



RIETI Discussion Paper Series 08-E-035

# **Productivity Dispersion: Facts, Theory, and Implications**

**AOYAMA Hideaki**

Kyoto University

**YOSHIKAWA Hiroshi**

RIETI

**IYETOMI Hiroshi**

Niigata University

**FUJIWARA Yoshi**

NiCT/ATR CIS, Applied Network Science Lab.



Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry

<http://www.rieti.go.jp/en/>

# PRODUCTIVITY DISPERSION: FACTS, THEORY, AND IMPLICATIONS

Hideaki Aoyama<sup>1\*</sup>, Hiroshi Yoshikawa<sup>2</sup>, Hiroshi Iyetomi<sup>3</sup>,  
and Yoshi Fujiwara<sup>4</sup>

<sup>1</sup> *Department of Physics, Kyoto University, Kyoto 606-8501, Japan*

<sup>2</sup> *Faculty of Economics, University of Tokyo, Tokyo 113-0033, Japan*

<sup>3</sup> *Department of Physics, Niigata University, Niigata 950-2181, Japan*

<sup>4</sup> *NiCT/ATR CIS, Applied Network Science Lab., Kyoto 619-0288, Japan*

## Abstract

We study productivity dispersions across workers, firms and industrial sectors. Empirical study of data on Japanese firms shows that they all obey the Pareto law, and also that the Pareto index decreases with the level of aggregation. In order to explain these two stylized facts, we propose a theoretical framework built upon the basic principle of statistical physics. In this framework, we employ the concept of superstatistics, which accommodates fluctuations of aggregate demand. Our analysis demonstrates that the allocation of production factors depends crucially on the level of aggregate demand. The frontier of the production possibility set is a never-never land. The higher the level of aggregate demand is, the closer the economy is to the frontier of production possibility set.

JEL No: E10, E30, O40

---

\* Corresponding author: [hideaki.aoyama@scphys.kyoto-u.ac.jp](mailto:hideaki.aoyama@scphys.kyoto-u.ac.jp), tel: 075-753-6777, fax: 075-753-6804.

## I. Introduction

The standard economic analysis postulates that the marginal products of a production factor such as labor are all equal across firms, industries, and sectors in equilibrium. Otherwise, there remains a profit opportunity, and this contradicts the notion of equilibrium. Factor endowment, together with preferences and technology, determines equilibrium in such a way that the marginal products are equal across sectors and firms.

A bold challenge to this orthodox theory was given by Keynes (1936). He pointed out that the utilization of production factors is not full, and therefore, factor endowment is not an effective determinant of equilibrium. In *the General Theory*, Keynes identified the less than full utilization of production factor with the presence of involuntary unemployment of labor. Many controversies revolved around the theoretically ambiguous notion of “involuntary” unemployment. However, for Keynes’ economics to make sense, the existence of involuntary unemployment is not necessary. More generally, it is enough to assume that there is *underemployment* in the economy in the sense that the marginal products are not uniform at the highest level. In effect, what Keynes said is that in the demand-constrained equilibrium, productivity of production factor differs across industries and firms.

There are, in fact, several empirical findings which strongly suggest that productivity dispersion exists in the economy. The celebrated Okun’s Law is an example. The standard assumption on the neoclassical production function entails that the elasticity of output with respect to labor input (or the unemployment rate) is less than one. Okun (1962), however, found this elasticity to be three for the U.S. economy. This finding turns out to be so robust that it has eventually become known as Okun’s Law.

Okun attempts to explain his own finding by resorting to several factors. They are (a) additional jobs for people who do not actively seek work in a slack labor market but nonetheless take jobs when they become available; (b) a longer workweek reflecting less part-time and more overtime employment; and (c) extra productivity. On cyclical changes in productivity, he argues as follows:

“I now believe that an important part of the process involves a downgrading of labor in a slack economy—high-quality workers avoiding unemployment by accepting low-quality and less-productive jobs. The focus of this paper is on the upgrading of jobs associated with a high-pressure economy. Shifts in the composition of output and employment toward sectors and industries of higher productivity boost aggregate productivity as unemployment declines. Thus the movement to full employment draws on a reserve army of the underemployed as well as of the unemployed. In the main empirical study of this paper, I shall report new evidence concerning the upgrading of workers into more productive jobs in a high-pressure economy.” (Okun (1962, p.208))

Okun well recognized and indeed, by way of his celebrated law, demonstrated that there exists dispersion of labor productivity in the economy.

Another example is wage dispersion, an observation long made by labor economists. Mortensen (2003), for example, summarizes his analysis of wage dispersion as follows.

“Why are similar workers paid differently? Why do some jobs pay more than others? I have argued that wage dispersion of this kind reflects differences in employer productivity. . . . Of course, the assertion that wage dispersion is

the consequence of productivity dispersion begs another question. What is the explanation for productivity dispersion? ” (Mortensen (2003, p.129))

To this question, Mortensen’s explanation is as follows:

“Relative demand and productive efficiency of individual firms are continually shocked by events. The shocks are the consequence of changes in tastes, changes in regulations, and changes induced by globalization among others. Another important source of persistent productivity differences across firms is the process of adopting technical innovation. We know that the diffusion of new and more efficient methods is a slow, drawn-out affair. Experimentation is required to implement new methods. Many innovations are embodied in equipment and forms of human capital that are necessarily long-lived. Learning how and where to apply any new innovation takes time and may well be highly firm-specific. Since old technologies are not immediately replaced by the new for all of these reasons, productive efficiency varies considerably across firms at any point in time.” (Mortensen (2003, p.130))

It is important to recognize that the processes Mortensen describes are intrinsically stochastic. Because there are millions of firms in the economy, it is absolutely impossible for economists to pursue the precise behavior of an individual firm although no doubt each firm tries to maximize profits, perhaps dynamically, under certain constraints. Therefore, we must explore such stochastic dynamics in the economy as a whole by a different method. In this paper, we pursue such a stochastic approach.

The “stochastic approach” was indeed once popular in diverse areas of study such as income distribution and firm (city) size: primary examples are Champernowne (1973) and

Ijiri and Simon (1977). However, the stochastic approach eventually lost its momentum.

According to Sutton (1997), the reason is as follows:

“It seems to have been widely felt that these models might fit well, but were “only stochastic.” The aim was to move instead to a program of introducing stochastic elements into conventional maximizing models.” (Sutton (1997, p.45))

This trend is certainly in accordance with the motto of mainstream macroeconomics. However, Sutton, acknowledging the importance of the stochastic approach, argues that<sup>1</sup>

“a proper understanding of the evolution of structure may require an analysis not only of such economic mechanisms, but also of the role played by purely statistical (independence) effects, and that a complete theory will need to find an appropriate way of combining these two strands.” (Sutton (1997, p.57))

Actually, as the system under investigation becomes larger and more complicated, the importance of the stochastic approach increases. In fact, it is almost the definition of macro-system to which the basic approach based on statistical physics can be usefully applied. In natural sciences such as physics, chemistry, biology and ecology, the “stochastic approach” is routinely taken to analyze macro-system. Why not in macroeconomics?

---

<sup>1</sup>Incidentally, Sutton (1997), evidently having the central limit theorem in mind, identifies what he calls “purely statistical” effects with the effects of *independent* stochastic variables. However, the method of statistical physics is effective even when stochastic variables are correlated, that is they are *not* independent.

In what follows, we first explain the concept of *stochastic macro-equilibrium*. Next, we present an empirical analysis of productivity dispersion across Japanese firms. Then, we present a theory which can explain the observed distribution of productivity. The theoretical analysis which we offer follows the standard method of statistical physics, and abstracts from microeconomic analysis of the behavior of an individual agent. We believe that as Sutton (1997) suggests, our analysis complements the standard search theoretical approach, which pays careful attention to microeconomic behavior. Finally, we discuss implications of our analysis and a possible direction of future research.

## II. The Stochastic Macro-equilibrium

Tobin (1972) proposed the concept of *stochastic macro-equilibrium* in his attempt to explain the observed Phillips curve. He argues that

“[it is] stochastic, because random intersectoral shocks keep individual labor markets in diverse states of disequilibrium; macro-equilibrium, because the perpetual flux of particular markets produces fairly definite aggregate outcomes. ” (Tobin (1972, p.9))

His argument remained only verbal. However, as we will see shortly, the fundamental principle of statistical physics, in fact, provides the exact foundations for the concept of macro-equilibrium. A case in point is productivity dispersion in the economy.

We consider the economy in which there are  $K$  firms with their respective productivities  $c_1, c_2, \dots$ . Without loss of generality, we can assume

$$(1) \quad c_1 < c_2 < \dots < c_K.$$

Labor endowment is  $N$ . We distribute  $n_k$  workers to the  $k$ -th firm. Thus, the following equality holds:

$$(2) \quad \sum_{k=1}^K n_k = N.$$

The neoclassical theory takes it for granted that  $n_K$  is  $N$  while  $n_k$  ( $k \neq K$ )'s are all zero. Instead, in what follows, we seek the most probable distribution of productivity across workers under suitable constraints.<sup>2</sup> The possible number of a particular configuration  $(n_1, n_2, \dots, n_K), w_{\{n\}}$  is

$$(3) \quad w_{\{n\}} = \frac{N!}{\prod_{k=1}^K n_k!}.$$

Because the total number of possible configurations is  $K^N$ , the probability of occurrence of such a configuration,  $P_{\{n\}}$ , on the assumption of equal probabilities for all configurations, is given by

$$(4) \quad P_{\{n\}} = \frac{1}{K^N} \frac{N!}{\prod_{i=1}^K n_i!}.$$

Following the fundamental principle of statistical physics, we postulate that the configuration  $\{n_1, n_2, \dots, n_K\}$  that maximizes  $P_{\{n\}}$  under suitable constraints is realized in equilibrium. The idea is similar to the method of maximum likelihood in statistics or econometrics.

It is extremely important to recognize that this analysis is consistent with the standard assumption in economics that economic agents maximize their objective functions. As each particle satisfies the Newtonian law of motion, and maximizes the Hamiltonian in physics, in the economy, to be sure, all the micro agents optimize. However, constraints

---

<sup>2</sup>The productivity  $c_k$  corresponds to allowed energy level, and workers to distinguishable particles in statistical physics.

facing these micro agents, and even their objective functions keep changing due to idiosyncratic shocks. The situation is exactly the same as in physics where we never know the initial conditions for all the particles. In the economy, the number of micro agents is less than the counterpart in physics, which is typically  $10^{23}$ , but still there are  $10^6$  firms and  $10^7$  households! It is simply impossible and meaningless to analyze micro behavior of agents in great detail. For studying the macro-system, we must regard the behaviors of micro agents as stochastic even though their behaviors are purposeful. It is the fundamental principle of statistical physics that we observe the state of the macro-system that maximizes the probability, (4).

Toward our goal, we define the following quantity;

$$(5) \quad S = \frac{\ln P_{\{n\}}}{N} + \ln K \simeq \frac{1}{N} \left( N \ln N - \sum_{k=1}^K n_k \ln n_k \right) = - \sum_{k=1}^K p_k \ln p_k,$$

where  $p_k$  is defined as

$$(6) \quad p_k \equiv \frac{n_k}{N}.$$

Here, we assume that  $N$  and  $n_k$  are large, and apply the Stirling formula:

$$(7) \quad \ln m! \simeq m \ln m - m \quad \text{for } m \gg 1.$$

The quantity  $S$  corresponds to the Boltzmann-Gibbs entropy. Note that the maximization of  $S$  is equivalent to that of  $P_{\{n\}}$ .

We maximize  $S$  under two constraints. One is the normalization condition

$$(8) \quad \sum_{k=1}^K p_k = 1.$$

This is, of course, equivalent to the resource constraint,  $\sum_{k=1}^K n_k = N$ . The other con-

straint requires that the total output (GDP)  $Y$  is equal to the aggregate demand  $\tilde{D}$ ,

$$(9) \quad N \sum_{k=1}^K c_k p_k = Y = \tilde{D}.$$

For convenience, we define aggregate demand relative to factor endowment,  $D$  as follows:

$$(10) \quad D = \frac{\tilde{D}}{N}.$$

In what follows, we simply call  $D$  the aggregated demand. As we will shortly see, the aggregate demand,  $D$  determines the state of stochastic macro-equilibrium.

We maximize the following Lagrangian form:

$$(11) \quad S - \alpha \left( \sum_{k=1}^K p_k - 1 \right) - \beta \left( \sum_{k=1}^K c_k p_k - D \right).$$

Differentiating (11) with respect to  $p_i$ , we obtain

$$(12) \quad \ln p_k + (1 + \alpha) + \beta c_k = 0.$$

This yields

$$(13) \quad p_k = e^{-(1+\alpha)} e^{-\beta c_k}.$$

The normalization condition or the resource constraint (8) determines  $\alpha$  so that the distribution we seek is obtained as follows:

$$(14) \quad p_k = \frac{1}{Z(\beta)} e^{-\beta c_k}.$$

Thus, the distribution which maximizes the probability (4) under two constraints, (8) and (9), is exponential. We may call it *the equilibrium distribution*. *The exponent of the equilibrium distribution is the Lagrangian multiplier  $\beta$  corresponding to the aggregate demand constraint in (11).*

This exponential distribution is called *the Maxwell-Boltzmann distribution*. Here,  $Z(\beta)$  is what is called the partition function in physics, and makes sure that the  $p_k$ 's sum up to one.

$$(15) \quad Z(\beta) \equiv \sum_{k=1}^K e^{-\beta c_k}.$$

The exponent  $\beta$  is equivalent to the inverse of *temperature*,  $1/T$  in physics. Equations (9) and (14) yield the following:

$$(16) \quad D = \frac{1}{Z(\beta)} \sum_{k=1}^K c_k e^{-\beta c_k} = -\frac{d}{d\beta} \ln Z(\beta).$$

This equation relates the aggregate demand  $D$  to the exponent of the distribution  $\beta$  by way of the partition function  $Z(\beta)$ . Note that at this stage, the distribution of productivity  $c_k$  is arbitrary. Once it is given, the partition function  $Z(\beta)$  is defined by equation (15), and it, in turn, determines<sup>3</sup> the relation between  $D$  and  $\beta$ .

Suppose that the distribution of productivity across firms is *uniform*, that is,  $c_k = k\Delta_c$ , where  $\Delta_c$  is a constant. Then, from (14), we obtain

$$(17) \quad Z(\beta) = \sum_{k=1}^K e^{-\beta k\Delta_c} = e^{-\beta\Delta_c} \frac{1 - e^{-\beta(K-1)\Delta_c}}{1 - e^{-\beta\Delta_c}} \simeq \frac{1}{\beta\Delta_c},$$

under the assumption that  $\beta\Delta_c \ll 1$  and  $\beta K\Delta_c \gg 1$ . Therefore, in this case, we know from equations (15) and (16) that the exponent of the equilibrium exponential distribution,  $\beta$  is equal to the inverse of aggregate demand:

$$(18) \quad D = -\frac{d}{d\beta} \ln Z(\beta) = -\frac{d}{d\beta} \ln \left( \frac{1}{\beta\Delta_c} \right) = \frac{1}{\beta}.$$

---

<sup>3</sup>This is exactly the same as in the standard analysis in statistical physics where the relation between the average energy ( $D$ ) and the temperature ( $T = 1/\beta$ ) is determined once the energy levels ( $c_k$ ) are given for the system under investigation.

Because the inverse of  $\beta$  is temperature, in this case, aggregate demand is equivalent to temperature.

In summary, under the assumption that the productivity dispersion *across firms* is uniform, the equilibrium distribution of productivity *across workers* becomes exponential, namely the Maxwell-Boltzmann distribution with the exponent equal to the inverse of the aggregate demand  $D$  (Yoshikawa (2003); Aoki and Yoshikawa (2007)):

$$(19) \quad p_k = \frac{n_k}{N} = \frac{e^{-\frac{c_k}{D}}}{\sum_{k=1}^K e^{-\frac{c_k}{D}}}.$$

When the aggregate demand  $D$  is high, the distribution becomes flatter, meaning that production factors are mobilized to firms or sectors with high productivity, and *vice versa*. Figure I shows two distributions of productivity corresponding to high aggregate demand  $D_1$  and low  $D_2$ .

The present analysis provides a solid foundation for Okun’s argument that in general, production factors are under-utilized, and that workers upgrade into more-productive jobs in a “high-pressure” economy. So far, so good. However, the Boltzmann-Gibbs distribution is precisely exponential. Does it really fit our empirical observation? We next turn to this question.

### III. Stylized Facts

In this section, we empirically explore how productivity is actually distributed. Before we proceed, we must point out that what we observe is the average productivity  $c$ , defined by  $c = Y/L$ , where  $Y$  is the output, and  $L$  the labour input. What matters theoretically is, of course, the unobserved *marginal* productivity  $c_M$  defined by

$$(20) \quad c_M = \frac{\partial Y}{\partial L}.$$

We will shortly discuss the relationship between the distributions of these two different productivities. There, we will see that the difference between  $c$  and  $c_M$  does not affect our results on the empirical distributions.

Another problem is that we measure  $L$  in terms of the number of workers, as such data is readily available. One might argue that theoretically, it is more desirable to measure  $L$  in terms of work-hours, or for that matter, even in terms of work efficiency units. The effects of these possible “measurement errors” can also be handled in a similar way, which we will explain in subsection B. That is, we can safely ignore the measurement error problem as well for our present purpose.

#### *A. Empirical Distributions*

We use the Nikkei-NEEDS database, which is a major representative database of the listed firms in Japan. The details of this database are given in Appendix A, together with a brief description of the data-fitting.

We study productivity distributions at three aggregate levels, namely, across workers, firms, and industrial sectors. In order to explain empirical distributions, we find the continuous model more convenient than the discrete model. Accordingly, we define, for example, the probability density function of firms with productivity  $c$  as  $P^{(F)}(c)$ . The number of firms with productivity between  $c$  and  $c + dc$  is now  $KP^{(F)}(c)dc$ . It satisfies the following normalization condition:

$$(21) \quad \int_0^{\infty} P^{(F)}(c) dc = 1.$$

In section II, we explained that equilibrium productivity dispersion *across workers* becomes the exponential distribution under the assumption that the distribution of pro-

ductivity *across firms* is uniform. Here,  $P^{(F)}(c)$  is not restricted to be uniform. Rather, the distribution  $P^{(F)}(c)$  is to be determined empirically. We analogously define  $P^{(W)}(c)$  for workers, and  $P^{(S)}(c)$  for industrial sectors. We denote their respective cumulative probability distributions by  $P_{>}^{(S,F,W)}(c)$ :

$$(22) \quad P_{>}^{(*)}(c) = \int_c^\infty P^{(*)}(c') dc' \quad (* = S, F, W).$$

Because the cumulative probability is the probability that a firm's (or worker's or sector's) productivity is larger than  $c$ , it can be measured by rank-size plots, whose vertical axis is the [rank]/[the total number of firms] and the horizontal axis  $c$ . The rank-size plot has advantages in that it is free from binning problems, which haunt probability density function (pdf) plots, and also that it has less statistical noise. For these reasons, in what follows, we will show the cumulative probability rather than the probability density function.

The productivity distribution across firms, to be exact  $\log P_{>}^{(F)}(c)$  obtained from the Nikkei-NEEDS data, is plotted in Figure II for the year 2005. The dots are the data points; each dot corresponds to a firm whose position is determined from its rank and the productivity  $c$ . For reference, the power, exponential, and log-normal distributions are shown in the diagram. The exponential and log-normal distributions are represented by their respective curves while the power-law by a straight line whose slope is equal to the power exponent.

Evidently, the uniform distribution implicitly assumed in the analysis in section II does not fit the actual data at all. For small  $c$  (low-productivity), the log-normal law (dash-dotted curve) fits well, and for large  $c$  (high-productivity), the power-law (broken line) fits well, with smooth transition from the former to the latter at around  $\log c \simeq 2$ .

The result shown in Figure II is for the year 2005, but the basically same result holds good for other sample periods.

Cumulative probability  $P_{>}^{(F)}(c)$  for large  $c$  can be, therefore, represented by the following *power distribution*:

$$(23) \quad P_{>}^{(F)}(c) \simeq \left(\frac{c}{c_0}\right)^{-\mu_F} \quad (c \gg c_0).$$

where  $c_0$  is a parameter that defines the order of the productivity  $c$ .<sup>4</sup> The power exponent  $\mu_F$  is called *the Pareto index*, and  $c_0$  the Pareto scale. The probability density function is then given by the following:

$$(24) \quad P^{(F)}(c) = -\frac{d}{dc}P_{>}^{(F)}(c) \simeq \mu_F \frac{c^{-\mu_F-1}}{c_0^{-\mu_F}} \quad (c \gg c_0).$$

Next, we explore the productivity dispersion *across industrial sectors*. The Nikkei-NEEDS database defines 33 sectors. The productivity distribution across these sectors for the year 2005 is plotted in Fig. III. We observe that once again, the power-law (a straight line in the diagram) fits the data pretty well. The same result holds true for all the sample years.

Finally, the productivity distribution *across workers* is plotted in Figure IV for the year 2005. Again, we observe that the power-law (broken line) fits the data very well for large  $c$ . A casual observation of Figure IV may make one wonder whether the share of workers for which the power-law fits well may be small. This impression is wrong. In fact, the fitted range is approximately, say,  $\log P_{>}^{(W)}(c) \in [-0.4, -3.1]$ , which translates to the rank of the workers

$$[10^{-0.4}, 10^{-3.1}] \times [\text{Total number of workers}] \simeq [1.52 \times 10^6, 3.04 \times 10^3]$$

---

<sup>4</sup>We note that  $c_0$  has the same dimension as  $c$ , so that  $P_{>}^{(F)}(c)$  is dimensionless, as it should be.

This means that some 1.52 million workers, that is, 39% of all workers fit the power-law.

The values of the Pareto indices obtained are shown in Figure V for three levels of aggregation. We note that the Pareto index decreases as the aggregation level goes up from workers to firms, and from firms to the industrial sectors.

### B. *Marginal vs. Average Productivity*

Before we conclude this section, we digress into the relationship between the marginal and average productivities of labor. Theoretically, the marginal productivity matters while what we observe and, therefore, used in the above analysis is the average productivity. To explore the relation between the two productivities,  $c$  and  $c_M$ , we assume the Cobb-Douglas production function:

$$(25) \quad Y = AK^{1-\alpha}L^\alpha \quad (0 < \alpha < 1).$$

Equation (25) leads us to the following relation:

$$(26) \quad c_M = \alpha c.$$

In general, the value of  $\alpha$  differs across firms. Therefore, the distribution of  $c_M$  is, in general, different from that of  $c$ .

However, thanks to (26), we can relate the pdf of  $c_M$ ,  $P_M(c_M)$  to the joint pdf of  $c$  and  $\alpha$ ,  $P_{c,\alpha}(c, \alpha)$  as follows:

$$(27) \quad \begin{aligned} P_M(c_M) &= \int_0^1 d\alpha \int_0^\infty dc \delta(c_M - \alpha c) P_{c,\alpha}(c, \alpha) \\ &= \int_0^1 \frac{d\alpha}{\alpha} P_{c,\alpha}\left(\frac{c_M}{\alpha}, \alpha\right). \end{aligned}$$

In general,  $P_{c,\alpha}(c, \alpha)$  can be written as follows,

$$(28) \quad P_{c,\alpha}(c, \alpha) = P^{(F)}(c)P(\alpha | c).$$

Here,  $P(\alpha | c)$  is the conditional pdf; it is the pdf of  $\alpha$  for the fixed value of productivity  $c$ . It is normalized as follows:

$$(29) \quad \int_0^1 P(\alpha | c) d\alpha = 1,$$

for any value of  $c$ . We have already seen that  $P^{(F)}(c)$  obeys the power-law for large  $c$ :

$$(30) \quad P^{(F)}(c) = \mu_F \frac{c^{-\mu_F-1}}{c_0^{-\mu_F}}.$$

From (30), (28), and (27), we obtain the following:

$$(31) \quad P_M(c_M) = \mu_F \frac{c_M^{-\mu_F-1}}{c_0^{-\mu_F}} \int_0^1 d\alpha \alpha^{-\mu_M} P\left(\alpha \mid \frac{c_M}{\alpha}\right).$$

Here, we assumed that  $P(\alpha | c)$  does not extend too near  $\alpha = 0$  so that  $c_M/\alpha$  stays in the asymptotic region. From (31), we can conclude that if  $\alpha$  and  $c$  are independent, namely

$$(32) \quad P(\alpha | c) = P_\alpha(\alpha),$$

then, the distribution of marginal productivity  $P_M(c_M)$  also obeys the power law with the *identical* Pareto index  $\mu_F$  as for the average productivity:

$$(33) \quad P_M(c_M) \propto c_M^{-\mu_F-1}.$$

In conclusion, *to the extent that  $\alpha$  and  $c$  are independent, the distribution of the unobserved marginal productivity obeys the power-law with the same Pareto index as for the observed average productivity.* A sample Monte-Carlo simulation result is shown in Figure VI. Because this is the log-log plot, the gradient of the straight region is equal to the Pareto index. The equality of the Pareto indices is clearly seen.

In our analysis,  $L$  is the number of workers. Theoretically, it would be desirable to measure  $L$  in terms of work-hours or for that matter, even in terms of work efficiency.

We can apply the above analysis to the case where the problem is measurement error by simply interpreting  $\alpha$  in (26) as such measurement error rather than a parameter of the Cobb-Douglas production function. Thus, to the extent that the average productivity  $c$  and measurement error  $\alpha$  are independent, the distribution of “true” productivity obeys the power-law with the same exponent as for the measured productivity. We conclude that the power-laws for productivity dispersion across workers, firms, and industry obtained above are quite robust.

### C. *Summary of Empirical Observations*

We summarize the empirical observations as two stylized facts:

- I. The distribution of productivity obeys the Pareto distribution (*i.e.* the power-law for the high-productivity group) at every level of aggregation, that is, across workers, firms, and industrial sectors.
- II. The Pareto index, namely the power exponent decreases as the level of aggregation goes up:  $\mu_W > \mu_F > \mu_S$  (Figure V).

As we explained in section II, under the assumption that the distribution of productivity *across firms* is uniform leads to the exponential distribution of productivity *across workers*. Obviously, this model does not fit the empirical observations. In the next section, extending the basic framework, we present a theoretical model that explains the above stylized facts.

## IV. Theory

In this section, we develop a new theoretical model that explains two stylized facts in three steps: First, we discuss the generic framework that explains the power-law distribution for firm's productivity. Secondly, we extend the basic model explained in section II by incorporating the stylized fact that productivity distribution across firms is not uniform, but rather obeys the power-law. Because the extended model still fails to explain the stylized facts, we take the third step; we propose the *superstatistics* framework. It can explain two stylized facts presented in section III.

### A. A Firm's Productivity Dispersion: A Model of Jump Markov Process

The standard economic analysis takes it for granted that all the production factors enjoy the highest marginal productivity in equilibrium. However, this is a wrong characterization of the economy. The fact is that production factors cannot be reallocated instantaneously in such a way that their marginal products are equal in all economic activities. Rather, at each moment in time, there exists a dispersion or distribution of productivity as shown in the preceding section. Evidently, an important reason why the marginal products of workers are not equal is, as Mortensen (2003) suggests, that there are differences in productivity across firms. To describe the dynamics of firm's productivity, we employ a continuous-time jump Markov process. Using the Markov model, we can show that the power-law distribution is a generic consequence under a reasonable assumption.

Suppose that a firm has a productivity denoted by  $c$ . In a small time interval  $dt$ , the firm's productivity  $c$  increases by a small amount, which we can assume is unity without loss of generality, with probability  $w_+(c) dt$ . We denote it as  $w_+(c)$  because this

probability  $w_+$  depends on the level of  $c$ . Likewise, it decreases by unity with probability  $w_-(c) dt$ . Thus,  $w_+(c)$  and  $w_-(c)$  are transition rates for the processes  $c \rightarrow c + 1$  and  $c \rightarrow c - 1$ , respectively.

We also assume that a new firm is born with a unit of productivity with probability  $p dt$ . On the other hand, a firm with  $c = 1$  will cease to exist if its productivity falls to zero; Thus the probability of exit is  $w_-(c = 1) dt$ . A set of the transition rates and the entry probability specifies the jump Markov process.

Given this Markov model, the evolution of the average number of firms having productivity  $c$  at time  $t$ ,  $n(c, t)$ , obeys the following master equation:

$$(34) \quad \frac{\partial n(c, t)}{\partial t} = w_+(c-1) n(c-1, t) + w_-(c+1) n(c+1, t) - w_+(c) n(c, t) - w_-(c) n(c, t) + p \delta_{c,1} ,$$

Here,  $\delta_{c,1}$  is 1 if  $c = 1$  and 0 otherwise. This equation shows that the change in  $n(c, t)$  over time is nothing but the *net* inflows to the state  $c$ .

The total number of firms is given by

$$(35) \quad K_t \equiv \sum_{c=1}^{\infty} n(c, t).$$

We define the aggregate productivity index  $C$  as follows:

$$(36) \quad C_t \equiv \sum_{c=1}^{\infty} c n(c, t).$$

It follows from (34) that

$$(37) \quad \frac{d}{dt} K_t = p - w_-(1) n(1, t),$$

$$(38) \quad \frac{d}{dt} C_t = p - \sum_{c=1}^{\infty} (w_-(c) - w_+(c)) n(c, t).$$

We consider the steady state. It is the stationary solution of (34) such that  $\partial n(c, t)/\partial t = 0$ . The solution  $n(c)$  can be readily obtained by the standard method. Setting (37) equal to zero, we obtain a boundary condition that  $w_-(1)n(1) = p$ . Using this boundary condition, we can easily show

$$(39) \quad n(c) = n(1) \prod_{k=1}^{c-1} \frac{w_+(c-k)}{w_-(c-k+1)} .$$

Next, we make an important assumption on the transition rates,  $w_+$  and  $w_-$ . Namely, we assume that the probabilities of an increase and a decrease of productivity depend on the firms current level of productivity. Specifically, the higher the current level of productivity, the larger a chance of a unit productivity change. This assumption means that the transition rates can be written as  $w_+(c) = a_+ c^\alpha$  and  $w_-(c) = a_- c^\alpha$ , respectively. Here,  $a_+$  and  $a_-$  are positive constants, and  $\alpha$  is greater than 1. Under this assumption, the stationary solution (39) becomes

$$(40) \quad n(c) = \frac{n(1)}{1 - n(1)/C_{(\alpha)}} \frac{(1 - n(1)/C_{(\alpha)})^c}{c^\alpha} \simeq c^{-\alpha} e^{-c/c^*} .$$

where

$$c^* \equiv (n(1)/C_{(\alpha)})^{-1} \quad \text{and} \quad C_{(\alpha)} \equiv \sum_{c=1}^{\infty} c^\alpha n(c) .$$

We have used the relation  $a_+/a_- = 1 - n(1)/C_{(\alpha)}$ , which follows from (37) and (38). The approximation in (40) follows from  $n(1)/C_{(\alpha)} \ll 1$ . The exponential cut-off works as  $c$  approaches to  $c^*$ . However, the value of  $c^*$  is practically quite large. Therefore, we observe the power-law distribution  $n(c) \propto c^{-\alpha}$  for a wide range of  $c$  in spite of the cut-off. We note that the power exponent  $\mu$  for the empirical distribution presented in section III is related to  $\alpha$  simply by

$$\mu = \alpha - 1 .$$

The present model can be understood easily with the help of an analogy to the formation of cities. Imagine that  $n(c, t)$  is the number of cities with population  $c$  at time  $t$ .  $w_+(c)$  corresponds to a birth in a city with population  $c$ , or an inflow into the city from another city. Similarly,  $w_-(c)$  represents a death or an exit of a person moving to another city. The rates are the instantaneous probabilities that the population of a city with the current population  $c$  either increases or decreases by one. They are, therefore, the entry and exit rates of *one person* times population  $c$ , respectively. And a drifter forms his own one-person city with the instantaneous probability  $p$ . In this model, the dynamics of  $n(c, t)$ , namely the average number of cities with population  $c$  is given by equation (34). In the case of population dynamics, one might assume that the entry (or birth) and exit (or death) rates of a person,  $a_+$  and  $a_-$  are independent of the size of population of the city in which the person lives. Then,  $w_+(c)$  and  $w_-(c)$  become linear functions of  $c$ , namely,  $a_+c$  and  $a_-c$ . Even in population dynamics, though, one might assume that the entry rate of a person into a big city is higher than its counterpart in a small city because of the better job opportunity or the attractiveness of “city life.” The same may hold for the exit and death rates because of congestion or epidemics.

It turned out that in the dynamics of firm productivity, both the “entry” and “exit” rates of an existing “productivity unit” are increasing functions of  $c$ , namely the level of productivity a part of which that particular unit happens to be; to be concrete, they are  $a_+c$  and  $a_-c$ . Thus,  $w_+(c)$ , for example, becomes  $a_+c$  times  $c$  which is equal to  $a_+c^2$ . Likewise, we obtain  $w_-(x) = a_-x^2$ . This is the case of  $\alpha = 2$ , the so-called Zipf law (see Sutton (1997)).<sup>5</sup>

---

<sup>5</sup>See Ijiri and Simon (1975); Marsili and Zhang (1998) for the formation of cities whose sizes obey the Zipf law. Our present model of productivity dynamics has, in fact, a very close analogy to the dynamics

There is also a technical reason why we may expect  $\alpha$  to be larger than two. From (40), we can write the total number of firms and the aggregate productivity index as

$$(41) \quad K = \frac{n(1)}{\Gamma(\alpha)} \int_0^\infty \frac{t^{\alpha-1}}{e^t - 1 + n(1)/C_{(\alpha)}} dt ,$$

$$(42) \quad C = \frac{n(1)}{\Gamma(\alpha - 1)} \int_0^\infty \frac{t^{\alpha-2}}{e^t - 1 + n(1)/C_{(\alpha)}} dt ,$$

where  $\Gamma(z)$  is the gamma function defined by  $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ . In the limit as  $n(1)/C_{(\alpha)}$  goes to zero, the integral in (41) is finite for  $\alpha > 1$ , while the integral in (42) tends to be arbitrarily large for  $1 < \alpha < 2$ . Therefore, for the finiteness of both (41) and (42), it is reasonable to assume that  $\alpha \geq 2$ .

In summary, under the reasonable assumption that the probability of a unit change in productivity is an increasing function of its current level  $c$ , we obtain the power-law distribution as we actually observe. Now, economists are prone to take changes in productivity as “technical progress.” That is why the focus of attention is so often on R & D investment. However, if productivity growth is always technical progress, its decrease must be “technical regression,” the very existence of which one might question. At the firm level, an important source of productivity change is actually a sectoral shift of demand; when demand for product A increases, for example, productivity at the firm producing A increases, and *vice versa*. Fay and Medoff (1985) indeed document such changes in firm’s labor productivity by way of changes in the rate of labor hoarding. Stochastic productivity changes, which our Markov model describes, certainly include technical progress, particularly in the case of an increase, but at the same time represent *allocative demand disturbances*, the importance of which Davis et al. (1996) have so

---

of city size. It is indeed mathematically equivalent to the model of Marsili and Zhang (1998) which analyzed dynamics of city size, except for an additional assumption that  $n(1)/C_{(\alpha)} = p$ .

persuasively demonstrated in their book entitled *Job Creation and Destruction*. The empirical observation on productivity dispersion and our present analysis suggest that the probability of a unit allocative demand disturbance depends on the current size of firm or sector.

### B. *The Stochastic Macro-equilibrium Once Again*

The basic framework explained in section II presumes that productivity dispersion across firms, namely  $P_{>}^{(F)}(c)$  is uniform. However,  $P_{>}^{(F)}(c)$  actually obeys the power-law. Having clarified the generic origin of power-law distribution for  $c$ , we now turn our attention to the productivity dispersion across workers. We must extend the theory of the stochastic macro-equilibrium explained in section II under the assumption that  $P_{>}^{(F)}(c)$  is the power-distribution.

In order to develop the new theoretical framework, it is better to adapt the continuous notation we introduced previously. An example is

$$(43) \quad K \sum_{k=1}^{\infty} n(c_k) \longrightarrow K \int_0^{\infty} P^{(F)}(c) dc.$$

In the continuous model, equations (5), (8) and (9) read

$$(44) \quad S = -K \int_0^{\infty} (p(c) \ln p(c)) P^{(F)}(c) dc,$$

$$(45) \quad K \int_0^{\infty} p(c) P^{(F)}(c) dc = 1,$$

$$(46) \quad K \int_0^{\infty} c p(c) P^{(F)}(c) dc = D,$$

respectively. Here, we have replaced  $p_k = n_k/N$  by continuous function  $p(c)$ . Note that because the distribution of productivity *across firms* is no longer uniform, but is  $P^{(F)}(c)$ ,

the corresponding distribution *across workers*,  $P^{(W)}(c)$ , is

$$(47) \quad P^{(W)}(c) \equiv K p(c) P^{(F)}(c).$$

Using  $P^{(W)}(c)$ , we can rewrite equations (44), (45) and (46) as follows:

$$(48) \quad S = - \int_0^\infty P^{(W)}(c) \ln \frac{P^{(W)}(c)}{P^{(F)}(c)} dc + [\text{const.}],$$

$$(49) \quad \int_0^\infty P^{(W)}(c) dc = 1,$$

$$(50) \quad \int_0^\infty c P^{(W)}(c) dc = D.$$

Here, [const.] in equation (48) is an irrelevant constant term.

We maximize  $S$  (equation (48)) under two constraints, (49) and (50), by means of calculus of variation with respect to  $P^{(W)}(c)$  to obtain

$$(51) \quad P^{(W)}(c) = \frac{1}{Z(\beta)} P^{(F)}(c) e^{-\beta c}.$$

Here, as in section II,  $\beta$  is the Lagrangian multiplier for the aggregate demand constraint, (46), and the partition function  $Z(\beta)$  is given by

$$(52) \quad Z(\beta) = \int_0^\infty P^{(F)}(c) e^{-\beta c} dc.$$

It is easy to see that constraint (49) is satisfied. Constraint (50) now reads

$$(53) \quad D = \frac{1}{Z(\beta)} \int_0^\infty c P^{(F)}(c) e^{-\beta c} dc.$$

This equation is equivalent to:

$$(54) \quad D = - \frac{d}{d\beta} \ln Z(\beta).$$

It is straightforward to see that equations (51) and (52) are the counterparts of equations (14) and (15), respectively, under the assumption that the productivity dispersion across firms is not uniform, but is  $P_{>}^{(F)}(c)$ .

Because the productivity distribution across firms,  $P^{(F)}(c)$ , obeys the Pareto law, the relation between the aggregate demand  $D$  and  $\beta$  (or the temperature) is not so simple as shown in equation (18), but is, in general, quite complicated. However, we can prove that the power exponent  $\beta$  is a decreasing function of the aggregate demand  $D$ . Thus, the fundamental proposition that when the aggregate demand is high, production factors are mobilized to firms and sectors with high productivity (Figure I), holds true in the extended model as well. We provide the proof in Appendix B.

Now, we explore productivity dispersion in this extended model. The productivity dispersion across workers  $P^{(W)}(c)$ , relates to that across firms  $P^{(F)}(c)$ , by way of equation (51). The latter obeys the power law. Figure VII shows examples of the productivity distributions across workers for several different values of  $D$ , given  $P^{(F)}(c)$ . The solid curve is for firms with  $\mu_F = 1.5$ . The dashed lines are the corresponding distributions for workers with different values of  $D$ . All of them have strong suppression for large  $c$  due to the exponential factor  $e^{-\beta c}$ . They are in stark contrast to the power-law. Note that the power-law corresponds to a straight line in the figure, which shows the relation between  $\ln P > (c)$  and  $\ln c$ . The only way to reconcile the distribution (51) with the observed power-law is to assume an extremely small value of  $\beta$  so that the Boltzmann factor  $e^{-\beta c}$  becomes close to one, and does not suppress  $P^{(W)}(c)$ . This trick, however, does not work because, in this case, it yields the Pareto index  $\mu_W$  of the worker's productivity distribution *equal to* that of the firm's distribution,  $\mu_F$ . It is inconsistent with the empirical observation that  $\mu_W > \mu_F$  (Figure V). Comparing Figures IV and VII, we conclude that the extended model still fails to explain the observations. We must seek a new theoretical framework.

### C. *A Worker's Productivity Dispersion under Fluctuating Aggregate Demand*

The theoretical framework explained so far implicitly assumes that the aggregate demand  $D$  is constant. Plainly, this is an oversimplification;  $D$  fluctuates.

A Macro-system under fluctuations of an external environment can be analyzed with the help of *superstatistics* or “statistics of statistics” in statistical physics (Beck and Cohen, 2003).<sup>6</sup> In this theory, the system goes through changing external influences, but is in equilibrium at the limited scale in time and/or space, in which the temperature may be regarded as constant and the Boltzmann distribution is achieved. In other words, the system is only locally in equilibrium; globally seen, it is out of equilibrium. In order to analyze such a system, superstatistics introduces averaging over the Boltzmann factors. Depending on the weight function used for averaging, it can yield various distributions, including the power-law (Touchette and Beck, 2005).

GDP of a particular year is certainly a scalar constant when the year is over. However, it actually fluctuates daily if those fluctuations cannot be practically measured. Accordingly, the macro environment surrounding firms keeps changing almost continuously. Differences by region, industry, and sector can also be taken care of by the superstatistics approach. When the aggregate demand  $D$  changes, the new stochastic process conditioned by new  $D$  allows production factors to move to a different equilibrium. Averaged over various possible equilibria, each of which depends on a particular value of  $D$ , the resulting distribution becomes different from any of each equilibrium distribution.

Specifically, in superstatistics, the familiar Boltzmann factor,  $\exp(-\beta c)$  is replaced by

---

<sup>6</sup>There are several model cases where superstatistics was applied successfully. Among them, the Brownian motion of a particle going thorough a changing environment provides a good analogy to our case (Beck, 2006).

the following weighted average:

$$(55) \quad B(c) = \int_0^\infty f(\beta) e^{-\beta c} d\beta,$$

Here, the weight factor  $f(\beta)$  represents the changing macroeconomic environment. Note that because  $\beta$  is a monotonically decreasing function of  $D$ , the weight factor  $f(\beta)$  corresponds to changes in the aggregate demand. With this weight factor, the probability distribution of worker's productivity, (51), is now replaced by the following:

$$(56) \quad P^{(W)}(c) = \frac{1}{Z_B} P^{(F)}(c) B(c).$$

Here, the partition function  $Z_B$  is also redefined as

$$(57) \quad Z_B = \int_0^\infty P^{(F)}(c) B(c) dc.$$

We now examine whether  $P^{(W)}(c)$  in equation (56) obeys the power-law for high productivity  $c$ . The integration in equation (55) is dominated by the small  $\beta$  (high demand) region for large  $c$ . We assume the following behavior of the pdf  $f(\beta)$  for  $\beta \rightarrow 0$ ,

$$(58) \quad f(\beta) \propto \beta^{-\gamma} \quad (\gamma < 1),$$

where the constraint for the parameter  $\gamma$  comes from the convergence of the integration in equation (55). The proportional constant is irrelevant because  $P^{(W)}(c)$  is normalized by  $Z_B$ . This leads to the following  $B(c)$  for large  $c$ :

$$(59) \quad B(c) \propto \Gamma(1 - \gamma) c^{\gamma-1}.$$

Substituting (59) into Equation (56), we find that the productivity distribution across workers obeys the power-law;

$$(60) \quad P^{(W)}(c) \propto c^{-\mu_w-1},$$

with

$$(61) \quad \mu_W = \mu_F - \gamma + 1.$$

Because of the constraint  $\gamma < 1$ , this leads to the inequality

$$(62) \quad \mu_W > \mu_F.$$

This agrees with our empirical observation (Figure V or stylized Fact II in section III.C).

Because  $\beta$  is related to  $D$  by way of equation (54), the pdf  $f_\beta(\beta)$  of  $\beta$  is related to pdf  $f_D(D)$  of the (fluctuating)  $D$  as follows:

$$(63) \quad f_\beta(\beta)d\beta = f_D(D)dD.$$

As is noted several times, small  $\beta$  corresponds to high aggregate demand  $D$ . In particular, the following relation holds<sup>7</sup> for  $\beta \rightarrow 0$ :

$$(64) \quad \langle c \rangle_0 - D \propto \begin{cases} \beta & \text{for } 2 < \mu_F; \\ \beta^{\mu_F - 1} & \text{for } 1 < \mu_F < 2. \end{cases}$$

This leads to<sup>8</sup>

$$(65) \quad f_D(D) \propto (\langle c \rangle_0 - D)^{-\delta},$$

with

$$(66) \quad \gamma - 1 = \begin{cases} \delta - 1 & \text{for } 2 < \mu_F; \\ (\mu_F - 1)(\delta - 1) & \text{for } 1 < \mu_F < 2. \end{cases}$$

Here, the parameter  $\delta$  is constrained by

$$(67) \quad \delta < 1$$

---

<sup>7</sup>The proof is to be given on request.

<sup>8</sup>At  $\mu_F = 2$ , we need additional logarithmic factors for  $f(\beta)$ , but the power of  $\beta$  is essentially the boundary case between the above two,  $\gamma = \delta$ .

from the normalizability of the distribution of  $f_D(D)$ , which is consistent with the constraint  $\gamma < 1$  and equation (66).

Equation (65) means that changes in the aggregate demand  $D$  follows the power-law. Gabaix (2005) indeed demonstrates that idiosyncratic shocks to the top 100 firms explain a large fraction (one third) of aggregate volatility for the U.S. economy. This is a characteristic of power-law. In any case,  $D$  is not constant, but fluctuates now. Accordingly, the problem is not a relation between the productivity dispersion across workers and the level of  $D$ , but rather how the distribution depends on the way in which  $D$  fluctuates.

Figure VIII shows an example of the relation between the distribution of  $D$ ,  $f_D(D)$  near the upperbound  $\langle c \rangle_0$  and the cumulative productivity distribution of workers  $P_{>}^{(W)}(c)$ . In the figure, the solid curves correspond to the large value of  $\delta$  whereas the broken curves to the small value. Figure VIII demonstrates that as the distribution of  $D$  becomes skewed toward large  $D$ , the tail of the productivity distribution becomes heavier. Roughly speaking, when  $D$  is high, the productivity dispersion becomes skewed toward the higher level, and *vice versa*.

Combining equations (61) and (66), we reach the following relation between the Pareto indices:

$$(68) \quad \mu_W = \begin{cases} \mu_F - \delta + 1 & \text{for } 2 < \mu_F; \\ (\mu_F - 1)(-\delta + 1) + \mu_F & \text{for } 1 < \mu_F < 2. \end{cases}$$

This relation between  $\mu_W$  and  $\mu_F$  is illustrated in Figure IX. As noted previously, because of the constraint  $\delta < 1$ , equation (68) necessarily makes  $\mu_W$  larger than  $\mu_F$ . This is in good agreement with our empirical finding. Incidentally, equation (68) has a fixed point at  $(\mu_W, \mu_F) = (1, 1)$ ; the line defined by equation (68) always passes through this point

irrespective of the value of  $\delta$ . The Pareto index for firms is smaller than that for workers, but it cannot be less than one, because of the existence of the fixed point (1,1).

The superstatistics framework presented above may apply for any adjoining levels of aggregation; instead of applying it for workers and firms, we may apply it for firms and industrial sectors. Then, we can draw the conclusion that as we go up from firms to industrial sectors, the Pareto index again goes down, albeit for a different value of  $\delta$ . This is illustrated in Fig. X. Because of the existence of the point (1, 1), as the aggregation level goes up, the Pareto index is driven toward 1, but not beyond 1. *At the highest aggregation level, it is expected to be close to one.* This is again in good agreement with our empirical finding that the Pareto index of the industrial sector  $\mu_S$  is close to one (see Figure V).

In summary, the superstatistics framework successfully explains two empirical findings we have summarized in section III.C. Furthermore, given the measured values of  $\mu_W$  and  $\mu_F$ , the relation (68) can be used to determine the value of  $\delta$ :

$$(69) \quad \delta = \begin{cases} \mu_F - \mu_W + 1 & \text{for } 2 < \mu_F; \\ \frac{\mu_F - \mu_W}{\mu_F - 1} + 1 & \text{for } 1 < \mu_F < 2. \end{cases}$$

The result is shown in Figure XI. Recall that  $\delta$  is the power exponent of the distribution of aggregate demand  $D$ . Therefore, low  $\delta$  means the relatively low level of the aggregate demand. In Figure XI we observe that the aggregate demand was high during the late 1980s, while beginning in the early 90s, it declined to the bottom in 2000-2001, and then turned up afterward. It is broadly consistent with changes in the growth rate during the period.

## V. Implications

The standard economic analysis takes it for granted that production factors move fast enough from low to high productivity firms and sectors, and that as a consequence, sooner or later they enjoy the same (highest) marginal productivity; otherwise, it contradicts the concept of equilibrium. However, we have some evidences suggesting that there is always productivity dispersion in the economy. As we referred to in the Introduction, Mortensen (2003) analyzing wage dispersion argued that there is productivity dispersion across firms. Okun (1973) also argued that a part of the reason why we obtain Okun's law is "the upgrading of workers into more-productive jobs in a high-pressure economy."

In this paper, we have provided a solid foundation for Okun's argument. The most important point of Okun's argument and also our present analysis is that *the allocation of production factors is not independent of the level of the aggregate demand*. Rather, it depends crucially on the aggregate demand. As we explained in section II, the fundamental principle of statistical physics indeed tells us that it is impossible for production factors to achieve the same (highest) level of productivity. Rather, we must always observe a distribution of productivity in the economy. Moreover, the theory indicates it is an exponential distribution (the Maxwell-Boltzmann), the exponent of which depends inversely on the level of aggregate demand. As the aggregate demand rises, production factors are mobilized to high-productivity firms and sectors just as Okun argued.

In section III, we showed that there does indeed exist distributions of productivity across workers, firms, and sectors. A serious problem for the theory of stochastic macro-equilibrium is, however, that the observed distribution of productivity across workers is not exponential, but obeys the power-law. We reconciled this empirical observation with

the basic theory by introducing the assumption that productivity dispersion across firms obeys the power-law rather than being uniform, and also another plausible assumption that the aggregate demand is not constant, but fluctuates.

In section IV, first, we explained how the power distribution of productivity across firms arises in a simple stochastic model. To obtain the power-law distribution, we need to make a crucial assumption that the higher the current level of productivity, the greater the probability of either a unit increase or decrease of productivity. Obviously, the probability of either a birth or a death is greater in a city with a large population than in a small city. By analogy, the above assumption seems quite natural.

The “size effect” in productivity (or TFP) growth has been much discussed in endogenous growth theory because its presence evidently contributes to endogenous growth. See, for example, Solow (2000). In growth theory, an increase in productivity is mostly identified with pure technological progress so that it is directly linked to R & D investment. However, to obtain the power distribution of productivity across firms, we must assume the significant probability of *decrease* in productivity. This suggests strongly that productivity changes facing firms are caused not only by technical progress, but also by *the allocative disturbances to demand*. Incidentally, Davis, Haltiwanger, and Schuh (1996) report that unlike job creation, job destruction for an industry is not systematically related to total factor productivity (TFP) growth; Namely, job destruction occurs in high TFP growth industries as frequently as in low TFP growth industries (their Table 3.7 on page 52). This fact also suggests the presence of the significant demand reallocation.

As Davis, Haltiwanger and Schuh (1996) rightly emphasize, the allocative demand shocks play a very important role in the macroeconomy. At the same time, the aggregate demand also plays a crucial role because the allocation of resources and production factors

depends crucially on the level of aggregate demand. Put simply, the frontier of the production possibility set is a never-never land. The higher the level of aggregate demand, the closer the economy is to the frontier.

### **Appendix A. Notes on Data used in section III.A**

As noted in section III, we used the Nikkei-NEEDS (Nikkei Economic Electronic Data-bank System) database for analysis of empirical distribution of productivity. This database is a commercial product available from Nikkei Media Marketing, Inc. (2008) and contains financial data of all the listed firms in Japan. As such, it is a well-established and representative database, widely used for various purposes from research to practical business applications. For our purpose, we used their 2007 CD-ROM version and extracted data for the period between 1980 and 2006. It covers some 1,700 to 3,000 firms and four million to six million workers.

We have found that in certain cases, the productivity calculated is unrealistically large. For example, firms that became stock-holding firms report huge reductions of the number of employees, while maintaining the same order of revenues in the year. This results in absurd values of labor productivity  $c$  for that year. Because of these abnormalities, we have excluded top-ten firms in terms of productivity each year. This roughly corresponds to excluding firms with productivity  $c > 10^9$  yen/person. We experimented with the analyses of several different cuts, *i.e.*, with cutting top-twenty firms, and so on. The results obtained remained basically the same as reported in the main text.

The values of the Pareto indices  $\mu_W, \mu_F, \mu_S$  given in this paper are determined by fitting the data with the GB2-distribution described in Kleiber and Kotz (2004) by the

maximum likelihood method.

## Appendix B. The Relationship between Aggregate Demand $D$ and the Exponent $\beta$

In this appendix, we prove the following three basic properties (i)–(iii) of the temperature-aggregate demand relation (54) in Macro-Equilibrium:

- (i) The temperature  $T = 1/\beta$  is a monotonically increasing function of the aggregate demand  $D$ . We can prove it using equation (54) as follows:

$$(70) \quad \frac{dD}{dT} = -\frac{1}{T^2} \frac{dD}{d\beta} = \beta^2 \frac{d^2}{d\beta^2} \ln Z(\beta) = \beta^2 (\langle c^2 \rangle_\beta - \langle c \rangle_\beta^2) \geq 0.$$

where  $\langle c^n \rangle_\beta$  is the  $n$ -th moment of productivity defined as follows:

$$(71) \quad \langle c^n \rangle_\beta \equiv \frac{1}{Z(\beta)} \int_0^\infty c^n P^{(F)}(c) e^{-\beta c} dc.$$

Comparing (53) and (56), we know that  $\langle c \rangle_\beta = D$ . This is a natural result. As the aggregate demand  $D$  rises, workers move to firms with higher productivity. It corresponds to the higher temperature according to the weight factor  $e^{-\beta c}$ .

- (ii) For  $T \rightarrow 0$  ( $\beta \rightarrow \infty$ ),

$$(72) \quad D \rightarrow 0.$$

This is evident from the fact that in the same limit the integration in Eq. (53) is dominated by  $c \simeq 0$  due to the factor  $e^{-\beta c}$ , and the integrand has the extra factor of  $c$  compared to the denominator  $Z(\beta)$ .

(iii) For  $T \rightarrow \infty$  ( $\beta \rightarrow 0$ ),

$$(73) \quad D \rightarrow \int_0^\infty c P^{(F)}(c) dc (= \langle c \rangle_0).$$

This can be established based on the property (i) because  $D = \langle c \rangle_\beta \rightarrow \langle c \rangle_0$  as  $\beta \rightarrow 0$  and  $Z(0) = 1$ .

An example of the relation between  $D$  and  $T \equiv 1/\beta$  is given in Figure XII.

### Acknowledgement

The authors would like to thank Dr. Y. Ikeda (Hitachi Research Institute) and Dr. W. Souma (NiCT/ATR CIS) for discussions. Part of this research was supported by a grant from Hitachi Research Institute. Yoshikawa's research is supported by the Research Institute of Economy, Trade and Industry. Yukawa Institute for Theoretical Physics (YITP) at Kyoto University supported this research in two ways; their computing facility was used for part of the calculations and discussions with participants at the YITP workshop YITP-W-07-16 "*Econophysics III –Physical approach to social and economic phenomena*" were helpful for completion of this work.

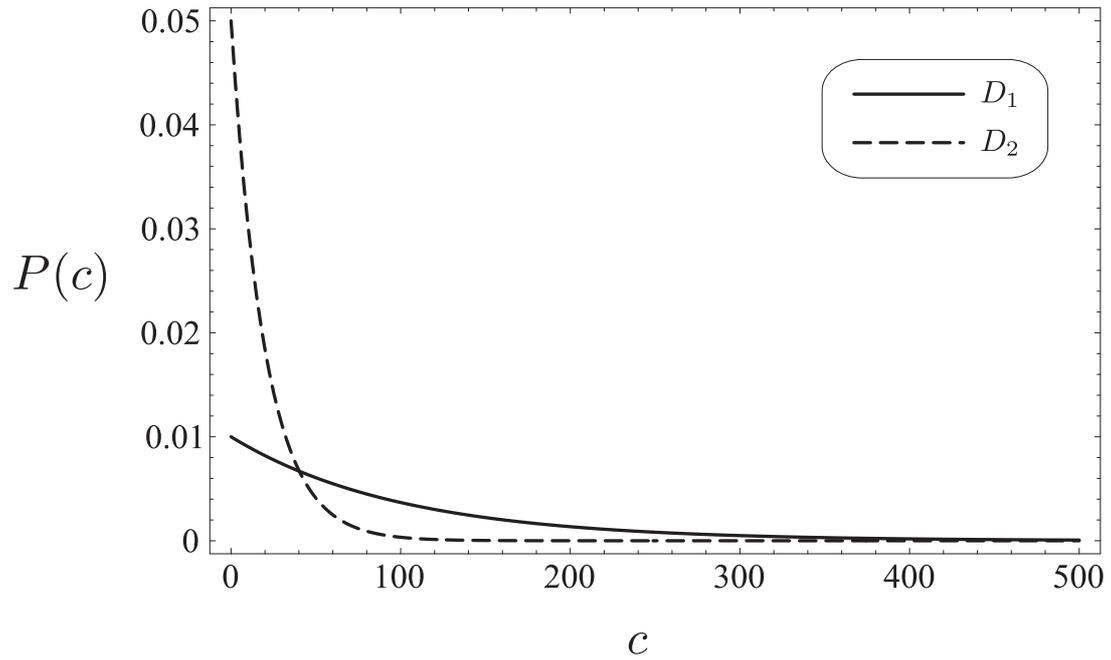
## References

- Aoki, Masanao and Hiroshi Yoshikawa**, *Reconstructing Macroeconomics – A Perspective from Statistical Physics and Combinatorial Stochastic Processes*, Cambridge, U.K.: Cambridge University Press, 2007.
- Beck, C.** 2006. “Superstatistical Brownian motion.” *Progress of Theoretical Physics. Supplement*, 162: 29–36.
- **and E. G. D. Cohen.** 2003. “Superstatistics.” *Physica A*, 322: 267–275.
- Champernowne, D.G.** 1973. *The Distribution of Income between Persons*, Cambridge: Cambridge University Press.
- Davis, S.J., J.C. Haltiwanger, and S. Schuh.** 1996. *Job Creation and Destruction*, Cambridge, MA: MIT Press.
- Fay, Jon A. and James L. Medoff.** 1985. “Labor and Output over the Business Cycle: Some Direct Evidence.” *American Economic Review*, 75: 63–655.
- Gabaix, Xavier.** 2005. “The Granular Origins of Aggregate Fluctuations.” MIT and NBER, DP.
- Ijiri, Yuji and Herbert A. Simon.** 1975. “Some Distributions Associated with Bose-Einstein Statistics.” *Proceedings of the National Academy of Sciences of the United States of America*, 72(5): 1654–1657.
- **and –** . 1977. *Skew Distributions and the Sizes of Business Firms*, Amsterdam: North-Holland Pub. Co..

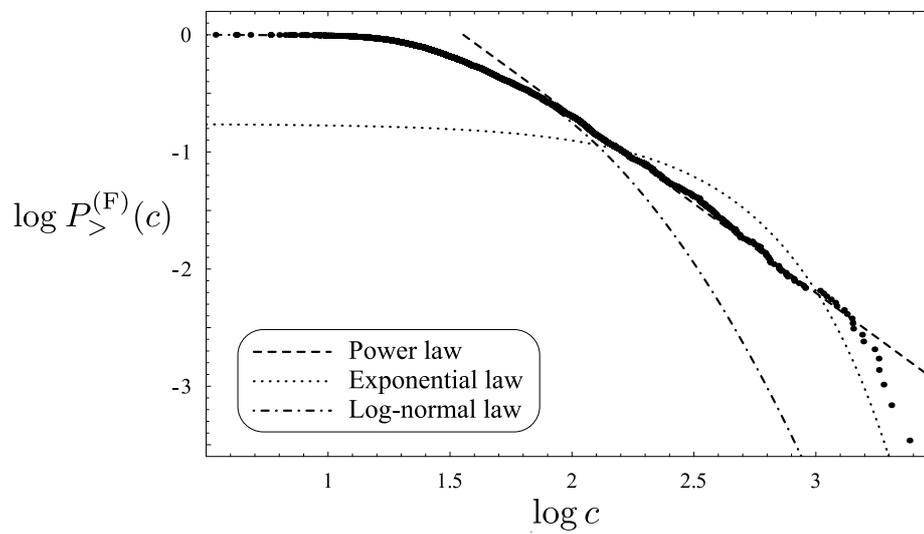
- Keynes, J.M.** 1936. *The General Theory of Employment, Interest, and Money*, London: Macmillan.
- Kleiber, C. and Samuel Kotz.** 2004. *Statistical Size Distributions in Economics and Actuarial Sciences*, Hoboken, New Jersey: John Wiley and Sons, Inc..
- Marsili, M. and Y.-C. Zhang.** 1998 “Interacting individuals leading to Zipf’s law.” *Physical Review Letters*, 80(12): 2741–2744.
- Mortensen, D. T.** 2003. *Wage Dispersion*, Cambridge, MA: MIT Press.
- Nikkei NEEDS CD-ROM.** 2008. Nikkei Media Marketing, Inc.. <http://www.nikkeimm.co.jp/english/index.html> and [http://www.nikkeimm.co.jp/service/macro/needs/con\\_needs\\_cd.html](http://www.nikkeimm.co.jp/service/macro/needs/con_needs_cd.html) (in Japanese, accessed May 12, 2008).
- Okun, A. M.** 1962. “Potential GNP: Its Measurement and Significance.” in “Proceedings of the Business and Economic Statistics Section” American Statistical Association, 98–104.
- . 1973. “Upward Mobility in a High-Pressure Economy.” *Brookings Papers on Economic Activity*, 1: 207–261.
- Solow, Robert M.** 2000. *Growth Theory: An Exposition*, second ed., New York: Oxford University Press.
- Sutton, John.** 1997. “Gibrat’s Legacy.” *Journal of Economic Literature*, March: 40–59.
- Tobin, J.** 1972. “Inflation and Unemployment.” *American Economic Review*, 62: 1–18.

**Touchette, H. and C. Beck.** 2005. "Asymptotics of superstatistics." *Physical Review E*, 71(1): 16131-1-6.

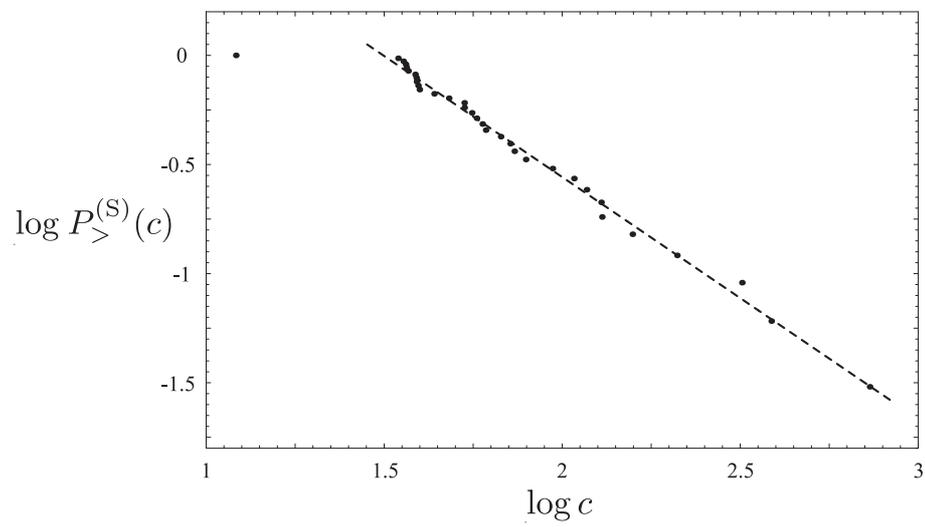
**Yoshikawa, Hiroshi,** "The Role of Demand in Macroeconomics," *The Japanese Economic Review*, 2003, 54, 1.



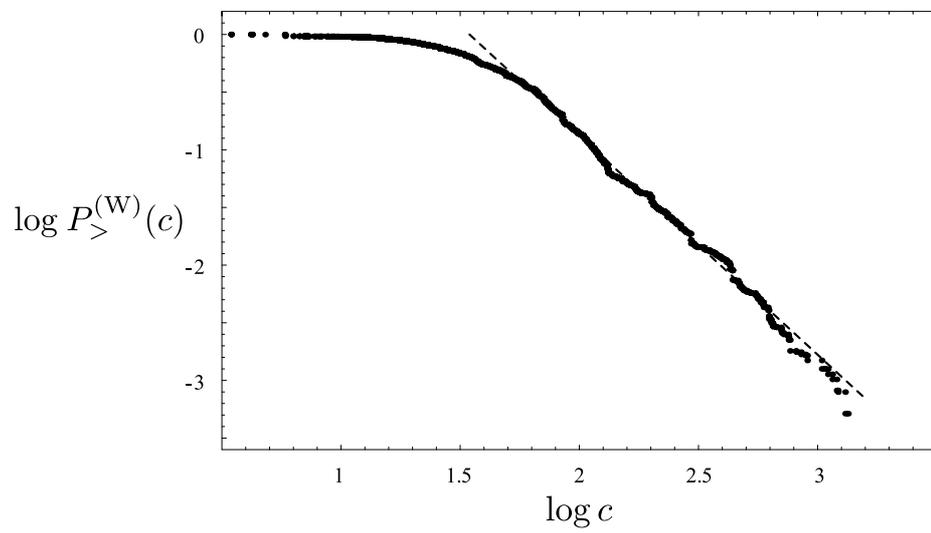
**Figure I:** PRODUCTIVITY DISTRIBUTION PREDICTED BY THE STOCHASTIC MACRO-EQUILIBRIUM THEORY FOR TWO DIFFERENT VALUES OF THE AGGREGATE DEMAND  $D$ ,  $D_1 > D_2$  Notes: The values of  $\beta$  are obtained from equation (18) for  $D_1 = 20$  and  $D_2 = 100$ .



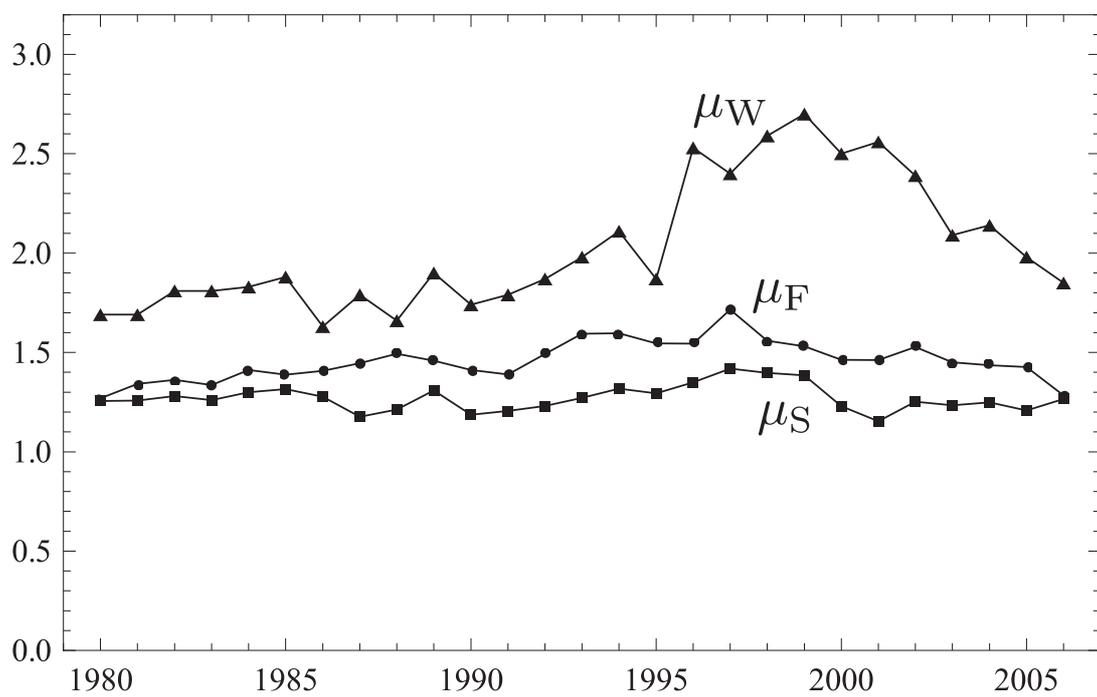
**Figure II:** PRODUCTIVITY DISTRIBUTION ACROSS FIRMS (2005) *Notes:* The productivity  $c$  is in the unit of  $10^6$ yen/person. The best fit for the exponential law and the power law is obtained for  $10 < c < 3000$ .



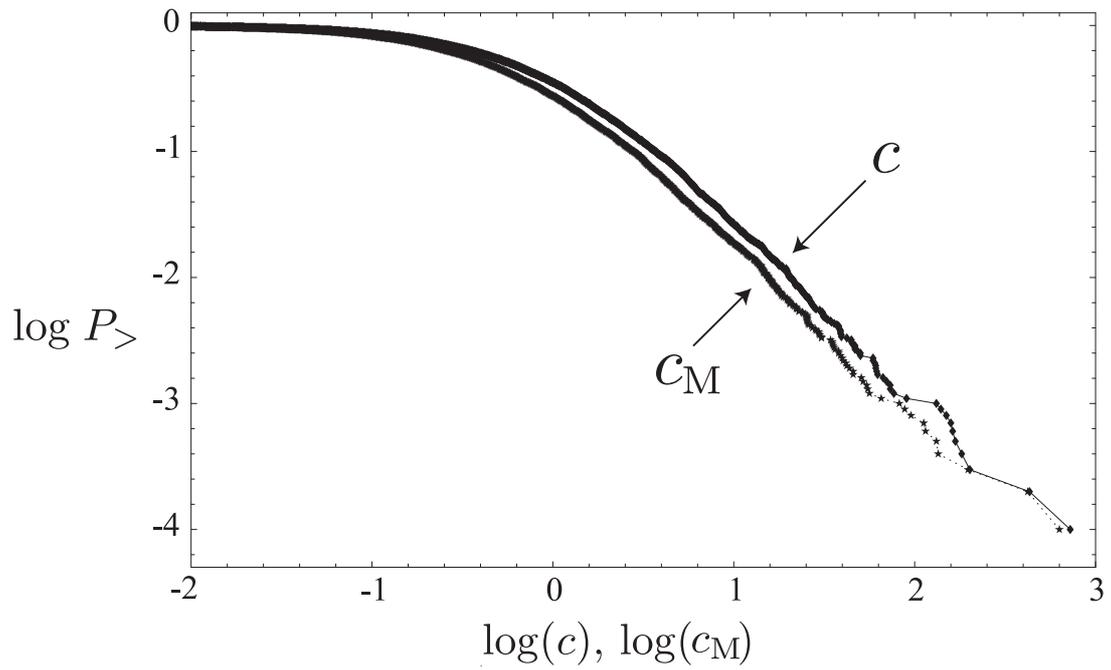
**Figure III:** PRODUCTIVITY DISTRIBUTION ACROSS *Business Sectors* (2005)



**Figure IV:** PRODUCTIVITY DISTRIBUTION ACROSS *Workers* (2005)

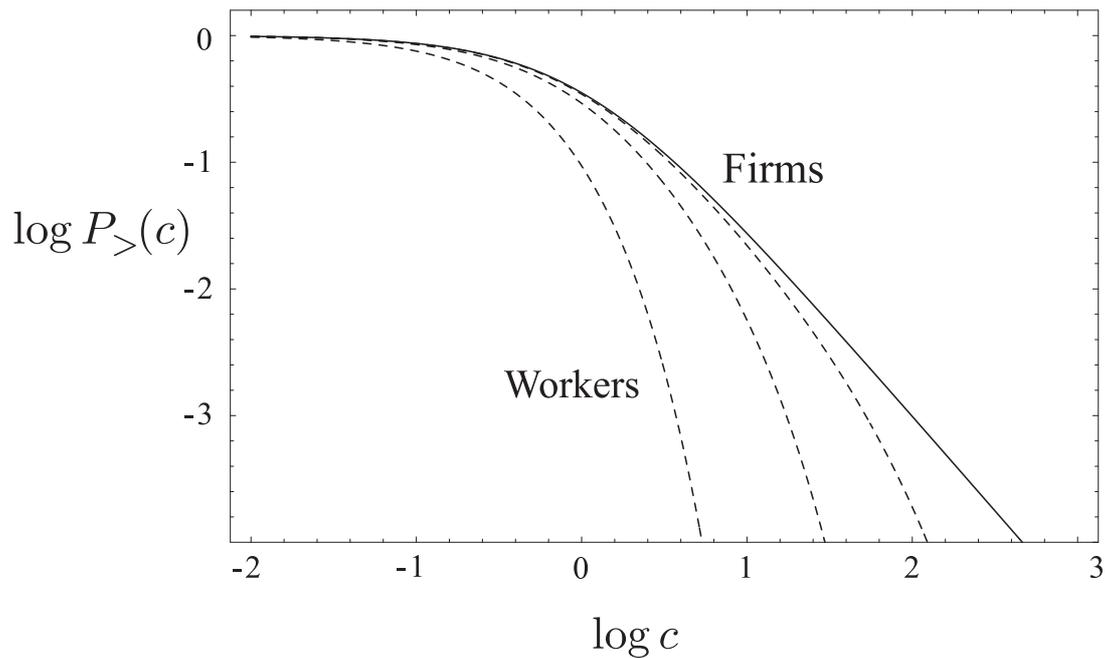


**Figure V:** PARETO INDICES OF THE PRODUCTIVITY DISTRIBUTIONS ACROSS WORKERS, FIRMS, AND INDUSTRIAL SECTORS



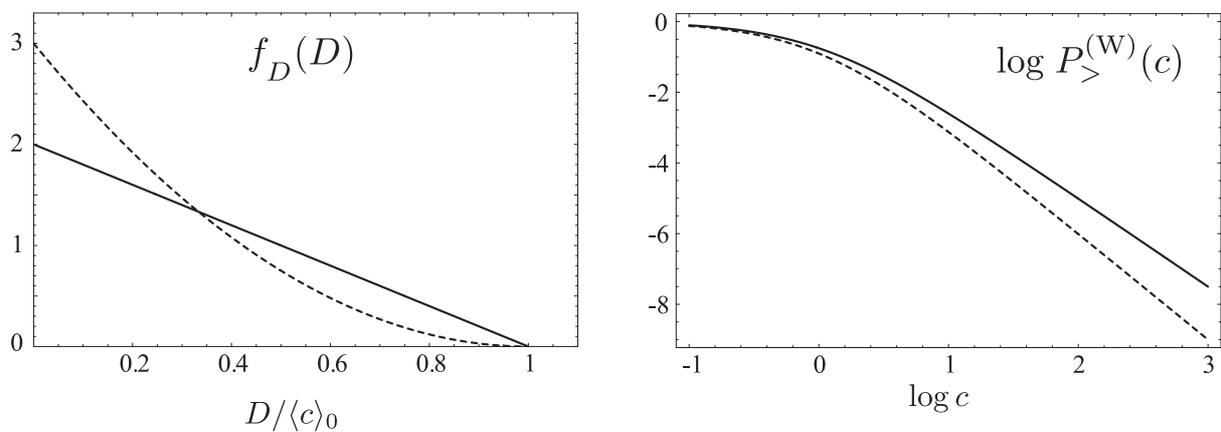
**Figure VI:** RANK-SIZE LOG-LOG PLOTS OF THE SIMULATION RESULTS *Notes:*

Distribution for  $c$  is for the Monte-Carlo data generated from a  $P(c)$  with  $\mu_M = 1.5$ . Data for  $c_M$  is then created with  $P_\alpha(\alpha)$  distributed from  $\alpha = 0.5$  to 1 uniformly.

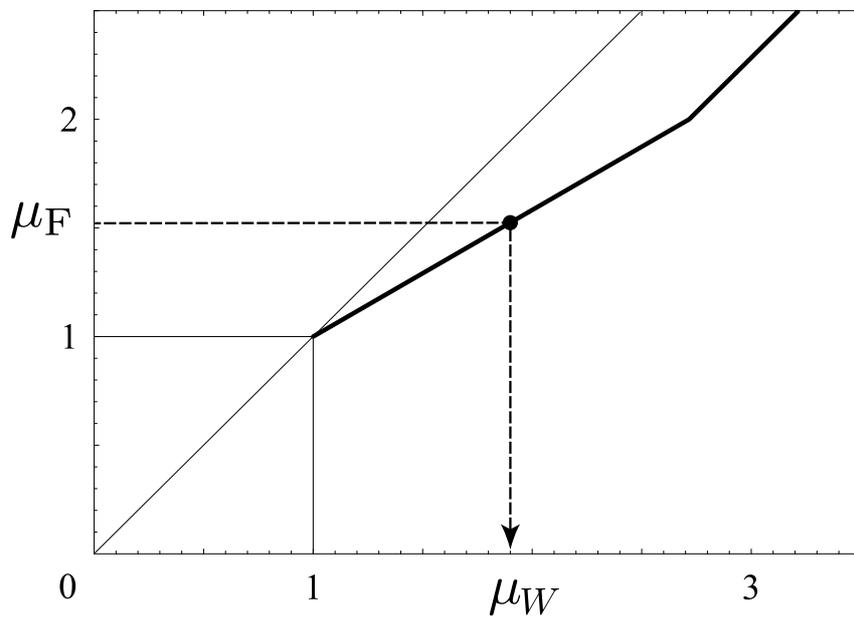


**Figure VII:** PRODUCTIVITY DISTRIBUTIONS ACROSS WORKERS AND FIRMS

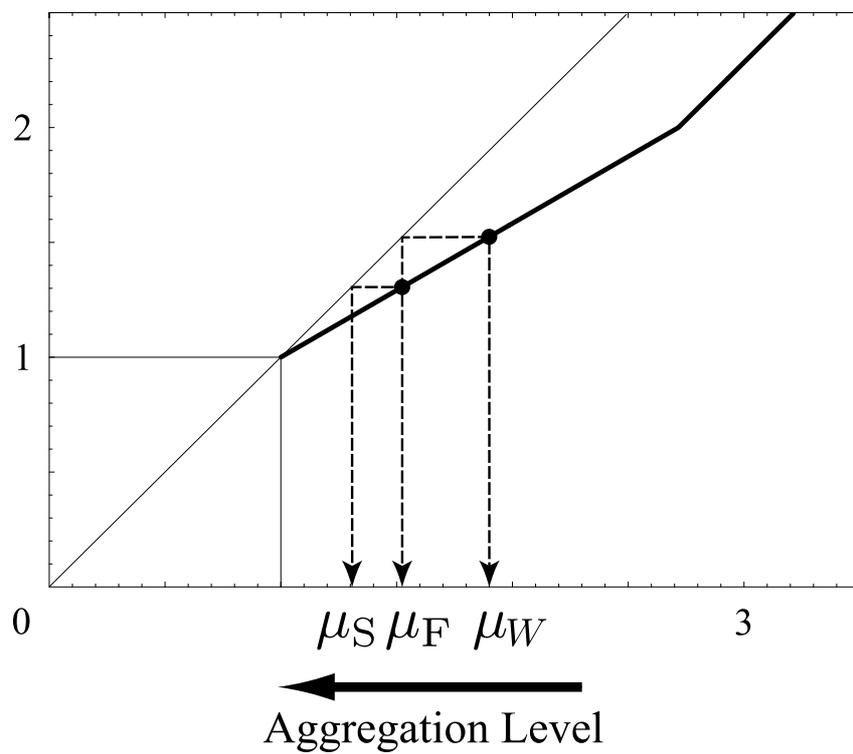
*Notes:* The solid curve is for firms with  $\mu_F = 1.5$ , while the dashed curves are for workers with  $\beta = 0.01, 0.1, 1$ , respectively. When  $\beta$  is small enough, the distribution across workers is close to that of firms. As  $\beta$  increases, the distribution across workers is suppressed for large  $c$  due to the exponential factor  $e^{-\beta c}$ .



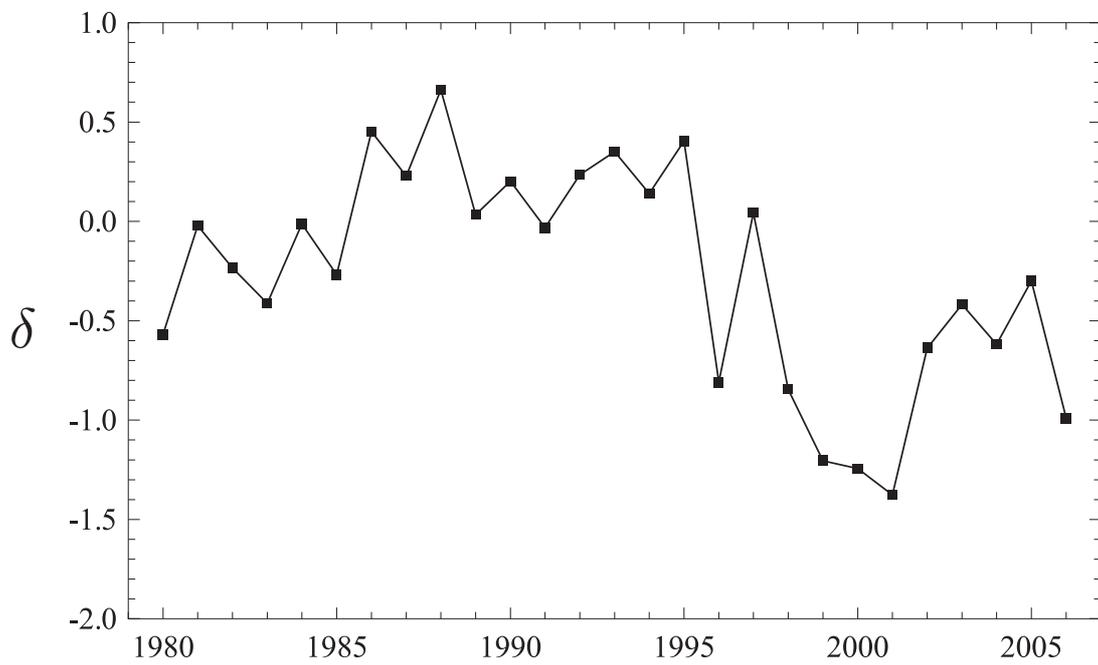
**Figure VIII:** THE DISTRIBUTION OF THE AGGREGATE DEMAND  $D$ ,  $f_D(D)$  (LEFT) AND THE CORRESPONDING CUMULATIVE PRODUCTIVITY DISTRIBUTION  $P_{>}^{(W)}(c)$  (RIGHT) *Notes:* The solid curves are for  $\delta = -1$  whereas the broken curves are for  $\delta = -2$ . The productivity distribution of firms is chosen to have  $\mu_F = 1.5$ .



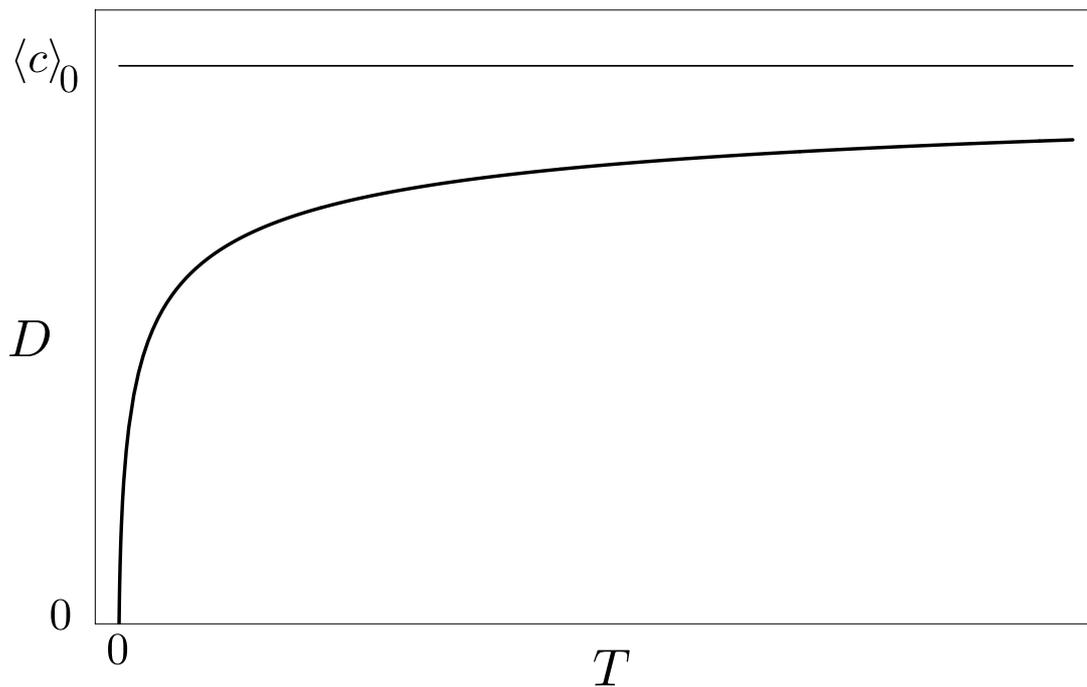
**Figure IX:** RELATION BETWEEN  $\mu_W$  AND  $\mu_F$  (68) *Notes:* The solid line is the relation (68), and the filled circle is the data.



**Figure X:** CHANGES OF THE PARETO INDEX AS THE AGGREGATION LEVEL CHANGES IN TWO STEPS, EACH WITH A DIFFERENT VALUE OF  $\delta$



**Figure XI:** THE VALUES OF  $\delta$  CALCULATED FROM EQUATION (68) FOR JAPANESE LISTED FIRMS



**Figure XII:** THE RELATION BETWEEN THE AGGREGATE DEMAND  $D$  AND THE TEMPERATURE  $T$  ( $= 1/\beta$ ) *Notes:* For a productivity distribution of firms with  $\mu_F = 1.5$ .