Innovative Combinational Analytics to Predict Stock Market

Karthikeyan. M,[1] Dheepan Cross. F,[2] Krishnan Nallaperumal[3] and K. Senthamarai Kannan[4]

Abstract

This paper deals with the Technical Analysis which helps the investors to discover hidden patterns from the historic financial and time series stock market data. These techniques have the probable capability to help the investors in their investment decisions by looking into the new hidden patterns and opportunities. It maximizes the prediction of financial stock market using time series quantitative analysis. Two analytics are devised by combining various financial indicators. A back test on the historic data is performed to calculate the percentage of profitable buy / sell signals generated by the combinational analytics. The Moving Average Crossover (MAC) Algorithm is taken as a benchmark to compare the percentage of profitable buy / sell signals generated by the proposed combinational algorithms. Both the proposed algorithms outperform the MAC algorithm.

[1]Research Scholar, the Centre for Information Technology and Engineering, Manonmaniam Sundaranar University, Tirunelveli-627 012, India.Email: karthikn_m@yahoo.co.uk

[2]Research Scholar, the Centre for Information Technology and Engineering, Manonmaniam Sundaranar University, Tirunelveli-627 012, India.

[3]Professor and Head, the Centre for Information Technology and Engineering, Manonmaniam Sundaranar University, Tirunelveli-627 012, India.

[4]Professor, Department of statistics, Manonmaniam Sundaranar University, Tirunelveli-627 012, India

Keywords : Financial Indicators, Moving Average Crossover Algorithm, Quantitative Analysis, Technical Analysis, and Time Series Data.

1. Introduction

The Competitive business pressures and a desire to leverage existing information technology investments have led many financial investment firms to explore the benefits of data mining technology. As of now, more than 70 per cent of trades in the US stock market are done by automated computer programs using data mining and predictive technologies.

Many fund management firms have invested heavily in information technology to help them manage their financial portfolios. Over the last three decades, large amounts of historical data have been stored electronically and this volume is expected to continue to grow considerably in the future. Yet despite this wealth of data, many fund managers have been unable to fully capitalize on their value. This is because information that is implicit in the data for the purpose of investment is not easy to discern.

For example, a fund manager may keep detailed information about each stock and its historic data but still it is difficult to pinpoint the subtle buying patterns until systematic explorative studies are conducted. This issue is addressed by applying data mining technology using quantitative analytical techniques to help to discover previously undetected patterns present in the historic data to determine the buying and selling points of equities.

Time series are temporal sequences of measures that can be mined for information [13]. The analysis of evolving time series data lead to the discovery of many interesting financial market predictive patterns [8]. Evolving time series data analysis maximizes the profit-loss value prediction of financial stock using time series quantitative analysis. The quantitative analytical techniques are used to discover previously undetected patterns present in the historic data to determine the buying and selling points of equities [1].

The approaches to predict stock market could be classified into two classes: Fundamental Analysis and Technical Analysis [16]. The Fundamental Analysis is based on macroeconomic data and the basic financial status of companies like money supply, interest rate, inflationary rates, dividend yields, earnings yield, cash flow yield, book to market ratio, price-earnings ratio and lagged returns. The Technical Analysis is based on the rationale that history will repeat itself and the correlation between price and volume reveals market behavior [16], [19]. Prediction is made by exploiting implications hidden in past trading activities and by analyzing patterns, which are essentially based on Technical Analysis, as most of the parameters of Fundamental Analysis are either static and well known or global in nature implying that they contribute only a little in predicting the behaviour of individual stocks.

Technical predictive approach includes time series quantitative analysis, stochastic analysis and heuristics. Time series quantitative analysis is used in this research work to predict the financial stock market. Number of technical indicators and time-series analytics used by financial experts are well documented [22], [12]. However, the criterion that is used to combine these analytics to arrive at predictive strategies is not understood well. Hence, a systematic study is performed to individually evaluate best known time series analytics for their predictability for profitable stock trades. Also, when multiple non-correlated time series analytics are combined, there is a possibility to improve the predictive power [18]. The combined performances of selective financial indicators are studied and the combinational patterns that maximize profitable trades as well as result in high return on investment are explored in [8].

2. Related Works

Garth Garner [7] combined five different methods of time series technical analysis [5] to predict the following day's closing price. This technique has resulted in the generation of 50% of correct prediction. Antonia Azzini [1] demonstrated that it is possible to extract meaningful and reliable models of financial time series from a collection of popular technical indicators [14] by means of combinational algorithms [21], [3],[2]. The combined selection of various technical indicators to decide the stock market trend is discussed in [9], [4]. Robert W. Stone [18] demonstrated the technique of producing the greatest single profit using the moving average algorithm combined with the RSI applied on 60 minute trading charts, closely followed by the moving average trading of 1 minute charts.

Stephen C.W. Chu [26] combined broad range of technical indicators, including Moving Averages, RSI, Stochastic and Bollinger Band. Many combinations of the financial indicators are studied and two of them are presented in this paper. The overall predictive capability is enhanced in the proposed combinational Time Series Quantitative Algorithms.

Hence, there is a necessity to study the potential of combinational uncorrelated existing analytics in order to improve the predictive power.

3. Financial Indicators

Various financial indicators used for the study and development of the proposed combinational algorithms are presented below.

A. Bollinger Bands Calculation

The Bollinger Bands [11], [10] are trading bands plotted around a price. This indicator is plotted as a grouping of 3 lines. The upper and lower lines are plotted according to market volatility. The middle line is the simple moving average between the two outer lines (bands). Market reversals occur near the upper and lower bands. The middle band acts as a support line.

B. Relative Strength Indicator

The RSI indicator [17] compares the number of days a stock finishes up with the number of days it finishes down. It is calculated for a certain time span usually between 9 and 15 days. The average number of up days is divided by the average number of down days. This number is added to one and the result is used to divide 100. This number is subtracted from 100. The RSI has a range between 0 and 100. A RSI of 70 or above can indicate a stock which is overbought and due for a fall in price. When the RSI falls below 30 the stock may be oversold and is a good they can vary depending on whether the market is bullish or bearish. RSI charted over longer periods tend to show less extremes of movement. Looking at historical charts over a period of a year or so can give a good indicator of how a stock price moves in relation to its RSI.

C. Chaikin Money Flow Indicator

Developed by Marc Chaikin, the Chaikin Money Flow indicator (CMF) [20] is calculated by summing Accumulation Distribution over the given period and then dividing by the sum of volume over the given period [11]. A period of 21 is recommended. The volume is essentially nothing more than volume times change divided by range. A positive CMF value signals accumulation, while a negative CMF value signals distribution. A reference line is drawn at zero to help quickly identify accumulation/distribution regions. The further the distance from the zero reference line, the stronger the signal.

D. True Strength Index

The True Strength Index (TSI) is calculated as detailed below.

A) Exponential Moving Average (MA) of Momentum for term1 periods

B) Exponential MA of (A) for term 2 periods

C) Exponential MA of Absolute Momentum for term1 periods

D) Exponential MA of (C) for term 2 periods.

E) TSI = B / D

The Momentum is an oscillator that measures the amount a stock price has changed over the observation period. It is a trend-follower.

Prices are the central quantities. Four types of prices are distinguished:

The open price O_t indicates the price of an asset at the start of trading day t.

- The high price H_t indicates the highest price at which an asset was traded during trading day t.
- The low price L_t indicates the lowest price at which an asset was traded during trading day t.
- The close price C_t indicates the price of an asset at the end of trading day t.

The Price at time t is defined by

$$P_t = \frac{w_1 O_t + w_2 H_t + w_3 L_t + w_4 C_t}{w_1 + w_2 + w_3 + w_4}$$

Default values for the non-negative weights are $w_1 = w_2 = w_3 = w_4 = 1$.

The Momentum measures the price change over a time period of length n,

Momentum(t) = $P_t - P_{t-n}$

The Momentum oscillates around the zero-line. Default value for n is 14 days.

E. Stochastic Momentum Index

The Stochastic Momentum Index (SMI) [17] is based on the Stochastic Oscillator. The difference is that the Stochastic Oscillator calculates where the close is relative to the high/low range, while the SMI calculates where the close is relative to the midpoint of the high/low range. The values of the SMI range from +100 to -100. When the close is greater than the midpoint, the SMI is above zero, when the close is less than the midpoint, the SMI is below zero. The SMI is interpreted the same way as the Stochastic Oscillator. Extreme high/low SMI values indicate overbought / oversold conditions. A buy signal is generated when the SMI rises above -50, or when it crosses above the signal line. A sell signal is generated when the SMI falls below +50, or when it crosses below the signal line.

F. Relative Momentum Index

The Relative Momentum Index (RMI) is a variation on the Relative Strength Index (RSI). To determine up and down days, the RSI uses the close compared to the previous close. The RMI uses the close compared to the close n days ago. An RMI with a time period of 1 is equal to the RSI. The RMI ranges from 0 to 100. Like the RSI, The RMI is interpreted as an overbought/ oversold indicator when the value is over 70/below 30. You can also look for divergence with price. If the price is making new highs/lows, and the RMI is not, it indicates a reversal.

G. Typical Price

The Typical Price indicator is calculated by adding the high, low, and closing prices together, and then dividing by three. The result is the average, or typical price.

H. Bollinger Signal

The Typical price (TP) and the Bollinger Bands for the days represented in terms of "i" are combined as follows to obtain the Bollinger Signal.

```
IF (TP[i] > Bollinger Lower Band[i]) &&
(TP[i-1] < Bollinger Lower Band[i-1])
 Generate a buy signal
ELSE
IF (TP[i] < Bolinger Upper Band[i]) &&
   (TP[i-1] > Bollinger Upper Band[i-1])
Generate a sell signal
         ENDIF
ENDIF
```

I. RSI Signal

The following algorithm, based on Relative Strength Indicator generates the RSI Signal, which is used to predict an increase or decrease in next days closing stock price.

```
If (RSI <=30)
Predict increase in tomorrow's closing price
Else
If (RSI >=70)
Predict decrease in tomorrow's closing price
Else
No change in tomorrow's closing price
End if
```

J. RMI Signal

The following algorithm, based on Relative Momentum Index generates the RMI Signal, which is used to predict an increase or decrease in next days closing stock price.

```
If (RMI <=20)
```

Predict increase in tomorrow's closing price
Else
If (RMI >=80)
Predict decrease in tomorrow's closing price
Else
No change in tomorrow's closing price
End if

K. SMI Signal

The following algorithm, based on Stochastic Momentum Index generates the SMI Signal, which is used to predict an increase or decrease in next days closing stock price.

If (SMI > 20)
Predict decrease in tomorrow's closing price
Else
If (SMI < -40)
Predict increase in tomorrow's closing price
Else
No change in tomorrow's closing price
End if

L. TSI Signal

The algorithm for the TSI signal based on the True Strength Index is given below.

```
IF (tsioutput[i] >-30 && tisoutput[i-1] <= -30
&& tsioutput[i-2] <= -30)
Generate a buy signal
ELSE
IF (tsioutput[i] < 30 && tsioutput[i-1] >= 30
&& tsioutput[i-2] >= 30)
Generate a Sell signal
        ENDIF
ENDIF
```

M. Moving Average Crossover Algorithm

The Moving Average shows the average price over a period of time [2]. For a 30 day moving average you add the closing prices for each of the 30 days and divide by 30. The most common averages are 20, 30, 50, 100, and 200 days. Longer time spans are less affected by daily price fluctuations. A moving average is plotted as a line on a graph of price changes. When prices fall below the moving average they have a tendency to keep on falling. Conversely, when prices rise above the moving average they tend to keep on rising.

4. Proposed Algorithm

Various methods of analyzing stocks were combined to generate a buy / sell signal resulting in profit, are mentioned below. All the methods are needed to be in agreement for the algorithm to predict profitable buy / sell.

a. Time Series Quantitative Algorithm 3 – TSQA3

The Chaikin Money Flow indicator is calculated from the daily readings of the Accumulation / Distribution Line [20]. The basic premise behind the Accumulation Distribution Line is that the degree of buying or selling pressure can be determined by the location of the Close relative to the High and Low for the corresponding period (Closing Location Value) [11]. The proposed TSQA3 analyzes the stock price using the optimized trading rule, which is evolved by combining the, Chaikin Money Flow Indicator as a confirming indicator, with the Relative Strength indicator, Stochastic Momentum Index, Bollinger Bands Calculation and RSI Signal. Experimental results show that this combinational analytic outperforms.

b. Time Series Quantitative Algorithm 4 – TSQA4

Piotr Lipinski [15] proposed a new trading rule by combining Stochastic Oscillator with Relative Strength Indicator. The author has not focused on profitable buy / sell signals generation, but proposed a technique of generating warning signals for specific situations relate to the rare circumstances of exceptional raises or drops of share prices. The Stochastic oscillator, a popular and dynamic indicator based on the premise that during an upward trading market, prices tend to close near their high, and during a downward trading market, prices tend to close near their low. TSQA4 is formed by combining a variation of this oscillator called Stochastic Momentum index with the confirming indicator Chaikin Money flow Indicator, Bollinger Bands Calculation, Relative Strength indicator and Stochastic Momentum Index Signal for achieving better results.

5. Back Testing

Most of the back-testing systems in commercial stock trading products do not have a converging signal generating system. Having well defined signal system is very important from many points of view such as presentation in charts as buy or signals and for a trading program in the trading system to make the right decision when such signals are generated. When signals are standardized, the meaning of the signals is well understood according to the application needs. From the very beginning of the design, we have made a decision to have converging signals with different grades of buy and sell signals will be central to make trading decisions. Thus, we have said, we have four buy and four sell signals and one neutral signal in our system. The idea behind the signal generation is described below

Stock trading decisions depend upon the validation of the trader's hypothesis based on objective or subjective functions. Standardizing the result of well conceived objective or subjective functions into well defined signals is the most important step in the automation of decision making processes and presentation in charts in user intelligible ways. Thus, any number of objective or subjective functions could be written for well conceived hypothesis and all of which can provide converged signals to a common automated trading system. When ideas or hypothesis driven by varying investment goals are innumerable, each one could be very carefully crafted into analytical processes in Mathematical programming language with an objective of generating the final decision points in the form of signals.

a. Data Model for Back Testing

When a number of equities are traded over a period of time, a trade consolidation can be done with the following data.

There are two transitions, one corresponds to the buy transaction and the other corresponds to the sell transaction. Each buy or sell transaction has its own set of values. The consolidation of the row will have total amount committed for buying and the total amount on sell closing. The combined consolidation will have Net = Total amount - Buy total amount. If the Net is negative, it is a loss and if the Net is positive it is a profit. The total of all net across all rows is the Total Net Profit / Loss.

Table 1 : Profitable Signals Generated By Moving Average Crossover Algorithm

Stock Name or Symbol	Total predictions made	No. of Profitable signals	Percentage of profitable signals
MSFT	96	46	47.9%
GOOG	83	39	46.9%
ORCL	70	38	54.2%
IBM	92	58	63.0%
YHOO	68	35	51.4%

Table 2: Profitable Signals Generated By TSQA3

Stock Name or Symbol	Total predictions made	No. of Profitable signals	Percentage of profitable signals
MSFT	48	33	68.7%
GOOG	42	30	71.4%
ORCL	57	38	66.67%
IBM	54	38	70.37%
YHOO	56	42	75.0%

6. Empirical Results

The bar data is available online in a format with six attributes. They are open, low, high, close prices, volume, and starting time of the business. Open is the opening price of the stock, Low is the lowest price of the stock during the business period, high is the highest price of the stock and close is the closing price of the stock for the business period. Volume represents the total number of stocks traded during the period. A bar data represents an abstract sample of trades that have taken place during the given period. Bar data can be formulated using different time intervals in minutes, hours, days, weeks or months. The day wise formulated bar data is used in this study.

Both the combinational analytics were tested on the historic bar data of 1009 symbols / stocks pertaining to the period from 01.01.2002 to 31.12.2006. This data is obtained from yahoo finance web site. The test results of the Moving Average Crossover algorithm, TSQA3 and TSQA4 are listed in the table 1 to table 3 respectively. The average percentage of

Table 4 : Profitable Signals

S.No	Algorithm	Average Percentage of profitable Buy / Sell signal generated
1	MAC	52.62%
2	TSQA3	70.19%
3	TSQA4	56.34%

profitable Buy / Sell signals generated by the proposed algorithms are listed in the Table 4.

The study of the proposed algorithms reveals the following.

1. The Moving Average Crossover (MAC) is a simple straight forward smoothing algorithm, which generates an average of 52.62% of profitable signals.
2. The percentage of profitable signals generated by the proposed TSQ3 is better for majority of the symbols / stocks.
3. TSQA3 outperforms over all the other algorithms considered for the study.

7. Conclusion and Future Work

The TSQA3 algorithm did very well on the stocks under study. In this case the predictions were correct or profitable for at least 70.19%. This is not to assure that this algorithm would make anyone rich, but it may be useful for trading analysis. This raises the following question.

How much could you lose before you actually won? The answer is that it is possible to win 70.19% of the time, but still lose a lot consecutively before we may actually win.

The main contribution of this work is demonstrating that it is possible to extract meaningful and reliable models of financial time series from a collection of popular technical indicators by means of combinational algorithms. The predictive power can be enhanced by combinational algorithms coupled with Artificial Neural Networks (ANN). A neuro-genetic approach may be introduced for the modeling of financial time series, which opens up opportunities for many extensions, improvements, and sophistications both on the side of indicators, and of the technicalities of the trading strategy. A natural extension would consist of enriching the data used for modeling with more sophisticated technical indicators. Novel indicators might be evolved or coevolved by means of other evolutionary techniques like genetic programming.

References

[1] Antonia Azzini and Andrea G. B. Tettamanzi, *"Evolving Neural Networks for Static Single-Position Automated Trading"*, Journal of Artificial Evolution and Applications, Article ID 184286, Vol. 2008.

[2] Azzini.A and A. G. B. Tettamanzi, *"A neural evolutionary approach to financial modeling"*, in Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation (GECCO '06), Vol. 2, PP. 1605–1612, Seattle, Wash, USA, July 2006.

[3] Azzini.A and A. G. B. Tettamanzi, *"Neuro-genetic single position day trading"*, in Proceedings of the Workshop Italiano di Vita Artificiale e Computazione Evolutiva (WIVACE '07), Sicily, Italy, September 2007.

[4] Brabazon.A and M. O'Neill, *"Biologically Inspired Algorithms for Financial Modelling"*, Springer, Berlin, Germany, 2006.

[5] Bruce M. Kamich, *"How Technical Analysis Works"*, Institute of Finance, New York, 2003.

[6] Caroline C. Hayes, *"Agents in a Nutshell - A Very Brief Introduction"*, IEEE Transactions on Knowledge and Data Engineering, Vol. 11, No. 1, Feb. 1999.

[7] Garth Garner, *"Prediction of Closing Stock Prices"*, Engineering Data Analysis and Modeling at Portland State University, fall term 2004.

[8] Graham Kendall and Yan Su, *"A Multi-agent Based Simulated Stock Market - Testing on Different Types of Stocks"*, Congress on Evolutionary Computation (CEC '03), Vol. 4, PP. 2298- 2305, IEEE 2003.

[9] Harris.L, *"Trading and Exchanges, Market Microstructure for Practitioners"*, Oxford University Press, New York, NY, USA, 2003.

[10] Jia-Hao Li, Jack J. W. Yang, *"A Trading Decision Support System Based on Neuro-Fuzzy Technique: Evidence from Asian Stock Market"*, Journal of International Management Studies, February 2008.

[11] John Bollinger, *"Using Bollinger Bands"*, Stocks & Commodities V. 10:2 (47-51), Technical Analysis Inc., 2005.

[12] Kaufman. P.J, *"Trading Systems and Methods"*, New York: Wiley, 1998.

[13] Martin.R and Yohai.V, *"Data Mining for unusual movements in temporal data"*, In proceedings of the KDD01 workshop on temporal data mining, San Fransisco, CA, 2001.

[14] Myungsook Klassen, *"Investigation of Some Technical Indexes in Stock Forecasting Using Neural Networks"*, The Third World Enformatika Conference WEC (5), 2005 .

[15] Piotr Lipinski and Jerzy Korczak, *"Early Warning in On-line Stock Trading Systems"*, 5th International Conference on Intelligent Systems Design and Applications (ISDA'05), IEEE 2005.

[16] Qiang Ye, Bing Liang and Yijun Li, *"Amnestic Neural Network for Classification: Application on Stock Trend Prediction"*, International Conference on Services Systems and Services Management, IEEE 2005, Vol. 2, PP. 1031- 1034.

[17] Robert W. Colby and Thomas A. Meyers, *"The Encyclopedia of technical Market Indicators"*, Dow Jones-Irwin, Homewood, IL, 2004.

[18] Robert W. Stone, *"Trading With A Day Job: Can Automated Trading Strategies Be Profitable?"*, October 2007.

[19] Rui Jiang, and K. Y. Szeto, *"Extraction of Investment Strategies based on Moving Averages: A Genetic Algorithm Approach"*, CIFEP' 03, Hong Kong, IEEE 2003.

[20] Stephen C.W. Chu, H.S. Ng and K.P. Lam, *"Intelligent Trading Advisor"*, ICMIT 2000, PP. 53-58.

[21] Subramanian.H, S. Ramamoorthy, P. Stone and B. J. Kuipers, *"Designing safe, profitable automated stock trading agents using evolutionary algorithms"*, in Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation (GECCO '06), Vol. 2, PP. 1777–1784, Seattle, Wash, USA, July 2006.

[22] Yung-Keun Kwon and Byung-Ro Moon, *"A Hybrid Neurogenetic Approach for Stock Forecasting"*, IEEE Transactions on Neural Networks, Vol. 18, No. 3, May 2007.

Author's Biography

Karthikeyan. M received B.E degree in Electronics and Communication Engineering from Bharathiar University, Coimbatore, India in 1990, M.Tech., degree in Computer and Information Technology from Centre for Information Technology and Engineering, Manonmaniam Sundaranar University, Tirunelveli, India. Currently, he is doing Ph. D., in Computer Science and Engineering at Faculty of Engineering, Manonmaniam Sundaranar University Tirunelveli. His research interests include Data mining and Image Processing. He is a Member of the IEEE.

F. Dheepan Cross received B.E degree in Computer Science and Engineering from Manonmaniam Sundaranar University, Tirunelveli India in 2002, M.Tech., degree in Computer and Information Technology from Centre for Information Technology and Engineering, Manonmaniam Sundaranar University, Tirunelveli, India. Currently, he is doing Ph. D., in Computer and Information Technology at Faculty of Engineering, Manonmaniam Sundaranar University Tirunelveli. His research interests include Data mining to predict Financial Market Using Quantitative Analysis.

Nallaperumal Krishnan received M.Sc., degree in Mathematics from Madurai Kamaraj University, Madurai, India in 1985, M.Tech degree in Computer and Information Sciences from Cochin University of Science and Technology, Kochi, India in 1988 and Ph.D., degree in Computer Science & Engineering from Manonmaniam Sundaranar University, Tirunelveli. Currently, he is heading Centre for Information Technology and Engineering of Manonmaniam Sundaranar University, Tirunelveli. His research interests include Signal and Image Processing, Remote Sensing, Visual Perception, Mathematical Morphology Fuzzy Logic, Data mining and Pattern recognition. He has authored three books, edited 18 volumes and published 25 scientific papers in Journals. He is a Senior Member of the IEEE.

K. Senthamarai Kannan is currently working as a Professor and Head in Statistics, at the Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu. He has more than 18 years of teaching experience at post-graduate level. He has published more than 30 research papers in international and national journals and authored four books. He has visited Turkey, Singapore and Malaysia. He has been awarded TNSCST Young Scientist Fellowship and SERC Visiting Fellowship. His area of specialization is 'Stochastic Processes and their Applications'. His other research interests include stochastic modeling in the analysis of birth intervals in human fertility, bio-informatics, data mining and precipitation analysis.