Stock Index Prediction Based on Grey Theory, ARIMA Model and Wavelet Methods

Zhao Yang Wu

A Thesis

in

the Department

of

Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Science

Concordia University

Montreal, Quebec, Canada

January, 2010

Your file Votre référence
ISBN: 978-0-494-80154-3
Our file Notre référence
ISBN: 978-0-494-80154-3

Abstract

Stock Index Prediction Based on Grey Theory, ARIMA Model and Wavelet Methods

Zhaoyang Wu

In this thesis, we develop a new forecasting method by merging traditional statistical methods with innovational non-statistical theories for the purpose of improving prediction accuracy of stock time series. The method is based on a novel hybrid model which combines the grey model, the ARIMA model and wavelet methods. First of all, we improve the traditional GM(1, 1) model to the GM(1, 1, μ, v) model by introducing two parameters: the grey coefficient μ and the grey dimension degree v. Then we revise the normal G-ARMA model by merging the ARMA model with the GM(1, 1, μ, v) model. In order to overcome the drawback of directly modeling original stock time series, we introduce wavelet methods into the revised G-ARMA model and name this new hybrid model WG-ARMA model. Finally, we obtain the WPG-ARMA model by replacing the wavelet transform with the wavelet packet decomposition. To keep consistency, all the proposed models are merged into a single model by estimating parameters simultaneously based on the total absolute error (TAE) criterion. To verify prediction performance of the models, we present case studies for the models based on the leading Canadian stock index: S&P/TSX Composite Index on the daily bases. The experimental results give the rank of predictive ability in terms of the TAE, MPAE

and DIR metrics as following: WPG-ARMA, WG-ARMA, G-ARMA, GM(1, 1, μ, v), ARIMA.

Acknowledgements

I would like to thank the Department of Mathematics and Statistics in Concordia University for providing such a perfect place to study. In particular, I would like to thank my supervisor, Dr.Wei Sun, for all his teaching, help, encouragement and patience over the last two and half years. Many of my teachers have inspired and enthused me throughout my education, none more so than Drs.Yogendra P. Chaubey, Xiaowen Zhou, Arusharka Sen, Lea Popovic and Pawel Gora. Thanks are due to all my professors, classmates and friends for their support and encouragement. Finally, I would like to thank my parents and my older sister for their patience and constant encouragement throughout this thesis.

Contents

List of Tables ix

List of Figures xi

List of Acronyms xii

1 Introduction 1

1.1 Choosing Models 2

1.2 Organization of the Study 4

2 Review of the Grey Model, the ARIMA Model and Wavelet Methods 7

2.1 ARIMA Modeling 8

2.1.1 Identify the Model and Find the Order of (p, q) 8

2.1.1.1 Autoregressive (AR) Process 8

2.1.1.2 Property of the ACF in the AR Model 9

2.1.1.3 Property of the PACF in AR Model 9

2.1.1.4 Moving Average (MA) Process 10

2.1.1.5 Property of the ACF in the MA Model 10

2.1.1.6 Property of the PACF in the MA Model 10

2.1.1.7 ARMA(p, q) Processes 11

2.1.2 Estimating in an ARIMA Model 12

2.1.3 Testing an ARMA Model 12

2.1.3.1 Significant Test for the Model 12

2.1.3.2 Significant Test for Parameters 13

2.1.4 Prediction of the ARMA Model 13

2.2 Grey Theory 13

2.2.1 Grey Modeling 14

2.2.2 Grey Generation 15

2.2.3 GM(1, 1) Algorithms 15

2.3 Wavelet Transform 16

2.3.1 Fourier Transform 16

2.3.2 Wavelet Transform 17

2.3.2.1 Definition of Wavelet 17

2.3.2.2 Continuous Wavelet Transform (CWT) 17

2.3.2.3 Discrete Wavelet Transform 18

2.3.3 Multiresolution Analysis 18

2.3.4 Deriving Wavelets from MRA 19

2.3.5 Wavelet Decomposition and Reconstruction 20

2.4 Wavelet Packets Transform 22

2.4.1 Wavelet Packets 23

2.4.2 Decomposition and Reconstruction Based on Wavelet Packets . 25

3 Choosing Metrics, Data and Software **27**

3.1 Choosing Metrics 27

3.2 Software 29

3.3 Data 31

4 Forecasting Based on the GM(1, 1, μ, v) Model 33

4.1 Background of the GM(1, 1, μ, v) Model 33

4.2 Constructing a GM(1, 1, μ, v) Model 35

4.3 Case Study: Prediction Using the GM(1, 1, μ, v) Model 36

4.4 Conclusions 39

5 ARIMA Model in Analyzing Stock Movement 42

5.1 Model Identification 42

5.2 Parameters Estimation and Diagnostic Checking 44

5.3 Forecasting with Constructed Model 44

5.4 Conclusions 45

6 Prediction with the Revised G-ARMA Model 49

6.1 Revised G-ARMA Model 49

6.2 Case Study: Prediction Using the Revised G-ARMA Model 51

6.3 Conclusions 53

7 Forecasting Stock Index Based on the WG-ARMA Model 55

7.1 WG-ARMA Model 55

7.2 A Case Study Using the WG-ARMA Model 57

7.3 Conclusions 60

8 Stock Index Forecasting with the WPG-ARMA Model 62

8.1 WPG-ARMA Model 62

8.2 Case Study: Forecasting with the WPG-ARMA Model 64

8.3 Conclusions 68

9 Conclusions **70**

9.1 Contributions 71

9.2 Future Works 71

9.2.1 Using Other Grey Models 72

9.2.2 Replacing the Linear ARIMA Model 72

9.2.3 Introducing Additional Wavelet Functions 72

References **73**

List of Tables

2.1 Criteria of Identifying the Model Based on the ACF and PACF 11

3.1 Descriptive Statistics of the S&P/TSX Index 31

4.1 TAE with Different Pairs of Parameters (μ, v) 37

4.2 S&P/TSX Data from 2008/12/22 to 2008/12/31 37

4.3 Prediction Results for 26 Daily Data Based on the GM(1, 1, μ, v) Model 39

4.4 Comparison with Different Grey Dimension Degree in terms of MPAE . 40

5.1 Autocorrelation Check for White Noise 43

5.2 Minimum Information Criterion 43

5.3 Conditional Lease Squares Estimations 44

5.4 One Step Forecasting Results 45

5.5 Prediction Results for 26 Daily Data Based on the ARIMA Model . . . 46

5.6 Comparison between Models in Terms of Three Metrics 47

5.7 Comparison Between BIC, AICC and TAE in Terms of MPAE 48

6.1 Prediction Results for 26 Daily Data Based on the revised G-ARMA Model 53

6.2 Comparison Between Models in Terms of the Three Metrics 54

7.1 The Serial Number of Wavelet Functions 58

7.2 Prediction Results for 26 Daily Data Based on the WG-ARMA Model . 60

7.3 Comparison Between Models in Terms of Three Metrics 61

8.1 The Serial Number of Related Wavelet Functions 66

8.2 Prediction Results for 26 Daily Data Based on the WPG-ARMA Model 68

8.3 Comparison Between Models in Terms of Three Metrics 69

List of Figures

2.1 Space Structure of MRA 19

2.2 Multiresolution Scheme and Orthogonal Projections 21

2.3 Decomposition Algorithm 22

2.4 Reconstruction Algorithm 22

2.5 MRA Space Decomposition 24

2.6 Wavelet Packets Space Decomposition 24

3.1 Plot of S&P/TSX Index 32

4.1 Comparison Between Actual and Predicted Value Based on the GM(1, 1, μ, v) Model 38

5.1 Comparison between Actual and Predicted Values Based on the ARIMA Model 45

6.1 Comparison Between Actual and Predicted Values Based on the G_ARMA Model 52

7.1 The Decomposition of Original Time Series X with the Wavelet Function 'Rbio3.3' 59

7.2 Comparison Between Actual and Predicted Values Based on the WG-ARMA Model 59

8.1 Decomposition Trees Table 65

8.2 The Decomposition of the Time Series X with the Wavelet Function 'Rbio3.3' 67

8.3 Comparison Between Actual and Predicted Values Based on the WPG-ARMA Model 69

List of Acronyms

ARIMA	Autoregressive Integrated Moving Average
ARCH	Autoregressive Conditional Heteroscedasticity
ARFIMA	Autoregressive Fractionally Integrated Moving Average
EMH	Efficient Market Hypothesis
DIR	Directional Errors
MPAE	Mean Absolute Percent Error
$GM(n, h)$	Grey Differential Equation Model With nth Order and h Parameters
$GM(1, 1)$	Grey Differential Equation Model With 1st Order and 1 Parameter
GM	Grey Modeling
(TAE)	Total Absolute Error Criterion
(AIC)	Akaike's Information Criterion
(BIC)	BIC is another criterion to correct the overfitting nature of the AIC
(AICC)	Akaike's Information Corrected Criterion
TAE	Total Absolute Error
PACF	Partial Autocorrelation function
ACF	Autocorrelation Function
ARMA	Mixed Autoregressive Mode and Moving Average Model

MA	Moving Average Model
AR	Autoregressive Model
WPG – ARMA	Hybrid Model with Wavelet Packets, Grey and ARMA model
WG – ARMA	Hybrid Model with Wavelet, Grey and ARMA model
G – ARMA	Hybrid Model with Grey and ARMA Model
S&P/TSX	TSX Composite Index
TSX	Toronto Stock Exchange
VERHULST	Grey Verhulst Model
DGM	Grey DGM Model
WP	Wavelet Packets
MRA	Multiresolution Analysis
FT	Fourier Transform
AGO	Accumulated Generating Operation

Chapter 1

Introduction

Mainly, there are two groups in stock market prediction research. One group believes that they can devise mechanisms to forecast the stock market. The other group believes the Efficient Market Hypothesis (EMH) [1], which generally claims that securities markets are extremely efficient in reflecting information, so when the information arises, the news spreads very quickly and is incorporated into the price of securities without delay. The EMH is associated with the idea of "random walk", which claims that tomorrow's price changes will reflect only tomorrow's news and will be independent of the price changes today. Therefore, it is impossible to consistently outperform the market by using information that is available on the market.

However, getting started from this century, many financial economists and statisticians began to believe that stock price was at least partially predictable. A new branch of economists emphasized psychological and behavioral element of stock-price determination, and they claimed that the future stock price was somewhat predictable based on the historical stock price.

The thesis assumes that stock price is partially predictable, and its goal is to predict stock returns based on historical data by using statistical and non-statistical models. Of course, it is not possible to generate a very accurate prediction for stock returns, but we can forecast at least, for example, the sign of tomorrow's returns.

1.1 Choosing Models

Choosing proper forecasting models is very important for forecasting accuracy. According to Makridakis et al. [2], each method is different in terms of accuracy, scope, time horizon and cost. To facilitate an adequate level of forecasting accuracy, developers have to determine if a particular method is appropriate for the studied object before applying it to real application. In the following, we will review some main points about forecasting models, study the main characteristics of stock market and then determine suitable models for predicting stock time series.

For forecasting models, we consider two things. One is statistical models vs. non-statistical models. The other is hybrid models vs. single models. Prediction models, because of the difference of modeling theories, can be divided into two types: the traditional prediction models based on the statistical theories and the innovational prediction models based on the non-statistical theories. The Autoregressive fractionally integrated moving average (ARFIMA) and the autoregressive conditional heteroscedasticity (ARCH) are both examples of the traditional prediction models, while the innovational prediction models include genetic algorithm, artificial neural networks, grey theory, wavelet methods, support vector machine etc. Although the traditional prediction methods were successfully used in time series in some researches [3,4,5], the nature of the stock market suggests that the innovational prediction models are more appropriate for forecasting stock time series. In fact, many researchers preferred the non-statistical methods when they analyzed time series [6,7]. However, some researchers reported that there was not clear evidence to show that the non-statistical methods could outperform the statistical models.

Some researchers [8,9] also reported that there was no such single forecasting method that gave an appropriate result in all situations. With the intention to improve forecasting accuracy, the combination of forecasting approaches were proposed by many researchers [10,11,12]. From their studies, they indicated that hybrid models outperformed single forecasting methods.

Now let us study the main characteristics of the stock market. First of all, the stock market can be treated as a huge system with many sub systems, and each sub system has its own characteristics. According to Peters [13], the financial markets including the stock market are consisted of investors in different time horizon, such as long-term investors, middle-term investors and short-term investors; and each group of investors has different investment habits. Therefore, it is reasonable to study the stock stochastic process by dividing it into several parts according to the time horizon.

The second feature of the stock market is that it is full of speculation, which makes well-known methods ineffective. In fact, if investors can find an even small statistical rule, they will immediately take advantage of it. Especially nowadays, sophisticated computer systems and complicated statistical models are widely used to scan the stock market for the purpose of finding any profitable opportunity. It tells us that if a method becomes well-known in the market, the market will adjust itself to make the technique ineffective sooner or later. In order to find an effective way to forecast the stock time series, we need to create a method which is not familiar yet in the market.

Third, the stock market is a system whose information is not completely known. This feature indicates that the stock market is a grey system and can be be analyzed by grey theory.

Finally, the stock market is a kind of system with not only statistical but also non-statistical behaviors. Therefore, it is better to model it with statistical and non-statistical methods together.

The above analysis suggests that we need to create new hybrid models by combining statistical and non-statistical models to improve forecasting accuracy for stock time series. It also help us narrow down theories to the grey model, the ARIMA model and wavelet methods due to their advantages to handle stock time series. In the following, we will give brief explanations of those theories used in stock prediction, and the details will be discussed in the next chapters.

The grey model is a forecasting method based on grey theory, which is a typical non-statistical theory and is regarded as a good model dealing with systems with partially known information. For investors, the stock market is a system where they may know some information, but there is also some information they do not know. So the stock market can be treated as a grey system. In fact, some researchers already used grey theory to study stock movements and claimed that the results were acceptable [14,15,16]. In the thesis, we will also adopt the grey model as one of our forecasting methods.

The ARIMA model is a classical method of processing time series with a solid statistical foundation. Since the sequence of stock price can be treated as a time series, some researchers think it a good idea to model the stock movement using the ARIMA model. They believe that historical price and stock trend have some effects on future movement of stock, and this relationship can be measured by the statistical correlation. For example, Poterba and Summers used the AR(1) model to study S&P 500 index [4]. French et al. specified the ARIMA (0, 1, 3) model to describe the volatility of S&P 500 index [5]. Zhu et al. used the ARIMA model to predict the Shanghai composite index [3]. In the thesis, we will use the ARIMA model to help us extract information based on the statistical correlation. We also believe that investors in the same time horizon have similar investment habits, and those habits can be observed from the statistical correlation.

Wavelet methods are our last theory and they have been successfully applied to analyze time series [17,18,19] in recent years. Some researchers already used wavelet methods to study the stock market [20,21,22]. Their main power comes from their ability to extract signal information from a non-stationary signal that is commonly found in a real-life signal and to decompose a complex signal into several relatively simple signals in different time horizon, Those special characteristics make wavelet theory a perfect candidate for our research, since stock time series can be treated as a super complicated signal combined with many sub-signals in different time horizon. It would make the analysis easier if we can study the stock time series after decomposing it.

1.2 Organization of the Study

This thesis is divided into 9 chapters. Here we present a brief outline for each chapter.

In the current chapter, we investigate the methods that have been used so far to predict the stock market, which lead us to focus on hybrid models based on the grey model, the ARIMA model and wavelet methods. This chapter also gives an outline for the whole thesis. Chapter 2 briefly reviews the basic concepts of the grey model, the ARIMA model and the wavelet and wavelet packets transforms that will be used in this thesis. Chapter 3 is concerned with choosing metrics, data and software.

In Chapter 4, we propose a revised model called the GM(1, 1, μ, v) model by introducing two new parameters (μ ,v) into the traditional GM(1, 1) model. We also present a method to find the optimal parameters (μ ,v) based on the total absolute error (TAE) criterion. Finally, we conduct a case study by using this new model and evaluate it through the three metrics introduced in Chapter 3.

In Chapter 5, an analysis using the traditional ARIMA model is presented. This analysis applies the (BIC) criterion to determine the optimum lags for the ARMA model with the same data used in Chapter 4. It shows that there only exists slight difference among the (BIC), (AICC), and (TAE) criteria for choosing the optimum lags of the ARMA model in terms of the MPAE metric. This conclusion gives us confidence to merge the ARMA model with other theories based on the (TAE) criterion.

In Chapter 6, we improve the G-ARMA model by merging the GM(1, 1, μ, v) model with the ARMA model. Different from the normal G-ARMA model, we construct the G-ARMA model by evaluating parameters simultaneously based on the (TAE) criterion. We then compare this revised G-ARMA model with a single GM(1, 1, μ, v) model and a singe ARIMA model through a case study based on the same data.

In Chapter 7, we propose a new hybrid model called the WG-ARMA model by introducing the wavelet transform into the revised G-ARMA model for the purpose of overcoming the drawback of directly modeling the raw data. The same technique of evaluating all the parameters will be applied in the effort to merge those three different theories into a single model. An analysis is applied to compare the GM(1, 1, μ, v) model, the revised G-ARMA model and the WG-ARMA model in terms of the predictive ability.

In Chapter 8, we replace the wavelet transform with the wavelet packet transform in order to make our model more flexible and more powerful. To keep consistency, we still use the (TAE) criterion to evaluate all the parameters, including a new parameter ρ, which represents the wavelet packet decomposition tree. At the end of the chapter, we conduct an analysis to empirically prove that the WPG-ARMA model has the lowest prediction error in terms of the three metrics introduced in Chapter 3.

Chapter 9 summarizes the conclusions that we have drawn as well as the contributions of the thesis. Finally, we present some suggestions for the future work.

In this thesis, most of our computations are performed by MATLAB R2008b [23]. Some calculations are done with SAS 9.1 [24] and ITSM2000 [25].

Chapter 2

Review of the Grey Model, the ARIMA Model and Wavelet Methods

This chapter attempts to give a brief review of the theories and concepts that would be used in the thesis.

The first theory is the ARIMA model, which is the most commonly used time series analysis method with a solid statistical foundation. The characteristics of the ARIMA model enable us to track information according to the correlations in a stock time series.

The second one is the grey model, which is a forecasting technique based on grey theory. Grey theory has been widely used in China since the last 2 decades. It is especially useful in dealing with a system with partially known and partially unknown information and gives a different view of looking at stock time series.

The wavelet transform and the wavelet packets transform are our last techniques. They have been widely used for signal compression, signal denoising and image compression. They are also useful mathematical tools for analyzing time series and are complementary to the existing analysis techniques (e.g., correlation and spectral analysis). This study will take advantage of the power that wavelets are well adapted to changes or singularity commonly found in real-life stock time series.

2.1 ARIMA Modeling

The ARIMA(p, d, q) model is the most general class of model for forecasting time series. It is designed to incorporate the autoregressive and moving average components, as well as the differencing transformation, where p represents the autoregressive components, d represents the differencing factor, and q represents the moving average components.

If a time series is not stationary, we cannot apply the stationary AR, MA, or ARMA process directly. One possible way of handling non-stationary time series is to apply the differencing technique to make it stationary. An ARIMA(p, d, q) model is just an ARMA(p, q) model after differencing d times. In the following, we will only briefly introduce the ARIMA method. For more information, we refer the reader the books [26-30].

Usually we follow four steps to build up a proper ARMA model

1. Identify the model and find the order of (p, q).
2. Estimate the parameters in the model.
3. Test the model.
4. Predict the time series with the constructed model.

2.1.1 Identify the Model and Find the Order of (p, q)

The autocorrelation function and the partial autocorrelation function, which are denoted by ACF and PACF, play an important role in identifying the ARMA model.

2.1.1.1 Autoregressive (AR) Process

The notation AR(p) refers to autoregressive process of order p, which means that variable X_t can be explained by the previous p order variables. The AR(p) model is defined by

$$\begin{cases} X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t, \\ \phi_p \neq 0, \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t, \\ E(X_s \varepsilon_t) = 0, \forall s < t. \end{cases}$$

2.1.1.2 Property of the ACF in the AR Model

The recursion formula of autocorrelation for stationary AR(p) model is

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}.$$

The autocorrelation function ρ_k can also be represented in terms of Green functions:

$$\rho(k) = \frac{\sum_{j=0}^{\infty} G_j G_{j+k}}{\sum_{j=0}^{\infty} G_j^2}.$$

The Green function is used to describe the impact of ε_t for the system. For example, the AR(1) model

$$X_t - \phi_1 X_{t-1} = \varepsilon_t$$

can also be written as

$$X_t = \sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j} = \sum_{j=0}^{\infty} G_j \varepsilon_{t-j},$$

where $G_j = \phi_1^j$ and is called Green function. Thus, for a stationary AR(p) model, if $|G_j| < 1$ and $\sum_{j=0}^{\infty} G_j < \infty$, the sequence of ρ_k decays exponentially. This property of ρ_k can help us identify an AR(p) model.

2.1.1.3 Property of the PACF in AR Model

Assuming that a time series is an AR(p) series, the p order autocorrelation is a measure of x_{t-k}'s influence by eliminating the variables $x_{t-1}, x_{t-2}, \cdots, x_{t-k+1}$ in the middle. The PACF is defined by

$$\alpha(0) = 1, \quad \alpha(k) = \phi_{kk}, \ k \geq 1.$$

where ϕ satisfies the following Yule-Walker equation

$$\begin{cases} \rho_1 = \phi_{k1}\rho_0 + \phi_{k2}\rho_1 + \cdots + \phi_{kk}\rho_{k-1}, \\ \rho_2 = \phi_{k1}\rho_1 + \phi_{k2}\rho_0 + \cdots + \phi_{kk}\rho_{k-2}, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots, \\ \rho_k = \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + \cdots + \phi_{kk}\rho_0. \end{cases}$$

If the real model is given by an AR(p) model, then we have $\alpha(p) = \phi_p$ and $\alpha(k) = 0$ for $k > p$. This is the property of p order cut off for PACF, which provides another way to identify an AR(p) model.

2.1.1.4 Moving Average (MA) Process

The notation MA(q) refers to the moving average model of order q. It has the following structure:

$$\begin{cases} X_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q}, \\ \theta_q \neq 0, \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t\varepsilon_s) = 0, s \neq t. \end{cases}$$

2.1.1.5 Property of the ACF in the MA Model

It can be shown that the ACF of an MA(q) process satisfies

$$\rho_k = \begin{cases} 1, & k = 0, \\ \frac{-\theta_k + \sum_{i=1}^{q-k} \theta_i\theta_{k+i}}{1+\theta_1^2+\cdots+\theta_q^2}, & 1 \leq k \leq q, \\ 0, & k > q. \end{cases}$$

The ACF equals zero for lags greater than q. Since $\theta_q \neq 0$, the ACF is not equal to zero at lag q. Thus, the ACF can be used to find the order of the process by considering the lags of the sample ACF which are significantly different from zero.

2.1.1.6 Property of the PACF in the MA Model

The definition of the PACF in a MA model is the same as a AR model and is still denoted by ϕ_{kk}. It can be shown that the PACF of an MA(q) model decays exponentially.

2.1.1.7 ARMA(p, q) Processes

A mixed AR(p) model and MA(q) model is abbreviated by ARMA(p, q) and is defined by

$$\begin{cases} X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}, \\ \phi_p \neq 0, \theta_q \neq 0, \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t, \\ E(X_s \varepsilon_t) = 0, \forall s < t. \end{cases}$$

By using the Green functions, we can get its ACF

$$\rho(k) = \frac{\sum_{j=0}^{\infty} G_j G_{j+k}}{\sum_{j=0}^{\infty} G_j^2}.$$

We can see that both ACF and PACF decay exponentially in an ARMA model. To sum up, we can use the graphs of the sample ACF and sample PACF to initially identify an ARMA model. The rule is given in Table 2.1.

Model	**Sample ACF**	**Sample PACF**
AR(p)	Decay exponentially	Cut off after lag p
MA(q)	Cut off after lag q	Decay exponentially
ARMA(p, q)	Decay exponentially	Decay exponentially

Table 2.1: Criteria of Identifying the Model Based on the ACF and PACF

Except identifying an ARMA model by using the ACF and PACF, we can also use more rigorous model selection criteria to find the optimal order of (p, q), such as the (AICC) and (BIC) criteria [31]. The (AICC) criterion is a bias-corrected version of the information of Akaike (1973) and is defined by

$$AICC(\beta) = -2 \ln L_x(\beta, S_x(\beta)/n) + 2(p+q+1)n/(n-p-q-2).$$

The (BIC) criterion is defined to be

$$BIC = (n-p-q)In\left[\frac{n\widehat{\sigma}^2}{(n-p-q)}\right] + n(1+\ln\sqrt{2\pi}) + (p+q)\ln\left[\frac{\left(\sum_{t=1}^{n} X_t^2 - n\widehat{\sigma}_2\right)}{(p+q)}\right].$$

In order to identify an ARMA(p, q) model, we first set upper bounds P and Q for the AR and MA models respectively. Then we fit all possible ARMA(p, q) models for $0 \le p \le P$ and $0 \le q \le Q$ by using the same sample size. The best models have

$$\min_{p\le P, q\le Q} AICC(p,q)$$

or

$$\min_{p\le P, q\le Q} BIC(p,q).$$

2.1.2 Estimating in an ARIMA Model

After determining the order of an ARMA(p, q) model, we need to estimate parameters. In the thesis, we use the lease square method to estimate parameters. For the ARMA (p, q) model, we denote parameters by

$$\beta = (\phi_1, ..., \phi_p, \theta_1, ..., \theta_q)^T.$$

Then the lease-square estimation satisfies

$$Q(\widehat{\beta}) = \min\{Q(\beta)\},$$

where

$$Q(\beta) = \sum_{t=\max(p,q)}^{N} (X - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} - \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q})^2.$$

We may calculate it by using the iteration algorithm with the help of computers.

2.1.3 Testing an ARMA Model

2.1.3.1 Significant Test for the Model

If the residual error is white noise, the established model has tracked all the information from the sample; otherwise, it still has some information left in the sample without tracking out, which means that the fitted model is not efficient enough. We may use the following statistic to test whether the residual error sequence is white noise

$$LB = N(N+2)\sum_{k=1}^{m} \frac{\widehat{\rho_k^2}}{N-k} \sim \chi^2(m).$$

2.1.3.2 Significant Test for Parameters

The purpose of testing parameters is to eliminate insignificant parameters in the final model. For a single parameter b_j, the significance can be tested by the following statistic:

$$t = \frac{b_j}{S(b_j)} \sim t(N-k),$$

where $S(b_j)$ is the standard deviation of parameter b_j and k is the number of parameters in the model.

2.1.4 Prediction of the ARMA Model

For stationary time series, its L steps forecasting can be calculated by its conditional expectation

$$\widehat{X}_t(l) = E(X_{t+1}\,|X_t, X_{t-1}, ...).$$

Here we only consider one step forecasting.

(1) For an AR (p) series, its one step prediction is

$$\widehat{X}_t(1) = E(X_{t+1}\,|X_t, X_{t-1}, ...) = \phi_1 X_t + \phi_1 X_{t-1}, ..., +\phi_p X_{t-p}.$$

(2) For an MA (q) series, its one step prediction is

$$\widehat{X}_t(1) = -\sum_{i=1}^{q} \theta_i \varepsilon_{t+1-i} \text{ for } q \geq 1.$$

(3) For an ARMA(p, q) series, since it can always be transferred into the form of

$$X_t = \sum_{j=1}^{\infty} I_j X_{t-j} + \varepsilon_t,$$

its one step prediction is

$$\widehat{X}_t(1) = \sum_{j=1}^{\infty} I_j X_{t+1-j}.$$

2.2 Grey Theory

Grey theory was first introduced by Professor Ju Long Deng in 1982 [32]. In grey theory, an information system can be classified into three categories: the white system, the grey system and the black system. If the information system is fully unknown, it is called the black system. On the contrary, if a system is completely known, it is called the white system. When a system is known between the white and the black systems, it is called the grey system [33].

In grey theory, each stochastic process is regarded as a grey quantity taking values on a certain range or changing on a certain range of time. Hence each stochastic process can be treated as a grey process [34]. We refer the reader to [35-37] for more details about grey theory.

2.2.1 Grey Modeling

In this thesis, we only focus on the grey model (GM). The GM is one of the main methods based on grey theory and has been widely used in modeling and forecasting agriculture, environment, and even stock indices for more than two decades [38 39]. It adopts the core of grey theory, and the advantages of grey modeling as following: (a) It can be applied in circumstances with relatively few data and in the system where the information is not clearly known. (b) It models a system by a differential equation and makes it easier to find the inherent law of a system. We can see that the grey model provides a brand new way to analyze a process where the probability distribution is not required. Those features make it an ideal tool in analyzing stock time series without knowing its statistical distribution.

The full name of the GM is the grey differential equation model. It is denoted by GM(n, h), where the parameter n represents the n-th order differential equation and the parameter h means h variables. So GM(1, 1) denotes the first order differential equation with one variable.

The grey model believes that the accumulation and release of energies usually satisfy an exponential pattern. To catch the pattern, the GM uses the grey differential equation to dynamically model small-sample discrete data series. The grey differential equation is actually a non-conventional differential equation. However, they use the same form. The GM assumes that the pattern of the discrete data series to be processed is exponential or can be transformed to an exponential pattern by some form of data pre-processing called grey generation.

2.2.2 Grey Generation

We assume that the original data series is denoted by

$$X^0 = \left(x^0(1), x^0(2), x^0(3), ..., x^0(n)\right).$$

In order to ensure the condition of exponential pattern, we should apply an operation called accumulated generating operation (AGO) to the original data series.

After applying the first-order AGO, the series X^0 becomes

$$X^1 = \left(x^1(1), x^1(2), x^1(3), ..., x^1(n)\right),$$

where

$$x^1(k) = x^0(1) + x^0(2) + \cdots + x^0(k).$$

The AGO can be carried out one or several times until the data series is suitable for building up the grey differential equation.

2.2.3 GM(1, 1) Algorithms

Given a non-negative time series X^0, after applying the first-order AGO, we get a new time series X^1. The time series X^1 is then fitted by a first-order differential equation given by

$$\frac{dx^1}{dt} + ax^1 = b, \tag{2.1}$$

where the parameter a and b are called the development coefficient and control variable respectively.

According to grey theory, the whitening of grey derivatives for discrete data with a unit time interval ($\Delta t = 1$) is given by

$$\frac{dX^1(t)}{dt} = x^1(k) - x^1(k-1) = x^0(k). \tag{2.2}$$

A new variable $z^1(k)$, which is known as the whitening value of $x^1(k)$, is defined by

$$z^1(k) = 0.5x^1(k) + 0.5x^1(k-1). \tag{2.3}$$

By substituting equations (2.2) and (2.3) into equation (2.1), the differential equation can be transformed into the discrete difference form:

$$x^0(k) + az^1(k) = b. \tag{2.4}$$

Then we can determine the grey parameters by using the lease square method

$$\widehat{\delta} = \left(B^T B\right)^{-1} B^T Y,$$

where

$$\widehat{\delta} = [\hat{a}, \hat{b}]^T; \quad Y = \begin{pmatrix} x^0(2) \\ x^0(3) \\ \cdots\cdots \\ x^0(n) \end{pmatrix}; \quad B = \begin{pmatrix} -z^1(2) & 1 \\ -z^1(3) & 1 \\ \cdots\cdots & \cdots \\ -z^1(n) & 1 \end{pmatrix}$$

and $\hat{a}, \hat{b}$ are identified as the whitening values of the grey parameters a, b. Now, we can use the estimated parameters and equation (2.1), (2.4) to get the following formula

$$\widehat{x}^1(k+1) = (x^1(1) - \frac{b}{a})e^{-ak} + \frac{b}{a}; k = 1, 2, \ldots, n.$$

Finally, by applying the inverse AGO, the predicted value of the raw data $x^0(k+1)$ is given by

$$\widehat{x}^0(k+1) = \widehat{x}^1(k+1) - \widehat{x}^1(k) = (1 - e^a) \cdot \left(x^1(1) - \frac{b}{a}\right) e^{-at}; k = 1, 2, \ldots, n. \tag{2.5}$$

The parameters of the GM(1, 1) model can be updated immediately when new data are obtained.

2.3 Wavelet Transform

The wavelet transform is a mathematical method to represent a function by using some special functions called wavelets. We refer the reader to [40-46] for more details.

2.3.1 Fourier Transform

Before introducing the wavelet transform, we will first review the Fourier transform (FT). Given a continuous function $f(t) \in L^2$, its FT is

$$F(w) = \int_{-\infty}^{+\infty} f(t)e^{-iwt}dt,$$

where $F(w)$ is called the Fourier coefficient. Given $F(w)$, $f(t)$ can be obtained by the inverse FT

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F(w)e^{iwt}dw.$$

After applying the FT, we can study the frequency content of a signal, which gives us a powerful tool to analyze the time series. However, the FT does not show how frequencies vary with time in the spectrum. Nevertheless, time-varying frequencies are quite common phenomena in the stock market. To investigate such phenomena, we need a transform that enables us to obtain the frequency content of a process locally in time.

2.3.2 Wavelet Transform

Unlike the Fourier transform, whose basis functions are sine and cosine, the wavelet transform is based on small waves, called wavelets. A signal can be represented by linear combination of wavelets. Wavelets are localized in time and are good building block functions for a variety of signals, including signals with features which change over time and signals which have jumps and other non-smooth features.

2.3.2.1 Definition of Wavelet

Given a function $\psi(t)$ satisfying the admissibility condition

$$C_\psi = \int_R \frac{\left|\widehat{\psi}(w)\right|^2}{|w|} dw < \infty, \tag{2.6}$$

where $\widehat{\psi}(w)$ is its Fourier transform, we call $\psi(t)$ a wavelet function. By scaling and translating the wavelet function $\psi(t)$, we can get a sequence of wavelets

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi(\frac{t-b}{a}); \quad (a, b \in R, a \neq 0),$$

where a is a scaling parameter which means the degree of scale, and b is a translation parameter which determines the location of wavelets.

2.3.2.2 Continuous Wavelet Transform (CWT)

Based on wavelets, we can define the wavelet transform. Let $f(t) \in L^2$. Its wavelet transform is defined by

$$W_f(a, b) =< f, \psi_{a,b} >= \frac{1}{\sqrt{a}} \int_R f(t)\overline{\psi(\frac{t-b}{a})}dt.$$

If the admissibility condition (2.6) is satisfied, then f can be reconstructed as

$$f(t) = \frac{1}{C_\psi} \int_{R^+} \int_R \frac{1}{a^2} W_f(a, b)\psi(\frac{t-b}{a})dadb.$$

We see that the wavelet transform can analyze a signal from the time zone and frequency zone at the same time, and can also measure a signal in detail by scaling its size.

2.3.2.3 Discrete Wavelet Transform

The obtained wavelet coefficients by CWT are highly redundant. We would like to remove the redundancy. The discrete wavelets can be obtained from the corresponding continuous version of wavelets by discretizing the parameters $a = a_0^j$; $b = ka_0^j b_0$; $j, k \in \mathbb{Z}$,

$$\psi_{a_0^j, kb_0}(t) = a_0^{-j/2}\psi(a_0^{-j}t - kb_0).$$

The discrete wavelet transform using discrete wavelets is defined as

$$W_f(a_0^j, kb_0) = \int_{-\infty}^{\infty} f(t)\overline{\psi_{a_0^j, kb_0}(t)}dt.$$

For computational efficiency, $a_0 = 2$ and $b_0 = 1$ are commonly used, which leads to the dyadic wavelets

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k), j, k \in \mathbb{Z}.$$

Then, the related wavelet transform is

$$W_f(j, k) = \int_{-\infty}^{\infty} f(t)\overline{\psi_{j,k}(t)}dt.$$

2.3.3 Multiresolution Analysis

The concept of multiresolution analysis (MRA) was initiated by Meyer [47] and Mallat [48]. It is formulated based on the study of orthonormal, compactly supported wavelet bases, which provides a natural framework for the understanding of wavelet bases and plays an important role in the formation of discrete wavelet transform.

An (MRA) is a nested sequence of closed subspaces

$$V_j \subset L^2(R), j \in \mathbb{Z}, \cdots V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset V_3$$

such that

1) $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$ and $\bigcup_{j\in\mathbb{Z}} V_j$ is dense in $L^2(R)$.

2)

$$\text{For any } f \in L^2(R) \text{ and any } j \in \mathbb{Z},\ f(x) \in V_j \text{ iff } f(2x) \in V_{j-1}. \tag{2.7}$$

3)

$$\text{For any } f \in L^2(R) \text{ and any } k \in \mathbb{Z},\ f(x) \in V_0 \text{ iff } f(x-k) \in V_0.$$

4) There exists a scaling function $\varphi \in V_0$ whose integer translations $\{\varphi_{0,k} : k \in \mathbb{Z}\}$ constitute an orthonormal basis of V_0.

2.3.4 Deriving Wavelets from MRA

The detail space W_j is defined to be the orthogonal complement of V_j in V_{j+1} (see Figure 2.1). Thus

$$V_{j+1} = V_j \oplus W_j,$$

where $V_j \perp W_j$ and $W_m \perp W_n, m \neq n$. By MRA, we can get

$$L^2(R) = \underset{j\in\mathbb{Z}}{\oplus} W_j. \tag{2.8}$$

Here V_j is called the scaling space and W_j is called the detail space.

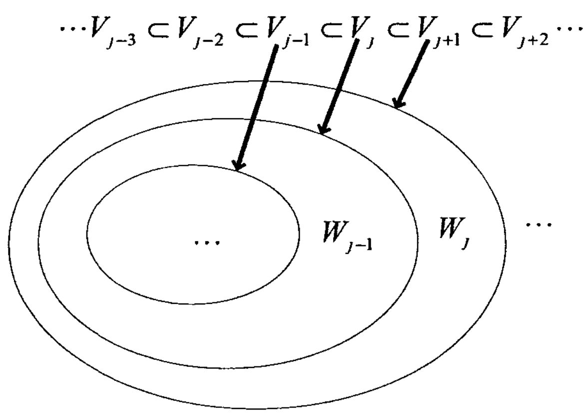


Figure 2.1: Space Structure of MRA

Thus, $L^2(R)$ may be decomposed into mutually orthogonal subspaces. Additionally, W_j inherits the scale-relating property (2.7) of V_j. In other words,

$$f(t) \in W_j \Leftrightarrow f(2t) \in W_{j+1}. \tag{2.9}$$

Therefore, if a function ψ is such that its integer translations forms an orthonormal basis of W_0, then the relations given in (2.8) and (2.9) ensure that

$$\left\{\psi_{j,k} : \psi_{j,k}(t) = 2^{j/2}\psi_{j,k}(2^{j/2}t - k)\right\}_{j,k\in\mathbb{Z}}$$

form an orthonormal basis of $L^2(R)$.

To derive a wavelet function ψ from the scaling function, we note that as $W_0 \in V_1$, hence $\psi(t) \in V_1$ and $\psi(t)$ may be represented as

$$\psi(t) = \sum_{k\in\mathbb{Z}} g_k \sqrt{2}\varphi(2t - k)$$

for some coefficients $\{g_k\}_{k\in\mathbb{Z}}$.

2.3.5 Wavelet Decomposition and Reconstruction

In order to perform the discrete wavelet transform more efficiently, Mallat provided a scheme called Mallat algorithm, which connects wavelets with MRA in order to efficiently perform discrete, wavelets-based transformation.

Let V_j be generated by some scaling function $\varphi \in L^2(R)$, and W_j be generated by some wavelet function $\psi \in L^2(R)$. According to MRA, we can always decompose V_j into detail space W_{j-1} and scaling space V_{j-1}, then decompose V_{j-1} into V_{j-2} and W_{j-2}... to get

$$\begin{aligned} V_j &= V_{j-1} \oplus W_{j-1} \\ &= V_{j-2} \oplus W_{j-2} \oplus W_{j-1} \\ &= \cdots \\ &= V_{j-m} \oplus W_1 \oplus W_2 \oplus \cdots \oplus W_{j-m}. \end{aligned} \tag{2.10}$$

Let $P_j f$ and $Q_j f$ be the orthogonal projections of the function on V_j and W_j respectively. From (2.10), the function $f(t)$ has a unique decomposition

$$
\begin{aligned}
f(t) \approx P_j f &= P_{j-1} f + Q_{j-1} f \\
&= P_{j-2} f + Q_{j-1} f + Q_{j-2} f \\
&= \cdots\cdots \\
&= P_{j-m} f + Q_{j-1} f \cdots\cdots + Q_{j-m} f,
\end{aligned}
\tag{2.11}
$$

where $P_j f$ is also called an approximation, and $Q_j f$ is also called a detail of function respectively at resolution level j. Figure (2.2) shows the multiresolution scheme and the related orthogonal projections.

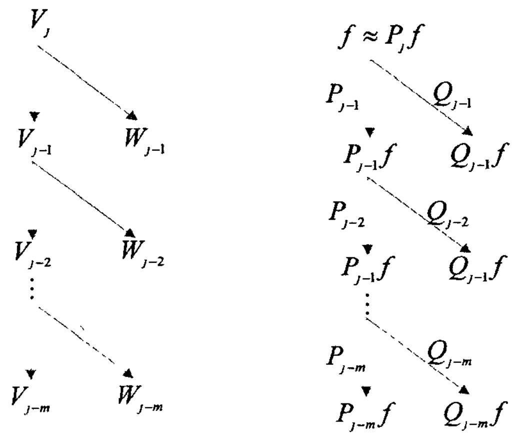


Figure 2.2: Multiresolution Scheme and Orthogonal Projections

Since both $\varphi(2t)$ and $\varphi(2t-1)$ are in V_1 and $V_1 = V_0 + W_0$, there exist four sequences $\{a_{-2k}\}, \{b_{-2k}\}, \{a_{1-2k}\}, \{b_{1-2k}\}; k \in \mathbb{Z}$, such that

$$\varphi(2t) = \sum_k \left[a_{-2k}\varphi(t-k) + b_{-2k}\psi(t-k)\right], \tag{2.12}$$

$$\varphi(2t-1) = \sum_k \left[a_{1-2k}\varphi(t-k) + b_{1-2k}\psi(t-k)\right] \tag{2.13}$$

for all $t \in R$. We can get the following formula by combining (2.12) and (2.13)

$$\varphi(2t-l) = \sum_k \left[a_{l-2k}\varphi(t-k) + b_{l-2k}\psi(t-k)\right], l \in \mathbb{Z}, \tag{2.14}$$

which is called the "decomposition relation" of the scaling function φ and the wavelet function ψ. The sequence of pairs $(\{a_k\}, \{b_k\})$ is called the decomposition sequence.

Since $V_j = \underset{k}{span}\{\varphi_{j,k}(t)\}$ and $W_j = \underset{k}{span}\{\psi_{j,k}(t)\}$, for $P_j f \in V_j$ and $Q_j f \in W_j$, we have a unique series representation

$$P_j f(t) = \sum_k c_k^j \varphi(2^j t - k), \tag{2.15}$$

$$Q_j f(t) = \sum_k d_k^j \psi(2^j t - k). \tag{2.16}$$

By applying (2.14), (2.15), (2.16), we can get the decomposition algorithm (see equation (2.17) and Figure (2.3)).

$$\begin{cases} c_k^{j-1} = \sum_l a_l - 2kc_l^j, \\ d_k^{j-1} = \sum_l b_l - 2kc_l^j. \end{cases} \tag{2.17}$$

$$\begin{tikzcd}
\cdots\cdots \arrow[r] & \{c_k^{j+1}\} \arrow[r] \arrow[d] & \{c_k^j\} \arrow[r] \arrow[d] & \{c_k^{j-1}\} \arrow[r] \arrow[d] & \cdots\cdots \\
& \{d_k^{j+1}\} \arrow[r] & \{d_k^j\} \arrow[r] & \{d_k^{j-1}\} \arrow[r] & \cdots\cdots
\end{tikzcd}$$

Figure 2.3: Decomposition Algorithm

Since both the scaling function $\varphi \in V_0$ and the wavelet function $\psi \in W_0$ are in V_1, and V_1 is generated by $\varphi_{j-1,k}(t) = 2^{1/2}\varphi(2t-k); k \in \mathbb{Z}$, there exist two sequences $\{p_k\}$ and $\{q_k\}$ such that

$$\varphi(t) = \sum_k p_k \varphi(2t - k), \tag{2.18}$$

$$\psi(t) = \sum_k q_k \varphi(2t - k) \tag{2.19}$$

for all $t \in \mathbb{R}$. The formulas (2.18) and (2.19) are called the "two-scale relation" of the scaling function and the wavelet function respectively.

By applying (2.15) (2.16) (2.18) (2.19), we get the reconstruction algorithm (see equation (2.20) and Figure (2.4)).

$$c_k^j = \sum_l \left[p_{k-2l} c_l^{j-1} + q_{k-2l} d_l^{j-1} \right]. \tag{2.20}$$

$$\begin{tikzcd}
\cdots\cdots \arrow[r] & \left\{c_k^{j-1}\right\} \arrow[r] & \left\{c_k^{j}\right\} \arrow[r] & \left\{c_k^{j+1}\right\} \arrow[r] & \cdots\cdots \\
\cdots\cdots \arrow[ur] & \left\{d_k^{j-1}\right\} \arrow[ur] & \left\{d_k^{j}\right\} \arrow[ur] & \left\{d_k^{j+1}\right\} & \cdots\cdots
\end{tikzcd}$$

Figure 2.4: Reconstruction Algorithm

2.4 Wavelet Packets Transform

In this section, we will give a brief explanation of wavelet packets as well as the decomposition and reconstruction algorithms based on wavelet packets. Wavelet packets (WP) were introduced by Coifman, Meyer, Quaker and Wickerhauser [49,50], which are a generalization of the wavelet decomposition. They offer a rich range of possibilities for signal analysis. In the wavelet analysis, the signal is split into an approximation and a detail. The approximation is then itself split into a second-level approximation and detail, and the process is repeated. In the wavelet packets analysis, the details as well as the approximations can be split repeatedly, so wavelet packets offer more flexibility. In particular, wavelet packets are better at representing signals that exhibit oscillatory or periodic behavior.

2.4.1 Wavelet Packets

Recall that V_j is the scaling space and W_j is the wavelet space. If we let

$$\begin{cases} U_j^0 = V_j \\ U_j^1 = W_j \end{cases}, \qquad j \in \mathbb{Z}.$$

Then $V_{j+1} = V_j + W_j$ can be represented as

$$U_{j+1}^0 = U_j^0 + U_j^1, \qquad j \in \mathbb{Z}. \tag{2.21}$$

Let $w_n(t) \in U_j^n$ and $w_{2n}(t) \in U_j^{2n}$ and $w_n(t)$ satisfy the following two-scale function

$$\begin{cases} w_{2n}(t) = \sqrt{2} \sum\limits_k h(k) w_n(2t - k), \\ w_{2n+1}(t) = \sqrt{2} \sum\limits_k g(k) w_n(2t - k), \end{cases} \tag{2.22}$$

where $\{h_k\}_{k \in Z} \in l^2, \{g_k\}_{k \in Z} \in l^2$. When $n = 0$, the above functions become

$$\begin{cases} w_0(t) = \sqrt{2} \sum_k h(k) w_0(2t - k), \\ w_1(t) = \sqrt{2} \sum_k g(k) w_0(2t - k). \end{cases} \tag{2.23}$$

We can see that $w_0(t)$, $w_1(t)$ are just a scaling function and a wavelet function.

Since (2.23) is equivalent to the representation (2.21), we get the equivalent representation of (2.22) as

$$U_{j+1}^n = U_j^{2n} + U_j^{2n+1}, \qquad j \in \mathbb{Z}, n \in \mathbb{Z}.$$

We can see that the wavelet packets space decomposition is different from the MRA space decomposition. We refer the reader to Figure (2.5) and Figure (2.6) for the difference.

To further understand the advantage of wavelet packets decomposition, we may look at Figure (2.6). For the space V_0, we can choose subspace $V_{31} \sim V_{34}$ and $W_{31} \sim W_{34}$

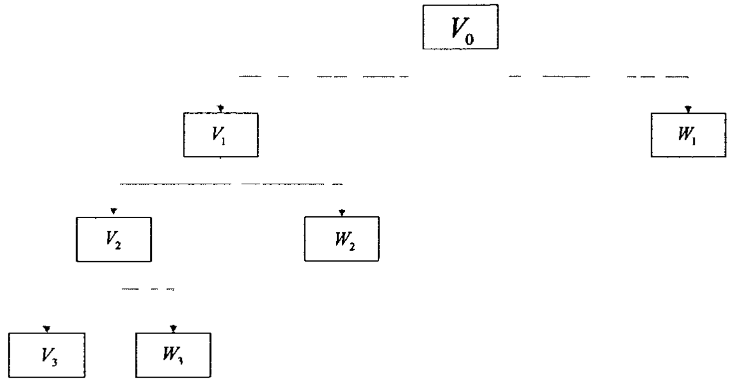


Figure 2.5: MRA Space Decomposition

to cover it, so the wavelet packets base is $\{w_0^8(t), w_1^8(t), \cdots, w_7^8(t)\}$. Alternatively, we can choose subspaces $V_{31}, W_{31}, W_{21}, W_{11}$ to cover V_0 and the wavelet packets base is $\{w_0^8(t), w_1^8(t), w_1^4(t), w_1^2(t)\}$. Thus, the the wavelet packets transform is more powerful than the wavelet transform when the proper wavelet packets tree best fits the characteristics of a signal is chosen.

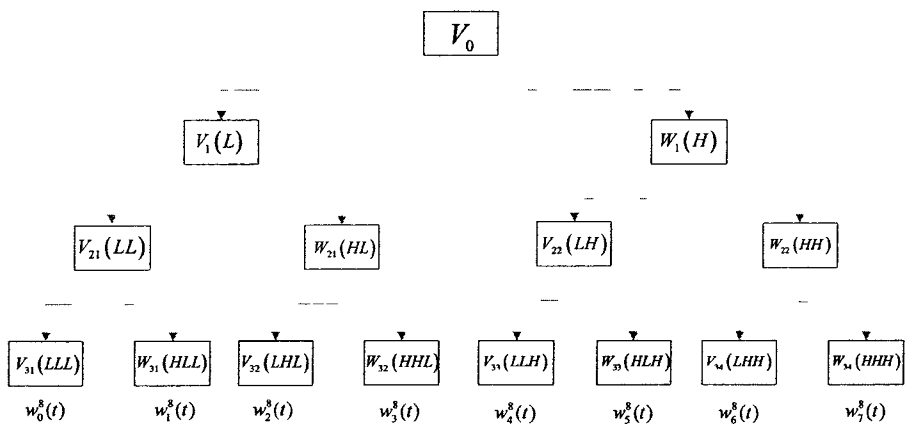


Figure 2.6: Wavelet Packets Space Decomposition

2.4.2 Decomposition and Reconstruction Based on Wavelet Packets

According to MRA, $L^2(R) = \otimes_{j\in\mathbb{Z}} W_j$, we can write W^n_{j+1} as U^n_{j+1}, so the decomposition for the wavelet space W^n_{j+1} is the following

$$W^n_{j+1} = U^n_{j+1} = U^{2n}_j \oplus U^{2n+1}_{j+1}.$$

By iteration,

$$\begin{aligned} W_j &= U^1_j = U^2_{j-1} \oplus U^3_{j-1}, \\ U^2_{j-1} &= U^4_{j-1} \oplus U^5_{j-1}, \\ U^3_{j-1} &= U^6_{j-1} \oplus U^7_{j-1} \cdots\cdots \end{aligned}$$

Thus, we get the general form for decomposing the wavelet space W_j as

$$\begin{aligned} W_j &= U^2_{j-1} \oplus U^3_{j-1}, \\ W_j &= U^4_{j-1} \oplus U^5_{j-1} \oplus U^6_{j-1} \oplus U^7_{j-1}, \\ &\cdots\cdots \\ W_j &= U^{2j}_0 \oplus U^{2j+1}_0 \oplus \cdots \oplus U^{2j+l-1}_0. \end{aligned}$$

We may denote this kind of subspace W_j as

$$U^{2^l+m}_{j-1}, m = 0, 1, \cdots, 2^l; l = 1, 2, \cdots, j; j = 1, 2, \cdots$$

The orthonormal basis for the subspace $U^{2^l+m}_{j-1}$ is

$$2^{-\frac{j-l}{2}} w_{2^l+m}\left(2^{j-l}t - k\right), k \in \mathbb{Z}.$$

We can see that when $l = 0$ and $m = 0$, the subspace $U^{2^l+m}_{j-1}$ becomes $U^1_j = W_j$ and the related orthonormal basis becomes $2^{-\frac{j}{2}} w_1\left(2^{-j}t - k\right) = 2^{-\frac{j}{2}}\psi\left(2^{-j}t - k\right)$, which is just the orthonormal wavelet family $\{\psi_{j,k}(t)\}$.

If we let $n = 2^l + m$, the wavelet packets can be denoted as $\psi_{j,k,n} = 2^{-\frac{j}{2}}\psi_n\left(2^{-j}t - k\right)$, where $\psi_n = 2^{\frac{l}{2}}w_{2^l+m}\left(2^l t\right)$. Compared with the wavelet $\psi_{j,k}(t) = 2^{-\frac{j}{2}}\psi\left(2^{-j}t - k\right)$, we can see that the wavelet packets have one more parameter n, which makes it more flexible than wavelets.

Let $g_j^n(t) \in U_j^n$. Then $\left\{g_j^n\right\}$ may be written as

$$g_j^n(t) = \sum d_l^{j,n} w_n(2^j t - l).$$

Then we can get $\left\{d_l^{j,2n}\right\}$ and $\left\{d_l^{j,2n+1}\right\}$ from $\left\{d_l^{j+1,n}\right\}$ by using the following decomposition algorithm

$$d_l^{j,2n} = \textstyle\sum_k a_{k-2l} d_k^{j+1,n},$$

$$d_l^{j,2n+1} = \textstyle\sum_k b_{k-2l} d_k^{j+1,n}.$$

We can also get $\left\{d_l^{j+1,n}\right\}$ from $\left\{d_l^{j,2n}\right\}$ and $\left\{d_l^{j,2n+1}\right\}$ based on the reconstruction algorithm as the following

$$d_k^{j+1,n} = \sum_k \left[h_{l-2k} d_k^{j,2n} + g_{l-2k} d_k^{j,2n+1}\right].$$

Chapter 3

Choosing Metrics, Data and Software

In this chapter, we are going to discuss about choosing metrics, data and software. These topics are important for the next chapters.

3.1 Choosing Metrics

It is very important to choose proper metrics to measure the performance of a given prediction system. Similar choosing suitable models, in order to find metrics for some specific situations, we need to understand not only the characteristics of the performance metrics but also the features of the studied object. In fact, there are a lot of metrics available for measuring the qualities of prediction methods. In the following, we will first study the stock market in terms of measurement, and then try to find some metrics which are suitable to measure prediction methods while dealing with stock time series.

In the stock market, people usually care about two things when they talk about prediction performance. The first one is the errors made between the actual and the predicted values. The second one is how many mistakes they have made in terms of the direction of movement.

Before finding suitable metrics according to the above features of the stock market, it is better to understand the basic features of a good metric. According to Armstrong and Callopy [51], to choose a proper metric for a certain circumstance, we need to consider several different aspects.

First of all, we need to consider the reliability of the metric, which means that the metric should be consistent for different data series. The second one is sensitivity, which measures the response of performance of the system to the variation of system parameters. We also need to consider the understandability, which means that the metric should be easy to understand and has a straightforward meaning for the situation. The fourth thing worthy of consideration is expense, which becomes critical if we need to measure the system in real time. The validity is our last thing to consider, which tells us that we should choose the metric measuring what is supposed to be measured.

The above analysis helps us finally narrow down the metrics to the total absolute error (TAE) metric, the mean absolute percentage error (MPEA) metric and the direction error (DIR) metric.

In terms of the TAE metric, we can see that it helps us to measure the absolute error between the actual value and the predicted value in terms of point, so it has a very clear financial meaning. It also satisfies all the above characteristics of being a good metric. It is one of our main metrics for verifying prediction performance and will also be used as a criterion due to less calculation cost.

For the MPEA metric, it is almost the most commonly used metric in measuring the magnitude of error. In order to compare with other researchers' work, in the thesis, we will use it as another main metric to measure prediction performance of our prediction models.

The above two metrics can only measure the magnitude of error; however, sometimes we are interested in measuring the direction of movement. The DIR metric will give us a sign about how many mistakes the forecasting model has made during a period of time in terms of direction.

Let x_j be the actual value, $\hat{x}_j$ be the prediction value and n be the total periods. Then the TAE, MPAE and DIR metrics are defined as follows.

1) TAE: Total Absolute Error

$$TAE = \sum_{j=1}^{n} |x_j - \hat{x}_j| .$$

TAE is the total absolute error value over n periods.

2) MAPE: Mean Absolute Percent Error

$$MPAE = \frac{100}{n} \sum_{j=1}^{n} \left| \frac{x_j - \hat{x}_j}{x_j} \right| .$$

MAPE represents the average error with respect to the true value over n periods.

The above two metrics can give the magnitude of the movement, but they are not ideal for decision-making. In order to measure the accuracy of predicting the direction of movement, we will use the following metric.

3) DIR: Directional Errors

$$DIR = \frac{100}{n} \sum_{j=1}^{n} d_j,$$

where

$$d_j = \begin{cases} 0, & (x_j - x_{j-1})(\hat{x}_j - \hat{x}_{j-1}) > 0 \\ 1, & \text{otherwise.} \end{cases}$$

Hence DIR=30% means that the predicted direction is not correct in 30% of the cases.

We point out that if the values of the metrics are lower, then performance of the model is better.

3.2 Software

A lot of software and computing languages can be used to deal with the grey model, the ARIMA model and the wavelet transform. For example, the software developed by Dr. B. Liu is dedicated to the grey models, which can easily calculate the GM(1, 1), the DGM model and the VERHULST model. For the ARIMA model, the statistical tools SAS, S-PLUS [52], SPSS[53] and R [54] can be used since they all have built-in modules to deal with time series. In terms of the wavelet transform, we may use MATLAB or R to decompose the time series. Hence, it won't be a big problem if we analyze stock time series by using the above software separately. However, when we have to merge the grey model, the ARIMA model and the wavelet transform into a single model by estimating all parameters simultaneously, choosing the right computing language will become crucial.

In fact, making a decision about what is the proper software to verify our hybrid models is one of the most challenging tasks in this thesis due to the lack of reference materials, the incompatibility of different software and no ready-made modules.

First of all, after searching the relevant literatures about hybrid models, we find that most researchers use manual ways to verify their hybrid models. Taking the normal G-ARMA model for example, researchers use one software to calculate the grey prediction value, then manually export results to another software for further analyzing. It may work for several G-ARMA models and won't take too much time to do that. However, if we have to verify thousands of G-ARMA models in order to find an optimal one, it will become inconvenient or even impossible. Unfortunately, our literature searching did not give us a satisfactory solution for calculating such kind of hybrid models.

In regard to the incompatibility of different software, it is commonly known that different software is usually developed by different companies and organizations. So we need to develop middle modules to connect them together. For example, we need to use SAS/C++ module [55] to call functions written by C++. The development of middle modules is very time consuming, which usually cannot be completed by only several researchers.

For the built-in modules, at this point, there is no single software which has all built-in modules to deal with the grey model, the ARIMA model and the wavelet transform. That means we have to develop some modules to help us achieve the goal of merging different models into one hybrid model.

After many attempts, we chose to use MATLAB as the computing language with the intention of avoiding the development of middle modules since developing middle modules is too hard to be done during very limited time. Our final strategy is to develop a relatively simple module to deal with the grey model while taking advantage of the power of built-in wavelet and ARIMA modules provided by the MATLAB company. Using this method, finally we get all the modules for the grey model, the ARIMA model and the wavelet transform, and we can write code to verify our hybrid models. Here, we want to especially thank Dr. Sheppard for his MATLAB time series tool box [56]. Although our final programs do not adopt his tool box, it is his tool box that gave us confidence at the beginning to focus on MATLAB. That saved us tons of time.

3.3 Data

The data used in this thesis is a stock index instead of a single stock. The reason is that a stock index has the function of evaluating the average moving of the stock market, which is an indicator of the national economics. Thus, it will be more meaningful to study a stock index.

The stock index data considered in this study are obtained from Yahoo Finance [57]. We are concerned with the Leading Canada Index: The S&P/TSX Composite Index. The S&P/TSX Composite Index is an index of the stock (equity) prices of the largest companies in the Toronto Stock Exchange measured by market capitalization. The Toronto Stock Exchange list companies in this index comprises about 70% of market capitalization for all Canadian-based companies [58].

Our S&P/TSX data consists of 152 daily observations from June 30, 2008 to February 6, 2009. Table (3.1) presents some basic statistics that describe the time series. The values of the time series against time are presented in Figure (3.1).

Statistic	TSX data
Mean	10781.5745
Standard Deviation	173.982593
Median	9779.28
Mode	8724.11
Standard Deviation	2145.00146
Sample Variance	4601031
Kurtosis	-1.6850951
Skewness	0.2705423
Range	6742
Minimun	7724.76
Maximun	14467.0
Count	152

Table 3.1: Descriptive Statistics of the S&P/TSX Index

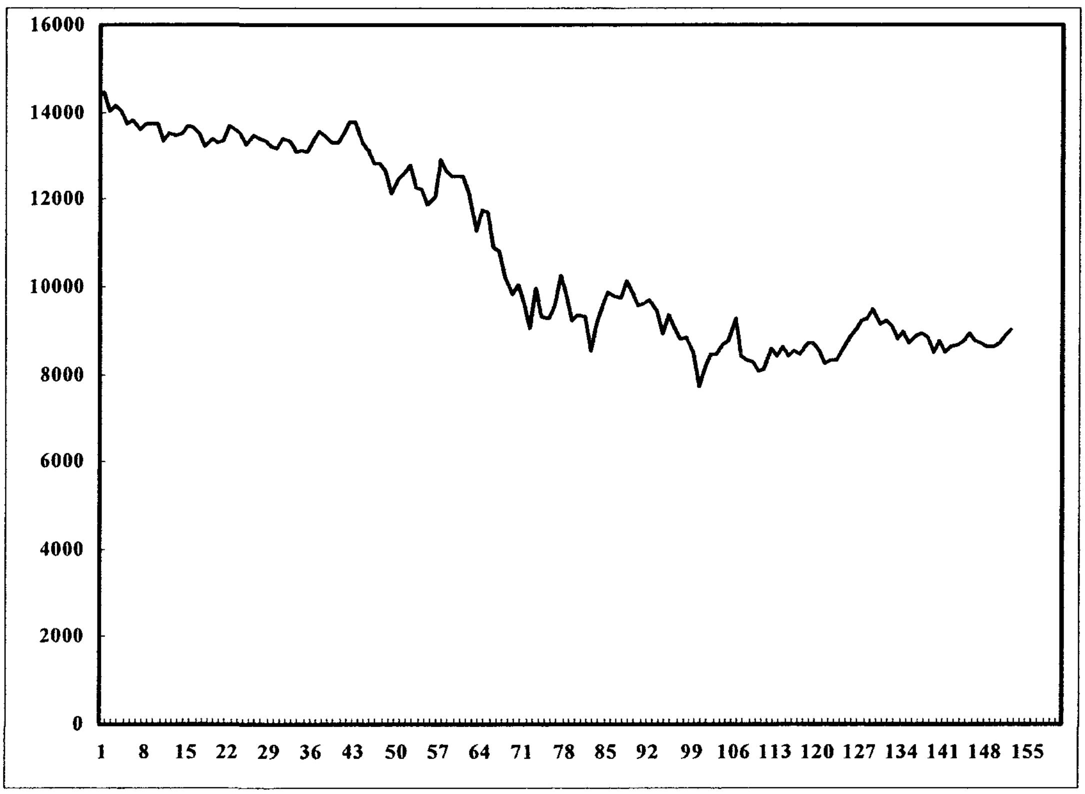


Figure 3.1: Plot of S&P/TSX Index

Chapter 4

Forecasting Based on the GM(1, 1, μ, v) Model

In Chapter 2, we studied the traditional GM(1, 1) model. In this chapter, we will first improve the GM(1, 1) model to the GM(1, 1, μ, v) model by introducing two more parameters. Then we will present the results which empirically show that the GM(1, 1, μ, v) model is better than the GM(1 ,1) model in terms of the MPAE metric. Finally, we conduct a case study based on the S&P/TSX stock index to show how to use the GM(1, 1, μ, v) model to predict time series. The results will be compared with other models in the future.

4.1 Background of the GM(1, 1, μ, v) Model

Recall that the GM(1, 1) model has been known as a good tool to model a system with partial information. Here, we will give additional reasons for choosing the GM model for predicting stock time series.

First of all, the grey model has been considered as a good way to model non-linear and non-stationary systems, which is one of main characteristics of the stock market. Second, in grey theory, it is not necessary to know the details of the system. So we can focus on the data itself, which makes it a good choice for people without much financial knowledge. Last but not least, in the stock market, information usually will not be extracted immediately. On the contrary, it is obtained gradually. This can be called "the function of historical information". In the grey model, we may use a coefficient called "system development coefficient" to catch the driving force.

In fact, some pioneers already used the GM(1, 1) model to study stock time series. For example, Wang et al.[59]used grey prediction methods to estimate the varying parameters in their stock forecasting model. Xu et al.[60]directly used the GM(1, 1) model to predict stock price.

Although the GM(1, 1) model can be used to predict time series, some researchers find that there still have space to improve the GM(1, 1) model by introducing new parameters and optimizing them.

Recall that, in most textbooks [33,36], the variable $z^1(k)$, the whitening value of $x^1(k)$, is often defined as

$$z^1(k) = 0.5x^1(k) + 0.5x^1(k).$$

Some researchers feel that the constant 0.5 may not be optimal, so they refine the variable $z^1(k)$ by replacing the constant 0.5 with a parameter: the grey coefficient μ,

$$z^1(k) = (1 - \mu)x^1(k) + \mu x^1(k + 1)$$

and they call this revised GM(1, 1) model the GM(1, 1, μ) model and propose to use genetic algorithm or particle swarm algorithm to find a proper μ [61 62]. Since both the developing coefficient a and the control variable b are determined by the variable $z^1(k)$, finding a proper μ will optimize the GM(1, 1) model.

Another parameter is the size of the sequence X^0 in the GM(1, 1) model. Traditionally, people just use the whole sample data as X^0 if the whole sample size is small, or arbitrarily choose several last data as X^0 if the whole sample size is big. For stock time series, we may get plenty of sample data and do not want to randomly choose the size of X^0, so we have to figure out a different way to choose the size of X^0. Actually, Hao et al. showed in [63] that the size of X^0 has significant influence on the prediction accuracy, and they claimed that people should count this factor when they tried to set up a GM(1, 1) model. But they did not give a solution to choose a proper size of X^0. In this chapter, we will study this problem and use a parameter named grey dimension degree v to represent the size of sequence X^0.

Until now, there is no published paper giving a method to optimize the two parameters simultaneously. In this thesis, we will try to optimize the two parameters at the same time. To distinguish this grey model from the traditional one, we call it the GM(1, 1, μ, v) model, where the first 1 represents the first order; the second 1 means one variable; the parameters μ and v are defined as above.

4.2 Constructing a GM(1, 1, μ, v) Model

Instead of finding the parameter μ or v separately, we will optimize the grey model by finding the optimal combination of μ and v simultaneously based on the total absolute error (TAE) criterion.

One reason for choosing the (TAE) criterion is that we want to integrate the GM(1, 1, μ, v) model with the ARIMA model and wavelet methods later based on the same criterion. The financial meaning of the (TAE) can be found in Chapter 3.

Generally speaking, in order to identify the GM(1, 1, μ, v) model, first we set $\mu \in (l, L)\,; v \in (r, R)$, and then fit all possible GM(1, 1, μ, v) models for $\mu \in (l, L)\,; v \in (r, R)$ using the same sample size. We claim that the optimal model satisfies the (TAE) criterion

$$\min(TAE); \mu \in (l, L)\,, v \in (r, R)$$

and the parameters $(\widehat{\mu}, \widehat{v})$ with minimum TAE are the optimal parameters and the $GM(1, 1, \widehat{\mu}, \widehat{v})$ model is the optimal GM model.

More specifically, for a given sequence $X = (x_1, \ldots, x_n)$ and fixed parameters $\mu \in (l, L)$; $v \in (r, R)$, we can find its prediction sequence $\widehat{X} = (\widehat{x}_1, \cdots, \widehat{x}_n)$ by using the $GM(1, 1, \mu, v)$ model. Then the (TAE) for a fixed parameter μ, v is

$$TAE = \sum_{k}^{n} |x_k - \widehat{x}_k|\,; k = v + 1, v + 2, ..., n.$$

Here we have to eliminate the first v data since it is used to get the prediction value $\widehat{x}_{v+1}$ from the sequence $X^0 = (x_1, \ldots, x_v)$. By repeating the above steps we can find all the TAEs for $\mu \in (l, L)\,, v \in (r, R)$. The parameters $(\widehat{\mu}, \widehat{v})$ with the lowest TAE are the optimal parameters, and the corresponding GM(1,1,$\widehat{\mu}$, $\widehat{v}$) model is the optimal GM model.

4.3 Case Study: Prediction Using the GM(1, 1, μ, v) Model

In this case study we apply the GM(1, 1, μ, v) model to predict the time series of S&P/TSX Composite Index. Our strategy is to use the first 126 daily data to construct the model, and then use the next 26 daily data to evaluate the model.

Before setting up a GM(1, 1, μ, v) model, we need to find the optimal parameters $(\widehat{\mu}, \widehat{v})$. First of all, we need to set the upper and lower bounds for the parameters (μ, v). The grey coefficient μ should belong to $[0, 1]$. For the grey dimension degree v, we know that the minimum number of data required to build up a GM(1, 1) model is 4, so we set the lower bound for v to be 4. According to [63], the forecasting error will increase significantly as the grey dimension degree is greater than 30. Hence, we set the upper bound for v to be 30. Therefore, the range for v is $[4, 30]$.

Second, we need to discretize the continuous parameters μ. Here we set the interval length as 0.01 since it will not make a big difference in forecasting. Now the parameter space becomes $\mu \in [0, 0.01, \ldots, 1], v \in [4, \cdots, 30]$, which includes a total of 2727 combinations of (μ, v). Our goal is to find the parameters $(\widehat{\mu}, \widehat{v})$ with the minimum TAE based on the same sample data.

Here we choose MATLAB to help us find the parameters $(\widehat{\mu}, \widehat{v})$ with the reason explained in Chapter 3. By using a program written by the computer language of MATLAB, we find that when $(\widehat{\mu}, \widehat{v}) = (1, 6)$, the TAE of the sequence $X = (x_1, \cdots, x_{126})$ reaches its minimum. This tells us that we should use the grey coefficient 1 and the grey dimension degree 6 to build up a GM(1, 1) model and this GM(1,1,1,6) model is the optimal GM(1, 1) model.

	μ				
v	0	0.3	0.5	0.7	1
4	43238	39202	36653	34469	31767
6	33932	32416	31527	30832	29903
9	34513	33819	33413	33006	32487
17	39130	38774	38538	38319	37997
23	42979	42686	42512	42337	42108
30	46026	45987	45960	45937	45908

Table 4.1: TAE with Different Pairs of Parameters (μ, v)

Table (4.1) lists part of the output TAEs with different pairs of parameters (μ, v) based on the sequence $X = (x_1, \ldots, x_{126})$. We can see that the difference is nontrivial. Horizontally, it is just the $GM(1, 1, \mu)$ model. Vertically, it is a traditional GM(1, 1) model when the grey coefficient μ is 0.5. In both cases, the difference cannot be omitted.

Now we will use $(\widehat{\mu}, \widehat{v}) = (1, 6)$ to build up a GM(1, 1) model and to make one step prediction. Since $\widehat{v} = 6$, we will choose the last 6 sample data. Table (4.2) lists the data which we will use to build up a GM(1, 1) model.

Date	12/22	12/23	12/24	12/29	12/30	12/31
Position	121	122	123	123	125	126
Data	8249.53	8311.91	8310.55	8637.29	8830.72	8987.7

Table 4.2: S&P/TSX Data from 2008/12/22 to 2008/12/31

The initial data series is

$$X^0 = \left(x_1^0, x_2^0, \ldots, x_6^0\right) = (8249.53, 8311.91, 8310.55, 8637.29, 8830.72, 8987.7)\,.$$

Step 1: We use the initial data series to perform the Accumulated Generating Operation (AGO) and obtain a new data series

$$X^1 = \left(x_1^1, x_2^1, ..., x_6^1\right) = (8249.53, 16561.44, 24871.99, 33509.28, 42340, 51327.7).$$

Step 2: Since $\mu = 1$, the whitening value of $x^1(k)$ is

$$z_k^1 = (1-\mu)x_k^1 + \mu x_{k+1}^1 = x_{k+1}^1; k = 2, 3, 4, 5, 6.$$

We may calculate the development coefficient a and the control variable b by using the following formula

$$\widehat{\delta} = \left(B^T B\right)^{-1} B^T Y,$$

where

$$Y = \begin{pmatrix} x_2^0 \\ x_3^0 \\ \cdots\cdots \\ x_6^0 \end{pmatrix}; B = \begin{pmatrix} -z_2^1 & 1 \\ -z_3^1 & 1 \\ \cdots\cdots & \cdots \\ -z_6^1 & 1 \end{pmatrix} = \begin{pmatrix} -x_2^1 & 1 \\ -x_3^1 & 1 \\ \cdots\cdots & \cdots \\ -x_6^1 & 1 \end{pmatrix}.$$

By using equation (2.5), we get the discrete time response equation

$$\widehat{x}_{k+1}^0 = (1-e^a).\left(x_1^1 - \frac{b}{a}\right)e^{-at} = (1-e^{-0.021545}).(8249.53 + 366163.28)\,e^{0.021545k}.$$

Let k=6, we get the predicted value $\widehat{x}_7^0 = 9081.85$. Thus, according to the GM(1, 1, 1, 6) model, the closed price of S&P/TSX for January 2, 2009 is predicted to be 9081.85.

To forecast the point 128 for January 3, 2009, we need to find the $(\widehat{\mu}, \widehat{v})$ first based on the new X with the sample size 127, and then use the above steps to find the predicted value of point 128. By repeating the above procedure, we can get all the 26 forecasting points.

Table (4.3) lists the results obtained from the case study and Figure (4.1) gives a visual view about the predicted values and actual values.

Time	Rel.	Abs.	Actual.	Pre.	AE.	Error%	Dir.
2009/1/2	127	1	9234.11	9081.8	152.31	1.649428	0
2009/1/5	128	2	9285.51	9344.3	58.79	0.633137	0
2009/1/6	129	3	9472.09	9417	55.09	0.581603	0
2009/1/7	130	4	9121.32	9554.8	433.48	4.752382	1
2009/1/8	131	5	9221.58	9346.3	124.72	1.35248	0
2009/1/9	132	6	9085.18	9220.7	135.52	1.49166	0
2009/1/12	133	7	8793.33	9075.2	281.87	3.205498	0
2009/1/13	134	8	8961.55	8790.4	171.15	1.909826	1
2009/1/14	135	9	8688.36	8799.5	111.14	1.279183	0
2009/1/15	136	10	8879.61	8677.6	202.01	2.274987	1
2009/1/16	137	11	8920.4	8828.8	91.6	1.02686	1
2009/1/19	138	12	8841.48	8877.1	35.62	0.402874	0
2009/1/20	139	13	8504.93	8932.1	427.17	5.022616	1
2009/1/21	140	14	8757.89	8547.7	210.19	2.400007	0
2009/1/22	141	15	8486.56	8591	104.44	1.230652	0
2009/1/23	142	16	8627.97	8486	141.97	1.645462	1
2009/1/26	143	17	8656.51	8614.9	41.61	0.480679	1
2009/1/27	144	18	8759.63	8600.3	159.33	1.818912	1
2009/1/28	145	19	8906.23	8801.4	104.83	1.177041	0
2009/1/29	146	20	8762.76	8924.9	162.14	1.85033	1
2009/1/30	147	21	8694.9	8864.1	169.2	1.945968	1
2009/2/2	148	22	8624.83	8714	89.17	1.033875	1
2009/2/3	149	23	8628.63	8563.7	64.93	0.752495	1
2009/2/4	150	24	8693.09	8583	110.09	1.266408	1
2009/2/5	151	25	8860.98	8660.1	200.88	2.267018	1
2009/2/6	152	26	9008.02	8856.6	151.42	1.680947	1

Table 4.3: Prediction Results for 26 Daily Data Based on the GM(1, 1, μ, v) Model

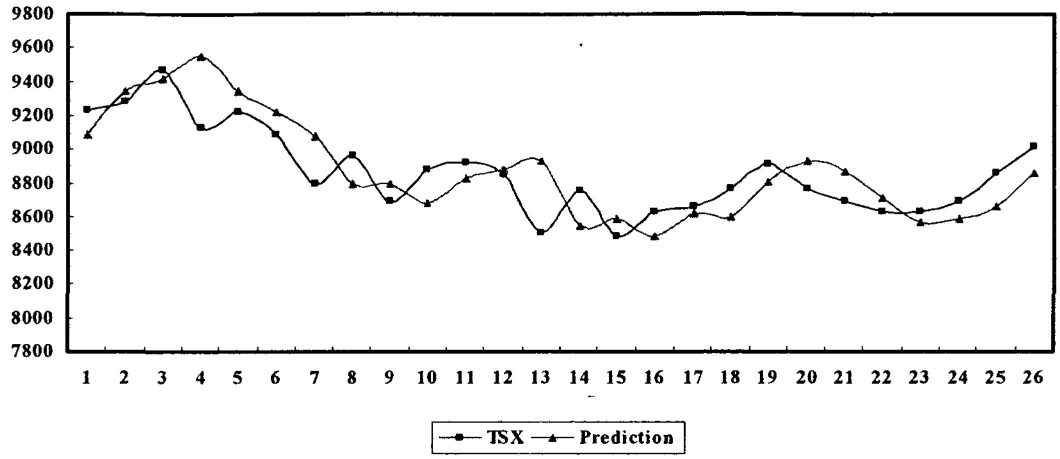


Figure 4.1: Comparison Between Actual and Predicted Value Based on the GM(1. 1, μ, v) Model

4.4 Conclusions

In this chapter, we propose a revised GM(1, 1) model called the GM(1, 1, μ, v) model. In order to verify that the GM(1, 1, μ, v) model is better than the traditional GM(1, 1) model in terms of the MPAE metric, we conduct a case study. Since it is not known which grey dimension degree v is the best one for building up a GM(1, 1) model, we consider all v which are greater than or equal to 4. Table (4.4) lists the MPAEs with different grey dimension degree from 4 to 30 for the data from 127 to 152. We can see that only $v = 5$ gives the lowest MPAE. This result shows that the GM(1,1,μ,v) model is better than the traditional GM(1, 1) model in term of the MPAE metric in most of cases. Also, it tells us that the model constructed with the lowest TAE based on the historical data cannot guarantee that the model gives the lowest MPAE in the future.

Although the revised GM model is better than the traditional one, the result in forecasting stock time series is not good enough. Especially the direction error (DIR)

Degree v	MPAE	Degree v	MPAE
4	1.80618	18	2.98026
5	1.66190	19	3.04270
6	2.01289	20	3.10810
7	2.19572	21	3.17413
8	2.33470	22	3.21093
9	2.42679	23	3.23813
10	2.62775	24	3.27287
11	2.61752	25	3.29390
12	2.63074	26	3.39137
13	2.57180	27	3.41689
14	2.65581	28	3.38728
15	2.80011	29	3.32885
16	2.91992	30	3.28684
17	2.97180		

Table 4.4: Comparison with Different Grey Dimension Degree in terms of MPAE

is more than 50 percent. One reason is that the GM model is not good for time series which oscillate intensively. The other reason is that stock index is more complex than other time series. It not only includes linear characteristics but also non-linear features. It also includes many sub-systems with different features. Some sub-systems have long term memory; some sub-systems oscillate intensively within a very short period time. Therefore, in order to dig the potential of the GM model, we will further introduce the ARIMA model which tracks information for stock time series according to the statistical correlation, as well as the wavelet transform which separates stock time series into long trend part with lower vibration and high frequency part which oscillate intensively. By letting the GM model focus on the lower vibration part, we will get a chance to dig the potential of the GM model.

Chapter 5

ARIMA Model in Analyzing Stock Movement

In this chapter, we will conduct a case study in detail with the ARIMA model. We will also compare the difference among the (AICC), (BIC) and (TAE) criteria for choosing the order of (p, q) in the ARMA model in terms of the MPAE metric. Our conclusions will be made at the end of the chapter.

The case study will closely follow the steps introduced in Chapter 2 by using the statistical software SAS 9.1. The global prediction scheme is the same as we did in the previous chapter, in which we split the collected data into two categories. The training set consists of the first 126 daily indexes (June 30, 2008-December 31, 2008), while the test set includes the remaining 26 daily indexes (January 2, 2009-February 6, 2009).

5.1 Model Identification

From Figure (3.1), it is clear that the S&P/TSX time series is not stationary since its movement has some trend. By using the IDENTIFY statement in SAS, we can see that the times series is non-stationary based on plots of the sample ACF and PACF. The original sequence also does not pass the augmented Dickey-Fuller test and Phillips-Perron test.

Since all those statistical tests and plots show that the time series is non-stationary, we need to transform it into a stationary series by differencing technique. That is, instead of modeling the S&P/TSX time series itself, we will model the differenced series. We denote the differenced time series $\mathrm{Y}_t = \mathrm{X}_\mathrm{t} - X_{t-3}$, where the time series X_t is the original S&P/TSX time series. The new time series Y_t has a constant mean and an approximate constant variance and it also passes the augmented Dickey-Fuller test and Phillips-Perron test with 99% confidence. From the white noise test showed in Table (5.1), we can see that the P value is smaller than 0.001. So it is not white noise and we need to build up an ARMA model for the series Y_t.

To Lag	Chi-Square	DF	Pr > Chi-Square
6	52.39	6	<.0001

Table 5.1: Autocorrelation Check for White Noise

To find an initial ARMA model, we may use plots of the sample PCF and PACF, the (AICC) criterion or (BIC) criterion. Here we will use the (BIC) criterion to find the initial ARMA model. Table (5.2) suggests that we should choose ARMA(0, 2) model as the initial model.

Lags	MA0	MA1	MA2	MA3	MA4	MA5
AR0	12.32329	11.96464	11.7846	11.80611	11.84467	11.87058
AR1	11.92197	11.89981	11.81911	11.83275	11.86778	11.8941
AR2	11.92608	11.93847	11.85397	11.87017	11.90575	11.93058
AR3	11.91593	11.8836	11.88548	11.90867	11.94046	11.96862
AR4	11.91718	11.91701	11.92073	11.93696	11.95845	11.97076

Table 5.2: Minimum Information Criterion

5.2 Parameters Estimation and Diagnostic Checking

After finding the initial order of (p, q), we may estimate parameters of the ARMA(0, 2) model by using the conditional least squares estimation. Table (5.3) shows that the parameters are all significant from zero and gives us the following equation

$$y_t = -174.93238 - 0.97567\varepsilon_{t-2} - 0.89807\varepsilon_{t-1} + \varepsilon_t.$$

The white noise test of this equation shows that this ARMA(0, 2) model is proper and tracks information sufficiently. Finally, we get the formula

$$x_t = x_{t-3} - 174.93238 - 0.97567\varepsilon_{t-2} - 0.89807\varepsilon_{t-1} + \varepsilon_t. \tag{5.1}$$

Parameter	Estimate	Standard Error	t Value	Pr > |t|
MU	-174.93238	84.02983	-2.08	0.0395
MA1,1	-0.97567	0.04017	-24.29	<.0001
MA1,2	-0.89807	0.04050	-22.18	<.0001

Table 5.3: Conditional Lease Squares Estimations

5.3 Forecasting with Constructed Model

Now we use equation (5.1) to do one step forecast. We will use the first 126 daily indexes to forecast the 127th daily index. The result is shown in Table (5.4). Therefore, according to the ARIMA model, the closing price of S&P/TSX for January 2, 2009 is predicted to be 8957.0959.

Obs	Forecast	Std Error	95% Confidence	95% Confidence
127	8957.0959	345.7572	8279.4243	9634.7675

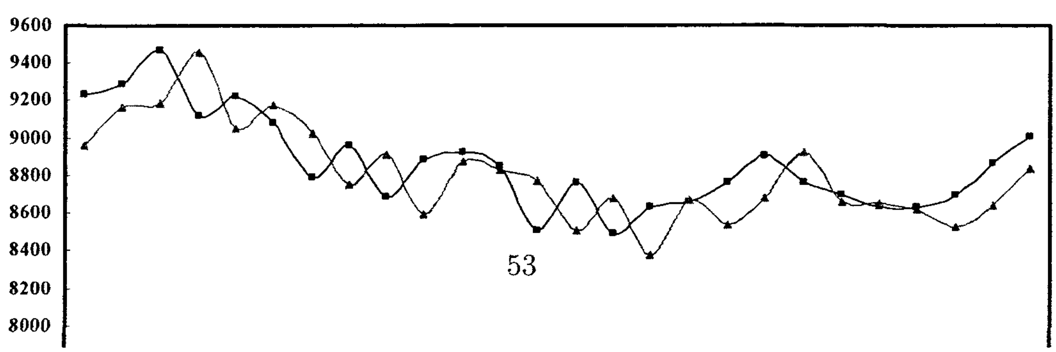
Table 5.4: One Step Forecasting Results

In order to forecast the 128th daily index for January 3, 2009, we will use the first 127 daily indexes and follow the same steps to get the predicted value of the 128th daily index. By repeating those steps, we can finally get all the 26 forecasting prices which are from January 2, 2009 to February 6, 2009.

Table (5.5) lists the results for this case study and Figure (5.1) is the plot of the predicted values and actual values.

Time	Rel.	Abs.	Org.	Pre.	AE.	Error%	Dir.
2009/1/2	127	1	9234.11	8957.096	277.0141	2.9999	1
2009/1/5	128	2	9285.51	9163.867	121.6431	1.310031	1
2009/1/6	129	3	9472.09	9184.672	287.418	3.034367	1
2009/1/7	130	4	9121.32	9457.804	336.4837	3.68898	0
2009/1/8	131	5	9221.58	9050.491	171.0888	1.855309	1
2009/1/9	132	6	9085.18	9172.382	87.2016	0.959822	0
2009/1/12	133	7	8793.33	9023.954	230.6242	2.622717	0
2009/1/13	134	8	8961.55	8752.44	209.1099	2.333412	1
2009/1/14	135	9	8688.36	8914.599	226.2385	2.603926	0
2009/1/15	136	10	8879.61	8593.012	286.5982	3.227599	1
2009/1/16	137	11	8920.4	8870.278	50.1216	0.561876	1
2009/1/19	138	12	8841.48	8826.923	14.5572	0.164647	0
2009/1/20	139	13	8504.93	8774.204	269.2742	3.166095	0
2009/1/21	140	14	8757.89	8506.378	251.5125	2.871839	0
2009/1/22	141	15	8486.56	8677.924	191.3636	2.254902	0
2009/1/23	142	16	8627.97	8378.017	249.9531	2.897009	1
2009/1/26	143	17	8656.51	8664.259	7.7488	0.089514	0
2009/1/27	144	18	8759.63	8538.22	221.4101	2.527619	1
2009/1/28	145	19	8906.23	8674.344	231.8856	2.603634	1
2009/1/29	146	20	8762.76	8920.363	157.6031	1.798555	1
2009/1/30	147	21	8694.9	8658.14	36.7596	0.422772	0
2009/2/2	148	22	8624.83	8643.47	18.6403	0.216124	0
2009/2/3	149	23	8628.63	8620.007	8.6234	0.099939	1
2009/2/4	150	24	8693.09	8528.82	164.2699	1.889661	1
2009/2/5	151	25	8860.98	8634.949	226.031	2.550858	1
2009/2/6	152	26	9008.02	8841.05	166.9705	1.853576	1

Table 5.5: Prediction Results for 26 Daily Data Based on the ARIMA Model



5.4 Conclusions

In this chapter, we describe in detail on how to use the ARIMA model to find the predicted values of stock index. Table (5.6) shows that the GM model gives lower TAE and MPAE while the DIR is the same. It is too early to make a judgment for the GM model and the ARIMA model using only Table (5.6) in terms of the ability of forecasting stock time series. However, we feel that the GM model may be more suitable than the ARIMA model for predicting stock time series due to its special characteristics.

Model	TAE	MPAE	DIR
ARIMA	4500.1446	1.94633%	57.69%
GM(1, 1, μ, v)	3990.67	1.73586%	57.69%

Table 5.6: Comparison between Models in Terms of Three Metrics

For example, the GM model uses the first-order differential equation to characterize the the behavior of stock time series while the ARIMA model uses the linear dependence on previous data to find predicted value. However the linear dependence may not be the main feature of stock market. The other example is when we use the GM model, we ignore the historical data before the grey dimension degree and count more weight on the latest information. On the contrary, the ARIMA model estimates the parameters based on the whole sample data, sometimes it gives too much weight to the old information.

Although the ARIMA model can track information from stock time series, we find that the results are not as good as the researchers stated in their papers [3]. The reason is that, as we discussed in the previous chapters, the stock market is a system with both statistical and non-statistical features and is integrated by many sub systems. If we only use the ARIMA model for modeling stock time series, the results will not be accurate enough.

The above analysis gives us the idea that, in order to dig the potential of the ARIMA model and improve prediction accuracy, we should let the ARIMA model focus on tracking the statistical linear relationship of stock time series, which is its specialty, and let the grey model to take care of the non-statistical part in stock movement.

Before going any further, we need to study the following phenomena. In Chapter 2, we know that we can find the order of parameters (p, q) in the ARMA model by looking at the sample ACF and PACF plots or depending on the (BIC) and (AICC) criteria. However, in order to merge the ARIMA model with other models, we have to find a different criterion which can also be used by other models since the (BIC) and (AICC) criteria are only used for the ARMA model. In the thesis, we will choose the (TAE) criterion to find the order of (p, q) with the reason explained in Chapter 3. Here, we do not attempt to discuss the (AICC) and (BIC) criteria in detail since it is out of the range of this thesis. We propose the (TAE) criterion only due to the special case of studying stock time series and its clear financial meaning.

The (TAE) criterion can be used to find the order of (p, q) by using the following method. We first set the upper bounds P and Q for the AR and MA model respectively, and then we fit all possible ARMA(p, q) models for $p \leq P$ and $q \leq Q$ using the same sample size. The best model has

$$\min_{p \leq P, q \leq Q} TAE(p, q)$$

and the corresponding parameters (p, q) are the optimal ones.

Now the natural question is how different the (AICC), (BIC) and (TAE) are in terms of the MPAE metric. To answer this question, we choose the order of (p, q)

Criterion	MAPE
AICC	1.793%
TAE	1.836%
BIC	1.946%

Table 5.7: Comparison Between BIC, AICC and TAE in Terms of MPAE

in the ARIMA model based on the (AICC), (BIC) and (TAE) criteria respectively to conduct case study . The strategy is the same as the previous studies, in which we use the first 126 daily data to set up the model, then use the last 26 daily data to evaluate the established model in terms of the MPAE metric. Table (5.7) shows that the difference among those criteria is insignificant. In fact, the MAPE error based on the (TAE) criterion is even lower than the error based on the (BIC) criterion, which empirically proves that the (TAE) criterion is reasonable for choosing the order of the ARMA model when we study the special time series: stock time series.

Chapter 6

Prediction with the Revised G-ARMA Model

In the previous chapters, we find out that a single GM model and a single ARIMA model do not give a good result. It may be better to integrate them together for digging the potential of each model and overcoming the drawbacks of both models. After searching literatures, we find that Hu et al. [64] have proposed a hybird G-ARMA model by integrating the GM model and the ARMA model. The main idea is to first get the predicted value based on the grey model, and then use the ARMA model to find the forecasting value of the residuals. The final predicted value is then a combination of the two predicted values. However, in Hu's G-ARMA model [64], they evaluate parameters of the two models separately and do not find an optimal G-ARMA model. In this chapter, we will try to find an optimal G-ARMA model by evaluating all the parameters of the two models at the same time, and we call it the revised G-ARMA model to distinguish with Hu's G-ARMA model.

6.1 Revised G-ARMA Model

In Hu's G-ARMA model, when they build up a grey model, they do not consider the influence from the ARMA model. Also, when they estimate the parameters of the ARMA model, they do not care how the grey model will affect the ARMA model. In order to deal with the interaction between the grey model and the ARMA model, we propose the revised G-ARMA model by estimating all parameters in both models at the same time. Before introducing the revised G-ARMA model, we will first explain the method of constructing a traditional G-ARMA model.

Given a time series $X = (x_1, \ldots, x_n)$, we first find its grey prediction sequence $\widehat{Z} = (\widehat{z}_1, \widehat{z}_2, \ldots, \widehat{z}_n)$ based on the GM(1, 1) model and the residual sequence $Y = (y_1, \ldots, y_n) = X - \widehat{Z}$. Then, for the sequence Y, we can use an ARMA model to find its prediction sequence $\widehat{Y} = (\widehat{y}_1, \ldots \widehat{y}_n)$. So the prediction sequence of X is $\widehat{X} = \widehat{Z} + \widehat{Y}$. We can see that the predicted value $\widehat{x}_t$ is a combination of $\widehat{z}_t$ and $\widehat{y}_t$ and can be obtained by using the following formula

$$\widehat{x}_t = [1 - \exp(a)] \cdot \left[x^{(0)}(1) - \frac{b}{a}\right] \exp\left[-a(k-1)\right] + \sum_{i=1}^{p} \varphi_k y(t-i) + \sum_{j=1}^{q} \theta_j \tau(t-j), \quad (6.1)$$

where

$$\widehat{z}_t = [1 - \exp(a)] \cdot \left[x^{(0)}(1) - \frac{b}{a}\right] \exp\left[-a(k-1)\right],$$

$$\widehat{y}_t = \sum_{i=1}^{p} \varphi_k y(t-i) + \sum_{j=1}^{q} \theta_j \tau(t-j).$$

In the revised G-ARMA model, we use the GM(1, 1, μ, v) model instead of the GM(1, 1) model. Different combinations of parameters (μ,v, p, q) give different G-ARMA models, we will find the optimal G-ARMA model. In the following, we present the detail for constructing the revised G-ARMA model.

The random variable x_t can be represented as

$$x_t = \widehat{x}_t + \varepsilon_t,$$

where $\widehat{x}_t$ is the predicted value of x_t calculated by the G-ARMA model and ε_t is the prediction error made by the G-ARMA model. In order to identify the revised G-ARMA, we first set the upper bound and lower bound for the parameters (μ, υ, p, q). Then we fit all possible G-ARMA models for $\mu \in (l, L)\,; \upsilon \in (r, R)\,; 0 \leq p \leq P; 0 \leq q \leq Q$ to find the prediction sequence $\widehat{X}$ with the formula (6.1) and the related TAE. We claim that the optimal G-ARMA model satisfies

$$\min(TAE); \mu \in (l, L)\,; \upsilon \in (r, R)\,; 0 \leq p \leq P; 0 \leq q \leq Q,$$

where TAE=$\sum_{t=k}^{n} |\varepsilon_t|\,; k = \max(P, Q) + R + 1$ and the corresponding parameters $(\widehat{\mu}, \widehat{\upsilon}, \widehat{p}, \widehat{q})$ are the optimal parameters. Here we find the order (p, q) of the ARMA model without using the (AICC) and (BIC) criteria for the purpose of merging the ARMA model with the GM model. The additional reasons can be found in the previous chapters.

6.2 Case Study: Prediction Using the Revised G-ARMA Model

We will attempt to predict the same stock index data using the revised G-ARMA model. The strategy is the same as the previous case studies, in which we use the first 126 daily data to set up a revised G-ARMA model, and then use it to do one step forecasting. Here, we denote the first 126 daily data as time sequence X, and we will use it to construct a revised G-ARMA model with the following steps.

The first step is to set reasonable bounds for the parameters (μ, v, p, q), so we can get relatively good parameters without too much calculations. Recall that we set the bounds for the grey dimension parameter $v \in [4, 30]$ in Chapter 4. But in this study, we only set $v \in [5, 9]$ based on our studies, which shows that almost all the best v is located within this bound in almost all cases. The other reason is that after introducing the ARMA model into our model, the calculation increases significantly, so we have to narrow down the parameter space in order to save the calculation time. For the order (p, q) of the ARMA model, we set the bounds $p \in [0, 3], q \in [0, 3]$. Although we may improve forecasting accuracy by increasing the order of (p, q), it will cost too much to calculate and make the model too complex.

The next step is to discretize the continuous parameter μ as did in Chapter 4. Here, the interval length is 0.1 instead of 0.01 for the purpose of saving the calculation time. Now we have a total of 880 combinations of (μ,v,p, q), which means that we have to build up 880 traditional G-ARMA models and find the optimal one among them based on the (TAE) criterion.

Now we are ready to write code to help us find the optimal G-ARMA model. A program written by MTALAB shows that when $(\mu, v, p, q) = (6, 0.4, 3, 1)$, the TAE$=\sum_{t=k}^{126} |\varepsilon_t|$; $k = \max(p, q) + v + 1$ gives the minimum value. So the G-ARMA model constructed with the parameters $(\widehat{\mu}, \widehat{v}, \widehat{p}, \widehat{q}) = (6, 0.4, 3, 1)$ is our optimal model.

After we get the optimal G-ARMA model, forecasting the value of x_{127} becomes straightforward. That is, we first use the $GM(1, 1, 6, 0.4)$ model to get the grey prediction sequence $\widehat{Z} = (\widehat{z}_1, \widehat{z}_2, \ldots, \widehat{z}_{126})$ and the predicted value $\widehat{z}_{127}$ by following the same steps introduced in Chapter 2, and then use the residual sequence $Y = X - \widehat{Z}$ to build up an ARMA(3, 1) model to find the predicted value $\widehat{y}_{127}$. Then the predicted value of

x_{127} is given by

$$\widehat{x}_{127} = \widehat{z}_{127} + \widehat{y}_{127} = 9261.42,$$

which means that the predicted value for point 127 corresponding to January 2, 2009 is 9261.42.

To forecast point 128 corresponding to January 3, 2009, we need to first find the revised G-ARMA model with the first 127 sample data, and then use the above steps to find the predicted value for point 128. Recursively, we can obtain all the predicted prices from January 2, 2009 to February 6, 2009 (26 days).

Table (6.1) lists the results obtained from the case study and Figure (6.1) gives a visual comparison between the predicted values and the actual values.

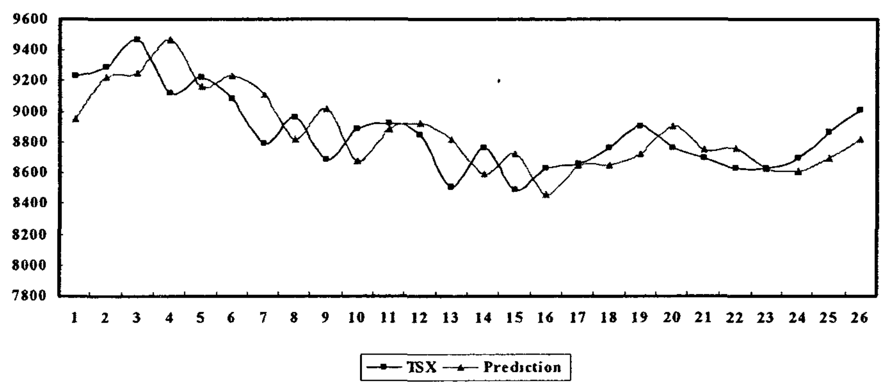


Figure 6.1: Comparison Between Actual and Predicted Values Based on the G_ARMA Model

6.3 Conclusions

In this chapter, instead of using a single grey model or a single ARIMA model, we use a hybrid model called the revised G-ARMA model to forecast the stock time series. From Table (6.2), we can see that the revised G-ARMA model has relatively lower prediction

Time	Rel.	Abs.	Org.	Pre.	AE.	Error%	Dir.
2009/1/2	127	1	9234.11	8953.2	280.91	3.042091	1
2009/1/5	128	2	9285.51	9222.8	62.71	0.675353	1
2009/1/6	129	3	9472.09	9248.9	223.19	2.356291	1
2009/1/7	130	4	9121.32	9469.3	347.98	3.815018	0
2009/1/8	131	5	9221.58	9162.7	58.88	0.638502	1
2009/1/9	132	6	9085.18	9236.3	151.12	1.663368	1
2009/1/12	133	7	8793.33	9110.6	317.27	3.608076	1
2009/1/13	134	8	8961.55	8817.1	144.45	1.611886	0
2009/1/14	135	9	8688.36	9017.1	328.74	3.783683	1
2009/1/15	136	10	8879.61	8675.1	204.51	2.303142	0
2009/1/16	137	11	8920.4	8885.8	34.6	0.387875	0
2009/1/19	138	12	8841.48	8924.3	82.82	0.936721	1
2009/1/20	139	13	8504.93	8821.9	316.97	3.726897	0
2009/1/21	140	14	8757.89	8595	162.89	1.859923	1
2009/1/22	141	15	8486.56	8719.6	233.04	2.745989	1
2009/1/23	142	16	8627.97	8456.4	171.57	1.988533	0
2009/1/26	143	17	8656.51	8652.6	3.91	0.045168	0
2009/1/27	144	18	8759.63	8643.9	115.73	1.321175	0
2009/1/28	145	19	8906.23	8719.7	186.53	2.094377	1
2009/1/29	146	20	8762.76	8898.4	135.64	1.547914	0
2009/1/30	147	21	8694.9	8747.6	52.7	0.606102	0
2009/2/2	148	22	8624.83	8764	139.17	1.613597	0
2009/2/3	149	23	8628.63	8630.7	2.07	0.02399	0
2009/2/4	150	24	8693.09	8608.5	84.59	0.973072	0
2009/2/5	151	25	8860.98	8699.3	161.68	1.824629	0
2009/2/6	152	26	9008.02	8815.1	192.92	2.141647	0

Table 6.1: Prediction Results for 26 Daily Data Based on the revised G-ARMA Model

error in terms of the TAE, MPAE and DIR metrics. However, we also notice that the error does not drop significantly. From the features of the revised G-ARMA model, we can see that we have tried very hard to dig the potential of both models. Why is the result only slightly better than the GM model? One of the reasons, as we discussed before, is that stock time series is very complicated time series with many sub-systems, and each sub-system has its own statistical features. So it is hard to improve predictive ability by directly modeling the raw data. Having this finding in mind, in the next chapters, we will try to find a way to divide the complex stock time series into several relatively simple sub time series. We believe that this new way can help us dig the potential of G-ARMA, GM and ARMA models.

Model	TAE	MPAE	DIR
ARMA	4500.1446	1.94633%	57.69%
$GM(1, 1, \mu, v)$	3990.67	1.73586%	57.69%
G-ARMA	3852.97	1.69%	53.84%

Table 6.2: Comparison Between Models in Terms of the Three Metrics

Chapter 7

Forecasting Stock Index Based on the WG-ARMA Model

The result obtained in Chapter 6 is still not satisfactory. In the present chapter, we will attempt to predict the stock time series by applying the wavelet transform to our model. As we stated in Chapter 1, wavelets have the power to extract information from a non-stationary signal and to decompose a complex signal into several relatively simple signals. Hence it would be natural to combine the wavelet transform with the ARIMA model and the GM model.

7.1 WG-ARMA Model

Before proposing our model, we need to consider two things: the type of the wavelet function and the decomposition level. In most papers, researchers decide them based on their experiences.

In terms of the wavelet function, it is easy to understand that we should choose a proper wavelet function to represent the unique characteristic of a signal. However, the movement of stock is so complex, it is not reasonable to guess what kind of wavelet function is proper for analyzing stock movement. In this chapter, we use the parameter m to represent the wavelet function, and decide the proper wavelet function by considering the TAE criterion. Still, we will estimate m with all the other parameters simultaneously. We use the parameter k to denote the decomposition level and will choose it based on the (TAE) criterion.

Now we are ready to propose the new model. According to the wavelet transform, given a sequence X, for a fixed wavelet function m and a fixed decomposition level k, we can decompose it as

$$X = A_m^0 + D_m^1 + D_m^2 + \cdots + D_m^k,$$

where A_m^0 is the low frequency sequence and D_m^1 ,$D_m^2, \ldots, D_m^k$ are high frequency sequences obtained from the wavelet function m. The low frequency sequence is the approximate sequence of X, and the high frequency sequences are the detail sequences of X. For each sequence, we will choose a proper model to fit its characteristics. In the thesis, we will use the G-ARMA model to describe the low frequency sequence since it is the approximation of the raw sequence, and use the ARMA models to describe high frequency sequences since they all have constant means and relatively stable variances. Thus, for the given sequence X, we write it in the following form

$$\begin{aligned} X = \hat{A}_m^0 + \hat{D}_m^1 + \hat{D}_m^2 + \cdots + \hat{D}_m^k + \varepsilon = [GM(1,1,v,\mu) + ARMA(p^0,q^0)]_m^0 \\ + [ARMA(p^1,q^1)]_m^1 + [ARMA(p^2,q^2)]_m^2 + \cdots + \left[ARMA(p^k,q^k)\right]_m^k + \varepsilon, \end{aligned} \tag{7.1}$$

where $\hat{A}_m^0$ is the forecasting sequence of A_m^0; $\hat{D}_m^1, \hat{D}_m^2, \ldots, \hat{D}_m^k$ are the forecasting sequences of $D_m^1, D_m^2, \ldots, D_m^k$ and ε is the residual sequence. We call this model the WG-ARMA model.

The method of constructing a WG-ARMA model is similar to that of constructing a revised G-ARMA model. First of all, we need to set the bounds for the parameters $(\mu, v, p^0, q^0, p^1, q^1, ..., p^k, q^k, m, k)$, such that

$$\begin{aligned} &\mu \in [l, L], v \in [r, R], p^0 \in [0, P^0], q^0 \in [0, Q^0], ..., \\ &p^k \in [0, P^k], q^k \in [0, Q^k], m \in [0, M], k \in [0, K]. \end{aligned} \tag{7.2}$$

Second, we need to discretize the continuous parameter μ, and then fit all the possible WG-ARMA models by using equation (7.1) for the combination of the parameters in (7.2). The best model satisfies

$$\min(TAE) = \left|\sum_{t=j}^{n} \varepsilon_t\right|; \; j = \max\left(P^0, \cdots, P^k, Q^0, \cdots, Q^k\right) + \text{R} + 1.$$

Here we use the (TAE) criterion to find the orders of $(p^0, q^0); \cdots; (p^k, q^k)$ without using the (BIC) and (AICC) criteria for the same reason stated in the previous chapters.

7.2 A Case Study Using the WG-ARMA Model

In this case study, the data used are the same as before. The period from June 30, 2008 to December 31, 2008 (126 days) is used to set up the first WG-ARMA model for predicting the one-step value, and then the whole testing data which are from January 2, 2009 to February 6, 2009 (26 days) can be found by following the same steps.

As mentioned before, we need to consider the cost of building up a model. Hence, the first step is to set proper bounds and discretize the continuous parameters with a reasonable interval length.

For the parameters $(\mu, v, p^0; q^0, p^1, q^1, ..., p^k, q^k)$, we will use the same settings in the revised G-ARMA model, which is

$$\mu \in [0, 0.1, \cdots, 1], v \in [5, \cdots, 9], \left(p^0, q^0, p^1, q^1, ..., p^k, q^k\right) \in [0, 3].$$

For the new parameter m, which represents the wavelet function, the bounds are limited by the availability of the wavelet functions provided by MATLAB. We give them in Table (7.1).

1	2	3	4	5	6	7	8
Haar	Db1	Db2	Db3	Db4	Db5	Db6	Db7
9	10	11	12	13	14	15	16
Db8	Db9	Db10	Sym2	Sym3	Sym4	Sym5	Sym6
17	18	19	20	21	22	23	24
Sym7	Sym8	Coif1	Coif2	Coif3	Coif4	Coif5	Bior1.1
25	26	27	28	29	30	31	32
Bior1.3	Bior1.5	Bior2.2	Bior2.4	Bior2.6	Bior2.8	Bior3.1	Bior3.3
33	34	35	36	37	38	39	40
Bior3.5	Bior3.7	Bior3.9	Bior4.4	Bior5.5	Bior6.8	Rbio1.1	Rbio1.3
41	42	43	44	45	46	47	48
Rbio1.5	Rbio2.2	Rbio2.4	Rbio2.6	Rbio2.8	Rbio3.1	Rbio3.3	Rbio3.5
49	50	51	52	53			
Rbio3.7	Rbio3.9	Rbio4.4	Rbio5.5	Rbio6.8			

Table 7.1: The Serial Number of Wavelet Functions

For the decomposition level parameter k, we only set $k = 2$ in the case study since the calculation time will increase significantly as k increases.

Now we have a total of 655875 combinations of $(v, \mu, k, m, p^0, q^0, p^1, q^1, p^2, q^2)$, which means that we need to set up 655875 temporary WG-ARMA models and calculate the TAE for each of them. The final WG-ARMA model has the lowest TAE and the corresponding parameters $\left(\hat{v}, \hat{\mu}, \hat{k}, \hat{m}, \hat{p}^0, \hat{q}^0, \hat{p}^1, \hat{q}^1, \hat{p}^2, \hat{q}^2\right)$ are the optimal ones.

After all the above considerations, we now use computers to do the calculations. A program written by MATLAB 2008b helps us find the optimal parameters as

$$\left(v = 1, \mu = 6, k = 1, m = 47, p^0 = 3, q^0 = 2, p^1 = 3, q^1 = 3\right),$$

which means that we need to decompose the sample stock time series into one lower frequency sequence $A^0 = (a_1^0, a_2^0, \ldots, a_{126}^0)$ and one high frequency sequence $D^1 = (d_1^1, d_2^1, \ldots, d_{126}^1)$ based on the wavelet function 'Rbio3.3' (see Figure 7.1). For the sequence A^0, we find its predicted value $\hat{A}^0$ by using the G-ARMA model with the parameters ($\upsilon = 1, \mu = 6, p^0 = 3, q^0 = 2$); for the sequence D^1, we use the ARMA model with the order ($p^1 = 3, q^1 = 3$) to find its prediction sequence $\hat{D}_1$. Therefore, the prediction sequence $\hat{X}$ is given by

$$\hat{X} = \hat{A}^0 + \hat{D}^1 = [GM(1,1,1,6) + ARMA(2,3)]^0 + [ARMA(3,3)]^1 .$$

After substituting the real data into the model, we get the predicted value for point 127

$$\hat{x}_{127} = \hat{a}_{127}^0 + \hat{d}_{127}^1 = 9248.1,$$

where $\hat{a}_{127}^0$ is the predicted value for the low frequency part and $\hat{d}_{127}^1$ is the predicted value for the high frequency part, which are given by the following equations

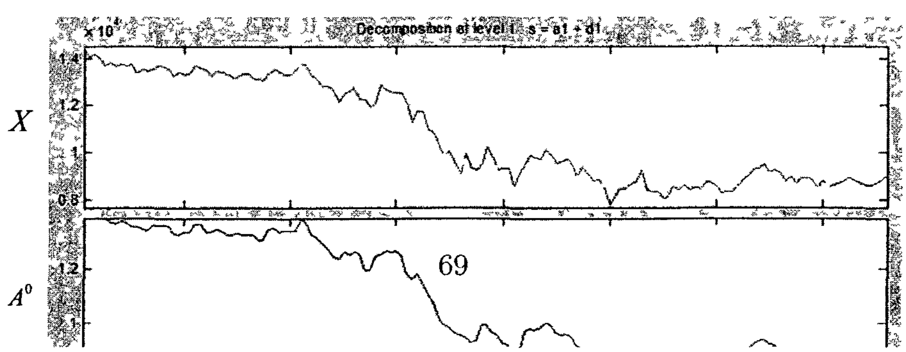
$$\hat{a}_{127}^0 = (1 - e^a)\left(a_{121}^0 - \tfrac{b}{a}\right) e^{-6a} + \phi_1^0(a_{126}^0 - \hat{a}_{126}^0) + \phi_2^0(a_{125}^0 - \hat{a}_{125}^0) + \phi_3^0(a_{124}^0 - \hat{a}_{124}^0)$$
$$+ \tau_{127}^0 - \theta_1^0\tau_{126}^0 - \theta_2^0\tau_{125}^0 - \theta_3^0\tau_{124}^0,$$

$$\hat{d}_{127}^1 = \phi_1^1(d_{126}^1 - \hat{d}_{126}^1) + \phi_2^1(d_{125}^1 - \hat{d}_{125}^1) + \phi_3^1(d_{124}^1 - \hat{d}_{124}^1) + \tau_{127}^1 - \theta_1^1\tau_{126}^1 - \theta_2^1\tau_{125}^1 -$$
$$\theta_3^1\tau_{124}^1.$$

For point 128, we follow the same strategy. We first construct a WG-ARMA model by using the first 127 daily data, and then get its one step predicted value. By repeating the same procedure, we can finally get all the 26 forecasting points which start at January 2, 2009 and end at February 6, 2009. Table (7.2) lists the results and Figure (7.2) gives the plot.

Time	Rel.	Abs.	Org.	Pre.	AE.	Error%	Dir.
2009/1/2	127	1	9234.11	9248.1	13.99	0.151504	0
2009/1/5	128	2	9285.51	9026.9	258.61	2.785092	1
2009/1/6	129	3	9472.09	9531.9	59.81	0.631434	0
2009/1/7	130	4	9121.32	9299.9	178.58	1.957831	0
2009/1/8	131	5	9221.58	9274.7	53.12	0.57604	0
2009/1/9	132	6	9085.18	9140.8	55.62	0.612206	0
2009/1/12	133	7	8793.33	9052.7	259.37	2.949622	0
2009/1/13	134	8	8961.55	8926.8	34.75	0.387768	0
2009/1/14	135	9	8688.36	8834.3	145.94	1.679719	0
2009/1/15	136	10	8879.61	8787.7	91.91	1.035068	0
2009/1/16	137	11	8920.4	8777.4	143	1.603067	1
2009/1/19	138	12	8841.48	8857.7	16.22	0.183453	0
2009/1/20	139	13	8504.93	8918.4	413.47	4.861533	1
2009/1/21	140	14	8757.89	8647.6	110.29	1.259322	0
2009/1/22	141	15	8486.56	8656	169.44	1.996569	0
2009/1/23	142	16	8627.97	8572.6	55.37	0 64175	0
2009/1/26	143	17	8656.51	8583.6	72.91	0.842256	1
2009/1/27	144	18	8759.63	8601.2	158.43	1.808638	1
2009/1/28	145	19	8906.23	8831.2	75.03	0.842444	0
2009/1/29	146	20	8762.76	8790.6	27.84	0.317708	0
2009/1/30	147	21	8694.9	8848.3	153.4	1.764253	1
2009/2/2	148	22	8624.83	8696.9	72.07	0.835611	1
2009/2/3	149	23	8628.63	8598.2	30.43	0.352663	1
2009/2/4	150	24	8693.09	8603	90.09	1.03634	1
2009/2/5	151	25	8860 98	8692.8	168.18	1.897984	1
2009/2/6	152	26	9008.02	8793.8	214.22	2.378103	1

Table 7 2: Prediction Results for 26 Daily Data Based on the WG-ARMA Model



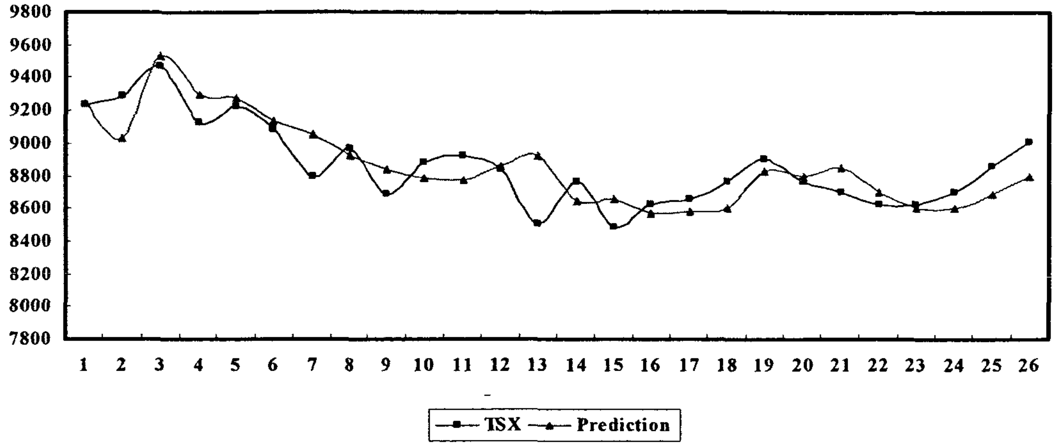


Figure 7.2: Comparison Between Actual and Predicted Values Based on the WG-ARMA Model

7.3 Conclusions

In this chapter, we propose a new hybrid model, the WG-ARMA model, which takes advantage of the power of the wavelet transform to first decompose the stock time series into several sub-time series in different time horizons, and then model the sub time series separately. Table (7.3) shows that the WG-ARMA model is much better than other models in terms of the TAE, MPAE, and DIR metrics. Especially, the DIR drops significantly. Thus, we conclude that predictive capacity of the WG-ARMA model is stronger than that of the ARIMA model, GM model and G-ARMA model.

Model	TAE	MPAE	DIR
ARMA	4500.1446	1.94633%	57.69%
GM(1, 1; μ, v)	3990.67	1.73586%	57.69%
G-ARMA	3852.97	1.69%	53.84%
WG-ARMA	3122.09	1.36108%	%42.31

Table 7.3: Comparison Between Models in Terms of Three Metrics

The finding of this chapter also empirically proves our assumption made in the previous chapters. That is, directly modeling stock time series may not dig the potential of the GM model and the ARIMA model. Recall that the stock market is a complex system with many sub-systems. Although the ARIMA model and the GM model are excellent tools for analyzing time series, they do not have power to deal with the time series with many sub-systems mixed together. However, even after we make a small work to decompose the stock time series into only 2 time series (see Figure 7.1), the ARMA model and the GM model can then extract more information from the stock series. The above analysis tells us that the ARMA model and the GM model also work for stock time series; however, we should use them in a proper way.

Although the wavelet transform significantly improves prediction accuracy, there are two drawbacks with the wavelet transform since it can only decomposes the low frequency part of a time series. The first one is that the wavelet transform has only one pattern, which may not give us more opportunity to find the best decomposition of the stock time series. The second one is that we may lose some information by modeling the high frequency sequences without further decomposing them. Keeping that analysis in mind, we will introduce the wavelet packets transform to overcome the two drawbacks since the wavelet packets transform can decompose the time series in both low and high frequencies.

Chapter 8

Stock Index Forecasting with the WPG-ARMA Model

In the previous chapter, we find that although the wavelet transform can improve prediction accuracy, there are still some drawbacks. For example, the wavelet transform lacks of flexibility since it can only decompose the low frequency parts of a signal in each time resolution. By contrast, the wavelet packets transform can decompose the low and high frequency parts simultaneously. So it gives us more power to represent stock time series. For example, sometimes the special features of a stock index are located around certain frequencies and bandwidths, The wavelets packets decomposition allows us to choose a proper decomposition tree to match the packets with appropriate frequencies and bandwidths. Due to this characteristic of wavelet packets, in this chapter, we will replace the wavelet transform with the wavelet packets transform. We give the new model the name of WPG-ARMA model.

8.1 WPG-ARMA Model

The WPG-ARMA model is similar to the WG-ARMA model. The only difference is that, in the WPG-ARMA model, we use the wavelet packets decomposition instead of the wavelet transform. As we stated in the introduction, to use the wavelet packets transform we have to first find the best decomposition tree based on some criteria [65,66]. However, those criteria may not be proper for a stock index and cannot be used when we integrate the wavelet packets transform with the other models. Hence, we have to figure out a different way to find a proper wavelet packets tree. To keep consistency, we use the (TAE) criterion to find the best tree. In the following, we will describe the algorithm in detail on how to build up a WPG-ARMA model.

First of all, let us introduce a new parameter ρ to represent the wavelet packets decomposition tree. An example is given in Table (8.1). Given a fixed wavelet function m and decomposition tree ρ, we may decompose the time series X as follows

$$
\begin{aligned}
&X = A_m^{1,\rho} + D_m^{2,\rho} + D_m^{3,\rho} + \cdots + D_m^{k,\rho},\\
&m : \text{wavelet function},\\
&\rho : \text{decompostion tree},\\
&k : \text{end branchs of the tree},
\end{aligned}
$$

where $A_m^{1,\rho}$ represents the low frequency sequence, and $D_m^{2,\rho}, D_m^{3,\rho}, \ldots, D_m^{k,\rho}$ are several high frequency sequences. For the approximate sequence $A_m^{1,\rho}$, we will use the G-ARMA model to describe it; for the high frequency sequences $D_m^{2,\rho}, D_m^{3,\rho}, \ldots, D_m^{k,\rho}$, we only use ARMA models to describe them. Thus, we can set up the WPG-ARMA model for the time series X as

$$X = \hat{A}_m^{1,\rho} + \hat{D}_m^{2,\rho} + \cdots + \hat{D}_m^{k,\rho} + \varepsilon = [GM(1,1,v,\mu) + ARMA(p^1,q^1)]_m^{1,\rho} \\ + [ARMA(p^2,q^2)]_m^{2,\rho} + \cdots + \left[ARMA(p^k,q^k)\right]_m^{k,\rho} + \varepsilon, \tag{8.1}$$

where the sequence $\hat{A}_m^{1,\rho}$ is the forecasting sequence for the low frequency sequence $A_m^{1,\rho}$; the sequences $\hat{D}_m^{2,\rho}, \hat{D}_m^{3,\rho}, \ldots, \hat{D}_m^{k,\rho}$ are the forecasting sequences for the high frequency sequences $D_m^{2,\rho}, D_m^{3,\rho}, \ldots, D_m^{k,\rho}$, and the sequence ε is the residual sequence. The total parameters needed to be estimated are $\left(v, \mu, p^1, q^1, ..., p^k, q^k, m, \rho\right)$. Similar to the previous chapter, we will first set the bounds and interval length for the parameters. Then we fit all possible WPG-ARMA models by using equation (8.1) for the following parameter space

$$v \in [r, R]\,; \mu \in [l, L]\,, p^1 \in [1, P^1]\,, q^1 \in [1, Q^1]\,, ...,$$
$$p^k \in \left[0, P^k\right], q^k \in \left[0, Q^k\right], m \in [0, M]\,, \rho \in [1, J]\,.$$

We claim that the best model satisfies

$$\min(TAE) = \left|\sum_{t=j}^{n} \varepsilon_t\right|; \; j = \max\left(P^1, \cdots, P^k, Q^1, \cdots, Q^k\right) + \text{R} + 1.$$

Here, we use the (TAE) criterion to find the orders of $(p^1, q^1); \ldots; \left(p^k, q^k\right)$ without using the (BIC) and (AICC) criteria.

8.2 Case Study: Forecasting with the WPG-ARMA Model

The current case study uses the same data of Chapter 7. The period from June 30, 2008 to December 31, 2008 (126 days) is used to set up the first WPG-ARMA model for predicting the one-step value, and then the predicted values for the rest testing data from January 3, 2009 to February 6, 2009 (25 days) can be found by following the same procedures.

As mentioned before, we need to consider the cost of building up a model. Hence, first of all, we need to set up reasonable bounds and interval length for the parameters. For the new parameter of the decomposition tree ρ, since the calculation time will significantly increase as the decomposition level increases, we only let the upper bound of ρ equal 8 here. That means that we have the total of 8 decomposition trees to be evaluated. Figure (8.1) lists the trees. Here we do not include the wavelet decomposition trees since we have studied them in Chapter 7.

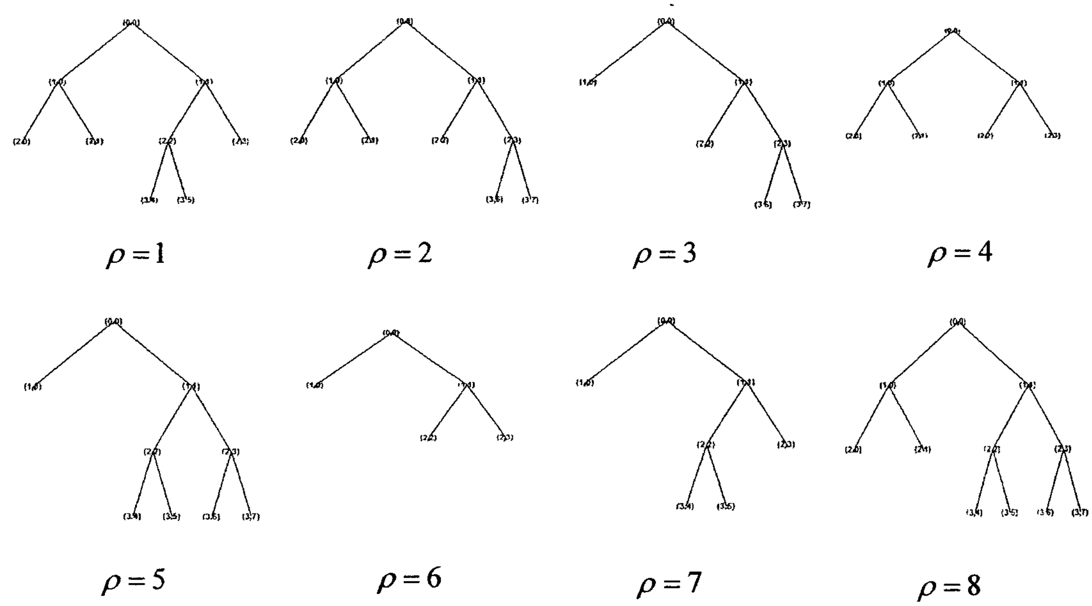


Figure 8.1: Decomposition Trees Table

For the parameters $(\mu, v, p^1, q^1, ..., p^k, q^k)$, we will use the same settings in the revised G-ARMA model, which is

$$\mu \in [0, 0.1, \cdots, 1], v \in [5, \cdots, 9], (p^1, q^1, ..., p^k, q^k) \in [0, 3],$$

where the parameter k is related to the decomposition tree ρ. After we find ρ, k will be fixed.

For the parameter m, we have to redefine it since as the layer becomes deeper, the $\hat{A}_m^{1,\rho}$ becomes unsuitable to build up a G-ARMA model. Table (8.1) lists the new serial number for the wavelet functions. To construct the final WPG-ARMA model, we need to set

1	2	3	4	5	6	7	8
Db2	Db4	Db5	Db6	Db7	Db8	Db9	Db10
9	10	11	12	13	14	15	16
Sym2	Sym4	Sym5	Sym6	Sym7	Sym8	Coif1	Coif2
17	18	19	20	21	22	23	24
Coif3	Coif4	Coif5	Bior2.2	Bior2.4	Bior2.6	Bior2.8	Bior3.1
25	26	27	28	29	30	31	32
Bior3.3	Bior3.5	Bior3.7	Bior3.9	Bior4.4	Bior5.5	Bior6.8	Rbio1.3
33	34	35	36	37	38	39	40
Rbio1.5	Rbio2.2	Rbio2.4	Rbio2.6	Rbio2.8	Rbio3.1	Rbio3.3	Rbio3.5
41	42	43	44	45			
Rbio3.7	Rbio3.9	Rbio4.4	Rbio5.5	Rbio6.8			

Table 8.1: The Serial Number of Related Wavelet Functions

up temporal WPG-ARMA modes for each combination of $\left(\mu, v, \rho, m, p^1, q^1, ..., p^k, q^k\right)$, and then calculate the TAE for all those models. The optimal WPG-ARMA model has the lowest TAE and the corresponding parameters $\left(\hat{\mu}, \hat{v}, \hat{\rho}, \hat{m}, \hat{p}^1, \hat{q}^1, ..., \hat{p}^k, \hat{q}^k\right)$ are the optimal parameters. We still write a program by MATLAB 2008b to do the calculations. The program finally helps us find the optimal parameters as

$$\left(\hat{\mu} = 1, \hat{v} = 5, \hat{\rho} = 9, \hat{m} = 39, \hat{p}^1 = 2, \hat{q}^1 = 3, \hat{p}^2 = 3, \hat{q}^2 = 3, \hat{p}^3 = 3, \hat{q}^3 = 3\right),$$

which means that we need to decompose the stock time series into one lower frequency sequence $A^1 = (a_1^1, a_2^1, \ldots, a_{126}^1)$, one high frequency sequence $D^2 = (d_1^2, d_2^2, \ldots, d_{126}^2)$

and one high frequency sequence $D^3 = (d_1^3, d_2^3, \ldots, d_{126}^3)$ based on the wavelet function 'Rbio3.3' (see Figure 8.1). For the sequence A^1, we may find its predicted value $\hat{A}^1$ by using the G-ARMA model with parameter $(\mu = 1, \upsilon = 5, p^1 = 3, q^1 = 2)$; for the sequence D^1, we use the ARMA model with the order $(p^2 = 3, q^2 = 3)$ to find its prediction sequence $\hat{D}^2$; for the sequence D^3, we find its prediction sequence $\hat{D}^3$ by using the ARMA model with the order $(p^3 = 3, q^3 = 3)$. Hence, the prediction sequence $\hat{X}$ can be represented as

$$\hat{X} = \hat{A}^1 + \hat{D}^2 + \hat{D}^3 = [GM(1,1,1,6) + ARMA(2,3)]^1 + [ARMA(3,3)]^2 + [ARMA(3,3)]^3 .$$

After substituting the real data into the model, we get the predicted value for point 127 is

$$\hat{x}_{127} = \hat{a}_{127}^1 + \hat{d}_{127}^2 + \hat{d}_{127}^3 = 9230.6,$$

where $\hat{a}_{127}^1$ is the one-step predicted value for the low frequency time series A^1, $\hat{d}_{127}^2$ is the one-step predicted value for the high frequency time series D^2 and $\hat{d}_{127}^3$ is the one-step predicted value for the high frequency time series D^3 at point 127. They are given by

$$\hat{a}_{127}^1 = (1-e^a)\left(a_{121}^1 - \frac{b}{a}\right)e^{-6a} + \phi_1^1(a_{126}^1 - \hat{a}_{1,126}) + \phi_2^1(a_{125}^1 - \hat{a}_{125}^1) + \tau_{127}^1 - \theta_1^1\tau_{126}^1 - \theta_2^1\tau_{125}^1 - \theta_3^1\tau_{124}^1,$$

$$\hat{d}_{127}^1 = \phi_1^2(d_{126}^2 - \hat{d}_{126}^2) + \phi_2^2(d_{125}^2 - \hat{d}_{125}^2) + \phi_3^2(d_{124}^2 - \hat{d}_{124}^2) + \tau_{127}^2 - \theta_1^2\tau_{126}^2 - \theta_2^2\tau_{125}^2 - \theta_3^2\tau_{124}^2,$$

$$\hat{d}_{127}^3 = \phi_1^3(d_{126}^3 - \hat{d}_{126}^3) + \phi_2^3(d_{125}^3 - \hat{d}_{125}^3) + \phi_3^3(d_{124}^3 - \hat{d}_{124}^3) + \tau_{127}^3 - \theta_1^3\tau_{126}^3 - \theta_2^3\tau_{125}^3 - \theta_3^3\tau_{124}^3.$$

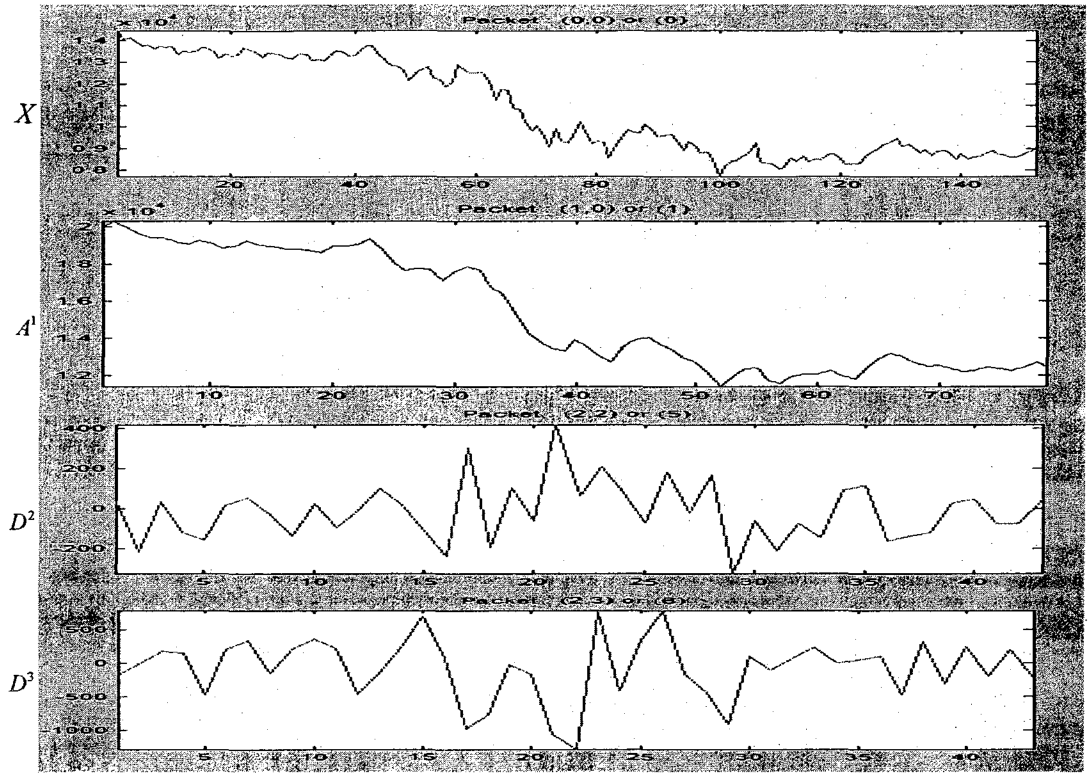


Figure 8.2: The Decomposition of the Time Series X with the Wavelet Function 'Rbio3.3'

Therefore, based on the historical data in conjunction with the WPG-ARMA model, the predicted value for January 2, 2009 is 9230.6, and the prediction error in terms the MPEA is only 0.038011%. After repeating the same procedure, we obtain the predicted values for all the other 25 forecasting points. The results are given in Table (8.2). And the plot is given in Figure (8.1) to visually compare the actual values and the predicted values.

Time	Rel.	Abs.	Org.	Pre.	AE.	Error%	Dir.
2009/1/2	127	1	9234.11	9230.6	3.51	0.038011	0
2009/1/5	128	2	9285.51	9020.2	265.31	2.857247	1
2009/1/6	129	3	9472.09	9527.6	55.51	0.586038	0
2009/1/7	130	4	9121.32	9295.9	174.58	1.913977	0
2009/1/8	131	5	9221.58	9263.9	42.32	0.458924	0
2009/1/9	132	6	9085.18	9133.7	48.52	0.534057	0
2009/1/12	133	7	8793.33	9058.6	265.27	3.016718	0
2009/1/13	134	8	8961.55	8945.3	16.25	0.18133	0
2009/1/14	135	9	8688.36	8850.7	162.34	1.868477	0
2009/1/15	136	10	8879.61	8803.1	76.51	0.861637	0
2009/1/16	137	11	8920.4	8783	137.4	1.54029	1
2009/1/19	138	12	8841.48	8852.1	10.62	0.120116	0
2009/1/20	139	13	8504.93	8903.4	398.47	4.685165	1
2009/1/21	140	14	8757.89	8639.9	117.99	1.347242	0
2009/1/22	141	15	8486.56	8662.8	176.24	2.076695	0
2009/1/23	142	16	8627.97	8583.6	44.37	0.514258	0
2009/1/26	143	17	8656.51	8589.1	67.41	0.77872	1
2009/1/27	144	18	8759.63	8616.2	143.43	1.637398	1
2009/1/28	145	19	8906.23	8844.9	61.33	0.688619	0
2009/1/29	146	20	8762.76	8780.9	18.14	0.207012	0
2009/1/30	147	21	8694.9	8824.9	130	1.495129	1
2009/2/2	148	22	8624.83	8689.4	64.57	0.748652	0
2009/2/3	149	23	8628.63	8596.5	32.13	0.372365	1
2009/2/4	150	24	8693.09	8597.9	95.19	1.095008	1
2009/2/5	151	25	8860.98	8702.4	158.58	1.789644	0
2009/2/6	152	26	9008.02	8791.5	216.52	2.403636	1

Table 8.2: Prediction Results for 26 Daily Data Based on the WPG-ARMA Model

8.3 Conclusions

In this chapter, we propose a new hybrid model, the WPG-ARMA model, which is an improved version of the WG-ARMA model. We use the WPG-ARMA model to conduct

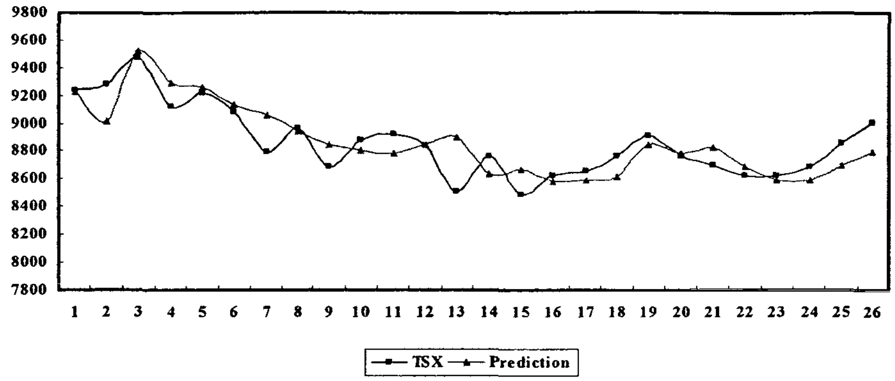


Figure 8.3: Comparison Between Actual and Predicted Values Based on the WPG-ARMA Model

a case study based on the same historical data and the same strategy as before. Table (8.3) shows that the WPG-ARMA model has the lowest error in terms of the TAE, MPAE and DIR metrics. This result enables us to conclude that the WPG-ARMA model is better than the WP-ARMA model since it has stronger capacity to represent stock time series. We also conclude that the WPG-ARMA model will be great helpful for investors with the MPAE being as low as 1.3% and DIR at 34.62%.

Model	TAE	MPAE	DIR
ARMA	4500.1446	1.94633%	57.69%
GM(1, 1, μ, v)	3990.67	1.73586%	57.69%
G-ARMA	3852.97	1.69%	53.84%
WG-ARMA	3122.09	1.36108%	%42.31
WPG-ARMA	2982.51	1.30063%	34.62%

Table 8.3: Comparison Between Models in Terms of Three Metrics

However, we also notice that there are still errors even though we have introduced many parameters in the WPG-ARMA model and try to extract information as much as possible from the historical data. The reason is that although the historical information can give great help for investors, there is no guarantee that investors can always make a right decision only based on available information in the market.

Chapter 9

Conclusions

In the thesis, we attempt to predict the stock market returns on the daily basis. In order to do that, we first propose the GM(1, 1, μ, v) model by introducing two parameters into the traditional GM(1, 1) model and estimate the new parameters (μ, v) based on the (TAE) criterion. We then revise the normal G-ARMA model by integrating the GM(1, 1, μ, v) model with the ARMA model and use the (TAE) criterion to estimate all the parameters simultaneously for the purpose of improving prediction accuracy. In order to overcome the drawbacks of directly modeling the original data, we introduce the wavelet transform into the revised G-ARMA model and call this new model WG-ARMA model. To keep consistency, we estimate all the parameters based on the (TAE) criterion. Finally, we get the final WPG-ARMA model by replacing the wavelet transform with the wavelet packets decomposition. For each of those models, we conduct a case study and then measure performance of those models in terms of the TAE, MPAE and DIR metrics. Based on the findings, we draw the following conclusions:

First of all, we conclude that prediction task in the stock market is feasible, although the result is not very accurate, but at least it can give a sign for tomorrow's returns.

Second, it is not a good idea to use the grey model and the ARIMA model directly on very complicated time series like stock time series even though they are both excellent tools for analyzing time series.

Finally, it is strongly suggested to use the wavelet and wavelet packets transforms to decompose the original stock time series properly before using the ARIMA model and the grey model. By this way, we can dig the potential of the ARIMA model and the grey model. Furthermore, the results also suggest that it is better to use the wavelet packets transform than the wavelet transform due to its stronger capacity to represent stock time series.

9.1 Contributions

The main contributions of the thesis are the following:

1. We improve the traditional GM(1, 1) model to the GM(1, 1, μ, v) model by introducing two parameters.

2. We improve the normal model G-ARMA by merging the GM(1, 1, μ, v) with the ARMA model.

3. We propose a new hybrid WG-ARMA model by introducing the wavelet transform.

4. The final model is the WPG-ARMA model which uses the wavelet packets transform to replace the wavelet transform.

5. The technique of constructing hybrid model is different from that of other researchers. Instead of analyzing the time series separately, we merge the models into a single model by estimating all the parameters simultaneously.

9.2 Future Works

In this section, we indicate the direction towards further improving the predictive ability of our models.

9.2.1 Using Other Grey Models

In the thesis, we only focus on the GM(1, 1) model. In the future work, we can consider introducing other prediction model based on grey theory, such as verhulst model. The verhulst model is mainly used to describe and study sigmoid processes. For example, this model is often used in the prediction of human populations, biological growth, reproduction, economic life span of consumable products, etc. [67]. Therefore, it may be more suitable than GM(1,1) model in analyzing stock time series.

9.2.2 Replacing the Linear ARIMA Model

In the thesis, we only use the linear ARIMA model and the GM model to analyze stock time series. However, we know that the stock time series is a combination of linear and non-linear systems. Although both linear model and non-linear model have been successfully applied to the stock market, in reality, stock time series data typically contain both linear and nonlinear patterns. So, a hybrid model which can integrate the linear and non-linear models should have stronger predictive ability. In the future, people should consider replacing the linear ARMA model with a non-linear model, such as the autoregressive conditional heteroscedasticity (ARCH) model, which could better simulate the clustering and time varying properties existing in the stock fluctuation [68].

9.2.3 Introducing Additional Wavelet Functions

According to the study in the thesis, we can see that the wavelet function has influence on prediction accuracy. However, due to the availability of wavelet function in MATLAB, we still have many wavelet functions left without using. In the future, we may introduce additional wavelet functions into our model or even find a wavelet function which is more suitable for the stock time series than the wavelet functions used in this thesis.

References

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

[2] Makridakis S., Anderson A., Carbone R., Fildes R., Hibdon M., Lewandowski R., Newton J., Parzen E. and Winkler R., *The Accuracy of Extrapolation(Time Series) Methods*, Results of a Forecasting Competition.J.Forecasting 1, pp.111-153, 1982

[3] Zhu N., Xu B. and Tong D. B., *Time-Series Prediction Model of Shanghai Composite Index*, Journal of Guilin University of Electronic Technology, Vol.26, No.2, April, 2006

[4] Poterba J.M. and Summers L.H., *The Persistence of volatility and Stock Market Fluctuations*, American Economic Review 76, pp.1143-1151, 1986

[5] French K.R., Schwert G.W. and Stambaugh R.F., *Expected Stock Returns and Volatilit*, Journal of Financial Economics 19, pp.3-29, 1987

[6] Yang Y.W. and Yang C.J., *Financial Time Series Forecasting Based on Support Vector Machine*, System Engineering: Theory, Method and Application, No.14(2), pp.176-181, 2005

[7] Wu W., Chen W.Q. and Liu B., *Prediction of Ups and Downs of Stock Market*

by BP Neural Networks, Journal of Dalian University of Technology, No.41(1), pp.9-12, 2001

[8] Chatfield C., *Time Series Forecasting*, Bocaraton, FL: Chapman & Hall/CRC, 2001

[9] Zhang G. P., Patuwo B.E. and Hu M.Y., *A Simulation Study of Artificial Neural Networks for Nonlinear Time-Series Forecasting*, Computers & Operations Research 28, pp.381-396, 2001

[10] Newbold P. and Granger C.W.J., *Experience with Forecasting Univariate Time Series and Combination of Forecast*, Journal of the Royal Statistical Society Series, Vol.137, pp.131-164, 1974

[11] Aladag C.H., Egrioglu E. and Kadilar C., *Forecasting Nonlinear Time Series with a Hybrid Methodology*, Applied Mathematics Letters 22, 1467-1470, 2009

[12] Xie J.X, Cheng C.T., Chau K.W. and Pei Y.Z., *A Hybrid Adaptive Time-Delay Neural Network Model for Multi-Step-Ahead Prediction of Sunspot Activity*, International Journal of Environment and Pollution, Vol.28, No.3-4, pp.364-381, 2006

[13] Peters E.E., *Fractal Market Analysis: Applying Chao Theory to Investment and Economics*, New York: John Wiely So Sons, pp. 39-50, 1996

[14] Li D., Su X.H. and Ma S.Y., *Stock Price Prediction Algorithms Based on Grey Markov Model*, Journal of Harbin Institute of Technology, 35(2), 2003

[15] Li G.D, Yamaaguchi D. and Nagai M., *The Development of Stock Exchange Simulation Prediction Modeling by a Hybrid Grey Dynamic Model*, The International Journal of Advanced Manufacturing Technology, Vol.36, No.1-2, 2008

[16] Lin C.T. and Yeh H.Y., *The Use of Grey Prediction to Forecast Taiwan Stock Index Option Prices*, Journal of Grey System, Vol.18, No.4, 2006

[17] Cazelles B., Chavez M., Magny G.C.D., Gugan J.F. and Hales S., *Time-Dependent Spectral Analysis of Epidemiological Time-Series with Wavelets*, Journal of the Royal Society Interface, Vol.4, No.15, 2007

[18] Campos-Cantn E. and Murgua J.S., *Wavelet Analysis of Chaotic Time Series*, Revista Mexicana de Fsica, 52(Regular+2), 2006

[19] Popivanov I. and Miller R.J., *Similarity Search Over Time-Series Data Using Wavelets*, 18th International Conference on Data Engineering (ICDE 2002), San Jose, California, USA, 26 February-March, 2002

[20] Yoshinori K. and Shozo T., *Prediction of Stock Trends by Using the Wavelet Transform and the Multi-stage Fuzzy Inference System Optimized by the GA*, IEICE Trams Fundamentals, E83-A(2), pp.357-366, 2000

[21] Taeksoo S. and Ingoo H., *Optimal Signal Multi-resolution by Genetic Algorithms Exchangerate to Support Artificial Neural Network Forecasting*, Expert System with Application (18), pp.257-269, 2000

[22] Ales A., Jonathan C. and Fionn M., *Wavelet Based Feature Extraction and Decomposition Strategies for Financial Forecasting*, Journal of Computational Intelligence in Finance (2), pp.5-12, 2004

[23] http://www.mathworks.com/

[24] http://www.sas.com/

[25] http://ideas.repec.org/a/eee/intfor/v18y2002i3p455-460.html

[26] Brockwell P.J. and Davis R.A., *Introduction to Time Series and Forecasting*, New York, USA: Springer-Verlag New York Inc., 2002

[27] Anderson O.D., *Time Series Analysis and Forecasting*, Butterworth & Co(Publishers)Ltd, 1975

[28] Janacek G. and Swift L., *Time Series Forecasting, Simulations, Applications*, Ellis Horwood Limited, 1993

[29] Pena D., Tiao G.C and Tsay R.S., *A Course in Time Series Analysis*, John Wiely & Sons, Inc., 2001

[30] Madsen H., *Time Series Analysis*, Taylor & Francis Group, LLC, 2008

[31] Brockwell P.J. and Davis R.A., *Introduction to Time Series and Forecasting*, New York, USA: Springer-Verlag New York Inc., pp.186-188, 2002

[32] Deng J.L., *Control Problems of Grey Systems*, Systems and Control Letters, Vol.1, No.5, pp.288-294, 1982

[33] Guo R., *Repairable System Modeling via Grey Differential Equations*, Journal of Grey System, Vol.8, No.1, pp.69-91, 2005

[34] Liu S., Liu Y., *An introduction to Grey Systems: Foundations, Methodology and Applications*, IIGSS Academic Publisher, Slippery Rock, 1998

[35] Deng J.L., *The Foundation of Grey System*, Wuhan: Center China University of science and technology press, 2002

[36] Deng J.L., *Introduction to Grey System Theory*, The Journal of Grey System, Vol.1, No.1, pp.1-24, 1989

[37] Deng J.L., *Grey System Theory Course [M]*, Wuhan: Center China University of science and technology press, 1990

[38] Yu L.Y., *The grey forecast for urban domestic wastes, Environment Pollution and Control*, Environment Pollution and Control, Vol.5, pp.7-9, 1986

[39] Zhang B. and Luo J., *Grey Dynamic model(GM) with Application to Grain Prediction*, System Engineering, Vol.4, pp.36-37, 1985

[40] Daubechies I., *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, 1992

[41] Cohen A. and Ryan R.D., *Wavelets and Multiscale Signal Processing*, Masson, Editeur, Paris, 1995

[42] Chui C.K., *Wavelets: A Mathematical Tool for Signal Analysis*, Society for Industrial and Applied Mathematics, 1997

[43] Aboufadel E. and Schlicker S., *Discovering Wavelets*, John Wiely & Sons Inc, 1999

[44] Ogden R.T., *Essential Wavelets for Statistical Applications and Data Analysis*, Birkhauser, 1997

[45] Amara G., *An Introduction to Wavelets*, IEEE Computational Sciences and Engineering, Vol.2, No.2, Summer , pp.50-61, 1995
(Available at www.amara.com/IEEEwave/IEEEwavelet.html)

[46] Mielniczuk J. and Wojdyho P., *Wavelets for time series analysis - a survey and new results*, Control and Cybernetics, Vol.34, No.4, 2005

[47] MEYER Y., *Wavelets and Operators*, Cambridge University, Cambridge, translated by D.H. Salinger, 1992

[48] MALLAT S., *A Theory for Multiresolution Signal Decomposition*, IEEE Trans. Pattern Anal. Machine Intell. (11), pp.674-693, 1989

[49] Wicherhause M.V., *Acoustic Signal Compression With Wavelet Packets*, In Wavelets: A Tutorial in Theory and Applications, C.K.Chui, Ed.New York: Academic, pp.679-700, 1992

[50] Coifman R., Meyer Y., Quake S. and Wickerhause V., *Signal Processing Compression with Wave Packets*, Numerical Algorithms Research Group, New Haven, CT:Yale University, 1990

[51] Armstrong J. and Collopy F., *Error Messages for Generalizing about Forecasting Method: Emperical Comparisons*, International Journal of Forecasting, Vol.8, pp.69-80, 1992

[52] *http://spotfire.tibco.com/Products/S-Plus-Overview.aspx*

[53] *http://www.spss.com/*

[54] *http://www.r-project.org/*

[55] *http://www.sas.com/products/sasc/index.html*

[56] *http://www.kevinsheppard.com/*

[57] $http://finance.yahoo.com/q?s = gsptse$

[58] $http://en.wikipedia.org/wiki/S26P/TSXCompositeIndex$

[59] Wang F. and Wen B., *A Study of Synthetic Time Series Forecast of Stock Price Fluctuation*, Economic Survey, No.2, 2006

[60] Xu W.W. and Gao F., *Application of Grey System Theory in Stock Price Forecasting*, Computer Simulation 24(11), 2007

[61] Hung K.C., Chien C.Y., Wu, K.J. and Hsu F.Y., *Optimal Alpha Level Setting in GM(1,1) Model Based on Genetic Algorithm*, Journal of Grey System, Vol. 12, No. 1, pp.23-32, 2009

[62] Liu H. and Zhang Q.S., *Life-span Prediction on Mechanical Products of GM(1,1,lamda)*, Journal of Machine Design, Vol.24(10), Oct, 2007

[63] Hao Y. H. and Wang X. M., *The Dynamic Model of Gray System and Its Application to Population Forecasting*, Mathematics in Practice and Theory, Vol.32 No.5, Sep, 2002

[64] Hu X.L. and He Z.W., *Forecasting about Annual Electric Consumption Based on GM-ARMA Combined Model*, Guangdong Electric Power, Vol.20, No.2, Feb, 2007

[65] Deng S.Z., Ru L., Du X.M. and Yang H., *Optimal wavelet packets Best-tree Searching Acquisition Easy Algorithm*, Journal of System Simulation, Vol.19(20), 2007

[66] Ramchandran K. and Vetterli M., *Best wavelet packets Bases Using Rate-Distortion Creteria*, 0-7803-0593-0/92, IEEE, 1992

[67] Liu S.F. and Liu Y., *Grey Information Theory and Practical Applications*, Springer-Verlag London Limited,2006

[68] Shen W., *Theory Survey of Stock Yield Prediction Models*, International Journal of Economics and Finance, Vol.1, No.1, February, 2009