Stock market time series forecasting with data mining methods*[1]

Milán Csaba Badics

A large number of data mining methods have been introduced over the past 20–25 years to forecast stock market time series, as ever-newer and increasingly sophisticated models have appeared in the literature and in market practice. This paper demonstrates the utility of the different data mining models for active portfolio management, specifically discussing the applicability of noise filtering and hybrid methods. The primary objective has been to present a price forecast-based trading strategy that is profitable even after transaction charges. The forecasting potential of the different data mining methods was tested using the times series of OTP's closing stock prices.

JEL-codes: C45, G14, G17

Keywords: Stock time sereies forecasting, Trading strategy, Neural networks, ICA, EMD, Data mining methods

Introduction

Investors have focused their attention on the changes in stock market time series for decades, and they have experimented with a variety of methods for forecasting them. This degree of interest has prompted more and more research in academic circles into the possibilities of these forecasts. Conventional statistical and econometric economic models were used at first, but these proved only moderately effective due to the special characteristics of the time series (such as, for instance, their nonlinearity, nonstationarity and high noise-to-signal ratio). It was at this time that interest shifted to nonparametric data mining and machine learning methods, which have fewer statistical constraints and are frequently used in engineering; this toolkit then opened a new perspective, offering

* The views expressed in this article are those of the author(s) and do not necessarily reflect the offical view ot the Magyar Nemzeti Bank.

Milán Csaba Badics, Ph.D student of the PADS PHD scholarship of the Magyar Nemzeti Bank.

1 I would like to thank to Tamás Ferenczi, who helped me to process and understand the relevant literature during my study. I also like to thank Zoltán Hans, Mihály Szoboszlai and Balázs Márkus, who supplied me with useful advices during drafting my research results. The original and longer version of this study won the first place in the competition of the X. Kochmeister-prize organized by the Hungarian Stock Exchange in May 2014.I am currently a student of the Magyar Nemzeti Bank PADS Phd scholarship program, and the other relevant issues mentioned in the study are still the subject of my research.

more efficient financial time series forecasting. Over the past 30 years, an increasing range of data mining methods have been introduced to analyse the changes in stock market data. Initially one of the most popular methods, different types of neural networks, was used to great advantage over the statistical methods. Since even the slightest improvement in forecasting accuracy may return huge extra profits, the search for and proper parametrisation of the optimal network continued to gain popularity in investor and academic circles alike. Nevertheless, the above-average potential of investment decisions predicated on this diminished over time, given the great success and the wide range of application. This can happen to any strategy built on forecasting, once many start to use it at the same time. This did not, however, prompt investment decision-makers to reject these methods. On the contrary, they invested more and more efforts in adopting these methodologies, which had been used successfully in miscellaneous fields of engineering, to time series forecasting. Among other things, they started to use modified versions of the rest of the data mining methodologies (SVR, Random Forest), as well as noise filtering (ICA, PCA) and decomposition-based (EMD, wavelet) techniques. Besides, the use of multi-step hybrid methods and the combining of different forecasts also became widespread. Countless methods and models have been developed to date, with the stock market time series forecast-based strategy being one of the most popular. Their use, however, represents a serious challenge, as efficient forecasting presupposes familiarity with the advantages and disadvantages of the different models.

For this reason, this paper focuses on the best-known data mining methods suitable for active portfolio management, as well as their advantages and disadvantages, discussing which should be used when and how, and also touching upon the most important current research trends. The objective is to present the entire process, from the selection of stock prices to forecasting (OTP and MOL daily closing prices in this paper), through the definition of the necessary input variables and available data mining methods right up to the execution of the trade, essentially giving the reader a roadmap for forecast-based active portfolio management.

1 Stock market time series characteristics, forecast methods and difficulties

Financial time series are difficult to forecast, as they tend to be noisy, nonstationary, nonlinear and chaotic, and often incorporate structural breaks as well (*Hall,* 1994; *Li,* et al., 2003; *Yaser* and *Atiya,* 1996; *Huang* et al., 2010; *Lu* et al., 2009; *Oh* and *Kim,* 2002; *Wang,* 2003). It is for these reasons that forecasting financial and stock market time series is one of the greatest challenges for market participants.

The forecasting methods used in studies fall into two categories: statistical/econometric and data mining/machine learning methods. Traditional statistical approaches include linear regression, calculation of moving average, exponential smoothing, and ARIMA, GARCH and VAR methods. These methods return good forecasting results if the financial time series are linear or nearly linear, although this is not typical in real life. Besides, the conventional statistical methods demand large volumes of historic data, which must also have a normal distribution as a precondition for a good forecasting result (*Cheng* and *Wei,* 2014).

Data mining methods forego these requirements, as they are better able to model the nonlinear structure of time series. Besides neural networks, these include Support Vector Machines (SVM) and the different types of decision trees. These data-driven and nonparametric methods can reveal and manage the unknown interconnections between empirical data, and they are therefore more efficient in forecasting the changes in complex and nonlinear stock market data (*Chen* et al., 2003; Chun and Kim, 2004; *Thawornwong* and *Enke,* 2004; *Enke* and *Thawornwong,* 2005; *Hansen* and *Nelson* 2002). The increasing number of data mining articles and applications appearing in recent years demonstrates that these applications are competitive and have significant advantages over the traditional methods (*Lu* et al., 2009; *Duan* and *Stanley,* 2011; *Huang* et al. 2010; *Ni* and *Yin* 2009).

Every data mining method has disadvantages, however, which is why the crossing (hybridising) of different data mining techniques is increasingly popular in academic circles. The basic idea is that hybrid methods will eliminate the disadvantages of the individual methods and create synergies to improve forecasting accuracy. Basically, the hybrid methods consist of three different types. The first is based on the "divide-and conquer" principle, which holds, essentially, that it may be worthwhile to divide complex problems into several smaller ones and resolve these sequentially. One of the most widespread applications of this method in stock market forecasting is Empirical Mode Decomposition (EMD) (*Cheng* and *Wei,* 2014). In the second type, an attempt is made to filter out the noise from the input variables of the models, facilitating a more accurate result. This is the most frequently used method, Independent Component Analysis (ICA), and it is based on the principle that creating independent components (ICs) from the input variables allows isolation of the component containing the noise; if this component is then removed, forecasting accuracy will improve (*Lu* et al., 2009). The third method involves combining the forecasts of the different data mining models, from simple aggregation through Bayesian averaging to Lasso regression. The methods of combining are predicated on the principle that forecasting variation may be reduced by taking all the methods into account at the same time (*Sermpinis* et al., 2012).

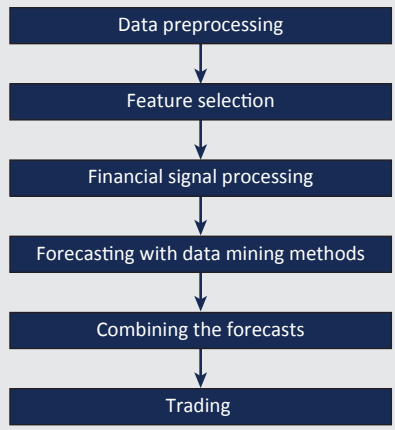
The following challenges and difficulties are faced when creating a stock market forecast-based active portfolio management strategy:

1. the selection of the appropriate explanatory variables (*feature selection*)
2. noise filtering, signal processing (*financial signal processing*)

3. forecasting with a data mining method, with the parameters optimised (*forecasting with data mining methods*)
4. combining different forecasts (*combining data mining techniques*)

Chart 1 shows the entire process, including the preparation and transformation of data and trading.

Chart 1
Structure of the model framework presented in this research



(edited by author)

2 Process of building a data mining model

2.1 Noise filtering and hybrid methods

Accurate forecasting presupposes identification of the latent variables underlying stock price movements and using them in the modelling process. Independent Component Analysis, a method widely used in engineering, is suitable for solving this type of problem. This process is able to reveal hidden components that drive the changes in data series, isolating them in such a way that they depend on one another as little as possible and their linear combination can be used for reconstructing the original data series (*Kapelner* and *Madarász,* 2012).

Independent Component Analysis allows locating and removing the noise component from the data used in the modelling process, thus improving the accuracy of the forecast (*Lu,* 2010). This is a method frequently used in engineering for *signal processing* (*Beckmann* and *Smith,* 2004), *noise filtering in facial recognition systems* (*Déniz* et al., 2003) and, of course, *stock market time series forecasting. Oja* et al. (2000) used Independent Component Analysis to reduce the noise/signal ratio in the model input data and then forecast exchange rates with an autoregressive model.

The EMD decomposition process uses a slightly different approach than noise filtering, as it tries to filter out the noise from the original time series rather than the input variables. The aforementioned "divide and conquer" principle is the essence of empirical mode decomposition, which was developed by *Huang* et al. (1998) and is based on the Hilbert-Huang transformation. This decomposes the original time series into a finite number of IMFs, which are easier to manage. Because they are highly correlated, it is simpler to forecast them one by one and then, after aggregation, arrive at the forecast for the original time series (*Cheng* and *Wei,* 2014). This method is often used for *decomposing earthquake signals* (*Vincent* et al., 1999), *forecasting wind speed* (*Guo* et al., 2012), and even *determining tourism demand* (*Chen* et al., 2012). In addition to ICA, therefore, this method is combined with a data mining model in the current study.

2.2 Possible data mining methods

The different neural networks are the most widespread and popular of the data mining methods used to forecast financial time series (*Cao* and *Parry,* 2009; Chang et al., 2009; *Chavarnakul* and *Enke,* 2008; *Enke* and *Thawornwong,* 2005). These data-driven, nonparametric methods do not require strong model assumptions or advanced statistical assumptions about the input data, and they are able to model any kind of nonlinear function (*Vellido* et al., 1999; *Zhang* et al., 1998). In their article reviewing nearly 100 studies, Atsalakis and Valavanis (2009) point out that, of all the different kinds of neural networks, feed-forward neural networks (FFNN) and recurrent neural networks (RNN) are the ones most frequently used by researchers to forecast financial time series. The most popular feed-forward neural network is the back-propagation neural network (BPN), whereas the Elman and Jordan networks are the most popular of the recurrent networks.

Further solutions to forecast time series include Support Vector Machines, decision trees and genetic algorithms. The large number of available methods may make it highly time-consuming to identify the most efficient option for a particular time series and, as we have seen, each also has its advantages and disadvantages; therefore, more than one method is often used in the modelling process and the results are then combined. A feed-forward neural network with one hidden layer can model any kind of complex problem (*Chauvin* and *Rumelhart,* 1995), which is why it is used here.

2.3 Combining data mining methods

One of the most interesting questions in the literature on forecasting concerns the ways in which different forecast techniques may be combined. Given the advantage of compensating for the deficiencies of individual methods, several researchers have pointed out the merits of combining different techniques, especially for short-term forecasts (*Zhang* and *Wu* 2009, *Armstrong* 1989). Although Timmermann (2006) notes that simple averaging might compete with the more sophisticated techniques, there are indeed situations when one method is much more accurate than others, making averaging insufficiently efficient. *Granger* and *Ramathan* (1984) recommended the regression technique as having promising results, whereas *Swanson* and *Zeng* (2001) advocated Bayesian averaging. Almost all of the authors have asserted that it is necessary to combine different forecasting methods, but there has been no agreement on when and which should be employed; therefore, several combinations have been used in this analysis.

3 Methods used in this paper

3.1 Independent Component Analysis

If data mining models are taught without taking into account their potential noise content, this may weaken the ability to generalise to the test set or lead to overfitting. Therefore, noise filtering of the input data is an essential task during the modelling process. Independent Component Analysis is used for this purpose here, but first the theory underlying it is explained.

Let $\mathbf{X}=[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m]^T$ be a multidimensional data matrix that is mxn in size, where m≤n and the $\mathbf{x}_i$ dimension of the observed mixed signals is 1xn i=1, 2, ..., m. Using the ICA model, this **X** matrix can be written out as the following equation:

$$X=AS=\sum_{i=1}^{m} a_i s_i ,$$

where $\mathbf{a}_i$ is column i of the mxm-sized unknown **A** mixing matrix and $\mathbf{s}_i$ is row i of the mxn-sized "source" **S** matrix. The $\mathbf{s}_i$ vectors are the latent data that cannot directly be observed from the mixed $\mathbf{x}_i$ data, but the latter can be written out as the linear combination of these latent data (Dai et al., 2012). The purpose of Independent Component Analysis is to find the mxm-sized W matrix (demixing matrix), for which it is true that

$$Y=WX,$$

where $\mathbf{y}_i$ is row i of the **Y** matrix, i=1, 2, …, m and these vectors are statistically independent (independent components). If the **W** matrix is the inverse of the **A** mixing matrix ($\mathbf{W=A^{-1}}$), then the independent components ($\mathbf{y}_i$) can be used to estimate the original latent signals $\mathbf{s}_i$ (Lu, 2010).

In the course of Independent Component Analysis, an optimisation problem is solved by selecting an objective function of the statistical independence of the independent components and using optimisation processes to find the **W** matrix (*Lu* et al., 2009). Several processes of this type have been designed and developed (*Bell* and *Sejnowski,* 1995; David and *Sanchez,* 2002; *Hyvärinen* et al., 2001). These tend to employ unsupervised teaching algorithms to maximise the statistical independence of the ICs. One of the most frequent ICA solutions is the FastICA algorithm (*Hyvärinen* et al., 2001), which I have also used to define the **W** matrix.

3.2 Empirical Mode Decomposition (EMD)

Empirical mode decomposition is a nonlinear signal transformation process developed by Huang et al. (1998) for the decomposing of nonlinear and nonstationary time series. The method decomposes the original time series into oscillating IMF (Intrinsic Mode Function) components of differing timescales (*Yu* et al., 2008). Each IMF must satisfy two conditions: first, the difference between the total number of local minimums and maximums and the number of zero positions of the function may not be more than one; second, the local average must be zero (*Cheng* and *Wei,* 2014). This algorithm is the following:

1. Define all local minimums and maximums of **x(t)**.
2. Define the lower envelope $\mathbf{x_u(t)}$ and the upper envelope $\mathbf{x_l(t)}$ of **x(t)**.
3. Using the upper and the lower envelope, set the time series average $\mathbf{m_1(t)=[x_u(t)+x_l(t)]/2}$.
4. Calculate the difference between the original time series **x(t)** and the average $\mathbf{m_1(t)}$ time series calculated in the previous step $\mathbf{h_1(t)=x(t)-m_1(t)}$, which will return the first IMF ($\mathbf{h_1(t)}$), if it satisfies the aforementioned two conditions.
5. Once the first IMF has been determined, continue with the same iteration algorithm until arrival at the final time series, the residual component **r(t),** which is a monotonic function signalling that the algorithm should be ended (*Huang* et al., 1998).

The original time series **x**(t) can be restored as the sum of the IMF components and the residual:

$$X(t)=\sum_{i=1}^{n} h_i(t)+r(t).$$

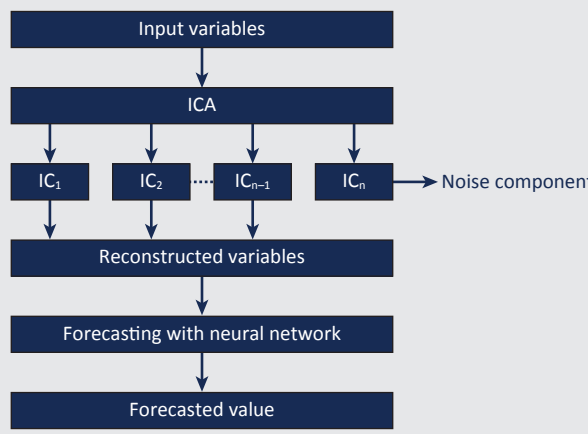
The resulting IMFs are nearly orthogonal to each other and their average is near zero (*Yu* et al., 2008). The residual is the trend component of the original time series, whereas the IMFs follow a decreasing order of ever-lower frequencies (*Cheng* and *Wei,* 2014).

3.3 Brief description of the ICA-BPN and EMD-BPN hybrid models

The first hybrid model used here comprises three steps: first, it uses the ICA method to determine the independent components (ICs) of the input variables; then it uses the TnA (Testing-and-Acceptance) method of Cheung and Xu (2001) to identify and filter out the noise component; and finally, it forecasts the time series with the help of the BPN neural network. Chart 2 demonstrates this process:

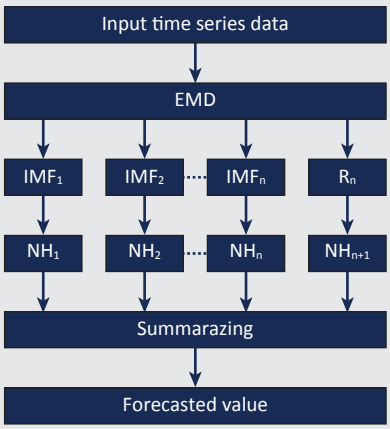
The other hybrid model used in this paper also comprises three steps: first, the EMD method is used to decompose the original time series into IMF components and the residual; then a BPN model is used to forecast the values for the next period for each IMF; and finally, the forecast value of the original time series is created as a sum of these. Chart 3 shows this hybrid method:

Chart 2
ICA-BPN model flowchart



(edited by author)

Chart 3
EMD-BPN model flowchart



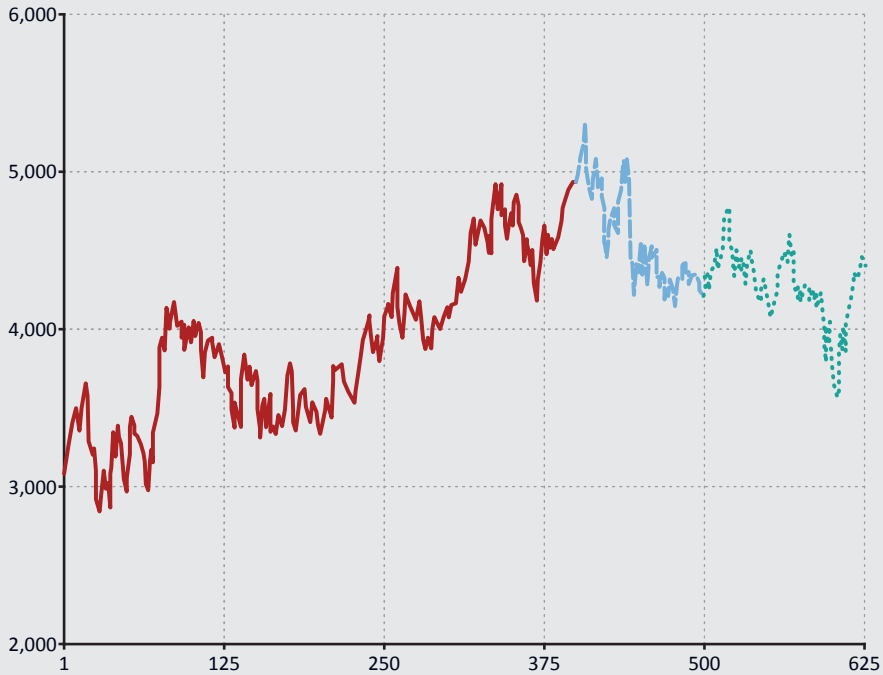
(edited by author)

4 Empirical analysis

4.1 Data and performance criteria

This research applies trading strategies based on the forecasting of the closing prices of OTP shares traded on the Budapest Stock Exchange. The period between 3 October 2011 and 11 April 2014 were selected for analysis. The time series was divided into learning, testing and validating data sets with respective ratios of 64%, 16% and 20% (the forecasting models were tested on the last six months of this time series of two and a half years). Chart 4 shows the price changes, with different sets marked in red, blue and green.

Chart 4
OTP stock prices over the period analysed



(edited by author)

Table 1
The forecasting results of the different methods

	Max.	Min.	Average	Deviation
Weighted MA	5,302	2,835	4,061.4	516
Momentum	789	-814	15.3	242.3
Stochastic K%	100	0	53.5	31.3
Stochastic D%	98.8	2.6	53.4	27.2
RSI	88.5	15.2	51.6	15.4
MACD	235.5	-231.4	3.4	76.3
LW R%	0	-100	-47	30.6
A/D Oszcillator	100	0	51.2	28.7

(edited by author)

The eight technical indicators selected for modelling have been widely and successfully used by Kara et al. (2011), among others. Table 1 contains the statistical features of the indicators in the period analysed.

4.2 The forecasting results of the different methods

In the empirical part of the research, three data mining models were used. The first step taken in the modelling was normalisation of the input data, which was a prerequisite for fast convergence of the algorithms. The data was transformed into a [0,1] interval for each variable by means of the following method:

$$x'_i = \frac{x_i - x_{min}}{x_{max} - x_{min}},$$

where x_{min} and x_{max} are the minimums and maximums of the individual variables in a given time series.

The first step in the modelling employed one of the most popular neural networks, the backpropagation neural network (BPN). The appropriate parameters (number of neurons in hidden layer, learning rate) were selected by means of the grid search process. The network's input layer comprised eight neurons (in accordance with the number of explanatory variables), whereas networks of 11, 12, 13 and 14 neurons were tested in the interim layer. The network had one output: the daily yield of the share. Relying on the study by *Lu* (2010), the models were tested at low learning rates (0.01, 0.02, 0.03, 0.04, 0.05) in the learning process. The convergence criterion used was a rule that the learning process would be halted if the RMSE indicator fell below 0.0001 or if the 1000th iteration was reached. The network topology with the lowest RMSE in the test set was chosen as optimal. Table 2 shows the performance measured in the test set at different neural network parameters; subsequently, when modelling in the validation set, this was relied on to use the network with 8-12-1 topology and 0.05 learning rate.

Chart 5 shows the original and the forecast prices, absolute error and sign accuracy in the validation period.

Since the financial time series is characterised by a high noise/signal ratio, Independent Component Analysis was used to filter out the noise from the input variables prior to using the BPN network in the second model. This first required generating the independent components (ICs) and, second, identifying the noise component with the help of the TnA algorithm.

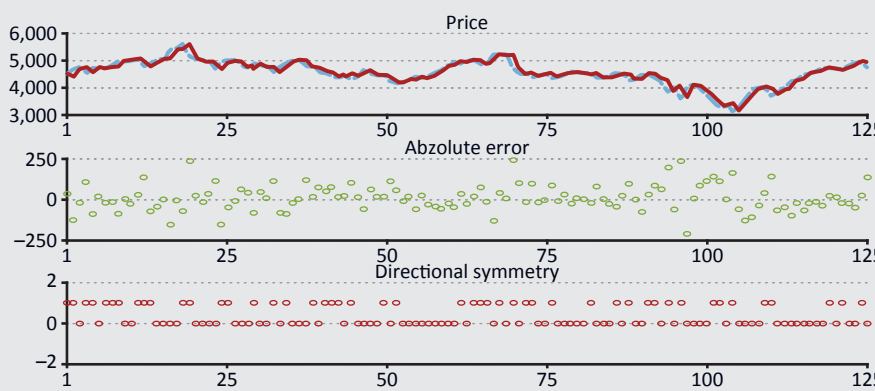
In a TnA algorithm, the individual ICs were discarded one by one; then the input matrix was restored and analysed in terms of the extent to which it differed from the original. Deviation was measured with the RHD indicator. As eight input variables were used, it was necessary to perform this operation seven times to find the noise component. These RHD values are shown in Table 3.

Table 2
Test set errors of BPN networks with different parameters

Number of neurons in hidden layer	Learning rate	Validation RMSE
11	0.01	0.124111
	0.02	0.120873
	0.03	0.119689
	0.04	0.119021
	0.05	0.118578
12	0.01	0.120424
	0.02	0.117532
	0.03	0.116893
	0.04	0.116581
	0.05	0.116369
13	0.01	0.124840
	0.02	0.123034
	0.03	0.121980
	0.04	0.121219
	0.05	0.120619
14	0.01	0.124489
	0.02	0.120798
	0.03	0.119771
	0.04	0.119247
	0.05	0.118872

(edited by author)

Chart 5
BPN model forecast accuracy in the validation period



(edited by author)

Table 3
RHD values of reconstructed input matrixes

Main components	RHD
IC1, IC2, IC3, IC4, IC5, IC6, IC7	4.3674
IC1, IC2, IC3, IC4, IC5, IC6, IC8	3.6260
IC1, IC2, IC3, IC4, IC5, IC7, IC8	4.4830
IC1, IC2, IC3, IC4, IC6, IC7, IC8	2.4118
IC1, IC2, IC3, IC5, IC6, IC7, IC8	3.7873
IC1, IC2, IC4, IC5, IC6, IC7, IC8	3.9655
IC1, IC3, IC4, IC5, IC6, IC7, IC8	7.1473
IC2, IC3, IC4, IC5, IC6, IC7, IC8	7.7748

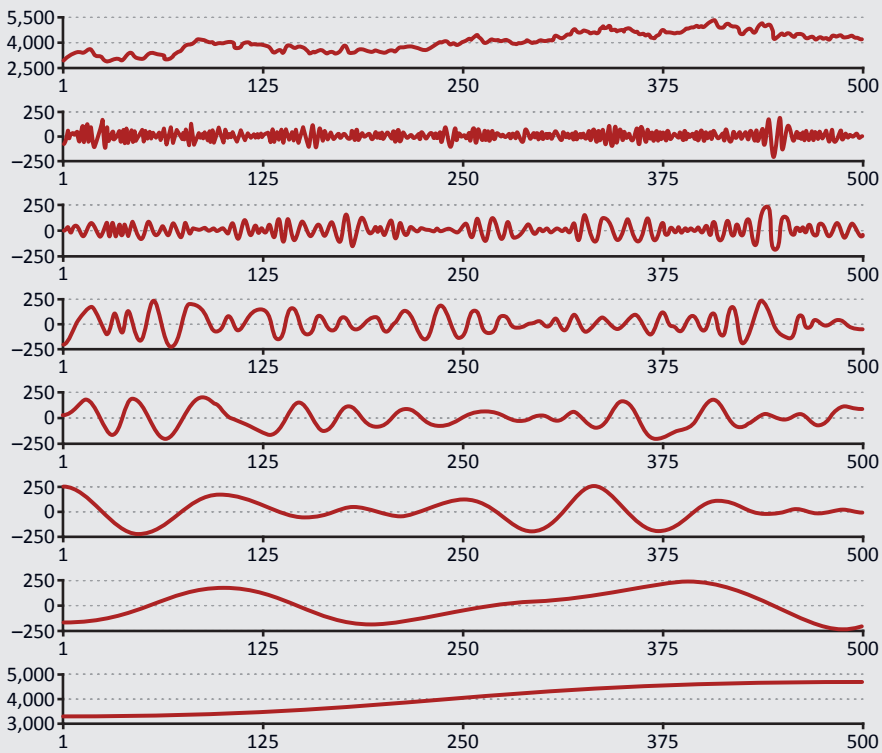
(edited by author)

The table demonstrates that the noise is in the fifth component. The second step in the model involved using the reconstructed variables to build a BPN network. The selection of the optimal parameter was closely similar to that described above; the optimal network turned out to be the one with 8-12-1 topology in the ICA-BPN model as well.

In the third method, EMD was used to decompose the original time series into IMFs, given the complex dynamics of the stock market time series. Each of these were forecast separately, then aggregated to arrive at the forecast value for the original time series. As done in a number of other studies (*Yu* et al., 2008; *Cheng* and *Wei,* 2014), stock prices were forecast with this method. Chart 6 shows the empirical mode decomposition of the OTP stock price.

I indicate the original time series in the top row, then the IMF's of ever decreasing frequencies (IMF1, IMF2, …, IMF6) below, and finally the residual of the trend in the last row. As the second step in the method, each IMF was forecast with neural networks of different parameters; the results were then aggregated to arrive at the OTP closing price for the following day. Because it was ultimately necessary to forecast eight time series in this instance and determine the number of optimal inputs for them (the number of lags needed in the NAR model), the process was much more complex and time-consuming than in the earlier two models. In order to keep the problem solvable, the process followed Mingming and Jinliang (2012) and the number of lags were set as 10 for each IMF. In this way, it was only necessary to find the optimal number of neurons and learning rate. Table 4 shows these for the different IMFs.

Chart 6
Profits achievable in the validation set using the various data mining and combinatory models
(OTP)



(edited by author)

Table 4
Optimal parameters of different IMF's

IMF	Number of neurons	Learning rate
1	12	0.05
2	12	0.05
3	12	0.05
4	12	0.025
5	12	0.025
6	12	0.025
7	13	0.025
8	13	0.025

(edited by author)

Once the optimal parameters of the three models were identified, they were used to create a forecast for the validation period (Table 5).

Table 5
Performances of different methods in the validation set
(OTP)

Modell	RMSE	MAPE (%)	DA (%)
BPN	0.018864	113.38	61.6
ICA–BPN	0.018738	107.79	60.8
EMD–BPN	0.026672	292.47	56.8

(edited by author)

The tables make it clear that the sign-forecasting rate of the more sophisticated hybrid methods is no better than that of the simple BPN model; nevertheless, it will be interesting to investigate in the following whether they surpass the first model in terms of profits achieved.

First, it was checked whether combining the three methods could improve the achievable forecast results. As mentioned above, combining helps eliminate disadvantages of the individual methods and thus facilitates better forecasting and higher profits. Table 6 summarises the forecasting results of three kinds of combinations of the three methods (simple average, Bayesian average, GRR).

Table 6
Results from combining the three methods
(OTP)

Modell	RMSE	MAPE (%)	DA (%)
Átlag	0.018854	144.82	64.8
Bayes-i átlag	0.018733	107.87	60.8
GRR	0.019087	151.21	61.6

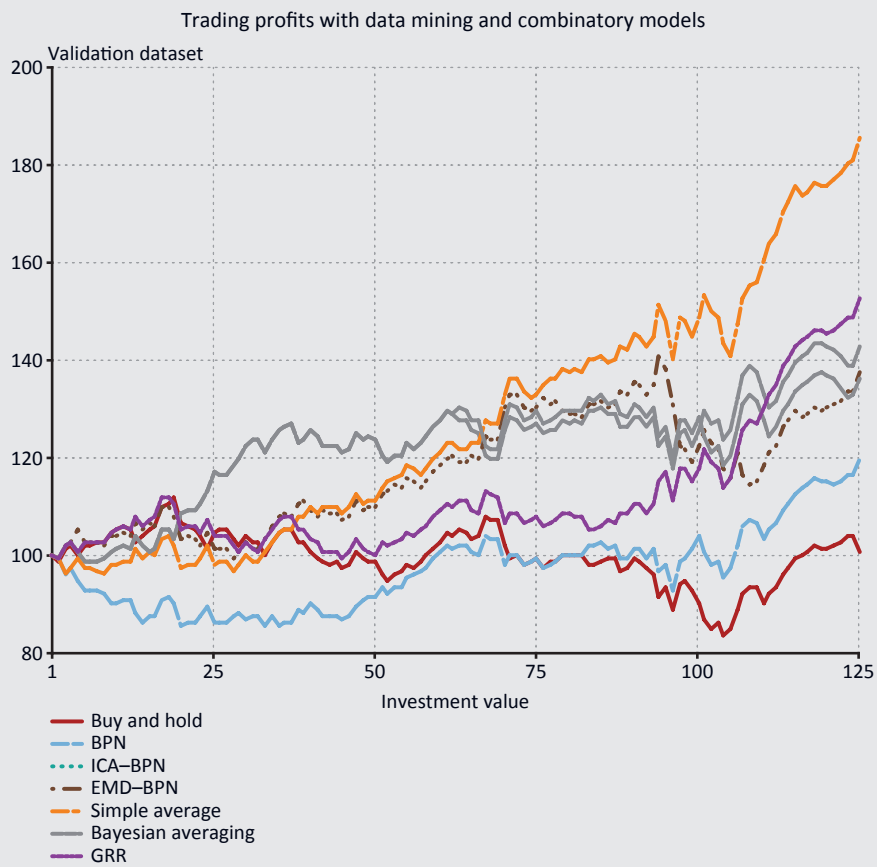
(edited by author)

Table 7
Profit generated by the 6 models in the validation set
(OTP)

	Buy and hold	BPN	ICA–BPN	ICA–EMD	Average	Bayes-average.	GRR
Annual yield	2.50%	36.46%	71.53%	64.01%	124.27%	62.71%	85.16%
Annual volatility	30.63%	30.22%	29.68%	29.92%	28.73%	29.81%	29.52%

(edited by author)

Chart 7
Az egyes adatbányászati és kombinációs modellek alkalmazásával elérhető profit a validációs halmazon
(OTP)



(edited by author)

Table 7 and Chart 7 show the profits from the three data mining models and three combination models on the validation set (incorporating a 0.1% transaction charge).

The table demonstrates that although the sign-forecasting ratio is worse in the pair of more complex data mining models than in the simple neural network, they greatly surpass the latter in terms of profits. Furthermore, all three models perform better than a "buy-and-hold" strategy. In addition, it is striking that out of all three combining methods, simple averaging delivers the best performance in terms of profits. This may appear surprising at first, since it is the least sophisticated method of averaging. If one considers, however, that the other two weight the models on the basis of errors in the learning and test sets, where the models performing better in terms of profits (ICA-BPN and ICA-EMD) had performed worse than the simple neural network and these are thus included with lower weights in the average, it becomes clear why these combinations result in lower profits. These two methods (ICA-BPN and EMD-BPN) forecast with nearly the same accuracy in the validation period as in the test and learning periods, whereas the simple neural network forecasts worse; thus, when the sophisticated models have higher weights in the combination (simple averaging), more profits are achieved. It would be worthwhile to analyse in subsequent research what would happen if the weights of the other two combination methods were to be selected on the basis of profits instead of RMSE.

5 Challenges in using the method, further opportunities for research, conclusion

As seen from the results of the previous chapter, active management that relies on data mining techniques will have more success than with a "buy and hold" strategy. Nevertheless, it is not straightforward to use models; there are certain difficulties and constraints. These include selecting the optimal methods and parameters, as well as recalibrating the models periodically (perhaps even daily), which is extremely time-consuming and requires extreme care.

For this reason, it is fitting to suggest a few opportunities for further progress and research directions. It would, of course, be interesting to investigate whether the same methods can return the most accurate forecasts for different securities, and, if not, which feature of the stock is responsible. Furthermore, it is also recommended to remember and rely on the data mining methods only briefly touched upon in this paper, from genetic algorithms through decision trees to textual data mining. The latter, which has been the most dynamically advancing area in the past two or three years, involves automated analysis

of market news, establishing expected impacts on the different securities (*Hagenau* et al., 2013).

As far as trading is concerned, it may be important to analyse the profits achievable with leverage, which is permitted on an extremely high number of markets. Sermpinis et al. (2012) presented a methodology for this, forecasting not only prices but also the volatility of securities and determining the rate of leverage according to how high this was (high leverage in the event of low volatility and vice versa). However, this is almost a unique example, as most researchers ignore this matter. It might, therefore, be worthwhile to investigate the possibilities for forecasting volatility, as well as the validity of the leverage rules formulated on this basis.

The reasons listed demonstrate that while data mining allows achieving significant extra yields compared to traditional investment strategies, its execution is highly complex and problematic, requiring considerable resources and expertise.

References

Armstrong, J. S. (1989): "Combining forecasts: the end of the beginning or the beginning of the end?", International Journal of Forecasting, 5, pp. 585–588.

Atsalakis, G. S. and Valavanis, K.P. (2009): "Surveying stock market forecasting techniques – Part II. Soft computing methods", *Expert Systems with Applications*, 36(3), pp. 5932–5941.

Beckmann, C. F. and Smith, S. M. (2004): "Probabilistic independent component analysis for functional magnetic resonance imaging", *IEEE Transactions on Medical Imaging*, 23(2), pp. 137–152.

Bell, A. J. and Sejnowski, T. J. (1995): "An information-maximization approach to blind separation and blind deconvolution", *Neural Computation*, 7, pp. 1129–1159.

Cao, Q. and Parry, M. E. (2009): "Neural network earnings per share forecasting models: A comparison of backward propagation and the genetic algorithm", *Decision Support Systems*, 47(1), pp. 32–41.

Chang, P-C., Liu, C-H., Lin, J-L., Fan, C-Y. and Ng, C. S. P. (2009): "A neural network with a case-based dynamic window for stock trading prediction", *Expert Systems with Applications*, 36(3), pp. 6889–6898.

Chauvin, Y. and Rumelhart, D. E. (1995): *Backpropagation: Theory, architectures, and applications*. New Jersey: Lawrence Erlbaum Associates.

Chavarnakul, T. and Enke, D. (2008): "Intelligent technical analysis-based equivolume charting for stock trading using neural networks", *Expert Systems with Applications*, 34(2), pp. 1004–1017.

Chen, A-S., Leung, M. T. and Daouk, H. (2003): "Application of neural networks to an emerging financial market: Forecasting and trading the Taiwan Stock Index", *Computers and Operations Research*, 30(6), pp. 901–923.

Chen, C.f., Lai, M.c. and Yeh, C.c. (2012): "Forecasting tourism demand based on empirical mode decomposition and neural network", *Knowledge-Based Systems,* 26, pp. 281–287.

Cheng, C-H. and Wei L-Y. (2014): "A novel time-series model based on empirical mode decomposition for forecasting TAIEX", *Economic Modelling*, 36, pp. 136–141.

Cheung, Y. M. and Xu, L. (2001): "Independent component ordering in ICA time series analysis", *Neurocomputing*, 41(1–4), 145–152.

Chun S-H. and Kim S.h. (2004): "Data mining for financial prediction and trading: application to single and multiple markets", *Expert Systems with Applications*, 26 (2), pp. 131–139.

Dai, W., Wu, J-Y. and Lu, C-J. (2012): "Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes", *Expert Systems with Application*s, 39(4), pp. 4444–4452.

David, V. and Sanchez, A. (2002): "Frontiers of research in BSS/ICA", *Neurocomputing*, 49(1), pp. 7–23.

Déniz, O., Castrillón, M. and Hernández, M. (2003): "Face recognition using independent component analysis and support vector machines", *Pattern Recognition Letters*, 24(13), pp. 2153–2157.

Duan, W-Q. and Stanley, H.e. (2011): "Cross-correlation and the predictability of financial return series", *Physica A: Statistical Mechanics and its Applications*, 390(2), pp. 290–296.

Enke, D. and Thawornwong, S. (2005): "The use of data mining and neural networks for forecasting stock market returns". *Expert Systems with Applications*, 29(4), pp. 927–940.

Granger, C.w.j. and Ramanathan, R. (1984): "Improved methods of combining forecasts", *Journal of Forecasting*, 3(2), pp. 197–204.

Guo, Z., Zhao, W., Lu, H. and Wang, J. (2012): "Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model", *Renewable Energy*, 37(1), pp. 241–249.

Hall, J. W. (1994): "Adaptive selection of US stocks with neural nets", *Trading on the edge: Neural, genetic and fuzzy systems for chaotic financial markets*, pp. 45–65.

Hansen, J. V., and Nelson, R. D. (2002): "Data mining of time series using stacked generalizers", *Neurocomputing*, 43(1), pp. 173–184.

Hagenau, M., Liebmann, M. and Neumann, D. (2013): "Automated news reading: Stock price prediction based on financial news using context-capturing features", *Decision Support Systems*, 55(3), pp. 685–697.

Huang, N.e., Shen, Z., Long, S.r., Wu, M.c., Shih, H.h., Zheng, Q., Yen, N.c., Tung, C.c. and Liu, H.h. (1998): "The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis", *Proceedings of the Royal Society of London A – Mathematical Physical and Engineering Sciences, Series A*, 454, pp. 903–995.

Huang, S-C., Chuang, P-J., Wu, C.f. and Lai, H-J. (2010): "Chaos-based support vector regressions for exchange rate forecasting", *Expert Systems with Applications*, 37(12), pp. 8590–8598.

Kara, Y., Boyacioglu, M. A. and Baykan, Ö. K. (2011): "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of Istanbul Stock Exchange", *Expert Systems with Applications*, 38, pp. 5311–5319.

Kapelner, T. and Madarász, L. V. (2012): *Független komponens analízis és empirikus tesztjei kötvényhozamok felhasználásával*. (*Independent component analysis and its empirical tests using bond yields.*) TDK (student research) paper.

Li, T., Li, Q., Zhu, S. and Ogihara, M. (2003): "A survey on wavelet applications in data mining", *SIGKDD Explorations*, 4(2), pp. 49–68.

Lu, C-J., Lee, T-S. and Chiu, C-C. (2009): "Financial time series forecasting using independent component analysis and support vector regression", *Decision Support Systems* 47(2), pp. 115–125.

Lu, C-J. (2010): "Integrating independent component analysis-based denoising scheme with neural network for stock price prediction", *Expert Systems with Applications*, 37(10), pp. 7056–7064.

Mingming, T. and Jinliang, Z. (2012): "A multiple adaptive wavelet recurrent neural network model to analyse crude oil prices", *Journal of Economics and Business*, 64(4), pp. 275–286.

Ni, H. and Yin, H. (2009): Exchange rate prediction using hybrid neural networks and trading indicators", *Neurocomputing*, 72(13–15), pp. 2815–2823.

Oja, E., Kiviluoto, K., and Malaroiu, S. (2000): "Independent component analysis for financial time series", in: *Proceedings of the IEEE 2000 adaptive systems for signal processing, communications, and control symposium*, Lake Louise, Canada. pp. 111–116.

Oh, K. J. and Kim, K.-J. (2002): "Analyzing stock market tick data using piecewise nonlinear model", *Expert System with Applications*, 22(3), pp. 249–255.

Sermpinis, G., Dunis, C., Laws, J. and Stasinakis, C. (2012): "Forecasting and trading the EUR/USD exchange rate with stochastic Neural Network combination and time-varying leverage", *Decision Support Systems*, 54(1), pp. 316–329.

Swanson, N.R. and Zeng, T. (2001): "Choosing among competing econometric forecasts: regression-based forecast combination using model selection", *Journal of Forecasting*, 20(6), pp. 425–440.

Timmermann A. (2006): "Chapter 4: Forecast Combinations", *Handbook of Economic Forecasting*, 1, pp. 135–196.

Thawornwong, S. and Enke, D. (2004): "The adaptive selection of financial and economic variables for use with artificial neural networks", *Neurocomputing*, 56, pp. 205–232.

Vincent, H.T., Hu, S-L.J. and Hou, Z. (1999): "Damage detection using empirical mode decomposition method and a comparison with wavelet analysis", *Proceedings of the Second International Workshop on Structural Health Monitoring*, Stanford, pp. 891–900.

Vellido, A., Lisboa, P. J. G. and Vaughan, J. (1999): "Neural networks in business: A survey of applications (1992–1998)", *Expert Systems with Applications*, 17(1), pp. 51–70.

Wang, Y-F. (2003): "Mining stock prices using fuzzy rough set system", *Expert Systems with Applications*, 24(1), pp. 13–23.

Yaser, S. A-M. and Atiya, A. F. (1996): "Introduction to financial forecasting", *Applied Intelligence*, 6, pp. 205–213.

Yu, L., Wang, S. and Lai, K.K. (2008): "Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm", *Energy Economics*, 30(5), pp. 2623–2635.

Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998): "Forecasting with artificial neural networks: The state of the art", *International Journal of Forecasting*, 14, pp. 35–62.

Zhang, Y. D., and Wu, L. N. (2009): "Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network", *Expert Systems with Applications*, 36, pp. 8849–8885.