Matching log-normal distribution with Pareto distribution under Zipf's law

July 30, 2024

1 Objective

The goal is to match the log-normal distribution with the Pareto distribution under Zipf's law, focusing on their mean, median, and the Herfindahl index, and to evaluate how well the log-normal distribution represents real-world data compared to the Pareto distribution.

To begin with, simulate the Herfindahl for Pareto distribution. The formulas are from Gabaix (2011).

$$h = \Big[\sum_{i=1}^{N} (\frac{S_{it}}{Y_t})^2\Big]^{0.5}$$

$$Y_t = \sum_{i=1}^{N} S_{it}$$

Except for $\zeta = 1$, also assume that ζ is 1.059 following Axtell(2001).

Moreover, this experiment generate Pareto distribution by inverse transform sampling. Suppose that the random variable X is Pareto distributed and its cumulative distribution function is

$$F_X(x) = \begin{cases} (\frac{x_m}{x})^{\zeta}, & \text{if } x \geq x_m \\ 0, & x < x_m \end{cases}$$

If we let $F_X(x) = u$, then random variable $U = u$ is uniform distributed, $U \sim \mathcal{U}[0, 1)$. Then, by the relationship

$$P(X \leq x) = P(F_X^{-1}(U) \leq x) = P(U \leq F_X(x)) = F_U(F_X(x)) = F_X(x),$$

we can obtain the Pareto random variable X by taking the inverse of uniform distributed U. More specifically, $x = F_X^{-1}(u) = x_m(1-u)^{-1/\zeta}$.

2 Match the mean and median provided that the log-normal distribution is cut off under some $\bar{x}$.

Given the Pareto distribution with tail index $\zeta > 1$, the density function is

$$f_P(x) = \begin{cases} \frac{\zeta \bar{x}^{\zeta}}{x^{\zeta+1}}, & \text{if } x \geq \bar{x} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Then, the corresponding mean and median are $\frac{\zeta \bar{x}}{\zeta - 1}$ and $\bar{x}2^{1/\zeta}$, respectively. The mean can be obtained by $\mathbb{E}(x) = \int x f_P(x)dx$ and the median is the root for $\frac{1}{2} = F_P(x)$ where $F_P(x) = 1 - (\frac{x}{\bar{x}})^{-\zeta}$ is the cumulative density function for $x \geq \bar{x}$.

If the log-normal distribution is $\ln(x) \sim \mathcal{N}(\mu, \sigma)$, then we can use the same approaches to find that its mean and median which are $\exp(\mu + \frac{\sigma^2}{2})$ and $\exp(\mu)$, respectively. In addition, the distribution for log-normal is

$$g(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\{-\frac{(\ln x - \mu)^2}{2\sigma^2}\}$$

Also, let $G(x)$ be the cumulative distribution function of log-normal which is also equal to the CDF of the standard normal $\Phi(\frac{\ln x - \mu}{\sigma})$.

$$G(x) = \Phi(\frac{\ln x - \mu}{\sigma})$$

Next, I rescale the distribution such that it is zero for $x < \bar{x}$ which has the same cutoff as Pareto distribution. After setting the minimum value $\bar{x}$ and re-scale the probability function such that $P(x \geq \bar{x}) = 1$, the distribution is

$$\hat{g}(x) = \begin{cases} \frac{g(x)}{1-G(\bar{x})}, & \text{if } x \geq \bar{x} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

Since $\hat{g}(x)$ is a scaled $g(x)$, the mean for density $\hat{g}$ is $\frac{1}{1-G(\bar{x})} \exp(\mu + \frac{\sigma^2}{2})$ which is a product of scaling constant and the mean of log-normal. For the median, using its CDF $\hat{G}(x) = \frac{G(x)-G(\bar{x})}{1-G(\bar{x})}$, the median should pin down the bellowing equation.

$$\frac{1}{2} = \frac{G(x) - G(\bar{x})}{1 - G(\bar{x})}$$

Thus, the median is

$$median = G^{-1}\big[\frac{1}{2}(1 - G(\bar{x})) + G(\bar{x})\big]$$

Now, we can try to match the scaled log-normal with Pareto distribution under Zipf's law. Since the first moment is infinite when $\zeta = 1$ for Pareto distribution, let ζ be very close to one but not one. Following the assumption for simulation in Gabaix (2011), assume $\bar{x} = 1$. Then, fixing the mean and median of Pareto distribution, if the mean and median for scaled log-normal distribution are equal to those of Pareto distribution, we have

$$\begin{aligned} &\text{(mean)} \quad \frac{1}{1 - G(\bar{x})} \exp(\mu + \frac{\sigma^2}{2}) = \frac{\zeta}{\zeta - 1} \\ &\text{(median)} \quad G^{-1}\big[\frac{1}{2}(1 + G(\bar{x}))\big] = 2^{1/\zeta} \end{aligned}$$

The median equation can also be written as

$$G^{-1}[2G(2^{1/\zeta}) - 1] = \bar{x}$$

The CDF $G(x)$ is determined by the shape parameters (μ, σ). To match the distribution, we need to solve the above equations and find the corresponding pair of root (μ, σ) which decides the shape of log-normal. Since $G(.)$ is a function of $\Phi(.)$, the median equation contains the inverse of $\Phi(.)$ function which is complicated to solve analytically. Therefore, I try to solve them numerically and first plot the equations (LHS - RHS) with respect to parameters (μ, σ) to find the potential roots.

Looking at the graphs, the roots are on the line with value 0.0. Since there is no intersection between the lines of root on these tow graphs, there is no pair of (μ, σ) which can pin down the roots for the mean and median equations at the same time.

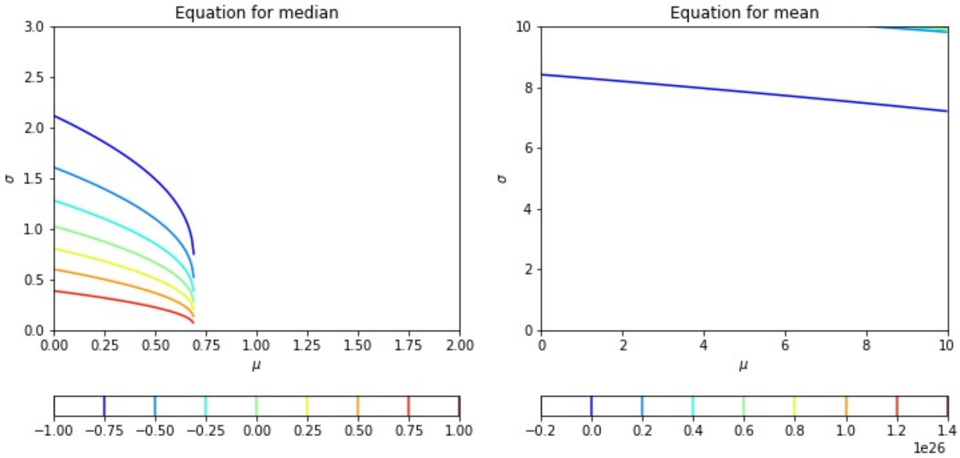

Figure 1: (LHS-RHS) of median and mean equations.

3 Match the mean and median without the cutoff for log-normal

Without scaling, the lognormal has the mean $\exp(\mu+\frac{\sigma^2}{2})$ and median $\exp(\mu)$. Then, if the mean is equal to the mean of Pareto distribution, we have

$$\frac{\zeta\bar{x}}{\zeta-1} = \exp(\mu + \frac{\sigma^2}{2})$$

Similarly for median,

$$\bar{x}2^{1/\zeta} = \exp(\mu)$$

Given ζ is close one and $\bar{x} = 1$, we can solve the equations for μ and σ. Therefore, $\mu = \ln(\bar{x}2^{1/\zeta})$ where $\bar{x} = 1$ and $\sigma = (2\ln(\frac{\zeta}{\zeta-1}) - \frac{2}{\zeta}\ln 2)^{1/2}$

3.1 Conclusion

$(\mu, \sigma) = (0.69315, 8.40839)$ is shape for matched log-normal under Zipf's law(ζ is close to one but not one). In this case, the Herfindahl is 59.42

When firm's growth rate is 12%, the deviation of growth rate $\sigma_{GDP} = h\sigma$ is around 7.2% for log-normal which is also higher than 1.4% of Pareto.

4 match

Match the mean and median provided that the log-normal distribution is cut off under some $\bar{x}$. Given the Pareto distribution with tail index $\zeta > 1$, the density function is

$$f_P(x) = \begin{cases} \frac{\zeta\bar{x}^\zeta}{x^{\zeta+1}}, & \text{if } x \geq \bar{x} \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Then, the corresponding mean and median are $\frac{\zeta\bar{x}}{\zeta-1}$ and $\bar{x}2^{1/\zeta}$, respectively. The mean can be obtained by $\mathbb{E}(x) = \int x f_P(x)dx$ and the median is the root for $\frac{1}{2} = F_P(x)$ where $F_P(x) = 1 - (\frac{x}{\bar{x}})^{-\zeta}$ is the cumulative density function for $x \geq \bar{x}$.

If the log-normal distribution is $\ln(x) \sim \mathcal{N}(\mu, \sigma)$, then we can use the same approaches to find that its mean and median. The distribution for log-normal is

$$g(x) = \frac{1}{\sigma x\sqrt{2\pi}} \exp\{-\frac{(\ln x - \mu)^2}{2\sigma^2}\}.$$

Thus, from $\mathbb{E}(x) = \int g(x)dx$, the mean is $\exp(\mu + \frac{\sigma^2}{2})$. In addition, the CDF is $G(x) = \Phi(\frac{\ln(x)-\mu}{\sigma})$ where $\Phi(.)$ is the CDF for standard normal. Then, using $\frac{1}{2} = G(x)$, we know that $\frac{\ln(x)-\mu}{\sigma} = 0$ and so the median is $\exp(\mu)$.

Now, we can try to match the log-normal with Pareto distribution under Zipf's law. Since the first moment is infinite when $\zeta = 1$ for Pareto distribution, let ζ be very close to one but not one. Following the assumption for simulation in (Gabaix, 2011), assume $\bar{x} = 1$. Then, fixing the mean and median of Pareto distribution (or fixing ζ and $\bar{x}$), if the mean and median for log-normal distribution are equal to those of Pareto distribution, we have

$$\begin{aligned} \text{mean} \quad & \frac{\zeta\bar{x}}{\zeta - 1} = \exp(\mu + \frac{\sigma^2}{2}) \\ \text{median} \quad & \bar{x}2^{1/\zeta} = \exp(\mu) \end{aligned}$$

Solve the equation of median given ζ and $\bar{x} = 1$, we have $\mu = \ln(2^{1/\zeta})$. Plug this into the mean equation, we can get $\sigma = (2\ln(\frac{\zeta}{\zeta-1}) - \frac{2}{\zeta}\ln 2)^{1/2}$.

This research simulate the Herfindahls and the standard deviations for GDP growth for both Pareto distribution and log-normal distribution for firm sizes under Zipf's law. The definition of Herfindahl, the choose of Pareto index ζ and the formulas for variance follow Gabaix (2011). In addition, the simulation method is Monte Carlo, the Pareto random variables are generated by the inverse transform sampling, and the parameters for log-normal are determined such that the corresponding mean and median are equal those of Pareto distribution.

In the beginning, the result of Herfindahl in section 2.3 of Gabaix is confirmed. In this simulation, the Herfindahl is 11.53% under $\zeta = 1.059$ which is close to 12% of Gabaix. Moreover, the the simulated volatilities for GDP growth are too small and not in the same magnitude for both $\zeta = 1.059$ and $\zeta = 1.5$. This also confirms the proposition one of Gabaix. This also confirms the proposition one of Gabaix, which shows that the GDP volatility decays in the rate of $\frac{1}{N^{1/2}}$ when the firm sizes have finite variance.

When firm's growth rate is 12%, the deviation of growth rate $\sigma_{GDP} = h\sigma$ is around 7.2% for log-normal which is also higher than 1.36% of Pareto (11.3% $\times$ 12%). Since the empirically measured macroeconomic fluctuation is around 1%, the log-normal distribution is not reasonable under this matching. In conclusion, the underlying distribution for firm sizes is more like Pareto distribution rather than log-normal distribution in this experiment.

References

Gabaix, X. (2011). The granular origins of aggregate fluctuations. 79(3):733–772.