Journal of
Risk and Financial Management

MDPI

Article

Modeling NYSE Composite US 100 Index with a Hybrid SOM and MLP-BP Neural Model

Adriano Beluco [1], Denise L. Bandeira [2] and Alexandre Beluco [3,*]

[1] Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS), Campus Viamão, Av Sen Salgado Filho, 7000, Bairro São Lucas, 94440-000 Viamão, RS, Brazil; adbeluco@gmail.com
[2] Escola de Administração, Universidade Federal do Rio Grande do Sul (UFRGS), Rua Washington Luiz, 855, Centro Histórico, 90010-460 Porto Alegre, RS, Brazil; dlbandeira@ufrgs.br
[3] Instituto de Pesquisas Hidráulicas (IPH), Universidade Federal do Rio Grande do Sul (UFRGS), Av Bento Gonçalves, 9500, Bairro Agronomia, 91501-970 Porto Alegre, RS, Brazil
* Correspondence: albeluco@iph.ufrgs.br; Tel.: +55-51-99956-7314

Academic Editor: Michael McAleer
Received: 22 August 2016; Accepted: 19 January 2017; Published: 5 February 2017

Abstract: Neural networks are well suited to predict future results of time series for various data types. This paper proposes a hybrid neural network model to describe the results of the database of the New York Stock Exchange (NYSE). This hybrid model brings together a self organizing map (SOM) with a multilayer perceptron with back propagation algorithm (MLP-BP). The SOM aims to segment the database into different clusters, where the differences between them are highlighted. The MLP-BP is used to construct a descriptive mathematical model that describes the relationship between the indicators and the closing value of each cluster. The model was developed from a database consisting of the NYSE Composite US 100 Index over the period of 2 April 2004 to 31 December 2015. As input variables for neural networks, ten technical financial indicators were used. The model results were fairly accurate, with a mean absolute percentage error varying between 0.16% and 0.38%.

Keywords: modeling financial indicators; NYSE indexes; self organizing maps; multilayer perceptron; back propagation algorithm; software Matlab

PACS: JEL-C53; JEL-E37

1. Introduction

The prediction of future values of time series has been an ongoing challenge for professionals involved in various fields of engineering and management. Among mathematical tools available, neural networks have shown characteristics of flexibility in modeling and fast response that allow an incredible balance between costs and benefits of their use [1].

Neural networks have been increasingly used in financial areas because of their characteristics of flexibility and responsiveness. The applications are quite diverse and include risk classification of investments, fixed and variable income investments, simulation of markets, selection and portfolio diversification, and economic forecasting [2].

Recently, considering the results from neural and stochastic models for time series forecasting, researchers have been dedicated to the composition of models with different combinations such as genetic algorithms and neural networks [3], and, specifically, self organizing maps (SOM) [4]; SOM and general regression neural networks, multilayer perceptron (MLP) and generalized autoregressive conditional heteroskedastic [5], among others [6].

Neural networks represent a concept of processing system where the models are based on neurophysiological processing principles. The brain is composed of differentiated cells called neurons,

which have a cell body (soma) where most of their organelles can be found. The soma of individual neurons extend axons and dendrites (inputs and outputs) and each neuron receives electrical impulses by their dendrites, which are processed in the soma and transmitted via axons to the dendrites of other neurons. The connections between neurons are defined as synapses [1], which are basic functional units for the formation of biological neural circuits.

In recent decades, neural networks have been applied to solve problems in many different areas, but most of these applications focuses on the use of a single network design, such as back propagation for time series modeling. Each model of an artificial neural network has specific characteristics that respond more appropriately to a given class of problem. Each design of a neural network seems to have a vocation to solve a particular problem. This feature seems to encourage a growing coverage of neural networks in the description of time series.

This article shows the time series modeling of the New York Stock Exchange (NYSE) indexes based on a hybrid model designed with the use of an SOM network followed by the use of an MLP-BP (multilayer perceptron with back propagation algorithm) network. It is an arrangement not yet adopted to analyze historical data or to predict future values of economic data. An SOM is quite suitable for the classification of the components of a time series into clusters, grouped according to common characteristics. An MLP is suitable for the identification of these features, even if they are very complex, intending a model to forecast future values of the series. The BP algorithm ensures better performance for this hybrid model.

The prediction of values of the stock exchange indexes is a next step to this study, and may involve other variables and should be able to predict parameters such as the direction of daily variation of the stock market, among others. The technique presented in this paper can also obviously be implemented to describe other notable time series, such as the historical series of oil barrel prices. The difficulty will be the identification of parameters that can be used as indicators of a trend of variation of the values of the series under study.

2. The NYSE Indexes

The data used in this article consist of daily rates of the index NYSE Composite US 100 in the period from 5 April 2004 to 31 December 2015. These data were obtained using the system Economatica (BSI Tecnologia, São Paulo, Brazil, 2015) and quotations provided by the NYSE [7]. The variables used for the database involve daily trading values, considering the range of daily fluctuation, so that the variables in the database include values of opening and closing, minimum and maximum values and daily trading volume.

Figure 1 shows the evolution over time of the daily trading volume, Figure 2 shows the values of opening and closing and Figure 3 shows the daily minimum and maximum values. These three figures show the period from 5 April 5 2004 to 31 December 2015. Figures 2 and 3 clearly show the consequences of the 2009 crisis on the data presented.

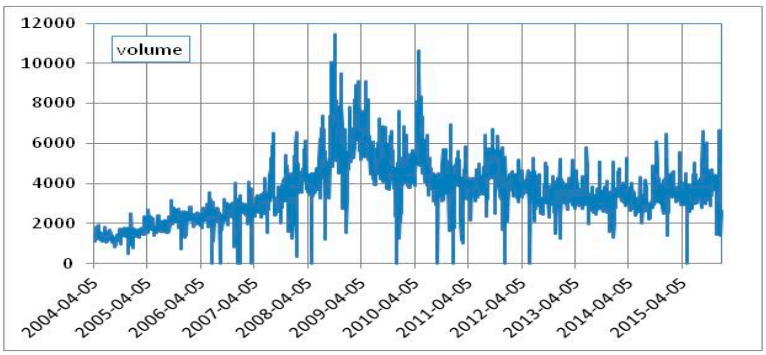

Figure 1. Evolution over time of the daily trading volume of the NYSE (New York Stock Exchange) Composite US 100 between 5 April 2004 and 31 December 2015.

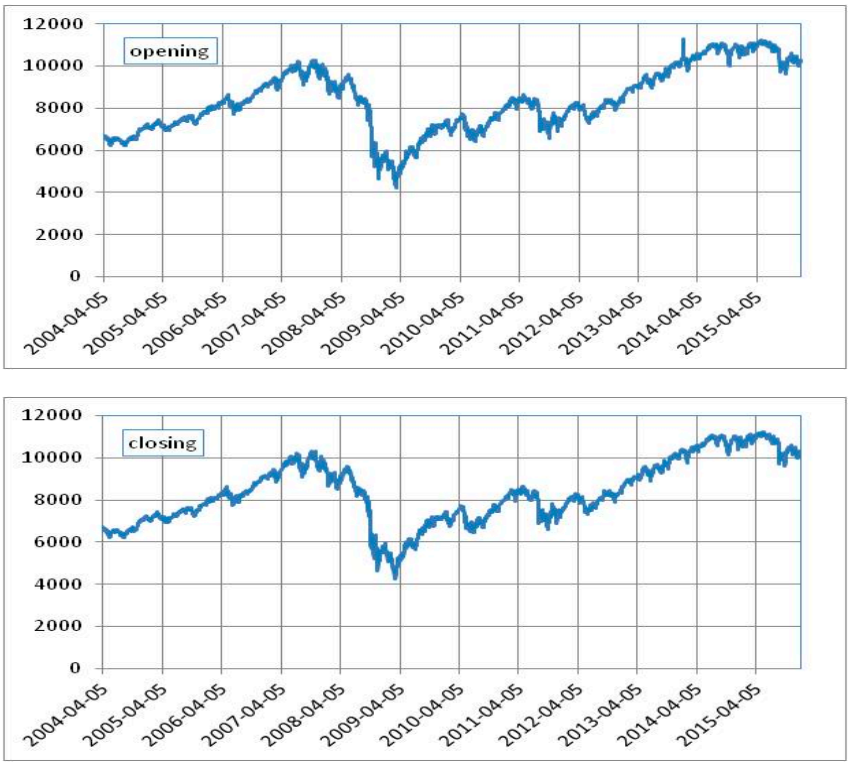

Figure 2. Evolution over time of the values of opening (above) and closing (below) of the NYSE Composite US 100 between 5 April 2004 and 31 December 2015.

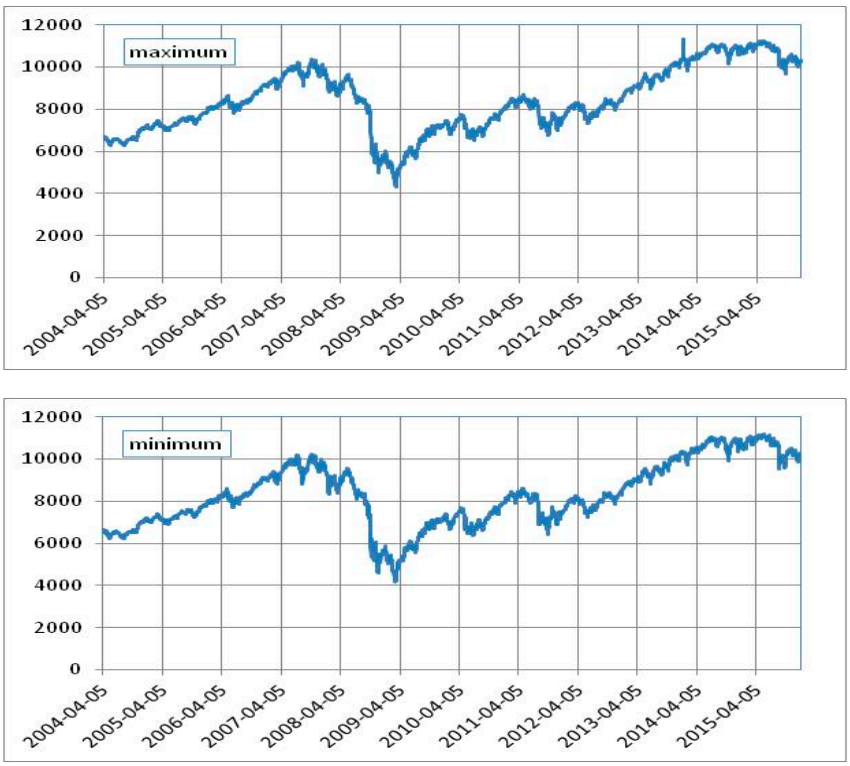

Figure 3. Evolution over time of the values of daily maximum (above) and minimum (below) of the NYSE Composite US 100 between 5 April 2004 and 31 December 2015.

The New York Stock Exchange established the NYSE Composite Index in 1966 to provide a comprehensive measure of the performance of all listed common shares. The NYSE Composite represents 77% of the total market capitalization of all publicly traded companies in the United States and 66% of the total market capitalization of all publicly traded companies in the world [8].

The NYSE Composite US 100 depicts the capitalization of the hundred largest companies in the index. It was developed to measure the performance of all common shares, consisting of more than two thousand North American and foreign stock exchanges. It is a measure of changes in the market value, adjusted to eliminate the effects of changes in capitalization.

For these reasons, four new indexes to help investors were launched in June 2002: Composite NYSE US 100, NYSE International 100, NYSE TMT and the NYSE World Leaders. In 2004, the index NYSE Financial, Energy and Healthcare was created. With broad participation of the largest companies in the world, the NYSE is the market reference for the positioning of global investors.

3. Self Organizing Maps

The SOM are neural networks for the specific purpose of grouping similar data together so as to form clusters. An SOM consists basically of an input and an output layer, represented by a grid of postsynaptic uni- or two-dimensional characteristics. The output layer is formed by a network of neurons connected to the neurons closest to them, where each neuron is a cluster of the mesh. However, the neurons forming the input layer are connected to all postsynaptic neurons.

An SOM network is a class of neural networks whose learning is so unsupervised. This class is also known as a Kohonen network [9]. The basic neurobiological motivation for mapping models of features is characterized by the ability of compressing the input data. The SOM network then becomes an input pattern topologically ordered in a uni- or two-dimensional discrete map [1].

The biological basis of a self-organizing map is based on the principle of sorted mapping of the cerebral cortex from sensory inputs [1]. The principle of formation of topographic maps through the correspondence between the spatial location of an output neuron in a topographic map and a specific feature of information taken from the input space was formulated in [10].

The first step of the algorithm is the random initialization of the synaptic weights of the grid. The random characteristic of this step ensures the absence of a preliminary organization of the map. The formation of the map occurs in three stages: competition, cooperation and adaptation.

At the competition stage, the values of a discriminant function corresponding to each input pattern are determined. The neuron that has the highest score for the discriminant function is called the winner neuron. Each input vector space of input data is subjected to a discriminant function that will be responsible for constituting the basis for competitiveness between neurons, where a neuron that receives the largest value of the discriminant function will be designated as the winner.

The cooperation process specifies the spatial location of the topological neighborhood of neurons in an excited state from the neuron considered the winner. The grid synaptic response may indicate the position of the winning neuron or synaptic weight vector closest to the input vector, considering the Euclidean distance. This step is based on neurobiological evidence of lateral interaction between excited neurons. Thus, it is clear that the topological neighborhood around a winner neuron is reduced as a lateral distance [11].

The stage of adapting enables stimulated neurons around the winner neuron increase the results of the discriminant function from the input patterns through adjustments in their synaptic weights [11].

The preview of the learning process of the SOM is required for the verification of the result of their topological sorting. The result can be viewed as unified distance matrix, known as U-Matrix. This matrix facilitates the visualization process, and can be represented by an image. The U-Matrix can be seen as an image in which the color of the pixels occurs according to the intensity of each component of the matrix. Thus, higher values correspond to neighboring neurons not similar and lower values correspond to similar neighboring neurons. Despite the U-Matrix generating a complex image,

it allows for visualization of the separation of topological groups. This representation is extremely useful when the dimensionality is greater than 3.

4. Multilayer Perceptrons with a Back Propagation Algorithm

Neural networks are models that make use of the connectionist paradigm, which seeks to understand and emulate the properties resulting from the high degree of parallelism and connectivity in solving certain kinds of problems. Thus, a network is constituted by a large number of processing elements, largely interconnected. Each link connects two processing elements in a single direction through a weight that determines the degree of connectivity between elements.

Processing is distributed across all network elements, each of which performs its function in an isolated way and in parallel, sending its result to the other units of the next layer through their connections. Each processor element normally has several inputs and one output. Their processing consists of transferring to output a value calculated from the values given in the entries by means of a transfer function.

Typically, the inputs are combined via a weighted sum being transferred to the output through a threshold function [12]. In addition to the threshold function, it is possible to make use of the sigmoid function or hyperbolic tangent function. The activation potential of an element is the value of output in a given time. The set of states of activation of each of the processing elements is defined as the activation function or activation state of the neural network.

The network structure is determined by the shape of the collation processing elements in layers. In general, the structure consists of an input layer, where they are simple presentation of data to the network, one or more intermediate layers, widely connected to the output layer, responsible for obtaining the results. The layers are processed in order of the input layer to the output layer, and there are no connections between components of the same layer.

The learning methodologies allow the modification of the standard interconnection of a neural network, enabling it to solve a given problem. Three mechanisms are generally used for learning: supervised when they are given the desired results, for reinforcement, an external parameter when a comparison is made, and unsupervised when the network itself is apt to adjust their operation.

In general, three phases are used for adapting a neural network to a problem. First, a training phase, which teaches the network to model a set of output patterns associated with input patterns. In a second stage, input patterns are presented to the network and the outputs are compared to desired outputs. In the final phase, the network is used to implement the solution.

A concept widely used by the MLP is the BP of error. In the propagation step, the synaptic weights that indicate the importance of the neuron in the final result are fixed and determined stochastically. The BP learning algorithm has become the standard for use in neural networks type multilayer perceptron as it concisely addresses the problem of assigning credit for back propagation through correction of errors in the results.

The topology of the MLP has characteristics extremely plastic for the design of the structure in relation to the number of artificial neurons to be used in each layer or even the number of intermediate layers to consider. In fact, there are no studies that specify the number of neurons per layer or number of layers to be used. The biological flexibility adopted by the human brain to solve problems in this regard is of great similarity.

The training algorithm has the following steps: initialization of weights and network parameters (coefficient of learning and parameter time); calculation of the potential of activation of neurons in the hidden layer; calculation of outputs of neurons in the hidden layer; calculation of the potential of activation for the neurons of the output layer; calculation of activation of the output neurons similar to the intermediate layer; calculation of the error terms (gradient location of the error); calculation of the error terms for intermediate units; update the weights in the output layer; update the weights in the hidden layer; repetition of steps to all standards.

For performance analysis, it is common to use the root mean square error (RMSE), the mean absolute error (MAE) and the mean absolute percentage error (MAPE). The RMSE is defined by Equation (1), below, where Y_p is an actual component of the cluster p, O_p is a result of the model for the cluster p and n is the number of components of the cluster p:

$$RMSE_p = \sqrt{\frac{1}{n}\sum_{k=1}^{M}\left(Y_{pk} - O_{pk}\right)^2}. \quad (1)$$

The MAE is defined by Equation (2). The MAE is a relatively common measure of the forecast error for time series analysis

$$MAE_p = \frac{1}{n}\sum_{k=1}^{M}\left|Y_{pk} - O_{pk}\right|. \quad (2)$$

The MAPE is defined by Equation (3). The MAPE is a relative measure for the error between the predicted value and the actual value

$$MAPE_p = \frac{1}{n}\sum_{k=1}^{M}\left|\frac{Y_{pk} - O_{pk}}{Y_{pk}}\right|. \quad (3)$$

These evaluation measures are typical in efficiency analysis of neural networks [1].

5. A Hybrid Model

The methods used in this study were based on the Operational Research and forecasting models using neural networks. Tasks based on Operational Research should delimit five main stages, which are (i) the problem statement, (ii) model building, (iii) model solution, (iv) model validation and (v) analysis of results. Obviously, these steps are not rigid and can be rearranged according to the characteristics of the problem to be analyzed.

Neural networks have the behavior of a black box, since it does not have access to specific details about its operation. Thus, the understanding of the process occurs through the use of input stimuli and analysis of outputs. References [4,13] list some of the key aspects in the modeling, which are (i) selection of the input variables, (ii) determining the amount of input variables, (iii) definition of the network topology, (iv) specification of the training algorithm and (v) prediction of the network output.

However, the study by Setyawati et al. [14] calls attention to the lack of rules for defining the topology of neural networks. There is no evidence to prove the prevalence of hexagonal architecture on the rectangular SOM networks, or even an inverse relationship. Likewise, it still prevails the questioning regarding the number of neurons in the hidden layer of a back propagation network or even the number of hidden layers. These issues may be subject to specific studies.

As the above assumptions, the scheme proposed in this study for the hybrid model to forecast financial indices consists of four phases. Each phase of the general layout of the hybrid model has several sub-stages of execution.

The first phase of the hybrid model is characterized by the collection of data and its preprocessing. After collecting the data, the calculation of financial indicators is made and the results are normalized to the interval $[-1, 1]$. The process of calculating the indexes and subsequent normalization are processed through the software MS Excel (Microsoft Corporation, Redmond, WA, USA) due to the size of the sample.

The second phase is characterized by clustering by means of the SOM. Initially, the segmentation of the database is done randomly in a training group (80% of the data), a test group (10% of the data) and a validation group (10% of the data). The random segmentation of networks for training is based on work published by several researchers [4,15–17].

After clustering, the normalized value of the variable closure is used for training the MLP-BP in each of the clusters. Finishing the second step, the insertion of the test group for a preliminary analysis of the results obtained by neural models is conducted.

The software Matlab (MathWorks, Natick, MA, USA, 2016) was used in both steps. First, the SOM toolbox developed by the Laboratory of Computer and Information Science (Department of Information and Computer Science, Helsinki University of Technology, Espoo, Finland). Then, the neural network toolbox, a Matlab (MathWorks) native tool, was used. SOM toolbox brings routines ready for use, but MLP analysis requires the user to build its own routines.

In the third phase, the model validation is performed. Selected models are optimized in each cluster to build the hybrid model, which will be inserted into the validation groups. The selection process is based on the results for the RMSE.

The fourth and final stage is devoted to the evaluation of the performance of the model built. The results generated by the model undergo reverse process of normalization to be compared with the data in the validation group. This comparison is performed with RMSE, MAE and MAPE.

6. Implementation of the Hybrid Model

The model is implemented through the stages of pre-processing of data, fitting of the model and its application. The preprocessing stage consists in arranging the data corresponding to the NYSE Composite US 100 in a manner suitable for evaluation by the hybrid model presented below.

The collected data consist of 2957 daily observations, from 5 April 2004 to 31 December 2015, with a total of five variables: opening and closing, as well as minimum and maximum value and the turnover of the index NYSE Composite US 100. The database collected is therefore organized in a matrix of order 2957×5 on the five variables of 2957 daily observations.

Input variables selected for the hybrid model predictive financial ratios derived from technical indices used by several authors [4,13,18,19] to evaluate the performance of stock exchange. There are other means to evaluate the behavior of the stock exchange, as discussed, for example, by Sandoval et al. [20], and this issue may also be the subject of a specific study.

The definitive database is organized in a matrix of order 2080×10, ten referring to the technical financial ratios calculated based on five variables of 2957 daily observations. Due to some of these indices that consider prior periods up to 26 days, values of all indexes in the period between 5 April 2004 and 10 May 2004 were discarded in order to minimize noise in the training of neural networks.

These ten technical financial ratios will be calculated from TV_i, the trading volume on day 1; OP_i and CP_i, values of opening and closing by day 1; HP_i and LP_i the maximum and minimum values of the NYSE Composite US 100 index on day 1.

The first of these ratios is MA 10, the moving average of the closing value of the NYSE index on the previous ten days, defined by Equation (4)

$$\text{MA } 10_i = \frac{\sum_{i-9}^{i} CP_i}{10}. \tag{4}$$

The index BIAS 20 calculates the difference between the closing value and the moving average of the closing value of the last 20 days, calculated using Equation (5). This index uses the index above, but is calculated for a twenty-day base

$$\text{BIAS } 20_i = \frac{CP_i - \text{MA } 20_i}{\text{MA } 20_i}. \tag{5}$$

The index of overbought and oversold of Williams is a momentum indicator based on the relationship between the difference of the maximum value and the closing value of the difference between the maximum and minimum values within the last nine days, calculated by Equation (6).

In this equation, HP 9 and LP 9 correspond respectively to the maximum and minimum values of the last nine days

$$\text{WMS\% R9}_i = \frac{\text{HP 9}_i - \text{CP}_i}{\text{HP 9}_i - \text{LP 9}_i}. \tag{6}$$

The stochastic index K for the last nine days, K 9, is defined by Equation (7), where HP 9 represents the maximum value within the last nine days and LP 9 the minimum value within the last nine days

$$\text{K 9}_i = \frac{2}{3}\,\text{K 9}_{i-1} + \frac{1}{3} \times 100 \times \frac{\text{CP}_i - \text{LP 9}_i}{\text{HP 9}_i - \text{LP 9}_i}. \tag{7}$$

A stochastic index related to the last nine days, D 9, may be calculated by Equation (8), which uses K 9 defined by the previous equation

$$\text{D 9}_i = \frac{2}{3}\,\text{D 9}_{i-1} + \frac{1}{3}\,\text{K 9}_i. \tag{8}$$

The MTM 10 index is an index that measures the time of changes in the value of closing within the last 10 days. It is calculated by Equation (9)

$$\text{MTM 10}_i = \text{CP}_i - \text{CP}_{i-10}. \tag{9}$$

The ROC 10 index represents a rate of change that measures the percentage changes between the current closing value and the closing value of 10 days, calculated by Equation (10)

$$\text{ROC 10}_i = \frac{\text{CP}_i - \text{CP}_{i-10}}{\text{CP}_{i-10}} \times 100. \tag{10}$$

The index called the Commodity Channel Index is used to identify cycles in the closing value of the commodities. It is calculated by Equation (11), which also states the variables determined by Equations (12)–(14)

$$\text{CCI 24}_i = \frac{\text{TP}_i - \text{SMATP 24}_i}{0,015 \times \text{MD 24}_I}. \tag{11}$$

A typical value is calculated by averaging the maximum, minimum and closing values, as indicated by Equation (12). Then, an average value of the last 24 days of this typical value is calculated by Equation (13) and the average deviation of the mean, calculated by Equation (14), is determined

$$\text{TP}_i = \frac{\text{HP}_i + \text{LP}_i + \text{CP}_I}{3}, \tag{12}$$

$$\text{SMATP 24}_i = \frac{\sum_{J=I-23}^{I} \text{TP}_J}{24}, \tag{13}$$

$$\text{MD 24}_i = \frac{\sum_{J=I-23}^{I} |\text{TP}_J - \text{SMATP 24}_i|}{24}. \tag{14}$$

The AR 26 is an index that tells the right time to buy and sell within the last 26 days, calculated by Equation (15). The BR 26 is an index that seeks to show the tendency to purchase and sale within the last 26 days, calculated by Equation (16)

$$\text{AR 26}_i = \frac{\sum_{j=i-25}^{i} \text{HP}_j - \text{OP}_j}{\sum_{j=i-25}^{i} \text{OP}_j - \text{LP}_j}, \tag{15}$$

$$\text{BR 26}_i = \frac{\sum_{j=i-25}^{i} \text{HP}_j - \text{CP}_{j-1}}{\sum_{j=i-25}^{i} \text{CP}_{j-1} - \text{LP}_j}. \tag{16}$$

Finally, these ten technical indices should then be normalized to the interval $[-1, 1]$.

The next step corresponds to the model fit. The use of SOM network aims to segment the database into multiple clusters, which have highlighted its features. The MLP network is used to build a predictive model for each cluster based on these ten financial indicators.

The SOM network consists of 10 neurons in an input layer that receives normalized financial ratios and several neurons arranged in a two-dimensional intermediate layer. The initial weight of each neuron is determined randomly and the total number of iterations in the learning stage is 62,400, about 30 times the number of records in the database [11]. The initial learning parameter is 0.06, decreasing gradually to 0.03 after 31,156 iterations of learning, reaching a value of 0.01 after 52,640 iterations of learning were created.

Figure 4 shows the U-Matrix and plans for each financial index. Seven clusters were identified, as can be seen in the U-Matrix, where the map of synaptic weights were generated randomly in the dimensions 25 × 9. The learning process used in this experiment is sequential. The figure on the bottom right is reproduced in Figure 5 and shows the result of applying the SOM network. Obviously, the identification of clusters after training the SOM network appears abbreviated and refers to the financial ratios presented above.

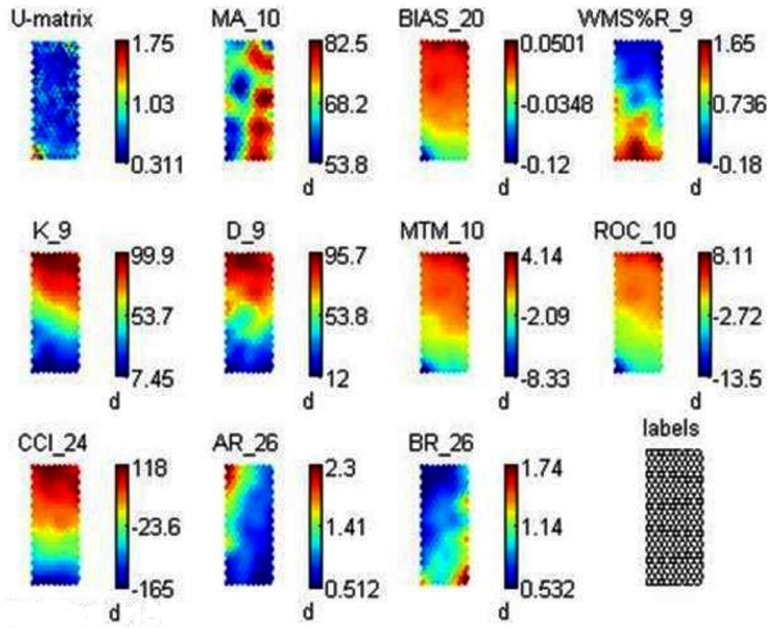

Figure 4. U-Matrix and planes of the SOM network. The results called "labels" on the bottom right are reproduced for clearness in Figure 5.

Table 1 shows the number of data in each cluster. Importantly, clustering has aimed to achieve a classification that leads to clusters with similar characteristics and that does not mean that the classes have the same amount of information. The process identifies in each cluster a texture that best describes the data of that cluster.

The MLP-BP is used to build supervised models for each cluster. The data will be divided into three groups, with 80% of the data used for training, 10% used for testing and 10% for validation. The segmentation of the database was random and the percentage composition of each group is used by many researchers [4,15–17].

The topology of the networks will be assembled from tests performed with the intermediate layer, with layers of five, 10 or 15 neurons. The input layer has 10 neurons, corresponding to 10 technical indices. The output layer has one neuron, which represents the projection to the value of the NYSE Composite US 100 index for the next day.

The determination of the number of neurons to be used in the intermediate layer has no specific rule [1]. Accordingly, the specification used in the experiments is that the number of neurons was lower, upper and equal to the size of the input layer.

Table 1. Results of the clustering process.

Cluster	Components
1	225
2	275
3	348
4	312
5	246
6	392
7	282

The activation function used was the hyperbolic tangent, which proves more efficient in the convergence of the results when the database is normalized to a range between −1 and +1.

Table 2 shows the results obtained in the training phase. The number of iterations was set at 2500, based on similar studies [4–13]. For each cluster, the number of neurons that resulted in a lower value for the RMSE is chosen. Thus, the result appears in Table 3.

Table 2. Analysis of the performance of the MLP neural model.

Cluster	Neurons in the Intermediate Layer	Iterations	RMSE
1	5	2500	0.5417
	10	2500	0.4220
	15	2500	0.5826
2	5	2500	0.3075
	10	2500	0.2619
	15	2500	0.1432
3	5	2500	0.2734
	10	2500	0.5092
	15	2500	0.1249
4	5	2500	0.3582
	10	2500	0.3103
	15	2500	0.3491
5	5	2500	0.2211
	10	2500	0.0995
	15	2500	0.4117
6	5	2500	0.3320
	10	2500	0.5273
	15	2500	0.2439
7	5	2500	0.3714
	10	2500	0.3178
	15	2500	0.4005

MLP: multi layer perceptron; RMSE: root mean square error.

Thus, with intermediate layers defined, the next step consists in testing the network, while applying the test data group. Table 4 shows the results, evaluated by the RMSE. These results can be considered satisfactory.

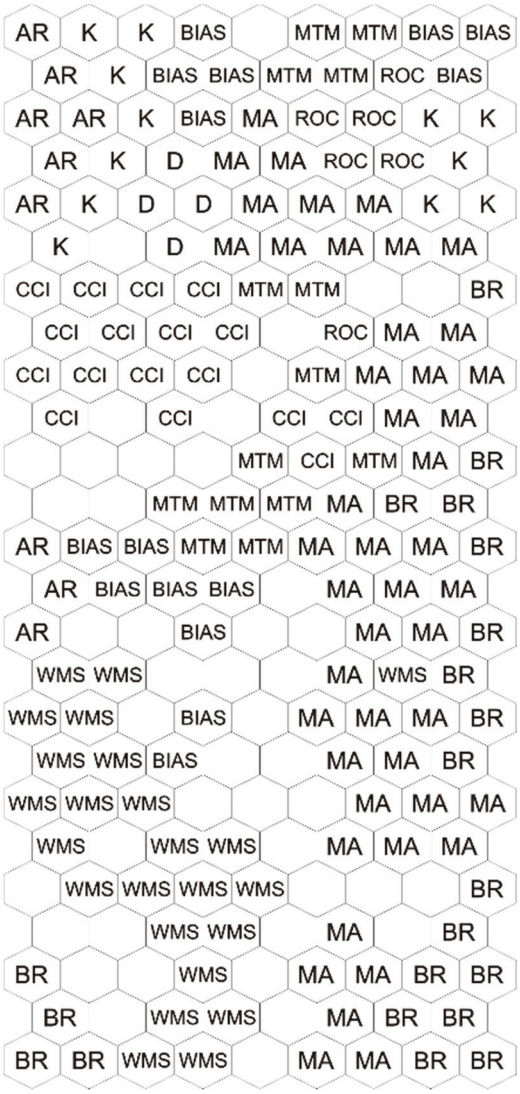

Figure 5. Visual identification of clusters after the network training. This figure is a reproduction of the results on the bottom right of Figure 4. The identification of clusters corresponds to the indexes defined in Equations (4)–(16).

It is important to emphasize that the parameters for the construction of the perceptron were established from previous works and references. The three numbers of neurons in the intermediate layer and the number of iterations were established a priori and a more detailed study could be undertaken for their optimization.

Table 3. Composition of the intermediate layer of the models with the best performance per cluster.

Cluster	1	2	3	4	5	6	7
Neurons in the intermediate layer	10	15	15	10	10	15	10

Table 4. Performance evaluation of neural models for the testing group.

Cluster	1	2	3	4	5	6	7
RMSE	0.4245	0.1563	0.1507	0.2950	0.2171	0.2402	0.3310

7. Results and Discussion

The performance of the model was evaluated with its application to data from the validation group. The performance, as stated earlier, was evaluated by determining the RMSE, the MAE and the MAPE. The results can be seen in Table 5.

Table 5. Performance evaluation of models for the validation group.

Hybrid Model SOM-MLP-BP	RMSE	MAE	MAPE
Cluster 1	0.5253	0.3034	0.3837%
Cluster 2	0.3242	0.1671	0.2140%
Cluster 3	0.2939	0.1536	0.2075%
Cluster 4	0.3288	0.2747	0.1643%
Cluster 5	0.2419	0.1654	0.1937%
Cluster 6	0.3555	0.2751	0.2562%
Cluster 7	0.4718	0.2421	0.2355%

The results for the RMSE are somewhat higher than those obtained with the test group. The neural models showed different results for the RMSE varying in a range between 0.25 and 0.50 approximately.

Regarding the MAE, the results are in a range between 0.14 and 0.30 approximately. Regarding the MAPE, the results are located in the range between 0.15% and 0.25%, with only one result out of range (0.3837%).

The successful reproduction of the data present in the validation group can be attributed to the objectives of the networks applied. The SOM network separates the data according to their characteristics and MLP reproduces these characteristics.

The model is composed of networks that lead to value forecasted for the next day for the NYSE Composite US 100 index, for each of the identified clusters. In this work, the performance of the model was evaluated based on the performance of the model with the group of validation data.

Naturally, the work should be continued with the application of the model to the data following the data set used here, for the period subsequent to that considered in this work.

8. Conclusions

This article showed the time series modeling of the NYSE Composite US 100 index based on a hybrid model designed with the use of a SOM network followed by the use of an MLP-BP network. SOM is suitable for the classification of a time series into clusters and MLP is suitable for the identification of the features of these clusters, intending a model to forecast future values of the series. In fact, the data were divided into three groups: one for training, one for testing and another one for validation.

Thus, the description has been validated by the application of the model on a portion of the total data set. The results indicate that the model shows quite satisfactory performance. The RMSE varies between 0.24 and 0.53, MAE varies between 0.15 and 0.30 and MAPE varies between 0.16% and 0.38%. The successful reproduction of the data present in the validation group can be attributed to the objectives of the networks applied. The SOM network separates the data according to their characteristics and MLP reproduces these characteristics.

Acknowledgments: The authors thank the *Journal of Risk and Financial Management* for the opportunity to publish this article as Open Access. The third author thanks CNPq for the support for his research work.

Author Contributions: Ad.B. built the hybrid model and performed the experiments with neural networks. D.L.B. guided the work. Ad.B., D.L.B. and Al.B. discussed the results. Ad.B. and Al.B. wrote the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Haykin, S.S. *Neural Networks, a Comprehensive Foundation*, 2nd ed.; Prentice-Hall International: Upper Saddle River, NJ, USA, 1998.
2. Amari, S. Dreaming of mathematical neuroscience for half a century. *Neural Netw.* **2013**, *37*, 48–51. [CrossRef]
3. Armano, G.; Marchesi, M.; Murru, A. A hybrid genetic-neural architecture for stock indexes forecasting. *Inf. Sci.* **2005**, *170*, 3–33.
4. Hsu, C.M. A hybrid procedure for stock price prediction by integrating self organizing map and genetic programming. *Expert Syst. Appl.* **2011**, *38*, 14026–14036.
5. Bildirici, M.; Ersin, Ö.Ö. Improving forecasts of GARCH family models with the artificial neural networks: an application to the daily returns in Istanbul stock exchange. *Expert Syst. Appl.* **2009**, *36*, 7355–7362. [CrossRef]
6. Dai, W.; Wu, J.Y.; Lu, C.J. Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes. *Expert Syst. Appl.* **2012**, *39*, 4444–4452. [CrossRef]
7. New York Stock Exchange Data Base. Available online: www.nyse.nyx.com (accessed on 11 November 2016).
8. New York Stock Exchange. Available online: www.nyse.com/trade (accessed on 16 December 2016).
9. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **1982**, *43*, 59–69. [CrossRef]
10. Kohonen, T. New developments and applications of self organizing maps. In Proceedings of the International Workshop on Neural Networks for Identification, Control, Robotics, and Signal/Image Processing (NICROSP), Venice, Italy, 21–23 August 1996; IEEE Computer Society Press: Los Alamitos, CA, USA, 1996; pp. 164–172.
11. Kohonen, T. Essentials of the self organizing map. *Neural Netw.* **2013**, *37*, 52–65. [CrossRef] [PubMed]
12. Rumelhart, D.E.; McClelland, J.L. *Parallel Distributed Processing, Volume 1. Explorations in the Microstructure of Cognition: Foundations*; MIT Press: Cambridge, MA, USA, 1986.
13. Mostafa, M.M. Forecasting stock exchange movements using neural networks: Empirical evidence from Kuwait. *Expert Syst. Appl.* **2010**, *37*, 6302–6309. [CrossRef]
14. Setyawati, B.R.; Creese, R.C.; Sahirman, S. Neural network for cost estimation. *AACE Intern. Trans.* **2003**, *14*, 1–10.
15. Wang, J.Z.; Wang, J.J.; Zhang, Z.G.; Guo, S.P. Forecasting stock indices with back propagation neural network. *Expert Syst. Appl.* **2011**, *38*, 14346–14355.
16. Lu, C.J.; Wu, J.Y. An efficient CMAC neural network for stock index forecasting. *Expert Syst. Appl.* **2011**, *38*, 15194–15201. [CrossRef]
17. Tsai, C.F.; Hsiao, Y.C. Combining multiple feature selection methods for stock prediction: Union, intersection and multi-intersection approaches. *Decis. Support Syst.* **2010**, *50*, 258–269. [CrossRef]
18. Liu, F.; Wang, J. Fluctuation prediction of stock market index by Legendre neural network with random time strength function. *Neurocomputing* **2012**, *83*, 12–21. [CrossRef]
19. Li, S.T.; Kuo, S.C. Knowledge discovery in financial investment for forecasting and trading strategy through wavelet-based SOM networks. *Expert Syst. Appl.* **2008**, *34*, 935–951. [CrossRef]
20. Sandoval, L., Jr.; Mullokandov, A.; Kenett, D.Y. Dependency relations among international stock market indices. *J. Risk Financ. Manag.* **2015**, *8*, 227–265.