Shape of Growth Rate Distribution Determines the Type of Non-Gibrat's Property

Atushi Ishikawa
Shouji Fujimoto
And
Takayuki Mizuno

March 9, 2010
(Revised: June 5, 2010)

Shape of Growth Rate Distribution
Determines the Type of Non-Gibrat's Property

Atushi Ishikawa[*] and Shouji Fujimoto[†]

Kanazawa Gakuin University, Kanazawa, Japan

Takayuki Mizuno[‡]

Hitotsubashi University, Tokyo, Japan

(Dated: June 5, 2010)

Abstract

In this study, the authors examine exhaustive business data on Japanese firms, which cover nearly all companies in the mid- and large-scale ranges in terms of firm size, to reach several key findings on profits/sales distribution and business growth trends. First, detailed balance is observed not only in profits data but also in sales data. Furthermore, the growth-rate distribution of sales has wider tails than the linear growth-rate distribution of profits in log-log scale. On the one hand, in the mid-scale range of profits, the probability of positive growth decreases and the probability of negative growth increases symmetrically as the initial value increases. This is called Non-Gibrat's First Property. On the other hand, in the mid-scale range of sales, the probability of positive growth decreases as the initial value increases, while the probability of negative growth hardly changes. This is called Non-Gibrat's Second Property. Under detailed balance, Non-Gibrat's First and Second Properties are analytically derived from the linear and quadratic growth-rate distributions in log-log scale, respectively. In both cases, the log-normal distribution is inferred from Non-Gibrat's Properties and detailed balance. These analytic results are verified by empirical data. Consequently, this clarifies the notion that the difference in shapes between growth-rate distributions of sales and profits is closely related to the difference between the two Non-Gibrat's Properties in the mid-scale range.

INTRODUCTION

Distributions with a power-law tail have been found in various fields of natural and social science. Examples of such studies include, for instance, avalanche sizes in a sandpile model [1], fluctuations in the intervals of heartbeats [2], fish school sizes [3], citation numbers of scientific papers [4], frequency of jams in Internet traffic [5], city sizes (see the recent review in Ref. [6]), land prices [7]–[9], stock market price changes [10], and firm sizes [11]. Here, variables (denoted by x) follow the probability density function (PDF):

$$P(x) \propto x^{-(\mu+1)} \quad \text{for} \quad x > x_{\text{th}} \tag{1}$$

over some size threshold x_{th}. This is called Pareto's Law, which was first observed in the field of personal income [12]. The index μ is called the Pareto index. Refer to Newman [13] for a useful description of Pareto's Law.

In statistical physics, the study of distributions with a power-law tail (1) is significant because the k-th moment $\langle x^k \rangle = \int dx P(x) x^k$ diverges in the case of $\mu \leq k$. It is impossible to describe the system by using the variance $\sigma^2 = \langle x^2 \rangle$ or the standard deviation σ in the case of $\mu \leq 2$. This feature comes from power-law behavior in the tail. Furthermore, it is worth noting that a large portion of the overall data are included in the power-law tail. For example, approximately 90% of total sales or profits in Japanese firms are included in the power-law tail. In economics (especially in macroeconomics), one of the major issues is the state of the entire economy. In this sense, it is important to clarify the nature of the power-law tail not only in physics but also in economics.

In general, the power-law breaks below the size threshold x_{th} to suppress the divergence of the PDF [14], [15]. There are many distributions that have a power-law tail. These include, for instance, Classical Pareto Distribution (Pareto Type I Distribution), Pareto Type II Distribution, Inverse Gamma Distribution, Inverse Weibull Distribution, q–Distribution, A–Distribution and B–Distribution [16]. In addition to these distributions, it has been hypothesized that many other distributions with a power-law tail follow the log-normal distribution for mid-sized variables below the size threshold x_{th}:

$$P(x) \propto \frac{1}{x} \exp\left[-\frac{1}{2\sigma^2} \ln^2 \frac{x}{\bar{x}}\right] \quad \text{for} \quad x_{\text{min}} < x < x_{\text{th}} . \tag{2}$$

Here, $\bar{x}$ is a mean value and σ^2 is a variance. A lower bound of the mid-scale range x_{min} is often related to the lower bound of an exhaustive set of data. A pseudo log-normal

distribution is approximately derived from A–Distribution or B–Distribution in the mid-sized range [16].

The study of distributions in the mid-scale range below the size threshold x_{th} is as important as the study of the power-law tail. In physics, we are interested not only in the mechanism generating a power-law tail but also in the reason for the tail breaking. In economics, we should note that the majority of firms are mid-sized. For instance, in sales or profits data, more than 90% of the total number of firms are in the mid-scale range. In this study, by examining exhaustive business data of Japanese firms that nearly cover the mid- and large-scale ranges, the authors investigate the relevant distributions with a power-law tail. This research is expected to be useful for understanding phenomena not only in economics but also in physics.

On the one hand, it has been shown that Pareto's Law and the log-normal distribution can be derived by assuming some model. For example, a multiplicative process with boundary constraints and additive noise can generate Pareto's Law [17]. On the other hand, by using no model, Fujiwara et al. have recently shown that Pareto's Law (1) is derived from Gibrat's Law and from the detailed balance observed in the large-scale range of exhaustive business data [18]. The relations among laws observed in exhaustive business data are important for examining the characteristics of distributions based on firm-size. For instance, in the study of Fujiwara et al., it was found that Pareto index μ is related to the difference between a positive growth-rate distribution and a negative one. Furthermore, along the lines of their study, one of the authors (A. I) has shown that the log-normal distribution (2) can be inferred from detailed balance and from Non-Gibrat's Property observed in the profits data of the mid-scale range [19]. The study of the growth-rate distribution is an interesting subject in itself, and an ongoing investigation into this issue has progressed recently [20].

Detailed balance means that the system is thermodynamically in equilibrium, the state of which is described as

$$P_J(x_T, x_{T+1}) = P_J(x_{T+1}, x_T) \ . \tag{3}$$

Here, x_T and x_{T+1} are firm sizes at two successive points in time. In Eq. (3), the joint PDF $P_J(x_T, x_{T+1})$ is symmetric under the time reversal exchange $x_T \leftrightarrow x_{T+1}$.

Gibrat's Law and Non-Gibrat's Property are observed in the distributions of firm-size growth rate $R = x_{T+1}/x_T$. The conditional PDF of the growth rate $Q(R|x_T)$ is defined

3

as $Q(R|x_T) = P_J(x_T, R)/P(x_T)$ by using the PDF $P(x_T)$ and the joint PDF $P_J(x_T, R)$. Gibrat's Law, which is observed in the large-scale range, implies that the conditional PDF $Q(R|x_T)$ is independent of the initial value x_T [21]:

$$Q(R|x_T) = Q(R) . \tag{4}$$

Sutton [22] provides an instructive resource for obtaining the proper perspective on Gibrat's Law.

Non-Gibrat's Property reflects the dependence of the growth-rate distribution on the initial value x_T. The following properties are observed in the mid-scale range of positive profits data of Japanese firms [19]:

$$Q(R|x_T) = d(x_T) \ R^{-t_+(x_T)-1} \qquad \text{for} \ \ R > 1 , \tag{5}$$

$$Q(R|x_T) = d(x_T) \ R^{+t_-(x_T)-1} \qquad \text{for} \ \ R < 1 , \tag{6}$$

$$t_\pm(x_T) = \pm\alpha \ \ln x_T + C_\pm . \tag{7}$$

Here, α and $C_\pm$ are positive constants. In this composite Non-Gibrat's Property (5)–(7), the probability of positive growth decreases and the probability of negative growth increases symmetrically as the initial value x_T increases in the mid-scale range. It is particularly noteworthy that the shape of the growth-rate distribution (5)–(6) uniquely determines the change in the growth-rate distribution (7) under detailed balance (3). Moreover, the rate-of-change parameter α appears in the log-normal distribution (2). We designate (5)–(7) as Non-Gibrat's First Property to distinguish it from another Non-Gibrat's Property that is observed in sales data.

The shape of the growth-rate distribution (5)–(6) is linear in log-log scale. This type of growth-rate distribution is observed in profits and income data of firms (for instance [23], [24], [25]). In contrast, it has been reported in various articles that the growth-rate distributions of assets, sales, number of employees in firms, and personal income have wider tails than those of profits and income in log-log scale (for instance [26], [18], [27], [28], [29], [25]). In this case, the shape of the growth-rate distribution is different from Eqs. (5) and (6). There must be, therefore, another Non-Gibrat's Property corresponding to this shape. In fact, it has been reported in several studies that a Non-Gibrat's Property different from Non-Gibrat's First Property exists in the mid-scale range of assets and sales of firms (for instance [30]–[32]).

In this study, we report the following findings by employing the sales data of Japanese firms, which include not only data in the large-scale range but also those in the mid-scale range.

1. Detailed balance (3) is confirmed in the mid- and large-scale ranges of sales data.

2. In not only the large-scale range but also the mid-scale range of sales data, the growth-rate distributions have wider tails than those of profits in log-log scale.

3. Under detailed balance (3), the allowed change of the growth-rate distribution in the mid-scale range is analytically determined by using empirical data. The change is different from that of profits. We call this Non-Gibrat's Second Property.

4. A log-normal distribution is derived from Non-Gibrat's Second Property and from detailed balance. This is verified with empirical data.

From these results, we conclude that the shape of the growth-rate distribution determines the type of Non-Gibrat's Property in the mid-scale range.

NON-GIBRAT'S FIRST PROPERTY

In this section, we review the analytic discussion in Ref. [19] and confirm it by applying the results to newly obtained data. In the analytic discussion, detailed balance (3) and the shape of the growth-rate distribution (5)–(6) lead uniquely to a change in the growth-rate distribution (7). In addition, Non-Gibrat's First Property and detailed balance derive a log-normal distribution (2) in the mid-scale range.

In this study, we employ profits and sales data supplied by the Research Institute of Economy, Trade and Industry, IAA (RIETI) [33]. In this section we analyze profits data, and sales data are analyzed in the next section. The data set, which was created by TOKYO SHOKO RESEARCH, LTD. [34] in 2005, includes approximately 800,000 Japanese firms over a period of three years: the current year, the preceding year, and the year before that. The number of firms is approximately the same as the actual number of active Japanese firms. This database is considered nearly comprehensive, at least in the mid- and large-scale ranges. In this study, we investigate the joint PDF $P_J(x_T, x_{T+1})$ and the distribution of the growth rate $R = x_{T+1}/x_T$. Therefore, by using data of each firm in the previous three
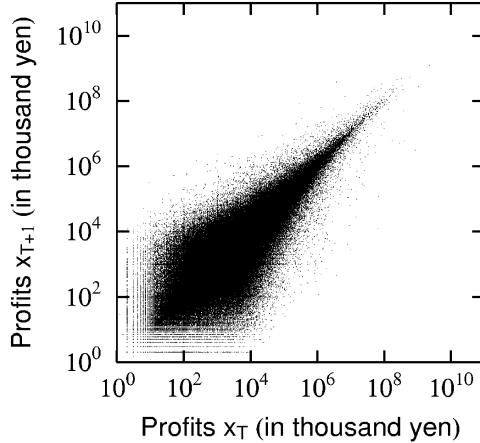
FIG. 1. Scatter plot of positive profits in the database. Here, x_T and x_{T+1} are positive profits of individual firms in consecutive years.

years, we analyze a data set that has two values at two successive points in time as follows: (x_T, x_{T+1}) = (data in preceding year, data in current year) $\cup$ (data in year before last, data in preceding year). Here, $\cup$ indicates set-theoretic union. This superposition of data is employed in order to secure a statistically sufficient sample size. This procedure is allowed in cases where the economy is stable, that is, thermodynamically in equilibrium. The validity is checked by detailed balance, as described below.

First, detailed balance (3) is observed in profits data. Note that only positive-profits data are analyzed here, since we assume that non-negligible negative profits are not listed in the database. Negative-profits data are thus not regarded as exhaustive. We employ "622,420" data sets (x_T, x_{T+1}) that have two positive profits at two successive points in time. Figure. 1 shows the joint PDF $P_J(x_T, x_{T+1})$ as a scatter plot of individual firms. Detailed balance (3) is confirmed by the Kolmogorov–Smirnov (KS), Wilcoxon–Mann–Whitney (WMW), and Brunner–Munzel (BM) tests. In the statistical tests, the range of x_T is divided into N bins as $i_0 \leq i_1 \leq \cdots \leq i_{n-1} \leq i_n \leq \cdots \leq i_N$ to approximately equalize the number of data in each bin "$x_T \in [i_{n-1}, i_n)$ and $x_T > x_{T+1}$." Here, i_0 and i_N are the lower and the upper bounds of x_T, respectively. We compare the distribution sample for "$P_J(x_T \in [i_{n-1}, i_n), x_{T+1})$ and $x_T > x_{T+1}$" with another sample for "$P_J(x_T, x_{T+1} \in [i_{n-1}, i_n))$ and $x_T < x_{T+1}$" ($n = 1, 2, \cdots, N$) by making the null hypothesis that these two samples are taken from the same parent distribution.
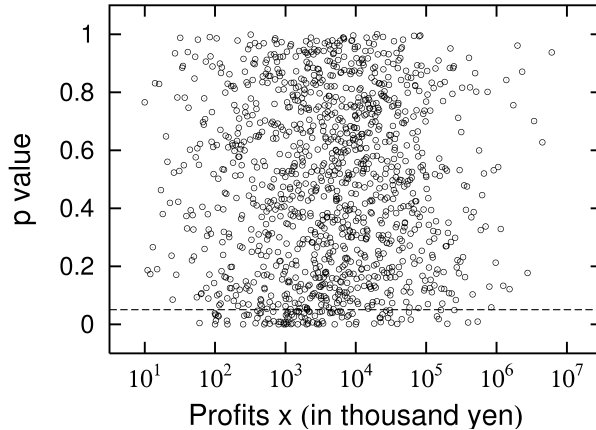
6

FIG. 2. Each p value of the WMW test for the scatter plot of positive-profits data points in Fig. 1.

Each p value of the WMW test for the case of $N = 2000$ is shown in Fig. 2. Note that the profits data contain a large number of same-value amounts, which are round numbers: 100, 200, $\cdots$, 1000, 2000, $\cdots$, 10000, 20000, $\cdots$. This phenomenon is frequently observed in economic data. A bin with a round-number amount may contain an exceptionally large number of data in this method of division. For the case of $N = 2000$, almost all bins typically contain 200 data; however, a bin with the round number of 5000, for instance, contains an exceptional 4437 data. In order to generally equalize the average amount of data in bins to the typical value, an appropriate number of empty bins are inserted at such bins of round-number amounts as needed (Fig. 3). In the case of $N = 2000$, there are 759 empty bins. P values with respect to the remaining 1241 bins are depicted in Fig. 2, in which 1141 p values exceed 0.05. Regardless of the division number N and the kind of test, p values exceed 0.05 in approximately 92% of bins. This means that the null hypothesis is not rejected within the 5% significance level in approximately 92% of the range. This result does not change in the case where the range of x_T is divided into logarithmically equal bins. Consequently, the detailed balance (3) in Fig. 1 is generally confirmed.

Second, we divide the range of the initial value x_T into logarithmically equal bins as $x_T \in [10^{1+0.4(n-1)}, 10^{1+0.4n})$ $(n = 1, 2, \cdots, 15)$ in order to identify the shape of the growth-rate distribution and the change as the initial value x_T increases. The conditional PDFs $q(r|x_T)$ of the logarithmic growth rate $r = \log_{10} R$ are shown in Figs. 4–6. In Figs. 5 and 6, the growth-rate distributions in the mid- and large-scale ranges are approximated by a
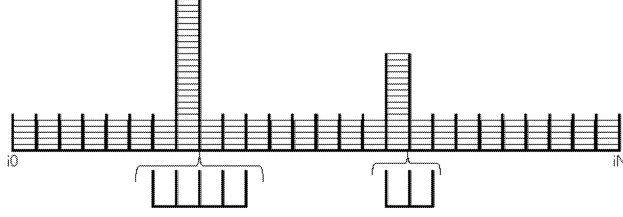
FIG. 3. A bin with a round-number amount contains an exceptionally large number of data. In order to generally equalize the average amount of data in bins to the typical value, empty bins are inserted at bins with round-number amounts as needed.
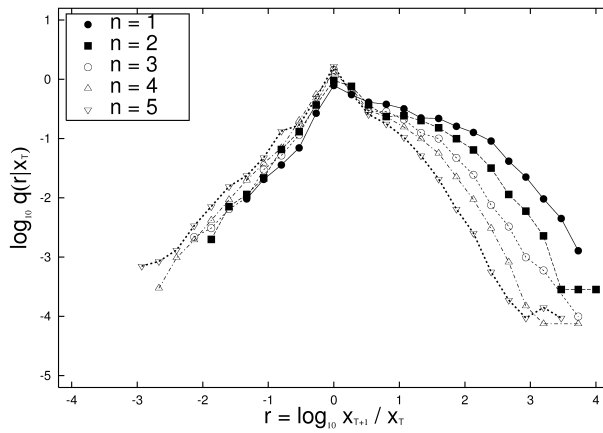

FIG. 4. Conditional PDFs of positive-profits growth rate in the low-scale range ($10^1 \leq x_T < 10^3$). Here, x_T and x_{T+1} are positive profits in consecutive years, in thousand yen.

linear function of r:

$$\log_{10} q(r|x_T) = c(x_T) - t_+(x_T)\, r \qquad \text{for}\ \ r > 0\ , \tag{8}$$

$$\log_{10} q(r|x_T) = c(x_T) + t_-(x_T)\, r \qquad \text{for}\ \ r < 0\ . \tag{9}$$

The approximation (8)–(9) is equivalent to Eqs. (5) and (6) by using relations $\log_{10} q(r|x_T) = \log_{10} Q(R|x_T) + r + \log_{10}(\ln 10)$ and $d(x_T) = 10^{c(x_T)}/\ln 10$. From $\int_0^\infty dR\, Q(R|x_T) = 1$, the normalization coefficient $d(x_T)$ (or the intercept $c(x_T)$) is determined as

$$\frac{1}{d(x)} = \frac{1}{t_+(x)} + \frac{1}{t_-(x)}\ . \tag{10}$$

Following the discussion in a previous work [19], we derive the change in the growth-rate distribution (7) from the shape of the growth-rate distribution (5)–(6) under detailed balance
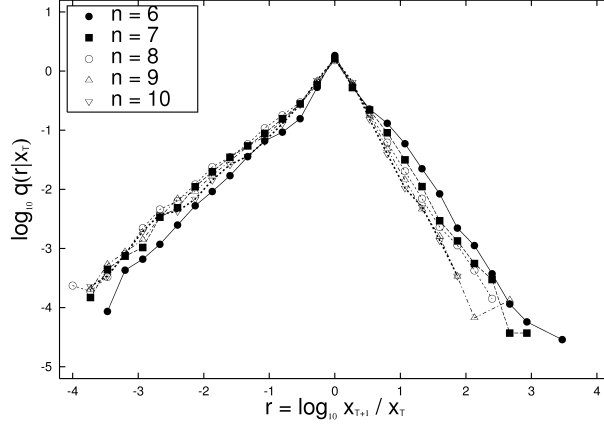
8

FIG. 5. Conditional PDFs of positive-profits growth rate in the mid-scale range ($10^3 \leq x_T < 10^5$). Here, x_T and x_{T+1} are positive profits in consecutive years, in thousand yen.
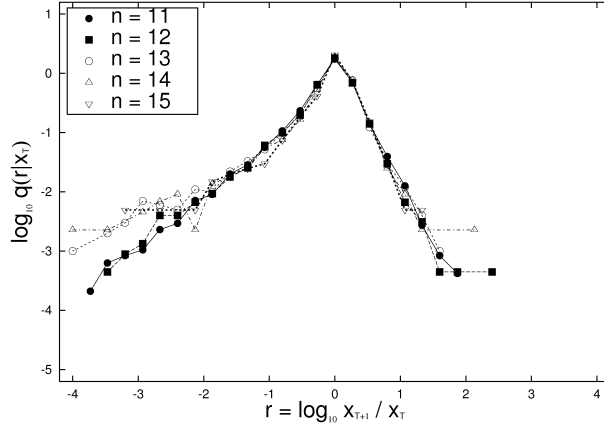

FIG. 6. Conditional PDFs of positive-profits growth rate in the large-scale range ($10^5 \leq x_T < 10^7$). Here, x_T and x_{T+1} are positive profits in consecutive years, in thousand yen.

(3) and then derive the log-normal distribution in the mid-scale range. Under the exchange of variables from (x_T, x_{T+1}) to (x_T, R), two joint PDFs $P_J(x_T, x_{T+1})$ and $P_J(x_T, R)$ are related to each other as $P_J(x_T, R) = x_T P_J(x_T, x_{T+1})$. Substituting the joint PDF $P_J(x_T, R)$ for the conditional PDF $Q(R|x_T)$ and using detailed balance (3), we obtain

$$\frac{P(x_T)}{P(x_{T+1})} = \frac{1}{R} \frac{Q(R^{-1}|x_{T+1})}{Q(R|x_T)} \ . \tag{11}$$

By substituting the conditional PDF for the shape of the growth-rate distribution (5)–(6),

9

another expression of detailed balance (11) is reduced to

$$\frac{\tilde{P}(x_T)}{\tilde{P}(x_{T+1})} = R^{+t_+(x_T)-t_-(x_{T+1})+1} \tag{12}$$

for the case of $R > 1$. Here, we denote $\tilde{P}(x) = d(x)\ P(x)$. By expanding Eq. (12) around $R = 1$ with $x_T \to x$ and $x_{T+1} \to R\ x$, the following three differential equations are obtained:

$$\left[1 + t_+(x) - t_-(x)\right]\tilde{P}(x) + x\ \tilde{P}'(x) = 0 \ , \tag{13}$$

$$t_+{}'(x) + t_-{}'(x) = 0 \ , \quad t_+{}'(x) + x\ t_+{}''(x) = 0 \ . \tag{14}$$

The same differential equations are obtained for $R < 1$. Equations (14) uniquely fix $t_\pm(x_T)$ as Eq. (7). Now, let us verify this by empirical data.

Figure 7 shows $t_\pm(x_T)$ and $c(x_T)$ estimated by fitting the approximation (8)–(9) to each growth-rate distribution in Figs. 4–6. In Fig. 7, $c(x_T)$ is fixed as the empirical value and $t_\pm(x_T)$ is estimated by using the least-squares method. In Fig. 4, the linear function (8)–(9) is difficult to approximate for each growth-rate distribution, and the values for $n = 1, 2, \cdots, 5$ in Fig. 7 are untrustworthy. In Fig. 5, however, the linear approximation (8)–(9) is appropriate. Applying the change in the growth-rate distribution $t_\pm(x_T)$ (7) to $n = 6, 7, 8$ ($10^3 \leq x_T < 10^{4.2}$) in Fig. 7, we obtain the rate-of-change parameter $\alpha = 0.11 \pm 0.02$ from $t_+(x_T)$ and $\alpha = 0.11 \pm 0.03$ from $t_-(x_T)$ by using the least-squares method. This
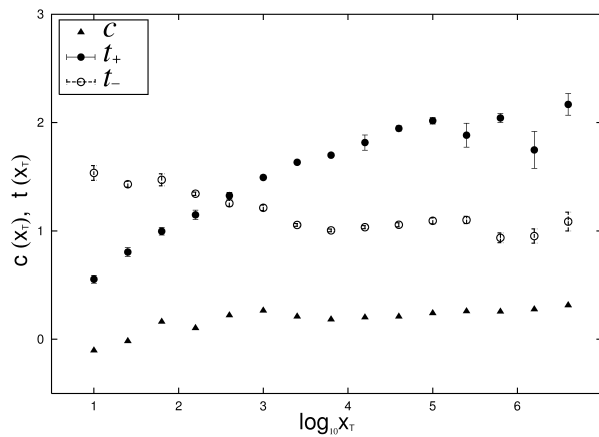

FIG. 7. Estimations of $c(x_T)$ and $t_\pm(x_T)$. Here, x_T is the lower bound of each bin, in thousand yen, and $c(x_T)$ is the original value of the growth-rate distribution. From left, each point on the graph represents $n = 1, 2, \cdots, 15$.
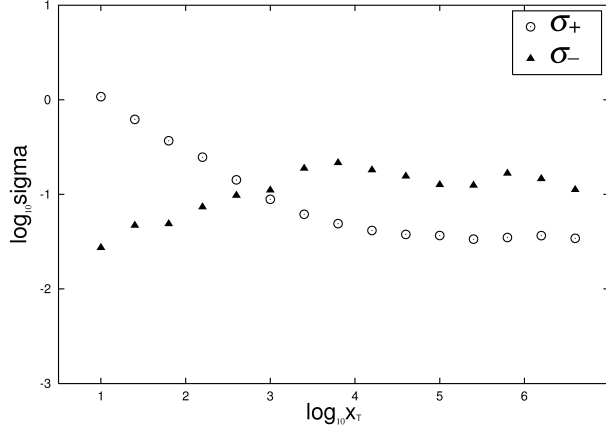
10

FIG. 8. Estimations of $\sigma_{\pm}(x_T)$. Here, x_T is the lower bound of each bin, in thousand yen. From left, each point on the graph represents $n = 1, 2, \cdots, 15$.

coincidence of two estimated values guarantees Non-Gibrat's First Property (5)–(7) in the empirical data. We regard $10^3 \leq x_T < 10^{4.2}$ as the mid-scale range.

In Fig. 6, the growth-rate distribution barely changes as n increases. This means that Gibrat's Law (4) is valid in the large-scale range. In Fig. 7, the values $t_{\pm}(x_T)$ vary in the large-scale range, since the number of data in Fig. 6 is statistically insufficient to estimate $t_{\pm}(x_T)$ by the least-squares method. However, by measuring the positive and negative standard deviations $\sigma_{\pm}$ of each growth-rate distribution in Figs. 4–6, we confirmed that the growth-rate distribution only slightly changes in the range $x_T \geq 10^5$ (Fig. 8). From Fig. 8, we regard $x_T \geq 10^5$ as the large-scale range and set $\alpha = 0$ in this range. Strictly speaking, a constant parameter α must not take different values. However, in the database, a large number of firms stay in the same range for two successive years. This parameterization is, therefore, generally suitable for describing the PDF.

In Fig. 7, $c(x_T) = \log_{10}(d(x_T) \ln 10)$ hardly changes in the mid- and large-scale ranges $x_T \geq 10^3$. This is consistent with $C_{\pm} >> \alpha \ln x_T$ in Eqs. (7) and (10). Consequently, by approximation we determine that the dependence of $d(x_T)$ on x_T is negligible in the mid- and large-scale ranges. Using $t_{\pm}(x)$ (7), Eq. (13) uniquely decides the PDF of x as

$$P(x) = C \ x^{-(\mu+1)} \ \exp\left[-\alpha \ln^2 x\right] \qquad \text{for} \quad x > x_{\min} \ . \qquad (15)$$

Here, we regard $d(x)$ in $\tilde{P}(x) = d(x) \ P(x)$ as a constant and denote $\mu = C_+ - C_-$. The solutions (7) and (15) satisfy Eq. (12) beyond perturbation around $R = 1$, and thus these

11

are not only necessary but also sufficient.

Figure 9 shows that the resultant PDF (15) fits correctly with the empirical profits data. In the large-scale range ($\alpha = 0$), the PDF (15) behaves as Pareto's Law (1). The Pareto index is estimated as approximately $\mu \sim 1$ in the large-scale range ($x \geq 10^5$) of Fig. 9. In the mid-scale range, the PDF (15) behaves as the log-normal distribution (2) with $\alpha = 1/(2\sigma^2)$, $\mu = -\ln \bar{x}/(\sigma^2)$. Applying the PDF (15) to the mid-scale range ($10^3 \leq x < 10^{4.2}$) of Fig. 9, we obtain the rate-of-change parameter $\alpha = 0.082 \pm 0.089$ by using the least-squares method. The error bar is not small because we have applied the least-squares method to the quadratic curve in log-log scale. The estimated value ($\alpha = 0.082 \pm 0.089$) is, however, consistent with the values estimated by the change in $t_\pm(x_T)$ ($\alpha = 0.11 \pm 0.02$ or 0.11 ± 0.03). From these results, we conclude that Non-Gibrat's First Property is confirmed by the empirical data.
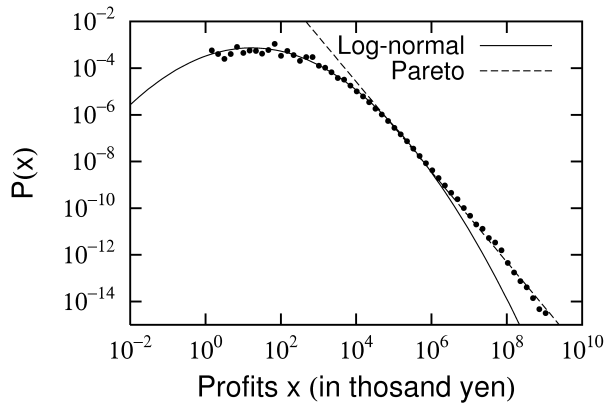

FIG. 9. A PDF of positive profits in the database. Pareto's Law is observed in the large-scale range ($x \geq 10^5$) and in the log-normal distribution in the mid-scale range ($10^3 \leq x < 10^{4.2}$).

NON-GIBRAT'S SECOND PROPERTY

In this section, we investigate another Non–Gibrat's Property observed in the mid-scale range of sales data. This is the main aim of this study. First, detailed balance (3) is also observed in sales data. Here, we employ "1,505,108" data sets (x_T, x_{T+1}) that have two sales at two successive points in time. Figure 10 shows the joint PDF $P_J(x_T, x_{T+1})$ as a scatter plot of individual firms. Detailed balance (3) is also confirmed by using the KS, WMW, and BM tests in the same manner as in the previous section. Figure 11 shows each p value of the
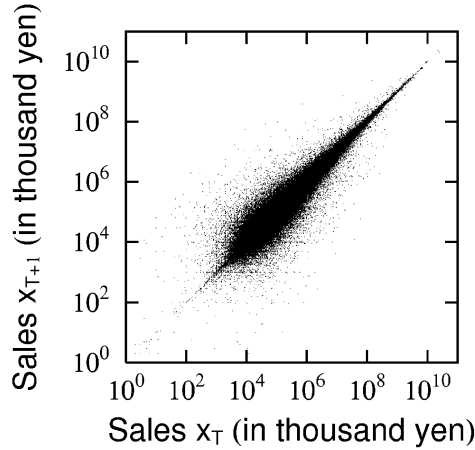
FIG. 10. Scatter plot of sales in the database. Here, x_T and x_{T+1} are sales of individual firms in consecutive years.
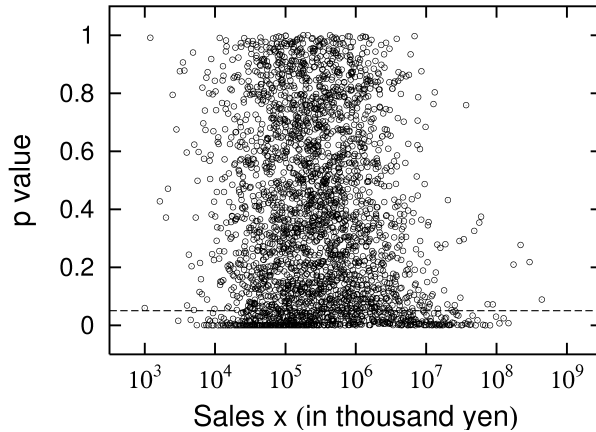

FIG. 11. Each p value of the BM test for the scatter plot of sales data points in Fig. 10.

BM test for the $N = 5000$ case. Regardless of the division number N and the kind of test, p values exceed 0.05 in approximately 82% of bins. This means that the null hypothesis is not rejected within the 5% significance level in approximately 82% of the range. Note that the sales data also contain a large number of same-value amounts, which are round numbers. P values of the statistical test for bins with a large number of round values are unusually small. In this situation, 82% is acceptable. The percentage is slightly higher in the case where the range of x_T is divided into logarithmically equal bins. We assume, therefore, that detailed balance (3) in Fig. 10 is generally verified.

13

Second, we divide the range of the initial value x_T into logarithmically equal bins as $x_T \in [10^{3+0.4(n-1)}, 10^{3+0.4n})$ $(n = 1, 2, \cdots, 15)$. The conditional growth-rate distributions $q(r|x_T)$ are shown in Figs. 12–14. Each growth-rate distribution in Figs. 12–14 has curvatures. It
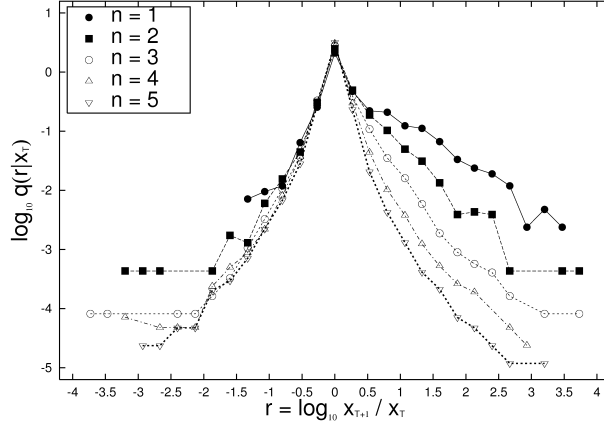

FIG. 12. Conditional PDFs of sales growth rate in the small- and mid-scale ranges ($10^3 \le x_T < 10^5$). Here, x_T and x_{T+1} are sales in consecutive years, in thousand yen.
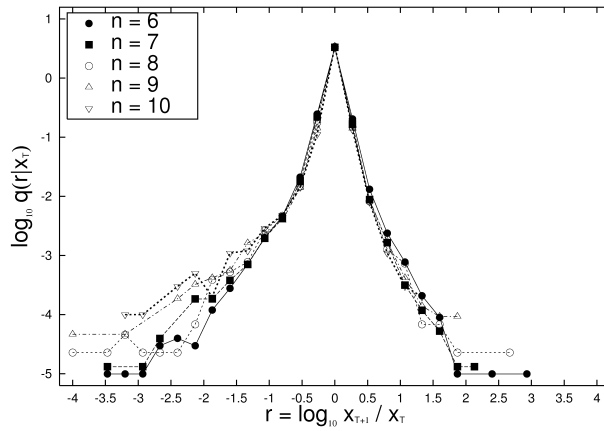

FIG. 13. Conditional PDFs of sales growth rate in the mid- and large-scale ranges ($10^5 \le x_T < 10^7$). Here, x_T and x_{T+1} are sales in consecutive years, in thousand yen.

is difficult to approximate the growth-rate distributions by the linear approximation (8)–(9) as in the profits case. As the simplest extension, we have added a second-order term with respect to r to express the curvatures as follows:

$$\log_{10} q(r|x_T) = c(x_T) - t_+(x_T)\, r + \ln 10\, u_+(x_T)\, r^2 \quad \text{for} \quad r_c > r > 0, \tag{16}$$
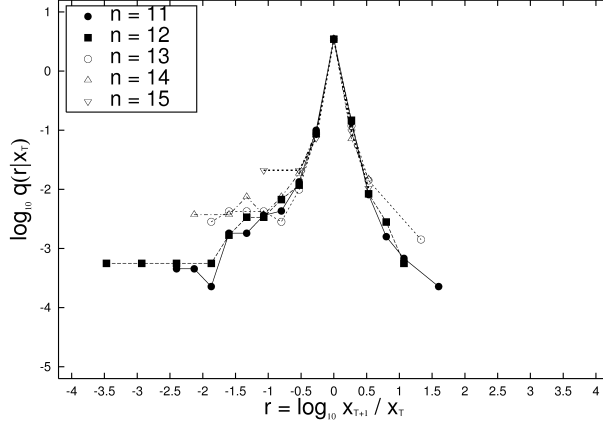
14

FIG. 14. Conditional PDFs of sales growth rate in the large-scale range ($10^7 \leq x_T < 10^9$). Here, x_T and x_{T+1} are sales in consecutive years, in thousand yen.

$$\log_{10} q(r|x_T) = c(x_T) + t_-(x_T) \, r + \ln 10 \, u_-(x_T) \, r^2 \qquad \text{for} \quad -r_c < r < 0 \, . \qquad (17)$$

Note that we must introduce a cut r_c in order to normalize the probability integration as $\int_{10^{-r_c}}^{10^{r_c}} dR \; Q(R|x_T) = 1$, since Eqs. (16) and (17) are quadratic with respect to r. From this normalization condition, $c(x_T)$ can be expressed by using $t_\pm(x_T)$, $u_\pm(x_T)$, and r_c. The expression is quite complicated, and it is later observed that $c(x_T)$ only slightly depends on x_T in the empirical data. Therefore, we do not describe the expression here.

The approximation (16)–(17) is rewritten as

$$Q(R|x_T) = d(x_T) \, R^{-1-t_+(x_T)+u_+(x_T)\ln R} \qquad \text{for} \quad R > 1 \, , \qquad (18)$$

$$Q(R|x_T) = d(x_T) \, R^{-1+t_-(x_T)+u_-(x_T)\ln R} \qquad \text{for} \quad R < 1 \, . \qquad (19)$$

By using this shape, in the case of $R > 1$, detailed balance (11) is reduced to

$$\frac{\tilde{P}(x_T)}{\tilde{P}(x_{T+1})} = R^{\;1+t_+(x_T)-t_-(x_{T+1})-[u_+(x_T)-u_-(x_{T+1})]\ln R} \; . \qquad (20)$$

By expanding Eq. (20) around $R = 1$ with $x_T \to x$ and $x_{T+1} \to R\,x$, the following five differential equations are obtained:

$$\left[1 + t_+(x) - t_-(x)\right]\tilde{P}(x) + x\,\tilde{P}'(x) = 0 \, , \qquad (21)$$

$$x\left[t_+{}'(x) + t_-{}'(x)\right] + 2\left[u_+(x) - u_-(x)\right] = 0 \, , \qquad (22)$$

$$2\,t_+{}'(x) + t_-{}'(x) + 6u_+{}'(x) + x\left[2\,t_+{}''(x) + t_-{}''(x)\right] = 0 \, , \qquad (23)$$

15

$$t_+{}'(x) + t_-{}'(x) + 3x\left[t_+{}''(x) + t_-{}''(x)\right] + x^2\left[t_+{}^{(3)}(x) + t_-{}^{(3)}(x)\right] = 0 \ , \tag{24}$$

$$t_+{}'(x) + 7x\, t_+{}''(x) + 6x^2\, t_+{}^{(3)}(x) + x^3\, t_+{}^{(4)}(x) = 0 \ . \tag{25}$$

The same differential equations are obtained for $R < 1$. Equations (22)–(25) uniquely fix the change in the growth-rate distribution $t_\pm(x)$, $u_\pm(x)$ as follows:

$$t_+(x) = \frac{\gamma}{3}\ln^3 x + \frac{\beta}{2}\ln^2 x + \alpha \ln x + C_1 \ , \tag{26}$$

$$t_-(x) = -\frac{\gamma}{3}\ln^3 x + \frac{\delta - \beta}{2}\ln^2 x + (\eta - \alpha)\ln x + C_2 \ , \tag{27}$$

$$u_+(x) = -\frac{\gamma}{6}\ln^2 x - \frac{\delta + \beta}{6}\ln x + C_3 \ , \tag{28}$$

$$u_-(x) = -\frac{\gamma}{6}\ln^2 x + \frac{2\delta - \beta}{6}\ln x + C_3 + \frac{\eta}{2} \ . \tag{29}$$

Now, let us confirm these solutions with the empirical data.

Figure 15 shows $t_\pm(x_T)$, $u_\pm(x_T)$ and $c(x_T)$ estimated by fitting the approximation (16)–(17) to each growth-rate distribution in Figs. 12–14. In Fig. 15, $c(x_T)$ is fixed as the empirical value and $t_\pm(x_T)$ and $u_\pm(x_T)$ are estimated by using the least-squares method. For $n = 13, 14, 15$ in Fig. 14, there are not sufficient data points to estimate $t_\pm(x_T)$, $u_\pm(x_T)$ for $n = 14, 15$ or to estimate the error bar for $n = 13$. Therefore, data points for $n = 13, 14, 15$ are not plotted in Fig. 15.

On the one hand, for $n = 9, 10, \cdots, 15$ ($x_T \geq 10^{6.2}$) in Figs. 13 and 14, the growth-rate
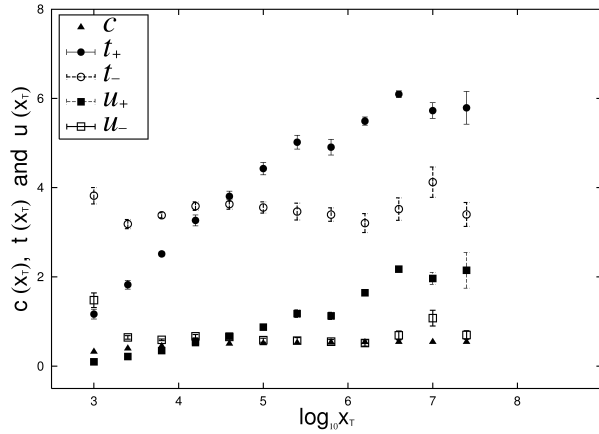

FIG. 15. Estimations of $c(x_T)$, $t_\pm(x_T)$, and $u_\pm(x_T)$. Here, x_T is the lower bound of each bin, in thousand yen, and $c(x_T)$ is the original value of the growth-rate distribution. From left, each point on the graph represents $n = 1, 2, \cdots, 12$.

16

distribution hardly changes as n increases. This means that Gibrat's Law (4) is verified by the empirical data. We regard $x_T \geq 10^{6.2}$ as the large-scale range and set $\gamma = \beta = \delta = \alpha = \eta = 0$ in this range because $t_\pm(x_T)$ and $u_\pm(x_T)$ do not depend on x_T. In Fig. 15, the values of $t_\pm(x_T)$ and $u_\pm(x_T)$ vary in this range because the number of data in Fig. 14 is statistically insufficient to estimate them by the least-squares method. However, by measuring positive and negative standard deviations $\sigma_\pm$ of each growth-rate distribution in Figs. 12–14, we confirmed that the growth-rate distribution hardly changes in the large-scale range $x_T \geq 10^{6.2}$ (Fig. 16).
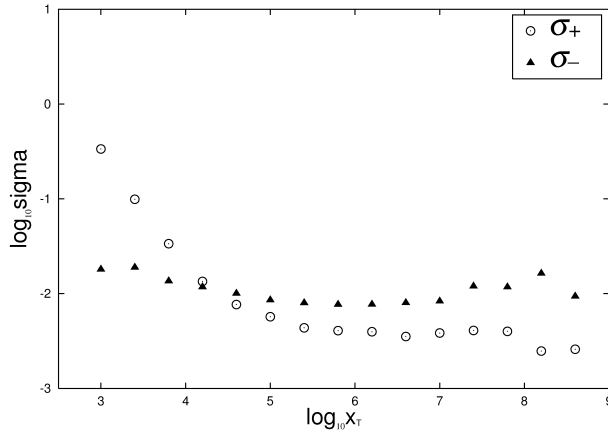

FIG. 16. Estimations of $\sigma_\pm(x_T)$. Here, x_T is the lower bound of each bin, in thousand yen. From left, each point on the graph represents $n = 1, 2, \cdots, 15$.

On the other hand, in Fig. 12, while the negative growth-rate distribution hardly changes as n increases, the positive growth-rate distribution gradually decreases. This is Non-Gibrat's Property in the mid-scale range of sales data. We should estimate parameters $\gamma, \beta, \delta, \alpha$ and η by applying the change in the growth-rate distribution (26)–(29) to Fig. 15. However, there are insufficient data points in Fig. 15 for using the least-squares method by polynomial functions (26)–(29). Consequently, as a first-order approximation, we assume that the negative growth-rate distribution does not depend on x_T, even in the mid-scale range. This approximation is guaranteed by Fig. 16 because the negative standard deviation σ_- hardly changes compared with the positive standard deviation σ_+.

In this approximation, the parameters are simplified as

$$\gamma = \delta = \beta = 0 \quad \text{and} \quad \eta = \alpha . \tag{30}$$

17

Only the change in the positive growth-rate distribution $t_+(x_T)$ depends on x_T as follows:

$$t_+(x) = \alpha \ln x + C_1 \,, \tag{31}$$

$$t_-(x) = C_2 \,, \quad u_+(x) = C_3 \,, \quad u_-(x) = C_3 + \frac{\alpha}{2} \,. \tag{32}$$

We call this Non-Gibrat's Second Property.

Applying $t_+(x_T)$ (31) to $n = 3, 4, 5, 6$ $(10^{3.8} \leq x_T < 10^{5.4})$ in Fig. 15, we obtain the rate-of-change parameter $\alpha = 0.68 \pm 0.03$ by the least-squares method. We regard $10^{3.8} \leq x_T < 10^{5.4}$ as the mid-scale range of sales. In this range, $t_-(x_T)$ and $u_\pm(x_T)$ hardly change compared with $t_+(x_T)$, so the approximation (32) is considered relevant. Nevertheless, the value α estimated by the difference between $u_+(x_T)$ and $u_-(x_T)$ disagrees with the value estimated by the change in $t_+(x_T)$. Most likely, this comes from a limitation of the second-order approximation with respect to r (16)–(17). To fix this discrepancy, we may add a third-order term with respect to r. We will consider this point in the conclusion. In addition, we should note that the intercept $c(x_T)$ only slightly depends on x_T in the mid- and large-scale ranges $x_T \geq 10^{3.8}$, as in the profits case.

Using $t_\pm(x)$ (26)–(27), Eq. (21) uniquely determines the PDF of x as

$$P(x) \propto x^{-(\mu+1)} \ \exp\left[-\frac{\gamma}{6} \ln^4 x + \frac{\delta - 2\beta}{6} \ln^3 x - \left(\alpha - \frac{\eta}{2}\right) \ln^2 x\right] \,. \tag{33}$$

Here, we regard $d(x)$ in $\tilde{P}(x) = d(x) \, P(x)$ as a constant and denote $\mu = C_1 - C_2$. The solutions (26)–(29) and (33) satisfy Eq. (20) beyond perturbation around $R = 1$, so these are not only necessary but also sufficient. In the approximation (30), the PDF is reduced to

$$P(x) \propto x^{-(\mu+1)} \ \exp\left[-\frac{\alpha}{2} \ln^2 x\right] \,. \tag{34}$$

Figure 17 shows that the resulting PDF (34) fits correctly with the empirical sales data. In the large-scale range $(\alpha = 0)$, the PDF (34) behaves as Pareto's Law (1). The Pareto index is estimated as approximately $\mu \sim 1$ in the large-scale range $(x \geq 10^{6.2})$ of Fig. 17. In the mid-scale range, the PDF (34) behaves as the log-normal distribution (2) in the same manner as in the profits case. Applying the PDF (34) to the mid-scale range $(10^{3.8} \leq x < 10^{5.4})$ of Fig. 17, we obtain the rate-of-change parameter $\alpha = 0.65 \pm 0.04$ by using the least-squares method. This is consistent with the value estimated by the change in $t_+(x_T)$ $(\alpha = 0.68 \pm 0.03)$. From these results, we conclude that Non-Gibrat's Second Property is also confirmed by the empirical data.
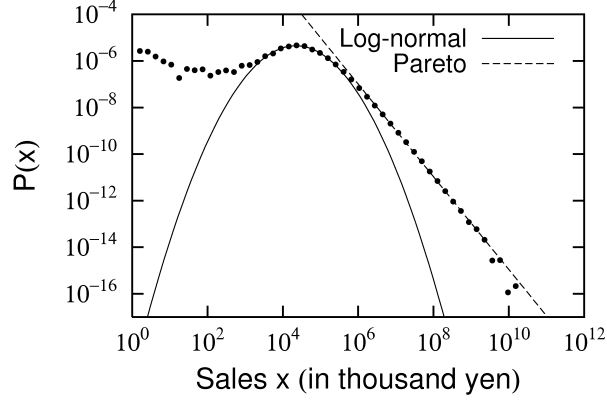
FIG. 17. A PDF of sales in the database. Pareto's Law is observed in the large-scale range ($x > 10^{6.2}$) and the log-normal distribution in the mid-scale range ($10^{3.8} \leq x < 10^{5.4}$).

CONCLUSION

In this study, we have employed exhaustive business data on Japanese firms that nearly cover not only the entire large-scale range but also the entire mid-scale range in terms of firm size. Using this newly assembled database, we first reconfirmed the previous analyses for profits data [19] as described below. In the mid-scale range, the log-normal distribution is derived from detailed balance and from Non-Gibrat's First Property. In Non-Gibrat's First Property, the probability of positive growth decreases and the probability of negative growth increases symmetrically as the initial value x_T increases. Under detailed balance, this change is uniquely reduced from the shape of the growth-rate distribution, which is linear in log-log scale.

Second, the following findings were reported with respect to sales data. Detailed balance is also observed in the mid- and large-scale ranges of sales data. The growth-rate distribution of sales has wider tails than the linear growth-rate distribution of profits in log-log scale. In the mid-scale range, while the probability of negative growth hardly changes as the initial value x_T increases, the probability of positive growth gradually decreases. This feature is different from Non-Gibrat's First Property observed in the profits data. We have approximated the growth-rate distribution with curvatures by a quadratic function. In addition, from an empirical observation, we have imposed the condition that the negative growth-rate distribution does not depend on x_T, even in the mid-scale range. Under detailed balance,

19

these approximations and conditions uniquely lead to a decrease in positive growth. We call this Non-Gibrat's Second Property. In the mid-scale range, the log-normal distribution is also derived from detailed balance and from Non-Gibrat's Second Property. These results are confirmed by the empirical data.

In this study, it was clarified that the shape of the growth-rate distribution of sales is different from that of profits. It was also demonstrated that this difference is closely related to the difference between two kinds of Non-Gibrat's Properties in the mid-scale range. The growth-rate distribution of income of firms is approximated by a linear function in log-log scale as in the profits data. The growth-rate distributions of assets, the number of employees, and personal income have wider tails than a linear function in log-log scale, as in the sales data. If we obtained exhaustive data that include the mid-scale range, Non-Gibrat's First Property would probably be observed in the income data of firms, while Non-Gibrat's Second Property would probably be observed in the assets, the number of employees, and the personal income data.

We have not determined what makes the difference between the shapes of the growth-rate distributions. However, this difference is probably related to the following factors [25]. Income and profits of firms are calculated by a subtraction of total expenditures from total sales in a rough estimate. Assets and sales of firms, the number of employees, and personal income are not calculated by any subtraction.

Let us consider the distribution of added values, the sum of which is GDP. Clearly, added values are calculated by some subtraction. If we obtained exhaustive data of added values, Non-Gibrat's First Property would certainly be observed. It has been reported that the growth-rate distribution of GDPs of countries is linear in log-log scale (for instance [35]). This report reinforces that speculation. The results in this paper should be carefully considered in cases where governments and firms discuss strategies of growth.

Finally, we consider a method to fix the inconsistency by which the rate-of-change parameter α is not estimated by the difference between $u_{\pm}(x_T)$ (32). Let us add not only the second-order term with respect to r but also a third-order term as follows:

$$\log_{10} q(r|x_T) = c(x_T) - t_+(x_T)\, r + \ln 10\ u_+(x_T)\, r^2 - \ln^2 10\ v_+(x_T)\, r^3 \qquad \text{for}\ \ r > 0 \ , \tag{35}$$

$$\log_{10} q(r|x_T) = c(x_T) + t_-(x_T)\, r + \ln 10\ u_-(x_T)\, r^2 + \ln^2 10\ v_-(x_T)\, r^3 \qquad \text{for}\ \ r < 0 \ . \tag{36}$$

In the same manner as in the previous section, under detailed balance, coefficients $t_{\pm}(x)$,

$u_\pm(x)$, and $v_\pm(x)$ are uniquely obtained as follows:

$$t_+(x) = \frac{\zeta}{5}\ln^5 x + \frac{\epsilon}{4}\ln^4 x + \frac{\gamma}{3}\ln^3 x + \frac{\beta}{2}\ln^2 x + \alpha\ln x + C_1 \ , \tag{37}$$

$$t_-(x) = -\frac{\zeta}{5}\ln^5 x + \frac{\kappa - \epsilon}{4}\ln^4 x + \frac{\theta - \gamma}{3}\ln^3 x + \frac{\delta - \beta}{2}\ln^2 x + (\eta - \alpha)\ln x + C_2 \ , \tag{38}$$

$$u_+(x) = -\frac{\zeta}{5}\ln^4 x - \frac{4\epsilon + 3\kappa}{20}\ln^3 x - (\lambda + \frac{2\gamma + \theta}{12})\ln^2 x + (\nu - \frac{\delta + \beta}{6})\ln x + C_3 \ , \tag{39}$$

$$u_-(x) = -\frac{\zeta}{5}\ln^4 x - \frac{4\epsilon - 7\kappa}{20}\ln^3 x - (\lambda + \frac{2\gamma - 5\theta}{12})\ln^2 x + (\nu + \frac{2\delta - \beta}{6})\ln x$$
$$+ C_3 + \frac{\eta}{2} \ , \tag{40}$$

$$v_+(x) = \frac{\zeta}{15}\ln^3 x + \frac{2\kappa + \epsilon}{20}\ln^2 x + \lambda\ln x + C_4 \ , \tag{41}$$

$$v_-(x) = -\frac{\zeta}{15}\ln^3 x + \frac{3\kappa - \epsilon}{20}\ln^2 x - (\lambda - \frac{\theta}{6})\ln x + C_4 + \mu \ . \tag{42}$$

By imposing the condition that the negative growth-rate distribution does not depend on x_T even in the mid-scale range, these are simplified as follows:

$$t_+(x) = \frac{\beta}{2}\ln^2 x + \alpha\ln x + C_1 \ , \qquad t_-(x) = C_2 \ , \tag{43}$$

$$u_+(x) = -\frac{\beta}{2}\ln x + C_3 \ , \qquad u_-(x) = C_3 + \frac{\alpha}{2} \ , \tag{44}$$

$$v_+(x) = C_4 \ , \qquad v_-(x) = C_4 - \frac{\beta}{6} \ . \tag{45}$$

The results in the previous sections (31) and (32) correspond to a special case $\beta = 0$, $C_4 = 0$ in Eqs. (43)–(45). In the previous section, it was difficult to estimate α by the difference in $u_\pm(x)$. In the expressions (43)–(45), this discrepancy is probably solved with a negative β. Note that Eqs. (26)–(29) cannot be reduced to Eqs. (43) and (44) in any parameterization.

It is technically difficult to estimate $t_\pm(x)$, $u_\pm(x)$, and $v_\pm(x)$ by approximating the growth-rate distribution by the cubic function (35)–(36) and to estimate β and α fitting Eqs. (43)–(45) by the least-squares method. At the same time, under the approximation by the cubic function (35)–(36), the integration $\int_0^\infty dR\ Q(R|x_T)$ converges without a cut r_c, as in the linear approximation. Because this work involves difficulties as well as advantages, we will investigate the above issues in the near future.

ACKNOWLEDGMENTS

---

* ishikawa@kanazawa-gu.ac.jp

† fujimoto@kanazawa-gu.ac.jp

‡ mizuno@ier.hit-u.ac.jp

[1] P. Bak, C. Tang and K. Wiesenfeld, Phys. Rev. Lett. 59 (1987) 381;

P. Bak, C. Tang and K. Wiesenfeld, Phys. Rev. A 38 (1988) 364.

[2] C.-K. Peng, J. Mietus, J. M. Hausdorff, S. Havlin, H. E. Stanley and A. L. Goldberger, Phys. Rev. Lett. 70 (1993) 1343.

[3] E. Bonabeau and L. Dagorn, Phys. Rev. E 51 (1995) R5220.

[4] S. Render, Eur. Phys. J. B4 (1998) 131.

[5] M. Takayasu, H. Takayasu and T. Sato, Physica A 233 (1996) 824.

[6] A. Saichev, Y. Malevergne and D. Sornette, *Theory of Zipf's law and beyond, Lecture Notes in Economics and Mathematical Systems*, p. 632 (Springer, 2009).

[7] T. Kaizoji, Physica A 326 (2003) 256.

[8] T. Yamano, Eur. Phys. J. B 38 (2004), 665.

[9] A. Ishikawa, Physica A 371 (2006) 525;

A. Ishikawa, Prog. Theor. Phys. Supple. No. 179 (2009) 103.

[10] R. N. Mantegna and H. E. Stanley, Nature 376 (1995) 46.

[11] M. H. R. Stanley, S. V. Buldyrev, S. Havlin, R. Mantegna, M. A. Salinger and H. E. Stanley, Economics Lett. 49 (1995) 453.

[12] V. Pareto, *Cours d'Economique Politique* (Macmillan, London, 1897).

[13] M. E. J. Newman, Contemporary Physics 46 (2005) 323;

A. Clauset, C. R. Shalizi and M. E. J. Newman, SIAM Review 51 (2009) 661.

[14] W. W. Badger, in B. J. West (ed.) *Mathematical Models as a Tool for the Social Science*, p. 87 (Gordon and Breach, New York, 1980).

[15] E. W. Montroll and M. F. Shlesinger, J. Stat. Phys. 32 (1983) 209.

[16] H. Aoyama, H. Iyetomi, Y. Ikeda, W. Souma and Y. Fujiwara, *ECONOPHYSICS* (Kyoritsu, Tokyo, 2008 in Japanese).

[17] M. Levy, S. Solomon, Int. J. Mod. Phys. C 7 (1996) 595;

H. Kesten, Acta Math. 131 (1973) 207;

D. Sornette, R. Cont, J. Phys. I 7 (1997) 431;

H. Takayasu, A.-H. Sato, M. Takayasu, Phys. Rev. Lett. 79 (1997) 966.

[18] Y. Fujiwara, W. Souma, H. Aoyama, T. Kaizoji and M. Aoki, Physica A 321 (2003), 598;

Y. Fujiwara, C. D. Guilmi, H. Aoyama, M. Gallegati and W. Souma, Physica A 335 (2004), 197.

[19] A. Ishikawa, Physica A 367 (2006) 425;

A. Ishikawa, Physica A 383 (2007) 79.

[20] M. Riccaboni, F. Pammolli, S. V. Buldyrev, L. Ponta and H. E. Stanley, Proc. Natl. Accad. Sci. USA 105 (2008) 19595.

[21] R. Gibrat, *Les inegalites economiques* (Sirey, Paris, 1932).

[22] J. Sutton, J. Econo. Lit. 35 (1997) 40.

[23] K. Okuyama, M. Takayasu and H. Takayasu, Physica A 269 (1999) 125.

[24] A. Ishikawa, Physica A 363 (2006) 367.

[25] A. Ishikawa, Economics 3 –Special Issue Reconstructing Macroeconomics, 2009–11.

[26] L. A. N. Amaral, S. V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M. A. Salinger, H. E. Stanley, and M. H. R. Stanley, J. Phys. I France 7 (1997) 621.

[27] K. Matia, D. Fu, S. V. Buldyrev, F. Pammolli, M. Riccaboni and H. E. Stanley, Europhys. Lett. 67 (2004) 498.

[28] D. Fu, F. Pammolli, S. V. Buldyrev, M. Riccaboni, K. Matia, K. Yamasaki and H. E. Stanley, Proc. Natl. Acad. Sci. 102 (2005) 18801.

[29] S. V. Buldyrev, J. Growiec, F. Pammolli, M. Riccaboni and H. E. Stanley, J. Eur. Economic Association 5 (2-3) (2007) 574.

[30] H. Aoyama, Ninth Annual Workshop on Economic Heterogeneous Interacting Agents (WEHIA 2004);

H. Aoyama, Y. Fujiwara and W. Souma, The Physical Society of Japan 2004 Autumn Meeting.

[31] H. Aoyama, H. Iyetomi, Y. Ikeda, W. Souma and Y. Fujiwara, *Pareto Firms* (Nihon Keizai Hyouronsha, Tokyo, 2007 in Japanese).

[32] H. Takayasu, New way of financing firms based on the fat-tailed distribution of growth rate. APFA7 & Tokyo Tech. Hitotsubashi Interdisciplinary Conference (2009).

[33] Research Institute of Economy, Trade and Industry, IAA (RIETI), http://www.rieti.go.jp/en/index.html.

[34] TOKYO SHOKO RESEARCH, LTD., http://www.tsr-net.co.jp/.

[35] D. Canning, L. A. N. Amaral, Y. Lee, M. Meyer and H. E. Stanley, Economics Lett. 60 (1998) 335.