

# A Computational Referencing Approach to Stocks Correlation Analysis

Ruibin Zhang

A thesis submitted to Unitec Institute of Technology  
in fulfillment of the requirements  
for the degree of Master of Computing

March, 2013

Department of Computing  
Faculty of Creative Industries & Business

Primary Supervisor: Dr. Shaoning Pang

## Abstract

The activity of the stock market is dynamic and complicated, with financial figures changing every minute. Thus, it is not ever an easy task for a professional investor to discover the beneficial knowledge that will advantage his investment intuition. Nevertheless, various studies have proved that a stock's volatility is always associated with several financial factors, even though the association is versatile and the strength of the association varies from one to another. By finding and referencing these factors, investors can either be led to profit, or misled to loss. It follows that stocks correlation presents an often positive influence to future stock movement prediction. Unlike typical technical and fundamental stocks correlation analysis, we developed in this research a hybrid method on the basis of technical analysis with attention to the theory of fundamental analysis. In other words, we deploy a mathematical model to numerically measure stocks correlation in parameterization of a selected fundamental economic factor. The approach promotes a two-tier correlation computation architecture. In the top tier, a Pearson product-moment correlation is derived to measure the direct connection of a stock to a pre-defined referencing factor. Next, the obtained results are used for referencing featured stocks correlation modeling. In contrast to traditional stocks correlation analysis, the employment of a referencing approach for computational correlation knowledge discovery will enhance the accuracy, credibility and intelligibility of the interrelationship between each pair of stocks. This is because those pre-defined referencing factors are characterized with remarkable stability and reputation in the stock market as well as global economy. We have applied the proposed referencing model to stocks correlation analysis of S&P 500 for the period of January 2000 to December 2011. In our case studies, we use the U.S. crude oil price and the S&P 500 indices price respectively as the referencing factor to compute correlation of a selected family of stocks, with them being compared to the results of same correlation analysis applied to random selected stocks within the scope of S&P500. The performance of the proposed model is demonstrated clearly through both numerical and visual observations generated from the experiment. The correlation knowledge extracted by the proposed approach will give investors a different way to interpret stock volatility in order to strengthen their investment confidence. However, the selection of referencing factors is subject to criticism on the ground of subjectivity and the arbitrary nature of the selection process - unavoidable problems in this area of research. This leave us a future re-

search direction that, by employing multiple referencing factors to ensure that the outcome developed by stocks correlation analysis is utterly impartial.

# Acknowledgments

Foremost, I would like to express the deepest appreciation to my supervisor, Dr Shaoning Pang, for his inspiration, patience and support, during my research at Unitec, not to mention his advice and unsurpassed knowledge of machine learning and pattern recognition. His perpetual energy and enthusiasm for research have motivated all his students, including me. In addition, Dr. Pang was always approachable and willing to help students with their research. Without his guidance and persistent help, this thesis would not have been completed.

I would also like to acknowledge all people of the Unitec Institute of Technology, who have helped and inspired me during my master study. I thank my fellow lab-mates in Decentralized Machine Learning Intelligence (DMLI) group: Yiming Peng, Lei Zhu and Lei Song for their friendship and assistance during my study, they have made our lab a convivial place to work.

I am very much indebted to my family, my wife Yan Liu, daughter Alva Enyu Zhang, who supported me in every possible way to see the completion of this work, and brought countless cheeriness. Last but not the least, I would like to thank my family: my parents Guoqiang Zhang and Ying Gao, who encouraged and helped me at every stage of my personal and academic life, I owe a lot to my parents, and this thesis would be simply impossible without them.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background . . . . .	1
1.2	Research Objectives . . . . .	2
1.3	Research Contributions . . . . .	3
1.4	Notation Declaration . . . . .	3
1.5	Thesis Structure . . . . .	4
<b>2</b>	<b>Literature Review of Fundamental &amp; Technical Correlation Analysis</b>	<b>5</b>
2.1	Efficient Market Hypothesis . . . . .	5
2.2	Fundamental Stocks Correlation Analysis . . . . .	6
2.2.1	The Relationship Between Changes in Commodity and Stock Price . . . . .	7
2.2.2	The Relationship Between U.S. macroeconomic news announcements and Stock Price . . . . .	10
2.2.3	The Relationship Between Changes in Money Policy and Stock Price . . . . .	11
2.3	Technical Stock Correlation Analysis . . . . .	11
2.3.1	Moving Average . . . . .	13
2.3.2	Data Mining driven Stock Correlation Analysis . . . . .	14
2.4	Summary . . . . .	15
<b>3</b>	<b>Existing Computational Correlation Analysis Methods</b>	<b>16</b>
3.1	Statistical Computational Correlation methods . . . . .	16
3.1.1	Similarity in Inner Product Space Measurement . . . . .	16
3.1.2	Pearson's product-moment Correlation Coefficients . . . . .	17

---

3.2	Graphical Computational Correlation Analysis . . . . .	20
3.3	Conditional Covariance Correlation Analysis . . . . .	23
3.4	Summary . . . . .	24
<b>4</b>	<b>The Proposed Correlation Analysis Method</b>	<b>25</b>
4.1	Introduction . . . . .	25
4.2	Motivation for the Presented Research . . . . .	25
4.3	Proposed Computational Referencing Approach . . . . .	26
4.4	Toy Sample Experiment . . . . .	28
4.5	Extension to Multiple Referencing Approach . . . . .	29
<b>5</b>	<b>Application to S&amp;P 500 Stock Market</b>	<b>33</b>
5.1	Introduction . . . . .	33
5.2	Data Pool . . . . .	33
5.3	Experiment Setup . . . . .	34
5.4	Referencing to S&P 500 Indices Price . . . . .	34
5.4.1	Datasets Selection . . . . .	34
5.4.2	Results . . . . .	35
5.4.3	Discussion . . . . .	38
5.5	Referencing to U.S. Crude Oil Price . . . . .	38
5.5.1	Datasets Selection . . . . .	38
5.5.2	Results . . . . .	40
5.5.3	Discussion . . . . .	42
5.6	Comparison between U.S. Crude Oil price and S&P 500 Indices price	42
5.6.1	Datasets Selection . . . . .	42
5.6.2	Results . . . . .	43
5.6.3	Discussion . . . . .	44
<b>6</b>	<b>Conclusion and Future Works</b>	<b>45</b>
	References . . . . .	46

# Chapter 1

## Introduction

### 1.1 Background

The financial market is always being expressed as a dynamic and complex system to ordinary people, whose thinking is that it is almost impossible to forecast the consequences of various financial products, especially the stock market; that it is an investment activity which involves a large number of participants. In fact, to those who are professional investors, it is possible to understand, analyse and forecast future movement.

Compared to others, professional investors will have an expanded response to all kinds of information, whether they are related directly or not. Based on their well-built knowledge base, they are able to predict the performance of a particular product in the near future. Professional stock analysts believe that the information from investors could be transformed into a real application, by using computational data mining techniques to discover the correlations between individual stock and a referencing factor.

In the stock world, the market acts like a river. The water flows through every single minute during the trading period. Whether it flows rapidly or calmly, it is really dominated by the resistance from the riverbed underneath. It is the same thing when we are concerned with stock market activities from a financial point of view. Geologists well understand that the behaviour of a river's flow depends on the knowledge of riverbed structure. It is the same in the stock market: professional investors always have a better knowledge of the market than the ordinary investors. Thus, they can make predictions for the next movements in the market. Unfortunately,

most traders do not have that kind of vision of certainty. Our job is to discern a better knowledge of stock market with well-known macroeconomic referencing factors, and demonstrate the relationships between each of the components within the structure in terms of a numerical model. Such a correlation analysis could be described as pattern recognition, which is an embranchment study of data mining.

Most of existing research in this field has mainly focused on direct stock correlation analysis. This research proposes to work differently, by introducing a referencing approach as a key feature in our stocks cross-correlation applications. The target of this research is to develop numerical methods for extracting and representing the interrelationship between stock products with respect to a pre-defined referencing factor. The proposed computational correlation model is trained over the historical data of the stock price and the referencing factor, and the generated stocks correlation indicates in practice the future movements of a stock or a group of stocks.

## 1.2 Research Objectives

Although it is too hard to accurately predict a stock's movement, it does not mean that the stock's price is generated randomly. There is certainly some concealed knowledge associated with the movement of stocks, and investors will benefit from such knowledge for their portfolio diversification once that knowledge is extracted (Zargham & Sayeh, 1999; Liu & Zargham, 2006). This research aims to develop referencing featuring linear cross-correlation extraction for performing an efficient and reliable stocks analysis. The proposed method is established on the basis of both fundamental and technical financial analysis, integrating with well-known macroeconomic factors as a parameter. The objectives included:

- Propose an cross-correlation algorithm which is able to analyze stocks cross-correlation with reference to specified macroeconomic factors;
- Verify the proposed algorithm by implementing it with real world data and applications.

## 1.3 Research Contributions

In contrast to traditional stocks correlation analysis, the employment of referencing approach for computational correlation knowledge discovery will enhance the accuracy, credibility and intelligibility of interrelationship inbetween each pair of stocks, because of those pre-defined referencing factors are characterised with remarkable stability and reputation in the finance field.

The contributions of this research are summarized as follows:

- The development of computational single referencing approach to stocks correlation analysis;
- The proposed referencing stock cross-correlation is extended into a multi-class scenario which improves the knowledge understanding;
- The effectiveness of the proposed algorithm for referencing stocks correlation analysis is verified by theoretical analysis and real-world application tests.

## 1.4 Notation Declaration

For the convenience of method derivation and clarity of presentation, we summaries most notations used in the paper in Table 1.1.

Notation	Descriptions
$\mathbf{X}$	Vector of Stock's price, $\mathbf{X} \in \mathbb{R}^p$
$\mathbf{Z}$	Vector of a referencing factor, $\mathbf{Z} \in \mathbb{R}^1$
$p$	Number of sample stock products
$k$	Number of referencing factors
$n$	Number of sample stock's price samples
$t$	Period of time
$\rho$	Correlation coefficient
$PA$	Channel pattern
$d_c$	Channel distance
$C$	Multiple Correlation Coefficient
$\hat{S}$	Diagonal Precision(Correlation) matrix of two sample stocks

Table 1.1: NOTATIONS

## 1.5 Thesis Structure

**Chapter 2** reviews previous studies on financial correlation analysis and discusses our motivation for the present research. In the review, both fundamental and technical stock cross-correlation analysis methods are underlined.

**Chapter 3** describes the principle and practice of three existing computational correlation techniques: statistical correlation analysis, graphical correlation analysis and conditional covariance correlation analysis.

**Chapter 4** presents the proposed computational referencing approach to stocks cross-correlation analysis. This chapter introduces first the mathematical ground of Pearson product-moment Correlation and Multiple Correlation Coefficient (MCC), which are two core elements of the proposed methodology. Then, a pair of stocks cross-correlation is modeled as an MCC calculation over the Pearson connections between individual stocks to a pre-defined referencing factor.

**Chapter 5** looks at the applications and analysis over S&P 500 stocks data that include two referencing featured stocks correlation analyses. In each application, we use the U.S. crude oil price and the S&P 500 indices price respectively as the referencing factor to compute correlation of a selected family of stocks, with them being compared to the results of same correlation analysis applied to random selected stocks within the scope of S&P500.

**Chapter 6** presents the conclusions of the thesis and directions for future work.

## Chapter 2

# Literature Review of Fundamental & Technical Correlation Analysis

### 2.1 Efficient Market Hypothesis

The Efficient Markets Hypothesis (EMH) should always be considered prior to any stock-related correlation analysis (Leigh, Purvis & Ragusa, 2002). Fama (Fama, 1991) stated that the prices of a stock market normally follow a random walk and are unpredictable from their past behavior. According to the principle of EMH, the market efficiency has been divided into three degrees: Strong form, semi-strong form and weak form. In the strong form of EMH, all the knowable information influences the price of the market immediately. The semi-strong form states that the market's price will be affected immediately, when all the related information becomes known to the public. Also, anyone with private information will be able to make profit on it. The weak form of EMH states that, by investigating the historical data, only information obtained from the investigation will be factored into the market's price immediately. There are numerous financial literatures revealing the most common anomalies that involve abnormal returns in relation to U.S. government monetary policy announcement, unexpected earnings announcements, recommendation from professional analysts and others (Leigh et al., 2002; Hong & Stein, 1999; Hong, Lim & Stein, 2000). Furthermore, Raghubir and Das (Raghubir & Das, 1999) classified the behavioral anomalies as time-series patterns, price and return effects, volume and volatility effects and miscellaneous effects.

## 2.2 Fundamental Stocks Correlation Analysis

In fundamental stocks correlation analysis, the centre of attention is the investigation of stocks' characteristics, regardless of any numerical financial figures calculation. Economic intuition supports the idea that firms in the same industry share high return correlations compared to firms in different industries. However, stocks can be categorized into homogeneous groups using criteria other than industry affiliation.

The existence of a relationship between fundamental economic variables and financial market returns is evidentially proved by numerous studies. In addition, microeconomic variables are also assigned as the explanatory factors of the volatility in financial market (Bilson, Brailsford & Hooper, 2001).

The existing literature provides that academic researchers and investment practitioners follow a variety of approaches to construct homogeneous stock groupings. Purely statistical procedures can be applied to the problem. Early heuristic approaches for partitioning stocks into similar groups are proposed by Farrell (Farrell, 1974), Elton and Gruber (Elton & Gruber, 1970). Brown and Goetzmann (Brown & Goetzmann, 1997) cluster mutual funds into distinct investment-style categories. Alternatively, stocks can be assigned to groups on the basis of a priory economic attribute such as market capitalization or operating performance.

Chan et al state that (Chan, Lakonishok & Swaminathan, 2007), financial researchers and analysts frequently grapple with the issue of identifying homogeneous groups of stocks. For example, when analyzing the consequences of events such as changes in financial policies. A common procedure is to group any stocks unrelated to the event, and similar in all other respects. The behavior of such stocks is then assessed against the reference group. Assessments of a stock price in terms of comparisons with others frequently appear in their financial statement to the public. Many investors follow strategies based on identifying stocks that are matched along economically relevant dimensions.

Fama and French(Fama, 1991) developed a list of industry assignments that have become broadly accepted in academic research studies. Their studies also include the measurement of homogeneity in terms of coincidence in stock price movements. Two firms with a similarity in economics are not necessarily experiencing a strong return covariation over short horizons, plus their co-movement may be drowned out by the idiosyncratic portion of returns, including noneconomic forces such as investor sentiment.

The above literature reviews only draw a brief understanding of the fundamental financial correlation analysis. However, since our research concerns are on specified financial problems, those general studies will not be sufficient for establishing our knowledge base. The relevance of various fundamental economic factors is demonstrated in extant literature, such as exchange rates, interest rates, monetary policy, commodity price, purchase power and regional stock market indices. Therefore, we extend our literature review further by studying the fundamental correlation between well-known economic indexes (commodity, macroeconomical news announcements and monetary policy) and stocks in particular.

### **2.2.1 The Relationship Between Changes in Commodity and Stock Price**

Matia et al (Matia, Amaral, Goodwin & Stanley, 2002; Matia, Ashkenazy & Stanley, 2003) have found that, unlike stock and currency markets, the commodity market is much less influenced than other financial securities. This is due to the commodities are always represents phisical products, which requires transportation and storage, as a result of this, the market activity may exhibit a slower response to change in demand. They also found that, during the same period, the returns of stocks and commoditie are exhibiting with a similar multifractal behavior, but the range of commodities' multifractal behavior is much extensive than stocks'.

Mensi et al (Mensi, Beljid, Boubaker & Managi, 2013) stated that, in the transmission of volatility between financial market, liberalization and cointegration are the key elements, especially between crude oil markets and stock markets. Moreover, Maghyreh and Al-Kandari (Maghyreh & Al-Kandari, 2007) concluded that the GCC countries' stock prices were directly impacted by crude oil prices, by using multivariate GARCH model and daily closing global crude oil and stock prices from 1994 to 2001. There are plenty of studies prove the association between commodity market and stock market in practical. Malik and Hammoudeh (Malik & Hammoudeh, 2007) examined the stocks volatility and realated change in commodity price together, as a result, significant volatility spillovers has been found between US stock market and global crude oil market. In the extension of their study, they revealed that, apart from Saudi Arabia, all the stock markets in the Gulf Cooperation Council Countries are connected with the global crude oil market tightly. In another work, Park and Ratti (Park & Ratti, 2008) analyzed the association between global crude oil

price and stock returns in United States and 13 European countries through VAR model over the period of 20 years, they have found that, global crude oil price had a powerful effect on stock returns in European countries, rather than U.S., which are out of their expectations. Furthermore, Arouri et al. (Arouri, Jouini & Nguyen, 2012) employ VAR-GARCH model to test the volatility spillovers between global crude oil price and European stock price from particular sector, and they have found the significant volatility spillovers between them. Similarly, Malik and Ewing (Malik & Ewing, 2009) have found the evidence of strong linkage between shock in crude oil market and specific U.S. stock market sectors in the period between 1992 and 2008, by using bivariate GARCH models.

As Creti, Joëts and Mignon (Creti, Joëts & Mignon, 2013) instructed in their research, figure 2.1, 2.2 and 2.3 illustrate the similarity in terms of volatility between S&P 500 stock returns and commodity price return, which gives us a clear picture of the correlation between price of stock product and price of commodity.

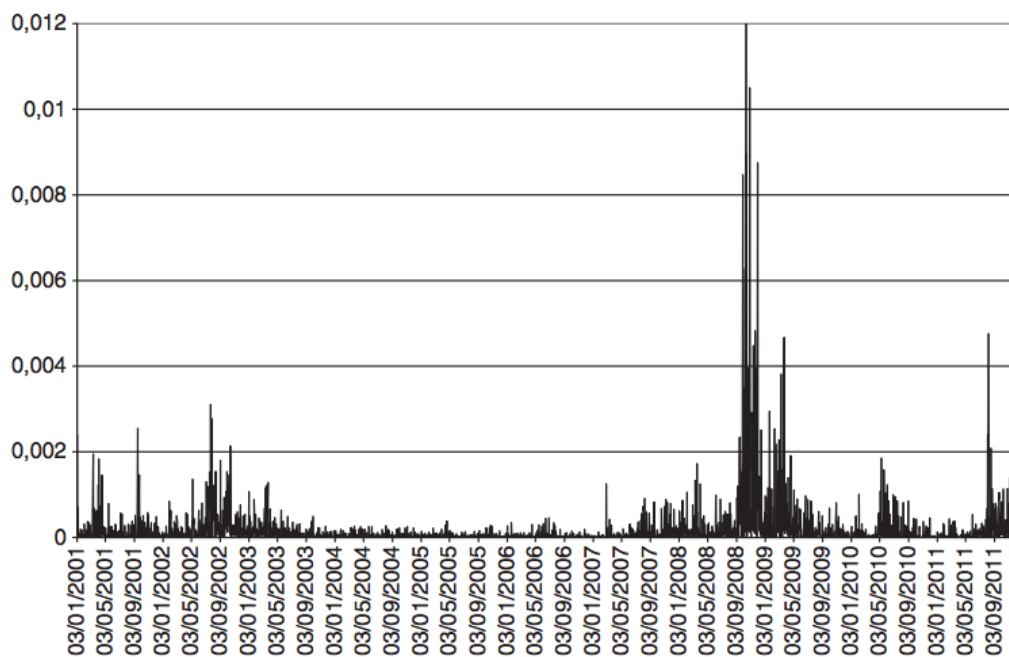


Figure 2.1: S&P 500 stock returns volatility (01/03/2001-28/09/2011).

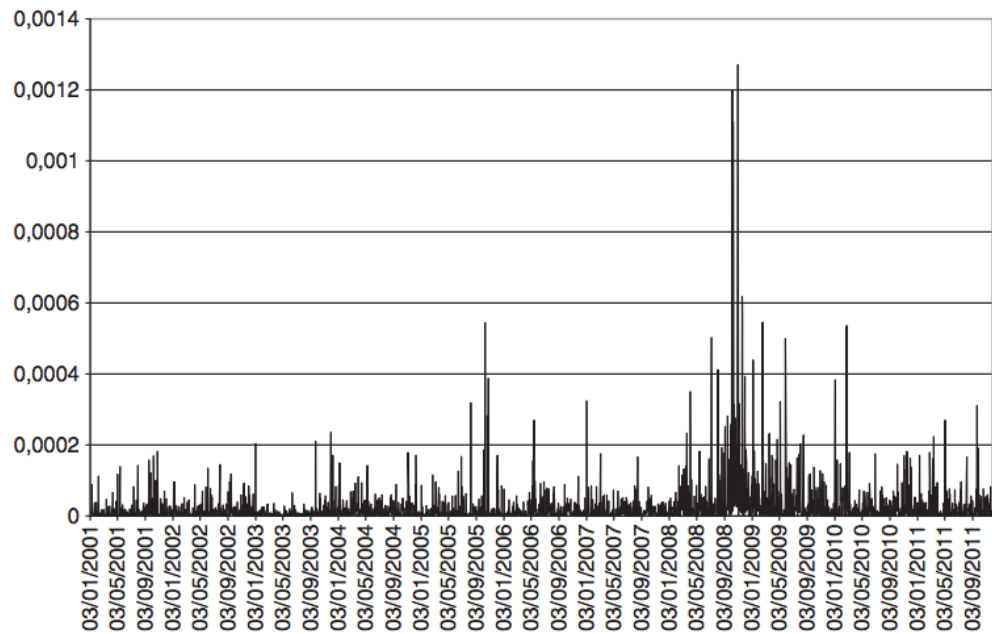


Figure 2.2: Commodity price returns volatility (01/03/2001 to 11/28/2011).

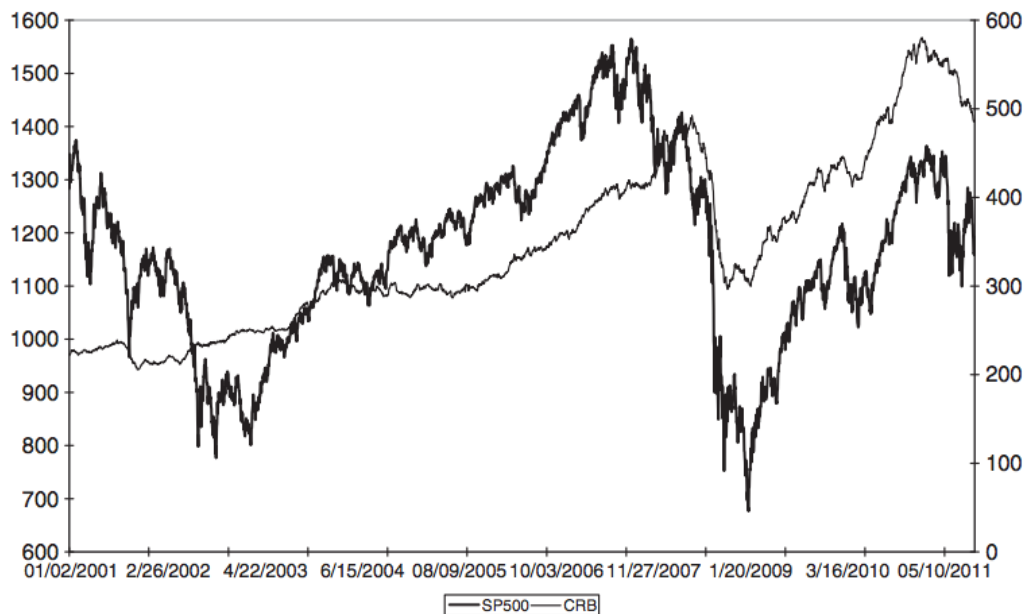


Figure 2.3: Evolution of S&P 500 and CRB indexes (01/03/2001 to 11/28/2011).

Such literatures are strongly supported our hypothesis on the associations between

US stock market and commodity market, especially the linkage between crude oil price and stock price.

### **2.2.2 The Relationship Between U.S. macroeconomic news announcements and Stock Price**

Nikkinen et al (Nikkinen, Omran, Sahlström & Äijö, 2006) stated that, all the investors around the world are interested in the status of the United States' economy because of its dominating position in the world economy. Consequently, macroeconomic news from the United States has become the main issues of interest on stock markets worldwide. They also concluded that, the significance of the scheduled news announcement varies for both United States stock market operating investors. Bollerslev, Cai, and Song (Bollerslev, Cai & Song, 2000) and Graham, Nikkinen, and Sahlström (Graham, Nikkinen & Sahlström, 2003) have conducted similar studies on this kind of spillover phenomenon. The impact of United States macroeconomic news on overseas stock market has been also discovered by various scholars that, similar to United States domestic stock market, the response of overseas stock market to United States news announcement varies. Such evidences for European stock markets have been documented by Becker, Finnerty, and Friedman (Becker, Finnerty & Friedman, 1995; Nikkinen & Sahlström, 2004). In addition, Connolly, Wang and Kim (Connolly & Wang, 2003; S.-J. Kim, 2003) have found the similar characters in Asia-Pacific stock markets.

Regarding to the effect of anticipated news announcement on volatility of stock's return, it is hard to investigate, unless several assumptions have been made (Nofsinger & Prucyk, 2003). Kim and Verrecchia (O. Kim & Verrecchia, 1994) presented a analysing model based on investors are not acquiring private information before the announcement, and they have found that, the stock will significantly affected after the announcement is made. In their another research (O. Kim & Verrecchia, 1991), they found the stock price changes proportionally by the unforeseen part of the news, by assuming investors are able to receive private information before the public announcement, and apply such private information according to their judgement.

### **2.2.3 The Relationship Between Changes in Money Policy and Stock Price**

Recently, more financial studies analysis the reaction of stock market when relevant monetary policy is changing (Hess, 2004). There is an interpretation has been made by Thorbecke (Thorbecke, 1997) to indicate that, the monetary policy makes significant effects on stock market activity, and such effects are primarily caused by risk premium (Patelis, 1997). However, there is no empirical evidence shows that, pattern exists between stock price and change in monetary policy. This is might due to the different monetary policy measures apply to different region around the globe, but this is debatable, as the global economic and financial integration (Bernanke & Mihov, 1998; Strongin, 1995). There also an argument regarding to this problem, Cover and Garcia (Cover, 1992; Garcia & Schaller, 2002) found that, the negative change in money policy change will results greater impact to the stock market, rather than positive changes, and also the scope of the reaction vary with the state of the economy at that particular time period. Such above concepts describe the significance of fundamental stock correlation analysis, despite subjective determination may causes minor insufficient performance in terms of accuracy. Therefore, analysis in technical perspective view has been brought to the front, and becomes a popular topic in this field.

## **2.3 Technical Stock Correlation Analysis**

Technical analysts are always concerned with the dynamics of the market price and volume behavior, rather than any fundamental economic nature of the securities. Charles Dow developed the original Dow Theory for technical analysis in 1884, and a modern explication is found in Edwards and Magee (Edwards, Magee & Bassetti, 2007; Leigh et al., 2002).

Figure 2.4 shows the general knowledge of technical analysis, and various widely used technical indicators in stock market analysis (Beat Grunder & Zingg, 2012).

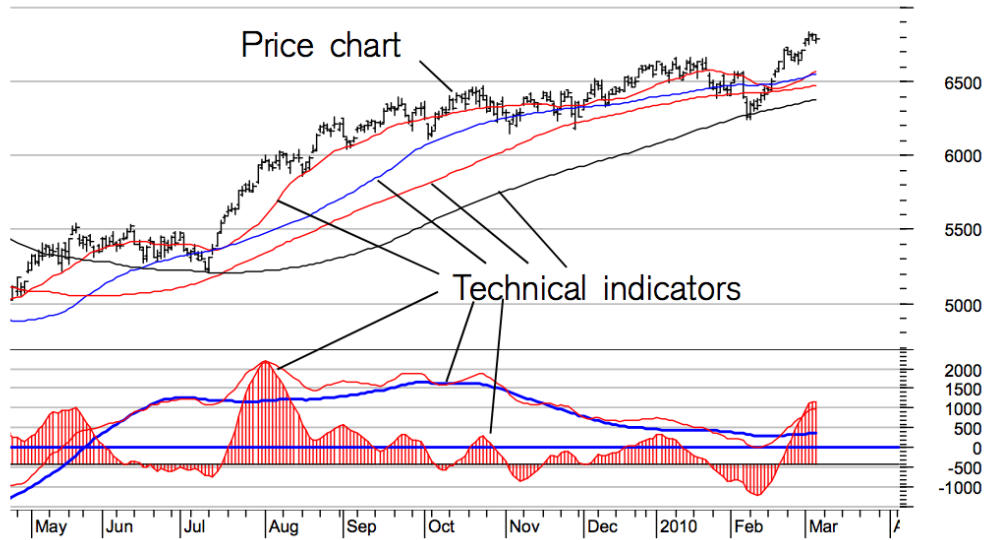


Figure 2.4: Commonly used technical indicators for stock analysis

According to Janssens's contribution (Cory Janssen & Murphy, 2013), prior to any technical analysis, there are assumptions that need to be made:

- The stock market discounts everything by assuming at any given time that the price of a stock reflects all the influencing factors, including fundamental factors. Therefore in a technical analysis, everything has been priced into the stock, and any fundamental factors are disregarded. Price movement is the only concern left.
- The trend of a stock's price movement is available. The reason that most technical trading strategies are based on this assumption is that, once the price trend has been established, the future price movement is most likely to follow the previous direction rather than move against it.
- An occurrence in the past is repeatable: since the participants are more likely to react consistently to similar market status, this kind of market psychology causes the repetitive characteristic of price movement. This assumption enables users to analyse and understand the market movement by using patterns from charts. This gives an opportunity to discover valuable knowledge from a chart, even it has been used over several decades, as a similar price movement may occur again.

### 2.3.1 Moving Average

Since the stock market movement varies over time, dynamically and evolutionarily, it is hard to track any particular rule that the market follows. However, similar market co-movements could be established from the historical data (Lei, 2010).

In order to capture the homogeneous moving pattern between the index and its components, the Moving Average (MA) needs to be introduced into the system. By employing MA, the noise oscillation information will be eliminated from the original curve. So the new moving average curve will be much more stable than the original, and hence the price trend is much easier to measure. In this paper, we are using a 120 days indicator, which is the long-term average index, also called the average line index (Huang, 2011; Zhao & Han, 2010).

Moving Average curve can be calculated simply as:

$$average(x_a) = \left\{ \frac{1}{k} \sum_{n=i-k+1}^{n=i} x_n, (i \geq k) \right\}, \quad (2.1)$$

where  $k$  is the number of days during which the average price is calculated. In the stock market, day unit of its data means a trade day.

Figure 2.5 illustrate the difference between 5-day moving average and 21 day moving average. As shown in the figure 2.5, the fluctuations of stock price have been smoothed out, As a result, the determination of the underlying trend has been made easier (Beat Grunder & Zingg, 2012).

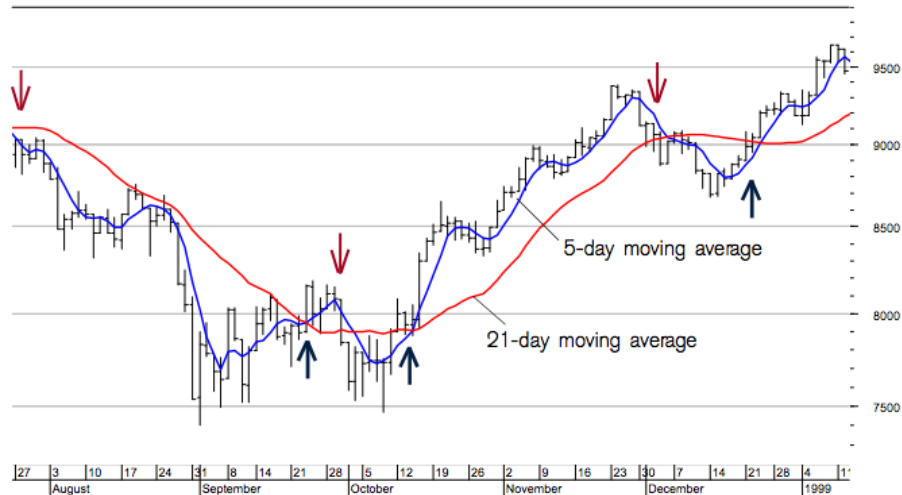


Figure 2.5: Difference between 5-day moving average and 21 day moving average.

### 2.3.2 Data Mining driven Stock Correlation Analysis

Along with the development of communication network and the trend of information disclosure, the detailed financial information is now easy to find. Efficient data mining techniques including machine learning and statistical mathematics are rapidly growing, and widely available (Hegland, 2001). The significance of the data mining is not limited to the modeling based on massive data, but the ability to capture complicated and ambiguous characteristics of the data.

In one of the previous studies, Washio et al (Washio, Shinnou, Yada, Motoda & Okada, 2007) used learning techniques provided in the data mining tool Weka for stocks correlation analysis, finding that every financial index in their experiment does not have any strong correlation with future profit, but most of the indices synthetically determine the future profit.

Beyond that, Fu and Chung (Ting, Fu & Chung, 2006) conducted a sequential (intra-stock mining) and non-sequential association analysis (inter-stock mining) by transforming numeric stock data to symbolic sequences. The research discovered frequently appearing stock price patterns, and the strong inter-relationship among several stocks.

Fiol-Roig et al (Fiol-Roig, Miró-Julià & Isern-Deyà, 2010) used data mining techniques to convert the past stock price data into a decision tree, which shows a graphical representation of a reasoning process and represents a sequence of decisions,

interrelationships between alternatives and possible outcomes.

Kumar and Kalia (Kumar & Kalia, 2012) have done another study to combine the use of symbolic representation and Euclidean distance measurement to establish relationships among various stocks. It has been found that symbolic representation provides an easier interpretation and helps to determine an overall pattern. Symbolic pattern is having a resemblance with price change patterns in numeric representation.

## 2.4 Summary

According to Fama (Fama, 1991), the selection of referencing factors is subject to criticism on the ground of subjectivity and the arbitrary nature of the selection process, which are unavoidable problems in this area of research. However, the selection of a referencing factor will be determined by the aspect of the research. Therefore, using either traditional fundamental or technical correlation analysis is insufficient to provide creditable outcomes. In the next chapter, we will review three existing computational correlation analysis methodologies, which will contribute a better solution to our problem with the combination of traditional correlation analysis methods.

# Chapter 3

## Existing Computational Correlation Analysis Methods

In recent years, financial correlation analysis methods have been widely used with extensive and productive efforts. The research has been expanding from simple to more advanced techniques both in depth and breadth. The computational correlation analysis is normally based on two financial variables.

### 3.1 Statistical Computational Correlation methods

#### 3.1.1 Similarity in Inner Product Space Measurement

The inner product space is a simple approach to computational correlation analysis. According to Emch's study (Emch, 1972), the inner product is a geometrical operation that measures the magnitudes of two equal-length sequences of vectors and the cosine of the angle between them. It is defined as:

$$x \cdot y = \sum_{i=1}^n x_i y_i, \quad (3.1)$$

according to the principle of trigonometry, the angle  $\Theta$  between vectors  $x$  and  $y$  is related to their inner product by

$$x \cdot y = \|x\| \|y\| \cos \Theta. \quad (3.2)$$

The normalized inner product will be given as an equation 3.3 after rearrangement of the equation

$$\cos(\theta) = \frac{X \cdot Y}{\|X\|_2 \|Y\|_2}, \quad (3.3)$$

where  $\cos(\theta)$  is the measurement of how well two variables point in the same direction, and identifies the similarity in trend. Figure 3.1 illustrates the concept of solving  $\theta$ .

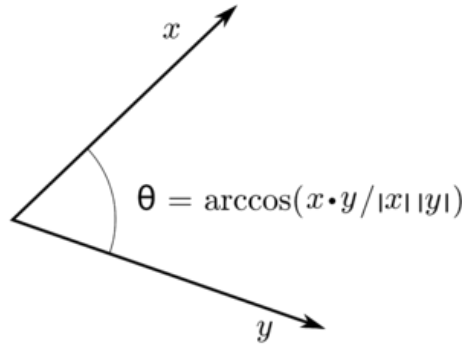


Figure 3.1: Illustration of Inner Product Angle  $\theta$

Since  $\cos \theta \in [-1, +1]$ , there will be three conditions for  $\cos \theta$ :

- $\cos \theta > 0$  indicates two vectors are pointing into the same half space of  $\mathbb{R}^n$ , and perfect alignment (where  $\theta = 0$ ) occurs when  $\cos \theta = 1$ ;
- $\cos \theta < 0$  indicates two vectors are pointing into separate half spaces, and they are pointing exactly opposite directions, under the condition of  $\cos \theta = -1$ ;
- $\cos \theta = 0$  due to  $\theta$  is an odd multiple of  $\frac{\pi}{2}$ , the two vectors are orthogonal. Therefore, the indication of the observation is uncorrelated.

### 3.1.2 Pearson's product-moment Correlation Coefficients

Pearson product-moment correlation is a widely-used measurement of the strength of linear dependence between two variables. The Pearson product correlation coefficient ( $\rho_{X,Y}$ ) has been defined as:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E((X - \mu_X)(Y - \mu_Y))}{\sigma_X \sigma_Y}, \quad (3.4)$$

where  $X$  and  $Y$  are two time series data:  $X = \{x_1, x_2, \dots, x_n\}$  and  $Y = \{y_1, y_2, \dots, y_n\}$  respectively;  $cov$  stand for covariance;  $\sigma_X$  and  $\sigma_Y$  are the standard deviations of variable  $X$  and variable  $Y$  respectively;  $\mu_X$  and  $\mu_Y$  are the mean value of the population; and  $E$  is the expected value operator (Pearson, 1895). In practice, p-value related hypothesis tests are always involved with the correlation calculation. p-value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed during the statistical significance testing. By assuming the null hypothesis of  $p = p_0$  is true, thus, the sample pairs are independent and identically distributed. P-value is crucial for assessing the statistical significance of the correlation, therefore, p-value is always calculated alongside the Pearson correlation coefficient, which can be interpreted as follows: (Rice, 1989).

- The correlation is statistically significant if the p-value is less than 0.05;
- The correlation is not statistically significant if the p-value is greater than 0.05.

By considering  $\mu_X = E(X)$ ,  $\sigma_X^2 = E[(X - E(X))^2] = E(X^2) - E^2(X)$  and in the same way for  $Y$ , also  $E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$ , the Pearson correlation equation 3.4 also can be expressed in terms of uncentered moments as:

$$\rho_{X,Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - (E(X))^2} \sqrt{E(Y^2) - (E(Y))^2}}. \quad (3.5)$$

Cohen (Cohen, 1988) provided a guideline for the interpretation of the size of Pearson correlation coefficient. The Pearson correlation contains a positive correlation and a negative correlation, which identify the range of the Pearson correlation coefficient  $\rho_{X,Y}$  is from -1 to +1. A positive correlation exists between two variables when both variables are moving in tandem. That means when one variable increases, the other variable also increases and vice versa. In case of negative correlation, one variable will decrease as other variable increases, and vice versa.  $\rho_{X,Y} = +1$  or  $\rho_{X,Y} = -1$  indicates a perfect positive correlation or a perfect negative correlation respectively.

Figure 3.2, illustrate the scatter diagrams with different values of correlation coefficient ( $\rho$ )

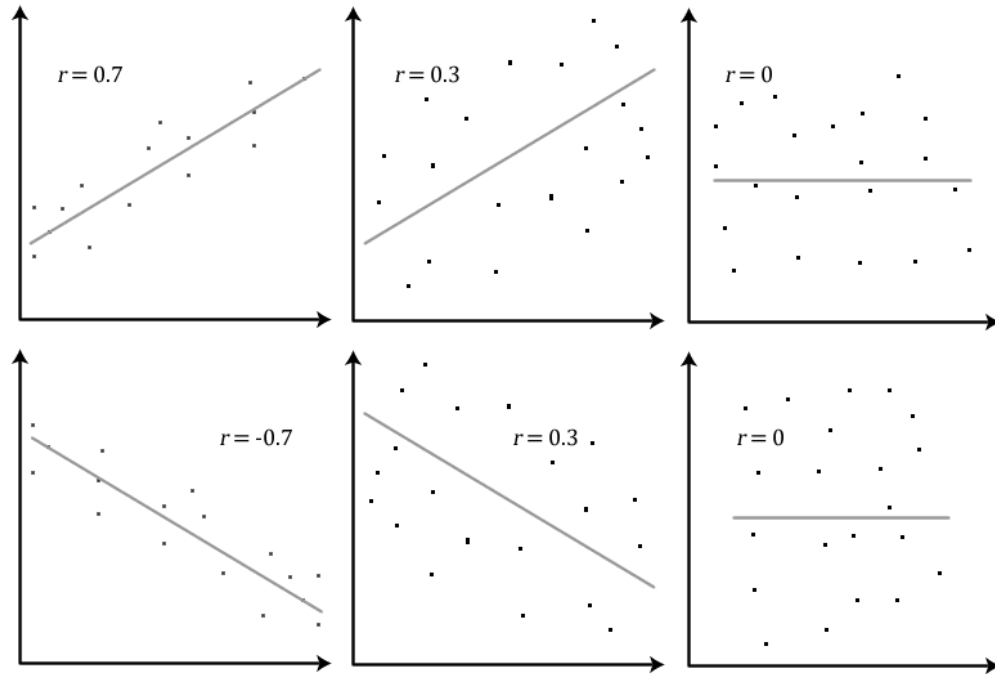


Figure 3.2: Positive, Negative and None association in terms of Pearson Correlation Coefficients

The benefit of implementing the Pearson product-moment correlation is revealed when the association between the two variables is solid. This returns the correlation prediction with high accuracy. The adeptness of the Pearson product-moment correlation application in financial analysis has been demonstrated by several literatures. Kondratenko and Kuperin (Kondratenko & Kuperin, 2003) applied the Pearson product-moment correlation to neural networks in order to forecast the future exchange rates between U.S. dollar and four other major currencies: Euro, British Pound, Swiss Franc and Japanese Yen. They concluded that their neural network application achieved much better results with the participation of the Pearson product-moment correlation extraction. In another similar study, the involvement of the Pearson product-moment correlation with a neural network model returns a better performance in terms of average internode distance (Kwapien, Gworek & Drozd, 2009). However, both studies indicate that the reliability of the Pearson product-moment correlations supported prediction on time series is doubtful.

## 3.2 Graphical Computational Correlation Analysis

Recently, computational graphical correlation analysis has begun a new era of correlation studies in the financial field. According to channel correlation extraction proposed by Pang et al (Pang, Song & Kasabov, 2011), the similarity of trend between two time series variables will be graphically approximated with a concrete arc.

Figure 3.3 shows 4 typical trend patterns: fast growing, slowly increasing, fast dropping and slowly decreasing.

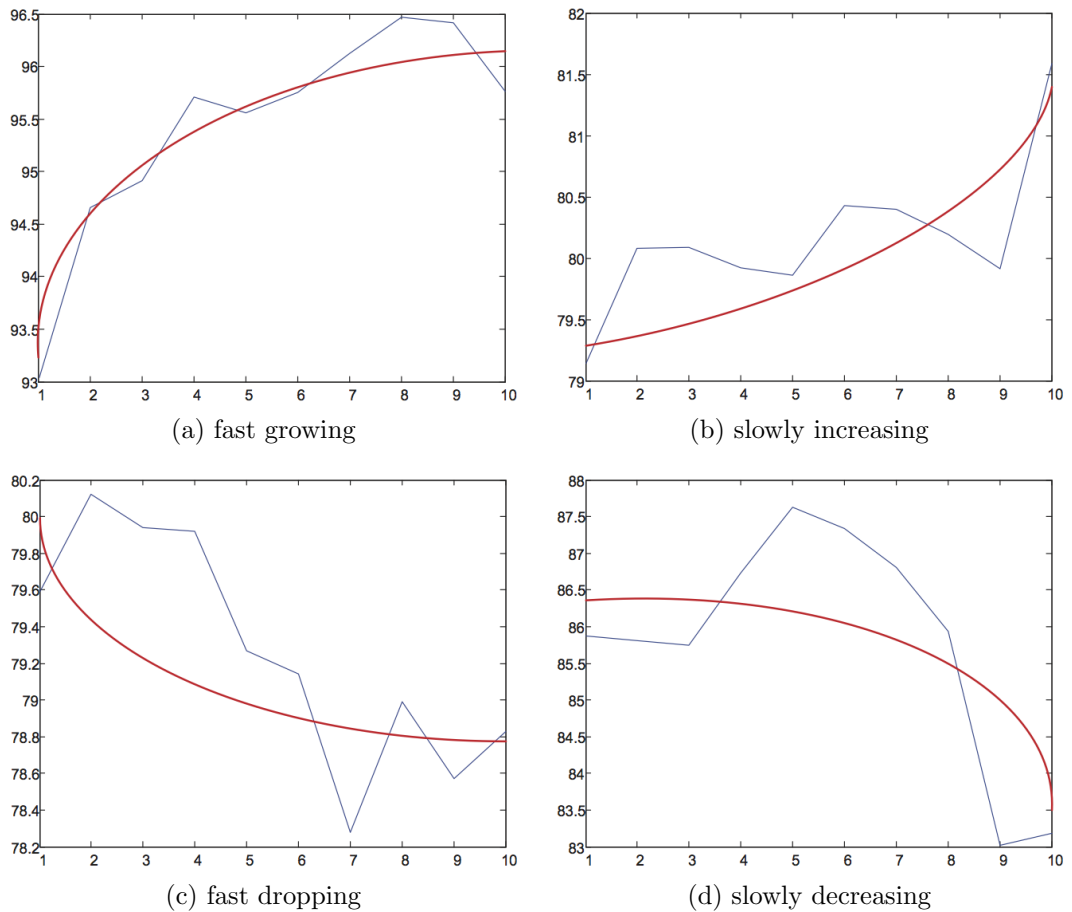


Figure 3.3: Four trend patterns used for channel approximation

The above 4 trend patterns have been approximated by 4 smooth arches, which is obvious and straightforward to understand. The functions for such 4 trend patterns can be obtained by the following equations respectively:

$$(x - x_0)^2 + (y - y_0)^2 = R^2 \left| \begin{array}{l} x_0 = 0, y_0 = R \\ x \in [0, \sin\alpha \cdot R\sqrt{2(1 - \cos 2\alpha)}] \end{array} \right. \quad (3.6)$$

$$(x - x_0)^2 + (y - y_0)^2 = R^2 \left| \begin{array}{l} x_0 = R, y_0 = 0 \\ x \in [0, \sin\alpha \cdot R\sqrt{2(1 - \cos(\pi - 2\alpha))}] \end{array} \right. \quad (3.7)$$

$$(x - x_0)^2 + (y - y_0)^2 = R^2 \left| \begin{array}{l} x_0 = 0, y_0 = 0 \\ x \in [0, (1 - \cos\alpha) \cdot R\sqrt{2(1 - \cos(\pi - 2\alpha))}] \end{array} \right. \quad (3.8)$$

$$(x - x_0)^2 + (y - y_0)^2 = R^2 \left| \begin{array}{l} x_0 = R, y_0 = R \\ x \in [0, (1 - \cos\alpha) \cdot R\sqrt{2(1 - \cos 2\alpha)}] \end{array} \right. \quad (3.9)$$

where  $\alpha \in (0, \pi/4)$ ,  $\angle\alpha$  is used to measure the rate of trend variation, and the radius  $R$  determines the length of the trend pattern corresponding to the time frame of observation. In practice, a discrete arc can be obtained according to the length of the time series for channel approximation.

Figure 3.4 gives a better illustration for interpreting the above 4 arc functions.

Given a vector (i.e. a stock closing price)  $X$  with period of time  $T$ , applying Eq. (3.6) - Eq.(3.9) to  $X$ , respectively, one of 4 types arc (i.e. functions) called ‘channel pattern’ is selected with its parameter  $\alpha$  tuned to best suit the time series under observation:

$$PA = \arg \min_{\alpha, i \in [1, 4]} \frac{\sum_{t=1}^T \|p_t^i - x_t\|}{T}. \quad (3.10)$$

In addition, based on the channel pattern ( $PA$ ) extracted by equation 3.10, the channel distance of an observation  $X$  to the referencing factor  $Z$  (i.e, WTI Crude Oil) can be defined as below:

$$d_c = \frac{\sum_{t=1}^{T-1} ((PA_{t+1}^y - PA_t^y) + (PA_{t+1}^x - PA_t^x))}{T - 1}. \quad (3.11)$$

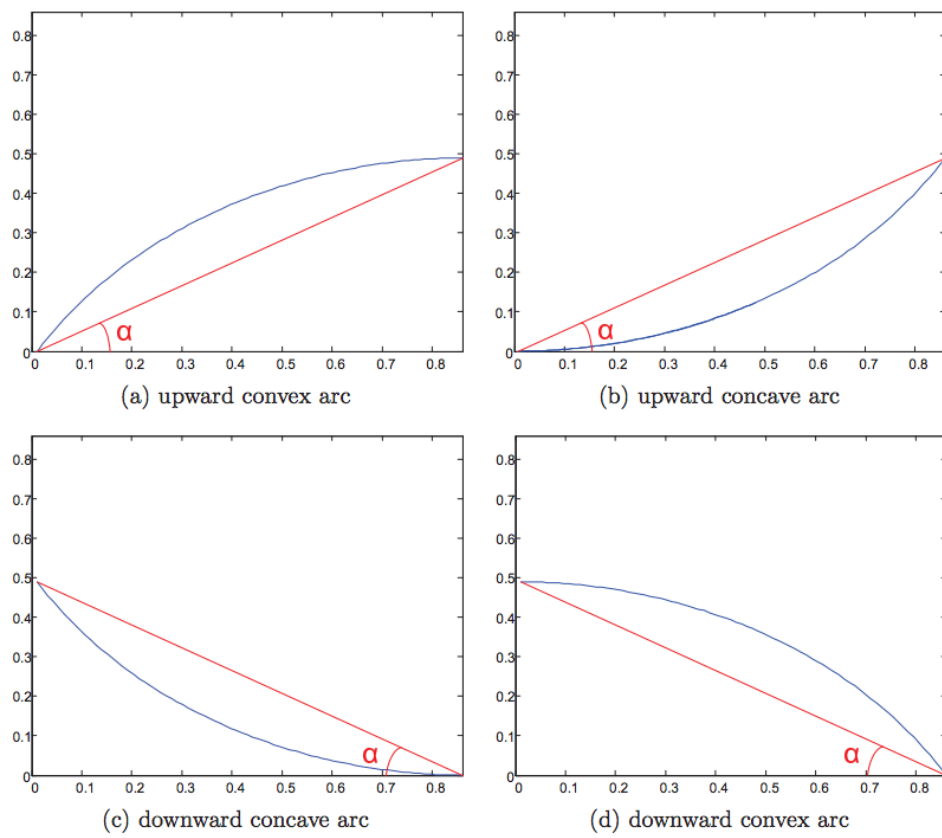


Figure 3.4: 4 arc rulers corresponding to 4 trend patterns shown in Figure 3.3

### 3.3 Conditional Covariance Correlation Analysis

Conditional covariance correlation analysis is another approach to analysis stocks correlation, which is a solution of the typical statistical problem: conditional covariance matrix estimation. In Kolar, Parikh and Xing's research (Kolar, Parikh & Xing, 2010), a penalized kernel estimator is employed to develop a new nonparametric model, for selecting non-zero elements of the precision covariance matrix, which conditionally vary to a reference variable. They defined the problem of conditional covariance selection, e.g.: measuring the cross-correlation between stocks with respect to a referencing factor can be expressed as the estimation of non-zero elements in the conditional precision matrix  $\Omega(Z)$ , where  $Z$  represents a given referencing factor variable. Since the cross-correlation between different stocks of  $X$  can be formulated in terms of partial correlation coefficient  $\rho_{ab}(z)$ , the problem can be computed as:

$$\rho_{ab}(z) = -\frac{\omega_{ab}(z)}{\omega_{aa}(z)\omega_{bb}(z)}. \quad (3.12)$$

Lauritzen (Lauritzen, 1996) found that, the partial correlation coefficients is related to a regression model, and for the above equation, it can be reformulated as:

$$X_a = \sum_{a \neq b} X_b c_{ab}(z) + \epsilon_a(z). \quad (3.13)$$

Since

$$c_{ab}(z) = -\frac{\omega_{ab}(z)}{\omega_{aa}(z)} = \rho_{ab}(z) \frac{\omega_{bb}(z)}{\omega_{aa}(z)}, \quad (3.14)$$

the relationship given in equation 3.14 is used for estimating any non-elements of the conditional precision matrix (Kolar et al., 2010).

At the beginning of estimation for stock a, matrix  $\tilde{C}_a^{(0)}$  needs to be obtained from the initiation with a specified threshold. After that, they built up a regression algorithm to literally update every elements of the conditional covariance matrix  $C_a$ . The loss function for updating stock a within the sample population  $p$  is defined as:

$$L_a(\mathbf{C}_a; D^n) := \sum_{z \in \{z^j\}_{j \in [n]}} \sum_{i \in [n]} (x_a^i - \sum_{b \neq a} x_b^i c_{ab}(z))^2 K_h(z - z^i) + 2\lambda \sum_{b \neq a} \|\tilde{c}_{ab}(\cdot)\|_2, \quad (3.15)$$

where  $D^n = \{(x^i, z^i)\}_{i \in [n]}$ ,  $a \in [p]$ , and  $C_a = (c_{a(z^1)}, \dots, c_{a(z^n)})$ ,  $c_{a(z^j)} \in \mathbb{R}^{p-1}$ , and

gradient decent progress is also taken in this place. In order to optimize the result, Karush-Kuhn-Tucker (KKT) condition 3.16 will be examined for eliminating any  $c_{ab}(z^j)$  that valued towards zero.

$$\frac{1}{2\lambda} \sum_{z \in z^j_{j \in [n]}} \left( \sum_{i \in [n]} x_b^i r_{ab}^i(z) K_h(z - z^i) \right)^2 \leq 1. \quad (3.16)$$

As the result, the minimizer of the loss for stock a will be defined as:

$$\hat{\mathbf{C}}_a := \underset{\mathbf{C} \in IR^{(p-1) \times n}}{\operatorname{argmin}} L_a(\mathbf{C}; D^n). \quad (3.17)$$

Once all single components of  $\{\hat{\mathbf{C}}_a\}_{a \in p}$  have been combined together, the following estimator will be formulated for obtaining all the non-zero elements of the precision(correlation) matrix between stock  $a$  and  $b$ :

$$\hat{S} := \{(a, b) : \max \{\|\hat{c}_{ab}(\cdot)\|_2, \|\hat{c}_{ba}(\cdot)\|_2\} > 0\}. \quad (3.18)$$

### 3.4 Summary

Most of existing studies on computational stocks correlation analysis have provided exceptional methodologies to discover the association between stocks, however most of them mainly focus on non-parametric stocks correlation analysis, without consideration of any referencing factors as the parameter. Our approach is going to fill the blank, as we believe the outcome of stocks correlation analysis by referencing a well-known macroeconomical factor will be remarkable.

# Chapter 4

## The Proposed Correlation Analysis Method

### 4.1 Introduction

One of the most critical stock market analyses is to find the correlations between individual products in every aspect. This could be done by measuring the strength of association between every of the stock products, in terms of both social and financial correlation coefficients. In order to construct a hierarchical map describing the functional grouping of large-scale stock products, an estimate of the similarities between all pairs of expression profiles is needed. The correlation matrix is essential to achieve the above feature, and also, from the review, it evidently plays an important role in classification of stock products. In the case of stock market analysis, one or more economic indicators of interest is introduced at the same time. In other words, the references are used to help us to determine the relationships among stocks, and such references are the conditions, that will influence the association between different stocks.

### 4.2 Motivation for the Presented Research

The motivation of the present research is that the stock market is changing dynamically due to the evolutions of many relevant economic and financial activities (Ciora, Munteanu, Hrinca & Ciobanu, 2011). It will be very hard for non-professional investors to capture useful knowledge amidst all the perplexity. However, it is worth

noting that we can always find some variable whose variation is more traceable and whose impact is much larger than any other individual stock in the market. We believe that using these variables as references (i.e., fundamental truth to decision making) allows us to discover/model the relation of any individual stock's volatility to reference; and this will guide us to interpret the stock market activities more efficiently, and benefit an investor's decision making.

### 4.3 Proposed Computational Referencing Approach

Let  $X = (X_1, \dots, X_\rho)^n \in IR^p$  denote a  $\rho$ -dimensional vector representing price of a series of stock products, and  $Z \in IR^1$  denote an index variable representing an economic indicator of references, e.g. the crude oil price. In this situation, a model is needed for generating a multiple correlation coefficient matrix, which is able to represent the related stocks in clusters, and the classification of the cluster should be close to reality. The calculation will be taken one by one, so each selected stock product will be represented as  $X_a (a \in [p])$ , where  $p$  denotes the population of all the stocks, and  $X_{\setminus a} := \{X_b : b \neq a, b \in [p]\}$  represents all the rest.

The data consists of pairs  $\{x^i, z^i\}$ , the vector of standardized stock daily close prices and the oil price, respectively, obtained over certain period of time. We analyze the data to recover the strength of associations between different stocks as a function of the oil price. The belief is that each stock is associated with a small number of other stocks and that the set of associations is fixed over a time-frame of interest, although the strengths may change.

The mathematical grounds we are using for our methods are Pearson product-moment correlation coefficient and sample multiple correlation coefficients. The combination of two such algorithms helps to measure the strength of association between the independent variables, which are subject to a dependent variable (referencing factor).

Initially, the linear correlation coefficient between the reference and every individual stock products, should be calculated by the Pearson product-moment correlation (4.1):

$$\rho_{\alpha,\beta} = \frac{cov(\alpha, \beta)}{\sigma_\alpha \sigma_\beta} = \frac{E((\alpha - \mu_\alpha)(\beta - \mu_\beta))}{\sigma_\alpha \sigma_\beta}. \quad (4.1)$$

As we are calculating the sample correlation coefficients, by substituting estimates

of the covariance and variances in the above formula and applying it to the practical problem, the formula will be derived as:

$$\rho_{X_a Z} = \frac{\sum_{i=1}^n (X_{a_i} - \bar{X}_a)(Z_i - \bar{Z})}{\sqrt{\sum_{i=1}^n (X_{a_i} - \bar{X}_a)^2} \sqrt{\sum_{i=1}^n (Z_i - \bar{Z})^2}}, \quad (4.2)$$

where  $X_a$  represents stock product  $a$ , and  $Z$  represents a referencing factor

By repeating equation (4.2), the linear correlation between reference  $Z$  and stock product  $X_b$  will be derived as:

$$\rho_{X_b Z} = \frac{\sum_{i=1}^n (X_{b_i} - \bar{X}_b)(Z_i - \bar{Z})}{\sqrt{\sum_{i=1}^n (X_{b_i} - \bar{X}_b)^2} \sqrt{\sum_{i=1}^n (Z_i - \bar{Z})^2}}. \quad (4.3)$$

Once the computation for all stocks has been completed, we should conclude with a  $p \times p$  dimensional correlation matrix:

$$\rho_{XZ} = \begin{bmatrix} \rho_{X_{a_1} Z_1}, \rho_{X_{b_1} Z_1}, \dots, \rho_{X_{p-1} Z_1}, \rho_{X_{p_1} Z_1} \\ \rho_{X_{a_2} Z_2}, \rho_{X_{b_2} Z_2}, \dots, \rho_{X_{p-1} Z_2}, \rho_{X_{p_2} Z_2} \\ \vdots \\ \rho_{X_{a_{n-1}} Z_{n-1}}, \rho_{X_{b_{n-1}} Z_{n-1}}, \dots, \rho_{X_{p-1} Z_{n-1}}, \rho_{X_{p_{n-1}} Z_{n-1}} \\ \rho_{X_{a_n} Z_n}, \rho_{X_{b_n} Z_n}, \dots, \rho_{X_{p-1} Z_n}, \rho_{X_{p_n} Z_n} \end{bmatrix}$$

Next, we are going to use the above outcomes to calculate cross-correlations between each pair of stocks. The mathematical theorem that has been employed here is Multiple Correlation Coefficient (MCC).

The MCC formula for computing two independent variables,  $X_a$  and  $X_b$  is stated as:

$$\rho_{X_a X_b} = \frac{\sqrt{\rho_{X_a Z}^2 + \rho_{X_b Z}^2 - 2\rho_{X_a Z}(\rho_{X_b Z})(\rho_{X_a X_b})}}{\sqrt{1 - \rho_{X_a X_b}^2}}, \quad (4.4)$$

where

- $\rho_{X_a Z}$  = correlation coefficient between  $X_a$  and  $Z$ ;
- $\rho_{X_b Z}$  = correlation coefficient between  $X_b$  and  $Z$ ;

- $\rho_{X_a X_b}$  = correlation coefficient between  $X_a$  and  $X_b$ .

Therefore, the algorithm of the proposed method is outlined as:

---

**Algorithm 1** MCC Matrix Calculation Algorithm for single referencing factor
 

---

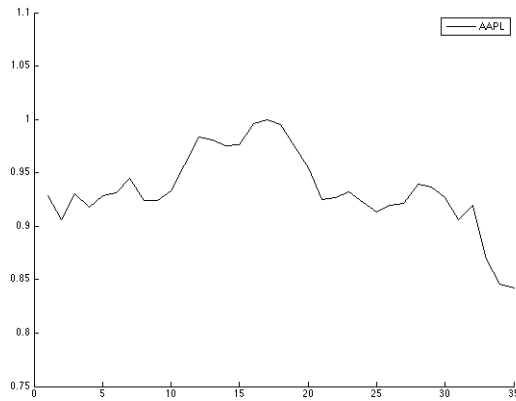
**Input:** Instance Matrix  $\mathbf{X} \in \mathbb{R}^{p \times n}$ ; Column Vector  $\mathbf{Z} \in \mathbb{R}^{n \times 1}$

**Output:**  $\mathbf{C} \in \mathbb{R}^{p \times p}$

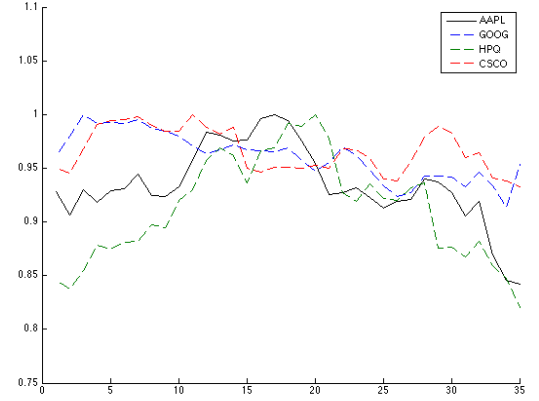
- 1: Combine the instance matrix and column vector into  $\mathbf{D}^n$ ,  $\mathbf{D}^n = \{(x_1, x_2, \dots, x_{p-1}, x_p)^n, z^n\}$
  - 2: Form identity matrix  $\mathbf{C} \in \mathbb{R}^{p \times p}$
  - 3: **for**  $i = 1$  to  $p$  **do**
  - 4:   **for**  $j = 1$  to  $p$  **do**
  - 5:     Compute  $\{C_{x_i x_j}^n(z^n)\}_{i,j \in [p]}$  according to formula (4.3)
  - 6:   **end for**
  - 7: **end for**
  - 8:  $\mathbf{C} = \{C_{x_i x_j}\}_{i,j \in [p]}$
- 

## 4.4 Toy Sample Experiment

At the beginning stage, we only consider a small toy example in order to demonstrate our algorithm's performance.



(a) Price movement of Apple Inc. (AAPL)



(b) Similar trend movements to 4.1a

Figure 4.1: Toy sample experiment for referencing featured stocks correlation analysis

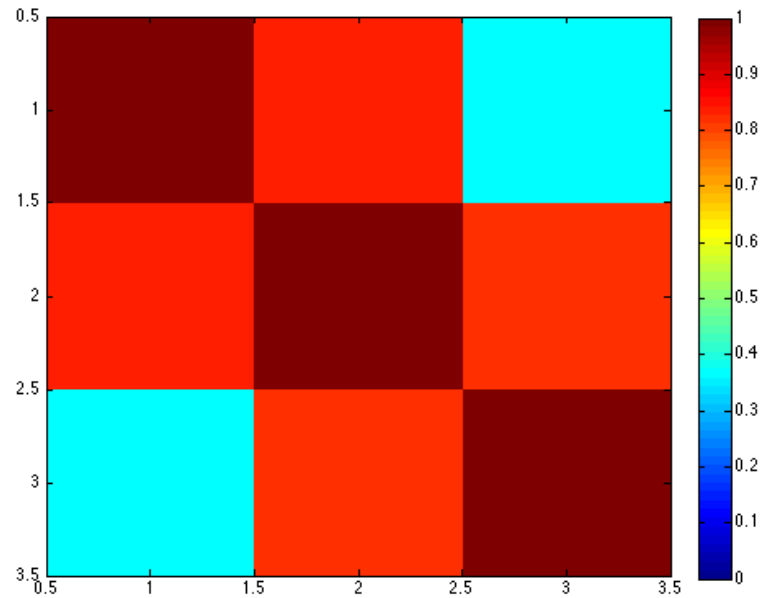


Figure 4.2: Stocks Cross-Correlation Heatmap

As shown in the Figure 4.1, the correlation knowledge between those three stocks with respect to AAPL is accessible from the original price data generated plots, but with limitations. By closely looking at the chart, positive and negative correlations will be drawn attention to at a specified time point. However, in practice, it is impossible to manually extract such correlation knowledge from time series financial data streams.

The proposed algorithm that has been developed in this research will efficiently solve this problem, as has been proved by the toy sample experiment. Figure 4.2 illustrate the outcome from the proposed model based on same input data. The heat map clearly indicates the interrelationship between each pair of stocks in terms of numerical measurement.

## 4.5 Extension to Multiple Referencing Approach

The existence of relationship between fundamental economic variables and financial market returns is evidently proved by numerous studies. In addition, microeconomic variables are also assigned as the explanatory factors of the volatility in financial market (Bilson et al., 2001). The relevance of various fundamental economic factors

are demonstrated by extant literatures; such as exchange rates, interest rates, monetary policy, oil price, purchase power and regional stock market indices. Fama (Fama, 1991) stated that, the selection of referencing factors is subject to criticism on the ground of subjectivity and the arbitrary nature of the selection process, which are unavoidable problems in this area of research. However, the selection of referencing factor will be determined by the aspect of research. According to such limitations and drawbacks, we are extending our research from single referencing approach to multiple referencing approach. It is like using mean value in most of the experiment for eliminating standard errors. Once we take multiple referencing approach into the consideration of the correlation analysis problem, it will be even complicated, as the dimension of the problem is raised from 2D to 3D. The complexity of the computation will be a challenge for us. Anyway, we include the works on multiple referencing approach we have done so far in the following paragraph.

As derived in the early part of this section, in the multiple referencing approach, the correlation coefficient will be calculated as:

$$\rho_{X_a Z^\alpha} = \frac{\sum_{i=1}^n (X_{a_i} - \bar{X}_a)(Z_i^\alpha - \bar{Z}^\alpha)}{\sqrt{\sum_{i=1}^n (X_{a_i} - \bar{X}_a)^2} \sqrt{\sum_{i=1}^n (Z_i^\alpha - \bar{Z}^\alpha)^2}}, \quad (4.5)$$

and

$$\rho_{X_a Z^\beta} = \frac{\sum_{i=1}^n (X_{a_i} - \bar{X}_a)(Z_i^\beta - \bar{Z}^\beta)}{\sqrt{\sum_{i=1}^n (X_{a_i} - \bar{X}_a)^2} \sqrt{\sum_{i=1}^n (Z_i^\beta - \bar{Z}^\beta)^2}}, \quad (4.6)$$

where  $X_a$  represents a stock product, and  $Z^\alpha$  and  $Z^\beta$  represent referencing factor  $\alpha$  and  $\beta$  respectively.

By repeating equation (4.5) and equation (4.6), the linear correlation between reference factors  $\alpha$ ,  $\beta$  and stock product  $X_b$  will be derived as:

$$\rho_{X_b Z^\alpha} = \frac{\sum_{i=1}^n (X_{b_i} - \bar{X}_b)(Z_i^\alpha - \bar{Z}^\alpha)}{\sqrt{\sum_{i=1}^n (X_{b_i} - \bar{X}_b)^2} \sqrt{\sum_{i=1}^n (Z_i^\alpha - \bar{Z}^\alpha)^2}}, \quad (4.7)$$

and

$$\rho_{X_b Z^\beta} = \frac{\sum_{i=1}^n (X_{b_i} - \bar{X}_b)(Z_i^\beta - \bar{Z}^\beta)}{\sqrt{\sum_{i=1}^n (X_{b_i} - \bar{X}_b)^2} \sqrt{\sum_{i=1}^n (Z_i^\beta - \bar{Z}^\beta)^2}}. \quad (4.8)$$

Once the computation for all stocks has been completed, we should conclude with a  $p \times p \times k$  three dimensional correlation matrix, where  $p$  is number of sample stock products, and  $k$  is number of referencing factors.

$$\rho_{XZ^\alpha} = \begin{bmatrix} \rho_{X_{a_1} Z_1^\alpha}, \rho_{X_{b_1} Z_1^\alpha}, \dots, \rho_{X_{p-1_1} Z_1^\alpha}, \rho_{X_{p_1} Z_1^\alpha} \\ \rho_{X_{a_2} Z_2^\alpha}, \rho_{X_{b_2} Z_2^\alpha}, \dots, \rho_{X_{p-1_2} Z_2^\alpha}, \rho_{X_{p_2} Z_2^\alpha} \\ \vdots \\ \rho_{X_{a_{n-1}} Z_{n-1}^\alpha}, \rho_{X_{b_{n-1}} Z_{n-1}^\alpha}, \dots, \rho_{X_{p-1_{n-1}} Z_{n-1}^\alpha}, \rho_{X_{p_{n-1}} Z_{n-1}^\alpha} \\ \rho_{X_{a_n} Z_n^\alpha}, \rho_{X_{b_n} Z_n^\alpha}, \dots, \rho_{X_{p-1_n} Z_n^\alpha}, \rho_{X_{p_n} Z_n^\alpha} \end{bmatrix}$$

$$\rho_{XZ^\beta} = \begin{bmatrix} \rho_{X_{a_1} Z_1^\beta}, \rho_{X_{b_1} Z_1^\beta}, \dots, \rho_{X_{p-1_1} Z_1^\beta}, \rho_{X_{p_1} Z_1^\beta} \\ \rho_{X_{a_2} Z_2^\beta}, \rho_{X_{b_2} Z_2^\beta}, \dots, \rho_{X_{p-1_2} Z_2^\beta}, \rho_{X_{p_2} Z_2^\beta} \\ \vdots \\ \rho_{X_{a_{n-1}} Z_{n-1}^\beta}, \rho_{X_{b_{n-1}} Z_{n-1}^\beta}, \dots, \rho_{X_{p-1_{n-1}} Z_{n-1}^\beta}, \rho_{X_{p_{n-1}} Z_{n-1}^\beta} \\ \rho_{X_{a_n} Z_n^\beta}, \rho_{X_{b_n} Z_n^\beta}, \dots, \rho_{X_{p-1_n} Z_n^\beta}, \rho_{X_{p_n} Z_n^\beta} \end{bmatrix}$$

$$\rho_{XZ^k} = \begin{bmatrix} \rho_{X_{a_1} Z_1^k}, \rho_{X_{b_1} Z_1^k}, \dots, \rho_{X_{p-1_1} Z_1^k}, \rho_{X_{p_1} Z_1^k} \\ \rho_{X_{a_2} Z_2^k}, \rho_{X_{b_2} Z_2^k}, \dots, \rho_{X_{p-1_2} Z_2^k}, \rho_{X_{p_2} Z_2^k} \\ \vdots \\ \rho_{X_{a_{n-1}} Z_{n-1}^k}, \rho_{X_{b_{n-1}} Z_{n-1}^k}, \dots, \rho_{X_{p-1_{n-1}} Z_{n-1}^k}, \rho_{X_{p_{n-1}} Z_{n-1}^k} \\ \rho_{X_{a_n} Z_n^k}, \rho_{X_{b_n} Z_n^k}, \dots, \rho_{X_{p-1_n} Z_n^k}, \rho_{X_{p_n} Z_n^k} \end{bmatrix}$$

In the next stage, the correlations with respect to referencing factors in-between each pair of stocks will be calculated through Multiple Correlation Coefficient(MCC), like stated in equation 4.3.

Due to the time restriction, the derivation of the most efficient MCC in terms of multiple referencing approach is not completed yet, and we will use the similar framework for developing multiple referencing algorithm, where the imcompleted

---

**Algorithm 2** MCC Matrix Calculation Algorithm for two referencing factors

---

**Input:** Instance Matrix  $\mathbf{X} \in \mathbb{R}^{p \times p}$ ; Column Vectors  $\mathbf{Z}^\alpha \in \mathbb{R}^{n \times 1}$  and  $\mathbf{Z}^\beta \in \mathbb{R}^{n \times 1}$ **Output:**  $\mathbf{C} \in \mathbb{R}^{p \times p}$ 

- 1: Combine the instance matrix and column vector into  $\mathbf{D}^n$ ,  $\mathbf{D}^{n|\alpha} = \{(x_1, x_2, \dots, x_{p-1}, x_p)^n, z^{n|\alpha}\}$   
and  $\mathbf{D}^{n|\beta} = \{(x_1, x_2, \dots, x_{p-1}, x_p)^n, z^{n|\beta}\}$
  - 2: Form two identity matrices  $\mathbf{C}^\alpha \in \mathbb{R}^{p \times p}$  and  $\mathbf{C}^\beta \in \mathbb{R}^{p \times p}$
  - 3: **for**  $i = 1$  to  $p$  **do**
  - 4:   **for**  $j = 1$  to  $p$  **do**
  - 5:     Compute  $\{C_{x_i x_j}^{n|\alpha}(z^{n|\alpha})\}_{i,j \in [p]}$  and  $\{C_{x_i x_j}^{n|\beta}(z^{n|\beta})\}_{i,j \in [p]}$  according to formula (4.3)
  - 6:   **end for**
  - 7: **end for**
  - 8:  $\mathbf{C}^\alpha = \{C_{x_i x_j}^{n|\alpha}\}_{i,j \in [p]}$ ,  $\mathbf{C}^\beta = \{C_{x_i x_j}^{n|\beta}\}_{i,j \in [p]}$
- 

algorithm for multiple referencing approach is outlined as:

As those noduses we have addressed for the multiple referencing approach development, we will continue working on this problem, and this is our future research direction indeed.

# Chapter 5

## Application to S&P 500 Stock Market

### 5.1 Introduction

The proposed algorithm is validated by using simulated datasets, to show dedicated graphical results competing with other cutting-edge algorithms. Theoretically, individual stock products' node should be represented as closely correlated, if the proposed algorithm is working perfectly, whereas the edge between each stock products indicates the amount of variation in terms of correlation. In this research we report the results from three S&P 500 case studies.

### 5.2 Data Pool

Like most typical stock market analysis represented, using real financial data will draw our attention into actual problem analysis, and would be useful to an economist studying the effect of various indicators on the market. In this case, we form the data pool by entire S&P 500 stocks, for the proposed computational correlation analysis experiments. The S&P 500 has been widely regarded as the best single gauge of the large cap U.S. equities market since the index was first published in 1957. The index has over US\$ 5.58 trillion benchmarked, with index assets comprising approximately US\$ 1.31 trillion of this total. The index includes 500 leading companies in leading industries of the U.S. economy, capturing 75% coverage of U.S. equities. The stock data were downloaded from Yahoo Finance. The data for each day consisted of Stock

---

name, Date, Open price, Close price, Daily highest price, Daily lowest price and the volume of stock traded on that day. Only close prices are used as input, and they are all normalized for efficient calculation. A total of 11 years data have been gathered from 2000 to 2011.

## 5.3 Experiment Setup

The proposed single referencing computational stocks correlated analysis model is implemented in MATLAB R2012a student version 7.14.0.739, on a 1.8GHz Intel Core i5 macintosh with 4GB RAM. For the experiment results, we applied MATLAB and Graphviz Version 2.29 to generate visualization for the superior illustration.

## 5.4 Referencing to S&P 500 Indices Price

### 5.4.1 Datasets Selection

In this correlation case study, proposed algorithm is applied for measuring of long-term associations between pairs of companies traded on the S&P 500 stock markets, in terms of referencing the S&P 500 indices. The companies were selected randomly; as long as they met the requirement of consisting of all consecutive daily close prices from 3rd of January, 2000 to 2nd of December, 2011.

### 5.4.2 Results

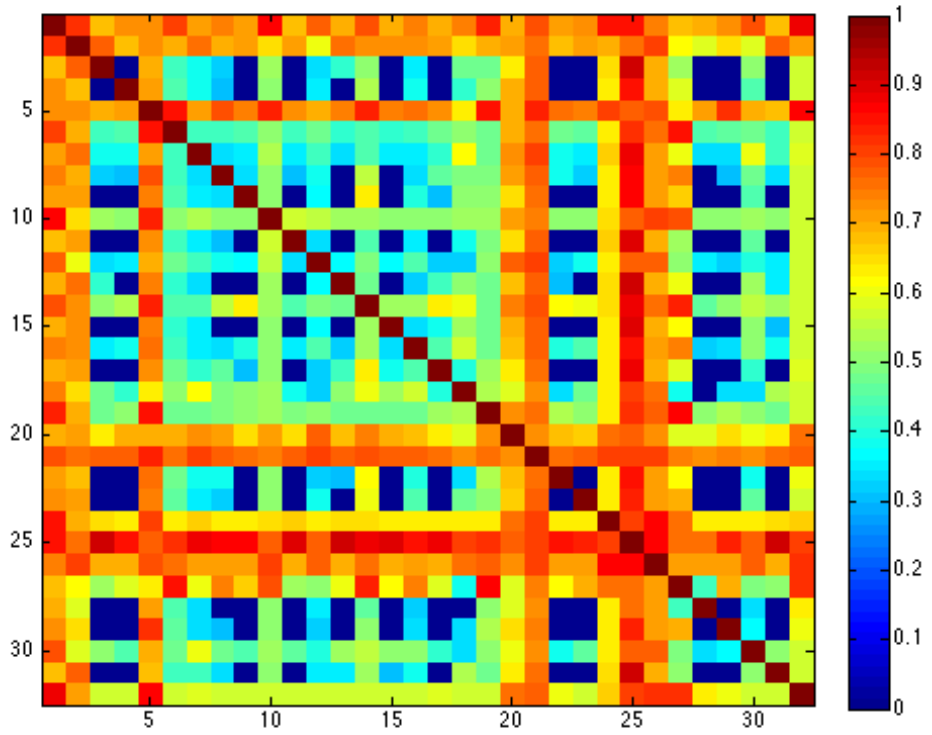
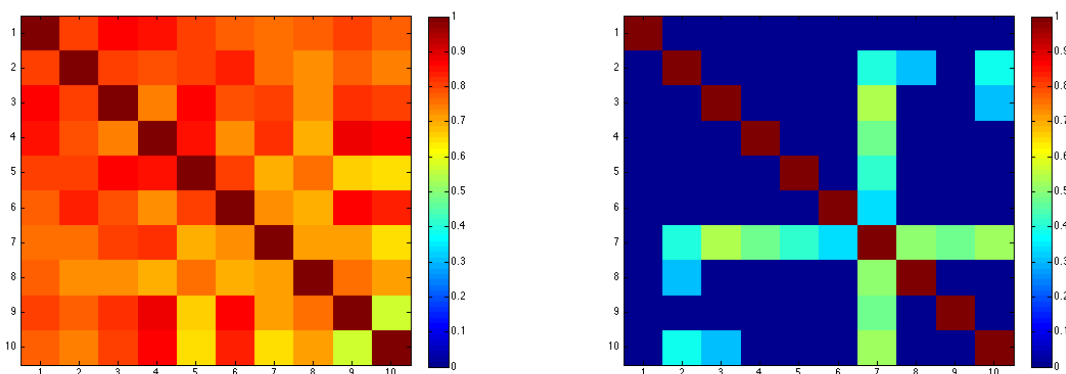


Figure 5.1: Multiple Correlation Coefficient between 32 S&P 500 Stock Products



(a) Most correlated stocks in parameteration of S&P 500 indices      (b) Least correlated stocks in parameteration of S&P 500 indices

Figure 5.2: Comparison of Multiple Correlation Coefficient heat map between sorted and unsorted dataset from S&P 500

Rank	Stock Name	Industry	Correlation Coefficient $\rho$
1	Morgan Stanley (MS)	Financial	0.7560
2	Xerox Corporation (XRX)	Information Technology	0.7289
3	Microsoft Corporation (MSFT)	Information Technology	0.6983
4	Agilent Technologies Inc. (A)	Health Care	0.6807
5	Nabors Industries Ltd. (NBR)	Energy	0.6327
6	The AES Corporation (AES)	Utilities	0.6196
7	Alcoa Inc. (AA)	Materials	0.5830
8	Yahoo! Inc. (YHOO)	Information Technology	0.5695
9	The McGraw-Hill Companies, Inc. (MHP)	Financial	0.5682
10	Citigroup, Inc. (C)	Financial	0.5069

Table 5.1: Most correlated stock products in parameteration of S&P 500 indices

Rank	Stock Name	Industry	Correlation Coefficient $\rho$
1	Mylan, Inc. (MYL)	Health Care	0.0431
2	Yum! Brands, Inc. (YUM)	Consumer Discretionary	0.1377
3	McDonald's Corp. (MCD)	Consumer Discretionary	0.1557
4	Apple Inc. (AAPL)	Information Technology	0.1754
5	Abbott Laboratories (ABT)	Health Care	0.1940
6	3M Company (MMM)	Industrials	0.1964
7	XL Group plc (XL)	Financial	0.2053
8	Apache Corp. (APA)	Energy	0.2267
9	NextEra Energy, Inc. (NEE)	Utilities	0.2355
10	CSX Corp. (CSX)	Industrials	0.2458

Table 5.2: Least correlated stock products in parameteration of S&P 500 indices

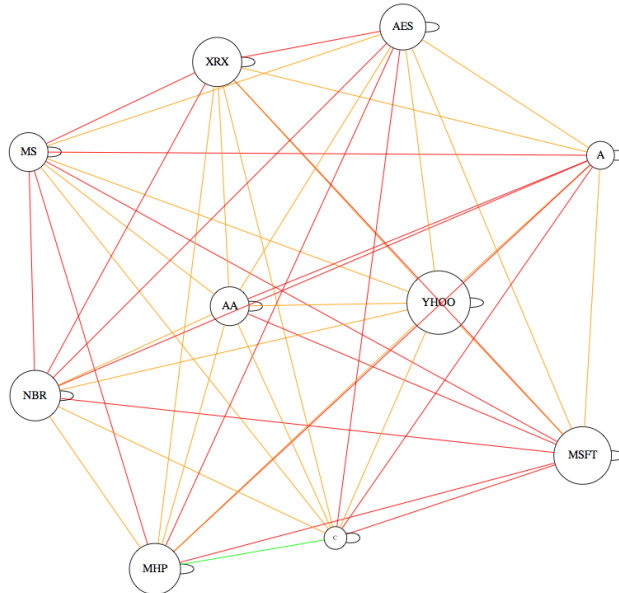


Figure 5.3: Dot graph for most correlated stock products in parameteration of S&P 500 indices

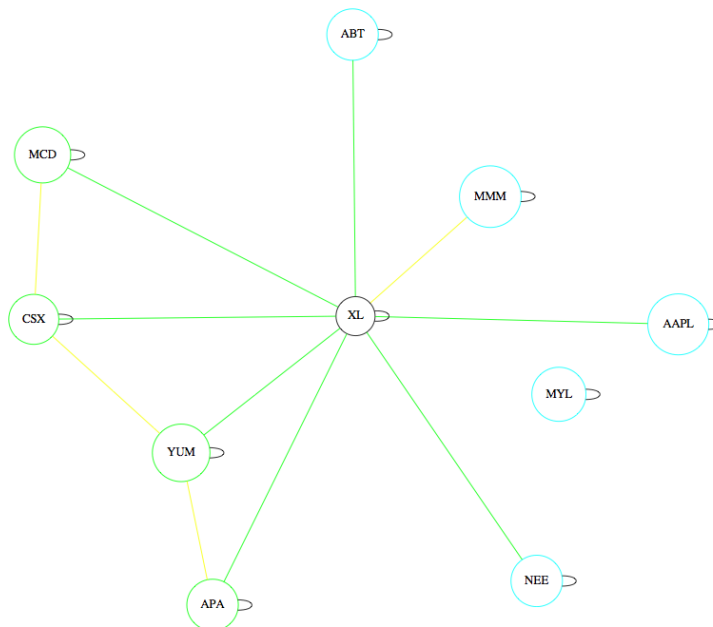


Figure 5.4: Dot graph for least correlated stock products in parameteration of S&P 500 indices

### 5.4.3 Discussion

As shown in Figure 5.1, we have obtained the matrix of empirical multiple correlation coefficients between 32 stock products from the S&P 500 index. The variation of colors indicates the strength associated across each pair of stocks. The strength of association between every pair of stocks varies as indicated by the colour scale. It is obvious, but still hard to interpret. Thus, by selecting 10 stocks from both top and bottom ends, the picture is much clearer for a user to distinguish the correlation between stocks based on index's activity. Figure 5.2 presents both most correlated and uncorrelated 10 stocks with respect to the index.

Also, in Table 5.1 and 5.2, the most likely related 10 stocks and the most unlikely related 10 stocks are both ranked by correlation coefficient  $\rho$  respectively. From the industry column in both Table 5.1 and 5.2, we could say that, since the distribution of the constituents is fairly homogeneous, taxonomy does not play a crucial factor in this case study. It is worth to note the hub stock XL in figure 5.4, even in the least correlated stocks group, it connect a set of diverse stocks that does not closely correlated. According to table 5.2, stock XL is corresponding to financial sector, it possible for such stock to be a good indicators of the status of the market. Such features will not be extracted by only analysing the correlation coefficients as listed in table 5.2.

## 5.5 Referencing to U.S. Crude Oil Price

### 5.5.1 Datasets Selection

In this case study, the proposed algorithm has been applied with a manipulated data set, to demonstrate the practical performance in regard to the fundamental correlation study. The exceptional data set will be selected by two divergent tactics: manipulated selection and random selection.

Manipulated selection is mainly focused on empirical study, which manually categorises the stocks' cluster. In order to ensure the end-result is well distributed, several factors have been taken into consideration, while making the decision of this type of selection.

First of all, stocks will be clustered by their constituents, which are determined by the Global Industry Classification Standard (GICS) (Bhojraj, Lee & Oler, 2003).

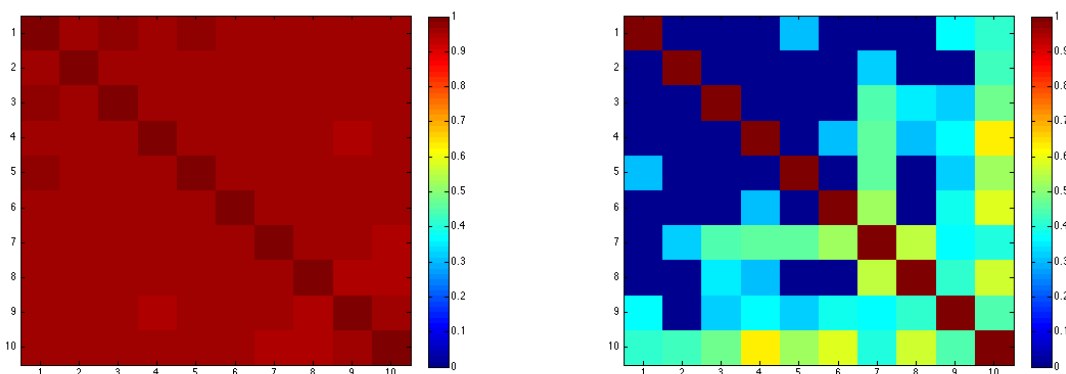
The GICS structure consists of sectors and industries into which S&P has categorized all major public companies.

In addition to the taxonomy investigation, the power of stocks will be ranked by market cap. Stocks with larger market cap indicate a greater contribution to the constituent of market (Yang & Zhu, 2005). We select market's most powered stocks to maximize the accuracy of the designed stocks correlation test.

Unlike the manipulated selection, random selection is mainly focused on a suppositions study. When randomly picking the set of stocks from the data pool, the impartial character will be most likely reflected; beyond that, a sparse outcome will be expected. In this scenario, stocks from the entire market will be numbered and a random number generator will be used for the outcome. While discussing the references, the prices of crude oil and gold are two commonly used indicators for mirroring world financial status in terms of commodities. Another influential benchmark is the stock market self-index. In front of those academic financial understanding, the standard could be set to one of those measures.

By inspecting the GICS sectors, the petroleum sector has been chosen. The significance of the selected sector is determined by its contribution of the market cap to entire indices. Once the sector has been targeted, we will select stocks from same GICS industry group as in the first place, since we believe that the measurement of GICS classification will align with their correlation. In general, the number of candidate stocks is always surplus to the number required; therefore, only top ranked stocks will be considered, whereas the ranking is based on individual stock's market cap. Once stocks have been selected, each one of them will be computed against the index data, a discrete arc can be obtained according to the length of time series for channel approximation. The method applied here is graphical channel correlation analysis. The graphical illustration will obviously identify the pattern similarity between the index and each stock, and verify our selection.

Therefore, 28 petroleum stocks classified by GICS in S&P 500 and 28 randomly selected stocks to against crude oil daily price respectively. The time window for those samples is between 4th of January 2000 and 2nd of December 2011. Data for both tests are sourced from YAHOO finance database.



(a) Petroleum Stock Products

(b) Random Stock Products

Figure 5.5: Comparison of Multiple Correlation Coefficient heat map between Petroleum and Random dataset

### 5.5.2 Results

Rank	Stock Name	Sub-Industry from GICS Energy Sector
1	Apache Corp. (APA)	Oil & Gas Exploration & Production
2	Murphy Oil Corporation (MUR)	Integrated Oil & Gas
3	CONSOL Energy Inc. (CNX)	Oil & Gas Exploration & Production
4	Devon Energy Corporation (DVN)	Oil & Gas Exploration & Production
5	Denbury Resources Inc. (DNR)	Oil & Gas Exploration & Production
6	Hess Corporation (HES)	Integrated Oil & Gas
7	EOG Resources, Inc. (EOG)	Oil & Gas Exploration & Production
8	National Oilwell Varco, Inc. (NOV)	Oil & Gas Equipment & Services
9	ConocoPhillips (COP)	Integrated Oil & Gas
10	Noble Energy, Inc. (NBL)	Oil & Gas Exploration & Production

Table 5.3: Most Correlated Petroleum Stocks with Crude Oil

Rank	Stock Name	GICS Sectors
1	Bank of America Corporation (BAC)	Financials
2	Alcoa Inc. (AA)	Materials
3	Cisco Systems, Inc. (CSCO)	Information Technology
4	Merck & Co. Inc. (MRK)	Health Care
5	General Electric Company (GE)	Industrials
6	The Home Depot, Inc. (HD)	Consumer Discretionary
7	Microsoft Corporation (MSFT)	Information Technology
8	Intel Corporation (INTC)	Information Technology
9	Wal-Mart Stores Inc. (WMT)	Consumer Staples
10	Verizon Communications Inc. (VZ)	Telecommunications Services

Table 5.4: Least Correlated Random Stocks with Crude Oil

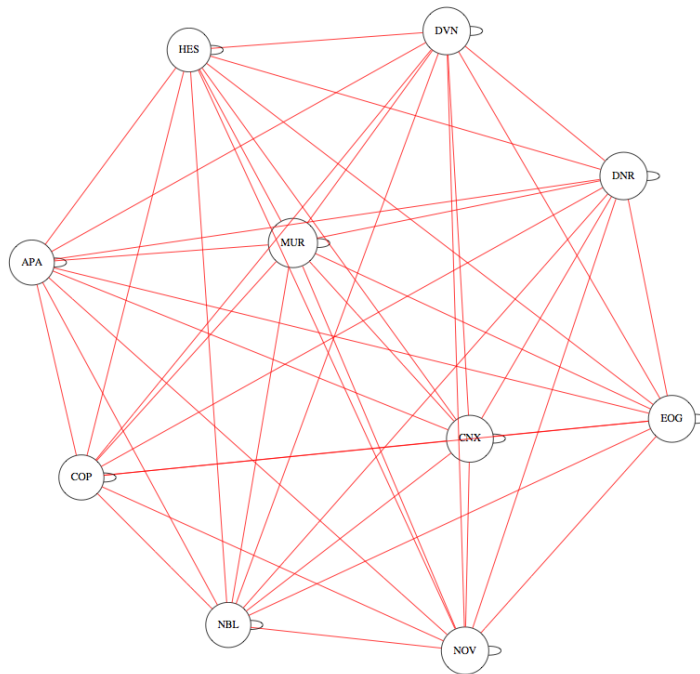


Figure 5.6: Dot Graph for Petroleum Stock Products

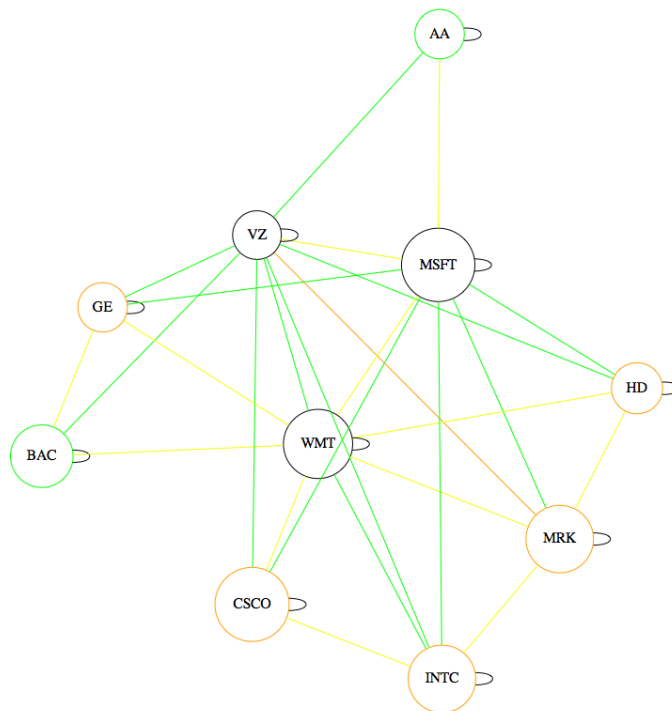


Figure 5.7: Dot Graph for Random Stock Products

### 5.5.3 Discussion

Figure 5.5 is the results obtained from stocks correlation analysis with respect to U.S. crude oil price. Since we used U.S. crude oil as the reference, the correlations between petroleum stock products are naturally all very strong, and randomly selected stocks are diversely correlated. Such outcomes could provide evidence to prove our concept on fundamental stock correlation, in which similar industry shares have similar characteristics in volatility.

Figure 5.6 and 5.7 are alternative illustrations of results acquired from the same stocks correlation analysis. Circled nodes indicate stock products, and edges between nodes represent the corresponding correlation and the colour of the edge defines correlation strength, while a darker colour means stronger correlation. From the result, we could conclude that there is strong and numerous cross-correlation between petroleum stocks. On the other hand, only fewer and weaker cross-correlations exist within the group of random picked stocks. By all means, when we analyzing stocks correlation with the consideration of U.S. crude oil, only energy related aspect will be examined at the first place. Regardless of two stocks' operational nature, as long as they are having similar features in petroleum industry, they are classified as highly correlated pair.

## 5.6 Comparison between U.S. Crude Oil price and S&P 500 Indices price

### 5.6.1 Datasets Selection

In this case study, the proposed algorithm has been applied to two scenarios. The first one is to compute the stocks correlation between all the stocks U.S. Crude Oil prices. Stocks included in this scenario are 28 petroleum stocks and 28 random stocks that were used in case study one, plus 32 random stocks that were used in case study two. The second scenario is to compute stocks correlation between all the above stocks and S&P 500 indices price. The time window for those data is between 4th of January 2000 and 2nd of December 2011 and it has been sourced from YAHOO finance database.

### 5.6.2 Results

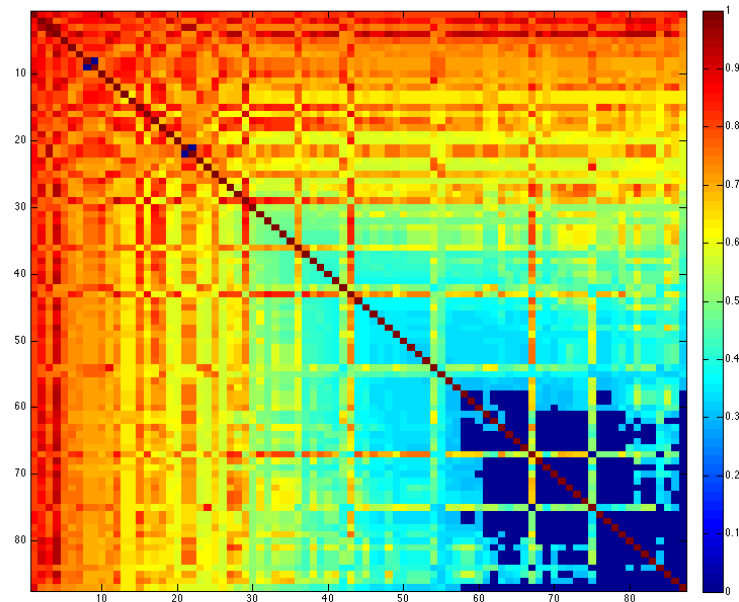


Figure 5.8: Stock Products vs S&P 500 Indices Price

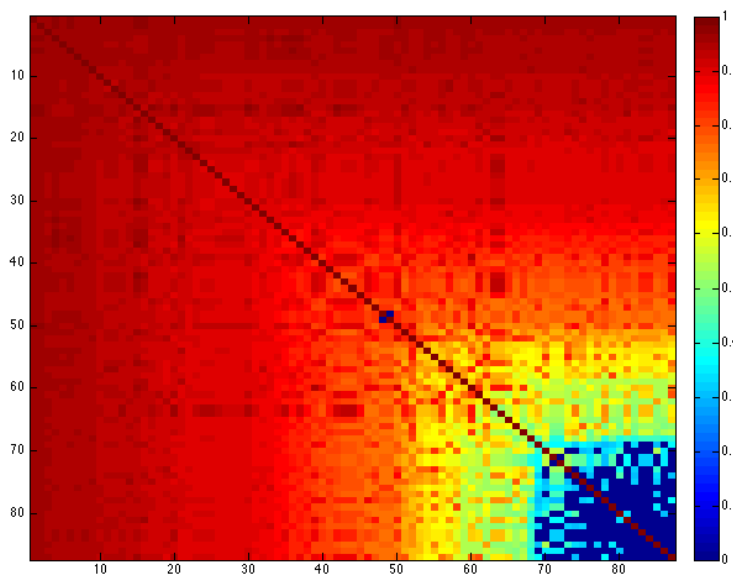


Figure 5.9: Stock Products vs Crude Oil Price

### 5.6.3 Discussion

As shown in Figure 5.8 and 5.9, the visualizations of the stocks correlation are partially different, even though the input data were identical. This outcome demonstrates our ground idea of the proposed referencing approach: the measurement of stocks correlation will be different with respect to various referencing factors. E.g. by referencing to U.S. crude oil price, the stocks correlation will be measured in terms of petroleum activities rather than simple price volatility. On the other hand, when we use indices price as the reference, stocks correlation will be determined by the similarity in the price movement during the time frame. The interpretation of such a finding is that investors will be able to advantage their investment intuition by choosing the appropriate referencing factors.

## Chapter 6

# Conclusion and Future Works

The stock market is always described as being a dynamic and complex financial system to analyse. Accurately forecasting a stocks future price movement has been classified as an unfeasible task for the non-professional investor. There are countless studies to discover the most appropriate methodology to extract correlations within stock market correctly, but most of them mainly focus on non-parametric stocks correlation analysis, without consideration of any referencing factors as the parameter. In our research, we proposed a hybrid method on the grounds of technical correlation analysis with respect to traditional fundamental correlation analysis. The merit of this research is that we develop a numerical referencing correlation analysis model for extracting and representing the most accurate and intelligible interrelationships between stock products in parameterization of a pre-defined referencing factor. The proposed computational stocks correlation analysis model is trained over the historical data of stock price and referencing factor, and the generated stocks correlation indicates the future movements of a stock or a group of stocks. In the experiment, the proposed algorithm accurately measured the sample S&P 500 stocks correlations, with the consideration of two referencing factors respectively (U.S. Crude Oil price, S&P 500 indices price). However, there are limitations of subjectivity and drawbacks of the arbitrary nature for current referencing factor selection. Even though it is an unavoidable problem in this area of research, this will leave us a future research direction: minimizing such disruption to ensure the outcome is utterly impartial by integrating multiple referencing factors.

## References

- Arouri, M. E. H., Jouini, J. & Nguyen, D. K. (2012). On the impacts of oil price fluctuations on european equity markets: Volatility spillover and hedging effectiveness. *Energy Economics*, 34(2), 611–617.
- Beat Grunder, C. S., Michael Macdonald & Zingg, P. (2012). *Technical analysis - explained* (Tech. Rep.). Credit Suisse Credit Suisse (Guernsey) Limited.
- Becker, K. G., Finnerty, J. E. & Friedman, J. (1995). Economic news and equity market linkages between the us and uk. *Journal of Banking & Finance*, 19(7), 1191–1210.
- Bernanke, B. S. & Mihov, I. (1998). Measuring monetary policy. *The Quarterly Journal of Economics*, 113(3), 869–902.
- Bhojraj, S., Lee, C. & Oler, D. K. (2003). What's my line? a comparison of industry classification schemes for capital market research. *Journal of Accounting Research*, 41(5), 745–774.
- Bilson, C. M., Brailsford, T. J. & Hooper, V. J. (2001). Selecting macroeconomic variables as explanatory factors of emerging stock market returns. *Pacific-Basin Finance Journal*, 9(4), 401–426.
- Bollerslev, T., Cai, J. & Song, F. M. (2000). Intraday periodicity, long memory volatility, and macroeconomic announcement effects in the us treasury bond market. *Journal of Empirical Finance*, 7(1), 37–55.
- Brown, S. J. & Goetzmann, W. N. (1997). Mutual fund styles. *Journal of financial Economics*, 43(3), 373–399.
- Chan, L. K., Lakonishok, J. & Swaminathan, B. (2007). Industry classifications and return comovement. *Financial Analysts Journal*, 56–70.
- Ciora, C., Munteanu, S. M., Hrinca, I. G. & Ciobanu, R. (2011). Uncertainty and fluctuations on the stock markets. In *International conference on financial management and economics* (Vol. 11). Singapore: IACSIT Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge Academic.
- Connolly, R. A. & Wang, F. A. (2003). International equity market comovements: Economic fundamentals or contagion? *Pacific-Basin Finance Journal*, 11(1), 23–43.
- Cory Janssen, C. L. & Murphy, C. (2013, 04). Retrieved from <http://www.investopedia.com/university/technical/>

- Cover, J. P. (1992). Asymmetric effects of positive and negative money-supply shocks. *The Quarterly Journal of Economics*, 107(4), 1261–1282.
- Creti, A., Joëts, M. & Mignon, V. (2013). On the links between stock and commodity markets' volatility. *Energy Economics*.
- Edwards, R. D., Magee, J. & Bassetti, W. (2007). *Technical analysis of stock trends*. CRC Press.
- Elton, E. J. & Gruber, M. J. (1970). Homogeneous groups and the testing of economic hypotheses. *Journal of Financial and Quantitative Analysis*, 581–602.
- Emch, G. G. (1972). *Algebraic methods in statistical mechanics and quantum field theory*. Wiley New York.
- Fama, E. F. (1991). Efficient capital markets: Ii. *The journal of finance*, 46(5), 1575–1617.
- Farrell, J. L. (1974). Analyzing covariation of returns to determine homogeneous stock groupings. *The Journal of Business*, 47(2), 186–207.
- Fiol-Roig, G., Miró-Julià, M. & Isern-Deyà, A. (2010). Applying data mining techniques to stock market analysis. In Y. Demazeau et al. (Eds.), *Trends in practical applications of agents and multiagent systems* (Vol. 71, p. 519-527). Springer Berlin Heidelberg. Retrieved from [http://dx.doi.org/10.1007/978-3-642-12433-4\\_61](http://dx.doi.org/10.1007/978-3-642-12433-4_61) doi: 10.1007/978-3-642-12433-4\_61
- Garcia, R. & Schaller, H. (2002). Are the effects of monetary policy asymmetric? *Economic Inquiry*, 40(1), 102–119.
- Graham, M., Nikkinen, J. & Sahlström, P. (2003). Relative importance of scheduled macroeconomic news for stock market investors. *Journal of Economics and Finance*, 27(2), 153–165.
- Hegland, M. (2001). Data mining techniques. *Acta Numerica 2001*, 10(1), 313–355.
- Hess, M. K. (2004). Dynamic and asymmetric impacts of macroeconomic fundamentals on an integrated stock market. *Journal of International Financial Markets, Institutions and Money*, 14(5), 455–471.
- Hong, H., Lim, T. & Stein, J. C. (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *The Journal of Finance*, 55(1), 265–295.
- Hong, H. & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance*, 54(6), 2143–2184.

- Huang, L. (2011). Moving average effect of the stock market in china: An analysis based on the textile section. In *E -business and e -government (icee), 2011 international conference on* (p. 1-3). doi: 10.1109/ICEBEG.2011.5882170
- Kim, O. & Verrecchia, R. E. (1991). Trading volume and price reactions to public announcements. *Journal of Accounting Research*, 29(2), 302–321.
- Kim, O. & Verrecchia, R. E. (1994). Market liquidity and volume around earnings announcements. *Journal of Accounting and Economics*, 17(1), 41–67.
- Kim, S.-J. (2003). The spillover effects of us and japanese public information news in advanced asia-pacific stock markets. *Pacific-Basin Finance Journal*, 11(5), 611–630.
- Kolar, M., Parikh, A. P. & Xing, E. P. (2010). On sparse nonparametric conditional covariance selection. In *Proceedings of the 27th international conference on machine learning*.
- Kondratenko, V. & Kuperin, Y. A. (2003). Using recurrent neural networks to forecasting of forex. *arXiv preprint cond-mat/0304469*.
- Kumar, M. & Kalia, A. (2012). Preprocessing and symbolic representation of stock data. In *Advanced computing communication technologies (acct), 2012 second international conference on* (p. 83-88). doi: 10.1109/ACCT.2012.89
- Kwapien, J., Gworek, S. & Drozd, S. (2009). Structure and evolution of the foreign exchange networks. *arXiv preprint arXiv:0901.4793*.
- Lauritzen, S. L. (1996). *Graphical models* (Vol. 17). Oxford University Press.
- Lei, S. (2010). *Informative correlation extraction from and for forex market analysis* (Unpublished doctoral dissertation). AUT University.
- Leigh, W., Purvis, R. & Ragusa, J. M. (2002). Forecasting the nyse composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: a case study in romantic decision support. *Decision Support Systems*, 32(4), 361–377.
- Liu, R. & Zargham, M. (2006). *Relationship between fundamental values in the financial market* (Unpublished doctoral dissertation). Southern Illinois University Carbondale.
- Maghyereh, A. & Al-Kandari, A. (2007). Oil prices and stock markets in gcc countries: new evidence from nonlinear cointegration analysis. *Managerial Finance*, 33(7), 449–460.
- Malik, F. & Ewing, B. T. (2009). Volatility transmission between oil prices and equity

- sector returns. *International Review of Financial Analysis*, 18(3), 95–100.
- Malik, F. & Hammoudeh, S. (2007). Shock and volatility transmission in the oil, us and gulf equity markets. *International Review of Economics & Finance*, 16(3), 357–368.
- Matia, K., Amaral, L. A. N., Goodwin, S. P. & Stanley, H. E. (2002). Different scaling behaviors of commodity spot and future prices. *Physical Review E*, 66(4), 045103.
- Matia, K., Ashkenazy, Y. & Stanley, H. E. (2003). Multifractal properties of price fluctuations of stocks and commodities. *EPL (Europhysics Letters)*, 61(3), 422.
- Mensi, W., Beljid, M., Boubaker, A. & Managi, S. (2013). Correlations and volatility spillovers across commodity and stock markets: Linking energies, food, and gold. *Economic Modelling*, 32, 15–22.
- Nikkinen, J., Omran, M., Sahlström, P. & Äijö, J. (2006). Global stock market reactions to scheduled us macroeconomic news announcements. *Global Finance Journal*, 17(1), 92–104.
- Nikkinen, J. & Sahlström, P. (2004). Scheduled domestic and us macroeconomic news and stock valuation in europe. *Journal of multinational financial management*, 14(3), 201–215.
- Nofsinger, J. R. & Prucyk, B. (2003). Option volume and volatility response to scheduled economic news releases. *Journal of Futures Markets*, 23(4), 315–345.
- Pang, S., Song, L. & Kasabov, N. (2011). Correlation-aided support vector regression for forex time series prediction. *Neural Computing and Applications*, 20(8), 1193–1203.
- Park, J. & Ratti, R. A. (2008). Oil price shocks and stock markets in the us and 13 european countries. *Energy Economics*, 30(5), 2587–2608.
- Patelis, A. D. (1997). Stock return predictability and the role of monetary policy. *the Journal of Finance*, 52(5), 1951–1972.
- Pearson, K. (1895). Contributions to the mathematical theory of evolution. ii. skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London. A*, 186, 343–414.
- Raghubir, P. & Das, S. R. (1999). A case for theory-driven experimental enquiry. *Financial Analysts Journal*, 56–79.

- Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution*, 43(1), 223–225.
- Strongin, S. (1995). The identification of monetary policy disturbances explaining the liquidity puzzle. *Journal of Monetary Economics*, 35(3), 463–497.
- Thorbecke, W. (1997). On stock market returns and monetary policy. *The Journal of Finance*, 52(2), 635–654.
- Ting, J., Fu, T.-c. & Chung, F.-l. (2006). Mining of stock data: intra-and inter-stock pattern associative classification. *Threshold*, 5(100), 5–99.
- Washio, T., Shinnou, Y., Yada, K., Motoda, H. & Okada, T. (2007). Analysis on a relation between enterprise profit and financial state by using data mining techniques. *New Frontiers in Artificial Intelligence*, 305–316.
- Yang, J. W., Jian; Kolari & Zhu, G. (2005). European public real estate market integration. *Applied Financial Economics*, 15(13), 895–905.
- Zargham, M. R. & Sayeh, M. R. (1999). A web-based information system for stock selection and evaluation. In *Advance issues of e-commerce and web-based information systems, wecwis, 1999. international conference on* (pp. 81–83).
- Zhao, E. & Han, Z. (2010). Analyze long & mid-term trends of stock with genetic programming on moving average and turning points. In *Advanced computer control (icacc), 2010 2nd international conference on* (Vol. 3, pp. 87–91).