Financial Stock Market Forecast using Data Mining Techniques

Article · March 2010
Source: DOAJ

4 authors, including:

Senthamarai Kannan Kaliyaperumal
Manonmaniam Sundaranar University
114 PUBLICATIONS   **697** CITATIONS

Mohamed Sathik
Sadakathullah Appa College
135 PUBLICATIONS   **963** CITATIONS

Paliah Arumugam
Manonmaniam Sundaranar University
19 PUBLICATIONS   **152** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   Comparative analysis of community discovery methods in social networks   View project

Project   neural networks   View project

Financial Stock Market Forecast using Data Mining Techniques

K. Senthamarai Kannan, P. Sailapathi Sekar, M.Mohamed Sathik and P. Arumugam

Abstract— The automated computer programs using data mining and predictive technologies do a fare amount of trades in the markets. Data mining is well founded on the theory that the historic data holds the essential memory for predicting the future direction. This technology is designed to help investors discover hidden patterns from the historic data that have probable predictive capability in their investment decisions. The prediction of stock markets is regarded as a challenging task of financial time series prediction. Data analysis is one way of predicting if future stocks prices will increase or decrease. Five methods of analyzing stocks were combined to predict if the day's closing price would increase or decrease. These methods were Typical Price (TP), Bollinger Bands, Relative Strength Index (RSI), CMI and Moving Average (MA). This paper discussed various techniques which are able to predict with future closing stock price will increase or decrease better than level of significance. Also, it investigated various global events and their issues predicting on stock markets. It supports numerically and graphically.

Index Terms— Data mining, Time series Analysis, Binomial test, Typical Price, Bollinger Bands, Relative Strength Index and Moving Average.

I. INTRODUCTION

Data mining can be described as "making better use of data". Every human being is increasingly faced with unmanageable amounts of data; hence, data mining or knowledge discovery apparently affects all of us. It is therefore recognized as one of the key research areas. Ideally, we would like to develop techniques for "making better use of any kind of data for any purpose". However, we argue that this goal is too demanding yet. Over the last three decades, increasingly 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

Dr.K.Senthamarai Kannan, is with Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli, India. (e-mail:senkannan2002@gmail.com,Ph: 04622330366, Fax: 04622334363)

P.Sailapathi Sekarc is with Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli, India.
(e-mail: pssekar@hotmail.com).

Dr.M.Mohamed Sathik is with Department of Computer Science, Sadakathullah Appa College, Tirunelveli, India
(e-mail: mmdsadiq@gmail.com).

Dr..P.Arumugam is with Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli, India.
(e-mail: sixfacemsu@gmail.com).

data, many fund managers have been unable to fully capitalize on their value. This paper attempts to determine if it is possible to predict if the closing price of stocks will increase or decrease on the following day. The approach taken in this paper was to combine six methods of analyzing stocks and use them to automatically generate a prediction of whether or not stock prices will go up or go down. After the predictions were made they were tested with the following day's closing price. If the following day's closing price can be predicted to increase or decrease 70% of the time at the 0.07 confidence level, then this analysis would be an easy and useful aid in financial investing. Furthermore, the results would show that the results are better than random at a reasonable level of significance. Many fund management firms have invested heavily in information technology to help them manage their financial portfolios. Over the last three decades, increasingly 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. The automated computer programs using data mining and predictive technologies do a fare amount of trades in the markets. Data mining is well founded on the theory that the historic data holds the essential memory for predicting the future direction. This technology is designed to help investors discover hidden patterns from the historic data that have probable predictive capability in their investment decisions. This is an attempt is made to maximize the prediction of financial stock markets using data mining techniques. Predictive patters from quantitative time series analysis will be invented fortunately, a field known as data mining using quantitative analytical techniques is helping to discover previously undetected patterns present in the historic data to determine the buying and selling points of equities. When market beating strategies are discovered via data mining, there are a number of potential problems in making the leap from a back-tested strategy to successfully investing in future real world conditions. The first problem is determining the probability that the relationships are not random at all market conditions. This is done using large historic market data to represent varying conditions and confirming that the time series patterns have statistically significant predictive power for high probability of profitable trades and high profitable

returns for the competitive business investment.

II. METHODOLOGY

Five methods of analyzing stocks were combined to predict if the following day's closing price would increase or decrease. All five methods needed to be in agreement for the algorithm to predict a stock price increase or decrease. The five methods were Typical Price (TP), Chaikin Money Flow indicator (CMI), Stochastic Momentum Index (SMI), Relative Strength Index (RSI), Bollienger Bands (BB), Moving Average(MA) and Bollienger Signal.

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.

    Algorithm:
1. Inputting High, Low, Close values of the daily share
2. Take an output array and add the values of H,L,C
3. Devide the total by 3

$$TP = \left[\frac{H + L + C}{3} \right]$$ where H=High; L=Low; C=Close. Where the TP greater than the bench mark we have to sell or to buy.

Chaikin Money Flow Indicator

Chaikin's money flow is based on Chaikin's accumulation/distribution. Accumulation/distribution in turn, is based on the premise that if the stock closes above its midpoint [(high+low)/2] for the day, then there was accumulation that day, and if it closes below its midpoint, then there was distribution that day. Chaikin's money flow is calculated by summing the values of accumulation/distribution for 13 periods and then dividing by the 13-period sum of the volume.

  It is based upon the assumption that a bullish stock will have a relatively high close price within its daily range and have increasing volume. However, if a stock consistently closed with a relatively low close price within its daily range with high volume, this would be indicative of a weak security. There is pressure to buy when a stock closes in the upper half of a period's range and there is selling pressure when a stock closes in the lower half of the period's trading range. Of course, the exact number of periods for the indicator should be varied according to the sensitivity sought and the time horizon of individual investor. An obvious bearish signal is when Chaikin Money Flow is less than zero. A reading of less than zero indicates that a security is under selling pressure or experiencing distribution.

An obvious bearish signal is when Chaikin Money Flow is less than zero. A reading of less than zero indicates that a security is under selling pressure or experiencing distribution. A second potentially bearish signal is the length of time that Chaikin Money Flow has remained less than zero. The longer it remains negative, the greater the evidence of sustained selling pressure or distribution. Extended periods below zero can indicate bearish sentiment towards

the underlying security and downward pressure on the price is likely. The third potentially bearish signal is the degree of selling pressure. This can be determined by the oscillator's absolute level. Readings on either side of the zero line or plus or minus 0.10 are usually not considered strong enough to warrant either a bullish or bearish signal. Once the indicator moves below -0.10, the degree selling pressure begins to warrant a bearish signal. Likewise, a move above +0.10 would be significant enough to warrant a bullish signal. Marc Chaikin considers a reading below -0.25 to be indicative of strong selling pressure. Conversely, a reading above +0.25 is considered to be indicative of strong buying pressure. The Chaikin Money Flow is based upon the assumption that a bullish stock will have a relatively high close price within its daily range and have increasing volume. This condition would be indicative of a strong security. However, if it consistently closed with a relatively low close price within its daily range and high volume, this would be indicative of a weak security.

The Following formula was used to calculate CMI.

$$CMI = \left[\frac{sum(AD,n)}{sum(VOL,n)} \right]; \quad AD = VOL \left[\frac{(CL - OP)}{(HI - LO)} \right]$$

    AD stands for Accumulation Distribution, Where n=Period; CL=today's close price; OP=today's open price; HI=High Value; LO=Low value

Stochastic Momentum Index

The Stochastic Momentum Index (SMI) 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. Also look for divergence with the price to signal the end of a trend or indicate a false trend.

  The Following formula was used to calculate SMI.

$$100 \times \left[\frac{[MOV[MOV[C - [5 \times [HHV(H,13) + LLV(L,13)]]],25,E],2,E]}{[5 \times [MOV[MOV[[HHV(H,13) + LLV(L,13)]]],25,E],2,E]} \right]$$

where HHV= Highest high value.
    LLV = Lowest low value.
      E = exponential moving avg.
Using the following formula, exponential moving average was calculated.

$$EMA = \left[\left[(Price(i) - prevMVG) \times \left(\frac{2}{N+1} \right) \right] \right] + prevMVG$$

Relative Strength Index

This indicator 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.

RSI=100–(100/1+RS);   RS=AG/AL
AG=[(PAG)x13+CG]/14; AL=[(PAL)x13+CL]/14
PAG = Total of Gains during past 14 periods/14
PAL = Total of Losses during past 14 periods/14

Where   AG=Average   Gain,   AL=Average   Loss
PAG=Previous   Average   Gain,   CG=Current   Gain
PAL=Previous Average Loss, CL=Current Loss

The following algorithm was used to calculate RSI:
Upclose = 0
DownClose = 0
      Repeat for nine consecutive days ending today
  If (TC > YC)
        UpClose = (Upclose + TC)
        Else if (TC < YC)
        DownClose = (Down Close + TC)
      End if

$$RSI = 100 - \left[\frac{100}{\left(1 + \left[\frac{UpClose}{DownClose} \right] \right)} \right]$$

The following figure explains the RSI graph of a sample data set
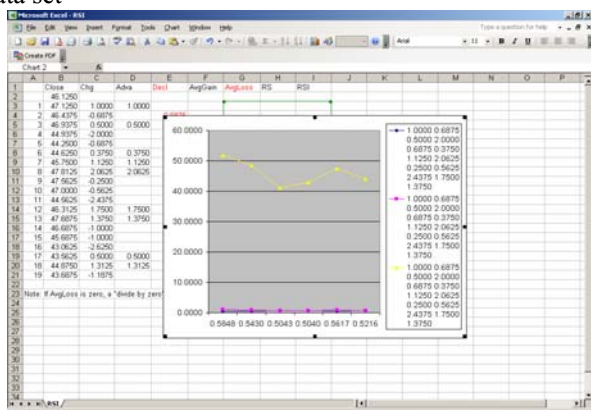

Figure 1: RSI graph

Bollinger Bands

Bollinger Bands are based upon a simple moving average. This is because a simple moving average is used in the standard deviation calculation. The upper band is two standard deviations above a moving average; the lower band is two standard deviations below that moving average; and the middle band is the moving average itself. This indicator is plotted as a grouping of 3 lines. The upper and lower lines are

plotted according to market volatility. When the market is volatile the space between these lines widens and during times of less volatility the lines come closer together. The middle line is the simple moving average between the two outer lines (bands). As prices move closer to the lower band the stronger the indication is that the stock is oversold the price should soon rise. As prices rise to the higher band the stock becomes more overbought meaning prices should fall. Bollinger bands are often used by investors to confirm other indicators. The wise technical analyst will always use a number of indicators before making a decision to trade a particular stock. Bollinger Bands (BB) are not a standalone indicators as they do not generate explicit buy or sell signals and are generally used to provide a form of guideline, indicating possible trend reversals. In this case, if the current price breaks through the lower bollinger band it is considered a buy signal, while if it breaks through the upper band it is considered a sell signal. The Upper and Lower Bands are calculated as

$$\text{stdDev} = \sum_{i=1}^{N}(price\,(i) - MA(N))^2$$

$$\text{Upperband} = MA + D\sqrt{\sum_{i=1}^{N} \left[\frac{(Price(i) - MA)2}{N} \right]}$$

$$\text{Lowerband} = MA - D\sqrt{\sum_{i=1}^{N} \left[\frac{(Price(i) - MA)2}{N} \right]}$$

where D= No. of standard deviations applied.

Moving Average

The most popular indicator is the moving average. This shows the average price over a period of time. 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.

Bollinger Signal

This indicator is plotted as a grouping of 3 lines. The upper and lower lines are plotted according to market volatility. When the market is volatile the space between these lines widens and during times of less volatility the lines come closer together. The middle line is the simple moving average between the two outer lines (bands). As prices move closer to the lower band the stronger the indication is that the stock is oversold the price should soon rise. As prices rise to the higher band the stock becomes more overbought meaning prices should fall. Bollinger bands are often used by investors to confirm other indicators. The wise technical analyst will always use a number of indicators before making a decision to trade a particular stock. Bollinger Bands (BB) are not a standalone indicators as they do not generate explicit buy or sell signals and are generally used to provide a form of guideline, indicating possible trend reversals. In this case, if the current price breaks through the lower bollinger band it is considered a buy signal, while if it breaks through the upper band it is considered a sell signal. The Upper and Lower Bands are calculated as

$$stdDev = \sum_{i=1}^{N}(price\ (i) - MA(N))^2$$

$$Upperband = MA + D\sqrt{\sum_{i=1}^{N}\left[\frac{(Price(i) - MA)2}{N}\right]}$$

$$Lowerband = MA - D\sqrt{\sum_{i=1}^{N}\left[\frac{(Price(i) - MA)2}{N}\right]}$$

where D = No. of standard deviations applied.

The following figure explains the Bollinger band crossover of the sample data

Figure 2: Bollinger band crossover BSRCTB
(Combinatorial Algorithm)

In this algorithm we are using the concepts of different techniques like SMI, RSI, CMI and Bollinger band. By using the advantages of all the above techniques we can make the net profit as high. The buy signal and sell signals can be produced by using the function Bollinger signals. By comparing with moving average crossover we can find out how much effective is the new technique. We are keeping the moving average crossover as the benchmark. This algorithm will overcome almost all the limitations by the above Papers.

Result Analysis

From this analysis I have got a profitable signal for Moving average as 52.62% and comparing to it, my algorithm BSRCTB have a profitable signal of 58.25%. The BSRCTB algorithm performs better than all the other algorithm. So refer that the result of our algorithm will provide a good response in the stock market prediction. The table below shows the result comparison of all the methods used in SBRC algorithm with the Moving Average and it shows the profitability and the success of each method. The result we have got from using all the methods are quoted in the above table. From this we can find out that all the methods are being used with a sample of 400 signals. When we take Moving Average (MA), it is found that the profitable signal produced is 60%. By using Bollinger Bands we are getting a profit of 84.24%. Using Chaikin Money Flow Indicator (CMI) we are getting a profit of 51.45%. The profit percentage of Relative Strength Index (RSI) method is 56.04%. The Stochastic Momentum Index produces a result of profitable signals as 100%. So we can find out that the SMI and Bollinger Bands could produce more profitable signals. The profit loss analysis table is shown bellow

Table 1: Profit Loss Analysis

Methods	Symbols Processed	% of profitable Signal	% of Annual Return
MAV	400	52.62	24.190
BollingerBands	400	84.24	
CMI	400	51.45	21.800
RSI	400	56.04	25.130
RSI Signal	400	0.00	
SMI Signal	400	100.00	
BSRCTB	400	58.25	32.178

III. CONCLUSION

The results show that this algorithm was able to predict if the following day's closing price would increase or decrease better than chance (50%) with a high level of significance. Furthermore, this shows that there is some validity to technical analysis of stocks. This is not to say that this algorithm would make anyone rich, but it may be useful for trading analysis. The algorithm performed well on half of the stocks and not so well on the other half of the stocks.
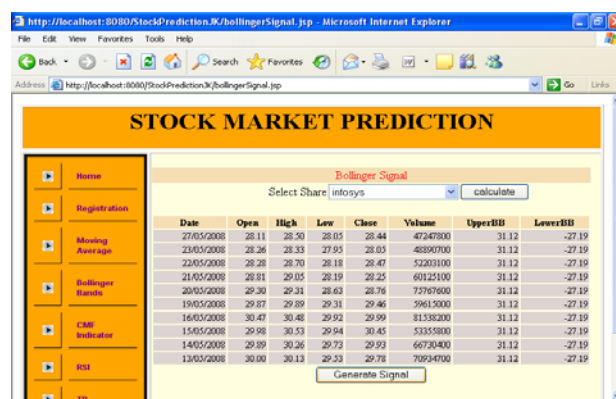

Figure 3: Implementation of BSRCTB (Combinatorial Technique)

In either case the prediction was correct at least 50% of the time. You could win 50% of the time, but still lose a lot consecutively before you actually won. The algorithm generated both increase and decrease predictions, but the predictions did not come very often. Therefore, if you trusted the indication of an increase as a buy signal you would not be able to use the algorithm as an indicator of when to sell because the algorithm is usually silent. This algorithm could perhaps be used as a buying or selling signal or it could be used to give confidence to a trader's prediction of stock prices.

REFERENCES

[1] Allen M. Poteshman, (2001). "Unusual Option Market Activity and the Terrorist Attacks of September".

[2] Amaury Lendasse, Francesco Corona, Antti Sorjamaa, Elia Liiti¨ainen, Tuomas K¨arn¨a, Yu Qi, Emil Eirola, Yoan Mich´e, Yongnang Ji, Olli Simula, (2006). "Time series prediction".

[3] Cyril Schoreels, Jonathan M. Garibaldi, (2000). "A preliminary investigation into multi-agent trading simulations using a genetic algorithm".

[4] Durbin Hunte, (2007). "Hedge Funds Do About 60% Of Bond Trading, Study Says", The Wall Street Journal.

[5] Eamonn Keogh, Stefano Lonardi, (2004). Chotirat Ann Ratanamahatana Towards Parameter-Free Data Mining, KDD '04.

[6] Garth Garner, Prediction of Closing Stock Prices, (2004). This work was completed as part of a course project for Engineering Data Analysis and Modeling at Portland State University.

[7] Hol, E.M.J.H. (2003): ING Group Credit Risk Management, Amsterdam, "The Netherlands Empirical studies on Volatility in International Stock Markets".

[8] Hossein Nikooa, Mahdi Azarpeikanb, Mohammad Reza Yousefib,c, Reza Ebrahimpourb,c, and Abolfazl Shahrabadia,d (2007). "Using A Trainable Neural Ntwork Ensemble for Trend Prediction of Tehran Stock Exchange" IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.12,

[9] http://www.wallstreettape.com\ bollinger signal \ Bollinger band crossover - Stock market timing signal.htm

[10] Jeremy Berkowitz (2004) "On Justifications for the ad hoc Black-Scholes method of Option Pricing"

[11] Jiarui Ni and Chengqi Zhang (2006). A Human-Friendly MAS for Mining Stock Data IEEE / WIC / ACM International Conference on Web Intelligence and Intelligent Agent Technology.

[12] Juliana Yim , Heather Mitchell (2005). A comparison of corporate distress prediction models in Brazil, nova Economia_Belo Horizonte, 15 (1), 73-93.

[13] Justin Wolfers and Eric Zitzewitz (2006). "Prediction Markets in Theory and Practice"NBER Working Paper No.12083.