Equilibrium Distribution of Agents by Types in a Market, and Existence of Power Laws

Masanao Aoki
Department of Economics, University of California Los Angeles
Los Angeles, Ca 90095-1477
and
Department of Economics and Information, Gifu Shotoku Gakuen University.
1-38 Nakauzura, Gifu City, 500-8288

December 7, 1999

Abstract

This paper considers a market in which shares of a single asset is traded by agents with many strategies. The market behavior is examined by regarding clusters of agents with same trading rules as random combinatorial partitions of agents by types. The equilibrium distribution is shown to be the Ewens sampling formula. When agents are positively correlated, it is shown that a small number of groupings such as two types of agents dominate the markets. Distributions for large price difference or return are shown to be power-laws if a certain set of conditions is satisfied.

Introduction

This paper examines behavior of an open market in which a large number of agents trade single asset such as stock index or shares of a holding company. We refer to this type of markets as share markets.[1] Traders may employ many different types of strategies or trading rules, and they may switch to different strategies at any time. The market is open because the traders may enter and exit the market at any time. We treat share markets rather than regular stock markets in which many different types of shares are traded in order to focus on stationary distributions of random groupings of agents by common strategies, or sentiments, and their effects on the prices and returns.

We apply the notion of random exchangeable partitions, due to Kingman (kin78),(king78) to examine the equilibrium behavior of the fractions of agents of various types in this share market. To keep the paper to a reasonable size we examine only the situations in which two dominant types of agents, that is two largest subgroups of agents, with the same strategy within each subgroup, emerge to determine the market-wide excess demands for the shares and hence changes in the share prices and returns.

[1] A deterministic share market with two agents has been analyzed in (day90).

Two kinds of state variables, one familiar and the other less familiar to economists, are used to describe the groups formed by agents with the same rules in the market. The first state vector lists the number of users of each strategy j, $j = 1, 2, \ldots, K$, where K is the number of strategies currently in use. For the moment, think of K as finite and known. This assupmtion is later abandoned. We identify agent types with the strategies they employ and speak of agents of type j. The number of strategies may vary with the number of participants. Then we use K_N, where N is the total number of agents when we wish to emphasize this point. In the language of the occupancy problem, think of N dislabelled or indistinguishable balls and K labelled or distinguished boxes. Although the use of this vector seems very natural, we later show the problem with this way of describing the state of the market.

It is more effective for our objective to use another state vector defined by $\mathbf{a} = (a_1, a_2, \ldots, a_n)$, where a_i is the number of types (strategies) with exactly i agents. Sachikov calls these two types of state vectors as the specifications of first and second kind, (sachikov97). The vector $\mathbf{a}$ is named as partition vector by Zabell (zabell92). Here, both balls and the boxes into which they fall are dislabelled.

In this paper we think of the patterns of groupings or clusterings of agents, each with the same type in the subsets of agents, as partitions of a set of agents by types. More precisely, we treat them as exchangeable random partitions in the sense of Kingman. Two partitions are exchangeable when the two partition vectors are identical. The two partitions are then equally probable by assumption. See Zabell.

We first describe the equilibrium distribution for the patterns in terms of the partition vector. After characterizing the distribution we derive the marginal distributions for the two largest fractions, which is then used to characterize market equilibirum. The distribution for returns with large magnitudes is then heuristically shown to have power-laws. The paper concludes with some observations on empirical tests.

Model of Agent Participation: Transition Rate Specifications

This section formulates a jump Markov process which governs the ways agents enter or exit the market, change their strategies over time, and cluster into several types in the market. When there are K types of agents, the state vector defined by $\mathbf{n} = (n_1, n_2, \ldots, n_K)$ is used first, where n_i is a non-negative integer which denotes the number of agents of type i, $i = 1, 2, \ldots, K$. It shows how agents are distributed over K types. Let $N = \sum_{i=1}^{K} n_i$. Then, $\mathbf{n}/N$ is the empirical distribution, also called frequencies. In the above and in what follows we omit the time variable from expressions such as $n_i(t)$.

Dynamics are described by the master equation and are determined once we specify transition rates between the states under the conditions to rule out pathological behavior by Markov processes. (The conditions rule out the process making an infinite number of jumps in a finite time.) The method for constructing Markov process with these transition rates is standard due to Feller.

For example, with $K = 2$ there are six admissible transitions of state: entry of each of two types of investors; departure of each of two types; a change of strategy by a type one agent into a type two and its reverse transition from type two to type one. These transitions are conveniently expressed by defining two vectors $\mathbf{e_1} = (1, 0)$, and $\mathbf{e_2} = (0, 1)$. Entry and exit by a type i agent are the events denoted by state transition $\mathbf{n} \to \mathbf{n} + \mathbf{e_i}$, and$\mathbf{n} \to \mathbf{n} - \mathbf{e_i}$, for $i = 1, 2$. The number of agents in the market changes by one in these two transitions. Changes of strategies by an agent from strategy i to j is denoted by

$$\mathbf{n} \to \mathbf{n} - \mathbf{e_i} + \mathbf{e_j},$$

where $i, j = 1, 2$, and $i \neq j$ since n_i is decreased by one and n_j increases by one. For example $\mathbf{n} - \mathbf{e_2} + \mathbf{e_1} = (n_1 + 1, n_2 - 1)$ means that one of type two participants changes into a type one agent. Note that unlike entries or exits, the total number of agents in the market remains the same in the type changes. With $K > 2$ there are more transitions. The notation $w(\mathbf{n}, \mathbf{n'})$ is used to denote the transition rate from state $\mathbf{n}$ to $\mathbf{n'}$.

We posit

$$w(\mathbf{n}, \mathbf{n} - \mathbf{e_i}) = \phi_i(n_i), \tag{1}$$

where $\phi(0) = 0$,

$$w(\mathbf{n}, \mathbf{n} + \mathbf{e_i}) = \psi_i(n_i), \tag{2}$$

and

$$w(\mathbf{n}, \mathbf{n} - \mathbf{e_i} + \mathbf{e_j}) = \lambda_{i,j}\phi_i(n_i)\psi_j(n_j), \tag{3}$$

$i \neq j$, and $i, j = 1, 2$.

Eqs. (1) through (3) specify that agents of either type may enter or change his or her type at rates which are influenced by the number of agents of the same type in the market. By appropriate choices of the function forms ϕ or ψ, we may model phenomena such as herd behavior, bandwagon effects, group sentiments and the like.

A special case in which the functions are linear is of particular interest, since this specification leads us to the Ewens sampling formula which is justly famous in the population genetics literature. Let us specify $\phi(n_i) = d_i n_i, d_i \neq 0$, and $\psi_i(n_i) = b_i + c_i n_i$. The process with these transition rates is sometimes called a birth-dealth with immigration process, since there is a term b_i which is independent of n_i in the entry transition rate.

Poisson or negative binomial distributions can be generated as equilibrium distributions by appropriate choices of the function forms for ϕ and ψ. We consider market behavior with these distributions to be an important and informative benchmark or limiting case.[2]

[2] Although we do not explicitly discuss microeconomic factors involved in decisions by agents to either enter, exit or change strategies such as cumulative profit or loss figures or explicit evaluations of merits of alternative choices, they are implicit in the specifications of transition rates we employ. See (aok96) (Sec. 3.3, 5.5, and 6.3) for further elaborations. In the face of uncertain consequences of any choice, these specifications may be thought of

Equilibrium Probability Distribution

In this section we describe the equilibrium distribution of patterns of agent clustering, first in terms of the empirical distribution, then, in terms of the partition vector, **a**. This alternative description of the model state leads to the distribution function due to Ewens, famous in the population genetics literature, (ewens90).

Equilibrium Distribution in terms of n

When we use the empirical distribution as state vector, the condition for the equilibrium distribution is stated as

$$\pi(\mathbf{n})\sum_{\mathbf{n}'} w(\mathbf{n},\mathbf{n}') = \sum_{\mathbf{n}'} \pi(\mathbf{n}')w(\mathbf{n}',\mathbf{n}), \tag{4}$$

where summation is over all possible next states; for example, six possible next states in the case of $K = 2$, and where $\pi(\mathbf{n})$ is the equilibrium probability of state vector $\mathbf{n}$. This equation states that the probability influx and outflux balance out in statinary states.

In general, stationary probability distributions cannot be obtained analytically. When the equality holds term-by term, that is, when we assume that the detailed balance conditions hold, we can sometimes solve for the equilibrium distributions. For example, with $K = 2$, there are six separate equations. When we impose the detailed balance condition, the solution of any one of the six equations turns out to satisfy the remainder of the equations with suitable conditions imposed on the model parameters. For example, we start with (2), or with (1) or (3). It does not matter how we begin.

We first verify for the case with $K = 2$ that the stationary probabilities in product form satisfy the detalied balance condition. That is, we posit $\pi(n_1, n_2) = \pi_1(n_1)\pi_2(n_2)$.

The detailed balance condition using (2) yields the first order difference equations

$$\pi_i(n_i + 1) = \frac{w(\mathbf{n}, \mathbf{n} + \mathbf{e_i})}{w(\mathbf{n} + \mathbf{e_i}, \mathbf{n})}\pi_i(n_i) = \frac{\psi(n_i)}{\phi(n_i + 1)}\pi_i(n_i),$$

$i = 1, 2$. Note that $w(\mathbf{n} + \mathbf{e_i}, \mathbf{n})$ is the departure rate when the number of type i agent is $n_i + 1$. This is a first order difference equation for the probabilities. Iterating this relation we obtain $\pi_i(n_i) = B_i \prod_{r=1}^{n_i} \frac{\psi_i(r-1)}{\phi_i(r)}$, where B_i is the normalizing constant. When we specialize $\psi_i(r)$ to ν_i and $\phi_i(r)$ to $\mu_i r$, we obtain Poisson distributions, given the linear transition rates $\phi_i(n_i) = \mu_i n_i$, and $\psi_i(n_i) = \nu_i$. It is straightforward to verify that these expressions satisfy the remaining terms of the set of the detailed balance equations, (3),which in the linear case becomes $\lambda_{i,j} n_i$, that is , the full equilibrium equation holds, provided we impose the condition that $\alpha_1\lambda_{12} = \alpha_2\lambda_{21}$.

as first order approximations since transition rates are likely to be complicated nonlinear functions of state vector components. Linearly specified transitions rates may be thought of as the first order approximations to nonlinear transition rates ϕ and or ψ. See (aok96) p.121 for higher order terms of transition rates.

Equilibrium Distribution in terms of Partition Vector

Here, we introduce another way of describing patterns of clusters of agents by types.

Let a_i be the number of types represented by i agents, $i = 1, 2, \dots, N$, and define the partition vector by $\mathbf{a} = (a_1, a_2, \dots, a_N)$. To understand this vector, think of indistinguishable or delabelled balls dropped randomly among K indistinguishable or delabelled boxes. The component a_i is the number of boxes with exactly i balls in each of them. Zabell characterizes these as as frequencies of the frequencies, where the components of the empirical distributions n_k/N, $k = 1, 2, \dots, K$, are the frequencies. They are also called abundances in the biological literature. Partition patterns of N agents over K types become exchangeable random partitions when two partitions with the same partition vectors are treated or assumed as equiprobable. This notion of Kingman generalizes the notion of exchangeable sequences or random variables of de Finetti. By definition, the total number of agents (balls) is $\sum_{i=1}^{N} i a_i = N$, and the number of types represented in a sample of size N (number of non-empty boxes) is $\sum_{i=1}^{N} a_i = k \leq K$. Usually $k = K$. In situations where a large number N of K types are in the population and n of them are sampled, the types represented by this sample of size n could be less than K.[3]

With the transition rates specified as the birth-death with immigration process mentioned above we solve for the equilibrium distribution of the partition vector, given linear transition rates with $\phi_i(n_i) = d_i n_i$, and $\psi_k(n_k) = b_k + c_k n_k$,

$$\pi(\mathbf{a}|N) = \frac{K!}{a_1! a_2! \cdots a_N! (K - \sum a_i)!} C^{-1}_{-Kf,N} \prod_k C^{a_k}_{-f,k},$$

where $f = b_k/c_k$ is taken to be the same for all k for simplicity, and where $C_{-f,k}$ is the negative binomial coefficient, which is equal to $C_{f+k-1,k}(-1)^k$, where $C_{n,j}$ is the binomial coefficient $n!/j!(n-j)!$.

Now let K become very large while $Kf = \theta$ is held fixed. The probability approaches

$$\pi(\mathbf{a}|N) = C^{-1}_{\theta+N-1,N} \prod_i (\frac{\theta}{i})^{a_i} \frac{1}{a_i!}. \qquad (5)$$

Note that for small f, $C_{-f,k} \approx (-1)^k f/k$. We recognize that $\prod_i C_{-f,n_i} = \prod (C_{-f,k})^{a_k}$ because a_k of ns are equal to k.

This distribution is known as the Ewens sampling formula. See (ewens90), and (joh97) for further detail of the sampling formula.

Multivariae Ewens Distributions

The exchangeable random partitions have the equilibrium distribution called MED in Johnson, Kotz, and Balakrishnan (joh97). We describe two ways of deriving this distribution in this paper, because each is instructive in its own

[3] Costantini and Garibaldi treat two interacting systems one of which is regarded as heat bath in this way (cost97).

way. One is via specifications of transition rates of agents in the market to generate negative binomial disstribution, discussed above. The other is via the Dirichlet distribution as the representing measure in the exchangeable random partitions leading to Kingman's Poisson - Dirichlet distributions, (kingman93).

Combinatorial Structures and the Ewens Distribution

We consider the following special case. There is a large number of agents of a large number of types as population. However, only two of the types are dominant, that is, the sum of the two largest fractions of agents is nearly one. In the above section, we have introduced the Ewens sampling formula as a limiting distribution from the negative binomial distribution. Here we present an alternative derivtion of the Ewens formula starting from the K dimensional multinomial distribution with probability vector $\mathbf{p} = (p_1, p_2, \dots p_K)$ with p_i being the probability of an agent being type i. Then assume that the prior distribution for $\mathbf{p}$ is given by a symmetric Dirichlet distribution with parameter ϵ (symmetric because agents are exchangeable, that is, the index given to agents are merely for convenience of reference and has no substance)

$$\Pr(p_1, p_2, \dots, p_K) = \frac{\Gamma(K\epsilon)}{\Gamma(\epsilon)^K}(p_1 \cdots p_K)^{\epsilon-1}, \tag{6}$$

where the probabilities are all assumed to be positive and sum to one.[4] We show that the order statistics from this distribution leads to the Ewens sampling formula. For a more elegant derivation of the Poisson-Dirichlet distribution which is related to this, see kingman93, chapt. 9.

To proceed, average the ps in the multinomial distribution with the Dirichlet distribution. The result is the Dirhlet-multinomial distribution,

$$\Pr(n_1, n_2, \dots, n_K) = \frac{N!}{n_1! \cdots n_K!} \frac{\Gamma(K\epsilon)}{\Gamma(\epsilon)^K} \frac{\Gamma(n_1+\epsilon) \cdots \Gamma(n_K + n_K)}{\Gamma(K\epsilon + N)}.$$

The joint probability distribution of the first k largest numbers of agents is

$$\Pr(n^{(1)}, n^{(2)}, \dots, n^{(k)}) = B \frac{\Gamma(K\epsilon)}{\Gamma(\epsilon)^k} \frac{\Gamma(n^{(1)}+\epsilon) \cdots \Gamma(n^{(k)}+\epsilon)}{\Gamma(K\epsilon + r)},$$

where $n^{(1)} \geq n^{(2)} \geq \dots$, and $r = n^{(1)} + \cdots + n^{(k)}$ is the number of agents in the first k largest types, and

$$B = \frac{r!}{n^{(1)}! \cdots n^{(k)}!} \times \frac{K!}{(K-k)! a_1! a_2! \cdots a_r!},$$

where a_i is the number of $n^{(j)}$, $j = 1, 2, \dots, k$ that are equal to i. By definition, $\sum_i a_i = k$, and $\sum_i i a_i = n$. These factors arise because of the order statistics.

[4] The reader may wonder why this particular distribution. The Dirichlet distribution is the correct mixing measure in representing the mesures for exchangeable random partitions. See [**?**].

We now examine the consequences of large K, while letting $K\epsilon$ converge to a positive number θ. Use the relation $\Gamma(\epsilon) = \Gamma(\epsilon+1)/\epsilon$ in letting ϵ go to zero. We have

$$\frac{K!}{(K-k)!\Gamma(\epsilon)^k} \to \theta^k.$$

Thus we arrive at the distribution of the first k order statistics

$$\Pr(n^{(1)}, \cdots, n^{(k)}) = \frac{n!\theta^k\Gamma(\theta)}{n^{(1)}n^{(2)}\cdots n^{(k)}a_1!\cdots a_r!\Gamma(\theta+r)}.$$

This may be written as $\Pr(n^{(1)}, n^{(2)}, \ldots, n^{(k)}) = \frac{n!}{\theta^{[n]}}\prod_1^n (\frac{\theta}{j})^{a_r}\frac{1}{a_r!}$. In the above we note that $n^{(1)}\cdots n^{(k)} = 1^{a_1}2^{a_2}\cdots n^{a_n}$. Note also that the ratio $\Gamma(\theta)/\Gamma(\theta+n)$ is equal to $1/[(\theta+n-1)(\theta+n-2)\cdots\theta] := 1/\theta^{[n]}$. In this form we clearly see a connection with the Cauchy formula for permutations in product of cycles, which is obtained by setting θ to 1. When θ is not equal to one, all permutations are not equi-probable. Probabilities are skewed or biased with weight $\theta^k/\theta^{[n]}$.

Effects of values of θ

It is useful to have some intuitive understanding of parameter θ. It is a measure of correlatedness of individual agents. To see this suppose that $N = 2$, $K = 2$, $a_1 = 0$ and $a_2 = 1$, that is, randomly select two agents, and suppose that they are both of the same type. From (5) the probability that two randomly sampled agents are of the same type is $1/(1+\theta)$.[5] This shows that two agent are more likely to be of the same type as θ becomes smaller. Conversely, as θ becomes larger, two randomly chosen agents are likely to be independent. We see from the distribution that parameter θ determines degree of correlatedness of agents.

The probability density of the number of types observed in a sample of size n, K_n, is

$$P_n(K_n = k) = \frac{1}{\theta^{[n]}}c(n,k)\theta^k, \tag{7}$$

and

$$E(K_n) = \sum_{k=1}^{n} k\theta^k c(n,k)/\theta^{[n]}, \tag{8}$$

where $c(n,k)$ is he signless or unsigned Stirling number of the first kind. One way to characterize it is by $\theta^{[n]} = \sum_k c(n,k)\theta^k$, where $\theta^{[n]} = \theta(\theta+1)\cdots(\theta+n-1)$. See (ewens90) or (hop87) for derivtion of these formulas.

We can use this formula to verify that the expected number of types increase with θ. As θ goes to infinity, the expected number of types approaches n, namely, total fragmentation of agents in the sample by types.

As we show shortly, the parameter value of θ reflects the degree of correlatedness of agent types. We can evaluate the effects of increasing correlations or mutual dependence on the size of Ea_j by taking partial derivative of

[5] This probability can also be calculated as $E(\sum_i p_i^2) = \int_0^1 x^2 g(x)dx$, where $g(x) = \theta x^{-1}(1-x)^{\theta-1}$ is the probability density that in an interval $(x, x+dx)$ there is a type. It is called the frequency spectrum by Ewens. This notion arises in other areas as well. See hig95 for example.

it with respect to θ: As θ increases, Ea_j for j much smaller than n increases linearly in θ.

For small values of θ, $E(K_n) = \sum_{j=0}^{N-1} \frac{\theta}{\theta+N-j} \approx 1 + \theta[\ln(N-1) + \gamma]$, where $\gamma = .577$ is the Euler's constant.

Combinatorial Aspects of a Large Number of Agents

What is most remarkable about the patterns of clusters of a large number of positively correlated agents is that some small subsets of configurations account for majority of possible patterns. That is, some small number of configurations are most likely to be realized or observed. This feature has been noticed in other context as well. See [**?**] for example. For example, suppose that θ is about 0.4, that is two randomly selected agents are of the same type with probability about 0.7. Then, $E(x) \approx .79$, and $E(y) \approx .20$. With $\theta = .4$, we also note that $E(K_{10}) = 2.1$, $E(K_{100}) = 3.0$, $E(K_{1000}) = 4.0$, $E(K_{10^5}) = 5.8$, and $E(K_{10^7}) = 7.7$. These figures indicate that there are several small fractions in addition to the two large ones when $N \geq 100$.

Let $E(x)$ and $E(y)$ be the expected values of the two laragest fractions, x and y, and suppose that $E(x) + E(y) \approx 1$, so that clusters of agents of other types are smaller fractions and may be ignored. Then, approximating $1/(1+\theta)$ by $E(x)^2 + E(y)^2$, we use Table III in (guess77) where numerical values of the expected values of the largest fraction is listed for different values of θ. For example, with $\theta = .4$, $E(x) = .79$. They also give an approximate formula for the expected value of the second largest fraction as

$$E(y) \approx (\theta \ln 2)E(x).$$

We can approximately calculate the variances of the two largest fractions. For example, following (watt76) we calculate $E(xy)/E(x) \geq \frac{\theta}{\theta+1}[1-(\frac{1}{\theta})^{\theta+1}]$, and use $y \approx 1-x$ to arrive at $var(x) \approx E(x)[1-\theta/(1+\theta)+(1/2)^{\theta+1}\theta/(\theta+1) - Ex]$. On substituting the numerical values of $\theta = 0.4$ and $E(x) = .79$, we see that

$$\sqrt{var(x)}/E(x) \approx .20.$$

In other words, the standard deviation of the largest fraction is about 20 per cent of its mean. See (watt76), and (guess77) for more precise calculation procedures.

Two Large Fractions of Agents

Market Excess Demands

In this section we derive approximate expression for the market excess demands with two large fractions x and y.

Using P to denote the price level of the shares, we let $d_x(P)$ denote individual excess demand of the type which happens to be the largest fraction x. When there are only two strategies, it is either $d_1(P)$ of type 1 agent for the shares, or $d_2(P)$ for type 2.[6] Do likewise with $d_y(P)$ for the excess

[6] They correspond to type α and β respectively of (day90).

demand of the type which happens to be the second largest fraction. If x is type 2, then d_y is the excess demand of type 1, and vice versa. For definiteness suppose that $d_x(P) = d_1(P)$, and $d_y(P) = d_2(P)$. The market excess demand is then given by summing over individual excess demands

$$d(P)/N = xd_x(P) + yd_y(P) = (n_1/N)d_1(P) + (n_2/N)d_2(P),$$

where n_i is the total number of agents of type i, $i = 1, 2$, at time t. For shorter notation, argument t is suppressed from $P(t)$ when no ambiguity is likely.

We use the same function form as in Day and Huang $d_1(P) = (u - P)f(P)$, and $d_2(P) = -(u - P)$, with $f(P) = [(P - m)(M - P)]^{-1/2}$ where we have set $a = b = 1$ in their specification and set $u = (M + m)/2$ without loss of generality. We note that the two excess demands are of opposite sign, i.e., the two types of agents are on the opposite side of the market.

The inequality $(M - m)/2 \geq (x/y)$ is important for market to have mutiple equilibria. If this does not hold, then there is only one P, namely at u at which the excess demand vanishes. If this inequality does hold, then there are three prices at which the excess demand vanishes, namely u, P^*, and P_*, $P_* < u < P^*$, where P_* and P^* are the roots of $f(P) = y/x$.

The derivative of the market excess demand with respect to P is $d'(u) = -f(u)x + y$, which is negative if $(x/y) > (M - m)/2$, that is there is a unique equilibrium which is locally stable.

When the opposite inequality holds, $P = u$ is locally unstable because $d'(P) > 0$ then. The other two prices at which the market excess demand vanish are locally stable $d'(P^*) = -(u - P^*)^2 f(P^*)^3 < 0$, and $d'(P_*) = -(u - P_*)^2 f(P_*)^3 < 0$.

Equilibrium Distributions

The probability density for x and y, $1 \geq x \geq y \geq 0$, and $x + y \leq 1$ have been calculated in (watt76), and (guess77).

The expressions for the densities are simple for certain ranges of the variables but are complicated in the remainder of the ranges:

$$f(x) = \theta x^{-1}(1 - x)^{\theta - 1},$$

for $1/2 < x$, and

$$g(x, y) = \theta^2 x^{-1} y^{-1} (1 - x - y)^{\theta - 1},$$

for $0 \leq y \leq x \leq 1$, and $x + 2y \geq 1$.

In regions of $x < 1/2$ and $x + 2y < 1$ more complicated formula obtains and we must resort to numerical determination. In the previous subsections we have given expressions which give us some explanation how much volatility is expected in the market excess demands

$$var(d(P)/N) \approx [d_x(P) + d_y(P)]^2 var(x),$$

where $y \approx 1 - x$ has been substituted in, and where $d_x(P) + d_y(P) = a(u - P)[f(P) - 1]$ is non-zero with $P \neq u$.

Approximate Dynamics for Price Differences

In this section we examine behavior of large price differences. Analysis is slightly more complicated for returns, but the same in essence. We fix time interval Δ and write the recurrent equation for small deviations of the price. We define price difference by $\Delta P_t = P_t - P_{t-\Delta}$ and return at time t, r_t, by $P_t = P_{t-\Delta} e^{r_t}$.

We determine the dynamics of price changes in a share market, and heuristically indicate how the distribution may exhibit power-law behavior.

Suppose that the prices change in proportion to the market excess demand, that is, according to

$$P_{t+\Delta} = P_t + \kappa d(P_t, n_1(t), n_2(t)),$$

where κ is a small positive constant, $c \times \Delta$, where c is a positive constant, and Δ is also a small positive constant. Since we examine prices at discrete time instants which are Δ apart, we use P_k rather than subscript t. $P_{t+\Delta}$ is now indicated by P_{k+1}, for example. The expression for the excess demand is given by

$$d(P, n_1, n_2) = n_1 d_1(P) + n_2 d_2(P),$$

in the case where the largest fraction is of type 1. We have suppressed time subscripts for shorter notation. In the opposite case where type 2 is the largest fraction we have

$$d(P, n_1, n_2) = n_2 d_1(P) + n_1 d_2(P).$$

To be definite, we assume the former, that is $x = n_1/N$, and $y = n_2/N$.

To illustrate we use the excess demand functions used by Day and Huang,

$$d_1(P) = (u - P) f(P),$$

for a type 1 agent, and

$$d_2(P) = -(u - P),$$

for a type 2 agent where

$$f(P) = [(P - m)(M - P)]^{-1/2},$$

where $0 < m < P < M$. This is a device used by Day and Huang to keep P from wondering off to infinity, or being stuck at 0 in their simulation runs.

It is convenient to change variable from P to

$$z = \frac{P - u}{R},$$

where we introduce two parameters; $R = (M - m)/2$, which is half of the total allowable range of the price, and we set the middle point of the price range as the equilibrium price, $u = (M + m)/2$.

Market excess demand is zero at $P = u$ or at P which satisfies

$$f(P) = n_2/n_1.$$

With the function $f(P)$ specified above, there are two real values denoted by $0 < P_* < u < P^*$ when $R > x/y$, given by $P_* = u - \sqrt{R^2 - (x/y)^2}$, and $P^* = u + \sqrt{R^2 - (x/y)^2}$, or $z_* = -\sqrt{1 - (\frac{x/y}{R})^2}$, and $z^* = -z_*$. When the opposite inequality holds, this equation has no real roots, and u is the unique equilibrium price.

From the price adjustment dynamics we derive a first order difference equation for the price difference scaled by R, $z_k = \Delta P_k / R$ as

$$z_{k+1} = F_k z_k + G,$$

where

$$F_k = 1 + \kappa d_P,$$

with

$$d_P = \partial d(P)/\partial P = n_2 - n_1 R^2 h(P_k),$$

and where

$$h(P) = f(P)^3.$$

From these relations we have

$$F_k = 1 + \kappa N\{y - \frac{x}{R}[1 - z^2]^{-3/2}\}.$$

The expression for G is

$$G_k = -\kappa z_k[\Delta n_1/[R\sqrt{1 - z_k^2} - \Delta n_2],$$

where Δn_i is the change in the number of agents of type i during the time interval Δ, $i = 1, 2$. Using the transition rates for the jump Markov process, we can calculate the expected values involving G_k.[7]

Near the critical values of $P = u, P_*, P^*$, that is $x = 0, x_*, and x^*$, we note that

$$F_k(0) = 1 + \kappa[n_2 - n_1/R] = 1 + \kappa N[y - x/R],$$

which is greater than 1 by our assumption. The critical point u is locally unstable. At the other two critical points,

$$F_k(z^*) = F_k(z_*) = 1 - \kappa N y[R^2(y/x)^2 - 1].$$

We then appeal to the results obtained by Vervaat (1979), Letac (1986), Goldie (1991), de Haan et. al (1989) if we can treat (F_k, G_k) as i.i.d. random sequence or use Brandt (1987) if they are regarded as strictly stationary and ergodic to conclude that there is a unique stationary distribution z_∞ which satisfies

$$z_\infty = F z_\infty + G$$

where F and G are either F_1 and G_1 in the former, or F_∞ and G_∞ in the latter. The assumption of i.i.d. random sequences is likely to hold near the critical points since the coefficients are independent of the prices there.

[7] Alternatively we can use the results of Arratia et al. (1992) which show that the Ewens distribution can be approximated by Poisson random variables, conditioned by the sum, where the mean of the random variables are θ/j, $j = 1, 2, \ldots$.

To show the existence of power laws, we verify the existence of positive parameter γ

$$E\,|F|^{\gamma} = 1,$$

among others. The other conditions, $E(|F|^{\gamma} log^{+}|F|) < \infty$, and $EG^{\gamma} < \infty$, are satisfied.

Heuristically, letting $\Phi(\rho) = \Pr(x_{\infty} > \rho)$, we have an integral equation

$$\Phi(\rho) = \int \phi(f)\Phi(\frac{\rho - G}{f})df,$$

where $\phi(f)$ is the density function of F. For large values of ρ, this equation is approximated by

$$\rho^{-\gamma} = \int \phi(f)(\rho/f)^{-\gamma}df$$

or

$$1 = \int \phi(f)f^{\gamma}df,$$

that is $EF^{\gamma} = 1$ for some positive γ.

Write $F = 1 - \kappa s$ and evaluate $E\exp[\gamma \ln(1 - \kappa s)]$, assuming that $|\kappa s| < 1$. The condition that $EF^{\gamma} = 1$ leads approximately to

$$\gamma \approx 1 + \frac{2}{\kappa}\frac{Es}{Es^2},$$

with $sn_2[R^2(n_2/n_1)^2 - 1] > 0$, near z_* and Z^*. Hence, approximately we determine

$$\gamma = 1 + \frac{2}{\kappa R^2 N}\frac{E(y^3/x^2)}{E(y^5/x^4)},$$

if $R^2 > En_2/E(n_2^3/n_1^2) \approx E(y)/E(y^3/x^2)$, where we approximately treat N as constant equal to EN.

We see that the product of κ which control the speeds of price adjustments and R which determines the allowable range of price movements affects γ as $1/\kappa R^2$. To obtain some idea about the size of γ, suppose that $y \approx 0.2$, and $x \approx 0.8$. Then γ is larger than 2 if $\kappa R^2 N < 32$. Recalling that κ is the product of time interval Δ and a constant for adjusting excess demand, this inequality will hold with small κ.

When $Ex > REy$, the distribution about $x = 0$ has the power law index

$$\gamma \approx 1 - \frac{2}{\kappa RN}\frac{y - x/R}{(y - x/R)^2}.$$

Empirical Issues

Whether the Ewens sampling formula, which is found to be useful in the population genetics field, is applicable to share market data is a question which requires empirical tests, and is outside the scope of this paper. One of the important implications of the Ewens sampling formula is the "neutrality" issue. It is also related to the question of size-biased sampling. Basically the issue is the following: After removing all of the agents of the type belonging to the largest fraction, renormalize the remaing agents and examine the joint distribution of the remaining fractions of types after renormalization. This is the basic idea behind size-biased resampling. See (kingman93). The neutrality issue is that the size-biased samples have the same structure of distribution as the original one. See (ewens90) and references cited therein, and on robustness of the Ewens sampling formula and on cares one must exercise in statistical testings.

Discussion

We have proposed the Ewens sampling formula as a useful characterization of a collection of agents of several types. As pointed out by Zabell the situation is exactly the same as that of the so-called sampling of species problem faced by statisticians. New types of agents emerge in the process and we cannot exclude this possibility. Kingman's approach assigns positive probability for this unanticipated event.

Whether the Ewens sampling formula turns out to be useful in economics is yet unanswered. We feel that the answer is yes, because we see that there are several other applications of the approach outlined in this paper. For example, Herfindahl index is a measure of concentration often used in the industrial organization literature. By interpreting the shares of markets by firms of type i to be fraction x_i which is positive and sum to one, it is defined to be $\sum x_i^2$. Interpreting x_i as the fraction, it is governed by the frequency spectrum. The expected vaue is then

$$E(\sum_i x_i^2) = \theta \int x^{2-1}(1-x)^{\theta-1}dx = 1/(1+\theta).$$

References

aok96

Aoki,Masanao, *New Approaches to Macroeconomic Modeling: Evolutionary Stochastic Dynamics, Multiple Equilibria, and Externalities as Field Effects*, New York: Cambrdige University Press, 1996.

Arratia, R., A.D. Barbour, and S. Tavaré, "Poisson Approximations for the Ewens Sampling Formula", *Ann. Appl.Ptob.*— **2**,*519–35 (1992(.*

bra86

Brandt, A., " The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ *with stationary coefficcients", Adv. App. Peib.* **18**, *211–20 (1986)*

cost97

Costantini, D., and Garibaldi, U., Stud. Hist. Phil. Mod. Phys. **28**, *483 (1997).*

day90

Day, R., and Huang, W., J. Econ. Behavior Org. **14** *299 (1990).*

ewens90

Ewens, W. J., in Mathematical and Statistical Developments of Evolutionary Theory S. Lessard. ed, London:Kluwer, 1990

gold91

Goldie, C. M., "Implicit renewal theory and tails of solutions of random

equations", Ann. Appl. Prob. **1**, *126–66,(1991).*
hig92
Higgs, P. G., "Frequency distributions in population genetics parallet those in statistical physics", Phy. Rev. **51**, *95–101 (1995).*
hop87
Hoppe, F. M., J. Math. Biol. **25** *123 (1987).*
joh97
Johnson, N. ,Kotz, S., Balashrishnan, A., Multivariate Discrete Distributions, New York: Wiley, 1997, ch.10.
kin78
Kingman,J.F. K.,Proc. Roy. Soc. **361**, *1 (1978).*
king78
Kingman, J, F. K., J. Lond. Math. Soc. **18**, *347 (1978a)*
kingman93
Kingman, J. F. K., Poisson Processes,chap.9, Oxford: Clarendon Press, 1993.
letac86
Letac, G., Contemporary Mathematics **50** *263, (1986).*
mek91
Mekjian, A., Z., Phy.Review A **44** *8361 (1991).*
sachikov97
Sachikov, V. N., Probabilistic Methods in Combinatrial Analysis, New York:Cambridge University Press pp.81-82, 1997.
taka97
Takayasu, H., and Sato, A., Phy. Rev. Lett. **79** *966 (1997).*
watt76
Watterson, G. A., J. Appl. Prob. **13** *639. (1976)*
guess77
*Watterson, G. A., and Guess, H. A., Theor.Popul. Biol.***11** *141 (1977)*
zabell92
Zabell, S., Synthese **90** *205. (1992).*