MULTI-FREQUENCY ANALYSIS FOR HIGH FREQUENCY TRADING

LIU HONGGUANG

Ph.D

The Hong Kong Polytechnic University

2017

The Hong Kong Polytechnic University

Department of Industrial and Systems Engineering

Multi-frequency Analysis for High Frequency Trading

Liu Hongguang

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

November 2016

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_______________________________________(Signed)

____________Liu Hongguang_______(Name of student)

ABSTRACT

High-Frequency Trading (HFT) in financial markets has been making media headlines. The 2010 Flash Crash in the US and the 2013 Everbright Securities' incident in China showed its dramatic impacts on the markets. However, as a relatively new phenomenon, most of the discussion on HFT is not backed by solid academic research. At the same time, current academic research on high-frequency trading focuses on its afterward influences, the motivation and the trading logic behind the HFT is rarely explored. Basically, there are two kinds of HFT, the first kind of HFT takes advantage of "time", the most advanced computers are placed right next to the exchanges to reduce the time delay of the receiving of market data and the execution of trading orders that aiming to capture a very small fraction of the profit on every trade. The second kind of HFT is conducted based on the analysis of the historical data of the related financial time series. This thesis focuses on the study of the second kind of HFT.

Multiple methods can be used in the design of the second kind of HFT. In this research, multi-frequency analysis and wavelet are combined with technical indicators and modern machine learning tools. Forecasting of the directions of the financial time series is crucial in the design of such kind of HFT systems, many economic and technical models and indicators have been built in the past, however, most of the past research merely analyze the data in time domain, the frequency

domain of the HFT is rarely explored. This research focuses on the multi-frequency predictions of the short-term movements of the financial time series and the design of the trading systems based on the forecast.

HFT systems based on moving averages and a simple trend following system are developed to set benchmarks for the multi-frequency related systems. An experiment on the performance of two-frequency ARIMA model is also conducted to show the prediction power of the multi-frequency analysis, as time series in different resolutions may convey different information on its characteristics, the empirical results indicated that multi-frequency could improve the forecast performance. After that, an intra-day trading system is designed based on the Genetic Programming (GP) and technical analysis, wavelet de-noise is introduced to improve the performance of the GP based system, the system with wavelet de-noise showed best performance in the empirical test. To explore the nonlinear relationship, artificial neural network (ANN) is applied in the prediction of the financial time series. Both Nonlinear AutoRegressive with eXogenous (NARX) and wavelet based Multi-layer perceptron models are utilized in the forecasting of the intra-day high-frequency time series, based on which, HFT systems are developed.

To test the performance of the HFT systems, the China index futures is selected as the experiment asset. Based on the experiments in this thesis, the HMA trading system shows the best performance among the tested moving averages trading systems; the two-frequency ARIMA beats the traditional single frequency models;

the GP systems trained using the wavelet de-noised data outperforms the GP systems trained using the original data, and the hard-threshold denoise method provides the best out-of-sample trading performance; the WMLP based trading model outperforms the NARX model in the out-of-sample trading test.

ACKNOWLEDGMENTS

This research would not have been so successful without the help of many people who took great support. I would like to express my special thanks to my supervisor, Prof. Ping Ji of Department of Industrial and Systems Engineering in The Hong Kong Polytechnic University, for his generous advice, inspiring guidance and encouragement throughout my research project. His encouragement, helpful suggestions have supported the development of the project. His deep understanding and wide knowledge have broadened my view in the field about this research. His kindness guided me through the hard times during my Ph.D. study and my life.

Besides, I must express my sincere gratitude to my family and friends for their tremendous support and encouragement. Especially my parents and my grandparents, I owe the most to them.

The financial support from The Hong Kong Polytechnic University enables this project possible. I would like to acknowledge all parties from the heart including those mentioned above together with those who helped me indeed but missed to thank before.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 – INTRODUCTION

1.1 Background

High-frequency trading (HFT) is a type of algorithmic trading, specifically the use of sophisticated technological tools and computer algorithms to rapidly trade securities. Another definition of high-frequency trading is that HFT is the kind of trading that rely on advanced computer systems to carry out the execution of trade at very high speed. Judging from the definitions, high-frequency trading has two key characteristics: the trading process is computerized and the time interval of trading is very short. High-frequency trading firms are believed to account for over 60% of all the US equity trading volume.

Regarding the trading interval and the trading logic, high-frequency trading can be divided into two categories: the first kind of high-frequency traders take advantage of "time," such kind of high-frequency trades are most likely to happen several times in a single second. To do so, the most advanced computers are placed right beside the exchanges to reduce the time delay of the orders' transaction and execution process. In most of the cases, such kind of trades aims to capture a very small fraction of profit on every trade. However, the extremely mass amount of trades could accumulate considerable profits.

Another kind of high-frequency trading is conducted based on the analysis of various kinds of related historical data of the targeting assets. Different algorithms are developed to form a better understanding of the statistical characteristics of the assets' prices. Trading strategies are created based on the forecasting of the movements of the assets prices. The speed and frequency of such kind of high-frequency trading are much lower than the first kind of HFT.

Although some of the tricks including the Graphrical Processing Unit(GPU) calculation boost and the location of the strategy execution computer are taken into consideration, the high-frequency trading that this research mainly focuses on the second category, such kind of HFT could also be called "algorithm trading", instead of taking advantage of the ultra-high speed of execution of trading orders, the trading actions are made based on the quantitative analysis of the historical data of the assets.

Regarding the statistical analysis of the historical data, financial time series forecasting needs to be addressed here as it has drawn the interest of both researchers and investors for decades. Many types of research have been conducted to uncover the mystery of the movements of financial time series. Both fundamental and technical information on the corresponding financial times series have been utilized in the prediction of the time series. An agreement has been reached that there is no perfect model that could be used to forecast all the time series accurately. All that researchers can do is to find optimal models that fit for a specific time series with certain characteristics.

Regarding financial time series forecasting, especially for the stock or futures forecast, most of the research frequencies are days, weeks and even months. Some models showed indeed outstanding performance in this kind of “long” term prediction of the time series. Past researchers showed the profitability of the trading strategies based on those kinds of forecasting. However, these kinds of theoretical profits are highly unlikely to be obtained from the real markets. This is because that most of the trading in futures markets are margin trading, some small movements of the futures prices mean a large amount of gains or losses. For most of the time, investors do not have enough money to wait for the market to go to the “right” price.

During the last decade, various types of machine learning tools have been used in the searching of better financial time series prediction models. Tools like genetic programming, support vector machine, neural network and deep neural network have been utilized in the design of forecasting models. However, most traditional forecasting models focus on the time domain analysis of the time series, for example, the AR (auto-regression) model takes the lag terms of the time series as the explanation variables, a clear drawback of such kind of analysis is that the characteristics of the frequency domain of the series are not considered. In terms of time series frequency analysis, wavelet showed its powerful advantage. With the introduction of wavelet de-noise, the useful part of the original time series can be efficiently separated from the noise part. By introducing wavelet decomposition into the analysis of the frequency domain, the original time series can be decomposed into different levels based the calculation of the characteristics of the series. Both continuous wavelet transformation (CWT) and discrete wavelet transformation (DWT) methods are described in the methodology part of this thesis. After the introduction of the powerful frequency analysis tool –wavelet, various combinations of wavelet transformation, wavelet de-noise and different traditional time series forecasting models will be used to generate trading strategies in this study.

To evaluate the performance of the proposed models in this research, the CSI 300 index futures is selected as the testing target. The emerging index futures market of China is one of the most actively traded markets. Past studies on index futures usually based on developed markets, such as the USA and the Hong Kong market, characteristics of emerging futures market like China index futures are rarely explored. The proposed strategies in this research are mainly tested using the intra-day 1-minute interval data.

As most of the high-frequency trades are traded at very high speed on the consideration that 'shorter holding period means lower risk,' programmed trading platforms are needed to execute the trading orders placed by the trading strategies. Trading platform like Tradestation, Multicharts have millions of users trading on them every day, some large banks and brokers even build their trading platform to protect their trading strategies from being leaked. In the first stage of this study, to simplify the testing process of the proposed trading strategies, Multicharts is selected as the platform to conduct the first round back test and real-time test due to its better data availability on China markets. However, as the connection speed with the exchange and the time delay of the calculations of the strategies have significant affection on the range of slippage and the possibility of order execution, co-located automated high-speed trading systems based on the Femas API provided by the China Financial Exchange and GPU Boost technology provided by Nvidia is developed in the second stage of this research to fully test the real market performance of the proposed strategies.

1.2 Motivation and Objectives

Although high-frequency trading is becoming increasingly popular in the real markets, most articles on high-frequency trading in literature are just some simulated trading experiments and the impact of HFT on the market. None of the models showed real profitability. Although there is no documented profitable trading strategy, increasingly institutions are putting higher and higher weight on high frequency programmed trading. Due to the protection of commercial secrets, none of those trading strategies is disclosed to the public.

It is believed that different frequency's financial time series carry different characteristics of that series. However, forecasting models that combined various frequency of the original series can rarely be found in the literature. During last decades, machine learning tools like Genetic Programing (GP) and Neural Network

(NN) have shown their power in the extraction of patterns from big amount of data, however, most of the past research ignore the un-stationary characteristic of the financial time series, denoise methods should be explored to its effect.

All of the past research on HFT found in literature shared a common drawback; that is past models are all trained with a long-term period of data, such as daily, weekly or even monthly data without consideration of the jumping points caused by the close market event. In other words, past forecasting systems ignored the impact of incoming information on the movement of the time series and tried to extract pattern from the series with various kinds of jumping points caused by the different financial event. Information like financial reports and change of interest rates are highly likely to affect the price movement of the target analyzing asset; jumping points occurred after the announcement of such kind of information. All the models in literature trained with long term data which includes many jumping points are unable to uncover the real characteristics of the time series. To overcome this problem, models in this study are trained using the intra-day data, at the same time, indicators that are used in the proposed models are all calculated using the data from the same trading day. Information that has dramatic impact on the market are expected to be digested during the market closed periods, therefore, jumping points barely existed in the intra-day data. True trends of the time series can be recognized with the proposed models trained using the intra-day data.

As an emerging index futures market, the China index futures is among the most extensively traded market in the world. Due to the difference in government regulation and the structure of the market, the China index futures is quite different from the developed markets. Few research has been carried out in this market, especially the intra-day high-frequency trading. The China Financial Exchange just released the Femas API, based on which ultra-high speed of order transmission and execution system can be developed.

The research objectives of this study are:

- To find more reliable forecasting models for high-frequency financial times series, based on which, profitable trading strategies could be made. In the forecasting process, different frequency of the original data will be taken into consideration to extract the true trend related characteristics that can be applied into the design of intra-day trading systems.

- To propose a way to design expert trading systems from the simplest system to rather complex systems based on different kinds of machine learning tools combined with denoise methods. The way to improve the performance of the designed trading systems will also be discussed.

- To test the market efficiency of the China Financial Index Future market. According to the efficient market hypothesis (EMH), if future prices can be predicted by analyzing prices from the past and excess return can be earned in the long run, the market efficiency is even weaker than the weak-form efficiency.

1.3 Scope of this thesis

This project is mainly devoted to the design of multi-frequency analysis based high-frequency trading systems. Various types of machine learning tools are combined with multi-frequency in this research to extract true patterns from the financial time series, which could then lead to real profit in the trading systems. The structure of the thesis is organized as follows.

In Chapter 2, an extensive literature review is conducted to demonstrate what have been studied on financial time series forecast, multi-frequency analysis, high-frequency trading. Since the Genetic Programing (GP), Wavelet de-noise and wavelet decomposition, Artificial Neural Network (ANN) are adopted in the searching of both linear and nonlinear combinations of various kinds of technical indicators in this project, the fundamental concepts and procedures of the pertinent exact algorithms and heuristic methods are also surveyed.

In Chapter 3, a detailed description of the related methodologies used in this thesis is conducted.

In Chapter 4, a series of trading systems based on simple moving average, exponential moving average and hull moving average are presented. As moving averages are among the most widely used technical tools by participants in the financial markets, the performance of these systems will mainly be used as the benchmarks for the following complex trading systems proposed in this thesis.

In Chapter 5, a manually two frequency Auto Regressive Integrated Moving Average (ARIMA) model is proposed to give the longer time interval one step ahead forecast. The empirical results of this model demonstrated the power of the joint frequency analysis. A simple trend following trading system is also presented in this chapter, just like the moving average based trading systems in Chapter 4, the performance of this simple trend following system is used as the benchmark for further developed systems.

In Chapter 6, an integrated trading system is proposed based on wavelet de-noise and Genetic Programming (GP). GP is used in this study to find the optimal linear combination of various types of technical indicators and orders. Taken the non-stationary and noisy characteristics of the time series into consideration, Wavelet Threshold (WT) is introduced to reduce the impact of the noise. The optimized trading system is programmed on the Multicharts systems to test its profitability in the real index futures market. The performance of the proposed models is compared with the trading systems from Chapter 4 and 5.

In Chapter 7, as the method GP introduced in Chapter 6 merely explored the linear combinations of the technical indicators, Artificial Neural Network is applied in this chapter to explore the nonlinear relationship space. A Nonlinear Autoregressive Exogenous Model (NARX) is proposed to perform one step ahead forecast, based on which trading systems developed. With the introduction of multi-frequency analysis into the design of forecasting model, a wavelet based Multi-Layer Perceptron (MLP) neural network is constructed. The performance of the proposed trading systems designed based on these models is compared with the trading systems from Chapter 4,5 and 6.

In Chapter 8, the distinctive achievements of this project are provided. Both the academic and industrial contributions of this research are concluded. Finally, some recommendations for future work are suggested.

CHAPTER 2 – LITERATURE REVIEW

2.1 Introduction

This chapter is organized as follows. Section 2.2 highlights the past studies on the multi-frequency forecast and high-frequency trading. In Section 2.3, the basic concepts and pertinent technologies used in the design of trading systems are discussed with the emphasis on the searching of better technical trading strategies. Afterward, the studies on the forecasting of financial time series are elaborately surveyed in Section 2.4. There are two categories of forecasting research, which is the fundamental forecast and the technical forecast. The fundamental Forecast is mostly used to identify the long-term trend of the asset, whereas technical forecast along with various kinds of machine learning tools is more useful in short term price movement prediction. Section 2.5 provides an overview of the powerful multi-frequency analysis tool—wavelet. Section 2.6 gives a short review on the valuation of forecast accuracy. Finally, some remarks concerning the reviews are summarized in Section 2.7.

2.2 Multi-frequency Forecasting and High-Frequency Trading

High-frequency trading has been popular in recent years for its immense profitability. Compared with traditional low-frequency trading, the key characteristic of high-frequency trading is the high ratio of capital turnover in rapid computer-driven response to fast-changing market conditions. Instead of holding trading positions for days or weeks as traditional money managers do, High-frequency traders execute multiple trades each day or even each minute.

High-frequency trading is quantitative trading that is characterized by short portfolio holding periods (Wilmott, 2008). All trading decisions are made by computerized quantitative models. Modern high-frequency trading was believed to start from 1999, after the authorization of electronic exchanges in 1998 by U.S. Securities and Exchange Commission (SEC). At the very beginning of the 21st century, the minimum time taken by high-frequency trading was several seconds, whereas by 2010 this had declined to a millisecond or even microsecond. According to the data from NYSE, the high-frequency trading volume soared by 164% from 2005 to 2009. Liquidity provision is another key factor that leading to the rapid growth of high-frequency trading. Many market makers introduced high-frequency trading to the market with lower volatility and wider bid-ask spreads.

A. Cartea and J. Penalva [Car12] explored the value of high-frequency trading. Three types of traders were considered in their model, namely, liquidity traders, professional traders and high-frequency traders. They found that high-frequency trades increase the microstructure noise of prices, and that trading volume increases dramatically with the intermediate of high-frequency trading. They also stated that although high-frequency traders did “rob” money from traditional professional traders, traditional professional traders are better off with higher liquidity discount in the market price.

High-frequency trading can be classified into four categories based on the typical holding period of the target assets:

Table 1.1 Four categories of high-frequency trading

Holding period	Description	Strategies utilized
Less than 1-minute	Quantitative algorithms for optimal pricing and execution of market-making positions	Automated liquidity provision
1-minute to 10 minutes	Identifying trading party order flows through reverse engineering of observed quotes	Market microstructure trading
10 minutes to 1 hour	Short-term trading on macro events	Event Trading
1 hour to 1 day	Statistical arbitrage of deviations from equilibrium: triangle trades, basis trades, etc.	Deviations arbitrage

High-frequency trading can also be distinguished by different trading goals, like speculation and market making. Another trend in high-frequency trading is to analyze the trend using tick-by-tick ultra-frequency financial time series data, whereas trading decisions are made consequently. Various kinds of models are used in the processing, filtration and analysis of those high-frequency data. In this research, multi-frequency analysis tools wavelet de-noise and wavelet transformation are used in the high-frequency short-term price forecast, based on which, expert trading system are developed.

Charles and Maureen [Cha97] conducted a survey of the issues and applications of high-frequency data in financial markets. They pointed out two important research questions that have not been fully explored. The first one is how the properties of the data set differ on different sampling rules; this questions could be interpreted as how the properties of different time of interval high-frequency data differ from each other. The other question is that what kind of information should be included in the development of futures high-frequency data sets. Does the price in the market include all the related information? Should volume, timing and order types be put into the high-frequency data sets? Those questions remain unsolved, which results in a more difficult high-frequency forecasting environment.

Taylor [Tay10] introduced five univariate exponentially weighted methods in the forecasting of intraday time series that contain both intra-week and intraday seasonal cycles. Several applications of these methods like the prediction of call center arrivals, email traffic and electricity loads were mentioned in this paper, it would be interesting to test these models with high-frequency financial series. Robert and Magdalena [Rob12] decomposed the volatility of high-frequency asset returns into multiplicative easily interpreted and estimated components and applied their volatility forecasting model to a comprehensive sample consisting of 10 minutes' returns on more than 2500 US equities. This novel intraday volatility forecasting model could be used in the design of computerized high-frequency trading which depends on volatility forecasts.

Jose and Juan [Jos12] examined the predictability of intraday stock market returns with both linear and nonlinear time series models using high-frequency data. They took the time horizons of 5,10, 20, 30 and 60 minutes into consideration and found that in total nonlinear models out performed linear models, this is consistent with the fact that financial series is nonlinear. They also stated that more flexible nonlinear models like support vector machines and artificial neural network (ANN) did not clearly beat other traditional nonlinear models like a smooth transition and ARMA models. The results provide weak evidence of intraday predictability of high and low volatility regarding statistical criteria but corroborate the superiority of nonlinear model predictability using economic criteria. Their work seems to cover most of high-frequency forecasting except for shorter time horizons, as 5-minutes itself is a rather long time in financial markets, sharp market movements can be accomplished with one minute. Andrea et al. [And12] proposed ARFIMA model in the forecast of the intraday market price of money, both the point and density forecasting. Their empirical results provide evidence that the proposed model outperformed random walk and autoregressive benchmarks.

Bollerslev and Jonathan [Bol01] showed in their paper how high-frequency data could be used to construct superior daily volatility forecasts. They proposed a simple volatility dynamics model with high-frequency data by simply fitting an auto regression to log-squared, squared, or absolute returns and found that when dealing with high-frequency intraday data, these simple auto regressions tend to outperform standard GARCH and EGARCH models in forecasting future volatility.

Carlin and Lam (2010) proposed a new method for the modeling of intraday volatility, which improved the subsequent ARCH in the sequential method by integrating the filtration process. In this model, they utilized the interaction effect between the periodicity and the heteroskedasticity in the searching of better ARCH parameters. Their empirical results of NASDAQ and S&P 500 indexes 10 minutes' returns provide strong evidence of better performance among other models. The superior modeling accuracy can be of great help in the computing of intrinsic value of financial instruments.

Yan and Li [Yan12] found out that ARMA-EGARCH (2,2) outperforms other models in simulation and forecast of the intraday volume time series of Shanghai security composite index. They are among the pioneer researchers who took 1-minute intraday high-frequency data into consideration. They observed that unlike most western countries, the volume time series in China exhibit W shape trend in most trading time.

Martin and Zein [Mar03] improved the forecast of financial volatility by shifting time-series models based on historical daily returns to option implied volatilities.

High-frequency data and long-term modeling were used in the study. They tested their model on S&P500, YEN/USD, and Light, Sweet Crude Oil; all the results indicated the superiority of their model.

Torben et al. (2005) conducted a study on the price driven effect of announcements of the U.S. macroeconomic news. They found that news surprises lead to conditional mean jumps on the high-frequency data of U.S, German and British stock, bond and foreign exchange market. This stated that market reacts differently in the different macroeconomic environment. This result could be used as a source of an external factor in the forecasting of certain asset.

Menkveld, Albert J. [Men13] characterized the trading strategy of a high-frequency market maker. His study indicated that the market maker incurred a loss on its HFT inventory and earned a profit on the bid-ask spread.

Kirilenko, Andrei A., et al. [Kir15] conducted a study on intraday market intermediation in an electronic market before and during a period of large and temporary selling pressure. They found that the trading pattern of the most active non-designated intraday intermediaries (classified as High-Frequency Traders) did not change when prices fell during the Flash Crash.

2.3 Trading Strategies

Neely and Weller [Nee03] tested the performance of two technical intraday trading strategies—a genetic program and an optimized linear forecasting model on the exchange market. They stated that no excess returns could be gained by either strategy when realistic transaction costs and trading hours are taken into consideration. Thus, they claimed the results to be consistent with market efficiency theory. However, as increasingly high-frequency trading team reported positive excess return over market return, there is a high possibility that their strategies' poor performance might be caused by the selected models and the specified assets. Exchange rate always tends to stay at a certain level. Thus linear forecasting model's bad performance may simply attribute to the special properties of the specified asset.

Kablan Abdalla [Abd09] introduced Sugeno model as an implementation of the ANFIS (Adaptive Neuro-Fuzzy Inferences Systems) in the forecasting of high-frequency financial series and the estimation of parameters. In this paper, Kablan proposed a system that performs a prediction on the financial times series in high frequency using intraday data (5-minutes intraday trading prices), buy and sell orders are placed based on the prediction of the movement of the prices. Although this system was observed to have a very high accuracy rate and open the right position at the right place, however, these trading results did not always translate into higher returns due to transaction costs. He then modified the trading strategy by holding the assets for a relatively "longer period" do not change position as long as the current position matches the latest prediction, the performance was improved due to the reduced number of trading, hence the reduction of transaction costs.

Kablan introduced a trigger point in the final modification of the system, the results of the systems' performance were very promising compared to the traditional buy and hold strategy, with positive Sharpe and Sortino ratios, which means that the system did not take higher risk for the higher amount of return gained.

Kleidon [Kle92], Kumar and Seppi [Kum94], and Holden [Hol95] highlighted the importance of execution risk in index arbitrage under special market conditions. With increasing investors' interest and regulatory concerns for high-frequency trading, understanding the role of execution risk in related strategies is very important. Roman and Wing [Rom12] explored the role of execution risk in high-frequency trading through arbitrage strategies, as those arbitrage strategies isolated other risks in the setting, an ideal platform to analyze execution risk is provided. They demonstrate the execution risk arising from the crowding effect of competing arbitrageurs placing the same trading order and inflicting negative externalities on each other. Roman used the triangular arbitrage in the FX market as the empirical experiment and showed that in contrast to the common notion that competition improves prices efficiency, competition among arbitrageurs can limit efficiency when they inflict negative externalities on each other. Brownlees and Gallo [Bro06] highlighted the importance of preprocessing of high-frequency financial series; tick-by tick trading series were used in their empirical study. Their study showed that "wrong" ticks of the series is likely to shorten the financial durations between substantial price movements and to alter the autocorrelation profile of the series. Different filter algorithms were used in the handling of outliers of the data to clean the original series. This paper posed a great problem in the analyzing of past data.

Frank [Fra10] argued in his paper that high-frequency trading may potentially have some harmful effects on the U.S. capital market. He showed that the high-frequency trading volume accounts for about 78% of total dollar trading volume in the first two-quarters of 2009, which is apparent too much, according to his statement. His study indicated that high-frequency trading is positively related to stock price volatility even after controlling for the volatility of stock's fundamentals and other volatility drivers. However, what is the underlying mechanism of high-frequency trading in price discovery and how the high-frequency trading affects the volatility of the stock remain not clear. Those unresolved questions increase the systematic risk of U.S. market as high-frequency trading constitutes the lion's share of trading volume in the capital market.

Ren S. Miller and Gary Shorter conducted a survey of the recent developments of high-frequency trading [Ren16]. Some recent HFT related regulations are covered in this study.

2.4 Financial time series forecasting

Since the introduction of stock, the prediction of the movements of the stock price has been one of the most important aspects of forecasting research. In this study, the forecasting literature will be divided into two categories, namely, the traditional financial time series forecasting and the high-frequency financial time forecasting.

The traditional forecasting models were used in the prediction of daily, weekly, monthly and even yearly value of financial time series. Basically, there are two kinds of forecasting methods, one is fundamental analysis, which mostly takes the financial statement of a certain company into consideration and makes prediction based on such kind of fundamental data; and another method is technical analysis, which is not concerned with any fundamental data of the underlies company, forecasting is made merely for the technical analysis of the historical price of the stock.

2.4.1 Fundamental Forecast

There are many papers focusing on fundamental forecasting in literature, this kind of forecasting can be used to identify the long-term trend of a certain asset. The sound fundamental analysis will help to discover firms that represent a good value. E.F. Fama [Fam70] published his famous paper of Market Efficiency Hypothesis (EMH). The market was defined into three kinds of efficiency, namely, weak form, semi-strong form, and strong form. Fundamental analysis is believed to be useless if the market passes the semi-strong form test, like all publicly available information has already been revealed in the market price. Much work has been done to test the market efficiency in different markets since the introduction of EMH.

As ratio analysis is started from the late 1800, Melvin C. O'Connor [Mel73] tried to identify the usefulness of this kind of financial ratio analysis. In his paper, O'Connor used financial ratios to predict the return rankings instead of returning itself, which turned out to be useless. S. Basu [Bas97] was among the first who questioned the market efficiency hypothesis, he tested the profitability of trading based on price-earnings (PE) ratios indicators and found that returns on stocks with low PE ratios tends to be larger than warranted by the underlying risks, costs. This means that the behavior of security prices over the empirical study period is not completely described by the efficient market hypothesis and that fundamental indicators like PE ratios could be useful in the prediction of stock prices.

[Ban80] Reported that firms with small size regarding market value had significantly larger risk-adjusted returns than large NYSE firms over a forty-years period. This size effect can be utilized by fundamental analyzers in their forecasting of stock returns. Another important ratio in fundamental analysis is Debt/Equity ratio. L.C. Bhandari [Bha88] conducted an empirical study of the relation between Debt/Equity ratio and stock return; he found that the expected returns on common stocks are positively related to the Debt/Equity ratio, with the control for the beta and firm size, both including and excluding January effect. J.A. Ou and S.H. Penman [Ou89] performed an extensive analysis of company fundamentals to derive a method to forecast the changes in the companies' future earnings; a portfolio was formed based on the method. The P_r, a scalar measure, was used to cover the useful information contained in the financial statement. They took different positions according to the value of P_r, (long position when P_r>=0.6, short position when P_r<=0.4). Their approach generated a significant market adjusted return of 14.53% over the holding period. J.L. Farrell, Jr. [Far89] showed the importance of macroeconomic environment in making fundamental forecasting. The original relationship between different assets will change if there is a significant variation in economic environment.

As Ou and Penman's [Ou89] finding prompted fundamental analysis as a worthwhile activity, Setiono and Strong [Set98] re-examined the value of fundamental analysis with two approaches, and found consistent evidence that an investor could have earned abnormal returns by exploiting published financial statement information to predict future earnings changes. R.W. Holthausen and D.F. Larcker [Hol92] took one more step ahead in the using of fundamental signals, instead of predicting the annual earnings-per-share (EPS), and they set the subsequent one-year excess return as their forecasting object. They examined the profitability of trading strategy based on the three excess return metrics, namely, the market-adjusted return, the excess returns computed using the Capital Asset Pricing Model (CAPM) and the size-adjusted returns; the trading strategy takes a long position in firms predicted to have positive excess returns and short position in firms predicted to have negative excess returns over the next 12 months, all of the three different models turned out to be profitable. E.F. Fama and K.R. French

[Fam92] published their famous paper on a cross-section of expected stock returns. In this paper, the book-market ratio was shown to be most related to the expected stock returns in the United States. Furthermore, they discovered that the combination of book-market ratio and market value of equity captures the explanatory power of the price-earnings ratio, financial leverage and beta for stock return. This paper formed the basis of modern fundamental analysis. Baruch Lev and S. Ramu Thiagarajan [Bar93] identified a set of fundamentals to evaluate firms' performance and estimate future earnings; analysts' descriptions guided those fundamentals instead of a statistical searching procedure, which significantly increased the economic meanings of the fundamentals.

J.S. Abarbanell and B.J. Bushee [Aba98] defined several fundamental signals and examined the relationship between security prices and those accounting-based fundamental signals. They found that investors, in general, may be inefficient in their fundamental analysis and that they cannot perceive the whole information contained in the financial statements. Sandip et al. [San97] conducted fundamental analysis of stock returns in Korea and found that annual stock returns were positively related to B/M (book to market ratio) and negatively related to firm size. J.S. Abarbanell and B.J. Bushee [Aba98] examined the profitability of trading based on fundamental analysis; they stated that fundamental signals provide information about future returns associated with future earnings news, their abnormal return gained from the fundamental trading is mostly generated around subsequent earnings announcements. P.M. Dechow et al. [Dec01] identified that short-sellers utilized fundamental signals (such as earning to book values ratio) to form their forecast and built their trading strategies based on this kind of prediction. They documented a strong relationship between the short-sellers' trading strategies and the fundamentals ratios of market prices and showed that short-sellers target securities have low fundamental analysis ratios and that the positions were changed as those ratios revert to normal levels.

Recently, fundamental factors are often used in the analysis of macroeconomic environment and the long-term movement of related stocks. Those factors show few contributions in the short-term stock price forecast. Technical and statistical analysis are more powerful and useful in such kind of short-term prediction scenarios.

2.4.2 Technical Forecast and Machine learning tools

Technical analysis is another aspect of financial analysis, instead of studying economic, industry and the company conditions to find the intrinsic value of a company's stock; technical analysis focuses on historical data in the discovery of useful information and prediction of future stock price. Even though it has been criticized and scorned by many academic researchers and practitioners, technical analysis has proved itself to be useful and powerful in both academia and practice.

F. Ferna´ndez-Rodr´ıguez et al. [Fer00], R.Gencay [Gen98] and T. Chavarnakul, D.

Enke [Cha08] presented extensive reviews on the relevance of technical analysis in financial markets and the profitability of technical patterns. A considerable amount of work proved technical trading rules to be capable of producing valuable economic signals in both stock markets and foreign exchange markets. F. Fernańdez-Rodŕıguez et al. [Fer00] tested the ANN forecast based trading strategy on Madrid stock market; significant profit was gained without the consideration of transaction cost.

The simplest technical forecast is the naïve expectation, P.M. Dechow and R.G. Sloan [Dec97] tested the profitability of contrarian investment strategies based on naïve expectations hypothesis. They found that the higher return on contrarian investment strategies came from investors' naïve reliance on analysts' forecasts of future earnings growth, other than the naïve extrapolation of past trends in earnings and sales growth. The random walk model is another simple but powerful model. Many researchers use this model as one of their forecasting benchmarks. R.G. Brown and R.F. Meyer [Bro60] developed the exponential smoothing process, which turned out to be a very efficient way to simplify the computations required in fitting higher order polynomials to data as a bias for the forecasts. Autoregressive moving average model (ARIMA) is one of the most famous linear time series forecasting models. Described by Peter Whittle in his thesis, the general ARMA model was popularized by George E.P. Box and Gwilym Jenkins (1971). Box-Jenkins method was expounded in the choosing and estimating of ARMA model.

ARIMA model was applied where the data show evidence of non-stationarity and that a differencing step could remove the non-stationarity. Numerous researchers have contributed to the implementation of this powerful method in the analysis of time series. Fuzzy regressions were combined with ARIMA model by F.M. Tseng et al. [Tse01]. With the possibility distribution of future values taken into consideration, their forecast model can provide prediction intervals instead of exact future values, which make it possible for the decision makers to forecast the best and worst possible situation.

As linear models cannot capture all of the patterns in real life problems, non-linear models were introduced to account for the processing of complex real-life time series. Bilinear Model (1978), threshold autoregressive (TAR, 1983) model and the autoregressive conditional heteroscedastic (ARCH, 1982) model were developed to meet the specific non-linear time series pattern. Among all the nonlinear forecasting models, Artificial Neural Networks (ANNs) have been extensively used in time series forecasting. There is no need to set a fixed model for using ANNs in modeling. This is the major advantage of ANNs over other nonlinear models. ANN models have very powerful properties in pattern recognition. The form of the model could adapt according to the special properties of the data sets, which makes it extremely suitable for the analysis of high-frequency financial data. G. Zhang et al. [Zha98] presented a review on some traditional modeling and applications of ANNs in the last two decades of the 20th century. Recently, many hybrid ANNs models appeared in the literature of forecasting.

J. Yao, C.C. Tan [Yao00] fed neural network with simple time series data and technical indicators, such as moving average in the prediction of foreign exchange rates. Their empirical results showed that useful prediction can be obtained and that acceptable paper profit can be generated in the consideration of transaction cost. Although their forecast results did have better perform than their selected benchmarks, the paper profit was not that persuasive, as no trader holds forex for weeks, most of the trading are done within a single day. G. Peter Zhang [Zha03] was among the first who combine ARIMA with ANNs in the financial time series forecasting; he showed that the forecasting accuracy of the hybrid model outperformed both models used separately. This hybrid model took advantage of the unique features of ARIMA and ANNs models in linear and nonlinear modeling. ANNs process all the information from the data in a parallel way, no pre-set priorities were calculated, the special characteristics of the original series determined the model. Their model was tested with different kinds of data, namely, Wolf's sunspot data, the Canadian lynx data, and the Pound/US Dollar exchange rate data, all of which showed better performance over benchmark models.

Similar to G. Peter Zhang [Zha03], other people also combined ARIMA and ANNs in the analysis of time series forecasting. P.F. Pai, C.S. Lin [Pai05] conducted a one-step ahead forecast using the hybrid ARIMA and support vector machine (SVM). In their study, daily opening and closing prices of certain stocks were used in the training of the model, the empirical results of the proposed model showed great improvement of prediction accuracy over each separated ARIMA and SVM models. H. Ince and T.B. Trafalis [Inc06] combined parametric and nonparametric techniques in the forecast of the exchange rate. Vector autoregressive (VAR) selection based support machine (SVR) model and ARIMA selection based ANNs model were compared in their study, both of which showed their advantages and disadvantages. H.Ince and T.B. Trafalis [Inc06] also showed that SVR (support vector machine) outperformed the MLP (multi-layer perceptron) method in finding the global optimum.

W. He et al. [He08] presented several ways to improve the performance of support vector regression (SVR or SVM) in the forecast of time series. Self-adaptive parameters, multiple kernel function, and feature selection were utilized to capture the complex attributes of the input data. Experimental forecasting of their model on the data of Mackey-Glass and Sunspot data showed appreciable results compared with the benchmark models. T. Chavarnakul, D. Enke [Cha08] combined generalized regression neural network model with technical indicator VAMA (volume adjusted moving average) in the prediction of future stock price. By taking trading volume into consideration, the trading strategies based on their models indeed had a better performance compared with benchmark strategies. To capture the local trend and tolerate the changing noise of the complex financial time series, H. Yang et al. [Yan09] proposed a novel localized support vector regression (LSVR) to forecast time series. Compared to the traditional support vector regression (SVR), the proposed LSVR captures the local information in the data. More importantly, LSVR is mathematically connected with several related models, for example, LSVR could be transformed into the standard SVR by setting some mild assumptions. With more information taken into consideration, the empirical results of LSVR demonstrate its advantages over the standard SVR. C.J. Lu et al. [Lu09] improved the SVR model by combing the independent component analysis and the standard support vector regression. In their study, the independent component analysis was used to detect the independent variables that contain noise and alleviate the influence of the noise in the standard SVR. The modified SVR model was shown to outperform the standard SVR model and a random walk model with the data of the Nikkei 225 opening cash index and TAIEX closing cash index. More specifically, H.Ni and H. Yin (2009) proposed a model that combines several local regressive models, including support vector regressive (SVR) and recurrent self-organizing map (RSOM) models, with influential trading indicators, for example, exponential moving average (EMA), in their model, and all those separated predictive factors were mixed with the help of the genetic algorithm (GA). The profitability of this model outperforms generalized autoregressive conditional heteroskedasticity (GARCH) model and several other benchmark models in their simulated trading process of foreign exchange rate. M. Khashei and M. Bijari [Kha11] improved G. Peter Zhang [Zha03]'s hybrid ANNs and ARIMA model by reducing possible inappropriate assumptions and changing the way of data processing in ANNs and ARIMA models. Better performance was achieved in the prediction of three well-known data sets -- Wolf's sunspot data, the Canadian lynx data, and the British pound/US dollar exchange rate data.

Along with ANN models, fuzzy technic is another important forecasting tool. The fuzzy theory was first introduced by Q. Song and B.S. Chissom [Son94] in the forecasting of university enrollments. In their meaningful work, basic definitions of fuzzy time series models were made, and the analysis framework was proposed to model fuzzy relationships among observations. Fuzzy regressions were combined with ARIMA model by F.M. Tseng et al. [Tse01], which created the FARIMA model. In consideration of possibility distribution of future values, the proposed forecast model can provide prediction intervals instead of fixed future values, which make it possible for the decision makers to forecast the best and worst possible situation. Weighted fuzzy time series models were proposed by H.K. Yu (2004) in the forecasting of Taiwan stock index, compared with former fuzzy time series models, this model assigned different weights to various fuzzy relationships which were used to be treated equally. The weighting scheme which was commonly used in local regression models was utilized in the specification of weights of different fuzzy relationships, where it is seen that the more recent ones have higher weights than the old ones. This model showed to outperform the conventional fuzzy time series models with equal weights in the forecasting of TAIEX. C.H. Cheng et al.

[Che06] improved H.K. Yu's work by introducing trend weighting scheme in the specifications of weighting parameters. At the same time, an adaptive value was assigned to the former models to make the forecasting results more reliable. This model was tested using the data of university enrolment and the TAIEX; better performance was achieved by the researchers with refined modeling processes. A hybrid fuzzy time series model was proposed by T.L. Chen et al. [Che07] by the combination of Q. Song and B.S. Chissom's [Son93] analysis framework, the concept of the Fibonacci sequence and the weighting method of H.K. Yu. The technical analysis indicator, namely the Elliott wave principle which applies the Fibonacci sequence to the prediction of the timing of stock price fluctuations was chosen in the establishment of this hybrid model. The proposed model showed lowest RMSE (Root mean square error) both in the prediction of a 5-year period of single stock TSMC and the 13-year period of TAIEX, compared with H.K. Yu and other conventional fuzzy time series models.

T.L. Chen et al. [Che08] improved their Fibonacci sequence-based model by proposing a comprehensive fuzzy time series model, in which linear relationship between recent periods of stock prices and fuzzy logical relationships mined from time series were factored into the forecasting process. One of the key contributions of this paper is the utilization of multi-period adaption method. This method is proved to be effective in improving the performance of conventional fuzzy time series models which mostly ignore the short-term factors that may affect the stock prices. T.L. Chen et al. [Che08]'s empirical study supports the point that there exist short-term stock price patterns in the Taiwan stock and Hong Kong stock markets by both traditional statistic method and their brand new proposed model. With the utilization of adaptive expectation method, C.H. Cheng et al. [Che08] developed a new model based on the conventional fuzzy time series models. As the recent forecasting errors were taken into consideration, the simulation process is more in line with the actual movements of the stock price, as reasonable investors will take certain actions to correct their "mistakes". The proposed model was proved to perform best in the prediction of TAIEX in the comparison of RMSEs (Root mean square error). P.C. Chang and C.H. Liu [Cha08] developed a Takagi-Sugeno-Kang (TSK) type fuzzy rule-based system in the prediction of the stock price. Different technical indicators were used as the input variables. The stepwise regression method was applied to select proper indicators, after which, a linear combination of those indicators were processed to form the final forecast. The proposed model was believed to have better performance compared to the back propagation neural (BPN) network models and the multiple regression analysis models in the forecasting of TSE index or a particular stock- MediaTek Inc.

H.H. Chu et al. [Chu09] revisited the mysterious relationship between the stock prices and the trading volumes with the help of the fuzzy dual factor models. By rightly selecting forecasting factors, which are highly correlated to the future stock index, the proposed model outperformed commonly multiple subjective factors fuzzy time-series models in the forecasting of the TAIEX and the NASDAQ index. G.S. Atsalakis and K.P. Valavanis [Ats09] developed a neuro-fuzzy based model to optimize the forecast of the next days' trends of certain stocks. Fifteen different combinations of past stock prices were calculated by them to identify the best type of stock prices in the prediction of the next day's trends with minimum root mean square error (RMSE). The proposed model was tested in both well-developed stock market and the emerging markets – the New York Stock Exchange (NYSE) and the Athens, both of which showed superior performance. Combining the support vector machine (SVM) and the fuzzy technic, R. Khemchandani et al. [Khe09] proposed a novel approach to forecast financial times series with noise and non-stationarity. This model was powerful and simple which requires only a single matrix inversion to calculated the regressor, and the empirical forecasting results showed acceptable normalized mean square error (NMSE). Similar to G.S. Atsalakis and K.P. Valavanis (2009), A. Esfahanipour and W Aghamiri [Esf10] studied the performance of the adapted neuro-fuzzy inference system, which based on the TSK fuzzy rule, in the analysis of stock market series. Technical indicators were tested and selected as the input variables of the proposed model, which were then linearly combined to get the forecast value of certain stock prices. The forecasting performance of this model on the Tehran Stock Exchange was showed to be superior to BPN and multiple regression analysis. Taking one more step ahead, E. Hadavandi et al. [Had10] added a genetic algorithm into the fuzzy artificial neural networks in the prediction of stock prices. In their model, stepwise regression analysis was firstly used to determine factors with significant influence on the movement of stock price, then self-organizing map was used to divide the raw data into *k* clusters, final forecasting was made by feeding all clusters into independent genetic fuzzy systems with the ability of rule base extraction and database tuning. One of the highlights of the empirical study is the separate forecasting of stock price based on different sectors. This complex model was shown to have the lowest Mean Absolute Percentage Error (MAPE) compared with all previous methods.

J.R. Chang et al. [Cha11] developed a hybrid adaptive network-based fuzzy inference system model based on AR and volatility in the analysis of Taiwan stock exchange capitalization weighted stock index. T.L. Chen [Che08] and H.K. Yu [Yu04] models were used as the benchmarks. This hybrid model increased the number of variables used in the time series analysis, at the same time, added a more economical explanation to the conventional ANNs (black box method) based forecasting models. The model outperformed benchmark models in the forecasting of TAIEX in consideration of Root Mean Square Error (RMSE). Similar to C.H. Cheng et al. [Che08] who added adaptive expectation to simple conventional fuzzy time series, L.Y. Wei [Wei13] hybridized the adaptive network-based fuzzy inference system model with adaptive expectation genetic algorithm to forecast TAIEX. With more adaptive functions added into the proposed model, the forecasting processed can take more instant responses to the recent forecasting errors. This model is shown to beat both H.K. Yu [Yu04] and T.L. Cheng's [Che08] model in the prediction of TAIEX regarding RMSE. A simple combination of fuzzy time series method and a genetic algorithm was made by Q.S. Cain and tested on the prediction of the TAIEX, the performance of the model is good regarding reducing the RMSE compared with conventional fuzzy time series.

Similar to fuzzy set series, rough sets theory is another powerful tool in the combination of different kinds of technical indicators, C.H. Cheng et al. [Che10] proposed a hybrid model based on rough sets theory and genetic algorithms. One of the advantages of this model, compared with traditional statistic ARMA and ARCH model, is the loose of statistical assumptions about variables, which makes the model closer to the reality. In this model, cumulative probability distribution approach and the minimize entropy principle approach were utilized in the construction of linguistic values. Rough sets theory algorithm was then used to extract linguistic rules from the constructed linguistic values. In the end, a genetic algorithm was used to refine the extracted rules and get a more reliable forecast of the target time series. An experimental test including simple stock return forecast and simulated trading was made in a 6-year period of TAIEX dataset. The results showed that the proposed model outperformed separated rough sets theory and genetic algorithm forecasting models.

2.5 Wavelet related literature

Most of the traditional models are dealing with the daily or weekly financial time series; these studies are of great importance in the long-run forecasting of the price of certain assets. With the development of the modern information technology, increasingly high-frequency data which used to be unable to be accessed by most investors and researchers, become available. Although many of the above models could be simply applied in the forecasting of high-frequency time series, the accuracy of such prediction may not be so good as they were in the low-frequency forecasting, since high-frequency data has its properties, which cannot be easily captured by those traditional low-frequency models. Wavelet models, which were mostly used for decades in digital signal processing, have special advantages in the predictions of short-term high-frequency time series.

Regarded as a “microscope” in mathematics, a wavelet is extensively used in signal processing, image compression, and other engineering disciplines. It is also very powerful in capturing nonlinearities in time series. Unlike most of the traditional models, which study time series only at a predetermined frequency, wavelet method regards the frequency components of the time series as a variable that affects the movement of the series. Compared to standard Fourier analysis, wavelets enable the researchers to analyze data at different scales. Information from both time domain and frequency domain is combined with each other in the analysis of the time series. One of the key advantages of wavelet-based model is that it does not make strong assumptions about the generating process of the original series. The model is to be inferred from the data, which makes it very flexible. The original time series is analyzed regarding general wavelet decomposition after the decomposition, both simple and complex conventional time series models can be fitted with the decomposed components to conduct further analysis. Another advantage of wavelet transformation based time series analysis is the capability of taking the trends and the local patterns of the time series simultaneously.

Over the last two decades, many researchers have contributed to the introduction of wavelet method to finance and economic fields. Large amounts of financial and econometric papers have been published with the application of wavelet to various areas. According to J. Ramsey [Ram96], the various applications can be categorized into four aspects. The first aspect is the processing of non-stationarity and the handling of complex functions in Besov space; the second aspect focus on the density estimation and the structural changes of economic system; the third aspect is the recognition of relationships between economic variables using time scale decomposition, as the series is analyzed under disaggregate level; the last aspect, which is also the aspect that this research is concerned of, is the utilization of wavelet in the forecasting of time series in disaggregate level. In this study, only the literature related with forecasting using wavelet is reviewed, some basic and mathematical literature is also listed for reference, namely books by I. Debauchies (1992), D. Walnut (2002), J. Walker (1999), D. Percival and A. Walden (2000) and

P. Addison (2002).

Z.R. Struzik [Str01] presented a study on the ability of the wavelet transformation in the characterization of the scale-free behavior of S&P Index through Holder exponent. In this paper, the continuous wavelet decomposition was chosen to obtain better accuracy and adaptive properties of the analyzing tool. The advantages of wavelet transformation in processing financial time series were also briefly discussed. A short-term load forecasting was conducted by I.K. Yu et al. [Yu00] using wavelet transformation. Daubechies wavelet transforms were adopted in this study to identify the load characteristics in electric systems. Instead of decomposing the historical load data, decomposed temperature data was used as the explanation variable. Each decomposed scale of data was analyzed with the AR model, which was then combined to form the final load prediction. The proposed model was proved to have good performance in the load prediction of there hours interval. Inspired by this model, some extra explanation variables can be added to the design of financial forecasting model, as there are much more data availables that are closely related to the movement of the asset price, trading volume, for instance. B.L. Zhang et al. [Zha01] investigated the effectiveness of wavelet transformation in the contribution of improving financial time series forecasting. In their study, a three stage prediction scheme, namely, the decomposition, the multi-layer perceptron, and the recombination, was raised. Another contribution of this paper is the introduction of Bayesian technique Automatic Relevance Determination (ARD) into the selection of relevant decomposed input data. The shining point of their study exists in the utilization of realistic money management system and trading model to the evaluation the forecasting performance. The trading system uses the proposed forecasting model was proved to be more profitable, which makes it the most convincing empirical experiment.

Similar to B.L. Zhang et al. [Zha01], O. Renaud et al. [Ren02] utilized wavelet decomposition into the forecasting of short (short-term movement) and long (trend) memory of the time series. An AR multi-scale model was proposed and tested with both simulated and real world data; the forecasting results were shown to outperform the model with neural network approaches. They concluded that a linear mapping of the multi-resolution data is more stable than more sophisticated nonlinear alternatives.

C. Schleicher [Sch02] was among the first to present a full practical introduction to wavelets for economists, some basic ideas and economic related applications of wavelet were illustrated. This working paper could be a good start point for researchers who want to apply wavelet to their study. P.M. Crowley [Cro05] complemented C. Schleicher's work by adding more examples of applications and more detailed explanations of different wavelets and wavelet transforms. Besides the intensive amount of literature reviewed in this paper, P.M. Crowley provided various software sources and practical guide for other researchers to conduct their empirical experiment.

S. Yousefi et al. (2005) utilized the wavelet transformation to investigate the market efficiency of crude oil futures markets. The Discrete Wavelet Transform (DWT) was used in this study to decompose the oil price into different timescale, and different models were assigned to each time scale to fit the characteristic of the data, after which the final forecasting results was achieved by the wavelet construction. The empirical results indicated that the wavelet-based forecasting procedure works best for large sample sizes, which makes it extremely suitable for the high frequency short-term stock index forecast.

J. Li et al. [Li06] designed a hybrid forecasting model by the combination of genetic programming and wavelet decomposition. Different scales of the decomposition of the original stock index were chosen as indicators to feed the financial genetic programming (FGP) model. The proposed model was used in the prediction of whether the Dow Jones Industrial Average Index (DJIA) rises by 2.2% or more within the next 21 trading days. Significant improvement of forecasting performance is observed compared to the conventional FGP.

P. Masset [Mas08] presented a full blueprint to conduct frequency domain analysis of financial time series. Starting from the identification of different frequency components of time series using spectral analysis, various filtering methods were tested. Wavelet method was introduced based on the filtering methods and Fourier analysis to overcome the limitations of conventional frequency models.

C. Stolojescu et al. [Sto10] tested a wavelet based forecasting method for time series. In this study, four prediction methodologies including Artificial Neural Networks (ANN), ARIMA, Linear regression and Random walk model were combined with the Stationary Wavelet Transform (SWT). Different mother wavelets were also applied to the decomposition process of the original data. The ANN method was shown to have the best performance compared to the rest. As the selected financial data was ERU-USD currency exchanges rate, which exhibits an almost constant tendency, whether the forecasting model is appropriate for the prediction of financial data with greater volatility remains unknown.

By replacing the conventional Gaussian kernel function with the wavelet kernel function into the support vector machine, L.B. Tang et al. [Tan09] were able to describe the stock time series both at various locations and at varying time scale. A hybrid wavelet support vector machine (WSVM) model was tested with real-world stock data. The proposed model outperforms the Gaussian kernel model in all aspects. Another sparkling point of their empirical test is the introduction of "fuzzy" theory into the preprocessing of the original data. The original series is transformed into a five-day relative difference in percentage (RDP) to improve the predictive power. In the Master's thesis of Chong Tan [Tan09], an intensive introduction and review of Neural Network (NN) and Wavelet Transform (WT) were presented. A hybrid wavelet de-noising based forecasting model was designed to fulfill the prediction task. In this model, both the Stationary Wavelet Transform (SWT) and the Discrete Wavelet Transform (DWT) was used to compare with each other to show the impact of shift-invariant. The proposed model was utilized in the forecasting of a bunch of exchange rates, whose performance was shown to beat the benchmark models only in short term forecast. Here demonstrates the feasibility of combing wavelet neural network and statistical methods, which could be modified to fit this study to conduct high-frequency short term financial series forecasting and high-frequency trading.

2.6 Forecast accuracy

Forecast accuracy is very important. A forecasting method is good or not is assessed by forecast accuracy. The root mean square error (RMSE), mean square error (MSE) have been used as a standard statistical metric to measure model performance in various kinds of studies. The mean absolute error (MAE) is another useful measure widely used in model evaluations. A.W. Alford and P.G. Berger [Alf99] formed simultaneous equations in the analysis of forecast accuracy. Several factors that affect the accuracy of forecast were taken into consideration, including the following degree of analysts and trading volume. They found that higher forecast accuracy is usually simultaneous associated with higher analysts following. Taken the over-fitting problem of ANNs models, a meta-learning technique was used by L. Yu et al. [Yu09] to improve the prediction accuracy of financial time series. In this study, different subsets were extracted from the original dataset to form separated ANNs forecast, the principal component analysis (PCA) was utilized to generate the optimal meta-model. The proposed model showed to have the smallest normalized RMSE (root mean square error) compared with ARIMA and SVM in both single and meta-models. Some more detailed forecast accuracy measurement will be described in the methodology section of this thesis.

2.7 Summary

In this chapter, an extensive literature review on HFT, financial time series forecast as well as wavelet has been conducted. A short survey on the evaluation of forecast accuracy was also conducted. Some remarks concerning the reviews can be itemized as follows.

1. Although high-frequency trading occupies a great amount of the volume in all kinds of markets, not so much literature could be found on the design of high-frequency trading systems. Some basic concepts and impact of HFT were surveyed by quite a few researchers (Charles and Maureen 1997, Wilmott 2008, A. Cartea & J. Penalva 2012). As high-frequency trading gradually gains important position in the capital market, high-frequency models become more and more popular, both investors and researchers are interested in the prediction of the financial time series in the next hour, minute and even second.

2. The empirical results from Neely and Weller (2003) and some other researchers are consistent with the efficient market hypothesis (EMH), while Kablan Abdalla [Abd09], etc. is not. Kleidon [Kle92], Kumar and Seppi [Kum94], and Holden

[Hol95] Roman and Wing [Rom12] highlight the importance of execution risk in index arbitrage under special market conditions. Frank [Fra10] and Ren S. Miller and Gary Shorter [Ren16] explored the possible impact of HFT on the market.

3. A series of fundamental related prediction was covered in this section. Different kinds of results were presented by various researchers. Some of them insisted that fundamental analysis is useful in the forecast of future returns, while other researchers proved it to be not. However, according to the movement of price in the FX markets, it is clear that the announcement of fundamentals does have a significant impact on the market. In this way, fundamental analysis could be used in the design of event trading systems.

4. Basic studies covering technic analysis were surveyed in this section. Both linear and nonlinear combinations of technical indicators systems were explored by numerous researchers from both academic and industrial background. The performance of their systems various a lot due to different reasons. Some machine learning tools including Genetic Programming (GP) and Artificial Neural Network (ANN) are also used by many researchers to recognize pattern of the financial time series.

5. I. Debauchies [Deb92], D. Walnut [Wal02], C. Schleicher [Sch02], J. Walker [Wal99], D. Percival and A. Walden [Per00] and P. Addison [Add02], gave a thorough introduction to wavelet and some related concepts. The literature on

wavelet decomposition and wavelet de-noise is also covered in this section. Wavelet is also combined with GP, ANN, and SVM models in the searching of better forecasting models. J. Li et al. [Li06] C. Stolojescu et al. [Sto10] L.B. Tang et al. [Tan09], Chong Tan [Tan09].

6. Some basic forecast accuracy valuation literature was covered in this section.

The review showed that many studies have been carried out trying to find more reliable forecasting methods for financial time series. Both investors and researchers are keen to reveal the core factor that dominant the movement of the financial time series. This study provides a statistical way of looking into the frequency domain of the time series that have a significant impact on the movements of the time series. The proposed forecast models in this thesis are used to build programmed trading strategies. With the application of multi-frequency analysis method, better forecasting performance can be achieved, based on which, modifications of the designed trading systems can be made.

In research aspect, this study contributes in finding more accurate high-frequency point and directional forecasting models. More importantly, a clear path of converting forecasting models into high-frequency trading systems and the way to improve the performance of the expert systems are introduced. The innovations of this research are described as follows:

(i) Past studies in the literature on the forecasting of financial time series are based on rather long time intervals, namely days, weeks even months. Few studies have been carried out on real-time high-frequency data. This research uses the high-frequency data of China index futures contracts in the test of the performance of the proposed forecast models and trading strategies. Technical indicators utilized in the forecasting models are calculated using only the data coming from the same trading day to avoid the impact of jumping points on the training process of the proposed models.

(ii) Genetic programming is used in this research to find the optimal linear combination of various kinds of technical indicators and order types, based on which, optimized intra-day trading system are developed on both Multicharts and the self-programmed Femas trading platform.

(iii) Taken the un-stationary characteristic of financial time series into consideration, wavelet de-noise is introduced to reduce the impact of noise on the training of the GP based trading systems. With the application of both soft-threshold and hard-threshold wavelet de-noise, better trading performance is achieved in the empirical experiments.

(iv) To explore the nonlinear relationships between various kinds of indicators, the Nonlinear Autoregressive Exogenous Model (NARX) is used in the design of nonlinear automated trading systems. Trading volume is taken as the driving (exogenous) series in the forecast of the future price. Positive returns are recoded in the back test experiments with the consideration of slippage and transaction cost.

(v) With the combination of wavelet and the Artificial Neural Network (ANN), a wavelet based Multi-Layer Perceptron (MLP) neural network is constructed to perform the next step forecast of the future price theory. Trading systems are developed based on the forecast results to test its performance.

(vi) Traditional trading related experiments always ignore the importance of holding period. In most of the studies, some futures contracts are assumed to be held for months, which is almost impossible for most futures traders – the holding risk can be so high that they can blow their account before the market reach their expected price. A maximum holding period is set in this study. For part of the studies in this thesis, all futures positions opened during the trading day are forced to close at the end of the specific day. In this way, the holding risk of the risky assets can be minimized, which makes the proposed trading strategies more reliable and practical.

(vii) In this thesis, different forecasting models are used in the design of expert trading systems. With the application of more complex and more accurate forecasting models, more efficient trading systems can be built. This research provides a clear path to covert forecasting models into practical trading systems.

(viii) With the application of multi-frequency analysis method, indicators bearing more useful information can be utilized in the decision-making process of trading orders. As more characteristics and pattern of the time series are extracted, the multi-frequency analysis provides a possible way to improve the trading performance of various kinds of current expert systems.

The next chapter will present a thorough description of the methodologies that will be used in this project. Basic concepts of the integrated models proposed in this research are also illustrated. After that, intra-day trading systems based on the proposed models will be developed and tested using the high-frequency data of the China index future.

CHAPTER 3 – RESEARCH METHODOLOGY

3.1 Introduction

In this chapter, a thorough description of the models that will be used in this project is presented. Section 3.1 covers a great deal of the traditional forecasting models that are used by past researchers. Machine learning tools like GP and ANN that are used to find optimal combinations of the technical indicators are listed in Section 3.2. Section 3.3 illustrates the multi-frequency analysis tool – wavelet, methodology about wavelet and wavelet decomposition is explained in this section. The design of trading systems based on the forecast results is shown in Section 3.4. The formulas for the evaluation of the performance of the proposed forecast models are presented in Section 3.5. Finally, some remarks concerning the methodologies used in this thesis are summarized.

3.2 Traditional Forecasting models used in the design of trend following systems

The forecasting models studied in this research are mainly used in the design of high-frequency trading systems. This sub-section covers most of the trend following related forecasting models.

3.2.1 Naïve Expection Model

Also known as the static expectation, this method was widely used as the benchmark in the early forecasting literature. It assumes that the series to be forecasted will stay at the same level as the previous one. Regarding stocks price or index value, it takes the following form:

$$E(P_t) = P_{t-1} \tag{1}$$

P_{t-1} is the previous value of the series, and P_t Is the predicted value. Nowadays, this method is always quoted as a benchmark for the newly proposed models.

3.2.2 A random walk model

The first difference of the original series is taken into consideration in this model

$$E(P_t) - P_{t-1} = \alpha \tag{2}$$

Rewrite Equation (2) in the following form by moving P_{t-1} to the right-hand side

$$E(P_t) = P_{t-1} - \alpha \tag{3}$$

Compared with the Naïve expectation, a random movement has been added into this model. Normally, α is white noise, which means it takes a zero mean and a fixed standard variance of δ. The value of the next periods equals to the last periods' value plus a constant, which representing the average changes between periods. The random walk model assumes that, from one period to the next, the original series takes a random "step" away from the previous value.

When the mean of α is zero, there is no drift in the series. However, for most financial time series, α always takes a non-zero value, which means there might be an upward trend (α>0) or a downward trend (α<0). The model with or without drift is shown in the following figures (The source code to generate the model will be shown in the Appendix).

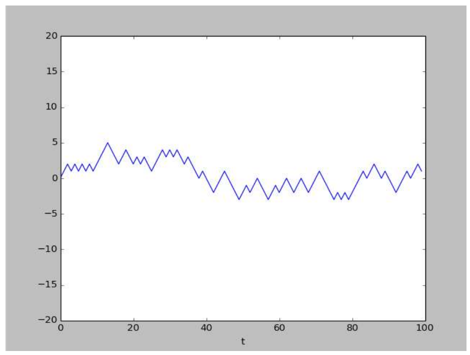

Figure 3.1 A random walk model without drift

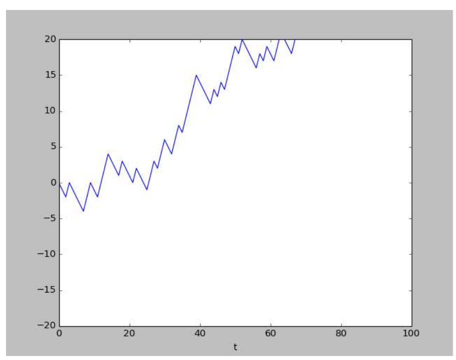

Figure 3.2 A random walk model with drift

3.2.3 An exponential smoothing model

Exponential smoothing model is another widely-used model in time series forecasting. Derived from the simple moving average (SMA) model, in which past observations are equally weighted, the exponential smoothing assigns exponentially decreasing weights over time. In the SMA model:

$$E(P_t) = \frac{1}{k}\sum_{n=0}^{k-1} P_{t-n} = P_{t-1} + \frac{P_t - P_{t-k}}{K} \tag{4}$$

The SMA has some disadvantages, the first one is that it cannot be used to forecast the first k-1 terms of value as there is no mean value available, another drawback is that some sudden peaks or troughs may largely affect the predicted value.

Another important model needs to be mentioned here is the weighted moving average model. In this model, special weighting factors $\{\omega_1, \omega_2, \dots, \omega_k\}$ is chosen to make the forecast of P_t.

$$\sum_{n=1}^{k} \omega_n = 1 \tag{5}$$

And the value of P_t is calculated by the following equation:

$$E(P_t) = \omega_1 P_{t-1} + \omega_2 P_{t-2} + \cdots + \omega_k P_{t-k} \tag{6}$$

In practice, more weight will be given to the most recent period of values and less weight to the older data in the time series, as the recent values are of greater importance to the value to be forecasted. Although the peak and trough effect is not perfectly solved in this model, the more reliable prediction could be made with proper weighting factors.

Taking one more step ahead, Robert Goodell Brown (1956) suggested the exponential smoothing model in the formulae:

$$E(P_1) = P_0 \tag{7}$$

$$E(P_t) = \alpha P_{t-1} + (1-\alpha)E(P_{t-1}), t > 1 \tag{8}$$

P_0 is a preset value to be used in the forecast process. Usually it takes the previous value of certain related value, α is the smoothing factor, it determines to how much degree that the latest period of the time series affects the predicted value. When α goes closer to one, recent terms of the series have more effect on the predicted value, on the other hand, when α goes closer to zero, the predicted value is less responsive to recent changes in the series. Practically, there is no formally correct procedure to choose α, the method of least squares might be a good choice to determine a proper α. By substituting the exponential smoothing model directly into itself, the exponential nature of the model is shown below:

$$E(P_t) = \alpha[P_{t-1} + (1-\alpha)^2 P_{t-2} + (1-\alpha)^3 P_{t-3} + \cdots] + (1-\alpha)^t P_0 \tag{9}$$

From the above formulae, the older the value, the less effect to the predicted value; the exponential smoothing model could be seen as a weighted moving average with an exponentially decayed weighting factor.

3.2.4 ARIMA model

3.2.4.1 Autoregressive model

AR(p) is the notation of autoregressive model of order p; it is written:

$$P_t = c + \sum_{i=1}^{p} \varphi_i P_{t-i} + \varepsilon_t \tag{10}$$

where P_t is the underlying time series, φ_i $(i = 1,2,\dots,p)$ are parameters to be estimated, c is a constant, and the random variable ε_t is white noise. Normally, the parameters are determined using the least square error method, and the order of the model is determined by autocorrelation test.

3.2.4.2 Moving average model

A moving average model is a linear regression of the current value against the current value and previous error terms. MA(q) is the notation of moving average model; q is the order of the model; as it is mentioned in the former model, the model can be illustrated in another form:

$$P_t = \mu + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t \qquad (11)$$

θ_i are the residual parameters to be estimated, μ is the expectation of the analyzed time series P_t, ε_t are the white noise residual terms. Same to the AR model, the parameters are also determined using the least square method, and the order of the model is determined regarding autocorrelation function and the partial autocorrelation function.

3.2.4.3 Autoregressive moving average model

ARMA model is a combination of the AR and MA model. It is one of the most widely used and most important time series analysis model. ARMA(*p,q*) is the notation of the following model:

$$P_t = c + \sum_{i=1}^{p} \varphi_i P_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t \tag{12}$$

p, q are the order of the contained AR and MA model, c is the constant term, ε_t is the white noise, φ_i and θ_i are the specified parameters to be estimated. ARMA models were popularized by Box and Jenkins because of their iterative estimating method (Box-Jenkins methods). Similar to the above-separated models, the orders *p, q* of the models are determined using the autocorrelation function and the partial autocorrelation function.

3.2.4.4 Autoregressive integrated moving average model

Autoregressive integrated moving average model is a generalization of the ARMA model to deal with the cyclicity characteristic of the underlying time series. It is mostly widely used in the cases that time series showed evidence of non-stationarity. By the approach of a differencing step of the original series, the non-stationarity could be removed. Notated as ARIMA (*p, d, q*), ARIMA model is a very important part of the Box-Jenkins time series modeling methods. *P, d*, and *q* are non-negative integers, *q* is the term of autoregressive, *d* is the difference frequency before a time sequence become smooth, and *q* is the term of moving average. For most of the time, ARIMA model is written in an operator form (Lagging operator and the difference operator).

The lagging operator noted as B, transforms the time series into its previous value, for time series P_t,

$$BP_t = P_{t-1} \tag{13}$$

The difference operator noted as ∇, transforms the time series into its first order difference,

$$\nabla P_t = P_t - P_{t-1} \tag{14}$$

From equation 13 and 14, it is clear that $\nabla = 1 - B$. In term of the lag operator, ARMA (p, q) model could be written in the following form:

$$\left(1 - \sum_{i=1}^{p} \varphi_i B^i\right) P_t = \left(1 + \sum_{i=1}^{q} \theta_i B^i\right) \varepsilon_t \tag{15}$$

B is the lag operator, p; q are the orders of the AR and MA model, φ_i and θ_i are the parameters to be estimated, ε_t is the error term, which normally follows a normal distribution with zero means: $\varepsilon_t \sim N(0, \delta^2)$, where δ^2 is the variance. Akaike information criterion (AIC, now this term is automatically generated by the most statistical program) is commonly applied to choose *p, q*. After the determination of p, q; ARMA models are usually fitted by the least squares regressions to find the values of φ_i and θ_i that minimize the error term.

Based on the formulae of the ARMA model, the ARIMA(*p,d,q*) model can be written in the form of two cascaded models:

$$X_t = (1 - B)^d P_t \tag{16}$$

$$\left(1 - \sum_{i=1}^{p} \varphi_i B^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i B^i\right) \varepsilon_t \tag{17}$$

d is the term of the difference, B is the lag operator, and X_t is a variation of the original series to form the stationarity. X_t in formulae (17) can be viewed as wide-sense stationary as the difference operation removed the non-stationarity in the original series P_t. Similar to the ARMA (*p,q*) model, AIC is mostly used in the determination of *p, q*. The least square regression remains an effective way to estimate the related parameters.

3.3 Machine learning tools related models

3.3.1 Artificial Neural Network

The ANN model processes the information in a parallel way that mimics the human brain. This model is usually presented in inter-connected (hidden layer) neurons that can process input values in a specified way by feeding information through the network. Most past studies use the multi-layer perceptron in the construction of ANN as shown in the following figure:

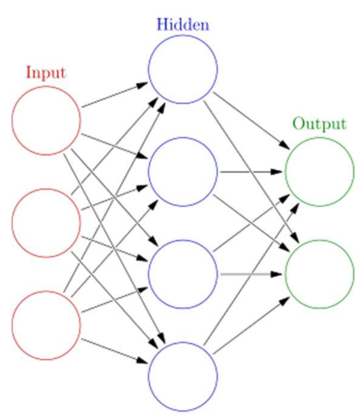

Figure 3.3 Structure of an ANN model

Normally, multi-layer perceptron consists of three layers, namely the input layer, the hidden layer and the output layer. The input layer is where the past information enters the system, and the output layer outputs the target data. There might be more than one hidden layer in the proposed model, however, with increased number of hidden layers, the difficulty to estimate the model increases geometrically. Single hidden layer feed-forward network is widely used and is enough to represent possible useful relationships between input variables regarding stock index.

The capability of machine learning and pattern recognition are the two key factors of an ANN model. As no assumption of the structure of the model is required, ANN models can approximate a large class of functions with high accuracy. The past data is divided into three parts, the feeding part, based on which the relationship parameters are calculated. The testing part, this part is used to determine the length of the feeding part to avoid the over-fitting problem. The last part is the validation part; this part is used to test the performance of the proposed model.

For general forecasting matters, ANN models can be written in the following form:

$$\mathrm{P_t} = f(x_1, x_2, \cdots x_p) \tag{18}$$

$\mathrm{P_t}$ is the valued to be forecasted, and $x_1, x_2, \cdots x_p$ are predictor variables, which are past observations in this case. Then, the model can be viewed as a nonlinear regression model:

$$\mathrm{P_t} = f(\mathrm{P_{t-1}}, \mathrm{P_{t-2}}, \cdots \mathrm{P_{t-p}}) \tag{19}$$

More specifically, the model can be written in the following form:

$$P_t = \omega_0 + \sum_{j=1}^{q} \omega_j \cdot g(\omega_{0,j} + \sum_{i=1}^{p} \omega_{i,j} \cdot P_{t-i}) + \varepsilon_t \qquad (20)$$

Where p is the number of input nodes and q is the number of the hidden nodes; $\omega_{i,j}$ (*i=0, 1, 2, …, p, j=1, 2, …, q*) and $\omega_j (j = 0, 1, 2, \ldots, q)$ are model parameters to be estimated using the past observations; ε_t is the error term. The ANN model performs a nonlinear mapping from the past observation to the future value through a certain function that is trained by past data. The training process is guided by the minimization of the following equation:

$$f = 1/2 \sum_{i=n+1}^{N} (P_i - E(P_i))^2 \qquad (21)$$

where $E(P_i)$ is the previously forecasted value P_i.

3.3.2 Genetic Algorithm and Genetic Programming

As a member of the larger class of Evolutionary Algorithms (EA), unlike ANN model, which is designed to function like neurons in the brain, a Genetic Algorithm (GA) utilizes the concepts of natural selection to determine the best solution for a problem. GA is now widely applied to all kinds of scientific fields in the adjustment of parameters to minimize or maximize some feedback measures.

Regarding financial forecasting, GA is most commonly utilized in the selection of independent variables and the corresponding weighting parameters of certain indicators that could be used in the forecasting of stock index values. The flow of the genetic algorithm is shown in Figure 3.4 below.

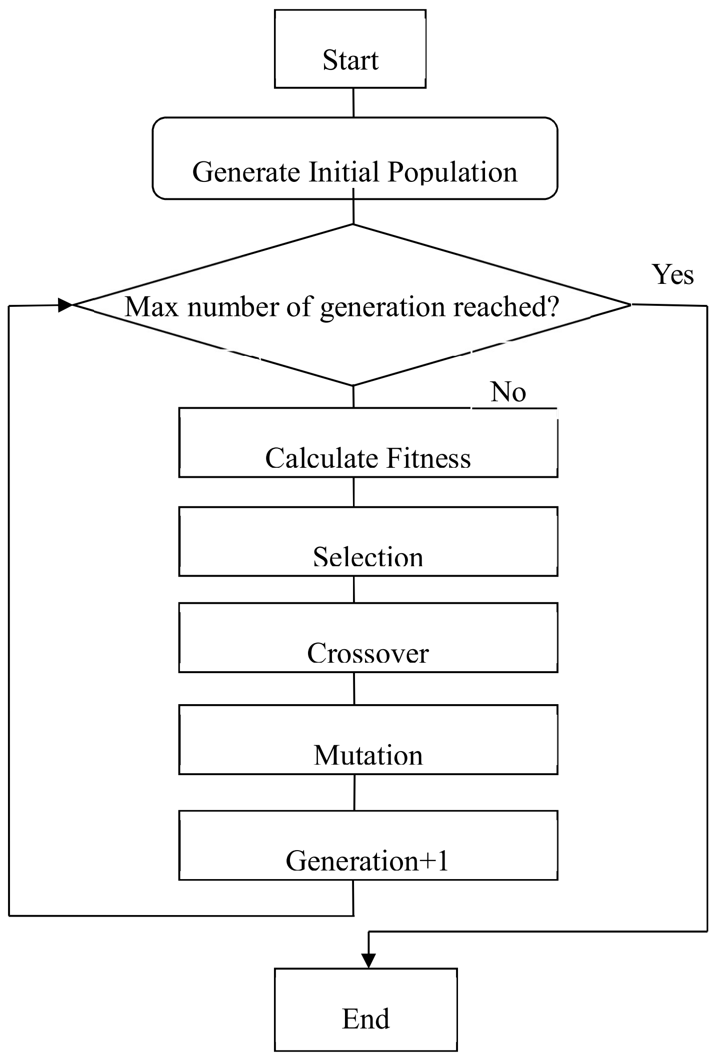

Figure 3.4 The selection process of the GA method

A GA method begins with a set of candidate solutions (parameters in this study) noted as population. A newer population is generated from the old in the hope of making a better population. The generating process is guided by the improvement of model fitness. The generating process of the population is repeated until some termination condition is met, for example, exceeding the maximum generation allowed.

In the generation of the initial population, binary encoding is the most common method for chromosomes. Chromosomes are represented by strings of 1s and 0s and each position in the chromosome represents a characteristic of the problem. The real number encoding system may also be employed if there are more than two aspects of parameters to be determined.

After the generation of the initial population, the fitness function is determined to select better population generated by genetic operators: Crossover and Mutation. The crossover operation represents the reproduction and biological crossover in biology, whereby a child takes on certain characteristics of its parents; this operation is supposed to keep the “good” parts of the parents’ chromosome as the “bad” parts are filtered out during the selection process followed by this operation. Mutations represent biological mutations and are used to maintain genetic diversity from one generation of a population to the next by introducing small random changes to the parents’ chromosomes. Offsprings created by the above crossover and mutation operations are then fed to certain fitness function to select improved results.

In this study, the fitness function is set to minimize the forecasting error; the fitness function F(x) is shown in the following equation:

$$\mathrm{F(x)} = |\text{Forecast value} - \text{Actual value}| \tag{22}$$

Over time, the process showing in Figure 3.4 will results in increasingly favorable chromosomes for the following forecasting process.

3.4 Wavelet related methodology

By overcoming the shortcomings of Fourier transform in the performing of multi-resolution analysis of signals, wavelet has been widely applied to science and engineering areas. Unlike Fourier analysis, wavelets are localized in both time and frequency scale. Mathematically, a wavelet can be defined as a function of a zero average:

$$\int_{-\propto}^{\propto} \psi(t)dt = 0 \tag{23}$$

The concept of wavelet transformation is to represent target function as a superposition of a set of wavelets which are small waves located at different times. The original series can be transformed and described in the content of different frequency over time at certain time-scale. Wavelet-based models are very suitable for non-stationary data as some transient trends, or hidden characteristics of the original series can be highlighted. Financial time series are well-known to be non-stationary, which makes wavelet transformation one of the most powerful tools in the analysis and forecasting of financial time series.

3.4.1 Continuous Wavelet Transform (CWT)

Like the Fourier transform, the continuous wavelet transform uses inner products to measure the similarity between a signal and an analyzing function. The analyzing function in the continuous wavelet transformation is a wavelet $\Psi(t)$. Shifted and compressed or stretched versions of $\Psi(t)$ is compared to the original signals to obtain a function of two variables. A continuous wavelet transform is used to divide a continuous-time function into wavelet. A continuous wavelet transforms of $f(t)$ is defined as:

$$CWT_{\psi}f(a,b) = |a|^{-0.5}\int_{-\infty}^{\infty} f(t)\Psi^{*}\left(\frac{t-b}{a}\right)dt \tag{24}$$

where * denotes complex conjugation and variables *a>0* and *b* ϵ R are the new dimensions of the wavelet transformation. Ψ(t) is a continuous function in both time domain and frequency domain called the mother wavelet. Ψ(t) is generated from the so-called mother wavelet ψ(t) by scaling and translation:

$$\Psi(t) = |a|^{-0.5}\psi\left(\frac{t-b}{a}\right) \tag{25}$$

Where *a>0* is a scale factor that reflects the width of a particular basis function, b ϵ R is a positioning factor that points out the translated position along the t axis and the |a|^(-0.5) is used for the energy normalization across the different scales. Not only do the values of scale and position affect the continuous wavelet transformation coefficients, but the choice of mother wavelet also affects the values of the coefficients. One of the strengths of wavelet analysis is that there are many different admissible wavelets can be used in the continuous wavelet transformation as different wavelet can be selected to suit the specific characteristics of a certain series.

3.4.2 Discrete Wavelet Transform (DWT)

For most practical studies, there is no need to calculate wavelet coefficients at every possible scale as it would be a fair amount of work to do that, which generates a great deal of data. The discrete wavelet transformation projects a time series onto a collection of orthonormal transformation. Unlike the continuous wavelet transformation that considers all the possible frequencies at continuous times, discrete wavelet transformation focuses on some specific frequencies at distinct times. The discrete wavelet is defined as:

$$\psi_{j,k}(t) = S_0^{-\frac{1}{j}} \psi\left(\frac{t - k\tau_0 S_0^j}{S_0^j}\right) \tag{26}$$

where *j* and *k* are integers, $S_0 > 1$ is a fixed dilation step and the translation factor τ_0 depends on the dilation step, when $S_0 = 2$, the discrete wavelet transformation is defined as dyadic wavelet transform. By definition, the scaling function and the wavelet function of discrete wavelet transformation are:

$$\varphi(2^j t) = \sum_{i=1}^{k} a_k \, \varphi(2^{j+1} t - k) \tag{27}$$

Equation 27 is the scaling function, also known as the scaled father wavelet, a, k are the corresponding coefficients, $k \epsilon Z$. The mother wavelet ψ is obtained by the linear combinations of the scaled father wavelet. Proper coefficients must be selected to maintain the orthogonality of the basis wavelets.

$$\psi(2^j t) = \sum_{i=1}^{k} b_k\, \varphi(2^{j+1}t - k) \quad (28)$$

The mother wavelet ψ are uniquely determined by their coefficients $\{b, k\}$, $k \epsilon$ Z.

And then, a signal f(t) can be written as:

$$f(t) = \sum_{i=1}^{k} a_{j,k}\, \varphi(2^{j+1}t - k) + \sum_{i=1}^{k} b_{j,k}\, \psi(2^{j+1}t - k) \quad (29)$$

A common example is the Haar mother wavelet ψ:

$$\psi(t) = \begin{cases} 1 & \text{for } t \in [0, 0.5] \\ -1 & \text{for } t \in [0.5, 1] \\ 0 & \text{for } t \notin [0, 1] \end{cases} \quad (30)$$

The corresponding father wavelet φ is:

$$\varphi(t) = \begin{cases} 1 & \text{for } t \in [0,1] \\ 0 & \text{for } t \notin [0,1] \end{cases} \quad (31)$$

In this case, the scaling equation for the father wavelet is:

$$\varphi(t) = \varphi(2t) + \varphi(2t - 1) \quad (32)$$

and $a_0 = a_1 = 1$, $a_k = 0$ for $k \neq 0,1$. To keep the orthogonality of the basis wavelets, $b_0 = 1$ and $b_1 = -1, b_k = 0$ for $k \neq 0, 1$.

3.4.3 The Wavelet decomposition

The original signal can be decomposed into different elements based on different scales using the cascade algorithm, as shown in the following equation:

$$f(t) = A_1(t) + D_1(t)$$

$$= A_2(t) + D_2(t) + D_1(t)$$

$$= A_n(t) + D_n(t) + D_{n-1}(t) + \cdots + D_1(t) \quad (33)$$

where $A_n(t)$ are the approximation coefficients at scale n, and $D_n(t)$ are the detail parts of the original signal at scale n. Prediction of the original signal can be made by the reconstruction of the different scales of wavelets with proper coefficients.

For signals that are regularly updated (i.e. live trading data), it is important to be clear which pixels of the input signal are used to calculate the last wavelet coefficient in the different scales as shown in Figure 3.5.

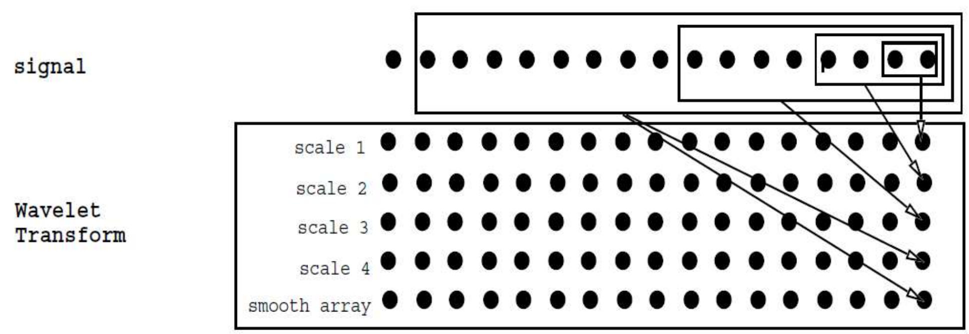

Figure 3.5 Multi-scale decomposition of the signal

3.5 Design of the trading systems based on the above models

It is obvious that the above models try to give a point or interval prediction of the time series shortly, trading actions can be taken based on the analysis of the forecast results. Regarding the automated trading system, three aspects need to be taken into consideration are:

3.5.1 Entry point

This is the point the buy or sells short order is placed into the market, after the execution of that kind of orders, the portfolio ends up with either holding long or short positions of the selected targets. With the application of the above forecasting methods, the entry point is quite easy to be selected. If the point prediction is above (below) the price of the specified targets at the current time, or the directional prediction of the financial time is up, an entry point of long (short) the specified asset is determined.

3.5.2 Leaving point

This is the point that the trader decides to leave the market either by the placement of short or long to cover orders. After the executions of such kind of orders, the portfolio ends up with zero position of a specified target. The leaving point can be reached by exceeding the maximum trading profit or maximum trading loss. It can also be triggered by the predictions of the above models, if the point prediction of the financial time series is below (above) the current price of the targets, a leaving point of the long (short) positions of the assets is reached.

3.5.3 Risk management

Regarding risk management in an automated trading system. Three aspects of the risks need to be considered, namely, execution risk, liquidity risk and the risk to go bankruptcy. The execution risk is the risk that the trading orders placed by the automated trading systems may not be executed due to the sharp movements of the price. The liquidity risk is quite similar to the execution risk. It is a kind of risk that a trader is not willing to buy the assets from the market. The bankruptcy risk is the risk that most traders do not have that much money to wait for the market to the "right" place, holding the assets could mean very serious risky. In this research, the execution risk is carefully taken care of as there is very little other kinds of risks needed to be considered as the index futures is being traded very extensively at all kinds of markets. The risk that should be taken good care of is to make sure the execution of the orders placed by the automated trading systems.

3.6 Evaluation criteria of the forecasting results

Three are many types of forecasting accuracy measurements, the list below shows the traditional prediction accuracy measurements that have been adopted in the comparison of the performance of different models.

1. Root mean squared error (RMSE), RMSE is defined as:

RMSE= $\sqrt{\frac{\sum_{i=1}^{N}(predicted\ value - actual\ value)^2}{N}}$ *N* is the number of predictions (35)

2. Mean square error(MSE), MSE is defined as:

MSE= $\frac{\sum_{\mathrm{i=1}}^{\mathrm{N}}(\mathrm{predicted\ value - actua\ \ value})^2}{\mathrm{N}}$ N is the number of predictions (36)

3. Mean absolute error(MAE), MAE is defined as:

MAE= $\frac{1}{N}\sum_{i=1}^{N}|predicted\ value - actual\ value|$ *N* is the number of predictions (37)

4. Mean absolute percentage error (MAPE), the above three measurements are all not scaled independent, which means they can only be used to assess accuracy in a single series. MAPE has the advantage of being scale independent; this method is widely used to compare forecast performance between different data series—this property makes it extremely suitable for this study as series with different frequency will be analyzed separately. MAPE is defined as:

$$\text{MAPE}=\frac{1}{\text{N}}\sum_{\text{i}=1}^{\text{N}}\left|\frac{\text{predicted value}-\text{actual value}}{\text{actual value}}\right| \text{ N is the number of predictions} \quad (38)$$

5. Another important measurement is the tracking signal (TS); this method is very powerful in the discovery of sudden changes in the series. The tracking signal is used to pinpoint forecasting models that need adjustment. TS is defined as:

$$\text{TS}_i = \frac{\sum_{i=1}^{N} Actual\ value - predict\quad value}{MAD} \quad (39)$$

N is the number of prediction and MAE is the mean absolute error. As long as the tracking is between - 4 and 4, the forecast model is assumed to be working correctly.

Listed above are some traditional measurements for forecasting accuracy. Some special evaluation criterias for the point forecasting of the time series will also be introduced to gain better understand of the forecasting results, based on which, further application of the forecasting results could be achieved efficiently.

6. The hit rate (HR) of the forecast is probability that the forecasted value is exactly the same as the actual value; HR is defined as:

$$\text{HR}=\frac{1}{\text{N}}(\text{Number of hit forecast}) \quad N \text{ is the total number of the forecast} \quad (40)$$

7. The over forecast rate (OFR) of the forecast, this indicator shows the rate of the forecast value is bigger than the actual value. OFR could be used to improve the performance of the proposed model by some proper modifications. OFR is defined as:

OFR= $\frac{1}{N}(\text{Number of over forecast})$ N is the total number of the forecast (41)

8. The under forecast rate (UFR), similar to the OFR, this indicator shows the rate of the forecast value that is smaller than the actual value. UFR is defined as:

UFR= $\frac{1}{N}(\text{Number of under forecast})$ N is the total number of the forecast (42)

The last and also the most useful way of the forecast accuracy measurement is the profitability of the programme trading based on the proposed direction and point forecast of the financial time series.

3.7 Summary

In this chapter, a detailed presentation of the related methodologies that are used in this project is conducted. Some mostly used traditional forecasting models are briefly introduced. Machine learning tools including Genetic Programming and Artificial Neural Network are listed. Multi-frequency analysis tool—wavelet is also covered. In the end, the evaluation criteria for the performance of the forecasting models and the basic step to design trading systems are described. The detailed content of the design and execution of programmed trading systems based on the proposed forecasting model will be discussed in the following chapters.

CHAPTER 4 –INTRADAY TRADING SYSTEMS BASED ON MOVING AVERAGES

4.1 Introduction

The ability of simple rules, such as technical analysis, to forecast asset price movement was a controversial subject, especially during the late 90s. Technical analysts were held in disdain by the academic community during that time, however, with the utilization of technical trading rules in high-frequency trading systems, more and more researchers are trying to find out the rational motivation behind technical analysis.

Moving averages may be the most universal of all technical analysis indicators. They do not predict time series' direction, but rather define the current direction with a lag. The lag exists because moving averages are based on past time series. Despite this lag, moving averages are the simplest and powerful tools to smooth time series and filter out most of the noise.

However, according to Fama's Efficient Market Hypothesis (EMH), it is impossible to make profits by exploiting all currently available information because assets prices have fully digested and reflected all private and public information in an efficient market. In 1978, three broad categories of hypotheses were developed by Jensen, namely

I: The weak-form of the EMH, in which asset prices reflect all market information and that the rates of return on the market should be independent, past rates of return have no effect on future rates.

II: The semi-strong form of EMH, in which asset prices reflect all publicly available information. The Semi-strong EMH incorporates the weak form EMH, and investors cannot benefit over and above the market by trading on the new information.

III: The strong-Form of the EMH, in which the asset prices reflect all information both public and private. No investor would be able to profit above the average investor even if he was given private information of the target asset.

In this project, intra-day trading systems are built based on moving averages, which are computed using historical information. The trading systems will be tested using the intraday data of the emerging China index futures market. If any systems showed consistent over market return, a conclusion could be made that the weak-form EMH won't hold in the China index futures market.

4.2 Data and Methodology

A great deal of literature on technical analysis was reviewed, and most of the past studies focusing on the long-term profitability and most the data sets utilized were past daily, weekly monthly or even yearly data. In this project, intraday day systems are built to fully exploit the short-term prediction advantage of technical analysis.

4.2.1 Dataset used in this chapter

Of all the financial time series, index future is one of the most active trading targets that enjoy high trading volumes in both developed and emerging market. For the empirical analysis, the intraday 1-minute data of the CSI 300 Index Futures is used. The CSI 300 Index Futures is electronically traded at the China Financial Futures Exchange. The underlying asset of this futures contract is CSI 300 index, and the contract multiplier is 300 per index point. As an emerging index futures markets, many arbitrage opportunities may exist since the contract was not introduced until 2010. The 1-minute data of IFLX0 from 20150901 to 20151030 will be used in this project to test the moving average trading rules. The first month of the data (20150910 to 20150930) will be used to train the model to find out the optimal lag value for each moving average. The second month of the data (20151001 to 20151030) will be used to conduct the out-of-sample test. The out of sample performance of each moving average will be compared with each other to find out the best available moving average. The daily candlestick figure of the data set with moving average line is shown in the following Figure 4.1.

Figure 4.1 Candle-stick figure of the data set

4.2.2 Three kinds of moving average

A simple moving average, also known as an arithmetic average is formed by computing the average price of an asset over a specific number of periods. Each price in the data series is equally weighted; no weighting factors are considered. As most moving averages are based on closing prices (some more representatives of the prices from a certain trading period will be considered in further study), a simple moving average is calculated by adding the closing price of the security for a number of time periods and then dividing this total by the number of time periods. Short-term averages respond quickly to changes in the price of the underlying, while long-term averages are slow to react.

Exponential moving average, known as EMA, is a moving average that is similar to a simple moving average, except that more weight is given to the latest data. EMA is a member of weighted moving average; EMA reduces the lag of SMA by applying more weight to recent prices. The weighting applied to the most recent price depends on the number of periods in the moving average. The three steps to calculate the EMA is shown below.

(1) Calculate the SMA

$$SMA = Sum(N \text{ periods of close prices})/N \tag{43}$$

Sum() is a function to calculate the sum of the values. N is the number of lag periods to be considered. SMA is used as the previous period's EMA in the first calculation.

(2) Calculate the weighting multiplier

Multiplier: 2/(N+1) (44)

(3) Calculate the EMA

EMA = (close-EMA(previous period))*Multiplier + EMA(Previous period) (45)

Where close is the close price for each period; Multiplier is calculated in the second step The last moving average used in this chapter is the Hull moving average.

The Hull Moving Average makes a moving average more responsive while maintaining a curve smoothness. The formula for calculating this average is as follows:

HMA(i)=SMA((2*SMA(close, N/2)-SMA(close, N), SQRT(N)) (46)

Where HMA(i) is the hull moving average, SMA() is the function to calculate the simple moving average, the first parameter is the data series, the second parameter is the lag value, N is the number of lag, close is the data series.

4.2.3 Moving average trading

Moving average trading is a practice of systematically or programmed trading of buying and selling whenever the price crosses its average. The trading logic is that at each point in time, the price is either in an uptrend or a downtrend. An uptrend is a period when prices are rising, and a downtrend is a period when prices are dropping. There are various kinds of moving average trading strategies in consideration of the number of the moving average line used. The most simple one is the one in which only one moving average is used, the current price is compared with the moving average value, if the price cut down through its average from above, a sell signal occurs, if the price cuts up through its average from below, a buy signal occurs. Another commonly used trading strategy is the double crossover, in this strategy, two moving average is used, namely the long-lag moving average and the short-lag moving average. For double crossover

strategy, when the short-lag moving average cut up through the long-lag moving average, a buy signal is indicated, similarly, when the short-lag moving average cut down through the long-lag moving average, a sell signal is indicated. Triple crossover strategy can be built by adding another moving average into the trading decision-making process, which is much more process. To simplify this study, only the one moving average method will be used in this project.

4.3 Empirical Test

To simplify the study, the trading volume and bid-ask spread are not considered in this project. For all of the three kinds of moving average, the first month's data will be used to find out optimal lag value for each moving average. Regarding the intraday trading strategy, no position is held overnight. The trading process is illustrated in the following figure 4.2.

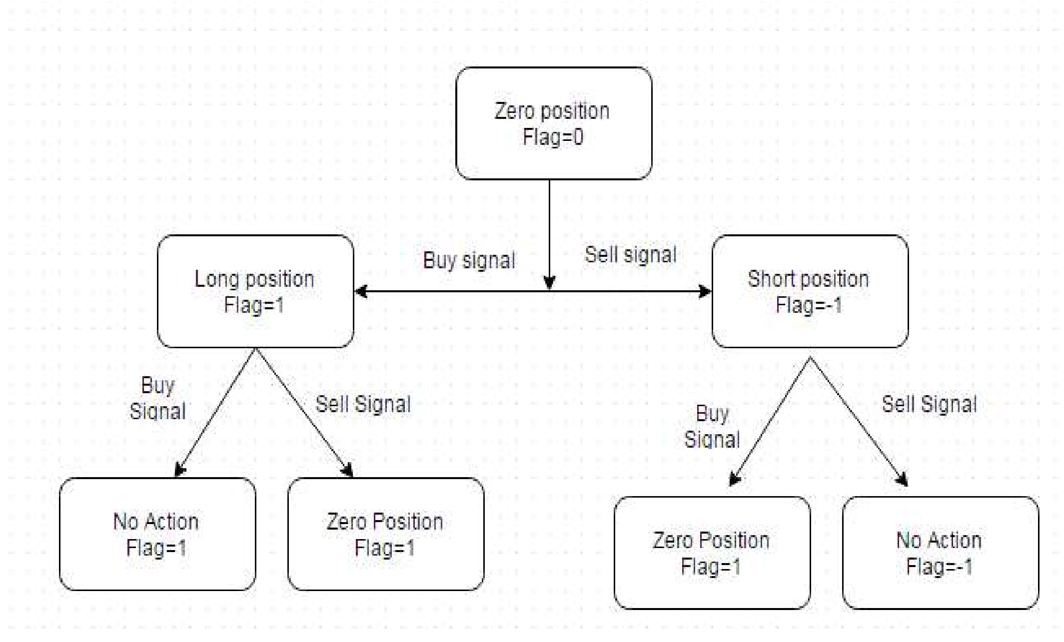

Figure 4.2 Trading Process Illustration

At the end of each trading day, all positions will be closed. The lag value for each moving average will be selected using the enumeration method from 3 to 30. The objective function of selecting lag value is to maximize the cumulative profit using the first month's data. The value selected, and the cumulative profit for each moving average is shown in the following tables. Top 5 Net Profit of the simple moving average is listed in Table 4.1.

Table 4.1 Top 5 Net Profit of in sample test of SMA

Net Profit	Gross Profit	Gross Loss	Total Trades	% Profitable	Winning Trades	Losing Trades	Max Intraday Drawdown	lag (SMA)
591420.00	1826580.00	-1235160.00	1850.00	43.46	804.00	1008.00	-57180.00	3.00
520140.00	1589040.00	-1068900.00	1544.00	42.55	657.00	862.00	-65040.00	4.00
387600.00	1143900.00	-756300.00	802.00	39.03	313.00	475.00	-95100.00	10.00
340080.00	1180020.00	-839940.00	922.00	37.09	342.00	559.00	-80520.00	9.00
291720.00	1375860.00	-1084140.00	1383.00	38.54	533.00	823.00	-85500.00	5.00

Top 5 Net Profit of exponential moving average is listed in Table 4.2.

Table 4.2 Top 5 Net Profit of in sample test of EMA

Net Profit	Gross Profit	Gross Loss	Total Trades	% Profitable	Winning Trades	Losing Trades	Max Intraday Drawdown	lag (EMA)
496020.00	1675680.00	-1179660.00	1753.00	41.07	720.00	997.00	-53040.00	3.00
342300.00	1430460.00	-1088160.00	1543.00	38.17	589.00	921.00	-78960.00	4.00
327120.00	1252500.00	-925380.00	1158.00	36.61	424.00	709.00	-87180.00	7.00
294780.00	1356780.00	-1062000.00	1425.00	37.33	532.00	860.00	-76260.00	5.00
293520.00	1291020.00	-997500.00	1288.00	35.87	462.00	796.00	-88020.00	6.00

Top 5 Net Profit of hull moving average is listed in Table 4.3.

Table 4.3 Top 5 Net Profit of in sample test of HMA

Net Profit	Gross Profit	Gross Loss	Total Trades	% Profitable	Winning Trades	Losing Trades	Max Intraday Drawdown	lag (HMA)
291180.00	1408860.00	-1117680.00	1149.00	41.86	481.00	647.00	-52980.00	16.00
102540.00	753720.00	-651180.00	359.00	42.34	152.00	206.00	-83700.00	26.00
69540.00	185880.00	-116340.00	20.00	55.00	11.00	9.00	-101520.00	5.00
69540.00	185880.00	-116340.00	20.00	55.00	11.00	9.00	-101520.00	6.00
69540.00	185880.00	-116340.00	20.00	55.00	11.00	9.00	-101520.00	7.00

In the one month out of sample trading test, the lag value with the highest net profit in the training period is selected. The detailed equity curve for out of sample simple moving average with a lag value of 3 is shown in the following figure 4.3.

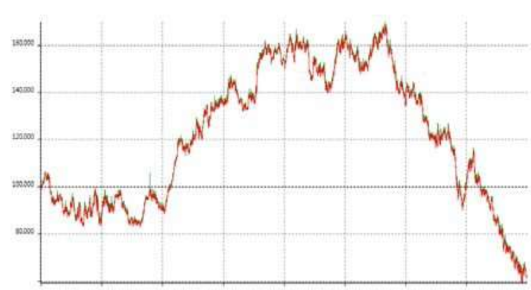

Figure 4.3 Out of sample equity curve of the SMA strategy

The detailed equity curve for out of exponential moving average with a lag value of 3 is shown in figure 4.4.

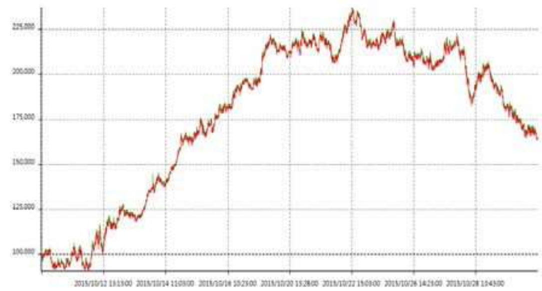

Figure 4.4 Out of sample equity curve of the EMA strategy

The detailed equity curve for out of hull moving average with a lag value of 3 is shown in figure 4.5.

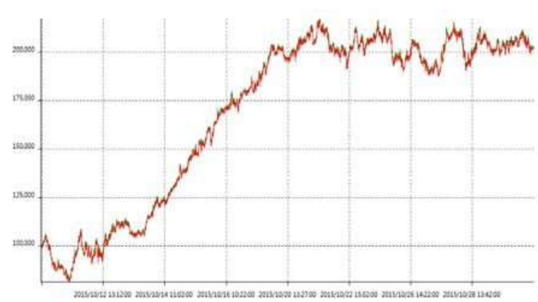

Figure 4.5 Out of sample equity curve of the HMA strategy

As shown in the above figures, the out of sample performance of SMA strategy suffered a great loss in the last few trading days, which resulting in a total of loss in the out of sample period. The drawdown of the EMA strategy is smaller than the SMA strategy and can maintain a profit position in the out of sample period. Compared with the other two moving average strategies, the HMA strategy has a much more stable performance during the out of sample period. The out of sample performance of EMA and HMA strategies are inconsistent with weak-from market efficiency hypothesis as the trading decisions are simply made based on historical data.

4.4 Summary

In this chapter, three moving average based trading systems are constructed and tested using the China Index future. Some remarks can be summarized as follows.

1. The advantages of using moving averages need to be weighed against the disadvantages. Moving averages are trend following, or lagging, indicators that will always be a step behind. As with most technical analysis tools, moving averages should not be used on their own, but in conjunction with other complementary tools.
2. Both the SMA and EMA strategy have a significant drawdown in the period of the out of sample test, and the drawdown of the SMA strategy is so big that the account went into total lose in the end. The performance of the HMA strategy is much more stable compared with the other two. The drawdown of the HMA strategy is much smaller than the SMA and EMA strategies.
3. The out of sample profit performance of EMA and HMA trading strategy indicate that the China CSI index futures market is not weak-form efficient.
4. Although the simple EMA and HMA trading strategies showed a cumulative profit from the market, much more experiments must be done before these trading systems are deployed into the real trading environment. This is because that the experiment results may suffer from overfitting problem and

that emerging market assets are heavily influenced by the government policy, past trading environment may be different with the current one.

For the design of trading systems based on moving averages, we have covered basic knowledge about this traditional forecast of financial time series in Chapter 3, the key steps to creating trading systems is also mentioned in Chapter 3. For the rest of this project, the performance of these strategies in this chapter is mainly used as the benchmark for some more complex trading systems. However, the out of sample performance of the HMA strategy is indeed quite convincing. This system could be possibly tested in a real market with some more consideration on the transaction cost and the slippage.

As only one interval of data is used and no multi-frequency technic is applied in this chapter, a manual combination of two frequencies data analysis will be presented in the next chapter. The forecasting performance of the proposed two frequency model will be presented, after that, a simple trend following trading system without any consideration of any forecast will be also showed.

CHAPTER 5 – TWO FREQUENCY ARIMA

5.1 Introduction

In this chapter, two frequencies analysis – an example of the proposed multi-frequency method is presented. The 1-minute and 3-minute data set are selected as the experiment data. The experiment data in this period is divided into two sub-periods. The first half of the data will be used for the in-sample model estimation, and the second half of the data will be used for out of sample model validation. The famous Auto-Regressive Moving Average (ARMA) model is used to make the short-term prediction for both the 1-minute and the 3-minute data set, a combination of the 1-minute and 3-minute data forecast results is made to form the final prediction value. After that, the performance of a simple trend following system is also illustrated in this chapter.

5.2 Data Description

As mentioned above, this project focuses on the prediction of high-frequency financial time series, so the target time series must have enough liquidity to match the high-frequency models. Of all the financial time series, Index futures is one of the most active trading targets that enjoy high trading volumes in both developed market and emerging market. Therefore, Index Futures in the China Securities Index (CSI) 300 Index Futures (IF) is selected as the target data set to test the proposed model.

The CSI 300 Index Futures is electronically traded at the China Financial Futures Exchange. The underlying asset of this futures contract is CSI 300 index, and the contract multiplier is 300 Yuan per index point. As an emerging index futures markets, many arbitrage opportunities may exist since the contract was not introduced until 2010. Different from the most of the markets in the world, which have no up and down limit, the CSI 300 index futures contracts has an up and down limit of 10 percent per day. The high-frequency data of CSI index futures contracts was collected using the software Touchance 3.0. The real time tick data is also provided by Touchance 3.0; programmed trading systems can be executed by the combination of Touchance 3 and Multicharts.

As this project focuses on the forecasting of very high-frequency time series, there is no need to introduce years of trading data into consideration. Besides, such amount of unnecessary data will cost a very long time to compute. The 1-minute, 3-minute, and trading data from 2013 January 1st to 2014 January 1st are chosen as the experiment period. There are 75618 1-minute data and 25200 3-minute data in the selected period. The figures of the data are shown below.

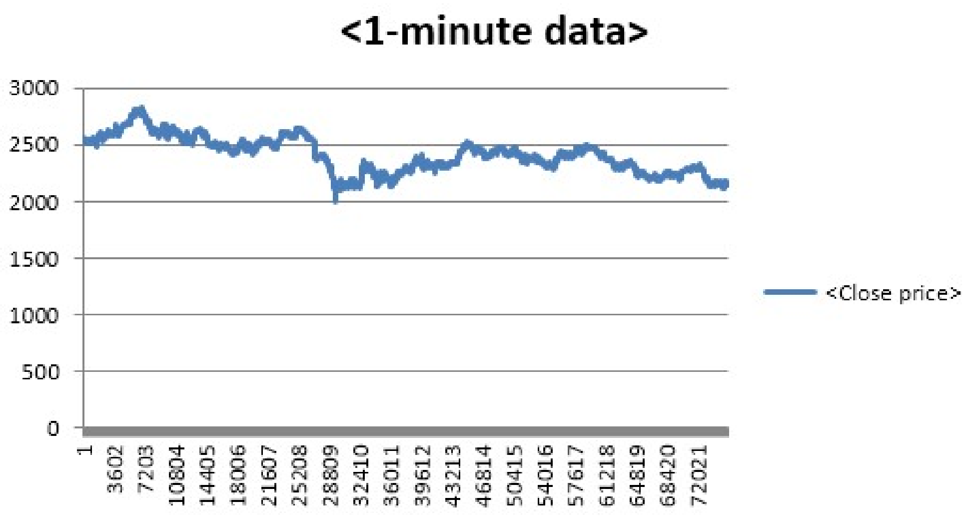

Figure 5.1 1-min data for the CSI index future

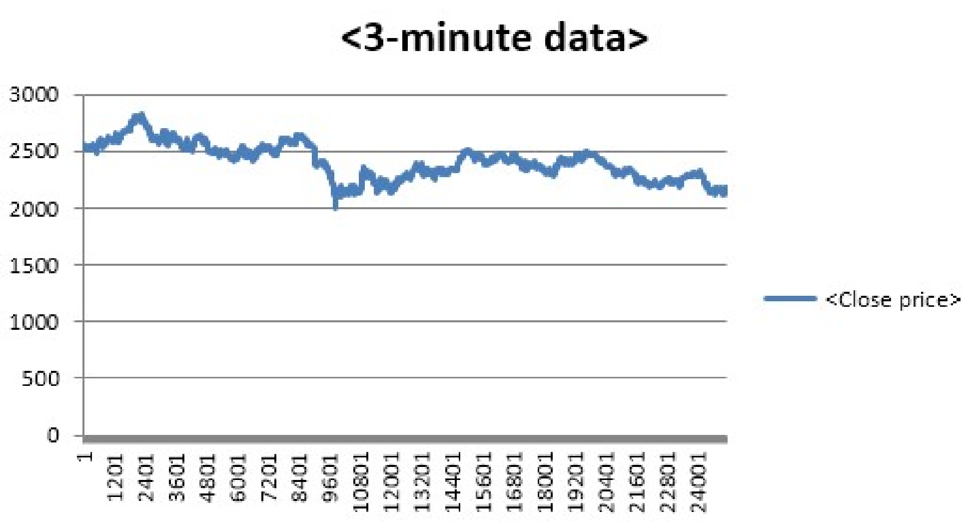

Figure 5.2 3-min data for the CSI index future

The only difference between Figure 5.1 and 5.2 is the frequency of the close price being collected from the trading market. The low frequency data set is used to capture the main trend of the series and the high frequency data is to reveal the hidden characteristics of the series.

The trading data from this period will be equally divided into two sub-periods. The first period of the data will be used to determine the parameters and the second period will be used to form the foresting results. After the selection of the most efficient way to forecast the index future, real time data will be introduced to make real time predictions of the index future, based on which, trading orders will be placed by programmed trading platform to test the profitability of the proposed model.

5.3 The proposed forecasting model and the forecast procedure

The prediction in this chapter is made based on the combination of two ARIMA models on different frequencies. To simplify the calculation part of this study, only the last month's trading data shown in Figure 5.1 and 5.2 is used as the experiment data in this chapter.

The prediction process of this chapter is shown in the following Figure 5.3.

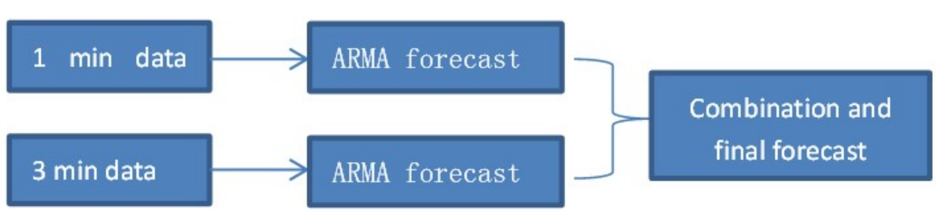

Figure 5.3 The prediction process for this study

For the simple ARIMA model, the autocorrelations of the time series need to be calculated to determine the parameters for ARIMA. *R* will be used to conduct the calculation.

Many researchers believe that the price of most financial assets move in a random pattern, and that it is impossible to make reasonable short term forecast of the financial time series. The auto correlation coefficient function is calculated here to show the predictability of such kind of series. Only the 3-minute data set is used in this calculation to minimize the computing, this is feasible because if the 3-minute data shows high auto correlation, the 1-minutes data must have even higher auto correlation, as it is an expanded collection of the 3-minute data.

Figure 5.4 The ACF of the 3-minute data set

Figure 5.4 showed that the auto correlation coefficient is nearly a constant at a lag of 20, so the high-frequency data hold a rather “long” period of memory. At the same time, all the ACF coefficients are above zero, a high monotonicity exists in the financial data set, even though there might be some sudden turning point. The correlation coefficient denies that the series moves randomly so that the short-term forecasting could be conducted based on thorough analysis of the characteristic of the series.

As mentioned before, in this study, only the last month’s data is used, the following figure showed the basic description characteristic of the data set that will be utilized in the following experiments.

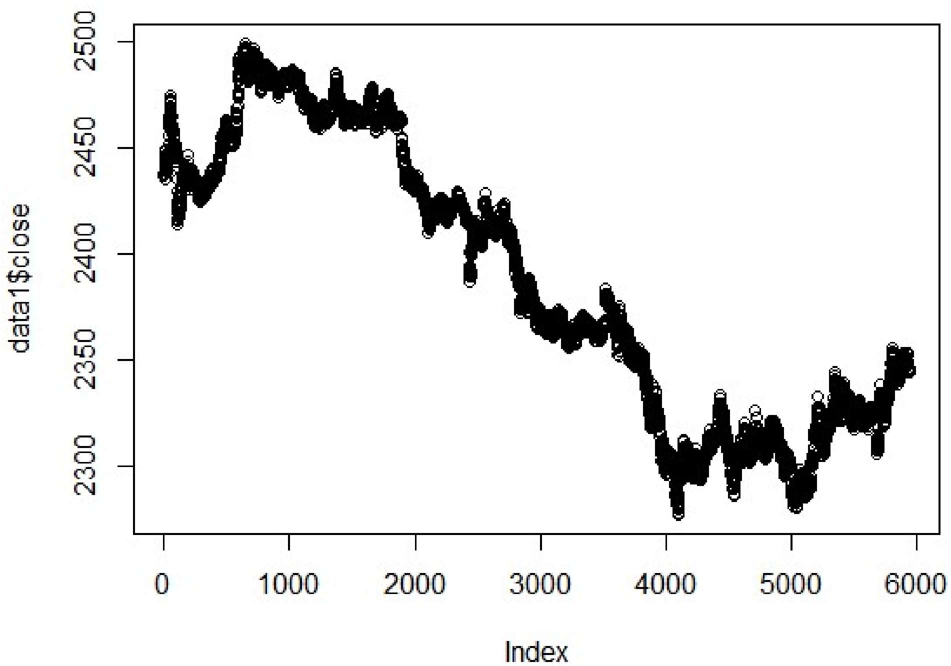

Figure 5.5 1-minute close price of the last month

The mean of this 1-minute data is 2386.896, the maximum price is 2498.8, the minimum price is 2277, the count of the price is 5939, and the standard difference is 64.94932.

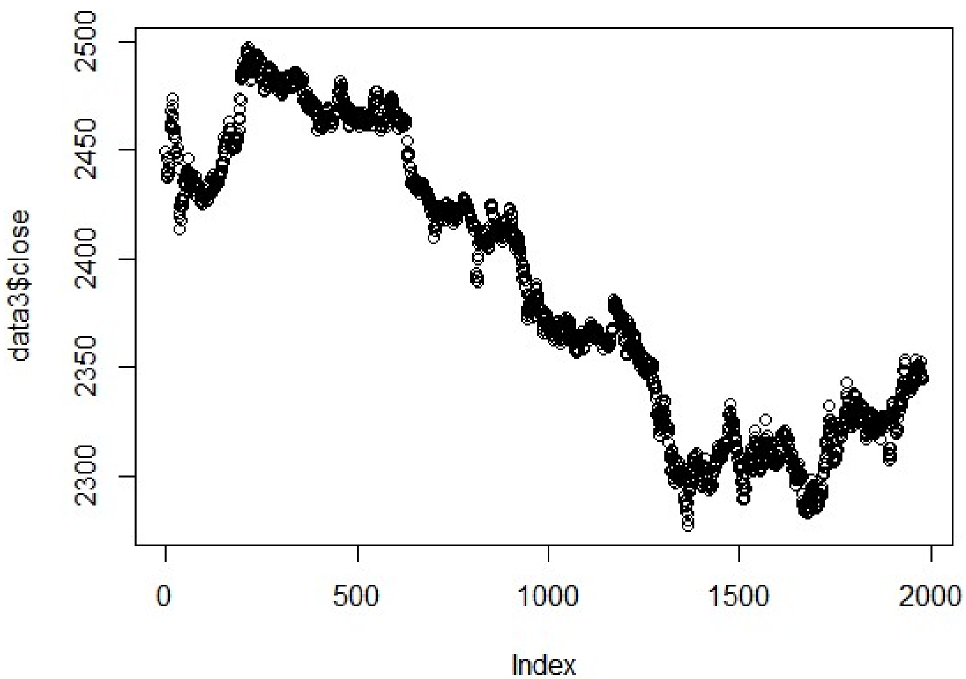

Figure 5.6 3-minute close price of the last month

Figure 5.6 shows that there is less number of prices, this is due to the decrease in sampling frequency. The mean of the 3-minute data is 2386.871, the maximum price is 2497.4, the minimum price is 2277, the count of the price is 1979, and the stand difference is 64.96562.

5.3.1 1-minute data ARIMA forecast

In this specific 1-minute data forecast, the first 4588 price points will be used to determine the ARMA model, and the rest of the data will be used to conduct out of sample forecast. The R package "forecast" is used in this model to simplify the model. The auto.arima() function is used to determine the basic model, the result is shown in the following figure.

```
> fit<-auto.arima(determine1$close)
> print(fit)
Series: determine1$close
ARIMA(4,1,3)

Coefficients:
          ar1     ar2     ar3      ar4     ma1      ma2      ma3
      -0.1948  0.2166  0.1856  -0.0133  0.1781  -0.2276  -0.1564
s.e.      NaN  0.7095  0.6469   0.0101     NaN   0.7444   0.6734

sigma^2 estimated as 2.312:  log likelihood=-8430.49
AIC=16876.97   AICc=16877   BIC=16928.42
```

Figure 5.7 The ARIMA model of 1-minute data

Figure 5.7 shows that the ARIMA(4,1,3) fitted the model best, a test of the validity of the model is shown in Figure 5.8

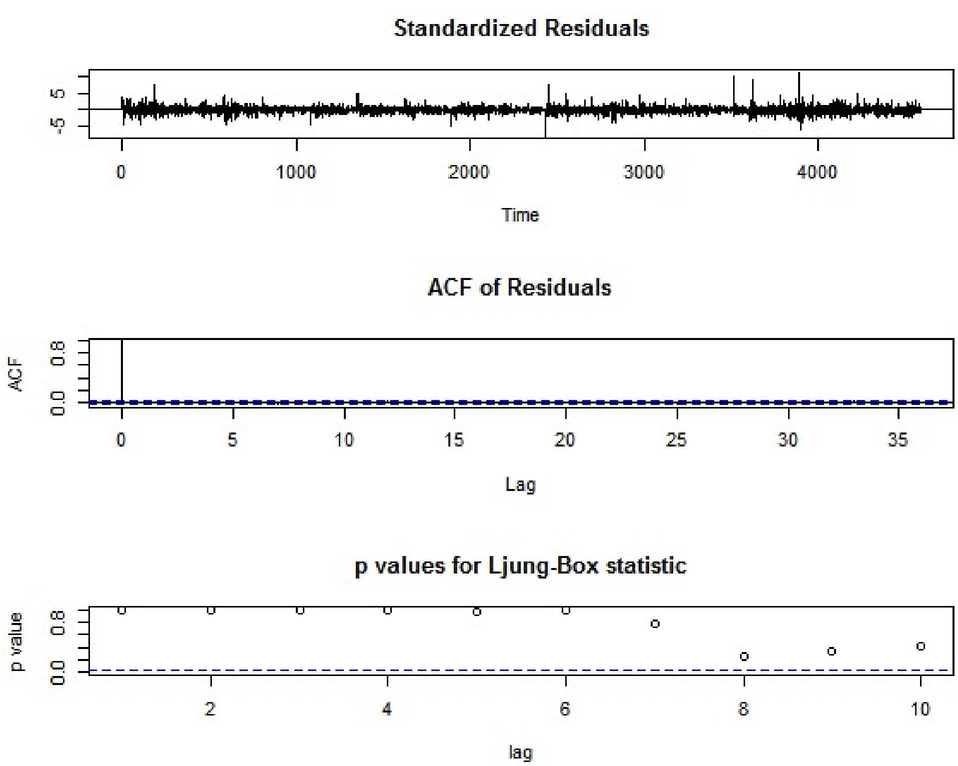

Figure 5.8 Validity test of the fitted model

The ACF test in Figure 5.8 indicates that the residuals do not show significant auto-correlation, and the P values for Ljung-Box statistic are greater than 0.1. The validity of the fitted model is acceptable.

After the determination of the model, the out sample forecast will be conducted to test the efficiency of the model.

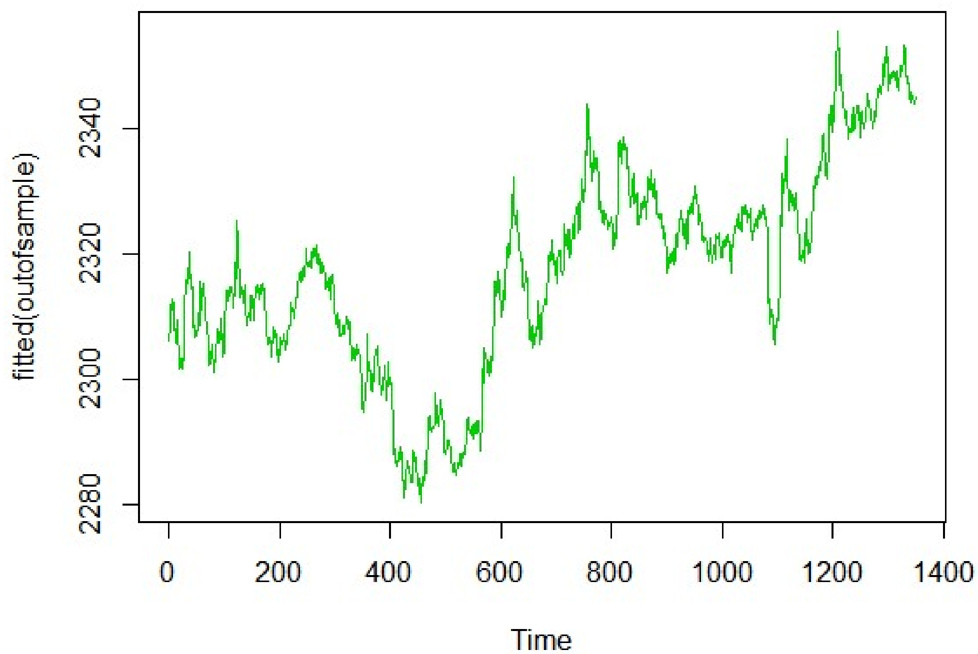

Figure 5.9 The out of sample forecast of 1-minute data

the green color (this figure can be better viewed in the electronical version of thesis) is used in Figure 5.9 to display the out of sample forecast to show the difference of real data and forecast data as they are very similar to each other.

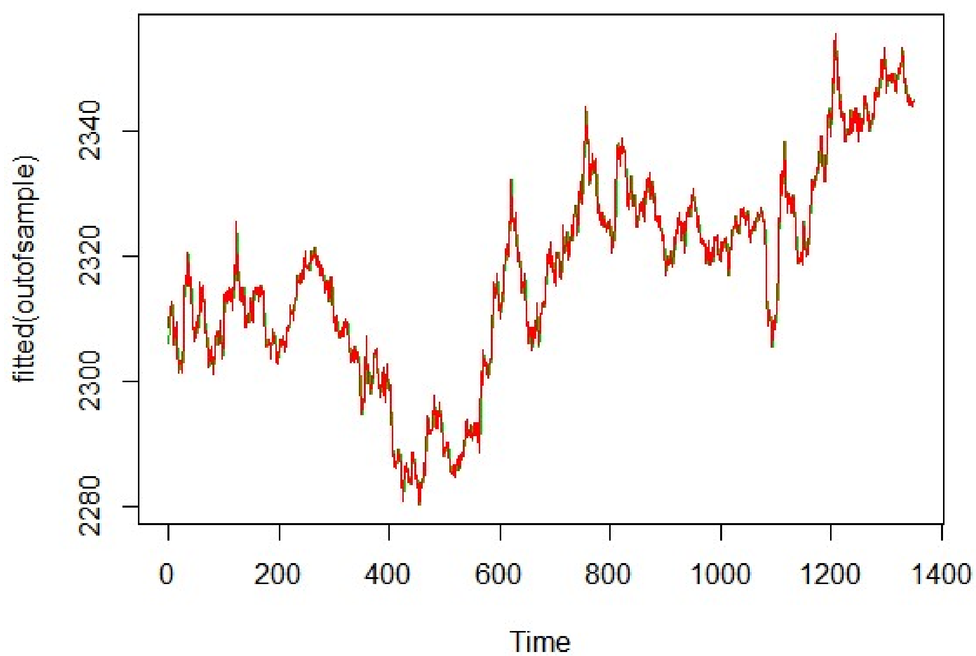

Figure 5.10 The out of sample data and forecast results

As shown in Figure 5.10, the red line shows the real out of sample data and the green line shows the forecast out of sample data. The accuracy of the model is shown in the following Figure 5.11.

```
> accuracy(fit)
                      ME     RMSE      MAE          MPE       MAPE       MASE
Training set -0.02844837 1.520653 1.071807 -0.001221243 0.04463527 0.02081607
> accuracy(outofsample)
                     ME     RMSE      MAE         MPE      MAPE       MASE
Training set 0.02909338 1.646306 1.213432 0.001225481 0.0523446 0.09242201
```

Figure 5.11 The accuracy of the forecast model

The accuracy of fit showed the in-sample accuracy results of the model, and the accuracy of out of sample showed the out of sample accuracy of the model.

To combine with the 3-minute data set, the 3 step out of sample forecast will be conducted in this experiment. The R code that produces the 3 step out of sample forecast is listed in the Appendix.

5.3.2 3-minute data ARIMA forecast

In this 3-minute data forecast, the first 1530 price points will be used to determine the ARIMA model, and the rest of the data will be used to conduct out of sample forecast.

```
> fit3<-auto.arima(determine3$close)
> fit3
Series: determine3$close
ARIMA(4,1,3) with drift

Coefficients:
         ar1      ar2     ar3      ar4      ma1     ma2      ma3    drift
      0.5667  -0.6011  0.8501  -0.0346  -0.5589  0.5713  -0.8534  -0.3716
s.e.  0.1225      NaN  0.0302   0.0295   0.1208     NaN      NaN   0.2060

sigma^2 estimated as 121.6:  log likelihood=-5835.63
AIC=11689.26   AICc=11689.38   BIC=11737.25
```

Figure 5.12 The 3-minute ARIMA model

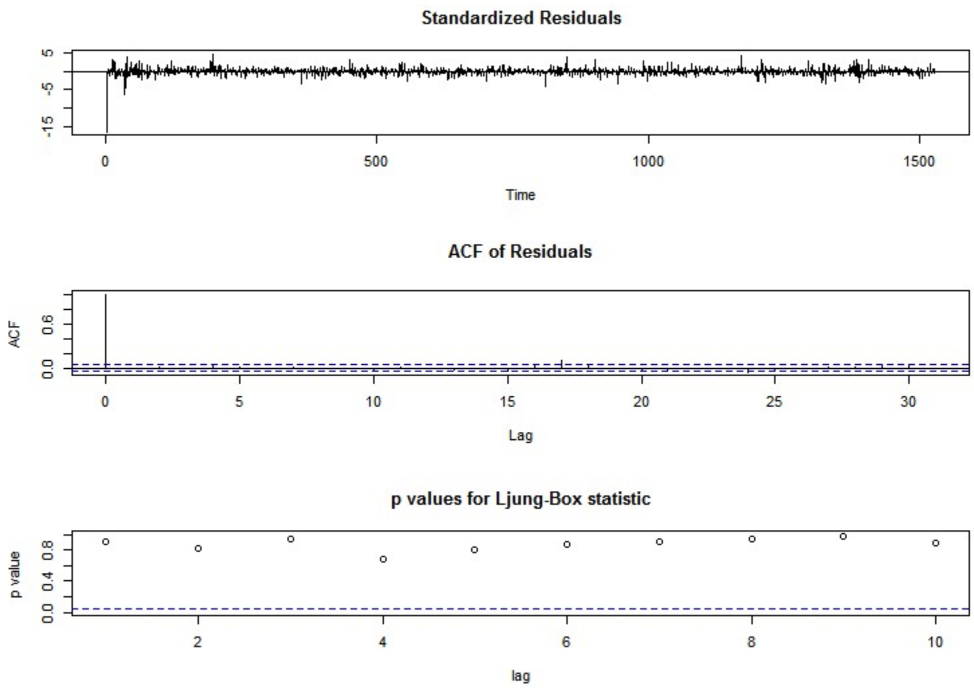

Figure 5.13 The Tsdiag of the determined Model

Figure 5.12 shows the 3-minute ARIMA model determined by the in sample data, and Figure 5.13 indicates that the model is validated. Similar to the 1-minute data, the out of sample predicting results will be illustrated in green color as shown in Figure 5.14, and the real out of sample data is shown in red color, the combination of the forecast and real data is shown in Figure 5.15.

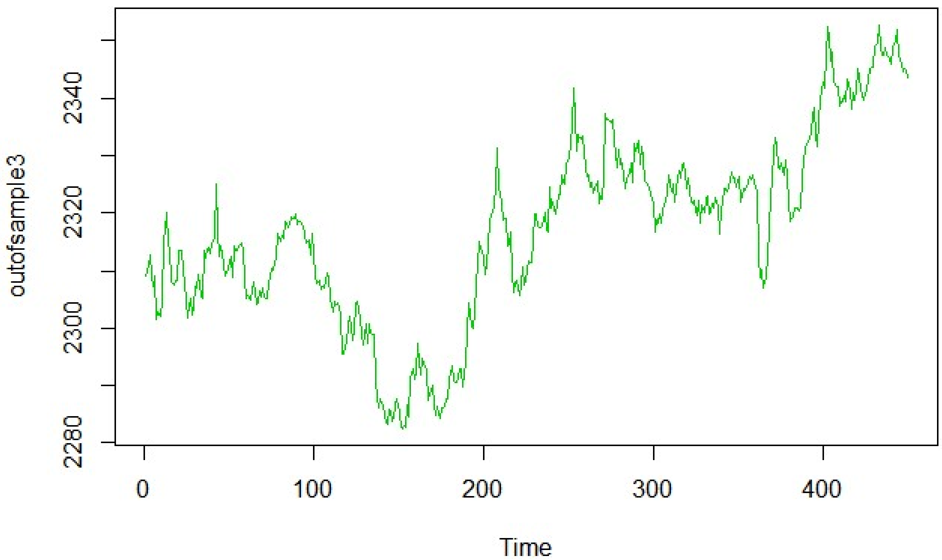

Figure 5.14 Out of sample forecasting of 3-minute data

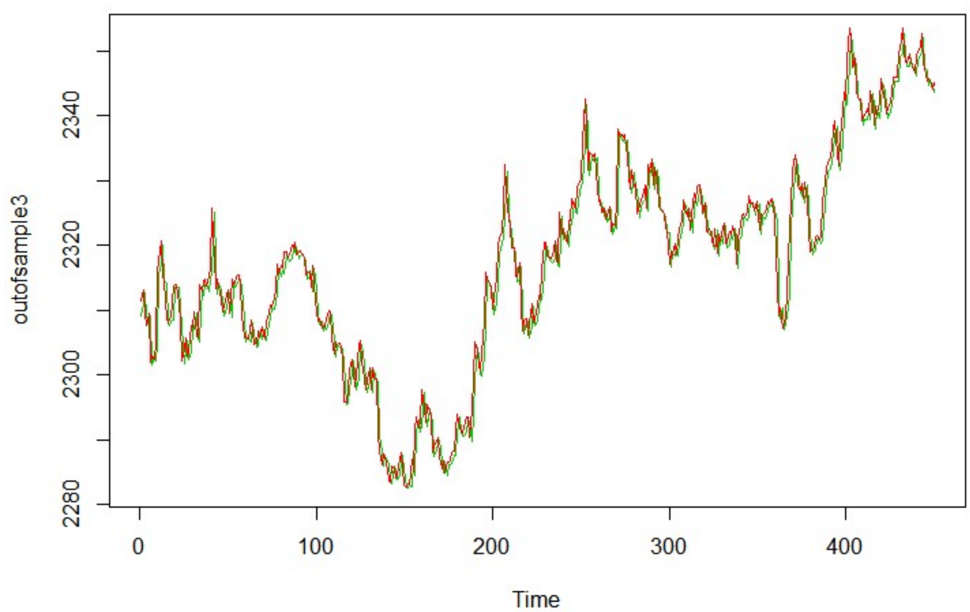

Figure 5.15 Combination of real data and forecast results

The accuracy of the in-sample and out-of-sample forecast is shown below.

```
> accuracy(fit3)
                      ME     RMSE      MAE        MPE     MAPE       MASE
Training set -0.03765271 11.02157 7.197861 -0.7315093 4.478907 0.04357899
> accuracy(forecast3)
                   ME     RMSE     MAE        MPE       MAPE      MASE
Training set 0.614322 2.885243 2.16515 0.02639345 0.09339285 0.1651514
```

Figure 5.16 The Accuracy of the in sample and out of sample forecast

As shown in Figure 5.16, the first line showed the in-sample forecast accuracy, and the second line showed the out of sample forecasting results. Those results will be used as benchmarks to test the efficiency of the proposed multi-frequency.

5.3.3 Combination of the forecast

In this part of the experiment, a combination of the 1-minute and 3-minute forecasting results will be made to create a multi-frequency forecast for the long period-3 min. For this special experiment, the data set will be divided into three sections, the first section will be used to determine the separate ARIMA model, and the second will be used to determine the combined model, and the last part will be used to test the efficiency of the combined model.

In this chapter, a simple linear combination of the two-frequency forecast will be made, to find out the best coefficients for each forecast, a numeration of ratio from 0.01 to 1 is calculated for the in-sample data to find out the lowest RMSE. The in-sample experiment results showed that the RMSE reduced as the weight of 3-minute 1 step ahead forecast is increased and the weight of 1-minute 3-step ahead forecast is decreased simultaneously. When the weight of the 1-minute 3-step ahead forecast is set to 0.01, and the weight of 3-minute 1-step ahead forecast to 0.99, the RMSE is reduced from 2.76 to 2.71.

The result of the above experiments proved the statements in the introduction section that useful information is contained in some specific frequencies of the original data, both multi-frequency analysis for forecasting high-frequency trading made sense as more information is taken into consideration compared to the single frequency analysis.

It is believed that better performance could be achieved with the application of wavelet and some other multi-frequency as the above experiment showed that different frequency of the same time series carries some special characteristics that can be used to improve the forecast performance.

5.4 The experiment on one simple trend following trading system

As discussed before, the most effective way to test the efficiency and accuracy of any financial forecasting model is to test the profitability of trading based on the proposed predict model. There are two ways to test the profitability. The first one is to calculate the possible gains and losses based on historical data, and the another one is to trade in real time market. The second way is clearly thought to be more reasonable as historical gains or losses are just hypothetical results.

In this study, the famous trading platform Multicharts 8.0 is used to conduct the real-time trading test based on the proposed models. Multicharts is one of the most popular trading software package for charting, back testing, and multi-broker automated trading. It is the best choice for the intra-day trading experiments.

To realize real-time trading, the instant data feed is needed. Touchance 3.0 provides us a perfect solution to this problem. With the combination of Multicharts and Touchance, real-time tick data can be received during trading hours of CSI, based on which, algorithm trading could be conducted.

A trend following trading system was created to test the execution performance of Multicharts and Touchance. The trading codes are shown in the appendix. In the following chapters of this thesis, more auto trading systems are proposed based on the multi-frequency directional forecast.

The summary of the performance of the basic trend following system is shown in the following figures.

Portfolio Performance Summary			
	All Trades	Long Trades	Short Trades
Net Profit	45120	20880	24240
Gross Profit	89400	51000	38400
Gross Loss	-44280	-30120	-14160
Account Size Required	8880	8700	8280
Return on Account	508.1081081	240	292.7536232
Return on Initial Capital	45.12	20.88	24.24
Profit Factor	2.01897019	1.693227092	2.711864407
Slippage Paid	0	0	0
Commission Paid	0	0	0
Open Net Profit	0	0	0
Max Portfolio Drawdown	-10500		
Max Portfolio Drawdown (%)	-7.960893855		
Max Portfolio Close To Close Drawdow	-8880		
Max Portfolio Close To Close Drawdow	-7.02		
Return on Max Portfolio Drawdown	4.297142857		
Portfolio Performance Ratios			
Portfolio Net Profit as % of Max Por	429.7142857		
Portfolio Time Analysis			
Trading Period	28 Dys, 5 Hrs, 9 Mins		
Time in the Market	4 Dys, 22 Hrs, 30 Mins		
Percent in the Market	17.4998154		
Max Run-up Date	3 Dys, 18 Hrs, 2 Mins		
Max Run-up Date	2014/6/27 15:07		
Max Portfolio Drawdown Date	2014/6/19 10:03		
Max Close To Close Drawdown Date	2014/6/19 9:59		

Figure 5.18 The portfolio performance summary

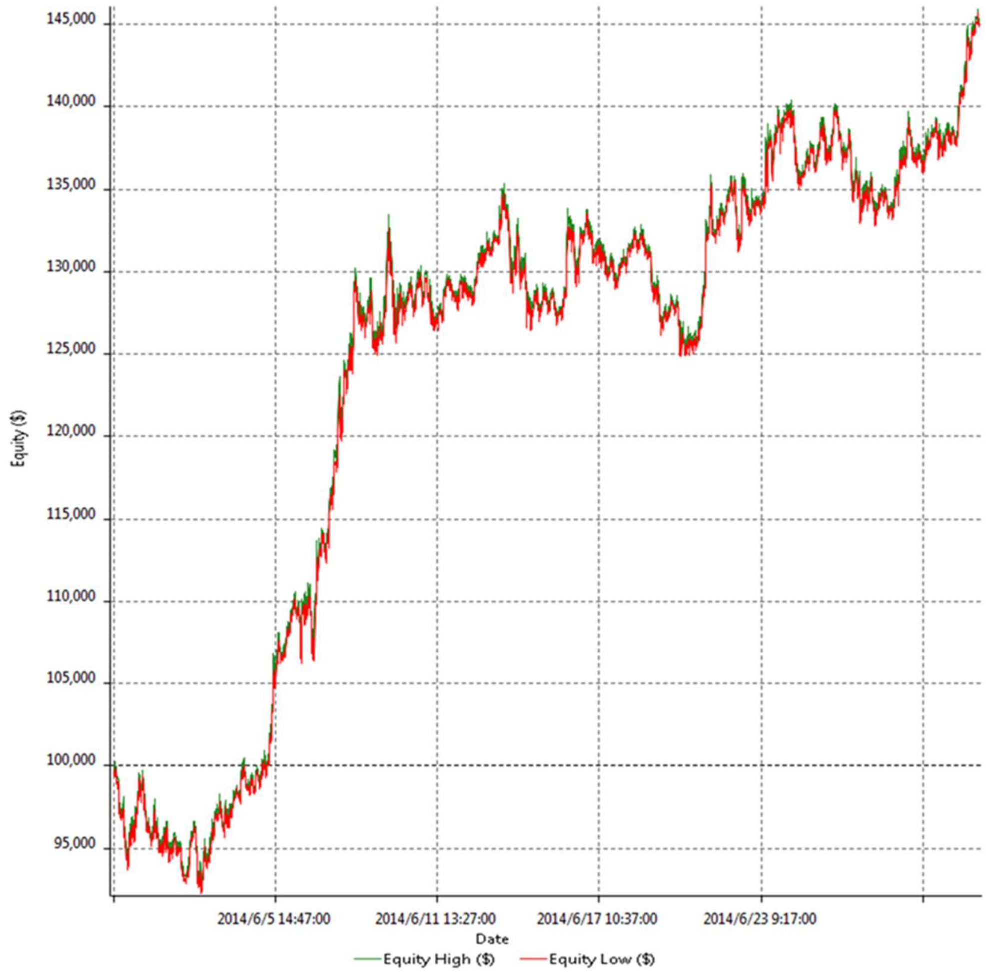

Figure 5.19 Equity Curve detailed

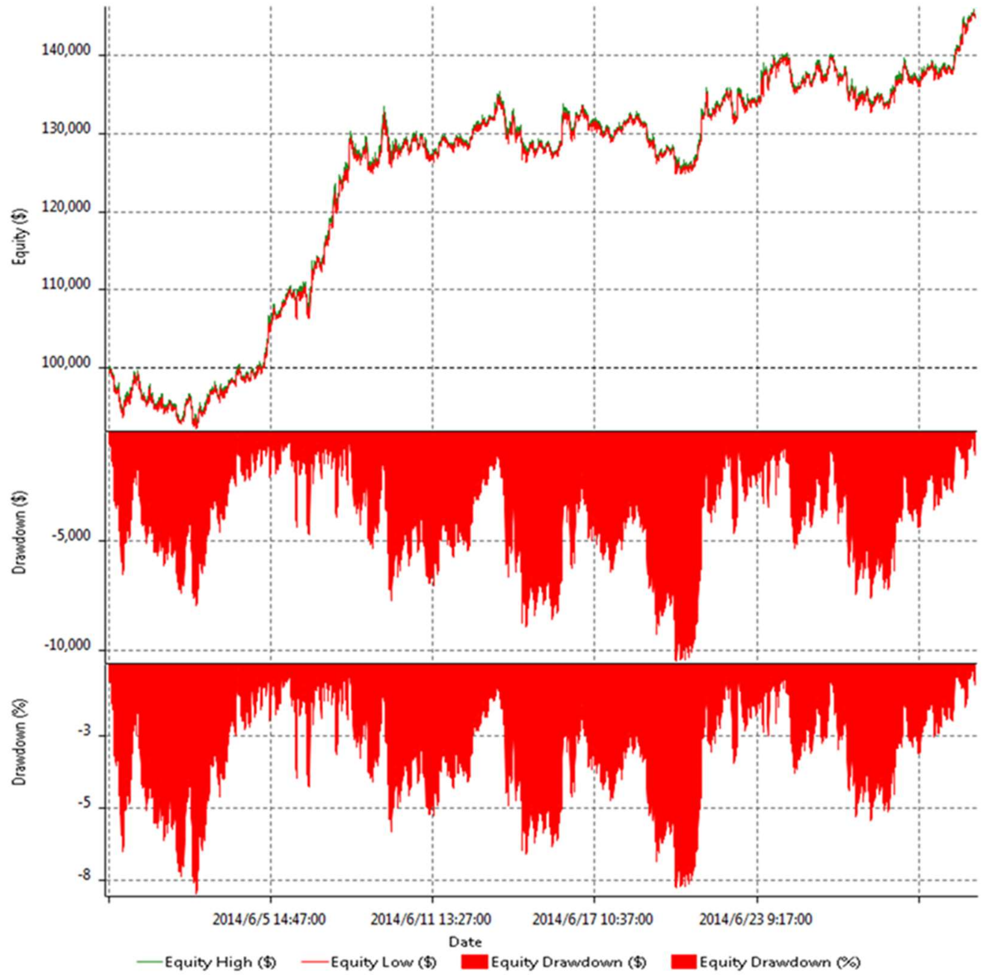

Figure 5.20 Equity Curve Detailed with DrawDown

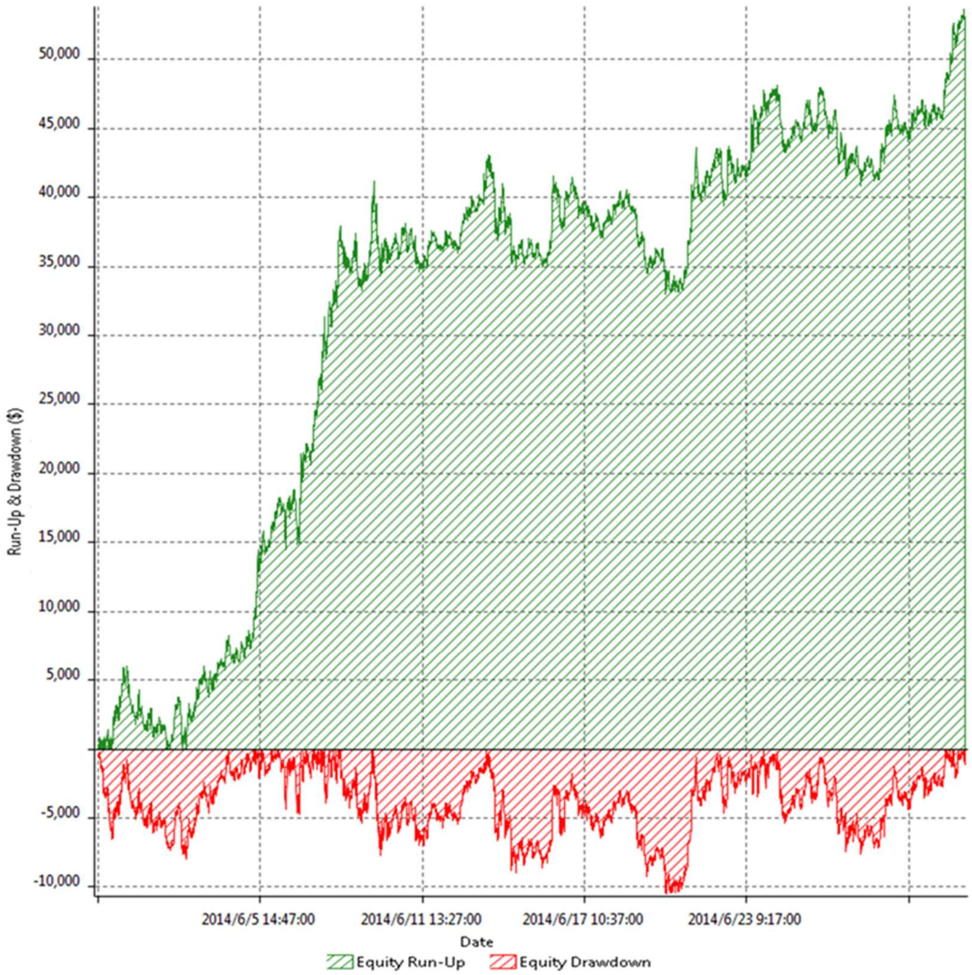

Figure 5.21 Equity Run-up & Drawdown

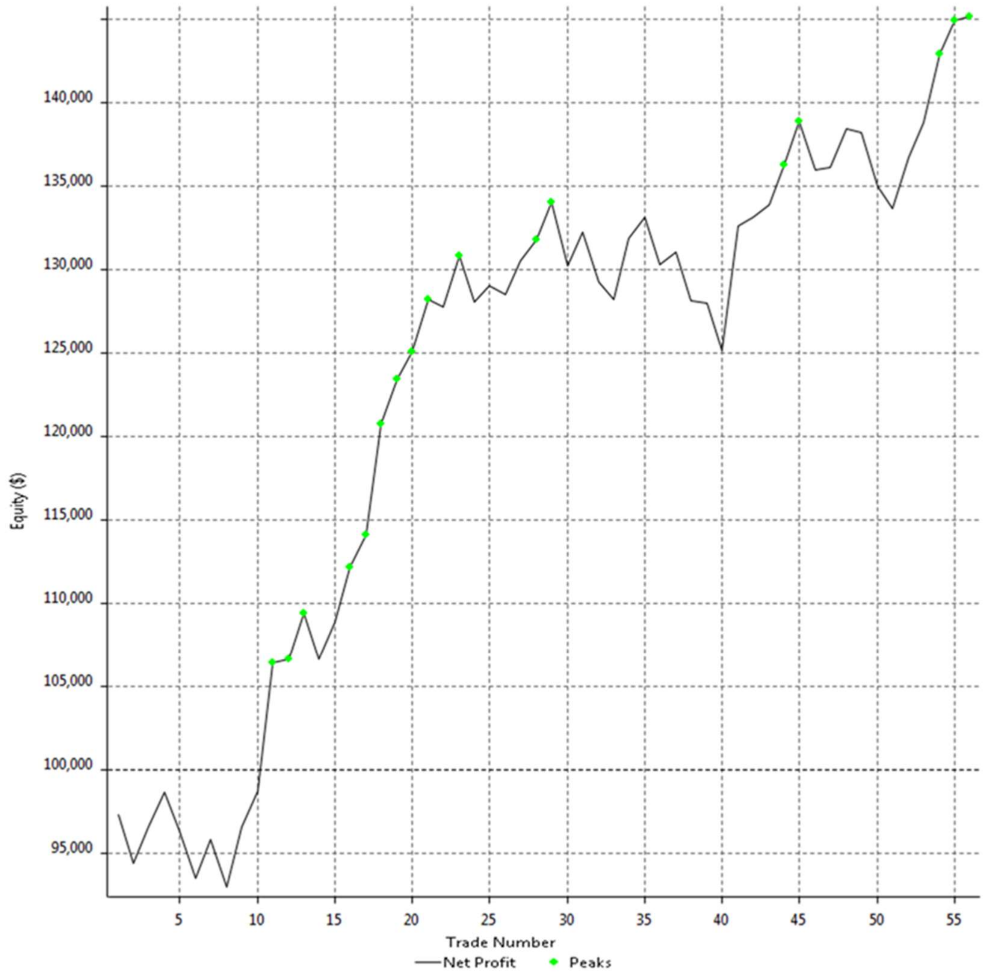

Figure 5.22 Equity Curve Close to Close

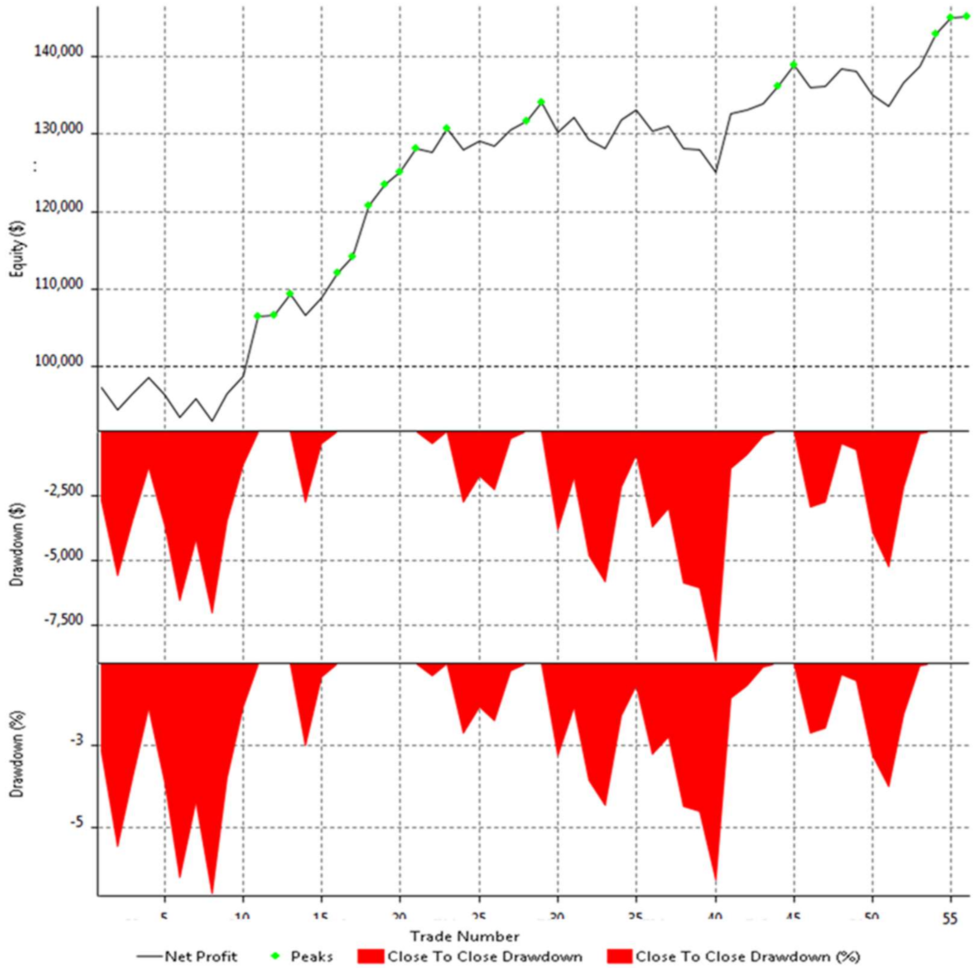

Figure 5.23 Equity Curve Close to Close with Drawdown

5.5 Summary

In this chapter, the multi-frequency ARMA is used to make the high frequency forecast of financial time series. A brief conclusion is drawn as follows.

1. The 1-minute and 3-minute data are linearly combined to form the final prediction of the series. As shown above, the forecasting results of the proposed model are proved to have smaller RMSE compared to the direct

ARMA forecast.

2. The meaning of the forecasting experiment in this chapter is the design of more efficient multi-frequency forecast models, and more importantly, the proof that the time series in different frequencies indeed carries some specific characteristics that could be utilized to improve the forecasting accuracy.

3. As the wavelet decomposition is the most powerful multi-frequency forecasting tool, it is believed that better forecasting performance could be achieved with the introduction of this tool. Based on the results that the combination of different frequencies of the same time series can improve the performance of the prediction, multi-frequency based trading systems could be built.

4. In the experiments of a simple trend following system, all the trades are held no more than one trading day. The results of the backtest showed promising profitability. However, this backtested system may suffer from overfitting problems. With the application of more advanced methods including the wavelet transformation and genetic algorithm, trading systems with better performance could be achieved with the consideration of overfitting problem.

In this chapter, a simple two frequency analysis on the financial time series is presented, the motivation for this study is to prove that different frequency of the time series conveys different characteristic of the series, that is the longer-term time series mainly carries the trend of the series and the shorter-term time series contains some local details of the time series. Some more sophisticated multi-frequency analysis based systems will be introduced in Chapter 6 and 7. Just like the trading systems in Chapter 4, the results of the models and trend following system will be mainly used as the benchmarks for the future study. However, the profitable trading system could be derived based on the simple trend following system if better trend detection indicator is discovered and utilized.

CHAPTER 6 – GP and WAVELET-DENOISE BASED GP

6.1 Introduction

Technical analysis related studies have been extensively conducted by researchers and investors for decades in various financial domains. Technical trading rules are generated from historical price and volume data, buying and selling signals are usually sent to exchanges over a very short period. To beat the market, many machine learning tools have been utilized in the design of technical trading systems. For linear analysis of the financial time series, one of the most commonly used tools is the genetic programming.

Genetic Programming (GP) is an automated method for creating a working computer program from a high-level problem statement of a problem. In this study, GP is used to find the optimized combinations of technical indicators that can generate profit from the market. Past researchers have applied GP to the design of trading systems. However, their models are all trained with a long-term period of data, such as daily, weekly or even monthly data. One of the largest drawbacks of these kinds of systems is that the incoming information effect is not considered. Information like financial reports and change of interest rates highly affect the price movement of the target analyzing asset; jumping points occurred after the announcement of such kind of information. GP models trained with long term data which includes many jumping points are unable to extract the real characteristics of the time series. To overcome this problem, in this study, all the models are trained using the intra-day data. Information that has dramatic impact on the market are expected to be digested during the market closed periods, therefore, jumping points barely exist in the intra-day data. True trends of the time series can be recognized with GP models trained using the intra-day data.

Another innovation point of this study is the noise reduction of the intra-day data. High frequency intra-day financial time series are highly correlated and un-stationary. Wavelet de-noise is introduced in this study to handle the noises. Wavelet transform leads to a sparse representation for many real-world signals, many features in different scales of the data can be localized. Wavelet transform concentrates signal in a few large-magnitude wavelet coefficients, wavelet coefficients which are small in value are typically noise and can be removed while preserving important signal features. Both the soft-thresholding and the hard-thresholding methods are used in this study to de-noise data, optimized GP models based on the original data and the proposed wavelet de-noised data are compared with each other using the out of sample performance.

In this chapter, the CSI 300 index futures is selected as the target training and testing asset. The index futures market of China is one of the most actively traded emerging markets. Past studies on index futures usually use the frequency of a day, a week or even a month, some of which indeed showed outstanding performance in the out of sample experiments. However, these kinds of theoretical profits are highly unrealistic in the real markets, because most of the trades in the futures markets are margin trading and small movements of the stock index might lead to a large amount of gains or losses to investors. For most of the time, investors do not have enough money to wait for the market to go to the right price that their model is expected. The proposed model in this study is trained using the intra-day 1-minute interval data, for each trading day, the data is separated into two different part, the first parts is used to train the GP model and the second part is used to perform the out of sample performance. In this way, no open position is held overnight, which could greatly reduce the risk caused by overnight unfavorable price jumps.

6.2 Related Methodology

6.2.1 Wavelet and Wavelet de-noise

To overcome the shortcomings of the Fourier transform in the performing of multi-resolution analysis of signals, the wavelet has been widely applied to different fields. Unlike the Fourier analysis, wavelets are localized in both time and frequency scale. Mathematically, a wavelet is defined as a function of a zero average:

$$\int_{-\propto}^{\propto} \psi(t)dt = 0 \tag{23}$$

The concept of Wavelet Transform (WT) is to represent the target function as a superposition of a set of wavelets which are small waves located at different times. The original series can be transformed and described in the content of different frequency over time at certain time-scale. Wavelet-based models are suitable for non-stationary data as some transient trends, or hidden characteristics of the original series can be highlighted. Financial time series are well-known to be non-stationary, which makes WT one of the most powerful tools for the analysis and forecasting of financial time series.

6.2.1.1 The Discrete Wavelet Transform

For the most practical studies, wavelet coefficients are not needed to be calculated at every possible scale as it would be a fair amount of work to do that, which will generate an awful lot of data. The DWT projects a time series onto a collection of orthonormal transformations. Unlike the continuous wavelet transformation that considers all possible frequencies at continuous times, the DWT focuses on some specific frequencies at distinct times. The discrete wavelet is defined as:

$$\psi_{j,k}(t) = S_0^{-\frac{1}{j}} \psi(\frac{t - k\tau_0 S_0^j}{S_0^j}) \qquad (26)$$

j and k are integers. $s_0 > 1$ is a fixed dilation step and the translation factor τ_0 depends on the dilation step. If $s_0 = 2$, the discrete wavelet transformation is defined as dyadic wavelet transform. By definition, the scaling function and the wavelet function of discrete wavelet transformation are:

$$\varphi(2^j t) = \int_{i=1}^{k} a_k \, \varphi(2^{j+1} t - k) \qquad (27)$$

Equation [eq45] is the scaling function, known as the scaled father wavelet, a_k are the corresponding coefficients, $k \epsilon Z$. The mother wavelet ψ is obtained by the linear combinations of the scaled father wavelet. Proper coefficients must be selected to maintain the orthogonality of the basis wavelets.

$$\psi(2^j t) = \int_{i=1}^{k} b_k \, \varphi(2^{j+1} t - k) \tag{28}$$

The mother wavelet ψ are uniquely determined by their coefficients $\{b_k\}$, $k \in Z$. Then, a signal $f(t)$ can be written as:

$$f(t) = \int_{i=1}^{k} a_{j,k} \, \varphi(2^{j+1} t - k) + \int_{i=1}^{k} b_{j,k} \, \psi(2^{j+1} t - k) \tag{29}$$

This is the basic idea of the discrete wavelet decomposition of the signal.

6.2.1.2 Wavelet de-noise

Separating noises from the original signal is the procedure of de-noise. Suppose $f(t)$ is the original time series, some parts of the series $s(t)$ represents the trending of the original series, while the others are just localized noises. Then the model of the de-noise process is:

$$s(t) = f(t) - e \tag{48}$$

e represents the noise series. If this process is combined with the discrete decomposition, the process of de-noising the original time series becomes the adjust coefficients of different scales. The most common method is the threshold rule.

The basic idea of this threshold rule is to truncate the insignificant coefficients since less amount of information is contained or only localized noise is contained. One of the key steps is how to select threshold parameters. If the threshold value is less than the noise level, unexpected noises are present in the processed series. However, if the threshold value is significantly bigger than the noise level, the information contained in the original series would lose. A universal threshold value is proposed as follows:

$$threshold = \sigma\sqrt{2log(N)} \tag{49}$$

σ is the standard deviation and N refers to the total number of the data.

There are two kinds of threshold methods, namely the hard threshold and the soft threshold. Both of these methods are conducted in this research to evaluate the performance with each other. The hard threshold is a straight forward technique to implement the wavelet de-noising process. The advantage of the hard threshold is easy to use and the better reconstruction of discontinuities. The coefficients adjustment is given as,

$$H(\beta, d) = \begin{cases} \beta, if|d| \leq \beta \\ 0, otherwise \end{cases} \tag{50}$$

d represents the wavelet coefficients and β is the threshold value. This method is not a continuous mapping, and the input coefficients is effective if it is less than or equal to the threshold.

The soft threshold is another way to adjust the wavelet coefficients. Instead of forcing wavelet coefficients to zero or leaving it untouched, the soft threshold pushes all coefficients towards zero, which can be defined as,

$$S(\beta, d) = \begin{cases} \beta, if|d| < \beta \\ Sign(d)(|d| - \beta), otherwise \end{cases} \quad (51)$$

$$Sign(d) = \begin{cases} 1, if d > 0 \\ 0, if d = 0 \\ -1, if d < 0 \end{cases}$$

Compared with the hard threshold, the different setting happens when the absolute value of wavelet coefficients is greater than the threshold. As all the coefficients are affected in this process, the soft threshold is a continuous mapping. As all wavelet coefficients are suppressed by the soft threshold, a smoother reconstructed time series is expected.

6.2.2 Generating trading rules with genetic programming

In this subsection, the process to conduct the adaptation of genetic programming is presented. The genetic programming will be used to generate trading rules based on technical indicators which are encoded as programs.

6.2.2.1 Encoding of the technical indicators

The initial programs, which are represented as tree structures, are recursively constructed from a predefined set of functions F and terminals T. Both functions F and terminal T change dynamically during the evolution process. Functions F and Terminal T are defined in the following.

Functions:

Arithmetic operators: $+,-,*,/$;

Boolean operators: and, or, not;

Relations operators: $<,>$;

Boolean functions: if-then-else.

Real functions (user defined functions, here are the technical indicators)

Technical indicators, variables represents for constant Price:

$Norm(r1, r2)$: absolute value of the difference between real number;

$Avg(s, n)$: average of price over the past n periods;

$EMA(s, n)$: exponential moving average of the past n periods;

$Max(s, n)$: maximum value of price over the past n periods;

$Min(s, n)$: minimum value of price over the past n periods;

$Lag(s, n)$: price value lagged by n periods;

$Valotility(n)$: variance in returns over the past n periods;

$RSI(n)$: relative strength index;

$ROC(n)$: rate of change.

$$ROC(n) = (\frac{closing\ price\ of\ the\ current\ minute}{closing\ price\ of\ n\ minutes\ ago} - 1) * 100 \tag{52}$$

$RSI(n) = 100 - (\frac{100}{1+RS(n)})$: $RS(n) = \frac{\int_{i \in D^{+}(n)} r_i}{-\int_{i \in D^{-}(n)} r_i}$, D^{+} is the set of minutes with rising prices, D^{-} is the set of minutes with falling prices and r_i is the return of minute i, which is positive when the price rising, and negative otherwise.

Terminals:

Constants: chosen in the interval $[0,270]$, where 270 is the approximate number of minutes in a single trading day;

Boolean: True, False;

Others: Price

Real variable: price of the current minute;

Order Types: Entry order and Exit order

Entry order: Market Entry order: enter into the market at market price;

Stop Entry order: these orders are placed above the market for a long entry and below the market for a short entry;

Limit Entry order: these orders are placed below the market for a long entry and above the market for a short entry;

Exit order: Exit at target profit, Exit at target percentage profit, Exit at Target price, Protective Stop, Trailing Stop, Exit after N Bars, Exit after N Bars profit, Exit after N Bars loss, Exit after certain time, Exit at Market, Exit End-of-Day.

It is obvious that functions and terminals sets chosen in this research violate the closure assumption of genetic programming. Some functions set cannot be an argument to another function in F, given that both Boolean and real functions are found in that set. The restrictions mean that the real functions are always in the lower part of the tree structures. It makes sense that all the trading decisions are made based on technical indicators while Boolean functions and relational operators are found on the upper part of the tree.

6.2.2.2 Fitness evaluation criteria

In most of the relevant studies, fitness values are calculated using the excess return over the markets. However, the net profit is a much simpler and efficient way to evaluate the profitability of trading strategies. In the trading period, the total net profit is the total profit (gross profit minus gross loss) for all closed trades, both wins and losses.

6.2.2.3 Initialization of the base population

There are two methods to generate the initial population of the trading blocks. The full method creates tree structures, in which the length of every path between a terminal and the root equals to a predefined depth. The growing method creates the tree structures with variable shapes. In this section, a combination of the full and grow methods is applied. Due to the particular structure of function set, the recursive construction has to follow the rules bellow. Firstly, the root of the tree should be selected from the Boolean functions and operators. It illustrates that the buy and sell signal must be generated at the top of the tree structure. Secondly, descendants of the tree can be selected among Boolean constants, Boolean functions, Boolean or Boolean operators. Thirdly, a relational operators' descendants should be selected among real functions or terminals.

After the generation of the initial population, more complex trading decisions are built using the initial trade blocks with some mutations and crossover.

6.2.2.4 Crossover, mutation, reproduction and selection

As the initial population usually has low fitness, genetic operators must be utilized to search better programs.

The crossover operator is used to combine the genetic material of two parents by swapping a part of one parent with a part of the other. In this section, the offspring resulted by crossover replaces the parent with less fitness. The mutation operator is used to operate on only one individual. Whenever one of the individuals in a particular population has been selected for mutation, a tree node is selected randomly, and the existing subtree is replaced with a randomly created sub-tree. The mutated individual is then put back to the original population to calculate the fitness. The reproduction operator is simply a clone of the current tree structure, which usually happens to a tree with rather high fitness value. This operator is controlled by the parameter reproduction probability.

In the selection of programs in the population, a ranking based approach proposed by Baker (1985) and Whitley (1981) is used. To use the ranking system, the fitness of all programs must be calculated and sorted from the best (rank 1) to the worst (rank n). A new value f_i is then associated with the program or rank i as follows:

$$f_i = Max - [(Max - min)\frac{i-1}{N-1}] \qquad (53)$$

In this way, the best program gets fitness Max, and the worst program gets fitness Min. Most importantly, the remaining programs are equally spaced between Min and Max. The super program problem that crowded out all the other programs can be avoided.

6.3 Design of the experiment in this study

The designed genetic programming method are applied to the original financial time series and the wavelet de-noised time series. The detailed process is shown in Figure 6.1.

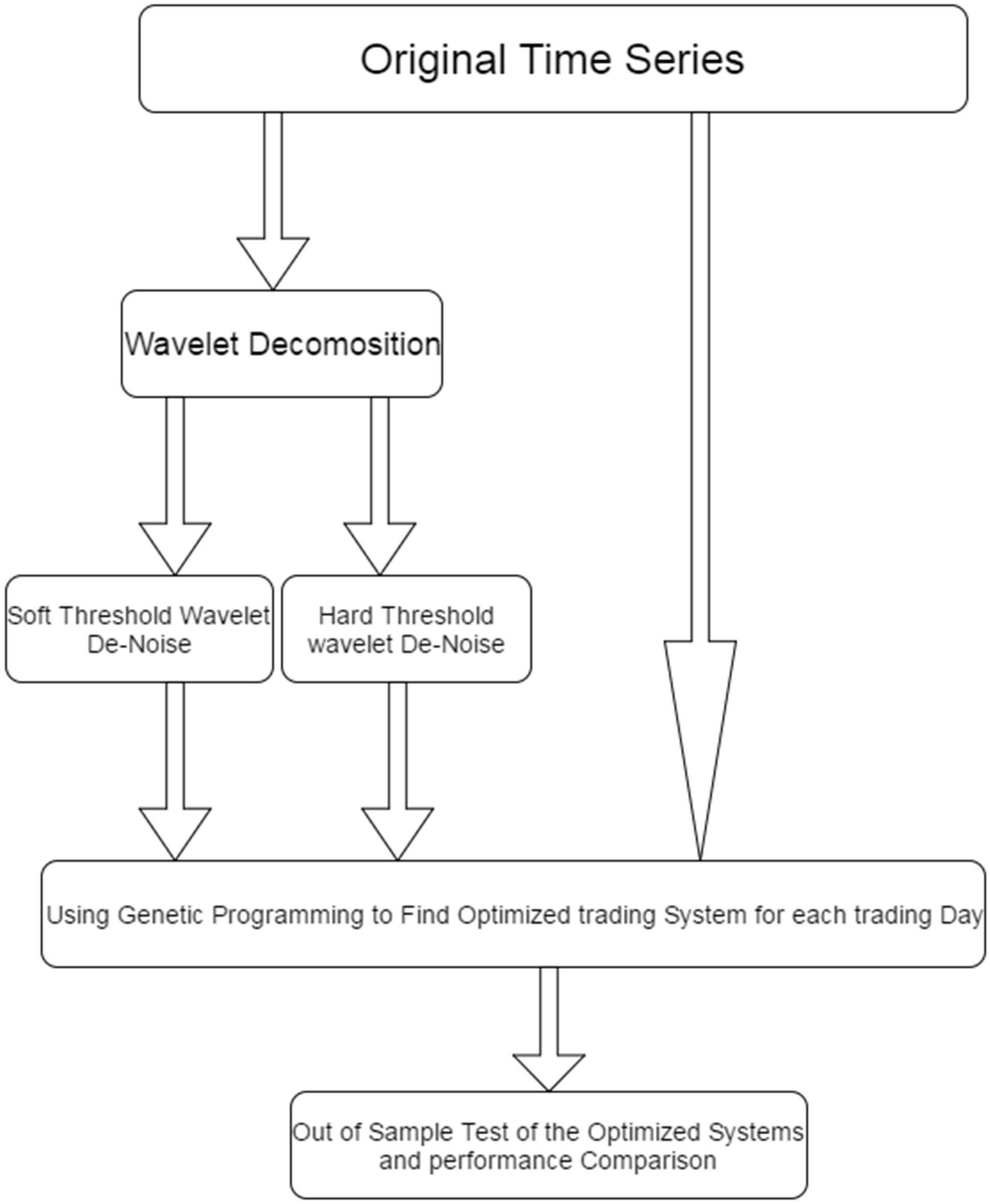

Figure 6.1 Flow chart of the experiment

Three experiments are conducted simultaneously. In the first experiment, technical indicators are calculated using the original 1-minute frequency data of the CSI 300 index future. After that, genetic programming are used to search for the best possible combinations of all the indicators and order types listed in the methodology.

Once the optimized trading strategies are determined for each trading day, out of sample test of the models is conducted. One of the key points in the design of these experiments is that all the technical indicators are calculated using the data from a single trading day and no historical trading information of the former day or before is used in the calculation of the indicators. It is a good try to reduce the effect of a price jump between two trading days. In the second experiment, the original data are firstly de-noised with two different methods, namely the soft threshold and hard threshold wavelet de-noise method. After that, the de-noised data are used to calculate the related indicators. The rest of the second experiment is similar to the first experiment. Furthermore, to explore the effect of different trading day jumps, another experiment is conducted. In the last experiment, multiple-day data are used instead of the single day data in the training of the models. The overall trading results of the third experiment are compared with the first and second experiment.

6.4 Empirical experiment and results

In this empirical experiment, the CSI 300 Index futures is selected as the target asset. Detailed information about the index futures is shown in the following Table 6.1. Data of 21 days of the CSI 300 index futures is used, each of which contains 270 data points. The original data points cover the period from 29/09/2014 to 03/11/2014. In each trading day, the first 128 data points are used to train the model, and the last 142 data points are used to test the out of sample performance of the model. The selection of 128 training data points is by the decomposition requirement of the wavelet analysis. As the morning session of a trading day consists of 130 minutes, the training processes were conducted in the morning session, and the out of sample testing process was conducted in the afternoon session.

Table 6.1 The CSI 300 Index

Underlying Index	CSI 300 Index
Contract Multiplier	CNY 300
Unit	Index point
Tick Size	0.2 point
Contract Months	Monthly: current month, next month, next two calendar quarters (four total)
Trading Hours	09:30 am - 11:30 am, 01:00 pm - 03:00 pm
Limit Up/Down	+/-10% of settlement price on the previous trading day
Margin Requirement	8% of the contract value
Last Trading Day	Third Friday of the contract month, postponed to the next business day if it falls on a public holiday
Delivery Day	Third Friday, same as "Last Trading Day"
Settlement Method	Cash Settlement
Transaction Code	IF
Exchange	China Financial Futures Exchange

To overcome the over fitting problem, a higher weight is assigned on the out of sample performance of the developed model. Only the out of sample performances are discussed as the in sample trading profit is highly unrealistic. As shown in the design of the experiment, the out of sample performance of the model optimized by the 128 in sample original data points will be compared with the model optimized by the wavelet de-noised 128 in sample data points. As described in the methodology section, there are two different de-noise methods, the hard threshold, and soft threshold. Both methods are applied in this chapter to compare with each other. Another comparative study about the trading with and without holding assets overnight is also tested. In the experiment of holding the assets overnight, the data of the first 14 days are used to train the model, and the data of the last seven days are used to test the out of sample profitability. Also, just like the first experiment, wavelets are utilized to de-noise the in-sample data.

Excessive preliminary experiments were conducted to optimize the parameters for the genetic training process. The final parameters are illustrated in Table 6.2.

Table 6.2 Parameters for genetic programming

Population size	300
Number of generations	30
Crossover percentage	60%
Mutation percentage	50%
Tree depth	5
Tournament size	5
Limit of entries per day	8
Wait for exit before entering new trade	Yes
Max bars looking back for Indicators	30

The detailed trading results of 21 days without holding assets overnight are shown in Table 6.3.

Table 6.3 Trading results for 21 days of intra-day data

Date	Out of sample P/L[1]	Out of sample Cumulative P/L[2]	Out of sample P/L[3]	Out of sample Cumulative P/L[4]	Out of sample P/L[5]	Out of sample Cumulative P/L[6]
2014/9/29	-960.00	-960.00	300.00	300.00	360.00	360.00
2014/9/30	-1140.00	-2100.00	2940.00	3240.00	900.00	1260.00
2014/10/8	2100.00	0.00	420.00	3660.00	1380.00	2640.00
2014/10/9	480.00	480.00	-240.00	3420.00	1380.00	4020.00
2014/10/10	6900.00	7380.00	-600.00	2820.00	-780.00	3240.00
2014/10/13	-840.00	6540.00	540.00	3360.00	-720.00	2520.00
2014/10/14	780.00	7320.00	1260.00	4620.00	-60.00	2460.00
2014/10/15	2100.00	9420.00	2160.00	6780.00	-3780.00	-1320.00
2014/10/16	-2640.00	6780.00	-3600.00	3180.00	-5160.00	-6480.00
2014/10/17	2640.00	9420.00	-2460.00	720.00	1560.00	-4920.00
2014/10/20	960.00	10380.00	-540.00	180.00	-2640.00	-7560.00
2014/10/21	0.00	10380.00	4560.00	4740.00	-2340.00	-9900.00
2014/10/22	-5640.00	4740.00	-5880.00	-1140.00	-5760.00	-15660.00
2014/10/23	540.00	5280.00	840.00	-300.00	2100.00	-13560.00
2014/10/24	0.00	5280.00	1320.00	1020.00	2160.00	-11400.00
2014/10/27	-540.00	4740.00	420.00	1440.00	1980.00	-9420.00
2014/10/28	-120.00	4620.00	-360.00	1080.00	-360.00	-9780.00
2014/10/29	-1740.00	2880.00	600.00	1680.00	2580.00	-7200.00
2014/10/30	-480.00	2400.00	300.00	1980.00	-2940.00	-10140.00
2014/10/31	3120.00	5520.00	13260.00	15240.00	9360.00	-780.00
2014/11/3	-3060.00	2460.00	-3480.00	11760.00	4500.00	3720.00
Min	-5640	-2100	-5880	-1140	-5740	-15660
Max	6900	10380	13260	15240	9360	4020
Std	2484	NULL	3619	NULL	3313	NULL

[1] Out of sample Profit/Loss of model trained with original data

[2] Cumulative Profit/Loss of out of sample performance of model trained with original data

[3] Out of sample Profit/Loss of model trained with hard threshold de-noised data

[4] Cumulative Profit/Loss of out of sample performance of model trained with hard threshold de-noised data

[5] Out of sample Profit/Loss of model trained with soft threshold de-noised data

[6] Cumulative Profit/Loss of out of sample performance of model trained with soft threshold de-noised data

As shown in Table 6.3, the out of sample performance of the training models varies during different trading days. The positive value represents gains, and the negative value represents losses. The sum of the total trading profits and losses for three different training methods are 2,460 for the GP strategy trained on the original data, 11,760 for the GP model trained on the hard threshold wavelet de-noised data and 3,720 for the GP strategy trained on the soft threshold wavelet de-noised data.

The intra-day training and testing results are shown in Figures 6.3, 6.4 and 6.5.

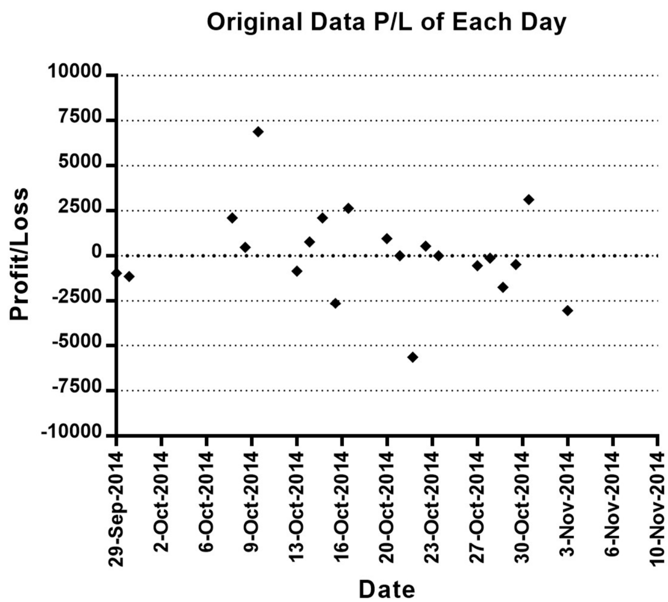

Figure 6.3 Original Data P/L of each trading day

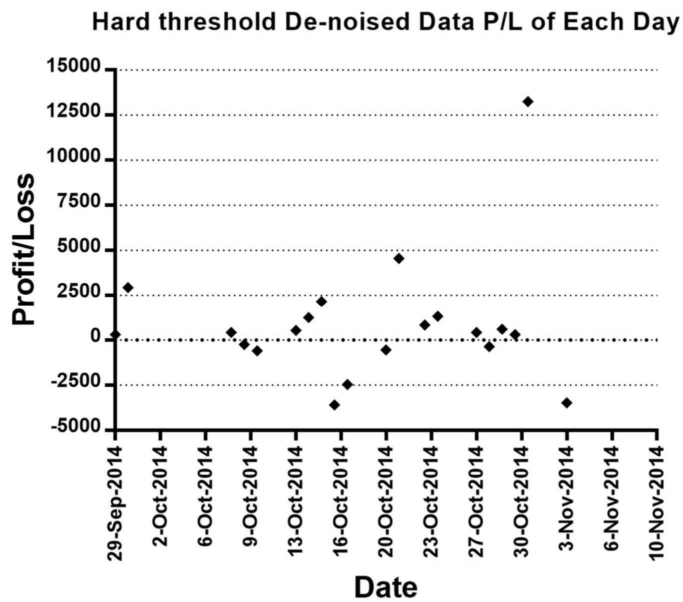

Figure 6.4 Hard threshold De-noised Data P/L of each day

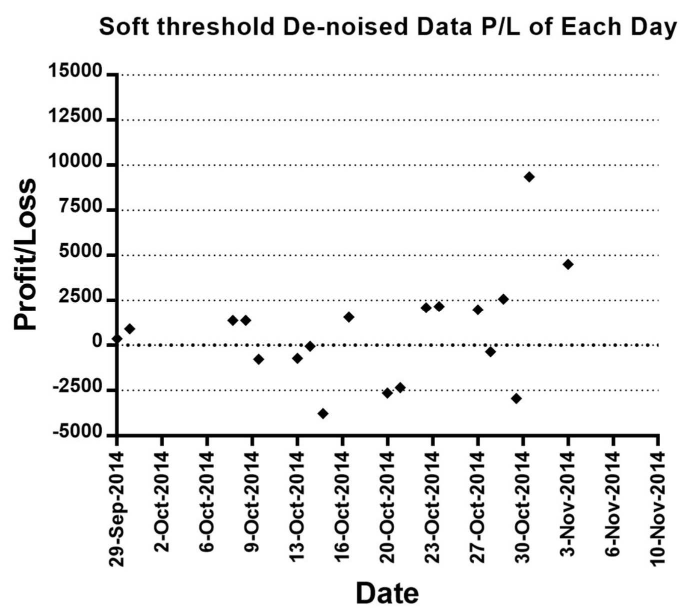

Figure 6.5 Soft threshold De-noised Data P/L of each day

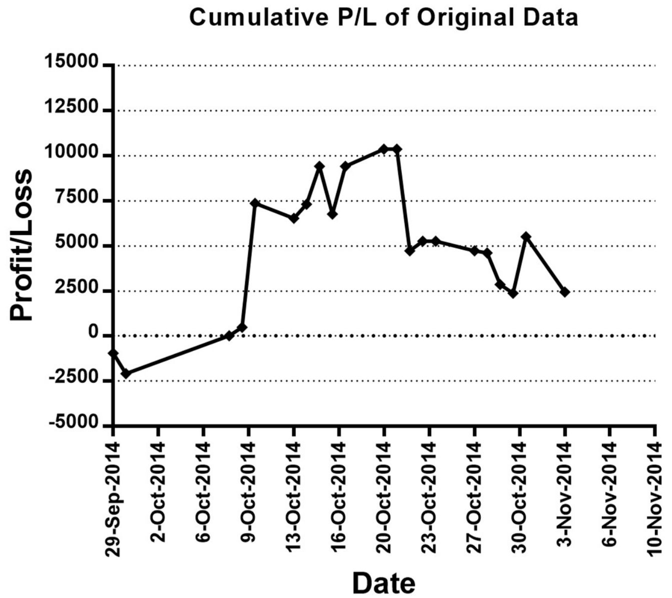

Figure 6.6 Cumulative P/L of Original Data

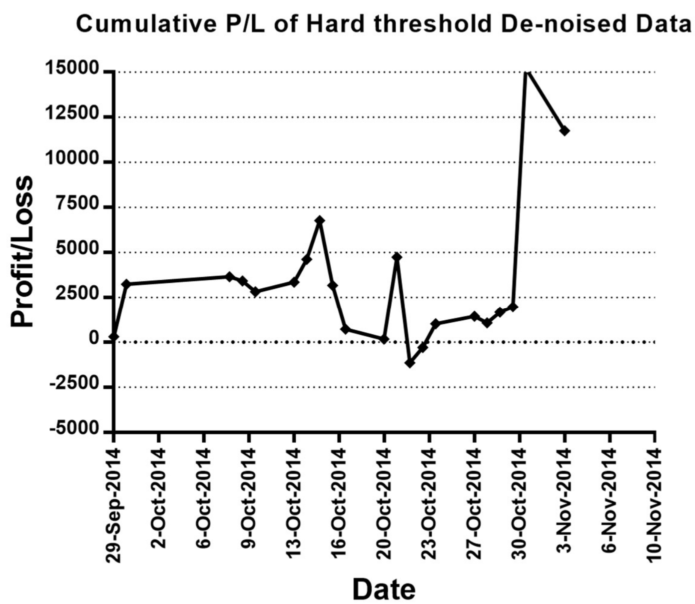

Figure 6.7 Cumulative P/L of Hard threshold De-noised Data

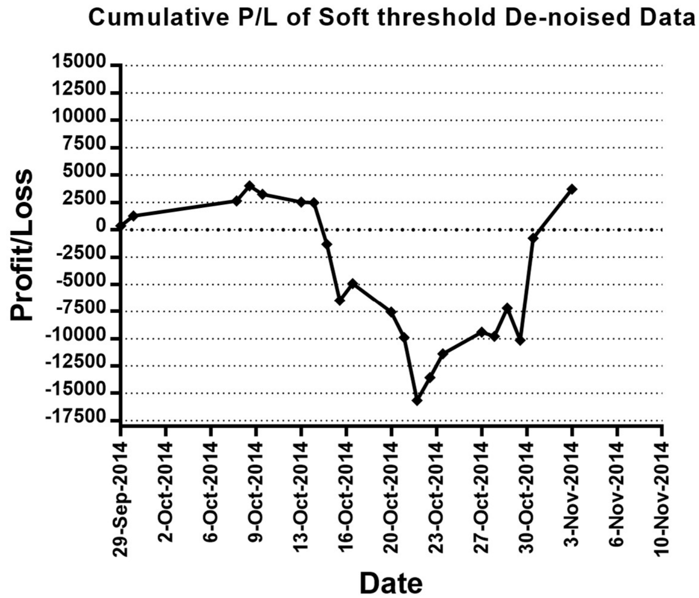

Figure 6.8 Cumulative P/L of Soft threshold De-noised Data

The results in Table 6.3 indicated that better out-of-sample performance are obtained if the GP model is trained with wavelet de-noised data sets and the hard threshold de-noise method has the best out-of-sample performance. At the bottom of Table 6.3, the min, max, and std of each profit and loss are listed, the results indicate that the max intra-day loss for the three experiments does not vary very much. The hard threshold has the highest possible intra-day loss of 5,880. However, it also has the highest possible intra-day profit of 13,260. The soft threshold has a higher possible loss and lowers possible profit than the performance from the original data. With a much higher standard deviation of the profit/loss, the out-of-sample results shows an interesting phenomenon that the wavelet de-noise process tends to magnify the gains and losses for most of the trading days. From Figures 6.6, 6.7 and 6.8, it is clear that the hard threshold de-noised data has the lowest max drawdown and the highest possible profit, which made it the best trading strategy in the experiments. Although the soft threshold de-noised data trained strategy has higher total profit than the original data trained strategy, the max drawdown is much higher than the other two strategies.

As the total trading profit of the wavelet de-noise data trained models is higher than the model trained from the original data, it can be concluded that with the consideration of the non-stationary and noisy characteristics of intra-day high-frequency time series, and the introduction of wavelet de-noise, better trading performance systems can be developed.

The trading results in Table 6.3 indicates that when multiple-day data are utilized to train the model, all the out of sample trading is subjects to severe losses. According to the severe losses in Table 6.4, it can be concluded that trading trend varies in different trading days and profitable trading strategies in a certain trading day might suffer from severe losses in other days. This results proved that the jumping points in the price of the index futures will influence the performance of GP trained models, and the models trained using long time intervals cannot extract the genuine trend of the series, as the jumping points which are caused by exogenous information are more likely to be recognized as significant features.

Table 6.4 Trading results for multi-days training methods

Date	Out of sample Profit/Loss of model trained with original data	Out of sample Profit/Loss of model trained with hard threshold de-noised data	Out of sample Profit/Loss of model trained with soft threshold de-noised data
2014-10-24 - 2014-11-03	-37320	-34140	-22260

6.5 Summary

In this chapter, to de-noise the original financial time series and to search profitable trading rules, an integration of the genetic programming and wavelet de-noise is utilized to generate intra-day programmed trading rules. The following observations can be made.

1. Relevant information that affects the movement of the time series is assumed to be fully digested during the market closed periods to avoid the jumping points of the daily or monthly data. Intra-day high-frequency time series is capable of exploiting more performance of the short-term forecasting advantage of technical analysis.

2. The results showed that the trading rules generated by genetic programming with the wavelet de-noised data points have better out of sample performance and that the hard threshold de-noise method outperforms the soft threshold methods.

3. The trading strategies trained with morning session data are most likely to

be profitable in the afternoon session of the same trading day. Also, the trading GP Strategies trained with multiple-day data may lead to severe losses in the following several trading days, which indicates that financial time series' trend tends to vary in different trading days and stays the same in a single trading day.

4. As trading strategies are trained separately using different single day's data, which makes the experiment more complicated, only 21-days of 1-minute data is tested in the experiment, more tests need to be conducted to evaluate the robustness of the GP trained strategies.

5. Although the trading cost is considered in this research, much more realistic problems, such as topological execution of orders and the execution delay of orders, should be taken into consideration before the deployment of the actual system.

In this chapter, the linear space of the related technical indicators is searched by genetic programming, at the same time the non-stationary time series is de-noised using the multi-frequency wavelet analysis. Although the systems' performance in this chapter is a little bit worse than those in Chapters 4 and 5, the risk of this system is much smaller than that in Chapters 4, and 5 as no position is held overnight.

Nonlinear space of the technical characteristics will be shown in the next chapter; wavelet based multi-layer perceptron will also be applied to explore the power of multi-frequency analysis.

CHAPTER 7 – NARX AND WMLP

7.1 Introduction

In this chapter, Artificial Neural Network (ANN) is used to explore the nonlinear space of the financial time series. Two different models will be used to perform the forecast, based on which, high-frequency trading systems are designed and tested.

The first model is the Nonlinear Autoregressive models with eXogenous input (NARX model), which is also called NARX recurrent neural networks. The NARX model is based on ARX model, which is commonly used in time series modeling. NARX is a class of powerful models that has been demonstrated to be well suited for modeling nonlinear systems and especially time series. Some important qualities of gradient-descending trained NARX model are (Eugen et al. 2008):

1. Learning is more effective in NARX networks than in another neural network;
2. Gradient descent networks converge much faster and generalize better than other networks.

The second model is the wavelet based multi-layer perceptron model. Conventional neural networks process signals only on their finest resolutions or some given resolutions. Wavelet decomposition produces a good local representation of the signal in both the time and frequency domains. Wavelet-based Multi-Layer

Perceptron (MLP) model allows for hierarchical, multi-resolution learning of input-output maps from data. The wavelet-based MLP neural network is an MLP with the wavelet decomposition as a feature extraction method to obtain time-frequency information. This model has been successfully applied to classification of biomedical signals, image, and speech; some researchers also used it to forecast financial time series. In this study, the wavelet MLP model is used to predict the intra-day high frequency time series, based on which, real-time trading systems are developed and tested.

7.2 Related Methodology

7.2.1 The Nonlinear Autoregressive with eXogenous model

The nonlinear autoregressive network with exogenous inputs is a recurrent dynamic network, with feedback connections enclosing several layers of the network. As the NARX model is based on the linear time series ARX model, it is well suited for the analysis of financial time series.

The defining equation for the NARX model is:

$$y(t)=f(y(t-1), y(t-2), \ldots, y(t-n_y), u(t-1), u(t-2), \ldots, u(t-n_u)) \tag{54}$$

where the next value of the dependent output time series *y(t)* is regressed on previous values of the output series and previous value of an independent (exogenous) input time series. In this study, the trading volume will be used as the exogenous value. The NARX model can be implemented using a feedforward neural network to approximate the function *f*. A diagram of the resulting network is shown below, where a two-layer feedforward network is used for the approximation.

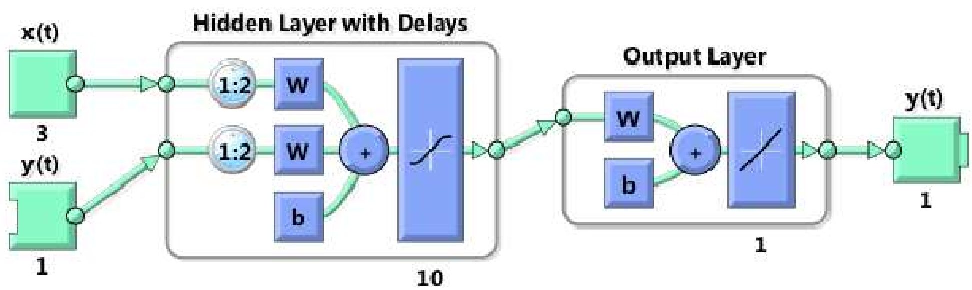

Figure 7.1 Two-layer feedforward network for NARX

There are two types of NARX model, as shown on the left of figure 7.2,

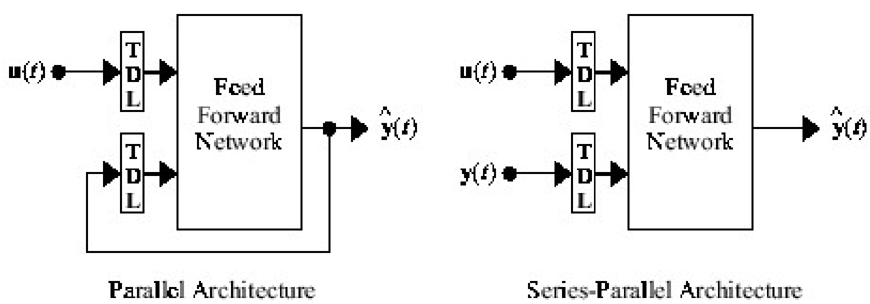

Figure 7.2 Two types of NARX model

(www.mathwork.com)

In the parallel architecture of NARX, the estimated *y(t)* is fed back into the training process. In the Series-Parallel Architecture of the NARX model, the real *y(t)* series is fed into the training process. The Series-Parallel Architecture has two advantages: the first is that the input to the feedforward network is more accurate as the *y(t)* is the original output time series. The second is that the resulting network has a pure feedforward architecture, and static backpropagation can be used for training. In this study, the Series-Parallel Architecture type of the NARX model is used to predict the intra-day financial time series. Matlab R2013b with the neural network toolbox is used to conduct the forecasting experiments.

7.2.2 The Wavelet based Multi-Layer Perceptron model

The wavelet-based Multi-Layer Perceptron (MLP) consists of three parts:

1. Input with a tapped delay line with *j* delays;
2. Wavelet decomposition;
3. A Multi-Layer Perception neural network;

The output X_{t+1} is the value of the time series at time *t*+1 and is assumed to be a function of the values of the time series at j previous time steps.

7.2.2.1 Wavelet and Wavelet decomposition

I: Wavelet

Mathematically, a wavelet can be defined as a function of a zero average:

$$\int_{-\propto}^{\propto} \psi(\mathrm{t})\mathrm{dt} = 0 \qquad (23)$$

The concept of wavelet transformation is to represent a target function as a superposition of a set of wavelets which are small waves located at different times.

The original series can be transformed and described in the content of different frequency over time at certain time-scale. Wavelet-based models are very suitable for non-stationary data as some transient trends, or hidden characteristics of the original series can be highlighted. Financial time series are well-known to be non-stationary, which makes wavelet transformation one of the most powerful tools in the analysis and forecasting of financial time series.

II: The Discrete Wavelet Transform

For the most practical studies, wavelet coefficients are not needed to be calculated at every possible scale as it would be a fair amount of work to do that, which will generate an awful lot of data. The DWT projects a time series onto a collection of orthonormal transformations. Unlike the continuous wavelet transformation that considers all possible frequencies at continuous times, the DWT focuses on some specific frequencies at distinct times. The discrete wavelet is defined as:

$$\psi_{j,k}(t) = S_0^{-\frac{1}{j}} \psi(\frac{t - k\tau_0 S_0^j}{S_0^j}) \qquad (26)$$

j and k are integers. $s_0 > 1$ is a fixed dilation step and the translation factor τ_0 depends on the dilation step. If $s_0 = 2$, the discrete wavelet transformation is defined as dyadic wavelet transform. By definition, the scaling function and the wavelet function of discrete wavelet transformation are:

$$\varphi(2^j t) = \int_{i=1}^{k} a_k \, \varphi(2^{j+1} t - k) \qquad (27)$$

Equation [eq3] is the scaling function, known as the scaled father wavelet, a_k are the corresponding coefficients, $k \epsilon Z$. The mother wavelet ψ is obtained by the linear combinations of the scaled father wavelet. Proper coefficients must be selected to maintain the orthogonality of the basis wavelets.

$$\psi(2^j t) = \int_{i=1}^{k} b_k \, \varphi(2^{j+1} t - k) \tag{28}$$

The mother wavelet ψ are uniquely determined by their coefficients $\{b_k\}$, $k \epsilon Z$. Then, a signal $f(t)$ can be written as:

$$f(t) = \int_{i=1}^{k} a_{j,k} \, \varphi(2^{j+1} t - k) + \int_{i=1}^{k} b_{j,k} \, \psi(2^{j+1} t - k) \tag{29}$$

This is the basic idea of the discrete wavelet decomposition of the signal.

III: Wavelet decomposition

The original signal can be decomposed into different elements based on different scales using the cascade algorithm, as shown in the following equation:

$$\begin{aligned} \mathrm{f(t)} &= \mathrm{A_1(t)} + \mathrm{D_1(t)} \\ &= \mathrm{A_2(t)} + \mathrm{D_2(t)} + \mathrm{D_1(t)} \\ &= \mathrm{A_n(t)} + \mathrm{D_n(t)} + \mathrm{D_{n-1}(t)} + \cdots + \mathrm{D_1(t)} \end{aligned} \tag{30}$$

where $A_n(t)$ are the approximation coefficients at scale n, and $\mathrm{D_n(t)}$ are the detail parts of the original signal at scale n. A prediction of the original signal can be made by the reconstruction of the different scales of wavelets with proper coefficients.

IV: Daubechies D4 wavelet:

The Daubechies wavelet is named after the mathematician Ingrid Daubechies. The Daubechies D4 wavelet transform has four wavelet and scaling function coefficients. The scaling function coefficients are:

$$h_0 = \frac{1+\sqrt{3}}{4\sqrt{2}} \tag{61}$$

$$h_1 = \frac{3+\sqrt{3}}{4\sqrt{2}} \tag{62}$$

$$h_2 = \frac{3-\sqrt{3}}{4\sqrt{2}} \tag{63}$$

$$h_3 = \frac{1-\sqrt{3}}{4\sqrt{2}} \tag{64}$$

Each step of the wavelet transform applies the scaling function to the data input. If the original data set has N values, the scaling function will be applied in the wavelet transform step to calculate N/2 smoothed values. In the ordered wavelet transform the smoothed values are stored in the lower half of the N element input vector.

The wavelet function coefficient values are:

$$g_0 = h_3 \tag{65}$$

$$g_1 = -h_2 \tag{66}$$

$$g_2 = h_1 \tag{67}$$

$$g_3 = -h_0 \tag{68}$$

Each step of the wavelet transform applies the wavelet function to the input data. If the original data set has N values, the wavelet function will be applied to calculate N/2 differences (reflecting change in the data). In the ordered wavelet transform the wavelet values are stored in the upper half of the N element input vector.

The scaling and wavelet functions are calculated by taking the inner product of the coefficients and four data values. The equations are shown below:

Daubechies D4 scaling function:

$$\alpha_i = h_0 s_{2i} + h_1 s_{2i+1} + h_2 s_{2i+2} + h_3 s_{2i+3} \tag{69}$$

Daubechies D4 wavelet function:

$$c_i = g_0 s_{2i} + g_1 s_{2i+1} + g_2 s_{2i+2} + g_3 s_{2i+3} \tag{70}$$

Each iteration in the wavelet transform step calculates a scaling function value and a wavelet function value. The index i is incremented by two with each iteration, and new scaling and wavelet function values are calculated.

7.2.2.2 The Multi-Layer Perceptron

The artificial neuron consists of a summing function with an internal threshold and "weighted" inputs as shown below.

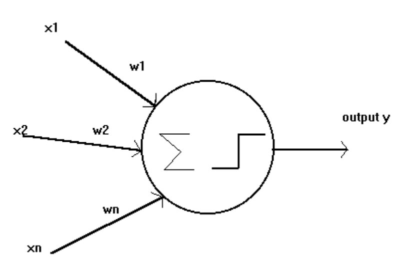

Figure 7.3 Simple Neuron demonstration

For a neuron receiving *n* inputs X_i ($i = 1, 2, \ldots, n$) is weighted by multiplying it with a weight W_i. The sum of the W_iX_i products give the net activation of the neuron. This activation value is subjected to a *transfer function* to produce the neuro's output.

There are many types of transfer functions used in ANNs to compute the output of a node for its net activation, the one that is used in this study is the Sigmoid function.

The *sigmoid* transfer function produces a continuous value in the range 0 to 1. It has the form:

$$output_i = \frac{1}{1 + e^{-activation_i}} \tag{71}$$

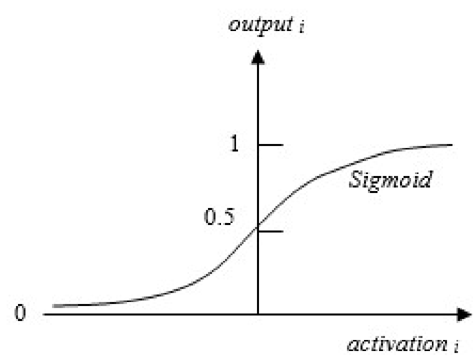

Figure 7.4 Functional form of sigmoid

To be able to solve nonlinearly separable problems, some neurons are connected in layers to build a multilayer perceptron. Each of the perceptrons is used to identify small linearly separable sections of the inputs. Outputs of the perceptron are combined into another perceptron to produce the final output. In a multilayer perceptron, the neurons are arranged in an input layer, an output layer, and one or more hidden layers.

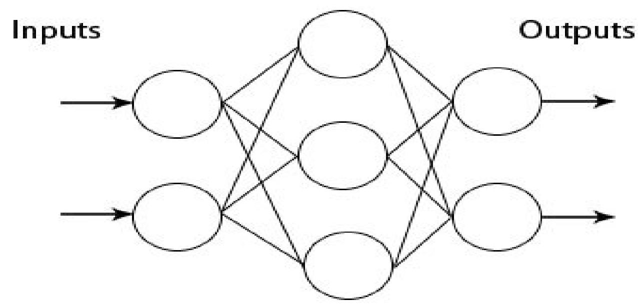

Figure 7.5 The multilayer perceptron model

The Levenberg-Marquardt (LM) method is chosen in this study as the training algorithm. The LM method is a popular method of finding the minimum of a function $F(x)$ that is a sum of squares of nonlinear functions,

$$\mathrm{F(x)} = \frac{1}{2}\sum_{i=1}^{m}[f_i(x)]^2 \tag{72}$$

Let the Jacobian of $f_i(x)$ be denoted $J_i(x)$, then the LM method searches in the direction given by the solution P to the equations:

$$(J_k^T J_k + \lambda_k I)p_k = -J_k^T f_k \tag{73}$$

where λ_k are nonnegative scalars and I is the identity matrix. The method has the nice property that, for some scalar Δ related to λ_k, the vector p_k is the solution of the constrained sub-problem of minimizing $||J_k p + f_k||_2^2$ subject to $||p||_2 \leq \Delta$.

The LM learning process in this study is executed using the Matlab toolbox.

7.2.2.3 The Wavelet Multilayer perceptron

The combination of wavelet decomposition and multilayer perceptron is shown in Figure 7.6.

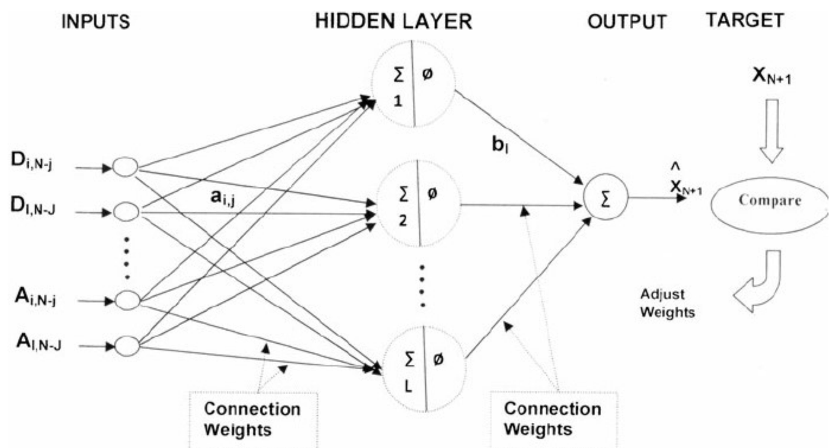

Figure 7.6 Wavelet based multi-layer perceptron

where $A_{i,N-j}$ are the approximation coefficients at scale n, and $D_{i,N-j}$ are the detail parts of the original signal at scale n; $a_{i,j}$ is the connection weights between the input layer and the hidden layer, b_i is the connection weights between the hidden layer and the output layer. The sum of the weighted value is computed to get the net activation, which will then feed to the transfer function to get the output.

7.2.3 The design of trading system based on the forecast models

The above NARX and Wavelet MLP are used to give the one step prediction of the financial time series; trading actions can be taken based on the analysis of the forecasting results. A basic trading strategy based on the forecast results is made to test the trading performance of the two models. The trading test logic is discussed.

7.2.3.1 Entry point

This is the point a buy or sells short order is placed into the market, after the execution of that kind of orders, the portfolio ends up with either holding long or short positions of the selected targets. With the application of the above forecasting methods, the entry point is quite easy to be selected. If the point prediction is above (below) the price of the specified targets at the current time, or the directional prediction of the financial time is up, an entry point of buying (selling) the specified asset is determined.

7.2.3.2 Leaving point

This is a point that the trader decides to leave the market either by the placement of sell or buy to cover orders. After the executions of such kind of orders, the portfolio ends up with zero position of a specified target. The leaving point can be reached by exceeding the maximum trading profit or maximum trading loss. It can also be triggered by the predictions of the above models, if the point prediction of the financial time series is below (above) the current price of the targets, a leaving point of the long (short) positions of the assets is reached and vice versa.

The testing strategy is shown below.

1. Assume the initial position is zero, let flag =0 represent the zero position.

   If *flag =0* and *forecast value- the real value> a*

(Here *a* is a positive threshold to decide whether an order should be put here.) Then long the assets, and the flag is set to represent the long position 1, if *flag =1* and the *real value – forecast value > a,* then short the asset and set *flag=-1*

2. If the *flag =1* and the *forecast value- the real value>0* then hold the long position, the flag is not changed here. If the *flag =1* and the *forecast value- the real value <0* and the *real value- forecast <a* then close the long position and set the *flag=0*, If the *flag =1* and the *forecast value – the real value <-a* then close the long position and open another short position and set *flag=-1*

3. If the *flag =-1* and the *real value - the forecast value >0* then hold the short position, the flag is not changed here. If the *flag =-1* and the *real value - the forecast value >0* and the *real value- forecast <a* then close the long position and set the *flag=0*, If the *flag =-1* and the *forecast value – the real value >a* then close the short position and open another long position and set *flag=1*

The program to conduct this test is provided in the Appendix.

7.3 Empirical experiment

The 1-minute time series of the China Financial Index futures is used as the experiment data. Similar to the experiment in Chapter 6, the CSI 300 Index futures is selected as the target asset for this study. Detailed information about the index futures is shown in Table 6.1. The 90-day data of the CSI 300 index futures are used, each of which contains 270 data points. The original data points cover the period from 4 January 2016 to 28 March 2016.

Table 6.1 The CSI 300 Index

Underlying Index	CSI 300 Index
Contract Multiplier	CNY 300
Unit	Index point
Tick Size	0.2 point
Contract Months	Monthly: current month, next month, next two calendar quarters (four total)
Trading Hours	09:30 am - 11:30 am, 01:00 pm - 03:00 pm
Limit Up/Down	+/-10% of settlement price on the previous trading day
Margin Requirement	8% of the contract value
Last Trading Day	Third Friday of the contract month, postponed to the next business day if it falls on a public holiday
Delivery Day	Third Friday, same as "Last Trading Day"
Settlement Method	Cash Settlement
Transaction Code	IF
Exchange	China Financial Futures Exchange

The data set is separated into three parts, namely the training part, the validation part and the out of sample testing part.

The data set used in both of the models is shown in Figure 7.7.

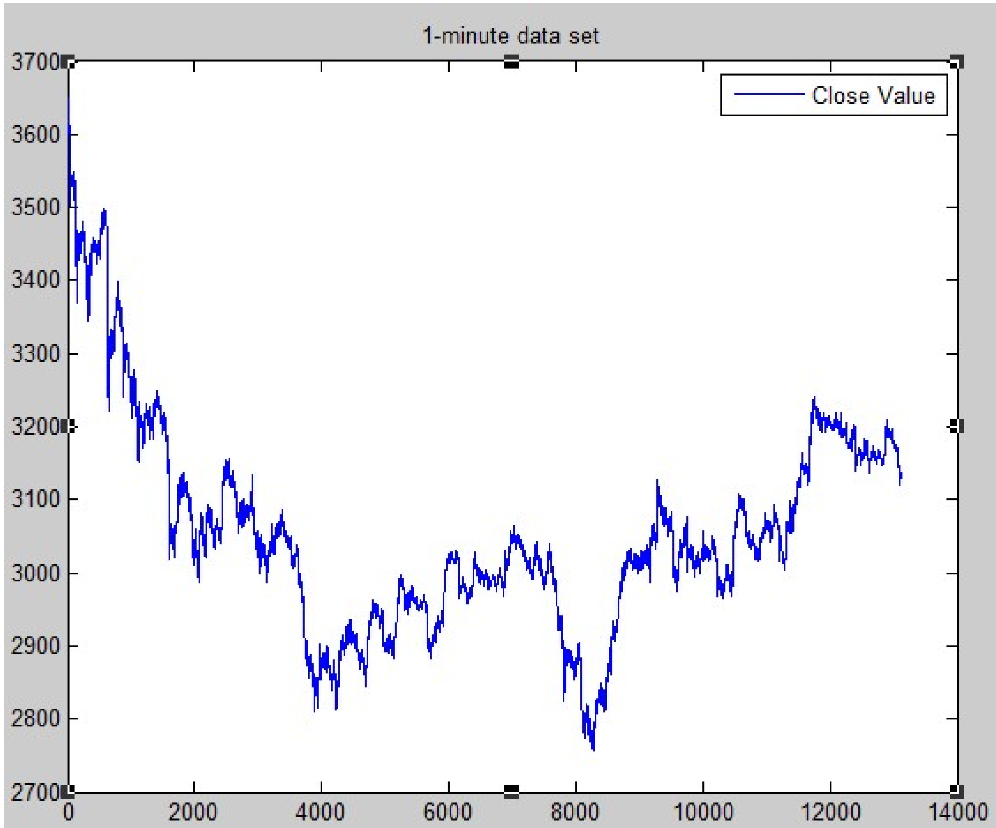

Figure 7.7 1-minute data set used in this chapter

7.3.1 High frequency forecasting with NARX model

In this model, the 1-minute close price is the time series to be forecasted; there is total 13115 data point and the separation of the training, validation and testing is shown in the Figure 7.8.

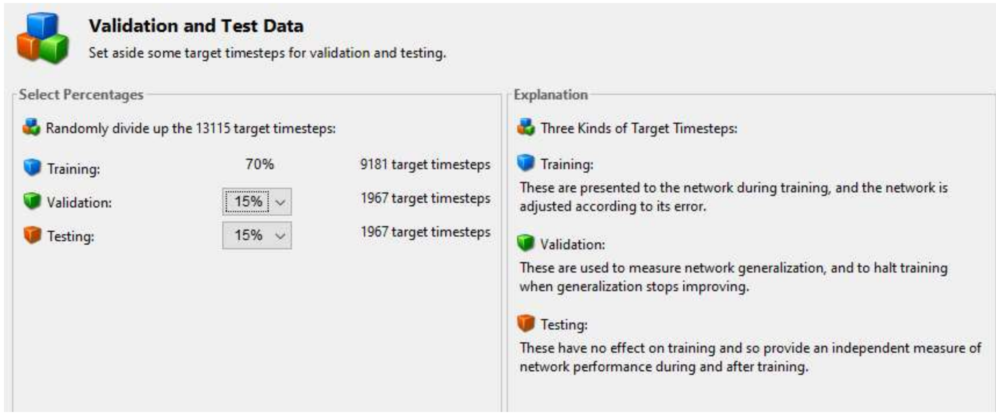

Figure 7.8 Data separation: Training, Validation, and Testing

Based on numbers of experiments' results, 10 is selected as the number of hidden neurons based on the mass of experiments, and the number of delays is set to 6. The form of the constructed neural network is shown in Figure 7.9.

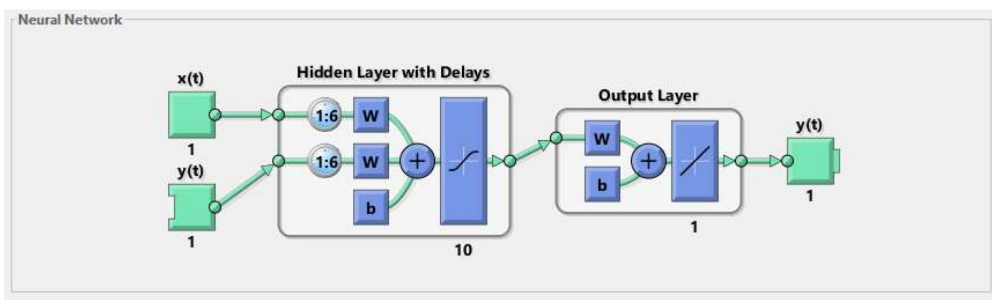

Figure 7.9 Neural Network for the NARX Model

The Levenberg-Marquardt backpropagation method is used in the training process.

The training results are shown in Figure 7.10.

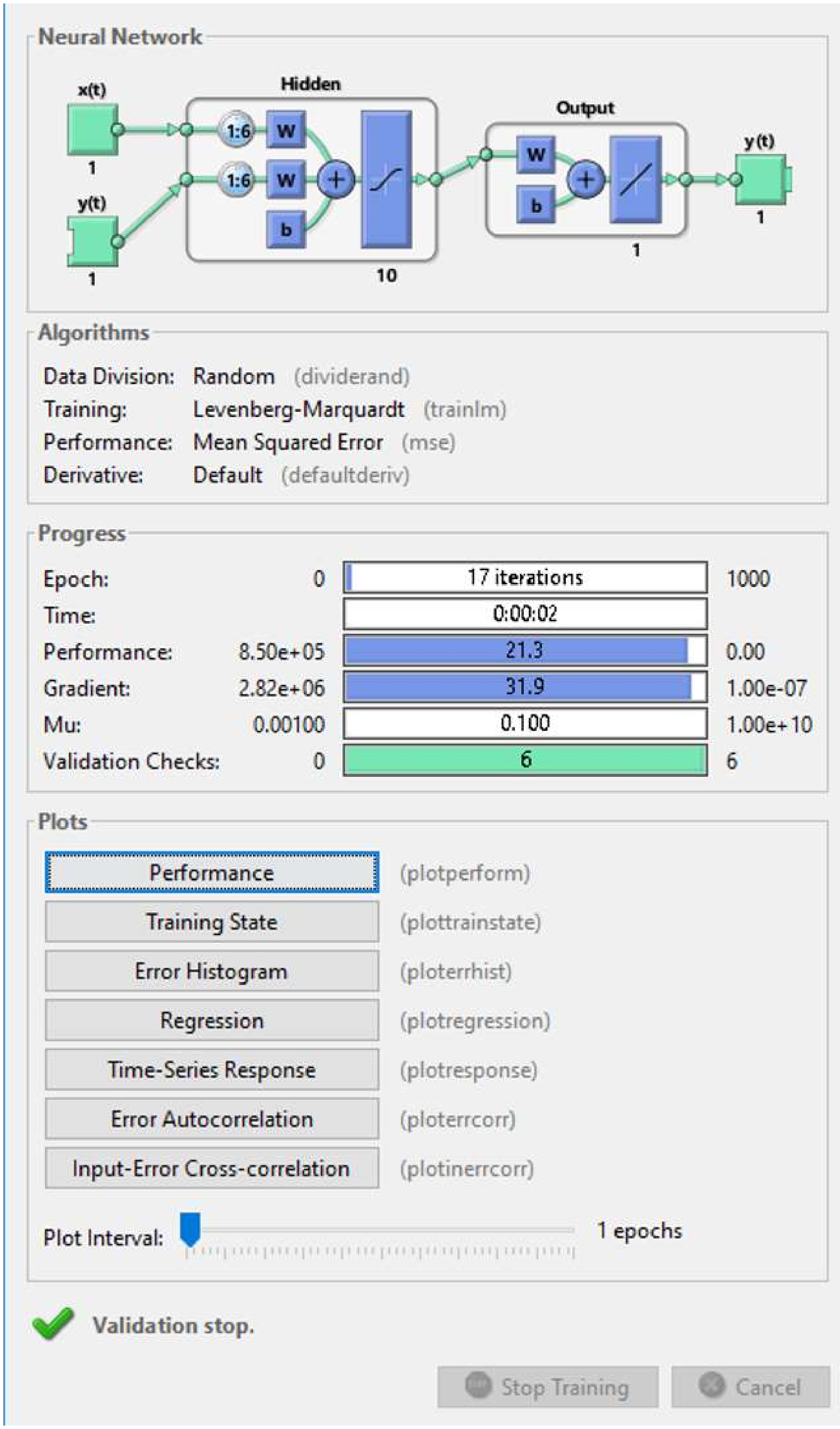
Neural Network
x(t)
Hidden
Output
y(t)
1:6
W
b
1
10
1
1
Algorithms
Data Division: Random (dividerand)
Training: Levenberg-Marquardt (trainlm)
Performance: Mean Squared Error (mse)
Derivative: Default (defaultderiv)
Progress
Epoch: 0 17 iterations 1000
Time: 0:00:02
Performance: 8.50e+05 21.3 0.00
Gradient: 2.82e+06 31.9 1.00e-07
Mu: 0.00100 0.100 1.00e+10
Validation Checks: 0 6 6
Plots
Performance (plotperform)
Training State (plottrainstate)
Error Histogram (ploterrhist)
Regression (plotregression)
Time-Series Response (plotresponse)
Error Autocorrelation (ploterrcorr)
Input-Error Cross-correlation (plotinerrcorr)
Plot Interval: 1 epochs
Validation stop.
Stop Training
Cancel

Figure 7.10 Training results for the NARX model

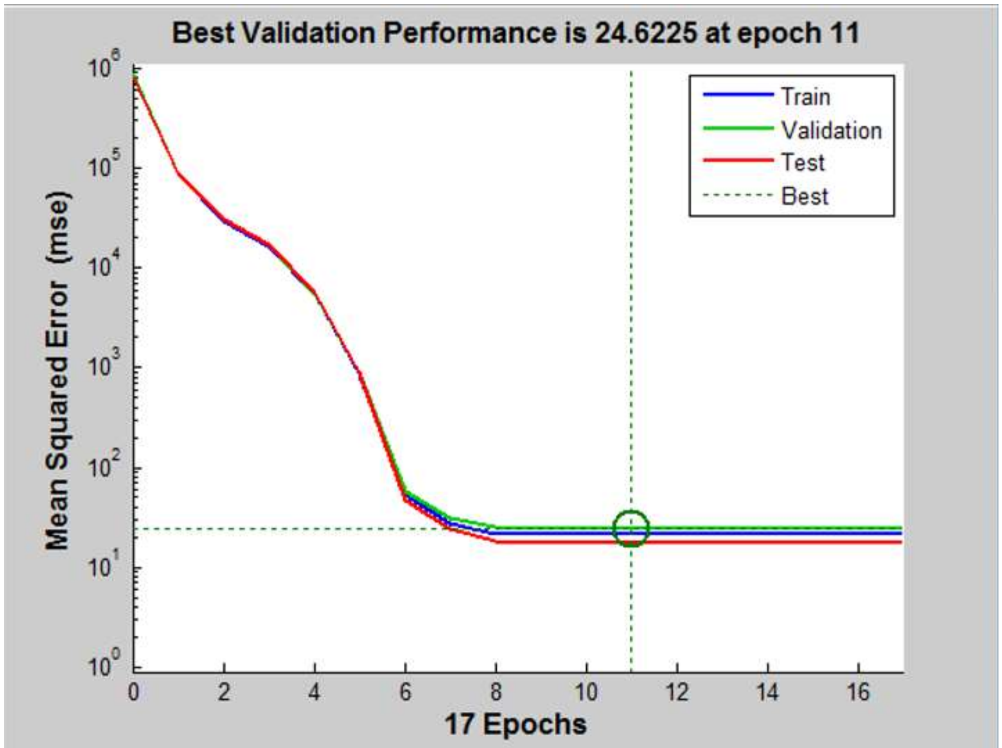

Figure 7.11 Network Training performance

The network training performance in illustrated in Figure 7.11. The training state of the network is indicated in Figure 7.12.

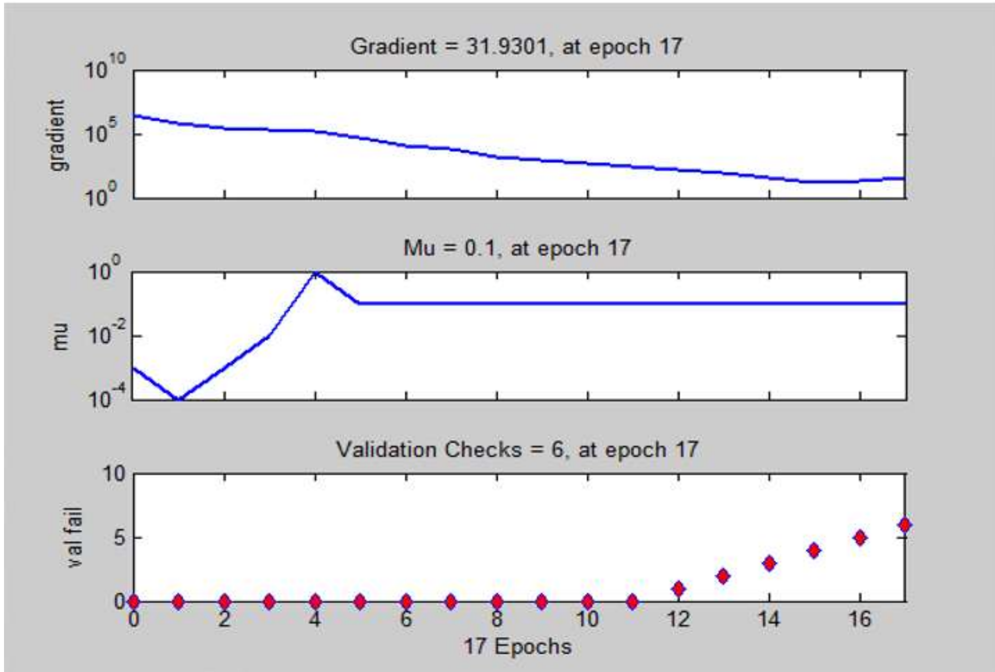

Figure 7.12 Network training state

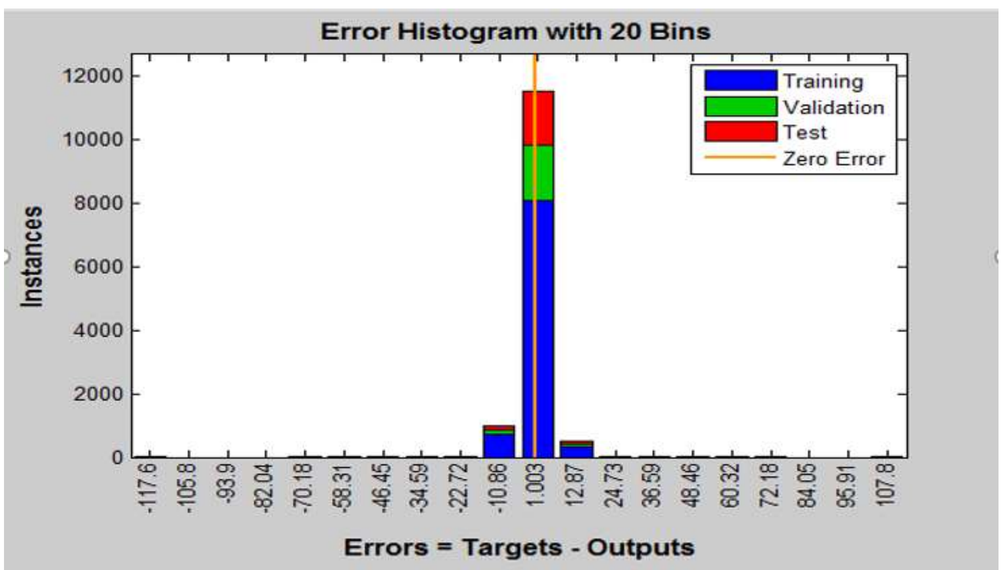

Figure 7.13 The Error Histogram for the NARX model

Judging from the errors in Figure 7.13, both the training and test set have a larger than zero errors.

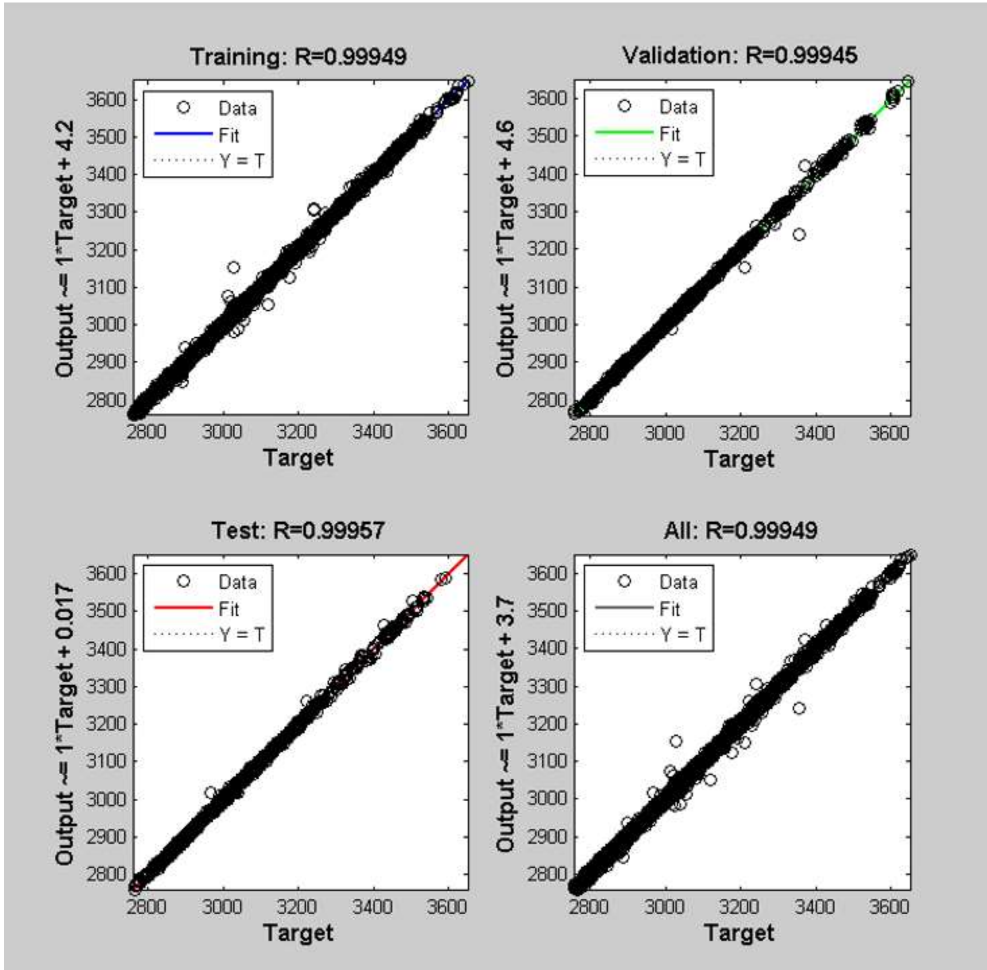

Figure 7.14 The Network Training Regression

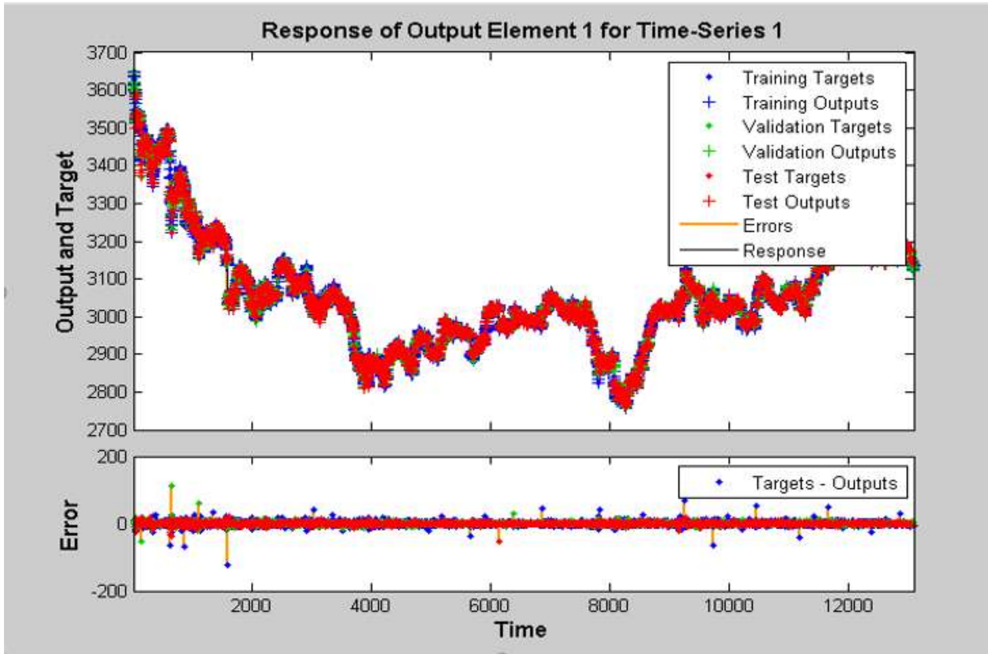

Figure 7.15 Time series Response

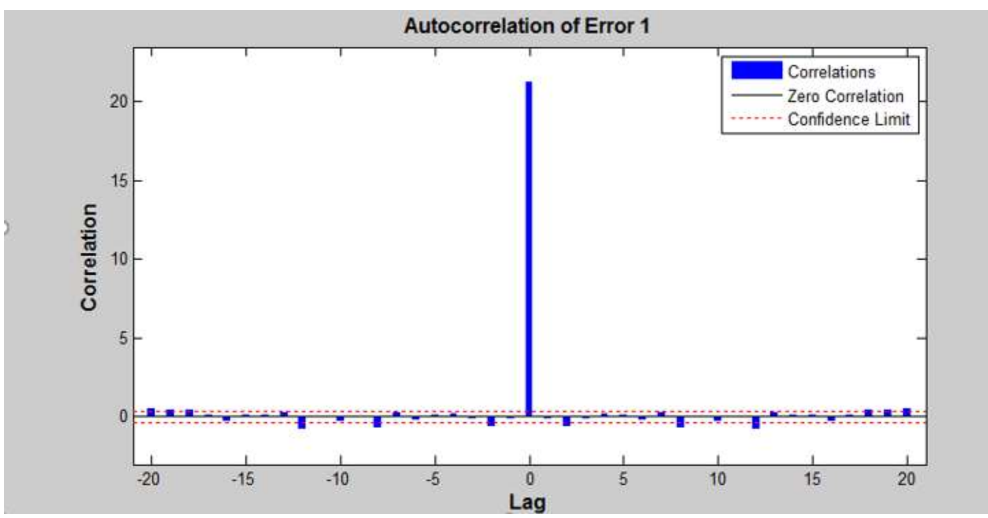

Figure 7.16 The Training Error Autocorrelation

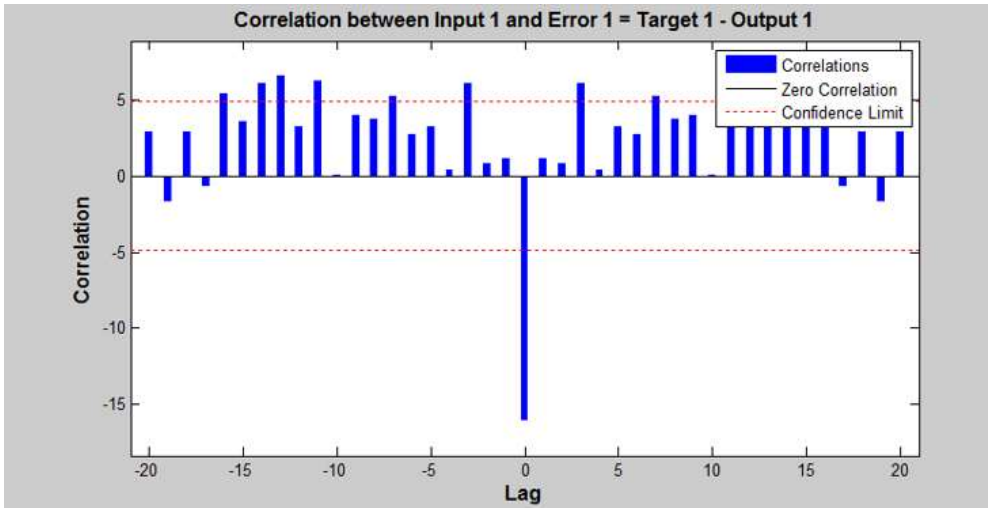

Figure 7.17 The Input -Error Cross-correlation

The networking training error autocorrelation is generally within the confidence limit.

The out of sample trading results based on the forecasting results and the proposed trading strategies are shown in Table 7.2:

Table 7.2 Out of sample trading results for NARX model

Value of *a*	Long trade No.(M)	Short trade No.(N)	Profit/Loss	Tran-cost (M+N)*70	Net profit
0	735	734	71.4*300	102830	-81410
0.2	512	558	-81.8*300	74900	-99440
0.4	231	259	13.2*300	34300	-30340
0.6	75	81	11.2*300	10920	-7560
0.8	34	39	19.6*300	5110	770
1.0	24	22	15.4*300	3220	1400
1.2	20	18	15.2*300	2660	1900
1.4	15	14	4*300	2030	-830
1.6	12	12	6.4*300	1680	240
1.8	9	9	7.8*300	1260	1080
2.0	6	8	-0.2*300	980	-1040
2.2	2	3	0	350	-350
2.4	2	3	0	350	-350

To simplify the calculation, a transaction cost of 70 yuan per trading is used in this experiment. From the results, it can be seen that the highest profit is obtained with a = 1.2. This result is positively related to the forecast errors.

7.3.2 High-frequency forecasting with wavelet based MLP model

In this experiment, the original 1-minute time series is firstly decomposed using the Db4 (details of Db4 wavelet is shown in Appendix I)wavelet at level 5, the decomposing results are shown in Figure 7.18

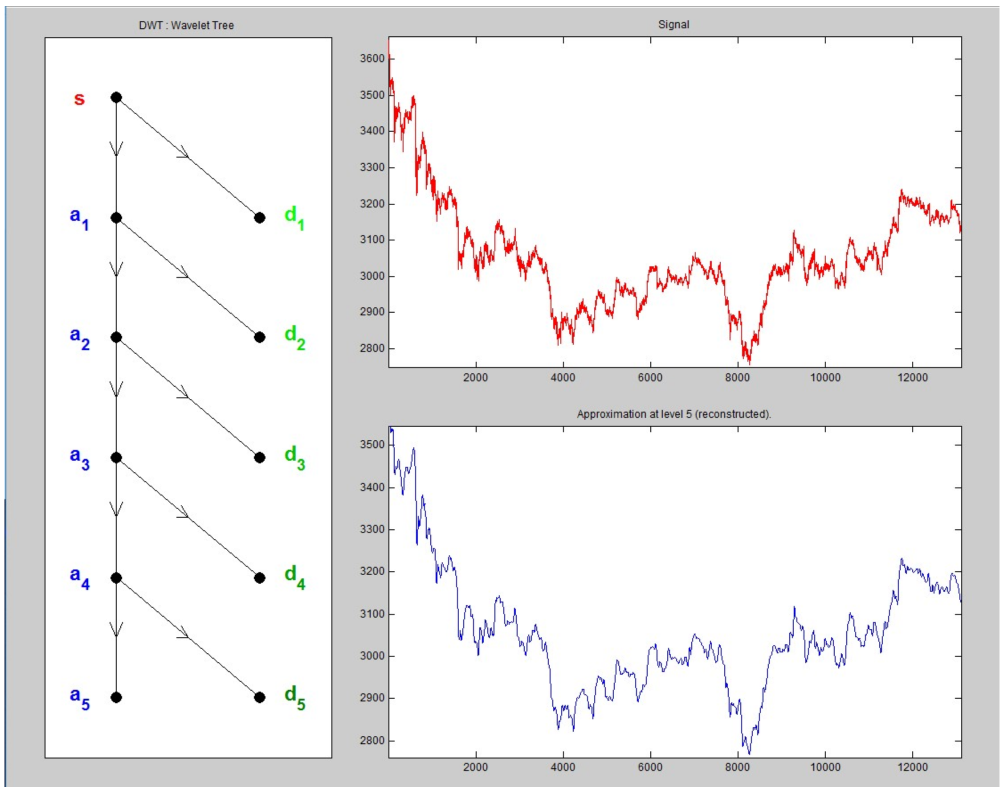

Figure 7.18 The db4 5 level decomposition

After the decomposition, both the approximations and the details part of the original time series are fed into a neural network to perform the 1-step ahead prediction.

The training process and results using Matlab are shown in Figures 7.19-7.23.

As the same data set in this model as NARX are used, the first 70% of the data are used to train the model, 15% of the samples are used to validate the model, and the last 15% of the samples are used to conduct out of sample test. Both the approximations and the details of the original time series are fed into the training process, there are total ten input variables for this model, and the number of hidden neurons is set to 10, the structure of the Neural Network is shown in Figure 7.19.

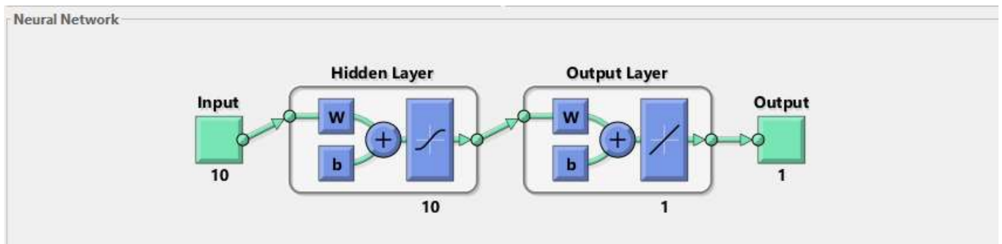

Figure 7.19 Structure of the Wavelet MLP Network

Figure 7.20 illustrated the training process of the neural network.

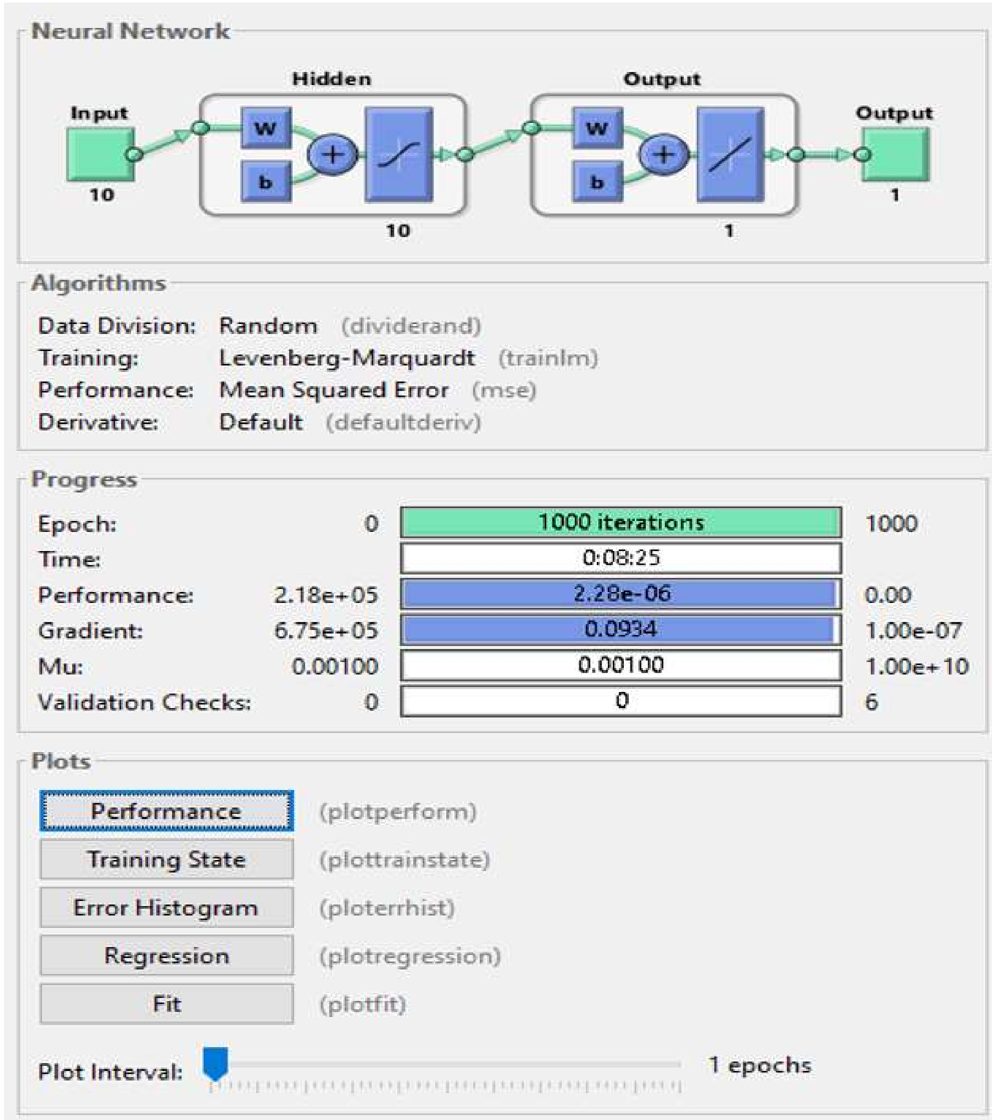

Figure 7.20 Training process of the Wavelet MLP network

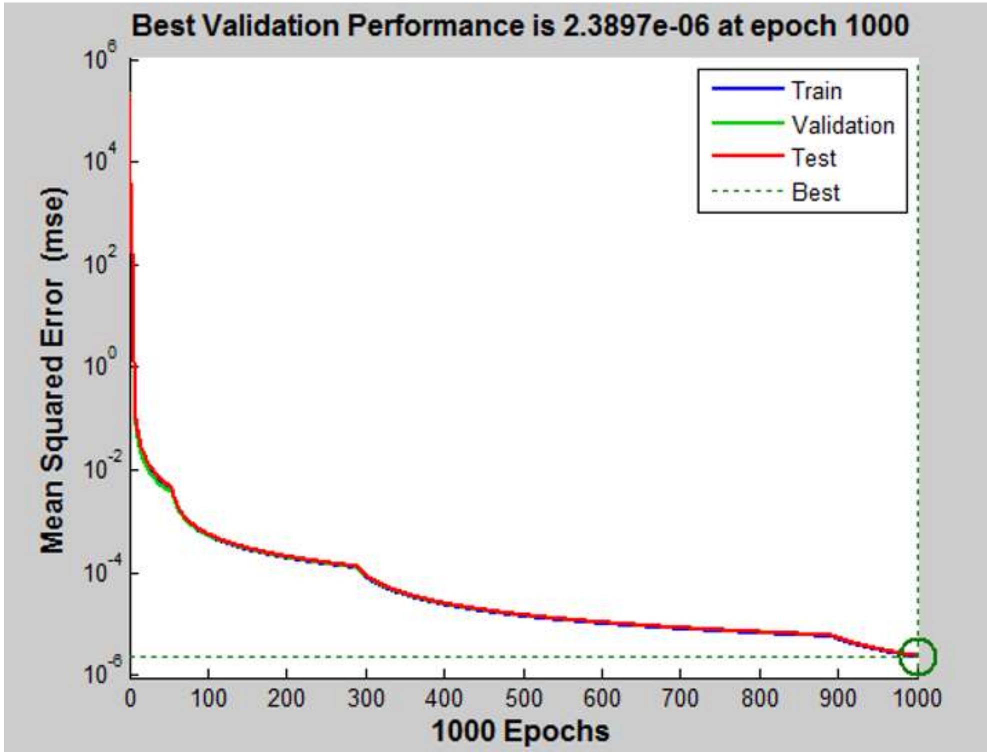

Figure 7.21 Performance of the Wavelet MLP network

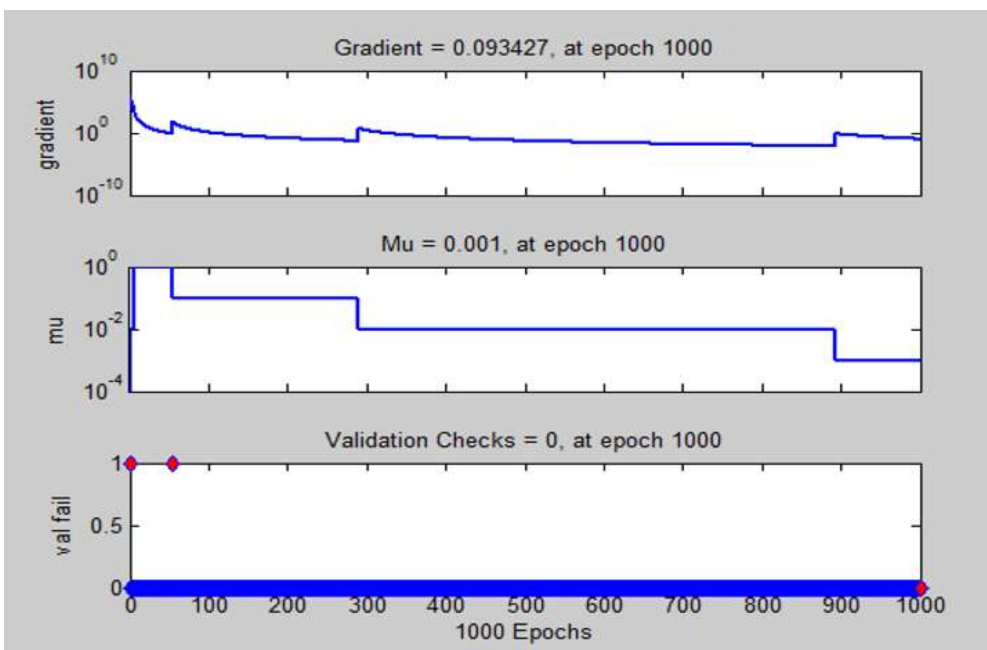

Figure 7.22 Training state of the Wavelet MLP network

From figure 7.22, it is clear that the network did not pass the validation check, and the training process was ended by reaching the max epochs. But the *R* value in Figure 7.23 is extremely high; this is mainly due to the reason that some of the input variables are highly correlated with each other in the output time series.

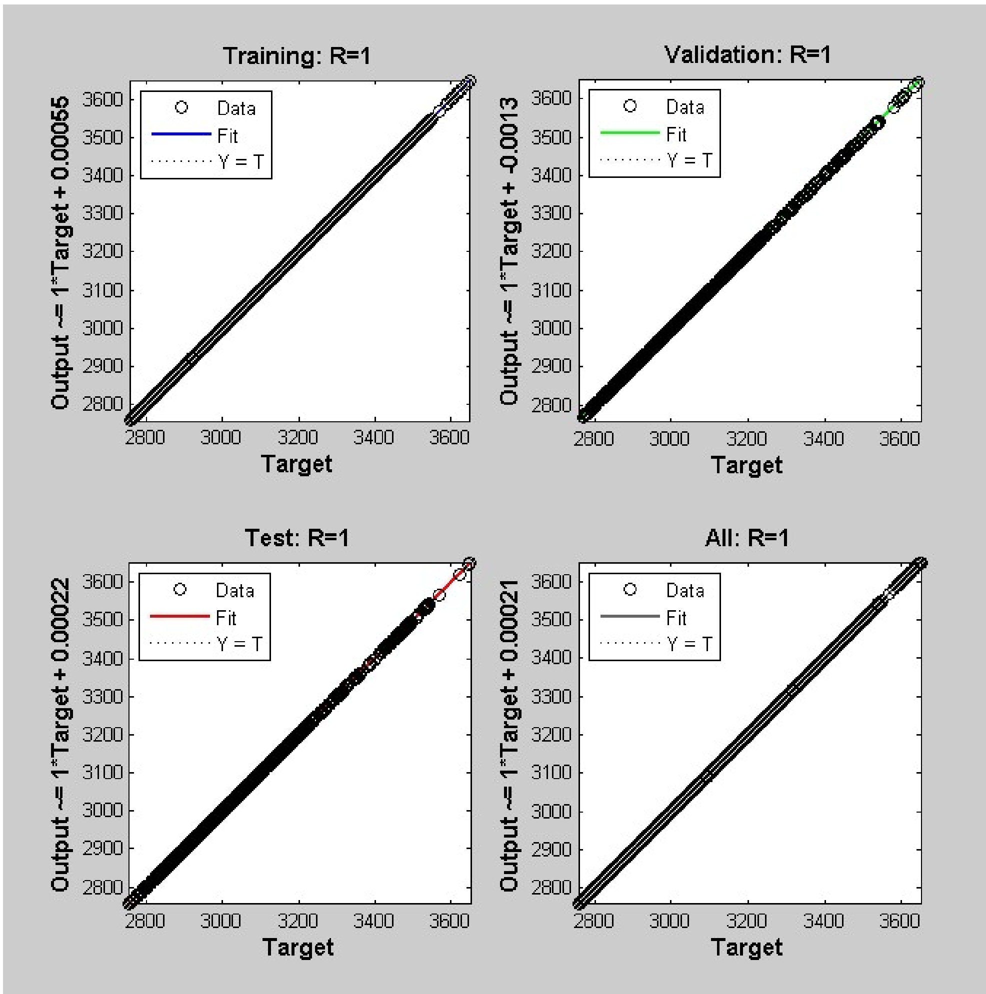

Figure 7.23 Training regression of the Wavelet MLP network

The out of sample trading results based on the forecasting results and the proposed trading strategies are shown in Table 7.3

Table 7.3 Out of sample trading results for Wavelet MLP model

Value of *a*	Long trade No.(M)	Short trade No.(N)	Profit/Loss	Tran-cost (M+N)*70	Net profit
0	1067	1066	4182.6	149310	1105290
0.2	1065	1056	4068.2	148470	1071990
0.4	1055	1052	3928.6	147490	1031090
0.6	1033	1038	3806.6	144970	997010
0.8	1012	1010	3590.8	141540	935700
1.0	982	975	3406.2	136990	884870
1.2	961	944	3249.6	133350	841530
1.4	933	900	3062.2	128310	790350
1.6	898	868	2830.4	123620	725500
1.8	844	820	2642.2	116480	676180
2.0	790	744	2494.2	107380	640880
2.2	741	721	2327.2	102340	595820
2.4	681	658	2200.8	93730	566510

The trading test results of this wavelet based MLP model are much better than that one in the NARX model. Although the performance of the Wavelet MLP based trading system is very promising, the profit shown in Table 7.3 is very difficult to obtain from the real market due to the following reasons. Firstly, no execution delay and slippage considered the assumption of the back test, Secondly, the China index futures market has a trading volume limit, that one investor can only trade no more than ten shares of contracts, which means, the total trade should be no more than

900. The trading results could be largely affected by this regulation.

7.4 Summary

In this chapter, two forecasting models, namely the Nonlinear Autoregressive with eXogenous inputs (NARX) and the Wavelet based Multilayer perceptron model are used to give the one-step ahead forecast, based on which, trading strategies are developed and tested. A brief conclusion is drawn as follows.

1. In the training of the NARX model, the volume series was used as the exogenous inputs. The out of sample performance of the trained model is quite promising. However, the high R value may be due to the correlation between the delayed input value, as shown in the correlation figure, the inputs are highly correlated with each other. The trading system based on the NARX forecast's performance relies on the value, which is the threshold to entry the market.

2. With the combination of wavelet decomposition and multilayer perceptron, the WMLP method is used to predict the intra-day high-frequency time series. Although the forecasting error is smaller compared to the NARX model, this model does not pass the validation test, which indicates the existence of the overfitting problem. However, the performance of the out of sample test's performance is promising.

3. To show the real profitability of the proposed trading strategy, execution delay, and slippage need to be taken into consideration. At the same time, specified regulations on the trading of index futures also have a dramatic impact on the

performance of the trading system.

4. The non-stationary and correlation problem is not well taken care of in this study, some methods to de-noise the time series and to reduce the correlation need to be carried out in the future.

The next chapter, which is the last part of this dissertation, will present an overall summary of this project and also provide some recommendations for future research.

CHAPTER 8 – Conclusions and future work

8.1 Distinctive Achievements

High-Frequency Trading (HFT) occupies a great amount of volume in both the developed and the emerging markets. Many researchers have been conducted to explore the HFT strategies in the developed market. This research focuses on the emerging index futures market of China. The intra-day trading systems were designed based on the multi-frequency analysis of the index futures time series. Various types of forecasting models have been applied in this research to for a better forecast that could be used in the design of algorithm trading systems. Machine learning tools like GP and ANN are also utilized in the searching of both linear and nonlinear combinations of technical indicators and forecasting models. To explore the frequency domain of the time series, wavelet decomposition and wavelet de-noise are used to extract useful information conveyed in a different frequency of the time series and to reduce the impact of the noise on the training of the systems.

In this research, the remarkable achievements in the HFT multi-frequency study are summarized in the following:

1. Studies on HFT in the literature mainly focused on the impact of the HFT on the market. Traditional financial time series forecasts are mostly based on the analysis of rather long time intervals, some of which even used yearly data. To find some forecasting models and indicators that could be used in the design

of HFT systems, this research uses the high-frequency data from the China Financial Exchange; the index futures is selected as the target assets. In the training process of some models, technical indicators utilized in the forecasting models are calculated using only the data coming from the same trading day to avoid the impact of jumping points on the training process of the proposed models.

2. A full examination of the performance of the moving average based trading systems is conducted using the intra-day high-frequency data. The performance of the SMA， EMA and HMA based trading systems are compared with other. The HMA trading strategy showed a most promising profit with the lowest drawdown.

3. A manually two-frequency analysis of the intra-day high-frequency data is conducted using the ARMA model, the combination of the forecasting results from different frequency showed to have smaller RMSE, which means a better prediction can be made with the integration of different frequency of data. A simple trend following system is also developed and presented.

4. To reduce the impact of noise on the training of the GP model, wavelet de-noise is introduced to process the original noisy and non-stationary financial time series. Both soft-threshold and hard-threshold wavelet de-noises are utilized in the analysis. The GP model trained based on the hard-threshold

wavelet de-noise has the better out-of-sample trading performance.

5. The GP model proposed in this research optimized the linear combinations of various types of technical indicators and the order types. To explore the nonlinear relationship of the financial time series, the NARX and the wavelet-based multi-layer perceptron network are trained to conduct the one step ahead forecast, based on which, the trading strategy is constructed and tested using the out of sample data set. The performance of the trading strategy indicates that wavelet decomposition—the multi-frequency analysis does have a significant impact on the performance of the trading strategies.

6. In this thesis, different forecasting models and the combinations of technical indicators are used in the design of expert trading systems. With the application of more complex and more accurate forecasting models, more efficient trading systems can be built. This research provides a clear path to apply multi-frequency analysis to the forecasting process of the financial time series and the trading process of the related assets.

8.2 Academic contributions

With the emphasis of high-frequency trading, some novel application of the multi-frequency analysis tool wavelet and machine learning tools GP and ANN were made in this research. Both linear GP and nonlinear NARX and MLP models are utilized in the searching of the relationship between independent variables and lagged value of the time series.

In this research, six academic contributions associated with the construction of HFT are made.

First of all, the Genetic Programing (GP) based trading system studied in this work performs a higher level of integration on technical indicators and intra-day trading. Traditionally, GP models are trained with long term interval data, during which, the information that may have an impact on the financial time series is released. As shown in most of the financial markets, there are overnight jumps for most of the trading targets. With the consideration of incoming information that leads to overnight jump points, the training bias was significantly reduced and a better performance was provided.

Secondly, past GP based trading systems were trained directly using the original time series from the target assets, the non-stationary characteristic of the time series is not well dealt with. In this project, both soft-threshold and hard-threshold wavelet de-noises of the original time series were conducted before the training of the GP model, the resulting performance of the trading systems indicates that the multi-frequency de-noise is capable of improving the performance of the GP model.

Thirdly, the wavelet de-noised GP model in this project was trained with much larger searching space compared with the GP models in the literature. Unlike the other GP based trading models that only used the technical indicators, different types of the orders were also added into the searching space, resulting in a much larger space of the combinations of both technical indicators and order types.

Fourthly, in the analysis of the nonlinear relationship of the financial time series, both the NARX and a wavelet decomposed based MLP neural network were trained simultaneously to construct trading strategies, the performance of the trading systems was compared with each other. In the training of the NARX model, both price and volume series were used, this neural network implies the nonlinear relationship between the price and volume interaction. At the same time a novel integrated wavelet decomposition based MLP neural network was trained to perform the prediction of the price series, better trading performance was achieved compared to the NARX based trading system with the application of multi-frequency analysis tool—wavelet.

Fifthly, two different ways of applying multi-frequency analysis into HFT were shown in this research, namely the wavelet de-noise and the wavelet decomposition. Both of these two methods provided a path to utilize the information hiding in the frequency domain of the high-frequency time series.

Finally, the performance of the trading systems proposed in this research indicates that the China index futures market is not weak form efficient, as cumulative returns are obtained based on the pure analysis of past trading data.

8.3 Possible benefits to industry

High-frequency trading is an emerging industry in China, many institutional investors and individual investors are working on the development of high-frequency trading systems. With some careful treatment on the execution delay and slippage, the trading systems proposed in this research can be easily deployed into a real production environment.

As a financial time series of different frequency is proved in this research to convey different information about the trend and detailed of the series, multi-frequency analysis can be used in different forms, like the directional prediction of the financial times using the combination forecast of 1-minute, 3-minute and 5-minute data is more convincing than a single frequency forecast result.

More machine learning tools could be integrated with multi-frequency analysis to form better prediction and to extract a pattern from the financial time series, based on which, trading strategy with better performance can be constructed.

8.4 Future work

Some possible further work related to this project in future is suggested as follows:

1. Implementing the HMA based trading system

   The HMA based trading system has a rather stable performance, unlike the SMA and EMA model, the drawdown of this system is quite small. To deploy this system into a real production environment, some more researches on the control of execution delay and slippage need to be conducted.

2. Enriching the multi-frequency based forecast model

   Although the two frequency analysis forecast of the time series showed the strength of multi-frequency analysis, more experiments including more frequency of data are needed to be conducted to test the robustness of this hypothesis. At the same time, ANN and other machine learning tools can be improved with multi-frequency analysis.

3. Improving the performance of the wavelet de-noised genetic programming's performance

   The proposed wavelet de-noised GP model is trained separately for each trading day, and different trading strategy is obtained for different trading days. One possible improvement for this study is to train the GP model simultaneously, which means the GP model is trained using a matrix of the data from the morning trading sessions, and the afternoon trading sessions' data is used to perform the out of sample test. An optimized trading strategy that lasts for a specific long time interval could be formed.

4. Extending the application of ANN to the high-frequency forecasting

For the NARX model in this research, the trading volume is used directly as the exogenous inputs. Some variants of the volume time series may be a better candidate, an NARX model training using the variant of the volume may have better performance. For the wavelet de-noised multilayer perceptron model, only the db4 5 level wavelet decomposition is applied in this research. There are much more wavelets like haar can be combined with the MLP to perform the prediction of the time series.

5. Making more use of the Volume and Bid-Ask book

   In this project, only the NARX neural network took the volume time series into consideration, the other models just focused on the single price time series. Some more information affecting the trend of the price of the time series could be extracted from the change of trading volume and the bid-ask spread. Research on the utilization of volume and bid-ask spread can be helpful in the design of more complex and reliable trading systems.

APPENDIX

Appendix I: Python program for the trading systems

The detailed program is provided in the electronic package of this thesis.

REFERENCES

[Aba98] Abarbanell, J.S. and Bushee, B.J.; 1998, Fundamental Analysis, Future Earnings and Stock Prices, *Journal of Accounting Research,* 35(1), pp. 1-24.

[Aba98] Abarbanell, J.S. and Bushee, B.J.;1998, Abnormal Returns to a Fundamental Analysis Strategy, *The Accounting review,* 73(1), pp. 19-45.

[Abd09] Abdalla, K.;2009, Adaptive Neuro-Fuzzy Inference Systems for High Frequency Financial Trading and Forecasting, *2009 Third International Conference on Advanced Engineering Computing and Applications in Sciences.*

[Add02] Addison, P.; 2002, *The Illustrated Wavelet Transform Handbook,* Institute of Physics, Bristol, UK.

[Alf99] Alford, A.W. and Berger, P.G.; 1999, A Simultaneous Equations Analysis of Forecast Accuracy, Analyst Following, and Trading Volume, *Journal of Accounting, Auditing & Finance,* 14(3), pp. 219-240.

[And05] Andersen, T.G.; Bollerslev, T.; Meddahi, N.; 2005, Correcting the Errors: Volatility Forecast Evaluation Using High Frequency data and Realized Volatilities, *Econometrica,* 73(1), pp. 279-296.

[Ane05] Anersen, T.G.; Bollerslev, T.; Diebold, F.X.; Vega, C.;2005, Real-time Price Discovery in Stock, Bond and Foreign Exchange Markets, (working paper).

[Ats09] Atsalakis, G.S. and Valavanis, K.P.; 2009, Forecasting stock market short-term trends using a neuro-fuzzy based methodology, *Expert Systems with Applications,*36, pp. 10696-10707.

[Bar93] Baruch Lev and S.Ramu Thiagarajan,; 1993, Fundamental Information Analysis, *Journal of Accounting Research* 31(2), pp. 190-215.

[Bas97] Basu, S.; 1997, Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis, *The Journal of Finance,* 32(3), pp.663-682.

[Bro61] Brown, R.G. and Meyer, R.F.; 1961,The Fundamental Theorem of Exponential Smoothing, *Operations Research*, 9(5) pp. 673-687.

[Bha88] Bhandari, L.C.; 1988, Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence, *The Journal of Finance,* 43(2), pp.507-528.

[Bla01] Blair, B.J; Poon, S.H; Taylor, S.J; 2001, Forecasting S&P 100 Volatility: The Incremental Information Content of Implied Volatilities and High Frequency Index Returns, *Journal of Econometrics,* 105, pp. 5-26.

[Ben81] Benz, R.W.; 1981, The Relation Between Return and Market Value of Common Stocks, *Journal of Financial Economics,* 9, pp. 3-18.

[Bro06] Brownlees, C.T.; 2006, Financial Econometric Analysis at Ultra-high Frequency: Data handling Concerns, *Computational Statistics & Data Analysis*, 51, pp. 2232-2245.

[Bol01] Bollerslev, T. and Jonathan, H.W.; 2001, High Frequency Data, Frequency Domain Inference, and Volatility Forecasting, *The Review of Economics and Statistics*, 83(4), pp. 596-602.

[Cai13] Cai, Q.S.; Zhang, D.F.; Wu, B.; Stehpen C.H. Leung.; 2013, A novel stock forecasting model based on fuzzy time series and genetic algorithm, *Procedia Computer Science*, 18, pp. 1155-1162.

[Car11] Carlin, C.F. and Lam, K.P.; 2011, Modeling Intraday Volatility: A New Consideration, *Journal of International Financial Markets, Institutions & Money,*21, pp. 388-418.

[Car12] Cartea, A. and Penalva, J.; 2012, Where is the Value of High Frequency Trading, *Quarterly Journal of Finance,*2(3), pp.1-46.

[Cha08] Chavarnakul, T. and Enke, D.; 2008, Intelligent technical analysis based equivoque charting for stock trading using neural networks, *Expert Systems with Applications,* 34, pp. 1004-1017.

[Cro05] Crowley, P.M.; 2005, An intuitive guide to Wavelets for Economists, Discussion papers of Bank of Finland Research.

[Cha97] Charles A.E. Goodhart and Maureen O'Hara; 1997, High Frequency data in Financial Markets: Issues and Applications, *Journal of Empirical Finance,* 4, pp. 73-114.

[Cha08] Chang, P.C.; Liu, C.H.; 2008, A TSK-type fuzzy rule-based system for stock price prediction, *Expert Systems with Applications,* 34, pp. 135-144.

[Cha11] Chang, J.R.; 2011, A hybrid ANFIS model based on AR and volatility for TAIEX forecasting, *Applied Soft Computing,* 11, pp. 1388-1395.

[Che06] Cheng, C.H.; Chen, T.L.; Chiang, C.H.; 2006, Trend-Weighted fuzzy time-series model for TAIEX forecasting, *Lecture Notes in Computer Science*, 4234, pp. 469-477.

[Che08] Chen, T.L.; Cheng, C.H.; Teoh, H.J.; 2008, High order fuzzy time series based on multi-period adaptation model for forecasting stock markets, *Physica A*, 387, pp. 876-888.

[Che07] Chen, T.L.; Cheng, C.H.; Teoh, H.J.; 2007, Fuzzy time-series based on Fibonacci sequence for stock price forecasting, *Physica A,* 380, pp. 377-390.

[Che08] Cheng, C.H.; Chen, T.L.; Teoh, H.J.; Chiang, C.H.; 2008, Fuzzy time-series based on adaptive expectation model for TAIEX forecasting, *Expert Systems with Applications,* 34, pp. 1126-1132.

[Che10] Cheng，C.H.； Chen, T.L.; Wei, L.Y.; 2010, A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting, *Information Sciences,* 180, pp.1619-1629.

[Chu09] Chu, H.H.; Chen, T.L.; Cheng, C.H.; Huang, C.C.; 2009, Fuzzy dual-factor time-series for stock index forecasting, *Expert Systems with Applications,* 36, pp. 165-171.

[Dan97] J. Danielsson. and C.G. de Vries; 1997, Tail Index and Quantile Estimation with very High Frequency data, *Journal of Empirical Finance,* 4, pp. 241-257.

[Dac01] Dacorogna, M.M.; Gencay, R.; Muller, U.A.; Olsen, R.B.; Pictet, O.V.; 2001, *An Introduction to High Frequency Finance*, Academic Press, San Diego.

[Dac96] Dacorogna, M.M.; Gauvreau, C.L.; Muller, U.A.; Olsen, R.B.; Pictet, O.V.; 1996, Changing Time Scale for Short-term Forecasting in Financial Markets, *Journal of Forecasting*, 15, pp. 203-227.

[Deb92] Debauchies, I.; 1992, Ten Lectures on Wavel0ets, *Society for Industrial and Applied Mathematics,* Capital City Press, Montpelier, VT, USA.

[Dec97] Dechow, P.M. and Sloan, R.G.; 1997, Returns to contrarian investment strategies: Tests of naïve expectations hypotheses, *Journal of Financial Economics,* 43(1), pp.3-27.

[Dec01] Dechow, P.M.; Hutton, A.P.; Meulbroek, L.; Sloan, R.S.; 2001, Short-sellers, Fundamental Analysis and Stock Returns, *Journal of Financial Economics,* 61(1), pp.77-106.

[Eng82] Engle, R.F.;1982, Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. in Dation, *Econometrica,* 50, pp. 987-1008.

[Eng00] Robert F. Engle; 2000, The Econometrics of Ultra-High Frequency Data, Econometrica, 68(1), pp. 1-22.

[Eng12] Engle, R.F. and Sokalska, M.E.; 2012, Forecasting Intraday Volatility in the US Equity Market, Multiplicative Component GARCH, Journal of Financial Economics, 10(1), pp. 54-83.

[Esf10] Esfahanipour, A. and Aghamiri, W.; 2010, Adapted Neuro-Fuzzy Inference System on indirect approach TSK fuzzy rule base for stock market analysis, *Expert Systems and Applications,* 37, pp.4742-4748.

[Far89] James L. Farrell, Jr.,; 1989, A Fundamental Forecast Approach to Superior Asset Allocation, *Financial Analysts Journal,* 45(3), pp. 32-37.

[Fam70] Fama, E.F.; 1970, Efficient Capital Markets: A Review of Theory and Empirical Work, *Journal of Finance,* 25(2), pp.383-417.

[Fer00] F. Ferna´ndez-Rodr´ıguez; C. Gonzalez-Martel; S. Sosvilla-Rivero; 2000, On the profitability of technical trading rules based on artificial neural networks: Evidence from the Madrid stock market, *Economics Letters,* 69, pp.89-94.

[Fra10] Zhang, X.F.; 2010, High-Frequency Trading, Stock Volatility, and Price Discovery.

[Gar12] Garcke, J.; Gerstner, T.; Griebel M.; 2012, Intraday Foreign Exchange Rate Forecasting Using Sparse Grids, Sparse Grids and Applications, pp.81-105.

[Gem98] Gemcay, R.; 1998, Optimization of Technical Trading Strategies and Profitability in Security Markets, *Economics Letters,*59, pp.249-254.

[Gen98] Gencay, R.; 1998, The predictability of security returns with simple technical trading rules, *Journal of Empirical Finance*, 5, pp.347-359.

[Gra78] Granger, C.W.J. and Anderson, A.P.; 1978, An Introduction to Bilinear Time Series Models, Vandenhoeck and Ruprecht, GYottingen .

[Had10] Hadavandi, E.; Shavandi, H.; Ghanbari, A.; 2010, Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting, *Knowledge-Based Systems,* 23, pp.800-808.

[Hol92] Holthausen, R.W. and Larcker, D.F.; 1992, The Prediction of Stock Returns Using Financial Statement Information, *Journal of Accounting and Economics,* 15, pp.373-411.

[He08] He, W.W.; Wang, Z.Z.; J, H.; 2008, Model optimizing and feature selecting for support vector regression in time series forecasting, *Neurocomputing*, 72, pp. 600-611.

[Ire13] Irene, A.; 2013, *High Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems,* John Wiley & Sons, Inc. Hoboken, New Jersey.

[Inc06] Ince, H.; Trafalis, T.B.; 2006, A hybrid model for exchange rate prediction, *Decision Support Systems,* 42, pp.1054-1062.

[Jos12] Jose M.M. and Juan C.R.; 2012, Forecasting Performance of

Nonlinear Models for Intraday Stock Returns, *Journal of Forecasting,* 31, pp. 172-188.

[Kir15] Kirilenko, Andrei A., et al. "The flash crash: The impact of high frequency trading on an electronic market." Available at SSRN 1686004 (2015).

[Kha11] Khashei, M. and Bijar, M. i; 2011, A novel hybridization of artificial neural networks and ARIMA models for time series forecasting, *Applied Soft Computing,* 11, pp. 2664-2675.

[Khe09] Khemchandani, R.; Jayadeva; Chandra, S.; 2009, Regularized least squares fuzzy support vector regression for financial time series forecasting, *Expert Systems with Applications*, 36, pp. 132-138.

[Li06] Li, J.; Shi, Z.; Li, X.L.; 2006, Genetic Programming with Wavelet-based indicators for Financial Forecasting, The Institute of Measurement and Control.

[Lu09] Lu, C.J.; Lee, T.S.; Chiu, C.C.; 2009, Financial time series forecasting using independent component analysis and support vector regression, *Decision Support System,* 47, pp. 115-125.

[Mar03] Martens, M. and Zein, M.; 2003, Predicting Financial Volatility: High Frequency Time Series Forecast Vis-À-Vis Implied Volatility, *The Journal of Futures Markets,* 24(11), pp. 1005-1028.

[Mas08] Masset, P.; 2008, Analysis of Financial Time-Series using Fourier and Wavelet Methods.

[Mel73] Melvin C. O'Connor; 1973, On the Usefulness of Financial Ratios to Investors in Common Stock, *The Accounting Review,* 48(2), pp.339-352.

[Men13] Menkveld, Albert J. "High frequency trading and the new market makers." Journal of Financial Markets 16.4 (2013): 712-740.

[Nee03] Neerly, C.J. and Weller, P.A.; 2003, Intraday Technical Trading in the Foreign Exchange Market, *Journal of International Money and Finance,*22, pp. 223-237.

[Ni09] Ni, H. and Yin, H.; 2009, Exchange rate prediction using hybrid neural networks and trading indicators, *Neurocomputing,* 72, pp.2815-2823.

[Ou89] Ou, J.A. and Penman, S.H.; 1989, Financial Statement Analysis and the Prediction of Stock Returns, *Journal of Accounting and Economics,* 11, pp.295-329.

[Pai05] Pai, P.F. and Lin, C.S.; 2005, A hybrid ARIMA and support vector machines model in stock price forecasting, *Omega*, 33, pp. 497-505.

[Per00] Percival, D. and Walden, A.; 2000, *Wavelet Methods for Time Series Analysis,* Cambridge University Press, UK.

[Pes94] Pesaran, M.H. and Timmermann A.; 1994, Forecasting Stock Returns: An Examination of Stock Market Trading in the Presence of Transaction Costs, *Journal of Forecasting,* 13, pp. 335-367.

[Reb10] Reboredo, J.C.; Matias, J.M.; Raquel Garcia-Rubio; 2010, Nonlinearity in Forecasting of High Frequency Stock Returns, *Computational Economics*, 40(3), pp. 245-264.

[Rom12] Roman, K. and Wing, W.T.;2012, Execution Risk in High Frequency Arbitrage, *Management Science*, 58(11), pp. 2131-2149.

[Yan12] Yan, R. and Li, H.D.; 2012, Modeling and Forecasting the Intraday Volume of Shanghai Security Composite Index, *2012 International Conference on Systems and Informatics.*

[Ren02] Renaud, O., Murtagh, F., & Starck, J. L. (2002). Wavelet-based forecasting of short and long memory time series. Université de Genève/Faculté des sciences économiques et sociales.

[Ren16] Ren S. Miller and Gary Shorter; 2016 High Frequency Trading: Overview of Recent Development, Congressional Research Service.

[Ram96] Ramsey, J.B.; 1996, The Contribution of Wavelets to the Analysis of Economic and Financial Data.

[San97] Sandip Mukherji; Manjeet S.Dhatt; Yong H.Kim; 1997, A Fundamental Analysis of Korean Stock Returns, *Financial Analysts Journal,* 53(3), pp. 75-80.

[Set98] Section, B. and Strong, N.; 1998, Predicting Stock Returns Using Financial Statement Information, *Journal of Business Finance & Accounting,* 25, pp. 631-656.

[She10] Shephard, N. and Sheppard, K.; 2010, Realizing the Future: Forecasting with High Frequency based Volatility Models, *Journal of Applied Econometrics,* 25, pp. 197-231.

[Sch02] Schleicher, C.; 2002, An Introduction to Wavelets for Economists, Working paper of Bank of Canada.

[Son93] Song, Q. and Chissom, B.S.; 1993 Forecasting enrollments with fuzzy time-series. Part I, *Fuzzy Sets Syst*, 54, pp. 1-10.

[Son94] Song, Q. and Chissom, B.S.; 1993 Forecasting enrollments with fuzzy time-series. Part II, *Fuzzy Sets Syst*, 54, pp. 1-8.

[Sto10] Stolojescu, C.; Railean, I.; Moga, S.; Isar, A.;Lenca, P.;2010, A Wavelet-Based Prediction Method for Time Series, Working Paper.

[Str01] Struzik, Z.R.; 2001, Wavelet Methods in (Financial) Time-series

Processing, *Physica A* ,Vol. 296, pp. 307-319.

[Tan09] Tan, C.; 2009, Financial Time Series Forecasting Using Improve Wavelet Neural Network, Master Thesis.

[Tan09] Tang, L.B. and Sheng, H.Y.; 2009, Wavelet Kernel Function for Stock Index Forecast, 2009 Second International Symposium on Electronic Commerce and Security.

[Tay10] Taylor, J.W.; 2010, Exponentially Weighted Methods for Forecasting Intraday Series with Multiple Seasonal Cycles, International Journal of Forecasting, 26, pp. 627-646.

[Ton83] Tong, H.; 1983, Threshold Models in Non-Linear Time Series Analysis, Springer-Verlag, New York.

[Tse01] Tseng, F.M.; Tzeng, G.H.; Yu, H.C.; Yuan, B.J.C.;2001, Fuzzy ARIMA model for forecasting the foreign exchange market, *Fuzzy Sets and Systems,* 118, pp. 9-19.

[Wal02] Walnut, D.; 2002, *An Introduction to Wavelet Analysis,* Birkhauser Boston, MA, USA.

[Wal99] Walker, J.; 1999, *A Primer on Wavelet and their Scientific Applications,* CRC Press, Boca Raton, FL, USA.

[Wei13] Wei, L.Y.; 2013, A hybrid model based on ANFIS and adaptive expectation genetic algorithm forecast TAIEX, *Economic Modelling*, 33, pp. 893-899.

[Yan09] Yang, H.Q.; Huang, K.Z.; King, I.; Lyu, M.R.;2009, Localized support vector regression for time series prediction, *Neurocomputing,*72, pp.2659-2669.

[Yu00] Yu, I.K.; Kim, C.I.; Song, Y.H.; 2000, A Novel Short-Term Load Forecasting Technique Using Wavelet Transform Analysis, *Electric Machines & Power Systems*, Vol. 28, pp. 537-549.

[Yu05] Yu, H.K.; 2005, Weighted fuzzy time series models for TAIEX forecasting, *Physica A*, 349, pp. 609-624.

[Yao00] Yao, J. and Tan ,C.L.; 2000, A case study on using Neural networks to perform technical forecasting of Forex, *Neurocomputing, 34, pp.79-98.*

[Yu09] Yu, L.; Wang, S.Y.; Lai, K.K.; 2009, A neural-network-based nonlinear metamodeling approach to financial time series forecasting, *Applied Soft Computing,* 9, pp.563-574.

[Zha01] Zhang, B.L.; Coggins, R.;Jabri, M.A.; Dersch, D.; Flower, B.; 2001, Multi-resolution Forecasting for Futures Trading Using Wavelet Decompositions, *IEEE Transactions on Neural Networks,* Vol. 12, No.4, pp.765-775.

[Zha03] Zhang, G. P.; 2003, Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing, 50, pp. 159-175.

[Zha98] Zhang, G.; Patuwo, B.E.; Hu, M.Y.; 1998, Forecasting with artificial neural networks: the state of the art, *International Journal of Forecasting,* 14, pp.35-62.