数理解析研究所講究録 1264 巻 2002 年 122-144

The Size Distribution of Firms, Economies of Scale and Growth

Hideyuki Adachi

Department of Economics, Kobe University, Kobe, 657, Japan

Abstract

The size distribution of firms in each industry will usually be highly skew, and empirical evidence shows that it is approximated closely by the Pareto distribution. In this paper we make an attempt to explain why the Pareto law applies to the size distribution of firms based on their innovation and investment behavior, and then develop a model of economic growth that takes into account this empirical law. First, we show that the Pareto distribution of firms is generated under the assumption that firms acquire the technology of operating efficiently on a larger scale through learning by doing, and expand their scale of operation through the accumulation of capital induced by profitability. Then, we set up a model of economic growth that is based on the Pareto distribution of firms and economies of scale. In our model the growth rate is determined endogenously, and it exhibits scale effects with respect to savings and population. Our model is different from the neoclassical growth model or the recently developed endogenous growth models in that it takes into account the size structure of firms, and it yields quite realistic predictions.

1. Introduction

Empirical laws are rare in economics, and one of such laws is the regular pattern of some statistical distributions, such as the distribution of persons according to the level of income or of business firms according to some measurement of size such as sales or the number of workers. Many of these distributions conform to the so-called the law of Pareto. Many economists attempted to explain the mechanisms that generate the Pareto distributions by constructing models with stochastic processes. Simon (1955), Champernowne (1953), Wold and Whittle (1957), Steindl (1965), etc. may be mentioned as pioneers of such models. The most ingenious model among them is the one developed by Simon (1955), which explains the Pareto distributions based on two simple and meaningful assumptions: one is 'the law of proportionate effect, and the other is the constancy of new entry. When his model is applied to the size distribution of firms, however, it is not clear how those assumptions are related to firms' behavior; Besides, there is no work, as far as I know, that make use of this interesting empirical evidence on the size distribution of firms to analyze macroeconomic problem such as economic growth or income distribution.

The purpose of this paper is first to explain why the size distribution of firms is approximated by the Pareto distribution based on the innovation and investment behavior of firms, and secondly to develop a model of economic growth that takes into account this empirical law. In our model we assume that new firms start their operation from the minimum size, because they lack not only the necessary know-how to operate efficiently at larger size but also sufficient finance to start on a large scale. They gradually acquire the technology of operating efficiently on a larger scale through learning by doing, and expand their scale of operation through accumulation of capital induced by profitability. We show the Pareto distribution of firms is generated under such assumptions.

Using this size distribution function and the learning function, we set up a model of economic growth embodying economies of scale. In this model the growth rate is determined endogenously, and it exhibits scale effect with respect to savings and population growth. Our model is different from the Solow growth model or the recently developed endogenous growth models in that it takes into account the size structure of firms.

The paper is organized as follows. Section 2 reviews the Simon's model and the generalization of it by Sato. Section 3 introduces learning by doing model to explain growth of firms. Section 4 discusses the determination of investment of firms, and shows that the Pareto distribution is generated through the process of learning by doing and capital accumulation. In Section 5, we construct a macroeconomic model based on the Pareto law and the learning by doing hypothesis, and analyze income distribution in this model. In Section 6, we extend it to a growth model. Section 7 analyzes the steady-state properties of this model. It is shown that the steady growth equilibrium exhibits scale effect, but it is unstable. In Section 8, we consider the substitutability between capital and labor, show that the steady growth equilibrium becomes stable in that case.

2. The Size Distribution of Firms

The size distributions of firms in U.S. and Germany are illustrated in the Appendix of Steindl's book (1965). [1] They approximate the Pareto distribution, especially in the upper tail, which is given by

$$N(k) = Ak^{-\rho}. \tag{2.1}$$

Here, k represents the size of firms, $N(k)$ the number of firms with the size in excess of k, and ρ is called the Pareto coefficient. The size of firms is measured by sales, capital or employment depending on the availability of data. The above equation implies that the number of firms with the size in excess of k, plotted against k on logarithmic paper, is a straight line. The size distribution of firms in the Japanese manufacturing industry, as shown by Fig.1, is also a beautiful illustration of the Pareto law. It is almost entirely a straight line on the logarithmic paper. [2]

The Pareto distribution is observed not only in the size distribution of firms but in many other fields, such as distributions of income by size, distributions of scientists by number of papers published, distributions of cities by population. [3] Why such a regular pattern is observed in many fields is a big puzzle. Many economists have challenged to reveal this puzzle. Among them, the solution given by Simon (1955) seems to me the simplest and the most ingenious.

Let us first review the Simon's model. His model was designed for a non-economic problem, namely the distribution of words in a book. Suppose that we read a book, classifying words that appear successively. Some words appear more often than others. Let the total number of words in a book already run through reached K. We designate by $f(k,K)$ the number of different words that have appeared k times. Then, we must have

$$\sum_{k=1}^{K} kf(k,K) = K. \tag{2.2}$$

Now, Simon makes the following two assumptions:

Assumption 1: The probability that the $(K+1)$-*st* word is a word that has already appeared exactly k times is proportional to $kf(k,K)$——that is, to the total number of occurrence of all the words that have appeared exactly k times.

Assumption 2: There is a constant probability, α, that the $(K+1)$-*st* word be a new word——a word that has not occurred in the first K words.

The first assumption is called the law of proportional effect, which was proposed by Gibrat (1930) to derive the log-normal distribution. With this assumption, the expected number of words that would have appeared k times after the $(K+1)$-*st* word has been drawn is determined by

$$E[f(k,K+1)] = f(k,K) + L(K)\{(k-1)f(k-1,K) - kf(k,K)\}, \quad k = 2,\cdots,K+1 \tag{2.3}$$

where $L(k)$ is the proportionality factor of the probabilities. The second assumption implies that the probability of a new entry of a word is constant. This assumption together with the first one gives the following equation:

$$E[f(1,K+1)] = f(1,K) - L(K)f(1,K) + \alpha. \tag{2.4}$$

Simon is concerned with "steady-state" distributions, so he replaces the expected values in the above two equations by the actual frequencies. In other words, the expectation operator E is dropped from (2.3) and (2.4) in order to have the steady state distribution. The definition of the steady-state distribution is given by

$$\frac{f(k,K+1)}{f(k,K)}=\frac{K+1}{K} \quad \text{for all } k \text{ and } K. \tag{2.5}$$

This means that all the frequencies grow proportionately with K, and maintain the same relative size. The relative frequencies denoted by $f^*(k)$ may be defined as

$$f^*(k)=\frac{f(k,K)}{\alpha K}, \tag{2.6}$$

where αK is the total number of different words.

With the above assumptions and the definition of the steady-state distribution, Simon shows that the relative frequency of different words in the steady state, which is denote by $f^*(k)$, is independent of K, and becomes as

$$f^*(k)=\frac{(k-1)(k-2)\cdots 2\bullet 1}{(k+\nu)(k+\nu-1)\cdots(2+\nu)}f^*(1)=\frac{\Gamma(k)\Gamma(\nu+2)}{\Gamma(k+\nu+1)}f^*(1) \tag{2.7}$$

Here,

$$\nu=\frac{1}{1-\alpha}, \qquad f^*(1)=\frac{1}{2-\alpha} \tag{2.8}$$

The expression (2.7) is a steady-state solution to equations (2.3) and (2.4), since it satisfies the latter two equations without the expectation operator E. Simon called the expression (2.7) the *Yule distribution.*

From the well-known asymptotic property of the Gamma function, we have

$$\Gamma(k)/\Gamma(k+\nu+1)\to k^{-(\nu+1)} \quad \text{as} \quad k\to\infty. \tag{2.9}$$

Hence, from (2.7), we have

$$f^*(k)\to\Gamma(\nu+2)f^*(1)k^{-(\nu+1)}=Ak^{-(\nu+1)} \quad \text{as} \quad k\to\infty. \tag{2.10}$$

We can confirm that $f^*(k)$ is a proper distribution function. For we have

$$\sum_{k=1}^{\infty}kf^*(k)\to\Gamma(\nu+2)f^*(1)\sum_{k=1}^{\infty}k^{-\nu}, \tag{2.11}$$

and this expression is convergent if $\nu>1$.

Thus, as (2.9) shows, the steady state distribution $f^*(k)$ obtained under the above two assumptions is identical with the Pareto distribution for large values of k. The value of the Pareto coefficient ν is determined by the probability of a new entry α according to (2.8).

It is easy to interpret Simon's model explained above in terms of the size distribution of firms. In this context, we may interpret K as the total assets accumulated in the economy, and $f(k,K)$ as the number of firms with assets k. The parameter α is the ratio of the assets of newly entering firms to the increment of assets of all firms above a certain minimum. The newly entering firms are those that pass beyond this minimum in the period in question. The greater the contribution of new firms to the total growth of assets is, the greater will be the Pareto coefficient. The greater Pareto coefficient implies less inequality of the distribution of firms.

K. Sato (1970) generalized Simon's model to include the case where the law of proportionate effect does not apply. Instead of *Assumption 1* above, he assumes the following:

Assumption 1': The probability that the $(K+1)$-*st* word is a word that has already appeared exactly k times is proportional to $(ak+b)f(k,K)$ under the condition that it also satisfies

$$\sum_{k=1}^{K}(ak+b)f(k,K)=\sum_{k=1}^{K}kf(k,K)=K. \tag{2.12}$$

With this assumption together with *Assumption 2* above, he shows that the steady-state distribution becomes as

$$f^{*}(k)=\frac{\Gamma\left(k+\frac{b}{a}\right)\Gamma\left(\frac{\nu+b}{a}+2\right)}{\Gamma\left(k+\frac{\nu+b}{a}+1\right)}f^{*}(1). \tag{2.13}$$

Here, $a+b>0$ is required for this value to be finite. This distribution becomes asymptotically as follows:

$$f^{*}(k)\to\left(k+\frac{b}{a}\right)^{-\frac{\nu}{a}-1} \quad \text{as} \quad k\to\infty. \tag{2.14}$$

This is called Pareto distribution of the second kind. This distribution function when plotted on logarithmic paper, is not exactly a straight line.

However, since $1+(b/ak)\to 1$ as $k\to\infty$ for any given value of b/a, the steady-state distribution (2.14) is asymptotic to Pareto distribution of the first kind, that is,

$$f^{*}(k)\to k^{-\frac{\nu}{a}-1} \quad \text{as} \quad k\to\infty. \tag{2.15}$$

The smaller the value of b/a, the more closely the steady-state distribution (2.14) is approximated by (2.15).

It is shown that a and b must satisfy the following relation with α:

$$b=\frac{1-a}{\alpha}. \tag{2.16}$$

From this relation, it is obvious that

$$a\gtrless 1 \quad \text{according as} \quad b\lessgtr 0. \tag{2.17}$$

The expected growth rate of k is proportional to $(ak+b)/k$, that is,

$$E\left(\frac{\Delta k}{k}\right)=L(K)\frac{ak+b}{k}=L(K)\left(a+\frac{1-a}{\alpha k}\right), \qquad k\in[1,K] \tag{2.18}$$

where $L(K)$ is the proportionality factor. The proportionality factor depends on the total number of words K. Equation (2.18) implies that the expected growth rate of k increases or decreases with k, depending on whether $a>1$ or $a<1$. When $a=1$, the expected growth rate of firms is independent of size. Thus, Sato obtains the following proposition:

Proposition 1: Under the *assumptions 1'* and *2* above, the size distribution is asymptotic to Pareto distribution, and following three cases occur.

(a) The case of proportionate growth ($a=1$ and $b=0$): In this case, the relative growth rate is independent of size. The Pareto coefficient is $\nu=1/(1-\alpha)$ as Simon demonstrated.

(b) The case of size-impeded growth ($a<1$ and $b>0$): In this case, the growth rate is stochastically proportional to $a+b$ at $k=1$, and proportionately declines towards a as $k\to\infty$. The Pareto coefficient ν/a exceeds ν.

(c) The case of size-induced growth ($a>1$ and $b<0$): In this case, the growth rate is stochastically proportional to $a+b$ at $k=1$, and proportionately increases towards a as $k\to\infty$. The Pareto coefficient ν/a is less than ν.

3. Learning by Doing and Economies of Scale

In the neoclassical theory of the firm, it is assumed that the U-shaped curve, *LAC*, illustrated in Figure 2 is the long-run average cost curve of all firms in a particular industry, freely available to all including to potential new entrants. It is not by empirical observation but by the assumption of perfect competition that the theory requires the long-run average cost curve to be U-shaped. If it is U-shaped, the size distribution of firms is expected to be a normal distribution around the optimum size at which the long-run average cost is minimum. But, as is shown by many data, the size distributions of firms in Japan as well as in U.S. and Germany are highly skewed, being approximated closely by the Pareto distribution. This implies that the neoclassical theory of the firm is inconsistent with empirical observations. [4]

In this section, we develop a different model of firms, which explains consistently the observed size distribution of firms—the Pareto distribution. Considering that the Pareto distribution is derived from *Assumptions 1* (or *1'*) and *2* above, our model should be consistent with those assumptions. In the context of size distributions of firms, *Assumption 1'* and *Assumption 2* may be restated as follows.

Assumption 1': When the aggregate stock of capital in the economy, K, is increased by one, the probability of a firm with size k being expanded by one is proportional to $(ak+b)f(k,K)$ under the condition that it also satisfy

$$\sum_{k=1}^{K}(ak+b)f(k,K)=\sum_{k=1}^{K}kf(k,K)=K. \qquad (3.1)$$

Assumption 2: When the aggregate stock of capital in the economy, K, is increased by one, the probability of this increment to be apportioned to newly entering firms is α.

Assumption 1' implies that the expected growth of firms with size k is proportional to $a+(b/k)$, while *Assumption 2* implies that the ratio of the capital stock of newly entering firms to the increment of total capital is α. These parameters a, b, α must satisfy (2.16). Depending on whether $a>1\,(b<0)$ or $a<1\,(b>0)$, the expected growth of firms increases or decreases with size k. When $a=1\,(b=0)$, the expected rate of growth of firms is independent of their size.

Following *Assumption 2*, we assume that new firms start their operations from the minimum size. There are two reasons to justify this assumption. The first is that new entrants do not have the necessary know-how to operate efficiently at larger sizes. The second is that new entrants usually cannot have sufficient finance to start on a large scale. But once they have acquired the necessary technology and finance, they will expect to grow in size. Firms with same size do not necessarily grow at the same rate. Profitable firms tend to grow faster than unprofitable firms. Their eventual growth will depend on successful experience—learning by doing—and the accumulation of profits, both of which take time.

Most firms believe that there are economies of scale to be gained, if they acquire necessary technology and necessary finance. In order to expand successfully in size, however, a firm has to master technology of operating efficiently on a large scale, and it is through a process of learning by doing that a firm can master such technology. Arrow (1962) formulated a model of economic growth based on the hypothesis of learning by doing. We follow him to explain productivity growth of firms. We assume that learning by doing worked through each firm's investment. Specifically, an increase in a firm's capital stock leads to an increase in its stock of knowledge, and therefore to its growth of productivity. But the rate of growth in productivity may be different among firms even with the same size. Some firms improve their efficiency better than others. Thus, though each firm follows a different path in learning by doing, we assume that the learning function of a typical firm with capital stock k is expressed as follows:[5]

$$\frac{l(k)}{k} = \gamma(k), \qquad \gamma'(k) < 0, \qquad k \in [1, K] \tag{3.2}$$

$$\frac{x(k)}{k} = \delta(k), \qquad \delta'(k) \geq 0, \qquad k \in [1, K]. \tag{3.3}$$

The notations are as follows: $l(k) \equiv$ amount of labor used in production by a typical firm with size k, $x(k) \equiv$ output capacity of a typical firm with size k. It is assumed that $\gamma(k)$ is a decreasing function, while $\delta(k)$ is a non-decreasing function. In this case, an expansion of the typical firm with size k definitely leads to a reduction in costs of production at any given wages and rental value of capital goods, since they save labor input per unit of output without increasing capital input per unit of output by expanding the size.

To simplify the analysis without losing reality, we will specify these functions as follows:

$$\gamma(k) = ck^{\lambda-1}, \quad \text{where} \quad 0 < \lambda < 1, \qquad k \in [1, K] \tag{3.4}$$

$$\delta(k) = dk^{\mu-1}, \quad \text{where} \quad \mu \geq 1, \qquad k \in [1, K]. \tag{3.5}$$

Then, we have

$$l(k) = ck^{\lambda}, \tag{3.6}$$

$$x(k) = dk^{\mu}. \tag{3.7}$$

These relations fit quite well to the data of Japanese manufacturing. [6]

4. Profitability and Expansion of Firms

The incentive of firms to expand arises from the prospect of improving their profitability by increasing their scale of operation. The accumulated profits can be used for further expansion, either directly or as security for raising external finance. Therefore, the rate of profit is a key variable as the determinants of the expected growth rate of firms in each size. Assuming that the learning function of a typical firm with size k is given by (3.4) and (3.5), we can express its profit rate as follows:

$$e(k) = \frac{x(k) - wl(k)}{k} = \frac{dk^{\mu} - wck^{\lambda}}{k} = dk^{\mu-1} - wck^{\lambda-1}, \qquad k \in [1 < K]. \tag{3.8}$$

Here, w denote the wage rate, which is assumed here to be the same for any size of firms.

In reality, the average wage per worker tends to be an increasing function of size of firm, although not to the same degree as decreases of labor input. One reason for this is that larger firms will usually have a more detailed division of labor, with a larger proportion of higher-paid skilled or managerial workers. Another reason is that trade unions are usually more powerful in larger firms, and may succeed in extracting part of extra profits created by economies of scale. Because of these reasons, we assume that the average wage per worker increases with size of firms as follow:[7]

$$w(k) = w(1)k^{\omega}, \quad \text{where} \quad \omega > 0, \quad k \in [1, K]. \tag{3.9}$$

To simplify the following analysis we assume that $\mu = 1$ in equation (3.7). This assumption is roughly supported by actual data.[8] With this assumption and (3.8), the rate of profit of a typical firm with size k becomes as follows:

$$e(k) = d - w(1)ck^{\lambda+\omega-1}, \qquad k \in [1, K]. \tag{3.10}$$

It is obvious from this function that, if $\lambda + \omega = 1$, the rate of profit is constant irrespective of firm size k. If $\lambda + \omega \neq 1$, on the other hand, the rate of profit increases or decreases with firm size k, depending on whether $\lambda + \omega < 1$ or $\lambda + \omega > 1$.[9]

As is mentioned above, the incentive of firms to expand arises from the prospect of improving their profitability by increasing their scale of operation. So, we assume that the expected rate of growth of a typical firm with size k depends on the rate of profit earned by that firm, $e(k)$. For simplicity, we assume it to be expressed by the following linear equation:

$$E\left(\frac{\Delta k}{k}\right) = M(K)\{\tau + \xi e(k)\}, \quad (\tau > 0, \xi > 0), \qquad k \in [1, K], \tag{3.11}$$

where $M(K)$ is the proportionality factor that depends on total capital stock, K.

Substituting (3.10) into (3.11), we can express equation (3.11) as follows:

$$E\left(\frac{\Delta k}{k}\right) = M(K)[\tau + \xi\{d - w(1)ck^{\lambda+\omega-1}\}] = M(K)(p - qk^{\lambda+\omega-1}), \qquad k \in [1, K]. \tag{3.12}$$

Here, $p \equiv \tau + \xi d$ and $q \equiv \xi w(1)c$, which are positive constants.

First, consider the case where $\lambda + \omega = 1$. In this case, the expected rate of growth becomes as

$$E\left(\frac{\Delta k}{k}\right) = M(K)(p - q), \qquad k \in [1, K], \tag{3.13}$$

where $p - q$ is constant. In other words, the relative growth rate of firms is independent of size k. This case corresponds to (a) in *Proposition 1,* and we have Pareto distribution.

Next, let us consider the case where $\lambda + \omega \neq 1$. In this case, as is obvious from (3.12), the expected growth of firms increases or decreases with size k depending on whether $\lambda + \omega < 1$ or $\lambda + \omega > 1$. In order to relate (3.12) to *Proposition 1* by Sato, let us rewrite equation (3.12) as

$$E(\Delta k) = M(k)(pk - qk^{\lambda+\omega}), \qquad k \in [1, K] \tag{3.14}$$

and linearize it around k^*. Then, we get

$$E(\Delta k) = M(k)[p(k-k^*) - (\lambda+\omega)q(k-k^*)]$$
$$= M(K)(p-q)\frac{p-(\lambda+\omega)q}{p-q}(k-k^*), \qquad k \in [1, K]. \tag{3.15}$$

This equation can be rewritten as

$$E(\Delta k) = M(K)(p-q)(ak+b), \tag{3.16}$$

where

$$a \equiv \frac{p-(\lambda+\omega)q}{p-q}, \quad b \equiv -\frac{p-(\lambda+\omega)q}{p-q}k^*. \tag{3.17}$$

Equation (3.16) implies that the expected increase in assets of a firm with size k is proportional to $ak+b$, which is exactly the same as the condition stated in *Assumption 1'* above. In addition, we assume that a and b defined by (3.17) satisfy (2.16). Then, the value of k^* is determined as

$$k^* = \frac{a-1}{\alpha a}. \tag{3.18}$$

So, a and b are determined by the parameters given in our model.

Comparing the above results with *Proposition 1* by Sato, we get the following proposition.

Proposition 2: Suppose that new firms are being born in the smallest-size class, and that they account for a constant rate α of the growth in total assets. Suppose also that a typical firm of each size class masters technology of operating more efficiently on a larger scale through learning by doing as represented by (3.6) and (3.7), and that its rate of expansion depends on the rate of profit as expressed by (3.11). Then, the size distribution of firms converges to the Pareto distribution of the form (2.14). Depending on the value of $\lambda+\omega$, we can distinguish the following three cases:

(a) If $\lambda+\omega=1$, then $a=1$ and $b=0$. In this case, the growth rate is independent of size, and the Pareto coefficient is equal to $\nu = 1/(1-\alpha)$.

(b) If $\lambda+\omega<1$, then $a>1$ and $b<0$. In this case, the growth rate increases with size, and the Pareto coefficient is less than ν.

(c) If $\lambda+\omega>1$, then $a<1$ and $b>0$. In this case, the growth rate decreases with size, and the Pareto coefficient exceeds ν.

5. Determinants of Income Distribution

In the previous sections we were concerned with the behavior of firms operating under potential economies of scale in an industry, and showed that the size distribution of firms is approximated by the Pareto distribution under quite realistic assumptions about the technology and investment behavior of firms. In this section we will turn to the analysis of the whole economy. It is assumed that, when industries are aggregated, there are persistent economies of scale over the whole range of firm sizes. While economies of scale in one industry may be limited, in another they are more extensive, and they extend right up to the largest observed size of firms in some industry. Thus, we may assume that the Pareto function applies to size distribution of firms in the whole economy. We also assume that the learning function of the form described by (3.6) and (3.7) is still applicable when we consider the behavior of the whole economy.

If we assume that the size distribution of firms is of the Pareto form over its entire range, we can express it by the frequency function as

$$n(k) = \rho A k^{-(\rho+1)}, \qquad (\rho > 1, \quad A > 0), \tag{5.1}$$

where k represents the size of firm measured by its capital stock, and $\rho = v/a$.

Suppose that the minimum size firm has capital stock k_0, and the maximum size firm k_T. Then, the total number of firms is given by

$$N(k_0, k_T) = \int_{k_0}^{k_T} n(k)dk = A({k_0}^{-\rho} - {k_T}^{-\rho}). \tag{5.2}$$

If we denote the ratio of k_T to k_0 by m, we have

$$k_T = mk_0. \tag{5.3}$$

We call m 'size ratio' in the following. Using this notation, we can rewrite (5.2) as follows:

$$N(k_0, k_T) = A(1 - m^{-\rho}){k_0}^{-\rho}. \tag{5.4}$$

Similarly, the total stock of capital is given by

$$K(k_0, k_T) = \int_{k_0}^{k_T} kn(k)dk = \frac{\rho A}{\rho - 1}(1 - m^{1-\rho}){k_0}^{1-\rho}. \tag{5.5}$$

Taking into account (3.6) and (3.7), we can also calculate the total employment and total output as follows:

$$L(k_0, k_T) = \int_{k_0}^{k_T} l(k)n(k)dk = \frac{\rho A c}{\rho - \lambda}(1 - m^{\lambda-\rho}){k_0}^{\lambda-\rho}, \tag{5.6}$$

$$X(k_0, k_T) = \int_{k_0}^{k_T} x(k)n(k)dk = \frac{\rho A d}{\rho - \mu}(1 - m^{\mu-\rho}){k_0}^{\mu-\rho}. \tag{5.7}$$

We assume here that the total output is defined by value added. In the following, we deal with the case where $\mu = 1$. In this case, (5.7) becomes as

$$X(k_0, k_T) = \frac{\rho A d}{\rho - 1}(1 - m^{1-\rho}){k_0}^{1-\rho} = dK(k_0, k_T). \tag{5.8}$$

Suppose that the minimum size firms (or we may call them "marginal firms") set product price with mark-up factor β on wage costs. We assume that marginal firms are under perfect competition, so that β is determined at the level that just covers capital costs. Then, the real wage rate of a typical marginal firm is given by

$$w(k_0) = \frac{1}{\beta}\frac{x(k_0)}{l(k_0)} = \frac{d}{\beta c}k_0^{\mu-\lambda}. \tag{5.9}$$

We also assume that the average wage per worker rises with size of firms, as is shown by (3.9). When the minimum size of firms is k_0, (3.9) is rewritten as

$$w(k) = w(k_0)\left(\frac{k}{k_0}\right)^{\omega}. \tag{5.10}$$

All the original entrants into the industry are small enterprise of minimum size. They will grow by improving their technology through experience. As successful firms expand their scale, they will, on the average, be able to reduce their costs by exploiting economies of scale. As long as $\lambda+\omega<1$ and $\mu\geq 1$, the larger firms attain more favorable profit margins than smaller firms.

From (5.6), (5.7), (5,9) and (5.10) the aggregate share of wages in value added becomes as

$$S_w = \frac{\rho-1}{\beta(\rho-\lambda-\omega)}\frac{1-m^{\lambda+\omega-\rho}}{1-m^{1-\rho}}. \tag{5.11}$$

Thus, the aggregate wage share in this model is determined by the Pareto coefficient ρ, scale parameters λ, ω, ρ, the size ratio m, and mark-up factor, β. This theory of income distribution is quite different from the orthodox marginal productivity theory. It can be shown straightforwardly that the aggregate wage share depends on those parameters or variables in the following way.

$$\frac{\partial S_w}{\partial \rho}>0,\quad \frac{\partial S_w}{\partial \lambda}>0,\quad \frac{\partial S_w}{\partial \omega}>0\quad \frac{\partial S_w}{\partial m}<0,\quad \frac{\partial S_w}{\partial \beta}<0. \tag{5.12}$$

6. A Model of Economic Growth with Economies of Scale

In this section, we construct a growth model to examine the dynamics of aggregate variables obtained above. Taking the time derivatives of equations (5.4), (5.5), (5.6) and (5.8), we can rewrite them in terms of the growth rates as follows:

$$\hat{N} = \hat{A} + \frac{\rho}{m^{\rho}-1}\hat{m} - \rho\hat{k}_0, \tag{6.1}$$

$$\hat{K} = \hat{A} + \frac{\rho-1}{m^{\rho-1}-1}\hat{m} - (\rho-1)\hat{k}_0, \tag{6.2}$$

$$\hat{L} = \hat{A} + \hat{c} + \frac{\rho-\lambda}{m^{\rho-\lambda}-1}\hat{m} - (\rho-\lambda)\hat{k}_0, \tag{6.3}$$

$$\hat{X} = \hat{A} + \hat{d} + \frac{\rho - 1}{m^{\rho-1} - 1}\hat{m} - (\rho - 1)\hat{k}_0, \qquad (6.4)$$

where $\hat{y} \equiv \dot{y} / y$ for any given variable y. Thus, the growth rate of the number of firms, $\hat{N}$, and the growth rate of capital, $\hat{K}$, are explained by the shifting rate of the Pareto curve, $\hat{A}$, the rate of increase in the size ratio, $\hat{m}$, and the growth rate of the minimum size firms, $\hat{k}_0$. The growth rate of labor employment, $\hat{L}$, depends not only on $\hat{A}$, $\hat{m}$ and $\hat{k}_0$ but also on $\hat{c}$, which represents the rate of change in labor input per unit of capital caused by exogenous technological change. As is obvious from (3.6), a decrease in c leads to a reduction in labor input per unit of capital for every size class of firms. Therefore, $\hat{c}$ represents technological change affecting every size class of firms, and normally takes negative value.

As mentioned above, we assume that new entrants start their operation at the minimum size k_0, and that the proportion α of the increment in total capital, ΔK, is apportioned to the new firms. In other words, we have

$$\alpha = \frac{k_0 \Delta N}{\Delta K}, \qquad (6.5)$$

which is rewritten as

$$\hat{N} = \alpha \frac{1}{k_0} \frac{K}{N} \hat{K} \qquad (6.6)$$

Substituting from (5.4) and (5.5), and taking into account the relation $\rho = 1/a(1-\alpha)$, we obtain the following relationship between the growth rate of capital and the growth rate of the number of firms:

$$\hat{N} = \frac{\rho - (1/a)}{\rho - 1} \frac{m^\rho - m}{m^\rho - 1} \hat{K}. \qquad (6.7)$$

We must have $\rho > 1/a$, as long as α is positive. Substituting this equation into (6.1), we can express the shifting rate of the Pareto curve as follows:

$$\hat{A} = \frac{\rho - (1/a)}{\rho - 1} \frac{m^\rho - m}{m^\rho - 1} \hat{K} - \frac{\rho}{m^\rho - 1} \hat{m} + \rho \hat{k}_0 \qquad (6.8)$$

We assume that labor grows at a constant rate, n, and we consider the case of full employment in the following analysis. Thus, we have

$$\hat{L} = n. \qquad (6.9)$$

To complete the model, we have to specify the equation for the capital accumulation. We assume here that a fraction s_p of profits and a fraction s_w of wages are saved and devoted to investment, and that s_p is larger than s_w.[10] We also assume that there is no depreciation of capital. Then, the growth rate of capital is expressed by the following equation:

$$\hat{K} = \frac{X}{K}[s_p(1-S_w) + s_w S_w], \tag{6.10}$$

where S_w is the wage share defined by (5.10). It is a decreasing function of m as is shown by (5.11), so we denote it as $S_w(m)$. In view of (5.8), we have $X/K = d$. In the following analysis, we assume d to be constant. Then, equation (6.10) is rewritten as

$$\hat{K} = d[(s_p - s_w)\{1 - S_w(m)\} + s_w], \text{ where } s_p > s_w \geq 0 \text{ and } S_w'(m) < 0. \tag{6.11}$$

Thus, the growth rate of capital $\hat{K}$ is an increasing function of m. Denoting it as $\hat{K}(m)$ for notational convenience, we have $\hat{K}'(m) > 0$.

Now, our model consists of 7 equations [*i.e.,* (6.2) through (6.4), (6.7), (6.8), (6.9), and (6.11)], which includes 7 variables [*i.e.,* N, X, L, K, A, m, k_0]. This complete model can be reduced to the system consisting of two equations as follows. Substituting (6.8) into (6.2) yields

$$\left(\frac{\rho-1}{m^{\rho-1}-1} - \frac{\rho}{m^{\rho}-1}\right)\hat{m} + \hat{k}_0 = \left(1 - \frac{\rho-(1/a)}{\rho-1}\frac{m^{\rho}-m}{m^{\rho}-1}\right)\hat{K}(m). \tag{6.12}$$

This equation represents the equilibrium condition for the capital goods market. Similarly, substituting (6.8) and (6.9) into (6.3) yields

$$\left(\frac{\rho-\lambda}{m^{\rho-\lambda}-1} - \frac{\rho}{m^{\rho}-1}\right)\hat{m} + \lambda\hat{k}_0 = (n-\hat{c}) - \frac{\rho-(1/a)}{\rho-1}\frac{m^{\rho}-m}{m^{\rho}-1}\hat{K}(m). \tag{6.13}$$

This equation represents the equilibrium condition for the labor market. The system consisting of equations (6.12) and (6.13) includes two variables, m and k_0, so that it is a complete system.

Eliminating $\hat{k}_0$ from (6.12) and (6.13), we obtain the dynamic equation for $\hat{m}$:

$$\hat{m} = \frac{1}{D(m)}[\Phi(m)\hat{K}(m) - (n-\hat{c})], \tag{6.14}$$

where,

$$D(m) = \lambda\left(\frac{\rho-1}{m^{\rho-1}-1} - \frac{\rho}{m^{\rho}-1}\right) - \left(\frac{\rho-\lambda}{m^{\rho-\lambda}-1} - \frac{\rho}{m^{\rho}-1}\right), \tag{6.15}$$

$$\Phi(m) = \lambda + (1-\lambda)\frac{\rho-(1/a)}{\rho-1}\frac{m^{\rho}-m}{m^{\rho}-1}. \tag{6.16}$$

It can be proved that there exists $\overline{m}$ such that

$$D(m) > 0 \quad \text{for} \quad m > \overline{m}. \tag{6.17}$$

The magnitude of $\overline{m}$ is sufficiently small compared to the relevant range of m, so that we may assume that $D(m) > 0$ always holds in our model. [11] The function $\Phi(m)$, on the other hand, has the following properties.

$$\Phi(m) > 0, \quad \Phi'(m) > 0. \tag{6.19}$$

Substituting (6.14) into ((6.12) and solving it with respect to $\hat{k}_0$, we have

$$\hat{k}_0 = \frac{1}{D(m)}[\Psi(m)(n-\hat{c}) - \Omega(m)\hat{K}(m)], \tag{6.20}$$

where

$$\Psi(m) = \frac{\rho-1}{m^{\rho-1}-1} - \frac{\rho}{m^{\rho}-1}, \tag{6.21}$$

$$\Omega(m) = \left(\frac{\rho-\lambda}{m^{\rho-\lambda}-1} - \frac{\rho}{m^{\rho}-1}\right) + \frac{\rho-(1/a)}{\rho-1}\frac{m^{\rho}-m}{m^{\rho}-1}\left(\frac{\rho-1}{m^{\rho-1}-1} - \frac{\rho-\lambda}{m^{\rho-\lambda}-1}\right). \tag{6.22}$$

It can be shown that [12]

$$\Psi(m) > 0, \text{ and } \Omega(m) > 0. \tag{6.23}$$

Thus, equation (6.14) determines the dynamic path of m starting from its initial value. Corresponding to the path of m, the growth rate of capital ($\hat{K}$) and the minimum size of firms (k_0) are determined. The growth rate of output ($\hat{X}$) is equal to the growth rate of capital ($\hat{K}$) under the assumption of fixed coefficient. This assumption will be relaxed later.

7. The Steady Growth and its Instability

In this section, we examine the properties of the steady state of the above model. In view of the dynamic equation (6.14), the steady growth equilibrium is attained at m^* that satisfies the following equation:

$$\hat{K}(m^*) = d[(s_p - s_w)\{1 - S_w(m^*)\} + s_w] = \frac{n-\hat{c}}{\Phi(m^*)}. \tag{7.1}$$

Since both $\hat{K}(m)$ and $\Phi(m)$ are increasing functions, it is straightforward that a rise in the saving rate (either s_p or s_w) will increase m^*, and also the steady growth rate of capital, $\hat{K}(m^*)$. In this respect, our model is different from the Solow growth model in which the steady growth rate does not depend on the saving rate. This result comes from the fact that, in our model, firms with different size grow over time by taking advantage of potential economies of scale through learning by doing. This feature of our model may seem somewhat similar to the endogenous growth model of the Arrow type. However, our model differs from the existing endogenous growth models in that it takes into account of the size distribution of firms.

We can also examine how the steady growth is affected by the structural parameters, such as the Pareto coefficient, ρ, the scale effect, λ, the wage structure, ω, or mark-up factor, β. Let us first examine the effects of a change in the Pareto coefficient, ρ. As is shown by (5.11), an increase in ρ leads to an increase in S_w. It means that the wage share function $S_w(m)$ in equation (7.1) shifts upward. Then, the steady state value of the size ratio, m^*, must increase, since $\Phi(m)$ in equation (7.1) is an increase function. Therefore, the steady growth rate of capital, $\hat{K}(m^*)$, will decrease. Note that the Pareto coefficient is determined by $\rho = 1/a(1-\alpha)$, where α is the share of new firms' investment in the total increment of capital, and $a=1$, $a>1$, or $a<1$ depending on whether $\lambda+\omega=1$, $\lambda+\omega<1$ or $\lambda+\omega>1$. An increase in α or a decrease in a brings about an increase in ρ, which implies higher equality in the distribution of firms. Thus, more equal size distribution leads to the lower wage share and to the lower growth rate. But it should be noted here that changes in ρ take a long period of time, since Pareto distribution is the steady-state distribution. Therefore, changes in ρ have effects on various variables only after a long period of time.

The effects of a change in λ may similarly be examined. As is shown by (5.11), an increase in λ affects S_w to the same direction as an increase in ρ. Therefore, it leads to the lower wage share and to the lower growth rate. An increase in ω also has the same effects both on the wage share and the growth rate.

Conversely, an increase in the mark-up factor, β, will increase the steady growth rate of capital, since it shifts the wage share function downwards and leads to a decrease in m as the result.

Next, we examine the stability of this steady growth equilibrium. For this purpose, let us focus on equation (6.14). It is a one-variable differential equation, which determines the time path of m. Since $D(m)$, $\Phi(m)$ and $\hat{K}(m)$ are all increasing functions, $\hat{m}$ is an increasing function with respect to m in the neighborhood of the steady state equilibrium, $m = m^*$. Hence, the steady state is unstable. Fig. 3 provides a graphical representation of this instability property. Suppose that $m > m^*$ holds initially. Then, m and $\hat{m}$ will increase over time, and so will $\hat{K}(m)$. In this case, the equilibrium condition for the labor market (6.13) will be violated sooner or later, since we must have $\hat{k}_0 \geq 0$ when k_0 reached its minimum value. Conversely, suppose that $m < m^*$ holds initially. Then, m and $\hat{m}$ will decrease over time, and so will $\hat{K}(m)$. In this case, the equilibrium condition for the capital goods market (6.12) will be violated sooner or later, since we must have $\hat{m} + \hat{k}_0 = \hat{k}_T \geq 0$ unless the largest firms shrink their size. Thus, the steady growth equilibrium will not be maintained, unless $m = m^*$ is satisfied initially.

Ⅷ. Factor Substitution and the Stability of the Steady State Equilibrium

So far we have assumed that the production process of firms with each size of capital is characterized by fixed coefficients, so that a fixed amount of labor is used and a fixed amount of output is obtained. In this section, we take into account the substitutability between labor and capital, and show that it stabilizes the system.

When there is substitutability between labor and capital, the production function of a typical firm with size k may be expressed as

$$x = F\left(\frac{\delta(k)}{\gamma(k)} l,\ \delta(k)k\right), \tag{8.1}$$

where $\gamma(k)$ and $\delta(k)$ are the learning functions defined by (3.2) and (3.3). Assuming that this production function exhibits constant returns to scale and other usual properties, we can rewrite (8.1) as follows:

$$x = \delta(k)k\phi\left(\frac{1}{\gamma(k)}\frac{l}{k}\right), \quad \text{where} \quad \phi(0) = 0, \quad \phi' > 0, \quad \phi'' < 0 . \tag{8.2}$$

A typical firm with size k is assumed to make a choice of technique to minimize the total cost, given output capacity and technological knowledge. Thus, the problem of the typical firm is formulated as follows:

$$\min\ wl + rk, \qquad s.t. \quad \bar{x} = \bar{\delta}k\phi\left(\frac{1}{\bar{\gamma}}\frac{l}{k}\right) \tag{8.3}$$

The first order condition for this minimization problem is

$$\frac{w}{r}=\frac{\phi'(l/\bar{\gamma}k)}{\phi(l/\bar{\gamma}k)-(l/\bar{\gamma}k)\phi'(l/\bar{\gamma}k)} \tag{8.4}$$

Solving this equation in with respect to l/k, we have

$$\frac{l}{k}=\bar{\gamma}\psi\left(\frac{w}{r}\right), \quad \text{where} \quad \psi'<0. \tag{8.5}$$

Let us consider the case where the learning function $\gamma(k)$ is specified as (3.4). Substituting (3.4) into (8.5), we have

$$l=c\psi(w/r)k^{\lambda}. \tag{8.6}$$

This function replaces (3.6). We also specify the function $\delta(k)$ as (3.5), and assume μ to be unity and d to be constant. Under these assumptions, substitution of (8.6) into (8.2) gives

$$x=d\phi(\psi(w/r))k \tag{8.7}$$

This function replaces (3.7). Thus, (8.6) and (8.7) represent the learning process that takes into consideration the substitutability between labor and capital.

In our model, the wage rate is endogenously determined by (5.9) and (5.10), but the rate of interest, is given exogenously. So, we assume r to be constant and put it equal to unity for convenience. In addition, we specify $\psi(w)$ as a function with constant elasticity, that is, $\psi(w)=w^{-\eta}$, where η is assumed to be less than unity. Then, (8.6) is rewritten as

$$l=cw^{-\eta}k^{\lambda}. \tag{8.8}$$

It should be noted here that the wage rate w is a function of k, as is shown by (5.10).

Substituting this (8.8) into (5.6) and taking (5.10) into consideration, we have

$$L(k_0,k_T)=\frac{\rho Ac\{w(k_0)\}^{-\eta}}{\rho-\tilde{\lambda}}(1-m^{\tilde{\lambda}-\rho})k_0{}^{\tilde{\lambda}-\rho}. \tag{8.9}$$

It is assumed here that $\tilde{\lambda}=\lambda-\eta\omega>0$. Then, the growth rate of the total employment is given by

$$\hat{L}=\hat{A}+\hat{c}-\eta\hat{w}(k_0)+\frac{\rho-\tilde{\lambda}}{m^{\rho-\tilde{\lambda}}-1}\hat{m}-(\rho-\tilde{\lambda})\hat{k}_0. \tag{8.10}$$

Substituting (8.8) into (5.9), we have the following equation that shows determination of the wage rate for marginal firms.

$$w(k_0)=\frac{1}{\beta}\frac{d}{c\{w(k_0)\}^{-\eta}}k_0{}^{1-\tilde{\lambda}} \tag{8.11}$$

Taking logarithmic differentiation of this equation and solving it with respect to $\hat{w}(k_0)$, we have the following equation:

$$\hat{w}(k_0) = \frac{1}{1-\eta}[(1-\lambda)\hat{k}_0 - \hat{c}] \tag{8.12}$$

Substituting this equation into (8.10), we have the following equation for the growth rate of the total employment.

$$\hat{L} = \hat{A} + \kappa\hat{c} + \frac{\rho - \tilde{\lambda}}{m^{\rho-\tilde{\lambda}} - 1}\hat{m} - (\rho - \tilde{\lambda} + \varepsilon)\hat{k}_0, \tag{8.13}$$

where

$$\kappa \equiv \frac{1}{1-\eta} > 0, \qquad \varepsilon \equiv \frac{\eta(1-\tilde{\lambda})}{1-\eta} > 0 \tag{8.14}$$

Thus, when we take into consideration the factor substitution in our model, the equation for the growth rate of total employment (6.3) is replaced by (8.13). In this case the dynamic equation for firm-size ratio, (6.14), is replaced by

$$\hat{m} = \frac{1}{\tilde{D}(m)}[\tilde{\Phi}(m)\hat{K}(m) - (n - \kappa\hat{c})], \tag{8.15}$$

where

$$\tilde{D}(m) \equiv (\tilde{\lambda} - \varepsilon)\left(\frac{\rho - 1}{m^{\rho-1} - 1} - \frac{\rho}{m^{\rho} - 1}\right) - \left(\frac{\rho - \tilde{\lambda}}{m^{\rho-\tilde{\lambda}} - 1} - \frac{\rho}{m^{\rho} - 1}\right) \tag{8.16}$$

$$\tilde{\Phi}(m) \equiv \tilde{\lambda} + (1 - \tilde{\lambda})\frac{\rho - (1/a)}{\rho - 1}\frac{m^{\rho} - m}{m^{\rho} - 1} \tag{8.17}$$

It can be shown that if ε is sufficiently large, then $\tilde{D}(m) < 0$.[13] In this case, the dynamic equation for m has negative slope on $m - \hat{m}$ plane, as is shown in Figure 4. So, the steady-state equilibrium of this system is stable.

The comparative analysis of the steady state equilibrium that we have carried out in the last section becomes actually meaningful for the model in this section, since its stability has been proved.

FOOTNOTES

1. See also Simon and Bonini (1958).
2. Taking logarithm of equation (2.1) and regressing it to the size distribution of firms in Japanese manufacturing industry as shown by Figure 1, we obtain the following results:

$$\log N = \underset{(0.06)}{6.38} - \underset{(0.027)}{1.17} \log k \qquad (R^2 = 0.995)$$

where the numerical values below each coefficient represent its standard error.
3. See Simon (1955) for such examples of the Pareto distribution.
4. Lydall (1998) criticizes the neoclassical theory of firms from this point of view, and proposes an alternative theory. Though his ideas presented in his book are quite interesting, he does not present any concrete model.
5. The form of the function assumed here is the same as Arrow's. However, he assumes that the learning enters at the production of new capital goods in aggregate, while we assume that it enters in the process of capital accumulation of each firm.
6. Taking logarithm of these equations and regressing them to the data of Japanese manufacturing industry, we obtain the following results:

$$\log l = \underset{(0.04)}{-0.41} + \underset{(0.012)}{0.83} \log k \qquad (R^2 = 0.998)$$

$$\log x = \underset{(0.03)}{0.13} + \underset{(0.009)}{0.99} \log k \qquad (R^2 = 0.999),$$

where the numerical value below each coefficient represents its standard error.
7. Regression of this equation to the data of Japanese manufacturing industry gives the following result.

$$\log w = \underset{(0.02)}{0.37} + \underset{(0.005)}{0.08} \log k \qquad (R^2 = 0.968).$$

where the numerical value below each coefficient represent its standard error. This result shows that the positive relation between the wage rate and the size of firms is statistically significant.
8. See the second regression equation in footnote 5, which shows $\mu = 0.99$.
9. In the case of $\lambda + \omega > 1$, the size of firms will be bounded above as follows:

$$k \le \left(\frac{w(1)c}{d} \right)^{\frac{1}{1-\lambda-\omega}}$$

For, the rate of profit, $e(k)$, will become negative unless k satisfies this inequality.

10. More orthodox approach to the determination of saving in recent macroeconomics is to assume that the households maximize lifetime utility. But it is too complicated to deal with our model by introducing this assumption.
11. We omit the proof to save space.
12. We omit the proof to save space.
13. We omit the proof to save space

REFERENCES

Arrow, K. J. (1962), "The Economic Implications of Learning by Doing," *Review of Economic Studies,* Vol.24 (June), pp. 155-173.

Champernowne, D. G. (1953), "A Model of Income Distribution," *Economic Journal,* Vol. 63 (June), pp.318-351.

Gibrat, R. (1930), *Les Inegalites Economique,* Paris: Librairie du Sirey.

Lydall, H. (1998), *A Critique of the Orthodox Economics,* Lomdon: Macmillan.

Sato, K. (1970), "Size, Growth, and Skew Distribution," *Discussion Paper,* No. 145, SUNY at Buffalo.

Simon, H. A. (1955), "On a Class of Skew Distribution," *Biometrica,* Vol. 82, pp. 145-164.

Simon, H. A. and Bonini, C. P. (1958), "The Size Distribution of Business Firms," *American Economic Review,* Vol. 48 (September), pp.607-17.

Steindl, J. (1965), *Random Processes and the Growth of Firms,* London: Griffin.

Wold, H. O. A. and Whittle, P. (1957), "A Model Explaining the Pareto Distribution of Wealth," *Econometrica,* Vol. 25 (October), pp.591-595.

Figure 1. Perfectly Competitive Equilibrium of the Firm

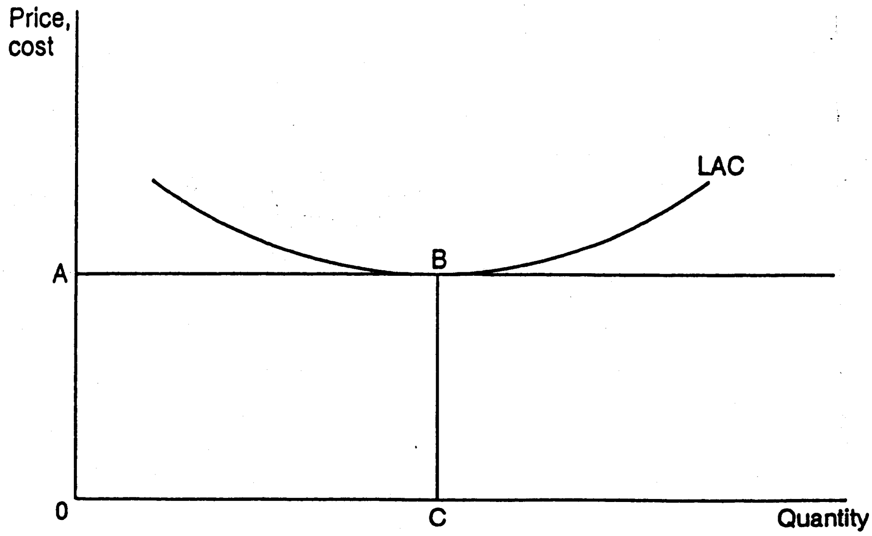
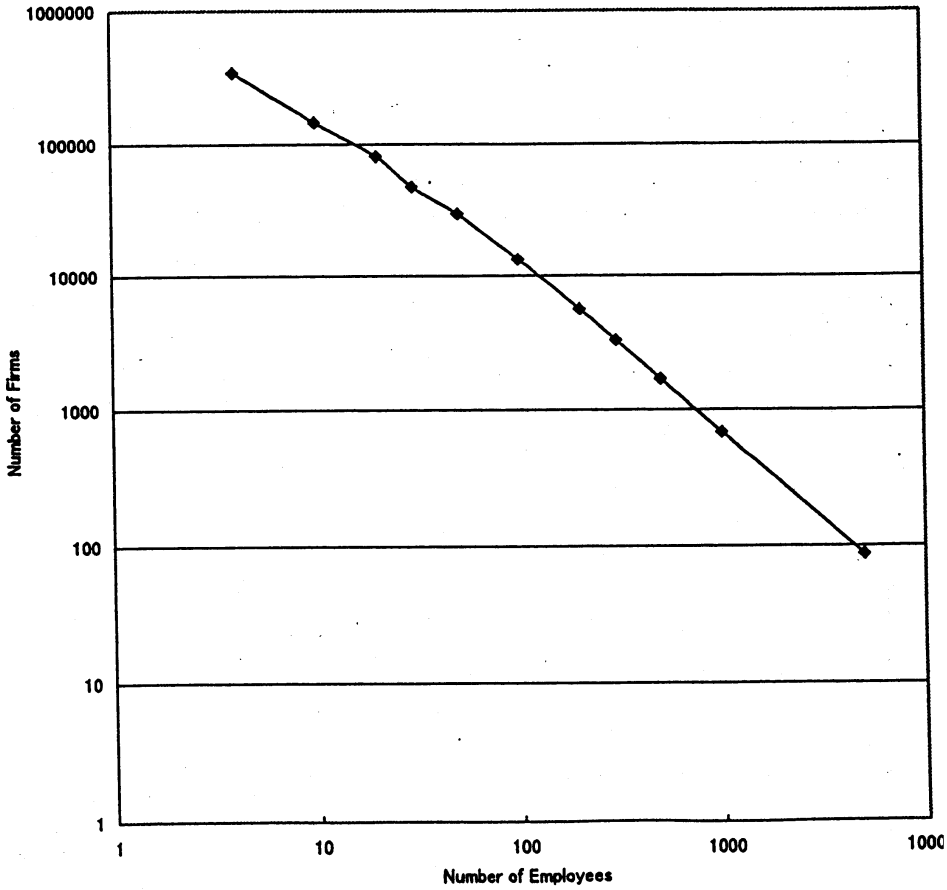


Figure 2. The Size Distribution of Firms in Japanese Manufacturing Industry



Source: Census of Manufactures, 1998.
(Ministry of International Trade and Industry)

Figure 3. Instability of the Steady Growth Equilibrium

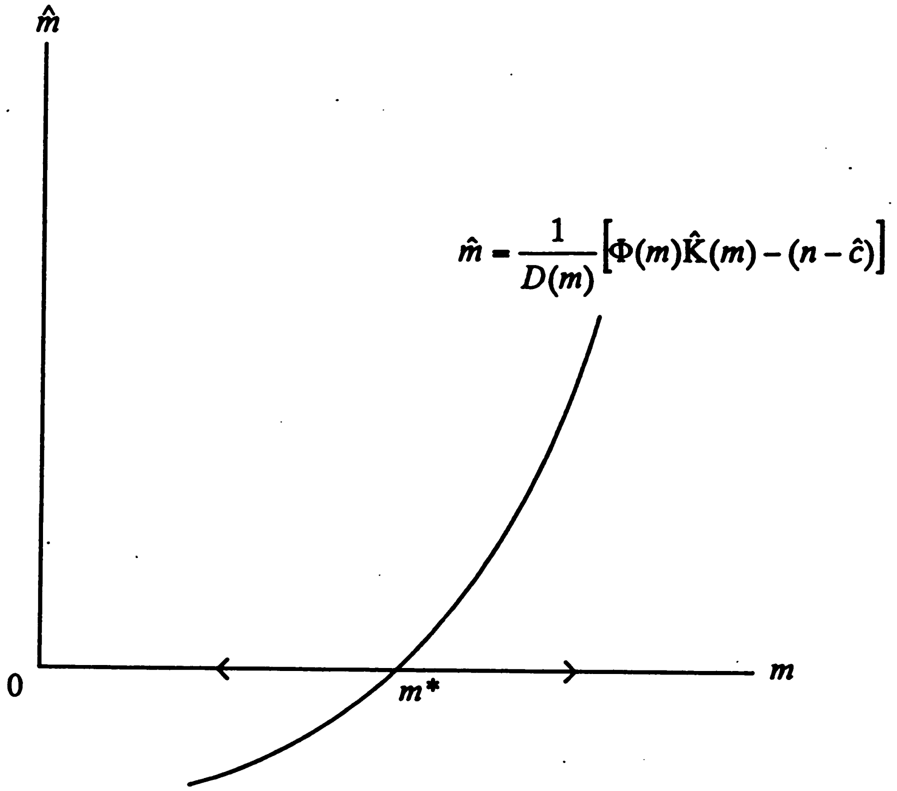


Figure 4. Stability of the Steady Growth Equilibrium