

## STUDY OF PREDICTION MODELS FOR TIME SERIES

Erik Bartoš<sup>1,a</sup>, Richard Pinčák<sup>2,b,c</sup><sup>a</sup>*Institute of Physics, Slovak Academy of Sciences  
Dúbravská cesta 9, SK-845 11 Bratislava, Slovakia*<sup>b</sup>*Institute of Experimental Physics, Slovak Academy of Sciences  
Watsonova 47, SK-043 53 Košice, Slovakia*<sup>c</sup>*Bogoliubov Laboratory of Theoretical Physics, Joint Institute for Nuclear Research  
141980 Dubna, Moscow Region, Russia*

Received 4 August 2017, accepted 9 October 2017

We present the concept which approaches the string theory to the field of time series forecast and data analysis through a transformation of currency rate data to the topology of physical strings and branes. We introduce new type of prediction models for financial time series based on string invariants. The performance of the first versions of prediction models is compared to support vector machines and artificial neural networks on an artificial and financial time series. We propose a string angular momentum as an another tool to analyze the stability of currency rates except the historical volatility. Next we investigate the fundamental properties of the space of time series data. We provide the proof that the space of time series data is a Kolmogorov space with  $T_0$ -separation axiom using the loop space of time series data.

PACS: 11.25.-w, 11.25.Wx, 05.45.Tp, 89.65.Gh

KEYWORDS: String Theory, Time-Series Analysis, Econophysics, Financial Market

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>String maps of time series</b>	<b>6</b>
2.1	One dimensional maps . . . . .	6
2.2	Two dimensional maps . . . . .	7
2.3	Symmetry with respect to direct and indirect quotes . . . . .	9
2.4	D2-branes map . . . . .	11
2.5	Partial compactification . . . . .	11
2.6	Intra-string statistical picture . . . . .	12
2.7	String polarized by the external field . . . . .	14
2.8	Inter-currency study: map onto rotating strings . . . . .	15
2.9	Differentials of string map . . . . .	16

<sup>1</sup>E-mail address: erik.bartos@savba.sk<sup>2</sup>E-mail address: pincak@saske.sk

<b>3</b>	<b>String prediction models and their utilization</b>	<b>18</b>
3.1	Prediction model based on string invariants . . . . .	18
3.2	Experimental analysis of PMBSI . . . . .	19
3.3	Prediction model based on the deviations from the closed string/pattern form . . .	24
3.4	Experimental analysis of PMBCS . . . . .	31
3.5	Regge slope parameter . . . . .	34
<b>4</b>	<b>Kolmogorov space and the concept of algebraic topology and their relation to time series data</b>	<b>39</b>
4.1	Loop space of time series . . . . .	42
4.2	Proof of the main theorem . . . . .	53
<b>5</b>	<b>Conclusions</b>	<b>56</b>
	<b>Acknowledgments</b>	<b>56</b>
	<b>Appendices</b>	<b>56</b>
<b>A</b>	<b>PMBSI model construction</b>	<b>57</b>
<b>B</b>	<b>The realisation of the PMBCS model</b>	<b>60</b>
<b>C</b>	<b>Sharpe ratio and volatility</b>	<b>68</b>
<b>D</b>	<b>Empirical data analysis of time series data by decomposition methods</b>	<b>70</b>
	<b>References</b>	<b>80</b>

## 1 Introduction

The idea of exploring the relationship between more intuitive geometric methods and financial data is not new. The discipline called the geometric data analysis [1] includes many diverse examples of the conceptual schemes and theories grounded on the geometric representation and properties of data. In this work we propose the concept that is based on projection data into higher dimensional vectors in the sense of the string theory, the theory which achieved a high degree of popularity and respect among the physicists [2, 3]. The reason lies in its inherent ability to unify theories that come from diverse physical spheres. The prime instrument of the unification represents the concept of extra dimension. The side-product of theoretical efforts can be seen in the elimination of the ultraviolet divergences of Feynman diagrams. However, despite the considerable achievements, there is a lack of the experimental verification of the original string theory. In contrast, in the present work we exploit financial time series which can build the family of the string motivated models of boundary respecting maps.

Classifying future direction of nonlinear [4] and nonstationary [5, 6] time series data, especially financial time series data, has become one of the intriguing topic and is extensively studied by researchers from different fields due to its commercial applications and attractive benefits that it has to offer. Algorithmic trading with large amount of processed data, the ticks in millisecond scale, requires new physical methods to describe the statistics of the return intervals on the short and large scales [7] and new geometric representation of data, e. g., new view on data statistics in higher dimensions [8]. Moreover, the global markets consist of a large number of interacting units and their time-averaged dynamics resemble the systems with many-body effects. The classical statistical instruments which treats the market as a whole, like the returns and volatility distributions [9, 10] and traditional autoregressive moving average models, must be enhanced by new phenomena from informational and social sciences. Theoretical interest is also oriented to the distribution of the occurrence of rare extreme events in historical time series data [11, 12]. Their clustering in data records indicates the existence of a long-term memory dependencies in financial time series, which is intensively studied [13], i. e., by multiplicative random cascade models. The future stock prices movements are modeling by multiresolution analysis techniques including wavelet analysis [14], empirical and variational mode decomposition [15], adaptive methods for regression [16], frequently in the context of training and predicting on artificial neural networks [17, 18]. New algorithms [19] and new proposed approaches covering the findings of the long memory effects of forex data [20] and its stochastic features in the presence of nonstationarity [21], the renormalization group approach [22], exploitation of genetic algorithms [23], open novel perspectives.

We want to introduce the concept which approaches the string theory to the field of time series forecast and financial data analysis through a transformation of currency rate data to the topology of physical strings and D2-branes [24, 25, 26]. The ideas have been practically demonstrated by a novel prediction method based on string invariants [27, 28] with genetic algorithm for optimization of method's parameters. The method has been tested on competition and real world data, its performance compared to artificial neural networks and support vector machines algorithms. Another interesting application has been the construction of trading algorithm based on 1-endpoint strings and the demonstration of model properties on real online trade system [29]. Stability of the algorithm on transaction costs for long trade periods has been confirmed and compared to benchmark prediction models and trading strategies.

The prediction models in trading on financial markets can be enhanced in the framework of a string theory. We propose to proceed from simple 1-endpoint and 2-endpoints strings to more complex objects, D2-branes. The D2-branes have the ability to smooth the movement of prices on the market and to process the preserved market memory with better efficiency than in the case of the strings, the study of a statistics of momenta of string objects reveal the perspectives of D2-branes. However, the simulations with prediction models based on string approach show that one can profit only on the regions with high stability. In real data, the fluctuations of forex market prices brake the statistics of the predictions [30] and one must build into the models various trading brakes, to deal with the rapid changes. The evaluation of a volatility [31, 32] serves as one of the sources of analyzing tools in pricing strategies. We introduce new methodics based on the analogy with the angular momentum in the string theory [33]. Its application into trading models can serve as complementary financial instrument in addition to a volatility. The changes of Regge slope parameter or the string tension can identify trends on the market, their understanding allows us to dynamically change an intra-string characterization (reduction a string length for a short period) and better predict the movement of prices. Especially in large market fluctuations, their exploitation needs further experimental verification.

Many theorists are trying to understand a dynamics of financial systems with modeling of stock market using empirical analysis of stock price [34]. The financial market can be realized as a topological space of underlying financial time series data in which typical econometric tool of linear regression models such as ARIMA [35], GARCH [36] and state space model cannot visualize all multiple processes of complex system such as the financial market [37]. On the other side, the scientists borrowed from the signal processing the data-mining tool such as neural network combined with wavelet transformation or they supported vector machines (SVM) with some extra datamining tools to predict financial time series with the overfitting and prior problems [38]. They believe that they can overcome prediction problem by finding a good risk factors to let a Bayesian system to learn [39] or by regression of those risk factors, but they do not realize the main problem connected with the defect of algebraic topological construction of the data-mining tools. The main defect of data-mining tools is based on a fitting problem with one parameter of learning from single stochastic process instead of infinite factors in which the influence of infinite stochastic processes governs on future expectation price. In other words one financial time series are composed of infinitely many random variables in which the average of all random variables not always converge to single Kolmogorov space over Euclidean space of time series data. When we add one point of future price and make a fitting curve using the data-mining tool for regression and state space model of Markov switching regime the coefficient of equation which we used for description of the historical data will update and change the historical path, so it leads to a nonrealistic situation.

All those problems of the prior effect and endeffect of time series data have intrinsic behavior of algebraic defect of topological space underlying time series data in separable  $T_0$ -axiom of Kolmogorov space between price and time. The problem of forecasting arose from a defect in algebraic and geometric construction of space of time series and it has a deep relationship to an empirical analysis problem of nonstationarity of time series data and volatility clustering phenomena in financial time series data so called stylized fact [40] and separation of hidden Markov transition probability state in quantum entanglement state with Hopf fibration. In nature of macroeconomic time series model, we assume equilibrium properties of dynamic stochastic model over stochastic process of deterministic dynamical system with many assumptions. The

precise definition of Kolmogorov topological space underlying the financial time series model in spinor field can hopefully introduce the better understanding of the macroeconomic models. The suitable algebraic reconstruction of space of time series, possible under a Kolmogorov space concept with consistent separation axiom, could help with analysis of the prior effect and endeffect of time series models. There are some indications that such a concept could be realized as a quotient topological space with a few hidden states in extradimensions of loop space of time series data. It is possible to connect hidden eight states in a Kolmogorov space with the empirically observed characteristic correlation structure patterns [41].

In the future we want to concentrate on further expansion of proposed ideas. We plan to build algorithms based on support vector machines and support spinor machine (SSM) [42] and to implement them not only for the analysis of time series data but also to find their application in interdisciplinary areas as machine learning, econophysics or biological sciences.

The paper is organized as follows. In Section 2 we study the projections of the real exchange rate dynamics onto the string-like topology. We introduce the general models of multi dimensional string models, their properties and comparison with previous models are discussed. Section 3 presents the string prediction models based on string invariants. The models are supplemented with their utilization on artificial and real world data and their performance compared to statistical and artificial intelligence methods. We also introduce the Regge alpha slope parameter for the investigation of the stability of currency rates. In Section 4 we specify the basic definition of Kolmogorov space and how the concept of algebraic topology is related to financial time series. We define a loop space in time series data by using of extradimensions of underlying topological space and prove that there exists a time series data in spinor field with underlying structure of Kolmogorov space in time series data. In the last Section 5 the conclusions are summarized and in Appendices we provide details of our data analyses.

## 2 String maps of time series

A time series is a series of discrete data points indexed by some index set in time order. Very often it is a sequence of spaced points taken in the equal time intervals. An example of a set of time series is

$$X = \{x_1, x_2, x_3, x_4, \dots, x_n\}, \quad i = 1, 2, \dots, n. \quad (2.1)$$

which induces a sequence of measured values

$$x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow \dots \rightarrow x_n \quad (2.2)$$

and a sequence of discrete time intervals between measurements

$$t_1 \rightarrow t_2 \rightarrow t_3 \rightarrow \dots \rightarrow t_{n-1}. \quad (2.3)$$

Real examples of time series are long-time physical events as lunar eclipses, occurrences of sunspots, earthquakes, El Niño events or short-time biological events in living bodies, as blood glucose concentration etc. In econophysics the stock market indices (e. g., NASDAQ, S&P 500), option derivatives or currency exchange rates are monitored. The center of our attention is focused on the exchange rates  $p(\tau)$  of main trading currencies, as EUR/USD, GBP/USD, USD/JPY, where  $\tau$  represents the time index. The mean  $p(\tau) = (p_{\text{ask}} + p_{\text{bid}})/2$  is determined from a bid-ask spread of the ask price  $p_{\text{ask}}$  and the bid price  $p_{\text{bid}}$  of an asset in the market.

By applying standard methodologies of detrending one may suggest to convert original time series of the mean currency exchange rate onto a series of returns defined as

$$\frac{p(\tau + h) - p(\tau)}{p(\tau + h)}, \quad (2.4)$$

where  $h$  denotes a tick lag between exchange rates  $p(\tau)$  and  $p(\tau + h)$ .

At this point we introduce the string maps originated from string theory framework. The concept of string maps is based on the connection of the currency quotes and the string objects. For the defined time series of currency exchange rates for the ask  $p_{\text{ask}}(\tau)$  and bid  $p_{\text{bid}}(\tau)$  values in time  $\tau$  one can construct the string maps with the typical length  $l_s$ . These non-local objects serve as the basic objects for further operations. In contrast to classical time series forecasting methods, e. g., autoregressive and moving average models, which forecast the variable of interest using a linear combination of past values or errors of the variable, the string maps carry the larger price history, thereafter the trends of irregular or untypical price changes can be caught with better accuracy.

### 2.1 One dimensional maps

Using the string theory let us first define the 1-endpoint open string map

$$P^{(1)}(\tau, h) = \frac{p(\tau + h) - p(\tau)}{p(\tau + h)}, \quad h \in \langle 0, l_s \rangle \quad (2.5)$$

where superscript (1) refers to the number of endpoints and  $l_s$  to the length of the string (string size), it is a positive integer. The tick lag  $h$  may be interpreted as a variable which extends along the extra dimension limited by the string size  $l_s$ .

A natural consequence of the transformation Eq. (2.5) is the fulfillment of the boundary condition

$$P^{(1)}(\tau, 0) = 0, \quad (2.6)$$

which holds for any tick coordinate  $\tau$ . To enhance the influence of rare events, a power-law  $q$ -deformed model is introduced

$$P_q^{(1)}(\tau, h) = f_q \left( \frac{p(\tau + h) - p(\tau)}{p(\tau + h)} \right), \quad h \in \langle 0, l_s \rangle \quad (2.7)$$

by means of the signum function

$$f_q(x) = \text{sign}(x) |x|^q, \quad q > 0. \quad (2.8)$$

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

Very interesting issue is the generalization of the 1-endpoint string map to the object which includes the effect of many length scales

$$P_q^{(N_{l_s})}(\tau, h; \{l\}) = \prod_{i=1}^{N_{l_s}} f_q \left( \frac{p(\tau + l_i) - p(\tau + h)}{p(\tau + h)} \right), \quad (2.9)$$

which relies on the sequence  $\{l\} \equiv \{l_i, i = 1, \dots, N_{l_s}\}$ , including the end points

$$\min_{i=1, \dots, N_{l_s}} l_i, \quad \max_{i=1, \dots, N_{l_s}} l_i, \quad (2.10)$$

as well as the  $N_{l_s} - 2$  interior node points that divide the string map into the sequence of unfixed segments of the non-uniform length.

## 2.2 Two dimensional maps

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

$$P_q^{(2)}(\tau, h) = f_q \left( \left( \frac{p(\tau + h) - p(\tau)}{p(\tau + h)} \right) \left( \frac{p(\tau + l_s) - p(\tau + h)}{p(\tau + l_s)} \right) \right), \quad h \in \langle 0, l_s \rangle. \quad (2.11)$$

The map is suggested to include boundary conditions of Dirichlet type

$$P_q^{(2)}(\tau, 0) = P_q(\tau, l_s) = 0, \quad \text{at all ticks } \tau. \quad (2.12)$$

One can see from Eq. (2.11) that the sign of  $P_q^{(2)}(\tau, h)$  composes the information about the behavior differences of  $p(\cdot)$  at three particular times  $(\tau, \tau + h, \tau + l_s)$ . The case  $P_q^{(2)}(\tau, h) < 0$  occurs for trends of the different sign, whereas  $P_q^{(2)}(\tau, h) > 0$  indicates the match of the signs.

In addition to the variable  $P_q^{(2)}(\tau, h)$  mapping the open string with 2-endpoints one can introduce the conjugate variable  $X_q^{(2)}(\tau, h)$  with the help of the recurrent summation

$$X_q^{(2)}(\tau, h + 1) = X_q^{(2)}(\tau, h) + P_q^{(2)}(\tau, h - 1) [t(\tau + h) - t(\tau + h - 1)] \quad (2.13)$$

where  $t(\cdot)$  stands for a time-stamp corresponding to the quotation index  $\tau$  in the argument. The above discrete form is suggested on the basis of the time-continuous Newton second law of motion

$$\dot{X}_q^{(2)}(t, h) = P_q^{(2)}(t, h)$$

(written here for an unit mass). The form is equivalent to the imposing of the quadratic kinetic energy term  $\frac{1}{2}(P_q^{(2)})^2$ . Thus, the Hamiltonian picture [43] can be reconstructed in the following way

$$\mathcal{H} = \frac{1}{2} \sum_{h=0}^{l_s} \left[ (P_q^{(2)}(\tau, h))^2 - [\phi_{\text{ext}}(\tau, h + 1) - \phi_{\text{ext}}(\tau, h)] X_q^{(2)}(\tau, h) \right], \quad (2.14)$$

where  $\phi_{\text{ext}}(\tau, h)$  is the external field term which depends on the transformation of currency rate, see, e. g., Eq. (2.11)]. We pass from the continuum to discrete theory by means of the functional form

$$\dot{P}_q^{(2)} = - \frac{\delta \mathcal{H}}{\delta X_q^{(2)}(h)} = \phi_{\text{ext}}(\tau, h + 1) - \phi_{\text{ext}}(\tau, h) = P_q^{(2)}(\tau, h + 1) - P_q^{(2)}(\tau, h), \quad (2.15)$$

where  $P_q^{(2)}(\tau, h)$  can be calibrated to equal  $\phi_{\text{ext}}(\tau, h)$ . The discrete conjugate variable meets the Neumann type boundary conditions

$$X_q^{(2)}(\tau, 0) = X_q^{(2)}(\tau, 1), \quad X_q^{(2)}(\tau, l_s - 1) = X_q^{(2)}(\tau, l_s), \quad (2.16)$$

which is illustrated in Fig. (2.1).

A more systematic way to obtain the 2-endpoint string map represents the method of undetermined coefficients. The numerator of  $q = 1$  can be chosen in the functional polynomial form of degree 2 with the coefficients  $\beta_0, \dots, \beta_5$  as

$$P_{q=1, \text{Num}}^{(2)}(\tau, h) = \beta_0 p^2(\tau + h) + \beta_1 p^2(\tau) + \beta_2 p^2(\tau + l_s) + \beta_3 p(\tau)p(\tau + h) + \beta_4 p(\tau)p(\tau + l_s) + \beta_5 p(\tau + h)p(\tau + l_s). \quad (2.17)$$

As in the previous case, the Dirichlet conditions  $P_{q=1, \text{Num}}^{(2)}(\tau, 0) = P_{q=1, \text{Num}}^{(2)}(\tau, l_s) = 0$  yield  $P_{q=1, \text{Num}}^{(2)} = \beta_0(p(\tau) - p(\tau + h))(p(\tau + l_s) - p(\tau + h))$  with arbitrary  $\beta_0$ . The overlooked denominator part of fraction  $P_{q=1}^{(2)}$  then serves as a normalization factor.

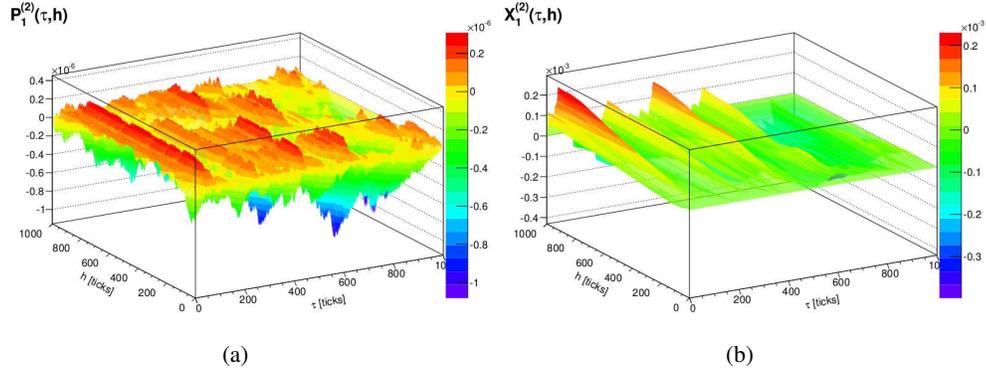


Fig. 2.1: The examples of the currency data map for EUR/USD OANDA data represented by the 2-end-point string map  $P_1^{(2)}(\tau, h)$  (a) and the conjugate variable  $X_1^{(2)}(\tau, h)$  (b). The calculation carried out for  $l_s = 1000$ ,  $q = 1$  at some time instant.

### 2.3 Symmetry with respect to direct and indirect quotes

The currency pairs can be separated into direct and indirect types. In a direct quote the domestic currency is the base currency, while the foreign currency is the quote currency. An indirect quote is just the opposite case. Therefore, it would be interesting to take this symmetry into account. Hence, one can say that this two-fold division of the market network admits duality symmetry. Duality symmetries are some of the most interesting symmetries in physics. The term duality is used to refer to the relationship between two systems that have different descriptions but identical physics (identical trading operations).

Let us analyze the 1-endpoint elementary string map when the currency changes from direct to indirect. The change can be formalized by means of the transformation

$$\hat{T}_{\text{id}} : P\{p(\cdot)\} \rightarrow \bar{P}\{p(\cdot)\} \equiv P\{1/p(\cdot)\}, \quad (2.18)$$

where we use the notation  $P\{p\}$  which emphasizes the functional dependence upon the currency exchange rate  $\{p\}$ . It should also be noted that the use of  $P$  highlights the canonical formal correspondence between the rate of return and the internal string momentum.

For the 1-endpoint map model of the string, see Eq. (2.7), we obtain

$$\hat{T}_{\text{id}} P_q^{(1)}(\tau, h) = \bar{P}_q^{(1)}(\tau, h) = f_q \left( \frac{p(\tau) - p(\tau + h)}{p(\tau)} \right). \quad (2.19)$$

Let us consider two-member space of maps  $V_P^{(1)} = \{P_q^{(1)}, \bar{P}_q^{(1)}\}$ . What is important, we see that  $\hat{T}_{\text{id}}$  preserves the Dirichlet boundary conditions and the identity operator  $\hat{T}_{\text{id}}^2$  leaves the elements of  $V_P^{(1)}$  unchanged. The space  $V_P^{(1)}$  is closed under the left action of  $\hat{T}_{\text{id}}$ . These ideas are straightforward transferable to the 2-endpoint string points.

Now we omit the notation details and proceed according to Eq. (2.18). The map  $P(\cdot)$  is

decomposable into a sum of symmetric and antisymmetric parts

$$P^S = \frac{1}{2}(P + \bar{P}), \quad P^A = \frac{1}{2}(P - \bar{P}), \quad (2.20)$$

respectively. Due to of normalization by  $1/2$ , we get the projection properties

$$\hat{T}_{\text{id}} P^S = P^S, \quad \hat{T}_{\text{id}} P^A = -P^A. \quad (2.21)$$

To be more concrete, we choose  $q = 1$  and obtain

$$P_{q=1}^{(1),S} = 1 - \frac{1}{2} \left[ \frac{p(\tau)}{p(\tau+h)} + \frac{p(\tau+h)}{p(\tau)} \right], \quad P_{q=1}^{(1),A} = \frac{1}{2} \left[ \frac{p(\tau)}{p(\tau+h)} - \frac{p(\tau+h)}{p(\tau)} \right]. \quad (2.22)$$

and

$$\begin{aligned} P_1^{(2),A} &= \frac{1}{2} \left[ \frac{p(\tau)}{p(\tau+l_s)} - \frac{p(\tau+l_s)}{p(\tau)} + \frac{p(\tau+h)}{p(\tau)} - \frac{p(\tau)}{p(\tau+h)} \right. \\ &\quad \left. + \frac{p(\tau+l_s)}{p(\tau+h)} - \frac{p(\tau+h)}{p(\tau+l_s)} \right], \quad (2.23) \\ P_1^{(2),S} &= 1 + \frac{1}{2} \left[ \frac{p(\tau+l_s)}{p(\tau)} + \frac{p(\tau)}{p(\tau+l_s)} - \frac{p(\tau)}{p(\tau+h)} - \frac{p(\tau+h)}{p(\tau)} \right. \\ &\quad \left. - \frac{p(\tau+h)}{p(\tau+l_s)} - \frac{p(\tau+l_s)}{p(\tau+h)} \right]. \end{aligned}$$

We see that the  $P_{q=1}^{(1),S}$  and  $P_{q=1}^{(2),S}$  maps acquire formal signs of the systems with  $T$ -dual symmetry [3]. When the world described by the closed string of the radius  $R$  is indistinguishable from the world of the radius  $\propto 1/R$  for any  $R$ , the symmetry manifests itself by  $(R \pm \text{const.}/R)$  terms of the mass squared operator. The correspondence with our model becomes apparent one assumes that  $R$  corresponds to the ratio  $p(\tau)/p(\tau+h)$  in Eq. (2.22). However, we must also refer that in our model we do not consider for the moment the compact dimension. One can also find in the option price dynamics some real examples of duality symmetry [44]. Concretely put–call duality which means “A call to buy foreign with domestic is equal to a put to sell domestic for foreign.” Also most questions will not spell out what is domestic or foreign but let you decide what is the underlying asset and which is the strike asset.

### $\hat{T}_{\text{id}}$ transform under the conditions of bid-ask spreads

The generalization can also be made with the currency variables which appear as a consequence of the transaction costs [45]. The occurrence of ask-bid spread complicates the analysis in several ways. Instead of one price for each currency, the task requires the availability of two prices. The impact of ask-bid spread on the time-series properties has been studied within the elementary model [46].

Thus, for the purpose of a thorough and more realistic analysis of the market information, it seems straightforward to introduce generalized transform

$$\hat{T}_{\text{id}}^{\text{ab}} P\{p_{\text{ask}}(\cdot), p_{\text{bid}}(\cdot)\} = \bar{P}\{1/p_{\text{bid}}(\cdot), 1/p_{\text{ask}}(\cdot)\}, \quad (2.24)$$

which converts to Eq.(2.19) in the limit of vanishing spread.

## 2.4 D2-branes map

More interesting way how to go beyond a string model is to extent the string lines towards the more complex maps including alternative spread-adjusted currency returns, the membranes called D2-branes. Practically, the generalized mapping onto the D2-brane with the  $(h_1, h_2) \in \langle 0, l_s \rangle \times \langle 0, l_s \rangle$  coordinates which vary along two extra dimensions could be proposed in the form

$$P_{D2,q}(\tau, h_1, h_2) = f_q \left( \frac{p_{\text{ask}}(\tau + h_1) - p_{\text{ask}}(\tau)}{p_{\text{ask}}(\tau + h_1)} \frac{p_{\text{ask}}(\tau + l_s) - p_{\text{ask}}(\tau + h_1)}{p_{\text{ask}}(\tau + l_s)} \right. \\ \left. \times \frac{p_{\text{bid}}(\tau) - p_{\text{bid}}(\tau + h_2)}{p_{\text{bid}}(\tau)} \frac{p_{\text{bid}}(\tau + h_2) - p_{\text{bid}}(\tau + l_s)}{p_{\text{bid}}(\tau + h_2)} \right). \quad (2.25)$$

The map constituted by the combination of ‘‘bid’’ and ‘‘ask’’ quotes is constructed to satisfy the Dirichlet boundary conditions

$$P_{D2,q}(\tau, h_1, 0) = P_{D2,q}(\tau, h_1, l_s) = P_{D2,q}(\tau, 0, h_2) = P_{D2,q}(\tau, l_s, h_2). \quad (2.26)$$

The above construction, Eq. (2.25), has been chosen as an explicit example, where the action of  $\hat{T}_{\text{id}}^{\text{ab}}$  becomes equivalent to the permutation of coordinates

$$\hat{T}_{\text{id}}^{\text{ab}} P_{D2,q}(\tau, h_1, h_2) = P_{D2,q}(\tau, h_2, h_1). \quad (2.27)$$

Thus the symmetry with respect to interchange of extra dimensions  $h_1, h_2$  can be achieved through  $P_{D2,q} + \hat{T}_{\text{id}}^{\text{ab}} P_{D2,q}$ . In a straightforward analogous manner one can get an antisymmetric combination  $P_{D2,q} - \hat{T}_{\text{id}}^{\text{ab}} P_{D2,q}$ .

At the end of this subsection, we consider the next even simple example, where mixed boundary conditions take place. Now let the 2-endpoint string to pass to the 1-endpoint string by means of the homotopy  $P_{q_1, q_2}^{(1,2)}(\tau, h, \eta) = (1 - \eta)P_{q_1}^{(1)}(\tau, h) + \eta P_{q_2}^{(2)}(\tau, h)$  driven by the parameter  $\eta$  which varies from 0 to 1. In fact, this model can be seen as a variant of the D2 brane with extra dimensions  $h$  and  $\eta$ .

## 2.5 Partial compactification

The numerical analysis of the string statistics have been performed qualitatively by the models of the maps of the exponential and periodic data inputs. Most of the numerical investigations have been obtained for the open topology; however, we describ briefly the ways to partial compactification (pc). The data structures can also be mapped by means of the curled dimension which arises as a sum of periodic data contributions. The idea of the compactified strings can be realized as well by the application of the inverse Fourier transform of the original signal.

In the framework of string theory, the 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, the compactification may be viewed as an act of the 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

$$\tilde{p}(\tau) = \frac{1}{N_m} \sum_{m=0}^{N_m-1} p(\tau + l_s m), \quad (2.28)$$

where the input of any open string (e. g., Eq. (2.5), (2.11)) 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, the deviations of  $\tilde{p}$  from the periodic signal become less pronounced.

For example, one might consider the construction of the  $(\tilde{D} + 1)$ -brane

$$f_q \left( \frac{p(\tau + h_0) - p(\tau)}{p(\tau + h_0)} \right) \prod_{j=1}^{\tilde{D}} f_q \left( \frac{\tilde{p}_j^{(\pm)}(\tau + h_j) - \tilde{p}_j^{(\pm)}(\tau)}{\tilde{p}_j^{(\pm)}(\tau + h_j)} \right) \quad (2.29)$$

maintained by combining  $(\tilde{D} + 1)$  1-endpoint strings, where partial compactification in  $\tilde{D}$  extra dimensions is supposed. Of course, the construction introduces auxiliary variables  $\tilde{p}_j^{(\pm)}(\tau) = \sum_{m=0}^{N_m, j-1} p(\tau \pm m l_{s, j})$ .

## 2.6 Intra-string statistical picture

The idea of the string related maps proposed here is the transformation of the original point object such as selected single price into a system of a prices from its admissible neighborhood. This changeover from a local to a non-local description directly extends econometrics belief that future prices are deducible from the price history of a given period.

### Time invariant strings and elementary statistics

To understand better the idea, let us suppose to deal with short-time evolution of the currency  $\ln p(\tau) = \ln p_0 + b\tau$  characterized by the linear-logarithmic parameter  $b$ . After the substitution into the string, Eq. (2.11), the expressions collapse to the invariant (independent of  $\tau$ ) form

$$\begin{aligned} P_1^{(1)}(\tau, h) &= 1 - \exp(-hb), \\ P_1^{(2)}(\tau, h) &= 1 - \exp[(h - l_s)b] - \exp(-hb) + \exp(-l_s b). \end{aligned} \quad (2.30)$$

It is quite interesting to look at the lowest order terms of Taylor series of this result around  $b = 0$ . Ignoring the terms of order  $b^4$  or higher gives  $P_1^{(1)} = bh - \frac{1}{2}b^2h^2 + \mathcal{O}(b^3)$ ,  $P_1^{(2)} = b^2h(l_s - h) + \mathcal{O}(b^3)$ . It demonstrates that the model of the 1-endpoint string is more sensitive to the sign of  $b$  variations.

The elementary qualitative statistical model of the string can be obtained by taking the unexamined assumption that  $b$  fluctuates with the Gaussian probability density  $(2\pi\sigma_b^2)^{-1} \exp(-\frac{b^2}{2\sigma_b^2})$ .

The averaging of Eq. (2.30) with this weight yields

$$\begin{aligned}\langle P_1^{(1)} \rangle_{\text{Gauss}}(h) &= 1 - \exp\left(-\frac{h^2\sigma_b^2}{2}\right), \\ \langle P_1^{(2)} \rangle_{\text{Gauss}}(h) &= 1 - \exp\left(-\frac{h^2\sigma_b^2}{2}\right) - \exp\left(-\frac{(h-l_s)^2\sigma_b^2}{2}\right) + \exp\left(-\frac{l_s^2\sigma_b^2}{2}\right).\end{aligned}\quad (2.31)$$

Eq. (2.31) predicts an increase in the  $h$  dependence as a consequence of the symmetric fluctuations in  $b$ .

### Mapping of the periodic input signal

Simultaneously with statistical averaging of data, it is instructive to briefly examine the string map of the signal of periodic form described by some elementary function. The input signal  $p(\tau) = a_1 + a_2 \cos(\omega\tau)$  can be suitable for this purpose. Subsequently, the analytic calculation for the 2-endpoint map can be carried out perturbatively under the requirement  $a_2 \ll a_1$ . The common form of average unifying formulas obtained for different integer  $q$  values can be written as

$$\langle P_q^{(2),(S)}(h) \rangle_{\text{cos}} = \left[ \cos(h\omega) + \cos((h-l_s)\omega) - \cos(l_s\omega) - 1 \right]^q \sum_{j=0} c_{q,j} \left( \frac{a_2}{a_1} \right)^{2(q+j)}, \quad (2.32)$$

where  $c_{q,j}$  are the numerical coefficients which are not critical for further reasoning. The intuitive idea that  $P_q^{(2),(A)}(\tau, h)$  discriminates fluctuations, which stems from the comparison of the symmetric and antisymmetric averages, is partially justified by the result  $\langle P_q^{(2),(A)}(h) \rangle_{\text{cos}} = 0$ .

Interestingly, the calculation highlights the idea of the presence of the resonant lengths  $l_s(n) = 2\pi n/\omega$  as  $n = 1, 2, \dots$ . This basic result motivated us to introduce the 2-endpoint string model, that has potential to identify characteristic dynamic scales represented here by  $1/\omega$ . It is important that anomalous aspect is absent in the statistical characteristics of the 1-endpoint strings.

### String map in the representation of internal Fourier modes

Following the analysis of ultra-high-frequency financial data using advanced Fourier transformation as in [47] one can assume that each tick  $\tau$  of the string may be represented by the sequence  $P_q^{(2)}(\tau, h)$ ,  $h = 0, 1, \dots, l_s - 1$  which can be transformed by means of the discrete Fourier transform

$$P_{\text{DFT},q}(k, \tau) = \sum_{h=0}^{l_s} P_q^{(2)}(\tau, h) \exp\left(-\frac{2\pi i k h}{l_s + 1}\right), \quad k = 0, 1, \dots, l_s. \quad (2.33)$$

Having done this, one can introduce the inverse transform

$$P_{\text{IDFT},q}(h, \tau) = \sum_{k=0}^{l_s} P_{\text{DFT},q}(\tau, k) \exp\left(\frac{2\pi i k h}{l_s + 1}\right), \quad (2.34)$$

which can be understood as a periodic extension of the input  $P_q^{(2)}(\tau, h)$  with a period of the  $(l_s + 1)$  ticks. Then  $P_{\text{IDFT},q}(h, \tau)$  can be viewed as a portion of the original signal which curls up along the compact dimension of the closed string. Thus, the integer  $h/(l_s + 1)$  has the meaning of a winding number of  $P_{\text{IDFT},q}(h, \tau)$ . The Fourier transform of the inherent string structure may serve to identify distinguishing features of currencies at selected time scale.

## 2.7 String polarized by the external field

In this section, we modify the 2-endpoint string model, Eq. (2.11), in order to account for the transaction costs. This is some kind of analogy with a charged string which can be polarized by an external electric field [48]. A natural way is to consider the relation for the spread-adjusted return  $(p_{\text{bid}}(\tau + h) - p_{\text{ask}}(\tau))/p(\tau + h)$  (written here for a long position). If we now routinely extend the 2-endpoint, Eq. (2.11), we obtain

$$P_q^{\text{ab}}(\tau, h) = f_q \left( \frac{p_{\text{bid}}(\tau + h) - p_{\text{ask}}(\tau)}{p(\tau + h)} \frac{p_{\text{bid}}(\tau + l_s) - p_{\text{ask}}(\tau + h)}{p(\tau + l_s)} \right). \quad (2.35)$$

However this clearly violates, the Dirichlet boundary conditions, see Eq. (2.12). The spread itself yields a negligible correction to the mean values.

The boundary conditions can be easily renewed by the subtraction  $\tilde{P}_q^{\text{ab}}(\tau, h) = P_q^{\text{ab}}(\tau, h) - P_q^{\text{ab}}(\tau, 0)$ . However, we show there exists a more fundamental alternative way which reflects a bid-ask difference and preserves the Dirichlet boundary conditions. The string states are polarized by the instant possibility to place a successful/unsuccessful buy order. For each  $h$  and  $Y = A, S$  we construct the inequality constrained sequence

$$P_{(q,+)}^{(2),Y}(\tau + 1, h) = \begin{cases} P_q^{(2),Y}(\tau, h), & p_{\text{bid}}(\tau + l_s) \geq p_{\text{ask}}(\tau), \\ P_{(q,+)}^{(2),Y}(\tau, h), & \text{otherwise,} \end{cases} \quad (2.36)$$

and non-buy contributions, respectively,

$$P_{(q,-)}^{(2),Y}(\tau + 1, h) = \begin{cases} P_q^{(2),Y}(\tau, h), & p_{\text{bid}}(\tau + l_s) < p_{\text{ask}}(\tau), \\ P_{(q,-)}^{(2),Y}(\tau, h), & \text{otherwise.} \end{cases} \quad (2.37)$$

In both cases it is supposed that  $P_q^{(2),Y}(\tau, h)$  is calculated according an unconditioned model defined by Eq. (2.11). Now, to characterize the arbitrage opportunities, we introduced the statistical polarization measure in the form

$$g_{q,Y} = \left\langle \frac{\sum_{h=0}^{l_s} |P_{(q,+)}^{(2),Y}(\tau, h) - P_{(q,-)}^{(2),Y}(\tau, h)|}{\sum_{h=0}^{l_s} |P_{(q,+)}^{(2),Y}(\tau, h) + P_{(q,-)}^{(2),Y}(\tau, h)|} \right\rangle, \quad Y = A, S. \quad (2.38)$$

These ideas could be extended to the case of market polarizations which are due to interactions amongst the agents in the market [49]. The agents could be characterized by the strings.

We continue the characterization of arbitrage opportunities by defining the (momentum) distance function between the strings as

$$d_q^{(Y)}(\tau) = \frac{1}{l_s + 1} \sum_{h=0}^{l_s} \left| P_{(q,+)}^{(Y)}(\tau, h) - P_{(q,-)}^{(Y)}(\tau, h) \right|. \quad (2.39)$$

In this case the statistics of string distances can be characterized by the customized variant of the well-known model of the correlation sum [50, 51, 52] or by the correlation structure patterns for identifying states of a financial market [41]. However, the motivation here differs from that given in these papers, where the intent was to analyze nonlinear relationships. The correlation sum shows the probability that the states of two strings or branes are localized within a certain distance. In our view we adjust the original formula to the string and brane models which, in addition, reflect the transactions involving profits. We define the measure

$$C_q^{(Y)}(\varepsilon) = \frac{\langle \Theta(\varepsilon - d_q^{(Y)}(\tau)) \rangle}{\int d\varepsilon' \langle \Theta(\varepsilon' - d_q^{(Y)}(\tau)) \rangle}, \quad (2.40)$$

where  $\varepsilon$  is the threshold distance,  $\Theta(\cdot)$  is the Heaviside step function; here  $l_s$  plays the role of so called embedding dimension. The key concept surrounding this measure is the concept of the fractal dimension. Further, one can extend the concept of distance for D2-branes as defined in Eq.(2.25). In this case we suggest generalization

$$d_{D2,q}(\tau) = \frac{1}{(l_s + 1)^2} \sum_{h_1=0}^{l_s} \sum_{h_2=0}^{l_s} \left| P_{D2,(q,+)}(\tau, h_1, h_2) - P_{D2,(q,-)}(\tau, h_1, h_2) \right|. \quad (2.41)$$

## 2.8 Inter-currency study: map onto rotating strings

The incorporating of the mutual relations between the pairs into the mapping procedure represents a very challenging task. Let us study trading activity in the  $(I, J)$  plane, where  $I, J$  stands for indices of the currency pair described by two 2-end-point strings. The real time data are used instead of tick by tick (Eq. (2.3)) in order to maintain the consistency of prices quoted.

By continuing examples of the generalized distance concept (Eq. (2.39) and Eq. (2.41)), we can introduce the inter-currency momentum distance function

$$d_{q,I,J}(t) = \frac{1}{l_s + 1} \sum_{h=0}^{l_s} \left| P_{q,I}^{(2)}(t, h) - P_{q,J}^{(2)}(t, h) \right|. \quad (2.42)$$

At higher dimension, it is tempting to deal with angular momentum

$$M_{q,I,J}(t) = \sum_{h=0}^{l_s} \left[ P_{q,I}^{(2)}(t, h) X_{q,J}^{(2)}(t, h) - P_{q,J}^{(2)}(t, h) X_{q,I}^{(2)}(t, h) \right]. \quad (2.43)$$

The momentum calculation can be interpreted as a measure of the rotational information flows between the currency pairs. Simultaneously, the concept of distance and moment has been extended

to analyze the impact of spread. Analogously, as in the previous cases, the distance between the ask and bid strings may be defined

$$d_{q,I}^{\text{ab}}(t) = \frac{1}{l_s + 1} \sum_{h=0}^l \left| P_{q,\text{ask}}^{(2)}(t, h) - P_{q,\text{bid}}^{(2)}(t, h) \right|, \quad (2.44)$$

$$M_{q,I}^{\text{ab}}(t) = \sum_{h=0}^{l_s} \left[ P_{q,\text{ask}}^{(2)}(t, h) X_{q,\text{bid}}^{(2)}(t, h) - P_{q,\text{bid}}^{(2)}(t, h) X_{q,\text{ask}}^{(2)}(t, h) \right]. \quad (2.45)$$

Here

$$P_{q,\text{ask}}^{(2)} \equiv P_q^{(2)}|_{p \rightarrow p_{\text{ask}}}, \quad P_{q,\text{bid}}^{(2)} \equiv P_q^{(2)}|_{p \rightarrow p_{\text{bid}}} \quad (2.46)$$

are obtained by substituting expressions above in Eq. (2.11). With the help of Eq.(2.13) and  $P_{q,\text{ask}}^{(2)}$ ,  $P_{q,\text{bid}}^{(2)}$  we construct iteratively  $X_{q,\text{ask}}^{(2)}$  and  $X_{q,\text{bid}}^{(2)}$ . The differences measured in terms of  $M_{q,I,J}$  are very subtle. The fundamental role in the string theory is played by the Regge slope parameter  $\alpha'$  (or inverse string tension). This has a proper analogy with our approach where we introduced a slope in terms of the angular momentum

$$\alpha'_{q,I,J} = \frac{\langle |M_{q,I,J}| \rangle}{2\pi l_s^2}. \quad (2.47)$$

It is worth to note that relation (2.47) should be understood as an estimate since there is no statistical mean of the type  $\langle |M_{\dots}| \rangle$  in the original specification. The problem of estimation of the slope parameter arises from the fact that in the original model nonaveraged angular momentum is divided by the square of the mass instead of  $l_s^2$ .

## 2.9 Differentials of string map

Gâteaux derivative is a generalization of the concept of a directional derivative in the differential calculus. In our study the concept can be viewed as a systematic way in the generation of more structured maps expressing more information about the structure of data we deal with. Given the string map  $P(\cdot)$ , the  $m$ -th Gâteaux derivative of  $P(\cdot)$  in the “direction” of  $\psi(\cdot)$  (unspecified yet series) is defined as follows

$$d^m P(\{p\}; \{\psi\})(\tau, h) = \frac{d^m}{d\varepsilon^m} P(\{p(\tau, h) + \varepsilon\psi(\tau, h)\}) \Big|_{\varepsilon \rightarrow 0}. \quad (2.48)$$

For  $q = 1$  the calculation gives

$$dP_1^{(1)}(\{p\}; \{\psi\})(\tau, h) = \frac{1}{p(\tau + h)} \left[ \frac{p(\tau)\psi(\tau + h)}{p(\tau + h)} - \psi(\tau) \right] \quad (2.49)$$

and

$$\begin{aligned} dP_1^{(2)}(\{p\}; \{\psi\})(\tau, h) &= \psi(\tau) \left( \frac{1}{p(\tau + l_s)} - \frac{1}{p(\tau + h)} \right) \\ &+ \psi(\tau + h) \left( \frac{p(\tau)}{p^2(\tau + h)} - \frac{1}{p(\tau + l_s)} \right) \\ &+ \frac{\psi(\tau + l_s)}{p^2(\tau + l_s)} (p(\tau + h) - p(\tau)). \end{aligned} \quad (2.50)$$

By going to the second order we obtained

$$d^2 P_1^{(2)}(\{p\}; \{\psi\})(\tau, h) = \frac{2\psi(\tau + h)}{p^2(\tau + h)} \left[ \psi(\tau) - \frac{p(\tau)\psi(\tau + h)}{p(\tau + h)} \right] \quad (2.51)$$

and

$$\begin{aligned} d^2 P_1^{(2)}(\{p\}; \{\psi\})(\tau, h) &= \frac{2\psi(\tau + l_s)}{p^2(\tau + l_s)} [\psi(\tau + h) - \psi(\tau)] \\ &+ \frac{2\psi(\tau + h)}{p^2(\tau + h)} \left[ \psi(\tau) - \frac{p(\tau)\psi(\tau + h)}{p(\tau + h)} \right] \\ &+ \frac{2\psi^2(\tau + l_s)}{p^3(\tau + l_s)} [p(\tau) - p(\tau + h)]. \end{aligned} \quad (2.52)$$

The generalized differentiation generates maps which satisfy the Dirichlet boundary conditions

$$d^m P_1^{(1)}(\{p\}; \{\psi\})(\tau, 0) = 0, \quad m = 1, 2; \quad (2.53)$$

$$d^m P_1^{(2)}(\{p\}; \{\psi\})(\tau, 0) = d^m P_1^{(2)}(\{p\}; \{\psi\})(\tau, l_s) = 0. \quad (2.54)$$

Many alternative ways exist to exploit the models with the auxiliary field  $\psi(\cdot)$ . The field can be related to, e. g.,

- (i) models which place emphasis on the currency margins determined by some adaption process,
- (ii) on the spread in a style of Subsec. 2.7 with  $\psi(\tau) = p_{\text{bid}}(\tau) - p_{\text{ask}}(\tau - l_s)$ ,
- (iii) the periodic function  $\psi(\tau)$  can model the action of the compact,
- (iv) the benchmark setting represents  $\psi = 1$ .

In the last case, one can see that the generalized derivative modifies the original map as follows

$$dP_1^{(1)}(\{p\}; \{\psi\})(\tau, h)|_{\psi=1} = -\frac{P_1^{(1)}(\tau, h)}{p(\tau + h)}, \quad (2.55)$$

$$dP_1^{(2)}(\{p\}; \{\psi\})(\tau, h)|_{\psi=1} = \left( \frac{1}{p(\tau + h)} + \frac{1}{p(\tau + l_s)} \right) P_1^{(2)}(\tau, h). \quad (2.56)$$

### 3 String prediction models and their utilization

#### 3.1 Prediction model based on string invariants

The meaning of invariant is that something does not change under transformation, e. g., such as some equations from one reference frame to another. We want to extend this idea also on the time-series forecast, to find some invariants in the finance data and utilize them to predict the future values. Let us introduce a positive integer  $l_{pr}$  denoting the prediction scale of how many steps ahead of  $\tau_0$  lies the predicted value. Let us introduce an auxiliary positive integer  $\Lambda$  and a condition

$$\Lambda = l_s - l_{pr}, \quad l_s > l_{pr}. \quad (3.1)$$

The power of the nonlinear string maps of time series data is to be utilized to establish a prediction model similarly as in [53,54,55]. The approach to define the string invariants is based on the correlation function (see Appendix A).

The first tests of prediction behavior of the prediction model based on string invariants (PMBSI) with and without transaction costs on the EUR/USD currency rate of the forex market were obtained by using the values of parameters  $l_s = 900$ ,  $l_{pr} = 1$ ,  $\eta_1 = 0$ ,  $\eta_2 = 0$ ,  $q = 6$ . Their graphical description is plotted in Figs. 3.1 and 3.2.

The main weakness of this model is its prediction length (the parameter  $l_{pr}$ ), in this case it is one tick ahead. The price was predicted correctly in 48.57% of all cases (16201 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  $l_{pr}$  induces a loss of the prediction

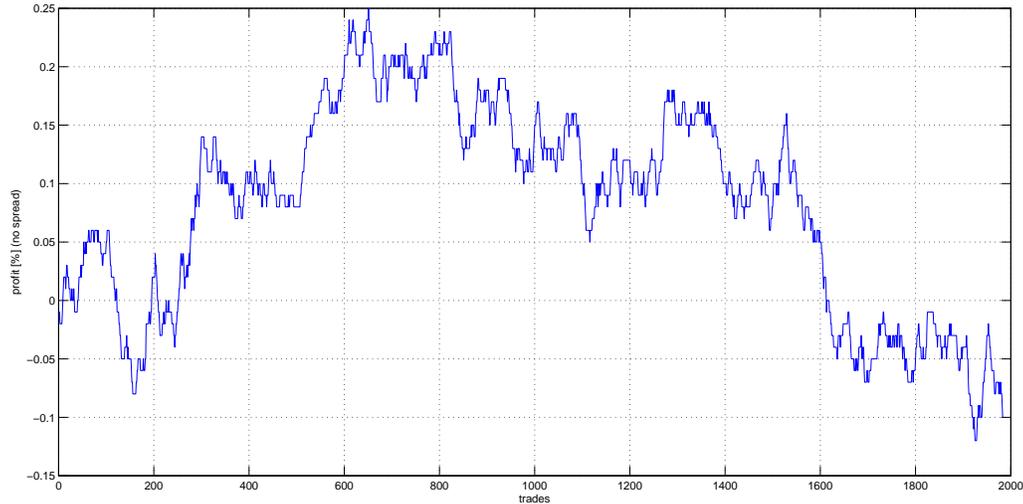


Fig. 3.1: The profit of PMBSI model on the EUR/USD currency rate without transaction costs included dependence on trades for one year period.

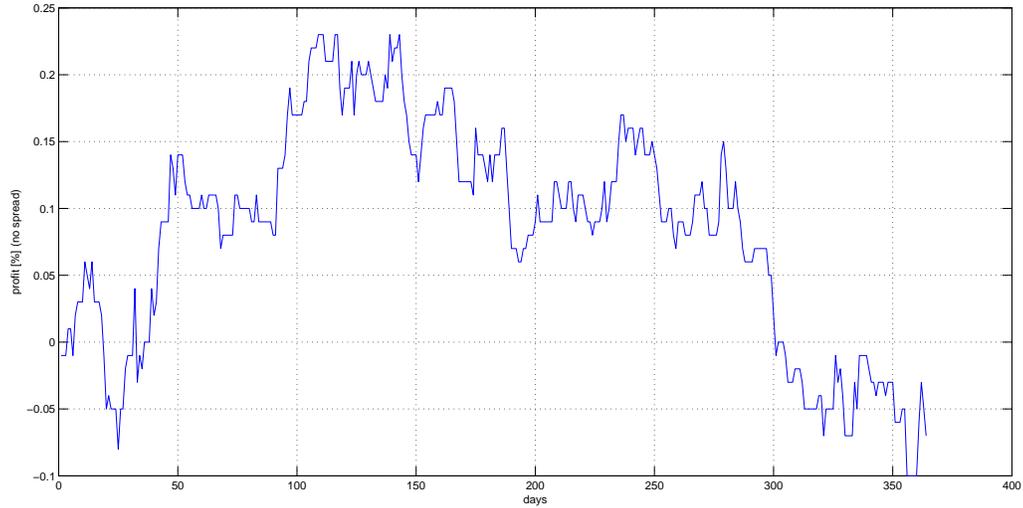


Fig. 3.2: The profit of PMBSI model on the EUR/USD currency rate without transaction costs included dependence on days for one year period.

strength of the model. For example when  $l_{pr} = 2$  (two ticks ahead) prediction strength decreases from around 50% to 15%.

The problem is that the invariant Eq. (A.1) is fulfilled only on the very short period of the time series due to the very chaotic nature of financial data behaviour. 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, 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 Eqs. (A.8), (A.9). The right and optimized weights should considerably extend the interval where Eq. (A.1) is fulfilled.

### 3.2 Experimental analysis of PMBSI

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 the period of 1295 days. The performance of PMBSI was compared to SVM and to naïve forecast. There were two error measures used, mean absolute error (MAE) and symmetric mean absolute percentage error (SMAPE) defined as

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |A_t - F_t|, \quad (3.2)$$

$$\text{SMAPE} = \frac{100}{n} \sum_{t=1}^n \frac{|A_t - F_t|}{0.5(|A_t| + |F_t|)}, \quad (3.3)$$

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, 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 forecast for the validation data (the most recent) that were never used in the model optimization process.

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 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 models.

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).

### Experimental results on the artificial time series

A single period of a sinusoid sampled by 51 regularly spaced samples was used. They were divided into subsets so that the positive half of the period was used for training and evaluation and the negative half for validation. This was done to assess the ability of PMBSI to extrapolate and generalize. For PMBSI the time series was shifted above zero by adding a positive constant. The constant was then subtracted from the forecast. SVM with linear kernel was used as a benchmark. The positive half of the period was divided 7/3 for training/validation. Predictions of 1, 2 and 3 steps ahead were made. It became obvious that PMBSI performs well in one step ahead prediction but for multiple steps ahead predictions its performance drops rapidly. Therefore, iterated prediction using the one step prediction model was made, improving the PMBSI results significantly. For the illustration, Fig. 3.3 shows the comparison of iterated versus the direct prediction using PMBSI. Experimental results on the evaluation and validation data are presented in Table 3.1, the results of the best performing models are highlighted.

The optimal  $l_{tw}$  for SVM was 3 for all predictions. Table 3.2 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.

### Experimental results on the financial time series

The financial time series was divided into subsets so that the most recent 40% of the data was used for validation and the remaining data were used for training/validation divided in the ratio of 6/4. While extrapolation of sigmoid was a relatively simple task, the financial time series was highly non-linear and chaotic. SVM with Gaussian RBF kernel was used as the benchmark and the predictions from 1 to 10 steps ahead were made. Table 3.3 shows a selection of the experimental results, the results of the best performing models are highlighted. Table 3.4 summarizes

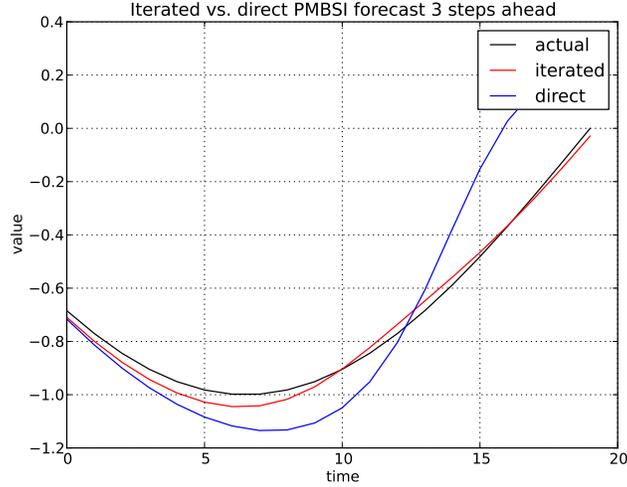


Fig. 3.3: Iterated and direct prediction using PMBSI on artificial data.

Method	$l_{pr}$	MAE eval	MAE valid	SMAPE valid
PMBSI	1	0.000973	<b>0.002968</b>	<b>8.838798</b>
	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	<b>0.007703</b>	<b>10.711573</b>
	3	0.012412	<b>0.007322</b>	<b>11.551324</b>
Naïve forecast	1	—	0.077947	25.345352
	2	—	0.147725	34.918149
	3	—	0.207250	41.972591

Tab. 3.1: Experimental results on the evaluation and validation of artificial time series data.

$l_{pr}$	$l_s$	$q$	$\eta_1$	$\eta_2$
1	2	0.30	0.80	-0.20
2	5	0.10	0.80	-0.60
3	8	0.10	0.80	-0.60

Tab. 3.2: Optimal values of PMBSI parameters for artificial time series data.

the optimal parameters found and states the percentual count of NaNs forecast by PMBSI. Interestingly, SVM preferred long time windows reaching the upper limit on the length while PMBSI utilized much less of the past data to make a forecast.

Method	$l_{pr}$	MAE eval	MAE valid	SMAPE valid
PMBSI	1	0.023227	0.023595	7.380742
	2	0.037483	0.036335	11.378275
	4	0.048140	0.046381	14.876330
	6	0.054556	0.049755	16.094349
	8	0.057658	0.056097	18.546008
	10	0.060192	0.058216	18.752986
Iterated PMBSI	1	—	—	—
	2	0.032706	0.031940	9.953547
	4	0.043134	0.042414	13.250729
	6	0.049916	0.047784	15.102693
	8	0.055326	0.051355	16.306971
	10	0.057802	0.052353	16.552731
SVM	1	0.021383	0.025546	8.046289
	2	0.027721	0.031878	10.046793
	4	0.036721	<b>0.039702</b>	<b>12.578553</b>
	6	0.041984	<b>0.044450</b>	<b>14.157343</b>
	8	0.044525	<b>0.047175</b>	<b>15.036534</b>
	10	0.046166	<b>0.050236</b>	<b>15.898355</b>
Naïve forecast	1	—	<b>0.023273</b>	<b>7.287591</b>
	2	—	<b>0.031486</b>	<b>9.822408</b>
	4	—	0.041811	13.078883
	6	—	0.047238	14.958371
	8	—	0.050788	16.148619
	10	—	0.051923	16.428804

Tab. 3.3: Experimental results on the financial time-series for PMBSI.

$l_{pr}$	SVM			PMBSI		NaN(%)
	$l_{tw}$	$l_s$	$q$	$\eta_1$	$\eta_2$	
1	51	2	20.3	0.0	0.0	20.43
2	51	5	15.5	0.0	-0.05	14.6
4	51	8	11.9	0.0	0.0	19.25
6	51	10	14.3	0.1	-0.05	21.42
8	51	18	16.4	0.4	0.10	25.69
10	51	14	11.9	0.3	0.05	21.32

Tab. 3.4: Optimal parameters on the financial time-series and percent of NaNs forecast by PMBSI.

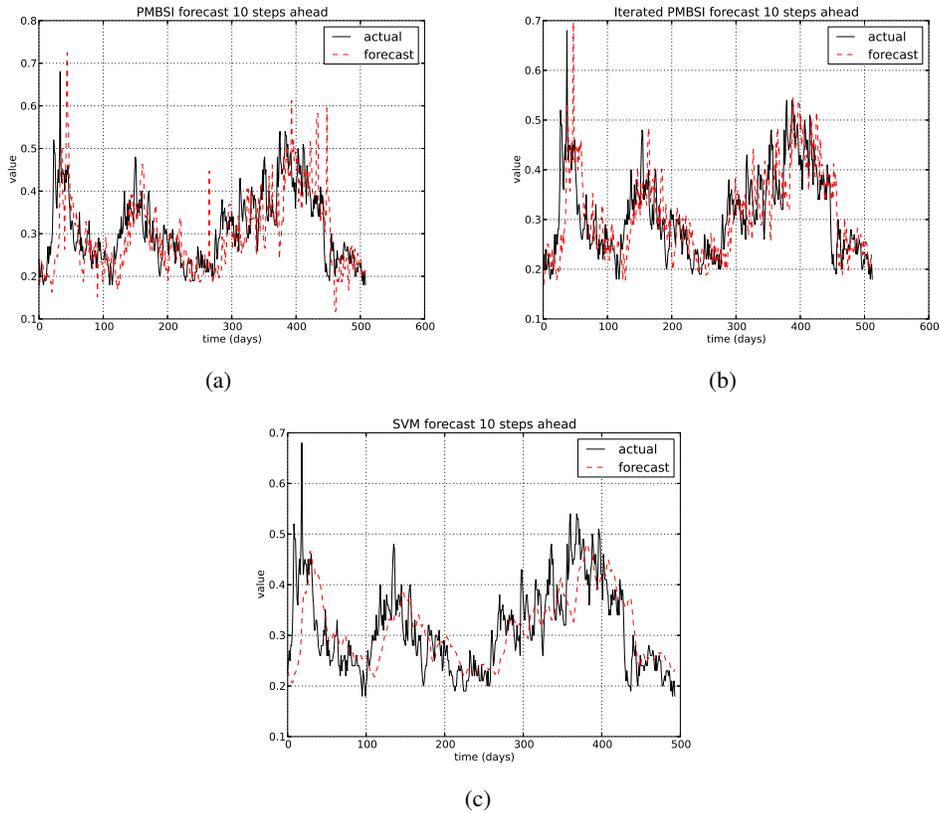


Fig. 3.4: The financial time-series and the forecast 10 steps ahead for PMBSI (a), iterated PMBSI (b), SVM (c) methods.

Again, the prediction accuracy of PMBSI was deteriorating significantly for longer forecasts and the results have improved significantly with iterated prediction. The longest prediction of 10 steps ahead was chosen to depict the experimental results on the financial data graphically (Fig. 3.4).

The goal of performed experimental analysis of the PMBSI method was to evaluate the prediction accuracy, the generalization performance, the convenience of the method in terms of the operators 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.

The prediction capability of PMBSI was proven and it was shown that it can match and even outperform SVM in some cases (see the results on the artificial data). On the financial data both methods, SVM and PMBSI struggled to match the naïve forecast. The reason for this is probably the complexity, intrinsic variability and chaotic nature of the system the time series is describing. Although the tests of PMBSI method on a larger set of time series are on the way, the presented

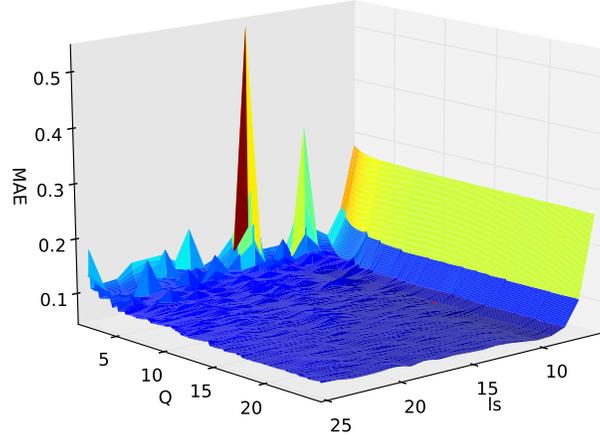


Fig. 3.5: MAE corresponding to various settings of  $l_s$  and  $q$  on the financial data. The red dot is the global minimum of MAE.

results have proven that PMBSI can be successfully used for single step forecast. The problem is that the more chaotic is the time series the shorter is the period when the invariant Eq. (A.17) is fulfilled. Therefore PMBSI is effective for the single step prediction because the probability of a significant change in the time series is lower. The situation is different for multiple steps prediction leading to small efficiency in such cases.

The way to improve the performance in multiple steps predictions is to choose more appropriate weighting coefficients (Eqs. (A.8), (A.9)). The optimized weights should considerably extend the interval where Eq. (A.17) is fulfilled. 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,  $\alpha$  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:  $l_s$ ,  $Q$ ,  $\eta_1$ ,  $\eta_2$ . This, clearly, is an advantage. On the other hand the optimal setting of the parameters is not easy to be found as there are many local minima on the error surface. In this analysis the optimal setting was found by testing of all combinations of parameters sampled from given ranges. Fig. 3.5 shows the Mean Absolute Error of the 5-steps ahead forecast of the financial time series corresponding to various settings of  $l_s$  and  $Q$  ( $\eta_1 = \eta_2 = 0$ ). But the figure makes also obvious that PMBSI's performance is approximately the same for a wide range of settings, making it unnecessary to explore the whole error surface. These are the reasons why PMBSI model can be fast to construct and to deploy.

### 3.3 Prediction model based on the deviations from the closed string/pattern form

For the next trading strategy we want to define some real values of the string sequences. Therefore we define the momentum  $M$  which acquired values from the interval  $(0, 1)$ . The momentum

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 place in time series 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 time-series has 55% of efficiency to predict movement of price but in the invariant place of our trading strategy where Eq. (3.4) is almost invariant the efficiency of our predictor increased to 80%. Therefore the idea to find invariant in time series plays a crucial role in our trading strategy but one still needs to find an appropriate expression for such prediction.

The momentum  $M$  of the string (the predictor) is proposed for the study of deviations of string maps from benchmark string sequence in the form

$$M(l_s, m, q, \varphi) = \left( \frac{1}{l_s + 1} \sum_{h=0}^{l_s} |P_N(\tau, h) - F_{CS}(h, \varphi)|^q \right)^{1/q}, \quad (3.4)$$

for  $m, q > 0$ ,  $l_s$  is the string length.  $P_N(\tau, h)$  represents the generalized  $N$ -points string map, e. g.  $P_N(\tau, h) \rightarrow P_q^{(1)}(\tau, h)$  (see (2.5) and [24] where the  $q$ -deformed prediction model based on the deviations from benchmark string sequence of 1-endpoint string map  $P_q^{(1)}(\tau, h)$  was thoroughly studied)

$$P_q^{(1)}(\tau, h) = \frac{p(\tau + h) - p_{\min}(\tau)}{p_{\max}(\tau) - p_{\min}(\tau)}, \quad P_q^{(1)} \in (0, 1), \quad (3.5)$$

$$p_{\max}(\tau) = \max_{h \in \{0, 1, 2, \dots, l_s\}} p(\tau + h), \quad p_{\min}(\tau) = \min_{h \in \{0, 1, 2, \dots, l_s\}} p(\tau + h). \quad (3.6)$$

Such model yields to the momenta values depicted in Fig. 3.6(a). The regular function  $F_{CS}(h, \varphi)$  could be substituted by various periodic functions, for more see Eqs. (3.14—3.20), we have used the form

$$F_{CS}(h, \varphi) = \frac{1}{2} (1 + \cos(\tilde{\varphi})), \quad \tilde{\varphi} = \frac{2\pi m h}{l_s + 1} + \varphi. \quad (3.7)$$

We have shown in [25] and confirmed practically, by the hundreds of thousands of simulations that we have executed on high precision simulation computing platform (more details in Appendix B), that regular function does play an influential role in predictors' decision making.

We have found out, that there is a relation between the time series data, that are streamed to algorithm, regular function and the mid/long term trends discovery process predictor is executing. We have been able to find right parameters for predictor to follow these trends on training data during a certain time period, but the problem was that real time series contains a lot of irregularities, which made the mid/long term trend following very difficult to use. We have used this predictor for live trading and while entering into new positions (opening a position, which is an operation of buy/sell an investment instrument like EUR/USD) was statistically sufficient, market irregularities in the form of sudden unpredictable events, which causes sudden price changes, made it very difficult to exit the position (close the position) in right time.

It depends on money and risk management strategies we have used at the exit rule. While statistically, a percentage of predictions have looked sufficient enough from entry rule point of view, it have been almost impossible to fine tune exit rule in a way, that it overcomes the market

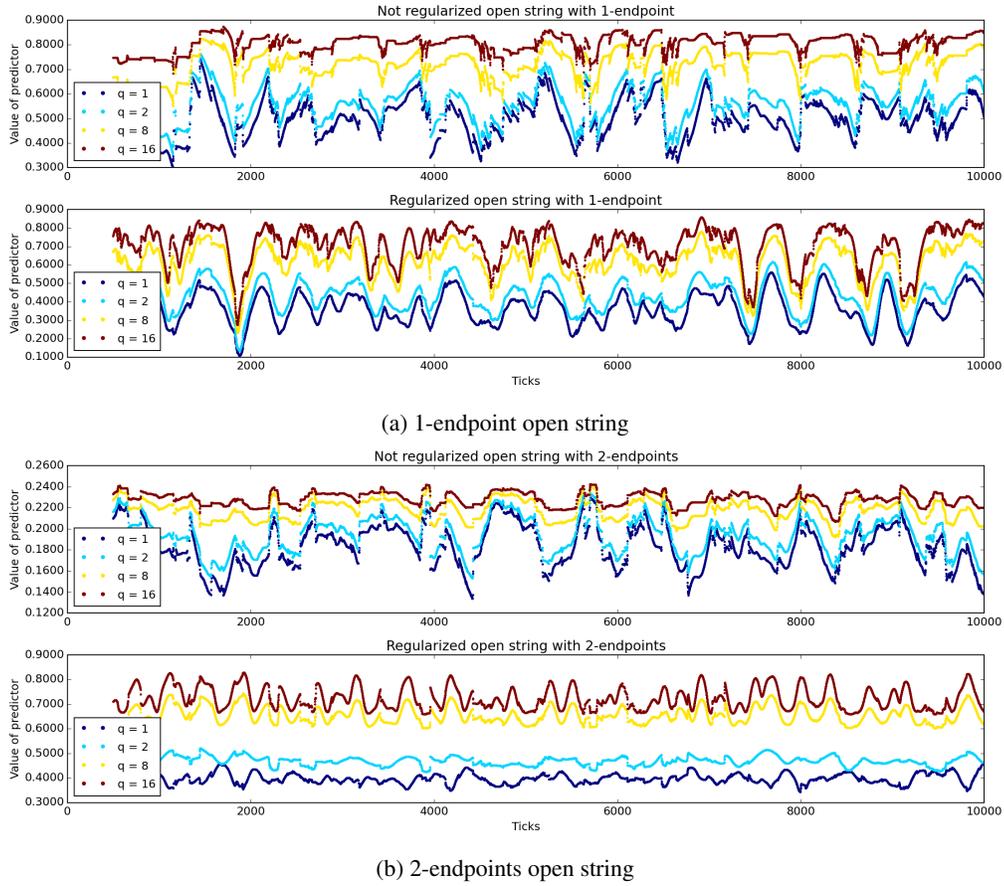


Fig. 3.6: Not regularized and regularized values of the momentum of the string for the sample of time series ticks in the case of 1-endpoint open string (a) and 2-endpoints open string (b) for typical values of  $q$  parameter.

irregularities. At the end, wrong set up of exit trade rule could cost more loss, than the entry rule based on predictors' decisions could cover.

Despite the simple approach, the results of the simulations with the prediction model in the OANDA market [29] have demonstrated the stability of the proposed trading algorithm on the transaction costs for the long trade periods. However, to avoid the situation with market irregularities lead us to find further possible string mappings.

### Open (polarized) string with 2-endpoints

In the next we study the influence of more complex string objects on the momentum behavior. At first we propose to incorporate a long-term trend by the nonlinear map corresponding to an

open string with 2-endpoints (see (2.11)) with the

$$P_N(\tau, h) \rightarrow P_q^{(2)}(\tau, h) = \left( \frac{p(\tau + h) - p(\tau)}{p_{\max}(\tau) - p_{\min}(\tau)} \right) \left( \frac{p(\tau + l_s) - p(\tau + h)}{p_{\max}(\tau) - p_{\min}(\tau)} \right), \quad (3.8)$$

with  $p_{\max}(\tau)$  and  $p_{\min}(\tau)$  from Eq. (3.6) or with the possible modifications of the string map to include spread-adjusted currency return  $(p_{\text{bid}}(\tau) - p_{\text{ask}}(\tau))/p(\tau, h)$ , i. e., the analogy with a charged string polarized by an external field, see Eq. (2.35)

$$P_N(\tau, h) \rightarrow P_q^{\text{ab}}(\tau, h) = \left( \frac{p^{\text{bid}}(\tau + h) - p^{\text{ask}}(\tau)}{p_{\max}(\tau) - p_{\min}(\tau)} \right) \left( \frac{p^{\text{bid}}(\tau + l_s) - p^{\text{ask}}(\tau + h)}{p_{\max}(\tau) - p_{\min}(\tau)} \right) \quad (3.9)$$

or

$$P_N(\tau, h) \rightarrow P_q^{\text{ab}} = \left( \frac{p^{\text{bid}}(\tau + h) - p^{\text{ask}}(\tau)}{p_{\max}^{\text{bid}}(\tau) - p_{\min}^{\text{ask}}(\tau)} \right) \left( \frac{p^{\text{bid}}(\tau + l_s) - p^{\text{ask}}(\tau + h)}{p_{\max}^{\text{bid}}(\tau) - p_{\min}^{\text{ask}}(\tau)} \right). \quad (3.10)$$

Fig. 3.6(b) shows that the effect of the regularization is notable in comparison with previous case of 1-endpoint open string, even for low values of  $q$  parameter. It allow us to focus on the predictor values which determine the stability of the algorithm or in other words they reflect the price changes on the scale of string length.

For the polarized string mapping, i. e., the replacement  $P_N(\tau, h) \rightarrow P_q^{\text{ab}}(\tau, h)$ , the regularized and nonregularized values of the momenta  $M$  looks identically to the previous case of open string (see Fig. 3.6(b)) and the simulations yield to the similar results.

To quantify the received predictor statistics one can construct the histograms for a spectrum of  $M$  momenta as shown in Fig. 3.7. Broader peaks of the distributions for regularized values of  $M$  for 1-endpoint and 2-endpoints strings suggests that the values are more smoothed than in the unregularized case and the aims to forecast the market trends are based on the sharper values of  $M$ , i. e., only the highest changes of a price on the market are taken into account and in this way they facilitate the evaluation of buy/sell orders.

### D2-brane model

More interesting way how to go beyond a string model is to extent the string lines towards the more complex maps, the membranes called D2-branes. Practically it can be realized with the mapping  $P_{\text{D2},q}(\tau, h_1, h_2)$  (see Eq. (2.25)) and the momentum of D2-brane model can be modified to

$$M(l_s, m, q, \varphi, \varepsilon) = \left( \frac{1}{(l_s + 1)^2} \sum_{h_1=0}^{l_s} \sum_{h_2=0}^{l_s} \left| P_{\text{D2},q}(\tau, h_1, h_2) - F_{\text{D2}}(h_1, h_2, \varphi, \varepsilon) \right|^q \right)^{1/q}, \quad (3.11)$$

the regular function depends also on more variables, e. g., it can has the form (3.20)

$$F_{\text{D2}}(h_1, h_2, \varphi, \varepsilon) = \frac{1}{2} (\sin(\tilde{\varphi}^2) \cos(\tilde{\varepsilon}^2)), \quad \tilde{\varphi} = \frac{2\pi m h_1}{l_s + 1} + \varphi, \quad \tilde{\varepsilon} = \frac{2\pi m h_2}{l_s + 1} + \varepsilon. \quad (3.12)$$

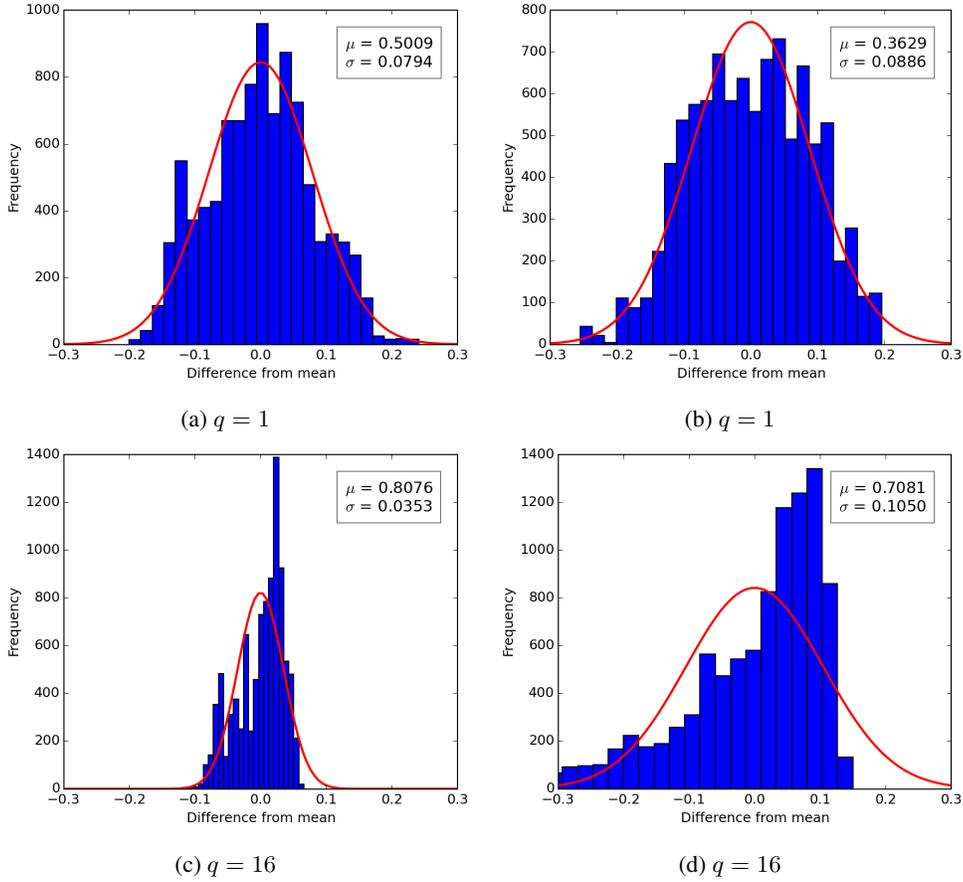


Fig. 3.7: Values of momenta for 1-endpoint open string without (left column) and with (right column) regularization function. Histograms show the difference of values from the mean of normal distribution function  $(\mu, \sigma)$ .

The effect of higher dimension D2-branes onto the  $M$  values in Eq. (3.11) is visible in Fig. 3.8. In comparison with 1-endpoint and 2-endpoints open strings the unregularized values are more smoothed. The regularization does not improve the spectrum so significantly as in the previous case of string models as it is visible from the histograms shown in Fig. 3.9. One can conclude that even the D2-branes model with basic configuration is suitable to capture the dynamic changes of prices on the financial market.

As another tool for evaluating of the different approaches represented by the string and D2-branes models can server the return volatility  $\sigma_{l_s/2}$ . In contrast to a historical volatility (the standard deviation of currency returns), the return volatility acts at the time scale  $l_s/2$  as string

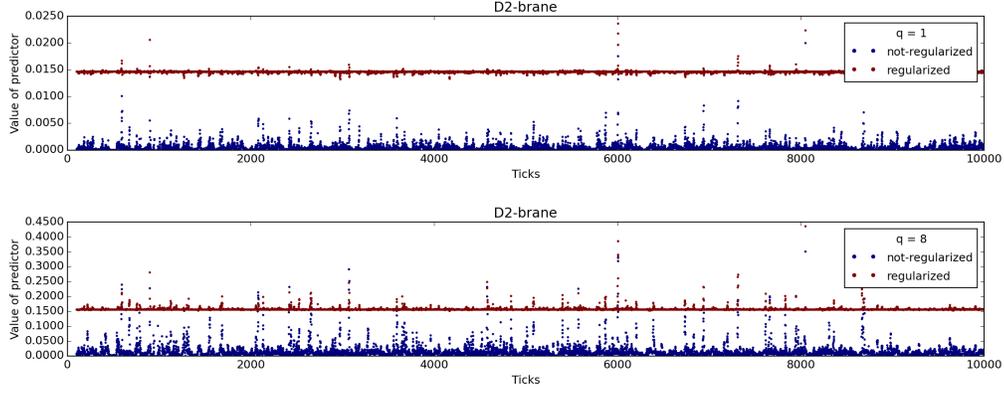


Fig. 3.8: Not regularized (blue) and regularized (red) values of the momenta for D2-brane. The sample of 10 thousand time series ticks,  $q = 1$  and  $q = 8$ .

statistical characteristic. It is defined as

$$\sigma_r(l_s/2) = \sqrt{r_2(l_s/2) - r_1^2(l_s/2)}, \quad (3.13)$$

$$r_m(l_s/2) = \sum_{h=1}^{l_s/2} [(p(\tau + h) - p(\tau + h - 1)) / (p(\tau + h))]^m,$$

for  $m = 1, 2$ . The scatterplot in Fig. 3.10 shows the relationship of return volatility at the scale of  $l_s/2$  to the changes in the price trends represented by the string amplitudes for 2-endpoints string  $P_i^{(2)}(\tau, l_s/2)$  and D2-brane  $P_{D2,i}(\tau, l_s/2, l_s/2)$ ,  $i = 1, 8$ .

The impact of high  $q$  to identify the rare events of volatility is visible in both cases, nevertheless, if one decides or does not decide to use the  $q$ -deformed model in favor of D2-branes it depends also on the technical conditions of real time calculations, because to receive the statistics and to make predictions with D2-branes requires more computing power.

### Regular functions

The formula for the basic regular function used thoroughly the work is defined as

$$F_{CS0}(h) = \frac{1}{2}(1 + \cos(\tilde{\varphi})), \quad \tilde{\varphi} = \frac{2\pi mh}{l_s + 1} + \varphi, \quad (3.14)$$

$\varphi$  is a phase of a periodic function  $\cos(\tilde{\varphi})$ .

The periodic function in the definition of the regular function could be substituted by different

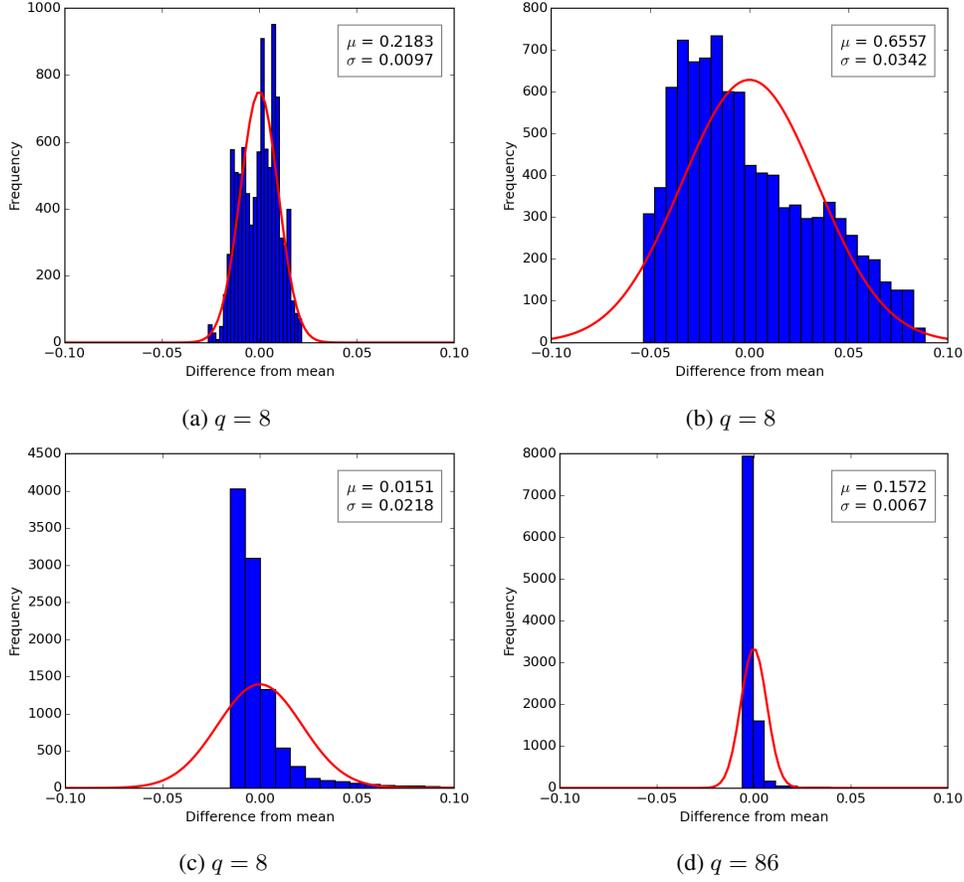


Fig. 3.9: Values of momenta for 2-endpoints open string ((a)–(b)) and D2-branes ((c)–(d)) without (left column) and with (right column) regularization function. Histograms show the difference of values from the mean of normal distribution function  $(\mu, \sigma)$ .

types of mathematical functions

$$F_{CS1}(h) = \frac{1}{2}(1 + \sin(\tilde{\varphi})), \quad (3.15)$$

$$F_{CS2}(h) = \frac{1}{2}(1 - \cos(\tilde{\varphi})), \quad (3.16)$$

$$F_{CS3}(h) = \frac{1}{2}(1 + \tanh(\tilde{\varphi})), \quad (3.17)$$

$$F_{CS4}(h) = \frac{1}{2}(\sin(q * \tilde{\varphi}) + \cos(\tilde{\varphi})). \quad (3.18)$$

The regular function enters all types of the predictors and modify their values.

More complex regular functions, e. g., used for partially compactified D2-brane can be writ-

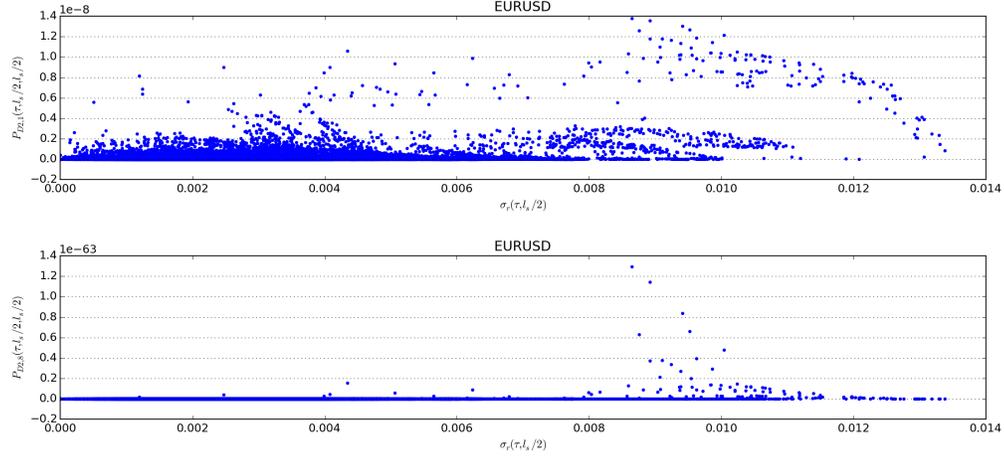


Fig. 3.10: Relationship of return volatility  $\sigma_r(l_s/2)$  and the string amplitudes for 2-endpoints string (first row) and D2-brane (second row). It shows the separating effect for  $q = 8$ . Calculated for 1min. EUR/USD ticks at time period 01 – 12/2015 and  $l_s = 1000$ .

ten as

$$F_{CS5}(h) = \frac{1}{2} (\sinh(q \cdot \tilde{\varphi}) + \cosh(\tilde{\varphi})), \quad (3.19)$$

$$F_{CS6}(h) = \frac{1}{2} (\sin(\tilde{\varphi}^2) \cdot \cos(\tilde{\epsilon}^2)), \quad \tilde{\epsilon} = \frac{2\pi m h}{l_s + 1} + \epsilon. \quad (3.20)$$

### 3.4 Experimental analysis of PMBCS

In addition to the theoretical modeling, our purpose was to study the financial forecasting on the real data. The first simulations have been done with the trading algorithm (Appendix B) on the OANDA data with the most realistic trade conditions [29]. We have illustrated the impact of the string parameters and the prediction behavior of our Self-learning model on the net asset value. As a result, the abilities of the self-learning model were shown to find the optimal string parameters for the final opening/closing of trade positions.

The string PMBCS Self-learning model was benchmarked against the basic time series forecasting models and trading strategies. For this purpose we chosen the scalping strategy of taking profits on small price changes (SCALPER) [56], a trend-following momentum indicator Moving Average Convergence Divergence (MACD) based on the exponential moving averages (EMA) [57] and finally the class of the autoregressive integrated moving average models (ARIMA) [35, 58], including ARIMA(0,0,0)+ $c$  – the mean constant model, ARIMA(0,1,0) – the random walk model and ARIMA(0,1,0)+ $c$  – the random walk with drift model (for different constants  $c$ ). All mentioned models were implemented into the trade system as the corresponding algorithms. The results of the simulations are presented in Tab. 3.5 and Fig. 3.11. As one can see the predicted NAV values after three months time period are close to zero profit for the most of

Model	Mean $\mu$	Sigma $\sigma$	NAV [%]
String PMBCS	741	1482	4.33
SCALPER	-1396	1340	-3.99
MACD	-383	1020	1.35
ARIMA(0,0,0) + $c_1$	95	249	0.40
ARIMA(0,1,0)	-247	805	-0.67
ARIMA(0,1,0) + $c_2$	-1699	1208	-3.04
ARIMA(0,1,0) + $c_3$	44	679	0.89

Tab. 3.5: The comparison of the results for the net asset values (NAV) of our model and basic time series forecasting models (ARIMA) and trading strategies (SCALPER, MACD) on the EUR/USD currency rate for the time period 2010/07/15 – 2010/10/15. Mean  $\mu$  is the average of the values (reference point  $10^5$ ),  $\sigma$  is the standard deviation and NAV is the percentage change of the start and end positions for the selected time period (see also Fig. 3.11).

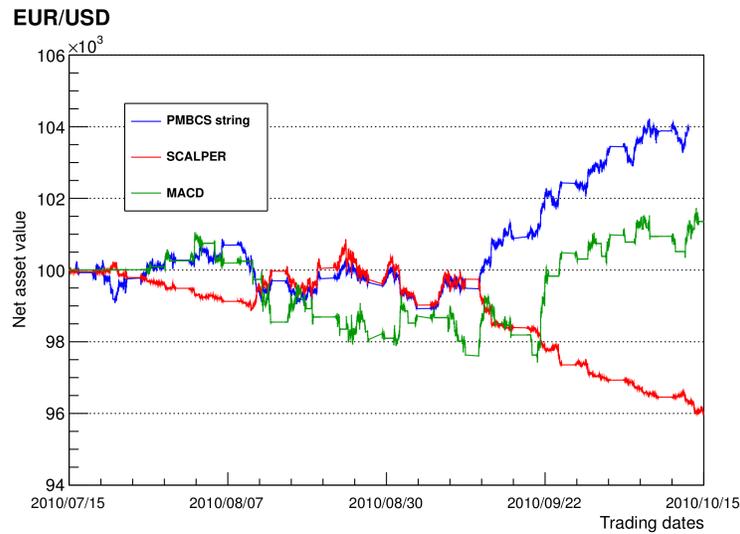
the cases, especially for ARIMA models. Also the mean  $\mu$  of NAV values oscillates around the starting point throughout the whole period. It is not surprising, one expects such behaviour for the random walk models where predicted values are equal to the last observed values. The reliability of the SCALPER method for examined period is also very small as it tends to the negative profit very rapidly. However, the scalping strategy is primary intended to take small profits for short time scales.

The stability of the algorithm on the transaction costs for long trade periods was demonstrated and the next logical step was to improve the Self-learning algorithm for the data evaluation, as the received results were encouraging.

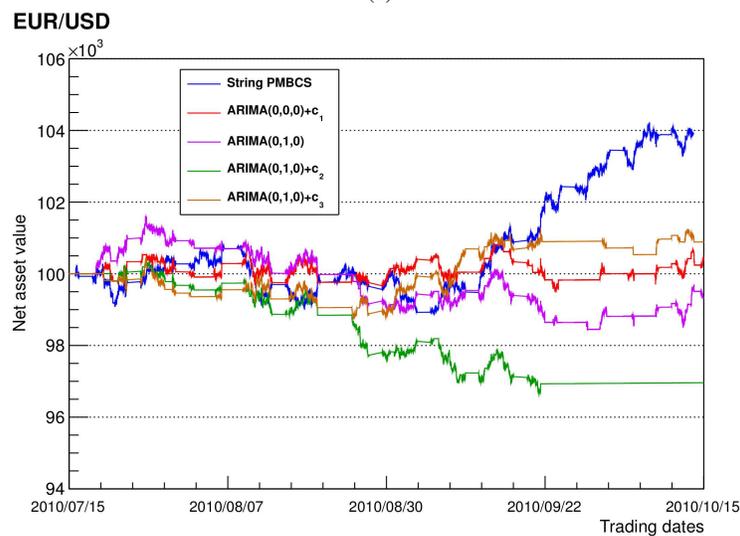
It was done with the new version of the trading algorithm. For the purpose to demonstrate the impact of different types of string maps on the net asset value (NAV) we performed numerical simulations with open strings with one and two endpoints, D2-branes and ARMA(p,q) type forecasting models on trade online system with build-in derived algorithms. The plot in Fig. 3.12 presents the results of the simulations for EUR/USD currency pair. In the simulations we have tried to keep all parameters the same as possible, the impact of string length  $l_s$  was tested on final result, OS1ep and OS2ep models have the same regularization function with  $q = 8$ , D2-brane model is not regularized. The study revealed the incapability of ARMA models to keep even zero profit. On the contrary, the results of the string models revealed improvement of NAV with the transition from 1-endpoint to to 2-endpoints open string and D2-branes. Moreover, the higher efficiency for the string models may be achieved by longer string  $l_s$  lengths.

To see the differences and perspectives of the proposed model from the Section 3.3, we have performed the real, i. e., not theoretical runs of demo sessions on the Interactive Brokers (IB) and LMAX Exchange (LMAX) market accounts [59, 60], which were done through the Librade online trade system [61]. The chosen currency pairs EUR/USD, CHF/JPY, AUD/CAD, AUD/JPY were simulated with new algorithm version and thereafter compared with the results from demo sessions (see Tab. B.2 and B.3).

Fig. 3.13 shows NAV plots for this real demo trading results and the simulation results (the



(a)



(b)

Fig. 3.11: The net asset value of the model on the EUR/USD currency rate for a selected time period compared with the values of basic time series forecasting models (ARIMA) and trading strategies (SCALPER, MACD) (see also Tab. 3.5).

NAV scales differ as they were initialized with different trade volumes). One can observe that all demo results for currency pairs follow the main trend of simulations for chosen time period, i. e., nearly two months. The best coincidence is clearly visible for currency pair EUR/USD

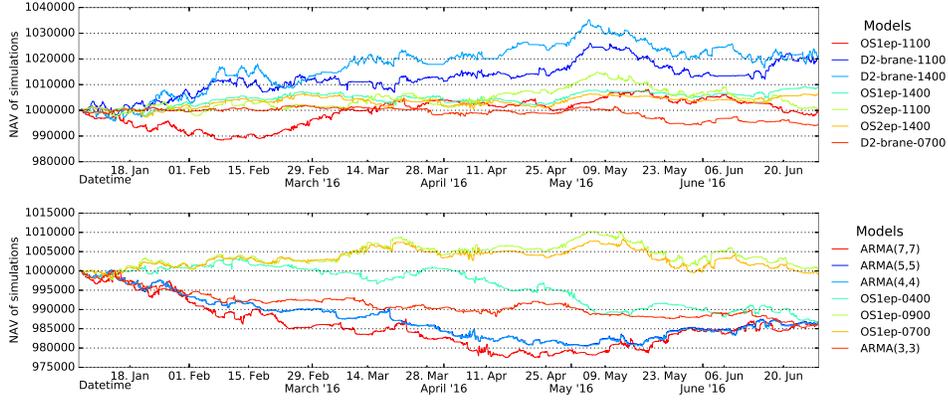


Fig. 3.12: Net asset value plots for model simulations on the EUR/USD currency rate for a time period 01 – 06/2016 as the dependence on string mapping (OS1ep – one string with one endpoint, OS2ep – open string with two endpoints, D2-branes model) compared to time series forecasting models of ARMA type. Number in string model denotes the value of string length  $l_s$ .

(IB-test-12 and LMAX-test-16 accounts), which was in the center of our interest. Also we found nice candidate for currency pair AUD-CAD as one can see in the case of LMAX-test-14 account.

### 3.5 Regge slope parameter

In this section we closely look at another quantity which has origin in the string theory, so called Regge slope parameter  $\alpha'$ . The connection of the slope parameter and the angular momentum makes it suitable for the investigation of the stability of currency rates as shown below.

For rotating open string, the parameter  $\alpha'$  or inverse of the string tension, is the constant that relates the angular momentum of the string  $J$  to the square of its energy  $E$

$$\alpha' = \frac{J}{\hbar E^2}. \quad (3.21)$$

In our analogy we introduce the slope parameter in terms of the angular momentum  $M_q^{\text{ab}}(\tau)$ .

For the time series of open-high-low-close (OHLC) values of currency rates  $p(\tau)$  one can construct separated ask and bid strings, in our case we use the open string with 2-endpoints and string length  $l_s$ , introduced via the nonlinear map in Eq. (2.11). Then the momentum distance function  $d_q^{\text{ab}}(\tau)$  between the ask string  $P_{q,\text{ask}}^{(2)}(\tau, h) \equiv P_q^{(2)}(\tau, h)|_{p \rightarrow p_{\text{ask}}}$  and the bid string  $P_{q,\text{bid}}^{(2)}(\tau, h) \equiv P_q^{(2)}(\tau, h)|_{p \rightarrow p_{\text{bid}}}$  has the form as in Eq. (2.44). In case of rotating open string, the nonvanishing component of angular momentum is  $M_{12}$ , and its magnitude is denoted by  $J = |M_{12}|$  (more in [3])

$$M_{12} = \int_0^{\sigma_1} (X_1 P_2^\tau - X_2 P_1^\tau) d\sigma \quad (3.22)$$

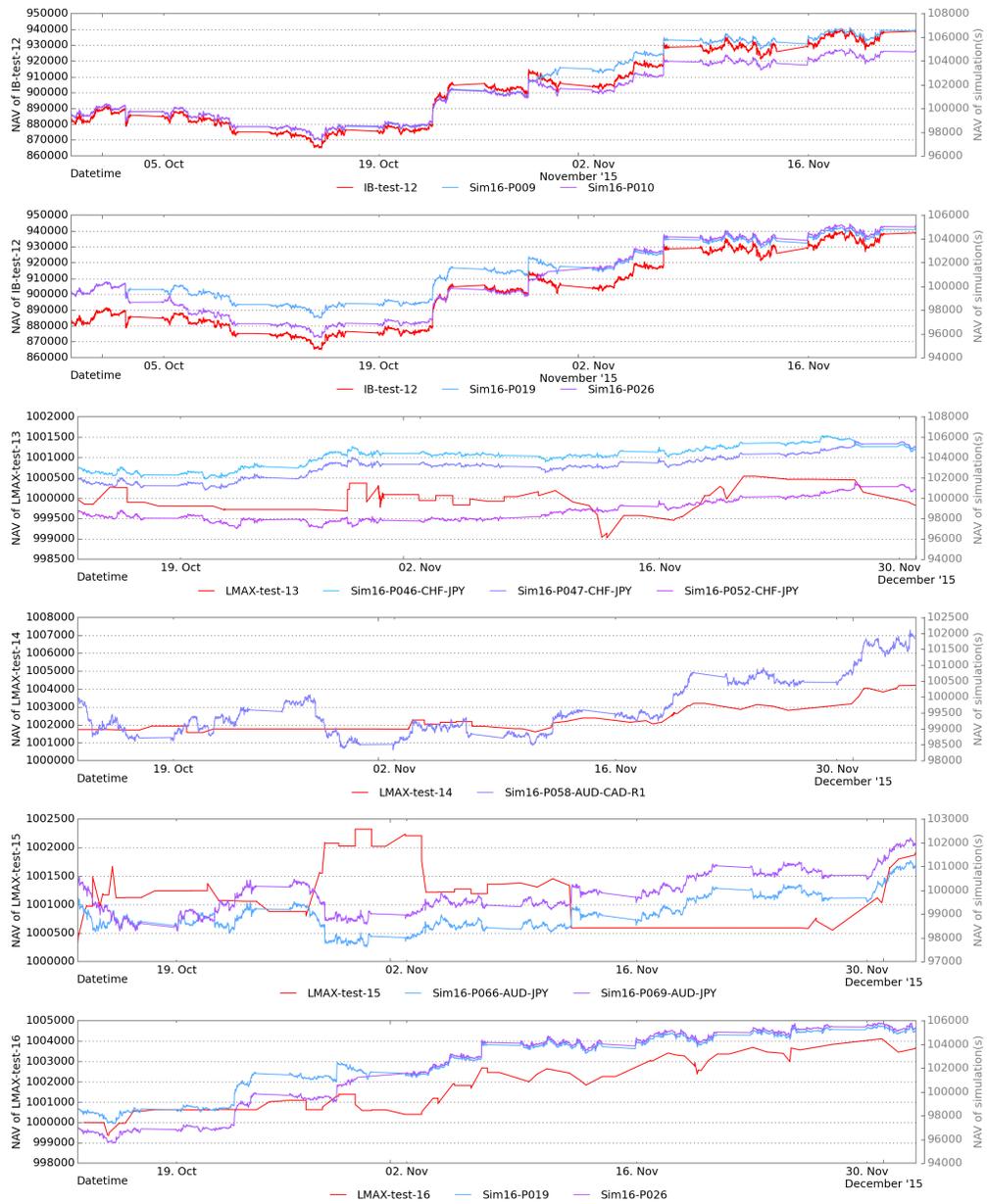


Fig. 3.13: Net asset value plots for opened demo sessions and the results of simulations performed with StringAlgo v.16 on real demo trading IB and LMAX accounts as presented in Tab. B.3.

String length	Volatility window [in min.]			
	5	10	40	60
10	0.5175	0.6072	0.3759	0.3531
20	0.3949	0.4574	0.4307	0.3865
30	0.4726	0.5022	0.5398	0.4935
40	0.4460	0.5098	0.6579	0.5843
50	0.4384	0.4184	0.5230	0.5240

Tab. 3.6: The Pearson product-moment correlation coefficients between the angular momentum (dependent on a string length  $l_s$ ) and the historical volatility (dependent on a time window) calculated for close ask 1 min. ticks of EUR/USD exchange rate on December 4th, 2015.

Currency pair	$\langle P_1^{(2)}(l_s/2) \rangle$ [ $\times 10^{-7}$ ]	$\alpha'_1$ [ $\times 10^{-13}(2\pi)^{-1}$ ]	$T_0$ [ $\times 10^{12}$ ]
AUD/CAD	3.6841	8.9764	1.1140
EUR/USD	0.3539	2.1890	4.5684
GBP/USD	-5.0099	5.4474	1.8357
USD/CAD	8.6794	12.0247	0.8316
USD/CHF	10.6082	10.6185	0.9418
USD/JPY	28.2180	6.9397	1.4410

Tab. 3.7: Average values of string amplitude of 2-endpoints string  $P_1^{(2)}(l_s/2)$ , slope parameter  $\alpha'_1$  and tension  $T_0$  for main currency pairs. One month (02/2016) tick data with 1 min. resolution, string length  $l_s = 1000$ .

for space and conjugate components  $P_i, X_i, i = 1, 2$  and  $\sigma_1 = E/T_0$ . It leads to the relation connecting the slope parameter  $\alpha'$  and  $T_0$  as the string tension  $T_0$

$$T_0 = \frac{1}{2\pi \alpha' \hbar c}. \quad (3.23)$$

In our notation, the angular momentum can be written as (2.45). The slope parameter has final form

$$\alpha'_q = \frac{\langle |M_q^{ab}(\tau)| \rangle}{2\pi l_s^2}. \quad (3.24)$$

Table 3.7 presents the typical values of slope parameter  $\alpha'_q$  together with the mean of string amplitude  $P_1^{(2)}(l_s/2)$  for 2-endpoints string mapping, the string tension  $T_0$  is estimated with the help of Eq. (3.23) ( $\hbar c = 1$ ). It is obvious that each currency pair operates with the own characteristic inter-string values. From the theory of D-branes is known generalized formula for the tension of  $D_p$ -brane [62]

$$T_{D_p} = \frac{1}{g_s (2\pi)^p l_s^{p+1}}, \quad (3.25)$$

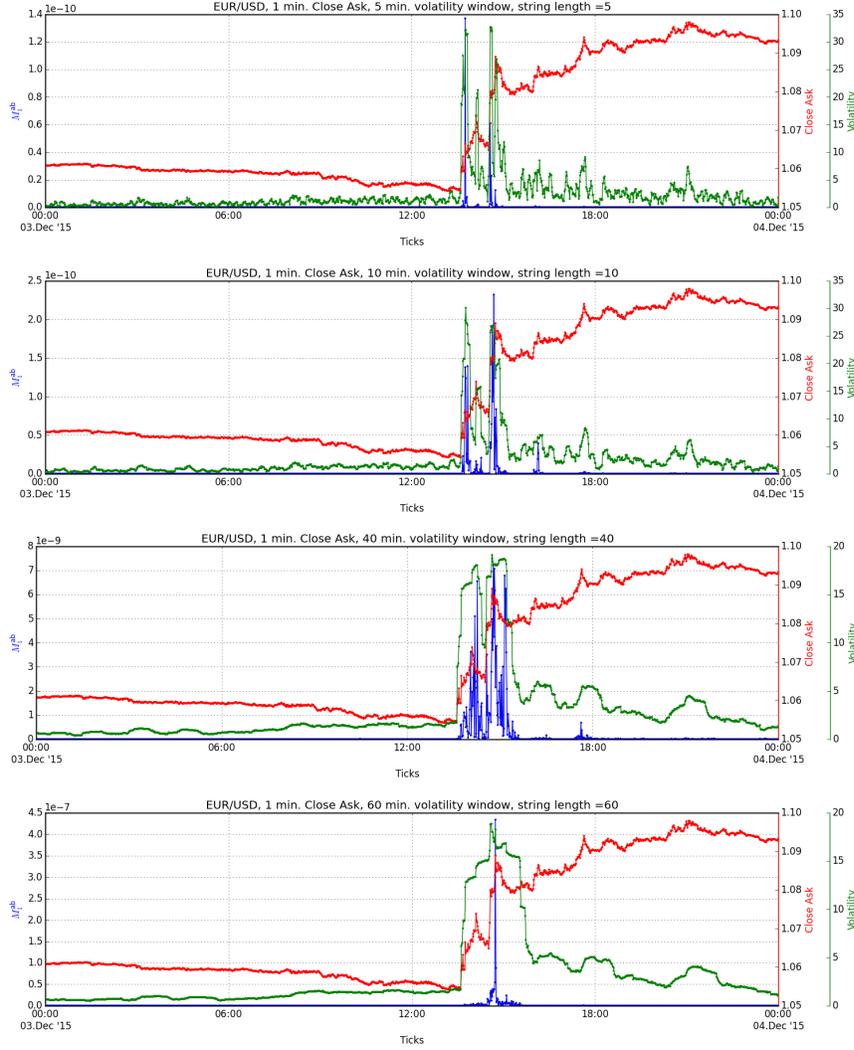


Fig. 3.14: Plot shows the close ask value of EUR/USD exchange rate (red) for 1 min. ticks on December 4th, 2015. The historical volatility in 5, 10, 40, 60 min. windows (green) is compared with the angular momentum  $M_q^{ab}(\tau)$  (blue) for  $q = 1$  and  $l_s = 10$  min.

$l_s$  is the familiar string length and  $g_s$  is the string coupling, which can be used for higher dimensions not considered in this work.

In such a way the angular momentum  $M_q^{ab}(\tau)$  can be applied as complementary tool to analyze the stability of currency rates except the historical volatility, as they are both compared in Fig. 3.14. For short string lengths the angular momentum indicates the same sharp changes in

exchange rate. The correlation between measures is highest for equal values of a string length and time window parameters, see Table 3.6. Although there exists a certain relation between those measures, for instance, the similar sensitivity in time, the memory effect of angular momentum seems to be lower. Therefore, it may provide a helpful indication of market changes or to serve as a trade brake in algorithms.

In connection with a slope parameter  $\alpha'$  and a string tension  $T_0$  we have compared their values for a set of currency pairs in Table 3.7 (we have chosen the main six trading pairs). One can deduce that an increase of slope parameter values (or decrease of tension) indicates the changes on a market and a volatility is increasing. Although the fall in prices can last for a short time, the trading algorithms must immediately respond on the situation to avoid large losses.

#### 4 Kolmogorov space and the concept of algebraic topology and their relation to time series data

In the previous paper [29] we had mentioned the idea of the proper algebraic and geometric construction of the space of time series, possible under a Kolmogorov space concept with consistent separation axiom, which could help with the analyses of nonstationary time series models, the volatility clustering phenomena in financial time series data or the separation of hidden Markov transition probability state in quantum entanglement state. A space of time series data was represented by the topological space with fixed point property. In this section we present a concept which could be realized as a topological space with a few hidden states in extradimensions of loop space of time series data.

As we know, no precise mathematical definition exists for time series and financial time series. We only know that time series are the observations ordered in a time (or space). For time series data, we can not use a set to define time series data directly. Since the data can have the same value in a set notation, the same value cannot be separated using the  $T_0$ -separation under discrete topology. Therefore, a pointed space of time series data is not a Kolmogorov space, one needs to define extradimensions [24] in time series data as loop space [63] of path lifting for the separation of data under  $T_0$ -separation axiom.

**Definition 1** (Kolmogorov space). *A Kolmogorov space  $X$  is a topological space fulfilling the  $T_0$ -separation axiom such that for any two points  $x, y \in X$ , there exists an open set  $U$  such that  $x \in U$  and  $y \notin U$  or  $y \in U$  and  $x \notin U$ .*

A Kolmogorov space [64] is a topological space fulfilling the  $T_0$ -separation axiom, in other words, it is a topological space [65] in which every pair of distinct points is topologically distinguishable. A fixed point space [66] need not be Hausdorff space necessarily, but it has to satisfy weaker  $T_0$ -separation axiom, it means that all fixed point spaces are Kolmogorov ones [67]. For that reason one must to verify a  $T_0$ -separation axiom property for a space of time series data, in order to declare it to be Kolmogorov space.

The most recent work in financial data analysis [68] is based on a definition of space of time series  $X = \{x_t \in \mathbb{R}, t \in \mathbb{N}\}$ , without separation axiom in the contribution. Let  $X$  be a set of ordered points of time series values. We can consider a set of time series data as an object in categories of SET with objects sets and morphisms – the injection functions between sets. A functor is a transformation of object of time series from categories of SET into another categories of TOP and GROUP. If we consider a sequence of data just a set, we can define a discrete topology on a set of data (see Eq. (2.1–2.3)). For a time series data, we cannot use a set to define a time series data directly, since the data can have the same value in a set's notation. E. g., if we have a time series data with following six numbers of sample data

$$A = \{1, 2, 3, 3, 3, 4\}, \quad (4.1)$$

the same values can not be separated. If we use set theory to induce a pointed set topology, we will fail to define an open set of data because,

$$A = \{1, 2, 3, 3, 3, 4\} = \{1, 2, 3, 4\}. \quad (4.2)$$

**Definition 2.** *Let  $A \neq \emptyset$  be a set. Let  $\tau = P(A)$  be the power set of  $A$ . Then  $\tau$  is called the discrete topology on  $A$  and  $(A, \tau) = (A, P(A))$  the discrete space on  $A$ , or just a discrete space.*

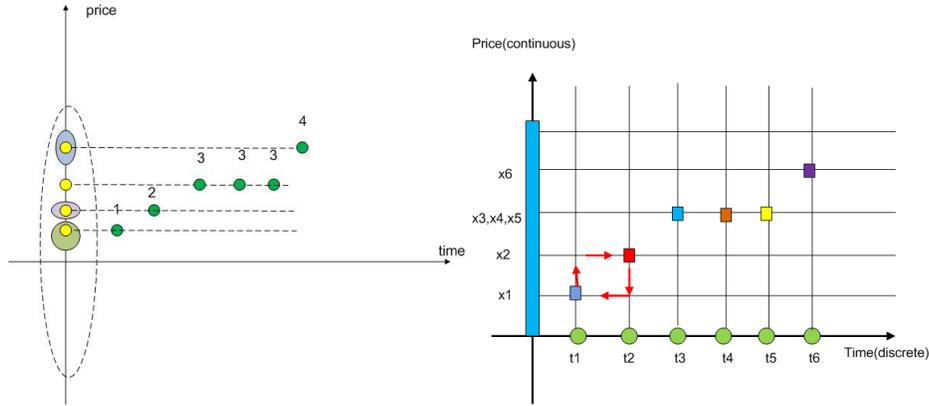


Fig. 4.1: On the left the discrete topology of a time series data with the open set visualized in the projective plane at  $y$ -axis. We can see that the projection of all three points with the value 3 is just one point at the same position, we cannot use an open set to separate these three points. The plot on the right shows that the point of time series data need to be embedded in the intersection point of perpendicular line.

**Definition 3** (Finite discrete topology). *If  $A$  is finite,  $\tau = P(A)$  is a finite discrete topology, and  $(A, \tau) = (A, P(A))$  is a finite discrete space.*

Let a sequence of data be

$$x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow x_4 \rightarrow x_5 \rightarrow x_6 \tag{4.3}$$

with the values

$$x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 3, x_5 = 3, x_6 = 4. \tag{4.4}$$

When we use discrete topology, we will get an open set with  $2^{n(A)} = 2^4 = 16$  open subsets. A sequence of similar values for  $x_3 = 3, x_4 = 3, x_5 = 3$  will not be separated (Fig. 4.1), an open set can not be used to separate all time series data. Therefore with discrete topology a space  $(A, \tau) = (A, P(A))$  is a finite discrete space, it is not a Kolmogorov space with  $T_0$ -separation axiom. Let us assume that a time series data is embedded in non-Euclidean plane with extradimension (Fig. 4.2). It allows us to embed a loop structure between all data of time series connected to each other as path components. By this redefinition of a financial time series data, one is allowed to use an equivalent class of loop space (fundamental group of time series) to separate the sequence of similar values for  $x_3, x_4, x_5$  in Eq. (4.4) by open set in path component.

For time series model a time is discrete. We use algebraic topology tool to change topology of discrete space  $\mathbb{N} \subset \mathbb{Z}$  to  $\mathbb{R}$  by using quotient topology of covering space  $\mathbb{R}/\mathbb{Z}$  with some fibre space. We glue a pointed space of each fibre to form a pointed space of time series data by using coproduct in topology. If we treat time as point in real axis, we will induce an open set. For every  $i = 1, \dots, n, t_i \in \mathbb{R}$  we will have an open set

$$(-\infty, t_1), (t_1, t_2), (t_2, t_3), \dots, (t_{n-1}, t_n), (t_n, \infty) \tag{4.5}$$

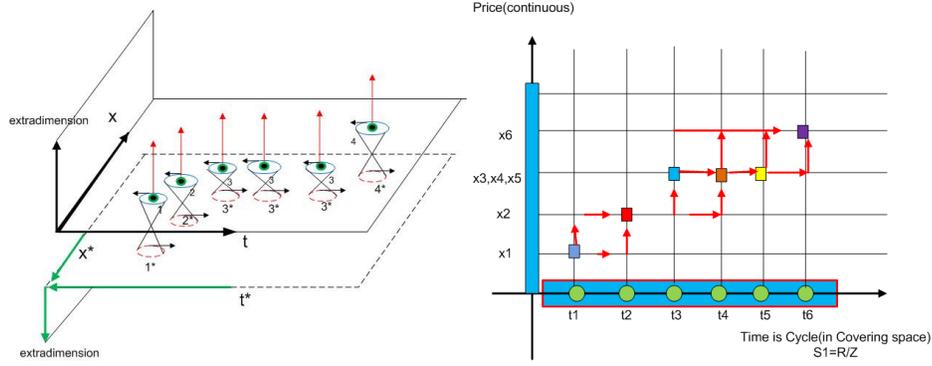


Fig. 4.2: On the left the visualisation of the extradimension for time series. There exists an induced field of cross product between price and time in covariant and contravariant tensor field in hidden plane in extradimension of Euclidean plane of time series data. The right figure shows CW complex of torus. The cell complex is in orientation state, if it is in nonorientation state the glue process will produce Möbius strip of time series data.

We call this open set a topology for time series data of time path. Let  $t \in A \subset \mathbb{R}$  be an index set of time series data. Let a pointed space of time series data be  $X_t = \{x_t\}$ . We define a family of pointed space of time series data by  $\{X_t | t \in A\}$ . The family of space underlying measurement data will induce a topological sum of space of time series data  $\coprod_{t \in A} X_t$  where  $\cup \{X_t \times \{t\} | t \in A\}$ .

We let a closed embedded map

$$i_\beta : X_\beta \rightarrow \coprod_{t \in A} X_t, x \mapsto (x, \beta) \quad (4.6)$$

for every  $\beta \in A$  we have  $i_\beta(X_\beta) \cap i_\alpha(X_\alpha) = \emptyset$ , if  $\alpha \neq \beta$ . One of the major problems in financial time series of stock price is how can we incorporate a behavior of trader into financial time series data directly. We solve this problem by using CW decomposition [65] of financial time series. We can attach buying and selling operation using a coproduct of topology for each cell decomposition of Euclidean space of time series data (see Fig. 4.3). This process induces a hidden dimension in Kolmogorov space of time series data.

All data which cannot be separated in Euclidean plane (Fig. 4.4) now can be separated by using path lifting to open set in covering space of physiology of time series. A ground space of time series data  $A$  can be realized as topological space by using disjoint union from

$$A = \{1\} \cup \{2\} \cup \{3\} \cup \{3\} \cup \{3\} \cup \{4\} \quad (4.7)$$

to a space of time series  $X$ ,

$$X = \{1\} \coprod \{2\} \coprod \{3\} \coprod \{3\} \coprod \{3\} \coprod \{4\}. \quad (4.8)$$

Let  $t \in I = [0, 1]$  be a time interval. We define a equivalent class of path  $\alpha : I \rightarrow X$  by

$$[I, X] = [I, \{1\}] \coprod [I, \{2\}] \coprod [I, \{3\}] \coprod [I, \{3\}] \coprod [I, \{3\}] \coprod [I, \{4\}]. \quad (4.9)$$

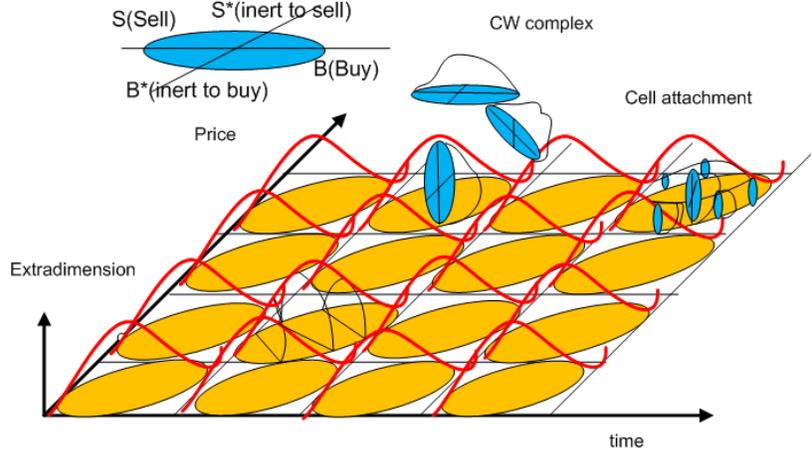


Fig. 4.3: Non-euclidean plane of time series. The demonstration of a mirror symmetry of dual price and dual time scale axis between price and time for covariant and contravariant tensor field in hidden plane in extradimensions of Euclidean plane of time series data.

Since  $S^1$  is homotopy equivalent to  $I/\partial I$ , we induce a fundamental group of space of time series data,

$$[S^1, X] = [S^1, \{1\}] \amalg [S^1, \{2\}] \amalg [S^1, \{3\}] \amalg [S^1, \{3\}] \amalg [S^1, \{3\}] \amalg [S^1, \{4\}]. \quad (4.10)$$

In next section we introduce a precise definition of time series data in loop space  $\Omega(X, x_0)$ . We use equivalent class of loop space in covering space of predefined 4 basis  $s_i$ ,  $i = 1, 2, 3, 4$  in  $S^1 = \{x \in \mathbb{C}, |x| = 1\}$  and another perpendicular  $S^{1*}$  for location of time series data (Fig. 4.5). These equivalent classes induce a group structure with a symmetry of time series equivalent to a orbital in quantum state. This is a spinor field of time series data in which one can classify a financial time series data. This new construction used to explain a precise definition of financial time series allows one to search for a new concept of a spinor field of time series data and a mirror symmetry in space of financial time series. In the prove that a space of time series data is a Kolmogorov space we use projective geometry of quaternionic field instead of a discrete topology.

#### 4.1 Loop space of time series

##### Morphology of financial time series

For a given time series  $x_t \in X$ , we induce two spaces of times series with the empirical mode decomposition method (EMD) [69], a complex time scale in 4-dimensional space and a physiology of time series in 4-dimensional vector space  $s_j \in V \simeq \mathbb{H}$ .

An end point of time series data  $\text{End}(x_t)$  is defined as

$$\text{End}(x_t) = \sum_{i=1}^4 \lambda_i g_{ij} s_j, \quad (4.11)$$

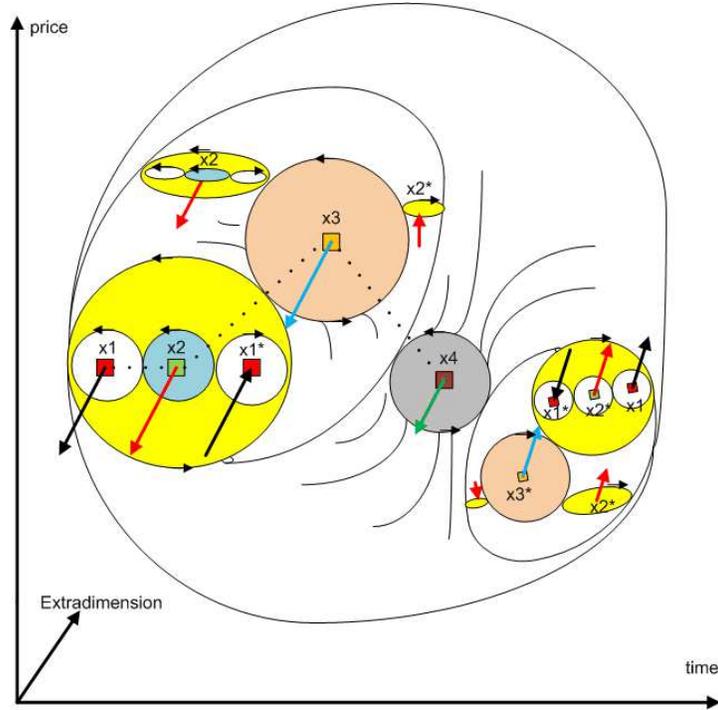


Fig. 4.4: The loop space of time series in which one can separate three same values “3” with a loop. One has  $x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 3, x_5 = 3, x_6 = 4$  in 3-dimensional blend complex surface model of spinor field of time series data in Kolmogorov space. The thick dark line connecting  $x_3$  to  $x_4$  to  $x_5$  is in a perspective view of 3-dimensional with extradimension add to Euclidean plane by induced equipotential spinor fields of times data of 3, 3, 3. This fields are propulsion to each other in model of loop space modelling and induce a straight line with equal slope blend to the direction of extradimension (projective of this line is still straight line in Euclidean plane). Therefore values 3, 3, 3 are completely separated by using  $T_0$ -separation axiom of Kolmogorov space of time series data.

with  $g_{ij}$  – a Jacobian of transformation of the coordinate system and  $\lambda_i \in \{0, 1\}$ . We have

$$\text{End}(x_t(x * \otimes t*)) = \lambda_1 g_{11} \begin{bmatrix} s_1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 g_{22} \begin{bmatrix} 0 \\ s_2 \\ 0 \\ 0 \end{bmatrix} + \dots + \lambda_4 g_{44} \begin{bmatrix} 0 \\ 0 \\ 0 \\ s_4 \end{bmatrix}. \quad (4.12)$$

The hidden direction is coming from a state of the end point for a skeleton of time series data so called (ITD – IMF)chain<sub>1</sub>. Let four hidden directions of a local state of physiology of end

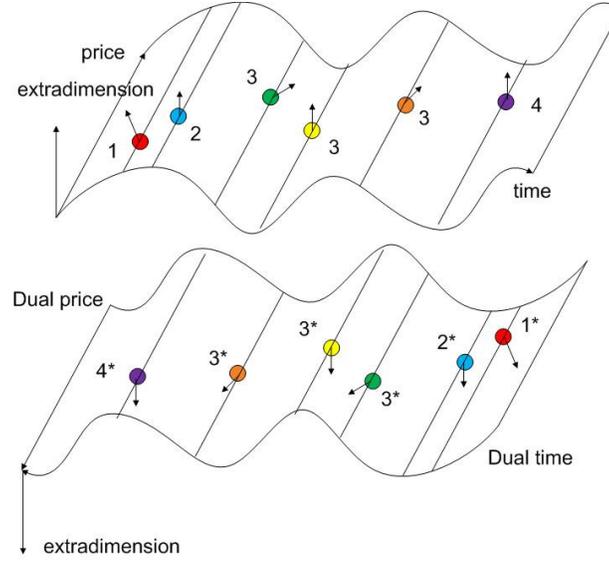


Fig. 4.5: A schematic view of the loop space of time series data with the covering space.

point of time series be

$$\begin{aligned}
 s_1(x_t) &= \text{mono}_{\text{up}}(x_t), \\
 s_2(x_t) &= \max x_t, \\
 s_3(x_t) &= \text{mono}_{\text{down}}(x_t), \\
 s_4(x_t) &= \min x_t.
 \end{aligned} \tag{4.13}$$

where  $\text{mono}_{\text{up}}(x_t)$  is a monotonic function up of time series data of the end point of time series data. It is the point between minimum point of time series data and maximum point of time series data. Sometimes this point does not exist.  $\text{mono}_{\text{down}}(x_t)$  is defined by a monotonic function down of time series data of the end point of time series data. It is the point between maximum point of time series data to a minimum point of time series data, sometimes this point also does not exist.

**Definition 4** (Cyclic coordinate of time scale). *Let  $T_1$  be a location of monotone function up, measured from a distance between  $s_4$  and  $s_1$  of time series data of  $(\text{ITD} - \text{IMF})\text{chain}_1$ .  $T_2$  be a location of maximum point, measured from distance between  $s_4$  to  $s_2$  of time series data of  $(\text{ITD} - \text{IMF})\text{chain}_1$ .  $T_3$  is a location of monotone function down, measured from distance between  $s_4$  to  $s_3$  of time series data of  $(\text{ITD} - \text{IMF})\text{chain}_1$ .  $T_4$  is a location of minimum point, measured from distance between  $s_4$  to next cycle of  $s_4$  of time series data of  $(\text{ITD} - \text{IMF})\text{chain}_1$ . If a restart of every cycle starts from zeros every cyclic time coordinate of  $t = (T_1, T_2, T_3, T_4)$  will be a circle of time scale in loop space.*

The details of EMD algorithm, the definition of  $(\text{ITD} - \text{IMF})\text{chain}_1$ , the skeleton of time

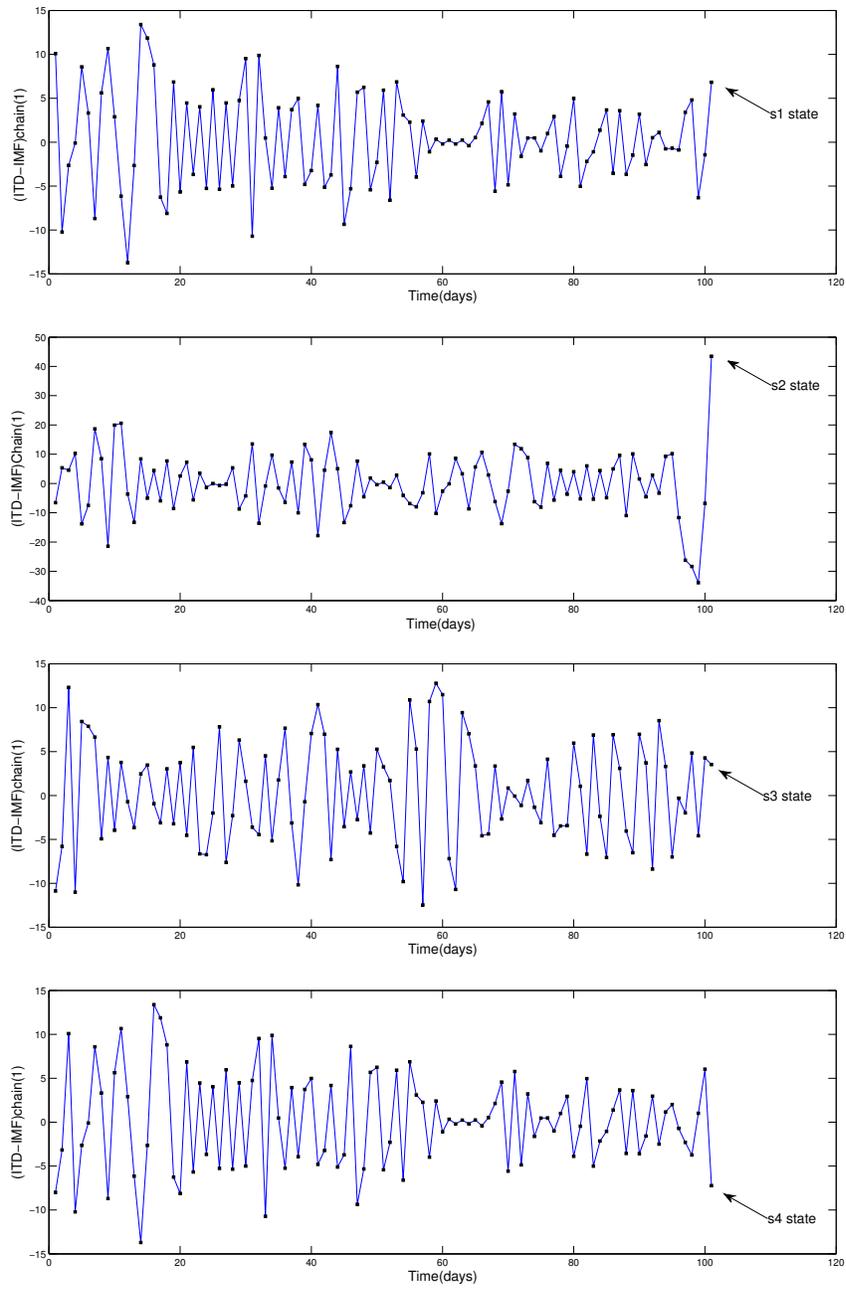


Fig. 4.6: The example of the shape of end point for  $(ITD - IMF)chain_1$  for  $s_1, s_2, s_3, s_4$  states of SET index within 100 days.

series data and the empirical work on financial time series data of cyclic time coordinate are shown in Fig. 4.6 and Appendix D.

### Covering space of time series data

Let

$$x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow \cdots \rightarrow x_n \quad (4.14)$$

be a time series data with underlying trivial topological based space  $X$  where

$$X = \{x_1\} \amalg \{x_2\} \amalg \{x_3\} \amalg \cdots \amalg \{x_n\}. \quad (4.15)$$

We define a tangent space of time series data by lifting path of covering space. We use a notation

$$\amalg_{i=1}^n T_{x_i} X = T_{x_1} X \amalg T_{x_2} X \amalg T_{x_3} X \amalg \cdots \amalg T_{x_n} X. \quad (4.16)$$

for a tangent space of time series data or covering space of time series data. The element of tangent space of time series data is defined by 4 states of morphology of time series data denoted by  $s_i, i = 1, 2, 3, 4$ . We have  $dx_i \in T_{x_i} X$  if  $x_i \in X$ ,

$$dx_i = \sum_{j=1}^4 p_i \frac{\partial x_i}{\partial s_j} ds_j. \quad (4.17)$$

where  $p_i$  is a probability to find a hidden state  $s_i$ . Giving a tangent space we also induce a dual tangent space of time series data  $T_x^* X$  and also differential form of time series data using wedge product  $\wedge T_x^* X$ .

### Tensor field of time series data

In definition of time series data we assume two independent spaces of two measurement systems, a vector space of price  $X$  and a space of time  $t \in S^1$ , respectively a complex unit sphere  $S^1$ . The space of time series data obtained by merging  $X$  and  $S^1$  is represented by the tensor product  $X \otimes S^1$  as a state space of time series data in tensor field. The states that can be broken into the tensor product of states from the constituent subsystems are called separable states, whereas states that are unbreakable are called entangled states of time series data. Let  $x \in X, t \in S^1$  we have a dual price  $x_* = \vec{t} \wedge \vec{x}, x^* = \vec{x} \wedge \vec{t}$ . We can define dual time scale [70] of time series data by

$$t^* = x_* \otimes x^*. \quad (4.18)$$

### Homotopy path of time series data

Let

$$x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow \cdots \rightarrow x_n, \quad (4.19)$$

be a sequence of points of financial time series in categories of SET with preorder relation of time ordering as morphism.

Let  $\pi_1 : \text{TOP} \rightarrow \text{GROUP}$  be a functor of fundamental group of chosen based point  $x_0$  of object in categories of SET with  $(X, x_0)$  a topological space  $X$  of underlying space of financial time series as object in TOP

$$\pi_1(X, x_1) \rightarrow \pi_1(X, x_2) \rightarrow \pi_1(X, x_3) \rightarrow \cdots \rightarrow \pi_1(X, x_n), \quad (4.20)$$

there is a sequence of equivalent class of loop in loop space of time series data

$$\Omega(X, x_1) \rightarrow \Omega(X, x_2) \rightarrow \Omega(X, x_3) \rightarrow \cdots \rightarrow \Omega(X, x_n). \quad (4.21)$$

There exists a one-to-one sequence of discrete time intervals between measurements

$$t_1 \rightarrow t_2 \rightarrow t_3 \rightarrow \cdots \rightarrow t_{n-1} \quad (4.22)$$

with a loop space of location (complex time scale coordinate of time series)

$$\Omega(X, t_1) \rightarrow \Omega(X, t_2) \rightarrow \Omega(X, t_3) \rightarrow \cdots \rightarrow \Omega(X, t_{n-1}). \quad (4.23)$$

Let 4-dimensional space  $\mathbb{H}$  be a space of cyclic time coordinate. Let

$$t(x) = T_1(x) + T_2(x)i + T_3(x)j + T_4(x)k \in \mathbb{H}. \quad (4.24)$$

where  $T_1$  is a time from the origin to the state of monotone up  $s_1$  of time series data of (ITD – IMF)chain<sub>1</sub>.  $T_2$  is a time period from the origin to the state of maximum  $s_2$  of time series data of (ITD – IMF)chain<sub>1</sub>.  $T_3$  is a time from the origin to the state of monotone down  $s_3$  of time series data of (ITD – IMF)chain<sub>1</sub>.  $T_4$  is a time period from the origin to the state of minimum  $s_4$  of time series data of (ITD – IMF)chain<sub>1</sub>.  $\mathbb{H}$  is a quaternion field with 3 complex numbers implying hidden states of time scale with  $i^2 = j^2 = k^2 = -1, ijk = -1$ .

For a given time series data set  $X = \{x_1, x_2, \dots, x_n\}$  we induce a set of location in time series data  $t_X = \{t_1, t_2, \dots, t_n\}$ ,  $|X| = |t_X|$ . Let a functor  $[S^1 \cdot] : \text{SET} \rightarrow h - \text{TOP}$ ,  $X \mapsto [S^1 X] = \pi_1(X)$  where  $S^1 = \{z \in \mathbb{C}, |z| = 1\}$ . An object of homotopy category  $h - \text{TOP}$  is a set of equivalent class of classifying space  $S^1$  over a pointed space of time series data. Let a homotopy path be

$$[\alpha] \in [S^1 X] \quad (4.25)$$

to partition  $X$  into 4 equivalent classes of equivalent location in physiology of time series  $[T_1]$ ,  $[T_2]$ ,  $[T_3]$  and  $[T_4]$ .

Let us consider a correlation problem of inert frame of reference. Most economists use a formula below to calculate a correlation on price not on time scale but on fix time scale  $\text{corr}(x_1, x_2) = \text{Corr}(x_1(t), x_1(t))$  with

$$\text{Corr}(x_1, x_2) = \sum \frac{(x_1 - \mu_1)^2}{\sigma_1} \frac{(x_2 - \mu_2)^2}{\sigma_2} \quad (4.26)$$

Now we use a transformation to interchange coordinates between price and time (transition path) by inversion (projection on time line and price line)  $t' = x(t)$  which induces a bijective map

$$i : t'(x_1, x_2) \rightarrow x(t_1, t_2). \quad (4.27)$$

The Jacobian on this coordinate transformation between space and time is so called Minkowski metric. It is opposite to Euclidean space concept by means the space and time are completely separated. We consider the correlation in waiting time  $t_1, t_2$  of the join return of rotational invariant and translation invariant of 2 returns  $x_1, x_2$

$$\text{Corr}(t_1(x), t_2(x)) = \sum \frac{(t_1(x) - \mu_1)^2}{\sigma_1} \frac{(t_2(x) - \mu_2)^2}{\sigma_2} \quad (4.28)$$

which induce a Jacobian matrix for the transformation between these two ways to calculate a correlation by using price and time coordinate

$$J = \begin{vmatrix} \frac{\partial t_1}{\partial x_1} & \frac{\partial t_1}{\partial x_2} \\ \frac{\partial t_2}{\partial x_1} & \frac{\partial t_2}{\partial x_2} \end{vmatrix} \quad (4.29)$$

where  $x = x(x_1, x_2)$ ,  $dr = J(dt)$ . In Euclidean space-time continuum concept it is impossible to use this kind of transformation. What we can do is to transform both space (price) and time all together because price and time of two stocks in which they are correlated to each other are evolved together and cannot be separated (isometry) so we have Jacobian of transformation as a part of isometry group

$$G\langle x, t \rangle = \langle x, t \rangle, \quad G = J^2 \quad (4.30)$$

so we have

$$d(x', t') = J(d(x, t)) \quad (4.31)$$

with  $J = -1$  to interchange a projection between price and time as inversion point of symmetry breaking. We define this induced Jacobian transformation in hidden coordinate of complex plane by introducing additional hidden coordinate of a projection to imaginary axis of complex plane by

$$J = \begin{vmatrix} x & t \\ x^* & t^* \end{vmatrix} = x \wedge t^* - x^* \wedge t. \quad (4.32)$$

The determinant commutes by using wedge product with hidden coordinate defined by induce cross product of vector of space and time and hidden space and time as double complex plane. The empirical work of correlation between price and time is given by correlation matrix between cyclic time coordinate  $t = (T_1, T_2, T_3, T_4)$  and cyclic state  $x = (s_1, s_2, s_3, s_4)$

$$\text{Corr}(x, t) = \begin{bmatrix} & T_1 & T_2 & T_3 & T_4 \\ s_1 & \text{Corr}(s_1, T_1) & \text{Corr}(s_1, T_2) & \text{Corr}(s_1, T_3) & \text{Corr}(s_1, T_4) \\ s_2 & \text{Corr}(s_2, T_1) & \text{Corr}(s_2, T_2) & \text{Corr}(s_2, T_3) & \text{Corr}(s_2, T_4) \\ s_3 & \text{Corr}(s_3, T_1) & \text{Corr}(s_3, T_2) & \text{Corr}(s_3, T_3) & \text{Corr}(s_3, T_4) \\ s_4 & \text{Corr}(s_4, T_1) & \text{Corr}(s_4, T_2) & \text{Corr}(s_4, T_3) & \text{Corr}(s_4, T_4) \end{bmatrix} \quad (4.33)$$

and is shown in Appendix D. We define each point of  $T_i$ ,  $i = 1, 2, 3, 4$  as disjoint pointed space embedded in complex projective space as cell decomposition of pointed space  $e_0$  of CW

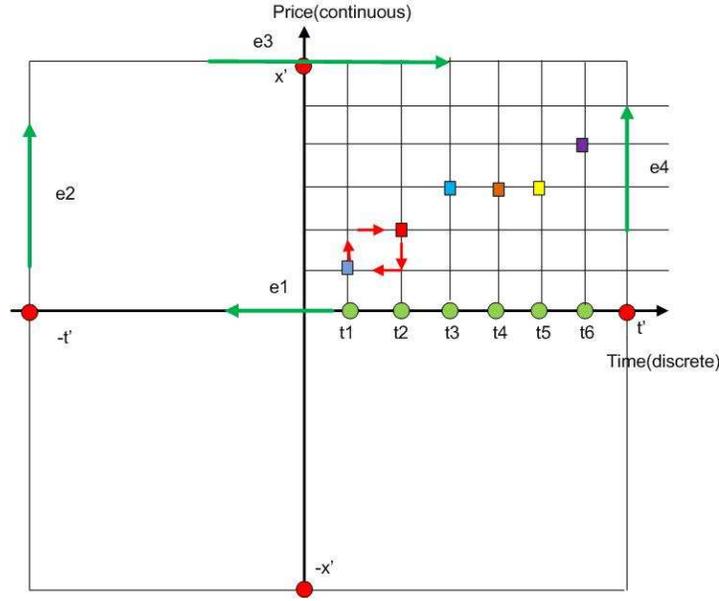


Fig. 4.7: CW complex of boundary of space of time series data. For every value  $x_t$  of time series data in Euclidean plane there exist an upper bound  $x' \in \mathbb{R}$  such that for every  $x_t \in X_t, x_t < x'$ . For index set of time  $t_n$  there exists  $t' > t_n$  for every  $n \in \mathbb{N}$ . Therefore we can define a closed boundary of Euclidean subspace of time series data by rectangle in this figure. After that we define cell decomposition for each corner of boundary in non-orientation state. Then we glue them into Möbius strip to induce spinor field of time series data.

complex decomposition. We can define a point of complex time scale as chosen basepoint for base space of fundamental group. When we collapse a Riemann sphere to base point of time scale of time series we get a cone space of time series as quotient topology. We have  $e^0 = \{T_i\}$ ,  $T_i \in S^0 \subset S^1 \subset S^2 \subset S^3 \subset \dots$ . CW-complex (Fig. 4.7) is required only 2 cells, one a point  $e_0$  and the other  $e^n = S^n - \{e^0\}$  which is homeomorph of ball  $\mathbb{B}^n$ . We define our translated Riemann sphere with relative coordinate of time series by attaching cell  $e^0 = \{T_i\}$  to a center of Riemann sphere as relative frame of time scale of Riemann sphere  $S^2$ .

If we take into account only equivalent class of loop in  $\pi_1(X, x_t)$ , a loop structure of time series data in empirically measured by Hilbert transform of (ITD – IMF)chain<sub>1</sub>. The result of Hilbert transform of time series data is a cycle in complex plane in which it is homotopy equivalent to  $S^1$ . We can explicitly define a physiology of time series data in loop space by using equivalent class of path in  $S^1$ . We separate  $S^1$  into 4 states of physiology of time series data in following way.

Let  $s_1 \in [s_1(x_t)]$  be an equivalent class of loop of time series from  $x_0$  to  $s_1(x_0) = e^{0i} \in S^1$  and  $s_1(x_1) = e^{i\frac{\pi}{2}} \in S^1$ , a covering space of time series with homotopy path (see Fig. 4.8)

$$h : S^1 \times [0, T_1] \rightarrow S^1, \quad h(t, 0) = e^{0i}, \quad h(t, T_1) = e^{i\frac{\pi}{2}}. \quad (4.34)$$

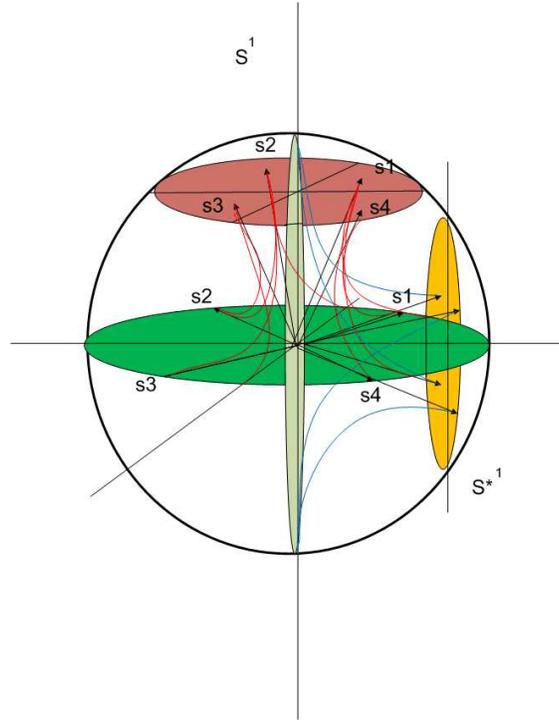


Fig. 4.8: A homotopy path of hyperbolic space of time series. One can see that hyperbolic space of time series data is inside Riemann sphere of time series data. One side of sphere is predictor state of time series data, the perpendicular side is predictant state of time series data. The hyperbolic line connects the predictor and predictant states.

Let  $s_2 \in [s_2(x_t)]$  be an equivalent class of loop of time series from  $x_0$  to  $s_2(x_0) = e^{i\frac{\pi}{2}} \in S^1$  and  $s_2(x_1) = e^{i\pi} \in S^1$ , a covering space of time series with homotopy path

$$h : S^1 \times [T_1, T_2] \rightarrow S^1, \quad h(t, T_1) = e^{i\frac{\pi}{2}}, \quad h(t, T_2) = e^{i\pi}. \quad (4.35)$$

Let  $s_3 \in [s_3(x_t)]$  be an equivalent class of loop of time series from  $x_0$  to  $s_3(x_0) = e^{i\pi} \in S^1$  and  $s_3(x_1) = e^{i\frac{3\pi}{2}} \in S^1$ , a covering space of time series with homotopy path

$$h : S^1 \times [T_2, T_3] \rightarrow S^1, \quad h(t, T_2) = e^{i\pi}, \quad h(t, T_3) = e^{i\frac{3\pi}{2}}. \quad (4.36)$$

Let  $s_4 \in [s_4(x_t)]$  be an equivalent class of loop of time series from  $x_0$  to  $s_4(x_0) = e^{i\frac{3\pi}{2}} \in S^1$  and  $s_4(x_1) = e^{i2\pi} \in S^1$ , a covering space of time series with homotopy path

$$h : S^1 \times [T_3, T_4] \rightarrow S^1, \quad h(t, T_3) = e^{i\frac{3\pi}{2}}, \quad h(t, T_4) = e^{i2\pi}. \quad (4.37)$$

An example of empirical analysis of Hilbert transform of  $s_4$  state of (ITD – IMF)chain<sub>1</sub> of financial data is shown in Fig. 4.9.

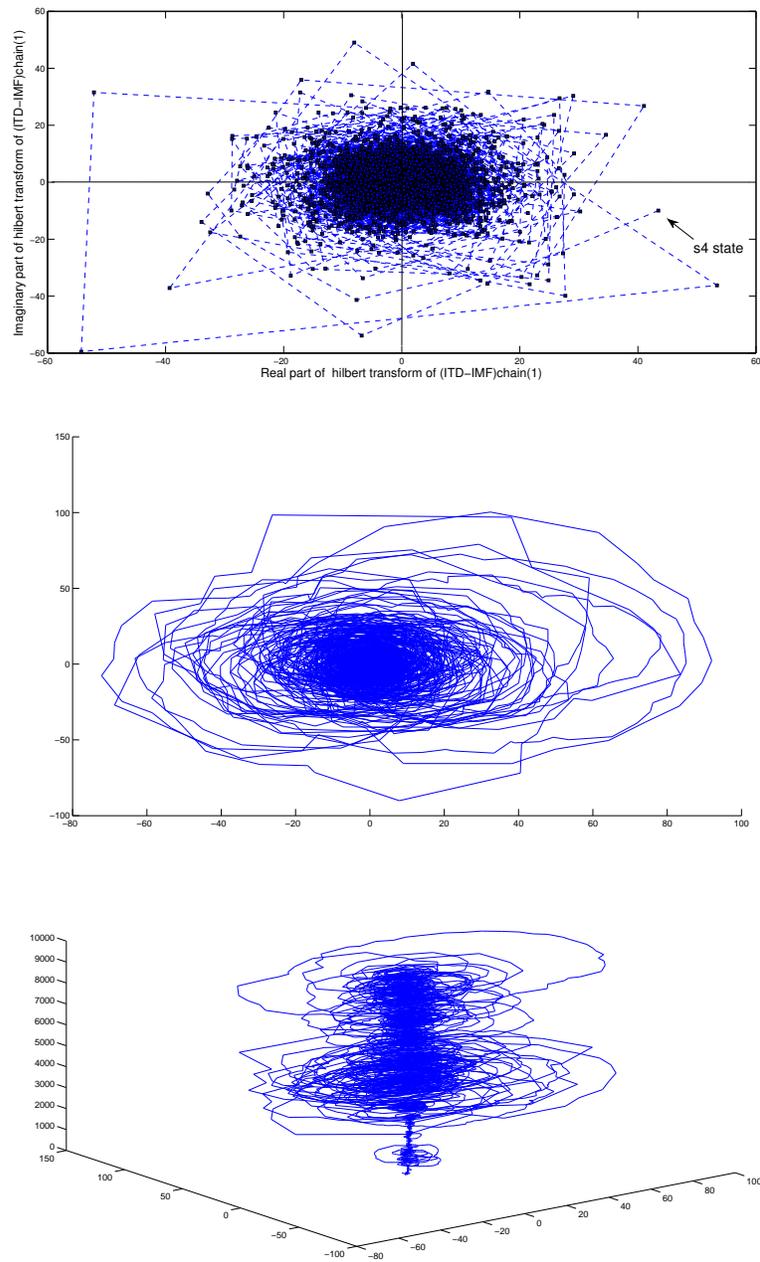


Fig. 4.9: The upper figure shows the graph of Hilbert transformation of  $(ITD - IMF)_{chain_1}$  of SET index of 100 daily closed price with the end point of time series in  $s_4(x_t)$  state. The bottom figure on the left represents Hilbert transformation of  $(ITD - IMF)_{chain_1}$  of SET index of 1000. The bottom right figure represents the same plot in 3-dimensional view.

An identification of equivalent class coming from an opposite direction of monotone function up and down and also maximum state and minimum state

$$[s_1] = -[s_3],$$

thus the inverse of  $s_1$  is  $s_3$ . Since

$$[s_2] = -[s_4]$$

we have an inverse of  $s_2$  to be  $s_4$ . We have

$$[s_1] + [s_3] = [0]$$

$$[s_2] + [s_4] = [0]$$

therefore

$$[s_1] + [s_3] + [s_2] + [s_4] = [0] \quad (4.38)$$

with  $[0]$  be a loop from the origin of time series in ground state to itself.

Therefore, in these constructions we allow all mixed states between  $s_i$  and expected path in hyperbolic space to  $s_i^*$  in complex structure of spinor field of 2 perpendicular cycles inducing a hidden field between each other as shown in Fig. 4.8. The suitable mathematical model can use a torus of time series data instead of Riemann sphere. The equipotential line of induced field between evolutionary feedback path between predictor state  $[s_i^*]$  and predictant state (a real state) is shown in Fig. 4.10.

### Quaternionic projective space

When we consider time series data in Euclidean plane we have an upper bound of value of measurement in real line  $x' > x_t$  for all  $t$ . In time coordinate we have a partial ordering of time scale so we have also an upper bound in time variable  $t' > t_i$  for all  $i$ . We will induce  $-t'$ , then we can connect  $t'$  with  $-t'$  and define cell decomposition as shown in Fig. 4.7. We glue  $e_1 \sim e_3$  and  $e_2 \sim e_4$ . Then we get a spinor field of time series data as Möbius trip of space of time series data.

Let  $D = \{1, i, j, k\}$  be the canonical basis for set of location  $\{T_1, T_2, T_3, T_4\}$  in  $\mathbb{R}^4$ . A real quaternion for time series is

$$x(t) = s_1(t) + s_2(t)i + s_3(t)j + s_4(t)k \in \mathbb{H}_{x(t)}, \quad (4.39)$$

this coordinate is a cyclic coordinate of value for time series.

Let  $\{s_1, s_2, s_3, s_4\}$  be a set of value in  $\mathbb{R}^4$ . A real quaternion for time series is

$$t(x) = T_1(x) + T_2(x)i + T_3(x)j + T_4(x)k \in \mathbb{H}_{t(x)}. \quad (4.40)$$

Let a mathematical definition of time series be a map between two quaternionic fields to quaternionic projective space  $\mathbb{H}_{x_t} \times \mathbb{H}_{t(x)} \rightarrow \mathbb{H}P^1 \simeq S^7/\text{Spin}(3)$ , where  $\text{Spin}(3)$  is a fibre state of time series with spinor field invariant property. It is an equivalent class of time series of glueing a state  $T_4$  with  $T_1$  state for a next cycle of cyclic coordinate.

For a given sequence  $x(t) \in \mathbb{R}$  it is known that  $\mathbb{H}P^1/\text{Spin}(3) = S^7$ . Let  $S^7 = \{\varphi = (x_1, x_2, \dots, x_8) \in \mathbb{R}^8, |\varphi| = 1\}$  be a hidden dimension of financial time series. In the next section we are going to prove that a space of time series in canonical form as defined above is a Kolmogorov space.

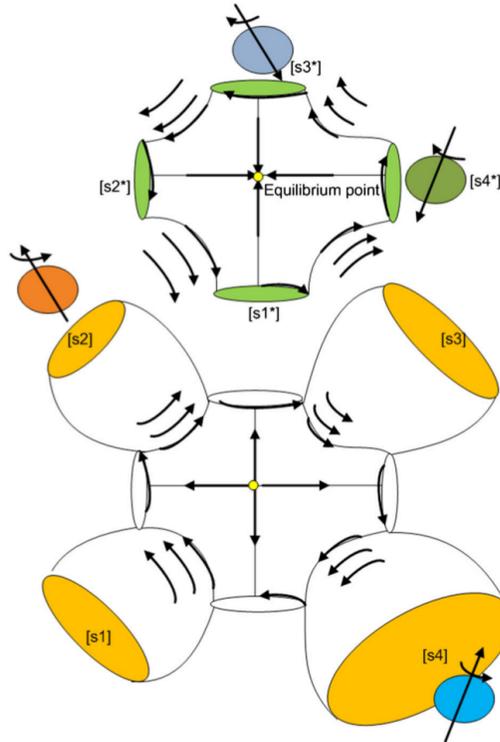


Fig. 4.10: A spin orbit coupling state between predictor and predictant states of spinor field of time series data induces an equipotential line as hyperbolic state for time series data. A path of expectation of physiology of time series is defined by this equipotential line of evolution feedback path between predictor and predictant. Uncoupling spinor fields of time series data can induce two types of fix pointed fields in Kolmogorov space for time series data. There exist 8 states of induced spinor field of time series data represented by red and green hyperbolic equipotential lines.

## 4.2 Proof of the main theorem

We knew that the generalization of Euclidean space of time series in  $\mathbb{R}^n$  is  $n$ -dimensional manifold. A local coordinate is defined as section of tangent of manifold together with Jacobian of coordinate transformation appeared as cocycle of group action over fibre bundle of tangent of manifold in which it is diffeomorphic to  $\mathbb{R}^n$ , when we assume that  $n$  data of time series data is embedded in  $n$ -dimensional manifold  $X$ , as underlying hidden topological space with value in tangent of manifold  $x_0 \in T_{x_0}X = p^{-1}(X)$ . In this case we will induce a sequence of tangent bundle of manifold of time series data in pointed space  $(X, x_0)$ , with disjoint union of covering space of time series data of chosen based points satisfying with sequence of time series data of

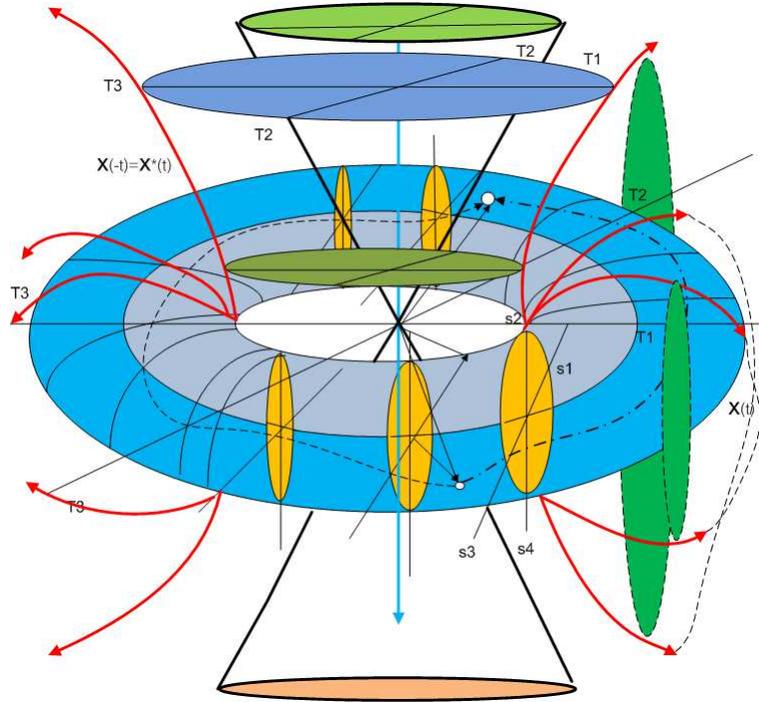


Fig. 4.11: A time series model of Kolmogorov space. The cyclic coordinate of time is perpendicular to physiology coordinate  $s_1, s_2, s_3, s_4$  of time series data. All points are separated and we prove that the complex projective plane of time series data in  $\mathbb{R}^8 \cup \{\infty\} = S^7$  is disjoint open set of separated point in  $CP^3$ . In this space a time series data can induce an equipotential line along hidden dimensions.

$$\begin{array}{ccc}
 \{x_0, x_1, \dots, x_n\} & & \\
 E = \bigsqcup_{i=0}^n T_{x_i} X & \longrightarrow & \{x_i\} \in U \times \mathbb{R}^n \\
 \downarrow p & & \downarrow \\
 x_i \in U \subset X & \longrightarrow & U \subset X
 \end{array}, \tag{4.41}$$

where  $E$  is a covering space of time series,  $X$  is a  $n$ -dimensional manifold of time series data with open set  $U_i$ ,  $F$  is a fibre space  $\mathbb{R}^n$  of time series data. Most people assume that a sequence of measurements is independent of period and substitute a discrete fibre from  $\mathbb{Z}$  with  $\mathbb{R}^n$ . We knew that a tangent space of  $\mathbb{R}^n$  is  $T\mathbb{R}^n = \mathbb{R}^n$  so we use covering space of  $X = \mathbb{R}^n$  for time series data in this case.

Let us assume that a time series data is embedded in non-Euclidean plane, a high dimensional sphere  $S^n$  with 2-extradimensions of  $S^1$  for fibre and hidden fibre induces hidden state  $t^* \in S^1$  and  $x^* \in S^1$  (see Fig. 4.11). In this case we model a time series in Riemann sphere  $S^2 = CP^1$

as based space with fibre in  $S^1$  and covering space in 4-dimensional space  $S^3$ . This is a principle bundle with discrete fibre  $S^1$ . This construction of principle Spin(3) bundle (spinor field of time series data) also allows us to define a loop structure between all data of time series connected to each other as path components and our time series model can be defined by a covering space of  $\mathbb{H}P^1$ .

**Theorem 1.** *A space of financial time series data is a covering space  $S^7$  with based space in  $X \simeq \mathbb{H}P^1$ . It is a Kolmogorov space with  $T_0$ -separation axiom.*

*Proof.*  $\mathbb{H}P^1$  is CW complex with one cell for each dimension  $k \leq 1$ . A cell complex is a Hausdorff space satisfying the separation axiom  $T_2$  which imply Kolmogorov space with lower  $T_0$ -separation axiom. We triangulate  $\mathbb{H}P^1$  into the union of disjoint subsapce  $\{e_\alpha, \alpha \in \Lambda\}$  called cells with  $e^n = S^n - \{(1, 0, 0, \dots)\} \subset \mathbb{R}^{n+1}$  with  $n$ -skeleton space of  $X$

$$X^n = \bigcup_{k \leq n} e_\alpha^k. \quad (4.42)$$

Let  $e_\alpha$  be an  $n$ -cell, then there exist a characteristic map of pairs

$$\chi_\alpha : (\mathbb{B}^n, S^{n-1}) \rightarrow (X, X^{n-1}) \quad (4.43)$$

which restricts to  $\mathbb{B}^n - S^{n-1}$  as homeomorphism onto  $e_\alpha$ .

Let  $q : S^7 \rightarrow \mathbb{H}P^1$  be the quotient map of principle Spin(3) bundle, and let  $u, v \in \mathbb{H}P^1$  with  $u \neq v$ , there are  $x, y \in S^7$  such that  $q^{-1}[\{u\}] = \{x, -x\}$  and  $q^{-1}[\{v\}] = \{y, -y\}$ . Let  $\epsilon = \frac{1}{3} \min\{\|x - y\|, \|x + y\|\}$  and set  $U = B(x, \epsilon) \cap S^7$  and  $V = B(y, \epsilon) \cap S^7$ , where the open balls are taken in  $\mathbb{R}^8$ . Then  $U, V, -U$  and  $-V$  are pairwise disjoint open neighborhoods of  $x, y, -x$  and  $-y$ , respectively, in  $S^7$ . Moreover,  $q^{-1}[q[U]] = -U \cup U$  and  $q^{-1}[q[V]] = -V \cup V$ . Therefore  $q[U]$  and  $q[V]$  are disjoint open neighborhoods of  $u$  and  $v$  in  $\mathbb{H}P^1$ .  $\square$

The real application of a Kolmogorov space of time series data is a directional prediction. We investigate a loop space of time series data of entanglement state of mixed direction between future direction and past direction in time series data. These states are suitable to open a short position or to open a long position in Stock Index Futures market. The ultimate goal of time series prediction is directional prediction. The typical output of a directional prediction is the prediction to up or down (or down with no direction change) with respect to the present value. We have tested the performance of our mathematical modeling using the forecasting methodology over nonlinear and nonstationary time series data of stock market price. There exist new tools for data analysis of nonlinear and nonstationary time series data so called Hilbert Huang transformation [69] and intrinsic time scale decomposition (ITD) [71]. These tools can be simple used together with the artificial neural network to predict the direction of stock price [72]. More about practical use of our concept and the preliminary results of the empirical data analysis with cyclic coordinate in financial time series data can be found in the Appendix D.

## 5 Conclusions

In the study we present new ideas from the contemporary field of econophysics, mainly we concentrate on the building of prediction models for time series in the framework of string theory.

We show, in the Section 2, how the string theory may motivate the adoption of the nonlinear techniques of the data analysis and how the numerical study recovers interesting fundamental statistical properties of the maps from the data onto string-like objects. The string maps give a geometric interpretation of the information value of the data. The model of the string allows one to manipulate the information stored along several extra dimensions. We start from the theory of the 1-endpoint and 2-endpoint strings, and we continue with more complex D2-branes. Most of the numerical investigations are obtained for the open topology, however, we describe briefly the ways to partial compactification and string polarized by the external fields.

The latter approach allow us to build the stable prediction models in trading in the financial market. 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. The real application of the multi-string structures is provided to demonstrate our ideas. A brief overview of the results and analysis is given in Section 3. The first model PMBSI 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 show the experimental results for the models on artificial and real world data with the most realistic trade conditions. For this purpose we have constructed trading algorithms, which were benchmarked against the most used class of the time series forecasting models and trading strategie. The stability of the algorithm on the transaction costs for long trade periods have been also demonstrated. The algorithms have been progressively enhanced by self-learning modules for the parametric optimization and a module for parallel evaluation of string moment values in the form of the genetics component which handles an autooptimization of algorithm in realtime.

In the Section 4, we provide the proof that the space of time series data over cyclic coordinate of location to local maximum and local minimum state of time series is a Kolmogorov space with  $T_0$ -separation axiom using the loop space of time series data. In our approach we define a cyclic coordinate of intrinsic time scale of time series data after empirical mode decomposition. A spinor field of time series data comes from the rotation of data around price and time axis by defining a new extradimension to time series data. We show that there exist hidden eight dimensions in Kolmogorov space for time series data. Our concept is realized as the algorithm of empirical mode decomposition and intrinsic time scale decomposition and it is subsequently used for preliminary analysis on the real time series data.

The received results led us to believe that our ideas and methodology can contribute to the solution of the open problems in the econophysics and we believe that our methods afford potential to be used in the practical applications.

## Acknowledgments

The work was partly supported by the grants VEGA 2/009/16, VEGA 2/0153/17 and APVV-0463-12. We would like to express our gratitude to Librade LTD for providing access to their flexible platform, team and community. Their professional insights have been extremely helpful during the development, simulation and verification of the algorithms.

### A PMBSI model construction

In our case we have the correlation between the strings in time series. As example we use 1-endpoint strings Eq. (2.7) with parameter  $q = 1$ . Usually 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

$$C(\tau, l_0) = \sum_{h=l_0}^l W(h) \left(1 - \frac{p(\tau - h)}{p(\tau - 1 - h)}\right) \left(1 - \frac{p(\tau - 1 - h)}{p(\tau - 2 - h)}\right), \quad (\text{A.1})$$

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

$$C(\tau, l_0) = C(\tau + 1, l_0). \quad (\text{A.2})$$

Now we want to find exact expression for the one step prediction  $p(\tau + 1)$ . Therefore we evaluate one step correlation invariant Eq. (A.2) with initial condition  $l_0 = 0$

$$\begin{aligned} W(0) \left(1 - \frac{p(\tau)}{p(\tau - 1)}\right) \left(1 - \frac{p(\tau - 1)}{p(\tau - 2)}\right) = \\ W(0) \left(1 - \frac{p(\tau + 1)}{p(\tau)}\right) \left(1 - \frac{p(\tau)}{p(\tau - 1)}\right) \\ + W(1) \left(1 - \frac{p(\tau)}{p(\tau - 1)}\right) \left(1 - \frac{p(\tau - 1)}{p(\tau - 2)}\right), \end{aligned} \quad (\text{A.3})$$

which can be rewritten in the more compact form

$$C(\tau, 0) = W(0) \left(1 - \frac{p(\tau + 1)}{p(\tau)}\right) \left(1 - \frac{p(\tau)}{p(\tau - 1)}\right) + C(\tau + 1, 1) \quad (\text{A.4})$$

and

$$\left(1 - \frac{p(\tau + 1)}{p(\tau)}\right) = \frac{C(\tau, 0) - C(\tau + 1, 1)}{W(0) \left(1 - \frac{p(\tau)}{p(\tau - 1)}\right)}. \quad (\text{A.5})$$

We finally obtain the prediction

$$p(\tau + 1) = p(\tau) \left(1 + \frac{C(\tau + 1, 1) - C(\tau, 0)}{W(0) \left(1 - \frac{p(\tau)}{p(\tau - 1)}\right)}\right), \quad (\text{A.6})$$

valid for  $p(\tau) \neq p(\tau - 1)$ . These are general definitions for the one step prediction correlation invariants.

Next in the construction of our PMBSI model, we suggest 2-endpoints mixed string model where one string is continuously deformed into the other with  $q > 0$ . The family of invariants is

written by use of the parametrization

$$\begin{aligned}
C(\tau, \Lambda) &= (1 - \eta_1)(1 - \eta_2) \sum_{h=0}^{\Lambda} W(h) \times \left(1 - \left[\frac{p(\tau)}{p(\tau + h)}\right]^q\right) \left(1 - \left[\frac{p(\tau + h)}{p(\tau + l_s)}\right]^q\right) \\
&\quad + \eta_1(1 - \eta_2) \sum_{h=0}^{\Lambda} W(h) \left(1 - \left[\frac{p(\tau)}{p(\tau + h)}\right]^q\right) \\
&\quad + \eta_2 \sum_{h=0}^{\Lambda} W(h) \left(1 - \left[\frac{p(\tau + h)}{p(\tau + l_s)}\right]^q\right), \tag{A.7}
\end{aligned}$$

where  $\eta_1 \in (-1, 1)$ ,  $\eta_2 \in (-1, 1)$  are variables that may be called homotopy parameters,  $q$  is a real valued parameter, and the weight  $W(h)$  is chosen in the bimodal single parameter form

$$W(h) = \begin{cases} 1 - W_0, & h \leq l_s/2, \\ W_0, & h > l_s/2. \end{cases} \tag{A.8}$$

and

$$W_0 = \frac{1}{\sum_{h'=0}^{l_s} e^{-h'/\lambda}}. \tag{A.9}$$

The above formulas do not represent the only setting of the weight parameters, the other settings can be tested. Then the term  $p(\tau_0 + l_{pr})$  is expressed in terms of the auxiliary variables

$$A_1(\Lambda, \tau) = (1 - \eta_1)(1 - \eta_2) \sum_{h=0}^{\Lambda} W(h) \left(1 - \left[\frac{p(\tau)}{p(\tau + h)}\right]^q\right), \tag{A.10}$$

$$A_2(\Lambda, \tau) = - (1 - \eta_1)(1 - \eta_2) \sum_{h=0}^{\Lambda} W(h) \left(1 - \left[\frac{p(\tau)}{p(\tau + h)}\right]^q\right) p^q(\tau + h), \tag{A.11}$$

$$A_3(\Lambda, \tau) = \eta_1(1 - \eta_2) \sum_{h=0}^{\Lambda} W(h) \left(1 - \left[\frac{p(\tau)}{p(\tau + h)}\right]^q\right), \tag{A.12}$$

$$A_4(\Lambda, \tau) = \eta_2 \sum_{h=0}^{\Lambda} W(h), \tag{A.13}$$

$$A_5(\Lambda, \tau) = - \eta_2 \sum_{h=0}^{\Lambda} W(h) p^q(\tau + h). \tag{A.14}$$

and the expected prediction form reads

$$\hat{p}(\tau_0 + l_{pr}) = \left[ \frac{A_2(\Lambda, \tau') + A_5(\Lambda, \tau')}{C(\tau_0 - l_s, \Lambda) - A_1(\Lambda, \tau') - A_3(\Lambda, \tau') - A_4(\Lambda, \tau')} \right]^{1/q}, \tag{A.15}$$

where we use the notation  $\tau' = \tau_0 + l_{\text{pr}} - l_s$ , ( $\tau' = \tau_0 - \Lambda$ ). The derivation is based on the invariance

$$C(\tau, l_s - l_{\text{pr}}) = C(\tau - l_{\text{pr}}, l_s - l_{\text{pr}}), \quad (\text{A.16})$$

and the model will be efficient if

$$C(\tau_0, \Lambda) \simeq C(\tau_0 + l_{\text{pr}}, \Lambda). \quad (\text{A.17})$$

Thus the model's free parameters are  $l_s, l_{\text{pr}}, \eta_1, \eta_2, q$  which must be set during the optimization phase.

PMBSI requires the time-series being processed to be non-negative. Otherwise the forecasts will not be defined (NaN). Even so PMBSI sometimes returns NaN values. This problem was fixed here by substitution of the NaN forecast by the most recent input for  $l_{\text{pr}}$  (naive prediction) and by the last valid forecast recorded for  $l_{\text{pr}} > 1$ .

## B The realisation of the PMBCS model

To demonstrate the behavior of physical ideas under the real trading conditions, we have constructed the trading algorithm based on the PMBCS model (Section 3.3). It has been intensively developed and tested on the trade online system [61] on different real financial data (OANDA, FOREX) and it provides highly professional algorithmization of the trading strategies with tick-data level accuracy of simulations.

For the PMBCS prediction model the trade system has worked with defined trading strategy, its trading algorithm is schematically reviewed in Fig. B.1.

Various sources serve as input data for the algorithm, e. g., real time data or historical data of currency rates, we have chosen the second one. Data are handled by the current prediction model (PMBCS Prediction Model). In the heart of the algorithm there lies a momentum calculator module (Moment Predictors). The module calculates the values of the momentum predictors,

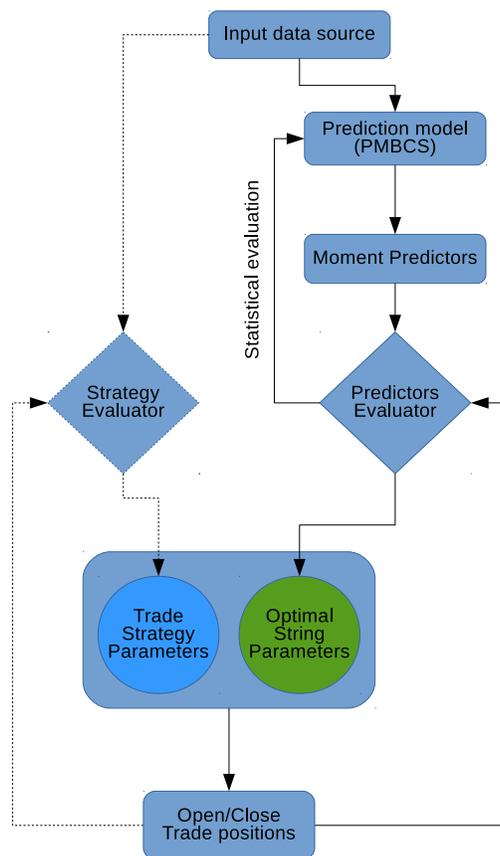


Fig. B.1: The scheme of the trading strategy for the prediction model based on the deviations from the closed string/pattern form.

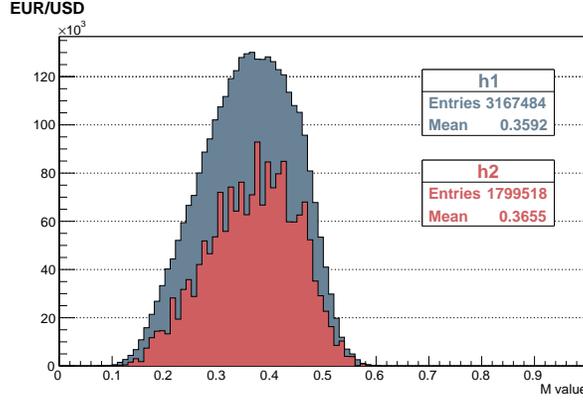


Fig. B.2: The typical distributions evaluated by the Predictors Evaluator module. The distributions of the incoming ( $h_1$ ) and outgoing ( $h_2$ ) momenta values  $M$ , normalized to the interval  $(0, 1)$ .

which determine the values for the string parameters of the algorithm (Optimal String Parameters). These values directly affect the trade positions (Open/Close Trade positions). This is the most direct way. But the algorithm works also in the parallel way. The values of the momentum predictors are statistically evaluated with the results of the open/close trade positions and compared with predicted values (Predictors Evaluator). Then the “optimal” values of the momenta are sent ahead to provide the “optimal” values of the string parameters (see Fig. B.2). The left side of the scheme outlines future Strategy Evaluator. Its purpose is to evaluate the trade strategy parameters and this way to control the risk in a more sophisticated way.

The first set of parameters, including length of moment string (number of ticks or time period), quotient or exponent of moment, frequency of momentum function, phase shift of momentum function. describes the momentum (simple scalar function of several variables from the interval  $(0, 1)$ ).

Beside the first set of parameters, there exists also the second set of parameters, which controls the risk of the algorithm. The risk can vary from zero (low risk but also low or zero number of trades) to the boundary values. These parameters are called the trade strategy parameters and in our case they are represented by a maximum number of simultaneously opened trades, skewness of momenta distribution and Sharpe ratio (C.1) of closed trades. Together, the trade strategy parameters and “optimal” string parameters determine the final opening and closing of trade positions.

An arbitrage opportunity taking advantage of the occurrence of difference in distribution. The opportunity is measured by Kullback-Liebler divergence

$$D_{KL} = \sum_{j(\text{bins})} \text{pdf}(M^+(j)) \log \left( \frac{\text{pdf}(M^+(j))}{\text{pdf}(M^-(j))} \right) \quad (\text{B.1})$$

where larger  $D_{KL}$  means better opportunities ( $D_{KL} > D_{\text{threshold}}$ ), e. g., when  $D_{KL} > D_{\text{threshold}}$

it means the buying Euro against USD could be more profitably. Statistical significance means the smaller the statistics accumulated into bins  $\text{pdf}(M^+(j))$ ,  $\text{pdf}(M^-(j))$ , the higher is the risk ( $M$  from the selected range should be widespread). The meaning of pdf in the definition above is the probability density function.

More generally we can construct the series of  $(l_s + 1)$  price tics  $[p(\tau), p(\tau + 1), \dots, p(\tau + l_s)]$  which are transformed into a single representative real value  $M(\tau + l_s)$ . Nearly stationary series of  $M(\tau + l_s)$  yields statistics which can be split into: a branch where  $M$  is linked with future uptrend/downtrend and a branch where  $M$  is linked with future profit/loss taking into account transaction costs. The accumulation of  $\text{pdf}(M_{\text{long}}^{+-})$  means (profit+/loss-) or  $\text{pdf}(M_{\text{short}}^{+-})$  (profit+/loss-),  $M^+$  in Eq. (B.1) brings profit and  $M^-$  loss.

From the trading point of view, we have tried various methods of predictor integration with components of trading algorithm, that is build in modular architecture and thus consists from various components, which follows the separation of concerns principle (every component provides a service and its state and logic is encapsulated from the rest of the components).

String predictor is now used as the foundation for entry rule component, that handles the points when the algorithm will enter into position (when the position will be opened). The entry rule has a process to evaluate and also make a decision of “how to open the position” based on predictors BUY/SELL signal. It also takes into consideration the size of already opened position, and it calculates size of new opening position based on money and risk management restrictions. Besides entry rules, there are components like exit rules, trade book, quantity calculator, trading brake, custom event component, genetics component, statistics.

The exit rules take care about when and how to exit from the position (close the position). There are various approaches which utilizes combinations of common trading techniques, like take profit, stop loss, trailing stop, etc. The algorithm internal trade book component keeps track of all open positions, creation of new orders, selection of trading technique (like usage of MARKET, LIMIT, STOP, STOPLIMIT orders to open/close position). The quantity calculator counts the size of the new opened position based on not solely money management restrictions like account and investment instrument used margin, but also internal logic of algorithm. The trading brake is a simple component which allows algorithm to keep track of market irregularities by continuous monitoring volatility and spread (e. g., spread tends to widen before big event hits the market), another possibility is the incorporation of the angular momentum values instead of the historical volatility (Fig. 3.14).

### PMBCS Self-learning model

The algorithm version StringAlgo v.15 has demonstrated the financial forecasting on the OANDA real data for the PMBCS Self-learning model.

The simulations were carried out on the OANDA data for EUR/USD currency rate for the time period of three months, from 2010/07/15 to 2010/10/15. The typical background of the simulations is presented in Figs. B.2–B.4. Figure B.3 shows the dependence in the three month simulation on the ask-bid data spread, i. e., on the transaction costs (Fig. B.4(a)). Also for long trade periods our algorithm behaves very stably, the values of NAV do not show the high dependence of the transaction costs. The distribution of trades is nearly uniform throughout all time period, it varied in the range from 0 to 14 per day (Fig. B.4(b)). There are no rapid increases and decreases of the amounts of the trades.

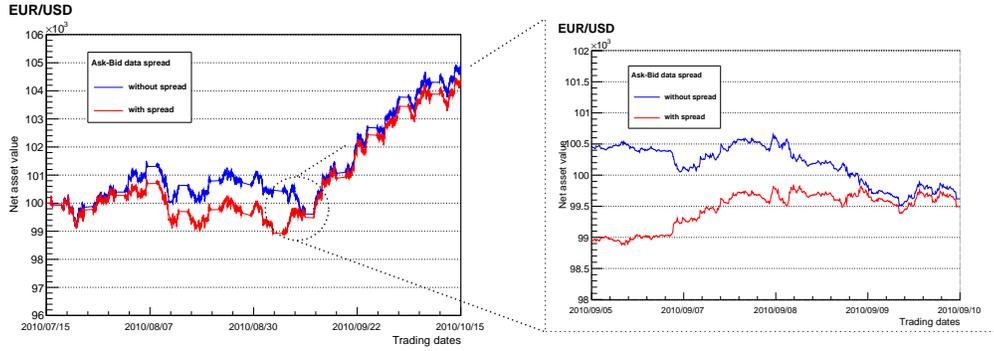


Fig. B.3: The net asset value of the simulation ( $l_s = 1000, Q = 1, m = 1$ ) on the EUR/USD currency rate for a selected time period as the dependence on the ask-bid data spread.

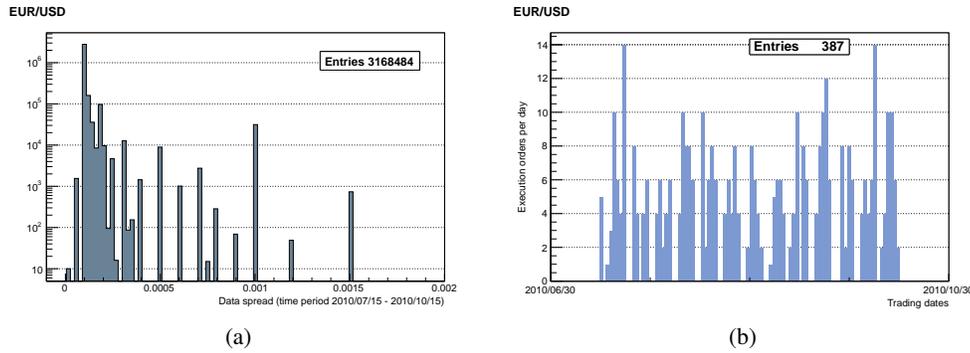


Fig. B.4: (a) The ask–bid spread in the evaluated OANDA data, (b) the typical distribution of the execution orders per day. Both histograms are for the time period from 2010/07/15 to 2010/10/15.

In Figs. B.5–B.7 we described some interesting results relating to the string parameters. The net asset value of the model dependence on the string length parameter  $l_s$  is presented in Fig. B.5(a). As one can see, the value of  $l_s = 900$  seems to be most promising, this value was fixed for the next predictions. The dependence of the model on the parameter  $Q$  – the quotient of the moment is described on the next figures. Figure B.5(b) represents the dependence on low values, i. e.,  $Q = 1, 2, 4, 8$ , Fig. B.6(a) represents the dependence on higher values, i. e.,  $Q = 1, 16, 24, 32$ . The comparison of the value  $Q = 24$ , which seems to be most suitable for the next forecasts, with the simultaneous use of three values is shown in Fig. B.6(b).

The interesting case is the choice of the regular function  $F_{CS}$ . The previous forecasts [25,27] were made for a trigonometric function  $\cos(x)$ . Here we are presented the results of the tests with other functions, as one can see in Fig. B.7. The comparisons of the forecast with the function  $\cos(x)$  were made for forecasts with functions  $\sin(x)$ ,  $\sinh(x)$  and  $\cosh(x)$  in Fig. B.7(a). The

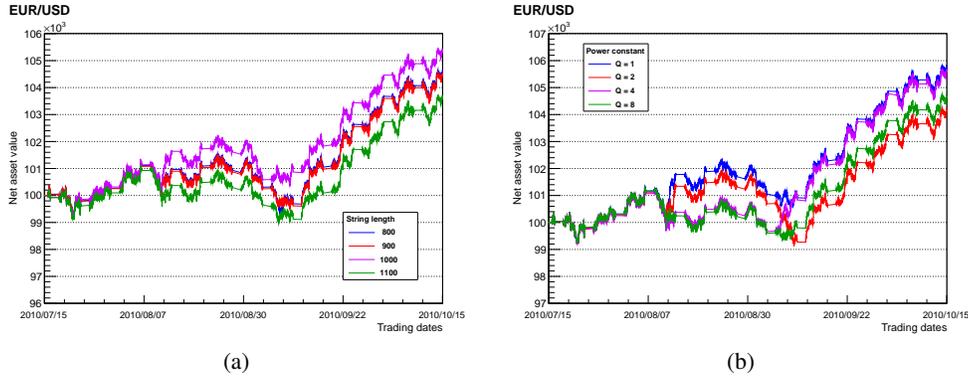


Fig. B.5: The net asset value of the model on the EUR/USD currency rate for a selected time period as the dependence on the string length parameter and the power constant

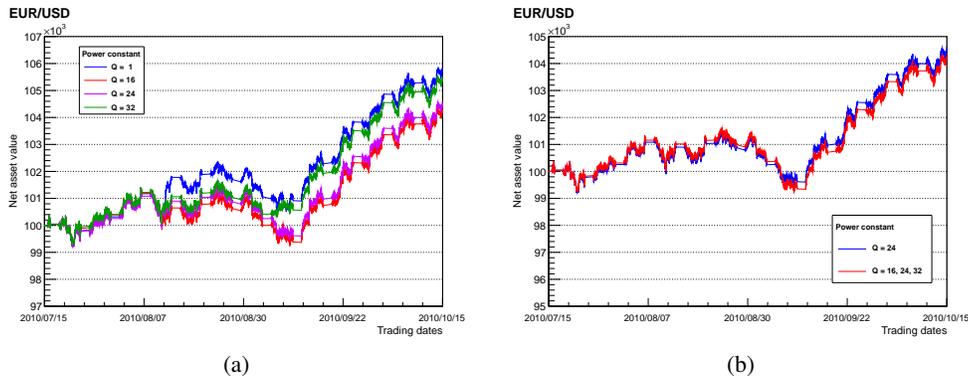


Fig. B.6: The net asset value of the model on the EUR/USD currency rate for a selected time period as the dependence on the power constant.

subfigure Fig. B.7(b) represents the comparison of forecasts for the function  $\sin(x)$  with the different arguments, i. e.,  $x$  and  $x + \phi$ , where  $\phi = 0, \pi$ .

The values of string parameters, which were used in previous simulations, are summarized in Table B.1 in a Simple model column. For each forecast we used one value for each type of parameters:  $l_s, Q, F_{CS}, m$  and  $\phi$ , respectively. In other words, only one set of string parameters is used, we have denoted it by  $n_s = 1$ . However, the algorithm can work with various sets of free parameters simultaneously ( $n_s \geq 1$ ). This is possible due to the fact that the model was enhanced to so called Self-learning model. In terms of the string parameters from Table B.1, all possible combinations of the values from the third column are taken into account and the corresponding momentum predictors are calculated.

The best combinations of string parameters is evaluated in Predictor Evaluator module by

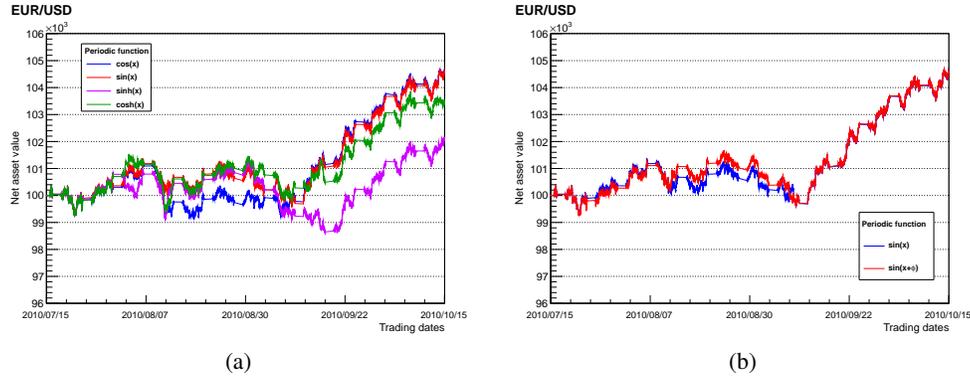


Fig. B.7: The net asset value of the model on the EUR/USD currency rate for a selected time period as the dependence on the periodic function  $F_{CS}$ , (a) the comparison of the forecasts for the functions  $\cos(x)$ ,  $\sin(x)$ ,  $\sinh(x)$ ,  $\cosh(x)$ , (b) the comparison of the forecasts for the functions  $\sin(x + \phi)$ , where  $\phi = 0, \pi$ .

means of Sharpe ratio. For this purpose, the Sharpe ratio needs a sufficient amount of predicted momenta to provide reliable optimization. The number of sets of string parameters  $n_s$  must be fixed at a sufficient number. The lower value of  $n_s$  means less statistics available for the finding of optimal parameters. On the other hand, very high values of  $n_s$  need the corresponding computing power. In Fig. B.8, we compared the effect of the higher value of  $n_s$  for the Self-learning model with  $n_s = 2$  (left column) and  $n_s = 16$  (right column). The histograms show the average values of the execution reports sent by the model. In the subfigures B.8(a)–B.8(b) (three months data), the effect is not seen, however, in the subfigures B.8(c)–B.8(d) (one year data) the number of the execution reports is approximately 100 times higher. It means that the algorithm is more flexible in trading for a longer period.

String parameters	PMBCS Simple model	PMBCS Self-learning model
$l_s$	800, 900, 1000, 1100	[900]
$Q$	1, 2, 4, 8, 16, 24, 32	[8, 16, 24, 32]
$F_{CS}$	$\cos(x)$ , $\sin(x)$ , $\sin(x + \phi)$ , $\sinh(x)$ , $\cosh(x)$	$[\cos(x + \phi)]$
$m$	0, 1, 2, 3	[0, 1, 2, 3]
$\phi$	0, 3.14	[0, 3.14]

Tab. B.1: The values of string parameters used for the PMBCS Simple and Self-learning models. The square brackets emphasize the fact that the Simple model works with exactly one value of the analyzed string parameter values, while the Self-learning model can work with sets of parameters simultaneously.

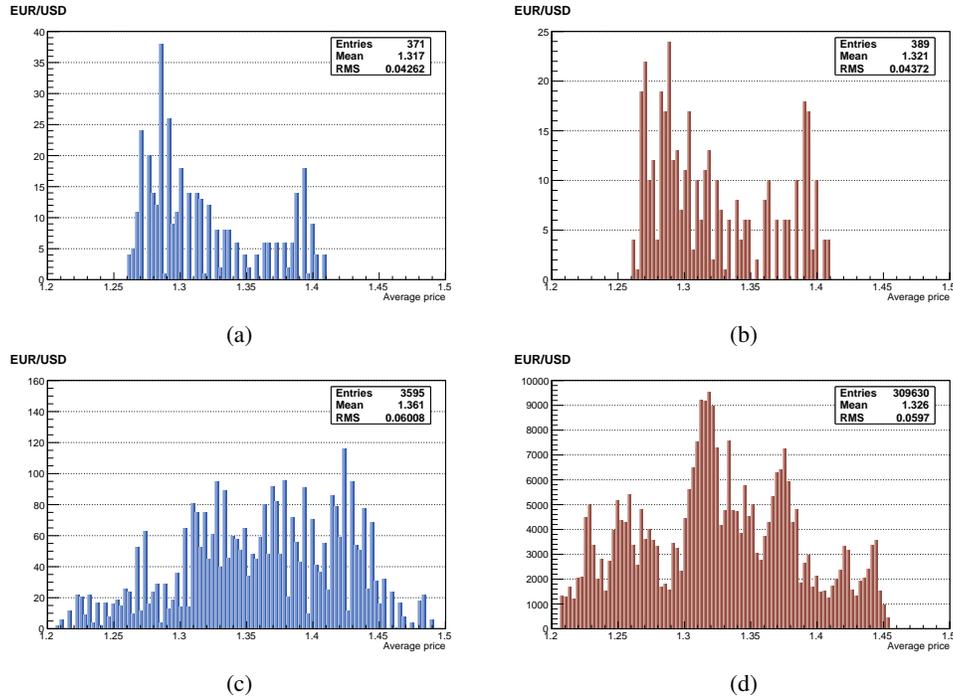


Fig. B.8: The histograms of the average values of the execution reports for the Self-learning model with  $n_s = 2$  (left column) and  $n_s = 16$  (right column). The time period of the simulation is three months for (a), (b) and one year for (c), (d).

The opening and closing of trade positions is determined by statistical evaluation of string momenta. In Fig. B.9, such a statistical procedure is demonstrated graphically for  $n_s = 16$ . Each new price tick leads to the evaluation of string momenta, in our case to sixteen values equal to  $-1, 0, 1$ . One can see the distribution of evaluated values in Fig. B.9(b) for a very short time period. A few positive values and one negative are clearly visible, the others are equal to zero. Then in Fig. B.9(b) the red dot represents the summarized value of the evaluated string momenta normalized to  $\pm 1$ . The blue dots are the EUR/USD currency rate ticks from 2010/07/15 10:00:00 up to the first  $5 \times 10^5$  ones. On the left of the subfigure one can see the instant of the “learning”, when the statistics is gained and the momenta do not predict any values.

### PMBCS live version model

The encouraging results with evolutionary algorithm for the parametric optimization [28] lead us to enhance the algorithm with a module for parallel evaluation of string moment values in the form of the genetics component which handles an autooptimization of algorithm in realtime. The genetics component is constantly executing and evaluating multiple inner simulations trying to evolve its internal parameters. The trends in price change, identified either with volatility,

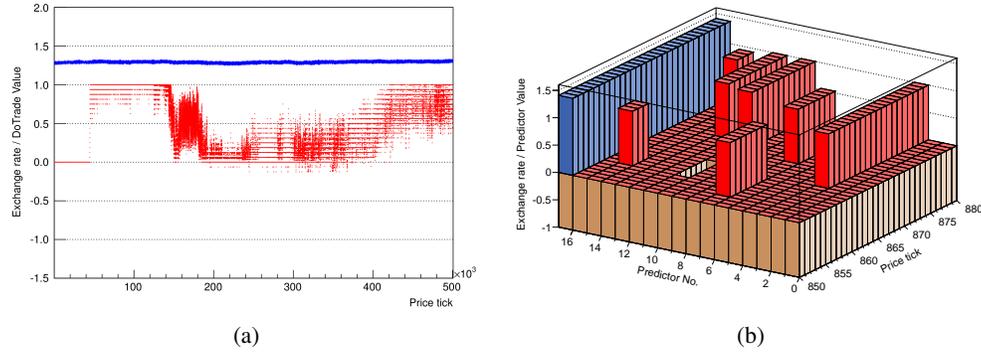


Fig. B.9: The EUR/USD currency rate ticks from 2010/07/15 10:00:00 (blue dots) with the evaluated values of summarized prediction (red dots), normalize to  $\pm 1$  (a). The detail of subfigure (a) for a very short time period (b), the evaluated string momenta values (red) with the currency rate ticks on the background (blue).

Simulation	Trade altitude	$l_s$	Bin	$F_{CS}$	Moment type
SIM16-P009	900	1400	0	4	0
SIM16-P010	700	1400	0	1	0
SIM16-P019	700	1400	2	4	0
SIM16-P026	850	1600	2	4	0
SIM16-P046	85000	1500	2	4	0
SIM16-P047	85000	1500	2	4	0
SIM16-P052	90000	1400	0	4	0
SIM16-P058	700	1900	0	0	1
SIM16-P066	70000	720	0	2	0
SIM16-P069	70000	1500	2	4	1

Tab. B.2: Summary of the string parameters for the simulations with StringAlgo v1.16.

yield to dynamic change of the parameters as a string length and a trade altitude. They are not keep constant, e. g., the trade altitude is lowered, so the algorithm can profit even under new conditions. The genetics component has predefined limits within which it selects the most suitable combination of parameters leading ultimately to BUY/SELL orders.

In contrast to previous case when only EUR/USD currency rate was used, new currency rates CHF/JPY, AUD/CAD, AUD/JPY were simulated with upgraded algorithm (version StringAlgo v.16) and thereafter compared with the results from demo sessions Tabs. B.2, B.3. The algorithm StringAlgo v.16 has builtin new proposed string maps Eq. (3.8), (3.9), (2.25), as well the modified Sharpe ratio (C.3) which serves as new statistical quantity to evaluate the value at risk.

Demo session	Currency pair	Session start	Session end	Simulation code
IB-test-12	EUR/USD	2015-09-28	2015-11-23	SIM16-P009
LMAX-test-16	EUR/USD	2015-10-13	2015-12-03	SIM16-P010 SIM16-P019 SIM16-P026
LMAX-test-13	CHF/JPY	2015-10-09	2015-12-01	SIM16-P046 SIM16-P047 SIM16-P052
LMAX-test-14	AUD/CAD	2015-10-12	2015-12-05	SIM16-P058
LMAX-test-15	AUD/JPY	2015-10-12	2015-12-03	SIM16-P066 SIM16-P069

Tab. B.3: Summary of opened sessions for real demo trading on the Interactive Brokers and LMAX Exchange market accounts.

### C Sharpe ratio and volatility

The Sharpe ratio for calculating risk-adjusted return

$$S = \frac{E[r_a - r_f]}{\sigma} = \frac{\mu - r_f}{\sigma}, \quad (\text{C.1})$$

where  $r_a$  is asset return,  $r_f$  is risk free rate of return,  $E[r_a]$  is mean asset return,  $E[r_a - r_f] = \mu - r_f$  is the expected value of the excess of the asset return over the benchmark return with standard notation

$$\left\{x_i\right\}_{i=1}^N, \quad \mu = \frac{1}{N} \sum_{i=1}^N x_i, \quad \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (\text{C.2})$$

$\mu$  is the mean,  $\sigma$  is the standard deviation.

The Sharpe ratio formula for the *modified value at risk*

$$S_{\text{MVaR}} = \frac{\mu - r_f}{\text{MVaR}}, \quad (\text{C.3})$$

with

$$\text{MVaR} = -(\mu + \sigma z_{cf}),$$

$$z_{cf} = z_c + \frac{1}{6} [(z_c^2 - 1)S] + \frac{1}{24} [(z_c^3 - 3z_c)K] - \frac{1}{36} [(2z_c^3 - 5z_c)S^2],$$

$z_c$  is the  $c$ -quantile of the standard normal distribution,  $S$  is the skewness of asset return and  $K$  is the excess kurtosis of asset return.

The volatility  $\sigma$  denotes the degree of variation of a trading price of a financial instrument for a specified time period as measured by the standard deviation of logarithmic returns (C.2). The generalized formulae for the calculation of volatility  $\sigma_T$  for time horizon  $T$  in years is

$$\sigma_T = \sigma\sqrt{T}. \tag{C.4}$$

The most common is annualized volatility  $\sigma_{\text{annualy}} = \sigma\sqrt{252}$ , where  $\sigma$  is 1-day historical volatility and 252 is the number of trading days in year. The monthly volatility is then  $\sigma_{\text{monthly}} = \sigma\sqrt{21}$ .

## D Empirical data analysis of time series data by decomposition methods

### Empirical mode decomposition and intrinsic time scale decomposition

The technique of Hilbert-Huang empirical mode decomposition (EMD) method [69] is a new tool for the analysis of complex time series data into elementary, almost orthogonal components that do not overlap in frequency. EMD algorithm is a novel application of Hilbert-Huang transformation technique for nonlinear and nonstationary time series. For a given signal  $x_t$  we define the EMD to be a transformation from spatial space to zero band space of zero crossings

$$x_t = \sum_{i=1}^n c_i(t) + r(t) \quad (\text{D.1})$$

$$\text{EMD}_i(x(t)) = c_i(t), \quad i = 1, \dots, n \quad (\text{D.2})$$

where  $c_i(t)$  is an  $i$ -th intrinsic mode function (IMF) creating an adaptive orthogonal basis of time series data and  $r(t)$  is an intrinsic trend.

The IMF satisfies two conditions: (1) in the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one; (2) at any point, the mean value of the envelopes defined by the local minima and the local maxima is zero. The problem of IMF is the local maximum and minimum point of original time series and IMF is not in the same location. These problem we called intermittency problem of EMD algorithm. For a given time series  $x_t$ , the stepwise procedure of EMD can be summarized as follows

- Step 1.** Identify all the local maxima and minima of time series  $x_t$ .
- Step 2.** Generate the upper and lower envelopes of  $x_t$ . Once the extrema are identified, all the local maxima are connected by a cubic spline line as the upper envelope. Repeat the similar procedure for the local minima to produce the lower envelope. The upper and lower envelopes should cover all the data between them.
- Step 3.** Calculate the point-by-point mean  $m_1$  from the upper and lower envelopes.
- Step 4.** Define the difference between the data and  $m_1$  as the first component of  $h_1$

$$x_t - m_1 = h_1. \quad (\text{D.3})$$

- Step 5.** The sifting process has to be repeated more times,  $h_1$  is treated as the data

$$h_t - m_1 = h_{11}. \quad (\text{D.4})$$

We can repeat this sifting procedure  $k$  times until  $h_{1k}$  is an IMF, that is

$$h_{1(k-1)} - m_{1k} = h_{1k}. \quad (\text{D.5})$$

Then the first IMF component from the data can be designated as

$$c_1 = h_{1k}, \quad (\text{D.6})$$

where  $c_1$  is separated from the rest of the data by

$$x_t - c_1 = r_1. \quad (\text{D.7})$$

**Step 6.** The residue  $r_1$  is treated as the new data and subjected to the same sifting process as described in Steps 1.—5. This procedure can be repeated on all the subsequence  $r_j$  and the result is

$$r_1 - c_2 = r_2, \dots, r_{n-1} - c_n = r_n. \quad (\text{D.8})$$

The sifting process can be stopped by any of the following predetermined criteria: either when the component  $c_n$  or the residue  $r_n$  become so small that it is less than the predetermined value of substantial consequence, or when the residue  $r_n$  become a monotonic function from which no more IMF can be extracted. By summing up all of the IMFs, we can finally obtain

$$x_t = \sum_{i=1}^n c_i + r_n. \quad (\text{D.9})$$

Thus, the original time series  $x_t$  has been decomposed into  $n$  empirical modes and a residua  $r_n$ .

### Ensemble empirical mode decomposition

In ensemble EMD (EEMD) procedure a white Gaussian noise (WGN) is added directly to the input signal of interest,  $x_t$  can be input time series, before applying EMD [73]. The perturbed signal is given by

$$x_v(t) = x_t + v_t, \quad (\text{D.10})$$

$v_t$  is the standard deviation of noise. The stepwise procedure of EEMD algorithm can be summarized as follows

**Step 1.** Perturb the input signal  $x_t$  as described by Eq. (D.10).

**Step 2.** Apply the EMD algorithm to  $x_v(t)$  to obtain IMF set  $\{c_i(t)\}_{i=1}^M$ .

**Step 3.** Repeat Step 1. and 2. for the signal realizations of WGN and estimate average IMF set

$$\{\overline{c_i(t)}\}_{i=1}^M = \frac{1}{S} \left( \{c_i(t)\}_{i=1}^M + \dots + \{c_i(t)\}_{i=1}^M \right).$$

### Intrinsic time scale decomposition

Intrinsic time scale decomposition (ITD) [71] decomposed the original signal into  $\text{ITD}_i(t)$  and monotonic trend  $r_1(t)$  by using a baseline function

$$x_t = \sum_{i=1}^n \text{ITD}_i(t) + r_1(t), \quad (\text{D.11})$$

with  $\text{ITD}_i(t)$  defined by using residua  $H_t$  after the recursive subtraction with baseline function  $L_t$ , similarly to EMD process,

$$x_t = L_t + H_t = Lx_t + (1 - L)x_t.$$

The baseline function  $L_t$  of  $\text{ITD}_i(t)$  is defined assuming

$$L_t = L_k + \frac{L_{k+1} - L_k}{x_{k+1} - x_k}(x_t - x_k), \quad (\text{D.12})$$

where  $L_t$  is an extremum location for  $t \in (\tau_k, \tau_{k+1}]$ . The recursive process of a decomposition of time series is coming from the calculation of forward looking baseline function with three input parameters

- the values of extrema  $(x_k, x_{k+1}, x_{k+2})$ ,
- the locations of extrema  $(\tau_k, \tau_{k+1}, \tau_{k+2})$ ,
- and the adjusting parameter  $\alpha$

with

$$L_{k+1} = \alpha \left[ x_k + \frac{\tau_{k+1} - \tau_k}{\tau_{k+2} - \tau_k}(x_{k+2} - x_k) \right] + (1 - \alpha)x_{k+1}. \quad (\text{D.13})$$

### Running (ITD – IMF)chain<sub>1</sub>

The EEMD transformation has a problem of the location of local extrema point of original signal if the location of input time series data and the height of IMF is not the same as the height of the original time series data. We solved this problem by performing the ITD, then sending the result of  $\text{ITD}_1$  to EEMD transformation and to get IMF without the mentioned problem above. The result is called running (ITD – IMF)chain<sub>1</sub> with

$$x_t = \sum_{i=1}^n \text{ITD}_i(t) + r_1(t). \quad (\text{D.14})$$

We select only  $\text{ITD}_1(t)$  to perform a further EEMD process with

$$\text{ITD}_1(t) = \sum_{i=1}^n c_i(t) + r_2(t), \quad (\text{D.15})$$

where we call

$$(\text{ITD} - \text{IMF})\text{chain}_1(t) = c_1(t). \quad (\text{D.16})$$

The important point of this result is a minimum structure of local maximum and local minimum state of time series data in which we can identify the minimum local structures of physiology of financial time series. The local maximum and minimum point of  $(\text{ITD} - \text{IMF})\text{chain}_1(t)$  lay at the same location of original signal  $x_t$  without intermittency problem. If we can predict the local maximum and minimum state of  $(\text{ITD} - \text{IMF})\text{chain}_1(t)$  it means we can overcome the prediction of local maximum and local minimum state of original time series data. In this research we use  $(\text{ITD} - \text{IMF})\text{chain}_1(t)$  as skeleton of time series.

**Definition 5.** We call  $(\text{ITD} - \text{IMF})\text{chain}_1(t)$  of time series  $x_t$  a skeleton of time series data  $x_t$ .

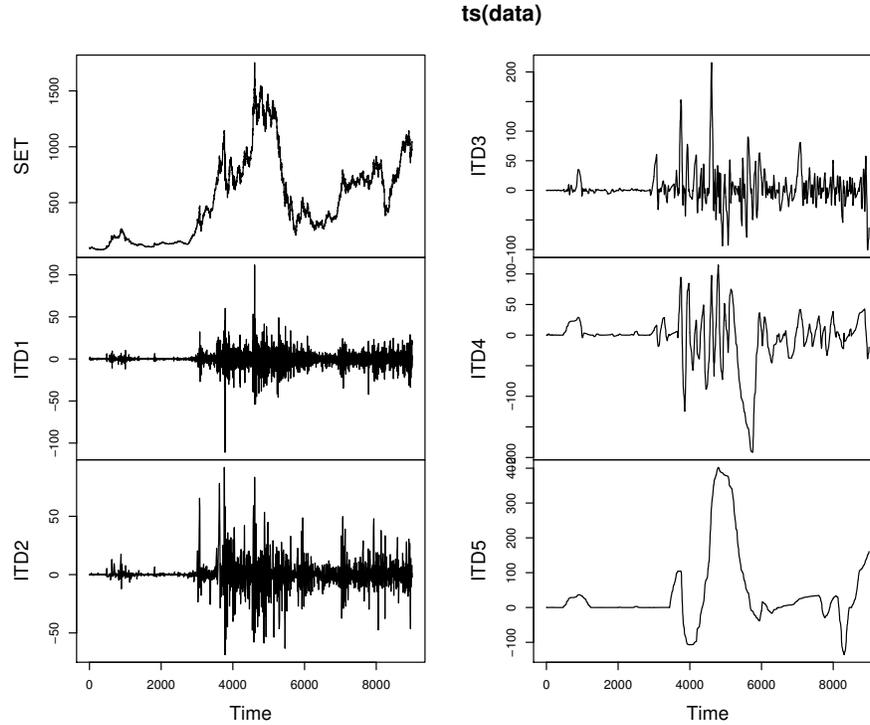


Fig. D.1: Intrinsic time scale decomposition  $ITD_1 - ITD_5$  of time series data of SET index daily closed price. A data showed here is a daily closed price of SET index between 2/5/1975 to 12/09/2011 with 9000 time series data and  $ITD_1 - ITD_5$  at the same periods between raw data 9000 data of SET index.

In this research, we have used financial time series data of daily closed price of SET, a Thai stock market index. We have mainly considered the daily closing prices of SET during the periods from the beginning of market on 2/5/1975 to 12/9/2011, totally of 9000 data points of time series for our data analysis. At first we have computed  $ITD_1 - ITD_5$  of SET index and result is shown in Fig. (D.1). At second we have used the result of  $ITD_1$  for the computation of EEMD in order to get  $(ITD - IMF)_{chain_1}$ . The EEMD computation was performed with the standard deviation of noise 0.05 with 1000 rounds of running and it has taken about 2 hrs. per sample point.

### Empirical analysis of cyclic time scale

We have used data of 9000 daily closed prices of SET. We have got only 2792 cycles in  $(ITD - IMF)_{chain_1}$  with zeros crossing of cyclic coordinate for time series data. That means we have 2792 maximum points which equal to minimum points of time series data. We have computed

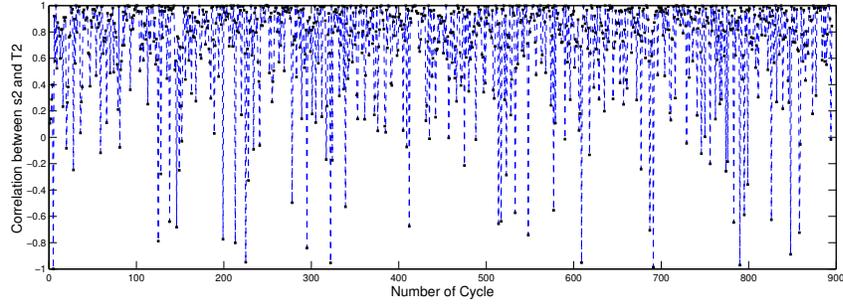


Fig. D.2: The correlation between  $s_2$  and  $T_2$  of  $(ITD - IMF)_{chain_1}$  of SET in the first 900 cycles.

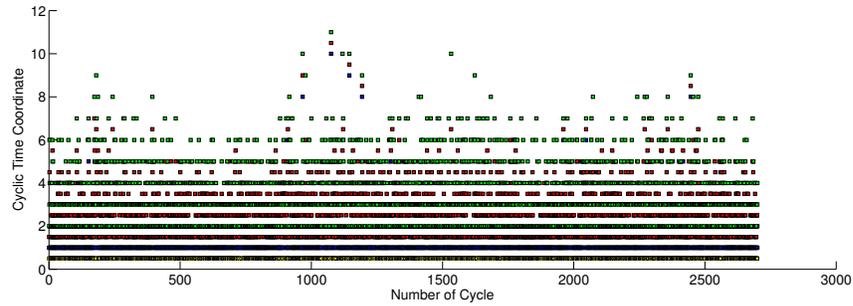


Fig. D.3: The cyclic time coordinate of  $(ITD - IMF)_{chain_1}$  of SET daily closed price data from 2/5/1975 to 12/09/2011 with 9000 data point. The higher dot on the vertical line is  $T_4$  cyclic time scale coordinate. It is a time from minimum to a next minimum point (labeled in green color). The lower dot on vertical line is  $T_1$ , a cyclic time scale from local minimum point to monotone function up (labeled in yellow color). The  $T_2$  is labeled with blue color, it is a cyclic time coordinate from minimum to maximum point. The  $T_3$  is labeled with red color, it is a cyclic time coordinate from minimum to monotone function down. For each cycle there are four points in the vertical line. The horizontal line represents time circle. The higher ITD will contain lower circle of time scale. We plotted only 2792 cycles in  $(ITD - IMF)_{chain_1}$  with zeros crossing of cyclic coordinate for time series data on the leftside of panel. The highest peak label represent about 12 days of cycle (labeled in blue color).

the correlation between state  $s_2$  of  $(ITD - IMF)_{chain_1}$  and cyclic time scale  $T_2$  (see Fig. D.2). The highest period in one smallest cycle  $T_4$  is 11 days. We can notice from the highest point of graph Fig. D.3 in cyclic time coordinate.

The graphs of Hilbert transformation of  $(ITD - IMF)_{chain_1}$  and  $ITD_1 - ITD_5$  of SET are shown in Figs. D.4, D.5, D.6.

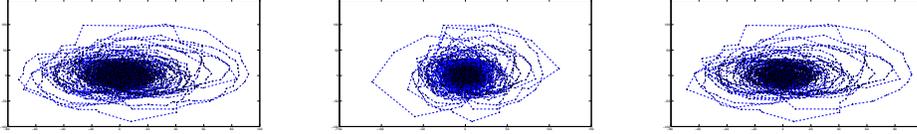


Fig. D.4: The graph of Hilbert transformation of  $(ITD - IMF)chain(1)$  on the left,  $ITD_1$  in the middle and  $ITD_2$  on the right.

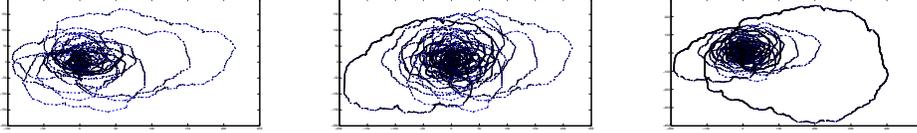


Fig. D.5: The graph of Hilbert transformation of  $ITD_3$  on the left,  $ITD_4$  in the middle and  $ITD_5$  on the right.

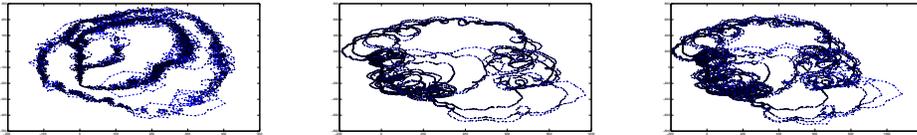


Fig. D.6: The graph of Hilbert transformation of  $ITD_1$  plus  $ITD_5$  on the left,  $ITD_2$  plus  $ITD_3$ ,  $ITD_4$ ,  $ITD_5$  in the middle and  $ITD_1$  plus  $ITD_2$ ,  $ITD_3$ ,  $ITD_4$ ,  $ITD_5$  on the right.

### Grey model

Grey system theory is an interdisciplinary scientific area that was introduced in early 80's by Deng [74]. Grey models require only a limited amount of data to estimate the behavior of unknown systems with its ability to deal with the systems that have partially unknown parameters. Grey models predict the future values of time series based only on a set of the most recent data depending on window size of the predictor. GM(1, 1) type of Grey model is the most widely used in the literature. The differential equations

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \quad (D.17)$$

have time varying coefficient  $(a(t), b(t))$ . Let us consider a time series sequence  $\mathbf{x}^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$ ,  $n \in \mathbb{N}$ , that denotes the close price of set whose accumulating operator (AGO) series is  $\mathbf{x}^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\}$ ,  $n \in \mathbb{N}$ , where

$$x^{(1)}(t) = \sum_{i=1}^t x^{(0)}(i). \quad (D.18)$$

In the above differential equation,  $[a \ b]^T$  is a sequence of parameters that can be found as

$$[a \ b]^T = (B^T B)^{-1} B^T Y \quad (\text{D.19})$$

where

$$Y = \{x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)\} \quad (\text{D.20})$$

and

$$B = \begin{bmatrix} -\frac{x^{(1)}(1)+x^{(1)}(2)}{2} & 1 \\ -\frac{x^{(1)}(2)+x^{(1)}(3)}{2} & 1 \\ -\frac{x^{(1)}(3)+x^{(1)}(4)}{2} & 1 \\ \vdots & \vdots \\ -\frac{x^{(1)}(n-1)+x^{(1)}(n)}{2} & 1 \end{bmatrix}. \quad (\text{D.21})$$

The solution of

$$\hat{x}^{(0)}(t+1) = \hat{x}^{(1)}(t+1) - \hat{x}^{(1)}(t) \quad (\text{D.22})$$

and

$$\hat{x}^{(1)}(t+1) = \left[ x^{(1)}(0) - \frac{b}{a} \right] e^{-at} + \frac{b}{a} \quad (\text{D.23})$$

since by definition of AGO  $x^{(1)}(0) = x^{(0)}(1)$ . Therefore

$$\hat{x}^{(0)}(t+1) = \left[ x^{(0)}(1) - \frac{b}{a} \right] e^{-at} (1 - e^a) \quad (\text{D.24})$$

and the predicted value of the primitive data at time  $(t+H)$  is

$$\hat{x}^{(0)}(t+H) = \left[ x^{(0)}(1) - \frac{b}{a} \right] e^{-a(t+H-1)} (1 - e^a). \quad (\text{D.25})$$

According the paper [75], the improved GM(1, 1) model was proposed, which has the architecture of GM-HHT. The equation

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \quad (\text{D.26})$$

can be obtained from theorem below.

**Theorem 2.** Let the original signal  $\mathbf{x}^{(0)} = \{x^{(0)}, \dots, x^{(n)}\}$ . The discretize of the first order derivative of GM(1, 1) can be obtained from

$$\mathbf{M}' = \begin{bmatrix} m_1 \\ m_1 \\ \vdots \\ m_m \end{bmatrix} = \begin{bmatrix} b - ax^{(1)}(1) \\ A^{-1}G \\ b - ax^{(1)}(n) \end{bmatrix} \quad (\text{D.27})$$

where  $m_i = \frac{dx^i}{dt}|_{t=i}$  and

$$A = \begin{bmatrix} 2 & 0.5 & 0 & \cdots & \cdots & 0 \\ 0.5 & 2 & 0.5 & \cdots & 0 & \vdots \\ \vdots & \cdots & \cdots & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & 0.5 & 2 & 0.5 \\ \vdots & \cdots & \cdots & 0 & 0.5 & 2 \end{bmatrix} \quad (\text{D.28})$$

$$G = \begin{bmatrix} g_2 - 0.5(b - ax^{(1)}(1)) \\ g_3 \\ g_4 \\ \vdots \\ g_{n-2} \\ g_{n-1} - 0.5(b - ax^{(1)}(n)) \end{bmatrix} \quad (\text{D.29})$$

where

$$g_j = 1.5(x^{(0)}(j) + x^{(0)}(j+1)), j = 2, 3, \dots, n. \quad (\text{D.30})$$

We can write the equation as

$$\mathbf{M}' + a\mathbf{x}^{(1)} = b. \quad (\text{D.31})$$

*Proof:* see [75].

The algorithm of a shifting process with endeffect solving with improving GM(1, 1) and ITD before sending to empirical mode decomposition (EMD) is shown in Fig. D.7.

### Performance test of directional prediction of SET index

The five days ahead forecast by using of improved GM(1, 1)-ITD-HHT-ANN is performed with 92 data set of out of sample test. The test data start from 2/3/2014 with data number 9528 to 24/6/2014 with data number 9622. 9622 means SET index of daily closed price of date number 9622 since market started. The result of calculation was used for the performance test of profit of short and long positions of IndexFutures Market. The result of the performance test of our prediction model is test over 60 sample data. The trading day start with data number 9528 to data number 9588. The graph of performance test is shown in Figs. D.8. For one day ahead forecast we can notice only which state is maximum state  $s_2$  in our time series data. Then we can open a short position from that state. We obtain average accuracy of this method by average all 5 days prediction in 92 days of out of sample test. We get average accuracy at 51.96% with standard deviation (SD) at 0.501430242.

In this work we also detected entanglement state of time series data after using (ITD – IMF) $\text{chain}_1$  transformation. We found that mostly the entanglement state occurred, when time series data is in the maximum state  $s_2$ .

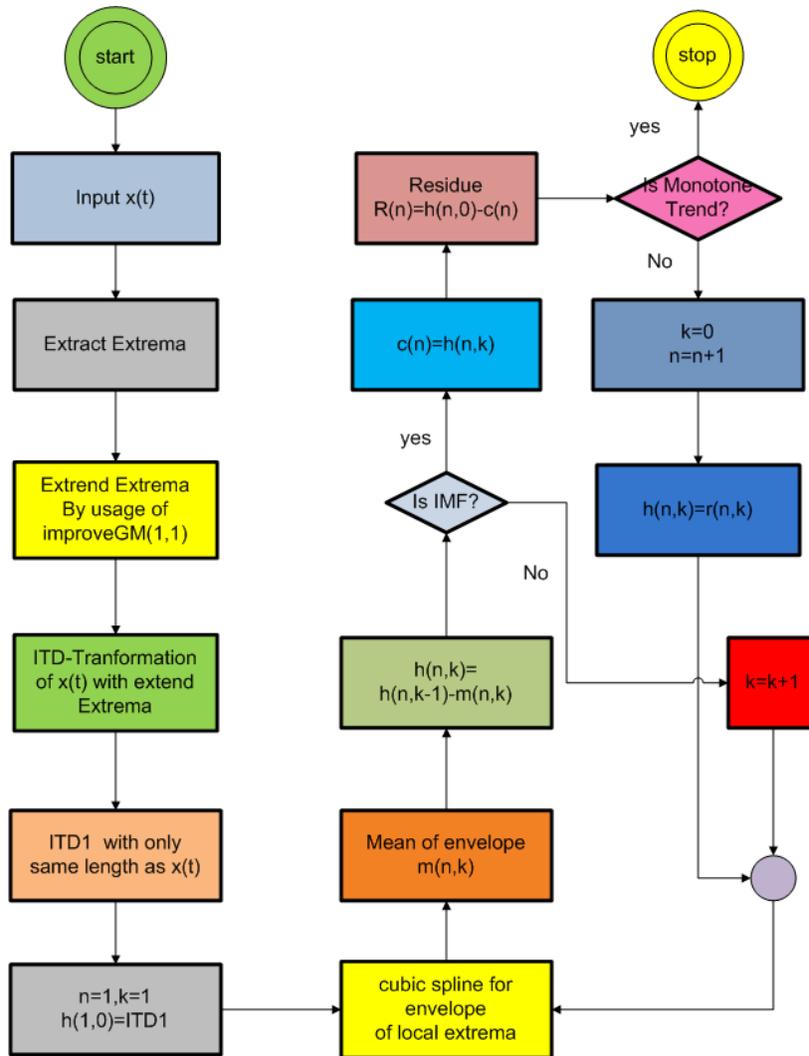


Fig. D.7: The flowchart of EMD algorithm with endeffect solving which improving GM(1, 1) and ITD. The algorithm predicts entanglement state of loopback between maximum and minimum state in time series.

The picture of down direction of stock index can be notice one day ahead in out of sample test within modeling of entanglement state in time series data of our empirical analysis is analogy with endeffect of our (ITD – IMF)chain<sub>1</sub>, shown in Fig. D.9.

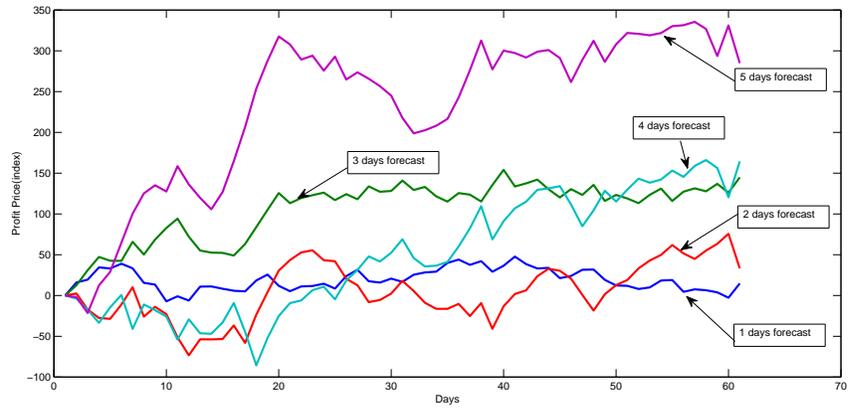


Fig. D.8: The performance plot of 5 days forecast with holding a position in five days as the best result. The second rank is 4 days forecast with holding a position within 4 days. The worst case is one day forecast.

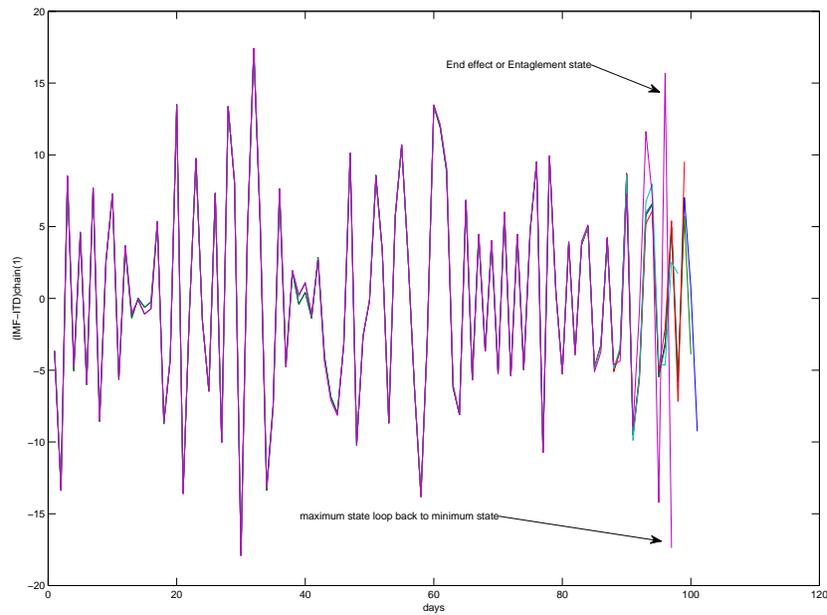


Fig. D.9: The end effect with entanglement state of loopback between maximum and minimum states. It can be used to forecast the maximum state one day ahead.

## References

- [1] B.L. Roux and H. Rouanet. *Geometric Data Analysis: From Correspondence Analysis to Structured Data Analysis*. Springer Netherlands, 2004.
- [2] J. Polchinski. *String Theory*. String Theory 2 Volume Hardback Set. Cambridge University Press, 2001.
- [3] B. Zwiebach. *A first course in string theory*. A First Course in String Theory. Cambridge University Press, 2004.
- [4] R.N. Mantegna and H.E. Stanley. Scaling behaviour in the dynamics of an economic index. *Nature*, 376(6535):46–49, 1995.
- [5] P. F. Verdes, P. M. Granitto, H. D. Navone, and H. A. Ceccatto. Nonstationary time-series analysis: Accurate reconstruction of driving forces. *Phys. Rev. Lett.*, 87:124101, 2001.
- [6] Guillermo L. Grinblat, Lucas C. Uzal, Pablo F. Verdes, and Pablo M. Granitto. Nonstationary regression with support vector machines. *Neural Computing and Applications*, 26(3):641–649, 2015.
- [7] Thomas Lux and Michele Marchesi. Volatility clustering in financial markets: a microsimulation of interacting agents. *International Journal of Theoretical and Applied Finance*, 03(04):675–702, 2000.
- [8] Kabin Kanjamapornkul and Richard Pinčák. Kolmogorov space in time series data. *Mathematical Methods in the Applied Sciences*, pages 4463–4483, 2016.
- [9] Fengzhong Wang, Shwu-Jane Shieh, Shlomo Havlin, and H. Eugene Stanley. Statistical analysis of the overnight and daytime return. *Phys. Rev. E*, 79:056109, 2009.
- [10] Fengzhong Wang, Kazuko Yamasaki, Shlomo Havlin, and H. Eugene Stanley. Multifactor analysis of multiscaling in volatility return intervals. *Phys. Rev. E*, 79:016103, 2009.
- [11] Armin Bunde, Jan F. Eichner, Jan W. Kantelhardt, and Shlomo Havlin. Long-term memory: A natural mechanism for the clustering of extreme events and anomalous residual times in climate records. *Phys. Rev. Lett.*, 94:048701, 2005.
- [12] P. Hartmann, S. Straetmans, and C.G. De Vries. Asset market linkages in crisis periods. *Review of Economics and Statistics*, 86(1):313–326, 2004.
- [13] Mikhail I. Bogachev and Armin Bunde. Memory effects in the statistics of interoccurrence times between large returns in financial records. *Phys. Rev. E*, 78:036114, 2008.
- [14] L. Bai, S. Yan, X. Zheng, and B.M. Chen. Market turning points forecasting using wavelet analysis. *Physica A*, 437:184–197, 2015.
- [15] S. Lahmiri. Intraday stock price forecasting based on variational mode decomposition. *Journal of Computational Science*, 12:23–27, 2016.
- [16] M.M. Rounaghi, M.R. Abbaszadeh, and M. Arashi. Stock price forecasting for companies listed on tehran stock exchange using multivariate adaptive regression splines model and semi-parametric splines technique. *Physica A: Statistical Mechanics and its Applications*, 438:625–633, 2015.
- [17] S. Lahmiri. Interest rate next-day variation prediction based on hybrid feedforward neural network, particle swarm optimization, and multiresolution techniques. *Physica A: Statistical Mechanics and its Applications*, 444:388–396, 2016.
- [18] M.A.L. Caetano and T. Yoneyama. An autocatalytic network model for stock markets. *Physica A: Statistical Mechanics and its Applications*, 419:122–127, 2015.
- [19] F. Chen, C. Gou, X. Guo, and J. Gao. Prediction of stock markets by the evolutionary mix-game model. *Physica A: Statistical Mechanics and its Applications*, 387:3594–3604, 2008.
- [20] Jose C. Nacher and Tomoshiro Ochiai. Foreign exchange market data analysis reveals statistical features that predict price movement acceleration. *Phys. Rev. E*, 85:056118, 2012.

- [21] Mehrnaz Anvari, Cina Aghamohammadi, H. Dashti-Naserabadi, E. Salehi, E. Behjat, M. Qorbani, M. Khazaei Nezhad, M. Zirak, Ali Hadjihosseini, Joachim Peinke, and M. Reza Rahimi Tabar. Stochastic nature of series of waiting times. *Phys. Rev. E*, 87:062139, 2013.
- [22] Marco Zamparo, Fulvio Baldovin, Michele Caraglio, and Attilio L. Stella. Scaling symmetry, renormalization, and time series modeling: The case of financial assets dynamics. *Phys. Rev. E*, 88:062808, 2013.
- [23] Rajkumar Venkatesan and V. Kumar. A genetic algorithms approach to growth phase forecasting of wireless subscribers. *International Journal of Forecasting*, 18(4):625 – 646, 2002.
- [24] Denis Horváth and Richard Pinčák. From the currency rate quotations onto strings and brane world scenarios. *Physica*, A391:5172–5188, 2012.
- [25] Richard Pinčák. The string prediction models as invariants of time series in the forex market. *Physica*, A392:6414–6426, 2013.
- [26] Richard Pinčák. With strings toward safety future on financial markets. In Prof. Mohsen Bahmani-Oskoev, editor, *Financial Markets: Recent Developments, Emerging Practices and Future prospects*. NOVA Science Publisher, New York, 2014.
- [27] Marek Bundzel, Tomáš Kasanický, and Richard Pinčák. Experimental analysis of the prediction model based on string invariants. *Computing and Informatics*, 32(6):1131–1146, 2013.
- [28] Marek Bundzel, Tomáš Kasanický, and Richard Pinčák. Using string invariants for prediction searching for optimal parameters. *Physica*, A444:680–688, 2016.
- [29] Richard Pinčák and Erik Bartoš. With string model to time series forecasting. *Physica*, A436:135–146, 2015.
- [30] K. Kanjamapornkul, R. Pinčák, and E. Bartoš. The study of Thai stock market across the 2008 financial crisis. *Physica*, A462:117–133, 2016.
- [31] Fengzhong Wang, Kazuko Yamasaki, Shlomo Havlin, and H. Eugene Stanley. Scaling and memory of intraday volatility return intervals in stock markets. *Phys. Rev. E*, 73:026117, 2006.
- [32] Giampiero M. Gallo and Edoardo Otranto. Forecasting realized volatility with changing average levels. *International Journal of Forecasting*, 31(3):620 – 634, 2015.
- [33] Erik Bartoš and Richard Pinčák. Identification of market trends with string and D2-brane maps. *Physica*, A479:57–70, 2017.
- [34] P. Rinn, Y. Stepanov, J. Peinke, T. Guhr, and R. Schäfer. Dynamics of quasi-stationary systems: Finance as an example. *EPL*, 110(6), 2015.
- [35] George Edward Pelham Box and Gwilym Jenkins. *Time Series Analysis, Forecasting and Control*. Holden-Day, Incorporated, 1990.
- [36] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307 – 327, 1986.
- [37] J.-C. Hung. Robust kalman filter based on a fuzzy garch model to forecast volatility using particle swarm optimization. *Soft Computing*, 19(10):2861–2869, 2015.
- [38] L. Xi, H. Muzhou, M.H. Lee, J. Li, D. Wei, H. Hai, and Y. Wu. A new constructive neural network method for noise processing and its application on stock market prediction. *Applied Soft Computing Journal*, 15:57–66, 2014.
- [39] J.L. Ticknor. A bayesian regularized artificial neural network for stock market forecasting. *Expert Systems with Applications*, 40(14):5501–5506, 2013.
- [40] C. Lönnbark. Asymmetry with respect to the memory in stock market volatilities. *Empirical Economics*, 2015.
- [41] Michael C. Münnix, Takashi Shimada, Rudi Schäfer, Francois Leyvraz, Thomas H. Seligman, Thomas Guhr, and H. Eugene Stanley. Identifying states of a financial market. *Nature Scientific Reports*, 2:644, 2012.

- [42] K. Kanjamapornkul, R. Pinčák, and E. Bartoš. Support Spinor Machine. *Digital Signal Processing*, accepted, 2017.
- [43] M. B. Green, J. H. Schwarz, and E.d Witten. *Superstring theory. Vol. 1: Introduction*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 1988.
- [44] José Fajardo and Ernesto Mordecki. Symmetry and duality in lévy markets. *Quantitative finance*, 6(3):219–227, 2006.
- [45] W. H. Wagner and M. Edwards. Best execution. *Financial Analysts Journal*, 49(1):65–71, 1993.
- [46] R. Roll. A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, 39:1127–1139, 1984.
- [47] I. Giampaoli, W. L. Ng, and N. Constantinou. Analysis of ultra-high-frequency financial data using advanced Fourier transforms. *Finance Research Letters*, 6:47–53, 2009.
- [48] P. O. Kazinski. Effective dynamics of an electrically charged string with a current. *Zh. Eksp. Teor. Fiz.*, 128:312–321, 2005. [J. Exp. Theor. Phys.101,270(2005)].
- [49] Sitabhra Sinha and Srinivas Raghavendra. Market polarization in presence of individual choice volatility. In *Advances in Artificial Economics*, pages 177–190. Springer Berlin Heidelberg, 2006.
- [50] Peter Grassberger and Itamar Procaccia. Measuring the strangeness of strange attractors. *Physica D: Nonlinear Phenomena*, 9(1):189 – 208, 1983.
- [51] John Y. Campbell, Andrew W. Lo, A. Craig MacKinlay, Petr Adamek, and Luis M. Viceira. *The Econometrics of Financial Markets*. Princeton University Press, Princeton, NJ, 1997. Attachment-Solution manual for exercises.
- [52] N. Bigdeli and K. Afshar. Characterization of Iran electricity market indices with pay-as-bid payment mechanism. *Physica A: Statistical Mechanics and its Applications*, 388(8):1577 – 1592, 2009.
- [53] Justin Wolfers and Eric Zitzewitz. Prediction markets. *Journal of Economic Perspectives*, 18(2):107–126, 2004.
- [54] J. D. Christiansen. Prediction markets: practical experiments in small markets and behaviours observed. *The Journal of Prediction Markets*, 1:17–41, 2007.
- [55] Ching-Chih Chang, Chin-Yuan Hsieh, and Yung-Chih Lin. A predictive model of the freight rate of the international market in Capesize dry bulk carriers. *Applied Economics Letters*, 19(4):313–317, 2012.
- [56] M. M. Dacorogna, R. Gencay, U. Muller, R. B. Olsen, and O. V. Pictet. *An introduction to high frequency finance*. Academic Press, New York, 2001.
- [57] G. Appel. *Technical Analysis: Power Tools for Active Investors*. FT Press, 2005.
- [58] Guillaume Weisang and Yukika Awazu. Vagaries of the euro: an introduction to arima modeling. *Case Studies In Business, Industry And Government Statistics*, 2(1), 2008.
- [59] IB. Interactive Brokers LLC, 2015.
- [60] LMAX Exchange. London Multi Asset Exchange, 2015.
- [61] Librade LTD. The online trade system, 2015.
- [62] C.V. Johnson. *D-Branes*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2003.
- [63] David Ben-Zvi and David Nadler. *Loop Spaces and Connections*, 2011.
- [64] Zbigniew Karno. On Kolmogorov Topological Space. *Journal of Formalized Mathematics*, 6(1):1–5, 1994.
- [65] K. Jänich and S. Levy. *Topology*. Undergraduate Texts in Mathematics. Springer New York, 1995.
- [66] Klaus Reiner Schenk-Hoppé and Björn Schmalfuß. Random fixed points in a stochastic solow growth model. *Journal of Mathematical Economics*, 36(1):19–30, 2001.

- [67] Markus Szymik. Homotopies and the universal fixed point property, 2013.
- [68] Bartosz Jablonski. Quaternion dynamic time warping. *IEEE Transactions on Signal Processing*, 60(3):1174–1183, 2012.
- [69] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences*, 454:903–995, 1998.
- [70] Samuel Dooley. Basic algebraic topology: The fundamental group of circle. preprint, University of Chicago, 2011.
- [71] Mark G Frei and Ivan Osorio. Intrinsic time-scale decomposition: time-frequency-energy analysis and real-time filtering of non-stationary signals. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 463(2078):321–342, 2007.
- [72] Mengjia Xu, Pengjian Shang, and Aijing Lin. Cross-correlation analysis of stock markets using emd and eemd. *Physica A*, 442:82–90, 2016.
- [73] Z. Wu and N. E. Huang. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv. Adapt. Data Anal.*, 1(1):1–41, 2009.
- [74] J.L. Deng. Introduction to grey system theory. *The Journal of Grey System*, 1(1):1–24, 1989.
- [75] Zhi He, Yi Shen, and Qiang Wang. Boundary extension for hilbert–huang transform inspired by gray prediction model. *Signal Processing*, 92(3):685 – 697, 2012.



**Erik Bartoš** born in Ružomberok, Slovakia, in 1976, received his Ph.D. in theoretical physics from the Comenius University in Bratislava in 2005. He has long been concerned with the problems of phenomenological models of nucleons and light mesons. In addition, he is experienced with study of subnuclear and heavy ion collisions processes gained during long-term and short-term visits at prominent international centers, for example JINR Dubna (Russia), Hamburg DESY (Germany), CERN (Switzerland). At present, he is a researcher at Institute of Physics, Slovak Academy of Sciences, Department of Theoretical Physics. In the last years he has been involved in the approach of the string theory to the field of time series forecast and data analysis. He is actively participating in several national and other research projects and he is also co-organizer of the International Conferences “Hadron Structure” and co-editor of their proceedings.



**Richard Pinčák** works at the Institute of Experimental Physics, Slovak Academy of Sciences; he was several years senior researcher in Bogoliubov Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, Dubna, Russia. He is an active theoretical physicist with a lot of branches of research as condensed matter physics, biophysics and also cosmology. The last time he is also trying to find the fundamental revolution principles of the behavior prediction of financial forex market and options pricing. He studies the projections of the real exchange rate dynamics onto the string-like topology. His approach in this field is inspired by the contemporary movements in the string theory.