DESIGN AND DEVELOPMENT OF A SOFTWARE TOOL FOR ECONOMIC FACTORS ANALYSIS

SUWICHAI GULTHAMMANUSORN

A RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE (COMPUTER SCIENCE) FACULTY OF GRADUATE STUDIES MAHIDOL UNIVERSITY 2008

Research Project
Entitled

DESIGN AND DEVELOPMENT OF A SOFTWARE TOOL FOR ECONOMIC FACTORS ANALYSIS

……
Mr. Suwichai Gulthammanusorn
Candidate

……
Lect. Songsri Tangsripairoj,
Ph.D.
Major-Advisor

……
Lect. Sudsanguan Ngamsuriyaroj,
Ph.D.
Co-Advisor

……
Lect. Srisupa Palakvangsa Na Ayudhya,
Ph.D.
Co-Advisor

……
Asst. Prof. Auemphorn Mutchimwong,
Ph.D.
Acting Dean
Faculty of Graduate Studies

……
Assoc.Prof. Supachai Tangwongsan
Ph.D.
Chair
Master of Science Programme in
Computer Science
Faculty of Science

Research Project
Entitled

DESIGN AND DEVELOPMENT OF A SOFTWARE TOOL FOR ECONOMIC FACTORS ANALYSIS

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Master of Science (Computer Science)
on
October 3, 2008

……
Mr. Suwichai Gulthammanusorn
Candidate

……
Lect. Waraporn Jirapanthong,
Ph.D.
Chair

……
Lect. Songsri Tangsripairoj,
Ph.D.
Member

……
Lect. Sudsanguan Ngamsuriyaroj,
Ph.D.
Member

……
Lect. Srisupa Palakvangsa Na Ayudhya,
Ph.D.
Member

……
Asst. Prof. Auemphorn Mutchimwong,
Ph.D.
Acting Dean
Faculty of Graduate Studies
Mahidol University

……
Prof. Skorn Mongkolsuk
Ph.D.
Dean
Faculty of Science
Mahidol University

ACKNOWLEDGEMENTS

The success of this thesis is attributed to the extensive support and assistance from my major advisor, Dr. Songsri Tangsripairoj ,my co-advisors, Dr. Sudsanguan Ngamsuriyaroj, Dr. Srisupa Palakvangsa Na Ayudhya, and Dr Waraporn Jirapantthong. I deeply thank them for their valuable advice and guidance in this research.

I am grateful to all the lecturers and staff of the Computer Science Department for their valuable advice and thanks also go to my friends in the Computer Science Department class for their kind support.

Finally, I am grateful to my family for their financial support, entirely care, and love. The usefulness of this research project, I dedicate to my father, my mother and all the teachers who have taught me since my childhood.

Suwichai Gulthammanusorn

DESIGN AND DEVELOPMENT OF A SOFTWARE TOOL FOR ECONOMIC FACTORS ANALYSIS

SUWICHAI GULTHAMMANUSORN 4637268 SCCS/M

M.Sc. (COMPUTER SCIENCE)

RESEARCH PROJECT ADVISORS: SONGSRI TANGSRIPAIROJ, Ph.D., SUDSANGUAN NGAMSURIYAROJ, Ph.D., SRISUPA PALAKVANGSA NA AYUDHYA, Ph.D.

ABSTRACT

Since such economic factors as oil price, gold bar price and Stock Exchange of Thailand value have constantly fluctuated over the past several years; it is interesting to investigate the relationship of these economic factors whether or how they have an impact on one another. Though currently economic factor data are disseminated through newspapers and web sites, they have never been compiled at one central place appropriate for analysis.

This research was conducted to gather data from different web sites on 6 significant economic factors, namely, those on oil price, gold bar price, interest rate, foreign exchange rate, External Stock Market and Stock Exchange of Thailand value between January 1, 1997 and December 31, 2007. The data were stored in a data warehouse and analyzed. The economic factor analysis used a developed tool which consisted of 3 parts as follows: 1) **OLAP tool** for retrieving interesting data in different views; 2) **Time Series Analysis tool** for studying attributes of time series data such as trend, seasonal variation, cyclical variation and irregular variation, and 3) **Similarity Analysis tool** for measuring the similarity of time series data sets needing comparison.

To demonstrate the capability of the economic factors analysis tool developed, 3 experiments were conducted. Experiment I was to find the relationship between oil price and Industrial Stock index. The consequence, based on the analysis, was that oil price had the most influence on the SERVICE industry index. Experiment II was to study the relationship of gold bar price and Interest rate. The result of the analysis was that gold bar price and Interest rate were very slightly similar or had very little influence on each other. Finally, Experiment III was to find the relationship among all economic factor data by selecting representatives of data groups of the 6 factors. The outcome of the analysis could be divided into 2 groups: 1) Oil price had an influence on gold bar price data and the ENERGY Sector index data; and 2) Gold bar price had an influence on the ENERGY Sector index data and money exchange rate.

KEY WORDS: ECONOMIC FACTOR / DATA MINING / DATA WAREHOUSE / TIME SERIES / SIMILARITY ANALYSIS

110 pp.

การออกแบบและการพัฒนาโปรแกรมสำหรับการวิเคราะห์ผลกระทบทางเศรษฐกิจ
(DESIGN AND DEVELOPMENT OF A SOFTWARE TOOL FOR ECONOMIC FACTORS ANALYSIS)

สุวิชัย กุลธรรมานุสรณ์ 4637268 SCCS/M

วท.ม. (คอมพิวเตอร์)

คณะกรรมการควบคุมสารนิพนธ์ : ทรงศรี ตั้งศรีไพโรจน์ Ph.D., สุดสงวน งามสุริยโรจน์ Ph.D., ศรีสุภา ปาลกะวงศ์ ณ อยุธยา Ph.D.

บทคัดย่อ

เนื่องจากปัจจัยทางเศรษฐกิจต่างๆ เช่น ราคาน้ำมัน ราคาทองคำ ตลาดหลักทรัพย์ในประเทศ มีความผันผวนตลอดเวลาที่ผ่านมา จึงเป็นสิ่งที่น่าสนใจที่จะศึกษาถึงความสัมพันธ์ของปัจจัยทางเศรษฐกิจเหล่านี้ว่ามีผลกระทบต่อกันหรือไม่ และถึงแม้ว่าในปัจจุบันข้อมูลปัจจัยทางเศรษฐกิจ ได้มีการเผยแพร่ตามสื่อต่างๆ แต่ไม่ได้มีการรวบรวมให้เป็นแหล่งข้อมูลกลางที่เหมาะสมสำหรับทำการวิเคราะห์

งานวิจัยนี้จัดทำขึ้นเพื่อรวบรวมข้อมูลปัจจัยทางเศรษฐกิจที่สำคัญ 6 ปัจจัย ได้แก่ ราคาน้ำมัน ราคาทองคำ อัตราดอกเบี้ย อัตราแลกเปลี่ยนเงินตรา ตลาดหลักทรัพย์ต่างประเทศ และตลาดหลักทรัพย์ในประเทศ ระหว่าง 1 มกราคม 2540 ถึง 31 ธันวาคม 2550 จาก Web Sites ต่างๆ มาจัดเก็บใน Data Warehouse และได้พัฒนา Economic Factor Analysis Tool เพื่อเป็นเครื่องมือสำหรับวิเคราะห์ข้อมูล ประกอบด้วย 3 ส่วน คือ 1) OLAP Tool สำหรับเรียกดูข้อมูลที่สนใจ 2) Time Series Analysis Tool สำหรับศึกษาลักษณะของข้อมูลอนุกรมเวลา และ 3) Similarity Analysis Tool สำหรับวัดความคล้ายคลึงกันของชุดข้อมูลอนุกรมเวลา

เพื่อแสดงความสามารถของเครื่องมือ จึงได้ทำการทดลอง 3 การทดลอง คือ1) เพื่อศึกษาความสัมพันธ์ของราคาน้ำมันกับกลุ่มอุตสาหกรรมหลักทรัพย์ ผลที่ได้ ราคาน้ำมันมีอิทธิพลต่อ SERVICE Industry มากที่สุด 2) เพื่อศึกษาความสัมพันธ์ของราคาทองคำกับอัตราดอกเบี้ย ผลที่ได้ ราคาทองคำกับอัตราดอกเบี้ยมีความคล้ายคลึงกันน้อยมากๆ และ 3) เพื่อศึกษาความสัมพันธ์ของข้อมูลปัจจัยทางเศรษฐกิจ 6 ปัจจัย จากเลือกตัวแทนข้อมูล ผลที่ได้แบ่งออกเป็น2 กลุ่ม คือ 1) ราคาน้ำมันจะมีอิทธิพลต่อราคาทองคำแท่งและ ENERG Sector และ 2) ราคาทองคำแท่งมีอิทธิพลต่อ ENERG Sector และอัตราแลกเปลี่ยนเงินตรา

110 หน้า

CONTENTS

CONTENTS (Cont.)

CONTENTS (Cont.)

CONTENTS (Cont.)

LIST OF TABLES

Page

LIST OF TABLES (Cont.)

Page

LIST OF TABLES (Cont.)

Page

LIST OF FIGURES

Page

LIST OF FIGURES (Cont.)

Page

LIST OF FIGURES (Cont.)

Page

CHAPTER I
INTRODUCTION

1.1 Research Motivation

Since the economic factors in Thailand have always fluctuated, they have impacts on the country's economic system, resulting in inflation and the higher cost of living. Such factors are, for instance, global oil price, gold price rise, foreign exchange rate with stronger Thai baht compared to US dollar, decreasing domestic interest rate on deposit and vice versa on loan, including the volatility of both Stock Exchange of Thailand and External Stock Markets.

Based on the above facts, the researcher was motivated to do the research project in order to analyze the economic factors. To fulfill this objective, the data on various economic factors which could be found on the web site of the central Bank of Thailand and other web sites concerned, were compiled. The data surveyed were analyzed to find their relationship and the similarity of economic factors, as well as examine the characteristic of time series data so as to find the influence that economic factors have on one another.

1.2 Problem Statement

The economic factors in Thailand have continuously fluctuated since 1997 during which the Bank of Thailand declared the float of the baht [2][4]. Since then the Thai economy has wavered. The volatile economic factors are, for example,

1. Domestic oil prices including the prices of global crude oil and refined oil.
2. Both domestic and global prices of gold bar.
3. Interest rates of domestic commercial banks, both on deposit and loan.
4. Foreign Exchange Rate.
5. External Stock Exchange index.

6. The values of purchases and sales of securities in the Stock Exchange of Thailand (SET).

Based on the fact that economic factors data have been scattered at different data sources such as web site, causing a variety of data storage pattern, which makes it difficult to analyze those data. For this season, they are compiled and stored in the same place and the same pattern so that it is easy to have as access to and make an analysis of economic factors data.

In Addition, since economic factors data significantly differ in terms of data scale, it is hard to see the data relationship. Therefore, the data scale will have to be adjusted so that it is in the same range, for an easy analysis of data.

Similarly, a technique of Data Mining is applied with respect to Time Series Similarity Analysis in order to study the similarity of Time Series data.

1.3 Objectives

This research project has the following objectives.

1. To compile economic factors data from different web sites so as to generate a data warehouse.
2. To collect important events that had occurred and affected Thai economic during 1997 and 2007, totaling 11 years.
3. To use data from the data warehouse to create data cubes so as to be used for Online Analytical Processing (OLAP).
4. To create a software tool for economic factors analysis.
5. To analyze data by applying a technique of Data Mining with regard to Time Series Similarity Analysis with an aim to :
    4.1 Study the data similarity to use it as a parameter for measuring the influence of economic factors data on one another.
    4.2 Investigate the changing trend and tendency of economic factors data.
    4.3 Find interesting aspects or significant events that happened to economic factors data.
6. To make a report to be used to present analytical data in various forms of graphs, including data tables complementary to those graphs.

1.4 Scope of the Project

This research project has the following scopes.

1. Compile economic factors data during 1997 and 2007.
2. Compile data on domestic oil prices of Diesel, Diesel Puresila, Gasohol 91, Benzene 91, Gasohol 95 and Benzene 95 [30].
3. Compile data on global crude oil and refined oil prices quoted by WTI, Brent, Rotterdam ARA, Singapore and Dubai [8].
4. Compile data on Gold bar prices, with domestic prices in Thai baht and global prices in US dollar [13][14].
5. Compile data on Interest rates, both on deposit and loan, of 5 major commercial banks, namely, Bangkok Bank Plc, Krung Thai Bank Plc, Bank of Ayudhaya Plc, Thai Farmers Bank Plc and Siam Commercial Bank Plc. The categories of Loans and Deposits [2][19] are as follows:
    5.1 The types of interest rates for loan are MOR, MRR, MLR, overdue interest, maximum interest, and maximum and minimum interest from credit card (ceiling).
    5.2 The types of interest rates for deposit are from savings, 3-month fixed deposit, 6-month fixed deposit, 12-month fixed deposit and 24-month fixed deposit, both at minimum and maximum ceilings.
6. Compile data on Foreign Exchange rates of 35 countries, which the Bank of Thailand had announced the exchange rates [2][29].
7. Compile data on External Stock index of 10 External Stock Markets, which the Bank of Thailand had collected [2][10].
8. Compile data on purchases and sales of stock from the Stock Exchange of Thailand. They consist of Industrial group's data, Business Sector data and Securities data [34].
9. Use the principle of the technique of Time Series Similarity Analysis to do the following :
    9.1 Find similarity value between time series data by means of Similarity Analysis [15][16][17][18][22].
    9.2 Find the data attribute by means of Time Series Analysis [33][38].

1.5 Expected Benefits

The benefits to be derived from this research project include the following.

1. The analyzer knows the similarity among economic factors data.
2. The analyzer knows the trend of economic factors movement, including the data attribute that has occurred over and over in the past 11 years.
3. The analyzer knows significant events affecting economic factors.
4. The analyzer knows the relationship between oil prices and data on Industrial Stock group.
5. The analyzer knows the relationship between gold prices and data on Interest rates of deposit commercial banks.
6. The analyzer knows the relationship among economic factors from factors data representatives.

CHAPTER II
LITERATURE REVIEW

This chapter discusses details of economic factors used for the analysis, for instance, oil prices, gold bar prices, Foreign Exchange rate, interest rates for loan and deposit, External Stock exchange index, and Stock Exchange of Thailand (SET). In addition, a review is also making on details of Data Warehouse and Data Mining, OLAP, Time Series Analysis, Similarity Analysis, and a survey of software tools for economic analysis and other research works related to the analysis of oil prices, stock exchange index, foreign external rate and other economic factors.

2.1 Economic Factors

Typically, an economic system consists of numerous data on integrated factors, each of which is differently significant, for example, Foreign Exchange rate, Interest rates for loan and deposit, External Stock exchange index and Stock Exchange of Thailand data. However, in the consumer market, the factors include oil prices and gold bar prices. These data were collected between January 1, 1997 and December 31, 2007.

2.1.1 Oil Factor

Oil prices are the factor that has the most impact on an economic system since they are related to all sectors of the country's economic system [2][8][30].

Table 2.1: Types of oil distributed in the country

Type	Meaning
Diesel	Diesel oil
Gasohol 91	Gasohol 91 oil
Benzene 91	Benzene 91 oil
Gasohol 95	Gasohol 95 oil

Table 2.1 reveals types of oil distributed inside the country. The data are derived from the web site of Shell Company of Thailand, at http://www.shell.com /home/content2/Thailand-th/shell_for_motorists/price_board/historyprice_1118.html. and are in the format of Excel File [30].

Table 2.2: Categories of global crude oil and refined oil

Type	Meaning
WTI	West Texas Intermediate, USA
Brent	Brent Blend, London Brent and Brent petroleum
Rotterdam ARA	Amsterdam-Rotterdam-Antwerp
Singapore	Singapore Crude Oil Manufacturers & Suppliers
Dubai	Dubai Crude, United Arab Emirates.

Table 2.2 displays categories of both global crude oil and refined oil. The data are obtained from the web site of Energy Information Administration, at http://tonto.eia.doe.gov/dnav/pet/pet_pri_spt_s1_d.htm, and are in the format of Excel File [8].

2.1.2 Gold Factor

Apparently, gold bar prices have a certain impact on the foundation of Thai economy since gold bar are used as reserve fund for making the country's banknote [2][13][14].

Table 2.3: Type of Gold Bar price

Type	Meaning
US Gold bars	US gold bar price
Thai Gold bars	Domestic gold bar price

Table 2.3 show types of gold bar price. For the US gold bar price data (US dollar), they are received from the web site of Kitco, which is the web site compiling precious minerals data, at http://www.kitco.com/londonfix/gold_londonfix07.html, and are in the format of HTML File [13]. With respect to domestic gold bar prices data (Baht), they are gained from the web site of gold shop, at http://www.ranthong.com, which is the web site compiling the data on domestic gold bar prices, and are in the format of HTML File [14].

2.1.3 Foreign Exchange Factor

Generally, the Foreign Money exchange rate is a rate comparing a currency (such as local currency-Bath) with a principal currency (US dollar) [2][29].

Table 2.4: Type of Currencies

No	Currency	Country Name
1	AED	UNITED ARAB EMIRATES
2	AUD	AUSTRALIA
3	BDT	BANGLADESH
4	BND	BRUNEI DARUSSALAM
5	CAD	CANADA
6	CHF	SWITZERLAND
7	CNY	CHINA
8	CZK	CZECH REPUBLIC
9	DKK	DENMARK
10	EUR	EURO
11	GBP	UNITED KINGDOM
12	HKD	HONG KONG
13	IDR	INDONESIA
14	INR	INDIA
15	JPY	JAPAN
16	KES	KENYA
17	KHR	CAMBODIA
18	KRW	KOREA
19	KWD	KUWAIT
20	LAK	LAOS
21	MMK	MYANMAR
22	MXN	MEXICO
23	MYR	MALAYSIA
24	NOK	NORWAY
25	NZD	NEW ZEALAND
26	PHP	PHILIPPINES
27	PKR	PAKISTAN
28	RUB	RUSSIAN FEDERATION
29	SAR	SAUDI ARABIA
30	SEK	SWEDEN
31	SGD	SINGAPORE
32	TWD	TAIWAN
33	USD	UNITED STATES
34	VND	VIETNAM
35	ZAR	SOUTH AFRICA

The data on Foreign Exchange rate is obtained from the web site of the Bank of Thailand, at http://www.bot.or.th/Thai/Statistics/FinancialMarkets/ExchangeRate/_layouts/Application/ExchangeRate/ExchangeRate.aspx, and are in the format of Excel File consisting of 35 countries [29] as revealed in Table 2.4.

2.1.4 Interest Rate Factor

The Bank Interest rate is divided into interest rates for loan and deposit [2][19].

1. **Deposit Rate** includes :
    1.1 Interest rate for saving.
    1.2 Interest rate for 3-month, 6-month, 12-month and 24-month fixed deposit
2. **Loaning Rate** includes :
    2.1 Medium Lending Rate (MLR) or minimum loan rate means the interest rate for Loan, for a period of time, which a commercial bank changes on its good large clients.
    2.2 Minimum Overdraft Rate (MOR) means overdraft interest rate that a commercial bank changes on its good large clients.
    2.3 Minimum Retail Rate (MRR) means the interest rate for loan which a commercial bank changes on its good small clients.

The Interest rates data are obtained from the web site of the Bank of Thailand, at http://www.bot.or.th/Thai/Statistics/FinancialMarkets/InterestRate/_layouts/Application/Interest_Rate/IN_Rate.aspx. The data between 1996 and 1999 were available only for month-ends. For other years, data were available only on the day when the Bank of Thailand was open, not every day, and were in the format of Text file [19].

2.1.5 External Stock Index Factor

The External Stock index data are received from the web site of the Bank of Thailand, at http://www.bot.or.th/Thai/Statistics/FinancialMarkets/CapitalMarket/Pages/StatCapitalMarket.aspx, and are in the format of Excel file [2][10]. Table 2.5 reveals a name list of External Stock exchanges compiled.

Table 2.5: Name list of External Stock Exchange [10]

Market Name	Country Name
Set Index	Thailand
Dow Jones	NewYork
Financial Times	London
Nikkei	Japan
Hang Seng	Hong Kong
Straits Times	Singapore
KLSE Composite	Malaysia
JSX Composite	Indonesia
PSE Composite	Philippines
KSE Composite	South Korea
ALL Ordinaries	Australia

2.1.6 Stock Exchange Factor

The Stock Exchange of Thailand (SET) is the place where the capitals of listed companies are mobilized, and where investors are interested to invest in stock instead of depositing their money in the bank or starting the business of their own [34].

Currently, the Stock Exchange of Thailand has publicized 8 Industries Stock group of principal industrial indices and 25 Sectors of Businesses Stock [34], as shown in Table 2.6.

Table 2.6: Structure of prices index in the Stock Exchange of Thailand [34]

No	Industry	Sector	Name
1.	Food and Agro (.AGRO)	Agro-Business	.AGRI
		Food and beverage	.FOOD
2.	Consumer goods (.CONSUMP)	Fashion	.FASHION
		Household and office articles	.HOME
		Personal effects and medical supplies	.PERSON
3.	Finance (.FINCIAL)	Banking	.BANK
		Capital and stock	.FIN
		Non-life and life insurance	.INSUR
4.	Industrial goods (.INDUS)	Motor vehicle	.AUTO
		Industrial material and Machinery	.IMM
		Packaging	.PKG
		Paper and printing material	.PAPER
		Petrochemistry and chemicals	.PETRO
5.	Property and construction (.PROPCON)	Construction materials	.CONMAT
		Property development	.PROP
6.	Resources (.RESOURC)	Energy and infrastructure	.ENERG
		Mining	.MINE
7.	Services (.SERVICE)	Commerce	.COMM
		Medication	.HELTH
		Media and publication	.MEDIA
		Special service	.PROF
		Tourism and recreation	.TOURISM
		Transport and logistics	.TRANS
8.	Technology (.TECH)	Electronic parts	.ETRON
		Information and communication technology	.ICT

2.2 Data Warehouse, OLAP and Data Mining

Data Warehouse is a system for the collection of data from different sources at a central place in order to be used for accurate and efficient data analysis [20][25].

2.2.1 Data Warehouse Characteristics

The properties of Data Warehouse consist of [25].

1. **Subject-Oriented**: Only the data to be used for analysis or decision making are stored. An emphasis is not put on the performance or process of database.
2. **Integration**: It is a compilation of data from various sources, in the format of database or file. Then, the data compiled are standardized while the parameter of data variable is defined to be in the same pattern.
3. **Time Variance**: The time period of data compilation will have to be determined since there has to be a comparison of data and time period for the analysis or decision making of data. This aims to know the time of occurrence of data.
4. **Nonvolatile**: The stability of data stored in the Data Warehouse must not be altered, whether it is an addition or modification of data. A user can have an access to data only.

2.2.2 Data Warehouse Architecture

Figure 2.1 illustrates the Data Warehouse architecture [20], which consists of 4 following parts.

1. **Data Source or Raw Data**: It is raw data collected from various sources through the ETL (Extract Transform Loading) process, which is a verification of data before being stored in the Data Warehouse.
2. **Data Warehouse Server**: It is the part storing summarized data derived from data processing, which emerges from daily routine or other data sources.
3. **OLAP Server or OLAP Engine**: It serves to retrieve data from Data Cube, generated and stored in the Data Warehouse, to client or application program.
4. **Front-End Tool or Application Program**: It is the part for displaying data, derived from the Data Warehouse through OLAP engine, in the form of report or graph.

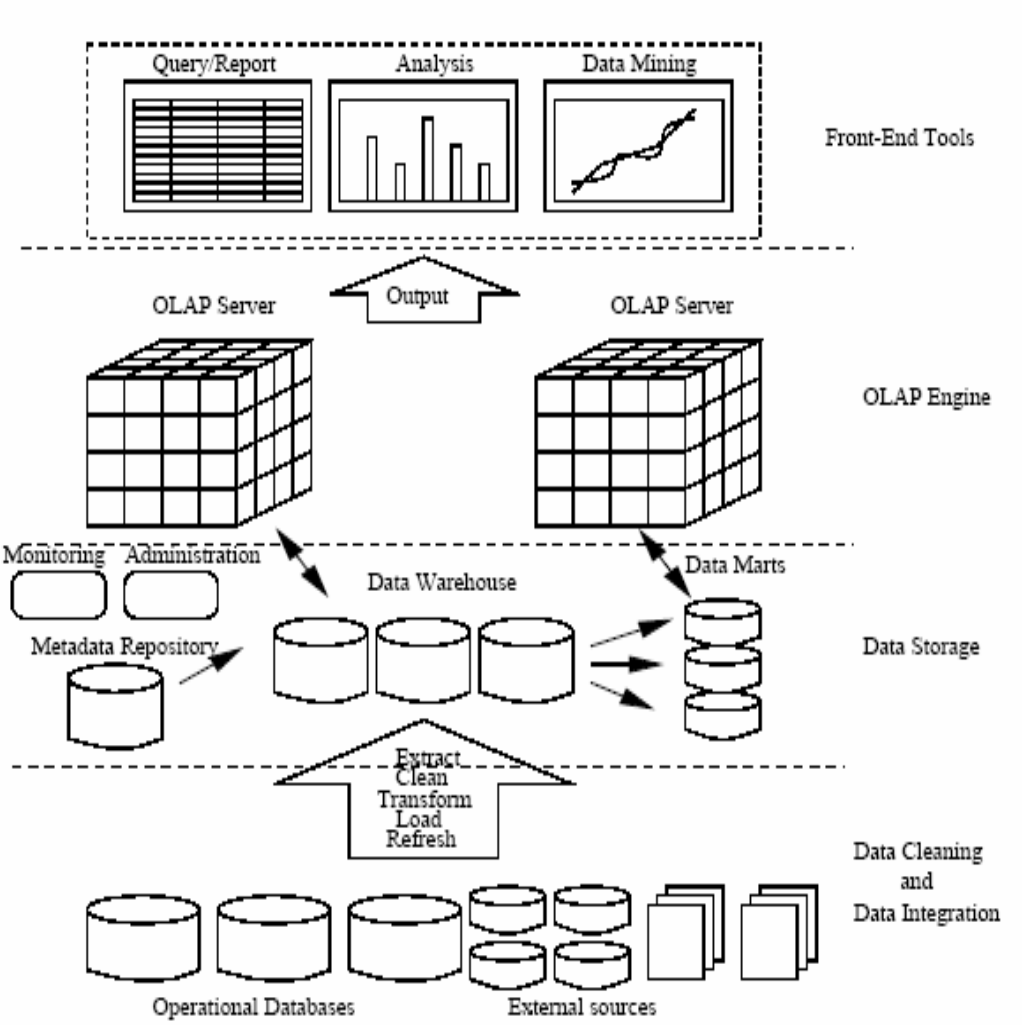


Figure 2.1: Data Warehouse Architecture [20].

2.2.3 Data Warehouse Design

The Data Warehouse stores data in the format of multidimensional data, which are stored in Fact table and Dimension table] [25].

1. **Fact Table** is a table recording numerical data to be analyzes. It is used to make a summary report, consisting of key connected with Dimension table, and numerical data.
2. **Dimension Table** is a table recording detail data complementary to a summary report, comprising key and attributes.

They are 2 popular patterns of the design of Data Warehouse schema [25], which reveals the relationship between the Fact table and Dimension table.

1. **Star Schema**: There is a Fact table as the center surrounded by the Dimension table, with keys connected with the Fact table, as shown in Figure 2.2.

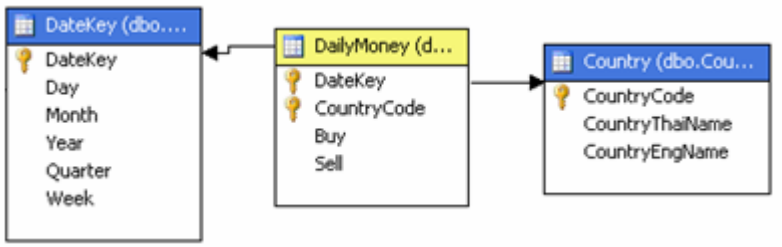


Figure 2.2: Star Schema

2. **Snowflake Schema**: It is similar to star schema but with several levels of Dimension table, and key connected with other Dimension tables, as displayed in Figure 2.3.

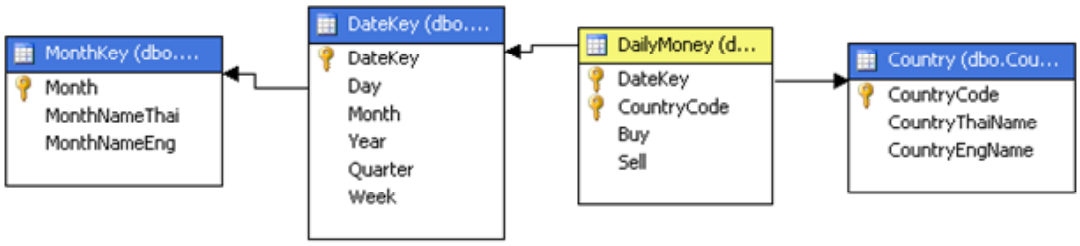


Figure 2.3: Snowflake Schema

2.2.4 ETL-Extract Transform Loading

It is the stage of transforming raw data into the data to be analyzed. Subsequently, they are stored in the Data Warehouse. Stages of Data Warehouse generation include the following [20][25].

1. **Extract**: It is a process of selecting data beneficial for the analysis, based on the subject (Subject-Oriented). This process includes
    1.1 **Data Cleaning** is an elimination of abnormal or incomplete data.
    1.2 **Data Integration** is a compilation of data in the same or approximate pattern so that they become the same set of data.
2. **Transform**: It is data transformation to be in the format ready for the analysis in accordance with the principle of Data Mining. The transformation process includes

2.1 **Data Selection** is the identification of data attribute that is pertinent to the defined purpose and then such data are selected.

2.2 **Aggregation** is a compilation of related data such as the sum of data.

3. **Load**: It is a process that serves to bring data into the system of Data Warehouse.

2.2.5 OLAP (On-Line Analytical Processing)

It is a tool used to extract data from Data Cube (Multidimensional data) in the Data Warehouse for the analysis. It is done by calculating in advance the sum that is frequently used. Thus, when the analysis needs to be made, the data can be easily and swiftly extracted [20][25].

The data management structure of OLAP consists of 3 following types [20][25].

1. **ROLAP (Relational On-Line Analytical Processing):** It is the management of relational database by using query processing, through the SQL (Structured Query Language). Data will be stored in the format of index structure in the Data Warehouse. The data of Fact table and Dimension table are also stored in relational database.
2. **MOLAP (Multidimensional On-Line Analytical Processing):** It is the management of database stored in the format of multidimensional structure. Data will be stored in the format of multidimensional sparse arrays, using index (cell) and hashing as data reference. Besides, in MOLAP, there is a command called Aggregate Functional, which is used for computation. For instance, the SUM command is used to calculate the sum of data. The consequence derived from the Aggregate functional command is stored in the measure of Fact table (which is relational database).
3. **HOLAP (Hybrid On-Line Analytical Processing):** It is a mixed structure of MOLAP and ROLAP. MOLAP is used to manage the process of high-level summary data which ROLAP is used to manage detail data.

OLAP Operations are basis used to extract data from Data Cube. Display of results derived from OLAP query [20][25] is, for example,

1. **Roll Up** is the stage of data display from detail data to summary data.
2. **Drill Down** is the stage of data display from summary data to detail data.
3. **Slice and Dice** is the stage of selecting to display only part of data or interesting dimension, which is the filler of data to be displayed.
4. **Pivot or Rotate** is the stage of changing the view data dimension from X axis to Y axis.

2.2.6 Data Mining

It is a process of searching for interesting information, for instance, relationship pattern, outstanding structure or abnormal attribute of data, all from a large amount of data stored in database [20][21][24[25][27].

Data Mining is classified into 2 types [21][25], namely,

1. **Descriptive Data Mining** is to find a model or knowledge in order to explain some attributes of data. Normally, it is to find relationship rule, and categorization or grouping of data.
2. **Predictive Data Mining** is the prediction of attributes or estimate of the value of data based on previous data.

There is a myriad of data mining techniques to solve the problems. Here some examples of widely used techniques are given.

1. **Association Rule** is the analysis to find the relationship rule from an enormous amount of existing data. Then the defined rule is used to explain the future emerging data, such as Market Basket Analysis [25].
2. **Classification** is the categorization or grouping of data by surveying outstanding attribute of data that occur and using those attribute for data classification [29]. When there are new data coming in, they can be classified based on the outstanding attribute defined [25].
3. **Clustering** is to cluster data. During the clustering, it is not known what the emerging data groups are composed of until the data search is complete [25].

4. **Time Series Analysis** is the analysis to find the attribute of time series data so as to see the trend of current data and data which will occur in the future. Similarly, it is to find the attribute of data that happen repetitiously over one year, or the movement corresponding with data cycle, or to find the attribute of abnormal data [21].

2.3 Data Normalization

It is an adjustment of data size to be in the same standard since the data to be analyzed are typically of different size. There are various techniques of Data Normalization. In this research 4 techniques are proposed.

2.3.1 Previous Data Change

It is the rate of data change compared to the previous data, and is an adjustment of data size and position to be in standard [17], through Equation 2.1.

$$R_t = \frac{X_t - X_{t-1}}{X_{t-1}} x100 \qquad 2.1$$

where R_t is the value of change rate at time t.

X_t is the value of data at time t.

X_{t-1} is the value of data at time t-1.

2.3.2 Moving Average Ratio

It is time-dependent data [27], which is a method of finding Moving Average Ratio of data according to the same time period, such as to find average ratio every 5 days, from all data in 365 days. It is an adjustment of size and rearrangement of position of data to be in the same range [27][28], through Equation 2.2 and 2.3.

$$MA(m) = \frac{\sum_{j=t-m+1}^{t} X_m}{m} \qquad 2.2$$

$$M_t = \frac{X_t - MA(m)}{MA(m)} x100 \qquad 2.3$$

where M_t is the average value of data adjusted at time t.

X_m is data in the time period t-m+1.

X_t is the value of data at time t-1.

$MA(m)$ is average ratio of data at time t.-m+1

m is time span needed.

Both Previous Data Change and Moving Average Ratio are appropriate for data that are slightly changed [27].

2.3.3 Standard Deviation Normalization

It is Z-Value normalization, which is a method of comparing the values of data to see how they differ. It is an adjustment of size and rearrangement of position of data to be in standard, using average ratio and standard deviation value for a calculation [25][27], through Equation 2.4.

$$\sigma = \sqrt{\frac{x_t - \overline{X}}{n}} \quad , Z_{xt} = \frac{x_t - \overline{X}}{\sigma(x)} \tag{2.4}$$

where Z_t is the average value of data adjusted at time t.

X_t is the value of data at time t-1.

$\overline{X}$ is average ratio of all data.

σ is standard deviation value of all data.

2.3.4 Min-Max Normalization

It is an approach of finding the value from maximum and minimum data range, and is an adjustment of size and rearrangement of position of data to be is standard [27] through Equation 2.5.

$$M_t = \frac{X_t - Min(X)}{Max(X) - Min(X)} \tag{2.5}$$

where M_t is the value of data adjusted at time t.

X_t is the value of data at time t.

Min is minimum value of all data.

Max is maximum value of all data.

2.4 Time Series Analysis

Time Series data means data occurring in time sequence, with the same time distance and continuous occurrence, such as daily oil price, The analysis will be make on the components of time series data, which consist of 4 patterns, namely, Trend, Seasonal Variation, Cyclical Variation and Irregular Variation [7][24][26][27][28][33][38][39].

2.4.1 Trend

The value of secular trend indicates the long-time or trend movement of data to observe that the alteration trend is linear (when data have a constant change rate) or curved (when data have an inconstant change rate) [33][38][39]. The calculation is made via Equations 2.6, 2.7 and 2.8.

Trend finding uses a technique of Moving Average Ratio in each time span [33].

$$MA = \frac{X_1 + X_2 + ... + X_n}{n}, \frac{X_2 + X_3 + ... + X_{n+1}}{n}, \frac{X_3 + X_4 + ... + X_{n+2}}{n}, ... \quad 2.6$$

$$Trend = MA(1), MA(2), MA(3), ... \quad 2.7$$

$$Center.Trend = \frac{MA(1) + MA(2)}{2}, \frac{MA(2) + MA(3)}{2}, \frac{MA(3) + MA(4)}{2}, ... \quad 2.8$$

2.4.2 Seasonal Variation

Seasonal Variation means seasonal alteration which occurs repetitiously over 1 year until it becomes the same pattern [33][38][39].

The analysis aims to find Seasonal Index to see how a season influences data. The values obtained are relational quantity or percentage. The calculation applies the ratio to simple average method. The values received are in percentage of monthly or quarterly average value [33].

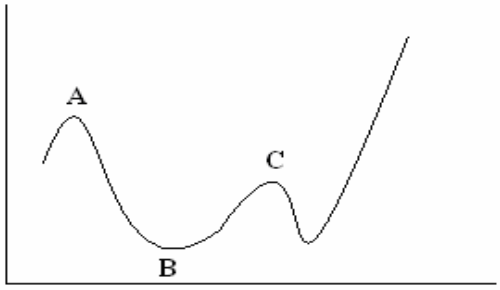
The ratio to Simple Average method [33] consists of the following stages.

1. Calculate the Average value of data in each year.
2. Divide the value derived from item 1 by monthly or quarterly seasonal data in that particular year, which is called specific Seasonal value.

3. Compute the specific Seasonal average value, which is called the approximate Seasonal Index value.
4. Adjust the value received from item 3 to become Seasonal Index value by multiplying it with 1,200 in case of monthly value, or with 400 in case of quarterly value, and then dividing it by the sum of approximate Seasonal Index values.

2.4.3 Cyclical Variation

Cyclical Variation means cyclical movement (such as business cycle). Typically, the feature of a cycle comprises a period showing prosperity, decline or recession, standstill or depression, and recovery. Each cycle of occurrences is not constant but varies in time and broadness range, as illustrated in Figure 2.4 [33].



No	Point	Phase
1	A	Prosperity
2	B	Depression
3	BC	Recovery
4	AB	Recession

Figure 2.4: Cyclical Phase

To calculate Cyclical Variation value [33], divide the Trend value by Seasonal Index value, through the following stage.

1. Calculate the Average value of data over 1 year.
2. Calculate the Trend value by Equation 2.8 in section 2.4.1.
3. Calculate the Seasonal Index value through the stage mentioned in section 2.4.2, and because the Seasonal Index value is in percentage, then divide it by 100.
4. Divide the trend value by the Seasonal Index value through Equation 2.9.

$$Cycle = \frac{Trend * 100}{SeasonalIndex}, \qquad 2.9$$

5. Use the Cyclical Variation value to calculate the Average value by weighting in the ratio of 1:2:1, with 3 values each on the average, in order to eliminate variation due to an abnormal occurrence (Irregular Variation).

2.4.4 Irregular Variation

Irregular variation is variable and unpredictable, for example, natural disasters, wars, work strikes, etc.

In practice, when there is the analysis of Time Series, Irregular Variation is generally not analyzed. Additionally, Irregular Variation can be viewed from a distribution chart of data if a data value is much higher or lower than a normal one [33].

2.5 Similarity Analysis

It is a method of measuring similarity of time series data and grouping similar data, such as the grouping of stocks having the same movement of prices, or for data classification [11][15][16][17][18][37].

Similarity analysis consists of approaches of similarity distance measurement and Regression and Correlation, using Simple Linear Regression equation. The similarity value obtained will be an indicator telling whether the 2 set of time series data have an approximate or similar attribute [15][16][17][18][37].

2.5.1 Similarity Distance

It is a method of measuring the distance between 2 sets of time series data, through a comparison of distance between points of data. These methods are, for instance, Euclidean distance and Dynamic Time Warping (DTW) [15][16][17][18].

2.5.1.1 Euclidean Distance

It is an approach of measuring the distance between points of data. In this regard, there are 2 sets of time series data, with the same number, namely, $X = \{x_1,.....,x_n\}$ and $Y = \{y_1,.....,y_n\}$. Then the distance between point X and point Y is computed through Equation 2.10.

$$d(i,j) = \sqrt{\sum_{i=1,j=1}^{n} | X_{i,n} - Y_{j,n} |^{2}} \quad 2.10$$

For similarity distance measurement, the value obtained from the sum of distances of all points is divided by the number of data (N), via Equation 2.11.

$$Sim(X,Y) = \frac{\sum_{i=1,j=1}^{n} d(i,j)}{N} \quad 2.11$$

In case that there are more than 2 sets of time series data, the Similarity distance measurement uses the least value of similarity set, which is the most similarity pair, through Equation 2.12.

$$Similarity = Min\left[Sim(X,Y), Sim(X,Z), Sim(Z,Y)\right] \quad 2.12$$

2.5.1.2 Dynamic Time Warping (DTW)

Dynamic Time Warping is a method of calculating data similarity, normally used in Signatures Similarity or Speed Recognition, and with 2 sets of time series data; $X = \{x_1,.....,x_m\}$ and $Y = \{y_1,.....,y_n\}$. By DTW method, a table of n-by-m matrix will have to be generated, with data in each cell being the distance between points x_i and y_j ($D(x_i, y_j)$), and calculated via an approach of Euclidean distance or Absolute distance. Subsequently, find the line passing through the points diagonally across the table, starting from the point at cell (1, 1) to the point at cell (n, m), called “Warping Path”. Later, a technique of dynamic programming is used, via Equation 2.13, to compute all possible values on the Warping Path, as illustrated in Figure 2.5 [15][16][17][18].

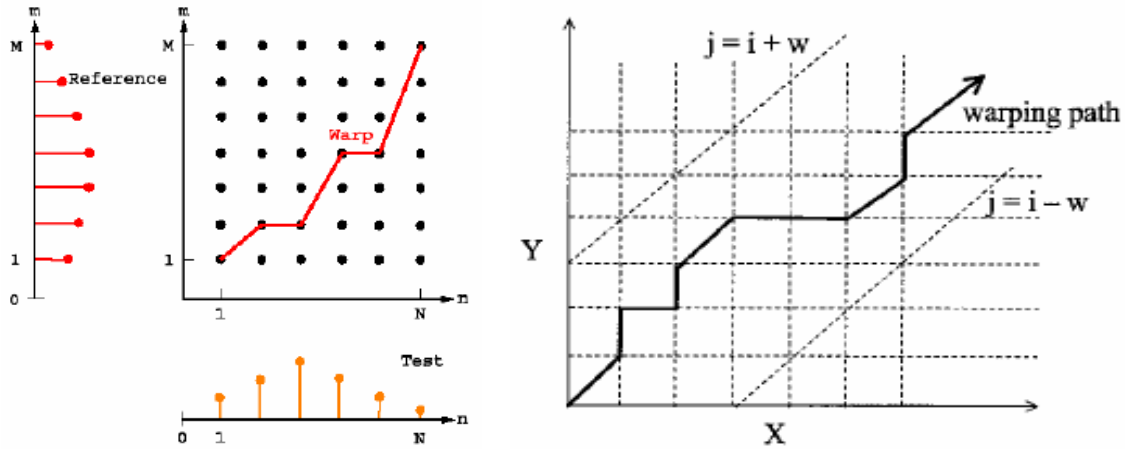


Figure 2.5: Dynamic Time Warping Path

The measurement of distance of $X = \{x_1,.....,x_i\}$ and $Y = \{y_1,.....,y_j\}$ is called the Cost of sequence X and sequence Y, and uses a technique of Absolute distance, through Equation 2.13.

$$Cost(i,j) = \left|x_i - y_j\right|^2 \qquad 1 \le i,j \le n \qquad 2.13$$

Dynamic Programming is a technique of calculating the least value of the sum of distances of point on the Warping Path, between point i to X axis and point j to Y axis, through Equation 2.14.

$$DTW(i,j) = \begin{cases} 0 & f(i,j) = 0 \\ Cost(i,j) + D(i,j-1) & f(i=0).and.(j>0) \\ Cost(i,j) + D(i-1,j) & f(i>0).and.(j=0) \\ Cost(i,j) + Min\{D(i-1,j), D(i-1,j-1), D(i,j-1)\} & if\,(i>0).and.(j>0) \end{cases} \qquad 2.14$$

To calculate DTW, the value called "Warping Windows (w)" will be determined, which is to define the scope of points to be computed on the Warping Path, with the condition being $|i-j| \le w$.

To measure the similarity value, the value of similarity set obtained from the sum of distances on the Warping Path is divided by warping length, via Equation 2.15.

$$Sim(X,Y) = \frac{\sum_{i=1,j=1}^{n} DTW(i,j)}{Warping\_Length} \quad 2.15$$

In case there are more than 2 sets of time series data, similarity measurement will use the least value of similarity set which is the most similar pair, via Equation 2.16.

$$Similarity = Min\left[Sim(X,Y), Sim(X,Z), Sim(Z,Y)\right] \quad 2.16$$

Euclidean and DTW are similarity measure with relative meaning [29]. That is the similarity distance cannot give answer without comparing it with other distances. Euclidean is distance measure used to determine the alignment between 2 same data sequence length and DTW can be used in 2 difference data sequence length.

2.5.2 Regression and Correlation

It is the analysis of relationship between independent variables and dependent variables to find whether or how they are related or not related, or whether they go in the same or opposite direction, using the value of Coefficient of Correlation, which indicates the level of relationship. In this connection, the values obtained are between -1 and 1, with maximum value being 1, which indicates highest relationship (same direction), and minimum value being -1, which points out least relationship (opposite direction). If the value is 0, it means that the variables have no relationship [16][17][18]. The calculation is made through Equation 2.17.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad 2.17$$

where x_i, y_i are any values of each variable that are matched.

$\bar{x}, \bar{y}$ are average values of each variable.

2.5.2.1 R-Square

R-Square or “Coefficient of Determination” is a technique of Simple Linear Regression, used to measure the relationship or similarity of 2 vectors of data or Time Series data [29]. The R-Square technique is appropriate for 1-dimensional sequence data [17][18].

Here, there are 2 sets of Time Series data: $X = \{x_1,.....,x_m\}$ and $Y = \{y_1,....., y_n\}$. By the R-Square technique, the distribution of data along the points $(x_1, y_1),(x_2, y_2),....,(x_n, y_n)$ on axes X and axes Y, calculated by Equation 2.18.

$$R - Square(X,Y) = \frac{\left[\sum_{i=1}^{N}(x_i - \overline{X})(y_i - \overline{Y})\right]^2}{\sum_{i=1}^{N}(x_i - \overline{X})^2 \sum_{i=1}^{N}(y_i - \overline{Y})^2} \qquad 2.18$$

where $\overline{X}$ and $\overline{Y}$ is average value of x, y sequences.

The closer the value of R-Square to 1 the stronger the linear relation of the two sequences or strong similarity, while R-Square is close 0 is means weak similarity or dissimilarity [17][18].

2.5.2.2 Regression Time Warping (RTW)

It is a method adapted from the R-Square technique and the Dynamic Time Warping technique, which are integrated together. The method was devised by Hansheng Lei, Srinivas Palla and Venu Govindaraj [31][36], from the R-Square technique, used with 1-dimensional sequence, while the DTW concept using the Warping Path to compute the distance combined, from the n-by-m matrix table, is the RTW technique. This technique uses points on the Warping Path, between point i of X axis and point j of Y axis for a calculation through a technique of Simple Linear Regression, as illustrated in Figure 2.6 [16][17][18], via Equation 2.19.

$$RTW = \frac{\left[\sum_{j=1}^{M}[\sum_{i=1}^{N}(x_{ij} - \overline{X}_j)(y_{ij} - \overline{Y}_j)]\right]^2}{\sum_{j=1}^{M}\sum_{i=1}^{N}(x_{ij} - \overline{X}_j)^2 \sum_{j=1}^{M}\sum_{i=1}^{N}(y_{ij} - \overline{Y}_j)^2} \qquad 2.19$$

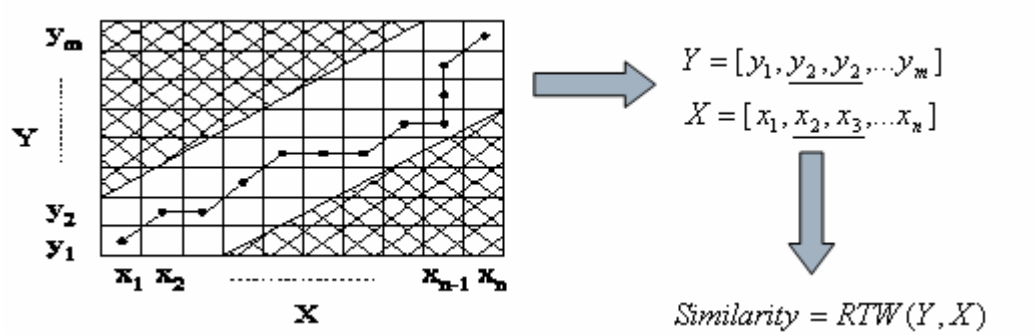


Figure 2.6: Regression Time Warping Path

R-Square is a good measure for 1-dimensional Data. RTW is extended from R-Square and DTW to measure multi-dimensional sequences similarity.

R-Square and RTW compute with amplitude scaling and offset shifting. R-Square allows the one-one mapping between the elements of two sequences. If two sequences have similar shape but they are out of phase, R-Square based on one-one mapping will not work. RTW can be used on this situation, which allows the many-one and one-many mapping between the elements of two sequences [18].

2.6 Related Works

Over the past years, there have been a large number of interested people, both inside and outside Thailand, who conduct a study on the impact of economic factors. In this section, researches on the relationship among economic factors are surveyed focusing mainly on those related to oil prices, stock exchange, exchange rate and other economic factors.

2.6.1 Oil prices and GDP

With respect to the domestic refined oil prices and the Thai Economy, Bundit Chaiwitchayachat (2007)[3][4] proposed in his articles entitled "**The impacts of domestic oil prices alteration on Thai economic system**" (2006)[3] and "**The impact of global oil prices on Thai economic system : macro-economic model**" (2007)[4]. In addition, based on other research surveys on global oil prices, demand for energy consumption in Thailand, and the amount of oil consumption and macro-economic variables (Economic Growth Rate – GDP, Inflation rate, interest rates policy, Dubai crude Oil prices, Benzene 91 retail prices, Benzene 95 retail prices and

Diesel retail prices), it has been found that oil prices alteration is one of the factors directly affecting the country's economy. The analysis for this research applies a statistical approach called "Granger Causality" and is based on the following assumptions.

For the relationship between oil prices and macro-economic variables, it is found that.

1. Oil prices and Gross National Product (GNP) are interrelated.
2. Oil prices, Interest rates and Inflation rates are not related.

With regard to the Thai economic reaction to oil prices variation, it is found that.

1. Oil prices fluctuation has an impact on Gross Domestic Product (GDP).
2. Interest rates and Inflation rate respond only slightly to oil prices variation.

Concerning the impact of oil prices on Thai economic system, it is found that.

1. Prices of all types of oil react to economic activities.
2. Oil Prices variation generally results in the Inflation rate.
3. Domestic oil prices variation gives rise to Interest rate fluctuation rather than global oil prices variation.

Furthermore, Siriwan Sukhanthapree's study (2006)[32] entitled, "**The study on impacts of oil prices variation on macro-economic variables of the Thai economic system**" investigates the demand for and production, import and price level of fuel oil of Thailand. The study analyzes the impacts of oil prices variation on such Thai macro-economic variables as Gross Domestic Product (GDP), the private sector consumption and investment, taxes, interest rates, the imports of goods and Services, as well as demands to conduct financial transactions. The analysis is made through the technique of Co-Integration and Error Correlation. The study findings reveal that oil prices are most related to the private sector investment, followed by Gross National Product (GNP), the imports of goods and services, the private sector consumption, demands to conduct financial transactions, and taxes and interest rate.

2.6.2 Oil prices and Stock Exchange

With respect to the relationship between securities prices and oil prices, Sarun Thawilwung (2004)[31] proposes in his article "**Impacts of oil prices on SET Index using Quantitative Analysis**" that, based on the study on the relationship between oil prices and SET index using the Correlation Coefficient technique, it is found that the Correlation Coefficient value between SET index and oil prices is 0.41, which is not high, since there are many factors affecting SET Index. However, between 1999 and 2000, during which there arose and oil prices crisis, in which oil prices increased while SET index diminished, the Correlation Coefficient value was -0.51, which was considered a high relationship level but in an opposite direction. Similarity, between January and August 2004, the Correlation Coefficient value obtained was -0.84, indicating that investors regarded oil prices as factor in investment decision making in SET index.

Concerning the relationship of oil prices and industrial group, Sudaporn Pengmol (2006)[35] states in her article, "**The study on the consequences of oil prices variation on securities index for individual industrial group**" using the Cointegration technique , that group of Food and Beverage, Life and Non-Life Insurance, Real estate developers, Energy and Infrastructures, Mining, Commerce. Hotel and Touristic services, Printing and Publications, Transports, Chemicals and Plastic, Communications, Construction Materials and Decoration, and Electric devices all have long-term relationship with oil prices rise. Nonetheless, Banking Industry as well as Motor Vehicles and Equipment Industry have no long-tern relationship with oil prices rise.

For the study on the relationship of oil prices and securities in Energy Industry doing oil-related business, using technique of Cointegration and Error Correlation, Woravit Suriyawong (2006)[40], in his article, "**The study on relationship between securities prices in Energy Industry conducting oil-related business and oil prices in global market**s" finds that for a decision to invest in PTT securities, data on Dubai oil prices variation should be used. Likewise, for a decision to invest in

PTTEP, SUSCO and TOP securities, data in Malaysian, Singaporean, and Singaporean oil prices variation should be used, respectively.

Similarity, the study abroad on oil prices and Stock Exchange, Syed A. Basher and Perry Sadorsky (2006)[36], in their article, "**Oil price risk and emerging stock market**" examine the imparts of oil prices variation on the remuneration from investments in Stock Markets, through finding the relationship among oil prices, Stock Markets in foreign countries and Money Exchange rate. It is found that, based on a test of the relationship of the 3 afore-mentioned variables via the Multiple Regression technique, the Money Exchange rate has no relationship with External Stock Markets while oil prices and External Stock Markets do have certain relationship.

Gerben Driesprong, Ben Jacobsen and Benjamin Maat (2004)[12], in their article "**Stock Market and Oil Prices**" examine the effects of oil prices variation on Stock Markets, in term of "**A rise in oil prices, lower future stock market returns**". The monthly data on oil prices at WTI (West Texas), Dubai, Brent and Arabian Light are used while the Regression technique is used to create a model to find the relationship between oil prices and External Stock Markets. The findings reveal that oil prices variation affects External Stock Markets in the opposite direction. The higher the oil prices, the less the profit derived from an investment in the Stock Markets.

2.6.3 Exchange Rate and GDP

Pakin Jitpokkasame (2007)[41], in their article "**Exchange Rate Pass-through to Consumer Price in Thailand**". The data analyzed are the export and import values (1995-2005), exchange rate, domestic consumer goods prices level and GDP. The analysis is done via models of correlation between exchange rate and consumer prices index, creating by the technique of Unit Root Test, Cointegration and Error Correlation Model.

From the analysis, the correlation coefficient in the models is equivalent to 0.18. That is, if the exchange rate is adjusted by 1%, the domestic consumer goods

prices level will be adjusted by 0.18%. And the long-run Cointegration analysis reveals a long-run positive equilibrium relationship between GDP and consumer prices index. With regard to the impact of oil price changes in every period, there is a positive equilibrium relationship between oil prices and consumer good prices.

Jakree Sakunbongkot (2001) [42], in their article "**The Impact of Exchange Rate on Thai Industrial Export Prices**" process the hypothesis that currency exchange rate is the major factor influencing the export price variation. When the exchange rate weakens the balance of trade will improve a result of export price decrease. The data analyzed are the data on export to USA, Australia, Singapore, Malaysia, Taiwan, The Netherlands, Belgium, Japan and Portugal, currency exchange rate, industrial goods data (Vehicle, Computer and IC board group), export price index and GDP. The data cover the year 1990 to 1999. The analysis is done via stationary test using Unit Root and the Cointegration technique or with VAR (Vector Autoregressive).

The result of the study is that in the vehicle industry, the exchange rate affects the export to the USA, Australia, Belgium and Portugal. But it does not affect the export to Japan. In the computer industry, the exchange rate affects the export to the USA and Singapore but does not affect the export to Japan, The Netherlands, and Malaysia. In the IC board industry, the export to the USA, Japan, The Netherlands, and Singapore are affected while the export to Taiwan is not.

2.6.4 Exchange Rate and Oil Prices

In the article entitled "**Oil Prices and the Rise and Fall of the U.S. Real Exchange Rate**" by Robert A. Amano and Simon van Norden (1997) [43] study the relationship between oil prices and US exchange rate; how oil prices explain the movement of US exchange rate. The analysis techniques used are Cointegration, Causality and Error-Correction. The data used cover the year 1972 to 1993. The result found from the study is that oil prices influence the change of US exchange rate (Cointegration – US Dollar (0.550) and Oil (0.686)).

Ingunn Fride Tvete and Anders Loland (2007) [44], study in "**Oil prices and exchange rates: A survey**". Based on the assumption that oil prices and profits from oil trading are in US dollar, the study is to find out whether there is a relationship in the opposite direction among the rise in oil prices, exchange rate and debts. The technique used are Cointegration, Causality and Natural Hadge.

The data used are oil prices, EUR and USD exchange rate covering the year 1975 to 1998. It is found that oil prices are related to exchange rate and the relationship is in the opposite direction.

Table 2.7 reveals a summary of the local and foreign studies on the relationship between oil prices and securities.

Table 2.7: A Summary of Literature Review

No	Researcher	Title	Methodology	Consequence
1	Bundit Chaiwitchayachat [3]	Impacts of oil prices variation on Thai Economic System	Granger Causality	Oil Prices and Gross National Product (GNP) are interrelated.
				Oil Prices, Interest rate and Inflation rate are not related.
		Journal of Economics ,Kasetsart University, Vol 1, 2006		Oil Prices variation affects GDP in the opposite direction.
2	Bundit Chaiwitchayachat [4]	Impacts of domestic oil prices variation on Thai Economic System.	Multiple Regression and Vector Autoregressive	Domestic Oil prices variation results in Interest rate fluctuation more than global oil prices variation does.
		Faculty of Economics ,Chulalongkorn University		Domestic oil prices rise reduces both short and long tern economic growth rate.
3	Siriwan Kunthapree [32]	Impacts of oil prices variation on Macro-Economic variables of Thai Economic System.	Cointegretion and Error Correlation	Oil prices have highest relationship with the private sector investment, followed by GNP, Imports, Goods & Services, Private Sector Consumption, and demand to hold money, Taxes and Interest rate.
		Economics Library, Faculty of Economics, Chiang Mai		

Table 2.7: A Summary of Literature Review (Cont.)

No	Researcher	Title	Methodology	Consequence
4	Sarun Thawilwung [31]	Impacts of oil prices on SET Index using Quantitative analysis	Correlation Coefficient	Industrial group (except banking as well as motor vehicles and equipment) have long-tern relationship with Oil prices rise.
		Far East Securities Company		Oil prices are a factor in decision to invest in SET Index.
5	Woravit Suriyawong [40]	Study on relationship between Securities prices in Energy group doing oil-related business and global markets oil prices	Cointegretion and Error Correlation	Dubai oil prices impacts PTT securities.
				Malaysian oil prices impact PTTEP securities.
				Singapore oil prices impact SUSCO securities.
		Economics Library, Faculty of Economics, Chiang Mai		Singapore oil prices impact TOP securities.
6	Syed A. Basher and Perry Sadorsky [36]	Oil price risk and Emerging Stock Markets	Multiple Regression and Vector Autoregressive	Money Exchange rate is not related to Foreign Stock Markets.
		Global Finance Journal		Oil Prices and Foreign Stock Markets are interrelated.
7	Gerben Driesprong, Ben Jacobsen and Benjamin Maat[12]	Stock Markets and Oil Prices	Regression	Oil prices variation influences Foreign Stock Markets in the opposite direction.
		Social Science Electronic Publishing		

Table 2.7: A Summary of Literature Review (Cont.)

No	Researcher	Title	Methodology	Consequence
8	Pakin Jitpokkasame [41]	Exchange Rate Pass-through to Consumer Price in Thailand Faculty of Economics, Kasetsart University	Unit Root Test, Cointegration and Error Correlation Model	Exchange rate is adjusted by 1%, the domestic consumer goods prices level will be adjusted by 0.18%.
9	Jakree Sakunbongkot [42]	The Impact of Exchange Rate on Thai Industrial Export Prices Faculty of Economics, Chulalongkom University	Unit Root and Cointegretion	The change in exchange rate does not increase the capability for export.
10	Robert A. Amano and Simon van Norden [43]	Oil Prices and the Rise and Fall of the U.S. Real Exchange Rate Research Department, Bank of Canada	Cointegration, Causality and Error Correction	Oil prices variation influences US Exchange Rate.
11	Ingunn Fride Tvete and Anders Loland[44]	Oil prices and exchange rates: A survey Norsk Regnesentral, Norwegian Computing Center	Cointegration, Causality and Natural Hedge	Oil prices variation influences Foreign Exchange Rate in the opposite direction.
12	Sudaporn Pengmol [6]	The study on the consequences of oil prices variation on securities index for individual industrial group.	Multiple Regression and Vector Autoregressive.	COMM, TOURISM, MEDIA, TRANS in SERVICE have long-term relationship with oil prices rise. BANK in FINCIAL and AUTO Sector have no long-tern relationship with oil prices rise.

In summary of the studies on the relationship between economic factors influencing the Thai Economy as discussed earlier, it is found that.

1. Most of the studies use a statistical approach to analyze data in order to find the relationship of data in terms of cause and effect, or to examine which data have an influence on other data, The consequences derived from the experiment become figures of which their values are used as references, This, of course, enables a comparison of figures and other standard values, making an easier comprehension of the analysis of consequences.
2. Economic factors that most researches used for analysis are oil prices data, GDP data, NGP data, price or index of stock exchange, external stock exchange index and foreign exchange rate.

2.7 A Survey of Software Tools for Economic Data Analysis

This topic discusses the software used in economical data analysis, which are:

1. **SPSS** [46] (Statistical Package for the Social Sciences) is a software for data management and analysis. SPSS processes input data in many types of file format (Excel, Text , dBase, ,SAS). Output is shown in summary tables, pivot tables, graphs, descriptive statistical summary reports and advance statistical summary reports. SPSS runs on Windows, Macintosh and UNIX system. Its main features are statistical function, data analysis function, and report and deploy function. SPSS is easy to use because statistical analysis is done in Excel spreadsheets and analysis functions are in command line. SPSS is commonly used in social science researches such as psychology, marketing, and etc.

Popular features of SPSS are data editor (data management similar to excel spreadsheets), ANOVA (data analysis), regression, forecasting and other advance statistical functions. Analysis outputs are in GUI graph, pivot tables and custom tables.

2. **Stata** [47] is a software based on written command. It has command line, statistical function, spreadsheet tool, and data management features (data editor). It is mostly used in biostatistics and epidemiology.

Main features of Stata are linear regression, logistic regression, Probit, point estimates, associated standard errors, contingency tables with Rao-Scott corrections of chi-squared, tests and confidence intervals.

3. **EViews** [48] is a software tool for immediately performing tasks ranging from data manipulation, to statistical and econometric analysis, to complex multivariate simulation, to construction of presentation. EViews has amount of recent statistical algorithms implemented (Time Series, Unit Root, Cointegration, Error Correction Model, etc). EViews is the most popular language used by academic statisticians. User can write a program to analyze data, but it dose not have data editor for input raw data.

Most popular statistical features of EViews are descriptive statistics, ARCH models, GARCH, Correlation analysis, Autocorrelation and partial autocorrelation analysis, Granger causality tests, unit root and panel unit root tests, Multiple Regression and OLS, and Vector Autoregression and Error Correction Models. Output of analysis is time series plots and distribution graphs (histograms, distribution plots, kernel density plots).

4. **SAS:** Statistical Analysis System [45]. SAS is used mostly in business. It is used by more than 3,000 financial institutions worldwide, including 97 percent of banks in the FORTUNE Global 500®. Most users use it for data management and data analysis. SAS is much better for large data sets, and for data preparation.

Main features of SAS are data management, multiple data sets, statistical analysis module, miner module, and integration module. Miner module has data preparation functional, Clustering algorithms, Decision tress, Memory-based Reasoning (MBR) and text mining process. Statistical analysis module is regression, correlation, forecasting, and time series model. Integration model can load data files in variety of formats and extract data from the document. SAS has great technical support, online help and tutorial.

A comparison of SPSS, Stata, SAS and EViews are given in Table 2.8.

Table 2.8: A Comparison of SPSS, Stata, SAS and EViews.

	***SPSS**	***Stata**	***SAS**	**EViews**
User	All user, Marketing researcher.	Academic user.	Power user, Statistic researcher.	Academic user
Industry	Marketing research	Academic institutions	Enterprise computing, Pharmaceutical, Database marketing	Academic institutions.
Ease of use	Commands are complex	Command structure is easy	Commands are not very easy to learn	Commands are not easy to learn or remember
Style	Menu-driven or programming	Command, Menu-driven or programming	Programming only	Programming only
Features	Greater statistical ,Regression, panel designs, time series, and diagnostics; User developed procedures Graphics for analysis output	Greater statistical, ANOVA, regression, forecasting, and analysis output in GUI graph, pivot tables and custom tables.	Data Management Multiple data sets statistical analysis module, miner module, and integration module.	Correlation and Autocorrelation, Partial autocorrelation, Granger causality tests, unit root and panel unit root tests, and Multiple Regression and OLS.
Weak	Hard to add-on modules need separate purchase	Weak on ANOVA, Factor Analysis	Not easy to use Difficult to learn	No technical support.

*Alan Acock. SAS, Stata, SPSS: A Comparison. [49]

CHAPTER III
SYSTEM ANALYSIS AND DESIGN

This chapter discusses System Analysis and Design, Data Collection methodology, stages of converting data into Data Warehouse, retrieval of data from Data Warehouse by OLAP engine, and analysis of data via Time Series Similarity Analysis. Details of each section are given below.

3.1 System Design

An overview for Economic Factors Analysis System is illustrated in Figure 3.1, starting from raw data to the display of data derived from the analysis.

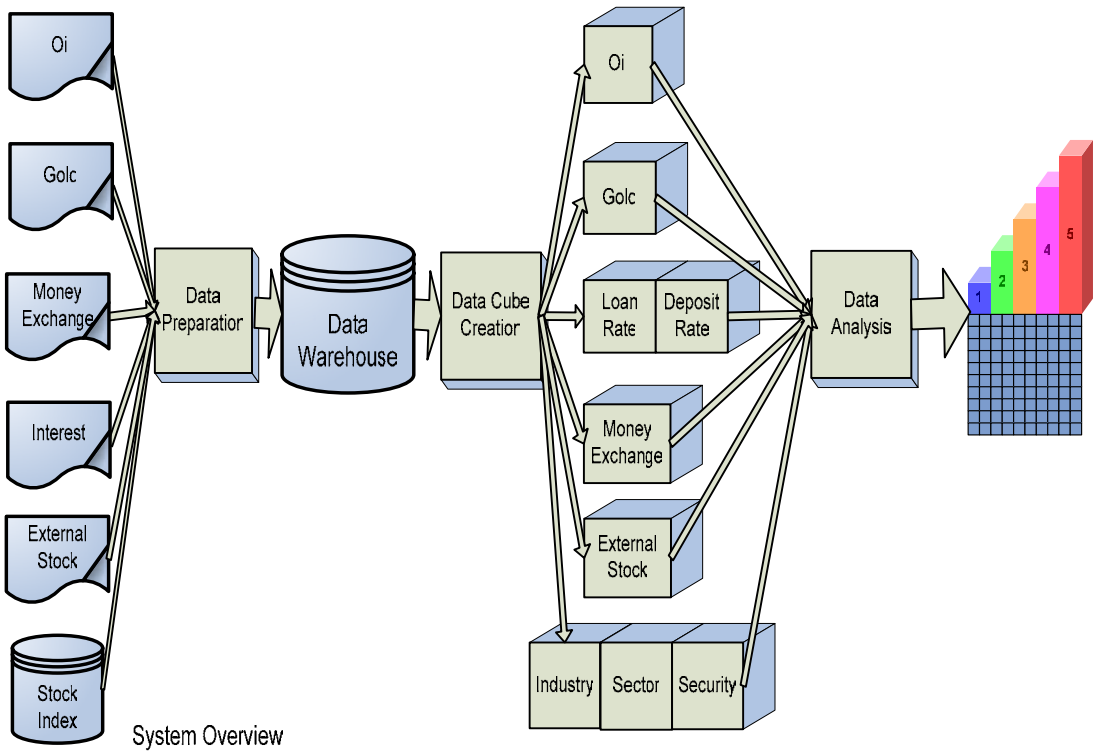


Figure 3.1: System Overview for Economic Factors Analysis System

Based on Figure 3.1, the system structure consists of 3 following working processes.

1. Stage of Data Preparation.
2. Stage of Data Cube Creation.
3. Stage of Data Analysis.

3.1.1 Structure Chart of the Economic Factors Analysis System

The structure chart of Economic Factors Analysis proposed in this research is composed of Data Preparation, Data Cube Creation and Data Analysis, as illustrated in Figure 3.2

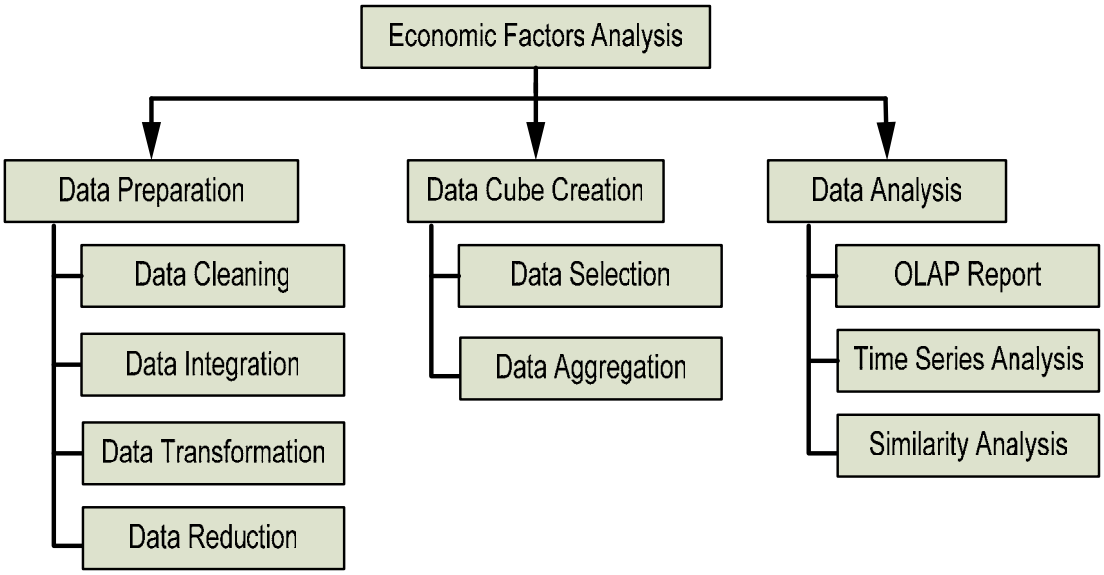


Figure 3.2: Structure Chart of Economic Factors Analysis System

3.1.2 Context Diagram

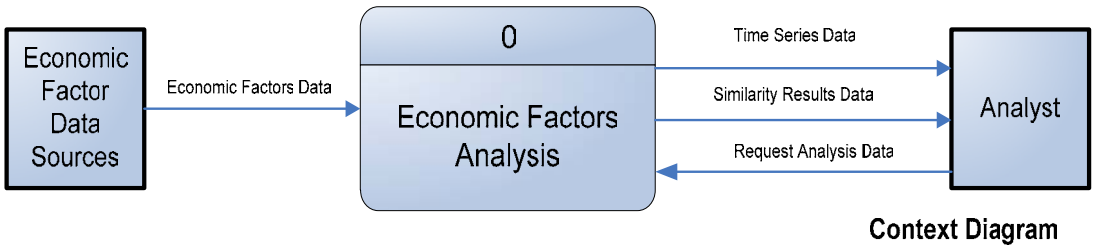


Figure 3.3: Context Diagram: Economic Factors Analysis System

3.1.3 First Level Data Flow Diagram

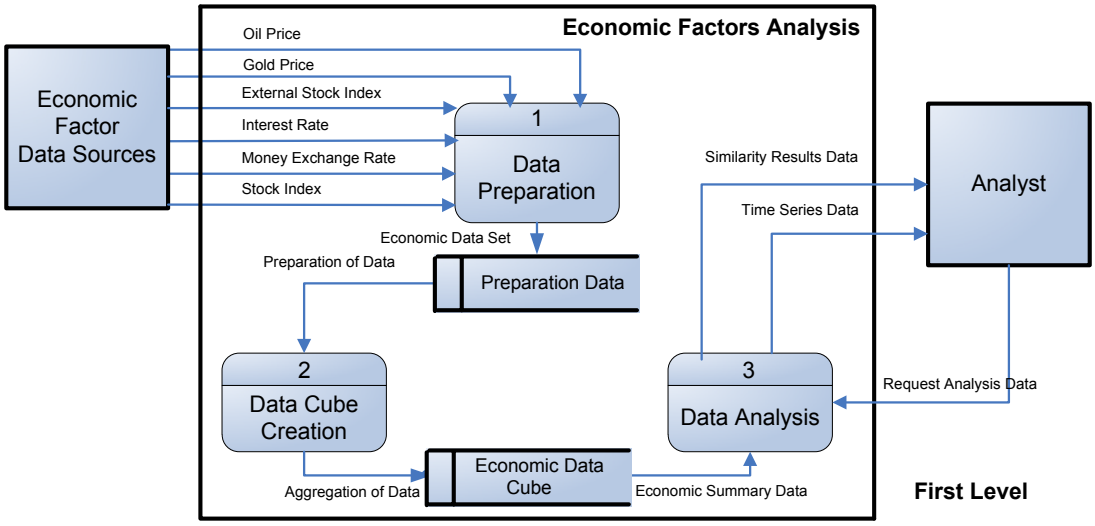


Figure 3.4: First Level: Economic Factors Analysis System

3.1.4 Second Level Data Flow Diagrams

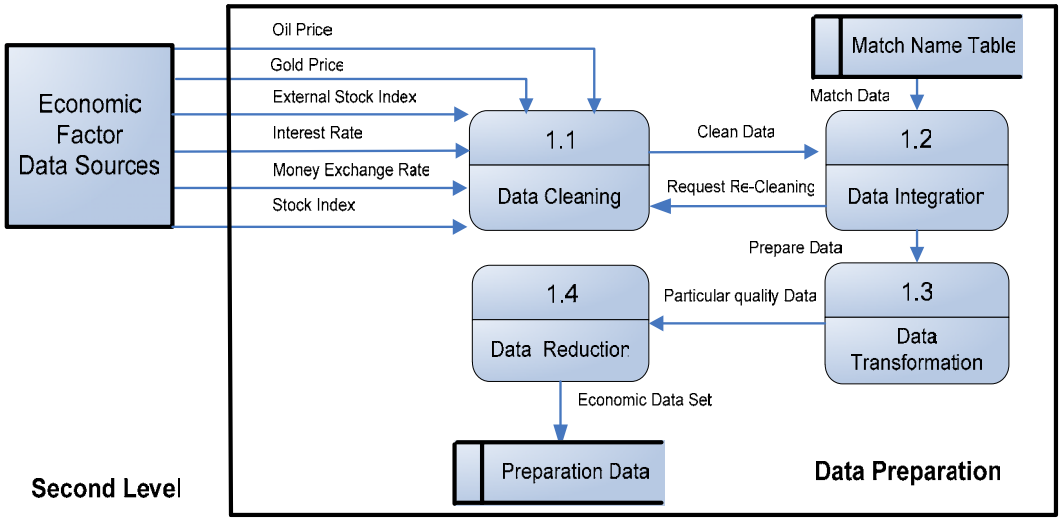


Figure 3.5: Second Level: Data Preparation

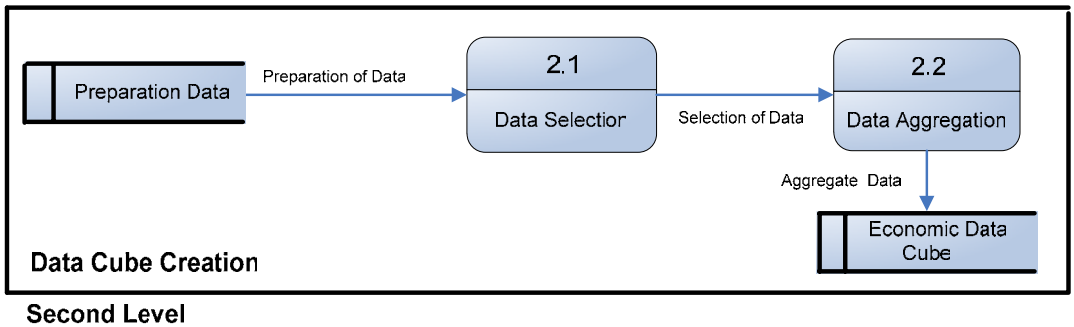


Figure 3.6: Second Level: Data Cube Creation

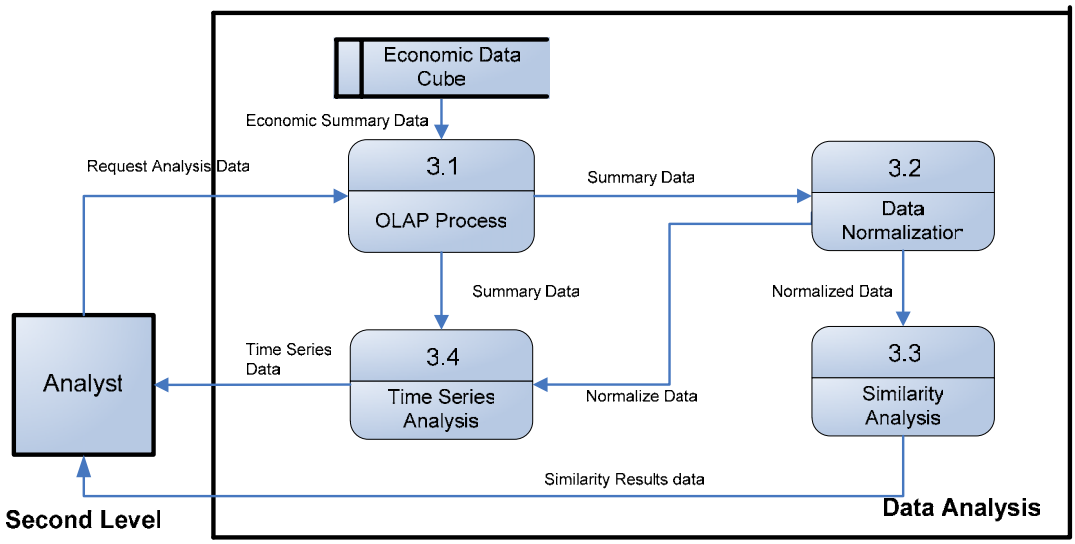


Figure 3.7: Second Level: Data Analysis

Table 3.1: Process Name 0: Economic Factors Analysis System

Process Name : 0	**Economic Factors Analysis**
Input Data Flow	Economic Factors Data
	Request Analysis Data
Output Data Flow	Time Series Data
	Similarity Results Data
Data Stored Used	
Description	Process of Economic Factors Analysis

Table 3.2: Process Name 1: Data Preparation

Process Name : 1	**Data Preparation**
Input Data Flow	Oil Price
	Gold Price
	External Stock Index
	Interest Rate
	Money Exchange Rate
	Stock Index
Output Data Flow	Economic Data Set
Data Stored Used	Preparation Data
Description	Process of preparing Raw Data

Table 3.3: Process Name 2: Data Cube Creation

Process Name : 2	**Data Cube Creation**
Input Data Flow	Preparation of Data
Output Data Flow	Aggregation of Data
Data Stored Used	Economic Data Cube
Description	Process of building Data Cube.

Table 3.4: Process Name 3: Data Analysis

Process Name : 3	**Data Analysis**
Input Data Flow	Summary Data
	Request Analysis Data
Output Data Flow	Time Series Data
	Similarity Results Data
Data Stored Used	Economic Data Cube
Description	Process of preparing analysis of economic factors

Table 3.5: Process Name 1.1: Data Cleaning

Process Name : 1.1	**Data Cleaning**
Input Data Flow	Oil Price
	Gold Price
	External Stock Index
	Interest Rate
	Money Exchange Rate
	Stock Index
	Request Re-Cleaning
Output Data Flow	Clean Data
Data Stored Used	
Description	Process of cleaning Raw Data

Table 3.6: Process Name 1.2: Data Integration

Process Name : 1.2	**Data Integration**
Input Data Flow	Clean Data
	Match Data
Output Data Flow	Prepare Data
	Request Re-Cleaning
Data Stored Used	Match Name Table
Description	Process of combining and grouping Raw Data

Table 3.7: Process Name 1.3: Data Transformation

Process Name : 1.3	**Data Transformation**
Input Data Flow	Prepare Data
Output Data Flow	Particular Quality Data
Data Stored Used	
Description	Process of transforming and converting Raw Data to Particular Quality Data

Table 3.8: Process Name 1.4: Data Reduction

Process Name : 1.4	**Data Reduction**
Input Data Flow	Particular Quality Data
Output Data Flow	Economic Data Set
Data Stored Used	Preparation Data
Description	Process of reducing data size

Table 3.9: Process Name 2.1: Data Selection

Process Name : 2.1	**Data Selection**
Input Data Flow	Preparation of Data
Output Data Flow	Selection of Data
Data Stored Used	
Description	Process of performing in selecting the data.

Table 3.10: Process Name 2.2: Data Aggregation

Process Name : 2.2	**Data Aggregation**
Input Data Flow	Selection of Data
Output Data Flow	Aggregate of Data
Data Stored Used	Economic Data Cube
Description	Process of Summary Data

Table 3.11: Process Name 3.1: OLAP Process

Process Name : 3.1	**OLAP Process**
Input Data Flow	Economic Summary Data
Output Data Flow	Summary Data
	Summary Data
Data Stored Used	Economic Data Cube
Description	Process of performing OLAP Engine for getting data from Data Warehouse.

Table 3.12: Process Name 3.2: Data Normalization

Process Name : 3.2	**Data Normalization**
Input Data Flow	Summary Data
Output Data Flow	Normalize Data
	Normalize Data
Data Stored Used	
Description	Process of converting Summary Data to join in normalizing

Table 3.13: Process Name 3.3: Similarity Analysis

Process Name : 3.3	**Similarity Analysis**
Input Data Flow	Normalized Data
Output Data Flow	Similarity Results Data
Data Stored Used	
Description	Process of calculating Similarity Analysis

Table 3.14: Process Name 3.4: Time Series Analysis

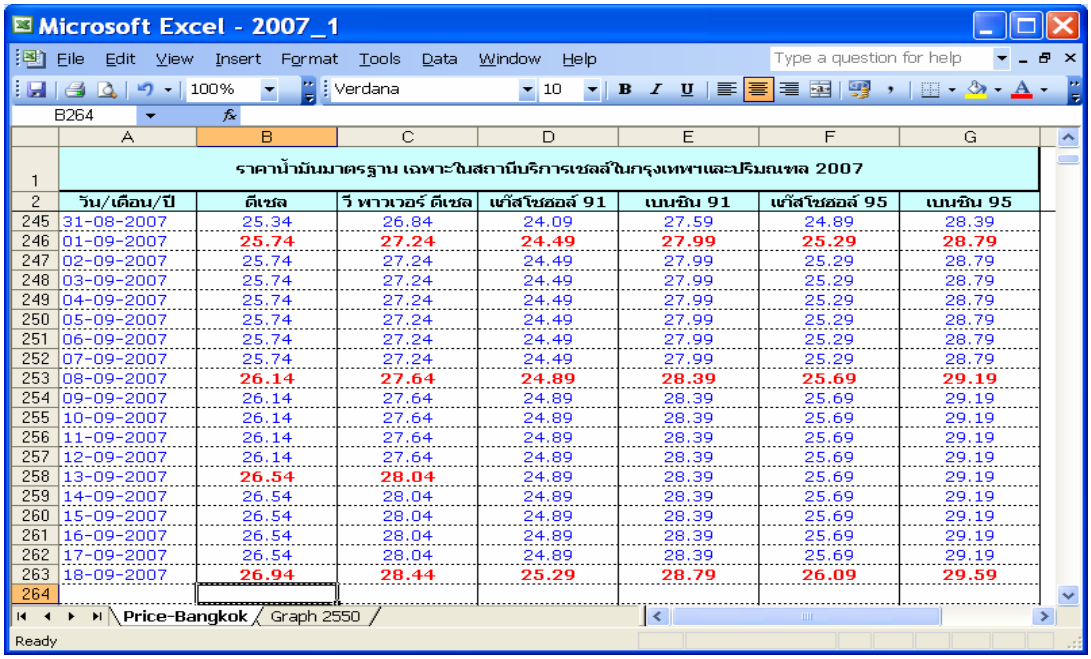
Process Name : 3.4	**Time Series Analysis**
Input Data Flow	Normalize Data
	Summary Data
Output Data Flow	Time Series Data
Data Stored Used	
Description	Process of performing in seeking part of Time Series.

3.2 Economic Factor Data Sources

They are data used for analysis and are in the format of raw data, whether they be Excel file, HTML file, Text file and Database. Data Sources are normally downloaded from various web sites and also from the Stock Exchange of Thailand (SET).

3.2.1 Oil File Structure

Data of prices of oil sold in the country are derived from the web site of Shell Company, Thailand. Files obtained are Excel files categorized by year, each of which consists of 2 sheets of Price-Bangkok and Graph YYYY and data start at 1 January 1996 – 30 January 2008, in record count is 26,214 records and file count is 13 files. The former compiles oil prices data which the latter displays graph of oil prices. An example of data is illustrated in Figure 3.8.



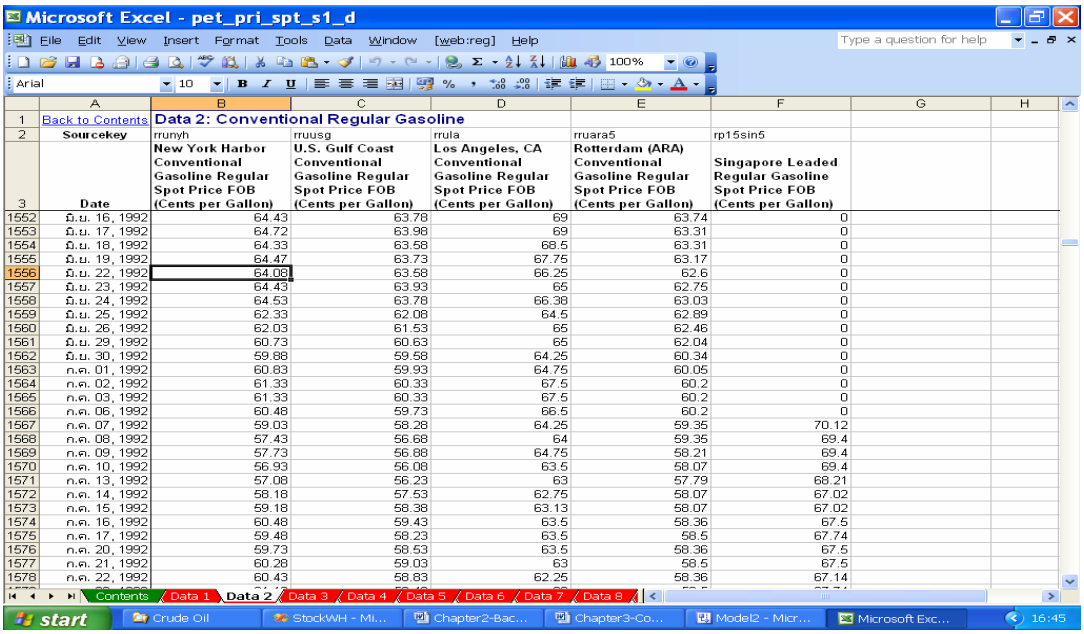
Microsoft Excel - 2007_1

ราคาน้ำมันมาตรฐาน เฉพาะในสถานีบริการเชลล์ในกรุงเทพฯและปริมณฑล 2007

	วัน/เดือน/ปี	ดีเซล	วี พาวเวอร์ ดีเซล	แก๊สโซฮอล์ 91	เบนซิน 91	แก๊สโซฮอล์ 95	เบนซิน 95
245	31-08-2007	25.34	26.84	24.09	27.59	24.89	28.39
246	01-09-2007	25.74	27.24	24.49	27.99	25.29	28.79
247	02-09-2007	25.74	27.24	24.49	27.99	25.29	28.79
248	03-09-2007	25.74	27.24	24.49	27.99	25.29	28.79
249	04-09-2007	25.74	27.24	24.49	27.99	25.29	28.79
250	05-09-2007	25.74	27.24	24.49	27.99	25.29	28.79
251	06-09-2007	25.74	27.24	24.49	27.99	25.29	28.79
252	07-09-2007	25.74	27.24	24.49	27.99	25.29	28.79
253	08-09-2007	26.14	27.64	24.89	28.39	25.69	29.19
254	09-09-2007	26.14	27.64	24.89	28.39	25.69	29.19
255	10-09-2007	26.14	27.64	24.89	28.39	25.69	29.19
256	11-09-2007	26.14	27.64	24.89	28.39	25.69	29.19
257	12-09-2007	26.14	27.64	24.89	28.39	25.69	29.19
258	13-09-2007	26.54	28.04	24.89	28.39	25.69	29.19
259	14-09-2007	26.54	28.04	24.89	28.39	25.69	29.19
260	15-09-2007	26.54	28.04	24.89	28.39	25.69	29.19
261	16-09-2007	26.54	28.04	24.89	28.39	25.69	29.19
262	17-09-2007	26.54	28.04	24.89	28.39	25.69	29.19
263	18-09-2007	26.94	28.44	25.29	28.79	26.09	29.59

Figure 3.8: Oil Price: Excel File

In addition, data on foreign crude oil and refined oil prices are received from the web site of Energy Information Administration Company. Files obtained are Excel files consisting of 10 sheets and data start at 2 June 1986 – 30 January 2008, in record count is 5,541 records and file count is 1 file. Data are cut out from data sheets 2 and 3, which are composed of date data as well as foreign crude oil and refined oil prices data from various sources, as illustrated in Figure 3.9.



Microsoft Excel - pet_pri_spt_s1_d

File Edit View Insert Format Tools Data Window [web:reg] Help

Back to Contents | Data 2: Conventional Regular Gasoline

	Sourcekey	rrunyh	rruusg	rrula	rruara5	rp15sin5
Date		New York Harbor Conventional Gasoline Regular Spot Price FOB (Cents per Gallon)	U.S. Gulf Coast Conventional Gasoline Regular Spot Price FOB (Cents per Gallon)	Los Angeles, CA Conventional Gasoline Regular Spot Price FOB (Cents per Gallon)	Rotterdam (ARA) Conventional Gasoline Regular Spot Price FOB (Cents per Gallon)	Singapore Leaded Regular Gasoline Spot Price FOB (Cents per Gallon)
1552	ม.ย. 16, 1992	64.43	63.78	69	63.74	0
1553	ม.ย. 17, 1992	64.72	63.98	69	63.31	0
1554	ม.ย. 18, 1992	64.33	63.58	68.5	63.31	0
1555	ม.ย. 19, 1992	64.47	63.73	67.75	63.17	0
1556	ม.ย. 22, 1992	64.08	63.58	66.25	62.6	0
1557	ม.ย. 23, 1992	64.43	63.93	65	62.75	0
1558	ม.ย. 24, 1992	64.53	63.78	66.38	63.03	0
1559	ม.ย. 25, 1992	62.33	62.08	64.5	62.89	0
1560	ม.ย. 26, 1992	62.03	61.53	65	62.46	0
1561	ม.ย. 29, 1992	60.73	60.63	65	62.04	0
1562	ม.ย. 30, 1992	59.88	59.58	64.25	60.34	0
1563	ก.ค. 01, 1992	60.83	59.93	64.75	60.05	0
1564	ก.ค. 02, 1992	61.33	60.33	67.5	60.2	0
1565	ก.ค. 03, 1992	61.33	60.33	67.5	60.2	0
1566	ก.ค. 06, 1992	60.48	59.73	66.5	60.2	0
1567	ก.ค. 07, 1992	59.03	58.28	64.25	59.35	70.12
1568	ก.ค. 08, 1992	57.43	56.68	64	59.35	69.4
1569	ก.ค. 09, 1992	57.73	56.88	64.75	58.21	69.4
1570	ก.ค. 10, 1992	56.93	56.08	63.5	58.07	69.4
1571	ก.ค. 13, 1992	57.08	56.23	63	57.79	68.21
1572	ก.ค. 14, 1992	58.18	57.53	62.75	58.07	67.02
1573	ก.ค. 15, 1992	59.18	58.38	63.13	58.07	67.02
1574	ก.ค. 16, 1992	60.48	59.43	63.5	58.36	67.5
1575	ก.ค. 17, 1992	59.48	58.23	63.5	58.5	67.74
1576	ก.ค. 20, 1992	59.73	58.53	63.5	58.36	67.5
1577	ก.ค. 21, 1992	60.28	59.03	63	58.5	67.5
1578	ก.ค. 22, 1992	60.43	58.83	62.25	58.36	67.14

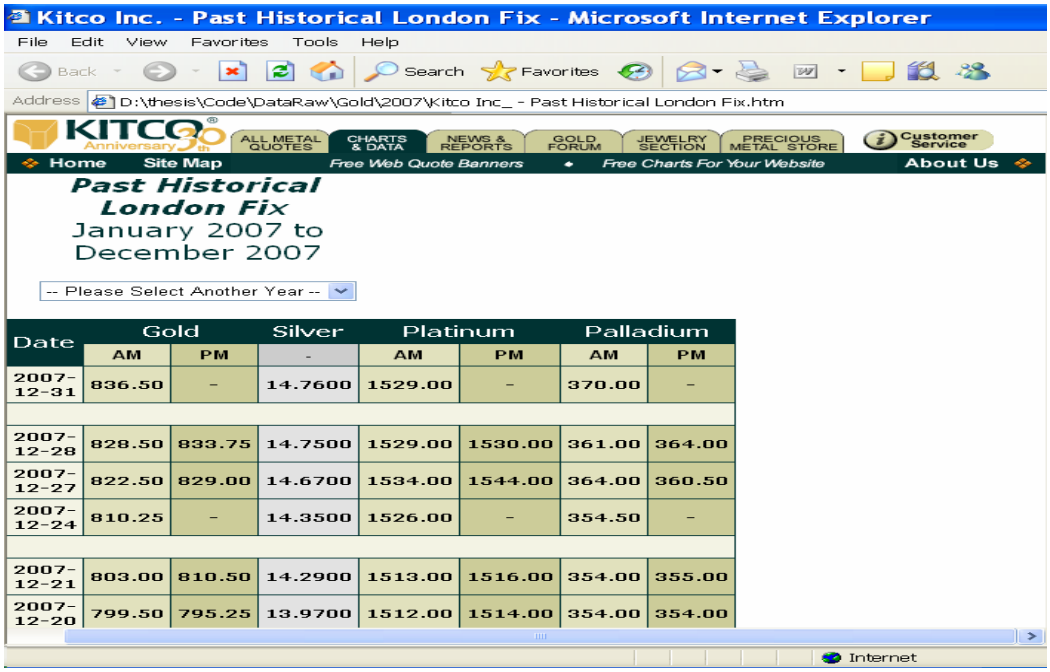
Contents | Data 1 | Data 2 | Data 3 | Data 4 | Data 5 | Data 6 | Data 7 | Data 8

start | Crude Oil | StockWH - Mi... | Chapter2-Bac... | Chapter3-Co... | Model2 - Micr... | Microsoft Exc... | 16:45

Figure 3.9: Crude Oil Price: Excel File

3.2.2 Gold File Structure

Data on foreign gold bar prices are derived from the web site of Kitco, which records precious minerals data. They are in the format of HTML file and comprise Gold, Silver, Platinum and Palladium data, of which prices are in US dollar and divided into morning and afternoon ones, data start at 6 January 1997 – 31 December 2007, in record count is 5,118 records and file count is 11 files, as revealed in Figure 3.10.



Kitco Inc. - Past Historical London Fix - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back | Search | Favorites

Address D:\thesis\Code\DataRaw\Gold\2007\Kitco Inc_ - Past Historical London Fix.htm

KITCO 30 Anniversary | ALL METAL QUOTES | CHARTS & DATA | NEWS & REPORTS | GOLD FORUM | JEWELRY SECTION | PRECIOUS METAL STORE | Customer Service

Home | Site Map | Free Web Quote Banners | Free Charts For Your Website | About Us

Past Historical London Fix

January 2007 to December 2007

-- Please Select Another Year --

Date	Gold		Silver	Platinum		Palladium	
	AM	PM	-	AM	PM	AM	PM
2007-12-31	836.50	-	14.7600	1529.00	-	370.00	-
2007-12-28	828.50	833.75	14.7500	1529.00	1530.00	361.00	364.00
2007-12-27	822.50	829.00	14.6700	1534.00	1544.00	364.00	360.50
2007-12-24	810.25	-	14.3500	1526.00	-	354.50	-
2007-12-21	803.00	810.50	14.2900	1513.00	1516.00	354.00	355.00
2007-12-20	799.50	795.25	13.9700	1512.00	1514.00	354.00	354.00

Internet

Figure 3.10: Gold in US Dollar: Html File

For data on prices of domestic gold bar, in Thai Baht, they are obtained from the web site of a local gold bar shop and are in the format of HTML file and data start at 1 January 1996 – 31 December 2007, in record count is 6,118 records and file count is 12 files, as displayed in Figure 3.11.

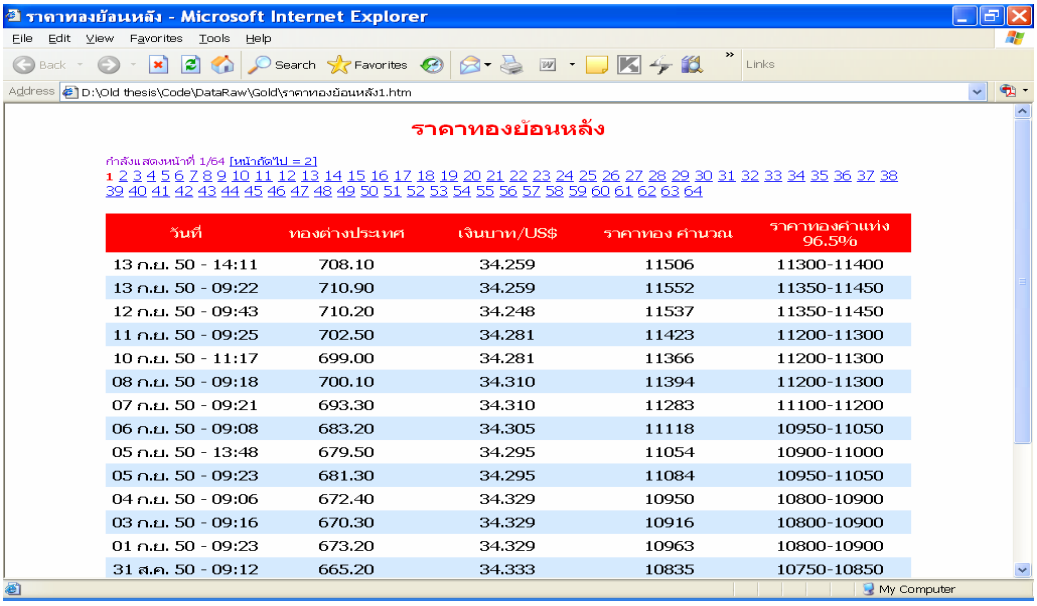


Figure 3.11: Gold in Baht Price: Html File

3.2.3 Foreign Exchange File Structure

The Foreign Exchange rate data are received from the web site of the Bank of Thailand and in the format of Excel file, and data start at 1 January 1996 – 31 December 2007, in record count is 895,650 records and file count is 132 files classified by month and year. The file consists of 6 sheets and 35 currencies, as shown in Table 3.15.

Table 3.15: Foreign Country

Sheet Name	Country
Sheet1 (5)	US. DOLLAR, POUND STERLING, Euro, YEN, HONGKONG DOLLAR
Sheet2 (5)	MALAYSIA RINGGIT, SINGAPORE DOLLAR, BRUNAI DOLLAR, PHILIPPINES PESO, INDONESIA RUPIAH
Sheet3 (5)	INDIA, SWITZERLAND, AUSTRALIA, NEWZEALAND, PAKISTAN
Sheet4 (5)	CANADA, SWEDEN, DENMARK, NORWAY, CHINA
Sheet5 (7)	MAXICO, SOUTH AFRICA, BURMA, KOREA, TAIWAN, KUWAIT, SAUDI ARABIA
Sheet6 (8)	UNITED ARAB EMIRATES, BANGLADESH, CZECHOSLOVAKIA, CAMBODIA, KENYA, LAOS, USSR, VIETNAM

In the first column of each sheet, date data are recorded which in the second column, average US dollar value is recorded. The other column, however, record purchase prices of each currency, which is divided into 2 prices of Sight Draft and Telegram (T/T). The prices selected are those of Sight Draft, of which the pattern is in sheets 1 through 4 while sheets 5 and 6 consist only of purchase prices. A data sample is revealed in Figure 3.12.

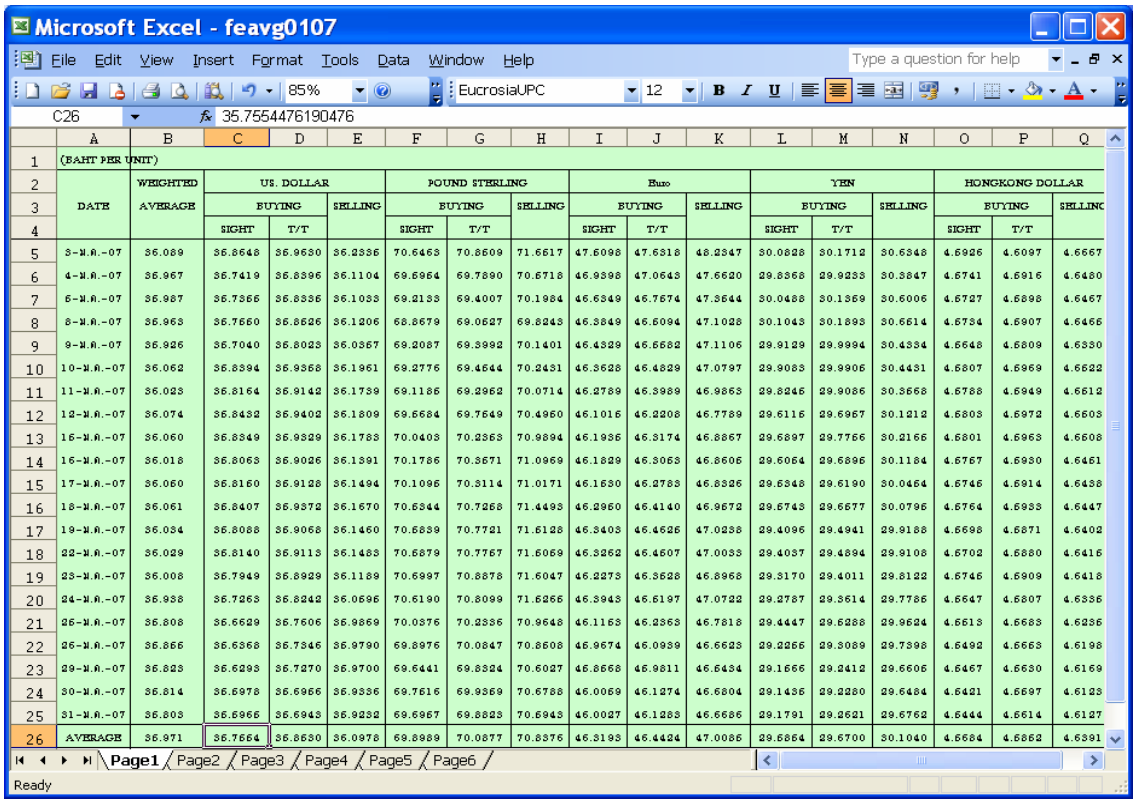


Figure 3.12: Foreign Exchange Rate: Excel File

3.2.4 Interest File Structure

The Bank Interest rate data are derived from the web site of the Bank of Thailand and in the form of Text file, classified by year and bank type. In this research project, only 5 banks are selected. They include Bangkok Bank Plc., Krung Thai Bank Plc., Bank of Ayudhya Plc., Thai Farmers Bank Plc., and Siam Commercial Bank Plc.

1. **Interest Rate for Loan** data consist of date, MOR, MLR, MRR, Highest, Miss an appointment, and Credit Card. A data start at 1 January 1997 – 31 December 2007, in record count is 68,705 records and file count is 59 files, the data sample is displayed in Figure 3.13.

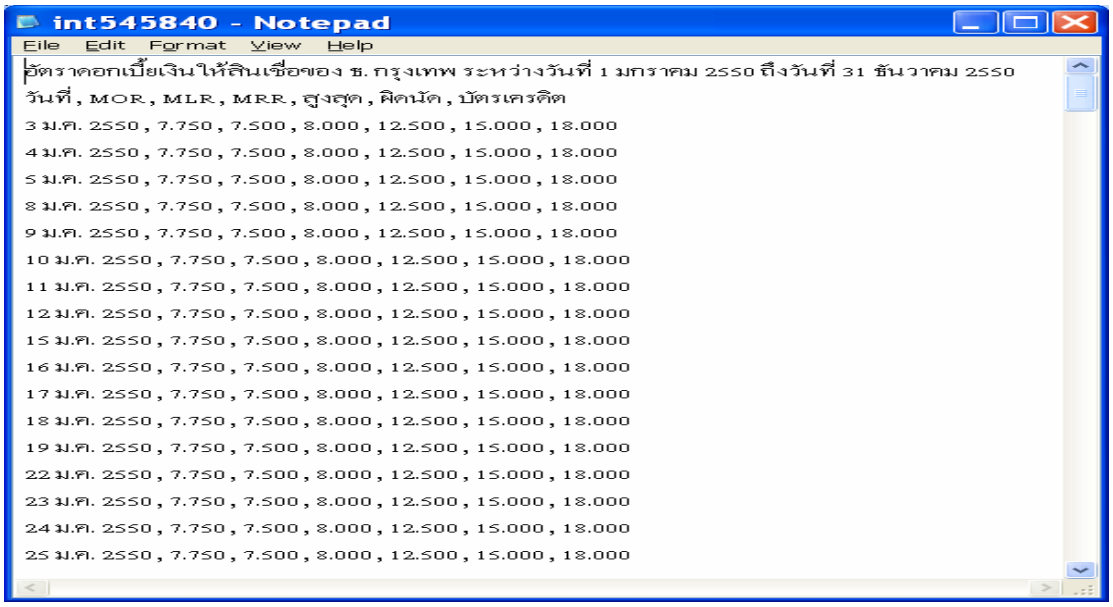


Figure 3.13: Loan Rate: Text File

2. **Interest Rate for Deposit** data are composed of date, Savings, 3-Month fixed, 6-Month fixed, 12-Month fixed, and 24-Month fixed. A data start at 1 January 1997 – 31 December 2007, in record count is 100,600 records and file count is 60 files, the data sample is shown in Figure 3.14.

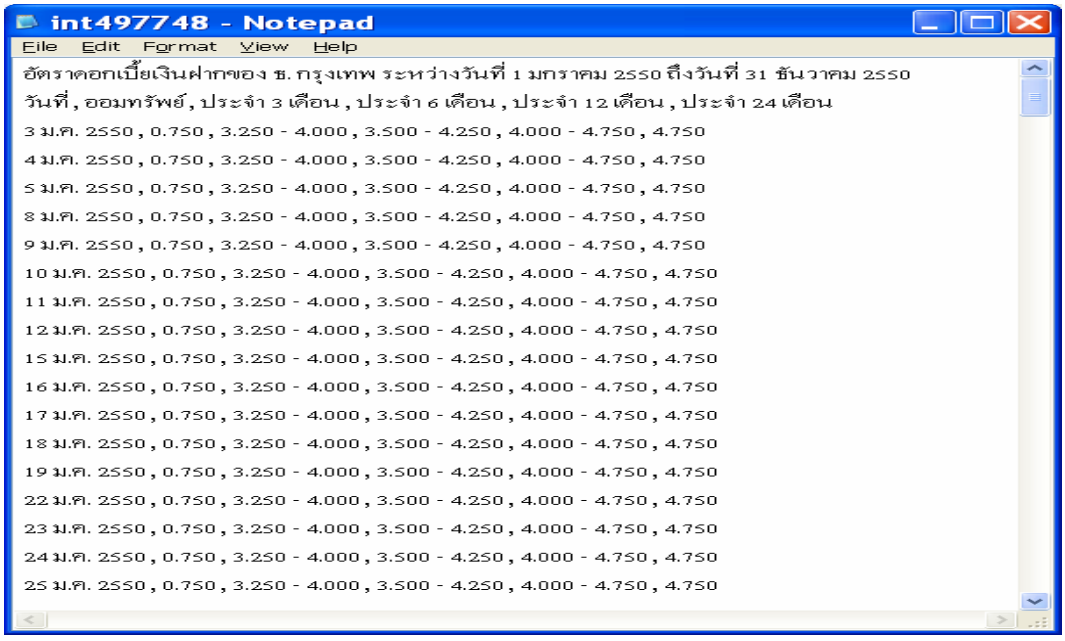


Figure 3.14: Deposit Rate: Text File

3.2.5 External Stock Index File Structure

The External Stock Exchange index data are gained from the web site of the Bank of Thailand and in the Excel file. A data start at 1 January 1996 – 31 December 2007, in record count is 38,900 records and file count is 11 files, the data sample is revealed in Figure 3.15.

	A	B	C	D	E	F	G	H	I	J	K	L
1		Historical SET & External Stock Market Indexes										
2		Source : The Stock Exchange of Thailand & Reuters										
3	DATE	SET	Dow Jones	Financial Times	Nikkei	Hang Seng	Straits Times	KLSE Composite	JSX Composite	PSE Composite	KSE Composite	ALL Ordinaries
4		(Thailand)	(New York)	(London)	(Japan)	(Hong Kong)	(Singapore)	(Malaysia)	(Indonesia)	(Philippines)	(South Korea)	(Australia)
5	1-ม.ค.-03	closed	closed	closed	closed	closed	closed	closed	closed	closed	closed	closed
6	2-ม.ค.-03	351.52	8,607.52	4,009.50	closed	9,365.52	1,335.98	632.43	409.13	1,008.64	635.17	2,996.20
7	3-ม.ค.-03	357.23	8,601.69	4,004.90	8,578.95	9,583.85	1,339.93	633.50	407.51	1,011.50	661.10	3,025.70
8	6-ม.ค.-03	364.15	8,773.57	4,001.40	8,713.33	9,665.96	1,331.62	627.89	398.25	1,015.07	666.71	3,044.80
9	7-ม.ค.-03	365.51	8,740.59	3,957.40	8,656.50	9,652.40	1,318.74	626.17	394.52	1,019.59	652.20	3,043.20
10	8-ม.ค.-03	360.41	8,595.31	3,924.80	8,517.80	9,688.21	1,332.32	625.13	389.41	1,040.42	651.72	3,042.40
11	9-ม.ค.-03	358.76	8,776.18	3,934.00	8,497.93	9,675.41	1,335.09	626.58	396.03	1,035.31	630.40	3,035.60
12	10-ม.ค.-03	358.76	8,776.18	3,934.00	8,497.93	9,675.41	1,335.09	626.58	396.03	1,035.31	630.40	3,035.60
13	13-ม.ค.-03	364.05	8,785.98	3,948.30	8,470.45	9,834.08	1,386.05	651.48	396.23	1,055.05	648.06	3,042.50
14	14-ม.ค.-03	373.33	8,842.62	3,945.60	8,553.06	9,796.31	1,401.37	661.41	407.18	1,061.33	650.05	3,049.00
15	15-ม.ค.-03	371.82	8,723.18	3,887.80	8,611.75	9,873.49	1,386.62	668.21	405.60	1,075.47	648.29	3,049.60
16	16-ม.ค.-03	370.48	8,697.87	3,881.80	8,609.17	9,743.23	1,379.52	670.14	397.11	1,074.45	648.69	3,030.30
17	17-ม.ค.-03	367.16	8,586.74	3,820.60	8,690.25	9,614.59	1,366.83	670.53	401.65	1,081.60	636.46	3,024.70
18	20-ม.ค.-03	371.45	closed	3,778.60	8,558.82	9,552.02	1,363.19	666.36	404.42	1,066.70	634.50	3,028.20
19	21-ม.ค.-03	375.91	8,442.90	3,736.70	8,708.58	9,568.47	1,365.32	675.87	405.16	1,060.07	632.86	3,024.10
20	22-ม.ค.-03	373.17	8,318.73	3,678.00	8,611.04	9,560.29	1,369.13	672.41	405.70	1,052.44	622.49	2,999.10
21	23-ม.ค.-03	376.56	8,369.47	3,622.20	8,790.92	9,584.70	1,364.41	670.78	406.77	1,048.53	625.18	2,997.80
22	24-ม.ค.-03	376.30	8,131.01	3,603.70	8,731.65	9,460.60	1,358.04	668.81	405.34	1,050.42	609.43	closed
23	27-ม.ค.-03	370.80	7,989.56	3,480.80	8,609.47	9,298.67	1,331.25	664.62	393.56	1,038.67	593.09	closed
24	28-ม.ค.-03	374.76	8,088.84	3,490.00	8,525.39	9,325.60	1,339.69	671.63	395.21	1,035.57	600.56	2,940.50

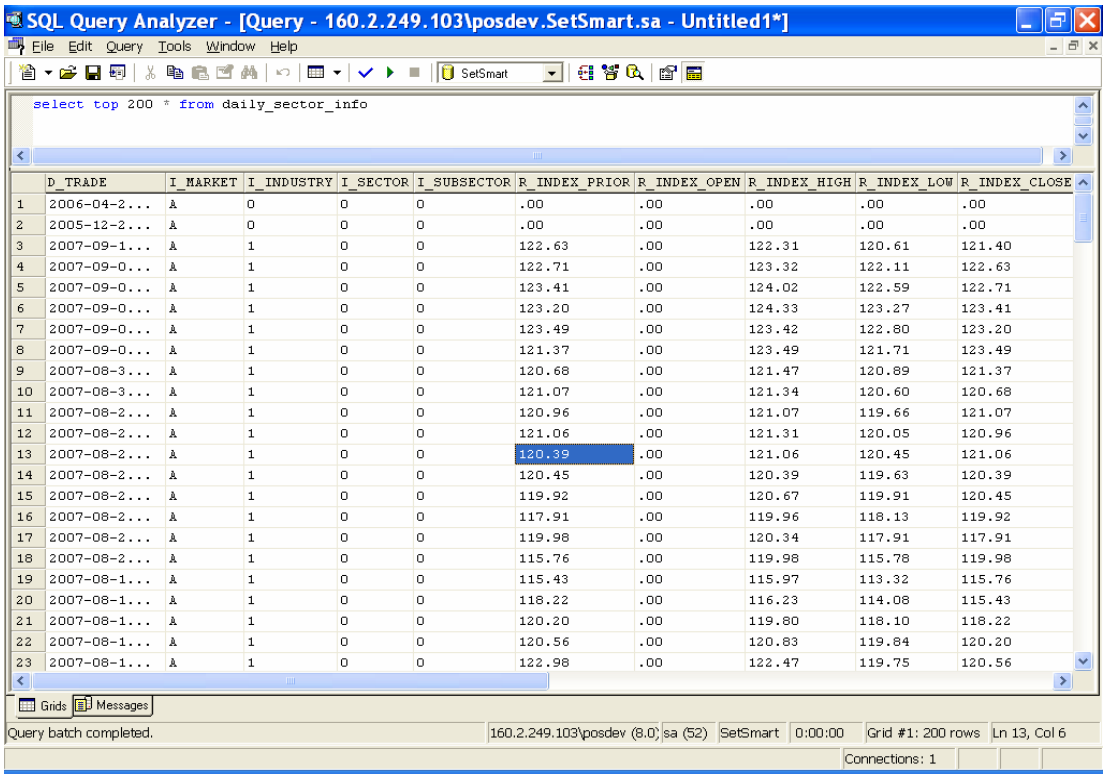
Figure 3.15: External Stock Index: Text File

3.2.6 Stock Index – SET SMART Database Structure

Data are gleaned from the Stock Exchange of Thailand (SET) "SET SMART" Database, which is a database system of stock purchase and sale that SET provides to companies purchasing and selling. 4 Major group of stocks as follows:

1. Company: is tables group related to listed companies.
2. Sector: is tables group related to business types of listed companies.
3. Security: is tables group related to stocks of listed companies.
4. Tables compiling general data.

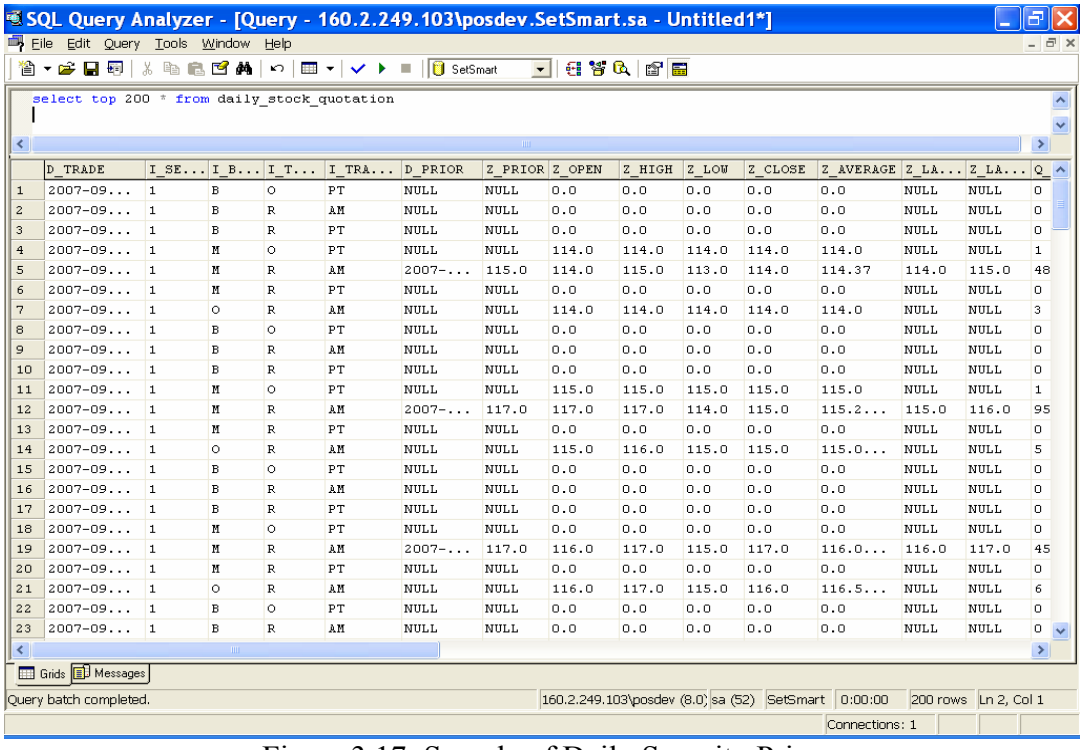
Figure 3.16 reveals a data sample of daily index of Industry and Sector derived from table Daily_Sector_Info and all record count is 200,340 records. Date start at 30/4/1975 to 10/9/2007.



	D_TRADE	I_MARKET	I_INDUSTRY	I_SECTOR	I_SUBSECTOR	R_INDEX_PRIOR	R_INDEX_OPEN	R_INDEX_HIGH	R_INDEX_LOW	R_INDEX_CLOSE
1	2006-04-2...	A	0	0	0	.00	.00	.00	.00	.00
2	2005-12-2...	A	0	0	0	.00	.00	.00	.00	.00
3	2007-09-1...	A	1	0	0	122.63	.00	122.31	120.61	121.40
4	2007-09-0...	A	1	0	0	122.71	.00	123.32	122.11	122.63
5	2007-09-0...	A	1	0	0	123.41	.00	124.02	122.59	122.71
6	2007-09-0...	A	1	0	0	123.20	.00	124.33	123.27	123.41
7	2007-09-0...	A	1	0	0	123.49	.00	123.42	122.80	123.20
8	2007-09-0...	A	1	0	0	121.37	.00	123.49	121.71	123.49
9	2007-08-3...	A	1	0	0	120.68	.00	121.47	120.89	121.37
10	2007-08-3...	A	1	0	0	121.07	.00	121.34	120.60	120.68
11	2007-08-2...	A	1	0	0	120.96	.00	121.07	119.66	121.07
12	2007-08-2...	A	1	0	0	121.06	.00	121.31	120.05	120.96
13	2007-08-2...	A	1	0	0	120.39	.00	121.06	120.45	121.06
14	2007-08-2...	A	1	0	0	120.45	.00	120.39	119.63	120.39
15	2007-08-2...	A	1	0	0	119.92	.00	120.67	119.91	120.45
16	2007-08-2...	A	1	0	0	117.91	.00	119.96	118.13	119.92
17	2007-08-2...	A	1	0	0	119.98	.00	120.34	117.91	117.91
18	2007-08-2...	A	1	0	0	115.76	.00	119.98	115.78	119.98
19	2007-08-1...	A	1	0	0	115.43	.00	115.97	113.32	115.76
20	2007-08-1...	A	1	0	0	118.22	.00	116.23	114.08	115.43
21	2007-08-1...	A	1	0	0	120.20	.00	119.80	118.10	118.22
22	2007-08-1...	A	1	0	0	120.56	.00	120.83	119.84	120.20
23	2007-08-1...	A	1	0	0	122.98	.00	122.47	119.75	120.56

Figure 3.16: Sample of Daily Sector / Industry Index

Figure 3.17 shown a data sample of daily stock purchase and sale derived from table Daily_Stock_Quotation and all record count is 2,284,851 records.



	D_TRADE	I_SE...	I_B...	I_T...	I_TRA...	D_PRIOR	Z_PRIOR	Z_OPEN	Z_HIGH	Z_LOW	Z_CLOSE	Z_AVERAGE	Z_LA...	Z_LA...	Q_
1	2007-09...	1	B	O	PT	NULL	NULL	0.0	0.0	0.0	0.0	0.0	NULL	NULL	0
2	2007-09...	1	B	R	AM	NULL	NULL	0.0	0.0	0.0	0.0	0.0	NULL	NULL	0
3	2007-09...	1	B	R	PT	NULL	NULL	0.0	0.0	0.0	0.0	0.0	NULL	NULL	0
4	2007-09...	1	M	O	PT	NULL	NULL	114.0	114.0	114.0	114.0	114.0	NULL	NULL	1
5	2007-09...	1	M	R	AM	2007-...	115.0	114.0	115.0	113.0	114.0	114.37	114.0	115.0	48
6	2007-09...	1	M	R	PT	NULL	NULL	0.0	0.0	0.0	0.0	0.0	NULL	NULL	0
7	2007-09...	1	O	R	AM	NULL	NULL	114.0	114.0	114.0	114.0	114.0	NULL	NULL	3
8	2007-09...	1	B	O	PT	NULL	NULL	0.0	0.0	0.0	0.0	0.0	NULL	NULL	0
9	2007-09...	1	B	R	AM	NULL	NULL	0.0	0.0	0.0	0.0	0.0	NULL	NULL	0
10	2007-09...	1	B	R	PT	NULL	NULL	0.0	0.0	0.0	0.0	0.0	NULL	NULL	0
11	2007-09...	1	M	O	PT	NULL	NULL	115.0	115.0	115.0	115.0	115.0	NULL	NULL	1
12	2007-09...	1	M	R	AM	2007-...	117.0	117.0	117.0	114.0	115.0	115.2...	115.0	116.0	95
13	2007-09...	1	M	R	PT	NULL	NULL	0.0	0.0	0.0	0.0	0.0	NULL	NULL	0
14	2007-09...	1	O	R	AM	NULL	NULL	115.0	116.0	115.0	115.0	115.0...	NULL	NULL	5
15	2007-09...	1	B	O	PT	NULL	NULL	0.0	0.0	0.0	0.0	0.0	NULL	NULL	0
16	2007-09...	1	B	R	AM	NULL	NULL	0.0	0.0	0.0	0.0	0.0	NULL	NULL	0
17	2007-09...	1	B	R	PT	NULL	NULL	0.0	0.0	0.0	0.0	0.0	NULL	NULL	0
18	2007-09...	1	M	O	PT	NULL	NULL	0.0	0.0	0.0	0.0	0.0	NULL	NULL	0
19	2007-09...	1	M	R	AM	2007-...	117.0	116.0	117.0	115.0	117.0	116.0...	116.0	117.0	45
20	2007-09...	1	M	R	PT	NULL	NULL	0.0	0.0	0.0	0.0	0.0	NULL	NULL	0
21	2007-09...	1	O	R	AM	NULL	NULL	116.0	117.0	115.0	116.0	116.5...	NULL	NULL	6
22	2007-09...	1	B	O	PT	NULL	NULL	0.0	0.0	0.0	0.0	0.0	NULL	NULL	0
23	2007-09...	1	B	R	AM	NULL	NULL	0.0	0.0	0.0	0.0	0.0	NULL	NULL	0

Figure 3.17: Sample of Daily Security Price

Table 3.16: A Summary Information of All Data Sources

Factor	Date	Files	Records
Oil - Internal	1/1/1996 - 30/1/2008	13	26,214
Oil - External	2/6/1986 - 30/1/2008	1	5,541
Gold Bar - Foreign	6/1/1997 - 31/12/2007	11	5,118
Gold Bar - Domestic	1/1/1996 - 31/12/2007	12	6,118
Foreign Exchange	1/1/1996 - 31/12/2007	132	895,650
Interest - Loan	1/1/1997 - 31/12/2007	59	68,705
Interest - Deposit	1/1/1997 - 31/12/2007	60	100,600
External Stock	1/1/1996 - 31/12/2007	11	38,900
Stock Exchange - Daily_Stock_Quotation	30/4/1975 - 10/9/2007	-	2,284,851
Stock Exchange - Daily_Sector_Info	30/4/1975 - 10/9/2007	-	200,340

A summary information of all data sources are given in Table 3.16.

3.3 Data Preparation

It is a stage of data preparation starting with reading all files in a folder defined. Before reading the file, file types have to be examined to find whether they are Excel, Text or HTML files. Then begin reading the file line by line and later put the derived data through the processes of Data Cleaning, Data Integration, Data Transformation and Data Reduction, respectively before finally recording data in data warehouse. All stages of data preparation are illustrated in Figure 3.18. Each stage is discussed in detail below.

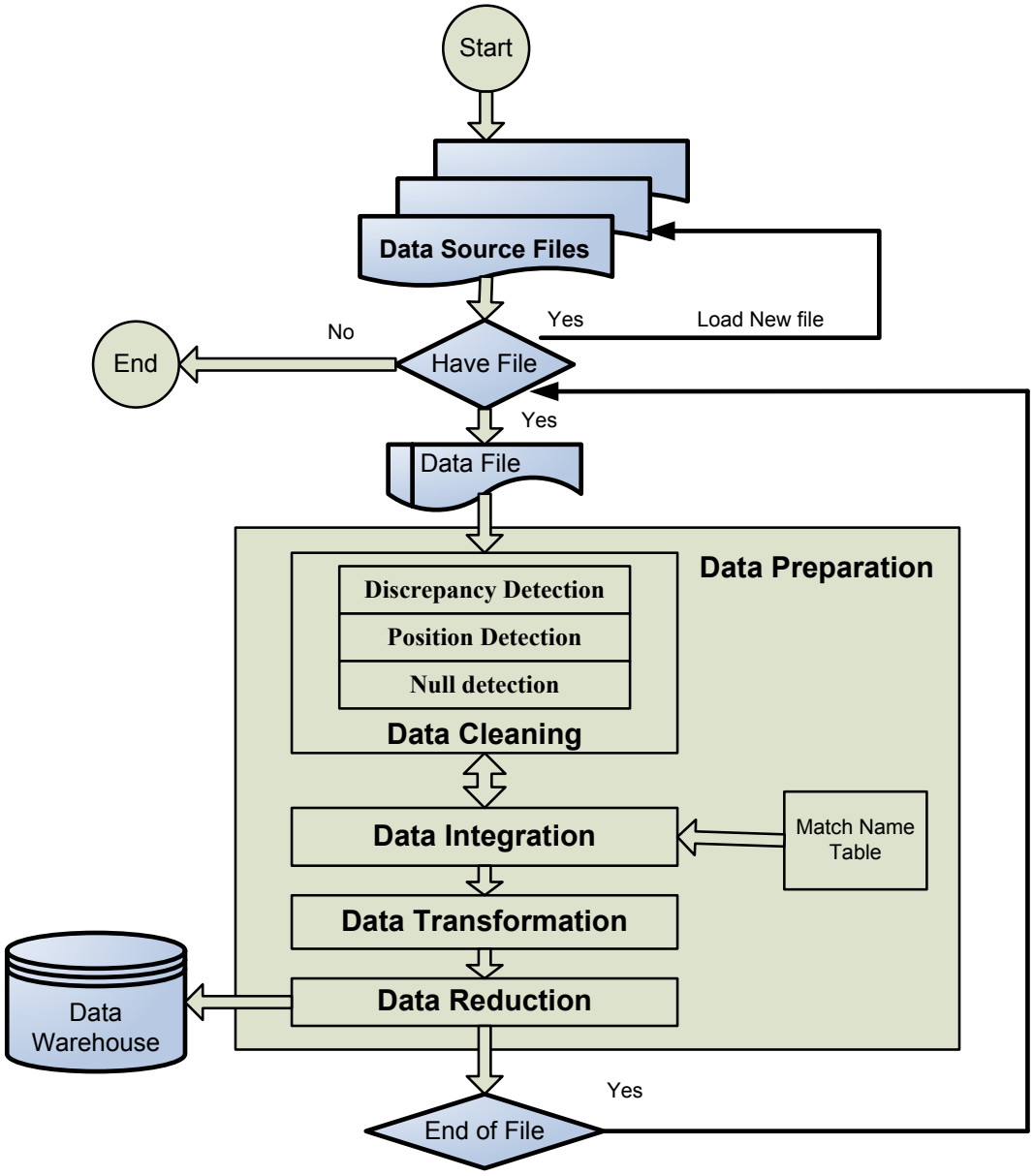


Figure 3.18: Data Preparation Process

1. **Data Cleaning** It is a stage of filtering raw data, only those that are needed, and comprises the following.
    1.1 **Discrepancy Detection** It is the detection of data which contradict, for instance, date pattern like “03-01-06” or “03 January 2006” etc. It needs to be verified whether the text is date data or numerical data. For example, in External Stock index, close index value, which should be figures, turns out to be text “Closed” or “n.a.” instead.
    1.2 **Position Detection** first the position of data in the file has to be identified to find whether in which row, column, page or sheet the data are. Then, the program will be able to cut the data. For example, for data in an Excel file, sheet, row and column have to be identified before cutting data.
    1.3 **Null Detection** It is the detection of data to find whether there is an error such as “20..0” or in the HTML file, Tab Html has to be cut off first so that there are only data left.
2. **Data Integration** It is a stage of compilation of data from different sources to be in the same set such as data matching, which is to integrate data of the same meaning but with different writing into the same set. For instance, oil data of different types such as Lead-free 95, Super or Benzene 95 are integrated to become Benzene 95. Pairs of data are also defined in the “Match Name Table”.
3. **Data Transformation** It is the transformation of the format of raw data into the format stored in database such as from format of date “03-01-06” to “03 Jan 2006”.
4. **Data Reduction** It is the detection to find data repetition since the data obtained may have a repetition value.

Data of the Stock Exchange of Thailand in the data preparation stage are typically different form other data. It is the transformation of data from database of SET SMART to that of data warehouse.

3.3.1 Oil Price Preparation

1. **Data Cleaning** is the detection of oil price data values compiled to find whether they are figures. If not, they are converted into zero.
2. **Data Integration** is the matching of data values because for data in each year, names of oil types are not identical. So, there has to be matching of oil types first, as revealed in Tables 3.17

Table 3.17: Matching of Oil types

Type	Name called for oil type
Diesel	Diesel and solar
Pura Diesel	Pura Diesel and V-Power diesel
Gasohol 91	Gasohol 91
Benzene 91	Benzene 91 and Benzene
Gasohol 95	Gasohol 95
Benzene 95	Benzene 95 and Super

3. **Data Transformation** is to transform date data from the format of "01-12-1998" or "01/12/2548" into the format of "1 December 1998" so that data can be recorded in data warehouse..

3.3.2 Gold Price Preparation

1. **Data Cleaning** is the detection of HTML tag and alteration of numerical data that are "closed", "-", or null to zero.
2. **Data Transformation** is to transform date data from the format of "December 24, 1997" into the format of "24 December 1997".
3. **Data Reduction** since data include prices of Gold, Silver, Platinum and Palladium while prices are divided into ones in the morning and afternoon, other data which are not morning gold price have to be cut off.

3.3.3 Foreign Exchange Rate Preparation

1. **Data Cleaning** is to verify data whether they are figure or not. If not, convert them into zero.
2. **Data Transformation** is to transform date data from the format of "04-ม.ค.-00" into the format of "4 January 2000".

3. **Data Reduction** since purchase prices data comprise 2 prices of Sight Draft and Telegram (T/T), the Telegram data will be cut off and there remain only Sight Draft data.

3.3.4 Interest Rate Preparation

1. **Data Cleaning** is to verify data whether they are figure or not. If not or null data, convert them into zero.
2. **Data Transformation** is to transform date data from the format of "3 ม.ค. 2550" into the format of "3 January 2007".

3.3.5 External Stock Index Preparation

1. **Data Cleaning** is to verify data whether they are figure or not. If not or null data, convert them into zero.
2. **Data Integration** is the matching of data and column in Excel file based on the data matching table of the External Stock exchange, as revealed in Tables 3.18

Table 3.18: Matching of Data of External Stock Index

Column	External Stock Market Name (Country)
C (No 3)	Dow Jones (NewYork)
D (No 4)	Financial Times (London)
E (No 5)	Nikkei (Japan)
F (No 6)	Hang Seng (Hong Kong)
G (No 7)	Straits Times (Singapore)
H (No 8)	KLSE Composite (Malaysia)
I (No 9)	JSX Composite (Indonesia)
J (No 10)	PSE Composite (Philippines)
K (No 11)	KSE Composite (South Korea)
L (No 12)	ALL Ordinaries (Australia)

3. **Data Transformation** is to transform date data from the format of "3-ม.ค.-03" into the format of "3 January 2003".

3.3.6 Stock Index Preparation

1. **Data Selection** is to select tables from database of SET SMART, which are tables of Daily_Sector_Info, Daily_Stock_Quotation, Sector, Security, Company and Market.
2. **Data Cleaning** is to convert null numerical data into zero and delete unused columns.
3. **Data Integration** is to convert data values from one table to another, as shown in Table 3.19

Table 3.19: List of Conversion Table

Set Smart Table	Data warehouse Table
Daily_Stock_Quotation	DailySecurity (Fact)
Daily_Sector_Info	DailyIndustry (Fact)
	DailySector (Fact)
Sector	Indystry (Dimension)
	Sector (Dimension)
Security	Security (Dimension)
Company	Company (Dimension)

3.4 Data Warehouse Design

Data Cubes are generated into 8 schemas while data of Stock Exchange of Thailand are divided into 3 schemas, namely, Stock Industry schema, Sector schema and Security schema.

3.4.1 Oil Schema

It is a Data Cube of oil prices consisting of 1 Fact table and 3 Dimension tables displayed in Tables 3.20 through 3.24 while schema structure is shown in Figure 3.19.

Table 3.20: List of Oil Price Table

No	Table	Type	Mean
1	DailyOil	Fact Table	Fact Table of daily oil price data.
2	DateKey	Dimension	The data of date.
3	OilDim	Dimension	The data of Oil kind.
4	MonthKey	Dimension	The data of Month name

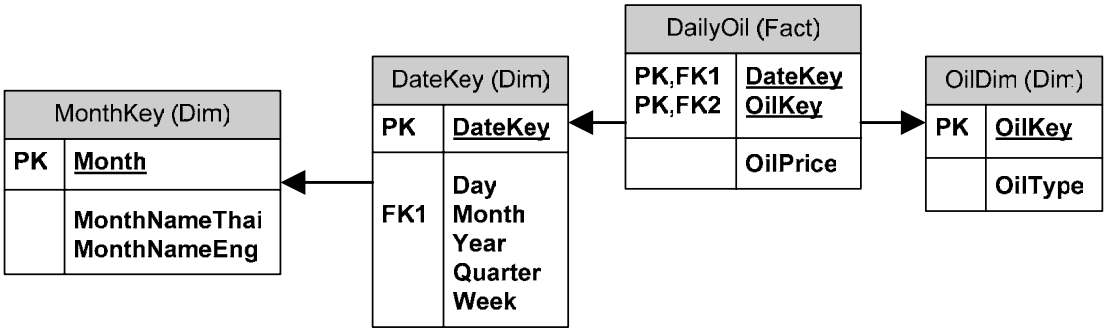


Figure 3.19: Oil Price Snowflake Schema

Table 3.21: DailyOil Fact Table

Fact Table		**DailyOil**		
No	**Field**	**Type**	**Key**	**Mean**
1	DateKey	DateTime	PK	Daily Transaction
2	OilKey	Int	PK	Oil Type Key
3	OilPrice	Float		Summary Measure

Table 3.22: Datekey Dimension Table

Dimension Table		**DateKey**		
No	**Field**	**Type**	**Key**	**Mean**
1	DateKey	DateTime	PK	Date
2	Day	Integer		Day
3	Month	Integer		Month
4	Year	Integer		Year
5	Quarter	Integer		Quarter
6	Week	Integer		Week

Table 3.23: OilDim Dimension Table

Dimension Table		**OilDim**		
No	**Field**	**Type**	**Key**	**Mean**
1	OilKey	Int	PK	Oil Type Key
2	OilType	nVarChar(50)		Type of oil

Table 3.24: Monthkey Dimension Table

Dimension Table		**MonthKey**		
No	**Field**	**Type**	**Key**	**Mean**
1	Month	Int	PK	Month Key
2	MonthNameThai	nVarChar(50)		Month Name in Thai
3	MonthNameEng	nVarChar(50)		Month Name in English

3.4.2 Gold Schema

It is a Data Cube of gold bar prices consisting of 1 Fact table and 3 Dimension tables revealed in Tables 3.25 through 3.27 while schema structure is shown in Figure 3.20.

Table 3.25: List of Gold Price Table

No	Table	Type	Mean
1	DailyGold	Fact Table	Fact Table of daily gold price data.
2	DateKey	Dimension	The data of date.
3	GoldDim	Dimension	The data of Gold kind.
4	MonthKey	Dimension	The data of Month Name

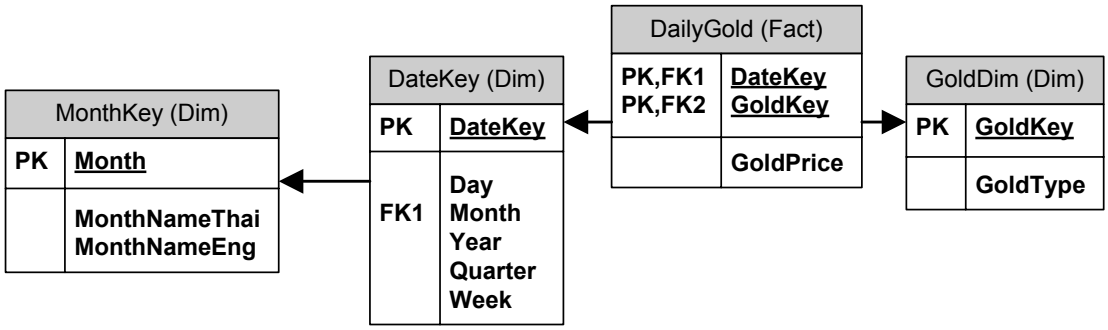


Figure 3.20: Gold Price Snowflake Schema

Table 3.26: DailyGold Fact Table

Fact Table		DailyGold		
No	**Field**	**Type**	**Key**	**Mean**
1	DateKey	DateTime	PK	Daily Transaction
2	GoldKey	Int	PK	Gold Type Key
3	GoldPrice	Float		Measure

Table 3.27: GoldDim Dimension Table

Dimension Table		GoldDim		
No	**Field**	**Type**	**Key**	**Mean**
1	GoldKey	Int	PK	Gold Type Key
2	GoldType	nVarChar(50)		Type of gold

DatekKey Dimension table has detail given in Table 3.22 while MonthKey Dimension table has detail given in Table 3.24.

3.4.3 Foreign Exchange Schema

It is a Data Cube of Foreign Exchange rates consisting of 1 Fact table and 3 Dimension tables shown in Tables 3.28 through 3.30 while schema structure is displayed in Figure 3.21.

Table 3.28: List of Foreign Exchange Table

No	Table	Type	Mean
1	DailyMoney	Fact Table	Fact Table of daily foreign exchange rate data.
2	DateKey	Dimension	The data of date.
3	Country	Dimension	The data of Country kind.
4	MonthKey	Dimension	The data of Month Name

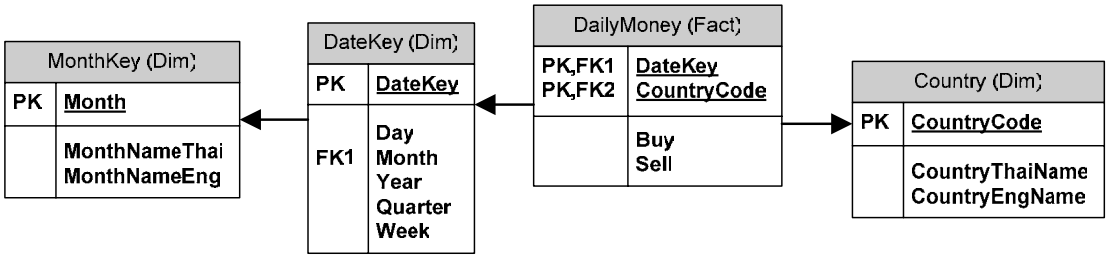


Figure 3.21: Foreign Exchange Snowflake Schema

Table 3.29: DailyMoney Fact Table

Fact Table	DailyMoney			
No	**Field**	**Type**	**Key**	**Mean**
1	DateKey	DateTime	PK	Daily Transaction
2	CountryCode	Int	PK	Country Code Key
3	Buy	Float		Measure
4	Sell	Float		Measure

Table 3.30: Country Dimension Table

Dimension Table	Country			
No	**Field**	**Type**	**Key**	**Mean**
1	Country Code	nVarChar(3)	PK	Country Key
2	Country Name Thai	nVarChar(50)		Country Name in Thai
3	Country Name Eng	nVarChar(50)		Country Name in English

DateKey Dimension table has detail given in Table 3.22 while MonthKey Dimension table has detail given in Table 3.24.

3.4.4 Interest Rate Schema

It is a Data Cube of Bank Interest rates consisting of 2 Fact tables and 5 Dimension tables shown in Tables 3.31 through 3.36 while schema structure is shown in Figure 3.22.

Table 3.31: List of Interest Rate Table

No	Table	Type	Mean
1	DailyDeposit	Fact Table	Fact Table of daily Deposit rate data.
2	DailyLoan	Fact Table	Fact Table of daily Loan rate data.
3	DateKey	Dimension	The data of date.
4	Bank	Dimension	The data of Bank Name kind.
5	MonthKey	Dimension	The data of Month Name
6	DepositDim	Dimension	The data of Deposit kind.
7	LoanDim	Dimension	The data of Loan kind.

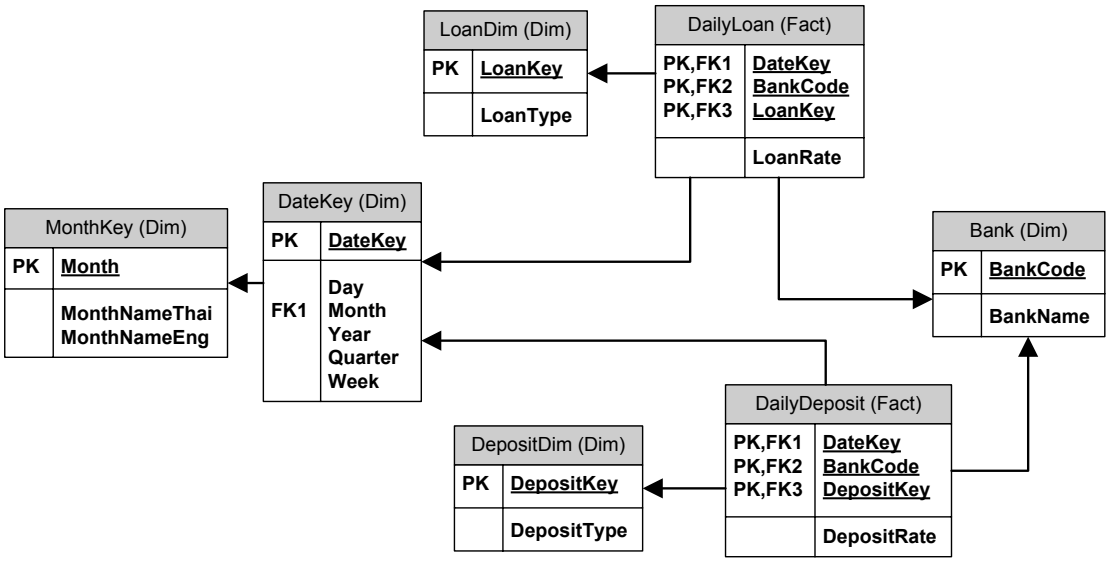


Figure 3.22: Interest Snowflake Schema

Table 3.32: DailyDeposit Fact Table

Fact Table		DailyDeposit		
No	**Field**	**Type**	**Key**	**Mean**
1	DateKey	DateTime	PK	Daily Transaction
2	BankCode	Int	PK	Bank Code Key
3	DepositKey	Int	PK	Deposit Type Key
4	DepositRate	Float		Measure

Table 3.33: DailyLoan Fact Table

Fact Table		**DailyLoan**		
No	**Field**	**Type**	**Key**	**Mean**
1	DateKey	DateTime	PK	Daily Transaction
2	BankCode	Int	PK	Bank Code Key
3	LoanKey	Int	PK	LoanType Key
4	LoanRate	Float		Measure

Table 3.34: Bank Dimension Table

Dimension Table		**Bank**		
No	**Field**	**Type**	**Key**	**Mean**
1	Bank Code	nVarChar(3)	PK	Bank Key
2	Bank Name	nVarChar(50)		Country Name in Thai

Table 3.35: DepositDim Dimension Table

Dimension Table		**DepositDim**		
No	**Field**	**Type**	**Key**	**Mean**
1	DepositKey	Int	PK	Deposit Key
2	DepositType	nVarChar(50)		Deposit Type

Table 3.36: LoanDim Dimension Table

Dimension Table		**LoanDim**		
No	**Field**	**Type**	**Key**	**Mean**
1	LoanKey	Int	PK	Loan Key
2	LoanType	nVarChar(50)		Loan Type

DateKey Dimension table has detail given in Table 3.22 while MonthKey Dimension table has detail given in Table 3.24.

3.4.5 External Stock Index Schema

It is a Data Cube of External Stock indices consisting of 1 Fact table and 4 Dimension tables displayed in Tables 3.37 through 3.39 while schema structure is given in Figure 3.23.

Table 3.37: List of External Stock Index Table

No	**Table**	**Type**	**Mean**
1	DailyExternal	Fact Table	Fact Table of Daily External index Data
2	DateKey	Dimension	The data of date.
3	Country	Dimension	The data of Country Name kind.
4	MonthKey	Dimension	The data of Month Name
5	ExternalDim	Dimension	The data of External Stock Name kind.

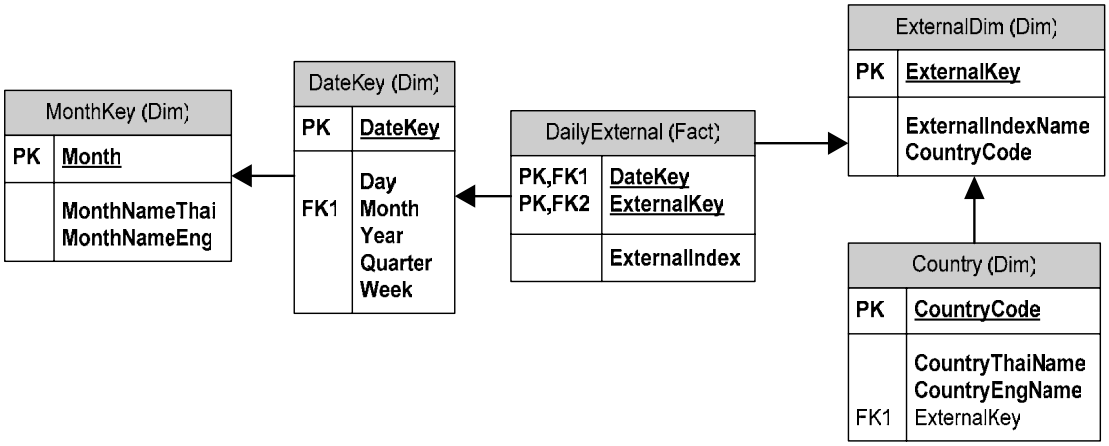


Figure 3.23: External Stock Index Snowflake Schema

Table 3.38: DailyExternal Fact Table

Fact Table		**DailyExternal**		
No	**Field**	**Type**	**Key**	**Mean**
1	DateKey	DateTime	PK	Daily Transaction
2	ExternalKey	Int	PK	External Type Key
3	ExternalIndex	Float		Measure

Table 3.39: ExternalDim Dimension Table

Dimension Table		**ExternalDim**		
No	**Field**	**Type**	**Key**	**Mean**
1	ExternalKey	Int	PK	External Key
2	ExternalIndexName	nVarChar(50)		External Index Name
3	CountryCode	nVarChar(3)		Country Key

DateKey Dimension table has detail given in Table 3.22 while MonthKey Dimension table has detail given in Table 3.24 and Country Dimension table has detail given in Table 3.30.

3.4.6 Industry Index Schema

It is a Data Cube of Stock Industry group consisting of 1 Fact table and 3 Dimension tables displayed in Tables 3.40 through 3.42 while schema structure is displayed in Figure 3.24.

Table 3.40: List of Industry Index Table

No	Table	Type	Mean
1	DailyIndustry	Fact Table	Fact Table of daily Industry index data
2	DateKey	Dimension	The data of date.
3	Industry	Dimension	The data of Industry kind.
4	MonthKey	Dimension	The data of Month Name

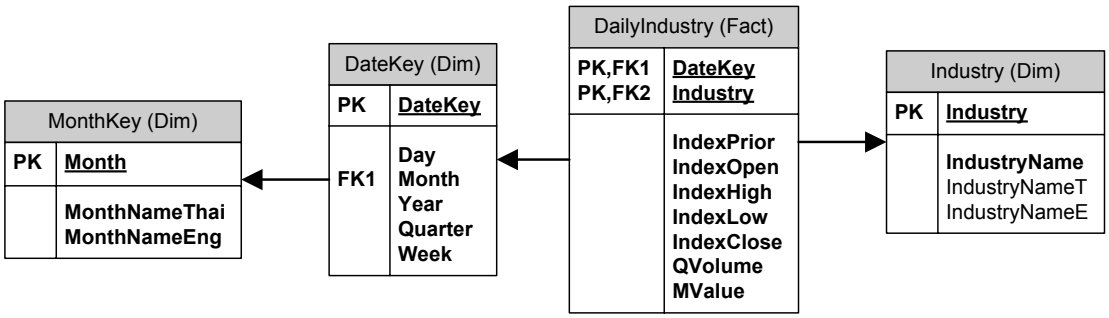


Figure 3.24: Industry Index Snowflake Schema

Table 3.41: DailyIndustry Fact Table

Fact Table	DailyIndustry			
No	**Field**	**Type**	**Key**	**Mean**
1	DateKey	DateTime	PK	Daily Transaction
3	Industry	Int	PK	Industry Code
4	IndexPrior	Float		Prior Index
5	IndexOpen	Float		Open Index
6	IndexHigh	Float		High Index
7	IndexLow	Float		Low Index
8	IndexClose	Float		Close Index
10	QVolume	Float		The quantity of the list trades all
11	MValue	Float		The cost of the list trades all

Table 3.42: Industry Dimension Table

Dimension Table	Industry			
No	**Field**	**Type**	**Key**	**Mean**
2	Industry	Int	PK	Industry Data
3	IndustryName	nVarChar(8)	PK	Industry Name
4	IndustryNameT	nVarChar(200)		Industry Full Name – Thai
5	IndustryNameE	nVarChar(200)		Industry Full Name – English

DateKey Dimension table has detail given in Table 3.22 while MonthKey Dimension table has detail given in Table 3.24.

3.4.7 Sector Index Schema

It is a Data Cube of Business Sector index consisting of 1 Fact table and 4 Dimension tables revealed in Tables 3.43 through 3.45 while schema structure is given in Figure 3.25.

Table 3.43: List of Sector Index Table

No	Table	Type	Mean
1	DailySector	Fact Table	Fact Table of daily Sector index data
2	DateKey	Dimension	The data of date.
3	Sector	Dimension	The data of Sector kind.
4	Industry	Dimension	The data of Industry kind.
5	MonthKey	Dimension	The data of Month Name

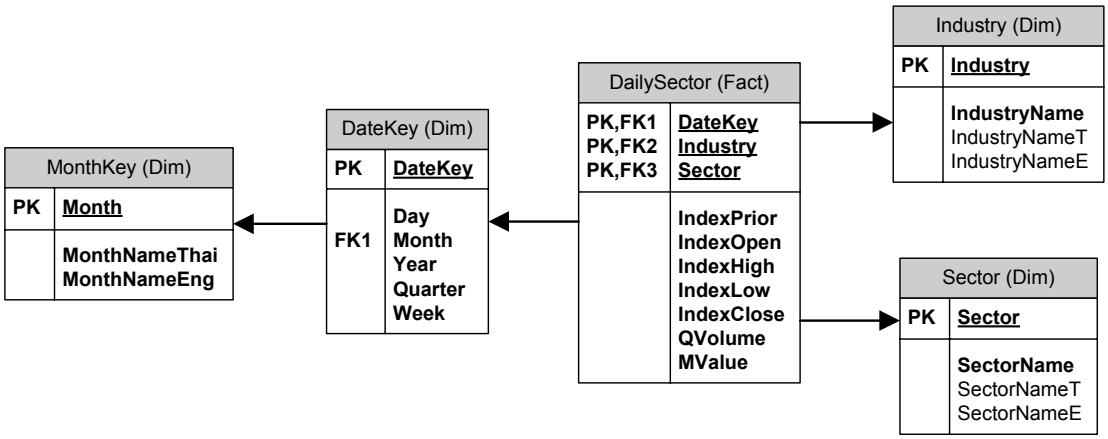


Figure 3.25: Sector Index Snowflake Schema

Table 3.44: DailySector Fact Table

Fact Table		DailySector		
No	**Field**	**Type**	**Key**	**Mean**
1	DateKey	DateTime	PK	Daily Transaction
2	Industry	Int	PK	Industry Code
3	Sector	Int	PK	Sector Code
4	IndexPrior	Float		Prior Index
5	IndexOpen	Float		Open Index
6	IndexHigh	Float		High Index
7	IndexLow	Float		Low Index
8	IndexClose	Float		Close Index
9	QVolume	Float		The quantity of the list trades all
10	MValue	Float		The cost of the list trades all

Table 3.45: Sector Dimension Table

Dimension Table		Sector		
No	**Field**	**Type**	**Key**	**Mean**
1	Sector	Int	PK	Sector Code
2	SectorName	nVarChar(8)	PK	Sector Code Name
3	SectorNameT	nVarChar(200)		Sector Full Name – Thai
4	SectorNameE	nVarChar(200)		Sector Full Name – English

DateKey Dimension table has detail given in Table 3.22 while MonthKey Dimension table has detail given in Table 3.24 and Industry Dimension table has detail given in Table 3.42.

3.4.8 Security Index Schema

It is a Data Cube of Securities purchase and sale index consisting of 1 Fact table and 4 Dimension tables displayed in Tables 3.46 through 3.49 while schema structure is shown in Figure 3.26.

Table 3.46: List of Security Index Table

No	Table	Type	Mean
1	DailySecurity	Fact Table	Daily Security Trade
2	DateKey	Dimension	The data of trend date
3	Security	Dimension	The data of Security
4	Company	Dimension	The data of Security Company
5	MonthKey	Dimension	The data of month name

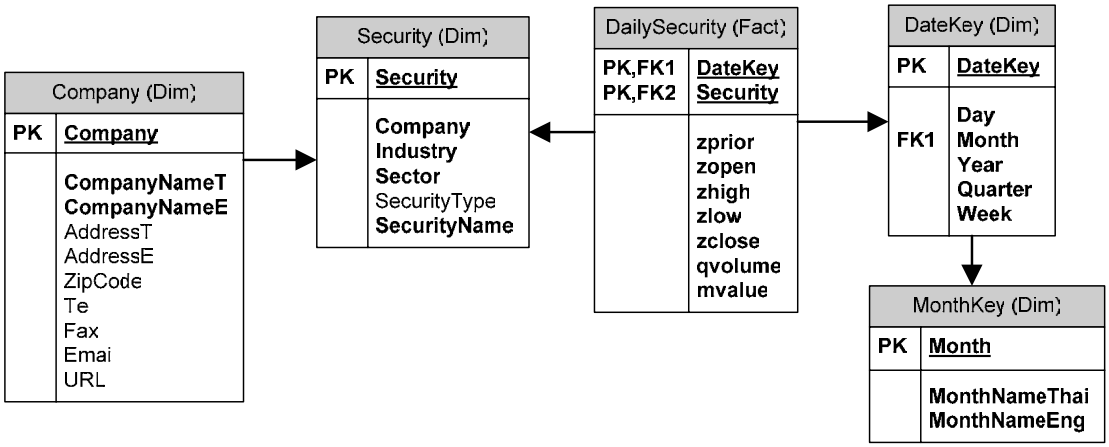


Figure 3.26: Security Snowflake Schema

Table 3.47: DailySecurity Fact Table

Fact Table		**DailySecurity**		
No	**Field**	**Type**	**Key**	**Mean**
1	DateKey	DateTime	PK	Daily Transaction
2	Security	Int	PK	Security Code
3	ZPrior	Float		Prior Price
4	ZOpen	Float		Open Price
5	ZHigh	Float		High Price
6	ZLow	Float		Low Price
7	ZClose	Float		Close Price
8	QVolume	Float		Volumn of daily trade
9	MValue	Float		Value of daily trade

Table 3.48: Security Dimension Table

Dimension Table		**Security**		
No	**Field**	**Type**	**Key**	**Mean**
1	Security	Int	PK	Security Code
2	Company	Int		Security Company
3	Industry	Int		Industry Code
4	Sector	Int		Sector Code
5	SecurityName	nVarChar(8)		Security Name
6	SecurityNameT	nVarChar(200)		Security Full Name – Thai
7	SecurityNameE	nVarChar(200)		Security Full Name – English

Table 3.49: Company Dimension Table

Dimension Table		**Company**		
No	**Field**	**Type**	**Key**	**Mean**
1	Company	Int	PK	Security Company
2	CompanyNameT	nVarChar(200)		Company Name – Thai
3	CompanyNameE	nVarChar(200)		Company Name – English
4	AddressT	nVarChar(200)		Company Address – Thai
5	AddressE	nVarChar(200)		Company Address – English
6	ZipCode	nVarChar(50)		Zip Code
7	Tel	nVarChar(50)		Telephone Number

DateKey Dimension table has detail given in Table 3.22 while MonthKey Dimension table has detail given in Table 3.24.

3.5 Data Analysis

It is the analysis of data by detecting data similarity values and movement attributes of time series data. The analysis structure includes 3 major processes of OLAP Report, Time Series Analysis and Similarity Analysis, as revealed in Figure 3.27.

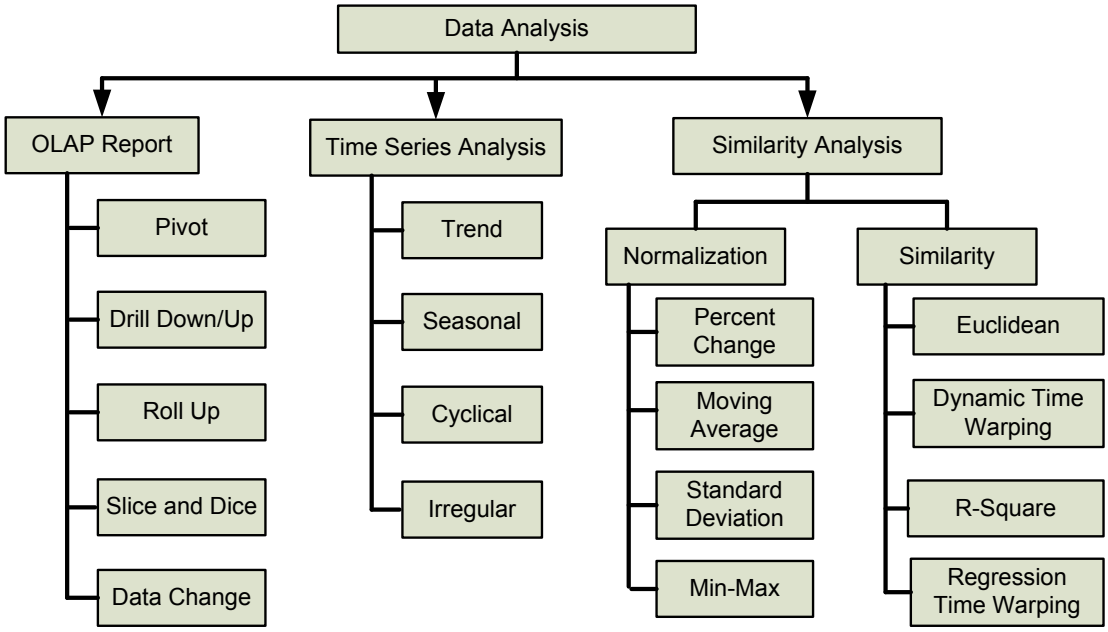


Figure 3.27: Data Analysis Structure

3.5.1 OLAP Report

It is a report on displaying the consequence of preliminary Data Analysis using statistic principle for descriptive analysis. The results are presented in the pattern of tables that are capable of OLAP operation (Pivot, Drill Down/Up, Roll Up as well as Slice and Dice), including data change graph displaying alteration of data compared to previous ones.

3.5.2 Time Series Analysis

It is the analysis of components of time series data as the following:

1. **Trend** is the analysis of the long-term movement of data to find whether it has an increasing or decreasing trend.
2. **Seasonal Variation** is the analysis of data alteration incurred by seasonal influences which occur repetitiously at the same time every year.

3. **Cyclical Variation** is the analysis of repeated movement of data, which is similar to Seasonal Variation analysis. The difference is that the data movement cycle is over year.
4. **Irregular Variation** is the analysis of data movement with no definite pattern. The data attribute stems from irregular events.

3.5.3 Similarity Analysis

It is to calculate similarity values of 2 sets of time series data to find how approximate they are stages of analysis include Data Normalization and Similarity method, of which detail are described below.

1. **Data Normalization** is a process of data scale adjustment so that data are in the same or similar scale, through the following methods.
    1.1 **Previous Data Change** is data scale adjustment via calculation of data change rate.
    1.2 **Moving Average** is data scale adjustment via calculation of data average values.
    1.3 **Standard Deviation** is data scale adjustment via calculation of data standard values.
    1.4 **Min-Max** is data scale adjustment by using minimum data value and maximum data value.
2. **Similarity Method** is a method of calculating data similarity values to find how approximate the 2 sets of data are. The method can be carried out by the following techniques.
    2.1 **Euclidean** is to compute the distance between 2 points.
    2.2 **Dynamic Time Warping** is to compute the distance between points in the pattern of matrix.
    2.3 **R-Square** is to calculate the relationship of 2 sets of data via Simple Linear Regression technique.
    2.4 **Regression Time Warping** is to calculate the relationship of 2 sets of data via an application of R-Square technique and Dynamic Time Warping technique.

CHAPTER IV
SYSTEM IMPLEMENTATION

This chapter discusses hardware and software used for implementation and the prototype implementation. Details of each section are described below.

4.1 Hardware and Software Used For Implementation

Hardware used to develop the system utilizes Intel Centrino Pentium M 1.3 GHz with 256 Mbyte RAM.

Software used for the system development includes Microsoft Visual Studio 2005 tool, with the system running on windows application. Other software consists of the following.

1. **Microsoft Window XP SP2**: is operation system used to develop the system.
2. **Microsoft SQL Server 2005 SP2**: is database system used to store data and Data Cubes of Data Warehouse.
3. **Microsoft Business Intelligence Service**: is a set of software command containing the operating program of Data Warehouse.
4. **Microsoft Analysis Service**: is warehouse software used to generate Data Cube.
5. **MDX – Multiple Dimensional Expression Language**: is warehouse language tool, which is the format of query expression language used to retrieve data from Data Cubes.
6. **Microsoft C-Sharp (C#) 2005**: is programming software used to create the program.

4.2 Economic Factors Analysis Prototype Implementation

The system structure consists of 3 main parts of Data Preparation, Data Cube Creation and Data Analysis, as illustrated in Figure 4.1. For Data Preparation and Data Analysis, they are developed with Microsoft C-Sharp 2005 in the program called "Economic Factors Analysis" shown in Figure 4.2 with respect to Data Cube Creation, it is developed with Microsoft Analysis Service in the program called "Economic Data Cube" displayed in Figure 4.3.

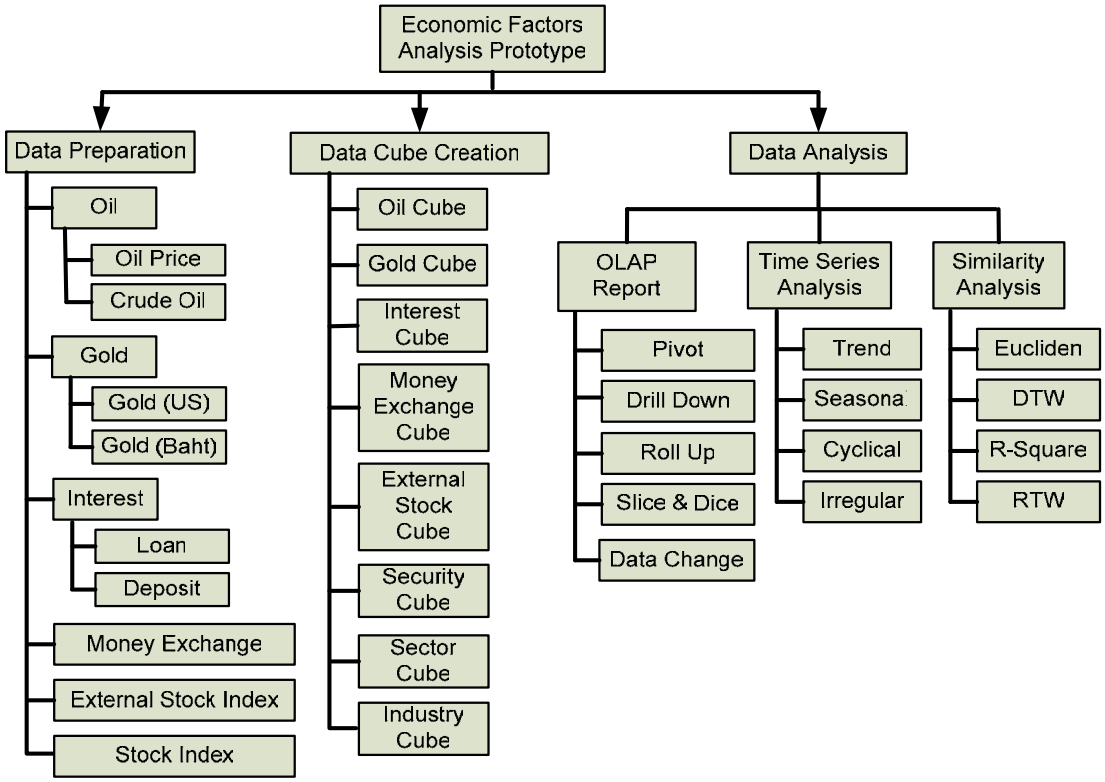


Figure 4.1: Prototype Structure

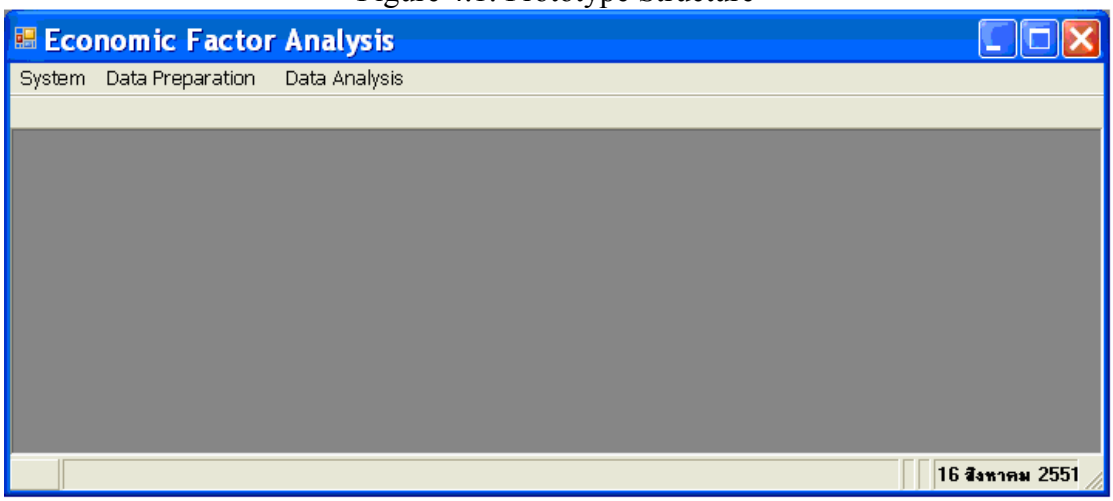


Figure 4.2: Economic Factors Analysis Program

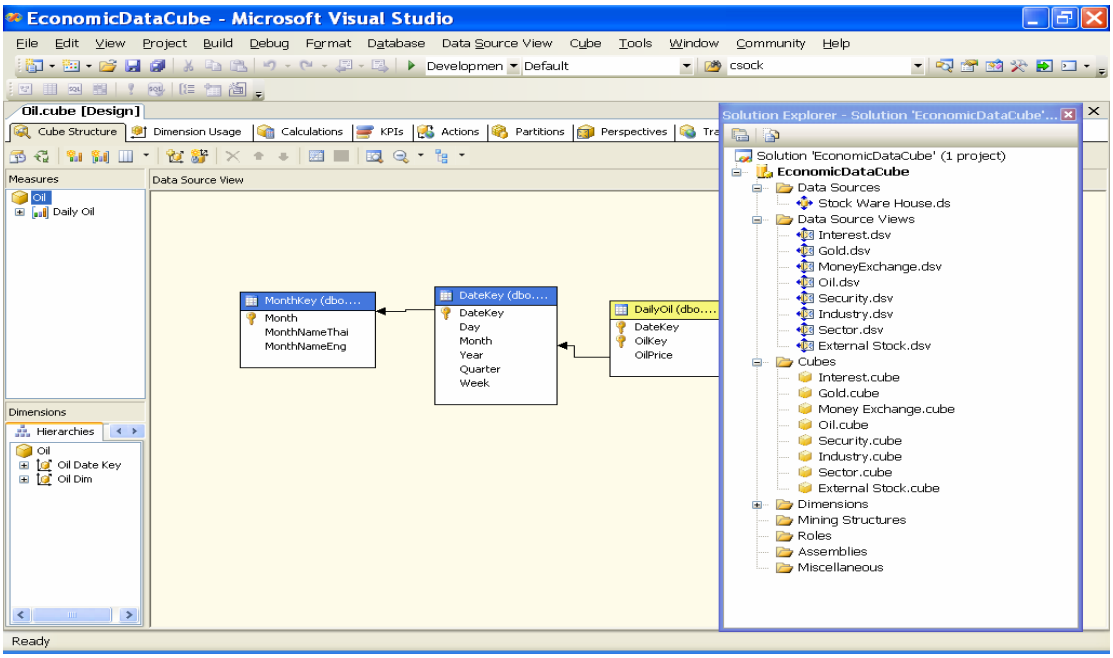


Figure 4.3: Economic Data Cube (Data Cube Creation) Program

4.2.1 Data Preparation

It is the part serving to prepare raw data to become data that can be recorded into Data Warehouse. The data obtained will be stored in temporary data before they are retrieved by the "Economic Data Cube" program to generate Data Cube in Data Warehouse, as illustrated in Figure 4.4. Data Preparation includes the following.

1. **Oil Preparation** is the part preparing oil price data consisting of domestic oil prices and global crude oil and refined oil prices.
2. **Gold Preparation** is the part preparing gold bar price data consisting of domestic gold bar prices (in Thai baht) and foreign gold bar prices (in US dollar).
3. **Interest Preparation** is the part preparing Bank Interest rate data consisting of Interest rate for loan and deposit.
4. **Foreign Exchange Preparation** is the part preparing Foreign Exchange rate data.
5. **External Stock Preparation** is the part preparing External Stock index data.

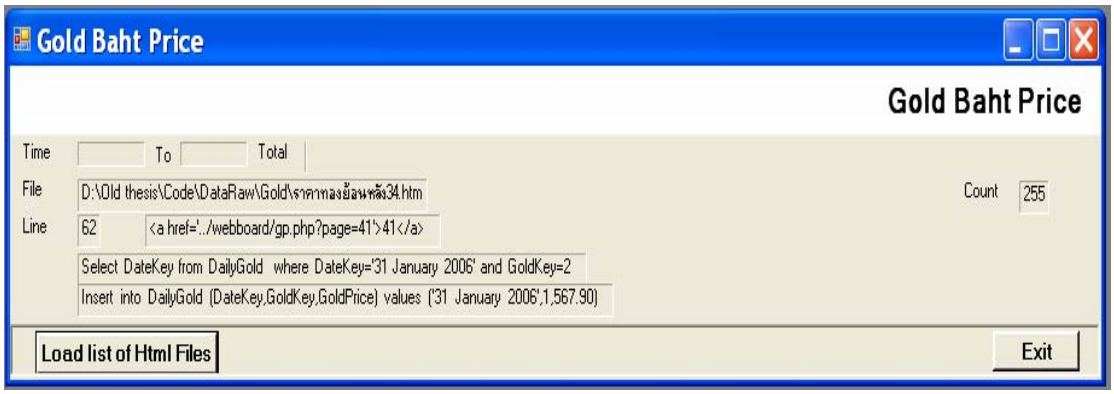


Figure 4.4: Data Preparation

4.2.2 Data Cube Creation

It is the part serving to generate Data Cubes from data received from the process of Data Preparation. This program is created from Microsoft Analysis Services, as illustrated in Figure 4.5.

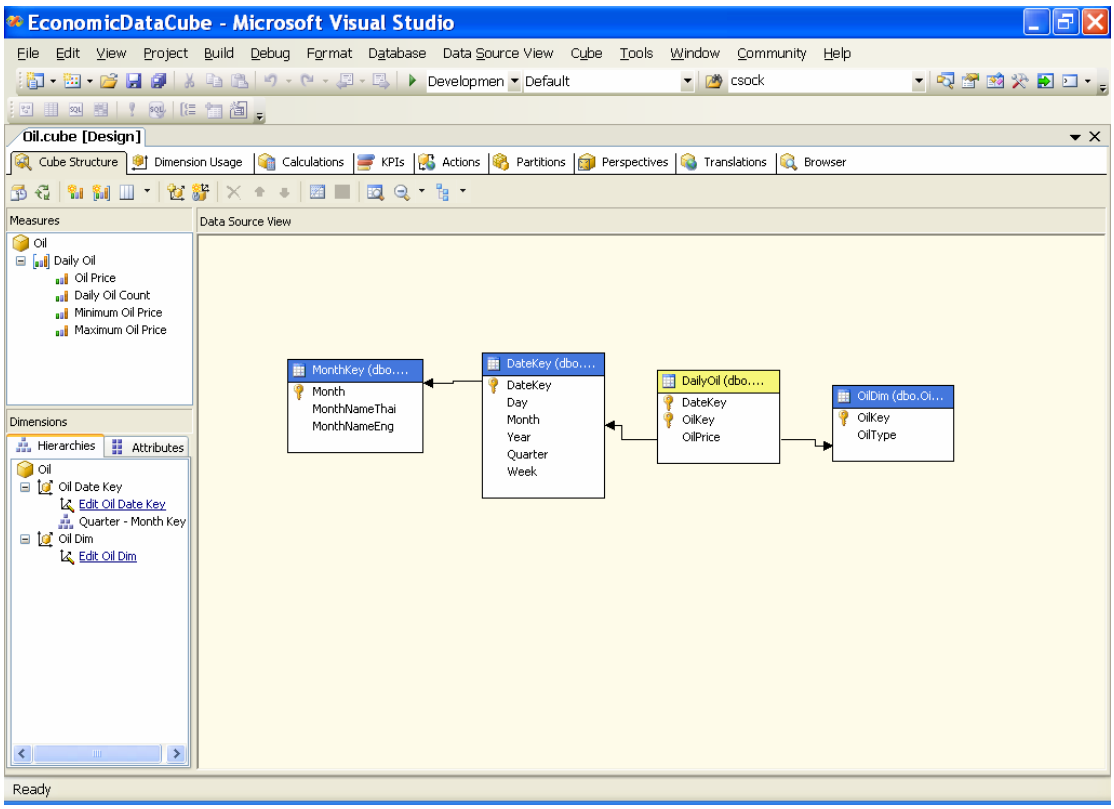


Figure 4.5: Data Cube Creation

.

4.2.3 Data Analysis

It is the analysis of data using the principle of Data Mining, and also applies the techniques of Time Series Analysis and Similarity Analysis.

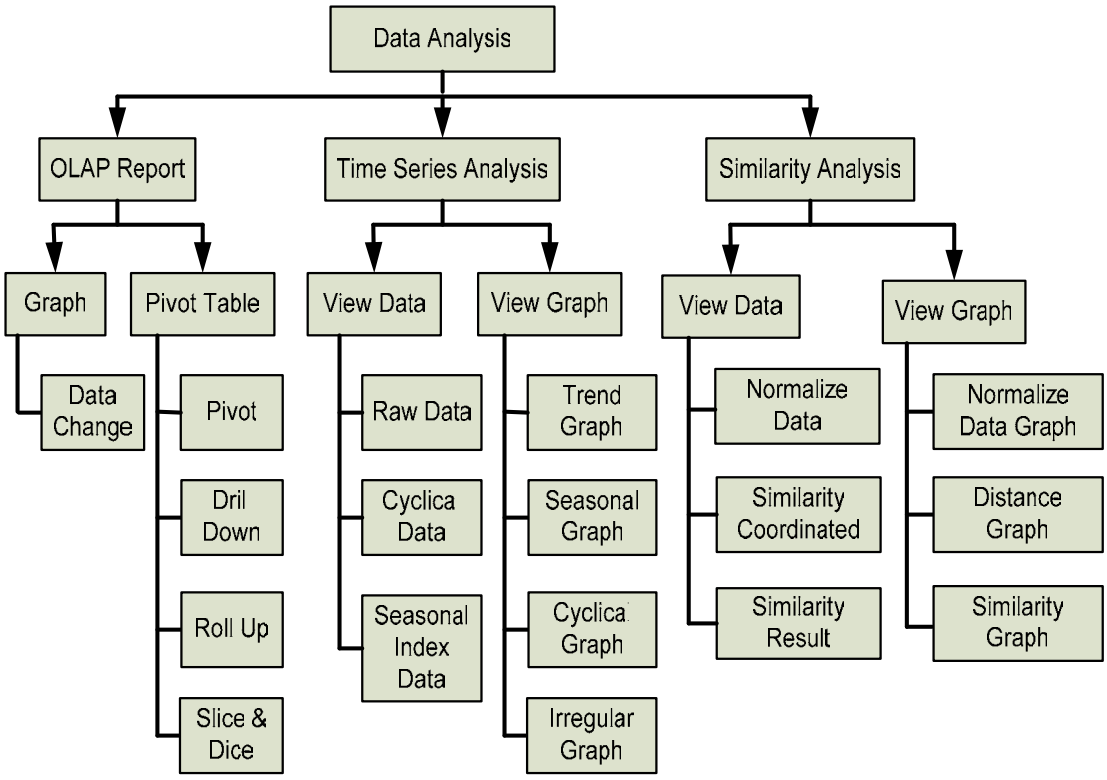


Figure 4.6: Data Analysis

Based on Figure 4.6, the structure of Data Analysis comprises 3 main parts of OLAP report, Time Series Analysis and Similarity Analysis. Each part is discussed in detail below.

4.2.3.1 OLAP Report

It is the part serving to make a report based on data obtained from the OLAP process. The consequences are in 2 forms of graph and pivot table.

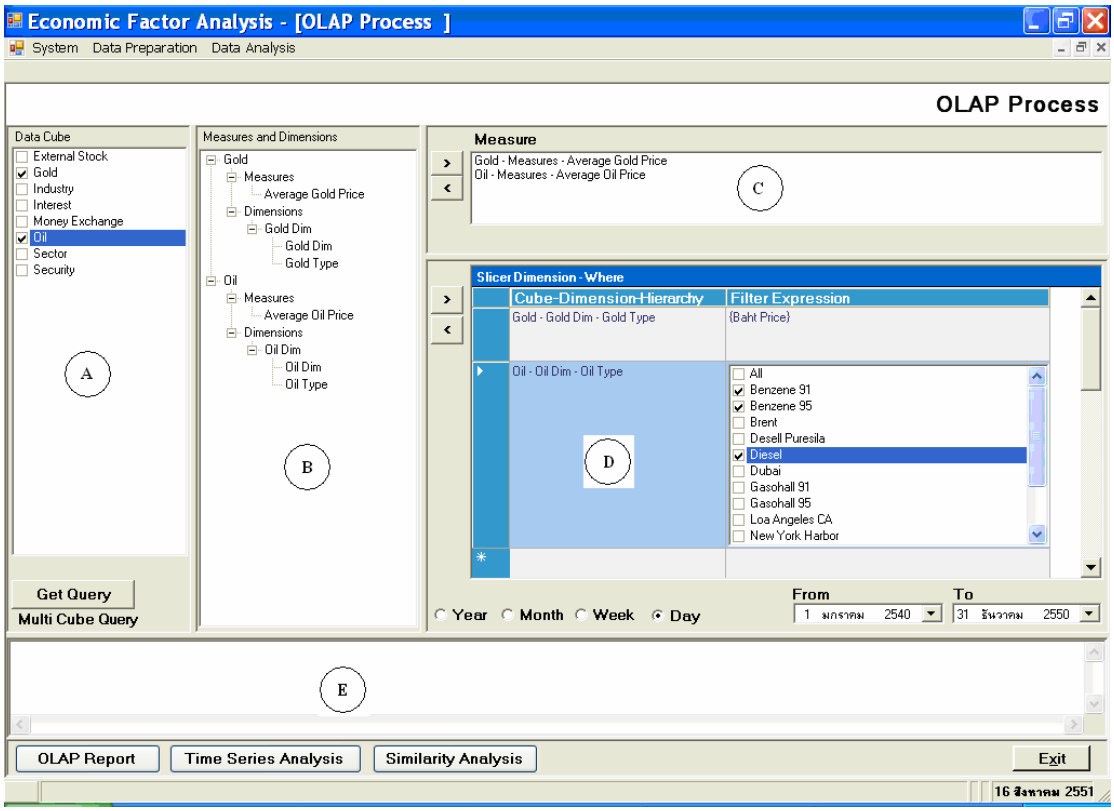


Figure 4.7: OLAP Process

An interface of OLAP reports, according to Figure 4.7, includes the following.

A. **Data Cube** is the part used to select Data Cubes. In this regard, many Data Cubes may be selected.
B. **Measure and Dimension** is the part displaying Fact tables and Dimension tables. The Measure or Dimension key may be selected to generate items.
C. **Measure Selection** is the part showing a name list of Measure keys.
D. **Dimension Selection** is the part used to select dimension field.
E. **MDX Cube Query Expression** is the part displaying query expression derived from the condition selection.

1. **Graph Data change graph** is the graph illustrating a change of data compared to previous data, as seen in Figure 4.8.

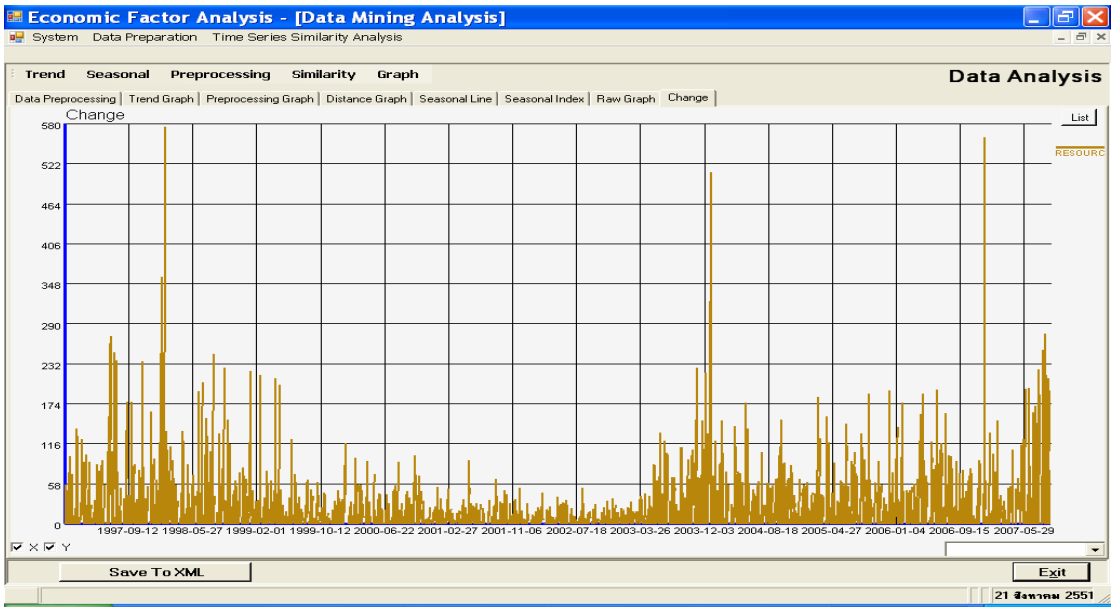


Figure 4.8: Percentage Change Graph

2. **Pivot Table** is the part displaying a pivotal view, allowing an alteration of (the angle of) data view. Besides, it is able to summarize data both in vertical and horizontal direction, and present graphs. Along with data tables, as illustrated in Figures 4.9 and 4.10.

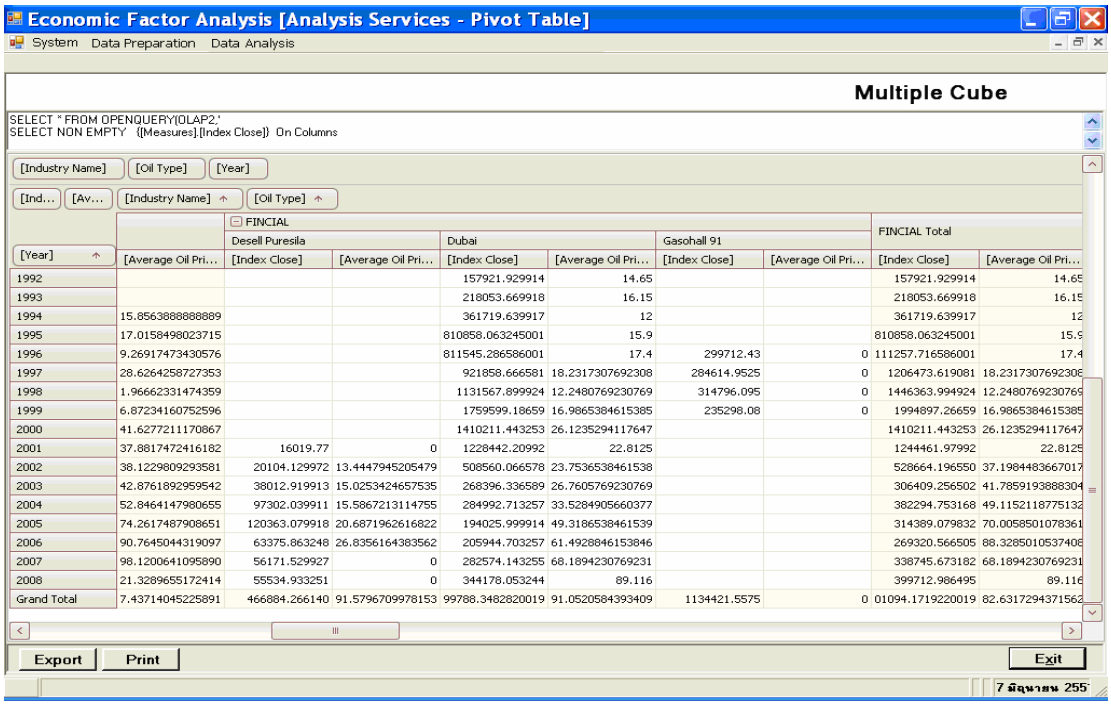


Figure 4.9: Pivot Table

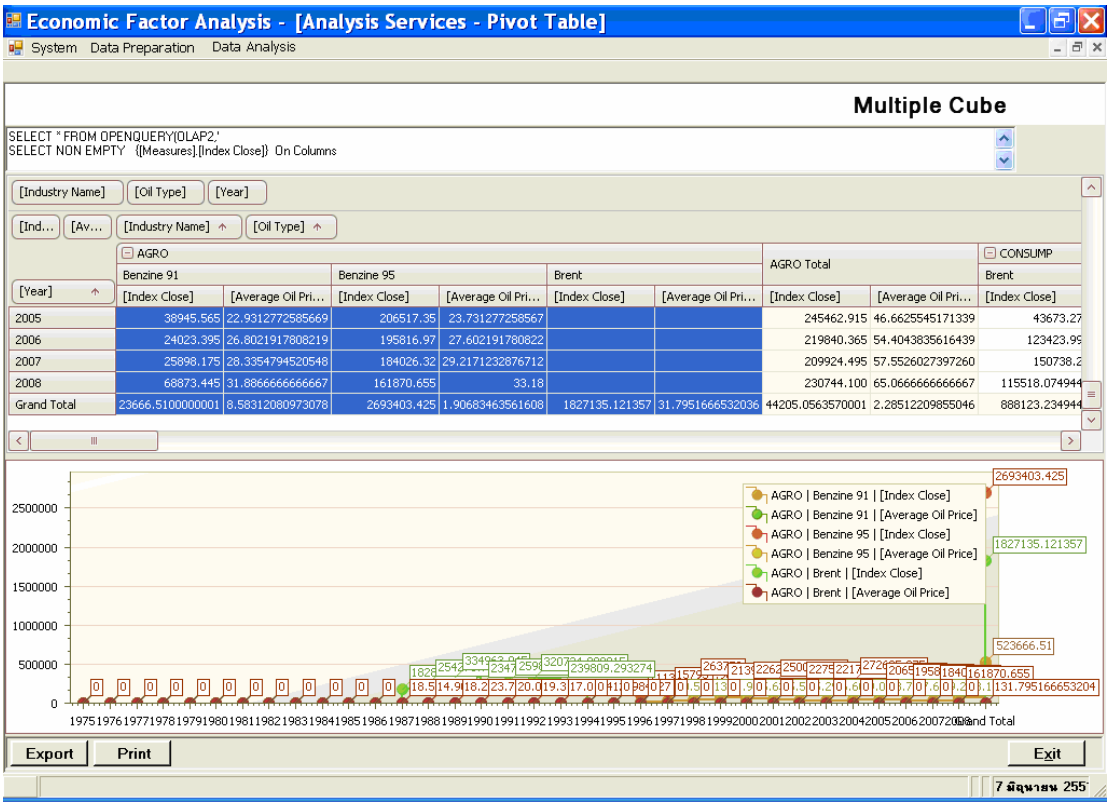


Figure 4.10: Pivot Table and Graph

4.2.3.2 Time Series Analysis

It consists of 2 patterns of Data View and Graph View, of which details are given below.

1. **Data View** is the part displaying data details, as revealed in Figure 4.11, and consists of 3 following parts.
    1.1 **Raw Data** is the part displaying data on daily average value received from Data Warehouse and through the OLAP process. Data can be filtered based on year and month.
    1.2 **Seasonal Index Data** is the part showing the consequence of the seasonal variation process. Data may be filtered based on Dimension key (column).
    1.3 **Cyclical Data** is the part revealing the outcome of the cyclical variation process. Likewise, data may be filtered based on Dimension key (column).

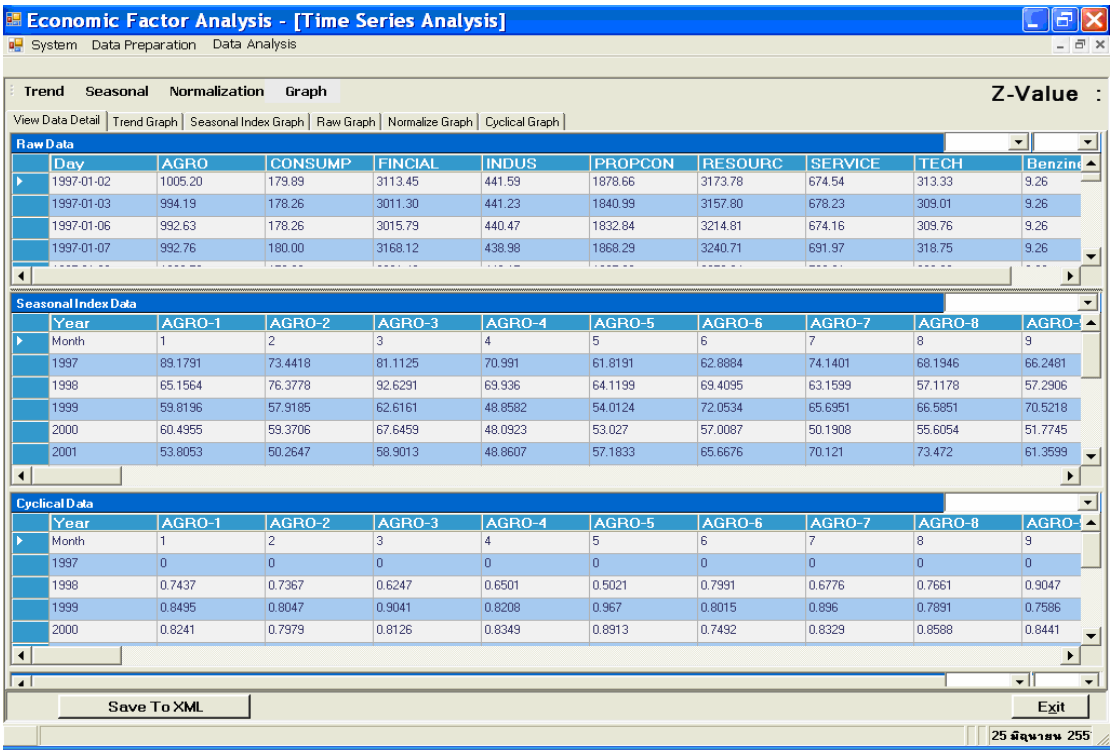


Figure 4.11: Data View

2. **Graph View** is the part displaying data graphs, which include the following.

2.1 **Trend Graph** is the graph showing the trend of data movement obtained from the trend process, as illustrated in Figure 4.12.

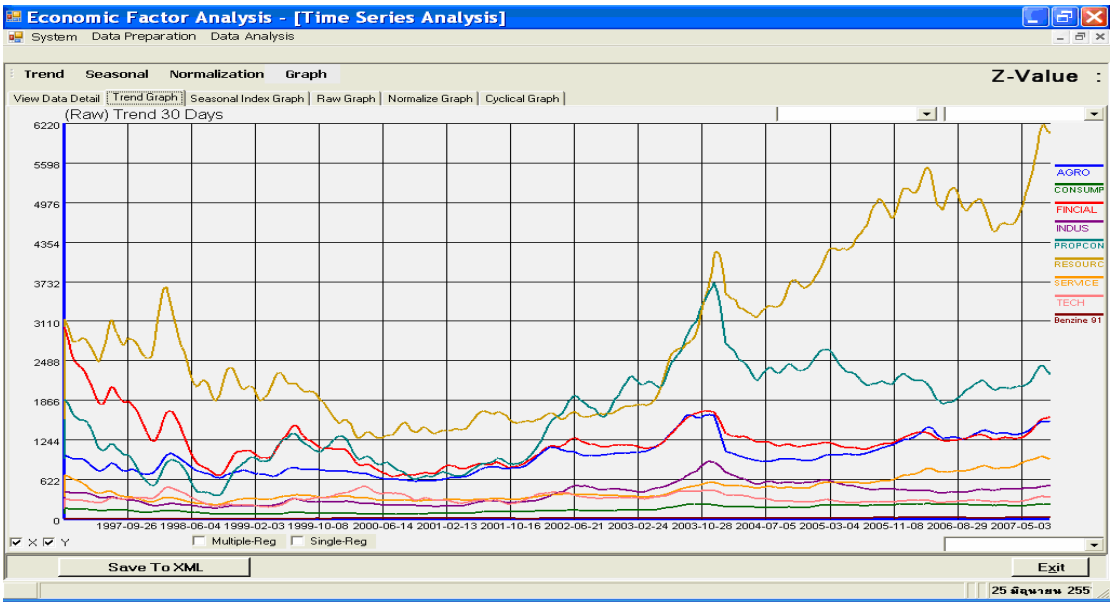


Figure 4.12: Trend Graph

2.2 **Seasonal Index Graph** is the graph showing Seasonal Index values derived from the seasonal variation process, as illustrated in Figure 4.13.

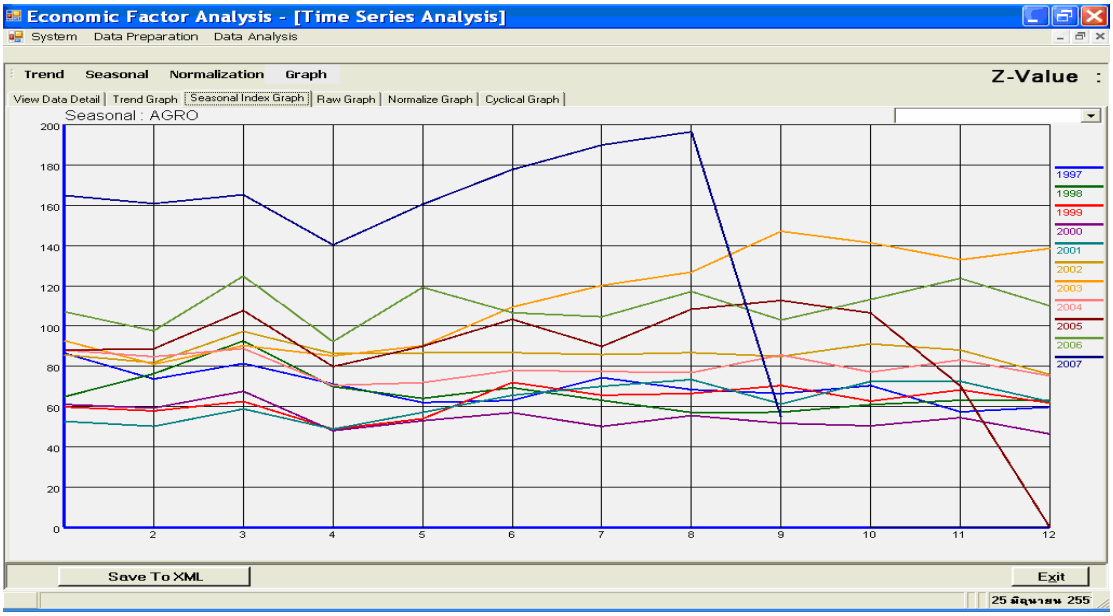


Figure 4.13: Seasonal Index Graph

2.3 **Cyclical Graph** is the graph revealing the variation of data cycle gained from the cyclical variation process, as illustrated in Figure 4.14.

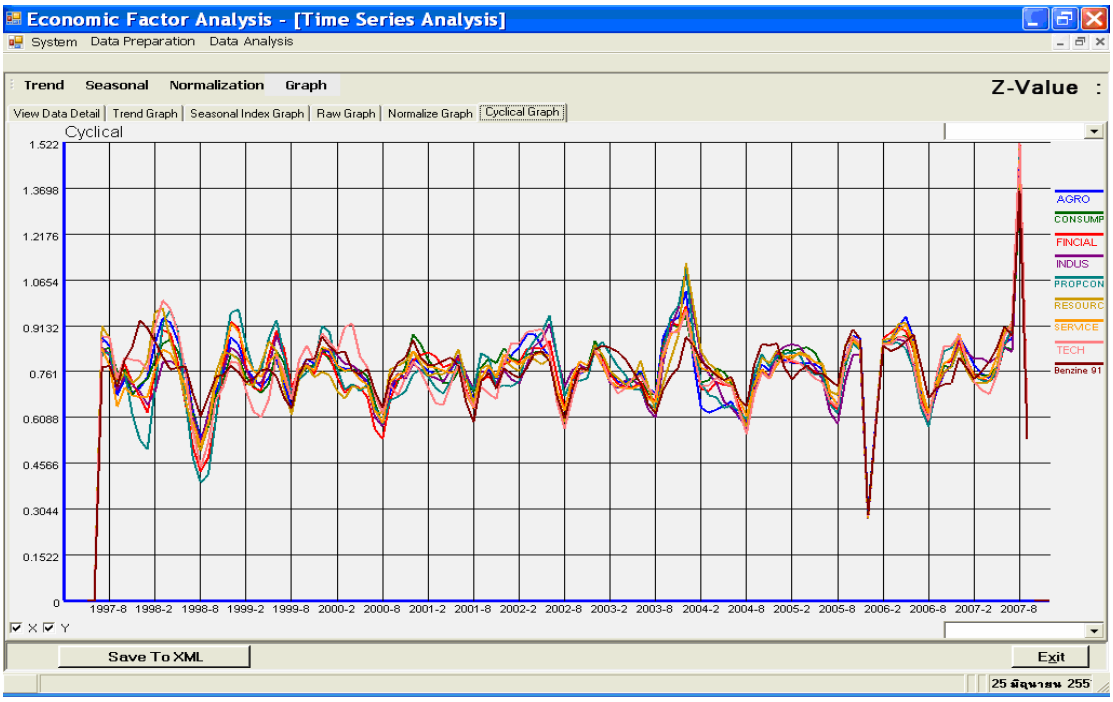


Figure 4.14: Cyclical Graph

2.4 **Irregular Graph** is the graph showing the movement of daily average value data, along with significant events that occur, as illustrated in Figure 4.15.

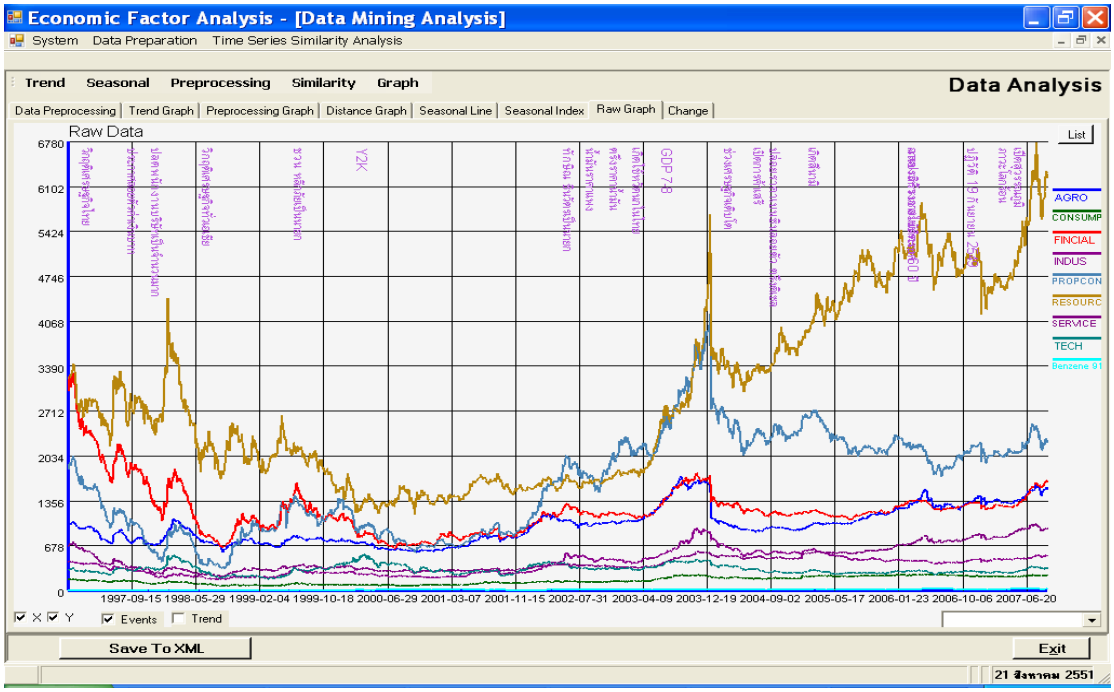


Figure 4.15: Irregular Data Graph

4.2.3.3 Similarity Analysis

It consists of data normalization, similarity method, data view and graph view. Details of each part are described below.

1. **Data Normalization** is the process of converting data in Data Warehouse into normalized data, which includes:
    1.1 **Previous Change** is the transformation of data based on previous data.
    1.2 **Moving Average** is the transformation of data of data by calculating the average value.
    1.3 **Standard Deviation Normalization** is data conversion with data standard values.
    1.4 **Min-Max Normalization** is data conversion through the difference between maximum data value and minimum data value.
2. **Similarity Distance** comprises the similarity technique used in the system, which include:

2.1 **Euclidean** is a method of calculating similarity value based on the distance between points of data.

2.2 **Dynamic Time Warping** is an approach of computing similarity value from the distance diagonally passing the distance table of all points of data in order to find minimum value.

2.3 **R-Square** is the method of calculating similarity value through the regression technique.

2.4 **Regression Time Warping** is an approach combining the R-Square technique and Dynamic Time Warping technique together.

3. **Data View** is the part displaying detail of data consequence, as shown in Figure 4.16, and includes the following:

3.1 **Raw Data** is the part showing data on daily average value obtained from warehouse. Data may be filtered based on year and month.

3.2 **Normalize Data** is the part displaying the outcome of the data normalization process. Data can be filtered based on year and month.

3.3 **Similarity Coordinate Matrix** is the display of the consequence of the similarity process.

3.4 **Similarity Result** is the display of similarity measure value. Data may be chosen to display the result and be prioritized.

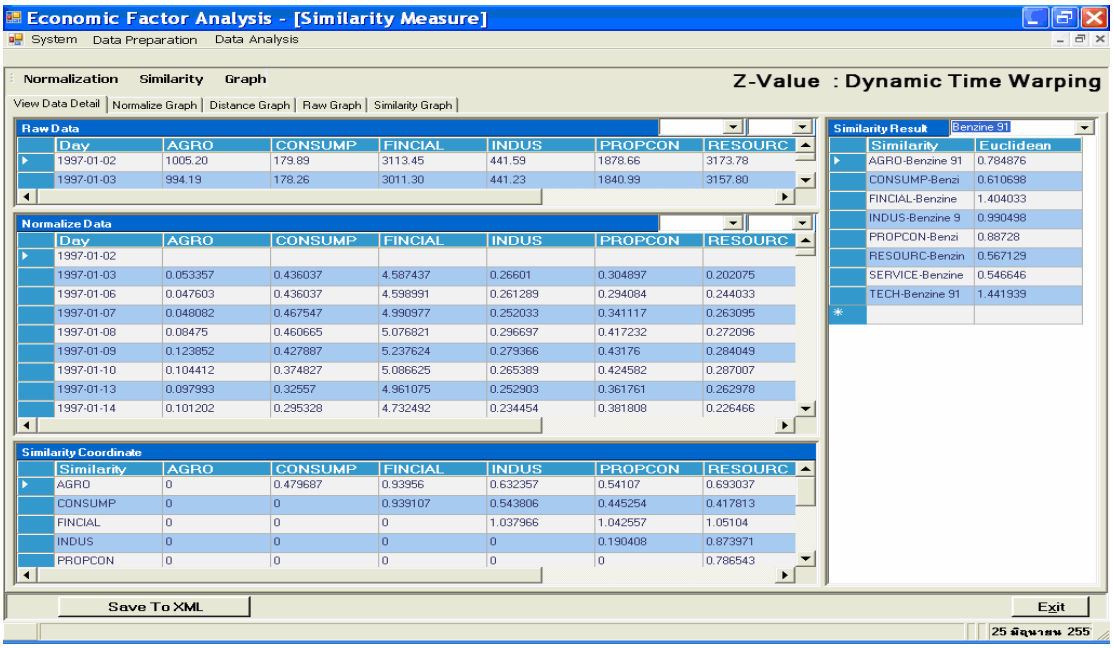


Figure 4.16: Data View Detail of Similarity Analysis

4. **Graph View** is the graph showing data result and consists of the following.

4.1 **Raw Data Graph** is the graph displaying the movement of daily average value data, as illustrated in Figure 4.17.

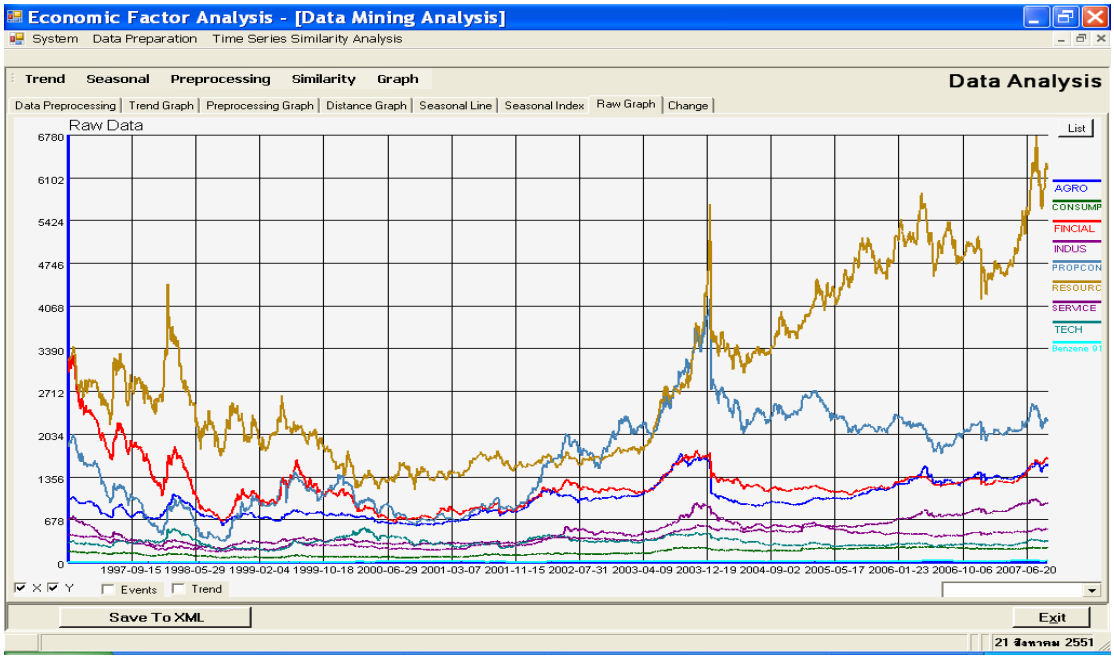


Figure 4.17: Raw Data Graph

4.2 **Data Normalization Graph** is the graph showing the movement of data of the normalization process, as illustrated in Figure 4.18.

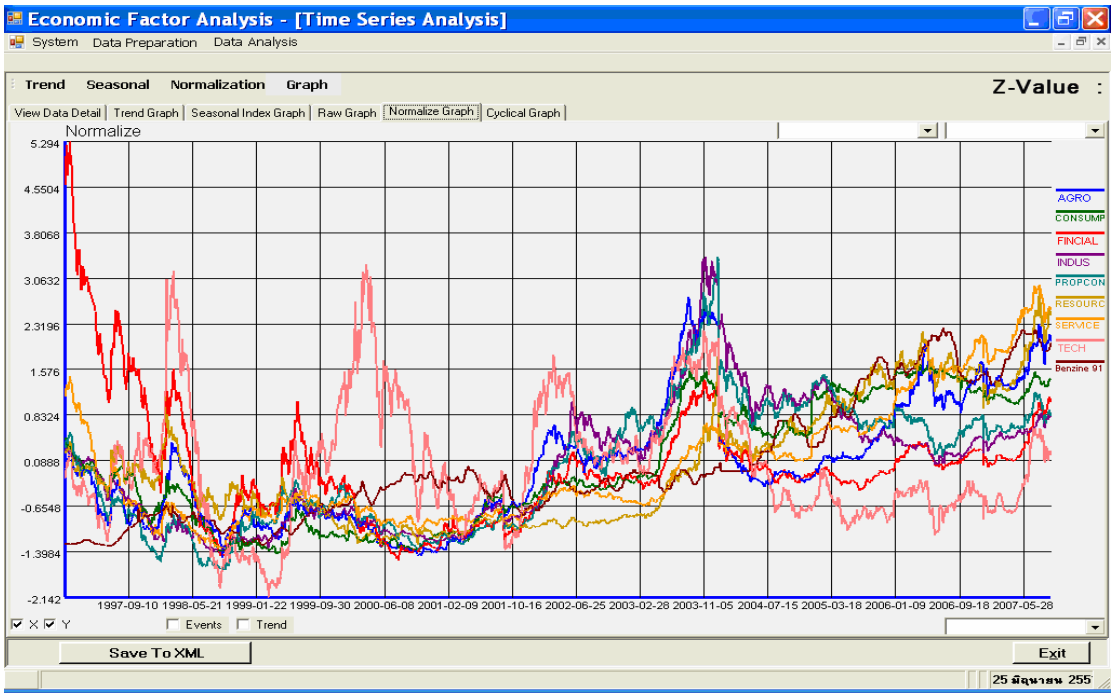


Figure 4.18: Normalize Data Graph

4.3 **Distance Graph** is the graph demonstrating the distance between data obtained from the consequence of the Euclidean technique or Dynamic Time Warping technique, as illustrated in Figure 4.19.

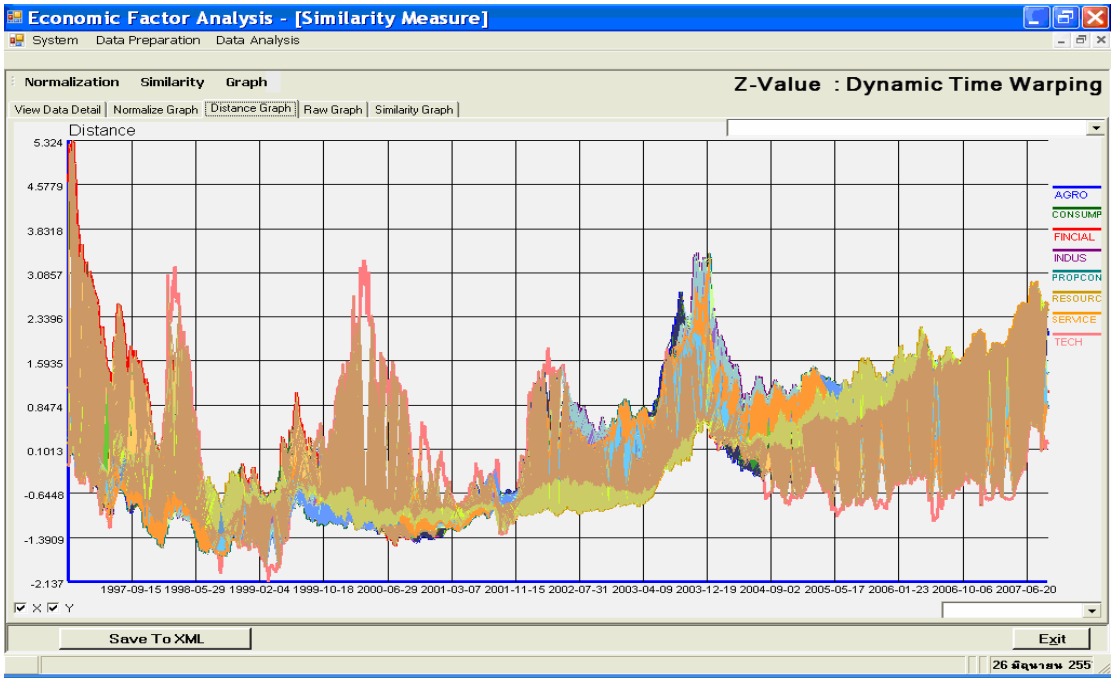


Figure 4.19: Distance Graph

4.4 **Similarity Graph** is the graph showing similarity measure value, as illustrated in Figure 4.20.

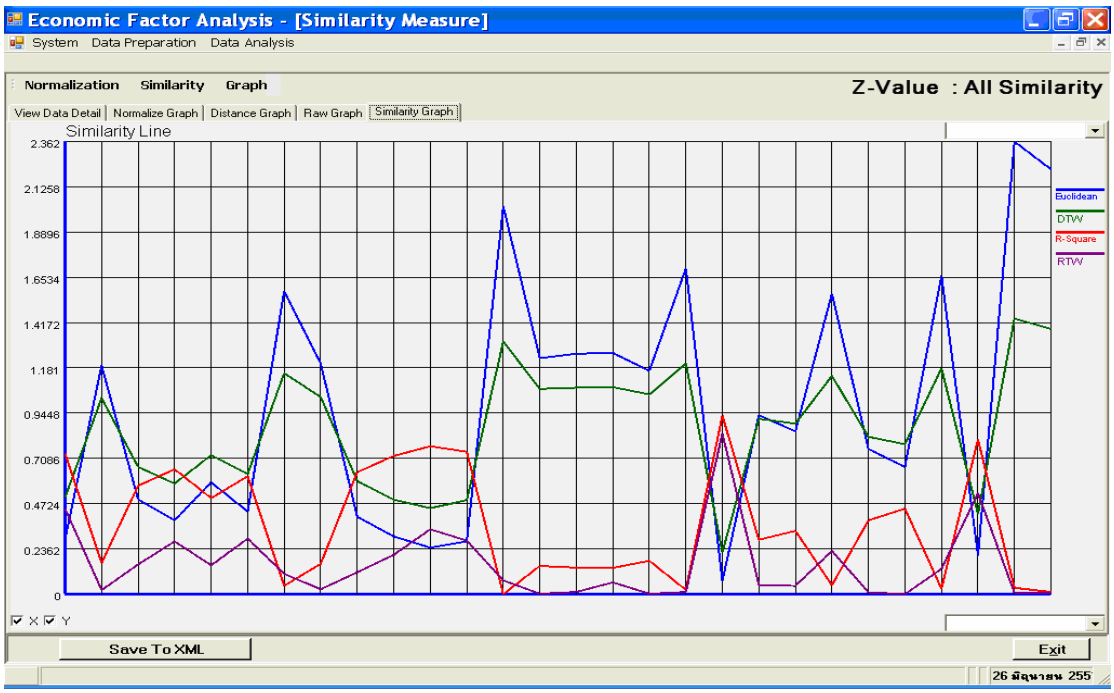


Figure 4.20: Similarity Measure Graph

CHAPTER V
EXPERIMENTAL RESULTS

This chapter discusses the consequence of the experiment and aims to investigate the relationship among various factors impacting Thai economy in terms of trend of movement of economic factors, including the similarity of time series data in each set in order to find out which set (pair) of data has an influence on others.

The experiment is divided into 3 experiments, of which details and described below.

1. Study of the relationship between oil prices data and security industry data with respect to the following :
    1.1 The Similarity between 2 sets of time series data.
    1.2 Trend of movement of data, seasonal variation of data, and significant events that have occurred to data. The study is based on time series data that are most similar and different.
    1.3 The attribute of variation (change) of data compared to previous data.
2. Study of the relationship between gold bar prices data and data on interest rates for loan and deposit in terms of :
    2.1 The similarity among the 3 sets of time series data.
    2.2 Trend of movement of data based on Time series data which are most similar and different.
3. Study of the relationship among data on all economic factors by selecting data group representation so as to find the similarity between time series data, and trend of data movement.
    3.1 For oil prices, data on Benzene 91 prices are used.
    3.2 For gold bar prices, data on domestic gold bar prices (in Thai baht) are used.

3.3 With regard to interest rates, data on 3-month fixed deposit rate and MLR loan rate are used.

3.4 Concerning Foreign Exchange rate, data on US dollar exchange rate are used.

3.5 For External Stock index, data on US Dow Jones Stock Exchange index are used.

3.6 For Stock Industry index, data on Financial Business group are used.

3.7 With respect to Stock Sector index, data on Energy and Infrastructure Sector are used.

5.1 Experiment I : An Analysis of Oil and Industry Factors

It is the study of relationship between oil prices data and major Stock Industries index data in terms of the similarity of time series data, as well as components of time series data with respect to Trend of data movement, Seasonal Variation of data, and data variation due to emerging significant incidences. Finally, the attribute of variation of data compared to previous data is studied.

5.1.1 Oil and Industry Similarity Analysis

It is to calculate the similarity of 2 sets of time series data (passing through the Standard Deviation Normalization process) using the technique of Regression Time Warping (RTW). The similarity values can indicate which pair of time series data is similar and which data have the most influence on the group. In this experiment, it is examined whether or not oil prices have an influence on Stock Industry group's data.

Table 5.1: Oil and Industry Similarity Matrix

RTW	AGRO	CONSUMP	FINCIAL	INDUS	PROPCON	RESOURC	SERVICE	TECH	Benzene 91	Benzene 95	Diesel
AGRO	0	0.478648	0.032794	0.170387	0.312387	0.170896	0.30305	0.138745	0.049482	0.037797	0.101145
CONSUMP	0	0	0.016136	0.131581	0.245874	0.362457	0.306204	0.084202	0.13582	0.13933	0.209055
FINCIAL	0	0	0	0.003271	0.014369	0.054802	0.004467	0.014764	0.000703	0.000232	0.000449
INDUS	0	0	0	0	0.854881	0.036725	0.035308	0.248858	0.220067	0.059599	0.185798
PROPCON	0	0	0	0	0	0.006136	0.000008	0.156707	0.098394	0.003945	0.081249
RESOURC	0	0	0	0	0	0	0.555754	0.002675	0.257811	0.107987	0.264985
SERVICE	0	0	0	0	0	0	0	0.006057	0.528525	0.401062	0.524916
TECH	0	0	0	0	0	0	0	0	0.007623	0.000184	0.000284
Benzene 91	0	0	0	0	0	0	0	0	0	0.71173	0.92466
Benzene 95	0	0	0	0	0	0	0	0	0	0	0.711745
Diesel	0	0	0	0	0	0	0	0	0	0	0

Table 5.2: Oil and Industry Similarity

Benzene91		Benzene95		Diesel	
Similarity	**RTW**	**Similarity**	**RTW**	**Similarity**	**RTW**
SERVICE-Benzene 91	0.528525	SERVICE-Benzene 95	0.401062	SERVICE-Diesel	0.524916
RESOURC-Benzene 91	0.257811	CONSUMP-Benzene 95	0.13933	RESOURC-Diesel	0.264985
INDUS-Benzene 91	0.220067	RESOURC-Benzene 95	0.107987	CONSUMP-Diesel	0.209055
CONSUMP-Benzene 91	0.13582	INDUS-Benzene 95	0.059599	INDUS-Diesel	0.185798
PROPCON-Benzene 91	0.098394	AGRO-Benzene 95	0.037797	AGRO-Diesel	0.101145
AGRO-Benzene 91	0.049482	PROPCON-Benzene 95	0.003945	PROPCON-Diesel	0.081249
TECH-Benzene 91	0.007623	FINCIAL-Benzene 95	0.000232	FINCIAL-Diesel	0.000449
FINCIAL-Benzene 91	0.000703	TECH-Benzene 95	0.000184	TECH-Diesel	0.000284

Based on Tables 5.1 and 5.2, it is found that the most similar between oil prices data and Stock Industry group data are those pairs of SERVICE Industry index and Benzene 91 (0.5285), SERVICE Industry index and Diesel (0.5249) and SERVICE Industry and Benzene 95 (0.4010). This means that oil prices data have the most influence on SERVICE Industry index.

For the pairs of data that are most different, they are pairs of TECH Industry index and Benzene 95 (0.000184), TECH Industry and Diesel (0.000284) and FINCIAL Industry index and Benzene91 (0.000703). This means that oil prices data have the least influence on TECH Industry index, followed by on FINCIAL Industry index.

In sum, Benzene 91 oil price has the most influence on SERVICE Industry index in Industry Stock Index group.

5.1.2 Oil and Industry Time Series Analysis

It is to compute the attribute of components of time series data with regard to Trend of data movement, attribute of Seasonal Variation of data occurring repetitiously over 1 year, and attribute of Irregular Variation.

To calculate the attribute of components of Time series data, the most similar data obtained from Table 5.2, which are SERVICE Industry index and Benzene 91, are used. For the most different data, they are also derived from Table 5.2, which are FINCIAL Industry index and Benzene 95 as well as TECH Industry index and Benzene 95.

1. Oil and Industry Trend

It is to view the trend of movement of data pair which is similar and different.

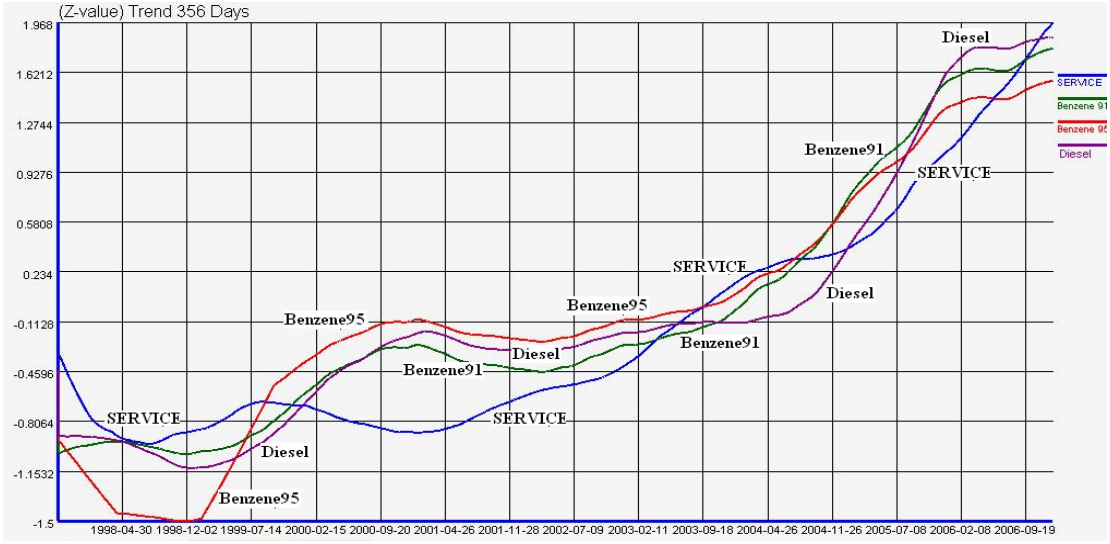


Figure 5.1: SERVICE and Oil Price Trend (1 Year) Graph (Similarity)

Figure 5.1 illustrates the trend of similar movement of data, calculated from data of 1 year. It can be seen that the trend of SERVICE Industry index and Benzene 91 price data goes in the same direction and in increasing.

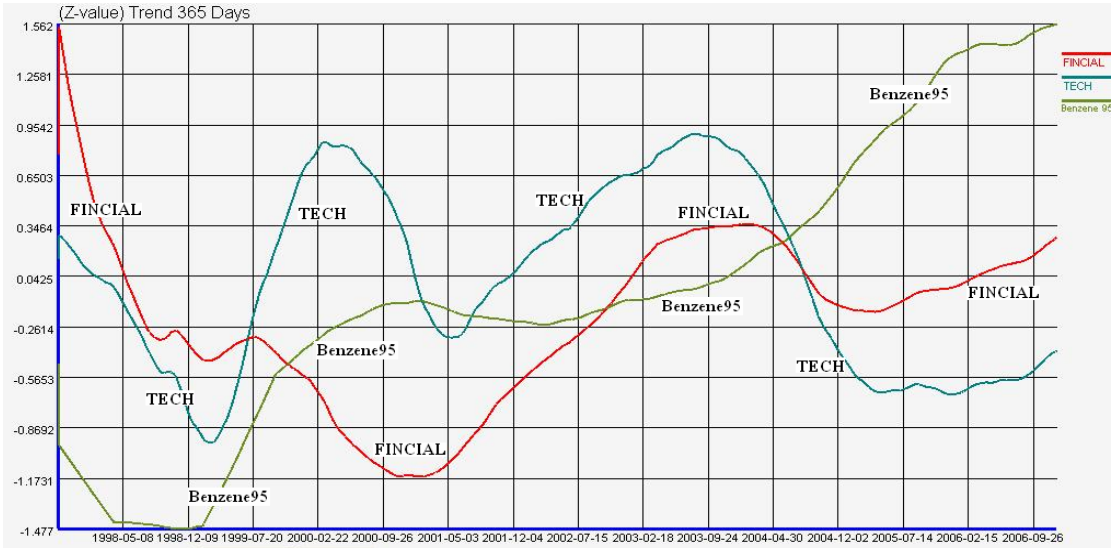


Figure 5.2: FINCIAL and Benzene 95 Trend (1 Year) Graph (Dissimilarity)

Figure 5.2 illustrates the trend of dissimilar movement of data, calculated from data of 1 year. It can be seen that the trend of FINCIAL Industry index and Benzene 95 price goes in the different direction. Benzene 95 price have an increasing trend while data on FINCIAL Industry index and TECH Industry index have an approximate movement and a stable trend.

2. Oil and Industry Seasonal Variation

It is to find data characteristic over 1 year to see how data vary in each month, based on data on Benzene 91 price, Benzene 95 price, SERVICE Industry index , FINCIAL Industry index and TECH Industry index .

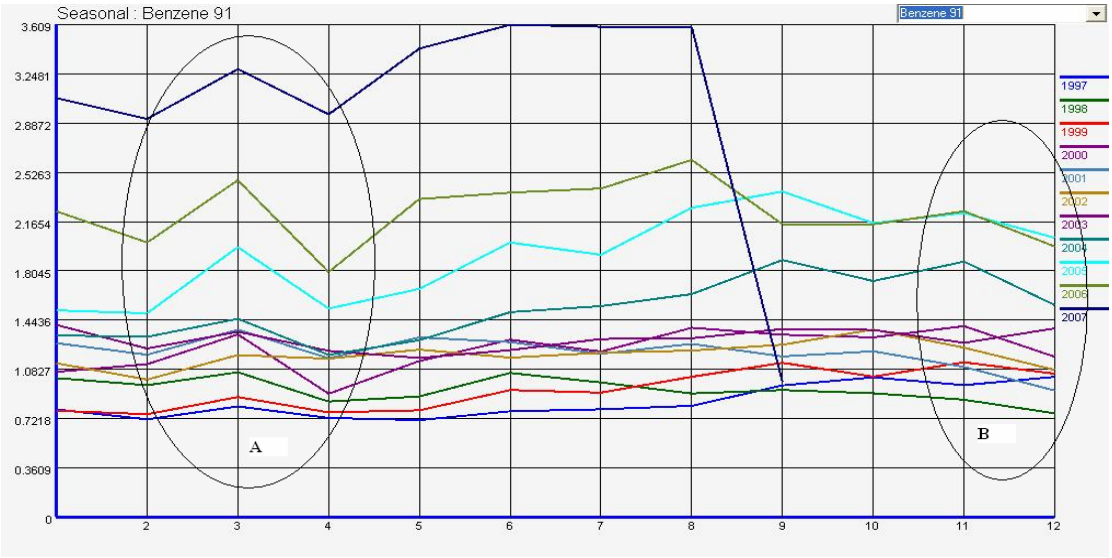


Figure 5.3: Benzene91 Seasonal Graph

Based on Figure 5.3, it is found that Benzene 91 price data at point A have movement characteristic of W shape. This means data decline in the first month and rise at the end of the second month before they fell again in the third month and rise at the end of the fourth month, except in 1997, 1999 and 2002. At point B, which is year-end, data diminish except in 1997 and 2003.

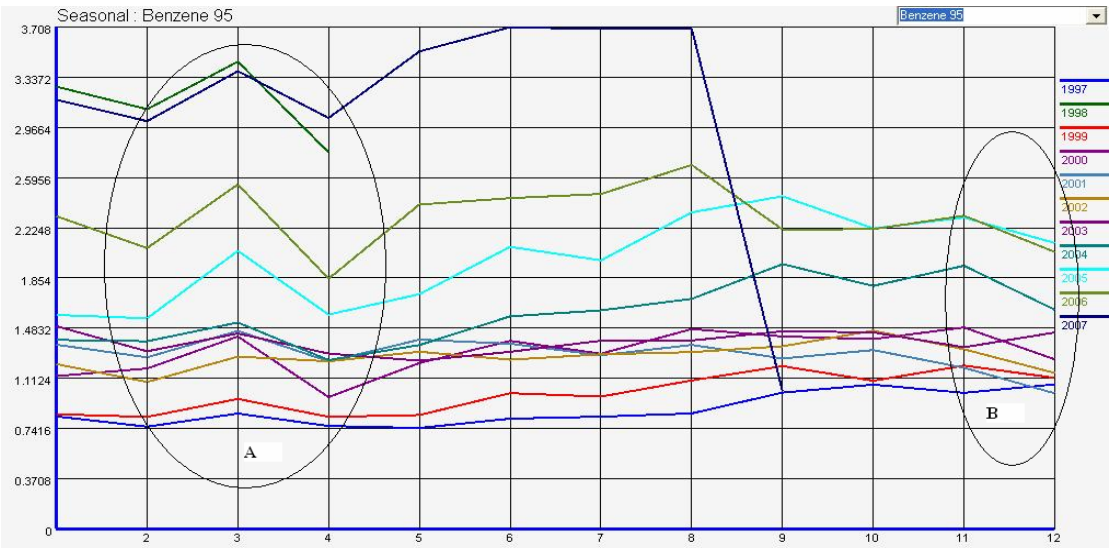


Figure 5.4: Benzene95 Seasonal Graph

Based on Figure 5.4, it is found that Benzene 95 price data have a movement similar to that of Benzene 91. This means data decline from the first month and rise at the end of the second month before they fell in the third month and rise again at the end of the fourth month except in 1997, 1999 and 2002. At point B, this is year-end, data decline except in 1997 and 2003.

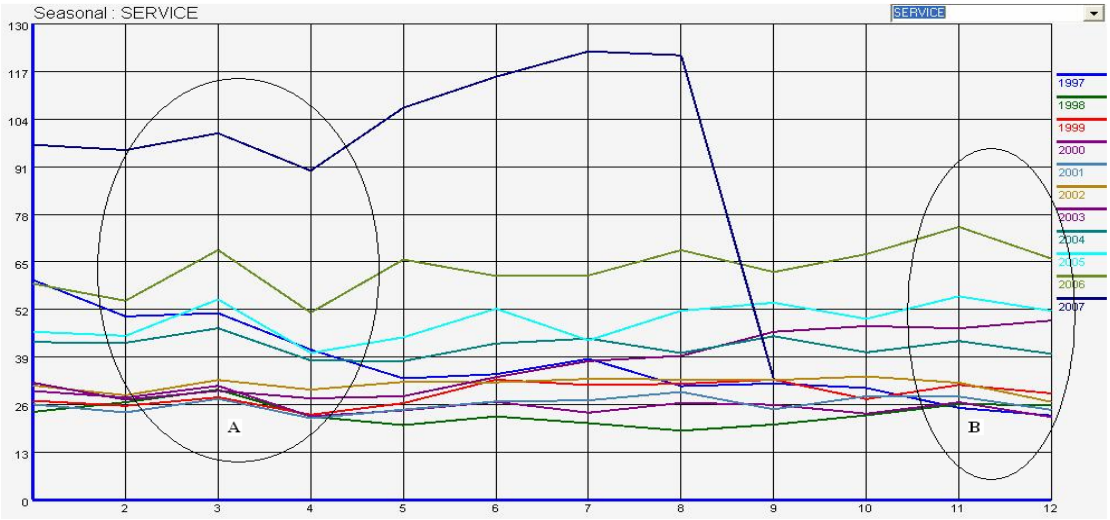


Figure 5.5: SERVICE Seasonal Graph

Based on Figure 5.5, it is found that SERVICE Industry index at point A have a movement characteristic of W shape, which means data decline from the first month and rise at the end of the second month before they fell again in the third month and increase at the end of the fourth month except in 2005, 2006 and 2006. At point B, this is year-end, data decline except in 2003.

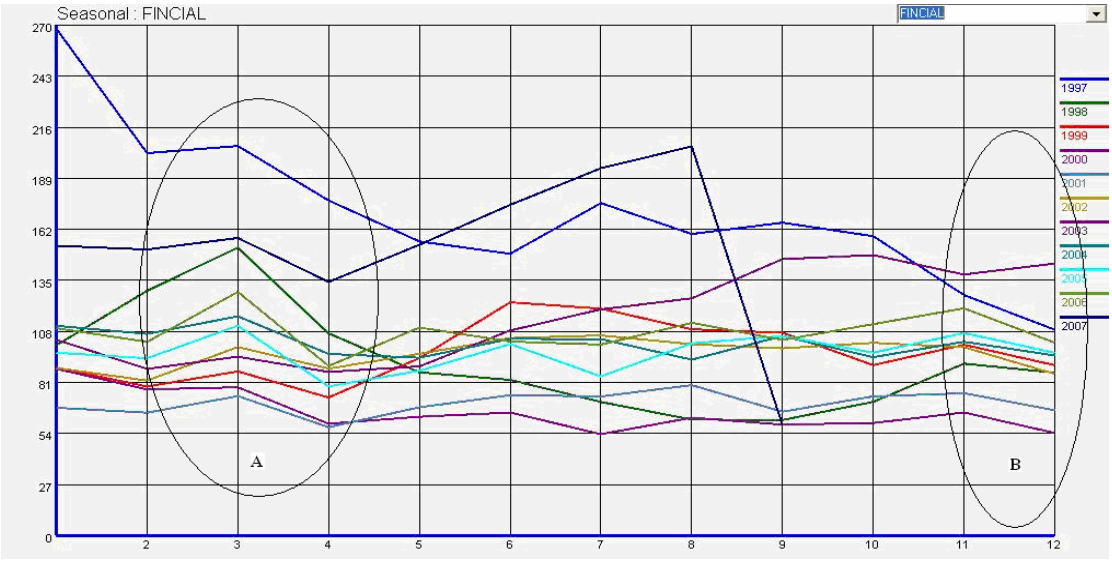


Figure 5.6: FINCIAL Seasonal Graph

Based on Figure 5.6, it is found that FINCIAL Industry index at point A have a movement characteristic of W shape, which means that decline from the first month and rise at the end of the second month before they fall down again in the third month and rise at the end of the fourth month in 2006 and 2007. At point B, this is year-end, data decline except in 2003.

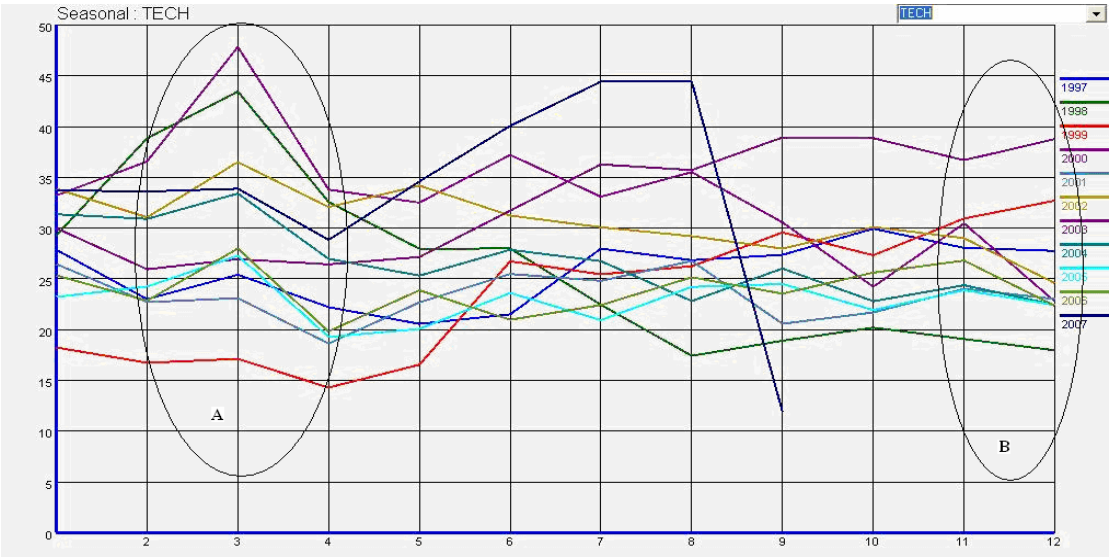


Figure 5.7: TECH Seasonal Graph

Based on Figure 5.7, it is found that TECH Industry index at point A have a movement characteristic of W shape, which means data decline from the first month and rise at the end of the second month before they fell down again in the third month and then rise at the end of the fourth month in 1998, 1999, 2000, 2005 and 2007. At point B, this is year-end, data decline except in 1997, 1999 and 2003.

In conclusion, the attribute of Seasonal Variation data can be divided into 2 patterns as follows:

1. Data in the first through fourth months had a declining characteristic from the first month and rose at the end of the second month before they fell down again in the third month and rose at the end of the fourth month. Oil prices data possessed this characteristic every year except in 1997, 1999 and 2002. Similarly, Stock Industry groups also had this trend every year except in 2003.
2. At year-end, data typically declined, with oil prices data having this trend every year except in 1997 and 2003. Likewise, Stock Industry group also had this trend in 2002, 2004, 2005 and 2006.

3. Oil and Industry Irregular Variation

It is to find the attribute of variation due to emerging significant occurrences of data, through viewing data on Benzene 91 price, Benzene 95 price as well as SERVICE, FINCIAL and TECH Industries index.

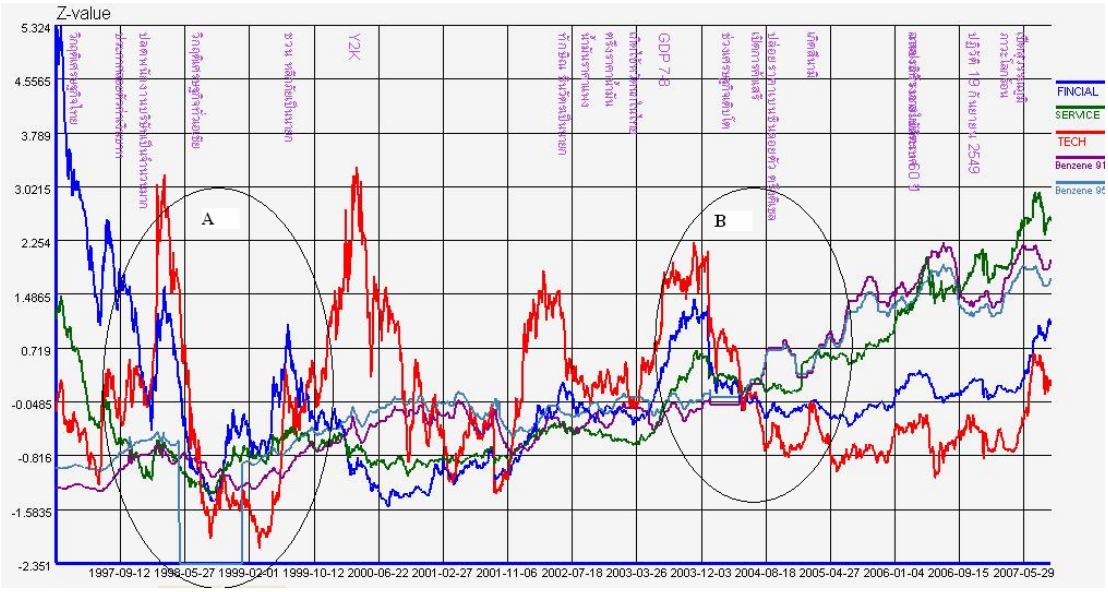


Figure 5.8: Oil and Industry Event Graph

Figure 5.8 illustrates significant incidences taking place between 1997 and 2007, during which 2 events outstanding attributes were found. At point A, Stock Industry index declined vertically (declaration of a float of baht in 1997). At point B, Stock Industry index surged vertically (the country's economy improved, with GDP between 7-8%).

5.1.3 Oil and Industry Data Change

It is to find an increase rate of data through a calculation of the previous data change value subtracted by the current data value.

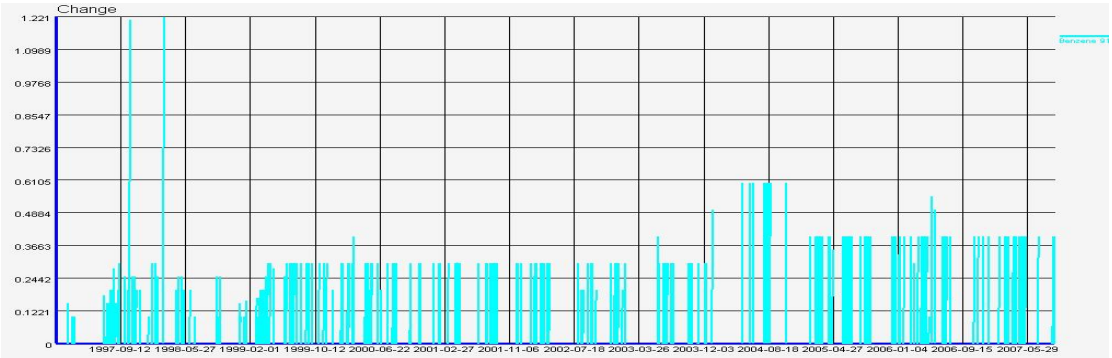


Figure 5.9: Benzene91 Change Graph

Based on Figure 5.9, it is found that data on Benzene 91 price had a stable increase rate between 1999 and 2003 and rose in 2004. Prices also had a stable increase rate between 2005 and 2007.

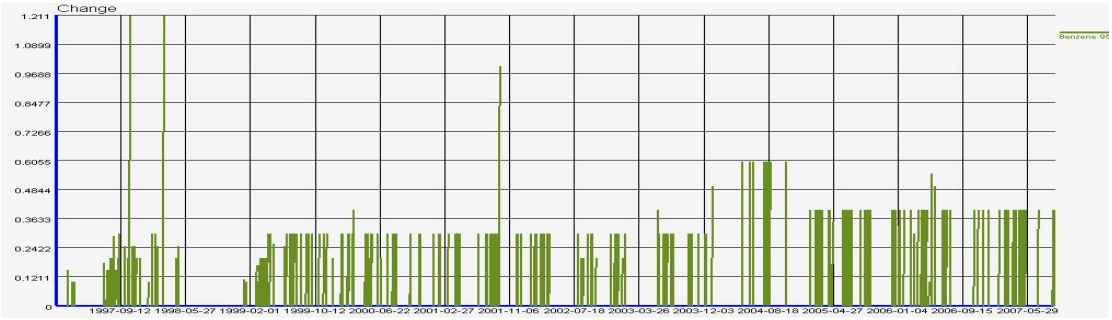


Figure 5.10: Benzene95 Change Graph

Based on Figure 5.10, it is found that data on Benzene 95 price had an increasing rate similar to that of Benzene 91 price, as shown in Figure 5.9.

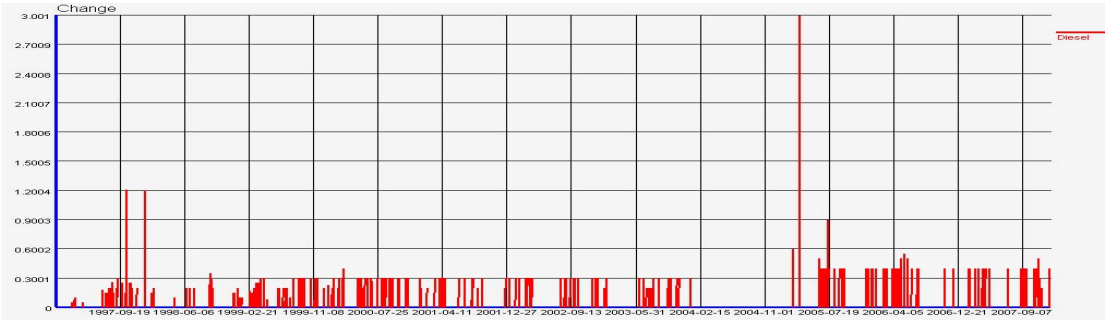


Figure 5.11: Diesel Change Graph

Based on Figure 5.11, it is found that data on Diesel price had a stable increase rate between 1999 and 2003 and rose in 2004 to Apr 2005. Prices also had a stable increase rate between 2005 and 2007.

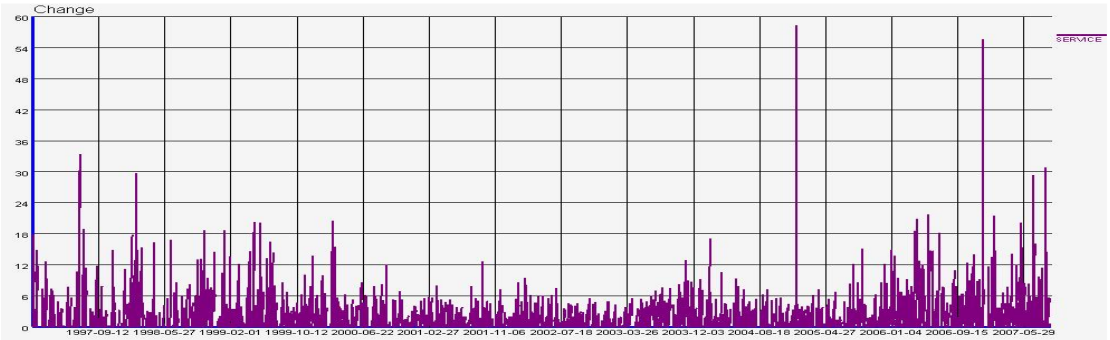


Figure 5.12: SERVICE Change Graph

Based on Figure 5.12, it is found that data on SERVICE Industry index had a variably increasing rate. Nevertheless, between 2000 and 2003, an increasing rate was the least.

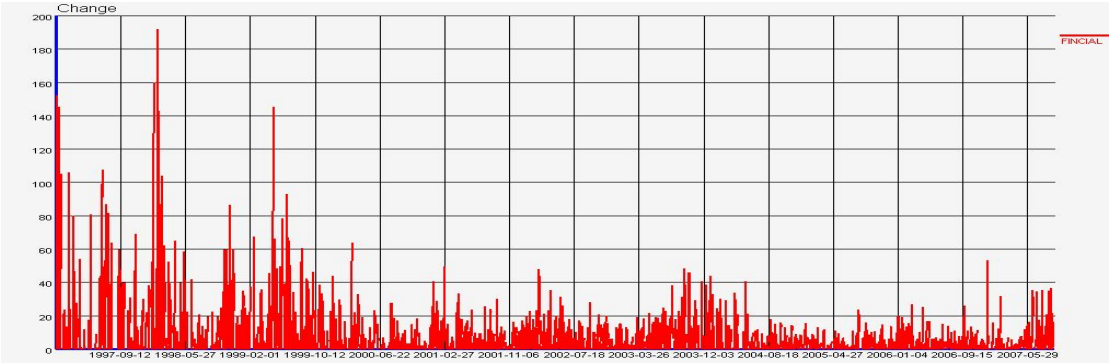


Figure 5.13: FINCIAL Change Graph

Based on Figure 5.13, it is found that data on FINCIAL Industry index had a constant change of increasing rate. Between 1997 and 2004, data underwent a steady rise and subsequently had a stable increasing rate.

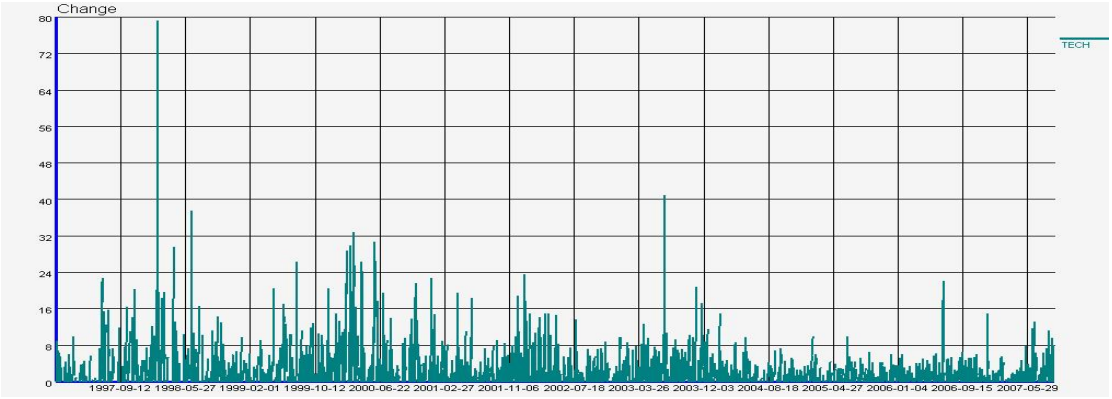


Figure 5.14: TECH Change Graph

Based on Figure 5.14, it is found that data on TECH Industry index had a constant change of rising rate. Nevertheless, between 2004 and 2007, data had a lower increasing rate than that between 1997 and 2004.

In sum, for the trend of data increase, it is found that oil prices data possessed a stable and long-term increasing trend while Stock Industry index had a constant change of rising trend.

5.2 Experiment II : An Analysis of Gold and Interest Factors

It is the study of relationship between gold bar price data and data on interest rates for loan and deposit with respect to the similarity of time series data trend of time series data movement.

5.2.1 Gold and Interest Similarity Analysis

It is to calculate the similarity of 2 sets of time series data (passing through the Standard Deviation Normalization process) using the technique of Regression Time Warping (RTW). The similarity values can indicate which data have the most influence on the group. In this experiment, it will be examined whether or not gold bar prices have any influence on bank interest rate.

Table 5.3: Gold and Interest Similarity Matrix

RTW	Baht Price	US Price	High Fix 3 M	High Fix 6 M	High Saving	MLR	MOR	MRR
Baht Price	0	0.851946	0.014743	0.012453	0.004414	0.093844	0.120075	0.090191
US Price	0	0	0.100762	0.094418	0.001436	0.21637	0.246121	0.205854
High Fix 3 M	0	0	0	0.988492	0.120316	0.754315	0.689471	0.75657
High Fix 6 M	0	0	0	0	0.093547	0.71403	0.647742	0.717449
High Saving	0	0	0	0	0	0.553254	0.587978	0.554808
MLR	0	0	0	0	0	0	0.99103	0.994404
MOR	0	0	0	0	0	0	0	0.989109
MRR	0	0	0	0	0	0	0	0

Table 5.4: Gold and Interest Similarity

Gold (Baht Price)		Gold (US Price)	
Similarity	**RTW**	**Similarity**	**RTW**
Baht Price-MOR	0.110021	US Price-MOR	0.250509
Baht Price-MLR	0.08492	US Price-MLR	0.20946
Baht Price-MRR	0.082184	US Price-MRR	0.199129
Baht Price-High Fix 3 Month Rate	0.007179	US Price-High Fix 3 Month Rate	0.038626
Baht Price-High Fix 6 Month Rate	0.005472	US Price-High Fix 6 Month Rate	0.030782
Baht Price-High Saving Level	0.003508	US Price-High Saving Level	0.011081

Based on Tables 5.3 and 5.4, it is found that a pair with the most similarity is one of foreign (US dollar) gold bar price and MOR loan rate (0.250509).

In contrast, a pair with the most difference is one of domestic gold bar price (in Baht) and savings deposit rate (0.0003508).

In sum, foreign gold bar price (in US dollar) and domestic gold bar price (in Thai baht) are the most similar to interest rate for loan. Nonetheless, since the value obtained is less than 0.3, data on gold bar prices and interest rates are very slightly similar.

The pair with the most difference is one of domestic gold bar prices (in Thai baht) and interest rates for deposit.

5.2.2 Gold and Interest Time Series Analysis (Trend)

It is to find the attribute of components of time series data with respect to data movement trend.

To calculate the trend of data movement, the most similar data obtained from Table 5.4 are used, which are foreign gold bar prices (in US dollar) and MOR loan rate. Similarly, the most different data are also derived from Table 5.4, which are domestic gold bar prices (in Thai baht) and saving deposit rate.

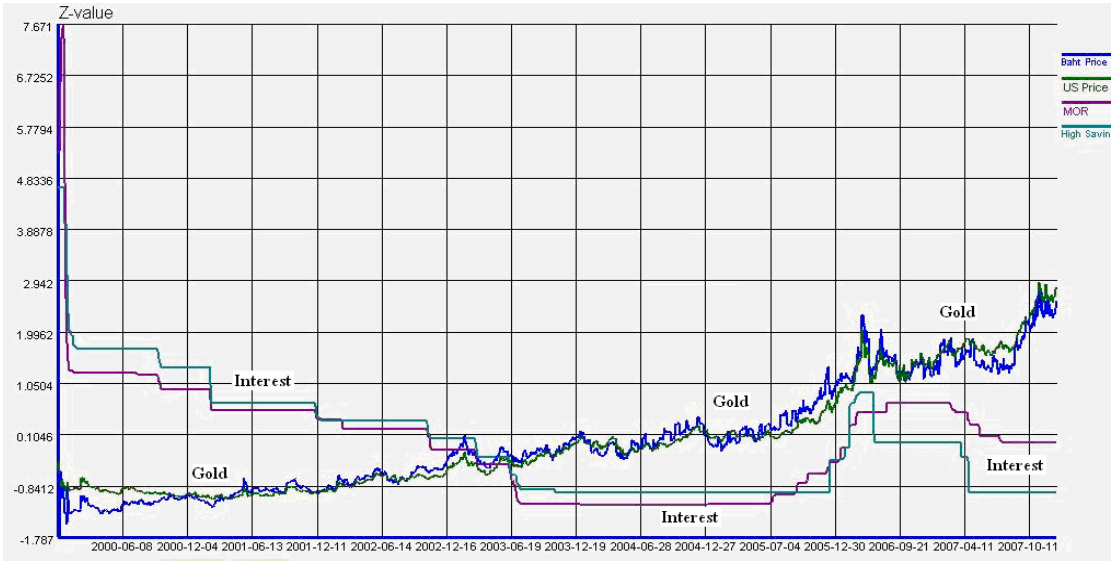


Figure 5.15: Gold and Interest Movement Graph

Figure 5.15 illustrates the characteristic of movement of gold bar prices data as well as data on MOR loan rate and saving deposit rate. Gold bar prices data have a movement of constant change while interest rate data have almost continuously stable movement.

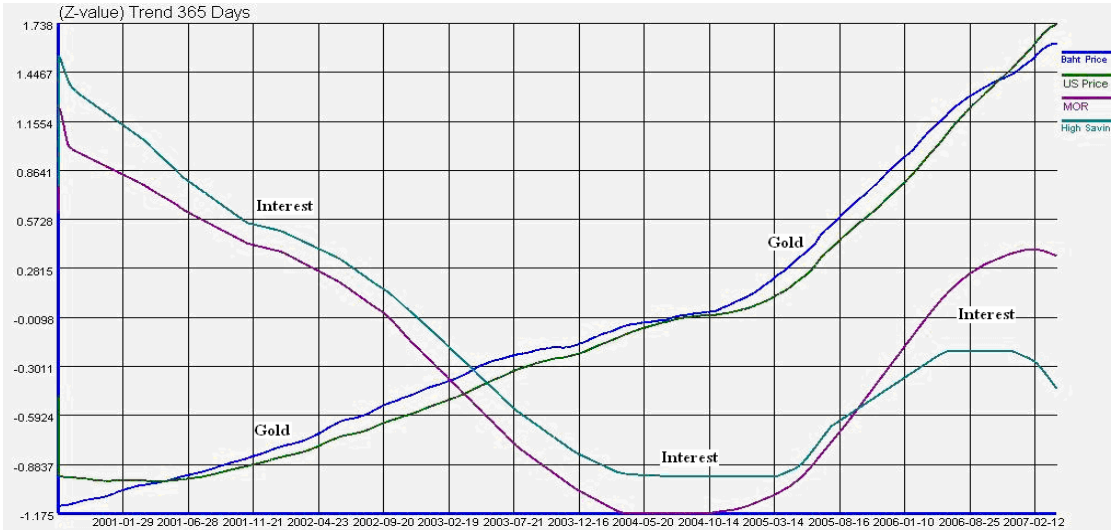


Figure 5.16: Gold and Interest Trend Graph

Figure 5.16 illustrates trend of data movement. It can be seen that gold prices have constantly increasing trend. Meanwhile, interest rate has passed through the lowest point and then rises. However, between 2006 and 2007, the interest rates encountered a declining trend.

5.3 Experiment III : An Analysis of All Economic Factors

It is to find the relationship of all economic factors by sampling for the representative of factors data in order to calculate data similarity and trend of movement of economic factors data.

The following are used as group representative. Benzene 91 price is used for oil price, domestic gold bar price for gold bar price, 3-month fixed deposit for deposit rate and MLR loan for loan rate, US dollar exchange rate for Foreign Exchange rate, Dow Jones Stock index for External Stock index, FINCIAL Industry index for Stock Industry index, and ENERG Sector index for Stock Sector index.

5.3.1 Economic Factors Similarity Analysis

It is to compute the similarity values of 2 sets of time series data. The values obtained can specify which pair of time series data is similar, and which data have the most influence on the group.

Table 5.5: Economic Factors Similarity Matrix

RTW	Dow Jones	Baht Price	FINCIAL	High Fix 3 Month	MLR	US	Benzene 91	ENERG
Dow Jones	0	0.072398	0.001839	0.00113	0.032749	0.127713	0.086069	0.052564
Baht Price	0	0	0.117122	0.014381	0.026929	0.375501	0.699158	0.628377
FINCIAL	0	0	0	0.000021	0.064777	0.058376	0.01752	0.005846
High Fix 3 Month	0	0	0	0	0.737241	0.011056	0.00602	0.014327
MLR	0	0	0	0	0	0.049779	0.031657	0.102214
US	0	0	0	0	0	0	0.25773	0.25997
Benzene 91	0	0	0	0	0	0	0	0.682398
ENERG	0	0	0	0	0	0	0	0

Table 5.6: Economic Factors Similarity / Dissimilarity

Similarity		Dissimilarity	
Similarity	**RTW**	**Similarity**	**RTW**
High Fix 3 Month Rate-MLR	0.737241	FINCIAL-High Fix 3 Month Rate	0.000021
Baht Price-Benzene 91	0.699158	Dow Jones-High Fix 3 Month Rate	0.00113
Benzene 91-ENERG	0.682398	Dow Jones-FINCIAL	0.001839
Baht Price-ENERG	0.628377	FINCIAL-ENERG	0.005846
Baht Price-US	0.375501	High Fix 3 Month Rate-Benzene 91	0.00602
US-ENERG	0.25997	High Fix 3 Month Rate-US	0.011056
US-Benzene 91	0.25773	High Fix 3 Month Rate-ENERG	0.014327
Dow Jones-US	0.127713	Baht Price-High Fix 3 Month Rate	0.014381
Baht Price-FINCIAL	0.117122	FINCIAL-Benzene 91	0.01752
MLR-ENERG	0.102214	Baht Price-MLR	0.026929
Dow Jones-Benzene 91	0.086069	MLR-Benzene 91	0.031657
Dow Jones-Baht Price	0.072398	Dow Jones-MLR	0.032749
FINCIAL-MLR	0.064777	MLR-US	0.049779
FINCIAL-US	0.058376	Dow Jones-ENERG	0.052564

Based on Tables 5.5 and 5.6, it is found that data pairs with the most similarity are those of 3-month fixed deposit rate and MLR loan rate (0.737241), Benzene 91 price and domestic gold bar price (0.699158), Benzene 91 price and ENERG Sector index (0.682398), domestic gold bar price and ENERG Sector index (0.628377), and domestic gold bar price and US dollar exchange value (0.375501).

On the contrary, data pairs with the most difference are those of FINCIAL Industry index and 3-month fixed deposit rate (0.000021), Dow Jones Stock index and 3-month fixed deposit rate (0.00113), Dow Jones Stock index and FINCIAL Industry index (0.001839), FINCIAL Industry index and ENERG Sector index (0.005846), and 3-month fixed deposit rate and Benzene 91 price (0.00602).

It can be summarized that similar data group, ranking from most to least, include interest rates for deposit and loan, followed by oil price, which is partly similar to gold bar price and ENERG Sector index, as well as by gold bar price, which is partly similar to ENERG Sector index and Foreign Exchange rate as shown in Table 5.7.

Table 5.7: Summary of Economic Factors Similarity

Similarity		
No	**Independent Variable**	**Dependent Variable**
1	Deposit interest rate	Loan interest rate
2	Oil Price	Gold bar price
		ENERG Sector Index
3	Gold bar price	ENERG Sector Index
		Foreign Exchange rate

For different data groups, they are interest rate for deposit and FINCIAL Industry index, External Stock index and oil price, and FINCIAL Industry index as well as External Stock index and ENERG Sector Index, as shown in Table 5.8.

Table 5.8: Summary of Economic Factors Dissimilarity

Dissimilarity		
No	**Independent Variable**	**Dependent Variable**
1	Deposit interest rate	FINCIAL Industry Index
		External Stock index
		Oil price
2	FINCIAL Industry Index data	External Stock index
		ENERG Sector Index

5.3.2 Economic Factors Time Series Analysis (Trend)

It is to find the attribute and trend of similar data movement from the representatives of economic factor data selected in section 5.3. The most similar data obtained from Table 5.7. Group of similarity data includes the following.

1. Loan and Deposit Similarity

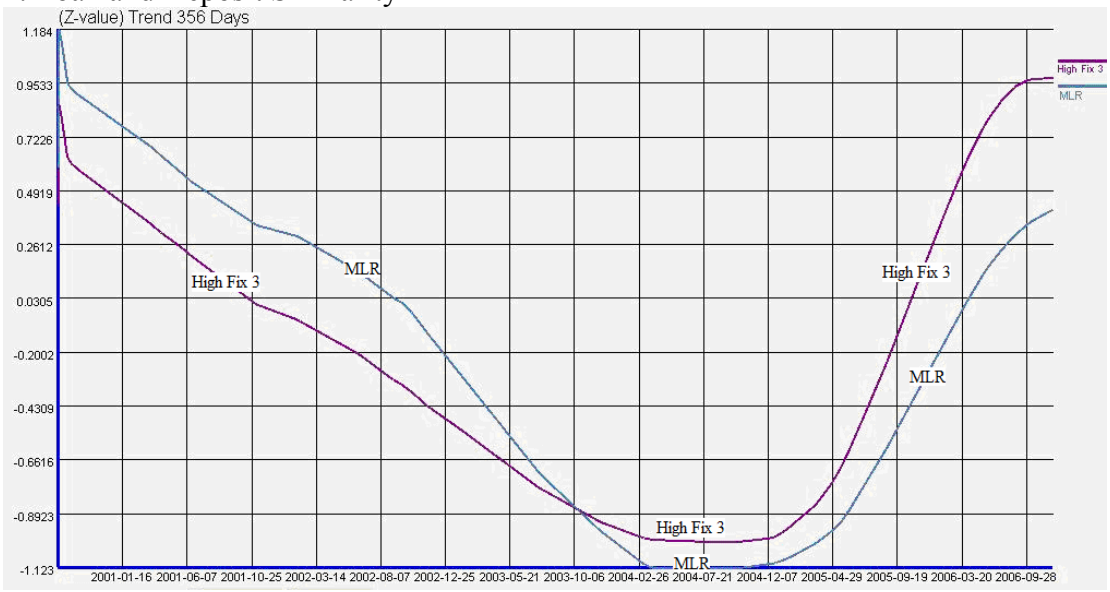


Figure 5.17: Loan and Deposit Similarity Graph

Figure 5.17 illustrates the trend of similar movement of data, calculated from data of 1 year. It can be seen that the trend of Loan interest rate and Deposit interest rate data goes in the same direction and in increasing.

2. Oil Price similarity between Gold Bar and ENERG Sector

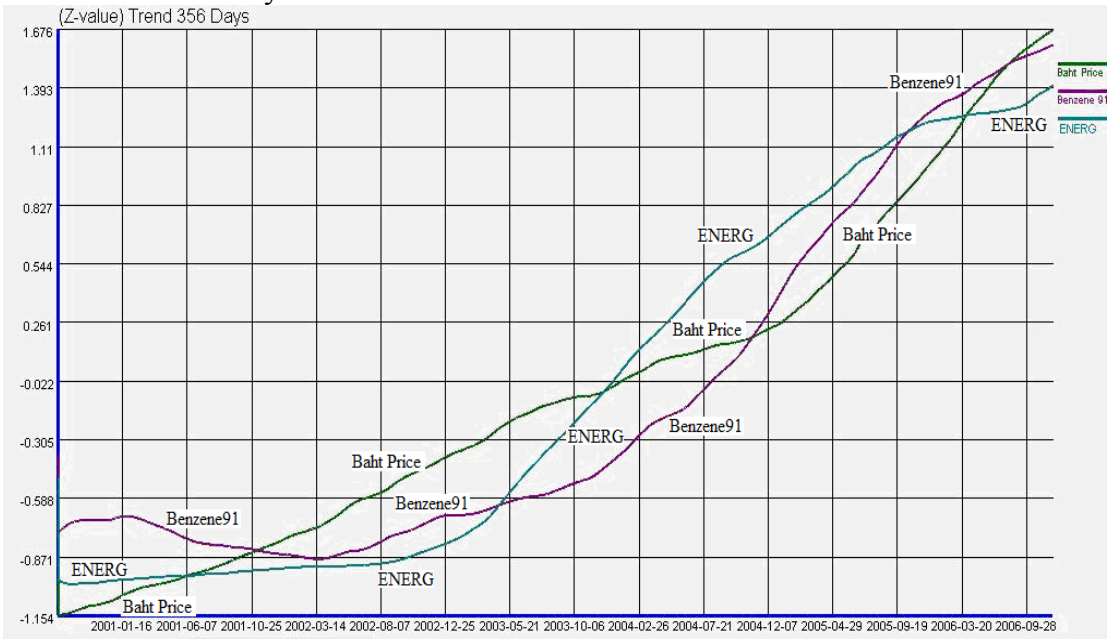


Figure 5.18: Oil Price, Gold Bar and ENERG Sector Similarity Graph

Figure 5.18 illustrates the trend of similar movement of data, calculated from data of 1 year. It can be seen that the trend of oil price, gold bar price data and ENERG sector index goes in the same direction and in increasing.

3. Gold Bar similarity between ENERG Sector and Foreign Exchange

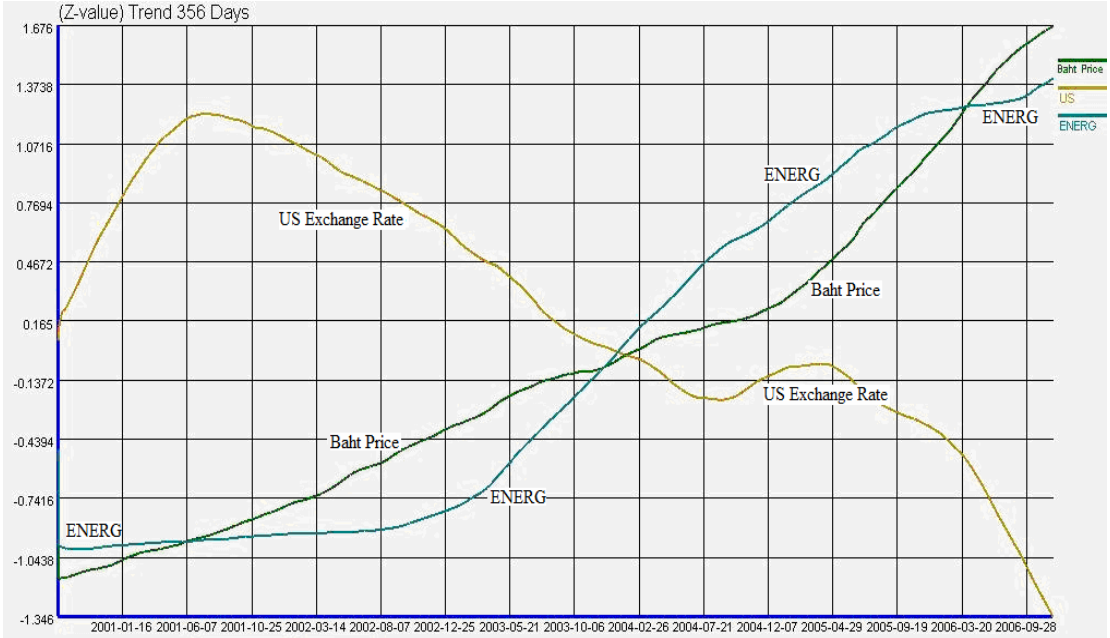


Figure 5.19: Gold Bar, ENERG Sector and Foreign Exchange Similarity Graph

Figure 5.19 illustrates the trend of similar movement of data, calculated from data of 1 year. It can be seen that the trend of gold bar price data and ENERG sector index goes in the same direction and in increasing. Except foreign exchange rate movement in cross direction from gold bar price data and ENERG sector index.

It is to find the attribute and trend of dissimilar data movement from the representatives of economic factor data selected in section 5.3. The most dissimilar data obtained from Table 5.8. Group of dissimilarity data includes the following.

1. Deposit dissimilarity between FINCIAL Industry, External Stock and Oil Price

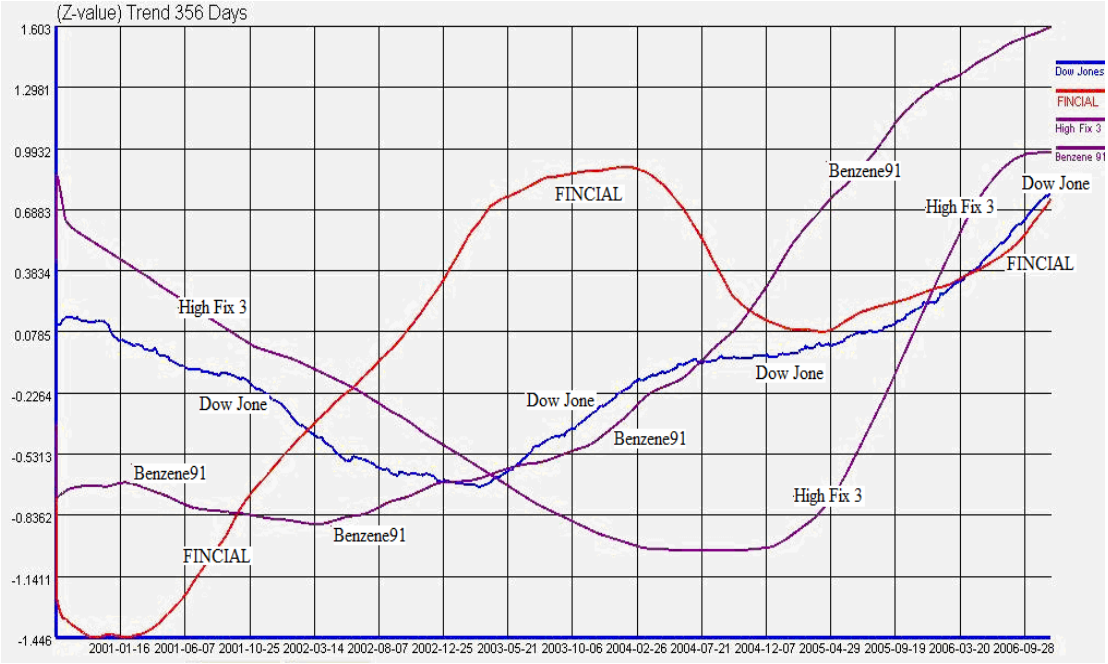


Figure 5.20: Deposit, FINCIAL, External Stock and Oil Price Dissimilarity Graph

Figure 5.20 illustrates the trend of dissimilar movement of data, calculated from data of 1 year. It can be seen that the trend of deposit interest rate, FINCIAL industry index, External Stock index and oil price goes in the different direction. But oil price and External Stock Index have goes in the same direction.

2. Financial Industry dissimilarity between External Stock and ENENG Sector

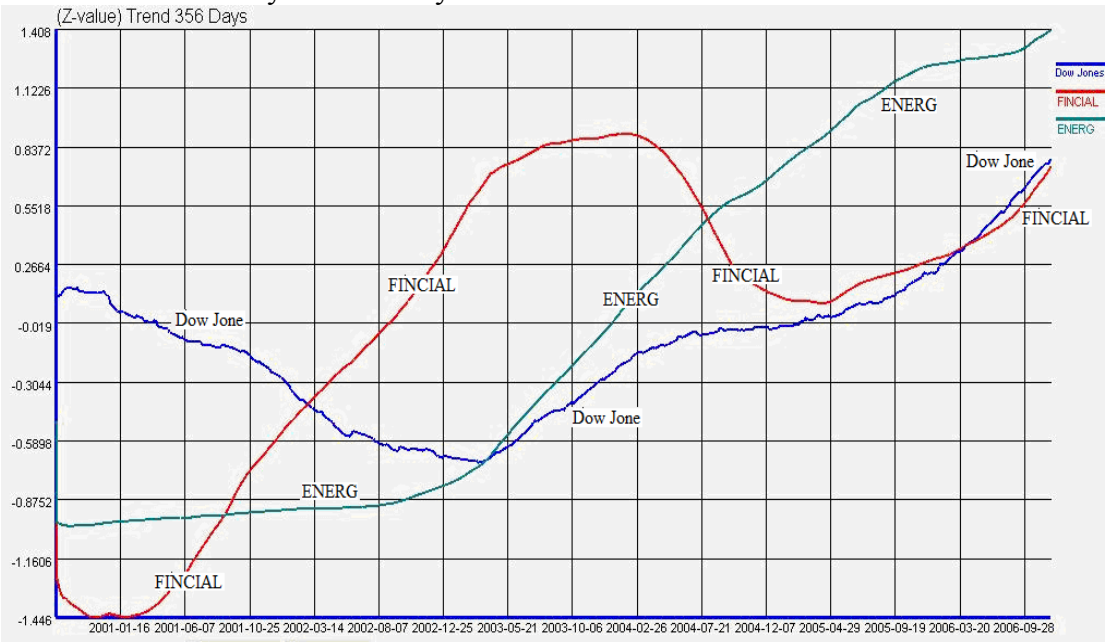


Figure 5.21: FINCIAL, External Stock and ENENG Dissimilarity Graph

Figure 5.21 illustrates the trend of dissimilar movement of data, calculated from data of 1 year. It can be seen that the trend of FINCIAL industry index, External Stock index and ENENG sector index goes in the different direction

5.4 Discussion

In conclusion, the data analysis begins with similarity verification of 2 sets of data, of which the scale has been adjusted with the Standard Deviation Normalization technique. Then the similarity of time series data is verified with the Regression Time Warping technique. After that the sets of time series data which are most similar and the sets which are most different are analyzed in regard to Trend Movement, Seasonal variation and Irregular variation.

To compare the experimental results with some related works, which are:

1. In Experimental I (oil price and stock industry index) from the most similar data obtained from Table 5.2 have, which are SERVICE Industry and oil price, or mean oil price, have relationship with SERVICE Industry. The result is consistent with Sudaporn Pengmol's research (2004)[6], proposed in her articles entitled "**The study on the consequences of oil prices variation on securities index for**

individual industrial group" was discovered in data of Commerce in COMM Sector stock, Hotel and Touristic services in TOURISM Sector, Printing and Publications in MEDIA Sector, and Transports in TRANS Sector that are inside a SERVICE Industry stock have long-term relationship with oil prices rise.

2. In Experimental I (oil price and stock industry index) from the most dissimilar data obtained from Table 5.2 have, which are FINCIAL Industry, TECH Industry, and oil price, or mean oil price have no relationship with FINCIAL and TECH Industry. The result is consistent with Sudaporn Pengmol's research (2004)[6] proposed in her articles entitled "**The study on the consequences of oil prices variation on securities index for individual industrial group**" was discovered in data of Banking in BANK Sector stock, that is inside a FINCIAL Industry stock have no long-tern relationship with oil prices rise.

3. In Experimental III (all economic factors) from the most dissimilar data obtained from Table 5.6 have, which are US dollar exchange rate was different from US Dow Jones stock exchange. The result is consistent with Syed A. Basher and Perry Sadorsky's research (2006)[37], in their articles entitled "**Oil price risk and emerging stock markets**" was discovered in money exchange rate have no relationship with many external stock exchange.

4. In Experimental III (all economic factors) from trend of the most dissimilar data graph as illustrated in Figures 5.17 and 5.18, which are oil price direction in the opposite money exchange rate from year 1999 to 2004. The result is consistent with Ingunn Fride Tvete and Anders Loland's research (2007)[44], in their articles entitled "**Oil prices and exchange rates: A survey**" was discovered in oil price had relationship opposite money exchange rate.

CHAPTER VI
CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

In this research, data on economic factors fluctuating from January 1, 1997 to December 31, 2007 consisted of 6 data groups of oil price, gold bar price, Foreign Exchange rate, Bank Interest rate, External Stock index, and Stock Exchange of Thailand (SET). They were compiled from different web sites and recorded in data warehouse through stages of Extract Transform Load (ETL). Later, the data were analyzed with the OLAP tool used for retrieving interesting data in different views.

For data analysis, the concept of Data Mining was used, which was a survey on new knowledge from the existing considerable data, using the technique of Time Series Analysis was applied for creating Time Series Analysis tool in order to be used to search for the attribute of components of time series data in terms of Trend, Seasonal Variation, Cyclical Variation and Irregular Variation. In addition the technique of Similarity Analysis to generate Similarity tool used as an instrument for searching for the similarity of 2 sets of time series data.

The following hypotheses have been proposed.

1. **To study the relationship of oil price and Stock Industry index data.**

Based on the experimental consequences, oil price data and Stock Industry index which are most similar include those pairs of SERVICE Industry index and Benzene 91 price (0.5285), SERVICE Industry index and Diesel price (0.5249), and SERVICE Industry index and Benzene 95 price (0.4010). Or Oil price group data have the most influence on SERVICE Industry index.

2. **To study the relationship of gold bar price and interest rate.**

Based on the experimental results, foreign (US) gold bar price and MOR loan rate (0.250509) or gold bar price data and interest rate have less than 30% similarity.

3. **To study the relationship of data on all economic factors based on the representatives of data for all 6 data groups.**

Based on the outcomes of the experiment on all economic factors data, it is found that the pairs of data with the most similarity are those of 3-month fixed deposit rate and MLR loan rate (0.737241), Benzene 91 price and domestic gold bar price (0.699158), Benzene 91 price and ENERG Sector index (0.682398), as well as domestic gold bar price and ENERG Sector index, as revealed in Table 6.1, a summary of economic factors similarity.

Table 6.1: Summary of Economic Factors Similarity

No	Independent Variable	Dependent Variable
1	Deposit interest rate	Loan interest rate
2	Oil Price	Gold bar price
		ENERG Sector Index
3	Gold bar price	ENERG Sector Index
		Foreign exchange rate

For data pairs with the most difference they are those of FINCIAL Industry index and 3-month fixed deposit rate (0.000021), US Dow Jones Stock index and 3-Month fixed deposit rate (0.00113), US Dow Jones Stock index and FINCIAL Industry index (0.001839), FINCIAL Industry index and ENERG Sector index (0.005846), as well as 3-month fixed deposit rate and Benzene 91 price, as displayed in Table 6.2, a summary of economic factors dissimilarity.

Table 6.2: Summary of Economic Factors Dissimilarity

No	Independent Variable	Dependent Variable
1	Deposit interest rate	FINCIAL Industry Index
		External Stock index
		Oil price
2	FINCIAL Industry Index	External Stock index
		ENERG Sector Index

6.2 Discussion

The results from the data analysis are not 100% accurate because of the following reasons:

1. There are no economists participating in this research project. An economist should be involved in selecting economic factors to be analyzed and in interpreting the analysis results.
2. Data classification during data preparation may give wrong meanings to the data. For example, the "Super" price data and the "Benzene95" price data are classified as the same group of data, "Benzene95". These two types of gasoline are both non-leaded (from the official web site of Bangchak Company http://www.bangchak.co.th/th/productStation.asp) but still they are different types of gasoline. Grouping them together may results in false meaning.
3. Lastly, the analysis based on daily data causes errors in data property analysis of time series, especially the data which have been "normalized" with Previous Data Change or Moving Average. The graph obtained shows data with abnormalities since the raw data gathered are not on daily basis. (There are no data as of Saturday, Sunday and public holidays.) Therefore, to achieve the correct data analysis, the analysis should be does based on monthly data instead of daily data.

6.3 Future Work

To have a more complete research project in the future, the following procedures should be added.

1. Since the similarity value of time series data is calculated based on data at a period of time to another period, there arises a problem of which time period is appropriate for computing similarity value of data. This research proposes collecting a myriad of statistical values of similarity of time series data at different period of time. Then, the consequences obtained should be analyzed by an economist to come to a conclusion which time period is the most suitable.

2. Beside, there is another problem of selecting economic factor variables to be analyzed so as to find which pair of variables is most appropriate for data analysis. This research thus proposes finding data pattern through the pattern recognition technique. Once an obvious pattern is derived, a clustering approach is used for grouping economic factors data. Subsequently, data within the same or different groups are analyzed.
3. The 6 economic factors included in the analysis may not cover all economic problems. Accordingly, for further study on this subject, there should be an economist who can recommend what additional economic factors should also be analyzed.
4. For the study in the future, for the sake of more precise findings of the study, a larger amount of time series data, like those of over 10 years, should be applied.
5. The patterns of raw data to be analyzed are fixed according to the program, so to have flexibility in analysis of other economic factors, the program should be able to recognize the raw data in configuration file which is XML file. Then the program is able to read the pattern of raw data in XML file instead of the pattern fixed by the program.
6. To achieve a more practical software tool the requirements of economists should be gathered. And the software tool should be tested and evaluated by the economists.

REFERENCES

1. Ad Feelders. (2003). Data Mining in Economic Science. Institute of Information & Computing Sciences. Retrieved January 17, 2007, from http://people.cs.uu.nl/ad/dmecon.pdf.
2. Bank of Thailand. (2007, February 15). Retrieved from http://www.bot.or.th.
3. Bundit Chaiwitchayachat. (2006). ผลกระทบจากการเปลี่ยนแปลงราคาน้ำมันภายในประเทศที่มีต่อระบบเศรษฐกิจไทย. *Journal of Economics* (Vol. 1). Kasetsart University: Faculty of Economics.
4. Bundit Chaiwitchayachat. (2007, February). ผลกระทบของราคาน้ำมันที่มีต่อระบบเศรษฐกิจไทย:ต้นแบบเศรษฐกิจมหภาค. Chulalongkorn University: Faculty of Economics.
5. Chotirat Ann Ratanamahatana, Jessica Lin, Dimitrios Gunopulos, Eamonn EamonnKeogh, Michail Vlachos and Gautam Das. (2005). Mining Time Series Data. *Data Mining and Knowledge Discovery Handbook*. Springer Publishers.
6. DaimlerChryslers AG. (2007). Cross–Industry Standard Process for Data Mining. Retrieved March 1, 2007, from http://www.crisp-dm.org.
7. Eamonn Keogh, Selina Chu, David Hart, and Michael Pazzani. (2001). Segmenting Time Series: A Survey and Novel Approach. Retrieved May 20, 2007, from http://www.ics.uci.edu/~pazzani/Publications/survey.pdf.
8. EIA, Crude Price File (2007, February 1). Retrieved from US Government: Energy Information Administration: http://tonto.eia.doe.gov.
9. Eugene Fink and Kevin B. Pratt. (2002). Indexing of Compressed Time Series. Retrieved March 20, 2007, from http://citeseer.ist.psu.edu/539566.html.

10. External Stock Index File (2007, May 1). Retrieved from Bank of Thailand: http://www.bot.or.th/Thai/Statistics /FinancialMarkets/CapitalMarket/Pages/StatCapitalMarket.aspx

11. Gautam Das and Dimitrios Gunopulos. (2003). Time Series Similarity and Indexing. *The Handbook of Data Mining*. New Jersey: Lawrence Erlbaum Associates: Mahwan Publishers.

12. Gerben Driesprong,Ben Jacobsen and Benjamin Maat. (2004, September). Stock Markets and Oil Prices. *Social Science Electronic Publishing*. Retrieved February 29, 2007, from Erasmus University. Retrieved May 20, 2007, from http://www1.fee.uva.nl/fm/PAPERS/Stock Markets and Oil Prices sept2004.pdf.

13. Gold US Price File. (2007, February 15). Retrieved from Kitco Inc - Past Historical London Fix**,** http://www.kitco.com.

14. Gold Baht Price File. (2007, February 15). Retrieved from Ranthong http://www.ranthong.com.

15. Hansheng Lei, Srinivas Palla and Venu Govindaraj. (2004). An Intuitive Similarity Measure for On-Line Signature Verification. Buffalo University: Center for Unified Biometrics and Sensors. Retrieved May 20, 2007, from http://portal.acm.org/citation.cfm?id=1387882.

16. Hansheng Lei and Venu Govindaraju. (2004). Matching and Retrieving Sequential Patterns under Regression. Retrieved May 20, 2007, from http://ieeexplore.ieee.org/iel5/9689/30573/01410787.pdf.

17. Hansheng Lei and Venu. Govindaraju. (2004). GRM: Generalized Regression Model for Clustering Linear Sequences. Retrieved May 20, 2007, from http://www.siam.org/proceedings/datamining/2004/sdm04_003leih.pdf.

18. Hansheng Lei and Venu. Govindaraju. (2004). Regression Time Warping for Similarity Measure of Sequence. Buffalo University: Center for Unified Biometrics and Sensors. Retrieved May 20, 2007, from http://portal.acm.org/citation.cfm?id=1025405.

19. Interest File. (2007, February 1). Retrieved from Bank of Thailand http://www.bot.or.th/thai/statistics /financialmarkets/interestrate/_layouts/application/interest_rate /IN_Rate.aspx.

20. Ina Naydenova and Kalinka Kaloyanova. (2005). Basic Approaches of Integration between Data Warehouse and Data Mining. Retrieved April 1, 2007, from http://www.nbu.bg/PUBLIC/IMAGES/File/departamenti/informatika.

21. Jiawei Han and Micheline Kamber. (2006) *Data Mining: Concepts and Techniques* (Second Edition). Morgan Kaufmann Publishers.

22. Magnus Lie Hetland. (2007). A Survey of recent methods for Efficient Retrieval of Similar Time Sequences. *Series in Machine Perception and Artificial Intelligence* (Vol.57). World Scientific Publishing.

23. Marc Labonte. (2004, June). The Effects of Oil Shocks on the Economy: A Review of the Empirical Evidence. Retrieved February 1, 2007, from The Library of Congress, Congressional Research Service: http://opencrs.com/getfile.php?rid=65075.

24. Mark Last, Abraham Kandel and Horst Bunke. (2004). *Data Mining in Time Series Databases*. World Scientific Press Publishing.

25. Mehmed Kantardzic. (2003). *Data Mining: Concepts, Models, Methods, and Algorithms*. John Wiley & Sons Publishing.

26. Mehmet Satay. (2006). Time Series Analysis and Mining Techniques. *Idea Group Publishing Encyclopedia of Data Warehousing and Mining* (pp1120-2006). USA: Hewlett-Packard Labs.

27. Michael Leonard and Brenda Wolfe (2003) Mining Transaction and Time Series Data. Retrieved February 10, 2007, from http://biocomp.cnb.csic.es/~coss/Docencia/ADAM/Sample/Session0_Leonard.pdf.

28. Michail Vlachos and Dimitrios Gunopulos. (2004) *Indexing Time-Series under Conditions of Noise*. World Scientific Publishing.

29. Money Exchange Rate File. (2007, March 1). Retrieved from Bank of Thailand, http://www.bot.or.th/Thai/Statistics/FinancialMarkets/ExchangeRate/_layouts/Application/ExchangeRate/ExchangeRate.aspx.

30. Oil Price File. (2007, March 1). Retrieved from Shell http://www.shell.com/home/content2/thailand-th/shell_for_ motorists/price _board/historyprice_1118.html.

31. Sarun Thawilwung.(2004). *ผลกระทบของราคาน้ำมัน ต่อ SET Index โดยใช้ Quantitative analysis*. Far East Securities Company.

32. Siriwan Kunthapree. (2006*). ศึกษาถึงผลกระทบของการเปลี่ยน แปลงราคาน้ำมันที่มีผลต่อตัวแปรทางเศรษฐกิจมหภาคของระบบเศรษฐกิจไทย*. Chiang Mai University: Faculty of Economics: Economics Library.

33. Somkiat Getaiem. (2005) *เทคนิคการพยากรณ์*, Thaksin University Publishing.

34. Stock Exchange of Thailand. (2007, February 1). Retrieved from http://www.set.or.th.

35. Sudaporn Pengmol. (2006). *ศึกษาถึงผลของการเปลี่ยนแปลงราคาน้ำมันต่อดัชนีหลักทรัพย์รายกลุ่มอุตสาหกรรมจำนวน 20*. Chiang Mai University: Faculty of Economics: Economics Library.

36. Syed A. Basher and Perry Sadorsky (2006). Oil price risk and emerging stock markets. *Global Finance Journal*, 17(2), 224-251. Retrieved February 1, 2007, from http://ideas.repec.org/p/wpa/wuwpif/0410003.html.

37. R. Agrawal and R. Srikant. (1995, March). Mining sequential patterns. Taiwan: Taipei: In Proc. of the International Conference on Data Engineering. Retrieved February 3, 2007, from http://citeseer.ist.psu.edu/agrawal95mining.html.

38. Richard J. Povinelli. (2004, December). Time Series Data Mining: Identifying Temporal Patterns for Characterization and Prediction of Time Series Events. Retrieved February 3, 2007, from http://citeseer.ist.psu.edu/povinelli99time.html.

39. Tak-chung Fu, Fu-lai Chung, Robert Luk and Chak-man Ng.(2007). Stock time series pattern matching:Template-based vs. rule-based approaches. Retrieved February 5, 2007, from http://portal.acm.org/citation.cfm?id=1230538.

40. Woravit Suriyawong. (2006). *การศึกษาความสัมพันธ์ระหว่างราคาหลักทรัพย์ในกลุ่มพลังงานที่ประกอบธุรกิจเกี่ยวกับน้ำมันกับราคาน้ำมันในตลาดต่างๆ*. Chiang Mai University: Faculty of Economics: Economics Library.

41. Pakin Jitpokkasame. (2007) *ผลกระทบของความผันผวนของอัตราแลกเปลี่ยนต่อระดับราคาสินค้าในประเทศไทย*. Kasetsart University: Faculty of Economics.

42. Jakree Sakunbongkot. (2001). *ผลกระทบของของอัตราแลกเปลี่ยนต่อราคาส่งออกสินค้าอุตสาหกรรมของไทย*. Chulalongkom University: Faculty of Economics.

43. Robert A. Amano and Simon van Norden. (1997, May). Oil Prices and the Rise and Fall of the U.S. Real Exchange Rate. Bank of Canada: Research Department. Retrieved March 1, 2007, from http://neumann.hec.ca/pages/simon.van-norden/wps/jimf98.pdf.

44. Ingunn Fride Tvete and Anders Loland. (2007 October). Oil prices and exchange rates: A survey. Norsk Regnesentral Norwegian Computing Center. Retrieved February 1, 2007, from http://neumann.hec.ca/pages/simon.van-norden/wps/RIEfxoil.pdf.

45. SAS (2007, October 20). Retrieved from SAS Institute http://www.sas.com.

46. SPSS (2007, October 20). Retrieved from SPSS Institute http://www.spss.com.

47. STATS (2007, October 20). Retrieved from StataCorp Company http://www.stata.com.

48. Eviews (2007, October 20). Retrieved from Quantitative Micro Software. http://www.eviews.com/index.html.

49. Alan Acock. (2005, November). SAS, Stata, SPSS: A Comparison. *Journal of Marriage and Family* (Vol. 67, pp. 10093-1101). Blackwell Publishing.

BIOGRAPHY

NAME	Mr. Suwichai Gulthammanusorn
DATE OF BIRTH	28 November 1969
PLACE OF BIRTH	Bangkok, Thailand
INSTITUTIONS ATTENED	Kasertsart University, 1988-1992: Bachelor of Science (Agriculture) Mahidol University, 1992-1994: Bachelor of Science (Computer Science) Mahidol University, 2003-2008: Master of Science (Computer Science)
POSITION&OFFICE	Freewill Solution Company Limited 28 th Fl Lumpini Tower, Rama IV Rd, Tungmahamek Sathorn, Bangkok, 10120 Thailand. Position: Expert Specialist Tel 086-609-8650 E-Mail suwichai_gul@freewillsolutions.com
HOME ADDRESS	107 Charunsanitworg 42, Bangyeekhun, Bangpad, Bangkok, Thailand Tel. 086-609-8650