Intraday Patterns and Local Predictability of High Frequency Financial Time Series

Lutz Molgedey [1] Werner Ebeling [2]

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

Abstract

The structure of high frequency time series of financial data taking the DAX future as an example is investigated with respect to the existence of local order on a time horizon of a few minutes. We will show that there might be special local situations where there exist local order, and where the predictability is considerably higher than in average. We discretize the time series and investigate the continuation frequency of definite words of length n first. Besides higher order Shannon entropies and conditional entropies (dynamic entropies) which yield mean values of the uncertainty/predictability we study the local values of the uncertainty/predictability and the distribution of these quantities. The local order significance is treated by means of surrogate sequences with identical short memory as the original data.

Key words: local predictability; entropy; symbol dynamics
PACS: 05.45.Tp, 02.50.Ey, 65.50.+m

1 Introduction

The prediction of future financial events is an important task and some individuals are more successful in predicting than others. Evidently there exist good and bad methods to make a prediction. The analysis of this problem has attracted a lot of interest [1–8].

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 [5–7]. 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 [8].

[1] lutz@molgedey.de
[2] ebeling@physik.hu-berlin.de

Here we concentrate on predictability and significance on an intraday time horizon using methods based on Shannon's concept of information entropy [9].

2 Conditional Entropy of Financial Time Series

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 [10,11]. 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 but the difference between the Shannon entropies for trajectories (words) of length $n+1$ and trajectories of length n:

$$\bar{h}_n = H_{n+1} - H_n \quad . \tag{1}$$

This conditional entropy (mutual information) measures the uncertainty of predicting a state one step into the future, given a history consisting of n states; i.e. the present state and the previous $n-1$ states are known [12]. Thus, the estimation of Shannon's n-gram entropies (block entropies) for a series of word length n is our basic problem. Predictability in this work is measured 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 $\bar{h}_n$. In general our expectation is that any long-range memory decreases the conditional entropies and improves the chances for predictions.

Let $A_1 \ldots A_n$ be the letters of a given subtrajectory of length $n \leq L$. The length of the alphabet is λ. Let further $p(A_1 \ldots A_n)$ be the probability to find a block (subtrajectory) with the letters $A_1 \ldots A_n$ in the total trajectory. The entropy per block of length n (the n-gram entropy) is defined as

$$H_n = - \sum_{\{A_1 \ldots A_n\}} p(A_1 \ldots A_n) \log_\lambda p(A_1 \ldots A_n) \quad , \tag{2}$$

where the sum is over all λ^n possible realizations $\{A_1 \ldots A_n\}$.

From the block entropies H_n we derive conditional entropies $\bar{h}_n$ (n-gram dynamic entropies) as the differences $\bar{h}_n = H_{n+1} - H_n$.

The maximum of the uncertainty (in units of $\log(\lambda)$) is $\bar{h}_n = 1$. Hence one can define the averaged predictability as the difference between the maximal and the actual uncertainty

$$\bar{r}_n = 1 - \bar{h}_n \quad . \tag{3}$$

In other words, predictability is related to the certainty that we have about 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 (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 $A_1 \ldots A_n$, 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. Extending the concepts of Shannon, the expression

$$h_n(A_1 \ldots A_n) = - \sum_{\{A_{n+1}\}} p(A_{n+1}|A_1 \ldots A_n) \log_\lambda p(A_{n+1}|A_1 \ldots A_n) \quad (4)$$

defines the next state's conditional uncertainty (1 step into the future) following the measured trajectory $A_1 \ldots A_n$ ($A_i \in$ alphabet). We note that in these units the inequality holds

$$0 \leq h_n(A_1 \ldots A_n) \leq 1 \quad . \quad (5)$$

The average of the local uncertainty

$$\bar{h}_n = \langle h_n(A_1 \ldots A_n) \rangle = \sum_{\{A_1 \ldots A_n\}} p(A_1 \ldots A_n) h_n(A_1 \ldots A_n)$$

leads us back to Shannon's uncertainty (n-gram dynamic entropy). Further we define

$$r_n(A_1 \ldots A_n) = 1 - h_n(A_1 \ldots A_n) \quad (6)$$

as the next state's predictability following a measured subtrajectory, which is a quantity between zero and one.

3 Entropy analysis of Financial Time Series

Our concept was previously demonstrated on meteorological strings [13,14], on nerve signals [15,16] and daily stock index data [17]. To calculate higher order entropies one needs a long sequence where one has to assume stationarity. In the case of daily data over several decades the stationarity assumption is problematic. The situation is much more comfortable for high frequency data.

In the following we employ tick by tick data of the german DAX future 1998/06/19 – 1998/12/18. The data are resampled first to equidistant prices S_t of 2 minutes. In the analyzed data, it is guaranteed that at least one trade has occurred in each 2 minute time window.

As for daily data we use logarithmic price changes

$$x_t = \ln(S_t) - \ln(S_{t-1}) \quad (7)$$

where the time unit is 2 minutes.

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 maximizing the entropy converging to the Kolmogorov–Sinai entropy.

However for strong noisy signals with short memory an equal letter frequency is nearly optimal.

One would like to choose a small alphabet in order to have a small statistical error in the calculation of the entropies. On the other hand a large alphabet is required for the backmapping of the predicted symbols A_{t+1} to the real values x_{t+1}.

To be concrete $\lambda = 3$ and $A_t = 0; x_t < -0.000263$ (strong decrease in the stock value), $A_t = 2; x_t > 0.000271$ (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 $\bar{h}_0 = 1$ by definition and one can discuss words up to 5 letters with statistical significance.

The small asymmetry in the partition is mainly due to a small skew in the distribution of price changes.

$p^{(2)}(\cdot\vert\cdot)$	$0\vert$	$1\vert$	$2\vert$
$\vert 0$	0.451	0.296	0.253
$\vert 1$	0.291	0.402	0.307
$\vert 2$	0.257	0.302	0.441

Table 1. Conditional probability $p^{(2)}(A_2|A_1)$ of the discretized price changes.

The price changes are weakly autocorrelated on a 2 minute time scale. This can also be seen from the two time conditional probability (frequency) listed in Table 1.

The result of the local uncertainty $h_n(A_1 \ldots A_n)$ for the next 2 minutes following after a pattern of n points $A_1 \ldots A_n$ according to equation (4) for $n = 5$ is plotted in Fig. 1. The local uncertainty is close to one, i.e. the local predictability is mostly very small. The value 1 means that the conditional probabilities for all three symbols are identical, whereas values

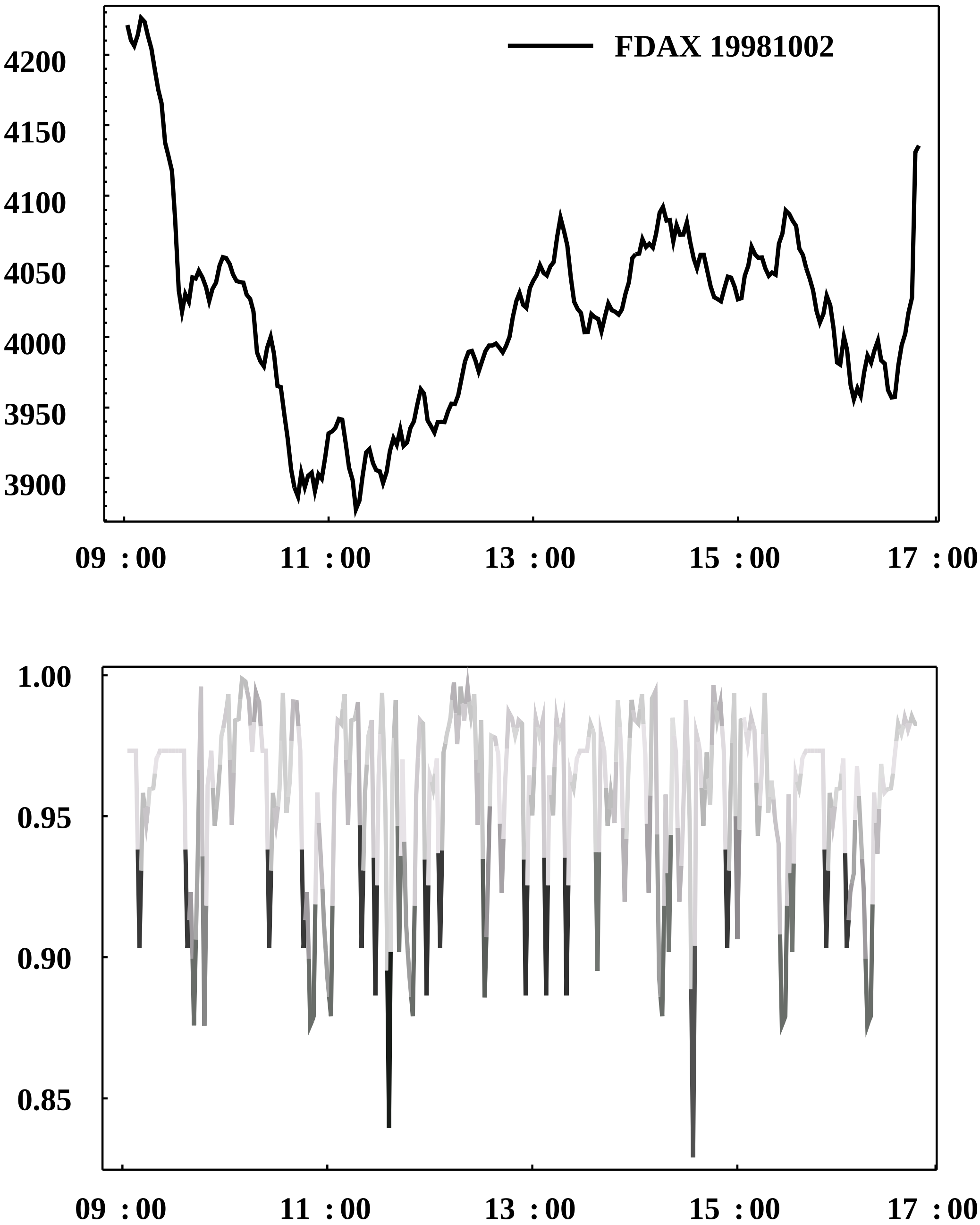

Fig. 1. DAX Future (upper curve) and local uncertainty h_5 of the prediction of the 6th symbol based on the 5 preceding symbols (lower curve) for a trading day with large fluctuations are shown. The greyvalue in the lower curve codes the level of significance calculated from surrogates with memory of one. 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).

smaller than 1 mean that the three symbols have different conditional probabilities – i.e. some prediction is possible. After certain patterns of stock movements $A_1 \ldots A_n$ the local predictability reaches 17%. This is a notable value for the stock market, which is usually purely random. The mean predictability over the full data

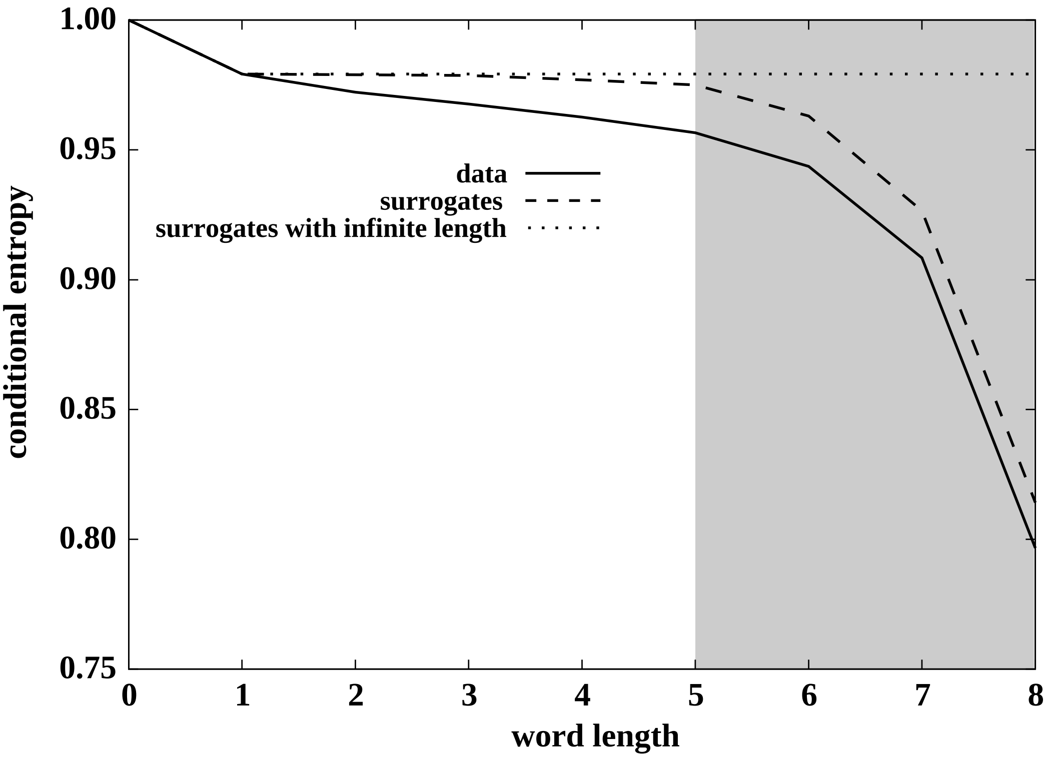

Fig. 2. Conditional entropy (uncertainty) $\bar{h}_n = H_{n+1} - H_n$ as a function of word length n. The uncertainties for the original data (solid curve) are always smaller than those calculated from the surrogate sequences (dashed curve) of the same length. The surrogates have a memory of one, i.e. for infinite surrogate sequences the uncertainties would be constant for $n \geq 1$ (dotted curve). Beyond $n = 5$ (grey region) the calculation of the conditional entropy is not reliable due to large statistical errors (finite length effects) [10,11].

set is approximately 3–4% (see Fig. 2).

The prediction significance is treated by calculating a distribution of local uncertainty $h_n^S(A_1 ... A_n)$ by means of surrogates [16,18–20]. Our surrogate sequences have the same two point probabilities $p^{(2)}(A_2|A_1)$ as the original sequence (Table 1) [19].

This short time Markovian memory explains at least half of the averaged predictability (Fig. 2) and accounts for effects such as persistence of the volatility (large fluctuations are likely followed bye large one) [21]. However there is no special reason to choose a memory of one for the surrogate sequences besides the large decrease in the conditional entropy $\bar{h}_n$ for $n = 1$. Furthermore the surrogate memory has to be shorter than the considered local histories.

The level of significance K is calculated as the difference of the measured local uncertainty $h_n(A_1 ... A_n)$ and the mean local uncertainty $\langle h_n^S(A_1 ... A_n)\rangle$ of the

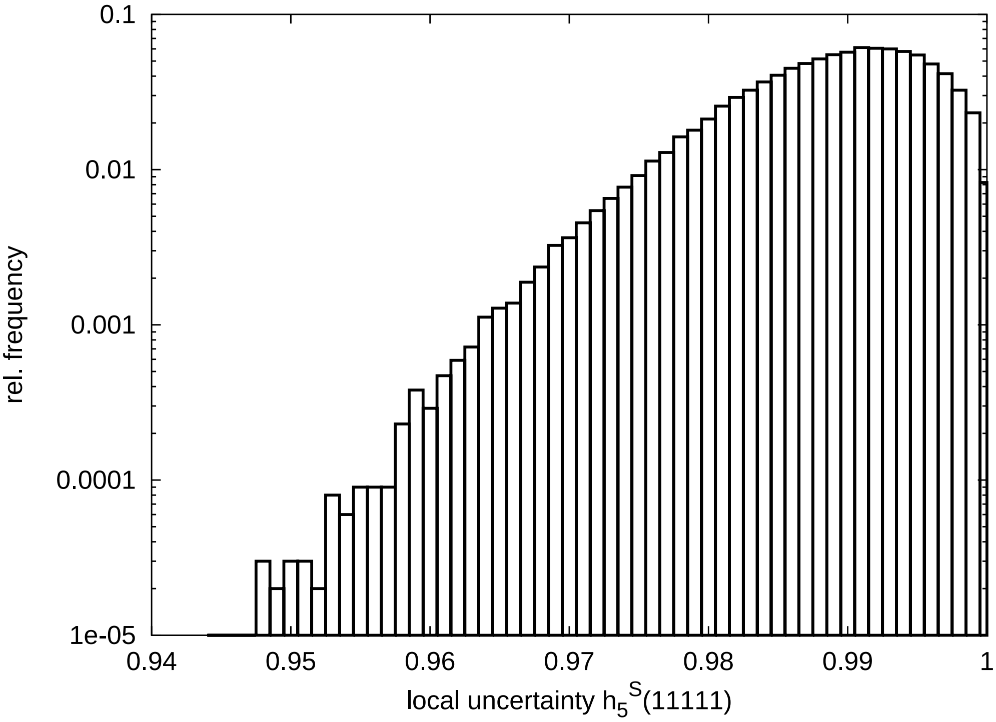

Fig. 3. Local uncertainty distribution of the surrogate sequence for the word 11111.

surrogate sequences in standard deviation units [16,20]:

$$K_n(A_1 \ldots A_n) = \frac{\left| h_n(A_1 \ldots A_n) - \langle h_n^S(A_1 \ldots A_n) \rangle \right|}{\sqrt{\langle h_n^S(A_1 \ldots A_n)^2 \rangle - \langle h_n^S(A_1 \ldots A_n) \rangle}} \quad , \tag{8}$$

where $\langle \cdots \rangle$ denotes the surrogate ensemble average.

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

Fortunately higher local predictabilities coincide with larger levels of significance as seen in Fig. 1 and Table 2.

For the prediction itself one can use the lookup Table 3 obtained from the historical data.

For daily data a probability symmetry for up and down movements was observed. This was connected to the persistence of the volatility (volatility clustering), i.e. large fluctuations were followed by large fluctuations.

Here, on a short time scale of 2 minutes most of the predictability is covered by the

word	h_3	K	word	h_4	K	word	h_5	K
102	0.925	7.8	1111	0.892	21.3	20120	0.829	5.1
111	0.928	21.1	0202	0.898	5.1	02020	0.830	4.1
120	0.935	5.3	0102	0.899	5.9	11102	0.834	8.4
110	0.937	6.0	1110	0.903	7.8	00102	0.839	7.0
020	0.938	4.4	2220	0.912	6.7	11111	0.854	19.1
112	0.940	7.1	2120	0.914	4.2	21102	0.862	5.8
220	0.943	4.9	1112	0.916	7.3	10212	0.871	4.4
002	0.947	5.4	2020	0.918	3.4	02120	0.876	3.3
210	0.954	2.5	0120	0.920	3.8	00202	0.876	4.1
202	0.955	2.7	1102	0.921	5.2	02022	0.879	4.1
012	0.961	2.5	0112	0.929	4.9	11110	0.881	6.3

Table 2
Words with the smallest uncertainty h_n (highest predictability $r_n = 1 - h_n$) have a good significance K. The significance K is on average decreasing with the word length n due to finite length effects.

continuation of the last letter (trend following). A small preference for the continuation is already included in the two point conditional probability $p^{(2)}(A_2|A_1)$ from Table 1, but for the most significant word this preference is amplified.

4 Out of Sample Performance Analysis

Surrogate based prediction significance analysis is always an in sample method assuming stationarity of the time series. In order to test that a prediction which was found to be significant in the sample is still significant out of the sample, we divide our data into a training set (first two thirds) and a test set (last third).

On the training set a prediction table including significance similar to Table 2 and 3 was built up. Then on the test set the following performance analysis was carried out. At each time step a prediction of the next symbol based on the local history was performed. If the in sample prediction significance had exceeded a threshold K, the prediction was compared with the realization. If the prediction was correct, a value of 2 was added to the performance. If the prediction was wrong, a value of 1 was subtracted from the performance. Since the frequency of the symbols is $1/3$, the mean performance without benefits from the predictions should be zero. If all predictions were correct the performance value per decision would be 2.

Trusting only the predictions found to be significant in the sample is evidently

word	a. freq.	r. freq. 0	r. freq. 1	r. freq.2
20120	68	0.64	0.18	0.18
02020	46	0.66	0.17	0.17
11102	84	0.07	0.49	0.44
00102	146	0.19	0.18	0.63
11111	810	0.18	0.61	0.21
21102	70	0.20	0.19	0.61
10212	71	0.10	0.41	0.49
02120	70	0.60	0.20	0.20
00202	128	0.19	0.22	0.59
02022	126	0.26	0.16	0.58
11110	297	0.38	0.50	0.12

Table 3
We list the empirically observed relative frequencies of a larger downturn (0), a roughly constant market (1) and a larger upswing (2) for the next 2 trading minutes, for a variety of histories (summarized by our words (absolute frequency)) of the preceding five symbols (10 minutes). These are the most predictable events from Table 2.

improving the performance per decision as shown in Fig. 4. However the number of decisions, i.e. trading possibilities, decreases exponentially with larger significance level.

5 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 local history observations.

Further we can conclude that there are specific substrings which rarely occur and for which the uncertainty is noticeably less than 1: the local predictability is better than 10%. 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 Rosario Mantegna, Frank Moss, Alexander Neiman and Dietrich

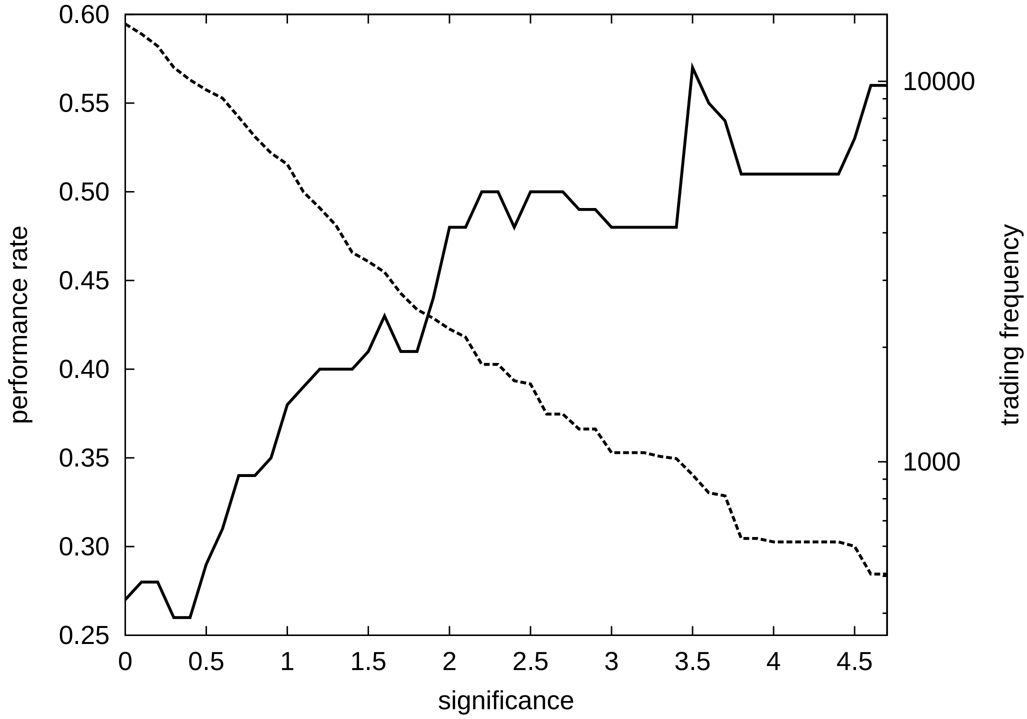

Fig. 4. Out of sample performance analysis (solid line) as a function of the minimal desired in sample prediction significance K is shown. The total number of out of sample predictions (dashed line) having an in sample significance value larger than K is a nearly exponentially decreasing function of K.

Stauffer for many fruitful discussions and the EUREX–exchange for providing high frequency time series.

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] T. Lux, The Economic Journal **105**, 881 (1995).

[6] D. Sornette, Phys. Rep. **297**, 297 (1998)

[7] A. Johansen, D. Sornette, Int. J. Mod. Phys. **C 10**, 563 (1999).

[8] L. Laloux, M. Potters, R. Cont, J.-P. Aguilar, J.-P. Bouchaud, Europhys. Lett. **45**, 1 (1999).

[9] C. Shannon, Bell Systems Tech. **30**, 50 (1951).

[10] W. Ebeling, G. Nicolis, Europhys. Lett. **14**, 191 (1991); Chaos, Solitons & Fractals **2**, 635 (1992).

[11] H. Herzel W. Ebeling, A. O. Schmitt, Phys. Rev. **E 50**, 5061 (1994); Chaos, Solitons & Fractals **4**, 97 (1994).

[12] W. Ebeling, G. Nicolis, Chaos, Solitons & Fractals **2**, 635 (1992).

[13] C. Nicolis, W. Ebeling, C. Baraldi, Tellus **49A**, 10 (1997).

[14] P. C. Werner, F.-W. Gerstengarbe, W. Ebeling, Theor. Appl. Climatol. **62**, 125 (1999).

[15] W. Ebeling, M.A. Jimenez-Montano, T. Pohl, *Entropy and Complexity od Sequences, Festschrift devoted to Jagat N. Kapur*, New Delhi (1999).

[16] X. Pei, F. Moss, Nature **379**, 618 (1996).

[17] L. Molgedey, W. Ebeling, Eur. Phys. J. **B 15**, 733 (2000).

[18] E. N. Trifonov, V. Brendel, *Gnomic – A Dictionary of Genetic Codes*, VCH, Weinheim (1987).

[19] A. O. Schmitt, W. Ebeling, H. Herzel Biosystems **37**, 199 (1996).

[20] K. Dolan, A. Witt, M. L. Spano, A. Neiman, F. Moss, Phys. Rev. **E 59**, 5235 (1999).

[21] A. Arneodo, J.-F. Muzy, D. Sornette, Eur. Phys. J. **B 2**, 277 (1998)