Forecasting with Entropy

Christopher A. Zapart

The Institute of Statistical Mathematics,
10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan
czapart@ism.ac.jp

Abstract

The paper describes an alternative approach to forecasting financial time series based on entropy (C. A. Zapart, On entropy, financial markets and minority games, Physica A: Statistical Mechanics and its Applications, 388 (7) 2009, pages 1157-1172). The research builds upon an earlier statistical analysis of financial time series with Shannon information entropy, published in (Molgedey, L and Ebeling, W, 2000, "Local order, entropy and predictability of financial time series", European Physical Journal B - Condensed Matter and Complex Systems). A novel generic procedure is proposed for making multistep-ahead predictions of time series by building a statistical model of entropy. The approach is applied to the prediction of Japanese Yen/US dollar intraday currency exchange time series. The study also reinterprets the Minority Game (Moro E, 2004, "The Minority Game: an introductory guide", Advances in Condensed Matter and Statistical Physics) within the context of physical entropy, and uses models derived from minority game theory as a tool for measuring the entropy of a model in response to time series. This entropy conditional upon a model is subsequently used in place of information-theoretic entropy in the proposed multistep prediction algorithm.

Subsequently the paper suggests using alternative entropy measures such as author's NeuroEntropy or Approximate Entropy, introduced in (S Pincus, Approximate entropy as a measure of system complexity, Proceedings of National Academy of Sciences USA 1991 (88) pages 2297-2301) and (S Pincus, Irregularity, volatility, risk and financial market time series, Proceedings of National Academy of Sciences USA 2004 101 (38) pages 13709-13714). Exponentially-Weighted Smooth Approximate Entropy is proposed to make it more sensitive to recent data.

1 Introduction

In many areas of science and engineering a common recurring task arises: forecasting ΔT steps ahead an observed time series $x_1, x_2, \ldots, x_N$ of some physical phenomena. Usually the process may involve finding a linear or non-linear regression function that takes as arguments p past values of $x(t)$. However, in cases where observations come from financial markets standard statistical techniques fail to capture sufficiently the underlying time series generator and its often non-stationary nature. This is evidenced by poor out-of-sample performance of models yielding apparently good in-sample fits [1]. The inadequacy of the existing approaches has been mentioned before in for example [2, 3], and some explanation of the perceived failures of the status quo has been contained in the works of Mandelbrot [4]. One of the reason for the failure is that established statistical time series forecasting models, both linear regression and non-linear neural networks, do not take into account the physical generative aspect of financial time series. Financial time series arise directly as a result of interactions between a large number of traders which provides a justification for applying various tools of statistical physics to computational finance [5, 6, 7]. As a consequence, from a physics point of view a much more attractive proposition is to try to approximate the underlying processes responsible for generating the time series in the first place.

The paper describes concisely an entropy-based approach to forecasting financial time series that can be seen in full detail in [7]. The indirect forecasting method is applied to multistep-ahead prediction of the Japanese Yen/US Dollar intraday currency futures data.

2 Forecasting with Entropy

We assume that there are N observations $x_1, x_2, \dots, x_N$ available from some physical process. The goal of the analysis is to produce an estimate for the change $x_{t+\Delta T} - x_t$, where $t \in \mathbb{I}$ denotes the current time step and ΔT is the forecasting horizon. The proposed algorithm can be outlined in the following sequence of steps:

(1)	start with an observation sequence leading up to x_t
	$\Downarrow$
(2)	extract entropy time series $H(t)$ from x_t
	$\Downarrow$
(3)	assume a statistical model for entropy
	$\Downarrow$
(4)	simulate possible future paths of x_{t+k} ΔT steps ahead
	$\Downarrow$
(5)	measure entropy for the simulated paths
	$\Downarrow$
(6)	discard paths inconsistent with entropy model identified in (3)
	$\Downarrow$
(7)	formulate a forecast for $x_{t+\Delta T}$ based on the retained paths

With respect to estimating entropy $H(t)$ of a given time series, a good starting point might be the Shannon n-gram (block) entropy suggested in [5, 6]. In a very general case, a given sequence of N observations $x_1, x_2, \ldots, x_N$ is first partitioned into subvectors of length L with an overlap of one time step, which are further divided into subtrajectories (delay vectors) of length $n < L$. Real-valued observations $x_i \in \mathbb{R}$ are discretised by mapping them onto λ non-overlapping intervals $A^{\{\lambda\}}(x_i)$. The precise choice of those intervals (also called *states*) denoted by $A^{\{\lambda\}}$ would depend on the range of values taken by x_i[1]. Hence a certain subtrajectory $x_1, x_2, \ldots, x_n$ of length n can be represented by a sequence of states $A_1^{\{\lambda\}}, A_2^{\{\lambda\}}, \ldots, A_n^{\{\lambda\}}$. The authors then define the n-gram entropy (entropy per block of length n) to be

$$H_n = -\sum_{\Omega} p\left(A_1^{\{\lambda\}}, A_2^{\{\lambda\}}, \ldots, A_n^{\{\lambda\}}\right) \log_\lambda p\left(A_1^{\{\lambda\}}, A_2^{\{\lambda\}}, \ldots, A_n^{\{\lambda\}}\right). \quad (1)$$

In the above equation the summation is done over all possible state sequences $\Omega \in \{A_1^{\{\lambda\}}, A_2^{\{\lambda\}}, \ldots, A_n^{\{\lambda\}}\}$. The probabilities $p\left(A_1^{\{\lambda\}}, A_2^{\{\lambda\}}, \ldots, A_n^{\{\lambda\}}\right)$ are calculated based on all subtrajectories $x_1, x_2, \ldots, x_n$ contained within a given subvector of the length L. Predictability of the time series, expressed as an uncertainty of predicting the next step given the past n states $A^{\{\lambda\}}$, is given by a conditional (dynamic) entropy (or differential block entropy)

$$h_n = H_{n+1} - H_n. \quad (2)$$

In the actual analysis [5, 6] of financial time series, instead of using the dynamic entropy as per equation (2) the authors use a local value of the uncertainty defined [5] as

$$h_n^{(1)} = -\sum_{\Omega'} p\left(A_{n+1}^{\{\lambda\}} | A_1^{\{\lambda\}}, A_2^{\{\lambda\}}, \ldots, A_n^{\{\lambda\}}\right) \log_\lambda p\left(A_{n+1}^{\{\lambda\}} | A_1^{\{\lambda\}}, A_2^{\{\lambda\}}, \ldots, A_n^{\{\lambda\}}\right), \quad (3)$$

where $\Omega' \in \{A_{n+1}^{\{\lambda\}}\}$ enumerates all possible λ states for $A_{n+1}^{\{\lambda\}}$. As noted in [5] $h_n^{(1)}$ satisfies

$$0 \leq h_n^{(1)} \leq 1. \quad (4)$$

At first the local dynamic entropy given by equation (3) will be used in this paper. Subsequently the information theoretic entropy will be replaced by a physical entropy extracted from minority game theory models [8].

[1]In subsequent sections financial time series will be used. In case of modelling financial returns the simplest choice of states $A^{\{\lambda\}}$ can be made by setting $\lambda = 2$. Then $A^{\{\mathrm{up}\}}$ corresponds to positive returns and $A^{\{\mathrm{down}\}}$ would represent negative returns.

In order to make predictions over a given time horizon ΔT all possible $\lambda^{\Delta T}$ future paths $\zeta_k = \{x^k_{N+1}, x^k_{N+2}, \dots, x^k_{N+\Delta T}\}, k = 1 \dots \lambda^{\Delta T}$ are generated assuming an equal probability[2] of occurrence of each state $A^{\{\lambda\}}$. The simulated paths are appended to the end of the original sequence $\{x_i\}$. Assuming the last available observed entropy to be H_N, for each path ζ_k one can note the corresponding entropy sequence

$$H_N \rightarrow H^k_{N+1} \rightarrow H^k_{N+2} \rightarrow \dots \rightarrow H^k_{N+\Delta T}. \tag{5}$$

The probability of the path ζ_k can be linked to the probability of encountering such an entropy sequence:

$$P(\zeta_k) = \prod_{i=1}^{\Delta T} P\left(H^k_{N+i} | H^k_{N+i-1}, H^k_{N+i-2}, \dots\right), \tag{6}$$

with the starting point $H^k_N = H_N$. In the simplest case, for estimating conditional probabilities $P\left(H^k_{N+i} | H^k_{N+i-1}, \dots\right)$ the following maximum-likelihood procedure is suggested. As entropy takes on values between 0 and 1 a simple logistic regression model of the order p is constructed to provide one-step ahead predictions of H^k_{N+i}:

$$\hat{H}^k_{N+i} = \frac{1}{1 + \exp(-\phi)} \tag{7}$$

$$\phi = w_0 + \sum_{j=1}^{p} w_j H^k_{N+i-j}. \tag{8}$$

For more complex entropy time series one could also consider using a committee of non-linear multilayer feedforward neural networks [9, 10]. As prediction errors are bounded by $\left|\hat{H}^k_{N+i} - H^k_{N+i}\right| \leq 1$, the following formulæ derived from a β probability density function is used to model the conditional probability from equation (6):

$$P\left(H^k_{N+i} | H^k_{N+i-1}, \dots\right) = \frac{\epsilon^{\alpha-1}(1-\epsilon)^{\beta-1}}{B(\alpha, \beta)} \tag{9}$$

$$\epsilon = \frac{\hat{H}^k_{N+i} - H^k_{N+i} + 1}{2} \tag{10}$$

with optimum parameters $\hat{\alpha}$ and $\hat{\beta}$ obtained by maximising the likelihood of observing the entropy time series $H(t)$:

$$\{\hat{\alpha}, \hat{\beta}\} = \operatorname*{argmax}_{\alpha>0, \beta>0} \prod_{j=p+1}^{t} P\left(H_j | H_{j-1}, H_{j-2}, \dots, H_{j-p}\right). \tag{11}$$

[2]The equal probability follows from the maximum entropy principle.

The reason for choosing a β probability density function in equation (9) is purely pragmatic. A β distribution can accommodate a range of empirical histograms obtained for the residual errors ϵ. To further simplify early experiments, a normal distribution

$$P\left(H^k_{N+i}|H^k_{N+i-1},\dots\right) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(\epsilon-\mu)^2}{2\sigma^2}\right) \tag{12}$$

has also been used to fit even simpler residual errors

$$\epsilon = \left(w_0 + \sum_{j=1}^{p} w_j H^k_{N+i-j}\right) - H^k_{N+i}. \tag{13}$$

For the entropy modelling purposes only the original observations $x_1, x_2, \dots, x_N$ are used together with their corresponding dynamic entropies. There is no one defined way of modelling the entropy with regression models. The choice, dependent on the actual entropy time series, has to be made on a case-by-case basis.

Let $x(\zeta_k)$ denote the final simulated observation $x^k_{t+\Delta T}$ coming from the kth ζ_k path. A pseudo-forecast[3] of $x_{t+\Delta T}$ is given by

$$\mu = E\left[x^k_{t+\Delta T}\right] = \frac{\sum_k P(\zeta_k)x(\zeta_k)}{\sum_k P(\zeta_k)} \tag{14}$$

with the corresponding variance

$$\sigma^2 = \mathrm{Var}\left(x^k_{t+\Delta T}\right) = \frac{\sum_k P(\zeta_k)\left(x(\zeta_k)-\mu\right)^2}{\sum_k P(\zeta_k)}. \tag{15}$$

Effectively the equations (14) and (15) discard paths ζ_k that are inconsistent with an identified statistical model for the entropy $H(t)$. The approach replaces setting up a regression model for the observations $\{x_i\}$ with performing linear (or non-linear) regression on the entropy time series and simulating the future evolution of the system alongside all possible observation trajectories.

3 Demonstration

The main reason for suggesting such an unusual entropy-based forecasting method was (perhaps "deemed to fail" from the beginning) an attempt to predict financial time series. The following two subsections show early results obtained using the two types of entropy: the local dynamic entropy given by equation (3) and a physical entropy extracted from minority game theory models.

[3]This is not a proper forecast of $x_{t+\Delta T}$ since the original time series x_t has been mapped onto λ discrete states $A^{\{\lambda\}}$. For small λ the transformation removes much of the vital information about the magnitude of each move.

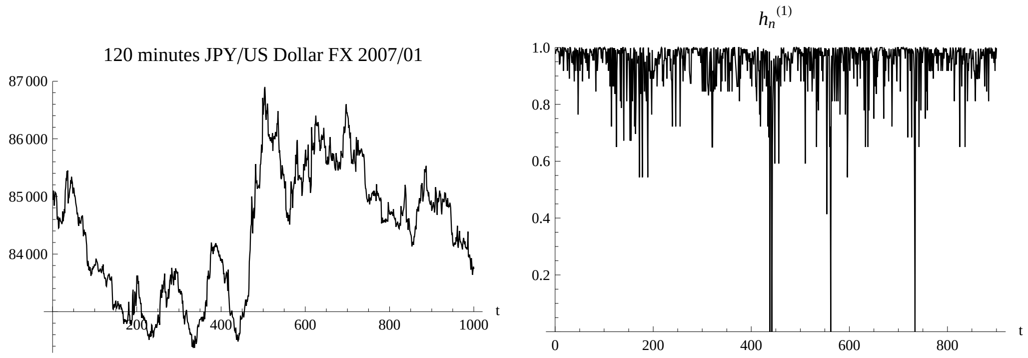

Figure 1: Intraday GLOBEX foreign exchange futures data for the Japanese Yen/US Dollar sampled at a 120 minute interval (left) with the corresponding entropy $h_n^{(1)}$ (right). The time series starts on 2007/01/02; only the first 1000 data points are shown.

3.1 Information-theoretic entropy

The analysis was performed on intraday Japanese Yen/US Dollar currency futures data collected from GLOBEX between 2$^{\text{nd}}$ January and 31$^{\text{st}}$ October 2007. The first 1000 prices, sampled at regular two hour intervals, are shown in figure (1). The same figure also shows the corresponding local entropy $h_n^{(1)}$. For simplicity only two states are used with $\lambda = 2$. The local dynamic entropy $h_n^{(1)}$ was estimated using windows (subvectors) of length $L = 100$. The observation time series was first transformed into a series of differences $\{x_i - x_{i-1}\}$ before being mapped onto two binary states: $A^{\{1\}} =$ "up$''$ and $A^{\{2\}} =$ "down$''$, which corresponds to the differenced series taking on either *positive* or *negative* values. Since at 100 time steps the estimation windows for n-gram (block) entropies are relatively short, accordingly short subtrajectories of length $n = 3$ were selected in order to reduce frequentist probability estimation errors associated with small sample sizes. As the currency exchange time series does not exhibit a strong exponential growth there was no need for taking logarithmic price changes. Instead normal differences $\{x_i - x_{i-1}\}$ were calculated.

The plot of the entropy time series resembles the one extracted in [5] from the Dow Jones Industrial Average stock market index with $\lambda = 3$. The system spends most of the time near the *maximum entropy* state of 1.0, only very briefly visiting low entropy configurations. The distribution of entropy values $h_n^{(1)}(t)$ has been fitted with a two-component β-mixture model

$$pdf_H(x) = \sum_{i=1}^{2} \pi_i \frac{x^{\alpha_i - 1}(1-x)^{\beta_i - 1}}{B(\alpha_i, \beta_i)}, \tag{16}$$

where the mixing coefficients π_i are probabilities satisfying $0 \leq \pi_i \leq 1$, $\sum_{i=1}^{2} \pi_i =$

1. The maximum likelihood EM (Expectation-Maximisation) algorithm [11] coupled with Simulated Annealing [12] (used in the M-step of the EM algorithm) was employed to find the optimum parameters $\{\pi, \alpha, \beta\}$. After fitting the most probable value (the peak of the probability density function) for $h_n^{(1)}$ was identified at 0.98 which is very close to the maximum entropy value of 1.0. At this point it is worth going back to statistical physics and recall that:

1. in a disordered state of maximum entropy a system *loses memory* of past events
2. literature ([3], [5], [6]) suggests existence of only limited *pockets of predictability* in real financial time series, to be found when entropy reaches temporary troughs.

In other words, although predicting financial time series may be very difficult most of the time, there may exist for very limited periods of time some occasional trading opportunities. These opportunities would typically be associated with low entropy states which do not persist for long, as can be observed in figure (1). Therefore it may be worth trying to exploit the mean-reversion exhibited by entropy time series through the concept of a binomial entropy tree, shown in figure (2). Starting from a low value of entropy $h_n^{(1)}$ all possible future paths ζ_k of the price time series x_t are simulated over a time horizon ΔT and entropy is calculated. The mainstream econometrics modelling framework (computational finance) assumes that prices of financial assets follow a stochastic Random Walk [13] process. As such, after adjusting for the drift component and a risk free rate, at any given time step financial time series have an equal probability of moving up or down. Therefore any random path of length ΔT would be equally probable. However, according to entropy mean reversion, depending on the current value of entropy certain future price trajectories — those leading to yet lower net entropy over the time horizon ΔT — are deemed "physically impossible" (or statistically very unlikely). Therefore they could be excluded from the Random Walk, which results in a "biased" Random Walk. Price forecasts are given by a distribution of future prices from the modified (biased) Random Walk that excludes low probability ("physically impossible") paths, as per equations (14) and (15). The effective forecasting horizon ΔT is itself governed by the relaxation time τ after which the entropy reverts to mean. Following a temporary trough in entropy it does not seem reasonable to be seeking forecasts long after the system has returned to its natural *maximum entropy* state. For relatively short relaxation times τ it is sufficient to use a binomial entropy tree instead of Monte Carlo simulations to simulate possible paths ζ_k. If the number of possible paths $\lambda^{\Delta T}$ is large a binomial tree becomes impractical. In general, for more complex cases with

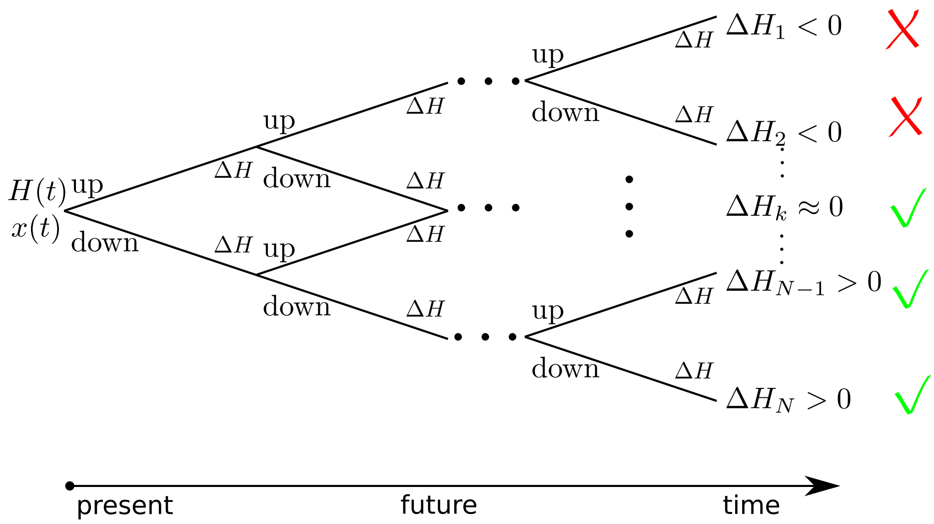

Figure 2: A conceptual binomial entropy tree corresponding to $\lambda = 2$, starting from a low value of $H(t)$. The paths k for which the net entropy ΔH_k is negative are deemed *physically impossible* (highly improbable), hence they are disallowed (assigned a very low probability). Labels "up" and "down" refer to possible up and down moves in the foreign exchange rate (or share price) x_t, not the entropy $H(t)$. The net entropy ΔH is measured along all possible future movements of a financial asset.

$\lambda > 2$ it may be beneficial to consider simulating an entropy-adjusted Random Walk[4]. Taking a lead from econophysics, a power-law distribution [14] could be used to model logarithmic returns of financial time series. Then paths ζ_k drawn in a Monte Carlo simulation would approximate possible trajectories a financial asset can follow between timesteps t and $t+\Delta T$. After discretely mapping ζ_k onto λ states $A^{\{\lambda\}}$ the entropy-based forecasting procedure would continue.

The aforementioned Japanese Yen/US Dollar time series contains 2544 samples of which the first 1000 quotes are used to initialise the entropy model and train a committee of small neural networks for modelling $h_n^{(1)}$. For those 1000 points, shown in figure (1), the average entropy relaxation time $\langle\tau\rangle$ has been estimated at 3 steps. In 95% of the cases the entropy reverts to mean within 7 time steps. The remaining 1544 prices serve as an *out-of-sample* test set, with the forecasting horizon accordingly set to $\Delta T = 7$ steps. The overall out-of-sample forecasting accuracy without applying any trading filters is just under 52% which is what could be expected from a purely random coin toss. The figure (3) shows what happens when only the *low-entropy* trades are included (the left chart). Improved forecasting accuracy can be obtained by selecting only forecasts with the largest magnitudes, as shown in the right chart in figure (3). This is achieved by linearly rescaling all the forecasts so that they lie within the $[-1, 1]$ interval. Afterwards a *forecasting threshold* is applied to the magnitudes of the rescaled model forecasts given by equation (14). By excluding forecasts with a magnitude below the forecasting threshold, which is varied between 0 and 1, one can estimate the forecasting accuracy of the remaining large magnitude (above-threshold) predictions. To calculate the forecasting accuracy of low-entropy trades, the forecasts that are taken into account are those where the initial entropy $H(t)$ (appearing in the left part of figure (2)) lies below an arbitrary entropy threshold varied between 0 and 1. What is striking is that indeed, as expected from statistical physics, forecasts initiated in low-entropy states have a higher than 50% chance of being correct, with the trading accuracy gradually reverting towards the average 50% after increasing the entropy cut-off threshold. The accuracy also increases especially for large-magnitude forecasts. However, these results have to be treated with extreme caution as the rather high forecasting accuracy has been achieved on very few samples, for example the number of cases found in the test data where $h_n^{(1)} \leq 0.5$ was only 23. The same pattern of above-average forecasting accuracy for low-entropy trades is repeated on out-of-sample data in previous years 2006 and 2005 (shown in figure (4)). Taken together, the limited study seems

[4]The fact that a Random Walk model is considered should not be seen as supporting its supposed validity in describing the "true" generative model of financial assets. Rather, it could be used in a Monte Carlo-style fashion to approximate possible future paths of a financial asset.

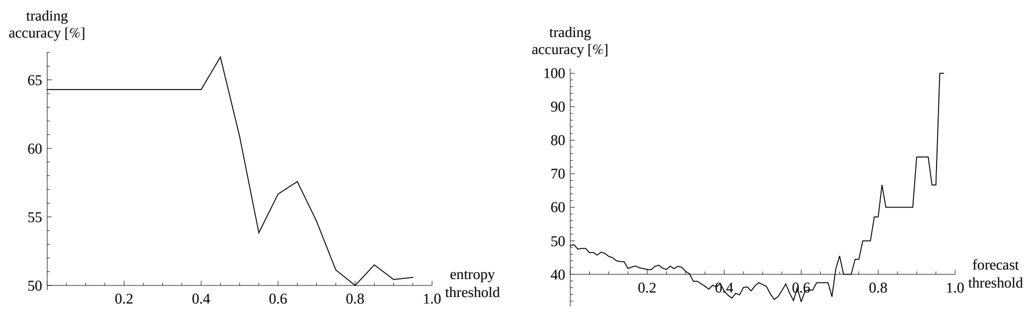

Figure 3: Out-of-sample forecasting accuracy for 120 minutes Japanese Yen/US Dollar year 2007 futures data. The left plot shows the percentage of successful forecasts starting from the initial entropy H_N below a preset threshold. The accuracy displayed in the chart to the right depends on the forecasting threshold applied to the output from equation (14), varied between 0 and 1.

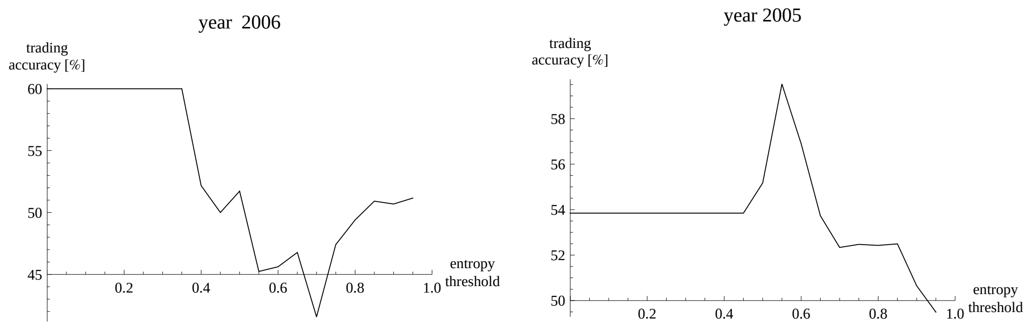

Figure 4: Percentage of successful out-of-sample low-entropy forecasts for years 2006 (left figure, $\Delta T = 5$) and 2005 (right figure, $\Delta T = 6$).

to corroborate conclusions reached by other authors ([3], [5], [6]) dealing with financial time series in that low entropy regions can be associated with trading opportunities.

3.2 Minority Games

The previous subsection demonstrated an unusual forecasting technique that relies on directly extracting local dynamic entropies from the time series of observations x_t. When applied to predicting financial time series the method showed a weak positive result mostly associated with rare low-entropy events. In contrast with direct entropy estimation [5, 6], in this subsection an attempt will be made at using an indirect Boltzmann physical entropy extracted from minority game theory models in response to the external time series x_t. This will replace the previously used information-theoretic entropy $h_n^{(1)}$.

Over the past decade there has been considerable interest expressed by the research community in applying Minority Games to study financial time series. The use of the Minority Game Theory in financial markets can be traced back to a seminal paper [15] published in 1997. Readers unfamiliar with minority games may find the following references [8] and [16] useful. A more detailed explanation of how to build realistic market models within the context of Minority Games is provided in [17]. At the root of all minority game theory models lies a simple observation: for a given level of risk in financial markets no single investment strategy or a trading system remains "excessively"[5] profitable in the long term. In other words, stock markets tend to be mean-reverting and exhibit zero-sum game characteristics. As soon as a highly successful investment strategy is discovered, increasing numbers of traders tend to adopt it. When a majority of market participants start using it, the profitability of such a strategy gradually diminishes. The only entities that benefit from the profitable strategy are early adopters able to take advantage of it in early stages when only a small minority of traders are aware of it. This results in a minority of traders benefiting at the expense of the majority.

The study published in [7] reinterprets Minority Games within the context of statistical physics using the concept of entropy, and uses it in an attempt to make practical predictions of financial time series. In most types of minority games there are N binary computational agents that buy or sell imaginary shares in a competitive setting. At each time step, of those N agents N_+ will be found placing buy orders and N_- will be sellers. From a physics perspective minority games resemble a spin glass — a disordered magnetic material with a high level of frustration. By frustration one usually means the inability of a system to remain in the lowest energetic state. The "buy" decisions are equivalent to an "up" magnetic spin orientation, and "sell" operations correspond to the "down" spin. It is possible to estimate the level of disorder exhibited by such a simple system by defining a Boltzmann entropy $S = k_B \ln W$, where the state multiplicity $W = \binom{N}{N_+}$ is the number of different realisations of a current state characterised by N_+ and N_-. The maximum theoretically allowed entropy value S_{max} corresponds to the case $N_+ = N_-$. Detailed studies of the internal dynamics of minority games have been presented before, for instance in ([18],[19]). It is therefore not the author's intention to include a similar analysis in this paper. Instead the current study focuses on the behaviour of entropy when minority game theory models are presented with real financial time series, and on incorporating this entropy within the previously described forecasting approach. In minority games it is possible to replace the internal binary time series generated by the game with external real financial time series as demon-

[5] "Excessive" profits can be thought of as being above the long-term growth rate of the economy.

strated in [20]. Agents take trading decisions based on the binary series $h_t = 0101101\ldots$ encoding the past winning groups ("1" for buyers and "0" for the sellers). The history of real financial time series x_t coming, for example, from the foreign exchange market, can be transformed into a binary form according to $h(t) = \mathrm{H}[x_t - x_{t-1}]$, where $\mathrm{H}[\cdot]$ denotes the Heaviside function, with $h(t)$ then replacing the internal series h_t. This is equivalent to performing a discrete mapping of x_t onto states $A^{\{\lambda\}}$ with $\lambda = 2$ for the purpose of calculating the Shannon n-gram (block) entropies.

The literature on minority game theory models describes different variations of minority games. However, in most versions of the game agents take trading decisions according to some sort of a risk criterion. For example, a common element of minority models is keeping track of the success rates for both the binary strategy tables used internally by the agents as well as the overall prediction accuracies of each agent. During trading agents only trade if their *confidence to trade* exceeds a risk threshold. The literature [17] suggests using two types of thresholds: either based on agent scores or strategy scores that can be used interchangeably. In [7] the following scheme was adopted. Over a time horizon $T = 100$ $N = 151$ agents collect points +1 or -1 each time they win or lose which then form the confidence to trade measure $c(t)$. In addition they also keep a separate record of those values of $c(t)$ measured over a horizon $T_c = 100$. Individual minimum trading thresholds r_{min} are set to $\mu + \lambda\sigma$ where $\mu = \langle c(t) \rangle$ and σ is the standard deviation of $c(t)$ calculated over T_c. The parameter λ controls the risk appetite of agents. Another difference over standard minority games was to try to average out the effects on recovered entropy of agents using different random strategy tables. Instead of using a single minority game theory model, the average entropy extracted from a committee of M models is used with M being as high as 100. Agents have two distinct random strategy tables and the length of the global price history set to $m = 3$.

After experimenting with different values of the risk parameter λ some very interesting entropy time series have been observed for $\lambda \approx 3$. Due to the computational cost[6] of simulating 100 separate minority games, in the actual simulation the physical entropy, shown in the next figure (5), was averaged over only 10 separate games. Prediction results using a long forecasting horizon $\Delta T = 20$ steps are shown in figures (6), (7) and tables (1), (2). It can be observed that indeed low entropy regions can be positively associated with a higher than average probability of directional forecasts being correct. However, obtaining a certain degree of consistency across the three datasets seems problematic. Overall, the forecasting accuracy for large-magnitude trades, shown in figure (7) and table (2), rises well above 50%.

[6] Although executing standard minority games on modern computers is very fast, the algorithm described in [7] involves significant overheads to do with simulating multiple paths, running neural networks and calculating path probabilities. This is carried out in a hybrid C/C++/Mathematica environment which introduces further overheads.

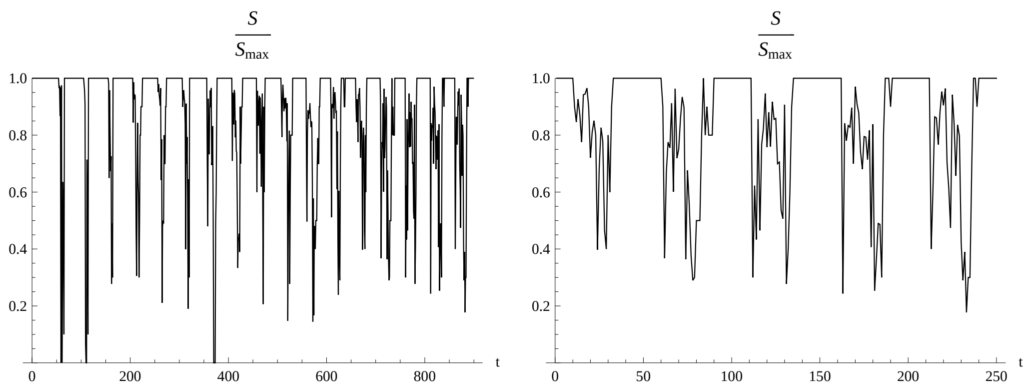

Figure 5: In-sample S/S_{max} entropy time series averaged over 10 games with a risk level $\lambda = 3.0$. The right plot shows a close-up snapshot of the last 250 points.

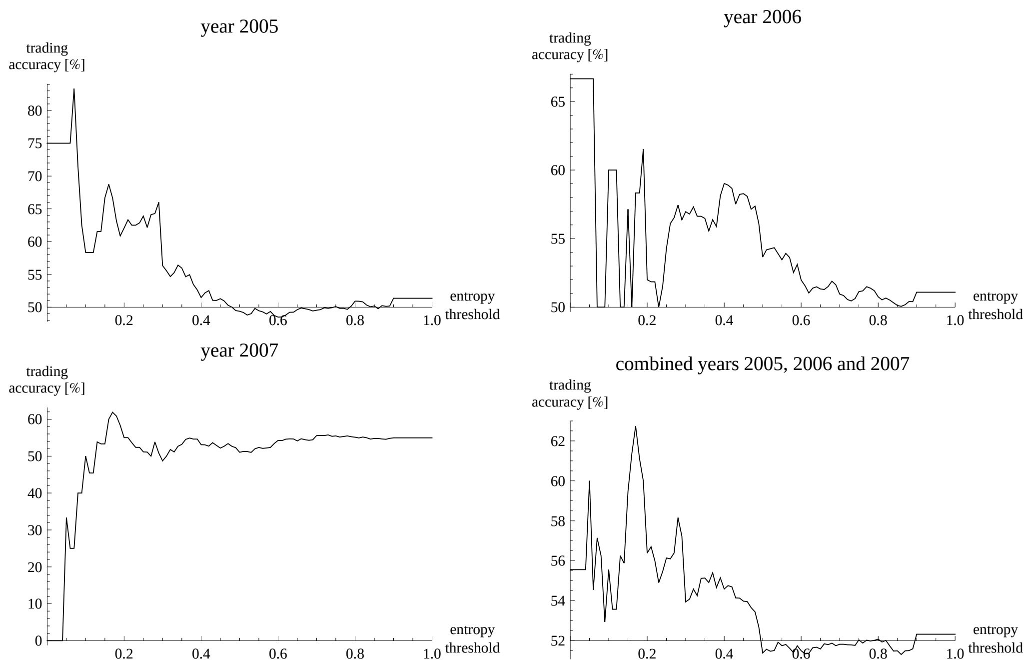

Figure 6: Forecasting accuracy of low-entropy trades using a forecasting horizon $\Delta T = 20$ steps for years 2005, 2006 and 2007. The bottom right plot shows a combined accuracy after collating together trades from those three years. Large fluctuations seen for low values of the entropy threshold are likely caused by the scarcity of data points in those regions.

Table 1: The accuracy and number of trades initiated when the entropy is at or falls below a set threshold (out-of-sample data for the years 2005, 2006 and 2007).

year 2005

entropy threshold	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
accuracy [%]	75	58	62	56	51	49	48	50	51	51
number of trades	4	12	29	71	136	239	363	525	709	999

year 2006

entropy threshold	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
accuracy [%]	67	60	52	57	59	54	52	51	51	51
number of trades	3	5	25	79	144	246	379	522	733	1010

year 2007

entropy threshold	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
accuracy [%]	0	50	55	49	53	51	54	56	55	55
number of trades	2	10	40	78	145	237	328	439	584	837

Table 2: Combined years 2005, 2006 and 2007 out-of-sample forecasting accuracy and number of trades initiated when the magnitude of the model forecast is above a set threshold.

combined years 2005, 2006 and 2006

output threshold	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
accuracy [%]	52	53	51	50	51	51	53	55	63	83
number of trades	2846	2153	1663	1229	890	599	386	206	87	12

4 NeuroEntropy

Potential improvements could be brought about by the use of neural networks in place of simple binary look-up tables (as suggested in for example [21, 22]) and perhaps also by using evolutionary minority games [23]. In the future research one could examine how these changes affect the entropy and the overall predictability of financial time series. For example, by taking advantage of the NVIDIA CUDA/OpenCL GPGPU programming environment, a new *NeuroEntropy* model has been developed. NeuroEntropy, being a form of an entropy-adjusted Random Walk model, can be used as a forward-looking short-term risk indicator in financial markets. However, the early version appearing in [7] used binary minority games which ignored the magnitude of returns in financial time series. This restriction can be removed by replacing binary agents with real-valued multilayer perceptron neural networks. Information-theoretic Shannon entropy is derived from a

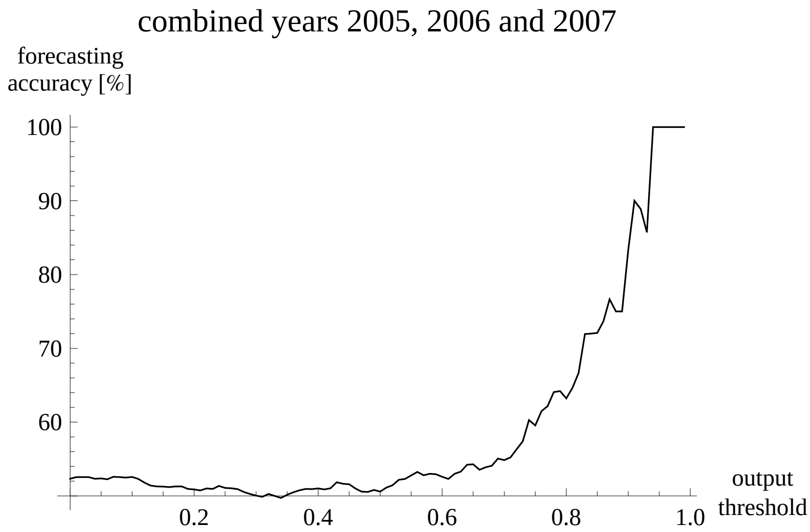

Figure 7: Combined years 2005, 2006 and 2007 out-of-sample forecasting accuracy as a function of the output threshold set between 0 and 1. As the output threshold is increased the number of samples (shown in table (2)) steadily decreases adversely affecting the accuracy of results.

set of N two-state neural networks using equation 17:

$$H = -\sum_{i=1}^{N} \left[P\left(y_i = +1\right) \log P\left(y_i = +1\right) + P\left(y_i = -1\right) \log P\left(y_i = -1\right) \right] \quad (17)$$

where $P\left(y_i = +1\right)$ and $P\left(y_i = -1\right)$ are probabilities of the ith neural network being in the $+1$ or -1 state. Tsallis or Rényi entropies can also be used. Figure 8 shows a $\Delta T = 5$ steps ahead reconstruction/prediction of a discretely-sampled sine function using NeuroEntropy. The subsequent figure 9 shows an example of an asymmetric forward risk profile yielded by NeuroEntropy.

Instead of using the Shannon information-theoretic entropy, one could replace it with Approximate Entropy, introduced in [24] and [25]. Exponentially-Weighted Smooth Approximate Entropy has also been developed by the author to make it sensitive to more recent data.

5 Conclusions

The study represents an imperfect attempt to utilise entropy in the hope of being able to predict financial time series. An alternative time series forecasting method has been demonstrated which relies on building a statistical model of entropy. Instead of predicting directly the underlying time series

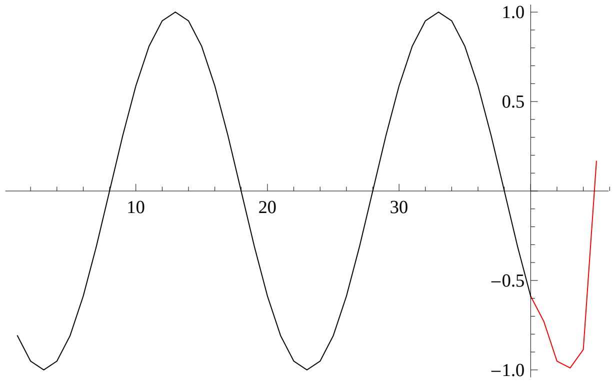

Figure 8: Toy problem: reconstructing a discretely-sampled sine function $\Delta T = 5$ steps ahead. The red curve shows the most probable future path as recovered by NeuroEntropy. Potential path candidates have been sampled randomly from a uniform distribution $U(-1, 1)$.

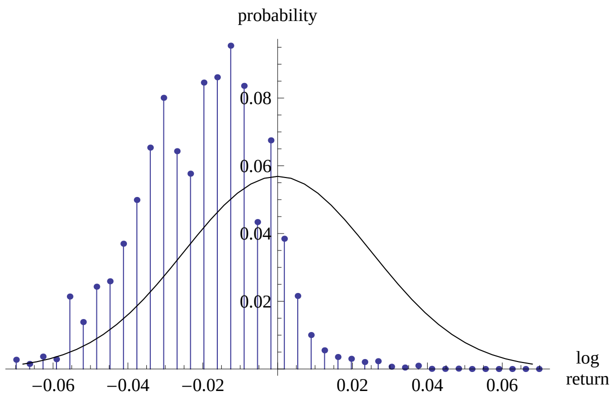

Figure 9: Entropy-adjusted Random Walk (vertical lines) versus standard Random Walk (black solid line) for future logarithmic returns of the Nikkei 225 index as of 2010/01/15, just before major falls lasting over two weeks. According to the NeuroEntropy model the forward risk is skewed to the downside, in contrast with a symmetrical risk profile offered by the standard Random Walk model.

the method first extracts the corresponding entropy, subsequently performing predictions on the entropy time series. A weak trading advantage has been found in financial forecasts of foreign exchange currency futures initiated in low entropy regions, which agrees with results from other, earlier econophysics studies. Conversely, predicting time series in high entropy regions is very difficult to achieve. This follows directly from statistical physics which teaches that in a disordered state of maximum entropy complex systems lose memory of past events.

Established statistical time series forecasting techniques, both linear regression and non-linear neural networks, do not take into account the physical generative aspect of financial time series. Such time series arise directly as a result of interactions between a large number of traders. As a consequence, from a physics point of view a much more attractive proposition is to try to approximate the underlying processes responsible for generating the time series in the first place. Therefore the paper attempted to replace an information-theoretic entropy with a physical entropy extracted from minority game theory models. According to literature [18] such models could provide a simplified approximation to the way real financial markets operate. One advantage of minority games is that they allow more control over the type of disorder (or complexity measure) being extracted from the time series. However, this comes at a price of having to decide how to choose the "correct" model configuration. As yet there is no principled way of dealing with this issue.

Summing up, the paper attempts to express quantitatively a qualitative claim that "low entropy regions lead to improved predictability of financial time series". In doing so it makes a positive contribution towards greater acceptance of econophysics by the mainstream computational finance. For more comprehensive description of the method outlined in this paper readers are encouraged to refer to [7].

References

[1] Franses P H Clements, M P and Swanson N R. Forecasting economic and financial time-series with non-linear models. *International Journal of Forecasting*, 20/2:169–183, April/June 2004.

[2] J D Farmer. Physicists attempt to scale the ivory towers of finance. *Computing in Science and Engineering*, pages 26–39, November/December 1999.

[3] G A Darbellay and D Wuertz. The entropy as a tool for analysing statistical dependences in financial time series. *Physica A*, 287:429–439, June 2000.

[4] B Mandelbrot. A multifractal walk down wall street. *Scientific American*, pages 50–53, February 1999.

[5] L Molgedey and W Ebeling. Local order, entropy and predictability of financial time series. *European Physical Journal B - Condensed Matter and Complex Systems*, 15/4:733–737, June 2000.

[6] L Molgedey and W Ebeling. Intraday patterns and local predictability of high-frequency financial time series. *Physica A*, 287:420–428, June 2000.

[7] C A Zapart. On entropy, financial markets and minority games. *Physica A: Statistical Mechanics and its Applications*, 388(7):1157–1172, April 2009.

[8] E Moro. The minority game: an introductory guide. *Advances in Condensed Matter and Statistical Physics*, February 2004.

[9] C Chatfield. *The Analysis of Time Series.* Chapman & Hall/CRC, 2004.

[10] S Haykin. *Neural Networks.* Maxwell Macmillan, 1994.

[11] A P Dempster, N M Laird, and D B Rubin. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of Royal Statistical Society B*, 39:1–38, 1977.

[12] C R Reeves. *Modern Heuristic Techniques for Combinatorial Problems.* Halsted Press, 1993.

[13] J Voit. *The Statistical Mechanics of Financial Markets.* Springer, 2005.

[14] X Gabaix, P Gopikrishnan, V Plerou, and H E Stanley. A theory of power-law distributions in financial market fluctuations. *Nature*, 423:267–270, May 2003.

[15] D Challet and Y C Zhang. Emergence of cooperation and organisation in an evolutionary game. *Physica A: Statistical Mechanics and its Applications*, 246:407, September 1997.

[16] D Challet, M Marsili, and Y Zhang. *Minority Games: Interacting Agents in Financial Markets (Oxford Finance).* Oxford University Press, 2005.

[17] P Jeffries, M L Hart, P M Hui, and N F Johnson. From market games to real-world markets. *European Physical Journal B - Condensed Matter and Complex Systems*, pages 493–501, April 2001.

[18] D Challet, M Marsili, and Y C Zhang. Stylized facts of financial markets and market crashes in minority games. *Physica A: Statistical Mechanics and its Applications*, 294:514–524, 2001.

[19] M L Hart, P Jeffries, and N F Johnson. Dynamics of the time horizon minority game. *Advances in Condensed Matter and Statistical Physics*, February 2001.

[20] D Lamper, S Howison, and N F Johnson. Predictability of large future changes in a competitive evolving population. *Advances in Condensed Matter and Statistical Physics*, May 2001.

[21] Metzler R. *Neural Networks, Game Theory and Time Series Generation.* PhD thesis, cond-mat/0212486, 2002.

[22] Luca Grilli and Angelo Sfrecola. A neural networks approach to minority game. Quaderni DSEMS 13-2005, Dipartimento di Scienze Economiche, Matematiche e Statistiche, Universita' di Foggia, June 2005.

[23] Metzler R and Horn C. Evolutionary minority game: the benefits of imitation. *Advances in Condensed Matter and Statistical Physics*, oai:arXiv.org:cond-mat/0109166, December 2002.

[24] Pincus S. Approximate entropy as a measure of system complexity. In *Proceedings of the National Academy of Sciences of the USA*, volume 88, pages 2297–2301, 1991.

[25] Pincus S and Kalman R E. Irregularity, volatility, risk and financial market time series. In *Proceedings of the National Academy of Sciences of the USA*, volume 101 (38), pages 13709–13714, September 2004.