Derivation of the distribution from extended Gibrat’s law

*Atushi Ishikawa
Kanazawa Gakuin University, Kanazawa 920-1392, Japan

Abstract

Employing profits data of Japanese companies in 2002 and 2003, we identify the non-Gibrat’s law which holds in the middle profits region. From the law of detailed balance in all regions, Gibrat’s law in the high region and the non-Gibrat’s law in the middle region, we kinematically derive the profits distribution function in the high and middle range uniformly. The distribution function accurately fits with empirical data without any fitting parameter.

PACS code: 04.60.Nc
Keywords: Econophysics; Pareto law; Gibrat law; Detailed balance

1 Introduction

The distribution of wealth, income or company size is one of important issues not only in economics but also in econophysics [1]. In these distributions, a cumulative number \( N(> x) \) obeys a power-law for \( x \) which is larger than a certain threshold \( x_0 \):

\[
N(> x) \propto x^{-\mu} \quad \text{for} \quad x > x_0 .
\]  

This power-law and the exponent \( \mu \) are called Pareto’s law and Pareto index, respectively [2] \(^1\). Here \( x \) is wealth, income, profits, assets, sales, the number of employees and etc.

The study of power-law distributions in the high range is quite significant. Because a large part of the total wealth, income or profits is occupied by persons or companies in the high region, although the number of them is a few percent. They have the possibility to influence economics. Power-law distributions are observed in fractal systems which have self-similarity. This means that there is self-similarity in economic systems. The power-law distribution in the high region is well investigated by using various models in econophysics.

Recently, Fujiwara et al. [5] find that Pareto’s law (and the reflection law) can be derived kinematically form the law of detailed balance and Gibrat’s law [6] which are observed in high

\(^1\) Recently Pareto’s law is checked with high accuracy (See [3, 4] for instance).
region $x > x_0$. In the proof, they assume no model and only use these two underlying laws in empirical data.

The detailed balance is time-reversal symmetry:

$$ P_{12}(x_1, x_2) = P_{12}(x_2, x_1). \quad (2) $$

Here $x_1$ and $x_2$ are two successive incomes, profits, assets, sales, etc, and $P_{12}(x_1, x_2)$ is a joint probability distribution function (pdf). Gibrat’s law states that the conditional probability distribution of growth rate $Q(R|x_1)$ is independent of the initial value $x_1$:

$$ Q(R|x_1) = Q(R). \quad (3) $$

Here growth rate $R$ is defined as the ratio $R = x_2/x_1$ and $Q(R|x_1)$ is defined by using the pdf $P(x_1)$ and the joint pdf $P_{1R}(x_1, R)$ as

$$ Q(R|x_1) = \frac{P_{1R}(x_1, R)}{P(x_1)}. \quad (4) $$

In Ref. [7], by using profits data of Japanese companies in 2002 ($x_1$) and 2003 ($x_2$), it is confirmed that the law of detailed balance (2) holds in all regions $x_1 > 0$ and $x_2 > 0$ (Fig. 1). We also find that Gibrat’s law still holds in the extended region $x_1 > x_0$ and $x_2 > 0$. From these observations, we have shown that Pareto index is also induced from the growth rate distribution empirically.

On the other hand, it is also well known that the power-law is not observed below the threshold $x_0$ [6, 8]. For instance, we show the profits distributions of Japanese companies in 2002 and 2003 (Fig. 2). We find that Pareto’s law holds in the high profits region and it fails in the middle one. The study of distributions in the middle region is as important as the study of power-law those. Because a large number of persons or companies is included in the middle region. Furthermore, it is interesting to study the breaking of fractal.

In order to obtain the distribution in the middle region, we examine data below the threshold $x_0$. It is reported that Gibrat’s law is valid only in the high region ([9] for instance). The recent works about the breakdown of Gibrat’s law is done by Stanley’s group [10]. Aoyama et al. also report that Gibrat’s law does not hold in the middle region by using data of Japanese companies [11].

In the analysis of Gibrat’s law in Ref. [7], we concentrate our attention to the region $x_1 > x_0$ and $x_2 > 0$. In this paper, we examine the growth rate distribution in all regions $x_1 > 0$ and $x_2 > 0$ and identify the law which is the extension of Gibrat’s law 2. We employ profits data of Japanese companies in 2002 and 2003 which are available on the database “CD Eyes” published by Tokyo Shoko Research, Ltd. [12].

By using the extended Gibrat’s law, we derive the distribution function in the high and middle profits region under the law of detailed balance. It explains empirical data with high accuracy. Notice that the distribution function has no fitting parameter. The parameters of the function are already decided in the extended Gibrat’s law.

---

2 In this paper, we call the law in the region $x_1 < x_0$ and $x_2 > 0$ non-Gibrat’s law.
2 Gibrat’s law and non-Gibrat’s law

In this section, we reconfirm Gibrat’s law in the high profits region and identify non-Gibrat’s law in the middle one. We divide the range of $x_1$ into logarithmically equal bins as $x_1 \in 4 \times 10^{2+0.2(n-1)}, 10^{2+0.2n}$ thousand yen with $n = 1, 2, \cdots, 15$. In Fig. 3, 4 and 5, the probability densities for $r$ are expressed in the case of $n = 1, \cdots, 5$, $n = 6, \cdots, 10$ and $n = 11, \cdots, 15$, respectively. The number of the companies in Fig. 3, 4 and 5 is “35,513”, “64,205” and “25,189”, respectively. Here we use the log profits growth rate $r = \log_{10} R$. The probability density for $r$ defined by the following equation:

$$\log_{10} q(r| x_1) = \log_{10} q(r| x_1) + \log_{10} (\ln 10) = \log_{10} q(r| x_1).$$

From Fig. 3, 4 and 5, we express the relation between $r$ and $q(r| x_1)$ as follows:

$$\log_{10} q(r| x_1) = c(x_1) - t_+(x_1) r \quad \text{for } r > 0,$$

$$\log_{10} q(r| x_1) = c(x_1) + t_-(x_1) r \quad \text{for } r < 0. \quad (6, 7)$$

By the use of Eq. (5), these relations are rewritten in terms of $R$ as

$$Q(R| x_1) = d(x_1) R^{-t_+(x_1) - 1} \quad \text{for } R > 1,$$

$$Q(R| x_1) = d(x_1) R^{t_-(x_1) - 1} \quad \text{for } R < 1,$$

with $d(x_1) = 10^{c(x_1)/\ln 10}$.

By applying the expressions (6) and (7) to data in Fig. 3, 4 and 5, the relation between $x_1$ and $t_\pm(x_1)$ is obtained (Fig. 6). In Fig. 6, $t_\pm$ hardly responds to $x_1$ for $n = 13, 14, 15$. This means that Gibrat’s law holds in the high profits region. On the other hand, $t_+$ ($t_-$) increases (decreases) linearly with $\log_{10} x_1$ for $n = 4, 5, \cdots, 11$. In the middle profits region, not Gibrat’s law but the other law (non-Gibrat’s law) holds as follows 4:

$$t_+(x_1) = t_+(x_0) - \alpha_+ \ln \frac{x_1}{x_0}, \quad \quad (10)$$

$$t_-(x_1) = t_-(x_0) + \alpha_- \ln \frac{x_1}{x_0}. \quad \quad (11)$$

In Fig. 6, $\alpha_+$ and $\alpha_-$ are estimated as

$$\alpha_+ \sim \alpha_- \sim 0 \quad \text{for } x_1 > x_0,$$

$$\alpha_+ \sim \alpha_- \sim 0.27 \quad \text{for } x_{\min} < x_1 < x_0,$$

where $x_0 = 4 \times 10^{2+0.2(13-1)} \sim 10^5$ thousand yen (= 100 million yen) and $x_{\min} = 4 \times 10^{2+0.2(4-1)} \sim 1,600$ thousand yen. In this paper we call the combination of Gibrat’s law ((10), (11) and (12)) and non-Gibrat’s law ((10), (11) and (13)) extended Gibrat’s law.

In Ref. [7], we only consider the case for $n = 11, 12, \cdots, 15$.

4In section 5, we examine another type of non-Gibrat’s law.
3 Profits distribution in the high and middle region

In Refs. [7], Pareto’s law (1) and the Pareto index $\mu$ can be derived from the detailed balance (2) and Gibrat’s law (3) in the high profits region. In this section, we derive profits distribution not only in the high profits region but also in the middle one by using the detailed balance and the extended Gibrat’s law.

Due to the relation of
\[ P_{12}(x_1, x_2)dx_1dx_2 = P_{1R}(x_1, R)dx_1dR \]
under the change of variables from $(x_1, x_2)$ to $(x_1, R)$, these two joint pdfs are related to each other,
\[ P_{1R}(x_1, R) = x_1P_{12}(x_1, x_2). \]

By the use of this relation, the detailed balance (2) is rewritten in terms of $P_{1R}(x_1, R)$ as follows:
\[ P_{1R}(x_1, R) = R^{-1}P_{1R}(x_2, R^{-1}). \]

Substituting the joint pdf $P_{1R}(x_1, R)$ for the conditional probability $Q(R|x_1)$ defined in Eq. (4), the detailed balance is expressed as
\[ \frac{P(x_1)}{P(x_2)} = \frac{1}{R} \frac{Q(R^{-1}|x_2)}{Q(R|x_1)}. \]

In the preceding section, the conditional probability $Q(R|x_1)$ is identified as (8) or (9). Under the change of variables $x_1 \leftrightarrow x_2$, the conditional probability $Q(R^{-1}|x_2)$ is expressed as
\[ Q(R^{-1}|x_2) = d(x_2) R^{-t_+(x_2)-1} \quad \text{for } R < 1, \]
\[ Q(R^{-1}|x_2) = d(x_2) R^{t_-(x_2)-1} \quad \text{for } R > 1. \]

From the detailed balance and the extended Gibrat’s law, one finds the following:
\[ \frac{P(x_1)}{P(x_2)} = R^{t_+(x_1)-t_-(x_2)+1} \]
\[ = R^{t_+(x_0)-t_-(x_0)+1+\alpha_+\ln \frac{x_0}{x_0} + \alpha_-\ln \frac{x_0}{x_0}} \]
for $R > 1$. Here we assume that the $x$ dependence of $d(x)$ is negligible in the derivation, the validity of which should be checked against the results. By expanding Eq. (20) around $R = 1$, the following differential equation is obtained
\[ \left[ +t_+(x_0) - t_-(x_0) + 1 + (\alpha_+ + \alpha_-) \ln \frac{x}{x_0} \right] P(x) + x P'(x) = 0, \]
where $x$ denotes $x_1$. The same differential equation is obtained for $R < 1$. The solution is given by
\[ P(x) = C x^{-(\mu+1)} e^{-\alpha \ln^2 \frac{x}{x_0}} \]
with $\alpha = (\alpha_+ + \alpha_-)/2$. Here we use the relation $t_+ (x_0) - t_- (x_0) = \mu$ which is confirmed in Ref. [7].
4 Data fitting

In the previous section, we derive the profits distribution function (22) in the high and middle profits region from the detailed balance and the extended Gibrat’s law. In this section, we directly examine whether it fits with profits distribution data.

In order to average the scattering of data points, we employ the cumulative number of companies (Fig. 2). For \( x_1 > x_0 \), by using Eqs. (12) and (22) the cumulative distribution in the high profits region is expressed as

\[
N_H(x) = N(x_0) \int_x^\infty dt \, P(t) = N(x_0) \left( \frac{x}{x_0} \right)^{-\mu}.
\]

On the other hand, by the use of Eqs. (13) and (22) the cumulative distribution in the middle profits region \( x_{\text{min}} < x_1 < x_0 \) is given by

\[
N_M(x) = \{N(x_{\text{min}}) - N(x_0)\} \int_x^{x_0} dt \, P(t) + N(x_0)
\]

\[
= \{N(x_{\text{min}}) - N(x_0)\} \frac{\text{Erf} \left( \frac{1}{2 \sqrt{\alpha}} \right) - \text{Erf} \left( \frac{\mu + 2\alpha \ln \frac{x_{\text{min}}}{x_0}}{2 \sqrt{\alpha}} \right)}{\text{Erf} \left( \frac{1}{2 \sqrt{\alpha}} \right) - \text{Erf} \left( \frac{\mu + 2\alpha \ln \frac{x_{\text{min}}}{x_0}}{2 \sqrt{\alpha}} \right)}
\]

\[
+ N(x_0).
\]

Here \( \text{Erf}(x) \) is error function defined by \( \text{Erf}(x) = 2 \int_0^x e^{-t^2} dt / \sqrt{\pi} \).

In Fig. 7, the distribution functions (24) in the high profits region \( x_1 > x_0 \) and (26) in the middle one \( x_{\text{min}} < x_1 < x_0 \) explain empirical data in 2003 with high accuracy. This guarantees the validity of the assumption in the previous section. Notice that there is no ambiguity in parameter fitting, because indices \( \mu, \alpha \) and the bounds \( x_0, x_{\text{min}} \) is already given in the extended Gibrat’s law \( ^5 \).

5 Another non-Gibrat’s law

In section 2, we present the non-Gibrat’s law (10), (11) and (13) as a linear approximation of \( t_\pm(x_1) \) in Fig. 6. In this section, we examine another linear approximation in Fig. 8, the vertical axis of which is the logarithm.

In Fig. 8, \( \log_{10} t_+ \) (\( \log_{10} t_- \)) increases (decreases) linearly with \( \log_{10} x_1 \) for \( n = 4, 5, \ldots, 11 \). This relation is expressed as

\[
t_+(x_1) = t_+(x_0) \left( \frac{x_1}{x_0} \right)^{+\beta_+},
\]

\[
t_-(x_1) = t_-(x_0) \left( \frac{x_1}{x_0} \right)^{-\beta_-}.
\]

\(^5\)Pareto’s law with \( \mu = 1 \) is especially called Zipf’s law [13].
where

\[ t_+ (x_0) - t_- (x_0) = \mu , \quad (29) \]

\[ \beta_+ \sim 0.35 , \quad \beta_- \sim 0.36 \quad \text{for} \quad x_{\text{min}} < x_1 < x_0 , \quad (30) \]

\[ \beta_+ \sim \beta_- \sim 0 \quad \text{for} \quad x_1 > x_0 . \quad (31) \]

Here \( x_0 \) and \( x_{\text{min}} \) are same values in section 2. From the detailed balance and this extended Gibrat’s law, one finds

\[ \frac{P(x_1)}{P(x_2)} = R^{+t_+ (x_0)} \left( \frac{x_1}{x_0} \right)^{\beta_+} \left( \frac{x_2}{x_0} \right)^{-\beta_-} + 1 \quad (32) \]

for \( R > 1 \).

By expanding this equation around \( R = 1 \), the following differential equation is obtained

\[ \left[ +t_+ (x_0) \left( \frac{x}{x_0} \right)^{\beta_+} - t_- (x_0) \left( \frac{x}{x_0} \right)^{-\beta_-} + 1 \right] P(x) + x P'(x) = 0 , \quad (33) \]

where \( x \) denotes \( x_1 \). The same differential equation is obtained for \( R < 1 \). In order to take \( \beta_\pm \to 0 \) limit into account, we rewrite this as follows:

\[ \left[ +t_+ (x_0) \left\{ \left( \frac{x}{x_0} \right)^{\beta_+} - 1 \right\} - t_- (x_0) \left\{ \left( \frac{x}{x_0} \right)^{-\beta_-} - 1 \right\} + 1 \right] P(x) + x P'(x) = 0 . \quad (34) \]

The solution is given by

\[ P(x) = C x^{-(\mu+1)} e^{-\beta \ln^2 \frac{x}{x_0}} + O(\beta^2) + O(\beta^2) \quad (35) \]

with \( \beta = (\beta_+ t_+(x_0) + \beta_- t_-(x_0)) / 2 \). Here we use the expansion \( x^{\beta-1}/(x_0)^{\beta-1} = \beta (\ln x/x_0)/x + O(\beta^2) \).

If we neglect \( O(\beta^2) \) terms in Eq. (35), the profits distribution function (35) can be identified with (22), because \( \beta \sim \alpha \) numerically. In other words, there is no essential difference between two expressions (10), (11) and (27), (28) in the middle region.

6 Conclusion

In this paper, we have kinematically derived the profits distribution function in the high and middle region from the law of detailed balance and the extended Gibrat’s law by employing profits data of Japanese companies in 2002 and 2003.

Firstly, we have reconfirmed Gibrat’s law in the high profits region \( (x > x_0) \). The value of \( x_0 \) is estimated to be about 100 million yen in Fig. 6 or 8. At the same time, we have found that Gibrat’s law fails and another law holds in the middle profits region \( (x_{\text{min}} < x < x_0) \). We have identified the non-Gibrat’s law and the value of \( x_{\text{min}} \) is estimated to be about 1,600 thousand yen in Fig. 6 or 8. We have called the combination of Gibrat’s law and the non-Gibrat’s law the extended Gibrat’s law.
Secondly, we have derived not only Pareto’s law in the high profits region but also the distribution in the middle one from the detailed balance and the extended Gibrat’s law. The derivation has been described uniformly in terms of the extended Gibrat’s law.

The profits distribution in the middle region is very similar to log-normal one:

\[ P(x) = C x^{-1} \exp \left[ -\frac{1}{2\sigma^2} \ln^2 \frac{x}{\bar{x}} \right], \]

where \( \bar{x} \) is mean value and \( \sigma^2 \) is variance. The difference between two distributions (22) and (36) is only the power of \( x \). It brings a translation along the horizontal axis in the log-log plot. In this sense, the distribution in the middle profits region obtained in this paper is essentially equivalent to the log-normal one. Notice that it has no fitting parameter that the log-normal distribution, prepared only for data fitting, has. Indies \( \mu, \alpha \) and the bounds \( x_0, x_{\text{min}} \) is already given in the extended Gibrat’s law.

In the derivation of the profits distribution, we have used the detailed balance and the extended Gibrat’s law. The detailed balance is observed in a relatively stable period in economy [5]. The extended Gibrat’s law is interpreted as follows. Companies, classified in small-scale profits category, have more (less) possibilities of increasing (decreasing) their profits than companies classified in large-scale one. In other words, it is probably difficult that companies gain large-scale profits in two successive years. This is the non-Gibrat’s law and it leads the distribution in the middle profits region.

This phenomenon is not observed above the threshold \( x_0 \). For \( x > x_0 \), companies, classified in small-scale profits category, have same possibilities of increasing (decreasing) their profits with companies classified in large-scale one. This is the Gibrat’s law and it leads the power distribution in the high profits region.

In this paper, we cannot mention the distribution in the low profits region \( x < x_{\text{min}} \), because no law is observed in the region (Fig. 6 or 8). This may be caused by insufficient data in the low region. Lastly, we speculate possible distributions in the region. If we do not take \( \beta_{\pm} \to 0 \) limit, we should solve the differential equation (33). The solution is given by

\[ P(x) = C x^{-1} \exp \left[ -\frac{1}{\beta_+} \left( \frac{x}{x_0} \right)^{\beta_+} - \frac{1}{\beta_-} \left( \frac{x}{x_0} \right)^{-\beta_-} \right]. \]

For the case \( \beta_+ \sim 1 \) and \( \beta_- \sim 0 \), the distribution (37) takes the exponential form in Refs. [4, 14]. On the other hand, for the case \( \beta_+ \sim 0 \) and \( \beta_- \sim 1 \), the distribution (37) takes Weibull form in Ref. [15]. We can decide the distribution in the low profits region if sufficient data in the region are provided.

We have showed that a distribution function is decided by underlying kinematics. For profits data we used, the distribution is power in the high region and log-normal type in the middle one. This does not claim that all the distributions in the middle region are log-normal types. For instance, the personal income distribution may take a different form, if the extended Gibrat’s law changes. Even in the case, the other extended Gibrat’s law will decide the other distribution by the use of the method in this paper.
Acknowledgments

The author is grateful to Professor H. Aoyama for useful discussions about his lecture.

References


Figure 1: The scatter plot of all companies, the profits of which in 2002 ($x_1$) and 2003 ($x_2$) exceeded 0, $x_1 > 0$ and $x_2 > 0$. The number of the companies is “132,499”.

Cumulative number distributions $N(x_1)$ and $N(x_2)$ for companies, the profits of which in 2002 ($x_1$) and 2003 ($x_2$) exceeded 0, $x_1 > 0$ and $x_2 > 0$. The number of the companies is “132,499”.

The probability density distribution $q(r|x_1)$ of the log profits growth rate $r = \log_{10} x_2/x_1$ from 2002 to 2003. The data points are classified into five bins of the initial profits with equal magnitude in logarithmic scale, $x_1 \in 4 \times [10^{2+0.2(n-1)}, 10^{2+0.2n}]$ ($n = 1, 2, \cdots, 5$) thousand yen. The number of companies in this regime is “35,513”.

Figure 4: The probability density distribution \( q(r|\text{x}_1) \) of the log profits growth rate \( r = \log_{10} \frac{\text{x}_2}{\text{x}_1} \) from 2002 to 2003. The data points are also classified into five bins of the initial profits with equal magnitude in logarithmic scale, \( \text{x}_1 \in 4 \times [10^{2+0.2(n-1)}, 10^{2+0.2n}] \) (\( n = 6, 7, \ldots, 10 \)) thousand yen. The number of companies in this regime is “64,205”.

Figure 5: The probability density distribution \( q(r|\text{x}_1) \) of the log profits growth rate \( r = \log_{10} \frac{\text{x}_2}{\text{x}_1} \) from 2002 to 2003. The data points are also classified into five bins of the initial profits with equal magnitude in logarithmic scale, \( \text{x}_1 \in 4 \times [10^{2+0.2(n-1)}, 10^{2+0.2n}] \) (\( n = 11, 12, \ldots, 15 \)) thousand yen. The number of companies in this regime is “25,189”.
Figure 6: The relation between the lower bound of each bin $x_1$ and $t_\pm(x_1)$. From the left, each data point represents $n = 1, 2, \cdots, 15$.

Figure 7: Cumulative number distributions $N(x_2)$ for companies, the profits of which in 2002 ($x_1$) and 2003 ($x_2$) exceeded 0, $x_1 > 0$ and $x_2 > 0$. The cumulative distribution function derived from the detailed balance and the extended Gibrat’s law accurately fits with the data. Indices $\mu$, $\alpha$ and the bounds $x_0$, $x_{\min}$ is already given in the extended Gibrat’s law.
Figure 8: The log-log plot of the relation between the lower bound of each bin $x_1$ and $t_{\pm}(x_1)$. From the left, each data point represents $n = 1, 2, \cdots, 15$. 