2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th Intl. Conference on Data Science and Systems

Stock price forecasting incorporating market state

Erhan Beyaz, Firat Tekiner, Xiao-jun Zeng, John A. Keane
School of Computer Science
University of Manchester, Manchester, UK
erhan.beyaz@postgrad.manchester.ac.uk

Abstract— **This paper evaluates the impact of identifying and accounting explicitly for the states of the stock market when forecasting stock prices using machine learning. Clustering methods are applied to market mood indicators (e.g. VIX) to identify various states of the stock market and forecasting models for the selected companies are developed for each market state. Tests using 85 companies from the S&P 500 indicate that, for a forecasting horizon of 126 days, accounting for the market mood improves forecasting accuracy (lower RMSE) in 47% of cases. This improvement from accounting explicitly for market states is also seen when only Technical indicators or Fundamental indicators are inputs to the forecasting models; the benefit decreases when these indicators are combined.**

Keywords—Clustering; kmeans++; ANN; SVR; Technical Analysis; Fundamental Analysis; Stock Price Forecasting; Market States

I. INTRODUCTION

In general, stock prices are determined by forces of supply and demand in the stock market which in turn is driven by traders' decisions to buy or sell a company's stock [1]. Technical and Fundamental analysis are the two main approaches used by Finance practitioners in forecasting stock prices and making trading decisions [1].

Stock markets tend to go through a variety of moods [17-18]. Various market sentiment indicators are used to measure the states of the market. One such mood indicator is the Volatility Index (VIX), also known as the fear index, which is widely used to represent the expectations of the stock market, where a high VIX means an expectation of high volatility in the stock prices and vice versa [23]. The prices of individual stocks are impacted by the market moods. For example, even if the main determinants of the company's financial performance might not have changed, the stock price of a company could be reduced because of a broad decline in the overall stock market. Stock prices are therefore intrinsically linked to movements in the overall market.

Machine learning methods have been applied to stock price forecasting [5-7]. A challenge presented by financial timeseries data is that it is non-stationary, where its statistical properties tend to change over time [5,21]. Among many other things this change over time can be driven by economic changes. Approaches to address its non-stationary nature have involved using the financial timeseries data of the company in question and apply clustering techniques to identify the point at which to recalibrate the forecasting model [19-20] to accommodate for the changes. However, these approaches tend to rely on the data of the company that is being forecast and not necessarily try to link it to the overall economic movements or moods of the stock market.

To address this opportunity, this paper presents a set of experiments which compares the performance of forecasting models accounting for the states of the overall stock market to forecasting models that do not, when forecasting change in the stock price of individual companies over a 126 and a 252 day horizon respectively. The outcome is an analysis of the benefits of identifying and considering the moods exhibited by the stock market when forecasting stock prices with machine learning models.

The rest of this paper is organized as follows: Section II reviews machine learning techniques in stock price forecasting; Section III describes the experimental setup; Section IV discusses results; finally, Section V presents conclusions.

II. BACKGROUND AND APPROACH

A. *Machine learning methods in stock price forecasting*

1) Inputs

The type of inputs used in forecasting stock prices are, to a large extent, dependent on the underlying investment analysis approach: Fundamental or Technical Analysis [5-7]. Technical analysis uses the historical stock prices of a company and trading volume information in both deciding what the stock price will be and making a trading decision [1, 3]. Fundamental analysis calculates the expected stock price based on a detailed study of the underlying business drivers (profitability, operational efficiencies, managerial expertise, etc.) related to the company, its products, its industry and the general economy [2,4]. Historically, work using machine learning methods has shown clear preference towards using Technical indicators as inputs [5-7]. This apparent over-reliance on using Technical indicators as inputs is attributed to the fact that machine learning approaches depend on large volumes of data and that Technical indicators were more available and easily accessible, especially on a daily basis; in contrast, Fundamental indicators become available on a less frequent basis (quarterly or yearly) and as such have been much less preferred. Most studies reviewed in the surveys [5-7] used Technical indicators only, a few used Fundamental indicators only, and a handful looked at combining the two to some extent.

2) Supervised Machine Learning Methods:

Due to non-linear nature of the financial forecasting problem, Neural Networks (NN) and Support Vector Regression (SVR) have been effectively utilized in the area [5-

978-1-5386-6614-2/18/$31.00 ©2018 IEEE
DOI 10.1109/HPCC/SmartCity/DSS.2018.00263

7]. In terms of a broad categorization, the majority of machine learning approaches using labeled data and supervised learning methods tend to fall into three main groups [5]: models that use a single machine learning technique, models that use a hybrid combination of techniques with optimization approaches, and models that are an ensemble of various single models.

3) Stock Market States and financial forecasting:

The stock market is known to exhibit Bull and Bear states, which are periods of "upward and downward trends of stock index or positive and negative stock index returns over a period of time" [17]. The relationship between stock price and its drivers tend to change over time [5,21], as the financial markets evolve over time. For example, according to [5], "The time series of stock prices of a company may change its behaviour due to changes in political and economic factors or due to changes in the investor psychology or expectations." This results in financial timeseries being non-stationary. Thus, identifying the drivers of stock valuation and developing static models and weights based on these inputs is inefficient, as the dynamic nature of the market makes it difficult to have one model that is always valid.

Such concept drift occurs in financial time series data [19]; one approach to addressing this is to recalibrate the models on a pre-determined basis (i.e. implicit); the other approach is to have a trigger (i.e. explicit) which "monitor some statistics of the data stream in order to detect concept drifts" [20]. Work in [19] simulated an approach with use of such an explicit drift detection in updating the NN resulted in improved speed whilst maintaining accuracy. Clustering methods, Perceptually Important Points (PIP) or Turning Points (TP) have been used for such segmentation of time series data to serve as triggers for concept drift [5]. Work in [21] used a two-layered, "divide and conquer" approach, where Self-Organizing Maps (SOMs) are used as an initial filter to break the historical price data into groupings that show similar characteristics. Once these groupings were created, an SVR was applied to each to establish the relationship between the independent variables (lagged closing prices) and the dependent variable (relative price change 5 days into the future). The results showed that the two-layered approach of SOM + SVR was superior to using SVR by itself, consistently across the 7 major market indices.

B. *Research Questions*

As covered in Section II-A, clustering techniques have been used to tackle the non-stationary nature of financial time series data [5,19-21]. However, these approaches have been limited to targeting the fluctuations exhibited within the stock price forecasted, rather than targeting the sensitivity of the stock in question to the various states of the market. The approach proposed in this paper is inspired by the existing clustering approaches, but rather seeks to establish a link between the overall state of the stock market and stock price forecasting. Given the dynamic nature of the stock market, we investigate whether identifying and accounting for the various states of the stock market within the stock price forecasting process would improve forecasting. This analysis therefore seeks to answer the following questions:

- How to capture the various moods / states of the overall stock market?
- What is the impact of accounting for the sensitivity of the stock price to various states of the market?

III. EXPERIMENTAL SETUP

Experiments were designed to predict the percentage change in a selected company's stock price in the future (252 and 126 trading days out), using NNs and SVR forecasting models exposed to Technical indicators, Fundamental indicators and their combination. Furthermore, market moods were captured by applying clustering to market sentiment indicators (as defined in [23]) Volatility Index (VIX) and Put-Call (P2C) Ratio, and forecasting models were developed based on the market state exhibited. For the forecasts generated, RMSE was calculated and captured by comparing against the actuals. The experiments were implemented using R [8] and the open source data mining tool WEKA [9] and involved 85 companies selected from the S&P 500 index[1], which represents 80% of US equities market by capital. Companies were selected based on having sufficient data available for a period of time (identified as Jan 1996-Dec 2015) which ensured incorporation of times where the stock market exhibited a variety states (turbulences, ups and downs, etc.).

A. *Input Data:*

For each company two sets of financial data, Technical and Fundamental, were collected and a daily financial timeseries data was created. Finally, these two sets were merged by using the dates to create a combined data set.

1) Technical Indicators:

For each company in the study the end of day stock price data (Open, High, Low, Close, Volume) was retrieved from Quandl[2] and the TTR package [10] was used to generate the 10 technical indicators. To address the issue of any missing data, the average of the data from the closest available trading days was used, as in [11]. Table I gives the list of technical indicators picked based on the coverage of Technical analysis in [1] and [13] and the parameters used (mainly defaults in [10]) in generating them where relevant.

TABLE I. TECHNICAL INDICATORS

Average True Range (ATR) over a period of 14 days.
Moving Average Convergence Divergence (MACD) with simple moving average method and 26 days & 12 days for the slow and fast periods respectively.
Money Flow Index (MFI) over a period of 14 days.
FastK and FastD values of Stochastic Oscillator using 14,3, and 3 days for FastK, FastD, SlowD respectively.
Directional Movement Index (DMI) using 14 days

[1] List retrieved from https://us.spindices.com/indices/equity/sp-500

[2] Price data retrieved from Quandl (https://www.quandl.com/product/WIKIP/WIKI/PRICES-Quandl-End-Of-Day-Stocks-Info)

Commodity Channel Index (CCI) using 20 days, and 0.015 as the constant to apply to the mean deviation.
Relative Strenght Index (RSI) using 14 days and weighted moving average.
Price Rate of Change (ROC) over 252 or 126 trading days.
The Chaikin Accumulation / Distribution (AD) line.

2) *Fundamental Indicators:*

Following [9], the fundamental indicators used in the experiments can be categorized into groups as follows: those relating to the performance of the company in question, those related to direct competitors, those related to the industry to which the company belongs, and macroeconomic indicators. The top two direct competitors of each company were identified based on market capitalization and picked out of the list generated by the Thomson One[3] database. The IBES[4] database contains the monthly forward-looking forecasts by financial analysts on companies as well as their recommendations on buying/holding/selling the stock. The median of the monthly estimates by financial analysts for Earnings per share (EPS) 1 year and 2 years out, as well as long-term expected growth percentage in EPS, were retrieved for each company and their competitors. Missing values were dealt with by using "the last observation carry forward" method [11]. As these estimates were only available on a monthly basis, their frequency was converted to daily by dividing it by the prior day's stock closing price data.

The industry designation for the company was determined using the classification available on the Yahoo Finance website[5]. The daily index price data for each corresponding industry was retrieved from the MSCI[6] website which were transformed into moving average convergence and divergence (MACD) indicators for short term (with 26 days and 12 days) and medium term (with 126 days and 12 days). Apart from these company-related data, macroeconomic indicators[7], which were the same for all companies in the study, were used. One such indicator was based on the foreign currency data, where the daily value of "Trade Weighted U.S. Dollar Index against Major Currencies" was transformed with MACD (126 days and 12 days). Another macroeconomic indicator used is the "S&P 500 futures data" whose daily value was transformed further using MACD (26 days and 12 days). The final macroeconomic indicator used was derived from the ratio of 10 year to 2 year constant maturity rate which was transformed using MACD (26 days and 12 days). Table II gives the list of fundamental indicators used.

[3] Thomson One, Retrieved from Wharton Research Data Service, 2016

[4] I/B/E/S, Retrieved from Wharton Research Data Service, 2016

[5] Yahoo Finance, available at: https://finance.yahoo.com/ [Accessed September 2016]

[6] MSCI USA IMI SECTOR INDEXES, available at: https://www.msci.com/msci-usa-imi-sector-indexes [Accessed September 2016]

[7] Retrieved from Federal Reserve Economic Database (FRED), available: https://fred.stlouisfed.org/series/DTWEXM [September 2016]

TABLE II. FUNDAMENTAL INDICATORS

Earnings Per Share (EPS) 1 year out for the company / Price
(Earnings Per Share (EPS) 2 years out) / (Earnings Per Share (EPS) 1 year out)
EPS long term growth rate percentage
Earnings Per Share (EPS) 1 year out for competitor 1 / Price for competitor 1
Earnings Per Share (EPS) 1 year out for competitor 2 / Price for competitor 2
Daily MSCI industry index prices (MACD, 252 days, 12 days)
Daily MSCI industry index prices (MACD, 26 days, 12 days)
S&P 500 Futures prices (MACD, 252 days, 12 days)
Daily Trade Weighted U.S. Dollar Index against Major Currencies (MACD, 252 days, 12 days)
10 year to 2 year constant maturity rate (MACD, 26 days, 12 days)

The experiments use two different forecasting horizons: 126 days (i.e. 6 trading months) and 252 days (i.e. 1 trading year). The input sets used for the different horizons were adjusted to ensure that the inputs stayed relevant. For example, the Price Rate of Change (ROC) of the technical indicators is calculated for 252/126 days depending on the forecasting horizon. Similarly, Fundamental indicators would be adjusted where the formulae for industry and macroeconomic indicators would match the forecasting horizon.

B. *Additional experiment details*

Following [12], the data was split into 80% training data and 20% testing data. A standard way to ensure robustness is to use K-fold cross validation [14]. However, as K-fold cross validation requires random sampling to form the test and training sets, it is unsuited for financial time-series forecasting, in which it is important to separate testing data from training data such that chronological order of data is preserved [15]. This ensures that the model is not prematurely exposed to information in the training phase (look ahead bias), potentially producing unrealistically good performance. Therefore, during simulations testing data was used chronologically after training data. Following [15], to ensure robustness 10 random starting points were generated, and from each starting point available data was split into training and test sets. Each random starting point resulted in 1892 training set observations and 473 test set observations. Fig. 1 overlays the 10 random points against the overall stock market performance; it can be seen that some test start points (marked in x) occur during market up swings, and others occur during market down turns.

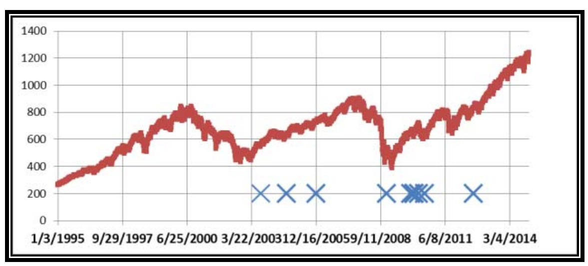

Fig. 1. Random Test points generated against backdrop of overall market

As covered in Section II, NN and SVR models were applied to similar problems, and have been selected for the experiments. To determine the parameters of the models, the training sets were further split into training and validation sets (again using the 80-20 split and preserving chronological order) and the model parameters which yielded the lowest forecasting error rate (RMSE) on the validation sets were set aside. Work in [16] determined the number of hidden neurons in NN by starting from a small number (the square root of the number of features) and incrementing until the performance of the model no longer improved. To configure the NN model, the ideal number of hidden neurons, the learning rate and momentum rate were decided based on validation set performance across the scenarios listed in Table III.

TABLE III. ANN PARAMETERS

# of Hidden Neurons	3,5,7
Learning Rate	0.05, 0.3, 0.6
Momentum	0.1, 0.3,0.7

For the SVR, the C and gamma values have been tested over several different scenarios, as shown in Table IV, to determine the optimum model calibration:

TABLE IV. SVR PARAMETERS

C Values	0.125, 0.5 , 2
Gamma Value	0.01953, 0.125 ,0.5 , 1

C. *Accounting for the states of the overall stock market*

The inclusion of the market states within the forecasting process was done by taking a different approach towards how the composition of the training set is determined. Models that did not account for market states (No State) used the most recent time series information (1892 instances) as the training data; models accounting for the state of the market (With State) used the training set that is formed by considering the state of the market. The forecasting models were trained and tested for each testing instance (i.e. model recalibration 1 day window). To account for the state of the market, firstly the date of the testing instance is used to retrieve the values of the market sentiment indicator on that particular date and back to the first available date (January 1st 1996). The value of the market sentiment indicator on the dates prior to the testing date are input to a clustering algorithm to create the various moods that the overall stock market exhibited. The next step is to identify the cluster to which the market sentiment indicator from the testing date belongs, thereby identifying the "active" state or the mood of the market on that date. All the dates from the market sentiment indicator training data set belonging to this active state of the market are mapped to the instances from the training set of the company's input set. This effectively filters the training set to contain only instances where the market mood exhibited is the same. Fig. 2 shows the approach taken towards capturing and reflecting the market states in the stock price modeling.

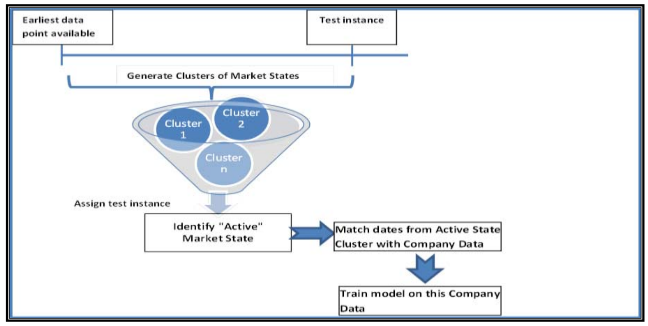

Fig. 2. Approach for explicitly accounting for the state of the stock market

This approach does result in a non-uniform set of instances in terms of size of the training set. In order to keep the training set size the same throughout the experiments, the training set generated by the active state layer had to be further adjusted to match the training set size (e.g. 1892). In the cases where the size of the training set matching the active market state was larger than the training set used in the remainder of the experiments, the suggested training set was reduced to the training set size used in the remainder of the experiments where the most recent portion of the data was kept. In the cases where the reverse is true, the training data matching the active market state was repeated until the training set size reached the same level as used in the remainder of the experiments. To capture the various moods / states of the market, VIX and Put-Call Ratio (P2C) used market sentiment indicators [23]. The state layer was implemented by applying "kmeans++" clustering [22]. One parameter decision to be made was the number of clusters. Markets can generally be considered as exhibiting an upwards trend (e.g. "Bull"), or downwards trend (e.g. "Bear"), or stationary (i.e. side-way movements). Using this simplistic view of market moods as a starting point and based on the number of states observed in [17] and [18], the number of clusters used was 3, 5 and 7. For each company input (Technical, Fundamental, Combined) and machine learning combination (ANN, SVR), forecasting models with a state layer using cluster sizes of 3, 5, and 7 were implemented.

IV. RESULTS

A. *Experiment 1: VIX vs P2C*

Experiment 1 compared the relative effectiveness of VIX and P2C in being able to capture the various moods of the overall stock market. Forecasting model performances were compared across input types (Technical, Fundamental, Combined) and machine learning methods (NN, SVR) and for forecasting horizons of 126 days and 252 days respectively. In the overall, VIX (RMSE of 0.12889 and 0.16722 for 126 days and 252 days) and P2C (RMSE of 0.13211 and 0.16745 for 126 days and 252 days) outperformed each other in 50% of the cases. Fig. 3 displays the average RMSE per industry when using VIX and P2C for 126 days and 252 days forecasting.

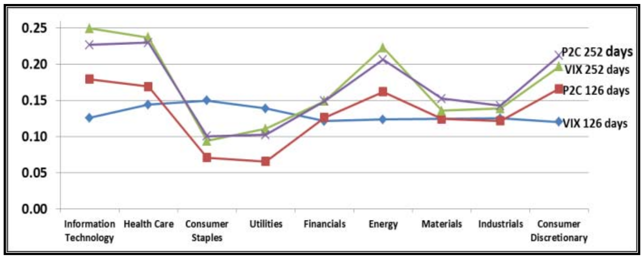

Fig. 3. Average RMSE of models using VIX and P2C to capture the market states

When the forecasting horizon is 252 days, the performance of VIX and P2C follow each other closely regardless of the industry. However, when the forecasting horizon is 126 days, the models using VIX to capture market states tend to outperform models using P2C for the companies operating in Information Technology, Health Care, Energy and Consumer Discretionary. The reverse is observed for companies operating in Consumer Staples and Utilities sectors.

Another analysis was drawn with regards to the impact of the assumption with regards to the number of states / moods, as represented by the number of clusters (3,5, and 7) used. Table V shows the average RMSE for each market mood indicator and forecasting horizon, as well as the % of cases where a specific number of clusters outperformed the rest. In the overall models using 3 states for the market mood indicator slightly outperformed the others (% of Best).

TABLE V. COMPARISON OF NUMBER OF CLUSTERS

Indicator	Avg. RMSE			% Best		
	3	5	7	3	5	7
VIX_126 days	0.1289	0.1282	0.1296	40	31	30
P2C_126 days	0.1340	0.1312	0.1311	32	30	38
Overall 126 days	0.1315	0.1297	0.1303	36	31	34
VIX_252 days	0.1676	0.1659	0.1682	35	35	30
P2C_252 days	0.1639	0.1700	0.1684	38	26	37
Overall 252 days	0.1658	0.1679	0.1683	36	30	34

B. Experiment 2: No State vs. With State

Experiment 2 compared the relative performance of the models explicitly accounting for the states of the overall stock market (With State) against those that did not (No State). For this part of the analysis best performing (lowest RMSE) market mood indicator and number of state layer definition for each company were used in defining the forecasting models "With State". In the overall, when the forecasting horizon was 252 days, models that did not account for the market states explicitly outperformed those that did in 82% of the cases considered and achieved a lower RMSE on the average (0.0696 vs. 0.0616). However, when the forecasting horizon was 126 days, the models that accounted for the market states explicitly outperformed those that did in 47% of the cases considered and achieved a lower RMSE on average (0.0476 vs. 0.0550). Therefore, the impact of the various states of the overall stock market are more pronounced when the forecasting horizon is 126 days and less so for longer forecasting horizons.

Further review was taken on a business sector level to investigate whether companies in certain industries were more sensitive to moods of the market than others. Fig. 4 displays the average RMSE per industry for a 126 days forecasting horizon. Companies in the Consumer Staples, Utilities, and materials industries are not sensitive to the states of the economy, and appears reasonable as these are all industries where consumers cannot shrink their spending on regardless of the market state in the short term. On the other hand, companies in the Information Technology, Health Care, Financials, Energy, and Consumer Discretionary are sensitive to the states of the economy.

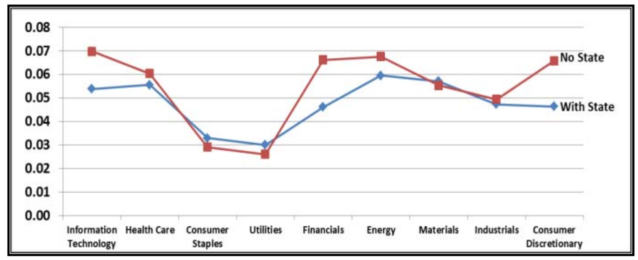

Fig. 4. Average RMSE of models With State against No State for 126 days forecasting

As covered in Section II, machine learning researchers tend to use technical indicators over Fundamental indicators in the domain of financial forecasting. Another analysis was drawn comparing the performance of the models accounting for the market states against those that did not from the perspective of the inputs utilized. Table VI compares the average RMSE of models "With State" against "No State" when using Technical indicators, Fundamental indicators, and a combination of the two when forecasting for 252 days. Table VI indicates that regardless of the industry to which a company belongs, models using solely Technical or Fundamental indicators do benefit (i.e. achieve lower RMSE) from explicitly accounting for the states of the market in the forecasting process. Table VI only displays results for 252 days forecasting yet identical observations were made in the case of 126 days forecasting. This provides additional insight to our earlier analysis that accounting for the states of the market becomes less impactful than forecasting horizon, in that the model inputs used are also a factor. When generating a forecast for 252 days, the synergy achieved by using a combination of Technical and Fundamental indicators does surpass the benefits that are achieved by explicitly accounting for the states of the market.

TABLE VI. AVERAGE RMSE BY INPUT TYPE OF MODELS THAT ACCOUNT FOR MARKET STATE (YES) AND MODELS THAT DO NOT ACCOUNT (NO)

Sector (# of Comp.)	**252 Days**					
	Technical		***Fundamental***		***Combined***	
	YES	***NO***	***YES***	***NO***	***YES***	***NO***
Inform. Technology (6)	0.2114	0.2425	0.1707	0.1810	0.0906	0.0775
Health Care (8)	0.2281	0.2848	0.1562	0.1600	0.0939	0.0736
Consum. Staples (9)	0.0922	0.1023	0.0560	0.0591	0.0350	0.0279
Utilities (6)	0.0811	0.0909	0.0550	0.0711	0.0331	0.0285
Financials (11)	0.1396	0.1716	0.1080	0.1180	0.0702	0.0626
Energy (9)	0.1902	0.2221	0.1768	0.2211	0.0890	0.0795
Materials (7)	0.1354	0.1581	0.0951	0.1053	0.0603	0.0546
Industrials (17)	0.1360	0.1652	0.0981	0.1198	0.0605	0.0585
Overall	**0.1851**	**0.2334**	**0.1502**	**0.1781**	**0.0890**	**0.0821**

V. CONCLUSION

Machine learning methods have been successfully applied to learn from past movements of a company's stock price and generate future forecasts. A challenge in financial timeseries forecasting is caused by the non-stationary nature of this data where the relationship between the drivers and forecasting models changes over time. Stock markets exhibit various moods and these moods have a bearing on the stock prices of the individual stock prices. This work has investigated the impact of accounting explicitly for the states of the stock market within the stock price forecasting approach.

Our experiments involving 85 companies from S&P 500 have shown that overall the benefit of accounting for the states of the market tend to decrease as the forecasting horizon increases (252 vs. 126 days). With a forecasting horizon of 126 days, approx. 47% of the cases in the study improved by accounting for the market states; this was more pronounced in specific industries such as Information Technology, Health Care, Financials, Energy, and Consumer Discretionary. Furthermore, our experiments have shown that the set of inputs used are influential as to whether accounting for the states of the market was beneficial. When using Technical indicators or Fundamental indicators solely as inputs to the models, the forecasting performance improves from accounting for the state of the market regardless of the forecasting horizon undertaken or the industry to which the company belongs. This improvement disappears when a combination of the Technical and Fundamental indicators is used.

REFERENCES

[1] M. C. Thomsett, Getting Started Getting Started in Stock Analysis, Illustrated Edition. Hoboken, Singapore: John Wiley Publishing, 2015.

[2] M. Krantz, Fundamental Analysis For Dummies. Indianapolis, Indiana, USA: Wiley Publishing, 2016.

[3] B. Rockefeller, Technical Analysis For Dummies. Wiley Publishing, Indianapolis, 2011.

[4] M. C. Thomsett, Getting Started in Fundamental Analysis. Wiley Publishing, Hoboken, 2006.

[5] C. R. Cavalcante, C. R. Brasileiro, L. F. V. Souza, P. J. Nobrega and A. L. Oliveira, "Computational Intelligence and Financial Markets: A Survey and Future Directions," Expert Systems with Applications, vol. 55, pp. 194-211, 2016.

[6] G. S. Atsalakis and K. P. Valavanis, "Surveying stock market forecasting techniques - Part II: Machine learning methods". Expert Systems with Applications, vol. 36, pp. 5932-5941, 2009.

[7] B. Krollner, B. Vanstone and G. Finnie "Financial time series forecasting with machine learning techniques: A survey". Paper presented at the European symposium on artificial neural networks: Computational and machine learning, Bruges, 2010.

[8] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/ (2013).

[9] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H: The WEKA Data Mining Software: An Update. *SIGKDD Explorations.* 11(1) (2009)

[10] J. Ulrich, J, TTR: Technical Trading Rules. R package version 0.23-0. http://CRAN.R-project.org/package=TTR, 2015.

[11] J. P. Romero and T. Balch, What Hedge Funds Really Do: An Introduction to Portfolio Management. Business Expert Press, 2014.

[12] R. J. Hyndman and G. Athanasopoulos, Forecasting: principles and practice. OTexts, 2014.

[13] J. Patel, S. Shah, P. Thakkar and K. Kotecha, "Predicting stock market index using fusion of machine learning techniques," Expert Systems with Applications, vol. 42 (4), pp. 2162–2172, 2015.

[14] M. Kuhn and K. Johnson, Applied Predictive Modeling.Springer, 2013.

[15] L. Torgo, Data mining with R : learning with case studies, 2nd Edition. Chapman & Hall/CRC, Florida, 2017.

[16] B. Vanstone and G. Finnie, " An empirical methodology for developing stockmarket trading systems using artificial neural networks," Expert Systems with Applications, vol. 36 (3), pp. 6668-6680, 2009.

[17] Jiang Y. Jiang, X. Fang, "Bull, bear or any other states in the us stock market?" Economic Modelling, vol. 44, pp. 54–58, 2015.

[18] M. Munnix, S. Takashi, S. Rudi, F. Leyvraz, T. Seligman, T. Guhr, E. H. Stanley, "Identifying States of a Financial Market". Scientific reports, vol. 2., pp 644, 2012.

[19] R. C. Cavalcante, A. L. Oliveira, "An approach to handle concept drift in financial time series based on Extreme Learning Machines and explicit Drift Detection," International joint conference on neural networks(IJCNN), 1–8, 2015.

[20] R. C. Cavalcante, A. L. Oliveira, "FEED: Feature Extraction for Explicit Concept Drift Detection in Time Series," International joint conference on neural networks (IJCNN), pp. 740-747, 2016.

[21] S. H. Hsu, "A two-stage architecture for stock price forecasting by integrating self-organizing map and support vector regression," Expert Systems with Applications , vol. 36 (4), pp. 7947-7951, 2009.

[22] D. Arthur, S. Vassilvitskii, "k-means++: the advantages of careful seeding," Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics Philadelphia, PA, USA. pp. 1027–1035, 2007.

[23] S. B. Achelis, Technical Analysis from A to Z, 2nd Edition. McGrawHill,2013.