A model for the size distribution of customer groups and businesses

Dafang Zheng[1,2,*], G. J. Rodgers[1] and P. M. Hui[3]

[1] *Department of Mathematical Sciences, Brunel University, Uxbridge, Middlesex UB8 3PH, UK*

[2] *Department of Applied Physics, South China University of Technology, Guangzhou 510641, P.R. China*

[3] *Department of Physics, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong*

Abstract

We present a generalization of the dynamical model of information transmission and herd behavior proposed by Eguíluz and Zimmermann. A characteristic size of group of agents s_0 is introduced. The fragmentation and coagulation rates of groups of agents are assumed to depend on the size of the group. We present results of numerical simulations and mean field analysis. It is found that the size distribution of groups of agents n_s exhibits two distinct scaling behavior depending on $s \leq s_0$ or $s > s_0$. For $s \leq s_0$, $n_s \sim s^{-(5/2+\delta)}$, while for $s > s_0$, $n_s \sim s^{-(5/2-\delta)}$, where δ is a model parameter representing the sensitivity of the fragmentation and coagulation rates to the size of the group. Our model thus gives a tunable exponent for the size distribution together with two scaling regimes separated by a characteristic size s_0. Suitably interpreted, our model can be used to represent the formation of groups of customers for certain products produced by manufacturers. This, in turn, leads to a distribution in the size of businesses. The characteristic size s_0, in

*e-mail: phdzheng@scut.edu.cn

this context, represents the size of a business for which the customer group becomes too large to be kept happy but too small for the business to become a brand name.

PACS Nos.: 05.65.+b, 87.23.Ge, 02.50.Le, 05.45.Tp

I. INTRODUCTION

Power law behavior of various kinds have been observed in a wide range of systems. Within the context of economics and finance, it has been found, for example, that the wealth of individuals [1], price-returns in stock markets [2,3], and company sizes in different countries all give non-trivial and interesting distributions [4–6]. By analyzing U.S. establishment and firm sizes, Nagel *et al.* [5] found that the size distributions n_s are power laws of the form $n_s \sim s^{-\tau}$ with $\tau = 2$ for firms with annual sales larger than some typical size. Income distribution of companies and size distribution of debts among bankrupt companies in Japan [7,8] also indicate a power law behavior with an exponent of about -2. However, studies on the size distribution of companies in different countries [6,9] indicated that the distributions as well as the exponents are different from country to country and hence are *not* universal.

Nagel *et al.* introduced several models of price formation using a two-dimensional spatial structure for information transmission in which consumers can only learn from their neighbors [5]. The formation of groups of consumers with shared opinion has direct impacts on the price-returns in stock markets [10,11] and on the size of businesses [5,12] as consumers of similar opinion may act collectively in making a transaction or a deal. Such herd formation and information transmission in a population have been a topic of active research in recent years. Recently, Eguíluz and Zimmermann [11] proposed and studied a simple model (henceforth referred to as the EZ model) of stochastic opinion cluster formation and information dispersal. It is a dynamical model in which there is a continual grouping and re-grouping of agents to form clusters of similar opinion. A group of agents act together in a transaction and the group dissolves after the transaction has occurred. When a group decides not to trade, it may combine with another group to form a larger group. Detailed numerical studies [11] and mean field analysis [13] revealed that the model could lead to a fat-tail distribution of price returns qualitatively similar to those observed in real markets [2,3]. The size distribution n_s of groups is shown to take on the form $n_s \sim s^{-5/2}$ for a range of group size s, followed by an exponential cutoff [11,13]. The model represents a dynamic generalization of the percolation-type model previously introduced by Cont and Bouchard [10] in which the behavior $n_s \sim s^{-5/2}$ is also observed. Interesting, it was recently found that the exponent characterizing this scaling behavior can actually be changed and made tunable by introducing a size-dependent dissociation and coagulation rates of groups [14] in the EZ model. The possibility of tuning the exponent makes the EZ model and its generalizations a good starting point for modelling herd formation and opinion sharing in different contexts.

In this paper, we propose and study an extension of the EZ model in which a characteristic size s_0 of group of agents (customers) is introduced. The rates of dissociation and coagulation of groups are assumed to take on different forms for $s \leq s_0$ and for $s > s_0$. The distribution in the size of groups of agents is found to exhibit two scaling regimes. Within the context of customers, the different behavior for $s \leq s_0$ and $s > s_0$ represent the situations in which a small group of customers may find themselves better served by businesses and customers tend to be happy when they find themselves to be part of a big group of customers. The distribution of groups of customers is unevitably related to the distribution of sizes of businesses that the customers support. In Sec.II, we introduce the model and present results of numerical simulations. Results of a mean field analysis are given in Sec.III and compared with numerical results. We summarize our results in Sec.IV.

II. THE MODEL

Our model is an extension of the EZ model [11], which, in turn, is a dynamical version of the static percolation model of Cont and Bouchaud [10]. Consider a system with a total of N agents (customers). These agents are organized into groups. They share the same information and hence they act collectively, i.e., customers belonging to a group decide whether to make a deal and what to do after a deal collectively. Initially, all the customers are isolated, i.e., each group has only one customer. At each time step, a customer i is selected at random. With probability a, all the customers belonging to the group of customer i

decide to make a deal, say, with a certain business or to buy a certain product from a manufacturer. After the deal, the group of customers will find themselves disappointed with probability $f(s_i)$. The group is then broken up into isolated customers and the corresponding business loses her customers. With probability $(1-f(s_i))$, the group of customers are pleased with the deal (service) and they remain in the group. With probability $1-a$, the chosen customer i and its group decide not to make a deal. In this case, another customer j is selected at random among the whole population of customers. With probability $c(s_i, s_j)$, the two groups of customers i and j coagulate to form a bigger group of customers, and with probability $(1-c(s_i, s_j))$ they remain separate. We study the size distribution of the groups of customers which is unevitably related to the size distribution of the businesses that the customers support.

In the present work, we take the fragmentation rate of group of customers to be

$$f(s_i) = \begin{cases} p(s_i), & s_i \leq s_0, \\ q(s_i), & s_i > s_0, \end{cases} \tag{1}$$

and the coagulation rate of groups of customers as

$$c(s_i, s_j) = \begin{cases} p(s_i)p(s_j), & s_i \leq s_0, s_j \leq s_0, \\ p(s_i)q(s_j), & s_i \leq s_0, s_j > s_0, \\ q(s_i)p(s_j), & s_i > s_0, s_j \leq s_0, \\ q(s_i)q(s_j), & s_i > s_0, s_j > s_0. \end{cases} \tag{2}$$

The functions $p(s)$ and $q(s)$ are taken to be increasing and decreasing power law functions:

$$p(s) = (\frac{s}{s_0})^\delta, \quad q(s) = (\frac{s_0}{s})^\delta, \tag{3}$$

where a characteristic group size of customers s_0 is introduced. From the standpoint of customers, a small group of customers may find themselves easier to be pleased by the service and the deal. Customers are also happier when they find themselves to belong to a big group of customers supporting the same business – a result of herd behavior. Groups of intermediate sizes near the characteristic size s_0 will be the hardest to please. When a group decides not to make a deal, members of the group tend to seek alternative opinion by combining with another group and there is a higher probability that groups with size around the characteristic size are combined. From the business point of view, businesses may find it easier to satisfy a small group of customers. Big businesses with a big group of customers enjoy the frame and customers are self-satisfied. Business with a characteristic size s_0 finds it difficult to satisfy its group of customers due to the diversity in demand. In the merging of businesses, it is more likely that businesses of size of the characteristic size merge so as to become a bigger business with possibly a better image to the customers.

The present model readily recovers previous extensions of the EZ model. In the EZ model [11], the fragmentation rate and the coagulation rate are simply taken to be $f(s_i) = c(s_i, s_j) \equiv 1$. The model leads to the size distribution $n_s \sim s^{-\tau}$ in the steady state with τ taking on a *robust* value of 5/2, together with an exponential cutoff setting in for large s [13]. Recently, it has been shown that if the fragmentation rate and the coagulation rate are taken to have a power law dependence on the size of the group(s) involved, i.e., $f(s_i) = s_i^{-\delta}$ and $c(s_i, s_j) = s_i^{-\delta} s_j^{-\delta}$, respectively, the exponent τ characterizing the distribution of group sizes becomes model-dependent and takes on the value $\tau = 5/2 - \delta$ [14].

Figure 1 shows the simulation results for the group size distribution n_s/n_1 as a function of s for a system with a total of $N = 10^4$ agents. The results are obtained within the time window between $t = 10^5 \sim 10^6$ steps, after the system approaches its steady state. Each data point represents an average taken over 32 independent runs. As scaling behavior is more conveniently observed for small value of a [11], we have taken $a = 0.01$. Figure 1(a) shows the results for different characteristic sizes of $s_0 = 1, 5, 20, 10^4$ with the parameter $\delta = 0.2$. For $s_0 = 1$, n_s has a power-law decay with an exponent of $\tau = 5/2 - \delta = 2.3$ throughout the whole range of s since only the $s > s_0$ regime exists. For $s_0 = N = 10^4$ corresponding to the largest possible characteristic size in the system, scaling is also found throughout the whole regime of s, but with a different value of the exponent of $\tau = 5/2 + \delta = 2.7$. For intermediate values of s_0, e.g., $s_0 = 5$ and $s_0 = 20$, two scaling regimes corresponding to the exponents 2.7 and 2.3 are found for $s < s_0$ and $s > s_0$, respectively. Figure 1(b) shows similar results for a larger value of $\delta = 0.5$ with $s_0 = 1, 5, 20, 10^4$. Crossover between the two scaling regimes are observed again at $s = s_0$ for the intermediate cases of $s_0 = 5$ and $s_0 = 20$. For $s_0 = N$, however, large groups of customers disappear and the scaling behavior no longer exist in the present case. This feature can be understood qualitatively as the coagulation rate is too small to allow large groups of customers to be formed [15].

III. MEAN FIELD ANALYSIS

The present model can be analyzed via a mean field theory [13,14]. Let $n_s(t)$ be the number of groups of agents with size s at time t. The master equations that govern $n_s(t)(s > 1)$ and $n_1(t)$ are

$$N\frac{\partial n_s}{\partial t} = -af(s)sn_s + \frac{(1-a)}{N}\sum_{r=1}^{s-1} c(r, s-r)rn_r(s-r)n_{s-r} - \frac{2(1-a)sn_s}{N}\sum_{r=1}^{\infty} c(s,r)rn_r, \quad (4)$$

and

$$N\frac{\partial n_1}{\partial t} = a\sum_{r=2}^{\infty} f(r)r^2 n_r - \frac{2(1-a)n_1}{N}\sum_{r=1}^{\infty} c(1,r)rn_r, \quad (5)$$

respectively. Here $f(i)$ and $c(i,j)$ are the fragmentation and coagulation rates defined in Eqs.(1) and Eq.(2). The first term on the right hand side of Eq.(4) describes the fragmentation of a group of customers of size s. The second term describes the coagulation of two groups to form a group of size s. The last term is the coagulation of a group of size s with another group. In Eq.(5), the first term on the right hand side describes the fragmentation of bigger groups into isolated customers and the second term represents the coagulation of a group of size unity with another group. In the steady state, $\frac{\partial n_s}{\partial t} = 0$. We have

$$n_s = \frac{1-a}{Nasf(s) + 2(1-a)s\sum_{r=1}^{\infty} c(s,r)rn_r}\sum_{r=1}^{s-1} c(r, s-r)r(s-r)n_r n_{s-r} \quad (6)$$

for $s > 1$, and

$$n_1 = \frac{Na}{2(1-a)\sum_{r=1}^{\infty} c(1,r)rn_r}\sum_{r=2}^{\infty} f(r)r^2 n_r. \quad (7)$$

The steady state equations (6) and (7) can be solved by using the generating function technique [14]. Following standard procedures [14,15], n_s can be solved to give the scaling behavior in the limit of $a \approx 0$ to get

$$n_s \sim N s^{-(\frac{5}{2}+\delta)} \tag{8}$$

for $s \leq s_0$, and

$$n_s \sim N s^{-(\frac{5}{2}-\delta)} \tag{9}$$

for $s > s_0$. For finite value of a, the scaling behavior is masked by an exponential cutoff. Equations (8) and (9) indicate that there are two scaling behavior for the size distribution of groups of customers. The exponents characterizing the power law decay of the size distribution of groups of customers are $\tau = 5/2 + \delta$ for $s \leq s_0$ and $\tau = 5/2 - \delta$ for $s > s_0$, respectively. These results are in agreement with the numerical results given in Fig.1.

IV. SUMMARY

In summary, we have presented an extension of the EZ model for opinion sharing and herd formation in a population of agents in which there is a characteristic size s_0. The model can be treated analytically via a mean field approach. Both results from numerical simulation and mean field theory indicate that the size distribution of groups of agents n_s exhibits two distinct scaling behavior depending on $s \leq s_0$ or $s > s_0$. For $s \leq s_0$, $n_s \sim s^{-(5/2+\delta)}$, while for $s > s_0$, $n_s \sim s^{-(5/2-\delta)}$. Here δ is a model parameter and hence the exponents characterizing the scaling behavior are tunable. Suitably interpreted, our model can be used to represent the formation of groups of customers for certain products produced by manufacturers. This, in turn, leads to the size distribution of businesses. Within this context, the characteristic size s_0 represents the size of a business for which the customer group becomes too "large" to be pleased but too "small" for the business to become a brand name. With its tunable exponent, the model can be used to fit to empirical data as it has been shown that markets in different countries are characterized by non-universal exponents [6]. Recently, it has also been shown empirically [5] that the distributions of U.S. establishment sizes and firm sizes in the retail sector are characterized by power laws with an exponent of -2. Within our model, such a behavior can be obtained for $\delta = 0.5$ and $s_0 = 1$. The introduction of the two parameters δ and s_0 thus makes the present model highly flexible and hence can be used to model a wide variety of phenomena in which herd formation and information transmission are important.

ACKNOWLEDGMENTS

DFZ, GJR and PMH acknowledge financial support from The China Scholarship Council, The Leverhulme Trust and the Research Grants Council of the Hong Kong SAR Government under grant CUHK4241/01P, respectively.

REFERENCES

[1] V. Pareto, *Cours d'économie politique* (Lausanne, 1897).

[2] R.N. Mantegna and H.E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, 2000).

[3] P. Gopikrishnan, V. Plerou, L.A.N. Amaral, M. Meyer, H.E. Stanley, Phys. Rev. E **60**, 5305 (1999).

[4] M.H.R. Stanley *et al.*, Econ. Lett. **49**, 453 (1995); L.A.N. Amaral *et al.*, J. Phys. I France **7**, 621 (1997).

[5] K. Nagel, M. Shubik, M. Paczuski and P. Bak, Physica A **287**, 546 (2000).

[6] J.J. Ramsden and Gy. Kiss-Haypál, Physica A **277**, 220 (2000).

[7] K. Okuyama, M. Takayasu and H. Takayasu, Physica A **269**, 125 (1999).

[8] H. Aoyama, W. Souma, Y. Nagahara, M.P. Okazaki, H. Takayasu and M. Takayasu, Fractals **8**, 293 (2000).

[9] H. Takayasu and K. Okuyama, Fractals **6**, 67 (1998).

[10] R. Cont and J.-P. Bouchaud, Macroeconom. Dyn. **4**, 170 (2000).

[11] V.M. Eguíluz and M.G. Zimmermann, Phys. Rev. Lett. **85**, 5659 (2000).

[12] R. D'Hulst and G.J. Rodgers, Euro. Phys. J. B **21**, 447 (2001).

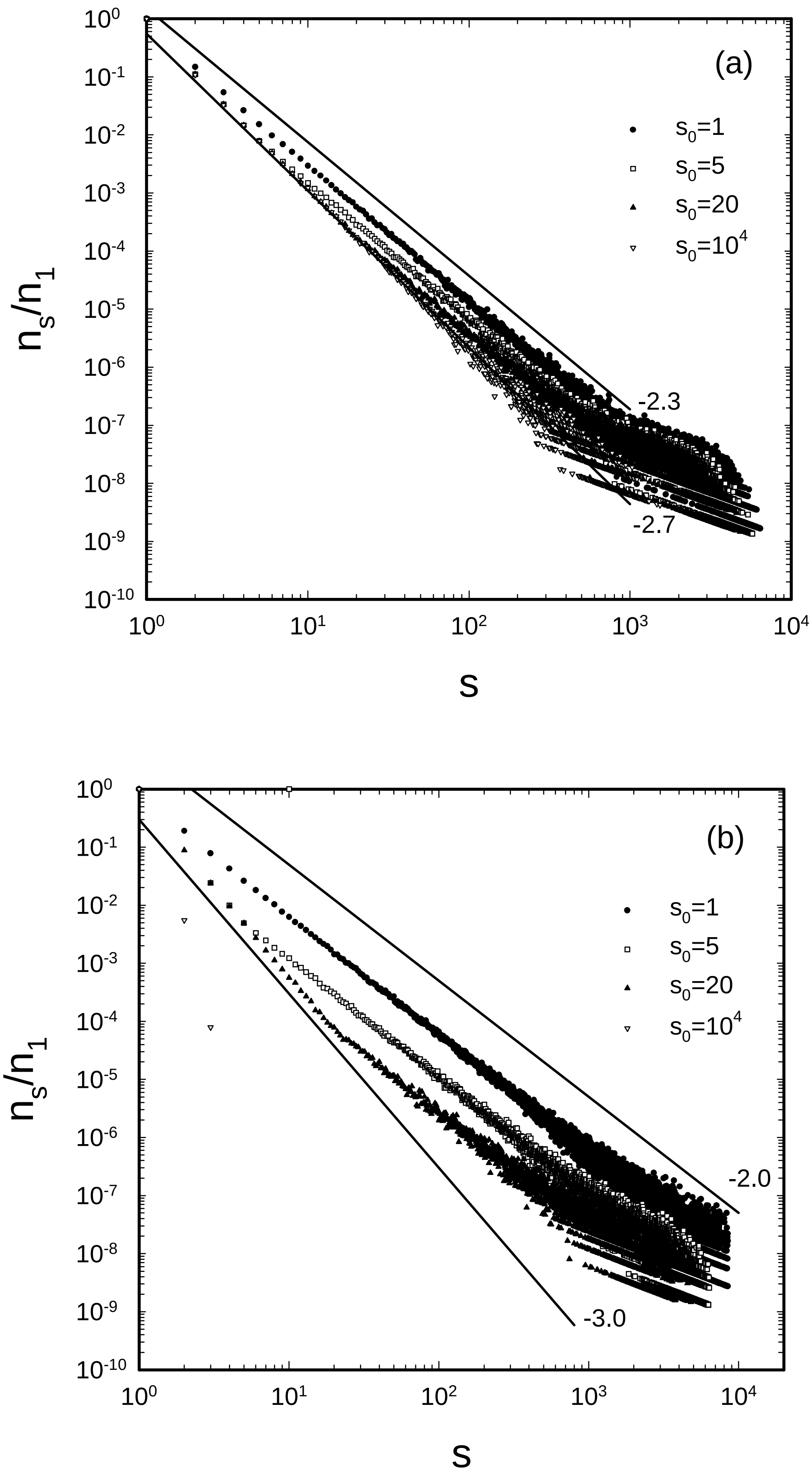
[13] R. D'Hulst and G.J. Rodgers, Int. J. Theor. Appl. Finance **3**, 609 (2000).

[14] D. Zheng, P.M. Hui and N.F. Johnson, cond-mat/0105474.

[15] D. Zheng, G.J. Rodgers, P.M. Hui and R. D'Hulst, Physica A **303**, 176 (2002).

FIGURE CAPTIONS

Fig. 1. The size distribution of groups of customers n_s/n_1 as a function of size s on a log-log scale for (a)$\delta = 0.2$ and (b)$\delta = 0.5$. The values of the characteristic size s_0 chosen in the calculations are: $s_0 = 1, 5, 20, 10^4$. The solid lines are a guide to the eye.



(a)
$s_0=1$
$s_0=5$
$s_0=20$
$s_0=10^4$
-2.3
-2.7
n_s/n_1
s
10^0 10^{-1} 10^{-2} 10^{-3} 10^{-4} 10^{-5} 10^{-6} 10^{-7} 10^{-8} 10^{-9} 10^{-10}
10^0 10^1 10^2 10^3 10^4
(b)
$s_0=1$
$s_0=5$
$s_0=20$
$s_0=10^4$
-2.0
-3.0
n_s/n_1
s
10^0 10^{-1} 10^{-2} 10^{-3} 10^{-4} 10^{-5} 10^{-6} 10^{-7} 10^{-8} 10^{-9} 10^{-10}
10^0 10^1 10^2 10^3 10^4

Figure 1