Data Mining Driven Computational Analysis of Stock Markets, Methods and Strategies

Anthony Lai\textsuperscript{a}, Shaoning Phang\textsuperscript{b}, Wayne Holmes\textsuperscript{a}

\textsuperscript{a}, Electrotechnology Department, Carrington Rd, Mt Albert, Auckland and 92025, New Zealand
\textsuperscript{b}, Computing Department, Carrington Rd, Mt Albert, Auckland and 92025, New Zealand

Abstract

The stock market is a complex, dynamic and non-linear environment. The prediction of any future market reaction is further complicated by huge amounts of often unstructured financial data and uncertainty due to the effects of unforeseen market events. The application of correlation analysis to significant market events is still seen as a useful tool in the prediction of future trends on the stock market in a global sense. This paper proposes the application of data mining computation correlation analysis to the stock market to enhance the durability of predictions. This method of correlation analysis is a combination of cross correlation, auto correlation and fundamental analysis that is further enhanced by Channel correlation, Weighted Pearson’s correlation and added correlation Support Vector Regression. Channel correlation traces the similarity of trends while the weighted Pearson’s correlation acts as a noise filter during the correlation extraction process.

Keywords : Computational correlation analysis; Cross correlation; Auto correlation
1. Introduction

The stock market is very uncertain and unpredictable. Some stocks will experience dramatic price increases and declines, some will barely fluctuate and others will remain largely unchanged. The following four guidelines are widely promoted by many investors: (1) do not follow the herd when investing in the stocks (2) do not invest in a stock until you understand its composition (3) do not believe in the rumors but confirm the facts first (4) determine your personal risk appetite by considering an acceptable risk-reward ratio before investing in stock. Although this is valid advice, it may be too limiting given the vast number of stocks available on the world stock market. It is just not possible for every investor to consider tens of thousands of companies mainly through analyzing their annual financial statements. Even after a substantial period of researching historic data, the investor will still be no closer to predicting the stock market’s future behavior.

Most correlation methods suffer from the inability to compensate for the effects of global economic events such as natural and manmade disasters. These may include tragic events such as the Tsunami in Japan, the earthquakes in Christchurch and the civil war in Libya.

2. Objectives

The aim of this study will propose the “data mining through computational correlation” method and analyze the results to uncover possible significant correlations in the stock market. The additional correlation analysis proposed in this paper will focus mainly on stock profitability ranking, seasonal factors related to the stock market and special events that are expected to have a meaningful influence on stock prices.

3. The importance of financial correlation analysis

The stock market is an important part of the economy of a country. Most of the time the stock market plays a vital role in the growth of industry and commerce of a country that eventually affects even unrelated aspect in the country as a whole. That is because the government and the central banks of the country keep a close watch on the events on the stock market. The stock market is important from both the industry point of view as well as the traders point of view. From the related literature, there seems to be evidence of economic forces in stock market behavior [1, 2]. Walter and Lopez implied that correlation is worth considering when it comes to foreign exchange [3]. Our works in an earlier publication shown the crude oil price has been widely sought has impact on stock market variation [11]. The following section illustrates the types of correlation and correlation classification of a country.
3.1. Types of correlation of a country

Figure 1 shows all the different types of correlation which will have an effect on the stock market. The correlation between government and bank is pervasive around the world and has significant consequences for economic and financial development especially in the stock market. A government can participate in the financing of firms in a variety of ways for example providing subsidies directly or encouraging private banks through regulation. The relationship between banks and the government is comparable to the relationship between the judiciary and the government.

Fig. 1. Different types of correlation exist between the government and banks

3.2. Correlation based on single stock product discovery

For stock product within a country, a few correlations were indicated in the picture as in general correlation of stock product. From the view point of the individual trader and small business trader, figure 2 shows the proposed general correlation of single stock product discovery.

Fig. 2. General correlation of stock product

First, one or few sample stocks will be chosen to find the most valuable references. Those samples either have strong financial statements or interest for particular investors. The reason might come from the historical data or information was made to known by investors.
3.3. Correlation based on fundamental discovery

It is practically very difficult for traditional technical and fundamental analysis approaches to discover the correlation. A graphical computational correlation analysis method is proposed for impact analysis [11]. The core of the proposed method is a synthesis of correlation extracted with the use of two correlation approximation: The channel correlation and the weighted Pearson’s correlation will be discussed in the next section. Fundamental analysis is based on macroeconomics factors. Some of the factors are effecting certain stocks in the market. A single stock could correlates with many factors, a few case studies are presented in the following to identify the target in order to perform the analysis.

3.4. Correlation based on existing technical analysis

Technical analysis employs models and trading rules based on price and volume transformations, such as the relative strength index, moving averages, regressions, inter-market and intra-market price correlations, business cycles, stock market cycles or, classically, through recognition of chart patterns. Technical analysis stands in contrast to the fundamental analysis approach to security and stock analysis. Technical analysis analyzes price, volume and other market information, whereas fundamental analysis looks at the facts of the company, market, currency or commodity. Most large brokerage, trading group, or financial institutions will typically have both a technical analysis and fundamental analysis team. Users hold that even if technical analysis cannot predict the future, it helps to identify trading opportunities.

4. Proposed computational correlation analysis methods

In recent years, correlation analysis methods have been used with extensive and productive efforts. The research has been expanding from simple to more advanced techniques both in depth and in breadth. For a correlation extraction method to be useful in predicting efficient correlation information for stock market analysis, the method is required to combine cross correlation, auto correlation and fundamental analysis. The method proposed in this study is based on this minimum requirement and further enhanced by channel correlation and weighted Pearson’s correlation. The computational correlation analysis is normally based on two variables.

3.1. Pearson’s Correlation

Pearson’s Correlation is the first formal correlation measure and it is still the most widely used measure of relationship [4]. It is a statistical measure of two variables movement relationship, which can be calculated as correlation coefficient

\[ \rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \]  \hspace{1cm} (1)

Where cov is the covariance; \( \sigma_X \) and \( \sigma_Y \) are standard deviations; \( E \) is the expected value operator. The advantage of using Pearson’s correlation is that more accurate prediction can be made when a strong correlation exists amongst variables.

3.2. Channel Correlation

The channel method for correlation extraction is used to model a concrete arc, approximating graphically the trend similarity between two time series. Base on the channel pattern, \( p \), from equation 2, the channel distance, \( d_c \) of an observation \( X \) (i.e., a stock closing price) to the reference \( Y \) within a time frame \( T \) can be defined as below.

\[ p = \arg \min_{\alpha, \epsilon \in [1,4]} \frac{\sum_{i=1}^{T} \| p_i^\epsilon - x_i \|}{T} \]  \hspace{1cm} (2)
3.3. Support Vector Regression (SVR) Time Series Prediction

Given a stock exchange time series \( x(t) \) where \( t \) represents the time point. Suppose the present time point is \( N \), then a prediction \( x \) for \( t > N \) is computed over the training data \( X(t) = \{ x(1), x(2), ..., x(N) \} \). Thus, the goal is to find a function \( f(x) \) that matches the actually obtained targets \( x(t) \) of the next time point for all the training data. According to [5], a non-linear estimation of \( f(x) \) is computed in

\[
d_c = \frac{\sum_{t=1}^{T-1}((p_{t+1}^y - p_t^y) + (p_{t+1}^y - p_t^y))}{T-1}
\]

(3)

To evaluate the validity of correlation knowledge, consider a comparison of original Support Vector Regression (SVR) time series prediction against a correlation-aided SVR (cSVR) time series prediction. The cSVR employs correlation data \( C \) in addition to the observed time series data \( X \) for regression. Let \( C \) be the correlation data to the observed time series \( X \), then the equation (1) is extended for cSVR as below [8].

\[
f(x) = (w \cdot \Phi([xC])) + b
\]

(4)

where “.” means a dot product [6] and \( \Phi(x) \) refers to the kernel function \( K(x, x') = \langle \Phi(x), \Phi(x') \rangle \), which enables performing a linear regression in higher dimensional feature space. The solution of the SVR does not rely on the input data, a kernel function that satisfies Mercer’s [7] conditions can be written as above. The table 1 below shows different types of SVR Kernel functions.

<table>
<thead>
<tr>
<th>SVM Kernel</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polynomial (homogeneous)</td>
<td>( K(x, x') = (x \cdot x')^d )</td>
</tr>
<tr>
<td>Polynomial (inhomogeneous)</td>
<td>( K(x, x') = (x \cdot x' + 1)^d )</td>
</tr>
<tr>
<td>Radial Basis Function (RBF)</td>
<td>( K(x, x') = \exp(-\gamma | x - x' |^2), for \gamma &gt; 0 )</td>
</tr>
<tr>
<td>Gaussian Radial Basis Function</td>
<td>( K(x, x') = \exp(-\frac{| x - x' |^2}{\sigma^2}) )</td>
</tr>
<tr>
<td>Sigmoid</td>
<td>( K(x, x') = \tanh(\kappa \cdot x^2 + c) ) for some(not every)( \kappa &gt; 0 ) and ( c &lt; 0 )</td>
</tr>
</tbody>
</table>

5. Special Factors Affecting Correlation

The following few correlations analyses proposed in our paper research namely:

- Stock product profitability ranking in stock market
- Seasonal factors related to stock market
- Special events expected influence on certain stock product
- Trends (interrelationship) from one stock product to others

Ranking is essential in order to rate or record performance in a system. The ranking order is based on a high change of price in a short period. Valuable stock portfolios have high quality businesses, good growth potential, robust balance sheets and a solid dividend. After many factors have put in for consideration, the system is ready to perform the ranking. The tagged target stocks will be the input. The following figure 3 shows the system which will calculate the scores for each individual stock input. The system will rank the target stocks based on the average
scores; top scorer means it has the highest potential of price increases in the defined time window. Obviously, the lowest scorer means either it might already be over priced or steady growing in its price, the systems will inform user what is the most likely circumstances for this product.

![Diagram of ranking system](image)

**Fig. 3. Propose the ranking system**

The next correlation analysis is about seasonal factors. The seasonal factors are the historical seasonal trends for individual stocks and the market indexes under different fundamental economic, monetary, business and political environments. Those seasonal tuned to different fundamental environments are the seasonal factors trends.

Start with the annual seasonal for every individual stock of each country stock exchange. Try to look for the seasonal price trends that consistently repeat themselves, in other words the most reliable trends. These repeated tendencies produce the upwards and downwards trend lines that show up in our seasonal charts. Sometimes certain stocks stop following their seasonal trends. These are due to an industry shake-up, technological breakthrough, public scandal or a variety of other reasons in the case studies, these would need to be eliminated from consideration as they are not following their seasonal trends. The stocks with tradable trends for the month and their price movements in the market are currently correlated to their historical seasonal tendency. Next confirm the seasonal price trends against the fundamental seasonal factor to see how much the trend has, as well as support from the seasonal factor trend charts of the current fundamental economic environment.

For the special events expected influence on certain stock product. One of the solid examples given here is the relationship between crude oil and its oil stock product. From the world news, US’s most recent military activities in the Middle East region has stirred up political instability in the respective regions. These military actions will continue affecting crude oil price in short and long term, as the unstable political situation will cause a certain detrimental effect on the crude oil production from Middle-Eastern countries. The WTI crude oil price fluctuated because Osama Bin Laden was killed in the Operation Neptune Spear on 02/05/2011. It appears that market have become more volatile and more intertwined once those special events happened[9, 10].

The following shows some characteristics of special events:

- Characteristics of market volatility changes and stock price reactions
- The nature and frequency of volatile changes
- Significant events associated with volatile changes
- Price level and expected return reactions to changes in volatility

The next correlation analysis is about the interrelationship from one stock product to others. The following shows the series of sectors is related to certain stocks market that will grow in either short or long term by rising living standards, changing consumer behaviour, new technologies and increased pressure on resources. These are:

- New technologies
- Rising demand for resources (energy and commodities)
- Growing middle class (rising income)
- Rising food consumption (increase population)
- Aging prolong (increase demand of health care products)
- Industrialization (growth in need for housing, power, road)
- Increasing global integration (world is getting closer)

Figure 4 below shows interrelationship between one stock to another stock or other commodities. The line at the centre indicates the mirror. Stock 1 and stock 2 could be related when something happen on their related issues.

Figure 4 shows the proposed inter-relationship design as follows:
- Stock vs trends change when the gold price is rising and falling
- Stock vs trends change when the oil price is rising and falling
- Stock vs trends change when inflation rate is rising and falling
- Stock vs trends change interest rate is rising and falling
- Stock vs trends change based on political cycles which party is in control of government

The above each of the case studies approach provides research opportunities.

5. Summary

The stock market is influenced by diverse factors from domestic economic, government, tragedy, conflict, international economic and complex environment. From the above few special correlation factors, we have analysed some specific problems on each of the case studies condition. By using the model, individual trader or small business investors should be able positively correlate among the autocorrelation, cross correlation and fundamental analysis together and compute better numerical value.

6. Conclusion

The correlation techniques are widely seen as an indicator of market variation for stock exchange analysis. It is nevertheless practically very difficult to discover comprehensive correlation knowledge from an observed market. Using traditional technical and fundamental analysis approaches, the variation of the market is influenced by diverse factors from domestic economic, historical market states, as well as international economic background. This paper proposes a computational correlation analysis for the automatic correlation extraction from available market and economic data. The proposed correlation is a synthesis of financial correlation and types of correlation in one country. The model would positively correlate among the autocorrelation, cross correlation and fundamental analysis together. The cSVR prediction is found sometime suffering unexpectedly far away from the truth value, which implies that despite the significance of the proposed correlation, how to use and fuse correlation into the present market data remains a challenge preventing us from enhancing further market understanding through computational analysis. In addition, the selection of a few correlations analyses mentioning above and the determination of time period $N$ for analysis are two computationally essential points worth addressing further for future stock market correlation analysis.
References


