J Econ Interact Coord (2010) 5:27–54
DOI 10.1007/s11403-010-0063-9

REGULAR ARTICLE

Productivity dispersion: facts, theory, and implications

Hideaki Aoyama · Hiroshi Yoshikawa · Hiroshi Iyetomi · Yoshi Fujiwara

Received: 2 October 2009 / Accepted: 18 March 2010 / Published online: 3 April 2010

Abstract This paper discusses the stylized facts, the theory, and the remaining problems of productivity dispersion, which is essentially related to the concept of equilibrium in the neoclassical theory. Empirical study of data relating to 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 on the basic principle of statistical physics and on the concept of superstatistics, an approach that accommodates fluctuations of aggregate demand. We show that the allocation of production factors depends crucially on the level of aggregate demand, and that the higher the level of aggregate demand, the closer the economy is to the frontier of the production possibility set.

The authors would like to thank Dr. Y. Ikeda and Dr. W. Souma for various discussions related to the topics discussed in this paper. We also wish to acknowledge the generous support of the Hitachi Research Institute and Hitachi Ltd for a grant towards part of our research, and the Research Institute of Economy, Trade and Industry for supporting Hiroshi Yoshikawa's work. We are grateful to the Yukawa Institute for Theoretical Physics (YITP) at Kyoto University for allowing us to use their computing facility, and to the participants at the YITP workshop YITP-W-07-16 *"Econophysics III Physical approach to social and economic phenomena"* for helpful comments while we were completing our investigation.

H. Aoyama (✉)
Department of Physics, Kyoto University, Kyoto 606-8501, Japan
e-mail: hideaki.aoyama@gmail.com

H. Yoshikawa
Faculty of Economics, University of Tokyo, Tokyo 113-0033, Japan

H. Iyetomi
Department of Physics, Niigata University, Niigata 950-2181, Japan

Y. Fujiwara
ATR Laboratories, Kyoto 619-0288, Japan

Keywords Labor productivity · Superstatistics · Aggregate demand

1 Introduction—why productivity dispersion?

Macroeconomics was once synonymous with Keynesian economics with its central themes of depression and unemployment, while the neoclassical theory or microeconomics, on the other hand, explains how resources are efficiently allocated in markets on the assumption that production factors such as labor and capital are fully employed. The pillars of neoclassical economics are rational choice or optimization by individual economic agent such as household and firm, and the price mechanism for efficient resource allocation. Together, macroeconomics and microeconomics constitute economic theory.

This consensus has been weakening over the last thirty years, and the pursuit of micro-foundations for macroeconomics has produced such theories and concepts as the natural rate of unemployment (Phelps 1970), rational expectations (Lucas 1972), and the real business cycle theory (Kydland and Prescott 1982). As a consequence, the profession has progressively turned away from Keynesian macroeconomics to a neoclassical micro-founded macroeconomics. In illustration, we can simply cite Lucas (1987)'s triumphant declaration:

> "The most interesting recent developments in macroeconomic theory seem to me describable as the reincorporation of aggregative problems such as inflation and the business cycle within the general framework of "microeconomic" theory. If these developments succeed, the term "macroeconomic" will simply disappear from use and the modifier "micro" will become superfluous. We will simply speak, as did Smith, Ricardo, Marshall and Warlras, of economic theory. If we are honest, we will have to face the fact that at any given time there will be phenomena that are well-understood from the point of view of the economic theory we have, and other phenomena that are not. We will be tempted, I am sure, to relieve the discomfort induced by discrepancies between theory and facts by saying that the ill-understood facts are the province of some other, different kind of economic theory. Keynesian "macroeconomics" was, I think, a surrender (under great duress) to this temptation. It led to the abandonment, for a class of problems of great importance, of the use of the only "engine for the discovery of truth" that we have in economics. Now we are once again, putting this engine of Marshall's to work on the problems of aggregate dynamics. (Lucas 1987, pp. 107–108.)"

Twenty years after ringing assertion the global financial crisis and the world-wide deep recession (2008–2009), which some even called depression, casted serious doubts on mainstream neoclassical micro-founded macroeconomics. In the 2009 Lionel Robbins lectures at the London School of Economics on June 10th, Paul Krugman, the 2008 Nobel laureate, feared that most macroeconomics of the past 30 years was "spectacularly useless at best, and positively harmful at worst". However, while the pendulum seems to be swinging back once again towards Keynesian economics, the fundamental problem remains unresolved. Namely, does the fundamental difference between Keynesian economics and the neoclassical theory lie, or to put it another way, what are the proper microeconomic foundations for Keynesian macroeconomics?

The standard neoclassical theory 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, in which production factors are fully employed. Factor endowment, together with preferences and technology, determines the equilibrium in such a way that the marginal products are equal across sectors and firms.

This orthodox theory was challenged by Keynes (1936), who pointed out that the utilization of production factors is not complete, and therefore, that factor endowment is not an effective determinant of equilibrium. In *the General Theory*, Keynes identified the incomplete utilization of production factor with the presence of involuntary unemployment of labor, and many controversies revolved around the theoretically ambiguous notion of "involuntary" unemployment. However, for Keynes' economics to make sense it is unnecessary to posit the existence of involuntary unemployment, since the real problem is not a binary distinction between work and unemployment, and employed workers in fact work at different levels of productivity. Rather the issue is whether 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 a 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 is a real phenomenon, for example the celebrated research of Okun. 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 1, because theoretically it should be equal to the share of labor income. Okun (1962), however, found this elasticity to be 3 for the U.S. economy, a finding that 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, namely (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 clearly recognized and indeed, by way of his celebrated law, demonstrated that there is 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 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)"

Morten answers this question thus:

> "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)"

Clearly, Okun and Mortensen both recognize that the real economy is not in a neoclassical equilibrium where marginal products of labor are equal across firms and sectors. Since, as explained above, the presence of distribution of labor productivity *constitutes* the micro-foundations for the Keynesian macroeconomics, productivity dispersion is not merely one of many curious distributions discovered by econophysicists, but is essentially related to the concept of equilibrium in neoclassical economics.

In what follows, we first explain the concept of *stochastic macro-equilibrium*. Then, drawing on our previous empirical investigations, we summarize the "stylized facts" of the labor productivity distribution, and the challenges they present. Finally, we discuss remaining problems and directions for future research.

2 The stochastic macro-equilibrium

Starting from the principles of statistical physics Yoshikawa (2003) proposed the Maxwell–Boltzmann distribution to describe labor productivity in equilibrium, a concept akin to the *stochastic macro-equilibrium* advanced by Tobin (1972) in his attempt to explain the observed Phillips curve. Tobin argued 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)"

However, though this proposal remained only verbal the fundamental principle of statistical physics, in fact, provides exact foundations for the concept of macro-equilibrium. A case in point is productivity dispersion, and in what follows we will explain the distribution of labor productivity in equilibrium by drawing on Aoki and Yoshikawa (2007).

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

$$c_1 < c_2 < \cdots < c_K. \tag{1}$$

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

$$\sum_{k=1}^{K} n_k = N. \tag{2}$$

The neoclassical theory takes it for granted that all workers enjoy the highest marginal labor productivity, i.e. n_K is N while $n_k (k \neq K)$'s are all zero. Instead, we will seek the most probable distribution of productivity across workers under suitable constraints.[1] The possible number of particular configurations $(n_1, n_2, \ldots, n_K)$, $w_{\{n\}}$ is

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

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

$$P_{\{n\}} = \frac{1}{K^N} \frac{N!}{\prod_{k=1}^{K} n_k!}. \tag{4}$$

Following the fundamental principle of statistical physics, we postulate that the configuration $\{n_1, n_2, \ldots, n_K\}$ that maximizes $P_{\{n\}}$ under suitable constraints is realized in equilibrium, an idea 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. Of course, the micro agents in an economy will only optimize their functions,

[1] It can be observed that the productivity c_k corresponds to the allowed energy level, and workers correspond to distinguishable particles.

but such agents face constraints, and even their objective functions change due to idiosyncratic shocks. The situation is identical to that in physics where we cannot know the starting conditions for all particles. Admittedly, the number of micro agents in an economy is less than the counterpart quantity in physics, which is typically 10^{23}, but there are still 10^6 firms and 10^7 households, and these are large numbers. Consequently, it is infeasible and would be meaningless to analyze the micro behavior of agents in great detail. Instead any study of a macro-system must regard the behaviors of micro agents as stochastic even though their behaviors are purposeful, just as it is the fundamental principle of statistical physics that we observe the state of a macro-system that maximizes probability (4).

Indeed, the "stochastic approach" was once popular in diverse areas of study such as income distribution and firm and city size, the primary examples being Champernowne (1973) and Ijiri and Simon (1977). However, this research movement eventually lost momentum, for reasons described by Sutton (1997):

> "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[2]

> "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)"

In fact, as the system under investigation becomes larger and more complicated, the importance of the stochastic approach increases, and it is practically the identifying characteristic of a macro-system to which the statistical physical approach can be usefully applied. Since the natural sciences routinely use the "stochastic approach" when analysing macro-systems there is little reason to neglect it in macroeconomics.

Toward our goal, we define the following quantity;

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

where p_k is defined as

$$p_k \equiv \frac{n_k}{N}. \tag{6}$$

[2] Incidentally, with the central limit theorem evidently in mind, Sutton (1997) identifies what he calls "purely statistical" effects with the effects of *independent* stochastic variables. However, the methods of statistical physics are effective even when stochastic variables are correlated, i.e. when they are *not* independent.

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

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

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, firstly the normalization condition

$$\sum_{k=1}^{K} p_k = 1. \tag{8}$$

This is, of course, equivalent to the resource constraint, $\sum_{k=1}^{K} n_k = N$. The second constraint requires that the total output (GDP) Y is equal to the aggregate demand $\widetilde{D}$,

$$N \sum_{k=1}^{K} c_k p_k = Y = \widetilde{D}. \tag{9}$$

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

$$D = \frac{\widetilde{D}}{N}. \tag{10}$$

For the sake of simplicity we will call D the aggregated demand, and, as we will shortly see, D determines the state of stochastic macro-equilibrium.

The maximization of the entropy S is a generic framework for analysing a macro-system consisting of a large number of micro units each subject to randomness, and deserves to be better known in economics, though one notable example can be found in Foley (1994) which applies the framework to the Walrasian general equilibrium.

We maximize the following Lagrangian form:

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

Differentiating Eq. (11) with respect to p_i, we obtain

$$\ln p_k + (1 + \alpha) + \beta c_k = 0. \tag{12}$$

This yields

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

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

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

Here, $Z(\beta)$ is what is called the partition function in physics, and ensures that the p_k's sum to 1.

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

Thus, the distribution which maximizes the probability (4) under two constraints, (8) and (9), is exponential, and is referred to in physics as the *the Maxwell–Boltzmann distribution*. We will refer to it here as the *the equilibrium distribution*, and note that *the exponent of the equilibrium distribution is the Lagrangian multiplier β corresponding to the aggregate demand constraint in Eq.* (11). In passing, it is interesting to note that the exponent β is equivalent to the inverse of *temperature*, $1/T$ in physics.

Equations (9), (10), (14) and (15) yield the following:

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

This equation relates the aggregate demand D to the exponent of the distribution β by way of the partition function $Z(\beta)$. Note that at this stage, the distribution of productivity c_k is arbitrary, but that once it is established the partition function $Z(\beta)$ is defined by Eq. (15), and it, in turn, determines the relation between D and β.[3]

Suppose that the distribution of productivity across firms is *uniform*, that is, $c_k = k\Delta_c$, where Δ_c is a constant. Then, from Eq. (15), we obtain

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

under the assumption that $\beta\Delta_c \ll 1$ and $\beta K \Delta_c \gg 1$. Therefore, in this case, we know from Eqs. (16) and (17) that the exponent of the equilibrium exponential distribution, β is equal to the inverse of aggregate demand:

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

In this case, because the inverse of β is temperature, aggregate demand is equivalent to temperature.

[3] This is identical to 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 established for the system under investigation.

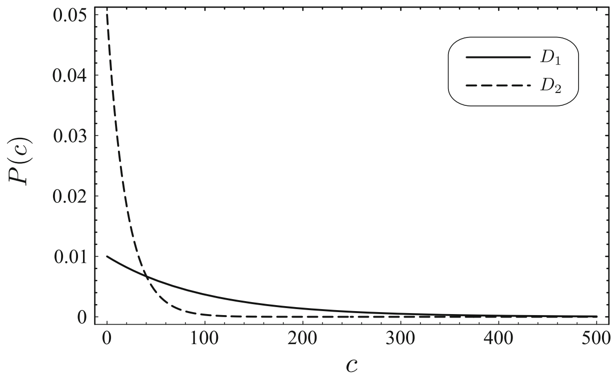

Fig. 1 Productivity distributions as predicted by the stochastic macro-equilibrium theory for two different values of the aggregate demand D, $D_1 > D_2$. *Notes* The values of β are obtained from Eq. (18) for $D_1 = 20$ and $D_2 = 100$

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

$$p_k = \frac{n_k}{N} = \frac{e^{-\frac{c_k}{D}}}{\Sigma_{k=1}^{K} e^{-\frac{c_k}{D}}}. \tag{19}$$

When the aggregate demand D is high, the distribution becomes flatter, meaning that production factors migrate to firms or sectors with high productivity, and vice versa. Figure 1 shows two distributions of productivity corresponding to high aggregate demand D_1 and low D_2.

This 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. It has been plausibly suggested that this provides the proper microeconomic foundation for Keynesian economics. However, the Maxwell–Boltzmann distribution is precisely exponential, and it must be asked whether this actually fits our empirical observations. In the following section we address this problem by turning to the stylized facts of the matter.

3 Stylized facts

Much research has explored the empirical distribution of labor productivity, for example Delli Gatti et al. (2008), analyzes the distribution of labor productivity in France and Italy. Aoyama et al. (2009) explores the same problem using the Nikkei-NEEDS database (Nikkei Media Marketing, Inc. 2008) for Japan, and Ikeda and Souma (2009) studies the empirical distribution of labor productivity for Japan and the United States. There is a broad consensus among these studies, and we will here summarize the results obtained by Aoyama et al. (2009).

3.1 Empirical distributions

Aoyama et al. (2009) studies productivity distributions at three aggregate levels, namely, across workers, firms, and industrial sectors. In order to explain the empirical distributions, we find that the continuous model is more convenient than the discrete model. Accordingly, we define, for example, the probability density function of firms with productivity c as $P^{(\mathrm{F})}(c)$. The number of firms with productivity between c and $c + dc$ is now $K P^{(\mathrm{F})}(c)dc$, which satisfies the following normalization condition:

$$\int_0^\infty P^{(\mathrm{F})}(c)\, dc = 1. \tag{20}$$

In Sect. 2, we explained that, grounded on the principle of statistical physics, equilibrium productivity dispersion *across workers* becomes the exponential distribution (19), *assuming that the distribution of productivity across firms is uniform*. Here, $P^{(\mathrm{F})}(c)$ is not defined as uniform, but, rather, the distribution $P^{(\mathrm{F})}(c)$ is to be determined empirically.

We analogously define $P^{(\mathrm{W})}(c)$ for workers, and $P^{(\mathrm{S})}(c)$ for industrial sectors. We denote their respective cumulative probability distributions by $P_>^{(\mathrm{S,F,W})}(c)$:

$$P_>^{(*)}(c) = \int_c^\infty P^{(*)}(c')\, dc' \quad (* = \mathrm{S, F, W}). \tag{21}$$

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, where the 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 we will restrict our discussion to the cumulative probability rather than the probability density function.

The productivity distribution across firms, or, to be precise, $\log P_>^{(\mathrm{F})}(c)$ based on the Nikkei-NEEDS data plotted against $\log c$ for the year 2005, is shown in Fig. 2 (Aoyama et al. 2009). The dots are the data points, each dot corresponding 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, where the exponential and log-normal distributions are represented by their respective curves, and 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 Sect. 2 does not fit the actual data at all. For small values of c (low-productivity) the log-normal law (dash-dotted curve) fits well, and for large values of c (high-productivity) the power law (broken line) fits well, with a smooth transition from the former to the latter at around $\log c \simeq 2$. The result shown in Fig. 2 is for the year 2005, but a basically similar result holds good for other sample periods.

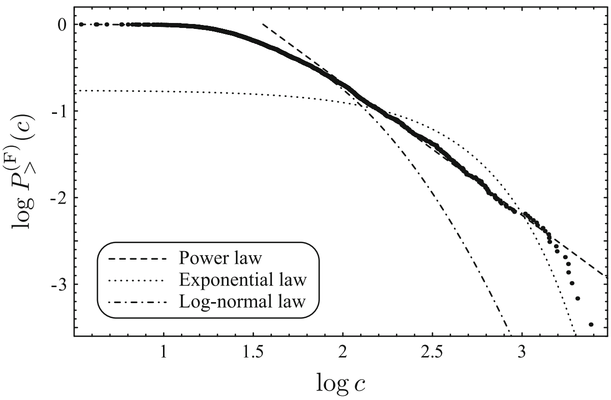

Fig. 2 Productivity distribution across firms (2005). *Notes* The productivity c is given in units of 10^6 yen/person, and the best fits for the exponential law and the power law are obtained for $10 < c < 3{,}000$. *Source* Aoyama et al. (2009)

The cumulative probability $P_>^{(F)}(c)$ for large values of c can be represented by the following power distribution:

$$P_>^{(F)}(c) \simeq \left(\frac{c}{c_0}\right)^{-\mu_F} \quad (c \gg c_0). \tag{22}$$

where c_0 is a parameter that defines the order of the productivity c.[4] The power exponent μ_F is called *the Pareto index*, and c_0 the Pareto scale. The probability density function is then given by the following:

$$P^{(F)}(c) = -\frac{d}{dc}P_>^{(F)}(c) \simeq \mu_F \frac{c^{-\mu_F - 1}}{c_0^{-\mu_F}} \quad (c \gg c_0). \tag{23}$$

In Aoyama et al. (2009), we also explored the productivity dispersion across the thirty-three industrial sectors defined by the Nikkei-NEEDS database. The productivity distribution across these sectors for the year 2005 is plotted in Fig. 3. 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 also analyzed in Aoyama et al. (2009), and is plotted in Fig. 4 for the year 2005. Again, we observe that the power law (broken line) fits the data very well for large values of c. A casual examination of Fig. 4 may make one wonder whether the share of workers for which the power law fits well is in fact small, however, this impression is mistaken. 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

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

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

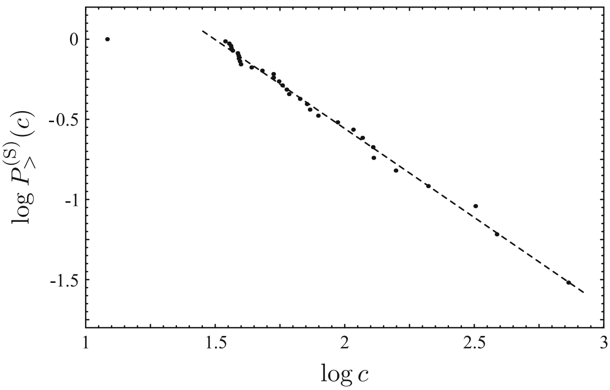

Fig. 3 Productivity distribution across industrial sectors (2005). *Source* Aoyama et al. (2009)

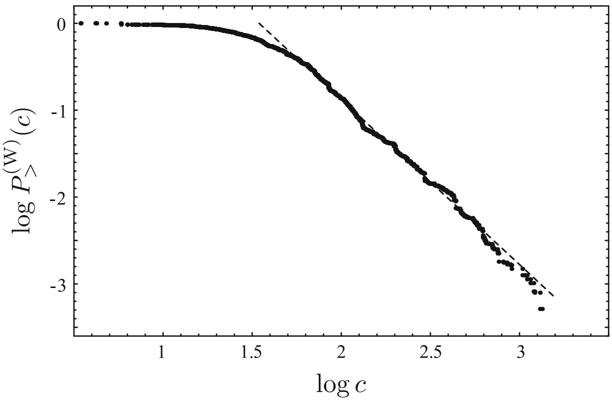

Fig. 4 Productivity distribution across Workers (2005). *Source* Aoyama et al. (2009)

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 Fig. 5 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.

3.2 Marginal versus average productivity

Before we conclude our discussion of the empirical distribution, it is extremely important to explore the relation between the *average* labor productivity and the *marginal* labor productivity. What we observe is the average productivity c, defined by $c = Y/L$, where Y is the output, and L the labor input. However, what matters theoretically is, of course, the unobserved *marginal* productivity c_{M} defined by

$$c_{\mathrm{M}} = \frac{\partial Y}{\partial L}. \tag{24}$$

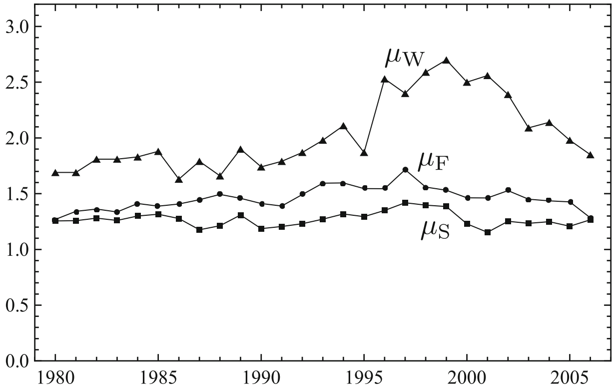

Fig. 5 Pareto indices of the productivity distributions across workers, firms, and industrial sectors. *Source* Aoyama et al. (2009)

Thus, we must explore the relationship between the distributions of these two different productivities.

Another problem is that in almost all empirical investigations, including our own in Aoyama et al. (2009), researchers measure L in terms of the number of workers, as such data is readily available. One might argue that it is theoretically more desirable to measure L in terms of work-hours, or for that matter, even in terms of work efficiency units. As we will see shortly, the effects of these possible "measurement errors" can also be handled in a similar way.

To explore the relation between the two productivities c and c_M we assume the Cobb-Douglas production function:

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

Under this assumption, the marginal productivity c_M is calculated as follows;

$$c_M = \frac{\partial Y}{\partial L} = \alpha \frac{Y}{L} = \alpha c. \tag{26}$$

Therefore, we arrive at the following relation:

$$c_M = \alpha\, c. \tag{27}$$

In general, the value of α differs across firms, and therefore, the distribution of c_M is, in general, different from that of c.

However, thanks to Eq. (27), we can relate the pdf of c_M, $P_M(c_M)$ to the joint pdf of c and α, $P_{c,\alpha}(c, \alpha)$ as follows:

$$P_{\mathrm{M}}(c_{\mathrm{M}}) = \int_0^1 d\alpha \int_0^\infty dc\, \delta(c_{\mathrm{M}} - \alpha c) P_{c,\alpha}(c,\alpha)$$
$$= \int_0^1 \frac{d\alpha}{\alpha}\, P_{c,\alpha}\left(\frac{c_{\mathrm{M}}}{\alpha}, \alpha\right). \tag{28}$$

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

$$P_{c,\alpha}(c,\alpha) = P^{(\mathrm{F})}(c) P(\alpha \mid c). \tag{29}$$

Here, $P(\alpha \mid c)$ is the conditional pdf, in other words it is the pdf of α for the fixed value of productivity c, which can be normalized as follows for any value of c:

$$\int_0^1 P(\alpha \mid c)\, d\alpha = 1. \tag{30}$$

We have already seen that $P^{(\mathrm{F})}(c)$ obeys the power law for large values of c:

$$P^{(\mathrm{F})}(c) = \mu_{\mathrm{F}} \frac{c^{-\mu_{\mathrm{F}}-1}}{c_0^{-\mu_{\mathrm{F}}}}. \tag{31}$$

From Eqs. (28), (29), and (31), we obtain the following:

$$P_{\mathrm{M}}(c_{\mathrm{M}}) = \mu_{\mathrm{F}} \frac{c_{\mathrm{M}}^{-\mu_{\mathrm{F}}-1}}{c_0^{-\mu_{\mathrm{F}}}} \int_0^1 d\alpha\, \alpha^{-\mu_{\mathrm{F}}} P\left(\alpha \,\middle|\, \frac{c_{\mathrm{M}}}{\alpha}\right). \tag{32}$$

Here, we have assumed that $P(\alpha \mid c)$ does not extend too near $\alpha = 0$, so that c_{M}/α stays in the asymptotic region. From Eq. (32) we can conclude that if α and c are independent, namely

$$P(\alpha \mid c) = P_\alpha(\alpha), \tag{33}$$

then, the distribution of marginal productivity $P_{\mathrm{M}}(c_{\mathrm{M}})$ also obeys the power law with the *identical* Pareto index μ_{F} as that for the average productivity:

$$P_{\mathrm{M}}(c_{\mathrm{M}}) \propto c_{\mathrm{M}}^{-\mu_{\mathrm{F}}-1}. \tag{34}$$

In conclusion, *to the extent that α 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 Monte Carlo simulation result is shown in Fig. 6. Because this is the log-log plot the gradient of the straight region is equal to the Pareto index. We observe that the equality of the Pareto indices is clearly visible.

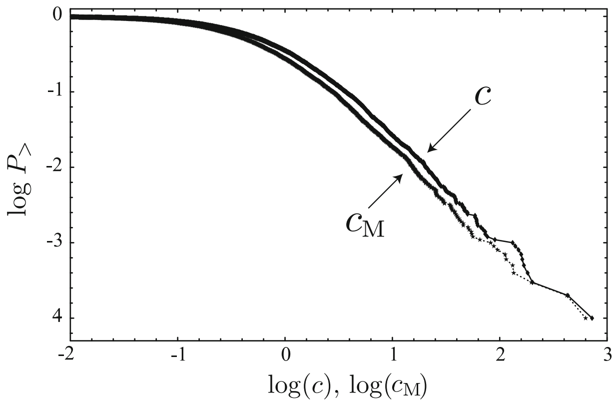

Fig. 6 The rank-size log-log plots of the simulation results. *Notes* The distribution for c is obtained from the Monte Carlo data generated from a $P(c)$ with $\mu_M = 1.5$. Data for c_M is then created with $P_\alpha(\alpha)$ uniformly distributed from $\alpha = 0.5$ to 1

The same analysis can be applied to the problem of measurement errors. In our analysis, L is the number of workers, and it would be theoretically 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 α in (27) as such a measurement error rather than a parameter of the Cobb-Douglas production function. Thus, to the extent that the average productivity c and measurement error α are independent, the distribution of "true" productivity obeys the power law with the same exponent as that for the measured productivity.

In summary, we can safely ignore the possible problems which might arise from the divergence between the marginal and average productivities, on the one hand, and the measurement errors in our empirical study of productivity dispersion, on the other.

3.3 Summary of empirical observations

The empirical observations made in Aoyama et al. (2009) can be summarized 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, namely across workers, firms, and industrial sectors.
II. The Pareto index or the power exponent decreases as the level of aggregation increases: $\mu_W > \mu_F > \mu_S$ (Fig. 5).

In the next section we extend the basic framework and present a theoretical model that explains these stylized facts.

4 Theory

The are several theoretical challenges before us. As we explained in Sect. 2, the assumption that the distribution of productivity across firms is uniform leads us to the exponential distribution of productivity across workers (19). Obviously, this model does not fit the empirical observations, and thus we need to explore the generic framework that explains the power law distribution for the productivity of firms. Furthermore, we need to extend the basic model of stochastic macro-equilibrium to explain two "stylized facts" described in the previous section, and in pursuit of this we will invoke the promising theoretical *superstatistics* framework explored by Aoyama et al. (2009).

4.1 The productivity dispersion across firms: a model of a jump Markov process

The standard economic analysis takes it for granted that all production factors enjoy the highest marginal productivity in equilibrium. However, this is a wrong characterization of the economy, and the fact of the matter 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. Adapting Marsili and Zhang (1998)'s Markov model of city size, 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 without loss of generality we can assume is unity, with probability $w_+(c)\,dt$. We denote this as $w_+(c)$ because this probability w_+ depends on the level of c. Likewise, it decreases by a unit with probability $w_-(c)\,dt$. Thus, $w_+(c)$ and $w_-(c)$ are transition rates for the processes $c \to c+1$ and $c \to c-1$, 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 of productivity c at time t, $n(c,t)$, obeys the following master equation:

$$\begin{aligned} \frac{\partial n(c,t)}{\partial t} &= w_+(c-1)\,n(c-1,t) + w_-(c+1)\,n(c+1,t) \\ &\quad - w_+(c)\,n(c,t) - w_-(c)\,n(c,t) + p\,\delta_{c,1}\,, \end{aligned} \tag{35}$$

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

The total number of firms is given by

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

We define the aggregate productivity index C as:

$$C_t \equiv \sum_{c=1}^{\infty} c\, n(c, t). \tag{37}$$

It follows from Eq. (35) that

$$\frac{d}{dt} K_t = p - w_-(1)\, n(1, t), \tag{38}$$

$$\frac{d}{dt} C_t = p - \sum_{c=1}^{\infty} (w_-(c) - w_+(c))\, n(c, t). \tag{39}$$

Let us consider the steady state, which is the stationary solution of Eq. (35) such that $\partial n(c, t)/\partial t = 0$. The solution $n(c)$ can be readily obtained by the standard method, and by setting (38) equal to zero, we obtain a boundary condition that $w_-(1)\, n(1) = p$. Using this boundary condition, we can easily show

$$n(c) = n(1) \prod_{k=1}^{c-1} \frac{w_+(c-k)}{w_-(c-k+1)}. \tag{40}$$

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 firm's 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 α is greater than 1. Under this assumption, the stationary solution (40) becomes

$$n(c) = \frac{n(1)}{1 - n(1)/C_{(\alpha)}} \frac{(1 - n(1)/C_{(\alpha)})^c}{c^{\alpha}} \simeq c^{-\alpha}\, e^{-c/c^*}. \tag{41}$$

where

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

We use the relation $a_+/a_- = 1 - n(1)/C_{(\alpha)}$, which follows from Eqs. (38) and (39). The approximation in Eq. (41) follows from $n(1)/C_{(\alpha)} \ll 1$, and the exponential cut-off works as c approaches to c^*. However, the value of c^* is practically large, and 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 μ for the empirical distribution presented in Sect. 3 is related to α simply by

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

The present model can be understood easily with the help of an analogy with 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. These rates are the instantaneous probabilities that the population of a city with the current population c either increases or decreases by 1. They are, therefore, the entry and exit rates of *one person* $\times$ population c, respectively. A drifter thus 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 Eq. (35). 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. In that case $w_+(c)$ and $w_-(c)$ become linear functions of c, namely a_+c and a_-c. However, even in population dynamics one might assume that the entry rate of a person into a large city is higher than its counterpart in a smaller one because of better job opportunities or the social attractiveness of such places, as encapsulated in the words of the song,"bright lights, big city". The same may hold for exit and death rates, because of congestion or epidemics.

It turns 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 in which that particular unit happens to be located; to be concrete, 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 Ijiri and Simon 1975; Sutton 1997; Saichev et al. 2009).

There is also a technical reason why we may expect α to be larger than 2. From Eq. (41), we can write the total number of firms and the aggregate productivity index as

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

$$C = \frac{n(1)}{\Gamma(\alpha - 1)} \int_0^\infty \frac{t^{\alpha-2}}{e^t - 1 + n(1)/C_{(\alpha)}} dt \ , \tag{43}$$

where $\Gamma(z)$ is the gamma function defined by $\Gamma(z) = \int_0^\infty t^{z-1} e^t dt$. In the limit, since $n(1)/C_{(\alpha)}$ goes to zero, the integral in Eq. (42) is finite for $\alpha > 1$, while the integral in Eq. (43) tends to be arbitrarily large for $1 < \alpha < 2$. Therefore, it is reasonable to assume that $\alpha \geq 2$, so that both Eqs. (42) and (43) are finite.

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 actually observed. Since economists are prone to take changes in productivity as "technical progress" they tend to focus on R & D investment. However, if productivity growth is always technical progress, its decrease must be "technical regress," the very existence of which one might question. At the level of firms an important source of productivity change is actually a sectoral shift of demand, that is to say when demand for product A increases productivity at the firm producing A increases, and vice versa. Indeed, Fay and Medoff (1985) documented such changes in a firm's labor productivity by way of changes in the rate of labor hoarding. Of course, stochastic productivity changes, as described in our Markov model, certainly include technical progress, particularly in the case of an increase, but at the same time they represent *allocative demand disturbances*, the importance of which Davis et al. (1996) have so persuasively demonstrated in their book *Job Creation and Destruction*. Empirical observation of productivity dispersion and our present analysis suggest that the probability of a unit allocative demand disturbance depends on the current size of the firm or the sector.

4.2 The stochastic macro-equilibrium revisited

The basic framework explained in Sect. 2 presumes that productivity dispersion across firms is uniform, whereas in fact it obeys the power law. Having clarified the generic origin of the power law distribution for c we can now turn our attention to the productivity dispersion across workers, and extend the theory of the stochastic macro-equilibrium explained in Sect. 2 under the assumption that $P_>^{(\mathrm{F})}(c)$ is the power-distribution.

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

$$\sum_{k=1}^{K} f_k \longrightarrow \int_0^\infty P^{(\mathrm{F})}(c) f(c) dc, \tag{44}$$

where the function $f(c)$ is a continuous version of the series f_k, so that $f(c_k) = K f_k$, as is evident from the equality of both sides in the case $P^{(\mathrm{F})}(c) = (1/K)\sum_{k=1}^{K} \delta(c - c_k)$. In the continuous model, Eqs. (5), (8) and (9) read

$$S = -\int_0^\infty P^{(\mathrm{F})}(c) p(c) \ln p(c) dc, \tag{45}$$

$$\int_0^\infty P^{(\mathrm{F})}(c) p(c) dc = 1, \tag{46}$$

$$\int_0^\infty P^{(\mathrm{F})}(c)\, c\, p(c) dc = D, \tag{47}$$

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

$$P^{(\mathrm{W})}(c) \equiv p(c) P^{(\mathrm{F})}(c). \tag{48}$$

Using $P^{(\mathrm{W})}(c)$, we can rewrite Eqs. (45), (46) and (47) as follows:

$$S = -\int_0^\infty P^{(\mathrm{W})}(c) \ln \frac{P^{(\mathrm{W})}(c)}{P^{(\mathrm{F})}(c)}\, dc + [\text{const.}], \tag{49}$$

$$\int_0^\infty P^{(\mathrm{W})}(c)\, dc = 1, \tag{50}$$

$$\int_0^\infty c\, P^{(\mathrm{W})}(c)\, dc = D. \tag{51}$$

Here, [const.] in Eq. (49) is an irrelevant constant term.

We maximize S (Eq. (49)) under two constraints, (50) and (51), by means of a calculus of variation with respect to $P^{(\mathrm{W})}(c)$ to obtain

$$P^{(\mathrm{W})}(c) = \frac{1}{Z(\beta)} P^{(\mathrm{F})}(c)\, e^{-\beta c}. \tag{52}$$

Here, as in Sect. 2, β is the Lagrangian multiplier for the aggregate demand constraint, (47), and the partition function $Z(\beta)$ is given by

$$Z(\beta) = \int_0^\infty P^{(\mathrm{F})}(c)\, e^{-\beta c} dc. \tag{53}$$

It is easy to see that constraint (50) is satisfied. Constraint (51) now reads

$$D = \frac{1}{Z(\beta)} \int_0^\infty c\, P^{(\mathrm{F})}(c)\, e^{-\beta c} dc. \tag{54}$$

This equation is equivalent to:

$$D = -\frac{d}{d\beta} \ln Z(\beta). \tag{55}$$

It is straightforward to see that Eqs. (52) and (53) are the counterparts of Eqs. (14) and (15), respectively, under the assumption that the productivity dispersion across firms is not uniform, but is $P^{(\mathrm{F})}(c)$.

Because the productivity distribution across firms, $P^{(\mathrm{F})}(c)$, obeys the Pareto law the relation between the aggregate demand D and β is not so simple as shown in Eq. (18), but is, in general, quite complicated. However, we can prove that the power exponent β is a decreasing function of the aggregate demand D. Thus, the fundamental proposition that when the aggregate demand is high, production factors are mobilized

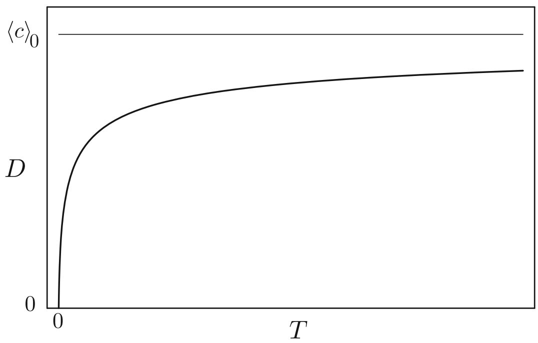

Fig. 7 The relation between the aggregate demand D and temperature T $(= 1/\beta)$. *Notes* For a productivity distribution of firms with $\mu_{\mathrm{F}} = 1.5$

to firms and sectors with high productivity (Fig. 1), holds true in the extended model as well, the proof of which can be found in Aoyama et al. (2009). An example of the relation between D and $T = 1/\beta$ is given in Fig. 7.

We can explore productivity dispersion in this extended model. The productivity dispersion across workers, $P^{(\mathrm{W})}(c)$, relates to that across firms, $P^{(\mathrm{F})}(c)$, by way of Eq. (52), which obeys the power law. Figure 8 shows examples of the productivity distributions across workers for several different values of D, given $P^{(\mathrm{F})}(c)$. The solid curve represents firms with $\mu_{\mathrm{F}} = 1.5$, and the dashed lines show the corresponding distributions for workers with different values of D. Due to the exponential factor $e^{-\beta c}$ all have strong suppressions for large values of c, which is 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 (52) with the observed power law is to assume an extremely small value for β so that the Boltzmann factor $e^{-\beta c}$ becomes close to 1, and does not suppress $P^{(\mathrm{W})}(c)$. However, this method fails because, in this case, the Pareto index μ_{W} of the worker's productivity distribution must be *equal to* that of the firm's distribution, μ_{F}. Thus, it is inconsistent with the empirical observation that $\mu_{\mathrm{W}} > \mu_{\mathrm{F}}$ (Fig. 5). Comparing Figs. 4 and 8, we conclude that the extended model still fails to explain the observations, and consequently we must seek a new theoretical framework.

4.3 Worker's productivity dispersion under fluctuating aggregate demand: superstatistics

The theoretical framework explained so far implicitly assumes that the aggregate demand D is constant. Plainly, this is an oversimplification; in fact, D fluctuates. A macro-system experiencing fluctuations in an external environment can be analyzed with the help of *superstatistics* or the "statistics of statistics", as employed in statistical physics (Beck and Cohen 2003).[5] In this section, we summarize the results of

[5] There are several model cases where superstatistics has been applied successfully. For example, the Brownian motion of a particle passing through a changing environment provides a good analogy to our case (Beck 2006).

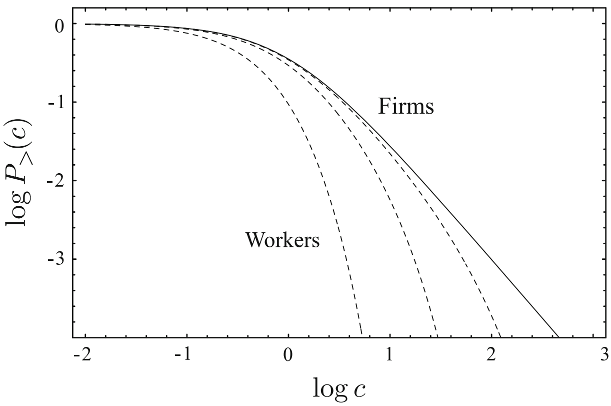

Fig. 8 Productivity distributions across workers and firms. *Notes* The *solid curve* represents firms with $\mu_F = 1.5$, while the *dashed curves* show workers with $\beta = 0.01, 0.1, 1$, respectively. When β is small enough the distribution across workers is close to that of firms. As β increases, the distribution across workers is suppressed for large values of c due to the exponential factor $e^{-\beta c}$

our analysis of superstatistics, and the reader is referred to Aoyama et al. (2009) for technical details.

In the theory of superstatistics, the system passes 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; seen from a global perspective it is out of equilibrium. In order to analyze such a system superstatistics averages over the Boltzmann factors, and depending on the weight function used for this averaging it can yield various distributions, including the power law (Touchette and Beck 2005).

The GDP of a particular year is certainly a scalar constant when the year is over, but it actually fluctuates daily, though those fluctuations cannot be practically measured. Accordingly, the macro environment surrounding firms changes almost continuously. Differences by region, industry, and sector can also be handled with a superstatistical approach. For example, when the aggregate demand D changes, the new stochastic process conditioned by the new value of 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 other equilibrium distribution.

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

$$B(c) = \int_0^\infty f(\beta)\, e^{-\beta c} d\beta. \tag{56}$$

Here, the weight factor $f(\beta)$ represents the changing macroeconomic environment. Note that because β 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, (52), is now replaced by the following:

$$P^{(W)}(c) = \frac{1}{Z_B} P^{(F)}(c)B(c). \tag{57}$$

Here, the partition function Z_B is also redefined as

$$Z_B = \int_0^\infty P^{(F)}(c)B(c)dc. \tag{58}$$

Aoyama et al. (2009) demonstrates that in this framework, $P^{(W)}(c)$ in Eq. (57) obeys the power law for high productivity c with the Pareto index μ_W, which is larger than μ_F as we actually observe. The assumption made to establish this result is the following behavior of the pdf $f(\beta)$ for $\beta \to 0$,

$$f(\beta) \propto \beta^{-\gamma} \quad (\gamma < 1). \tag{59}$$

The constraint for the parameter γ is drawn from the convergence of the integration in Eq. (56).

Because β is related to D by way of Eq. (55), the pdf $f_\beta(\beta)$ of β is related to pdf $f_D(D)$ of the (fluctuating) D as follows:

$$f_\beta(\beta)d\beta = f_D(D)dD. \tag{60}$$

It can be shown that given a pdf of β (59), the pdf of D is as follows:

$$f_D(D) \propto (\langle c \rangle_0 - D)^{-\delta}. \tag{61}$$

In this case $\langle c \rangle_0$ is the maximum level of aggregate demand, and the exponent δ is defined as follows:[6]

$$\gamma - 1 = \begin{cases} \delta - 1 & \text{for } 2 < \mu_F; \\ (\mu_F - 1)(\delta - 1) & \text{for } 1 < \mu_F < 2. \end{cases} \tag{62}$$

Equation (61) means that changes in the aggregate demand D follow the power law. Indeed, Gabaix (2005), demonstrates that idiosyncratic shocks to the top 100 firms explain a large fraction (one third) of aggregate volatility for the U.S. economy, which is a notable characteristics of power law. In any case, D is not constant, but fluctuates. Accordingly, the problem is now not a relation between the productivity dispersion across workers and the level of D, but rather the manner in which the distribution depends on the way in which D fluctuates.

[6] The parameter δ is constrained by $\delta < 1$ from the normalizability of the distribution of $f_D(D)$, which is consistent with the constraint $\gamma < 1$ and Eq. (61).

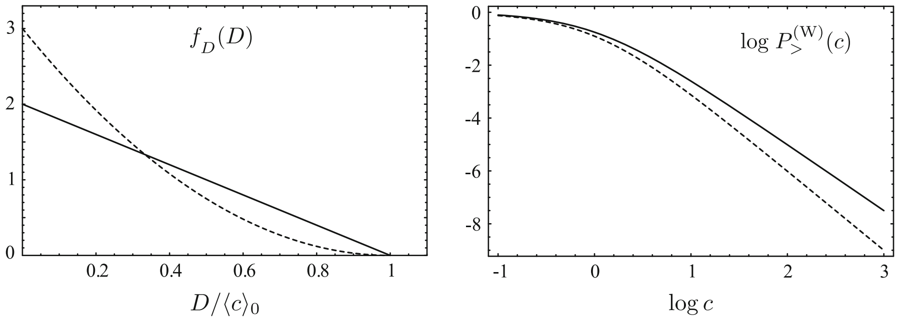

Fig. 9 The Distribution of the aggregate demand D, $f_D(D)$ (*left*) and the corresponding cumulative productivity distribution $P_>^{(\mathrm{W})}(c)$ (*right*). *Notes* The *solid curves* are for $\delta = -1$ whereas the *broken curves* are for $\delta = -2$. The productivity distribution of firms is chosen to have $\mu_{\mathrm{F}} = 1.5$

Figure 9 illustrates, by way of an example, the relation between the distribution of D, $f_D(D)$ near the upper bound $\langle c\rangle_0$ and the cumulative productivity distribution of workers $P_>^{(\mathrm{W})}(c)$. In the figure the solid curves correspond to the large value of δ, whereas the broken curves represent the small value. Figure 9 demonstrates that as the distribution of D becomes skewed toward large values of 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.

Thus, using the superstatistics framework, we can show not only that the distribution of productivity across workers obeys the power law in the high productivity region, but also that when the level of aggregate demand is high workers are mobilized towards high productivity firms. Furthermore, it can be shown that the Pareto index μ goes down as we go up from workers to firms, and from firms to industrial sectors;

$$\mu_{\mathrm{W}} > \mu_{\mathrm{F}} > \mu_{\mathrm{S}}. \tag{63}$$

This is exactly the first of the two stylized facts (Fig. 5) presented in Sect. 3. The reader interested in the details of the analysis is referred to Aoyama et al. (2009).

Finally, given the measured values of μ_{W} and μ_{F}, we can determine the value of δ: The result is shown in Fig. 10. Recall that δ is the power exponent of the distribution of aggregate demand D, or more precisely the deviation of D from its maximum level, $\langle c\rangle_0$ (see Eq. (61)). Therefore, low δ means a relatively low level of aggregate demand. In Fig. 10 we observe that the aggregate demand was high during the late 1980s, while beginning in the early 90s, it declined to lows in 2000–2001, and then turned upwards, which is broadly consistent with changes in the growth rate of the Japanese economy during that period.

5 Remaining problems

Productivity dispersion is very much more than an interesting example of the many curious distributions that econophysicists have found using large economic data (Stanley et al. 2004; Aoyama et al. 2010). Indeed, it is essentially related to the

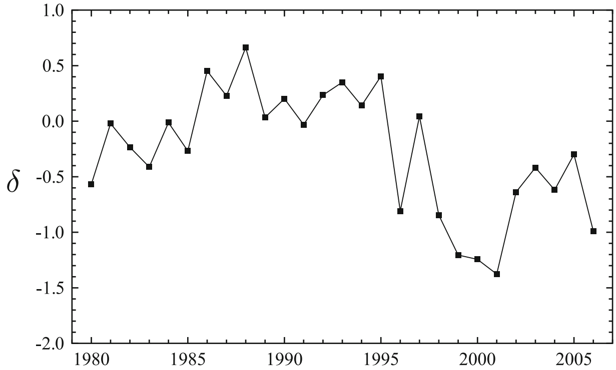

Fig. 10 The values of δ calculated based on the measured value of μ_{W} and μ_{F}. *Source* Aoyama et al. (2009)

concept of equilibrium in the neoclassical theory, which presumes 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, the concept of equilibrium would be contradicted. This is the definition of "full employment" of resources.

However, we have significant evidence suggesting that there is always productivity dispersion in an economy. As referred to in the Introduction, Mortensen (2003)'s analysis of wage dispersion argues that there is productivity dispersion across firms. Okun (1973) also argued that a part of the reason why we obtain his law is "the upgrading of workers into more-productive jobs in a high-pressure economy." Furthermore, in Sect. 3, we have summarized our own previous studies which demonstrate that the distribution of labor productivity obeys the Pareto distribution.

The fundamental principle of statistical physics 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. However, the standard theory indicates that it is an exponential distribution (the Maxwell–Boltzmann), the exponent of which depends inversely on the level of aggregate demand. This is in contradiction to the robust empirical finding that the distribution of labor productivity is a power distribution. We have shown that we can reconcile 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. The latter, namely, the superstatistical framework, which is still little known, seems very promising.

In what follows, we list remaining important problems which await further investigation. First, in Sect. 4 we explained that 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, namely that the higher the current level of productivity, the greater the probability of either a unit increase or decrease of productivity. Since there are some positive "size effects" this relation must be more than proportional.

The "size effects" in productivity (total factor productivity or TFP) growth have 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, the observed power distribution of productivity across firms implies that there is a significant probability of *decrease* in productivity in an economy. This, in turn, strongly suggests that productivity changes are caused not only by technical progress, but also by *the allocative disturbances in demand*. Incidentally, Davis et al. (1996) report that unlike job creation, job destruction for an industry is not systematically related to growth in TFP. Namely, job destruction occurs in high TFP growth industries as frequently as in low TFP growth industries (see the Table 3.7 on page 52). This fact also suggests the presence of significant demand reallocation. As Davis et al. (1996) rightly emphasize, 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. Contrary to the postulate of neoclassical economics, the real economy is not on the frontier of the production possibility set. The higher the level of aggregate demand, the closer the economy is to the frontier. The relation between the allocative disturbances and the distribution of productivity is a crucial problem awaiting further research.

Secondly, employing the superstatistical framework to explain the power law distribution of productivity across workers requires the assumption that the distribution of aggregate demand D obeys the power law. On the one hand, a finding made by Gabaix (2005) that idiosyncratic shocks to the top 100 firms explain one third of the aggregate volatility of the US economy strongly suggests the power law. On the other hand, Canning et al. (1998) report that the distribution of the growth rates of GDP is exponential. Superficially, the volatility of asset prices and the real economy appear to differ by nature (Aoki and Yoshikawa 2007, Chap. 10). Obviously, it is extremely important to explore the distribution of GDP in greater detail.

Thirdly, the standard principle of statistical physics suggests that whether the distribution is exponential or a power law, in any case the shares of low productivity workers of firms are greater than those of high productivity workers or firms. In contrast, the standard economic reasoning is that resources flow from low productivity sectors or firms to high productivity sectors or firms. In fact, neoclassical economics presumes that all workers enjoy the highest (not the lowest!) marginal productivity in equilibrium. In the neoclassical paradigm, the standard approach to dynamic adjustment is a stability analysis such as Arrow et al. (1959). Behind the standard stability analysis, resources are reallocated from low productivity sectors to high productivity sectors. It suggests that the shares of high productivity workers (firms) are greater than those of low productivity workers (firms). Thus, the principle of statistical physics and that of neoclassical economics are in contradiction. Indeed, we have evidence that in some cases, the distribution of productivity is locally skewed toward the high productivity segment (see Fig. 11).

The distribution of productivity emerges from persistent disturbances and accompanying adjustments in the economy. The principle of maximum entropy puts more resources in low productivity sectors while economic motives mobilize resources to

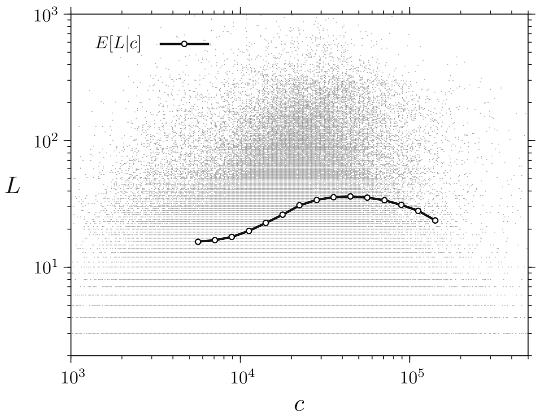

Fig. 11 The productivity c and labor L. *Note* The *gray dots* show firms in (c, L) plane. The *circles* show the expectation value of L for given c, $E[L|c]$, which increases as c increases in the middle section. For higher c, $E[L|c]$ declined with c, as expected from the superstatistics theory

high productivity sectors. Meanwhile the economy is incessantly subject to various disturbances. To build a unified theory to explain these phenomena requires further investigation.

References

Aoki M, Yoshikawa H (2007) Reconstructing macroeconomics—a perspective from statistical physics and combinatorial stochastic processes. Cambridge University Press, Cambridge

Aoyama H, Yoshikawa H, Iyetomi H, Fujiwara Y (2009) Labour productivity superstatistics. Prog Theor Phys Suppl 179:80–92

Aoyama H, Fujiwara Y, Ikeda Y, Iyetomi H, Souma W (2010) Econophysics and companies—statistical life and death in complex network. Cambridge University Press, Cambridge

Arrow K, Block H, Hurwicz L (1959) On the stability of the competitive equilibrium, II. Econometrica 27:82–109

Beck C (2006) Superstatistical Brownian motion. Prog Theor Phys Suppl 162:29–36

Beck C, Cohen EGD (2003) Superstatistics. Physica A 322:267

Canning D, Amaral L, Lee Y, Meyer M, Stanley H (1998) Scaling the volatility of GDP growth rates. Econ Lett 60(3):335–341

Champernowne D (1973) The distribution of income between persons. Cambridge University Press, Cambridge

Davis S, Haltiwanger J, Schuh S (1996) Job creation and destruction. MIT Press, Cambridge, MA

Delli Gatti D, Gaffeo E, Gallegati M, Giulioni G, Palestrini A (2008) Emergent macroeconomics. Springer, Berlin

Fay JA, Medoff JL (1985) Labor and output over the business cycle: some direct evidence. Am Econ Rev 75:63–655

Foley D (1994) A statistical equilibrium theory of markets. J Econ Theory 62(2):321–345

Gabaix Z (2005) The granular origins of aggregate fluctuations. MIT and NBER, DP

Ijiri Y, Simon HA (1975) Some distributions associated with Bose–Einstein statistics. Proc Natl Acad Sci USA 72(5):1654–1657

Ijiri Y, Simon HA (1977) Skew distributions and the sizes of business firms. North-Holland Pub., Co., Amsterdam

Ikeda Y, Souma W (2009) International comparison of labor productivity distribution for manufacturing and non-manufacturing firms. Prog Theor Phys Suppl 179:93–102

Keynes J (1936) The general theory of employment, interest, and money. Macmillan, London

Kydland F, Prescott E (1982) Time to build and aggregate fluctuations. Econometrica 50:1345–1370

Lucas R Jr (1972) Expectations and the neutrality of money. J Econ Theory 4:103–124

Lucas R Jr (1987) Models of business cycles. Basil Blackwell, New York

Marsili M, Zhang YC (1998) Interacting individuals leading to Zipf's law. Phys Rev Lett 80(12):2741–2744
Mortensen DT (2003) Wage dispersion. MIT Press, Cambridge
Nikkei Media Marketing, Inc. (2008) Nikkei NEEDS CD-ROM. http://www.nikkeimm.co.jp/english/index.html
Okun AM (1962) Potential GNP: its measurement and significance. In: Proceedings of the business and economic statistics section, American Statistical Association, pp 98–104
Okun AM (1973) Upward mobility in a high-pressure economy. Brookings Pap Econ Act 1:207–261
Phelps E (1970) Microeconomic foundations of employment and inflation theory. Norton, New York
Saichev A, Malevergne Y, Sornette D (2009) Theory of Zipf's Law and Beyond. Springer, Berlin
Solow RM (2000) Growth theory: an exposition. 2nd ed. Oxford University Press, New York
Stanley H, Gopikrishnan P, Plerou V (2004) Statistical physics and economic fluctuations. The complex dynamics of economic interaction: essays in economics and econophysics, p 109
Sutton J (1997) Gibrat's legacy. J Econ Lit 35:40–59
Tobin J (1972) Inflation and unemployment. Am Econ Rev 62:1–18
Touchette H, Beck C (2005) Asymptotics of superstatistics. Phys Rev E 71(1):16131
Yoshikawa H (2003) The role of demand in macroeconomics. Jpn Econ Rev 54:1–27