (or the bottom) of income distribution.

# Towards Pareto Distributions of Capital Income

- We will now explicitly model the incomes of all agents in the economy resulting from stochastic accumulation of capital as well as other sources of stochasticity incomes.
- To give the basic idea, consider an economy consisting of a continuum of measure 1 of heterogeneous individuals.
- Suppose that each individual  $i$  is infinitely lived and consumes a constant fraction  $\beta$  of her wealth,  $A_{it}$ .
- She has a stochastic (possibly serially correlated) labor income  $Z_{it}$  (with  $\mathbb{E}Z_{it} \in (0, \infty)$  and finite variance), and has a stochastic rate of return equal to  $r + \varepsilon_{it}$ , where  $\varepsilon_{it}$  is a stochastic, return term that is also possibly serially correlated (with the unconditional mean  $\mathbb{E}\varepsilon_{it}$  equal to zero as a normalization).

# Asset Dynamics

- Therefore, the law of motion of the assets of individual  $i$  is

$$A_{it+1} = (1 + r - \beta + \varepsilon_{it})A_{it} + Z_{it}.$$

- Dividing both sides of this equation by GDP (also average income per capita),  $Y_t$ , we obtain

$$\tilde{a}_{it+1} = \frac{1 + r - \beta + \varepsilon_{it}}{1 + g} \tilde{a}_{it} + \tilde{z}_{it},$$

where  $\tilde{a}_{it} \equiv A_{it} / Y_t$  and  $\tilde{z}_{it} \equiv Z_{it} / Y_t$ .

- Suppose next that  $\tilde{a}_{it}$  converges to a stationary distribution (we verify this below).
- Then let  $\mathbb{E}\tilde{a}$  be the average (expected) value of  $\tilde{a}_{it}$  in the stationary distribution.

# Pareto Tail

- Now divide both sides of the previous equation by  $\mathbb{E}\tilde{a}$  to obtain

$$a_{it+1} = \frac{1 + r - \beta + \varepsilon_{it}}{1 + g} a_{it} + z_{it}, \quad (3)$$

where  $a_{it} \equiv \tilde{a}_{it} / \mathbb{E}\tilde{a}$  and  $z_{it} \equiv \tilde{z}_{it} / \mathbb{E}\tilde{a}$ , and of course  $\mathbb{E}a_{it+1} = \mathbb{E}a_{it} = 1$  in the stationary distribution. This also implies that  $\mathbb{E}z_{it} \in (0, 1)$ .

- Equation (3) is an example of a Kesten process introduced above.
- Because  $\frac{1+r-\beta}{1+g} < 1$ , (3) will converge to a stationary distribution with a Pareto tail as we have seen, and thus the right tail of the distribution, corresponding to  $a \geq \bar{a}$  for  $\bar{a}$  sufficiently large, will be approximated by  $1 - \Gamma a^{-\lambda}$  for some endogenously-determined Pareto shape parameter  $\lambda \geq 0$ .



## Heuristic Derivation

- The intuition is the same as the one provided above. Let us focus on the case in which  $z$  and  $\varepsilon$  are iid.
- Define the counter-cumulative density function of (normalized) wealth in this economy to be  $\mathbf{G}(a) \equiv 1 - \Pr[a_{it} \leq a]$ . Then

$$\begin{aligned} \Pr[a_{it+1} \geq a] &= \mathbb{E} \left[ \mathbf{1}_{\{\gamma a_{it} + z \geq a\}} \right], \\ &= \mathbb{E} \left[ \mathbf{1}_{\{a_{it} \geq (a-z)/\gamma\}} \right], \end{aligned}$$

where  $\mathbf{1}_{\{\mathcal{P}\}}$  is the indicator function for the event  $\mathcal{P}$ , and

$$\gamma \equiv \frac{1 + r + \varepsilon - \beta}{1 + g}$$

and the indices for  $z$  and  $\gamma$  have been suppressed.

- Then, by the definition of a stationary distribution  $\mathbf{G}$ , we have

$$\mathbf{G}(a) = \mathbb{E} \left[ \mathbf{G} \left( \frac{a-z}{\gamma} \right) \right].$$

## Heuristic Derivation (continued)

- Now let us conjecture a Pareto tail with shape parameter  $\lambda$ , i.e.,  $G(a) = \Gamma a^{-\lambda}$  for large  $a$ . Then for large  $a$ , we have

$$\Gamma a^{-\lambda} = \Gamma \mathbb{E}(a - z)^{-\lambda} \left[ \gamma^\lambda \right],$$

or

$$1 = \mathbb{E} \left( \frac{a - z}{a} \right)^{-\lambda} \left[ \gamma^\lambda \right].$$

- Since  $\mathbb{E}z < \infty$  and has finite variance, we can write  $\lim_{a \rightarrow \infty} \mathbb{E} \left( \frac{a - z}{a} \right)^{-\lambda} = 1$ , which confirms the conjecture and defines  $\lambda$  as the positive solution to

$$\mathbb{E} \left[ \gamma^\lambda \right] = 1. \quad (4)$$

- This equation also explains why  $\mathbb{E}\gamma = \frac{1+r-\beta}{1+g} < 1$  is necessary for convergence to a stationary distribution (as otherwise the wealth distribution would diverge).

# Interest Rates and Pareto

- What is the relationship between  $r - g$  (or  $r - g - \beta$ ) and  $\lambda$ ?
- Do higher interest rates or lower growth rates lead to a rightward shift in the stochastic distribution of  $\gamma$  thus reducing  $\lambda$  and leading to greater scope inequality?
- The answer is in general ambiguous but in some special cases, there is such a relationship.

## Interest Rates and Pareto (continued)

- Suppose again that  $\varepsilon_{it}$  and  $z_{it}$  are iid, and we have

$$a_{it+1} = \gamma_{it} a_{it} + z_{it}.$$

- Taking expectations on both sides, using the fact that  $\gamma_{it}$  is iid and that in the stationary distribution  $\mathbb{E}a_{it+1} = \mathbb{E}a_{it} = 1$ , we have

$$\mathbb{E}\gamma = 1 - \bar{z},$$

where  $\bar{z} = \mathbb{E}z_{it} \in (0, 1)$ , as noted above.

- This equation also implies that  $\mathbb{E}\gamma \in (0, 1)$ .
- To determine the relationship between  $r - g$  and  $\lambda$ , we consider two special cases.

## Interest Rates and Pareto (continued)

- *First* suppose that  $\gamma$  (or  $\varepsilon$ ) is log normally distributed.
- In particular, suppose that  $\ln \gamma$  has a normal distribution with mean  $\ln(1 - \bar{z}) - \sigma^2/2$  and variance  $\sigma^2 > 0$  (so that  $\mathbb{E}\gamma = 1 - \bar{z}$ ).
- Then we have

$$\mathbb{E}[\gamma^\lambda] = \mathbb{E}[e^{\lambda \ln \gamma}],$$

which is simply the moment generating function of the normally distributed random variable  $\ln \gamma$ , which can be written as

$$\mathbb{E}[e^{\lambda \ln \gamma}] = e^{\lambda[\ln(1-\bar{z})-\sigma^2/2]+\lambda^2\sigma^2/2}.$$

## Interest Rates and Pareto (continued)

- Then the definition of  $\lambda$ ,  $\mathbb{E}[\gamma^\lambda] = 1$ , implies that

$$\lambda[\ln(1 - \bar{z}) - \sigma^2/2] + \lambda^2\sigma^2/2 = 0,$$

which has two roots,  $\lambda = 0$  (which is inadmissible for the stationary distribution), and the relevant one,

$$\lambda = 1 - \frac{\ln(1 - \bar{z})}{\sigma^2/2} > 1.$$

- Finally, for small values of  $r - g - \beta < 0$ , we can write

$$\gamma \approx 1 + r - g - \beta + \varepsilon,$$

and thus from the relationship that  $\mathbb{E}\gamma = 1 - \bar{z}$ , we have that  $\bar{z} = -(r - g - \beta) > 0$ , so that

$$\lambda \approx 1 - \frac{\ln(1 + r - g - \beta)}{\sigma^2/2}.$$

## Interest Rates and Pareto (continued)

- Therefore  $\lambda$  is decreasing in  $r - g - \beta$ , thus implying that higher  $r - g$  and lower marginal propensity to consume out of wealth,  $\beta$ , lead to greater top inequality.
- The same conclusion follows without the approximation  $\gamma \approx 1 + r - g - \beta + \varepsilon$ .
- In this case, we would simply have

$$\lambda = 1 - \frac{\ln\left(1 + \frac{1+r-\beta}{1+g}\right)}{\sigma^2/2},$$

which yields the same comparative statics.

## Interest Rates and Pareto (continued)

- *Second*, a similar relationship can be derived even when  $\gamma$  is not log normally distributed, but only when  $\bar{z}$  is small (and we will see why this may not be very attractive in the context of the stationary distribution of wealth).
- Let us start by taking a first-order Taylor expansion of  $\mathbb{E}[\gamma^\lambda] = 1$  with respect to  $\lambda$  around  $\lambda = 1$  (which also corresponds to making  $\bar{z}$  lie close to zero).
- In particular, differentiating within the expectation operator, we have

$$\mathbb{E}[\gamma + \gamma \ln \gamma(\lambda - 1)] \approx 1,$$

where this approximation requires  $\lambda$  to be close to 1.

- (Formally, we have  $\mathbb{E}[\gamma + \gamma \ln \gamma(\lambda - 1) + o(\lambda)] = 1$ ).



## Interest Rates and Pareto (continued)

- Then again exploiting the fact that  $\mathbb{E}\gamma = 1 - \bar{z}$ , we have

$$\lambda \approx 1 + \frac{\bar{z}}{\mathbb{E}[\gamma \ln \gamma]} > 1.$$

(where the fact that  $\mathbb{E}[\gamma \ln \gamma] > 0$  follows from the fact that  $\bar{z}$  is close to zero).

- More specifically, note that  $\gamma \ln \gamma$  is a convex function and apply Jensen's inequality,

$$\mathbb{E}[\gamma \ln \gamma] > \mathbb{E}\gamma \cdot \ln \mathbb{E}\gamma = (1 - \bar{z}) \ln(1 - \bar{z}).$$

- For  $\bar{z}$  close enough to zero,  $(1 - \bar{z}) \ln(1 - \bar{z}) \approx -\bar{z}$ , and thus  $\mathbb{E}[\gamma \ln \gamma] > 0$ .
- This expression clarifies why  $\lambda$  is close to 1 when  $\bar{z}$  is close to 0.

## Interest Rates and Pareto (continued)

- Moreover, note that the derivative of  $\gamma \ln \gamma$  is  $1 + \ln \gamma$ .
- For  $\bar{z}$  small,  $\ln \gamma > -1$  with sufficiently high probability, and thus  $\mathbb{E}[\gamma \ln \gamma]$  increases as  $\gamma$  shifts to the right (in the sense of first-order stochastic dominance).
- Therefore, when  $\lambda$  is close to 1 or equivalently when  $\bar{z}$  is close to 0, a higher  $r - g - \beta$  increases  $\mathbb{E}[\gamma \ln \gamma]$  and reduces the shape parameter  $\lambda$ , raising top inequality.
- However, it should also be noted that this case is much less relevant for stationary wealth distributions which have Pareto tails much greater than 1.

## Pareto and Social Mobility

- A derivations similar to the one provided here applies also with some limited forms of social mobility.
- In particular, Benhabib, Bisin and Zhu (2011) extend the result on the Pareto-tail of the wealth distribution to a setup with finitely-lived agents with bequest motives.
- In this case, the tail of the distribution is in part driven by which individuals have been accumulating for the longest time.
- They also derive the consumption choices from optimization decisions, consider the equilibrium determination of the interest rate, and confirm the results derived heuristically here.
- In addition, they show that one type of social mobility—related to the serial correlation of  $\varepsilon$ , thus making financial returns less correlated for a household over time—tends to make the tail less thick, hence reducing top inequality.

# Critiques of Theories of Redistribution of Capital Income

- There are several reasons why these models may not be entirely satisfactory as models of top inequality, however.
- First, to the extent that very rich individuals have diversified portfolios, variability in rates of returns as a driver of the tail of the distribution may not be the most dominant factor.
- Second, the structure of these models implies that labor income plays no role in the tail of the stationary wealth distribution, but this may be at odds with the importance of wages and salaries and “business income” in the top 1 percent or even top 0.1 percent share of the national income (Piketty and Saez, 2003).

# Critiques of Pareto Distribution Theories

- Third and relatedly, these models having no or little role for entrepreneurship, which is one of the important aspects of the interplay between labor and capital income.
- Fourth, these models do not feature social mobility (except the limited type of social mobility related to the correlation of financial returns considered in Benhabib, Bisin and Zhu, 2011).
- Finally, in more realistic versions such as Benhabib, Bisin and Zhu (2011) and Jones and Kim (2014), a key determinant of the extent of top inequality turns out to be the age or some other characteristic of the household which determines how long the household has been accumulating (e.g., experience above).
  - But this is also at odds with the salient patterns of the tail of the income and wealth distribution in the United States, whereby successful entrepreneurs or professionals are more likely to be represented at this tale than individuals or households that have been accumulating capital for a long while.