Obtaining and Using A-Correlation Information on Stocks via the Internet

Liqian GAO

School of Computer Science

McGill University, Montreal

May, 2001

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements of the degree of Master of Science

To

Yang

and

My Parents

Acknowledgments

I would like to express my sincere appreciation to my supervisor, Dr. Monty Newborn for his invaluable guidance, his encouragement, his care and support throughout the course of this thesis work.

I would also express my thanks to Lingyang Zhang and Fang Liao for their helps and friendship.

A special thank to Sam-Sang Suy and Yanjun Liu for translating the abstract into French.

Thanks are also due to Xing Yu, Yun Hu for their friendship and encouragement.

Finally, I would like to express a special acknowledgment to my husband Yang Wu, my mother Yumei Liu, my father Jinyue Gao and my brother Dawei Gao for their constant support, understanding, and making all of this possible.

Abstract

This thesis introduces two stock analysis software packages. The software packages are designed to gather stock data over the Internet, and process the gathered data to determine an approximate correlation (A-correlation) between a reference stock and a set of comparison stocks.

In this thesis, the A-correlation is obtained from historical stock data using data mining techniques. The data mining techniques include a data retrieval and analysis algorithm. First, a large pool of stock data is retrieved from the CNBC web site by employing an Internet data retrieval algorithm. Then an analysis algorithm is applied to determine the A-correlation between a reference stock and each of a set of the comparison stocks. The analysis algorithm is based on the accumulation of individual stock movements (up/down) over a period of time.

We have designed and implemented the two programs using Microsoft Excel/VBA. The two programs share common functionalities such as providing the A-correlation as a function of time, ranking the A-correlation results in a predefined order, and allowing the user to spot possible A-correlation trends. The difference between the two programs is that the first program analyses A-correlation with no time delay and the second program

deals with the time delayed A-correlation. Each program can be useful in revealing different aspects of the relationships between stocks. These two programs can be useful tools for stock market analysis.

Résumé

Cette thèse présente deux progiciels d'analyse boursière. Les progiciels sont conçus pour recueillir les données boursières à travers l'Internet, et traitent les données recueillies pour déterminer une corrélation approximative (A-corrélation) entre des références de donnée boursière courante et un ensemble de comparaison de donnée boursière.

En cette thèse, la A-corrélation est obtenue à partir d'un historique des données courantes en utilisant des techniques d'extraction de données. Les techniques d'extraction de données incluent un algorithme d'extraction et d'analyse de données. D'abord, un grand regroupement des données courantes est recherché à partir du site Internet de CNBC en utilisant un algorithme d'extraction de données dans le réseau l'Internet. Ensuite, un algorithme d'analyse est appliqué pour déterminer la A-corrélation entre des références de donnée boursière courante et chacune un ensemble de comparaison de donnée boursière. L'algorithme d'analyse est basé sur l'accumulation des différents mouvements boursiers (haut/bas) sur une certaine période de temps.

Nous avons conçu et avons mis en application les programmes en utilisant Microsoft Excel/VBA. Les deux programmes partagent des fonctionnalités communes telles que fournir la A-corrélation en fonction du temps, rangeant les résultats de A-corrélation dans une période prédéfinie, et permettre à l'utilisateur de repérer la tendance possible de la A-corrélation. La différence entre les deux programmes est que le premier programme analyse la A-corrélation sans le délai et le deuxième programme traite la A-corrélation à un certain délai. Chaque programme peut être très utile pour indiquer des différents aspects de fluctuation de la bourse. Ces deux programmes d'utilitaires peuvent être utiles pour analyser le marché boursier.

Abbreviation

What typed	Replacement Text
Approximate correlation	A-correlation

Table 1.1 Abbreviation overview

Contents

Acknowledgments i

Abstract ii

Résumé iv

Abbreviation vi

1 Introduction 1

1.1 Overview 1

1.2 Motivation 2

1.3 Background 4

1.3.1 Stock Market 4

1.3.2 Data Mining and Stock A-Correlation 8

1.4 Thesis Outline 8

2 A-correlation Between Two Stocks 10

2.1 Introduction 10

2.1.1 Data Mining and Statistical Correlation 10

2.1.2 Degree of Correlation 11

2.1.3 Data Mining and Existing Knowledge 12

2.1.4 Data Mining Requirements 12

2.1.5 A-correlation Model 15

2.2 Implementation Issues 16

2.2.1 A Brief Introduction to Excel 2000/VBA 17

2.2.1.1 Using Excel 17

2.2.1.2 Visual Basic for Applications (VBA) 20
2.2.1.3 Summary 22
2.2.2 Implementation Using Excel/VBA 22
2.2.2.1 Correlation Analyzer (Program I) 23
2.2.2.1.1 Data Retrieval Via Internet (Block I) 24
2.2.2.1.2 Most A-correlated Stocks (Block II) 28
2.2.2.1.3 A-correlation Converging Trend (Block III) ... 33
2.2.2.1.4 A-correlation Diverging Trend (Block IV) 35
2.2.2.2 Time Delayed Correlation Analyzer (Program II) 36
2.3 Summary 38

3 Stock A-correlation Sample Output 39

3.1 Introduction 39
3.2 Sample Output 39
3.2.1 Sample Output from Correlation Analyzer (Program I) 39
3.2.1.1 Most A-correlated Stocks 40
3.2.1.2 A-correlation Converging Trend 43
3.2.1.3 A-correlation Diverging Trend 46
3.2.2 Sample Output from Time Delayed Correlation Analyzer (Program II) 47
3.2.2.1 One-day Delayed A-correlation 47
3.2.2.2 Twenty-days Delayed A-correlation 49
3.3 Summary 51

4 User Manual 52

4.1 Correlation Analyzer (Program I) 52
4.1.1 Open Microsoft Excel Application 52
4.1.2 Running Correlation Analyzer 54
4.1.3 Viewing the Results 57
4.1.4 Analyzing Another Stock 59
4.1.5 Closing Application 60

4.2 Time Delayed Correlation Analyzer (Program II) 61
4.2.1 Open Microsoft Excel Application 61
4.2.2 Running Time Delayed Correlation Analyzer 62
4.2.3 Viewing the Results 63
4.2.4 Analyzing Another Stock 64
4.2.5 Closing the Application 64

5 Summary 65

6 Bibliography 68

List of Tables

Table 1.1 Abbreviation overview.....vi

Table 2.1 Examples of objects used in the programs.....19

Table 2.2 Example of stock information in Sheets "StockData".....28

Table 3.1 Stocks overview.....40

Table 3.2 Stocks overview (Time Delayed Correlation Analyzer).....47

List of Figures

Figure 1.1 Conceptual diagram of A-correlation programs 3

Figure 2.1 An example of web query instructions 18

Figure 2.2 Correlation Analyzer program flow chart 23

Figure 2.3 Data retrieval flow chart (Block I) 25

Figure 2.4 Flow chart of function GetMovingDirection 29

Figure 2.5 Flow chart of function IfMovingSameDirection 30

Figure 2.6 Flow chart of function DataSum 32

Figure 2.7 Flow chart for obtaining A-correlation converging trend 34

Figure 2.8 Flow chart of function IfMovingSameDirection 37

Figure 3.1 Result of most A-correlated stock to the reference stock 41

Figure 3.2 Ranking result of most A-correlated stock to the reference stock 43

Figure 3.3 Result of A-correlation converging trend 44

Figure 3.4 Chart of A-correlation converging trend 45

Figure 3.5 A-correlation diverging trend 46

Figure 3.6 Result of one-day delayed A-correlation 48

Figure 3.7 Rank result of one-day delayed A-correlation 49

Figure 3.8 Result of twenty-days delayed A-correlation 50

Figure 3.9 Rank result of twenty-days delayed A-correlation 51

Figure 4.1 Open Excel application 52

Figure 4.2 Open Correlation Analyzer program 53

Figure 4.3 Choose Correlation Analyzer program 53
Figure 4.4 Enable Macros in opening Correlation Analyzer program 54
Figure 4.5 Initiate Macro command 54
Figure 4.6 Execute Macro 55
Figure 4.7 Input stock symbols 55
Figure 4.8 Stock symbols entered by user 56
Figure 4.9 Continuing analysis dialog 56
Figure 4.10 A-correlation result 57
Figure 4.11 Ranking A-correlation result 57
Figure 4.12 A-correlation converging trend result 58
Figure 4.13 Chart of A-correlation converging trend 58
Figure 4.14 A-correlation diverging trend 59
Figure 4.15 Continue to analyse stock of IBM 59
Figure 4.16 A-correlation result, IBM is the selected reference stock 60
Figure 4.17 Close Correlation Analyzer program 60
Figure 4.18 Saving results 60
Figure 4.19 Choose Time Delayed Correlation Analyzer 61
Figure 4.20 Enter number of days delay 62
Figure 4.21 Result of time delayed up/down price A-correlation 63
Figure 4.22 Rank result of time delayed up/down price A-correlation 63

Chapter 1

Introduction

1.1 Overview

In this thesis, two stock analysis software programs have been designed and implemented in Microsoft Excel/VBA environment. Both stock analysis programs are intended to provide an approximate correlation (A-correlation) between a reference stock and each of a set of comparison stocks. The two programs share common functionalities such as providing the A-correlation as a function of time, ranking the A-correlation results in a predefined order, and allowing the user to spot possible A-correlation trends. The difference between the two programs is that one of them analyzes the time delayed A-correlation (the second program), and the other deals with no time delayed A-correlation (the first program). Each program can be useful in revealing different aspects of the relationships between stocks.

In Chapter Two, the theory behind the A-correlation is explained. Emphasis is placed on demonstrating how the A-correlation and the time delayed A-correlation is obtained through the use of data mining techniques. The data mining techniques employed in this thesis include a data retrieval algorithm and an analysis algorithm. The data retrieval

algorithm is designed to format proper Internet data request queries and gather stock data. The analysis algorithm acts on the gathered stock data and calculates A-correlation between the stocks. The algorithms are routinely used throughout the programs to provide accurate and reliable information on the stock A-correlation.

A user manual is also provided as a part of a complete stock analysis package.

1.2 Motivation

Money and personal wealth have always been driving forces in our society, and will remain to be so for the foreseeable future. These activities include providing labour work, engaging commercial activities, and careful investment. Investment is a relatively new activity, as compared to the former two. Since the inception of industrial revolution, investment activity has flourished throughout the entire world. There are number of means by which investments are made and carried out, with the most popular and essential form being through stock securities. This thesis is dedicated to demonstrating how modern technologies such as the Internet, data mining, and statistics can be utilized to facilitate the stock investment activity.

To understand why modern technologies are required to help in stock investing, one has to realize the complexity and unpredictability of the stock market. Ideally, stock prices obey basic supply-demand principles [Jones 97]. The demand for a stock is associated with its return on investment. However, in the real world, many of the purchasing and selling decisions of a particular stock are based on investors' own assessments. With

millions of investors presenting their own assessment, the day-to-day stock market can be described by one word: chaotic. To increase one's return on investment, one must use all the help they can obtain.

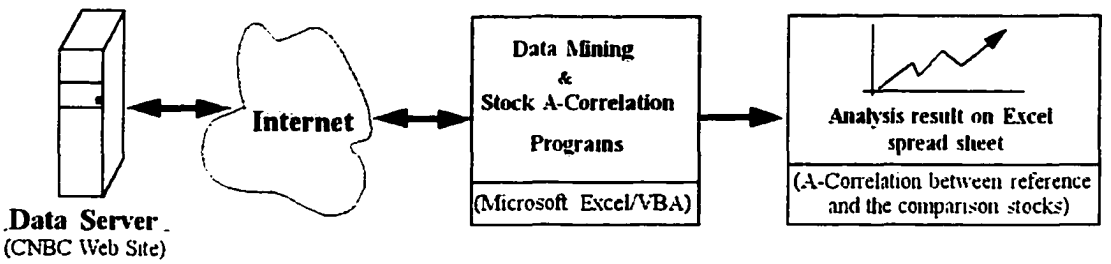


Figure 1.1 Conceptual diagram of A-correlation programs

In this thesis, we devise a method that will help investors to spot patterns and relationships among stocks. This method, which uses data mining, is implemented using Internet and Microsoft Excel/VBA programs, as shown in Figure 1.1. The implemented method retrieves stock data (prices or volume) over the Internet, analyses the data, and determines the A-correlation between a reference stock and each of the comparison stocks. The A-correlation can be very beneficial for investors as it provides clues on the relationship in price (or volume) movement between stocks. The sample outputs from the execution of the implemented programs are presented to demonstrate the usefulness of the method.

1.3 Background

Before delving into the programs that are presented in this thesis, a brief introduction into the history and basic operation of stock market is necessary. In addition, data mining and statistics knowledge, which is the essential part of the thesis. is also briefly presented.

1.3.1 Stock Market

The modern North American stock market was initially founded in 18th century in Manhattan in New York City [Mcma 96]. It was designed to help industrialists and manufacturers in raising capital without incurring debt. Since then, stock markets have spread to all corners of the world. Each day, billions of dollars change hands, from investors looking for profits, to companies that desperately seek cash to fund their operations. In the stock market, a company is referred by a ticker name, which typically consists of no more than four letters. In each trading day, the opening price is the first trading price when stock market opens. The closing price refers to the last trading price when the stock market closes. Volume refers to number of shares changing hands each trading day.

Since the founding of the stock market, many forms of analysis have been introduced to help in valuing investment activities. In addition to conducting fundamental analyses [Bell 63][Bran 98][Lorie 73] such as obtaining the overall economical situation and company performance, major institutional investors and floor traders also developed technical analyses [Schw 99][Wein 96][Crow 77], which are handy in the actual trading

processes. Next, we shall look into both the fundamental analysis and technical analysis in detail.

Fundamental analysis is the standard old-fashioned techniques of looking the horse in the mouth—inquiring into such basic matters as interest rates, dividends, earnings prospects, and the many influences in economics and politics which might have a bearing on the stock market. Fundamental analysis deals with the indispensable parts of investment process. Typically, they offer insight into the long-term performance of a stock. [Bell 63]

Fundamental analysis is performed preliminary to the technical analysis. In an ideal world, one should perform the fundamental analyses on a desired stock before proceeding to conduct technical analyses.

Technical analysis is primarily concerned with measuring psychology, the character and quality of the buying and selling in the market, the demand for and supply of securities, the momentum and trends of price movements. For a floor trader, technical analysis offers insight into the short-medium term profitability of stocks [Bell 63].

Technical analysis consists of a series of measurements and studies on stocks. The measurements and studies include the moving average [Schw 99], the volatility factor [Schw 99], the momentum [Schw 99], and so on. These measurements and studies are popular, used frequently by floor traders and investors.

The moving average is probably the best-known and most versatile indicator in the analyst's tool chest. It can be used with the analyst's price of choice (highs, closing prices, etc.) and can also be applied to other indicators, helping to smooth out volatility. As the name implies, the moving average is the average of a given amount of data over a specified time period. For example, a 14-day moving average of closing prices is calculated by adding the last 14 closing prices and dividing by 14. The result is noted on a chart. The next day the same calculations are performed with the new result being connected (using a solid or dotted line) to that of the previous day's. And so forth. Variations of the basic moving average are the Weighted and Exponential moving averages. [Schw 99][Wein 96]

Volatility analysis is based on the idea that stocks bottom from "panic" selling, after which a rebound is imminent. One way of measuring this phenomenon is to observe a widening range between the high and low prices each day. In general, a progressively wider range, observed over a relatively short period of time, can indicate that a bottom is near. Price ceilings are generally reached at a more leisurely pace and can be characterized by a narrowing of the price range. This measure of the trading range takes place over a specified period in order to determine whether or not an issue is being "dumped" and is approaching a bottom. A pre-requisite to a valid bottom is an increase in the volatility line above the reference line. In a similar manner, an indication of an imminent top would be a decrease in the volatility line below the reference line. As long as volatility is rising, in all probability a stock will not approach a top. It should be noted that this study should be

used in conjunction with trend following analysis and momentum oscillators for confirmation and accuracy [Schw 99][Wein 96].

Momentum provides an analysis of changes in prices (as opposed to changes in price levels). Changes in the rate of ascent or descent are plotted. The momentum line is graphed positive or negative to a straight line representing time. The position of the time line is determined by the price at the beginning of the momentum period. Traders use this analysis to determine overbought and oversold conditions. When a maximum positive point is reached, the market is said to be overbought and a downward reaction is imminent. When a maximum negative point is reached, the market is said to be oversold and an upward reaction is indicated [Schw 99][Wein 96].

In this thesis, we describe two programs that determine the A-correlation between a reference stock and each of a set of comparison stocks. It is implemented using Microsoft Excel/VBA. The user enters a reference stock and a set of comparison stocks. Then the programs retrieve the stocks' prices or volumes over the last two years via the Internet. Upon receiving the data, the program determines the stocks' direction of movement (up/down) for each trading day, and determines whether or not each of the comparison stocks moved in the same direction as the reference stock. Finally, the programs count the number of days that each of the comparison stock moved in the same direction (up/down) as the reference stock in last X days, where X is a power of 2 ranging from 0 to 9. The higher the calculated number, the more A-correlated the comparison stock is said to be with the reference stock.

1.3.2 Data Mining and Stock A-Correlation

A wealth of information exists on the Internet. The information ranges from news, food, arts, scientific reports, to financial information. The information can be retrieved easily. However, due to the amount of information and lack of standard, it is difficult to find exactly what one is looking for. To answer the need of finding out useful information quickly and efficiently, data mining has been developed to suit this need.

In recent years, the Internet has become quite prominent. Data mining is becoming increasingly popular as well. Many papers and books [Asog 94][Azhar 94][Cios 98][Diel 80][Gime 99][Golan 95][Iba 2000][Jang 91][Mats 91][Wang 96][Wuth 98] have been written on the subject. In the case of stock market technical analysis, one is faced with the task of sorting through tons of stock data looking for relevant information. Data mining is designed to overcome this obstacle. In essence, A-correlation analysis, as previously introduced, can be considered as one of the applications of data mining. The details of data mining and statistical methods are discussed in greater detail in Chapter Two.

1.4 Thesis Outline

This thesis is divided into five chapters. The first chapter introduces the motivation behind the proposed programs. Other stock analyses are also presented. The second chapter, the core of the entire thesis, deals mainly with the design and implementation of the two programs. Chapter Two starts with an introduction to the theory and model behind the A-correlation. It then explains how data mining and Microsoft Excel/VBA can be combined to achieve efficiency in obtaining the A-correlation. In Chapter Two, the data

structure, program flow, and each sub-function used in the two programs are documented. In Chapter Three, sample outputs from executing the two programs are presented. The sample outputs are tabulated, graphed, and interpreted to show how they can be useful to investors and traders. In order to facilitate the future improvement of the programs, a well-documented user manual is presented in Chapter Four. The manual provides step-by-step instructions on how to use the two programs, which are named Correlation Analyzer and Time Delayed Correlation Analyzer. Each instruction comes with a diagram indicating how the instruction should be carried out. Finally, in the last chapter, Chapter Five, a summary is given. In the summary, we discuss what has been designed and implemented, and what we have succeeded in achieving.

Chapter 2

A-correlation Between Two Stocks

2.1 Introduction

In this chapter, we first discuss the concept of data mining, and requirements on the design and implementation of data mining applications. And then, armed with knowledge of data mining, we define A-correlation. Lastly, we describe the programs that have been implemented.

2.1.1 Data Mining and Statistical Correlation

Correlation is a notion from statistics and probability [Fell 70][Garc 94]. It is a measure of the extent to which two random variables track one another. For example, the German Mark and the Dutch Guilder are positively correlated. They both tend to appreciate or depreciate relative to other currencies at the same time. Stock prices and interest rates tend to be negatively correlated. When one rises, the other tends to fall.

Correlation can be very useful in stock trading as well. By tracking the correlation between selected stocks, additional investment strategies may be adopted. Other investment strategies may range from trading one company off of another's news, spread

trading, to sector trading [Crow 77][Wein 96]. These strategies could potentially result in a higher return on investment ratio.

As it is evident from the above discussion, it is extremely beneficial to obtain the correlation between individual stocks of interest. The process of obtaining the correlation requires sorting through a large amount of data. In essence, one aspect of data mining is finding out statistical correlation among a large collection of data. Through the various methods and processes as previously described, the data that possesses a high degree of correlation with the subject of interest, is collected and grouped, and presented in a suitable manner.

2.1.2 Degree of Correlation

When stocks are positively or negatively correlated, movement in one stock is paralleled by movement in other correlated stocks. The degree of correlation, positive or negative, will determine the degree to which prices of correlated stocks will move together. And in the same way that different stocks have some degree of correlation to one another, different sectors, which are made up of groups of related stocks, are correlated to one another.

Correlation describes the strength, or degree to which two stocks are related to each other. That is, correlation lets us specify to what extent the two variables behave alike [Fell 70]. Correlation is typically expressed as a decimal ranging from negative one to positive one, with a result of zero indicating no correlation at all.

2.1.3 Data Mining and Existing Knowledge

Data mining deals with finding useful patterns in data. The words, "useful" and "pattern" are extremely subjective. The data *miner* defines "usefulness" of the collected data, which is typically obtained by applying algorithms. The algorithms search for some pre-defined "pattern", which helps a data *miner* in expanding his/her knowledge. A pattern can be understood as an entity representing (describing, characterizing) an abstract concept or a physical object. A recognized object is characterized by a description, representing information about the object, commonly called a pattern.

The patterns discovered by data mining are useful because they extend existing knowledge in useful ways. But "new" knowledge is not created "in a vacuum": it builds on existing knowledge, and this existing knowledge is in the mind of the expert. The expert therefore plays a critical role in data mining, both as an essential source of input (knowledge) and as the consumer of the results of data mining.

Once the basic concept of data mining is understood, the requirements of data mining still have to be defined before one can proceed to develop data mining applications. In the next section, we look at the requirements for data mining applications.

2.1.4 Data Mining Requirements

The requirement of data mining to be accessible to end-users shapes the requirements for data mining tools. These end-user oriented requirements can be described in many

different ways, but here we focus on three key properties: data mining tools must be interactive, incremental and iterative [Cios 98].

Interactive: Modern "desktop" applications are highly interactive, but here we focus on a deeper interpretation of "interactive": the user must be enabled to interact with the data, and not just with the technology. The user interface of data mining tools should be designed to highlight the properties of data, and play down the details of technology, whether the technology be database links, efficient indexing, visualization display parameters or machine learning algorithms.

Incremental: The data mining process is incremental. Each successive investigation builds on the results of the previous one. Therefore, the principle learning from experience applies not just to the data mining exercise as a whole but also to each step within it. Data mining tools must be designed to encourage this re-use of results as the data *miner*, in a step-by-step manner, builds up a picture of the patterns in the data. This means that data mining tools must be highly integrated. Queries must naturally lead to visualization, visualization to data transformation and modeling, and modeling to visualization or further queries. These transitions are merely examples. Overall, the process must appear seamless, with the effective methods of investigation at any point also being the most obvious, and without intervention of technological barriers or distractions.

Iterative: Data mining is seldom a simple linear process. Successive steps not only build on one another's results, but also refine the approach of earlier steps. For example, the results of modeling may show that the data should be further refined and that the modeling be repeated, or may point to areas for closer examination in an earlier data exploration phase. Any result may point to earlier steps, refining not only the data but also the process itself. Each step also has the potential to open up entirely new avenues of enquiry. It should be emphasized that the process is not organized into discrete steps concerned with different types of knowledge; rather the discovery of detailed properties of the data proceeds alongside a gradual refinement of the business concepts involved, and the unfolding of key patterns to be utilized.

In summary, the data mining process must be driven by those with expert knowledge of the domain. This has many implications for the process and for the tools that support it: the process must be thoroughly domain-oriented rather than technically oriented, and the tools must support an interactive, incremental and iterative style of work.

Data mining, the extraction of hidden information from large databases, is a powerful new technology with great potential to help people focus on the most important information in their data warehouses. Data mining tools can sometimes even predict future trends and behaviors, allowing one to make proactive, knowledge-driven decisions. The automated prospective analysis offered by data mining move beyond the analysis of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer the questions that traditionally were too time-consuming to be resolved. In this

thesis, data mining is used to determine stock A-correlation. As we will see later in this paper, knowledge of data mining lends itself well in the development of the stock A-correlation programs. We also demonstrate how the stock A-correlation programs are designed and implemented to be interactive, incremental and iterative, which are demanded by typical end-users.

2.1.5 A-correlation Model

Investors commonly ask how likely it is that a stock will go up/down given that a reference stock is going up/down. Our approach to solving this problem is to assume that the probability of one stock following another stock is related to the correlation coefficient [Garc 94] between the two. If the correlation coefficient between one stock and a reference stock is positive, then the stock is likely to move in the same direction as the reference stock. If the correlation coefficient is near zero, then it implies that the stock could move in either direction regardless of the reference stock. If the correlation coefficient is negative, it indicates that the stock is more likely to move in the opposite direction of the reference stock.

The key, as seen from the above discussion, is in finding the correlation coefficient between stocks S_1 and S_2. However, rather than calculating this correlation coefficient, we approximate it by using an alternative measure, which we call A-correlation. For each day over a range D of days, we assign a "1" if two stocks are moving in the same direction, and a zero if the opposite is true. The A-correlation is expressed as $N(S_1, S_2, D, d)$. $N(S_1, S_2, D, d)$ is the number of days beginning with the most recent day and going back D days

in which S_1 and S_2 moved in the same direction. D is a power of 2 ranging from 2 to 9. d is the number of days delayed. When d is zero, the A-correlation is referred to as no time delayed A-correlation. When d is non-zero, the A-correlation is referred to as d-days delayed A-correlation. It should be clear that A-correlation is not mathematically equivalent to the correlation coefficient, but it still generally reveals the relationship between the movement of two stocks.

Obviously, the A-correlation is a function of days, *D*. A-correlation between a reference stock and a set of other stocks is shown in the sample outputs in Chapter Three. In addition, the movements indicate that either a stock is becoming more A-correlated to the reference stock, or A-uncorrelated to the reference stock, as are shown in the sample outputs.

The programs have been implemented to obtain the A-correlation $N(S_1, S_2, D, d)$ between stocks for a wide range of time periods. The first program is implemented to obtain A-correlation provided d as zero. The second program is implemented to obtain A-correlation provided d as non-zero.

2.2 Implementation Issues

Based on the theory introduced previously, two programs, Correlation Analyzer (Program I) and Time Delayed Correlation Analyzer (Program II), have been implemented as Microsoft Excel/VBA applications. In this section, the implementation of both programs is discussed in detail.

2.2.1 A Brief Introduction to Excel 2000/VBA

2.2.1.1 Using Excel

- **Introduction to Excel**

Excel was originally designed by Microsoft as a spreadsheet program to facilitate the process of making large spreadsheets, accounting, and data processing. The program was initially introduced in 1985 [Dodge 95]. Over the past years, Excel has constantly been improved on. With each version, the designers of Excel have come up with brilliant new ways to make their spreadsheet more versatile and easier to use. The newest version, Excel 2000, has more powerful features than any older versions. For example, Excel 2000 walks users through the steps of creating new Web queries, and allows for automatic refreshes. Users can set up Web Queries to import real-time data—such as stock quotes and currency exchange rates—from the Web into your spreadsheet or database. Figure 2.1 shows an example of web query instructions.

Microsoft Excel 2000 provides developers and users with improved Web-enabled collaboration and information sharing, enhanced analysis tools, and software that are easier to manage and use. Those new features make Excel 2000 more suitable for accessing information on the Internet and analyzing the data [Microsoft 98]. The following three paragraphs demonstrate those new features in detail.

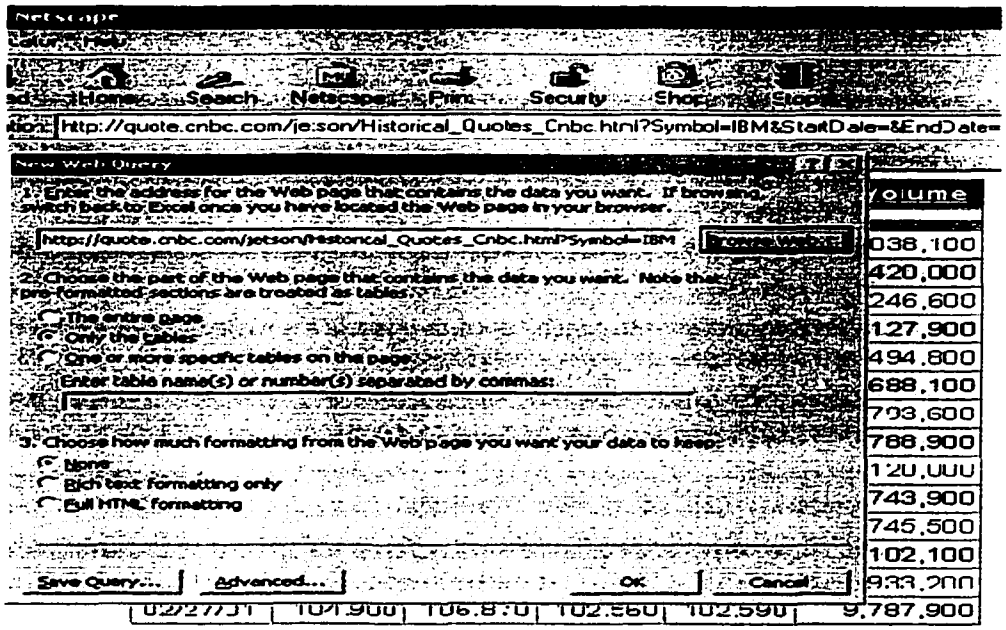


Figure 2.1 An example of web query instructions

Web-enabled collaboration and information sharing: Excel 2000 offers a new Web featured work style that integrates core productivity tools with the Web to streamline the process of sharing information and working with others. Excel 2000 is the preferred tool for working with tabular data on the Internet. It excels at creating, publishing, viewing and analyzing this data on the Web.

Rich analysis tools for better decision-making: Excel 2000 makes it easier to use an Internet/Intranet to access vital business information and provides innovative analysis tools that help users make better, more timely business decisions.

Software that is easy to use: Excel 2000 is very easy to use. It provides spreadsheets to organize and analyze mass data easily. It also provides a recording macro function for

developers to design the VBA program. It enables developers and users to get up running quickly and achieves great results with fewer resources.

- **Significant Features in Excel**

Object Models

The Microsoft Excel object model contains several dozen objects that developers can manipulate through VBA code. Developers can access most of Excel functionalities through the user interface alone. For example, developers can directly manipulate objects through VBA. However, the Excel user interface is limited. Complicated tasks such as automated data retrieving, pattern searching and repetitive processes are better to be carried out through the use of VBA. In most of cases, software developers prefer to work with VBA because it is powerful and easy to use.

Objects in Excel	Objects and its operation used in correlation program (Examples)
Cells	Cells(35, 1).Value = -1 (Set the value of cell(35, 1) is -1)
Ranges	MyRange = "A1:A10" Range(MyRange).Select (Select cells ranging from A1 to A10
Sheets	Sheets("StockData").Select (Select Sheets "StockData")
Charts	Charts.Add (Add a new charts)

Table 2.1 Examples of objects used in the programs

When using VBA to work with Excel objects either from within Excel itself or from another Office application, developers still maintain the access to every aspect of Excel.

The objects, which developers typically work with, include cells, ranges, sheets, workbooks, and charts. It should be noted that an object represents every element in Excel. Developers using VBA, can manipulate the object. The table 2.1 is an example of Excel objects, which are used in the development of the stock correlation program.

Web Queries

Excel 2000 provides a new web page query and a text file importation enhancement. Developers are able to import real-time data from a remote web site into spreadsheets, formulas or charts. The new web query dialog makes it easy to transfer data from the web onto an Excel spreadsheet. The dialog walks users through the process of bringing data from a web page into Excel, helping them create a query file as they choose the URL and parameters for how to import the data. Web query pages can be refreshed automatically on a scheduled basis.

Programmers can set up web queries to import real-time data, such as stock quotes and currency exchange rates, from a remote web site into a spreadsheet. Excel 2000 can help developers to go through the steps of creating new Web queries, and allows for automatic refreshes.

2.2.1.2 Visual Basic for Applications (VBA)

- **What is VBA?**

Microsoft Visual Basic for Applications (VBA) is subset of Visual Basic Language introduced by Microsoft. VBA is an embeddable programming environment designed to

enable developers to build custom solutions using the full power of Microsoft Visual Basic. Developers, who use applications that host Visual Basic for Applications, can automate and extend the application functionality, shortening the development cycle of custom business solutions.

- **Differences between Visual Basic and VBA (Why choose VBA, instead of Visual Basic?)**

Visual Basic is a standalone tool for creating independent software components such as executable programs, COM components and ActiveX® controls. Visual Basic is useful when you must build a specialized solution from scratch.

VBA offers equally powerful tools as Visual Basic in the context of an existing application. VBA is one of the best solutions for customizing software that already meets most of people's needs. For example, Excel provides numerous function calls, which enable complex data processing. Through Excel/VBA, a user can use powerful VBA object models perform complicated tasks such as automated data retrieving.

- **VBA Security Features**

Visual Basic for Applications includes ability to prevent users from viewing or modifying code. Developers can password-protect their VBA code so it is encrypted and cannot be viewed. For Excel/VBA, users can view and change code easily if code is not protected.

That's very important for an application to be protected from corrupting intentionally or unintentionally.

2.2.1.3 Summary

The implementation of our programs requires the retrieving of Internet data. In addition, the implementation demands spreadsheet capability to store stock data, which can then be quickly and easily organized and analyzed. Because of its features, Excel 2000/VBA was chosen as the preferred application environment and language.

2.2.2 Implementation Using Excel/VBA

In this section, we present the implementation of two programs: Correlation Analyzer and Time Delayed Correlation Analyzer. The first program demonstrates how to obtain stock A-correlation without time delay, and most A-correlated stock with respect to a reference stock. The second program demonstrates how to obtain stock A-correlation with d-days' delay.

As previously mentioned, the "reference stock" represents the stock that all other stocks are compared to. The "comparison stocks" represents the other stocks that user is interested in comparing to the reference stock.

2.2.2.1 Correlation Analyzer (Program I)

Correlation Analyzer is implemented to obtain A-correlation $N(S_1, S_2, D, d)$ provided the time delay, d, is zero. Figure 2.2 is a flow chart that demonstrates the structure of the program.

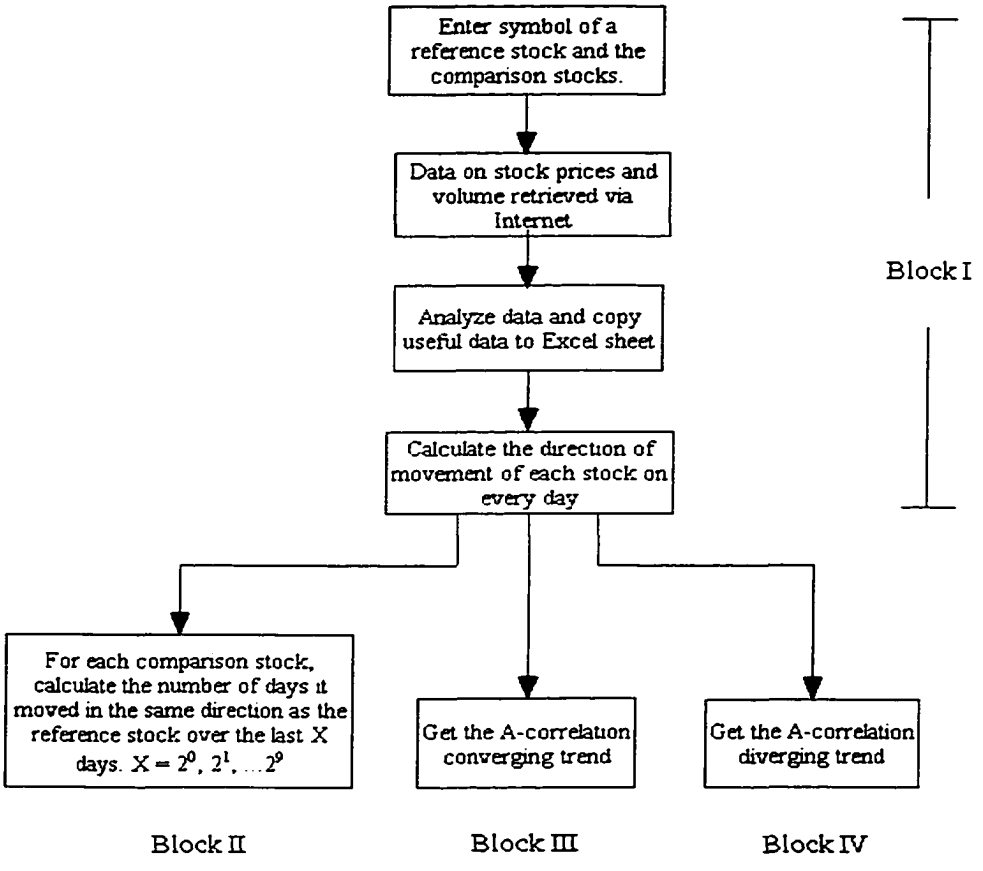


Figure 2.2 Correlation Analyzer program flow chart

The overall program can be divided into four functional blocks as indicated in Figure 2.2 Block I carries out the retrieval of the stock data from Internet. Block II performs preliminary calculations on the data retrieved and obtains the A-correlation between reference stock and each of the comparison stocks. Block III and IV deal with issues such as getting A-correlation converging trend and A-correlation diverging trend. A-correlation

converging trend indicates which one of the comparison stocks is becoming more A-correlated to the reference stock. A-correlation diverging trend indicates which one of the comparison stocks is becoming less A-correlated to the reference stock.

The following four sections are dedicated to an explanation and discussion of each of the four functional blocks.

2.2.2.1.1 Data Retrieval Via Internet (Block I)

Figure 2.3 shows the structure of the program (Block I), which retrieves stock data from Internet. The Internet site, which was used in this thesis, is http://www.cnbc.com.

With the Internet becomes increasingly popular, stock information, such as price, volume, P/E ratio, and company news and earning, can be found on numerous sites. These sites not only provide static information, but also real time stock trading information that can be extremely useful for regular day-time traders. The most popular sites include www.nasdaq.com (NASDAQ), www.amex.com (AMEX), and www.nyse.com (New York Stock Exchange). In addition, there are number of Internet sites that provide not only stock information, but also financial management, for a fee.

In this thesis, the closing price and trading volume of the stocks are used in calculating A-correlation. By forming a web query, the stock A-correlation program is capable of automating the access of the web site. The access of the web sites includes retrieving historical stock quotes that are demanded by users.

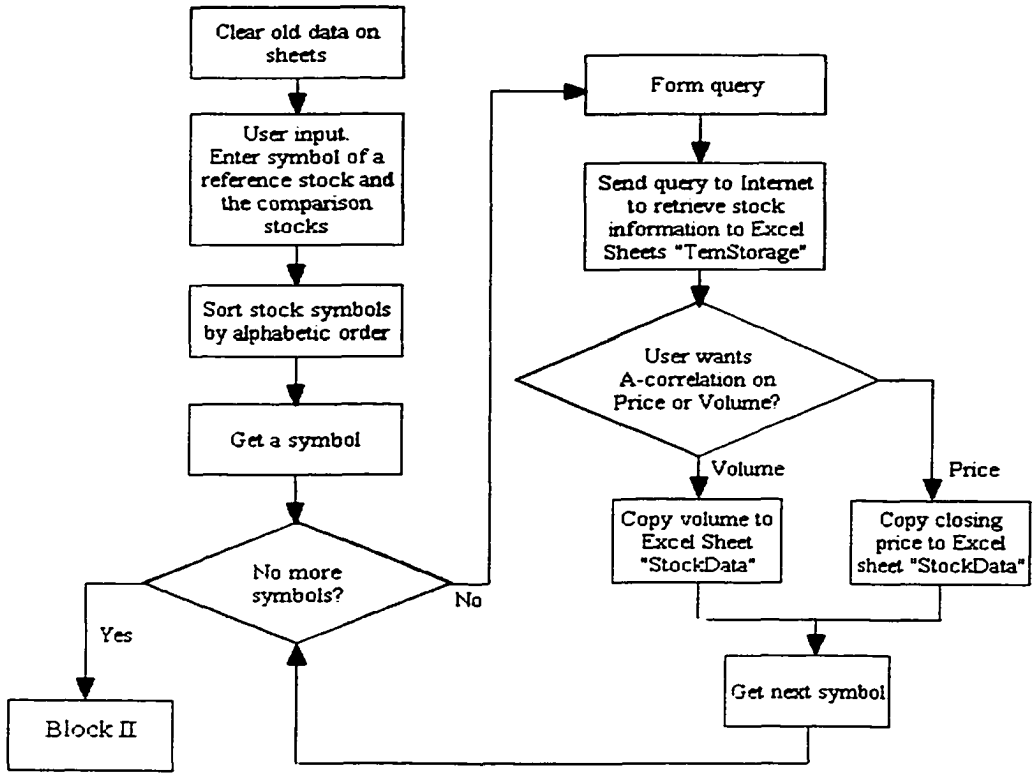


Figure 2.3 Data retrieval flow chart (Block I)

In the program, six worksheets have been used. The sheet "TemStorage" stores historical data temporarily and stores stock symbols. The Sheet "StockData" stores useful stock data and calculation results. The sheet "Result" presents A-correlation results. The sheet "Rank_Results" shows ranked A-correlation results. Finally, the sheet "C_Trend" shows A-correlation converging trend and the sheet "D_Trend" shows A-correlation diverging trend.

Typically, it is assumed that before retrieving the stock data from a web site, the worksheets are not empty. This is a problem that the existing obsolete data could corrupt

the latest retrieved data. The program clears all the worksheets by calling a function **ClearSheetsSub** before execution.

After clearing all work sheets, the user is asked to enter symbols (or ticker name) of a reference stock. Then the comparison stocks, which are to be A-correlated with the reference stock, have to be entered by the user. Once the comparison stocks and reference stock have been entered, the program will automatically gather the necessary data associated with the stock symbols over Internet. The retrieved data will be analyzed and the A-correlation between the reference stock and each comparison stock will be calculated.

Two functions have been implemented for the purpose of stock data retrieving: Function **GetStockInformation** has been implemented to function as mentioned above. Function **SortStocks** sorts stocks in alphabetic order to present them properly.

Once the reference stock and a set of comparison stocks have been entered, the program forms the appropriate web queries and then retrieves the stock information from the CNBC (http://www.cnbc.com) web site. The CNBC web site contains a large database of historical stock information. The retrieval of the data is achieved by utilizing an HTTP format query. The program will form an HTTP format query automatically and wait for a response from the CNBC web site. To obtain the correct stock information, the format of the query to the CNBC web site must contain a file path, stock symbol, start-date and end-date. The end date refers to the current day, and the start-date may be exactly two years

earlier from the current day. In this program, two years (504 trading days, 24 months) of data has been collected and processed. Function **StartDateContent** and **EndDateContent** are responsible for forming the proper date format, which is required by CNBC web site. Function **RetrieveData** forms the final query based on the stock symbols entered by user.

A typical query format, for example, will look like the following: *http://quote.cnbc.com/jetson/Historical_Quotes_Cnbc.html?Symbol=IBM&StartDate=02%2F05%2F99&EndDate=02%2F08%2F01&Type=0&Button=Get+Quotes&Format=0&sort=%7Cdate*. The program uses the above query to access CNBC web site to obtain stock quotes. Then the stock data is copied to Sheets "TemStorage", which is used to store information temporarily.

A stock quote from CNBC web site contains the following information: date, open, high, low, close, and volume. Among the available information, we can use date, closing price, or volume depending on the user's interest. The function **GetPriceArray** determines if closing price or volume information is used according to user input.

For each stock symbol entered by the users, the program repeats the above functions to collect all the stock information. Table 2.2 is an example of a stock information table, which has been sorted from raw Internet data. "X" row represents stock symbols (reference stock and comparison stocks). "Y" column represents date, and the cells in "Z" area represent the closing price of stock.

	pfe	adi	aol	brcm	cal
2/8/2001	44.75	50.47	48.76	78.875	51.19
2/7/2001	44.5	49.4	48.2	82.812	51.78
2/6/2001	45.85	52.44	48.85	92.5	52.58
2/5/2001	45.75	53.85	49.37	95.25	51.21
2/2/2001	44.9	55.2	47.79	99.125	54
2/1/2001	45.21	59.2	49.83	107	51.5
1/31/2001	45.15	62.6	52.56	109.937	52.15

X

Y

Z

Table 2.2 Example of stock information in Sheets "StockData"

The next section explains how the program uses the retrieved data to calculate A-correlation between a reference stock and the comparison stocks.

2.2.2.1.2 Most A-correlated Stocks (Block II)

The most A-correlated stock is defined as the single comparison stock which has the highest A-correlation with the reference stock. This is achieved by implementing three functions. The three functions have been designed to work together to find out the A-correlation between the reference stock and each of the comparison stocks, and most importantly, to determine that which one of the comparison stocks is most A-correlated to the reference stock.

Function GetMovingDirection

Input: Sheet "StockData"; Number of Stocks; Number of Data (for each stock)

Output: a) The direction of movement of each stock in everyday; b) Index, j (a row number of Sheets "StockData". j indicates a position that is the end of stock data)

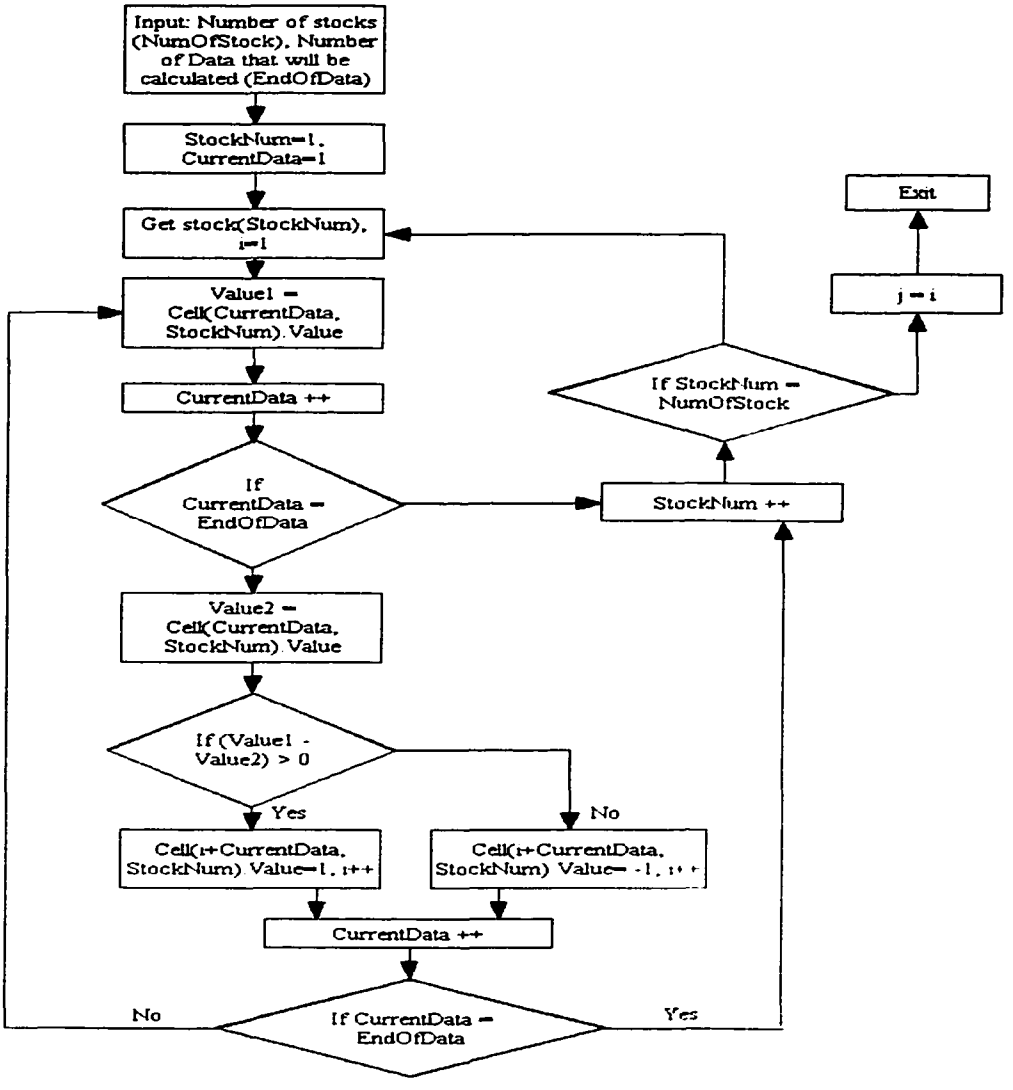


Figure 2.4 Flow chart of function GetMovingDirection

GetMovingDirection is used to obtain the direction of movement for each stock every day. A daily change in stock is calculated by subtracting the previous day's closing price from the closing price of the current day. The daily change in stock would be checked to find a direction of movement. If the change is positive, we use a "1" to represent the upward movement. If the change is negative, we use a "–1" to represent the downward

movement. This process is repeated until each stock's direction is obtained for each day. The value of the direction of movement is stored in unused cells in sheets "StockData".

Function IfMovingSameDirection

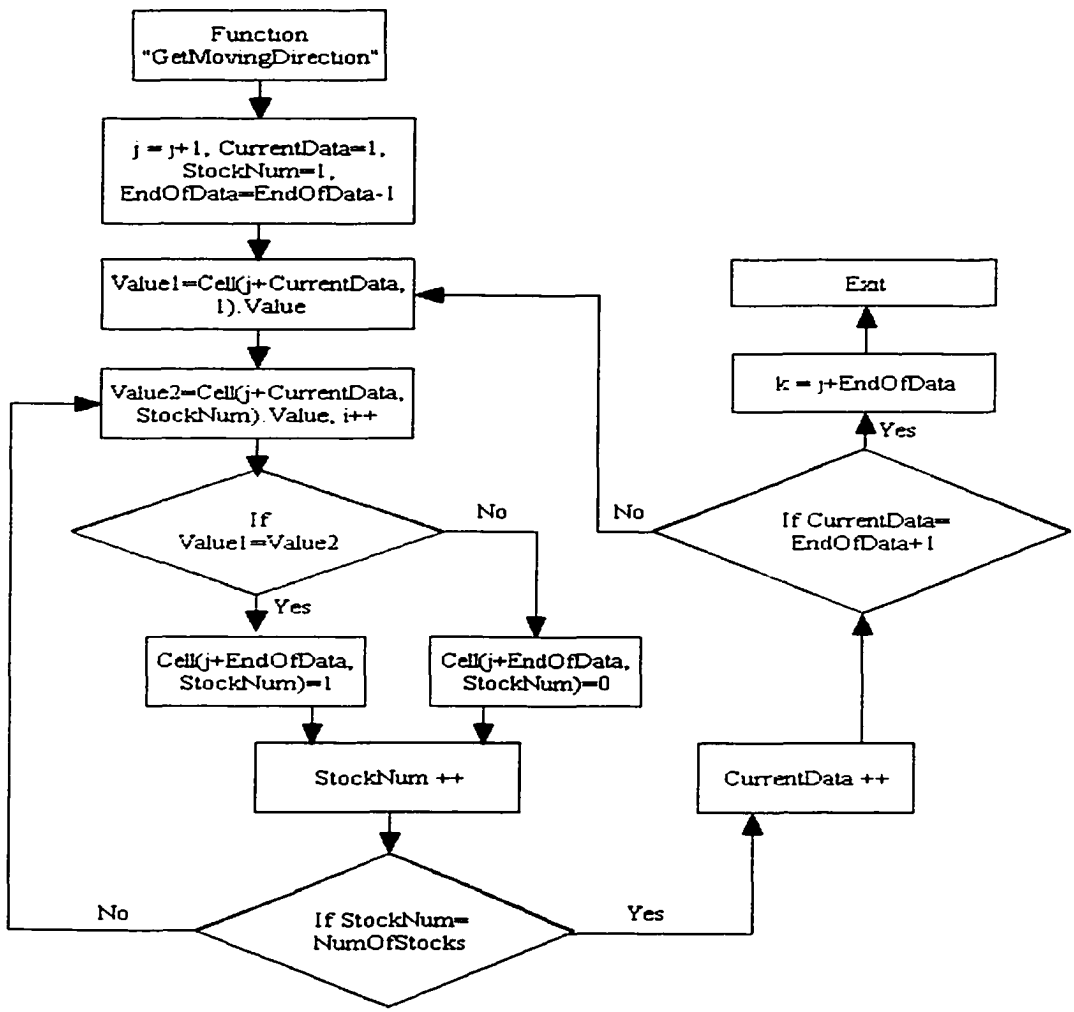


Figure 2.5 Flow chart of function IfMovingSameDirection

Input: a) Sheet "StockData"; The direction of movement of each stock (the reference stock and the comparison stocks) in everyday; b) Index, j (which is a row number of Sheets "StockData". j indicates a position that is the end of stock data.)

Output: a) Result of checking if one of the comparison stock moved the same direction as the compared stock in each day; b) Index, k (a row number of Sheets "StockData". k indicates a position that is the end of calculated data of the direction of movement.)

IfMovingSameDirection is used to compare the direction of movement of one of the comparison stocks with respect to the reference stock. If the value of the direction of movement is the same, we store a "1" in an unused cell; Otherwise, a "0" to represent that the direction of movement is different. This step is repeated until there is no more data to be compared.

Function DataSum

Input: a) Result of checking if the comparison stocks moved in the same direction as the reference stock each day; b) Index, k (a row number of Sheets "StockData". k indicates a position that is the end of calculated data of the direction of the movement.)

Output: A-Correlation on each one of the comparison stocks

DataSum is used to find out which one of the comparison stocks is most A-correlated to the reference stock. This function will count the number of days the reference stock moved in the same direction as each of the comparison stocks over the last *X* trading days. The result is stored in the worksheet "StockData" temporarily, and then copied to the worksheet "Result". In this program, we use the sum of data within the last 1 day, the last 2 days, the last 4 days...the last 256 days and the last 512 days.

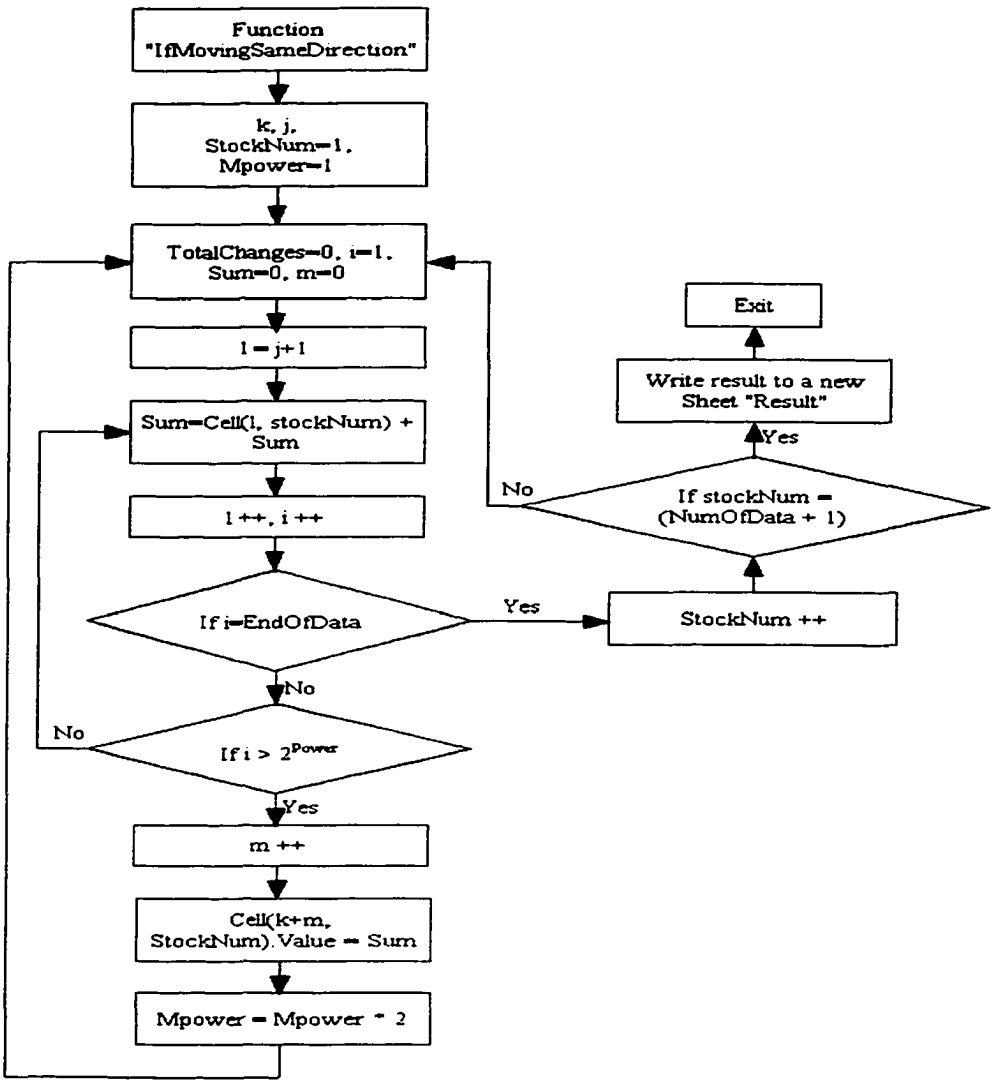


Figure 2.6 Flow chart of function DataSum

Function CopySub copies the result in Sheet "StockData" into a table "Up/Down Price A-correlation" in the worksheet "Result" for users to read.

Function RedRankMax is used to rank the maximum number in each column of "Up/Down Price A-correlation" red. The largest number in red indicates the one of the comparison stock that is the most A-correlated with the reference stock. The sample output in this thesis uses a square to denote the red number.

Function RankResult is used to rank the numbers in each column of "Up/Down Price A-correlation" and save the result on sheets "Rank_Results". The ranking in A-correlation between stocks can give user a clear picture of which stock is most A-correlated to the reference stock, and which isn't.

2.2.2.1.3 A-correlation Converging Trend (Block III)

This section explains how the program spots a converging trend in the stock A-correlation. The trend indicates which one of the comparison stocks is becoming more A-correlated to the reference stock.

There is one main function **DataSumCertainDays**, which is designed to obtain A-correlation converging trend.

Function DataSumCertainDays

Input: a) Sheets "StockData"; Result of checking if the comparison stocks are moving in the same direction as the reference stock in each day; b) Index, k (a row number of Sheets "StockData". k indicates a position that is the end of calculated data of the direction of movement.)

Output: A-correlation converging trend

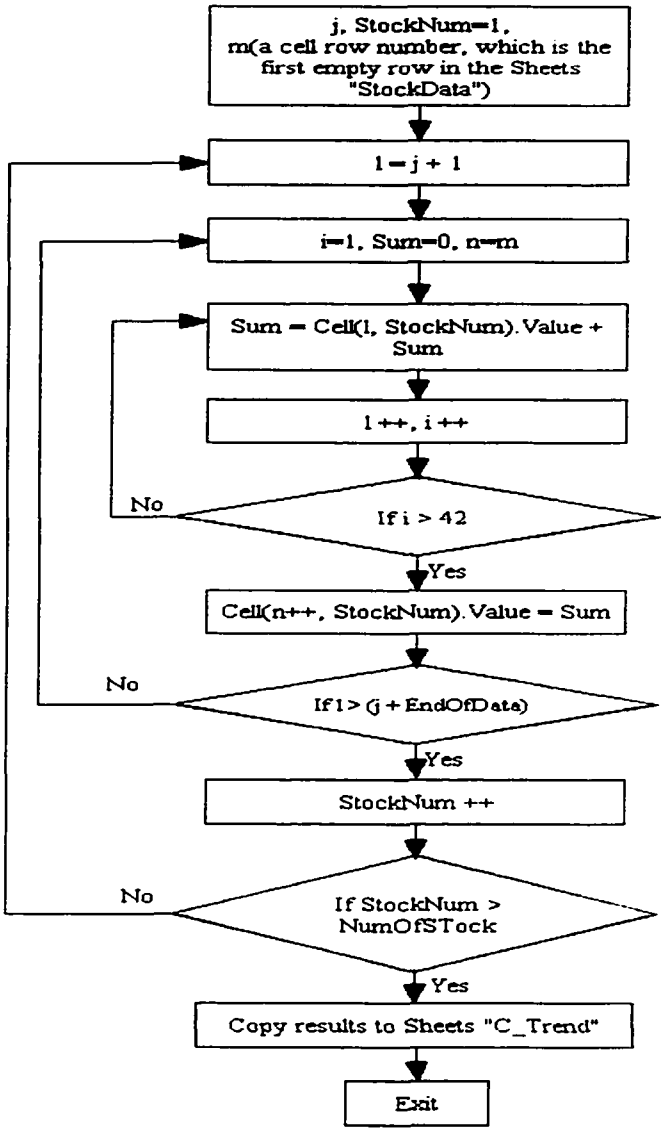


Figure 2.7 Flow chart for obtaining A-correlation converging trend

DataSumCertainDays is used to count the number of days from the past to the most recent day. The number of days counted is divided into intervals. Each interval has the same number of days. By checking the correlation in each interval, the trend in correlation can be obtained. Figure 2.7 is the flow chart showing how the function **DataSumCertainDays** is organized.

Function DrawChart is used to draw a chart showing the A-correlation converging trend, which gives the user a visual presentation of the converging trend.

Function CopySub is used to copy result of A-correlation converging trend in Sheets "StockData" into the table "Up/Down Price A-correlation Converging Trend" in sheets "C_Trend" for user reading ease.

Function RedRankMax is used to rank the numbers in red in each column of "Up/Down Price A-correlation Converging Trend". The largest number in red indicates which one of the comparison stocks is becoming more A-correlated with the reference stock.

2.2.2.1.4 A-correlation Diverging Trend (Block IV)

Implementation of A-correlation diverging trend is similar to the implementation of the A-correlation converging trend. The only difference is that Function **RedRankMin** is used instead of **RedRankMax** function.

Function **RedRankMin** ranks the number in red in each column of "Up/Down Price A-correlation Diverging Trend". The smallest number in red can inform the user on which one of the comparison stocks is becoming less and less A-correlated with the reference stock.

2.2.2.2 Time Delayed Correlation Analyzer (Program II)

In this section, implementation of the A-correlation N(S1, S2, D, d) with d-days delay is introduced. The idea behind this program is to compare the direction of movement of each of the comparison stocks with the reference stock in one previous day or several previous days, then calculate numbers of days in which both are moving in the same direction to obtain A-correlation.

The implementation of time-delayed stock A-correlation is similar to the implementation of the A-correlation without time delay. The only difference lies in the implementation of the function **IfMovingSameDirection**. Figure 2.8 is the flow chart of function **IfMovingSameDirection** in the Time Delayed Correlation Analzyer program. The function is explained in greater detail below.

Function IfMovingSameDirection

Input: a) Sheet "StockData"; The direction of movement of each stock (A and B_i) in everyday; b) Index, j (which is a row number of Sheets "StockData". j indicates a position that is the end of stock data.); c) Number of days delay (NumofDaysDelay)

Output: a) Result of checking if the comparison stocks are moving in the same direction as the reference stock in one previous day or several previous days; b) Index, k (a row number of Sheets "Data". k indicates a position that is the end of calculated data of the direction of movement.)

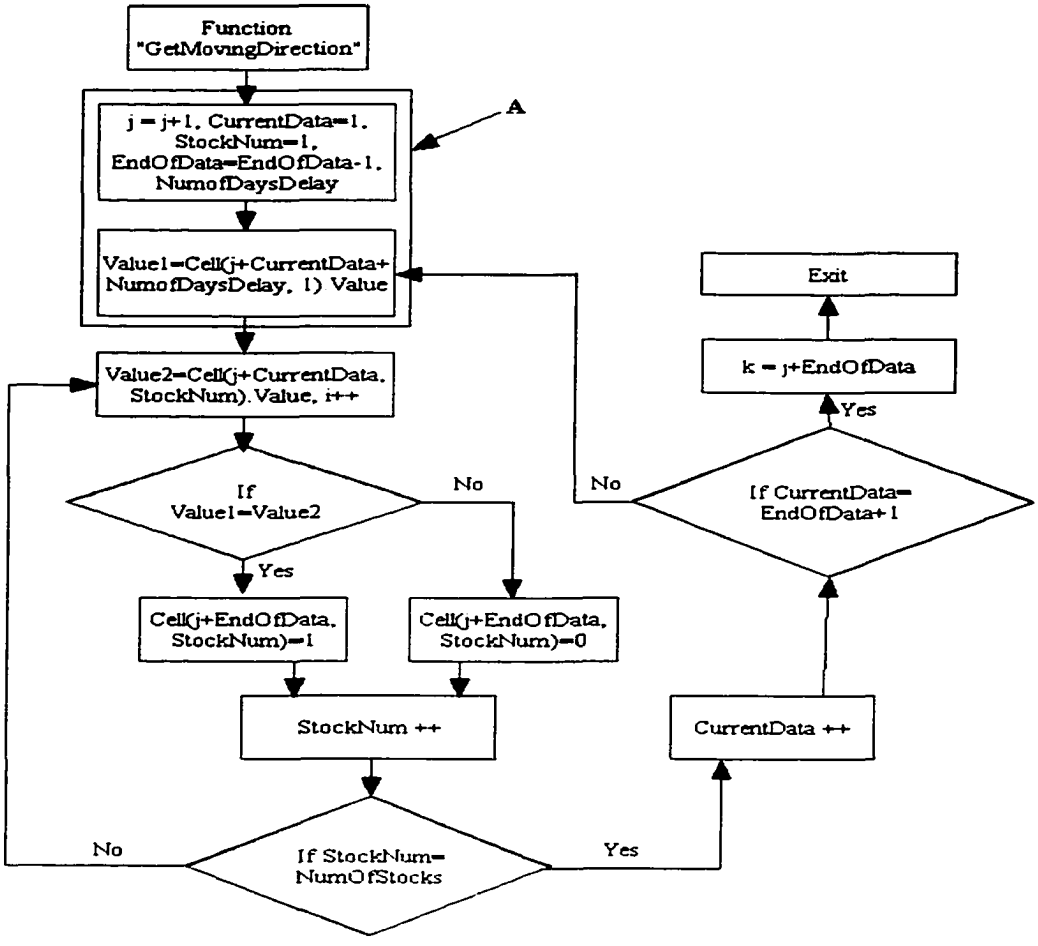


Figure 2.8 Flow chart of function IfMovingSameDirection

(Time Delayed Correlation Analyzer Program,

Similar to Figure 2.5 except for blocks pointed to by "A")

IfMovingSameDirection is used to compare the direction of movement of the comparison stock in d-days' delay with respect to the reference stock. If the direction of movement is the same, we store a "1" into an unused cell; otherwise, a "0" to represent that the direction of movement is different. This step is repeated until there is no more data to be compared.

2.3 Summary

This chapter introduces the theory behind data mining and how the A-correlation is related to data mining. Most of Chapter Two dedicates to the design and implementation of the Correlation Analyzer program and Time Delayed Correlation Analyzer program. Each function and associated structural flow is explained in detail. In the next chapter, sample outputs generated from running the above two programs are presented and discussed.

Chapter 3

Stock A-correlation Sample Output

3.1 Introduction

In this chapter, the sample outputs of the two programs are illustrated. We present how the programs are capable of providing useful information on the A-correlation among stocks not only without time delay, but also with d-days' delay.

3.2 Sample Output

3.2.1 Sample Output from Correlation Analyzer (Program I)

In this section, three sets of sample output from the Correlation Analyzer are presented. The sample output include obtaining the most A-correlated stocks, finding out which one of the comparison stocks is becoming more A-correlated to the reference stock, and which one is becoming less A-correlated.

In this example, the reference stock is PFE (PFIZER, INC). PFIZER is a drug company. The comparison stocks can be separated into six categories: drug companies, airline companies, chip design companies, computers, auto, and media companies. Table 3.1

gives a brief explanation of each symbol in each category. The variety of selected stocks provides interesting results.

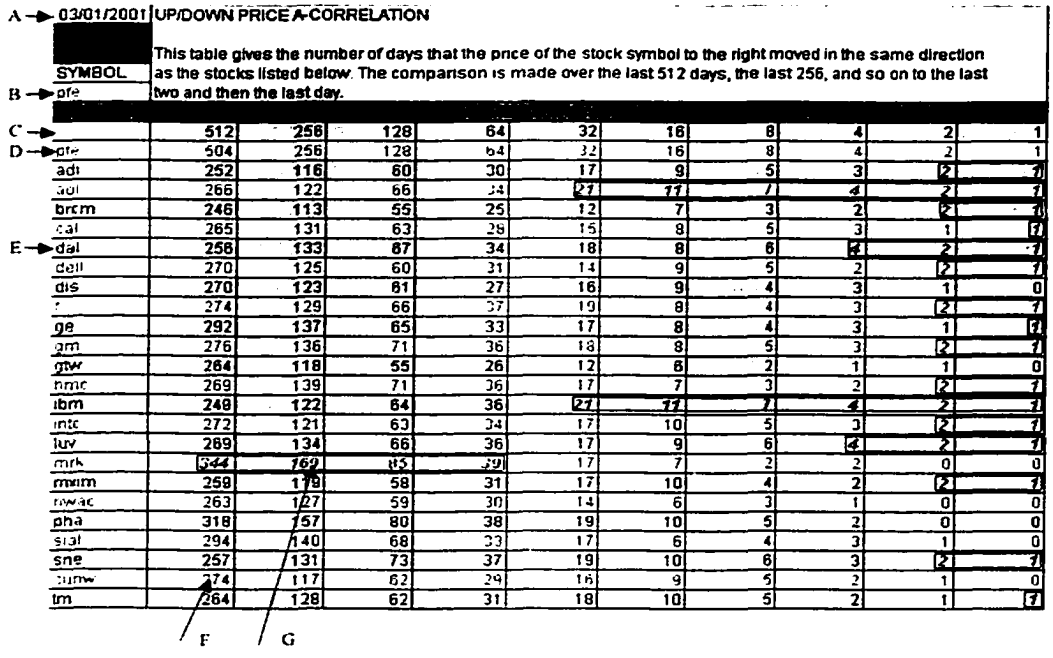
Drugs	Airlines	Chips	Computers /Peripherals:	Automotive	Media & Entertainment
PFE Pfizer Inc.	**DAL** Delta Airlines	**MXIM** Maxim Integrated Prod.	**DELL** Dell Computer Corp.	**F** Ford Motor Co.	**GE** General electric
MRK Merck & Co.	**NWAC** Northwest Airlines	**ADI** Analog Devices	**GTW** Gateway Inc.	**HMC** Honda Motor Adr.	**AOL** AOL Time Warner
PHA Pharmacia Corp.	**CAL** Continental Airlines Inc. CL. B	**BRCM** Broadcom Corp.	**IBM** International Business Machines	**GM** General Motors	**SNE** Sony Corp. Adr.
SIAL Sigma-Aldrich	**LUV** Southwest Airlines	**INTC** Intel Corp.	**SUNW** Sun Microsystems	**TM** Toyota Motor Ads.	**DIS** Disney

Table 3.1 Stocks overview

3.2.1.1 Most A-correlated Stocks

The Figure 3.1 shows which one of the comparison stocks is most A-correlated to the reference stock.

The following explains the first sample output, which is shown in the Figure 3.1. In the Figure 3.1, the value in cell "A" represents the day on which the Correlation Analyzer was run. The program will process over two years (504 days, see row "C", starting from current day back to two years) trading days' stock data to obtain the A-correlation. The value in cell "B" represents the symbol of the reference stock, in this example, PFE, the drug company.



A →	03/01/2001	UP/DOWN PRICE A-CORRELATION								
		This table gives the number of days that the price of the stock symbol to the right moved in the same direction								
	SYMBOL	as the stocks listed below. The comparison is made over the last 512 days, the last 256, and so on to the last								
B →	pfe	two and then the last day.								
C →		512	256	128	64	32	16	8	4	2	1
D →	pfe	504	256	128	64	32	16	8	4	2	1
	adi	252	116	60	30	17	9	5	3	2	1
	aol	266	122	66	34	21	11	7	4	2	1
	brcm	246	113	55	25	12	7	3	2	2	1
	cal	265	131	63	28	15	8	5	3	1	1
E →	dal	258	133	67	34	18	8	6	4	2	1
	dell	270	125	60	31	14	9	5	2	2	1
	dis	270	123	61	27	16	9	4	3	1	0
	f	274	129	66	37	19	8	4	3	2	1
	ge	292	137	65	33	17	8	4	3	1	1
	gm	276	136	71	36	18	8	5	3	2	1
	gtw	264	118	55	26	12	6	2	1	1	0
	hmc	269	139	71	36	17	7	3	2	2	1
	ibm	248	122	64	36	21	11	7	4	2	1
	intc	272	121	63	34	17	10	5	3	2	1
	luv	269	134	66	36	17	9	6	4	2	1
	mrk	244	169	85	39	17	7	2	2	0	0
	msft	258	118	58	31	17	10	4	2	2	1
	nwac	263	127	59	30	14	6	3	1	0	0
	pha	318	157	80	38	19	10	5	2	0	0
	sial	294	140	68	33	17	6	4	3	1	0
	sne	257	131	73	37	19	10	6	3	2	1
	sunw	274	117	62	29	16	9	5	2	1	0
	tm	264	128	62	31	18	10	5	2	1	1

F G

Figure 3.1 Result of most A-correlated stock to the reference stock

The row "C" represents the last one day (2^0 day), the last 2 days (2^1 days), ..., until the last 512 days (2^9 days). Actually, the program only uses the last 504 days, which can be divided by 42 evenly. This information is used in getting the second set of sample output. The row "D" represents A-correlation between the reference stock and itself. The value "E" cell represents the symbol of one of the comparison stocks. The value in cell "F" represents A-correlation between the reference stock and one of the comparison stocks, starting from the last one day until the last 504 days. The value in cell "G" is using an Italic and Bold format with a square. The special format implies this value is the highest in the column. The meaning of the special format is that during the last X days, the value of the cell in the special format is the highest A-correlation between the reference stock and the comparison stock.

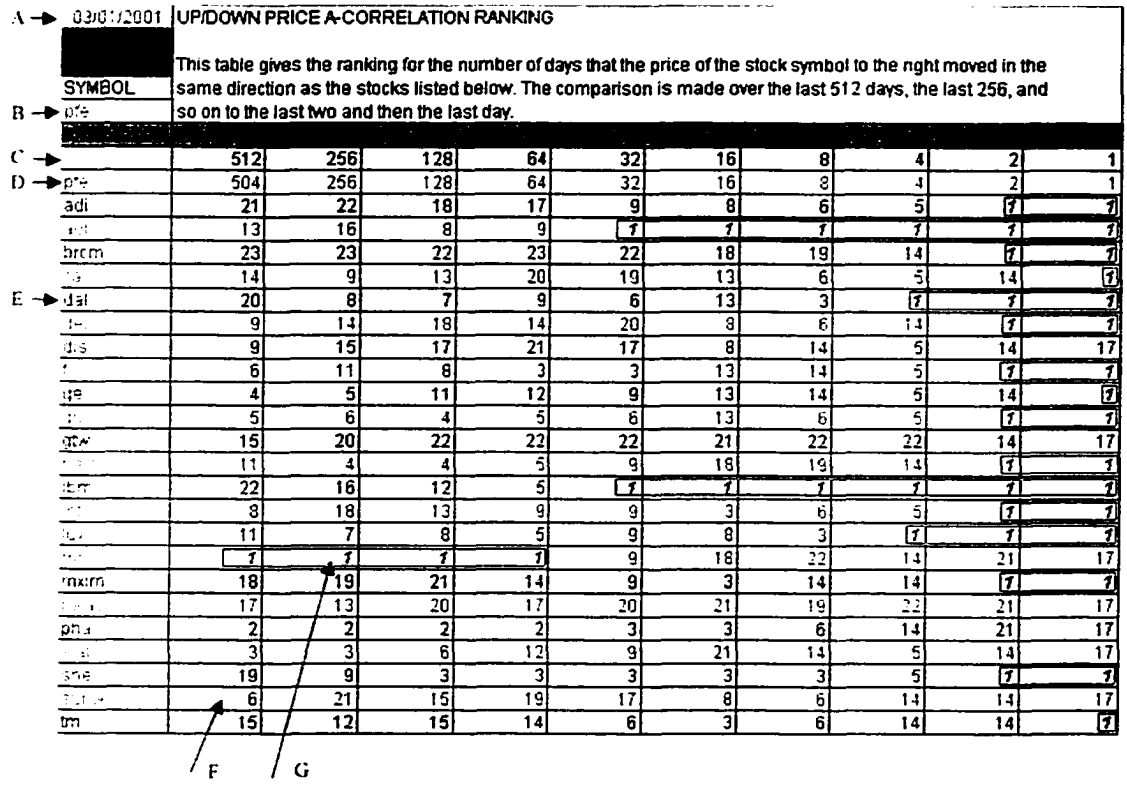
Figure 3.1 shows the A-correlation between the reference stock and the comparison stocks, and the most A-correlated stock with the reference stock. For example, the number, 169, in the cell that "H" is pointing to, is in the special format. This means during the last 256 days, MRK (Merck, a drug company) has moved 169 days in the same direction of movement as PFE (Pfizer Inc., a drug company too). Merck was most correlated to Pfizer during the last 256 days among all the 23 comparison stocks.

Figure 3.1 shows that during the last one day, there are sixteen stocks moved in the same direction (either up or down) as Pfizer. During the last two days, there are thirteen stocks moved in the same direction as Pfizer. Starting from the last four days, the special format becomes rare. From the last eight days to the last 32 days, AOL (AOL Time Warner) and IBM were the most correlated to Pfizer. During the last 64 days until the last 504 days (two years), Merck was the most A-correlated stock to Pfizer. Both of them are drug companies.

Another result from Figure 3.1 is that in the short term, AOL Time Warner and IBM were the most A-correlated stocks to Pfizer. In the long term, Merck was the most A-correlated to Pfizer.

The following Figure 3.2 shows the same result but through a different angle. It ranks every column in table "Up/Down Price A-correlation" according to which stock is most A-correlated to the reference stock. Not only do the most A-correlated stocks show clearly, but also the most A-uncorrelated stocks.

In the short term, GTW (Gateway Inc., a computer company) was the most A-uncorrelated to Pfizer. In the long term, BRCM (Broadcom Corp., a chip design company) was the most A-uncorrelated to Pfizer.



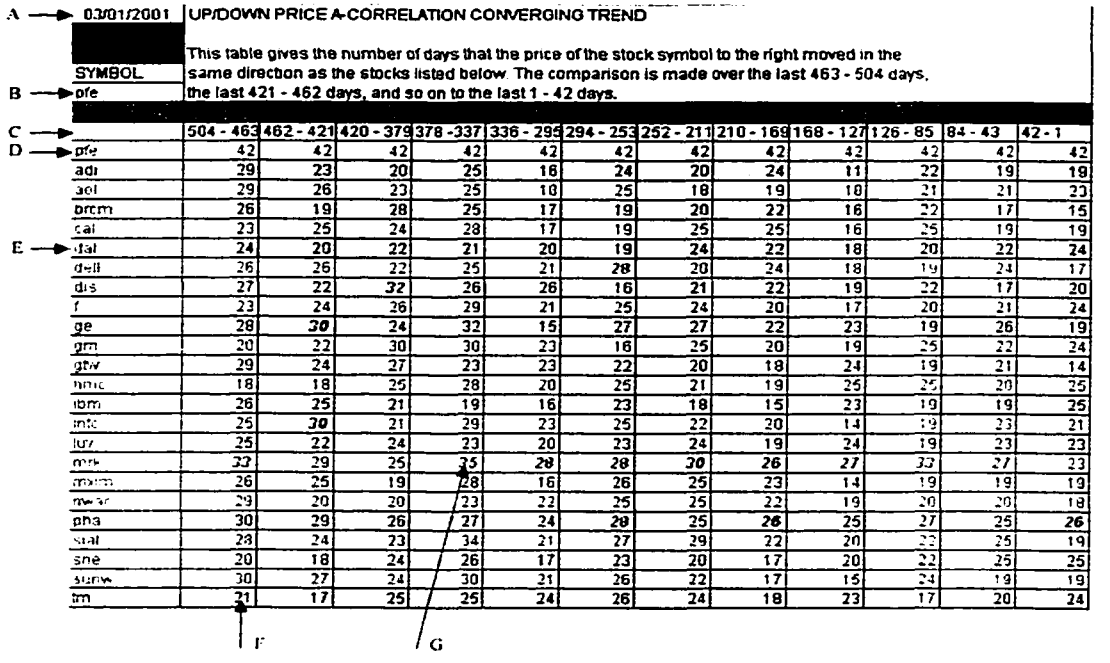
SYMBOL	512	256	128	64	32	16	8	4	2	1
pfe	504	256	128	64	32	16	8	4	2	1
adi	21	22	18	17	9	8	6	5	1	1
[illegible]	13	16	8	9	1	1	1	1	1	1
brcm	23	23	22	23	22	18	19	14	1	1
[illegible]	14	9	13	20	19	13	6	5	14	1
dal	20	8	7	9	6	13	3	1	1	1
[illegible]	9	14	18	14	20	8	6	14	1	1
[illegible]	9	15	17	21	17	8	14	5	14	17
[illegible]	6	11	8	3	3	13	14	5	1	1
[illegible]	4	5	11	12	9	13	14	5	14	1
[illegible]	5	6	4	5	6	13	6	5	1	1
gtw	15	20	22	22	22	21	22	22	14	17
[illegible]	11	4	4	5	9	18	19	14	1	1
ibm	22	16	12	5	1	1	1	1	1	1
[illegible]	8	18	13	9	9	3	6	5	1	1
[illegible]	11	7	8	5	9	8	3	1	1	1
[illegible]	1	1	1	1	9	18	22	14	21	17
mxim	18	19	21	14	9	3	14	14	1	1
[illegible]	17	13	20	17	20	21	19	22	21	17
phm	2	2	2	2	3	3	6	14	21	17
[illegible]	3	3	6	12	9	21	14	5	14	17
[illegible]	19	9	3	3	3	3	3	5	1	1
[illegible]	6	21	15	19	17	8	6	14	14	17
tm	15	12	15	14	6	3	6	14	14	1

Figure 3.2 Ranking result of most A-correlated stock to the reference stock

3.2.1.2 A-correlation Converging Trend

In this section, the sample output of A-correlation converging trend is illustrated. Figure 3.3 shows the sample output. The meaning of the different part of Figure 3.3 is explained.

"A" and "B" have the same meaning as explained in Figure 3.1. Typically, a total of 504 days is divided into twelve intervals. Each interval has a same number of days (42 days). Row "C" represents the consecutive intervals starting from the last 1 - 42 days, until the last 463 – 504 days.



A →	03/01/2001	UP/DOWN PRICE A-CORRELATION CONVERGING TREND										
		This table gives the number of days that the price of the stock symbol to the right moved in the										
	SYMBOL	same direction as the stocks listed below. The comparison is made over the last 463 - 504 days,										
B →	pfe	the last 421 - 462 days, and so on to the last 1 - 42 days.										
C →		504 - 463	462 - 421	420 - 379	378 -337	336 - 295	294 - 253	252 - 211	210 - 169	168 - 127	126 - 85	84 - 43	42 - 1
D →	pfe	42	42	42	42	42	42	42	42	42	42	42	42
	adi	29	23	20	25	18	24	20	24	11	22	19	19
	aol	29	26	23	25	10	25	18	19	10	21	21	23
	brcm	26	19	28	25	17	19	20	22	16	22	17	15
	cal	23	25	24	28	17	19	25	25	16	25	19	19
E →	dal	24	20	22	21	20	19	24	22	18	20	22	24
	dell	26	26	22	25	21	28	20	24	18	19	24	17
	dis	27	22	32	26	26	16	21	22	19	22	17	20
	f	23	24	26	29	21	25	24	20	17	20	21	24
	ge	28	30	24	32	15	27	27	22	23	19	26	19
	gm	20	22	30	30	23	16	25	20	19	25	22	24
	gtw	29	24	27	23	23	22	20	18	24	19	21	14
	hmc	18	18	25	28	20	25	21	19	25	25	20	25
	ibm	26	25	21	19	16	23	18	15	23	19	19	25
	intc	25	30	21	29	23	25	22	20	14	19	23	21
	luv	25	22	24	23	20	23	24	19	24	19	23	23
	mrk	32	29	25	35	28	28	30	26	27	33	27	23
	msft	26	25	19	28	16	26	25	23	14	19	19	19
	nwac	29	20	20	23	22	25	25	22	19	20	20	18
	pha	30	29	26	27	24	28	25	26	25	27	25	26
	stal	28	24	23	34	21	27	29	22	20	23	25	19
	sne	20	18	24	26	17	23	20	17	20	22	25	25
	sunw	30	27	24	30	21	26	22	17	15	24	19	19
	tm	21	17	25	25	24	26	24	18	23	17	20	24

F G

Figure 3.3 Result of A-correlation converging trend

Row "D" represents the A-correlation between Pfizer (PFE) and itself during each interval. Row "E" represents same meaning as the results in Figure 3.1. The cells of "F" and "G" have the same meaning as the result in Figure 3.1.

Figure 3.4 is the graphical representation of Figure 3.3, which gives a more directly view of which the comparison stock is becoming more A-correlated to the reference stock.

A-correlation between a given group of stocks and a reference stock is a complex function of stocks category, time, and other factors (Economic indexes, company policy, and etc). A-correlation can vary significantly from time to time. Upon close examination of the conducted sample output, we have arrived at several conclusions, which are discussed below.

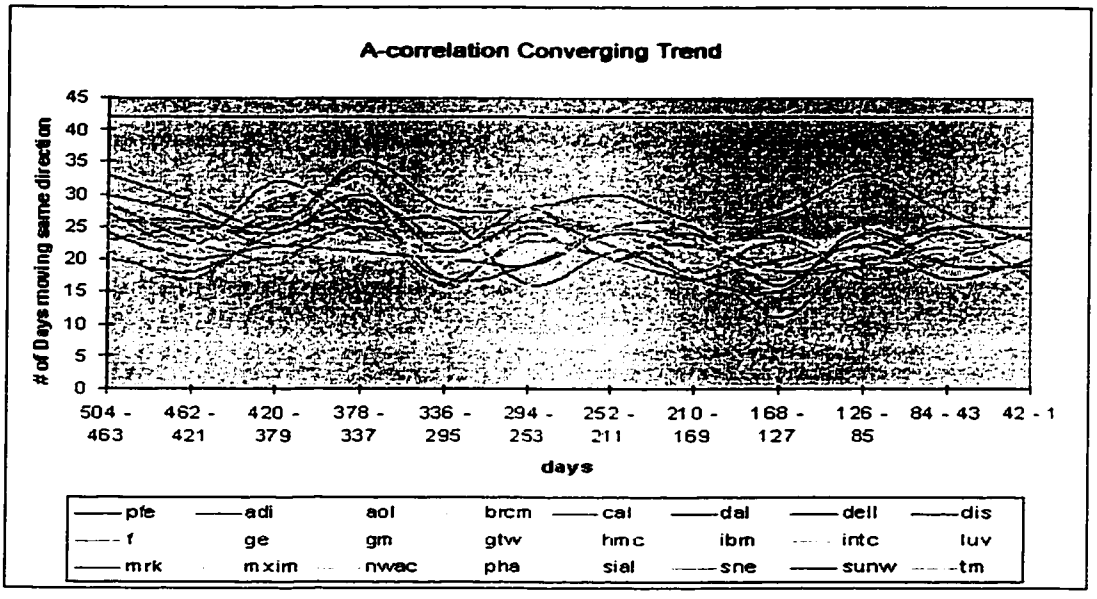


Figure 3.4 Chart of A-correlation converging trend

First of all, stock A-correlation is not static, which can be seen from both Figure 3.3 and Figure 3.4. The stock A-correlation varies not only from one period of time to another period of time, but also over different period lengths. The variation in A-correlation could be a result from several reasons such as company management policies, change in financial environment, and technology innovation. Due to the chaotic nature of the stock market, it is difficult to pin down one specific reason for the change in the A-correlation.

In addition, stocks in the same industry as the reference stock typically exhibit much higher A-correlation over a long period of time. This is not surprising as the company in the same industry shares common products and markets, in turn, the ups and downs of their stock value.

3.2.1.3 A-correlation Diverging Trend

A → 03/01/2001 UP/DOWN PRICE A-CORRELATION DIVERGING TREND

This table gives the number of days that the price of the stock symbol to the right moved in the same direction as the stocks listed below. The comparison is made over the last 463 - 504 days, the last 421 - 462 days, and so on to the last 1 - 42 days.

B → SYMBOL pfe

C →

D →	504 - 463	462 - 421	420 - 379	378 -337	336 - 295	294 - 253	252 - 211	210 - 169	168 - 127	126 - 85	84 - 43	42 - 1
pfe	42	42	42	42	42	42	42	42	42	42	42	42
adi	29	23	20	25	16	24	20	24	11	22	19	19
aol	29	26	23	25	18	25	18	19	18	21	21	23
brcm	26	19	28	25	17	19	20	22	16	22	17	15
cal	23	25	24	28	17	19	25	25	16	25	19	19
E → dal	24	20	22	21	20	19	24	22	18	20	22	24
dell	26	26	22	25	21	28	20	24	18	19	24	17
dis	27	22	32	26	26	16	21	22	19	22	17	20
f	23	24	26	29	21	25	24	20	17	20	21	24
ge	28	30	24	32	15	27	27	22	23	19	26	19
gm	20	22	30	30	22	18	25	20	19	25	22	24
gtw	29	24	27	23	23	22	20	18	24	19	21	14
hmc	18	18	25	28	20	25	21	19	25	25	20	25
ibm	25	25	21	19	16	23	18	15	23	19	19	25
intc	25	30	21	29	23	25	22	20	14	19	23	21
luv	25	22	24	23	20	23	24	19	24	19	23	23
mrk	33	29	25	35	28	28	30	26	27	33	27	23
mxim	26	25	19	28	16	26	25	23	14	19	19	19
nwac	29	20	20	23	23	25	25	22	19	20	20	18
pha	30	29	26	27	24	28	25	26	25	27	25	26
slal	28	24	23	34	21	27	29	22	20	22	25	19
sne	20	18	24	26	17	23	20	17	20	22	25	25
sunw	30	27	24	30	21	26	22	17	15	24	19	19
tm	21	17	25	25	24	26	24	18	23	17	20	24

F G

Figure 3.5 A-correlation diverging trend

Figure 3.5 shows stock A-correlation diverging trend. The meaning of mark "A", "B", "C", "D", "E", "F" is the same as Figure 3.3. The numbers in the special format with a square, which "G" is pointing to, represents a stock that is least A-correlated with reference stock over a period of days.

3.2.2 Sample Output from Time Delayed Correlation Analyzer (Program II)

In this section, two sets of sample output are displayed. Both of them are the output of looking for the most A-correlated stocks. The first set is obtained by calculating one-day delayed up/down price A-correlation. The second set is obtained by calculating twenty-days-delayed up/down price A-correlation. When conducting this example, the stocks that were used were slightly different from the ones used by the previous program. A brief description is given in table 3.2 for each stock different from previous program.

Chips	Computers/Peripherals:
AMD Advanced Micro Device	**SGI** Silicon Graphics
	MU Micron Technology
	WDC Western Digital Corp.

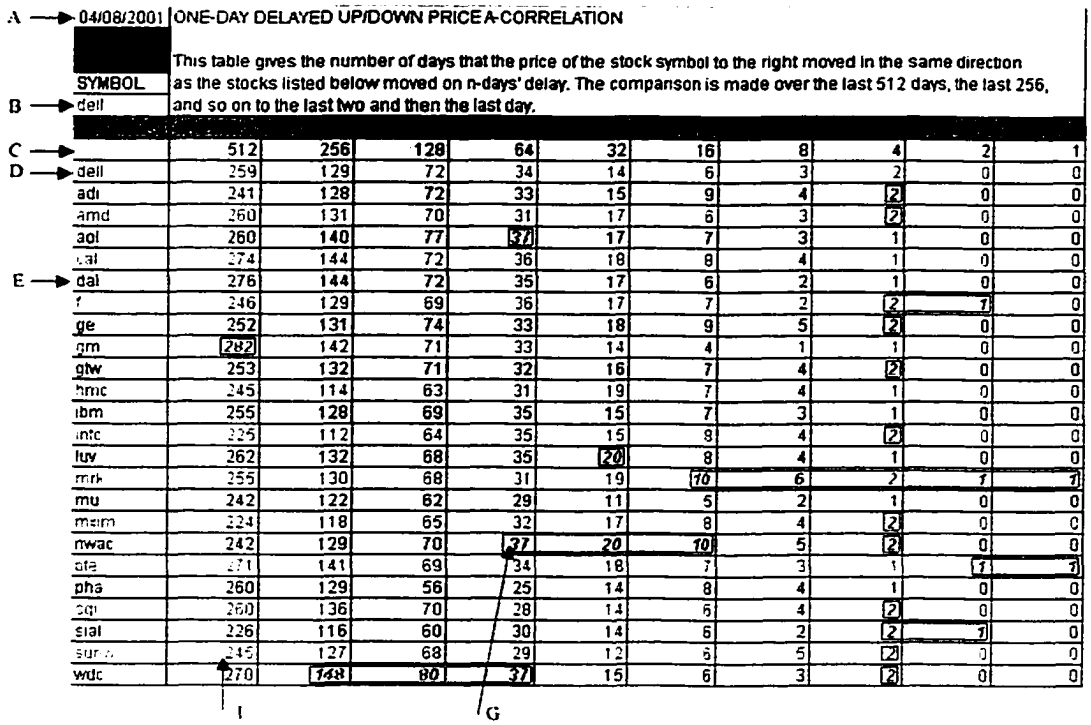
Table 3.2 Stocks overview (Time Delayed Correlation Analyzer)

Even though TM (Toyota Motor Ads.), SNE (Sony Corp. Adr.), and DIS (Disney) were removed from the list of stocks used by the previous experiments, the variety in the stock selections used for this example should still provide a reliable result.

3.2.2.1 One-day Delayed A-correlation

In this section, the sample output of the most A-correlated stocks with one-day delay is explained. The purpose of obtaining one-day delayed A-correlation is to find out which stock is the most A-correlated with the reference stock on the previous day. This is useful

as it helps to understand if there is a time-delayed relationship between the comparison stocks and the reference stock.



A → 04/08/2001 ONE-DAY DELAYED UP/DOWN PRICE A-CORRELATION

This table gives the number of days that the price of the stock symbol to the right moved in the same direction as the stocks listed below moved on n-days' delay. The comparison is made over the last 512 days, the last 256, and so on to the last two and then the last day.

B → SYMBOL: dell

C →

D →	512	256	128	64	32	16	8	4	2	1
dell	259	129	72	34	14	6	3	2	0	0
adi	241	128	72	33	15	9	4	2	0	0
amd	260	131	70	31	17	6	3	2	0	0
aol	260	140	77	37	17	7	3	1	0	0
cal	274	144	72	36	18	8	4	1	0	0
E → dal	276	144	72	35	17	6	2	1	0	0
f	246	129	69	36	17	7	2	2	1	0
ge	252	131	74	33	18	9	5	2	0	0
gm	282	142	71	33	14	4	1	1	0	0
gtw	253	132	71	32	16	7	4	2	0	0
hmc	245	114	63	31	19	7	4	1	0	0
ibm	255	128	69	35	15	7	3	1	0	0
intc	225	112	64	35	15	8	4	2	0	0
luv	262	132	68	35	20	8	4	1	0	0
mrk	255	130	68	31	19	10	6	2	1	1
mu	242	122	62	29	11	5	2	1	0	0
msm	224	118	65	32	17	8	4	2	0	0
nwac	242	129	70	37	20	10	5	2	0	0
ote	271	141	69	34	18	7	3	1	1	1
pha	260	129	56	25	14	8	4	1	0	0
sqi	260	136	70	28	14	6	4	2	0	0
stal	226	116	60	30	14	6	2	2	1	0
sun	245	127	68	29	12	6	5	2	0	0
wdc	270	148	80	37	15	6	3	2	0	0

F G

Figure 3.6 Result of one-day delayed A-correlation

The meaning of "A", "B", "C", "D", "E", "F", "G" in Figure 3.6 is the same as Figure 3.1 in Section 3.2.1.1.

The result of Figure 3.6 shows that from the last 1 day to the last 16 days, MRK has been the most A-correlated to DELL. From the last 16 days to the last 64 days, NWAC has been the most A-correlated to DELL. From the last 64 days to the last 256 days, WDC has been the most A-correlated to DELL; and during the last entire 512 days, GM was the most A-correlated to DELL.

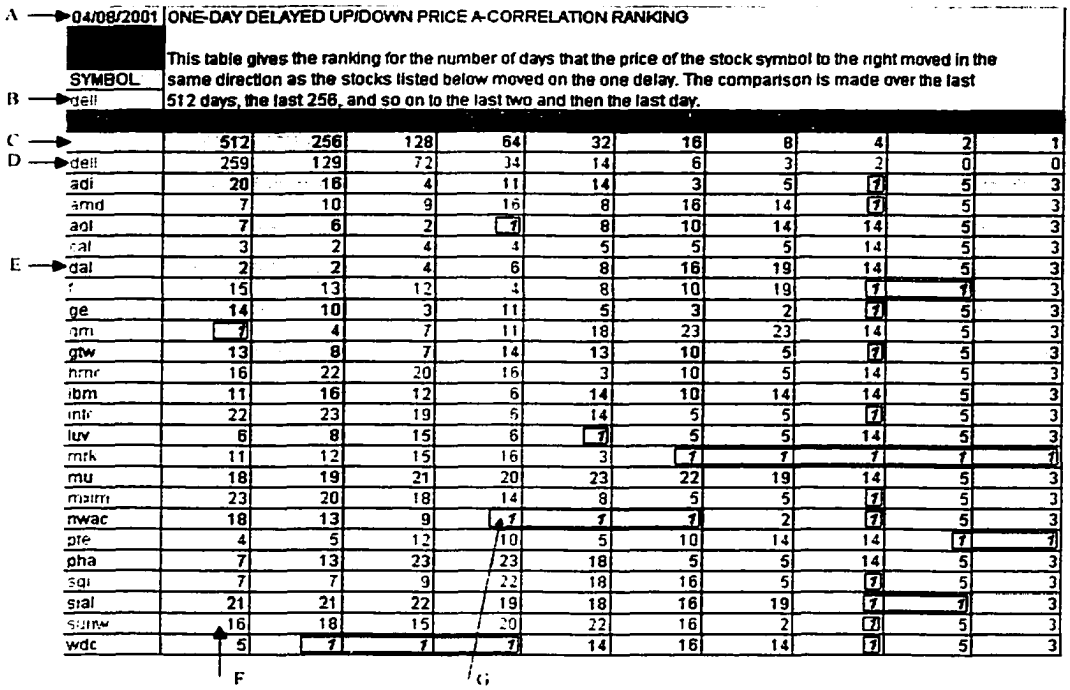


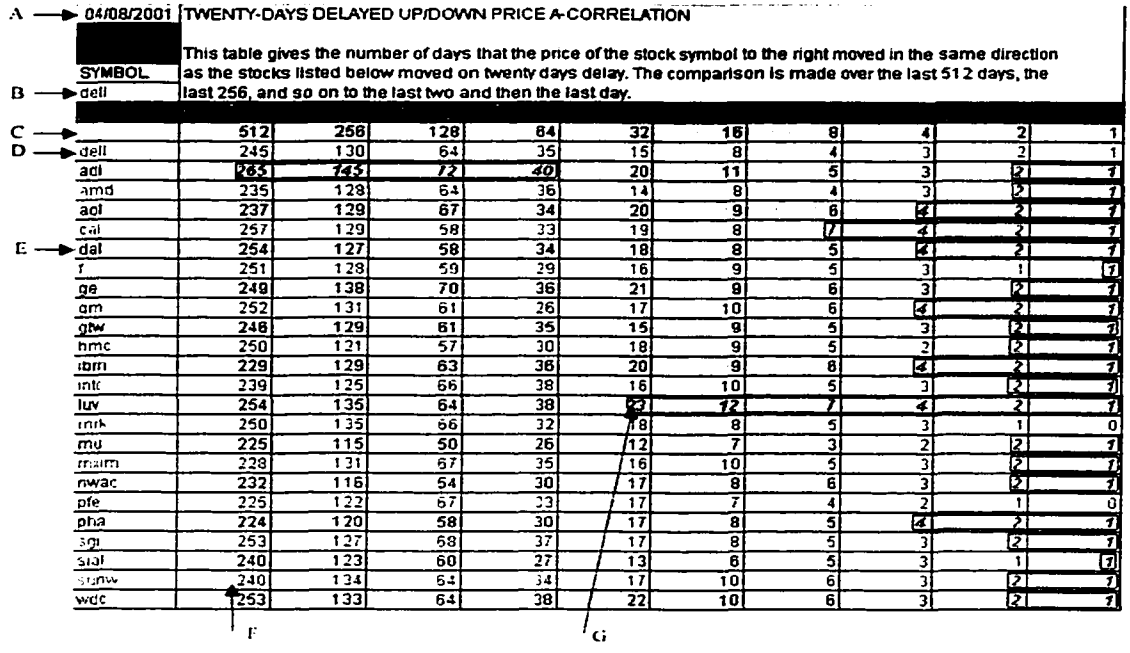
Figure 3.7 Rank result of one-day delayed A-correlation

The Figure 3.7 shows the same result in a different format. It ranks every column in table "One-day Delayed Up/Down Price A-correlation" according to which stock is most A-correlated to the reference stock.

3.2.2.2 Twenty-days Delayed A-correlation

In this section, the most A-correlated stock sample output of twenty-days delayed up/down price A-correlation is demonstrated. The purpose of obtaining twenty-days delayed A-correlation is to find out which stock is the most A-correlated with the reference stock on twenty previous days. This measures the time-delayed relationship between the comparison stocks and the reference stock within a medium term.

The meaning of "A", "B", "C", "D", "E", "F", "G" in Figure 3.8 is the same as Figure 3.1 in Section 3.2.1.1.



A → 04/08/2001 TWENTY-DAYS DELAYED UP/DOWN PRICE A-CORRELATION

This table gives the number of days that the price of the stock symbol to the right moved in the same direction as the stocks listed below moved on twenty days delay. The comparison is made over the last 512 days, the last 256, and so on to the last two and then the last day.

B → SYMBOL dell

C →

D →

E →

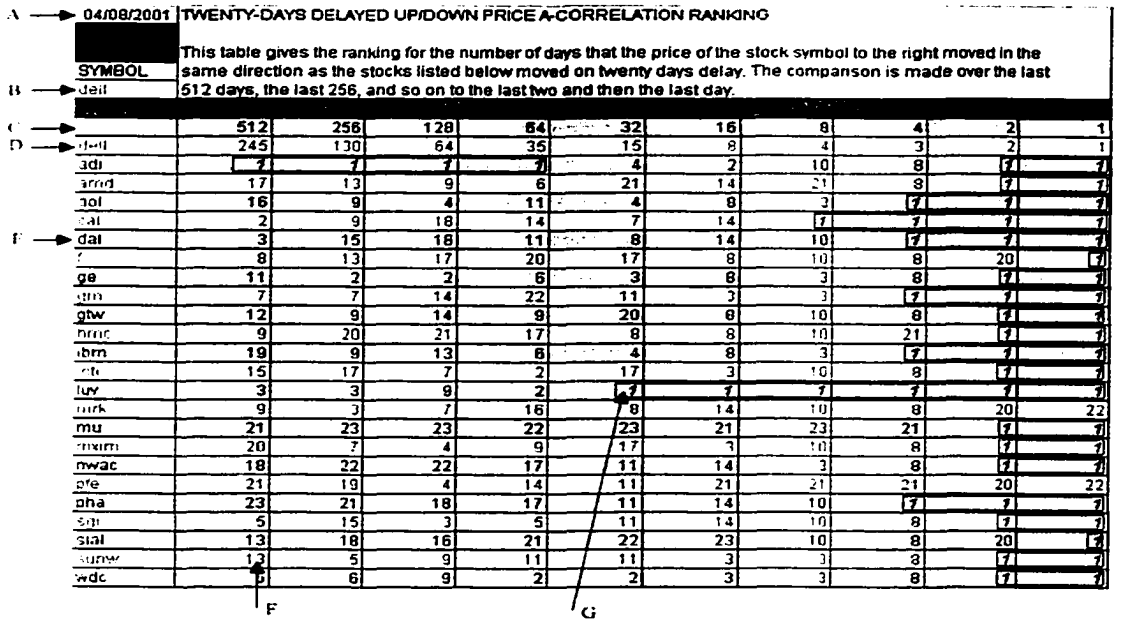
	512	256	128	64	32	16	8	4	2	1
dell	245	130	64	35	15	8	4	3	2	1
adi	265	145	72	40	20	11	5	3	2	1
amd	235	129	64	36	14	8	4	3	2	1
aol	237	129	67	34	20	9	6	4	2	1
cal	257	129	58	33	19	8	7	4	2	1
dal	254	127	58	34	18	8	5	4	2	1
f	251	128	59	29	16	9	5	3	1	1
ge	249	138	70	36	21	9	6	3	2	1
gm	252	131	61	26	17	10	6	4	2	1
gtw	246	129	61	35	15	9	5	3	2	1
hmc	250	121	57	30	18	9	5	2	2	1
ibm	229	129	63	36	20	9	6	4	2	1
intc	239	125	66	38	16	10	5	3	2	1
luv	254	135	64	38	22	12	7	4	2	1
mrk	250	135	66	32	18	8	5	3	1	0
mu	225	115	50	26	12	7	3	2	2	1
msft	228	131	67	35	16	10	5	3	2	1
nwac	232	116	54	30	17	8	6	3	2	1
pfe	225	122	67	33	17	7	4	2	1	0
pha	224	120	58	30	17	8	5	4	2	1
sgi	253	127	68	37	17	8	5	3	2	1
stal	240	123	60	27	13	8	5	3	1	1
sunw	240	134	64	34	17	10	6	3	2	1
wdc	253	133	64	38	22	10	6	3	2	1

F G

Figure 3.8 Result of twenty-days delayed A-correlation

The result of Figure 3.8 shows that from the last 1 day to 8 days, CAL has been the most A-correlated to DELL. From the last 1 day to 32 days, LUV has been the most A-correlated to DELL; and from the last 64 days to 512 days, ADI has been the most A-correlated to DELL.

The Figure 3.9 shows the same result in a different format. It ranks every column in table "Twenty-days Delayed Up/Down Price A-correlation" according to which the comparison stock is the most A-correlated to the reference stock.



A →	04/08/2001	TWENTY-DAYS DELAYED UP/DOWN PRICE A-CORRELATION RANKING									
		This table gives the ranking for the number of days that the price of the stock symbol to the right moved in the									
	SYMBOL	same direction as the stocks listed below moved on twenty days delay. The comparison is made over the last									
B →	dell	512 days, the last 256, and so on to the last two and then the last day.									
C →		512	256	128	64	32	16	8	4	2	1
D →	dell	245	130	64	35	15	8	4	3	2	1
	adi	1	1	1	1	4	2	10	8	1	1
	amd	17	13	9	6	21	14	21	8	1	1
	aol	16	9	4	11	4	8	3	1	1	1
	bal	2	9	18	14	7	14	1	1	1	1
F →	dal	3	15	18	11	8	14	10	1	1	1
	?	8	13	17	20	17	8	10	8	20	1
	ge	11	2	2	6	3	8	3	8	1	1
	gm	7	7	14	22	11	3	3	1	1	1
	gtw	12	9	14	9	20	8	10	8	1	1
	hmc	9	20	21	17	8	8	10	21	1	1
	ibm	19	9	13	6	4	8	3	1	1	1
	intc	15	17	7	2	17	3	10	8	1	1
	luv	3	3	9	2	1	1	1	1	1	1
	mrk	9	3	7	16	8	14	10	8	20	22
	mu	21	23	23	22	23	21	23	21	1	1
	msft	20	7	4	9	17	3	10	8	1	1
	nwac	18	22	22	17	11	14	3	8	1	1
	pfe	21	19	4	14	11	21	21	21	20	22
	pha	23	21	18	17	11	14	10	1	1	1
	sgp	5	15	3	5	11	14	10	8	1	1
	sial	13	18	16	21	22	23	10	8	20	1
	sunw	12	5	9	11	11	3	3	8	1	1
	wdc	6	6	9	2	2	3	3	8	1	1

F G

Figure 3.9 Rank result of twenty-days delayed A-correlation

3.3 Summary

Several conclusions have been reached from the sample output of the two programs introduced in this chapter. The programs answer the following questions:

A) For the reference stock, which comparison stock is most A-correlated over the last X days, where X is a power of 2, such as $2^0, 2^1, \ldots, 2^9$,

B) For the reference stock, which the comparison stock is becoming more A-correlated.

C) For the reference stock, which the comparison stock is becoming less A-correlated.

Chapter 4

User Manual

4.1 Correlation Analyzer (Program I)

In this section, the user manual of the **Correlation Analyzer** program is presented. The user manual contains five sections, which explains in detail how to use the program designed and implemented in this thesis.

4.1.1 Open Microsoft Excel Application

In this section, the start up of the application is explained step by step.

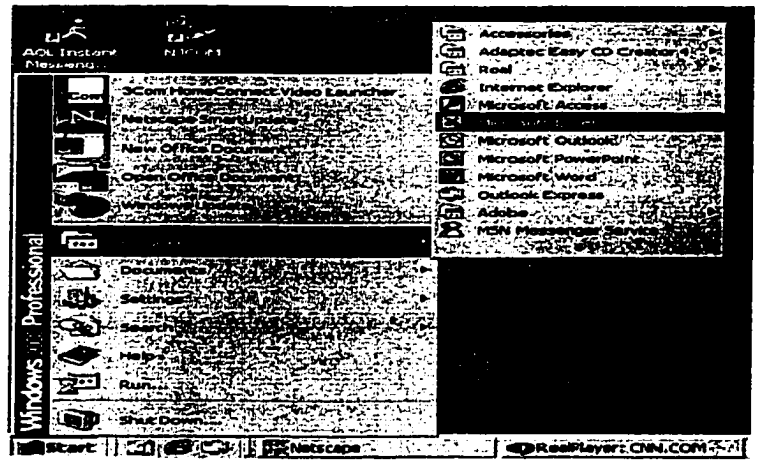


Figure 4.1 Open Excel application

Step 1: Go to **Start** button, then **Programs**. Click on the application Microsoft Excel. (Figure 4.1) Microsoft Excel is launched.

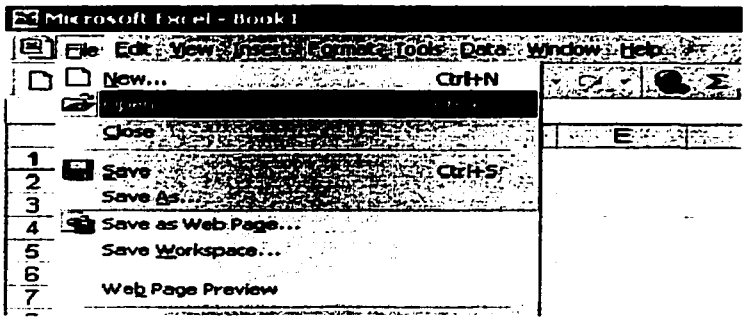


Figure 4.2 Open Correlation Analyzer program

Step 2: Click on **File** in **Menu** bar, then click on **Open** button. (Figure 4.2)

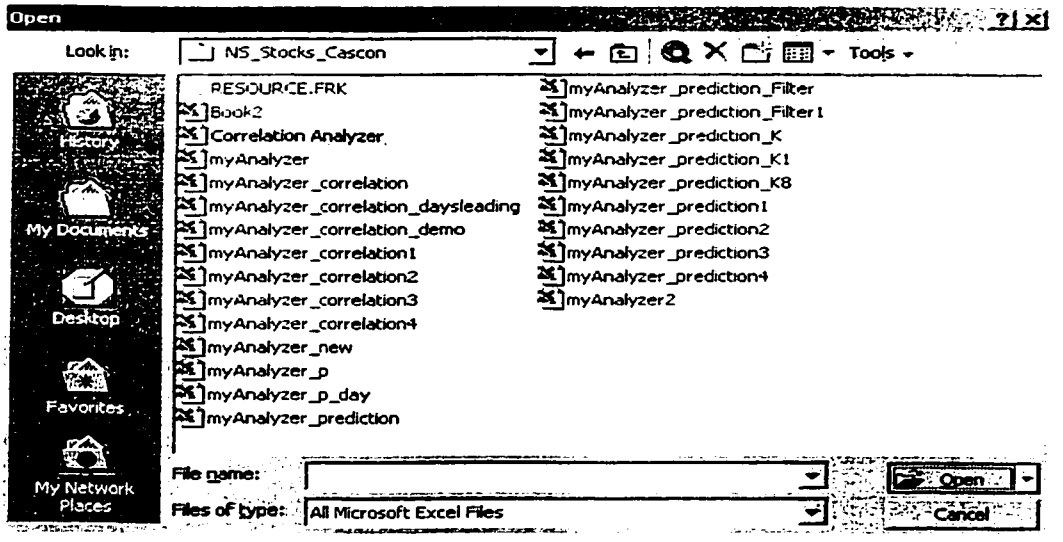


Figure 4.3 Choose Correlation Analyzer program

Step 3: Choose **Correlation Analyzer**, the Excel application, and then click on the **Open** button. (Figure 4.3) A Microsoft Excel interactive dialog appears.

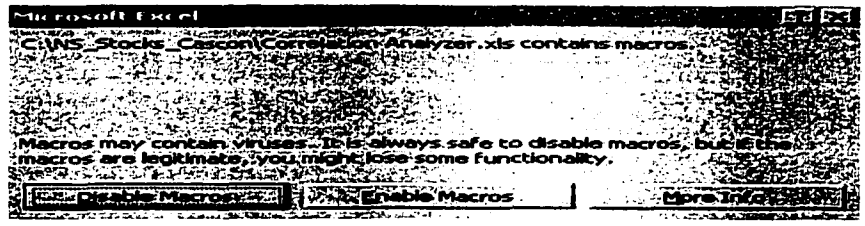


Figure 4.4 Enable Macros in opening Correlation Analyzer program

Step 4: Click on **Enable Macros** button. (Figure 4.4)

After Step 4, the Correlation Analyzer is opened.

4.1.2 Running Correlation Analyzer

In this section, the explanation of Correlation Analyzer manual continues. Following step 4, in this section, how to launch Correlation Analyzer program is shown.

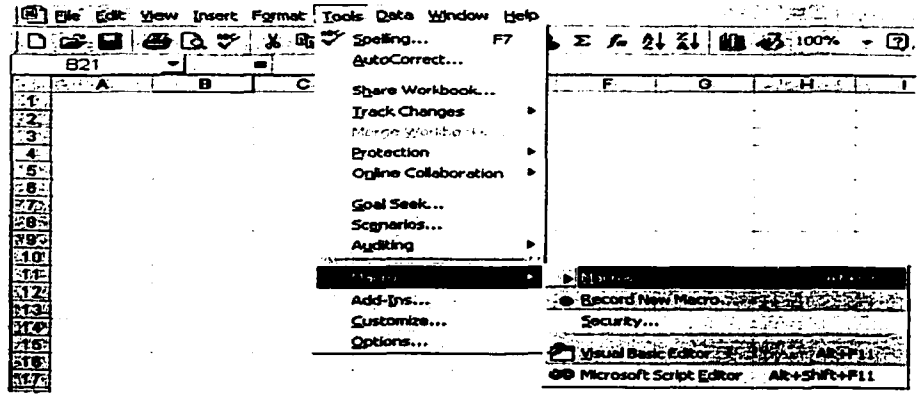


Figure 4.5 Initiate Macro command

Step 5: Go to **Tools** in **Menu** bar, then **Macro**. Click on **Macros**. (Figure 4.5), and **Macros dialog** shows up.

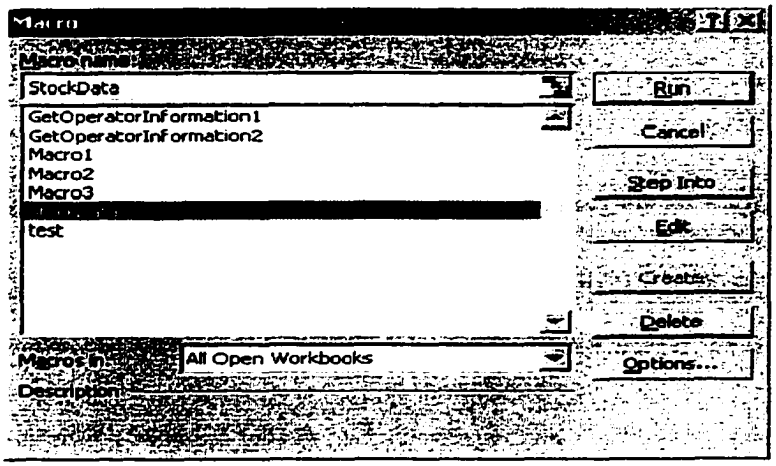


Figure 4.6 Execute Macro

Step 6: Choose **StockData** Macro from the list, then click on **Run** button. (Figure 4.6)

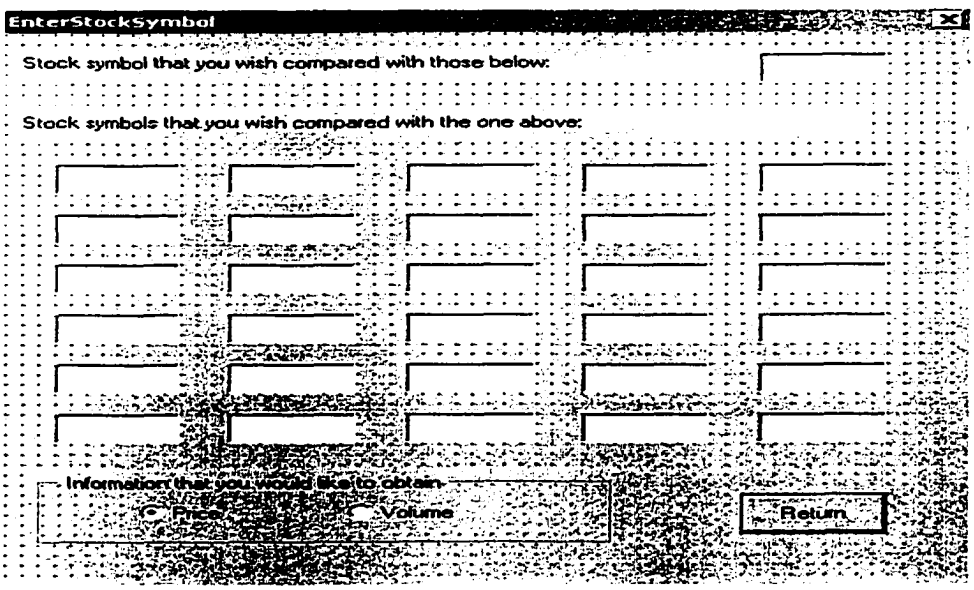


Figure 4.7 Input stock symbols

Step 7: A user dialog box appears and waits for user's input. (Figure 4.7)

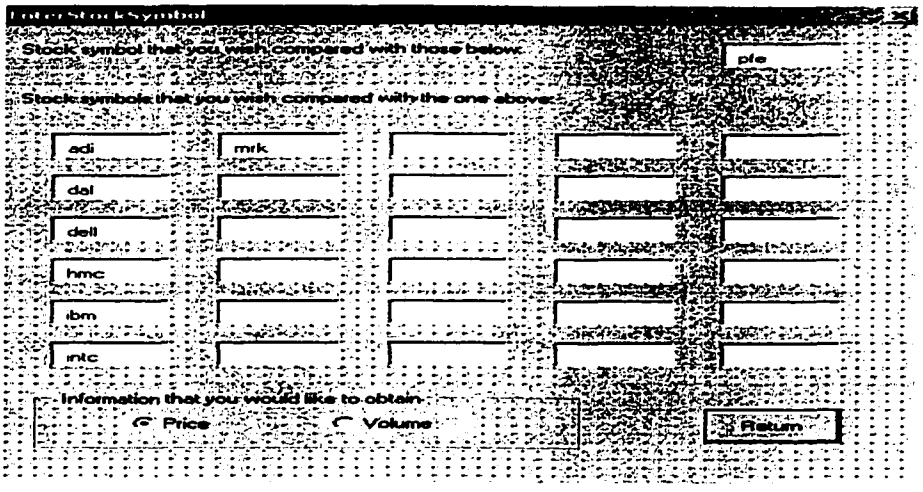


Figure 4.8 Stock symbols entered by user

Step 8: (Figure 4.8) In the right-top textbox, the user can input a stock symbol that the user is interested in comparing to comparison stocks. In the other textboxes, the user can input up to 30 stock symbols. These 30 stocks are compared to the stock entered in the right-top textbox. After entering all of the desired stock symbols, the user can click on **Return** button to start correlation program.

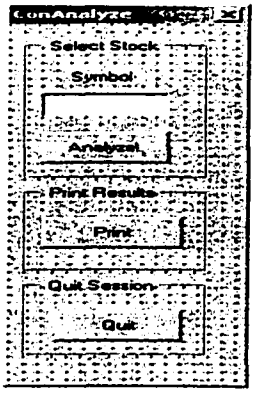


Figure 4.9 Continuing analysis dialog

Step 9: As a last step, a prompt to continue analyzing dialog shows up. (Figure 4.9) The user can click on **Quit** to view current running result. In section 4.1.4, how to analyze another stock is shown.

4.1.3 Viewing the Results

This section shows how to view the results.

03/05/2001	UP/DOWN PRICE A-CORRELATION									
SYMBOL	This table gives the number of days that the price of the stock symbol to the right moved in the same direction as the stocks listed below. The comparison is made over the last 512 days, the last 256, and so on to the last two and then the last day.									
pfe										
	512	256	128	64	32	16	8	4	2	1
pfe	504	256	128	64	32	16	8	4	2	1
adi	252	116	60	30	15	9	4	2	0	0
dal	256	133	67	35	18	9	5	3	1	0
dell	270	125	61	31	15	8	5	3	1	0
hmc	270	139	72	37	18	9	5	4	2	1
ibm	248	122	64	37	21	11	7	3	1	1
intc	272	120	63	34	17	8	4	2	0	0
mrk	343	168	84	38	17	7	3	1	1	0

Figure 4.10 A-correlation result

Step 10: By clicking on the sheet **Result**, the correlation results (Figure 4.10) shows up. The user can print out this sheet or save it to a file.

03/05/2001	UP/DOWN PRICE A-CORRELATION RANKING									
SYMBOL	This table gives the ranking for the number of days that the price of the stock symbol to the right moved in the same direction as the stocks listed below. The comparison is made over the last 512 days, the last 256, and so on to the last two and then the last day.									
pfe										
	512	256	128	64	32	16	8	4	2	1
pfe	504	256	128	64	32	16	8	4	2	1
adi	6	7	7	7	6	2	5	5	6	3
dal	5	3	3	4	2	2	2	2	2	3
dell	3	4	6	6	6	2	2	2	2	3
hmc	3	2	2	2	2	2	2	1	1	1
ibm	7	5	4	2	1	1	1	2	2	1
intc	2	6	5	5	4	2	5	5	6	3
mrk	1	1	1	1	4	7	7	7	2	3

Figure 4.11 Ranking A-correlation result

Step 11: By clicking on the sheet "Rank_Results", the ranking of A-correlation results (Figure 4.11) shows up.

03/05/2001	UP/DOWN PRICE A-CORRELATION CONVERGING TREND											
SYMBOL	This table gives the number of days that the price of the stock symbol to the right moved in the same direction as the stocks listed below. The comparison is made over the last 463 - 504 days, the last 421 - 462 days, and so on to the last 1 - 42 days.											
pfe												
	504 - 463	462 - 421	420 - 379	378 -337	336 - 295	294 - 253	252 - 211	210 - 169	168 - 127	126 - 85	84 - 43	42 - 1
pfe	42	42	42	42	42	42	42	42	42	42	42	42
adi	31	21	21	24	18	24	20	25	11	22	18	19
dal	24	21	21	21	20	19	24	22	18	19	24	23
dell	27	25	21	25	22	28	19	26	17	18	24	18
hmc	19	17	26	27	19	26	20	21	23	25	22	25
ibm	26	25	20	19	17	23	16	15	24	19	19	25
intc	27	30	21	28	24	23	22	22	13	19	22	21
mrk	32	29	26	34	29	27	31	25	28	31	28	23

Figure 4.12 A-correlation converging trend result

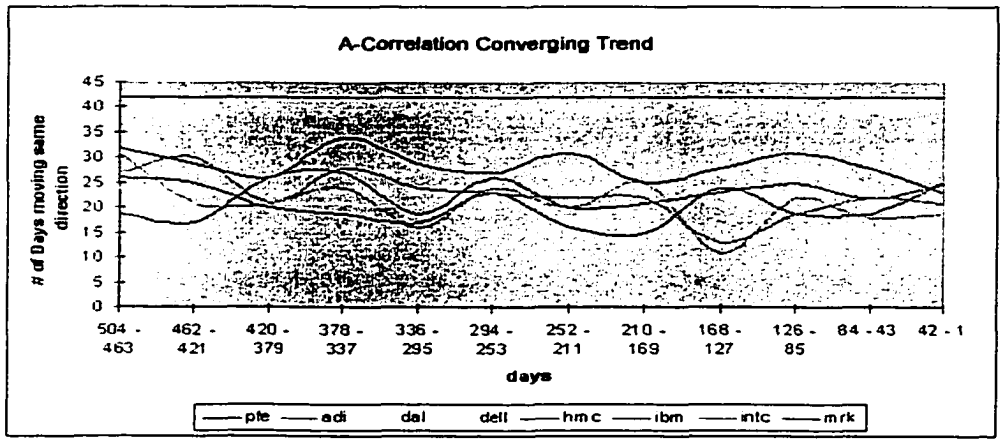


Figure 4.13 Chart of A-correlation converging trend

Step 12: By clicking on the sheet **C_Trend**, A-correlation converging trend result shows up. (Figure 4.12) Accompany with the trend result, there is a trend chart, (Figure 4.13) which shows A-correlation converging trend in a more convenient way.

03/05/2001	UP/DOWN PRICE A-CORRELATION DIVERGING TREND											
	This table gives the number of days that the price of the stock symbol to the right moved in the											
SYMBOL	same direction as the stocks listed below. The comparison is made over the last 463 - 504 days,											
pfe	the last 421 - 462 days, and so on to the last 1 - 42 days.											
	504 - 463	462 - 421	420 - 379	378 - 337	336 - 295	294 - 253	252 - 211	210 - 169	168 - 127	126 - 85	84 - 43	42 - 1
pfe	42	42	42	42	42	42	42	42	42	42	42	42
adl	31	21	21	24	***16***	24	20	25	***11***	22	***18***	19
dal	24	21	21	21	20	***19***	24	22	18	19	24	23
dell	27	25	21	25	22	28	19	26	17	***18***	24	***18***
hmc	***19***	***17***	26	27	19	26	20	21	23	25	22	25
ibm	26	25	***20***	***19***	17	23	***16***	***15***	24	19	19	25
intc	27	30	21	28	24	23	22	22	13	19	22	21
mrk	32	29	26	34	29	27	31	25	28	31	28	23

Figure 4.14 A-correlation diverging trend

Step 13: By clicking on the sheet **D_Trend**, stock A-correlation diverging trend result shows up. (Figure 4.14)

4.1.4 Analyzing Another Stock

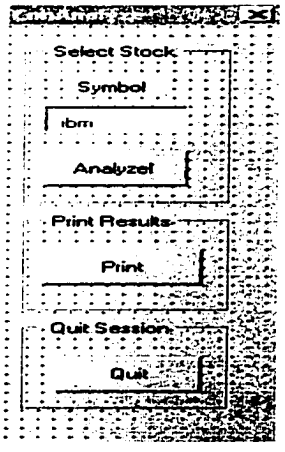


Figure 4.15 Continue to analyse stock of IBM

Step 14: To analyze another stock among those selected, the user can enter a stock symbol in **ConAnalyze**, which must be one of the comparison stocks. (Figure 4.15) Then click on **Analyze!** Button.

03/05/2001	UP/DOWN PRICE A-CORRELATION									
SYMBOL	This table gives the number of days that the price of the stock symbol to the right moved in the same direction as the stocks listed below. The comparison is made over the last 512 days, the last 256, and so on to the last two and then the last day.									
ibm										
	512	256	128	64	32	16	8	4	2	1
ibm	504	256	128	64	32	16	8	4	2	1
adi	308	156	82	41	20	11	5	3	1	0
nsl	292	143	77	36	17	9	4	2	0	0
dell	322	159	83	40	22	11	4	2	0	0
hmc	262	133	68	32	17	9	4	3	1	1
pfe	248	122	64	37	21	11	7	3	1	1
intc	322	164	85	43	24	13	5	3	1	0
mrk	257	124	52	25	14	6	2	0	0	0

Figure 4.16 A-correlation result, IBM is the selected reference stock

Step 15: An A-correlation result similar to Figure 4.11 (Figure 4.16) shows up.

4.1.5 Closing Application

In this section, it is shown how to close the application.

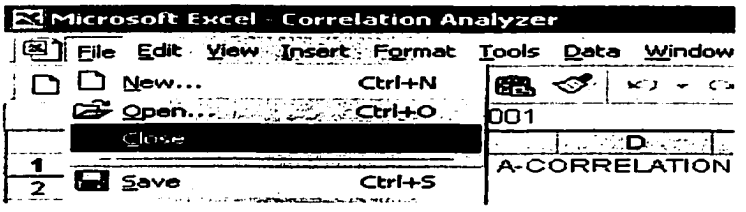


Figure 4.17 Close Correlation Analyzer program

Step 16: Go to **File** in **Menu** bar and click on **Close**. This operation closes the Correlation Analyzer program. (Figure 4.17)

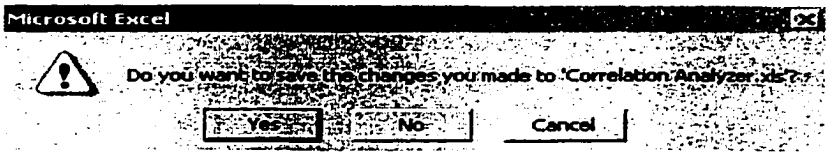


Figure 4.18 Saving results

Step 17: A Microsoft Excel dialog shows up asking whether the user wants to save the results. (Figure 4.18) By clicking **Yes** to save all the changes made to the A-correlation application. Then A-correlation application closes.

4.2 Time Delayed Correlation Analyzer (Program II)

In this section, the user manual of **Time Delayed Correlation Analyzer** is presented. Since this program is very similar to **Correlation Analyzer**, only the differences in the instructions are presented.

4.2.1 Open Microsoft Excel Application

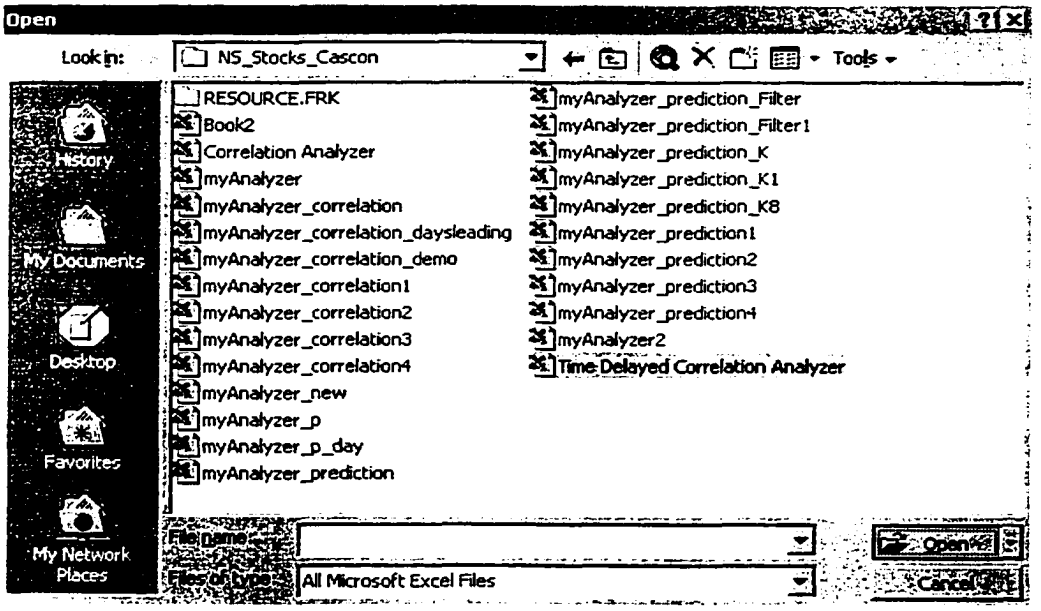


Figure 4.19 Choose Time Delayed Correlation Analyzer

Step 1 - Step 2: The same as Step 1 and Step 2 in Section 4.1.1.

Step 3: Choose **Time Delayed Correlation Analyzer**, the Excel application, then click on **Open** button. (Figure 4.19)

Step 4: The same as Step 4 in Section 4.1.1.

4.2.2 Running Time Delayed Correlation Analyzer

Step 5 ~ Step 8: The same as Step 5 ~ Step 8 in Section 4.1.2.

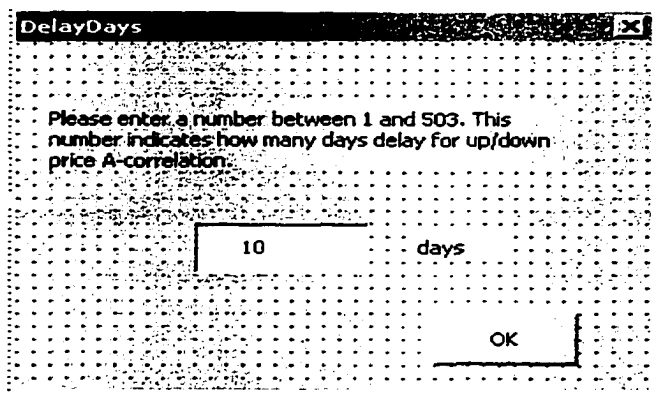


Figure 4.20 Enter number of days delay

Step 9: Figure 4.20 shows up. Enter a number between 1 and 503. This number indicates time delay between the reference stock and the comparison stocks in number of days. Then click on **OK** button.

Step 10: The same as Step 9 in Section 4.1.2.

4.2.3 Viewing the Results

04/08/2001	TEN-DAYS DELAYED UP/DOWN PRICE A-CORRELATION									
SYMBOL	This table gives the number of days that the price of the stock symbol to the right moved in the same direction as the stocks listed below moved on ten days delay. The comparison is made over the last 512 days, the last 256, and so on to the last two and then the last day.									
dell										
	512	256	128	64	32	16	8	4	2	1
dell	256	136	69	36	17	9	3	2	1	0
adi	257	133	71	33	18	8	4	2	1	0
amd	277	144	73	35	16	9	3	2	1	0
aol	254	127	62	27	14	8	5	3	1	0
cal	236	127	63	32	15	7	4	3	1	0
dal	238	131	67	31	16	7	4	3	1	0
f	251	118	62	28	14	8	4	2	0	0
ge	235	118	56	25	11	6	3	2	1	0
gm	235	119	66	35	21	9	5	3	1	0
gtw	242	127	62	32	13	8	2	2	1	0
hmc	250	127	62	31	16	6	2	1	1	0
ibm	263	129	64	31	14	8	3	3	1	0
intc	276	151	83	39	20	9	4	2	1	0
luv	230	119	67	31	13	7	4	3	1	0
mrk	259	133	73	41	22	11	6	4	2	1
mu	252	131	71	33	18	8	2	1	1	0
mxim	281	143	76	38	20	9	4	2	1	0
nwac	235	118	63	29	13	5	3	2	1	0
pfe	244	129	70	32	17	6	3	3	2	1
pha	244	128	63	35	17	7	4	3	1	0
sgi	248	127	71	40	21	11	4	2	1	0
sial	258	131	69	32	17	7	4	2	0	0
sunw	260	132	73	39	19	11	3	2	1	0
wdc	230	123	61	31	16	9	3	2	1	0

Figure 4.21 Result of time delayed up/down price A-correlation

04/08/2001	TEN-DAYS DELAYED UP/DOWN PRICE A-CORRELATION RANKING									
SYMBOL	This table gives the ranking for the number of days that the price of the stock symbol to the right moved in the same direction as the stocks listed below moved on ten days delay. The comparison is made over the last 512 days, the last 256, and so on to the last two and then the last day.									
dell										
	512	256	128	64	32	16	8	4	2	1
dell	256	136	69	36	17	9	3	2	1	0
adi	8	4	6	9	7	9	4	10	3	3
amd	1	2	3	6	12	4	14	10	3	3
aol	9	13	18	22	17	9	2	2	3	3
cal	18	13	15	11	16	14	4	2	3	3
dal	17	7	11	15	12	14	4	2	3	3
f	11	21	18	21	17	19	4	10	22	3
ge	19	21	23	23	23	19	14	10	3	3
gm	19	19	13	6	2	4	2	2	3	3
gtw	16	13	18	11	20	9	21	10	3	3
hmc	12	13	18	15	12	19	21	22	3	3
ibm	3	10	14	15	17	9	14	2	3	3
intc	2	1	1	3	4	4	4	10	3	3
luv	22	19	11	15	20	14	4	2	3	3
mrk	6	4	3	1	1	1	1	1	1	1
mu	10	7	8	9	7	9	21	22	3	3
mxim	4	3	2	5	4	4	4	10	3	3
nwac	19	21	15	20	20	23	14	10	3	3
pfe	14	11	9	11	9	19	14	2	1	1
pha	14	11	15	6	9	14	4	2	3	3
sgi	13	13	6	2	2	1	4	10	3	3
sial	7	7	10	11	9	14	4	10	22	3
sunw	5	6	3	3	6	1	14	10	3	3
wdc	22	18	22	15	12	4	14	10	3	3

Figure 4.22 Rank result of time delayed up/down price A-correlation

Step 11: This step is very similar to Step 10 in Section 4.1.3. Figure 4.21 shows the results of executing the program.

Step 12: This step is similar to Step 11 in Section 4.1.3. Figure 4.22 shows a ranked result from executing the program.

Step 13: The similar result like Figure 4.12, Figure 4.13, and Figure 4.14 shows up.

4.2.4 Analyzing Another Stock

Step 14 - Step 15: The same as Step 14 - Step 15 in Section 4.1.4

4.2.5 Closing the Application

Step 16 - Step 17: The same as Step 16 - Step 17 in Section 4.1.5.

Chapter 5

Summary

This thesis has described two stock market analysis software packages that are useful for determining the A-correlation between stocks. These packages can be seen as an application of data mining on the Internet. The CNBC web site is used as the source of the data. Using Microsoft Excel/VBA, the data is imported from the web site onto a spreadsheet. From there, the data is analysed to determine the A-correlation between the selected stocks. The A-correlation is obtained by counting the number of days that the selected stocks moved in the same direction (up/down). In this thesis, a large amount of trial runs of the two programs have been demonstrated. The sample outputs show that the programs are capable of helping investors in spotting patterns among stocks.

This entire thesis has been divided into five major chapters. In Chapter One, the basic financial knowledge such as stock market history, stock ticker, open and closing price, and volume are introduced. The basic terms used in thesis are also clarified. In addition, a literature survey is presented on the popular stock trading methods and forms of analysis. In Chapter Two, an introduction on data mining is presented. The introduction discusses the history, evolution, and popular data mining techniques. The advantages of data mining via Internet/Intranet are fully explained. Also, in this chapter, an A-correlation model is

proposed to find the basis for the programs, which have been implemented in the thesis. The A-correlation is based on the simple and basic statistical knowledge. Finally, the core of the thesis, the design and implementation of the stock A-correlation is fully discussed and explained in Chapter Two. The design and implementation starts with explanation of why Microsoft Excel is chosen as the preferred choice of the application platform. Microsoft Excel is chosen for its simplicity, powerful database and data processing ability, ease of Internet interaction and Internet data query, as well as the functionalities and objects offered by VBA. Then, the flow chart and structure of the entire programs are presented. Each designer defined function within the flow chart is shown in detail, with input and output parameters, functionality, and hierarchy fully defined in an easy to read format. Overall, Chapter Two demonstrates the motive, design, and implementation of the programs in Microsoft Excel.

Once the programs are discussed in Chapter Three, the sample output is shown and analysed to demonstrate the success and usefulness of the implemented programs. The sample output is conducted using historical data retrieved from the CNBC web site. The test subjects include a group of stocks, which belong to different industries. The programs are performed on the chosen stocks for variable amount of time. The A-correlation results are tabulated and graphed. In addition, the results are analysed to explain the behaviour of the stock A-correlation. The movement of stock A-correlation as function of time, and stock A-correlation as a function of time delay is also studied.

To facilitate the future project based on the programs implemented in this thesis, a user manual is presented in Chapter Four. The manual is written in a fashion that leads users step by step to obtain the desirable stock A-correlation information. The manual is fully graphical, and contains many hints on how to use the program efficiently.

At last, in Chapter Five, a conclusion is given and possible future improvement is also discussed.

Future improvement

In the future, it will likely be possible to offer the service of providing stock A-correlation on the Internet. This is very beneficial, as people don't need expensive software package, as demonstrated through the capable use of Microsoft Excel.

In addition, it may also be beneficial to offer growth rate A-correlation. The growth A-correlation can be a powerful tool in technical analysis. This analysis can help investors and traders to better understand growth A-correlation among stocks.

Chapter 6

Bibliography

[Aber 73] John W. Aber, "Beta Coefficients and Models of Security Return", Lexington Books, 1973

[Asog 94] Shizuo Asogawa, Norio Akamatsu, "An Efficient Algorithm for Clustering Data and Best Model Selection Using AIC", IEEE International Conference on Neural Networks, Page(s): 540 –545, vol.1, 1994.

[Azhar 94] Salman Azhar, Greg J. Badros, Arman Glodjo, Ming-Yang Kao, John H. Reif, "Data Compression Techniques for Stock Market Prediction", Proceedings of Data Compression Conference, Page(s): 72 –82, 1994

[Bell 63] Douglas H. Bellemore, "The Strategic Investor", Simmons-Boardman Publishing Corporation, New York, 1963

[Bost 98] J.R. Boston, "A Measure of Uncertainty for Stock Performance" Proceedings of the IEEE/IAFE/INFORMS 1998 Conference on Computational Intelligence for Financial Engineering (CIFEr), Page(s): 161 –164, 1998

[Bran 98] Charles H. Brandes, "Value Investing Today", McGraw-Hill, 1998

[Brea 83] R. A. Brealey, "An Introduction to Risk and Return from Common Stocks", Second Edition, The MIT Press, Cambridge, Massachusetts, 1983

[Cios 98] Krzysztof J. Cios, Witold Pedrycz, Roman W. Swiniarski, "Data Mining Methods for Knowledge Discovery", Kluwer Academic Publishers, 1998

[Crow 77] Richard A. Crowell, "Stock Market Strategy", McGraw-Hill Book Company, 1977

[Diel 80] Terry E. Dielman, "Pooled Data for Financial Markets", UMI Research Press, 1980

[Dodge 95] Mark Dodge, Chris Kinata, Craig Stinson, the Cobb Group, "Running Microsoft Excel for Windows 95", Microsoft Press, 1995

[Ellis 71] Charles D. Ellis, Allan E. Young, "The Repurchase of Common Stock", The Ronald Press Company, New York, 1971

[Erkip 98] Elza Erkip, Thomas M. Cover, "The Efficiency of Investment Information", IEEE Transactions on information theory, Vol. 44, No. 3, May 1998

[Fell 70] William Feller, Eugene Higgins, "An Introduction to Probability Theory and its Applications", Vol. I, Third Edition, John Wiley & Sons, Inc., 1970

[Garc 94] Alberto L. Garcia, " Probability and Random Processes for Electrical Engineering", 2nd Edition, May, 1994

[Ghos 96] Sabyasachi Ghoshray, "Foreign Exchange Rage Prediction by Fuzzy Inferencing on Detrministic Chaos", Proceedings of the IEEE/IAFE 1996 Conference on Computational Intelligence for Financial Engineering, Page(s): 96 –102, 1996

[Gime 99] V. Gimenez-Martinez, J. Castellanos, L. F. Mingo, "A New Tool for the Market Research using a Modified Auto-Associative Memory", Proceedings of the 1999 IEEE Signal Processing Society Workshop, Page(s): 507 –514, 1999

[Golan 95] Robert H. Golan, Wojciech Ziarko, "A Methodology for Stock Market Analysis utilizing Rough Set Theory", Proceedings of the IEEE/IAFE 1995 Computational Intelligence for Financial Engineering, Page(s): 32 -40, 1995

[Hiem 94] Ypke Hiemstra, "A Stock Market Forecasting Support System Based on Fuzzy Logic", Proceedings of the Twenty-Seventh Annual Hawaii International Conference on System Sciences, Page(s): 281 –287, IEEE 1994

[Iba 2000] Hitoshi Iba, Nikolay Nikolaev, "Genetic Programming Ploynomial Models of Financial Data Series", Proceedings of the 2000 Congress on Evolutionary Computation, Page(s): 1459 –1466, vol.2, 2000

[Ibbo 82] Roger G. Ibbotson, Rex A. Sinquefield, "Stocks, Bonds, Bills, and Inflation: The Past and The Future", The Financial Analysts Research Foundation, 1982

[Jang 91] Gia-Shuh Jang, Feipei Lai, Bor-Wei Jiang, Li-Hua Chien, "An Intelligent Trend Prediction and Reversal Recognition System Using Dual-module Neural Networks", Proceedings of first International Conference on Artificial Intelligence on Wall Street, Page(s): 42 –51, 1991

[John 96] George H. John, Peter Miller, Randy Kerber, "Stock Selection Using Rule Induction", IEEE Expert [see also IEEE Intelligent Systems], Page(s): 52 – 58, Vol 11, Issue: 5, Oct. 1996

[Jones 1997] Charles I. Jones, "Introduction to Economic Growth", Dec 1997

[Kona 2000] Prabhudev Konana, Aloysius K. Mok, chan-Gun Lee, Honguk Woo, Guangtian Liu, "Implementation and Performance Evaluation of a Real-Time E-Brokerage System", Proceedings of the 21st IEEE Real-Time Systems Symposium, Page(s): 109 –118, 2000

[Krey 72] Erwin Kreyszig, "Advanced Engineering Mathematics", John Wiley and Sons, Inc., 1972

[Lorie 73] James H. Lorie, Mary T. Hamilton, "The Stock Market – Theories and Evidence", Richard D. Irwin, Inc., 1973

[Mark 87] Harry M. Markowitz, "Mean-Variance Analysis in Portfolio Choice and Capital Markets", Basil Blackwell, 1987

[Mats 91] Ikuo Matsuba, "Neural Sequential Associator and Its Application to Stock Price Prediction", Proceedings of International Conference on Industrial Electronics, Control and Instrumentation, Page(s): 1476 –1479, vol.2, 1991

[Mcma 96] Bernard F. McMahon, "History of stock market", Columbia Publishing House, 1996

[Merr 69] Robert D. Merritt, "Financial Independence through Common Stocks", Revised and Updated Edition, Simon and Schuster, New York, 1969

[Microsoft 98] Microsoft Excel 2000 Product Enhancements Guide, 1998

[Mira 97] Fernando Gonzalez Miranda, Johan Knif, Kenneth Hogholm, "Ranked Market Information as a Stock Return Indicator", Proceedings of the IEEE/IAFE 1997 Computational Intelligence for Financial Engineering (CIFEr), Page(s): 195 –201, 1997

[Regn 98] Mireille Regnier, Wojciech Szpankowski, "On the Approximate Pattern Occurrences in a Text", Proceedings of Compression and Complexity of Sequences 1997, Page(s): 253 –264, 1998

[Rich 94] Matthew Richardson, Tom Smith, "A Unified Approach to Testing for Serial Correlation in Stock Returns", Journal of Business, Vol. 67, no.3, 1994

[Rigo 2000] Isidore Rigoutsos, "The emergence of pattern discovery techniques in computational biology", Journal of Metabolic Engineering, Vol. 2, no.3, Page(s): 159-177, July 2000

[Schi 2000] Christian Schittenkapf, Polien Tino, Georg Dorffner, "The Profitability of Trading Volatility Using Real-valued and Symbolic Models", Proceedings of the IEEE/IAFE/INFORMS 2000 Conference on Computational Intelligence for Financial Engineering (CIFEr), Page(s): 8 –11, 2000

[Schw 99] Jack D. Schwager, "Getting Started in Technical Analysis", John Wiley & Sons, Inc., 1999

[Srik 99] V. Srikanth, "Intelligent Trading Systems: a Multi-agent Hybrid Architecture", Proceedings of the IEEE/IAFE 1999 Conference on Computational Intelligence for Financial Engineering, (CIFEr), Page(s): 64 –73, 1999

[Wang 96] Jung-Hua Wang, Jia-Yann Leu, "Stock Market Trend Prediction Using ARIMA-based Neural Networks", IEEE International Conference on Neural Networks, Page(s): 2160 –2165, vol.4, 1996

[Wein 96] Neal T. Weintraub, "Tricks of the floor trader – Insider Trading Techniques for the Off-the-Floor Trader", IRWIN Professional Publishing, New York, 1996

[West 71] Richard R. West, Seha M. Tinic, "The Economics of the Stock Market", Praeger Publishers, 1971

[Wuth 98] B. Wuthrich, V. Cho, S. Leung, D. Permunetilleke, K. Sankaran, J. Zhang, W. Lam, "Daily Stock Market Forecast from Textual Web Data", 1998 IEEE International Conference on Systems, Man, and Cybernetics, Page(s): 2720 -2725 vol.3, 1998

[Zari 99] Thaleia Zariphopoulou, "Asset valuation with unhedgeable risks", Proceedings of the 38th Conference on Decision & Control, Page(s): 2794 –2798, vol.3, IEEE 1999