

Forecasting performance analyses of stock prices considering time series based yearly trends

Heedong Hong¹ and Hyunsoo Lee^{2*}

¹Dept. of Electronics Commerce,
Semyung University, Jecheon,
27136, Korea

myclub114@semyung.ac.kr

²School of Industrial Engineering,
Kumoh National Institute
of Technology,

Gumi, Korea, 730-701, Korea

hsl@kumoh.ac.kr

Abstract

Background/Objectives: This research checks the effectiveness of past trend data for forecasting a future stock data. The corroborated examinations from this research guide the more right decision making for stock investments.

Methods/Statistical Analysis: In order to check the effectiveness of past trend, three prediction models are provided. An autoregressive model only considering recent data is compared with a model incorporating past trend. Then, a trend data is compensated and is incorporated to the second model. Multi-layer perceptron, Partial least square method and Support vector regression analyses are applied for testing these prediction models and the total forecasting errors are compared.

Findings: As test results, Partial least square method is considered as the best forecasting analyzer. However,

the analyzers' performances are different in each forecasting model. Support vector regression shows the best forecasting performance in testing the autoregressive model. Then, the consideration of past trend gives much worse forecasting errors than a general time-series model without trend components. It implies that the accurate extraction of past trend is difficult and a general past data fails to characterize current stock data. As a result, a recent data based stock prediction is considered as a much better stock investment strategy.

Improvements/Applications: This research results are used as a decision-making guidance for stock investment. In order to increase the forecasting accuracy, partial considerations of accurate trends are expected.

Key Words : Time Series, Trend Analyses, Forecasting Performance, Partial Least Square, Multi-Layer Perceptron, Support Vector Regression.

1 Introduction

Forecasting of stock prices has been wide issues in many research studies, markets and related applications. While many forecasting methods have been proposed, the embedding uncertainties on stock prices make it possible to predict stock indices inaccurately. In particular, sudden political issues or other events might influence heavily on the changes of stock prices. This study focuses on issues handling the past trend in its forecasting. While many regression based forecasting methods attempt to predict stock prices with predefined decision variables, they are limited in the fact that insufficient number of decision variables are used for its prediction. As stock prices are influenced by many economic, political, social and other aspects, it is impossible to consider overall aspects and to incorporate those in a prediction model. For this reason, time series based prediction approaches have been used popularly. When time series based methods are applied, the basic assumption is that past data trends influence on current data. It indicates that similar aspects are included already in the past trend and a future data can be analyzed using the consideration of the trends.

In usual, past data can be classified into two types: recent data and past trend data. When several seasonal effects¹ are included and they invoke cycles, the past trend data have to be considered with a recent data. This research examines the influence level of the past trend in a stock prices' forecasting. While there are huge beliefs that the past trend is important for considering a future stock index, this paper corroborates its effectiveness with case studies. These investigations help to guide a right decision making for stock investments. In order to check the influence degree, this research provides three types models: a prediction model without past trends, a forecasting model considering past trend and a compensated trend based model. In order to test and compare them, different analyzing methods are applied.

The following section provides the related background and literature reviews. Section 3 describes three types of forecasting models and Section 4 introduces the applied analyzing methods. Then, numerical studies are provided in Section 5 for examining the effectiveness of a past trends' consideration in the stock prices prediction.

2 Background and Literature Reviews

In general, a stock data is considered a typical non-stationary data². Figure 1 shows a Korea Composite Stock Price Index (KOSPI) data which is acquired from the site (<http://global.krx.co.kr>) of Korea Exchange (KRX).

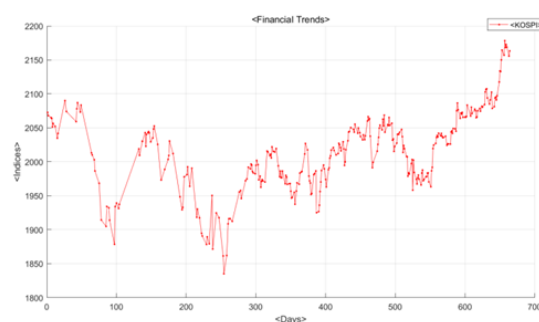


Figure 1: KOSPI Data from June 4th, 2015 to March 28th, 2017

As shown in Figure 1, the data is a heavily fluctuating and

volatile data. In order to analyze these kinds of data, Granger³ uses nonlinear function based approaches. However, these variables based regression models fails to incorporate overall various aspects which influence on the stock prices.

For this reason, time series based forecasting is applied for their forecasting. As usual time series analyses, Autoregressive model (AR), Moving Average model (MA) and its integration ARMA methods are applied for analyzing stationary data. Subha⁴ examines the stationarity on a stock data and attempt to analyze using these methods.

Autoregressive integrated moving average (ARIMA) model, Autoregressive conditional heteroscedasticity (ARCH) model and a generalized ARCH (GARCH) models are considered as representative analyzing methods for nonstationary data. Many research papers including Arowolo⁵ use these methods for analyzing stock data and for forecasting a future data.

As stock data incorporates many volatile changes, the preprocessing methods are required. The representative method is incorporating smoothing method. Hong and Lee⁶ uses moving average based smoothing methods. The usage of smooth methods helps to find a trend.

This research focuses on the effectiveness of the detected past trend in stock indices prediction.

3 Forecasting Model of Stock Data

This section provides analyses of the targeted stock data. As shown in Figure 1, the stock prices are fluctuating comparatively heavily. It indicates that it is difficult to forecast future stock data using stationary data based analyses such as AR, MA and ARMA. Even, several non-stationary data analyses (e.g. ARIMA, NARCH, GARCH) fail to forecast a future prediction comparatively accurately. One of the main reasons is the embedding trends in the data. For instance, a year' KOSPI data might have seasonal trends (e.g. semiannually, quarterly or monthly). Similarly, overall KOSPI data might have a yearly trend as shown in Figure 2.

This paper examines how yearly trend influence on the forecasting of stock data using several forecasting methods. As a target

data, KOSPI data with the same period (June 11th ~ December 28th) in 2015 and 2016 are considered. As shown in Figure 2, it seems that these data have a partial trend. In particular, the latter half of the data has a more explicit trend than the first half of the data.

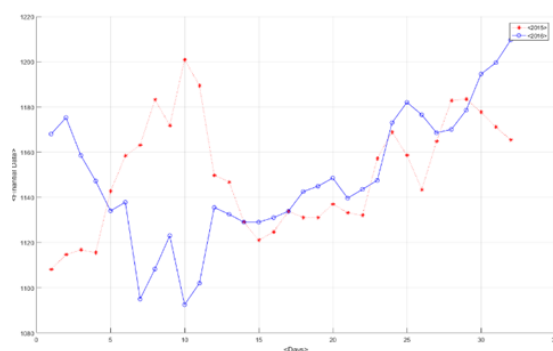


Figure 2: KOSPI Data from June 11th to December 28th in 2015 and 2016, respectively

In order to check the trend and its usages for a stock data forecasting, this paper tests three types of models: 1) a time series model without a yearly trend, 2) a time series model with a yearly trend and 3) a time series model with an adjust yearly trend. The first model is defined using (1).

$$X_{t+1} = \phi_1 \cdot X_t + \dots \phi_m \cdot X_{t-(m-1)} + \omega_{t+1} \quad (1)$$

ω_{t+1} = the white noise of the forecasting model

As shown in (1), the forecasting target index (X_{t+1}) at time is estimated using the sum of linear combinations with the previous data and their coefficients. (1) indicates a AR(m) model, where the white noise is modelled using a Gaussian distribution with mean 0. Its sigma is estimated using the average variance of the data. The second model is defined using (2) with a yearly trend.

$$X_{t+1} = \phi_1 \cdot X_t + \dots \phi_m \cdot X_{t-(m-1)} + \phi_1 \cdot T_{t+1} + \dots + \phi_n \cdot T_{t-n} + \omega_{t+1} \quad (2)$$

As defined in (2), T_{t+1} is the data at time (t+1) in the previous

year and ϕ_1 is its coefficient. The model integrates the first model and the trend data. As a trend data, the data from $(t+1)$ time to $(t-n)$ in the previous year is used. The reason for including T_{t+1} is resulted from the fact that the maximal likelihood trend data of X_{t+1} is T_{t+1} .

The final model is the variation of the second model with modified trend. For example, the trend is considered as relative data. As shown in (3), the current data and the previous data has a strongly positive trend as shown in figure 3.

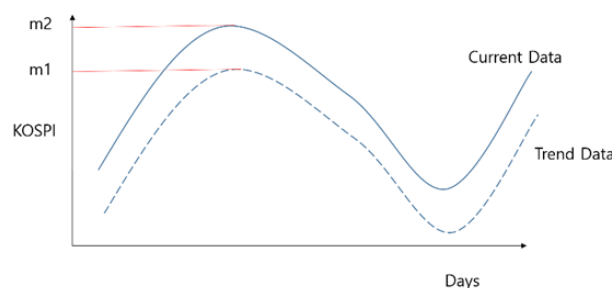


Figure 3: Stock Data with a relative trend data

In this situation, the compensation value $(m2/m1)$ is multiplied to the trend data, it closed to the current data. It indicates that the compensated trend data might advantageous for forecasting the future data. The final model (4) borrows this concept.

$$X_{t+1} = \sum_{i=t-(m-1)}^t \phi_i \cdot X_i + \theta \cdot \sum_{i=t-n}^{i=t+1} \varphi_i \cdot T_i + \omega_{t+1} \quad (4)$$

where θ = compensation value.

While (3) uses each x and each T , it is different from the fact that (4) handles each x and $\theta \bullet T$ for predicting a future data.

This research compares three types of forecasting models. The primal objective of this research study is to validate the usage of a yearly trend data for predicting a stock index. In order to estimate each parameters and forecasting the future data, three methods are applied. The following section provides the detailed features and characteristics of these methods.

4 Forecasting Methods and Analyses

As discussed in the previous section, this paper tests three types of forecasting model: AR model, AR model with trend data and a time-series model with compensated trend data. In order to find each modeling parameters and predict a stock index using them, three analyzing methods (Neural network, Partial least squares and Support vector regression) are applied. Neural network⁷ is a kind of network consisting input layer, output layer, hidden layer. The connection between two layers is achieved using weight vectors and an activation function. (5) shows a simple relationship between a input layer (X) and its output layer(Y).

$$Y_j = \phi \left(\sum_{i=1}^m (w_{ji} \cdot x_i) + \theta b \right) \quad (5)$$

where x_i = the i^{th} input data, y_j = the j^{th} output data, w_{ji} = the weight from x_i to y_j , b = bias vector, θ = the coefficient of bias vector and $\phi(\bullet)$ = activation function.

In order to determine each parameter of the neural network model, backpropagation method⁸ using Levenberg - Marquardt method⁹ is used. When a neural network is used for estimating the third model (4), the overall architecture is constructed as shown in Figure 4.

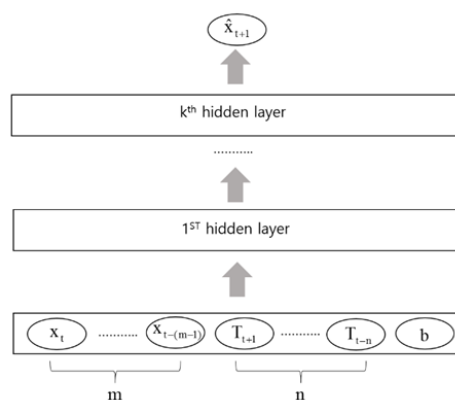


Figure 4: Multi-layer Perceptron model with past trend data in its input node

As shown in (4), its input node consists of three types of data: current AR(m) node, the trend data with the size n and its bias. The model is used for predicting the estimated x_{t+1} .

The second analyzing method is Partial Least Squares (PLS) method. The primal usage of PLS method is a projection method using latent variables to a lower dimension from a high dimension with process variables (k) and quality indices (m). The detailed algorithm and usages are provided in Rosipal and Kramer¹⁰ and Mevik and Wehrens¹¹.

The final comparing analysis is Support Vector Regression (SVR) which is driven from Support Vector Machine (SVM). As a common SVM is used for classifying a supervised data, SVR is considered as an integrated model with SVM and a regression method. This paper uses epsilon-SVR¹² and nu-SVR¹³ among many SVR method, for predicting x_{t+1} .

The following section provides the test results and comparisons with three types of predicting models and three analyzing methods.

5 Numerical Experiments and Analyses

This section provides numerical experiments with the target data which is described in Section 3. Table 1 provides the parameters of three types of predicting model.

Table 1: Parameters for each predicting model

Content	Prediction model 1	Prediction model 2	Prediction model 3
Parameters	- m = 3	- m=3 - n = 3	- m=3 - n = 3 - θ =0.9985

As provided in Table 1, the first prediction model is considered as an AR(3) model. In the second model, the same size of trend data are used for predicting x_{t+1} with AR(3). The compensation value (0.9985) is calculated with the comparison between the trend data and current data. Then, it is multiplied to the trend data for balancing with the current data.

Table 2 provides control parameters and detailed methods of three analyzing techniques.

Table 2: Control parameter and detailed method of each analyzing technique

Analyzing method	MLP	PLS	SVR
Parameters and Methods	- Number of hidden layer : 1 - Nodes in a hidden layer : 4 - Activation function : TanSig - Optimizing method : Levenberg Marquardt method - Number of iterations : 100	- Number of latent variables : 1	- SVR type : Epsilon – SVR - Kernel type : RBF - Gamma value : 0 - C : 1 - Epsilon : 0.1

In order to test and analyze them, a big data analyzer, ECMiner is used as shown in Figure 5. As three analyzing methods use a training data, 70% of overall data is used as a training data and 30% of data is used as a test data.

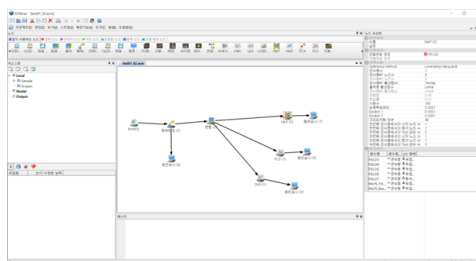


Figure 5: Tests and analyses using ECMiner©

The predicting performance is measured using (6).

$$error_{total} = \sum_{i=1}^{|N|} |X_i - \hat{X}_i| \quad (6)$$

Figure 6 shows a plot of total errors using three analyzing methods with the first model - AR(3) model. As shown in Figure 6, SVR is considered as an analyzer with the lowest forecasting error.

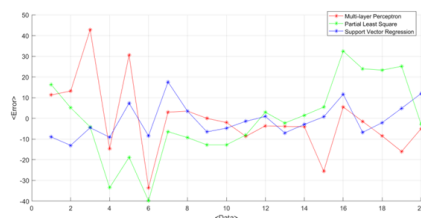
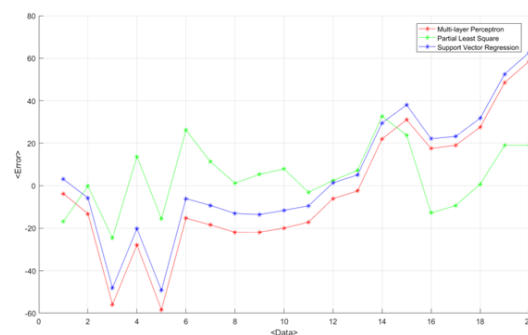
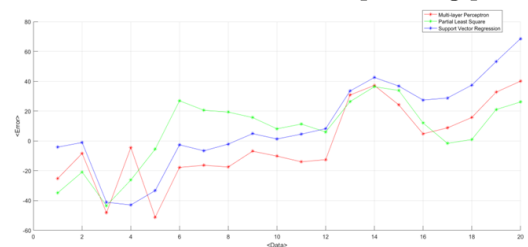


Figure 6: Total predicting errors of AR model

Figure 7(a) and 7(b) show the total errors of the second (AR model incorporating past trends) and the third model (AR model with compensated trends).



(a) Total errors of AR model incorporating past trend



(b) Total errors of AR model with compensated past trend

Figure 7: Total predicting errors considering past trend

The overall total errors are provided in Table 3. As provided in Table 3, PLS method is considered as a best analyzing method with respect to the total errors.

Table 3: Total error of each prediction model using three analyzing methods

Content	Total error using MLP	Total error using PLS	Total error using SVR
AR model	11.835	14.335	6.7037
AR model incorporating trend data	25.3348	12.6569	22.7908
AR model with compensated trend data	21.3402	19.8626	24.0640

However, the performances of methods are changed by each analyzed model. For instance, SVR is considered as the most suitable

model for analyzing the first model. It indicates that the consideration of trend data influences on the forecasting methods and their results very sensitively. One of most important results is the fact that the consideration of trend data leads to much worse forecasting results in this research. It indicates that past trend fail to characterize the current trend successfully. Then, the recent data has more valuable for forecasting future stock data.

6 Conclusion

This paper examines the effectiveness of past trends for forecasting a future stock data. While the consideration of trends might be helpful for more accurate predictions, this research corroborate that a recent data is more important and trend data have a negative impact on the prediction. In order to prove the corroboration, three types of prediction models are provided: AR model, AR model with past trend and AR model incorporating compensated trend data. The compensation is calculated using the comparisons between current average stock index and past average stock index.

In order to test these model, three different analyzing methods are applied. As results, PLS method is considered as a good analyzers in general. However, the performances are different in each analyzed model. The important fact is that a better forecasting can be achieved without the consideration of past trends. It indicates that it is difficult to extract the meaningful trend in stock data and the recent data based approaches are used for making right decision in stock investments. This research study corroborates this examination.

As further studies, the more segmented trend analyses are expected. While a past trend has to be considered carefully in the prediction of stock index, a partial incorporation of trend data might help to characterize current data. And, many other analyzing methods are expected to be applied for accurate forecasting.

7 Acknowledgment

This paper was supported by Research Fund, Kumoh National Institute of Technology.

References

- [1] Arsham H, Shao JR S P, *Seasonal and Cyclic Forecasting for the Small Firm*, American Journal of Small Business, 1985, 11(4), pp.46-57.
- [2] Murthy V, Washer K, Wingender J, *Are Stock Prices in the US nonstationary? Evidence from Contemporary Unit Root Tests*, Journal of Applied Financial Economics, 2011, 21(1), pp. 1703-1709.
- [3] Granger C W J, *Forecasting Stock Market Prices: Lessons for Forecasters*, 1992, 8(1), pp. 3-13.
- [4] Subha M V, *A Study on Stationarity of Global Stock Market Indices*, Journal of Contemporary Research in Management, 2010, 1(1), pp. 41-46.
- [5] Arowolo W B, *Predicting Stock Prices Returns using Garch Model*, International Journal of Engineering and Science, 2013, 2(5), pp. 32-37.
- [6] Hong H, Lee H, *Trend Analyses of KOSPI Data using Moving Average based Smoothing Filter*, The 4th International Conference on Digital Policy and Management, 2017, pp. 535-536.
- [7] Haykin S O, *Neural Networks and Learning Machines*, 3rd , Pearson, 2008, pp. 1-936.
- [8] Johansson E M, Dowla F U, Goodman D M, *Backpropagation Learning for Multilayer Feed-forward Neural Networks using the Conjugate Gradient Method*, International Journal of Neural Systems, 1991, 2(4), pp.291-301.
- [9] Suratgar A A, Tavakoli M B, Hoseinabadi A, *Modified Levenberg-Marquardt Method for Neural Networks Training*, International Journal of Computer, Electrical, Automation, Control and Information Engineering, 2007, 1(6), pp.1745-1747.
- [10] Rosipal R, Kramer N, *Overview and Recent Advances in Partial Least Squares*, Lecture Notes of Computer Science, 2006, 3940(1), pp. 34-51.

- [11] Mevik B, Wehrens, The pls Package: *Principal Component and Partial Least Squares Regression in R*, 2007, 18(2), pp.1-23.
- [12] Wang G, Yeung D, Lochofsky, F H, *A New Solution Path Algorithm in Support Vector Regression*, IEEE Transactions on Neural Network, 2008, 19(10), pp. 1753-1767.
- [13] Chang C C, Lin C, *Training Epsilon-Support Vector Regression: Theory and Algorithm*, Neural Computation, 2002, 14(8), pp. 1959-1977.