Stochastic Macro-equilibrium
and
Microfoundations for the Keynesian Economics*

Hiroshi Yoshikawa
Faculty of Economics, University of Tokyo
yoshikawa@e.u-tokyo.ac.jp

June 2012

Abstract

This paper begins with pointing out limitations of the standard labor search theory. It then presents an alternative concept of stochastic macro-equilibrium based on the principle of statistical physics. This concept of equilibrium is motivated by the presence of all kinds of unspecifiable micro shocks in the macroeconomy. They make the use of optimization exercises on representative agent assumptions dim. We present a model which mimics the empirically observed distribution of labor productivity. The distribution of productivity depends crucially on aggregate demand. When aggregate demand rises, more workers are employed by firms with higher productivity. At the same time, the unemployment rate declines. The result provides a proper micro-foundation for Keynes’ principle of effective demand.

Key words: Labor Search Theory, Microeconomic Foundations, Stochastic Macro-equilibrium, Effective Demand

JRL: D39, E10, J64

* This work is supported by the Program for Promoting Methodological Innovation in Humanities and Social Sciences by Cross-Disciplinary Fusing of the Japan Society for the Promotion of Science. The simulation presented in the paper was carried out by Mr. Yoshiyuki Arata. The author is grateful to him for his excellent research assistance. He is also grateful to Professors Hideaki Aoyama, Jean-Philippe Bouchaud, Hiroshi Iyetomi, Bertrand Roehner, Seung-Gyu Sim, Robert M. Solow and seminar participants at École normale supérieure for their useful comments.
1. Introduction

The purpose of this paper is to present a new concept of stochastic macro-equilibrium, and explain that it provides solid micro-foundations for the Keynesian theory of effective demand.

The most influential concept of equilibrium in economics is the Walrasian general equilibrium as represented by Arrow and Debreu (1954). Though it is a grand concept, and well established in the profession, it cannot be more different from the real economy. Diamond (2011) states as follows:

“Arrow-Debreu theory does not contain a mechanism or process for an economy to achieve its equilibrium allocation. In the 1960s there was ongoing work to find a hypothetical process that would converge to this equilibrium, with a focus on equations for price adjustment based on excess demands and supplies at tentative prices (referred to as tâtonnement). It struck me that the wrong question was being asked. Rather than asking whether a process could be found that would converge to a standard competitive equilibrium, I chose to work on the question of finding the allocation to which a plausible process would converge. (Diamond, 2011; p.1047)”

Rather than adhering to the imaginary Walrasian general equilibrium, Diamond proposes a research strategy to seek “a plausible process” or a realistically dynamic description of the economy, and then explore the concept of equilibrium which corresponds to such a dynamic process. Diamond (1994) dubs it “real time economics”.

The most successful example of real time economics is equilibrium search theory surveyed by its pioneers Diamond (2011), Mortensen (2011), and Pissarides (2000, 2011). Search theory stems from realizing limitations of the Walrasian equilibrium. The standard general equilibrium abstracts itself altogether from the search and matching costs which are always present in the actual markets. By explicitly exploring search frictions, search theory has succeeded in shedding much light on the workings of labor market. What has been done is well known.

While acknowledging the achievement of search theory, this paper presents an alternative concept of stochastic equilibrium of the macroeconomy based on the basic method of statistical physics. Section 2 points out limitations of search theory. After a brief explanation of the basic idea of statistical physics in Section 3, Section 4 presents a model of stochastic macro-equilibrium. The model is broadly consistent with the empirically observed distribution of labor productivity. Section 5 then explains that the stochastic macro-equilibrium provides a proper micro-foundations for Keynes’ principle of effective demand. The equilibrium to which a plausible process would converge is nothing but the old Keynesian equilibrium (Tobin (1993))! The final section offers brief
concluding remarks.

2. Limitations of Search Theory

The search theory starts with the presence of various frictions and accompanying matching costs in market transactions. Once we recognize these problems, we are led to heterogeneity of economic agents and multiple outcomes in equilibrium. In the simplest retail market, for example, with search cost, it would be possible to obtain high and low (more generally multiple) prices for the same good or service in equilibrium. This break with the law of one price is certainly a big step toward reality. Frictions and matching costs are particularly significant in labor market. And the analysis of labor market has direst implications for macroeconomics. In what follows, we will discuss labor search theory.

In search equilibrium, potentially similar workers and firms experience different economic outcomes. For example, some workers are employed while others are unemployed. In this way, search theory well recognizes heterogeneity of workers and firms. Despite this recognition, when it comes to model optimizing behavior of economic agent such as worker, it, in effect, presumes the representative agent. For example, it is explicitly stated in a classical paper by Diamond (1982) that “all individuals are assumed to be alike.” As a consequence, what is called the “critical production cost” which is essentially the reservation wage is the same for all the workers looking for employment in the model. The representative worker assumption is made in recent papers such as Mortensen (2010). Thus, in the Diamond – Mortensen – Pissarides search model of unemployment, the reservation wage $R$ is the same for all the workers.

On this assumption, the standard analysis typically goes as follows. In the equilibrium labor market, we must obtain

$$fu = s(1-u)$$  \hspace{1cm} (1)

where $u$ is the unemployment rate, and $s$ and $f$ are the separation rate and the job finding rate, respectively. Equation (1) makes sure the balance between in- and out-flows of the unemployment pool. If $\lambda$ is the offer arrival rate and $F(w)$ is the cumulative distribution function of wage offers, the job finding rate $f$ is

$$f = \lambda(1-F(R)).$$  \hspace{1cm} (2)

Equation (2) is the standard equation in the literature. However, it is important to note that this equation is a macro equation, and that the offer arrival rate $\lambda$ and the cumulative distribution function of wage offers $F$ make sense in macro equation only
on the assumption that they are common to all the workers. Otherwise, we cannot relate \( \lambda \) and \( F(R) \) to the job finding rate \( f \), a macro-variable. It is obvious that in reality \( \lambda \) and \( F(R) \) differ significantly across workers. It is actually difficult to imagine that workers of different educational attainments face the same probability distribution of wage offers. Education is only one element. Workers differ with respect to all sorts of qualifications, and, therefore, face different probability of wage offers.

From equations (1) and (2), we obtain

\[
u = \frac{s}{s + \lambda(1 - F(R))}.
\] (3)

Though this analysis may be a useful first step, plainly extremely strong assumptions are made. As we pointed out above, it is very doubtful if observed wage differences arise due only to differences in realized wage offers for workers with the same offer arrival rate and reservation wage. Besides, although wages are one of the most important elements in any job offer, workers care not only wages but other factors such as job quality, tenure, and location. Preferences for these other factors which define a job offer certainly differ widely across workers. Preference for job of each worker cannot be represented by a stochastic variable, and a common probability distribution is hard to define.

The similar problems arise for models in which the behavior of firm is explicitly analyzed. In the first place, a firm does not usually employ workers with the same quality, but rather different kinds of workers, say white-collar and blue-collar workers or full-time and part-time workers. Their wages and separation rates differ, of course. These problems are assumed away in the standard search models. In summary, standard models are built on extremely unrealistic assumptions such as identical workers, risk neutral economic agents, and the stationary distribution of productivity.

The point is not that we must explicitly introduce all these complexities characterizing labor market into analytical model. It would simply make model intractable. Rather, we must fully recognize that it is absolutely impossible to trace the microeconomic behaviors of workers and firms in detail. From this viewpoint, the merits of doing sophisticated optimization exercises based on representative agent assumptions are dim. The following argument made by Arrow (1986) applies to macroeconomics in general, but is particularly relevant to labor market.

“In the aggregate, the hypothesis of rational behavior has in general no implications; that is, for any set of aggregate excess demand functions, there is a choice of preference maps and initial endowments, one for each individual in the economy, whose maximization implies the given aggregate excess demand functions.
The implications of the last two remarks are in contradiction to the very large bodies of empirical and theoretical research, which draw powerful implications from utility maximization for, respectively, the behavior of individuals, most especially in the field of labor supply, and the performance of the macoconomy based on “new classical” or “rational expectations” models. In both domains, this power is obtained by adding strong supplementary assumptions to the general model of rationality. Most prevalent of all is the assumption that all individuals have the same utility function (or at least that they differ only in broad categories based on observable magnitudes, such as family size). But if agents are different in unspecifiable ways, then remark 3 above shows that very little, if any, inferences can be made. (Arrow, 1986; pp.S388-S389)

In labor market, microeconomic shocks are indeed unspecifiable. For example, a worker may suffer from illness. This amounts to shocks to his utility function on one hand, and to his resource constraint on the other. His preference for job offer including the reservation wage, the location, and working hours necessarily changes. Note that these micro shocks are not to cancel out each other in the nature of the case. Nor we are assured that the standard central limit theorem holds true so that we can safely focus on the means; see Aoki and Yoshikawa (2012) for non-self-averaging.

Thus, for the purpose of the analysis of the macoconomy, sophisticated optimization exercises based on representative agent assumptions do not make much sense (Aoki and Yoshikawa (2007)). This is actually partly recognized by search theorists themselves. The recognition has led the introduction of the “matching function” into the analysis. The matching function relates the rate of meetings of job seekers and firms to the numbers of the unemployed and job vacancies. The idea behind it is explained by Pissarides (2011) as follows.

“Although there were many attempts to derive an equilibrium wage distribution for markets with search frictions, I took a different approach to labor market equilibrium, that could be better described by the term “matching”. The idea is that the job search underlying unemployment in the official definitions is not about looking for a good wage, but about looking for a good job match. Moreover, it is not only the worker who is concerned to find a good match, with the firm passively prepared to hire anyone who accepts its wage offer, but the firm is also as concerned with locating a good match before hiring someone.

The foundation for this idea is that each worker has many distinct features, which make her suitable for different kinds of jobs. Job requirements vary across firms too, and employers are not indifferent about the type of worker that they hire, whatever the wage. The process of matching workers to jobs takes time, irrespective of the wage offered by each job. A process whereby both workers and firms search for each other and jointly either accept or reject the match seemed to be closer to reality.

……. It allowed one to study equilibrium models that could incorporate real-world features like differences across workers and jobs, and differences in the institutional structure of labor markets.

The step from a theory of search based on the acceptance of a wage
offer to one based on a good match is small but has far-reaching implications for the modeling of the labor market. The reason is that in the case of searching for a good match we can bring in the matching function as a description of the choices available to the worker. The matching function captures many features of frictions in labor markets that are not made explicit. It is a black box, as Barbara Petrongolo and I called it in our 2001 survey, in the same sense that the production function is a black box of technology. (Pissarides, 2011; pp.1093-1094)

What job seeker is looking for is not simply a good wage, but a good job offer which cannot be uniquely defined but differs significantly across workers. It is simply unspecifiable. Pissarides recognizes such “real-world features” as differences across workers and jobs. Then, at the same time, he recognizes that we need a macro black box. The matching function is certainly a black box not explicitly derived from micro optimization exercises.

The matching function is based on a kind of common sense in that the number of job matching would increase when there are a greater number of both job seekers and vacancies. However, it still abstracts itself from an important aspect of reality. As Okun (1973) emphasizes, the problem of unemployment cannot be reduced only to numbers.

“The evidence presented above confirms that a high-pressure economy generates not only more jobs than does a slack economy, but also a different pattern of employment. It suggests that, in a weak labor market, a poor job is often the best job available, superior at least to the alternative of no job. A high-pressure economy provides people with a chance to climb ladders to better jobs.

The industry shifts are only one dimension of ladder climbing. Increased upward movements within firms and within industries, and greater geographical flows from lower-income to higher-income regions, are also likely to be significant. (Okun, 1973; pp.234-235)"

Dynamics of unemployment cannot be separated from qualities of jobs, or more specifically distribution of productivity on which we focus in the present paper. To explicitly consider these problems, we face greater complexity and, therefore, need a “greater macro black box” than the matching function. This is the motive for stochastic macro-equilibrium we explain in the next section.

3. Stochastic Macro-equilibrium —— The Basic Idea

Our vision of the macroeconomy is basically the same as standard search theory. Workers are always interested in better job opportunities, and occasionally change their jobs. Job opportunities facing workers are stochastic. Unemployment, a great challenge to any economy, deserves special attention in economic analysis, but actually workers
on the job also search. In fact, by far a major of workers change their jobs without experiencing any spell of unemployment. For example, a Japanese data reveals that in the year of 1981, 19.3 million workers changed their jobs without experiencing any spell of unemployment whereas only 1.6 million could reach their respective job sites out of the pool of unemployment. Though the standard matching function focuses on unemployment, plainly, unemployment is not at all a prerequisite for job turnover. On the job search obviously complicates micro optimization exercises on realistic assumptions. The importance of on the job search also means that job turnover depends crucially on the distribution of qualities of jobs in the economy. Though on the job search has been analyzed in search models, to date the analyses have been pursued only on unrealistic assumptions such as identical workers (Garibaldi and Moen (2010)).

While workers search for suitable jobs, firms also search for suitable workers. Firm’s job offer is, of course, conditional on its economic performance. The present analysis focuses on the firm’s labor productivity. The firm’s labor productivity increases thanks to capital accumulation and technical progress or innovations. However, those job sites with high productivity remain only potential unless firms face high enough demand for their products. In fact, because labor adjustment always takes time, when demand suddenly falls, productivity necessarily falls.

Though the standard literature tends to focus on wages, workers are interested not only in wages but also other job characteristics such as tenure, fringe benefits, and various work conditions. We assume that firms with higher productivity make more attractive job offers to workers. This assumption means that whenever possible, workers move to firms with higher productivity. However, for workers to move to firms with higher productivity, it is necessary that those firms must decide to fill the vacant job sites, and post enough number of vacancy signs. They post such vacancy signs only when they face high enough demand for their products. We will see in section 4 that in the economy as a whole, the most important factor determining the distribution of job qualities is aggregate demand. Before we turn to the model, we first explain the basic idea.

The question we ask is what is the distribution of employed workers across firms whose productivities differ. As we argued in the previous section, because microeconomic shocks to both workers and firms are so complex and unspecifiable, optimization exercises based on representative agent assumptions do not help us much. It is just impossible to pursue micro behavior of individual economic agent. This recognition is precisely the starting point of the fundamental method of statistical
physics. At first, one might think that allowing too large a dispersion of individual characteristics leaves so many degrees of freedom that almost anything can happen. However, it turns out that the methods of statistical physics provide us not only with qualitative results but also with quantitative predictions.

Foley (1994), in his seminal application of this approach to general equilibrium theory, advanced the idea of “statistical equilibrium theory of markets”. It works as follows.

“The market begins with agents defined by offer sets reflecting their information, technical possibilities, endowments, and preferences. The market distributes the agents over their sets to maximize the entropy of the market transaction distribution. This is the most disorderly, or in other words, the most decentralized, allocation possible given the offer sets of the agents. The aim of this theory is not to pick out a particular transaction in the sense of predicting what will happen to each agent, but to characterize the most likely statistical distribution of agents over feasible outcomes. (Foley, 1994; p.322)”

Recall that the Walrasian equilibrium determines the precise micro behaviors of all the agents. It is just as if we study gas by pursuing the exact behaviors of countless molecules. For the reasons explained in the previous section, it is actually impossible for us to describe micro behavior of each agent. Instead, we derive the most likely distribution of agents over feasible outcomes using the method of statistical physics. Following the lead of Foley (1994), Yoshikawa (2003) proposed the notion of stochastic macro-equilibrium based on the principle of statistical physics. The fundamental constraint in the model is aggregate demand, \( D \). The basic idea can be explained with the help of the simplest case.

Suppose that \( n_k \) workers belong to firms whose productivity is \( c_k (c_k < c_{k'}, \text{where } k < k') \). There are \( K \) levels of productivity in the economy. The total number of workers \( N \) is given.

\[
\sum_{k=1}^{K} n_k = N
\]  

(4)

A vector \( n = (n_1, n_2, \cdots, n_K) \) shows a particular allocation of workers across firms with different productivities. The combinatorial number \( W_n \) of obtaining this allocation, \( n \), is equal to that of throwing \( N \) balls to \( K \) different boxes. Therefore, we obtain

\[
W_n = \frac{N!}{\prod_{k=1}^{K} n_k!}
\]  

(5)

Because the number of all the possible ways to allocate \( N \) identical balls to \( K \) different boxes is \( K^N \), the probability that a particular allocation \( n = (n_1, n_2, \cdots, n_K) \) is obtained
is
\[
P_n = \frac{W_n}{K^N} = \frac{1}{K^N} \frac{N!}{\prod_{k=1}^{K} n_k!}.
\]  

(6)

It is the fundamental postulate of statistical physics that the state or the allocation \( n = (n_1, n_2, \ldots, n_K) \) which maximizes the probability \( P_n \) or (6) under macro-constraints is to be actually observed in the economy. The basic idea is similar to maximum likelihood in statistics/econometrics. It can be easily understood with the help of a simple example. When we throw a pair of dice, the possible sum ranges from two to twelve. Six is much more likely than twelve simply because the combinatorial number of the former is five whereas that of the latter is only one.

Maximizing \( P_n \) is equivalent to maximizing \( \ln P_n \). Applying the Stirling formula for large number \( x \)
\[
\ln x! \approx x \ln x - x,
\]  

(7)

we find that the maximization of \( \ln P_n \) is equivalent to that of \( S \).

\[
S = -\sum_{k=1}^{K} p_k \ln p_k \quad \text{where} \quad p_k = \frac{n_k}{N}.
\]  

(8)

\( S \) is the Shannon entropy, and captures the combinatorial aspect of the problem. As the above example of dice shows, the combinatorial consideration summarized in the entropy play a decisive role for the final outcome.

It is essential to understand that the present approach does not regard economic agents’ behaviors as random. Certainly, firms and workers maximize their profits and utilities. The economic behavior is purposeful, not random. Randomness underneath this analysis comes from the fact that both the objective functions of and constraints facing a large number of economic agents are constantly subject to unspecifiable micro shocks. We must recall that the number of households is of order \( 10^7 \), and the number of firms, \( 10^6 \). Therefore, there is nothing for outside observers, namely economists analyzing the macroeconomy but to regard a particular allocation under macro-constraints as equi-probable. Then, it is most likely that the allocation of workers which maximizes the probability \( P_n \) or (6) under macro-constraints is realized.

This method has been time and again successful in natural sciences when we analyze object comprising many micro elements. Economists might be still skeptical of the validity of the method in economics saying that inorganic atoms and molecules comprising gas are essentially different from optimizing economic agents. Every student of economics knows that behavior of dynamically optimizing economic agent
such as the Ramey consumer is described by the Euler equation for a problem of calculus of variation. On the surface, such a sophisticated economic behavior must look remote from “mechanical” movements of an inorganic particle which only satisfy the law of motion. However, every student of physics knows that the Newtonian law of motion is actually nothing but the Euler equation for a certain variational problem. It is called the principle of least action: see Chapter 19 of Feynmann (1964)’s Lectures on Physics, Vol. II. Therefore, behavior of dynamically optimizing economic agent and motions of inorganic particle are on a par to the extent that they both satisfy the Euler equations for respective variational problems. The method of statistical physics can be usefully applied not because motions of micro units are “mechanical,” but because object under investigation comprises many micro units individual movements of which we are unable to know.

To repeat, the present model doe not regard micro behavior of economic agent as random. If micro behaviors of workers and firms satisfy the Euler equations for respective variational problems, we cannot pin them down because they are subject to unspecifiable micro shocks as explained in section 2. The method of statistical physics perfectly fits this situation. We should expect that the allocation of workers which maximizes the probability $P_n$ or the Shannon entropy $S$ under the following two macro-constraints is realized in the real economy.

The first macro-constraint concerns the labor endowment, (4). The second macro-constraint concerns the effective demand.

$$\sum_{k=1}^{K} c_k n_k = D \quad (9)$$

Here, aggregate demand $D$ is assumed to be given. By maximizing $P_n$ or $S$ under two macro-constraints (4) and (9), we obtain the following exponential distribution (Aoki and Yoshikawa (2007, p.79-81)):

$$\frac{n_k}{N} = \frac{e^{\frac{Nc_k}{D}}}{\sum_{k=1}^{K} e^{\frac{Nc_k}{D}}} \quad (10)$$

This distribution depends crucially on the level of the aggregate demand $D$ which corresponds to temperature $T$ in physics. According to (10), the share of workers at firms whose productivity is high gets larger when the aggregate demand $D$ - to be precise, the aggregate demand relative to the endowment, namely $D/N$ - becomes greater, and vice versa (Figure 1). As Okun (1973) puts it, workers climb up the ladder of productivity when the aggregate demand is high, and vice versa.
4. The Model

The above analysis clarifies the relation between the distribution of workers at firms with different productivities and the level of aggregate demand. It turns out, however, that the basic model in the previous section is too simple to explain the empirically observed distribution of labor productivity.

Empirical Distribution of Productivity

Figures 2 (a) and (b) based on Iyetomi (2011) show the distributions of workers at different productivity levels for the Japanese manufacturing and non-manufacturing industries. The data used are the Nikkei Economic Electric Database (NEEDS, http://www.crd-office.net/CRD/english/index.html) and the Credit Risk Database (CRD, http://www.crd-office.net/CRD/english/index.html) which cover more than a million large and medium/small firms for 2004.

The “productivity” here is simply value added of the firm divided by the number of employed workers, that is, the average labor productivity. Theoretically, we should be interested in unobserved marginal productivity, not the average productivity. Besides, proper “labor input” must be in terms of work hour, or for that matter even in terms of work efficiency units rather than the number of workers. For these reasons, the average labor productivity shown in the figure is a crude measure of theoretically meaningful unobserved marginal productivity. However, Aoyama et. al. (2010; p.38-41) demonstrates that when the average productivity and measurement errors are independent, the distribution of true marginal productivity obeys the power law with the same exponent as that for the measured average productivity. In other words, distribution is robust with respect to measurement errors in the present case. Incidentally, we also note that the Pareto efficiency of the economy pertains to marginal products of production factors, not to total factor productivity (TFP) some economists focus on.

Figure 2 drawn on the double logarithm plane broadly shows that (1) the distribution of labor productivity is single-peaked, (2) in the low productivity (left) region, it is upward-sloping exponential whereas (3) in the high productivity (right) region, it obeys downward-sloping power-law (Aoyama et. al., (2010)). Ikeda and Souma (2009) find a similar distribution of productivity for the U.S. while DelliGatti et. al. (2008) find power-law tails of productivity distribution for France and Italy. In what follows, we present an extended model based on the principle of statistical physics for
explaining the broad shape of this empirically observed distribution.

The standard analysis in physics presumes that whenever possible, particles tend to achieve a low energy level. In contrast, workers strive for acquiring more attractive jobs possibly with higher wages. Attractive jobs are most likely offered by firms with high productivity. To the extent that we identify the level of productivity with the level of energy, our situation in economics is just opposite to that in physics. Iyetomi (2011) suggests that we handle this problem by introducing the “negative temperature”; On negative temperature, see Appendix E of Kittel and Kroemer (1982, 460-463). With the negative temperature, the downward-sloping distribution drawn in Figure 1 turns upward-sloping meaning that more particles (workers) tend to move toward the higher levels of energy (productivity). In physics, negative temperature does not make sense under normal conditions because the level of energy is unbounded; If temperature is negative, total energy diverges. However, in economics, negative temperature though it may sound odd, is nothing but an indicator of the level of aggregate demand, and as such it makes perfect sense because the level of productivity is bounded. The correspondence between aggregate demand $D$ and negative temperature $T$ is shown in Table 1.

With negative temperature, if possible workers always move to higher productivity firms, but deficient aggregate demand prevents them from climbing up to the firm with the highest level of productivity. Economics behind it is as follows. When aggregate demand is low, relatively a small number of firms with high productivity post job vacancy signs, and, therefore, given frictions, incomplete information and search costs, less workers can reach such job sites than otherwise. When aggregate demand is high, just the opposite holds true. The textbook neoclassical equilibrium in which all the workers enjoy the highest marginal product corresponds to the (negative) zero temperature, $0^-$ or the maximum level of aggregate demand. However, such a state remains imaginary; The zero temperature state is simply impossible to be realized.

By introducing the negative temperature, we can explain the broad pattern of the left-hand side of the distribution shown in Figures 2 (a) and (b), namely an upward-sloping exponential distribution (Iyetomi (2011)). However, we cannot explain the right-hand side downward-sloping power distribution. To explain it, we need to make an additional assumption that the number of potentially available jobs decreases as the level of productivity rises.

**Dynamics of Potential Job Creation/Destruction**
Potential jobs are created by firms by accumulating capital and/or introducing new technologies, particularly new products. At the same time, they are destroyed by firms’ losing demand for their products permanently. Schumpeterian innovations by way of creative destruction raise the levels of some potential jobs, but at the same time lower the levels of others. In this way, the number of potential jobs with a particular level of productivity keeps changing. They remain potential because firms do not attempt to fill all the job sites with workers. To fill them, firms either keep the existing workers on the job or post job vacancy signs, but it is an economic decision, and depends crucially on the economic conditions facing the firms. The number of potential job sites, therefore, is not exactly equal to, but rather imposes a ceiling on the sum of the number of filled job sites, namely employment and the number of job openings or vacancy signs.

The statistical theory we will discuss later explains how employment is determined. In this sub-section, we first consider dynamics of potential job creation and destruction. Causes of creation/destruction of potential job sites are micro-shocks, and as explained in the previous section, unspecifiable. The best way to describe them is Markov model. Good examples in economics are Champernowne (1950) on income distribution, and Ijiri and Simon (1979) on size distribution of firms. Here, we adapt the model of Marsili and Zhang (1998) to our own purpose. The goal of the analysis is to derive a power-law distribution such as the one for the tail of the empirically observed distribution of labor productivity.

Suppose that there are $f_j$ ”potential” jobs with productivity, $c_j$ ($c_j < c_j'$, when $j < j'$). In small time interval $dt$, the level of productivity $c$ of a potential job site increases with probability $w_+(c) dt$ by a small amount which we can assume is unity without loss of generality. We denote this probability as $w_+(c)$ because $w_+$ depends on the level of $c$. Likewise, it decreases by a unit with productivity $w_-(c) dt$. Thus, $w_+(c)$ and $w_-(c)$ are transition rates for the processes $c \to c + 1$ and $c \to c - 1$, respectively. As noted above, the productivity of job site changes for many reasons. It may reflect technical progress or innovations. Given significant costs of adjusting labor, productivity also changes when demand for firm’s products change. For example, when demand for firm’s products falls, labor productivity necessarily declines. The decline of productivity in this case reflects labor hoarding (Fay and Medoff (1985)). These changes are captured by transition rates, $w_+(c)$ and $w_-(c)$ in the model.
We also assume that a new job site is born with a unit of productivity with probability \( p \, dt \). On the other hand, a job site with productivity \( c = 1 \) will cease to exist if \( c \) falls to zero. Thus the probability of exit is \( w_-(c = 1) \, dt \). A set of the transition rates and the entry probability specifies a jump Markov process.

Given this Markov model, the evolution of the average number of job site of productivity \( c \) at time \( t \), \( f(c, t) \), obeys the following master equation:

\[
\frac{\partial f(c, t)}{\partial t} = w_+(c - 1) f(c - 1, t) + w_-(c + 1) f(c + 1, t) - w_+(c) f(c, t) - w_-(c) f(c, t) + p \delta_{c,1},
\]

(11)

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

We consider the steady state or the stationary solution of equation (11) such that \( \frac{\partial f(c, t)}{\partial t} = 0 \). The solution \( f(c) \) can be readily obtained by using the boundary condition that \( w_-(1) \, f(1) = p \) :

\[
f(c) = f(1) \prod_{k=1}^{c-1} \frac{w_+(c-k)}{w_+(c-k+1)}.
\]

(12)

Here, 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 current level of productivity of job site. Specifically, the higher the current level of productivity, the larger a chance of 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 (12) becomes

\[
f(c) = \frac{f(1)}{1 - f(1)/C(\alpha)} \left( \frac{1 - f(1)/C(\alpha)}{c^\alpha} \right)^c \approx c^{-\alpha} e^{-c/c^*}.
\]

(13)

where

\[
c^* = (f(1)/C(\alpha))^{-1} \quad \text{and} \quad C(\alpha) = \sum_{c=1}^\infty c^\alpha f(c).
\]

We use the relation \( a_+ / a_- = 1 - f(1)/C(\alpha) \). The approximation in equation (13)
follows from $n(1)/C_1 << 1$, and the exponential cut-off works as $c$ approaches to $c^\star$. However, the value of $c^\star$ 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. Therefore, under the reasonable assumption, we obtain power law distribution for job sites, $f_j$.

The above model can be understood easily with the help of an analogy with the formation of cities. Imagine that $f(c, t)$ 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 the population $c$ of the city, respectively. In addition, a drifter forms his own one-person city with the instantaneous probability $p$. In this model, the dynamics of $f(c, t)$ or the average number of cities with population $c$, is given by equation (11). 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 true for exit and death rates because of congestion or epidemics.

In turns out that in the dynamics of productivity of job site, both the “entry” and “exit” rates of an existing “productivity job site” are increasing functions of $c$, namely the level of productivity in which that particular job site happens to be located; to be concrete, $a_+c$ and $a_-c$. Thus, $w_+(c)$ becomes $a_+c$ times $c$ which is equal to $a_+c^2$. Likewise, we obtain $w_-(c) = a_-c^2$. This is the case of the so-called Zipf law (Ijiri and Simon (1975)). Thus, under the reasonable assumption that the probability of a unit change in productivity is an increasing function of its current level $c$, we obtain power law distribution for job sites $f_j$.

Economists are prone to presume that changes in productivity are caused by supply-side factors such as technical progress and entry/exit of firm alone. However, an
important source of productivity change is actually a sectoral shift of demand. Indeed, Fay and Medoff (1985) documented such changes in firm’s labor productivity by way of changes in the rate of labor hoarding. 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. By their careful study of industry-level productivity and worker flows, Davis et al. (1996) found that while job creation is higher in industries with high total factor productivity growth, job destruction for an industry is not systematically related to productivity growth; job destruction is actually highest in the industries in the top productivity growth quintile (Table 3.7, Davis et. al. (1996, P.52)). This finding suggests the importance of negative demand shocks for job destruction.

**Distribution of Productivity**

Distribution of potential job sites with high productivity obeys downward-sloping power law. However, the determination of employment by firms with various levels of productivity is another matter. To fill potential job sites with workers is the firm’s economic decision. For this, the firm can either keep the existing workers on the job or post vacancy signs toward successful job matching. Such actions of the firms and job search of workers are not random but purposeful. However, micro shocks affecting firms and workers are just unspecifiable. Then, how are workers actually employed at firms with various levels of productivity? This is the problem we consider in what follows.

The number of workers working at the firms with productivity, \( c_j \), namely \( n_j \) is
\[
   n_j \in \{0, 1, \ldots, f_j\}, \quad (j = 1, 2, \ldots K). \tag{14}
\]
Here, \( f_j \) is the number of potential jobs with productivity \( c_j \), and puts a ceiling on \( n_j \): Garibaldi and Scalas (2010) suggest that we study the problem by Markov model with similar constraints. Here, in the high productivity region, \( f_j \) obeys power-law distribution as we have analyzed above.

The total number of employed workers is simply the sum of \( n_j \):
\[
   N = \sum_{j=1}^{K} n_j. \tag{15}
\]
In the basic model explained in section 3, the total number of employed workers, \( N \) is exogenously given (equation (4)). In the present model, \( N \) is assumed to be variable.
$N$ is smaller than the exogenously given total number of workers or labor force, $L$ ($N < L$). The difference between $L$ and $N$ is the number of the unemployed, $U$:

$$U = L - N.$$  \hfill (16)

As in the basic model, the total output is constrained by the aggregate demand, $D$ (Equation (9)). However, in the basic model, $D$ is literally given. Accordingly, total output is also given by equation (9), and is constant. In the present model, we more realistically assume that following fluctuations of aggregate demand, total output $Y$ also fluctuates. Specifically, $Y$ defined by

$$Y = \sum_{k=1}^{K} c_k n_k$$  \hfill (17)

is determined in such a way that $Y$ is on average equal to constant $D$. That is, we have

$$<Y> = D$$  \hfill (18)

where $<Y>$ is the average of $Y$. Aggregate demand constrains total output in the sense of its expected value. Under this assumption, the probability that the level of total output is $Y$ turns out to be exponential; The density function $g(Y)$ is

$$g(Y) = \frac{e^{-\beta Y}}{\sum_{i} e^{-\beta Y}}$$  \hfill (19)

This result is obtained by the method of Gibbs’ canonical ensemble. Gibbs established the statistical mechanics by introducing the concept of “canonical ensemble” which is a collection of macro states, $Y$ in our present case. Suppose that there are $K$ possible levels of $Y$ denoted by $Y_1, \cdots Y_K$. For the moment, we reinterpret $n_k$ as the number of cases where $Y$ takes the value $Y_k$ ($k = 1, \cdots, K$). The sum of $n_k$, $N$ is given. Then, $n_k$ satisfies equation (4). We assume that the average of $Y$ is equal to constant $D$.

$$\sum_{k=1}^{K} Y_k \left( \frac{n_k}{N} \right) = D$$  \hfill (20)

Replacing $c_k$ by $Y_k$, we observe that equation (20) is equivalent to equation (9). Thus, we can apply the exactly same entropy maximization as we did in the basic model in section 3. It leads us to equation (19). In (19), $\beta$ is the Lagrange multiplier for constraint (20) in the maximization of entropy (8), and is equal to the inverse of the negative temperature: $\beta = 1/T$. 
Obviously, $Y$ constrained by aggregate demand $D$ affects the distribution of productivity, $n_k$ (equation (17)). In the present model, the number of employed workers $N$ is not constant, but changes causing changes in unemployment. Besides, the number of potential job sites with high productivity, $f_j$ is constrained in number, and follows power-law. In this model, we seek the state which maximizes the probability $P_n$ or equation (6).

Before we proceed, it is useful to explain partition function because it is rarely used in economics, but we will use it intensively in the subsequent analysis. When a stochastic variable $X$ is exponentially distributed, that is, its density function $g(Y)$ is given by equation (19), partition function $Z$ is defined as

$$Z = \sum_ie^{-\beta i}.$$  

This function is extremely useful as a moment generating function. For example, the first moment or the average of $Y$ can be simply found by differentiating $\text{log } Z$ with respect to $\beta$.

$$-\frac{d}{d\beta} \text{log } Z = -\frac{d}{d\beta} \text{log}(\sum_ie^{-\beta i}) = -\frac{\sum(-Y_i)e^{-\beta_i}}{\sum e^{-\beta_i}}$$

$$= \sum Y_i(\frac{e^{-\beta_i}}{\sum e^{-\beta_i}}) = \sum Y_i g(Y_i) = E(Y_i)$$  

In our case, $\beta$ in equation (19) is the inverse of temperature $T$, and is, therefore, negative. It is an indicator of the level of aggregate demand, $D$. The correspondence between $D$ and $T$ or $\beta$ is shown in Table 1. The higher the level of aggregate demand, the more likely higher level of aggregate output is realized.

As in the basic model, we want to find the state which maximizes the probability, $P_n$ or equation (6). We have two macro-constraints, equations (15) and (17). The total number of workers employed $N$ is not constant but variable. The aggregate output $Y$ is also not constant but obeys the exponential distribution, namely equation (19).

Because the level of total output depends on the total number of employed workers $N$, we denote $Y_i$ as $Y_i(N)$. Then, the canonical partition function $Z_N$ can be written as

$$Z_N = \sum_i e^{-\beta_i(N)}.$$  

(23)
Using equation (17), we can rewrite this partition function as follows:

\[ Z_N = \sum_{\{n_i\}} \exp(-\beta \sum_{i=1}^{K} c_i n_i). \]  

(24)

Unfortunately, it is generally difficult to carry out the summation with respect to \( \{n_i\} \) under constraint (15), namely \( N = \sum n_i \). Rather than taking \( N \) as given, we better allow \( N \) to be variable as we do here, and consider the grand canonical partition function \( \Phi \).

The grand canonical partition function is defined as

\[ \Phi = \sum_{N=0}^{\infty} z^N Z_N \]  

(25)

where

\[ z = e^{\beta \mu}. \]  

(26)

In physics, the parameter \( \mu \) is called the chemical potential, and measures the marginal contribution in terms of energy of an additional particle to the system under investigation. In the present model, \( N \) is the number of employed workers, and, therefore, \( \mu \) measures the marginal product of a worker who newly acquired job out of the pool of unemployment. Put another way, \( \mu \) is equal to the reservation wage, or more generally the “reservation job offer” of the unemployed. When \( \mu \) is high, the unemployed worker is choosy, and vice versa.

Substituting equation (24) into equation (25), the grand canonical partition function \( \Phi \) becomes as follows:

\[ \Phi = \sum_{N=0}^{\infty} z^N \sum_{n_j} \exp\{-\beta \sum_{j} n_j c_j\} \text{ where } z = e^{\beta \mu} \]  

(27)

Using the definitions of \( z \), (26), and also \( N \), (15), we have

\[ \Phi = \sum_{N=0}^{\infty} e^{\beta \mu (n_1 + \cdots + n_K)} \sum_{n_j} \exp\{-\beta \sum_{j} n_j c_j\} = \prod_{j=1}^{K} \sum_{n_j} \exp[\beta(\mu - c_j)n_j]. \]  

(28)

Because there is ceiling \( f_j \) for \( n_j \) (constraint (14)), (28) can be rewritten as follows:

\[ \Phi = \prod_{j=1}^{K} \left[1 + e^{\beta(\mu - c_f)} + \cdots + e^{f_j \beta(\mu - c_f)}\right] = \prod_{j=1}^{K} \left[1 - e^{(f_j + 1)\beta(\mu - c_f)}}{1 - e^{\beta(\mu - c_f)}}\right] \]  

(29)

With this grand canonical partition function \( \Phi \), we can easily obtain the expected value of the total number of employed workers \( N \), \( <N> \) by differentiating \( \log \Phi \) with respect to \( \mu \) which corresponds to the reservation wage of the unemployed worker.
This can be seen by differentiating (25) and noting the definition of \( z \), (26).

\[
\frac{1}{\beta} \left[ \frac{\partial}{\partial \mu} \log \Phi \right] = \frac{1}{\beta} \left[ \frac{\partial}{\partial \mu} \log \left( \sum_{N=0}^{\infty} e^{\beta N} Z_N \right) \right] = \frac{1}{\beta} \left[ \frac{\beta \sum_{N=0}^{\infty} Ne^{\beta N} Z_N}{\sum_{N=0}^{\infty} e^{\beta N} Z_N} \right] =< N > .
\] (30)

In the present case, \( \Phi \) is actually given by equation (29). Therefore, we can find \(< N >\) as follows.

\[
\langle N \rangle = \frac{1}{\beta} \left[ \frac{\partial}{\partial \mu} \log \Phi \right] = \frac{1}{\beta} \sum_{j=1}^{K} \left\{ \log(1 - e^{(\beta j + 1) \beta (\mu - e_j)}) - \log(1 - e^{\beta (\mu - e_j)}) \right\} = \sum_{j=1}^{K} \left[ (f_j + 1)e^{(\beta j + 1) \beta (\mu - e_j)} - e^{\beta (\mu - e_j)} \right] \left( \frac{e^{(\beta j + 1) \beta (\mu - e_j)} - 1}{e^{\beta (\mu - e_j)} - 1} \right). \] (31)

The expected value of the number of workers employed on the job sites with productivity \( c_j \), \(< n_j >\) is simply the corresponding term in the summation of \(< N >\) or equation (31).

\[
\langle n_j \rangle = (f_j + 1)e^{(\beta j + 1) \beta (\mu - e_j)} - e^{\beta (\mu - e_j)} \left( \frac{e^{(\beta j + 1) \beta (\mu - e_j)} - 1}{e^{\beta (\mu - e_j)} - 1} \right). \] (32)

Equation (32) determines the distribution of productivity in our stochastic macro-equilibrium.

Figure 3 shows how this model works. On one hand, there is dynamics of creation and destruction of potential job sites with various levels of productivity (Figure 3 (a)). We presented a Markov model which leads us to power-law tail of productivity distribution in the steady state. At the same time, there is another dynamics of job matching which the standard search theory analyzes. Convolution of two dynamics determines the distribution of workers at job sites with various productivities. The result is equation (32).

This distribution is fundamentally conditioned by aggregate demand. When the level of aggregate demand is high, high productivity firms make more job openings. They attract not only the unemployed but also workers currently on the inferior jobs. As Okun (1973) vividly illustrates, “a high pressure economy provides people with a chance to climb ladders to better jobs.” And people actually climb ladders in such circumstances.
A Simulation

With the help of a simple numerical example, we can better understand how equation (32) looks like, and also how the present model works. Figure 4 shows the distribution of $n_j$ given by equation (32). In the figure, the level of productivity $c_i = 1, \cdots c_{200} = 200$ are shown horizontally. The number of potential jobs or the ceiling at each productivity level, $f_j$, is assumed to be constant at 10 for $c_i, \cdots c_{50}$, while it declines for $c_j (j = 50, \cdots, 200)$ as $c_j$ increases. Specifically, for $c_j (j = 50, \cdots, 200)$, $f_j$ obeys a power distribution: $f_j \sim 1/c_j^2$. This assumption means that low productivity jobs are potentially abundant whereas high productivity jobs are limited. In the figure, the number of potential jobs is shown by a dotted line.

In this example, we have two cases; Case A corresponds to high aggregate demand whereas Case B to low aggregate demand. Specifically, $\beta = 1/\bar{T}$ is assumed to be (A) -0.01, and (B) -0.007. As explained in Table 1, Case (A) $\beta = -0.01$ corresponds to high demand $D$ whereas Case (B) $\beta = -0.007$ to low demand. In both cases, the number of workers or labor force $L$ is assumed to be 680. The number of employed workers $N$ is endogenously determined. The chemical potential $\mu$ is assumed to be -1.

In Figure 4, we observe that $n_j$ increases up to $j = 50$, and then declines from $j = 50$ to 200 in both cases. Broadly, this is the observed pattern of productivity dispersion among workers (Figure 2). What happens in this model is as follows. Whenever possible, workers strive to get better jobs offered by firms with higher productivity. That is why the number of workers $n_j$ increases as the level of productivity rises in the relatively low productivity region. The number of workers $n_j$ becomes an increasing function of $c_j$ because potential jobs with low productivity are abundant. Note that the number of potential jobs or the ceiling is not an increasing function of $c_j$ but constant in this region.

The number of workers $n_j$ turns to be a decreasing function of productivity $c_j$ in the high productivity region simply because the number of potentially available jobs $f_j$ declines as $c_j$ rises. Note, however, that $n_j$ is not equal to $f_j$. The ratio of the number of employed workers to the potential jobs, $n_j / f_j$, is much higher in the high productivity region than in the low productivity region (Figure 5). Again, this reflects the fact that workers always strive to get better jobs offered by firms with higher productivity.

In this model, firms are assumed to be monoplistically competitive facing downward sloping individual demand curve for their own products. Job offers made by such firms depend ultimately on the aggregate demand (Solow (1986)). In Figure 4, two distributions of $n_j$ are shown: Case (A) and Case (B). They depend on high and low
levels of aggregate demand $D$, respectively. When aggregate demand rises, the distribution of productivity as a whole goes up. Figure 5 indeed shows that more workers are employed by firms with higher productivity. Attractiveness of job is not simply determined by wages. However, to the extent that wages offered by firm are proportional to the firm’s productivity, the distribution shown in Figure 4 corresponds to the wage offer distribution function in the standard search literature. It depends crucially on the level of aggregate demand.

When aggregate demand $D$ goes up, the number of employed workers $N$ which corresponds to the area below the distribution curve, increases. Specifically, $N$ is 665 in case (A) while it is 613 in case (B). It means that given labor force $L = 680$, the unemployed rate $U/L = (L - N)/L$ declines when aggregate demand $D$ rises. In this example, the unemployment rate is 2.2% in case (A) while it is 9.9% in case (B).

**Summary**

Let us summarize economics of the present model. In the first place, the number of potential job sites with various levels of productivity is assumed to be given. They are determined not only by capital accumulation and technical progress but also by allocative disturbances to demand. The existing stock of capital and technology only slowly change, but allocative demand disturbances can swiftly change the economic values of the job sites associated with those capital and technology. Creative destructions due to Schumpeterian innovations raise the levels of productivity of some job sites, but at the same time, lower the levels of productivity of others. We consider a Markov model to describe dynamics of creations and destructions of potential job sites, and derive the conditions for the stationary state. The number of job sites with high productivity in the stationary state turns out to be power distribution. The important point is that job sites with relatively low productivities are abundant whereas those with high productivities are limited following power distribution. In passing, Houthakker (1955) shows that the aggregate production function becomes Cobb-Douglas on the assumption that the distribution of productivity (labor and capital coefficients in his model) is the power distribution: See also Sato (1974).

Now, let us visualize a potential job site as a “box”. Each job site or box is associated with a particular level of productivity. It is either empty or occupied by a worker. We can consider that a firm is nothing but a cluster of many job sites or boxes. It is actually conceivable that productivity differs across boxes in a single firm.

To fill a job site or a box with a worker is a firm’s economic decision. To do so, the firm can either keep the existing worker on the job or post a vacancy sign for
workers searching better jobs. In the economy as a whole, there are tens of million potential job sites with different levels of productivity. Employment of a worker on a job site with a particular level of productivity is the outcome of a successful random matching of a firm and a worker. We considered how employment of worker on the job sites with various levels of productivity is determined.

Because micro shocks facing firms and worker are so complex and unspecifiable, we cannot usefully pursue the micro behaviors. The single matching function for the economy as a whole will not do because there are many different levels of productivity. Every vacancy sign a firm posts is associated with a particular level of productivity, and a firm is looking for a suitable worker. Similarly, a worker is not looking for any job, but a kind of job which he/she thinks is suitable to him/her. Wages are obviously important, but attractiveness of job is not fully determined by wages, but depends on many factors; workers are interested in tenure, location, fringe benefits and other work conditions. The relative importance of these factors differs across workers. Likewise, the type of workers the firm is looking for cannot be simply defined, but depends on many factors. Again, the relative importance of these factors differs across firms. Besides, economic conditions facing workers and firms keep changing in unspecifiable ways. Given such complexity, optimization exercises based on representative agent assumptions do not help us find the outcome of random matching of workers and firms. Here comes the method of statistical physics.

The basic assumption of the model is that firms with higher productivity can afford to make more attractive job offers to workers. Firms are monopolistically competitive in that they face downward-sloping individual demand curves for their products. Then, their production and employment decisions depend ultimately on the level of aggregate demand; see Solow (1986). Given frictions, imperfect information, and search costs, the micro-behaviors of all the workers and firms are not precisely determined, however. Rather, when the level of aggregate demand is high, it is more likely that high productivity firms keep more workers on the job and post more vacancy signs to attract good workers. Workers are certainly aware of it; we know that quit rates go up in booms and down in recessions. As a result of such actions of both firms and workers, the distribution of productivity tilts to the direction of high productivity (Figures 4 and 5). As Okun (1973) argues, when aggregate demand rises, workers on the job “climb ladders to better jobs” without experiencing any spell of unemployment. At the same time, more workers currently in the unemployment pool find acceptable jobs. Employment $N$ increases, and the unemployment rate declines.

In the standard search theory, the matching function relates the successful
worker-job meetings to the number of workers unemployed and job vacancies. As a first step, it may be a useful analytical concept. However, it abstracts itself from qualities of job seekers and vacancies, and depends only on numbers. It also focuses only on the unemployed despite of the fact that by far a majority (more than 90 percent) of workers actually find their new jobs without experiencing any spell of unemployment. In these respects, it misses important aspects of labor market. The present analysis nicely captures these aspects. The most important message of the present analysis is, however, that the fundamental variable affecting the matching of workers and firms is aggregate demand.

5. The Principle of Effective Demand

According to Keynes (1936), the aggregate demand determines the level of output in the economy as a whole. Factor endowment and technology may set a ceiling on aggregate output, but the actual level of output is effectively determined by demand causing endogenous changes in utilization of production factors.

Modern macroeconomics is ready to regard factor endowment, technology, and preferences as exogenous in the short-run, but rejects the idea that even a part of demand is exogenous. This is in stark contrast to Keynesian economics. Keynes did not regard all the demand exogenous. Plainly, a significant part of demand is endogenously created by production implying more agents with greater purchasing power in the economy. This is, of course, the basic idea behind consumption function. Keynes’ principle of effective demand, however, regards a part of aggregate demand as exogenous, that is, almost impossible to explain within the framework. This exogenous demand is a fundamental determinant of the level of aggregate production. Keynes himself emphasized the volatility of investment which is in turn caused by volatility of “long-term expectations” (Keynes (1936, Chapter 12)).

The point was well taken in the review of the General Theory written by Hicks (1936) immediately after the publication of the book.

“There thus emerges a peculiar, but very significant, type of analysis. If we assume given, not only the tastes and resources ordinarily assumed given in static theory, but also people’s anticipations of the future, it is possible to regard demands and supplies as determined by these tastes, resources and anticipations, and prices as determined by demands and supplies. Once the missing element — anticipations — is added, equilibrium analysis can be used, not only in the remote stationary conditions to which many economists have found themselves driven back, but even in the real world, even in the real world in “disequilibrium.” …

From the standpoint of pure theory, the use of the method of
expectations is perhaps the most revolutionary thing about this book; but Mr. Keynes has other innovations to make, innovations directed towards making the method of anticipations more usable. (Hicks 1936, p.240)

It is a big irony that after forty years, a new method of expectations — rational expectations came to turn Keynesian economics. However, in the review, Hicks (1936) accepts that we assume given “people’s anticipations of the future” for fruitful economic analysis. It means that we regard a part of aggregate demand as exogenous. This assumption leads us to the old Keynesian economics.

“The central Keynesian proposition is not nominal price rigidity but the principle of effective demand (Keynes, 1936, Ch. 3). In the absence of instantaneous and complete market clearing, output and employment are frequently constrained by aggregate demand. In these excess-supply regimes, agents’ demands are limited by their inability to sell as much as they would like at prevailing prices. Any failure of price adjustments to keep markets cleared opens the door for quantities to determine quantities, for example real national income to determine consumption demand, as described in Keynes’ multiplier calculus. …

In Keynesian business cycle theory, the shocks generating fluctuations are generally shifts in real aggregate demand for goods and services, notably in capital investment.(Tobin, 1993)”

Real aggregate demand is, of course, $D$ in the model presented in the previous section. Figure 6 shows indices of exports and industrial production of Japan during the post-Lehman “great recession”. It is most reasonable to regard a sudden fall of exports as exogenous real demand shocks to the Japanese economy (Iyetomi et. al. (2011)). The principle of effective demand is alive and well! This paper shows that the principle of effective demand has good micro-foundations.

6. Concluding Remarks
Tobin (1972) advanced the idea of “stochastic macro-equilibrium.” 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.91)”

The equilibrium search theory analyzes a kind of stochastic macro-equilibrium. While acknowledging the achievement of search theory, we must point out that it rests on many unrealistic assumptions. The theory is, in fact, a curious hybrid of two different assumptions. On one hand, it presumes many heterogeneous agents, but on the other representative agents. Namely, it emphasizes heterogeneity of agents arising from the
presence of various market frictions and matching costs while at the same time it carries out optimization exercises, in effect, on representative agent assumptions. To get closer to reality, we must abandon the pursuit of the exact micro behavior of economic agent. This is the motive for the introduction of the matching function in the standard literature. Though it is a useful first step, it still remains half way. In the first place, the matching function focuses only on the unemployed, but by far a majority (more than 90 percent) of workers actually find their new jobs without experiencing any spell of unemployment. Clearly, focus only on the unemployed is inadequate for the purpose of studying matching of firms and workers. Secondly, firms differ with respect to productivity while workers differ with respect to their human capital and job preferences. Indeed, these micro differences are the ones that make matching of firms and workers difficult in the economy. Nonetheless, the standard matching function abstracts itself from these micro differences. Besides, economic environments facing firms and workers keep changing because of micro shocks. Once we recognize that these micro shocks are truly unspecifiable, we need a “greater macro black box” than the standard matching function.

To eschew pursuing the precise behavior of micro unit for understanding macro system is the basic principle of statistical physics. This paper presented a model of stochastic macro-equilibrium based on this principle. The concept of stochastic macro-equilibrium is motivated by the presence of all kinds of unspecifiable micro shocks. At first, one might think that allowing all kinds of unspecifiable micro shocks leaves so many degrees of freedom that almost anything can happen. However, the methods of statistical physics — maximization of entropy — actually provide us with the quantitative prediction about the equilibrium distribution of productivity, namely equation (32).

It is important to recognize that the present approach does not regard behaviors of workers and firms as random. They certainly maximize their objective functions perhaps dynamically. Randomness which plays a crucial role in our analysis comes from the fact that micro-shocks are so complex and unspecifiable that those of us who analyze the macroeconomy must take micro behaviors as random.

We just assume that workers move to better jobs if they are offered, and that firms try to fill potential job sites with workers by making job offers compatible with their respective levels of productivity. The firms are assumed to be monopolistically competitive facing downward sloping individual demand curves for their products. Then, their production and employment decisions depend on the level of aggregate demand; see Solow (1986). Specifically, when the level of aggregate demand is high, it
is most likely that high productivity firms put more vacancy signs than in the period of low demand. Workers become aware of such a change. It is demonstrated by the fact that quit rates are higher in high-demand periods. In this way, whether or not better jobs are really offered and workers move to those jobs depends ultimately on the level of aggregate demand. The most probable outcome of random matching of firms and workers is given by equation (32) which broadly coincides with empirically observed distribution of productivity.

This theory provides a correct micro-foundation for the Keynesian economics. Keynes’ theory has been long debated in terms of unemployment or “involuntary” unemployment. Though unemployment is one of the most important economic problems in any society, to focus only on unemployment is inadequate for the purpose of providing micro-foundations for the Keynesian economics. As the problem of changing capacity utilization reminds us, the real issue is whether or not there is any room for mobilizing production factors to high productivity jobs, firms, or sectors. The famous Okun’s law demonstrates that there is always such a room in the economy (Okun (1962)). Based on the method of statistical physics, the present paper quantitatively showed how labor is mobilized when the aggregate demand rises.

Modern micro-founded macroeconomics — real business cycle theory, endogenous growth theory, and new Keynesian models — begins with optimizing behavior of representative agent. However, this methodology is fundamentally flawed; see Kirman (1992) for a forceful criticism. In fact, representative agent assumptions trivialize the problems of uncertainty, imperfect information and search frictions in the real economy. We cannot understand the macroeconomy by simple homothetic enlargement of representative agents.

To take into account countless unspecifiable micro shocks, we must resort to the method of statistical physics. It leads us to Keynes’ principle of effective demand. The level of aggregate demand is the ultimate factor conditioning the outcome of random matching of firms and workers. By so doing, it changes the distribution of productivity, and consequently the level of aggregate production. The level of aggregate demand \( D \) is taken as exogenous in our model. It is, of course, an important research agenda to inquire how \( D \) is determined, but it is precisely what the old macroeconomics is all about (Tobin (1993)).
Table 1: Negative Temperature and Aggregate Demand

<table>
<thead>
<tr>
<th>Negative Temperature $T$</th>
<th>$-\infty$ ...... $0^\circ$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta = 1/T$</td>
<td>$0^\circ$ ...... $-\infty$</td>
</tr>
<tr>
<td>Aggregate Demand $D$</td>
<td>Low ...... High</td>
</tr>
</tbody>
</table>
Figure 1: Distribution of Labor across Sectors with Different Productivity

Source: Aoki and Yoshikawa (2007, p.82)
Figure 2: Distribution of Labor Productivity in Japan

(a) Manufacturing Sector in 2004

(b) Nonmanufacturing Sector in 2004

Source: Iyetomi (2012)
Figure 3: Model of Stochastic Macro-equilibrium

(a) Dynamics of Creation and destruction of Potential Jobs

Note: Both productivity and the number of potential job sites are in the natural logarithm. The straight line as drawn in the figure means that the distribution of productivity is power-law.

(b) Stochastic Macro-equilibrium

Aggregate Demand $D$ or $\beta$
Figure 4: Distribution of labor Productivity

The number of employed workers

Note: (A) High Aggregate Demand, (B) Low Aggregate Demand
See the main text for details.
Figure 5: Percentage of Potential Job Sites Occupied by Employed Workers

Note: (A) High Aggregate Demand, (B) Low Aggregate Demand
See the main text for details.
Figure 6: Indices of exports and the Industrial Production, normalized to 100 for 2005

Source: Cabinet Office, Ministry of Economy, Trade and Industry
References


Cambridge: Cambridge University Press.


