The Application of Chemometrics Derived Pattern Recognition Methods to Futures Market Analysis

A dissertation presented

by

Tao Yu

to
The Department of Chemistry and Chemical Biology

In partial fulfillment of the requirements for the degree of
Doctor of Philosophy

in the field of

Chemometrics Market Analysis

Northeastern University
Boston, Massachusetts
March, 2009

The Application of Chemometrics Derived Pattern Recognition Methods to Futures Market Analysis

By

Tao Yu

ABSTRACT OF DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry
in the Graduate School of Arts and Sciences of
Northeastern University, March 2009

ABSTRACT

Cycle-theory-based market analysis is the main focus of this thesis, in which I try to find systematic methods to recognize and utilize market patterns, especially cyclic ones, to obtain a correct understanding of market movements. The KNN algorithm, a pattern recognition method extensively used in chemometrics, has been employed to recognize similarities of current market movements and historical markets to permit market forecasts. Bayesian analysis, another pattern recognition method, has been used to infer longer-term market trends based on observable shorter-term market behaviors and to improve the real-time application of the KNN algorithm. An artificial neural network method, an example of a non-linear information processing system, has also been applied in this research to combine cycle-relative information to market behavior modeling. The promising overall results show that there exists a correlation between current and historical price movements and shows the possibilities of utilizing pattern recognition methods to obtain correct market forecasts.

Also, a novel use of moving averages, especially suitable for an oscillating market, has been introduced and successfully applied in market prediction. In this thesis, the S&P 500 futures market has been chosen as the market to study.

Acknowledgements

The completion of this dissertation could not have been possible without the support and guidance of my advisor, Prof. Bill C. Giessen, who introduced me to the fascinating world of chemometrics-based market analysis. His insightful vision, timely encouragement, detailed and constructive suggestions help me to proceed through the doctoral program and complete my dissertation. Also, Prof. Giessen's perpetual energy and enthusiasm in research have deeply motivated me and will become a source of inspiration to me.

I want to express my warm and sincere thanks to the members of the dissertation committee, Prof. Graham Jones, Prof. Max Diem, Prof. Kamran Dadkhah and Prof. Nicole Boyson, for their expertise and advice. Special thanks to Prof. Max Diem for his detailed corrections and comments.

I also want to thank Mr. Martin Wolf and Cambridge Market Analysis Corporation (CMAC) for the source of funding to this project.

Furthermore, I would like to express thankfulness to my colleagues and friends who helped me throughout the course of this dissertation. Many thanks to Dr. Jian Yao, Dr. Ke Xu, Dr. Zhaoyang Zhao and Dr. Haijiang Peng for their help and collaboration.

Finally, none of my success would have been possible without the love and support of my family members: my dear wife Yan Tian, my dear mother Ping Xiao and my dear father Yongdeng Yu.

TABLE OF CONTENTS

List of Figures

List of Tables

List of Abbreviations

FA	Fundamental analysis
TA	Technical analysis
FTVision	An in-house trading display system
C1	1-day cycle
C2	2.5-day or half-week cycle
C5	1-week cycle
C10	Half-month cycle
C20	1-month cycle
CQ	1-quarter cycle
CY	1-year cycle
SRT	Simulated real-time
CPL	Cumulative profit and loss
TIM	Time-in-market
FOM2	A figure-of-merit
KNN	K nearest neighbors
TFx	A set of forecasting models, including TF1, TF2, TF3, TF4, TF5, TV1 and TF2B used in this thesis
UD	Up or down move
LS	Leg-length-ratio
ANN	Artificial neural network
MA	Moving average
SMA	Simple moving average
WMA	Weighted moving average

Chapter 1 General Introduction

1.1 Introduction to Futures and Futures Markets

A futures contract is a type of financial contract or agreement in which two parties agree to transact a set of financial instruments or physical commodities at a future date and at a specific price, known as futures price. I will briefly explain the rational for futures use in this case.

Before the appearance of futures contracts, farmers would grow crops and then bring them to the market at harvest time. Due to lack of anticipating information, mismatches of supply and demand could have severe consequences; when supply exceeded what was needed, unsold crops were left to rot; conversely, when crops were in short supply, downstream processors had to pay high prices.

With the aim of providing an efficient and effective mechanism for the management of price risks, in the mid-nineteenth century, a futures market for grain commodities was established in Chicago, USA, for farmers to bring their crops and sell them to customs, like millers, either for immediate delivery (spot trading) or for future delivery (forward trading). For example, if a farmer thinks the price of wheat is going to fall by harvest time, he can sell a futures contract in wheat to lock in the price. That way, if the cash price of wheat does fall by harvest time, losing the farmer money, he will make back the cash-loss by profiting on the sale of the futures contract. On the other hand, if the cash price of wheat goes up, he still needs to sell at futures price. This process reduces the loss of crops for farmers and helps to stabilize supply and prices.

In the process two parties are involved, one party agrees to deliver a commodity (taking a short position), and the other party agrees to receive a commodity (taking a long position). In above scenario, the farmer would be the holder of the short position, while the miller would be the holder of the long position.

Futures markets are centralized marketplaces where buyers and sellers from around the world meet (now often electronically) and enter into futures contracts. Today's futures market is a global marketplace for not only agricultural goods, but also for currencies (forex) and financial instruments such as stock indexes, and treasury bonds. Because the futures market is both highly active and central to the global marketplace, it's a good source for vital market information and sentiment indicators. Since futures market prices use a continuous flow of information from around the world, factors such as weather, war, debt default, land reclamation and the way people process information all have a major effect on supply and demand, and as a result, the futures price.

Futures markets are also a mechanism for people to reduce risk and uncertainty[1-4]. Because the price is pre-set, participants know how much they will need to buy or sell; therefore, the risk of transacting at an unfavorable future price is reduced.

A futures contract is a standardized, transferable, exchange-traded contract. In every futures contract, the terms such as the quantity and quality of the commodity, the specific price per unit, and the date and method of delivery are specified. For example, a gold futures contract (GC) holds 100 troy ounces of 24-carat pure gold; and a Crude Oil futures contract holds 1000 barrels of crude oil of a certain quality. The price of a futures contract is represented by the agreed-upon price of the underlying commodity or financial instrument that will be delivered in the future.

There are a number of futures contracts traded in the futures markets, but in this thesis, for the purpose of study, we only focus on the S&P e-mini contract. Each “point” in the S&P e-mini contract represents $50. Therefore, if you buy 1 S&P mini contract at 1,500 and it moves to 1,506, representing a 6 point move, then you have made $300.

Generally, there are two types of futures contracts, one type provides for physical delivery of a particular commodity or product, and the other requires a cash settlement. However, participating in the futures market does not necessarily mean that a participant will be responsible for receiving or delivering possibly large inventories of physical commodities; buyers and sellers in the futures market enter into futures contracts primarily to hedge risk or to speculate rather than to exchange physical goods, the latter is the primary activity of the cash/spot market. The vast majority of market participants choose to buy or sell offsetting futures contracts prior to the delivery date to realize gain or loss, which for a "long" is the difference between the buying and selling price, respectively.

Spurred by the need to manage price and interest rate risks that exist in virtually every type of modern business, today's futures markets have also become major financial markets. Generally speaking, there are two main types of futures trader[1]: hedgers and speculators.

A hedger can be a producer of the commodity, i.e., a farmer, an oil company, or a mining company, who buys or sells in the futures market to secure the future price of a commodity intended to be sold at a later date in the cash market. This helps protect against price risks. Other hedgers include banks, insurance companies and pension fund companies who use futures to hedge against any fluctuations in the cash price of their investments at future dates.

Speculators include independent floor traders and private investors. Unlike hedgers, speculators don’t hold any commodities or underlying financial instruments and have no desire to own the commodity. Rather, they enter the market seeking profits by buying or selling now in anticipation of rising or declining prices. In other words, they invest in futures in the same way they might invest in stocks by buying at a low price and selling at a higher price.

Hedgers want to minimize their risk no matter what they're investing in, while speculators want to maximize their profits at the price of increasing risk. In order to achieve protection against unfavorable price changes, hedgers are willing to give up the opportunity to gain from favorable price movements. Speculators aim to profit from the very price change that hedgers are protecting themselves against. The interaction of hedgers and speculators helps to maintain active, liquid and competitive markets.

Different from the stock market, there are two distinct characteristics of futures trading in the futures market: low margin and high leverage. In the futures market, margin refers to the initial deposit made into an account in order to trade a futures contract. Rather than providing a down payment, the margin required to buy or sell a futures contract is solely a deposit that can be drawn on by the brokerage firm to cover losses that the client may incur in the course of futures trading. Minimum margin requirements for a particular futures contract at a particular time are set by the exchange on which the contract is traded. They are typically about five percent of the current value of the futures contract. Exchanges continuously monitor market conditions and risks, and raise or reduce their margin requirements if necessary.

Generally, there are two kinds of margins: initial margin and maintenance margin. Initial margin is the amount of funds that will be required to be available in the futures account to purchase futures contracts. On any single day, if there are profits on the open positions, they will be added to the balance in the margin account. Conversely, the losses will be deducted from the balance in the margin account.

Maintenance margin specifies the minimum amount of money that must be in the margin account. It is used to maintain the futures postion. If the funds remaining available in the margin account are reduced by losses to below the maintenance margin requirement, the broker will require additional funds to bring the account back to the level of the initial margin, also known as margin calls. If no additional funds are deposited into the margin account, the broker has the right to liquidate the open positions. Table 1.1 shows examples of the margin requirements for some futures contracts.

Table 1.1 Margin requirements of some futures contracts

Commodity	Exchange	Contract Size	Initial Margin	Maintenance Margin
E-mini S&P 500	CME	$50 x S&P index	$6,188	$4,950
T-Bonds	CBOT	$100,000	$4,320	$3,200
Gold	COMEX	100 oz.	$5,808	$4,302
Silver	COMEX	5,000 oz.	$8,640	$6,400
Euro	CME	€125,000	$6,345	$4,700

CME: Chicago Mercantile Exchange
COMEX: Commodity Exchange, NYC

In the futures market, leverage refers to having control over large cash values of commodities with comparatively small levels of capital. In other words, with a relatively small amount of cash, traders can enter into a futures contract that is worth much more than they initially have to pay (initial margin). Because of this low initial margin requirement (set by the futures exchange), futures positions are highly leveraged. The smaller the margin in relation to the cash value of the futures contract, the higher the leverage.

1.2 General Aspects of Financial Analysis

In financial markets, including futures markets, the price of a security represents the consensus of market participants, both hedgers and speculators, to trade at the level of a momentary equilibrium for the security[1]. The price changes with time, either going up or down, but it is axiomatic that it will converge to a consensus value after some price variations. Investors trade the security primarily based on their expectations. If they think the current price is higher or lower than what the security is worth, they will sell or buy it, respectively. These expectations are always changing, causing prices to oscillate between overbought and oversold levels. Fluctuations in prices are a natural process of changing expectations and lead to cyclical patterns.

Financial analysis is the process of analyzing the financial data of the security or market statistics of price movements to estimate the underlying value and form an expectation[5-20]. By comparing to the current security price, investors can make investment decisions assuming that the market will correct and move towards the estimated value at some point in the future.

There are two categories of financial analysis: fundamental and technical analysis. Fundamental analysis involves an examination of the economy, a particular industry, supply and demand data, and financial company data in order to lead to an estimate of value for a company or commodity. By contrast, technical analysis relies on price history in order to predict the future.

Fundamental analysis utilizes a much wider range of information than does technical analysis and relies on traditional financial statement analysis. Technical analysis, on the other hand, concerns itself with attempting to identify patterns in past price movements. Although both fundamental and technical analysts agree that the price of a security is determined by the interaction of supply and demand, technical analysts and fundamental analysis have different opinions on the influence of irrational factors. A technical analyst might expect that an irrational influence may persist for some time, whereas other market analysts would expect only a short-run effect with rational beliefs prevailing over the long run.

In this thesis, we focus on the use of Technical Analysis, based on data processing methods that an analytical chemist (and some physical ones) may be familiar with, including chemometrics, pattern recognition, statistical (Bayesian) inference, and artificial intelligence, etc. The security being studied is S&P futures, but the techniques developed here can also be applied to other financial products.

For completeness sake, we touch on fundamental analysis first, in the next section.

1.2.1 Fundamental Analysis

Fundamental analysis is the process of looking at a considered transaction at the basic or fundamental financial level, rising from company, to industry and country level[5-9]. The goal is to determine a current intrinsic value and potential profit from future price movements. At the company level, usually, information such as financial statements, management, business concept and competition will be examined. At the industry level, analysis of industry factors such as total sales, price levels, the effects of competing products, foreign competition, entry or exit from the industry, will be conducted. For the national economy, macroeconomic data, i.e. GDP growth rates, inflation, exchange rate, etc, will be evaluated to assess the present and future growth of the economy.

Continuing to take stock prices as an example, generally, fundamental analysts employ the top-down method to study the stock market. They will first start with national economy by analyzing economic indicators. After getting the picture of the current economic situation and future development, they will narrow down to industry analysis. In the end, they will search for the best business in a promising industry. By combining country, industry, and company analysis together, fundamental analysts try to derive a stock's current fair value and forecast future value. The difference between the current and fair values reflects their assessment of the stock’s potential as an investment opportunity. Believing that prices do not accurately reflect all available information, fundamental analysts look to capitalize on perceived price discrepancies.

For long-term investment, fundamental analysis is helpful as it takes long-term economic development into consideration. It is useful for the thorough understanding of the business. Sound fundamental analysis will help identify companies that represent a good value and uncover companies with valuable assets, a strong balance sheet, and strong competition power.

Fundamental analysis may offer excellent insights, but it can be extraordinarily time-consuming. Valuation techniques vary depending on the industry group and specifics of each company. For this reason, a different technique and model is required for different industries and different companies. This can be quite time-consuming, which can limit the amount of research that can be performed. Meanwhile, since a lot of data were involved in fundamental analysis, not many individual investors can easily access to that kind of resources, thus, it limits the application of fundamental analysis to general investors.

The numbers that a fundamentalist analyzes are only released over long periods of time. Financial statements are filed quarterly and changes in earnings per share don't emerge on a daily basis like price and volume information. Part of the reason that fundamental analysts use a long-term timeframe, therefore, is because the data they use to analyze a stock is generated much more slowly than the price and volume data used by technical analysts.

Fundamental analysis is suitable for long-term investment but not as good for the short-term. The market price variations in the short term can’t be explained by fundamental analysis since there is nearly no change in economic factors in the short term.

Furthermore, fair value evaluation is based on assumptions and can be very subjective. Any changes to growth or multiplier assumptions can greatly alter the ultimate valuation.

Also, as recently seen, emotional factors enter in a major way that is difficult to assess.

1.2.2 Technical Analysis

Technical analysis is a method of evaluating securities by analyzing the statistics generated by past market activity, such as past prices and volume. Unlike fundamental analysis that attempts to measure a security's intrinsic value, technical analysis relies on the use of charts and mathematical techniques to examine various aspects of a security's price movement to identify patterns that can suggest future activity. It studies supply and demand in a market in an attempt to determine what direction, or trend, will continue in the future. In other words, technical analysis attempts to understand the emotions in the market by studying the market itself, as opposed to its components.

The field of technical analysis is based on three assumptions derived from traders' experiences: the market discounts everything, price moves in trends and history tends to repeat itself[21].

Technical analysis assumes that, at any given time, a stock's price reflects everything that has or could affect the company, including fundamental factors. Technical analysts believe that the company's fundamentals, along with broader economic factors and market psychology, are all priced into the stock or commodity, removing the need to actually consider these factors separately. Since all information is already reflected in the price, which is viewed as a product of supply and demand for a stock or commodity, it represents the fair value, and should form the basis for analysis.

In technical analysis, price movements are believed to follow trends. This means that after a trend has been established, the future price movement is more likely to be in the same direction as the trend than to be against it. Figure 1.1 illustrates an example of a trend line for Apple Computer Inc. from 10/2004 to 02/2006. In the graph, an up-trending line is being plotted based on the price movements.

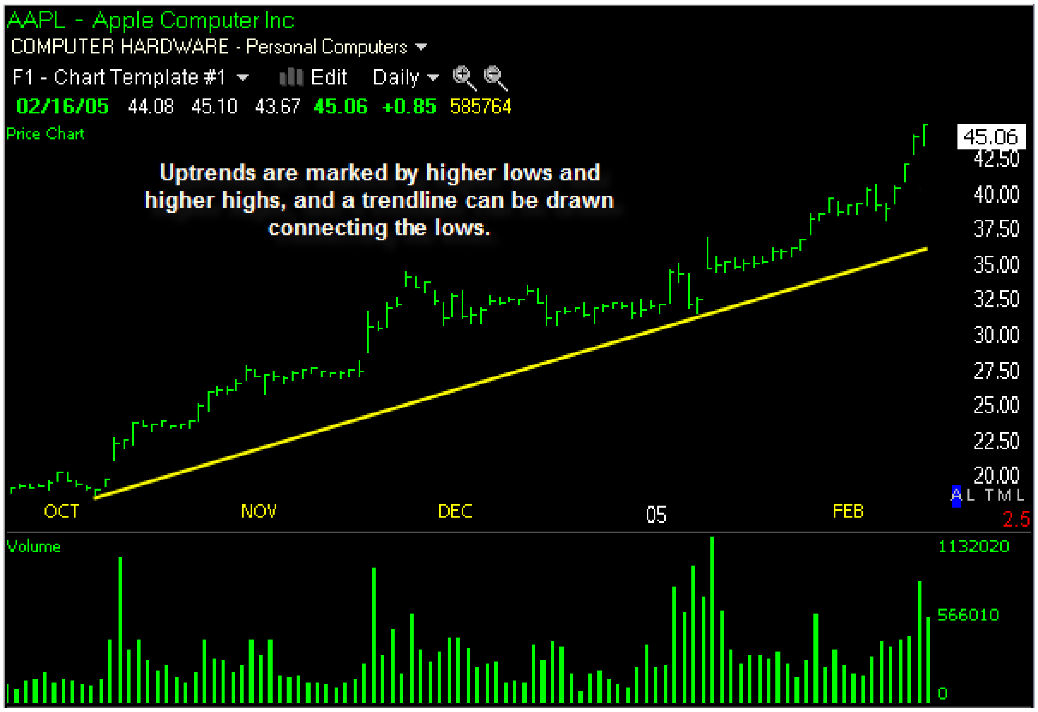


Figure 1.1 The price chart of AAPL[21] from 10/2004 to 02/2005.

Another important idea in technical analysis is that history tends to repeat itself, mainly in terms of price movement. The repetitive nature of price movements is attributed to market psychology; in other words, market participants tend to provide a consistent reaction to similar market stimuli over time. Technical analysis uses chart patterns to analyze market movements and understand trends. A double top, a kind of chart pattern[23-24], is shown in Figure 1.2. Combined with use of an up-trend line, a double top pattern provides a reversal signal to the stock. This chart pattern has been used for a long time, but it is still believed to be relevant because the pattern often repeats itself.

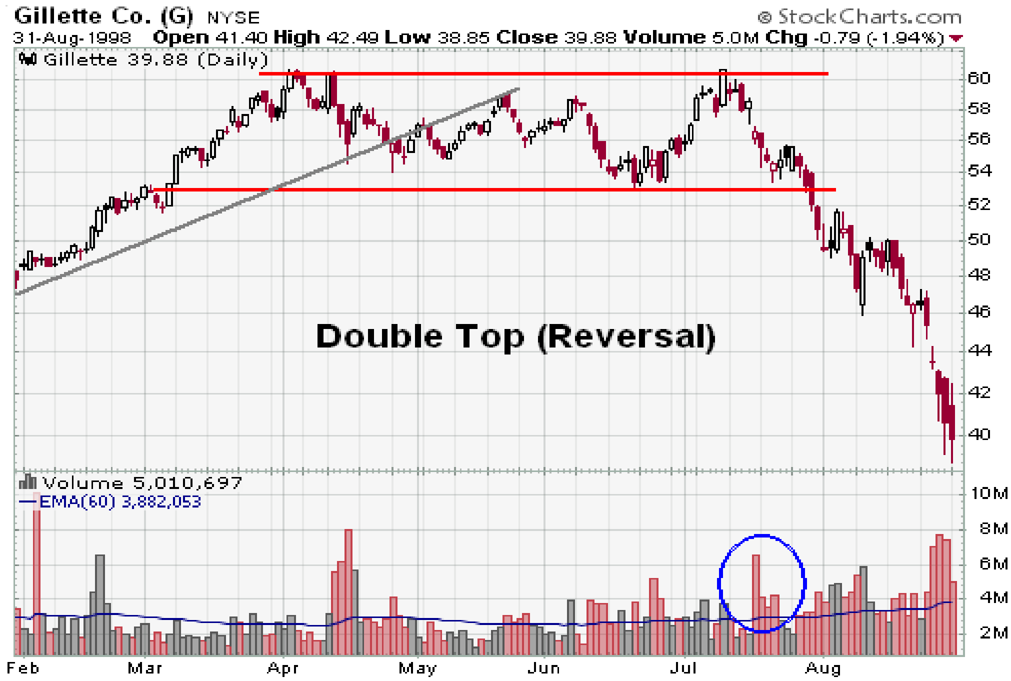


Figure 1.2 Double top chart pattern for Gillette Co. stock[22] from 02/1998 to 08/1998.

Technical analysis on this visual level is relatively easy to understand and operate, even for individual investors. There is no need to collect and study massive economic, industry and company data; instead, technical analysts only focus on price and volume charts, and rely on either patterns or indicators to determine the future development of the security.

Technical analysis is good for short-term investment because it focuses more on short-term price movement, and tries to find out whether the present price level is overbought or oversold. In common practice, fundamental analysis and technical analysis can complement each other. For example, some fundamental analysts may use technical analysis techniques to figure out the best time to enter into an undervalued security. By timing entry into a security correctly, the gains on the investment can be greatly improved.

Just as fundamental analysis, technical analysis is also subjective and personal biases can be reflected in the analysis[25-27]. Even though there are standards, often different technicians will look at the same chart and come up with different scenarios or see different patterns.

Much of the criticism of technical analysis is based on the “efficient market hypothesis” (EMH)[28-29], which has three versions: the strong version, the semi-strong version and the weak version. I will only give the basic tenet of the weak version. The theory says that the market's price is always the correct one - any past trading information is already reflected in the price of the stock and, therefore, any analysis to find undervalued securities is useless. It is concluded that past price information is useless for forecasting purposes and it is impossible to make a profit based on previous patterns. Although the theory is supported by some market statistics, there are also many items of evidence to disprove its universal validity in financial markets[30-37]. Prof.Lo[30-31] from MIT argues that market trends and chart patterns exist in many stock and futures instruments, which can’t be explained by the EMH theory.

The goal of our group’s research[38-44] at Northeastern University (sponsored by Cambridge Market Analysis Corporation (CMAC)), continued and brought to some degree of conclusion in this dissertation, is to find systematic methods to recognize patterns, especially cyclical ones, and create a trading system proven to reflect a correct predictive understanding of (cyclic) market movements by using profitability as a criterion.

1.3 Introduction to Chemometrics

Chemometrics is an interdisciplinary science that combines mathematics and chemistry[45]. The International Chemometrics Society (ICS) gives the following definition: “Chemometrics is the science of relating measurements made on a chemical system or process to the state of the system via application of mathematical or statistical methods. “

The history of Chemometrics dates back to 1969[46-48] when researchers attempted to apply a linear learning machine to classify low-resolution mass spectrometric data. The initial research focused on the qualitative or semi-qualitative relationship among data. Later, with the development and application of Chemometrics, increasingly advanced mathematical methods have been utilized to explore quantitative connection in multivariate data.

Chemometrics applies mathematical and statistical methods to common chemical measurements, and can extract information from observed chemical data in a faster, cheaper, better way than previous, informal methods. In recent years, with the advance of computer technology, it achieves even greater success in multivariate-based research areas such as spectroscopic measurements and data calibration. New chemometrics-related methods and applications are being developed continuously from industrial applications to scientific research.

Pattern recognition is an important tool in the chemometrics arsenal[45,49-58]. It is the research area that seeks out similarities and regularities in the vast amount of observation data, such as chromatographic or mass spectral data. Pattern recognition methods play an important role in the process of data evaluation. Generally, there are two groups of pattern recognition methods: supervised methods and unsupervised methods. Supervised methods, including K-Nearest Neighbor (KNN), Bayesian Classification, SIMCA and Linear Learning Machine, refer to techniques that classify patterns based on a prior knowledge of properties. Conversely, without the indication of properties, unsupervised methods try to group or separate data by clustering or projecting them from high-dimensional space to a low-dimensional space. Principal Component Analysis (PCA), Factor Analysis (FA), and Hierarchical Cluster Analysis (HCA) are typical examples of unsupervised methods.

The modern-day chemometrics not only considers mathematical and statistical approaches, but also introduces artificial intelligence for the inductive reasoning process[59-60]. Artificial Neural Network (ANN), Fuzzy theory and Genetic Algorithms (GA) are being used widely in the data handling to extract information from massive data.

In the past two decades, there has been a significant increase in the number of areas of interdisciplinary research, with scientists from different disciplines conducting collaborative research, combining new visions into complex systems, which was hoped to generate innovative solutions to tackle previous intractable problems[61-63]. The methods, developed and employed in Chemometrics, have been used in many interdisciplinary studies, including atmospheric, biological, clinical, environmental, forensic, geological and pharmaceutical, etc[64-67].

Since chemometrics is proven successful in handling many types of data, we try to extend the techniques from chemometrics to futures market data analysis. It is clear that chemometrics methods can also help define and extract information from large chunks of market data, and put theory and model to the test in the investigation and forecast of market movements. The KNN method and Bayesian Classification, belonging to supervised pattern recognition methods, will be utilized in the futures cycle analysis in this thesis. In addition, an artificial neural network will also be used to incorporate multiple source of information to facilitate the cycle analysis. As will be shown, chemometrics methods are well suitable to put each current period in a historical context and to better reveal their effective components, and thus facilitate the modeling and prediction of market behavior. Meanwhile, the current multidisciplinary study of chemometrical pattern recognition applications in the financial markets could also be some day to shed light on other new, non-traditional approaches to market behavior modeling and prediction, such as chaos theory, by recognizing periods of truly unpredictable, chaotic behavior.

1.4 Cycle Theory and Cycle Analysis

The presence of cycles in futures markets[68] (price cycles) is one of the major theoretical foundations for the research conducted in our research group and in this thesis. The existence of cycles and recognition of cycle-related patterns had already been discussed in theses of other group members[38-44]. A brief introduction to cycle theory and cycle analysis is presented here.

A cycle is the regular occurrence of an event at specified time and with a particular size. Cycles are prevalent in all aspects of life. They range from the very short term, like the life cycle of a June bug, which lives only a few days, to the life cycle of a planet, which takes billions of years. Everything in nature moves in cycles. The planets move in exact and predictable cycles around the sun, with the rotation of the earth producing the cycle of day and night. Many of the functions of the human body also revolve around cycles.

Cycles allow us to accurately predict events in nature: bird migrations, the tides, planetary movements, etc. They tend to move from a low point to a high point, and back to a low point again in fairly regular time spans. Because cycles constitute repetitive phenomena, this leads to their most important aspect; namely they can become predictable.

Each market or financial instrument develops its own profile of cycles that influence price direction and course. Analysis reveals that there exist long dominant cycles influencing the trend, and shorter period cycles, that control the price ups and downs that take place within the major trend. These short-term cycles can be used to pinpoint tops and bottoms, and thus entry and exit points.

Here, we discuss some factors, including the business cycle, the political cycle, and the seasonal cycle, that are generally believed to contribute to the long dominant cycles of the market.

The business cycle refers to the periodic fluctuation of an economy[69]. Every business cycle has an expansion and contraction period, and is reflected in the market movement. Depending on the time scale of the cycle, it can range from the Kitchen inventory cycle (3-5 years) to the Kondratieff wave (50-54 years), which is represented in Figure 1.3. From the plot, it is easy to observe the change of the economy with time, from expansion to contraction, then to expansion again.

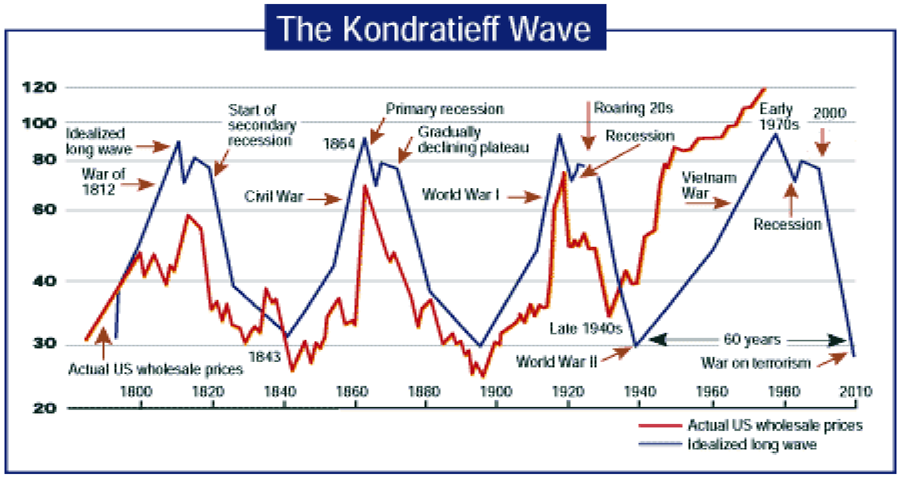


Figure 1.3 The plot of the Kondratieff Wave representing the business cycle with roughly 50-54 years as a whole cycle[70].

The political cycle refers to the effect of the four-year presidential cycle on the stock market, real estate, bonds and commodities[71]. One theory about this cycle suggests that economic sacrifices are generally made during the first two years of a president's mandate. As the election comes closer, administrations tend to do everything they can to stimulate the economy so that voters go to the polls with jobs and a feeling of economic well being.

The seasonal cycle is another factor that affects the market[69]. For example, some commodities, such as wheat and soybean, grow and mature at a certain season. The supply of those commodities will be affected by the inventory and incoming harvest; therefore, it will have some impact on the corresponding prices. The holiday season for the retail stores is another example.

Practical cycle theory assumes that cycles that have consistently persisted in data will continue to occur in the future. Cycles that have performed well in the past should continue this regularity into the future, if they are "genuine", therefore, cycles can be used for forecasting purposes.

Today, cycles remain one of the most mentioned, trading-related topics. This can be ascribed to the cyclical nature "obviously" present in most markets, as well as to the relative simplicity of cycles' structure. The technical analysis based on chemometrics presented in this thesis heavily relies on cycle analysis to get correct market entry or exit signal, but the tools developed here are believed to be much "sharper" than the methods traditionally used. We will discuss the details of cycle analysis in later chapters.

1.5 Overview of FTVision System

1.5.1 FTVision System

To study futures markets with our pattern recognition based methods, we created an in-house trading system named the FTVision system. It was first developed by Dr. Jun Chen[38], and further improved by members of this research group[39-41]. Just like the lab bench for a chemist, the FTVision system serves as a virtual lab bench in our research group to provide great flexibility in the model development.

Instead of using commercial trading systems, we chose to develop our own platform. Commercial trading systems usually provide widely used technical indicators, graphic data representation and an on-line market data connection. Some also provide a sort of flexibility allowing users to specify variables and conduct optimizations for some parameters. However, it limits the users to pre-defined algorithms from the system, and doesn't support user-developed model application. Therefore, to avoid the restriction of limited functionality from commercial systems and facilitate our research, we built our own trading platform and continue to add to it. It is an open system, which can be utilized to test and apply in-house models to historical and real-time market data.

The FTVision system provides lots of functionalities valuable to the development of market trading models. It receives and stores real-time tick data from eSignal, the retail arm of Interactive Data Corporation (IDC). We can operate market data on different time scale, i.e., 1-minute, 10-minute, or 1-hour, etc, which gives us large flexibility in data analysis. The FTVision system is an open environment, in which we can conduct research and test our models on historical and real-time data. Meanwhile, it is also very easy to integrate our system with other tools, e.g. Matlab, to leverage some existing useful algorithms. It is also a trading system, which can automatically generate a trading signal based on the models, and places an order through an outside broker (here, Interactive Brokers, Inc (IB)).

The main graphic user interface of the FTVision system is illustrated in Figure 1.1. It provides a convenient way to represent market data, show market cycles and display market trend line, trading line, etc. There are five major parts in the interface, which are menu, toolbar, price chart, cycle lines and status bar.

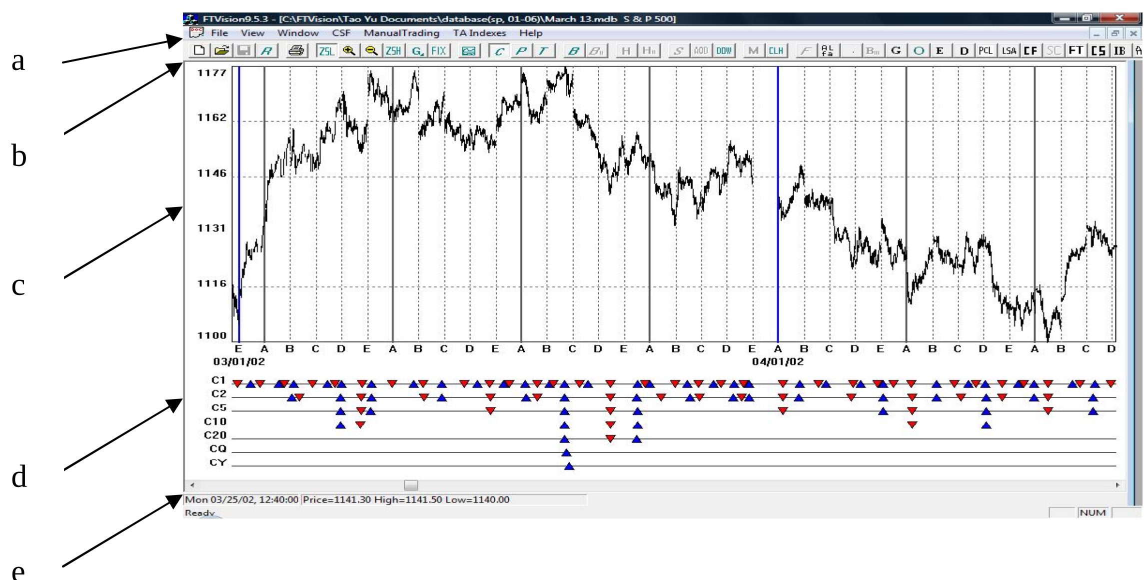


Figure 1.4 FTVision chart sample of S&P futures from 03/10/2002 to 04/17/2002. Part a, b, c, d and e represent menu, toolbar, price chart, cycle lines and status bar, respectively.

Price Chart

Part c in Figure 1.4 represents the price chart, which is the basic and the most important part of the system. Through the study of price movements in the price chart, we can obtain meaningful information, i.e. trends, price patterns, etc, to conduct market forecast for future development.

The price chart in Figure 1.4 shows the example of the S&P futures price changes for about two months. The X and Y coordinates indicates the time and price. In the X coordinate, the letters A, B, C, D and E represent Monday, Tuesday, Wednesday, Thursday and Friday, respectively. Meanwhile, for ease of recognition, the blue vertical line, the solid gray vertical line, and the dotted vertical line in the price chart indicate the start of a month, a week and a day, separately.

In the FTVision system, we store futures market data in our in-house data server, in which the last price of every 1-minute interval is being saved. Since we have our own market database, we can show it in the price chart with different resolution, from 1-minute, 10-minute, 1-hour, and daily to weekly and monthly. In real-time, the price chart will be updated automatically with the incoming tick data.

On the price chart, not only price movements, but other important information, i.e., trend lines, smooth lines, market patterns, etc, can be shown. Meanwhile, the historical trades will also be presented on it with different color lines representing information such as start/end, long/short and profit/loss for better review and analysis.

Cycle Lines

As discussed before, cycle theoretical model is the most important part of research in our group and the cycle recognition and analysis are the foundation in our pattern-recognition based market forecast. Viewed as a sub-structure of market movement, market cycles include peaks and valleys, which represent the local maximum and minimum prices during a certain period.

Corresponding to the price chart, part d in Figure 1.4 shows the cycle lines with different time lengths. On the cycle line, the triangles represent the futures cycles. The downward, solid red triangle means the cycle valley, and the upward, solid blue triangle indicates the cycle top. Currently, there are 7 cycle lines with different lengths in the system: 3 short-term cycles, including C1 (1-day), C2 (2.5-day, half-week) and C5 (1-week); 2 mid-term cycles, including C10 (half-month) and C20 (1-month); 2 long-term cycles, including CQ (1-quarter) and CY (1-year). Usually, in our system, the short-term cycles are used as the trading cycles, whereas the long-term cycles are regarded as control cycles, which are a proxy of the market trend.

According to the definition of the cycles and the observation of above figure, it is easy to find out the relationships existing between cycles with different lengths. In terms of the cycle length, a CY is composed of four CQs, and a CQ includes four C20s. For C20, C10, C5, C2 and C1, each longer-length cycle includes two adjacent shorter-length cycles. In terms of cycle extremes, the top or valley of a longer-length cycle usually is also the top or valley of the short-length cycles. For example, the top of a CY will also be the top of CQ, C20, C10, C5, C2 and C1. The relationship between different cycles will greatly help its recognition and analysis. Along with the price chart, it provides useful information to investigate the market development.

The construction of the cycles, discussed in Dr. Chen's thesis[38], was originally conducted by experienced human being with a set of rules. Later, Dr. Xu[40] created a fully automatic algorithm to generate the cycles, which was reported in his thesis.

Menu

The menu is shown as part a in Figure 1.4. It includes some basic items such as file operations (open, save, print, etc.), view operations (zoom in, zoom out, etc.) and window operations (new, cascade, etc.). It also contains some system-related items such as the TFx methods and ManualTrading (the manual trading setting). TA indexes form another menu item which includes some common technical analysis tools, i.e., moving average (MA), relative strength index (RSI), Bollinger bands, and moving average convergence divergence (MACD).

Toolbar

The toolbar is shown as part b in Figure 1.4. It contains shortcuts for some commonly used commands that have the corresponding item in the menu.

Status bar

Part e in Figure 1.4 is the status bar, which displays some useful information about the price chart. When we put the mouse on the price chart, the left side of the status bar will show the date and time for the specific mouse pointer, and the right side of that will show the price, high price and low price during the time interval, which can be set as 1-minute, 10-minute, 1-hour, etc. When we move the mouse alone the price chart, the content of the status bar will change accordingly.

1.5.2 Important Components of FTVision System

In addition to the main graphic user interface introduced in the above section, some other components have also been developed and are used in my research. Those components were described in detail in Dr. Chen's and Dr. Yao's theses. Here, I only briefly introduce some of them.

Bookie Box

Figure 1.5 shows a sample of the bookie box. This dialog box allows the user to choose the trading parameters, such as the trading entry (auto or manual), the trading mode, which will be discussed later, and others. It also shows the trading results in the middle part of the box, which contain information of every single trade including time, price, action, position and CPL (cumulative profit and loss). The detailed trading results can be saved in files for future retrieve and study. The AutoTrade button is used for automatic trade, whereas Buy and Sell buttons are for manual trade use.

As shown in the mode dropdown list box of the bookie box, there are four modes to be chosen: real-time, historic, simulated real-time and automatic simulated real-time. These four modes represent different utilities for our research and trading purposes.

The historic mode is the very first mode in which to begin the research. In this mode, all price and cycle data stored in our data server can be accessed through the price chart and cycle lines of the main GUI, while there is no price update from the real-time tick data. By applying some statistical tools on historic data, we can obtain useful pre-trade information such as cycle distribution, market pattern, etc. Meanwhile, we also plot the detailed trading results on the price chart and conduct a post-trade analysis. Both are fairly important for us to analyze market rhythm and further improve our market models.

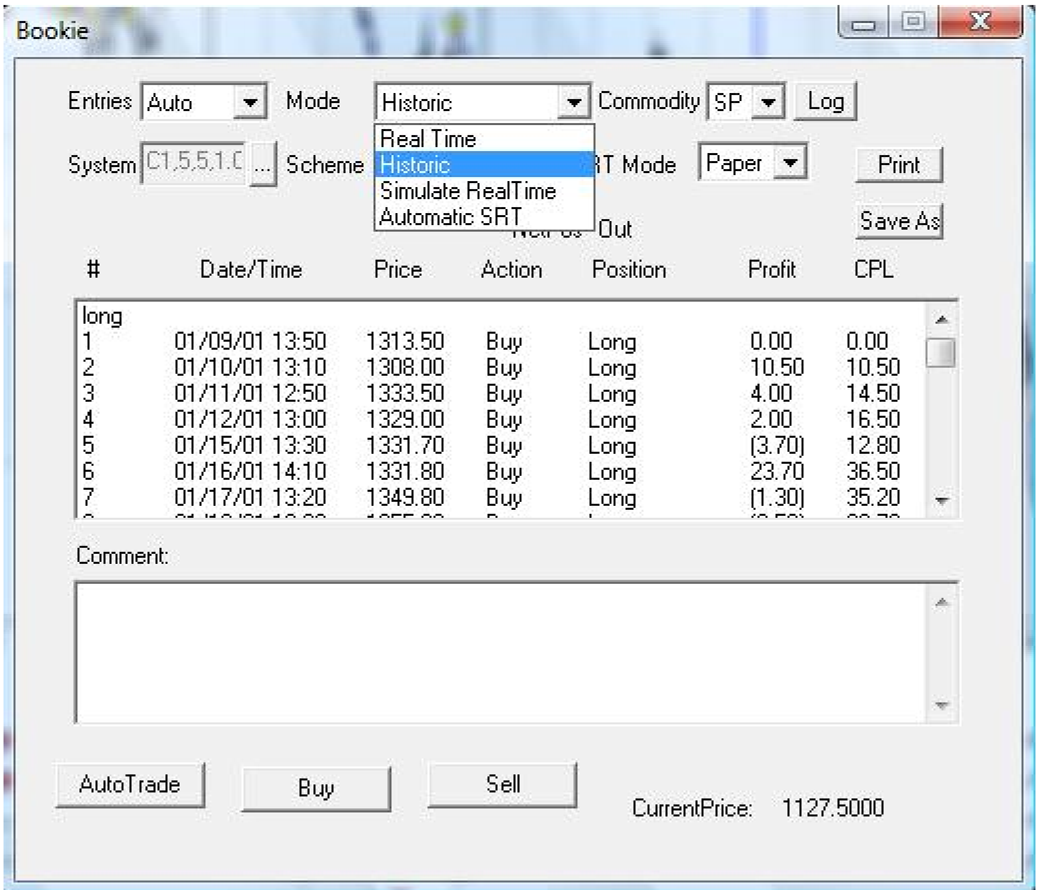


Figure 1.5 Sample of “bookie” historic trading result for C1 trades of S&P futures using the Phi-N method. Only long trades are shown in the list, while short trades are shown in the bottom of the list.

The real-time mode is the real trading mode, in which the real-time tick data are fed into the system, and the price chart are updated accordingly, meantime, all pre-set trading indicators will also be updated based on the new price information. By choosing either paper or real in the RT mode dropdown list box of the bookie box, we can test and apply our trading strategy in real-time. If it is the real trade, we can conduct it either manually via broker call or automatically through our system.

The simulated real-time mode is a very important feature allowing us to develop and test our models on historical data in a real-time fashion. Unlike other data, real-time financial market data are limited and need time to accumulate. Therefore, there exists a dilemma that as the limited source of real-time data, the model developed based on historical data can’t be tested very well in real-time environment, while subsequently, however, the model has to be applied in real-time. The simulated real-time mode provides a valuable way to conduct a stress test and fine-tune our model just like the real-time situation. In this mode, we still use historical data stored in our database, and assume that at a certain time, only the data before, not after it, are known to us. We can move the “certain time” forward to update market data; meantime, we also can choose not to update data by pausing the “certain time”.

Different from the human-driven feature of the simulated real-time mode, the automatic simulated real-time mode has a computer-driven feature. In the system, there are two kinds of update speeds to choose: fast (1 update per 1 second) and slow (1 update per 10 seconds). The automatic simulated real-time mode provides the same benefits as simulated real-time mode, but is more suitable for longer-period examination of the trading strategy.

Summary Box

In the FTVision system, there are two places to show trading results. One is the bookie, where the detailed trade-by-trade information is shown. The other is the summary box, where the graphs of the trading results along with some key numbers are displayed.

As shown in Figure 1.6, a sample of the summary box for S&P C5 trading is presented. In the figure, part a and b show the histograms of long and short trades with time, respectively. The part c brings long trades and short trades together. In those graphs, the blue bar means profits, while the red bar represents losses. The blue curve, red curve and green curve represent cumulative-profit, cumulative-loss and cumulative-profit-and-loss (CPL), respectively. The part d is a set of key numbers for the long and short trades, which include time-in-market (TIM), net-profit, run-down, etc. The detailed explanation of those numbers had been given in Dr. Yao's thesis[39]. Here, I give a simple definition of the numbers used in my thesis.

Time-in-market (TIM) represents the ratio of the time in a long or short position to the total trading period, and is a value between 0 and 1. For the same profit, the smaller the TIM, the better the system.

Figure-of-merit (FOM2) is the ratio of profit to the sum of profit and absolute loss shown in Equation 1.1. It is am important risk measure for the system, which also lies in 0 and 1. The bigger FOM2, the better the system.

$$\text{FOM2} = \text{Profit} / (\text{Profit} + |\text{Loss}|) \tag{1.1}$$

Net-Profit represents the cumulative-profit-and-loss (CPL), and is am important number to measure the profitability of the system.

Run-down measures the maximum cumulative-loss at any time during the trading period. It is another important risk parameter, especially for futures due to margin requirement.

Through the graphs and key numbers of the trades, the summary gives us an intuitive view of the trading performance and an easy way to evaluate relative performances of different trading strategies.

The reader is asked to bear in mind at times that trading performance is a stand-in for level of correctness of the forecasting algorithm used and not a purpose into itself.

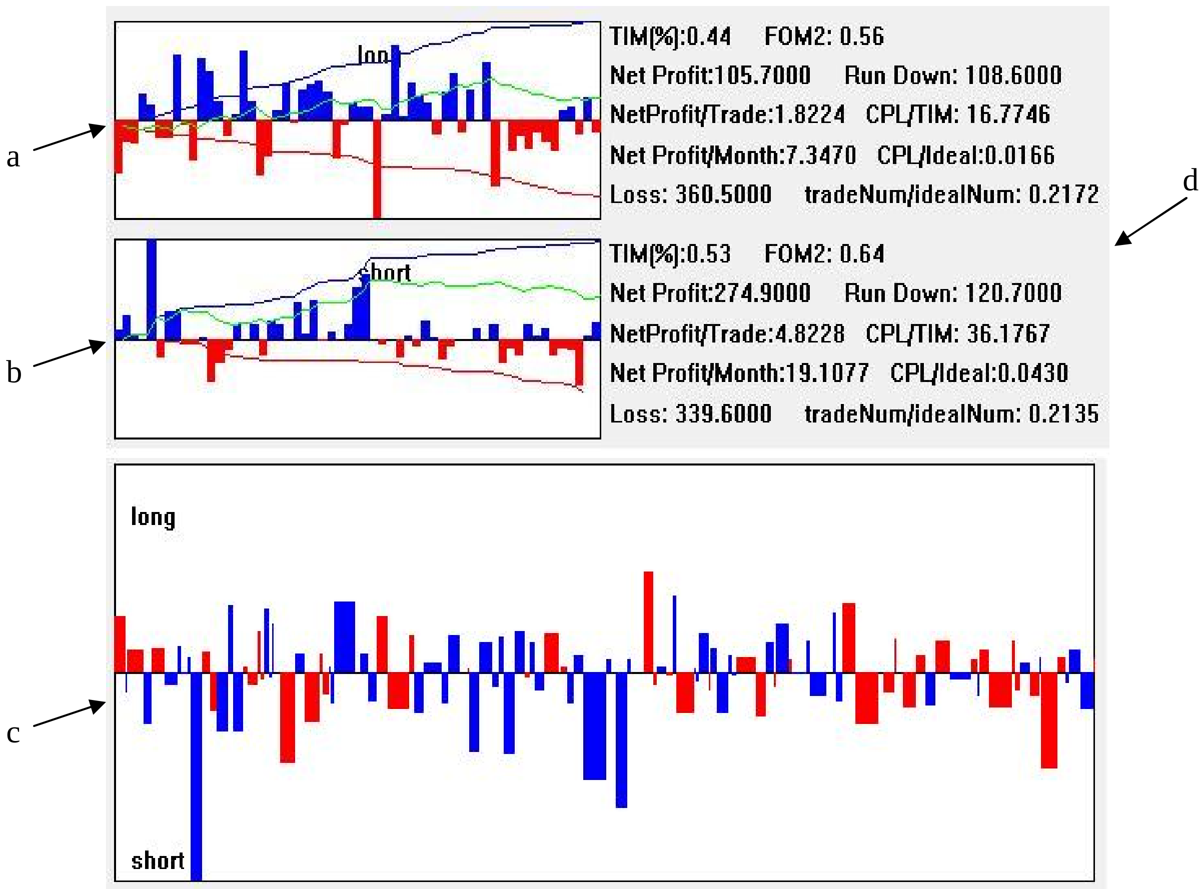


Figure 1.6 Summary of S&P C5 trading results from 02/10/2001 to 03/25/2002 with the TV1 method. Part a, b and c shows the graphs of long, short and combined trades, respectively. Part d shows some key numbers for long and short trades.

Chapter 2 Use of K-Nearest Neighbor (KNN) in Financial Market Analysis

2.1 Overview

Cycles are present everywhere in nature where they range from the regularity of seasonal cycles to the irregularity of monsoon. The existence of market cycles, which are repetition of price fluctuations – a low, a rally and a new low and a new rally, has long been recognized and is also a key assumption in our trading system design.

Based on the fact that market history tends to repeat itself, and, especially that the most recent market cycle is viewed as relevant to cyclic market action in the immediate future, we have used pattern recognition methods to characterize recent relevant market movements and patterns of similar market response to similar stimuli. Because of their perceived greater level of regularity, we mainly focus on the prediction of short-term cycles (half-weekly and weekly), and use long-term cycles (quarterly and yearly) to provide trend guidance. Each cycle has its valley and top, at which, respectively an up-leg and down-leg form. Figure 2.1 is a sketch of market cycles that show the cycle top, valley, up-leg and down-leg.

Dr. Jun Chen[38], in his thesis, discussed the existence and detection of market cycles for silver futures and used cycle statistics to demonstrate the rhythm of cycles. He found that the cycle lengths show a roughly normal distribution around a standard length, and demonstrated that the cycle valley-to-valley distance has more regularity (a smaller standard deviation) than the cycle top-to-top time distance. Following Dr. Chen's initial work, Dr. Jian Yao[39] further explored the presence of cycles in financial futures, extended the Phi-N-Alpha forward-counting method, and introduced the K-Nearest Neighbor (KNN) method, a well-known pattern recognition algorithm to cycle analysis research. In

Dr. Ke Xu's thesis[40], which followed Dr. Yao's, he demonstrated the existence of market cycles operationally by studying the relationship between the cycle model lengths and their trading profitability.

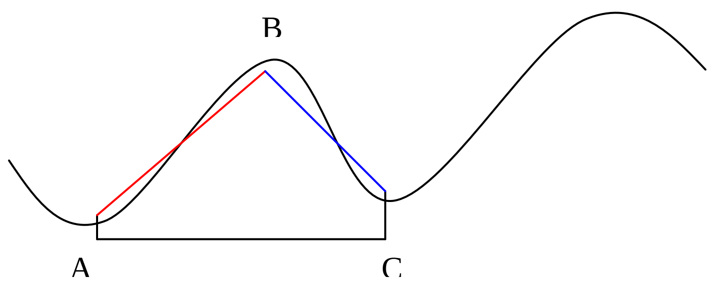


Figure 2.1 A schematic example of market cycles. The time from Point A to point C is a full market cycle. Points A and C are the cycle valleys and point B is the cycle top. The line AB (red line) forms the cycle up-leg, and the line BC (blue line) forms the cycle down-leg.

In this chapter, I will present a further application of the KNN method to the cycle-based market analysis studied in this group previously. With the introduction of two new TFx trading methods, I will show the application of the TFx methods in the S&P futures market for short-term cycles (C2 and C5). TFx is a generic designation of cycle-based futures price forecast. Meanwhile, an optimization of the trading rule in the TFx methods will be provided, and use of control cycles, here the "man-made" long-term cycles (CQ and CY), will be added to the TFx methods. The results of these innovations show that the cycle-based market prediction methods incorporating the KNN concept are valid, and give good predicative results which become even better when the control cycles are added.

Section 2.2 gives an introduction of pattern recognition and the KNN method. Section 2.3 describes the TFx methods, which are a group of cycle-based market prediction methods based on the KNN method. Section 2.4 describes the trading results obtained with the family of TFx methods in historical S&P futures markets. Also an optimization of the trading rule is introduced. Section 2.5 and 2.6 shows the effect of control cycles on the TFx methods. Section 2.7 summarizes the results of this chapter and draws some conclusions.

2.2 The K-Nearest Neighbor (KNN) Method in Pattern Recognition

2.2.1 Pattern Recognition

A pattern of an object is the collection of its characteristic features. Pattern recognition is the research area that studies the operation and design of systems that recognize patterns in data. It first collects and preprocesses original observations, then analyzes and extracts information from those data and, at the end, classifies observations into different pattern groups based on either a priori knowledge or statistical information extracted from the patterns.

Generally speaking, pattern recognition falls into two categories[45,54-55]: unsupervised leaning and supervised learning. Without the prior knowledge on the part of supervisors, unsupervised learning will group observations either by means of cluster methods or by projecting higher- dimensional data onto a lower-dimensional space. Cluster methods include distance and similarity measures or hierachical and non-hierarchical cluster analysis, etc; and projection methods include, e.g., principal component analysis (PCA) and factor analysis (FA). In contrast to unsupervised learning methods, supervised learning methods know in advance the particular clusters of which objects are members; discriminant analysis, k-nearest neighbor analysis (KNN) and “soft independent modeling of class analogies” (SIMCA) belongs to the category of pattern recognition methods.

In the technical analysis of financial market, the object of pattern recognition is to detect the pattern correctly and use it in real time, at the right moment. In my research presented here, I advance further the use of KNN method, and add use of Bayes' Theorem and an artificial neural network, in the search for cycle-based market patterns, and the conduct of futures forecasts for future market movement.

2.2.2 The KNN Method

The KNN method is a supervised learning algorithm where the unknown object is classified based on the known categories of its k nearest neighbors in a multi-dimensional property coordinated space[45]. Although the KNN method is conceptually simple, it is a very powerful classification technique which has been used in many applications such as data mining, image processing and others.

An example of the unknown object classification by use of the KNN method is presented in Figure 2.2. By comparison of the distances between the unknown object and the training samples, whose closeness are known in advance, k nearest neighbors with shortest distances are selected, and the classification type represented by a simple majority of the k nearest neighbors will be assigned to the unknown object. In this example, k is chosen as 7, and the unknown object is classified as a red triangle because there are more red triangles than blue squares among the 7 closest in the neighbors.

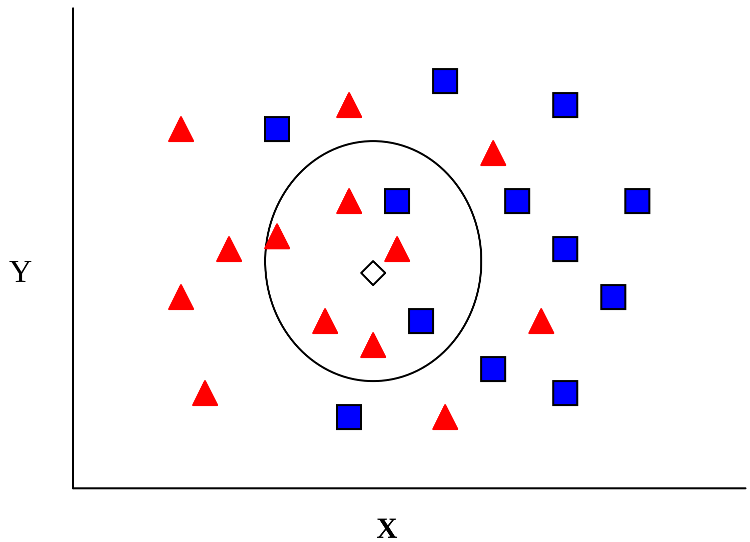


Figure 2.2 The classification of an unknown object (open diamond) in a 2-D parameters space using the KNN method. When k is chosen as 7, the unknown object is classified as red triangle because there are 5 red triangles versus 2 blue squares. The majority of nearest neighbors are red triangles.

I give here the detailed step procedure on how to use the KNN algorithm in the classification:

1. Collect data including the unknown object and the training samples.
2. Select parameter k, which is the number of nearest neighbors.
3. Calculate the distances between the unknown object and the training samples. The Euclidean distance is the popular algorithm to use for this purpose. Suppose the unknown object and the training sample have the coordinate (X_u, Y_u) and (X_t, Y_t) respectively, then the squared Euclidean distance can be calculated as follows: $D^2 = (X_u - X_t)^2 + (Y_u - Y_t)^2$.
4. Sort the distances and choose the k neighbors with the shortest distances from the object.
5. Gather the types of the k nearest neighbors.

6. Assign the type of simple majority of the k nearest neighbors to the unknown object.

The advantage of the KNN method is that it is robust to noisy training data and effective if there is a large enough training sample set. Also, it allows the inclusion of experiences from a long time base into the decision process. A disadvantage is that one needs to determine the optimal value of parameter k (number of nearest neighbors). The accuracy of the classification is usually affected by k, which needs to be optimized for good classification. As discussed below, this task had been attacked in the recent Ph.D. thesis of Dr. Zhao[41], where an optimal value of k as 10 was demonstrated for a specific, related case). As the pattern of an object is determined by its characteristic features, the selection of features used in a two-parameter combination will dramatically affect the classification result. This selection and results will be discussed in the following sections.

2.3 TFx Trading Methods

One of the foundations of technical analysis is the simple observation that history will repeat itself in similar circumstance. This implies that the market has a memory of past behaviors that affect its future movement. The recognition and application of this pattern of repeated similar response to similar challenge through which past market actions have relevance to future price actions is the main focus of our research.

In this thesis, a series of pattern recognition methods historically collectively named TFx were created to seek out market pattern. We focus on two specific aspects: recent market movements and prior patterns of market response to similar stimuli. The first aspect is represented by cycles with different time scales. Based on the notion that market dynamics represent the collective behavior of market participants, and can be considered as a natural phenomenon (analogous to the subjects of the physical sciences), with a statistical internal structure amenable to analysis, the second aspect can provide a guide to future price action.

In the TFx methods, the unknown object, which is awaiting the market at "now" point, is the extreme of the current cycle, and the training samples are historical cycle points. By placing the current cycle into the context of its historical analogs, we can use the KNN algorithm to identify those analogs that are most "like" the current cycle in terms of relevant, short-term contexts and other aspects of market history. Then, the features of those analogs can be used to detect the current cycle price movement. The main difference among the TFx methods is the selection of cycles' specific characteristic features (the markers).

Definitions and description of the use of part of TFx methods (TF1, TF2, TF3, TF4 and TF5) were presented in Dr. Yao's, Dr. Xu's and Dr. Zhao's theses[39-41]. The basic definitions of the five methods are given in Appendix A. In the following sections, I introduce two other TFx methods, named here TV1 and TF2B, with different markers as the construction elements, and conduct a comparison of their effectiveness to that of the above TFx methods.

2.3.1 The TV1 Method

Like the TFx methods, the TV1 method uses the KNN algorithm, and determines the status of current cycle by reference to historically similar cycle scenarios. Figure 2.3 is an idealized sketch showing the specific market situation to be analyzed and the markers used in the TV1 method. Starting from a "confirmed" cycle valley C, we are looking for current cycle top. The present point in time, the "now" point, is marked as A. Recent market action in the cycle is idealized as an "up-leg" ED from a cycle valley E to a top D, and a subsequent “down-leg” DC. It is assumed that points E, D and C have been confirmed by subsequent market movements. Between the “confirmed” cycle valley C and the “now” point A, point B is the temporary cycle top based on the price. Thus, CB becomes a tentative “up-leg”.

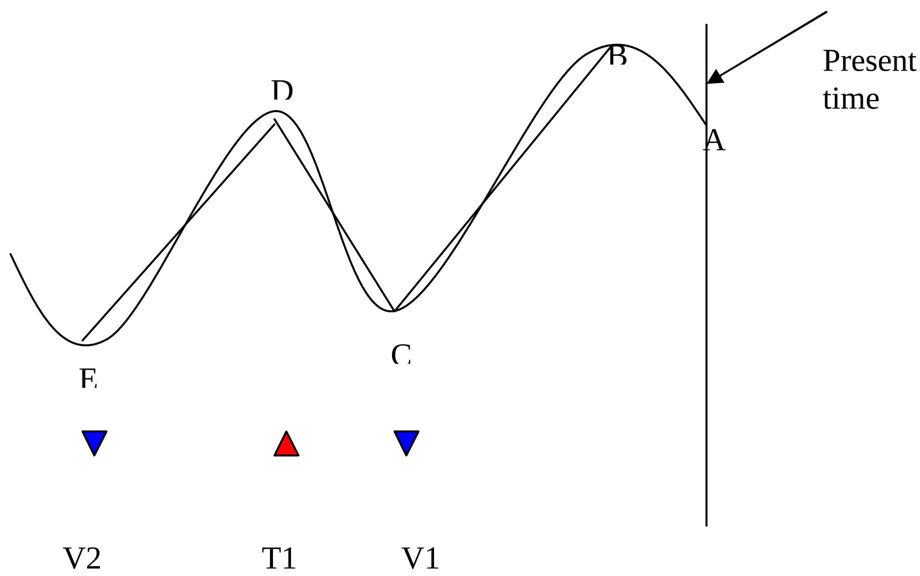


Figure 2.3 Idealized sketch of market situation and markers used in the TV1 method. Point A is “Now” point, representing the present time. Point B is the temporary top of current cycle. Points C, D and E are assumed to be “confirmed” cycle extremes.

The ratio of price change in the tentative up-leg CB, down-leg DC and up-leg ED are used as two dimensionless ratios X and Y as follows:

$$X = (P_B - P_C) / (P_D - P_C) \tag{2.1}$$

$$Y = (P_D - P_C) / (P_D - P_E) \tag{2.2}$$

Where

P: market price at the specific times B, C, D and E, respectively.

These ratios are then used as coordinates of an X-Y plot (Figure 2.4). This plot contains analogous ratios for analogous cycles from historical database. We are now seeking guidance for the finding of the most recent cycle top from this database for an optimal exit from a long trade which is in progress since point C. The meaning of the points in Figure 2.4 is as follows: cycles which at this time point of their history (time difference between points A and C) have already passed their cycle top are marked as red; those cycles for which this event still lies in the future are marked as blue. It can be seen that red and blue points are clustered in specific, different regions of the plot, indicating that similar histories predict similar futures.

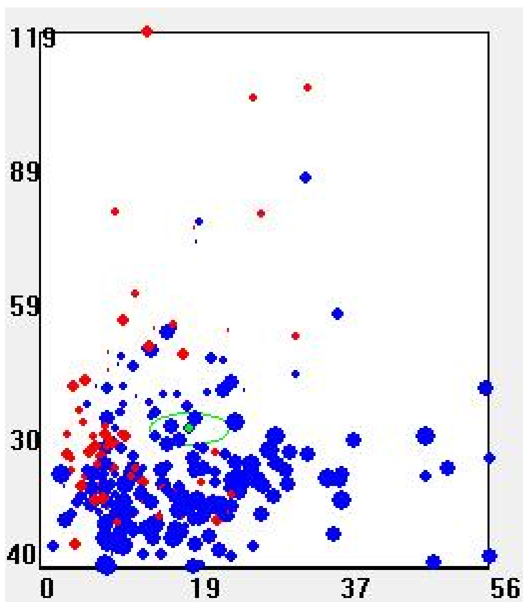


Figure 2.4 X-Y plot for the TV1 method. Coordinates for each point represent relative price amplitudes at given time for a historical cycle; a red point in the graph means that the cycle top has passed, and a blue point indicates the cycle top hasn't passed yet; the green point represents the present case. The green ellipse includes the k (10 in this case) nearest neighbors.

The "present-case" point is shown in the plot by the green point; the surrounding circle (shown here, scaled, as a green ellipse) is drawn to include k neighbors, using

Euclidean distances in this two-dimensional coordinate system. To “predict” the future of the present-case point based on its k analogs, the market histories of the k analogs (here k = 10) are then retrieved and plotted against time (Figure 2.5) and the information is recorded and displayed when each of these analogs reached their next top. Those k nearest neighbors can be considered as cohorts because of their similar price trajectories.

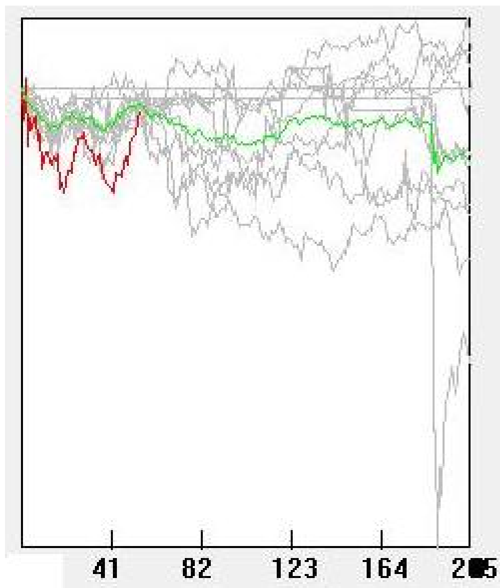


Figure 2.5 Price trajectories of present cycle and cohort cycle members. The red line is the price movement of a real-time C5 up-leg. The gray lines are the price trajectories of a cohort of 10 C5 historical cycles, selected by TV1 method. The green line is the average price movement of all gray lines and suggests the future price movement of the red line beyond present time.

The determination of cycle top can be based on the percentage of those analogous cycles passing their next cycle top. Figure 2.6 shows a trade snippet using the passing rate 60%, which means at least 6 out of 10 analogous historical cycles already passed their next cycle extremes, as the trading rule. When the next extreme of the current cycle is determined, it will close the previous trade and immediately initiate a reverse trade.

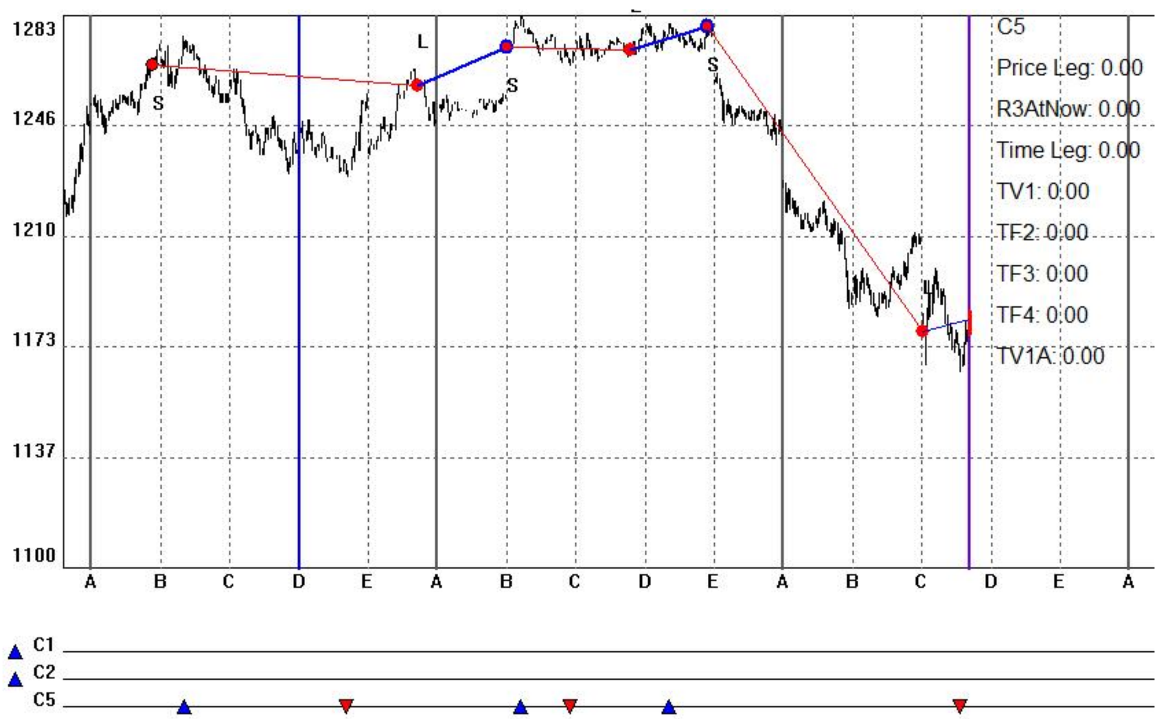


Figure 2.6 Trade snippet of the TV1 method trading results with 60% as the passing rate. The trade will close and then reverse when the majority of cohort cycle member passing next cycle point is larger than 60%.

2.3.2 The TF2B Method

In the same way as other members of the TFx family, the TF2B method also uses information on preceding cycles, including time and price, as the markers. However, it uses three pieces of information from those preceding cycles, and hence requires the use of a three-dimensional Euclidean distance to choose the nearest neighbors rather than two-dimensional distance as the other TFx method did. Figure 2.7 shows an idealized sketch with the markers used in the TF2B method.

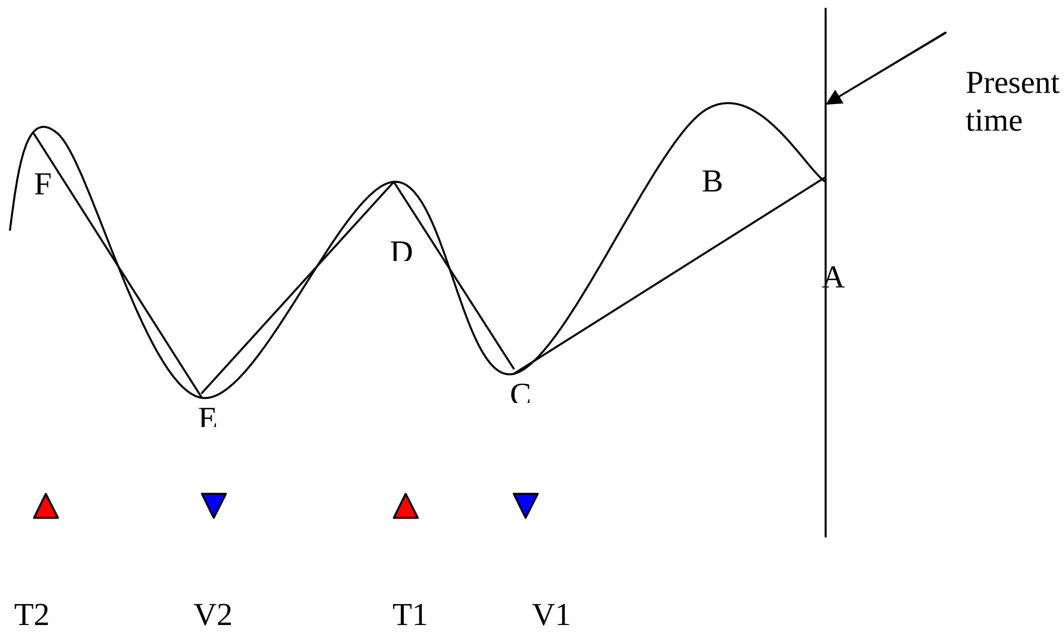


Figure 2.7 Idealized sketch of market situation and markers used in the TF2B method. Point A represents the present time. Point B is the temporary top of current cycle we are trying to confirm. Points C, D, E and F are “confirmed” prior cycle extremes.

The TF2B method is similar to the TF2 method in that both employ the same X and Y coordinates, but the TF2B method reaches back further in time in that it also uses the price ratio of the earlier down-leg FE and the up-leg ED as a third coordinate Z. The three dimensionless ratios are as follows:

$$X = (P_A - P_C) / (P_D - P_C) \qquad 2.3$$

$$Y = (P_D - P_C) / (P_D - P_E) \qquad 2.4$$

$$Z = (P_D - P_E) / (P_F - P_E) \qquad 2.5$$

A three-dimensional X-Y-Z plot, which is not shown here, is then constructed based on the three dimensionless ratios X, Y and Z. Historical cycle points will be put into the plot as the candidates and three-dimensional Euclidean distances will be calculated to find out k cohort members. The classification and decision processes are same as for the TV1 method described above.

2.4 Optimization of the Passing Rate for the TFx Methods

The TFx methods are a series of pattern recognition models for market forecast based on the KNN algorithm and cycle theory. In section 2.3.1, we introduced a rule for automatic trading, the "voting" method, in which the determination of next cycle extreme is based on the passing rate, the percentage of nearest neighbors (analogous cycles) that have passed their next cycle extremes. When the passing rate exceeds a set threshold, a trading signal is generated to close the current position and initiate a new reverse position. Meanwhile, a new cycle will form and automatically adjust based on the time and price constrains. (Details of the automation of cycle recognition and confirmation were described in Dr. Xu's thesis[40]). This algorithm is used here, and because of its fundamental significance for the TFx methods, the detail is given in Appendix B. Incorporating the new cycle information, a new X-Y plot will be constructed and used for the next cycle forecast.

In such a "voting" method, there are two crucial elements affecting the effectiveness of the TFx method. One is the cohort size k, which decides how many historical cycles should be considered as a cohort. Too large a k value will include many dissimilar historical cycles, while too small a k value will give poor statistics. In Dr. Zhao's thesis[41], she refined the k value and found that 10 is the best cohort size. Therefore, in this thesis, I will also use 10 as the operational cohort size to choose cohorts. The other critical element is the passing rate (voting ratio), which will determine the category of current cycle, and generate the corresponding trading signal.

In my research about the TFx methods, I conduct an optimization of the passing rate to investigate its effect. S&P 500 Futures are chosen as the subject of my thesis research. The half-weekly (C2) and weekly (C5) cycles are chosen as the testing cycles because they are short-term cycles so that there is large enough background pool of candidates for selection as cohort members. The testing and background periods are the same, both running from 01/2001 to 04/2006. It should be mentioned that the optimization was conducted in the "simulated real time" (SRT) mode, in which we can only access market information before the historical "now" time points, and don't use knowledge of the future market after that, except to evaluate the efficacy of the parameters being tested. The whole database (five and half years) is used to provide historical cycle background. Figure 2.8 shows the price chart of S&P 500 futures during the period. According to the chart, S&P futures experienced a strong downward trend in the first half of the period, and a strong upward trend in the second half, so there is no bias from a unidirectional market. The FTVision, an automatic trading platform described in Chapter 1, is being used to conduct the study.

Since my study of futures markets was not intended to provide practical investment tools without further improvement, the trading results presented in the following sections and chapters are theoretical results, which don't consider any transaction cost. In addition, contrary to convention in financial industry, we use S&P futures points rather than dollars in the analysis. As we know, one point in S&P futures is equal to $50 per contract so that the total amount of money in dollars is 50 times the number of S&P futures points.

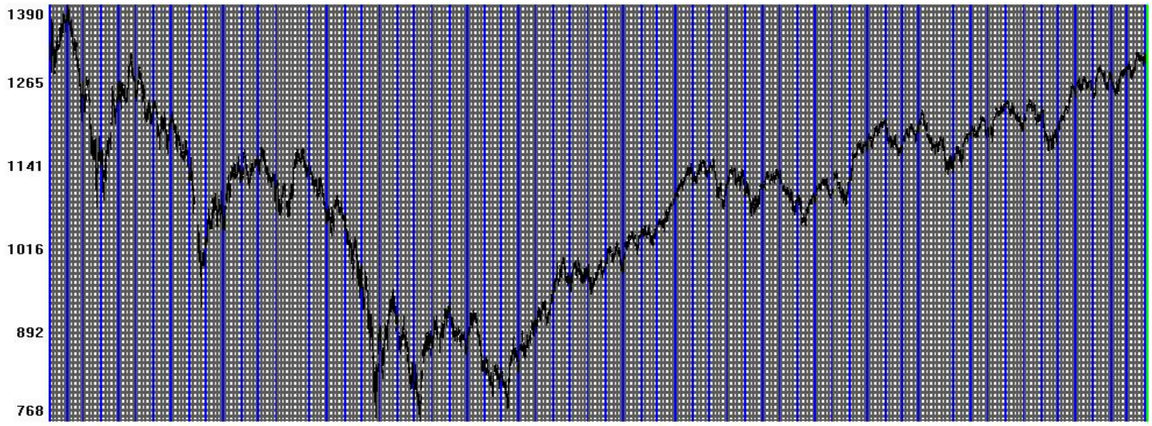


Figure 2.8 S&P 500 futures price chart from 01/2001 to 04/2006.

2.4.1 C2 Cycle Optimization Results

As discussed above, in the “voting” method, the trade signal is generated based on the present passing rate of cohort members. For instance, if the current passing rate (the percentage of k voters needed to make a decision) is 60% (6 out of 10 members), when 60% of the cohorts have passed their next cycle points, either a buy or sell signal will be generated. In this study, we conduct an optimization of the passing rate to investigate its effect on performance of the specific TFx method studied, and to examine their performances in market forecasts. The half-weekly cycle (C2) is used as the testing cycle, and all historical, “man-made” C2 cycles during the test period from 01/2001 to 04/2006 are used as background.

Table 2.1.1, 2.1.2 and 2.1.3 contain trading results of simulated real time C2 cycle testing of S&P 500 futures using the TFx methods. Among them, Table 2.1.1 and 2.2.2 shows the trading results for the passing rates between 30% and 80%, and Table 2.1.3 provides the best passing rate for each TFx method. In the tables, the “Long” represents the profit-loss for all long-side trades, and the “Short” for all short-side trades. The

“Total”, also called CPL (cumulative-profit-and-loss) in our term, shows the sum of the profit-loss for all long and short trades, which is used as the performance measure.

Overall, according to the trading results in Table 2.1.1 and 2.1.2, all TFx methods are profitable during the testing period. Considering that the long test period of over 5 years contained market periods of diverse levels of directions of “trend”, this suggests the general validity of the TFx methods in market forecasts also at other time.

Based on the results in Table 2.1.3, the best passing rate is different for each TFx method. The reason is that every TFx method defines its own pattern so that they may have different best rates in the application of pattern recognition. For the evaluation period as a whole, the TF1 and TV1 methods have the relatively best trading results, among the TFx methods, the TF4, TF5 and TF2B methods show the least good results. A further discussion of their results in the context of the cycle studied follows in section 2.4.3.

Table 2.1.1 Results of simulated real time C2 cycle testing of S&P 500 futures using the TFx methods (passing rates from 30% to 50%) with cycle database during 01/2001 and 04/2006

Method	30%			40%			50%		
	Long	Short	Total	Long	Short	Total	Long	Short	Total
TF1	39.9	89.2	129.1	79.4	116.3	195.7	197.9	248	445.9
TF2	-87.4	-60.2	-147.6	142	189.6	331.6	377.6	427.3	804.9
TF3	-102.8	-28.5	-131.3	370.6	423	793.6	314.4	356.4	670.8
TF4	176.3	260.5	436.8	163.5	240.5	404	-116.3	-47.9	-164.2
TF5	21.7	98.9	120.6	82.2	149.4	231.6	189.7	265.7	455.4
TV1	408.7	498.2	906.9	547.2	618.9	1166.1	402.5	469.6	872.1
TF2B	-27.1	10.2	-16.9	30.8	89.6	120.4	22.8	83.7	106.5

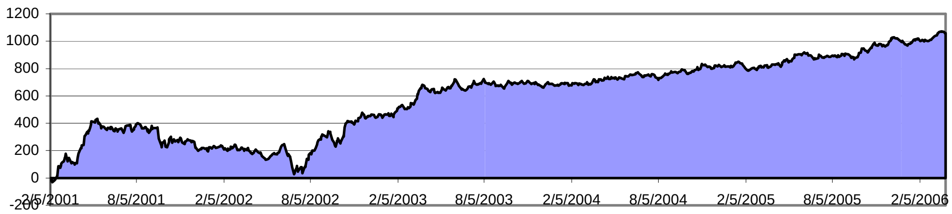
Table 2.1.2 Results of simulated real time C2 cycle testing of S&P 500 futures using the TFX methods (passing rates from 60% to 80%) with cycle database during 01/2001 and 04/2006

Method	60%			70%			80%		
	Long	Short	Total	Long	Short	Total	Long	Short	Total
TF1	511	546.2	1057.2	509.3	550.4	1059.7	264.7	305	569.7
TF2	160.9	209.6	370.5	88.2	133.2	221.4	19.8	55	74.8
TF3	196.9	231.6	428.5	268	295.3	563.3	111.7	185.6	297.3
TF4	25.8	98.5	124.3	47.7	93.6	141.3	14.3	51.4	65.7
TF5	28.3	62.5	90.8	20.1	50.8	70.9	1.8	8.5	10.3
TV1	285	351	636	14.1	65	79.1	111.8	148.4	260.2
TF2B	203.5	244.8	448.3	-14.7	41	26.3	189	249.6	438.6

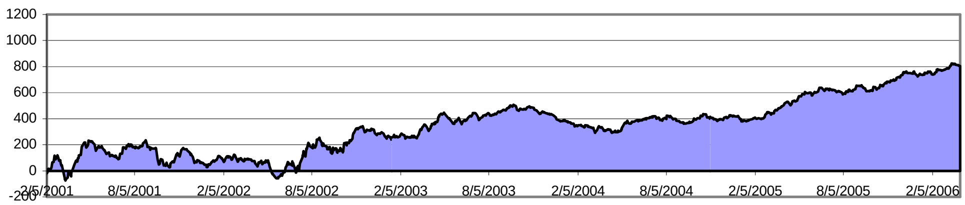
Table 2.1.3 Results of best passing rate for simulated real time C2 cycle testing of S&P 500 futures using the TFx methods with C2 cycle database during 01/2001 and 04/2006

Method	Best Passing Rate (%)	Long	Short	Total
TF1	70	509.3	550.4	1059.7
TF2	50	377.6	427.3	804.9
TF3	40	370.6	423	793.6
TF4	30	176.3	260.5	436.8
TF5	50	189.7	265.7	455.4
TV1	40	547.2	618.9	1166.1
TF2B	60	203.5	244.8	448.3

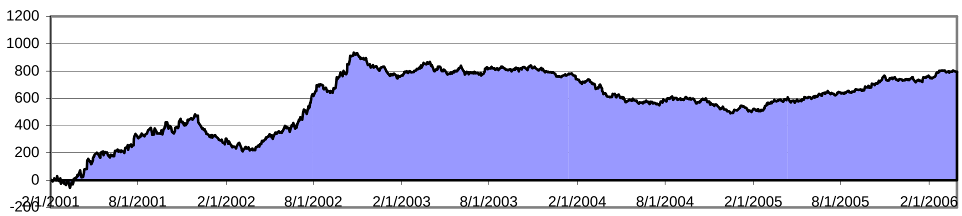
TF1



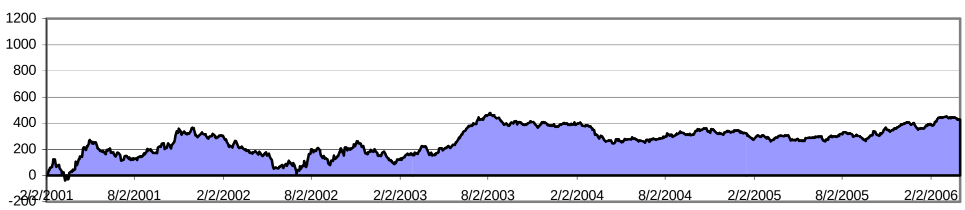
TF2



TF3



TF4



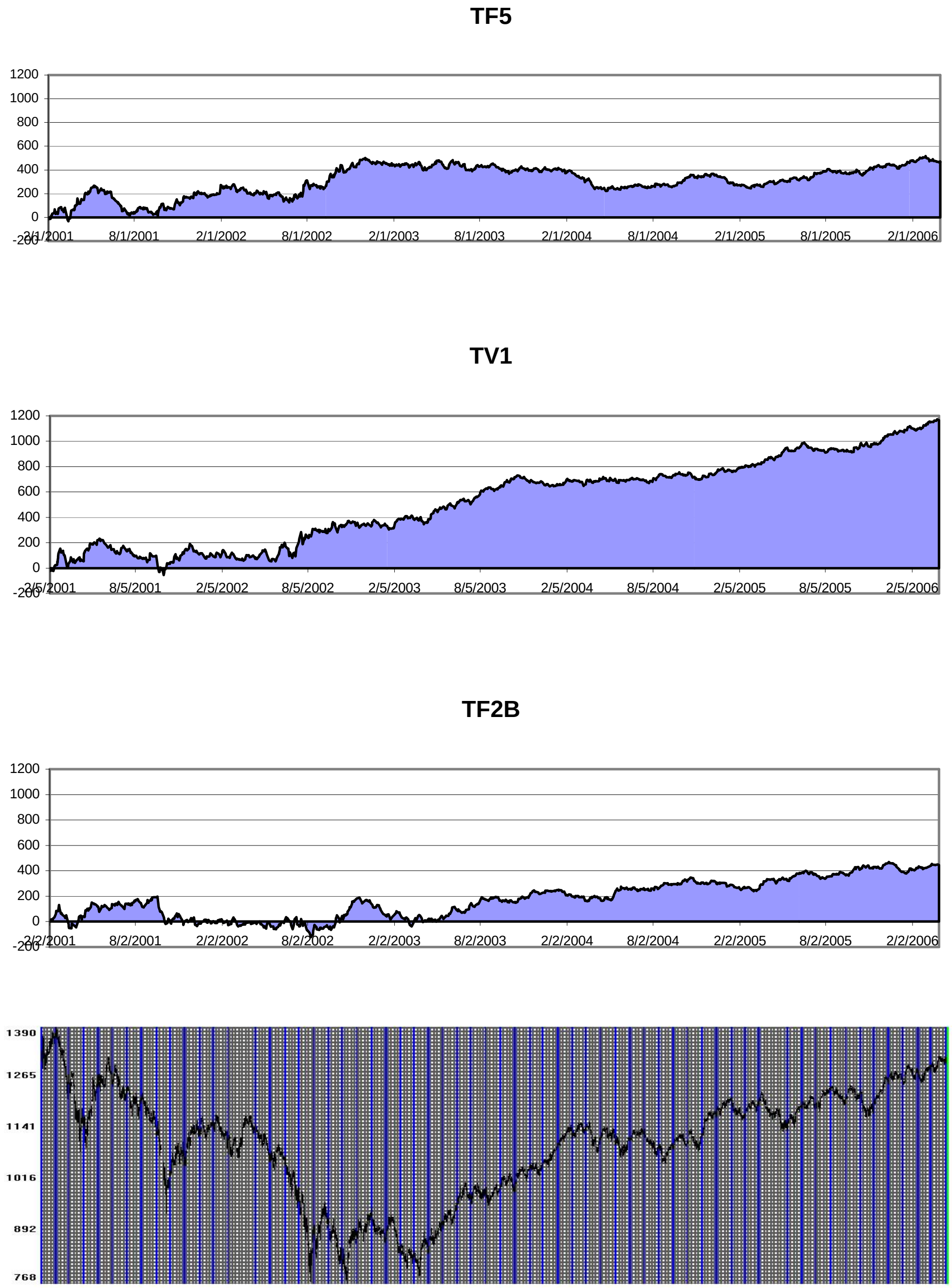


Figure 2.9 Histograms of TFx trading results with best passing rates from 01/2001 to 04/2006 and S&P price chart at the same period. The graphs from top to bottom represent TF1, TF2, TF3, TF4, TF5, TV1 and TF2B, respectively.

Figure 2.9 presents a visualization of the trading results assuming use of the best passing rates for all TFx methods; this figure also shows the corresponding S&P market. It shows histograms of cumulative profits and illustrates the relative independence of the profitability results on the rising or falling market. Figure 2.10 displays the “summary” plot of the trading result for the TF1 method with the optimized passing rate; it is shown as a representative of all the TFx methods. While profitability over the total 5-year period is excellent, year-by-year performance varies greatly. The most uniform performance is seen for TV1, introduced first in this thesis. According to the histograms, the TF1, TF2 and TV1 methods have the best overall performances during the 5-year period treated. They work better in the up-trending market, but so no perform well in the down-trending market, while the TF3 and TF5 methods perform relatively better in the down-trending market than the up-trending market. The new TF2B method does not perform well during the period. A further discussion of the relative performances is given in section 2.4.3.

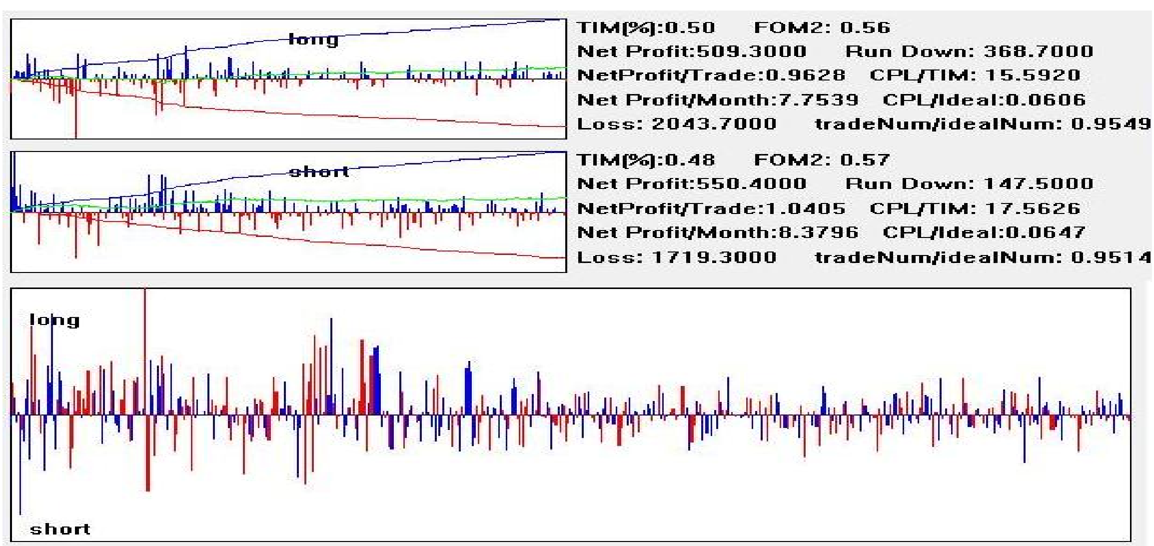


Figure 2.10 C2 cycle summary plot of TF1 trading results with a passing rate of 70% for S&P futures during 01/2001 and 04/2006.

As shown above, every TFX method displays different market performance during the 5-year test period. As a further updating test, I applied the TFx methods with the optimal passing rates to a more recent market period, from 06/2007 to 05/2008. The summary plot for the TF1 method, again taken as a representative, is shown in Figure 2.11, and the price chart for this period is also presented. The first half-year is a period of large market oscillations, then, after a big drop, the market became oscillating again. Table 2.2 contains the simulated real-time trading results for all TFx methods for this more recent period. All of them except TF3 and TF5 make a profit, and, in accordance with the general market trend in this period, short trades perform better than long trades. Among the various methods, the TV1 method performs best with a CPL over 500, and takes the lead in every rating category: FOM2, Net-profit/trade and lowest Run-down. The performance of the best TFx methods in the latest period is thus quite consistent with those in the earlier, longer test period. This may be a good place to put the financial significance of those results into context: The TV1 CPL for this recent one-year period of 501 S&P points (@$50/point) corresponds to a one-year trading profit of $25,000 (less commission of ~10%), which requires a margin of $5000 (+ back-up-holdings).

Table 2.2 Results of simulated real time C2 cycle testing of S&P 500 futures using the TFX methods with best passing rate during 06/2007 and 05/2008

Method	CPL			TIM		FOM2		NetProfit/Trade		Run Down	
	Long	Short	Total	Long	Short	Long	Short	Long	Short	Long	Short
TF1	64	245.75	309.75	0.59	0.39	0.52	0.62	0.6598	2.5077	148	117.25
TF2	15.5	204.25	219.75	0.55	0.43	0.51	0.59	0.1192	1.5592	149.5	90.75
TF3	-183.8	-6.5	-190.25	0.47	0.51	0.45	0.5	-1.0265	-0.0363	276.5	140.5
TF4	-10	166	156	0.43	0.54	0.5	0.55	-0.0441	0.7313	267.5	156.5
TF5	-180.3	46.25	-134	0.52	0.45	0.44	0.52	-1.1859	0.3023	284.5	157.75
TV1	151.5	350.5	502	0.5	0.47	0.55	0.63	1.1565	2.6756	105.5	79.75
TF2B	88.25	274.75	363	0.54	0.44	0.53	0.6	0.8023	2.4752	171	130

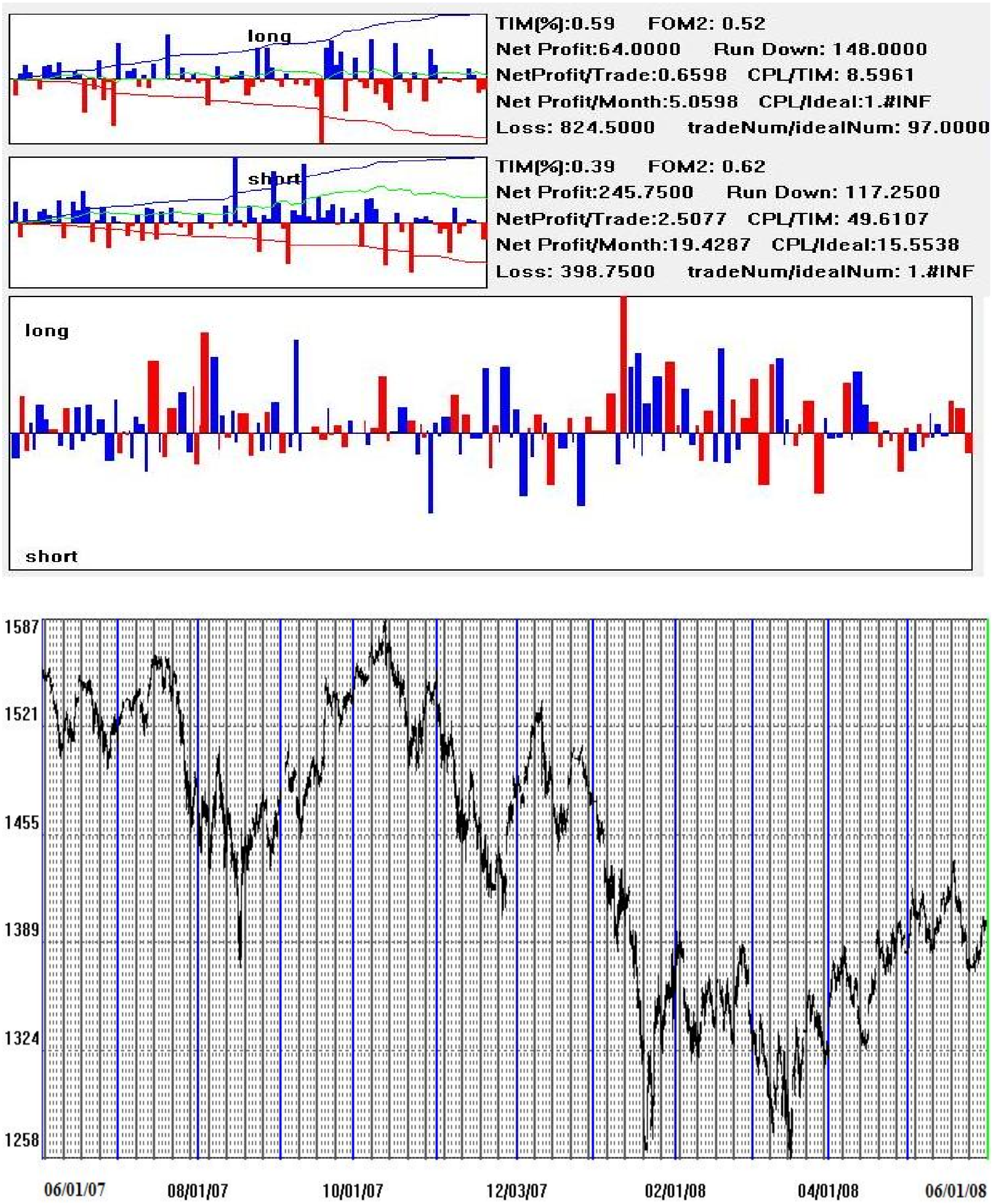


Figure 2.11 The top graph is the summary plot of TF1 trading with 70% passing rate from 06/2007 to 05/2008. The bottom graph is the price chart of the period from 06/2007 to 05/2008.

2.4.2 C5 Cycle Optimization Results

In addition to C2, the weekly cycle (C5), for which we also have a large cycle point background library and which is relatively well expressed in the cycle data presented in previous theses data for S&P, was chosen as a research cycle.

Table 2.3.1 and 2.3.2 contain simulated real time C5 cycle trading results for S&P futures using the TFx methods with cycle database during 01/2001 and 04/2006. Table 2.3.3 summarizes the best passing rates for every TFx method in terms of CPL.

Like C2 trading results, all TFx methods are profitable during the testing period, but the best passing rates are different from those for C2. According to Table 2.3.3, the TF2 and TV1 methods have the relatively best trading results, while the TF4 method has the relatively least good result, as also seen for C2 above. (Such differences are viewed as due to an indicator being coincident or trailing as discussed, below)

Table 2.3.1 Results of simulated real time C5 cycle testing of S&P 500 futures using the TFX methods (passing rates from 30% to 50%) with cycle database during 01/2001 and 04/2006

Method	30%			40%			50%		
	Long	Short	Total	Long	Short	Total	Long	Short	Total
TF1	237.1	245.2	482.3	-81.2	-58.2	-139.4	258.5	278.2	536.7
TF2	325.4	376.5	701.9	102.3	141.7	244	286.4	323.2	609.6
TF3	249.6	291.5	541.1	-229.8	-141.3	-371.1	153.9	209.9	363.8
TF4	108	175.2	283.2	105.8	101.3	207.1	133.1	221.8	354.9
TF5	238.2	271.7	509.9	29.3	35.1	64.4	205.3	239.4	444.7
TV1	368.9	401	769.9	189.9	227.8	417.7	400.7	480.9	881.6
TF2B	284.6	343.8	628.4	361.4	364.5	725.9	359	407	766

Table 2.3.2 Results of simulated real time C5 cycle testing of S&P 500 futures using the TFX methods (passing rates from 60% to 80%) with cycle database during 01/2001 and 04/2006

Method	60%			70%			80%		
	Long	Short	Total	Long	Short	Total	Long	Short	Total
TF1	147.9	184.9	332.8	354.4	420	774.4	-17.1	5.3	-11.8
TF2	442.7	501.9	944.6	22.9	120.3	143.2	3.3	95.9	99.2
TF3	130.8	154.1	284.9	59.4	121.7	181.1	-55.9	-10.8	-66.7
TF4	48.3	112	160.3	-136.1	-83.5	-219.6	-136.3	-36.4	-172.7
TF5	136.9	195.3	332.2	-73.3	5.5	-67.8	-385.4	-368.3	-753.7
TV1	307.1	350.1	657.2	41.4	98.8	140.2	164.3	192.1	356.4
TF2B	336.6	415.4	752	338.9	372.5	711.4	277.4	304.2	581.6

Table 2.3.3 Results of best majority passing rate for simulated real time C5 cycle testing of S&P 500 futures using the TFx methods with cycle database during 01/2001 and 04/2006

Method	Best Passing Rate (%)	Long	Short	Total
TF1	70	354.4	420	774.4
TF2	60	442.7	501.9	944.6
TF3	30	249.6	291.5	541.1
TF4	50	133.1	221.8	354.9
TF5	30	238.2	271.7	509.9
TV1	50	400.7	480.9	881.6
TF2B	50	359	407	766

2.4.3 Comparison of the Forecasting Performances of the TFx Methods

TF1:

$X = P_B - P_C$
$Y = P_C - P_D$

TF2:

$X = (P_A - P_C) / (P_D - P_C)$
$Y = (P_D - P_C) / (P_D - P_E)$

TF3:

$X = (P_B - P_A) / (P_B - P_C)$
$Y = (P_B - P_C) / (P_D - P_C)$

TF4:

$X = P_B - P_A$
$Y = P_B - P_C$

TF5:

$X = P_A - P_B$
$Y = P_B - P_C$
$Z = P_C - P_D$

TV1:

$X = (P_B - P_C) / (P_D - P_C)$
$Y = (P_D - P_C) / (P_D - P_E)$

TF2B:

$X = (P_A - P_C) / (P_D - P_C)$
$Y = (P_D - P_C) / (P_D - P_E)$
$Z = (P_D - P_E) / (P_F - P_E)$

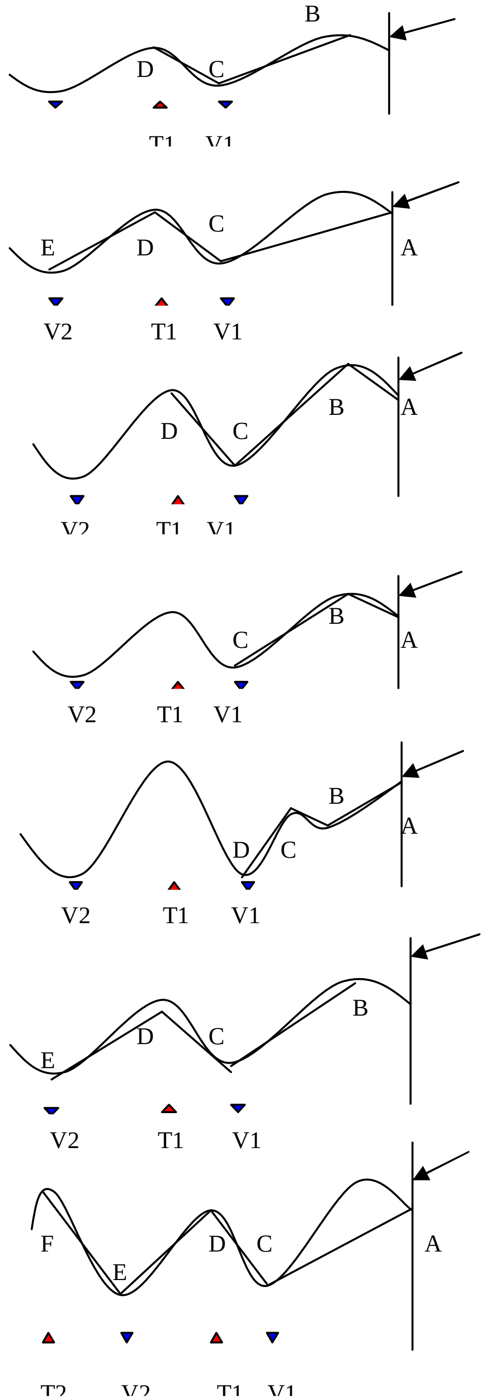


Figure 2.12 The sketch of the TFx methods.

As discussed above, the TFx methods shown in Figure 2.12 are built upon a cycle view of the market, and make use of the KNN algorithm to select analogous past cycles as guidance for dealing with the present cycle. Previous studies conducted by Dr. Yao, Dr. Xu and Dr. Zhao[39-41] had partial results about individual TFx methods (i.e. TF1, TF2, TF3, TF4 and TF5) and conclusions. In my research reported here I build on this work, and include two new TFx methods (TV1 and TF2B) and conduct an optimization of the passing rate. In the following, I will review the differences in results among those methods and review their strengths and weaknesses.

Based on the characteristic features used in the methods and using the language of economics, we can roughly classify them into two groups as coincident and trailing indicators. The first type, coincident indicators, includes TF1, TF2, TF2B and TV1. Trailing indicators comprise TF3, TF4 and TF5. A further differentiation is the use of absolute price differences vs. the use of ratios of price differences as coordinates to construct X-Y or X-Y-Z plot. Among the TFx methods, TF1, TF4 and TF5 belong into the former type, while the others use ratios. Furthermore, three-dimensional X-Y-Z plot has only been used in the TF2B and TF5 methods, while a two-dimensional X-Y plot has been applied to other TFx methods.

In the TF1 method, the current market situation is inferred from one previous cycle leg and the current market cycle segment. The TV1 method uses two previous cycle legs and the current market cycle segment. These two methods are considered as coincident indicators because they tend to move with changes in the market (cycle direction) and thus have the potential attribute of entering or exiting the market too early before an actual market extreme appears. In terms of the passing rate, a larger number may help the market cycle forecast since the requirement of more passing cycles in the cohorts can avoid too early action. The optimization results from the above sections indicate that for the TF1 method, the best passing rates are 70% for both C2 and C5; and for the TV1 method, the rates are 40% for C2 and 50% for C5. The best passing rates for the TF1 method lend some support for the coincident attribute, but not much for the TV1 method.

Both the TF2 and TF2B methods use two or three previous cycle legs and the current price leg as the plot coordinates to select cohort members; therefore, these two methods are also classified as coincident indicators because they use the state of the current market that may or may not have passed the actual cycle extreme. Based on the results from the optimization, the best rates are 50% for C2 and 60% for C5 to the TF2 method; the reverse best rate values are obtained for the TF2B method, which are 60% for C2 and 50% for C5. The results illustrate that half or a little more than half passing cycles in the cohorts will give best market forecast for those methods.

In the TF3, TF4 and TF5 methods, the retracement between last "confirmed" cycle extreme and the present time are used as one of coordinates in the plot. Since a certain market retracement is necessary to confirm next cycle extreme, it is imaginable that the trade signal will be late compared to coincident indicators. According to the optimization results, for the TF3 method, the best rates are 40% for C2 and 30% for C5; for the TF4 method, the best rates are 30% and 50% separately; and, for the TF5 method, the best rates are 50% and 30% for C2 and C5. Those best passing rates are all equal to or lower than 50%. It provides some evidences for the trailing attribute of above methods.

When comparing the trading results of the TFx methods, we find that coincident indicators (TF1, TF2 and TV1) tend to perform better than trailing indicators. The mediocre performances by trailing indicators (TF3, TF4 and TF5) may be due to their trailing nature. Those indicators usually don't provide the trading signal until a sizable market retracement occurs. Therefore, it is difficult for them to make a large profit. However, the coincident indicators utilize present and past cycle information in a valid manner so that they have better performances.

Since different TFx methods have their own market forecast attributes, either coincident or trailing, it is feasible and potentially useful to combine some of the methods to complement each other. In Dr. Zhao's thesis[41], she presented some work about the combination of two TFx methods. I will further explore such combinations by use of an artificial neural network in Chapter 4.

2.5 Improvement of TFx Short-Cycle Forecasts by Data Sorting

The short-term cycle TFx forecasts discussed here use a raw cycle leg data bank that represents an unsorted combination of data from both rising and falling markets. Since an assumption about the phase of long-term cycles such as CQ, will introduce only a quantifiable additional uncertainty while potentially improving the short-term cycle forecasts substantially, a study of the effect of sorted input data on the forecast quality was undertaken.

In order to study the effect of the background on the TFx methods, we use short-term cycles as the trading cycles, and long-term cycles as control sorting cycles. Based on the control cycle phase, the background is now split into two parts; one is for an up-trending market (control cycle up-leg), and the other is for a down-trending market (control cycle down-leg). When the present time lies in the up-trending market, the short-term historical cycles located in the up-trending market will be used as the background, while those short-term cycles in the down-trending market form the other background.

2.6 Optimization of the Passing Rate for the TFx Methods under Longer-Term Control Cycle

As discussed above, in the TFx methods, future market action is forecast by the behavior of a fraction of K nearest neighbors with similar pattern background. Since the cohorts are chosen from historical cycle background, they play an important role in the determination of the trading signal of the TFx methods. Generally, the more analogous the background is to the current market, the more accurate the forecast.

As we know, the shapes of short-term cycles (i.e., C2 and C5) will be affected by longer-term market trend. For example, in the up-trending market, the length of C5 up-leg tends to be longer than that of its down-leg. In our research, the phase of long-term cycles (i.e., CQ and CY) can be viewed as the long-term market trend.

In the following study, “man-made” long-term cycles (CQ and CY) are used as the control cycles that will determine the background based on their cycle phases, while short-term cycles (C2 and C5) are the trading cycles. Here, “man-made” cycles are determined by experienced human operators based on the observation of the market, with knowledge of the whole market movement and the specific rules about the cycle length and local market extreme prices. “Man-made” cycles are idealized cycles which are not available in real-time.

2.6.1 C2 Cycle Optimization Results under Long-Term Cycle Control

In section 2.4, we presented the trading results for the TFx methods without use of the control cycle. Here, we will apply the same procedure to the optimization of the passing rate for the TFx methods with “man-made” (i.e., visually selected) long-term cycles as control cycles. The S&P futures market during 01/2001 to 04/2006 is still the subject being tested.

Table 2.4.1 and 2.4.2 present the trading results of the TFx methods when the trading cycle is C2, and the control cycle is CQ. As mentioned above, the background has been split into an up-trending part when the trading cycle C2 lies in the CQ up phase, and a down-trending part when C2 is in the CQ down phase. Table 2.4.3 shows the best passing rate for every TFx method.

Figure 2.13 shows the corresponding histograms of trading results, which are from Table 2.4.3. Basically, every TFx method performs well and the profit increasing is very encouraging, with no major loss period. The performance increase is ascribed to the use of the “man-made” control cycle, which provides a better candidate pool to choose from and then better guidance to future market action.

Table 2.4.1 Results of simulated real time C2 cycle testing of S&P 500 futures using the TFX methods (passing rates from 30% to 50%, CQ ideal cycle as the control cycle) with cycle database during 01/2001 and 04/2006

Method	30%			40%			50%		
	Long	Short	Total	Long	Short	Total	Long	Short	Total
TF1	724.4	785.3	1509.7	1000.9	1051.4	2052.3	903.9	932.9	1836.8
TF2	911.3	958.1	1869.4	766.7	824.5	1591.2	909.6	977.7	1887.3
TF3	878.7	944	1822.7	665	735.6	1400.6	887.4	930.5	1817.9
TF4	1014.8	1092.9	2107.7	918	985.2	1903.2	797.1	858	1655.1
TF5	815.3	853	1668.3	696.6	766.3	1462.9	809.7	851.5	1661.2
TV1	1054.8	1104.6	2159.4	1328.8	1382.1	2710.9	1079.1	1152	2231.1
TF2B	611.1	674.9	1286	677.3	734	1411.3	712.9	769	1481.9

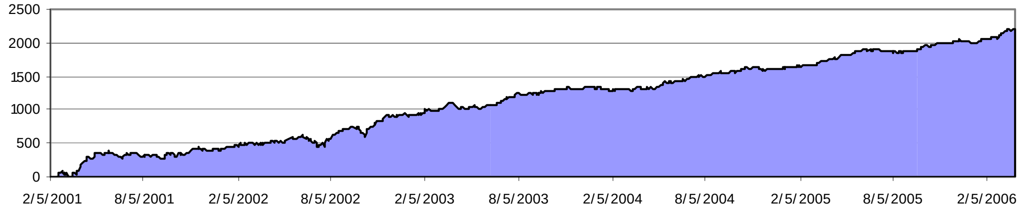
Table 2.4.2 Results of simulated real time C2 cycle testing of S&P 500 futures using the TFX methods (passing rates from 60% to 80%, CQ ideal cycle as the control cycle) with cycle database during 01/2001 and 04/2006

Method	60%			70%			80%		
	Long	Short	Total	Long	Short	Total	Long	Short	Total
TF1	1080.5	1114.3	2194.8	491.2	544.2	1035.4	575.7	630.8	1206.5
TF2	782.5	825.1	1607.6	403.5	452.2	855.7	468.3	519.7	988
TF3	842.1	892.1	1734.2	760.3	797.2	1557.5	523.2	575.2	1098.4
TF4	668.5	730.9	1399.4	565.1	615.1	1180.2	336.9	376.9	713.8
TF5	716	749.5	1465.5	755.6	782	1537.6	427.2	469.7	896.9
TV1	671.1	727.4	1398.5	618.4	706.1	1324.5	437.4	516.8	954.2
TF2B	1035.5	1107.8	2143.3	415.7	474.6	890.3	226.2	282.9	509.1

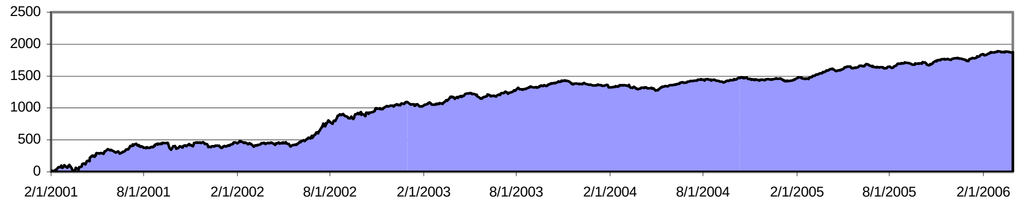
Table 2.4.3 Results of best passing rate for simulated real time C2 cycle testing of S&P 500 futures using the TFx methods and a control cycle with cycle database during 01/2001 and 04/2006, "man-made" CQ cycle is used as the control cycle

Method	Best Passing Rate (%)	Long	Short	Total
TF1	60	1080.5	1114.3	2194.8
TF2	50	909.6	977.7	1887.3
TF3	30	878.7	944	1822.7
TF4	30	1014.8	1092.9	2107.7
TF5	30	815.3	853	1668.3
TV1	40	1328.8	1382.1	2710.9
TF2B	60	1035.5	1107.8	2143.3

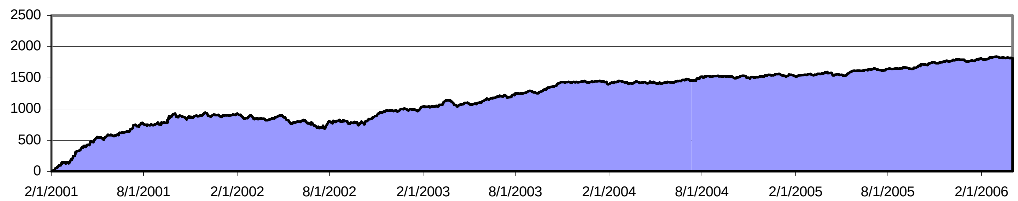
TF1



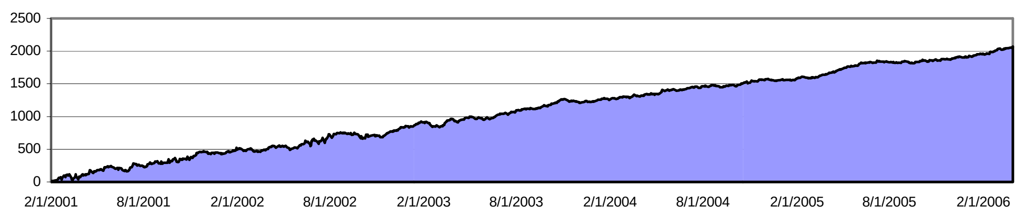
TF2



TF3



TF4



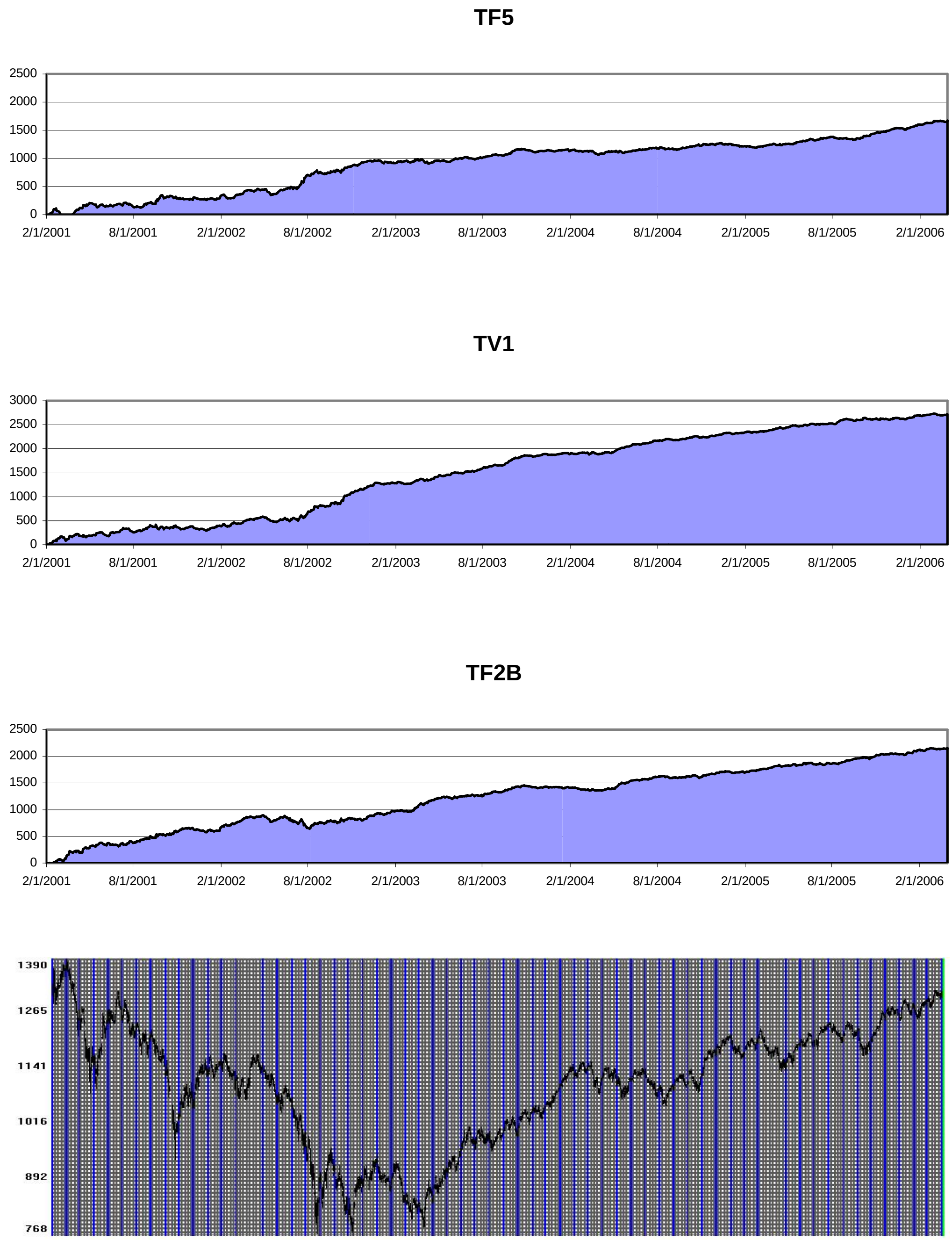


Figure 2.13 Histograms of TFx trading results with best passing rates during 01/2001 to 04/2006 and S&P price chart at the same period. The graphs from top to bottom represent TF1, TF2, TF3, TF4, TF5, TV1 and TF2B, respectively.

Table 2.5.1 Results of simulated real time C2 cycle testing of S&P 500 futures using the TFX methods (passing rates from 30% to 50%, CY ideal cycle as the control cycle) with cycle database during 01/2001 and 04/2006

Method	30%			40%			50%		
	Long	Short	Total	Long	Short	Total	Long	Short	Total
TF1	98.6	167.8	266.4	454.8	513.3	968.1	423.1	475.9	899
TF2	346.4	414.3	760.7	415.6	481.4	897	454.6	508.1	962.7
TF3	210.3	273.9	484.2	246.9	318.5	565.4	560	617.5	1177.5
TF4	415.6	513.7	929.3	307.1	402.2	709.3	485.1	549.6	1034.7
TF5	290.6	364.3	654.9	409.7	464.3	874	357.5	420.1	777.6
TV1	497.5	584.8	1082.3	614.5	689	1303.5	563.6	646.7	1210.3
TF2B	332.6	375	707.6	163.4	232.4	395.8	229.9	292.6	522.5

Table 2.5.2 Results of simulated real time C2 cycle testing of S&P 500 futures using the TFX methods (passing rates from 60% to 80%, CY ideal cycle as the control cycle) with cycle database during 01/2001 and 04/2006

Method	60%			70%			80%		
	Long	Short	Total	Long	Short	Total	Long	Short	Total
TF1	728	789.7	1517.7	428	505.9	933.9	179.4	243	422.4
TF2	504.4	561.3	1065.7	72.8	128.5	201.3	261.3	320.4	581.7
TF3	346.5	390.2	736.7	343.6	396.9	740.5	281.5	340.6	622.1
TF4	467.9	496.5	964.4	346.3	374.7	721	180.3	200.3	380.6
TF5	371.5	430.2	801.7	266.9	296	562.9	232.3	255.6	487.9
TV1	363	426.6	789.6	184.3	235.3	419.6	155.9	233.7	389.6
TF2B	332	402.8	734.8	309.2	382.4	691.6	229.5	301.4	530.9

Table 2.5.3 Results of best passing rate for simulated real time C2 cycle testing of S&P 500 futures using the TFx methods and a control cycle with cycle database during 01/2001 and 04/2006, "man-made" CY cycle is used as the control cycle

Method	Best Passing Rate (%)	Long	Short	Total
TF1	60	728	789.7	1517.7
TF2	60	504.4	561.3	1065.7
TF3	50	560	617.5	1177.5
TF4	50	485.1	549.6	1034.7
TF5	40	409.7	464.3	874
TV1	40	614.5	689	1303.5
TF2B	60	332	402.8	734.8

Table 2.5.1, 2.5.2 and 2.5.3 contain the trading results with CY as the control cycle. The best trading results of the TFx methods with or without use of the control cycle, which are from Table 2.1.3, 2.5.3 and 2.5.3, are displayed in Figure 2.14. We can draw the conclusion that the control cycle has positive impact on the performance of the TFx methods. The CPL of every TFx method had been improved to larger or smaller extent. With a “man-made” CQ as the control cycle, the results for any TFx methods are roughly one to three times better than those without use of the control cycle. When CY is used as the control cycle, the improvement is not that significant as with CQ, but the result is still a quarter to one times better than that without use of a control cycle. It is realized, of course, that this improvement has been the result of relieving some of the uncertainty of the trading cycle analysis by the assumption of a correct phase analysis for the correct cycle. An appropriate follow-up study would be to combine a CQ or CY phase forecast with the trading cycle analysis given here.

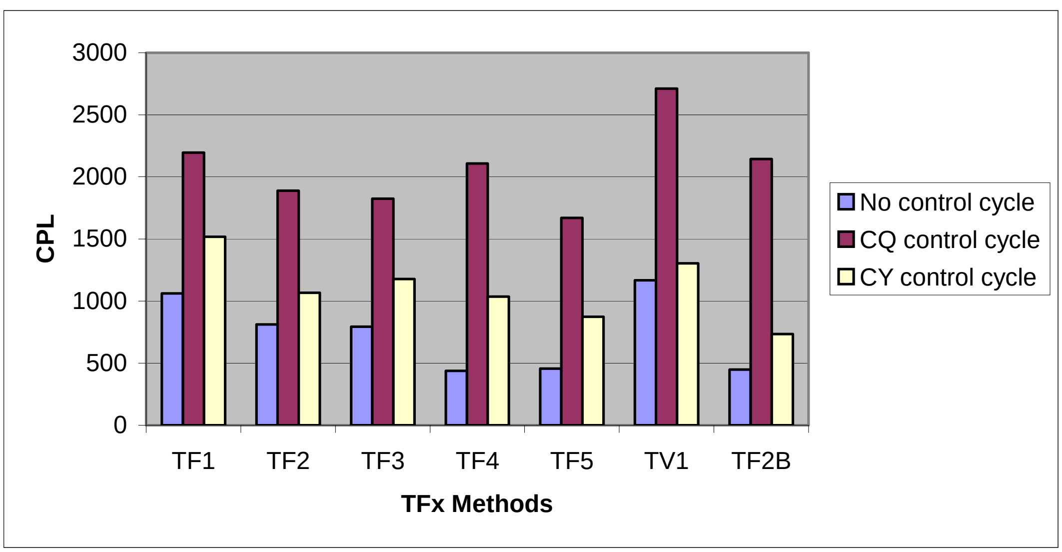


Figure 2.14 C2 cycle Best Trading Results for the TFx Methods with or without use of Control Cycle. Data are from Table 2.1.3, 2.4.3 and 2.5.3.

2.6.2 C5 Cycle Optimization Results under Long-Term Cycle Control

In this section, we conduct the optimization of the passing rate for the TFx methods with C5 cycle as the trading cycle. This study is parallel to that in the preceding section 2.5.1 except that C5 replaces C2 as the trading cycle. Table 2.6.1, 2.6.2 and 2.6.3 show the trading results when "man-made" CQ is the control cycle. Table 2.7.1, 2.7.2 and 2.7.3 contain the results with "man-made" CY as the control cycle.

The comparison of the best C5 cycle trading results using the TFx methods with or without use of the control cycle are illustrated in Figure 2.15 that uses data from Table 2.3.3, 2.6.3 and 2.7.3. According to the graph, we can reach the same conclusion drawn above for C2 that the control cycle has a pronounced effect on the performance of the TFx methods. Again, use of a "man-made" CQ control cycle has a larger impact than use of an analogous CY control cycle, while the impact of control cycle use on C5 is on average 1.3 times larger than on C2.

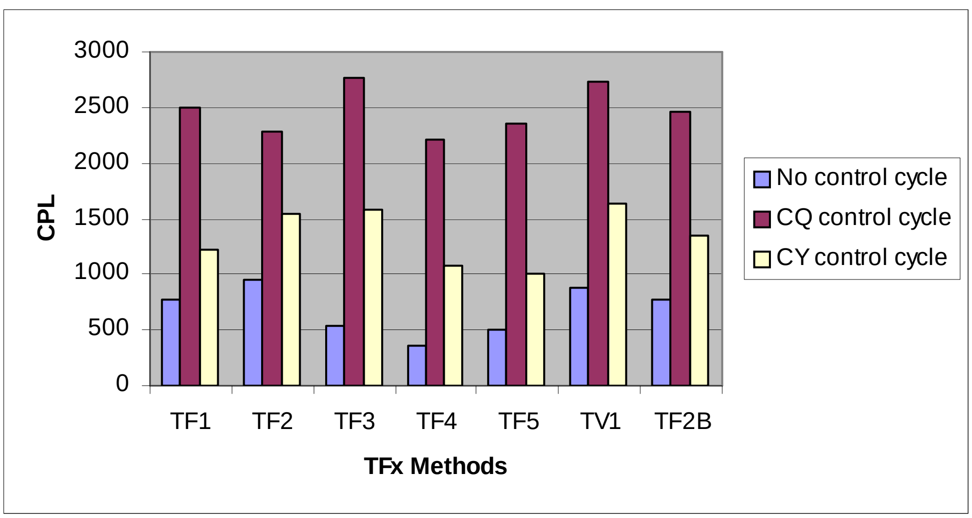


Figure 2.15 C5 cycle Best Trading Results for the TFx Methods with or without use of the Control Cycle. Data are from Table 2.3.3, 2.6.3 and 2.7.3.

Table 2.6.1 Results of simulated real time C5 cycle testing of S&P 500 futures using the TFX methods (passing rates from 30% to 50%, CQ ideal cycle as the control cycle) with cycle database during 01/2001 and 04/2006

Method	30%			40%			50%		
	Long	Short	Total	Long	Short	Total	Long	Short	Total
TF1	1068.9	1233.1	2302	1177.8	1236.7	2414.5	1059.7	1190.6	2250.3
TF2	1048.5	1132.3	2180.8	1110	1179.1	2289.1	1084.7	1180.7	2265.4
TF3	802.4	906.7	1709.1	1023.9	1136.4	2160.3	1347.4	1415.7	2763.1
TF4	1082.3	1121.3	2203.6	865.7	899.5	1765.2	866.8	936	1802.8
TF5	1167.3	1186.7	2354	957.8	1056	2013.8	875	911.4	1786.4
TV1	1177.9	1261.9	2439.8	1211	1323.4	2534.4	1307.5	1418.8	2726.3
TF2B	1091	1181.4	2272.4	1151.9	1194.5	2346.4	977.6	1029.5	2007.1

Table 2.6.2 Results of simulated real time C5 cycle testing of S&P 500 futures using the TFX methods (passing rates from 60% to 80%, CQ ideal cycle as the control cycle) with cycle database during 01/2001 and 04/2006

Method	60%			70%			80%		
	Long	Short	Total	Long	Short	Total	Long	Short	Total
TF1	1206.1	1289.9	2496	822.6	952.2	1774.8	617.8	754.4	1372.2
TF2	992.8	1071	2063.8	835.8	949.4	1785.2	630.4	723.3	1353.7
TF3	989	1081.7	2070.7	723.2	795.2	1518.4	468.4	576.8	1045.2
TF4	766.9	816.7	1583.6	399.6	488.6	888.2	472.2	509.5	981.7
TF5	662.5	686.4	1348.9	628.5	658.1	1286.6	584.9	656.6	1241.5
TV1	1102.3	1183.8	2286.1	975	1127.6	2102.6	501.1	579.6	1080.7
TF2B	1208.2	1246.5	2454.7	949.6	1024.5	1974.1	682.4	764.3	1446.7

Table 2.6.3 Results of best passing rate for simulated real time C5 cycle testing of S&P 500 futures using the TFx methods and a control cycle with cycle database during 01/2001 and 04/2006, “man-made” CQ cycle is used as the control cycle

Method	Best Passing Rate (%)	Long	Short	Total
TF1	60	1206.1	1289.9	2496
TF2	40	1110	1179.1	2289.1
TF3	50	1347.4	1415.7	2763.1
TF4	30	1082.3	1121.3	2203.6
TF5	30	1167.3	1186.7	2354
TV1	50	1307.5	1418.8	2726.3
TF2B	60	1208.2	1246.5	2454.7

Table 2.7.1 Results of simulated real time C5 cycle testing of S&P 500 futures using the TFX methods (passing rates from 30% to 50%, CY ideal cycle as the control cycle) with cycle database during 01/2001 and 04/2006

Method	30%			40%			50%		
	Long	Short	Total	Long	Short	Total	Long	Short	Total
TF1	505.2	553	1058.2	557.3	566.6	1123.9	611.9	606.6	1218.5
TF2	579.4	642.6	1222	643.1	676.9	1320	492	540	1032
TF3	390.2	461.9	852.1	778.5	805.5	1584	531.4	571.1	1102.5
TF4	539.1	546.5	1085.6	521.5	539.8	1061.3	463	501.3	964.3
TF5	482.2	518.4	1000.6	450.2	530.7	980.9	220.2	273.4	493.6
TV1	771.4	812.9	1584.3	747.5	782	1529.5	816.4	812.2	1628.6
TF2B	525.2	520.2	1045.4	623.2	624.7	1247.9	394.6	404.9	799.5

Table 2.7.2 Results of simulated real time C5 testing of S&P 500 futures using TFX methods (passing rates from 60% to 80%, CY ideal cycle as the control cycle) with cycle database during 01/2001 and 04/2006

Method	60%			70%			80%		
	Long	Short	Total	Long	Short	Total	Long	Short	Total
TF1	533.1	565.4	1098.5	383.8	429.4	813.2	574.9	610	1184.9
TF2	747.8	789.1	1536.9	415.9	534.2	950.1	216.1	307.6	523.7
TF3	410.4	486	896.4	316.7	372.6	689.3	143.5	197.5	341
TF4	256.2	329	585.2	-14.5	105.3	90.8	-61	8.7	-52.3
TF5	115.3	126.8	242.1	57.8	111.3	169.1	-174.4	-113.8	-288.2
TV1	712.8	704.5	1417.3	543.3	623.7	1167	445.5	543.5	989
TF2B	627	718.2	1345.2	323.5	382.3	705.8	393.2	460	853.2

Table 2.7.3 Results of best passing rate for simulated real time C5 cycle testing of S&P 500 futures using the TFx methods and a control cycle with cycle database during 01/01/2001 and 04/04/2006, “man-made” CY cycle is used as the control cycle

Method	Best Passing Rate (%)	Long	Short	Total
TF1	50	611.9	606.6	1218.5
TF2	60	747.8	789.1	1536.9
TF3	40	778.5	805.5	1584
TF4	30	539.1	546.5	1085.6
TF5	30	482.2	518.4	1000.6
TV1	50	816.4	812.2	1628.6
TF2B	60	627	718.2	1345.2

Based on the results of Figure 2.13 and Figure 2.14, it has been demonstrated that selecting cohorts from correct backgrounds can improve the forecast accuracy of the TFx methods considerably. The reason is that the split of background based on the control cycle phase will group more analogous cycles together, while excluding some cycles that will behave differently under different control cycle phase. Therefore, when current market trend is known in advance, there is a bigger chance to obtain more analogous past cycles from the split background and provide better guidance for future market action.

The results show that the C5 trading results with CQ as the control cycle are better than with CY as the control cycle. The reason is similar to that given above for C2 trading. In terms of cycle length, CQ is shorter than CY. Thus, CQ classifies the market more accurately than CY, and can provide a better background than CY, which will benefit the selection of cohort members. A comment about the shift of uncertainty by use of control cycles was given in 2.5.

2.7 Summary

In this chapter, the TFx methods, which are based on a cycle model and the KNN method, are introduced and discussed. Two new TFx methods, TV1 and TF2B, are presented. Based on the characteristic features used in the TFx methods, they can be classified as belonging to two types: coincident and trailing indicators. The strengths and weaknesses of the TFx methods have been discussed.

The TFx methods have been applied to the forecast of S&P 500 futures market. The trading results demonstrate that by selecting historical analogous cycles correctly, a profitable trading system can be constructed. It also validates the axiomatic statement that the market's historical pattern repeats itself and thus can provide some sort degree of valid advice to the current market.

We also conducted the optimization of the passing rate, which is the percentage of cohort members assuming to have passed their next cycle extreme, for the TFx methods, with C2 and C5 as the trading cycles. The optimization results provide better operational parameters for each method, and confirm their classification as coincident or trailing indicators.

Since the behavior of short-term cycles can be affected by the market trend, we also conducted tests for the TFx methods with “man-made” long-term cycles (CQ and CY) as the control cycles. Compared to trading results without use of the control cycle, use of “man-made” control cycles significantly improves the forecast, especially for CQ as a control cycle. The reason is that the splitting of the background data based on the market trend (control cycle phase) will combine more analogous cycles by excluding cycles with different regime.

Although “man-made” control cycles contribute a lot to the improvement of the forecast quality of the TFx methods, they are not available in real-time. In the next chapter, I attempt to replace “man-made” control cycles with ones inferred from Bayes’ Theorem by using real-time observable short-term market characteristics, such as daily up-or-down sequences.

Chapter 3 Analysis of Control Cycle Phase from Short-Term Market Actions Using Bayesian Statistics

3.1 Overview

As discussed in Chapter 2, the TFx methods, which are cycle-based pattern recognition methods, generate a trade signal by polling k cohort members selected from a background composed of historical, “man-made” cycles based on perceived similarities between these historical situations and the current one. By comparing the trading results with and without use of the control cycle, we found that use of a longer-period control cycle pre-selected to have the correct phase significantly improves the shorter-period trading cycle market forecast. However, this (historical-based) background cycle phase selection used above is based on an after-the-fact analysis, which uses posterior knowledge and must be replaced with a real-time phasing method using on-time data, as discussed below.

In this chapter, I am going to use Bayes’ Theorem to infer control cycle phase from real-time observable sequences of short-term market actions and apply it to a probabilistic market background pre-selection in the TFx methods. Four short-term observable market actions containing weekday, week and short-term cycle (C2, C5) information will be utilized in this approach and evaluated to produce CQ and CY cycle classification data via Bayesian Statistics for use in the trading cycle (C2, C5) market forecasts. These relations are schematically shown in Figure 3.1.

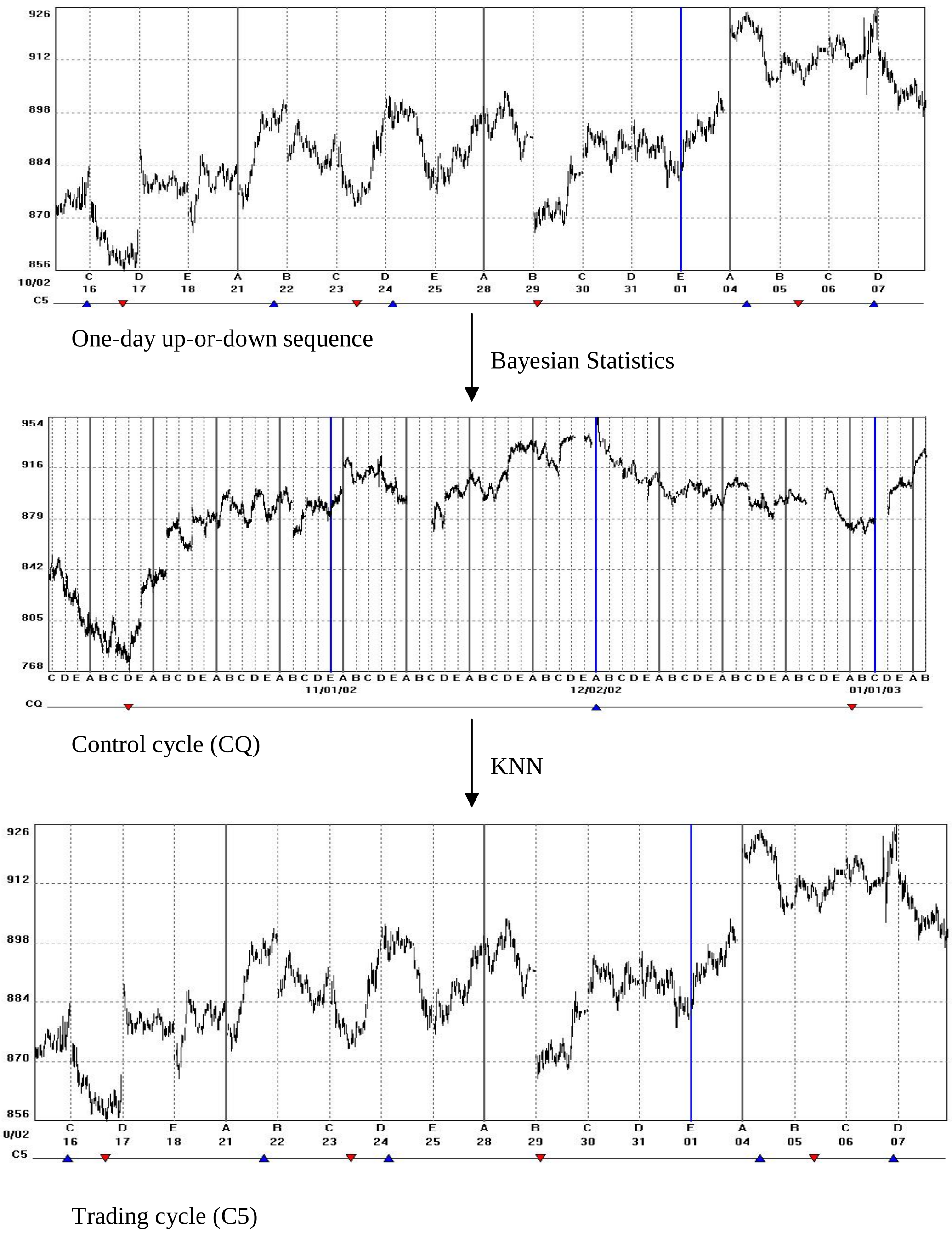


Figure 3.1 Conceptual steps in C5 cycle analysis (C5 is chosen as a paradigm for a short-term cycle) using a quarter-year-cycle (CQ) as the control cycle, the phasing of which is inferred from short-term (e.g. one-day) market action sequences via Bayesian statistics.

Section 3.2 gives an introduction to Bayes' Theorem. Section 3.3 presents four types of market action statistics that can be used to infer long-term cycle (CQ or CY) phases. The four sets of data used are daily up-or-down sequences, weekly up-or-down sequences, short-term cycle (C2 or C5) leg-length-ratio sequences, and short-term cycle (C2 or C5) extreme up-or-down sequences. Section 3.4 shows the trading results of the different TFx methods when long-term cycle phases predicted by use of Bayesian statistics from the above-selected data are used in the background pre-selection. Section 3.5 discusses the results and draws some conclusion.

3.2 Review of Bayes' Theorem

Bayes' Theorem is a theorem of probability theory originally stated by the Reverend Thomas Bayes in 1764[72]. It provides a way to calculate the conditional probability of the correctness of a hypothesis (in this case, the phase of a long-term cycle) based on its a priori probability, the conditional probabilities of observing certain data (in this case, the above-listed sequences) given the correctness of the hypothesis, and the observed data themselves (in this case, the sequence of market actions)[72-74]. Here, I present a brief derivation of this much-used relationship as below.

Bayes' Theorem arises directly out of the definitional relationship between conditional and joint probabilities shown in equation 3.1

$$P(A, a) = P(a|A) * P(A) \qquad \textbf{3.1}$$

Where
$P(A, a)$: the joint probability of the simultaneous occurrences of events A and a
$P(a|A)$: the probability of the occurrence of event a given that event A has occurred
$P(A)$: the probability of occurrence of event A

If events A and a are independent, then the occurrence of event A tells us nothing about the occurrence of event a, and we can get the equation: $P(a|A) = P(a)$, so that above equation reduces to $P(a,A) = P(a) * P(A)$. However, if events A and a are dependent, then the fact that event A has occurred changes our expectations regarding the probability of event a, and we have to take this into account in our computation of the joint probability.

It is clear that we can also write the joint probability in terms of the conditional probability $P(A|a)$ as:

$$P(A, a) = P(A|a) * P(a) \qquad \textbf{3.2}$$

Where
$P(A|a)$: the probability of the occurrence of event A given that event a has occurred
$P(a)$: the probability of occurrence of event a

The above two equations (equation 3.1 and 3.2) represent the same joint probability, thus, equating them yields a relationship between the two conditional probabilities, $P(a|A)$ and $P(A|a)$, in equation 3.3:

$$P(A|a) = P(a|A) * P(A) / P(a) \qquad \textbf{3.3}$$

Equation 3.3 is a form of Bayes' Theorem, which gives us a means for deriving one form of conditional probability from another. Quite often event A represents some hypothesis that is not directly observable and event a represents an observable consequence. In particular, P(A), also known as the prior probability of event A, represents our best estimate of the probability of event A, prior to consideration of the new piece of information. $P(A|a)$, which is the probability that event A is true given that event a is true, is also known as the posterior probability of event A.

3.3 Bayesian Statistics in Cycle Analysis

Bayes' Theorem basically relates the conditional and a priori probabilities of two dependent events to compute posterior probabilities given observations. As to cycle analysis, the probabilities of long-term cycle phase or market trend can be viewed as posterior probabilities, and Bayes' Theorem provides a possible approach to infer control cycle phase from prior probabilities and observation of some specific market characteristics such as day-to-day market direction.

As shown in Chapter 2, the long-term cycle phase or market trend will affect the development of short-term cycle or market movement. For instance, in an up-trending market, we will observe that the market will keep moving up, although occasionally there is also some sort of retracement. Therefore, in this case, the next short-term cycle top will be higher than the present cycle top, and the next short-term cycle valley will also higher than the present cycle valley. When the market reverses to down-trend, the corresponding relationship will change in the opposite way, which is that the next short-term cycle top will be lower than that of the present cycle top, and the next short-term cycle valley will be also lower than that of the present cycle valley. Since long-term market trend and short-term market trend can be considered as dependent events, the prerequisite of Bayes' Theorem has been met and it can be applied in our research.

As stated here, in terms of equation 3.3, the long-term cycle phase is event A, which is difficult to recognize in real-time, while short-term market characteristics, which are easy to observe in real-time, forms event *a*. In this section, we choose four different short-term market characteristics, which are: daily up-or-down sequences, weekly up-or-down sequences, short-term cycle (C2 or C5) leg-length-ratio sequences, and short-term cycle (C2 or C5) extreme up-or-down sequences, in our research. For each pair of long-term cycle phase and short-term market characteristics, we can obtain the corresponding Bayesian Statistics, which is crucial in the determination of the present long-term cycle phase.

In our research, the long-term cycle refers to CQ or CY, meanwhile, the short-term cycle means C2 or C5. The market used in this study is S&P futures during the period 01/2001 and 04/2006.

3.3.1 Daily Up-or-Down (ud) Sequence Statistics

In financial markets, the price of securities change almost every day, either up or down, and rarely remain unchanged. The price variation is affected by market trend; for example, in an up-trending market, the price will be trending up gradually despite of some retracements. Thus, weekday price movements and market trend are two dependent events with have positive correlation. Therefore, we can apply Bayes' Theorem to obtain statistics about the two events.

Following the same terms, event A is defined as the phase of a long-term cycle, CQ or CY; meantime, event *a* are the daily price up-or-down sequences. In this study, a day is defined as an up day when the present day's closing price is higher than that of the previous day; otherwise, it is classified as a down day. u is used to represent an up day, and d represents a down day. Thus, the weekday price variation sequence can be labeled as a combination of u and d. For instance, a four-day ud sequence can be described as uuud, which means that the first three days are up days, and the fourth day is a down day. In this research, we only focus on three-day and four-day ud sequences.

As a guidance to the reader, I discuss the specific case of the calculation of the conditional probability of having a CQ up-leg if a uuu sequence is observed. Table 3.1.1 contains three-day ud sequences statistics with CQ as the control cycle for S&P from 01/2001 to 04/2006. In the table, the “Cycle Phase” column specifies the up-or-down characteristics of the control cycle; the “Number” column represents the total number of three-day sequences in the corresponding cycle phase; the “P(A)” column shows the probability of a certain control cycle phase (Equation 3.4).

$$P(A) = N / N_T \tag{3.4}$$

Where
P(A): the probability of certain control cycle phase
N: the number of sequences in certain control cycle phase
N_T: total number of all sequences

The probability of a CQ up-leg period is 58.4%, whereas that of a down-leg period is only 41.6% for S&P. It means there is one third more up-period than down-period in terms of CQ.

The “Sequence” column shows the three-day price variations represented by the combination of u and d. There are eight possible sequences, which are uuu, uud, udu, udd, duu, dud, ddu and ddd.

Table 3.1.1 Three-day ud sequence statistics ("man-made" CQ as control cycle) for S&P futures during 01/2001 and 04/2006. Event *a* is a specific three-day ud sequence, and event A is a specific CQ cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CQ Up	719	0.584	uuu	137	0.137	0.191	0.814
			uud	101	0.122	0.14	0.670
			udu	120	0.141	0.167	0.692
			duu	101	0.122	0.14	0.670
			udd	72	0.124	0.1	0.471
			dud	88	0.144	0.122	0.495
			ddu	73	0.128	0.102	0.465
			ddd	27	0.083	0.038	0.267
CQ Down	512	0.416	uuu	32		0.063	0.191
			uud	49		0.096	0.327
			udu	53		0.104	0.307
			duu	49		0.096	0.327
			udd	81		0.158	0.530
			dud	89		0.174	0.503
			ddu	84		0.164	0.533
			ddd	75		0.146	0.732

The statistics is based on a "man-made" CQ cycle database. First, we label each day as u or d by comparison of closing prices of the present day and its previous day. Then, we produce three-day ud sequences for the whole period by replacing each day with the corresponding label and combining them. Once three-day sequences are available, under same control cycle phase, we look for three-day cases with same sequence, and add them together to get the total occurrence number of the specific sequence. For example, for a three-day sequence uuu in a CQ up-leg period, we find all uuu in that period, then, adding those together to get the number of 137, which means there are 137 occurrences of the uuu sequence in a CQ up-leg period.

The “P(*a*)” column shows the occurrence probability of a specific sequence in the test period. For example, P(uuu) is the occurrence probability of a uuu sequence, which represents three continuous up days. The total number of uuu sequence appearances in the test period is the sum of the number of uuu in CQ up-leg and down-leg periods (Equation 3.5). Therefore, the total number of uuu’s is 169, which is the sum of 137 and 32.

$$N_S = N_{SU} + N_{SD} \qquad \textbf{3.5}$$

Where

N_S: total number of a specific sequence

N_{SU}: number of this sequence in control cycle up-leg

N_{SD}: number of this sequence in control cycle down-leg

The total number of three-day sequences is the sum of all sequence occurrences in CQ up and CQ down phase (Equation 3.6). There are total 1231 cases of three-day ud sequences, which is the sum of 719 and 512.

$$N_A = N_{AU} + N_{AD} \qquad \textbf{3.6}$$

Where

N_A: total number of all sequences

N_{AU}: number of all sequences in control cycle up-leg

N_{AD}: number of all sequences in control cycle down-leg

P(*a*) is the occurrence probability of a specific sequence in the test period. Using Equation 3.7, and still taking the uuu sequence as an example, we can get the probability of having a uuu sequence, which is 0.137, the quotient of 167 over 1231.

$$P(a) = N_S / N_A \qquad \textbf{3.7}$$

$P(a|A)$ is the occurrence probability of a specific sequence given the control cycle phase. For instance, P(uuu | Up phase) means the appearance probability of three continuous up days under the circumstance of CQ up phase. Based on Equation 3.8, we can get its value, 0.191, which is the quotient of 137 over 719.

$$P(a|A) = N_{SU} / N_{AU} \quad \text{or} \quad P(a|A) = N_{SD} / N_{AD} \qquad \textbf{3.8}$$

We are now ready to proceed to the desired result. $P(A|a)$ represents the probability of control cycle phase given a specific three-day ud sequence. It is the posterior probability we try to obtain to help the TFx forecast. For example, P(Up phase | uuu) means the occurrence probability of a CQ up phase when the present three-day sequence is uuu. Based on the Bayesian Equation 3.9, the probability is calculated as 0.814.

$$P(A|a) = P(a|A) * P(A) / P(a) \qquad \textbf{3.9}$$

In Table 3.1.1, five sequences, uuu, uud, udu, duu and ddd, have the value of $P(A|a)$ over 60%. It indicates that there is an over 60% chance that the control cycle phase (or that long-period cycle market trend prediction) is correct when one of those five sequences occurs. For example, when the present three-day sequence is uuu, there exists an 81.4% chance that the current CQ is in an up-leg period.

Table 3.1.2 Four-day ud sequence statistics ("man-made" CQ as control cycle) for S&P futures during 01/2001 and 04/2006. Event *a* is a specific four-day ud sequence, and event A is a specific CQ cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CQ Up	697	0.587	uuuu	71	0.072	0.102	0.831
			uuud	59	0.063	0.085	0.792
			uudu	61	0.065	0.088	0.794
			uduu	58	0.063	0.083	0.773
			duuu	59	0.064	0.085	0.779
			uudd	38	0.057	0.055	0.566
			udud	59	0.078	0.085	0.639
			uddu	48	0.072	0.069	0.562
			duud	37	0.058	0.053	0.536
			dudu	55	0.077	0.079	0.602
			dduu	41	0.061	0.059	0.559
			uddd	22	0.052	0.032	0.348
			dudd	33	0.069	0.047	0.395
			ddud	29	0.068	0.042	0.351
			dddu	23	0.052	0.033	0.372
			dddd	4	0.028	0.006	0.126
CQ Down	491	0.413	uuuu	15		0.031	0.178
			uuud	16		0.033	0.216
			uudu	16		0.033	0.210
			uduu	17		0.035	0.230
			duuu	17		0.035	0.226
			uudd	30		0.061	0.442
			udud	34		0.069	0.366
			uddu	38		0.077	0.442
			duud	32		0.065	0.463
			dudu	36		0.073	0.392
			dduu	31		0.063	0.427
			uddd	40		0.081	0.644
			dudd	49		0.1	0.599
			ddud	52		0.106	0.644
			dddu	39		0.079	0.628
			dddd	29		0.059	0.871

We now proceed to the discussion of four-day sequences. Table 3.1.2 shows four-day ud sequence statistics. Different from three-day ud sequences, four-day ud sequences have 16 different forms, which are uuuu, uuud, uudu, uudd, uduu, udud, uddu, uddd, duuu, duud, dudu, dudd, dduu, ddud, dddu and dddd. The acquisition and calculation of data is similar to that of Table 3.1.1. In this table, there are 12 sequences which have an over 60% chance to make a correction phase prediction. Among those sequences, uuuu, uuud, uudu, uduu, duuu, udud and dudu can be utilized to forecast a CQ up phase, meanwhile, uddd, dudd, ddud, dddu and dddd are daily sequences to hint that the current market is in a CQ down phase.

Using “man-made” CY as the control cycle and same methodology as for CQ, we obtain the three-day and four-day ud sequence statistics shown in Table 3.2.1 and 3.2.2. For CY, more than two-third of time (69.2%) is in an up period, and only less than one-third of time (30.8%) is in a down period. Compared to data in Table 3.1.1, there is more market up time in CY than in CQ.

In three-day ud sequences statistics with CY as the control cycle, 7 out of 8 sequences except ddd predict that the present time is in a CY up-leg period and their probabilities are larger than that with CQ as the control cycle. For example, when the present three-day sequence is uuu, then there is 83.1% chance that it is in CY up-leg period, compared to 81.4% chance that it is in CQ up-leg period. As to four-day ud sequences statistics, 13 out of 16 sequences except uddd, dddu and dddd point to CY up-leg period. The number of sequences and the probabilities in CY phase forecast are still larger than their counterparts in CQ phase prediction. That there are more up periods for

CY than there are for CQ is the reason that there is a stronger preference for CY up phase under the same daily ud sequence.

Table 3.2.1 Three-day ud sequence statistics (“man-made” CY as control cycle) for S&P futures during 01/2001 and 04/2006. Event *a* is a specific three-day ud sequence, and event A is a specific CY cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(a)	P(a|A)	P(A|a)
CQ Up	916	0.692	uuu	148	0.135	0.162	0.831
			uud	120	0.124	0.131	0.731
			udu	137	0.14	0.15	0.742
			duu	120	0.125	0.131	0.726
			udd	107	0.126	0.117	0.643
			dud	123	0.141	0.134	0.658
			ddu	108	0.127	0.118	0.643
			ddd	53	0.082	0.058	0.490
CQ Down	407	0.308	uuu	31		0.076	0.173
			uud	44		0.108	0.268
			udu	48		0.118	0.259
			duu	45		0.111	0.273
			udd	60		0.147	0.359
			dud	64		0.157	0.343
			ddu	60		0.147	0.356
			ddd	55		0.135	0.506

Table 3.2.2 Four-day ud sequence statistics ("man-made" CY as control cycle) for S&P futures during 01/2001 and 04/2006. Event *a* is a specific four-day ud sequence, and event A is a specific CY cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CQ Up	910	0.695	uuuu	80	0.071	0.088	0.861
			uuud	68	0.065	0.075	0.802
			uudu	69	0.066	0.076	0.800
			uduu	67	0.065	0.074	0.791
			duuu	67	0.065	0.074	0.791
			uudd	51	0.059	0.056	0.659
			udud	69	0.075	0.076	0.704
			uddu	68	0.073	0.075	0.714
			duud	50	0.059	0.055	0.648
			dudu	67	0.074	0.074	0.695
			dduu	52	0.06	0.057	0.660
			uddd	38	0.053	0.042	0.550
			dudd	56	0.068	0.062	0.633
			ddud	54	0.066	0.059	0.621
			dddu	40	0.054	0.044	0.566
			dddd	14	0.027	0.015	0.386
CQ Down	400	0.305	uuuu	13		0.033	0.142
			uuud	17		0.043	0.202
			uudu	17		0.043	0.199
			uduu	18		0.045	0.211
			duuu	18		0.045	0.211
			uudd	26		0.065	0.336
			udud	29		0.073	0.297
			uddu	28		0.07	0.293
			duud	27		0.068	0.352
			dudu	30		0.075	0.309
			dduu	27		0.068	0.346
			uddd	31		0.078	0.449
			dudd	33		0.083	0.373
			ddud	33		0.083	0.384
			dddu	31		0.078	0.441
			dddd	22		0.055	0.622

3.3.2 Weekly Up-or-Down (ud) Sequence Statistics

In the previous section, daily ud sequences were used as short-term market characteristics to infer the current long-term market trend. As we know, the long-term market trend not only affects daily price variation, but also has an impact on the price change of a somewhat longer period, i.e. one week. In order to study the relation between long-term control cycle and short-term market price variation, we extend the time length to one week and conduct similar research to that in above section.

For simplicity, Monday to Friday is viewed as a week. u is assigned to a week when previous week's closing price (last Friday's closing price) is lower than this week's closing price (this Friday's closing price). In the same fashion, d is defined when closing price of previous week is higher than this week's closing price.

Table 3.3.1 and 3.3.2 contain three-week and four-week ud sequence statistics for S&P futures during 01/2001 and 04/2006, respectively. The "man-made" CQ is used as the control cycle.

Table 3.4.1 and 3.4.2 present three-week and four-week ud sequence statistics for S&P futures during 01/2001 and 04/2006. The "man-made" CY is used as the control cycle.

In these tables, event A is still the phase of long-term cycle, either CQ or CY, and event a are weekly ud sequences. The same principles as for daily ud sequences are applied in data collection. All those cells with $P(A|a)$ value over 60% can be used to make predictions for CQ or CY phases.

One important characteristic of weekly ud sequence statistics is that the probabilities of long-term cycle phases for certain sequences are high, especially for three-week ud sequences. Compared to that, the probabilities in daily ud statistics discussed before are weak. For example, in three-week ud sequence statistics with CQ as the control cycle, all sequences have strong prediction preference for CQ cycle phase, with most probabilities over 80%. Meanwhile, in three-day ud sequences statistics, still with CQ as the control cycle, the probabilities are around mid 60% or even lower. The reason may be that the longer observation time can lead to more accurate prediction for long-term cycle phase. In our statistics, three or four weeks are much longer time than three or four days; thus, they provide extra time to observe the current market trend and usually gives more accurate forecast.

Table 3.3.1 Three-week ud sequences statistics ("man-made" CQ as control cycle) for S&P futures during 01/2001 and 04/2006. Event *a* is a specific three-week ud sequence, and event A is a specific CQ cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CQ Up	98	0.636	uuu	32	0.208	0.327	1.000
			uud	20	0.149	0.204	0.871
			udu	20	0.143	0.204	0.908
			duu	14	0.11	0.143	0.827
			udd	4	0.084	0.041	0.311
			dud	4	0.097	0.041	0.269
			ddu	4	0.11	0.041	0.237
			ddd	0	0.097	0	0.000
CQ Down	56	0.364	uuu	0		0	0.000
			uud	3		0.054	0.132
			udu	2		0.036	0.092
			duu	3		0.054	0.179
			udd	9		0.161	0.697
			dud	11		0.196	0.735
			ddu	13		0.232	0.767
			ddd	15		0.268	1.005

Table 3.3.2 Four-week ud sequence statistics ("man-made" CQ as control cycle) for S&P futures during 01/2001 and 04/2006. Event *a* is a specific four-week ud sequence, and event A is a specific CQ cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CQ Up	81	0.664	uuuu	13	0.107	0.16	0.993
			uuud	12	0.098	0.148	1.003
			uudu	16	0.131	0.198	1.004
			uduu	11	0.09	0.136	1.003
			duuu	8	0.066	0.099	0.996
			uudd	1	0.016	0.012	0.498
			udud	4	0.033	0.049	0.986
			uddu	3	0.057	0.037	0.431
			duud	5	0.066	0.062	0.624
			dudu	3	0.041	0.037	0.599
			dduu	3	0.041	0.037	0.599
			uddd	0	0.016	0	0.000
			dudd	2	0.074	0.025	0.224
			ddud	0	0.057	0	0.000
			dddu	0	0.049	0	0.000
			dddd	0	0.057	0	0.000
CQ Down	41	0.336	uuuu	0		0	0.000
			uuud	0		0	0.000
			uudu	0		0	0.000
			uduu	0		0	0.000
			duuu	0		0	0.000
			uudd	1		0.024	0.504
			udud	0		0	0.000
			uddu	4		0.098	0.578
			duud	3		0.073	0.372
			dudu	2		0.049	0.402
			dduu	2		0.049	0.402
			uddd	2		0.049	1.029
			dudd	7		0.171	0.777
			ddud	7		0.171	1.008
			dddu	6		0.146	1.001
			dddd	7		0.171	1.008

Table 3.4.1 Three-week ud sequence statistics ("man-made" CY as control cycle) for S&P futures during 01/2001 and 04/2006. Event *a* is a specific three-week ud sequence, and event A is a specific CY cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CY Up	169	0.725	uuu	38	0.18	0.225	0.907
			uud	30	0.146	0.178	0.884
			udu	28	0.137	0.166	0.879
			duu	29	0.142	0.172	0.879
			udd	12	0.099	0.071	0.520
			dud	11	0.099	0.065	0.476
			ddu	13	0.107	0.077	0.522
			ddd	8	0.09	0.047	0.379
CY Down	64	0.275	uuu	4		0.063	0.096
			uud	4		0.063	0.119
			udu	4		0.063	0.126
			duu	4		0.063	0.122
			udd	11		0.172	0.477
			dud	12		0.188	0.522
			ddu	12		0.188	0.483
			ddd	13		0.203	0.620

Table 3.4.2 Four-week ud sequence statistics ("man-made" CY as control cycle) for S&P futures during 01/2001 and 04/2006. Event *a* is a specific four-week ud sequence, and event A is a specific CY cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CY Up	163	0.738	uuuu	19	0.095	0.117	0.908
			uuud	15	0.077	0.092	0.881
			uudu	22	0.104	0.135	0.957
			uduu	18	0.086	0.11	0.943
			duuu	14	0.072	0.086	0.881
			uudd	7	0.041	0.043	0.774
			udud	8	0.05	0.049	0.723
			uddu	6	0.059	0.037	0.463
			duud	15	0.077	0.092	0.881
			dudu	6	0.041	0.037	0.666
			dduu	11	0.054	0.067	0.915
			uddd	5	0.041	0.031	0.558
			dudd	5	0.063	0.031	0.363
			ddud	3	0.05	0.018	0.266
			dddu	6	0.05	0.037	0.546
			dddd	3	0.041	0.018	0.324
CY Down	58	0.262	uuuu	2		0.034	0.094
			uuud	2		0.034	0.116
			uudu	1		0.017	0.043
			uduu	1		0.017	0.052
			duuu	2		0.034	0.124
			uudd	2		0.034	0.218
			udud	3		0.052	0.273
			uddu	7		0.121	0.538
			duud	2		0.034	0.116
			dudu	3		0.052	0.333
			dduu	1		0.017	0.083
			uddd	4		0.069	0.442
			dudd	9		0.155	0.646
			ddud	8		0.138	0.724
			dddu	5		0.086	0.451
			dddd	6		0.103	0.659

3.3.3 Short-Term Cycle Leg-Length-Ratio Sequence Statistics

As before, in this study also, CQ and CY are considered as long-term cycles, whose phases can be viewed as the market trend background for which Bayesian guidance is sought while sequences of the leg-length-ratios of the short-term cycles C2 and C5 are used as the observables providing this guidance. Based on their standard cycle lengths, generally, a CY consists of four CQ; a CQ comprises 13 C5; a C5 includes 2 C2. Meanwhile, the cycle top of CY is also a cycle top of some CQ, C5 and C2; the same holds for the cycle valley. Therefore, long-term and short-term cycles have a certain connection, and the long-term cycle will affect the development of the short-term cycle. For example, in an up-trending market (CY up-leg), the short-term cycle (C2 or C5) is usually skewed to a situation with longer up-leg and shorter down-leg.

Using the same terms as before, event A is the phase of a long-term cycle, CQ or CY; meanwhile, event *a* comes from an observation on a short-term cycle, C2 or C5, as explained in the following. As we know, every cycle has an up-leg and down-leg, and the relative length relation between them is an observable property, which can be used as a short-term market characteristics. For the purpose of consistency, u is assigned to a cycle when the cycle up-leg length is longer than down-leg length; and d is assigned to a cycle when the up-leg length is shorter than down-leg length. Therefore, each short-term cycle is labeled as either u or d based on the intrinsic length relation between its up-leg and down-leg. Like daily and weekly ud sequences, three-cycle and four-cycle leg-length-ratio sequences can be obtained by assigning the corresponding label to each cycle. For example, uudd, a four-cycle leg-length-ratio sequence, means that for four contiguous cycles, the first two cycles have longer up-legs than down-legs, and the following two cycles have shorter up-legs than down-legs.

Using similar data acquisition and calculation methodology as in daily and weekly ud statistics, we obtain cycle leg-length-ratio statistics, with the working cycle as C2 or C5 and the control cycle as CQ or CY, shown in Table 3.5.1, 3.5.2, 3.6.1, 3.6.2, 3.7.1, 3.7.2, 3.8.1 and 3.8.2. The difference between daily and weekly ud sequences statistics is that, in the latter statistics, ud represents a cycle leg length relation. Here, it is used to present C2 or C5's leg-length-ratio feature to infer the CQ or CY phase from it.

Based on the results shown in the following tables, generally, the probabilities of a correct call of the long-term cycle phase with C5 as the working cycle are higher that that with C2 as the working cycle. C5 is longer than C2, thus, when C5 is the underlying cycle, it can provide a longer time to form a better forecast than C2. Meanwhile, when the same working cycle is used, the probabilities of CQ phase recognition under a certain sequence are higher than those for CY. The reason may be that because CQ is a shorter cycle than CY, some specific sequences, i.e., uuu and ddd, are more likely stay in the same for CQ than for CY.

Table 3.5.1 Three-cycle leg-length-ratio sequence statistics ("man-made" C2 as working cycle, "man-made" CQ as control cycle) for S&P futures during 01/2001 and 04/2006. Event *a* is a specific three-cycle leg-length-ratio sequence, and event A is a specific CQ cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CQ Up	260	0.577	uuu	43	0.124	0.165	0.767
			uud	42	0.133	0.162	0.702
			udu	44	0.153	0.169	0.637
			duu	45	0.137	0.173	0.728
			udd	23	0.102	0.088	0.497
			dud	24	0.135	0.092	0.393
			ddu	24	0.12	0.092	0.442
			ddd	15	0.095	0.058	0.352
CQ Down	191	0.423	uuu	13		0.068	0.232
			uud	18		0.094	0.299
			udu	25		0.131	0.363
			duu	17		0.089	0.275
			udd	23		0.12	0.498
			dud	37		0.194	0.609
			ddu	30		0.157	0.554
			ddd	28		0.147	0.655

Table 3.5.2 Four-cycle leg-length-ratio sequence statistics (“man-made” C2 as working cycle, “man-made” CQ as control cycle) for S&P futures during 01/2001 and 04/2006. Event *a* is a specific four-cycle leg-length-ratio sequence, and event A is a specific CQ cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CQ Up	245	0.585	uuuu	18	0.05	0.073	0.854
			uuud	20	0.072	0.082	0.666
			uudu	22	0.067	0.09	0.785
			uduu	31	0.093	0.127	0.798
			duuu	22	0.072	0.09	0.731
			uudd	18	0.062	0.073	0.688
			udud	11	0.062	0.045	0.424
			uddu	14	0.06	0.057	0.555
			duud	20	0.064	0.082	0.749
			dudu	19	0.088	0.078	0.518
			dduu	11	0.048	0.045	0.548
			uddd	9	0.041	0.037	0.528
			dudd	5	0.043	0.02	0.272
			ddud	11	0.074	0.045	0.356
			dddu	9	0.06	0.037	0.361
			dddd	5	0.045	0.02	0.260
CQ Down	174	0.415	uuuu	3		0.017	0.141
			uuud	10		0.057	0.329
			uudu	6		0.034	0.211
			uduu	8		0.046	0.205
			duuu	8		0.046	0.265
			uudd	8		0.046	0.308
			udud	15		0.086	0.576
			uddu	11		0.063	0.436
			duud	7		0.04	0.260
			dudu	18		0.103	0.486
			dduu	9		0.052	0.450
			uddd	8		0.046	0.466
			dudd	13		0.075	0.724
			ddud	20		0.115	0.645
			dddu	16		0.092	0.637
			dddd	14		0.08	0.738

Table 3.6.1 Three-cycle leg-length-ratio sequence statistics ("man-made" C2 as working cycle, "man-made" CY as control cycle) for S&P futures during 01/2001 and 04/2006. Event *a* is a specific three-cycle leg-length-ratio sequence, and event A is a specific CY cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CY Up	362	0.688	uuu	48	0.116	0.133	0.789
			uud	52	0.129	0.144	0.768
			udu	55	0.144	0.152	0.726
			duu	55	0.135	0.152	0.775
			udd	40	0.114	0.11	0.664
			dud	43	0.135	0.119	0.607
			ddu	42	0.127	0.116	0.629
			ddd	27	0.099	0.075	0.521
CY Down	164	0.312	uuu	13		0.079	0.212
			uud	16		0.098	0.237
			udu	21		0.128	0.277
			duu	16		0.098	0.226
			udd	20		0.122	0.334
			dud	28		0.171	0.395
			ddu	25		0.152	0.373
			ddd	25		0.152	0.479

Table 3.6.2 Four-cycle leg-length-ratio sequence statistics ("man-made" C2 as working cycle, "man-made" CY as control cycle) for S&P during 01/2001 and 04/2006. Event *a* is a specific four-cycle leg-length-ratio sequence, and event A is a specific CY cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CY Up	358	0.694	uuuu	21	0.048	0.059	0.853
			uuud	25	0.066	0.07	0.736
			uudu	26	0.064	0.073	0.791
			uduu	37	0.085	0.103	0.841
			duuu	27	0.068	0.075	0.765
			uudd	25	0.064	0.07	0.759
			udud	18	0.06	0.05	0.578
			uddu	23	0.068	0.064	0.653
			duud	26	0.064	0.073	0.791
			dudu	28	0.081	0.078	0.668
			dduu	18	0.052	0.05	0.667
			uddd	17	0.05	0.047	0.652
			dudd	15	0.05	0.042	0.583
			ddud	24	0.076	0.067	0.612
			dddu	19	0.06	0.053	0.613
			dddd	9	0.043	0.025	0.403
CY Down	158	0.306	uuuu	4		0.025	0.159
			uuud	9		0.057	0.264
			uudu	7		0.044	0.211
			uduu	7		0.044	0.159
			duuu	8		0.051	0.230
			uudd	8		0.051	0.244
			udud	13		0.082	0.418
			uddu	12		0.076	0.342
			duud	7		0.044	0.211
			dudu	14		0.089	0.336
			dduu	9		0.057	0.336
			uddd	9		0.057	0.349
			dudd	11		0.07	0.429
			ddud	15		0.095	0.383
			dddu	12		0.076	0.388
			dddd	13		0.082	0.584

Table 3.7.1 Three-cycle leg-length-ratio sequence statistics ("man-made" C5 as working cycle, "man-made" CQ as control cycle) for S&P futures during 01/2001 and 04/2006. Event *a* is a specific three-cycle leg-length-ratio sequence, and event A is a specific CQ cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CQ Up	92	0.634	uuu	24	0.166	0.261	0.998
			uud	12	0.103	0.13	0.801
			udu	13	0.097	0.141	0.922
			duu	10	0.083	0.109	0.833
			udd	11	0.124	0.12	0.614
			dud	10	0.117	0.109	0.591
			ddu	10	0.138	0.109	0.501
			ddd	2	0.172	0.022	0.081
CQ Down	53	0.366	uuu	0		0	0.000
			uud	3		0.057	0.202
			udu	1		0.019	0.072
			duu	2		0.038	0.167
			udd	7		0.132	0.389
			dud	7		0.132	0.412
			ddu	10		0.189	0.501
			ddd	23		0.434	0.922

Table 3.7.2 Four-cycle leg-length-ratio sequence statistics ("man-made" C5 as working cycle, "man-made" CQ as control cycle) for S&P during 01/2001 and 04/2006. Event *a* is a specific four-cycle leg-length-ratio sequence, and event A is a specific CQ cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CQ Up	77	0.67	uuuu	14	0.122	0.182	0.999
			uuud	6	0.052	0.078	1.004
			uudu	5	0.043	0.065	1.012
			uduu	4	0.035	0.052	0.995
			duuu	4	0.035	0.052	0.995
			uudd	4	0.061	0.052	0.571
			udud	5	0.052	0.065	0.837
			uddu	9	0.096	0.117	0.816
			duud	3	0.043	0.039	0.607
			dudu	5	0.052	0.065	0.837
			dduu	5	0.043	0.065	1.012
			uddd	2	0.035	0.026	0.497
			dudd	6	0.087	0.078	0.600
			ddud	4	0.078	0.052	0.446
			dddu	1	0.061	0.013	0.143
			dddd	0	0.104	0	0.000
CQ Down	38	0.33	uuuu	0		0	0.000
			uuud	0		0	0.000
			uudu	0		0	0.000
			uduu	0		0	0.000
			duuu	0		0	0.000
			uudd	3		0.079	0.428
			udud	1		0.026	0.165
			uddu	2		0.053	0.182
			duud	2		0.053	0.407
			dudu	1		0.026	0.165
			dduu	0		0	0.000
			uddd	2		0.053	0.500
			dudd	4		0.105	0.399
			ddud	5		0.132	0.559
			dddu	6		0.158	0.856
			dddd	12		0.316	1.004

Table 3.8.1 Three-cycle leg-length-ratio sequence statistics (“man-made” C5 as working cycle, “man-made” CY as control cycle) for S&P futures during 01/2001 and 04/2006. Event *a* is a specific three-cycle leg-length-ratio sequence, and event A is a specific CY cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CY Up	168	0.73	uuu	32	0.148	0.19	0.938
			uud	18	0.1	0.107	0.782
			udu	17	0.083	0.101	0.889
			duu	19	0.104	0.113	0.794
			udd	22	0.139	0.131	0.688
			dud	21	0.126	0.125	0.725
			ddu	23	0.143	0.137	0.700
			ddd	16	0.157	0.095	0.442
CY Down	62	0.27	uuu	2		0.032	0.058
			uud	5		0.081	0.218
			udu	2		0.032	0.104
			duu	5		0.081	0.210
			udd	10		0.161	0.312
			dud	8		0.129	0.276
			ddu	10		0.161	0.303
			ddd	20		0.323	0.555

Table 3.8.2 Four-cycle leg-length-ratio sequence statistics ("man-made" C5 as working cycle, "man-made" CY as control cycle) for S&P during 01/2001 and 04/2006. Event *a* is a specific three-cycle leg-length-ratio sequence, and event A is a specific CY cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CY Up	163	0.744	uuuu	19	0.087	0.117	1.001
			uuud	10	0.055	0.061	0.825
			uudu	6	0.032	0.037	0.861
			uduu	5	0.027	0.031	0.855
			duuu	12	0.064	0.074	0.861
			uudd	10	0.064	0.061	0.709
			udud	11	0.055	0.067	0.907
			uddu	14	0.082	0.086	0.781
			duud	7	0.046	0.043	0.696
			dudu	9	0.046	0.055	0.890
			dduu	13	0.064	0.08	0.930
			uddd	8	0.059	0.049	0.618
			dudd	12	0.082	0.074	0.672
			ddud	10	0.078	0.061	0.582
			dddu	9	0.064	0.055	0.640
			dddd	8	0.096	0.049	0.380
CY Down	56	0.256	uuuu	0		0	0.000
			uuud	2		0.036	0.167
			uudu	1		0.018	0.144
			uduu	1		0.018	0.170
			duuu	2		0.036	0.144
			uudd	4		0.071	0.284
			udud	1		0.018	0.084
			uddu	4		0.071	0.221
			duud	3		0.054	0.300
			dudu	1		0.018	0.100
			dduu	1		0.018	0.072
			uddd	5		0.089	0.386
			dudd	6		0.107	0.334
			ddud	7		0.125	0.410
			dddu	5		0.089	0.356
			dddd	13		0.232	0.618

3.3.4 Short-Term Cycle Extreme Up-or-Down (ud) Sequence Statistics

In section 3.3.3, one of the cycle features, the cycle leg-length-ratio property, was used as a market characteristic to infer the long-term market trend. Similar to this ratio, another cycle feature, ud, can also be employed to represent a present-market circumstance. Instead of comparing the time lengths of cycle up-leg and own-leg, we compare two continuous cycles' price variation. For the purpose of consistency, u is assigned to a cycle when the price of previous cycle valley is lower than the price of current cycle valley; conversely, d represents a cycle whose cycle valley price is lower than that of previous cycle. Therefore, three-cycle and four-cycle ud sequences for the test period can be obtained by assigning each cycle to the corresponding label. For example, uddd, a four-cycle ud sequence, means that for four continuous cycles, the first cycle is a price-up cycle, and the following three cycles are price-down cycles.

Similar to the above sections, event A is still the long-term cycle phase, but event *a* is one of the continuous cycle ud sequences. Given the present cycle ud sequence, we can infer the probability of the current long-term cycle phase.

As in the above sections, we apply similar data acquisition and calculation methods to cycle ud statistics, with the working cycle as C2 or C5 and the control cycle as CQ or CY. The results are shown in Table 3.9.1, 3.9.2, 3.10.1, 3.10.2, 3.11.1, 3.11.2, 3.12.1 and 3.12.2, respectively.

As both cycle ud statistics and cycle leg-length-ratio statistics use short-term cycle features to represent current market characteristics, we have similar observations based on the results of this study. Generally, the probabilities of long-term cycle phase with C5 as the working cycle are higher than those with C2 as the working cycle; meantime, when same working cycle is used, the probabilities of CQ phase identification under certain sequences are higher than those for CY.

When comparing the probabilities of correct identification of a control cycle phase under a certain cycle sequence, generally, the preference in cycle ud statistics is a little bit stronger than that in cycle leg-length-ratio statistics. The reason may be that, the cycle price variation is affected more by control cycle phase than cycle leg length relation. Therefore, when certain cycle ud sequence appears, it points to control cycle phase identification with a larger confidence level than for same cycle leg-length-ratio sequence.

Table 3.9.1 Three-cycle ud sequence statistics ("man-made" C2 as working cycle, "man-made" CQ as control cycle) for S&P futures during 01/2001 and 04/2006. Event *a* is a specific three-cycle ud sequence, and event A is a specific CQ cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CQ Up	260	0.577	uuu	52	0.115	0.2	1.003
			uud	40	0.111	0.154	0.800
			udu	56	0.173	0.215	0.716
			duu	40	0.113	0.154	0.786
			udd	18	0.1	0.069	0.398
			dud	33	0.166	0.127	0.441
			ddu	18	0.113	0.069	0.352
			ddd	3	0.109	0.012	0.063
CQ Down	191	0.423	uuu	0		0	0.000
			uud	10		0.052	0.198
			udu	22		0.115	0.282
			duu	11		0.058	0.217
			udd	27		0.141	0.597
			dud	42		0.22	0.561
			ddu	33		0.173	0.648
			ddd	46		0.241	0.936

Table 3.9.2 Four-cycle ud sequence statistics ("man-made" C2 as working cycle, "man-made" CQ as control cycle) for S&P futures during 01/2001 and 04/2006. Event a is a specific four-cycle ud sequence, and event A is a specific CQ cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CQ Up	245	0.585	uuuu	21	0.05	0.086	1.006
			uuud	25	0.06	0.102	0.994
			uudu	31	0.084	0.127	0.884
			uduu	25	0.072	0.102	0.828
			duuu	24	0.057	0.098	1.005
			uudd	10	0.036	0.041	0.666
			udud	27	0.103	0.11	0.624
			uddu	15	0.057	0.061	0.626
			duud	14	0.057	0.057	0.585
			dudu	22	0.091	0.09	0.578
			dduu	12	0.043	0.049	0.666
			uddd	3	0.041	0.012	0.171
			dudd	7	0.062	0.029	0.274
			ddud	6	0.072	0.024	0.195
			dddu	3	0.06	0.012	0.117
			dddd	0	0.057	0	0.000
CQ Down	174	0.415	uuuu	0		0	0.000
			uuud	0		0	0.000
			uudu	4		0.023	0.114
			uduu	5		0.029	0.167
			duuu	0		0	0.000
			uudd	5		0.029	0.335
			udud	16		0.092	0.371
			uddu	9		0.052	0.379
			duud	10		0.057	0.415
			dudu	16		0.092	0.420
			dduu	6		0.034	0.328
			uddd	14		0.08	0.810
			dudd	19		0.109	0.730
			ddud	24		0.138	0.796
			dddu	22		0.126	0.872
			dddd	24		0.138	1.005

Table 3.10.1 Three-cycle ud sequence statistics ("man-made" C2 as working cycle, "man-made" CY as control cycle) for S&P futures during 01/2001 and 04/2006. Event *a* is a specific three-cycle ud sequence, and event A is a specific CY cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CY Up	362	0.688	uuu	52	0.112	0.144	0.885
			uud	48	0.118	0.133	0.776
			udu	67	0.163	0.185	0.781
			duu	48	0.118	0.133	0.776
			udd	34	0.108	0.094	0.599
			dud	53	0.156	0.146	0.644
			ddu	35	0.116	0.097	0.575
			ddd	25	0.108	0.069	0.440
CY Down	164	0.312	uuu	7		0.043	0.120
			uud	14		0.085	0.225
			udu	19		0.116	0.222
			duu	14		0.085	0.225
			udd	23		0.14	0.404
			dud	29		0.177	0.354
			ddu	26		0.159	0.427
			ddd	32		0.195	0.563

Table 3.10.2 Four-cycle ud sequence statistics ("man-made" C2 as working cycle, "man-made" CY as control cycle) for S&P futures during 01/2001 and 04/2006. Event *a* is a specific four-cycle ud sequence, and event A is a specific CY cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CY Up	358	0.694	uuuu	23	0.05	0.064	0.888
			uuud	29	0.064	0.081	0.878
			uudu	30	0.074	0.084	0.788
			uduu	30	0.07	0.084	0.833
			duuu	27	0.06	0.075	0.867
			uudd	18	0.045	0.05	0.771
			udud	35	0.093	0.098	0.731
			uddu	20	0.056	0.056	0.694
			duud	19	0.056	0.053	0.657
			dudu	36	0.091	0.101	0.770
			dduu	18	0.05	0.05	0.694
			uddd	14	0.052	0.039	0.520
			dudd	15	0.062	0.042	0.470
			ddud	18	0.064	0.05	0.542
			dddu	15	0.06	0.042	0.486
			dddd	11	0.052	0.031	0.414
Down	158	0.306	uuuu	3		0.019	0.116
			uuud	4		0.025	0.120
			uudu	8		0.051	0.211
			uduu	6		0.038	0.166
			duuu	4		0.025	0.128
			uudd	5		0.032	0.218
			udud	13		0.082	0.270
			uddu	9		0.057	0.312
			duud	10		0.063	0.344
			dudu	11		0.07	0.236
			dduu	8		0.051	0.312
			uddd	13		0.082	0.483
			dudd	17		0.108	0.533
			ddud	15		0.095	0.455
			dddu	16		0.101	0.515
			dddd	16		0.101	0.595

Table 3.11.1 Three-cycle ud sequence statistics ("man-made" C5 as working cycle, "man-made" CQ as control cycle) for S&P futures during 01/2001 and 04/2006. Event *a* is a specific three-cycle ud sequence, and event A is a specific CQ cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CQ Up	92	0.634	uuu	26	0.179	0.283	1.003
			uud	14	0.11	0.152	0.877
			udu	19	0.145	0.207	0.906
			duu	13	0.103	0.141	0.869
			udd	6	0.103	0.065	0.400
			dud	8	0.131	0.087	0.421
			ddu	5	0.097	0.054	0.353
			ddd	1	0.131	0.011	0.053
CQ Down	53	0.366	uuu	0		0	0.000
			uud	2		0.038	0.126
			udu	2		0.038	0.096
			duu	2		0.038	0.135
			udd	9		0.17	0.603
			dud	11		0.208	0.580
			ddu	9		0.17	0.641
			ddd	18		0.34	0.949

Table 3.11.2 Four-cycle ud sequence statistics ("man-made" C5 as working cycle, "man-made" CQ as control cycle) for S&P futures during 01/2001 and 04/2006. Event *a* is a specific four-cycle ud sequence, and event A is a specific CQ cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CQ Up	77	0.67	uuuu	13	0.113	0.169	1.001
			uuud	7	0.061	0.091	0.999
			uudu	10	0.087	0.13	1.000
			uduu	10	0.087	0.13	1.000
			duuu	6	0.052	0.078	1.004
			uudd	2	0.026	0.026	0.670
			udud	6	0.07	0.078	0.746
			uddu	4	0.052	0.052	0.670
			duud	4	0.052	0.052	0.670
			dudu	6	0.07	0.078	0.746
			dduu	2	0.035	0.026	0.497
			uddd	1	0.026	0.013	0.335
			dudd	3	0.087	0.039	0.300
			ddud	2	0.043	0.026	0.405
			dddu	1	0.061	0.013	0.143
			dddd	0	0.078	0	0.000
CQ Down	38	0.33	uuuu	0		0	0.000
			uuud	0		0	0.000
			uudu	0		0	0.000
			uduu	0		0	0.000
			duuu	0		0	0.000
			uudd	1		0.026	0.330
			udud	2		0.053	0.250
			uddu	2		0.053	0.337
			duud	2		0.053	0.337
			dudu	2		0.053	0.250
			dduu	2		0.053	0.500
			uddd	2		0.053	0.674
			dudd	7		0.184	0.699
			ddud	3		0.079	0.607
			dddu	6		0.158	0.856
			dddd	9		0.237	1.004

Table 3.12.1 Three-cycle ud sequence statistics ("man-made" C5 as working cycle, "man-made" CY as control cycle) for S&P futures during 01/2001 and 04/2006. Event *a* is a specific three-cycle ud sequence, and event A is a specific CY cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CY Up	168	0.73	uuu	36	0.157	0.214	0.996
			uud	23	0.113	0.137	0.886
			udu	24	0.126	0.143	0.829
			duu	24	0.117	0.143	0.893
			udd	17	0.117	0.101	0.631
			dud	16	0.13	0.095	0.534
			ddu	16	0.122	0.095	0.569
			ddd	12	0.117	0.071	0.443
CY Down	62	0.27	uuu	0		0	0.000
			uud	3		0.048	0.115
			udu	5		0.081	0.173
			duu	3		0.048	0.111
			udd	10		0.161	0.371
			dud	14		0.226	0.469
			ddu	12		0.194	0.429
			ddd	15		0.242	0.558

Table 3.12.2 Four-cycle ud sequence statistics ("man-made" C5 as working cycle, "man-made" CY as control cycle) for S&P futures during 01/2001 and 04/2006. Event *a* is a specific four-cycle ud sequence, and event A is a specific CY cycle phase.

Cycle Phase	Number	P(A)	Sequence	Number	P(*a*)	P(*a*|A)	P(A|*a*)
CY Up	163	0.744	uuuu	20	0.091	0.123	1.006
			uuud	13	0.059	0.08	1.009
			uudu	13	0.059	0.08	1.009
			uduu	13	0.064	0.08	0.930
			duuu	14	0.064	0.086	1.000
			uudd	9	0.055	0.055	0.744
			udud	9	0.059	0.055	0.694
			uddu	9	0.064	0.055	0.640
			duud	10	0.059	0.061	0.770
			dudu	9	0.064	0.055	0.640
			dduu	10	0.055	0.061	0.825
			uddd	7	0.055	0.043	0.582
			dudd	8	0.068	0.049	0.536
			ddud	7	0.068	0.043	0.471
			dddu	7	0.055	0.043	0.582
			dddd	5	0.059	0.031	0.391
CY Down	56	0.256	uuuu	0		0	0.000
			uuud	0		0	0.000
			uudu	0		0	0.000
			uduu	1		0.018	0.072
			duuu	0		0	0.000
			uudd	3		0.054	0.251
			udud	4		0.071	0.308
			uddu	5		0.089	0.356
			duud	3		0.054	0.234
			dudu	5		0.089	0.356
			dduu	2		0.036	0.167
			uddd	5		0.089	0.414
			dudd	7		0.125	0.470
			ddud	8		0.143	0.538
			dddu	5		0.089	0.414
			dddd	8		0.143	0.620

3.4 Use of Phase Identification Results Obtained by Application of Bayesian Statistics in Market Cycle Analysis

In Chapter 2, we introduced the TFx methods, which are cycle-based pattern recognition methods. The basic idea is that, based on the assumed similarity of some historical cycles with the present cycle, k cohort members are selected to provide guidance as to current cycle movement. We conducted the optimization test for the TFx methods with and without use of a "man-made" and therefore correctly phased control cycles. We found that use of a cycle background pre-selected by correct control cycle would significantly improve selection of cohort members, and trading results.

As discussed in Section 3.2, by using Bayes' Theorem we can infer longer-term market action (longer-term cycle phases) based on present, observable short-term market characteristics. In Section 3.3, four different kinds of such present observable short-term market characteristics are introduced, namely daily ud sequences, weekly ud sequences, short-term cycle leg-length-ratio sequences, and short-term cycle extreme ud sequences, and these corresponding Bayesian statistics are presented. Based on the statistics, we can have certain levels confidence (in terms of probability) in the correctness of the inferred long-term cycle phase forecasts.

We can now apply the inferred long-cycle phases to the TFx methods as background. For instance, if based on a present weekly ud sequence, we predict that present market is in a CQ up-leg period, then, cohort members for C2 or C5 can be chosen from the historical CQ cycle up-leg cycle database, instead of the historical whole cycle database. If there is no strong preference for a current long-term cycle phase, the historical whole cycle database is still used as the background. Hence, long-term cycle information provides some refinement to cohort member selection.

The following sections present the trading results for the TFx methods, where long-term cycle derived from Bayesian statistics is applied in the background pre-selection.

3.4.1 Trading Results with Use of Daily ud Sequence Statistics

In this section, we investigate the effect of an inferred long-term cycle phase to short-term cycle (C2 or C5) trading with TFx as the trading methods. Instead of using a "man-made" control cycle, we apply daily ud sequence statistics in the determination of control cycle phase. Based on data from Section 3.1.1, for a certain daily ud sequence, we can obtain the probability of a certain control cycle phase. For example, when the present daily sequence is uuu, the probability of being in a CQ up-leg period is 81.4%. Thereupon, it gives us some preference for long-term cycle phase and for the use of the CQ-up database in TFx cycle leg analysis for C2 and C5.

We now apply the same trading procedures described in Chapter 2 in the optimization of the TFx methods, with use of the control cycle predicted by daily ud sequence statistics. In doing this, we must set limits for use of probabilities applied in the study. Since there are different probabilities for different daily ud sequence, in the test, only probabilities over 60% are chosen as an indication of control cycle phase. When a certain daily ud sequence, whose $P(A|a)$ value is larger than 60%, occurs in present market, we will pre-select the background based on the suggested long-term cycle phase. Otherwise, we still use the whole database as the background.

Take a three-day ud sequence as an example, and refer to Table 3.1.1. First, the background is separated into two parts: CQ up and CQ down. When the present three-day ud sequence is one of the four sequences uuu, uud, udu or duu, we infer that the present market is in a CQ up period, thus, only CQ up part of the database is used as the background. Likewise, when the present three-day ud sequence is ddd, which suggests it is in a CQ down period, then, the CQ down part is used as the background. When other sequences than the above-mentioned appear, the whole background will be utilized because of the uncertainty of the control cycle phase.

Table 3.13.1 presents the optimized trading results of simulated real time C2 cycle testing of S&P 500 futures using the TFx methods with cycle database during 01/2001 and 04/2006. The CQ phase is predicted by the daily ud sequence statistics, as discussed.

In the table, the "Long" column represents the profit-loss for all long-side trades, and the "Short" column means all short-side trades. The "Total" column, also called CPL (cumulative profit-loss), shows the sum of long and short trades, which is used as the performance measure for the TFx methods. The best passing rates for the individual TFx methods are also shown in the table.

Table 3.13.1 Optimized trading results for simulated real time C2 cycle testing of S&P 500 futures using the TFx methods (CQ predicted by daily ud sequences as control cycle) with cycle database during 01/2001 and 04/2006

Method	Best Passing Rate (%)	Long	Short	Total
TF1	60	375.9	407.2	783.1
TF2	50	372.7	419.9	792.6
TF3	60	316.7	359.3	676
TF4	40	186.6	262.4	449
TF5	50	89	150.9	239.9
TV1	40	590.4	647.5	1237.9
TF2B	60	275.5	307.7	583.2

Table 3.13.2 shows the optimized trading results of simulated real time C2 cycle testing of S&P 500 futures using the TFx methods with cycle database during 01/2001 and 04/2006. The CY cycle phase is predicted by daily ud sequence statistics.

Table 3.13.3 and 3.13.4 display the optimized trading results of simulated real time C5 cycle testing of S&P 500 futures using the TFx methods with cycle database during 01/2001 and 04/2006. Again, CQ and CY cycle phases are predicted by daily ud sequence statistics.

A discussion of these data is deferred to Section 3.4.5 when they are reviewed together with data based on the other short-term (present-time) Bayesian indicators.

Table 3.13.2 Optimized trading results for simulated real time C2 cycle testing of S&P 500 futures using the TFx methods (CY predicted by daily ud sequences as control cycle) with cycle database during 01/2001 and 04/2006

Method	Best Passing Rate (%)	Long	Short	Total
TF1	60	627.6	658	1285.6
TF2	60	432	476.2	908.2
TF3	40	376.3	420	796.3
TF4	40	257	334.2	591.2
TF5	40	300.7	367.4	668.1
TV1	40	514	595.2	1109.2
TF2B	60	421.5	450.3	871.8

Table 3.13.3 Optimized trading results for simulated real time C5 cycle testing of S&P 500 futures using the TFx methods (CQ predicted by daily ud sequences as control cycle) with cycle database during 01/2001 and 04/2006

Method	Best Passing Rate (%)	Long	Short	Total
TF1	70	219.8	303.1	522.9
TF2	60	439.8	523.7	963.5
TF3	30	239.3	286.7	526
TF4	50	120.1	179.3	299.4
TF5	30	179.3	197.3	376.6
TV1	30	314.5	357.7	672.2
TF2B	80	547.5	589.7	1137.2

Table 3.13.4 Optimized trading results for simulated real time C5 cycle testing of S&P 500 futures using the TFx methods (CY predicted by daily ud sequences as control cycle) with cycle database during 01/2001 and 04/2006

Method	Best Passing Rate (%)	Long	Short	Total
TF1	70	222.3	316.9	539.2
TF2	60	303	362.6	665.6
TF3	30	298.3	377.1	675.4
TF4	50	282.8	344.5	627.3
TF5	40	164.6	207.9	372.5
TV1	30	491.5	540.9	1032.4
TF2B	40	538.6	538.2	1076.8

3.4.2 Trading Results with Use of Weekly ud Sequence Statistics

In this section, we apply weekly ud sequence statistics to the determination of the control cycle phase. The same trading procedures as described in the previous section have been used in the optimization test of the TFx methods, with the trading cycle as C2 or C5 and the control cycle as CQ or CY. The results are presented in Table 3.14.1, 3.14.2, 3.14.3 and 3.14.4, respectively.

Table 3.14.1 Optimized trading results for simulated real time C2 cycle testing of S&P 500 futures using TFx methods (CQ predicted by weekly ud sequences as control cycle) with cycle database during 01/2001 and 04/2006

Method	Best Passing Rate (%)	Long	Short	Total
TF1	60	403.7	446.1	849.8
TF2	50	497.7	545.6	1043.3
TF3	60	429.2	493.5	922.7
TF4	30	351.6	412	763.6
TF5	30	410.6	485.3	895.9
TV1	60	359.8	407.3	767.1
TF2B	50	782.1	836.4	1613.8

Table 3.14.2 Optimized trading results for simulated real time C2 cycle testing of S&P 500 futures using TFx methods (CY predicted by weekly ud sequences as control cycle) with cycle database during 01/2001 and 04/2006

Method	Best Passing Rate (%)	Long	Short	Total
TF1	60	520.7	566	1086.7
TF2	60	424.9	479.4	904.3
TF3	40	305.1	358.5	663.6
TF4	40	240.2	318.4	558.6
TF5	40	325.5	387.8	713.3
TV1	40	664.9	740.5	1405.4
TF2B	60	196	214.1	410.1

Table 3.14.3 Optimized trading results for simulated real time C5 cycle testing of S&P 500 futures using TFx methods (CQ predicted by weekly ud sequences as control cycle) with cycle database during 01/2001 and 04/2006

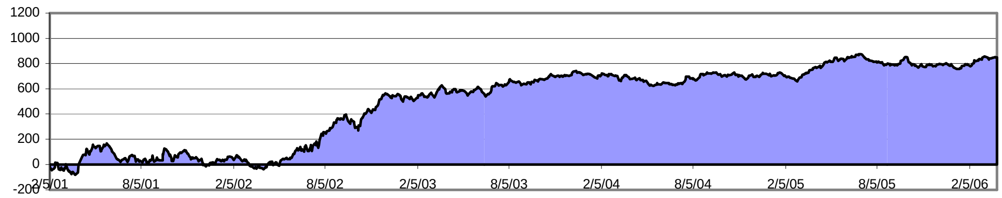
Method	Best Passing Rate (%)	Long	Short	Total
TF1	70	368.2	459.2	827.4
TF2	60	524.6	521.5	1046.1
TF3	60	264.2	295.7	559.9
TF4	40	335.7	367.5	703.2
TF5	60	387.7	452.8	840.5
TV1	70	464.3	504.5	968.8
TF2B	50	316.9	304.3	621.2

Table 3.14.4 Optimized trading results for simulated real time C5 cycle testing of S&P 500 futures using TFx methods (CY predicted by weekly ud sequences as control cycle) with cycle database during 01/2001 and 04/2006

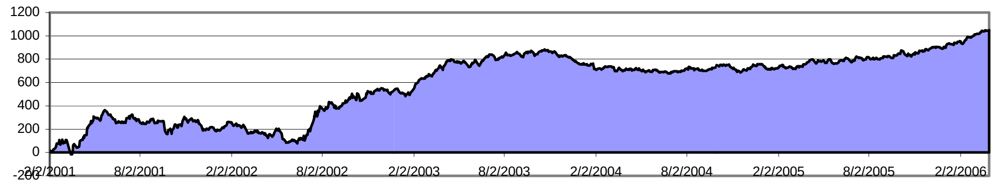
Method	Best Passing Rate (%)	Long	Short	Total
TF1	60	356.1	400	756.1
TF2	60	509.9	587.2	1097.1
TF3	30	227.4	268.3	495.7
TF4	50	246	300.6	546.6
TF5	50	361.9	419.6	781.5
TV1	30	464.1	540.4	1004.5
TF2B	40	476.5	502.3	978.8

Figure 3.2 presents histograms of cumulative profits of all TFx methods when weekly ud sequence statistics are used to infer the CQ cycle phase, also showing the corresponding S&P market. According to the graph, the TF2B method is seen to work well in most of the market and has the best performance. For other methods, the down-trending market (up to 05/2002) is not so amenable to them. It seems like most methods work best during the bottoming period (07/2002~02/2003) and steadily climbing markets (08/2002~08/2003 and 05/2005~04/2006); they fail to be profitable during the intervening, stagnating period. The TF2 method made most of profit in mid and very last of the test period. The TF4 and TV1 methods are relatively least good performers.

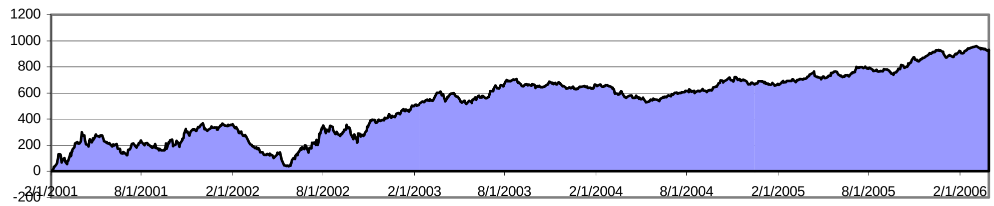
TF1



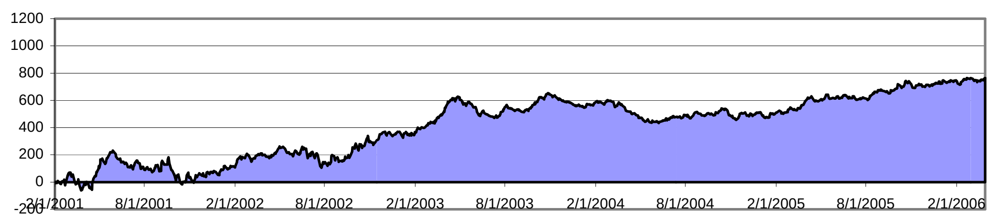
TF2



TF3



TF4



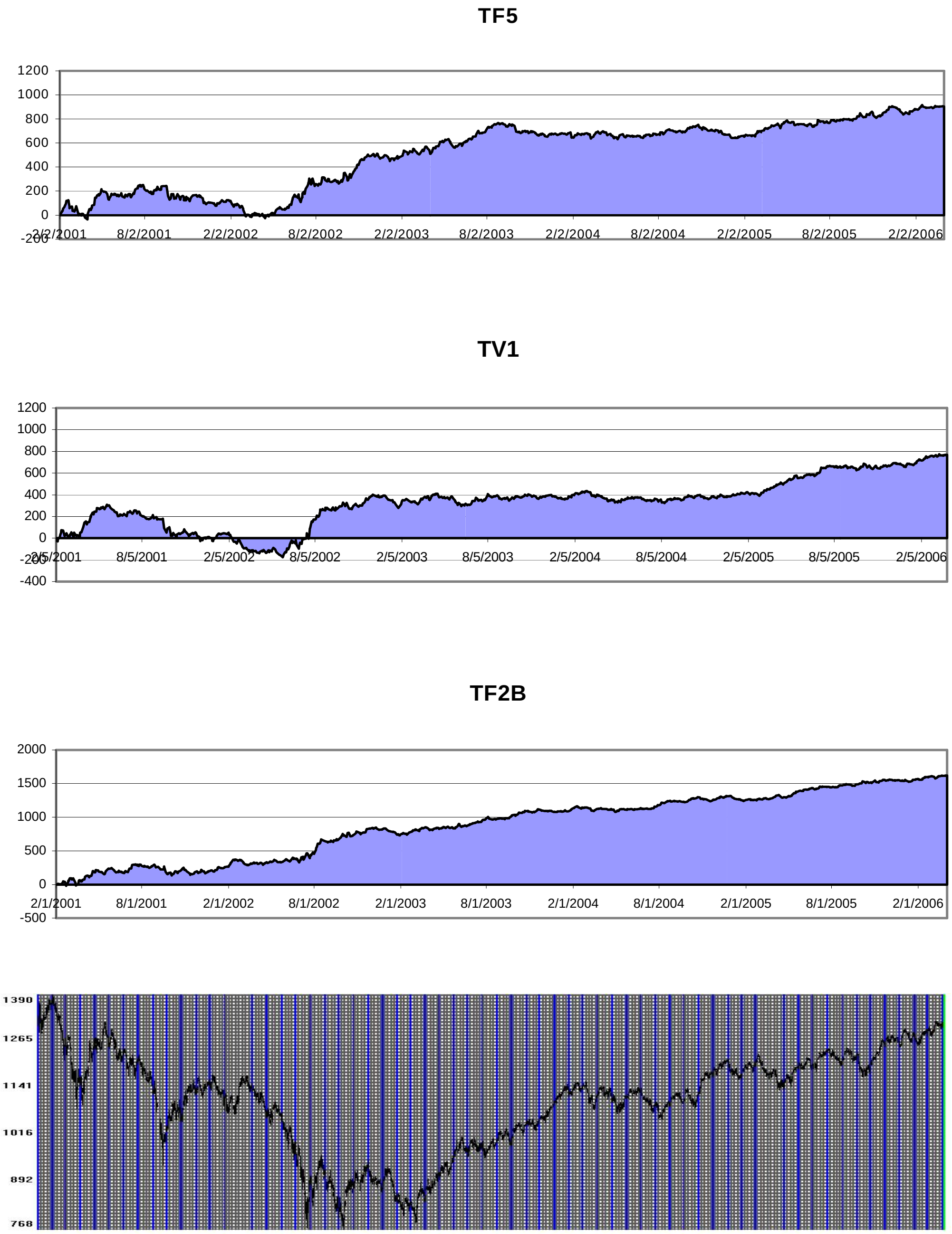


Figure 3.2 Histograms of TFx trading results with best passing rates under CQ control cycle predicted by weekly ud sequences statistics during 01/2001 to 04/2006 and S&P price chart at the same period. The graphs from top to bottom represent TF1, TF2, TF3, TF4, TF5, TV1 and TF2B, respectively.

Since the above results are for period during 01/2001 to 04/2006, I updated this study by applying the TFx methods with best passing rate (derived from the earlier period) to a recent market period from 06/2007 to 05/2008. Table 3.15 contains the simulated real-time trading results for all TFx methods. As a representative, the summary plot for the TF2 method and the price chart of that period are shown in Figure 3.3. Based on data in the table, all TFx methods make a profit and, as would be expected for an overall falling market, short trades perform better than long trades. Among the seven methods, the TF2 method performs best with a CPL over 700, and is also best in the other measures: FOM2, Net Profit/trade and Run-down. The TF5 and TF2B methods also perform well with CPLs of 498 and 542, respectively. Compared to the results for this recent period shown in Chapter 2 without use of a control cycle, the results (in S&P points / time) are improved for every TFx method except the TV1 method, showing that the effect of applying weekly ud sequence statistics to the TFx trading is mostly positive; therefore, this method can be used to help real-time trading.

Table 3.15 Results of simulated real time C2 cycle testing of S&P using the TFX methods with best passing rate under weekly ud sequence statistic during 06/2007 and 05/2008.

Method	CPL			TIM		FOM2		NetProfit/Trade		Run-Down	
	Long	Short	Total	Long	Short	Long	Short	Long	Short	Long	Short
TF1	77	266	343	0.57	0.41	0.52	0.61	0.7064	2.4404	139.25	96.25
TF2	259.25	449.5	708.75	0.55	0.42	0.58	0.66	1.9492	3.3545	113	91.5
TF3	78.5	262	340.5	0.50	0.47	0.53	0.59	0.6331	2.096	134.5	87.25
TF4	-71.5	122.5	51	0.42	0.55	0.48	0.53	-0.3341	0.5724	239	108.25
TF5	151.25	347.25	498.5	0.51	0.46	0.55	0.63	1.022	2.3305	114.25	67.5
TV1	-42.75	151.5	108.75	0.48	0.49	0.48	0.55	-0.4071	1.4292	148.5	110.75
TF2B	170.75	371.75	542.5	0.52	0.45	0.56	0.65	1.377	2.974	128.75	84.25

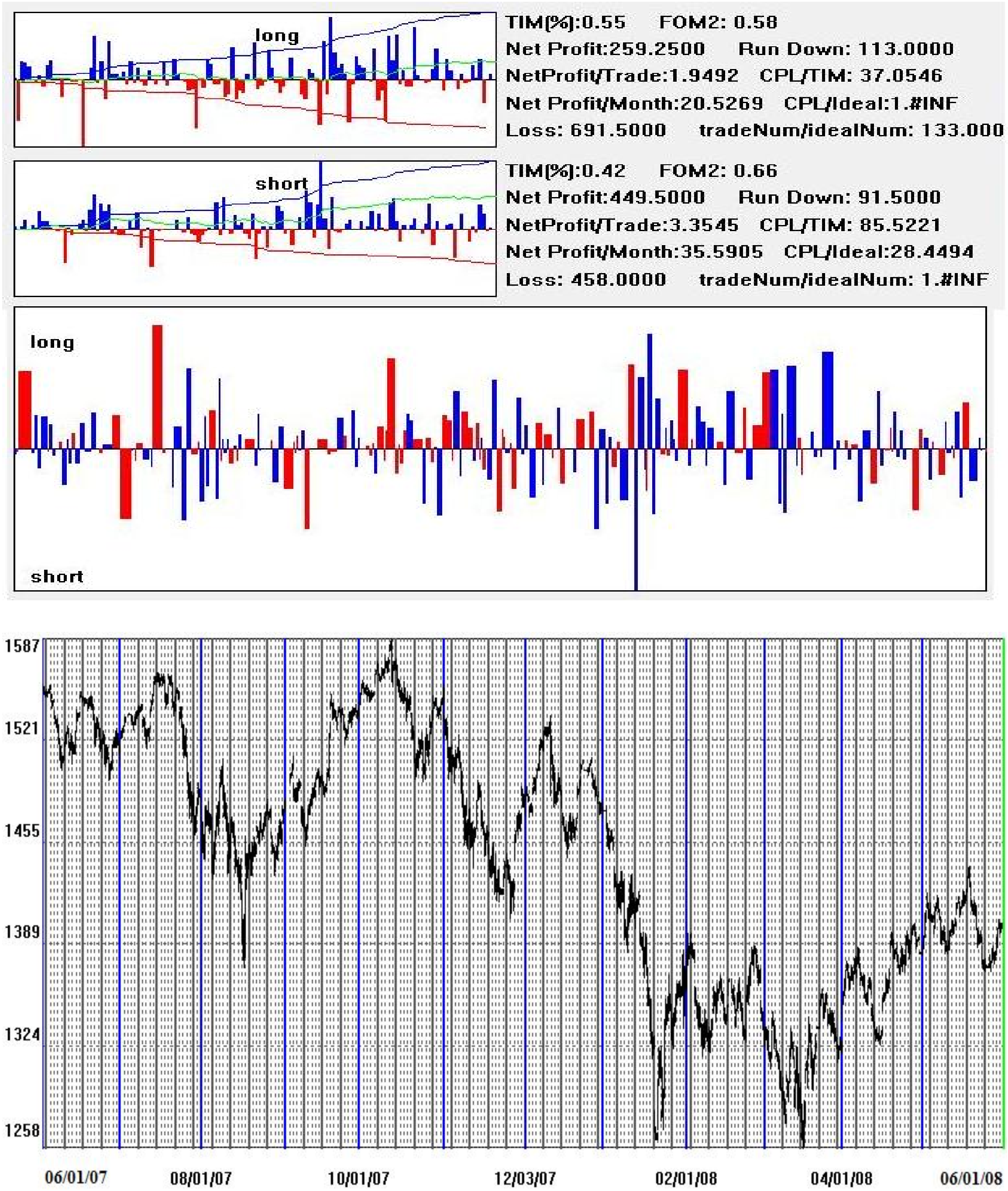


Figure 3.3 The top graph is the summary plot of TF2 trading with 50% passing rate under weekly ud sequence statistic from 06/2007 to 05/2008. The bottom graph is the price chart of the period from 06/2007 to 05/2008.

3.4.3 Trading Results with Use of Short-Term Cycle Leg-Length-Ratio Sequence Statistics

The cycle leg-length relation is one of the properties used above to represent short-term market characteristic. In previous section, we obtained Bayesian statistics about long-term control cycle phase derived from short-term cycle leg-length-ratio sequence. As in the work presented in the previous two sections, we also apply the statistics in the same test environment to find the effect of short-term cycle leg-length-ratio sequence to the TFx trading.

Table 3.16.1, 3.16.2, 3.16.3 and 3.16.4 contain the optimization results of the TFx methods for S&P during 01/2001 and 04/2006. In this study, C2 or C5 is used as the trading cycle, and CQ or CY is used as the control cycle whose phase is inferred from C2 or C5 cycle length leg-length-ratio sequences by Bayesian statistics.

Table 3.16.1 Optimized trading results for simulated real time C2 cycle testing of S&P 500 futures using the TFx methods (CQ predicted by C2 cycle leg-length-ratio sequences as control cycle) with cycle database during 01/2001 and 04/2006

Method	Best Passing Rate (%)	Long	Short	Total
TF1	60	419.2	452.5	871.7
TF2	50	720.4	782.6	1503
TF3	50	414.1	452.4	866.5
TF4	70	215.8	277.7	493.5
TF5	50	318.7	394.2	712.9
TV1	40	606.1	677.2	1283.3
TF2B	60	376.4	424.7	801.1

Table 3.16.2 Optimized trading results for simulated real time C2 cycle testing of S&P 500 futures using the TFx methods (CY predicted by C2 cycle leg-length-ratio sequences as control cycle) with cycle database during 01/2001 and 04/2006

Method	Best Passing Rate (%)	Long	Short	Total
TF1	70	498.8	531.3	1030.1
TF2	50	487.1	537.2	1024.3
TF3	40	189.9	271.3	461.2
TF4	30	188.2	276.3	464.5
TF5	70	259.6	288.2	547.8
TV1	50	498.6	568.5	1067.1
TF2B	60	389.5	419.5	809

Table 3.16.3 Optimized trading results for simulated real time C5 cycle testing of S&P 500 futures using the TFx methods (CQ predicted by C5 cycle leg-length-ratio sequences as control cycle) with cycle database during 01/2001 and 04/2006

Method	Best Passing Rate (%)	Long	Short	Total
TF1	70	342.5	375.8	718.3
TF2	30	455.8	555.6	1011.4
TF3	30	156	256.6	412.6
TF4	50	282.9	393.7	676.6
TF5	50	270.1	316.6	586.7
TV1	30	140.2	274.1	414.3
TF2B	50	305.3	368.8	674.1

Table 3.16.4 Optimized trading results for simulated real time C5 cycle testing of S&P 500 futures using the TFx methods (CY predicted by C5 cycle leg-length-ratio sequences as control cycle) with cycle database during 01/2001 and 04/2006

Method	Best Passing Rate (%)	Long	Short	Total
TF1	60	391.5	443.2	834.7
TF2	60	363.2	425.7	788.9
TF3	30	422.9	469.5	892.4
TF4	50	247.7	331.6	579.3
TF5	40	251.1	284.9	536
TV1	30	222.8	286.1	508.9
TF2B	70	554.9	558	1112.9

3.4.4 Trading Results with Use of Short-Term Cycle Extreme Up-or-Down (ud) Sequence Statistics

Short-term cycle (C2 or C5) extreme up-or-down (ud) sequence statistics is one of the four observables treated with Bayesian statistics that are presented in the study. Based on the statistics table, we can obtain the probability of control cycle phase under certain short-term cycle extreme ud sequences. In this section, we conduct an optimization test for the TFx methods when short-term cycle extreme ud sequence statistics is used to provide the control cycle phase.

Table 3.17.1, 3.17.2, 3.17.3 and 3.17.4 contain the optimization results for S&P during 01/2001 and 04/2006. Same as the settings in section 3.4.3, C2 or C5 is used as the trading cycle, and CQ or CY is used as the control cycle deduced from the application of short-term cycle extreme ud sequence statistics.

Table 3.17.1 Optimized trading results for simulated real time C2 cycle testing of S&P 500 futures using the TFx methods (CQ predicted by C2 short-term cycle extreme ud sequences as control cycle) with cycle database during 01/2001 and 04/2006

Method	Best Passing Rate (%)	Long	Short	Total
TF1	60	686.3	737.7	1424
TF2	80	435.1	490.7	925.8
TF3	70	473.6	510	983.6
TF4	30	241.2	316.1	557.3
TF5	40	145.9	212.6	358.5
TV1	70	408.3	476.8	885.1
TF2B	50	356.4	403.3	759.7

Table 3.17.2 Optimized trading results for simulated real time C2 cycle testing of S&P 500 futures using the TFx methods (CY predicted by C2 short-term cycle extreme ud sequences as control cycle) with cycle database during 01/2001 and 04/2006

Method	Best Passing Rate (%)	Long	Short	Total
TF1	60	559	600.2	1159.2
TF2	60	334.5	382.2	716.7
TF3	40	262.1	319.3	581.4
TF4	40	255.5	325.4	580.9
TF5	40	158.7	235.6	394.3
TV1	40	476.1	548.1	1024.2
TF2B	40	132.4	198.4	330.8

Table 3.17.3 Optimized trading results for simulated real time C5 cycle testing of S&P 500 futures using the TFx methods (CQ predicted by C5 short-term cycle extreme ud sequences as control cycle) with cycle database during 01/2001 and 04/2006

Method	Best Passing Rate (%)	Long	Short	Total
TF1	50	420.2	497.7	917.9
TF2	60	559.8	641.5	1201.3
TF3	70	290.9	349.6	640.5
TF4	30	301.8	391.1	692.9
TF5	30	320.7	346.7	667.4
TV1	40	436.3	523.6	959.9
TF2B	50	591.6	663.4	1255

Table 3.17.4 Optimized trading results for simulated real time C5 cycle testing of S&P 500 futures using the TFx methods (CY predicted by C5 short-term cycle extreme ud sequences as control cycle) with cycle database during 01/2001 and 04/2006

Method	Best Passing Rate (%)	Long	Short	Total
TF1	60	228.4	285	513.4
TF2	60	536	566.2	1102.2
TF3	60	196.5	249.8	446.3
TF4	50	280.2	362.5	642.7
TF5	60	280.2	362.5	642.7
TV1	50	444.2	542.3	986.5
TF2B	40	557.2	552.4	1109.6

3.4.5 Results and Discussion

In the above sections, we conducted the optimization tests for the TFx methods with use of a control cycle with phases inferred from short-term observables by Bayesian statistics. Four different Bayesian statistics, including daily ud sequences, weekly ud sequences, short-term cycle leg-length-ratio sequences and short-term cycle extreme ud sequences, were applied in the optimization for that purpose. During the study, C2 or C5 was used as the trading cycle, and CQ or CY was utilized as the control cycle. In order to evaluate the effect of Bayesian statistics, we plot those data together; meanwhile, for comparison purposes, the trading results, with or without use of man-made control cycle presented in Chapter 2, are also included in the graph.

The best C2 trading results for the TFx methods used with or without CQ as the control cycle are presented in Figure 3.4. For each TFx method, the trading results of six different variants, which are: without use of CQ, with use of man-made CQ, and with use of predicted CQ by four different Bayesian statistics, are aligned together for the convenience of comparison.

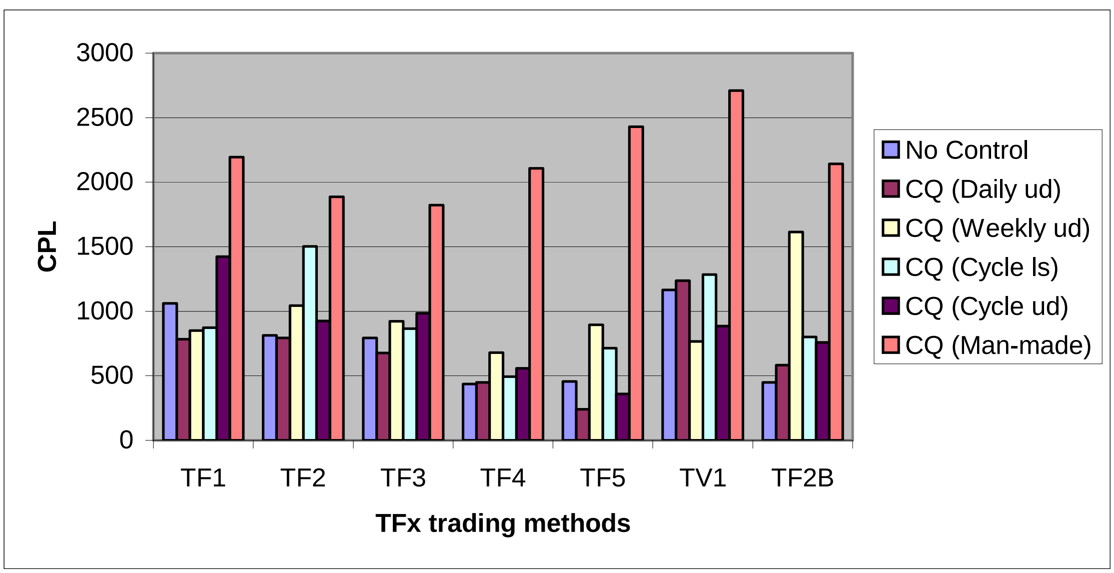


Figure 3.4 C2 best trading results for the TFx Methods with or without use of CQ as the control cycle for S&P 500 futures during 01/2001 and 04/2006. Data are from Table 2.1.3, 2.4.3, 3.13.1, 3.14.1, 3.16.1 and 3.17.1. (Cycle ls signifies cycle leg-length-ratio).

From Chapter 2, we already know that the trading results with use of man-made CQ control cycle are much better than those of without use of the CQ control cycle, and it is easy to see this in the above graph. According to the graph, we also find that the man-made CQ control cycle has a large impact than any of the CQ control cycle phases predicted by those statistics.

Further study of the graph shows that the CQ control cycle phase predicted by different statistics may or may not benefit the market forecasts. For the TF1 method, only the short-term cycle extreme ud sequence statistics improves the trading results. However, for the TF2 and TF3 methods, every statistics (except the daily ud sequences) helps the forecast, especially for TF2. The TF4 and TF2B methods benefit from every statistics, but the impact for the TF4 method is much less than for the TF2B method. For the TF5 method, weekly ud sequences and short-term cycle leg-length-ratio statistics are helpful, while irrationally, the other two even hurt the performance. As to the TV1 method, only daily ud sequences and short-term leg-length-ratio statistics helps slightly.

Figure 3.5 shows comparison of the best C2 trading results for the TFx methods with or without use of the long-term CY cycle as the control cycle. Again, six different scenarios for each TFx method are presented in the graph.

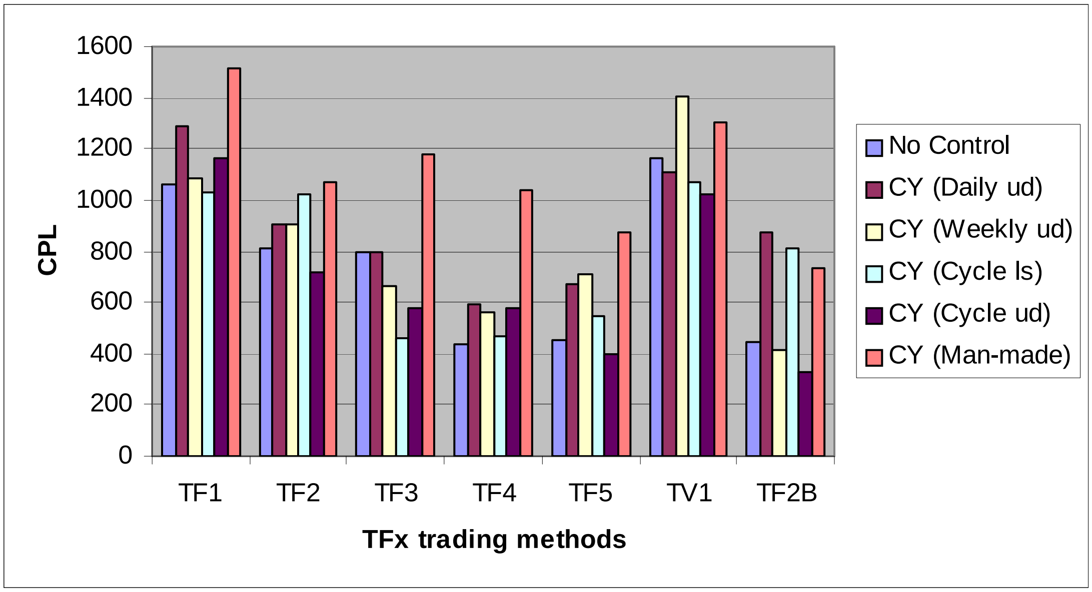


Figure 3.5 C2 best trading results for the TFx methods with or without use of CY as control cycle for S&P 500 futures during 01/2001 and 04/2006. Data are from Table 2.1.3, 2.4.3, 3.13.2, 3.14.2, 3.16.2 and 3.17.2. (Cycle ls signifies cycle leg-length-ratio).

Based on the graph, for every single TFx method, the effect of CY control cycle is not as significant as it was for CQ control cycle above, while the effect of man-made CY control cycles is still larger than that of most of the predicted CY control cycles, it is remarkable that at least in some cases, the effect of the Bayesian-derived CY control cycle phase is as great or (marginally) greater that the effect of the "man-made" CY control cycle phase. For the TF1 method, except for the short-term cycle leg-length-ratio sequence statistics, the remaining three improve the trading results. Short-term cycle extreme ud sequence statistics have a worse effect on the TF2 and TF5 methods, but the other three have a better impact. No statistics have a positive effect on the TF3 method, but they do improve the TF4 method. For the TV1 method, only the weekly ud sequence statistics boost the performance, up to the point where it is even better than that of the man-made control cycle. As to the TF2B method, only daily ud sequence and short-term cycle leg-length-ratio sequence statistics improve the trading results, also to a degree better than the man-made CY control cycle.

Switching from C2 to C5 cycle performance, Figure 3.6 shows a comparison of the best C5 trading results for the TFx methods with or without use of CQ as the control cycle. The format is same as in the previous two figures. Overall, as for C2 cycle, the effect of the predicted CQ control cycle is much less than that of the "man-made" one.

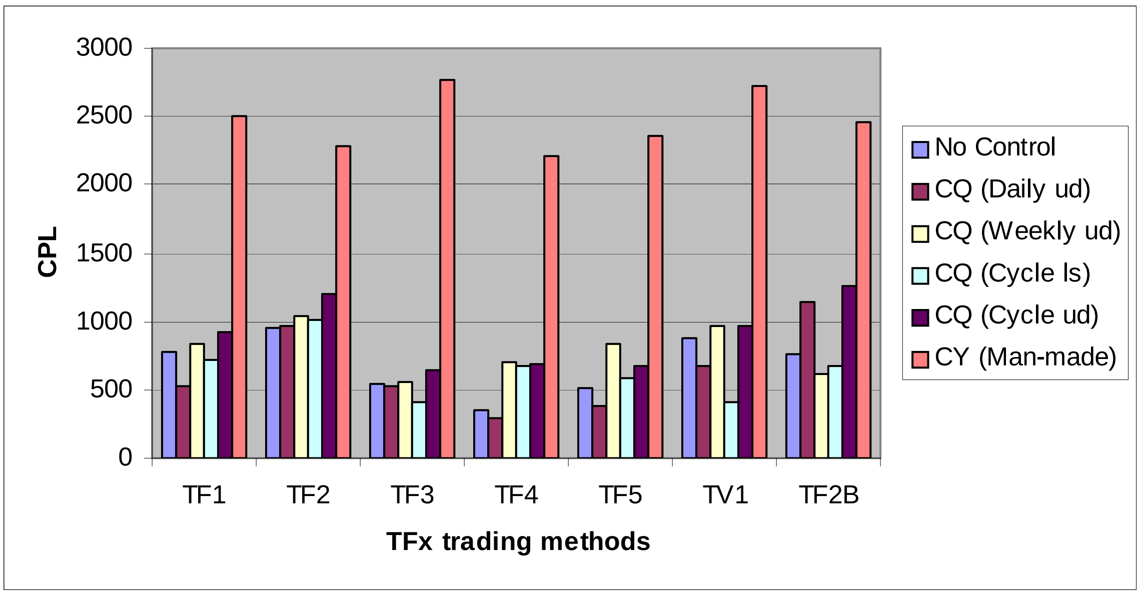


Figure 3.6 C5 best trading results for the TFx methods with or without use of CQ as control cycle for S&P 500 futures during 01/2001 and 04/2006. Data are from Table 2.3.3, 2.5.3, 3.13.3, 3.14.3, 3.16.3 and 3.17.3. (Cycle ls signifies cycle leg-length-ratio).

For the TF1, TF3 and TV1 methods, weekly ud sequences and short-term cycle extreme ud sequences statistics help improve trading results. All four statistics give a little boost to the performance of the TF2 method. Three statistics, except daily ud sequences, work for the TF4 and TF5 methods. As for the TF2B method, daily ud sequences and short-term cycle extreme ud sequences statistics are good for the market forecast.

Continuing on C5 cycle, but changing from CQ as control cycle to CY, Figure 3.7 presents a comparison of the best C5 trading results for the TFx methods with or without use of CY as the control cycle. Here, the trading results obtained with the help of "man-made" CY control cycles are better than any of the Bayesian counterparts.

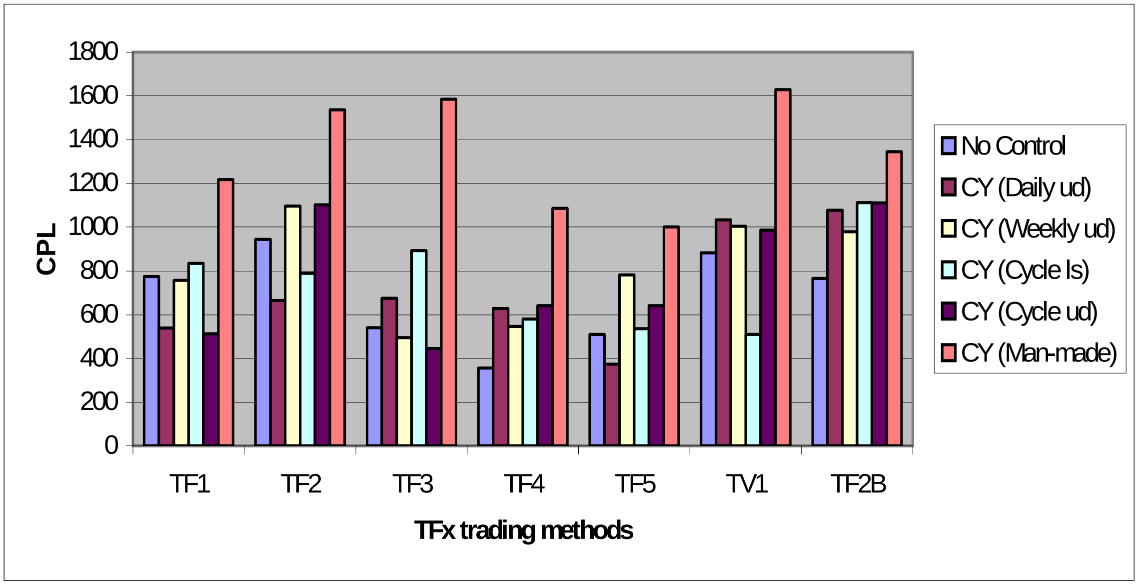


Figure 3.7 C5 best trading results for the TFx methods with or without use of CY as control cycle for S&P 500 futures during 01/2001 and 04/2006. Data are from Table 2.3.3, 2.5.3, 3.13.4, 3.14.4, 3.16.4 and 3.17.4. (Cycle ls signifies cycle leg-length-ratio).

To break down the effect of each statistics on the TFx methods, we take a closer look at the graph. Only short-term cycle leg-length-ratio sequence statistics works for the TF1 method, while weekly ud sequence and short-term cycle extreme ud sequence are useful to the TF2 and TF3 method. Meanwhile, all statistics work for the TF4 and TF2B methods. For the TF5 method, three statistics, not including daily ud sequences, are helpful. As to the TV1 method, the statistical methods generally also helps, only short-term cycle extreme ud sequence statistics contributes worse trading results.

To sum up, the control cycles with phases predicted by use of Bayes' Theorem has a mixed effect to the TFx methods. The effect of predicted control cycles on the TFx trading is usually less than that of "man-made" control cycle; sometimes, the results are even worse than a benchmark, which doesn't use any control cycle. It shows that the cycle phase predicted by Bayes' Theorem is not as accurate as that of the "man-made" cycle; therefore, it can't match the effect of "man-made" cycle.

Compared to benchmark results, some statistics, including weekly ud sequences and short-term cycle extreme ud sequences, do improve the trading in most cases. Part of the reason is that, in these two statistics, the probabilities of control cycle phase under certain sequences are greater than those of the other two statistics. It means weekly ud sequence and short-term cycle extreme ud sequence statistics may provide a more accurate control cycle phase prediction than other two, therefore, help corresponding trading.

As for predicted CQ control cycle, its effect is usually not close to that of "man-made" CQ. However, for predicted CY control cycle, the effect is comparable to that of the "man-made" CY. This result may rest on the fact that CQ is shorter than CY in terms of cycle length. As a result, a CQ cycle phase change is quicker than for CY, and it requires a quicker reaction to the change. Generally, the prediction of control cycle phase has a certain lag so that it is easier to get a false judgment, especially for a shorter cycle, which is CQ in our case.

When comparing the effects of predicted CQ and CY control cycles, there is no big difference between them. This may be explained by the interplay of two factors. One factor is that there are more false CQ phase predictions than CY, as mentioned above. The other factor is that the probabilities for CQ phase are usually higher than those of CY; meanwhile, a correct CQ pre-selected background has a more positive impact than

CY. Because of a partial offset of the two factors, sometimes, a predicted CQ has larger impact than a predicted CY; at other times, the impact is the opposite.

3.5 Summary

Bayes' Theorem is a theorem of probability theory, which is often used to compute posterior probabilities, given observations.

In this chapter, Bayes' Theorem has been used in the predictions of phase of long-term cycles (CQ and CY). Four different short-term market features, including daily up-or-down sequences, weekly up-or-down sequences, short-term cycle (C2 or C5) leg-length-ratio sequences, and short-term cycle (C2 or C5) extreme up-or-down sequences, have been utilized to infer long-term cycle phases. The statistics for those four short-term market features have been presented. Based on these short-term market features and the corresponding statistics, we can get a Bayesian-derived probability about the current long-term cycle phase.

Next, the control cycle phase predicted by Bayes' Theorem from short-term market observations has been used to help market cycle forecasts. By comparing the trading results of the TFx methods with or without use of the control cycle, we find that the "man-made" control cycle (obviously only with hindsight) causes the largest improvement to trading results, and the predicted control cycle has mixed effect. Compared to benchmark results, which doesn't use the control cycle, some statistics, weekly ud sequences and short-term cycle extreme ud sequences, do improve the trading in most cases. As for the predicted CQ control cycle, its effect is usually not close to that of a "man-made" CQ. However, for predicted CY control cycle, the effect is comparable to that of man-made CY. There is no big difference between the effects of predicted CQ and CY.

Chapter 4 Financial Market Cycle Analysis by Artificial Neural Networks

4.1 Overview

Artificial Neural Networks (ANNs)[75-78] are information processing systems that have seen an explosion of interest in recent years, and have been successfully applied across an extraordinary range of problem domains, in areas as diverse as finance, medicine, engineering, geology and physics.

The motion of stock prices and stock price indices is an example of a complex, but in some circumstances at least partially deterministic phenomenon to which ANNs can be applied. Although ANNs have been used in technical analysis to make forecasts of stock prices[79-81], this use was based upon factors such as past comparative performances of other stocks and fundamental quantities such as various economic indicators, I show here a first use in connection with cyclic analysis of market indexes.

In Chapter 2, we introduced and discussed the TFx methods, which are a family of cycle-based pattern recognition methods for cycle analysis. The performances of seven TFx methods have been presented and reviewed. As the TFx methods can be categorized as belonging to the second and third of three classic types: leading, coincident and trailing, it becomes reasonable to combine information from indicators with different characteristics with respect to those criteria to come up with a new combination trading method. In Dr. Zhao's thesis[41], she explored the possibility of combining TFx methods by simply adding up the passing rates of these TFx methods to set up new trading rules, of which at least one yielded superior results, but only two TFx methods had been used in her thesis line of research.

In this chapter, I try to use a different methodology, the ANN, to combine TFx methods with each other as well as with other cycle information. By pre-treating inputs/target data and choosing a suitable ANN structure, training parameters, and output function, we first train the network in a training phase to obtain optimal parameters; then, we apply the trained network to trading in the test phase. A moving time window optimization has been used in the process of training and testing.

Section 4.2 introduces the principle and the components of an ANN. Section 4.3 presents some details of the ANN used in the study. Section 4.4 applies the sigmoid network in the S&P futures market and discusses the results. Section 4.5 gives a summary and draws some conclusions.

4.2 Introduction of Artificial Neural Networks for Use in Market Cycle Analysis

4.2.1 Fundamentals of Artificial Neural Networks

An artificial neural network is an information processing system with a design inspired by studies of the brain and the nervous system. The human brain is composed of billions of special cells called neurons which are connected by complex networks. Therefore, the brain can be viewed as a collection of neural networks. Information is first received by some neurons, then passed to other neurons in the network. During the process, a signal will be generated. Based on the information and path of the signal, the brain analyzes and categories the information.

In 1943, McCulloch and Pitts[82] first introduced artificial neurons, which were conceptualized on the basis of the functionality of biological neurons. Since then, the field of ANNs has been much studied and great progress has been made. At present, due to the advance of computer technology and better understanding of the mechanisms of the brain, artificial neural networks are emerging as the technology of choice for many applications, such as pattern recognition, prediction, system identification, and control.

An ANN is a collection of artificial neurons, which are connected through a specific structure. In the network, the artificial neuron is the basic element which receives inputs, processes them, and delivers an output. This process is shown in Figure 4.1. As stated, a neuron receives a set of inputs, which are assigned corresponding weights. By adjusting the weight for each input, we can change the relative importance of the inputs. The summation function, a linear combination, adds up all the weighted inputs. Finally, an activation function controls the amplitude of the output of the neuron, which lies in the range between 0 and 1, or alternatively -1 and 1.

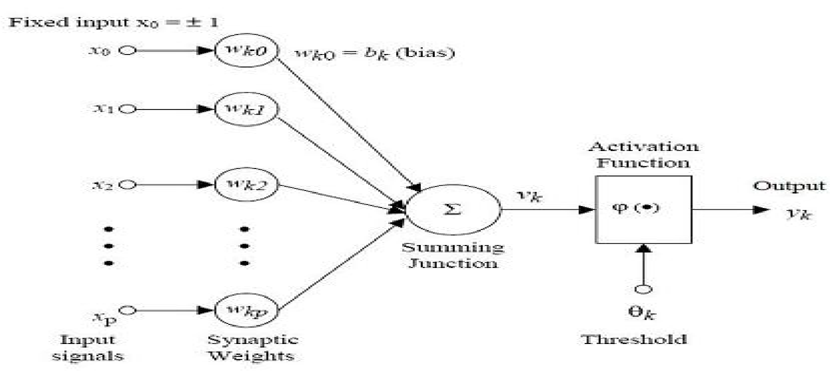


Figure 4.1 Processing information in an artificial neuron.

The summation of the weighted inputs can be shown by the equation 4.1.

$$v_k = \sum_{j=1}^{p} w_{kj} x_j \qquad \textbf{4.1}$$

where
V_k: the result of the function
W_{kj}: the weight for the input X_j
X_j: the input j
P: the number of the inputs

4.2.2 Structure of the Network

An ANN consists of a set of neurons, which are grouped together in different layers. In terms of the connection pattern and the propagation of signals, there are two major types of networks: feed-forward and feedback neural networks.

Feed-forward networks are a type of networks that allows signals to travel only one way, from input to output. The data processing can extend over multiple layers, but there are no feedback connections, which mean the output of any layer does not affect that same layer or previous layers. Feed-forward networks tend to associate inputs with outputs directly, and are extensively used in pattern recognition.

Rather than connecting only one way, feedback networks allow signals to travel in both directions by introducing loops in the network. Feedback networks are very powerful and can be extremely complicated. The dynamical properties of the network are important, and their states are changing continuously until they reach an equilibrium point. Feedback architectures are also referred to as interactive or recurrent. The latter term is often used to denote learning feedback connections in single-layer organizations.

4.2.3 Learning of the Network

When an ANN is constructed, it needs to be configured such that a set of inputs leads to the desired outputs. Usually, the ANN is trained by feeding learning patterns, and adjusting the weights of the inputs according to pre-set learning rules. Based on the training data and methods, the learning can be divided into two distinct categories: supervised and unsupervised learning.

Supervised learning uses as training data a set of inputs, whose desired outputs are known ahead of time. Figure 4.2 illustrates supervised learning by the ANN. At first, a set of inputs is fed into the network, and corresponding outputs are generated after their manipulation by the network. Then, based on the differences of outputs and targets, an error vector will be calculated, and used in a supervised learning algorithm to adjust the parameters of the network, i.e., the weights of neurons and threshold values, etc. The process will be repeated until a certain error criterion or a preset number of training steps has been reached. Backpropagation is a classic and widely used supervised learning process.

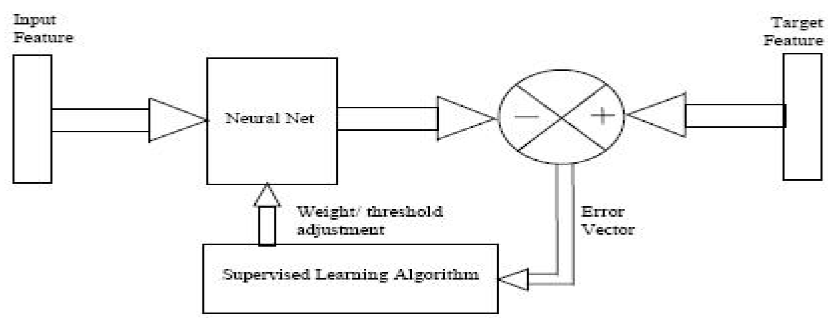


Figure 4.2 Supervised learning of artificial neural network.

In contrast to supervised learning, in unsupervised learning there are only inputs and no corresponding targets. The network is trained to discover the patterns of the inputs. Since no knowledge is available about the correctness of the training, it is possible for the network to generate meaningless results, in this case, which does not apply to the following.

4.2.4 Activation Function

The behavior of an ANN depends on the structure of the network, the weights of neurons, and also the activation function (refer to Figure 4.1). As mentioned previously, the aim of the activation function is to control the output in a certain range, usually between 0 and 1, or -1 and 1.

In general, three types of activation functions, denoted by $\varphi(v)$, are widely used.

The first type is the threshold function (shown as Equation 4.2). If the value of summed inputs is less than a certain threshold value, i.e., 0, it will take on a value of 0; otherwise, the output will be 1.

$$\varphi(v) = \begin{cases} 1 & \text{if } v \geq 0 \\ 0 & \text{if } v < 0 \end{cases} \qquad \textbf{4.2}$$

where
$\varphi(v)$: activation function
v : the summed inputs

The second type is the Piecewise-Linear function (shown as Equation 4.3). Based on the value of the summed inputs, the output can have a value between a certain thresholds; the interval (0,1) is used in this thesis.

$$\varphi(v) = \begin{cases} 1 & v \geq \frac{1}{2} \\ v & -\frac{1}{2} > v > \frac{1}{2} \\ 0 & v \leq -\frac{1}{2} \end{cases} \qquad \textbf{4.3}$$

The third type is the sigmoid function, which can generate any value between 0 and 1. The hyperbolic tangent function (shown as Equation 5.4) is an example of a sigmoid function.

$$\varphi(v) = \tanh\left(\frac{v}{2}\right) = \frac{1-\exp(-v)}{1+\exp(-v)} \qquad \textbf{4.4}$$

4.3 Design of Present Artificial Neural Network for Market Cycle Analysis

As discussed before, the TFx methods, a family of cycle-based pattern recognition methods, can be used to provide guidance for present market development. As every TFx method belongs to one of two types of indicators, namely coincident and trailing indicators, it is reasonable to incorporate these together to complement each other for a better market forecast.

Generally, an ANN is an adaptive, most often nonlinear system that learns to perform a function from data. Since the ANN provides an effective means to deal with multiple inputs, we try to design an ANN to take signals from the TFx methods and produce an output that can be used for a market forecast.

The ANN is built with a systematic step-by-step procedure to optimize a performance criterion or to follow some internal constraint, which is commonly referred to as the learning rule. After the training phase, the fine-tuned system is being used to produce outputs based on the inputs.

In order to get results from our network, we use supervised learning to train it. Therefore, the training data composed of inputs and corresponding desired outputs are very important, because they provide the necessary information to discover the optimal network parameters. In supervised learning, the error information, derived from the difference between the desired output and the system output, is fed back to the system and the system parameters are adjusted in a systematic fashion. The process is repeated until the performance is acceptable. It is clear that the performance rests heavily on the quality of the inputs and desired outputs data.

In order to design an appropriate ANN for our research, we need to prepare inputs and desired outputs training data, choose the network structure, the learning rule, etc. It is described in the following sections.

4.3.1 Inputs

Inputs are data fed into the network. It is crucial to send appropriate inputs into the network, because when irrelevant or wrong information is used as inputs, it may decrease the effectiveness of the network or even cause failure.

In this study, we mainly focus on cycle-based market forecast, and therefore nine different cycle-related measurements have been selected as inputs. Among them, seven inputs are from the TFx methods, which are the passing rates from use of TF1, TF2, TF3, TF4, TF5, TV1 and TF2B (refer to Chapter 2).

The other two cycle-related inputs are time leg ratio and price leg ratio. The time leg is defined as the time interval between the present time and the last cycle extreme. As a result, the time leg ratio is the ratio of the time leg and a cycle-related length unit, which is 80 “time points” for C2, and 200 “time points” for C5. In this study, one “time point” is equal to a 10-minute time interval. The price leg refers to the price difference between the present price and the price at the last cycle extreme. Accordingly, the price leg ratio is the ratio of the price leg and a cycle-related value, which is 100 S&P points for C2, and 200 S&P points for C5.

In general, the inputs are scaled between 0 and 1. As we know, the passing rate of every TFx method is between 0 and 100% so that there is no need for input pre-treatment. For the time leg ratio, the value starts from 0 and increases gradually with advancing time. When the ratio is larger than 1, it is still assigned the value 1. A similar pre-treatment is applied to the price leg ratio. When the ratio is less than 0, it will be assigned to 0; if the ratio is larger than 1, it will be set equal to 1.

4.3.2 Desired Outputs

In the study, supervised learning will be utilized in the training of the network so that training data with inputs/desired outputs pair is necessary for the training. The goal of the network training is to match the output with the desired output based on the inputs. Since the output is usually scaled between 0 and 1, the desired outputs also need to be adjusted between 0 and 1.

As the aim of our ANN is to provide a trading signal that predicts the next cycle extreme, the desired outputs will contain cycle-related information. Here, desired outputs are quantitative numbers which represent the closeness of the "now" point to the next cycle extreme. In order to get the desired outputs corresponding to the inputs, we use the following procedures (shown in Figure 4.3).

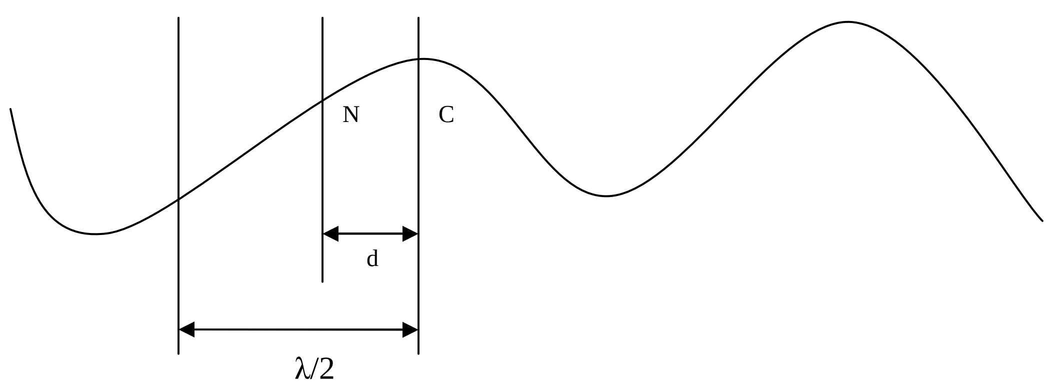


Figure 4.3 The schematic graph of the desired output based on the half-cycle length λ/2. N is the historical "now" point; C is the historical next cycle extreme. d is the time distance from N to C; λ/2 is the average half-cycle length.

We start from a man-made cycle extreme, and choose the time point that is a certain distance away from that cycle extreme. At that "time point", we can obtain the values of inputs: the passing rates of the TFx methods, time leg ratio and price leg ratio. Then, the corresponding desired output is calculated in the following:

We first calculate t, the intermediate value of the desired output, which is the ratio of the time difference from the present "time point" to the next cycle extreme, and the half-cycle length through Equation 4.5. Depending on the cycle, the half-cycle length will be defined 100 for C5; and it is set to 40 for C2.

$$t = d \,/\, (\lambda/2) \qquad \textbf{4.5}$$

where
t: the intermediate value of the desired output
d: the time distance from the "now" point to the next cycle extreme
$\lambda/2$: the half-cycle length

Then, the intermediate desired output is being limited to a value between –1 and 1 in Equation 4.6.

$$t = \begin{cases} 1 & t > 1 \\ t & 1 \geq t \geq -1 \\ -1 & t < -1 \end{cases} \qquad \textbf{4.6}$$

Finally, the desired output between 0 and 1 is obtained by converting the intermediate desired output through Equation 4.7. For example, if present "time point" is exactly the next cycle extreme, then the intermediate desired output will be 0 and the desired output will be 0.5.

$$T = (t + 1) / 2 \tag{4.7}$$

Where
T: the desired output

For every historical man-made cycle, we only pick up three sets of inputs/desired output pair as training data. For C2, the three sets of data are 20 points, 40 points and 60 points away from the last cycle extreme, respectively; for C5, those training data will be 50, 100 and 150 points from the last cycle extreme.

4.3.3 Structure of the Network

Due to its simplicity and strong ability in pattern recognition, a feed-forward three-layer network is chosen as our test network. Three layers, a popular setting, include input layer, hidden layer and output layer.

The activity of the input layer represents the original inputs that are fed into the network. The activity of the hidden layer is determined by the activity of the input layer and the weights on the connections between the input and the hidden layers. Likewise, the behavior of the output layer depends on the activity of the hidden layer and the weights between the hidden and output layers.

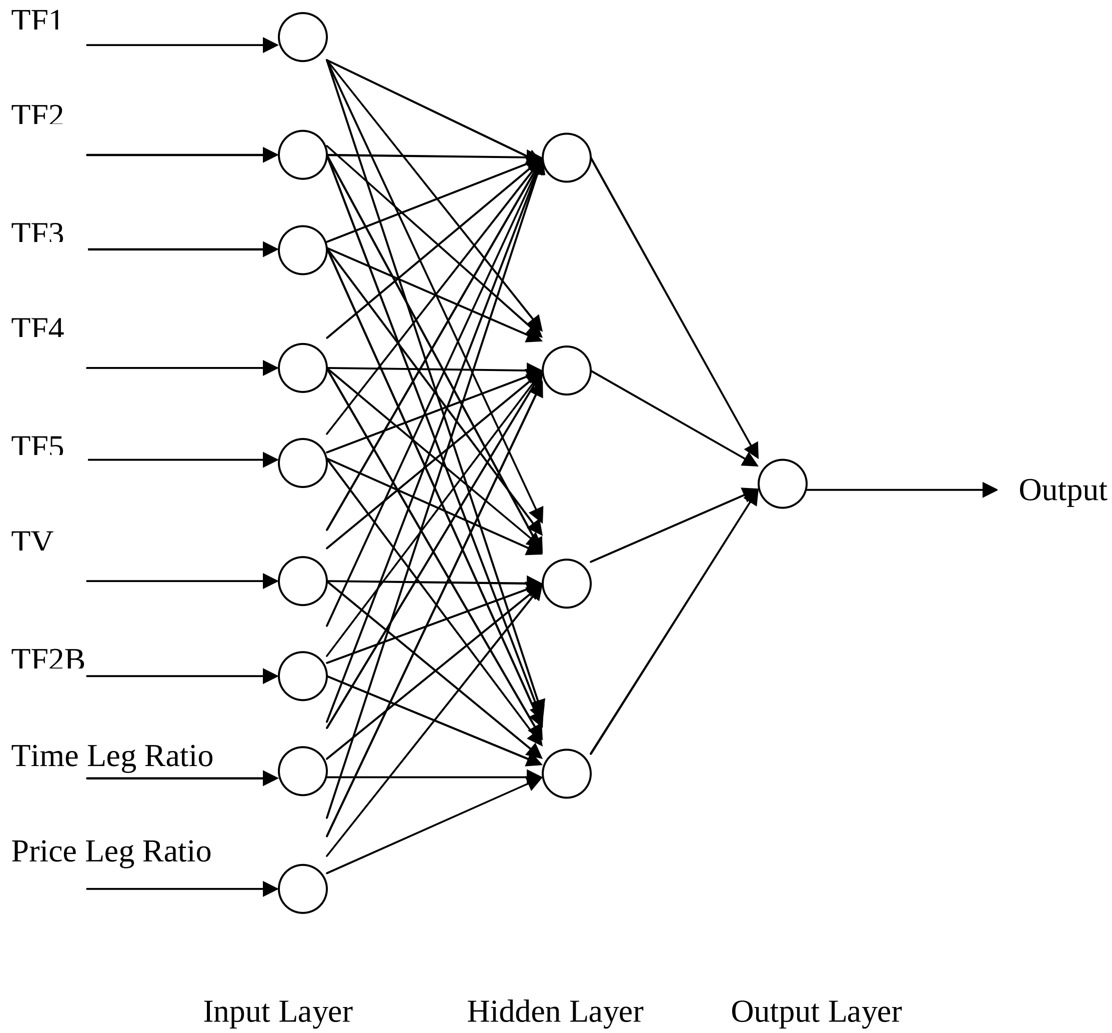


Figure 4.4 The Neural Network structure in this study. There are three layers: input, hidden and output. Nine units are in the input layer, and four units in the hidden layer, only one unit in the output layer. The nine inputs are TF1, TF2, TF3, TF4, TF5, TV1, TF2B, Time Leg Ratio and Price Leg Ratio.

The network used in the study is illustrated in Figure 4.4. In the input layer, there are 9 inputs, and 4 intermediate outputs that become the inputs for the hidden layer. Through the hidden layer, there are another 4 intermediate outputs that will be fed into the output layer. In the end, only one output will be generated as the indicator we will use to make market forecast.

4.3.4 Learning of the Network

Through the procedure discussed in above sections, we can obtain inputs/target sets for the training of the network. In this study, backpropagation, a sort of supervised learning, will be applied to the training. During the process, a learning rate of 0.1 will be used to adjust the weights of neuron units and find the best weights setting for the network.

4.3.5 Activation Function

In the output layer, a sigmoid function will be utilized as the activation function because sigmoid units bear a greater resemblance to real neurons than do linear or threshold units. For sigmoid units, the output varies continuously but not linearly as the input changes.

4.4 Application of the Artificial Neural Network to Market Cycle Analysis

As described above, I designed a network that belongs to the class of feed-forward networks, and contains three layers. In this section, analogous to the tests conducted in previous chapters, we apply the network to the forecast of the S&P futures market during 02/2002 and 02/2006.

As discussed, this application has two phases: the training phase and the test phase; before applying the network to the test period, we need to train it first by feeding it with inputs/desired outputs pairs from the training period. A moving time window method is applied to the network training and testing. We first use a certain length of time window, i.e. 6-month, as the training period to train the network; then, we apply the trained network in the following test period, i.e. 3 months. After the test, the test period can be used again in the training phase, and replace the first 3-month of the old training period. The process will be repeated until the end of the whole test period. Through this method, we can always absorb new information and adjust the network parameters accordingly.

In the study, the test window is set as 3-month, while the training window will be 6-month, 9-month and 1-year, respectively, in order to test the effect of training period length.

According to the definition of the desired output, which is between 0 and 1, 0.5 is viewed as the next cycle extreme. Therefore, when applying the network in the testing, we formulate the trading rule as follows: when the output of the network is between 0.45 and 0.55 (i.e., near-optimal), a trade signal will be generated; otherwise, when the time leg ratio is 1, a trade signal will be generated as well, since even a non-optimal trade must be ended in any case at some time.

4.4.1 C2 Cycle Trading Results

In this section, we apply the network to C2 cycle trading for S&P futures during 02/2002 to 02/2006. The inputs are the nine C2-related parameters discussed above.

Table 4.1, 4.2 and 4.3 present the C2 trading results with the training window set as 6-month, 9-month and 1-year, respectively. In the table, "CPL" is the sum of Long and Short trading results in the test window using the network trained in the training window. "CPL Sum" calculates the cumulative sum of the CPLs of all periods to date; thus, the last value of "CPL Sum" is the result for the whole test period.

In terms of CPL, the best result, just above 1000 S&P points, is found for the network trained with 6-month as the training window. The result for the 1-year training window is a little bit better than that of the 9-month training window, but both are inferior to the 6-month result. The reason may rest on the greater closeness of the training window period to that of the test window. The closer the training window length is to that of the test window, the more similar those two periods seem to be; therefore, shorter trading windows seem to provide better network parameters to the short test period. For a proof, an extension of the training window length to 3-month is required.

The graphic comparison of the C2 trading results with different training windows is illustrated in Figure 4.5. The related price chart is also shown in the graph. The top three graphs correspond to 6-month, 9-month and 1-year, respectively. The graphs show that those three networks have similar performance that works better at the beginning and end of the test period, but not the middle.

Compared to the C2 trading results for a single TFx method (refer to Chapter 2), the present ANN combination method is better than that of any other TFx methods except the TV1 method. It means that the combination of the TFx methods using current ANN setting is partially successful because it provides an effective means of combining the TFx methods with different characteristics and produces a decent result compared to the component methods; however, results of this method don’t excel the best single method.

Table 4.1 C2 cycle trading results for S&P 500 futures from 02/2002 to 02/2006. The training period is 6-month, and the test period is 3-month. The test period is always 3-month ahead of the training period. Long is the trading result for the long side when the trained network is applied to the test period. Short is for the short side. CPL is the sum of Long and Short. “CPL Sum” is the cumulative sum of previous CPLs, i.e., the last value is the cumulative CPL over the total period studied.

Number	Training Period	Test Period	Long	Short	CPL	CPL Sum
1	08/01/01~01/31/02	02/01/02~04/30/02	-34.7	-21.7	-56.4	-56.4
2	11/01/01~04/30/02	05/01/02~07/31/02	-9.5	175.3	165.8	109.4
3	02/01/02~07/31/02	08/01/02~10/31/02	96.7	42.5	139.2	248.6
4	05/01/02~10/31/02	11/01/02~01/31/03	-24.6	52.1	27.5	276.1
5	08/01/02~01/31/03	02/01/03~04/30/03	123.6	53.9	177.5	453.6
6	11/01/02~04/30/03	05/01/03~07/31/03	133.7	76.5	210.2	663.8
7	02/01/03~07/31/03	08/01/03~10/31/03	58.7	-20.7	38	701.8
8	05/01/03~10/31/03	11/01/03~01/31/04	32.6	-39.2	-6.6	695.2
9	08/01/03~01/31/04	02/01/04~04/30/04	-54.9	-19.7	-74.6	620.6
10	11/01/03~04/30/04	05/01/04~07/31/04	39.9	69.8	109.7	730.3
11	02/01/04~07/31/04	08/01/04~10/31/04	74	-18.8	55.2	785.5
12	05/01/04~10/31/04	11/01/04~01/31/05	15	22	37	822.5
13	08/01/04~01/31/05	02/01/05~04/30/05	31.7	49.9	81.6	904.1
14	11/01/04~04/30/05	05/01/05~07/31/05	28.3	-27	1.3	905.4
15	02/01/05~07/31/05	08/01/05~10/31/05	21	64.1	85.1	990.5
16	05/01/05~10/31/05	11/01/05~01/31/06	50.9	-10.7	40.2	1030.7

Table 4.2 C2 cycle trading results for S&P 500 futures from 02/2002 to 02/2006. The training window is 9-month, and the test window is 3-month. “CPL Sum” is the sum of previous CPLs.

Number	Training Period	Test Period	Long	Short	CPL	CPL Sum
1	05/01/01~01/31/01	02/01/02~04/30/02	-58.2	-46.4	-104.6	-104.6
2	08/01/01~04/30/02	05/01/02~07/31/02	-75	119.4	44.4	-60.2
3	11/01/01~07/31/02	08/01/02~10/31/02	128.4	42.6	171	110.8
4	02/01/02~10/31/02	11/01/02~01/31/03	29.5	86.3	115.8	226.6
5	05/01/02~01/31/03	02/01/03~04/30/03	13.8	-48	-34.2	192.4
6	08/01/02~04/30/03	05/01/03~07/31/03	92.1	48.1	140.2	332.6
7	11/01/02~07/31/03	08/01/03~10/31/03	86.8	8.7	95.5	428.1
8	02/01/03~10/31/03	11/01/03~01/31/04	7.7	-50.8	-43.1	385
9	05/01/03~01/31/04	02/01/04~04/30/04	-14	16.3	2.3	387.3
10	08/01/03~04/30/04	05/01/04~07/31/04	4.3	35.5	39.8	427.1
11	11/01/03~07/31/04	08/01/04~10/31/04	23.2	-69.7	-46.5	380.6
12	02/01/04~10/31/04	11/01/04~01/31/05	5.4	7.4	12.8	393.4
13	05/01/04~01/31/05	02/01/05~04/30/05	1.4	20.6	22	415.4
14	08/01/04~04/30/05	05/01/05~07/31/05	99	42.3	141.3	556.7
15	11/01/04~07/31/05	08/01/05~10/31/05	6.9	57.4	64.3	621
16	02/01/05~10/31/05	11/01/05~01/31/06	45.9	-15.2	30.7	651.7

Table 4.3 C2 cycle trading Results for S&P 500 futures from 02/2002 to 02/2006. The training period is 1-year, and the test period is 3-month. “CPL Sum” is the sum of previous CPLs.

Number	Training Period	Test Period	Long	Short	CPL	CPL Sum
1	02/01/01~01/31/02	02/01/02~04/30/02	13.5	20.8	34.3	34.3
2	05/01/01~04/30/02	05/01/02~07/31/02	-80.4	110.3	29.9	64.2
3	08/01/01~07/31/02	08/01/02~10/31/02	57.7	3.4	61.1	125.3
4	11/01/01~10/31/02	11/01/02~01/31/03	54.4	118	172.4	297.7
5	02/01/02~01/31/03	02/01/03~04/30/03	44.5	-19.6	24.9	322.6
6	05/01/02~04/30/03	05/01/03~07/31/03	74.5	31.7	106.2	428.8
7	08/01/02~07/31/03	08/01/03~10/31/03	28.8	-46	-17.2	411.6
8	11/01/02~10/31/03	11/01/03~01/31/04	10	-58.5	-48.5	363.1
9	02/01/03~01/31/04	02/01/04~04/30/04	-38.4	-1.9	-40.3	322.8
10	05/01/03~04/30/04	05/01/04~07/31/04	16.9	47.8	64.7	387.5
11	08/01/03~07/31/04	08/01/04~10/31/04	77.2	-14.8	62.4	449.9
12	11/01/03~10/31/04	11/01/04~01/31/05	-2.5	8.2	5.7	455.6
13	02/01/04~01/31/05	02/01/05~04/30/05	10.4	27	37.4	493
14	05/01/04~04/30/05	05/01/05~07/31/05	89.5	30.3	119.8	612.8
15	08/01/04~07/31/05	08/01/05~10/31/05	26.2	71.1	97.3	710.1
16	11/01/04~10/31/05	11/01/05~01/31/06	40.7	-22.8	17.9	728

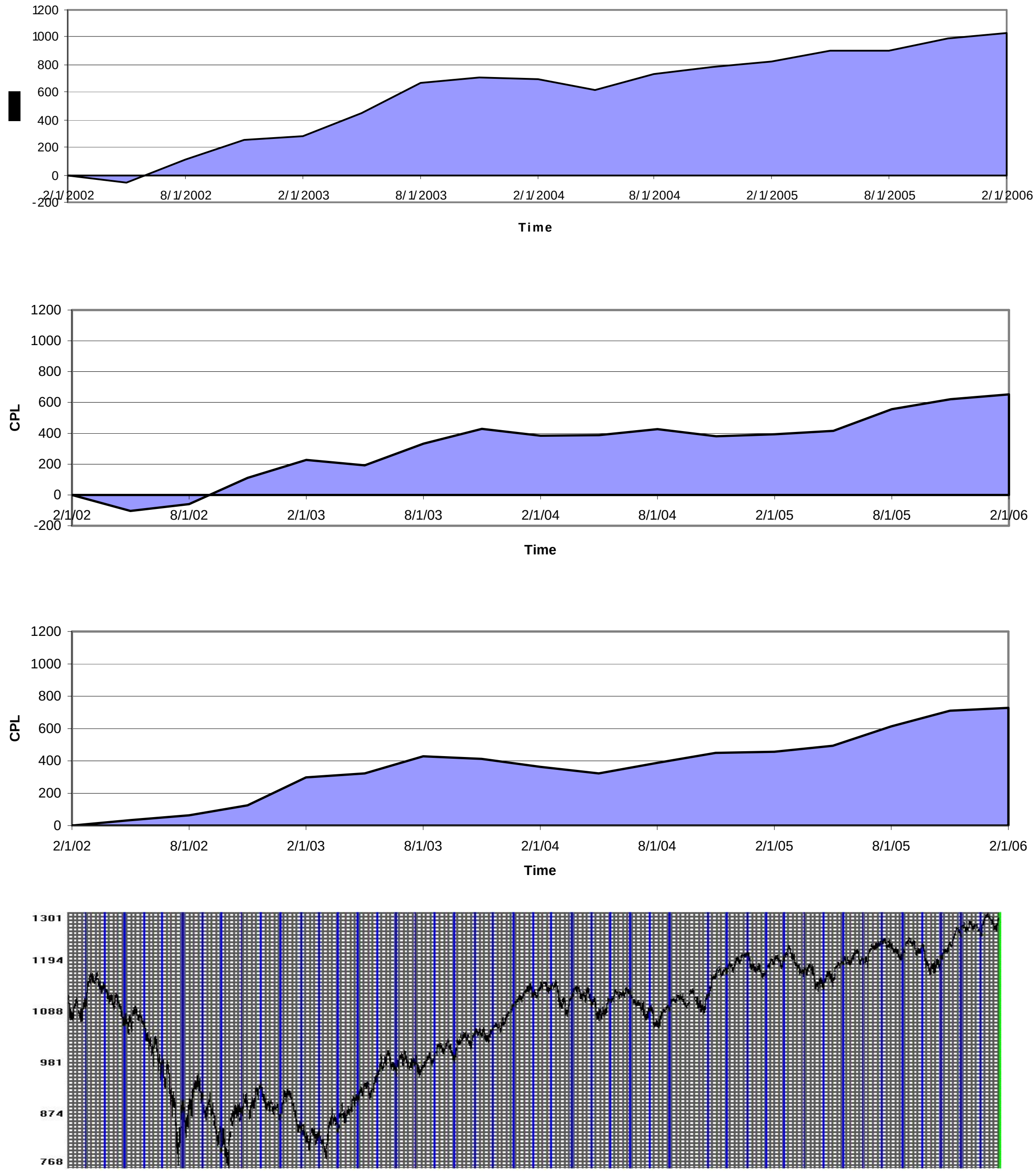


Figure 4.5 Comparison of C2 cycle trading results with neural network for S&P 500 futures from 02/2002 to 02/2006. The top three graphs show the results for 6-month, 9-month and 1-year training windows, respectively. The bottom graph shows the relative price chart.

4.4.2 C5 Cycle Trading Results

In this section, we give the results of applying the network to C5 cycle trading. The trading and test are same as C2 cycle trading. The C5 cycle trading results with the training window as 6-month, 9-month and 1-year are presented in Table 4.4, 4.5 and 4.6, respectively. As before, the graphic comparison of the C5 trading results with different training windows is illustrated in Figure 4.6.

According to the results, the network with the 9-month training window is better than the networks with the 6-month and 1-year as the training windows. This result is substantially different from that for C2 cycle trading when the 6-month window outperforms the longer ones. The reason may have something to do with the number of training data. As described before, we get three sets of inputs/desired outputs training data for every trading cycle. Since the length of C5 cycle (one-week cycle) is twice as that of C2 cycle (remember that C2 is a code for a half-week cycle), the number of C5 training data is only about half of that of C2 during the same training window period. In any case, the average performance of the ANN applied to C5 cycle is substantially worse than that for C2 cycle, suggesting that the ANN can capture the rhythm of the C2 cycles better than that of the weekly (C5) cycle.

Compared to the C5 cycle trading results for the TFx methods (refer to Chapter 2), the current ANN method is better than the TF3, TF4 and TF5 methods, but worse than the other TFx methods. This suggests that current ANN design doesn’t combine the TFx methods well enough to take advantage of their strengths and overcome their weaknesses for C5 cycles.

The ANN is a complicated system, which involves many components affecting its effectiveness. Since this was an initial effort to apply the ANN to combine those TFx methods, there still remains a lot to explore in future research. We can try to use different network structure, i.e. feed-backward structure, different inputs, different learning, etc.

Table 4.4 C5 cycle trading results for S&P 500 futures from 02/2002 to 02/2006. The training period is 6-month, and the test period is 3-month. "CPL Sum" is the sum of previous CPLs.

Number	Training Period	Test Period	Long	Short	CPL	CPL Sum
1	08/01/01~01/31/02	02/01/02~04/30/02	-25.8	3.2	-22.6	-22.6
2	11/01/01~04/30/02	05/01/02~07/31/02	144.1	331.1	475.2	452.6
3	02/01/02~07/31/02	08/01/02~10/31/02	-71	-131	-202	250.6
4	05/01/02~10/31/02	11/01/02~01/31/03	19.2	108.7	127.9	378.5
5	08/01/02~01/31/03	02/01/03~04/30/03	52.7	-20.3	32.4	410.9
6	11/01/02~04/30/03	05/01/03~07/31/03	48.2	-9.5	38.7	449.6
7	02/01/03~07/31/03	08/01/03~10/31/03	37.6	-15.5	22.1	471.7
8	05/01/03~10/31/03	11/01/03~01/31/04	10.8	-57	-46.2	425.5
9	08/01/03~01/31/04	02/01/04~04/30/04	-58.7	-30.7	-89.4	336.1
10	11/01/03~04/30/04	05/01/04~07/31/04	-21.2	-0.2	-21.4	314.7
11	02/01/04~07/31/04	08/01/04~10/31/04	30.7	-60.3	-29.6	285.1
12	05/01/04~10/31/04	11/01/04~01/31/05	36.9	42.9	79.8	364.9
13	08/01/04~01/31/05	02/01/05~04/30/05	-21.1	0.6	-20.5	344.4
14	11/01/04~04/30/05	05/01/05~07/31/05	32	-29.2	2.8	347.2
15	02/01/05~07/31/05	08/01/05~10/31/05	-8.8	40.8	32	379.2
16	05/01/05~10/31/05	11/01/05~01/31/06	-1.4	-39.2	-40.6	338.6

Table 4.5 C5 cycle trading results for S&P 500 futures from 02/2002 to 02/2006. The training period is 9-month, and the test period is 3-month. “CPL Sum” is the sum of previous CPLs.

Number	Training Period	Test Period	Long	Short	CPL	CPL Sum
1	05/01/01~01/31/01	02/01/02~04/30/02	11	17	28	28
2	08/01/01~04/30/02	05/01/02~07/31/02	164	318.5	482.5	510.5
3	11/01/01~07/31/02	08/01/02~10/31/02	19.6	-35.4	-15.8	494.7
4	02/01/02~10/31/02	11/01/02~01/31/03	15.9	72.2	88.1	582.8
5	05/01/02~01/31/03	02/01/03~04/30/03	66.6	-4.5	62.1	644.9
6	08/01/02~04/30/03	05/01/03~07/31/03	50.1	0.7	50.8	695.7
7	11/01/02~07/31/03	08/01/03~10/31/03	2.3	-46.6	-44.3	651.4
8	02/01/03~10/31/03	11/01/03~01/31/04	0.5	-65.8	-65.3	586.1
9	05/01/03~01/31/04	02/01/04~04/30/04	-58.8	-30.4	-89.2	496.9
10	08/01/03~04/30/04	05/01/04~07/31/04	-25.1	-9.1	-34.2	462.7
11	11/01/03~07/31/04	08/01/04~10/31/04	46.5	-45.5	1	463.7
12	02/01/04~10/31/04	11/01/04~01/31/05	56.4	72	128.4	592.1
13	05/01/04~01/31/05	02/01/05~04/30/05	6.2	45.3	51.5	643.6
14	08/01/04~04/30/05	05/01/05~07/31/05	57	-3.4	53.6	697.2
15	11/01/04~07/31/05	08/01/05~10/31/05	-18.8	29.2	10.4	707.6
16	02/01/05~10/31/05	11/01/05~01/31/06	3.5	-29.4	-25.9	681.7

Table 4.6 C5 cycle trading results for S&P 500 futures from 02/2002 to 02/2006. The training period is 1-year, and the test period is 3-month. CPL sum is the sum of previous CPLs.

Number	Training Period	Test Period	Long	Short	CPL	CPL Sum
1	02/01/01~01/31/02	02/01/02~04/30/02	-12.1	21.4	9.3	9.3
2	05/01/01~04/30/02	05/01/02~07/31/02	89.6	259.1	348.7	358
3	08/01/01~07/31/02	08/01/02~10/31/02	-2.6	-70.4	-73	285
4	11/01/01~10/31/02	11/01/02~01/31/03	-33.2	41	7.8	292.8
5	02/01/02~01/31/03	02/01/03~04/30/03	65.4	2.1	67.5	360.3
6	05/01/02~04/30/03	05/01/03~07/31/03	21.8	-25	-3.2	357.1
7	08/01/02~07/31/03	08/01/03~10/31/03	-3.7	-46	-49.7	307.4
8	11/01/02~10/31/03	11/01/03~01/31/04	4.9	-63.3	-58.4	249
9	02/01/03~01/31/04	02/01/04~04/30/04	-41.6	-14	-55.6	193.4
10	05/01/03~04/30/04	05/01/04~07/31/04	-34.7	7.3	-27.4	166
11	08/01/03~07/31/04	08/01/04~10/31/04	22.5	-70.6	-48.1	117.9
12	11/01/03~10/31/04	11/01/04~01/31/05	38.5	51.4	89.9	207.8
13	02/01/04~01/31/05	02/01/05~04/30/05	-18.8	4.5	-14.3	193.5
14	05/01/04~04/30/05	05/01/05~07/31/05	58	3.8	61.8	255.3
15	08/01/04~07/31/05	08/01/05~10/31/05	-19.3	28.4	9.1	264.4
16	11/01/04~10/31/05	11/01/05~01/31/06	0.1	-63.1	-63	201.4

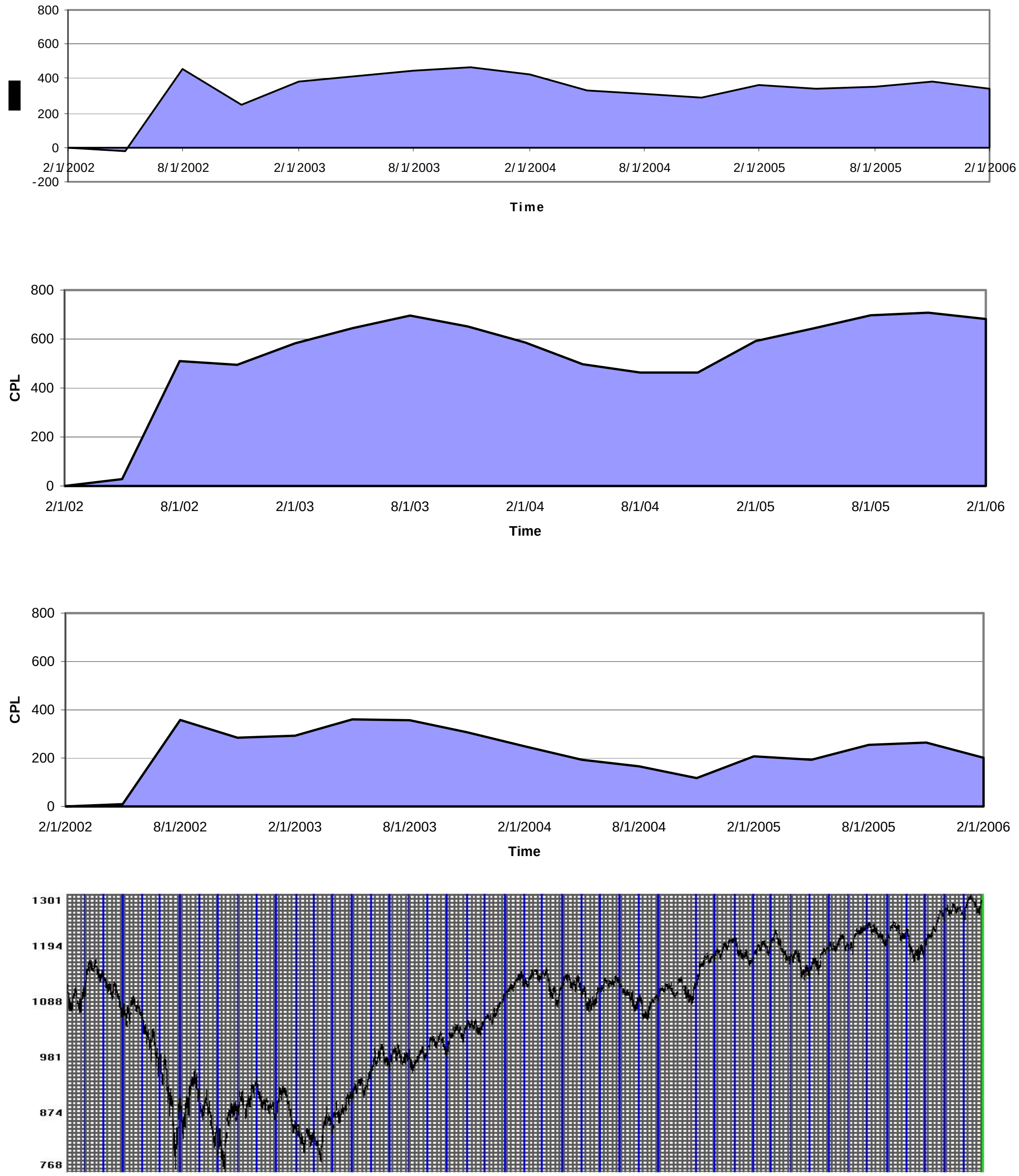


Figure 4.6 Comparison of C5 cycle trading results with neural network for S&P 500 futures from 02/2002 to 02/2006. The top three graphs show the results for 6-month, 9-month and 1-year training windows, respectively. The bottom graph shows the relative price chart.

4.5 Summary

In this chapter, we made initial efforts to apply a neural network to a combination of cycle-based market forecasts. As described in previous chapters, we used a set of previously developed TFx methods, which are cycle-based pattern recognition methods. As the artificial neural network provides a great way to incorporate different information non-linearly, it is used in this research to combine the TFx methods, and thus to come up with a new trading strategy, at least for the C2 cycle, the results are promising.

Summarizing this study, we chose the sigmoid, a feed-forward network, as the network due to its simplicity. In addition to the passing rates from the TFx methods, two other cycle-related parameters, time leg ratio and price leg ratio, were utilized as the inputs of the network. Supervised learning has been applied in the network training because of the availability of the inputs/desired outputs training data, which are pre-treated to be within the range of 0 and 1.

We applied the network to C2 and C5 trading for S&P futures during 02/2002 and 02/2006. A moving time window has been used in the training and testing of the network. The test window is 3-month, while the training window is 6-month, 9-month and 1-year, respectively. According to the trading results, the network trained with the 6-month window work best for C2, while for C5 trading where the results are less promising overall, the performance of the 9-month training window is incrementally better than the results of other two.

In this study, we only use the simple network in our trading system. There are many parameter choices which can be subject to further variation. For example, different network structures can be utilized, and the inputs can come from other sources, etc.

Chapter 5 Use of Moving Averages in Financial Market Analysis

5.1 Overview

The moving average is one of the most popular technical indicators used in financial market analysis[83-85]. By smoothing out price fluctuations as “noise” in market data, use of suitable moving average makes it possible to detect the market trend, which can be helpful for trading in a volatile market. Moving averages can also be used to provide support and resistance level identification, and an (imperfect) trading signal, etc.

Generally, the markets can be categorized as trending, cycling or oscillating. While it can be profitable over the whole 6-year period studied, to model the market as a whole as cycling as discussed in Chapters 2~4, it can’t be optimal because both alternative market modes (trending and oscillating) do not show identifiable cycles. One would like to be able also to forecast markets that are in these modes. This is where the moving average can come in helpfully. Based on the definition of the moving average, it will always be behind the current market movement; thus, it can be viewed as a trailing indicator, and belongs to the category of trend-following indicators**.** When the market is in trending mode, a correctly chosen moving average usually works well; however, such a moving average will provide many misleading signals when the market is in one oscillation mode, and this limits its usefulness severely.

Many trading rules based on moving averages have been developed and applied in financial market forecasts. A common trading rule is that a buy or sell signal is generated when the present market passes above (or below) a chosen moving average. The shortcoming of this basic trading rule is that too many false signals are generated during oscillating periods, and the performance of the method is negative. A need exists to forecast oscillating market.

In this chapter, I am going to introduce an innovative use of the moving average[86] to cover this need. Contrary to the common trading rule, two carefully chosen moving averages with different time length will be used to produce a trading signal in a “reverse” way. The method is applied to S&P futures market, and surprisingly, provides a powerful predictor over the 5-year period covered.

Section 5.2 introduces the principle and usage of moving averages. Section 5.3 presents novel use of the moving average in the reverse mode and discusses the trading results. Section 5.4 gives a summary for the method and draws some conclusions.

5.2 General Characteristics of Moving Averages

A time series is a sequence of well-defined data items obtained through repeated measurements over time. For example, S&P futures data with 10-minute as the interval, used in the study of the thesis, form a time series. Time series analysis is often used to project future values by observing how the value of a variable has changed in the past.

Usually, there is some form of “random” variation in financial data time series. An often-used technique is smoothing, which when properly applied, reveals more clearly an underlying trend. Forming moving averages is a kind of smooth technique, which smoothes out price fluctuations as "noise", and allows recognition of the direction of a trend where one exists.

Generally, there are two types of moving averages: the Simple Moving Average (SMA) and the Weighted Moving Average (WMA). SMA is the unweighted mean of the previous n data points, as shown in equation 5.1.

$$SMA = \Sigma P_t / n \qquad \textbf{5.1}$$

Where
P_t: value of data point t
n: the number of data points

The calculation is repeated for each data point of the time series. The averages are then joined to form a smooth curving line - the moving average line. Figure 5.1 shows a 20-day moving average line superimposed on a JP Morgan stock price chart from 12/26/2007 to 12/26/2008. (Note that this recent period also contains a strong downward trend which could, e.g. be brought out by a longer-term moving average.)

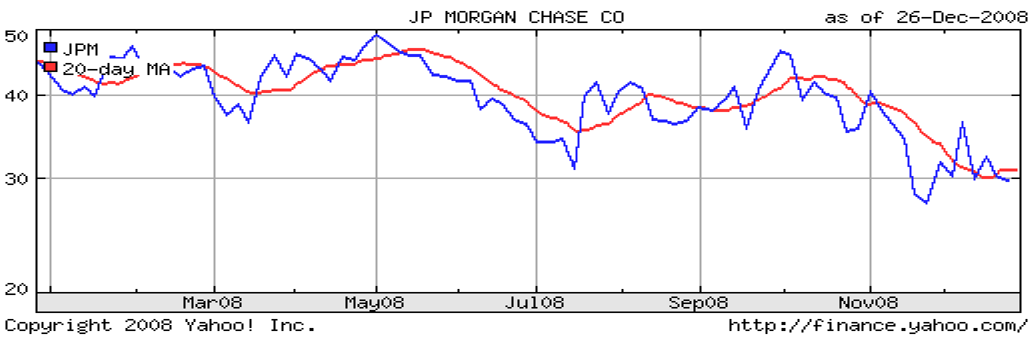


Figure 5.1 JP Morgan stock price chart from 12/26/2007 to 12/26/2008. 20-day simple moving average line (red line) is on top of stock price line (blue line).

The WMA is an average that applies different weights to different data points as shown in equation 5.2.

$$WMA = \Sigma P_t W_t / \Sigma W_t \qquad \textbf{5.2}$$

where
W_t: weight of data point t

The Exponential Moving Average (EMA), a popular moving average, is a special case of a WMA. The weighting factor for each older data point decreases exponentially, giving much more importance to recent observations while still not discarding older observations entirely. The equation 5.3 shows the formula for EMA.

$$EMA(t) = EMA(t - 1) + K \times [P(t) - EMA(t - 1)] \qquad \textbf{5.3}$$

Where
t: current data point
t –1: previous data point
K: weighting factor, set often equal to 2 / (n + 1)
P(t): current data point value

As the price pay for its smooth nature, a moving average lags behind the latest data point. Depending on the number N of data points being used, the SMA will lag by N/2 points, and is proportionately influenced by old data points. An EMA reduces the lag by applying more weight to recent data points relative to older data ones; the closer an old data point is to present data point, the more weight will be applied. Compared to an SMA, an EMA has a relatively smaller lag. As such, when applied to a financial market, an EMA will react quicker to recent price changes than an SMA, but at the price of greater noise of forecasting.

Moving averages are utilized in three areas of forecasting: trend identification, trade signal, and identification of support and resistance levels.

A common trend identification technique uses the direction of a suitable moving average to determine the trend. When a moving average is rising or declining, the trend is correspondingly considered as up or down. The direction of a moving average can be determined by its slope or by looking at a plot of the moving average.

As mentioned above, the most commonly used way to generate a trade signal is to compare the moving average price to the underlying price or as a proxy of the latter, a fast moving (short) moving average. When the underlying price rises above its moving average, a buy signal appears; if the price falls below its moving average, a sell signal is generated. Such moving average signals are most effective in a trending market and less effective when the market is oscillating. Once in a trend, moving averages will stay in until a breakout occurs, but they are trend following indicators that will always be a step behind the market movement, thus, they give late entry and exit signal.

A third, more technical use of moving averages is to identify support and resistance levels. This is not treated here.

5.3 Application of Moving Averages

5.3.1 Alternative Novel Use of Moving Averages to Operate in Oscillating Markets

As discussed above, moving averages can be used to generate a trading signal. In the standard method, the intersection between the moving average and the underlying price is used to produce buy or sell signals. Sometimes, two moving averages with different lengths are used to generate the trading signal. Figure 5.2 shows an example of two slow-moving (long-time) moving average lines, EMA(30 days) and EMA (100 days) applied to an Inter-Tel stock price chart. When the 30-day moving average moves above the 100-day moving average, a buy signal appears; conversely, when the 30-day moving average moves below the 100-day moving average, it creates a sell signal. Superficial examination of this chart shows this very show MA indicators to be generally “right”; however, detailed study shows that at the intersection time the actual market has already advanced or retreated considerably making this signal much less useful; in any case it does not relate to fine features (local oscillations) of the market.

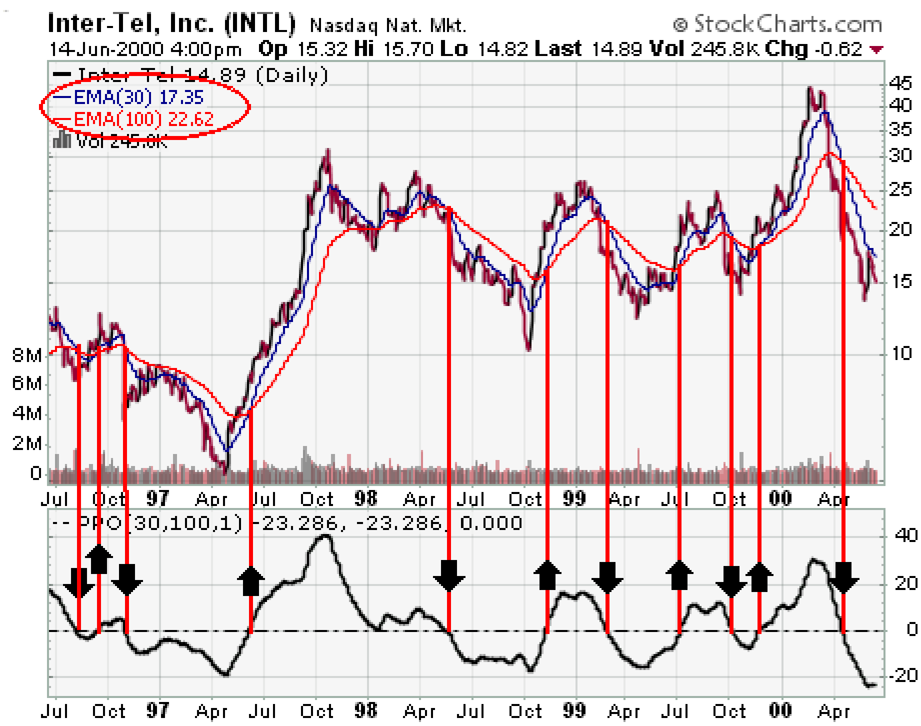


Figure 5.2 The Inter-Tel, Inc stock chart with two moving averages[85]. The blue line represents 30-day moving average, and the red line means 100-day moving average. When the blue line crosses the red line from below, it signals a buy. Conversely, when the blue line crosses the red line from above, it signals a sell. The difference of both MAs is plotted below the chart; a market action is taken when this difference goes through zero.

To obtain trading signals based on fast-direction-changing (oscillating) market moves, moving average signals of this kind are generally out of step with the market and produce false signals and losses. Therefore, we reverse the standard way of signal generation, where we obtain buy or sell signals when the shorter-term moving average (faster) is crossing the longer-term (slower) moving average from below or above. As before, we generate moving averages, which are SMA in the study, and record their intersections, but in our novel forecasts use of the moving average, we utilize the reverse mode, which is opposite to the standard mode (the direct mode), to generate the trading signal. When MA's of the order of hours rather than days are used, this approach produces forecasts over extended periods.

I then turned to tuning the moving average parameters used, since two moving averages are used in this prediction system, we conduct an optimization test of the reverse mode combination of two moving averages with different lengths. Table 5.1 shows a rectangle map for moving average combinations from 1 to 59 hours for S&P futures from 02/01/2001 to 07/01/2001. In the map, the values of the vertical axis represent the length of the short moving average, while the values of the horizontal axis corresponding to the length difference between the long and short moving averages. For instance, (4, 6) means that the short moving average is 4 hours, and the long moving average is 10 hours. Therefore, the length of the short moving average is between 1 and 30 hours, while the length of the long one in this test is between 2 and 60 hours. The number in the cells of the map represents Cumulative Profit & Loss (CPL), which is obtained by using the direct trading mode (described above), for the corresponding moving average combination in the given period. Positive numbers in the CPL cell signify that the direct decision method would have produced the listed CPL profit and negative numbers show that the reverse decision would have been profitable instead. Examination of Table 5.1 shows that these are extended regions of such negative CPL values on the chart for which the reverse trading mode would be the correct one to use. The complex relationship of direct and reverse parameter fields is not the subject of my study here; instead, I want to establish whether the reverse mode MA parameters optimized in a given period can be used successfully in a subsequent period. This is described in the next section.

From the map, we find that in terms of CPL, the best combination in the terms of the coordinates of this chart is (1, 4) (i.e., MA’s 1 and 5 hrs) with CPL as 308.2. As discussed before, we can apply the reverse mode in the moving average trading. The only different between direct and reverse mode is the trading direction, which means buy signal in direct mode will become sell signal in reverse mode, vice verse. Thus, the trading results should be exactly opposite. Consequently, from the same table, we obtain the worst combination for the direct mode, at (29, 25) (i.e., MA's 29 and 54 hrs) with negative CPL as -303.3. However, in the reverse mode, the same combination becomes the best one with a positive CPL of 303.3.

Table 5.1 Trading results of moving average combination in direct mode for S&P 500 futures from 02/01/2001 to 07/01/2001. The values of the vertical axis represent the lengths of the short moving average, and the values of the horizontal axis represent the length differences between long and short moving averages. The numbers in the cells are CPL numbers obtained by market entries in the direct mode; negative numbers indicate that the reverse mode would have been profitable.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	-10.3	89.5	160	308.2	180	123.2	149.8	194.4	166.2	161	141.6	101.6	75.7	28.5	148.7	149.6	163.3	72.3	67.3	122.1	155.2	156.2	172.2	214.6	199.9	178.7	148.1	93.8	29.2	19
2	-37.9	194.5	253.2	108.2	144.2	141.3	83.9	123.7	58.1	18.5	-7.9	97.5	167.6	157.7	153.1	86.1	105.3	100.6	98.5	127.7	105.4	97.9	73.5	95.7	77.1	45.8	64.2	53.8	48.5	6.7
3	180.5	270.6	105.7	113	131.2	97.5	127.3	-69.9	8.8	108.7	131.3	142.4	106.1	91.7	110	100.2	81.8	133.6	162.9	170	84.2	80.7	68.8	95.2	36.2	4.4	-11.7	7.4	-37.4	-18
4	-19.5	117.1	198.8	161.4	104.6	109.5	22.8	45.5	117.3	24.5	72.7	72.5	73	74.1	126.5	122.5	90.6	102	102.8	25.2	56.7	59.8	42.2	-2.4	15.4	17.4	-21.2	-78.3	-52.3	-41.6
5	193	275.6	198.3	269.5	46.9	-13.5	18.1	85.1	102.1	56.3	108.4	69.9	104.7	126.4	108	130	89.9	73.2	65.2	72.9	108.6	38.9	12.1	35.2	28.2	-34.9	-27.6	-61.6	-28.9	-26
6	248.3	225.1	112.4	-3.3	47.3	24.6	28.8	71.6	63.5	43.8	123.3	113.1	143.4	145.7	106.6	51.7	72.4	92.5	60.5	63.6	29.4	-0.7	-43.1	-26.8	-16	-5	-38.2	-47.5	-76.7	-111
7	113.2	-79.1	-116.2	-19.8	-41.3	-85	39	-13	41.9	90	142.6	72	76.7	39.3	27.3	14.8	-22.5	-24.3	-38.2	-8.4	-19	-64.7	-22.3	-32.1	-25.6	-31.7	-43.3	-74	-140.3	-123.6
8	-190.8	-41.4	31.4	-24.7	-11.4	-19.5	-51.3	-8.3	-20.4	19.3	-5.6	22.9	-48.3	-86.6	-55.4	-67.3	-33.1	-30	-58.4	-33.5	23.6	25.3	40.8	11.2	-32.5	-45	-46.2	-57.5	-30.5	0.6
9	-54	23.8	3.1	-52.5	-73.5	-69.4	-42.2	-90.7	-32.7	-60.6	-42.9	-115.7	-81.4	-108.6	-134.8	-12.3	11.6	-8.6	70.9	124.5	84.5	107.5	81.3	47.7	31.9	50.7	75.3	83.4	53.3	67.8
10	23.1	41.5	-80.4	-57.3	-76.4	-119.8	-136.3	-70.2	-74.2	-65.2	-136.3	-88.3	-82.8	-29.2	4.1	47.9	152	176.5	183.9	143.6	121.3	97.5	116.9	130.4	146.4	183.2	170.5	175.7	133.2	108.2
11	-65.4	-33.1	-93.6	-161.4	-171.9	-127.1	-164.6	-142.2	-139.1	-139.3	-80.3	-107.4	52.4	34.9	147.3	178.7	128.4	151.3	113.4	116.5	126.1	147.7	194.6	202.3	245.6	171.1	186.2	157.2	147.1	122.4
12	27.5	-142.5	-187.2	-213.5	-148.9	-77.7	-161.7	-153.7	-100.1	3.3	38	-13.5	118.2	199.1	216.9	182.3	203.3	211.7	212.1	215.3	185.7	221.2	166.9	117	139.1	148.5	160.5	163.6	136.1	147.7
13	-138.6	-156.2	-127.7	-149.8	-145.9	-167.3	15.1	32.1	-27.5	41.8	114.2	145.9	207.5	199.7	234	249.6	221.4	239.9	223.2	182	154.5	199.8	124.2	145.1	159.9	124.8	144.7	153.8	191.4	83.4
14	-144.1	-75.3	-91.9	-196.6	-168.5	-91.1	-16.9	-17.8	-2.6	90.3	164.1	210.6	192.5	201.9	201.4	215.4	205.8	225.1	164.4	192.3	160	128.9	145.2	135.7	83.8	83.8	57.3	51.6	47.2	67.9
15	-172.5	-207.8	-150.5	-115.5	-80.9	66.5	52.7	56.6	75.5	155.7	124.9	155.8	203.9	178.1	209.2	205.2	192.2	123.9	155.9	140	114	115.5	104.2	55.7	6.7	54.7	73.8	41.1	54.5	21.8
16	-164	-116.3	-72.5	53.1	179.6	134.4	69.6	116	175.2	182.4	186.8	192.5	218.8	189.1	157.8	153.4	145.3	101.7	70.8	61.4	21.7	0.4	-18.9	15.5	23.3	19.3	30.2	63.1	46.6	73.2
17	-26.6	-25.3	81.8	141.4	92.7	89.5	239.4	232.5	253.7	214.2	269.3	198.1	160	166.3	122.6	102.5	60.3	2.6	-50.5	-53.7	-10.6	24.5	44.3	35.5	44	56.7	-17.3	-7.3	3.7	-2.1
18	74.6	117.3	147	133.4	186.4	249	229.9	294	224.4	228.8	185.9	177.2	68.2	122.3	92.1	27.4	23.5	45	47.3	40.2	30.9	49.2	0	18.6	-1.2	-31.7	-32.2	-35	-50	-45.8
19	177.7	184.1	73.7	145.4	222.2	225.2	150.2	248.7	260.6	176.8	142.5	97.7	36	14.9	11.8	20.6	68.2	68.3	14.4	-13.8	0.8	26.1	0.7	17.1	-31.8	-29.4	-60.1	-31.3	-48.5	-70
20	172.6	41.3	170.5	139	163.1	164.3	245.2	198.6	177.5	88.4	52.9	-0.3	-23.5	28.8	55.3	89.3	57.8	19.3	16	31.5	18.7	7.9	-24.8	-37.4	-103.9	-107.2	-73.5	-66.8	-80.8	-118.1
21	124	166.1	194.1	186.7	236.2	197.4	195	124	70.7	46.1	-5.3	1.3	-36.1	23.3	10.2	21.4	17.2	42.4	-22.3	13	-27	-11.5	-80.3	-132.3	-149.1	-145	-156.5	-151.1	-177	-165
22	237.8	203.5	226.9	179.9	169.3	164	18.1	-13.8	3	-75.1	-101.3	-38.2	-10	51.9	33.1	27.8	33.9	-31.1	-17	-25.8	-85.3	-127.4	-160.2	-202.5	-199.7	-216.7	-212.8	-202.6	-213.4	-202.1
23	176.2	245.1	284.8	147.5	57.1	-43.7	-32.3	-75.6	-48.7	-51	-31.6	-49.7	42	44.7	14.3	-20	1.7	-53.7	-114.4	-99.6	-147.2	-116.4	-200.5	-115.9	-140.7	-160.9	-161.2	-159.9	-178.6	-186.9
24	249.9	240.2	98.6	-57.8	-116.1	-10.8	-24.4	-50.7	-40.5	-57.6	-39.3	-35.5	-28.4	-21.5	-20.5	-38.4	-86.3	-122.1	-122.9	-55.7	-60.9	-60	-52.2	-58.9	-79	-105.5	-132.4	-169.4	-199.2	-209.9
25	141.3	37.8	-87.9	-3.9	-7.5	-25.9	-10	-31.1	-30.7	-96.3	-21	-104.8	-63.7	-41.6	-65.9	-85.4	-88.1	-32.1	-55.4	-60	-87.2	-88	-87.4	-118.5	-155.3	-152.1	-196.3	-164.2	-194.2	-217.5
26	-72.8	-114.5	-160	-25.7	-9.4	-6.9	-23.9	-46	-82.1	-86	-81.1	-21.4	-66.2	-88.9	-78.2	-89.4	-97.8	-69.1	-51.6	-123.2	-92.8	-104.3	-156.4	-136.5	-179.9	-207.2	-198.9	-244.1	-231.2	-205.5
27	-184.1	-172.3	-50.1	-49.1	-17.7	12.7	-32	-108	-79	29.7	-34.9	-34.4	-33.2	-53.1	-37.1	-36.2	-75.6	-124.2	-170.3	-133.1	-159.2	-199.3	-241.1	-203.6	-237.6	-267.2	-250.9	-242.7	-275	-294.1
28	-111	-52.1	-33.3	-39.1	-89.2	-74.2	-39.2	-59	-46.3	-43.6	-63.7	-108.9	-72.5	-66.5	-93.8	-117.1	-147.3	-164.9	-174.1	-162.4	-180.1	-195.3	-246.3	-239	-237	-254.7	-248	-257.8	-266.4	-259.6
29	-59.3	-72.5	-31.6	-64.9	-46.6	-91.5	-73.2	-48.4	-90.1	-122.9	-117.4	-100	-139.5	-161.3	-142.1	-135.1	-115.8	-144.7	-153.4	-193.2	-225.4	-248.7	-254.1	-248	-303.6	-232.6	-234.2	-199.8	-229.9	-219.3
30	-60.5	-52.1	-19.7	-67	-91.6	-144.8	-168.3	-115.4	-149.7	-124.8	-143.3	-143.5	-178	-168.2	-132.9	-136	-165.4	-134.6	-147.9	-181.9	-190.3	-281.4	-277.7	-239.9	-225.8	-236.4	-214.7	-225.6	-196.9	-163.9

5.3.2 Results and Discussion

As seen above, in addition to the standard direct moving average trading method, in addition to the direct mode, we develop a reverse mode that operates in the opposite way. As discussed above, inspection of Table 5.1 had shown that both methods (trading modes) are capable, for a given quarter, of yielding nearly equal, large profits (correct forecasts) for their best parameters, we now want to compare the effectiveness of these two trading modes, we apply both of them to the S&P futures markets from 02/01/2001 to 01/31/2006 under equal conditions. Futures data will be separated into successive training period and testing period. In the training period, we conduct the optimization of the moving average combination, and obtain the “best” and “worst” combinations in terms of CPL. Then, we will apply these optimal combinations to the direct and reverse mode, respectively.

As to the optimization, a moving time window will be applied during the process. For example, we set the training period length as 6-month, and the test period length as 3-month. The procedure is as follows: In the optimization phase, we first use a 6-month period as the training period to obtain the “best” and “worst” moving average combination; then, in the test phase, we apply these optimal combinations to the following 3-month period. Once the trading in the test period is completed, we move to the optimization phase again; at this time, the 3-month test period will be added in to form the new training period, and the first 3-month in the old training period will be dropped to keep total training period still at 6-month. The training and testing will keep going until the end of the test database. By this way, we always use the 6-month right before the test period as the training period in order to include the latest market information.

In the following study, all optimizations in the training phase are processed with the “best’ and “worst” values obtained for the direct trading mode, and the tests can be conducted with the corresponding direct or reverse mode.

Table 5.2 contains the trading results with use of moving average direct trading mode for S&P 500 futures from 08/01/2001 to 01/31/2006. The training time window is 6-month, and the test time window is 3-month. “Training Period Best CPL” is obtained from the optimization of moving average combination described in section 5.3.1. For instance, for the training period from 02/01/2001 to 07/31/2001, the best combination is (1, 5) with a CPL of 308.2 S&P points. “Test Period CPL” shows the trading results when we apply this optimal combination to the test period. For example, the CPL of the test period from 08/01/2001 to 10/31/2001 is 142 by using the parameter combination (1, 5). “Test Period CPL Sum” is the sum of CPL of previous test periods. According to the table, in the direct mode, the direct CPL for the whole test period is –347.3 showing that despite its direct success in the initial 3 months period, the direct mode fails over the total period.

Table 5.3 presents the trading results using the moving average reverse trading mode. The training periods and testing periods are same as for the direct mode in Table 5.2. Instead of collecting the best CPL in each training period, Table 5.3 shows the worst direct mode combination and the corresponding worst CPL. As the reverse mode result is exactly the opposite of the direct mode result, the worst combination and its CPL in the direct mode will become the best combination and its CPL in the reverse mode. Based on the table, a positive CPL of 671.2 had been obtained for the whole test period.

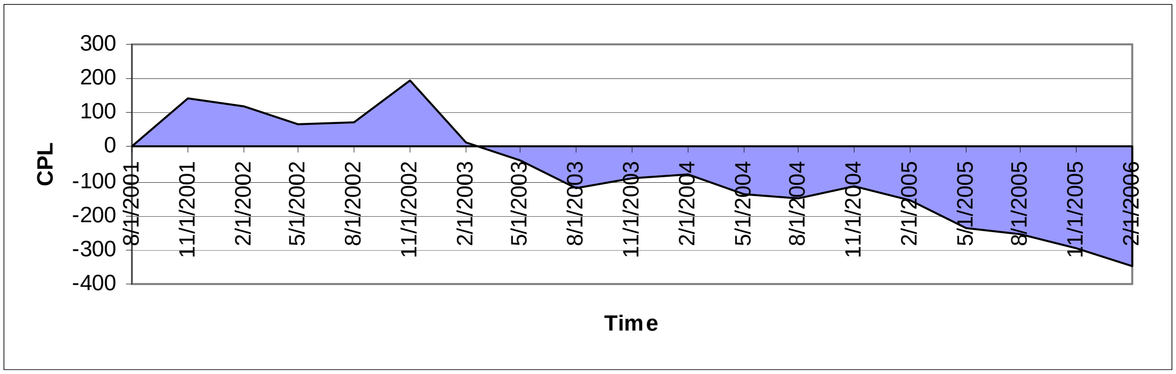
In order to visually compare the trading results difference between the direct and reverse modes, we plot them in Figure 5.3. The top graph shows the trading results for the direct mode, with data coming from “Test Period CPL SUM” in Table 5.3. The middle graph shows the reverse trading results from Table 5.3. The bottom graph shows the price chart for S&P for this period, from 08/01/2001 to 01/31/2006. From the graphs, it is easy to spot that the reverse mode performs much better than the direct mode. For the whole test period from 08/01/2001 to 01/31/2006, the reverse mode produces a profit as 671.2, which is about 1000 points more than the direct mode, which leads to loss for the last 2 years. The direct mode only makes profit in six out of 18 single test periods, while the reverse mode produces positive CPL in two thirds of the single test periods. The reverse mode works well in the middle of the test period from 11/01/2001 to 01/31/2005, then flattens out in the last year of the test period.

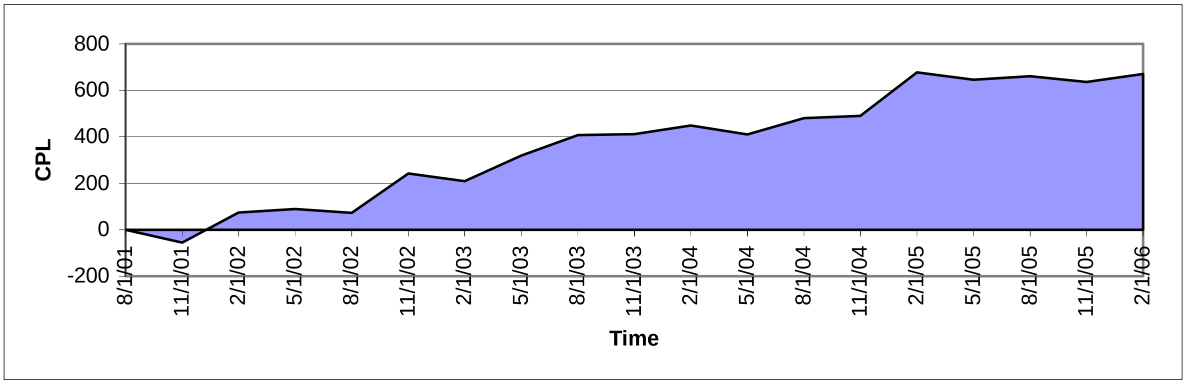
Table 5.2 Trading results with use of moving average combination in direct mode for S&P 500 Futures from 08/01/2001 to 01/31/2006. Optimal combination is the best combination, in terms of CPL, obtained in the training period. Test period CPL is obtained when the optimal combination is applied in the test period. Test period CPL sum is the sum of pervious CPL. Training period is always previous 6-month corresponding to test period.

Number	Training Period	Test Period	Optimal Combination	Training Period Best CPL	Test Period CPL	Test Period CPL Sum
1	02/01/01~07/31/01	08/01/01~10/31/01	(1, 5)	308.2	142	142
2	05/01/01~10/31/01	11/01/01~01/31/02	(5,9)	403.6	-23.4	118.6
3	08/01/01~01/31/02	02/01/02~04/30/02	(5,10)	267.1	-51.4	67.2
4	11/01/01~04/30/02	05/01/02~07/31/02	(7,19)	102.5	6	73.2
5	02/01/02~07/31/02	08/01/02~10/31/02	(3,4)	347.6	121.6	194.8
6	05/01/02~10/31/02	11/01/02~01/31/03	(3,4)	480.3	-181.2	13.6
7	08/01/02~01/31/03	02/01/03~04/30/03	(30,38)	240.3	-54.8	-41.2
8	11/01/02~04/30/03	05/01/03~07/31/03	(18,32)	182.8	-77.5	-118.7
9	02/01/03~07/31/03	08/01/03~10/31/03	(28,30)	120	26.5	-92.2
10	05/01/03~10/31/03	11/01/03~01/31/04	(25,26)	195.9	11	-81.2
11	08/01/03~01/31/04	02/01/04~04/30/04	(27,28)	161.5	-55.2	-136.4
12	11/01/03~04/30/04	05/01/04~07/31/04	(30,32)	139.1	-11.3	-147.7
13	02/01/04~07/31/04	08/01/04~10/31/04	(30,45)	78.5	33.3	-114.4
14	05/01/04~10/31/04	11/01/04~01/31/05	(26,28)	129.8	-39.9	-154.3
15	08/01/04~01/31/05	02/01/05~04/30/05	(10,11)	142.1	-81.6	-235.9
16	11/01/04~04/30/05	05/01/05~07/31/05	(5,6)	69.2	-20.8	-256.7
17	02/01/05~07/31/05	08/01/05~10/31/05	(27,36)	129.7	-36.7	-293.4
18	05/01/05~10/31/05	11/01/05~01/31/06	(30,38)	116.5	-53.9	-347.3

Table 5.3 Trading results with use of moving average combination in reserve mode for S&P 500 Futures from 08/01/2001 to 01/31/2006. Training period represents the optimization period, in which optimal moving average combination with worst training period CPL (using the direct mode) is obtained. Test period CPL (using the reverse mode) is obtained when the optimal combination is applied in the test period. Test period CPL sum is the sum of pervious CPL. Training period is always previous 6-month corresponding to test period.

Number	Training Period	Test Period	Optimal Combination	Training Period Worst CPL	Test Period CPL	Test Period CPL Sum
1	02/01/01~07/31/01	08/01/01~10/31/01	(29, 54)	-303.6	-55.8	-55.8
2	05/01/01~10/31/01	11/01/01~01/31/02	(26,38)	-343.4	129.5	73.7
3	08/01/01~01/31/02	02/01/02~04/30/02	(17,31)	-350.9	16	89.7
4	11/01/01~04/30/02	05/01/02~07/31/02	(24,54)	-245.4	-16.6	73.1
5	02/01/02~07/31/02	08/01/02~10/31/02	(22,23)	-213.5	168.5	241.6
6	05/01/02~10/31/02	11/01/02~01/31/03	(22,23)	-345.1	-31.9	209.7
7	08/01/02~01/31/03	02/01/03~04/30/03	(1,2)	-430.3	109.5	319.2
8	11/01/02~04/30/03	05/01/03~07/31/03	(1,4)	-297.9	87.9	407.1
9	02/01/03~07/31/03	08/01/03~10/31/03	(12,19)	-289.7	4.6	411.7
10	05/01/03~10/31/03	11/01/03~01/31/04	(2,12)	-263.2	37.5	449.2
11	08/01/03~01/31/04	02/01/04~04/30/04	(4,6)	-223.1	-39.4	409.8
12	11/01/03~04/30/04	05/01/04~07/31/04	(2,4)	-215	70.6	480.4
13	02/01/04~07/31/04	08/01/04~10/31/04	(7,30)	-239.8	9.5	489.9
14	05/01/04~10/31/04	11/01/04~01/31/05	(1,4)	-203.9	187.1	677
15	08/01/04~01/31/05	02/01/05~04/30/05	(1,4)	-344.6	-31.6	645.4
16	11/01/04~04/30/05	05/01/05~07/31/05	(1,3)	-183.1	15.6	661
17	02/01/05~07/31/05	08/01/05~10/31/05	(8,18)	-232.3	-24.5	636.5
18	05/01/05~10/31/05	11/01/05~01/31/06	(4,33)	-219.2	34.7	671.2





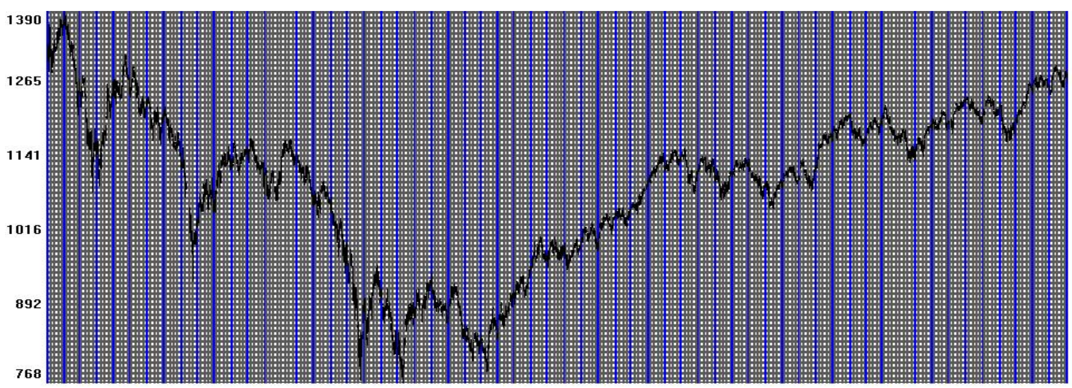


Figure 5.3 Trading results of moving average combination in direct and reserve modes for S&P 500 futures and price chart during the related time period from 08/01/2001 to 01/31/2006. The top graph shows the results in direct mode (Table 5.2), and the middle graph shows the results in reverse mode (Table 5.3). The training time window is 6-month, and the test time window is 3-month. The bottom graph present the corresponding price chart for S&P from 08/01/2001 to 01/31/2006.

Next, I studied the effect of the length of the training period on the result. The trading results, presented in Table 5.2 and 5.3, were based on the 6-month training period length. In order to evaluate the effect of the length of the training period, I used a 1-year as the training period length.

Table 5.4 corresponds to Table 5.2 in that it shows the trading results with use of moving average direct trading mode from 02/01/2002 to 01/31/2006, but instead of 6-month, 1-year is used as the length of the training period. This generates a loss of -328.7 (equal to (328.7) in financial parlance) in the whole test period.

Table 5.5 which corresponds to Table 5.3 presents the trading results using the reverse mode rule with 1-year as the training period length. From the table, this mode makes profit of 472.9 during the whole test period.

Again, we present the 1-year training rule trading results for the direct and reverse mode in Figure 5.4, which corresponds to Figure 5.3 for the shorter, 6-month training period. Same observation can be made that the reverse mode is still much better than the direct mode, by almost 800 S&P points more in terms of CPL. The direct mode only produces a positive CPL in one quarter of the 3-month test periods, while the reverse mode makes a fortune in 13 out of the 16 single test periods. The reverse mode works well in the period from 02/01/03 to 01/31/2006, but not well during the 6-quarter period before it.

Compared to the direct mode, the reverse mode performs significantly better in both test scenarios. Meanwhile, the 6-month training length scenario has a better performance than the 1-year training length scenario. The reason may be that shorter training length may contain more similar behavior to the test period, and therefore, it can provide a better combination guild for the test period.

While use of different training and testing windows and procedures (not done beyond the training periods studied above) may lead to somewhat different results, the trading result comparison shows that the reverse mode performs better than the direct mode, and in the absence of prior knowledge about the currently correct market mode (trending, cycling or oscillating), can be a mode of market forecast alternative to the cycle modes discussed in Chapter 2~4.. The reason can be explored by considering the nature of the moving average, which is time delayed. As we know, the moving average is obtained by adding up previous data, and the average then set as the current value. Therefore, it will always take past data into consideration and be time delayed. The longer the length of moving average, the larger the time delay will be. When the market is in an oscillation mode, there are many quick turning points, which are not easily caught by the moving average direct mode due to this time delay. However, when two relatively short moving averages are considered instead, it is very easy to get crossover signals between them due to rapid oscillations of the market, and usually give false trading signals. However, the reverse trading mode based on relatively long moving averages will validly capture the situation, and produce better results in this market environment.

Table 5.4 Trading results with use of moving average combination in direct mode for S&P 500 futures from 02/01/2002 to 01/31/2006. Optimal combination is the best combination, in terms of CPL, obtained in the training period. Test period CPL is obtained when the optimal combination is applied in the test period. Test period CPL sum is the sum of pervious CPL. Training period is always previous 1-year corresponding to test period.

Number	Training Period	Test Period	Optimal Combination	Training Period CPL	Test Period CPL	Test Period CPL Sum
1	02/01/01~01/31/02	02/01/02~04/30/02	(5,9)	511.6	-64.9	-64.9
2	05/01/01~04/30/02	05/01/02~07/31/02	(5,9)	377.2	295.9	231
3	08/01/01~07/31/02	08/01/02~10/31/02	(1,7)	601.8	-32.3	198.7
4	11/01/01~10/31/02	11/01/02~01/31/03	(3,4)	418.5	-181.2	17.5
5	02/01/02~01/31/03	02/01/03~04/30/03	(1,22)	351.5	-71.4	-53.9
6	05/01/02~04/30/03	05/01/03~07/31/03	(3,4)	411.4	-78.2	-132.1
7	08/01/02~07/31/03	08/01/03~10/31/03	(28,30)	293.5	26.5	-105.6
8	11/01/02~10/31/03	11/01/03~01/31/04	(28,30)	332.4	78.4	-27.2
9	02/01/03~01/31/04	02/01/04~04/30/04	(28,30)	254.1	-27.5	-54.7
10	05/01/03~04/30/04	05/01/04~07/31/04	(28,29)	267.1	-17.3	-72
11	08/01/03~07/31/04	08/01/04~10/31/04	(30,32)	173.1	18.8	-53.2
12	11/01/03~10/31/04	11/01/04~01/31/05	(30,33)	202.5	-96.6	-149.8
13	02/01/04~01/31/05	02/01/05~04/30/05	(4,13)	126.1	-41.3	-191.1
14	05/01/04~04/30/05	05/01/05~07/31/05	(17,25)	135.9	-81.1	-272.2
15	08/01/04~07/31/05	08/01/05~10/31/05	(28,36)	158	-13.3	-285.5
16	11/01/04~10/31/05	11/01/05~01/31/06	(12,15)	131.8	-43.2	-328.7

Table 5.5 Trading results with use of moving average combination in reserve mode for S&P 500 futures from 02/01/2002 to 01/31/2006. Training period represents the optimization period, in which optimal moving average combination with best training period CPL (using the direct mode) is obtained. Test period CPL (using the reverse mode) is obtained when the optimal combination is applied in the test period. Test period CPL sum is the sum of pervious CPL. Training period is always previous 1-year corresponding to test period.

Number	Training Period	Test Period	Optimal Combination	Training Period CPL	Test Period CPL	Test Period CPL Sum
1	02/01/01~01/31/02	02/01/02~04/30/02	(13,19)	-392.5	-123.7	-123.7
2	05/01/01~04/30/02	05/01/02~07/31/02	(26,38)	-547.5	27.5	-96.2
3	08/01/01~07/31/02	08/01/02~10/31/02	(18,34)	-491	17.4	-78.8
4	11/01/01~10/31/02	11/01/02~01/31/03	(22,37)	-388.1	-81.8	-160.6
5	02/01/02~01/31/03	02/01/03~04/30/03	(14,16)	-265.7	70.5	-90.1
6	05/01/02~04/30/03	05/01/03~07/31/03	(14,17)	-378.4	65.3	-24.8
7	08/01/02~07/31/03	08/01/03~10/31/03	(12,13)	-465.8	-3.1	-27.9
8	11/01/02~10/31/03	11/01/03~01/31/04	(1,4)	-418.6	92	64.1
9	02/01/03~01/31/04	02/01/04~04/30/04	(12,19)	-371	34	98.1
10	05/01/03~04/30/04	05/01/04~07/31/04	(1,4)	-331.2	70.9	169
11	08/01/03~07/31/04	08/01/04~10/31/04	(1,6)	-342	74.8	243.8
12	11/01/03~10/31/04	11/01/04~01/31/05	(2,3)	-330.1	132.6	376.4
13	02/01/04~01/31/05	02/01/05~04/30/05	(1,3)	-458.4	1.1	377.5
14	05/01/04~04/30/05	05/01/05~07/31/05	(1,4)	-353.1	31	408.5
15	08/01/04~07/31/05	08/01/05~10/31/05	(1,3)	-337.9	33.2	441.7
16	11/01/04~10/31/05	11/01/05~01/31/06	(5,19)	-278.4	31.2	472.9

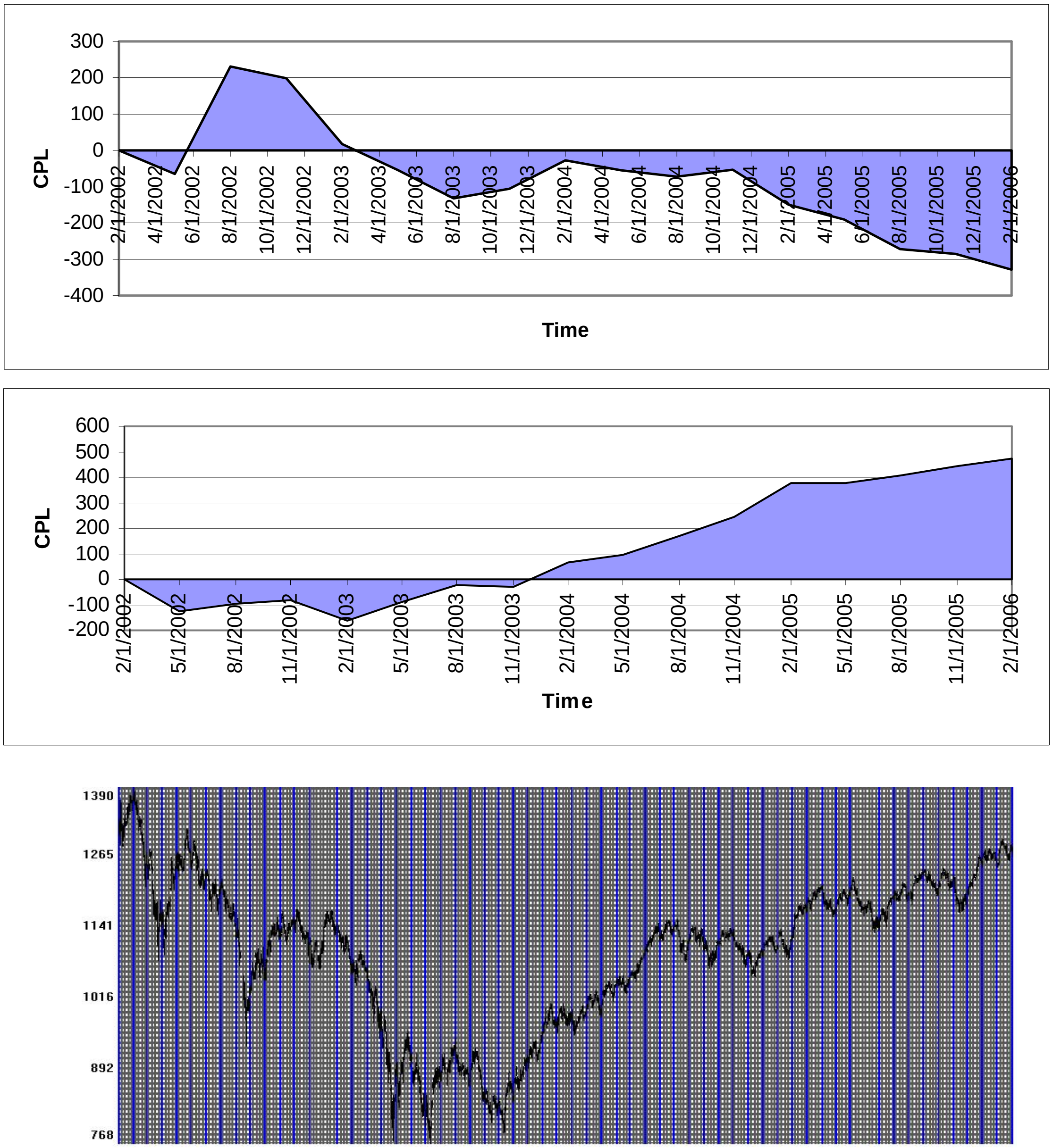


Figure 5.4 Trading results of moving average combination in direct and reserve modes for S&P 500 futures and price chart during the related time period from 08/01/2001 to 01/31/2006. The top graph shows the results in direct mode (Table 5.5), and the middle graph shows the results in reverse mode (Table 5.6). The training time window is 1-year, and the test time window is 3-month. The bottom graph present the corresponding price chart for S&P from 08/01/2001 to 01/31/2006.

5.4 Summary

In this chapter, I have discussed the development and application of a novel use of moving averages. Usually, a moving average is used in the direct mode, which uses the crossover of two moving averages to generate a trading signal such that when the short-term moving average is crossing above or below the long-term moving average, a buy or sell signal will be generated respectively. We created a reverse mode for the moving average, which generates the totally opposite trading signal to the direct mode.

Before using the moving average trading rules, we first optimize the moving average combinations to find the best and worst in terms of CPL in the direct mode. This optimization is done for a moving time window for the S&P futures market. Here, once the test is completed, the test period will be added into the new training period, while dropping a previous old training period with the same length. Based on the optimization results, we apply the direct and the reverse mode to the test period. Using either 6-month or 1-year as the length of the training period, we found that the reverse mode not only performs better that the direct mode, but we show that it captures essentials of the S&P index market behavior during a 5-year period.

Chapter 6 Summary and Outlook for Future Development

6.1 Summary

The research reported in this thesis applied chemometrics-derived pattern recognition and artificial intelligence methodologies to the analysis of a futures market (S&P index as an example). We have achieved significant progress in the area of futures market cycle analysis where many have doubts about its effectiveness due to apparent "randomness" in the market price movements. In our research, we find that the market shows some cyclic nature and can be predicted to a considerable extent through pattern recognition and other well-developed models.

In chapter 1, I conduct a review of futures and futures markets, fundamental and technical analysis, chemometrics and some of its derived methods, and cycle theory, as used in our research. Meanwhile, we also introduce the FTVision system, an in-house, flexible and versatile "workbench", where our research and market analysis are conducted.

In chapter 2, I present a family of TFx forecast models, which are cycle-based pattern recognition methods, and demonstrate the predictability of futures market through the models. The K-Nearest Neighbors (KNN) method, a chemometrics-derived pattern recognition algorithm, has been applied in the forecast models. By choosing K analogous historical cycles and using a "voting" rule, we can make a forecast to the present cycle development. The results show that there exists a correlation between historical price trajectories and the possibilities of discovering and utilizing such features in S&P futures market forecasts. A further investigation on the effectiveness of prior knowledge of market trends shows that correct control cycle phase control can significantly improve the performance of the TFx methods because it pre-selects correct historical cycle backgrounds.

In chapter 3, I apply Bayes' Theorem to the use of short-term observables in the determination of control cycle phase, and prove the validity of this procedure in the performance enhancement to the TFx methods. Four short-term market characteristics, namely daily up-or-down sequences, weekly up-or-down sequences, short-term cycle leg-length-ratio sequences and short-term cycle extreme up-or-down sequences, have been utilized to infer the present market long-term cycle phase through Bayesian classification, another pattern recognition method. The predicted long-term cycle has been applied in the background selection of the TFx methods. We obtain some encouraging results, especially for weekly up-or-down and short-term cycle extreme up-or-down, in the TFx trading with the aid of a predicted control cycle.

In chapter 4, I introduce an artificial neural network into our cycle-based market analysis. As the TFx cycle analysis methods can be classified as coincident and trailing indicators, the artificial neural network has been used to incorporate signals from many cycle phase indicators with the aim of improving their strengths and reducing their weaknesses. As an initial effort, a three-layer, feed-forward neural network with nine inputs, which are seven TFx methods and two other cycle-related measures, has been used for this purpose. The results are comparable to the performance of the single best TFx method. Further research about the use of the artificial neural network is recommanded.

In chapter 5, I describe an novel use of moving averages, one of the most popular used technical indicators, in the futures market analysis. Through moving time window optimization, two moving averages with different lengths are chosen in the forecast, but, contrary to common use of moving average combinations, a reverse mode has been employed in the trading. The initial success of the reverse mode trading shows that it partly avoids the weakness of moving average, which is too many false signals in oscillation market periods, trading in instead an inability to deal with trending markets and moving the trading uncertainty to a decision between two decision modes.

The methodologies developed in this thesis are innovative and effective in futures market analysis. Although the results are not yet perfect for all models, we hope that the approaches employed in our study shed light on future cycle-based market analysis.

6. 2 Future Development

Based on the research conducted in this thesis, there remain many opportunities in further considerations of present work. The further development (some of which are already being worked on by my colleagues) can be summarized as follows:

1. Improving accuracy of long-term cycle prediction by combining some Bayes' statistics or using other statistics. For example, other criteria than up-or-down can be used, i.e., comparison of average price of the day or week; instead of a week, the number of days can be changed to obtain better statistics.
2. Exploring further the use of artificial neural network in the combination of cycle-related information. Many network components such as network

structure, layers, learning rules and inputs, etc., can be changed to find appropriate ones.

3. Enhancing the application of moving averages in futures market analysis Introducing pattern recognition in the selection of optimal moving average to include more relative information for decision-making.

4. Combining cycle-based analysis (i.e., the TFx methods) and oscillation-based analysis (i.e., moving average) to reduce the weakness of the individual methods. When the market is in oscillating model, more weight will be assigned to oscillation-based indicators; while in a cyclic market, more emphasis needs to be given to cycle-based indicators.

5. Applying other chemometrics-derived methods such as component analysis, factor analysis to seek dominant features of market information. Meanwhile, other pattern recognition methods such as support vector machine (SVM) can also be applied in cycle-based market analysis.

References

1. Hull J.C., *Options, Futures, and Other Derivatives.* 2003: Prentice-Hall.
2. Mishkin F.S. and Eakins S.G., *Financial Markets and Institutions*. 2000, New York: Addison-Wesley.
3. Reilly F.K. and Brown K.C., *Investment Analysis and Portfolio Management.* 2003, U.S.:Thomson South-Weatern.
4. Kolb R.W., Overdahl J., and Bertola G., *Understanding Futures Markets*. 2006, Blackwell Publishing Professional.
5. Schwager J.D., *A Complete Guide to the Futures Markets: Fundamental Analysis, Technical Analysis, Trading, Spreads, and Options*. 1984, New York: Wiley.
6. Stein J.L., *The Economics of Futures Markets*. 1986, Cambridge: Billing $ Sons Ltd.
7. Gwartney J.D., Stroup R.L., Sobel R.S. and Macpherson D.A., *Economics Private & Public Choice*. 2003, U.S.: Thomoson South-West.
8. Bisset S., *Putting Energy into the Market.* Futures, 2002: p.66-69.
9. O'Sullivan T., *Dollar/Swiss: Still a Safe-Haven Trade*? Futures, 2002: p.34-35.
10. Lo A., Mamaysky H., and Wang J., *Foundation of Technical Analysis: Computational Algorithms, Statistical Interface, and Empirical Implementation.* The Journal of Finance, 2000, LV(4): p. 1705-1765.
11. Blume L., Easley D., and O'Hara M., *Market Statistics and Technical Analysis: The Role of Volume*. Journal of Finance, 1994, 49(1): p.153-181.
12. Brown D.P. and Jenings R.H., *On Technical Analysis*. Review of Financial Studies, 1989, 2(4): p.527-551.
13. Curcio R., Goodhart C., Guillaume D., and Rayne R., *Do Technical Trading Rules Generate Profits? Conclusions from the Intra-Day Foreign Exchange Market*. Int. J. Fin. Econ., 1997, 2(4).
14. Edwards R.d., and Magee J., *Technical Analysis of Stock Trends*. 1992: John Magee Inc.
15. Ring G., *Is the 'Flow' Turning Back to Commodities*. Futures, 2002: p. 40-43.

16. Ramazan G., *Linear, Nonlinear and Essential Foreign Exchange Rate Prediction with Simple Technical Trading Rules*. Journal of International Economics, 1999, 47(1): p.91-107.

17. Lukac L.P., Brorsen B.W., and Irwin S.H., *The Usefulness of Historical Data in Selecting Parameters for Technical Trading Systems.* Journal of Futures Markets, 1989, 9: p.55-65.

18. Neftci S.N., *Native Trading Rules in Financial Markets and Wiener-Kolmogorov Prediciton Theory: A Study of "Technical Analysis"*. Journal of Business, 1991, 64(4): p.549-571.

19. Neftci S.N., and Policano A.J., *Can Chartists Outperform the Market? Market Efficiency Tests for "Technical Analysis"*. Journal of Futures Markets, 1981, 4:p.465-478.

20. Roberts H.V., *Stock Market "Patterns" and Financial Analysis: Methodological Suggestions*. Journal of Finance, 1959, 14(1): p.1-10.

21. http://www.thestockbandit.com/Uptrend-stock.htm accessed by March 15, 2009

22. http://stockcharts.com/school/doku.php?id=chart_school:chart_analysis:chart_patterns:double_top_reversal accessed by March 15, 2009

23. Murphy J.J., *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*. 1999, New York: New York Institute of Finance.

24. Bulkowski T.N., *Encyclopedia of Chart Patterns*. 2005, Wiley.

25. Silber W., *Technical Trading: When It Works and When It Doesn't*. The Journal of Derivatives, 1994, (spring): p.39-44.

26. Jobman d., *How Traders Can Exploit Volatility in the Energy Markets*. Futures, 2002: p. 26-32.

27. Plummer, T., *The Psychology of Technical Analysis: Profiting from Crowd Behavior and the Dynamics of Price.* 1993, McGraw-Hill.

28. Malkiel B.G., *A Random Walk Down Wall Street*. 2003, W.W.Norton & Company.

29. Jensen M., and Benington G.A., *Random Walks and Technical Theories: Some Additional Evidence.* Journal of Finance, 1969, 25(2): p.469-492.

30. Lo A.W., and MacKinlay A.C., *A Non-Random Walk Down Wall Street*. 2001, Princeton University Press.

31. Lo A.W., *The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective.* 2004: Boston.

32. Bessembinder H. and Chan K., *Market Efficiency and the Returns to Technical Analysis.* Financial Management, 1998, 27(2): p.5-17.

33. Treynor J.L. and Ferguson R., *In Defense of Technical Analysis*. Journal of Finance, 1985, 40(3): p.757-773.

34. Pruitt S.W. and White R.E., *The Crisma Trading System: Who Says Technical Analysis Can't Beat the Market?* Journal of Portfolio Management, 1998: p.55-58.

35. Osler C., Support for Resistance: *Technical Analysis and Intraday Exchange Rates.* FRBNY Economic Policy Review, 2000, p.53-69.

36. Luca C., *Trading Global Currency Markets*. 2000, New York Institute of Finance: New York.

37. Stevenson R., and Bear R., *Commodity Futures: Trends or Random Walk*. Journal of Finance, 1970, 25: p.65-81.

38. Chen J., *An Introductory Study of Methods Derived from Chemometrics Applied to the Technical Analysis of Silver Futures Market Movements.* 2003, Northeastern University: Boston.

39. Yao J., *Statistical Analysis and Forecast of Futures Markets Using Chemometrics Derived Methods.* 2006, Northeastern University: Boston.

40. Xu K., *The Application of Chemometrics-Derived Methods to Better Cycle-Based Technical Analysis of Financial Market Movements*. 2007, Northeastern University: Boston.

41. Zhao Z., *Chemometrics-Derived Methods and Statstical Techniques to Model and Forecast Futures Markets*. 2008, Northeastern University: Boston.

42. Giessen B.C., Zhao Z., Chen J., Yao J., Xu K. and Yu T., *Successful Price Cycle Forecasts in S&P Futures using "TFX", a family of Pattern Recognition Algorithms Based on the KNN Method.* The Third Nikkei Econophysics Symposium. 2004, Tokyo, p. 116-120.

43. Xu k., Chen J., Yao J., Zhao Z., Yu T. and Giessen B.C., *Short Time Segment Price Forecasts Using Spline Fit Interactions.* The Third Nikkei Econophysics Symposium. 2004, Tokyo, p. 111-115.

44. Yao J., Chen J., Xu K., Zhao Z., Yu T., Giessen B.C. and Dadkhah K., *Market Cycle Turning Point Forecasts by an Interactively Learning Algorithm as a Trading Tool for S&P Futures*. The Third Nikkei Econophysics Symposium. 2004, Tokyo, p.

45. Otto M., *Chemometrics: Statistical and Computer Application in Analytical Chemistry*. 1999, New York: Wiley-Vch.

46. Jurs P.C., Kowalski B.R., Lenhour T.L., and Reilly C.N., *Computerized Learning Machines Applied to Chemical Problems.* Anal. Chem., 1969, 41: p.690-695.

47. Jurs P.C., Kowalski B.R., Lenhour T.L., and Reilly C.N., *Computerized Learning Machines Applied to Chemical Problems:Interpretation of Infrared Spectrrometry Data.*. Anal. Chem., 1969, 41: p.1945-1953.

48. Jurs P.C., Kowalski B.R., Lenhour T.L., and Reilly C.N., *Computerized Learning Machines Applied to Chemical Problems: Multicategory Pattern Classification by Least Squares.* Anal. Chem., 1969, 41: p.695-700.

49. Massart D.L., *Handbook of Chemometrics and Qualimetrics*. 1997, New York: Elsevier.

50. Yang M.H., Kriegman D.J., and Ahuja N., Detecting Faces in Images: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(1): p.34-58.

51. Zhang J., and Tan T., *Brief Review of Invariant Texture Analysis Methods*. Pattern Recognition, 2002, 35(3): p.735-747.

52. Gutierrez-Osuna R., *Pattern Analysis for Machine Olfaction: A Review*. IEEE Sensors Journal, 2002, 2(3): p.189-202.

53. Samal A., and Iyengar P.A., *Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey. Pattern Recognition*, 1992, 25(1): p.65-77.

54. Theodoridis S., and Koutroumbas K., Pattern Recognition. 2006, Elsevier.

55. Duda R.O., Hart P.E., and Stork D.G., *Pattern Classification*. 2006, New York: Wiley.

56. Kowalski B.R. and Bender C.F., *Pattern Recognition. A Powerful Approach to Interpreting Chemical Data*. J. Am. Chem. Soc., 1972, 94: p.5632-5639.

57. Sasic S. and Ozaki Y., *Statistical Two-Dimensional Correlation Spectroscopy: Its Theory and Applications to Sets of Vibration Spectra.* Anal. Chem., 2001, 1073(10): p. 2294-2301.

58. Lavine B.K., Davidson C.E. and Westover D.J., *Spectral Pattern Recognition Using Self-Organizing MAPS*. J. Chem. Inf. Comput. Sci., 2004, 44(3): p. 1056-1066.

59. Mendelsohn L.B., *Leading-Edge Quantitative Approaches – Chaos Theory, Neural Networks, and Genetic Algorithms*. Association for Investment Management and Research Annual Conference, 1992.

60. Wasserman P.D., Neural Computing. 1989, New York: Van Nostrand Reinhold.

61. Zubritsky E., *Feature-How Analytical Chemists Saved the Human Genome Project: January 2002*. Anal. Chem., 2002, 74(1): p.23A-26A.

62. Epstein J.R. and Ferguson J.A., *Combinational Decoding: An Approach for Universal DNA Array Fabrication.* J. Am. Chem. Soc., 2003, 125(45): p.13753-13759.

63. Figeys F., *Proteomics Approaches in Drug Discovery*. Anal. Chem., 2002, 74(15): p.413A-419A.

64. Mohimen A., Dobo A., Hoerner J.K, and Kaltashov I.A., *A Chemometric Approach to Detection and Characterization of Multiple Protein Conformers in Solution Using Electrospray Ionization Mass Spectrometry*. Anal. Chem., 2003, 75(16): P.4139-4147.

65. Hasegawa T., *Detection of Minute Chemical Species by Principal Component Analysis.* Anal. Chem., 1999, 71(15): p.3085-3091.

66. Grate J.W. and Wise B.M., *A Method for Chemometric Classification of Unknown Vapors from the Responses of an Array of Volume-Transducing Sensors.* Anal. Chem., 2001, 73(10): p.2239-2244.

67. Zwanziger H.W. and Sarbu C., *Validation of Analytical Methods Using a Regression Procedure*. Anal. Chem., 1998, 70(7).

68. Bernstein J., *Cyclic Analysis in Futures Trading*. 1998, John Wiely & Sonss, Inc. p.16.

69. Barsky R.B., and Miron J.A., *The Seasonal Cycle and the Business Cycle*. Journal of Political Economy, 1989, 97(3): p.503-534.

70. http://www.kwaves.com/kond_overview.htm accessed by March 15, 2009.

71. Alesina A., Roubini N., and Cohen G., *Political Cycles and the macroeconomy*. 1997, MIT Press.

72. Bayes, T., *An Essay Toward Solving a Problem in the Doctrine of Chances*. Philosophical Transactions of the Royal Society of London,1764, **53**, p.370-418.

73. Hartigan, J. A., *Bayes Theory*. 1983, New York: Springer-Verlag.

74. Swinburne, R., *Bayes' Theorem*. 2002, Oxford: Oxford University Press.

75. Haykin S., *Neural Networks.* 1999, Prentice Hall.

76. Gurney K., *An Introduction to Neural Networks.* 1997, UCL Press.

77. Fausett L., *Fundamentals of Neural Networks*, 1994, Prentice-Hall.

78. Rojas, R., *Neural Networks: A Systematic Introduction.* 1996, Berlin:Springer.

79. Trippi, R.R., Turban, E., *Neural Networks in Finance and Investing.* 1993, Irwin Professional Publishing.

80. Fishman M.B., Barr D.S., Loick W.J., *Using Neural Nets In Market Analysis*, 1991, Technical Analysis of Stocks & Commodities.

81. Yao J., Tan C.L, and Poh H.L., *Neural Network for Technical Analysis: A Study on KLCI.* International Journal of Theoretical and Applied Finance, 1999, 2(2): p.221-241.

82. McCulloch, W. S. and Pitts, W. H. *Bulletin of Mathematical Biophysics*, 1943, 5:115-133.

83. Fong W., and Yong L., Chasing Trends: Recursive Moving Average Trading Rules and Internet Stocks. Journal of Empirical Finance, 2005, 12(1): p.43-76.

84. Alessio E., Carbone A., Castelli G., Frappietro V., *Second-Order moving average and scaling stochastic time series*. The European Physical Journal B, 2002, 27(2), p. 197-200.

85. http://stockcharts.com/school/doku.php?id=chart_school:technical_indicators:moving_averages accessed by March 15, 2009

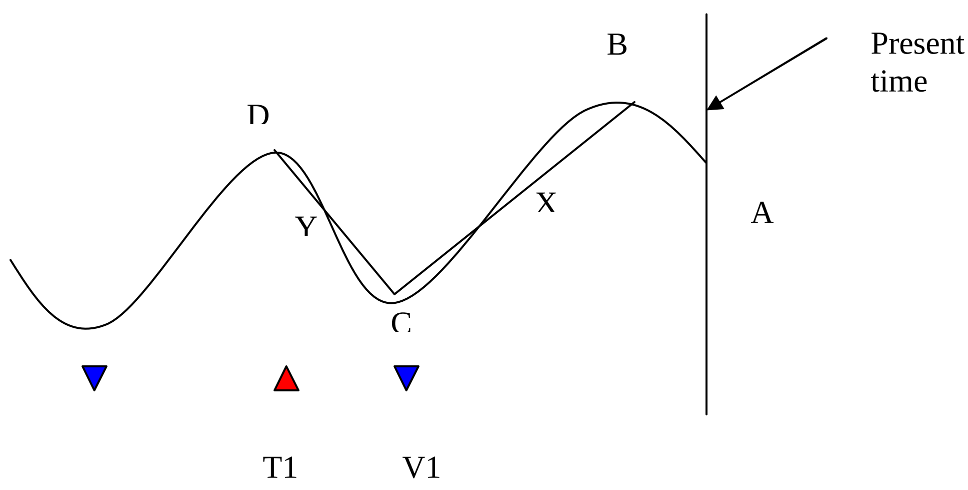
86. Giessen, B.C., Liu S.L. and Yu. T., U.S. Patent Application. 2006, Northeastern Unversity:Boston.

Appendix A. Definition of TFx Methods (TF1, TF2, TF3, TF4 and TF5)

In Chapter 2, we introduced a family of the TFx models, which are cycle-based pattern recognition methods. The K-Nearest Neighbors (KNN), a kind of pattern recognition algorithms, is applied in the models to choose historical analogous price trajectories to make forecast for present market movements. Based on the specific characteristics chosen and the manner of patter analysis, a family of the TFx methods has been developed. Among them, the TV1 and TF2B methods were introduced in chapter 2, and the other methods had been presented in Dr. Yao, Dr. Xu and Dr. Zhao's theses[39-41]. Here we give the simple introduction of those methods.

In the TF1 methods (Figure A.1), the X and Y coordinates are obtained as follows:

$X = P_B – P_C$
$Y = P_C – P_D$
Where:
P: market price at the specific time.



(TF1 method)

Figure A.1 Idealized sketch of markers used in the TF1 method.

In the TF2 methods (Figure A.2), the X and Y coordinates are obtained as follows:

$X = (P_A - P_C) / (P_D - P_C)$
$Y = (P_D - P_C) / (P_D - P_E)$

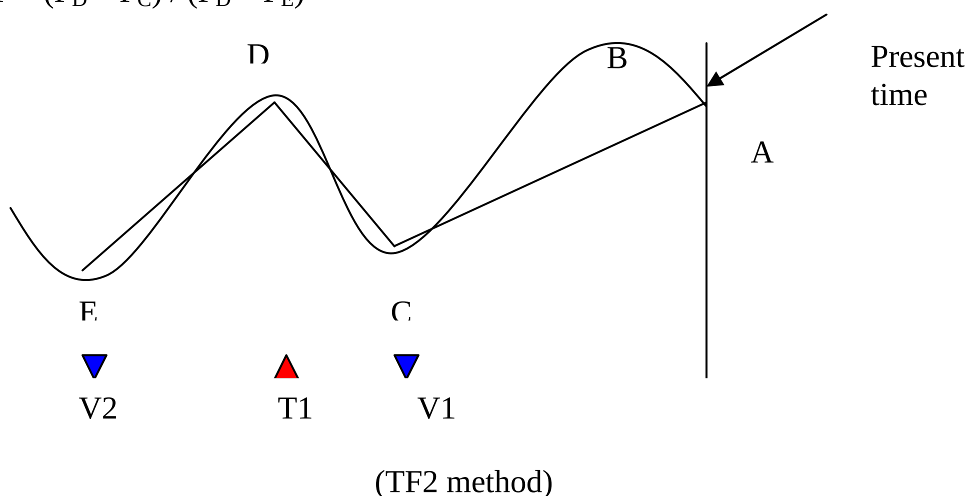


Figure A.2 Idealized sketch of markers used in the TF2 method.

In the TF3 methods, the X and Y coordinates are obtained as follows:

$X = (P_B - P_A) / (P_B - P_C)$
$Y = (P_B - P_C) / (P_D - P_C)$

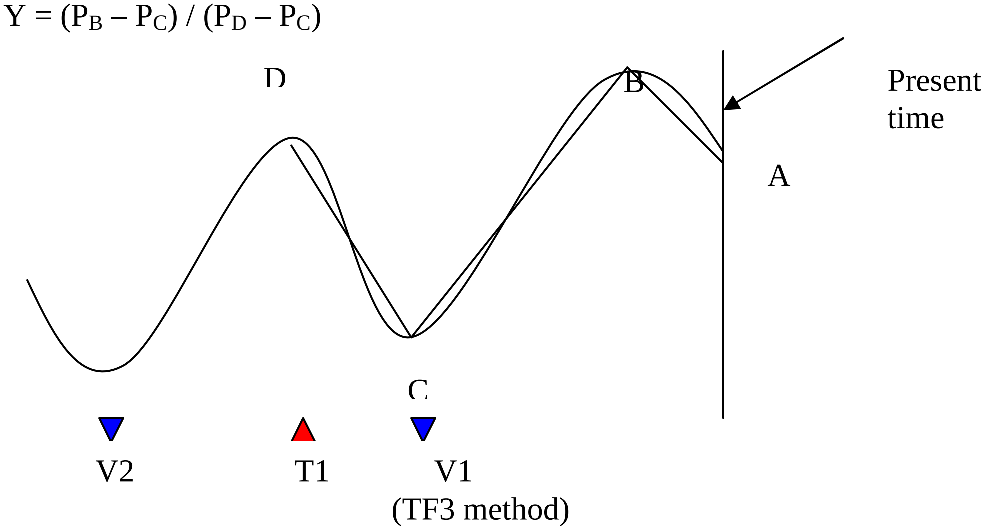


Figure A.3 Idealized sketch of markers used in the TF3 method.

In the TF4 methods, the X and Y coordinates are obtained as follows:

$X = P_B - P_A$
$Y = P_B - P_C$

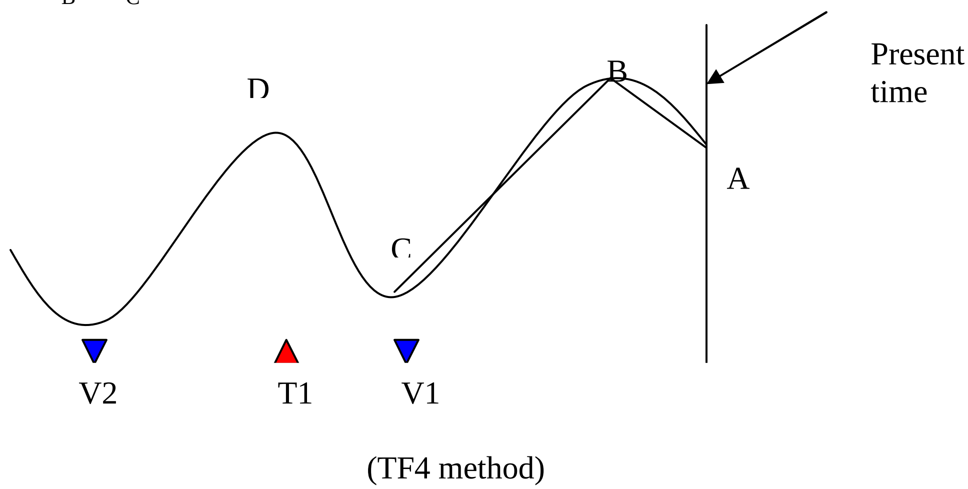


Figure A.4 Idealized sketch of markers used in the TF4 method.

In the TF5 methods, the X ,Y and Z coordinates are obtained as follows:

$X = P_A - P_B$
$Y = P_B - P_C$
$Z = P_C - P_D$

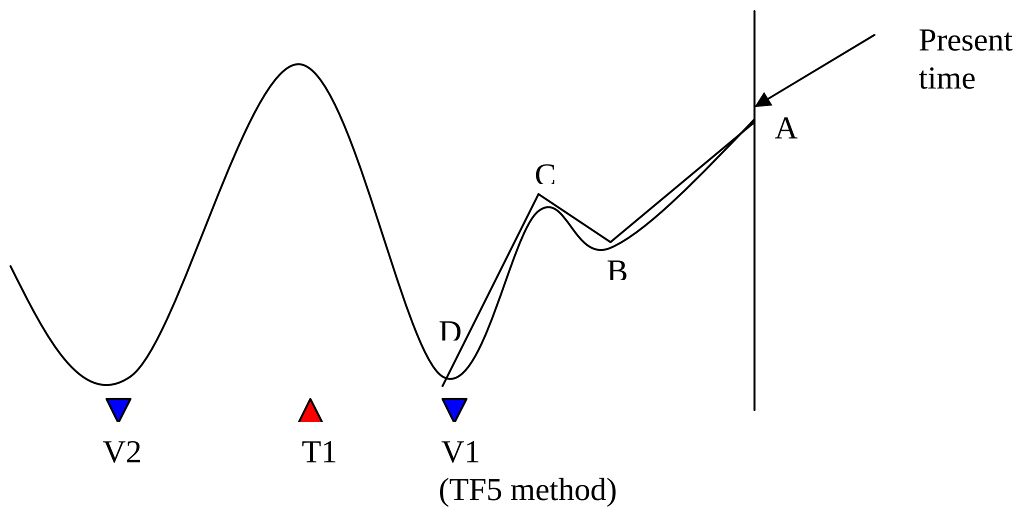


Figure A.4 Idealized sketch of markers used in the TF5 method.

Appendix B. The Trading Procedure of The TFx Methods (Cycle Determination in Real Time Mode)

In the TFx methods, we use historical cycle database as the pool so that the analogous cycles can be selected and used to guild present cycle movements. Here historical cycles come from “man-made” cycles that are designated by experienced cycle observer following some specific time and price rules.

The market markers used in the TFx methods are from real time self-determined cycles. In real time automatic trading, we can’t rely on human to recognize and update the previous several cycle points. The first reason is that it is not good for automatic trade because it needs intervention of people. The second reason is that there is no repeatability because every time different people may have different view to cycles, even same person can have different view when having incomplete information.

We developed a set of real time cycle recognition rule to help the TFx trade. Using the tools, such as spline fit and cycle score, developed by Dr. Xu[40], the real time cycle determination procedure is described with the following real time example using TF1 method:

1. In real time mode, the cycle recognition starts from some known cycles. As shown in Figure B.1, the vertical green line indicates the present time of the system. At that time, we already know previous top T_{-1} and V_{-1}. The system plans to predict the next top.

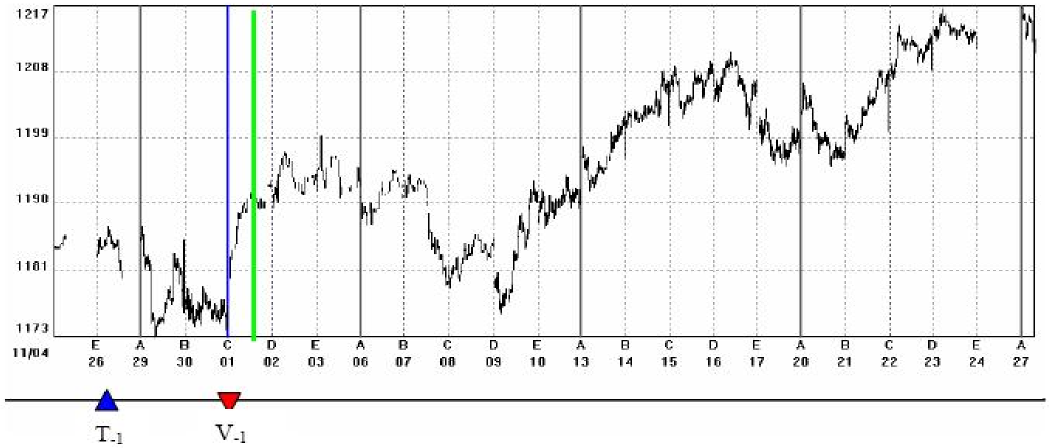


Figure B.1 Example of real time C5 cycle auto-recognition and self-adjustment.

2. At the time shown by green line in Figure B.2, the system generated a trade signal that predicted the present time is the cycle top based on TF1. Thus, the system closed the present long trade and immediately initiated a short trade. At the same time, the system identified a cycle top T_0 based on the specific spline fitting method and used this top to predict next cycle valley.

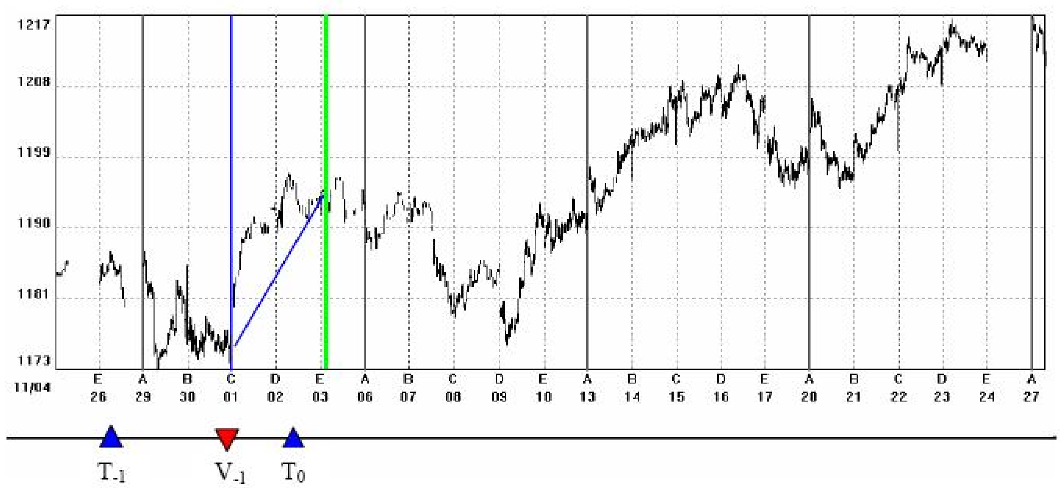


Figure B.2 Example of real time C5 cycle auto-recognition and self-adjustment.

3. At the time shown by green line in Figure B.3, the system generated another trade signal that determined the present time is the cycle valley. Then, the system closed the short trade and started a long trade. It also added a new valley V_0. Using the Triangle method, the system adjusted top T_0 to a new position.

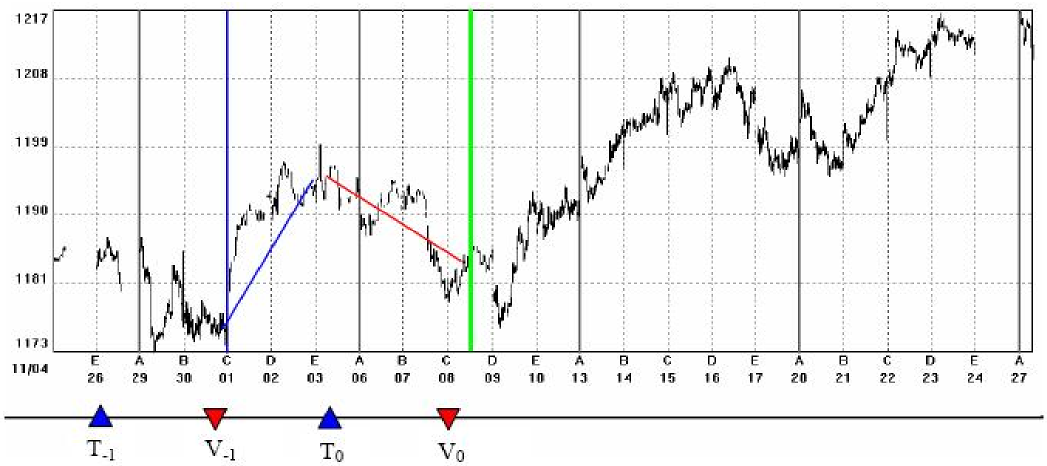


Figure B.3 Example of real time C5 cycle auto-recognition and self-adjustment.

4. At the time shown by green line in Figure B.4, the cycle valley V_0 was adjusted to a new position based on the price preference and seven-day cycle threshold time frame. Given the new position of V_0, T_0 was also reevaluated and readjusted.

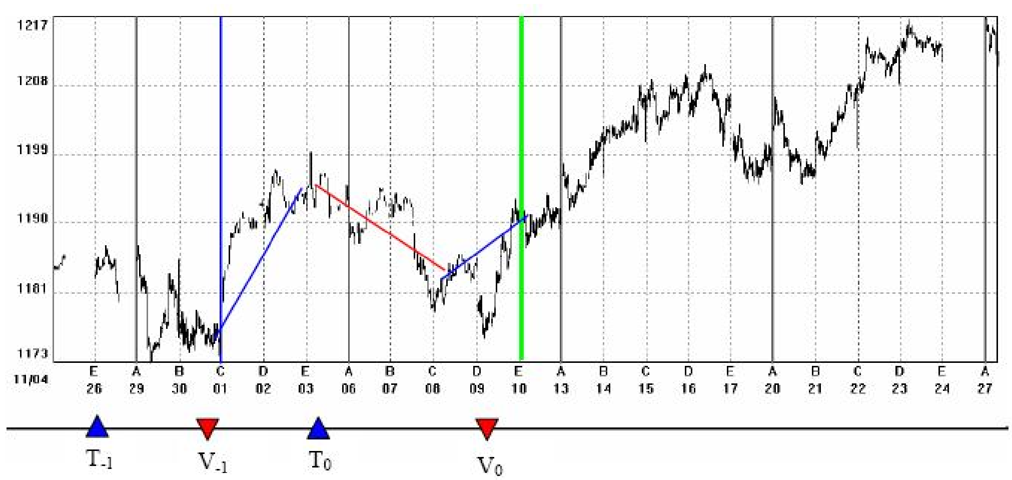


Figure B.4 Example of real time C5 cycle auto-recognition and self-adjustment.

5. At the time shown by green line in Figure B.5, the system created another trade signal that considered present time the cycle top. Then, the system reversed the trade and did a short trade. A new top T_1 is being put in the figure.

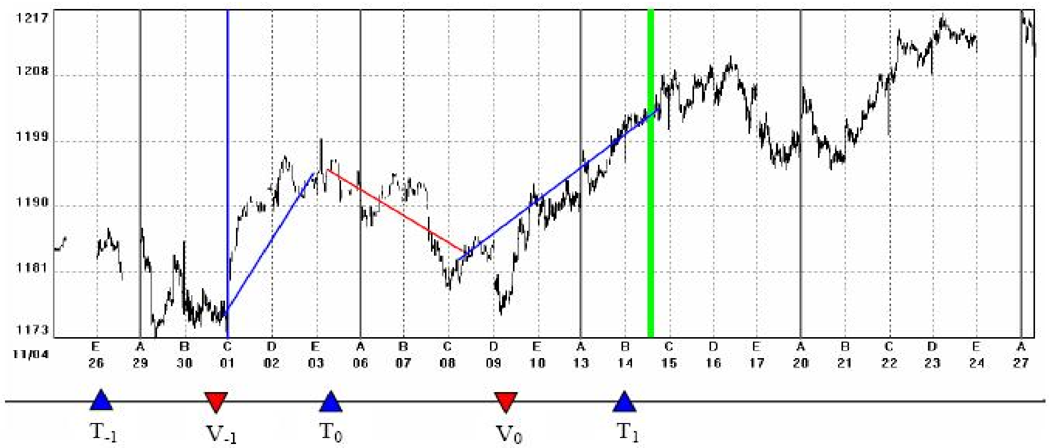


Figure B.5 Example of real time C5 cycle auto-recognition and self-adjustment.

6. At the time shown by green line in Figure B.6, based on the trade signal from TF1, the short trade was closed and a long trade was initiated. A new valley V1 is generated. At the same time, the Triangle method was used again to readjust the previous T1 to better represent new market information.

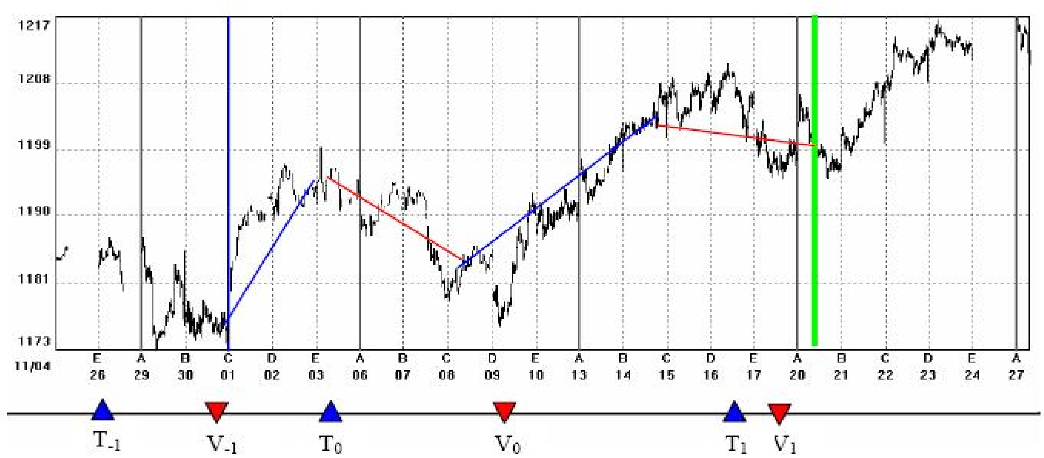


Figure B.6 Example of real time C5 cycle auto-recognition and self-adjustment.

7. At the time shown by green line in Figure B.7, considering the price preference and cycle threshold and the optimal-three-valley rule, the cycle valley V1 was adjusted to a new position.

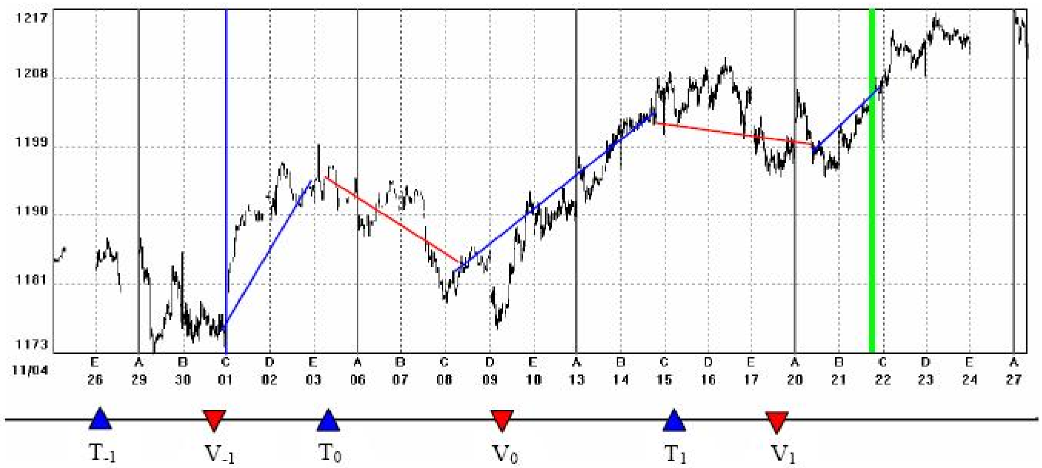


Figure B.7 Example of real time C5 cycle auto-recognition and self-adjustment.

8. Step 1 to 7 is repeated and the market markers are generated to use in the TFX methods.

Real time cycle auto-recognition is an important part of our cycle-related TFx methods. Using historical man-made cycle database and auto-recognition market makers, the TFx methods selected k cohort cycle members to guild prediction of present cycle. The majority of cohort cycle members will determine whether the present time is the present cycle point.