Predictability of large future changes in a competitive evolving population

D. Lamper and S. Howison

Oxford Centre for Industrial and Applied Mathematics,

Oxford University, Oxford, OX1 3LB, U.K.

N. F. Johnson

Physics Department, Oxford University, Oxford, OX1 3PU, U.K.

(May 14, 2001)

Abstract

The dynamical evolution of many economic, sociological, biological and physical systems tends to be dominated by a relatively small number of unexpected, large changes ('extreme events'). We study the large, internal changes produced in a generic multi-agent population competing for a limited resource, and find that the level of predictability actually *increases* prior to a large change. These large changes hence arise as a predictable consequence of information encoded in the system's global state.

01.75.+m, 02.50.Le, 05.40.+j, 87.23.Ge

Populations comprising many 'agents' (e.g. people, species, data-packets, cells) who compete for a limited resource, are believed to underlie the complex dynamics observed in areas as diverse as economics [1–4], sociology [5], internet traffic [6], ecology [7] and biology [8,9]. The reliable prediction of large future changes ('extreme events') in such complex systems would be of enormous practical importance, but is widely considered to be impossible [10].

In this paper, we examine the predictability of large future changes produced within an evolving population of agents who compete for a limited resource. We find that the level of predictability in the system actually *increases* prior to a large change. The implication is that such a large change arises as a predictable consequence of information encoded in the system's global state, as opposed to being triggered by some isolated random event.

We consider a generic multi-agent system comprising a population of N_{tot} agents where only a maximum of $L < N_{tot}$ agents can be winners at each timestep; an everyday example would be a popular bar with a limited seating capacity L [5]. For the purpose of this paper, we consider a specific case of such a limited-resource problem with $L = (N_{tot} - 1)/2$ with N_{tot} being odd [11], hence there are more losers than winners. We note that similar dynamics can also occur for more general L values [12]. Each agent is therefore seeking to be in the minority group: for example, a buyer in a financial market may obtain a better price if more people are selling than buying; a driver may have a quicker journey if she chooses the route with less traffic. At each timestep, an agent decides whether to enter a game where the choices are option 0 (e.g. buy, choose route A) and option 1 (e.g. sell, choose route B). Each agent holds a finite number of strategies and only a subset $N = N_0 + N_1 \leq N_{tot}$ of the population, who are sufficiently confident of winning, actually play: N_0 agents choose 0 while N_1 choose 1. If $N_0 - N_1 > 0$, the winning decision (outcome) is '1' and vice versa. If $N_0 = N_1$ the tie is decided by a coin-toss. Hence N and the 'excess demand' $N_{0-1} = N_0 - N_1$ both fluctuate with time. In contrast to the basic Minority Game (MG) [11], this variable-N model has the realistic feature of accounting for agents' confidence [13,14]. Furthermore the variable-N model can be used to generate statistical and dynamical features similar to those observed in financial markets (archetypal examples of complex systems) [13,2]. Therefore, demonstration of predictability of extreme events in the present multi-agent model would open up the exciting possibility of predictability of extreme events in real-world systems.

The only global information available to the agents is a common bit-string 'memory' of the m most recent outcomes. The agents can thus be said to exhibit 'bounded rationality' [5]. Consider $m = 2$; the $2^m = 4$ possible history bit-strings are $00, 01, 10$ and 11. A strategy consists of a response, i.e. 0 or 1, to each possible bit-string; hence there are $2^{2^m} = 16$ possible strategies. At the beginning of the game, each agent randomly picks q strategies and after each turn assigns one (virtual) point to a strategy which would have predicted the correct outcome. Agents have a time horizon T over which virtual points are collected and a threshold probability level τ; strategies with a probability of winning greater than or equal to τ, i.e. having $\geq T\tau$ virtual points, are available to be used by the agent. We call these *active* strategies. Agents with no active strategies within their individual set of q strategies do not play at that timestep. Agents with one or more active strategies play the one with the highest virtual point score; any ties between active strategies are resolved using a coin-toss. The 'excess demand' N_{0-1}, which can be identified as the output from the model system, can be expressed as

$$N_{0-1} = \sum_i \left\{ 1 - 2x_i s_i \right\} \tag{1}$$

where s_i is the prediction of the i-th strategy, e.g. 0 or 1, and x_i is the number of agents using this strategy, with the summation taken over the set of active strategies at that timestep.

Because of the feedback in the game, any particular strategy's success is short-lived. If all the agents begin to use similar strategies, and hence make the same decision, such a strategy ceases to be profitable. The game can be broadly classified into three regimes. (i) The number of strategies in play is much greater than the total available: groups of traders will play using the same strategy and therefore crowds should dominate the game [15]. (ii) The number of strategies in play is much less than the total available: grouping behaviour is therefore minimal. (iii) The number of strategies in play is comparable to the total number available: this represents a transition regime and is of most interest, since it produces seemingly random dynamics with occasional large movements. Even if complete knowledge of the state of the game were available at any timestep, it seems impossible that subsequent outcomes should be predictable with significant accuracy since the coin-tosses which are used to resolve ties in decisions (i.e. $N_0{=}N_1$) and active-strategy scores inject stochasticity into the game's time-evolution. Remarkably, however, we find that *large* changes over several consecutive timesteps can be predicted with surprising accuracy *without* any detailed knowledge of the game itself.

Suppose we are given an analogue time series $H(t)$ generated by a physical, sociological, biological or economic system, e.g. a financial market [13], whose dynamics are well-described by the multi-agent game for a fixed *unknown* parameter set m, N, τ, T and an *unknown* specific realization of initial strategy choices. We call this our 'black-box' game. Our goal is to identify 'third-party' games which can be used to predict large future changes in $H(t)$, where $\Delta H(t)$ is defined to be directly proportional to the excess demand N_{0-1}. For example, $\Delta H(t)$ could be the price change in a financial market, or may instead be a quantity which is derived from the system output using a known non-linear function. For the remainder of this article, we focus on the following game parameters for the 'black-box' game: $N = 101$, $m = 3$, $q = 2$, $T = 100$, $\tau = 0.53$, although our conclusions are more general [16]. Since $\tau > 0.5$ an agent will not participate unless she believes she has a better than average chance of winning.

We start by running $H(t)$ through a trial third-party game in order to generate an estimate of S_0 and S_1 at each timestep, the number of active strategies predicting a 0 or 1 respectively. This is obtained from the strategy space, or the pool of all available strategies in the third party game, and is independent of the distribution of agents. We wish to predict $\Delta H(t)$, i.e. N_{0-1} ; we will do this by linking S and N through an appropriate probability distribution. Provided the strategy space in the black-box game is reasonably well covered by the agent's random choice of initial strategies, any bias towards a particular outcome in the active strategy set will propagate itself as a bias in the value of N_{0-1} away from zero. Thus we expect N_{0-1} to be approximately proportional to $S_0 - S_1 = S_{0-1}$. This is equivalent to assuming an equal weighting x_i on each strategy in Eq. (1), indicating the exact distribution of strategies among the individual agents is unimportant in this regime [17]. In addition, the number of agents taking part in the game at each timestep will be related to the total number of active strategies $S_0 + S_1 = S_{0+1}$, hence the error (i.e. variance) in the prediction of N_{0-1} using S_{0-1} will be dependent on S_{0+1}. Based on extensive statistical analysis of known simulations for the multi-agent game [16], we have confirmed that it is reasonable to model the relationship by

$$N_{0-1} = bS_{0-1} + \varepsilon[0, f(S_{0+1})]$$

where ε is a noise term with mean zero and variance proportional to S_{0+1}, and b is a constant. In particular, N_{0-1} is well described by a Normal distribution of the form $N_{0-1} \sim \mathrm{N}(bS_{0-1}, cS_{0+1})$, where c is a constant. The variance of our forecast density function can be minimized by choosing a third-party game that achieves the maximum correlation between N_{0-1} and our explanatory variable S_{0-1}, with the unexplained variance being characterized by a linear function of S_{0+1}. We focus on the parameter regime known to produce realistic statistics (e.g. fat-tailed distribution of returns in financial markets). Within this parameter space we run an ensemble of third-party games through the black-box series $H(t)$, calculating the values of S_{0-1} from the reconstructed strategy space. We then identify the configuration that achieves the highest correlation between S_{0-1} and N_{0-1} produced by the original black-box game. As shown in Fig. 1, the third-party game that achieves the highest correlation is the one whose parameters coincide with the black-box game. From a knowledge of just $H(t)$, and hence N_{0-1}, we have therefore used next-step prediction to recover all the parameters of relevance to produce a 'model' game for prediction purposes.

We now extend this forecast to an arbitrary number j of timesteps into the future, in order to address the predictability of large changes in $H(t)$ arising over several consecutive timesteps. This is achieved by calculating the net value of S_{0-1} along all the $k = 2^{j-1}$ possible future routes of the third party game, weighted by appropriate probabilities. In order to assign these probabilities, it is necessary to calculate all possible S_{0-1} values in the next j timesteps. This is possible since the only data required to update the strategy space between timesteps is knowledge of the winning decision, and hence the third party game can be directed along a given path independent of the predictions of the individual agents in the black box game. The change in N_{0-1} along a path indexed by k is given by a convolution of the predictions over the j individual steps

$$\mathrm{N}(\mu_k, \sigma_k) \sim \mathrm{N}\left(b \sum S_{0-1}, c \sum S_{0+1}\right),$$

where the summation is taken along the path represented by k. In general, the pdf for the change in N_{0-1} during the next j timesteps is a mixture of Normals:

$$\mathrm{P}[\Delta N_{0-1}(i; i+j)] = \sum_{k=0}^{2^j-1} p_k \mathrm{N}(\mu_k, \sigma_k), \tag{2}$$

where p_k is the probability of path k being taken.

To test the validity of the density forecast, we perform a statistical evaluation using the realized variables. The one-step-ahead forecasts are normal distributions, and we define the test statistic Z_i as

$$Z_i = \frac{x_i - \mu_i^x}{\sigma_i^x} \tag{3}$$

where μ_i^x and σ_i^x are the mean and variance of the forecast distribution, and x_i is the realised value of N_{0-1} at the timestep i. The Z_i were found to be independent uniform $N(0,1)$ variates for 1000 out-of-sample predictions, confirming that the predicted distributions are correct. To compare the forecasts to a naive 'no-change' prediction, we calculate the Theil coefficient [18] which is the sum of squared prediction errors divided by the sum of squared errors resulting from the naive forecast. A coefficient of less than one implies a superior performance compared to the naive prediction; calculated values were typically in the region of 0.4. There is no accepted method in the literature for evaluating multi-step-ahead forecasts [19]. However, the density function for an arbitrary time horizon is a mixture of Normal distributions, see Eq. (2), each of which can be roughly characterised in terms of a single mean and variance:

$$\mathrm{E}[X] = \sum_{i=1}^{n} p_i \mu_i$$
$$\mathrm{Var}[X] = \sum_{i=1}^{n} p_i(\sigma_i^2 + \mu_i^2) - \left(\sum_{i=1}^{n} p_i \mu_i\right)^2$$

Hence the same test statistic as Eq. (3) can be calculated. Again, the predictions were found to be reliable.

Given that we can derive accurate distributions for the future changes in $H(t)$, these will be of most practical interest in situations where there is likely to be a substantial, well-defined movement. We characterise these moments by seeking distributions with a high value of $|\mu|$ and a low value of σ at a future timestep, or over a specified time horizon. In Fig. 2 we plot $|\mu|$ vs. σ for a number of separate forecasts, and take a fraction of points that are furthest from the average trend indicated by the regression line, i.e. we are interested in the outliers. The point with the highest residual is thus a candidate for the game to be in a highly predictable phase. We call these time periods *predictable corridors*, since comparatively tight confidence intervals can be drawn for the future evolution of the excess demand, a typical example of which is shown in Fig. 3. We subject these points to an identical test as described earlier to ensure these potential outliers are well described by our probability distributions, and this is found to be true.

We performed extensive numerical simulations to check the validity of these predictive corridors [16]. Our procedure is to take a sample of 5000 timesteps, then fit parameters using the first 3000 steps. We then look at the largest changes (extreme events) in our out-of-sample region. Extreme events are ranked by the largest movements in $H(t)$ over a given window size W. Hence we consider the top twenty extreme events and calculate the probability integral transform z_t of the realized variables with respect to the forecast densities. The z_t are found to be approximately uniform U[0,1] variates, confirming that the forecast distribution is essentially correct - see supplementary material for full details. About 50% of large movements occur in periods with tight predictable corridors, i.e. a large value of $|\mu|/\sigma$. Both the magnitude and sign of these extreme events are therefore predictable.

The remainder correspond to periods with very wide corridors. Although the magnitude of the future movement is now uncertain, the present method predicts with high probability the actual direction of change. Even this more limited information would be invaluable for assessing future risk in the physical, economic, sociological or biological system of interest. Our predictions generated from the third-party game were consistent with all such extreme changes in the actual (black-box) time series $H(t)$. Finally we note that some empirical support for our claim of enhanced predictability prior to extreme movements, has very recently appeared for the case of financial markets [20].

We are very grateful to Michael Hart, Paul Jefferies, Pak Ming Hui and Jeff Dewynne for their involvement in this project. D.L. thanks EPSRC for studentship.

REFERENCES

[1] T. Lux and M. Marchesi, Nature **397**, 498 (1999).

[2] See the Econophysics website http://www.unifr.ch/econophysics.

[3] R. Mantegna and H. Stanley, *Econophysics* (Cambridge University Press, Cambridge, 2000).

[4] J. Bouchaud and M. Potters, *Theory of Financial Risks* (Cambridge University Press, Cambridge, 2000).

[5] W. Arthur, American Economic Review **84**, 406 (1994).

[6] B. Huberman, P. Pirolli, J. Pitkow, and R. Lukose, Science **280**, 95 (1998).

[7] J. L. Tella, M. A. R. de Cara, O. Pla, and F. Guinea, in *Sixth Granada Lectures in Computational Physics*, edited by P. L. Garrido and J. Marro (American Institute of Physics).

[8] M. Nowak and R. May, Nature **359**, 826 (1992).

[9] D. Sornette, Physics World **12**, 57 (1999). See also A. Johansen and D. Sornette, e-print cond-mat/0010050.

[10] See, for example, B. Malkiel in *A Random Walk Down Wall Street* (Norton, New York, 1985).

[11] D. Challet and Y.-C. Yhang, Physica A **246**, 407 (1997); D. Challet and Y.-C. Yhang, Physica A **256**, 514 (1998). See Ref. [2] for a full bibliography of the Minority Game.

[12] N. F. Johnson, P. M. Hui, D. Zheng, and C. W. Tai, Physica A **269**, 493 (1999).

[13] P. Jefferies, N. F. Johnson, M. Hart, and P. M. Hui, *From market games to real-world markets* (Eur. J. Phys. B, 2001). See also I. Giardina, J.P. Bouchaud and M. Mezard, e-print cond-mat/0105076.

[14] N. F. Johnson, M. Hart, P. M. Hui, and D. Zheng, Int. J. Theo. Appl. Fin. **3**, 443 (2000).

[15] N. F. Johnson, M. Hart, and P. M. Hui, Physica A **269**, 1 (1999).

[16] Further details of the statistical tests and parameter sets are available from authors.

[17] An accurate estimate of x_i can be recovered using Kalman filters. (Work to be presented at the 7th International Conference of the Society for Computational Economics, Yale University, June 2001).

[18] T. Hellström and K. Holmström, Technical Report IMa-TOM-1997-07, Mälardalen University, Sweden.

[19] A. S. Tay and K. T. Wallis, J. Forecasting **19**, 235 (2000).

[20] R. Mansilla, e-print cond-mat/0104472.

FIGURES

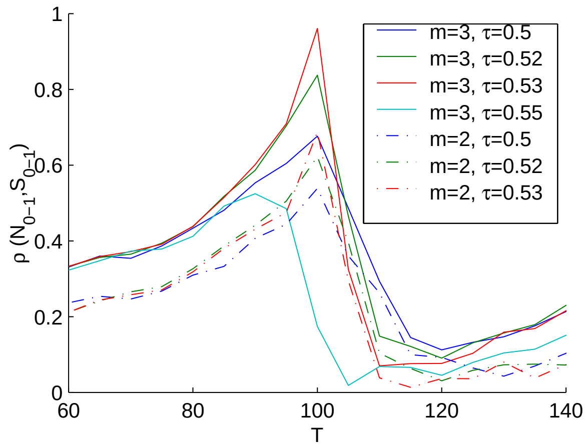

FIG. 1. Estimation of the parameter set for the black-box game. The correlation between N_{0-1} and S_{0-1} is calculated over 200 timesteps for an ensemble of candidate third-party games. The third-party game that achieves the highest correlation is the one with the same parameters as the black-box game.

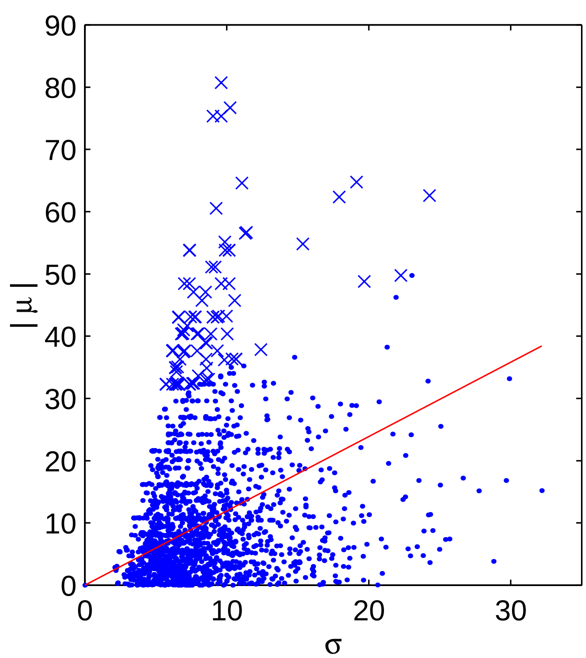

FIG. 2. A plot of $|\mu|$ vs. σ for 500 separate four-step density forecasts. Items marked by "x" are forecasts with an unusually large value of $|\mu|/\sigma$. At these moments, the game is likely to be in a highly predictable phase.

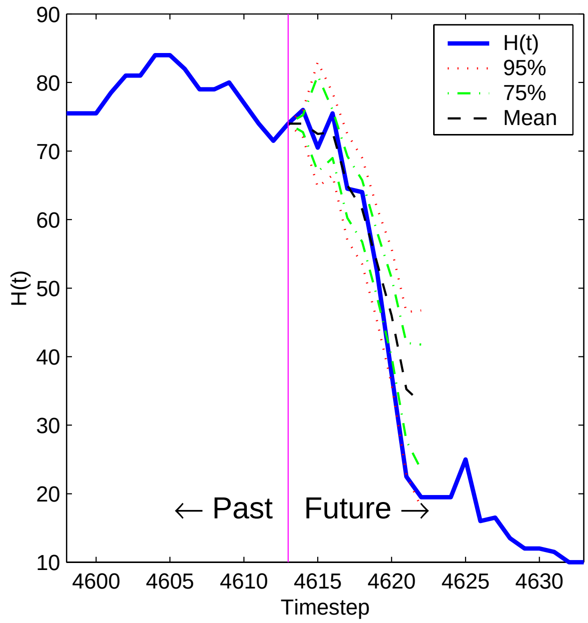

FIG. 3. Comparison between the forecast density function and the realised time series $H(t)$ for a typical large movement. The large, well-defined movement is correctly predicted.