Computational Modeling of Collective Human Behavior: The Example of Financial Markets

Andy Kirou[1], Błażej Ruszczycki[1], Markus Walser[2], and Neil F. Johnson[1]

[1] Department of Physics, University of Miami,
P.O. Box 248046, Coral Gables, FL 33124 USA
[2] Landesbank Baden-Württemberg, Am Hauptbahnhof 2, 70173 Stuttgart, Germany
A.Kirou@umiami.edu

Abstract. As a result of the increased availability of higher precision spatiotemporal datasets, coupled with the realization that most real-world human systems are complex, a new field of computational modeling is emerging in which the goal is to develop minimal models of collective human behavior which are consistent with the observed real-world dynamics in a wide range of systems. For example, in the field of finance, the fluctuations across a wide range of markets are known to exhibit certain generic stylized facts such as a non-Gaussian 'fat-tailed' distribution of price returns. In this paper, we illustrate how such minimal models can be constructed by bridging the gap between two existing, but incomplete, market models: a model in which a population of virtual traders make decisions based on common global information but lack local information from their social network, and a model in which the traders form a dynamically evolving social network but lack any decision-making based on global information. We show that a combination of these two models – in other words, a population of virtual traders with access to both global and local information – produces results for the price return distribution which are closer to the reported stylized facts. Going further, we believe that this type of model can be applied across a wide range of systems in which collective human activity is observed.

Keywords: complex systems, socio-economic systems, virtual traders, collective behavior.

1 Introduction

Socio-economic systems were traditionally treated from the points of view of game theory or traditional economic theory. These approaches – while undoubtedly successful in terms of gaining insight into core features – are unable to address the issue of how and why such systems produce the fluctuating external signals that they do [1,2,3,4,5]. Human systems as diverse as traffic, Internet downloads, and financial markets, are all known to produce large-scale fluctuations – for example, in the number of cars taking a certain road, or the number of people accessing a certain website, or the number of people trying to sell a stock at certain times [1,2,3,4,5]. In previous decades, there was typically an insufficient amount of reliable data available for researchers to address such problems

of dynamics. Nowadays, with the increase in online logging of data – from social, governmental and commercial sectors – this area of modeling now becomes very attractive. However few advances are likely to be made analytically, since any meaningful explanation of the dynamics must be related back to what the collection of individual objects are doing. In other words, it is what physicists call a many-body problem – one in which the objects are subjected to endogenous and exogenous feedback and nonlinear interactions – and it is known that such many-body problems are in general intractable. Given the additional feature that the objects themselves may be semi-autonomous (i.e. they each have some form of independent decision-making ability such that a given external input may yield various possible outputs depending on some internal state of the object itself) the only realistic route toward advancing our understanding of such systems must surely be driven by computational modeling and simulation. In short, such socio-economic systems are complex – and the key to understanding the dynamics of such complex systems is provided by computer simulation.

Among the wide range of socio-economic fields that have attracted the interest of complexity scientists, is the goal of trying to understand financial market dynamics [1,2,3,4,5]. This is partly because financial markets produce such high frequency data, and partly because the possible actions of an individual trader are quite simple: buy, sell or do nothing at any timestep. A wide range of interesting stylized facts have emerged based on analysis of the real market data over a wide range of timescales – from seconds through to days, weeks and months. In particular, it has been found that the distribution of price returns (i.e. changes in price between a given time t and a time $t + \Delta t$ later) do not follow the simple distribution expected from a random walk. The standard model of a financial market – based on the efficient market hypothesis – is that price-changes are like the toss of a coin. They are supposedly independent – hence if we assume that each trader trades according to the toss of a coin, then the probability of buying and selling would a priori be $\frac{1}{2}$ if we ignore the 'do nothing' option. Counting a head as $+1$ in terms of price change, and a tail as -1, the probability distribution for having a given price-change $\Delta P = N_{\text{buy}} - N_{\text{sell}}$ is simply the probability of obtaining N_{buy} heads and N_{sell} tails from N coin-tosses, such that $N = N_{\text{buy}} + N_{\text{sell}}$. This is a binomial expression, which then approaches a Gaussian in the large-N limit. We note in passing that it is more common in the finance literature to consider the logarithm of price-changes rather than the price-change itself, hence the distribution of returns in the ideal, random-walk market is log-normal. However the two distributions are essentially indistinguishable in most financial markets, since the returns are typically much smaller than the prices themselves.

2   Collective Non-randomness in the Human World

As a rough first approximation, financial market behavior is not far from the Gaussian model. However, many independent detailed empirical studies of financial market returns have confirmed that major deviations begin to arise

in the tails of the distribution [1,2,3,5]. Specifically, the distribution of price-changes ΔP deviates from Gaussian behavior even at moderate values of ΔP. In particular, the probability of intermediate-to-large price-changes is larger than the random coin-toss model would suggest. This leads to the so-called 'fat tail' terminology which is often used to describe real financial markets when they 'misbehave'. In fact, such fat-tailed behavior is common across a wide range of socio-economic domains [1]. In many scientific settings such as physics, such deviations from Gaussian behavior in the tail of the distribution might be considered a mere detail, since the average behavior of the system is typically what counts. However real-world finance is all about risk – in particular, the risk of large unexpected price-changes – and so these deviations are actually the most important features of the distribution, contributing to abnormally high estimates of the moments of the distribution such as the variance.

The fact that large price-changes are more likely than expected – and in more general socio-economic settings, that large traffic jams or heavy Internet downloads are more likely than expected – suggests that the population is unintentionally behaving in a coordinated way. It is as though the supposedly independent coin-tosses of the N traders are not in fact independent: when one comes up heads, they are all more likely to come up heads, and vice versa with tails. It is as though the population of traders was inter-connected in some way. The fact that getting the best price in a financial market is a competitive activity – in the same way that managing to grab space on a busy road, or a download on a busy website, are also competitive – means that such coordination is very unlikely to have arisen through some intentional population-level decision making. There is no central controller – and even if there were, the fact that individual objects are competing to win means that no central controller would necessarily be listened to or followed by individual members of the population.

This coordination observed in many scenarios where populations of humans are competing for some limited resource, thereby leading to larger than random probability for large events, is characteristic of many human systems [1,4]. But if so, what causes it and how can we provide a quantitative model of it? Because of the generic nature of the fat-tail statistics, any useful model should not depend on the details of the particular market or type of trader, or road or type of car, or website or type of computer. Instead it must be some fairly general feature of collective human activity.

The fundamental question as to what underlying model might best represent such collective coordination, has inspired a new breed of computer-based scientific investigation involving physical, biological and social scientists. At its root, a system such as a financial market, traffic system, or the Internet, involves agents (i.e. people) deciding between a few options (e.g. buy, sell, do nothing) based on some limited information – which may be global or locally generated – and then competing with the remaining agents for the available resource or reward. Any collective coordination will require some form of trader inter-connectedness. One way in which this could have arisen, is if subsets of the traders form social groups such that they and their immediate friends or associates, coordinate

their actions (i.e. bias their decisions and hence effectively connect their coins during the coin-toss). In fast-moving financial markets, such groups are likely to change fairly rapidly, and should at least be accounted for using dynamical models of such group formation. This idea has led to a particular class of models based on dynamical cluster formation. Notable examples include the dynamical clustering model of Eguíluz and Zimmermann [7] and of others in biology [6], [8]. A second way in which coordination could have arisen, concerns how agents react to a particular piece of common information. Models of this form, of which a notable example is the so-called Grand Canonical Minority Game [9], feature agents whose actions are dictated by a strategy (or set of strategies). In the first case of real clusters, the grouping is intentional, while in the second case agents form unintentional groups (i.e. crowds) as a result of using the same strategy at the same time and hence acting identically over a short period of time. These two classes of model are complementary. Each of these models is 'minimal' in the sense that they are the simplest known examples which seem to capture the essential ingredients of clustering and decision-making respectively. To date, the two classes of model have been studied separately – however, they should clearly both be combined in order to understand the interplay of local and global information on collective group formation and hence the collective dynamics.

In this paper, we take the first steps in this direction of analyzing a collective human system in which there are *local* interactions as observed in the Eguíluz and Zimmermann [7] model (i.e. E-Z model) and *global* interactions as observed in the Grand Canonical Minority Game [9] (i.e. GCMG). The focus of this paper is to show how adding global interactions to the E-Z model, does indeed improve the fit with the known empirical distributions of financial market returns. Our construction is a modified E-Z model in which the agents are randomly assigned the trading strategies and apply their strategy based on the last two price movements. Unlike the full GCMG, the model does not include an explicit strategy score. Instead, the system behaves as a relative majority voting system.

3   Relative Majority Vote

The relative majority vote system is specified by two parameters, the total number of agents N and the consensus parameter x. To set up a timescale, we need to prescribe some form of timescale constant. At each time-step, an agent is chosen at random and the group to which the agent belongs is identified. The size of this group is denoted as s. Once a particular group has been identified, it makes a decision in the following way: Each agent votes either to *sell*, *buy* or *wait* (see Fig. 1) with approximate probability $1/3$ for each of the listed options. If the number of votes for the most popular decision exceeds the threshold, which is defined as

$$\mathcal{T} \equiv x \cdot s \tag{1}$$

the consensus is reached and the decision is performed. Otherwise there is no consensus and the group fragments into single individuals. The group that decides to wait does not trade, but instead merges with another group by choosing
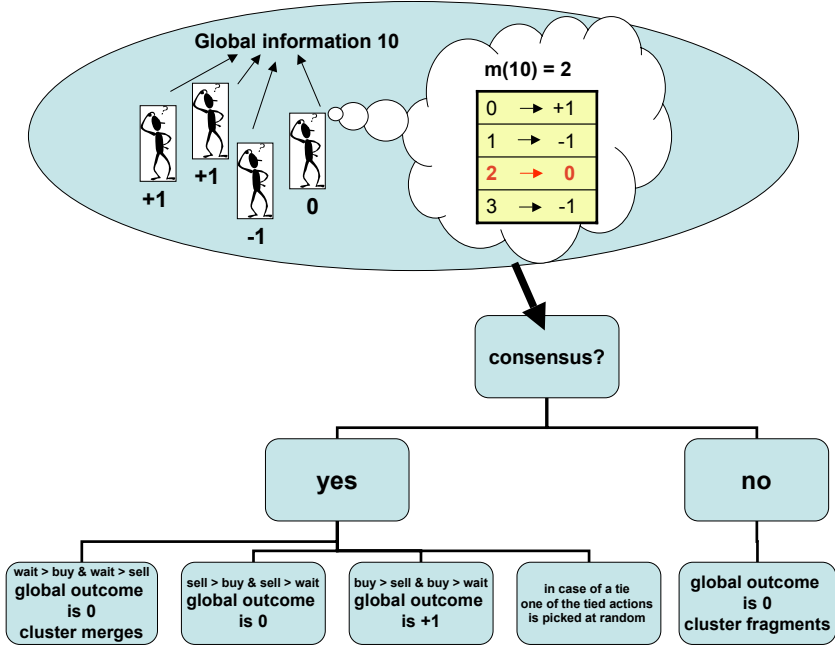
Fig. 1. Schematic diagram of the decision process in our multi-agent model

randomly any other agent from the entire population and then joining with the group to which this agent belongs. For a particular group of size s, let us denote the number of agents who vote to sell, buy or wait (and merge with another group) as $\mathcal{S}, \mathcal{B}, \mathcal{W}$ respectively. The conditions for the group decision are as follows:

$$
\begin{array}{ll}
\text{Fragments:} & (\mathcal{W} < \mathcal{T}) \wedge (\mathcal{S} < \mathcal{T}) \wedge (\mathcal{B} < \mathcal{T}) \\
\text{Buys:} & (\mathcal{B} \geq \mathcal{T}) \wedge (\mathcal{B} > \mathcal{S}) \wedge (\mathcal{B} > \mathcal{W}) \\
\text{Sells:} & (\mathcal{S} \geq \mathcal{T}) \wedge (\mathcal{S} > \mathcal{B}) \wedge (\mathcal{S} > \mathcal{W}) \\
\text{Merges:} & (\mathcal{W} \geq \mathcal{T}) \wedge (\mathcal{W} > \mathcal{B}) \wedge (\mathcal{W} > \mathcal{S})
\end{array}
\tag{2}
$$

We also need to account for the fact we may have a tied number of votes. This is resolved by randomly picking one of the two tied decisions, e.g. if

$$(\mathcal{S} \geq \mathcal{T}) \wedge (\mathcal{S} = \mathcal{B}) \wedge (\mathcal{S} > \mathcal{W})$$

then the group either sells or buys with equal probability. The decisions presented in Eq. (2) are exclusive, therefore the corresponding conditional probabilities (on the condition that the particular group is chosen) satisfy the equation

$$\hat{p}_{\text{frg}} + \hat{p}_{\text{sell}} + \hat{p}_{\text{buy}} + \hat{p}_{\text{merge}} = 1 \,, \tag{3}$$

where $\hat{p}_{\text{frg}}$ is the probability that the group fragments. Note that the above conditional probabilities depend on the group size s. From the symmetry of Eqs. (2) and (3) we see that it is sufficient to know $\hat{p}_{\text{frg}}$ since
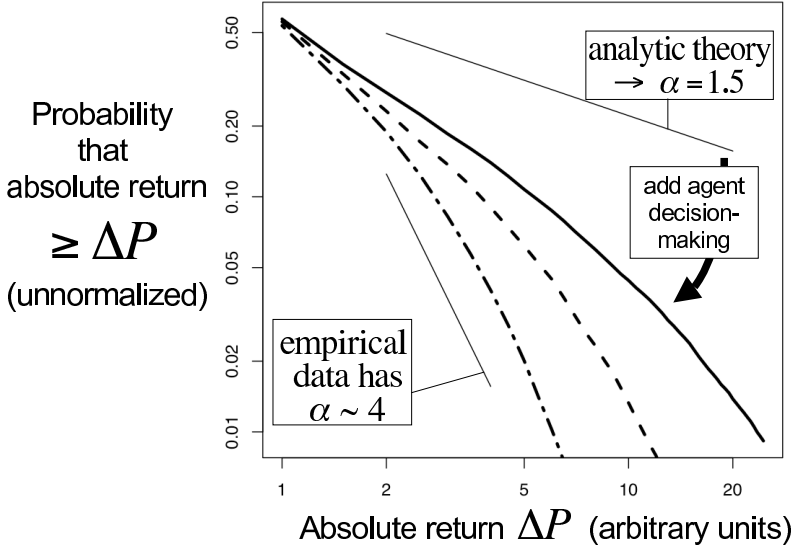
Fig. 2. Cumulative distribution of price returns ΔP in arbitrary units, for the consensus parameter $x = 37\%$ (thicker solid line), $x = 41\%$ (dashed line) and $x = 47\%$ (dashed-dotted line). The number of agents is 10000. The thin solid lines indicate the scaling behavior observed for the cumulative probability distribution for returns ΔP in the pure E-Z model and for typical empirical market data [2]. The underlying probability distribution for price-changes $p(\Delta P)$, follows a power-law $p(\Delta P) \sim (\Delta P)^{-\alpha}$ over a range of ΔP values in both the E-Z and real market data, with $\alpha = 1.5$ for the pure E-Z model and $\alpha \sim 4$ for empirical market data.

$$\hat{p}_{\text{sell}} = \hat{p}_{\text{buy}} = \hat{p}_{\text{merge}} = \frac{1 - \hat{p}_{\text{frg}}}{3} \, . \tag{4}$$

We calculate the combinatorial expression as

$$\hat{p}_{\text{frg}}(s) = \frac{s!}{3^s} \sum_{\mathcal{W}=0}^{s-1} \sum_{\mathcal{B}=s-\mathcal{T}-\mathcal{W}}^{\min(\mathcal{T}-1, s-\mathcal{W})} \frac{1}{\mathcal{W}! \mathcal{B}! (s - \mathcal{B} - \mathcal{W})!} \tag{5}$$

The system may be described by mean field theory (disregarding the fluctuations and finite size effects). We denote by n_s the average number of groups of size s. For the steady state, the set of master equations is semi-recursive[1] and is written as

$$-\frac{s}{N} \hat{p}_{\text{frg}}(s)(1 - \delta_{s1})n_s - \left[\frac{1}{N} \hat{p}_{\text{merge}}(s) + \frac{1}{N^2} \sum_{s'=1}^{N} s' \hat{p}_{\text{merge}}(s') \right] s\, n_s$$

$$+ \sum_{s'=1}^{s-1} s'\, n_{s'} (s - s')n_{s'} \hat{p}_{\text{merge}} s' + \delta_{s,1} \sum_{s'=s+1}^{N} \hat{p}_{\text{frg}}(s') s'^2 n_{s'} = 0 \, . \tag{6}$$

---

[1] The equation for n_l depends on n_i for all $i = 1 \ldots l - 1$.

The above set may be solved numerically[2]. In order to include corrections from the fluctuations and the finite size effect, we need to perform direct simulations of the system described by Eq. (6).

We are interested in the situation when the consensus parameter is within the range of $x = (33\,{}^1\!/_3\,\%, 50\,\%)$. If $x < 33\,{}^1\!/_3\,\%$ there is no fragmentation. If $x > 50\,\%$ we have the absolute majority vote condition. When $x \rightarrow 33\,{}^1\!/_3\,\%$ from above, we expect the return distribution to approach the one for the original Eguíluz-Zimmermann system, which shows a power law with exponent 1.5 over a large scale of return sizes. As we increase x slightly above $33\,{}^1\!/_3\,\%$, the dominant behavior of the buy/sell probabilty $\hat{p}_{\text{buy}} = \hat{p}_{\text{sell}}$ for the large groups (i.e. of order $s \gtrsim 100$) yields an exponential cut-off, while for smaller s we have the finite size effect. This modifies the model in two ways, by changing the group distribution (since the conditional probabilities enter Eq. (6)) and the trade mechanism. Unlike the E-Z system [7] where mostly the large groups trade, we expect to have the trades coming from the actions of the small groups with the exponential cut-off due to the behavior of Eq. (5).

4   Simulation Results

The simulations were performed for a system with $N = 10^4$ agents, $m = 2$, and 10^6 time steps, with three different values of the consensus parameter. The initial state of information was $(1, 1)$. After 10^5 timesteps, in order to allow the system to reach equilibrium, the returns where computed as follows: if a cluster of size s decides to buy, the return is $+s$. If a cluster of size s decides to sell, the return is $-s$. After the simulation was complete, the time was rescaled by adding the returns of two consecutive timesteps since on average a transaction occurred once every two timesteps. Thus the results in Fig. 2 are effectively for $9 \times 10^5/2$ timesteps. It is observed that indeed most of the trades come from the action of the small groups. As the consensus parameter is increased, the distribution of returns can be seen to fall more sharply due to the increasing dominance of the exponential cut-off. Our results demonstrate that the feature of allowing agents access to global information and subsequent decision-making, when built into a model focused on local group formation (i.e. E-Z), leads to a hybrid model which can better capture features of the known empirical distributions. In short, both local group formation *and* global information are important when building a minimal computational model of financial markets. By extension, the same statement should hold for collective human activity in *any* domain in which competition exists between a collection of interconnected agents.

5   Discussion

We have proposed a simple model system that represents a first step in the quest to develop minimal, individual-based computational models of real

---

[2] The numerical procedure for solving Eq. 6 is effective for modest values of N, since at least ${}^1\!/_2\,N^2$ iteration steps are required.

socio-economic systems in which both local and global interactions are featured. Such minimal models aim to incorporate the minimum number of rules, and hence parameters, such that individuals' behavior and interactions still appear credible; yet at the same time, the emergent dynamics should remain consistent with the maximum possible number of empirical stylized facts based on real-world data. In our particular case, we have incorporated global interactions via the heterogeneity of strategies held by the agents, as well as agent memory in the locally interacting system via the grouping mechanism.

Our specific results are as follows. The scenario in which the agents are allowed to vote introduces an exponential cut-off starting on the scale for which effects connected to the discrete nature of the system may be neglected. Our results show that those who usually trade are the small groups, and that there are no trades coming from the large groups. By contrast in the original E-Z model, the conditional probabilities were constant and any particular large group trades more often than a particular small group. The most realistic minimal model (which is as yet undiscovered) should lie somewhere in between. Any voting scenario is a Poisson process which introduces an exponential cut-off into the system. This exponential cut-off appears on a scale where the number of individuals involved is sufficiently large that we may disregard the discrete nature of the system. The modeling and the computational challenge is therefore as follows: How can the present model be further enhanced such that it reflects the more complex behavior of the individuals through the possession of memory, behavior based on past experience, and the passing of information between groups concerning whether to trade or not?

In terms of more general issues of computational modeling, we have tried to highlight the need to develop minimal computational models of real socio-economic systems through individual-based behavior. Future theoretical developments in such fields lie beyond simply integrating some form of phenomenological equation. Moreover, this sort of socio-economic modeling is an application of computation that is set to boom in the future given the growing availability of high-frequency data from socio-economic systems – and the fundamental philosophical need for theories which treat *dynamical fluctuations* in addition to mean behavior. One particular example in which this philosophy is now being developed, is in improving our understanding of human conflict – by looking at the stylized facts of conflict dynamics in exactly the same way as has been done for financial markets. Indeed, we have recently shown that remarkably similar minimal computational models can be built, exhibiting equally satisfying agreement with empirical data, simply by combining together global and local interactions among agents. This work on human conflict will be discussed in more detail elsewhere.

References

1. See for example, the wide range of publications and conferences around this common theme of computational modeling of socio-economic systems,
   `http://www.unifr.ch/econophysics`

2. Bouchaud, J.-P., Potters, M.: Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management, 2nd edn. Cambridge University Press, Cambridge (2004)
3. Mantegna, R.N., Stanley, H.E.: An Introduction to Econophysics: Correlations and Complexity in Finance. Cambridge University Press, Cambridge (1999)
4. Johnson, N.F.: Two's company, three is complexity. Oneworld, New York (2007)
5. Johnson, N.F., Jefferies, P., Hui, P.M.: Financial Market Complexity. Oxford University Press, Oxford (2003)
6. Gueron, S., Levin, S.A.: The Dynamics of Group Formation. Mathematical Biosciences 128, 243–246 (1995)
7. Eguíluz, V.M., Zimmermann, M.G.: Transmission of Information and Herd Behaviour: An Application to Financial Markets. Phys. Rev. Lett. 85, 5659–5662 (2000)
8. Cont, R., Bouchaud, J.-P.: Herd Behavior and Aggregate Fluctuations in Financial Markets. Macroeconomic Dynamics 4, 170–196 (2000)
9. Challet, D., Zhang, Y.C.: Emergence of Cooperation and Organization in an Evolutionary Game. Physica A 246, 407–418 (1997)