Physica A 392 (2013) 6414–6426

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

The string prediction models as invariants of time series in the forex market

R. Pincak*

Institute of Experimental Physics, Slovak Academy of Sciences, Watsonova 47, 043 53 Kosice, Slovak Republic
Bogoliubov Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, 141980 Dubna, Moscow Region, Russia

HIGHLIGHTS

- We apply a new approach of string theory to the real financial market.
- The models are constructed with an idea of prediction based on string invariants.
- The models are compared to support vector machines and artificial neural networks.
- Comparisons were done on artificial and a financial time series.
- The presented string models could be useful for portfolio creation and risk management.

ARTICLE INFO

Article history:
Received 13 May 2013
Available online 9 August 2013

Keywords:
Finance forex market
Nonlinear statistics
String theory
Trading strategy
Financial forecasting

ABSTRACT

In this paper we apply a new approach of string theory to the real financial market. The models are constructed with an idea of prediction models based on the string invariants (PMBSI). The performance of PMBSI is compared to support vector machines (SVM) and artificial neural networks (ANN) on an artificial and a financial time series. A brief overview of the results and analysis is given. The first model is based on the correlation function as invariant and the second one is an application based on the deviations from the closed string/pattern form (PMBCS). We found the difference between these two approaches. The first model cannot predict the behavior of the forex market with good efficiency in comparison with the second one which is, in addition, able to make relevant profit per year. The presented string models could be useful for portfolio creation and financial risk management in the banking sector as well as for a nonlinear statistical approach to data optimization.

1. Introduction

Time series forecasting is a scientific field under continuous active development covering an extensive range of methods. Traditionally, linear methods and models are used. Despite their simplicity, linear methods often work well and may well provide an adequate approximation for the task at hand and are mathematically and practically convenient. However, the real life generating processes are often non-linear. This is particularly true for financial time series forecasting. Therefore the use of non-linear models is promising. Many observed financial time series exhibit features which cannot be explained by a linear model.

There are plenty of non-linear forecast models based on different approaches (e.g. GARCH [1], ARCH [2], ARMA [3], ARIMA [4] etc.) used in financial time series forecasting. Currently, perhaps the most frequently used methods are based on Artificial Neural Networks (ANN, which covers a wide range of methods) and Support Vector Machines (SVM). A number of

* Correspondence to: Institute of Experimental Physics, Slovak Academy of Sciences, Watsonova 47, 043 53 Kosice, Slovak Republic. Tel.: +421 907936147; fax: +421 55 633 62 92.
E-mail addresses: pincak@saske.sk, pincak@sors.com.

0378-4371/$ – see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.physa.2013.07.048

research articles compare ANN and SVM to each other and to other more traditional non-linear statistical methods. Tay and Cao [5] examined the feasibility of SVM in financial time series forecasting and compared it to a multilayer Back Propagation Neural Network (BPNN). They showed that SVM outperforms the BP neural network. Kamruzzaman and Sarker [6] modeled and predicted currency exchange rates using three ANN based models and a comparison was made with the ARIMA model. The results showed that all the ANN based models outperform the ARIMA model. Chen et al. [7] compared SVM and BPNN taking the auto-regressive model as a benchmark in forecasting the six major Asian stock markets. Again, both the SVM and BPNN outperformed the traditional models.

While the traditional ANN implements the empirical risk minimization principle, SVM implements the structural risk minimization [8]. Structural risk minimization is an inductive principle for model selection used for learning from finite training data sets. It describes a general model of capacity control and provides a trade-off between hypothesis space complexity and the quality of fitting the training data (empirical error). For this reason SVM is often chosen as a benchmark to compare other non-linear models to. Also, there is a growing number of novel and hybrid approaches, combining the advantages of various methods using for example evolutionary optimization, methods of computational geometry and other techniques (e.g. Refs. [9,10]).

In this paper we apply the string model and approaches described in Ref. [11] to the real finance forex market. This is an extension of the previous work [11] into the real finance market. We derive two models for predictions of EUR/USD prices on the forex market. This is the first attempt for real application of string theory in the field of finance, and not only in high energy physics, where it is established very well. Firstly we describe briefly some connections between these different fields of research.

We would like to transfer modern physics ideas into the neighboring field called econophysics. The physical statistical viewpoint has proved to be fruitful, namely in the description of systems where many-body effects dominate. However, the standard, accepted by physicists, bottom-up approaches are cumbersome or outright impossible to follow the behavior of complex economic systems, where autonomous models encounter intrinsic variability.

The modern digital economy is founded on data. Our primary motivation comes from the actual physical concepts [12,13]; however, our realization differs from the original attempts in various significant details. Similarly as with most scientific problems, the representation of data is the key to efficient and effective solutions. The string theory development over the past 25 years has achieved a high degree of popularity among physicists [14].

The underlying link between our approach and string theory may be seen in switching from a local to a non-local form of data description. This line passes from the single price to the multivalued collection, especially the string of prices from the temporal neighborhood, which we term here as the string map. It is the relationship between more intuitive geometric methods and financial data. Here we work on the concept that is based on projection data into higher dimensional vectors in the sense of the works [15,16].

The present work exploits time series which can build the family of string-motivated models of boundary-respecting maps. The purpose of the present data-driven study is to develop statistical techniques for the analysis of these objects and moreover for the utilization of such string models onto the forex market. Both of the string prediction models in this paper are built on the physical principle of the invariance in time series of the forex market. Founding of a stationary state in the time series of the market was studied in Ref. [17].

2. Definition of the strings

By applying standard methodologies of detrending we suggest to convert the original series of quotations of the mean currency exchange rate $p(\tau)$ onto a series of returns defined by

$$\frac{p(\tau+h)-p(\tau)}{p(\tau+h)}, \tag{1}$$

where h denotes a tick lag between currency quotes $p(\tau)$ and $p(\tau+h)$, τ is the index of the quote. The mean $p(\tau) = (p_{\text{ask}}(\tau)+p_{\text{bid}}(\tau))/2$ is calculated from $p_{\text{ask}}(\tau)$ and $p_{\text{bid}}(\tau)$.

In the spirit of string theory it would be better to start with the 1-end-point open string map

$$P^{(1)}(\tau,h) = \frac{p(\tau+h)-p(\tau)}{p(\tau+h)}, \quad h \in \langle 0, l_s \rangle \tag{2}$$

where the superscript (1) refers to the number of endpoints.

The variable h may be interpreted as a variable which extends along the extra dimension limited by the string size l_s. For the natural definitions of the string to be fulfilled the boundary condition

$$P^{(1)}(\tau,0) = 0, \tag{3}$$

holds for any tick coordinate τ. We want to highlight the effects of rare events. For this purpose, we introduce a power-law Q-deformed model

$$P_q^{(1)}(\tau,h) = \left(1 - \left[\frac{p(\tau)}{p(\tau+h)}\right]^Q\right), \quad Q > 0. \tag{4}$$

The 1-end-point string has defined the origin, but it reflects the linear trend in $p(\cdot)$ at the scale l_s. Therefore, the 1-end-point string map $P_q^{(1)}(\cdot)$ may be understood as a Q-deformed generalization of the *currency returns*.

The situation with a long-term trend is partially corrected by fixing $P_q^{(2)}(\tau, h)$ at $h = l_s$. The open string with two end points is introduced via the nonlinear map which combines information about trends of p at two sequential segments

$$P_q^{(2)}(\tau, h) = \left(1 - \left[\frac{p(\tau)}{p(\tau + h)}\right]^Q\right)\left(1 - \left[\frac{p(\tau + h)}{p(\tau + l_s)}\right]^Q\right), \quad h \in \langle 0, l_s\rangle. \tag{5}$$

The map is suggested to include boundary conditions of *Dirichlet type*

$$P_q^{(2)}(\tau, 0) = P_q(\tau, l_s) = 0, \quad \text{at all ticks } \tau. \tag{6}$$

In particular, the sign of $P_q^{(2)}(\tau, h)$ comprises information about the behavior differences of $p(\cdot)$ at three quotes $(\tau,\ \tau + h,\ \tau + l_s)$.

Now we define partially compactified strings. In the frame of string theory, compactification attempts to ensure compatibility of the universe based on the four observable dimensions with twenty-six dimensions found in the theoretical model systems. From the standpoint of the problems considered here, compactification may be viewed as an act of information reduction of the original signal data, which makes the transformed signal periodic. Of course, it is not very favorable to close strings by the complete periodization of real input signals. Partial closure would be more interesting. This uses pre-mapping

$$\tilde{p}(\tau) = \frac{1}{N_m}\sum_{m=0}^{N_m-1} p(\tau + l_s m), \tag{7}$$

where the input of any open string (see e.g. Eqs. (2), and (5)) is made up partially compact.

Thus, data from the interval $\langle \tau, \tau + l_s(N_m - 1)\rangle$ are being pressed to occupy "little space" $h \in \langle 0, l_s\rangle$. We see that as N_m increases, deviations of $\tilde{p}$ from the periodic signal become less pronounced. The corresponding statistical characteristics of all the strings and branes described above were displayed in detail in Ref. [11]. The prediction models presented in the paper were tested on the tick by tick one year data of EUR/USD major currency pair from the ICAP market maker. More precisely, we selected the period from October 2009 to September 2010.

3. Correlation function as invariant

The meaning of invariant is that something does not change under transformation, such as some equations from one reference frame to another. We want to extend this idea also on the finance market, find some invariants in the finance data and utilize this as the prediction for the following prices. Unfortunately this model is able to define only one step prediction, see the definition below.

We suppose the invariant is in a form of correlation function

$$C_{(t,l_0)} = \sum_{h=l_0}^{h=l} w_h \left(1 - \frac{p_{t-h}}{p_{t-1-h}}\right)\left(1 - \frac{p_{t-1-h}}{p_{t-2-h}}\right), \tag{8}$$

with

$$w_h = \frac{e^{-h/\lambda}}{\sum_{h'=0}^{l} e^{-h'/\lambda}}, \tag{9}$$

including dependence on the time scale parameters l, l_0 and λ. The relative weights satisfy automatically $\sum_{h=0}^{l} w_h = 1$.

A correlation function is a statistical correlation between random variables at two different points, in our case the strings in time series. For simplicity as an example we used only one point strings equation (4) with parameter $Q = 1$. Ordinary the correlation function is defined as $C(\tau, l_0) = \langle P^1(\tau, l_0)P^1(\tau + 1, l_0)\rangle$. We suppose the invariant in the form of the correlation function

$$C(\tau, l_0) = \sum_{h=l_0}^{h=l} W(h)\left(1 - \frac{p(\tau - h)}{p(\tau - 1 - h)}\right)\left(1 - \frac{p(\tau - 1 - h)}{p(\tau - 2 - h)}\right), \tag{10}$$

with weight $W(h)$ defined above. We assume the condition of the invariance between close strings in τ and at the next step $\tau + 1$ in time series (It is the exact meaning of the one step prediction) in the form

$$C(\tau, l_0) = C(\tau + 1, l_0). \tag{11}$$

Now we want to find the exact expression for the one step prediction $p(\tau+1)$. Therefore we evaluate one step correlation invariant equation (11) with initial condition $l_0 = 0$

$$W(0)\left(1-\frac{p(\tau)}{p(\tau-1)}\right)\left(1-\frac{p(\tau-1)}{p(\tau-2)}\right) = W(0)\left(1-\frac{p(\tau+1)}{p(\tau)}\right)\left(1-\frac{p(\tau)}{p(\tau-1)}\right) + W(1)\left(1-\frac{p(\tau)}{p(\tau-1)}\right)\left(1-\frac{p(\tau-1)}{p(\tau-2)}\right), \quad (12)$$

which can be rewritten in the more compact form

$$C(\tau,0) = W(0)\left(1-\frac{p(\tau+1)}{p(\tau)}\right)\left(1-\frac{p(\tau)}{p(\tau-1)}\right) + C(\tau+1,1) \quad (13)$$

and

$$\left(1-\frac{p(\tau+1)}{p(\tau)}\right) = \frac{C(\tau,0)-C(\tau+1,1)}{W(0)\left(1-\frac{p(\tau)}{p(\tau-1)}\right)}. \quad (14)$$

We finally obtain the prediction

$$p(\tau+1) = p(\tau)\left(1+\frac{C(t+1,1)-C(t,0)}{W(0)\left(1-\frac{p(\tau)}{p(\tau-1)}\right)}\right), \quad (15)$$

valid for $p(\tau) \neq p(\tau-1)$. These are general definitions for the one step prediction correlation invariants. In the next section similar equations can be found also for 2-end-point and 1-end-point mixed string models with $Q > 0$.

3.1. Prediction model based on the string invariants (PMBSI)

Now we want to take the above-mentioned ideas onto the string maps of finance data. We would like to utilize the power of the nonlinear string maps of finance data and establish some prediction models to predict the behavior of the market similarly as in the works [18–20]. We suggest the method where one string is continuously deformed into the other. We analyze 1-end-point and 2-end-point mixed string models. The family of invariants is written using the parameterization

$$C(\tau,\Lambda) = (1-\eta_1)(1-\eta_2)\sum_{h=0}^{\Lambda} W(h) \quad (16)$$

$$\times\left(1-\left[\frac{p(\tau)}{p(\tau+h)}\right]^Q\right)\left(1-\left[\frac{p(\tau+h)}{p(\tau+l_s)}\right]^Q\right) + \eta_1(1-\eta_2)\sum_{h=0}^{\Lambda} W(h)\left(1-\left[\frac{p(\tau)}{p(\tau+h)}\right]^Q\right) + \eta_2\sum_{h=0}^{\Lambda} W(h)\left(1-\left[\frac{p(\tau+h)}{p(\tau+l_s)}\right]^Q\right), \quad (17)$$

where $\eta_1 \in (-1, 1)$, $\eta_2 \in (-1, 1)$ are variables (variables which we may call homotopy parameters), Q is a real valued parameter, and the weight $W(h)$ is chosen in the bimodal single parameter form

$$W(h) = \begin{cases} 1-W_0, & h \le l_s/2, \\ W_0, & h > l_s/2. \end{cases} \quad (18)$$

We plan to express $p(\tau+l_s)$ in terms of the auxiliary variables

$$A_1(\Lambda) = (1-\eta_1)(1-\eta_2)\sum_{h=0}^{\Lambda} W(h)\left(1-\left[\frac{p(\tau)}{p(\tau+h)}\right]^Q\right), \quad (19)$$

$$A_2(\Lambda) = -(1-\eta_1)(1-\eta_2)\sum_{h=0}^{\Lambda} W(h)\left(1-\left[\frac{p(\tau)}{p(\tau+h)}\right]^Q\right)p^Q(\tau+h), \quad (20)$$

$$A_3(\Lambda) = \eta_1(1-\eta_2)\sum_{h=0}^{\Lambda} W(h)\left(1-\left[\frac{p(\tau)}{p(\tau+h)}\right]^Q\right), \quad (21)$$

$$A_4(\Lambda) = \eta_2\sum_{h=0}^{\Lambda} W(h), \quad (22)$$

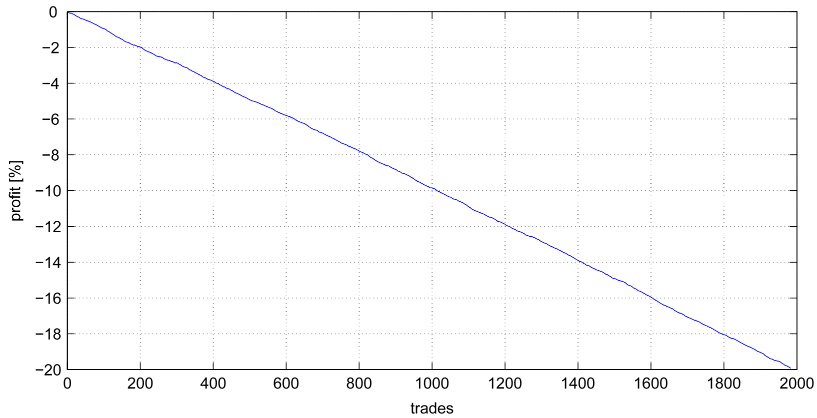

Fig. 1. The profit of the model on the EUR/USD currency rate with transaction costs included dependence on trades for one year period.

$$A_5(\Lambda) = -\eta_2 \sum_{h=0}^{\Lambda} W(h)\, p^Q(\tau + h). \tag{23}$$

Thus the expected prediction form reads

$$\hat{p}(\tau_0 + l_{\rm pr}) = \left[\frac{A_2(\Lambda) + A_5(\Lambda)}{C(\tau_0 - l_s, \Lambda) - A_1(\Lambda) - A_3(\Lambda) - A_4(\Lambda)}\right]^{1/Q}, \tag{24}$$

where we use the notation $\tau = \tau_0 + l_{\rm pr} - l_s$. The derivation is based on the invariance

$$C(\tau, l_s - l_{\rm pr}) = C(\tau - l_{\rm pr}, l_s - l_{\rm pr}), \quad \Lambda = l_s - l_{\rm pr}, \tag{25}$$

where $l_{\rm pr}$ denotes the prediction scale.

The model was tested for various sets of parameters l_s, $l_{\rm pr}$, η_1, η_2, Q and the new parameter ϵ which is defined as

$$\epsilon = \left|C(\tau, l_s - l_{\rm pr}) - C(\tau - l_{\rm pr}, l_s - l_{\rm pr})\right| \tag{26}$$

and describes the level of invariance in real data. The best prediction (the best means that the model has the best ability to estimate the right price) is obtained by using the following values of parameters

$$\begin{aligned} l_s &= 900, \\ l_{\rm pr} &= 1, \\ \eta_1 &= 0, \\ \eta_2 &= 0, \\ Q &= 6, \\ \epsilon &= 10^{-10}. \end{aligned} \tag{27}$$

The graphical descriptions of prediction behavior of the model with and without transaction costs on the EUR/USD currency rate of the forex market are described in Figs. 1–4. During a one year period the model lost around 20% of the initial money. It executed 1983 trades (Fig. 1) where only 10 were suggested by the model (and earned money) and the rest of them were random (which can be clearly seen in Figs. 3, and 4). The problem of this model is its prediction length (the parameter $l_{\rm pr}$), in this case it is one tick ahead. The price was predicted correctly in 48.57% of all cases (16 201 in one year) and from these 48.57% or numerally 7869 cases only 0.13% or numerally 10 were suitable for trading. This small percentage is caused by the fact that the price does not change too often one tick ahead. One could try to raise the prediction length to find more suitable cases for trading. This is only partly successful because the rising parameter $l_{\rm pr}$ induces a loss of the prediction strength of the model. For example when $l_{\rm pr} = 2$ (two ticks ahead) the prediction strength decreases from around 50% to 15%.

The problem is that the invariant equation (10) is fulfilled only on the very short period of the time series due to the very chaotic nature of financial data behavior. Therefore the PMBSI is effective only on the one step prediction where there is very low probability that time series change significantly. The situation, however, is different for more steps prediction where there is, on the contrary, a very high probability of big changes in time series to occur, and the following predictions have rather small efficiency in such cases. The only way how to establish better prediction also for more steps prediction is to choose the right weights equation (9). The right and optimized weights should considerably extend the interval where Eq. (10) is fulfilled. Therefore it is also our task in future work.

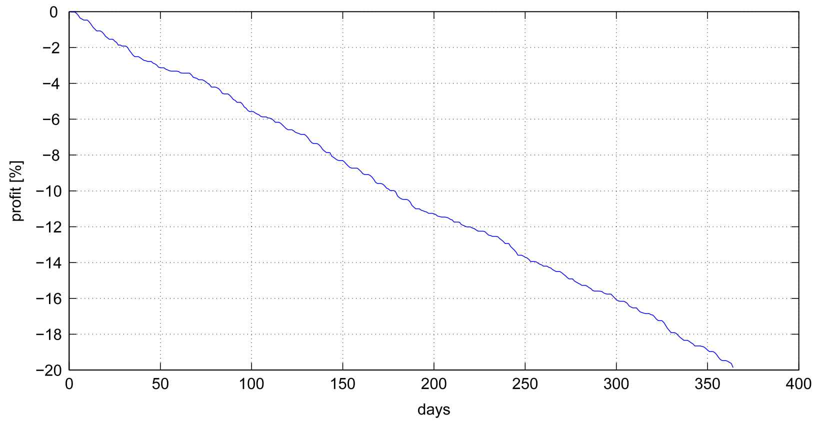

Fig. 2. The profit of the model on the EUR/USD currency rate with transaction costs included dependence on days for one year period.

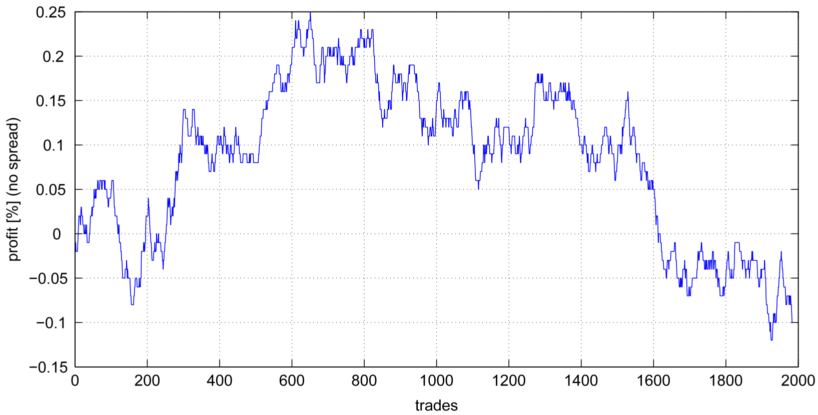

Fig. 3. The profit of the model on the EUR/USD currency rate without transaction costs included dependence on trades for one year period.

3.2. Experimental setup

The experiments were performed on two time series. The first series represented artificial data namely a single period of a sinusoid sampled by 51 regularly spaced samples. The second time series represented proprietary financial data sampled daily over a period of 1295 days. The performance of PMBSI was compared to SVM and to naive forecast. There were two error measures used, mean absolute error (MAE) and symmetric mean absolute percentage error (SMAPE) defined as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n} |A_t - F_t|, \tag{28}$$

$$\mathrm{SMAPE} = \frac{100}{n}\sum_{t=1}^{n} \frac{|A_t - F_t|}{0.5(|A_t| + |F_t|)}, \tag{29}$$

where n is the number of samples, A_t is the actual value and F_t is the forecast value. Each time series was divided into three subsets: training, evaluation and validation data. The time ordering of the data was maintained; the least recent data were used for training, while the more recent data were used to evaluate the performance of the particular model with the given parameters' setting. The best performing model on the evaluation set (in terms of MAE) was chosen and made to forecast for the validation data (the most recent) that were never used in the model optimization process. Experimental results on the evaluation and validation data are presented below. The parameters of the models were optimized by trying all combinations of parameters sampled from given ranges with a sufficient sampling rate. Naturally, this process is slow but it enabled us to get an image of the shape of the error surface corresponding to the given settings of parameters and ensured that local minima are explored. The above approach was used for both PMBSI and SVM. The SVM models were constructed so that the present value and a certain number of the consecutive past values comprised the input to the model. The input vector corresponds to what will be referred to here as the *time window* with the length l_{tw} (representing the equivalent of the length of the string map l_s by PMBSI).

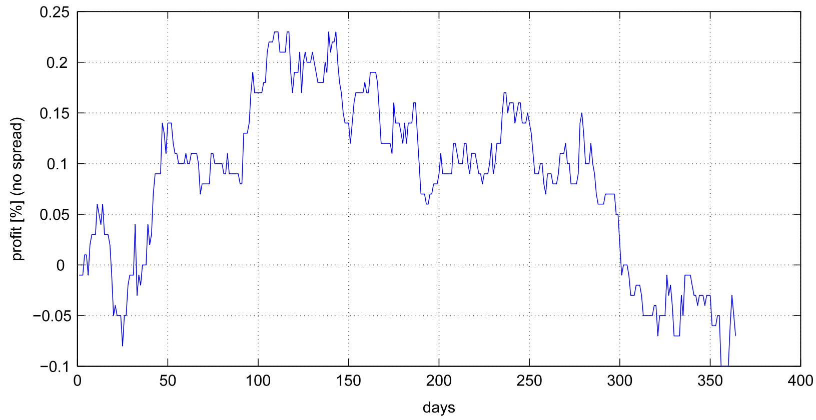

Fig. 4. The profit of the model on the EUR/USD currency rate without transaction costs included dependence on days for one year period.

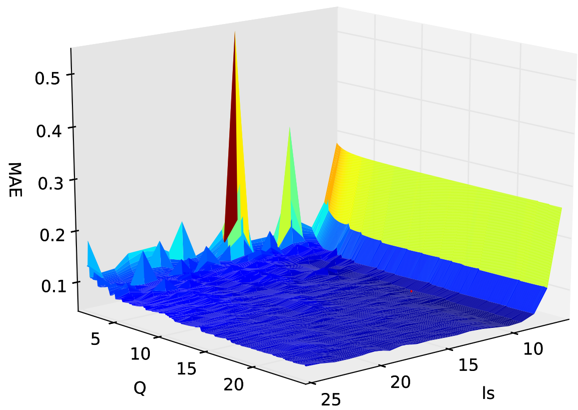

Fig. 5. MAE corresponding to various settings of *ls* and *Q* on the financial data. The red dot is the global minimum of MAE. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Comparison

There was a preliminary experimental analysis performed of the PMBSI method. The goal was to evaluate the prediction accuracy, generalization performance, convenience of the method in terms of the operator' effort needed to prepare a working model, computational time and other aspects of the PMBSI method that may have become obvious during the practical deployment. SVM was chosen as a benchmark. The experimental data comprised two sets: artificial data (a single period of a sinusoid) and real world data (financial, price development). We will provide a brief conclusion of the analysis here. Each time series was divided into three subsets for training, testing and validation. The results were calculated on the validation sets that have been entirely absent in the process of optimization of parameters.

The PMBSI predictor does not undergo a training process that is typical for ANN and SVM where a number of free parameters must be set (synaptic weights by ANN, α coefficients by SVM). PMBSI features a similar set of weights (W) but often very small and calculated analytically. The parameters to be optimized are only four: ls, Q, η_1, η_2. This, clearly, is an advantage. On the other hand the optimal setting of the parameters is not easy to find as there are many local minima on the error surface. In this analysis the optimal setting was found by testing all combinations of parameters from given ranges. Fig. 5 shows the Mean Absolute Error (MAE) of the 5-steps ahead forecast of the financial time series corresponding to various settings of ls and Q (η_1, $\eta_2 = 0$). But the figure makes it also obvious that PMBSI's performance is approximately the same for a wide range of settings on this data.

For PMBSI to work the elements of time series must be non-zero, otherwise the method will return *not a number* forecasts only. The input time series must then be modified by adding a constant and the forecast by subtracting the same constant. Even so the algorithm returned a *not a number* forecast in approx. 20% of the cases on the financial data. In such cases the last valid forecast was used. Due to reasons that are presently being evaluated the accuracy of PMBSI is matching and even outperforming SVM for single step predictions but rapidly deteriorates for predictions of more steps ahead. Iterated prediction of several steps ahead using the single step PMBSI predictor improves the accuracy significantly. The sinusoid used for experiments was sampled by 51 points, the positive part of the wave was used for optimization of the parameters

Table 1
Experimental results on artificial time series.

Method	l_{pr}	MAE eval	MAE valid	SMAPE valid
PMBSI	1	0.000973	**0.002968**	**8.838798**
	2	0.006947	0.034032	14.745538
	3	0.015995	0.161837	54.303315
Iterated PMBSI	1	–	–	–
	2	0.003436	0.011583	10.879313
	3	0.008015	0.028096	14.047025
SVM	1	0.011831	0.007723	10.060302
	2	0.012350	**0.007703**	**10.711573**
	3	0.012412	**0.007322**	**11.551324**
Naive forecast	1	–	0.077947	25.345352
	2	–	0.147725	34.918149
	3	–	0.207250	41.972591

Table 2
Optimal PMBSI parameters.

l_{pr}	l_s	Q	η_1	η_2
1	2	0.30	0.80	−0.20
2	5	0.10	0.80	−0.60
3	8	0.10	0.80	−0.60

and the rest for validation (approx. 50–50 division). Fig. 6 shows the comparison of iterated versus the direct prediction using PMBSI. Table 1 shows the experimental results. The results of the best performing models are highlighted.

The optimal l_{tw} for SVM was 3 for all predictions. Table 2 shows the optimal settings found for PMBSI. For $l_{pr} = 1$ when PMBSI outperformed linear SVM the optimal length of the string map was shorter than the optimal time window for SVM; in the remaining cases it was significantly longer.

5. Prediction model based on the deviations from the closed string/pattern form (PMBCS)

For the next trading strategy we want to define some real values of the string sequences. Therefore we define the momentum which acquired values from the interval (0, 1). The momentum M is not strictly invariant as in the previous model of the time series in its basic definition. It is a trading strategy to find such a place in the forex time series market where M is exactly invariant or almost invariant and we can predict increasing or decreasing of prices with higher efficiency. For example our predictor somewhere in the time series has 55% efficiency to predict the movement of price but in the invariant place of our trading strategy where Eqs. (26), and (30) are almost invariant the efficiency of our predictor increased to 80%. Therefore the idea to find the invariant in time series plays a crucial role in our trading strategy but one still needs to find an appropriate expression for such a prediction.

To study the deviations from the benchmark string sequence we define momentum as

$$M_{(l_s,m;Q,\varphi)} = \left(\frac{1}{l_s+1} \sum_{h=0}^{l_s} \left| \frac{p(\tau+h) - p_{\min}(\tau)}{p_{\max}(\tau) - p_{\min}(\tau)} - \frac{1}{2}\left(1 + \cos\left[\frac{2\pi m h}{l_s+1} + \varphi\right]\right) \right|^Q \right)^{1/Q} \tag{30}$$

where

$$p_{\text{stand}}(\tau; h; l_s) = \frac{p(\tau+h) - p_{\min}(\tau; l_s)}{p_{\max}(\tau; l_s) - p_{\min}(\tau; l_s)}, \quad p_{\text{stand}} \in (0, 1),$$

and

$$p_{\max}(\tau; h; l_s) = \max_{h \in \{0,1,2,\ldots,l_s\}} p(\tau+h), \qquad p_{\min}(\tau; h; l_s) = \min_{h \in \{0,1,2,\ldots,l_s\}} p(\tau+h),$$

and φ is a phase of periodic function. The momentum defined above takes the values from the interval $M_{(l_s,m;Q,\varphi)} \in (0, 1)$. The periodic function $\cos(\tilde{\varphi})$ in the definition of Eq. (30) could be substituted by other types of mathematical functions. The results with different kinds of functions could be different.

5.1. Elementary trading strategy based on the probability density function of M

The purpose is to take advantage of it whenever the market conditions are favorable. As in the previous model we are detrending forex data into the one dimensional topological object "strings" with different parameters. The trading strategy is based on the description of rate curve intervals by one value called the moment of the string. These moments are statistically

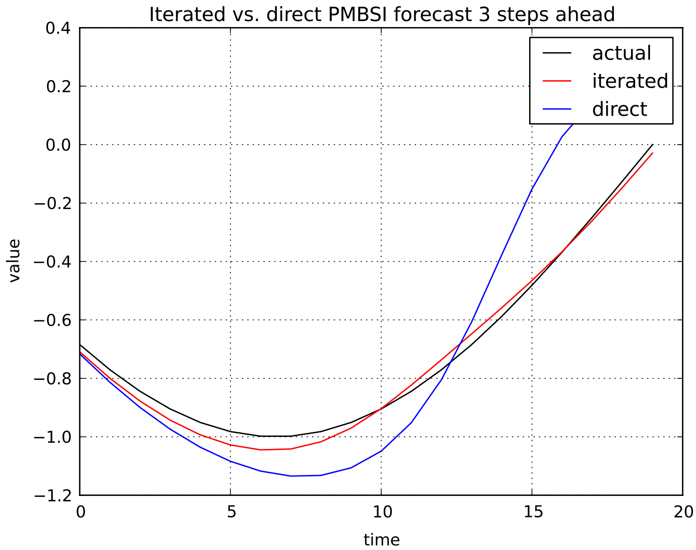

Fig. 6. Iterated and direct prediction using PMBSI on artificial data.

processed and some interesting values of moments are found. The values directly affect the opening and closing of trade positions. The algorithm works in two complementary phases. The first phase consists of looking for "good" values of moments followed by the second phase which uses results from the first phase and opening/closing of trade positions occur. Simultaneously the first phase is looking for new "good" values of moments.

Risk is moderated by a number of allowed trades that the algorithm can open during a certain period. Also it is moderated by two parameters which affect the selection of suitable moments for trading. The maximum number of trades is 10/h. The algorithm is tested on various periods of historical data. The number and period of simultaneously opened trades are monitored all the time.

The first set of parameters describes the moment (simple scalar function of several variables from the interval ⟨0, 1⟩). The first set consists of these parameters: length of moment string (number of ticks or time period), quotient or exponent of moment, frequency of moment function, and phase shift of moment function. The second set of parameters controls trading strategy and consists of these variables: maximum number of simultaneously opened trades, skewness of moments distribution and Sharpe ratio of closed trades. As soon as the algorithm calculates the value of the moment and finds out that the value is "good", then it immediately carries out an appropriate command.

The risk of the algorithm is governed by the second set of parameters and can vary from zero (low risk but also low or zero number of trades) to the boundary values controlled by the model parameters. These boundary values are unlimited but could be easily affected by the skewness and Sharpe ratio. These parameters can limit loss to a certain value with accuracy ±2% but also limit overall profit significantly if low risk is desired.

An arbitrage opportunity is taking advantage of the occurrence of a difference in distribution. Opportunity is measured by *Kullback–Leibler* divergence

$$D_{\mathrm{KL}} = \sum_{j(bins)} \mathrm{pdf}(M^+(j)) \log\left(\frac{\mathrm{pdf}(M^+(j))}{\mathrm{pdf}(M^-(j))}\right) \tag{31}$$

where larger D_{KL} means better opportunities ($D_{\mathrm{KL}} > D_{\mathrm{threshold}}$) e.g. when $D_{KL} > D_{threshold}$ it means the buying of Euro against USD could be more profitable. Statistical significance means the smaller the statistics accumulated into bins $\mathrm{pdf}(M^+(j))$, $\mathrm{pdf}(M^-(j))$, the higher is the risk (M from the selected range should be widespread). The meaning of pdf in the definition of equation above is the probability density function.

More generally we can construct the series of $(l_s + 1)$ price ticks $[p(\tau), p(\tau + 1), \dots, p(\tau + l_s)]$ which are transformed into a single representative real value $M(\tau + l_s)$. Nearly stationary series of $M(\tau + l_s)$ yield statistics which can be split into: branch where M is linked with future uptrend/ downtrend and branch where M is linked with future profit/ loss taking into account transaction costs. Accumulation of $\mathrm{pdf}(M^{+-}_{\mathrm{long}})$ means (profit+/ loss−) or $\mathrm{pdf}(M^{+-}_{\mathrm{short}})$ (profit+/ loss−). M^+ in Eq. (31) describes when Eq. (30) brings profit and M^- loss.

As in the previous section the model was again tested for various sets of free parameters l_s, h, Q, φ. This model can make "more-tick" predictions (in tests it varies from 100 to 5000 ticks). Therefore it is much more successful than the previous model. It is able to make a final profit of around 160% but this huge profit precedes a fall of 140% of the initial state. It is important to emphasize that all profits mentioned here and below are achieved by using leverage (borrowing money) from 1 to 10. The reason for leverage is the fact that the model could simultaneously open up to 10 positions (one position means one trade i.e. one pair of buy–sell transactions). If one decides not to use any leverage the final profit decreases 10 times.

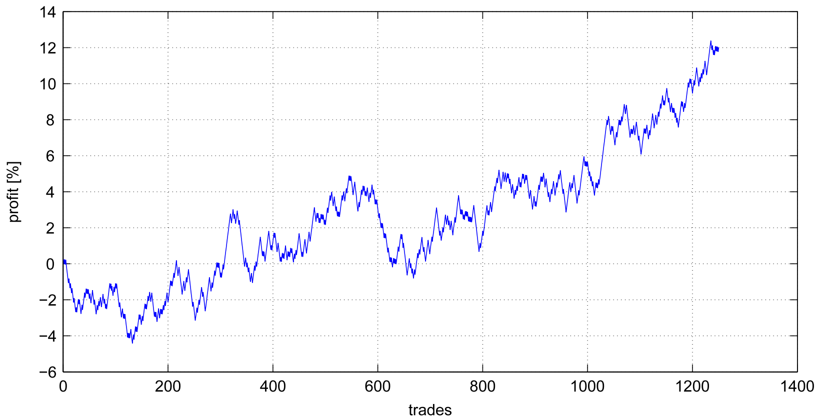

Fig. 7. The profit of the model on the EUR/USD currency rate with transaction costs included dependence on trades for one year period.

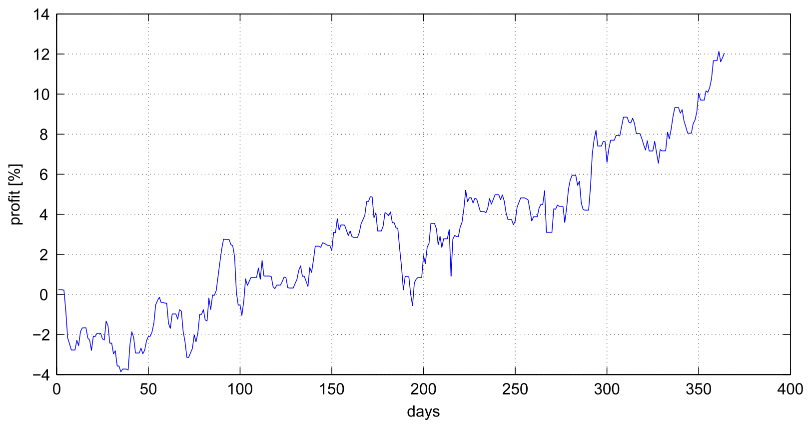

Fig. 8. The profit of the model on the EUR/USD currency rate with transaction costs included dependence on days for one year period.

On the other hand, with using the leverage 1 to 20 the final profit doubles itself. Of course, the use of higher leverages is riskier as dropdowns are also higher. There is, for example, in Fig. 7 a dropdown approx. 6% around 600 trades. With the use of leverage 1 to 20 this dropdown rises to 12%.

128 000 combinations of the model's parameters have been calculated. Figs. 7–10 describe some interesting cases of the prediction behavior of the model with the transaction cost included on the EUR/USD currency rate of the forex market. Figs. 7 and 8 describe the model (one set of parameters) under conditions that the fall down must not be higher than 5%. The best profit achieved in this case is 12%.

In order to sort out the best combinations of parameters it is helpful to use the statistical quantity called the Sharpe ratio. The Sharpe ratio is a measure of the excess return per unit of risk in a trading strategy and is defined as

$$S = \frac{E(R - R_f)}{\sigma}, \qquad (32)$$

where R is the asset return, R_f is the return on a benchmark asset (risk free), $E(R - R_f)$ is the mean value of the excess of the asset return over the benchmark return, and σ is the standard deviation of the excess of the asset return. You mention the Sharpe ratio. The values of the Sharpe ratio for the best fit are e.g. for Fig. 10 it is the value 1.896 and for Fig. 11 it is the value 1.953, where as a reference profit we choose a bank with 5% profit.

Fig. 9 shows the case where the Sharpe ratio has the highest value from all sets of the calculated parameters. One year profit is around 26% and the maximum loss is slightly over 5%. Fig. 10 describes the case requiring a high value of Sharpe ratio and with the aim to gain profit of over 50%.

There exist sufficiently enough cases with high Sharpe ratio which leads to enhancement of the model to create a self-education model. This enhancement takes some ticks of data, finds out the best case of parameters (high Sharpe ratio and also high profit) and starts trading with these parameters for some period. Meanwhile, the trading with previously found parameters model is looking for a new best combination of parameters. Fig. 11 describes this self-education model where parameters are not chosen and the model itself finds the best one from the financial data and is subsequently looking for the best values for the next trading strategy.

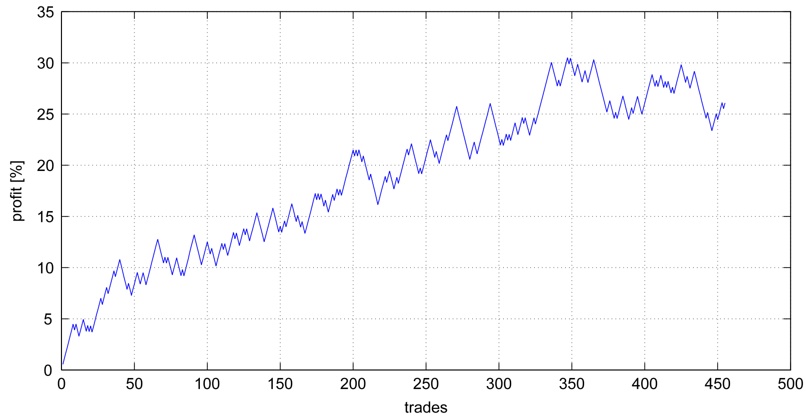

Fig. 9. The profit of the model on the EUR/USD currency rate with transaction costs included dependence on trades for one year period.

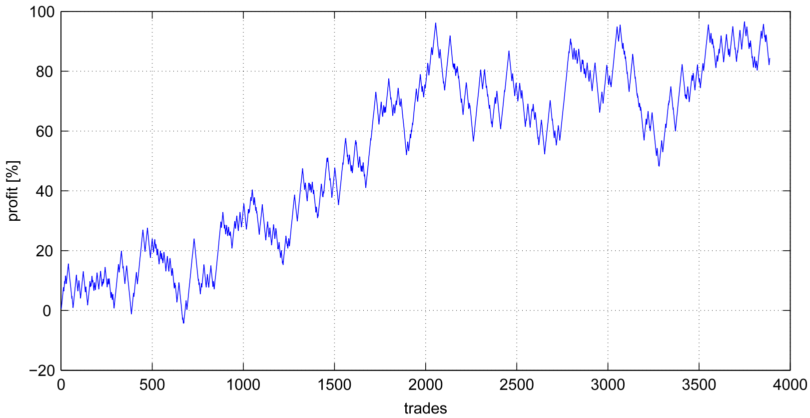

Fig. 10. The profit of the model on the EUR/USD currency rate with transaction costs included dependence on trades for one year period.

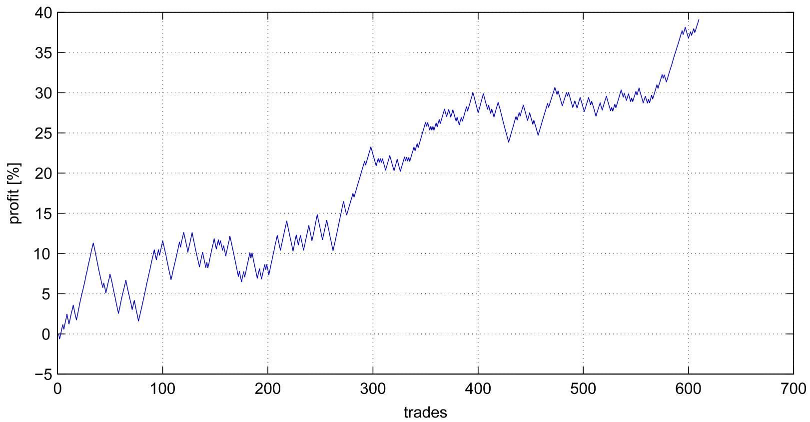

Fig. 11. The profit of the self education model on the EUR/USD currency rate with transaction costs included dependence on trades for one year period.

6. Conclusions

The model of strings allows one to manipulate with the information stored along several extra dimensions. We started from the theory of the 1-end-point and 2-end-point open string and continued with partially compactified strings that satisfy the Dirichlet and Neumann boundary conditions. We have 5 free parameters in our model. We have also tried out-of-sample tests, however, only using small data samples. We have not encountered "overfitting" due to the fact that parameters are stable enough within our string theory approach to produce profit even if we slightly change them. For all computations in the second model we are taking bid–offer spreads into account. We are calculating with real values of bid–offer spreads from historical data and it is dependent on where we are simulating on Oanda or Icap etc. A number of trades per day varies from 2 to 15 depending on fit strategy.

We have shown that string theory may motivate the adoption of the nonlinear techniques of data analysis with a minimum impact of justification parameters. The numerical study revealed interesting fundamental statistical properties of the maps from the data onto string-like objects. The main point here is that the string map gives a geometric interpretation of the information value of the data. The results led us to believe that our ideas and methodology can contribute to the solution of the problem of robust portfolio selection. The financial market invariants could be some other form of definition of scaling laws found in Ref. [21].

We established two different string prediction models to predict the behavior of the forex financial market. The first model PMBSI is based on the correlation function as an invariant and the second one PMBCS is an application based on the deviations from the closed string/pattern form. We found the difference between these two approaches. The first model cannot predict the behavior of the forex market with good efficiency in comparison with the second one which, moreover, is able to make relevant profit per year. From the results described we can conclude that the invariant model as one step price prediction is not sufficient for big dynamic changes of the current prices on the finance market. As can be seen in Figs. 3 and 4 when the transaction costs are switched off the model has some tendency to make a profit or at least preserve a fortune. It means that it could also be useful for other kinds of data, where the dynamics of changes are slower, e.g. for energetic [22] or seismographic data [23] with longer periods of changes. Finally the PBMSI in the form presented in this paper should be applicable with good efficiency only to other kinds of data with smaller chaotic behavior in comparison with financial data.

Moreover PMBSI is a method under development. Unlike SVM or ANN, at this stage PMBSI does not require a training process optimizing a large number of parameters. The experimental results indicate that PMBSI can match or outperform SVM in one step ahead forecasts. Also, it has been shown that finding optimal settings for PMBSI may be difficult but the method's performance does not vary much for a wide range of different settings. Besides the further testing of PMBSI we consider that fast methods for optimization of parameters must be developed. Because of the character of the error surface we have chosen to use evolutionary optimization as the method of choice. After a fast and successful parameters' optimization method is developed, optimization of the weighting parameters (Eqs. (9), and (14)) will be included into the evolutionary process.

The profit per year from the second prediction model was obtained from approximately 15% and more depending on the parameter set from the data we have chosen. This model is established efficiently on the finance market and could be useful to predict future prices for the trading strategy.

To summarize the second prediction algorithm we can conclude that we detrended forex data into the one dimensional topological object "strings" with some special parameters. Trading strategy is based on a description of rate curve intervals by one value called the moment of the string. These moments are statistically processed and some interesting values of moments are found. These values directly affect the opening and closing of trade positions. Risk is moderated by the number of allowed trades that the algorithm can open during some period.

Of course the model still needs to be tested further. With the flow of new financial data the model can be more optimized and also it could become resistant to a crisis. In the future research we would like to use evolutionary algorithms for the better optimization of the model's parameters. The presented models are universal and could also be used for predictions of other kinds of stochastic data. The self-educated models presented in Fig. 11 are very useful because they are able to find on their own the best parameter set from data in addition to data found in Eqs. (27), learn about the prices and utilize these pieces of information for the next trading strategy. These models could also be very helpful for portfolio optimization and financial risk management in the banking sector. Finally we very much hope that the presented approach will be very interesting and useful for a broad spectrum of people working on the financial market.

Acknowledgment

The work was supported by the Slovak Academy of Sciences in the framework VEGA Grant No. 2/0037/13.

References

[1] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31 (1986) 307–327.
[2] R. Engle, Autoregressive conditional heteroskedasticity with estimates of United Kingdom inflation, Econometrica 50 (1982) 987–1008.
[3] M. Deistler, The structure of ARMA systems in relation to estimation, in: P.E. Caines, R. Hermann (Eds.), Geometry and Identification, Proceedings of APSM Workshop on System Geometry, System Identification, and Parameter Estimation, Systems Information and Control, vol. 1, Math Sci Press, Brookline, MS, 1983, pp. 49–61.
[4] G.E.P. Box, G.M. Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco, 1970.
[5] F.E.H. Tay, L. Cao, Application of support vector machines in financial time-series forecasting, Omega 29 (2001) 309–317.
[6] J. Kamruzzaman, R. Sarker, Forecasting of currency exchange rates using ANN: a case study, in: Proc. IEEE Intl. Conf. on Neur. Net. & Sign. Process., ICNNSP03, China, 2003.
[7] W-H. Chen, J-Y. Shih, S. Wu, Comparison of support-vector machines and back propagation neural networks in forecasting the six major Asian stock markets, International Journal of Electronic Finance 1 (1) (2006) 49–67.
[8] Vladimir N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag New York, Inc., ISBN: 0-387-94559-8, 1995.
[9] M. Bundzel, P. Sincak, Combining gradient and evolutionary approaches to the artificial neural networks training according to principles of support vector machines, in: International Joint Conference on Neural Networks, 2006, IJCNN '06, 2006, pp. 2068–2074. http://dx.doi.org/10.1109/IJCNN.2006.246976.

[10] Marek Bundzel, Tomáš Kasanický, Baltazár Frankovič, Building support vector machine alternative using algorithms of computational geometry, in: Proceedings of the Third International Conference on Advances in Neural Networks—Volume Part I, ISNN'06, Springer-Verlag, 2006, pp. 955–961. ISBN: 3-540-34439-X, 978-3-540-34439-1.
[11] D. Horváth, R. Pincak, Physica A 391 (2012) 5172.
[12] D. McMahon, String Theory Demystified, The McGraw-Hill Companies, Inc., 2009.
[13] B. Zwiebach, A First Course in String Theory, Cambridge University Press, 2009.
[14] Joseph Polchinski, String Theory, Cambridge University Press, 1998.
[15] P. Grassberger, I. Procaccia, Physica D 9 (1983) 189.
[16] Y. Ding, X. Yang, A.J. Kavs, J. Li, International Journal of Trade, Economics and Finance 1 (2010) 320.
[17] M.C. Mnnix, T. Shimada, R. Schfer, F.L Thomas, H. Seligman, T. Guhr, H.E. Stanley, Scientific Reports 2 (2012). http://dx.doi.org/10.1038/srep00644. Article number: 644.
[18] Jed D. Christiansen, The Journal of Prediction Market 1 (2007) 17.
[19] Ch.Ch. Chang, Ch.Y. Hsieh, Y.Ch. Lin, Applied Economics Letters iFirst 1 (2011).
[20] J. Wolfers, E. Zitzewitz, Journal of Economic Perspectives 18 (2004) 107.
[21] J.B. Glattfelder, A. Dupuis, R.B. Olsen, Quantitative Finance 11 (4) (2011) 599.
[22] A. Lazzarettoa, A. Toffolo, Applied Thermal Engineering 28 (2008) 2405.
[23] R. Stefansson, M. Bonafede, G.B. Gudmundsson, Bulletin of the Seismological Society of America 101 (4) (2011) 1590.