Eur. Phys. J. B **15**, 733–737 (2000)

THE EUROPEAN
PHYSICAL JOURNAL B
EDP Sciences

Local order, entropy and predictability of financial time series

L. Molgedey[a] and W. Ebeling

Institute of Physics, Humboldt–University Berlin, Invalidenstrasse 110, 10115 Berlin, Germany

Received 19 January 2000

Abstract. We consider time series of financial data as the Dow Jones Index with respect to the existence of local order. The basic idea is that in spite of the high stochasticity in average there might be special local situations where there local order exist and the predictability is considerably higher than in average. In order to check this assumption we discretise the time series and investigate the frequency of the continuation of definite words of length n first. We prove the existence of relatively long-range correlations under special conditions. The higher order Shannon entropies and the conditional entropies (dynamical entropies) are calculated, characteristic fluctuations are found. Instead of the dynamic entropies which yield mean values of the uncertainty/predictability we finally investigate the local values of the uncertainty/predictability and the distribution of these quantities.

PACS. 05.45.Tp Time series analysis – 02.50.Ey Stochastic processes – 65.50.+m Thermodynamic properties and entropy

1 Introduction

One of the basic problems of economy and finance is the predictability of future events. We know from our everyday experience that in spite of the complexity of these processes some individuals are more successful in predicting than others.

Within a large scale of uncertainty predictions are indeed possible and there exist good and bad methods to make a prediction. The analysis of financial time series from the point of predictability has attracted a lot of interest [1–6].

Usually one is interested in the prediction of frequent events on a short time horizon [4] or of the rare events (crashes, bubbles, anti–bubbles) on a longer time horizon [5]. Since predictability is far from being perfect one has to address the significance of the analysis. For rare events there is no methodology to deal with mispredictions [6].

Here we concentrate on predictability and significance on a daily time horizon using methods which are based on Shannons concept of information entropy [7].

We consider for simplicity one-dimensional time series of events in discrete state space and discrete time. Let us discuss several characteristic cases.

As the basic quantity for estimating predictability we study the local probability distribution and the Shannon entropies H for certain subtrajectories, in particular conditional (dynamical) entropies [8,9]. Assuming that an observation has provided us a certain trajectory of length n (an n–word), we may ask for the uncertainty of predicting the next state (letter). This is nothing else than the difference between the Shannon entropies for trajectories (words) of length $n+1$ and trajectories of length n:

$$h_n = H_{n+1} - H_n. \tag{1}$$

This conditional entropy measures the uncertainty of predicting a state one step in the future, given a history consisting of n states, *i.e.* the present state and the previous $n-1$ states is known [10]. Thus the estimation of Shannons n-gram entropies, which often are called block entropies for a series of word length n, is our basic problem. Predictability is measured in this work by differences of Shannon entropies, in other words by conditional entropies. The existence of long correlations is expressed by long decreasing tails of the conditional entropies h_n. In general our expectation is that any long-range memory decreases the conditional entropies and improves the chances for predictions.

2 Conditional entropy of financial time series

This section is devoted to the introduction of several basic terms stemming from information theory which were mostly used by Shannon already. Let us assume that the processes or structures to be studied are modelled by trajectories on discrete state spaces having the total length L. Let λ be the length of the alphabet. Further let $A_1 \dots A_n$ be the letters of a given subtrajectory of length $n \leq L$. Let further $p^{(n)}(A_1 \dots A_n)$ be the probability to find

[a] e-mail: lutz@molgedey.de

a block (subtrajectory) with the letters $A_1 \dots A_n$ in the total trajectory. Then we may introduce the entropy per block of length n (the n-gram entropy):

$$H_n = -\sum p^{(n)}(A_1 \dots A_n) \log_\lambda p^{(n)}(A_1 \dots A_n). \tag{2}$$

From the block entropies we derive conditional entropies (n-gram dynamic entropies) as the differences $h_n = H_{n+1} - H_n$. Further we define

$$r_n = 1 - h_n \tag{3}$$

as the average predictability of the state following after a measured n-trajectory. We remember that $h_n = 1$ is the maximum of the uncertainty (in units of $\log(\lambda)$), so the predictability is defined as the difference between the maximal and the actual uncertainty. In other words, predictability is the information we get by exploration of the next state in the future in comparison to the available knowledge.

The limit of the dynamic n-gram entropies for large n is the entropy of the source (called dynamic entropy or Kolmogorov–Sinai entropy). The predictability of processes is closely connected to these dynamic entropies. Let us consider a certain section of length n of the trajectory, a time series, or another sequence of symbols $A_1 \dots A_n$, which is often denoted as a subcylinder. We are interested in the uncertainty of the predictions of the state trailing this particular subtrajectory of length n. Considering the concepts of Shannon again we define now the expression

$$h_n^{(1)}(A_1 \dots A_n) = -\sum p(A_{n+1}|A_1 \dots A_n) \log p(A_{n+1}|A_1 \dots A_n) \tag{4}$$

as the conditional uncertainty of the next state (1 step into the future) following behind the measured trajectory $A_1 \dots A_n$ ($A_i \in$ alphabet). Henceforth all logarithms are measured in λ-units. We note that in these units the inequality holds:

$$0 \le h_n^{(1)}(A_1 \dots A_n) \le 1. \tag{5}$$

Further we define

$$r_n^{(1)}(A_1 \dots A_n) = 1 - h_n^{(1)}(A_1 \dots A_n) \tag{6}$$

as the predictability of the next state following after a measured subtrajectory, which is a quantity between zero and one. We note that the average of the local uncertainty

$$h_n = h_n^{(1)} = \left\langle h_n^{(1)}(A_1 \dots A_n) \right\rangle = \sum p(A_1 \dots A_n) h_n^{(1)}(A_1 \dots A_n)$$

leads us back to Shannons uncertainty (n-gram dynamic entropy). A possible generalisation concerns the case that we want to predict the state which follows not immediately after the observed n-string, but only after k steps into the future [11]. We define

$$h_n^{(k)}(A_1 \dots A_n) = -\sum p(A_{n+k}|A_1 \dots A_n) \log p(A_{n+k}|A_1 \dots A_n)$$

as the uncertainty of the state which occurs k steps into the future after the observation of an n-block, or symbolically

$$[A_1 \dots A_n](k-1 \text{ states})[A_{n+k} = ?].$$

Further we accordingly define the predictabilities

$$r_n^{(k)}(A_1 \dots A_n) = 1 - h_n^{(k)}(A_1 \dots A_n). \tag{7}$$

For $n = 1$ the predictability is closely related to the transinformation (mutual information) which may be expressed as [9]

$$I(k) = r_1^{(k)} + (h_0 - 1).$$

For systems with long memory it makes sense to study the whole series of predictabilities with increasing n-values (where n denotes the lower index)

$$r_1^{(k)}, r_2^{(k)}, r_3^{(k)}, \dots, r_m^{(k)}.$$

Here m is an estimate for the length of the memory. Due to the inequality

$$r_{n+1}^{(k)} \ge r_n^{(k)}$$

the average predictability may be improved by taking longer blocks into account. In other words, one can gain advantage for predictions by basing the predictions not only on actual states but on whole trajectory blocks which represent the actual state and its history.

2.1 Entropy analysis of financial time series

Prediction of strong noisy data using classical linear methods usually fails to give an accurate and reliable confidence level of the prediction. Moreover the linear methods are dominated by the most frequent events. However predictability may not be constant in time and even higher for seldom events. The concept of entropy and local predictability in combination with classical methods is a good candidate to give reliable results. Applications of these concepts to meteorological strings were given in [11,12] and to nerve signals in [13,16].

In the following our concept will be demonstrated on daily stock index data S_t: Dow Jones 1900-1999 (27 044 trading days). Since the stock index itself has an exponentially growing trend (see the inset of Fig. 1) one uses daily logarithmic price changes

$$x_t = \ln(S_t) - \ln(S_{t-1}). \tag{8}$$

A direct application of the entropy concept requires a partitioning of the real value data x_t into symbols A_t of an alphabet having the length λ. Finding an optimal partition and alphabet is a process of maximising the entropy converging to the Kolmogorov–Sinai entropy.

However for strong noisy signals with short memory an equal frequency of the letters is nearly optimal. For statistical reasons one would like to choose a small alphabet

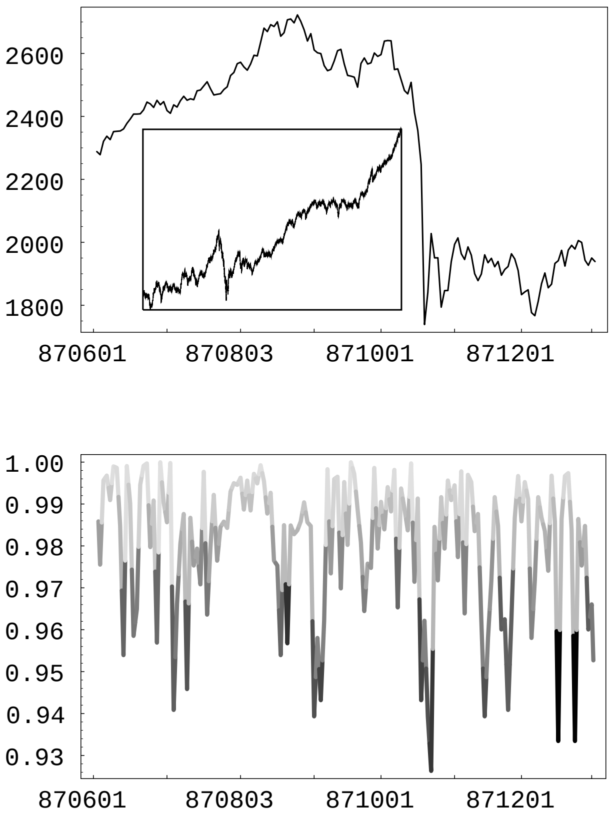

Fig. 1. Dow Jones (upper curve) and local uncertainty h_5 of the 6th symbol when seen 5 symbols (lower curve) for the second half of 1987. The inset shows the Dow Jones in a logarithmic scale for the full period 1900–1999. The greyvalue in the lower curve codes the level of significance calculated from surrogates with memory of 2. Dark represents a large deviation from the noise level (good significance). There is no trivial coherence between the price evolution (upper curve) and predictability (lower curve). However following a larger downturn the predictability is likely to increase (for instance the higher predictability following the October crash).

but a large alphabet for the backmapping of the predicted symbols A_{t+1} to the real values x_{t+1}.

To be concrete $\lambda = 3$ and $A_t = 0; x_t < -0.0025$ (strong decrease in the stock value), $A_t = 2; x_t > 0.0034$ (strong increase), $A_t = 1$ (intermediate) were chosen. With this partition the one symbol entropy is $H_1 = 1$ as well as the uncertainty without prior knowledge is $h_0 = 1$ by definition and one can discuss words up to 6 letters with statistical significance.

The asymmetry in the partition is due to the exponentially growing trend expressed in a positive mean logarithmic price change $\langle x \rangle = 0.0002$ and a small skew in the distribution of price changes.

The result of the calculation of the local uncertainty $h_n(A_1 \ldots A_n)$ for the next trading day following behind an observation of n trading days $A_1 \ldots A_n$ according to equation (4) for $n = 5$ is plotted in Figures 1 and 4. The local uncertainty is close to one, *i.e.* the local predictability is almost very small. The value one means that the conditional probabilities for all three symbols are identical, whereas values smaller than one mean that the three symbols have different conditional probabilities – *i.e.* some prediction is possible. Behind certain patterns of stock movements $A_1 \ldots A_n$ the local predictability reaches 8% – a notable value for the stock market, which is usually purely random. The mean predictability over the full data set is less then 2% (see Fig. 2).

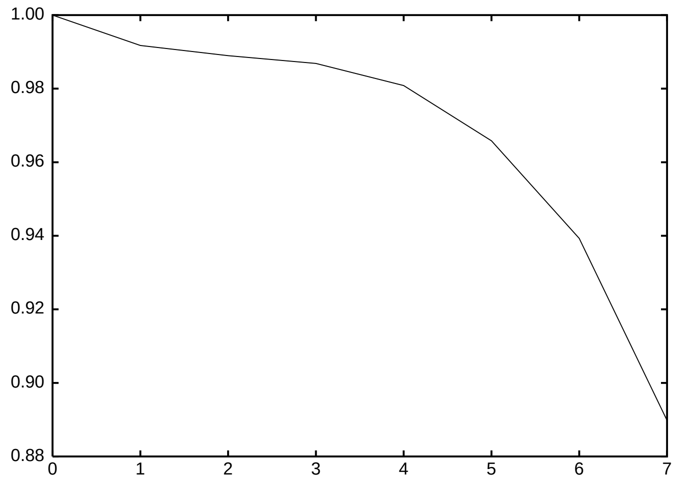

Fig. 2. Conditional entropy $h_n = H_{n+1} - H_n$ as a function of word length n. Beyond $n = 5$ the calculation of the conditional entropy is not reliable due to large statistical errors [8,9].

Looking at the in a time window averaged local uncertainty (Fig. 4) one finds certain periods of higher averaged predictabilities. In the last few decades the averaged predictability is relatively small – this could be connected to the beginning of computerised trading. However, a future increase of the averaged predictability seems to be expected during some crisis.

The significance's prediction is treated by calculating a distribution of local uncertainty $h_n^S(A_1 \ldots A_n)$ by help of surrogates [14–17]. Our surrogate sequences have the same two point probabilities $p^{(2)}(A_2|A_1)$ as the original sequence [15]. The level of significance K is calculated as

$$K_n(A_1 \ldots A_n) = \left| \frac{h_n(A_1 \ldots A_n) - \langle h_n^S(A_1 \ldots A_n) \rangle}{\sigma} \right|, \quad (9)$$

where $\langle h_n^S(A_1 \ldots A_n) \rangle$ is the mean and σ is the standard deviation of the local uncertainty distribution for the word $A_1 \ldots A_n$ [16,17].

Assuming Gaussian statistics $K \geq 2$ represents confidence greater then 95%. However the local uncertainty distribution (Fig. 3) is more exponential. Therefor larger K–values are required to guarantee significance. For the analysed data set a word length up to 5 seems to give reliable results.

Fortunately higher local predictabilities coincides with larger levels of significance as seen in Figure 1 and Table 1 (left).

Table 1. Left: words with the smallest uncertainty h_n (highest predictability $r_n = 1 - h_n$) have a good significance K. The significance K is decreasing with the word length n due to finite size effects. Right: we list the empirically observed relative frequencies of a larger downturn (0), a roughly constant market (1) or a larger upswing (3) from one trading day to the next, for a variety of histories (summarised by our words (absolute frequency)) of the preceding five trading days. These are the most predictable events from the left side.

word	h_3	K	word	h_4	K	word	h_5	K
020	0.971	27.9	1112	0.954	14.3	11110	0.919	9.5
112	0.971	30.5	0000	0.957	12.0	11120	0.926	9.1
110	0.977	23.4	1110	0.958	15.0	20000	0.926	6.9
120	0.981	29.4	0110	0.960	14.9	11112	0.931	8.7
000	0.982	18.5	0020	0.961	12.5	10120	0.933	5.2
212	0.983	19.5	1102	0.962	12.4	22202	0.933	9.4
202	0.984	22.9	2020	0.966	10.6	00000	0.934	6.4
111	0.985	20.1	0200	0.968	10.6	11011	0.937	9.7
121	0.985	12.4	0202	0.969	14.6	02000	0.939	5.7
012	0.987	14.2	0120	0.971	12.3	02020	0.941	5.1
102	0.988	10.5	2112	0.971	9.0	00020	0.943	6.4

word	a. freq.	r. freq. 0	r. freq. 1	r. freq. 2
11110	164	0.37	0.48	0.15
11120	71	0.41	0.44	0.15
20000	157	0.41	0.16	0.43
11112	153	0.16	0.42	0.42
10120	84	0.48	0.36	0.17
22202	144	0.29	0.21	0.50
00000	186	0.42	0.16	0.41
11011	144	0.22	0.51	0.27
02000	155	0.41	0.17	0.42
02020	87	0.51	0.21	0.29
00020	182	0.46	0.18	0.36

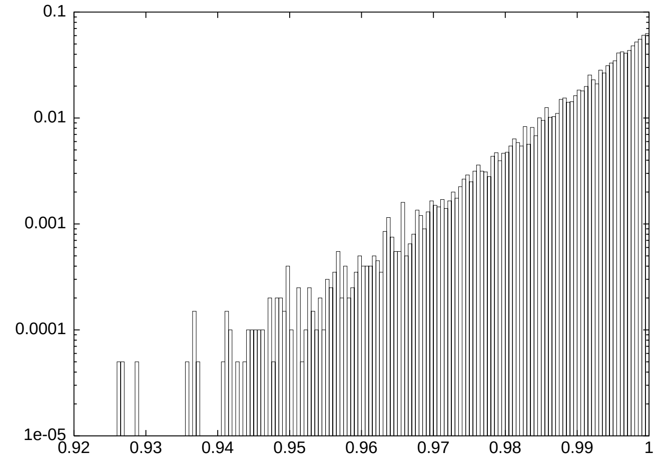

Fig. 3. Local uncertainty distribution of the surrogate sequence for the word 11110.

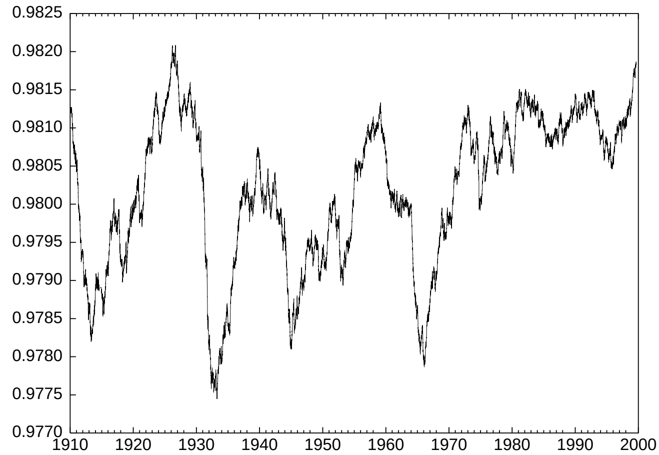

Fig. 4. Moving exponential average of the local uncertainty h_5 with an halflife period of 5 years for the full dataset.

For the prediction itself one can use the lookup Table 1 (right) obtained from the historical data. Most of the predictability is covered by the persistence of the volatility (fluctuations), *i.e.* following a period of large price movements (best seen for the words 20000, 00000 and 02000) a further strong movement is likely – but the direction of the movement (up or down) is unclear.

The interpretation is similar to that of weather forecasts: "tomorrows weather is likely to be the same as todays weather" [12] reads in the financial context "tomorrows volatility is likely to be the same as todays volatility". Improvements of this poor man's predictability need to consider longer histories.

According to the overall statistics assuming a Markovian process with a short memory one would expect nearly a symmetry between the probability of a strong increase (symbol 2) and the probability of a strong decrease (symbol 0) of the stock index following a certain pattern. These expectations are also covered by the models in the financial community – ARCH and GARCH [18]. Notably this symmetry is broken for most of the significant words. Deviations from the Markovian expectations improves the local predictability as well as the significance.

3 Conclusions

Our results show that local analysis is an appropriate tool for studying the predictability of financial time series. Of particular interest are local studies of the continuations and predictabilities of certain local histories. Local correlations are of specific interest since they improve the local predictability. Hence one can in principle improve the predictions at certain time instants by basing the predictions on observations of local histories.

Further we can conclude that there are specific substrings, which are relatively seldom, where the uncertainty is noticeably less than 1 – the predictability is better than 5%. In other words, there are specific situations where the predictability is better than the average predictability. However the effect is quite small and shows that the discussed financial time series is nearly random, but not fully random and shows some order at specific subtrajectories.

The authors thank R. Mantegna, F. Moss, A. Neiman and D. Stauffer for many fruitful discussions.

References

1. R.N. Mantegna, H.E. Stanley, *Scaling Approach to Finance* (Cambridge University Press, Cambridge, 2000).
2. S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, Y. Dodge, Nature **381**, 76 (1996).
3. S. Moss de Oliveira, P.M.C. de Oliveira, D. Stauffer, *Evolution, Money, War, and Computers* (Teubner, Stuttgart, 1999).
4. Y.C. Zhang, Physica A **269**, 30 (1999).
5. D. Sornette, Phys. Rep. **297**, (1998); A. Johansen, D. Sornette, Int. J. Mod. Phys. C **10**, 563 (1999).
6. L. Laloux, M. Potters, R. Cont, J.-P. Aguilar, J.-P. Bouchaud, Europhys. Lett. **45**, 1 (1999).
7. C. Shannon, Bell Systems Tech. **30**, 50 (1951).
8. W. Ebeling, G. Nicolis, Europhys. Lett. **14**, 191 (1991); Chaos, Solitons & Fractals **2**, 635 (1992).
9. H. Herzel, W. Ebeling, A.O. Schmitt, Phys. Rev. E **50**, 5061 (1994); Chaos, Solitons & Fractals **4**, 97 (1994).
10. W. Ebeling, G. Nicolis, Chaos, Solitons & Fractals **2**, 635 (1992).
11. C. Nicolis, W. Ebeling, C. Baraldi, Tellus **49A**, 10 (1997).
12. P.C. Werner, F.-W. Gerstengarbe, W. Ebeling, Theor. Appl. Climatol. **62**, 125 (1999).
13. W. Ebeling, M.A. Jimenez-Montano, T. Pohl, *Entropy and Complexity od Sequences, Festschrift devoted to Jagat N. Kapur* (New Delhi, 1999).
14. E.N. Trifonov, V. Brendel, *Gnomic – A Dictionary of Genetic Codes* (VCH, Weinheim, 1987).
15. A.O. Schmitt, W. Ebeling, H. Herzel, Biosystems **37**, 199 (1996).
16. X. Pei, F. Moss, Nature **379**, 618 (1996).
17. K. Dolan, A. Witt, M.L. Spano, A. Neiman, F. Moss, Phys. Rev. E **59**, 5235 (1999).
18. T. Bollerslev, R.Y. Chou, K.F. Kroner, J. Econometrics **52**, 5 (1992).